METSIKÖN LÄPIMITTAJAKAUMAN ENNUSTAMINEN METSIKKÖTUNNUSTEN JA JÄRJESTYSTUNNUSLUKUPUIDEN AVULLA

Joensuun Yliopisto
Yhteiskuntatieteellinen tiedekunta
Tilastotieteen laitos
Tilastotieteen sivuainelaudatur-työ
Syksy 2004/ Jukka Nyblom

Lauri Mehtätalo
129 782

SISÄLLYS

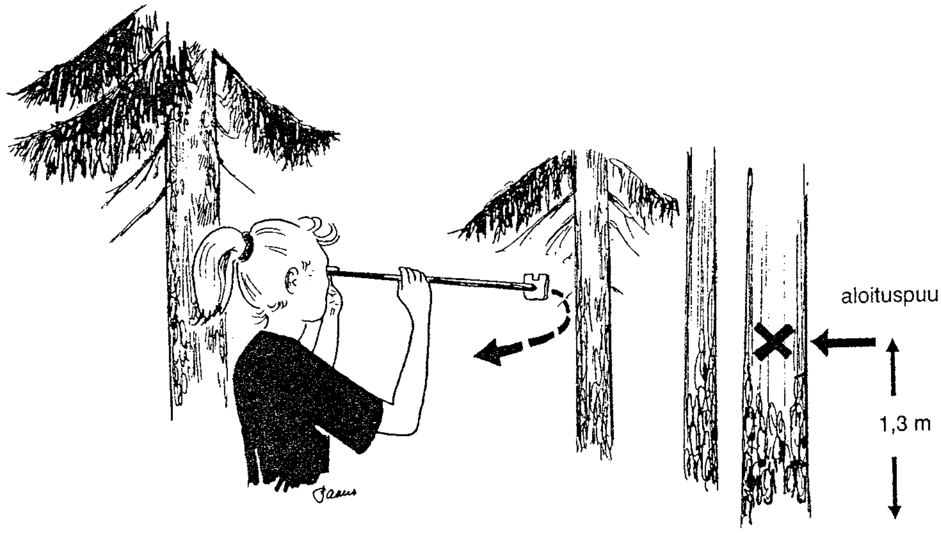
1. Johdanto

Metsäalueen käsittelyn suunnittelussa tarvitaan perustietoa suunnittelun kohteena olevasta metsäalueesta. Ajatellaan vaikka että olemme laatimassa metsäsuunnitelmaa 100 hehtaarin metsäalueelle. Ensimmäiseksi tulee ratkaista, kuinka yksityiskohtaista tietoa tarvitaan ja kuinka tarvittava tieto voidaan kerätä niin, että maastotiedon keruun kustannukset eivät kohoa liian korkeiksi mutta saatu tieto on niin yksityiskohtaista, että sitä voidaan hyödyntää suunnittelussa. Jos metsäalueen puusto kuvataan pelkästään keskitunnusten avulla, ei esimerkiksi suunniteltuja hakkuita voida kohdentaa tiettyyn metsäalueen osaan. Toisaalta puukohtaisen tiedon kerääminen tulisi aivan liian kalliiksi saavutettuun hyötyyn nähden. Siksi metsäalueen suunnittelussa metsäalue jaetaan pieniin alueisiin, jotka ovat suunnittelun ja maastotiedon keruun perusyksiköitä. Näitä alueita kutsutaan metsiköiksi. Metsikkö on maapohjan ja puuston ominaisuuksien suhteen sisäisesti homogeeninen, sopivan kokoinen (0.5-5 ha) metsäalueen osa, joka on sopiva käsittely-yksikkö metsässä tehtävissä toimenpiteissä.

Eräs tärkeä metsikköä kuvaava suure on sen läpimittajakauma. Sillä tarkoitetaan metsikön puiden läpimittojen jakaumaa. Läpimittajakauman kertymäfunktio on funktio $F(d)=P(D\leq d)$, jossa D on metsiköstä satunnaisesti valitun puun läpimitta 1.3 metrin korkeudelta (ns. rinnan korkeudelta). Läpimittajakauma täyttää todennäköisyysjakauman ominaisuudet, mutta yleensä läpimittajakaumaa ei tulkita todennäköisyysjakaumaksi vaan tiheysjakaumaksi, joka kertoo kuinka suuri osuus metsikön puista on läpimitaltaan pienempiä kuin annettu läpimitta d.

Koska suuret puut ovat taloudellisesti arvokkaampia kuin pienet, käytetään yleensä pohjapinta-alalla painotettua läpimittajakaumaa, jossa puun pohjapinta-alalla tarkoitetaan puun poikkileikkauksen pinta-alaa rinnankorkeudelta. Pohjapinta-alalla painotettu läpimittajakauma kertoo, kuinka suuri osuus metsikön puiden yhteenlasketusta pohjapinta-alasta on pienempiä kuin läpimitta d. Jakauman todennäköisyystulkinnassa pohjapinta-alalla painotettu läpimittajakauma kertoo millä todennäköisyydellä PPS (Propability Proportional to Size) -otannalla valitun puun läpimitta on pienempi kuin annettu läpimitta d, kun yksittäisen puun kokoa kuvataan sen pohjapinta-alalla. Jatkossa läpimittajakaumalla tarkoitetaan pohjapinta-alalla painotettua läpimittajakaumaa.

Jos oletetaan, että eri kokoiset puut ovat sijoittuneet metsikköön täysin satunnaisesti, voidaan PPS-otanta toteuttaa metsikössä poimimalla satunnaisen pisteen ympäristöstä sellaiset puut, joiden rinnankorkeusläpimitan ja etäisyyden suhde on suurempi kuin joku vakio. Tällainen otos voidaan poimia relaskoopilla. Relaskooppi on laite, jossa kiinteän varren päähän on kiinnitetty muovilevy johon on leikattu kiinteän kokoinen hahlo. Laitetta käytetään asettamalla varren toinen pää mittaajan poskelle ja tähtäämällä satunnaisen pisteen ympäristön puita hahlon läpi rinnan korkeudelle. Pisteessä pyörähdetään täysi ympyrä ja ne puut, jotka näyttävät hahlon läpi katsottaessa sitä suuremmalta, tulevat mukaan otokseen, jota kutsutaan relaskooppiotokseksi tai relaskooppikoealaksi (Kuva 1).



Kuva 1. Relaskoopin käyttö (Savolainen 2004).

Relaskooppiotokseen tulevien puiden lukumäärä on suoraan verrannollinen metsikön puiden poikkileikkauksen yhteenlaskettuun pinta-alaan (puuston pohjapinta-alaan) hehtaarilla mittauspisteen ympäristössä. Yleisesti käytetään relaskooppia, jossa varren pituus on 50 kertaa hahlon leveys. Tällöin relaskooppikerroin on yksi, mikä tarkoittaa että otokseen kuuluvien puiden lukumäärä on suoraan puuston pohjapinta-ala (m^2/ha) mittauspisteen ympäristössä. Yleensä relaskooppikoealalta mitataan myös puuston pohjapinta-alamediaanipuun läpimitta (DGM) arvioimalla mikä otokseen kuuluvista puista on mediaanipuu ja mittaamalla sen läpimitta.

Läpimittajakauma on metsikön kuvauksen perusta. Jos tunnetaan metsikön läpimittajakauma ja metsikön pohjapinta-ala, voidaan puukohtaisen tilavuusmallin avulla laskea esimerkiksi metsikön kokonaistilavuus tai tietyn puustonosan tilavuus. Esimerkiksi tukkikokoisten runkojen tilavuus on hyvin kiinnostava metsikön tunnus, jonka laskemiseksi tarvitaan läpimittajakaumaa Läpimittajakaumasta poimittuja puita voidaan kasvattaa puukohtaisten kasvumallien avulla, jolloin voidaan arvioida puuston tulevaa kasvua. Jakauman avulla voidaan tutkia hakkuussa poistuvan ja jäävän puuston rakennetta. Metsäsuunnittelussa muodostetaan suunniteltavan alueen metsiköille erilaisia mahdollisia hakkuuohjelmia ja arvioidaan niiden vaikutuksia eri kriteerimuuttujiin, joita voivat olla esimerkiksi hakkuutulot, jäävän puuston määrä, arvo tai rakenne tai vaikkapa alueen monimuotoisuudesta tai virkistysarvosta kertovat indeksit. Hakkuuohjelmia vertailemalla etsitään sellainen hakkuuohjelma, joka toteuttaa metsänomistajan tavoitteet parhaiten.

Metsikön läpimittajakauma voidaan estimoida sovittamalla metsiköstä poimittuun otokseen joku teoreettinen jakauma ja estimoimalla parametrit esimerkiksi suurimman uskottavuuden menetelmällä. Tällöin ajatellaan, että metsikön läpimittajakauman parametrivektori on kiinteä, tuntematon vektori, joka estimoidaan mitatusta otoksesta. Metsäsuunnittelua varten tehtävässä tiedonkeruussa ei kuitenkaan ole resursseja mitata otosta.

Tässä työssä läpimittajakauman parametrivektori ajatellaan satunnaisvektoriksi, joka saa arvon kussakin metsikössä. Tällöin laajasta metsikköaineistosta voidaan mallittaa läpimittajakauman parametrien vaihtelua metsiköiden välillä. Valittu lähestymistapa muistuttaa bayesilaista lähestymistapaa siinä, että parametri ajatellaan satunnaismuuttujaksi jolla on tietty jakauma. Parametrien ennustamiseksi on laadittu valtakunnallisia regressiomalleja, joissa metsikön läpimittajakauman parametrien suurimman uskottavuuden estimaatteja selitetään muutaman nopeasti mitattavissa olevan metsikkötunnuksen (ikä, pohjapinta-ala, DGM) avulla. Tarvittavat mittaukset saadaan metsiköstä mitatuilta relaskooppikoealoilta koealamittausten keskiarvoina. Jos oletetaan, että mallit ovat oikeita, mallinnettavat parametrien suurimman uskottavuuden estimaatit ovat oikeita, mallien kertoimien arvot tunnetaan ja metsikkömittaukset ovat virheettömiä, on mallien antama ennuste metsikön läpimittajakauman parametrien ehdollinen odotusarvo $E(\boldsymbol{\theta}|\mathbf{x})$, jossa $\boldsymbol{\theta}$ on jakauman parametrien vektori ja $\mathbf{x}$ havaittu metsikkötunnusten vektori. Metsikön läpimittajakauman ennuste saadaan sijoittamalla ennustetut parametrien arvot käytettyyn jakaumaan. Useissa sovelluksissa näin saatu jakaumaennuste on kuitenkin liian epätarkka, koska läpimittajakauman muotoon vaikuttaa moni sellainen tekijä jota malleissa ei voida ottaa huomioon. Tällaisia tekijöitä ovat esimerkiksi hakkuut ja metsätuhot.

Tässä tutkimuksessa tavoitteena on kehittää menetelmä, jolla metsikkötunnusten avulla saatua jakaumaennustetta voidaan tarkentaa metsiköstä poimitun relaskooppiotoksen otosjärjestystunnuslukujen (engl. sample order statistic) avulla. Lisäksi tavoitteena on näyttää numeerisen esimerkin avulla kuinka laskenta toteutetaan ja tutkia, kuinka paljon otosjärjestystunnuslukupuiden tarkalla mittaamisella voidaan pienentää metsikön tilavuusennusteen virhettä. Otosjärjestystunnuslukujen käyttäminen lisämittauksina on perusteltua siksi, koska tietyn puun järjestysluvun määräämiseksi kaikkien otokseen sattuvien puiden läpimittoja ei tarvitse mitata vaan voidaan silmävaraisesti arvioida, mitkä otoksen puista ovat valittua otosjärjestystunnuslukupuuta suurempia ja mitkä pienempiä. Siksi otosjärjestystunnuslukujen mittaaminen metsiköstä on luultavasti huomattavasti nopeampaa kuin koko otoksen mittaaminen. Otosjärjestystunnuslukupuun mittaus on yleistys otoksen DGM:n mittauksesta.

2. LÄPIMITTAJAKAUMA

2.1. Metsikön läpimittajakauman kehittyminen

Metsikön läpimittajakaumaan vaikuttavat monet eri asiat. Eräs olennainen läpimittajakauman muotoon vaikuttava tekijä on metsikön ikärakenne. Jos metsikkö on eri-ikäisrakenteinen, mikä tarkoittaa että metsikössä on kaiken ikäisiä puita ja uudistuminen tapahtuu niin että vanhimpia puita poistetaan poimintahakkuissa ja tilalle syntyy uutta taimikkoa, läpimittajakauma on käänteisen j-käyrän muotoinen. Eri-ikäisrakenteisia metsiä suomessa on kuitenkin hyvin vähän ja pääasiassa suomalaiset metsiköt ovat tasaikäisrakenteisia. Niissä kaikki puut ovat suunnilleen saman ikäisiä ja uudistuminen tapahtuu avo- tai siemenpuuhakkuiden avulla. Tässä työssä keskitytään tasaikäisrakenteisten metsiköiden läpimittajakauman ennustamiseen.

Metsikön puulaji vaikuttaa läpimittajakaumaan huomattavasti. Aikaisemmissa tutkimuksissa (esim Bailey 1973, Maltamo 1997) on havaittu, että vain vähän varjoa sietävillä ns. valopuulajeilla, joita Suomessa ovat lehtipuut ja mänty, tasaikäisrakenteisen metsikön läpimittajakauma on usein yksihuippuinen ja kohtuullisen symmetrinen. Sen sijaan kuusella, joka sietää enemmän varjoa ja kasvaa siksi usein alikasvoksena, läpimittajakauma on leveä, lattea, vino ja usein monihuippuinen. Koska eri puulajien läpimittajakaumat poikkeavat huomattavasti toisistaan, estimoidaan läpimittajakaumat yleensä sekametsiköissäkin erikseen männylle, kuuselle ja lehtipuille. Metsikön läpimittajakauma saadaan sitten puulajiositteiden läpimittajakauman summana.

Maastossa kohtuullisen helposti arvioitavia metsiköiden ominaisuuksia ovat aiemmin mainittujen pohjapinta-ala ja pohjapinta-alamediaanipuun läpimitan lisäksi tasaikäisrakenteisen metsikön ikä, maaperän viljavuudesta kertova metsätyyppiluokka sekä puuston keskipituus. Aiemmissa tutkimuksissa on havaittu, että näiltä ominaisuuksiltaan samantyyppisissä metsiköissä läpimittajakauman muoto on samantyyppinen. Nämä ominaisuudet evät kuitenkaan selitä edes puolta läpimittajakaumien vaihtelusta. Esimerkiksi Maltamon (1997) tutkimuksissa, jossa metsiköistä mitattuihin läpimittajakaumiin sovitettiin Weibull jakauma suurimman uskottavuuden menetelmällä ja saatuja parametriestimaatteja selitettiin metsikön ominaisuuksilla, estimoitujen regressioyhtälöiden selitysasteet vaihtelivat välillä 0.213-0.896. Tyypillisesti selitysasteet ovat kohtuullisen korkeita skaala- ja sijaintiparametrilla, mutta muotoparametrin selitysasteet ovat matalia (Rennolls ym. 1985). Se osa läpimittajakaumien vaihtelusta, jota metsikön ominaisuudet eivät selitä, voi johtua esimerkiksi metsikön kehityshistoriasta, topografiasta tai tuhoista.

Koska metsiköstä voidaan mitata tunnuksia, jotka ovat johdettavissa läpimittajakaumasta, ei kaikille parametreille tarvita regressiomalleja, vaan osa parametreista voidaan johtaa analyyttisesti. Suomessa on käytetty lähestymistapaa, jossa kaksi Weibull-jakauman kolmesta parametrista ennustetaan regressiomalleilla ja kolmas ratkaistaan mitatun mediaanipuun läpimitan ja pohjapinta-alan avulla. Analyyttisesti ratkaistava parametri vaihtelee eri tutkimuksissa (esim. Kilkki ja Päivinen 1986, Maltamo 1997).

2.2. Läpimittajakauman parametrien yleinen malli

Edellä kerrotun perusteella tasaikäisrakenteisen metsikön yhden puulajiositteen läpimittajakauman parametreille metsikössä m voidaan olettaa malli

$$\boldsymbol{\theta}_m = E\left[\boldsymbol{\theta} \middle| \mathbf{x}_m\right] + \mathbf{e}_m \,, \tag{1}$$

jossa $\boldsymbol{\theta}_m$ on metsikön m puulajiositteen läpimittajakauman parametrien vektori, $\mathbf{x}_m$ metsikkötunnusten vektori ja $\mathbf{e}_m$ satunnaisten virhetermien vektori, jolle pätee E($\mathbf{e}_m$)=0 ja var($\mathbf{e}_m$)=$\mathbf{D}$. Eri parametrien virhetermit voivat olla korreloituneita, joten matriisin $\mathbf{D}$ rakenteesta ei oleteta muuta kuin että se on positiivisesti definiitti. Vektori $\mathbf{e}_m$ kertoo kuinka paljon läpimittajakauman parametrien arvot tässä metsikössä poikkeavat ehdollisista odotusarvoistaan $E\left[\boldsymbol{\theta} \middle| \mathbf{x}_m\right]$ ja sen elementtejä kutsutaan jatkossa metsikkövaikutuksiksi. Merkintöjen selvyyden vuoksi alaindeksiä m ei käytetä jatkossa, vaikka metsikkövaikutuksista tai metsikön mittauksista puhuttaessa tarkoitetaan aina tietyn metsikön tunnuksia.

2.3. Läpimittajakauman parametrien ennustaminen

Jos metsiköstä tunnetaan metsikkötunnusten vektori $\mathbf{x}$ on mallin (1) avulla saatu ennuste paras ennustin ehdollinen odotusarvo, joka saadaan mallin kiinteän osan avulla

$$\hat{\boldsymbol{\theta}} = E\left[\boldsymbol{\theta} \middle| \mathbf{x}\right]. \tag{2}$$

Oletetaan, että jokin vektorin $\boldsymbol{\theta}$ komponenteista voidaan mitata ja käytetään siitä merkintää θ. Mittausta ei voida kuitenkaan tehdä tarkasti, vaan mittaukseen liittyy joku epävarmuus niin, että parametrin θ mittaus $\tilde{\theta}$ noudattaa mallia

$$\tilde{\theta} = E(\tilde{\theta}) + \varepsilon = \theta + \varepsilon = E(\theta|\mathbf{x}) + e + \varepsilon,$$

jossa e on parametrin θ metsikkövaikutus jonka varianssi on matriisin **D** diagonaalin parametria θ vastaava alkio, σ_e^2, ja ε on satunnainen mittausvirhe jonka varianssi on σ_ε^2. On luonnollista olettaa metsikkövaikutukset ja mittausvirhe korreloimattomiksi. Tällöin

$$\text{cov}(\boldsymbol{\theta}, \tilde{\theta}|\mathbf{x}) = \mathbf{c} = \text{cov}\left(E[\boldsymbol{\theta}|\mathbf{x}] + \mathbf{e}, E(\theta|\mathbf{x}) + e + \varepsilon\right) = \text{cov}(\mathbf{e}, e),$$

joka on matriisin **D** parametria θ vastaava sarake. Tällöin vektorin $\begin{bmatrix}\boldsymbol{\theta} & \tilde{\theta}\end{bmatrix}'$ odotusarvo ja varianssi ehdolla että metsikkötunnukset **x** on havaittu ovat

$$\begin{bmatrix}\boldsymbol{\theta}|\mathbf{x} \\ \tilde{\theta}|\mathbf{x}\end{bmatrix} \sim \left(\begin{bmatrix}E(\boldsymbol{\theta}|\mathbf{x}) \\ E(\theta|\mathbf{x})\end{bmatrix}, \begin{bmatrix}\mathbf{D} & \mathbf{c} \\ \mathbf{c}' & \sigma_e^2 + \sigma_\varepsilon^2\end{bmatrix}\right) \tag{3}$$

Parametrivektorin $\boldsymbol{\theta}$ paras ennustin on

$$BP(\boldsymbol{\theta}) = E[\boldsymbol{\theta}|\mathbf{x}, \tilde{\theta}].$$

Parhaan ennustimen laskemiseksi meidän täytyisi tuntea vektorin $\begin{bmatrix}\boldsymbol{\theta} & \tilde{\theta}|\mathbf{x}\end{bmatrix}'$ jakauma. Koska tätä ei tunneta, rajoitamme tarkastelun ennustimiin, jotka ovat muotoa

$$LP(\boldsymbol{\theta}) = \mathbf{a} + \mathbf{b}\tilde{\theta},$$

eli etsimme parasta lineaarista ennustinta, missä lineaarisuus tarkoittaa lineaarisuutta mittauksen $\tilde{\theta}$:n suhteen. Parasta lineaarista ennustinta käytettäessä riittää, että tunnemme satunnaisvektorin-vektorin $\begin{bmatrix}\boldsymbol{\theta} & \tilde{\theta}|\mathbf{x}\end{bmatrix}'$ odotusarvon ja varianssi-kovarianssimatriisin. Paras lineaarinen ennustin saadaan kaavalla (McCulloch ja Searle 2000, s. 251)

$$BLP(\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{x}) + \frac{\tilde{\theta} - E(\theta|\mathbf{x})}{\sigma_e^2 + \sigma_\varepsilon^2}\mathbf{c}, \tag{4}$$

ja sen ennustevirheen varianssi on

$$\operatorname{var}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)=\mathbf{D}-\frac{1}{\sigma_e^{\,2}+\sigma_\varepsilon^{\,2}}\mathbf{cc}'. \tag{5}$$

Nämä voidaan laskea jos kaavan (3) odotusarvot ja varianssit tunnetaan. Käytännössä tunnemme parhaimmillaankin vain niiden estimaatit. Kun estimaatit kirjoitetaan kaavoihin (4) ja (5), saadaan estimoitu paras lineaarinen ennustin (McCulloch ja Searle 2000, s. 257).

Ennustin (4) hyödyntää mittauksen $\tilde{\theta}$ lisäksi vektorin **x** sisältämää prioritietoa metsikön läpimittajakaumasta sekä informaatiota näiden epävarmuudesta. Siten lähestymistavalla on yhteyksiä bayesilaiseen lähestymistapaan, jossa parametrien priorijakaumia tarkennetaan otoksen avulla. Puhtaassa bayesilaisessa lähestymistavassa priorijakauman ajatellaan kuvaavan mallittajan ennakko-oletuksia parametreista, eivätkä ne tyypillisesti perustu aineistoon. Tässä lähestymistavassa priorijakauma saadaan metsikkötunnusten avulla regressiomallilla, eikä lähestymistapaa voida siksi pitää puhtaana bayesilaisena lähestymistapana. Kirjallisuudessa tällaisesta lähestymistavasta on käytetty nimitystä empiirinen bayesilainen lähestymistapa (Casella ja Berger 2002: 371).

2.4. Läpimittajakauman kuvaaminen prosenttipisteiden avulla

Viimeaikaisissa tutkimuksissa on havaittu, että teoreettiset jakaumat (esim. Weibull- ja beta-jakauma) ovat liian jäykkiä kuvaamaan metsikön läpimittajakaumaa. Siksi on etsitty eri menetelmiä joissa päästään eroon rajoittavista jakaumaoletuksista. Lupaavimpia menetelmiä on ns. prosenttiosuusmenetelmä (Borders ym. 1987), jossa jakauma määritellään jakauman prosenttipisteiden avulla ja jatkuva kertymäfunktio muodostetaan interpoloimalla.

Olkoon metsikön läpimittajakauma F_Y. Tätä approksimoidaan paloittain lineaarisella funktiolla, joka määritellään vektorien $\mathbf{d}=(d_1,d_2,\ldots,d_k)'$ ja $\mathbf{p}=(p_1,p_2,\ldots,p_k)'$ avulla. Vektorin **d** elementit ovat läpimittoja jotka vastaavat jotain kiinteitä metsikön läpimittajakauman kertymäfunktion F_Y, arvoja. Nämä arvot ovat vektorissa **p**. Edelleen p_1=0, p_k=1 ja vektorit **p** ja **d** on järjestetty kasvavasti. Tällöin läpimitta d_i on jakauman F_Y 100* p_i. prosenttipiste. Oletetaan vielä että läpimittajakauman kertymäfunktio peräkkäisten prosenttipisteiden välillä on lineaarinen. Tämä johtaa paloittain määritettyyn tasaiseen jakaumaan. Sen kertymäfunktio määritellään

$$F_Y(y) \approx F_{\mathbf{p}}(y|\mathbf{d}) = \begin{cases} 0 & y < d_1 \\ a_i + b_i y & d_i \le y < d_{i+1}, i = 1,...,k-1 \\ 1 & y \ge d_k \end{cases}, \tag{6}$$

jossa

$$b_i = \frac{p_{i+1} - p_i}{d_{i+1} - d_i}$$

ja

$$a_i = p_i - b_i d_i.$$

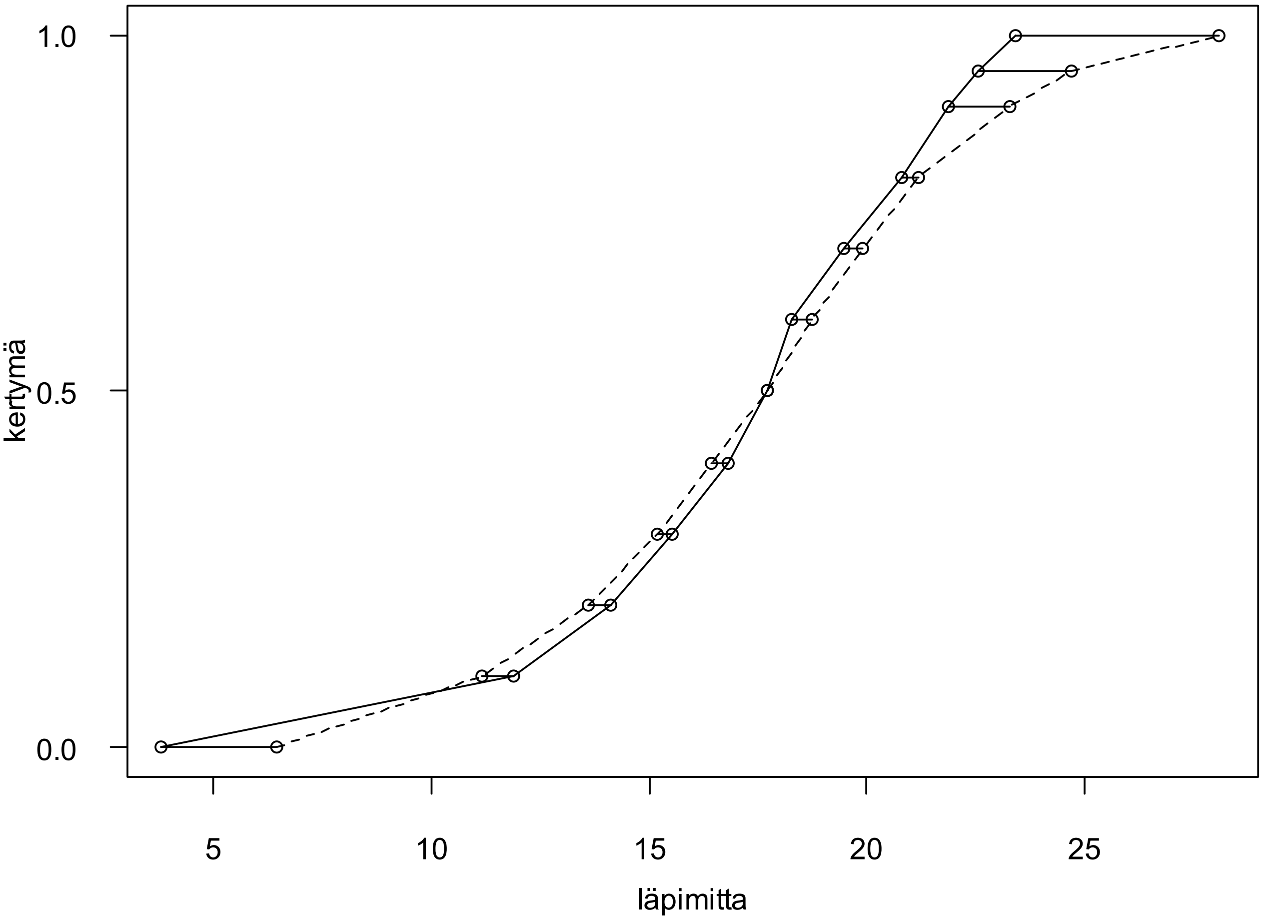
Approksimaation tarkkuus on sitä suurempi, mitä enemmän prosenttipisteitä käytetään. Helposti nähdään, että funktio $F_{\mathbf{p}}(y|\mathbf{d})$ täyttää yleiset todennäköisyysjakauman kertymäfunktiolle asetetut vaatimukset (Casella ja Berger 2002, s.31). Merkintä $F_{\mathbf{p}}(y|\mathbf{d})$ korostaa sitä, että $\mathbf{p}$ on ennalta määritetty kiinteä vektori, joka määrittelee käytettävän jakaumaperheen. Varsinaiset metsikön läpimittajakauman parametrit ovat vektorissa $\mathbf{d}$ joka siis prosenttiosuusjakauman tapauksessa korvaa merkinnän $\boldsymbol{\theta}$. Parametreilla on selkeä tulkinta: d_i on läpimittajakauman $100p_i$. prosenttipiste. Jakauman tiheysfunktio on

$$f_{\mathbf{p}}(y|\mathbf{d}) = \begin{cases} 0 & y < d_1 \\ b_i & d_i \le y < d_{i+1}, i = 1,...,k-1 \\ 0 & y \ge d_k \end{cases} \tag{7}$$

Prosenttiosuusjakauman tapauksessa malli (1) voidaan kirjoittaa muodossa

$$\mathbf{d} = E[\mathbf{d}|\mathbf{x}] + \mathbf{e}, \tag{8}$$

jossa ensimmäinen termi sisältää prosenttipisteiden ehdolliset odotusarvot kun metsikön metsikkötunnusten vektori $\mathbf{x}$ tunnetaan ja jälkimmäinen termi prosenttipisteiden metsikkövaikutukset. Kuvassa 2 on esitetty esimerkki prosenttiosuusjakauman avulla kuvatusta läpimittajakaumasta. Prosenttiosuusjakauman tapauksessa mitattu mediaanipuun läpimitta (DGM) on suoraan 50. prosenttipiste ja regressiomallit tarvitaan muille käytetyille prosenttipisteille.



Kuva 2. Prosenttipisteiden avulla kuvattu läpimittajakauman kertymäfunktio. Katkoviiva kuvaa jakauman joka perustuu prosenttipisteiden ehdollisiin odotusarvoihin, $F_{\mathbf{p}}\left(y \middle| E\left[\mathbf{d} \middle| \mathbf{x}\right]\right)$ ja yhtenäinen viiva kuvaa todellisen jakauman $F_{\mathbf{p}}\left(y \middle| E\left[\mathbf{d} \middle| \mathbf{x}\right] + \mathbf{e}\right)$. Vaakaviivat kuvaavat metsikkövaikutuksia $\mathbf{e} = \mathbf{d} - E\left[\mathbf{d} \middle| \mathbf{x}\right]$.

3. LÄPIMITTAJAKAUMAN ENNUSTAMINEN MITATTUJEN OTOSJÄRJESTYSTUNNUSLUKUJEN AVULLA

3.1. Prosenttipisteiden mittaaminen

Järjestystunnuslukupuu on otokseen kuuluva puu, josta tunnetaan läpimitta sekä sen järjestysluku otoksessa. Jos oletetaan, että metsikön puiden tilajärjestys on satunnainen, saadaan relaskoopin avulla riippumaton ja samoin jakautunut otos metsikön pohjapinta-alalla painotetusta läpimittajakaumasta. Tällöin järjestystunnuslukupuu voidaan mitata esimerkiksi valitsemalla ensimmäinen relaskooppikoealaan kuuluva puu järjestystunnuslukupuuksi ja mittaamalla sen läpimitta. Tämän jälkeen samalla kun luetaan relaskooppikoealaan kuuluvia puita, kustakin puusta arvioidaan, onko se pienempi vai suurempi kuin valittu järjestystunnuslukupuu. Jos varmaa arviota ei voida tehdä silmävaraisesti, määritetään suuruusjärjestys mittaamalla läpimitta.

Oletetaan, että tunnemme riippumattomasta metsikön puiden läpimittojen otoksesta järjestystunnusluvun. Järjestystunnusluku on satunnaismuuttuja, josta käytetään merkintää $Y_{r:n.}$ Olkoon $Y_1, \ldots, Y_n$ riippumaton otos läpimittoja metsikön läpimittajakaumasta F_Y. Otoksen *r.* suurinta havaintoa kutsutaan otoksen *r.* järjestystunnusluvuksi ja siitä käytetään merkintää $Y_{r:n,}$. Tunnemme siis otoskoon *n* sekä tiedämme, että tämän otoksen *r.* suurimman puun läpimitta on $Y_{r:n}$. Oletetaan, että metsikön läpimittajakauma $F_Y(y)$ tunnetaan, vaikka todellisuudessa sitä ei tunneta. Tähän oletukseen palataan myöhemmin. Tällöin $Y_{r:n}$:n tiheysjakauma on (Casella ja Berger 2002: 229, Reiss 1989: 21)

$$f_{r:n}(y) = \beta_0(n,r) f_Y(y) \left[F_Y(y)\right]^{r-1} \left[1 - F_Y(y)\right]^{n-r},$$

jossa

$$\beta_0(n,r) = \frac{n!}{(r-1)!(n-r)!}$$

ja $f_Y(y)$ on satunnaismuuttujan Y tiheysfunktio.

Sijoittamalla kaavaan prosenttiosuusjakauman tiheys ja kertymäfunktiot (6) ja (7) saadaan prosenttiosuusjakaumasta poimitun otoksen *r.* suurimman puun läpimitan jakaumaksi

$$f_{r:n}(y) = \begin{cases} \beta_0(n,r) b_i \left[a_i + b_i y\right]^{r-1} \left[1 - a_i - b_i y\right]^{n-r} & , i = 1, \ldots, k-1 \\ 0 & , muulloin \end{cases}. \tag{9}$$

Jakauman odotusarvo saadaan kaavalla

$$E(Y_{r:n}) = \int_{d_1}^{d_k} y f_{r:n}(y)\, dy = \sum_{i=1}^{k-1} \int_{d_i}^{d_{i+1}} y f_{r:n}{}^{i}(y) dy\ , \tag{10}$$

jossa $f_{r:n}{}^{i}$ on tiheysfunktio (9) välillä $[d_i,d_{i+1})$. $E(Y_{r:n})$ on metsiköstä m poimitun n:n kokoisen otoksen r. suurimman puun läpimitan odotusarvo. Seuraavaksi etsitään arvo p^*, jota vastaava metsikön läpimittajakauman arvo tuottaa odotusarvon (9)

$$E(Y_{r:n}) = F_Y^{-1}(p^*)$$

ratkaisemalla yllä olevasta yhtälöstä p^* saadaan

$$p^* = F_Y\left[E(Y_{r:n})\right]. \tag{11}$$

Tämä tarkoittaa, että havaittu järjestystunnuslukupuun läpimitta $Y_{r:n}$ on p^*. prosenttipisteen harhaton estimaattori.

Koska järjestystunnusluku on mitattu otoksesta, sisältyy siihen otantavirhettä, jonka varianssi on jakauman (9) varianssi. Se saadaan laskemalla ensin

$$E\left(Y_{r:n}{}^2\right) = \int_{d_1}^{d_k} y^2 f_{r:n}(y)\, dy = \sum_{i=1}^{k-1} \int_{d_i}^{d_{i+1}} y^2 f_{r:n}{}^{i}(y) dy$$

ja soveltamalla sitten kaavaa

$$\operatorname{var}(Y_{r:n}) = E\left(Y_{r:n}{}^2\right) - \left[E(Y_{r:n})\right]^2. \tag{12}$$

3.2. Prosenttiosuusjakauman parametrien ennustaminen

Prosenttiosuusjakauman tapauksessa läpimittajakauman parametrit ovat prosenttipisteitä. Jos tunnetaan vain metsikkötunnuksia, jotka vaikuttavat prosenttipisteisiin mallin kiinteän osan kautta, ennustamisessa voidaan soveltaa suoraan kaavaa (2). Edellisessä alaluvussa totesimme, että mitattu otoksen järjestystunnusluku voidaan tulkita prosenttipisteen mittaukseksi. Siten prosenttipisteiden avulla kuvatun läpimittajakauman parametreja voidaan mitata ja luvussa (2.3) esitettyä parasta lineaarista ennustinta voidaan soveltaa parametrien ennustamisessa.

Kaava (3) voidaan nyt kirjoittaa muodossa

$$\begin{bmatrix} \mathbf{d}|\mathbf{x} \\ \tilde{d}^*|\mathbf{x} \end{bmatrix} \sim \left(\begin{bmatrix} E(\mathbf{d}|\mathbf{x}) \\ E(d^*|\mathbf{x}) \end{bmatrix}, \begin{bmatrix} \mathbf{D} & \mathbf{c} \\ \mathbf{c}' & \sigma_e^2 + \sigma_\varepsilon^2 \end{bmatrix} \right),$$

jossa $E(\mathbf{d}|\mathbf{x})$ on mallin (8) avulla saatu ennuste, $\mathbf{D}$ on mallin (8) jäännösten (metsikkövaikutusten) varianssi-kovarianssimatriisi ja $\tilde{d}^*|\mathbf{x}$ havaittu järjestystunnuslukupuun läpimitta. Oletetaan, että otosjärjestystunnusluku on mitattu satunnaisotoksesta, mittaus on virheetön ja järjestysluku on määritetty oikein. Tällöin mittausvirhe ε koostuu kokonaan otantavirheestä, ja sen varianssi σ_ε^2 on suoraan otosjärjestystunnusluvun jakauman varianssi, joka saadaan kaavalla (12). Jos arvo p^* (kaava 11) sattuu olemaan tarkalleen joku vektorin $\mathbf{p}$ arvo, tunnetaan myös $\mathbf{c}$, σ_e^2 ja $E(d^*|\mathbf{x})$: $\mathbf{c}$ on matriisin $\mathbf{D}$ arvoa p^* vastaava sarake, σ_e^2 matriisin $\mathbf{D}$ diagonaalin vastaava alkio ja $E(d^*|\mathbf{x})$ vektorin $E(\mathbf{d}|\mathbf{x})$ vastaava alkio. Nyt luvussa kaavoja (4) ja (5) voidaan soveltaa suoraan vektorin $\mathbf{d}|\mathbf{x}$ ennustamiseen.

Koska arvo p^* on mikä tahansa luku väliltä (0,1), ei se koskaan saa tarkalleen mitään vektorin $\mathbf{p}$ arvoista, eivätkä $\mathbf{c}$, σ_e^2 ja $E(d^*|\mathbf{x})$ ole tunnettuja. Niitä voidaan kuitenkin approksimoida interpoloimalla. Interpoloinneissa käytetään lineaarista interpolointia. Mitatun prosenttipisteen ehdollinen odotusarvo $E(d^*|\mathbf{x})$ saadaan interpoloimalla vektoria $E(d^*|\mathbf{x})$, arvoa p^* vastaava varianssi interpoloimalla matriisin $\mathbf{D}$ diagonaalia ja kovarianssivektori $\mathbf{c}$ interpoloimalla matriisin $\mathbf{D}$ kovariansseja. Näiden approksimointien jälkeen kaavoja (4) ja (5) voidaan soveltaa vektorin $\mathbf{d}|\mathbf{x}$ ennustamiseen. Vektorin $\mathbf{d}|\mathbf{x}$ paras lineaarinen ennustin saadaan kaavalla

$$\hat{\mathbf{d}} = E\left(\mathbf{d}|\mathbf{x}\right) + \frac{\left[\tilde{d}* - E\left(d*|\mathbf{x}\right)\right]}{\left(\sigma_e^2 + \sigma_\varepsilon^2\right)}\mathbf{c}, \tag{13}$$

ja sen ennustevirheen varianssi on

$$\mathrm{var}\left(\hat{\mathbf{d}} - \mathbf{d}\right) = \mathbf{D} - \frac{1}{\sigma_e^2 + \sigma_\varepsilon^2}\mathbf{cc}'. \tag{14}$$

Toinen ongelma on, että kuten luvussa 3.1 todettiin, arvon $p*$ laskemiseksi olisi tunnettava metsikön todellinen jakauma. Tätä ei kuitenkaan tunneta, vaan juuri sitä ollaan ennustamassa. Tämä ongelma ratkaistaan etsimällä metsikkövaikutusten ennusteet iteratiivisesti ja käyttämällä kussakin iteroinnissa parasta käytössä olevaa arvausta metsikön jakaumaksi. Ensimmäisessä iteroinnissa paras arvaus on mitattuun metsikkötunnusten vektoriin perustuva jakauma $F_{\mathbf{p}}\left(y\middle|E\left[\mathbf{d}|\mathbf{x}\right]\right)$. Iterointi toteutetaan seuraavalla tavalla.

1. Asetetaan j=1 ja $\hat{\mathbf{d}}^{(j)} = E\left[\mathbf{d}|\mathbf{x}\right]$
2. Lasketaan $p*$ kaavalla (11) olettaen että metsikön läpimittajakauma on $F_{\mathbf{p}}\left(y\middle|\hat{\mathbf{d}}^{(j)}\right)$. Lasketaan $\hat{\mathbf{d}}^{(j+1)}$ kaavalla (13).
3. Jos $\left(\hat{\mathbf{d}}^{(j+1)} - \hat{\mathbf{d}}^{(j)}\right)'\left(\hat{\mathbf{d}}^{(j+1)} - \hat{\mathbf{d}}^{(j)}\right)$ poikkeaa merkittävästi nollasta, asetetaan j=j+1 ja palataan kohtaan 2. Muutoin asetetaan $\hat{\mathbf{d}} = \hat{\mathbf{d}}^{(j+1)}$.

3.3. Yleistys usean otosjärjestystunnusluvun tapaukseen

Edellä esitetyt tulokset voidaan yleistää tapaukseen, jossa metsiköstä on mitattu useita otosjärjestystunnuslukuja, merkataan mittausten määrää q:lla. Saman metsikön mittaukset kirjoitetaan peräkkäin vektoriin $\mathbf{d}*_{q\times 1}$ ja vektori **c** korvataan matriisilla $\mathbf{C}_{k\times q}$, jonka sarakkeina ovat kunkin mittauksen kovarianssivektorit **c**. Skalaari σ_e^2 korvataan matriisilla $\mathbf{D}*_{q\times q}$, joka saadaan matriisista **D** interpoloimalla lineaarisesti niin, että ensin interpoloidaan diagonaalille varianssit ja sen jälkeen diagonaalin ulkopuolelle kovarianssit. Skalaari σ_ε^2 korvataan matriisilla $\mathbf{R}_{q\times q}$, jonka diagonaalille tulevat yksittäisten mittausten otantavirheiden varianssit σ_ε^2. Jos otosjärjestystunnusluvut mitattiin kukin eri koealalta, niiden otantavirheiden kovarianssit ovat nollia ja **R** on diagonaalimatriisi. Nyt havaittujen ja ennustettujen prosenttipisteiden vektorien ominaisuudet voidaan kirjoittaa muodossa

$$\begin{bmatrix} \mathbf{d}|\mathbf{x} \\ \tilde{\mathbf{d}}^*|\mathbf{x} \end{bmatrix} \sim \left(\begin{bmatrix} E(\mathbf{d}|\mathbf{x}) \\ E(\mathbf{d}^*|\mathbf{x}) \end{bmatrix}, \begin{bmatrix} \mathbf{D} & \mathbf{C} \\ \mathbf{C}' & \mathbf{D}^*+\mathbf{R} \end{bmatrix} \right).$$

Matriisimuotoon yleistetty parhaan lineaarisen ennustimen kaava voidaan kirjoittaa muodossa

$$\hat{\mathbf{d}} = E(\mathbf{d}|\mathbf{x}) + \mathbf{C}(\mathbf{D}^*+\mathbf{R})^{-1}\left[\tilde{\mathbf{d}}^* - E(\mathbf{d}^*|\mathbf{x})\right] \tag{15}$$

ja ennustevirheen varianssi muodossa

$$\mathrm{var}(\hat{\mathbf{d}} - \mathbf{d}) = \mathbf{D} - \mathbf{C}(\mathbf{D}^*+\mathbf{R})^{-1}\mathbf{C}'. \tag{16}$$

Prosenttipisteiden ennustamisessa sovelletaan samaa iterointialgoritmia kuin yhden otosjärjestystunnusluvun tapauksessa, mutta nyt vaiheessa 2 metsikkövaikutukset ennustetaan kaavalla (15).

Jos samalta koealalta on mitattu useita järjestystunnuslukupuita, ovat näiden mittausten otantavirheet korreloituneita. Tässä tapauksessa matriisi **R** on lohkodiagonaalinen, jossa yhden koealan mittaukset muodostavat kukin yhden lohkon. Kovarianssien laskemiseksi tarvitaan kahden samasta otoksesta mitatun järjestystunnusluvun yhteistiheysjakauma. Yhteistiheysjakauma on yleisessä muodossa (Casella ja Berger 2002: 230, Reiss 1989: 30-31)

$$f_{r_1:n,r_2:n}(y_1,y_2) = \beta_1(r_1,r_2,n) f_Y(y_1) f_Y(y_2)\left[F_Y(y_1)\right]^{r_1-1}\left[F_Y(y_2)-F_Y(y_1)\right]^{r_2-r_1-1}\left[1-F_Y(y_2)\right]^{n-r_2},$$

jos $y_1<y_2$ ja 0 muulloin. Kaavassa $r_1<r_2$ ja

$$\beta_1(r_1,r_2,n) = \frac{n!}{(n-r_2)!(r_2-r_1-1)!(r_1-1)!}.$$

Kun kirjoitetaan kaavaan prosenttipisteiden avulla kuvatun läpimittajakauman kertymä- ja tiheysfunktiot (6) ja (7), saadaan kahden samasta otoksesta mitatun järjestystunnuslukupuun läpimitan yhteistiheysfunktioksi

$$f_{r:n}\left(y_1,y_2\right)=\begin{cases}\beta_1\left(n,r_1,r_2\right)b_ib_j\left(a_i+b_iy_1\right)^{r_1-1}\left(a_j+b_jy_2-a_i-b_iy_2\right)^{r_2-r_1-1}\left(1-a_j-b_jy_2\right)^{n-r_2} & ,i,j=1,...,k-1;i\le j,y_1\le y_2\\ 0 & ,muulloin\end{cases}.$$

Otosjärjestystunnuslukupuiden läpimittojen tulon odotusarvo saadaan kaavalla

$$E\left(Y_{r_1:n}Y_{r_2:n}\right)=\sum_{j=1}^{k-1}\sum_{i=1}^{k-1}\int_{d_i}^{d_{i+1}}\int_{d_j}^{d_{j+1}} y_1y_2f_{r_1:n,r_2:n}{}^{i,j}\left(y_1,y_2\right)dy_2dy_1$$

ja otantavirheiden välinen kovarianssi edelleen kaavalla

$$\operatorname{cov}\left(Y_{r_1:n,}Y_{r_2:n}\right)=E\left(Y_{r_1:n}Y_{r_2:n}\right)-E\left(Y_{r_1:n}\right)E\left(Y_{r_2:n}\right), \qquad (17)$$

jossa jälkimmäisen termin laskennassa käytetään kaavaa (11). Kaavalla (17) lasketut kahden järjestystunnusluvun kovarianssit kirjoitetaan matriisiin **R** niille kuuluviin paikkoihin. Tämän jälkeen ennustamisessa sovelletaan kaavaa (15).

4. Menetelmän testaus

4.1. Prosenttipisteiden regressiomallit

Kangas ja Maltamo (2000) esittivät mallit metsikön pohjapinta-alalla painotetun läpimittajakauman prosenttipisteille. Selitettävät prosenttipisteet valittiin niin, että k=11 ja

$$\mathbf{p} = \begin{pmatrix} 0,00 & 0,10 & 0,20 & 0,30 & 0,40 & 0,60 & 0,70 & 0,80 & 0,90 & 0,95 & 1,00 \end{pmatrix}' .$$

50 prosentin pistettä ei mallitettu, koska mallin selittäjänä oleva pohjapinta-alamediaanipuun läpimitta on mitattu 50. prosenttipiste. Residuaalien homogenisoimiseksi mallitettiin prosenttipisteiden logaritmeja. Kankaan ja Maltamon (2000) mallit ovat lineaarisia ja muotoa

$$\ln(\mathbf{d}) = E\left[\ln(\mathbf{d}) \middle| \mathbf{x}\right] + \mathbf{e} = \mathbf{Bx} + \mathbf{e} . \tag{18}$$

jossa edellinen termi on parametrivektorin odotusarvo ehdolla että metsikkötunnusten vektori **x** tunnetaan ja jälkimmäinen on metsikkövaikutusten vektori, jolle pätee E(**e**)=**0** ja var(**e**)=$\mathbf{D}_{kxk}$. Matriisin **B** *i.* rivi sisältää *i.* prosenttipisteen mallin kertoimet.

Mallit estimoitiin samanaikaisesti SUR-menetelmällä. Matriisin **B** estimaatti on taulukossa 1 ja Matriisin **D** estimaatti taulukossa 2.

Taulukko 1. Logaritmisten prosenttipisteiden mallin kertoimien estimaatit (Kangas ja Maltamo 2000). Mallin selittäjät DGM=metsikön kuusten pohjapinta-alalla painotettu mediaaniläpimitta (=50. prosenttipiste), T=metsikön ikä, vuotta, G=metsikön kuusten pohjapinta-ala, m^2/ha, soil=maapohjan hyvyyttä kuvaava valemuuttuja, joka saa arvon 1 kun maapohja on rehevä ja arvon 0 kun maapohja on karu tai normaali.

		mallin kertoimet					
i	p_i	vakio	ln(DGM)	ln(T)	ln(T/G)	ln(G)	soil
1	0	-0.3561	0.8351	0	-0.1178	-0.1261	0
2	0.1	-0.2120	0.8830	0	-0.0736	0	0
3	0.2	-0.1667	0.9679	0	-0.0789	0	0
4	0.3	-0.3199	1.0528	0	-0.0379	0	0
5	0.4	-0.1315	1.0266	0	-0.0313	0	0
6	0.6	0.1766	0.9688	0	0	0	0
7	0.7	0.3237	0.8964	0.0348	0	0	0
8	0.8	0.4768	0.8381	0.0603	0	0	0
9	0.9	0.7771	0.7502	0.0792	0	0	-0.0392
10	0.95	0.9005	0.7016	0.1014	0	0	-0.0409
11	1	1.3823	0.6241	0.0832	0	0	-0.0682

Taulukko 2. Taulukon 1 mallien metsikkövaikutusten estimoitu varianssi-kovarianssimatriisi (Kangas A, henkilökohtainen tiedonanto)

i\i	1	2	3	4	5	6	7	8	9	10	11
1	0.1617	0.0501	0.0223	0.0107	0.0069	0.0002	-0.0027	-0.0055	-0.0066	-0.0067	-0.0098
2	0.0501	0.0746	0.0349	0.0156	0.0088	-0.0027	-0.0051	-0.0073	-0.0081	-0.0082	-0.0090
3	0.0223	0.0349	0.0293	0.0150	0.0094	-0.0026	-0.0039	-0.0059	-0.0066	-0.0070	-0.0077
4	0.0107	0.0156	0.0150	0.0142	0.0093	-0.0015	-0.0024	-0.0036	-0.0045	-0.0048	-0.0054
5	0.0069	0.0088	0.0094	0.0093	0.0098	-0.0009	-0.0013	-0.0025	-0.0033	-0.0035	-0.0037
6	0.0002	-0.0027	-0.0026	-0.0015	-0.0009	0.0032	0.0030	0.0030	0.0031	0.0031	0.0028
7	-0.0027	-0.0051	-0.0039	-0.0024	-0.0013	0.0030	0.0059	0.0058	0.0058	0.0060	0.0062
8	-0.0055	-0.0073	-0.0059	-0.0036	-0.0025	0.0030	0.0058	0.0087	0.0083	0.0084	0.0086
9	-0.0066	-0.0081	-0.0066	-0.0045	-0.0033	0.0031	0.0058	0.0083	0.0110	0.0110	0.0109
10	-0.0067	-0.0082	-0.0070	-0.0048	-0.0035	0.0031	0.0060	0.0084	0.0110	0.0135	0.0138
11	-0.0098	-0.0090	-0.0077	-0.0054	-0.0037	0.0028	0.0062	0.0086	0.0109	0.0138	0.0250

4.2. Numeerinen esimerkki

Tässä luvussa esitetään numeerinen esimerkki läpimittajakauman ennustamisesta, kun virheettömien metsikkötunnusten lisäksi on mitattu yksi järjestystunnuslukupuu. Kuusikon DGM on 20 cm, pohjapinta-ala 22 m^2/ha, ikä 64 vuotta, metsikön kasvupaikan viljavuus normaali. Siten metsikkötunnusten vektori on

$$\mathbf{x} = \begin{pmatrix} 1 & \ln 20 & \ln 64 & \ln \frac{64}{20} & \ln 22 & 0 \end{pmatrix}'.$$

ja kiinteän osan logaritmisten prosenttipisteiden ennusteiksi saadaan

$$E(\mathbf{d}|\mathbf{x}) = \mathbf{Bx} = \mathbf{d}^{(1)} = (1.63 \quad 2.35 \quad 2.65 \quad 2.79 \quad 2.91 \quad 3.08 \quad 3.15 \quad 3.24 \quad 3.31 \quad 3.38 \quad 3.53)'.$$

Näiden ennusteiden perusteella interpoloitu läpimittajakauma on piirretty katkoviivalla kuvaan 3. Lisäksi oletetaan että 13 puun relaskooppiotoksen 3. pienimmän puun läpimitan on mitattu olevan 8.5 cm. Tämä tarkoittaa, että on havaittu, että

$$Y_{3:13} = \ln(8.5) = 2.140.$$

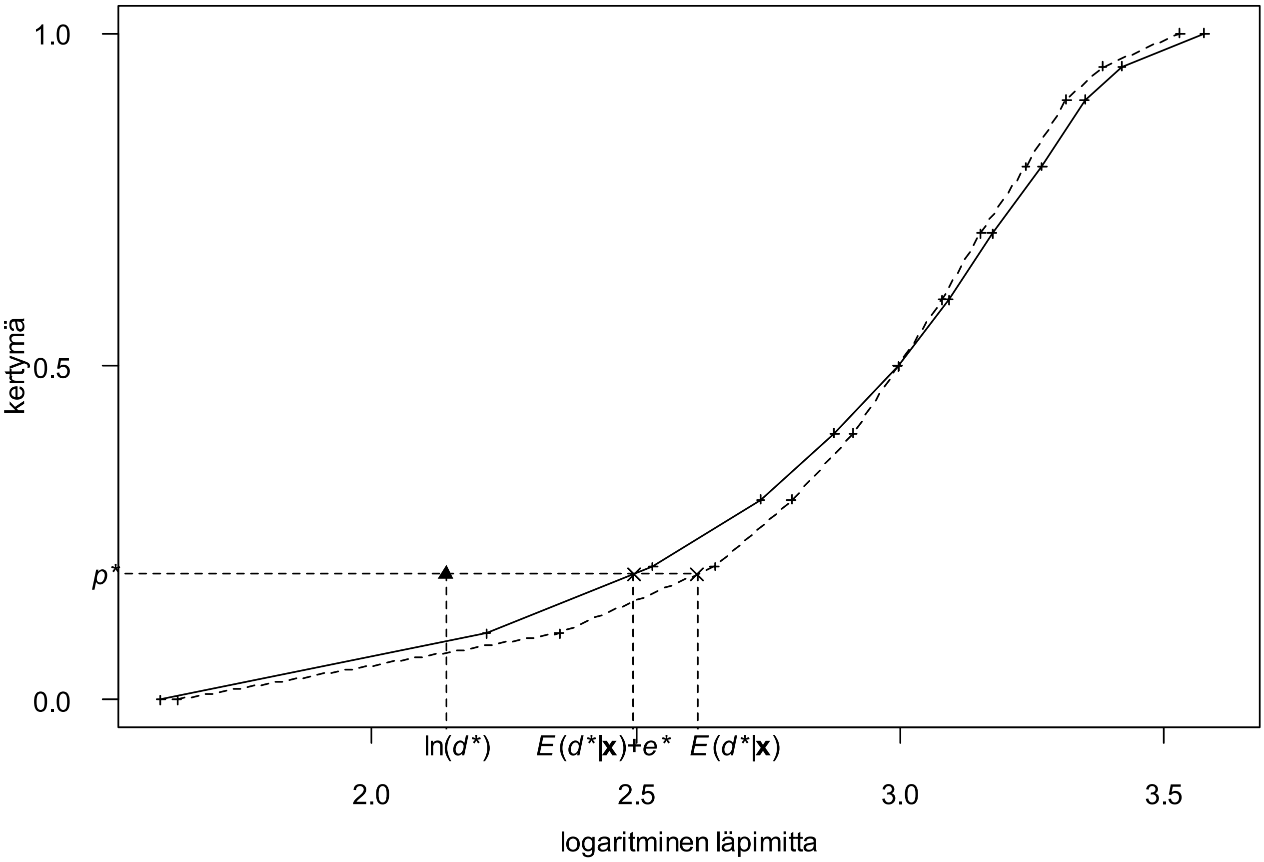
Iteroinnin ensimmäisessä vaiheessa oletetaan, että metsikkövaikutusten vektori $\mathbf{e}$=0, jolloin metsikön jakauman oletetaan olevan $F_{\mathbf{p}}(y|\mathbf{d}^{(1)})$. Tällöin kaavalla (10) saadaan mitatun otosjärjestystunnusluvun odotusarvoksi (Taulukko 3)

$$E(Y_{3:13}) = 2.595.$$

Sijoittamalla tämä kaavaan (11) saadaan arvo

$$p^* = 0.182,$$

mikä tarkoittaa että mitattu logaritminen läpimitta on pohjapinta-alalla painotetun läpimittajakauman 18.2 prosentin pisteen mittaus.



Kuva 3. Mallin kiinteän osan avulla ennustettu jakauma (katkoviiva) ja jakaumaennuste kun metsikkövaikutukset on ennustettu mitatun järjestystunnuslukupuun (▲) avulla (yhtenäinen viiva) numeerisen esimerkin metsikössä.

Seuraavaksi ennustetaan prosenttiosuusjakauman parametrien metsikkövaikutukset. Metsikkövaikutusten laskemiseksi tarvitaan mitatun prosenttipisteen otantavirheen ja metsikkövaikutuksen varianssit σ_ε^2 ja σ_e^2 sekä mitatun prosenttipisteen ja jakauman parametrien metsikkövaikutusten väliset kovarianssit **c**. Otantavirheen varianssiksi saadaan kaavalla (12)

$$\sigma_\varepsilon^2=0.0577.$$

Mitatun prosenttipisteen metsikkövaikutuksen varianssi lasketaan interpoloimalla matriisin **D** diagonaali arvolle 0.182. Lineaarisella interpoloinnilla saadaan

$$\sigma_e^2= 0.0746+(0.0293-0.0746)/0.10\times(0.182-0.1)=0.0376.$$

Kovarianssit **c** lasketaan myös interpoloimalla lineaarisesti matriisia **D**

$$\mathbf{c} = \begin{pmatrix} 0.0501+\left(0.0223\text{-}0.0501\right)/0.10\times\left(0.182\text{-}0.1\right) \\ \vdots \\ \text{-}0.00903+\left(\text{-}0.00774+0.00903\right)/0.10\times\left(0.182\text{-}0.1\right) \end{pmatrix} = \begin{pmatrix} 0.0274 \\ \vdots \\ -0.00798 \end{pmatrix}$$

Nyt prosenttipisteet voidaan ennustaa kaavalla (13). Prosenttipisteiden ennusteiksi ensimmäisen iteroinnin jälkeen saadaan

$$\hat{\mathbf{d}}^{(2)} = \begin{pmatrix} 1.63 \\ \vdots \\ 3.53 \end{pmatrix} + \begin{pmatrix} 0.0274 \\ \vdots \\ -0.00798 \end{pmatrix} \frac{2.140 - 2.595}{0.0376 + 0.0577} = \begin{pmatrix} \text{-}0.131 \\ \vdots \\ 0.0381 \end{pmatrix}.$$

Seuraavaksi laskenta toistetaan, mutta oletetaan että metsikön läpimittajakauma on $F_{\mathbf{p}}\left(y \middle| \mathbf{d}^{(2)}\right)$ ja lasketaan uudet arvot järjestystunnusluvun odotusarvolle ja varianssille. Iterointi lopetetaan kun iteroinnin kolmannen vaiheen kriteerimuuttujan arvo on pienempi kuin 10^{-8}. Taulukosta 3 nähdään kuinka otosjärjestystunnusluvun odotusarvo varianssi ja arvo $p*$ muuttuvat iteroinnin kuluessa ja kuvaan 3 on piirretty yhtenäisellä viivalla ennustettu jakauma iteroinnin jälkeen. Havaittu läpimitta on piirretty kuvaan 3 pisteeseen (2.140,0.189).

Taulukko 3. Numeerisen esimerkin iteroinnin kulku.

toisto	$E(Y_{3:13})$	$p*$	σ_ε^2
1	2.5948	0.1817	0.05771
2	2.6157	0.1888	0.08136
3	2.6143	0.1883	0.07566
4	2.6150	0.1886	0.07673
5	2.6149	0.1885	0.07651
6	2.6149	0.1886	0.07655

Ennustevirheen varianssi-kovarianssimatriisi on taulukossa 4. Jos ennustamisessa käytettäisiin vain metsikkötunnuksia, olisi ennustevirheen varianssi-kovarianssimatriisi matriisi **D** (Taulukko 1). Kummassakaan tapauksessa ei huomioida mallin (18) parametrien estimointivirheiden vaikutusta ennustevirheisiin. Taulukkojen diagonaalialkioiden vertailu osoittaa, että järjestystunnuslukupuun käyttö ennustamisessa pienentää ennustevirheiden varriansseja.

Taulukko 4. Numeerisen esimerkin ennustevirheen varianssi-kovarianssimatriisi.

$i \backslash i$	1	2	3	4	5	6	7	8	9	10	11
1	0.1558	0.0411	0.0154	0.0073	0.0048	0.0008	-0.0018	-0.0041	-0.0050	-0.0051	-0.0080
2	0.0411	0.0606	0.0243	0.0103	0.0055	-0.0017	-0.0037	-0.0051	-0.0057	-0.0056	-0.0062
3	0.0154	0.0243	0.0212	0.0109	0.0068	-0.0019	-0.0028	-0.0043	-0.0047	-0.0051	-0.0056
4	0.0073	0.0103	0.0109	0.0121	0.0081	-0.0012	-0.0019	-0.0027	-0.0036	-0.0038	-0.0043
5	0.0048	0.0055	0.0068	0.0081	0.0090	-0.0007	-0.0010	-0.0020	-0.0027	-0.0029	-0.0030
6	0.0008	-0.0017	-0.0019	-0.0012	-0.0007	0.0031	0.0029	0.0028	0.0030	0.0029	0.0026
7	-0.0018	-0.0037	-0.0028	-0.0019	-0.0010	0.0029	0.0058	0.0056	0.0055	0.0057	0.0060
8	-0.0041	-0.0051	-0.0043	-0.0027	-0.0020	0.0028	0.0056	0.0083	0.0080	0.0080	0.0082
9	-0.0050	-0.0057	-0.0047	-0.0036	-0.0027	0.0030	0.0055	0.0080	0.0106	0.0105	0.0104
10	-0.0051	-0.0056	-0.0051	-0.0038	-0.0029	0.0029	0.0057	0.0080	0.0105	0.0131	0.0132
11	-0.0080	-0.0062	-0.0056	-0.0043	-0.0030	0.0026	0.0060	0.0082	0.0104	0.0132	0.0244

4.3. Otosjärjestystunnuslukupuiden vaikutus metsikön tilavuusennusteen tarkkuuteen

Menetelmää testattiin 43 metsikön aineistossa, jossa metsikön koko vaihteli 600 ja 3000 m^2 välillä (Pukkala ym.1994). Aineiston metsiköt olivat mänty-kuusisekametsiä, ja testissä ennustettiin metsikön kuusiositteen läpimittajakaumaa. Metsikön kaikkien puiden läpimitat tunnettiin. Lisäksi tunnettiin metsikön ikä, metsätyyppi sekä pohjapinta-alamediaanipuun läpimitta (DGM) ja metsikön pohjapinta-ala, jotka laskettiin mitatuista metsikön puista. Metsiköihin simuloitiin relaskooppikoealoja seuraavasti. Ensin laskettiin kuusen *i* todennäköisyys kuulua metsikköön perustettuun relaskooppikoealaan kaavalla

$$\pi_i = \frac{1}{A}\pi\left(\frac{d_i}{2}\right)^2,$$

jossa d_i on puun *i* läpimitta ja *A* metsikön pinta-ala. Sen jälkeen arvottiin kaikille metsikön kuusille satunnaisluku tasaisesta jakaumasta väliltä [0,1]. Ne puut, joiden arvottu satunnaisluku oli pienempi kuin otantatodennäköisyys π_i, otettiin mukaan koealaan. Näin saatiin riippumattomia otoksia metsikön läpimittajakaumasta. Kuhunkin metsikköön simuloitiin 6 koealaa joista kustakin poimittiin satunnaisesti kaksi puuta järjetystunnuslukupuuhavainnoiksi.

Metsikön läpimittajakauman ennustamisessa käytettiin tunnettuja metsikkötunnuksia sekä 0, 1, 2, 3, 4, 5 tai 6 koealan järjestystunnuslukupuita. Koska puita poimittiin kaksi kultakin koealalta, oli järjestystunnuslukupuiden määrä vastaavasti eri testiajoissa 0, 2, 4, 6, 8, 10 tai 12. Otantavirheen vaikutuksen eliminoimiseksi laskenta toistettiin 100 kertaa kussakin metsikössä kullakin koepuumäärällä.

Iteroinnin konvergointikriteerinä oli, että metsikkövaikutusten arvojen muutosten itseisarvojen summa peräkkäisten iterointien välillä oli alle 10^{-6}. Jos iterointi ei konvergoinut 50 toiston jälkeen, laskenta keskeytettiin. Koska käytetty menetelmä ei takaa ennusteiden monotonisuutta, voi jakaumaennusteesta tulla jonkin iterointiaskeleen jälkeen ei-monotoninen. Myös tällaisessa tilanteessa laskenta jouduttiin keskeyttämään. Kun kustakin metsiköstä käytettiin vain kahta puuta, onnistui estimointi 98 prosentissa metsiköistä ja tarvittavien iterointikierrosten lukumäärä oli keskimäärin 8.28. Kun käytettyjen puiden määrä oli 12, onnistui estimointi enää 88 prosentissa metsiköistä ja iterointeja tarvittiin keskimäärin 12.0. Iteroinnin keskeytymisistä 93 prosenttia johtui siitä että jakaumaennuste tuli ei-monotoniseksi ja vain 7 prosenttia siitä, että iterointi ei konvergoinut 50 iteroinnin aikana. Epäonnistuneet iteroinnit ohitettiin laskennassa poimimalla metsiköstä uusi otos niin monta kertaa että ennustaminen onnistui.

Metsikön kuusten kokonaistilavuuden laskemiseksi logaritmisten prosenttipisteiden ennusteet palautettiin aritmeettiseen skaalaan eksponenttimuunnoksella. Ennen takaisinmuunnosta logaritmiseen prosenttipisteen ennusteeseen lisättiin ennustevirheen varianssin puolikas, joka saatiin kaavalla (16) lasketun matriisin diagonaalilta. Näin saadut ennusteet ovat harhattomia aritmeettisessa skaalassa, jos logaritmisten ennusteiden ennustevirheet ovat normaalisia (Flewelling ja Pienaar 1981).

Aritmeettiseen skaalaan muutetusta läpimittajakaumasta laskettiin metsikön kokonaistilavuuden ennuste kaavalla

$$\hat{V} = \sum_{i=1}^{K} \frac{\hat{g}_i}{\pi r_i^2},$$

jossa $\hat{g}_i$ on läpimittaluokan i ennustettu pohjapinta-ala ja r_i tämän läpimittaluokan keskikohtaa vastaavan puun säde. Lisäksi laskettiin läpimittajakauman virheindeksi kaavalla

$$e = \sum_{i=1}^{K} \left| \hat{g}_i - g_i \right|,$$

jossa g_i on läpimittaluokan i todellinen pohjapinta-ala. Tilavuuden laskennassa käytettiin 1 cm:n ja virheindeksin laskennassa 2 cm:n läpimittaluokkia. Virheindeksin luokkaleveys oli suurempi, koska aineiston metsiköt olivat suhteellisen pieniä ja todellisiin jakaumiin sisältyy satunnaisvaihtelua, joka vaikuttaa virheindeksin arvoon pienillä luokkaleveyksillä.

Läpimittajakaumaennusteen hyvyyden mittarina käytettiin virheindeksin keskiarvoa kullakin mittauslukumäärällä. Lisäksi laskettiin metsikön puuston tilavuuden suhteellinen keskineliövirhe

$$RMSE(\%) = 100\% \times \sqrt{\frac{\sum_{m=1}^{4300}\left(V_m - \hat{V}_m\right)^2}{4300}} \Big/ \bar{V} ,$$

ja harha

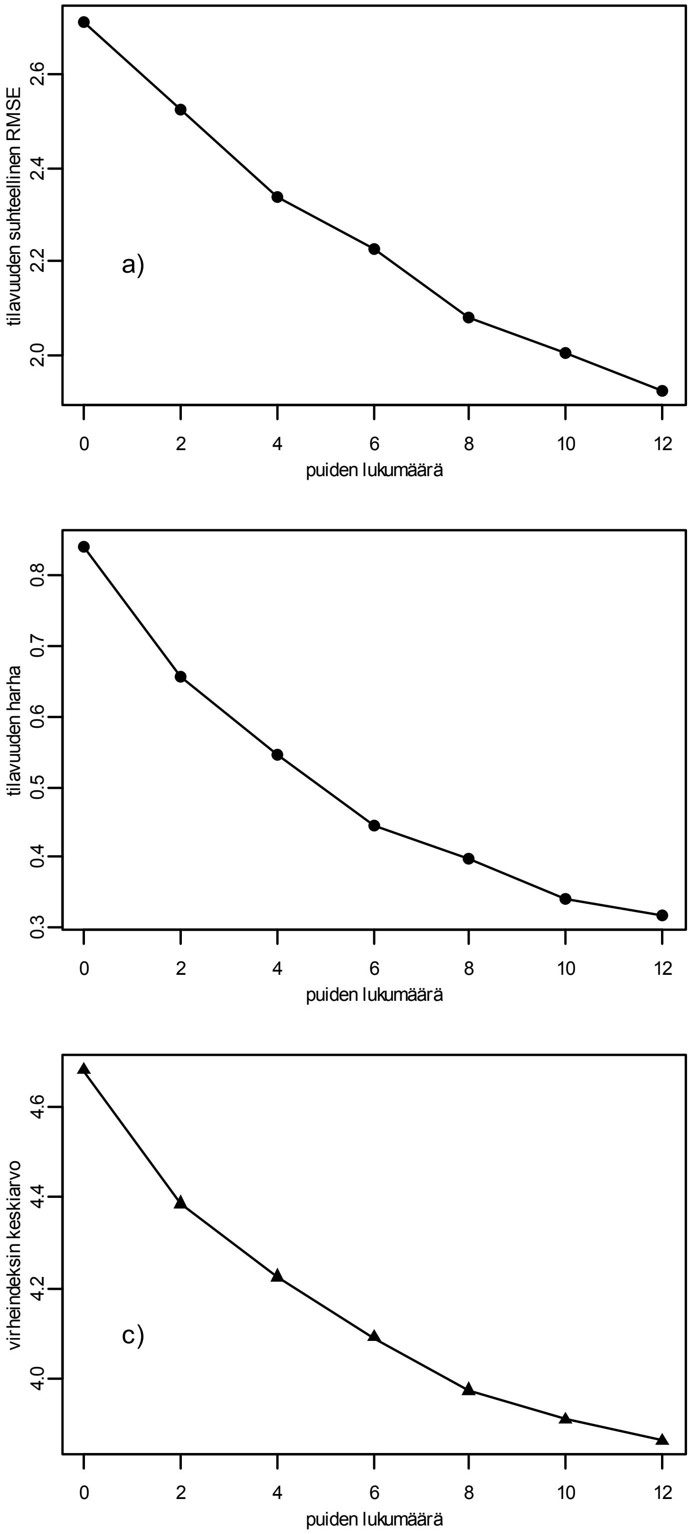
$$harha = \frac{\sum_{m=1}^{4300}\left(\hat{V}_m - V_m\right)}{4300} ,$$

jossa V_m on metsikön m todellinen tilavuus, $\hat{V}_m$ ennustetun läpimittajakauman avulla laskettu tilavuus ja $\bar{V}$ metsiköiden todellinen keskitilavuus. Sekä toden että ennustetun tilavuuden laskennassa käytettiin Laasasenahon (1982) tilavuusmalleja, joilla voidaan ennustaa yksittäisen puun tilavuus läpimitan avulla. Metsikön todellinen tilavuus laskettiin metsikön puiden tilavuusennusteiden summana.

Laskennat tehtiin R ympäristössä (www.r-project.org, Venables ja Ripley 2002). Järjestystunnuslukujen jakaumien integroinnissa käytettiin IMSL-aliohjelmakirjaston numeerisen integroinnin aliohjelmia, (IMSL 1997: 615-619), joka linkitettiin R-ympäristöön.

Testiajon tuloksista (Kuva 4) nähdään, että otosjärjestystunnuslukupuiden mittaamisella voidaan pienentää virheindeksin keskiarvoa, metsikön tilavuusennusteen virhettä ja tilavuusennusteen harhaa. Jo kahden puun mittaaminen pienensi virheindeksin arvon 4.68 m^2/ha:stä 4.38 m^2/ha:iin, tilavuuden keskineliövirheen 2.71%:sta 2.52%:iin ja tilavuuden harhan 0.84 m^3/ha:stä 0.66 m^3/ha:iin.

Koepuiden määrän lisääminen paransi ennusteita tasaisesti ja 12 puun tapauksessa virheindeksin arvo oli 4.02 m^2/ha, tilavuuden keskineliövirhe 1.92 % ja tilavuuden harha 0.32 m^3/ha.



Kuva 4. Otosjärjestystunnuslukupuiden määrän vaikutus tilavuuden suhteelliseen keskineliövirheeseen (a), tilavuuden negatiiviseen harhaan (b) ja virheindeksiin (c).

5. TULOSTEN TARKASTELUA

Tässä työssä esiteltiin menetelmä, jolla voidaan parantaa läpimittajakaumaennusteen tarkkuutta ennustamalla läpimittajakauman prosenttipisteiden metsikkövaikutukset otosjärjestystunnuslukupuiden avulla. Menetelmä pohjautuu lineaarisen ennustamisen teoriaan ja ennustettujen prosenttipisteiden lisäksi saadaan myös estimoitua ennustevirheiden varianssi-kovarianssimatriisi. Menetelmän soveltamista esiteltiin numeerisen esimerkin avulla. Menetelmää testattiin pienessä metsikköaineistossa, jossa se toimi lupaavasti. Menetelmän taustalla on joukko oletuksia, jotka ovat välttämättömiä laskennan suorittamisen kannalta. Seuraavassa tarkastellaan tehtyjen oletusten paikkansapitävyyttä.

Koko metsikköajattelun pohjana on oletus että metsäalue voidaan jakaa metsiköihin, joiden sisällä metsikön tunnukset eivät ole spatiaalisesti korreloituneita. Metsämaiseman tarkastelu esimerkiksi ilmakuvalta osoittaa, että oletus on perusteltu: metsäalueelta voidaan löytää selkeästi erilaisia metsiköitä, jotka poikkeavat huomattavasti ympäristöstään. Nämä erot voivat johtua maapohjan eroista (esim. suon ja kangasmetsän raja), mutta useimmiten ne johtuvat alueella tehdyistä hakkuista. Avohakatun alueen puusto on lähtötilanteessa hyvin samanlainen, varsinkin jos koko alueelle viljellään samaa puulajia ja luontaista taimiainesta ei ole. Metsikön sisällä kuitenkin esimerkiksi maapohjan vaihtelun ja mäkisyyden vuoksi puut kasvavat eri nopeudella ja metsikön sisällä kasvu on spatiaalisesti korreloitunut (esim. Pukkala ym. 1998). Samoin tilajärjestys ei ole useinkaan satunnainen (Tomppo 1986, Lin 2003). Tämän vuoksi metsiköstä mitatun koealan otos ei ole riippumaton ja todellisiin metsiköihin perustetuilta koealoilta mitatut järjestystunnusluvut eivät ole riippumattoman otoksen järjestystunnuslukuja. Siten niistä todellisuudessa saavutettu hyöty lienee pienempi kuin tämän tutkimuksen tulokset osoittavat. Mehtätalon ja Kankaan (2004) tutkimus kuitenkin osoitti, että todellisista metsiköistäkin mitatut järjestystunnusluvut parantavat läpimittajakaumaennusteen tarkkuutta huomattavasti. Metsikön spatiaalisen autokorrelaation vaikutusta voitaneen ainakin osittain eliminoida poimimalla useita pieniä otoksia eri puolilta metsikköä sen sijaan että poimittaisiin muutamia suuria otoksia.

Toinen oletus oli, että metsikkötunnukset ja järjestystunnuslukupuut voidaan mitata virheettömästi. Metsikkötunnuksiin sisältyy kuitenkin virhettä, jonka hajonta voi olla kymmeniä prosentteja metsikkötunnusten arvosta (Kangas ym. 2004) ja järjestystunnuslukupuitakaan tuskin voidaan mitata virheettömästi. Metsikkötunnusten mittausvirheiden vuoksi prosenttipisteiden ennustevirheiden varianssit ovat suurempia kuin matriisi $\mathbf{D}$ osoittaa. Jos käytetyt mallit ovat epälineaarisia metsikkötunnusten suhteen, aiheuttavat virheet lisäksi harhaa ennusteisiin (Mehtätalo ja Kangas 2004). Jos läpimittajakauman ennustamisessa käytettyihin metsikkötunnuksiin sisältyy virhettä eikä sitä huomioida laskennassa, ovat saadut ennusteet liian lähellä pelkkiin metsikkötunnuksiin perustuvaa jakaumaennustetta ja estimoidut ennustevirheiden varianssit ovat liian pieniä. Järjestystunnuslukupuiden mittausvirheet puolestaan vaikuttavat matriisiin **R**. Edellä oletettiin, että matriisi **R** koostuu kokonaan otantavirheistä, mutta mittausvirheet kasvattavat sen varianssejä. Siten oletus, että järjestystunnuslukupuiden mittauksiin ei sisälly virhettä, saa aikaan sen että läpimittajakaumaennusteet pyrkivät tulemaan liian lähelle mitattuja prosenttipisteitä, jos ja kun virhettä todellisuudessa esiintyy. Lisäksi ennustevirheiden varianssit tulevat aliarvioitua. Kaiken kaikkiaan metsikkötunnusten ja järjestystunnuslukupuiden mittausvirheet vievät saatua piste-estimaattia eri suuntiin ja kumoavat toistensa vaikutusta. Siten saadut läpimittajakaumaennusteet voivat olla hyvin lähellä oikeaa, mutta saadut ennustevälit ovat liian kapeita.

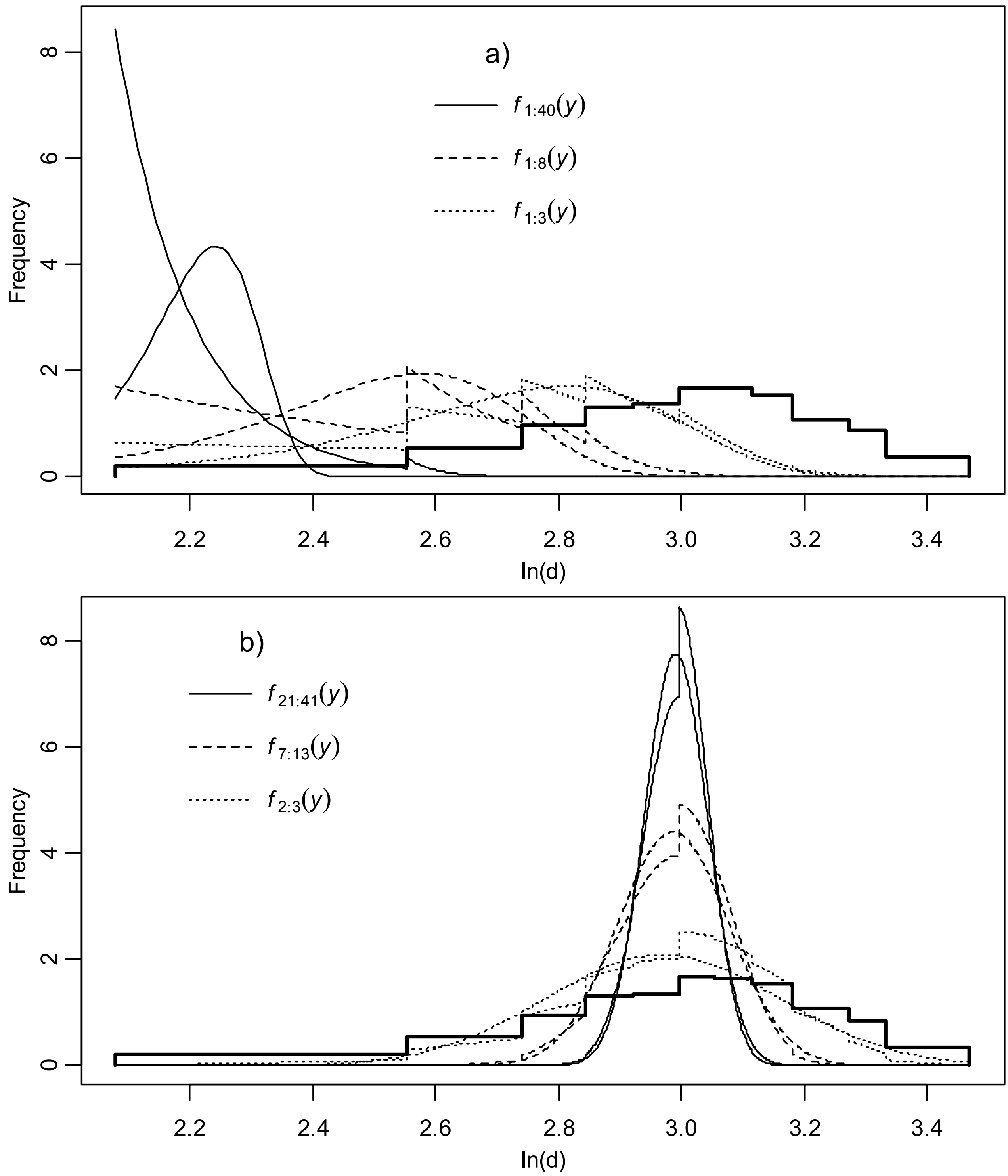
Metsikön läpimittajakaumaa approksimoitiin tässä tutkimuksessa prosenttiosuusjakaumalla, joka on huomattavasti joustavampi kuvausmenetelmä kuin esimerkiksi Weibull-jakauma kun prosenttipisteitä otetaan riittävän paljon. Joustavuuden hintana on luonnollisesti parametrien määrän kasvaminen. Jakauma-approksimaatiosta aiheutuu luonnollisesti epätarkkuutta tuloksiin, jota ei tässä tutkimuksessa voitu tarkastella. Käytetty approksimaatio lähestyy metsikön todellista jakaumaa kun prosenttipisteiden määrää lisätään. Siten approksimoinnista aiheutuva virhe ei luultavasti ole kovin merkittävä, koska käytetty prosenttipisteiden määrä oli kohtalaisen suuri.

Prosenttipisteiden mallien (Kangas ja Maltamo 2000) oletettiin olevan oikeita ja parametrit oletettiin tunnetuiksi. Edellinen oletus joudutaan yleensä aina tekemään mallinnusongelmien yhteydessä, eikä sen tarkastelu ilman alkuperäistä aineistoa ole mahdollista. Mallin parametrit ovat aina estimaatteja, joihin liittyy estimointivirhettä. Kankaan ja Maltamon (2000) laadinta-aineistona oli 251 metsikköä Etelä- ja Keski-Suomesta ja mallien kerrointen estimointivirheitä ei raportoitu. Tässä sovelluksessa parametrien estimointivirheet vaikuttavat samalla tavalla kuin mittausten virheet, eli tuovat ennusteita liian lähelle pelkkiin metsikkötunnuksiin perustuvaa jakaumaennustetta ja saavat aikaan liian kapeita ennustevälejä.

Mitattujen otosjärjestystunnuslukujen odotusarvoa ja varianssia laskettaessa oletettiin metsikön jakauma tunnetuksi. Toteutettu iterointi mahdollisti sen, että näiden tunnuslukujen laskenta perustuu siihen jakaumaan, joka on ennustamisen lopputulos ja siten paras käytettävissä oleva ennuste. Se ei kuitenkaan ole metsikön tosi jakauma, minkä vuoksi otosjärjestystunnuslukupuun odotusarvot, varianssit ja kovarianssit eivät ole täysin oikeita. Niiden virheitä ei tässä yhteydessä tarkastella enempää.

Lineaarisessa ennustamisessa ennustekaavat johdettiin käyttäen todellisia varianssi-kovarianssimatriiseja. Tämän jälkeen todet matriisit korvattiin estimaateilla, eikä estimointivirheiden vaikutuksia huomioitu. Tämä on hyvin yleinen tapa ja käytetyistä yhtälöstä käytetään nimitystä estimoitu parasta lineaarinen ennustin. Matriisien ennustevirheiden huomioiminen johtaa monimutkaisiin matriisilaskentoihin (ks McCulloch ja Searle 2000, s. 170-171).

Tässä tutkimuksessa käytettiin järjestystunnuslukutilastotieteen eksakteja tuloksia. Niiden johtaminen paloittain määritetyn tasaisen jakauman tapauksessa on kohtuullisen yksinkertaista ja johtaa otosjärjestystunnuslukujen paloittain määritettyyn betajakaumaan, joka on itse asiassa polynomi. Tämän jakauman odotusarvon laskeminen on kuitenkin ongelmallista. Odotusarvoja, variansseja ja kovariansseja (kaavat 10, 12 ja 17) laskettaessa joudutaan integroimaan polynomi, jonka asteluku on otoskoko-1 tai otoskoko-2. Relaskooppikerrointa 1 käytettäessä otoskoko on yleensä 15-30. Kun otoskoko nousi 25:een, alkoi laskennassa esiintyä tietokoneen laskentatarkkuudesta johtuvia ongelmia, koska hyvin suuria lukuja jouduttiin korottamaan hyvin pieneen potenssiin. Siksi tässä työssä käytettiin numeerisia integrointimenetelmiä, joiden ongelmana on hitaus. Esimerkiksi luvussa 4.3 raportoitu ajo kesti yhteensä noin vuorokauden. Eksaktien tulosten sijasta voitaisiin käyttää myös järjestystunnuslukutilastotieteen asymptoottisia tuloksia Asymptoottisesti otoksen minimi- ja maksimi lähestyvät ääriarvojakaumaa ja muut järjestystunnusluvut lähestyvät normaalijakaumaa (Reiss 1989, s. 109-110). Laskennat voisivat perustua eksaktien tulosten sijasta näiden jakaumien odotusarvoihin ja variansseihin, jotka voidaan laskea analyyttisesti hyvin nopeasti. Asymptoottisten tulosten käyttöä testattiin tapauksessa, jossa kultakin koealalta mitattiin yksi järjestystunnuslukupuu. Asymptoottisiin tuloksiin perustuvan laskennan avulla saatiin selvästi parempia läpimittajakaumaennusteita kuin ilman järjestystunnuslukumittauksia. Saavutettu läpimittajakaumaennusteiden tarkkuus oli kuitenkin selvästi huonompi kuin eksakteja tuloksia käytettäessä. Tämä oli odotettavissakin, koska käytetyllä relaskooppikertoimella otoskoot ovat niin pieniä, että järjestystunnuslukujen asymptoottiset odotusarvot ja varianssit poikkeavat merkittävästi eksakteista arvoistaan (Kuva 5).



Kuva 5. Otoksen minimin (a) ja mediaanin (b) eksakti jakauma (murtoviivat) eri kokoisissa otoksissa, kun otos poimittiin prosenttiosuusjakaumasta (paksu viiva). Kuvaan on piirretty lisäksi eksaktin jakauman odotusarvon ja varianssin perusteella lasketut asymptoottiset jakaumat (sileät viivat), jotka ovat kuvassa (a) ääriarvojakaumia ja kuvassa (b) normaalijakaumia.

LÄHTEET

1. Bailey, R.L. ja Dell, T.R. (1973). Quantifying diameter distributions with the Weibull function. *Forest Science* **19** 97-104.
2. Borders, B.E., Souter, R.A., Bailey, R.L. ja Ware, K.D. (1987). Percentile-based distributions characterize forest stand tables. *Forest Science* **33** 570-576.
3. Casella, G. ja Berger, R.L. (2002). *Statistical Inference*, 2nd Edition. Pacific Grove: Duxbury.
4. Flewelling, J.W. ja Pienaar, L.V. (1981). Multiplicative regression with lognormal errors. *Forest Science* **18** 241-245.
5. IMSL (1997). *Fortran Subroutines for Mathematical Applications*. Math/Library. Volumes 1&2. Visual Numerics.
6. Kangas, A. ja Maltamo, M. (2000). Percentile-Based Basal Area Diameter Distribution Models for Scots Pine, Norway Spruce and Birch Species. *Silva Fennica* **34** 371-380.
7. Kangas, A., Heikkinen, E. and Maltamo, M. (2004). Accuracy of partially visually assessed stand characteristics – A case study of Finnish inventory by compartments. *Canadian Journal of Forest Research* **34** 916-930.
8. Kilkki, P. ja Päivinen, R. (1986). Weibull function in the estimation of the basal area dbh distribution. *Silva Fennica* **20** 149-156.
9. Laasasenaho, J. (1982). Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae* **108** 1-74.
10. Lappi, J. (1991). Calibration of Height and Volume Equations with Random Parameters. *Forest Science* **37** 781-801.
11. Lin, C. (2003). *Generating forest stands with spatio-temporal dependencies*. Joensuun yliopiston yhteiskuntatieteellisiä julkaisuja 64.
12. McCulloch, C. E. ja Searle, S. R. (2001). *Generalized, Linear and Mixed Models.* New York: Wiley.
13. Maltamo, M. (1997). Comparing basal area diameter distributions estimated by tree species and for the entire growing stock in a mixed stand. *Silva Fennica* **31** 53-65.
14. Mehtätalo, L. ja Kangas, A. (2004). An approach to optimizing field data collection in an inventory by compartments. *Canadian Journal of Forest Research* (in press).
15. Pukkala, T. Vetteenranta, J, Kolström, T. ja Miina, J. (1994). Productivity of mixed stands of Pinus sylvestris and Picea abies. *Scandinavian Journal of Forest Research* **9** 143-153.
16. Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics With Applications to Nonparametric Statistics.* New York: Springer-Verlag.
17. Rennols, K., Geary, N. and Rollison, T.J.D. (1985). Characterizing diameter distribution by the use of Weibull distribution. *Forestry* **58** 57-66.
18. Savolainen, S. (2004). *Metsänmittausohjeita.* Suomen metsäyhdistys, moniste. < URL http://www.oppimispolku.fi > Luettu 22.11.2004.
19. Tomppo, E. (1986). Malleja ja menetelmiä puiden tilajärjestyksen analysoimiseksi. Models and methods for analysing spatial patterns of trees. *Communicationes Instituti Forestalis Fenniae* **138** 1-65

20. Venables, W.N. ja Ripley, B D. (2002). *Modern Applied Statistics with S.* 4th edition. New York: Springer-Verlag.