

Näslundin pituuskäyrään perustuvia Bayes-malleja puun
läpimitan ja pituuden väliselle riippuvuudelle

Joel Kostensalo

(Ohjaajat: Matti Vihola (JyU), Mari Myllymäki (Luke))

Pro gradu -tutkielma

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Kevät 2020

Tiivistelmä: J. Kostensalo *Näslundin pituuskäyrään perustuvia Bayes-malleja puun läpimitan ja pituuden väliselle riippuvuudelle*, pro gradu -tutkielma, 38 s., Jyväskylän yliopisto

Matematiikan ja tilastotieteen laitos

Kevät 2020.

Tässä tutkielmassa tutkittiin puiden pituusmallintamista Bayes-tilastotieteeseen perustuen. Aineistona käytettiin Parkanossa sijaitsevista mäntykokeista tehtyjä mittauksia. Metsikkö on perustettu vuonna 1929 ja puuston mittaukset on tehty keväällä 1963, syksyllä 1968 ja 1975, keväällä 1987 ja syksyllä 1999, 2006 ja 2011. Pituuden ja läpimitan todettiin olevan puutasolla korreloituneita mittauskertojen välillä. Toisin sanoen keskimääräistä ohuemmat tai paksummat puut olivat todennäköisesti myös tulevaisuudessa keskimääräistä ohuempia tai paksumpia. Niinpä puutason satunnaisvaikutus todettiin tarpeelliseksi mallinnuksessa, kun samasta puusta on käytössä useita mittauksia. Mallinnus perustui Näslundin pituuskäyrään, jonka parametrien tarkastelu ajan funktiona osoitti, että mallin vakioterminä toimivalla parametrilla b_0 ei ole selkeää aikariippuvuutta, mutta läpimitaan liittyvän parametrin b_1 arvon havaittiin pienenevän metsän varttuessa.

Mallien posteriorikeskiarvojen määräytymistä pienellä määrällä mittauksia tutkittiin valitsemalla aineistosta 100 kertaa kolme satunnaista puuta, joiden mittauksiin perustuen parametrien posteriorikeskiarvot määritettiin. Jopa hyvin heikon prioritiedon todettiin tarkentavan estimointia tasaamalla huonosta koepuiden valintaonnesta johtuvia poikkeuksellisia parametrien arvoja. Lisäksi Bayes-mallin pituusestimaattien todettiin tarkentuvan nopeasti, kun käytettävissä on myös aiempien vuosien mittauksia. Vertailukohtana käytetty linearisointiin perustuva malli osoittautui selvästi herkemmäksi poikkeaville havainnoille. Lisäksi asiantuntijatiedon tuominen mukaan parametrien rajaukseen käy varsin luontevasti Bayes-mallissa, jolloin muutamalla puulla tehdyt sovitukset eivät ole alttiita poikkeaville havainnoille samassa määrin kuin pelkästään mittausdataan perustuvissa malleissa.

Sisällys

Johdanto	1
Luku 1. Aineisto	3
1.1. Metsikkö	3
1.1.1. Runkoluku ja lajisto	3
1.2. Puuston pituus ja läpimitta	6
Luku 2. Menetelmät	13
2.1. Näslundin pituusmalli	13
2.2. Bayes-mallinnus	14
2.2.1. Yksinkertainen malli vakiovarianssilla	14
2.2.2. Yksinkertainen malli läpimitasta riippuvalla varianssilla	15
2.2.3. Virheanalyysi ja aikariippuvuus	15
2.2.4. Malleja aikariippuvalla b_1 -parametrilla	16
2.2.5. Aika- ja keskiläpimittariippuvat mallit satunnaisvaikutuksilla	18
Luku 3. Kokeellinen osuus	21
3.1. Teholliset otoskoot ja konvergenssi	21
3.2. Bayes-mallit	23
3.2.1. Linearisointi ja Bayes-malli	23
3.2.2. Läpimitasta riippuva varianssi	24
3.2.3. Aikariippuva b_1 -parametri	24
3.2.4. Sekamallit	26
3.3. Pieni data ja priorit	29
3.4. Sovellus metsien inventointeihin	31
Luku 4. Johtopäätökset	35
Kirjallisuutta	37

Johdanto

Talousmetsän arvon ja tulevien tuottojen simuloimiseksi tarvitaan luotettavaa tietoa puiden kokonaistilavuudesta. Tilavuuden arvioimiseksi tarvitaan tietoa puiden pituus- ja läpimittajakaumista [1, 2]. Puun pituuden määrittäminen on selvästi läpimittaa työläämpää, mistä johtuen läpimitta kirjataan ylös tavallisesti joukosta (tämän työn aineiston tapauksessa kaikista) jotain minimiläpimittaa paksumpia puita (nk. lukupuut), mutta ajan säästämiseksi pituus mitataan vain pienestä osasta lukupuuta (nk. koepuut). Lukupuiden pituudet estimoidaan pituus–läpimitta-relaatiota kuvaavalla pituuskäyrällä, jonka parametrit, joita on tavallisesti kaksi tai kolme, sovitetaan koepuiden avulla. Pituuskäyränä on käytetty useita funktioita, kuten Curtis [3], Schumacher [4], Prodan [5] ja Korf [6]. Mehtätalon *ym.* laajassa vertailussa [1] mikään tutkituista kuudestatoista funktiosta ei osoittautunut ylivoimaisesti parhaaksi. Testatuista malleista kaksiparametrinen Näslundin pituusmalli [12] näytti kuitenkin toimivan kaikissa käytössä olleissa aineistoissa varsin hyvin.

Näslundin pituuskäyrään perustuvia malleja on sovitettu useissa suomalaisissa tutkimuksissa [7, 8, 9, 10]. Näissä tutkimuksissa parametrien oletettiin olevan lineaarisesti riippuvia koealan puiden keskiläpimitasta ja -pituudesta. Oletus parametrien lineaarisesta keskiläpimittariippuvuudesta aiheuttaa kuitenkin harhaa pituusennusteisiin erityisesti nuorissa metsissä, joissa keskiläpimitan logaritmi näyttäisi toimivan paremmin [11]. Logaritminen riippuvuus osoittautui toimivaksi myös nuoressa loblollymäntykohteessa (*Pinus taeda*) Mehtätalon *ym.* tutkimuksessa [1].

Aiemmissä tutkimuksissa malleja on sovitettu epälineaarilla regressiolla tai vaihtoehtoisesti linearisoimalla malli epälineaarilla muunnoksella, minkä jälkeen malli on lokalisoitu eli kalibroitu yhden tai useamman havainnon avulla. Pituuskäyrän määrittämiseen vähäisellä määrällä havaintoja liittyy kuitenkin merkittävä riski siitä, että valitut puut ovat poikkeuksellisen ohuita tai paksuja suhteessa pituuteensa [14]. Linearisointiin perustuvan menetelmän yhtenä hankaluutena on myös se, että epälineaarisen muunnoksen käyttö aiheuttaa harhaa, mutta tätä ongelmaa voidaan lieventää tekemällä takaisinmuunnoksessa harhankorjaus [15].

Tämän tutkielman tarkoituksena on kehittää Bayes-tilastotieteeseen perustuvia malleja puiden pituusmallinnukseen. Kun parametrien ajatellaan olevan satunnaismuuttujia, ei niiden posteriorijakaumien simulointi vaadi mallin linearisoimista. Epälineaariseen regressioon perustuvaan frekventistiseen estimointiin nähden potentiaalisena etuna on se, että asiantuntijatietoa voidaan ottaa luontevasti mukaan priorien muodossa. Tällöin poikkeavat havainnot eivät korjaa pituuskäyrää niin voimakkaasti poikkeavan havainnon suuntaan, ja täten parametrien estimointi pienellä määrällä mittauksia ei ole niin riskialtista.

Aineistona käytetään Parkanossa sijaitsevista neljästä mäntykokeesta tehtyjä mittauksia. Kokeissa puuston harvennusvoimakkuus vaihtelee harventamattomasta voimakkaasti harvennettuun ja mittaukset on tehty metsän ollessa 34–83-vuotias. Mallinnuksen lähtökohtana käytetään Näslundin pituuskäyrää. Kiinnostuksen kohteena on funktion parametrien riippuvuus muuttujista, kuten metsikön iästä ja keskiläpimitasta. Tavoitteena on myös selvittää, voidaanko asiantuntijatietoa käyttäen saavuttaa käytännössä merkittävää etua frekventistisiin sovituksiin verrattuna, kun käytössä on vähäinen määrä mittauksia.

Tutkielma on jaettu neljään kappaleeseen. Kappaleessa 1 esitellään työssä käytettävä aineisto. Kappaleessa 2 esitellään Näslundin pituuskäyrä ja mallin rakennus vaiheittain alkaen yksinkertaisesta mallista ja päättyen hierarkkiseen sekamalliin. Kappaleessa 3 esitellään tulokset, ja kappale 4 sisältää lyhyen yhteenvedon sekä johdotukset.

LUKU 1

Aineisto

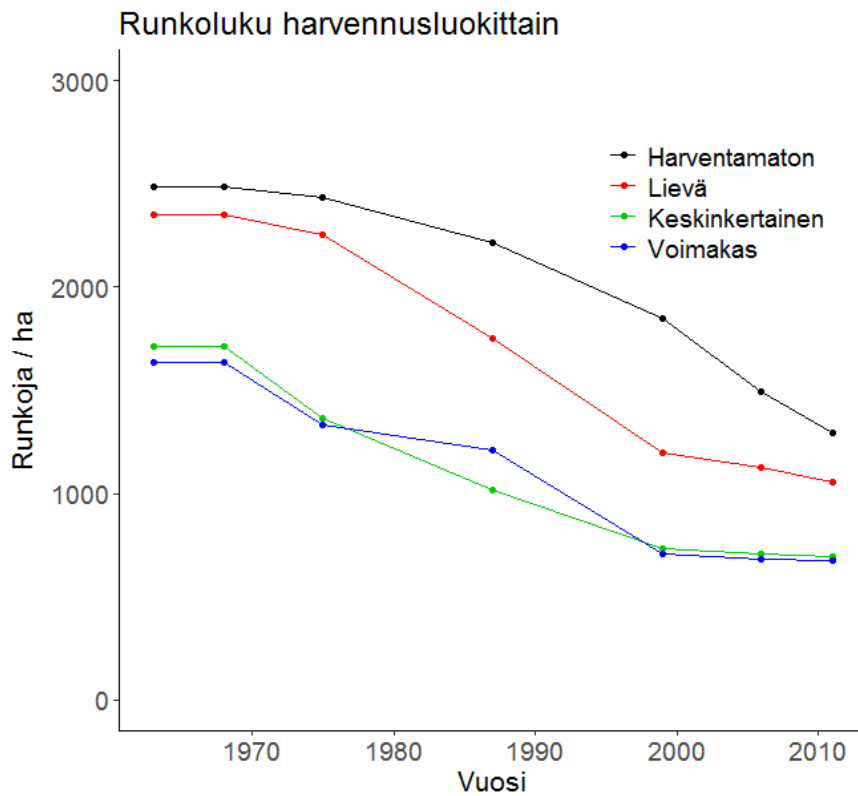
Tässä työssä aineistona käytetään mittauksia Parkanossa sijaitsevista mäntykokeista. Kokeisiin liittyy 17 tuhannen neliömetrin (0.1 ha) suuruista koealaa eli ruutua, jotka sijaitsevat kaikki samalla VT-kankaalla (Vaccinium-tyyppi eli puolukkatyyppi). Ruutujen harvennusvoimakkuus vaihtelee harventamattomasta voimakkaasti harvennettuun. Voimakkaasti harvennettuja ruutuja on viisi, ja muissa harvennusluokissa on kussakin neljä ruutua. Metsikkö on perustettu vuonna 1929 ja kokeet aloitettu 1963. Puuston mittaukset on tehty keväällä 1963 metsikön kasvuttua 34 kasvukautta (ksk), syksyllä 1968 (40 ksk) ja 1975 (47 ksk), keväällä 1987 (58 ksk), sekä syksyllä 1999 (71 ksk), 2006 (78 ksk) ja 2011 (83 ksk). Poikkeuksena on neljäs harventamaton ruutu, joka on mitattu ensimmäisen kerran vasta 1968. Syksyn 1999 mittauksia jatkettiin keväälle 2000, mutta kasvukausien lukumäärä oli mittausten ajan sama. Kaikista ruutujen puista on määritetty mittauksen yhteydessä läpimitta, ja lisäksi puulaaji on kirjattu ylös. Osa puista on valittu koepuiksi, joista on mitattu lisäksi puun korkeus ja latvustotunnuksia.

1.1. Metsikkö

Jokaisesta koealasta on tiedossa runkoluku, keskiläpimitta, valtapituus (100 paksuimman puun keskipituus), puuston kuitu- ja tukkitilavuus sekä harvennusluokka. Tiedot on kirjattu ylös koko puustolle, metsikköön harvennuksen yhteydessä jääville puille, leimatuille puille sekä luonnonpoistuman vuoksi kadonneille puille erikseen jokaisen mittauskerran yhteydessä.

1.1.1. Runkoluku ja lajisto. Runkoluku ajan funktiona eri harvennusluokille on esitetty kuvassa 1.1. Runkoluvun perusteella keskinkertaisesti ja voimakkaasti harvennetut metsät eivät eroa toisistaan merkittävästi. Lievä harvennus vastaa kokeen alkupuolella harventamatonta tapausta, mutta metsän varttuessa runkoluku pienee lievästi harvennetussa metsässä harventamatonta nopeammin. Harventamattoman metsän runkoluvun käytös poikkeaa harvennetuista: runkoluku laskee selvästi hitaammin alussa kiihtyen ajan myötä. Vuonna 2011 runkoluku oli 1300 runkoa/ha, mikä on muita harvennusluokkia selvästi korkeampi. Keskinertainen ja voimakas harvennus ovat johtaneet samankaltaisiin kehityksiin runkojen tiheydessä.

Vaikka kyseessä on neljä mäntykoetta, esiintyy metsiköissä myös pieniä määriä muita puulajeja. Lukupuiden lajijakauma on esitetty taulukossa 1. Kokeen alussa lukupuina toimivia mäntyjä oli 3031 kappaletta, kun taas vuonna 2011 lukupuumäntyjä oli enää 1340. Muiden lajien lukumäärä on pysynyt koko kokeen ajan varsin tasaisena. Poikkeuksena tästä ovat rauduskoivut, joiden lukumäärä on voimakkaasti pienentynyt, mahdollisesti harvennuksen seurauksena. Toiseksi yleisin puulaaji metsiköissä on



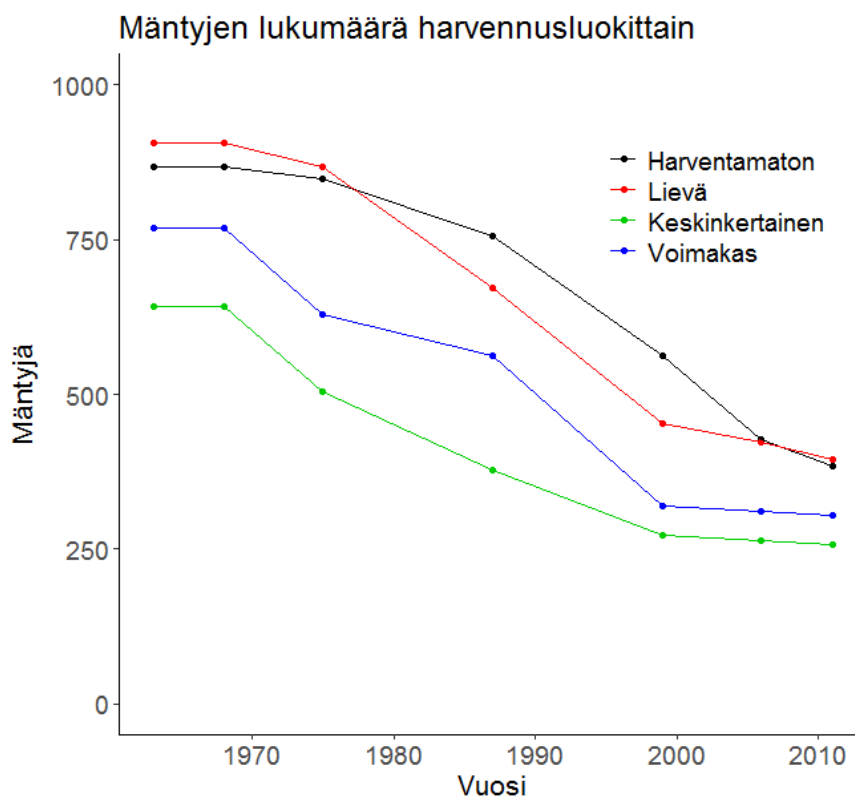
KUVA 1.1. Runkoluvun kehitys metsän varttuessa eri harvennusluokissa.

TAULUKKO 1. Lukupuiden kokonaismäärä Parkanon kokeissa puulajeittain ajan funktiona.

Vuosi	Mänty	Kuusi	Rauduskoivu	Hieskoivu	Haapa	Muu
1963	3031	34	38	177	2	0
1968	3184	35	38	179	2	0
1975	2849	35	25	179	2	0
1987	2367	32	6	192	2	0
1999	1606	40	5	207	2	5
2006	1423	40	5	198	2	5
2011	1340	39	5	166	0	2

TAULUKKO 2. Muiden puiden kuin mäntyjen keskimääräinen lukumäärä ja vaihteluväli harvennusluokittain.

Harvennus	Keskiarvo	Vaihteluväli
Harventamaton	141.6	(124, 178)
Lievä	30.6	(26, 35)
Keskinertainen	30.9	(20, 43)
Voimakas	39.7	(32, 49)

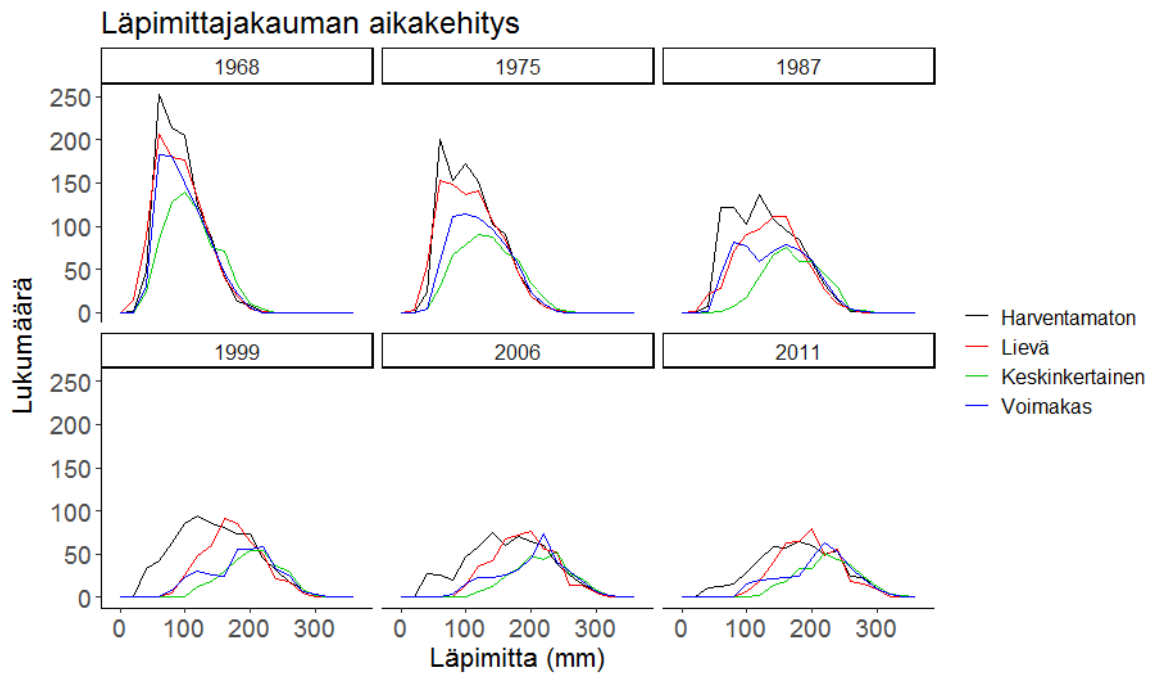


KUVA 1.2. Mäntyjen kokonaismäärä ajan funktiona.

TAULUKKO 3. Koepuiden lukumäärä ja osuus männyistä.

Vuosi	Koepuiden lkm	Osuus
1963	801	26.4 %
1968	865	27.2 %
1975	828	29.1 %
1987	661	27.9 %
1999	531	33.1 %
2006	512	36.0 %
2011	502	37.5 %

hieskoivu, joita on noin 10% puista. Taulukossa 2 puulajisto on esitetty harvennusluokittain. Muut puulajit näyttäisivät esiintyvän lähes yksinomaan harventamattomassa metsässä. Lisäksi lajiston ajallinen tarkastelu paljastaa, että muiden puulajien määrä näyttäisi ajan myötä kasvavan harventamattomassa metsässä, kun taas kaikissa muissa harvennusluokissa trendi on selvästi laskeva. Muiden lajien lukumäärä on siis varsinkin harvennettujen ruutujen kohdalla hyvin pieni. Kuvan 1.2 perusteella mäntyjen lukumäärät harventamattomien ja lievästi harvennettujen ruutujen välillä ovat varsin lähellä toisiaan koko tarkasteluajana. Erot runkoluvuissa johtuvat siis lähinnä harventamattomissa ruuduissa useammin esiintyvistä muista puulajeista.



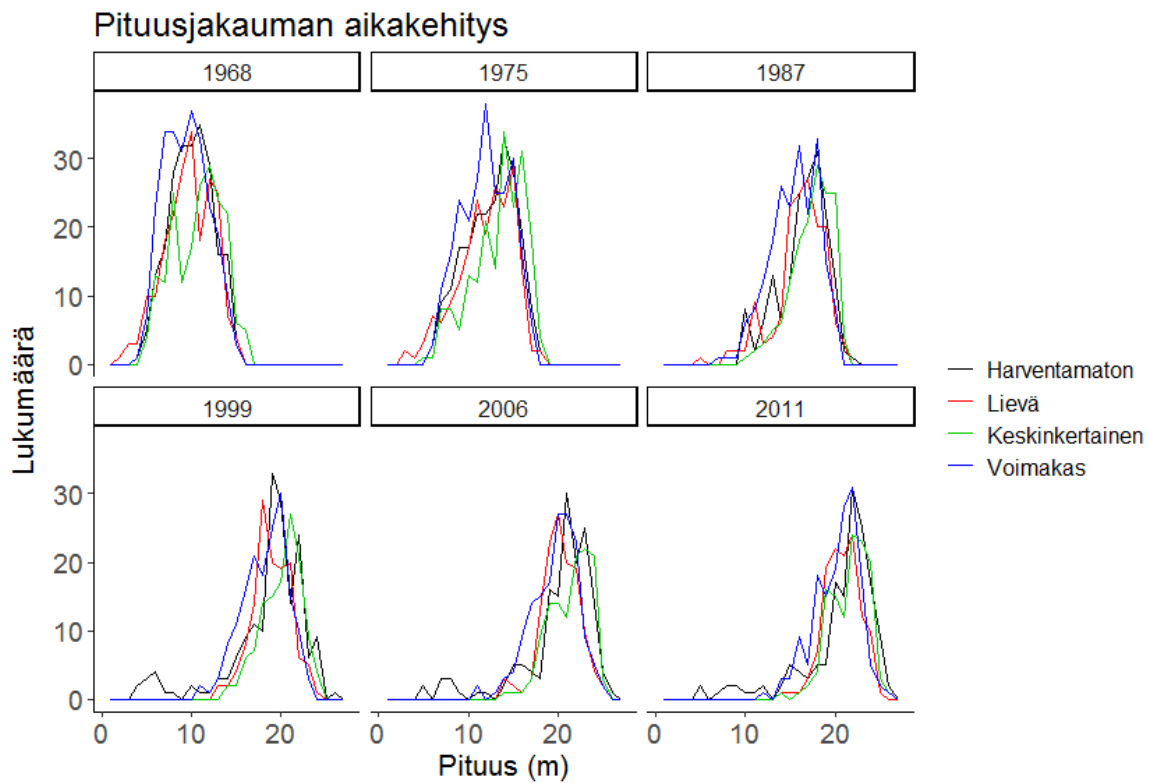
KUVA 1.3. Puiden halkaisijoiden jakauman aikakehitys harvennusluokittain.

Keskinkertaisesti ja voimakkaasti harvennettujen ruutujen runkoluvut ovat hyvin lähellä toisiaan, mutta voimakkaasti harvennettuja ruutuja on yksi enemmän, jolloin runkojen kokonaismäärä voimakkaasti harvennetuissa ruuduissa on suurempi. Erot runkoluvuissa harventamattomien ja lievästi harvennettujen ruutujen välillä johtuvat lähinnä muista puulajeista, joita esiintyy harventamattomissa ruuduissa selvästi enemmän.

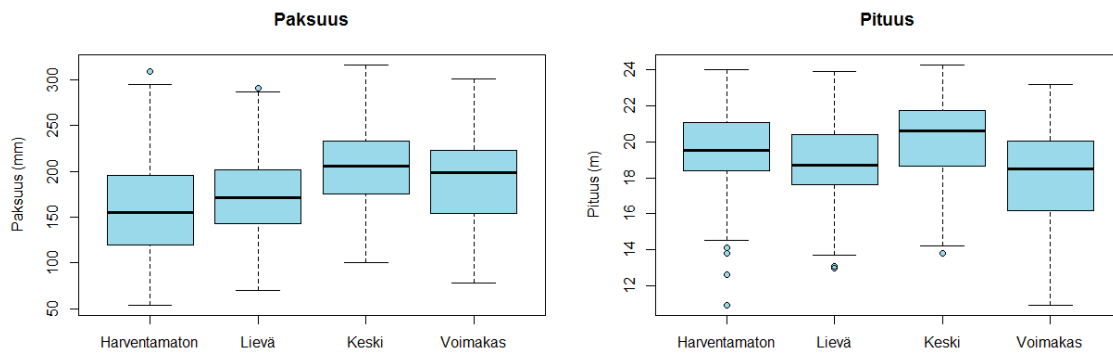
Koepuiden lukumäärät vuosittain on esitetty taulukossa 3. Koepuiden lukumäärä vaihteli tutkimusaikana välillä 502–865 ja sisälsi 26–38% lukupuista. Koepuiden osuus lukupuista kasvoi metsikön vanhetessa. Osa koepuista poistui kokeen edetessä, ja toisaalta osalle koepuista ei ole tehty pituusmittauksia ensimmäisestä mittauserästä alkaen vaan pituus on määritetty vasta jostakin myöhemmästä mittauskerrasta eteenpäin.

1.2. Puuston pituus ja läpimitta

Lukupuiden läpimittajakauman kehitys on esitetty kuvassa 1.3 ja koepuiden pituusjakauman evoluutio kuvassa 1.4. Nuoressa metsässä läpimittajakauma on voimakkaasti oikealle vino, mutta metsän ikääntyessä keskiläpimitta kasvaa ja jakauman muoto lähenee normaalijakaumaa. Keskinertaisen ja voimakkaan harvennusluokan jakaumat lähenevät toisiaan metsän varttuessa, mutta harventamattomien ja lievästi harvennettujen ruutujen läpimittajakaumat pysyvät ohuemmalla puolella. Pituusjakauma puolestaan on melko symmetrinen 40-vuotiaassa metsässä, mutta metsän varttuessa keskipituus kasvaa ja jakauma muuttuu vasemmalle vinoksi. Läpimittajakaumassa harvennuksen vaikutukset jakaumaan kasvavat metsikön varttuessa. Harventamattomassa metsässä rungot jäävät ohuemmiksi, kun taas keskinkertaisesti ja



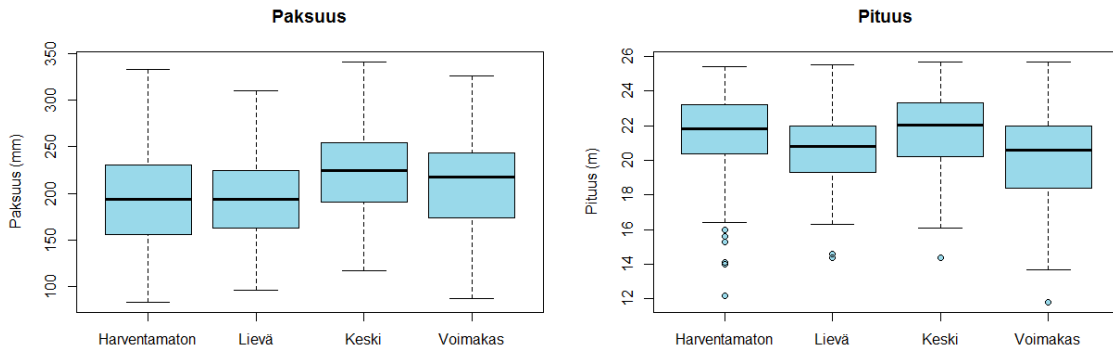
KUVA 1.4. Koepuiden pituusjakauman aikakehitys harvennusluokittain.



KUVA 1.5. Pituus- ja läpimittajakaumat harvennusluokittain vuonna 1999 viiksilaatikoilla esitettyinä.

voimakkaasti harvennetuissa ruuduissa puut ovat paksuimpia. Lievästi harvennetujen ruutujen jakauma sijoittuu näiden jakaumien välimaastoon. Pituusjakaumissa erot näyttäisivät ennemminkin tasoittuvan metsikön varttuessa, sillä pituusjakaumat ovat selvästi lähempänä toisiaan vuonna 2011 kuin vuonna 1968. Selkein poikkeavuus jakaumissa on se, että harventamattomissa ruuduissa esiintyy pieni määrä hyvin lyhyitä, alle 10 metrin pituisia puita.

Hakkuukypsän metsän pituus- ja läpimittajakaumien lisäksi talousmetsän tuottoa ajatellen on olennaista tietää, kuinka kauan puuston nopea kasvuvaihe kestää.

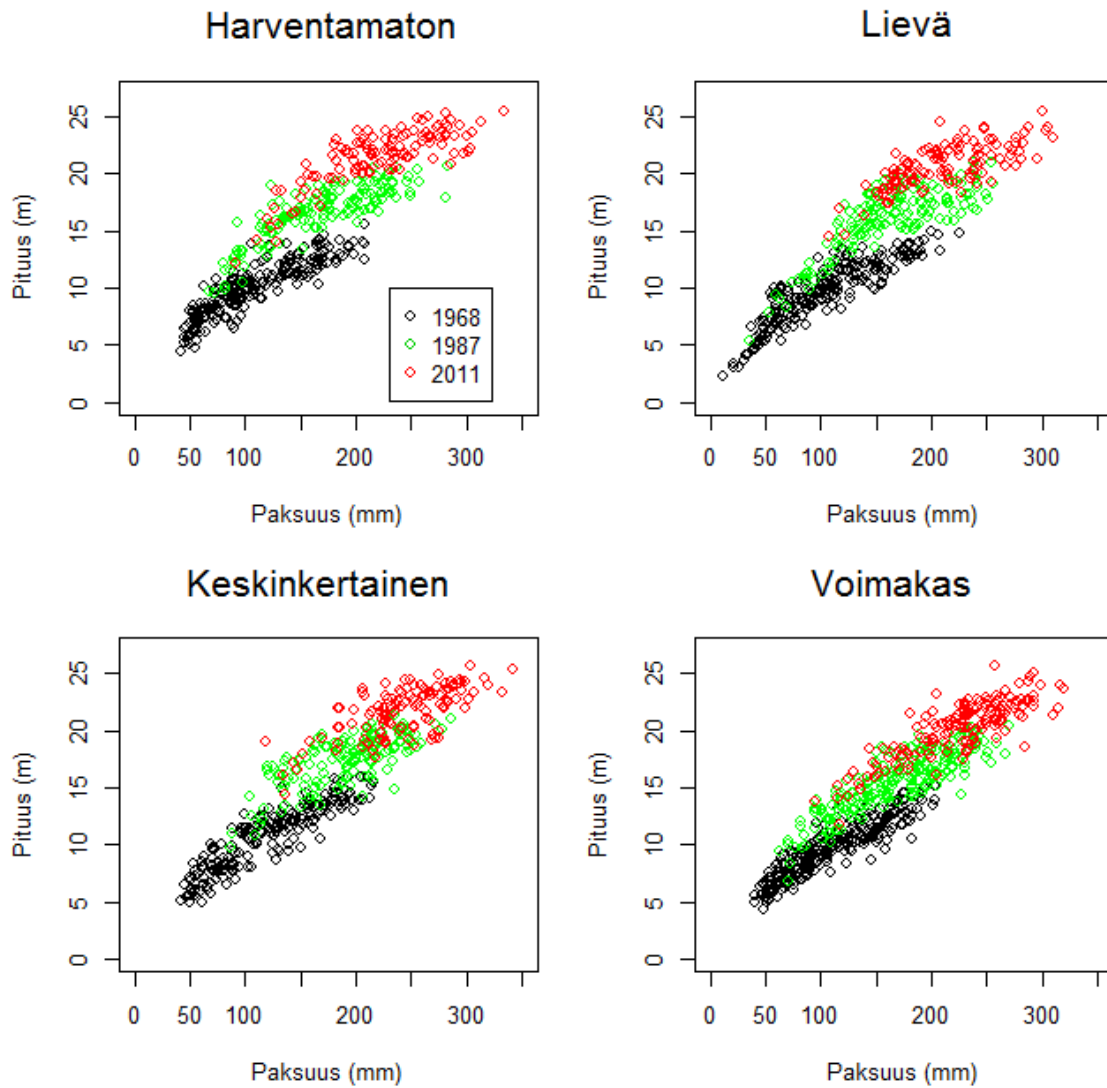


KUVA 1.6. Pituus- ja läpimittajakaumat harvennusluokittain vuonna 2011 viiksilaatikoilla esitettyinä.

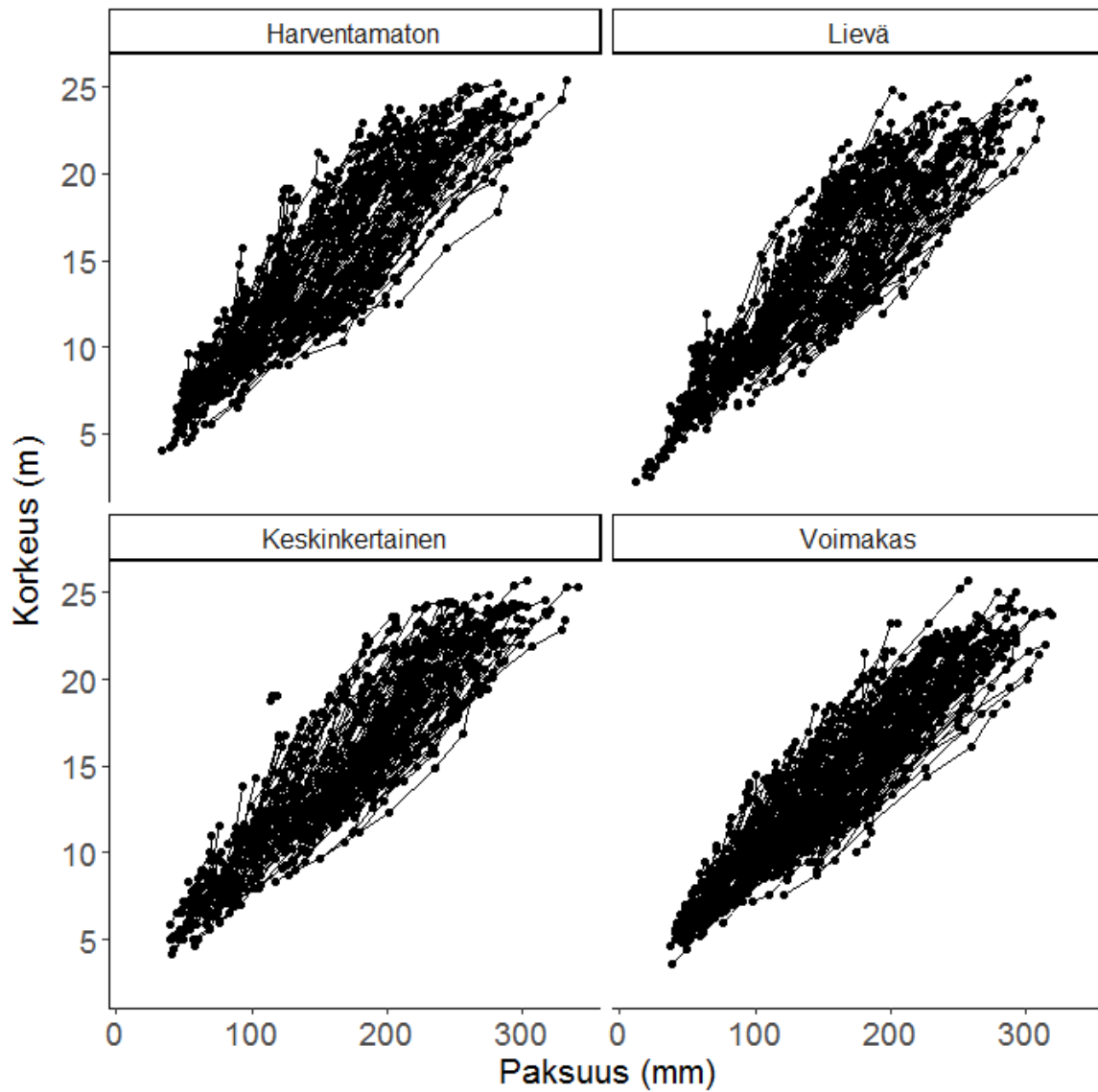
Pituus- ja paksuusjakaumat on esitetty kuvissa 1.5 ja 1.6 vuosina 1999 (71 kasvukautta) ja 2011 (83 kasvukautta). Suurin ero luokkien välillä on harventamattoman ja keskinkertaisesti harvennetun metsän välillä. Harventamattoman metsän mediaanipaksuus oli vain 155.5 millimetriä vuonna 1999, mutta kasvoi seuraavaan mittaushetkeen mennessä voimakkaasti ollen 193.5 millimetriä vuonna 2011. Keskinkertaisella harvennuksella kehitystä mediaanipaksuudessa oli 205.5 millimetristä 225.0 millimetriin samalla ajanjaksolla. Vastaavasti mediaanipuun pituus harventamattomassa metsässä oli vuonna 1999 18,55 metriä ja vuonna 2011 21.80 metriä. Keskinkertaisesti harvennettujen koalojen mediaanipituus puolestaan kasvoi 20.60 metristä 22.05 metriin samana ajanjaksona. Voimakkaasti harvennetussa metsässä läpimittajakauma on lähellä keskinkertaisesti harvennettua metsää, mutta puut ovat hieman lyhyempiä.

Koepuut on esitetty pituus-läpimitta-koordinaateissa vuosina 1968, 1987 ja 2011 kuvassa 1.7. Harvennustoimia ei aloitettu ennen vuotta 1968, joten vuoden 1968 jakaumat vastaavat harventamatonta metsää. Pituus-läpimitta-jakaumasta nähdään, että kehitys on samansuuntaista kaikissa harvennusluokissa. Voimakkaasti harvennetuilla koaloilla jakauma näyttäisi olevan hieman yhtenäisempi, mutta tämä saattaa johtua myös vuoden 1968 jakauman muita luokkia yhtenäisemmästä jakaumasta.

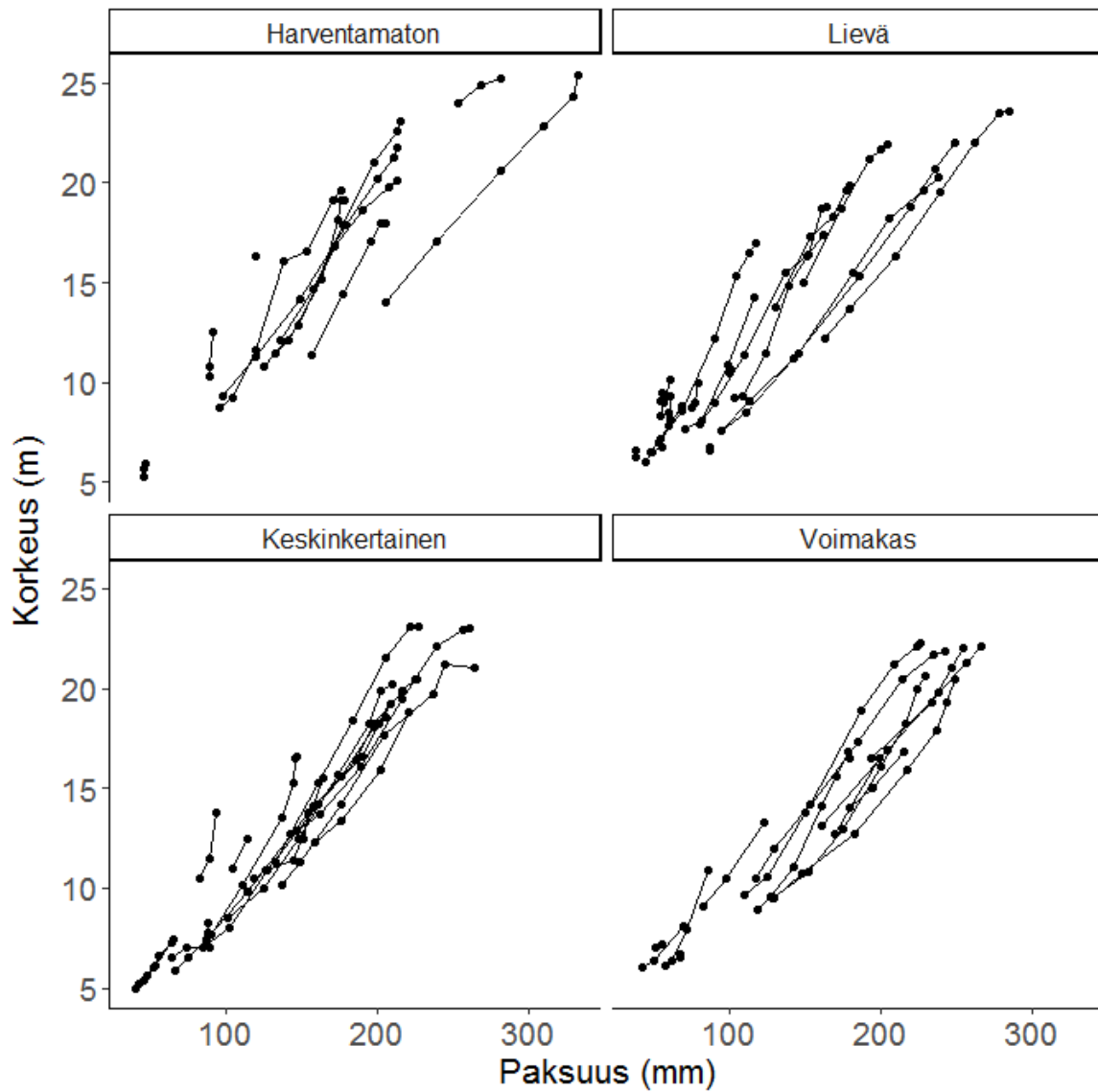
Yksittäisten koepuiden aikakehitys näkyy kuvasta 1.8. Lievästi ja keskinkertaisesti harvennettujen metsikköjen aikakehitys on hyvin samankaltaista. Harventamattomassa metsässä näkyy hieman enemmän hajontaa. Voimakkaasti harvennetussa metsässä pituutta nopeasti kasvaneet, suhteellisen ohuet puut on poistettu nopeammin koelalalta, mikä näkyy jakauman tasaisempina ylälaitana. Eri mänty-yksilöitä on kuvassa 1.8 yhteensä 981, mikä tekee yksittäisten puiden erottamisen hankalaksi. Kuvassa 1.9 puista on valittu yksinkertainen 60 yksilön satunnaisotos, jotta yksittäisten puiden elinkaaret pituus-läpimitta-koordinaatistossa erottuvat paremmin. Yksittäisten puiden pituus-läpimitta-suhde suhteessa muihin metsikön puihin on voimakkaasti korreloitunutta, eli puu on joko ohut tai paksu suhteessa muihin metsikön puihin vuosikymmenestä toiseen. Ohuempia puita poistetaan harvennuksessa enemmän, jolloin näiden puiden käyrät jäävät lyhyemmiksi, mutta pituus-paksuus-suhde näyttäisi näissäkin tapauksissa säilyvän samankaltaisena suhteessa muihin puihin mittauskerrasta toiseen.



KUVA 1.7. Koepuiden pituus ja läpimitta vuosina 1968, 1987 ja 2011 eri harvennusluokille.



KUVA 1.8. Yksittäisten koepuiden pituuden ja läpimitan kehitys vuosina 1963–2011.



KUVA 1.9. Sama kuin kuva 1.8, mutta tähän kuvaan on valittu satunnaisesti 60 puun käyrät. Harvennusluokkaa kohti on siis keskimäärin 15 puuta.

LUKU 2

Menetelmät

2.1. Näslundin pituusmalli

Näslundin pituusmalli antaa relaation puun pituudelle ja läpimitalle. Pituuskäyrä on muotoa [12]

$$(2.1) \quad h = \frac{d^2}{(b_0 + b_1 d)^2} + 1.3 \text{ m},$$

jossa h on puun pituus, b_0 ja b_1 ovat sovitettavat parametrit, ja d on puun läpimita rinnankorkeudelta (1.3 m) mitattuna. Tässä työssä läpimita mitataan metreissä, jolloin parametrien dimensiot ovat $[b_0] = \text{m}^{1/2}$ ja $[b_1] = \text{m}^{-1/2}$. Malli on alun perin kehitetty männyille ja toimiikin sellaisenaan hyvin valopuulajeille, kuten männyille ja koivulle. Varjoa sietävillä puulajeilla, kuten kuusella, pituuden ja läpimitan vaihtelut ovat suurempia, ja niiden tapauksessa parempi malli saadaan korvaamalla eksponentit luvulla kolme [14].

Laskennallisesti helppo lähestymistapa parametrien määrittämiseksi on linearisoida yhtälö (2.1) muunnoksella

$$(2.2) \quad y = \frac{d}{\sqrt{h} - 1.3} = b_0 + b_1 d + \epsilon.$$

Linearisaation yhteydessä tehtävä epälineaarinen muunnos parametrille h aiheuttaa takaisinmuunnoksessa harhaa, jota voidaan korjata harhankorjauksella [15]

$$(2.3) \quad y = \frac{d^2}{(b_0 + b_1 d)^2} + \frac{3d^2}{(b_0 + b_1 d)^4} \sigma_y^2 + 1.3 \text{ m}.$$

Jos jäännösvirhe ϵ yhtälössä (2.2) oletetaan homoskedastiseksi ja normaalijakautuneeksi, saadaan takaisinmuunnoksen yhteydessä jäännösvarianssi, joka kasvaa läpimitan kasvaessa [14]. Varianssin kasvaminen pituuden funktiona on Mehtätalon *ym.* mukaan varsin yleinen havainto pituusmalleja sovitettaessa [1], tosin heidän omista malleissaan tämä ei näyttänyt toteutuvan kaikille tutkimuksessa mukana olleille aineistoille. Jäännösvarianssia voidaan mallintaa varianssifunktiolla

$$(2.4) \quad \text{var}(\epsilon_{ij}) = \sigma^2 \|w_{ij}\|^{2\delta},$$

jossa ruudun i puuhun j liittyvä paino w_{ij} voi olla esimerkiksi pituuden estimaatti \hat{h}_{ij} , läpimita d_{ij} tai katkaistu suhteellinen läpimita

$$(2.5) \quad w_{ij} = \max \left(1, \frac{d_{ij} - m_i}{s_i} + 3 \right),$$

jossa m_i on ruudun läpimitan keskiarvo ja s_i otosvarianssi. Tämän painotuksen todettiin toimivan hyvin useille koealoille Mehtätalon *ym.* artikkelissa [1]. Erityisesti

eksponentin δ arvo oli kolmessa tapauksessa neljästä negatiivinen, eli jäännösvarianssi pieneni läpimitan kasvaessa, jos pienimmille puille oletettiin vakiovarianssi. Artikkelissa oli kuitenkin esitetty parametrille δ ainoastaan estimaatti eikä keskivirhettä, joten etumerkin vaihtelu saattaa olla myös sattumasta johtuvaa.

2.2. Bayes-mallinnus

Pituusmallit on aiemmissa tutkimuksissa sovitettu käyttäen usein frekventististä ajattelutapaa, jossa sovitettujen parametrien ajatellaan olevan estimaatteja todellisista parametreista. Bayes-tilastotieteeseen perustuvassa lähestymistavassa parametreja pidetään satunnaismuuttujina [16]. Bayes-mallinnuksessa linearisaatio ei ole tarpeen ja lisäksi kaikille parametreille saadaan helposti virhearviot posteriorijakaumien kautta. Myös aiempien mittausten ja asiantuntijatiedon ottaminen mukaan onnistuu luontevasti. Bayes-tilastotiede perustuu ehdolliseen jakaumaan, joka voidaan laskea Bayesin kaavalla

$$(2.6) \quad p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta},$$

jossa priorijakauma $p(\theta)$ kuvaa parametreihin liittyvää ennakkotietoa, $p(y|\theta)$ on valittu malli ja $p(\theta|y)$ on posteriorijakauma.

Pituus- ja läpimittajakaumien perusteella kaikki harvennusluokat on syytä käsitellä yksittäin tai ottaa harvennusluokka mukaan muuttujana. Koska harvennuksen vaikutus puun kasvuun ei ole tutkimuksen varsinainen aihe, voidaan aineistoa ajatella neljänä pienempänä aineistona.

Mallien 1 ja 2 avulla voidaan vertailla linearisoinnin ja Bayes-mallinnuksen eroavaisuuksia ja rajata parametreja priorijakaumien avulla siten, että monimutkaisempien Bayes-mallien estimointi onnistuu. Sovittamalla mallit jokaiselle mittauskerralle erikseen saadaan karkea kuva siitä, miten parametrien arvot kehittyvät ajassa.

2.2.1. Yksinkertainen malli vakiovarianssilla. Ensimmäisenä kiinnostuksen kohteena ovat linearisointia käyttäen saatujen parametrien ja Bayes-mallin estimoitujen parametrien erot. Eroavaisuuksien tutkimiseksi linearisoidun mallin parametrit sovitettiin jokaiselle mittauskerralle ja harvennusluokalle erikseen. Vastaavat sovitukset tehtiin myös Bayes-mallilla

$$\begin{aligned} h_{ij}|\theta_{ij} &\sim N(\theta_{ij}, \sigma^2) \\ \theta_{ij} &= \frac{d_{ij}^2}{(b_0 + b_1 d_{ij})^2} + 1.3 \text{ m} \\ b_0, b_1 &\sim \text{logNormal}(0, 10^5) \\ \sigma^2 &\sim \text{Tas}(10^{-6}, 10^6), \end{aligned} \quad (\text{Malli 1})$$

jonka priorijakaumat ovat hyvin epäinformatiivisia. Parametrien b_0 ja b_1 prioreiksi voitaisiin valita myös positiivisen reaaliakselin mittainen tasajakauma, mutta epäai-toja prioreja, jotka eivät ole todennäköisyysjakaumia, ei voi tässä työssä käytetyssä JAGS-ohjelmistossa [18] käyttää.

2.2.2. Yksinkertainen malli läpimitasta riippuvalla varianssilla. Malli 2 on mallin 1 yleistys, jossa pituuden varianssin sallitaan riippuvan läpimitasta. Parametri δ voi olla joko positiivinen tai negatiivinen, joten varianssi voi olla läpimitaltaan suuremmilla puilla kapeampia puita suurempi tai pienempi.

$$\begin{aligned} h_{ij}|\theta_i &\sim N(\theta_{ij}, \sigma^2 d_{ij}^{2\delta}) \\ \theta_{ij} &= \frac{d_{ij}^2}{(b_0 + b_1 d_{ij})^2} + 1.3 \text{ m} \\ b_0, b_1 &\sim \text{logNormal}(0, 10^5) \\ \sigma^2 &\sim \text{Tas}(10^{-6}, 10^6) \\ \delta &\sim N(0, 10^5) \end{aligned} \quad (\text{Malli 2})$$

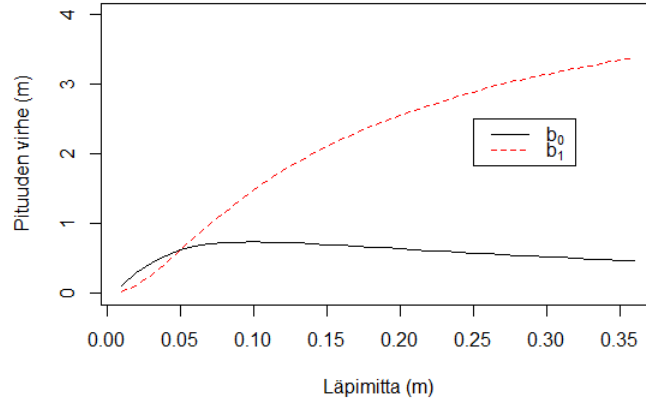
2.2.3. Virheanalyysi ja aikariippuvuus. Näslundin pituuskäyrässä (2.1) on kaksi parametria, mutta parametrien estimoinnin tarkkuuden vaikutus pituuden virheeseen ei ole molemmilla parametreilla yhtä suuri. Tilanteen tarkastelemiseksi tutkitaan Näslundin pituuskäyrän parametreihin liittyviä differentiaaleja (kts. esim. [19]). Parametrin b_0 arvo on harvennusluokasta ja metsikön iästä riippumatta noin 0.01 ja parametrin b_1 arvo noin 0.2 (kts. mallien 1 ja 2 sovitusten tulokset luvusta 3). Tällöin 10 % virhe yksittäisessä parametrissa aiheuttaa pituuden epätarkkuutta

$$(2.7) \quad \begin{aligned} \Delta h_{b_0} &\approx \frac{\partial h}{\partial b_0} \cdot \Delta b_0 \approx \frac{-2d_{ij}^2}{(0.01 + 0.2d_{ij})^3} \cdot 0.001 \\ \Delta h_{b_1} &\approx \frac{\partial h}{\partial b_1} \cdot \Delta b_1 \approx \frac{-2d_{ij}^3}{(0.01 + 0.2d_{ij})^3} \cdot 0.02. \end{aligned}$$

Pituusennusteen virheen suuruutta läpimitan funktiona on havainnollistettu kuvassa 2.1. Parametrin b_0 aiheuttama pituuden virhe on suurimmillaan (n. 0.75 m) läpimitalla 0.10 m, minkä jälkeen virhe lähtee laskuun lähestyen asymptoottisesti nolaa suurilla läpimitan arvoilla. Parametrin b_1 aiheuttama virhe on puolestaan aidosti kasvava lähestyen asymptoottisesti arvoa $2/b_1 \approx 10$ m. Jakamalla yhtälöparin (2.7) ylempi yhtälö alemmalla nähdään, että virheiden suhde käyttäytyy

$$\frac{\Delta h_{b_0}}{\Delta h_{b_1}} \approx \frac{1}{20d_{ij}}.$$

Parametriin b_0 liittyvä virhe dominoi siis alle 5 cm halkaisijoilla ja parametrin b_1 tarkkuus on merkittävämpi tätä suuremmilla läpimitan arvoilla. Mallintamisen kannalta nuoria puita kiinnostavampia ovat varttuneet puut, joten parametrin b_1 estimoinnin tarkkuus on mallin rakennuksessa etusijalla. Lisäksi mallin 1 sovitusten perusteella muutokset parametrissa b_0 ovat koko tarkasteluajankana noin 10 % ja parametrissa b_1 noin 25 % (kts. luku 3). Laskennallisista syistä mallin yksinkertaisuus on merkittävä etu, joten aikariippuvuutta ei kannata ottaa mukaan molempiin parametreihin, mikäli tämä ei paranna mallin ennustuskäytännön selvästi. Näiden tarkastelujen perusteella seuraava askel mallin yleistämisessä on aikariippuvuuden lisäksi muuttujaan b_1 .



KUVA 2.1. Yksittäisten parametrien 10 % suuruisen virheen vaikutus pituusennusteeseen läpimitan funktiona, kun toisen parametrin arvo on kiinnitetty arvoon $b_0=0.01$ tai $b_1=0.2$.

2.2.4. Malleja aikariippuvalla b_1 -parametrilla. Mallien 1 ja 2 sovitusten perusteella parametrilla b_0 ei ole selkeää trendiä, mutta parametrilla b_1 on selvä laskeva trendi ajan funktiona. Harvennusluokasta riippuen trendi noudattelee suurin piirtein lineaarista tai eksponentiaalista muotoa. Mallissa 3a oletetaan lineaarinen trendi, kun taas mallissa 3b oletetaan eksponentiaalinen trendi. Näissä malleissa Ika_i on metsikön i ikä mittaushetkellä, jolloin $(Ika_i - 34)$ on ensimmäisestä mittauskerrasta kulunut aika.

$$h_{ij}|\theta_{ij} \sim N(\theta_{ij}, \sigma^2 d_{ij}^{2\delta})$$

$$\theta_{ij} = \frac{d_{ij}^2}{(b_0 + b_1 d_{ij})^2} + 1.3 \text{ m}$$

$$b_1 = \alpha_1 + \eta \cdot (Ika_i - 34)$$

$$b_0 \sim \text{logNormal}(0, 10^5)$$

$$\alpha_1 \sim \text{logNormal}(0, 10^5)$$

$$\eta \sim N(0, 10^2)$$

$$\sigma^2 \sim \text{Tas}(10^{-6}, 10^6)$$

$$\delta \sim N(0, 10^5)$$

(Malli 3a)

$$\begin{aligned}
h_{ij}|\theta_i &\sim N(\theta_{ij}, \sigma^2 d_{ij}^{2\delta}) \\
\theta_{ij} &= \frac{d_{ij}^2}{(b_0 + b_1 d_{ij})^2} + 1.3 \text{ m} \\
b_1 &\sim \exp(\alpha_1 + \eta \cdot [\text{Ika}_i - 34]) \\
b_0 &\sim \text{logNormal}(0, 10^5) \\
\alpha_1 &\sim N(0, 10^5) \\
\eta &\sim \text{Tas}(-0.1, 0) \\
\sigma^2 &\sim \text{Tas}(10^{-6}, 10^6) \\
\delta &\sim \text{Tas}(-5, 5)
\end{aligned} \tag{Malli 3b}$$

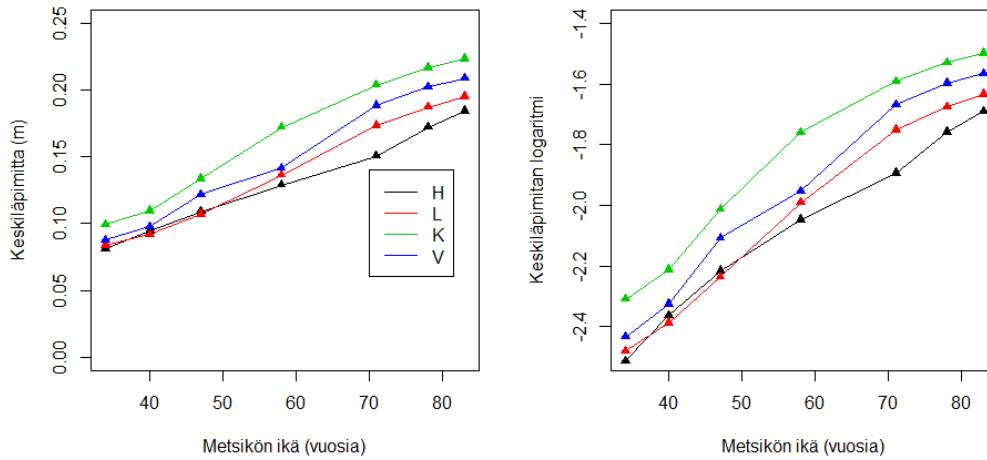
Mikäli eksponentiaalinen malli saadaan estimoitua hyvin, niin etuna on malliin 3a verrattuna se, että b_1 pysyy aina aidosti positiivisena. Linearisessa mallissa parametrin b_1 arvo laskee lähelle nollaa varttuneessa metsässä, minkä jälkeen se muuttuu negatiiviseksi, vaikka Näslundin pituuskäyrässä parametrien arvot oletetaan positiivisiksi. Lisäksi rajatapaus $b_1 \rightarrow 0$ on ongelmallinen mallissa, jossa b_0 on vakio (tai ei mene nolnaan yhtä nopeasti kuin b_1), sillä tällöin esimerkiksi läpimitaltaan 20 cm puulle pituusennuste lähestyy arvoa

$$\hat{h}_{ij} \rightarrow \frac{d_{ij}^2}{b_0^2} + 1.3 \text{ m} \approx \frac{d_{ij}^2}{(0.01 \text{ m}^{1/2})^2} + 1.3 \text{ m} \approx 10^4 \text{ m}^{-1} \cdot (0.2 \text{ m})^2 + 1.3 \text{ m} \approx 400 \text{ m}.$$

Toisaalta asymptoottisia tarkasteluja voidaan käyttää parametrien priorijakaumien rajaamiseen siten, että epärealistisilta parametrien arvoilta vältytään, mikä on eräs Bayes-mallinnuksen eduista. Esimerkiksi mallissa 3b eksponentti $\eta \leq 0$, jolloin trendi on joko vakio tai laskee ajan myötä. Lisäksi η :n priorijakauma on rajattu siten, että η ei saa pienempiä arvoja kuin -0.1 , jolloin parametrin b_1 puoliintumisaika on vähintään $10 \ln(2) \approx 7$ vuotta. Sopivan priorin voi päätellä asymptoottisen käyttäytymisen perusteella, jos parametrien b_0 ja b_1 suuruudet ajanhetkellä $t = 0$ ovat tiedossa. Olkoon t_{asy} aika, jota ennen asymptoottiseen käytökseen ei haluta toteutuvan, ja d_{max} on suurin läpimitan arvo. Rajoite perustuu siihen, että b_0 :n voidaan katsoa dominoivan, kun $b_1 d_{ij} < 0.1 b_0$, jolloin lähestytään epärealistista käytöstä $d^2/b_0^2 + 1.3 \text{ m}$. Tällöin η :n on oltava vähintään

$$\begin{aligned}
b_1 \cdot d_{\text{max}} \exp(t_{\text{asy}} \eta) &= 0.1 b_0 \\
\exp(t_{\text{asy}} \eta) &= \frac{0.1 b_0}{b_1 d_{\text{max}}} \\
\eta &= \frac{1}{t_{\text{asy}}} \ln \left(\frac{0.1 b_0}{b_1 d_{\text{max}}} \right).
\end{aligned}$$

Parkanon aineistossa tällä päättelyllä saadaan η :lle alaraja -0.085 , jos valitaan $t_{\text{asy}} = 50$. Parametrin η priorin lisäksi parametrin δ prioria on rajattu mallissa 3b siten, että arvot pysyvät välillä $(-5, 5)$. Artikkelin [1] sovitusten perusteella parametrin δ

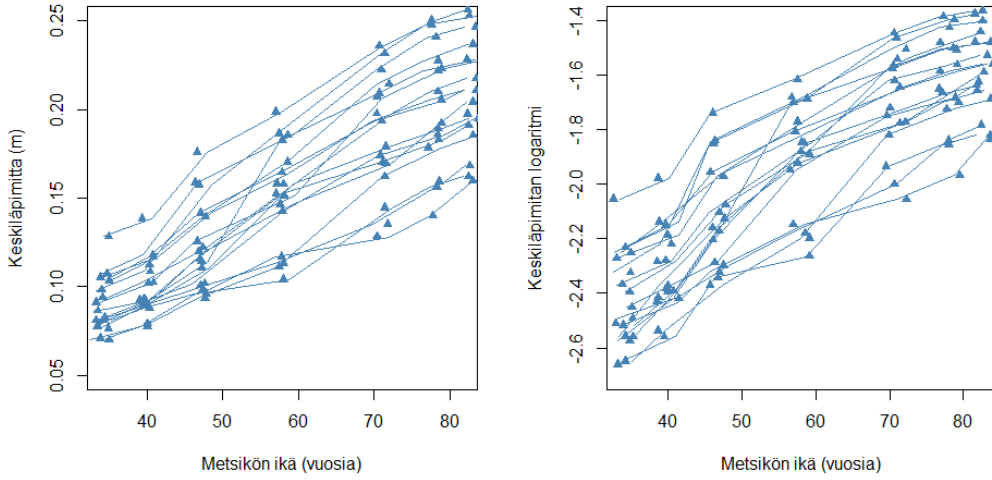


KUVA 2.2. Puiden keskiläpimitan ja sen logaritmin riippuvuus iästä harvennusluokittain. Harvennusluokat ovat (H)arventamaton, (L)ievä, (K)eskinkertainen ja (V)oimakas.

arvo on todennäköisesti välillä $(-2, 2)$, mutta väli on valittu konservatiivisesti hieman leveämmäksi.

2.2.5. Aika- ja keskiläpimitariippuvat mallit satunnaisvaikutuksilla. Luvun 1.2 tarkastelujen perusteella lopullisessa mallissa tulisi olla mukana puutason korrelaattorakenne. Aiempien tutkimusten perusteella satunnaisvaikutuksen pitämiseksi läpimitasta riippumattomana tulisi mallissa olla mukana koealan keskiläpimitasta riippuva termi [1, 20]. Esimerkiksi Mehtätalon *ym.* artikkelissa [1] metsämännuille (*Pinus sylvestris* kuten Parkanon aineistossa) sovitetussa logistisessa mallissa tämä oli huomioitu parametreissa kohteen keskiläpimitasta lineaarisesti riippuvalla termillä. Nuoreen loblollymäntykohteeseen (*Pinus taeda*) sovitetussa Näslundin mallissa puolestaan käytettiin koealan keskiläpimitan logaritmia. Läpimitan ja metsikön iän välillä on harvennusluokittain varsin selkeä lineaarinen riippuvuus, kuten kuvasta 2.2 voi havaita. Nuoressa metsässä läpimitan logaritmi ja ikä ovat myös miltei täydellisesti lineaarisesti riippuvia, joten yhtenä vaihtoehtona on käyttää keskiläpimitan sijaan metsikön ikää. Toisaalta, jos asiaa tarkastellaan koealatasolla, niin ruutujen välillä on selvää hajontaa, kuten kuvasta 2.3 voi havaita. Metsikön ikä on looginen valinta, jos mallia käytetään tulevien havaintojen ennustamiseen. Lisäksi läpimitatiedon saaminen vaatii mittauksia, kun taas tasaikäisen talousmetsän ikä tunnetaan tarkasti nyt ja tulevaisuudessa.

Mallissa 4a on yleistys mallista 3b, jossa on mukana satunnaisvaikutus parametrille b_1 . Aikariippuvuus on mallin 3b tapaan mukana ainoastaan parametrissa b_1 . Priorijakaumia on pyritty rajaamaan mallien 1–3 sovitusten perusteella (kts. kappale 2). Mallissa 4a parametrin δ arvo on rajattu artikkelin [1] sovitusten perusteella välille $(-2, 2)$ ja mallissa 4b rajaus on mallin 4a sovituksen (kts. luku 3) perusteella välille $(-1, 1)$.



KUVA 2.3. Puiden keskiläpimitan ja sen logaritmin riippuvuus iästä harvennusluokittain, mutta jaottelu on tehty harvennusluokkien sijaan ruuduittain.

$$\begin{aligned}
 h_{ij} | \theta_i &\sim N(\theta_{ij}, \sigma^2 d_{ij}^{2\delta}) \\
 \theta_{ij} &= \frac{d_{ij}^2}{(b_0 + b_1 d_{ij})^2} + 1.3 \text{ m} \\
 b_{1ij} &\sim \exp(\alpha_1 + \eta \cdot (\text{Ika}_i - 34) + u_{ij}) \\
 b_{0ij} &\sim \text{logNormal}(0, 10^3) \\
 \alpha_1 &\sim N(0, 10^3) \\
 \eta &\sim \text{Tas}(-0.1, 0) \\
 u_{ij} &\sim N(0, \sigma_1^2) \\
 \sigma_1^2 &\sim \text{Tas}(10^{-6}, 10) \\
 \sigma^2 &\sim \text{Tas}(10^{-6}, 10) \\
 \delta &\sim \text{Tas}(-2, 2)
 \end{aligned}
 \tag{Malli 4a}$$

Malli 4b sallii parametrien b_0 ja b_1 riippua ruudun keskiläpimitasta (ja siten ajasta). Lisäksi molempiin parametreihin sisältyvät puutason satunnaisvaikutukset. Lisäksi parametrien b_{0ij} ja b_{1ij} jakaumien leveyttä on pienennetty mallin 4a sovitusten perusteella (kts. luku 3), sillä varianssi 10^2 tekee jakaumista jo varsin epäinformatiivisia.

$$\begin{aligned}
h_{ij}|\theta_{ij} &\sim N(\theta_i, \sigma^2 d_{ij}^{2\delta}) \\
\theta_{ij} &= \frac{d_{ij}^2}{(b_{0ij} + b_{1ij}d_{ij})^2} + 1.3 \text{ m} \\
b_{1ij} &\sim \exp(\alpha_1 + \eta_1 \bar{d}_i + u_{ij}) \\
b_{0ij} &\sim \exp(\alpha_0 + \eta_0 \bar{d}_i + v_{ij}) \\
\alpha_0, \alpha_1 &\sim N(0, 10^2) \\
\eta_1 &\sim \text{Tas}(-10, 0) \\
u_{ij} &\sim N(0, \sigma_1^2) \\
\sigma_1^2 &\sim \text{Tas}(10^{-3}, 10) \\
\eta_0 &\sim \text{Tas}(-100, 0) \\
v_{ij} &\sim N(0, \sigma_0^2) \\
\sigma_0^2 &\sim \text{Tas}(10^{-3}, 10) \\
\sigma^2 &\sim \text{Tas}(10^{-3}, 10) \\
\delta &\sim \text{Tas}(-1, 1)
\end{aligned}
\tag{Malli 4b}$$

LUKU 3

Kokeellinen osuus

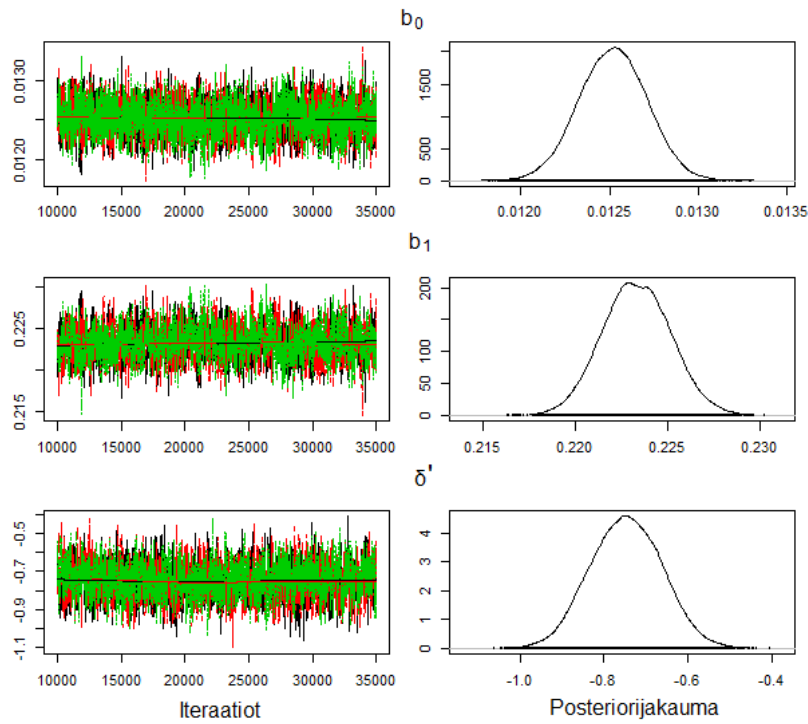
3.1. Teholliset otoskoot ja konvergenssi

Mallit sovitettiin R-ohjelmistolla [17] käyttäen JAGS:ia (*Just Another Gibbs Sampler*) [18] kolmella Markovin ketjulla (Gibbsin otanta on esitelty artikkelissa [21]). Mallille 1 iteraatiokierroksia tehtiin 25 000, joista 15 000 hylättiin lämmittelyjaksona. Mallille 2 tehtiin 50 000 iteraatiota, joista ensimmäiset 40 000 poistettiin lämmittelyjaksona. Malleille 3a ja 3b iteraatiokierroksia tehtiin 35 000 lämmittelyjakson ollessa 15 000. Mallille 4a vastaavat luvut olivat 45 000 ja 20 000 ja mallille 4b 50 000 ja 28 000. Iteraatiokierrosten määrän valinnassa pyrittiin pääsemään mahdollisimman suuriin tehollisiin otoskokoihin järkevällä laskennallisella kuormituksella. Toteutuneet teholliset otoskoot laskettiin rjags-kirjaston *effectivesize*-funktiolla ja tulokset on koottu taulukkoon 1. Erityisesti liian lyhyet lämmittelyjaksot osoittautuivat malleissa erityisen ongelmallisiksi johtuen epäinformatiivisista priorijakaumista, joiden vuoksi parametrien alkuarvot ketjuissa saattoivat olla kaukana tasapainojakaumasta. Ketjujen konvergenssi tarkastettiin graafisesti. Kaikkien mallien ketjujen konvergenssia hyvin edustavana esimerkkinä on esitetty voimakkaasti harvennettujen koealojen dataan sovitettun mallin 3a parametrien iteraatioketjuja kuvassa 3.1.

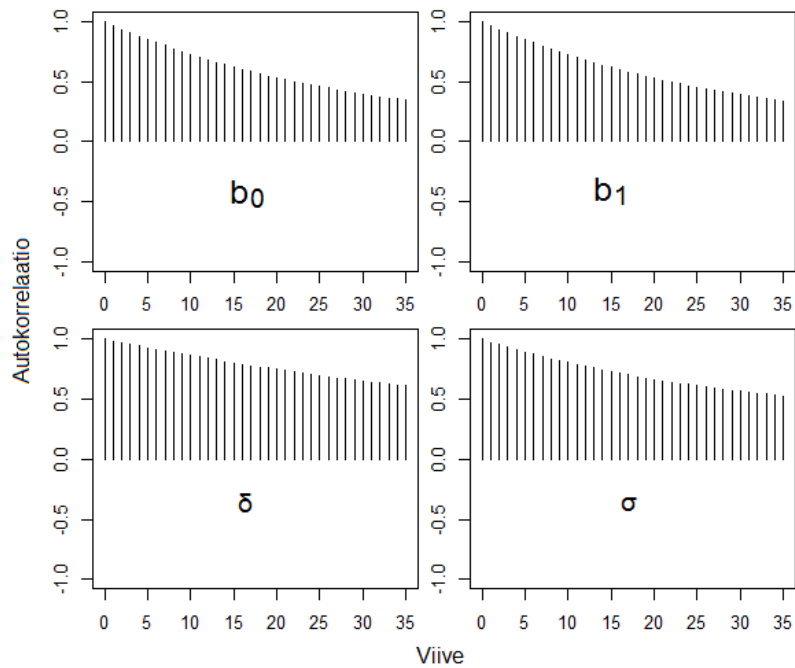
Parametrien tehollisiin otoskokoihin vaikuttavat Markovin ketjujen autokorrelaatiot. Kuvassa 3.2 on esitetty esimerkkinä yhden mallin 2 sovitukseen liittyvän Markovin ketjun autokorrelaatiot mallin eri parametreille. Erityisesti varianssifunktion parametreilla σ ja δ autokorrelaatio on varsin voimakasta.

TAULUKKO 1. Sovitettujen mallien parametrien teholliset otoskoot keskinkertaisesti harvennettujen koealojen dataan sovitettuina. Mallit 1 ja 2 on sovitettu vuoden 2011 dataa käyttäen. Puutason satunnaisfektien u ja v teholliset otoskoot olivat kaikille puille vähintään 2 500 mallissa 4a ja mallissa 4b vähintään 2 000. Malleille 4a ja 4b parametreja b_0 ja b_1 vastaa $\exp(\alpha_0)$ ja $\exp(\alpha_1)$.

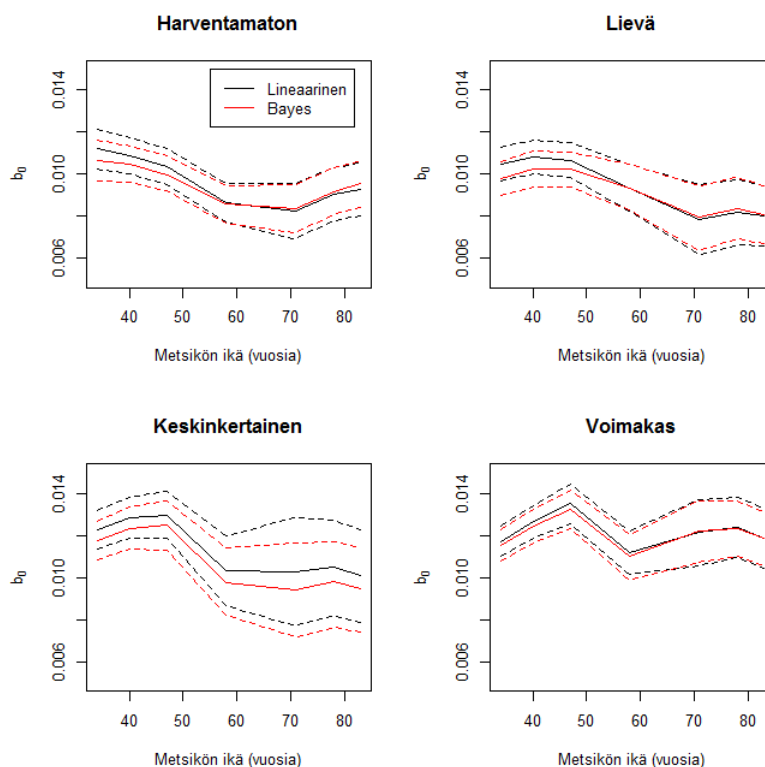
Malli	b_0	b_1	σ	δ	η_0	η/η_1	σ_0	σ_1
1	488	493	38 117					
2	1 059	1 072	522	452				
3a	1 227	991	1 455	1 451		1 375		
3b	1 340	1 165	1 527	1 513		2 094		
4a	190	117	723	403		403		387
4b	182	60	698	666	101	138	>40 000	1 747



KUVA 3.1. Mallin 3a parametrien b_0 , β_1 ja $\delta' = -2\delta$ Markovin ketjut voimakkaaseen harvennusluokkaan sovitettuina.



KUVA 3.2. Mallin 2 parametrien autokorrelaatiot voimakkaan harvennusluokan vuoden 2011 dataan sovitettuina.



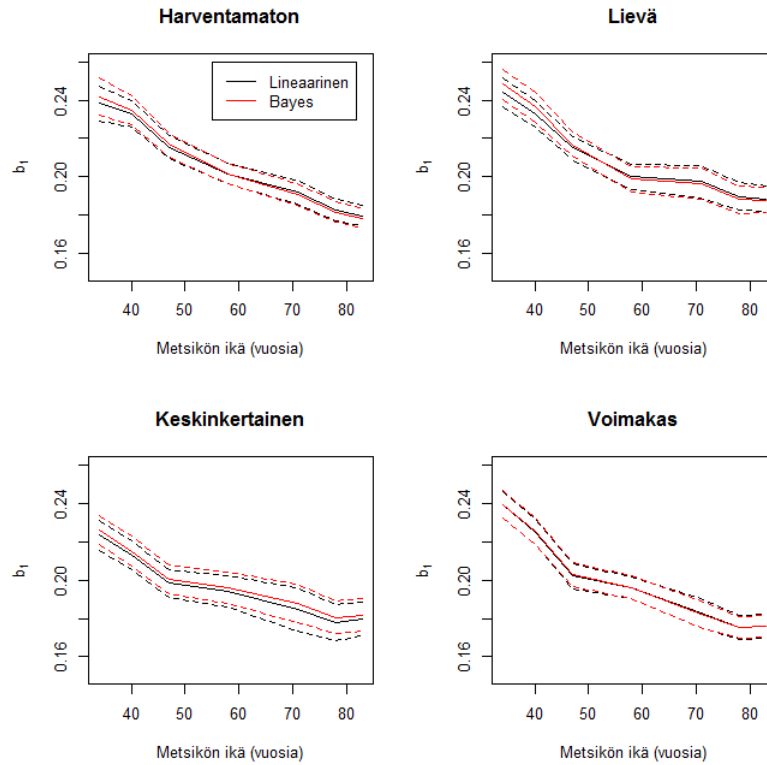
KUVA 3.3. Näslundin käyrän parametrin b_0 aikakehitys sekä linearisoidun yhtälön lineaarisella regressiolla että epäinformatiivisilla prioreilla sovitettuun Bayes-malliin (malli 1) perustuen.

3.2. Bayes-mallit

Tässä kappaleessa esitellään kappaleen 2 mallien sovitusten tulokset. Mallien 1 ja 2 tuloksia on käytetty valittaessa malleja 3a ja 3b, joiden tuloksia on puolestaan käytetty mallien 4a ja 4b luomiseen.

3.2.1. Linearisointi ja Bayes-malli. Kuvissa 3.3 ja 3.4 on esitetty Näslundin pituuskäyrän (yhtälö 2.1) parametrien aikakehitys linearisoidulla mallilla ja Bayes-mallilla (malli 1) määritettynä. Malleissa on oletettu vakiovarianssi. Bayes-malliin verrattuna linearisoitu malli estimoii parametrin b_0 arvon suuremmaksi ja vastaavasti parametrin b_1 arvon pienemmäksi nuoressa metsässä. Varttuneelle metsälle molemmat mallit antavat samanlaisia tuloksia lukuun ottamatta keskinertaisesti harvennettua metsää, jossa parametrin b_0 arvo on koko tarkasteluajan suurempi linearisoidussa mallissa ja parametrin b_1 arvo vastaavasti pienempi.

Harventamattomassa, lievästi harvennetussa sekä keskinertaisesti harvennetussa metsikössä parametrilla b_0 näyttäisi olevan mahdollisesti hitaasti laskeva trendi. Voimakkaasti harvennetussa metsikössä selvää trendiä ei näy. Posteriorikeskiarvon/estimaatin vaihtelu tarkasteluajana on noin 10 %. Vaihtelu on suurin piirtein samaa luokkaa kuin 95 %:n posterioriväli (tässä työssä *Equally Tailed Interval* (ETI) [16]), joten sopivia vaihtoehtoja aikakehitykseksi voisivat olla aikariippumaton vakioarvo (mallit 3a, 3b ja 4a) tai hitaasti laskeva eksponentiaalinen trendi (malli 4b).

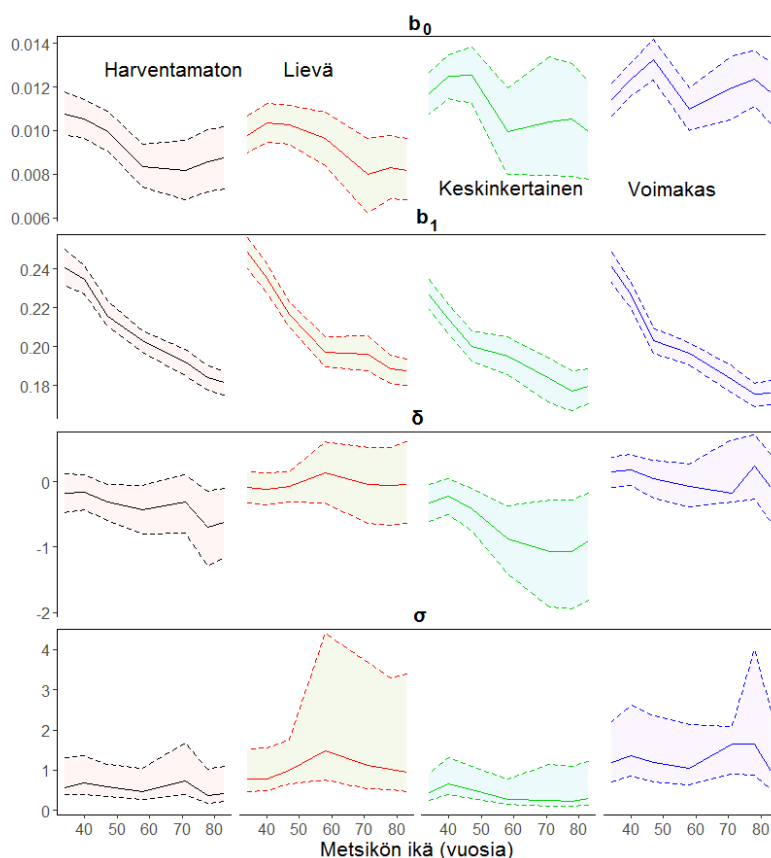


KUVA 3.4. Näslundin käyrän b_1 parametrin aikakehitys sekä frekventistiseen linearisointiin että epäinformatiivisilla prioreilla sovitettuun Bayes-malliin (malli 1) perustuen.

Parametrilla b_1 on jokaisessa harvennusluokassa selvästi laskeva trendi. Harventamattomalle metsälle trendi näyttäisi olevan melko lähellä lineaarista (malli 3a), mutta muille luokille käytös vaikuttaisi olevan ennemminkin eksponentiaalisesti laskevaa (mallit 3b, 4a ja 4b).

3.2.2. Lämpimitasta riippuva varianssi. Malli 2 on mallin 1 yleistys, jossa varianssin sallitaan riippuvan lämpimitasta. Parametrien posteriorikeskiarvot ja 95 %:n posteriorivälit sovitettuna eri mittauskerroille ja harvennusluokille on esitetty kuvassa 3.5. Parametrien b_0 ja b_1 arvot ja aikakehitys eivät poikkea selvästi mallin 1 arvoista. Eksponentti δ :lle ei näytä löytyvän tarkkaa arvoa, mutta posteriorikeskiarvot ovat joko lähellä nollaa tai negatiivisia.

3.2.3. Aikariippuva b_1 -parametri. Mallit 3a ja 3b sovitettiin harvennusluokittain kaikkia mittauspisteitä käyttäen. Näissä malleissa ei kuitenkaan ollut mukana puutason satunnaisvaikutuksia. Parametrien posteriorikeskiarvot ja 95 %:n posteriorivälit on koottu taulukoihin 2 ja 3. Parametrien b_0 posteriorikeskiarvot ovat systemaattisesti joitain prosentteja suurempia mallille 3a kuin mallille 3b, ja vastaavasti parametrien b_1^0 (eli b_1 ensimmäisellä mittauskerralla vuonna 1963) posteriorikeskiarvot ovat mallissa 3a muutaman prosentin pienempiä. Erot harvennusluokkien välillä ovat suhteellisesti suurempia parametrille b_0 kuin b_1^0 .



KUVA 3.5. Mallin 2 parametrien posteriorikeskiarvot ja 95 %:n posteriorivälit harvennusluokittain.

TAULUKKO 2. Mallien 3a ja 3b parametrien b_0 ja b_1^0 posteriorikeskiarvot ja 95 %:n posteriorivälit. Merkintä b_1^0 viittaa parametrin b_1 arvoon ensimmäisellä mittauskerralla vuonna 1963.

Malli	Harv.	$b_0(10^{-2})$	$b_0(95\%)$	b_1^0	$b_1^0(95\%)$
3a	H	1.04	(0.99, 1.08)	0.224	(0.220, 0.228)
	L	1.02	(0.99, 1.06)	0.235	(0.231, 0.239)
	K	1.22	(1.16, 1.27)	0.216	(0.211, 0.220)
	V	1.25	(1.21, 1.29)	0.224	(0.220, 0.227)
3b	H	1.01	(0.98, 1.05)	0.227	(0.224, 0.231)
	L	1.00	(0.97, 1.05)	0.239	(0.234, 0.243)
	K	1.20	(1.15, 1.26)	0.219	(0.214, 0.223)
	V	1.24	(1.20, 1.28)	0.227	(0.223, 0.231)

Molemmissa malleissa parametrin η 95%:n posterioriväli on kokonaisuudessaan negatiivinen kaikissa harvennusluokissa, joten on selvää näyttöä siitä, että parametrin b_1 arvo pienenee metsän varttuessa malleissa 3a ja 3b. Linearisessa mallissa laskua on noin 10^{-3} vuosittain, joten epärealistisen pieniin parametrin arvoihin päätyminen tapahtuisi vasta noin 200 vuoden iässä. Eksponentiaalisessa mallissa parametrin η

TAULUKKO 3. Mallien 3a ja 3b parametrien η , δ ja σ posteriorikeskiarvot ja 95 %:n posteriorivälit.

Malli	Harv.	$\eta(10^{-3})$	$\eta(95\%)$	δ	$\delta(95\%)$	σ	$\sigma(95\%)$
3a	H	-1.26	(-1.32, -1.20)	0.14	(0.04, 0.25)	1.78	(1.48, 2.24)
	L	-1.26	(-1.33, -1.20)	0.30	(0.21, 0.38)	2.53	(2.12, 3.07)
	K	-1.01	(-1.07, -0.95)	0.21	(0.12, 0.31)	2.18	(1.83, 2.63)
	V	-1.11	(-1.16, -1.05)	0.37	(0.28, 0.45)	2.56	(2.16, 3.03)
3b	H	-6.6	(-6.9, -6.4)	0.12	(0.02, 0.23)	1.71	(1.39, 2.12)
	L	-6.4	(-6.7, -6.2)	0.27	(0.18, 0.36)	2.36	(1.95, 2.85)
	K	-5.5	(-5.8, -5.5)	0.20	(0.10, 0.30)	2.10	(1.77, 2.54)
	V	-5.9	(-6.1, -5.6)	0.37	(0.28, 0.45)	2.51	(2.11, 3.00)

arvo on kaikissa harvennusluokissa noin $6 \cdot 10^{-3}$, joten parametrin arvo laskee vuosittain noin 0.6 %. Parametrin puoliintumiseen tässä mallissa menee noin 115 vuotta. Parametrin δ arvo malleissa on positiivinen posteriorikeskiarvojen vaihdellessa välillä 0.12–0.37 harvennusluokasta ja mallista riippuen. Tämä on linjassa useiden aiempien havaintojen kanssa [14]. Toisaalta mittauskerroille erikseen sovitetussa mallissa 2 posteriorikeskiarvot olivat todennäköisemmin negatiivisia. Varianssi näyttäisi kasvavan puuston varttuessa, jolloin suurempi läpimitta ennustaisi suurempaa varianssia, mutta yksittäisellä ajanhetkellä koealojen paksuimpien puiden varianssi on kuitenkin mahdollisesti ohuempia pienempi tai merkittävää riippuvuutta ei ole. Tämä saattaa osaltaan selittää, miksi artikkelissa [1] osalle aineistoista saatiin positiivisia ja osalle negatiivisia δ :n arvoja. Mallin 3b jäännösvariانسsit ovat systemaattisesti hieman mallia 3a pienempiä (vrt. parametrien σ ja δ arvot taulukossa 3). Eksponentiaalinen aikakehitys saattaisi siis olla parempi valinta monimutkaisempiin malleihin. Jäännösvariانسsin suuruus riippuu valitusta harvennuksesta ja puun läpimitasta, mutta esimerkiksi läpimitaltaan 25 senttimetrin paksuiselle puulle keskinkertaisesti harvennetussa metsässä pituuden keskihajonta on noin

$$\sqrt{2.18^2 \cdot (0.25 \text{ m})^{2 \cdot 0.21}} \approx 1.63 \text{ m} \quad (\text{Malli 3a})$$

$$\sqrt{2.10^2 \cdot (0.25 \text{ m})^{2 \cdot 0.20}} \approx 1.59 \text{ m}. \quad (\text{Malli 3b})$$

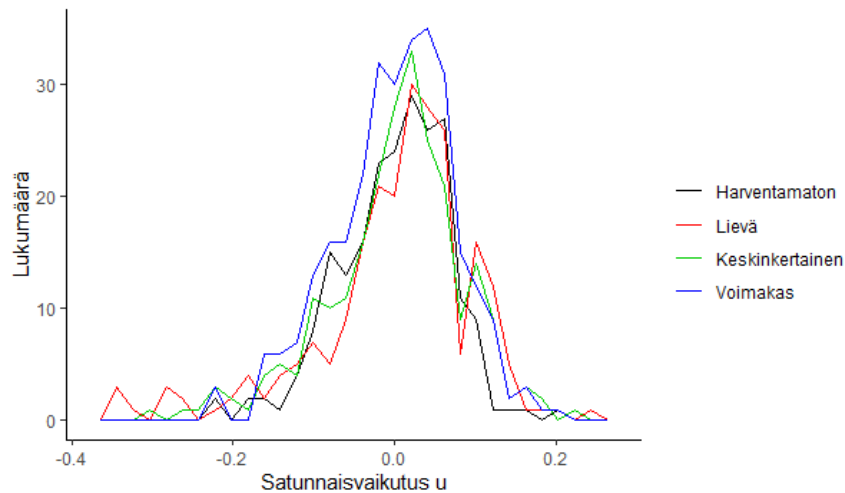
3.2.4. Sekamallit. Mallin 4a parametrien posteriorikeskiarvot ja posteriorivälit on koottu harvennusluokittain taulukoihin 4 ja 5. Parametri b_0 jakaa havainnot kahteen ryhmään: harventamattomalle ja lievästi harvennetulle metsälle posteriorikeskiarvo on 0.0125 ja keskinkertaiselle sekä voimakkaalle harvennukselle 0.0141. Parametrin b_1^0 arvo on noin 0.2 ja normaalijakautuneen satunnaisvaikutuksen u keskihajonta noin 0.09. Malliin 3b verrattuna parametrin η posteriorikeskiarvot ovat noin 10 % pienemmät, kun puutason korrelaatio on otettu mukaan. Lisäksi parametrin δ arvo on pienempi ja 95 %:n posterioriväli sisältää nollan harventamattoman ja keskinkertaisen harvennuksen tapauksessa. Parametrin σ posteriorikeskiarvo putoaa noin puoleen mallin 3b arvosta, mutta toisaalta osa vaihtelusta siirtyy puukohtaiseen vaihteluun. Puukohtaista satunnaisvaihtelua kuvaavan parametrin u posteriorikeskiarvot ovat melko lähellä normaalijakaumaa, kuten kuvasta 3.6 voi havaita.

TAULUKKO 4. Mallin 4a parametrien b_0^0 , b_1^0 (satunnaisvaikutus $u_{ij} = 0$ ja ajanhetki 1963) ja σ_1 posteriorikeskiarvot ja 95 %:n posteriorivälit.

Harv.	$b_0(10^{-2})$	$b_0(95\%)$	b_1^0	$b_1^0(95\%)$	$\sigma_1(\text{est.})$	$\sigma_1(95\%)$
H	1.25	(1.18, 1.33)	0.212	(0.204, 0.217)	0.075	(0.061, 0.092)
L	1.25	(1.18, 1.32)	0.211	(0.202, 0.219)	0.103	(0.083, 0.127)
K	1.41	(1.35, 1.48)	0.199	(0.193, 0.205)	0.091	(0.077, 0.105)
V	1.41	(1.36, 1.57)	0.208	(0.202, 0.214)	0.080	(0.070, 0.092)

TAULUKKO 5. Mallin 4a parametrien η , δ ja σ posteriorikeskiarvot ja 95 %:n posteriorivälit.

Harv.	$\eta(10^{-3})$	$\eta(95\%)$	$\delta(\text{est.})$	$\delta(95\%)$	$\sigma(\text{est.})$	$\sigma(95\%)$
H	-6.1	(-6.3, -5.8)	0.039	(-0.102, 0.176)	0.94	(0.72, 1.22)
L	-5.9	(-6.1, -5.7)	0.239	(0.081, 0.390)	1.30	(0.95, 1.72)
K	-5.2	(-5.4, -5.0)	0.111	(-0.042, 0.259)	0.91	(0.68, 1.17)
V	-5.5	(-5.7, -5.3)	0.156	(0.039, 0.280)	0.96	(0.76, 1.21)

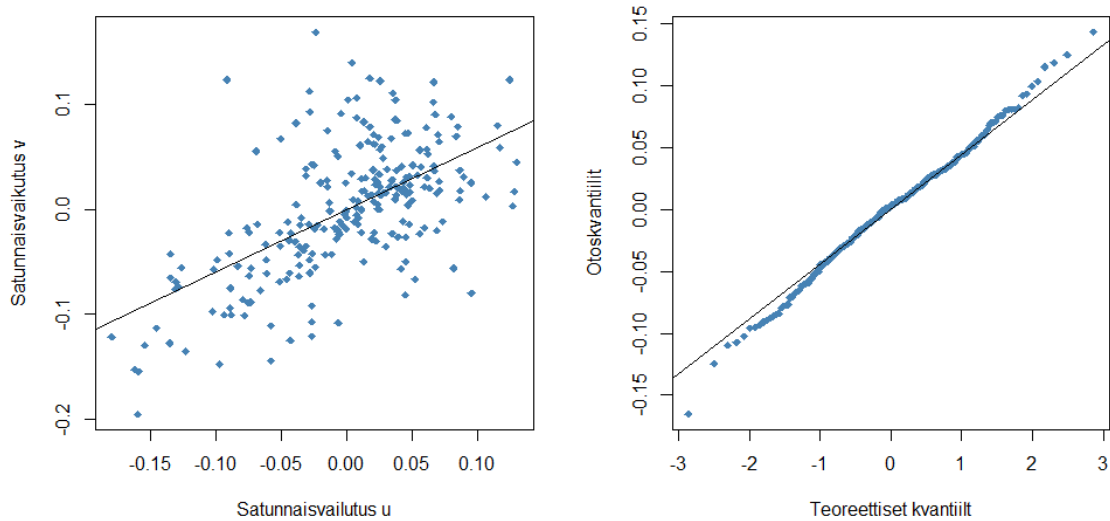


KUVA 3.6. Mallin 4a satunnaisvaikutusten u posteriorikeskiarvojen jakautuminen harvennusluokittain noudattaa melko hyvin normaalijakaumaa.

Mallin 4b posteriorikeskiarvot ja 95 %:n posteriorivälit on koottu taulukkoon 6. Mallin laskennallisen kuormittavuuden vuoksi sovitus tehtiin ainoastaan keskinkertaisesti harvennettujen koealojen tietoja käyttäen. Parametrien b_i^0 arvot ovat noin 20 % suuremmat kuin mallin 4a tapauksessa. Parametrin δ posterioriväli sisältää arvon 0, joten ei ole selvää näyttöä siitä, että varianssi riippuisi läpimitasta tässä mallissa. Parametrien η_0 ja η_1 arvot ovat negatiivisia, joten parametrien b_0 ja b_1 arvot pienenevät metsän varttuessa. Satunnaisvaikutusten hajontakuviota on esitetty kuvassa 3.7. Satunnaisvaikutusten u ja v korrelaatio on $\rho = 0.61$. Kvantiili–kvantiili-kuvaajan perusteella sovituksen residuaalit ovat normaalijakautuneet.

TAULUKKO 6. Mallin 4b parametrien posteriorikeskiarvot ja 95 %:n posteriorivälit. Laskennallisen vaativuuden vuoksi malli sovitettiin ai-noastaan keskinkertaisesti harvennettujen koealojen tietoja käyttäen.

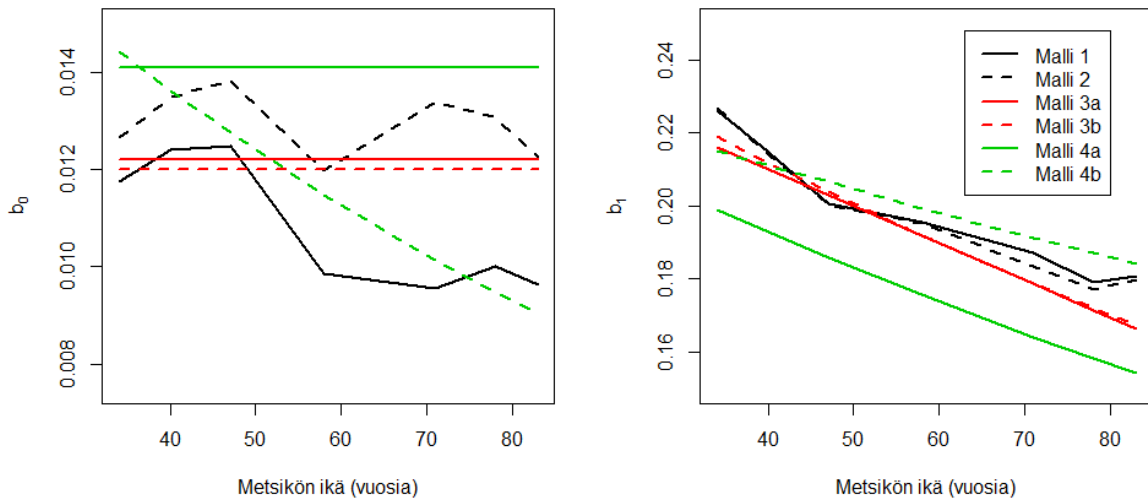
Parametri	Posteriorikeskiarvo	95 % ETI
b_0^0	0.019	(0.018, 0.021)
b_1^0	0.236	(0.226, 0.248)
δ	0.110	(-0.036, 0.255)
η_0	-3.44	(-4.11, -2.77)
η_1	-1.15	(-1.26, -1.04)
σ	0.71	(0.54, 0.92)
σ_0	0.063	(0.032, 0.205)
σ_1	0.083	(0.073, 0.093)



KUVA 3.7. Mallin 4b satunnaisvaikutusten u ja v hajontakuviio ja kvantiili–kvantiili-kuvaaja keskinkertaisesti harvennettuihin koealoihin sovitettuna. Satunnaisvaikutusten posteriorikeskiarvojen välillä on lineaarinen riippuvuus.

Kaikkien sovitettujen mallien parametrien b_0 ja b_1 posteriorikeskiarvojen aikakehitykset on koottu keskinkertaisesti harvennetulle metsälle kuvaan 3.8. Tutkituista malleista 4a:n parametrit poikkeavat selvästi muista malleista. Parametrin b_0 arvo on muita malleja noin 20 % suurempi ja vastaavasti parametri b_1 muita malleja noin 10 % pienempi. Mallien 4a ja 4b b_0 -parametrit ovat sopusoinnussa nuoressa metsässä, mutta aikariippuvuuden ja satunnaisvaikutuksen lisääminen parametriin b_0 näyttää muuttavan posteriorikeskiarvoja selvästi.

Sekamallien sopivuutta dataan voidaan tutkia jäännösvarianssin avulla. Kuten mallien 3a ja 3b tapauksissa, jäännösvarianssi riippuu harvennusluokasta ja puun läpimitasta. Jos verrataan edellisen luvun tapaan jäännösvarianssia läpimitaltaan



KUVA 3.8. Mallien 1, 2, 3a, 3b, 4a ja 4b parametrien aikakehitys keskinkertaisesti harvennetussa metsässä. Mallille 4b keskiläpimitan on oletettu olevan $0.08 + 0.00277 \cdot [\text{Ika}_i - 34]$ m.

25 senttimetrin paksuiselle puulle, jolle satunnaissvaikutus on nolla, saadaan keskinkertaisesti harvennetussa metsässä

$$\sqrt{0.91^2 \cdot (0.25 \text{ m})^{2 \cdot 0.111}} \approx 0.78 \text{ m} \quad (\text{Malli 4a})$$

$$\sqrt{0.71^2 \cdot (0.25 \text{ m})^{2 \cdot 0.110}} \approx 0.61 \text{ m}. \quad (\text{Malli 4b})$$

Mallilla 3a ja 3b jäännösvarianssit olivat vastaavasti 1.63 m ja 1.59 m, joten malleilla saadaan tarkempia ennusteita, mikäli puukohtainen satunnaisvaikutus puulle pystytään arvioimaan tarkasti.

3.3. Pieni data ja priorit

Frekventistisessä lähestymistavassa ongelmana on, että mallin parametrien määrittäminen on puhtaasti näiden muutaman mittauksen varassa, kun taas Bayes-menetelmällä ennakkotietoa voidaan ottaa mukaan priorien muodossa. Luvun 2 sekamallit 4a ja 4b sopivat koko aineistoon ainakin jäännösvarianssitarkastelujen valossa hyvin, kun puukohtainen satunnaisvaikutus saadaan arvioitua luotettavasti. Potentiaalinen ongelmatilanne voisi tulla eteen pienellä datalla, jossa satunnaisvaikutusten tarkka arviointi on ongelmallista vähäisen puumäärän vuoksi. Sekamallien toimivuutta pienellä datalla testattiin arpomalla 100 kertaa ensimmäisellä mittauskerralla mukana olleista keskinkertaiseen harvennusluokkaan kuuluneista koepuista kolme (kts. perustelu lukumäärälle luvusta 3.4) ja ottamalla mukaan ainoastaan näistä puista olevat mittaukset. Vertailukohtana pienellä datalla saatuihin posteriorikeskiarvoihin käytetään koko datalla tehdyllä sovituksella saatuja arvoja. Prioritiedon merkityksen tutkimiseksi kappaleen 2 mallien hyvin epäinformatiivisten priorien lisäksi mallit sovitettiin kolmella puulla käyttäen heikosti informatiivisia prioreja. Ero epäinformatiivisiin prioreihin on, että heikot priorit rajaavat parametrien kertaluokan oikeaksi, kun taas epäinformatiiviset priorit ovat yleisesti ottaen kertaluokkaa leveämmät. Esimerkiksi mallissa 4a

parametri δ on rajattu välille $(0, 0.5)$ ja η välille $(-0.01, 0)$, kun taas epäinformatiiviset priorit näille olivat $\text{Tas}(-2, 2)$ ja $\text{Tas}(-0.1, 0)$.

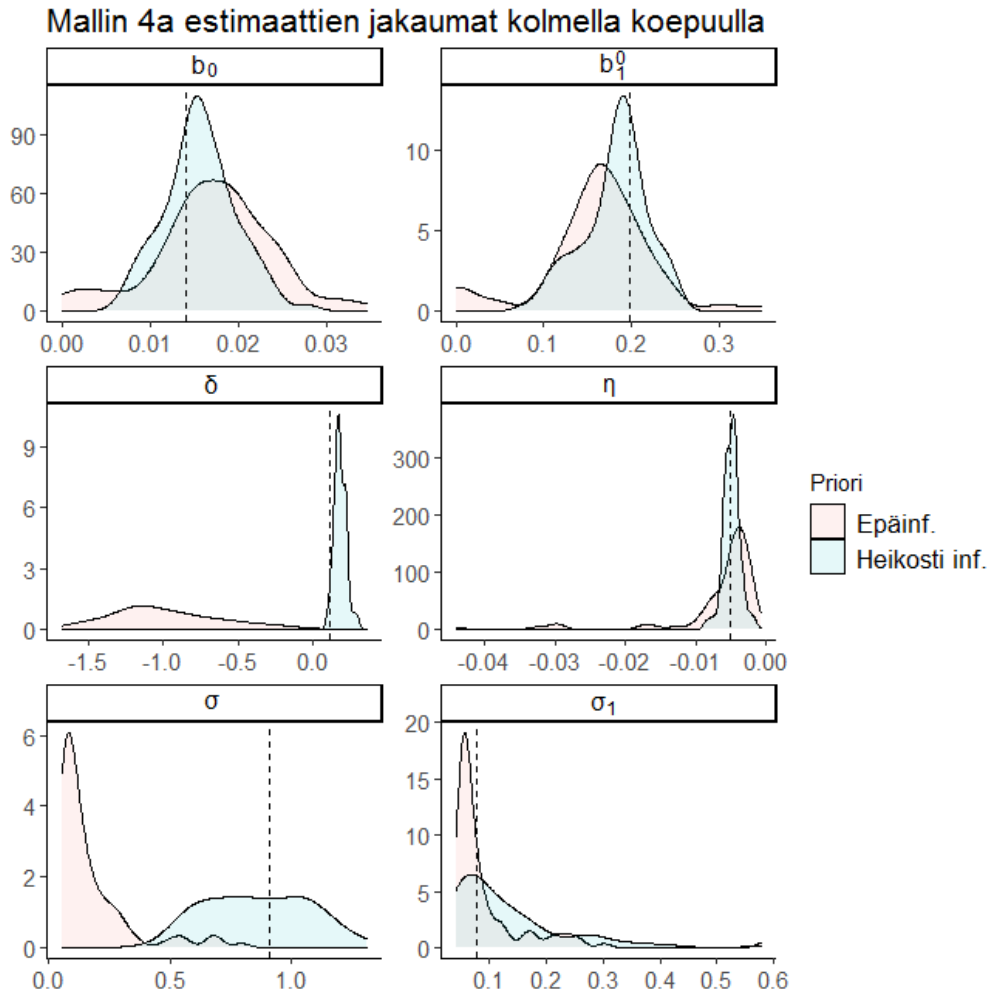
$$\begin{aligned}
 \alpha_1 &\sim N(-5, 5^2) \\
 \alpha_0 &\sim N(-2, 2^2) \\
 \sigma^2 &\sim \text{Tas}(10^{-3}, 4) \\
 \sigma_1^2 &\sim \text{Tas}(10^{-3}, 1) \\
 \delta &\sim \text{Tas}(0, 0.5) \\
 \eta &\sim \text{Tas}(-0.01, 0)
 \end{aligned}
 \tag{Heikot priorit 4a}$$

$$\begin{aligned}
 \alpha_1 &\sim N(-5, 5^2) \\
 \alpha_0 &\sim N(-2, 2^2) \\
 \sigma^2 &\sim \text{Tas}(10^{-3}, 10) \\
 \sigma_0^2 &\sim \text{Tas}(10^{-3}, 1) \\
 \sigma_1^2 &\sim \text{Tas}(10^{-3}, 1) \\
 \delta &\sim \text{Tas}(-0.1, 1) \\
 \eta_1 &\sim \text{Tas}(-3, 0) \\
 \eta_0 &\sim \text{Tas}(-6, 0)
 \end{aligned}
 \tag{Heikot priorit 4b}$$

Posteriorikeskiarvojen jakaumat mallille 4a on esitetty kuvassa 3.9. Heikot priorit korjaavat parametrien b_0 ja b_1^0 jakaumia lähemmäs koko datasta saatua posteriorikeskiarvoa. Lisäksi jakaumat ovat hieman kapeampia ja poikkeavia arvoja on epäinformatiivisia prioreja käyttäen saatuihin arvoihin verrattuna selvästi vähemmän. Parametrilla δ ero jakaumien välillä on hyvin merkittävä: epäinformatiivisella priorilla jakauma on lähes kokonaisuudessaan negatiivisen puolella, mutta ei-negatiiviseksi rajattu heikosti informatiivinen priorin vetää jakauman hyvin lähelle koko aineistolla saatua posteriorikeskiarvoa. Parametrilla η heikko priorin rajaa pois poikkeavimmat arvot ja parametrilla σ siirtää odotusarvon lähelle koko datan posteriorikeskiarvoa. Parametrilla σ_1 priorin merkitys on hieman epäselvämpi.

Posteriorikeskiarvojen jakaumat mallille 4b on puolestaan esitetty kuvassa 3.10. Myös mallilla 4b heikko prioritieto johtaa selvästi parempiin arvioihin lukuun ottamatta parametreja η_1 ja σ_0 , jolle epäinformatiiviset priorit eivät tuota selkeästi parempia tuloksia.

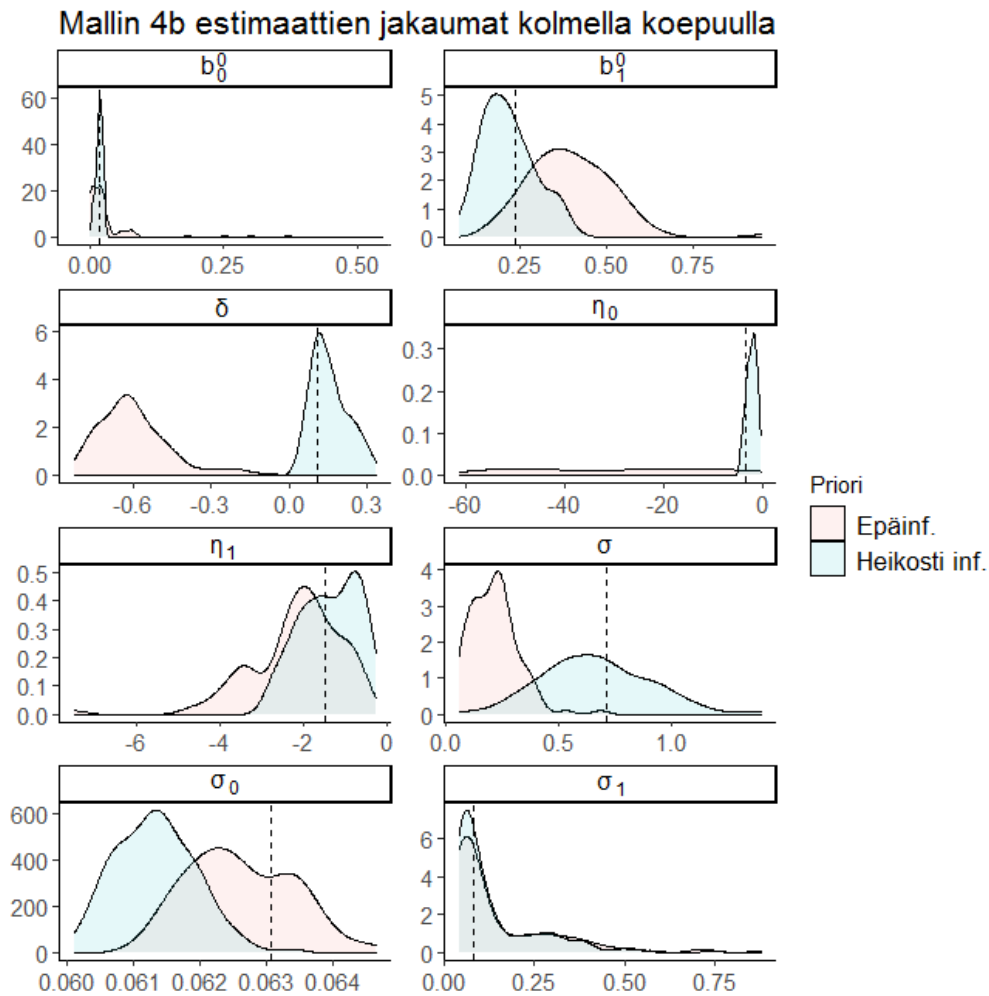
Tämän kappaleen tarkastelut osoittavat, että näin pienellä datalla mallit 4a ja 4b ovat ongelmallisia. Parametrien posteriorikeskiarvot riippuvat voimakkaasti valituista puista. Erityisenä ongelmana ovat puut, joista ei ole mittaustietoja kaikilta mittauskerroilta poistuman vuoksi. Sekamallien sijaan tällaisessa tilanteessa lienee parasta käyttää mallia 3b.



KUVA 3.9. Mallin 4a parametrien posteriorikeskiarvojen jakaumat kolmella satunnaisesti valitulla puulla luvun 2 epäinformatiivisilla prioreilla ja luvun 3.4 heikosti informatiivisilla prioreilla. Katkoviiva kuvaa posteriorikeskiarvoja, kun malli on sovitettu koko aineistoa käyttäen.

3.4. Sovellus metsien inventointeihin

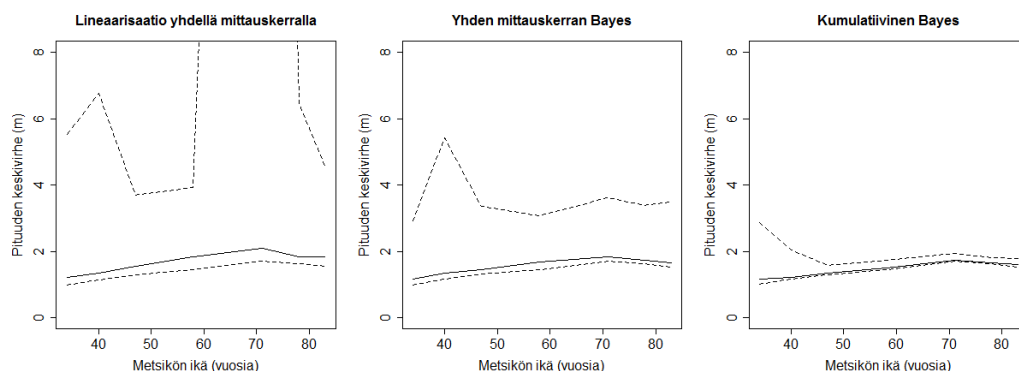
Bayes-mallinnuksen etua linearisointiin perustuvaan mallinnukseen verrattuna tutkittiin keskinkertaisesti harvennettujen ruutujen koepuiden tietojen avulla, sillä keskinkertainen harvennus vastaa parhaiten nykyaikaista harvennuskäytäntöä [22]. Tavallisen metsän inventaariokäytänteiden mukaisesti koepuita on yleensä 0–3 kappaletta koealaa kohti (esim. valtakunnan metsien inventointi [23]), joten vertailu tehtiin arpomalla kolme koepuuta kaikkien keskinkertaisesti harvennettujen ruutujen koepuiden joukosta ja vertaamalla saatuja pituuden ennusteita mitattuihin pituuksiin. Luvun 1 aineiston tarkastelujen perusteella noin puolet puista katoaa tarkastelujakson aikana luonnollisen poistuman ja harvennusten vuoksi. Niinpä ei voida olettaa, että valitut kolme koepuuta olisivat metsässä vielä seuraavalla mittauskerralla. Lisäksi satunnaisefektejä ei voida määrittää ensimmäisellä mittauskerralla, sillä samasta



KUVA 3.10. Mallin 4a parametrien posteriorikeskiarvojen jakaumat kolmella satunnaisesti valitulla puulla luvun 2 epäinformatiivisilla prioreilla ja luvun 3.4 heikosti informatiivisilla prioreilla. Katkoviiva kuvaa posteriorikeskiarvoja, kun malli on sovitettu koko aineistoa käyttäen.

puusta ei ole vielä useampaa mittausta. Tästä johtuen mallit 4a ja 4b eivät ole sopivia tässä tapauksessa, vaan Bayes-malleista on valittu malli 3b.

Linearisointiin perustuva sovitus sekä Bayes-mallin 3b sovitus tehtiin ensin käyttäen kunkin mittauskerran kolmea puuta. Tämän jälkeen Bayes-mallin sovitus tehtiin käyttäen lisäksi kaikkien aiempien mittauskertojen puuta (kolme per mittauskerta). Mallin 3b prioreina käytettiin jakaumia



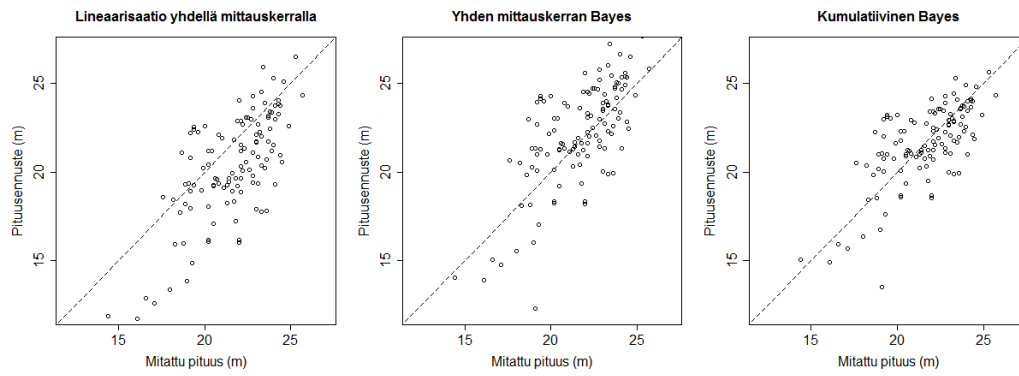
KUVA 3.11. Keskinäisesti harvennettujen koepuiden pituusennusteiden keskiarvo Bayes-mallilla 3b ja linearisointiin perustuvalla mallilla sovitettuna kolmella puulla. Yhtenäinen viiva vastaa 100 sovituksen mediaania, ja katkoviivat 5 %:n ja 95 %:n kvantiileja.

$$\begin{aligned}
 \alpha_1 &\sim N(-4.5, 1^2) \\
 b_0 &\sim \text{logNormal}(-1.5, 1^2) \\
 \sigma^2 &\sim \text{Tas}(10^{-3}, 4) && \text{(Informatiiviset priorit 3b)} \\
 \delta &\sim \text{Tas}(0, 1) \\
 \eta &\sim \text{Tas}(-0.01, 0).
 \end{aligned}$$

Priorien valinnassa on pyritty siihen, että huonosta tuurista koepuiden valinnassa johtuvat epäuskottavat parametrien arvot on rajattu pois, mutta parametrit määräytyvät kuitenkin suurimmaksi osaksi mittausdatan eivätkä priorien perusteella. Mallin sovitus tehtiin 100 kertaa arpomalla kolme koepuuta jokaiselta mittauskerralta dataksi. Bayes-mallin sovitukset tehtiin 20 000 iteraatiolla ja 10 000 pituisella lämmitteilyjaksolla.

Saatujen vertailujen tulokset on esitetty kuvassa 3.11. Yhtenäinen viiva vastaa pituuden jäännöskehajonnan mediaania ja katkoviivat 5 %:n ja 95 %:n kvantiileja. Priorijakaumien ottaminen mukaan näyttäisi lieventävän kaikkein poikkeavimpia tuloksia pitäen pituusennusteen keskihajonnan 95 % kvantiilin yhtä mittauskertaa lukuun ottamatta alle neljän metrin. Linearisoidussa mallissa puolestaan 95% kvantiili karkaa joillain mittauskerroilla yli kahdeksan metrin. Huomion arvoista tässä on se, että sovitukset on tehty käyttäen täsmälleen samoja mittauksia. Aikaisempien mittauskertojen tulosten ottaminen mukaan pienentää ennusteen virhettä huomattavasti, ja tulokset näyttäisivät konvergoivan kolmen mittauskerran jälkeen.

Esimerkkinä pituusennusteista kolmella satunnaisesti valitulla puulla on esitetty kuvassa 3.12. Tässä tapauksessa linearisointi antoi harhaisen estimaatin (huom. harhankorjauksen jälkeen), kun taas prioritiedon ottaminen mukaan johti selvästi vähemmän harhaisiin ennusteisiin. Lisäksi aiempien mittauskertojen ottaminen mukaan pienensi virhettä selvästi.



KUVA 3.12. Keskinertaisesti harvennettujen koepuiden pituusennusteet ja mitatut pituudet vuonna 2011. Estimoinnit on tehty käyttäen kolmea satunnaisesti valittua puuta vuoden 2011 mittauksesta. Kumulatiiviselle Bayes-mallille on lisäksi käytetty kolmea satunnaista puuta kaikilta aiemmilta mittauskerroilta.

Johtopäätökset

Tässä tutkielmassa tutkittiin puiden pituusmallintamista Bayes-tilastotieteen keinoin. Aineistona käytettiin Parkanossa sijaitsevista mäntykokeista tehtyjä mittauksia. Kohteita oli neljä, ja niissä oli yhteensä 17 tuhannen neliön (0.1 ha) suuruista koealaa. Kaikissa kohteissa harvennusvoimakkuus vaihteli harventamattomasta voimakkaasti harvennettuun. Metsikkö oli mitattu yhteensä seitsemän kertaa.

Tässä työssä rakennetut mallit perustuvat Näslundin pituusikäykäyrään [12], joka on kaksiparametrinen funktio, joka antaa puulle pituusennusteen läpimitan funktiona. Käyrän parametrien tarkastelu ajan funktiona osoitti, että parametrilla b_0 ei ole selkeää aikariippuvuutta, mutta mikäli arvo kehittyy johonkin suuntaan, se todennäköisimmin pienenee metsän varttuessa. Parametrin b_1 arvo puolestaan pienenee selvästi metsän varttuessa. Parametrin b_1 aikariippuvuutta voidaan mallintaa esimerkiksi eksponenttifunktion avulla. Parametrin b_1 tarkkuuden osoitettiin olevan kahdesta parametrusta tärkeämpi tarkkojen pituusennusteiden johtamiseen, joten laskeva (esim. eksponentiaalinen) aikariippuvuus parametrissa b_1 on olennainen ominaisuus Näslundin pituusikäykäyrään perustuvassa mallissa. Tässä työssä kehitetyissä malleissa 3a, 3b ja 4a (ks. luku 2) aikariippuvuus on mukana metsikön iän kautta. Tämä on toimiva ratkaisu tasaikäisille metsille, mutta mikäli metsikkö koostuu eri-ikäisistä puista, ei muuttujalle ole selkeää tulkintaa. Tässä tapauksessa voidaan käyttää aikamuuttujan paikalla metsikön läpimittaa (esim. malli 4b) tai sen logaritmia, joka on iän kanssa voimakkaasti korreloitunut. Työssä käytetty aineisto on hyvin homogeeninen. Tulosten yleistämiseksi erilaisille metsätyypeille ja ilmastoille voi olla tarpeen ottaa nämä muuttujat huomioon parametreissa. Lisäksi suurissa aineistoissa, joissa koealoja on paljon, voi olla tarpeen ottaa mukaan myös metsikkötason satunnaistekijät molempiin parametreihin.

Pituuden ja läpimitan todettiin graafisten tarkastelujen perusteella (ks. luku 1) olevan puutasolla korreloituneita mittauskertojen välillä. Toisin sanoen keskimääräistä ohuempia tai paksumpia puut olivat todennäköisesti myös tulevaisuudessa keskimääräistä ohuempia tai paksumpia. Tarkassa mallinnuksessa parametreihin b_0 ja b_1 otettiin mukaan puutason satunnaisefektit (malli 4b). Ottamalla puutason satunnaisefektit mukaan päädyttiin selvästi pienempiin jäännöskehajontoihin kuin malleilla 3a ja 3b, mikäli puutason satunnaisefekti oletettiin tarkasti tunnetuksi. Hyvin pieneen aineistoon malli puutason satunnaisefekteillä ei kuitenkaan sovi, sillä yksittäisistä puista pitäisi olla useita havaintoja, ja lisäksi puuta on oltava useita, jotta satunnaisvaikutusten määrittäminen tarkasti onnistuisi.

Mallien posteriorikeskiarvojen määritymistä pienellä määrällä mittauksia tutkittiin valitsemalla 100 kertaa kolme satunnaista puuta, joiden mittauksiin perustuen

posteriorikeskiarvot määritettiin. Jopa hyvin heikon prioritiedon todettiin tarkentavan estimointia tasaamalla huonosta koepuiden valintaonnesta johtuvia poikkeuksellisia parametrien arvoja. Hyvin määritetyillä priorijakaumilla voitaisiin siis laajentaa monimutkaisempien, puutason satunnaisvaikutuksia sisältävien mallien toimivuutta myös pienempiin aineistoihin. Ennen kuin tarvittavaa prioritietoa saadaan lisää, on malli ilman puutason satunnaisvaikutuksia kuitenkin toimivampi aineistoissa, joissa samasta metsiköstä on mitattu ainoastaan muutama puu. Malleja, joissa ei ole satunnaisefektejä, voidaan käyttää myös aineistoihin, joissa toistomittauksia ei ole tehty. Tällaista tilannetta vastaava sovelluskohde voisi löytyä valtakunnan metsien inventoinneista. Pituusennusteita tutkittaessa Bayes-mallin osoitettiin antavan selvästi parempia pituusestimaatteja, kun käytettävissä on myös mittauksia aiemmilta vuosilta. Esimerkiksi valtakunnan metsien inventoinnissa osasta koealoja on tehty mittauksia useana ajanhetkenä [23], joten aiempia mittauksia hyödyntävistä malleista voisi olla erityistä hyötyä tässä yhteydessä.

Bayes-mallinnuksen keskeisenä ongelmana on laskennallinen vaativuus. Puutason satunnaisefektejä sisältävien mallien sovitus tämän työn varsin rajalliseen aineistoon vei useita tunteja. Niinpä käytännön sovelluksia ajatellen laskennallisen vaativuuden pienentäminen olisi tärkeää. Erilaisten laskennallista vaativuutta vähentävien approksimaatioiden, kuten INLA:n (*Integrated Nested Laplace Approximation*) [24], käyttökelpoisuutta pituuskäyrien sovituksessa voitaisiin tutkia. Erityistä tarvetta laskennallisen vaativuuden pienentämiseen on, jos malleja päätetään soveltaa esimerkiksi valtakunnan metsien inventointiin, jossa koealoja on runsaasti, mutta prioritieto on tärkeää yksittäisten koealojen vähäisen koepuiden määrän vuoksi. Aineistoissa, joissa samalta koealalta on tehty runsaasti mittauksia, ei prioritiedosta ole merkittävää hyötyä, jolloin sovitukset voi tehdä ilman prioreja regressiolla. Mallit voidaan sovitaa suoraan epälineaarilla regressiolla, mutta tällöin satunnaisvaikutusten ennustaminen vaatii iteratiivisen ratkaisun [13], joten menettelytapa ei ole selvästi Bayes-mallinnusta yksinkertaisempi. Toisena vaihtoehtona on linearisoidun mallin käyttäminen. Koko tutkimusaineistoon tehtyjen sovitusten perusteella tavallisella lineaarisella regressiolla sovitetun linearisoidun mallin estimaatit parametreille vastasivat hyvin Bayes-mallin estimaatteja. Tämä laskennallisesti kevyt menetelmä sopii siis ainakin sellaisiin kerran mitattuihin koealoihin, joissa koepuita on yli sata kappaletta.

Kirjallisuutta

- [1] MEHTÄTALO L., DE-MIGUEL S. ja T.G. GREGOIRE: *Modeling height-diameter curves for prediction*, Canadian Journal of Forest Research, 4, 826-837, (2015).
- [2] HUANG S., PIERCE D. ja WIENS D.P.: *Development of ecoregion-based height-diameter models for white spruce in boreal forests*, Forest ecology and management, 129, 125–141 (2000).
- [3] CURTIS R.O.: *Height-diameter and height-diameter age equations for second-growth Douglas fir*, Forest Science, 13, 4, 365–375 (1967).
- [4] SCHUMACHER F.X.: *A new growth curve and its application to timber yield studies*, Journal of Forestry, 37, 819–820, (1939).
- [5] STRAND L.: *The accuracy of some methods for estimating volume and increment on sample plots*, Medd. norske Skogfors, 15, 4, 284–392, (1959).
- [6] LUNDQVIST B.: *On the height growth in cultivated stands of pine and spruce in Northern Sweden*, Medd. frstatens skogforsk, 133 (1957).
- [7] MALTAMO M., EERIKÄINEN K., PACKALÉN P. ja HYYPPÄ J.: *Estimation of stem volume using laser scanning-based canopy height metrics*, Forestry, 79, 2 (2006).
- [8] KUUSISTO L. ja KANGAS A.: *Harhakomponentit kuvioittaisen arvioinnin puuston tilavuuden laskentaketjussa*, Metsätieteen aikakauskirja 3/2008, 177–190 (2008).
- [9] MUSTONEN J., PACKALÉN P ja KANGAS A.: *Automatic segmentation of forest stands using a canopy height model and aerial photography*, Scandinavian Journal of Forest Research, 23, 534–545 (2008).
- [10] KANGAS A., MÄKINEN H. ja LYHYKÄINEN H.T.: *Value of quality information of Scots pine stands in timber bidding*, Canadian Journal of Forest Research, 40, 1781–1790, (2010).
- [11] SIIPILEHTO J.: *Methods and applications for improving parameter prediction models for stand structures in Finland*, Dissertationes Forestales, 124 (2011).
- [12] NÄSLUND M.: *Skogsförsöksanstaltens gallringsförsök i tallskog*, Meddelanden från Statens Skogsförsöksanstalt 29, (1936).
- [13] SIRKIÄ S., HEINONEN J., MIINA J. and EERIKÄINEN K.: *Subject-specific prediction using a nonlinear mixed model: consequences of different approaches*, Forest Science, 61, 2, 205–212, (2015).
- [14] SIIPILEHTO J. and KANGAS A.: *Näslundin pituuskäyrä ja siihen perustuvia malleja läpimitan ja pituuden välisestä riippuvuudesta suomalaisissa talousmetsissä*, Metsätieteen aikakauskirja 4/2015 205–236, (2015).
- [15] K. EERIKAINEN ja K.T. KORHONEN: *Pituusmallit*, Julkaisussa *Metsää kuvaava malli*, Silva Carelica, (2001).
- [16] GELMAN A, CARLIN J. B., STERN H. S., DUNSON D. B., VEHTARI A. ja RUBIN D. B.: *Bayesian Data Analysis*, Chapman & Hall, CRC 3. painos (2013).
- [17] R CORE TEAM: *A Language and Environment for Statistical Computing*. , R Foundation for Statistical Computing, Vienna, Austria (2019). <https://www.R-project.org/>.
- [18] PLUMMER M.: *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*, Julkaisussa Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria (2003).
- [19] ADAMS R.A. ja ESSEX C.: *Calculus: A Complete Course*, 8. painos, Pearson Canada Inc., (2013).
- [20] MEHTÄTALO L. ja LAPPI J.: *Biometry for Forestry and Environmental Data with Examples in R*, Kirja valmistumassa (2020).

- [21] GELFAND A. ja SMITH A.: *Sampling-Based Approaches to Calculating Marginal Densities*, Journal of the American Statistical Association, 85, 398-409 (1990).
- [22] HARRI MÄKINEN, LUONNONVARAKESKUS (SUUL.)
- [23] TOMPPO E. YM. *Designing and conducting a forest inventory-case: 9th National Forest Inventory of Finland*, Springer, New York (2011).
- [24] RUE, H., MARTINO, S., ja CHOPIN, N.: *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71, 2, 319-392 (2009).