

Mat-2.4177 Operaatiotutkimuksen projektityöseminaari

# Jokisysteemin vesivoimatuotannon simulointi

Loppuraportti

26. toukokuuta 2014

Vili Ojala (Projektipäällikkö)

Viivi Halla-aho

Sampo Kaukonen

Jukka Koskenranta

# Sisältö

Symbolit ja lyhenteet	iii
<b>1 Johdanto</b>	<b>1</b>
<b>2 Aiempi tutkimus</b>	<b>2</b>
2.1 Fysikaaliset mallit . . . . .	2
2.2 Tilastolliset ja muut mallit . . . . .	4
2.3 Aiemman tutkimuksen hyödyntäminen . . . . .	5
<b>3 Menetelmät</b>	<b>7</b>
3.1 Aikasarja-analyysi . . . . .	7
3.2 Mallin rakentamisen vaiheet . . . . .	8
3.2.1 Aineiston esikäsittely . . . . .	8
3.2.2 Mallin rakenneparametrien valinta ja muuttujien ker- toimien estimointi . . . . .	9
3.2.3 Mallin validointi . . . . .	10
3.2.4 Käytännön toteutus . . . . .	10
<b>4 Aineisto</b>	<b>13</b>
4.1 Aineiston esikäsittely . . . . .	14
4.2 Mallin rakenteen valinta . . . . .	14
<b>5 Tulokset</b>	<b>15</b>
5.1 Tasainen profiiliajo . . . . .	15
5.1.1 Vesivoimalan B juoksutusten ennustaminen . . . . .	15
5.1.2 Vesivoimalan C pinnankorkeuden ennustaminen vuo- sien 2010 ja 2013 aineistoilla . . . . .	17
5.1.3 Vesivoimalan C pinnankorkeuden ennustaminen vuo- den 2009 aineistolla . . . . .	19
5.1.4 Vesivoimalan D pinnankorkeuden ennustaminen . . . . .	20
5.2 Jääkansi . . . . .	21
5.2.1 Vesivoimalan C pinnankorkeuden ennustaminen . . . . .	21
5.2.2 Vesivoimalan D pinnankorkeuden ennustaminen . . . . .	24
5.3 Niukka vesitilanne . . . . .	25
5.3.1 Vesivoimalan C pinnankorkeuden ennustaminen . . . . .	25
5.3.2 Vesivoimalan D pinnankorkeuden ennustaminen . . . . .	28
<b>6 Yhteenveto</b>	<b>30</b>
<b>Viitteet</b>	<b>32</b>

A Tasainen profiaajo: vesivoimalan B juoksutusten ennustaminen	35
B Tasainen profiaajo: vesivoimalan C pinnankorkeuden ennustaminen vuosien 2010 ja 2013 aineistoilla	39
C Tasainen profiaajo: vesivoimalan C pinnankorkeuden ennustaminen vuoden 2009 aineistolla	42
D Tasainen profiaajo: vesivoimalan D pinnankorkeuden ennustaminen	45
E Jääkansi: vesivoimalan C pinnankorkeuden ennustaminen	47
F Jääkansi: vesivoimalan D pinnankorkeuden ennustaminen	52
G Niukka vesitilanne: vesivoimalan C pinnankorkeuden ennustaminen	56
H Niukka vesitilanne: vesivoimalan D pinnankorkeuden ennustaminen	60
I Itsearvio	63

# Symbolit ja lyhenteet

## Symbolit

$A$	joen poikkileikkauksen pinta-ala
$x$	paikka mitattuna joen yläjuoksulta
$t$	aika
$v$	virtauksen keskinopeus
$b$	uoman yläosan leveys
$h$	virtauksen korkeus
$g$	putoamissyvyys
$i$	joenpohjan nousukulman sini
$j$	energiahäviö, joka syntyy kun yksiköllinen nestettä liikkuu matkan pituusyksikön verran
$y(t)$	ulostulo hetkellä $t$
$u(t)$	sisäänmeno hetkellä $t$
$e(t)$	valkoinen kohina hetkellä $t$
$n_a$	napojen lukumäärä
$n_b$	nollien lukumäärä +1
$n_k$	kuollut aika
$n_c$	kertoimien lukumäärä

## Lyhenteet

HEC-RAS	Hydrologic Engineering Centers River Analysis System
ARX	Autoregressive Model with external input
ARMAX	Autoregressive Moving Average with external input Model
SARX	Seasonal AR-model with external input
SARMAX	Seasonal ARMA with external input, kausivaihtelu-ARMA-malli ulkoisella selittäjällä
AIC	Akaike Information Criterion
MDL	Minimum Description Length

# 1 Johdanto

Tämä työ tehtiin keväällä 2014 Aalto-yliopiston perustieteiden korkeakoulun kurssilla Operaatiotutkimuksen projektityöseminaari. Tehtävän asettaja oli UPM ja aiheena ”Jokisysteemin vesituotannon simulointi ja optimointi”. Projektin kohteena oli jokisysteemi, missä UPM:llä oli useita vesivoimalaitoksia ja UPM:n tavoitteena oli niiden tuotannon optimointi.

Sähkömarkkinoilla hinnat vaihtelevat nopeasti ja on tärkeää saada tuotettua mahdollisimman paljon sähköä, kun hinta on korkealla. Veden kulkeminen koko jokisysteemin läpi vie useita tunteja ja on toivottavaa, että vettä olisi mahdollisimman paljon käytössä, kun sähkön hinta on korkea. Kyseisestä jokisysteemistä tekee ongelmallisen se, että vettä ei voi juurikaan varastoida laitosten välille.

Työn tavoitteena oli mallintaa veden liikkeitä jokisysteemeissä. Tavoitteena oli veden pinnankorkeuksien ennustaminen tietyissä pisteissä, kun tunnetaan aikaisemmat pinnankorkeudet ja juoksutukset. Pinnankorkeuksille ja juoksutuksille on tietyt rajat, mitkä asettavat vaatimuksia ennustamisen tarkkuudelle. Tällöin veden liikkeet täytyy tuntea tarkkaan, jotta juoksutuksia voi optimoida ilman riskiä pinnankorkeuksien tai virtauksien rajojen ylittämisestä.

Matemaattisia malleja veden virtauksen ennustamista varten pyydettiin luomaan erityisesti kolmeen tyypilliseen tapaukseen, joissa malleista tulisi todennäköisesti olemaan hyötyä. Näiksi tapauksiksi valikoituivat niukka vesitilanne joessa, virtaus jääkannen alla ja tasainen profiliajo, kun vettä on runsaasti. Mallinnuksessa käytetään ARX-aikasarjamalleja.

Raportissa esitellään aiempaa tutkimusta kirjallisuuskatsauksen muodossa, käydään läpi käytetyt menetelmät ja esitellään saadut tulokset. Lopuksi pohditaan saatuja tuloksia ja käytettyjen menetelmien sopivuutta kyseessä olevaan tehtävään. Lisäksi kerrotaan, mitkä olivat mallinnuksen haasteet ja annetaan suosituksia tehtävän asettajalle jatkotoimenpiteitä ajatellen. Liitteenä on myös itsearvio projektista.

## 2 Aiempi tutkimus

Tarkoituksena on tuottaa matemaattinen malli joen virtaaman ennustamiseksi. Millaisia menetelmiä samankaltaisten ongelmien ratkaisuun on käytetty aiemmissä tutkimuksissa? Mahdolliset lähestymistavat voidaan jakaa kahteen eri joukkoon: fysikaalisiin ja tilastollisiin malleihin.

Aiemmissä tutkimuksissa mallinnettava alue on ollut usein hieman erilainen kuin meidän kohteenamme oleva vesistö. Luonnonjokimalleihin on kuulunut usein pato tai muu veden varastointiin soveltuva allas ja toisinaan malleja on tehty myös keinotekoisille vesiväylille. Lisäksi toisinaan huomioon on otettu myös vuorovesi ja veden varastoituminen ympäristöön.[21][22] Mallinnettavat vesistöt eroavat siis usein suuresti omasta mallinnuskohteestamme. Myös hyvin samankaltaisia tapauksia löytyy, kuten kahden vesivoimalan välisen jokiosuuden mallinnus Sohlbergin ja Sernfältin artikkelissa.[15]

Seuraavaksi esitellään fysikaalisia, tilastollisia ja muita malleja tarkemmin ja arvioidaan valmiiden mallinnustapojen sopivuutta meidän ongelmaamme.

### 2.1 Fysikaaliset mallit

Fysikaalisissa malleissa käytetään hyväksi veden virtausta kuvaavia differentiaaliyhtälöitä. Näiden mallien on havaittu olevan hyvin paikkansapitäviä ja luotettavia, mutta samalla monimutkaisia, laskennallisesti haastavia ja suuria datamääriä vaativia. Tarkan fysikaalisen mallin toteutus vaatii tarkat tiedot joen geometriasta.[2] Fysikaalisia malleja tarvitaan virtausten arviointiin erityisesti silloin kun joen rakenteeseen suunnitellaan merkittäviä muutoksia kuten uusien siltojen tai patojen rakentamista.[16]

Eräs fysikaalinen tapa mallintaa virtauksia on Saint Venant -differentiaaliyhtälöitä käyttämällä. Yhtälöryhmä koostuu kahdesta yhtälöstä, joista toinen on jatkuvuusyhtälö

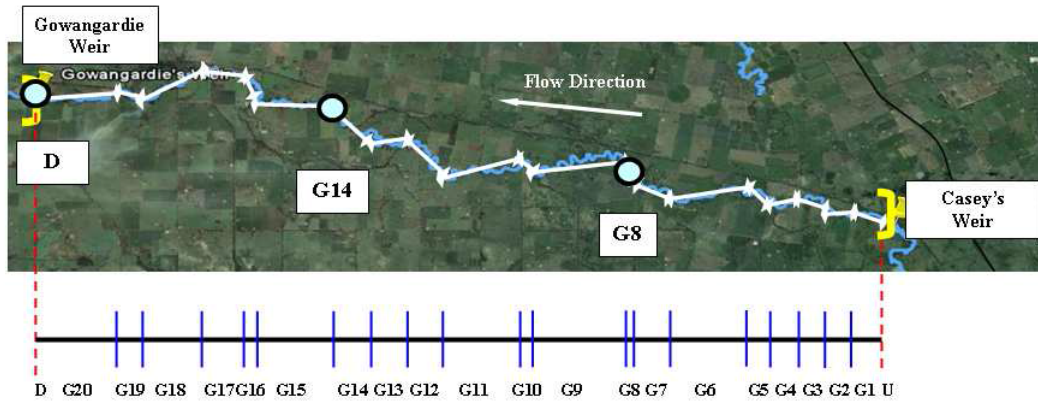
$$v \frac{\partial A}{\partial x} + A \frac{\partial v}{\partial x} + b \frac{\partial h}{\partial t} = 0 \quad (1)$$

ja toinen momenttiyhtälö

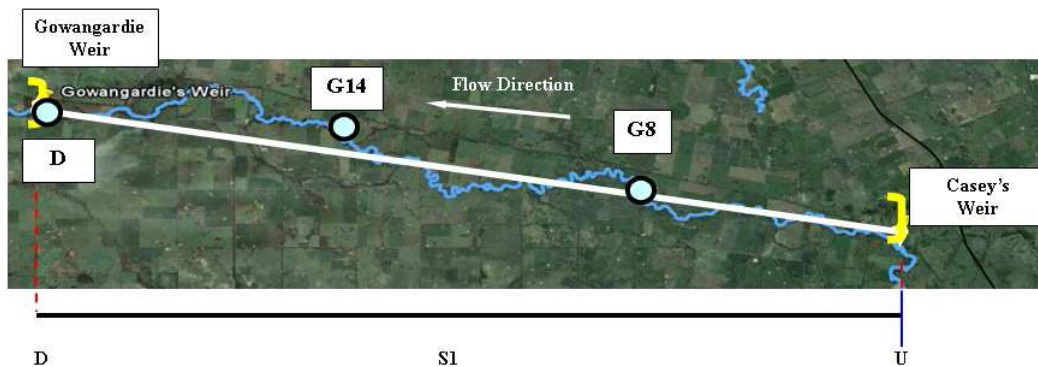
$$g \frac{\partial h}{\partial x} + v \frac{\partial v}{\partial x} + \frac{\partial v}{\partial t} = g(i - j), \quad (2)$$

joissa  $A$  on joenpätjän poikkileikkauksen pinta-ala,  $v$  virtauksen keskinopeus,  $x$  paikka mitattuna yläjuoksulta,  $t$  aika,  $b$  uomien yläosan leveys,  $h$  virtaaman syvyys ja  $g$  putoamiskiihtyvyys.  $i$  on joenpohjan nousukulman

sini ja  $j$  energiahäviö per joenpätjän pituuden yksikkö per nesteen painoyksikkö. Tässä yhtälöt on esitetty yksiulotteisessa tapauksessa. St. Venantin yhtälöissä veden ajatellaan virtaavan ainoastaan joen uoman suuntaisesti.[3]



Kuva 1: Australialaisen Broken Riverin virtausta mallinnettiin St. Venant-yhtälöiden avulla, ja joen geometrian oletettiin olevan vakio kullakin kuvaan merkityllä välillä. Kuvan lähde [3].



Kuva 2: Sama joki kuin kuvassa 1, mutta nyt joen geometriaa approksimoitiin pelkästään yhtenä geometrialtaan vakiona välinä. Kuvan lähde [3].

1D yhtälöillä voidaan mallintaa yksinkertaisten jokien virtaamia melko tarkasti. Esimerkiksi keinotekoisissa kanavissa vesi virtaa usein melko suora-  
viivaisesti, jolloin approksimaatio ei aiheuta suurta virhettä. Luonnonjoissa virtaukset ovat kuitenkin usein selvästi monimutkaisempia, minkä vuoksi fy-  
sikaalinen mallinnus approksimoitujen Saint Venantin yhtälöiden avulla on-  
nistuu paremmin keinotekoisille vesiväylille kuin luonnonvesistöille.[22] Esi-  
merkiksi mutkaisissa joissa vesi virtaa  $xy$ -tasossa, jolloin 1D approksimaatio

ei ole kovin tarkka.[3] Lisäksi esimerkiksi pohjan karkeus voi aiheuttaa merkittäviä virtauksia myös syvyysuunnassa, jolloin edes 2D-yhtälöt eivät välttämättä ole riittävän tarkkoja. Usein kuitenkin 1D-approksimaation avulla saadaan ainakin oikean suuntaisia virtausprofileja.

1D St. Venant yhtälöiden avulla on mallinnettu esimerkiksi australialaisen Broken Riverin virtausta.[3] Kyseisessä tutkimuksessa testattiin ennusteen hyvyttä erilaisilla approksimaatioilla joen geometrialle, joka on esitetty kuvassa 1. Joki jaettiin tiettyyn määrään segmenttejä, ja joen geometrian oletettiin olevan vakio kullakin välillä. Kyseisessä tapauksessa havaittiin, että alajuoksun pinnankorkeusarvio oli suunnilleen yhtä hyvä riippumatta siitä, mallinnettiinko koko tarkasteltavaa väliä useana segmenttinä kuten kuvassa 1, vai ainoastaan yhtenä homogeenisena segmenttinä kuten kuvassa 2.

Toinen laki, jota veden virtauksen mallintamisessa usein käytetään, on niin sanottu mass balance -yhtälö eli massan säilymistä kuvaava yhtälö. Se kuvaa systeemiin tulevaa ja systeemistä lähtevää massaa, tässä tapauksessa vettä. Periaate huolehtii siitä, ettei vesimassaa häviä systeemistä olemattomiin.[19] Massatasapainoyhtälöt eivät kuitenkaan aina pysty kuvaamaan kaikkia joki-systeemien fysikaalisia ilmiöitä. Esimerkiksi veden haihtumista, paluuvirtaamia ja säätelemättömiä tulovirtaamia on vaikea sisällyttää massatasemalleihin.[9]

St. Venantin differentiaaliyhtälöryhmää ei voida eksplisiittisesti ratkaista ilman merkittäviä oletuksia, jotka eivät päde useimmissa tapauksissa. Tämän vuoksi ratkaisut joudutaan etsimään numeerisesti.[14] Kun fysikaalisia malleja sovelletaan, joudutaan siis usein tekemään merkittäviä approksimaatioita ongelman ratkaisemiseksi kohtuullisessa ajassa.

## 2.2 Tilastolliset ja muut mallit

Tilastolliset mallit perustuvat fysikaalisten parametrien arvoista luotuihin aikasarjoihin, joiden tilastollisia ominaisuuksia tutkitaan. Tilastolliset mallit ovat matemaattisesti vähemmän haastavia kuin fysikaaliset, mutta ne eivät ole niin tarkkoja.[2] Toisinaan myös aikasarjoihinkin perustuvat mallit toteuttavat osan fysikaalisista laeista, kuten Saint Venantin yhtälöt. Näin on usein nk. grey box -mallinnuksessa, jossa todellista mallia ei tunneta kuin joiltain osin, mutta jotain tietämystä mallin rakenteesta on. Black box -mallinnuksessa prosessia taas ei tunneta.

Artikkelissaan Rivas-Perez et al. esittelevät veden virtausta keinokastelujärjestelmässä kuvaavan ARMAX-aikasarjamallin. Malli ennustaa vedenpinnan



korkeutta järjestelmän eri osissa. Mallin toimivuus perustuu robustiin, rekursiiviseen parametrien estimointiin pienimmän neliösumman (PNS) menetelmällä. Parametrit estimoidaan siis uudestaan kun uutta dataa on saatavilla. Robustisuus paikkaa epävarmuutta, kun parametrit varioivat esimerkiksi aineiston huonon laadun vuoksi.[13]

Campolo et al:n tutkimuksessa on taas käytetty neuroverkkoihin perustuvaa tilastollista menetelmää joen virtauksen ennustamiseen enintään kuuden päivän päähän. Menetelmän etuja ovat yksinkertaisuus ja melko suuri tarkkuus. Menetelmässä sisäänmenomuuttujista saadaan laskettua epälineaaristen yhteyksien kautta ulostulomuuttujat. Kyseisessä tapauksessa virtausta selittäviksi muuttujiksi oli valittu vesivarannoista laskeva vesi, sateet ja senhetkinen vesistön tilanne.[2]

Muitakin mahdollisuuksia toki on kuin tukeutua pelkästään tilastolliseen tai fysikaaliseen malliin. Esimerkiksi Sohlbergin ja Sernfältin malli pohjautui joen virtauksen fysikaalisiin ominaisuuksiin, manuaalisen säädön tuomaan kokemukseen systeemin riippuvuussuhteista ja aineistona käytetyistä vedenpinnan korkeuden mittauksista ja virtausestimaateista saatuun näkemykseen. Monenlaista tietoa on siis yhdistetty mallin luomiseksi.[15]

Usein virtausta kuvaisi parhaiten epälineaarinen malli. Mikäli kuitenkin halutaan saada aikaan lineaarinen malli, virtaaman epälineaarisuuksia voidaan arvioida monin tavoin. Bolea et al. yrittivät löytää hyvin toimivan lineaarisen ohjauksen yksinkertaiselle kanaalisysteemille. He jakoivat systeemin altaan pinnankorkeuksien perusteella eri toimintatiloihin. Mallin parametrit estimointiin erikseen kullekin toimintatilalle ja lopulliset kertoimet interpolointiin polynomifunktioiden avulla.[1]

## 2.3 Aiemman tutkimuksen hyödyntäminen

Matemaattisen mallin luominen vaatii paljon tietämystä aiheesta, mitä löytyy usein vain alan ammattilaisilta. Projektin aikana on toki saatu apua UPM:n työntekijöiltä, jolta tätä ammattitaitoa löytyy. Aiemminkin on luotu malleja vedenpinnan korkeuksia ja estimoituja veden virtauksia vesivoimalan läpi käyttäen, joten sen ei pitäisi olla mahdotonta meillekään. Kuten sanottua, lähtökohdat meidän ja muiden mallinnuskohteiden välillä ovat usein erilaiset. Periaatteita ja ideoita voidaan kyllä soveltaa meidän tapauksemme, mutta valmista ratkaisua ei aikaisemmista tutkimuksista tulla saamaan.

Fysikaaliset mallit vaatisivat niin paljon tietoa vesistöstä, että meillä ei ole mahdollisuuksia saada kaikkea käyttöömmme. Täytyisi tuntea joen geometriaa

ja tuntea tai estimoida sen fysikaalisia ominaisuuksia kuvaavia parametreja, kuten pohjan kitkakerroin. Lisäksi ei ole syytä tähdätä tarpeettoman monimutkaiseen malliin, jos yksinkertaisempikin tekee tehtävänsä.

Fysikaaliset mallit ovat epälineaarisia ja niiden ratkaisemiseksi tarvitaan merkittäviä approksimaatioita. Systemin identifiointi on helppo tapa mallin luomiseksi, kun kunnollinen fysikaalinen malli osoittautuu liian monimutkaiseksi toteuttaa.[13] Esimerkiksi tilastollisilla menetelmillä, kuten aikasarjamallilla, saadaan helposti aikaan toivottu lineaarinen malli. Näiden seikkojen vuoksi fysikaalisen mallin muodostamista ei tässä tapauksessa pidetä hyvänä vaihtoehtona.

Voisiko sitten aikasarjamallin laatua tai käytettävyyttä parantaa? R. Rivas Perez et al.:n artikkelissa käytettiin rekursiivista, robustia menetelmää parametrien estimoinnissa. Parametrien estimoinnin robustisuuden lisääminen on hyvä idea, jos esimerkiksi aineistossa esiintyvät virheet ovat uhka hyvän mallin löytämiselle. Yksi esimerkki robustisuuden lisäämisen keinoista on aineiston normalisointi. Robustisuudesta on hyötyä erityisesti silloin, kun on approksimoitu epälineaarista prosessia lineaarisella mallilla. Robustisuus voisi tehdä malleistamme parempia.[13]

Mallimme käytettävyyttä taas voitaisiin ehkä lisätä algoritmin rekursiivisuudella. Rekursiivisessa parametrien estimoinnissa voidaan luoda uudet parametrien arvot kun uutta dataa on saatavilla. Tarkan mallin aikaansaaminen on tärkeää reaaliaikaista päätöksentekoa ajatellen. Tästä voisi olla apua vesivoimaloiden optimaalisten juoksutustenkin suunnittelussa. Virtausta mallinnettaessa olosuhteet muuttuvat jatkuvasti, joten päivittyvä malli voisi kuvata tilannetta paremmin kuin yksi malli, joka sopii jotenkuten kaikkiin tilanteisiin ja vuodenaikoihin.[13]

Lisäksi jatkossa samanlaista tilannetta tutkittaessa kannattaa huomioda myös mallin luomiseen käytettävä aineisto. Ominaisuuksiltaan erilaiset osat mallinnettavassa alueessa on helpompi ottaa huomioon, jos mittauspisteet on sijoitettu vesistön erilaisiin kohtiin.

### 3 Menetelmät

Tässä osiossa esitellään mallinnuksessa käytetyt menetelmät. Matemaattinen malli luotiin aikasarja-analyysin keinoin ja sen peruseriaatteet ja käytetyt aikasarjamallit on esitelty lyhyesti ensin. Tämän jälkeen käydään läpi peruseriaatteet itse mallin rakentamisen vaiheista aineiston käsittelystä mallin validointiin.

#### 3.1 Aikasarja-analyysi

Aikasarjaksi kutsutaan jostain havaintoyksiköstä tehdyistä peräkkäisistä havainnoista muodostettua aineistoa. Aikasarja-analyysissa tavoitteena on tutkia aikasarjoja ja luoda malli, joka kuvaa aineiston taustalla olevaa satunnaisprosessia. Mallin avulla voidaan esimerkiksi ennustaa prosessin tulevia arvoja.[17] Erilaisia aikasarjamalleja on olemassa useita. Tässä työssä käytettiin pääasiassa ARX-aikasarjamalleja, joiden rakenne esitellään tarkemmin seuraavaksi.

ARX eli autoregressiivinen malli ulkoisella selittäjällä malli on muotoa

$$\begin{aligned} y(t) + a_1y(t-1) + a_2y(t-2) + \dots + a_{n_a}y(t-n_a) \\ = b_1u(t-n_k) + \dots + b_{n_b}u(t-n_b-n_k+1) + e(t), \end{aligned} \quad (3)$$

jossa  $y(t)$  on ulostulo hetkellä  $t$ ,  $u(t)$  sisäänmeno hetkellä  $t$ ,  $n_a$  napojen lukumäärä,  $n_b$  nollien lukumäärä  $+1$ ,  $n_k$  niinsanottu kuollut aika, eli niiden sisäänmenojen lukumäärä, jotka ehtivät tapahtua ennen kuin sisäänmeno vaikuttaa ulostuloon,  $y(t-1), \dots, y(t-n_a)$  edelliset ulostulot,  $u(t-n_k), \dots, u(t-n_b-n_k+1)$  edelliset sisäänmenot, joista ulostulon nykyinen arvo riippuu ja  $e(t)$  valkoinen kohina hetkellä  $t$ .

Usealle sisäänmenolle malli voidaan kirjoittaa seuraavasti

$$\begin{aligned} y(t) + A_1y(t-1) + A_2y(t-2) + \dots + A_{n_a}y(t-n_a) \\ = B_0u(t-n_k) + B_1u(t-1) + \dots + B_{n_b}u(t-n_k-n_b) + e(t), \end{aligned} \quad (4)$$

jossa  $n_b$  ja  $n_k$  ovat vektoreita, joiden  $i$ :s elementti kuvaa viivettä, joka liittyy sisäänmenoon  $i$ . [5]

Autoregressiivisessä mallissa ulkoisilla selittävillä tekijöillä eli ARX-mallissa ulostuloa hetkellä  $t$  selitetään ulostulon aiemmilla arvoilla, ulkoisten selittäjien nykyisillä ja edellisillä arvoilla sekä valkoista kohinaa olevan satunnaisuuttujan arvolla  $e(t)$ . Kukin selittäjä saa eteensä kertoimen, jolla sen arvoa

kerrotaan. Se, kuinka suurilla viiveillä eri selittäjät halutaan malliin mukaan voidaan valita.[11]

Merkintätapa  $ARX(n_a, [n_{b1}n_{b2}], [n_{k1}n_{k2}])$  tarkoittaa kahden ulkoisen selittäjän ARX-mallia, jossa  $n_a$  on sen napojen lukumäärä,  $n_{b1}$  on nollien lukumäärä ulkoiselle selittäjälle 1 ja  $n_{b2}$  sama ulkoiselle selittäjälle 2 ja  $n_{k1}$  sekä  $n_{k2}$  ovat kuolleet ajat ulkoisille selittäjille 1 ja 2. Jatkossa parametrien suuruudella viitataan  $n_a$ ,  $n_b$  ja  $n_k$  suuruuteen. Kun puhutaan mallin muuttujien kertoimista, tarkoitetaan viitataan estimoituihin kertoimiin  $a_1, \dots, a_{n_a}$  ja  $b_1, \dots, b_{n_b}$ . [5]

Muita malleja, jotka voisivat sopia kuvaamaan ainakin osaa tilanteista ovat ARMAX, SARX ja SARMAX. ARMAX:illa tarkoitetaan autoregressiivistä liikkuvan keskiarvon mallia. SARX eli Seasonal ARX ja SARMAX eli Seasonal ARMAX, ovat muuten samanlaisia kuin ARX ja ARMAX, mutta niihin on lisätty kausivaihteluelementti.

## 3.2 Mallin rakentamisen vaiheet

Esitellään seuraavaksi lyhyesti aikasarjamallin muodostamisen vaiheet eli aineiston esikäsittely, mallin rakenteen valinta, mallin parametrien estimointi ja mallin validointi.

### 3.2.1 Aineiston esikäsittely

Ennen varsinaisen mallin rakentamista, on syytä tutustua käytössä olevaan aikasarja-aineistoon ja tehdä sille tarvittava esikäsittely.

Aineistoon voi tutustua esimerkiksi tekemällä siitä kuvaajia ja tutkimalla sitä silmämääräisesti ja laskemalla erilaisia tilastollisia tunnuslukuja aineistolle. Jo tässä vaiheessa on mahdollista tehdä huomiota prosessin luonteesta ja huomata esimerkiksi mahdolliset poikkeavat havainnot tai erilaiset ajanjaksot aineistossa. Graafien avulla voi tutkia aikasarjan ominaisuuksia, kuten onko sillä trendiä tai kausivaihtelua.[10] Aineistoon ja ilmiöön tutustumalla voi saada vihiä siitä, millainen malli voisi kuvata ilmiötä parhaiten esimerkiksi aikasarjamallin viivetermien osalta.

Aineiston esikäsittelyssä on tarkoitus poistaa karkeat virheet ja varmistaa, että aineisto on mallien oletuksien mukaista.

Karkeat virheet on syytä poistaa, sillä ne voivat häiritä muuttujien kertoimien estimointia merkittävästi. Tämä voi johtaa siihen, ettei malli lopulta

kuvaa hyvin aineiston taustalla olevaa ilmiötä. Usein tämä ilmenee mallin huonona selityksasteena. Lisäksi puuttuvat havainnot on hyvä ottaa huomioon tässä vaiheessa.

Aikasarja-analyysissä lähtöoletuksena on, että aikasarja on jonkin stationaarisen stokastisen prosessin realisaatio [10]. Stationaarisessa stokastisessa prosessissa taas ei ilmene trendiä, varianssin (systemaattista) vaihtelua, kausivaihtelua tai sisäisen riippuvuusrakenteen vaihtelua [12]. Täytyy siis varmistaa, että nämä ehdot täyttyvät. Aikasarjan stationarisoimiseksi on kuitenkin olemassa työkaluja, jotka ovat usein yksinkertaisia matemaattisia operaatioita.

Aineisto on myös tapana jakaa estimointi- ja validointiaineistoon. Estimointiaineistoksi valittiin vähintään puolet koko käytettävästä aineistosta. Estimointiaineistoa käytetään mallin parametrien estimoinnissa ja saadun mallin hyvyys tarkistetaan validointiaineistolla, jolloin voidaan nähdä sopivatko mallin antamat tulokset oikeisiin havaintoihin.

### 3.2.2 Mallin rakenneparametrien valinta ja muuttujien kertoimien estimointi

Ennen mallin parametrien estimoinnista täytyy valita malli, joka estimoidaan. Aineistoon tutustumalla voi saada näkemystä sopivasta rakenteesta eli voisiko kyseessä olla esimerkiksi AR- tai ARMA-malli ja minkälaisilla parametreilla. Myös ulkoisten selittäjien tarve tulee selvittää. Erilaisia malleja voi myös kokeilla ja tehdä lopullinen päätös vasta tämän jälkeen. Tämä voi olla tarpeen, sillä silmämääräisesti sopivan mallityypin ja sen rakenteen hahmottaminen aikasarjasta on haastavaa.

Mallin rakenne on syytä säilyttää mahdollisimman yksinkertaisena, eli parametrien suuruus tulisi pitää mahdollisimman matalana. Mitä yksinkertaisemmalla mallilla voidaan kuvata hyvin kyseessä olevaa ilmiötä sitä parempi.

ARX-mallin muuttujien kertoimien estimoinnissa käytetään pienimmän neliösumman menetelmää. Toisin sanoen etsitään mallille sellaisia kertoimia, että todellisen havainnon ja mallin antaman sovituksen erotuksen neliöiden summa olisi mahdollisimman pieni. [5]

### 3.2.3 Mallin validointi

Mallin parametrien estimoinnin jälkeen on syytä tutkia aikaansaadun mallin hyvyttä eli validoida sitä. Ennen kaikkea on syytä tarkistaa, antaako malli ulos järkeviä tuloksia, eli sopiiko malli kuvaamaan kyseessä olevaa ilmiötä. Mallin validoinnissa tärkeä osa on havaintojen simulointi tai ennustaminen ja näiden tuloksien vertailu oikeisiin havaintoihin eli validointiaineistoon. [8] Simulaation ja validointiaineiston yhteensopivuudesta kertoo mallin selitysvaiva. Matlab käyttää selitysvaivasta nimeä *fit* ja sen laskentakaava on seuraava

$$fit = \left[ 1 - \frac{\|y - \hat{y}\|}{\|y - MEAN(y)\|} \right] \cdot 100, \quad (5)$$

jossa  $y$  on todelliset havainnot sisältävä vektori ja  $\hat{y}$  sovitevektori.  $MEAN(y)$  tarkoittaa havaintovektorin arvojen keskiarvoa. Todellisen havainnon ja ennusteen välistä erotusta,  $y - \hat{y}$ , kutsutaan residuaaliksi eli jäännöstermiksi. [4]

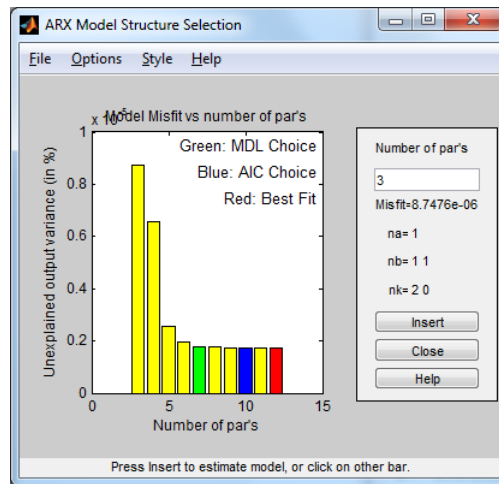
Hyvän mallin residuaalien pitäisi olla valkoista kohinaa. Toisin sanoen niiden pitäisi olla normaalijakautuneita odotusarvolla nolla ja varianssilla  $\sigma^2$ , jossa varianssi on vakio kaikille residuaaleille. Residuaalien autokorrelaatiota on syytä tutkia esimerkiksi tekemällä niistä kuvajia ja arvioimalla niiden merkittävyyttä.

Tässä projektissa mallien validointi ja järjestäminen paremmuusjärjestykseen tapahtuu selitysvaivan ja residuaalianalyysin perusteella. Muita keinoja vertailla mallien hyvyttä on erilaisten informaatiokriteerien käyttäminen ja niiden arvojen vertailu eri malleille, jonka jälkeen voidaan valita parhaat kriteerien arvot antava malli. [6]

System Identification Toolbox vertailee malleja Akaiken informaatiokriteerillä (AIC), Rissanen MDL-kriteerillä sekä validointiaineiston ja mallin ulostulon erotuksen perusteella ja ilmoittaa kullakin kriteerillä parhaan mallin [7]. Akaiken informaatiokriteeri on kompromissi sovitteen hyvyden ja mallin monimutkaisuuden eli parametrien lukumäärän välillä [18]. Rissanen MDL kriteeri eli Minimum Description Length perustuu taas tiedon pakkaamiseen mahdollisimman pieneen määrään merkkejä [20].

### 3.2.4 Käytännön toteutus

Mallien rakentamisen pääasiallinen työkalu on Matlab ja erityisesti sen System Identification Toolbox. Kyseisessä Matlab-lisäosassa on monipuolinen



Kuva 3: Kuvankaappaus mallinvalinnasta System Identification Toolbox:ssa käyttäen hyväksi erilaisia kriteerejä.

graafinen käyttöliittymä jonka avulla tehdään suurin osa analyysistä ja kuvien tuottamisesta. Graafinen käyttöliittymä on kuitenkin aika rajoittunut, ja osaan analyysistä käytetäänkin System Identification Toolbox:ia ilman graafista käyttöliittymää.

Aiemmin esiteltyt menetelmät mallien rakentamisesta ja validoinnista kirjoitettiin erityisesti pitäen silmällä System Identification Toolbox:n käyttämiä menetelmiä, sillä se on pääasiallisesti käytetty työkalu tässä projektissa. System Identification Toolbox:in käyttäminen ilman graafista käyttöliittymää nopeuttaa mallintamista, sillä samaa analyysiä voidaan ajaa moneen kertaan nopeasti eri aineistolle. Tätä käytetään erityisesti aineiston esikäsittelyyn.

System Identification Toolboxille voi antaa jokaiselle mallin parametrille välin, jonka sisällä sen arvo voi vaihdella. Ohjelma käy tämän jälkeen läpi kaikki parametrien kombinaatiot ja estimoi erilaisille malleille muuttujien kertoimet. Mallin parametrien arvoja on käyty läpi prioritietämyksen ja laskentaajan järkevyyden pohjalta noin nollan ja kymmenen väliltä. Tarkasteltujen mallien lukumäärät ovat siis olleet välistä riippuen kymmenestä tuhannesta satoihin tuhansiin. System Identification Toolbox esittää mallit siten, että niiden joukosta voi valita tietyillä informaatiokriteereillä ja muilla kriteereillä parhaan vaihtoehdon ja vertailla viiveparametrien suuruutta eri malleissa. Kuvassa 3 näkyy miten ohjelmisto ilmoittaa mallinvalintakriteerit malleille. Ohjelma ei esitä huonoimpia malleja. Kriteerit ovat suuntaa antavia merkkejä mallin hyvyyden arviointiin. Ohjelman antamiin parhaisiin malleihin ei pidä luottaa sokeasti, vaan niiden avulla saa käsityksen minkä suuruisia pa-

parametrit voisivat olla. Lopullinen malli täytyy etsiä käsin kokonaisvaltaisen validoinnin ja prioritetämyksen pohjalta, eikä minkään yksittäisen kriteerin perusteella.

Validointivaiheessa ohjelmaa käytetään laskemaan selitysvoima sekä estimointiaineistolle että validointiaineistolle, piirtämään kuvat residuaaleista ja ennusteista ja residuaalien autokorrelaatioista. Ennustamiseen valittiin sellainen tapa, että joka pisteessä ennustetaan 24 tuntia eteenpäin ja tallennetaan tämä kyseinen tulos. Seuraavalla askeleella ennustetaan myös 24 tuntia eteenpäin, mutta nyt käytössä on yksi tunti enemmän historia-aineistoa, ja niin edelleen. Lopputuloksena on joka tunnilta 24 tunnin päähän tehty ennuste ja näiden ennusteiden hyvyttä arvioidaan. Myös muita ennustepituuksia testataan, mutta pituutta 24 on käytetty kaikissa raporttiin liitetyissä kuvissa. Tähän päädyttiin, sillä tehtävänasettaja totesi seuraavan vuorokauden olevan keskeisin aikaikkuna ennusteen käyttötarkoituksen kannalta.

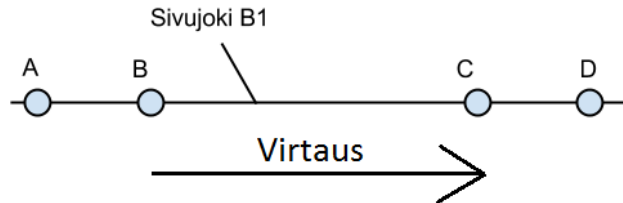
Mallia validoitaessa kiinnitettiin erityistä huomiota sovittujen hyvyyden ja residuaalien autokorrelaation lisäksi muuttujille estimoitujen kertoimien merkkiin ja itseisarvoon. Jokaisella ulkoisella selittäjällä on tietty merkitys ja tiedetään korottaako vai alentaako kyseinen ohjausmuuttuja pinnan korkeutta. Positiivinen kerroin pitäisi olla kaikilla ohjausmuuttujilla, jotka tuovat vettä laitokselle ja vastaavasti negatiivinen sellaisilla jotka vievät vettä pois laitokselta. Jos kertoimen merkki oli vääränmerkkinen, parametrien arvoja pyrittiin muuttamaan siten, että kaikki kertoimet ovat oikean merkisiä. Itseisarvoltaan arvoltaan liian pienet kertoimet poistettiin, sillä niiden vaikutus malliin on hyvin pieni ja siten ne olivat tarpeettomia. Poistetut kertoimet olivat suuruusluokaltaan 10-100 kertaa pienempiä kuin mallin muut kertoimet. Parametrien muutoksen jälkeen malli estimoitui uudelleen ja tutkittiin paranivatko tulokset.

System Identification Toolboxin tulostuksista käytetään myös residuaalien autokorrelaatiokuvaajia. Niiden perusteella arvioidaan, antavatko residuaalit näyttöä mallin virheellisyydestä. Autokorrelaatiokuvaajissa näkyvät myös haluttua merkitsevyytensä vastaavat luottamusrajat. Työssä käytetään Matlabin oletusarvoisesti käyttämää 99%:n merkitsevyytensä. Mikäli arvot osuvat tälle välille, pitäisi residuaalien autokorrelaation olla merkityksettömiä systeemille tällä merkitsevyytensä.

Lopullisen mallin löytäminen on iteratiivinen prosessi, jossa joudutaan tekemään kompromisseja. Mahdollisten mallien joukosta ei aina löydy yhtä, joka on ylitse muiden kaikkien kriteerien suhteen. Tällöin saatetaan joutua joustamaan jostain kriteereistä, sillä parempaa mallia ei ole saatavissa.



## 4 Aineisto



Kuva 4: Kaaviokuva vesivoimaloiden A, B, C ja D sijainnista toisiinsa nähden. Joki virtaa vasemmalta oikealle. Jokeen voimalaitosten B ja C välissä virtaava sivujoki B1 on myös merkitty kuvaan.

Tarkastelemme neljää vesivoimalaa, joista A sijaitsee yläjuoksulla ja muut ovat järjestyksessä sen jälkeen kuvan 4 mukaisesti. Laitosten B ja C välissä jokeen laskee useita sivujokia, mutta mukaan otetaan niistä vain suurin, B1, joka kattaa noin 50% kaikkien sivujokien virtaamista.

Käytettävissä on kahdenlaista aineistoa mallinnettavasta joesta: pinnan korkeuksia ja virtaamia vuosilta 2008-2010 ja 2013. Pääosin aineiston resoluutio on tunti, mutta vuoden 2013 laitosten B ja C välissä olevan sivujoen B1 virtaama on ilmoitettu päivän keskiarvoina.

Erityistä huomiota pyydettiin kiinnittämään laitoksen C pinnan korkeuteen, sillä sen ennustaminen on historiassa ollut kaikkein haastavinta, johtuen muun muassa pitkästä etäisyydestä B:n ja C:n välillä. Laitos D on toiseksi tärkein ja vähiten tärkeämmäksi koettiin laitos B. Laitos B toimii pintaohjauksella, eli sen juoksutusta säädelään automaattisesti, jotta vedenpinnan korkeus pysyy vakiona. Sen juoksutusta ei siis optimoida suurimmassa osaa ajasta ollenkaan, jolloin veden liikkeen mallinnukselle ei ole suurta tarvetta. Laitoksen A vedenpinnan korkeutta ei ennustettu, sillä se toimi systeemin alkupisteenä.

Alkuperäinen tavoite oli löytää yksi malli kuvaamaan joen jokaista osuutta. Aika nopeasti huomattiin että sama väli tarvitsee erilaisen mallin riippuen vesitilanteesta ja juoksutusstrategiasta. Tästä syystä UPM antoi kolme erillistä tapausta joille luoda malli: profiliajo, niukka vesitilanne ja jääkansi. Ne olivat tyypillisiä olosuhteita, joiden vallitessa matemaattiselle mallille tulisi todennäköisesti olemaan käyttöä. Nämä olivat aikavälin 24.10.-10.12.2013 profiliajo, aikavälin 21.10.-19.11.2009 niukka vesitilanne ja aikavälien 1.2.-15.3.2009 sekä 20.1.-1.3.2014 jääkannen peittämä virtaus.

Laitokselle C luotiin annetun ajan lisäksi profiliajomallit ajoilta 3.2.2009-30.3.2009 ja 28.01.2010-19.02.2010. Tähän päädyttiin, sillä vuoden 2013 aineistosta saatu profiliajomalli oli huono. Tämä saattoi johtua sivujoen virtaaman arvojen epätarkkuudesta. Sivujoen virtauksen arvot vaihtelivat myös voimakkaasti tänä aikana, joten sivujoen vaikutus pinnan korkeuteen oli suuri. Vuorokauden keskiarvot eivät vain riittäneet hyvän mallin muodostamiseen. Huonoihin malleihin ei tyydytä, sillä laitoksen C vedenpinnan korkeuden ennustaminen oli annettu projektin päätehtäväksi. Kyseisten aikavälien ominaisuuksista ja saaduista malleista kerrotaan tarkemmin tulokset osiossa.

## 4.1 Aineiston esikäsittely

Alkuperäisestä aineistosta siis otettiin vain pieniä osia, joita analysoitiin erikseen. Näissä alueissa oletuksena oli että juoksutusstrategiat ja vesitilanne pysyivät samoina, jolloin trendinpoistolle ei ollut tarvetta. Stationaarisuutta myös analysoitiin jokaisen ajanjakson kohdalla erikseen. Osassa aineistoa oli selvästi epästationaarisia ajanjaksoja, jolloin tarkasteltavaa pätkeä saatettiin hiukan lyhentää. Epästationaarisuutta aiheutti esimerkiksi juoksutusstrategian muutos tai viikonloput, jolloin juoksutus on keskimäärin selkeästi pienempää kuin arkipäivisin. Lopullisiin aikaväleihin jäi myös epästationaarisia kohtia, minkä poistolle olisi ehkä ollut aihetta, mutta ajan puutteen takia ei ehditty etsimään korvaavia ajanjaksoja tilalle.

## 4.2 Mallin rakenteen valinta

Päädyimme ARX-malleihin sillä, ne ovat ryhmän jäsenille tuttuja ja melko yksinkertaisia. Ne sopivat fysikaaliseen tulkintaan systeemistä ja niiden alustavat tulokset olivat melko hyviä. Lisäksi UPM:n yhteyshenkilö piti niitä sopivina. ARX-mallien muodostaminen System Identification Toolbox:illa oli myös erityisen helppoa. Ohjelmalla oli mahdollista estimoida samanaikaisesti useita ARX-malleja, mutta ei monimutkaisempia malleja kuten SARMAX-malleja. Jossain mielessä aineisto tuki myös SARX tai SARMAX malleja, koska juoksutuksissa ja pinnanvaihtelussa näkyi päivittäistä vaihtelua. Viikonloppu eroaa myös selvästi arkipäivistä. Nämä mallit jätimme pois ensinnäkin ajan puutteen takia ja lisäksi vaihtelu sekä päivittäin että viikoittain tuntui vaihtelevan merkittävästi lyhyelläkin aikavälillä.

## 5 Tulokset

Jokaiselle laitokselle B, C ja D luotiin ARX-aikasarjamallit aikaisemmin mainituilla aikaväleillä. Tulokset osiossa käydään läpi jokainen aikaväli, niille luodut mallit ja mallien ominaisuuksia. Yhteenvedossa ja pohdinnoissa verrataan mallien antamia tuloksia ja malleja ylipäänsä.

Laitoksen vedenpinnankorkeutta ennustetaan käyttämällä aikasarjamallia  $ARX(na, [nb1\ nb2\ nb3], [nk1\ nk2\ nk3])$ . Selittävät muuttujat ovat järjestyksessä: edellisen laitoksen juoksutus, tarkasteltavan laitoksen juoksutus ja sivujoen virtaama. Sivujokea ei ole jokaisessa mallissa, jolloin muuttujien järjestys pysyy samana. Tällöin ensimmäinen muuttuja on edellisen laitoksen juoksutus ja toinen muuttuja on tarkasteltavan laitoksen juoksutus. Tällöin malli on muoto  $ARX(na, [nb1\ nb2], [nk1\ nk2])$ .

### 5.1 Tasainen profiliajo

Profiliajolla tarkoitetaan karkeasti ottaen tilannetta, jolloin vettä on saatavilla ja päivällä juoksutus on lähellä laitoksen maksimia, eli tyypillisesti n. 200-400 m<sup>3</sup>/s, ja yöllä reilusti vähemmän eli vähintäänkin puolet päiväjuoksutuksesta. Lisäksi päivällä voi olla havaittavissa aamu- ja iltapäivän juoksutuspiikit, jolloin sähkön kysyntä on korkeimmillaan. Pienen sähkön kysynnän takia juoksutus laskee viikonloppuisin merkittävästi, UPM:ltä saatu esimerkkiaineisto profiliajolle löytyy ajalta 24.10-10.12.2013, mutta laitokselle C luotiin profiliajomallit myös ajoille 3.2.2009-30.3.2009 ja 28.01.2010-19.02.2010.

#### 5.1.1 Vesivoimalan B juoksutusten ennustaminen

Tarkastellaan laitosten A ja B väliä. Tämän välin mallinvalintaan liittyvät kuvat löytyvät liitteestä A. Laitoksen A juoksutusta säädelään normaalisti, mutta laitos B toimii pääasiassa pintajuoksutuksella. Tämä tarkoittaa sitä, että laitoksen B pinnankorkeus pidetään lähes vakiona, sen juoksutusta muuttamalla. Ei ole järkevää ennustaa laitoksen B pinnankorkeutta laitosten A ja B juoksutuksilla, sillä vedenpinnassa tapahtuu hyvin vähän muutoksia. Tämän takia ennustetaan laitoksen B juoksutusta laitoksen A juoksutuksella. Tarkastellaan aikaväliä 24.10-10.12.2013, jolloin ollaan ajettu tasaista profiliajtoa.

Kuvassa 5 on esitetty laitosten juoksutukset tältä aikaväliltä. Laitoksen A:n juoksutusten keskiarvo on 164,52 m<sup>3</sup>/s ja laitoksen B 153,37 m<sup>3</sup>/s. Vettä

joko häviää matkalla tai, sitten virtauksen suuruuden arvioinnissa on epävarmuutta.

Aineiston ensimmäiset 790 tuntia valitaan estimointidataksi ja loput 360 tuntia validointidataksi. Aineistosta estimoitiin erilaisia ARX-malleja, siten että niiden parametrien  $n_a$ ,  $n_b$  ja  $n_k$  arvot vaihtelivat välillä 1-10. Saatujen mallien hyvyyttä arvioitiin aluksi selitysvoiman ja havaintojen varianssin selittämätömän osuuden avulla. Tämän jälkeen tutkittiin parhaiden mallien residuaalien autokorrelaatiota. Parhaaksi malliksi valittiin alustavasti ARX(4,5,1).

Diskreettiaikainen ARX-malli on muotoa  $A(z)y(t) = B(z)u(t) + e(t)$ , jossa

$$A(z) = 1 - 0.9377z^{-1} + 0.004858z^{-2} + 0.002018z^{-3} + 0.06676z^{-4} \quad (6)$$

ja

$$B(z) = 0.1792z^{-1} + 0.1109z^{-2} - 0.07125z^{-3} - 0.05733z^{-4} - 0.03331z^{-5} \quad (7)$$

Mallin ARX(4,5,1) ennuste validointiaineistolle on esitetty kuvassa 6. Mallin selitysvoima on hyvin korkea 93,48, mutta se ennustaa lähes säännönmukaisesti liian suurta juokсутusta. Kuvasta 7 nähdään, että ennusteen residuaalit ovat jonkin verran autokorreloituneita.

Mallia ARX(5,4,1) validoitiin koko vuoden 2013 aineistolla ja sen ennuste on esitetty kuvassa 8. Malli näyttää toimivan kohtuullisesti koko vuoden ympäri ja sen selitysvoima on 93,67. Ennusteen residuaalit on esitetty kuvassa 9 ja siitä nähdään, että koko vuoden residuaalit ovat autokorreloituneita. Mallissa on myös melko paljon parametreja. Lisäksi laitoksen B juokсутus riippuu osittain negatiivisesti laitoksen A juokсутuksesta. Tämä ei ole järkevää, sillä laitoksen A juokсутuksen lisäyksen ei pitäisi vähentää laitoksen B juokсутusta millään viiveellä. Vähennetään parametrien määrää, kunnes saadaan järkevä malli.

Parhaaksi malliksi osoittautui ARX(1,2,1). Sen ennuste validointiaineistolle on esitetty kuvassa 10 ja sen selitysvoima on 93,36. Kuvassa 11 on esitetty ennusteen residuaalien autokorrelaatio. Kuvasta nähdään, että ne ovat jonkin verran autokorreloituneita. Mallin kertoimet ovat

$$A(z) = 1 - 0.6326z^{-1} \quad (8)$$

ja

$$B(z) = 0.1795z^{-1} + 0.167z^{-2}. \quad (9)$$

Kuvassa 12 on esitetty mallin ARX(1,2,1) ennuste koko vuodelle 2013 ja kuvassa 13 sen residuaalit. Residuaalit ovat keskimäärin melko pieniä, mutta

ennuste on epävarma muutamassa kohtaa. Erityisen huonosti malli ennustaa ajanjaksoa 8500h-9000h. Kuvassa 14 on koko vuoden ennusteen residuaalien autokorrelaation kuvaaja. Siitä nähdään, että residuaalit ovat voimakkaasti autokorreloituneita.

ARX(1,2,1) osoittautui vertailussa lähes yhtä hyväksi malliksi kuin ARX(4,5,1) ja siinä on huomattavasti vähemmän parametreja. ARX(1,2,1):ssä Laitoksen B virtausta ennustetaan, laitoksen A yksi ja kaksi tuntia sitten tapahtuneella juoksutuksella, sekä laitoksen B tuntia aikaisemmalla juoksutuksella. UPM:n arvioi veden virtaavan laitokselta A laitokselle B noin kahdessa tunnissa. Molempien mallien parametrit tukevat tätä tulkintaa, sillä niissä molemmissa on viiveterminä yksi tunti.

ARX(1,2,1) osoittautui parhaimmaksi ARX-malliksi, mitä profiliajolle saatiin aikavälillä 24.10-10.12.2013. Malli ei kuitenkaan pysty selittämään ulostuloa täydellisesti ja tekee systemaattista virhettä. Sitä pystyttiin käyttämään kohtuullisen hyvin koko vuoden ajalle. Tämä ei ole itsestäänselvyys, sillä oletamme joen dynamiikan muuttuvan veden määrään ja lämpötilan mukaan.

### 5.1.2 Vesivoimalan C pinnankorkeuden ennustaminen vuosien 2010 ja 2013 aineistoilla

Laitoksen C vedenpinnankorkeutta ennustetaan käyttämällä aikasarjamallia  $ARX(na, [nb1 \ nb2 \ nb3], [nk1 \ nk2 \ nk3])$ . Selittävät muuttujat ovat järjestyksessä: laitoksen B juoksutus, laitoksen C juoksutus ja sivujoen virtaama. Sivujoen virtaukselle on olemassa arvioita eri tarkkuudella. 2013 aineistossa sivujoen virtaamalle on annettu arvio päivän tarkkuudella. Vuosien 2008-2010 aineistossa sivujoen virtaama on taas annettu tunneittain.

Vuoden 2013 sivujoen virtauksen aikasarjan arvot on annettu 24 tunnin välein ja mallin muiden aikasarjojen muuttujien arvot on annettu tunnin välein. Jos virtaukset annettaisiin vuorokauden keskiarvoina, voisi joka keskiyö sivujoen arvo muuttua voimakkaasti, vaikka todellisuudessa muutos on tapahtunut pitkin päivää. Tämä ongelma yritettiin ratkaista laskemalla jokaiselle tunnille sivujoen virtaus liukuvalla keskiarvolla, jolloin sivujoen aikasarjan resoluutio saataisiin muutettua muiden kanssa sopivaksi.

UPM:n ehdottamalle profiliajopätkälle 24.10-10.12.2013 yritettiin luoda ARX-mallia, jossa sivujoen ohjauksessa on hyödynnetty liukuvaa keskiarvoa. Paras saatu malli oli muotoa  $ARX(3, [1 \ 2 \ 1], [6 \ 1 \ 1])$ . Sillä saatiin huonoja tuloksia. Epäonnistumisen syyksi epäiltiin epätarkkoja sivujoen virtaaman arvoja ja

liukuvan keskiarvon epäluotettavuutta. Tämä oli erityisen haitallista, sillä sivujoen arvot vaihtelivat kyseisenä aikana 10-100  $m^3/s$  välillä, jolloin sivujoen virtauksella on suuri vaikutus systeemiin. Epätarkan aineiston takia siirryttiin tutkimaan vuosia 2008-2010, jossa sivujoen virtaamasta oli saatavilla tarkempia arvioita.

Parhaiten analyysiin sopiva aikaväli oli 28.01.2010-19.02.2010. Kuvassa 15 on esitetty selittävien muuttujien kuvaajat. Kuvaajasta nähdään, että sivujoen virtaus on hyvin tasaista. Aineistosta näkee selkeästi arkipäivät ja viikonloput. Viikonloppuina juoksutus on alhaista ja lähes vakiota. Se eroaa selkeästi muusta aineistosta ja se saattaa vaikeuttaa mallin estimointia. Arkipäivinä on aamupäivän ja iltapäivän juoksutuspiikit näkyvissä selkeästi. Keskimääräinen juoksutus on tuona aikana laitoksella B 105,79  $m^3/s$ , laitoksella C 104,25  $m^3/s$  ja sivujoessa 8,29  $m^3/s$ . Täten joko vettä katoaa jonnekin, tai mittaukset ovat epätarkkoja. Kuvassa 16 on esitetty laitoksen C vedenpinnan korkeus. Toisenkin hyvännäköinen profiliajopätkä löytyi vuoden 2009 aineistosta ja sitä käsitellään seuraavassa alaluvussa.

Jaetaan aineisto estimointidataan, johon kuuluu ensimmäiset 342 tuntia ja validointidataan johon kuuluu loput 109 tuntia. Parhaassa mallissa ARX(1,[1 1 0],[1 0 0]) ei ollut mukana sivujoen virtaamaa. Sen muuttujien kertoimet ovat

$$A(z) = 1 - z^{-1}, \quad (10)$$

$$B_1(z) = 0.0006917z^{-1} \quad (11)$$

ja

$$B_2(z) = -0.0007695z^{-1}. \quad (12)$$

Yhtä hyviä tuloksia saatiin myös mallilla ARX(1,[1 1 1],[1 0 1]), jossa on mukana sivujoen virtaama. Tässä mallissa sivujoen muuttujan kerroin oli itseisarvoltaan hyvin pieni  $9,052 \cdot 10^{-5}$ , joten se poistettiin mallista tarpeettomana. Sivujoen virtaama on tänä ajankohtana hyvin tasaista ja vähäistä, joten sen poisto mallista on fyysikaalisen tulkinnankin mukaan järkevää. Validointiaineiston todelliset arvot ja mallin ARX(1,[1 1 0],[1 0 0]) ennuste ovat kuvassa 17 ja residuaalit kuvassa 18. Residuaalien autokorrelaation kuvaaja on esitetty kuvassa 19.

Mallin selitysvaikutus on 65,2, joka on melko hyvä tulos. Kuvasta 17 nähdään, että malli onnistuu ennustamaan laitoksen vedenpinnan korkeutta hyvin. Residuaalit eivät kuvan 19 perusteella ole autokorreloituneita. Laitoksen B juoksutuksen komponentti on positiivinen ja laitoksen C juoksutuksen komponentti on negatiivinen. Tämä tulos on järkevä, sillä laitoksen B juoksutuksen

täytyy nostaa vedenpintaa ja laitoksen C juoksutuksen täytyy laskea sitä. Laitoksen C juoksutuksen viiveeksi saatiin 0 tuntia, eli laitoksen C juoksutus vaikuttaa viiveettömästi vedenpinnan korkeuteen. Laitoksen B juoksutuksen viive on erikoisen pieni, sillä laitosten B ja C välimatka on useita kymmeniä kilometrejä. On epätodennäköistä, että vesi ehtii kulkea tämä matkan näin lyhyessä ajassa. Suuremman viiveen pakottaminen ja parametrien kasvattaminen eivät tuottaneet parempia tuloksia. Tällä mallilla saatiin hyviä tuloksia myös täydellä simulaatiolla, jossa mallille ei annettu todellisia arvoja 24 tunnin takaa, vaan mallille annettiin vain alkuarvot ja ohjaukset. Tällöin mallien selitysvoimaksi saatiin 59,74. Saatuun malliin voidaan siis olla tyytyväisiä.

### 5.1.3 Vesivoimalan C pinnankorkeuden ennustaminen vuoden 2009 aineistolla

Tämän tapauksen mallinvalintaan liittyvät kuvat löytyvät liitteestä C. Kuvassa 20 on esitetty laitoksen C vedenpinnan korkeutta selittävien muuttujien arvot siltä ajalta. Keskimääräiset juoksutukset laitoksella B on  $245,59 \text{ m}^3/\text{s}$ , laitoksella C  $235,49 \text{ m}^3/\text{s}$  ja sivujoella  $8,99 \text{ m}^3/\text{s}$ . Laitosten välillä on muitakin sivujokia, mutta käytettävissämme on vain suurimman sivujoen virtauksen tiedot. Se kattaa karkeasti mitattuna noin 50% sivujokien kokonaisvirtauksesta. Laitoksen C juoksutuksen voisi siis olettaa olevan noin  $263 \text{ m}^3/\text{s}$ , jotta sisään tulevan ja poistuvan veden määrä olisi sama. Ero voi selittyä juoksutusten mittauksen epätarkkuudella tai sitten vettä valuu matkan varrella pois. Kuvassa 21 on esitetty laitoksen C vedenpinnan korkeus.

Parhaaksi malliksi saatiin  $\text{Arx}(1, [2 \ 1 \ 1], [1 \ 1 \ 1])$ , jossa mallin kertoimet olivat

$$A(z) = 1 - z^{-1}, \quad (13)$$

$$B_1(z) = 1.511 \cdot 10^{-5} z^{-1} + 0.0005811 z^{-2}, \quad (14)$$

$$B_2(z) = -0.000637 z^{-1} \quad (15)$$

ja

$$B_3(z) = 6.905 \cdot 10^{-6} z^{-1}. \quad (16)$$

Muuttujien kertoimet ovat järkevän merkkisiä. Hyvin useissa malleissa sivujoen kerroin on negatiivinen ja sen itseisarvo on kaikissa malleissa pieni. Voi olla järkevää jättää se kokonaan pois, sillä se ei vaikuta merkittävästi vedenpinnan korkeuteen, jos sen virtaus pysyy vakiona. Viive on yksi tunti ja se on epäilyttävän lyhyt laitoksen B juoksutukselle ja sivujoen virtaamalle. Matka

on hyvin pitkä ja tämä näkyy ehkäpä myös siinä, että laitoksen B juoksu-  
tuksen toinen termi on huomattavasti suurempi kuin ensimmäinen. Ensimmäinen  
termi oli hyvin pieni, joten sen poistaminen voisi tulla kyseeseen. Fysikaalista  
tulkintaa tukevat suurempiviiveiset mallit tuottivat huonoja tuloksia, joten  
saatu malli oli ainut järkevä vaihtoehto. Kuvassa 22 on esitetty mallin sovite  
ja kuvassa 23 sen residuaalit. Mallin selitysvaivo on 70,57. Kuvassa 22 sovite  
näyttää seuraavan hyvin todellista arvoa, mutta residuaaleja tarkastellessa  
huomataan virheiden olevan korkeimmillaan 15 senttimetrin luokkaa.

Sivujoen virtaus on tällä ajanjaksolla melko tasaista mutta mukana on kaksi  
huippua. Aineisto ei ole ihanteellista sivujoen virtauksen vaikutuksen estimointiin.  
Lisäksi sivujoen kerroin on useissa estimoiduissa malleissa negatiivisia ja  
itseisarvoltaan hyvin pieniä. Tämän takia voisi olla perusteltua poistaa se  
mallista. Näin tehtiin ja parhaaksi malliksi saatiin ARX(1,[2 1],[1 1]), jonka  
kertoimet ovat

$$A(z) = 1 - z^{-1}, \quad (17)$$

$$B_1(z) = 1.506 \cdot 10^{-5} z^{-1} + 0.0005811 z^{-2} \quad (18)$$

ja

$$B_2(z) = -0.0006369 z^{-1}. \quad (19)$$

Sen antama ennuste oli erittäin lähellä aikaisempaa mallia ja residuaalien  
autokorrelaatio on molemmissa käytännössä samat. Mallin ARX(1,[2 1],[1 1])  
fit on aavistuksen parempi 70,66, joten se on parempi malli tälle välille.  
Lisäksi mallien parametrien arvot ovat sivujokea lukuun ottamatta samat ja  
muuttujien kertoimet lähes identtiset. Sivujoen voi siis jättää pois mallista,  
jos sen virtaus on pientä ja tasaista.

#### 5.1.4 Vesivoimalan D pinnankorkeuden ennustaminen

Tämän ajanjakson mallintamiseen liittyvät kuvat löytyvät liitteestä D. Laitoksen  
D vedenpinnankorkeutta ennustetaan ARX-mallilla, jossa selittäjinä ovat  
laitosten C ja D juoksu-  
tuksut. Niiden kuvaajat aikaväliltä 24.10-10.12.2013 löytyvät kuvasta 25. Laitoksen C  
vedenpinnankorkeus samalta ajalta löytyy kuvasta 26. Keskimääräiset juoksu-  
tuksut tuona aikana on laitoksella C 202,88  $m^3/s$  ja laitoksella D 200,17  $m^3/s$ .  
Juoksu-  
tusten välinen ero ei ole merkittävä.

Kyseinen aineisto on 1154 tuntia pitkä, josta estimointiin käytettiin ensimmäiset  
800 tuntia ja validointiin viimeiset 354 tuntia. Parhaaksi malliksi



osoittautui ARX(1,[2 1],[1 1]), jonka kertoimet muuttujille olivat

$$A(z) = 1 - z^{-1}, \quad (20)$$

$$B_1(z) = 0.0003379z^{-1} + 0.0003324z^{-2} \quad (21)$$

ja

$$B_2(z) = -0.0006652z^{-1}. \quad (22)$$

Kuvassa 27 on mallin sovite validointidatalle ja kuvassa 28 sen residuaalit. Kuvassa 29 on residuaalien autokorrelaation kuvaaja, josta nähdään että ne eivät ole autokorreloituneet. Mallin selitysvoima on 64,12. Tämä on hyvin yksinkertainen malli ja se sopii systeemin fysikaalisen tulkintaan. Laitosten C ja D välimatka on noin 14 kilometriä ja malli ottaa huomioon kaksi ja kolme tuntia sitten tapahtuneet juoksutukset laitoksella C. Tämä viive vaikuttaa järkevältä. Lisäksi muuttujien kertoimet ovat oikean merkkisiä.

Parametrien kasvattaminen antoi joissakin tapauksissa aavistuksen paremman sovituksen, mutta silloin osa muuttujien kertoimista saattoivat olla väärän merkkisiä. Esimerkiksi jokin laitoksen D juoksutuksen kertoimista oli positiivinen, joka viittaisi siihen, että laitoksen D juoksutuksen kasvattaminen nostaisi vedenpintaa jollakin viiveellä. Tämä ei ole mahdollista, joten ne mallit eivät kuvaa systeemin todellista toimintaa. Tämän takia laitoksen D juoksutus on mallissa mukana vain yhdellä komponentilla.

## 5.2 Jääkansi

Tapaus jääkansi koskee aikaväliä 1.2.-15.3.2009. Se on talvelta, kun joessa on jääpinta ja vettä ja juoksutusta on reippaasti. Se siis edustaa talven tavanomaista tilannetta, kun ei ole ongelmia ja vettä on paljon. Saimme UPM:ltä myös toisen aikavälin kyseiselle tilanteelle: 20.1.-1.3.2014. Vuoden 2014 aineistoa ei käytetty ajan puutteen takia ja vuoden 2009 aineisto oli laajempi joten sitä käytettiin ensin.

### 5.2.1 Vesivoimalan C pinnankorkeuden ennustaminen

Vuoden 2009 aineisto on mielenkiintoisempaa, koska myös sivujoen virtaamat löytyvät siitä tuntitasolla. Eli tärkeämpänä tutkimuskohteena pidetään vuoden 2009 kevättalvea ja tuloksista riippuen analyysiä jatketaan vuoden 2014 kevättalvelle.

Tämän tapauksen mallintamiseen liittyvät kuvat löytyvät liitteestä E. Kuvassa 30 näkyy selvästi vuorokauden mittaista kausivaihtelua. Vaihtelu vuorokauden sisällä on suurimmillaan lähes yhtä suurta kuin suurin vaihtelu koko ajanjakson sisällä. Pinnankorkeuden käyttäytyminen ei ole ihan samanlaista alku- ja loppupuoliskossa, joten aineiston jakaminen kahtia ja validointi loppupuoliskolla ei välttämättä tuota hyviä tuloksia.

Kuvan 31 perusteella virtaukset laitoksilla B ja C ovat keskimäärin yhtä suuria ja sivujoen B1 virtaus merkittävästi pienempi. Laitoksen B virtaama ei vaihtelee vuorokauden sisällä, kun taas laitoksen C virtaama vaihtelee voimakkaasti. Sivujoen B1 virtaama on niin paljon pienempi, että tuntuu lähes kuin sen voisi jättää kokonaan pois analysista. Keskimääräiset virtaamat kyseessäolevalla aikavälillä olivat  $227 \text{ m}^3/\text{s}$  laitoksella B,  $253 \text{ m}^3/\text{s}$  laitoksella C ja sivujoessa B1  $10 \text{ m}^3/\text{s}$ . Näistä tiedoista voidaan havaita, että suurin osa laitoksen C virtaamasta selittyy pelkästään laitoksen B virtaamalla, mutta noin  $10 \text{ m}^3/\text{s}$  laitoksen C virtauksesta tulee jostain muualta. Sivujoki B1 selittää aika pienen osan laitoksen C virtauksesta.

Testataan seuraavaksi erilaisia ARX-malleja. Ensin rakensimme mallin ilman sivujoen B1 huomioimista. Tällöin tulokseksi saatiin malli  $\text{Arx}(3, [1 \ 3], [4 \ 0])$ . 24 tunnin päähän tapahtuvan ennusteen selitysvoima 54,52 aineiston loppupuoliskossa ja 65,99 aineiston alkupuoliskossa, josta malli myös estimoitii. Mallin ennuste ja ennustevirheet näkyvät kuvassa 32 ja mallin ennustevirheiden autokorrelaatiot näkyvät kuvassa 36. Mallissa ohjaukset olivat järjestyksessä: B1 = laitoksen B juoksutus, B2 = laitoksen C juoksutus. Muuttujien kertoimet olivat

$$A(z) = 1 - 1.323z^{-1} + 0.1962z^{-2} + 0.1264z^{-3}, \quad (23)$$

$$B_1(z) = 0.0002849z^{-4} \quad (24)$$

ja

$$B_2(z) = -0.0006377 - 0.0003395z^{-1} + 0.0007039z^{-2}. \quad (25)$$

Seuraavaksi rakennettiin malli, jossa oli mukana sivujoki B1. Tämän tuloksena saimme mallin  $\text{ARX}(5, [1 \ 4 \ 2], [2 \ 0 \ 4])$ . 24 tunnin päähän tapahtuvan ennusteen selitysvoima 54,43 aineiston loppupuoliskossa ja 71,08 aineiston alkupuoliskossa, mistä malli myös estimoitii. Tässä aineistossa sivujoen B1 huomioiminen ei kovin paljoa auttanut mallissa. Tätä saattoi jo epäillä kuvan 2 perusteella, kun sivujoen B1 arvot eivät vaihtelee juurikaan. Mallin ennuste ja ennustevirheet näkyvät kuvassa 33 ja mallin ennustevirheiden autokorrelaatiot näkyvät kuvassa 35. Mallin ohjaukset olivat muuten samat kuin edellisessäkin mallissa, mutta niihin lisättiin sivujoen virtaama muuttujaksi B3.

Muuttujien kertoimiksi saatiin

$$A(z) = 1 - 1.184z^{-1} - 0.05537z^{-2} + 0.2504z^{-3} - 0.08557z^{-4} + 0.07448z^{-5}, \quad (26)$$

$$B_1(z) = 0.0002422z^{-2}, \quad (27)$$

$$B_2(z) = -0.0006335 - 0.0004299z^{-1} + 0.0006928z^{-2} + 0.0001421z^{-3} \quad (28)$$

ja

$$B_3(z) = -0.0001595z^{-4} + 0.0002533z^{-5}. \quad (29)$$

Edellisen mallin selitysvoima oli edelleen huono, joten annoimme mallille enemmän kertoimia. Näin saatiin malli ARX(8,[3 8 6],[0 0 0]). Tässä mallissa on jo jonkin verran parempi selitysvoima, mutta kertoimia on jo liikaa. 24 tunnin päähän tapahtuvan ennusteen selitysvoima 65,25 aineiston loppupuoliskossa ja 70,96 aineiston alkupuoliskossa, mistä malli myös estimoitii. Mallin ennuste ja ennustevirheet näkyvät kuvassa 34. Ohjaukset olivat samat kuin edellisessäkin mallissa ja muuttujien kertoimiksi saatiin

$$A(z) = 1 - 1.133z^{-1} - 0.07012z^{-2} + 0.2105z^{-3} - 0.05081z^{-4} - 0.05099z^{-5} + 0.07144z^{-6} + 0.05992z^{-7} - 0.03682z^{-8}, \quad (30)$$

$$B_1(z) = -0.0001472 + 5.92 \cdot 10^{-5}z^{-1} + 0.0003492z^{-2}, \quad (31)$$

$$B_2(z) = -0.0006311 - 0.0004696z^{-1} + 0.0006493z^{-2} + 0.0001477z^{-3} + 1.729 \cdot 10^{-6}z^{-4} + 5.737 \cdot 10^{-5}z^{-5} + 7.455 \cdot 10^{-5}z^{-6} - 7.561 \cdot 10^{-5}z^{-7} \quad (32)$$

ja

$$B_3(z) = 0.0001733 - 0.0001144z^{-1} - 0.0001333z^{-2} + 0.0001654z^{-3} - 0.0001447z^{-4} + 0.000211z^{-5}. \quad (33)$$

Mikä malleista siis olisi paras? Ensimmäinen malli, ARX(3,[1 3],[4 0]), vaikuttaa parhaalta, mutta siinä ei millään tavalla huomioida sivujokea B1. Se toimii, kunhan sivujoen virtaus pysyy vakiona. Malli ARX(5,[1 4 2],[2 0 4]) on paras kun sivujoki huomioidaan, mutta malli ei välttämättä ole luotettava jos sivujoen B1 arvot vaihtelevat, sillä se on estimoitu aineistosta jossa sivujoki pysyy vakiona. Tämän perusteella malli ARX(3,[1 3],[4 0]) on suositeltavin.

### 5.2.2 Vesivoimalan D pinnankorkeuden ennustaminen

Mallin rakentamiseen liittyvät kuvat löytyvät liitteestä F. Tutustutaan ensin aineistoon. Kuvasta 37 näkyy, että virtaukset käyttäytyvät koko ajanjakson hyvin samantapaisesti. Kuvasta 38 erot näkyvät tarkemmin. Molempien laitosten juoksutukset vaihtelevat voimakkaasti vuorokauden sisällä, mutta vaihtelu on hyvin samanlaista molemmilla laitoksilla. Pääsääntöisesti laitoksen D virtaus seuraa laitoksen C juoksutusta pienellä viiveellä. Pinnankorkeuksissa tai virtauksissa ei vaikuttaisi olevan nousevaa tai laskevaa trendiä. Pinnankorkeus noin tunnin 850 jälkeen poikkeaa merkittävästi aiemmista, joten jätämme sen pois analyysistä. Keskimääräiset juoksutukset ovat lähes yhtä suuret molemmilla laitoksilla: laitoksella C se oli  $253 \text{ m}^3/\text{s}$  ja laitoksella D  $242 \text{ m}^3/\text{s}$ . Tästä voidaan päätellä, että lähes kaikki vesi laitokselle D tulee laitokselta C.

Testataan erilaisia malleja. Ensimmäinen testattu ARX-malli oli  $\text{ARX}(1, [2 \ 5], [1 \ 0])$ , jonka 24 tunnin päähän tapahtuvan ennusteen selitysvoima 78,13 aineiston loppupuoliskossa ja 79,96 aineiston alkupuoliskossa, mistä malli myös estimoitiin. Mallin ennuste ja ennustevirheet näkyvät kuvassa 40. Ohjaukset järjestyksessä: B1 = laitoksen C juoksutus, B2 = laitoksen D juoksutus.

Diskreetti aikainen ARX-malli on muotoa  $A(z)y(t) = B(z)u(t) + e(t)$ , jossa

$$A(z) = 1 - z^{-1}, \quad (34)$$

$$B_1(z) = 0.0004312z^{-1} + 0.0002912z^{-2} \quad (35)$$

ja

$$B_2(z) = -0.0005653 - 0.0003483z^{-1} + 7.867 \cdot 10^{-5}z^{-2} + 2.559 \cdot 10^{-5}z^{-3} + 1.456e - 05z^{-4}. \quad (36)$$

$\text{ARX}(1, [2 \ 5], [1 \ 0])$  mallissa on turhan paljon viipeitä laitoksen D juoksutuksella, joten estimoimme myös mallin, missä parametri  $n_{b2}$  on pienempi. Ei tunnu uskottavalta että laitokselta pois lähtevä vesi vaikuttaa monen tunnin viipeellä veden pinnankorkeuteen laitoksen yläjuoksulla. Tuloksena saimme mallin  $\text{ARX}(3, [6 \ 2], [1 \ 0])$ . 24 tunnin päähän tapahtuvan ennusteen sovite 80,71 aineiston loppupuoliskossa ja 80,39 aineiston alkupuoliskossa, mistä malli myös estimoitiin. Mallin ennuste ja ennustevirheet näkyvät kuvassa 41. Ohjaukset järjestyksessä: B1 = laitoksen C juoksutus, B2 = laitoksen D juoksutus.

Mallin kertoimet olivat

$$A(z) = 1 - 0.9603z^{-1} + 0.05183z^{-2} - 0.09183z^{-3}, \quad (37)$$

$$B_1(z) = 0.0004221z^{-1} + 0.0003125z^{-2} + 5.03 \cdot 10^{-5}z^{-3} + 4.348 \cdot 10^{-5}z^{-4} - 4.41 \cdot 10^{-8}z^{-5} + 1.883 \cdot 10^{-5}z^{-6} \quad (38)$$

ja

$$B_2(z) = -0.0005623 - 0.0003592z^{-1}. \quad (39)$$

Jatkokehitimme molempia malleja vielä niin että poistimme niistä vääränmerkkiset muuttujien kertoimet. Laitoksen C juoksutusten pitäisi näkyä positiivisella kertoimella pinnankorkeutena laitoksella D ja laitoksen D juoksutusten pitäisi näkyä negatiivisella kertoimella laitoksen D pinnankorkeudessa. Näistä malleista parempi oli ARX(1,[2 2],[1 0]). ARX(1,[2 2],[1 0]), 24 tunnin päähän tapahtuvan ennusteen selitysvaivoima 68,72 aineiston loppupuoliskossa ja 74,95 aineiston alkupuoliskossa, mistä malli myös estimoitiin. Mallin ennuste ja ennustevirheet näkyvät kuvassa 42. Kuvassa 43 näkyy mallin ennustevirheiden autokorrelaatiot. ARX(1,[2 2],[1 0]) on yksinkertaisin malli, mutta se ei merkittävästi ennusta huonommin kuin muut mallit ja sen ennustevirheiden autokorrelaatiotkin ovat pieniä. Sille mallin kertoimet olivat

$$A(z) = 1 - z^{-1}, \quad (40)$$

$$B_1(z) = 0.0003652z^{-1} + 0.0004102z^{-2} \quad (41)$$

ja

$$B_2(z) = -0.0005736 - 0.0002869z^{-1}. \quad (42)$$

## 5.3 Niukka vesitilanne

Niukan vesitilanteen aikana vettä on vähän. Käytetty ajanjakso on 21.10.-19.11.2009. Vaikka vettä on vähän, täytyy juoksutusten ja pinnankorkeuksien pysyä tietyissä rajoissa. Tällöin säädön ja tuotannon optimoinnin mahdollisuudet ovat aika pienet. Toisaalta kun vettä on vähän, sähkön hinta saattaa olla korkea ja tällöin on tärkeää optimoida tuotanto mahdollisimman hyvin.

### 5.3.1 Vesivoimalan C pinnankorkeuden ennustaminen

Aloitetaan tutustumalla aineistoon. Kyseisen ajanjakson mallinnukseen liittyvät kuvat löytyvät liitteestä G. Kuvassa 44 näkyy selvästi vuorokauden mittaista kausivaihtelua. Vaihtelu vuorokauden sisällä on suurimmillaan lähes yhtä suurta kuin suurin vaihtelu koko ajanjakson sisällä. Pinnankorkeuksien käyttäytymisessä ei ole merkittävää eroa ajanjakson aikana. Tämä mahdollistaa esimerkiksi aineiston jakamisen kahtia ja alkupuolella estimoidulla aineistolla pitäisi pystyä ennustamaan loppupuoliskon käyttäytyminen.

Kuvassa 45 näkyy virtaukset laitoksilla B ja C ja sivujoessa B1 niukan vesitilanteen aikana. Sen perusteella virtaukset laitoksilla B ja C ovat keskimäärin yhtä suuria ja sivujoen B1 virtaus merkittävästi pienempi. Laitoksen B virtaama ei vaihtelee vuorokauden sisällä, kun taas laitoksen C virtaama vaihtelee voimakkaasti ja sivujoen B1 hieman vuorokauden sisällä. Aineiston lopussa laitoksen C ja sivujoen B1 juoksutukset kasvavat merkittävästi suuremmiksi kuin aikavälillä muuten. Se saattaa aiheuttaa ongelmia. Keskimääräiset virtaamat aikavälillä olivat  $49,03 \text{ m}^3/\text{s}$  laitoksella B,  $61,2 \text{ m}^3/\text{s}$  laitoksella C ja sivujoen virtaama oli  $5,6 \text{ m}^3/\text{s}$ . Tästä voidaan havaita, että laitoksen C virtaus on keskimäärin hieman suurempi kuin sivujoen B1 ja laitoksen B virtaukset yhteensä. Sivujoen B1 virtaus vaikuttaisi selittävän siis noin puolet kaikkien sivujokien virtauksista.

Testataan seuraavaksi erilaisia ARX-malleja. Sekä pinnankorkeus, että kaikki virtaukset vaikuttavat stationaarisiltä kuvissa 44 ja 45. Tästä syystä emme näy tarvetta aineiston stationarisoinnille, vaan käytämme aineistoa sellaisenaan mallin muodostamiseen. Mallin muodostus ja simulointi on myös helpompaa ilman stationarisointia. Mallit muodostettiin jakamalla aineisto puoliksi ja alkuosalla estimoitiin malli ja loppuosalla katsottiin kuinka hyvin malli ennustaa.

Ensimmäinen malli rakennettiin niin, ettei siinä käytetty hyväksi sivujokea B1. Tällä tavalla saatiin estimoitua malli  $\text{ARX}(6, [1 \ 3], [0 \ 0])$ . 24 tunnin päähän tapahtuvan ennusteen selitysvaivoima 45,18 aineiston alkupuoliskossa ja 54,86 aineiston alkupuoliskossa, mistä malli myös estimoitiin. Mallin ennuste ja ennustevirheet näkyvät kuvassa 46. Käytetyt ohjaukset järjestyksessä olivat  $B_1 =$  laitoksen B juoksutus ja  $B_2 =$  laitoksen C juoksutus. Muuttujien kertoimiksi saatiin

$$A(z) = 1 - 0.8289z^{-1} - 0.2998z^{-2} + 0.2703z^{-3} - 0.1783z^{-4} - 0.1115z^{-5} + 0.1482z^{-6}, \quad (43)$$

$$B_1(z) = 0.0003435 \quad (44)$$

ja

$$B_2(z) = -0.0005791 + 7.359 \cdot 10^{-5}z^{-1} + 0.0001951z^{-2}. \quad (45)$$

Seuraavaan malliin otettiin mukaan sivujoki B1. Käytössä olevassa aineistossa myös sivujoessa on päivittäistä vaihtelua mikä tukee sitä että sen lisäämisestä voi olla oikeasti hyötyä. Tällä tavalla saatiin malli  $\text{ARX}(6, [2 \ 3 \ 1], [0 \ 0 \ 6])$ . 24 tunnin päähän tapahtuvan ennusteen selitysvaivoima 63,85 aineiston alkupuoliskossa ja 83,03 aineiston alkupuoliskossa, mistä malli myös estimoitiin. Mallin ennuste ja ennustevirheet näkyvät kuvassa 47. Ohjaukset olivat

järjestyksessä: B1 = laitoksen B juoksutus, B2 = laitoksen C juoksutus, B3 = sivujoen B1 virtaus. Muuttujien kertoimet olivat

$$A(z) = 1 - 0.7505z^{-1} - 0.2748z^{-2} + 0.2227z^{-3} - 0.1906z^{-4} - 0.1203z^{-5} + 0.1135z^{-6}, \quad (46)$$

$$B_1(z) = -0.0005029 + 0.0009314z^{-1}, \quad (47)$$

$$B_2(z) = -0.0005771 + 3.08 \cdot 10^{-5}z^{-1} + 0.0001445z^{-2} \quad (48)$$

ja

$$B_3(z) = 0.0005584z^{-6}. \quad (49)$$

Kolmas malli on kehitellympi versio edellisestä mallista. Siitä poistettiin vääränmerkkiset muuttujien kertoimet. Laitoksen C juoksutuksella pitäisi olla negatiiviset kertoimet, laitoksen B juoksutuksella positiiviset ja sivujoella B1 positiiviset kertoimet. Näin saimme mallin ARX(5,[1 1 1],[1 0 6]). 24 tunnin päähän tapahtuvan ennusteen selitysvaivoima 60,73 aineiston alkupuoliskossa ja 72,05 aineiston alkupuoliskossa, mistä malli myös estimoitui. Mallin ennuste ja ennustevirheet näkyvät kuvassa 48 ja ennustevirheiden autokorrelaatiot kuvassa 49. Ohjaukset olivat samat kuin edellisessäkin mallissa. Muuttujien kertoimiksi saatiin

$$A(z) = 1 - 0.7051z^{-1} - 0.1031z^{-2} + 0.105z^{-3} - 0.2553z^{-4} - 0.0414z^{-5}, \quad (50)$$

$$B1(z) = 0.0005992z^{-1} \quad (51)$$

$$B2(z) = -0.0005629 \quad (52)$$

$$B3(z) = 0.0006406z^{-6} \quad (53)$$

Kaikki mallit ennustivat pääsääntöisesti hyvin, mutta lopussa kaikilla tulee verrattain iso virhe. Kyseisessä kohtaa näkyy poikkeama muusta aineistosta jo alkuperäisessä aineistossa. Mallit toimivat vain, jos tilanne ei muutu merkittävästi. Kolmas malli, ARX(5,[1 1 1],[1 0 6]), vaikuttaa parhaalta, sillä siinä on vähemmän parametreja kuin mallissa ARX(6,[2 3 1],[0 0 6]), mutta selitysvaivoima on lähes yhtä hyvä. Lisäksi sivujoesta B1 tuntuu olevan merkittävää hyötyä joten se on hyvä pitää mukana. Kuvasta 48 kuitenkin näkee, että kolmannen mallin ennustevirheet ovat autokorreloituneita. Loppuosan virhe voisi selittää osan autokorreloituneisuudesta.

### 5.3.2 Vesivoimalan D pinnankorkeuden ennustaminen

Tämän tapauksen mallinvalintaan liittyvät kuvat löytyvät liitteestä H. Tutustutaan ensin aineistoon. Kuvasta 50 näkyy että virtaukset käyttäytyvät koko ajanjakson hyvin samantapaisesti. Kuvasta 51 erot näkyvät tarkemmin. Molempien laitosten juoksutukset vaihtelevat voimakkaasti vuorokauden sisällä, mutta vaihtelu on hyvin samanlaista molemmilla laitoksilla. Pääsääntöisesti laitoksen D virtaus seuraa laitoksen C juoksutusta pienellä viiveellä. Pinnankorkeuksissa tai virtauksissa ei vaikuttaisi olevan nousevaa tai laskevaa trendiä.

Keskimääräiset juoksutukset ovat lähes yhtä suuret molemmilla laitoksilla: laitoksella C virtaama oli  $61,2 m^3/s$  ja laitoksella D  $61,6 m^3/s$ , joten lähes kaikki vesi laitokselle D tulee laitokselta C.

Testataan seuraavaksi erilaisia malleja. Ensimmäinen malli oli  $ARX(4, [2 \ 2], [1 \ 0])$ , jonka 24 tunnin päähän tapahtuvan ennusteen selitysvaivoima 33,17 aineiston loppupuoliskolla. Mallin ennuste ja ennustevirheet näkyvät kuvassa 53. Ohjaukset mallissa olivat B1 = laitoksen C juoksutus, B2 = laitoksen D juoksutus. Muuttujien kertoimiksi saatiin

$$A(z) = 1 - 0.8017z^{-1} - 0.344z^{-2} + 0.2424z^{-3} - 0.09686z^{-4}, \quad (54)$$

$$B1(z) = 0.0008578z^{-1} - 0.0001793z^{-2} \quad (55)$$

ja

$$B2(z) = -0.0004822 - 0.000219z^{-1}. \quad (56)$$

Malli ennustaa muuten ihan hyvin, mutta aikavälillä 75-125  $h$  on suuri poikkeama. Sama poikkeama näkyy alkuperäisessä aineistossa noin 500  $h$ :n kohdalla. Alkuperäisessä aineistossa näkyy lähes vastaavankaltaista poikkeamaa viikon välein, mitä malli ei siis ennusta. Tähän voisi auttaa kausivaihtelun huomioiminen mallissa.

Ensimmäisessä mallissa oli vääränmerkkisiä kertoimia jotka on poistettu tässä seuraavassa mallissa. Laitoksen D juoksutusten kertoimien pitäisi olla negatiivisia ja laitoksen C juoksutusten pitäisi olla positiivisia kun ennustetaan pinnankorkeutta laitoksella D, näin saatiin malli  $ARX(4, [1 \ 2], [1 \ 0])$ . 24 tunnin päähän tapahtuvan ennusteen selitysvaivoima 31,07 aineiston loppupuoliskolla. Mallin ennuste ja ennustevirheet näkyvät kuvassa 54. Ennustevirheiden autokorrelaatiot ovat kuvassa 55. Ohjaukset olivat samat kuin ensimmäisessäkin mallissa ja muuttujien kertoimiksi saatiin

$$A(z) = 1 - 0.6551z^{-1} - 0.4537z^{-2} + 0.1811z^{-3} - 0.07237z^{-4}, \quad (57)$$



$$B_1(z) = 0.0008234z^{-1} \quad (58)$$

ja

$$B_2(z) = -0.0004426 - 0.0003913z^{-1}. \quad (59)$$

Tässä mallissa ei ole negatiivista arvoa ylemmän laitoksen juoksutukselle, mikä tuntuu järkevältä. Malli ennustaa muuten ihan hyvin, mutta aikavälillä 75-125  $h$  on suuri poikkeama. Sama poikkeama näkyy alkuperäisessä aineistossa noin 500  $h$  kohdalla. Alkuperäisessä aineistossa näkyy lähes vastaavankaltaista poikkeamaa viikon välein, mitä malli ei siis ennusta. Tähän voisi auttaa kausivaihtelun huomioiminen mallissa.

Molemmat mallit ovat huonoja selitysvoinan näkökulmasta, toisaalta poikkeamat kaikkialla muualla paitsi 75-125  $h$  kohdalla ovat pieniä, eikä poikkeama kyseisessä kohdassakaan ole merkittävän iso. Malli ARX(4,[1 2],[1 0]) on parempi, koska siinä on vähemmän parametreja ja kaikki kertoimet ovat järkevän merkkisiä. Kuvan 50 virtaamissa ei näy merkittävää eroa samassa kohdassa, kuin pinta-aineistossa. Poikkeama mikä näkyy sekä ennustetuissa pinnankorkeuksissa, että alkuperäisessä aineistossa noin kohdalla 500  $h$ , ei vaikuttaisi löytyvän virtaamista kuvassa 50. Tästä syystä aineistossa saattaa olla virhe. Toisaalta ero voi olla niin pieni että sitä ei huomaa virtaamista.

## 6 Yhteenveto

Tässä projektissa tutkittiin mahdollisuuksia mallintaa joen virtausta ARX-aikasarjamallien avulla. Tärkeimpänä tuloksena huomasimme, että joen mallinnukseen ei riitä yksi lineaarinen aikasarjamalli, vaan eri laitosten välit tarvitsevat erilaiset mallit ja lisäksi sama väli tarvitsee erilaisen mallin riippuen vesitilanteesta ja juoksutusstrategiasta. Taulukosta 1 näkyvät rakennetut mallit ja mallien selitysvoimat kaikille tilanteille. Siitä havaitsee, että sekä parametrit, että selitysvoimat vaihtelevat merkittävästi eri tapauksissa. Usein huonoissa malleissa oli itse estimointiaineistossa jotakin epämääräistä ja paremmat mallit edellyttivät että ainakin juoksutusstrategia pysyy muuttumattomana. Taulukosta 1 nähdään myös, että profiliajolle onnistuttiin löytämään pieniparametriset mallit, kun taas niukka vesitilanne ja jääkansi -tilanteet edellyttivät suurempia parametreja. Profiliajon malleja rakennettiin ehkä hieman pidempään, mikä saattaa selittää sen, että löydettiin pieniparametriset mallit.

UPM oli erityisen kiinnostunut laitosten B ja C välistä. Siltä väliltä saatiin ristiriitaisia tuloksia sivujoen B1 vaikutuksesta. Niukan vesitilanteen aikana se paransi mallia merkittävästi, mutta muissa tilanteissa siitä ei ollut hyötyä. Niukka vesitilanne oli myös ainoa tilanne, jossa sivujoessa B1 oli päivän sisäistä vaihtelua. Tuntuu erikoiselta, että joessa ei olisi päivän sisäistä vaihtelua jollain aikavälillä jos siinä kuitenkin on vaihtelua jollain toisella aikavälillä. Jos aineistossa on virhe ja sivujoessa B1 on oikeasti myös päivän sisäistä vaihtelua, pystytään sen avulla parantamaan malleja laitosten B ja C välille.

Mahdollisia virheitä on voinut aiheuttaa virheellinen juoksutusaineisto. Juok-

Taulukko 1: Koonti luotujen ARX-mallien parametreista eri joenväleille ja valituille ajanjaksoille.

Aikaväli	Voimala	Parametrit	Selitysvoima
Tasainen profiliajo	B	(1,2,1)	93,36
	C	(1,[2 1 0],[1 1 0])	70,66
	C	(1,[1 1 0],[1 0 0])	65,2
	D	(1,[2 1],[1 1])	64,12
Niukka vesitilanne	C	(5,[1 1 1],[1 0 6])	60,73
	D	(4,[1 2],[1 0])	33,07
Jääkansi	C	(3,[1 3 0],[4 0 0])	54,52
	D	(1,[2 2],[1 0])	68,72

sutusten arvot on estimoitu sähköntuotannon perusteella ja siinä tiedetään olevan jonkin verran virhettä. Tämä tietenkin aiheuttaa virhettä mallin toiminnassa. Lisäksi aikasarjoja ei esikäsitelty juurikaan ennen mallinnusta, vaan oletettiin aikasarjat stationaarisiksi. Koska mallinnus tehtiin erikseen erilaisille ajanjaksoille, jotka kuitenkin olivat itsessään varsin homogeenisia, yksinkertaistuksien vaikutus ei ehkä ole niin suuri.

Aikasarjamalleista tarkasteltiin pelkästään ARX ja hiukan ARMAX-malleja, mutta muitakin olisi voinut kokeilla. Esimerkiksi kausivaihtelua sisältävät mallit, kuten SARX ja SARMAX olisivat voineet sopia tutkittuihin tilanteisiin. Erityisesti tasaisen profiliajon tapauksessa aikasarjaan sisältyi selkeä vuorokaudenaikojen mukainen kuvio, jonka kausivaihtelu olisi voinut selittää. Myös muiden lineaaristen mallien tarkempi tutkiminen olisi voinut tuottaa hyviä tuloksia. Lisäksi kirjallisuuskatsauksessa esille nousseet fyysiset mallit olisivat olleet yksi vaihtoehto, mutta huomattavasti monimutkaisempi.

Saadut mallit on estimoitu yhdeltä tietyltä ajalta, eikä niiden toimintakyvystä ole takeita muina aikoina. On todennäköistä, että mallien parametrit ja kertoimet täytyy päivittää, kun systeemin olosuhteet muuttuvat. Aikaisemmin estimoitua mallia voidaan esimerkiksi käyttää, kunnes sen antamat tulokset huononevat, jolloin luodaan uusi malli ja siirrytään käyttämään sitä.

Jos kehitettyjä aikasarjamalleja halutaan tarkentaa, pitää mallit estimoida sellaisilta ajanjaksoilta, jolloin juoksutukset eivät muutu merkittävästi päivien välillä. Osittain tästä syystä epäillään, että mallit toimivat hyvin, kun juoksutusstrategia pysyy samana. Epäillään myös, että mallit toimivat sitä huonommin, mitä enemmän juoksutusstrategiaa muutetaan. Kun malleja käytetään optimoinnissa täytyy huomioida, että ennustetarkkuus huononee, mitä enemmän juoksutusstrategia muuttuu.

Malleja tarkennettaessa täytyy kiinnittää enemmän huomiota myös stationaarisuuteen ja kausivaihteluun. Viikonloput ja arkipäivät erottuvat usein hyvin selkeästi toisistaan, jolloin aineisto ei ole stationaarista. Malleihin voisi ehkä saada parannusta jaksollisuuden huomioon ottavilla malleilla kuten SARMAX-aikasarjamallilla.

Joka tapauksessa aikasarjamallien rajat tulevat pian vastaan, eikä niistä saada välttämättä kovin paljon parempia malleja. Seuraava askel on siirtyä grey box malleihin, ottaa osia fyysisistä malleista mukaan tilastollisiin malleihin. Tällaisista malleista voi saada esimerkiksi viitteitä siitä minkälaisia epälineaarisia riippuvuuksia joesta löytyy. Kun riippuvuuden muoto on tiedossa, voidaan niiden ympärille kehittää tilastollisia malleja, lopputuloksena grey box -malli.

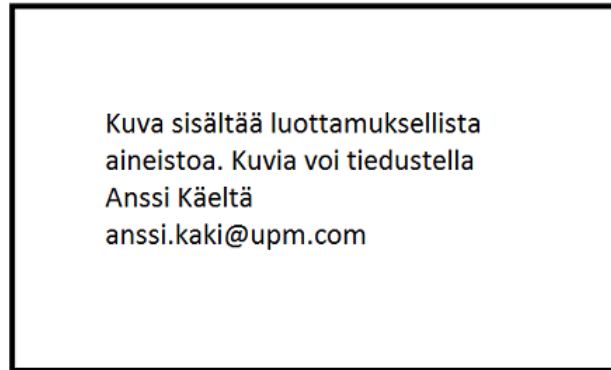
## Viitteet

- [1] Y. Bolea, R. Martinez-Gonzales, A. Grau, ja H. Martinez-Garcia. An LPV fractional model for canal control. *System Identification*, 15:1786–1791, 2009. Haettu 24.4.2014 osoitteesta <http://www.ifac-papersonline.net/Detailed/39842.html>.
- [2] M. Campolo, A. Soldati, ja P. Andreussi. Forecasting river flow rate during low-flow periods using neural networks, 1999. Haettu 24.4.2014 osoitteesta <http://onlinelibrary.wiley.com/doi/10.1029/1999WR900205/abstract>.
- [3] M. F. L. Foo. Modelling and control design of river systems, 2012. Haettu 24.4.2014 osoitteesta [http://dtl.unimelb.edu.au//exlibris/dtl/d3\\_1/apache\\_media/L2V4bGlicmlzL2R0bC9kM18xL2FwYWNoZV9tZWRpYS8yNzc3NzM=.pdf](http://dtl.unimelb.edu.au//exlibris/dtl/d3_1/apache_media/L2V4bGlicmlzL2R0bC9kM18xL2FwYWNoZV9tZWRpYS8yNzc3NzM=.pdf).
- [4] MathWorks. Compare model output and measured output -matlab compare- mathworks nordic. Haettu 12.5.2014 osoitteesta <http://www.mathworks.se/help/ident/ref/compare.html>.
- [5] MathWorks. Estimate parameters of ARX or AR model using least squares -matlab arx- mathworks nordic. Haettu 28.4.2014 osoitteesta <http://www.mathworks.se/help/ident/ref/arx.html>.
- [6] MathWorks. Identifying input-output polynomial models - matlab & simulink- mathworks nordic. Haettu 12.5.2014 osoitteesta <http://www.mathworks.se/help/ident/ug/identifying-input-output-polynomial-models.html>.
- [7] MathWorks. Identifying linear models using system identification tool -matlab & simulink- mathworks nordic. Haettu 12.5.2014 osoitteesta <http://www.mathworks.se/help/ident/gs/identify-linear-models-using-the-gui.html>.
- [8] MathWorks. Validating models after estimation - matlab & simulink- mathworks nordic. Haettu 28.4.2014 osoitteesta <http://www.mathworks.se/help/ident/ug/validating-models-after-estimation.html#bq4f9uo>.
- [9] M. Maxwell ja S. Warnick. Modeling and identification of the sevier river system. *American Control Conference, 2006*, pages 5342–5347, 2006. Haettu 24.4.2014 osoitteesta <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1657572>.

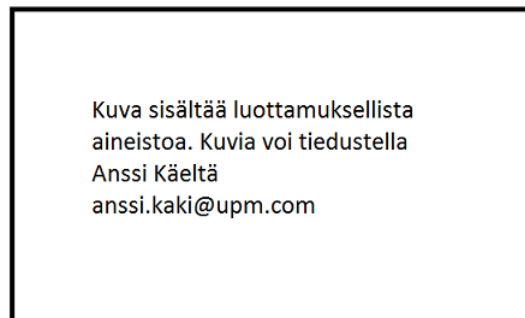
- [10] I. Mellin. Aikasarjat. Haettu 28.4.2014 osoitteesta [https://noppa.aalto.fi/noppa/kurssi/ms-c2128/luennot/MS-C2128\\_aikasarjat.pdf](https://noppa.aalto.fi/noppa/kurssi/ms-c2128/luennot/MS-C2128_aikasarjat.pdf).
- [11] I. Mellin. Arma-mallit. Haettu 28.4.2014 osoitteesta [https://noppa.aalto.fi/noppa/kurssi/mat-2.3128/luennot/Mat-2\\_3128\\_arma-mallit.pdf](https://noppa.aalto.fi/noppa/kurssi/mat-2.3128/luennot/Mat-2_3128_arma-mallit.pdf).
- [12] I. Mellin. Stationaariset stokastiset prosessit. Haettu 28.4.2014 osoitteesta [https://noppa.aalto.fi/noppa/kurssi/ms-c2128/luennot/MS-C2128\\_stationaariset\\_stokastiset\\_prosessit.pdf](https://noppa.aalto.fi/noppa/kurssi/ms-c2128/luennot/MS-C2128_stationaariset_stokastiset_prosessit.pdf).
- [13] R. R. Perez, V. F. Batlle, ja L. S. Rodriguez. Robust system identification of an irrigation main canal. *Advances in Water Resources*, 30:1785–1796, 2007. Haettu 24.4.2014 osoitteesta <http://www.sciencedirect.com/science/article/pii/S0309170807000164>.
- [14] P. Sleigh ja I. M. Goodwill. The st venant equations, 2000. Haettu 24.4.2014 osoitteesta <http://www.efm.leeds.ac.uk/CIVE/CIVE3400/stvenant.pdf>.
- [15] B. Sohlberg ja M. Sernfält. Grey box modelling for river control. *Journal of Hydroinformatics*, 4:265–280, 2002. Haettu osoitteesta <http://www.iwaponline.com/jh/004/0265/0040265.pdf>.
- [16] H. E. C. US Army Corps of Engineers. Hec-ras river analysis system applications guide, 2010. Haettu 12.5.2014 osoitteesta [http://www.hec.usace.army.mil/software/hec-ras/documentation/HEC-RAS\\_4.1\\_Applications\\_Guide.pdf](http://www.hec.usace.army.mil/software/hec-ras/documentation/HEC-RAS_4.1_Applications_Guide.pdf).
- [17] Wikipedia. Aikasarja-analyysi. Haettu 28.4.2014 osoitteesta <http://fi.wikipedia.org/wiki/Aikasarja-analyysi>.
- [18] Wikipedia. Akaike information criterion -wikipedia, the free encyclopedia. Haettu 12.5.2014 osoitteesta [http://en.wikipedia.org/wiki/Akaike\\_information\\_criterion](http://en.wikipedia.org/wiki/Akaike_information_criterion).
- [19] Wikipedia. Mass balance. Haettu 24.4.2014 osoitteesta [http://en.wikipedia.org/wiki/Mass\\_balance](http://en.wikipedia.org/wiki/Mass_balance).
- [20] Wikipedia. Minimum description length -wikipedia, the free encyclopedia. Haettu 28.4.2014 osoitteesta [http://en.wikipedia.org/wiki/Minimum\\_description\\_length](http://en.wikipedia.org/wiki/Minimum_description_length).
- [21] Q. Wu, S. Amin, S. Munier, A. M. Bayen, X. Litrico, ja G. Belaud. Parameter identification for the shallow water equation using modal decom-

- position. *Decision and Control, 2007 46th IEEE Conference on*, pages 1584 – 1590, 2007. Haettu 24.4.2014 osoitteesta <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4434678&tag=1>.
- [22] X. Zhuan, G. Zheng, ja X. Xia. A modelling methodology for natural dam-river network systems. *Control Engineering Practice*, 17:534–540, 2009. Haettu 24.4.2014 osoitteesta <http://www.sciencedirect.com/science/article/pii/S0967066108001810>.

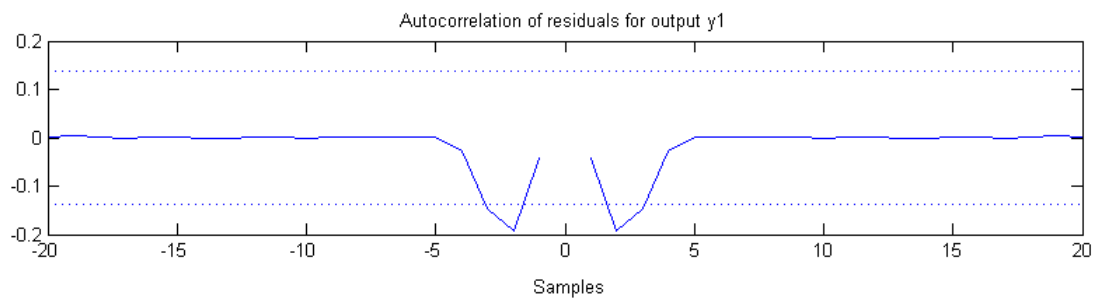
## A Tasainen profiiajo: vesivoimalan B juoksu- tusten ennustaminen



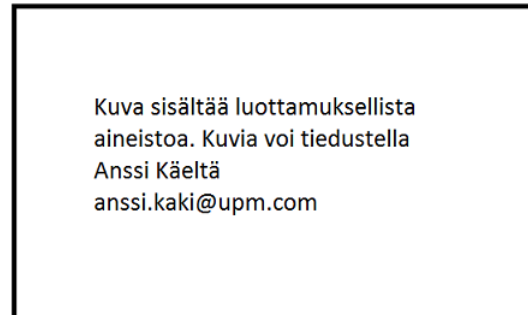
Kuva 5: Tasainen profiiajo välillä AB 24.10.-10.12.2013



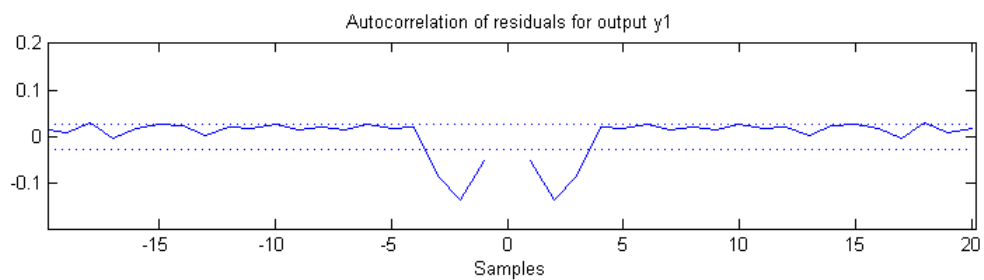
Kuva 6: Mallin ARX(4,5,1) ennuste laitoksen B juoksutukselle ja sen todelliset arvot validointiaineistossa.



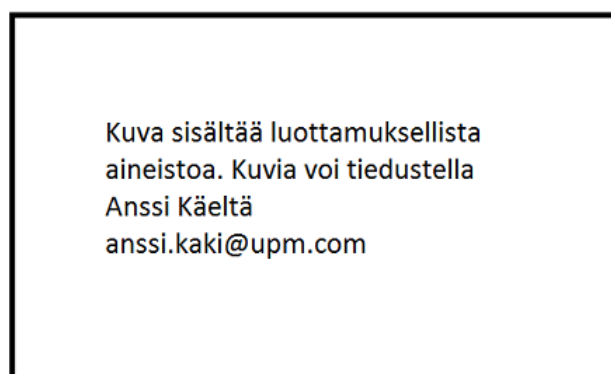
Kuva 7: Mallin ARX(4,5,1) validointiaineiston ennusteen residuaalien auto-korrelaatio.



Kuva 8: Mallin ARX(4,5,1) ennuste koko vuodelle 2013.

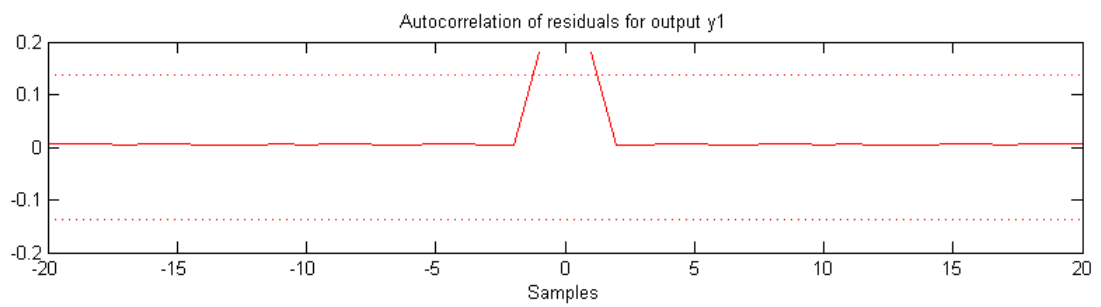


Kuva 9: Mallin ARX(4,5,1) vuoden 2013 residuaalien autokorrelaatio.



Kuva 10: Mallin ARX(1,2,1) ennuste laitoksen B juoksutukselle ja sen todelliset arvot validointiaineistossa.

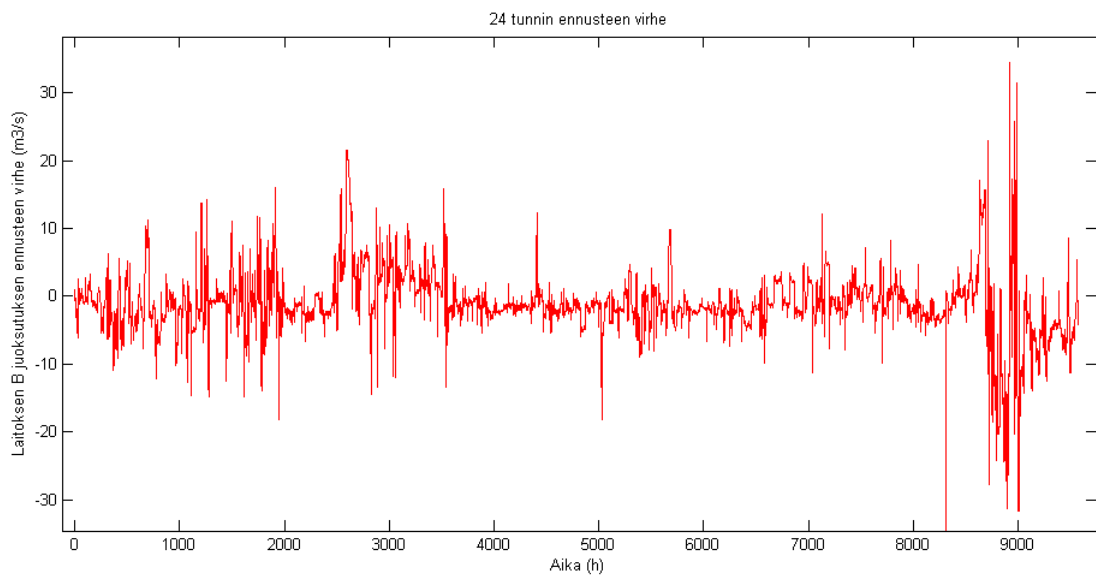




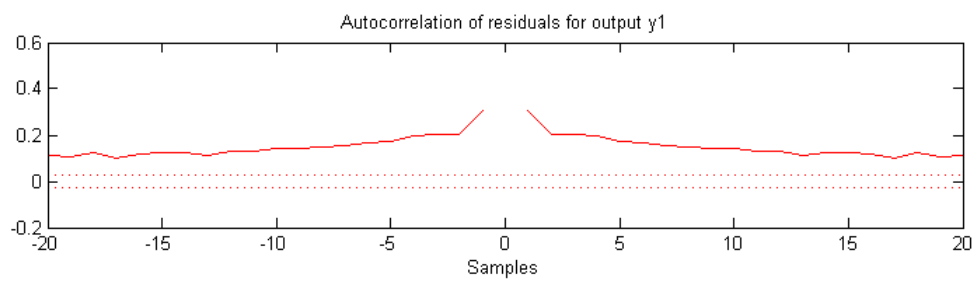
Kuva 11: Mallin ARX(1,2,1) validointiaineiston ennusteen residuaalien auto-korrelaatio.

Kuva sisältää luottamuksellista  
 aineistoa. Kuvia voi tiedustella  
 Anssi Käeltä  
 anssi.kaki@upm.com

Kuva 12: Mallin ARX(1,2,1) ennuste koko vuodelle 2013.

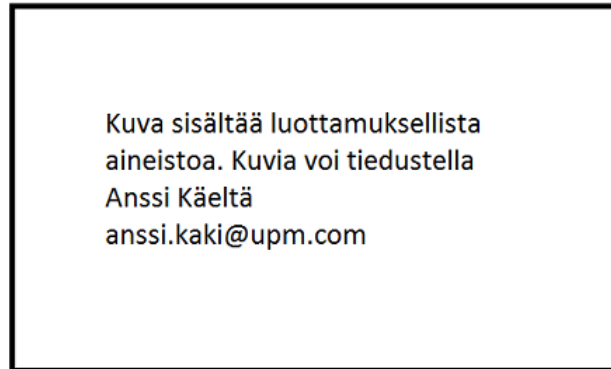


Kuva 13: Mallin ARX(1,2,1) ennusteen residuaalit koko vuodelle 2013.

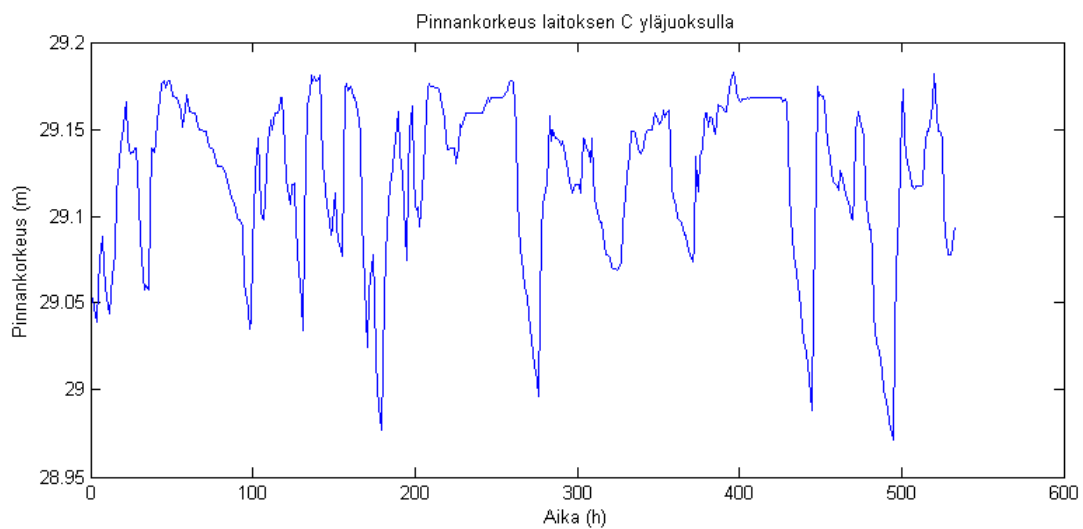


Kuva 14: Mallin ARX(1,2,1) vuoden 2013 residuaalien autokorrelaatio.

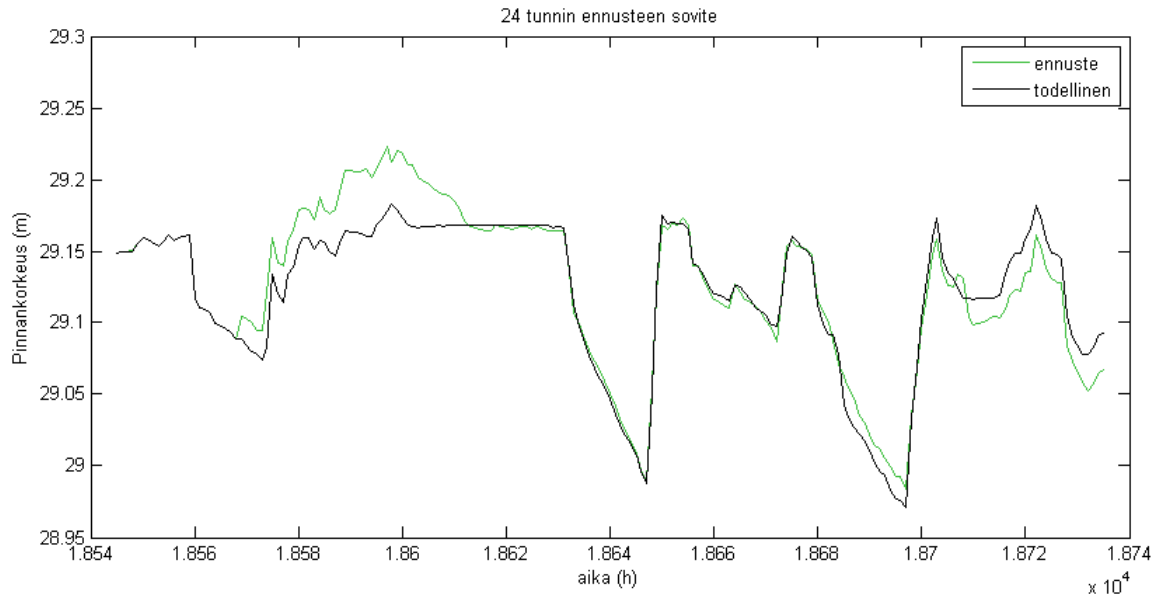
## B Tasainen profiajo: vesivoimalan C pinnan- korkeuden ennustaminen vuosien 2010 ja 2013 aineistoilla



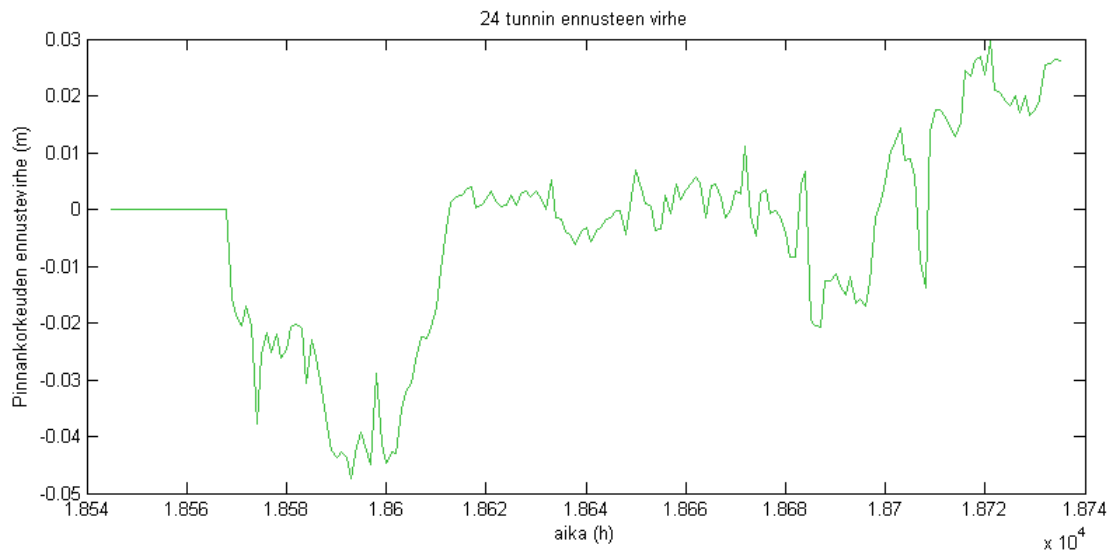
Kuva 15: Selittävien muuttujien juoksutukset ja virtaama 28.01.2010-19.02.2010.



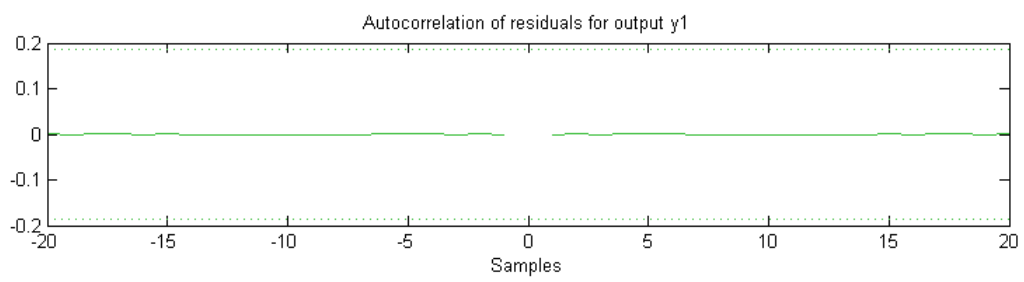
Kuva 16: Laitoksen C vedenpinnan korkeus 28.01.2010-19.02.2010.



Kuva 17: Mallin  $ARX(1,[1 \ 1 \ 0],[1 \ 0 \ 0])$  ennuste laitoksen C vedenpinnan korkeudelle ja sen todelliset arvot validointiaineistossa.



Kuva 18: Mallin  $ARX(1,[1 \ 1 \ 0],[1 \ 0 \ 0])$  residuaalit validointiaineistossa.

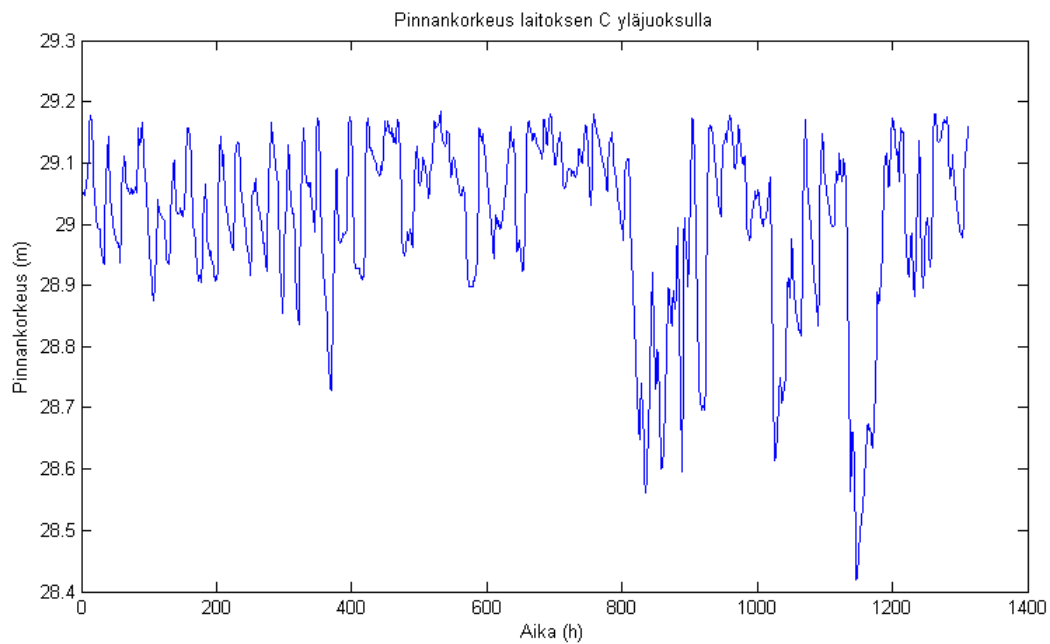


Kuva 19: Mallin  $ARX(1, [1 \ 1 \ 0], [1 \ 0 \ 0])$  validointiaineiston ennusteen residuaalien autokorrelaatio.

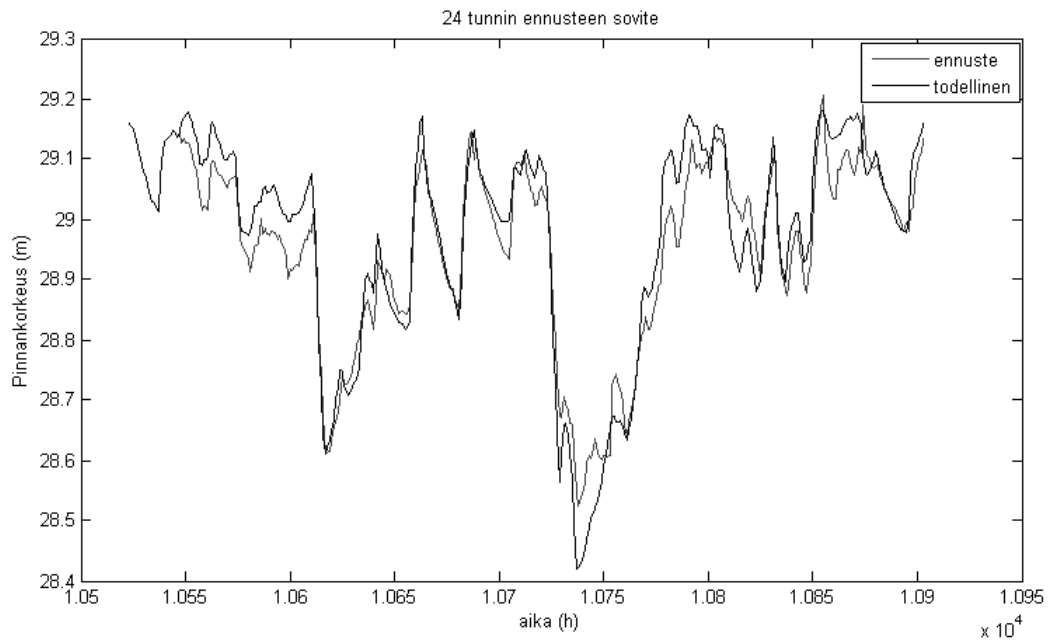
## C Tasainen profiajo: vesivoimalan C pinnan- korkeuden ennustaminen vuoden 2009 aineis- tolla

Kuva sisältää luottamuksellista  
aineistoa. Kuvia voi tiedustella  
Anssi Käeltä  
anssi.kaki@upm.com

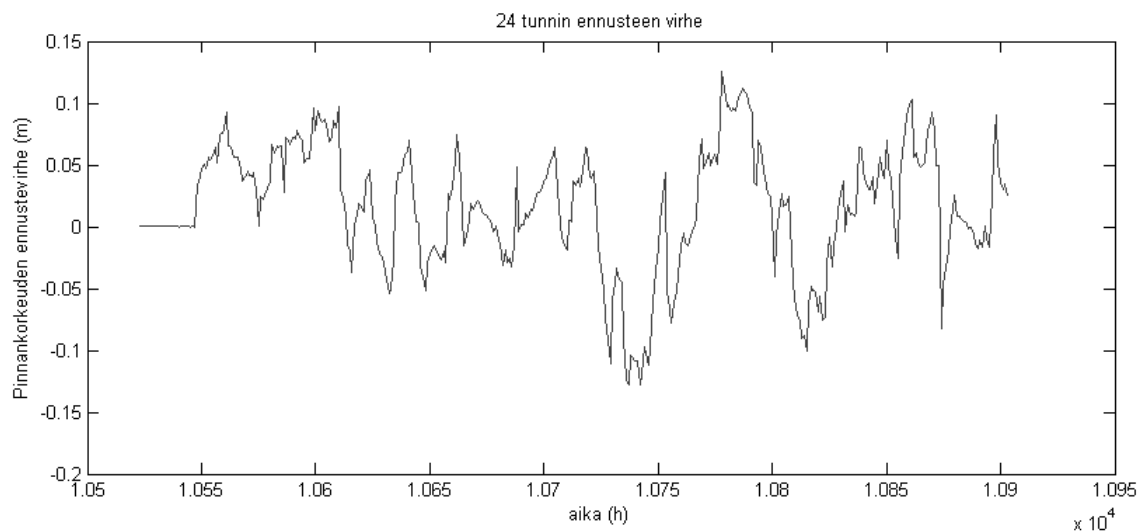
Kuva 20: Selittävien muuttujien juoksutukset ja virtaukset 3.2.2009-30.3.2009.



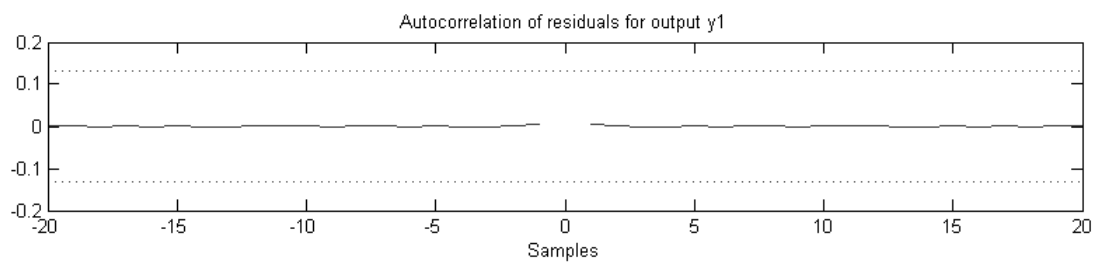
Kuva 21: Laitoksen C vedenpinnan korkeus 3.2.2009-30.3.2009.



Kuva 22: Mallin  $ARX(1,[2 \ 1 \ 1],[1 \ 1 \ 1])$  ennuste laitoksen C juoksutukselle ja sen todelliset arvot validointiaineistossa.



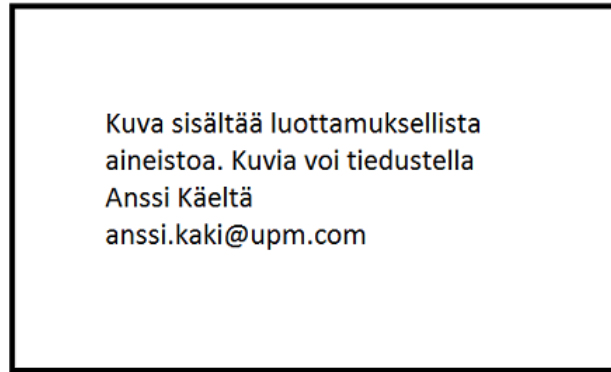
Kuva 23: Mallin  $ARX(1,[2 \ 1 \ 1],[1 \ 1 \ 1])$  residuaalit validointiaineistossa.



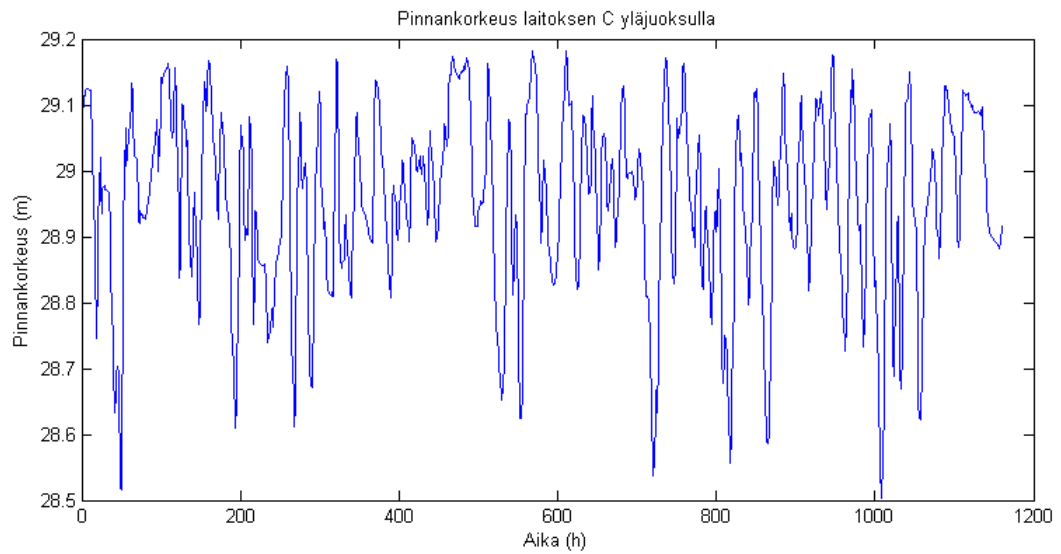
Kuva 24: Mallin ARX  $(1, [2 \ 1 \ 1], [1 \ 1 \ 1])$  validointiaineiston ennusteen residuaalien autokorrelaatio.



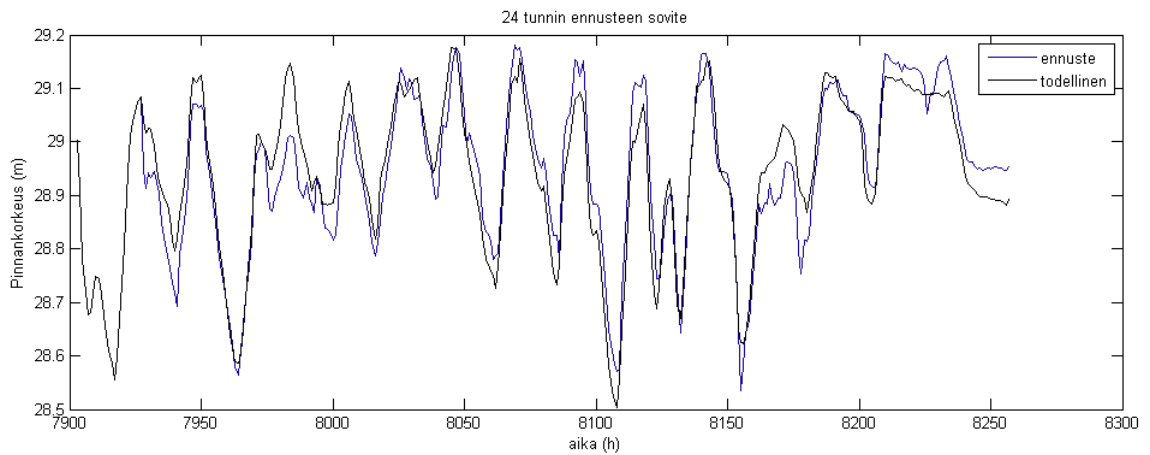
## D Tasainen profiaajo: vesivoimalan D pinnan- korkeuden ennustaminen



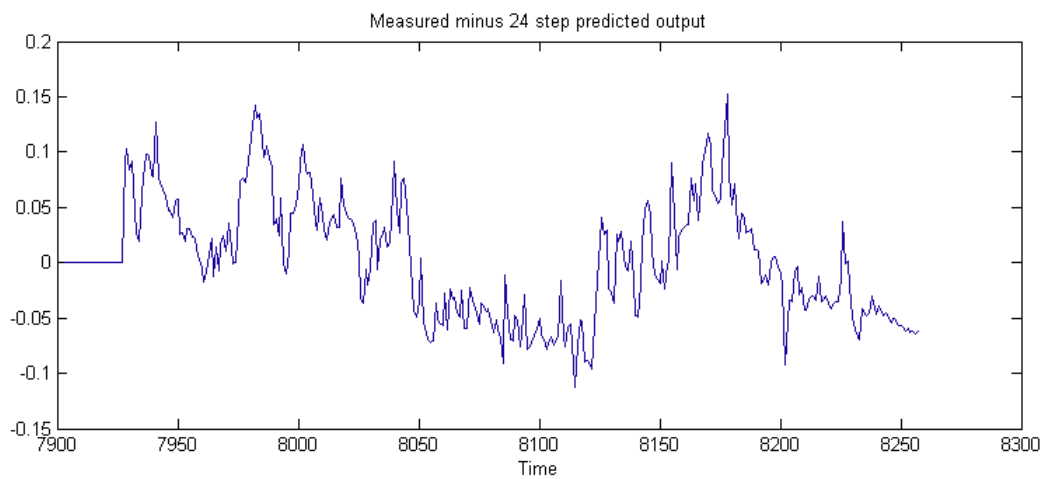
Kuva 25: Laitosten C ja D juoksutukset ajalta 24.10. - 10.12.2013.



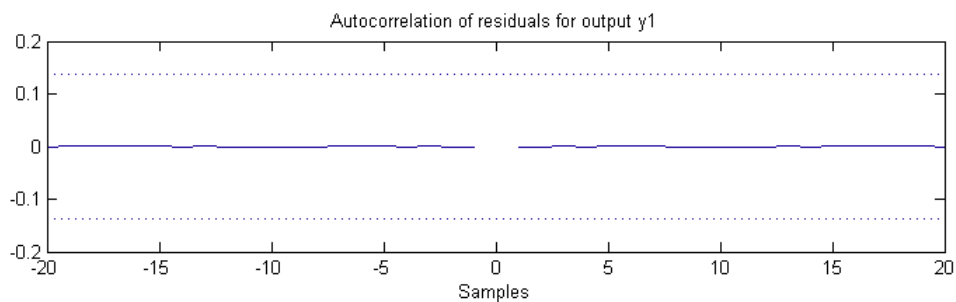
Kuva 26: Laitoksen D vedenpinnankorkeus ajalta 24.10. - 10.12.2013.



Kuva 27: Mallin  $ARX(1,[2 \ 1],[1 \ 1])$  ennuste laitoksen D vedenpinnan korkeudelle validointiaineistolla.

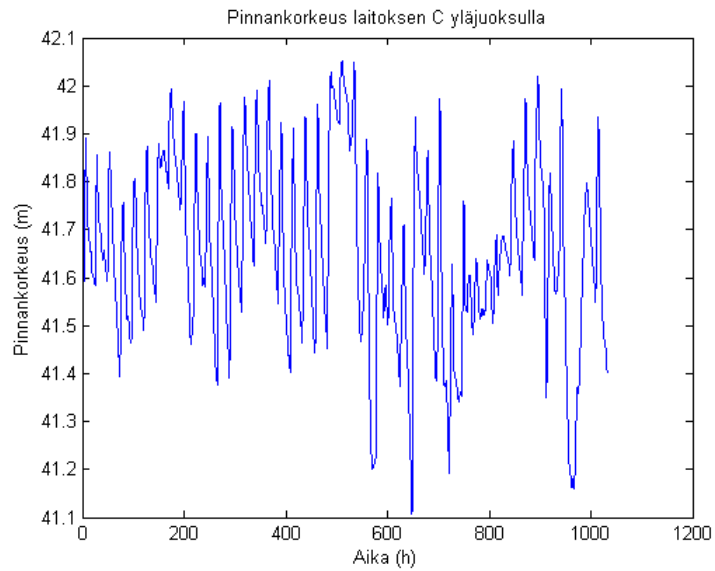


Kuva 28: Mallin  $ARX(1,[2 \ 1],[1 \ 1])$  validointiaineiston sovitetten residuaalit.

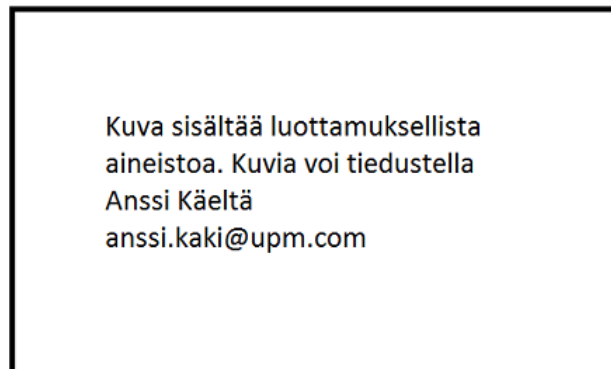


Kuva 29: Mallin  $ARX(1,[2 \ 1],[1 \ 1])$  residuaalien autokorrelaatio.

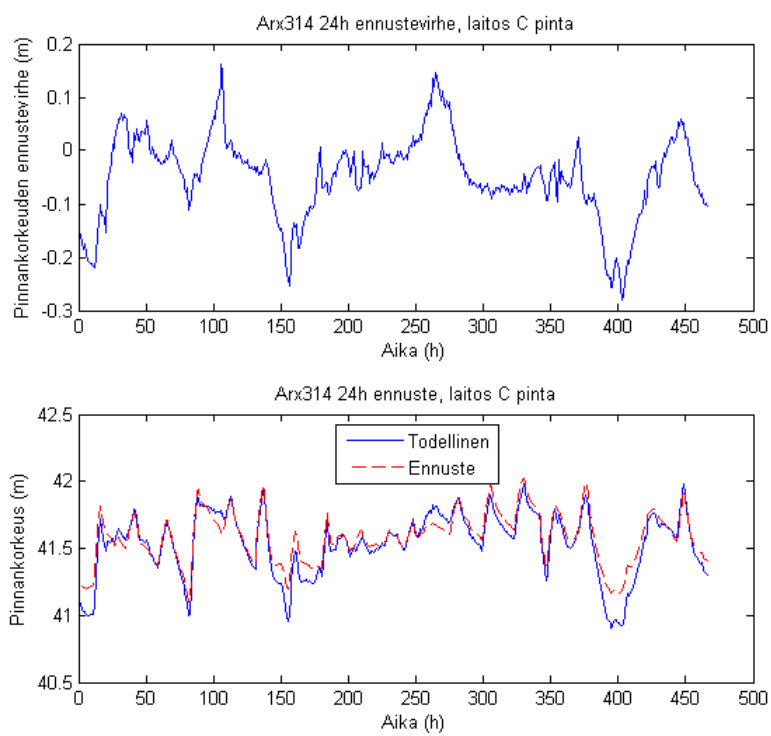
## E Jääkansi: vesivoimalan C pinnankorkeuden ennustaminen



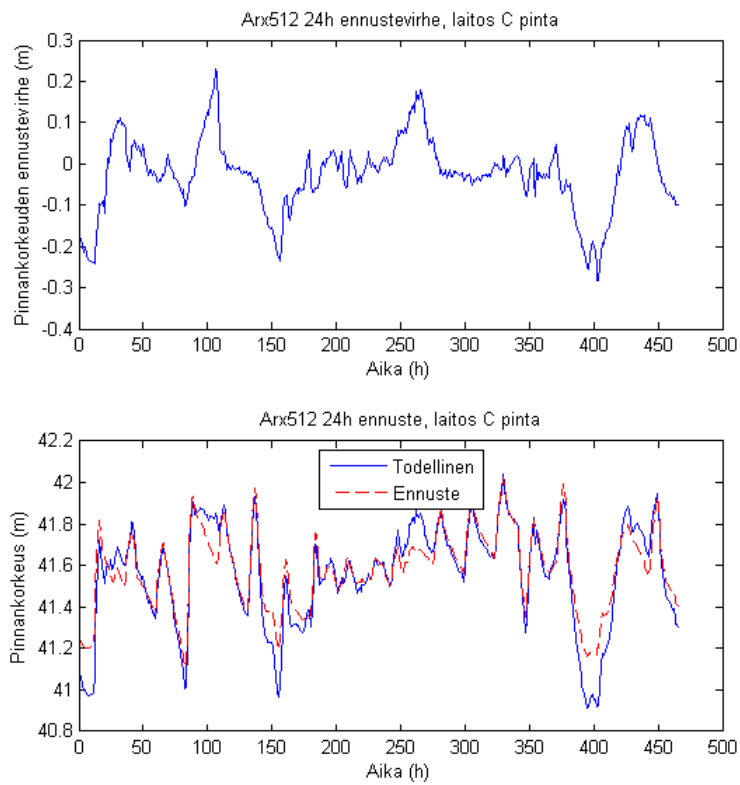
Kuva 30: Laitoksen C pinnankorkeudet laitoksen yläjuoksulla jääkannen alla, aikavälillä 1.2-15.3.2009.



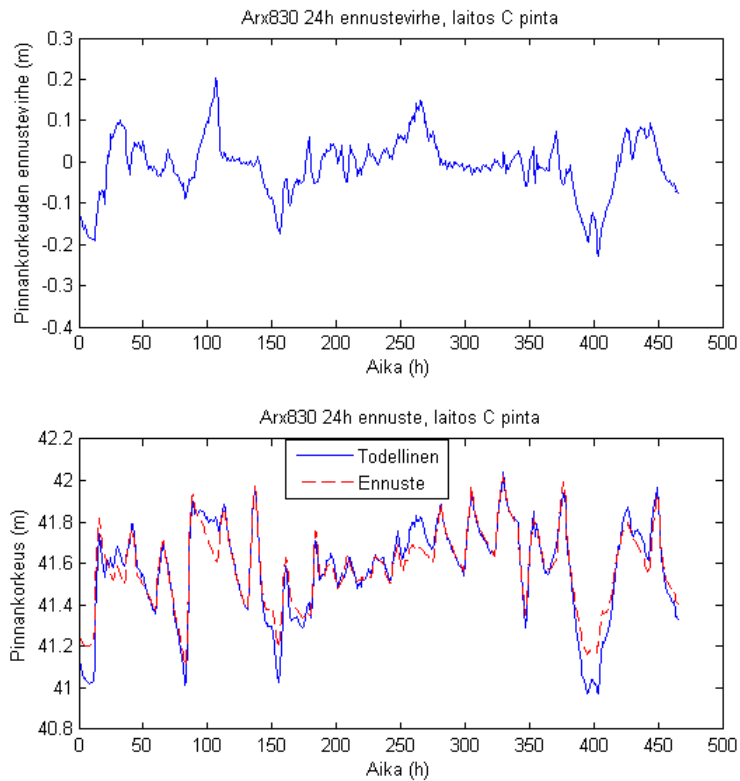
Kuva 31: Laitosten B, C, ja sivujoen B1 virtaukset jääkannen alla, aikavälillä 1.2-15.3.2009.



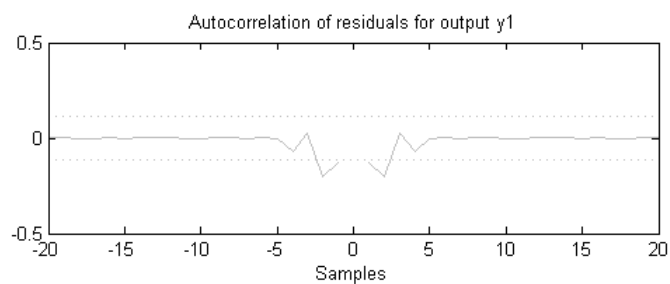
Kuva 32: ARX(3,[1 3],[4 0]), laitoksen C pinnankorkeuden ennustaminen ilman sivujoen B1 huomioista.



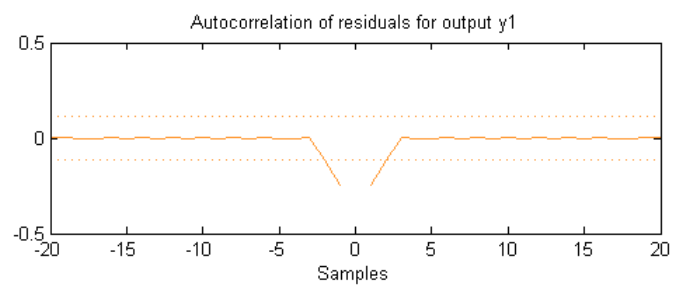
Kuva 33: ARX(5,[1 4 2],[2 0 4]), laitoksen C pinnankorkeuden ennustaminen kun sivujoki B1 huomioidaan.



Kuva 34: ARX(8,[3 8 6],[0 0 0]), laitoksen C pinnankorkeuden ennustaminen kun sivujoki B1 huomioidaan ja mukana enemmän parametreja kuin aikaisemmassa mallissa.

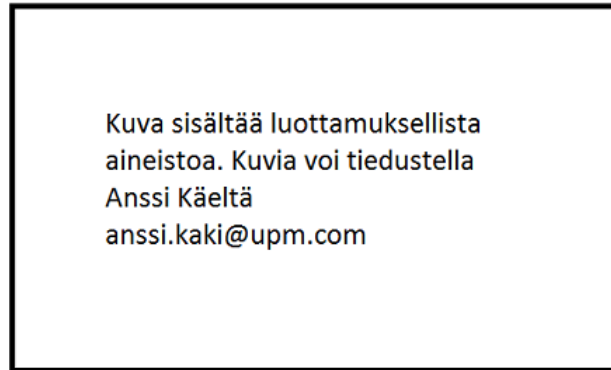


Kuva 35: ARX(5,[1 4 2],[2 0 4]), laitoksen C pinnankorkeuden ennustaminen kun sivujoki B1 huomioidaan, ennustevirheiden autokorrelaatiot.

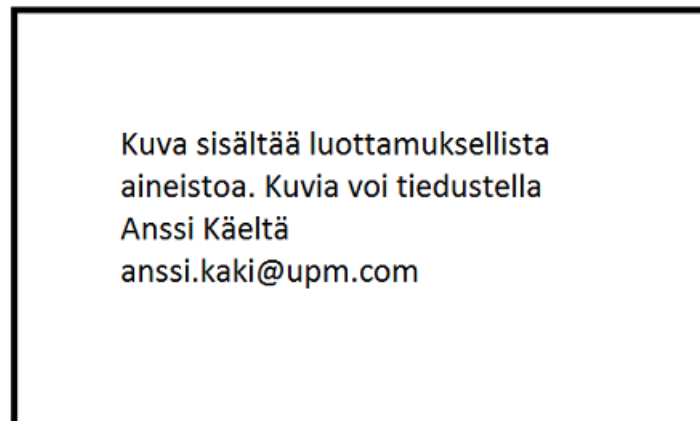


Kuva 36: ARX(3,[1 3],[4 0]), laitoksen C pinnankorkeuden ennustaminen ilman sivujokea B1, ennustevirheiden autokorrelaatiot.

## F Jääkansi: vesivoimalan D pinnankorkeuden ennustaminen

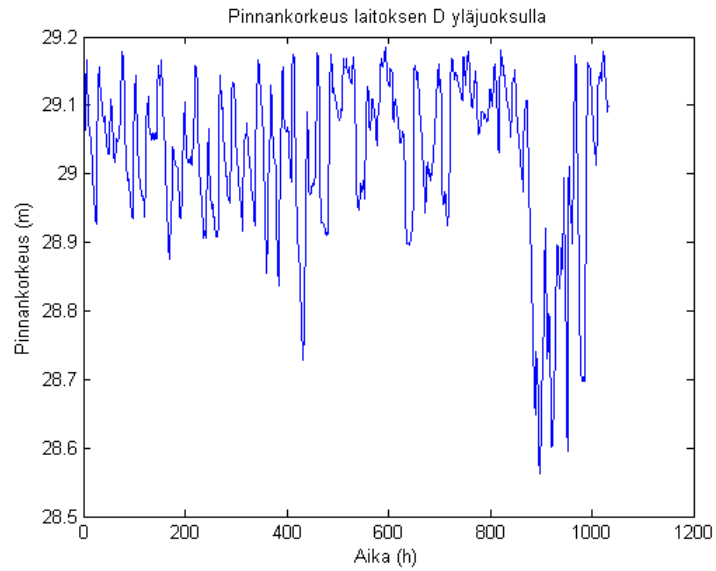


Kuva 37: Laitosten C ja D juoksutukset jääkansitilanteen aikana, aikavälillä 1.2-15.3.2009.

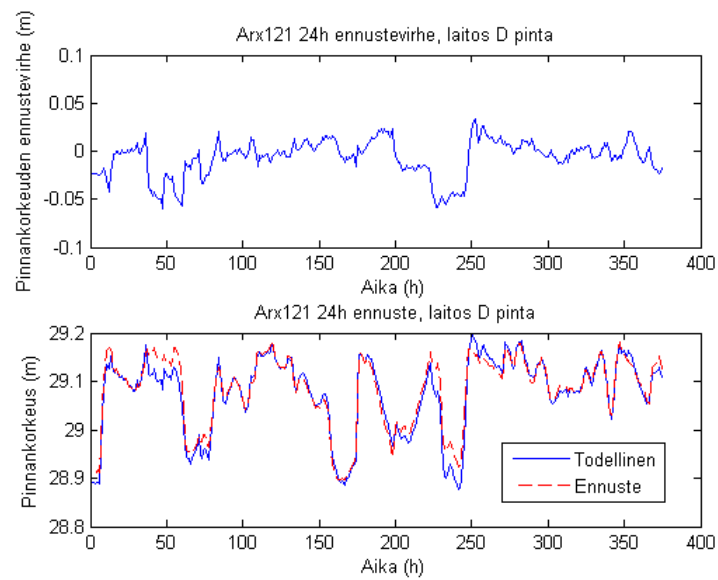


Kuva 38: Laitosten C ja D virtauksien vertaaminen tarkemmin jääkansitilanteen aikana.

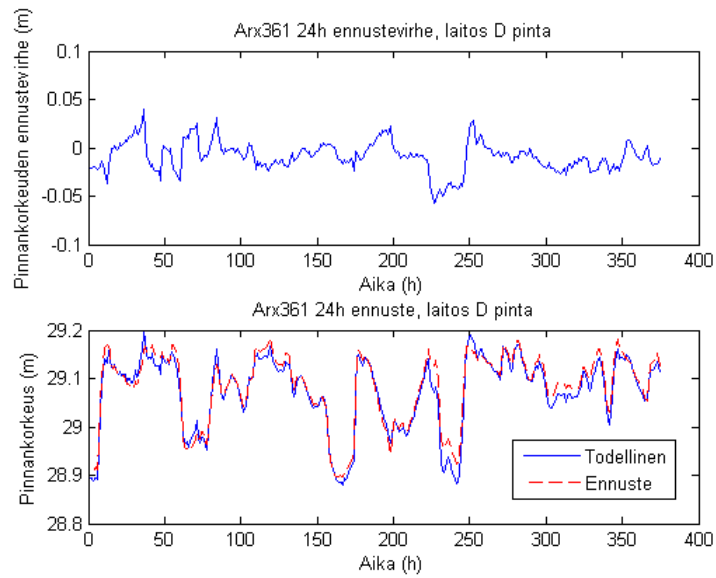




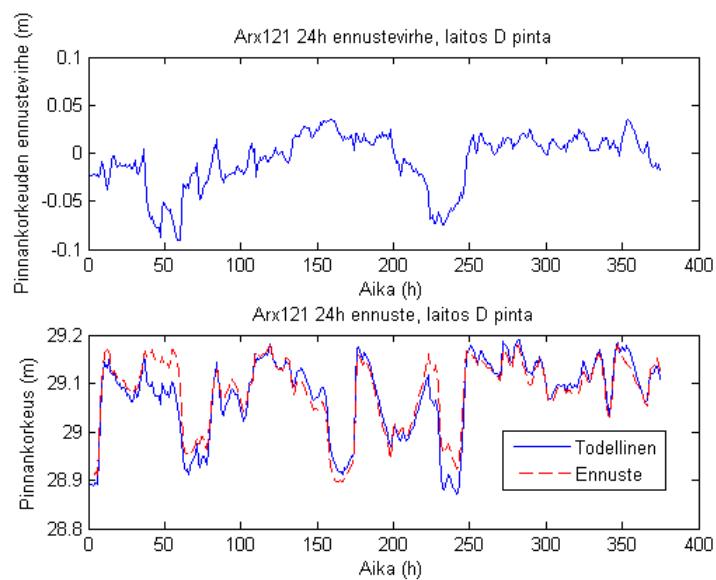
Kuva 39: Pinnankorkeus laitoksen D yläjuoksulla jääkansitilanteen aikana.



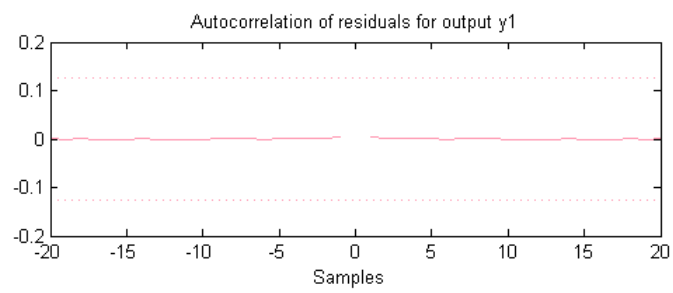
Kuva 40: ARX(1,[2 5],[1 0]), laitoksen D pinnankorkeuden ennustaminen.



Kuva 41: ARX(3,[6 2],[1 0]), laitoksen D pinnankorkeuden ennustaminen, vähemmän viipeitä laitoksen D juoksutuksella.

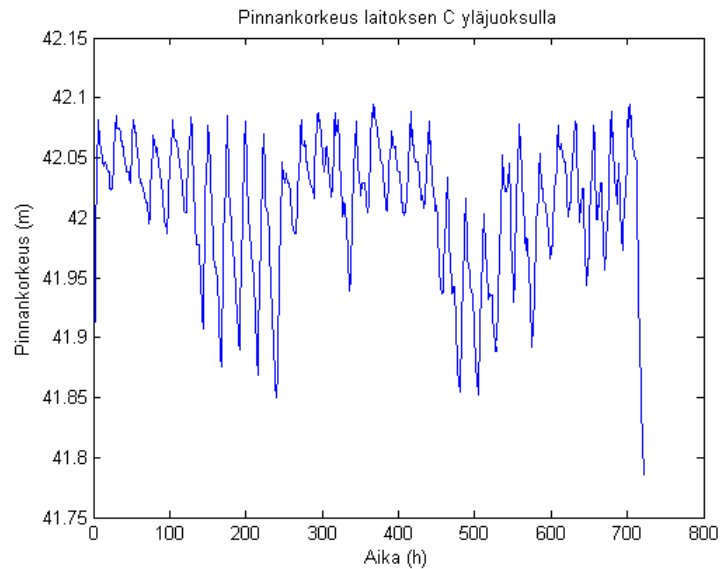


Kuva 42: ARX(1,[2 2],[1 0]), laitoksen D pinnankorkeuden ennustaminen, ei vääränmerkkisiä parametrien arvoja.

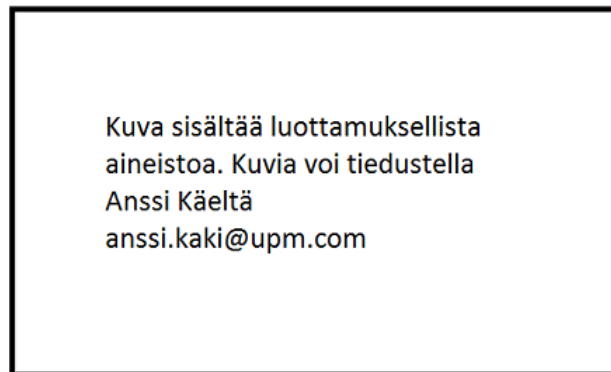


Kuva 43: ARX(1,[2 2],[1 0]), laitoksen D pinnankorkeuden ennustaminen, autokorrelaatiot.

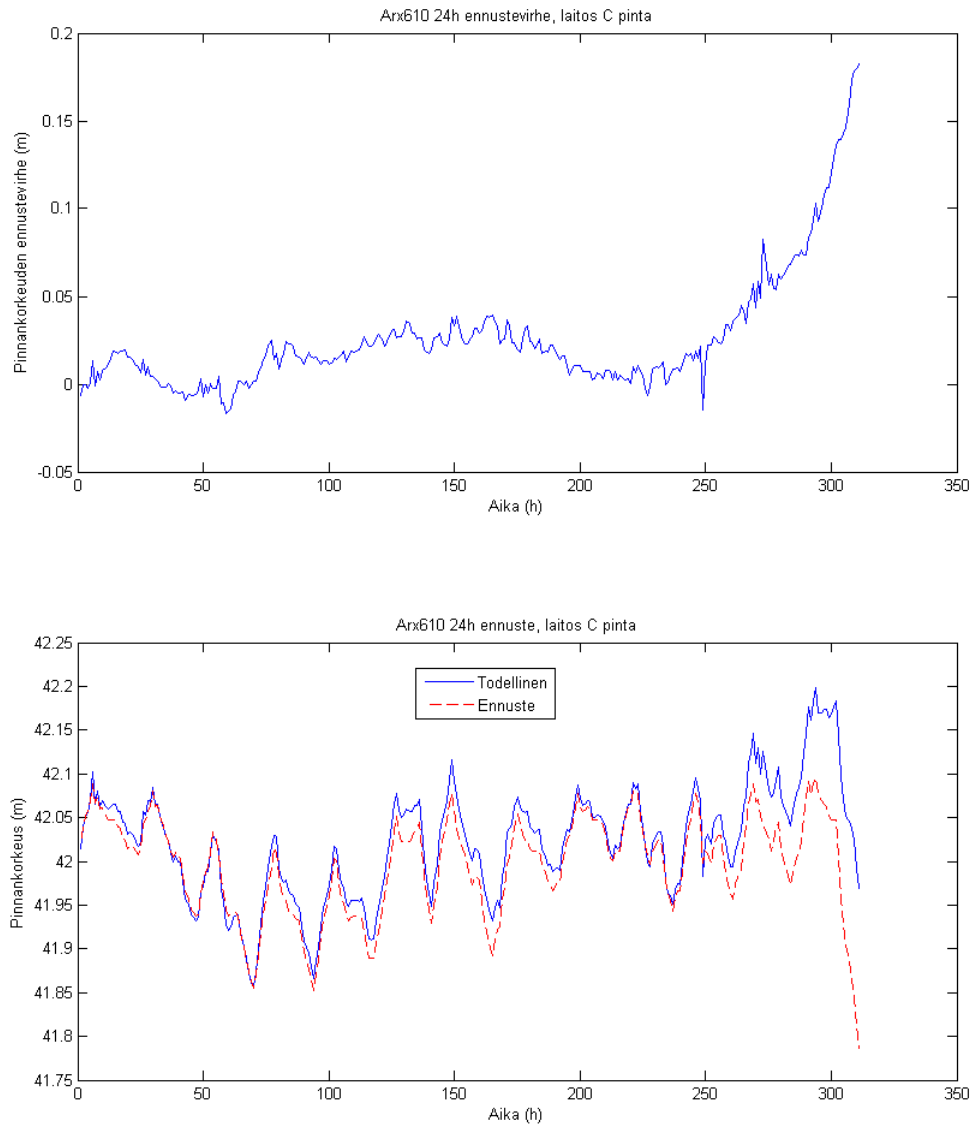
## G Niukka vesitilanne: vesivoimalan C pinnan- korkeuden ennustaminen



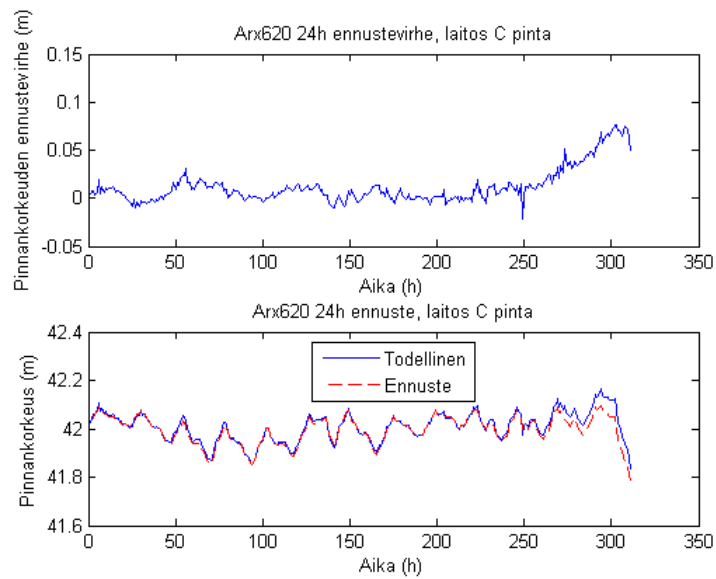
Kuva 44: Laitoksen C pinnankorkeudet laitoksen yläjuoksulla niukan vesitilanteen aikana, aikavälillä 21.10–19.11.2009.



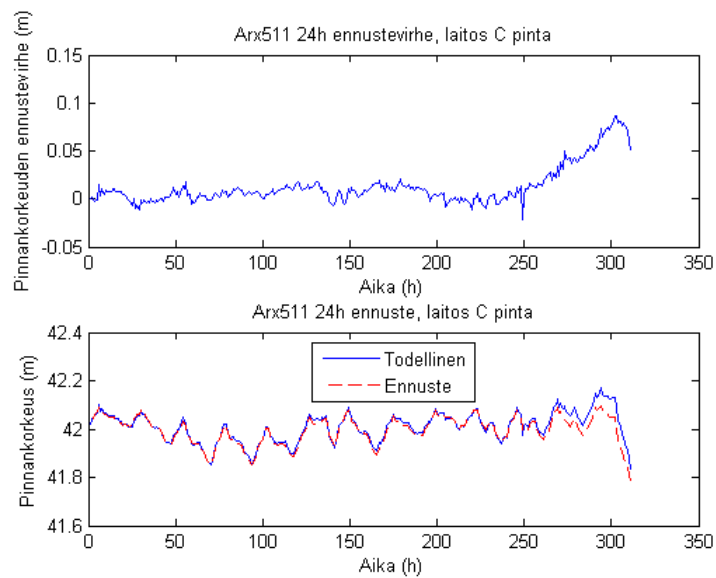
Kuva 45: Laitoksen C pinnankorkeudet laitoksen yläjuoksulla niukan vesitilanteen aikana, aikavälillä 21.10–19.11.2009



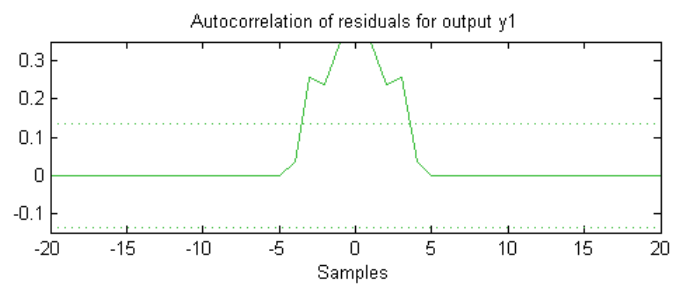
Kuva 46: ARX(6,[1 3],[0 0]), laitoksen C pinnankorkeuden ennustaminen ilman sivujoen B1 huomioimista.



Kuva 47: ARX(6,[2 3 1],[0 0 6]), laitoksen C pinnankorkeuden ennustaminen, kun sivujoki B1 huomioidaan.

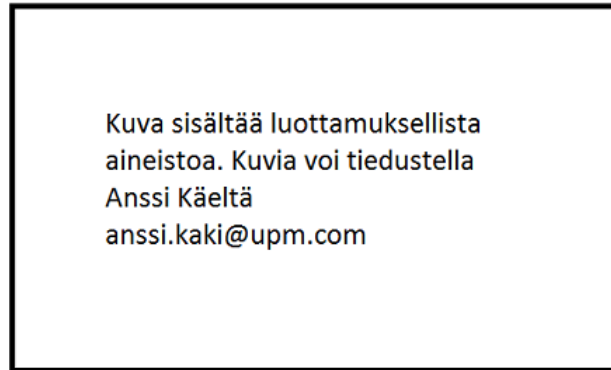


Kuva 48: ARX(5,[1 1 1],[1 0 6]), laitoksen C pinnankorkeuden ennustaminen, kun sivujoki B1 huomioidaan ja mallista poistettu väärän merkkiset parametrit.

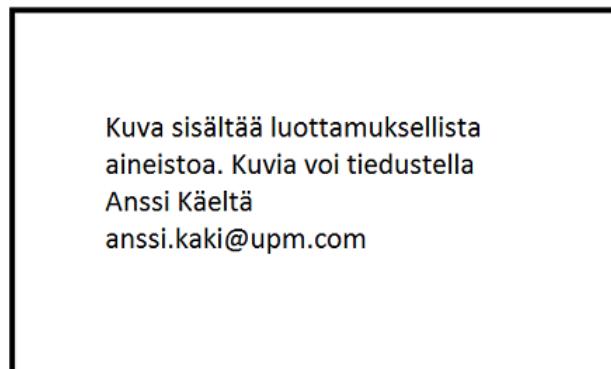


Kuva 49: ARX(5,[1 1 1],[1 0 6]), laitoksen C pinnankorkeuden ennustaminen, kun sivujoki B1 huomioidaan ja mallista poistettu väärän merkkiset parametrit, ennustevirheiden autokorrelaatiot.

## H Niukka vesitilanne: vesivoimalan D pinnan- korkeuden ennustaminen

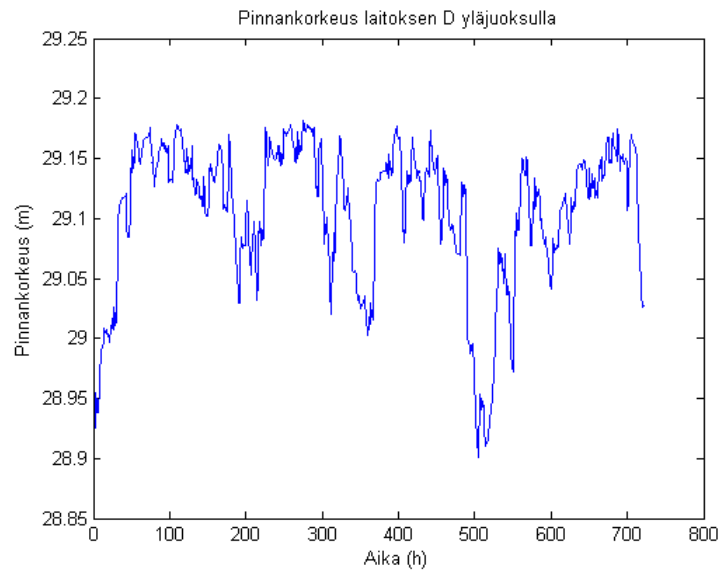


Kuva 50: Laitosten C ja D juoksutukset niukan vesitilanteen aikana.

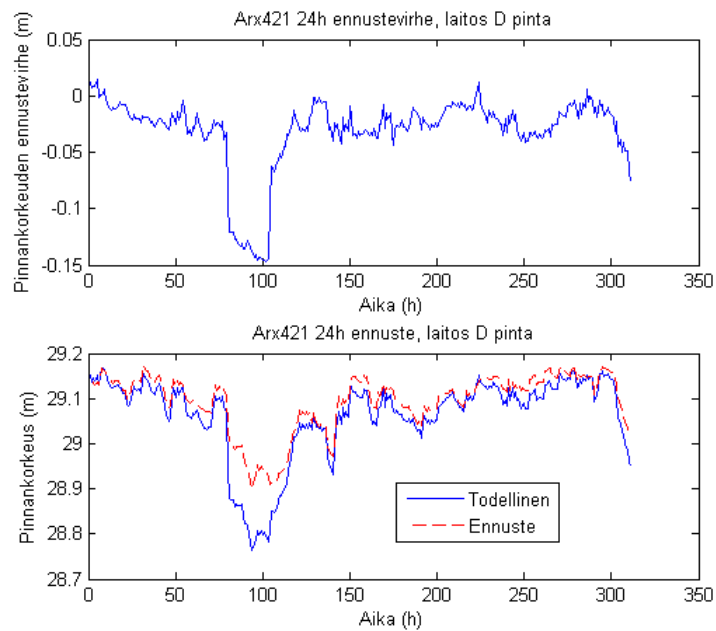


Kuva 51: Laitosten C ja D virtauksien vertaaminen tarkemmin niukan vesitilanteen aikana.

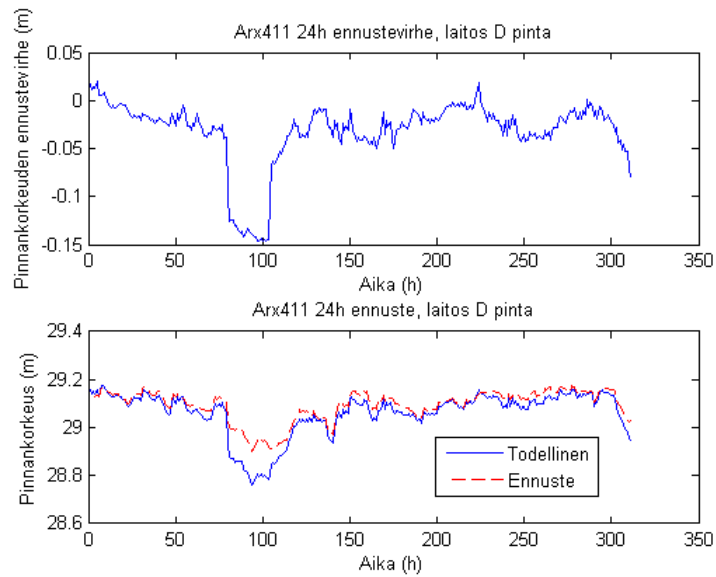




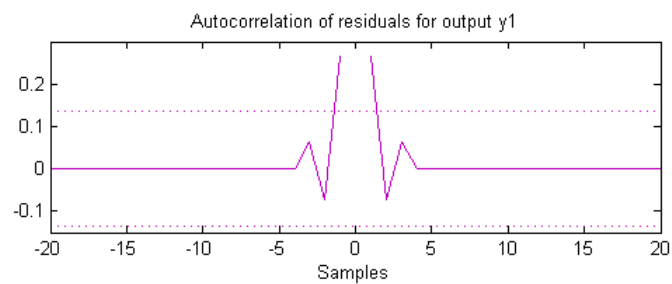
Kuva 52: Pinnankorkeus laitoksen D yläjuoksulla niukan vesitilanteen aikana.



Kuva 53: Laitoksen D pinnan ennustaminen mallilla  $ARX(4,[2 \ 2],[1 \ 0])$ .



Kuva 54: Laitoksen D pinnan ennustaminen mallilla  $ARX(4, [1 \ 2], [1 \ 0])$ . Tässä mallissa on järkevämmät kertoimien arvot.



Kuva 55: Laitoksen D pinnan ennustaminen  $ARX(4, [1 \ 2], [1 \ 0])$  mallilla, ennustevirheiden autokorrelaatiot.

## I Itsearvio

Projektityötä tehtiin niin, että ryhmä kokoontui kaksi kertaa viikossa tekemään sitä yhdessä. Raportin kirjoituksen aikana työtahti tiivistyi. Työ koitettiin jakaa tasaisesti kaikkien kesken, mutta aina se ei ollut helppoa. Aluksi kaksi ryhmäläistä keskittyi kirjallisuuskatsauksen tekoon ja toiset kaksi aloittivat itse mallinnusta ja tutustuivat aineistoon. Kun kirjallisuuskatsaus oli valmis, oli vaikea saada katsauksen tekijät mukaan mallinnukseen, jonka toiset olivat jo ehtineet aloittaa. Raportin kirjoittamisvaiheessa kaikki kuitenkin pystyivät osallistumaan tasapuolisemmin työntekoon.

Työmäärä koostui ajasta, jolloin tehtiin projektia yhdessä ja itsenäisestä työskentelystä, jonka osuus oli kuitenkin huomattavasti pienempi. Oli hyvä, että yhteiset ajat sovittiin, sillä sen ansiosta työ eteni tasaiseen tahtiin. Tavoitteena oli käyttää projektin tekoon kahdeksan tuntia viikossa. Tavoite saavutettiin pääosin. Välillä oli viikkoja, jolloin kukin vuorollaan teki vähemmän tunteja, mutta toisaalta kaksi viimeistä viikkoa töitä tehtiin huomattavasti enemmän, joten keskiarvo pysyi noin kahdeksassa tunnissa per viikko. Kurssin pituus oli 17 viikkoa, joten töitä tehtiin noin 136 tuntia per henkilö. Viiden opintopisteen työpanos vastaa arviolta 135 tunnin työtä. Käytimme siis aikaa ohjeistuksen verran, mikä on harvinaista, sillä useimmista kursseista saa hyvän arvosanan huomattavasti ohjeistusta pienemmällä työmäärällä.

Käytettävissä olevan ajan määrä rajasi sitä, kuinka syvällisesti pystyimme kuhunkin asiaan perehtymään ja tulokset ovat sen mukaisia. Projektin edessä jouduttiin tekemään joitain kompromisseja juuri jäljellä olleen ajan ja käytettyjen ohjelmistojen vuoksi. Alkuperäisestä suunnitelmasta poiketen ei esimerkiksi tutustuttu mallinnuksen ja optimoinnin yhdistämiseen kirjallisuuskatsauksessa. Käytetyn aineiston esikäsittely pidettiin myös mahdollisimman vähäisenä, sillä ei ollut selvää miten System Identification Toolbox sen teki ja oliko siihen luottaminen.

Käytetyt ohjelmat aiheuttivat yllättävän paljon ongelmia. Päätimme käyttää Matlabia ja erityisesti sen Identification Toolboxia, sillä se oli osalle ryhmäläisistä tuttu ja tiedettiin, että sillä pystyy aikasarjamalleja käsittelemään. Lisäksi alkuperäisessä tehtävänannossa oli toivottu lopullista mallin toteutusta Matlabin lisäosassa Simulinkissä. Matlab ei välttämättä ollut paras vaihtoehto, vaan esimerkiksi R-ohjelmisto olisi voinut sopia paremmin projektin toteutukseen. Erityisesti käyttämämme System Identification Toolbox -lisäosan graafisen käyttöliittymän kanssa oli ongelmia, sillä se oli toiminnoiltaan aika rajoittunut. Työskentelyn nopeuttamiseksi yritimme tehdä mahdollisimman suuren osan työstä Matlabin komentorivillä, mutta tässä ei täysin onnistuttu.

Kun nämä ongelmat huomattiin, oli jo liian myöhäistä alkaa opettelemaan uuden työkalun, kuten R:n, käyttöä. Matlabin dokumentaation tutkimiseen ja työskentelyn automatisointiin käytettiin hyvin paljon aikaa, mutta hyviä tuloksia ei saatu ja jouduimme palaamaan suppean ja hitaan käyttöliittymän käyttöön.

Olimme tyytyväisiä siihen, miten lähdimme ongelmaa ratkaisemaan. Perehdyimme jonkin verran fysikaalisiin malleihin kirjallisuuskatsauksessa, mutta lopulta aika nopeasti päädyimme aikasarja-analyysiin. Vaikka aikasarjamallit ovat meille tuttuja, niin silti aika oli loppua kesken. Viimeisillä viikoilla-kin vielä validoitiin malleja ja mietittiin mikä malli on millekin välille paras. Lisäksi esimerkiksi kausivaihtelun huomioimista SARMAX-malleilla emme testanneet ollenkaan, vaikka aineiston perusteella sille olisi voinut olla aihetta. SARMAX-malleja ei löytynyt System Identification Toolboxista ja tämä olikin yksi sen ongelmista.

Aikasarjamallit olivat ryhmäläisille ennestään tuttuja ja siksi turvallinen valinta mallinnustavaksi, mutta toisaalta myös jonkin uuden menetelmän opettelu olisi voinut olla erittäin mielenkiintoista. Aikasarjamallienkin luominen oli tosin yllättävän vaikeaa oikealle aineistolle verrattuna Ennustamisen ja aikasarja-analyysin kurssin harjoitustehtäviin.

Onnistuimme rajaamaan aiheemme hyvin yhteistyössä tehtävänantajan kanssa. Projektin tavoitteet päivittyivät sen edetessä realistisemmiksi. Tärkeimpiä huomioita oli, ettei yksi aikasarjamalli tulisi kuvaamaan erilaisia tilanteita hyvin vaan päivitettiin tavoite yhdestä mallista kolmen keskeisen, erilaisen tapauksen tutkimiseen. Alkuperäisiin tavoitteisiin ei ehkä päästy, mutta ne havaittiin epärealistisiksi projektin keston ja resursseihin nähden.