



LUNDS UNIVERSITET
Ekonomihögskolan

Tidsseriemodellering av fyra oreglerade älvars vattenföring

En explorativ studie med GARCH- och Tröskelteknik

Kandidatuppsats, 15 ECTS
Statistiska institutionen, Lunds universitet

Handledare: Peter Gustafsson

Henrik Bengtsson
Dennis Pedersen

Abstract

This paper studies the possibility to model the water flows of four Swedish unregulated rivers between 1980-2015 with ARMA, GARCH and SETAR models. By examining the amount of water that flows through the river and what the dependence in the process looks like in general, a further understanding for effects on for example wildlife and climate changes can be developed. The process of the time series is treated as an additive process, where components of dependence can be eliminated one by one until only white noise remains. The initial distribution of the observations is fairly symmetrical after a logarithmic transformation, although it is not normally distributed, but rather multimodal. The rivers are situated in the northern parts of Sweden, which induces heavy seasonal patterns when the water freezes as well as the ice and snow melts. These seasonal patterns make it problematic to apply the ARMA+GARCH on daily observations, and it is only when monthly observations are used that the seasonality can be eliminated with S-ARIMA. After that, the observed volatility clustering in the variance can be modeled using traditional GARCH. Daily observations can be modeled using SETAR and GARCH since the threshold technique captures the majority of the seasonal patterns.

The conclusive results are that it is possible to eliminate the structural dependence and achieve approximate white noise in the process with the selected models. Though it is difficult to distinguish any difference in the parameter choosing between the four rivers despite large differences in water volume and geographical location. Modeling with monthly observations does not seem to reflect the structure of the water flows in a highly detailed resolution, but rather, it tends to indicate a smoothing effect that generates simple models with white noise. Modeling with daily observations gives a more complex structure, where SETAR in some cases is incapable of reacting to the quick changes in the process. This leaves a desire for a better integrated model to efficiently separate the complex dependence structure in the river water flows.

Keywords: ARMA, GARCH, SETAR, threshold, unregulated rivers, water flows

Sammanfattning

Denna uppsats undersöker möjligheten att modellera fyra svenska oreglerade älvars vattenföring mellan 1980-2015 med hjälp av ARMA-, GARCH- och SETAR-modeller. Genom att studera den mängd vatten som transporteras i vattendrag och hur dess beroendestruktur ser ut i allmänhet kan förståelse om bland annat djurliv och klimataffekter fördjupas. Tidsseriens process behandlas som ett additivt system där komponenter kan elimineras succesivt så att endast vitt brus kvarstår. Grundfördelningen är efter logtransformation relativt symmetrisk. Dock är den inte normalfördelad, utan snarare multimodal. Älvarna är belägna i norra Sverige, vilket medför kraftigt säsongsberoende när vattnet fryser respektive isen smälter. Detta säsongsberoende gör det problematiskt att applicera ARMA+GARCH på dygnsobservationer, och det är enbart när månadsobservationer används som säsongen kan elimineras med S-ARIMA. Därefter kan observerat volatilitetskluster i variansen modelleras med traditionell GARCH. Dygnsobservationer kan modelleras med SETAR och GARCH då tröskeltekniken fångar upp större delen av säsongsberoendet.

Resultatet som framkommer är att det går att eliminera beroendestrukturen och uppnå approximativt vitt brus i processen för valda modeller. Det går dock inte att särskilja någon direkt skillnad i modellval mellan de fyra älvarna trots olik omfattning på vattenföringen och geografiskt läge. Modellering av månadsdata verkar inte återspegla vattenföringens struktur på ett högupplöst sätt, utan snarare observeras en utslätningseffekt och enkla modeller är tillräckligt för att uppnå vitt brus. Modellering av dygnsdata ger en mer komplicerad struktur där SETAR i vissa lägen har svårt att reagera på snabba förändringar i processen, vilket lämnar ett behov av en bättre integrerad modell för att effektivt separera älvarnas komplicerade beroendestruktur.

Nyckelord: ARMA, GARCH, SETAR, tröskeleffekter, oreglerade älvar, vattenföring

Innehåll

1. Inledning.....	1
1.1. Syfte och frågeställning.....	1
1.2. Avgränsning	2
2. Data	3
3. Metod	9
3.1. Grundantaganden	9
3.2. Praktiskt tillvägagångssätt.....	9
3.3. Modeller med ARMA+GARCH.....	10
3.3.1. ARMA.....	11
3.3.2. GARCH.....	11
3.4. Tröskelmodellering	12
3.4.1. Detektion av icke-linjäritet av tröskeltyp	12
3.4.2. SETAR	13
4. Teori	15
4.1. Modeller	15
4.1.1. ARMA.....	15
4.1.2. S-ARIMA	15
4.1.3. GARCH.....	16
4.1.4. TAR och SETAR	17
4.2. Modellvalidering	18
4.2.1. Ljung-Box	18
4.2.2. McLeod-Li	18
4.2.3. Li-Mak.....	18
4.2.4. AIC och BIC.....	19
5. Resultat.....	20
5.1. ARMA+GARCH för månadsobservationer	20
5.2. SETAR och ARMA+GARCH för dygnsobservationer	25
6. Diskussion	32
7. Slutsatser och vidare forskning	37
7.1. Slutsatser	37
7.2. Vidare forskning.....	38
Litteraturförteckning	40

Bilagor	41
Bilaga A - Grundobservationer	42
Bilaga B - Månadsobservationer	44
Bilaga C - Dygnsobservationer för perioden 1980-1985	47
Bilaga D - Modellkoefficienter	48

1. Inledning

Det svenska klimatet har ett obestridligt nyckfullt beteende. Detta avlånga land har stora temperaturskillnader från norr till söder och ständigt skiftande väderförhållanden. Samtidigt påverkar klimatförändringarna vädret över tid. Detta kan ha inneburit, alternativt komma att innebära förändringar i den temperatur och nederbörd som Sverige utsätts för och därmed orsaka förändringar i älvars vattenföring. Med vattenföring menas den mängd vatten i kubikmeter per sekund som rinner fram i ett vattendrag (SMHI, 2009).

Sveriges meteorologiska och hydrologiska institut, SMHI, har till uppgift att bevaka dessa förändringar. En av institutets viktigaste uppgifter är att mäta vattenföring i Sveriges vattendrag och varna om detta uppnår nivåer som kan leda till översvämningar. Orsaker till dessa höga flöden varierar beroende på geografiskt läge i landet. I södra Sverige är skyfall den främsta orsaken, vilket kan uppstå året om. För vattendragen i norra Sverige utgör även där regnet en variationsfaktor, men ihållande kalla vintrar och därmed betydande snöbildning som leder till snö- och issmältning på våren, utgör den största orsaken till skiftande vattenföring (SMHI, 2013).

Att modellera floder och älvars vattenföring är därför viktigt i arbetet med att förstå hur naturen beter sig och påverkar omgivningen. Delvis på grund av risken för översvämning eller torka, men även för att en stor del av djurlivet som till exempel hur lax och örings populationsutveckling påverkas av vattenföringen.

SMHI tillhandahåller mätningar ner till dygnsupplösning för majoriteten av alla svenska älvar och sjöar, där en del mätstationer har observationer daterade till tidigt nittonhundratals (SMHI, 2014). Det skapar intressanta möjligheter att undersöka vattendragens underliggande struktur och egenskaper.

1.1. Syfte och frågeställning

Idag analyserar SMHI flödesobservationerna med hjälp av extremvärdesanalyser, där de beräknar återkomsttiden för en viss händelse (SMHI, 2015). Detta innebär att SMHI fokuserar på de extrema värdena och risken för översvämning. Dock verkar det finnas lite forskning för att förstå fundamentala strukturer i flödesdata över tid, speciellt inom Sverige. Det är därför denna uppsats ämnar att studera vattenföringen från fyra svenska oreglerade älvar med statistiska metoder som möjliggör ett mer holistiskt perspektiv. Målet är att få en bredare bild över hur vattenföringen och dess beroendestruktur beter sig, vilket kan vara intressant ur ett samhällsnyttigt perspektiv.

Då det finns många olika sätt att studera tidsserier har denna uppsats fokuserats mot modellering med klassisk ARMA, betingade varianstekniken GARCH, samt tröskelvärdestekniken SETAR. Dessa väljs delvis för att det bland annat finns en misstanke om att strukturen i älvarnas vattenföring påminner om strukturen för finansiella tidsserier. Det vill säga skiftande varians över tid och volatila rörelser.

Inom finansmodellering är GARCH en särskilt värdefull metod (Cryer & Chan, 2008). Dessutom verkar det finnas tröskel effekter för flödet i älven då till exempel modelleringen kan se annorlunda ut beroende på om marken runt älven är vattenmättad eller inte. Dessa modeller kan förhoppningsvis fånga upp underliggande beroendestrukturer på ett tillfredsställande sätt och generera nya kunskaper och insikter.

Resonemanget i syftet mynnar ut i följande två frågeställningar:

- Hur ser beroendestrukturen ut för en svensk oreglerad älv?
- Kan en svensk oreglerad älv modelleras med ARMA, GARCH eller SETAR?

1.2. Avgränsning

De fyra svenska älvar som undersöks är Sveriges kvarstående stora oreglerade älvar i den bemärkelsen att det vatten som rinner ner i älven inte samlas upp eller stoppas av mänskliga regleringar (Lindh, 1998). Detta görs för att underlätta analysen av vattenföringen och för att få en klar bild av hur dess struktur ser ut i sin renaste form. Dessutom saknas offentlig och detaljerad information över hur vattenföringen i reglerade älvar påverkas av bland annat kraftverk och dämningar.

Observationer för både dygns- och månadsupplösning finns tillgängligt hos SMHI (u.å) och fokus ligger på att undersöka dygnsdata eftersom detta kan ge mest information om vattenföringens struktur. Dessutom är månadsupplösning inte lämplig för tröskeltekniken eftersom de snabba svängningarna i den dagliga vattenföringen inte hinner tas upp vid vissa tidsförskjutningar i modelleringen. Då den snabba ökningen sker under en månad och försvinner igen till nästa månad så medför det att endast en observation skulle finnas i en eventuell högregim vid tröskelmodellering. Detta bör kunna modelleras bättre med ARMA och GARCH då de inte är känsliga på samma sätt mot snabba förändringar.

Datamaterialet analyseras med de klassiska teknikerna för SETAR- och GARCH-modeller. Med det syftas att det finns många olika varianter och tillägg på metoderna som modellerar data på annat sätt, men som inte behandlas i denna uppsats.

2. Data

Det datamaterial som används är hämtat från SMHI:s vattenweb med följande specifikationer i tabell 1 (SMHI, u.å). Avrinningsarea innebär den yta i kvadratkilometer som leder vatten till den givna älven (SMHI, 2013). Älvarna och stationernas placering finns illustrerad i Figur 1.

Tabell 1. Mätstation för respektive älv

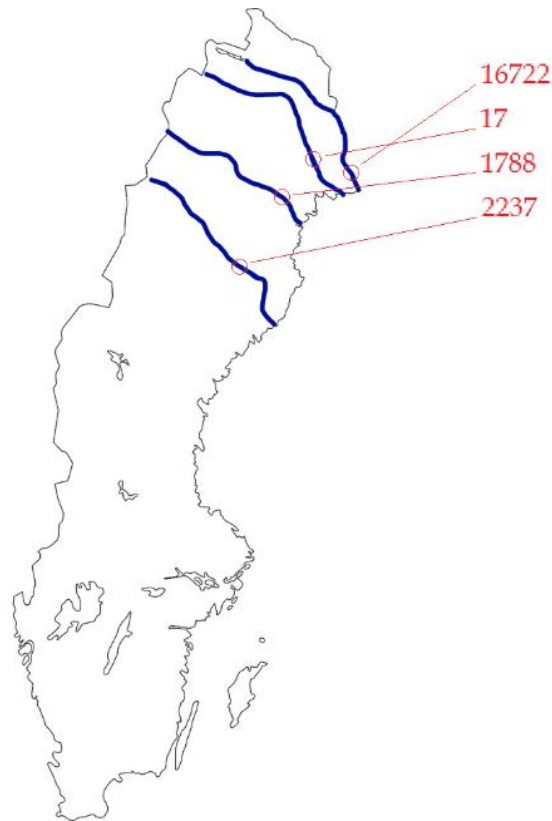
Vattendrag	Stationsnummer	Avrinningsarea [km ²]
Torneälven	16722	33 929,6
Kalixälven	17	23 102,9
Piteälven	1788	10 816,6
Vindelälven	2237	11 846,4

Mätningar som inhämtats är från 1/1-1980 till 31/12-2014, vilket innebär 12784 dygnsobservationer och 420 observationer för månadsdata. Någon hänsyn har inte tagits för de nio skottår som infaller under denna period.

Vindelälven används som exempel i uppsatsen då älvarna generellt liknar varandra till utseende och struktur. På så sätt blir inte läsaren överöst med figurer, och om någon älv avviker kraftigt kommenteras detta i texten samt kompletteras eventuellt med figur.

Inledningsvis exemplifieras tidsseriernas struktur i Figur 2 där det går att se klara säsongeffekter i materialet och det skiljer sig extremt mycket mellan vattenföringens minsta, respektive högsta flöde. Det ser även ut att vara en asymmetri i kurvorna där det sker en gradvis nedtrappning av vattenföringen, jämfört med uppgången som är närmast vertikal. Detta bör rimligen vara på grund av att snö- och ismältning samt sporadiska skyfall leder till en hastig ökning av vattenföringen, vilken skapar en tratteffekt där vattnet inte hinner rinna undan.

I Figur 3 har månadsdata för Vindelälven illustrerats och första bokstaven i varje månad har märkts ut. Som det tydligt går att se har vattenföringen sitt högsta läge mellan maj och juni, vilket borde kunna tolkas som att snö- och ismältningen har skett då. I Vindelälvens lägsta position är månaderna januari till mars representerade vilket kan anses rimligt då mycket vatten är bundet som is och snö under denna period. Detta mönster är liknande för samtliga fyra älvar och är att förvänta då de alla ligger långt norrut där vintrarna generellt är långa och kalla. En annan intressant iakttagelse är att vattenföringen för månadsdata har sitt maxvärde kring 600 m³/sekund, medan dygnsobservationerna nästan når det dubbla för samma period. Detta har givetvis att göra med att månadsdata baseras på medelvärden och att höga flöden inte varar särskilt länge. Slutsatsen av detta är att dygnsdata bäst beskriver älvarnas struktur och att månadsobservationer främst bör ses som en sekundär undersökningskälla.



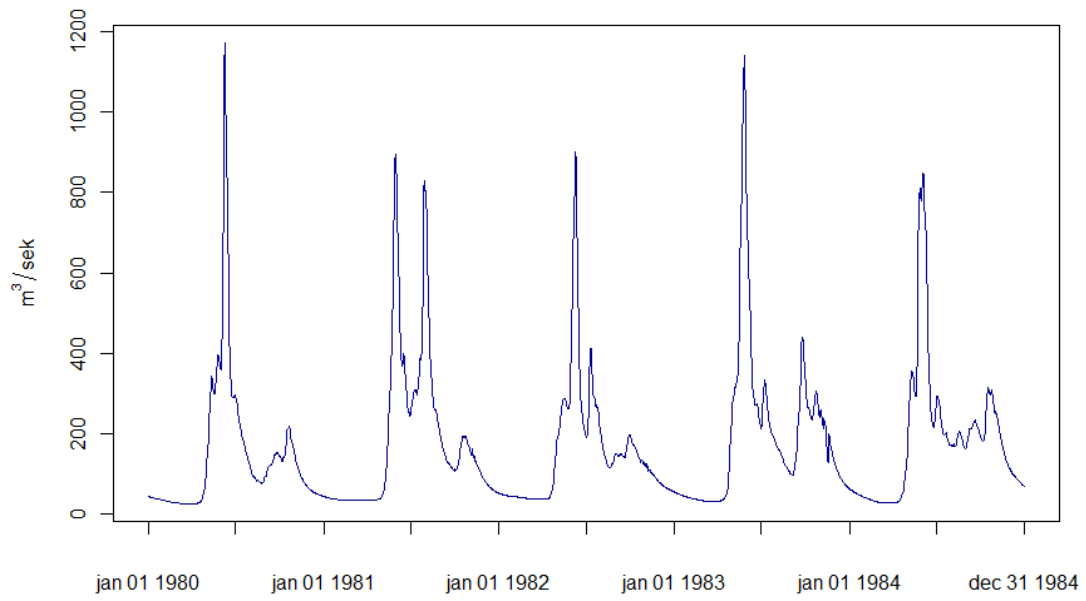
Figur 1. Karta över Sverige och de fyra mätstationerna

Vidare studeras fördelningen av respektive älvs vattenföring i Figur 4. De observerade täthetsfunktionerna har ett fåtal mycket stora värden och många mindre värden. Täthetsfunktionerna skattas genomgående med kärntätheten och bandbredden väljs av algoritmen i *R*. Liknande struktur gäller även för månadsdata men där kurvans rörelse inte är fullt så skarp. För att uppnå en mer symmetrisk fördelning som är lättare att modellera och för att skala ner stora, samt skala upp små värden logaritmeras datamaterialet. På så sätt får även observationer med lägre värde en större innebörd och fördelningens asymmetri förbättras. Figur 5 illustrerar fördelningen efter logaritmering av observationerna och Figur 6 visar tidsseriens beteende efter logaritmering. När materialet redovisas härnäst görs detta med logaritmerade observationer.

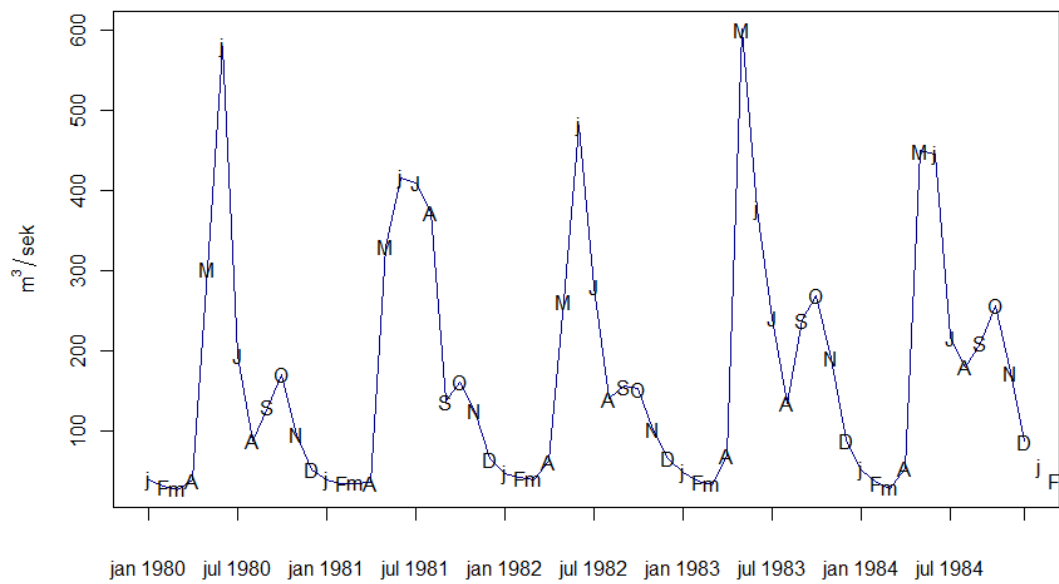
Stationäritet för de transformerade tidsserierna kontrolleras grafiskt och även formellt med Augmented Dickey-Fullers test (ADF) (Cryer & Chan, 2008). Testet har följande hypoteser:

- H_0 – Processen har en entalsrot ($\alpha=1$), det vill säga processen är inte stationär.
- H_1 – Processen har inte en entalsrot ($|\alpha| \neq 1$), processen kan vara stationär.

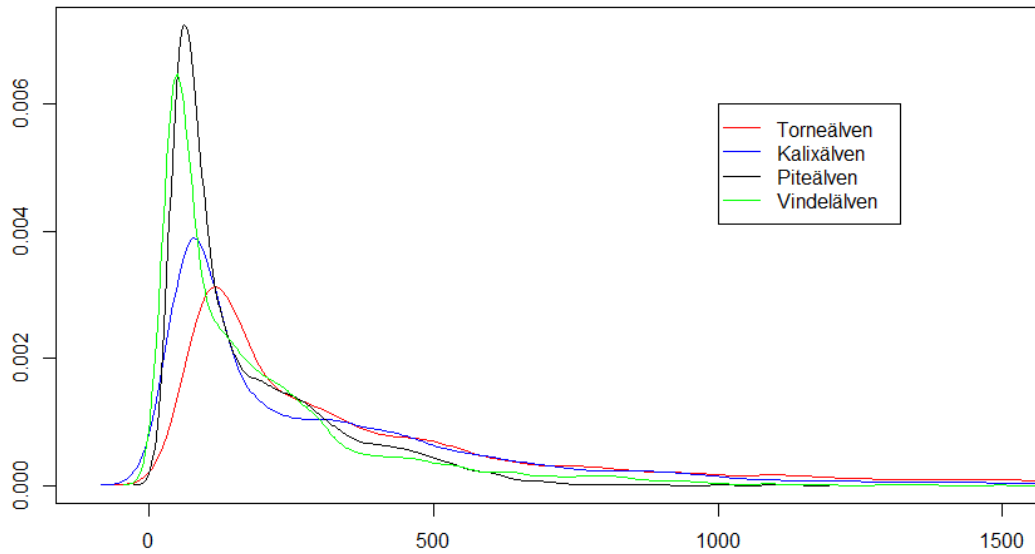
Båda metoderna konstaterar att samtliga fyra älvar och dess tidsperioder kan vara stationära i processen då nollhypotesen kan förkastas. Detta innebär även att materialet inte uppvisar en explosivitet i processvärdet.



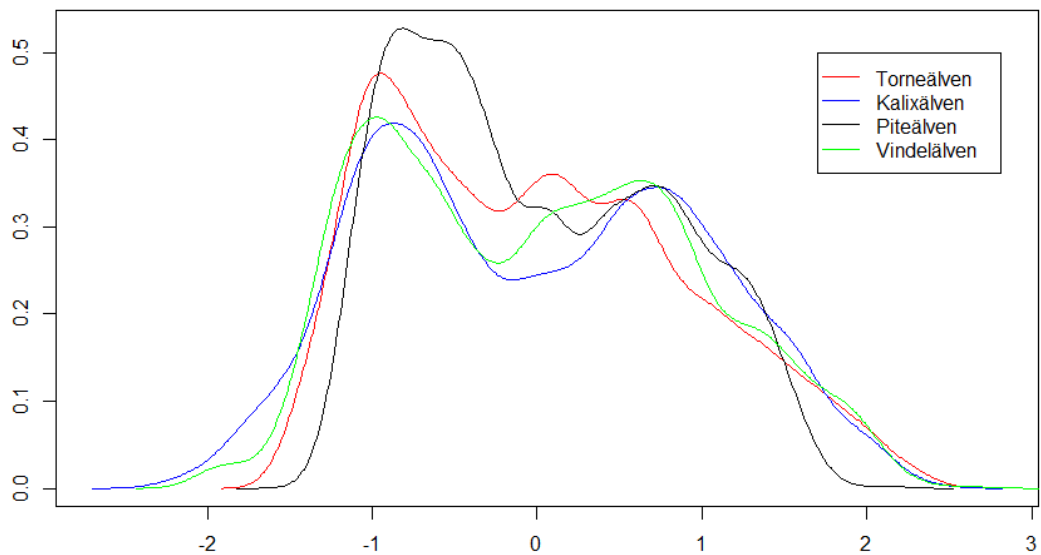
Figur 2. Grundobservationer för daglig vattenföringsmätning av Vindelälven 1980-1985



Figur 3. Illustration av månadsdata för Vindelälven 1980-1985 med månadsnotation

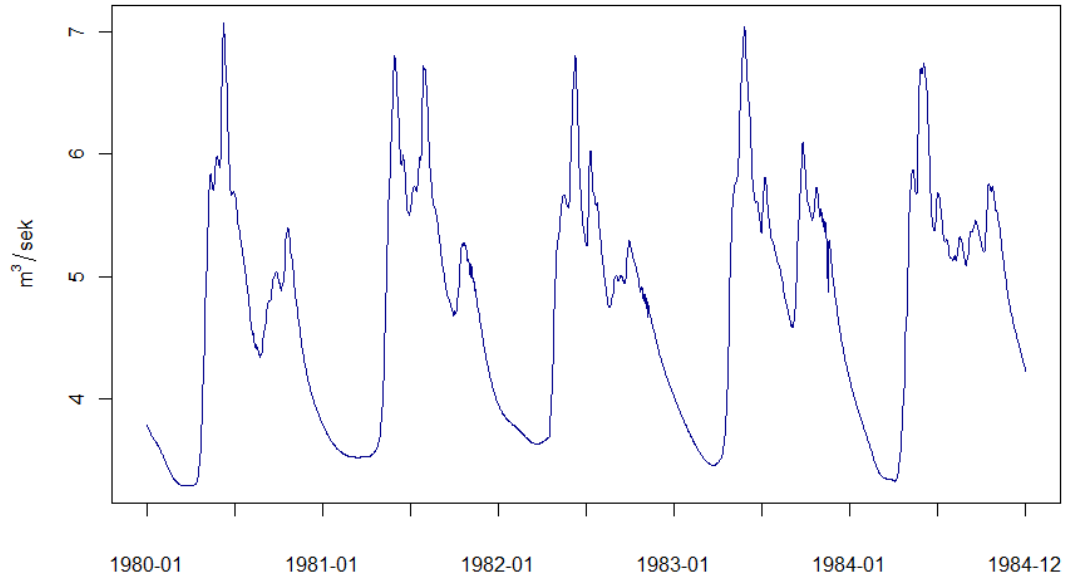


Figur 4. Täthetsfunktioner för de fyra älvarnas dygnsobservationer

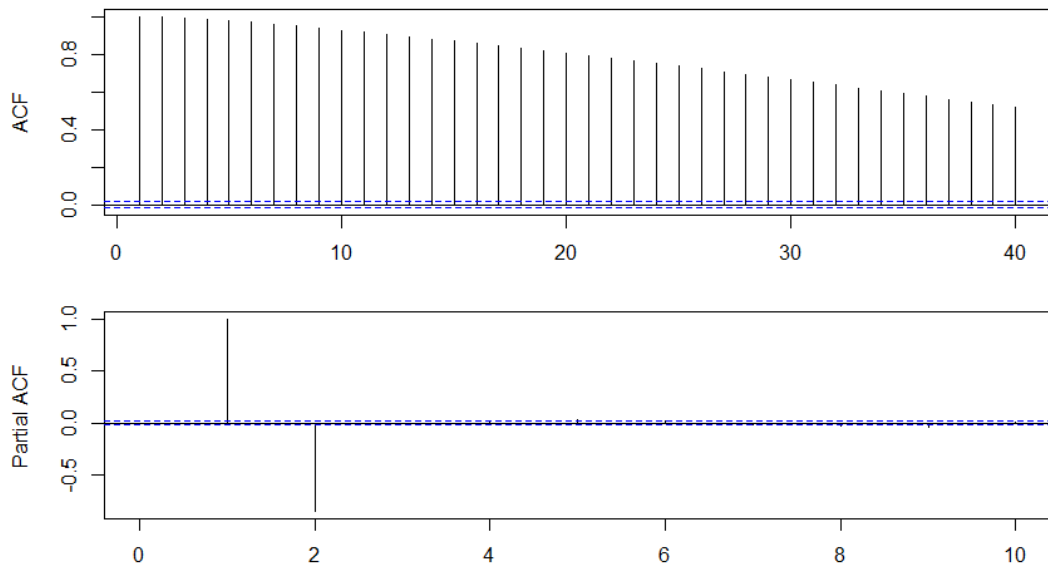


Figur 5. Täthetsfunktioner för de fyra älvarnas logaritmerade dygnsobservationer

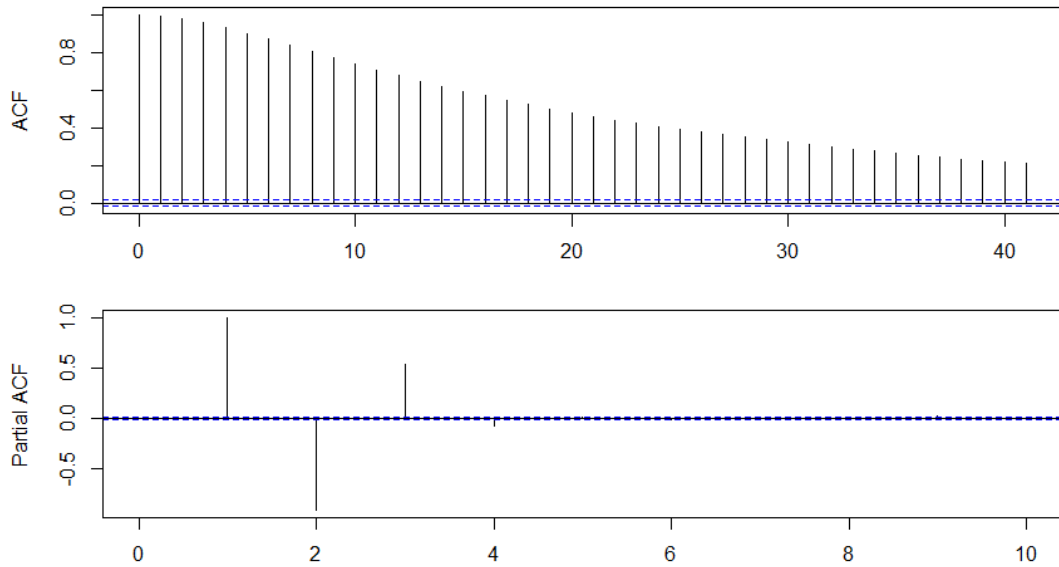
Av ACF och PACF i Figur 7 syns en tydlig beroendestruktur i datamaterialet. Ytterligare beroendestruktur av högre ordning finns även för de kvadrerade observationerna i Figur 8, vilket indikerar på betingad heteroskedasticitet, även uttryckt ARCH, se avsnitt 4.1.3. GARCH. McLeod-Li's test används även för att leta efter ARCH-effekter (McLeod & Li, 1983). Resultatet redovisas i Figur 9, där alla punkter befinner sig under linjen och nollhypotesen att det inte finns ARCH i materialet förkastas därmed. Detta resultat är samma för samtliga älvars tidsperioder.



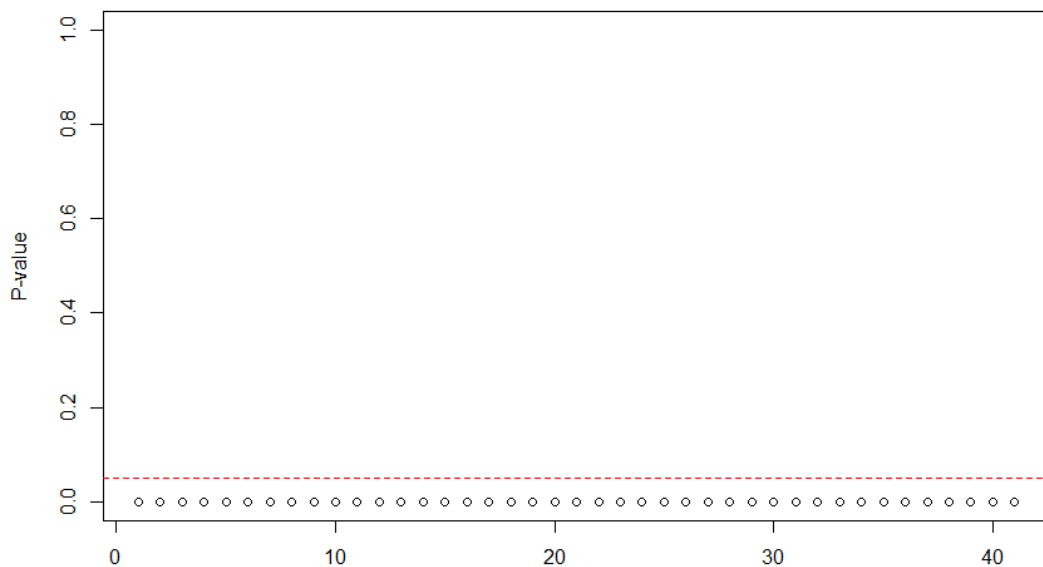
Figur 6. Logaritmerade dygnsobservationer för Vindelälven under perioden 1980-1985



Figur 7. ACF och PACF för samtliga av Vindelälvens dygnsobservationer



Figur 8. ACF och PACF för samtliga Vindelälvens kvadrerade dygnsobservationer



Figur 9. McLeod-Li's test för heteroskedasticitet i observationernas varians

Med denna egenskap stärks motiveringen till logaritmisk transformation av observationerna ytterligare då maximum likelihood-skattningarna som görs i en GARCH-modell ska vara approximativt oberoende av liknande skattningar i ARMA-modellen, vilket inte är fallet om residualerna har en asymmetrisk fördelning (Cryer & Chan, 2008).

Signifikansnivån för arbetet är genomgående 5 procent, men då det rör sig om empirisk data kan även 10 procent vara intressant och i förekommande fall särredovisas detta. Tidsserien för hela datamaterialet, samt exemplifiering av mindre perioder med 1980-1985 finns figurerat i Bilaga A - Grundobservationer.

3. Metod

Tillvägagångssättet som ligger till grund för de resultat som presenteras bygger på ett antal grundantaganden, praktiska tillvägagångssätt, samt olika avvägningar mellan modeller. Dessa presenteras nedan och syftar till att föra in läsaren i vedertaget resonemang. Om man däremot inte är bekant med ARMA, SETAR, GARCH och begreppet icke-linjäritet rekommenderas att kapitel fyra läses innan kapitel tre.

3.1. Grundantaganden

Analysen grundas på ett antagande om att strukturen i processen bygger på additiva egenskaper där det finns sanna koefficienter och ordningar, givet att processen modelleras med de metoder som beskrivs nedan. Detta är en förenkling som görs för att underlätta analysen av datamaterialet och anses därför som nödvändig. Detta medför att superpositionsprincipen kan användas för att identifiera denna struktur. Det vill säga att linjära system kan genom addition brytas ner i mindre delar, samt att multiplikativa komponenter kan brytas ner i additiva komponenter. Detta anses vara ett av de få sätten att demontera en tidsserie i mindre bitar och detaljstudera.

Om de *störande* komponenterna är rensade från processen bör endast en slumpkomponent e_t finnas kvar, där e_t kommer från en viss fördelning, exempelvis normalfördelningen. Slumpkomponenten har då samma egenskaper likt det Cryer och Chan (2008) beskriver som vitt brus.

Tong (1990) menar att detta bör vara möjligt att kombinera så kallade första generationens modeller som exempelvis SETAR eller ARCH. Dessa nya och mer komplexa modeller kallas då för andra generationens modeller, exempelvis SETAR+ARCH. Lägg ytterligare en grundmodell till blir det en tredje generationens modell och så vidare. Dessa flergenerationsmodeller kan fånga upp och modellera komplexa strukturer.

I Figur 10 nedan visas den praktiska implikationen av superpositionsprincipen och antagandet att Y_t kan beskrivas som ett linjärt system.



Figur 10. Processen som ett linjärt, delbart system

Där e_t är det vita bruset med kvarvarande beroendestruktur i variansen. Beroendestrukturen antas vara av ARMA- eller SETAR-typ. I komponenten övrig struktur ingår beroendestrukturer som inte fångas upp av de andra komponenterna.

3.2. Praktiskt tillvägagångssätt

Det angreppssättet som används för att finna innovationstermen är att eliminera de olika komponenterna med målet att det enbart ska återstå ett vitt brus. Därefter kan de uttagna komponenterna studeras och slutsatser dras om det som kännetecknar

den observerade processen. Innovationstermen är det som driver processen och gör den stokastisk, där detta i de vanligaste tidsseriemodellerna enbart är slumpstal från en fördelning. I innovationstermen kan det finnas strukturer i sig, som i exempelvis GARCH-modellen (Cryer & Chan, 2008).

Denna uppsats närmar sig problemet med modellering från två vinklar. Det första är med ARMA-modeller och där den betingade variansen modelleras med GARCH. Det andra angreppssättet är via tröskeltekniken, SETAR, där eventuell kvarvarande beroendestruktur modelleras med ARMA och GARCH. Den grundläggande TAR-modellen förutsätter att gränsvärdena är kända, vilket innebär att det blir problematiskt att avgöra när vattenföringen ska växla regim. Därför används SETAR istället som metod då tröskelvärdet är en endogen variabel.

Analysen grundar sig på den metod som Box-Jenkins tagit fram för att studera modeller (Cryer & Chan, 2008). Den går i korthet ut på att studera om tidsserien är stationär i både medelvärde och beroende, och därefter undersöks eventuellt säsongberoende. Slutligen tillpassas en modell för beroendestrukturen och om modellens residualer är vitt brus anses analysen av tidsserien vara avklarad. Tillägget utöver denna ansats är att beroendestrukturen i den betingade variansen studeras och att ingen differentiering görs om medelvärdet i processen inte är konstant. Istället tillpassas en klassisk linjär regression med tiden som förklarande variabel för att eliminera eventuella icke-stationäritet i medelvärdet.

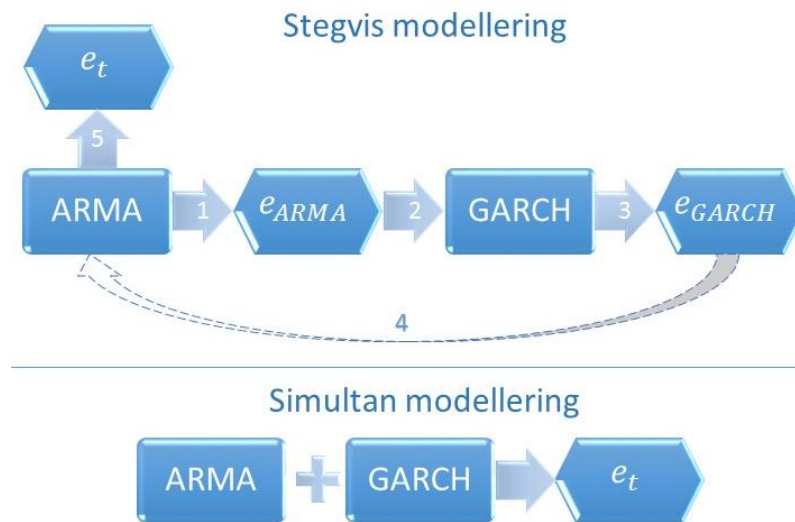
Tre stycken femårsperioder separeras från dygnsobservationerna för att studera om beroendestrukturen förändras över tid. Dessa är 1980-1985, 1995-2000 och 2010-2015. Månadsobservationer delas dock inte upp i perioder av anledningen att antalet observationer för en period blir för få och genererar opålitliga resultat.

3.3. Modellering med ARMA+GARCH

För att modellera en tidsserie behöver hänsyn tas till dess betingade medelvärde genom ARMA-modellering och även dess betingade varians med en GARCH-modell. Till detta används en simultan ARMA+GARCH-modell (Cryer & Chan, 2008). Parametrarna för ARMA-modellen kan skattas genom att studera ACF och PACF. Parameterskattningen för GARCH fås från de kvadrerade residualerna av den anpassade ARMA-modellen (Cryer & Chan, 2008). Slutligen kan en maximum likelihood-skattning för parametrarna i modellen fås genom att modellera en simultan ARMA+GARCH (Enders, 2014).

Det är även möjligt att undersöka tidsseriens betingade väntevärde och varians utan att göra det simultant. Enligt Engle (1982) och hans grundarbete om ARCH går det att först konstruera en ARMA-modell och därefter en GARCH-modell som baseras på residualerna från ARMA-modellen. Tidsserien kan med andra ord modelleras separat utan att tappa asymptotisk effektivitet i skattningen. Förutsättningen för detta är dock att den initiala ARMA-modellen skattas om med residualerna från GARCH för att erhålla parameterskattningar som inte är påverkade av

heteroskedasticitet. Detta arbete har för avsikt att skatta simultan ARMA+GARCH där det är möjligt. Figur 11 illustrerar det ovan beskrivna förloppet.



Figur 11. Illustration av GARCH-modellering

3.3.1. ARMA

För att kunna skatta en ARMA+GARCH till tidsserierna behöver säsongseffekten rensas bort. Detta görs med en säsongsdifferentierad ARMA-modell, kallad S-ARIMA.

Efter modellering kontrolleras residualerna för den säsongrensade modellen. ACF och PACF undersöks för att se om det är vitt brus och att beroendestrukturen har fångats upp. Detta kontrolleras även med Ljung-Box-test som tittar på de individuella laggen och dess signifikans som en grupp (Cryer & Chan, 2008). Residualernas normalfördelning kontrolleras grafiskt via normalfördelningsfigur där observationernas fördelning jämförs med en teoretisk normalfördelning (Chambers, Cleveland, Kleiner, & Tukey, 1983). Även en figur över residualernas täthetsfunktion skapas där fördelningen studeras.

3.3.2. GARCH

För att sedan skatta den betingade variansen behöver datamaterialets övre beroendestruktur studeras. Detta görs initialt genom att titta på ACF och PACF för de kvadrerade residualerna. Där skattas antalet parametrar som GARCH kan tänkas bestå av (Cryer & Chan, 2008). Att materialet innehar någon form av ARCH-effekter har formellt redan konstaterats med McLeod-Li:s test i dataavsnittet.

Olika parameteralternativ för GARCH testas där den med flest signifikanta parametrar och bästa kombination av AIC och BIC väljs ut för vidare analys. Om två modeller har signifikanta parametrar och liknande AIC samt BIC menar Enders (2014) att den med färre antal parametrar bör väljas ut.

De residualer som skapas ur GARCH-paketet är de standardiserade residualerna. Dessa skiljer sig från vanliga residualer i bemärkelsen att varje residual har dividerats med sin skattade och betingade standardavvikelse (Ruppert & Matteson, 2011).

De standardiserade residualerna från den anpassade GARCH-modellen kontrolleras för återstående ARCH-effekter genom ACF och PACF för kvadrerade residualer likt tidigare. Istället för testet McLeod-Li används nu istället testet av Li och Mak (1994) som är ett robust sätt att undersöka den kvarvarande ARCH-effekten i de standardiserade residualerna, se avsnitt 4.2.3. Li-Mak

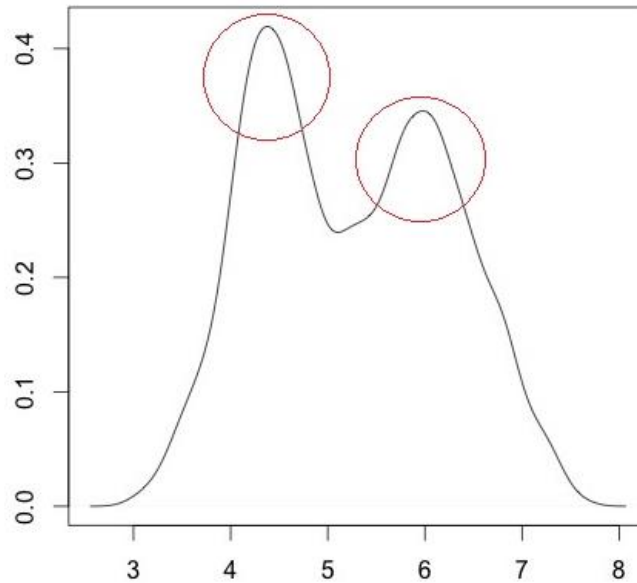
3.4. Tröskelmodellering

3.4.1. Detektion av icke-linjäritet av tröskeltyp

Det finns olika metoder för att detektera icke-linjäritet av tröskeltyp där i huvudsak två test, LR och Tsays test används. Nollhypotesen för LR-testet är att en AR-process observeras, och mothypotesen är att processen är en TAR-modell med homogena och normalfördelade feltermen (Chan & Ripley, 2012). Nollhypotesen för Tsays test är även den en AR-process, samt mothypotesen att processen är icke-linjär. Tsays test har även kritiserats i litteratur av Cryer och Chan (2008) för att inte upptäcka icke-linjäritet av tröskeltyp. LR testet för icke-linjäritet av tröskeltyp är att anse som osäkert då datamaterialets fördelning inte är helt klarlagt, samt att innovationstermens fördelning och homogenitet inte är känd. Detta medför att de inte kan användas som diagnostiska verktyg för aktuellt tidsseriematerial, utan får ses som en extra indikation. Därför utförs en grafisk analys där det dessvärre finns få tillgängliga metoder. Ett sådant tillvägagångssätt måste därför tas fram för att studera icke-linjäritet av tröskeltyp visuellt.

För att göra detta tas en utgångspunkt i vad en icke-linjäritet av tröskeltyp innebär, exempelvis via en SETAR-modell. Det finns en underliggande beroendestruktur som bestämmer processens beteende beroende på vilket värde processen har i en specifik tidpunkt t . Detta bygger på Tongs (1990) tröskelvärdesprincip. När det finns regimkonstanter i SETAR-modellen bör de observerade värdena samlas kring dessa. Därav bör inte en vanlig fördelning observeras utan snarare en multimodal fördelning. Det vill säga en fördelning med fler än en topp i sin sannolikhetsfunktion. Givet att dessa regimkonstanter finns samt att de är tillräckligt skilda från varandra bör SETAR indirekt kunna detekteras när histogram och empiriska fördelningsfunktioner studeras.

Det är på grund av detta som en visuell analys kan göras genom att studera både fördelningen samt den betingade fördelningen av observationerna, där betingningen sker på medianen. Om icke-linjäritet av tröskeltyp finns bör detta enligt resonemanget ovan innebära att fler observationer finns nära de underliggande processnivåerna istället för att vara enhetligt fördelade kring medelvärdet för den observerade processen, se Figur 12.



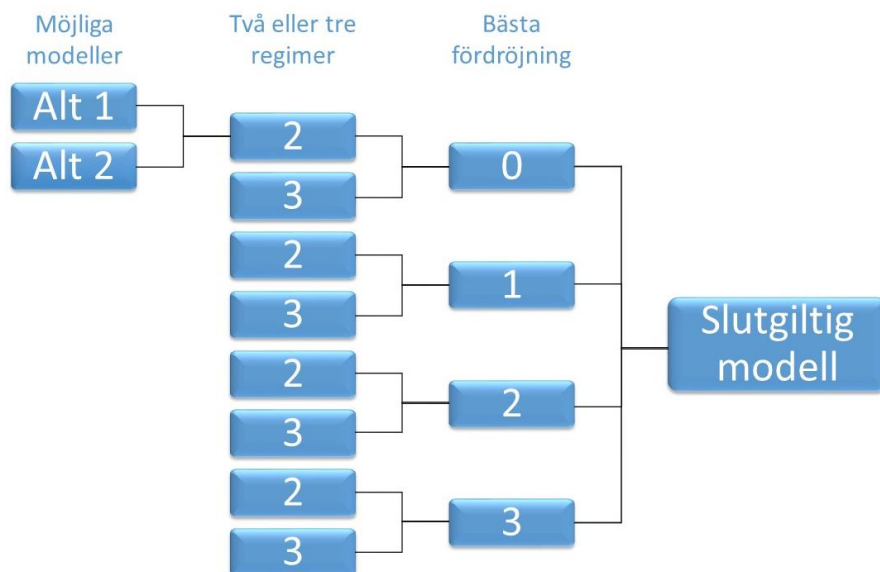
Figur 12. Exempel på multimodal fördelning

Om den observerade processen delas upp genom att betinga på vilken sida medianen är, så kan fördelningarna närmare studeras. För att säkerställa att denna visuella analys inte ger felaktiga resultat har ett stort antal normalfördelade slumpstal genererats i *R* med samma väntevärde och varians som den observerade processen. Detta är ett försök att minimera risken för en felaktig slutsats gällande om det finns två eller fler underliggande processer. För att säkerställa att metoden kan finna SETAR-modeller har simuleringar av olika typer av beroendestruktur studerats. Noterbart är att det finns svårigheter i att upptäcka SETAR-strukturer om regimkonstanterna är mycket nära varandra. Ett annat problem är att avgöra hur många regimer som bör användas för att skapa en korrekt modell, och därför bör denna metod endast användas för att konstatera om det eventuellt finns underliggande processer. Dessa processer bör även kunna observeras vid plot av datamaterialet i enlighet med Tongs (1990) tröskelprincip.

3.4.2. SETAR

Med hänsyn till den stora mängd parametrar i SETAR-modellen som skattas behövs en mer uttänkt angreppssätt användas för att producera en skattad modell. Genomgående används funktionen *setar* (i *tsDyn* paketet) för att identifiera varje modell. Där har algoritmen ändrats så att minst 5 procent av observationerna måste finnas i en regim. Detta görs för att fånga upp de extremregimer som består av de högsta eller lägsta vattenföringsvärdena och som utgör en liten del av den totala mängden observationer. Följande tre steg har använts för att skatta en bra modell. Det första steget är att ta fram en mängd möjliga modeller. Genom successiv elimination skattas ordningen för AR-komponenten i SETAR-modellen givet vissa värden på fördröjningen och antalet regimer. När det inte längre finns ej signifikanta parametrar avslutas eliminationen. I vissa fall skapas en något enklare modell för att se om denna kan ge en bättre tillpassning. Detta har gjorts för samtliga älvar och tidsperioder, innehållandes två och tre regimer, samt ett värde för

fördröjningsparametern i laggen från noll till tre. Detta resonemang illustreras i Figur 13 där olika alternativ för möjliga modeller slutligen mynnar ut i en slutgiltig modell.



Figur 13. Illustration av valet av SETAR-modell

Vid granskning av materialet förefaller en fördröjningsparameter på över tre vara osannolikt, med hänvisning till processens snabba skiftningar i vattenföring. Detta innebär sammanlagt att 256 möjliga modeller skapas. För varje grupp av modeller där minst två är likvärdiga väljs den enklare modellen. I andra steget jämförs de kvarstående 96 modellerna med det lämpliga antalet regimer genom att undersöka två-regimsmodellerna med tre-regimsmodellerna för varje älv och tidsperiod, vilket resulterar i att endast 48 modeller finns kvar efter detta steg.

I det tredje och slutgiltiga steget väljs fördröjningsparametern genom att jämföra de fyra möjliga modellerna för varje älv och tidsperiod. Efter detta steg återstår endast en modell per älv och denna kan ses som det slutgiltiga modellförslaget för hur SETAR-beroendestrukturen kan modelleras.

För att utvärdera vilka modeller som är bättre än andra har signifikanta parametrar, modellens komplexitet, samt AIC och BIC studerats. Dessa har använts för att göra en samlad bedömning. I det sista steget har även modellens residualer utvärderats, och använts som bedömningskriterier. Detta görs för att kunna dra slutsatser om vald metod och söka efter kvarvarande beroende, såsom GARCH.

4. Teori

Tidsserieanalysen studerar sekvenser av tal som genereras av en stokastisk process över tid. Teoretisk drivs de flesta stokastiska processer av det som kallas för vitt brus (Cryer & Chan, 2008). Vitt brus, betecknat e_t , är en sekvens av slumpantal som tas från en fördelning med ett väntevärde och varians, och är oberoende. Dessa bildar en serie som ser ut som följande: e_1, e_2, \dots, e_t . Då korrelationen mellan e_t och e_{t+1} är lika med noll finns det ingen systematisk beroendestruktur i det vita bruset. Det är alltså komponenten e_t som driver den stokastiska processen Y_t oavsett vilken beroendestruktur Y_t har. Följande avsnitt behandlar de olika typerna av modeller, beroendestrukturer och de test som används vid valideringen.

4.1. Modellering

4.1.1. ARMA

De enklaste typerna av stokastiska processer med beroendestrukturer är linjära funktioner av e_t . De enklaste typerna av beroendestrukturer är glidande medelvärdesstrukturer, (benämns här efter som MA med ordningen q) och autoregressiva beroendestrukturer (benämns här efter som AR med ordningen p). Beroendestrukturen för en generell MA(q) skrivs enligt ekvation (1):

$$Y_t = \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t \quad (1)$$

Det vill säga att tidigare slumpantal e_{t-q} vägs in med vikten θ i Y_t . En påfallande egenskap för MA(q)-modeller är att beroendestrukturen upphör helt efter tidslagg q .

AR-modellers beroendestruktur är konstruerad enligt följande ekvation (2):

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t \quad (2)$$

Detta medför att en successivt avtagande beroendestruktur framkommer som inte slutar efter lagg p . MA- och AR-beroendestrukturer kan kombineras till en ARMA(p, q) modell. Denna modell är enligt Cryer och Chan (2008) mycket användbar för att modellera verklig data. Den kombinerar egenskaperna från både MA- och AR-modellerna och har följande form (3):

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t \quad (3)$$

4.1.2. S-ARIMA

För tidsserier där det finns säsongvariationer räcker det ibland inte att enbart anta en deterministisk säsongstrend, utan det krävs en stokastisk säsongmodell för att eliminera autokorrelationen (Cryer & Chan, 2008). Säsongen kan sträcka sig bland annat över kvartal, månader eller till och med dagar.

Det är sällsynt att en tidsserie enbart har autokorrelation vid säsongslagg, utan det förekommer ofta i kombination med autokorrelation för vanliga ARIMA-modeller.

Då behövs en säsonganpassad ARIMA-modell, även kallad S-ARIMA (seasonal ARIMA). Därför antas den multiplikativa säsongmodellen $ARIMA(p, d, q) \times (P, D, Q)_s$, där s betecknar periodintervallet. Den generella modellen definieras som ekvation (4):

$$\phi(B)\Phi(B)\nabla^d\nabla_s^D Y_t = \theta(B)\Theta(B)e_t \quad (4)$$

Här är ∇ en differentiering av trend och B innebär *backshift operator* enligt Cryer och Chan (2008). Skillnaden mellan ARMA och ARIMA är differentiering (I) av processen om denna inte är stationär. Differentiering är inte nödvändigt för modellering av vattenföringen i de fyra älvarna. ARIMA behandlas därför enbart ur en teoretisk aspekt i detta avsnitt.

4.1.3. GARCH

För att modellera en icke konstant varians som en ARMA-modell inte klarar av har Engle (1982) skapat en modell för autoregressiv betingad heteroskedasticitet, även kallad ARCH-modell (autoregressive conditionally heteroskedastic). Detta är särskilt vanligt i finansiella tidsserier där volatilitetskluster uppstår. Med andra ord är den betingade variansen för en observation givet en tidigare observation inte konstant. Följande ekvation (5) beskriver den betingade variansen:

$$\sigma_{t|t-1}^2 = \omega + \alpha_1 y_{t-1}^2 + \dots + \alpha_q y_{t-q}^2 \quad (5)$$

där $y_t = \sigma_{t|t-1} \varepsilon_t$

ω är produkten av en konstant och variansen över tid, α är en okänd parameter, q anger ordning av ARCH och ε_t är oberoende, likafördelade slumpvariabler som kallas innovationstermer.

Denna metod har förfinats av bland annat Bollerslev (1986), där p antal laggar av den betingade variansen har lagts till i modellen. Detta för att kunna mäta längre beroendestrukturer. Denna generella modell kallas för GARCH (general autoregressive conditionally heteroskedastic) och den betingade variansen har följande form:

$$\sigma_{t|t-1}^2 = \omega + \alpha_1 y_{t-1}^2 + \dots + \alpha_q y_{t-q}^2 + \beta_1 \sigma_{t-1|t-2}^2 + \dots + \beta_p \sigma_{t-p|t-p-1}^2 \quad (6)$$

Där p är antalet laggar för den betingade variansen och β är den viktade variansen för det specifika lagget i ekvation (6).

En GARCH-modell kräver att det betingade medelvärdet är konstant, vilket ibland inte är fallet för tidsserier. Därför måste en ARMA(u,v)+GARCH(p,q) modelleras simultant för att ta hänsyn till både betingat väntevärde och varians. Denna process blir då som ekvation (7):

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_n Y_{t-u} + \theta_0 + e_t + \theta_1 e_{t-1} + \dots + \theta_v e_{t-v} \quad (7)$$

Här kan ARMA-modellen skattas med hjälp av processen Y_t och ordningen för GARCH skattas med de kvadrerade residualerna från ARMA-modellen (Cryer & Chan, 2008). GARCH-modellen återfinns i termen e_t i ekvation (7).

Notera här att den tidigare enhetsbenämningen för parametrarna i ARMA har döpts om till u och v vid ARMA+GARCH-modellering.

4.1.4. TAR och SETAR

Autoregressiva tröskelmodeller TAR (threshold autoregressive model) är en vidareutveckling av den autoregressiva modellen. Detta bygger på tröskelvärdesprincipen, vilket innebär att komplexa stokastiska system kan brytas upp i mindre och enklare subsystem eller i underliggande processer (Tong, 1990).

I denna modell införs olika regimer för AR-komponenterna enligt Cryer och Chan (2008), det vill säga att beroendestrukturen förändras beroende på vilket värde processen Y_t har i tidpunkten t . Förändringen i beroendestrukturen sker i form av vilka ordningar och koefficienter som AR-komponenterna har. TAR betecknas TAR(k,p) för att indikera antalet regimer (k) och antalet laggar (p). Det enklaste sättet att beskriva TAR-modeller är att visa ett exempel med en TAR(2,1), det vill säga en modell med två regimer och ett tidslaggar i varje regim. Konstanten c är det tröskelvärde som används för att avgöra vilken beroendestruktur som processen har i den givna tidpunkten t . Ekvation (8) formaliserar exemplet:

$$Y_t = \begin{cases} \varphi_{1,1} * Y_{t-1} + e_t & \text{om } Y_t \geq c \\ \varphi_{2,1} * Y_{t-1} + e_t & \text{om } Y_t < c \end{cases} \quad (8)$$

Konstanten c bestämmer i vilken av regimerna som processen är aktiv. TAR-modellen kan generaliseras för både fler regimer och fler AR-komponenter, där komponenterna inte behöver vara samma för de olika regimerna. En vidareutveckling av TAR-modellen är den autoregressiva modellen av självutvinnande tröskeltyp, SETAR (self exciting threshold autoregressive model) (Zivot & Wang, 2006). Denna liknar till stor del TAR, där skillnaden är att tröskelvärdet c är en exogen variabel i TAR och endogen i SETAR. Dessutom tillkommer även en fördröjningsparameter i SETAR som fördröjer övergången till en annan regim med d stycken tidpunkter.

Zivot och Wang (2006) lägger även in olika konstanter för regimerna, vilket medför att SETAR-modellen skrivs på följande sätt med sina parametrar SETAR(k,p,d) och ser ut som ekvation (9) för en SETAR(3,2,1):

$$Y_t = \begin{cases} \mu_1 + \varphi_{1,1} * Y_{t-1} + \varphi_{1,2} * Y_{t-2} + e_t & \text{om } Y_{t-1} \geq c_2 \\ \mu_2 + \varphi_{2,1} * Y_{t-1} + \varphi_{2,2} * Y_{t-2} + e_t & \text{om } c_1 < Y_{t-1} < c_2 \\ \mu_3 + \varphi_{3,1} * Y_{t-1} + \varphi_{3,2} * Y_{t-2} + e_t & \text{om } Y_{t-1} \leq c_1 \end{cases} \quad (9)$$

För både SETAR och TAR är e_t ett slumpstal från någon fördelning karakteriserad av några parametrar. Samtliga parametrar är positiva heltal och kan skattas med

MK- eller ML-metoder (Cryer & Chan, 2008; Franses & van Dijk, 2000). e_t är således den slumpterm som driver processen likt fallet för den enkla AR-modellen.

4.2. Modellvalidering

4.2.1. Ljung-Box

Detta test undersöker korrelationen för residualer vid olika tidslaggar, där det sammanväger residualerna till ett mått (Cryer & Chan, 2008). Testet är en vidareutveckling av Box-Pierce Q-statistic och används främst för att kontrollera om residualerna från en ARMA-modell är vitt brus.

$$Q = n(n + 2) \sum_{i=1}^n \hat{\rho}_i^2 / (n - i) \quad (10)$$

När n korrelationer har skattats i ekvation (10) kommer frihetsgraderna att minska med antalet skattade koefficienter. Detta innebär att Q är χ^2 -fördelad med $n-p-q$ frihetsgrader enligt Enders (2014) och har följande hypoteser:

H_0 – Ingen korrelation mellan skattade residualer, det vill säga vitt brus.

H_1 – Korrelation mellan skattade residualer, det vill säga beroende kvarstår.

4.2.2. McLeod-Li

McLeod och Li har tagit fram ett test för att kontrollera närvaron av ARCH-effekter i en tidsserie (McLeod & Li, 1983). Testet bygger på Ljung-Box ekvation (10) där kvadrerade residualer används istället för enkla. Residualerna (x_t) kvadreras och differensen $x_t^2 - x_{t-k}^2$ undersöks om korrelationen för något k är positivt enligt följande.

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^N (\hat{x}_t^2 - \hat{\sigma}^2)(\hat{x}_{t-k}^2 - \hat{\sigma}^2)}{\sum_{t=k+1}^N (\hat{x}_t^2 - \hat{\sigma}^2)^2}, \text{ där } k \text{ är antal lagg och } \hat{\sigma}^2 = \sum_{t=1}^N \frac{\hat{x}_t^2}{N} \quad (11)$$

Vid tillräckligt stora stickprov kan det påvisas med hjälp av Ljung-Box-statistikan i ekvation (10) att testet är asymptotiskt χ^2 -fördelad med följande hypoteser:

H_0 – Ingen betingad heteroskedasticitet/ ARCH existerar ej

H_1 – Det finns en betingad heteroskedasticitet/ ARCH existerar

Med andra ord studerar McLeod-Li samma residualer som Ljung-Box, men genom att kvadrera residualerna kan den högre beroendestrukturen testas för vitt brus. Om det finns en icke konstant varians över tid kommer testet ge utslag då $\hat{\rho}_k$ blir stort.

4.2.3. Li-Mak

För att testa om de standardiserade residualerna från en anpassad GARCH-modell har kvarvarande ARCH-effekter finns det så kallade generella test för felspecifikation. Li och Mak (1994) har skapat en asymptotisk fördelning genom

att undersöka de kvadrerade och standardiserade residualerna enligt följande ekvation (12):

$$\hat{C}_k = \frac{1}{n} \sum_{t=k+1}^n (\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-k}^2 - 1) \quad (12)$$

Där korrelationskoefficienten \hat{r}_k för den standardiserade residualen $\hat{\eta}_t$ i kvadrat definieras som ekvation (13):

$$\hat{r}_k = \frac{\sum_{t=k+1}^n (\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-k}^2 - 1)}{\sum_{t=1}^n (\hat{\eta}_t^2 - 1)^2} \quad (13)$$

Om modellen specificeras korrekt kommer teststatistikan få en χ^2 -fördelning med m frihetsgrader och hypoteserna:

H_0 – GARCH-modellen är rätt specificerad

H_1 – GARCH-modellen är fel specificerad

Testet Li-Mak används på grund av att McLeod-Li-statistikan inte är applicerbar i varken det en- eller flerdimensionella fallet. Detta är på grund av att testet inte anses få en asymptotisk χ^2 -fördelning efter GARCH-modellering (Andersen, Davis, Kreiß, & Mikosh, 2009).

4.2.4. AIC och BIC

AIC är framställd för att jämföra olika modeller som baseras på log-likelihoodvärden. AIC står för Akaike's information criterion enligt Cryer och Chan (2008) och beräknas med ekvation (14):

$$AIC = -2 \log(\text{maximum likelihood}) + 2k \quad (14)$$

Där k står för det sammanlagda antalet skattade AR(p) och MA(q) termer. Det är den modell med lägst AIC som anses vara bäst ur informationssynpunkt.

Syftet med $2k$ är enligt Cryer och Chan (2008) att straffa de modeller som har fler parametrar utan att tillföra någon extra information relativt modeller med färre parametrar. Detta kommer medföra att modeller med fler parametrar får ett större AIC.

Ett liknande test som framställts är Bayesian information criterion (BIC):

$$BIC = -2 \log(\text{maximum likelihood}) + k * \log(n) \quad (15)$$

Skillnaden mellan ekvation (14) och (15) är sättet som de olika metoderna straffar större modeller (Cryer & Chan, 2008). BIC multiplicerar antalet parametrar med den logaritmerade effektiva stickprovsstorleken, vilket leder till att mer komplexa modeller med fler parametrar kommer få ett högre värde.

5. Resultat

Resultatet redovisas i enlighet med det tillvägagångssätt som tas upp i metoden där komponenter elimineras stegvis. Månadsobservationernas modellering redovisas i 5.1 och därefter tas dygnsobservationerna upp i 5.2.

Olika modeller har skapats beroende på upplösning. För månadsaterialet används huvudsakligen S-ARIMA och GARCH. Det finns ingen signifikant deterministisk trend för månadsaterialet och därför görs ingen rensning.

Samtliga älvar och tidsperioder för dygnsobservationer uppvisar emellertid signifikanta deterministiska trender ($\beta \neq 0$), med undantag för Vindelälven samt Piteälven 2010-2015. Till skillnad från tillvägagångssättet för månadsaterialet rensas specifikt inte säsongen bort utan anses huvudsakligen rensas bort med hjälp av olika regimkonstanter i SETAR-modellen. Därefter tillpassas en GARCH-modell.

5.1. ARMA+GARCH för månadsobservationer

Som det går att se i dataavsnittet och Figur 6 finns det en klar säsongpåverkan i materialet. Det problem som uppstår vid behandling av stark säsongseffekt är att det på dygnsupplösning av observationerna är svårt att eliminera en signifikant säsong. Även om det teoretiskt är möjligt att undersöka tidsserien med dygnsdata, försvåras analysen av att programvaran har begränsningar och inte stödjer den uppdelning av intervall som behövs för att analysera tidsserien på dygnsupplösning. Vidare finns det enbart en mätning per dygn tillgängligt från SMHI, vilket ytterligare begränsar en eliminering av säsongskomponenter på dygnsupplösning. Dessutom finns det inget uppenbart tillvägagångssätt för att skatta en S-ARIMA+GARCH simultant i *R*, vilket leder till att S-ARIMA-modellen måste skattas först och att analysen av GARCH baseras på dess residualer.

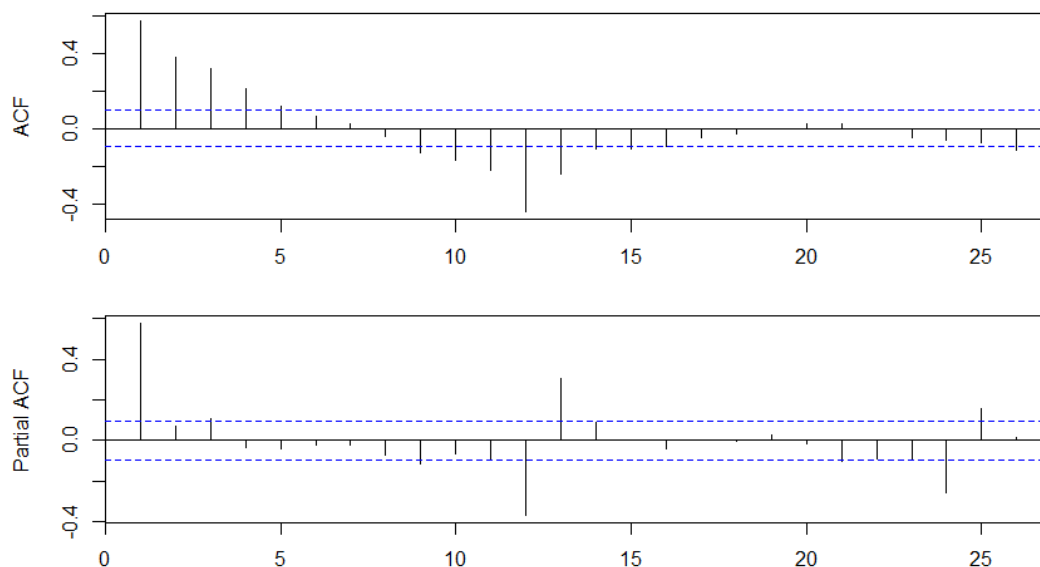
Av de ovan givna anledningarna är det därför olämpligt att anpassa en ARMA+GARCH till tidsserier där det finns markanta säsongeffekter i beroendestrukturen som inte kan rensas bort vid dygnsupplösning. Angreppssättet med ARMA+GARCH överges därför som lösning till huvudsyftet att modellera älvarna med dygnsobservationer och denna kraftansträngning fokuseras istället mot tröskeltekniken.

Det är däremot meningsfullt att undersöka om säsongeffekten kan tas i hänsyn för månadsupplösning och därefter anpassa en ARMA och GARCH. Då kan en modell med detta angreppssätt skapas för älvarna, även om det inte är på önskad observationsupplösning.

Analysen inleds med en kontroll av månadsobservationernas ACF och PACF. Vindelälven redovisas i Figur 14, där PACF:en indikerar på en AR(1)-komponent och ACF:en visar att det eventuellt kan röra sig om en MA(1).

Därefter görs en säsongrensning av datamaterialet med hjälp av S-ARIMA. Tabell 2 redovisar parameterintervalen för de fyra vattendragen. De modeller som anges

är de alternativ med lägst AIC, BIC och med signifikanta parametrar. Som tabellen visar är modellskattningen samma för älvarna med undantag för Kalixälven, där ytterligare en MA-komponent krävs. Parameterkoefficienter återfinns i Bilaga B - Månadsobservationer.



Figur 14. ACF och PACF för Vindelälven innan tillpassning av S-ARIMA

Tabell 2. Lämpliga modeller med signifikanta parametrar lägst AIC samt BIC

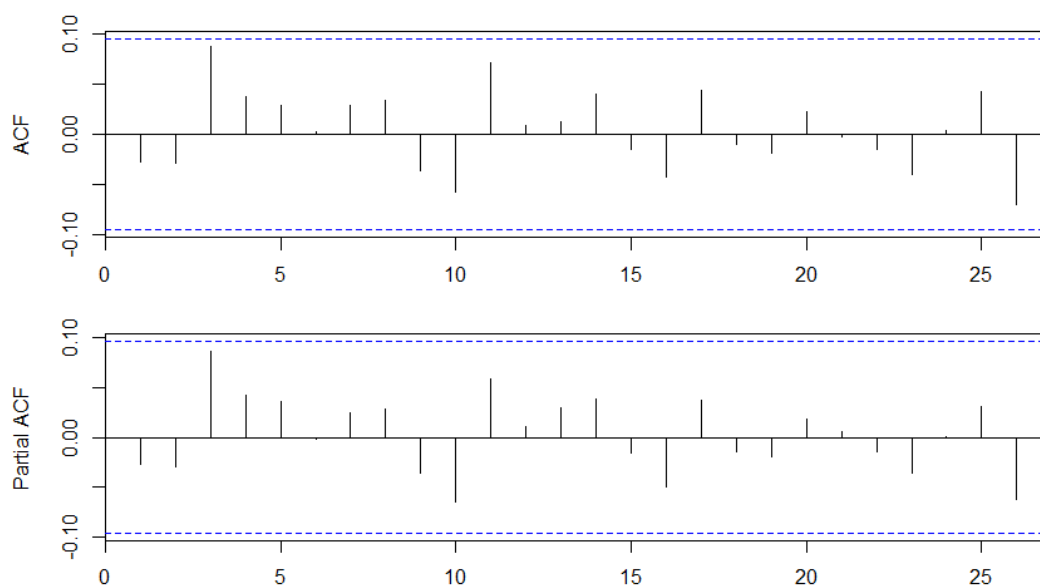
Vattendrag	S-ARIMA	AIC	BIC
Torneälven	$(1,0,0) \times (0,1,1)_{12}$	238,10	252,22
Kalixälven	$(1,0,1) \times (0,1,1)_{12}$	177,61	195,78
Piteälven	$(1,0,0) \times (0,1,1)_{12}$	73,85	87,97
Vindelälven	$(1,0,0) \times (0,1,1)_{12}$	227,37	241,50

Vidare undersöks modellvalidering där vitt brus och normalfördelning kontrolleras.

Den ACF och PACF som illustreras i Figur 15 visar att det inte finns några signifikanta laggar kvar i Vindelälvens anpassade S-ARIMA-modell. Således kan det grafiskt konstateras att beroendestrukturen är vitt brus.

Ljung-Box testar det vita bruset formellt och är beskrivet i tabell 3. För samtliga älvar kan nollhypotesen om ingen autokorrelation ej förkastas, vilket innebär att deras beroendestruktur kan antas vara vitt brus. I samma tabell finns även Shapiro-Wilks normalitetstest med resultatet att ingen älv har normalfördelad månadsdata, vilket konstaterats i dataavsnittet. Att de fyra älvarnas residualer efter säsonganpassning inte är normalfördelat konstateras även med en normalfördelningsfigur och en illustration av täthetsfunktionen som återfinns i Bilaga B - Månadsobservationer. Figurerna ser relativt raka ut men de uppvisar

avvikande *tunga* svansar, vilket är karaktäristiskt för data med ARCH-effekter (Cryer & Chan, 2008).



Figur 15. ACF och PACF för Vindelälven efter tillpassning av S-ARIMA

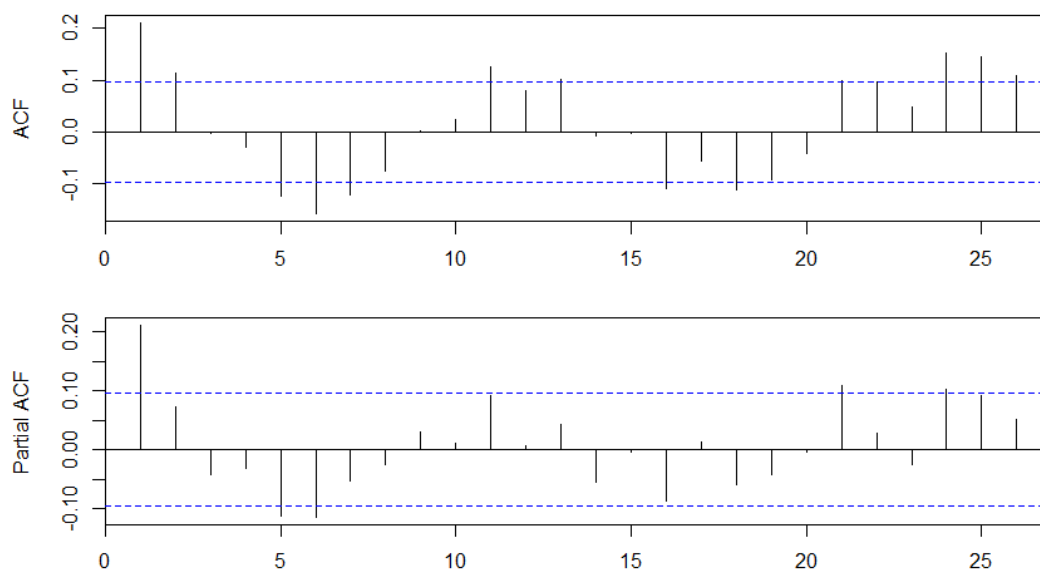
Tabell 3. Ljung-Box-test och Shapiro-Wilks p-värde för älvarna

Vattendrag	Ljung-Box [p-värde]	Shapiro-Wilks [p-värde]
Torneälven	0,2475	9,019e-07
Kalixälven	0,1226	2,644e-08
Piteälven	0,1066	2,112 e-03
Vindelälven	0,4657	2,609e-05

Därefter kontrolleras ACF och PACF för Vindelälvens kvadrerade residualer i Figur 16. Det verkar vara lämpligt med en ARCH(1), alternativt en GARCH(1,1). I tabell 4 redovisas de GARCH-modeller som har anpassats till vattendragen där enklast möjliga modeller med lägst AIC, BIC och signifikanta parametrar är bestämmande faktorer. Viktigt att notera är att R-paketet *rugarch* dividerar AIC och BIC med antalet observationer för att normalisera testet. Detta gör att skillnaderna kan verka små mellan modellerna, men när värdet multipliceras tillbaka med antalet observationer blir skillnaden desto större (Ruppert & Matteson, 2011).

ML-skattningen för GARCH-parametrarna grundas initialt på normalfördelning, men bättre resultat kan uppnås genom att justera till en fördelning som avspeglar materialet i högre grad. Det finns dock ett begränsat antal fördelningar tillgängligt i *rugarchpaketet*, vilket gör att det inte är fullt möjligt att finna den optimala fördelningen. Observationernas fördelning är inte heller fastställd. Student-t-fördelningen visar sig dock passa GARCH-modellerna mycket bättre relativt normalfördelningen och väljs därför som föredragen fördelning. I Figur 17

poängteras skillnaden mellan att använda student-t-fördelning framför normalfördelning vid beräkning av GARCH-parametrar för Vindelälven. De tunga svansarna fångas upp på ett mycket bättre sätt i höger QQ-plot för student-t. Dessutom förbättras värdet för AIC,BIC och ACF:en får färre signifikanta laggar.



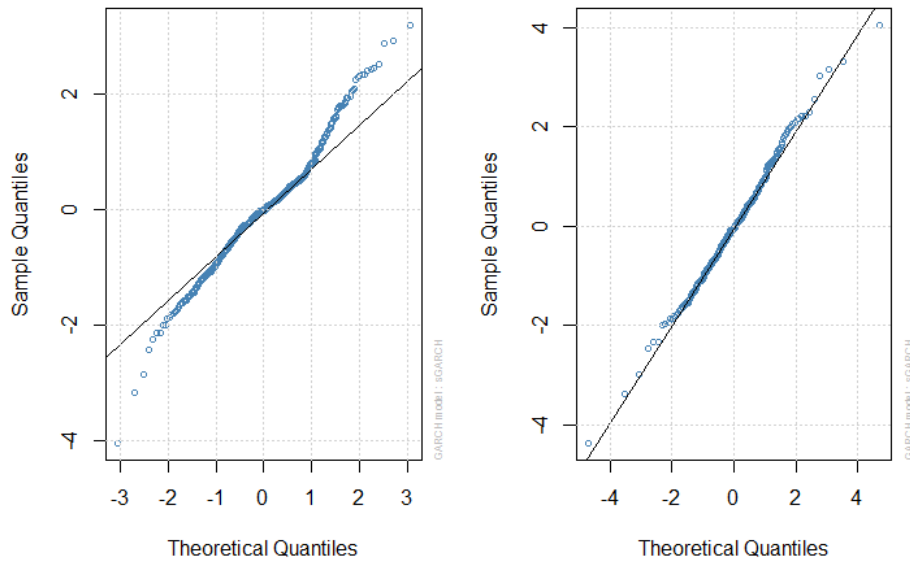
Figur 16. ACF och PACF för kvadrerade residualer från Vindelälven

Li-Mak-testet kan inte förkasta nollhypotesen efter tillpassning av GARCH och det går därför inte att urskilja mer GARCH-effekter i materialet. De standardiserade residualerna för Vindelälven illustreras i Figur 18 och Figur 19, där det ser ut att vara approximativt vitt brus i vanlig och högre beroendestruktur. De tre övriga älvarna har analyserats på samma sätt och motsvarande figurer samt koefficienter återfinns i Bilaga B - Månadsobservationer.

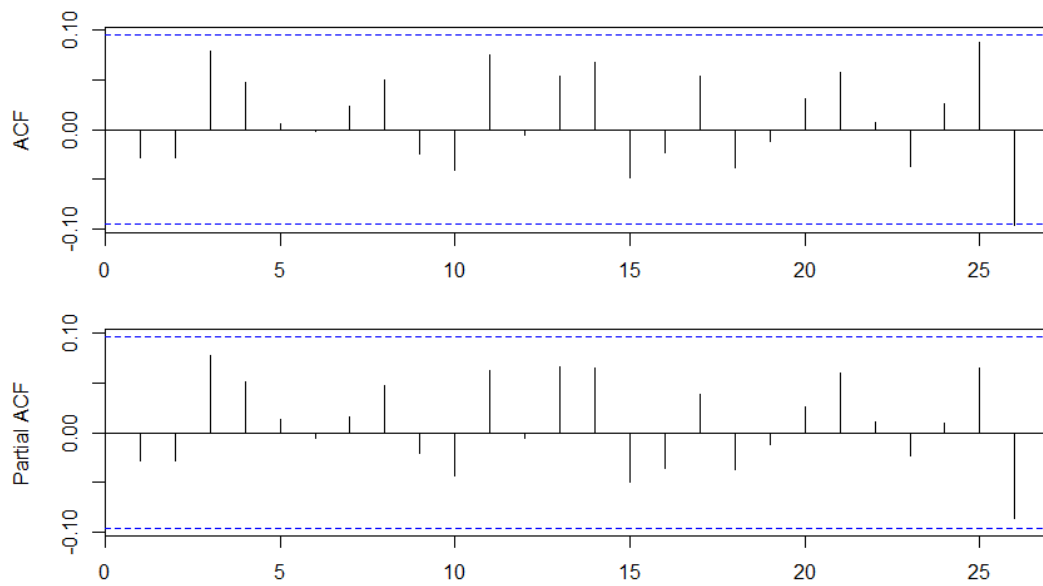
Slutligen kontrolleras S-ARIMA-modellen igen, där det för Vindelälven nu enbart består av approximativt vitt brus. Det kan således tolkas som att den initiala skattningen av parametrar har varit tillräcklig och poängen är visad; Det går att modellera älvarnas månadsobservationer med S-ARIMA och GARCH.

Tabell 4. Valda GARCH-modeller för månadsobservationer

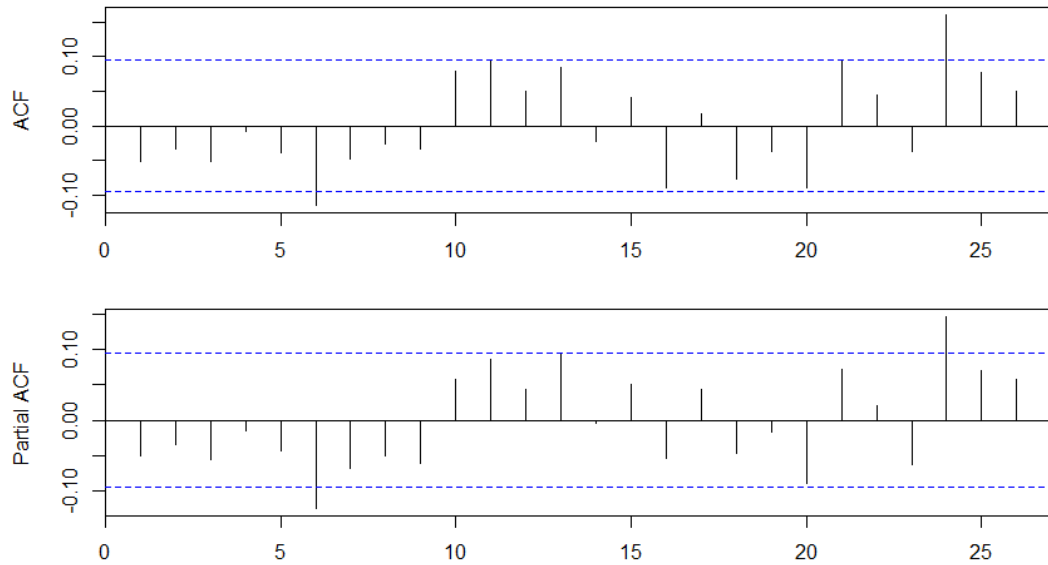
Vattendrag	Modell	AIC	BIC	Li-Mak [p-värde]
Torneälven	ARCH(1)	0,3409	0,3698	0,2924
Kalixälven	ARCH(1)	0,2167	0,2456	0,4428
Piteälven	ARCH(1)	0,0409	0,0698	0,8832
Vindelälven	ARCH(2)	0,3268	0,3653	0,3406



Figur 17. Skillnaden mellan att anpassa GARCH-modellen till normalfördelning (vänster), kontra student-t (höger)



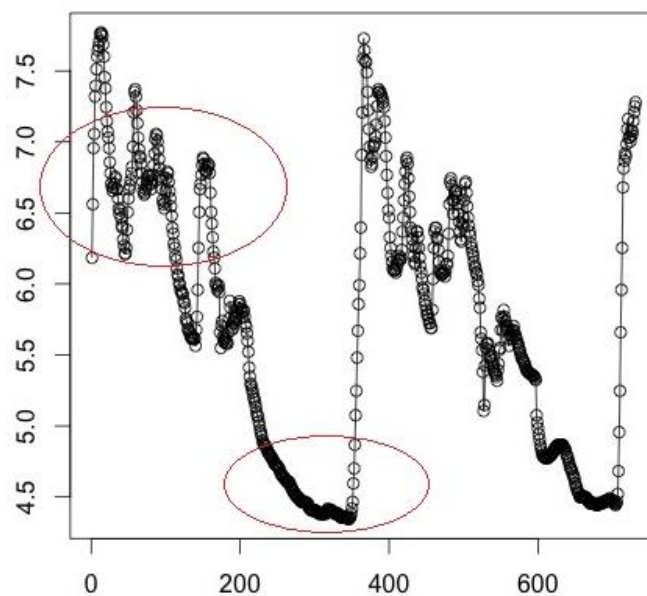
Figur 18. ACF och PACF för Vindelälvens standardiserade residualer efter GARCH



Figur 19. ACF och PACF för Vindelälvens kvadrerade observationer efter anpassning av GARCH

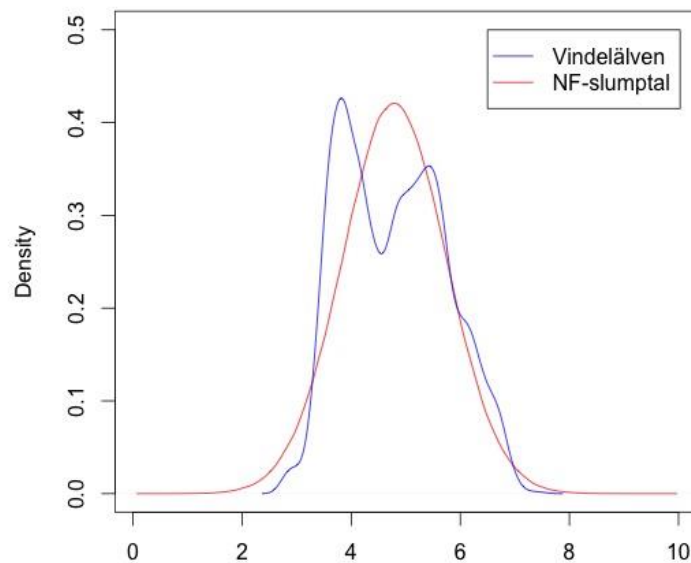
5.2. SETAR och ARMA+GARCH för dygnsobservationer

För att studera icke-linjäritet av tröskeltyp används tillvägagångssättet beskrivet i avsnitt 3.4.1. Initialt studeras datamaterialet där bland annat de individuella observationerna ritas ut. Då syns snabba uppgångar från perioder där värden ligger nära föregående värden till perioder som verkar vara mer instabila men där värden ligger i paritet med föregående värde. Det verkar alltså finnas två eller flera underliggande processer. Detta stöder användandet av en SETAR-modell enligt Tongs (1990) tröskelvärdesprincip. Det illustreras i Figur 20, där de röda ringarna pekar på vad som skulle kunna vara återkommande underliggande processer.

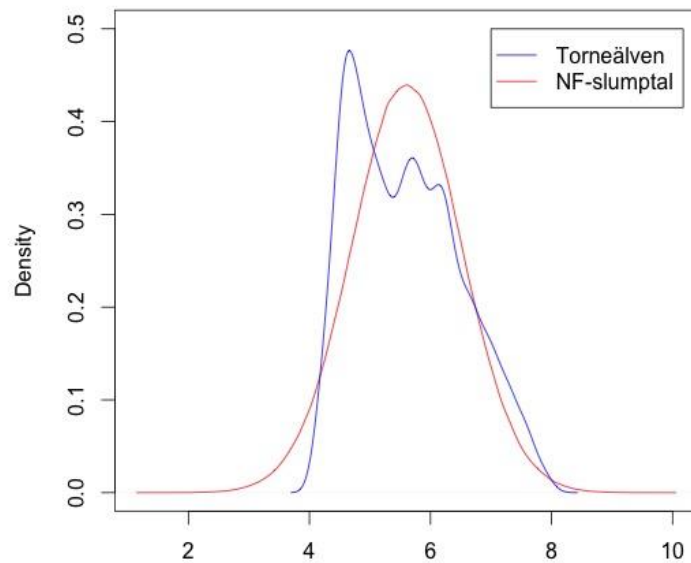


Figur 20. Exempel på underliggande processer

För Vindelälven och Kalixälven syns två tydliga *toppar* illustrerat i Figur 21 och kan indikera på att icke-linjäritet av tröskeltyp finns. Däremot uppvisar till exempel Torneälven något som mer kan liknas vid en plåtå i Figur 22. Det går vid undersökning av de betingade täthetsfunktionerna att konstatera olika processkonstanter. När de empiriska täthetsfunktionerna ritas upp mot en normalfördelning med samma väntevärde och varians blir skillnaderna tydliga och liknande struktur uppnås för de andra älvarnas och dess olika tidsperioder. Både LR testet och Tsays test förkastar sina respektive nollhypoteser, vilket som nämnt tidigare inte är helt tillförlitligt för befintligt datamaterial.



Figur 21. Skillnaden mellan en normalfördelning och den empiriska fördelningen för Vindelälvens samtliga observationer



Figur 22. Skillnaden mellan en normalfördelning och den empiriska fördelningen för Torneälvens samtliga observationer

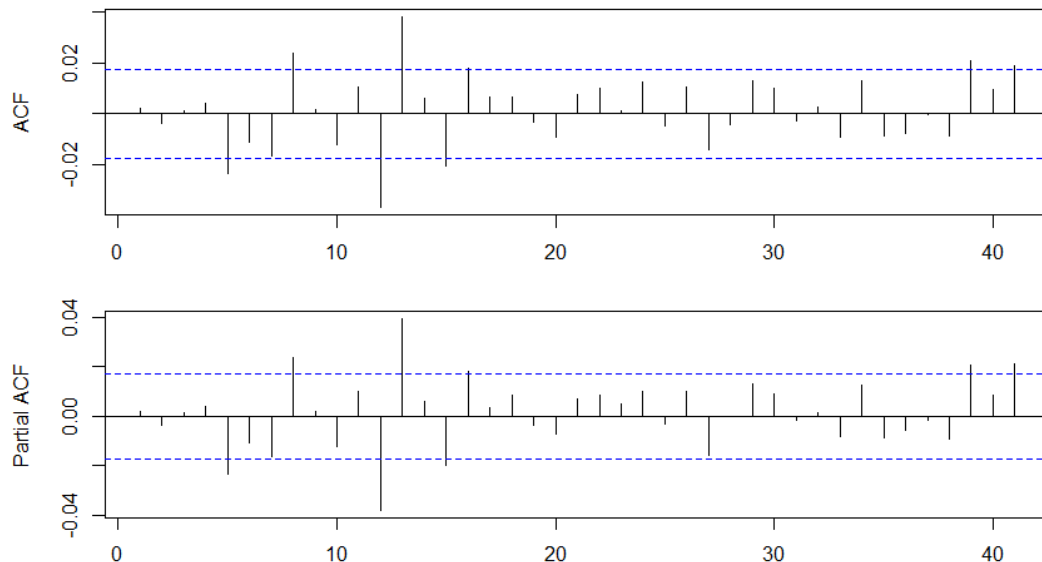
Det kan konstateras att det finns SETAR-effekter i datamaterialet då en tydlig dubbeltoppighet syns. Därför tillpassas SETAR-modeller för dygnsobservationer. En sammanställning av de valda modellerna finns nedan i tabell 5. Här återges huvudsakligen modellordningarna, medan koefficienter finns redovisade i Bilaga D - Modellkoefficienter.

Tabell 5. Valda SETAR-modeller för dygnsobservationer

Älv/Period	Antal regimer och AR-parametrar i respektive regim					Antal observationer i respektive regim [%]			Ljung-Box [p-värde]
	Regimer	Hög	Mellan	Låg	Fördröjning	Hög	Mellan	Låg	
1980-2015									
Torneälven	3	4	5	3	2	29	60	12	0,8699
Kalixälven	2	8	-	4	1	59	-	41	0,6968
Piteälven	2	3	-	2	0	45	-	55	0,8850
Vindelälven	3	4	4	5	1	25	33	42	0,8120
1980-1985									
Torneälven	3	2	3	4	3	48	46	5	0,6535
Kalixälven	2	4	-	2	2	57	-	43	0,4808
Piteälven	2	6	-	1	0	62	-	38	0,5277
Vindelälven	2	4	-	4	0	23	-	77	0,5354
1995-2000									
Torneälven	2	3	-	4	0	54	-	46	0,8726
Kalixälven	2	3	-	4	0	40	-	60	0,8638
Piteälven	2	2	-	3	1	41	-	59	0,3236
Vindelälven	2	4	-	5	0	75	-	25	0,8569
2010-2015									
Torneälven	2	4	-	3	1	27	-	73	0,8829
Kalixälven	2	4	-	4	1	22	-	78	0,8088
Piteälven	2	3	-	4	2	46	-	54	0,8193
Vindelälven	2	4	-	5	0	31	-	69	0,9910

Som tabellen visar är det framför allt två-regimmodeller som har valts för att studera datamaterialet. Samtliga AR-koefficienter är också signifikant skilda från noll på minst 5 procentsnivån, med undantag för en AR-koefficient i Piteälven 1980-1985. Fördröjningsparametern verkar vara relativt låg då endast en av sexton modeller har en fördröjningsparameter på tre. Istället för att presentera ett tröskelvärde ges antalet observationer i varje regim som ett procenttal. Detta är för att tydliggöra för läsaren var tröskelvärdena befinner sig om observationerna sorteras från minst till störst. Samtliga tröskelvärden är listade i Bilaga D - Modellkoefficienter.

För att studera hur bra dessa modeller är på att fånga upp beroendestrukturer i materialet studeras ACF och PACF för de olika älvarna och tidsperioderna. Beroendestrukturen liknar huvudsakligen vitt brus. Detta kan illustreras med Vindelälven 1980-2015 i Figur 21. ACF och PACF för tidsperioden 1980-1985 återfinns i Bilaga C - Dygnsobservationer för perioden 1980-1985.



Figur 21. ACF och PACF för Vindelälvens hela tidsintervall efter tillpassning av en SETAR-modell

Som det går att se i Figur 21 finns det några kvarstående signifikanta laggar. Dock verkar de existera i både ACF:en och PACF:en, vilket medför att de troligtvis inte är orsakade av AR- eller MA-beroendestrukturer. Den allmänna granskningen av ACF och PACF för älvarna indikerar att en större del av beroendestrukturen som fanns i det ursprungliga materialet är borta. De fåtal kvarvarande signifikanta laggen är mycket små och kan anses vara försumbara i detta empiriska datamaterial. Ljung-Box testar för vitt brus och återfinns i tabell 5 ovan. Samtliga älvars tidsperioder förkastar nollhypotesen och indikerar på vitt brus.

När det gäller konstanter är alla inte signifikant skilda från noll. Av 35 konstanter är 16 stycken signifikanta på minst femprocentsnivån, 13 är inte signifikanta och 6 är signifikanta på tioprocentnivån. En lista med vilka som är signifikanta finns i Bilaga D - Modellkoefficienter. Det är framförallt i datamaterialet från 1980-2015 som regimkonstanterna är signifikanta. De är också till stor del det i materialet från 2010-2015, vilket kan ställas i kontrast till de regimkonstanterna från 1980-1985 och 1995-2000 som verkar vara i mindre grad signifikanta.

Residualerna efter SETAR-anpassning innehåller GARCH-effekter, vilket kräver ytterligare modellering för att nå vitt brus i högre beroendestrukturer. Modelleringen av GARCH redovisas i tabell 6.

Som nämnt är residualerna innan modelleringen med GARCH approximativt vitt brus för vanlig ARMA-struktur. Däremot är de standardiserade residualerna efter

GARCH, där innovationsbruset divideras med respektive betingad standardavvikelse, inte vitt brus. För att åtgärda detta krävs en ARMA+GARCH-modell för vissa av perioderna.

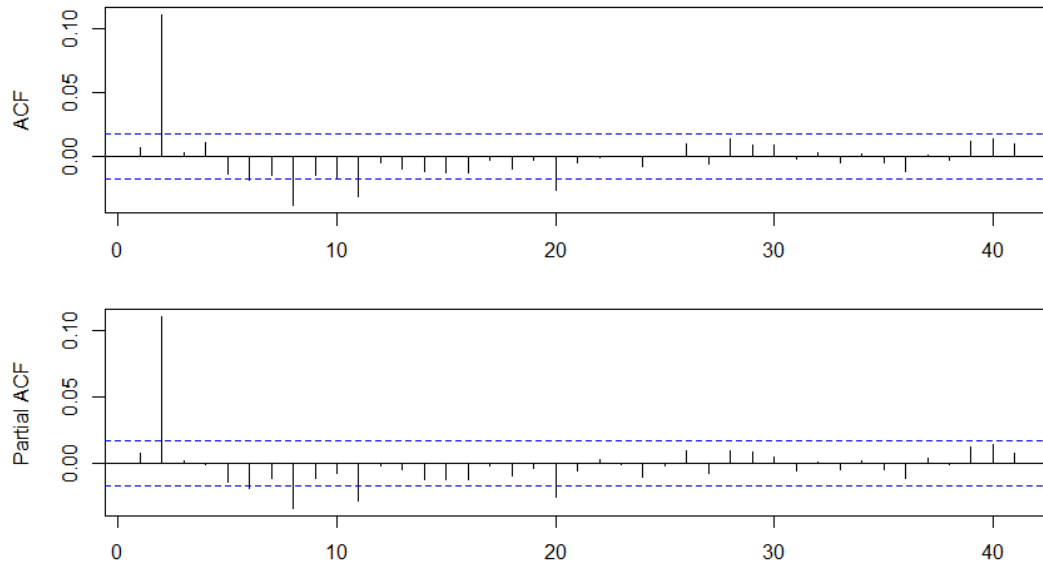
Även här används student-t-fördelningen, där den i de flesta fall fångar upp observationernas fördelning väl och ger en bättre QQ-plot relativt normalfördelningen. För bland annat Vindelälvens material i sin helhet finns det dock en kvarvarande struktur i residualerna efter GARCH-anpassning som orsakar tunga svansar i fördelningen. Dessutom finns det enstaka signifikanta laggar vid analys av ACF och PACF i Figur 22 och Figur 23 för vanlig och högre beroendestruktur över tid.

Ett annat resultat värt att påpeka är att femårsperioden 1980-1985 har en generellt sett enklare modellstruktur. Det är enbart i Vindelälven som ARMA-parametrar behöver läggas till för att uppnå vitt brus. Dessutom är Torneälvens observationer för 1980-1985 nästan helt fria från GARCH-effekter, vilket är intressant med tanke på de andra periodernas skarpa signifikans i högre beroenden. Det skiljer sig generellt lite i modelleringen mellan de tre resterande tidsperioderna och är svårt att hävda någon egentlig skillnad.

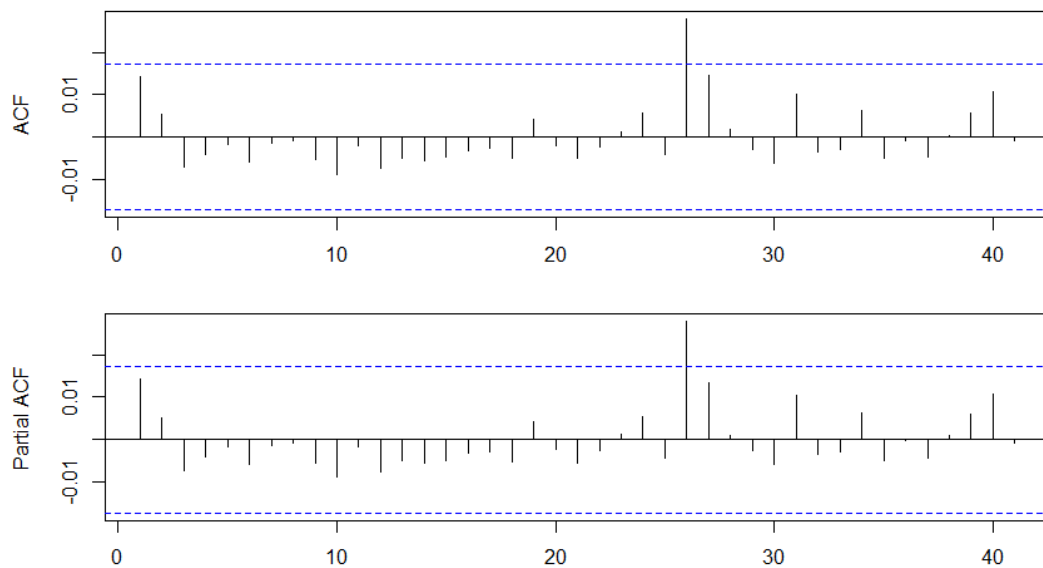
Figur 24 illustrerar fördelningen efter modellerna tillpassats. Det kan konstateras att det inte är normalfördelat. Tabell 7 redovisar de samlade modellerna för älvarnas dygnsobservationer där SETAR anges med maximal regimordning.

Tabell 6. De anpassade ARMA+GARCH-modellerna av residualerna från SETAR-modellering

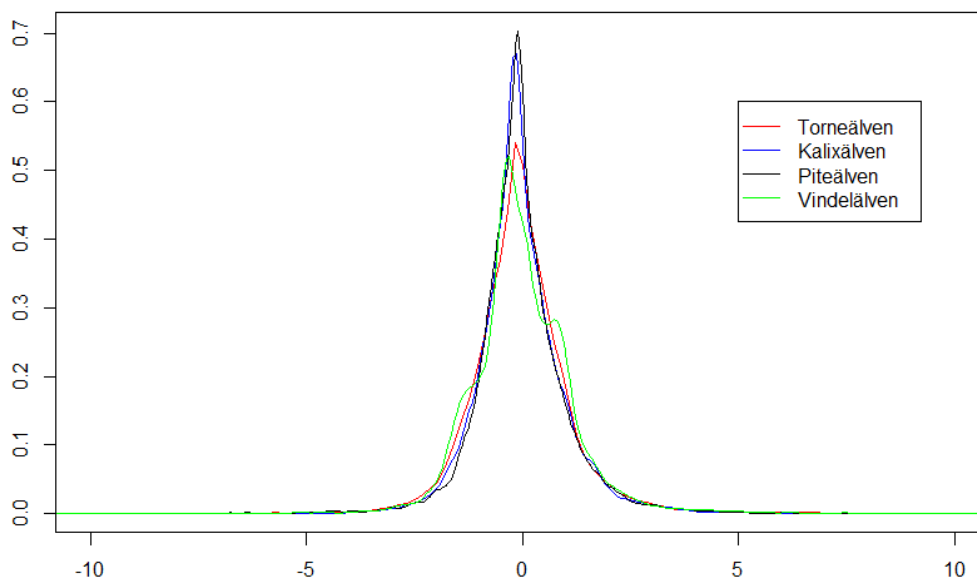
Älv/Period	ARMA(u,v)+GARCH(P,Q)	AIC	BIC	Li-Mak [p-värde]
1980-2015				
Torneälven	ARMA(1,1)+GARCH(1,1)	-4,4594	-4,4559	1
Kalixälven	ARMA(0,2)+GARCH(1,1)	-4,9838	-4,9803	0,7940
Piteälven	ARMA(1,2)+GARCH(1,1)	-4,0891	-4,0851	0,0574
Vindelälven	ARMA(1,1)+GARCH(1,1)	-6,7876	-6,7841	0,6115
1980-1985				
Torneälven	ARMA(0,0)+GARCH(1,0)	-4,2819	-4,2729	0,9989
Kalixälven	ARMA(0,0)+GARCH(1,1)	-4,7870	-4,7749	0,7162
Piteälven	ARMA(0,0)+GARCH(1,1)	-3,8166	-3,8045	0,5684
Vindelälven	ARMA(1,1)+GARCH(1,1)	-6,9437	-6,9256	0,7031
1995-2000				
Torneälven	ARMA(2,0)+GARCH(3,0)	-4,3490	-4,3278	0,8365
Kalixälven	ARMA(1,2)+GARCH(1,1)	-4,8902	-4,8691	0,5972
Piteälven	ARMA(1,1)+GARCH(1,1)	-4,2635	-4,2453	0,2152
Vindelälven	ARMA(1,1)+GARCH(1,1)	-7,0038	-6,9857	0,9296
2010-2015				
Torneälven	ARMA(1,1)+GARCH(1,1)	-4,6294	-4,6112	0,4531
Kalixälven	ARMA(1,1)+GARCH(1,1)	-5,0460	-5,0278	0,9458
Piteälven	ARMA(1,1)+GARCH(1,1)	-4,3748	-4,3567	0,0386
Vindelälven	ARMA(1,1)+GARCH(1,1)	-6,8933	-6,8751	0,9991



Figur 22. Vindelälvens hela material som exempel på kvarvarande beroendestruktur efter GARCH-anpassning



Figur 23. Vindelälvens hela material som exempel på kvarvarande beroendestruktur i högre beroendestrukturen efter anpassning av GARCH-modell



Figur 24. Täthetsfunktioner för samtliga älvars totala material efter anpassning av GARCH

Tabell 7. Slutgiltiga modeller för samtliga älvar och perioder

Älv/Period	SETAR(k,p,d)	ARMA(u,v)+GARCH(P,Q)
1980-2015		
Torneälven	SETAR(3,5,2)	ARMA(1,1)+GARCH(1,1)
Kalixälven	SETAR(2,8,1)	ARMA(0,2)+GARCH(1,1)
Piteälven	SETAR(2,3,0)	ARMA(1,2)+GARCH(1,1)
Vindelälven	SETAR(3,5,1)	ARMA(1,1)+GARCH(1,1)
1980-1985		
Torneälven	SETAR(3,4,3)	ARMA(0,0)+GARCH(1,0)
Kalixälven	SETAR(2,4,2)	ARMA(0,0)+GARCH(1,1)
Piteälven	SETAR(2,6,0)	ARMA(0,0)+GARCH(1,1)
Vindelälven	SETAR(2,4,0)	ARMA(1,1)+GARCH(1,1)
1995-2000		
Torneälven	SETAR(2,4,0)	ARMA(2,0)+GARCH(3,0)
Kalixälven	SETAR(2,4,0)	ARMA(1,2)+GARCH(1,1)
Piteälven	SETAR(2,3,1)	ARMA(1,1)+GARCH(1,1)
Vindelälven	SETAR(2,5,0)	ARMA(1,1)+GARCH(1,1)
2010-2015		
Torneälven	SETAR(2,4,1)	ARMA(1,1)+GARCH(1,1)
Kalixälven	SETAR(2,4,2)	ARMA(1,1)+GARCH(1,1)
Piteälven	SETAR(2,4,1)	ARMA(1,1)+GARCH(1,1)
Vindelälven	SETAR(2,5,0)	ARMA(1,1)+GARCH(1,1)

6. Diskussion

Denna statistiska analys som gjorts ger upphov till flera olika diskussionsämnen. För det första skapas kunskap om beroendestrukturen i det specifika datamaterialet som eventuellt kan generaliseras till liknande tidsseriestrukturer. För det andra kan metoden som används kommenteras och dess för- och nackdelar diskuteras. Slutligen kan ett resonemang föras gällande vilka komponenter modellen fångar upp och vad den missar.

Något som framkommer tydligt när vattenföringen studeras är att datamaterialet uppvisar en hög grad av komplexitet i beroendestrukturen för processvärdet samt i den betingade variansen. Utöver detta behöver en deterministisk trend och stora säsongsvariationer tas i hänsyn, vilket medför att denna typ av data är problematisk att modellera. Grundobservationerna uppvisar initialt en mycket skev fördelning som är svår att bestämma. Även efter en logtransformation består komplexiteten i datamaterialet. Fördelningen är multimodal, men upplevs vara mer symmetrisk efter logtransformation.

De tillpassade modellerna uppvisar ett antal intressanta drag. Ett av dessa är att när beroendestrukturen för processvärdet rensas bort återstår tydliga volatilitetskluster. Dessa sammanfaller med de perioder då flödet förändras hastigt och processen skiftar regim i SETAR-modellerna. Ett annat intressant drag är att ordningen för antalet ϕ -koefficienter i SETAR-modellen är generellt högre för de modellerna som tillpassats för perioden mellan 1980-2015 än för de femårsperioder som studerats. Detta bör vara rimligt då en större mängd data kan fånga upp en mer komplicerad struktur som förändras under tidsserien.

Vidare observeras att komplexiteten i beroendestrukturen är enklare för vissa av tidsperioderna när processen befinner sig i den lägre regimen. Särskilt i de modeller som berör hela datamaterialet och perioden 1980-1985. För de andra perioderna verkar antalet ϕ -koefficienter vara likvärdiga i både hög- och lågregim. Detta medför att det kan finnas en högre ordnings processdynamik. Med andra ord kan processens struktur ha förändrats under tiden som processen studeras. Detta kan indikera på att koefficienterna och modellparametrarna inte är konstanta utan förändras över tid, vilket kan medföra att parametrarna i sig är slumpmässiga eller följer en deterministisk struktur. Detta motsäger antagandet om sanna koefficienter. Alternativet kan även vara att metoden för att bestämma modellerna inte är tillräckligt precis.

Det finns dessutom en mycket stark återkoppling i form av värdet på ϕ -koefficienterna överlag i SETAR-modellerna. Det går även att se vissa problem med skattningstekniken i SETAR och den använda algoritmen då tröskelvärdet sätts olika för olika tidsperioder. Detta kan innebära svårigheter i att dra slutsatser om tidsseriens beroendestruktur har ändrats eller om andra skattningar har gjorts av tröskelvärdet, vilket medför att modellordningarna skiljer sig åt.

Det finns som tidigare nämnt en stark beroendestruktur i flera av älvarnas lägre regimer, där summan av ϕ -koefficienterna i SETAR-modellens regim är cirka ett, eller strax över ett. Detta medför att processens stationaritet i lågregimerna ifrågasätts. SETAR-modellens stationaritetsområde är något som fortfarande utforskas (Franses & van Dijk, 2000).

De mycket höga ϕ -koefficienterna i lågregimen för Torneälven 1980-1985 kan sannolikt sammankopplas med att den minsta mängden observationer i varje regim har satts till fem procent. Dock verkar denna gräns inte vara ett problem för de andra modellerna. Detta exemplifierar risken för felspecificeringen av modellordningen med det valda tillvägagångssättet.

När det gäller GARCH-komponenten i modelleringen uppstår det en del problem som är svåra att hantera. Först och främst är det den komplicerade fördelningen för datamaterialet som avviker ganska mycket från den ständigt förmodade normalfördelningen i litteratur. Det gör att all form av skattning där fördelningen spelar in behöver justeras, så som ML-skattningen i GARCH-modellering. Detta tas i hänsyn med en student-t-fördelning, men det finns fördelningar som troligtvis skulle passa ännu bättre och generera effektivare skattningar.

Det tillvägagångssätt som används för att analysera datamaterialet kan anses vara en tillämpning av det principiella sätt som Box-Jenkins tagit fram för att tillpassa modeller till tidsseriedata. Modellvalideringen består därför av att studera om residualerna är vitt brus.

Det som framkommer är att nästan all beroendestruktur kan rensas bort från processvärdet och från den betingade variansen. Enstaka laggar är signifikant skilda från noll i vissa av tidsperioderna, och dessa kvarstår oavsett val av modellordning. Detta kan vara en följd av brister i skattningstekniken för autokorrelationsstrukturen (Cryer & Chan, 2008). Det kan även vara ett problem med att komponenter successivt elimineras. Även om vissa laggar fortfarande är signifikanta så har en mycket stor del av beroendestrukturen eliminerats, och det ska påpekas att det ibland är mycket svårt att uppnå vitt brus i empirisk data (Cryer & Chan, 2008). Det finns inga signifikanta laggar som är större än drygt $\pm 0,1$ i skattningen av autokorrelationen för processvärdet vilket kan anses vara ett bra resultat med tanke på nivån beroendet startade på och avsaknaden av integrerade modeller.

Metoden bör därför anses vara framgångsrik i den aspekten att den möjliggör en förståelse av hur beroendestrukturen ser ut samt att den kan skapa en modell som till stor del eliminerar detta beroende. Metoden att successivt eliminera komponenter medför dock att variansen ökar i det brus som finns kvar för varje komponent som elimineras.

För att exemplifiera kommer eliminationen av en ϕ -koefficient medföra att: $E(\phi - \hat{\phi}) \approx 0$ och $V(\phi - \hat{\phi}) = V(\hat{\phi})$. Detta bör vara fallet för en AR(1)-modell, men bör även vara analogt för SETAR-modelleringen. Detta leder till att det kan finnas *rester* av tidigare koefficienter när nästa komponenter ska skattas, vilket

sannolikt försämrar koefficientskattningen och leder till en lägre precision i själva skattningen. När samtliga komponenter är eliminerade kommer tillvägagångssättet leda till att $V(e_t) \leq V(\hat{e}_t)$. Detta bygger på synsättet som presenteras i avsnitt 3.1. Det vill säga att det finns ett av naturen givet och sant ϕ -värde som därmed är en konstant för att resonemanget ovan ska stämma. Det finns andra metoder för att hitta de modeller som identifierats, exempelvis genom att lägga till koefficienter tills ej signifikanta parametrar uppkommer. Detta kommer troligtvis inte ge samma modellordningar. Vidare är inte kovariansstrukturen känd för skattningarna, vilket är ytterligare ett problem som påverkar metoden och kan påföra fel i skattningarna av koefficienterna. Detta medför att metoden troligtvis behöver undersökas och utvecklas vidare.

En viktig aspekt är frågan vad detta tillvägagångssätt fångar upp för komponenter. Genom att modellkomponenter upptäcks och elimineras successivt bör detta medföra att tillvägagångssättet har en god förmåga att upptäcka olika komponenter, givet att två komponenter inte direkt motverkar varandra.

En annan punkt att nämna är de strukturer i datamaterialet som de valda komponenterna fångar upp. Det är mycket svårt att ha kännedom om samtliga typer av strukturer som kan förekomma. Därför behandlas de vanligaste typerna av strukturer i hopp om att det ger ett satisfierat resultat.

För dygnsobservationerna används SETAR-modellens inbyggda konstanter som bör fånga upp vissa säsongeffekter. Denna modelltyp fångar dock endast upp beroenden av AR-typ då den inte innehåller några MA-komponenter (se ekvation (9)). En fördel med SETAR-modellen är att olika AR-strukturer kan fångas upp i en process som kan bete sig olika beroende på processvärdet i en specifik tidpunkt.

För vissa tidsperioder observeras en mer komplex beroendestruktur i den högre regimen. Detta är troligt då processen verkar vara mer instabil när processens värde är högre. Det kan ses som ytterligare ett argument för att en SETAR-modell är att föredra framför ARMA då SETAR kan fånga upp dessa föränderliga beroendestrukturer som beror på processens värde. Då tydliga multimodala fördelningar syns i datamaterialet kan det indikera på underliggande processer med olika beroendestruktur och behovet av SETAR.

Det starka säsongsmönster som finns i datamaterialet är ett av de största problemen som stöts på vid bearbetningen. Säsongen försvårar en traditionell ARMA+GARCH-anpassning av dygnsobservationer. Det verkar inte finnas några etablerade lösningar för detta problem i GARCH-sammanhang utan där tas andra metoder till hjälp. Dessutom begränsas möjligheterna av att SMHI har dygnsobservationer som lägsta upplösning och att det därför är svårt att se en lösning på säsongproblemet även om en bra metod för bearbetning framställs.

Månadsobservationerna för de fyra älvarna verkar kunna modelleras med S-ARIMA och GARCH. Några signifikanta laggar kvarstår i dess autokorrelation,

men kan bero på likväl spökkorrelation som att det finns en kvarvarande beroendestruktur i materialet. De olika älvarna kan dessutom modelleras med en liknande beroendestruktur och det går inte att se någon distinkt skillnad trots att storleken på älvarnas vattenföring skiljer sig markant. En fråga ställs därför hur informativ denna modellering är. Som det noterades i dataavsnittet skiljer sig värdena mycket mellan dygns- och månadsobservationer. Det finns alltså en tendens att beroendestrukturen i vattenföringen slätas ut med månadsdata, vilket rimligtvis bör ge resultatet att de olika älvarna inte skiljer sig särskilt mycket åt i hur de modelleras.

En problematisk aspekt är de snabba uppgångarna i vattenföringen som SETAR inte förklarar särskilt väl. Dessa uppgångar sker på ett säsongliknande sätt i alla perioder. Den mycket starka återkopplingen i de lägre regimerna bör medföra att flera sekventiella innovationstermer måste få värden som är mycket osannolika i förhållande till dess fördelning för att byta regim och uppvisa det mönster av snabb tillväxt som ger ett skifte av regimer, givet att modellen är korrekt specificerad. Detta medför att det kan finnas något annat som orsakar regimskifte men som inte fångas upp i de använda modellerna. Det kan vara en ny komponent eller något som SETAR-modellens regimkonstanter inte lyckas eliminera. Om det är en ny komponent kan den förslagsvis aktiveras under vissa tidsperioder och med någon sannolikhet. När komponenten aktiveras följer den snabba uppgång som observeras i processen.

Detta verkar också medföra att variansen förändras, det vill säga att GARCH-effekter verkar uppkomma. Det går att spekulera i vilka fysikaliska fenomen som orsakar detta, där en kan vara is- och snösmältningen på våren. Det kan således vara varmt väder eller högre temperaturer som orsakar att den snö och is som finns i avrinningsområdet smälter och börjar rinna ut mot havet. Det går troligtvis att studera och modellera temperaturväxlingarna i älvarnas avrinningsområden för att dra slutsatser om när detta sker. På så sätt bör denna komponent som orsakar det höga flödet gå att förklara med detta.

Detta medför att, givet att det är regn, snö- och issmältning som orsakar det snabbt växande flödet, borde en bättre modell kunna skapas genom att tillföra den information som finns i temperaturen vid avrinningsområdena. Med befintlig information om vattenföring i kombination med information gällande temperaturer borde modelleringen av aktiveringskomponenten kunna bli mer precis och därmed fånga upp de snabba flödesförändringarna bättre än med enbart information om vattenföring.

Gällande om det går att generalisera resultaten till övriga svenska älvar krävs ytterligare studier då många som nämnts tidigare har regleringar som styr vattenföringen på annat sätt. Troligtvis är det enklare modellering och bearbetning av reglerade älvar då de bör ha ett jämnare flöde för att till exempel vattenkraftverk ska kunna maximera sin produktion. En annan potentiell skillnad är att många älvar i södra delarna av landet inte upplever samma ihängande vinterkyla och därför inte

har lika omfattande töperiod under våren. Det är till och med möjligt att dygnsobservationerna för söderbelägna älvar kan modelleras med en simultan ARMA+GARCH tack vare ett svagt säsongberoende. Dessa potentiella skillnader genererar dock enklare modelleringar, vilket bör innebära att GARCH och SETAR även är lämpliga för dem. De oreglerade älvarna i norra delarna av Sverige bör rimligtvis skapa de mest komplicerade strukturerna och därför vara bäst som referenstest.

Slutligen kan det konstateras att beroendestrukturen i dygnsobservationerna verkar gå att modellera med en kombination av SETAR och GARCH. Visst restberoende finns kvar och frågan huruvida de snabba uppgångarna kan förklaras kvarstår dock efter att modellerna tillpassats.

7. Slutsatser och vidare forskning

7.1. Slutsatser

I inledningen ställs följande två frågor:

- Hur ser beroendestrukturen ut för en svensk oreglerad älv?
- Kan en svensk oreglerad älv modelleras med ARMA, GARCH eller SETAR?

De slutsatser som kan dras är följande - För det första går det att modellera älvarnas vattenföring med de ovan nämnda modelltyperna. Dock behöver dessa kombineras med varandra för att få ut det drivande vita bruset. De komponenter som verkar behövas för att modellera älvorna är följande: trendkomponent, säsongskomponent, beroendekomponent och en varianskomponent. Det kan även finnas ytterligare en komponent som triggar de snabba uppgångarna. Denna har dock inte studerats på djupet. Majoriteten av vattendragen behöver samtliga komponenter. Fokus i denna uppsats har varit på beroendekomponenten och varianskomponenten.

När traditionella beroendemodeller tillpassas på datamaterialet som exempelvis ARMA finns det tydliga indikationer på att en mer komplex beroendestruktur behövs för att till fullo modellera beroendet i vattenföringen från älvorna. Beroendekomponenten verkar vara av tröskel- och GARCH-typ.

För dessa komponenter kan beroendestrukturen modelleras på ett liknande sätt i samtliga av de studerade femårsperioderna. Beroendestrukturerna för tröskelmodellering uppvisar överlag en liknande komplexitet i både den höge och lägre regimen. Generellt är fördröjningen låg, vilket verkar logiskt när hänsyn tas till de snabba skiftningarna i processen. Vidare är det två regimer i nästan samtliga femårsperioder, medan det oftare förekommer tre regimer för hela tidsperioden. Antalet laggar är också högre när hela perioden modelleras vilket tillsammans med tendensen för att få tre regimer ger en slutsats att den extra komplexiteten i modellerna för hela datamaterialet krävs för att fånga upp den komplicerade strukturen som kan finnas under de 35 år som serien sträcker sig över.

Samtliga älvors vattenföring innehar någon form av varians som inte är konstant över tid och är alltså i behov att modelleras med GARCH. För månadsobservationerna räcker enkla ARCH-modeller för att nå vitt brus. Däremot kräver dygnsobservationerna något mer komplicerade modeller och det finns även mer restberoende i dessa innovationstermer. Den enda skillnaden som uppmärksammas mellan de olika tidsperioderna är att enklare ARMA+GARCH-modeller räcker för 1980-1985, där endast Vindelälven är i behov av en kompletterande ARMA-komponent.

Observationerna har en asymmetrisk grundfördelning och en bättre förklaring av beroendestrukturen kan uppnås genom att skattningarna beräknas med en fördelning som stämmer bättre överens än normal- respektive student-t-fördelningen.

Det går att modellera de fyra älvarnas månadsobservationer med S-ARIMA och GARCH, vilket genererar ett liknande modellval för älvarna. Denna medelvärdesbildning ger dock troligtvis inte en särskilt kvalitativ information om älvarnas verkliga beroendestruktur och det behövs sannolikt en större noggrannhet och bättre upplösning för att till exempel kunna varna för höga vattenflöden.

Att studera hur vattenföringen i de svenska älvarna beter sig kan betecknas som grundforskning. Den praktiska nyttan är just nu inte uppenbar. Kunskaperna kan dock användas vid senare tillfällen för att skapa en bättre förståelse kring exempelvis regleringsbehov eller påverkan på djur- och växtfaunan i de oreglerade älvarna. Kunskapen bör även ge en indikation på vad som skulle hända om exempelvis en nedmontering av olika älvregleringar skulle ske. Även om huvudfokus i denna uppsats inte har varit att prediktera vattenföringen kan en bättre förståelse och eventuella vidareutvecklingar av dessa resultat vara till nytta. Slutligen har denna uppsats poängterat de teoretiska svårigheterna med att modellera icke linjära tidsserier.

7.2. Vidare forskning

Först och främst bör det konstateras att just tidsserier inte är den enda möjligheten som existerar för att modellera vattenföringens beroendestruktur i älvarna. Andra metoder så som artificiella neurala nätverk skulle troligtvis även kunna användas.

Metoden och tillvägagångssättet i denna uppsats har gett upphov till fler frågor än den svarar på. I detta avsnitt behandlas de största av dessa. Engle (1982) presenterade ARCH-modellen som sen generaliserades till GARCH-modellen. Under de följande årtiondena har grundmodellen utvecklats och nya modeller har skapats. Andra GARCH-modeller än den traditionella borde kunna generera bättre skattningar. Till exempel kan någon modifierad variant av GJR-GARCH användas för att kompensera för de positiva chocker som uppstår i tidsserierna när is och snö smälter. Även TGARCH som är en form av tröskel-GARCH skulle kunna passa väl in på de fyra oreglerade älvarna.

Problematik gällande modellens metod och vad den fångar, respektive inte fångar upp har diskuterats. Ett förslag till vidare forskning är att arbeta vidare med en integrerad modell som bygger på den schematiska uppdelning som använts för att studera älvsystemen, se Figur 10. Att utveckla skattningstekniken för denna modell borde medföra att den diskuterade successiva ökningen i variansen när varje modellkomponent elimineras, minskar. Det borde vidare medföra att prediktionsförmågan förbättras om en integrerad skattningsteknik för parametrarna utvecklas. Detta skulle dock vara ett mycket komplicerat minimerings- eller maximeringsproblem i flera dimensioner beroende på om en ML- eller MK-metod används. En integrerad modell kan se ut på följande sätt:

$$\begin{aligned}
& Y(t) = \mu_t + S_t + \\
& \left\{ \begin{array}{l} \mu_1 + \varphi_1 * Y_{(t-1)} + \dots + \varphi_{1p} * Y_{(t-p)} + \theta_1 * e_{1(t-1)} + \dots + \theta_{1p} * e_{1(t-q)} + e_{1t}, \text{ om } Y(t)_{t-d} > C_1 \\ \mu_2 + \varphi_1 * Y_{(t-1)} + \dots + \varphi_{2p} * Y_{(t-p)} + \theta_1 * e_{2(t-1)} + \dots + \theta_{2p} * e_{2(t-q)} + e_{2t}, \text{ om } C_1 \leq Y(t)_{t-d} \leq C_2 \\ \mu_k + \varphi_1 * Y_{(t-1)} + \dots + \varphi_{kp} * Y_{(t-p)} + \theta_1 * e_{k(t-1)} + \dots + \theta_{kp} * e_{k(t-q)} + e_{kt}, \text{ om } C_{k-1} < Y(t)_{t-d} \end{array} \right. \quad (16)
\end{aligned}$$

Där $\mu_t = \alpha + \beta * t$, $e_{kt} = \sigma_{(t|t-1)}^2 * \varepsilon_t$ och $\varepsilon_t \sim f(E(\varepsilon_t) = 0, V(\varepsilon_t) = K)$

Här innebär f någon fördelning, till exempel normalfördelning.

Denna integrerade modell ska ses som ett förslag till vidare studier och har inte behandlats i denna uppsats då det inte var huvudsyftet. Modellen inkluderar de komponenter som illustreras i Figur 10. I säsongskomponenten kan exempelvis en aktiveringsfunktion ingå likt den beskrivna i diskussionen. Då borde denna integrerade modell, givet att skattningstekniken finns, troligtvis fånga upp de beroendestrukturer som observeras i datamaterialet. Genom att studera hur olika modeller kan kombineras bör möjligheten att undersöka en mängd olika datamaterial framträda. För att det ska ske måste dock skattningstekniken förbättras, inte bara för den föreslagna modellen utan även för generella kombinationer av modelltyper. Slutligen bör även den stationära regionen för SETAR-modellen utvecklas och generaliseras till en SETAR(k,p,d)-modell, för att bättre bedöma stationäritet. Samt att studera närmare hur icke-linjäritet av tröskeltyp kan upptäckas grafiskt.

Litteraturförteckning

Andersen, T. G., Davis, R. A., Kreiß, J.-P., & Mikosh, T. (2009). Handbook of financial time series. Berlin: Springer.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31, 307-327.

Chambers, J., Cleveland, W., Kleiner, B., & Tukey, P. (1983). Graphical Methods for Data Analysis. Wadsworth: Duxbury Press.

Chan, K.-S., & Ripley, B. (2012). Time series analysis. Hämtad 09-01-2016, från <https://cran.r-project.org/web/packages/TSA/TSA.pdf>

Cryer, J. D., & Chan, K.-S. (2008). Time Series Analysis. Iowa: Springer.

Enders, W. (2014). Applied econometric time series. Alabama: Wiley.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987-1007.

Franses, P. H., & van Dijk, D. (2000). Nonlinear time series models in empirical finance. Cambridge: Cambridge university press.

Li, W. K., & Mak, T. K. (1994). On the squared residual autocorrelations in nonlinear time series with conditional heteroskedasticity. Journal of Time Series Analysis (15), 627-636.

Lindh, A. (1998). Svar på skriftliga frågor. Hämtad 02-01-2016, från http://www.riksdagen.se/sv/Dokument-Lagar/Fragor-och-anmalningar/Svar-pa-skriftliga-fragor/nationalalvarna_GL12400/

McLeod, A. I., & Li, W. K. (1983). Diagnostics checking ARMA time series models using squared-residual autocorrelations. Journal of time series analysis, 4(4), 269-274.

Ruppert, D., & Matteson, D. S. (2011). Statistics and data analysis for financial engineering . New York: Springer.

SMHI. (2009). Vattenföring. Hämtad 02-01-2016, från <http://www.smhi.se/kunskapsbanken/hydrologi/vattenforing-1.6705>

SMHI. (2013). Hydrologiska begrepp. Hämtad 02-01-2016, från <http://www.smhi.se/kunskapsbanken/hydrologiska-begrepp-1.29125>

SMHI. (2014). Vattenföringsmätningar. Hämtad 02-01-2016, från <http://www.smhi.se/kunskapsbanken/hydrologi/smhis-vattenforingsmatningar-1.80833>

SMHI. (2015). Återkomsttider. Hämtad 02-01-2016, från <http://www.smhi.se/kunskapsbanken/aterkomsttider-1.89085>

SMHI. (u.å.). Vattenwebb. Hämtad 02-01-2016, från <http://vattenwebb.smhi.se/station/#>

Zivot, E., & Wang, J. (2006). Modeling financial time series with S-PLUS. New York: Springer.

Tong, H. (1990). Non-linear time series: a dynamical system approach. Oxford: Clarendon.

Bilagor

Databearbetning i denna rapport har utförts med programvaran R-studio (version 0.98.1102). Förutom de standardpaket som ingår har följande paket använts:

rugarch	(version 1.3-6)	För modellering av GARCH
tsDyn	(version 0.9-43)	För modellering av SETAR
xts	(version 0.9-7)	För tidsserier och grafik
forecast	(version 6.2)	För skattning av modeller
nsRFA	(version 0.7-12)	För modellverifiering

Datalogg:

2015-11-19

Vindelälven:

Observation ”2007-01-09” ändrad från ca 350 000 000 000 till 71,3 (enligt referensdata) pga. fel inmatning av okänd anledning.

Observation ”2009-04-25” ändrar från ca 2 390 000 till 60,4 (enligt referensdata) pga. fel inmatning av okänd anledning.

2015-11-20

Vindelälven:

Data (månad och dagar) är verifierad vid jämförelse av summa mellan CSV-filen och den beskärda referensfilen.

Piteälven:

Data (månad och dagar) är verifierad vid jämförelse av summa mellan CSV-filen och den beskärda referensfilen.

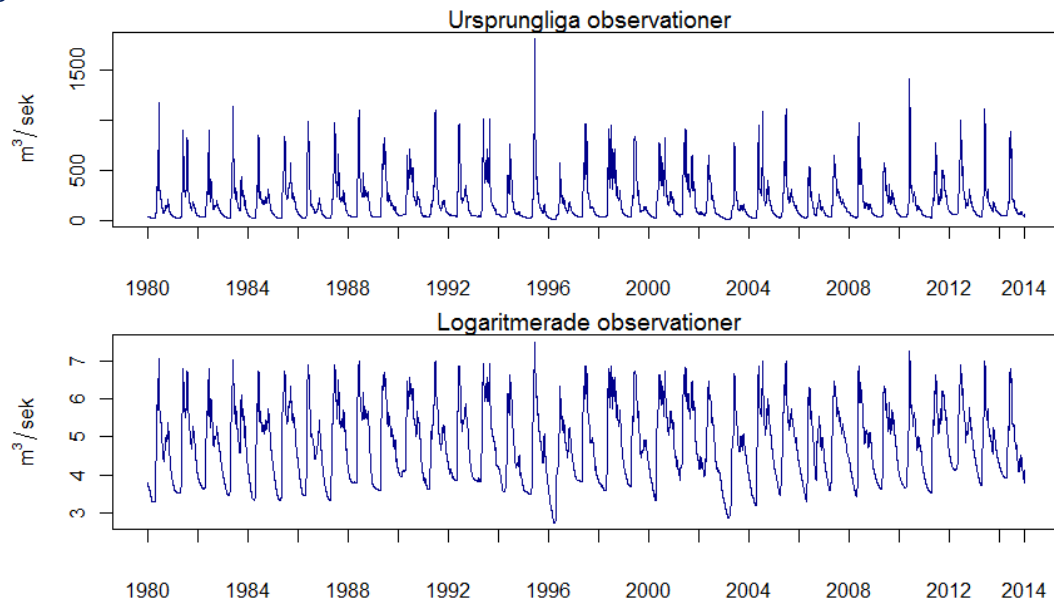
Torneälven:

Data (månad och dagar) är verifierad vid jämförelse av summa mellan CSV-filen och den beskärda referensfilen.

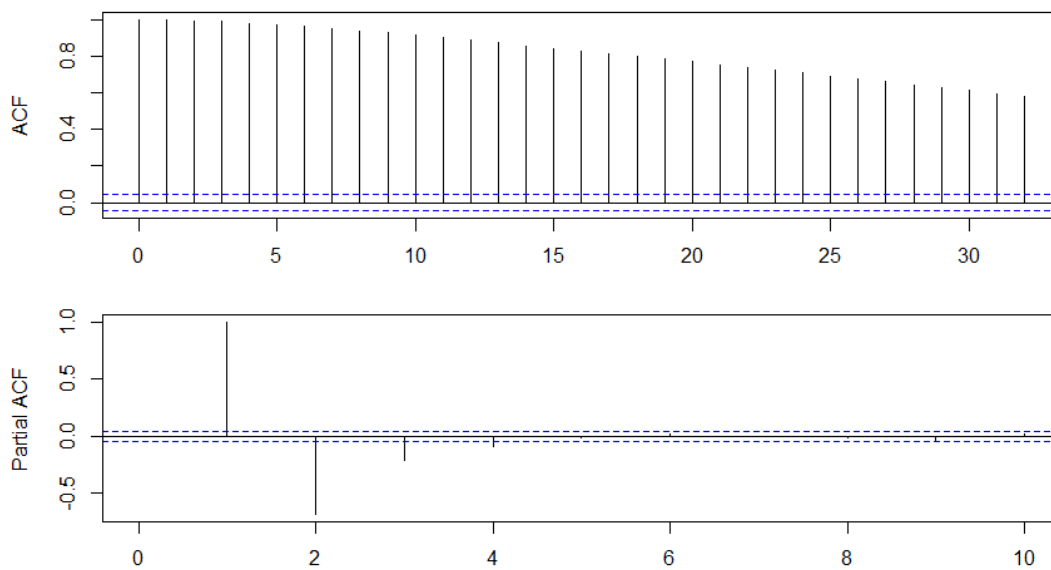
Kalixälven:

Data (månad och dagar) är verifierad vid jämförelse av summa mellan CSV-filen och den beskärda referensfilen.

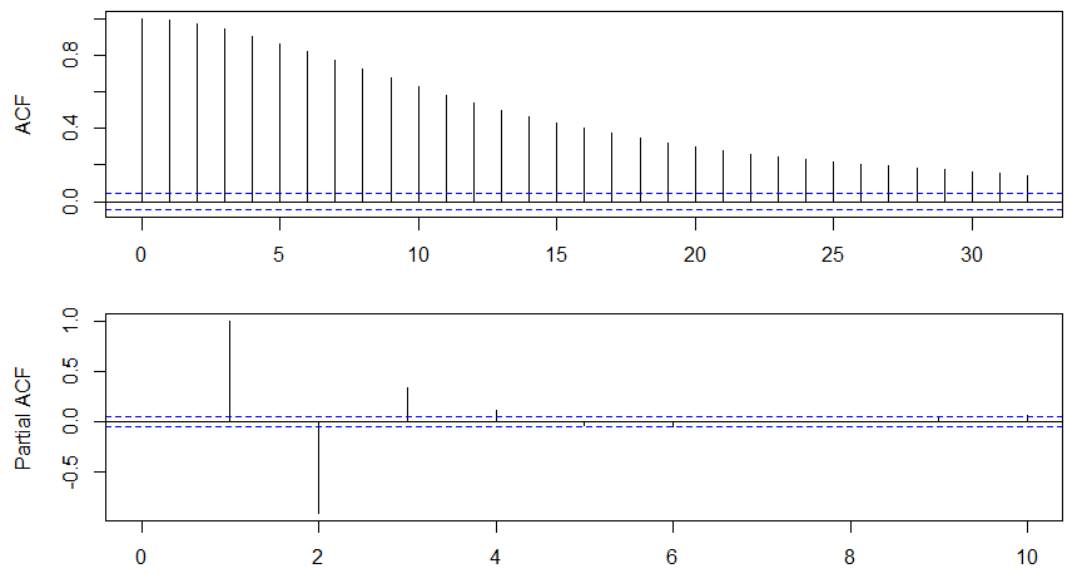
Bilaga A - Grundobservationer



Figur 25. Den ursprungliga tidserien samt den logaritmerade för Vindelälven

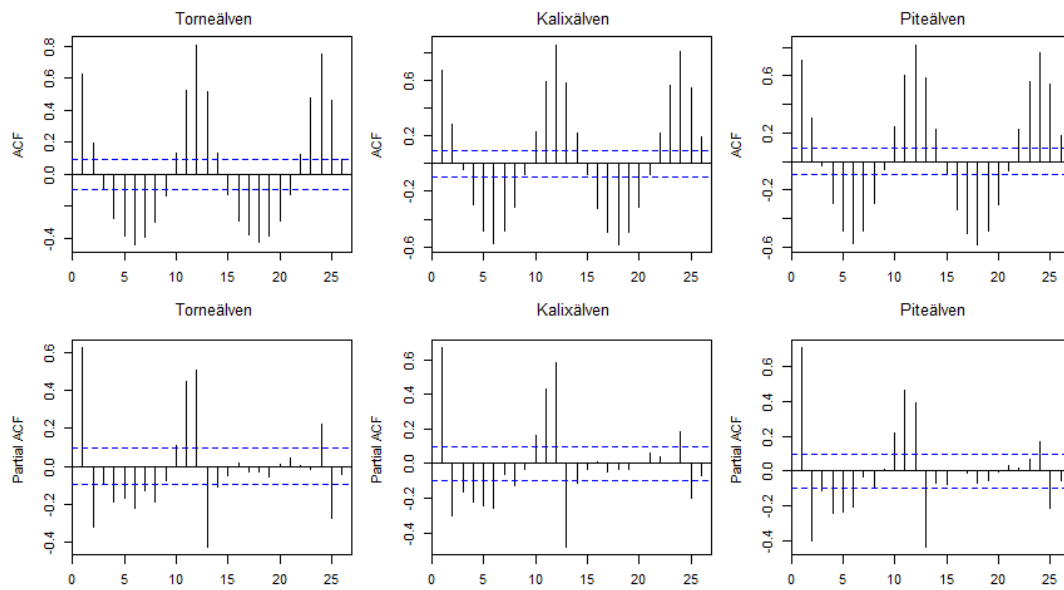


Figur 26. Vindelälvens ACF och PACF för dygnsobservationer under perioden 1980-1985

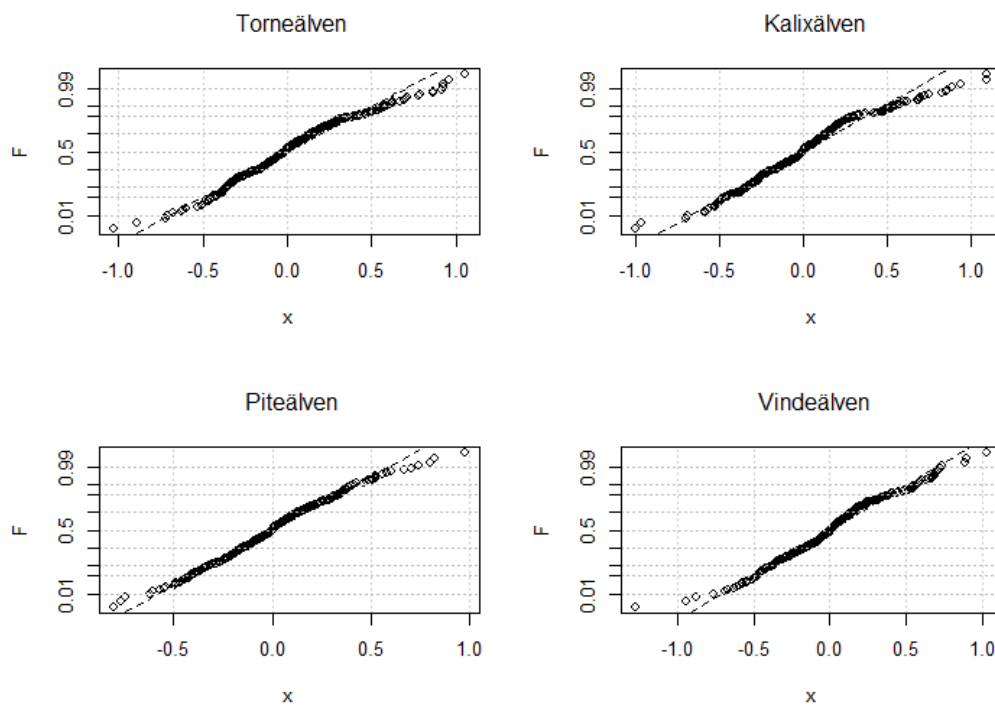


Figur 27. Vindelälvens ACF och PACF för kvadrerade dygnsobservationer under perioden 1980-1985

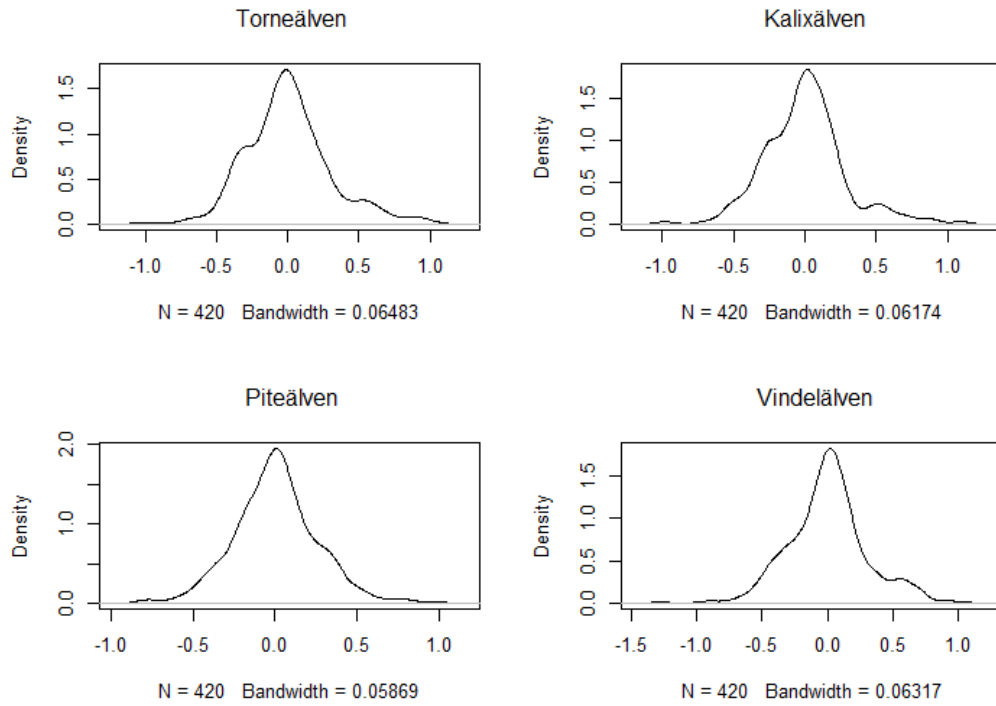
Bilaga B - Månadsobservationer



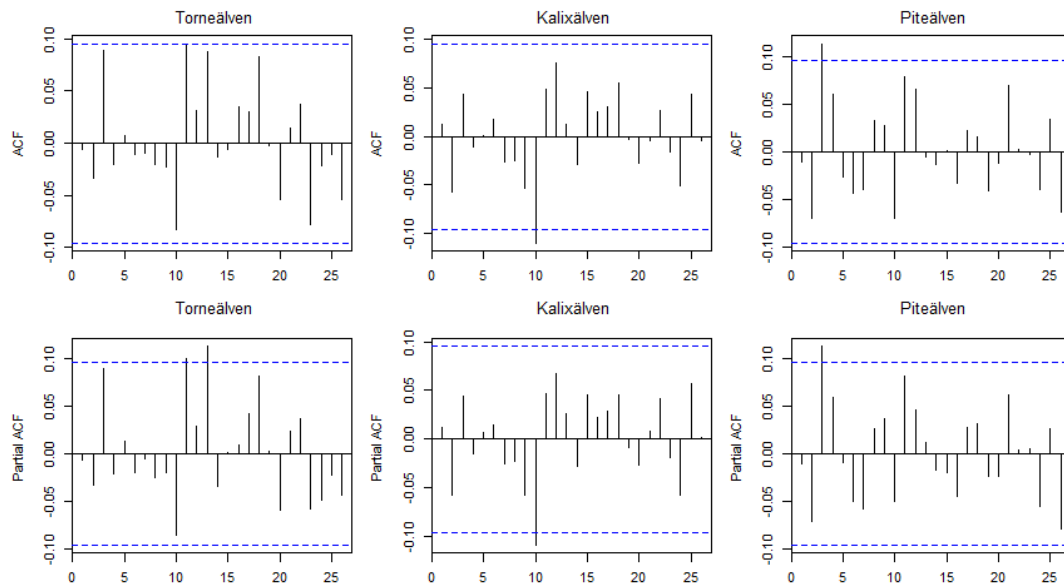
Figur 28. ACF och PACF för månadsobservationer innan säsongrensning



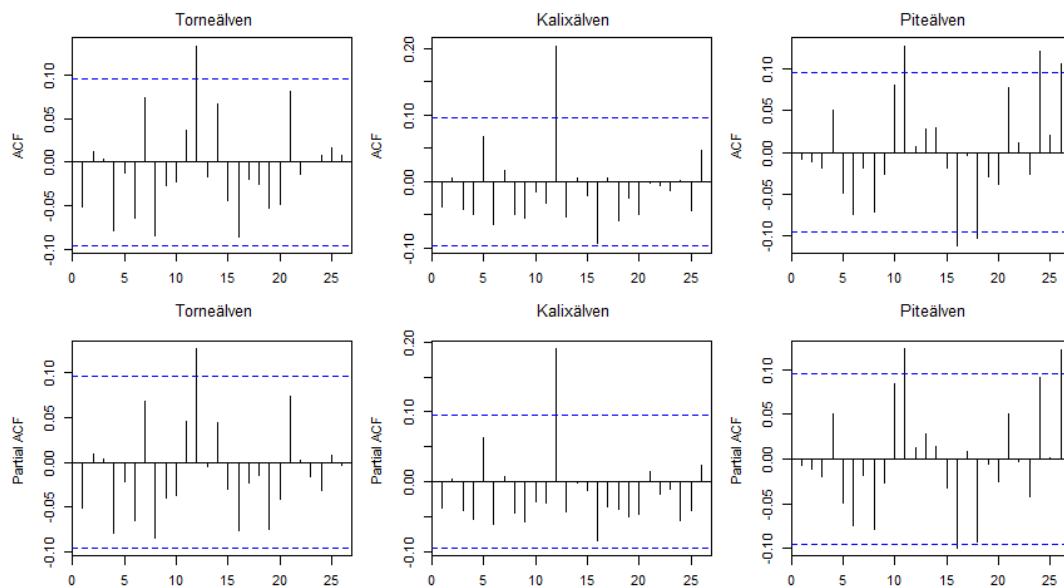
Figur 29. Normalfördelningsfigur över de olika vattendragen



Figur 30. Täthetsfunktioner för samtliga älvar



Figur 31. ACF och PACF för standardiserade residualer efter GARCH-anpassning

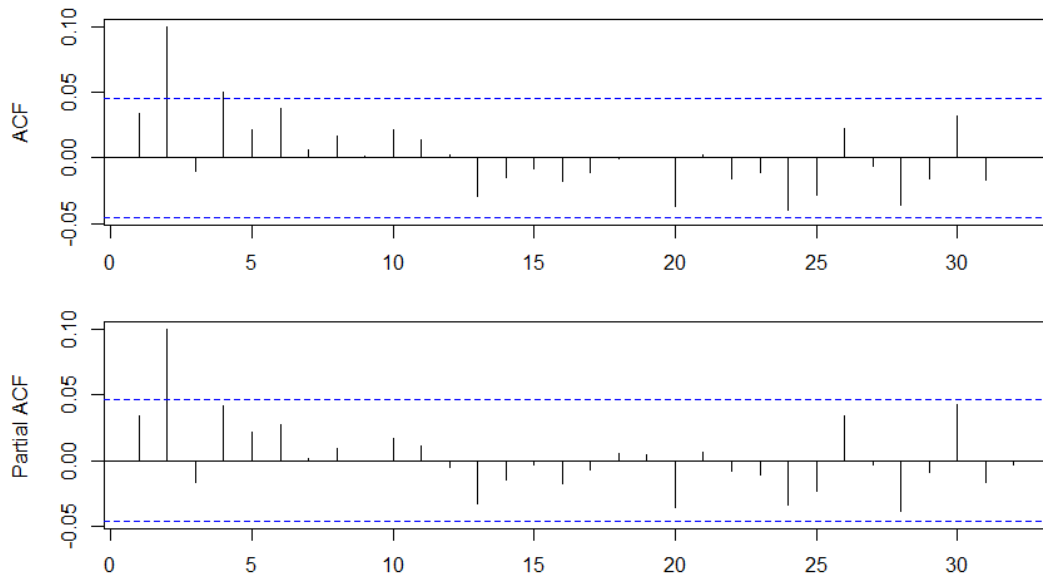


Figur 32. ACF och PACF för standardiserade residualer i kvadrat efter GARCH-anpassning

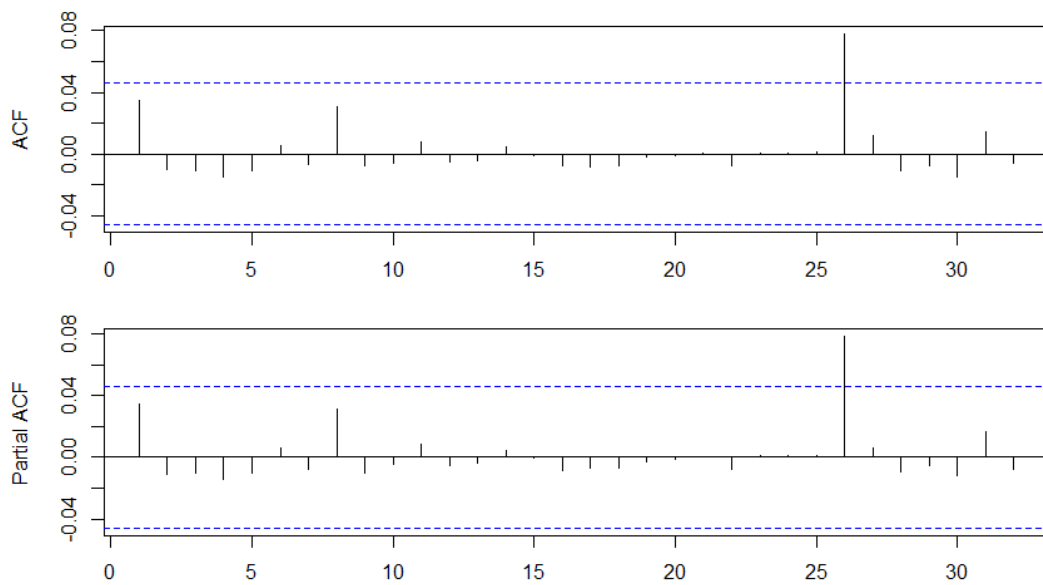
Tabell 8. Koefficienter för modellering av S-ARIMA och GARCH med månadsobservationer

Parametrar	Vattendrag							
	Torneälven		Kalixälven		Piteälven		Vindelälven	
	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm
S-ARIMA								
Φ	0,5857	0,0400	0,7432	0,0568	0,6500	0,0377	0,6052	0,0394
Θ	-	-	-0,2258	0,0847	-	-	-	-
D	1	-	1	-	1	-	1	-
Θ	-1	0,065	-0,9206	0,0340	-0,9101	0,0335	-0,9702	0,0644
GARCH								
ω	0,0586***	0,0087	0,0606***	0,0102	0,0496***	0,0062	0,0459***	0,0090
α_1	0,5063***	0,1626	0,4240***	0,1643	0,2727***	0,1044	0,4700**	0,1775
α_2	-	-	-	-	-	-	0,2704*	0,1168

Bilaga C - Dygnsobservationer för perioden 1980-1985



Figur 33. Vindelälvens ACF och PACF för residualer efter anpassning av GARCH för perioden 1980-1985



Figur 34. Vindelälvens ACF och PACF för residualer i kvadrat efter GARCH för perioden 1980-1985

Bilaga D - Modellkoefficienter

Nedan redovisas samtliga koefficienter för modellering av dygnsobservationer. Tabellerna innehåller värden för deterministisk trend, samt SETAR- och GARCH-parametrar.

Ett principutseende för modellerna beskrivs i detalj under teoriavsnittet med återges även här för överskådlighet.

$$\text{SETAR-modell: } Y_t = \begin{cases} (\mu_1 + \varphi_{1,1} * Y_{t-1} + \varphi_{1,2} * Y_{t-2} + e_t) & \text{om } Y_{t-1} \geq c_2 \\ (\mu_2 + \varphi_{2,1} * Y_{t-1} + \varphi_{2,2} * Y_{t-2} + e_t) & \text{om } c_1 < Y_{t-1} < c_2 \\ (\mu_3 + \varphi_{3,1} * Y_{t-1} + \varphi_{3,2} * Y_{t-2} + e_t) & \text{om } Y_{t-1} \leq c_1 \end{cases}$$

$$\text{ARMA+GARCH: } Y_t = \phi_1 Y_{t-1} + \dots + \phi_n Y_{t-n} + \theta_0 + e_t + \theta_1 e_{t-1} + \dots + \theta_v e_{t-v}$$

$$\text{Där } e_t = \sigma_{t|t-1} \varepsilon_t \text{ och } \sigma_{t|t-1}^2 = \omega + \alpha_1 e_{t-1}^2 + \dots + \alpha_q e_{t-q}^2 + \beta_1 \sigma_{t-1|t-2}^2 + \dots + \beta_p \sigma_{t-p|t-p-1}^2$$

Notation:

<i>Blankt</i>	<i>Ej signifikant</i>
<i>'</i>	<i>5% < P-värde < 10%</i>
<i>*</i>	<i>5%</i>
<i>**</i>	<i>1%</i>
<i>***</i>	<i>0,1%</i>
<i>-</i>	<i>Ingen parameter</i>

Tabell 9. Koefficienter för dygnsmodellering av Torneälven.

Parametrar	Period							
	1980-1985		1995-2000		2010-2015		1980-2015	
	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm
Deterministisk trend								
α	5,2450***	0,0434	5,1620***	0,0462	5,5360***	0,0411	5,5190***	0,0160
β	3,2E-4***	4,1E-5	4,0E-4***	4,4E-5	1,4E-4***	3,9E-5	1,4E-5***	2,2E-6
SETAR								
Lågregim								
Konstant	-0,4261*	0,2061	0,0205***	0,0053	0,0017	0,0015	0,0103	0,0119
φ_1	1,0670***	0,0603	1,6880***	0,0594	1,4440***	0,0288	1,9715***	0,0645
φ_2	3,8128***	0,4009	-0,4737***	0,1074	-0,1840***	0,0520	-0,4686***	0,1375
φ_3	-5,8946***	0,9983	0,2462*	0,1109	-0,2613***	0,0294	-0,4971***	0,0850
φ_4	1,6467***	0,7473	-0,4453***	0,0679	-	-	-	-
Tröskelvärde	-1,1534	-	-0,1619	-	0,6149	-	-1,0527	-
Mellanregim								
Konstant	0,0021	0,0047	-	-	-	-	0,00355***	0,00007
φ_1	1,5749***	0,0417	-	-	-	-	1,4582***	0,0112
φ_2	-0,3375***	0,0705	-	-	-	-	-0,2912***	0,0200
φ_3	-0,2377***	0,0363	-	-	-	-	-0,0613**	0,0208
φ_4	-	-	-	-	-	-	-0,0534*	0,0216
φ_5	-	-	-	-	-	-	-0,0516***	0,0131
Tröskelvärde	-0,0449	-	-	-	-	-	0,5485	-
Högregim								
Konstant	0,0055'	0,0030	0,0039	0,0027	0,0155**	0,0057	0,0082***	0,0023
φ_1	1,6905***	0,0225	1,4980***	0,0259	1,7935***	0,0380	1,6199**	0,0145
φ_2	-0,7018***	0,0227	-0,3613***	0,0465	-0,7114***	0,0776	-0,5538***	0,0273
φ_3	-	-	-0,1462***	0,0254	-0,2814***	0,0768	-0,1315***	0,0262
φ_4	-	-	-	-	0,1836***	0,0368	0,0524***	0,0129
ARMA+GARCH								
Φ_1	-	-	0,0888***	0,0263	0,9913***	0,0024	0,9922***	0,0007
Φ_2	-	-	0,2373***	0,0242	-	-	-	-
Θ_1	-	-	-	-	-0,9612***	0,0028	-0,9686***	0,0002
ω	0,0003***	3,6E-5	0,0001***	1,8E-5	9E-6***	2E-6	1,1E-5***	1E-6
α_1	0,9990***	0,0843	0,4803***	0,0508	0,2570***	0,0236	0,3124***	0,0102
α_2	-	-	0,2336***	0,0381	-	-	-	-
α_3	-	-	0,2851***	0,0432	-	-	-	-
β	-	-	-	-	0,7420***	0,0204	0,6866***	0,0092

Tabell 10. Koefficienter för dygnsmodellering av Kalixälven.

Parametrar	Period							
	1980-1985		1995-2000		2010-2015		1980-2015	
	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm
Deterministisk trend								
α	4,8670***	0,0454	4,7500***	0,0500	5,2410***	0,0412	5,1150***	0,0174
β	3,4E-4***	4,3E-5	4,7E-4***	4,70E-05	1,7E-4***	3,90E-05	2,1E-5***	2,4E-6
SETAR								
Lågregim								
Konstant	0	0,0044	0,0034'	0,0018	0,0009	0,0100	0,0032*	0,0013
φ_1	2,0488***	0,0285	2,0704***	0,0366	2,1029***	0,0264	2,0916***	0,020
φ_2	-1,0520***	0,0293	-1,5211***	0,0836	-1,6419***	0,0601	-1,2908***	0,0448
φ_3	-	-	0,6378***	0,0848	0,7373***	0,0618	0,2719***	0,0455
φ_4	-	-	-0,1866***	0,0386	-0,2001***	0,0283	-0,0718***	0,0217
Tröskelvärde	-0,3952	-	0,3373	-	0,8015	-	-0,4394	-
Högregim								
Konstant	0,0030'	0,0017	0,0035	0,0032	0,0165**	0,0063	0,0032***	0,0006
φ_1	2,1599***	0,0286	2,3240***	0,0382	2,3822***	0,0527	2,1928***	0,0098
φ_2	-1,6480***	0,0662	-1,8155***	0,0728	-2,0931***	0,1250	-1,7537***	0,0239
φ_3	0,5861***	0,0654	0,4864***	0,0371	0,8326***	0,1183	0,7861***	0,0295
φ_4	-0,1054***	0,0275	-	-	-0,1377**	0,0458	-0,3595***	0,0305
φ_5	-	-	-	-	-	-	0,1897***	0,0306
φ_6	-	-	-	-	-	-	-0,1019***	0,0297
φ_7	-	-	-	-	-	-	0,0739**	0,0240
φ_8	-	-	-	-	-	-	-0,0341***	0,0098
ARMA+GARCH								
Φ_1	-	-	0,4641***	0,173	0,9095***	0,1638	-	-
Θ_1	-	-	-0,5254**	0,1696	-0,7803***	0,2772	0,0381***	0,0103
Θ_2	-	-	0,1566***	0,0254	-	-	0,1232***	0,009
ω	1,4E-4***	2,5E-5	7,4E-5***	1,5E-5	3,3E-5	3,9E-5	5,9E-5***	0
α_1	0,6294***	0,089	0,5210***	0,0635	0,4530***	0,0877	0,5174***	0,0214
α_2	-	-	-	-	-	-	-	-
β	0,3696***	0,0546	0,4780***	0,0463	0,5460***	0,1768	0,4816***	0,0182

Tabell 11. Koefficienter för dygnsmodellering av Piteälven

Parametrar	Period							
	1980-1985		1995-2000		2010-2015		1980-2015	
	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm
Deterministisk trend								
α	4,4150***	3,6E-02	4,4820***	3,8E-02	4,9170***	3,4E-02	4,7420***	1,4E-02
β	3,5E-4***	3,5E-05	3,6E-4***	3,6E-05	-1,7E-05	3,3E-05	1,5E-5***	1,9E-06
SETAR								
Lågregim								
Konstant	0,0027	0,0098	-0,0017	0,0071	0,0383*	0,0173	-0,0044***	0,0012
φ_1	0,9997***	0,0120	1,5589***	0,0416	1,5255***	0,028	1,3918***	0,0108
φ_2	-	-	-0,8376***	0,0719	-0,6263***	0,0488	-0,4015***	0,0108
φ_3	-	-	0,2717***	0,0418	0,09218**	0,0289	-	-
Tröskelvärde	-0,5287	-	-0,4452	-	4,9345	-	0,0246	-
Högregim								
Konstant	0,0004	0,0021	-0,0005	0,0017	0,0488*	0,0193	0,0027*	0,0013
φ_1	1,4371***	0,0268	1,7171***	0,0217	1,9628***	0,0406	1,5865***	0,0137
φ_2	-0,3111***	0,0465	-0,7180***	0,0218	-1,3978***	0,0844	-0,5338***	0,0242
φ_3	-0,01256**	0,0470	-	-	0,6153***	0,0816	-0,0595***	0,0132
φ_4	-0,070	0,0470	-	-	-0,1893***	0,0378	-	-
φ_5	0,1935***	0,0465	-	-	-	-	-	-
φ_6	-0,1301***	0,0268	-	-	-	-	-	-
ARMA+GARCH								
Φ_1	-	-	0,8723***	0,0606	0,9320***	0,0189	0,9144***	0,0112
Θ_1	-	-	-0,8093***	0,0748	-0,8382***	0,0292	-0,8029***	0,0158
Θ_2	-	-	-	-	-	-	-0,0536***	0,0108
ω	0,0001'	7,70E-05	1E-04***	1,90E-05	8E-5***	1,70E-05	1,3E-4***	1,10E-05
α	0,3788***	0,0616	0,4440***	0,0544	0,4530***	0,0516	0,4810***	0,0207
β	0,6202***	0,096	0,5550***	0,0406	0,5459***	0,0444	0,5180***	0,0181

Tabell 12. Koefficienter för dygnsmodellering av Vindelälven.

Parametrar	Period							
	1980-1985		1995-2000		2010-2015		1980-2015	
	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm	Skattning	Felterm
Deterministisk trend								
α	4,3942***	0,0422	4,2970***	0,0469	4,8340***	0,0393	4,7410***	0,0168
β	3,5E-4***	4,10E-05	4,4E-4***	4,4E-5	-2,7E-6	3,7E-6	7,7E-6***	2,3E-6
SETAR								
Lågregim								
Konstant	0,0014'	0,0080	0,0006	0,0009	0,0102*	0,0048	0,0021**	0,0007
φ_1	1,6450***	0,0302	1,9632***	0,0437	1,6712***	0,0284	2,0790***	0,0196
φ_2	-0,6667***	0,0590	-1,0547***	0,0998	-0,7858***	0,0558	-1,3257***	0,0452
φ_3	0,2786***	0,0593	0,2405*	0,1137	0,3422***	0,0604	0,5607***	0,0538
φ_4	-0,2583***	0,0310	-0,2653*	0,1072	-0,1554**	0,0563	-0,3844***	0,0538
φ_5	-	-	0,1146*	0,0512	-0,0743*	0,0291	0,0709**	0,0268
Tröskelvärde	0,6955	-	0,698	-	5,273	-	-0,3253	-
Mellanregim								
Konstant	-	-	-	-	-	-	-0,0004	0,0004
φ_1	-	-	-	-	-	-	1,7389***	0,0122
φ_2	-	-	-	-	-	-	-0,7981***	0,0249
φ_3	-	-	-	-	-	-	0,2172***	0,0245
φ_4	-	-	-	-	-	-	-0,1559***	0,0118
Tröskelvärde	-	-	-	-	-	-	0,7326	-
Högregim								
Konstant	0,00755'	0,0039	0,0072'	0,0037	0,0369***	0,0104	0,0087***	0,0011
φ_1	2,5493***	0,0656	2,4364***	0,0561	2,5147***	0,0393	2,4494***	0,0160
φ_2	-2,3393***	0,1703	-2,1352***	0,1345	-2,3245***	0,0978	-2,1584***	0,0402
φ_3	0,9118***	0,1688	0,8535***	0,1290	0,9766***	0,0962	0,8301***	0,0397
φ_4	-0,1291*	0,0639	-0,1608**	0,0504	-0,1733***	0,0377	-0,1291***	0,0154
ARMA+GARCH								
Φ_1	0,9973***	0,0014	1,0000***	0,0016	0,9994***	0,0014	0,9686***	0,0039
Θ_1	-0,9116***	0,0123	-0,9468***	0,0072	-0,9333***	0,0082	-0,9029***	0,0096
ω	0	1E-6	0	1E-6	0	1E-6	1E-6	0
α	0,3835***	0,0239	0,3707***	0,0273	0,3613***	0,0252	0,3475***	0,0097
β	0,6155***	0,0177	0,6283***	0,0225	0,6377***	0,0207	0,6515***	0,0079