

Marianne Thorsnæs

Kolnes

Oppgave for graden master i statistikk

Matematisk Statistikk

Universitetet i Bergen, Norge

24. november 2009

Kalmanfilteret anvendt
på en problemstilling
innen seismikk



Denne oppgaven er skrevet i L^AT_EX_{2 ϵ} med dokumentklassen «uib-mi-master», laget av Karl Ove Hufthammer. Den ble kompilert med pdfTeX-1.40.3 den 24. november 2009. Brødteksten er satt i 11 punkts URW Palladio med kapiteler. Matematikken er satt i URW Palladio og Pazo Math, overskrifter i HV Math og programkode i Bera Mono.

Takk

Jeg vil først takke hovedveileder, Hans Julius Skaug, for god veiledning. En stor takk går også til veileder, Peder Solheim, for mange gode ideer til problemstillingen og god støtte.

Tusen takk, Svein Olav Knudsen, som introduserte meg til Peder.

Jeg vil også gjerne takke min fantastiske familie og nære gode venner, for oppmuntring og støtte. En spesiell takk til min kjære mor for at hun alltid stiller opp, og til min kjære far for korrekturlesing med innslag av positive tilbakemeldinger.

Ellers har jeg satt stor pris på hjelp og støtte fra medstudenter, og profesorer her på Kroepeliens hus. En spesiell takk fortjener Karl Over Hufthammer, for eksperthjelp i Latex. Også, Andreas Stordal, Kristian Thulin og Arne Holmin for ideer til oppgaven.

Til slutt vil jeg takke Asgeir Sørensen og Egil Pedersen som har svart på spørsmål rettet mot kybernetikk, og Henning Omre som henviste meg til de.

Innhold

1	Innledning	1
1.1	Seismiske undersøkelser	1
1.2	Problemstilling	2
1.3	Statistiske metoder	3
2	Bayesiansk inferens	7
2.1	Bayes' formel	8
2.1.1	A priori-fordelingen, $p(x)$:	9
2.1.2	Data-fordelingen, $p(y x)$:	9
2.1.3	A posteriori-fordelingen, $p(x y)$:	10
2.1.4	Marginal-fordelingen, $p(y)$:	11
2.1.5	A posteriori prediktiv-fordeling	11
2.2	Normalfordelingen	12
2.2.1	Bayes' formel og Normalfordelingen	13
2.3	Multivariabel normalfordeling	18
2.3.1	Bayes' Formel og Multivariabel normalfordeling	19
3	Tidsrekkeanalyse	26
3.1	Tidsrekke-teori	26
3.2	Stasjonære tidsrekker	28
3.2.1	ARMA(p,q)-modeller	30
3.2.2	Momenter i AR(1)-modellen	34
3.2.3	Prediksjon	35
3.3	Ikke-stasjonære tidsrekker	36
3.3.1	Tidsrekke 1 og 2	36
3.3.2	Trend	39
3.3.3	Parabelmetoden	41
3.3.4	ARIMA(p,d,q)-modellen	45

3.3.5	Akaikes' og Bayesian informasjonskriterium	45
3.4	Box-Jenkins teori	46
3.4.1	Identifisere ARIMA-modellen	46
3.4.2	Estimerer ARIMA-modellen	54
3.4.3	Unit root	62
3.4.4	Evaluerer ARIMA-modellen	63
3.5	Generalisering av ARMA-modellen	66
4	Sekvensielle metoder	68
4.1	Gaussiske State-space modeller	68
4.1.1	Generalisering av Gaussisk State-Space modell	71
4.2	Bayes' formel og State-space modellen	71
4.2.1	Filtrering og Prediksjon	73
4.3	Kalmanfilter	74
4.3.1	Likningene i Kalmanfilteret	74
4.3.2	Utleddning av likningene	75
4.3.3	Kalmanfilter-Algorithmen	77
4.3.4	Kalmanfilteret på Tidsrekke 1 og 2	77
4.4	Generalisering av Kalmanfilteret	79
4.4.1	Likningene i Fler-dimensjonalt Kalmanfilteret	80
4.4.2	3-dimensjonalt Kalmanfilter på Tidsrekke 1	81
4.5	ARMA-modeller på State-space form	83
5	Konklusjoner	86
5.1	Prediksjon av posisjon	86
5.2	Estimat av akselerasjon	87
A	Tabeller	89
	Litteratur	97

1

Innledning

Denne masteroppgaven er en monografi om en gitt problemstilling fra seismikken. Problemstillingen ble først kjent i samarbeid med veileder Peder, og senere fullstendig bestemt, sammen med veileder Hans Julius Skaug.

1.1 Seismiske undersøkelser

Geofysikere bruker fire forskjellige metoder for å analysere de ulike lagene i undergrunnen, den såkalte stratigrafien. Disse er basert på: seismikk, elektromagnetisme, gravitasjon og magnetisme. Alle har et felles mål om å finne olje. Problemstillingen er knyttet til marin seismikk, som er en analyse av stratigrafien under havet. I slike seismiske undersøkelser samles data inn med seismikk-kabler, som slepes bak en båt. Det blir brukt mellom 1-20 kabler, som slepes parallelt med en separasjon på ca 100 meter. Kablene ligger på ca. 8 meters dyp, og er mellom 3000-8000 meter lange. Dybden til kablene blir justert, opp eller ned, til ønsket dybde med såkalte *birds*. Disse er utstyrt med vinger og elektromotorer, som tilføres strøm gjennom kabelen. Det finnes også birds som styrer kablene horisontalt.

Fra en luftkanon på båten, sendes det ut trykkbølger som forplanter seg mot havbunnen og de geologiske lagene under. Deler av en trykkbølge blir

reflektert opp igjen når den treffer et nytt lag i stratigrafien. Desto dypere dette laget ligger, desto lenger tid bruker trykkbølgen på å returnere. Tiden blir registrert av *hydrofoner* som er festet langs kablene. Parameteren tid blir så konvertert til en dybdeverdi.

Posisjonen til seismikk-kablene blir målt hvert tiende s med stor nøyaktighet takket være akustiske målinger. Punktene på kablene som blir målt, kaller vi for *noder*. Avstanden mellom hver node på kabelen varierer mellom 100-600 meter.

Når vi vil skyte seismikk i et bestemt område, på et *preplot*, vil vi at kablene skal følge en bestemt rute til enhver tid i undersøkelsen. Ofte vil vi undersøke nøyaktig samme område som har blitt analysert tidligere, for så å sammenligne disse og spore opp eventuelle feil. Da vil vi styre kablene nærmest mulig posisjonen de hadde ved forrige undersøkelse, kontinuerlig.

1.2 Problemstilling

En fordel med seismikk-data, i motsetning til brønnplukk, som er observasjoner av stratigrafien tatt fra brønner, er at de dekker store områder, så og si kontinuerlig. En ulempe er derimot at de er ganske usikre. En potensiell forbedring i usikkerheten til seismikken er å øke evnen vår til å styre seismikk-kablene dit de skal ligge. For å kontrollere kablene må vi, i tillegg til å vite nåværende posisjon, vite hvor kablene sannsynligvis vil ligge i nærmeste fremtid. Med denne informasjonen, er målet å styre kablene, med birds, til «ønsket» posisjon dersom de ligger, eller vil ligge, på avveie.

I dette studiet fokuserer vi på posisjonen noden har i én dimensjon. Posisjonen i den horisontale, tverrgående retning. Denne retningen referer vi til som øst-vest for den definerte *krysslinjen*. Der ønsket posisjon ligger på krysslinjen. Dybdeverdier til kabelen analyserer vi ikke i dette studiet. Heller ikke nord-sør retning, siden kabelen blir dratt etter båten og posisjonen i denne retningen er forutsigbar, se figur 1.2. Dessuten tar vi ikke stilling til hvor sikker og rask styringen er, men det bør nevnes at dersom standardavviket til styringen, f. eks. 2 meter, er større enn avviket kabelen har fra ønsket posisjonen, f. eks. 1 meter, styrer vi selvsagt ikke kabelen. Usikkerheten i styringen er leverandøren av birdsene ansvarlig for.

I sin helhet er dette en kompleks problemstilling. Vi ser på problemstillingen fra en rent matematisk synsvinkel. Som sagt vet vi nåværende posisjon til

seismikk-kablene. De fremtidige posisjonene derimot må vi predikere med en gunstig statistisk modell. Gitt en gunstig modell er hovedmålet å finne en metode som predikerer posisjonen til seimikk-kabler, fortløpende. Vi vil vite hvor sikkert vi kan predikere posisjonen. Vi starter med å estimere og predikere evolusjonen til én enkelt node. Basert på samtale med Peder, antar vi at 6 tidssteg, dvs. 60 s, er tilstrekkelig å predikere, for å rekke å styre kablene til ønsket posisjon.

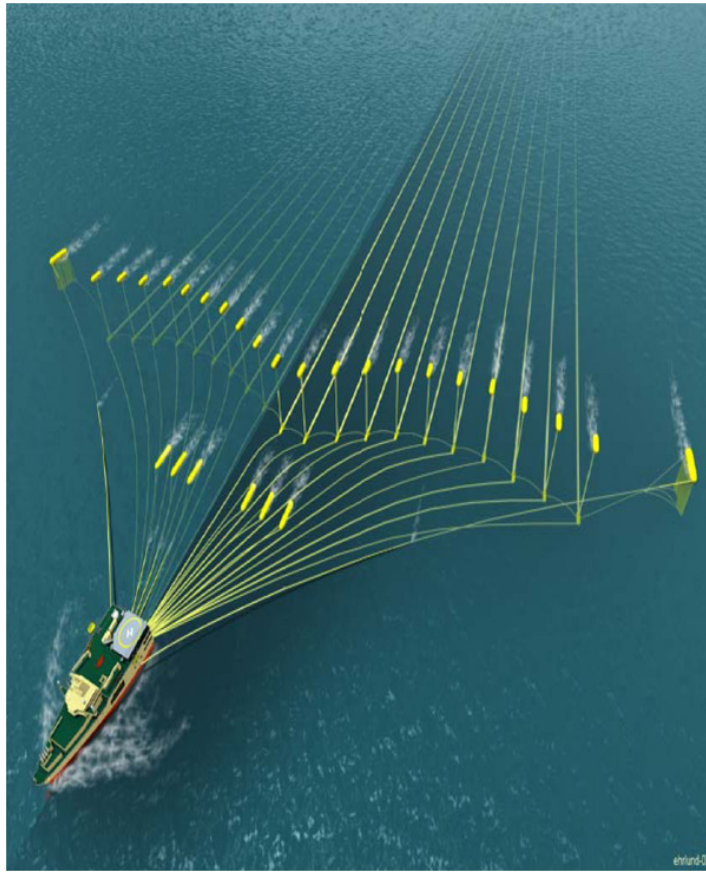
Et delmål i studiet er å finne ut om nøyaktigheten i prediksjonen kan økes ved å inkludere målinger av akselerasjon. Måling av akselerasjon er interessant, fordi det er mulig å feste et akselerometer på kabelen, og måle akselerasjon direkte, i motsetning til fart. Om dette vil være gunstig, avhenger av usikkerheten i akselerometeret, og hvor sikkert vi kan estimere akselerasjon basert på målinger av posisjon. Et mål på usikkerheten i akselerometeret kan vi få fra en leverandør. Men først vil vi finne ut hvor sikkert vi kan estimere akselerasjonen. Estimerer vi den mer sikkert enn et akselerometer vil måle den, da inkluderer vi neppe akselerasjonsmålinger.

1.3 Statistiske metoder

Dersom en seismisk undersøkelse analyserer i feil område, er store summer gått tapt. Dette er imidlertid ikke den eneste motivasjonen i dette studiet. Basert på vårt litteratursøk, er vi de første som angriper problemstillingen med følgende statistiske metoder. På forhånd visste vi derfor lite om de valgte metodene anvendt på lignende problemstillinger.

Modellen spiller en sentral rolle i prediksjon. Siden denne problemstillingen er kompleks, inneholder en reell matematisk modell svært mange og usikre variabler. Havstrømmer, bølger, tidevann etc. kan være vanskelig å modellere. Selv om værvarselet kan forutsi været relativt nøyaktig, er havvarsel et ganske nytt fenomen, og mer usikkert. Vi har derfor valgt å ikke modellere posisjonen til seismikk-kablene basert på det reelle dynamiske systemet. For øvrig kan avansert modellering av seimikk-kabler studeres i Turkyilmaz (2004), Pedersen og Sørensen (2001) og Pedersen (2000).

Vi har valgt en alternativ fremgangsmåte. Hydrodynamiske parametre o.l betrakter vi som stokastiske parametre, som ikke inngår i modellen, men som stokastisk støy. Videre antar vi at støyen i modellen er gaussisk med forventning 0. Det kan tenkes, at en dominerende havstrøm etc., vil føre til at

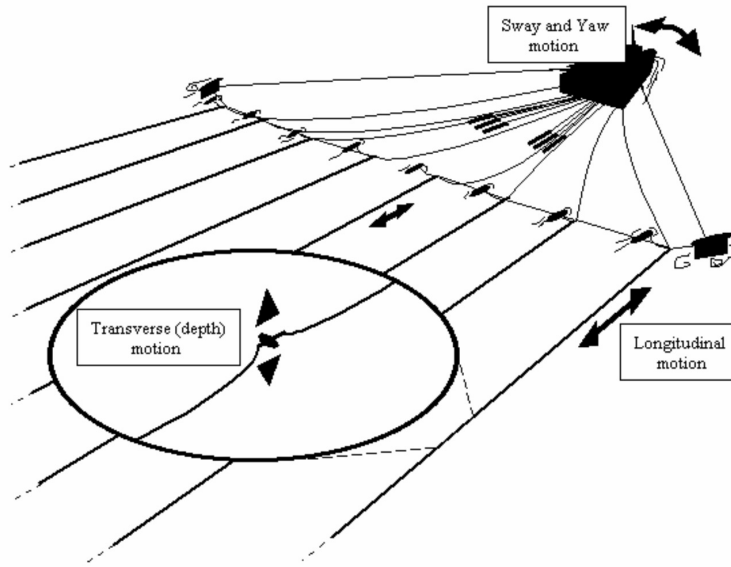


Figur 1.1: Seismikk-kabler slepes bak en båt.

disse antakelsene svikter. Likevel er det naturlig å anta symmetri her, vi vet ikke hvilken retning strømmen isåfall vil ha, og da er det meningsløst å anta at posisjonen til noden har en skjev normalfordeling.

Basert på disse antakelsene, modellerer vi evolusjonen til noden, med ulike tidsrekkemodeller. Dermed drar vi nytte av den sterke autokorrelasjonen til posisjonen, og sikre målinger. Først tilpasser vi en *lokal kvadratisk trend*-modell. Deretter tilpasser vi ulike *ARIMA*-modeller, og finner den «optimale» modellen ved hjelp av modelltilpasningskriterier. Til slutt tilpasser vi *state-space* modeller, assosiert med Kalmanfilteret.

Riktignok er dette et inverst problem, hvor parametrene i modellene blir estimert med tilgjengelig data. Alle de nevnte modellene blir estimert ut fra et klassisk frekventistisk perspektiv, der modellparametrene blir bestemt på grunnlag av *maksimum likelihood* (ML) estimater. Siden ML-estimering er en



Figur 1.2: Illustrasjon av ulike bevegelser til båten og seismikk-kablene. Bevegelsen til seismikk-kabelen i horisontal tverrgående retning, som vi analyserer, er ikke spesifikt merket på figuren.

sentral metode vi bruker i store deler av studiet vil vi definere Likelihood-funksjonen her.

Definisjon 1.3.1: Likelihood-funksjonen

Likelihood-funksjonen for n uavhengige data, $\mathbf{y}_{1:n}$, der alle har lik fordeling $p()$ med ukjente parametre gitt i vektoren $\boldsymbol{\theta}$, er definert ved:

$$L(\boldsymbol{\theta}, \mathbf{y}_{1:n}) = \prod_{i=1}^n p(y_i, \boldsymbol{\theta})$$

der L står for Likelihood-funksjon.

I Likelihood-funksjonen er det $\boldsymbol{\theta}$ som variere mens målingene $\mathbf{y}_{1:n}$ er gitt. Definisjonen er funnet i Berger R. L (2002). Siden vi antar at fordelingen til målingene er gaussisk, $Y_i \sim \mathcal{N}(\mu, \sigma^2) \forall i$, er Likelihood-funksjonen gitt ved:

$$L(\boldsymbol{\theta}, \mathbf{y}_{1:n}) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right) \quad (1.1)$$

Ved å maksimere (1.1) med hensyn på de forskjellige parametrene, finner vi ML-estimatene $\hat{\mu}, \hat{\sigma}^2$. Merk at når fordelingen er gaussisk, er $\hat{\mu}$ ekvivalent med estimatet vi finner med *minste kvadraters* metode (MKM).

Videre predikerer vi posisjon og akselerasjon til en node sekvensielt gitt ulike tidsrekke-modeller. Vi anvender rekursive prediksjons-algoritmer som er optimale gitt de tilhørende modellene. Prediksjon basert på ARIMA-modellen er klassisk frekventistisk inferens, mens prediksjon med Kalmanfilteret basert på gaussiske state-space modeller er Bayesiansk inferens. Kalmanfilteret, som i prinsippet er data-assimilasjon, forener målinger med en matematisk modell. Dette er interessant fordi vi håper å finne en konsistent modell som beskriver den sanne kursen til noden. Ensemble Kalmanfilter Evensen (1996) ble også vurdert i problemstillingen, men siden den endelige modellen i dette studiet er lineær, ble det ikke nødvendig å anvende denne metoden.

En fordel med state-space modeller er den enkle stukturen de er bygd opp av, og at det er enkelt å inkludere en eventuell kontroll-input i modellen. Siden kybernetikk åpenlyst inngår i problemstillingen har vi undersøkt hvilken kontroll-input som er mest passende under disse forholdene, og forsikret oss om at en kontroll-input kan inngå i metoden vi bruker.

Kort oppsummert er hovedmålet å predikere posisjoner til seismikk-kablene best mulig, slik at kablene kan styres nærmest mulig ønsket posisjon, fortløpende, i seismiske undersøkelser. Et delmål er å finne ut om et akselerometer vil styrke inferensen til posisjonen til noden. Da jeg satt meg inn i problemstillingen ble spesielt Richard C. Aster og Thurber og J. Meinhold, lagt vekt på. Den grunnleggende teorien i kapittel 2 og 3 bygger henholdvis på Gelman (2004) og Brockwell og Davis (2002). Da jeg begynte å lese om Kalmanfilter fant jeg mye nyttig stoff i Christopher K. Wikle (2006) og Wikle (2005), som teorien i kapittel 4 er basert på. Da jeg studerte Ensemble Kalmanfilter hentet jeg mest stoff fra Thulin (2005), Mandel (2006) og Skjervheim (2007). I forbindelse med utregninger har jeg funnet mye støtte i Petersen (2008).

Under studiet er programpakken R, versjon 2.9.2 2009, benyttet. Der pakkene fmek, fnkie, psfn og sdri har blitt mest brukt. Plottene er plottet i ggplot2. Datasettene jeg har tatt utgangspunkt i er levert av Geograf AS. Oppgaven er skrevet i Latex, versjon Latex2e.

2

Bayesiansk inferens

I dette kapitlet presenterer vi Bayesiansk inferens, siden dette er den grunnleggende teorien som Kalmanfilteret bygger på. Ofte deles sannsynlighet inn i to hoveddefinisjoner, klassisk frekventisme og Bayesiansk inferens. Frekventisme er ren empirisk statistikk mens Bayesiansk inferens kombinerer empiri med en subjektiv kunnskap, f. eks en fysisk modell. Det subjektive bidraget kan også være resultater basert på tidligere data fra lignende situasjoner etc. Uansett, kan det være feil å styrke sannsynlighet med subjektiv «tro». Forskere tar feil iblant, og noen mener at sannsynlighet skal være objektivt, og derfor kun bygge på empiri. Vi har valgt begge alternativ. I kapittel 3 utfører vi klassisk frekventisisk inferens, mens her Bayesiansk.

Metoder som i prinsippet er data assimilasjonen, bygger ofte på Bayesiansk inferens, blant andre kriging og Kalmanfilter. Ideén bak data assimilasjon er å oppdatere eksisterende kunnskap, ofte i form av matematiske modeller, med målinger. Vi er interessert i å oppdatere *a priori*-fordelingen til posisjonen til noden, med målinger av posisjonen. Med *a priori* mener vi at fordelingen ikke betinger på tidligere data. Målet er å ende opp med en mer nøyaktig *a posteriori*-fordeling, ved å benytte Bayes' formel. Med *a posteriori* mener vi at fordelingen betinger på målinger.

Først estimerer vi a posteriori-fordelingen gitt målinger av posisjon. Deretter estimerer vi a posteriori-fordelingen gitt målinger av posisjon og akselerasjon, siden vi er interessert i å vite i hvor stor grad målinger av akselerasjon kan styrke inferensen. For å inkludere målinger av akselerasjon i inferensen, utvider vi a priori-fordelingen til en 3-dimensjonal fordeling med parametrene: posisjon, fart og akselerasjon. Vi fokuserer på gaussiske a posteriori-fordelinger, siden dette er en forutsetning i et optimalt Kalmanfilter.

2.1 Bayes' formel

I Bayesiansk inferens deler vi ofte stokastiske variabler inn i følgende:

X =latent variabel

Y =observerbar variabel.

Videre opererer vi hovedsakelig med tre forskjellige fordelinger:

$p(x)$ - a priori-fordeling til X

$p(y | x)$ - data-fordelingen til Y gitt x

$p(x | y)$ - a posteriori-fordelingen til X gitt y .

Her antar jeg at leseren er kjent med marginalfordeling til X , $p(x)$, simultanfordeling til X og Y , $p(x, y)$, og betinget fordeling til X gitt y , $p(x | y)$. Stokastiske variabler skrives med store bokstaver, mens ikke-stokastiske føres med liten bokstav. Videre bruker jeg $p()$ for tetthets-funksjon, og fet skrift for vektorer og matriser.

Bayesiansk inferens er et resultat av Bayes' formel gitt i definisjonen under.

Definisjon 2.1.1: Bayes' Formel

$$p(x | y) = \frac{p(y | x)p(x)}{p(y)}$$

når $p(y) = \int p(x)p(y | x) dx \neq 0$.

Når y er en konstant, er $p(y)$ en normaliseringskonstant, og vi skriver

$$p(x | y) \propto p(y | x)p(x). \quad (2.1)$$

For å finne a posteriori-fordelingen må vi først finne den simultane fordelingen til den virkelige prosessen x , og målingen y , $p(x, y)$. Denne fordelingen kan uttrykkes som $p(x, y) = p(y | x)p(x) = p(x | y)p(y)$, siden formelen for betinget fordeling er $p(y | x) = \frac{p(x, y)}{p(x)}$. Det er viktig å merke seg at $p(x, y)$ må være konsistent. Dette krever pålitelig kunnskap om x og riktig innsamling av data. Noe som er karakteristisk for Bayesiansk statistikk er at inferensen om x uttrykkes som en sannsynlighetsfordeling. Det er derfor meningsfylt å undersøke sannsynligheten for regn imorgen, på lik linje som sannsynligheten for at en mynt vil lande med krone opp. De forskjellige leddene i Bayes' Formel presenterer jeg kort nedenfor, siden det øker forståelsen av bayesiansk inferens.

2.1.1 A priori-fordelingen, $p(x)$:

Valg av a priori-fordeling er en viktig del av bayesiansk inferens fordi den representerer vår forståelse av den ukjente, x , før nye målinger er tilgjengelige. Oftest er den et resultat av historiske data, tidligere erfaring, eller vitenskapelige modeller. Når fordelingen er konstruert, gir den oss et tilfeldig utvalg av sannsynlige verdier til x . Noen ganger velger man en ikke-informativ a priori-fordeling hvis kunnskapen om x er veldig usikker. Da er fordelingen lik 1, eller tilsvarende for flere variabler identitetsmatrisen, I . Velger vi en ikke-informativ a priori-fordeling er a posteriori-estimatet lik maksimum likelihood-estimatet. Dette er klassisk frekventistisk inferens.

2.1.2 Data-fordelingen, $p(y | x)$:

Denne fordelingen blir også kalt for observasjons-fordelingen, og er fordelingen til dataene gitt den virkelige prosessen x . Dataene påvirker bare a posteriori-fordelingen gjennom data-fordelingen. Anta at Y er en måling med feil, og x er den virkelige posisjonen til en node. Da er $p(y | x)$ fordelingen til målefeilen i de observerte målingene, der årsaken kan være feil ved opplesning eller i instrument, såvel som at X avviker fra sann posisjon. Dersom fordelingen anses som en funksjon av X , mens y er konstant, er den kjent som likelihood funksjonen, $L(x | y)$. Når y er konstant og X varierer, finner vi hvilken x som gir maksimum likelihood.

2.1.3 A posteriori-fordelingen, $p(x | y)$:

Vi så i (2.1) at a posteriori-fordelingen er proporsjonal med produktet av a priori-fordelingen og data-fordelingen. A posteriori-fordelingen er en oppdatering av vårt a priori-estimat gitt de tilgjengelige målingene. Således er bayesiansk inferens vitenskapelig konsistent siden den oppdaterer teori med hensyn på data. Vi kan se en sammenheng mellom forventningen til a priori-fordelingen og forventningen til a posteriori-fordelingen ved

$$\mathbb{E}(x) = \mathbb{E}(\mathbb{E}(x | y)) \quad (2.2)$$

der $\mathbb{E}(x) = \int \int xp(x, y) dx dy = \int \int xp(x | y)p(y) dx dy = \int \mathbb{E}(x | y)p(y) dy$. Den innerste forventningen finner gjennomsnittet av X betinget på y , mens den ytterste finner gjennomsnittet av Y . A priori-forventningen er lik gjennomsnittet av de mulige forventningene til a posteriori-fordelingen med hensyn på alle mulige data fra data-fordelingen. Tilsvarende kan vi uttrykke variansen ved forventning av betinget varians og varians av betinget forventning slik:

$$\begin{aligned} \text{Var}(x) &= \mathbb{E}(\text{Var}(x | y)) + \text{Var}(\mathbb{E}(x | y)) \\ &= \mathbb{E}(\mathbb{E}(x^2 | y) - (\mathbb{E}(x | y))^2) + \mathbb{E}(\mathbb{E}(x | y)^2) - (\mathbb{E}(\mathbb{E}(x | y)))^2 \\ &= \mathbb{E}(x^2) - \mathbb{E}(\mathbb{E}(x | y)^2) + \mathbb{E}(\mathbb{E}(x | y)^2) - \mathbb{E}(x^2) \\ &= \mathbb{E}(x^2) - \mathbb{E}(x)^2. \end{aligned}$$

Siden a posteriori-fordelingen er en oppdatert a priori-fordeling, er det naturlig å forvente at variansen til den førstnevnte er minst. Snur vi litt på likheten over ser vi at:

$$\mathbb{E}(\text{Var}(x | y)) = \text{Var}(x) - \text{Var}(\mathbb{E}(x | y)). \quad (2.3)$$

Altså er a posteriori-variansen gjennomsnittlig mindre enn a priori-variansen, og desto større variansen til a posteriori-forventningen er, desto større er reduseringspotensialet i a posteriori-variansen. Bayesiansk inferens er gitt ved a posteriori-fordelingen, siden den inneholder all tilgjengelig informasjon om x . Siden inferensen er en sannsynlighets-fordeling fremstilles den ofte grafisk. Inferensen kan også gis ved deskriptive data så lenge de er suffisiente. Er den gaussisk gis inferensen med forventning og varians. Når fordelingene er gaussiske finnes det en analytisk løsning.

2.1.4 Marginal-fordelingen, $p(y)$:

Fordelingen er gitt ved $p(y) = \int p(x)p(y | x) dx$ når vi antar kontinuerlig X . Dersom X er diskret, finnes det tilsvarende likhet med summasjon istedenfor integrasjon. Først og fremst blir fordelingen tenkt på som normaliseringskonstanten i Bayes' formel når y er konstant. Men fordelingen blir også kalt a priori prediktiv - fordeling, fordi den bidrar til å finne a posteriori prediktiv-fordelingen som omtales senere. Integralet over kan sjeldent løses analytisk, avhengig av fordelingen til dataene og a priori-fordelingen.

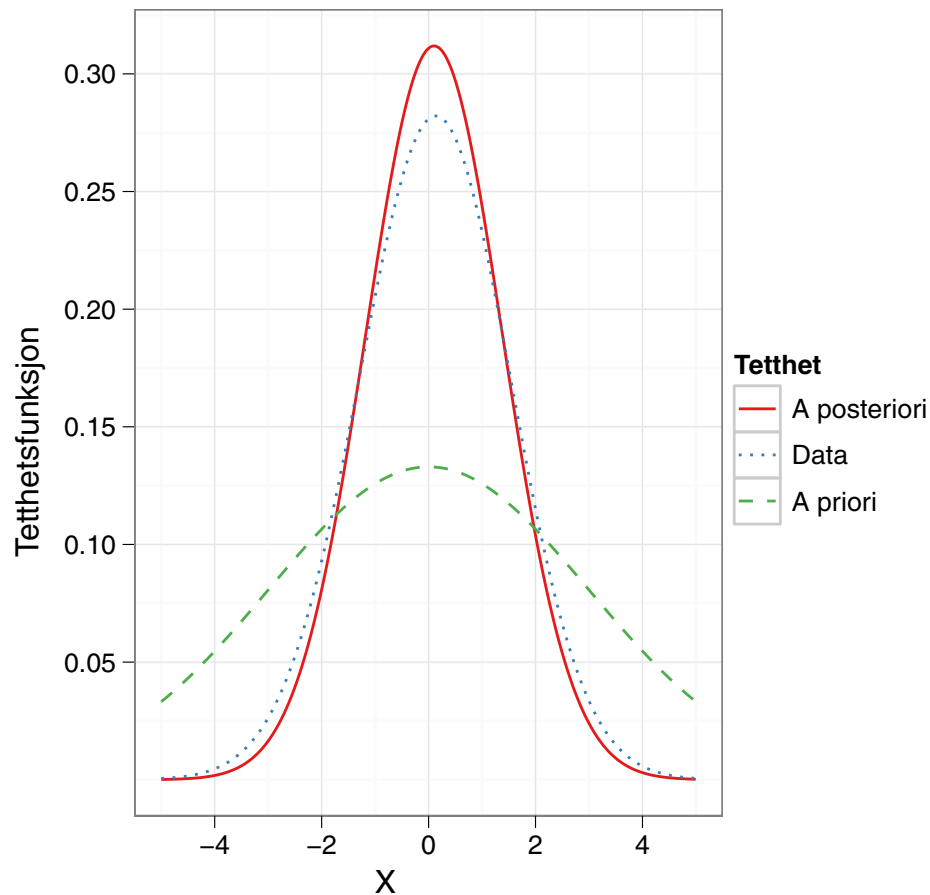
Figur 2.1 under, viser de tre førstnevnte fordelingene grafisk, basert på data fra eksempel 2.2.1. Legg merke til at forventningen til a posteriori-fordelingen er et kompromiss av de to andre, og at a posteriori-fordelingen er mest presis.

2.1.5 A posteriori prediktiv-fordeling

Dersom målinger mangler fra ved et tidspunkt, kan vi predikere målingen, og erstatte målingen med et estimat. Inferens om en ukjent observerbar, \tilde{y} , kalles prediktiv inferens. Før målingen y er tilgjengelig, så er fordelingen lik $p(y) = \int p(x)p(y | x) dx$. Fordelingen blir kalt a priori prediktiv-fordeling. Prediktiv fordi det er fordelingen til en observerbar parameter. Når data har blitt observert, kan vi predikere en ukjent observerbar måling \tilde{y} fra samme prosess. Fordelingen til \tilde{y} gitt y kalles a posteriori prediktiv-fordelingen. Den er gitt ved:

$$\begin{aligned} p(\tilde{y} | y) &= \int p(\tilde{y}, x | y) dx, \\ &= \int p(\tilde{y} | x, y)p(x | y) dx, \\ &= \int p(\tilde{y} | x)p(x | y) dx. \end{aligned} \tag{2.4}$$

I første linje over integrerer vi ut x . Andre linje er et resultat av Bayes' formel. Legg merke til at $p(u, v, w) = p(u | v, w)p(v | w)p(w)$. Tredje likhet fordi \tilde{y} gitt x ikke betinger på tidligere data.



Figur 2.1: A posteriori-fordelingen, data-fordelingen og a priori-fordelingen fra eksempel 2.2.1.

2.2 Normalfordelingen

Normalfordelingen er fundamental i mange lineære statistiske metoder, blant disse Kalmanfilteret og Ensemble Kalmanfilteret. Når vi antar at a priori-fordelingen og data-fordelingen er gaussiske, er Kalmanfilteret og Ensemble Kalmanfilteret optimalt. Dessuten fører dette til at også a posteriori-fordelingen er gaussisk, og derfor fullstendig forklart av forventning og variansen. Dette avsnittet er hovedsakelig basert på Wikle (2005) som også antar at a priori-fordelingen og data-fordelingen er gaussiske. Hvis a priori- eller data-fordelingen *ikke* er gaussiske, kan de bli transformert til gaussiske fordelinger, for så å bli transformert tilbake igjen etterpå. Som kjent er normalfordelingen

gitt ved:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad -\infty < x < \infty \quad (2.5)$$

der μ og σ^2 henholdsvis er forventningen og variansen til X .

2.2.1 Bayes' formel og Normalfordelingen

Nå skal vi anvende Bayes'formel på gaussisk a priori- og data-fordeling, siden det er dette likningene i Kalmanfilteret bygger på. Anta at posisjonen til en node har a priori-fordeling, $X \sim \mathcal{N}(\mu, \sigma^2)$. Betinget på den virkelige prosessen, x , har vi n uavhengige målinger med feil, $\mathbf{Y} = (Y_1, \dots, Y_n)'$. Dersom data-fordelingen til i 'te måling, gitt x , er $Y_i | x \sim \mathcal{N}(x, \tau^2)$, så er:

$$p(\mathbf{y} | x) = \frac{1}{(2\pi)^{\frac{n}{2}} \tau^n} \exp\left(-1/2 \sum_{i=1}^n (y_i - x)^2 / \tau^2\right).$$

Siden $p(\mathbf{y} | x) = \prod_{i=1}^n p(y_i | x)$ når observasjonene gitt x er uavhengige. Som et resultat av Bayes' formel kan vi vise at når de to fordelingene er gaussiske, er a posteriori-fordelingen også gaussisk. Bayes' formel gir da:

$$\begin{aligned} p(x | \mathbf{y}) &\propto \prod_{i=1}^n p(y_i | x) p(x) \\ &= \exp\left(-1/2\left(\sum_{i=1}^n (y_i - x)^2 / \tau^2 + (x - \mu)^2 / \sigma^2\right)\right) \\ &= \exp\left(-1/2\left(\sum_{i=1}^n (y_i^2 - 2x y_i + x^2) / \tau^2 + (x^2 - 2x\mu + \mu^2) / \sigma^2\right)\right). \end{aligned}$$

For å vise at a posteriori-fordelingen er gaussisk, samler vi alle variablene i eksponenten, mens konstantene uttrykkes som C . Siden konstanter ikke påvirker sannsynlighetsfordelingen, kan vi addere og subtrahere dem for å

fullføre kvadratet som vist nedenfor:

$$\begin{aligned}
&= \exp\left(-1/2(x^2(n/\tau^2 + 1/\sigma^2) - 2x(\sum_{i=1}^n y_i/\tau^2 - \mu/\sigma^2) + C)\right) \\
&= \exp\left(-\frac{(n/\tau^2 + 1/\sigma^2)}{2}(x^2 - 2x(\sum_{i=1}^n y_i/\tau^2 - \mu/\sigma^2)(n/\tau^2 + 1/\sigma^2)^{-1} + C)\right) \\
&\propto \exp\left(-\frac{(n/\tau^2 + 1/\sigma^2)}{2}(x - (\sum_{i=1}^n y_i/\tau^2 - \mu/\sigma^2)(n/\tau^2 + 1/\sigma^2)^{-1})^2\right).
\end{aligned}$$

Når vi sammenligner dette med eksponenten i (2.5), ser vi at a posteriori-fordelingen er gaussisk og gitt ved:

$$X | \mathbf{y} \sim \mathcal{N}((n/\tau^2 + 1/\sigma^2)^{-1}(\sum_{i=1}^n y_i/\tau^2 - \mu/\sigma^2), (n/\tau^2 + 1/\sigma^2)^{-1}).$$

Forventningen kan også skrives slik:

$$\begin{aligned}
\mathbb{E}(x | \mathbf{y}) &= (n/\tau^2 + 1/\sigma^2)^{-1}(\sum_{i=1}^n y_i/\tau^2 - \mu/\sigma^2) \\
&= \frac{\tau^2\sigma^2}{\tau^2 + n\sigma^2}(\sum_{i=1}^n y_i/\tau^2 - \mu/\sigma^2) \\
&= w_y \bar{y} + w_\mu \mu
\end{aligned}$$

der $\bar{y} = \sum_{i=1}^n y_i/n$, $w_y = \frac{n\sigma^2}{\tau^2 + n\sigma^2}$ og $w_\mu = \frac{\tau^2}{\tau^2 + n\sigma^2}$. Merk at $w_y + w_\mu = 1$, og at a posteriori-forventningen derfor er et vektet gjennomsnitt av a priori-forventningen og gjennomsnittet til målingene. Forventningen avhenger av de tre størrelsene n , σ og τ . Når $\sigma \rightarrow \infty$ dominerer data-fordelingen a posteriori-fordelingen, og $p(x | \mathbf{y}) \rightarrow \mathcal{N}(\bar{y}, \tau/n)$. Når $n \rightarrow \infty$ vil også data-fordelingen dominere. Alternativt, dersom $\sigma \rightarrow 0$ vil a priori-fordelingen være ganske lik a posteriori-fordelingen. Spesielt når dataene er usikre. A posteriori-forventningen uttrykkes ofte ved korreksjonen K , som sørger for en optimal

oppdatering av a priori-forventningen med hensyn på n , τ og σ slik:

$$\begin{aligned}
 \mathbb{E}(x | \mathbf{y}) &= \frac{\tau^2 \sigma^2}{\tau^2 + n\sigma^2} \left(\sum_{i=1}^n y_i / \tau^2 - \mu / \sigma^2 \right) \\
 &= \mu - \mu + \frac{\tau^2 \sigma^2}{\tau^2 + n\sigma^2} \left(\sum_{i=1}^n y_i / \tau^2 - \mu / \sigma^2 \right) \\
 &= \mu - \frac{\tau^2 + n\sigma^2}{\tau^2 + n\sigma^2} \mu + \frac{\tau^2 \sigma^2}{\tau^2 + n\sigma^2} \left(\sum_{i=1}^n y_i / \tau^2 - \mu / \sigma^2 \right) \\
 &= \mu + K(\bar{y} - \mu)
 \end{aligned} \tag{2.6}$$

hvor $K = \frac{n\sigma^2}{\tau^2 + n\sigma^2}$. Vi ser at a priori-forventningen blir justert mot gjennomsnittet til målingene. Derfor vektlegges dataene mye dersom $K \rightarrow 1$, mens når $K \rightarrow 0$ er a posteriori-forventningen ganske lik a priori-forventningen. A posteriori-variansen uttrykt ved K gir:

$$\begin{aligned}
 \text{Var}(x | \mathbf{y}) &= \frac{\tau^2 \sigma^2}{\tau^2 + n\sigma^2} + \frac{n\sigma^4}{n\sigma^2 + \tau^2} - \frac{n\sigma^4}{n\sigma^2 + \tau^2} \\
 &= \left(\frac{\tau^2 + n\sigma^2}{\tau^2 + n\sigma^2} - \frac{n\sigma^2}{\tau^2 + n\sigma^2} \right) \sigma^2 \\
 &= (1 - K)\sigma^2
 \end{aligned}$$

hvor vi ser at a posteriori-variansen er en oppdatert a priori-varians avhengig av K . Når vi uttrykker forventningen og variansen ved K , kommer det tydelig frem at dette i prinsippet er data assimilasjon. Justeringen K er optimal når a posteriori-fordelingen er gaussisk. Optimal i den forstand at den fører til at forventningen er forventningsrett, samtidig som den minimerer variansen. Dette kan bli vist ved å sette forventningen, i (2.6), inn i $\text{Var}(x | \mathbf{y})$. Så finne K der den deriverte er lik null. Hvordan vi oppdaterer en matematisk modell optimalt på sekvensielt vis kommer vi til senere i avsnittet om Kalmanfilter. I eksemplene under demonstrerer vi hvilken betydning variansen i dataene versus variansen i a priori-fordelingen har på a posteriori-fordelingen. Dette er eksperimentelle eksempler, der parameterverdiene ikke er basert på reelle data.

Eksempel 2.2.1: Presise data

For å gi et eksempel på hvordan vi vil estimere posisjonen til seismikkablene skal vi her se på en enkelt node. Vi skal estimere posisjonen noden har i krysslinjeretningen. Ønsket posisjon ligger på krysslinjen,

som tilsvarer verdien $x = 0$. Dette er også forventet posisjon. Én meter fra ønsket posisjon i retning vest, har verdien 1, og én meter fra ønsket posisjon retning øst har verdien -1 osv. Vi antar at $X \sim \mathcal{N}(0, 3^2)$, og $Y_i | x \sim \mathcal{N}(x, 2^2)$. Videre er målingene $\mathbf{y} = (0.1, 0.15)$, og $\bar{Y} | x \sim \mathcal{N}(x, 2^2/2)$. Da er a posteriori-fordelingen, definert ved forventning og varians gitt ved:

$$\begin{aligned}\mathbb{E}(x | \mathbf{y}) &= 0 + \frac{2 \cdot 9}{4 + 2 \cdot 9}(0.125 - 0) = 0.1023 \\ \text{Var}(x | \mathbf{y}) &= (1 - 18/22) \cdot 9 = 1.6364\end{aligned}$$

hvor $n = 2$, $\bar{y} = 0.125$ og $K = 18/22$. I figur 2.1 vises de tre forskjellige fordelingene grafisk.

Eksempelet over illustrerer hvilken konsekvens sikre målingene og usikker a priori-fordelingen har på a posteriori-fordelingen. Stor a priori-variens fører til stor K . Dette fører til at målingene blir vektlagt mye, og forventningen ligger derfor nærmest forventningen i data-fordelingen. Dessuten er a posteriori-variansen betraktelig redusert. Faktisk til ca. 19 % av variansen i a priori-fordelingen. Hadde flere målinger vært tilgjengelige, ville K vært enda høyere. Dermed har vi illustrert at på tross av en usikker modellen, vil sikre målinger føre til at estimatet av x blir sikkert.

Eksempel 2.2.2: Usikre data

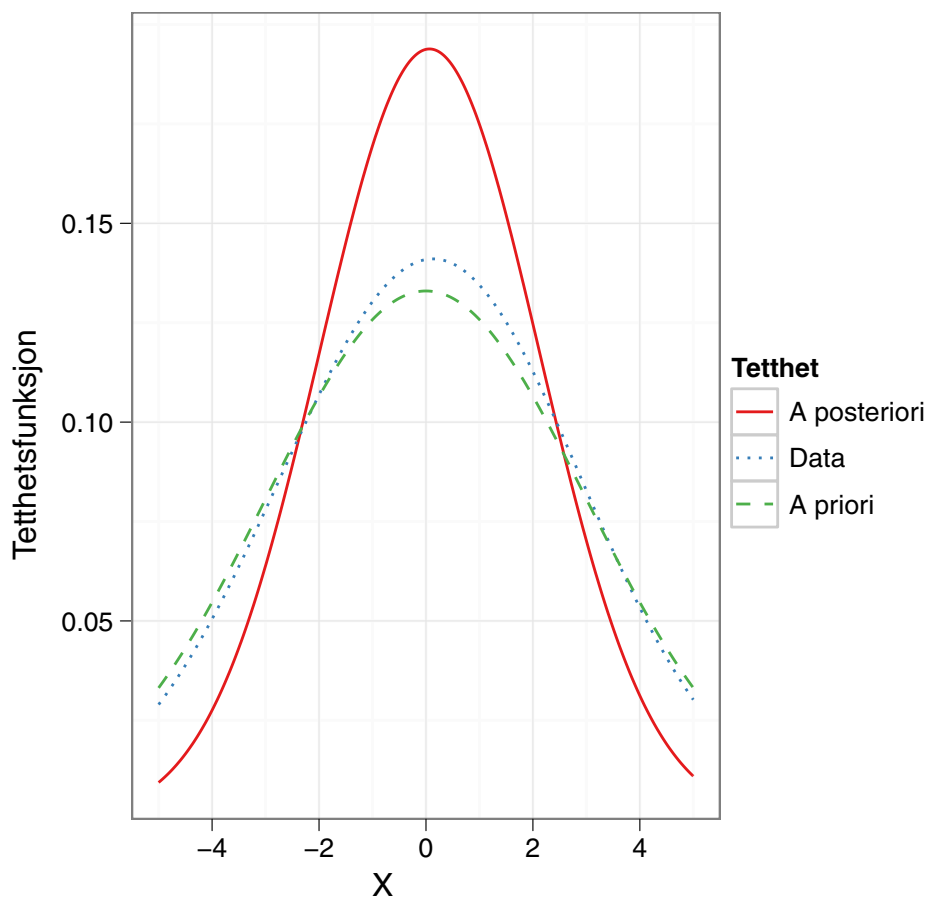
Vi beholder a priori-fordelingen over, og \mathbf{y} fra forrige eksempel, men antar at målingene er mer usikre. La $Y_i | x \sim \mathcal{N}(x, 4^2)$ og $\bar{Y} | x \sim \mathcal{N}(x, 4^2/2)$. Da får vi

$$\begin{aligned}\mathbb{E}(x | \mathbf{y}) &= 0 + \frac{2 \cdot 9}{16 + 2 \cdot 9}(0.025 - 0) = 0.0662 \\ \text{Var}(x | \mathbf{y}) &= (1 - 18/34) \cdot 9 = 4.2352.\end{aligned}$$

Den økte usikkerheten i dataene gir lavere $K = 18/34$, og a posteriori-forventningen er mer lik a priori-forventningen, som vist i figur 2.2. Variansen er redusert, men i mindre grad enn i forrige eksempel.

I avsnitt 2.2.1 antok vi at

$Y_i | x \sim \mathcal{N}(x, \tau^2)$ og $X | \mathbf{y}_i \sim \mathcal{N}(\mu + K(y_i - \mu), (1 - K)\sigma^2)$. Hvis vi setter



Figur 2.2: A posteriori-fordelingen, data-fordelingen og a priori-fordelingen fra eksempel 2.2.2.

$y_i = y$ nedenfor gir (2.4):

$$p(\tilde{y} | y) = \int \exp\left[-\frac{1}{2\tau^2}(\tilde{y} - x)^2\right] \exp\left[-\frac{1}{2\text{Var}(x | y)}(x - \mathbb{E}(x | y))^2\right] dx.$$

Siden eksponenten over er en kvadratisk funksjon av (x, \tilde{y}) , og $p(x, \tilde{y})$ er gaussisk, så er den marginale a posteriori-fordelingen til \tilde{y} også gaussisk. Siden vi vet at $\mathbb{E}(\tilde{y} | x) = x$ og $\text{Var}(\tilde{y} | x) = \tau^2$, så er forventningen og variansen til $p(\tilde{y} | y)$ gitt ved:

$$\mathbb{E}(\tilde{y} | y) = \mathbb{E}(\mathbb{E}(\tilde{y} | x, y) | y) = \mathbb{E}(x | y) = \mu + K(\bar{y} - \mu)$$

og:

$$\begin{aligned}\text{Var}(\tilde{y} | y) &= \mathbb{E}(\text{Var}(\tilde{y} | x, y) | y) + \text{Var}(\mathbb{E}(\tilde{y} | x, y) | y) \\ &= \mathbb{E}(\tau^2 | y) + \text{Var}(x | y) = \tau^2 + \text{Var}(x | y) = \tau^2 + (1 - K)\sigma^2.\end{aligned}$$

Altså er forventningen lik forventningen til a posteriori-fordelingen, mens variansen er lik variansen til a posteriori-fordelingen pluss variansen i data-fordelingen.

2.3 Multivariabel normalfordeling

I de fleste praktiske problemer er det flere enn én ukjent parameter, og det er spesielt da at fordelene med Bayesiansk inferens blir synlige. Selv om et praktisk problem inneholder flere parametre, er det oftest bare én parameter vi vil estimere eller predikere. I slike tilfeller er målet med Bayesiansk inferens å finne den marginale a posteriori-fordelingen til parameteren av primær interesse. Først finner vi simultan-fordelingen til alle parametrene, så integrere vi over nuisance parametrene, som kun er nødvendige for å konstruere en realistisk modell. I vår problemstilling kan fart og akselerasjon inngå i modellen som nuisance parametrene, der posisjonen er vår primære interesse.

Multivariabel gaussisk a priori- og data-fordeling gir sammen med Bayes' formel, grunnlaget for et multi-dimensjonalt Kalmanfilter. Den multivariable normalfordelingen er en generalisering av normalfordelingen, fra én dimensjon, $p = 1$, til flere dimensjoner, $p \geq 2$. Eksponenten i (2.6) som er kvadratet av avstanden mellom x og μ delt på standardavviket, kan uttrykkes litt annerledes slik:

$$J(x) = \left(\frac{x - \mu}{\sigma}\right)^2 = (x - \mu)(\sigma^2)^{-1}(x - \mu).$$

videre kan eksponenten, $J(x)$, generaliseres til flere dimensjoner

$$J(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (2.7)$$

der \mathbf{X} er en $p \times 1$ stokastisk vektor, som representerer et fler-variabelt utfall fra den multivariable normalfordelingen. Vektoren $\boldsymbol{\mu}$ er en $p \times 1$ vektor med forventningene til \mathbf{X} , og $\boldsymbol{\Sigma}$ er en $p \times p$ kovarians-matrise til de p variablene. Erstatte vi $J(x)$ med $J(\mathbf{x})$ i 2.7, og i tillegg bytter konstanten $\sqrt{2\pi\sigma^2}$ ut med

normaliseringskonstanten, $(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}$, ender vi opp med en multivariabel normalfordeling for vektoren $\mathbf{X} = (X_1, \dots, X_p)$ gitt ved:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp -(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2$$

der $-\infty < x_i < \infty$ for $i = 1, \dots, p$. Vi bruker $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ som notasjon for en p -dimensjonal multivariabel normalfordeling. De multivariable dataene fra n utfall organiseres i en $n \times p$ datamatrise slik

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & \dots & x_{1,p} \\ \vdots & \dots & \vdots \\ x_{n,1} & \dots & x_{n,p} \end{pmatrix}$$

der hver rekke presenterer et p -dimensjonalt utfall, og hver kolonne representerer n utfall på én variabel. En generell kovarians matrise for p variabler ser slik ut

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,p} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \sigma_{2,p} \\ \vdots & \dots & \dots & \vdots \\ \sigma_{p,1} & \sigma_{p,2} & \dots & \sigma_{p,p} \end{pmatrix}$$

der diagonalen består av variansene til de p variablene, $\sigma_{i,i} = \text{Var}(x_i)$, og ellers av kovarianser mellom de ulike variablene, $\sigma_{i,j} = \text{Cov}(x_i, x_j)$. Kovariansmatrisen er symmetrisk, som medfører at $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'$ siden $\sigma_{i,j} = \sigma_{j,i}$. Den er også positiv definit, $\mathbf{x}' \boldsymbol{\Sigma} \mathbf{x} > 0 \forall \mathbf{x} \neq 0$.

2.3.1 Bayes' Formel og Multivariabel normalfordeling

I avsnitt 2.6 fant vi ut at a posteriori-fordelingen for en variabel var gaussisk når a priori-fordelingen og data-fordelingen var gaussisk. Vi skal vise at det samme gjelder for flere enn én variabel, der a priori-fordelingen og datafordelingen er multivariable og gaussiske. Anta at vi er interessert i en $p \times 1$ vektor \mathbf{X} , for eksempel posisjonen til p forskjellige noder med ulik plassering på seismikk-

kabelen. Eller posisjon, fart og akselerasjon til en node med $p=3$. Fordelingen til \mathbf{X} er gitt ved a priori-fordelingen under:

$$\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{P}) \quad (2.8)$$

der $\boldsymbol{\mu}$ og \mathbf{P} antas å være kjent. I tillegg observerer vi en $k \times 1$ datavektor som antas å være gitt ved data-fordelingen

$$\mathbf{Y} | \mathbf{x} \sim \mathcal{N}_k(\mathbf{H}\mathbf{x}, \mathbf{R}) \quad (2.9)$$

hvor den $k \times p$ observasjonsmatrisen \mathbf{H} , og kovariansmatrisen \mathbf{R} , er kjente. Da er a posteriori-fordelingen gitt ved:

$$\mathbf{X} | \mathbf{y} \sim \mathcal{N}_p((\mathbf{H}'\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})^{-1}(\mathbf{H}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{P}^{-1}\boldsymbol{\mu}), (\mathbf{H}\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})).$$

Vi skal vise at dette er sant slik som vi gjorde for én variabel. Først multipliserer vi a priori-fordelingen og data-fordelingen i (2.8) og (2.9) og samler alle variablene i eksponenten mens alle konstantene inngår i C slik:

$$\begin{aligned} p(\mathbf{x} | \mathbf{y}) &\propto \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{P}^{-1}(\mathbf{x} - \boldsymbol{\mu}) + (\mathbf{y} - \mathbf{H}\mathbf{x})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x})\right) \\ &= \exp\left(-\frac{1}{2}(x'\mathbf{P}^{-1}x - x'\mathbf{P}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{P}^{-1}x + \boldsymbol{\mu}'\mathbf{P}^{-1}\boldsymbol{\mu} \right. \\ &\quad \left. + \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} - \mathbf{y}'\mathbf{P}^{-1}(\mathbf{H}\mathbf{x}) - (\mathbf{H}\mathbf{x})'\mathbf{R}^{-1}\mathbf{y} + (\mathbf{H}\mathbf{x})'\mathbf{R}^{-1}\mathbf{H}\mathbf{x})\right) \\ &= \exp\left(-\frac{1}{2}(x'(\mathbf{H}'\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})x - x'(\mathbf{H}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{P}^{-1}\boldsymbol{\mu}) \right. \\ &\quad \left. - (\mathbf{H}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{P}^{-1}\boldsymbol{\mu})'x + C)\right). \end{aligned}$$

Vi vet at eksponenten til en multivariabel normalfordeling skal se slik ut: $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$. Setter vi $\mathbf{A}^{-1} = (\mathbf{H}'\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}^{-1})$ og $\mathbf{B} = (\mathbf{H}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{P}^{-1}\boldsymbol{\mu})$, ser vi tydeligere at:

$$\begin{aligned} x'\mathbf{A}^{-1}x - x'\mathbf{B} - \mathbf{B}'x + C &= x'\mathbf{A}^{-1}x - x'\mathbf{A}^{-1}\mathbf{A}\mathbf{B} - (\mathbf{A}\mathbf{B})'\mathbf{A}^{-1}x + C \\ &= (x - \mathbf{A}\mathbf{B})'\mathbf{A}^{-1}(x - \mathbf{A}\mathbf{B}). \end{aligned}$$

I utledningen nedenfor, hvor forventningen og variansen er uttrykt ved \mathbf{K} , har jeg brukt Woodbury likheten fra Petersen (2008) der:

$$(\mathbf{P}^{-1} + \mathbf{H}'\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{P}^{-1}\boldsymbol{\mu} = (\mathbf{P} - \mathbf{P}\mathbf{H}'(\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}')^{-1}\mathbf{H}\mathbf{P})\mathbf{P}^{-1}\boldsymbol{\mu} \text{ og}$$

$$(\mathbf{P}^{-1} + \mathbf{H}'\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}'\mathbf{R}^{-1}\mathbf{y} = \mathbf{P}\mathbf{H}'(\mathbf{H}\mathbf{P}\mathbf{H}' + \mathbf{R})^{-1}\mathbf{y}. \text{ Den siste likningen gjel-}$$

der kun når P og R er positivt definitte matriser. Vi kunne brukt den siste likheten i begge tilfellene siden både P og R er kovariansmatriser, men jeg bruker begge to her for enkelthetens skyld der:

$$\begin{aligned}
\mathbb{E}(x | y) &= (P^{-1} + H'R^{-1}H)^{-1}(P^{-1}\mu + H'R^{-1}y) \\
&= (P^{-1} + H'R^{-1}H)^{-1}P^{-1}\mu \\
&\quad + (P^{-1} + H'R^{-1}H)^{-1}H'R^{-1}y \\
&= (P - PH'(R + HPH')^{-1}HP)P^{-1}\mu \\
&\quad + PH'(HPH' + R)^{-1}y \\
&= \mu - (PH'(R + HPH')^{-1}HP)P^{-1}\mu \\
&\quad + PH'(HPH' + R)^{-1}y \\
&= \mu - KH\mu + Ky \\
&= \mu + K(y - H\mu)
\end{aligned}$$

hvor $K = PH'(HPH' + R)^{-1}$ er en $p \times k$ matrise, som korrigerer optimalt når a posteriori-fordelingen er gaussisk. Variansen uttrykt ved K , der Woodbury likheten benyttes igjen, er gitt ved:

$$\begin{aligned}
\text{Var}(x | y) &= (H'R^{-1}H + P^{-1})^{-1} \\
&= P - PPH'(R + HPH')^{-1}HP \\
&= (I - KH)P.
\end{aligned}$$

Hverken eksempel 2.3.1 og 2.3.2 er basert på reelle data. For øvrig, siden eksemplene er tatt med for å illustrere at målinger av akselerasjon gir sikrere inferens om posisjonen til noden, så har vi valgt disse parameter-verdiene uten å reflektere spesielt grundig over dette valget. Viere har vi ikke undersøkt i hvor stor grad parametrene, posisjon, fart og akselerasjon, faktisk er korrelert. Men de korrelert, siden akselerasjon er endring i fart osv.

Eksempel 2.3.1: Én måling av posisjon

Vi er fremdeles ute etter å estimere posisjonen til en node på seismikk-kabelen. For å prøve å styrke inferensen om posisjonen til noden, inkluderer vi parametrene, fart og akselerasjon i a priori-fordelingen. La posisjon (m), fart (m/s) og akselerasjon (m/s^2) til noden i krysslinjeretning, $X = [X_p, X_f, X_a]$, ha a priori-fordeling, $X \sim \mathcal{N}_3(\mu, P)$, der $\mu = [0, 2, 0]$ og

$$\mathbf{P} = \begin{bmatrix} 1 & 0.02 & 0.001 \\ 0.02 & 1 & 0.01 \\ 0.001 & 0.01 & 0.1 \end{bmatrix}.$$

Vi antar at det er én måling tilgjengelig av posisjonen til noden, $y_p = 0.2$, som har data-fordeling, $Y_p | \mathbf{x} \sim \mathcal{N}(\mathbf{H}\mathbf{x}, \sigma_p^2)$, der

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \text{ og } \sigma_p^2 = 1.$$

Vi starter med å finne justeringen \mathbf{K} :

$$\begin{aligned} \mathbf{K} &= \begin{bmatrix} 1 & 0.02 & 0.001 \\ 0.02 & 1 & 0.01 \\ 0.001 & 0.01 & 0.1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0.02 & 0.001 \\ 0.02 & 1 & 0.01 \\ 0.001 & 0.01 & 0.1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0.02 \\ 0.001 \end{bmatrix} \left(\mathbf{1} + \begin{bmatrix} 1 & 0.02 & 0.001 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 0.5 \\ 0.01 \\ 0.0005 \end{bmatrix}. \end{aligned}$$

Som forventet er vekten i rad 1, som vektet parameteren det er måling av, vesentlig større enn vekten i rad 2 og 3. Vekten er lik 0.5 siden det er én måling, med samme varians som a-priori-fordelingen. Da blir a posteriori-forventningen lik gjennomsnittet av målingen og a priori forventningen. Verdiene i \mathbf{K} som vektet fart og akselerasjon er ulik null fordi fart og akselerasjon er korrelert med posisjon. Naturlig nok vektet fart, som er mest korrelert med posisjon, målingen mest. Gitt \mathbf{K} finner vi a posteriori-

forventningen gitt ved:

$$\begin{aligned} \mathbb{E}(x | y) &= \mu + K(y_p - H\mu) \\ &= \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.01 \\ 0.0005 \end{bmatrix} (0.2 - [1 \ 0 \ 0] \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}) \\ &= \begin{bmatrix} 0.1000 \\ 2.0020 \\ 0.0001 \end{bmatrix}. \end{aligned}$$

Forventningen er gjennomsnittet av målingen og a priori-forventningen. Fart og akselerasjon blir også justert etter i hvor stor grad de er korrelert med posisjon, og har en høyere forventning. Under, ser vi at variansen til posisjonen er halvvvert. Variansen til fart og akselerasjon er også redusert, siden målingen tilfører ekstra informasjon til disse parametrene også. Variansen er gitt ved:

$$\begin{aligned} \text{Var}(x | y) &= (I - KH)P \\ &= \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5 \\ 0.01 \\ 0.0005 \end{bmatrix} [1 \ 0 \ 0] \right) \begin{bmatrix} 1 & 0.02 & 0.001 \\ 0.02 & 1 & 0.01 \\ 0.001 & 0.01 & 0.1 \end{bmatrix} \\ &= \begin{bmatrix} 0.5 & 0.01 & 0.0005 \\ 0.01 & 0.9980 & 0.0099 \\ 0.0005 & 0.0099 & 0.0999 \end{bmatrix}. \end{aligned}$$

Ergo gir a posteriori-fordelingen sikrere estimat enn a priori-fordelingen for alle parametrene gitt én måling av posisjonen.

I neste eksempel illustrerer vi at målinger av akselerasjon gir forbedret estimat av posisjon. I hvor stor grad akselerasjonsmålingene vil øke sikkerheten er avhengig av hvor sikre målingene er, og i hvor stor grad m er korrelert med m/s^2 . I dette eksempelet gis det to målinger. I praksis vil det kun være en måling tilgjengelig ved hvert tidspunkt, og vi antar at målingene er uavhengige.

Eksempel 2.3.2: Målinger av posisjon og akselerasjon

I dette eksempelet er a priori fordelingen lik den over, men nå er to målinger av både posisjon og akselerasjon tilgjengelige. De er gitt ved:

$$\mathbf{y} = \begin{bmatrix} y_{1,p} & y_{1,a} \\ y_{2,p} & y_{2,a} \end{bmatrix} = \begin{bmatrix} 1 & 0.02 \\ 0.8 & 0.01 \end{bmatrix}.$$

Datafordelingen til $\mathbf{Y} = [Y_p, Y_a]$ er gitt ved $\mathbf{Y} | \mathbf{x} \sim \mathcal{N}(\mathbf{H}\mathbf{x}, \mathbf{R})$, der:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ og } \mathbf{R} = \begin{bmatrix} 1 & 0.02 \\ 0.02 & 0.1 \end{bmatrix}.$$

Vanligvis antar vi at korrelasjonen mellom målinger er lik 0. I denne oppgaven derimot, er akselerasjonsdata ikke tilgjengelige. Derfor er ikke målingene av akselerasjon uavhengige av målingene av posisjon. Vi finner først \mathbf{K} gitt ved:

$$\begin{aligned} \mathbf{K} &= \begin{bmatrix} 1 & 0.02 & 0.001 \\ 0.02 & 1 & 0.01 \\ 0.001 & 0.01 & 0.1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \left(\begin{bmatrix} 1 & 0.02 \\ 0.02 & 0.1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.02 & 0.001 \\ 0.02 & 1 & 0.01 \\ 0.001 & 0.01 & 0.1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \\ &= \begin{bmatrix} 1 & 0.001 \\ 0.02 & 0.01 \\ 0.001 & 0.1 \end{bmatrix} \\ &= \left(\begin{bmatrix} 1 & 0.02 \\ 0.02 & 0.1 \end{bmatrix} + \begin{bmatrix} 1 & 0.02 & 0.001 \\ 0.001 & 0.01 & 0.1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \\ &= \begin{bmatrix} 0.5005 & -0.04755 \\ 0.0095 & 0.0490 \\ -0.0048 & 0.5005 \end{bmatrix}. \end{aligned}$$

Når vi har tre parametre og to-dimensjonale målinger, blir \mathbf{K} en 3×2 matrise. I første rad har \mathbf{K} en høy verdi og en lav verdi. Den høye vektor gjennomsnittet av målingene til posisjonen, mens den lave vektor gjennomsnittet av akselerasjonen. I andre rad er begge vektene lave, siden det ikke er måling av fart. Til slutt, i tredje rad ser vi at den siste vekten, som korresponderer til akselerasjonsmålingen, er høy. Videre finner vi a

posteriori-fordelingen definert ved :

$$\begin{aligned} \mathbb{E}(x | y) &= \mu + K(\bar{y} - H\mu) \\ &= \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.5005 & -0.04755 \\ 0.0095 & 0.0490 \\ -0.0048 & 0.5005 \end{bmatrix} \left(\begin{bmatrix} 0.9 \\ 0.06 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} \right) \\ &= \begin{bmatrix} 0.4476 \\ 2.0115 \\ 0.0258 \end{bmatrix} \end{aligned}$$

Vi ser at forventningen til alle parametrene er større enn de var i a priori-fordelingen, grunnet høyere verdier i målingene. Parametrene, posisjon og akselerasjon, har vektet sine tilhørende målinger mest. Som vi så ovenfor vektet parametrene målinger på grunnlag av korrelasjonen til målingen. Målingen av akselerasjon gir derfor ekstra informasjon til posisjonen osv.

$$\begin{aligned} \text{Var}(x | y) &= (I - KH)P \\ &= \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5005 & -0.04755 \\ 0.0095 & 0.0490 \\ -0.0048 & 0.5005 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 & 0.02 & 0.001 \\ 0.02 & 1 & 0.01 \\ 0.001 & 0.01 & 0.1 \end{bmatrix} \\ &= \begin{bmatrix} 0.4995 & 0.0105 & 0.0053 \\ 0.0105 & 0.9993 & 0.0051 \\ 0.0053 & 0.0051 & 0.0500 \end{bmatrix} . \end{aligned}$$

Desto mer målingene blir vektet, desto større er reduseringspotensialet i a posteriori-variansen til de tilhørende parametrene. I vårt tilfelle er det $K[1,1]$ og $K[3,2]$ som sørger for at variansen til posisjon og akselerasjon omtrent blir halvert. Variansen til fart blir også redusert, men vesentlig mindre.

3

Tidsrekkeanalyse

Vi bruker tidsrekkemodeller til å modellere bevegelsen en gjenstand har over tid, der vi vil modellere posisjonen en node har over tid. Vi skiller mellom stasjonære og ikke-stasjonære tidsrekker. Først predikerer vi posisjon og akselerasjon på ikke-stasjonære tidsrekker, med Parabelmetoden. Så transformeres tidsrekkene slik at de blir stasjonære. Videre brukes Box-Jenkins metodikken til å finne passende $ARIMA(p, d, q)$ -modeller, der vi velger endelig modell basert på informasjonskriteriene, AIC og BIC . Med disse modellene predikerer vi akselerasjon og posisjon til noden. Til slutt sammenligner vi prediksjonsevnen til de ulike tidsrekke-modellene med Parabelmetoden.

3.1 Tidsrekketeori

En tidsrekke er en følge med observasjoner der hver observasjon, y_t , er målt ved en bestemt tid t . La

$$\mathbf{y}_{1:T} = \{y_1, \dots, y_T\}$$

være er en realisasjon av den stokastiske tidsrekken,

$$Y_{1:T} = \{Y_1, \dots, Y_T\}.$$

Den tilhørende stokastiske tidsrekken til den latente prosessen er gitt ved:

$$X_{0:T} = \{X_0, \dots, X_T\}$$

der X_0 er a priori-forventningen til den latente variabelen før målinger er tilgjengelig. En realisasjon av tidsrekken er gis ved:

$$x_{0:T} = \{x_0, \dots, x_T\}$$

der intervallene er uniforme, og T indikerer lengden på tidsintervallet. Dette er en diskret-tid prosess, siden tidene til observasjonene er diskrete. Kontinuerlig-tid prosess, er en prosess der tidene er kontinuerlige på et område, f. eks $T \in [0,1]$. Vi forholder oss til diskrete tidsrekker med uniforme tidsintervall.

En tidsrekke med en følge stokastiske variabler, $\{X_1, \dots, X_T\}$, kan uttrykkes som simultan-fordelingen til den stokastiske vektoren, (X_1, \dots, X_T) slik:

$$p(X_1 < x_1, \dots, X_T < x_T), \quad -\infty < x_1, \dots, x_T < \infty, \quad T \geq 1.$$

Simultanfordelingen for hele tidsrekken er sjelden kjent i tidsrekke-teori, siden den oftest inneholder for mange parametre å estimere på grunnlag av tilgjengelige data. Stort sett estimeres kun første og andre ordens-momentene til simultanfordelingen, ofte kalt for *andre-ordens* egenskapene, $\mathbb{E}(X_t)$ og $\mathbb{E}(X_t X_{t+h})$ for $t = 1, 2, \dots, T$ og $h = 0, 1, \dots, T - 1$. Generelt mister vi informasjon ved å kun se på andre-ordens egenskapene, men dersom simultanfordelingen definert over er en multivariabel normalfordeling, blir tidsrekken fullstendig beskrevet av disse. I definisjon 3.1.1 er andre-ordens egenskapene definert. Forventningsfunksjonen er forventningen til tidsrekken ved tiden t , og autokovariansfunksjonen er et mål på den lineære avhengigheten mellom to punkt på tidsrekken ved ulike tidspunkt, r og s .

Definisjon 3.1.1: Autokovariansfunksjon

La X_t være en tidsrekke hvor $\mathbb{E}(X_t^2) < \infty$. Forventnings- og variansfunksjonen til X_t er da gitt ved

$$\begin{aligned} \mu_X(t) &= \mathbb{E}(X_t) \\ \sigma_X^2(t) &= \text{Var}(X_t) \end{aligned}$$

Autokovariansfunksjonen til X_t er

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = \mathbb{E}(X_r - \mu_X(r))(X_s - \mu_X(s)), \quad (3.1)$$

for alle r og s .

Deler vi 3.1 på $\sigma_X(r)\sigma_X(s)$ finner vi autokorrelasjonsfunksjonen $\rho_X(r, s)$. Vi kommer tilbake til autokorrelasjonsfunksjonen når vi omtaler stasjonære tidsrekker.

3.2 Stasjonære tidsrekker

En stasjonær tidsrekke har egenskaper som er konstante i tid. Mange tidsrekker er ikke-stasjonære, men de fleste kan allikevel bli det ved å transformere dem. I prediksjon gir stasjonærhet store fordeler, fordi man predikerer at de konstante egenskapene forblir konstante i fremtiden. Vi deler stasjonærhet inn i svak, og sterk. Når vi bruker stasjonærhet i oppgaven mener vi alltid svak stasjonærhet. La oss først definere svak stasjonærhet:

Definisjon 3.2.1: Svak stasjonærhet

Tidsrekken X_t er svakt stasjonær hvis

$$i) \mathbb{E}(X_t^2) < \infty$$

og hvis

$$ii) \mathbb{E}(X_t)$$

$$iii) \text{Cov}(X_t, X_{t+h})$$

er uavhengig av tid for alle h .

Svak stasjonærhet blir ofte kalt; andre-ordens stasjonærhet, siden det bare er andre-orden egenskapene som antas å være stasjonære her. Sterk stasjonærhet medfører at hele simultan-fordelingen til tidsrekken er uavhengig av tid. Når fordelingen er gaussisk, er svak og sterk stasjonærhet ekvivalent.

Definisjon 3.2.2: Sterk stasjonærhet

Tidsrekken X_t er sterkt stasjonær hvis simultan-fordelingen til $X_{t:t+h}$ ikke avhenger av t for noen $h \geq 0$.

Når X_t er stasjonær, så er autokovariansfunksjonen (ACVF) bare avhengig av h , og kan uttrykkes slik:

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}).$$

Det samme gjelder for autokorrelasjonsfunksjonen (ACF) der

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

der $\gamma(0)$ er variansen. ACF'en måler styrken på lineær avhengighet mellom to variabler i tidsrekken, på en skala fra -1 til 1. Når variablene har tendens til å variere på samme side av forventningen, er ACF'en positiv. Men varierer de på motsatt side av forventningen er den negativ. Når vi estimerer ACVF'en bør vi bruke minst $T = 50$ ledd fra tidsrekken. Empirisk forventning, AVCF og ACF er henholdsvis gitt ved:

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{t=1}^T x_t \\ \hat{\gamma}(h) &= \frac{1}{T} \sum_{t=1}^{T-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad h < T \\ \hat{\rho}(h) &= \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad h < T.\end{aligned}$$

For identiske uavhengige residualer (iid) og hvit støy (WN), se definisjon 3.2.3, er $\hat{\rho}(h)$ tilnærmet $\mathcal{N}(0, 1/T)$, når T er stor. Derfor forventer vi at 95% av autokorrelasjonene til støyen ligger innenfor konfidensintervallet, $\pm 1.96/\sqrt{T}$. Vi bruker konfidensintervall på ACF, for å sjekke om residualene er ukorrelerte, og for å sjekke tilpasningsdyktigheten til en modell. I tidsrekker med trend, avtar ACF'en sakte mot 0. I tidsrekker med sesongkomponent, vil ACF'en ha regelmessige svingninger. ACF'en omtales mer i avsnittet om Box Jenkins teori. Den partielle ACF'en spiller også en sentral rolle i tidsrekkeanalyse. Her er den gitt ved:

$$\text{corr}(X_t, X_{t+h} \mid \mathbf{x}_{t+1:t+(h-1)}) = \frac{\text{Cov}(X_t, X_{t+h} \mid \mathbf{x}_{t+1:t+(h-1)})}{\sqrt{\text{Var}(X_t \mid \mathbf{x}_{t+1:t+(h-1)}) \text{Var}(X_{t+h} \mid \mathbf{x}_{t+1:t+(h-1)})}},$$

der $\mathbf{X}_{t:t+h} = \{X_t, X_{t+1}, \dots, X_{t+h}\}$ og $h > 0$. Vi ser at PACF'en er korrelasjonen mellom to variabler gitt de mellomliggende variablene. Fullstendig definisjon av PACF gis i definisjon 3.2.8. Denne spiller en viktig rolle når man identifiserer AR-parametre i ARMA-modellen.

Residualer, ofte referert til som støy, er avviket mellom observert verdi og verdien modellen estimerer. I definisjon 3.2.5 er Z_t residualen. For en stasjonær tidsrekke skal en akseptabel modell ha residualer som er iid eller hvit støy. Begge er definert under.

Definisjon 3.2.3: Hvit og iid støy

Tidsrekken Y_t er en hvit støyprosess hvis den er svakt stasjonær med forventning null, og variablene er ukorrelerte. Vi skriver:

$$Y_t \sim WN(0, \sigma^2).$$

Tidsrekken Y_t er en iid støyprosess hvis den er strengt stasjonær med forventning lik null, og variablene er uavhengige. Vi skriver:

$$Y_t \sim iid(0, \sigma^2).$$

Residualene skal ha gjennomsnitt omtrent lik 0, konstant varians og være ukorrelerte eller uavhengige. Dersom 95 % av verdiene til ACF'en faller innenfor konfidensintervallet, tyder dette på at residualene er ukorrelerte, eller uavhengige. Vi ønsker å teste om residualene er tilnærmet normalfordelt. Dette tester vi ved å plote residualene i et histogram og med Kolmogorov-Smirnov test. Dersom residualene avviker fra disse antakelsene, og inneholder en struktur som modellen ikke beskriver, da må vi finne en ny modell, eventuelt legge til flere parametre i modellen.

3.2.1 ARMA(p, q)-modeller

Vi skal se på en stor klasse lineære tidsrekkemodeller kalt *Autoregressive Moving average modeller*, ARMA(p, q)-modeller. ARMA-modeller brukes på lineære stasjonære tidrekker. Dersom tidsrekken er ikke-lineære kan den ofte lineariseres, f. eks. ved å ta logaritmen $\log(X \cdot Y) = \log(X) + \log(Y)$. Men ofte er det naturlige forholdet mellom variabler tilnærmet lineært. En lineær prosess defineres slik:

Definisjon 3.2.4: Lineær prosess

En *lineær prosess* er en prosess som kan skrives på formen:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

hvor $Z_t \sim WN(0, \sigma^2)$ og $\{\psi_j\}$ er en følge med konstanter slik at $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

ARMA-modeller er lineære prosesser som blir tilpasset stasjonære tidsrekker. Den er gitt ved:

Definisjon 3.2.5: ARMA(p, q)-modellen

Tidsrekken X_t er en *ARMA-modell* hvis den er stasjonær og tilfredstiller:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad \forall t,$$

hvor $Z_t \sim WN(0, \sigma^2)$ og polynomene $(1 - \phi_1 z - \dots - \phi_p z^p)$ og $(1 + \theta_1 z + \dots + \theta_q z^q)$ ikke har noen felles faktorer.

Legg merke til at X_t må være svakt stasjonær. Dette kravet er oppfylt dersom $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \forall |z| = 1, z \in \mathbf{C}$. Altså må $\phi(z)$ være ulik null for alle z som ligger på enhets sirkelen i det komplekse planet (\mathbf{C}). Da har X_t også en unik løsning. Det er vanlig å skrive ARMA(p, q)-modellen på en kompakt form slik:

$$\phi(B)X_t = \theta(B)Z_t,$$

hvor $B^j X_t = X_{t-j}$ og $B^j Z_t = Z_{t-j}$. Videre er polynomene $\phi(\cdot)$ og $\theta(\cdot)$ lik $\phi(h) = (1 - \phi_1 z - \dots - \phi_p z^p)$ og $\theta(h) = (1 + \theta_1 z + \dots + \theta_q z^q)$. Når $\phi(z) \equiv 1$ blir tidsrekken kalt for en *autoregressiv-modell* av orden p , $AR(p)$, og når $\theta(z) \equiv 1$ en *glidende gjennomsnitt-modell* av orden q , $MA(q)$.

Dersom ARMA-modellen er *kausalt* kan X_t uttrykkes ved Z_s for $s \leq t$, og X_t er uavhengig av de fremtidige residualene.

Definisjon 3.2.6: Kausalitet

ARMA(p, q)-modellen, X_t , er *kausalt* hvis det eksisterer konstanter ψ_i slik at $\sum_{i=0}^{\infty} |\psi_i| < \infty$ og

$$X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i} \quad \forall t.$$

Dette er sant hvis:

$$\phi(z) = (1 - \phi_1 z - \dots - \phi_p z^p) \neq 0 \quad |z| \leq 1 \quad z \in \mathbf{C}.$$

Det vil si at når alle røttene til $\phi(z)$ ligger utenfor enhets sirkelen, er ARMA(p, q)-modellen kausal. Vi ser at en kausal modell kan uttrykkes som MA(∞). Vi ser også at kausalitet er en forutsetning i prediksjon.

Definisjon 3.2.7: Invertibilitet

ARMA(p, q)-modellen, X_t er invertibel hvis det eksisterer konstanter π_i slik at $\sum_{i=0}^{\infty} |\pi_i| < \infty$ og

$$Z_t = \sum_{i=0}^{\infty} \pi_i X_{t-i} \quad \forall t.$$

Dette er sant hvis

$$\theta(z) = (1 + \theta_1 z + \dots + \theta_q z^q) \neq 0 \quad |z| \leq 1 \quad z \in \mathbf{C}.$$

Parametrene i definisjonene over, ψ_i og π_i , er gitt ved:

$$\begin{aligned} \psi_i - \sum_{k=1}^p \phi_k \psi_{i-k} &= \theta_i \quad j = 0, 1, \dots \\ \pi_i + \sum_{k=1}^q \theta_k \pi_{i-k} &= \phi_j \quad j = 0, 1, \dots \end{aligned}$$

hvor $\theta_0 = 1$, $\theta_j = 0$ for $j > p$ og $\psi_j = 0$ for $j < 0$. Videre er $\phi_0 = -1$, $\phi_j = 0$ for $j > p$ og $\pi_j = 0$ for $j < 0$.

I forrige avnitt introduserte vi PACF'en. For ARMA-modeller er partiell ACF fullstendig definert ved:

Definisjon 3.2.8: PACF

PACF'en til en ARMA-modell, X_t , er funksjonen $\alpha(\cdot)$ definert ved:

$$\begin{aligned} \alpha(0) &= 1 \\ \alpha(h) &= \phi_{hh} \quad h \geq 1 \end{aligned}$$

hvor ϕ_{hh} er den siste komponenten i

$$\phi_h = \Gamma_h^{-1} \gamma_h,$$

der

$$\Gamma_h^{-1} = \begin{pmatrix} \gamma(0) & \dots & \gamma(h-1) \\ \vdots & \dots & \vdots \\ \gamma(h-1) & \dots & \gamma(0) \end{pmatrix} \quad (3.2)$$

og

$$\gamma_h = [\gamma(1), \dots, \gamma(h)].$$

Lite overraskende, er det stor sammenheng mellom verdiene i PACF'en og AR-parametrene. Parametrene i en kausal AR(p)-prosess, ϕ_h for $h = 1, \dots, p$, er like $\alpha(h)$, der $\alpha(h) = 0$ for $h > p$. Således gir PACF'en verdifull informasjon om hvor mange AR-parametre som skal inkluderes i modellen.

Ofte viser det seg at ACF'en avtar sakte, mens PACF'en avtar raskt mot null. Autokorrelasjonen mellom x_t og x_{t-1} er antakeligvis lik autokorrelasjonen mellom x_{t-1} og x_{t-2} . Derfor forventer vi også å finne autokorrelasjonen mellom x_t og x_{t-2} , mens den partielle autokorrelasjonen kan være tilnærmet null her. Det gir derfor bare mening å regressere X_t på x_{t-1} og x_{t-2} , dersom den partielle autokorrelasjonen mellom X_t og disse leddene er ulik null. Vi kommer tilbake til dette i Box Jenkins analysen.

Siden AR(1)-modellen er en viktig komponent i Kalmanfilteret, ser vi på noen av egenskapene til denne modellen.

Eksempel 3.2.9: AR(1)-modellen

Tidsrekken X_t er en AR(1)-modell hvis den er stasjonær og tilfredstiller:

$$X_t = \phi X_{t-1} + Z_t \quad \forall t$$

hvor $Z_t \sim WN(0, \sigma^2)$, og $|\phi| \neq \pm 1$.

3.2.2 Momenter i AR(1)-modellen

Her ser vi på andre-ordens egenskapene til AR(1). Som sagt er modellen kausal og stasjonær når $|\phi| < 1$. Stasjonæritet fører til at andre-ordens egenskapene til X_t er kjent $\forall t = 1, \dots, T - 1$ dersom de er kjent for X_0 . Samme gjelder for η_t , siden vi antar at residualene er iid. Dersom X_0 er normalfordelt, med $\mathbb{E}(X_0) = 0$ og $\text{Var}(X_0) = \frac{\tau^2}{1-\phi^2}$, og $\eta_t \sim \mathcal{N}(0, \tau^2)$, finner vi forventningen til X_t rekursivt $\forall t$ slik:

$$\mathbb{E}(X_t) = \mathbb{E}(\phi X_{t-1}) + \mathbb{E}(\eta_t) = \phi \mathbb{E}(X_{t-1}) = \phi 0 = 0.$$

Resultatene for variansen og kovariansen nedenfor, er hentet fra Petersone (2009). Variansen til X_t er også uavhengig av tiden, og derfor gitt ved:

$$\begin{aligned}\text{Var}(X_t) &= \mathbb{E}(X_t^2) \\ &= \mathbb{E}((\phi X_{t-1} + \eta_t)^2) \\ &= \mathbb{E}((\phi X_{t-1})^2 + \eta_t^2 + 2\eta_t \phi X_{t-1}) \\ &= \phi^2 \mathbb{E}(X_{t-1}^2) + \mathbb{E}(\eta_t^2) \\ &= \phi^2 \mathbb{E}(X_{t-1}^2) + \tau^2 \\ (1 - \phi^2) \text{Var}(X_t) &= \tau^2\end{aligned}$$

der $\mathbb{E}(\eta_t^2) = \text{Var}(\eta_t) + \mathbb{E}(\eta_t)^2$ og $\mathbb{E}(X_{t-1}^2) = \text{Var}(X_{t-1}) + \mathbb{E}(X_{t-1})^2$ og:

$$\text{Var}(X_t) = \frac{\tau^2}{1 - \phi^2}.$$

Vi har funnet forventningen og variansen til en AR(1)-modell, og vi ser at begge er uavhengig av tid. Kovariansen og autokorrelasjonen forandrer seg heller ikke i tid, gitt tids-uavhengig forventning og varians.

$$\begin{aligned}\text{Cov}(X_t, X_{t+1}) &= \mathbb{E}(X_t X_{t+1}) - \mathbb{E}(X_t) \mathbb{E}(X_{t+1}) \\ &= \mathbb{E}(X_t (\phi X_t + \eta_{t+1})) \\ &= \phi \mathbb{E}(X_t^2).\end{aligned}$$

$$\text{Cov}(X_t, X_{t+1}) = \frac{\phi \tau^2}{1 - \phi^2}.$$

Siden $\text{Var}(X_t) = \frac{\tau^2}{1-\phi^2}$, får vi et pent uttrykk for autokorrelasjonen:

$$\text{corr}(X_t, X_{t+1}) = \frac{\text{Cov}(X_t, X_{t+1})}{\text{Var}(X_t)} = \phi.$$

Vi ser at $\rho(1) = \phi$, som viser at AR-modellen vekter leddene etter i hvor stor grad de er korrelerte med X_t . Denne er gitt ved PACF'en når modellen er kausal.

3.2.3 Prediksjon

Gitt en stasjonær tidsrekke $X_{0:n} = \{X_0, \dots, X_n\}$, med forventning μ og kovariansfunksjon γ , vil vi predikere X_{n+h} for $h > 0$. Målet er å finne en lineær kombinasjonen av $1, X_n, \dots, X_1, P_n X_{n+h}$, som minimerer:

$$\mathbb{E}(X_{n+h} - P_n X_{n+h})^2, \quad (3.3)$$

der $P_n X_{n+h} = a_0 + a_1 X_n + \dots + a_n X_n$. Vi finner beste lineære prediktor ved å derivere (3.3) med hensyn på alle vektene i P_n .

For rene AR- og MA-modeller brukes henholdsvis, Durbin Levinson's- og Innovasjons-Algorithmen i prediksjon. Innovasjons-Algorithmen brukes også på ARMA-modeller, der den forenkles ved å transformere X_t . Vi skal ikke gå mer detaljert inn på dette. Endelig, predikerer vi kausale ARMA-modeller rekursivt med:

$$P_n X_{n+h} = \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}), \quad (3.4)$$

for $h \geq 1$. Prediksjonen er en lineær kombinasjon av tidligere ledd og residualer i tidsrekken. Legg merke til at det gir bare mening å predikere med en kausal ARMA-modell. Som sagt er stasjonæritet viktig i prediksjon, siden vi predikerer at de konstante egenskaper forblir konstante. I tilfeller hvor dataene må transformeres for å bli stasjonære, blir predikerte verdier transformert tilbake. Vi kommer tilbake til dette under ikke-stasjonære tidsrekker. Der sammenligner vi prediksjonsevnen til ARIMA-modeller versus Parabelmetoden.

3.3 Ikke-stasjonære tidsrekker

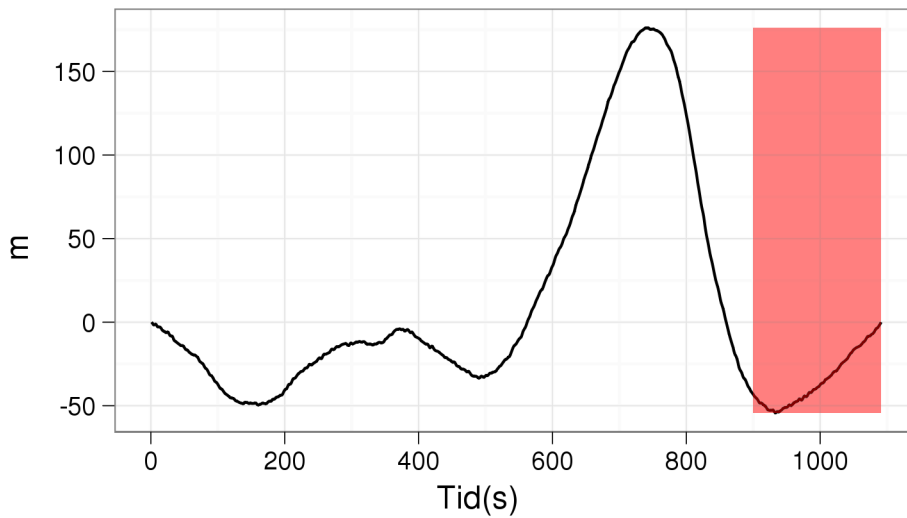
Mange tidsrekker er opprinnelig ikke-stasjonære, og det kreves ofte minst én matematisk transformasjon før de blir stasjonære. ARIMA-modeller er egnet på slike tidsrekker. Målet med ARIMA-modellen er å *differensiere* tidsrekken inntil den er stasjonær. Når dette er oppnådd, blir en ARMA-modell tilpasset de differensierte dataene. Antall parametre i ARMA-modellen finner vi på en systematisk måte i henhold til Box-Jenkins metodikken. Den optimale modellen velger vi ut fra informasjonskriteriene *AIC* og *BIC*. Parametrene i den valgte modellen, ϕ , ψ og σ^2 , er maksimum likelihood estimat. Til slutt, tester vi om residualene er tilnærmet normalfordelte.

3.3.1 Tidsrekke 1 og 2

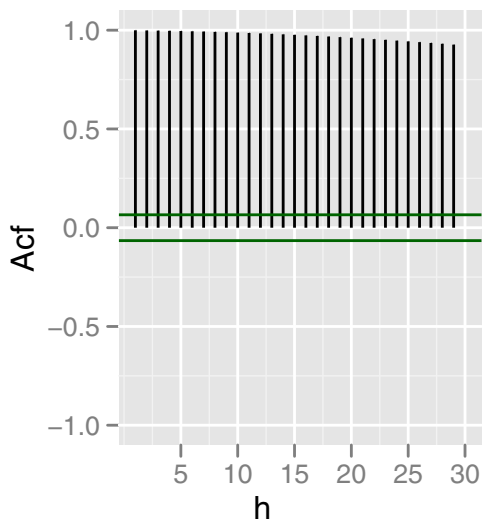
I figur 3.1 og 3.2 er henholdvis Tidsrekke 1 og Tidsrekke 2 plottet, med sine tilhørende ACF og PACF. Dataene er samlet inn i 2005 utenfor Vietnam under en seimisk undersøkelse. Tidsrekkene viser evolusjonen til en node over tid. Posisjonene til noden er avstanden noden har til den definerte krysslinjen, $x = 0$. Posisjonene er målt i meter. Som sagt, positive verdier, $x > 0$, svarer til verdier som ligger vest for krysslinjen. Det opprinnelige observasjonsgrunnlaget er GPS, rGPS, akustiske avstander og kompassretninger. Dette er derfor nøyaktige målinger. Det er 10s mellom hver målte posisjon. Totalt viser derfor figur 3.1 evolusjonen til noden i løpet av ca. 3 timer, og figur 3.2 ca. 7 timer og 22 min. Dette tilsvarer langt mindre enn én prosent av den totale undersøkelsen. Tidsrekke 1 representerer posisjoner til en node som ligger ca. midt på seimikk-kabelen. Mens noden i Tidsrekke 2 er plassert ganske langt ute på seismikk-kabelen. Dette er blant annet årsaken til at den er mindre glatt enn Tidsrekke 1. I Grindheim (2008) vises det at støyen til posisjonen er høyere for noder som ligger langt ute på kabelen. Dette virker naturlig, siden noden er mest bevegelig lengst ute på kabelen.

Vi ser bort fra den overordnede geometrien i datasettet. Krysslinjen er derfor ikke definert i samsvar med ønsket posisjon. Vi har definert krysslinjen som den rette linjen mellom første og siste punktet i tidsrekken. Dette er lettere å forholde seg til, og metodene vi bruker blir ikke påvirket av dette valget. Skulle vi modellert posisjonen basert på det reelle dynamiske systemet, måtte vi selvsagt definert krysslinjen annerledes.

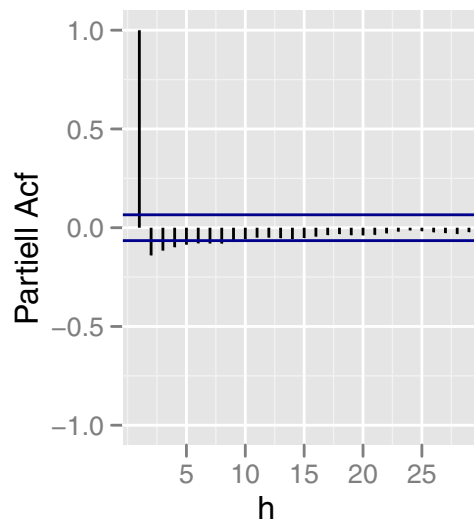
Tidsrekke 1



(a) 1091 posisjoner av node.



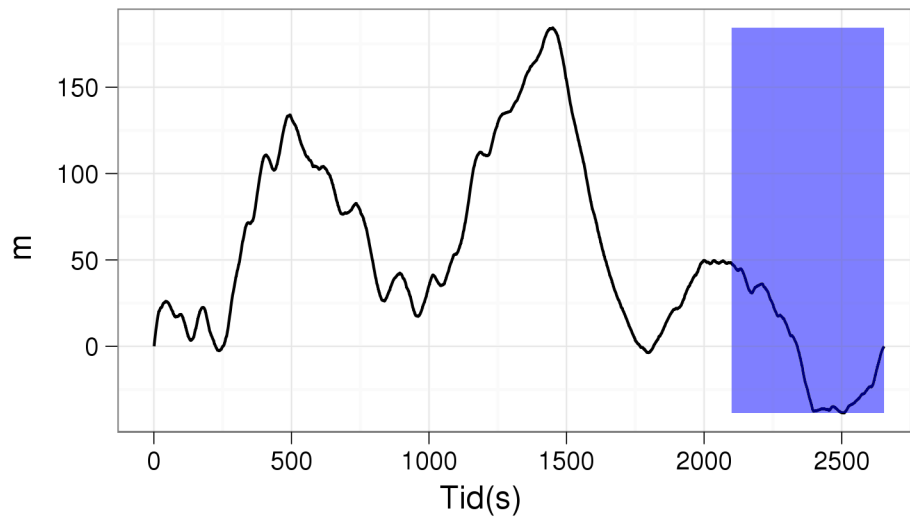
(b) Autokorrelasjonsfunksjon.



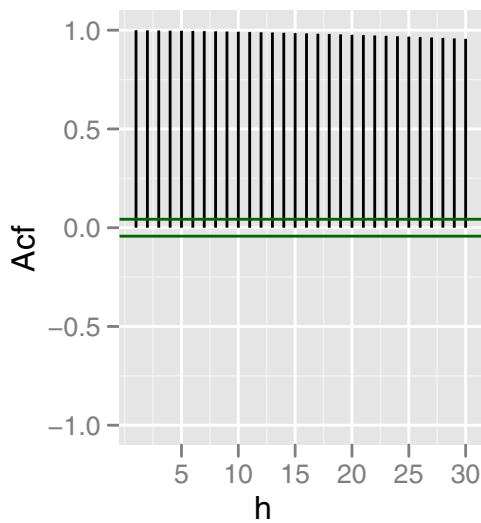
(c) Partiell autokorrelasjonsfunksjon.

Figur 3.1: 1091 posisjoner til noden i meter fra krysslinjen, $x = 0$. Krysslinjen er definert som en rett linje mellom første og siste punkt i tidsrekken. Markert område, 20% av tidsrekken, blir brukt senere til å evaluere tidsrekkemodeller. Disse dataene ble derfor ikke tatt med når vi regnet ut tilhørende ACF og PACF.

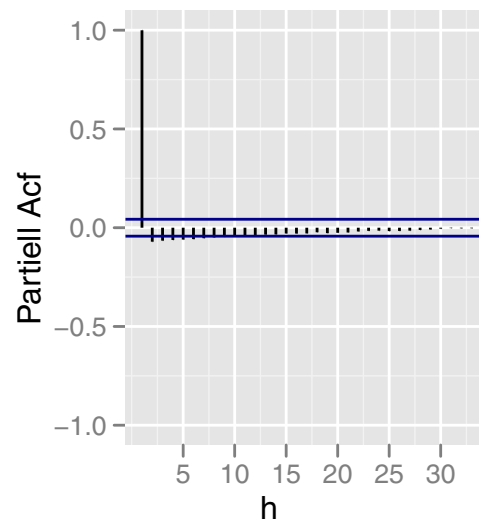
Tidsrekke 2



(a) 2654 posisjoner av node.



(b) Autokorrelasjonsfunksjon.



(c) Partiell autokorrelasjonsfunksjon.

Figur 3.2: 2654 posisjoner av en node i meter fra krysslinsen. Krysslinsen er definert på samme vis som i figur 3.1. Markert område, 20% av tidsrekken, blir senere brukt til evaluering av tidsrekkemodeller. Derfor har ikke disse dataene inngått i beregningen av ACF og PACF.

Hverken Tidsrekke 1 eller 2 er stasjonære. Men før vi transformerer de ikke-stasjonære tidsrekkene, estimerer vi trenden i de tidsrekkene lokalt med *Parabelmetoden*.

3.3.2 Trend

De deterministiske komponentene i en tidsrekke kalles oftest for *trend* og *sesongkomponent*. Sesongkomponenten innebærer regelmessige svingninger i tidsrekken, og ser ikke ut til å være tilstede i tidsrekkene, i figur 3.1 og 3.2. Begge tidsrekkene ser derimot ut til å inneholde en tidsvarierende trend. En tidsrekke med trend kan skrives på den strukturelle formen

$$X_t = m_t + W_t, \quad (3.5)$$

hvor m_t er en deterministisk trend og W_t er den stokastiske støyen. Vi gjenkjenner trend i tidsrekken med en lineær avtagende eller økende tidsrekke. Trenden kan også være kvadratisk, eller et polynom av enda høyere grad. Videre kan den være tidsvarierende som er tilfelle her. ACF'en til Tidsrekke 1 og 2, avtar sakte uten svingninger, som bekrefter at trend er tilstede. Når trenden varierer med tiden gir det ikke mening å estimere en deterministisk trend for hele tidsrekken. Her estimerer vi den lokalt. I de 40 første dataene i Tidsrekke 1 og 2, ser den kvadratiske trenden mer eller mindre konstant ut. Denne trenden estimerer vi med en andre-ordens kurvetilpasning gitt ved:

$$m_t = a_0 + a_1 t + a_2 t^2$$

der a_0 , a_1 og a_2 minimerer $\sum_{i=1}^n (x_i - m_t)^2$. I likhet med, Geograf AS refererer vi til metoden som Parabelmetoden. I eksempelet under finner vi a_0 , a_1 og a_2 , for de første 40 dataene i tidsrekken i figur 3.1 (a).

Eksempel 3.3.1: Kvadratisk trend

Vi partiell-deriverer $L = \sum_{i=1}^{40} (x_i - m_t)^2$ med hensyn på a_0 , a_1 og a_2 og finner minimum for alle parametrene slik

$$\begin{aligned} \frac{\partial L}{\partial a_0} &= \sum_{i=1}^{40} 2(a_0 + a_1 t_i + a_2 t_i^2 - x_i) \\ n a_0 + a_1 \sum_{i=1}^{40} t_i + a_2 \sum_{i=1}^{40} t_i^2 - \sum_{i=1}^{40} x_i &= 0 \\ \hat{a}_0 &= \frac{\sum_{i=1}^{40} x_i - a_1 \sum_{i=1}^{40} t_i - a_2 \sum_{i=1}^{40} t_i^2}{n} \end{aligned}$$

$$\begin{aligned}\frac{\partial L}{\partial a_1} &= \sum_{i=1}^{40} 2(a_0 + a_1 t_i + a_2 t_i^2 - x_i) t_i \\ a_0 \sum_{i=1}^{40} t_i + a_1 \sum_{i=1}^{40} t_i^2 + a_2 \sum_{i=1}^{40} t_i^3 - \sum_{i=1}^{40} x_i t_i &= 0 \\ \hat{a}_1 &= \frac{\sum_{i=1}^{40} x_i t_i - a_0 \sum_{i=1}^{40} t_i - a_2 \sum_{i=1}^{40} t_i^3}{\sum_{i=1}^{40} t_i^2}\end{aligned}$$

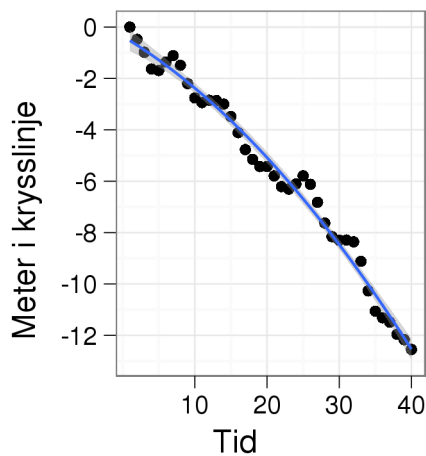
$$\begin{aligned}\frac{\partial L}{\partial a_2} &= \sum_{i=1}^{40} 2(a_0 + a_1 t_i + a_2 t_i^2 - x_i) t_i^2 \\ a_0 \sum_{i=1}^{40} t_i^2 + a_1 \sum_{i=1}^{40} t_i^3 + a_2 \sum_{i=1}^{40} t_i^4 - \sum_{i=1}^{40} x_i t_i^2 &= 0 \\ \hat{a}_2 &= \frac{\sum_{i=1}^{40} x_i t_i^2 - a_0 \sum_{i=1}^{40} t_i^2 - a_1 \sum_{i=1}^{40} t_i^3}{\sum_{i=1}^{40} t_i^4}\end{aligned}$$

der løsningen til $\min_{a_0, a_1, a_2} \sum_{i=1}^n (x_t - m_t)^2$ er $\hat{a}_0 = -0.3569$, $\hat{a}_1 = -0.1678$ og $\hat{a}_2 = -0.0034$.

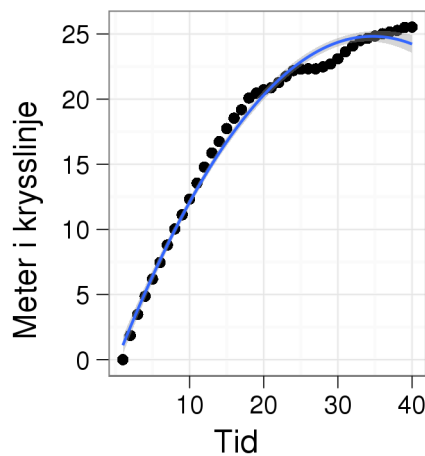
Løsningen til Parabelmetoden er plottet i figur 3.3 for begge tidsrekkene, der løsningen til (b) er $\hat{a}_0 = -0.3228$, $\hat{a}_1 = 1.4503$ og $\hat{a}_2 = -0.0209$. Legg merke til at når $\mathbb{E}(W_t) = 0$ og $\mathbb{E}(X_t) = m_t = a_0 + a_1 t + a_2 t^2$, så er a_0 lik posisjonen X_{t-1} . Videre er a_1 er proporsjonal med farten til X_{t-1} , i krysslinjeretning, og a_2 er proporsjonal med akselerasjonen til X_{t-1} . Det er vanlig å skrive $x_t = x_{t-1} + v_{t-1} \Delta t + \frac{1}{2} a_{t-1} \Delta t^2$.

Kurvene i figur 3.3 er estimert med alle 40 punktene. I figur 3.3 (a) ser den kvadratiske trend mer eller mindre konstant ut. Den tilpassede kurven er derfor omtrent like presis på hele plottet, og det kan tenkes at (3.5) er en god modell for disse 40 dataene, der støyen iid. Regresjonslinjen i 3.3 (b) derimot, er mest presis i de 15 første leddene, der trenden er omtrent konstant. Men i siste del endrer trenden form. Dataene varierer ikke jevnt rundt den estimerte trenden, og 3.20 er en dårlig modell her.

Tidsrekkemodellen i (3.5) blir gjenkjent som annen-grads regresjon, $Y_t = \mathbf{X}_t \boldsymbol{\beta} + W_t$, der $\boldsymbol{\beta}' = (a_0, a_1, a_2)$ og $\mathbf{X}_t = (1, t, t^2)$, der residualene, W_t , antas å være iid. Dersom W_t ikke er iid, som ser ut til å være tilfelle i 3.3 (b), kan vi gjøre et forsøk på å tilpasse en ARIMA-modell til W_t , gitt ved 3.2.5, og håpe at Z_t er iid.



(a) Tidsrekke 1



(b) Tidsrekke 2

Figur 3.3: En kvadratisk regresjonslinje er tilpasset de 40 første posisjonene.

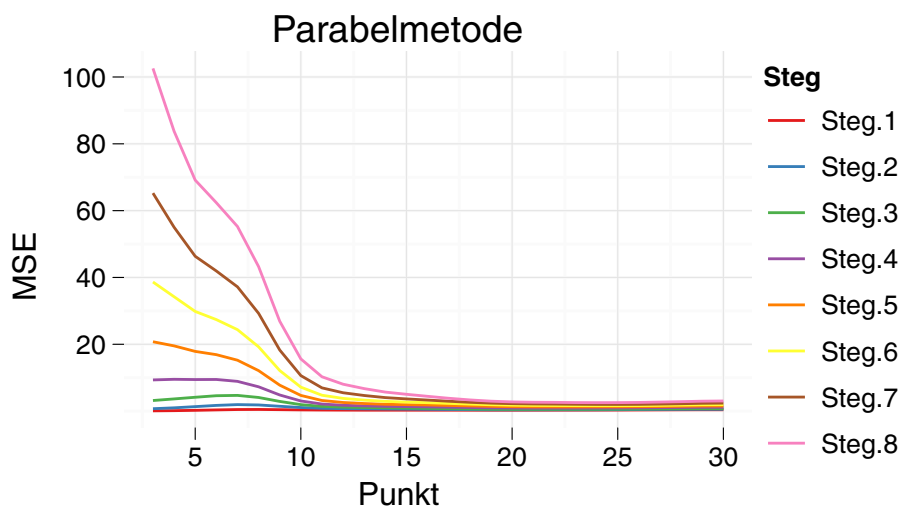
En tidsrekkemodell som estimerer den kvadratiske trenden lokalt, er nødvendig her. I avsnitt 4.4.2 blir modellen uttrykt på state-space form. Der antar vi at m_t ikke er deterministisk. I figur 3.3 (b) har de siste dataene er klart avvik fra den estimerte trenden, og her ville trolig en kvadratisk trend estimert med 3-5 punkter, fungert bedre. I neste avsnitt predikerer vi ved å tilpasse en lokal annengrads-kurve på tidsrekkene.

3.3.3 Parabelmetoden

Vi ble introdusert til Parabelmetoden gjennom en samtale med Peder. Det ble enighet om at vi skulle studere antall punkt i metoden som gir best prediksjon på Tidsrekke 1 og 2. Vi har testet hvor mange punkt som gjennomsnittlig gir best prediksjonsevne i Parabelmodellen. Vi har brukt 3,4,...,30 punkt i metoden til å predikere 1,2,...,8 steg frem i tid. Følgende antall punkt ble valgt fordi 3 er minimum, og den totale lengden på tidsrekken begrenser antall punkt. Dessuten viser det seg at 30 punkt, likevel ikke gir optimal prediksjon. Vi predikerer 8-steg fordi dette er tilstrekkelig når vi kun er interessert i å predikere akselerasjon 6-steg frem i tid.

Først testet vi metoden på hele Tidsrekke 1. Vi estimerte en annen-grads kurve med 3-30 punkt, og predikerte så 1 til 8 steg med alle kombinasjoner. Dette

gjorde vi på 30 ulike seksjoner i tidsrekken. Det vil si, de predikerte dataene er $30 \cdot i + j$ for $i = 1, \dots, 30$ og $j=1, \dots, 8$. I figur 3.4 er den gjennomsnittlige prediksjonsfeilen (MSE), $\sum_{i=1}^n (x_{t+i} - m_t)^2 / (n - 1)$ for $n = 30$ og $i = 1 \dots 8$, plottet for Tidsrekke 1. Tabell A.1, i første appendiks, viser også resultatene. I tabell A.1 ser vi at 1-steps prediksjon er mest presis når 3 punkt blir brukt



Figur 3.4: Gjennomsnittlig prediksjonsfeil (MSE) i Parabelmetoden med ulikt antall punkt i Tidsrekke 1.

i metoden. Mens 2-steps prediksjon gjør det best med 20 punkter. Videre gir 21 punkt best 3- og 4-steps prediksjon. For 5-steg er 23 punkt best. For 6- og 7-steg, 25 punkter. Og til slutt er 24 punkter er best for 8-steg. I figur 3.4 ser det ut som prediksjonsfeilen konvergerer. Men den øker faktisk igjen, for alle steg, etter 25 punkt. Resultatene antyder at det kreves færre punkt når vi predikerer få steg frem. Trolig fordi Parabelmetoden mister sin evne til å estimere lokalt med mange punkt. Derimot trengs flere punkter i Parabelmetoden når vi vil predikere langt frem i tid. Da betyr den globale formen på tidsrekken mer.

Senere skal vi predikere de markerte dataene i figur 3.1 og 3.2 med tidsrekke-modeller. Vi vil gjerne sammenligne den prediksjonen med Parabelmetoden. Derfor bruker vi også Parabelmetoden på de markerte dataene. Figur 3.5 viser tilsvarende resultater, som figur 3.4, for dataene i Tidsrekke 1 og 2. For Tidsrekke 1 har vi utført 1-8 prediksjoner med 3-30 punkter, som før. Men her har vi gjort det for kun 6 seksjoner. På Tidsrekke 2 har vi funnet gjennomsnittet av 18 seksjoner. Det vil si, de predikerte dataene i Tidsrekke 1 er $900 + 30 \cdot i + j$ for $i = 1, \dots, 6$ og $j=1, \dots, 8$. På lignende vis for Tidsrekke 2, er de predikerte

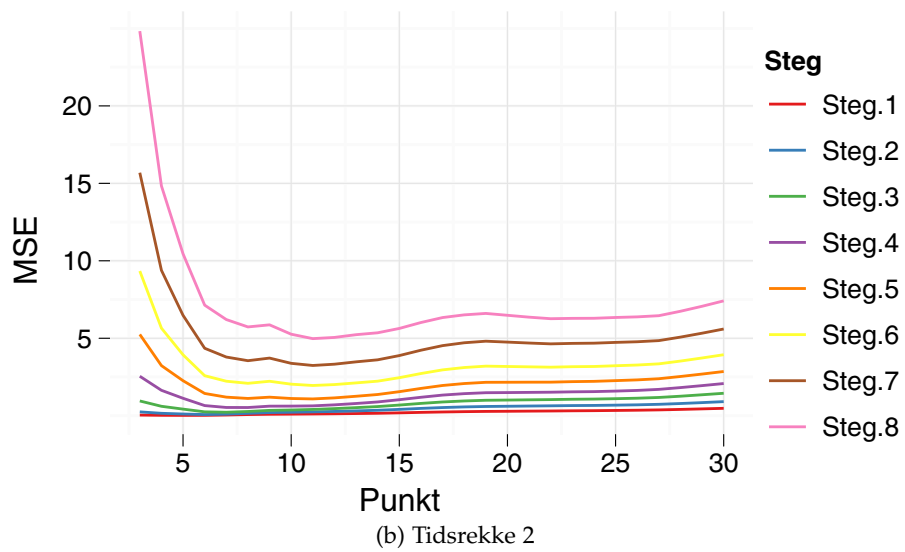
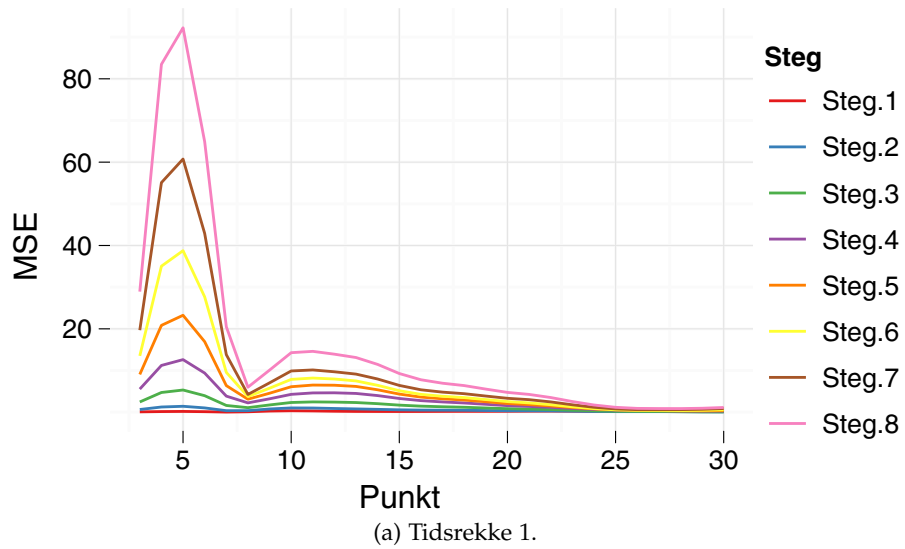
dataene $2100 + 30 \cdot i + j$ for $i = 1, \dots, 18$ og $j=1, \dots, 8$ Som sagt har tidsrekkene ulike egenskaper. Dette gjelder også de siste 20% av dataene. Tidsrekke 1 er veldig glatt her, og følger nesten uten unntak en konstant trend. Dessuten endrer posisjonene seg totalt 50 meter her, mens for Tidsrekke 2 er farten større. Tidsrekke 2 har også en trend som varierer, og større avvik fra denne.

Resultatene på Tidsrekke 1, i figur 3.5 (a), er litt ulike fra da vi predikerte over hele tidsrekken. Men nå er egenskapene i tidsrekken her, ganske ulike egenskapene i resten tidsrekken. Parabelmetoden har et klart maksimum i 5 punkter for 2-steps prediksjoner og oppover. Dette kan virke som en tilfeldighet, ved at en tilpasset kurve med 5 punkter ga vesentlig stor prediksjonfeil i noen få tilfeller. Metoden har et lokalt minimum i 8-9 punkt, avhengig av antall steg. For så å minke jevnt med antall punkt til et globalt minimum. I tabell B.2 ser vi at 7 punkt gir best 1-steps prediksjon. Videre gir 30 punkt best 2- og 3- prediksjon. 7-steg er best å predikere med 27 punkt. 28 punkt gir best 6-steps prediksjon. For 4- og 5-steg er 29 punkt best. Resultatene bekrefter at for de markerte dataene i Tidsrekke 1 er et høyt antall punkt, 27-29, å foretrekke i prediksjonen med unntak for 1-steps prediksjoner. En årsak til dette er fordi trenden, i slutten av tidsrekken, omtrent er konstant. Dette betyr at lokal estimering ikke er best egnet. Heller ikke for få steg.

Som forventet er resultatene på Tidsrekke 2 i figur 3.5 (b) annerledes, siden denne tidsrekken har en hyppigere varierende trend, og er mindre glatt. Dette ser vi tydeligst hvis vi sammenligner de to markerte områdene i tidsrekkene. Tidsrekke 2 har mange lokale svingninger og er ikke særlig glatt. I figur 3.5 (b) ser vi at færre punkt enn for Tidsrekke 1 gir minimum MSE. For 5- og 8-steps prediksjoner er 11 punkt optimalt. For 1-steg er 5 best. For 2-steg, 6 punkter. For 3- og 4-steg er 7 og 8 punkter best. Dette illustrere at når tidsrekken endrer seg lokalt, favoriserer Parabelmetoden færre punkter i prediksjonen.

Prediskjonevnen til Parabelmetoden er avhengig av egenskapene i tidsrekken. Vi har sett at antall punkt som gir optimal prediksjonevne avhenger av trenden i tidsrekken, hvor mye støy det er i tidsrekken, og hvor mange steg vi vil predikere. Å fjerne trenden i Tidsrekke 1 og Tidsrekke 2 ved å estimere den lokalt over hele tidsrekken er ikke en effektiv måte å oppnå en stasjonær tidsrekke på. Til det skal vi *differensiere* tidsrekken i avnittet under.

Parabelmetoden



Figur 3.5: Gjennomsnittlig prediksjonsfeil (MSE) i Parabelmetoden med ulikt antall punkt. Parabelmetoden er kun anvendt på de markerte dataene i begge tidsrekker, se figur 3.1 og 3.2.

3.3.4 ARIMA(p, d, q)-modellen

ARIMA(p, d, q)-modellen er en generalisering av ARMA(p, q)-modeller som også omfatter ikke-stasjonære prosesser. Målet er å differensiere dataene, et bestemt antall ganger, til prosessen er stasjonær. Da kan vi tilpasse en ARMA-modell til den differensierte tidsrekken. ARIMA er en forkortelse av, *Autoregressiv Integrated Moving Average*. Parameteren d er antall ganger tidsrekken blir differensiert. Antall AR-ledd, differensieringer og MA-ledd gis henholdsvis ved p, d og q . Modellen defineres slik:

Definisjon 3.3.2: ARIMA(p, d, q)

Hvis d er et positivt heltall, da er X_t en ARIMA(p, d, q)-modell hvis $Y_t := (1 - B)^d X_t$ er en kausal ARMA(p, q)-modell.

Definisjonen over sier at X_t oppfyller en differanseligning på formen

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t \quad Z_t \sim WN(0, \sigma^2),$$

der $\phi(z)$ og $\psi(z)$ er karakteristiske polynom av orden p og q , og $\phi(z) = (1 - \phi_1 z - \dots - \phi_p z^p) \neq 0 \forall |z| = 1, z \in \mathbb{C}$. Kausalitet ble definert i definisjon 3.2.5. Polynomiet $\phi^*(B) = \phi(B)(1 - B)^d$ har en rot av orden d i $z = 1$, og X_t er derfor bare stasjonær når $d = 0$. Da er ARIMA-modellen en ARMA-modell. For $d > 0$ er X_t en integrert versjon av den stasjonære tidsrekken.

3.3.5 Akaike's og Bayesian informasjonskriterium

Når vi kommer til Box Jenkins teori, skal vi anvende disse informasjonskriteriene når vi velger endelig ARIMA-modell for Tidsrekke 1 og 2. *Akaike's informasjonskriterie* (AIC) blir brukt til å finne hvilken orden p og q i ARIMA(p, \cdot, q)-modellen som tilpasser en gitt tidsrekke best med færrest mulig parametre. ARIMA-modeller med ulike parameterkombinasjoner blir rangert etter deres AIC-verdi. Den med minst AIC-verdi er vinneren. AIC-verdien belønner ikke bare tilpasningsdyktighet, ved å øke antall parametre øker man nemlig AIC-verdien. AIC gis ved:

$$AIC = 2k - 2\log(L) \tag{3.6}$$

der k er antall parametre i modellen, n antall data i tidsrekken og L er maksimum likelihood til den estimerte modellen. AIC kan være negativ. Det er

tilfelle når maksimum likelihood verdien er større enn én. Som i vårt tilfelle, der AIC-verdiene er negative for alle kombinasjoner av parametre i ARIMA-modellen, se figur 3.10 og 3.13.

Et annet informasjonskriteret kalt *Bayesian informasjonskriterium* (BIC) er gis ved:

$$BIC = \log(n)k - 2\log(L). \quad (3.7)$$

Vi ser at BIC straffer økt antall parametre i modellen i større grad enn AIC, for $n > 7$. Dette fører til at modellen ikke er like tilpasningsdyktig på eksisterende data. Derimot fører BIC ofte til bedre prediksjon.

3.4 Box-Jenkins teori

Box-Jenkins er en metodikk som går ut på å finne en passende ARIMA-modell på en systematisk måte. Metodikken er delt inn i tre hovedfaser:

- 1 Identifisere ARIMA-modellen
- 2 Estimere ARIMA-modellen
- 3 Evaluere ARIMA-modellen

3.4.1 Identifisere ARIMA-modellen

Differensierer tidsrekken

Det første og viktigste man gjør er å bestemme antall *differensieringer* som trengs for at tidsrekken blir stasjonær. Når $d = 0$ er modellen stasjonær. Ved å differensiere én gang, $\Delta X = X_t - X_{t-1}$, kan vi fjerne lineær trend i tidsrekken, siden én differensiering tilsvarer å derivere tidsrekken i alle punkt, hvis $\Delta t = 1$. Dersom krumning er tilstede i tidsrekken kreves en dobbeltdifferensiering, $\Delta^2 X = X_t - 2X_{t-1} + X_{t-2}$, for at disse egenskapene også skal bli konstante. Siden våre tidsrekker er posisjoner, vil én differensiering være endring i posisjon, som er proporsjonalt med fart. Når vi deler på tidsenhet, som er 10 s mellom hvert tidssteg, får vi fart. Videre, vil en dobbeltdifferensiering være proporsjonal med akselerasjon.

Vanligvis er korrekt antall differensieringer det laveste antallet som fører til at tidsrekken varierer noenlunde konstant rundt gjennomsnittet. Samtidig som ACF'en til residualene, mellom tidsrekken og $ARIMA(0,d,0)$, avtar raskt mot null. Hvis ACF'en har høye positive verdier for h større enn 10 eller mer, tyder det på at det trengs å differensiere flere ganger. Da reduserer vi autokorrelasjonen. Hvis ACF'en for $h=1$ er 0 eller negativ, trengs ikke flere differensieringer. Hvis ACF'en for $h=1$ er under -0.5 , kan den være overdifferensiert. Det er viktig å ikke differensiere for mange ganger. Det kan føre til overtilpasning, som igjen fører til behov for økt antall AR- og MA- parametre. Et plott av en overdifferensiert tidsrekke vil kanskje se veldig stasjonær ut ved første blick, men hvis man ser nærmere på tidsrekken vil man ofte se et mønster med hyppig skift fra positive til negative verdier.

Tidsrekke 1

Tidsrekke 1, i figur 3.1 (a), er ikke-stasjonær. Tydelig nok er det en trend tilstede. I Parabelmetoden så vi at trenden varierer, og mest mellom $t = 600$ til $t = 800$. Det er ikke noe regelmessig mønster, som tyder på at perioder er tilstede. ACF'en signaliserer også at det er trend tilstede, siden den beholder høye verdier for $h > 10$. PACF'en viser at denne sakte avtagende ACF'en hovedsakelig skyldes korrelasjonen mellom x_t og x_{t-1} . Naturlig nok er posisjonen ved tidspunkt t mest avhengig av posisjonen ved tidspunkt $t - 1$. Man kunne tenke seg at en AR(1)-modell er passende her, men dersom parameteren i en AR(1)-modell nesten er lik én, er den ikke stasjonær, og derfor ingen ARMA-modell. Stasjonærhet favoriseres. Neste fornuftige skritt for denne tidsrekken er derfor å tilpasse en $ARIMA(0,1,0)$. Som sagt, ved å differensiere en gang får vi data som er proporsjonale med farten i krysslinje retning. Vi ser, i figur 3.6(a), at farten er lik null der hvor stigningstallet i figur 3.1(a) er 0. Naturlig nok er verdiene i 3.6(a) betraktelig redusert. Farten i krysslinjen avviker mest fra 0 mellom $t = 600$ og $t = 800$ som skyldes den store endringen i posisjoner her. I figur 3.6 ser vi at farten varierer ganske jevnt rundt en trend. Men trenden varierer fortsatt. Siste del av tidsrekken avviker mest fra stasjonærhet. At det fremdeles er trend tilstede er også synlig i ACF'en til dataene. PACF'en viser at korrelasjonen hovedsakelig skyldes korrelasjon mellom variabler for $h = 1, \dots, 7$. Siden PACF'en avtar hurtigst har tidsrekken en *AR-signatur*. Det kommer jeg tilbake til senere i avsnittet. Selv om den er underdifferensiert, er det ikke sikkert den trenger å blir diffe-

rensiert en hel gang til. En partiell differensiering med AR-parametre holder kanskje. Ut fra PACF'en ser en AR(7)-modell passende ut. Den eliminerer korrelasjonen mellom residualene i ACF'en. Dette kunne vært en god modell. Men siden tidsrekken enda ikke ser stasjonær ut, som vi vil den skal være når vi predikerer, og siden vi er ute etter å predikere akselerasjon direkte, differensierer vi en gang til. I figur 3.7 er de dobbeltdifferensierte dataene plottet, som er proporsjonale med akselerasjon i krysslinje. Vi ser at de varierer omtrent konstant rundt gjennomsnittet, ARIMA(0,2,0). Det finnes ikke lenger synlig drift som varierer med tiden. Dataene vi skal predikere ser ut til å ha like egenskaper som resten av tidsrekken. ACF-en avtar raskt, som forventet når det ikke er trend i tidsrekken. Når den blir negativ for liten h , som her, er det tegn på at den ikke trengs å bli differensiert flere ganger. Tidsrekken har fremdeles en AR-signatur, men den er ikke like tydelig.

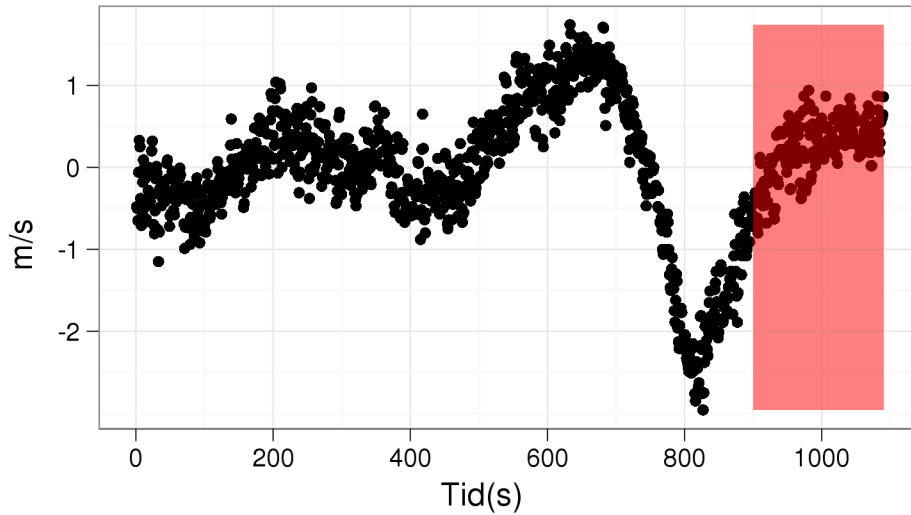
Tidsrekke 2

Vi ser at Tidsrekke 2 heller ikke er stasjonær. Denne tidsrekken er ikke fullt så glatt som Tidsrekke 1. Vi ser at selv om den ikke er stasjonær så ser egenskapene i dataene som skal predikeres, likere ut egenskapene i resten av tidsrekken, enn de gjorde i Tidsrekke 1. Svingninger, med ulik amplitude, gjentar seg. ACF-en er også høy for $h > 10$ her. Vi differensierer tidsrekken én gang. I figur 3.8 er den plottet med tilhørende ACF og PACF. Tidsrekken ser fremdeles ikke stasjonær ut. Svingninger og varierende trend er tilstede fortsatt. Farten i Tidsrekke 2 skifter oftere retning. Verdiene i ACF'en avtar raskere her, enn den gjorde for posisjonsverdiene, men den har fremdeles store verdier for $h > 10$. Videre har tidsrekken en klar AR-signatur. Av samme grunner som for Tidsrekke 1, differensierer vi en gang til. Den dobbeltdifferensierte tidsrekken i figur 3.9 ser stasjonær ut. Den varierer konstant rundt gjennomsnittet i tidsrekken. ACF'en bekrefter at den er stasjonær, når verdien for $h = 1$ er negativ. Den skifter tegn hyppig, men siden ACF'en ikke har for høy negativ verdi for $h = 1$, tyder det på at den ikke er overdifferensiert. Dessuten har den ikke en klar MA-signatur.

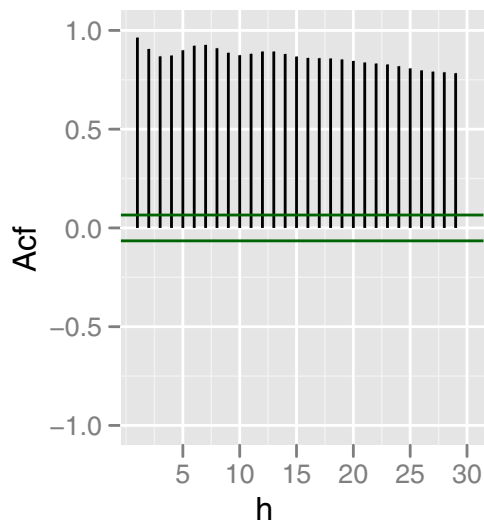
Identifiserer orden p,q

ARIMA-modellen tillater oss å øke antall parametre for å eliminere korrelasjonen i ACF'en på en lettvinnt måte. Når d er bestemt, skal vi bestemme hvilken

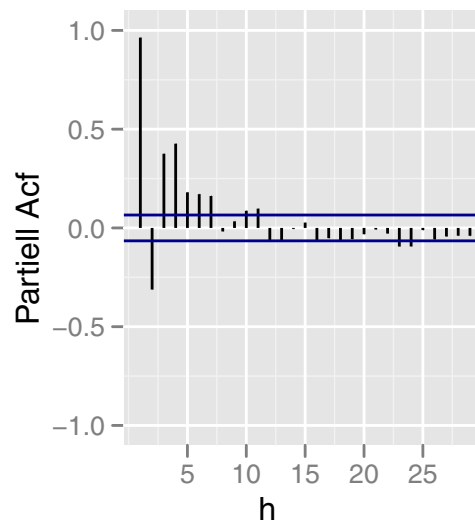
Tidsrekke 1: Differensierte data



(a) Tidsrekke 1 er differensiert én gang, og er proporsjonal med fart i krysslinjeretning. Verdiene er ikke dividert på 10s.



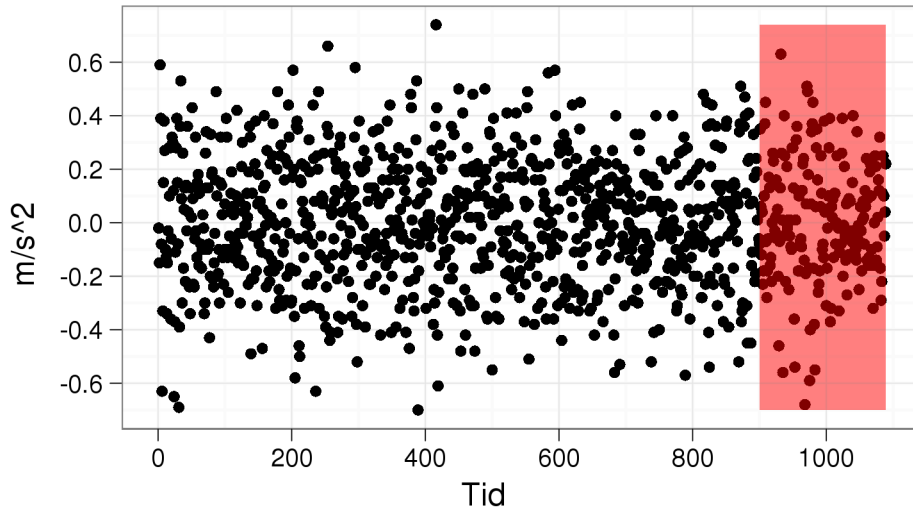
(b) Tilhørende autokorrelasjonsfunksjon, der h er antall tidssteg.



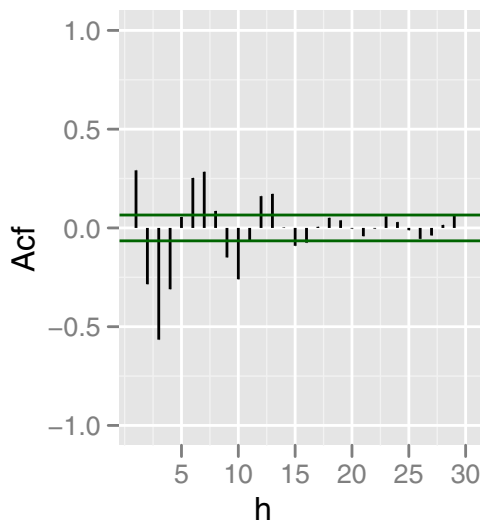
(c) Tilhørende partiell autokorrelasjonsfunksjon, der h er antall tidssteg.

Figur 3.6: Tidsrekke 1 er differensiert én gang, $\Delta X = X_t - X_{t-1}$. Tilhørende ACF og PACF er beregnet med, de første, 80% av dataene.

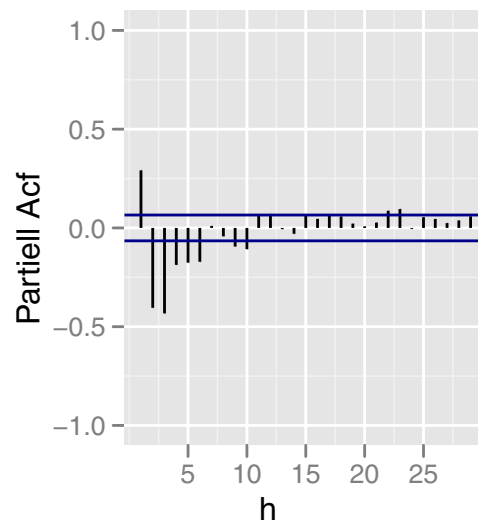
Tidsrekke 1: Dobbeltdifferensierte data



(a) Tidsrekke 1 er dobbeltdifferensert, og proporsjonal med akselerasjon i krysslinjeretning. Verdiene er ikke dividert på 100s.



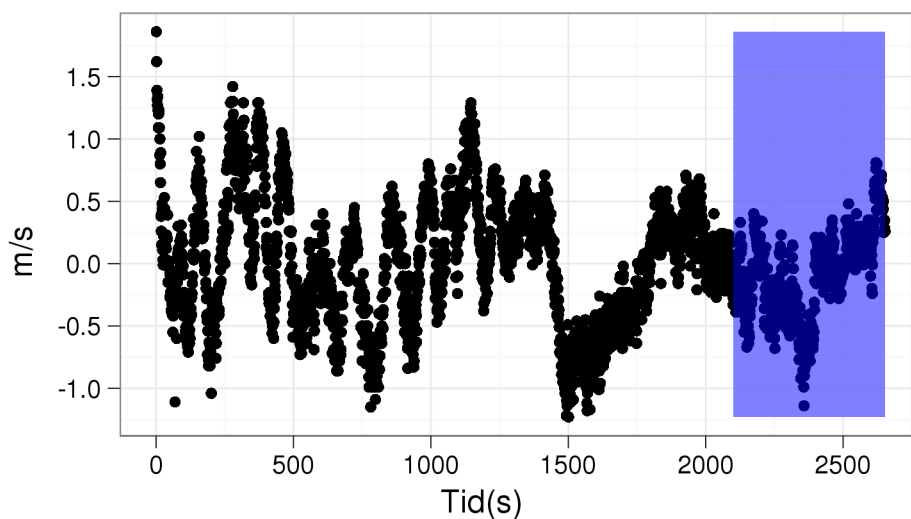
(b) Tilhørende autokorrelasjonsfunksjon, der h er antall tidssteg.



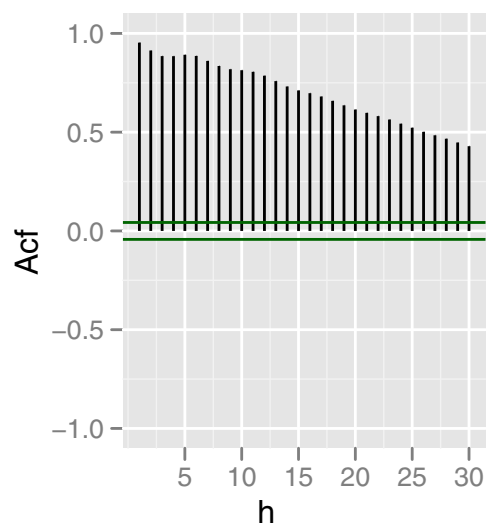
(c) Tilhørende partiell autokorrelasjonsfunksjon, der h er antall tidssteg.

Figur 3.7: Tidsrekke 1 er dobbeltdifferensert, $\Delta^2 X = X_t - 2X_{t-1} + X_{t-2}$. Tilhørende ACF og PACF er beregnet med de første 80% av dataene.

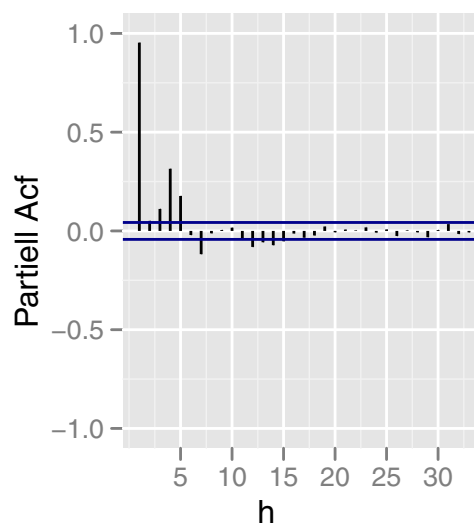
Tidsrekke 2: Differensierte data



(a) Tidsrekke 2 er differensiert én gang, og proporsjonal med fart i krysslinjeretning. Verdiene er ikke dividert på 10s.



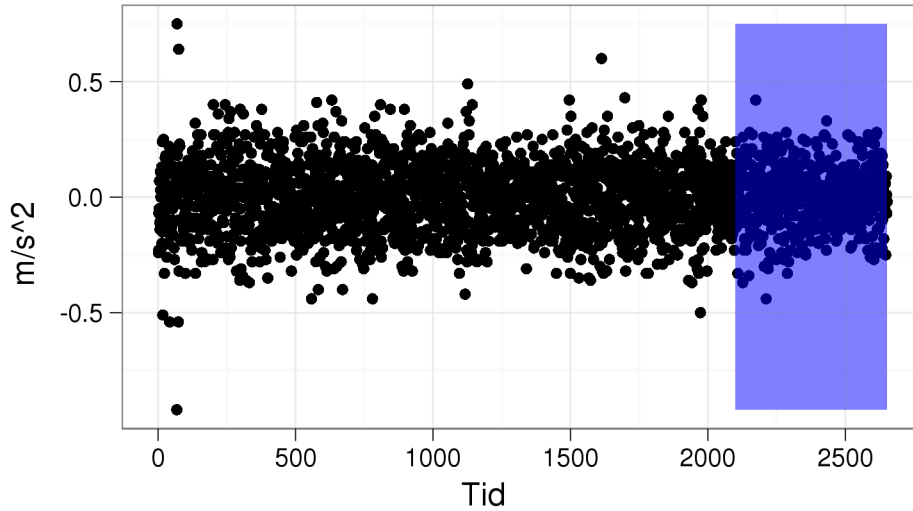
(b) Tilhørende autokorrelasjonsfunksjon, der h er antall tidssteg.



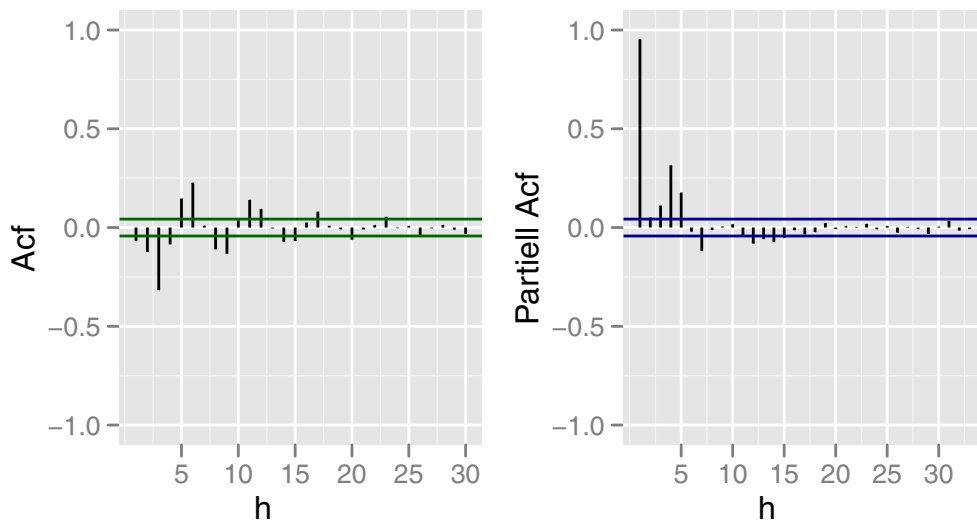
(c) Tilhørende partiell autokorrelasjonsfunksjon, der h er antall tidssteg.

Figur 3.8: Tidsrekke 2 er differensiert én gang, $\Delta X = X_t - X_{t-1}$. Tilhørende ACF og PACF er beregnet med de første 80% av dataene.

Tidsrekke 2: Dobbeltdifferensierte data



(a) Tidsrekke 2 er dobbeltdifferensiert, og proporsjonal med akselerasjon i krysslinjeretning. Verdiene er ikke dividert på 100s.



(b) Tilhørende ACF, der h er antall tidssteg. (c) Tilhørende PACF, der h er antall tidssteg.

Figur 3.9: Tidsrekke 2 er dobbeltdifferensiert, $\Delta^2 X = X_t - 2X_{t-1} + X_{t-2}$. Tilhørende ACF og PACF er beregnet med, de første, 80% av dataene.

orden MA- og AR-ledd som er mest tilpasningsdyktig. Det kan gjøres på en systematisk måte, ved å identifisere riktig antall AR- og MA-parametre ved hjelp av ACF'en og PACF'en. Dersom PACF'en har en knekk for ganske liten h , mens ACF'en avtar sakte, sier vi at tidsrekken har en *AR-signatur*. Det vil si at autokorrelasjonen mellom residualene kan beskrives mest effektivt med AR-parametre. AR-signatur forekommer ofte i tidsrekker hvor ACF'en er positiv for $h = 1$, og i svakt underdifferensierte tidsrekker, siden AR-ledd virker som en partiell differensiering. En AR(1) er f. eks lik en ARIMA(0,1,0), mens når koeffisienten er mellom 0 og 1, differensierer den partielt. Andre ganger er det mer effektivt å øke antall MA-ledd for å beskrive autokorrelasjonen mellom residualene. ACF'en spiller samme rolle for MA-leddene, som PACF'en gjør for AR-leddene. ACF'en antyder hvor mange MA-ledd som trengs for å fjerne uønsket autokorrelasjon. I tilfeller der ACF'en avtar raskt, mens PACF'en synker langsomt, sier vi at tidsrekken har en *MA-signatur*. MA-signatur assosieres ofte med en negativ ACF for $h = 1$, og er ofte tilfelle når tidsrekken er svakt overdifferensiert. MA-ledd eliminerer nemlig partiell differensiering. Vi skal basere det endelige valget av orden på AIC- og BIC-kriteriet, se (3.6) og (3.7).

Tidsrekke 1

ACF'en og PACF'en til den dobbeltdifferensierte tidsrekken viser hverken en klar AR- eller MA-signatur, men allikevel blir tidsrekken mest effektivt forklart av AR-parametre, siden PACF'en avtar raskest. ACF'en er tilnærmet lik null etter $h = 13$, men har kun signifikante verdier for h opp til 10. Dette tyder på at det kreves ganske mange MA-parametre i en ren MA-modell. Hvis vi velger en blandet modell, vil kravet på antall MA-parametre minke med antall AR-parametre. PACF'en tyder på at 6 AR-parametre er tilstrekkelig i en ren AR-modell.

Tidsrekke 2

I figur 3.9 ser vi at ACF'en har lave verdier, men avtar likevel ikke raskt mot null. PACF'en starter derimot ut med høye verdier, og avtar raskt mot null. Ingen klar AR-signatur her heller, men en snever en. I en blandet modell vil trolig p være høyere enn q . I en ren AR(p) ser $p=7$ ut til å være tilstrekkelig.

3.4.2 Estimerer ARIMA-modellen

Ofte er en ren AR- eller MA- modell det beste valget, siden parametrene har en tendens til å nøytralisere hverandres effekt i en blandet modell. For rene AR-modeller kan parametrene estimeres med *Yule-Walker* eller *Durbin-Levinson*. For MA-modeller med *Innovasjon-Algorithmen* eller *Hannan Rissanen*. For ARMA-modeller er det generelt litt mer komplisert å estimere parametrene. Som sagt bruker vi ML-estimering (1.1), til å finne $\hat{\phi}$, $\hat{\theta}$ og $\hat{\sigma}^2$. Vi definerer multivariabel gaussisk likelihood.

Definisjon 3.4.1: Multivariabel gaussisk likelihood

Anta at tidsrekken $\mathbf{X}_{1:T}$ er multivariabel normalfordelt med forventning $\boldsymbol{\mu} = 0$, og ikke-singulær kovariansmatrise Γ_n (3.2), da er Likelihoodfunksjonen gitt ved:

$$L(\Gamma_n, \mathbf{X}_{1:T}) = 2\pi^{-\frac{n}{2}} |\Gamma_n|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{X}'_{1:T} \Gamma_n^{-1} \mathbf{X}_{1:T}\right). \quad (3.8)$$

Vi finner $\hat{\Gamma}_n(\phi, \theta, \sigma^2)$, ved å maksimere (3.8), gitt $\mathbf{X}_{1:T}$. Men vi vil helst unngå å regne ut inversen til kovariansmatrisen for stor n . Vi uttrykker derfor Likelihooden annerledes, ved prediksjonfeilen $X_j - \hat{X}_j$ og tilhørende varians σ_j^2 for $j = 1, \dots, n$. Da kan (3.8) beregnes rekursivt med Innovasjons-Algorithmen. Således er gaussisk Likelihood for en ARMA-prosess, $L(\phi, \theta, \sigma^2)$, uttrykt annerledes. Men vi velger å ikke gå mer i detalj her, og lar R utføre estimeringen, når antakelsene over er oppfylt.

Til slutt velges orden p, q , basert på modellene som har lavest AIC- og BIC-verdi. Vi vurderer Tidsrekke 1 først.

Tidsrekke 1

Vi finner hvilken kombinasjon $p = 0, \dots, 10$ og $q = 1, \dots, 10$, som gir lavest AIC- og BIC-verdi. Vi kaller modellene med minimal AIC- og BIC-verdi for AIC- og BIC-modellen. I figur 3.10 gir ARIMA(6,2,7) lavest AIC-verdi = 583.1. Figuren viser at AIC-verdiene er relativt lave for mange kombinasjoner. En sum på 4 parametre, eller flere, vil gi en relativt god modell. Det er ikke alltid det svarer seg å velge modellen med lavest AIC-verdi. God tilpasningsdyktighet kan være en overtilpasning og fungere dårlig i prediksjon. Derfor viser det seg at BIC-modellen ofte gjør det best i prediksjon. En modell med færre

parametre kan derfor predikerte minst like bra som AIC- og BIC-modellen. Prediksjonsevnen til ARIMA(3,2,1)-modellen, som er en modell i knekkpunktet i figur 3.10 (a), var derimot dårlig sammenlignet med AIC- og BIC-modellen. Modellen med lavest BIC-verdi = 524.4 er ARIMA(5,2,7) (se figur 3.10 (b)). Studerer vi parameterverdiene til AIC-modellen i tabell 3.1 nærmere, ser vi at det karakteristiske polynomet til AR-koeffisientene har enhetsrøtter innenfor enhets sirkelen i det komplekset planet, (se tabell 3.5). Dette betyr at ARMA-modellen til de dobbeltdifferensierte dataene ikke er kausal, og X_t er derfor ikke en ARIMA(6,2,7)-modell. Dette betyr at vi ikke kan predikere posisjoner med AIC-modellen. Allikevel er AIC-modellen stasjonær, og derfor en ARMA-modell for akselerasjonsdataene. Det gir lite mening å predikere med en ikke-kausal ARMA-modell, men til tross for dette har vi testet prediksjonsevnen til AIC-modellen versus BIC-modellen. I figur 3.10 (b) ser vi

AR	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	
Par	1.1092	-0.7687	-0.1898	0.9554	-0.7466	0.5120	
s.e	0.0761	0.1042	0.1012	0.1035	0.0929	0.0583	
MA	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
Par	0-1.063	0.3376	0.0537	-0.8102	0.6046	-0.3655	0.3344
s.e	0.075	0.1029	0.1098	0.0930	0.0854	0.0625	0.0410

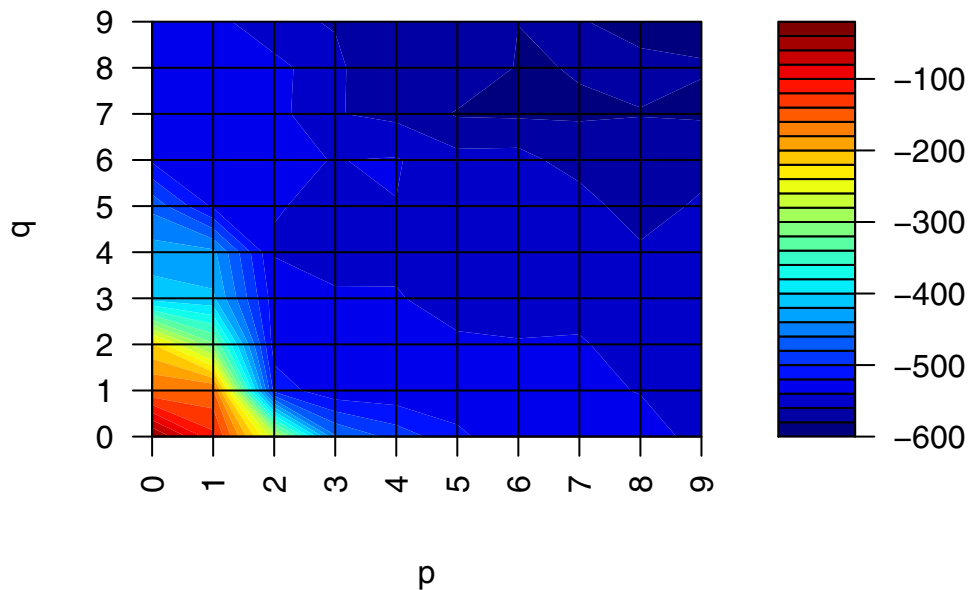
Tabell 3.1: Parameterverdier til ARIMA(6,2,7) på tidsrekke 1. Videre er $\sigma^2 = 0.0295$, log likelihood = 306.55 og AIC = -583.1

at BIC-verdiene er relativt lave for flere kombinasjoner enn for AIC-verdiene. Vi har valgt modellen, ARIMA(5,2,7), med minimum BIC. Tabell 3.2 viser de tilhørende parameter-verdiene. Som forventet er likelihood verdien mindre, og σ^2 større for denne modellen. AR-parametrene i BIC-modellen har hverken

AR	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5		
Par	1.9184	-2.5137	2.2349	-1.2939	0.5671		
s.e	0.1143	0.1971	0.2007	0.1324	0.0597		
MA	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
Par	-1.8726	2.0541	-1.9954	1.1548	-0.5480	0.1580	0.109
s.e	0.1176	0.2011	0.1869	0.1624	0.1181	0.1132	0.069

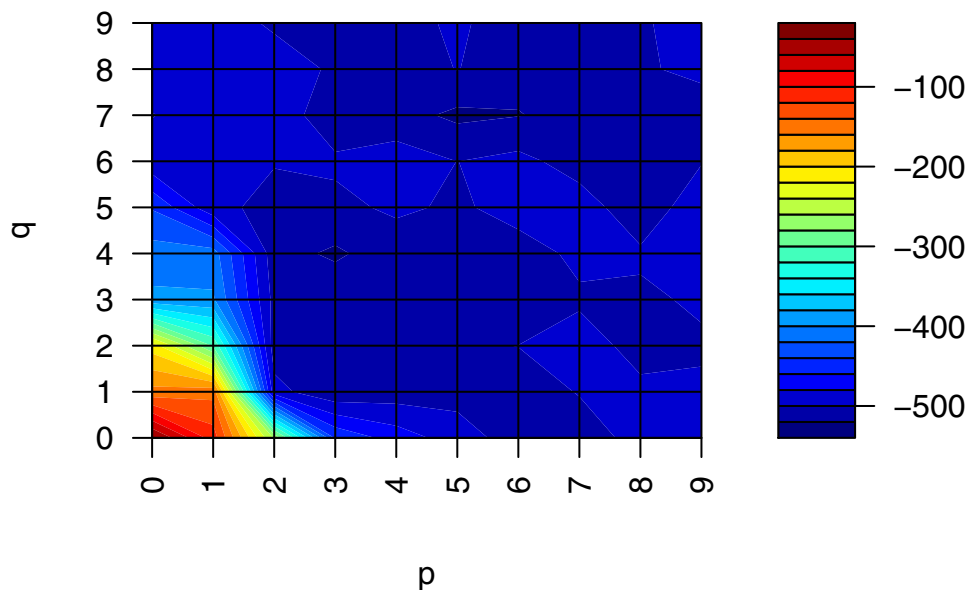
Tabell 3.2: Parameterverdier til ARIMA(5,2,7) på tidsrekke 1. Videre er $\sigma^2 = 0.0296$, log likelihood = 304.99 og AIC = -581.99.

Aic-verdier for ulike Arima(p,2,q)



(a) AIC-verdier for de ulike ARIMA($p,2,q$) modellene. Minimum AIC-verdi er -583.1 i $p, q = 6,7$.

Bic-verdier for ulike Arima(p,2,q)



(b) BIC-verdier for de ulike ARIMA($p,2,q$) modellene. Minimum BIC-verdi er -524.4 i $p, q = 5,7$.

Figur 3.10: AIC- og BIC-verdier for ARIMA($p,2,q$)-modeller tilpasset Tidsrekke 1.

røtter på eller innenfor enhetssirkelen i det komplekse plan, (se tabell 3.5). Den er derfor både stasjonær og kausal. Dette fører til at BIC-modellen tilfredstiller definisjon 3.3.2, og X_t en ARIMA(5,2,7)-modell. Dette fører til at vi kan predikere posisjoner, såvel som akselerasjon med BIC-modellen. Denne modellen blir derfor satt i fokus for Tidsrekke 1. Figur 3.11 viser residualene til X_t og ARIMA(5,2,7) for de første 80% av tidsrekken. Gjennomsnittet er omtrent 0, og variansen ser tilnærmet konstant ut. Dessuten er autokorrelasjonen eliminert. Residualene ser tilnærmet normalfordelte ut (se figur 3.12), til tross for at Kolmogorov Smirnov test forkaster hypotesen om at de er gaussisk fordelt med $D = 0.6542$, p -verdi $< 2.2 \cdot 10^{-16}$ på $\alpha = 0.05$ signifikansnivå. Der D er den maksimale avstanden mellom den kumulative normalfordeling med forventning og standard avvik lik residualene, og residualene i stigende orden. Siden n er stor, skal ikke avviket fra normalfordelingen være stort før testen forkaster hypotesen, og siden residualene ser normalfordelte ut i figur 3.12, godtar vi dette. Så langt tilfredstiller BIC-modellen forutsetningene for en god modell, men vi evaluerer modellen når vi predikere med den senere.

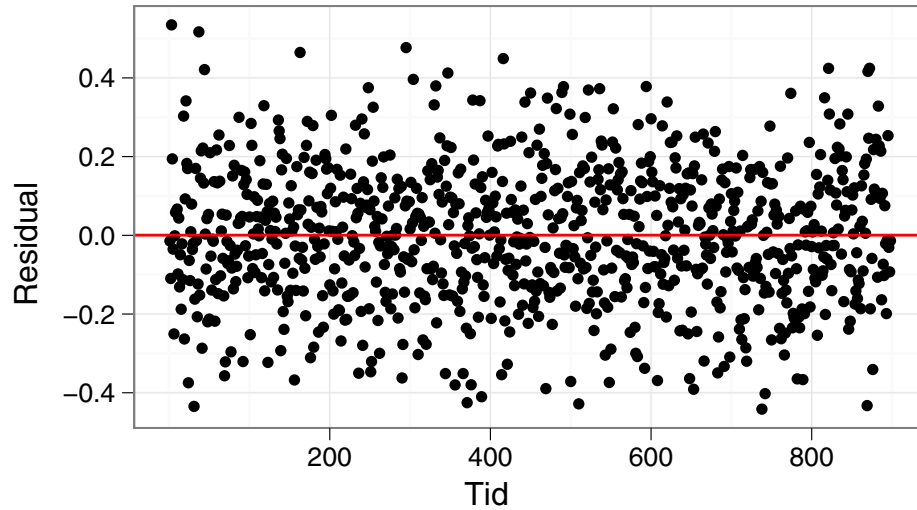
Tidsrekke 2

AIC- og BIC-verdiene, for de dobbeltdifferensierte dataene i Tidsrekke 2, finner vi i figur 3.13. ARIMA(9,2,7) gir minimum AIC-verdi = -2466.91 , og ARIMA(5,2,5) gir minimum BIC-verdi = -2397.03 . Parameterverdiene til ARIMA(9,2,7) gis i tabell 3.3. I likhet med AIC-modellen for Tidsrekke 1, er AIC-modellen for

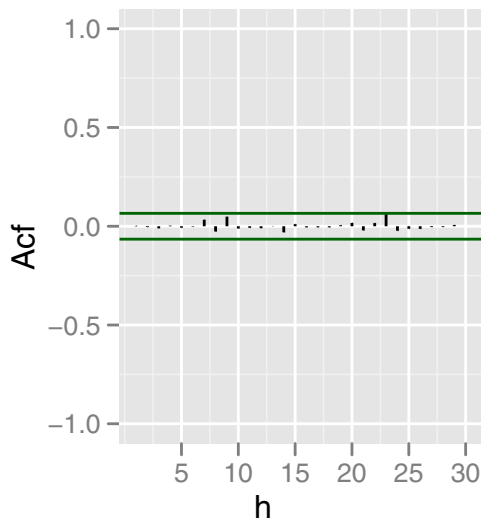
AR	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7
Par	1.6844	-0.8102	0.3758	-0.7894	1.3965	-1.0519	0.0386
s.e	00.1220	0.1889	0.0809	0.0645	0.1091	0.1478	0.0712
AR	ϕ_8	ϕ_9					
Par	0.1976	-0.0502					
s.e	0.0770	0.0485					
MA	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
Par	-1.9017	1.0366	-0.6013	1.1957	-1.5297	1.2572	-0.4540
s.e	0.1200	0.2081	0.0803	0.0448	0.1717	0.1643	0.0516

Tabell 3.3: Parameterverdier til ARIMA(9,2,7) på Tidsrekke 2. Videre er $\sigma^2 = 0.01774$, log likelihood = 1251.46, AIC = -2466.91 .

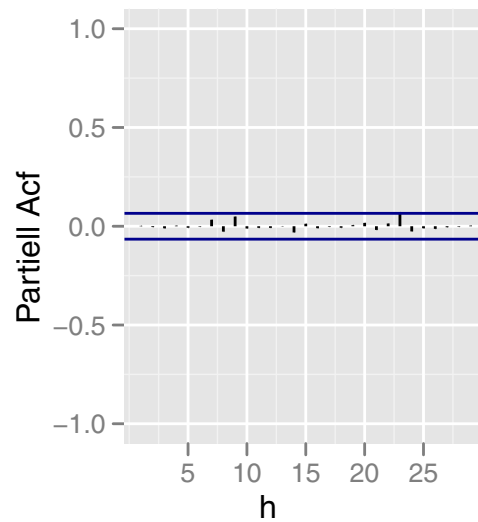
Tidsrekke 2, stasjonær, men ikke kausal. Dette fører oss til samme dilemma. Vi



(a) Residualene mellom den dobbeldifferensierte Tidsrekke 1 og ARIMA(5,2,7).

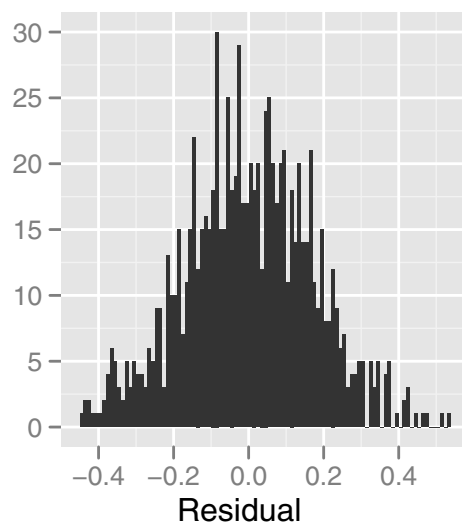


(b) Tilhørende ACF, der h er tidssteg.



(c) Tilhørende PACF, der h er tidssteg.

Figur 3.11: Residualene til BIC-modellen med tilhørende ACF og PACF. Merk at det er kun for de første 80% av tidrekken.



Figur 3.12: Histogram over residualene til BIC-modellen tilpasset Tidsrekke 1.

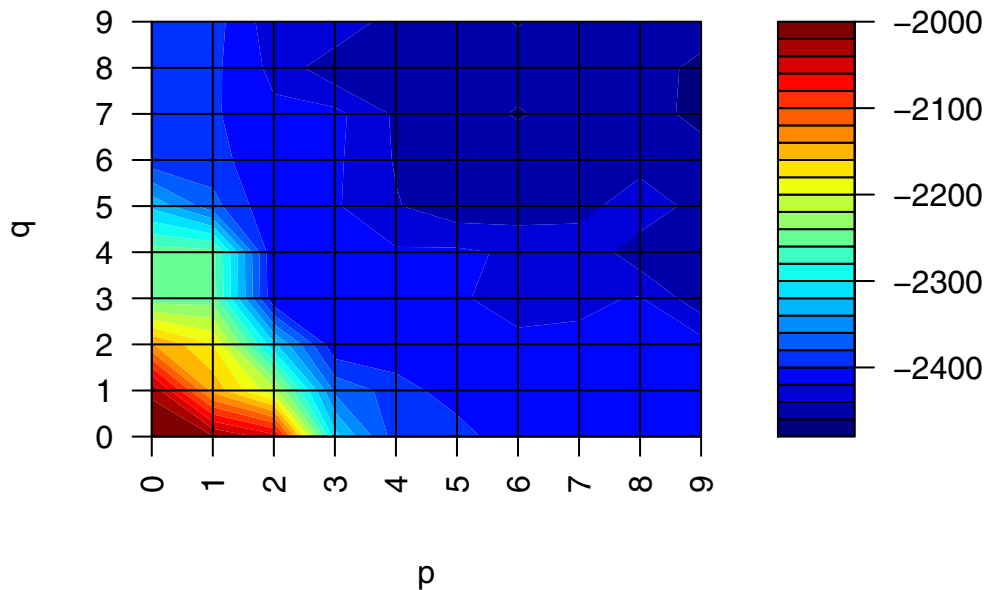
kan ikke bruke AIC-modellen til å predikere posisjoner. Samtidig gir det ikke mening å predikere akselerasjon med den. Parameterveridene til BIC-modellen på Tidsrekke 2 gis i tabell 3.4. I tabell 3.5 ser vi at BIC-modellen er stasjonær

AR	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
Par	2.0278	-1.8619	0.8226	0.1419	-0.2018
s.e	0.2340	0.4569	0.4854	0.3101	0.0806
MA	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
Par	-2.2368	2.1545	-1.2209	0.3058	0.0562
s.e	0.2354	0.5075	0.5416	0.3775	0.1292

Tabell 3.4: Parameterverdier til ARIMA(5,2,5) på Tidsrekke 2. Videre er $\sigma^2 = 0.01797$, log likelihood = 1238.76, AIC = -2453.51.

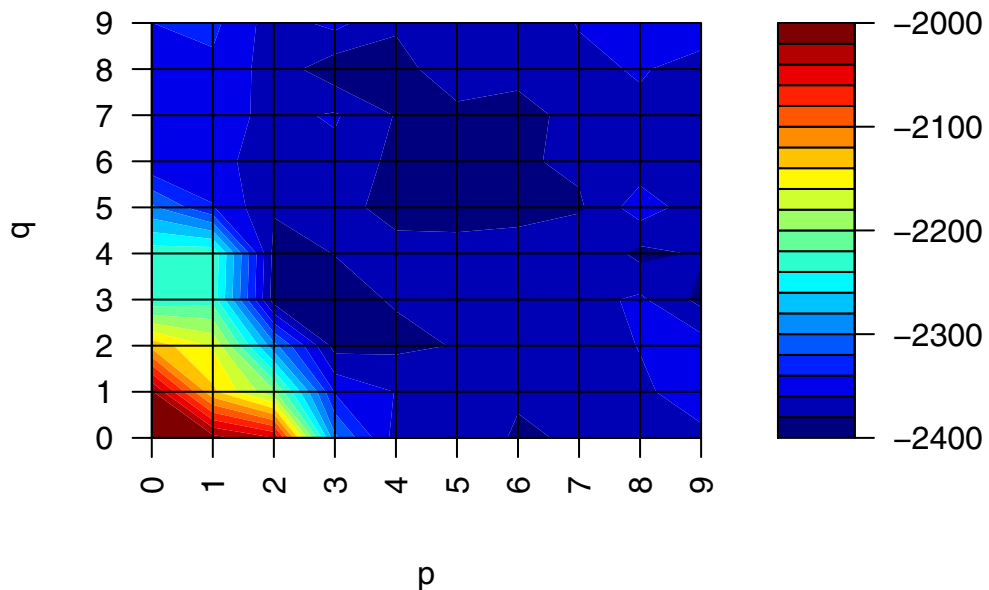
og kausal. I figur 3.14 er residualene plottet, med tilhørende ACF og PACF. Residualene har gjennomsnitt tilnærmet lik 0, og en varians som ser rimelig konstant ut, med unntak fra noen få uteliggere. Autokorrelasjonen er eliminert av modellen. Residualene ser normalfordelte ut i histogrammet (se figur 3.15), selv om Kolmogorov Smirnov testen forkaster hypotesen med $D = 0.6607$, p -verdi $< 2.2 \cdot 10^{-16}$ på et 0.05 signifikansnivå.

Aic-verdier for ulike Arima(p,2,q)



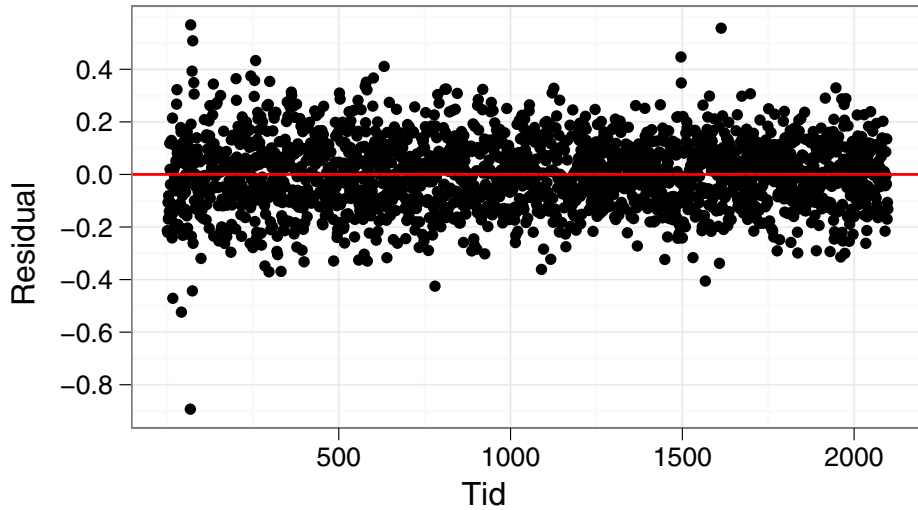
(a) AIC-verdier for de ulike ARIMA($p,2,q$) modellene. Minimum AIC-verdi er -2466.91 i $p, q = 9,7$.

Bic-verdier for ulike Arima(p,2,q)

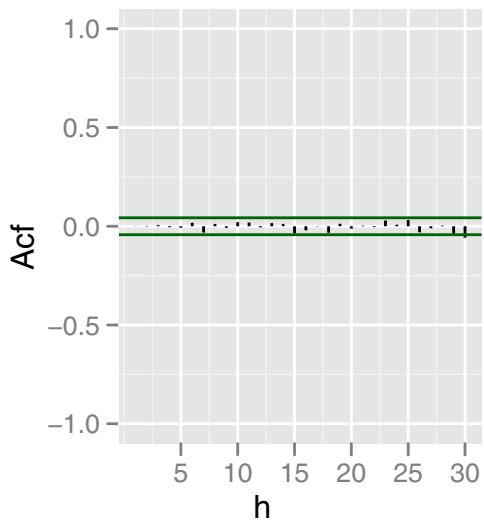


(b) BIC-verdier for de ulike ARIMA($p,2,q$) modellene. Minimum BIC-verdi er -2397.03 i $p, q = 5,5$.

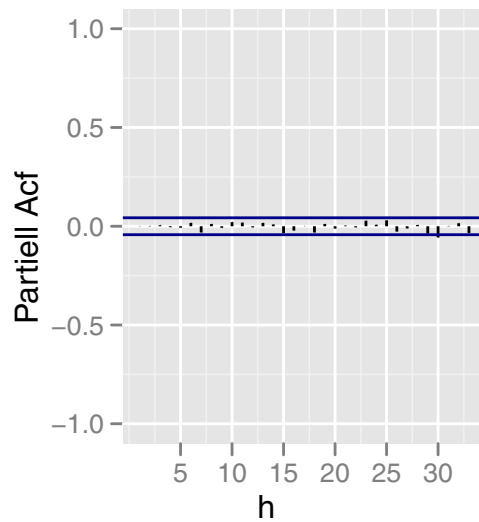
Figur 3.13: AIC-og BIC-verdier til ARIMA($p,2,q$)-modeller tilpasset på de dobbeltidifferensierte dataene i Tidsrekke 2.



(a) Differensiert tidsrekke $\Delta X = X_t - X_{t-1}$.

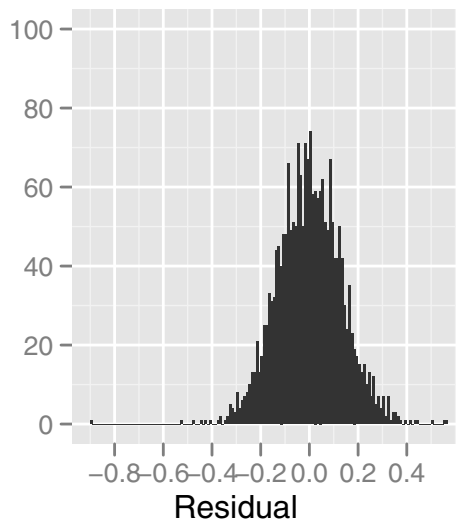


(b) Autokorrelasjonsfunksjonen til ΔX .



(c) Tilhørende PACF, der h er tidssteg.

Figur 3.14: Residualene til BIC-modellen tilpasset Tidsrekke 2, med tilhørende ACF og PACF. Merk det er kun for de første 80% av dataene.



Figur 3.15: Histogram viser residualene i BIC-modellen tilpasset Tidsrekke 2.

3.4.3 Unit root

En enhetsrot (unit root) i de estimerte AR- eller MA-parametrene kan antyde om modellen er over- eller underdifferensiert, og om en full differensiering må bli kansellert eller tilføyd. Det er en unit root i AR-delen hvis $\sum_{i=1}^p \phi_i - 1 = 0$, altså når summen av AR-parametrene er tilnærmet lik 1. Tilsvarende, når summen av MA-parametrene er nær én, så har MA-delen en enhetsrot. Generelt, når AR-delen har en enhetsrot, reduseres antall AR-parametre, mens antall differensieringer økes. Når MA-delen har en unit root, begrenses antall MA-parametre, og antall differensieringer reduseres. Dersom prediksjoner oppfører seg ustabil eller uberegnelig, kan det være en unit root, i MA- eller AR-Parametrene. For å bestemme om AR- eller MA-parametrene har en enhetsrot, må vi ta hensyn til standardavvikene til de estimerte parametrene. Dickey Fuller test, er en hypotesetest, som tester om det er signifikante enhetsrøtter tilstede i modellen. Som sagt, nøyer vi oss med å differensiere tidsrekken to ganger, så vi skal ikke utføre denne testen. Men i AIC- og BIC-modellene for begge tidsrekkene er summen av AR-parametrene ganske nær 1, som tyder på at begge tidsrekkene er underdifferensiert.

I definisjon 3.2.5 så vi at røttene i AR-polynomet indikerer om ARMA-modellen er stasjonær, eventuelt kausal. Tabell 3.5 viser at alle modellene er stasjonære, mens bare BIC-modellene er kausale. Her har usikkerheten i

parametrene ikke blitt tatt hensyn til, men røttene til $\phi(z)$ i BIC-modellene ligger langt utenfor enhetssirkelen.

Modell\Røtter	1	2	3	4	5	6	7	8
AIC 1	1.25	0.86	1.25	1.87	1.87			
BIC 1	1.73	1.95	1.95	1.73				
AIC 2	1.53	0.92	0.92	1.53	2.61	5.98	6.07	6.07
BIC 2	2.07	2.07	7.09	3.32				

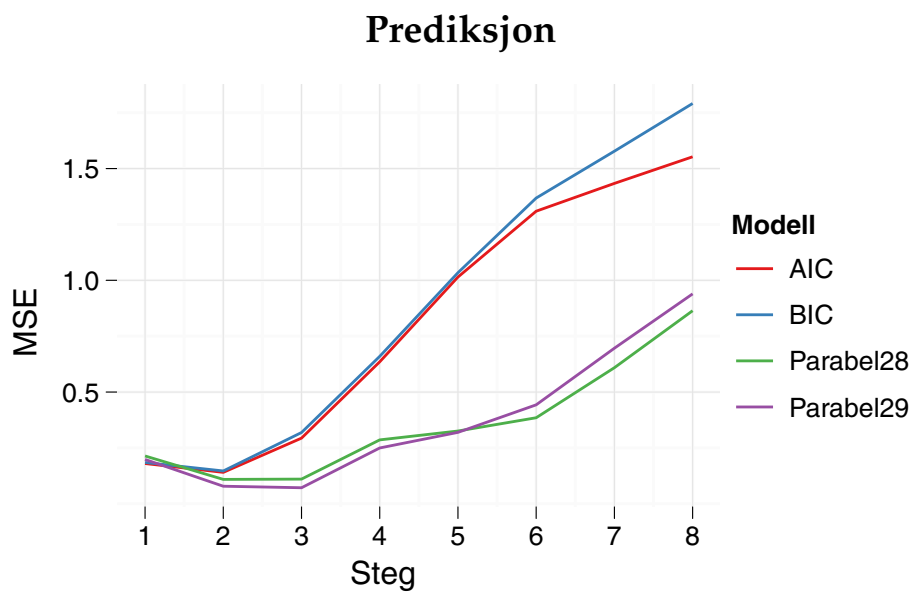
Tabell 3.5: Lengden til røtter, $z \in \mathbb{C}$, i $\phi(z)$ til de valgte modellene. AIC 1 er AIC-modellen på Tidsrekke 1 osv.

3.4.4 Evaluerer ARIMA-modellen

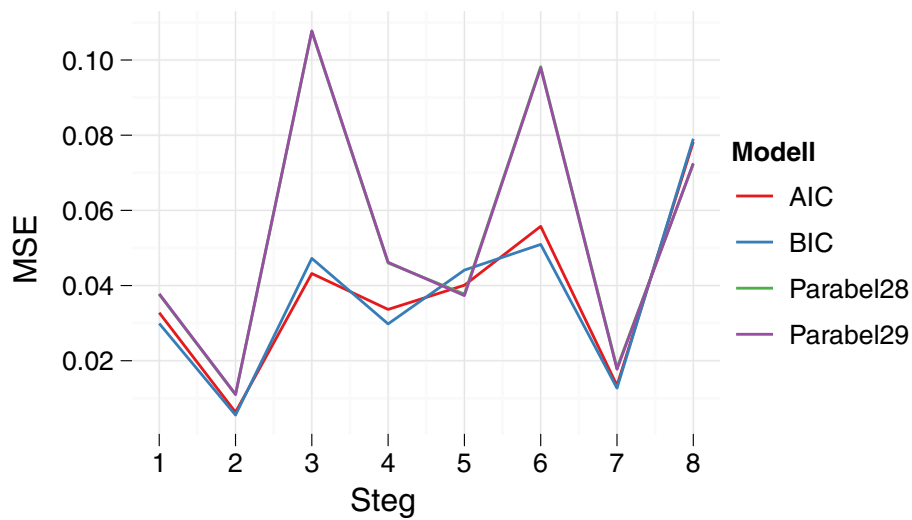
Nå skal vi evaluere ARIMA-modellene i prediksjon. I 3.2.3 så vi at prediksjonen, $P_n X_{n+h}$, for en kausal ARMA-prosess er gitt ved (3.4). Hvis, $(1 - B)^d X_t = \Delta^d X$, er en kausal ARMA-modell bruker vi (3.4) på de differensierte dataene, og transformerer så prediksjonene tilbake til X_t . For øvrig må vi anta at (X_{1-d}, \dots, X_0) er ukorrelert med $(1 - B)^d X_t$ for $t \geq 0$. Men siden vi bruker posisjonene, X_0, X_1 og X_2 , til å finne akselerasjonen ved $t = 2$, $(1 - B)^2 X_2$, holder ikke denne antakelsen her. Vi ser bort fra dette, siden det ble oppdaget sent i studiet. Jeg har predikert med funksjonen `predict()` gitt en estimert `arima()` i R. I figurene 3.16 og 3.17 vi har sammenlignet prediksjonsevnen til AIC- og BIC-modellene for Tidsrekke 1 og 2 med Parabelmetoden. Først har vi sammenlignet MSE-verdier for predikerte posisjoner. Det er nødvendig og integrere opp prediksjonsverdiene til ARIMA-modellene, fra akselerasjon til posisjon. Vi så at bare BIC-modellene er konsistente ARIMA-modeller for posisjonene, men tross dette skal vi sammenligne med AIC-modellene også. Vi har også sammenlignet prediksjonene på akselerasjonsnivå. Da har vi differensiert prediksjonsverdiene i Parabelmetoden.

Tidsrekke 1

I figur 3.1 så vi at Tidsrekke 1 er spesielt glatt i de markerte dataene. Vi mente dette er en hovedårsak til at Parabelmetoden er en god tilpasning her, spesielt god med 28 og 29 punkt. I figur 3.16 (a), ser vi at prediksjonsevnen til Parabelmetoden slår begge ARIMA-modellene på posisjonsnivå. Trolig fordi Tidsrekke 1 ser ut til å endre egenskaper i siste del av tidsrekken. Da



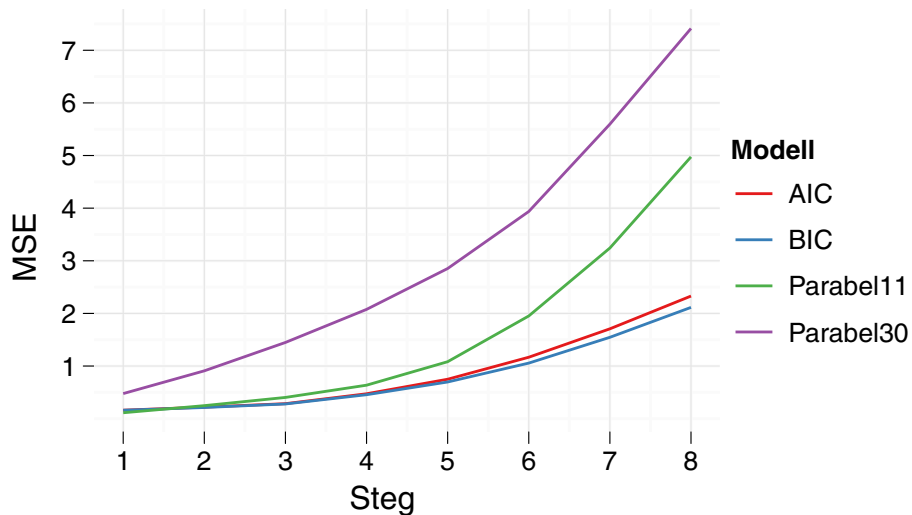
(a) Predikerer posisjon.



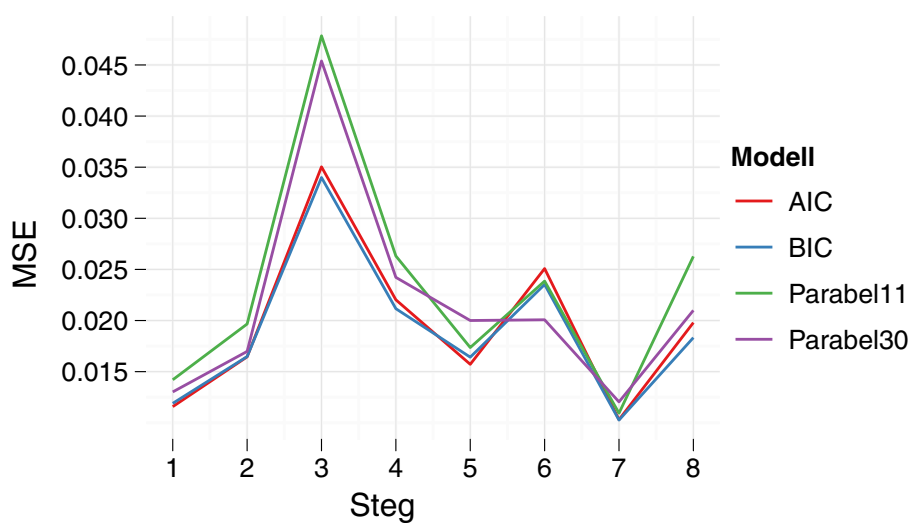
(b) Predikerer akselerasjon.

Figur 3.16: Sammenligner Parabelmetoden og AIC- og BIC-modellene i prediksjon på Tidsrekke 1, der Parabel28 indikerer at 28 datapunkter er brukt i tilpasning av parabelen osv.

Prediksjon



(a) Predikerer posisjon.



(b) Predikerer akselerasjon.

Figur 3.17: Sammenligner Parabelmetoden og AIC- og BIC-modellen i prediksjon på Tidsrekke 2, der Parabel11 indikerer at 11 datapunkter er brukt i tilpasning av parabelen osv.

hjelper det ikke om ARIMA-modellene er tilpasningsdyktige. AIC-modellen predikerer faktisk bedre enn BIC-modellen her. På akselerasjonsnivå, forholder tidsrekkemodellene og Parametoden seg annerledes. En årsak til at ARIMA-modellene slår Parabelmetoden på akselerasjonsnivå, skyldes at vi differensierer predikjonsverdiene i Parabelmetoden. Siden vi må predikere p_1, \dots, p_{10} for å finne tilhørende a_1, \dots, a_8 , der p_t og a_t er posisjon og akselerasjon ved tiden t , er Parabelmetoden mer usikker her. Parabel28 er gjemt bak Parabel29.

Dersom vi har grunn til å tro at evolusjonen til en node vil være veldig glatt, lønner det seg å bruke Parabelmetoden til å predikere posisjonen. På akselerasjonsnivå derimot, lønner det seg å bruke ARIMA-modellene selv når Parabelmetoden er overlegen på posisjonsnivå. For 5-steps prediksjon på akselerasjonsnivå, gjør faktisk Parabelmetoden det best.

Tidsrekke 2

ARIMA-modellene utkonkurrerer Parabelmetoden. Som forventet, siden prediksjonene med Parabelmetoden var dårligere her. Dessuten er egenskapene i de markerte dataene, likere egenskapene tidligere i tidsrekken, enn de var for Tidsrekke 1. Feilen i BIC-modellen for Tidsrekke 2, $\sigma^2 = 0.017$, er lavere enn feilen i BIC-modellen på Tidsrekke 1, $\sigma^2 = 0.029$ Ellers gjelder samme drøftelse, som for Tidsrekke 1.

3.5 Generalisering av ARMA-modellen

En mulig forbedring av den én-dimensjonale modellen er å inkludere parametrene, fart og akselerasjon, i en tre-dimensjonal tidsrekke-modell. I eksempel 2.3.1 og 2.3.2 illustrerte vi at korrelasjonen mellom parametrene gir sikrere estimat. En mulighet er å tilpasse en p -dimensjonal tidsrekke-modell, til p noder langs seismikk-kabel. Der avstanden mellom nodene varierer mellom 100-600 meter. Det er også korrelasjon mellom nodene langs kabelen fører til sikrere estimat. Å tilpasse en m -dimensjonal tidsrekke-modell for m noder som ligger parallelt med samme avstand fra båten er også en mulighet. Å kombinere disse multivariable-modellene er også mulig. Vi velger å ikke gå mer i detalj på disse modellene, siden de ikke er blitt nøye studert. Likevel definerer vi en multivariabel ARMA-modell, for vi vil trekke paralleller til denne modellen i avsnitt 4.4.2. La $X_t = (X_{t1}, \dots, X_{tp})'$ være en p -dimensjonal

stasjonær tidsrekke, med tilhørende forventning lik:

$$\boldsymbol{\mu}_t = \mathbb{E}(\mathbf{X}_t) = (\mathbb{E}(X_{t1}), \dots, \mathbb{E}(X_{tp}))'.$$

Videre er tilhørende kovariansmatrisen, $\Gamma(h) = \mathbb{E}(\mathbf{X}_{t+h} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})$ gitt ved:

$$\Gamma(h) = \begin{pmatrix} \gamma_{1,1}(h) & \gamma_{1,2}(h) & \dots & \gamma_{1,p}(h) \\ \gamma_{2,1}(h) & \gamma_{2,2}(h) & \dots & \gamma_{2,p}(h) \\ \vdots & \dots & \dots & \vdots \\ \gamma_{p,1}(h) & \gamma_{p,2}(h) & \dots & \gamma_{p,p}(h) \end{pmatrix}.$$

Betingelser for stasjonærhet i multivariable tidsrekker, er en rett frem utvidelse av stasjonærhet for én-dimensjonale tidsrekker. Det vil si, forventningen og kovariansmatrisen, over, er uavhengig av tiden. Når \mathbf{X}_t er stasjonær er, X_{ti} stasjonær $\forall i = 1, \dots, p$.

Definisjon 3.5.1: VARMA(p, q)-modellen

\mathbf{X}_t er en VARMA-modell hvis \mathbf{X}_t er stasjonær og tilfredstiller

$$\mathbf{X}_t - \Phi_1 \mathbf{X}_{t-1} - \dots - \Phi_p \mathbf{X}_{t-p} = \mathbf{Z}_t + \Theta_1 \mathbf{Z}_{t-1} + \dots + \Theta_q \mathbf{Z}_{t-q} \quad \forall t$$

hvor $\mathbf{Z}_t \sim WN(\mathbf{0}, \Sigma)$, $(\mathbf{I} - \Phi_1 z - \dots - \Phi_p z^p)$ og $(\mathbf{I} + \Theta_1 z + \dots + \Theta_q z^q)$ er polynomer, og \mathbf{I} er identitetsmatrisen.

Vi så at stasjonærhet og kausalitet for én variabel avhenger av røttene i ϕ . For flere variable er ARMA-modellen kausal og stasjonær dersom egenverdiene til Φ er mindre enn én i absoluttverdi. Vi kan også skrive denne modellen på kompaktform slik:

$$\Phi(B)\mathbf{X}_t = \Theta(B)\mathbf{Z}_t.$$

4

Sekvensielle metoder

I praksis er vi ute etter å estimere sekvensielt. Dette gir oss muligheten til å estimere og predikere posisjonen til noden fortløpende, slik at vi kan styre seismikk-kablene dit de skal ligge. Dette gjorde vi da vi predikerte posisjon og akselerasjon til noden rekursivt med (3.18), men det var klassisk frekventistisk inferens. Nå skal vi predikere med Kalmanfilteret, som er et Bayesiansk rekursivt filter, og derfor tillater oss å inkludere en a priori-modell i inferensen.

Vi utleder Bayes' formel, på sekvensielt vis, gitt en *gaussisk state-space* modell. Videre anvender vi Kalmanfilteret på state-space modeller tilpasset posisjonsdata og dobbeltdifferensierte data. Vi illustrerer óg hvordan andre tidsrekke-modeller kan skrives på state-space form.

4.1 Gaussiske State-space modeller

State-space modeller er den mest generelle tidsrekke-modellen vi omtaler i dette studiet, og omfatter blant annet ARIMA-modeller. Det er en lineær dynamisk modell som er en forutsetning for Kalmanfilteret, når støyen i modellen er gaussisk. Modellen ble utviklet fordi man ønsket å kontrollere lineære systemer. Blant annet, styring av kursen til raketter. Si en lineær modell beskriver kursen til en rakett, med kjente parametre. Underveis blir posisjonen

til raketten målt, der målingen er gitt ved den sanne verdien pluss støy slik:

$$Y_t = x_t + \epsilon_t.$$

Modellen som beskriver kursen til raketten blir oppdatert basert på disse målingene. Dersom måling indikerer at raketten er på avveie, blir den styrt mot riktig kurs. Evolusjonen til raketten over tid blir modellert slik:

$$X_t = \phi_t x_{t-1} + B_t u_t + \eta_t$$

der variabelen u_t signaliserer størrelsen på avviket, som må korrigeres for med hensyn på B_t . Parametrene, som beskriver kursen til raketten, er kjente og varierer med tiden. Vi kan trekke paralleller her, selv om parametrene som beskriver evolusjonen til noden er ukjente. Nedenfor definerer vi den gaussiske state-space modellen, siden vi antar at støyen i modellen er gaussisk.

Definisjon 4.1.1: Gaussisk State-Space modell

En state space-modell for tidsrekken Y_t består av to ligninger, gitt ved:

$$Y_t = \psi x_t + \epsilon_t \quad (4.1)$$

$$X_t = \phi x_{t-1} + \eta_t \quad (4.2)$$

for $t = 1, \dots, T$, der $\epsilon_t \sim \mathcal{N}(0, \tau^2)$, $\eta_t \sim \mathcal{N}(0, \sigma^2)$. Parametrene ψ og ϕ , skalarer.

Ligningene, (4.1) og (4.2), er kjent som, *observasjons-likningen* og *evolusjons-likningen*. Den første representerer en én-dimensjonal observasjon Y_t som er en lineær funksjon av den én-dimensjonale X_t pluss et feilledd. Evolusjons-ligningen er en Markov-prosess, som bestemmer X_t basert på x_{t-1} . Videre antar vi at X_0 er ukorrelert med alle støyleddene, og at $\mathbb{E}(\epsilon_t \eta_t) = 0$. Mer generelle state-space modeller, tillater korrelasjon mellom feilleddene i likningene. Vi kan også inkludere en kontroll-input, $B_t u_t$, i evolusjons-ligningen dersom vi er interessert i å styre prosessen.

Vi vil også styre noden, og derfor er det aktuelt å inkludere en kontroll-input i modellen. Likevel, siden vi estimerer parametrene i state-space modellen ut fra gitte data, og antar at disse er tidsuavhengige, vil vi først prøve å predikere noden uten å ta hensyn til styring. Uansett, i samtale med professorer innen kybernetikk og marin teknologi, blant dem Asgeir Sørensen ved Institutt for Marin teknologi ved NTNU, er en *PI-regulator* blitt anbefalt som styrings-input

i state-space modellen, der regulatorfunksjonen gis et signal u_t som er en funksjon av avviket noden har fra ønsket posisjon ved tiden t . Dette avviket, u_t , tilsvarer det signalet som gis birdsene. Dette kan studeres i Turkyilmaz (2004).

Vi har estimert state-space modeller på Tidsrekke 1 og 2, både på posisjonsdata og dobbeltdifferensierte data. Modellparametrene, $\hat{\phi}$, $\hat{\tau}^2$ og $\hat{\sigma}^2$, er estimert med *ML*, ved å maksimere en variant av den gaussiske Likelihood-funksjonen i (3.8), der forventningene i ϵ_t og η_t antas å være lik 0 og $\psi = 1$.

Tidsrekke 1

Først har vi estimert state-space modell på Tidsrekke 1:

$$Y_t = x_t + \epsilon_t \quad (4.3)$$

$$X_t = 0.999x_{t-1} + \eta_t \quad (4.4)$$

for $t = 1, \dots, 900$, der $\epsilon_t \sim \mathcal{N}(0, 0.016^2)$, $\eta_t \sim \mathcal{N}(0, 0.879^2)$ og initial-verdien til noden er $\hat{x}_0 = -43.14$ meter i krysslinje. Parameterverdiene er beregnet på de 900 første dataene i Tidsrekke 1 (se tabell A.8). Vi har også estimert følgende state-space modell på de dobbeltdifferensierte dataene i tidsrekken:

$$Y_t = x_t + \epsilon_t \quad (4.5)$$

$$X_t = 0.292x_{t-1} + \eta_t \quad (4.6)$$

for $t = 1, \dots, 900$, der $\epsilon_t \sim \mathcal{N}(0, 0.019^2)$, $\eta_t \sim \mathcal{N}(0, 0.225^2)$ og $\Delta^2 \hat{x}_0 = -5.11 \cdot 10^{-7}$. Parameterverdiene er beregnet på de 900 første dobbeltdifferensierte dataene i Tidsrekke 1 (se tabell A.9).

Tidsrekke 2

Tilsvarende har vi estimert en state-space prosess på dataene i Tidsrekke 2:

$$Y_t = x_t + \epsilon_t \quad (4.7)$$

$$X_t = 0.999x_{t-1} + \eta_t \quad (4.8)$$

for $t = 1, \dots, 2100$, der $\epsilon_t \sim \mathcal{N}(0, 0.011^2)$, $\eta_t \sim \mathcal{N}(0, 0.509^2)$ og $\hat{x}_0 = 48$. Parameterverdiene er beregnet på de 2100 første dataene i Tidsrekke 1 (se tabell A.10). State-space modellene for de dobbeltdifferensierte dataene i

Tidsrekke 2 gis ved:

$$Y_t = x_t + \epsilon_t \quad (4.9)$$

$$X_t = -0.076x_{t-1} + \eta_t \quad (4.10)$$

for $t = 1, \dots, 2100$, der $\epsilon_t \sim \mathcal{N}(0, 0.054^2)$, $\eta_t \sim \mathcal{N}(0, 0.139^2)$ og $\Delta^2 \hat{x}_0 = -9.38 \cdot 10^{-6}$. Parameterverdiene er beregnet på de 2100 første dobbeltdifferensierte dataene i Tidsrekke 1 (se tabell A.11). I avsnitt 4.3.4 anvender vi Kalmanfilteret på disse state-space modellene.

4.1.1 Generalisering av Gaussisk State-Space modell

Definisjon 4.1.2: Generalisering av Gaussisk State-Space modell

En state space-modell for den fler-dimensjonale tidsrekken Y_t består av to ligninger, gitt ved:

$$Y_t = Hx_t + \epsilon_t \quad (4.11)$$

$$X_t = Mx_{t-1} + \eta_t \quad (4.12)$$

for $t = 1, \dots, T$, der $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ og $\eta_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$.

Siden definisjonen over er en rett fram utvidelse av den én-dimensjonale state-space modellen, påpeker vi ikke flere betingelser ved denne modellen.

4.2 Bayes' formel og State-space modellen

Vi skal nå se på egenskapene til Bayes' formel på sekvensielt vis, gitt state-space modellen i definisjon 4.1.1. Ved Bayes' formel kan a posteriori-fordelingen til $X_{0:T}$ gitt $y_{1:T}$ uttrykkes ved:

$$p(x_{0:T} | y_{1:T}) \propto p(y_{1:T} | x_{0:T})p(x_{0:T}), \quad (4.13)$$

der $p(x_{0:T})$ representerer a priori-kunnskapen til den latente variabelen, x , og $p(y_{1:T} | x_{0:T})$ er den simultane data-fordelingen. A priori-prosessen er en Markovprosess når $p(x_t | x_{t-1}) = p(x_t | x_{t-1}, x_{t-2}, \dots, x_1)$. Dette er betinget uavhengighet, der utfallet ved tidspunkt t betinget på utfallet ved tidspunkt $t - 1$, er uavhengig av tidligere ledd i tidsrekken. Med Markov-egenskapen

kan den simultane a priori-fordelingen uttrykkes slik:

$$p(\mathbf{x}_{0:T}) = p(x_0) \prod_{t=1}^T p(x_t | x_{t-1}) \quad (4.14)$$

Bevis:

$$\begin{aligned} p(\mathbf{x}_{0:T}) &= p(x_0, x_1, \dots, x_t) \\ &= p(x_t | x_{t-1}, \dots, x_0) p(x_{t-1}, x_{t-2}, \dots, x_0) \\ &= p(x_t | x_{t-1}, \dots, x_0) p(x_{t-1} | x_{t-2}, \dots, x_0) \dots p(x_1 | x_0) p(x_0) \\ &= p(x_t | x_{t-1}) p(x_{t-1} | x_{t-2}) \dots p(x_1 | x_0) p(x_0) \\ &= p(x_0) \prod_{t=1}^T p(x_t | x_{t-1}) \end{aligned}$$

der $p(x_t | x_{t-1})$ er en overgangstetthet, og $p(x_0)$ er a priori-fordelingen til det initiale utfallet. En nødvendig antakelse er at målingene er uavhengige gitt den virkelige prosessen, som fører til at:

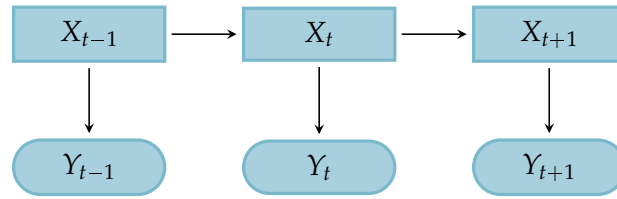
$$p(\mathbf{y}_{1:T} | \mathbf{x}_{0:T}) = \prod_{t=1}^T p(y_t | x_t). \quad (4.15)$$

Med (4.14) og (4.15) kan vi uttrykke (4.13) slik:

$$p(\mathbf{x}_{0:T} | \mathbf{y}_{1:T}) \propto p(x_0) \prod_{t=1}^T p(y_t | x_t) p(x_t | x_{t-1}).$$

Når a posteriori-fordelingen uttrykkes slik ser vi at $p(\mathbf{x}_{0:T} | \mathbf{y}_{1:T})$ oppdateres rekursivt, hver gang en ny måling er tilgjengelig. Når state-space modellen har disse egenskapene, er modellen en skjult Markov-prosess. Det vil si at x_t betinget på x_{t-1} er uavhengig av $x_{0:t-2}$, samtidig som y_t bare avhenger av verdien til x_t (se figur 4.1). Dette betyr selvsagt ikke at x_t er uavhengig av $x_{t-2}, x_{t-3}, \dots, x_1$.

Over så vi på den simultane a posteriori-fordelingen til $\mathbf{X}_{0:T}$, nå skal vi se på den marginale a posteriori-fordelingen til nåværende utfall X_t . Når vi finner



Figur 4.1: Skjult Markov prosess, der x_t er den skjulte variabelen og y_t er tilhørende observasjon.

$p(x_t | \mathbf{y}_{1:T})$ er vi interessert i tre ulike tilfeller:

$t = T \rightarrow$ filtrering

$t > T \rightarrow$ prediksjon

$1 < t < T \rightarrow$ glatting.

Vi ser nærmere på *filtrering* og *prediksjon*, som gir grunnlaget for Kalmanfilter. *Glatting* gir grunnlaget for Kalmansmoother, som vi ikke skal benytte.

4.2.1 Filtrering og Prediksjon

Ved filtrering antar vi at $p(x_{t-1} | \mathbf{y}_{1:t-1})$ er kjent. Når $t = 1$ er fordelingen lik $p(x_0)$. Dette benytter vi til å finne prediksjons-fordelingen $p(x_t | \mathbf{y}_{1:t-1})$, og deretter analyse-fordelingen $p(x_t | \mathbf{y}_{1:t})$. Basert på Bayes' formel og Markov-egenskapen er prediksjons-fordelingen gitt ved:

$$\begin{aligned}
 p(x_t | \mathbf{y}_{1:t-1}) &= \int p(x_t, x_{t-1} | \mathbf{y}_{1:t-1}) dx_{t-1}, \\
 &= \int p(x_t | x_{t-1}, \mathbf{y}_{1:t-1}) p(x_{t-1} | \mathbf{y}_{1:t-1}) dx_{t-1}, \\
 &= \int p(x_t | x_{t-1}) p(x_{t-1} | \mathbf{y}_{1:t-1}) dx_{t-1}. \tag{4.16}
 \end{aligned}$$

Analyse-fordelingen finner vi ved Bayes'formel slik:

$$\begin{aligned}
 p(x_t | \mathbf{y}_{1:t}) &= p(x_t | y_t, \mathbf{y}_{1:t-1}), \\
 &\propto p(x_t, y_t | \mathbf{y}_{1:t-1}) p(x_t | \mathbf{y}_{1:t-1}), \\
 &= p(y_t | x_t, \mathbf{y}_{1:t-1}) p(x_t | \mathbf{y}_{1:t-1}), \\
 &= p(y_t | x_t) p(x_t | \mathbf{y}_{1:t-1}). \tag{4.17}
 \end{aligned}$$

Repeterer vi (4.16) og (4.17) hver gang en ny måling er tilgjengelig, ender vi opp med parene:

$p(x_1), p(x_1 | y_1), p(x_2 | y_1), p(x_2 | \mathbf{y}_{1:2}), \dots, p(x_T | \mathbf{y}_{1:T-1}), p(x_T | \mathbf{y}_{1:T})$. Siden vi antar at alle de nevnte fordelingene er gaussiske, forenkles utregningene. Senere skal vi finne forventningen og variansen til prediksjons- og analysefordelingen.

4.3 Kalmanfilter

Kalmanfilteret(KF) er et Bayesiansk rekursivt filter, som estimerer x optimalt gitt en gaussisk state-space modell. KF er oppkalt etter Rudolf E. Kalman som publiserte en artikkel om KF i 1960. Senere utviklet Stratonovich et mer generelt filter, også for ikke-linære modeller. Året etter i 1961, publiserte Kalman og Bucy et oppdatert KF. Dette forklarer at KF også blir kalt, Stratonovich-Kalman-Bucy filteret. KF har siden utviklet seg i ulike retninger. De mest kjente er Extended Kalman filter (EKF), og Ensemble Kalmanfilter (EnKF).

4.3.1 Likningene i Kalmanfilteret

KF kan deles inn i to hovedfaser: prediksjonsfasen og analysefasen. I prediksjonsfasen predikerer vi den ukjente, x_t , gitt alle målingene til og med y_{t-1} , $\mathbf{y}_{1:t-1}$, og vi søker derfor $p(x_t | \mathbf{y}_{1:t-1})$. Forventningen til denne fordelingen kaller vi *a priori estimatet*, og vi bruker notasjonen $x_{t|t-1}$. Straks en ny måling er tilgjengelig, y_t , oppdateres fordelingen i prediksjonsfasen til a posteriorifordelingen, $p(x_t | \mathbf{y}_{1:t})$. Dette estimatet kaller vi *a posteriori estimatet* og vi skriver $x_{t|t}$. Vi oppsummerer med følgende:

$$\begin{aligned} x_{t|t-1} &= \mathbb{E}(x_t | \mathbf{y}_{1:t-1}) \\ x_{t|t} &= \mathbb{E}(x_t | \mathbf{y}_{1:t}). \end{aligned} \quad (4.18)$$

Videre kan vi definere de betingede kovariansmatrisene til feilledet, $x_t - x_{t|t}$, $\mathbf{P}_{t|t}$, og til feilledet, $x_t - x_{t|t-1}$, $\mathbf{P}_{t|t-1}$, slik:

$$\begin{aligned} \mathbf{P}_{t|t-1} &= \mathbb{E}((x_t - x_{t|t-1})(x_t - x_{t|t-1})' | \mathbf{y}_{1:t-1}) \\ \mathbf{P}_{t|t} &= \mathbb{E}((x_t - x_{t|t})(x_t - x_{t|t})' | \mathbf{y}_{1:t}). \end{aligned} \quad (4.19)$$

Siden forventningen i (4.18) er lineær og forventningsrett, samtidig som den er minimerer (4.19), er $x_{t|t}$ et såkalt *Best Linear Unbiased estimate* (BLUE). Selv om forventningene er forventningsrette, betyr ikke dette at estimatoren er troverdig, samtidig må feilleddet varierer lite rundt 0. Endelig, siden vi antar at $X_t | \mathbf{y}_{1:t-1}$ er gaussisk får vi:

$$\begin{aligned} X_t | \mathbf{y}_{1:t-1} &\sim \mathcal{N}(x_{t|t-1}, \mathbf{P}_{t|t-1}) \\ X_t | \mathbf{y}_{1:t} &\sim \mathcal{N}(x_{t|t}, \mathbf{P}_{t|t}). \end{aligned}$$

4.3.2 Utleddning av likningene

Prediksjonsfasen

Når vi utleder definisjonene i avsnittet over bruker vi Bayes' Formel. I (4.16) så vi at prediksjons-fordelingen gis ved Bayes' Formel slik: $p(x_t | \mathbf{y}_{t-1}) = \int p(x_t | x_{t-1})p(x_{t-1} | \mathbf{y}_{t-1}) dx_{t-1}$. Første sannsynlighetsfordeling på høyre side er AR(1)-modellen, $X_t | x_{t-1} \sim \mathcal{N}(\phi x_{t-1}, \sigma^2)$, og neste, a posterior-fordelingen ved forrige tidssteg, $X_{t-1} | \mathbf{y}_{t-1} \sim \mathcal{N}(x_{t-1|t-1}, \mathbf{P}_{t-1|t-1})$. Siden normalfordelingen er kjent ved forventning og varians, trengs ikke integralet løses for å finne prediksjonsfordelingen. La oss heller anvende regler for betinget forventning (2.3) og betinget varians, (2.4) der:

$$\begin{aligned} x_{t|t-1} &= \mathbb{E}(x_t | \mathbf{y}_{1:t-1}) = \mathbb{E}(\mathbb{E}(x_t | x_{t-1}) | \mathbf{y}_{1:t-1}) \\ &= \mathbb{E}(\phi x_{t-1} | \mathbf{y}_{1:t-1}) = \phi x_{t-1|t-1}. \end{aligned}$$

Videre er variansen gitt ved:

$$\begin{aligned} \mathbf{P}_{t|t-1} &= \text{Var}[x_t | \mathbf{y}_{1:t-1}] \\ &= \mathbb{E}(\text{Var}(x_t | x_{t-1}) | \mathbf{y}_{1:t-1}) + \text{Var}(\mathbb{E}(x_t | x_{t-1}) | \mathbf{y}_{1:t-1}) \\ &= \mathbb{E}(\sigma^2 | \mathbf{y}_{1:t-1}) + \text{Var}(\phi x_{t-1} | \mathbf{y}_{1:t-1}) \\ &= \sigma^2 + \phi^2 \mathbf{P}_{t-1|t-1}. \end{aligned}$$

Analysefasen

I (4.17) så vi at den gaussiske a posterior-fordelingen i analysefasen, er gitt ved: $p(x_t | \mathbf{y}_{1:t}) \propto p(y_t | x_t)p(x_t | \mathbf{y}_{t-1})$. Den første sannsynlighetsfordelingen er data-fordelingen, $Y_t | x_t \sim \mathcal{N}(\psi x_t, \tau^2)$, og neste prediksjons-fordelingen fra

prediksjons-fasen, $X_t | \mathbf{y}_{t-1} \sim \mathcal{N}(x_{t|t-1}, \mathbf{P}_{t|t-1})$. I avsnitt 2.2.1 fant vi fordelingen til $X | \mathbf{y}$. Vi bruker samme fremgangsmåte her til å finne $p(x_t | \mathbf{y}_{1:t})$, hvor a priori-fordelingen vår byttes ut med prediksjons-fordelingen $p(x_t | \mathbf{y}_{1:t-1})$. Vi får:

$$\begin{aligned}
 p(\mathbf{y}_{1:t} | x_t) &= \frac{1}{(2\pi)^{\frac{T}{2}} \tau^T} \exp\left(-1/2 \sum_{t=1}^T (y_t - \psi x_t)^2 / \tau^2\right) \\
 p(x_t | \mathbf{y}_{1:t}) &\propto \exp\left(-1/2 \left(\sum_{t=1}^T (y_t - \psi x_t)^2 / \tau^2 + (x_t - x_{t|t-1})^2 / P_{t|t-1}\right)\right) \\
 &= \exp\left(-1/2 (x_t^2 (\psi^2 T / \tau^2 + 1 / P_{t|t-1})^2 \right. \\
 &\quad \left. - 2x_t (\sum_{t=1}^T \psi y_t / \tau^2 - x_{t|t-1} / P_{t|t-1}) + C)\right) \\
 &\propto \exp\left(-\frac{(\psi^2 T / \tau^2 + 1 / P_{t|t-1})}{2} \right. \\
 &\quad \left. (x_t - (\psi \sum_{t=1}^T y_t / \tau^2 - x_{t|t-1} / P_{t|t-1})) (\psi^2 T / \tau^2 + 1 / P_{t|t-1})^{-1}\right)^2.
 \end{aligned}$$

A posteriori-fordelingen er dermed gitt ved:

$$\begin{aligned}
 X_t | \mathbf{y}_t &\sim \mathcal{N}\left((\psi^2 T / \tau^2 + 1 / P_{t|t-1})^{-1} \right. \\
 &\quad \left. (\psi \sum_{t=1}^T y_t / \tau^2 - x_{t|t-1} / P_{t|t-1}), (\psi^2 T / \tau^2 + 1 / P_{t|t-1})^{-1}\right).
 \end{aligned}$$

I avsnitt 2.2.1 ble forventningen og variansen uttrykt ved gainet K . Tilsvarende, uttrykker vi forventningen over ved *Kalmangainet* K_t slik:

$$\begin{aligned}
 \mathbb{E}(x_t | \mathbf{y}_{1:t}) &= (\psi^2 T / \tau^2 + 1 / P_{t|t-1})^{-1} (\psi \sum_{t=1}^T y_t / \tau^2 - x_{t|t-1} / P_{t|t-1}) \\
 &= \frac{\tau^2 P_{t|t-1}}{\tau^2 + \psi^2 T P_{t|t-1}} (\psi \sum_{t=1}^T y_t / \tau^2 - x_{t|t-1} / P_{t|t-1}) \\
 &= x_{t|t-1} + K_t (\bar{y} - \psi x_{t|t-1}),
 \end{aligned}$$

der $K_t = \frac{\psi^2 TP_{t|t-1}}{\tau^2 + \psi^2 TP_{t|t-1}}$. På lignende vis som i avsnitt 2.2.1 uttrykker vi også variansen over slik:

$$\begin{aligned}\text{Var}(x_t | \mathbf{y}_{1:t}) &= \frac{\tau^2 P_{t|t-1}}{\tau^2 + \psi^2 TP_{t|t-1}} + \frac{\psi^2 TP_{t|t-1}^2}{\tau^2 + \psi^2 TP_{t|t-1}} - \frac{\psi^2 TP_{t|t-1}^2}{\tau^2 + \psi^2 TP_{t|t-1}} \\ &= \left(\frac{\tau^2 + \psi^2 TP_{t|t-1}}{\tau^2 + \psi^2 TP_{t|t-1}} - \frac{\psi^2 TP_{t|t-1}}{\tau^2 + \psi^2 TP_{t|t-1}} \right) P_{t|t-1} \\ &= (1 - \psi K_t) P_{t|t-1}.\end{aligned}$$

Kalmangainet, K_t , optimaliserer a posteriori-fordelingen gitt en gaussisk state-space prosess, dette vises i N.A.Thacker og A.J.Lacey (2006).

4.3.3 Kalmanfilter-Algoritmen

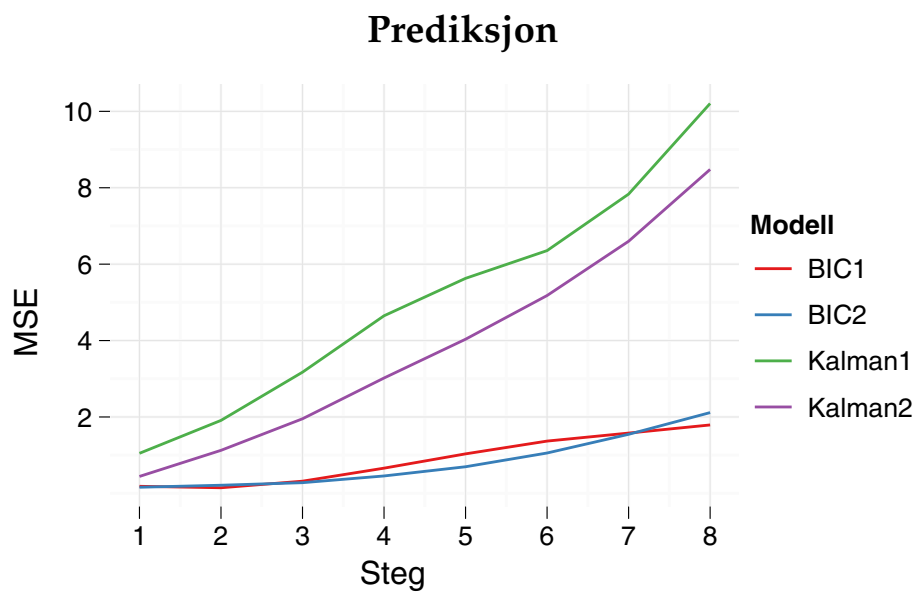
KF-algoritmen er rekursiv og består av noen få operasjoner. Gitt den gaussiske state-space modell i (4.1) og (4.2), der $\psi = 1$, og \hat{x}_0 er KF-algoritmen gitt ved:

$$\begin{aligned}x_{t|t-1} &= \phi x_{t-1|t-1} \\ P_{t|t-1} &= \sigma^2 + \phi^2 P_{t-1|t-1} \\ K_t &= \frac{TP_{t|t-1}}{\tau^2 + TP_{t|t-1}} \\ x_{t|t} &= x_{t|t-1} + K_t (y_t - \psi x_{t|t-1}) \\ P_{t|t} &= (1 - K_t) P_{t|t-1},\end{aligned}$$

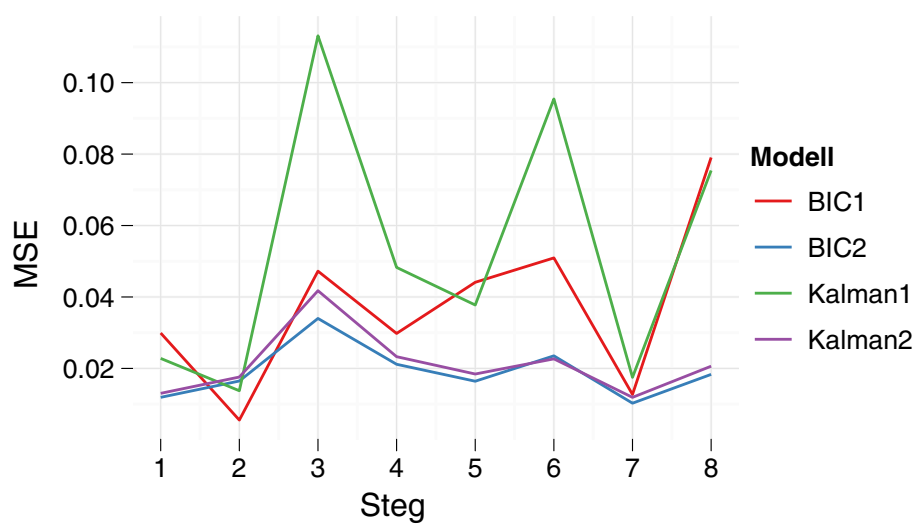
der T er antall målinger og K_t oppdateres hver gang en ny måling er tilgjengelig. Dersom det mangler observasjoner, y_t , utelukkes de tre siste likhetene.

4.3.4 Kalmanfilteret på Tidsrekke 1 og 2

Siden vi antar at evolusjonen til noden er et lineært dynamisk system på state-space form med gaussisk støy, har vi riktig grunnlag til å anvende Kalmanfilteret. Vi har kjørt KF på state-space modellene, tilpasset posisjonsdata og dobbeltdifferensierte data, på Tidsrekke 1 og 2. Det blir predikert 1- til 8-steg. Videre er prediksjonene utført på samme data som med ARIMA-modellene og Parabelmetoden, for sammenligningen sin del. Kalmanfilteret fungerer bra som optimal estimator når målinger og/eller modellen er nøyaktig. I prediksjon derimot, er KF avhengig av at modellen representerer tidsrekken godt. Desto



(a) Predikerer posisjon.



(b) Predikerer akselerasjon.

Figur 4.2: Sammenligner KF anvendt på state-space modellene med de respektive BIC-modellene i prediksjon på Tidsrekke 1 og 2. BIC1 betyr BIC-modellen på Tidsrekke 1 osv.

lengre frem vi predikerer, desto nærmere vil prediksjonen ligge gjennomsnittet i modellen. Hvor raskt prediksjonene tilnærmer seg gjennomsnittet i modellen, avhenger av usikkerheten i målingene versus usikkerheten i modellen.

Når vi predikerer posisjon på Tidsrekke 1 og 2 med KF anvendt på state-space modellene i (4.3) (4.4) og (4.7) (4.8) repektivt, predikerer KF best på Tidsrekke 2. Videre predikerer de repektive BIC-modellene best (se figur 4.2 (a)).

I state-space modellen som er tilpasset posisjonsdata på Tidsrekke 1 er $\tau = 0.016$ og $\sigma = 0.879$. Tilsvarende er $\tau = 0.011$ og $\sigma = 0.509$ i state-space modellen tilpasset Tidsrekke 2. Den lavere usikkerheten i sistnevnte modell, bekrefter hvorfor KF predikerer best basert på denne modellen. Feilen i de respektive BIC-modellene er $\sigma = 0.539$ for Tidsrekke 1 og $\sigma = 0.412$ for Tidsrekke 2, altså lavere enn feilen i evolusjons-ligningene til de tilhørende state-space modellene. Da er det heller ikke overraskende at BIC-modellene gjør det best i prediksjon av posisjon her.

Vi har også predikert de dobbeltdifferensierte dataene, som er proporsjonale med akselerasjon, i Tidsrekke 1 og 2 (se figur 4.2 (b)). I (4.5) og (4.6) så vi at feilen i state-space modellen på Tidsrekke 1 er gitt ved $\tau = 0.019$ og $\sigma = 0.225$. Tilsvarende er $\tau = 0.054$ og $\sigma = 0.139$ feilen i (4.9) og (4.10) på Tidsrekke 2. For Tidsrekke 1 antar state-space modellen at målingene er relativt sikre, mens modellen er usikker. State-space modellen på Tidsrekke 2 antar det motsatte, at målingene er usikre i forhold til modellen. Dette fører til at prediksjonene på Tidsrekke 2 er tilnærmet lik gjennomsnittet i modellen, $\phi\hat{x}_{t|t}$, allerede i 2-steps prediksjonen, mens prediksjonene på Tidsrekke 1 velger å støtte seg på målingene i flere steg. Sett i forhold til BIC-modellene, predikerer KF nesten like bra på Tidsrekke 2, mens relativt dårlig på Tidsrekke 1.

I alle tilfeller predikere BIC-modellene best. Dette er ikke overraskende når vi ser hvilken state-space modell BIC-modellene representerer i (4.30) og (4.31), nemlig en state-space modell på vektorform, som selvsagt bør beskrive mer av strukturen til den virkelige prosessen, x .

4.4 Generalisering av Kalmanfilteret

Ideelt sett inngår det flere enn én parameter i problemstillingen, vi har begrenset oss til tre, nemlig posisjon, fart og akselerasjon. Siden vi modellerer posisjon, fart og akselerasjon i krysslinjeretning er verdiene små, der akselera-

sjonen er tilnærmet null. Det som skaper akselerasjon og fart er blant annet havstrømmer bølger og trekkraft fra kabelen. Vi kan tenke oss at hvis punktet på kabelen ved $t = 0$ ligger på krysslinsen, så vil neste punkt også ligge her dersom akselerasjonen og farten er 0. De tre parameterene er derfor svært korrelerte, der posisjonen kan uttrykkes ved akselerasjon, og motsatt.

4.4.1 Likningene i Fler-dimensjonalt Kalmanfilteret

I avsnitt 3.2.1 fant vi a posteriori-fordelingen til $\mathbf{X} \mid \mathbf{y}$. På samme vis finner vi uttrykk for a posteriori-fordelingen, for flere dimensjoner, sekvensielt gitt state-space modellen i (4.11) og (4.12) her. Siden vi har utledet tilsvarende ligninger før, kommenterer vi utledningene i liten grad her.

$$\begin{aligned} \mathbf{X}_t \mid \mathbf{y}_t &\sim \\ \mathcal{N}((\mathbf{H}'\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}_{t|t-1}^{-1})^{-1}(\mathbf{H}'\mathbf{R}^{-1}\mathbf{y}_t + \mathbf{P}_{t|t-1}^{-1}\mathbf{x}_{t|t-1}), (\mathbf{H}\mathbf{R}^{-1}\mathbf{H} + \mathbf{P}_{t|t-1}^{-1})) \end{aligned}$$

Der a priori-fordelingen og data-fordelingen er gitt ved:

$$\begin{aligned} \mathbf{X}_t &\sim \mathcal{N}(x_{t|t-1}, \mathbf{P}_{t|t-1}) \\ \mathbf{Y}_t \mid \mathbf{x}_t &\sim \mathcal{N}(\mathbf{H}\mathbf{x}_t, \mathbf{R}). \end{aligned}$$

A priori-estimatet finner vi ved å anvende regler for betinget forventning slik:

$$\begin{aligned} \mathbf{x}_{t|t-1} &= \mathbb{E}(\mathbf{x}_t \mid \mathbf{y}_{t-1}) = \mathbb{E}(\mathbb{E}(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \mid \mathbf{y}_{t-1}) \\ &= \mathbb{E}(\mathbf{M}\mathbf{x}_{t-1} \mid \mathbf{y}_{t-1}) = \mathbf{M}\mathbf{x}_{t-1|t-1}. \end{aligned}$$

Tilhørende kovariansmatrise finner vi ved å anvende regler for betinget varians slik:

$$\begin{aligned} \mathbf{P}_{t|t-1} &= \text{Var}(\mathbf{x}_t \mid \mathbf{y}_{t-1}) \\ &= \mathbb{E}(\text{Var}(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \mid \mathbf{y}_{t-1}) + \text{Var}(\mathbb{E}(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \mid \mathbf{y}_{t-1}) \\ &= \mathbb{E}(\mathbf{Q} \mid \mathbf{y}_{t-1}) + \text{Var}(\mathbf{M}\mathbf{x}_{t-1} \mid \mathbf{y}_{t-1}) \\ &= \mathbf{Q} + \mathbf{M}\mathbf{P}_{t-1|t-1}\mathbf{M}'. \end{aligned}$$

der

$$\begin{aligned} P_{t-1|t-1} &= \mathbb{E}(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1})' \\ &= \mathbb{E} \left(\begin{array}{cccc} \mathbf{x}_{1,t-1} - \mathbf{x}_{1,t-1|t-1} & & & \\ & \dots & & \\ & & \mathbf{x}_{p,t-1} - \mathbf{x}_{p,t-1|t-1} & \\ & & & \dots \end{array} \right) \begin{pmatrix} \mathbf{x}_{1,t-1} - \mathbf{x}_{1,t-1|t-1} \\ \dots \\ \mathbf{x}_{p,t-1} - \mathbf{x}_{p,t-1|t-1} \\ \dots \end{pmatrix} \end{aligned}$$

Endelig finner vi a priori- og a posteriori-estimatet uttrykt ved \mathbf{K}_t .

$$\begin{aligned} \mathbf{x}_{t|t} &= \mathbf{x}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}\mathbf{x}_{t|t-1}) \\ P_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{H})P_{t|t-1}. \end{aligned}$$

der $\mathbf{K}_t = P_{t|t-1}\mathbf{H}'(\mathbf{H}P_{t|t-1}\mathbf{H}' + \mathbf{R})^{-1}$. Videre er Kalman-algoritmen gitt state-space modellen i (4.11) og (4.12) og $\hat{\mathbf{x}}_0$ gitt ved:

$$\begin{aligned} \mathbf{x}_{t|t-1} &= \mathbf{M}\mathbf{x}_{t-1|t-1} \\ P_{t|t-1} &= \mathbf{Q} + \mathbf{M}'P_{t-1|t-1}\mathbf{M} \\ \mathbf{K}_t &= P_{t|t-1}\mathbf{H}'(\mathbf{H}P_{t|t-1}\mathbf{H}' + \mathbf{R})^{-1} \\ \mathbf{x}_{t|t} &= \mathbf{x}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}\mathbf{x}_{t|t-1}) \\ P_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{H})P_{t|t-1}. \end{aligned}$$

4.4.2 3-dimensjonalt Kalmanfilter på Tidsrekke 1

I avsnitt 3.3.2 så vi på X_t , uttrykt ved deterministisk trend og støy. Modellen var lite passende, fordi den faktiske trenden varierer. Hvis vi tillater trenden å utvikle seg stokastisk slik:

$$M_t = M_{t-1} + V_t \quad V_t \sim WN(0, \sigma_v^2), \quad (4.20)$$

gir dette oss state-space modellen: (1) $Y_t = M_t + W_t$ og (2) $M_t = M_{t-1} + V_t$. En slik, *lokal trend*-modell, blir brukt dersom det forventes at trenden vil ligge innenfor visse grenser. Modellen kan også utvides ved å inkludere en lokal kvadratisk trend. La fart F_t og akselerasjon A_t være gitt ved:

$$F_t = F_{t-1} + U_t \quad U_t \sim WN(0, \sigma_u^2), \quad (4.21)$$

$$A_t = A_{t-1} + Q_t \quad Q_t \sim WN(0, \sigma_q^2), \quad (4.22)$$

der $M_t = M_{t-1} + F_{t-1} + A_{t-1} + V_t$. For å uttrykke en lokal kvadratisk trendmodell på state-space form skriver vi $\mathbf{X}_t = (M_t, F_t, A_t)'$ og

$$\mathbf{X}_t = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{X}_{t-1} + \mathbf{P}, \quad (4.23)$$

der \mathbf{P} er tilhørende kovariansmatrise, med variansene σ_u^2 , σ_q^2 og σ_w^2 på diagonalen og ellers kovarianser. Målingen Y_t blir representert ved observasjonslikningen:

$$\mathbf{Y}_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{X}_{t-1} + W_t \quad W_t \sim WN(0, \sigma_w^2). \quad (4.24)$$

Eller ved:

$$\mathbf{Y}_t = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \mathbf{X}_{t-1} + \mathbf{W}_t \quad \mathbf{W}_t \sim WN(\mathbf{0}, \mathbf{R}), \quad (4.25)$$

hvis målinger av akselerasjon er tilgjengelig, der \mathbf{R} er en 2×2 kovariansmatrise. Dette er ikke en stasjonær state-space representasjon, siden egenverdiene til \mathbf{A} ligger på enhetssirkelen i \mathbb{C} , men stasjonærhet er hverken en forutsetning i state-space modeller eller i Kalmanfilteret. Merk at verdiene på diagonalen i \mathbf{A} tilsvarer koeffisientene i (4.20), (4.21) og (4.22). Ellers er det lineære systemet bestemt i henhold til Newtons' lover der $\Delta t = 1$.

Eksempel 4.4.1: KF på en Lokal Kvadratisk-Trend modell

Vi illustrerer her at modellen i, (4.23) og (4.25), gir sikrere prediksjon enn, (4.23) og (4.24). For å påpeke dette har vi anvendt KF på to ulike a priori-modeller. I den første, er kun måling av posisjon gitt, der:

$$F_t = 0.96F_{t-1} + U_t, \quad (4.26)$$

$$\mathbf{X}_t = \begin{bmatrix} 0.99 & 1 & \frac{1}{2} \\ 0 & 0.96 & 1 \\ 0 & 0 & 0.29 \end{bmatrix} \mathbf{X}_{t-1} + \mathbf{P}, \quad (4.27)$$

$$\mathbf{Y}_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{X}_{t-1} + W_t, \quad (4.28)$$

videre er $U_t \sim WN(0, 0.054^2)$ og P er en 3×3 kovariansmatrise med variansene σ_u^2 , σ_q^2 og σ_w^2 på diagonalen, og null ellers. Den øverste modellen i (4.26) er en gaussisk state-space modell tilpasset de første 900 differensierede dataene i Tidsrekke 1. Parameterverdiene ellers er tatt fra state-space modellene i (4.3) (4.4) og (4.5) (4.6). A priori-fordelingen over har vi sammenlignet med en tilsvarende modell, bortsett fra at observasjons-ligningen er gitt ved:

$$Y_t = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} X_{t-1} + W_t, \quad (4.29)$$

der $R = \begin{bmatrix} 0.015^2 & 0 \\ 0 & 0.019^2 \end{bmatrix}$. Siden målinger av akselerasjon ikke var tilgjengelig under studiet, har vi erstattet målinger med estimat av akselerasjon. Eksempelet er derfor ikke reelt.

Begge modellene er stasjonære, i likhet med modellene tidligere i studiet. Vi observerer at kovariansene i P raskt blir ulik 0 i begge tilfeller. Alt i alt, predikerer KF anvendt på den sistnevnte modellen mer sikkert enn anvendt på modellen med kun måling av posisjon, (se tabell 4.1).

M \ S	1	2	3	4	5	6	7	8
1	1.0631	1.7650	2.4846	3.2389	4.0285	4.8507	5.7020	6.5791
2	0.9908	1.5523	2.1105	2.7097	3.3609	4.0625	4.8101	5.5982

Tabell 4.1: \sqrt{MSE} -verdier ved prediksjon på Tidsrekke 1 med 3-dimensjonalt KF gitt a priori-modellene gitt i eksempel 4.4.1. M står for modell, og S står for steg. M=1 tilsvarer (4.27) og (4.29), mens M=2 tilsvarer (4.28) og (4.29).

4.5 ARMA-modeller på State-space form

ARMA-modeller kan skrives på state-space form og motsatt, men state-space modellen er mer generell. Modellen har en enkel struktur, hvor evolusjonslikningen er en Markov-prosess. Ofte er evolusjons-ligningen på vektorform,

og vi vet f.eks at AR(2)-modellen kan skrives slik:

$$\begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} + \begin{pmatrix} Z_t \\ 0 \end{pmatrix}$$

Vi skal nå se hvordan vi kan uttrykke en kausal ARMA(p, q)-modell på state-space form. Merk at det ikke finnes en unik state-space form for ARMA-modeller. La ARMA(p, q)-modellen være definert slik som i definisjon 3.2.5. Videre la $r = \max(p, q + 1)$, og $\phi_j = 0$ $j > p$, $\theta_j = 0$ når $j > q$, og $\theta_0 = 1$. Hvis U_t er en kausal AR(p)-modell som tilfredstiller $\phi(B)U_t = Z_t$, da er $Y_t = \theta(B)U_t$, siden $\phi(B)Y_t = \phi(B)\theta(B)U_t = \theta(B)Z_t$. Vi får at:

$$Y_t = [\theta_{r-1}, \theta_{r-2}, \dots, \theta_0] \mathbf{X}_t$$

hvor

$$\mathbf{X}_t = \begin{bmatrix} U_{t-r+1} \\ U_{t-r+2} \\ \vdots \\ U_t \end{bmatrix}$$

og evolusjons-ligningen, på en generell form, er gitt ved:

$$\mathbf{X}_t = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \phi_r & \phi_{r-1} & \phi_{r-2} & \dots & \phi_1 \end{bmatrix} \mathbf{X}_{t-1} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} Z_{t+1}$$

Vi kan derfor skrive BIC-modellen for Tidsrekke 1 på state-space form slik:

$$Y_t = [0.11, 0.16, -0.55, 1.16, -2.00, 2.05, -1.87, 1] \mathbf{X}_t \quad (4.30)$$

og

$$\mathbf{X}_t = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0.57 & -1.29 & 2.24 & -2.51 & 1.92 \end{bmatrix} \mathbf{X}_{t-1} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} Z_{t+1}. \quad (4.31)$$

For øvrig, merke at state-space prosessen i (4.1) og (4.2) kan uttrykkes på VAR(1) form slik:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{bmatrix} 0 & \phi \\ 1 & 0 \end{bmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix}.$$

Det skal også finnes en ARMA(p, q)-modell for (4.1) og (4.2), men det ble dessverre ikke tid til å studere denne transformasjonen tilstrekkelig. Det hadde vært en fin mulighet til å sammenligne ARIMA(p, q)-modellene med state-space modellene.

5

Konklusjoner

5.1 Prediksjon av posisjon

Hovedmålet med denne oppgaven var å studere hvor sikkert vi kan predikere posisjonen til seismikk-kabler. Dette fordi vi vil styre seimikk-kablene nærmest mulig preplot til enhver tid. Vårt fokus har vært å predikere posisjonen til en node 1- til 6-steg frem i tid. Vi har anvendt optimale prediksjons-algoritmer, gitt de ulike tidsrekke-modeller som vi har estimert.

Hvilken modell og prediksjonsalgoritme bør vi velge? Vi har valgt å tilpasse relativt enkle modeller, og kan ikke forvente at modellene predikerer den sanne posisjonen til noden. Begrensningen i prediksjonen ligger i modellen, siden algoritmene er optimale gitt at modellen er korrekt. Da vi predikerte posisjoner på Tidsrekke 1 var Parabelmetoden best, mens på Tidsrekke 2 var BIC-modellen best. Vi diskuterte hvorfor i avnitt 3.4.4. Generelt antar jeg at en ARIMA-modell basert på BIC-kriteriet predikerer posisjonen til noden bedre enn Parabelmetoden, med mindre det forventes at noden følger en ekstremt glatt kurve.

Første rad i tabell 5.1 viser \sqrt{MSE} -verdier ved prediksjon av posisjon til noden i meter fra krysslinje. \sqrt{MSE} -verdier er valgt basert på de metodene som ga sikrest prediksjon i hvert steg. For Tidsrekke 1 tilsvarer \sqrt{MSE} -verdiene

prediksjon med Parabel29 for 1- til 5-steps prediksjon, og med Parabel28 for 6- til 8-steps prediksjon. For Tidsrekke 2 tilsvare \sqrt{MSE} -verdiene prediksjon med Parabel11 for første steg, og med BIC-modellen for 2- til 8 steg. Hvorvidt disse prediksjonsverdiene er sikre nok, er avhengig av hvor sikkert man kan styre noden og hvor presis posisjonen til seimikk-kablene bør være.

Tabell 5.1: \sqrt{MSE} -verdier ved prediksjon av posisjon til node på Tidsrekke 1 og 2, der T står for tidsrekke, S for steg.

T\S	1	2	3	4	5	6	7	8
1	0.4454	0.2800	0.2678	0.4997	0.5652	0.6205	0.7803	0.9294
2	0.3353	0.4632	0.5287	0.6748	0.8347	1.0276	1.2433	1.4541

Å forbedre prediksjonen av posisjon er mulig, dersom det tilpasses en modell som beskriver evolusjonen til noden bedre. Basert på dette studiet virker det naturlig å tilpasse en VARMA(p, q)-modell på flere noder, siden ARIMA-modellene predikerer posisjon best, når vi ser bort i fra at noden følger en ekstremt glatt kurve. En fordel med VARMA(p, q)-modellen er at den drar nytte av korrelasjonen mellom ulike noder på kableen. Dessuten er vi ute etter å predikere posisjonen til hele seimikk-kabler.

Dersom en VARMA(p, q)-modell skulle vise seg å være en akseptabel modell for seimikk-kabler i flere tilfeller, kan Kalmanfilter være riktig metode å bruke. Ved å transformere VARMA(p, q)-modellen på state-space form, kan modellen inngå som a priori modell i Kalmanfilteret, og forhåpentligvis styrke inferensen.

5.2 Estimat av akselerasjon

Et delmål med studiet var å studere hvor sikkert vi kan estimere akselerasjon sekvensielt, for å finne ut om et akselerometer vil styrke inferensen til posisjonen til noden. Kongsberg Seatex AS bekrefter i Kongsberg (2009) at støyen i deres akselerometer har et standardavvik på $\sigma_a = 0.0005m/s^2$. Klarte vi å estimere akselerasjon minst like sikkert? På akselerasjonsnivå predikerte BIC-modellene bedre enn state-space modellene på begge tidsrekkene. Likevel er det naturlig å sammenligne σ_a med den estimerte feilen i observasjonslikningene i state-space modellene (4.5) og (4.6). Dette fordi vi antar at feilen η_t , er feilen i målingen Y_t . I (4.5) tilpasset på Tidsrekke 1 er $\tau = 0.019$. Siden state-space modellen er tilpasset dobbeltdifferensierte data, er Y_t og tilhøren-

de standardavvik proporsjonale med akselerasjon. Vi må derfor dividere på $s^2 = 100$ for å sammenligne med akselerometeret. Dette gir $\tau_a = 0.0002$ og dermed er $\tau_a < \sigma_a$. Tilsvarende for Tidsrekke 2, får vi $\tau_a = 0.0005 = \sigma_a$ siden $\tau = 0.054$. Dette betyr at med en enkel state-space modell, som viste seg å ikke predikere best blant tidsrekke-modellene, estimerer akselerasjon minst like bra som et akselerometer måler den.

A

Tabeller

Tabellene i dette kapitlet inneholder data det blir referert til i oppgaven.

Tabell A.1: MSE-verdier for Parabelmetoden på Tidsrekke 1, i figur 3.1 (a). Første kolonne i tabellen viser MSE-verdi for estimatene der 10 punkt er brukt i Parabelmetoden. Videre inneholder kolonnene 2-7 MSE-verdi der 9,8,...3 punkt er brukt. Radene representerer hvor mange steg predikeres, første rad er 1-steg osv.

S \ P	3	4	5	6	7	8	9
1	0.0807	0.1346	0.2307	0.3533	0.4661	0.4894	0.4289
2	0.7204	0.9820	1.3368	1.7064	1.9477	1.8413	1.0649
3	3.1709	3.6466	4.1447	4.6090	4.6971	4.0757	1.9381
4	9.3292	9.5163	9.4459	9.4715	8.9038	7.2747	3.03028
5	20.7585	19.5171	17.8701	16.8916	15.2304	12.1073	4.6936
6	38.6367	34.2270	29.8270	27.3884	24.3786	19.2326	7.1696
7	65.2432	55.0318	46.3328	41.9344	37.1908	29.2512	10.6245
8	102.5721	83.7649	69.1278	62.3650	55.2498	43.244	15.6266
S \ P	10	11	12	13	14	15	16
1	0.3592	0.3145	0.2908	0.2792	0.2775	0.2803	0.2801
2	0.8387	0.7355	0.6792	0.6792	0.6371	0.5953	0.5405
3	1.4142	1.1935	1.0782	1.0782	0.9917	0.9183	0.8264
4	2.1114	1.7468	1.5623	1.5623	1.4295	1.3542	1.2737
5	3.1752	2.5756	2.2681	2.2681	2.0425	1.9320	1.8316
6	4.7501	3.7956	3.2881	3.2881	2.9083	2.7020	2.5183
7	6.9547	5.4909	4.6712	4.6712	4.0344	3.6447	3.3005
8	10.2734	8.0455	6.7295	6.7295	5.6945	5.0251	4.4180
S \ P	17	18	19	20	21	22	23
1	0.2755	0.2688	0.2584	0.2592	0.2694	0.2804	0.2824
2	0.4831	0.4369	0.3990	0.3863	0.3916	0.3996	0.3954
3	0.7270	0.6432	0.5761	0.5430	0.5379	0.5431	0.5384
4	1.1735	1.0741	0.9801	0.9166	0.8923	0.8941	0.8958
5	1.7006	1.5573	1.4165	1.3115	1.2576	1.2464	1.2438
6	2.3081	2.0962	1.9093	1.7772	1.7053	1.6772	1.6450
7	2.9527	2.6365	2.3836	2.2229	2.1395	2.1013	2.0391
8	3.8353	3.3459	2.9712	2.7480	2.6477	2.6192	2.5565
S \ P	24	25	26	27	28	29	30
1	0.2905	0.3090	0.3335	0.3600	0.3783	0.3944	0.3987
2	0.4030	0.4305	0.4676	0.5047	0.5234	0.5390	0.5390
3	0.5481	0.5785	0.6210	0.6696	0.7004	0.7261	0.7247
4	0.9129	0.9432	0.9825	1.0382	1.0866	1.1329	1.1365
5	1.2577	1.2778	1.3072	1.3665	1.4319	1.5030	1.5215
6	1.6306	1.6294	1.6499	1.7178	1.8009	1.8964	1.9280
7	2.0018	1.9911	2.0173	2.1055	2.2118	2.3393	2.3938
8	2.5326	2.5384	2.5914	2.7074	2.8228	2.9689	3.0347

Tabell A.2: MSE-verdier for Parabelmetoden på de siste 20 % av data i Tidsrekke 1, i figur 3.1 (a). Første kolonne i tabellen viser MSE-verdi for estimatene der 10 punkt er brukt i Parabelmetoden. Videre inneholder kolonnene 2-7 MSE-verdi der 9,8,...3 punkt er brukt. Radene representerer hvor mange steg predikeres, første rad er 1-steg osv.

S \ P	3	4	5	6	7	8	9
1	0.0702	0.1457	0.1764	0.1162	0.0207	0.0922	0.2647
2	0.6371	1.2411	1.4138	1.0179	0.3562	0.3585	0.7668
3	2.4506	4.7192	5.3170	3.9528	1.6104	1.0774	1.7337
4	5.5409	11.2132	12.6080	9.3844	3.8527	2.2298	3.2069
5	9.0564	20.8305	23.2500	16.9179	6.3759	3.1106	4.5439
6	13.5027	35.0387	38.7548	27.6693	9.5553	3.7515	5.7546
7	19.6741	55.0903	60.7395	42.8919	13.7968	4.2490	7.0304
8	28.9311	83.4719	92.2384	64.9094	20.510	15.9876	10.0860
S \ P	10	11	12	13	14	15	16
1	0.3355	0.2931	0.2208	0.1744	0.1610	0.1562	0.1654
2	1.0124	0.9923	0.8818	0.7796	0.6726	0.5579	0.4958
3	2.3197	2.4610	2.4194	2.3063	2.0417	1.6889	1.4402
4	4.2735	4.6182	4.6475	4.4935	3.9813	3.2830	2.7699
5	6.1054	6.4993	6.4608	6.1746	5.3678	4.3366	3.6136
6	7.8627	8.1867	7.9546	7.4925	6.4392	5.1443	4.2747
7	9.8882	10.1306	9.6878	9.1325	7.9559	6.4173	5.3651
8	14.2724	14.5943	13.9116	13.1088	11.4557	9.2762	7.7732
S \ P	17	18	19	20	21	22	23
1	0.1842	0.1980	0.1967	0.2082	0.2345	0.2476	0.2248
2	0.4734	0.4520	0.3916	0.3385	0.3270	0.3022	0.2321
3	0.4734	0.4520	0.3916	0.3385	0.7411	0.6288	0.4693
4	2.4528	2.2050	1.8909	1.5673	1.3752	1.1559	0.8921
5	3.2010	2.9019	2.5027	2.0801	1.8168	1.4913	1.1138
6	3.8119	3.4983	3.0263	2.5362	2.2220	1.7906	1.2855
7	4.8015	4.4329	3.8640	3.3275	2.9985	2.4601	1.7697
8	6.9428	6.3793	5.5349	4.7420	4.2578	3.5084	2.5516
S \ P	24	25	26	27	28	29	30
1	0.1985	0.1977	0.2089	0.2186	0.2139	0.1984	0.1947
2	0.1615	0.1328	0.1312	0.1315	0.1091	0.0785	0.0654
3	0.3197	0.2288	0.1856	0.1574	0.1104	0.0717	0.0608
4	0.6545	0.4961	0.4097	0.3531	0.2859	0.2497	0.2546
5	0.7827	0.5625	0.4435	0.3807	0.3255	0.3195	0.3683
6	0.8528	0.5684	0.4300	0.3910	0.3850	0.4427	0.5710
7	1.1712	0.7794	0.6044	0.5847	0.6088	0.6961	0.8711
8	1.7164	1.1615	0.90874	0.8649	0.8638	0.9392	1.1255

Tabell A.3: MSE verdier for Parabelmetoden på de siste 20 % av data i Tidsrekke 2, i figur 3.1 (b). Første kolonne i tabellen viser MSE-verdi for estimatene der 10 punkt er brukt i Parabelmetoden. Videre inneholder kolonnene 2-7 MSE-verdi der 9,8,...3 punkt er brukt. Radene representerer hvor mange steg predikeres, første rad er 1-steg osv.

S \ P	3	4	5	6	7	8	9
1	0.0414	0.0247	0.0201	0.0236	0.0449	0.0731	0.0931
2	0.2480	0.1545	0.1102	0.0732	0.0972	0.1525	0.2027
3	0.9523	0.6068	0.4248	0.2503	0.2237	0.2720	0.3448
4	2.5468	1.6400	1.1249	0.6544	0.5316	0.5288	0.6170
5	5.2488	3.2342	2.2585	1.4364	1.1976	1.1140	1.1974
6	9.3395	5.6514	3.9321	2.5832	2.2264	2.0912	2.2236
7	15.6865	9.3860	6.4808	4.3526	3.7864	3.5504	3.7190
8	24.8263	14.8342	10.4495	7.1362	6.2062	5.7312	5.8650
S \ P	10	11	12	13	14	15	16
1	0.1015	0.1124	0.1237	0.1393	0.1563	0.1836	0.2143
2	0.2237	0.2476	0.2760	0.3115	0.3508	0.4073	0.4713
3	0.3684	0.4037	0.4575	0.5219	0.5910	0.6868	0.7933
4	0.6191	0.6359	0.7005	0.7887	0.8887	1.0327	1.1905
5	1.1100	1.0822	1.1501	1.2560	1.3685	1.5522	1.7622
6	2.0347	1.9516	2.0096	2.1180	2.2306	2.4545	2.7225
7	3.3809	3.2429	3.3254	3.4763	3.6078	3.8822	4.2270
8	5.2610	4.9745	5.0529	5.2294	5.3520	5.6325	6.0149
S \ P	17	18	19	20	21	22	23
1	0.2411	0.2621	0.2770	0.2848	0.2954	0.3052	0.3171
2	0.5272	0.5677	0.5964	0.6100	0.6254	0.6380	0.6380
3	0.8869	0.9530	0.9967	1.0107	1.0270	1.0407	1.0616
4	1.3303	1.4233	1.4836	1.4934	1.5055	1.5163	1.5411
5	1.9510	2.0771	2.1565	2.1596	2.1640	2.1664	2.1990
6	2.9582	3.1108	3.2002	3.1778	3.1555	3.1328	3.1631
7	4.5241	4.7079	4.8100	4.7504	4.6907	4.6373	4.6686
8	6.3355	6.5100	6.6026	6.4821	6.3607	6.2612	6.2836
S \ P	24	25	26	27	28	29	30
1	0.3259	0.3384	0.3537	0.3746	0.4054	0.4398	0.4766
2	0.6635	0.6810	0.7032	0.7354	0.7875	0.8459	0.9087
3	1.0741	1.0983	1.1292	1.1770	1.2590	1.3509	1.4484
4	1.5548	1.5892	1.6318	1.6988	1.8149	1.9425	2.0765
5	2.2184	2.2641	2.3132	2.3875	2.5313	2.6886	2.8530
6	3.1776	3.2262	3.2725	3.3478	3.5266	3.7259	3.9366
7	4.6816	4.7329	4.7742	4.8478	5.0730	5.3263	5.5964
8	6.2836	6.2883	6.3415	6.3804	6.7413	7.0641	7.4140

Tabell A.4: MSE-verdier ved prediksjon av posisjon på Tidsrekke 1, plottet i figur 3.16. M står for modell, og S tidssteg.

M\S	1	2	3	4	5	6	7	8
BIC:5,7	0.1849	0.1469	0.3189	0.6596	1.0339	1.3680	1.5777	1.7912
AIC:6,7	0.1801	0.1406	0.2937	0.6353	1.0145	1.3091	1.4335	1.5527
Parabel28	0.2139	0.1091	0.1104	0.2859	0.3255	0.3850	0.6088	0.8638
Parabel29	0.1984	0.0785	0.0717	0.2497	0.3195	0.4427	0.6961	0.9392

Tabell A.5: MSE-verdier ved prediksjon av akselerasjon på Tidsrekke 1, plottet i figur 3.16. M står for modell, og S tidssteg.

M\S	1	2	3	4	5	6	7	8
BIC:5,7	0.0299	0.0055	0.0472	0.0297	0.0441	0.0509	0.01270	0.0790
AIC:6,7	0.0327	0.0063	0.0431	0.0336	0.0400	0.0557	0.0133	0.0782
Parabel28	0.0376	0.0109	0.1075	0.0460	0.0377	0.0982	0.0180	0.0723
Parabel29	0.0378	0.0110	0.1077	0.0461	0.0373	0.0979	0.0177	0.0725

Tabell A.6: MSE-verdier ved prediksjon av posisjoner på Tidsrekke 2, plottet i figur 3.17. M står for modell, og S tidssteg.

M\S	1	2	3	4	5	6	7	8
BIC:5,5	0.1601	0.2126	0.2795	0.4554	0.6968	1.0560	1.5460	2.1145
AIC:9,7	0.1613	0.2190	0.2851	0.4730	0.7492	1.1674	1.7074	2.3299
Parabel11	0.1124	0.2476	0.4037	0.6359	1.0822	1.9516	3.2429	4.9745
Parabel30	0.4766	0.9087	1.4484	2.0765	2.8530	3.9366	5.5964	7.4140

Tabell A.7: MSE-verdier ved prediksjon av akselerasjon på Tidsrekke 2, plottet i figur 3.17. M står for modell, og S tidssteg.

M\S	1	2	3	4	5	6	7	8
BIC:5,5	0.0119	0.0164	0.0339	0.0211	0.0164	0.0235	0.0102	0.0183
AIC:9,7	0.0115	0.0164	0.0350	0.0220	0.0157	0.0250	0.0103	0.0197
Parabel11	0.0142	0.0196	0.0478	0.0263	0.0173	0.0238	0.0109	0.0262
Parabel30	0.0130	0.0169	0.0453	0.0242	0.0200	0.0200	0.0120	0.0209

Tabell A.8: Parameterverdier til state-space modell estimert på Tidsrekke 1, der $T=900$ tilsier at modellparametrene er tilpasset $X_{1:900}$ osv.

T\S	ϕ	$\log(\sigma)$	τ	x_0
900	0.9997	-0.1286	0.0158	-43.1698
930	0.9998	-0.1412	0.0157	-52.6499
960	0.9997	-0.1545	0.0156	-48.8630
990	0.9997	-0.1655	0.0155	-40.3801
1020	0.9997	-0.1771	0.0154	-29.8301
1050	0.9997	-0.1868	0.0153	-16.4702

Tabell A.9: Parameterverdier til state-space modell estimert på dobbeltdifferensierte data i Tidsrekke 1, der $T=900$ tilsier at modellparametrene er tilpasset $\Delta^2 X_{1:900}$ osv.

T\S	ϕ	$\log(\sigma)$	τ	x_0
900	0.2916	-1.4905	0.0194	-5.11e-07
930	0.2928	-1.4969	0.0193	-9.34e-07
960	0.2950	-1.4916	0.0191	-8.08e-07
990	0.3057	-1.4831	0.0187	-1.11e-06
1020	0.3038	-0.1771	0.0187	-8.67e-08
1050	0.2981	-1.4927	0.0191	-4.58e-08

Tabell A.10: Parameterverdier til state-space modell estimert på Tidsrekke 2, der $T=2100$ tilsier at modellparametrene er tilpasset $X_{1:2100}$ osv.

T\S	ϕ	$\log(\sigma)$	τ	x_0
2100	0.9999	-0.6763	0.0110	48.0001
2130	0.9999	-0.6822	0.0110	44.5799
2160	0.9999	-0.6846	0.0110	33.9802
2190	0.9999	-0.6895	0.0109	34.5799
2220	0.9999	-0.6953	0.0109	34.3399
2250	0.9999	-0.6985	0.0109	24.4001
2280	0.9999	-0.7022	0.01088	17.3701
2310	0.9999	-0.7053	0.0108	7.9401
2340	0.9999	-0.7084	0.0108	-0.8896
2370	0.9999	-0.7010	0.0108	-22.1398
2400	0.9999	-0.6994	0.0109	-37.4801
2430	0.9999	-0.7053	0.0108	-36.3498
2460	0.9999	-0.7110	0.0108	-36.2301
2490	0.9999	-0.7160	0.0107	-37.5499
2520	0.9999	-0.7218	0.0107	-36.4301
2550	0.9999	-0.7268	0.0107	-31.9400
2580	0.9999	-0.7319	0.0106	-27.4600
2610	0.9999	-0.7367	0.0106	-23.2401
2640	0.9999	-0.7336	0.01068	-5.8903

Tabell A.11: Parameterverdier til state-space modell estimert på dobbeltdifferensierte data i Tidsrekke 2, der $T=2100$ tilsier at modellparametrene er tilpasset $\Delta^2 X_{1:2100}$ osv.

T\S	ϕ	$\log(\sigma)$	τ	x_0
2100	-0.0757	-1.9723	0.0541	-9.38e-06
2130	-0.0765	-1.9716	0.0534	2.21e-05
2160	-0.0808	-1.9677	0.0519	1.18e-05
2190	-0.0822	-1.9684	0.0516	4.31e-06
2220	-0.0842	-1.9668	0.0512	1.47e-05
2250	-0.0836	-1.9679	0.0514	9.64e-06
2280	-0.0852	-1.9679	0.0511	6.28e-06
2310	-0.0846	-1.9687	0.0513	5.21e-06
2340	-0.0826	-1.9740	0.0521	-1.36e-05
2370	-0.0805	-1.9763	0.0528	8.21e-06
2400	-0.0798	-1.9805	0.0533	2.04e-05
2430	-0.0805	-1.9821	0.0528	-1.66e-05
2460	-0.0814	-1.9818	0.0522	8.25e-06
2490	-0.0824	-1.9846	0.0518	4.48e-06
2520	-0.0825	-1.9883	0.0515	-8.42e-06
2550	-0.0861	-1.9881	0.0505	2.25e-05
2580	-0.0869	-1.9901	0.0500	-3.74e-06
2610	-0.0880	-1.9895	0.0496	7.27e-06
2640	-0.0919	-1.9859	0.0482	-1.29e-07

Tabell A.12: MSE-verdier ved prediksjon av posisjon og akselerasjon med KF, gitt en-dimensjonale state-space modeller:

Tidsrekke\Steg	0	1	2	3	4
AKS1	0.0003	0.0449	0.0477	0.0493	0.0498
AKS2	0.0013	0.0167	0.0170	0.0170	0.0170
POS1	5.74e-05	1.96e-01	1.94e-01	1.92e-01	1.90e-01
POS2	5.61e-05	1.29e-01	1.29e-01	3.74e-01	1.28e-01
Tidsrekke\Steg	5	6	7	8	
AKS1	0.0499	0.0500	0.0500	0.0500	
AKS2	0.0170	0.0170	0.0170	0.0170	
POS1	1.88e-01	1.87e-01	1.85e-01	1.84e-01	
POS2	1.27e-01	1.27e-01	1.27e-01	1.26e-01	

Litteratur

- Berger R. L C.G. (2002). *Statistical Inference*. Duxbury. Referert til på side 5.
- Brockwell P.J. og Davis R.A. (2002). *Introduction to Time Series and Forecasting*. QA280.B757. Springer, andre utgave. Referert til på side 6.
- Christopher K. Wikle L.B. (2006). «A bayesian tutorial for data assimilation». *Rend. Circ. Mat. Palermo*, side 2–8. Referert til på side 6.
- Evensen G. (1996). «Sequential data assimilation with nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics». *J. Geophys. Res.* 99, side 26. Referert til på side 6.
- Gelman A. (2004). *Bayesian Data Analysis*. World Student Series. Chapman and Hall/Crc, andre utgave. Referert til på side 6.
- Grindheim J.V. (2008). «Autokorrelasjon i tidsdomenet». side 1–2. Referert til på side 36.
- J. Meinhold N.S. (). «Understanding the kalman filter». *The American Statistician*, Vol. 37, No. 2, p. 123-127. Referert til på side 6.
- Kongsberg (2009). *MRU 5+: The Ullimate Marine Motion Sensor*. Kongsberg Seatex AS. Referert til på side 87.
- Mandel J. (2006). «Efficient implmentation of the ensemble kalmanfilter». side 1–6. Referert til på side 6.
- N.A.Thacker og A.J.Lacey (2006). «The likelihood interpretation of the kalman filter». side 4–5. Referert til på side 77.
- Pedersen E. (2000). «Drag model for towed seismic cables». side 1–5. Referert til på side 3.

- Pedersen E. og Sørensen A.J. (2001). «Modelling and control of towed marine seismic streamer cables». side 1–6. Referert til på side 3.
- Petersen K.B. (2008). «Matrix cookbook». side 17–18. Referert til på side 6 og 20.
- Petersone V. (2009). *Estimering av Stokastisk Volatilitet*. Masteroppgave, University of Bergen. Referert til på side 34.
- Richard C. Aster B.B. og Thurber C.H. (). *Parameter Estimation and Inverse Problems*. International Geophysics. Første utgave. Referert til på side 6.
- Skjervheim J.A. (2007). «Continuous updating of a coupled reservoir-seismic model using an ensemble kalman filter technique». side 1–48. Referert til på side 6.
- Thulin K. (2005). *Ensemble Kalman Filter for Field Estimation*. Masteroppgave. Referert til på side 6.
- Turkylmaz Y. (2004). «Modeling and control of towed seismic cables». *Doktorgrad ved NTNU*. Referert til på side 3 og 70.
- Wikle C.K. (2005). «Statistics for model/data fusion: An introduction». *Rend. Circ. Mat. Palermo*, side 26. Referert til på side 6 og 12.