

Vuoroveden ja pyydyskalastuksen
vaikutus Tenon lohen nousuun nousun
aktiivisena aikana

Datatiteen pro gradu-tutkielma
Antti Rätty
2499172
Matemaattisten tieteiden tutkinto-ohjelma
Oulun yliopisto
Talvi 2022

Tiivistelmä

Osana vuosittaista elinkiertoaan, Atlantin lohi nousee jokiin kutemaan. Nousu ajoittuu kesälle ja sen aikana lohet uivat jokea vastavirtaan kymmenien kilometrien matkoja. Löydettyään sopivan paikan, lohet kutevat ja palaavat mereen. Lohen nousu on monimutkainen ja tarkka prosessi, johon vaikuttaa monet ympäristömuuttujat ja niiden yhdistelmät. Tenojoki on yksi tärkeimmistä Atlantin lohien nousujoista.

Vuosina 2018 ja 2019 Tenolla suoritettiin kaikuluotauseurannat. Kaikuluotaimet sijoitettiin Pulmankiin, 55 kilometrin päähän Tenonvuonolta. Vuonna 2018 käytettiin yhtä kaikuluotainta ja vuonna 2019 kahta kaikuluotainta. Kaikuluotaimella havaitut kalat tallennettiin ja saatiin aineisto nousevien lohien määristä ja nousun sijoittumisesta ajallisesti.

Tässä lopputyössä tutkitaan, onko vuoroveden vaiheella ja pyydyskalastuksella vaikutusta lohien nousun intensiteettiin (kpl/h) nousun aktiivisena aikana. Lopputyö perustuu vuoden 2019 kaikuluotausaineistoon. Ilmiötä tutkitaan bayesiläisen tilastotieteen menetelmin ja ilmiötä kuvaamaan luodaan Bayes-malli. Malli on monivaiheinen ja se kuvaa lohien matkan Tenonvuonolta lohien kaikuluotauspaikalle.

Mallin tuloksien perusteella vuorovesi vaikuttaa lohien nousun intensiteettiin. Lohien nousun intensiteetti kasvaa lähestyttäessä vuoroveden huippua ja on yleisesti ottaen korkeampaa laskuveden aikaan. Pyydyskalastuksella ei ole merkittävää vaikutusta nousun intensiteettiin. Tulokset ovat uskottavia, mutta eroavat ennakkoarjoista pyydyskalastuksen suhteen.

Sisällys

1	Johdanto	4
2	Aineisto	6
2.1	Aineiston keruu	6
2.2	Aineiston käsittely	7
3	Tilastomenetelmät	9
3.1	Bayes-päätely	9
3.2	Bayesiläisen analyysin käsitteet ja Bayesin kaava	10
3.3	MCMC -simulointi	11
3.3.1	Markovin ketju	11
3.3.2	Monte Carlo -integrointi	12
3.3.3	Metropolisin ja Hastingsin algoritmi	13
3.3.4	Gibbsin otanta	14
3.3.5	Simuloitujen ketjujen suppenemisen tarkastelu	15
3.4	Tilasto-ohjelmat	16
3.4.1	R	16
3.4.2	JAGS	17
3.5	Hierarkkinen Bayes-malli	17

3.5.1	Vaihe yksi: Siirtymä Tenonvuonolta Tenojokeen	20
3.5.2	Vaihe kaksi: Siirtymä Tenojoessa kaikuluotauspaikalle .	21
3.5.3	Vaihe kolme: Havaituksi tuleminen	22
3.6	Priorit	23
4	Tulokset	31
4.1	Analyysin suorittaminen	31
4.2	Vaiheen yksi tulokset	31
4.3	Vaiheen kaksi tulokset	39
4.4	Vaiheen kolme tulokset	42
5	Pohdinta	44

1 Johdanto

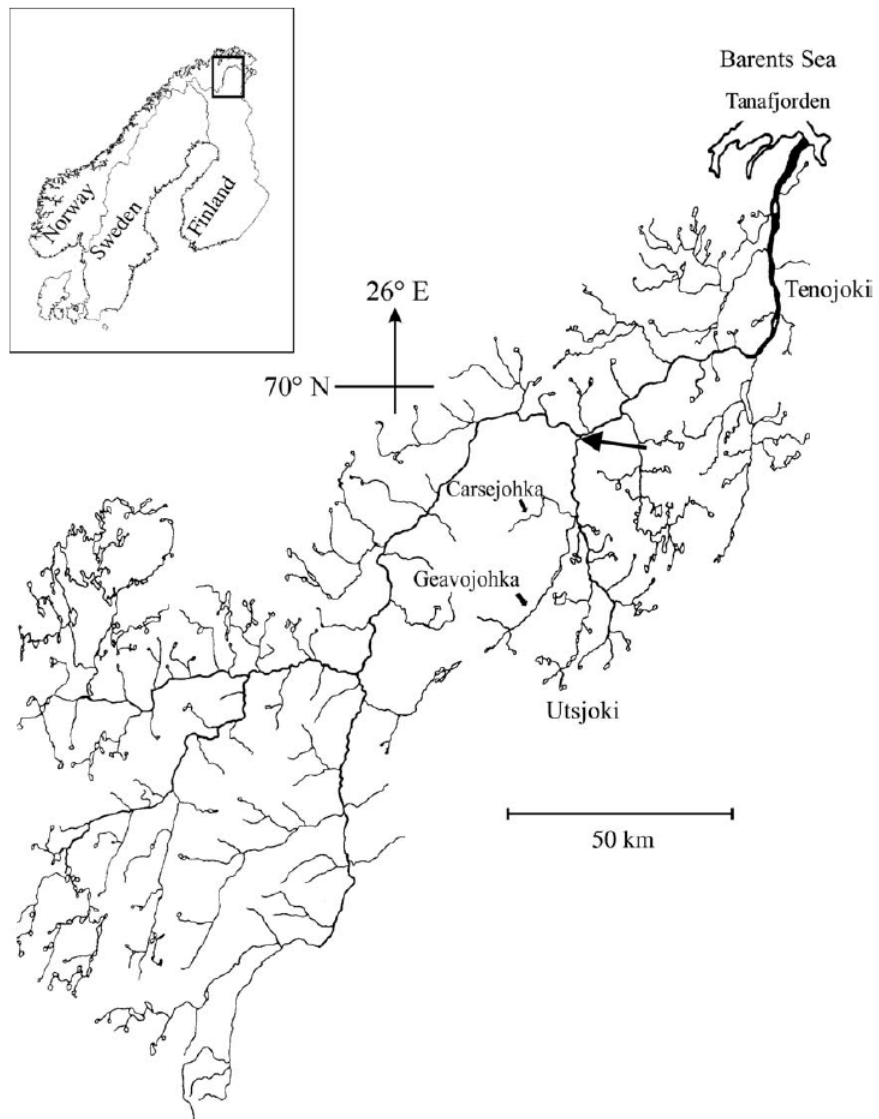
Tenojoki (Teno) kulkee Skandinavian niemimaan koillisosissa. Teno saa alkunsa Barentsinmeren rannalta ja laskee etelään toimien Suomen ja Norjan välisenä rajajokena. Teno haarautuu matkansa aikana pääuomastaan moneen sivu-uomaan. Tenojoen pääuoman pituudeksi on mitattu noin 361 kilometriä, mutta laskettaessa mukaan myös sivu-uomat Tenon pituutta voidaan pitää mittaamattomana (kuva 1).

Teno on yksi maailman tärkeimmistä Atlantin lohen lisääntymisjoista. Tenossa on yli 1200 kilometriä lohen lisääntymiselle suotuisaa aluetta, ja joessa tavataan yli 30 perinnöllisesti toisistaan eroavaa lohikantaa. Nämä kannat mahdollistavat kalastuksen Barentsinmereltä itse joelle asti. Pelkästään Tenojoesta saatu saalis vastaa noin kolmasosaa kaikkien Norjan lohijokien (yli 400 kappaletta) saaliista (Pohjola ym., 2020).

Tämä lopputyö käsittelee Atlantin lohen lisääntymistä ja siihen liittyvää lohen nousua Tenojoessa. Tarkoituksena on selvittää ulkoisten tekijöiden vaikutusta lohen nousun intensiteettiin perustuen vuoden 2019 kaikuluotausseurannan aineistoon. Tutkimuskysymys kuuluu: ”Onko pyydyskalastuksella ja vuorovedellä vaikutusta lohen nousun intensiteettiin nousun aktiivisena aikana?”.

Patokalastuksessa joki padotaan tiettyyn pisteeseen saakka ja padon päähän asetetaan pyydys, esimerkiksi verkko (Erkinaro ym., 1999a). Patojen ja pyydysten oletetaan aiheuttavan haittaa lohen liikkumiselle laskien nousun intensiteettiä.

Kalastajat ovat havainnoineet laskuveden olevan suotuisampaa aikaa kalastukselle. Aiemmassa tutkimuksessa laskuveden on huomattu olevan lohen vaelluksen kannalta suotuisampaa (Erkinaro ym., 1999a) ja (Karppinen ym., 2004). Tässä tutkielmassa tarkastellaan, onko samaa havaittavissa myös vuoden 2019 aineistosta.



Kuva 1: Tenojen vesistö, jossa näkyy Tenon pääuoma ja lukuisat sivu-uomat (Orell ym., 2007).

2 Aineisto

Lohikantoja säädellään kalastusrajoituksilla. Säätelyn tarkoituksena on ylläpitää ja varmistaa lohikantojen hyvinvointia ja säilyvyyttä (Pohjola ym., 2020). Säättely toteutetaan yhdessä Suomen ja Norjan välillä. Säätelyn toteuttamiseksi laaditaan suomalais-norjalainen Tenojoen kalastussopimus ja -sääntö, jota varten tehdään paljon tutkimusta ja seurantaa.

2.1 Aineiston keruu

Vuosina 2018 ja 2019 Tenojoessa suoritettiin kaikuluotausseurannat. Yhtenä taustasyynä seurannoille oli vuonna 2017 hyväksytty uusi kalastussopimus ja -sääntö, jota varten oli tärkeää saada luotettavaa ja ajantasaista tietoa nousevien lohien määristä (Pohjola ym., 2020).

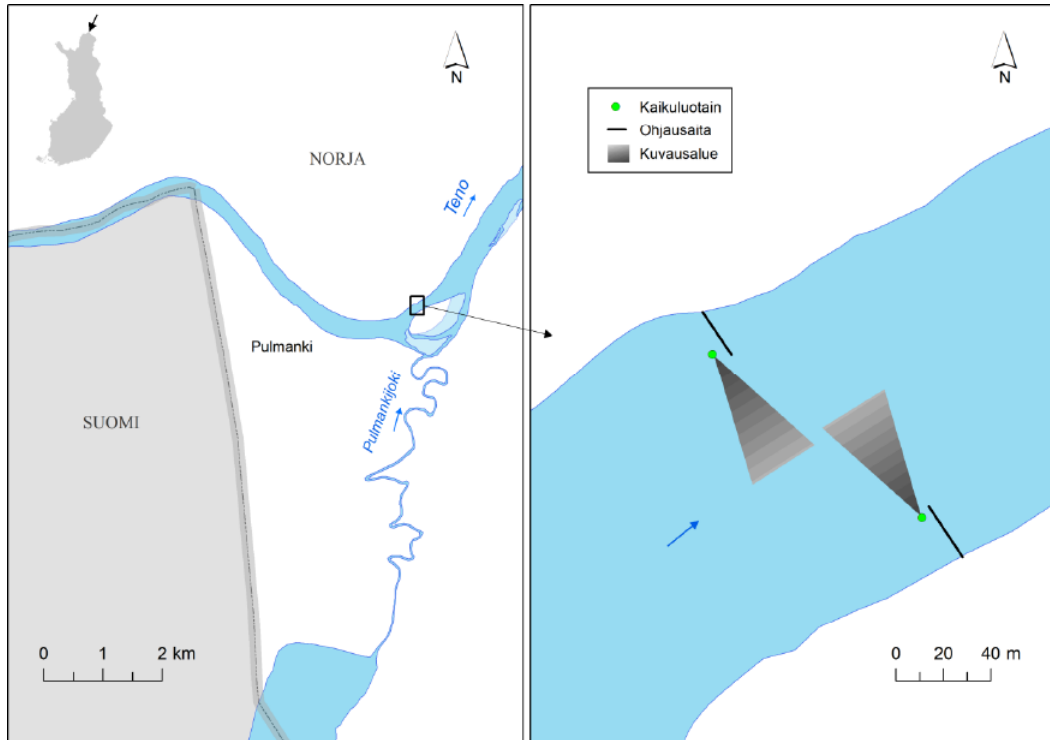
Vuonna 2018 käytössä oli yksi luotain, mutta vuonna 2019 käytettiin kahta luotainta, jolloin tarkkuus parani (Pohjola ym., 2020). Suuremman tarkkuuden takia tämä tutkielma perustuu vuoden 2019 seurantaan ja seuraavat kappaleet kuvaavat sen laitteistoa ja menetelmiä.

Luotaimet sijoitettiin Tenon pääuomaan Pulmangissa, noin 55 kilometrin päähän Tenonvuonolta. Veden ollessa matalimmillaan, joen leveys oli noin 120 metriä ja syvin kohta noin neljä metriä. Luotaimet asetettiin joen vastakkaisille rannoille siten, etteivät luotainten keilat mene päällekkäin. Luodattavan alueen kaventamiseksi käytettiin ohjausaitoja, joiden avulla luodattavan alueen leveydeksi jäi 100 metriä (Pohjola ym., 2020) kuva 2.

Aineiston keräämiseksi käytettiin Sound Metrics Aris Explorer 1200-luotainta, jota käyttää kahdella taajuudella 0,7 MHz ja 1,2 MHz. Korkeampi taajuus tuottaa laadukkaampaa kuvaa, muttei yhtä kaukaa (Pohjola ym., 2020). Tenojoella käytettiin matalampaa taajuutta ja noin 50 metrin luotausmatkaa.

Luotaimen tuottama materiaali tallennettiin ulkoisille kiintolevyille ja se käsiteltiin hyödyntäen luotaimen omaa tietokoneohjelmaa Arisfish. Arisfish-ohjelma helpottaa materiaalin käsittelyä esimerkiksi korostamalla liikkuvat kohteet luotainkuvasta. Materiaalin käsittely on kuitenkin työlästä, sillä jo-

kainen havainto tulee mitata ja varmistaa oikeaksi. Kerätty aineisto on tarkkaa havaintomäärien kannalta, mutta kalojen pituuksissa voi olla 10 senttimetrien heittoja (Pohjola ym., 2020).



Kuva 2: Tenojoen kaikuluotauspaikka Pulmangissa 2019 (Pohjola ym. 2020).

2.2 Aineiston käsittely

Kaikuluotauseurannoista saatu aineisto oli kasattu Excel-taulukkoon. Taulukossa jokainen rivi oli yksi luotaimella havaittu kala. Riville oli tallennettu havainnon kelloaika, pituus, uintisuunta (ylävirtaan/alavirtaan) ja muuta tietoa kuten joen virtaama. Aineistosta poistettiin alavirtaan liikkuvat havainnot.

Jokainen havaittu kala oletettiin Atlantin loheksi, eikä kaloja jaoteltu eri lajeihin pituuden perusteella. Erottelu olisi ollut mahdollisesti ongelmallista, koska ensimmäisen merivuoden Atlantin lohi on sekoitettavissa muihin pienempiin lajeihin. Lyhyitä havaintoja oli myös suhteessa vähäinen määrä (alle 25%) pidempiin Atlantin lohihavaintoihin nähden.

Aineisto aggregoitiin tunneittain ja päädyttiin muotoon, jossa yhdellä rivillä on tuntikohtainen luotaimella havaittu kalamäärä, yksikössä lohta per tunti. Muokattuun aineistoon lisättiin tieto vuoroveden vaiheesta ja paikallisesta kalastuskäytännöstä.

Norjalaiset tutkijat olivat kirjanneet vuoroveden huipun ajankohdat Tenonvuonolla taulukkoon, jonka perusteella aineistoon lisättiin vuorovesisarake. Vuorovesisaraketta varten luotiin vuorovesimuuttaja, joka huomioi vuoroveden huipun ajankohdan, sekä kuinka kauan aikaa on seuraavasta tai edellisestä huipusta.

Tätä työtä varten muodostettu vuorovesimuuttaja saa arvoja $[0, 100]$. Huipun aikaan muuttujan arvo on 100, jonka jälkeen muuttujan arvot laskevat arvoon 0 seuraavan kuuden tunnin aikana (laskuvesi). Tämän jälkeen arvot lähtevät tasaiseen nousuun kohti arvoa 100, kunnes uusi huippu saavutetaan (nousuvesi). Tämä arvojen muuttuminen perustuu oletukseen, jonka mukaan lohet ovat aktiivisempia vuoroveden huipun aikaan ja sen lähetyvillä (Erkinaro ym., 1999b). Lisäksi muodostettiin indikaattorimuuttuja, jonka avulla laskuvesi ja nousuvesi erotellaan toisistaan.

Aineistoon lisättiin paikallista kalastuskäytäntöä kuvaava sarake. Saraketta varten muodostettiin kalastusmuuttuja. Muuttuja saa arvoja $[0, 1]$, paikallisen kalastussäätelyn perusteella. Paikallisen kalastussäännön mukaan verkko- ja muuta pyydyskalastusta saa harjoittaa aikavälillä 1.6-31.7. Lisäksi pyydyskalastus on sallittua tällä ajanjaksolla vain joka viikon maanantaita kello 19 saman viikon keskiviikkoon kello 19. Muuttuja saa arvon 1, kun pyydyskalastus on sallittua ja arvon 0 muulloin.

3 Tilastomenetelmät

Tämän luvun tarkoitus on esitellä bayesiläinen analyysi ja siihen liittyvät käsitteet. Lisäksi paneudutaan analyysin suorittamisen metodeihin ja lopuksi esitellään tässä lopputyötä varten luotu malli.

3.1 Bayes-päätely

Bayesiläinen analyysi (Bayes-päätely) on tilastollisen päättelyn menetelmä, jonka katsotaan saaneen alkunsa englantilaisen pappismiehen Thomas Bayesin (1702-1761) vuonna 1763 postuumisti julkaistusta artikkelista (Bayes, 1763). Bayes-päätelyssä uusia tutkimustuloksia saadaan yhdistämällä aiempia tuloksia tämänhetkiseen tutkittavaan havaintoaineistoon.

Bayes-päätely eroaa perinteisestä frekventistisestä tilastotieteestä, suurin ero näiden välillä liittyy todennäköisyyden käsitteeseen. Frekventistisessä tilastotieteessä todennäköisyyden ajatellaan olevan kytkettynä ilmentymään, esimerkiksi toistoihin. 50% todennäköisyys onnistumiselle tarkoittaa frekventistisessä mielessä 100 toiston sarjassa 50 onnistumista, 1000 sarjassa 500 onnistumista ja niin edelleen (van de Schoot ym., 2013). Frekventistisessä ajattelutavassa todennäköisyys on objektiivista, se näyttäätyy kaikille samalla tavalla ja jokainen tekee siitä samat havainnot.

Bayes-päätelyssä todennäköisyyttä käsitellään uskomuksen asteena (degree of belief). Uskomuksen aste on subjektiivinen ja se voi muuttua. Bayesiläisessä mielessä 50% todennäköisyys toiston onnistumiselle on uskomus siitä, että seuraava toisto onnistuu 50% todennäköisyydellä. Jos toistojen suorittamista jatkettaisiin eikä onnistuneita toistoja havaittaisi alkuperäisen uskomuksen mukaisesti, uskomusta voitaisiin päivittää pienemmäksi. Tämä päivitys on subjektiivinen ja perustuu yhden havainnoijan (henkilön) uskomukseen ja sen muuttumiseen eikä toinen havainnoija välttämättä tekisi samaa päätelmää (van de Schoot ym., 2013).

Uskomuksen aste on keskeinen käsite Bayes-päätelyssä. Tällainen episteemisen tulkinnan lähtökohta antaa hyvät mahdollisuudet epävarmuuden selittämiseen ja tulkintaan. Parametristimaattien oletetaan noudattavan määrät-

tyä jakaumaa, joka sitoo sekä estimaatin että siihen liittyvän epävarmuuden samaan yhteyteen. Tarvittavat tulkinnat esimerkiksi posterioriväleistä (credible intervals) tehdään suoraan tästä jakaumasta (van de Schoot ym., 2013).

3.2 Bayesiläisen analyysin käsitteet ja Bayesin kaava

Bayes-päätelyssä ajatellaan saatujen havaintojen eli datan, $y = (y_1, \dots, y_n)$, noudattavan kiinnostavalle parametrille θ ehdollistettua jakaumaa. Datasta saatava näyttö (evidence) ilmenee sen pohjalta muodostettavasta uskottavuusfunktioista $L(\theta; y) = f(y|\theta)$ eli uskottavuudesta (van de Schoot ym., 2013).

Parametria θ käsitellään satunnaismuuttujana, jonka oletetaan noudattavan ennalta määrättyä jakaumaa. Ennalta määrättyä jakaumaa kutsutaan priorijakaumaksi. Priorin tarkoituksena on kuvastaa parametriin θ liittyvää tämänhetkistä varmuutta ja epävarmuutta. Priori muodostetaan datasta riippumattoman ulkopuolisen informaation pohjalta (van de Schoot ym., 2013).

Priorin tiheysfunktion $p(\theta)$ ja uskottavuuden $f(y|\theta)$ avulla muodostetaan posteriorijakauma (posteriori) $p(\theta|y)$. Posteriorin tiheysfunktion $f(\theta|y)$ avulla parametrin θ epävarmuutta tarkastellaan tämänhetkisen datan pohjalta (van de Schoot ym., 2013).

Parametrin θ estimaatti määrätään posterista. Estimaattina voidaan käyttää esimerkiksi posteriorin keskiarvoa tai mediaania. Parametrin epävarmuutta tarkastellaan posteriorivälin avulla, yleisin on 95% posterioriväli. Posterioriväli informoi välin $[a, b]$, jolle parametrin estimaatti kuuluu annetulla todennäköisyydellä. Esimerkiksi, jos parametrin θ 95% posterioriväli olisi $[0, 1]$, parametrin θ estimaatti kuuluu 95% todennäköisyydellä välille $[0, 1]$ (van de Schoot ym., 2013).

Posteriorijakauma muodostetaan:

posterioritiheys = $c \times$ **priori** \times **uskottavuus**,

jossa c normalisoiva tekijä. Normalisoiva tekijä varmistaa, jotta kun posterioritiheyttä integroidaan arvoalueensa ylitse, integraali on yksi ja täten tiheysfunktion määritelmä täyttyy. Yleensä posterioria kuvataan suoraan Bayesin kaavalla:

$$p(\theta|y) = \frac{p(\theta)f(y|\theta)}{f(y)}. \quad (1)$$

jossa $p(\theta|y)$ on posteriori, $p(\theta)$ on priori, $f(y|\theta)$ on havainnot generoivalle prosessille muodostetun tilastollisen mallin pistetodennäköisyys- tai tiheysfunktion pohjalta muodostettu uskottavuusfunktio ja $f(y)$ on datan reunajakauman pistetodennäköisyys- tai tiheysfunktio. Tämä kaava toimii sekä diskreeteille että jatkuville muuttujille.

Datan reunatiheyden $f(y) = \int_{-\infty}^{+\infty} p(\theta)f(y|\theta)d\theta$ laskeminen on hankalaa siirtäessä yksinkertaisimmista tilanteista realistisiin ja moniparametrisiin tilanteisiin. Useassa tapauksessa analyttisiä lausekkeita $f(y)$:lle ei ole olemassa, jolloin turvaudutaan likimääräisiin laskentamenetelmiin kuten MCMC-simulointiin (Robert & Casella, 2009).

3.3 MCMC -simulointi

MCMC -simulointia hyödynnetään korkeaulotteisten integraalien likimääräisessä laskennassa. MCMC:llä on monia hyödyntämiskohteita muun muassa fysiikan puolella, mutta tässä yhteydessä relevantein käyttötarkoitus on posteriorijakauman simulointi. MCMC:n nimi tulee kahdesta erillisestä tekniikasta: Markovin ketju (Markov chain) ja Monte Carlo -integrointi.

3.3.1 Markovin ketju

Markovin ketju on Markovin prosessista saatu havaintosarja. Markovin prosessi kuvaa stokastista prosessia, jossa tulevaan tilaan vaikuttaa vain ja ainoastaan nykyinen tila (Robert & Casella, 2009). Mahdolliset tulevat tilat ja niiden jakauma saadaan laskettua nykyisen tilan ja mahdollisten siirtymien avulla.

Olkoon X Markovin prosessin mukainen satunnaismuuttuja, x_t muuttujan arvo hetkellä t ja $\pi(t) = \mathbb{P}(X = x_t)$ on arvojen jakauma. Olkoon P siirtymätodennäköisyysmatriisi, jossa todennäköisyydet ovat ehdollistettuja muuttu-

jan arvolle. Nyt muuttujan X mahdollisten arvojen jakauma hetkellä $t + 1$ eli $\pi(t + 1)$ saadaan

$$\pi(t + 1) = \pi(t)P, \quad (2)$$

kun samaa päättelyä jatketaan taaksepäin, saadaan lopulta

$$\pi(t + 1) = \pi(t)P = \pi(t - 1)P^2 = \dots = \pi(0)P^{t+1}, \quad (3)$$

jossa $\pi(0)$ on ketjun alkuarvo.

Kun Markovin ketju etenee tarpeeksi pitkään, ketju voi päätyä kohdejakaumaan π^* , jossa mahdollisten arvojen todennäköisyydet eivät muutu enää juurikaan, eli

$$\pi^* = \pi^*P. \quad (4)$$

Tämä on haluttua käytöstä, sillä tässä vaiheessa ketju on konvergoinut. Konvergoineessa ketjussa muuttujan nykyinen arvo ei enää vaikuta muuttujan arvojen jakaumaan. Ketjun alkuarvot ovat unohtuneet ja eri alkuarvoilla päädyttäisiin samaan jakaumaan.

3.3.2 Monte Carlo -integrointi

Monte Carlo-integrointi on fyysikoiden kehittämä tekniikka, jonka alkuperäinen käyttötarkoitus liittyy moniulotteisten integraalien laskentaan hyödyntäen satunnaislukugenerointia (Robert & Casella, 2009).

Monte Carlo-integroinnin peruseriaate on seuraava. Olkoon $\int_a^b h(x)dx$ moniulotteinen integraali yli välin $[a, b]$. Nyt $h(x)$ on esitettävissä funktion $f(x)$ ja tiheysfunktion $p(x)$ tulona, jolloin aiemmin esitetty integraali muokkautuu muotoon

$$\int_a^b h(x)dx = \int_a^b p(x)f(x)dx = \mathbb{E}(f(x)), \quad (5)$$

josta saadaan lopputuloksena odotusarvo $\mathbb{E}(f(x))$. Tämä odotusarvo voidaan useimmissa tilanteissa approksimoida

$$\mathbb{E}(f(x)) \simeq \frac{1}{n} \sum_{i=1}^n f(x_i), \quad (6)$$

jossa $f(x_i)$ ovat satunnaislukugeneroituja funktion $h(x)$ arvoja.

Tällä tavalla saadaan riippumaton otos $h(x)$:n arvoja halutulta väliltä, joiden avulla $h(x)$:n integraali selvitetään.

3.3.3 Metropolisin ja Hastingsin algoritmi

Monte Carlo -integroinnin yleinen ongelma on $p(x)$ arvojen generointi siinä tilanteessa, kun $p(x)$ jakauma on monimutkainen. Tämän toteuttamiseksi kehitettiin Metropolisin algoritmi, jota myöhemmin yleistettiin ja syntyi Metropolisin ja Hastingsin algoritmi (Hastings, 1970). Kumpikin algoritmi on iteratiivinen prosessi ja tuottaa Markovin ketjun. Tästä johtuen tuotettu otos on riippuvainen toisin kuin Monte Carlo -integroinnissa.

Olkoon $p(\theta)$ hankala jakauma, josta halutaan generoida lukuja ja joka voidaan esittää $p(\theta) = f(\theta)/K$. K on normalisoiva tekijä, joka on hankalasti määritettävissä. Metropolisin algoritmi etenee seuraavaan tapaan.

- 1 Valitaan alkuarvo θ_0 , joka täyttää ehdon $f(\theta_0) > 0$.
- 2 Generoidaan ehdokasarvo θ^* ehdokasjakauman $q(\theta_1, \theta_2)$ ja nykyisen arvon avulla.
- 3 Lasketaan α eli tiheysfunktion arvojen suhde ehdokasarvon θ^* ja nykyisen arvon θ_{t-1} välillä.

$$\alpha = \frac{p(\theta^*)}{p(\theta_{t-1})} = \frac{f(\theta^*)}{f(\theta_{t-1})} \quad (7)$$

- 4 Jos $\alpha > 1$ eli tiheys on suurempaa ehdokaspisteessä, ehdokas hyväksytään. Jos $\alpha < 1$ eli tiheys on pienempää ehdokaspisteessä, ehdokas hyväksytään todennäköisyydellä α . Kummankin vaiheen jälkeen palataan kohtaan 2.

Algoritmin lopputulemana saadaan Markovin ketju $(\theta_0, \theta_1, \dots, \theta_k, \dots)$, joka mahdollisesti riittävän lämmittelyjakson (burn-in period), esimerkiksi k - iteraation, jälkeen saavuttaa konvergenssin. Saavutettuaan konvergenssi, generoidut luvut $(\theta_{k+1}, \theta_{k+2}, \dots)$ noudattavat $p(\theta)$ jakaumaa.

Metropolisin ja Hastingsin algoritmissa α on määritetty

$$\alpha = \min\left(\frac{f(\theta^*)q(\theta^*, \theta_{t-1})}{f(\theta_{t-1})q(\theta_{t-1}, \theta^*)}, 1\right), \quad (8)$$

joka $q(\theta_1, \theta_2)$ ollessa symmetrinen toimii kuin Metropolisin algoritmi.

3.3.4 Gibbsin otanta

Gibbsin otanta on Gemanin veljesten vuonna 1984 esittelemä erikoistapaus Metropolisin ja Hastingsin algoritmista (Geman & Geman, 1984). Siinä yhdistetty jakauma muodostetaan ehdollistettujen jakaumien avulla. Gibbsin otanta eroaa Metropolisin ja Hastingsin algoritmista myös sillä tapaa, että $\alpha = 1$, eli ehdokasarvo hyväksytään aina.

Gibbsin otannassa moniparametrisesta yhdistetystä jakaumasta $f(x, y, z)$ luodaan täysin ehdollistetut (full conditional) yhteisjakaumat, joissa parametrit ovat täysehdollistetut (fully conditional) toisilleen $f(x|y, z)$, $f(y|x, z)$, $f(z|x, y)$. Täysin ehdollistetut jakaumat poikkeavat yhteisjakaumasta ja voivat noudattaa jotain tunnettua jakaumaa. Gibbsin otannassa tuntematonta yhdistettyä jakaumaa ei tarvitse operoida, vaan tarvittavat operaatiot suoritetaan täysin ehdollistetuille jakaumille.

Jos täysin ehdollistetut jakaumat ovat tunnettuja, Bayes-analyysissä voidaan hyödyntää konjugaattijakaumia. Jakauman ja sen konjugaattijakauman tulo on jokin tunnettu jakauma. Jos tunnetulle täysin ehdollistetulle jakaumalle valitaan konjugaattipriorijakauma, normalisoivaa tekijää ei tarvita, koska priorin ja uskottavuuden tulo noudattaa jotain toista tunnettua jakaumaa.

Olkoon $f(y, z)$ kaksiparametrinen reunatiheysfunktio. Tarkoituksena on selvittää $f(y, z)$, joten siitä muodostetaan kaksi täysin ehdollistettua jakaumaa $y \sim f(y|z)$ ja $z \sim f(z|y)$. Käsittely etenee seuraavalla tapaa.

- 1 Poimitaan alkuarvo y_0 .
- 2 Poimitaan z_0 hyödyntäen y_0 : $z \sim f(z|y = y_0)$.
- 3 Poimitaan y_1 hyödyntäen z_0 : $y \sim f(y|z = z_0)$.
- 4 Poimitaan z_1 hyödyntäen y_1 : $z \sim f(z|y = y_1)$.

Tätä jatketaan, kunnes haluttu määrä otoksia on saavutettu.

Tämä prosessi tuottaa sarjan $(y_0, z_0, y_1, z_1, \dots, y_k, z_k, \dots, y_n, z_n)$, jota kutsutaan Gibbsin sarjaksi. Gibbsin sarja on Markovin ketju, joka riittävän lämmittelyajanjakson jälkeen, esimerkiksi k -iteraatiota, mahdollisesti konvergoi ja alkaa noudattamaan $f(y, z)$ jakaumaa.

3.3.5 Simuloitujen ketjujen suppenemisen tarkastelu

Tärkeä osa MCMC-simulointia on sen oikea-aikainen päättäminen. Jos simulaatio keskeytetään liian aikaisin, voi olla, ettei kohdejakaumaa ole saavutettu. Simulaation liian pitkällinen jatkaminen ei paranna tuloksia, mutta vie ylimääräistä aikaa. Ketjun tarkastelulla voidaan tehdä päätelmiä simulaation pituuden riittävyydestä (Robert & Casella, 2009).

Konvergoinut ketju on saavuttanut kohdejakauman. Tätä on hankala todistaa, koska kohdejakaumat ovat usein monimutkaisia eivätkä välttämättä tiedossa. Useamman yhtäaikaisen rinnakkaisen ketjun tarkastelu parantaa tilannetta. Rinnakkaisten ketjujen tunnuslukuja, esimerkiksi keskiarvoja, voidaan verrata toisiinsa. Jos kohdejakauma on saavutettu, keskiarvot ovat lähellä toisiaan (Robert & Casella, 2009).

Ergodisuus on termi, joka kuvaa simuloitujen ketjun keskiarvosuppenevuutta (Robert & Casella, 2009). Keskiarvosuppenevuuden periaatteen mukaan, simuloitujen ketjun tulee löytää kohdejakauman moodit. Ketju ei välttämättä löydä kaikkia moodeja, jos se ei pääse vapaasti käymään lävitse kohdejakauman tukialuetta.

Ketjun peräkkäisten pisteiden välinen korrelaatio haittaa ketjun vapaata liikkumista tukialueella. Ohentamalla ketjua, pisteiden välinen korrelaatio las-

kee. Ohentamisessa simuloidusta ketjusta (x^1, x^2, \dots, x^n) poimitaan osaotos kokonaisluvun k välein $(x^{k-1}, x^{k-2}, x^{k-3}, \dots)$. Syntyneessä osaotoksessa ketjun peräkkäisten pisteiden välinen korrelaatio on vähäisempää.

Simuloiduista ketjuista on tyypillistä piirtää diagnostiikkakuvia. Kuvien avulla ketjujen käytöstä tarkastellaan yleisellä tasolla ja etenkin ei-toivottu käytös on huomattavissa. Jälkikuva kuvastaa ketjun kulkemista arvoavaruudessa. Siinä ketjun pisteet asetetaan peräkkäin ja yhdistetään viivalla (Robert & Casella, 2009).

Ketjun pisteiden avulla muodostetaan simuloitu tiheysjakauma. Tiheysjakauman kuvaajaa tarkastelemalla voidaan havaita ketjun huonoa käytöstä. Huonosti käyttäytyvän ketjun tiheysjakauman kuva voi näyttää hyvin epätasaiselta ja piikikkäältä.

Ketjuihin ja niiden tarkasteluun liittyy puuttuvan tiheyden ongelma (missing mass problem). Ongelma syntyy, koska Markovin ketjun avulla voi tarkastella vain arvoja, joissa ketju on käynyt. Puhutaan "olet nähnyt vain sen, missä olet käynyt" (you've only seen where you've been)-ilmiöstä (Robert & Casella, 2009). Ilmiötä ehkäistään käyttämällä useaa rinnakkaista ketjua, joille annetaan toisistaan poikkeavat alkuarvot. Ketjujen jälkikuvia arvioimalla nähdään, jos ketjut vierailevat erialueilla.

3.4 Tilasto-ohjelmat

Lopputyön tilastolliset analyysit suoritettiin pääosin R-ohjelmalla ja sen eri lisäpaketeilla. Käytössä oli myös JAGS Bayes-analyysin suorittamiseen. Alkuvaiheen datan käsittelyssä käytettiin myös Exceliä, mutta käyttö oli vähäistä.

3.4.1 R

R on avoimen lähdekoodin ohjelmointikieli ja ympäristö tilastollisen laskennan suorittamiseen. R-kieli perustuu Beckerin, Chambersin ja Wilksin S-kielen, sekä Sussmanin Schema-kielen (Crawley, 2007). Kielen syntaksia pi-

detään helppona oppia myös sellaisille käyttäjille, joilla ei ole aiempaa kokemusta koodaamisesta.

R on suosittu työkalu tilastollisen analyysin suorittamiseen. Osasyys suosioon johtuu lukuisista lisäpaketeista, joita R-ohjelmaa varten on tehty ja tehdään edelleen. Lisäpakettien avulla R-ohjelmasta saa muokattua useaan tarkoitukseen sopivan.

3.4.2 JAGS

JAGS on avoimen lähdekoodin ohjelma bayesiläisen analyysin suorittamiseen käyttäen MCMC-metodiikkaa ja Gibbsin otantaa. JAGS on lyhennelmä sanoista "Just Another Gibbs Sampler".

JAGS on helposti käytettävissä R-ohjelman lisäpaketin runjags avulla, joka mahdollistaa koko analyysin suorittamisen R ympäristössä.

3.5 Hierarkkinen Bayes-malli

Mallin tavoitteena on kuvata lohen nousun dynamiikkaa ja ympäristötekijöiden vaikutusta ilmiöön mahdollisimman realistisesti. Malli on yhdistelmä kolmesta erillisestä mallista, jotka hyödyntävät toistensa tuloksia. Mallin vaiheet ja eri vaiheisiin vaikuttavat muuttujat esitellään kuvassa 3.

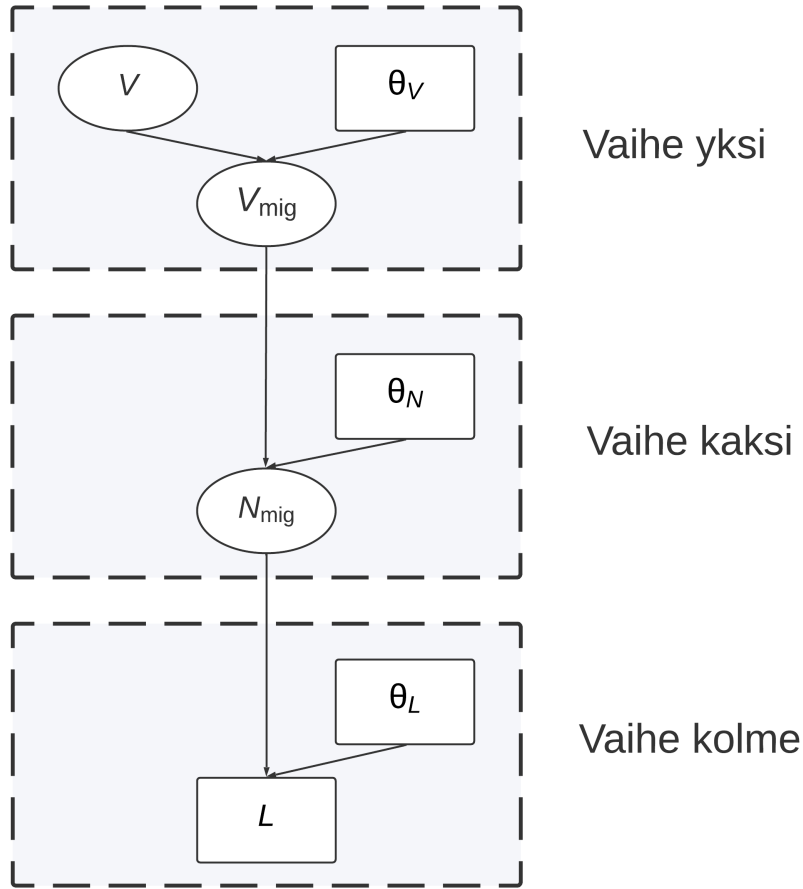
Mallin vasteena on lohen nousun intensiteetti lohien lukumäärä tunnissa ja selittävinä tekijöinä vuoroveden vaihe ja kalastuksen säätely. Mallia varten lohen nousu Tenonvuonolta kaikuluotaimelle on jaettu kolmeen eri vaiheeseen. Edellisen vaiheen tulosta hyödynnetään seuraavassa vaiheessa. Mallin aika-askel on yksi tunti ja analyysissä käytetty ajanjakso on vuoden 2019 viikko 25 maanantaista perjantaihin. Viikko 25 valittiin, koska tällöin lohen nousu on aktiivisimmillaan eikä joen virtaamassa tai lämpötilassa tapahdu suuria muutoksia.

Lohen nousua kuvaavaa ilmiöltä voi ajatella usean tankin järjestelmänä. Tankkeja yhdistävät venttiilit. Ajan kuluessa lohi voi oleilla samassa tan-

kissa tai siirtyä venttiilin kautta toiseen tankkiin. Kun lohi siirtyy tankista toiseen, se lähestyy kaikuluotainta. Venttiilien avonaisuutta pyritään arvioimaan mallin avulla. Avonaisuus rajoittaa lohien siirtymistä tankkien välillä. Selittävät tekijät vaikuttavat avonaisuuteen.

Eri vaiheiden (tankki-venttiili-pari) toimintaperiaate on sama. Kuhunkin tankkiin tulee lohia ja sieltä lähtee lohia. Tankkissa hetkellä t olevien lohien määrä muodostetaan tulevien ja lähtevien lohien erotuksena. Lähtevien lohien määrä saadaan siirtymän todennäköisyyden $[0, 1]$ (venttiilin avonaisuus) ja tankkissa olevan lohimäärän tulona.

Alkuun tankit ovat tyhjiä. Ajan kuluessa lohet alkavat liikkua tankkien välillä. Lohet voivat jäädä oleilemaan tankkeihin ja syntyy viivettä. Viive kuvaa lohien vaellukseen kuluvaa aikaa.



Kuva 3: Mallin rakennetta kuvaava suunnattu syklitön verkko. Verkossa on esitetty eri vaiheet ja niihin vaikuttavat muuttujat. Nelikulmion kuvaa muuttujaa, josta on aineistoa ja pyöreä muoto muuttujaa, josta ei ole aineistoa. Vaiheeseen yksi vaikuttavat V , joka on Tenonvuonolla oleva lohimäärä ja θ_V , joka on vuoroveden vaikutus. Vaiheessa yksi ratkaistaan muuttujan V_{mig} yhteisjakauma muuttujan θ_V ja V kanssa. Muuttuja V_{mig} kuvaa vuonolta jokeen siirtyvää lohimäärää. Vaiheeseen kaksi vaikuttaa vaiheessa yksi ratkaistun V_{mig} yhteisjakauma ja θ_N , joka on pyydyskalastuksen vaikutus lohen aktiivisuuteen. Vaiheessa kaksi ratkaistaan muuttujan N_{mig} yhteisjakauma muuttujan θ_N ja edellisen vaiheen muuttujien kanssa. Muuttuja N_{mig} kuvaa luotauspaikalle siirtyvän lohimäärää. Vaiheeseen kolme vaikuttaa vaiheessa kaksi ratkaistu N_{mig} yhteisjakauma ja θ_L , joka on pyydyskalastuksen vaikutus lohen saaliiksi jäämiseen. Vaiheesta kolme ratkaistaan muuttujien posteriojakauma aineistoon L perustuen.

3.5.1 Vaihe yksi: Siirtymä Tenonvuonolta Tenojokeen

Vaihe yksi kuvastaa Tenonvuonoa. Lohet tulevat vuonolle Barentsinmereltä ja siirtyvät Tenojokeen. Siirtymään vuonolta joelle vaikuttaa vuorovesi:

$$p(V_{\text{mig}}, \theta_V, V) = p(\theta_V)p(V)p(V_{\text{mig}}|\theta_V, V), \quad (9)$$

jossa $p(\theta_V)$ on vuoroveden vaikutuksen priorijakauma ja $p(V)$ on vuonolla olevan lohimäärän jakauma. Siirtyvän lohimäärän, vuoroveden vaikutuksen ja vuonolla olevan lohimäärän yhteisjakauma on $p(V_{\text{mig}}, \theta_V, V)$.

Tenovuonolla olevan lohimäärän oletetaan noudattavan log-normaalijakaumaa:

$$V[t] \sim \text{LogN}(\log(15) - 0.5 \times \log(0.2^2 + 1), \quad 0.5 \times \log(0.2^2 + 1)), \quad (10)$$

jossa $V[t]$ on vuonolla oleva lohimäärä, jonka odotusarvo on 15 ja variaatio-kerroin on 0.2.

Vuoroveden vaikutus on jaettu kahteen osaan: nousuveteen ja laskuveteen. Vaikutusmekanismi mukaillee lineaarisen lauseen kaavaa.

$$\begin{aligned} \psi[t] &\sim N(\psi_0[t] \times (1 - \text{tide_i}[t]) + \psi_1[t] \times \text{tide_i}[t], \quad \sigma_V) \\ \psi_0[t] &= \alpha_0 + \beta_0 \times \text{tide}[t] \\ \psi_1[t] &= \alpha_1 + \beta_1 \times \text{tide}[t], \end{aligned} \quad (11)$$

jossa $\psi[t]$ on vuoroveden vaikutus siirtymään, $\text{tide}[t]$ on vuorovesimuuttujan arvo hetkellä t . Indikaattori-muuttuja $\text{tide_i}[t]$ erottaa nousuveden ja laskuveden toisistaan. Parametrit α_0 , β_0 , α_1 ja β_1 kuvaavat vuoroveden vaikutusta. Laskuveden vaikutusta kuvaavat α_1 ja β_1 , nousuveden vaikutusta α_0 ja β_0 .

Vuoroveden vaikutus $\psi[t]$ on välimatka-asteikollinen. Vaihtelun lisäämiseksi suureen $\psi[t]$ oletetaan noudattavan normaalijakaumaa keskihajonnalla σ_V .

Suurelle $\psi[t]$ tehdään expit-muunnos, jolla vaikutus muunnetaan $[0, 1]$ yksikköväliä.

$$\theta_V[t] = \text{expit}(\psi[t]), \quad (12)$$

jossa $\theta_V[t]$ on vuonosta jokeen siirtymän todennäköisyys ja expit-muunnos on logit-muunnoksen käänteismuunnos.

Siirtymän todennäköisyyden $\theta_V[t]$ ja vuonon lohimäärän $V[t]$ avulla muodostetaan siirtyvän lohimäärän odotusarvo, joka noudattaa Poissonin jakaumaa.

$$\begin{aligned} V_{\text{mig}}[t] &\sim \text{Poisson}(\mu_V[t]) \\ \mu_V[t] &= \theta_V[t] \times V[t], \end{aligned} \quad (13)$$

jossa $V_{\text{mig}}[t]$ on vuonosta jokeen siirtyvä lohimäärä ja $\mu_V[t]$ on siirtyvän lohimäärän odotusarvo.

3.5.2 Vaihe kaksi: Siirtymä Tenojoessa kaikuluotauspaikalle

Vaihe kaksi kuvastaa matkaa Tenojokea pitkin kaikuluotauspaikalle ennen havaituksi tulemistä:

$$p(N_{\text{mig}}, \theta_N, V_{\text{mig}}, \theta_V, V) = p(\theta_N)p(V_{\text{mig}}, \theta_V, V)p(N_{\text{mig}}|\theta_N, V_{\text{mig}}, \theta_V, V), \quad (14)$$

jossa $p(\theta_N)$ on pyydyskalastuksen vaikutuksen priorin ja $p(N_{\text{mig}}, \theta_N, V_{\text{mig}}, \theta_V, V)$ on kaikuluotauspaikalle siirtyvän lohimäärän, pyydyskalastuksen vaikutuksen, vuonolta jokeen siirtyneen lohimäärän ja vuoroveden vaikutuksen yhteisjakauma.

Lohien määrä muodostetaan vaiheessa Tenonvuonolta tulevien ja kaikuluotaimelle siirtyvien lohimäärien erotuksena.

$$N[t] = N[t - 1] + V_{\text{mig}}[t] - N_{\text{mig}}[t], \quad (15)$$

jossa $N[t]$ on Tenojossa oleva lohimäärä, V_{mig} on Tenonvuonolta tuleva lohimäärä ja $N_{\text{mig}}[t]$ on kaikuluotauspaikalle siirtyvä lohimäärä.

$N_{\text{mig}}[t]$ oletetaan noudattavan Poissonin jakaumaa ja se muodostetaan siirtymän todennäköisyyden ja joessa olevan lohimäärän avulla.

$$\begin{aligned} N_{\text{mig}}[t] &\sim \text{Poisson}(\mu_N[t]) \\ \mu_N[t] &= \theta_N[t] \times N[t - 1] \\ \theta_N[t] &= \theta_N^1 \times \text{vk}[t] + \theta_N^0 \times (1 - \text{vk}[t]), \end{aligned} \quad (16)$$

jossa $\theta_N[t]$ on siirtymän todennäköisyys ja $\text{vk}[t]$ on pyydyskalastusmuuttuja. θ_N^1 on siirtymän todennäköisyys pyydyskalastuksen ollessa sallittua ja θ_N^0 siirtymän todennäköisyys pyydyskalastuksen ollessa kiellettyä. μ_N on siirtyvän lohimäärän odotusarvo ja $N_{\text{mig}}[t]$ on joelta kaikuluotauspaikalle siirtyvä lohimäärä.

3.5.3 Vaihe kolme: Havaituksi tuleminen

Vaihe kolme kuvastaa kaikuluotaimella havaituksi tulemistä ja kalastuksesta aiheutuvaa kalojen saaliiksi jäämistä:

$$\begin{aligned} p(N_{\text{mig}}, V_{\text{mig}}, V | L, \theta_L, \theta_N, \theta_V) &\propto \\ p(\theta_L) p(N_{\text{mig}}, \theta_N, V_{\text{mig}}, \theta_V, V) p(L | \theta_L, N_{\text{mig}}, \theta_N, V_{\text{mig}}, \theta_V, V) \end{aligned} \quad (17)$$

jossa $p(\theta_L)$ on pyydyskalastuksen takia saaliiksi jäävän lohimäärän osuuden prioriksi. Aineistosta saatu uskottavuusfunktio on $p(L | \theta_L, N_{\text{mig}}, \theta_N, V_{\text{mig}}, \theta_V, V)$ ja $p(N_{\text{mig}}, V_{\text{mig}}, V | L, \theta_L, \theta_N, \theta_V)$ on parametrien posteriorijakauma.

Jokainen kaikuluotaimelle siirtyvä lohi tulee havaituksi, ellei se päädy saaliiksi. Saaliiksi jääneiden lohien määrä muodostetaan joesta siirtyvien lohien ja saaliiksi jäämisen todennäköisyyden avulla.

$$\begin{aligned} L_r[t] &= \theta_L[t] \times N_{\text{mig}}[t] \\ \theta_L[t] &= \theta_L^1 \times \text{vk}[t] + \theta_L^0 \times (1 - \text{vk}[t]), \end{aligned} \tag{18}$$

jossa θ_L^1 on saaliiksi jäämisen todennäköisyys pyydyskalastuksen olleessa sallittua ja θ_L^0 saaliiksi jäämisen todennäköisyys pyydyskalastuksen ollessa kiellettyä. $L_r[t]$ on saaliiksi jäänyt lohimäärä.

Havaittu lohimäärä muodostetaan saaliiksi jääneen lohimäärän ja kaikuluotaimelle saapuvan lohimäärän erotuksena. Lohimäärän oletetaan noudattavan Poissonin jakaumaa.

$$\begin{aligned} L[t] &\sim \text{Poisson}(\mu_L[t]) \\ \mu_L[t] &= N_{\text{mig}}[t] - L_r[t], \end{aligned} \tag{19}$$

jossa $\mu_L[t]$ on kaikuluotaimella havaittavan lohimäärän odotusarvo ja $L[t]$ on kaikuluotaimella havaittu lohimäärä.

3.6 Priorit

Prioritiheyksien muodostuksessa hyödynnettiin sekä aiempaa tutkimustietoa että asiantuntija-arviota. Kaikissa tilanteissa informatiivisen prioritiheyden muodostaminen ei ollut mahdollista, koska aiempaa tietoa ei ollut olemassa, tai sitä oli vaikeaa muuntaa matemaattiseen formaattiin. Prioritieto saatiin pääasiassa aiemmista tutkimustuloksista, mutta pyydyskalastuksen kohdalla hyödynnettiin asiantuntija-arviota.

Vuoroveden vaikutuksen priorijakaumiin on hankala soveltaa aiempaa tutkimustietoa, johtuen itseluodusta vuorovesimuuttujasta. Aiemman tutkimus-

tiedon mukaan vuoroveden vaihe vaikuttaa lohen aktiivisuuteen. Lohi on aktiivisimmillaan vuoroveden huipun, sekä laskuveden aikaan (Erkinaro ym., 1999b).

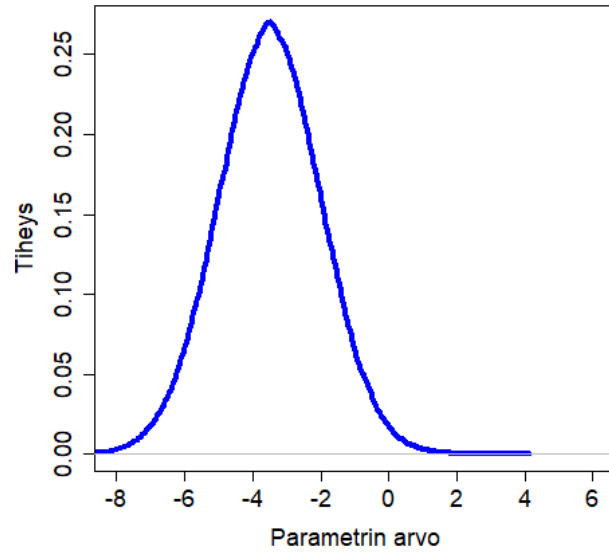
Priorijakaumat muodostettiin tutkien vuorovesimuuttujan arvoista aiheutuva siirtymän todennäköisyyttä perustuen parametrien α, β ja σ_V priorijakaumiin. Tarkoitusta varten piirrettiin kuvaaja (kuva 7). Kuvaajassa x-akselilla on vuorovesimuuttujan arvot ja y-akselilla siitä aiheutuva siirtymän todennäköisyys.

Parametrin β avulla säädettiin siirtymän todennäköisyyden kasvun jyrkkyyttä ja parametrilla α arvoaluetta, jolla kasvu tapahtuu. Parametrilla σ_V lisättiin epävarmuutta. Parametrin σ_V prioritiheyttä kuvataan typistetyllä jakaumalla. Typistetty jakauma saa arvoja vain ennaltamääritetyllä arvoalueella. Typistystä merkitään lisäämällä $T(\text{alaraja}, \text{yläraja})$ jakauman määrittelyn päätteeksi. Esimerkiksi standardinormaalijakaumaa, joka typistetään välille $[0, 10]$, merkittäisiin $N(0, 1)T(0, 10)$. Jos ylä- tai alarajana on äärettömyys, kyseistä ylä- tai alarajaa ei merkitä.

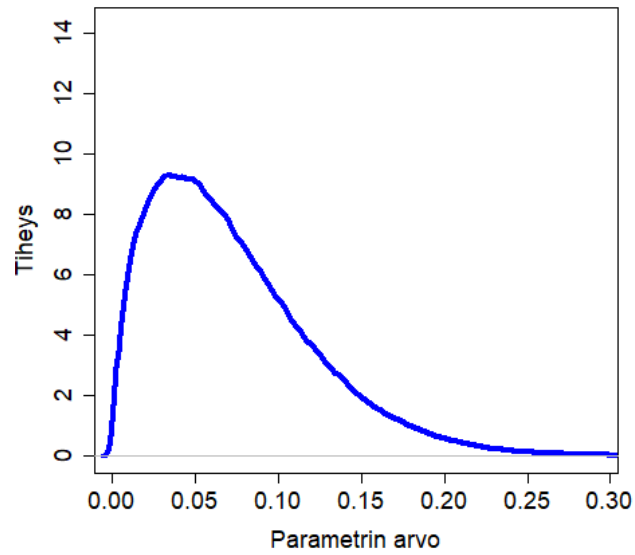
Epävarmuus luotiin parametrien α, β ja σ_V priorijakaumien avulla. Epävarmuuden kuvaamiseksi käytettiin laatikkokuvioita. Laatikkokuviossa laatikko kuvaa arvojen kvartiiliväliä. Laatikossa oleva poikkiviiva on arvojen mediaani. Laatikon päistä alkavat janat kuvaavat muuttujan arvoaluetta. Epävarmuuden ajatellaan olevan suurimmillaan muuttujana arvojen keskivaiheella. Muuttujan arvojen ääripäissä epävarmuus suppenee, mutta pysyy kohtuullisen laajana.

Nousuvettä ja laskuvettä kuvaaville muuttujille oletettiin samat priorit. Nousuveden aikana, vuorovesimuuttujan arvot nousevat, mutta eivät saavuta maksimiarvoa 100. Laskuveden aikana vuorovesimuuttuja arvot laskevat.

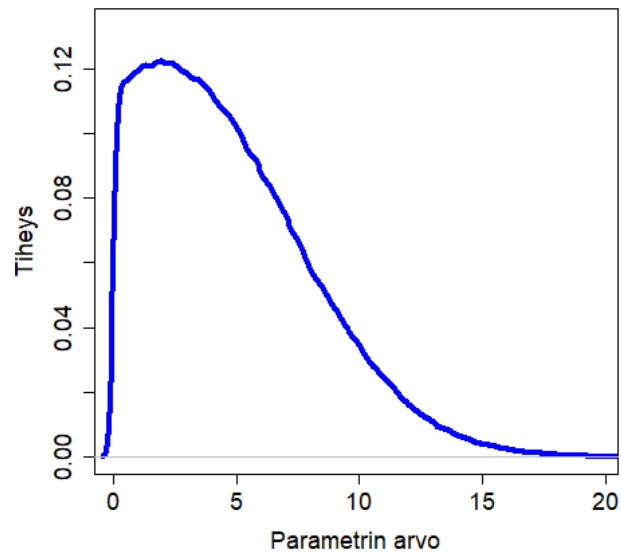
$$\begin{aligned}\alpha &\sim N(-3.5, 3) \\ \beta &\sim \text{Weibull}(1.5, 0.08) \\ \sigma_V &\sim N(2, 5)T(0,)\end{aligned}\tag{20}$$



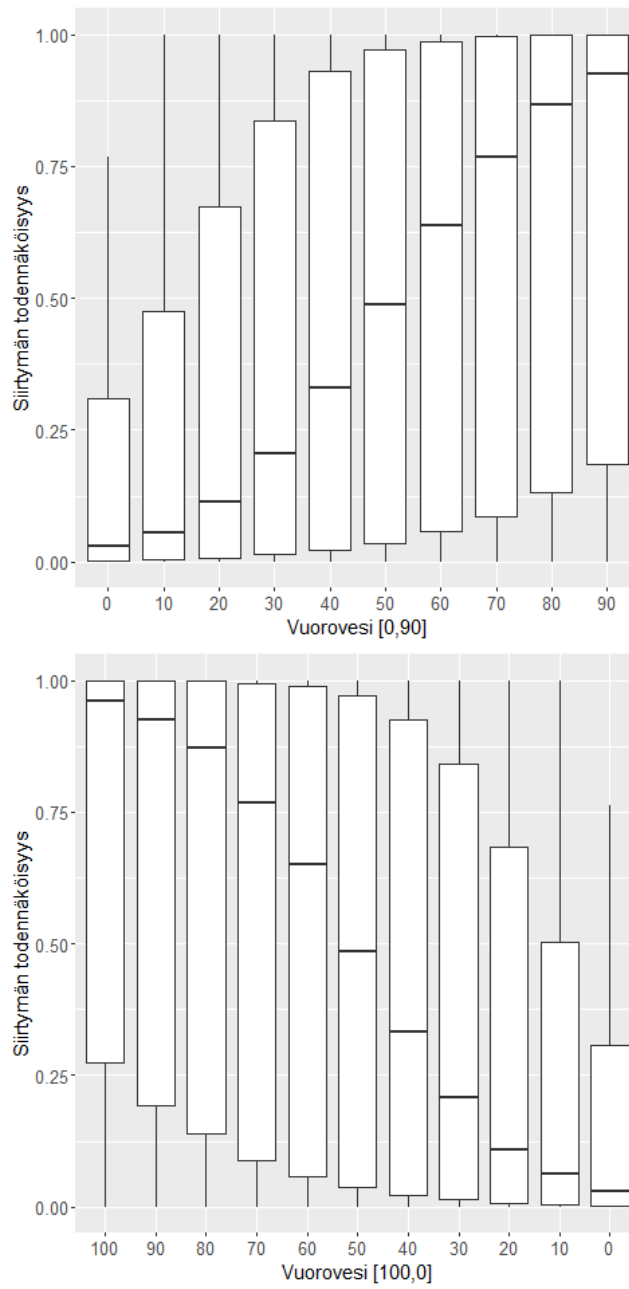
Kuva 4: Parametrin α prioritiheys. Priori on epäinformatiivinen ja sen valitsemisessa hyödynnettiin kuvaa 7



Kuva 5: Parametrin β prioritiheys. Priori on epäinformatiivinen ja sen valitsemisessa hyödynnettiin kuvaa 7



Kuva 6: Parametrin σ_V prioritiheys. Priori on epäinformatiivinen ja sen valitsemisessa hyödynnettiin kuvaa 7



Kuva 7: Vuorovedestä aiheutuva siirtymän todennäköisyys perustuen prioritiheyksiin. Yläpuolinen kuva kuvaa nousuvettä ja alapuolinen laskuvettä. Todennäköisyys alkaa nolasta ja siirtyy kohti yhtä, siten että vuorovesimuuttujan arvon ollessa 50 todennäköisyys on noin 0.5.

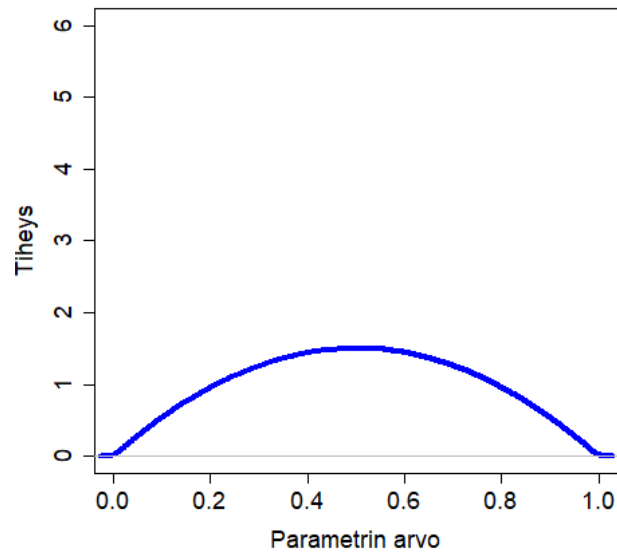
Kuvaa 7 hyödynnettiin parametrien α , β ja σ_V priorien valinnassa. Edellä mainittujen parametrien priorit valittiin siten, että kuvaan saatiin laaja epävarmuus.

Pyydyskalastuksen liittyvät priorit muodostettiin oletukselle, että pyydyskalastuksen ollessa sallittua lohen aktiivisuus on pienempää. Tätä varten priorijakaumien välille muodostettiin riippuvuussuhde hyödyntäen ylimääräistä muuttujaa θ^* , jolle asetetaan oma priorijakauma.

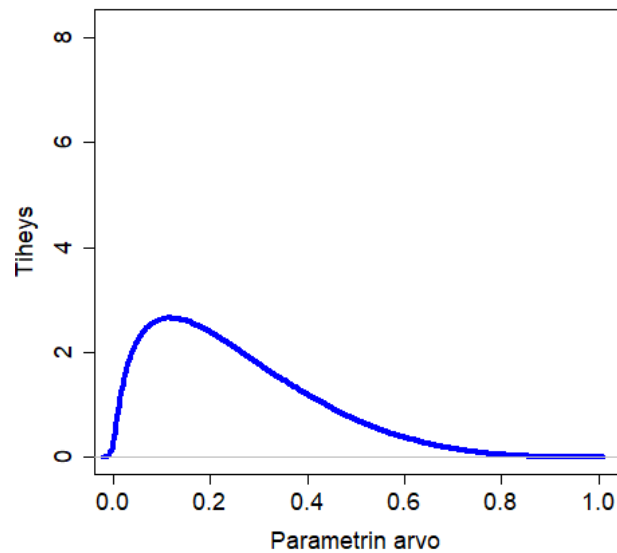
Tilanteessa, jossa pyydyskalastus on kiellettyä θ_N^0 , priorin on laaja ja epävarma se kattaa koko arvoalueen $[0, 1]$. Kun pyydyskalastus on sallittua θ_N^1 , priorin halutaan painottuvan pienempiin arvoihin. Muuttujalla θ^* kerrotaan parametrin θ_N^0 prioria ja saadaan parametrin θ_N^1 priorin.

$$\begin{aligned}\theta_N^0 &\sim \text{Beta}(2, 2) \\ \theta^* &\sim \text{Beta}(2, 2) \\ \theta_N^1 &= \theta_N^0 \times \theta^*\end{aligned}\tag{21}$$

Kuva A

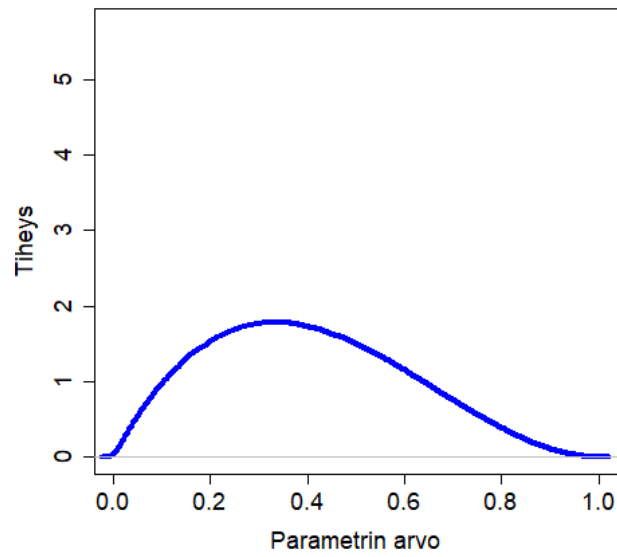


Kuva B



Kuva 8: Prioritiheydet pyydyskalastuksen vaikutukselle. Kuva A kuvaa tilannetta, jossa pyydyskalastus on kiellettyä θ_N^0 ja kuva B tilannetta, jossa pyydyskalastus on sallittua θ_N^1 . Kuvan A priori on epäinformatiivinen ja kattaa koko parametrin arvoalueen $[0, 1]$. Kuvan B priori perustuu oletukseen, että pyydyskalastus haittaa lohien liikkuemista, jonka takia se painottaa arvoalueen pienempiä arvoja.

Aiempien tutkimustulosten perusteella kalastettujen kalojen osuus Tenojessa on noin 40% (Erkinaro ym., 1999a). Parametrin θ_L^1 prioritiheys muodostettiin tämän perusteella. Lähes kaiken lohisaaliin oletetaan tulevan pyydyskalastuksen kautta.



Kuva 9: Priorijakauma saaliiksi jäävien kalojen osuudelle θ_L^1 . Priori perustuu aiempiin tutkimustuloksiin (Erkinaro ym., 1999a).

4 Tulokset

Tässä luvussa esitellään analyysin tulokset. Saatujen posteriorijakaumien tiheysfunktiot esitetään ja niihin liittyviä jälkikuvia tarkastellaan. Analyysin tulokset esitellään vaihe kerrallaan. Kappaleen lopussa esitellään yleisiä huomioita tuloksista.

4.1 Analyysin suorittaminen

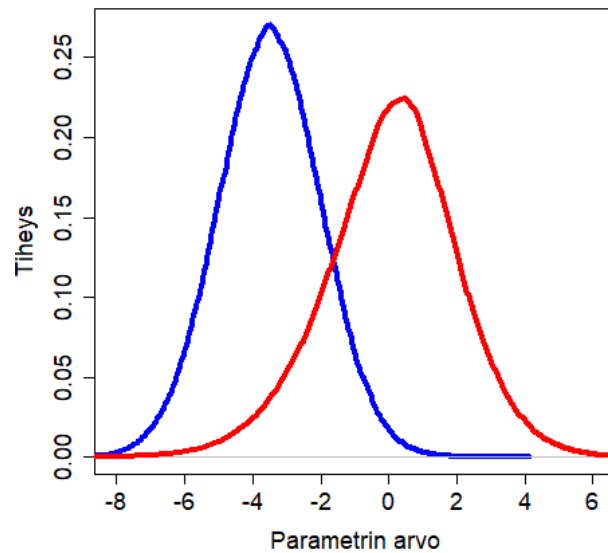
Analyysin suorittaminen vie paljon aikaa suuremmilla havaintomäärillä. Tätä tutkielmaa varten analyysissä käytettiin viiden päivän jaksoa maanantaista perjantaihin viikolla 25. Jaksoon sisältyy 120 havaintoa. Posteriorijakaumasta otettiin 200000 suuruinen otos, käyttäen ohennusta vakiolla 30. Otoksen suuruus oli kokonaisuudessaan $200000 \cdot 30 = 6000000$ kappaletta. Otoksen tuottamiseen kului aikaa noin viisi vuorokautta. Analyysin suorittamiseen käytettiin erityisesti MCMC-ketjun laskentaa varten rakennettua tietokonetta.

4.2 Vaiheen yksi tulokset

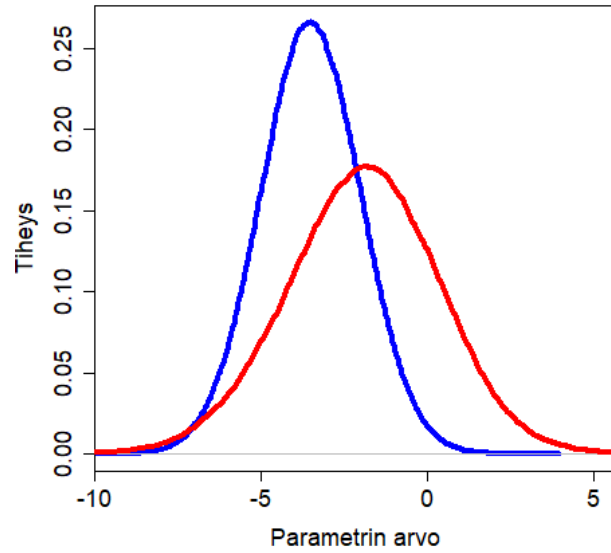
Vaiheessa yksi estimoitiin vuoroveteen liittyvien parametrien α , β ja σ_v posteriorijakaumat. Posteriorijakaumat esitetään graafisesti ja niitä verrataan priorijakaumiin. Lisäksi esitetään graafinen esitys parametrien vaikutuksesta siirtymän todennäköisyyteen kummassakin tilanteessa.

Taulukko 1: Vaiheessa yksi estimoitujen parametrien posteriorikeskiarvot ja 95% posteriorivälit. Tunnusluvut arvioitiin 200000 pituisesta MCMC-ketjusta.

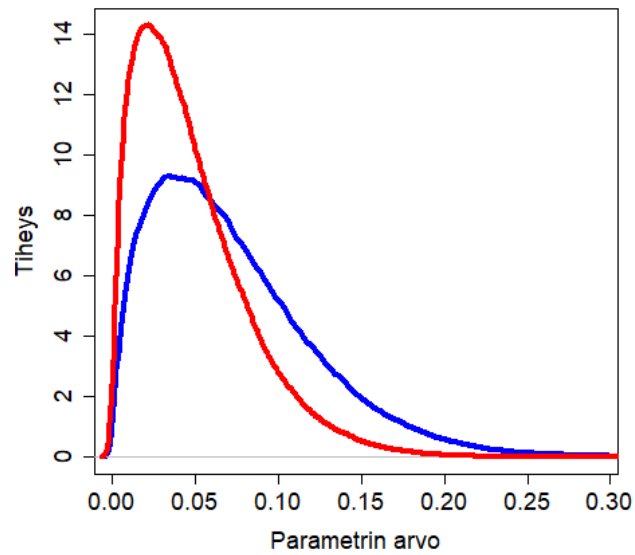
	Posteriorikeskiarvo	95 % Posterioriväli
α_0	0.08	[-4.00, 3.76]
α_1	-1.96	[-6.57, 2.46]
β_0	0.05	[0.00, 0.13]
β_1	0.09	[0.02, 0.19]
σ_v	8.29	[3.23, 15.00]



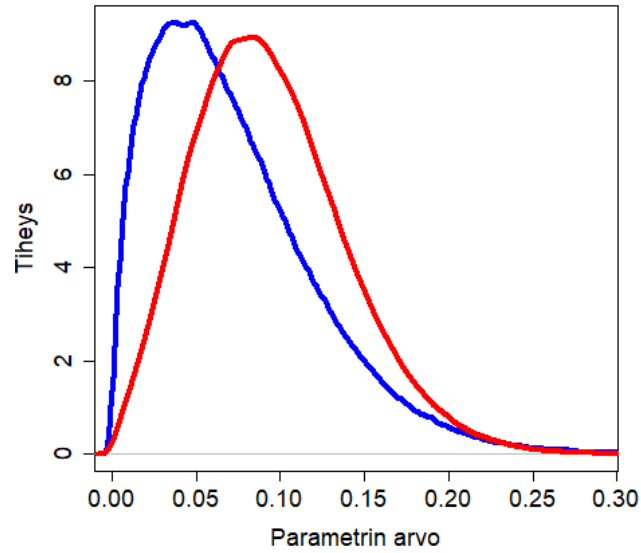
Kuva 10: Parametrin α_0 priori- ja posterioritiheys. Prioritiheyttä on merkitty sinisellä viivalla ja posterioritiheyttä punaisella.



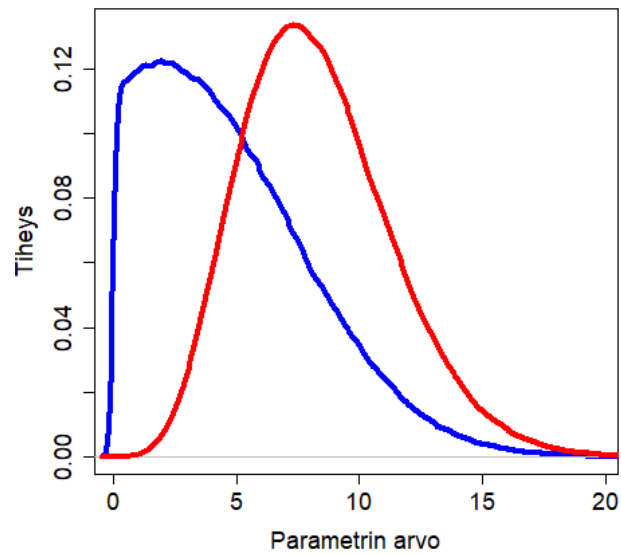
Kuva 11: Parametrin α_1 priori- ja posteriorijakauma. Prioritiheyttä on merkitty sinisellä viivalla ja posterioritiheyttä punaisella.



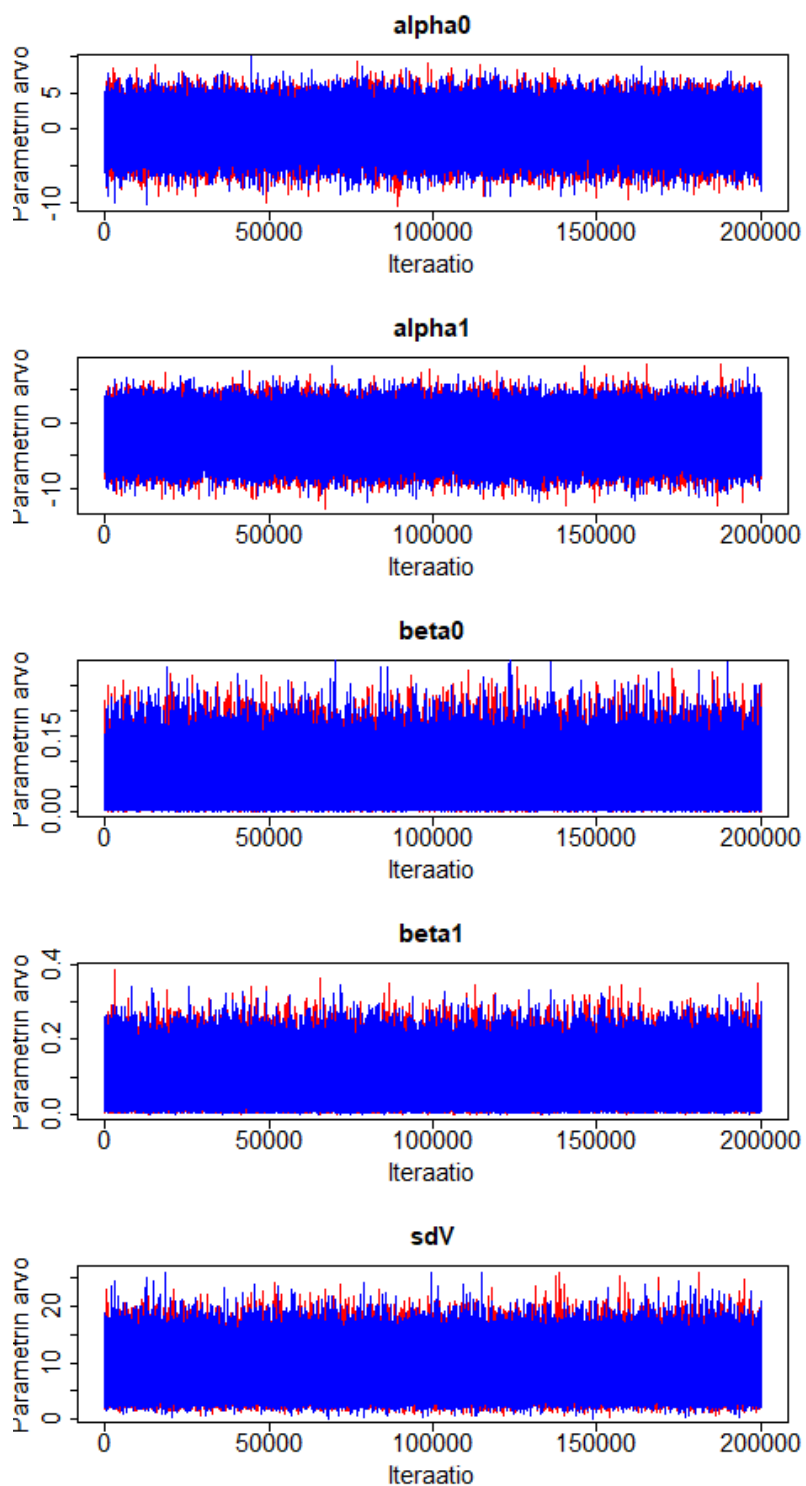
Kuva 12: Parametrin β_0 priori- ja posteriorijakauma. Prioritiheyttä on merkitty sinisellä viivalla ja posterioritiheyttä punaisella.



Kuva 13: Parametrin β_1 priori- ja posteriorijakauma. Prioritiheyttä on merkitty sinisellä viivalla ja posterioritiheyttä punaisella.



Kuva 14: Parametrin σ_V priori- ja posteriorijakauma. Prioritiheyttä on merkitty sinisellä viivalla ja posterioritiheyttä punaisella.



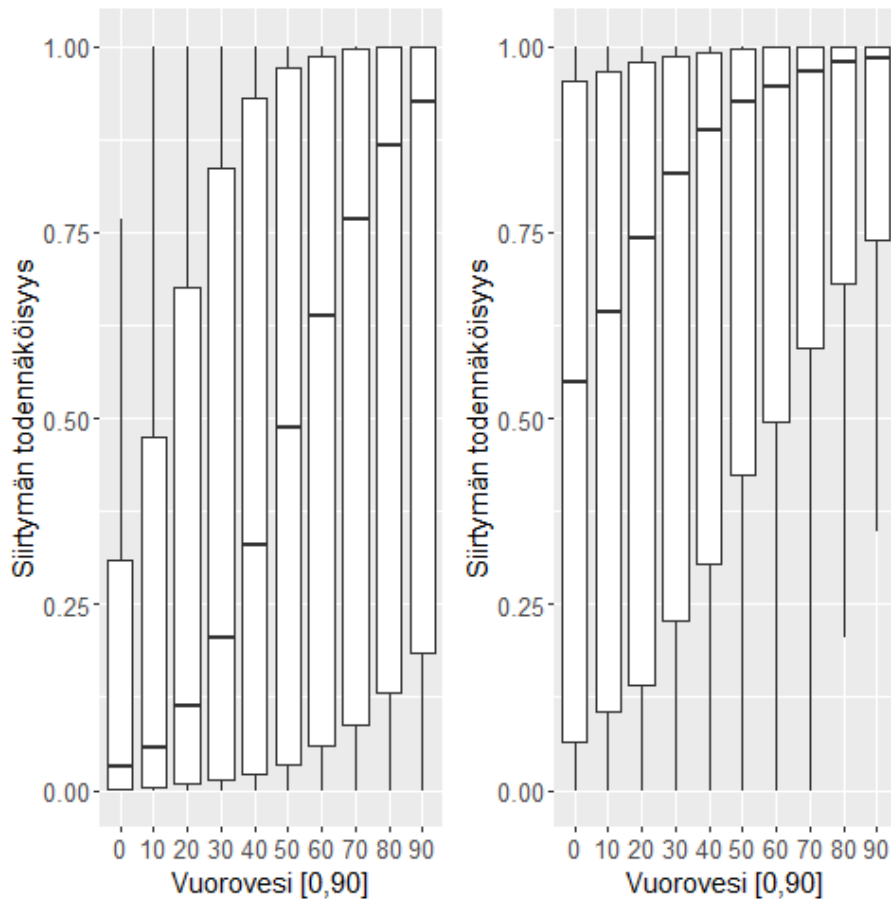
Kuva 15: Parametrien α_0 , α_1 , β_0 , β_1 ja σ_V jälkikuvat.

Kuvasta 10 ja 11 nähdään parametrien α_0 ja α_1 posteriorijakaumien päivittyneen oikealla, suhteessa priorijakaumaan. Nousuveteen liittyvä α_0 näyttää päivittyneen oikealla voimakkaammin kuin laskuveteen liittyvä α_1 .

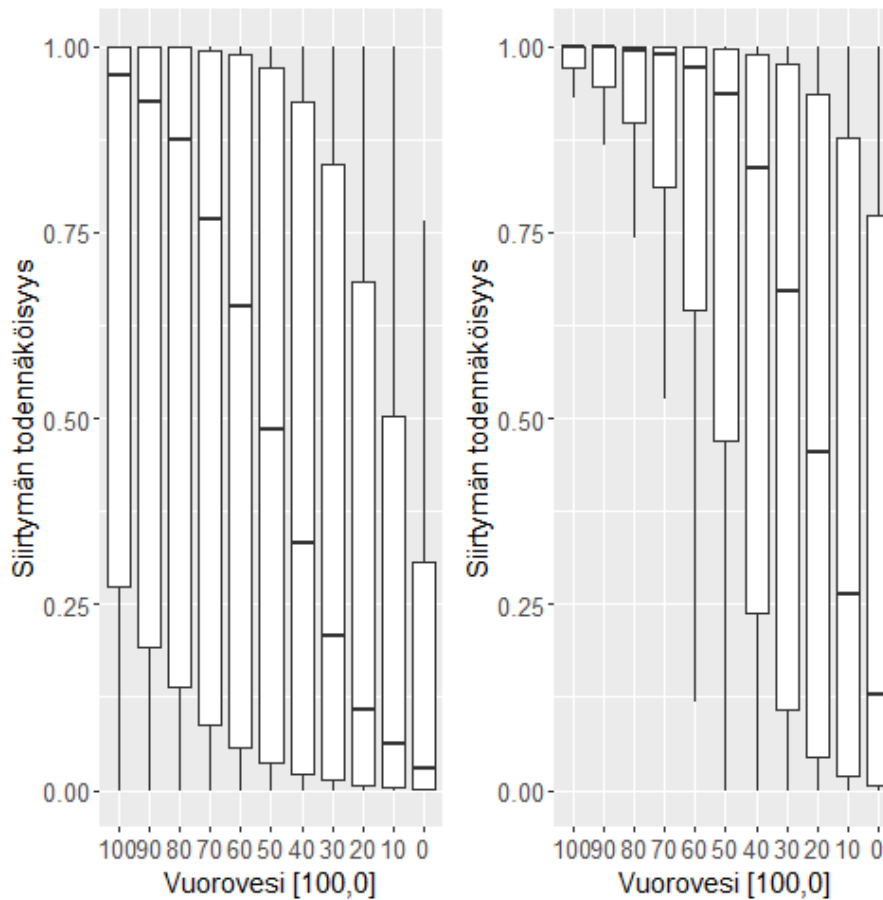
Kuvien 12 ja 13 perusteella, parametrien β_0 ja β_1 posteriorijakaumat ovat päivittyneet toisistaan poiketen. Nousuveteen liittyvä β_0 on päivittynyt lähemmäs nollaa. Laskuveteen liittyvä β_1 on päivittynyt priorijakaumaansa verrattuna oikealle, nolasta poispäin.

Kuvasta 14 nähdään vuoroveteen liittyvän hajontaparametrin σ_V päivittyneen prioristaan oikealle. Posteriori kattaa samankaltaisen alueen kuin prior, mutta todennäköisyysmassa sijaitsee enemmän suuremmilla arvoilla.

Kuvan 15 perusteella nähdään, että eri parametrien erilliset ketjut (liila ja sininen) ovat sekoittuneet, kattavat samankaltaisen arvoalueen kunkin parametrin kohdalla ja liikkuvat satunnaisesti arvoalueidensa sisällä. Tämän perusteella ketjut ovat konvergoineet ja simulaation pituus on ollut riittävä.



Kuva 16: Nousuveden aikana parametreista α_0 ja β_0 aiheutuvan siirtymän todennäköisyyden priori ja posteriori.



Kuva 17: Laskuveden aikana Parametreista α_1 ja β_1 aiheutuvan siirtymän todennäköisyyden priorin ja posteriorin.

Kun prioria verrataan posterioriin, kuvissa 16 ja 17, nähdään epävarmuuden pienentyneen vuorovesimuuttujan suurilla arvoilla ja lisääntyneen vuorovesimuuttujan pienillä arvoilla. Vuorovesimuuttujan arvon noustessa sekä laskuveden nousuveden posteriorit suppenevat siirtymän todennäköisyyden arvon 1 lähelle. Laskuveden posteriorissa suppeneminen on voimakkaampaa kuin nousuveden posteriorissa.

Kuvista 10, 11, 12, 13 ja 14 nähdään, että parametrien $\alpha_0, \alpha_1, \beta_0, \beta_1$ ja σ_V posteriorijakaumat ovat päivittyneet, mutta eivät juurikaan tarkentuneet. Posteriorijakaumat eroavat prioreistaan muodoltaan ja sijainniltaan, mutta jakaumaan liittyvä hajonta on pysynyt samankaltaisena kuin priorissa.

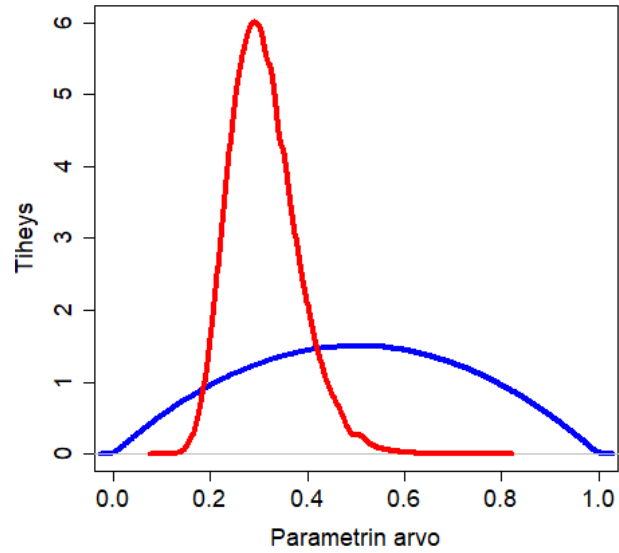
Saatujen tulosten perusteella vuorovedellä ja sen vaiheella näyttää olevan vaikutusta lohen nousun intensiteettiin. Vaikutus näyttää olevan suurempaa, mitä lähempänä vuoroveden huippua ollaan ajallisesti. Laskuvedestä aiheutuva vaikutus näyttää olevan voimakkaampaa kuin nousuvedestä aiheutuva.

4.3 Vaiheen kaksi tulokset

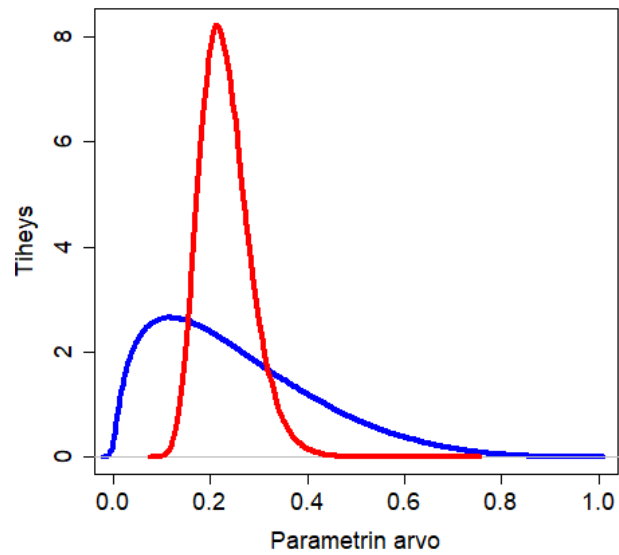
Vaiheessa kaksi estimoitiin verkkokalastuksen vaikutukseen liittyvien parametrien θ_0 ja θ_1 posteriorijakaumat. Posteriorijakaumat esitetään graafisesti ja niitä verrataan priorijakaumiin. Posteriorijakaumia verrataan myös keskenään.

Taulukko 2: Vaiheessa kaksi estimoitujen parametrien posteriorikeskiarvot ja 95% posteriorivälit. Tunnusluvut arvioitiin 200000 pituisesta MCMC-ketjusta.

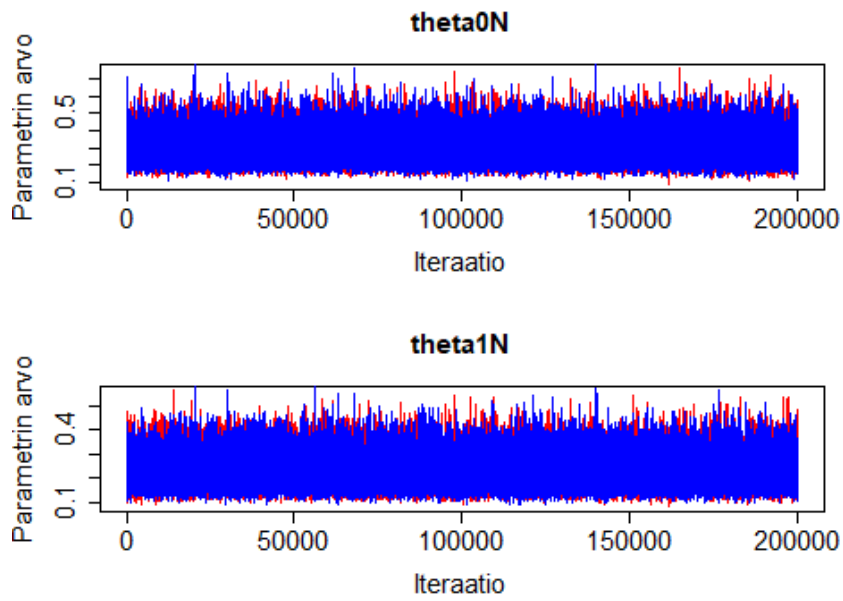
	Posteriorikeskiarvo	95 % Posterioriväli
θ_N^0	0.31	[0.19, 0.47]
θ_N^1	0.23	[0.15, 0.35]



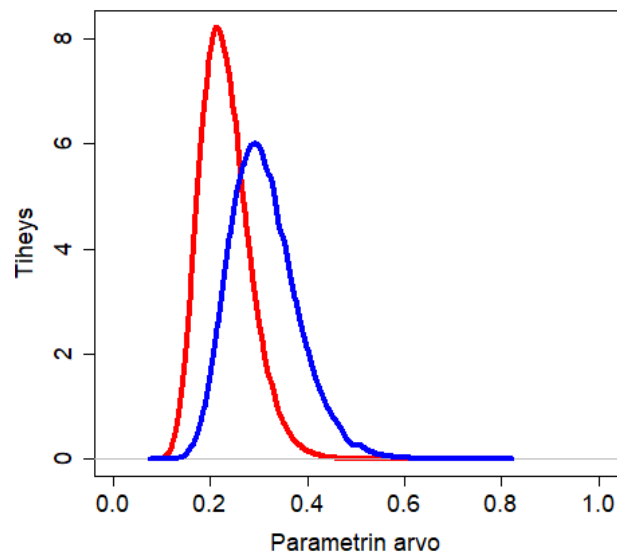
Kuva 18: Parametrin θ_N^0 priori- ja posterioritiheys. Prioritiheyttä on merkitty sinisellä viivalla ja posterioritiheyttä punaisella.



Kuva 19: Parametrin θ_N^1 priori- ja posterioritiheys. Prioritiheyttä on merkitty sinisellä viivalla ja posterioritiheyttä punaisella.



Kuva 20: Parametrien θ_N^0 ja θ_N^1 jälkikuvat.



Kuva 21: Parametrien θ_N^0 ja θ_N^1 posterioritiheydet. Parametria θ_N^1 on merkitty punaisella ja parametria θ_N^0 sinisellä.

Kuvasta 18 nähdään parametrin θ_N^0 päivittyvän pääosin arvon 0.5 alapuolelle. Posteriori on tarkentunut verrattuna laajaan ja epävarmaan prioriinsa.

Kuvan 19 perusteella parametrin θ_N^1 posteriorijakauma on tarkentunut ja päivittynyt priorijakaumansa huipusta oikealle, lähemmäs arvoa 0.5.

Kuvan 20 jälkikuviosta nähdään erillisten ketjujen sekoittuneen, kattavan samankaltaisen arvoalueen ja liikkuvan satunnaisesti arvoalueensa sisällä. Tämän perusteella ketjut ovat konvergoineet ja simulaation pituus on ollut riittävä.

Kun parametrin θ_N^0 ja θ_N^1 posteriorijakaumia verrataan keskenään kuvassa 21, nähdään molempien saavan arvoja pääasiassa väliltä $[0.1, 0.5]$. Parametrin θ_0 posteriori jakauman huippu on lähempänä arvoa 0.5, kun taas parametrin θ_1 posteriorin huippu on lähempänä arvoa 0.1.

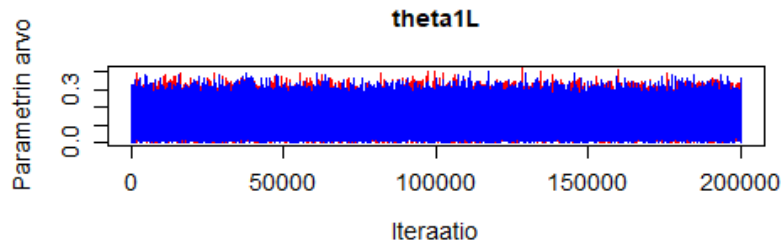
Saatujen tulosten perusteella kalastuksesta aiheutuva vaikutus lohen aktiivisuuteen ei näytä eroavan suuresti eri kalastusmuotojen välillä. Posteriorijakaumissa on kuitenkin havaittavissa eroja, mutta ne ovat pieniä ja voivat selittyä esimerkiksi prioritiheyksien erolla.

4.4 Vaiheen kolme tulokset

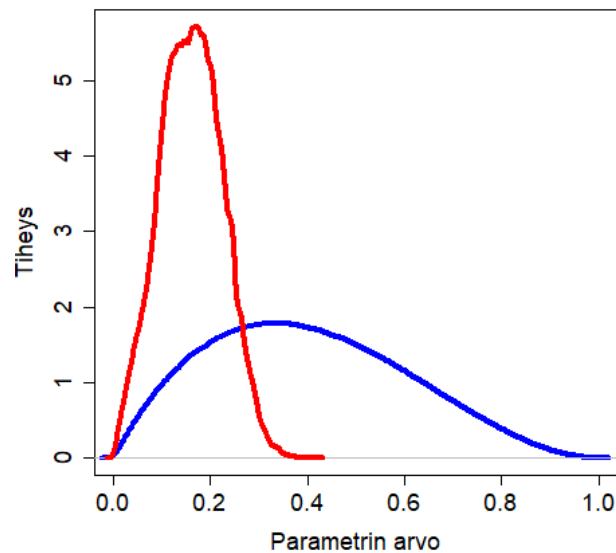
Vaiheessa kolme estimoitiin verkkokalastuksesta johtuvaa kalojen poistumista. Poistuneiden kalojen osuutta kuvaa parametri θ_L^1 . Posteriorijakauma esitetään graafisesti ja sitä verrataan priorijakaumaan.

Taulukko 3: Vaiheessa kolme estimoitujen parametrin posteriorikeskiarvot ja 95% posteriorivälit. Tunnusluvut arvioitiin 200000 pituisesta MCMC-ketjusta.

	Posteriorikeskiarvo	95 % Posterioriväli
θ_L^1	0.16	[0.04, 0.29]



Kuva 23: Parametrin θ_L^1 jälkikuva.



Kuva 22: Parametrin θ_L^1 priori- ja posterioritiheys. Prioritiheyttä on merkitty sinisellä viivalla ja posterioritiheyttä punaisella.

Kuvan 23 jälkikuvan perusteella ketjut ovat sekoittuneet, kattavat samankaltaisen arvoalueen ja liikkuvat satunnaisesti arvoalueensa sisällä. Tämän perusteella ketjut ovat konvergoineet ja simulaation pituus on ollut riittävä.

Kuvasta 22 nähdään parametrin θ_L^1 posteriorin tarkentuneen ja päivittyneen prioristaan vasemmalle. Posterioritiheys saa arvoja pääosin alueelta $[0, 0.3]$. Estimaatti joesta pyydyskalastuksen kautta poistuvien lohien osuudelle on 0.16.

5 Pohdinta

Lopputyön tavoitteena oli selvittää, onko vuoroveden vaiheella ja pyydyskalastuksella vaikutusta lohen nousun intensiteettiin nousun aktiivisella ajanjaksolla. Saatujen tulosten perusteella vuoroveden vaiheella on vaikutusta lohen nousun intensiteettiin. Laskuveden aikaan lohi on aktiivisempi kuin nousuveden aikaan ja aktiivisuus lisääntyy, kun vuoroveden huippu lähenee. Pyydyskalastuksella ei osoittautunut olevan merkittävää vaikutusta lohen nousun intensiteettiin.

Hyvän MCMC -otoksen tuottaminen on raskasta ja hidasta aineiston havaintomäärällä. Bayes-mallien ajamista varten rakennetulla tietokoneella hyvän otoksen tuottamiseen kuluu noin viisi vuorokautta. Hyvän otoksen tuottamiseen kuluva aika riippuu havaintomäärästä. Havaintomäärän kasvaessa otoksen tuottamiseen kuluva aika kasvaa lähes eksponentiaalisesti. Tämä on tyypillistä Bayes-malleille, koska havaintomäärän kasvaessa myös simulaation pituutta tulee kasvattaa. Aikaa voisi lyhentää vähentämällä havaintomäärää. Tämä voitaisiin saavuttaa esimerkiksi aggregoimalla dataa voimakkaammin.

Saadut tulokset ovat uskottavia, mutta osa tuloksista poikkeaa ennakkoojatuksista ja aiemmista tutkimuksista. Osasyynä tälle on todennäköisesti se, että koko nousun aktiivista jaksoa yritetään selittää viiden päivän aineiston perusteella. Olosuhteet voivat olla hyvin erilaiset viikkojen ja päivien välillä. Olosuhteiden vaihtelut vaikuttavat paljon lohen käytökseen ja aktiivisuuteen (Erkinaro ym., 1999b). Ongelmaa voisi yrittää ratkaista pidentämällä analyysissä käytettävää ajanjaksoa. Ajanjakson pidentäminen nostaisi analyysiin kuluvaa aikaa entisestään, jonka takia sitä ei tehty.

Tällä hetkellä malli on yhdistelmä kolmesta erillisestä mallista, jotka eivät yhtäaikaisesti riipu toisistaan. Jotta osat olisivat yhtäaikaisesti riippuvaisia toisistaan, yhteismalli (joint model) olisi ihanteellinen toteutus. yhteismallit ovat kuitenkin hyvin haastavia ja laskennallisesti raskaita, jonka vuoksi sitä ei toteutettu.

Malliin voisi lisätä selittäjiä, mahdollisia uusia selittäjiä olisivat esimerkiksi joen virtaama ja veden lämpötila. Aluksi nämä selittäjät oli tarkoituksena sisällyttää malliin, mutta ne jäivät pois, kun analyysissä käytettäväksi ajanjaksoksi valikoitui viisi päivää. Tällä ajanjaksolla kyseisissä muuttujissa ei

tapahdu suuria muutoksia eikä niiden lisääminen malliin olisi järkevää. Jos analyysissä hyödynnettäisiin pidempää ajanjaksoa, nämä selittäjät kannattaisi sisällyttää malliin.

Lohen luonnollinen parikäytös tulisi huomioida mallissa. Parvikäytös aiheuttaa hetkellistä ylidispersiota. Yhdessä mallin kehitysvaiheessa parvikäytös huomioitiin negatiivisen binomijakauman avulla (Lindén & Mäntyniemi, 2011), mutta tämä johti suppenemisiongelmiin.

Mallia tulisi myös tarkastella herkkyysanalyysillä. Tämä antaisi tärkeää tietoa siitä, miten mallin tulokset muuttuvat riippuen annetuista lähtöoletuksista ja prioritiheyksistä. Herkkyysanalyysin suorittaminen olisi hidasta, koska yhden otoksen tuottamiseen menee useita päiviä. Tästä syystä herkkyysanalyysiä ei sisällytetty tähän tutkielmaan.

Mallin avulla tulisi myös ennustaa lohen aktiivisuutta. Ennustettua aktiivisuutta verrattaisiin havaittuun ja tällä tavalla mallin toimivuutta voitaisiin tarkastella.

Yleisesti tarkasteltuna malli toimii ja tuottaa uskottavia tuloksia. Malli on hyvä lähtökohta tulevalle kehitykselle. Mielestäni analyysin suorittaminen pidemmällä ajanjaksolla olisi hyödyllistä. Näin saatuja tuloksia voitaisiin verrata lyhyemmän ajanjakson tuloksiin.

Toinen mallin hyödyntämistapa olisi suorittaa useampia analyysejä lyhyille ajanjaksoille. Esimerkiksi yksi analyysi kesän jokaiselta viikolta. Näiden tuloksien vertailu ja yhdistäminen toisi lisäinformaatiota ilmiöstä.

Viitteet

- Bayes, T. (1763) LII. An essay towards solving a problem in the doctrine of changes. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Royal Society*, 53, 370-418.
- Casella, G. George, E. I. (1992) Explaining the Gibbs Sampler. *The American Statistician*, 46, 167-174.
- Crawley, M. (2007) *The R book*. Wiley, Chichester.
- Erkinaro, J., Økland, F., Moen, K., Niemelä, E. & Rahiala, M. (1999a) Return migration of the Atlantic salmon in the Tana river: distribution and exploitation of radiotagged multi-sea-winter salmon. *Boreal Environment Research*, 4, 115-124.
- Erkinaro, J., Økland, F., Moen, K., Niemelä, E. & Rahiala, M. (1999b) Return migration of Atlantic salmon in the river Tana: the role of environmental factors. *Journal of Fish Biology*, 55, 506-516.
- Geman, S. & Geman, D. (1984) Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 720-741.
- Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.
- Karppinen, P., Erkinaro, J., Niemelä, E., Moen, K. & Økland, F. (2004) Return migration of one-sea-winter Atlantic salmon in the river Tana. *Journal of Fish Biology*, 64, 1179-1192.
- Lindén, A. & Mäntyniemi, S (2011) Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92, 1414-1421.
- Orell, P., Erkinaro, J., Svenning, M. A., Daviden, J. G. & Niemelä, E. (2007) Synchrony in the downstream of smolts and upstream migration of adult Atlantic salmon in the subarctic River Utsjoki. *Journal of Fish Biology* 71, 1735-1750.
- Pohjola, J-P., Orell, P., Kuusela, J. & Lilja, J. (2020) Teno- ja Inarijokeen nousevien lohien kaikkuluotauseurannat v. 2018-2019. *Luonnovara- ja Biotalouden Tutkimus 22/2020*. Luonnonvarakeskus, Helsinki.

- Robert, C. & Casella, G. (2009) *Introducing Monte Carlo Methods with R*. Springer, New York.
- van de Schoot, R., Denissen, J., Neyer, F., Kaplan, D., Asendorpf, J. & van Aken, M. (2013) A gentle introduction to Bayesian analysis: applications to developmental research. *Child Development*, 85, 842-860.