

KOEPUUTIEDON KÄYTTÖ INVENTOINTITULOSTEN LASKENNASSA

Kari T. Korhonen

Metsäntutkimuslaitoksen tiedonantoja 569

N:o 569

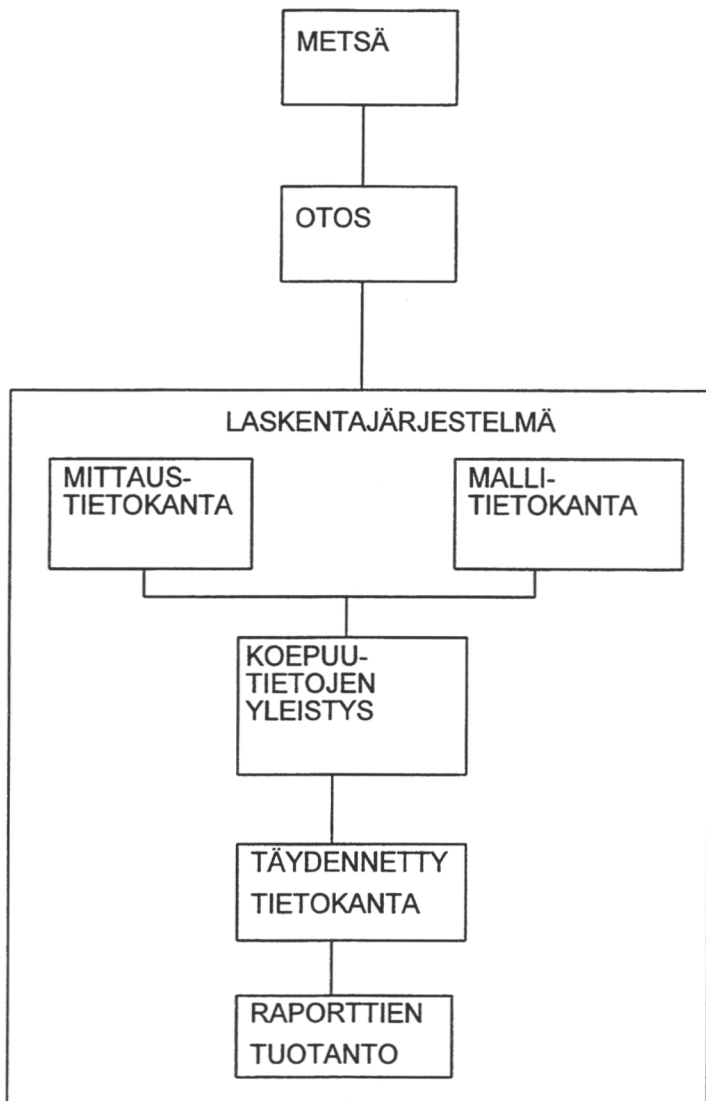
Kari T. Korhonen

KOEUUTIEDON KÄYTTÖ INVENTOINTITULOSTEN LASKENNASSA

AKATEEMINEN VÄITÖSKIRJA

Esitetään Joensuun metsätieteellisen tiedekunnan suostumuksella julkisesti tarkastettavaksi yliopiston päärakennus Carelian auditoriossa C 2, Yliopistokatu 2, perjantaina 15. päivänä joulukuuta 1995 klo 12.

Joensuun yliopisto
Joensuu 1995



KOEPUUTIEDON KÄYTTÖ INVENTOINTITULOSTEN LASKENNASSA

Kari T. Korhonen

ISSN 0358-4283
ISBN 951-40-1469-3

Gummerus Kirjapaino Oy
Saarijärvi 1995

TIIVISTELMÄ

Tutkimuksessa kootaan yhteen neljän osajulkaisun keskeiset tulokset ja päätelmät. Osajulkaisut käsittelevät suuralueen inventoinnin laskentamenetelmiä. Ensimmäisessä osajulkaisussa esitetään sekamallitekniikkaan perustuvat runkokäyrämallit männylle, kuuselle ja koivulle. Runkokäyrämallien antamia tilavuusestimaatteja verrataan kiinteäparametrisilla yhtälöillä saataviin estimaatteihin valtakunnan metsien inventoinnin (VMI) koepuuaineistossa. Toisessa osajulkaisussa tutkitaan vanhan ja uuden koepuutiedon yhdistämistä sekaestimoinnilla männyn tilavuusyhtälöiden laatimiseksi. Sekaestimaattoria verrataan tavanomaiseen pienimmän neliösumman estimaattoriin, joka ei hyödynnä ennakkoinformaatiota. Kolmannessa osajulkaisussa esitetään sekaestimaattoriin ja tavanomaisiin regressiomalleihin perustuva inventoinnin laskentajärjestelmä. Neljännessä osajulkaisussa tutkitaan lähimmän naapurin menetelmän soveltamista koepuutunnusten yleistämiseen lukupuille. Tutkimuksessa todetaan, että käytetyt, sekamallitekniikkaan perustuvat runkokäyrämallit eivät ole tarpeeksi joustavat tarkentuakseen riittävästi metsikkökohtaisesti. Sekamallitekniikka ja sekaestimaattori ovat kuitenkin tehokkaita menetelmiä ennakkoinformaation hyödyntämiseksi inventoinnissa. Ei-parametrinen lähimmän naapurin menetelmä voidaan suositella laskentajärjestelmiin, joissa koepuutunnusten estimaatteja käytetään jatkolaskennassa selittäjinä (esimerkiksi puuston tulevan kehityksen simulointiin).

Avainsanat: metsäninventointi, tilavuus, sekamallit, sekaestimointi, ei-parametriset mallit, lähimmän naapurin menetelmä.

ALKUSANAT

Tässä työssä esiteltävät tutkimukset on tehty Metsäntutkimuslaitoksen Valtakunnan metsien inventoinnin hankkeessa. Ensimmäisen osajulkaisun aikoihin hankkeen johtajana oli prof. Pekka Kilkki. Prof. Kilkin esittämät ajatukset olivat pohjana myös myöhempiä osajulkaisuja tehdessäni. Prof. Erkki Tomppo on Kilkin seuraajana vaikuttanut merkittävästi tutkimusteni edistymiseen.

Valtakunnan metsien inventoinnin hankkeessa työskentelevät henkilöt ovat myötävaikuttaneet siihen, että käytettävissäni on ollut erinomainen VMI:n maastoaineisto. Erityisesti edesmennyt mt Matti Kujala, MMK Arto Ahola ja MH Sakari Salminen ovat avustaneet aineistoon ja inventoinnin käytäntöön perehtymisessä. VTT Juha Lappi perehdyttänyt sekamallitekniikkaan ja vaikuttanut myös kolmannen ja neljännen osajulkaisun sisältöön. Keskustelut prof. Juha Alhon ja FK Jaakko Heinosen kanssa ovat auttaneet tilastollisten menetelmien soveltamisessa.

Osajulkaisuja ovat töiden eri vaiheissa kommentoineet jo mainittujen henkilöiden lisäksi MMT Helena Henttonen, MMT Jouko Laasasenaho, MMT Risto Ojansuu, MMT Risto Päivinen, prof. Simo Poso ja MMT Jussi Saramäki. Yhteistyö MMT Annika Kankaan kanssa on mahdollistanut neljännen osajulkaisun syntymisen.

Metsäntutkimuslaitoksen Joensuun tutkimusasema ja joensuulainen metsätiedeyhteisö ovat tarjonneet innostavan ilmapiirin tutkimuksen tekemiseen. Kiitän lämpimästi kaikkia työni sisältöön ja edistymiseen vaikuttaneita henkilöitä.

Syyskuussa 1995

Kari T. Korhonen

LUETTELO OSAJULKAISUISTA

I

Korhonen, K.T. 1991. Sekamallitekniikalla laadittujen runkokäyrämallien käyttö metsäinventoinnissa. Summary: Using taper curve models based on mixed linear models in forest inventory. *Folia Forestalia* 774. 27 s.

II

Korhonen, K.T. 1993. Mixed estimation in calibration of volume functions of Scots pine. Tiivistelmä: Sekaestimointi männyn tilavuusyhtälöiden kalibroinnissa. *Silva Fennica* 27(4):269-276.

III

Korhonen, K.T. 1994. Calculation system for large-scale forest inventory. *Metsäntutkimuslaitoksen tiedonantoja* 505. Joensuun tutkimusasema. 36 s.

IV

Korhonen, K.T. & Kangas, A. 1995. Application of nearest-neighbor regression for generalising sample tree information. *Scandinavian Journal of Forest Research* -sarjaan toimitettu käsikirjoitus. 16 s.

SISÄLLYS

1. JOHDANTO	5
1.1 Metsäninventoinnin tavoitteet	5
1.2 Tutkimuksen tavoite	7
2. TILASTOLLISET MENETELMÄT SUURALUEEN INVENTOINNISSA ...	8
2.1 Otanta	8
2.2 Runkokäyrämallit koepuiden tilavuuksien estimoinnissa	10
2.3 Tilavuuskasvun estimointi koepuutunnuksista	12
2.4 Koepuutunnusten yleistäminen lukupuille	13
2.4.1 Suorat menetelmät	13
2.4.2 Vaiheittaiset menetelmät	15
2.4.3 Ennakkoinformaation käyttö	16
3. PÄÄTELMÄT	23
KIRJALLISUUS	25
LUETTELO OSAJULKAISUJEN VIRHEISTÄ	29
OSAJULKAISUT	

1. JOHDANTO

1.1 Metsäinventoinnin tavoitteet

Cunia (1991) luokitteli metsäinventoinnit neljään ryhmään: 1) toiminnan suunnitteluun tähtäävät inventoinnit (yleensä pienille alueille), 2) metsänhoidon suunnittelemiseksi pienille metsäalueille tehtävät inventoinnit, joissa seurataan muutaman keskeisen parametrin kehitystä, 3) suuraluetason ja valtakunnallisen tason inventoinnit, jotka poikkeavat edellisestä vain suuremman pinta-alansa vuoksi ja 4) maapallon kattavat inventoinnit, joiden päätavoite on seurata metsien pinta-alan kehittymistä maapallon laajuisesti ja tuottaa tietoa metsien suojelemiseksi. Tässä kirjoituksessa suuralueen inventoinnilla tarkoitetaan kolmatta luokkaa Cunian (1991) ryhmittäessä.

Perinteisesti suuralueen inventoinnin keskeisin tavoite on ollut metsien tilan seuranta. Esimerkiksi Suomen ensimmäinen valtakunnan metsien inventointi (VMI) aloitettiin 1920-luvulla, koska nousevan sahateollisuuden ja tervanpolton pelättiin hävittäneen maan puuvarat. Sama pelko oli toisen VMI:n taustalla. Myöhemmin metsien terveydentilan ja kasvillisuuden muutosten seuranta ovat tulleet osaksi VMI:n tavoitteita. Jo ensimmäisten inventointien tietoja käytettiin myös tulevien hakkuumahdollisuuksien arviointiin (Ilvessalo 1943).

Metsävarojen tilaa on seurattu estimoimalla inventointiaineistoista muutamia keskeisiä parametreja Suomen metsäpopulaatiolle (esimerkiksi Kuusela & Salminen 1991). Keskeisimmistä näistä parametreista ovat olleet metsämaan pinta-ala, puuston eri ikäluokkien osuudet, puuston kokonaistilavuus ja puuston kasvu. Näiden parametrien avulla on voitu estimoida myös tulevia hakkuumahdollisuuksia (Ilvessalo 1943, Kuusela & Nyysönen 1962).

Tietojenkäsittelyn kehittyessä ja matemaattisten menetelmien tuntemuksen levitessä hakkuumahdollisuuksien arviointiin käytettävät suunnittelumenetelmät ovat muuttuneet. Myös inventointimenetelmien kehittyminen on osaltaan mahdollistanut kehityksen. Nykyaikaisen suunnittelumenetelmän keskeinen ydin on simulaattori, jolla voidaan ennustaa metsien tuleva kehitys erilaisilla käsittelyvaihtoehdoilla. Kun toiminnan tavoitteet ja rajoitteet on asetettu, optimointimenetelmillä voidaan poimia käsittelyvaihtoehdoista tavoitteet parhaiten toteuttava vaihtoehto (esimerkiksi Kilkki & Siitonen 1975, Jonsson ym 1993, Siitonen 1993).

Suuralueen inventoinnin laskentajärjestelmän on ensisijaisesti pystyttävä tuottamaan metsien

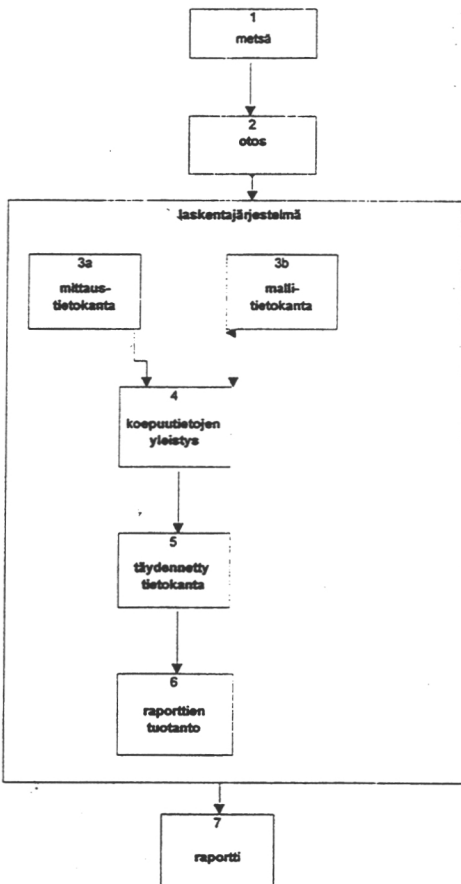
seurannassa käytettäville parametreille täsmälliset ja tarkat estimaatit. Jos inventoinnilla on muita tavoitteita, nämä on otettava huomioon aineiston keruussa ja käsittelyssä. Esimerkiksi, jos tavoitellaan metsien tulevan kehityksen ennustamista, otosyksiköt (metsiköt tai koealat) on kuvattava tavalla, joka mahdollistaa puuston kasvun ja kuoleamisen ennustamisen.

Suuralueen inventoinnissa käytetään yleensä monivaiheista otantaa (Pelz 1991). Tällöin inventoinnin laskentajärjestelmässä voidaan erottaa seuraavat osat.

1. tilavuuksien ja kasvujen estimointi koealuille,
2. koeputunnuksista johdettujen tunnusten yleistäminen lukupuille ja
3. aluekohtaisten tulosten summaus.

Vaiheet 1 ja voivat olla myös päinvastaiset kuin edellä:

1. koeputunnusten yleistäminen lukupuille ja
2. parametrien (esimerkiksi tilavuus) estimointi koeputunnusten estimaateista.



Kuva 1. Inventoinnin kuvaus systeeminä.

Kuvassa 1 on esitetty edellä kuvatun laskentajärjestelmän sijoittuminen metsäinventoinnin systeemiin. Kuvan 1 esitys noudattaa pääpiirteissään Päivisen (1987) inventointisysteemin kuvausta. Inventoinnin laskentajärjestelmän keskeisin tuloste on tilasto tai raportti. Järjestelmässä tuotettava täydennetty mittaustietokanta (vaihe 5 kuvassa 1) voi toimia myös lähtöaineistona muille järjestelmille, esimerkiksi suunnittelusysteemille. Tällaisella menettelyllä voidaan turvata se, että inventoinnin laskentajärjestelmä ja suunnittelujärjestelmä tuottavat yhdenmukaisia estimaatteja.

1.2 Tutkimuksen tavoite

Tässä työssä tutkitaan menetelmiä, joilla suuralueen inventoinnin mittausaineistosta voidaan tuottaa estimaatit laskenta-alueen perusparametreille - puuston kokonaistilavuudelle, puutavaralajeittaisille tilavuuksille ja eri ositteiden pinta-aloille. Tutkimus keskittyy puutavaralajeittaisten tilavuuksien estimointiin ja koepuutietojen yleistämiseen lukupuille. Tutkimuksen osajulkaisuissa käsitellään lyhyesti myös puuston menneen kasvun estimointia ja tulosten summausta suuralueelle tai sen ositteelle.

Tutkimuksen lähtökohtana on, että inventointi toteutetaan otosinventointina. Otantavirheen tai inventoinnin muiden virheiden estimointi eivät ole tutkimuksen kohteena. Menetelmien valinnassa ja testauksessa on kiinnitetty erityistä huomiota menetelmien käyttökelpoisuuteen käytännön laskentajärjestelmissä.

2. TILASTOLLISEN MENETELMÄT SUURALUEEN INVENTOINNISSA

2.1 Otanta

Populaation kokonaistutkimus (complete enumeration) on käyttökelpoinen menetelmä vain suhteellisen pienissä (=vähän alkioita) populaatioissa (Liedes & Manninen 1975). Metsällisiä esimerkkejä täydellisestä numeroinnista ovat leimikon pystymittaus ja kuvioittainen arviointi. Leimikon mittauksessa inventoitava alue on pieni, jolloin kaikkien puiden luku (mittaus) on mahdollista. Kuvioittaisessa arvioinnissa populaation alkioksi on valittu puun sijasta kuvio, jolloin alkioiden määrä on suurellakin alueella kohtuullisen pieni.

Suuralueen inventoinnissa (kuten VMI) on käytettävä otantaa, jotta mittauskustannukset pysyisivät kohtuullisina (Pelz 1991). Kun käytettävissä ei ole ennakkoinformaatiota mitattavasta alueesta, on tavanomaista soveltaa yksinkertaisia otantamenetelmiä. Systemaattinen otanta on yleisempi kuin satunnaisotanta. Mittauskustannusta pienennetään usein ryvästämällä otosyksiköitä, jolloin voidaan soveltaa yksi- tai kaksiaseteista ryväotantaa.

Kun käytettävissä on ennako- tai lisäinformaatiota maastoaineiston ohella, on mahdollista käyttää tehokkaampia otantamenetelmiä. Karttatietoa (esimerkiksi kasvillisuuskartat) voidaan hyödyntää inventointialueen osittamisessa, jolloin voidaan soveltaa ositettua otantaa. Kaukokartoitustietoa voidaan hyödyntää osittamisessa, yhtenä otannan vaiheena tai regressiomenetelmän tai referenssikoealamenetelmän osana (Poso 1973, Tokola 1990, Tomppo 1992).

Osittaminen jakaa mitattavan populaation osiin, joissa otanta voidaan tehdä toisistaan riippumatta (Liedes & Manninen 1975). Ositus voi olla joko etu- tai jälkikäteen ositusta. Etukäteen osituksessa maastokoealat voidaan allokoida tehokkaasti esimerkiksi kaukokartoitustiedon ohjaamattomalla luokittelulla (Tokola 1990). Jos kaukokartoitustietoa käytetään jälkikäteen ositukseen, kaukokartoitustiedon ainoa tehtävä on painottaa maastokoealoja ositteiden kokoestimaattien mukaisesti. Kullekin ositteelle on laskettava maastokoealoista perusparametrit aivan vastaavalla tavalla kuin pelkkään maasto-otokseen perustuvassa inventointimenetelmässä lasketaan parametrit laskenta-alueille.

Kun kaukokartoitustietoa käytetään yhtenä otannan vaiheena (Tokola 1990, Tomppo 1992), kaukokartoituksen tehtävä on sama kuin jälkikäteen osituksessa: koealojen painottaminen

toistensa suhteen. Referenssikoeala- ja regressiomenetelmissä koealoille lasketut puuston parametrit yleistetään satelliittikuvan kuva-alkioille tai niistä segmentoiduille kuviolle (Tokola 1990). Esimerkiksi Suomen monilähteisen VMI:n tulosten laskennassa satelliittikuvan kuva-alkiot ovat ensimmäisen vaiheen täydellisesti numeroitu 'otos', jolle maastokoealojen tiedot yleistetään (Tomppo 1992). Käytännössä tämä merkitsee koealojen painottamista kuva-alkioiden avulla. Suomen monilähteisessä VMI:ssa satelliittikuvien toinen tehtävä on rasterimuotoisten teemakarttojen tuotanto (Tomppo 1992).

Edellä käydystä otantamenetelmien sovellusten lyhyestä katsauksesta ilmenee, että pelkkään maasto-otantaan perustuvan inventoinnin ja esimerkiksi kaukokartoitusta hyödyntävän inventoinnin laskentamenetelmät voivat olla suurelta osin samat. Kaikkiin inventointimenetelmiin liittyvän laskentamenetelmän yhtenä vaiheena on parametrien laskenta jollekin laskentayksikölle pelkkää maastoaineistoa käyttäen. Laskentayksikkö voi olla menetelmästä riippuen koeala tai koealajoukon muodostama osite.

Useimmissa inventoinneissa ei ole taloudellisesti mahdollista mitata suoraan sitä muuttujaa (esimerkiksi tilavuus tai biomassa), jonka tilastollista parametria (esimerkiksi summa) ollaan estimoimassa. Tämän vuoksi kaksi- tai monivaiheista otantaa sovelletaan yleisesti myös pelkkään maastomittaukseen perustuvissa inventoinneissa (Kilkki 1979, Cunia 1986a, Poso ym 1990, Ranney & Svensson 1991). Seuraavankaltainen sovellus on yleinen. Ensimmäisen vaiheen otoksena mitataan suhteellisen suuri otos puita (lukupuut), joista mitataan läpimitta rinnankorkeudella (d) ja joitakin muita helposti mitattavia tunnuksia. Toisessa vaiheessa poimitaan lukupuista alaotos (koeput), johon kuuluvat puut mitataan täydellisemmin kuin lukupuut. Tyypillisiä koepuista mitattavia tunnuksia (koeputunnuksia) ovat pituus (h), ikä (t), lisäläpimitta (yleensä d_6 eli 6 m korkeudelta mitattu läpimitta), eri puutavaralajien pituudet rungossa ja läpimitan kasvu. Kolmannen vaiheen otos tarvitaan tilavuuksien ja kasvujen estimoimiseksi koeputunnuksista. Usein tätä kolmannen vaiheen otosta ei pidetä osana inventointia vaan koeputlaskennassa käytettäville malleille annetaan luonnonlakien asema.

2.2 Runkokäyrämallit koepuiden tilavuuksien estimoinnissa (I)

Koepuiden tilavuudet estimoidaan mitatuista koepuutunnuksista yleensä yksinkertaisilla regressiomalleilla. Mallit perustuvat joko inventoinnissa tarkasti mitattuun koepuuaineistoon tai aiemmin mitattuun enemmän tai vähemmän subjektiivisesti valittuun aineistoon. Eri rungon osien tai puutavaralajien tilavuuksien estimointiin käytetään joko runkokäyrämalleja (esimerkiksi Laasasenaho 1982, Lappi 1986, I) tai tilavuusmalleja, jotka estimoivat tilavuuden kannosta annettuun vakioläpimittaan (esimerkiksi Gregoire & Schabenberg 1994).

Runkokäyräytälöt ovat edellä mainituista kahdesta vaihtoehdosta käyttökelpoisempia silloin, kun tilavuuden on estimoitava puutavaralajeille, joiden pituus- ja läpimittavaatimukset vaihtelevat. Runkokäyrämalleilla voidaan estimoida läpimitta annetulla korkeudella ja tilavuus annetulle rungon osalle voidaan laskea joko analyttisellä tai numeerisella integroinnilla (Lahtinen & Laasasenaho 1979, Laasasenaho 1982). Runkokäyrämallien estimointi onkin niiden käyttökelpoisuuden vuoksi ollut suosittu tutkimuskohde metsänarvioimistieteessä (esimerkiksi Cajanus 1911, Kuusela 1965, Laasasenaho 1982, Lappi 1986, Sterba 1990).

Runkokäyrämalleissa on tyypillisesti mallitettu suhdetta d_l/d_x l:n funktiona, missä d_l on läpimitta (suhteellisella) korkeudella l ja d_x on valittu perusläpimitta. Esimerkiksi Laasasenahon (1982) malleissa perusläpimittana on läpimitta suhteellisella korkeudella 20 % ja runkomuotoa estimoidaan polynomiyhtälöllä (1).

$$d_l/d_{2h} = a_1x + a_2x^2 + a_3x^3 + a_4x^5 + a_5x^8 + a_6x^{13} + a_7x^{21} + a_8x^{34}, \quad (1)$$

missä d_{2h} = läpimitta 20 % suhteellisella korkeudella,

d_l = läpimitta suhteellisella korkeudella l ja

$x = 1-l/h$ (suhteellinen etäisyys latvasta).

Edellä esitetyn suhteellisten korkeuksien läpimittoihin perustuvan runkokäyrämallin merkittävin puute on, että puun pituus on tunnettava malleja käytettäessä. Jos pituutta ei tunneta, se voidaan estimoida erillisellä mallilla. Tällöin pituusestimaatissa oleva satunnaisvirhe aiheuttaa harhaa runkokäyräestimaattiin. Yleensä tätä virhelähdettä ei ole otettu huomioon inventoinnin laskentajärjestelmissä.

Kilkki ym (1978) ja Laasasenaho (1982) ovat esittäneet simultaanisiin yhtälöihin perustuvan runkokäyrämallin. Niissä pituus on yhtenä estimoitavana rungon dimensiona. Muut estimoitavat dimensiot ovat läpimittoja suhteellisilla korkeuksilla. Tämän vuoksi myös Kilkin ym (1978) ja Laasasenahon (1982) simultaaniyhtälöihin perustuvat runkokäyrämallit tosiasiallisesti olettavat puun pituuden tunnetuksi suureeksi.

Lappi (1986) esitti napakoordinaatistoon perustuvan runkokäyrämallin. Napakoordinaatistosysteemissä puun runkokäyrä kuvataan läpimittoina eri kulmissa (kuva 1 osajulkaisussa I). Läpimitta voidaan ilmaista kulman funktiona yhtälön 2 mukaisesti.

$$d(u) = \cos(u)R(u), \quad (2)$$

missä u = napakoordinaatioston kulma ja

$R(u)$ = kulmaan u liittyvä säde (hypotenuusa).

Kun puun runkokäyrä kuvataan napakoordinaatistossa, puun pituus voidaan pitää yhtenä estimoitavana muuttujan eikä pituutta tarvitse käyttää rungon läpimittojen korkeuksien määrittämiseen. Lapin (1986) runkokäyrämalli sisälsi 14 yhtälöä, joista 12 estimoi eri kulmien läpimittoja, yksi pituutta ja yksi latvusrajan korkeutta. Jokainen yhtälö sisälsi kaksi komponenttia: kiinteän osan ja satunnaisen osan (yhtälö 3).

$$d_{ki}(u) = a_0(u) + a_1(u)s_{ki} + a_2(u)s_{ki}^2 + a_3s'_{ki} + v_k(u) + e_{ki}(u), \quad (3)$$

missä

$d_{ki}(u)$ = logaritminen läpimitta kulmassa u puulle i metsikössä k ,

$a_0(u)$, $a_1(u)$, $a_2(u)$ ja $a_3(u)$ ovat kiinteiden selittäjien kertoimia,

s_{ki} = puun koko,

s'_{ki} = puun suhteellinen koko (suhteessa muihin metsikön puihin),

$v_k(u)$ = koealakohtainen satunnaisvaikutus ja

$e_{ki}(u)$ = puukohtainen satunnaisvaikutus.

Osajulkaisussa I on käytetty Lapin (1986) kehittämää mallintamistapaa männyn, kuusen ja koivun runkokäyrän mallittamiseen. Nämä sekamalleihin perustuvat runkokäyrämallit havaittiin rungon tilavuuden estimoimisessa luotettavuudeltaan likimäärin samantasoisiksi kuin

Laasasenahon (1982) esittämät polynomimuotoiset runkokäyrämallit, kun d ja h tai d , d_6 ja h ovat selittäviä muuttujia (taulukko 1).

Taulukko 1. Todellisten ja estimoitujen puittaisten tilavuuksien (dm^3) keskiarvo (\bar{x}) ja keskihajonta (sd) laadinta-aineistoissaan, kun estimointimenetelmä on ollut polynomimuotoinen (POL) ja napakoordinaatistoon ja sekamalleihin perustuva (MIX) runkokäyrämalli. (Lähde: Laasasenaho 1982 ja I.)

d ja h mitattu						
	Mänty		Kuusi		Koivu	
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
POL	0.40	7.31	0.55	7.49	0.53	8.79
MIX	0.85	7.29	0.53	7.40	0.50	8.69

d, d_6 ja h mitattu						
	Mänty		Kuusi		Koivu	
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
POL	0.31	3.46	0.39	3.41	0.61	4.95
MIX	0.33	3.50	-0.32	3.23	-0.36	5.38

2.3 Tilavuuskasvun estimointi koepuutunnuksista

Tilavuuskasvun suora mittaaminen on vielä hankalampaa kuin rungon tilavuuden suora mittaaminen. Metsänarvioimisessa menneeseen kasvuun liittyvänä tunnuksena on yleisesti mitattu rinnankorkeusläpimitan kasvu. Havupuilla käytetään myös pituuskasvua, joka useilla havupuulajeilla sopivissa oloissa on määritettävissä kasvukiehkuroiden avulla. Rungon tilavuuskasvun johtamiseksi läpimitan ja pituuden kasvusta on kehitetty useita menetelmiä. Svensson (1988) ryhmitteli menetelmät kolmeen ryhmään (katso myös Kangas & Päivinen 1994, sivu 48):

1. komponenttimenetelmät
2. erotusmenetelmät
3. suora kasvun mallitus.

Komponenttimenetelmässä tilavuuskasvu jaetaan kolmeen osaan: rinnankorkeuden poikkileikkausalan kasvuun, pituuskasvuun ja muotoluvun muutokseen. Metsäinventoinnissa poikkileikkausalan ja pituuden kasvu voidaan mitata (joissakin tapauksissa) ja muotoluvun muutos estimoidaan malleilla.

Erotusmenetelmässä oletetaan, että tilavuutta selittävät muuttujat on mitattu kasvunlaskentajakson alussa ja lopussa. Yleensä selittävinä muuttujina ovat läpimitta rinnankorkeudella ja pituus. Menetelmässä lasketaan tilavuus jakson alussa ja lopussa tilavuusyhtälöiden avulla, tilavuuskasvu saadaan näiden tilavuusestimaattien erotuksena. Erotusmenetelmä perustuu oletukseen, että tilavuusyhtälöllä voidaan riittävän luotettavasti estimoida rungon muodon muutos kasvunlaskentajakson aikana.

Kasvun mallitukseen perustuvassa menetelmässä tilavuuskasvu ennustetaan regressioyhtälöllä käyttäen selittäjinä rungosta mitattuja tunnuksia. Esimerkiksi Svensson (1988) on laatinut Ruotsin valtakunnan metsien inventointia varten runkoanalyysiaineistosta tilavuuskasvua estimoivat yhtälöt. Yhtälöissä ovat selittäjinä mm. läpimitta, pituus, läpimitan kasvu, puun ikä ja latvusrajan korkeus. Tutkimuksessaan Svensson (1988) osoitti, että mallitukseen perustuvaa menetelmää käytettäessä on varmistettava mallien laadinta-aineiston edustavuus.

Suomen VMI:ssa puun tilavuuskasvu lasketaan menetelmällä, joka on lähinnä erotusmenetelmää Svenssonin (1988) ryhmittelyssä. Kasvua kuvaavina tunnuksina mitataan läpimitan kasvu, pituuskasvu ja kuoren paksuus. Rungon muodon muuttumista ja kuoren paksuuskasvua ei mitata. Nämä muutokset otetaan huomioon Salmisen (1978) ja Kujalan (1980) kuvaamalla tavalla (katso myös osajulkaisu III). Tässä menetelmässä oletetaan, että kuorellisen tilavuuden ja rinnankorkeuden kuorettoman poikkileikkausalan suhteen (v/g_i) muutos kasvunlaskentajakson aikana voidaan estimoida funktiolla ($v/g_i=f(h)$), jonka kertoimet on estimoitu inventoinnissa mitatuista mittaushetken rungon kuorellisista tilavuuksista, pituuksista ja kuorettomista poikkileikkausaloista.

2.4 Koepuutietojen yleistäminen lukupuille

2.4.1 Suorat menetelmät (I, III)

Koepuutietojen yleistäminen lukupuille on ollut metsänarvioimistieteessä runkokäyrämallien ohessa suosittu tutkimusaihe (Burk & Ek 1982, Kilkki 1983, Pekkonen 1983, Green ym. 1985, Cunia 1986b, 1986c, 1986d, Lappi 1986, 1991, Lappi & Bailey 1988, Meng ym 1990, Ranney & Svensson 1991). Tässä kirjoituksessa käsitettä 'suorat menetelmät' käytetään menetelmistä, joiden perusajatuksena on paikallisten tilavuustaulukoiden tai -yhtälöiden (local tariff tables) laadinta (Loetsch ym 1973, Päivinen 1987, Kangas ym 1990).

Yksinkertaisimmillaan koepuutietojen yleistys tehdään seuraavasti.

1. Estimoidaan koepuiden tilavuudet, puutavaralajeittaiset tilavuudet ja kasvut koepuutunnusten ja mallien avulla.
2. Lasketaan tilavuuksien ja kasvujen keskiarvot valituissa ositteissa, esimerkiksi kasvupaikka- ja läpimittaluokittain.
3. Siirretään jokaiselle lukupuulle ko. ositteen koepuiden tilavuuksien ja tilavuuskasvun keskiarvo.

Edellä kuvatun menetelmän etuna on harhattomuus ja yksinkertaisuus. Menetelmä on kuitenkin tehoton ja edellyttää suuria koepuumääriä, jotta jokaiseen ositteeseen saadaan riittävä määrä havaintoja. Jos koepuita on vähän, ositteita on myös oltava vähän, jolloin puukohtaisten estimaattien jäännösvirhe jää suureksi. Menetelmää voidaan tehostaa käyttämällä ositteittaisten keskiarvojen tilalla jatkuvia regressioyhtälöitä, kuten on tehty osajulkaisussa III. Yhtälöiden laadinnassa tavallinen pienimmän neliösumman menetelmä (ordinary least squares, OLS) on käyttökelpoinen yksinkertaisuutensa vuoksi.

Regressiomalleihin perustuvan menetelmän heikkoutena on, että eri tunnuksien estimaattien yhteisvaihtelu ei ole samalainen kuin kyseisten muuttujien todellisten arvojen yhteisvaihtelu. Regressiomallit voivat antaa siten luotettavat estimaatit esimerkiksi tilavuudelle ja tilavuuskasvulle, mutta estimaateista johdettu tilavuuskasvuprosentti ei välttämättä ole realistinen (Holm ym 1979). Jos inventoinnin laskentajärjestelmää käytetään pelkästään pinta-alojen, puutavaralajeittaisten tilavuuksien ja menneen kauden kasvun estimoimiseen laskenta-alueen eri ositteille, edellä kuvattu puukohtaisten tunnusten estimaattien yhteisvaihtelun vääristyminen ei ole ongelma. Jos inventointiaineistoa käytetään metsien tulevan kehityksen simuloimiseen, on simuloinnissa lähtökohtana käytettävien muuttujien estimaattien yhteisvaihtelun oltava 'luonnollista'. Estimoitavien muuttujien harmonisointi voidaan toteuttaa monimuuttujamenetelmillä (esimerkiksi Lappi 1986, Ojansuu 1993, I).

2.4.2 Vaiheittaiset menetelmät (IV)

Tässä yhteydessä vaiheittaisilla menetelmillä tarkoitetaan koepuutietojen yleistysmenetelmiä, joissa estimoidaan ensin lukupuille koepuutunnukset ja sen jälkeen estimoidaan lopulliset kiinnostavat tunnuksot käyttäen koepuutunnusten estimaatteja selittäjinä (Kilkki 1979). Yksinkertaisimmillaan näissä menetelmissä esimerkiksi tilavuuden estimointi lukupuille tehdään seuraavasti:

1. estimoidaan koepuista mallit pituudelle ja yläläpimitalle,
2. estimoidaan em. malleilla lukupuille pituudet ja yläläpimitat,
3. estimoidaan yleisillä tilavuusmalleilla lukupuille tilavuudet käyttäen selittäjinä pituus- ja yläläpimittaestimaatteja.

Kuvatussa menetelmässä on kahdentyyppisiä ongelmia:

1. kun pituus- ja yläläpimittaestimaatteja käytetään selittäjinä, ovat tilavuusestimatit harhaisia, ellei estimaattien satunnaisvirhettä oteta huomioon (Kilkki 1979) ja
2. pituus- ja yläläpimittaestimaattien välinen suhde ei ole välttämättä realistinen (Kilkki 1979, Ranneby & Svensson 1991, IV).

Kun estimoitavia koepuutunnuksia on useita, on edellä kuvattujen ongelmien ratkaiseminen hankalaa. Esimerkiksi VMI:n tulostenlaskenta edellyttäisi seuraavien koepuutunnusten mallittamista: pituus, yläläpimitta, kuoren paksuus, 5 vuoden läpimitan kasvu, pituuskasvu ja rungon tukkisaantoon vaikuttavat viat. Näiden kaikkien muuttujien estimointi parametrisilla malleilla siten, että estimaattien varianssi ja kovarianssi on todenmukainen lienee käytännössä mahdotonta. Käyttökelpoinen ratkaisu voisi olla Rannebyn & Svenssonin (1991) esittämä menetelmä, jossa yksittäisten mallien estimaattien antamiin tunnusten estimaatteihin lisätään satunnaislukuja siten, että estimaattien varianssit ja kovarianssit sekä koealojen välillä että sisällä on luonnollisen kaltaista. Satunnaislukuja käyttöön perustuva inventoinnin laskentamenetelmä ei kuitenkaan ole houkutteleva vaihtoehto.

Holm ym (1979) esittivät niin sanotun grid-menetelmän koepuutietojen yleistämiseksi lukupuille. Menetelmässä koepuuaineisto jaetaan soluihin (grids) muutamien avainmuuttujien avulla. Avainmuuttujina voivat olla kaikki lukupuista ja koealoista mitatut muuttujat, esimerkiksi puulaji, läpimitta, kasvupaikka ja maantieteellinen sijainti. Lukupuun koepuutunnus estimoidaan samaan soluun kuuluvien koepuiden koepuutunnusten keskiarvona. Menetelmällä on seuraavia

etuja:

1. koepuutunnusten luonnollinen vaihtelu säilyy lukupuiden vastaavissa estimaateissa (mikäli solut ovat pieniä) ja
2. eri koepuutunnusten välinen yhteisvaihtelu säilyy luonnollisena (mikäli solut ovat pieniä).

Osajulkaisussa IV on tutkittu ei-parametrasta lähimmän naapurin menetelmää, joka muistuttaa Holmin ym (1979) grid-menetelmää (Altman 1992). Lähimmän naapurin menetelmässä kullekin luku puulle etsitään k kappaletta koepuita, jotka ovat eniten samanlaisia kuin ko. luku puu. Puiden "etäisyyttä" (samanlaisuutta) mitataan etäisyysfunktioilla. Osajulkaisussa IV käytettiin Euklidista etäisyyttä. Etäisyysfunktion ulottuvuuksina (selittäjinä) olivat seuraavat tunnuksat: läpimitta, sijainti, puuston pohjapinta-ala, puuston ikä ja keskiläpimitta. Lähimmän naapurin menetelmän todettiin täyttävän hyvin koepuutunnusten yleistämismenetelmälle asetettavat vaatimukset.

2.4.3 Ennakkoinformaation käyttö (I, II, III)

Kun koepuutunnuksia tai niistä laskettuja tunnuksia yleistetään luku puille, käytettävissä on inventoinnin koepuuaineiston lisäksi usein muuta aineistoa, esimerkiksi edellisten inventointien koepuuaineistoja. Tällaisissa tilanteissa ennakkoinformaatiota voidaan hyödyntää esimerkiksi bayesilaisella estimoinnilla tai sekaestimoinnilla. Bayesilaisessa estimoinnissa ennakkoinformaationa on tuntemattoman parametrivektorin todennäköisyysjakauma. Esimerkiksi Green & Strawderman (1985) käyttivät bayesilaista estimointia puun tilavuusyhtälön laadinnassa. Tutkimuksessa yhtälön parametrien priorijakauma otettiin aiemmin julkaistuista saman puulajin tilavuusyhtälöistä. Tutkimuksessa osoitettiin, että ennakkotiedon käytöllä voidaan parantaa merkittävästi tilavuusyhtälön luotettavuutta tai vaihtoehtoisesti vähentää mitattavaa koepuun määrää. Käytännön sovelluksissa ongelmaksi muodostuu parametrien priorijakauman määrittäminen.

Sekaestimointi on käyttökelpoinen tilanteissa, joissa estimoitavan mallin parametrit vaihtelevat ajan tai jonkin (muun) luokittävän muuttujan suhteen. Mallin parametrit voidaan tulkita luokittaisiksi satunnaismuuttujiksi, jolloin päädytään satunnaisparametri- tai sekamalliin. Jos parametrit tulkitaan kiinteiksi ja niiden estimoitavaksi käytettävissä oleva aineisto stokastiseksi, voidaan esimerkiksi eri ajankohtien aineistoja käyttää parametrien estimoinnissa niin sanotulla sekaestimoinnilla. Lähestymistapa on otantateoreettinen (Teräsvirta 1981).

Sekaestimoinnin kuvaamiseksi merkitään:

$$y = \mathbf{X}\beta + \varepsilon \quad (4)$$

$$r = \mathbf{R}\beta + \phi \quad (5)$$

Yhtälö (4) on yleinen lineaarinen malli, jossa y on selitettävän muuttujan arvot sisältävä vektori, \mathbf{X} selittävien muuttujien arvot sisältävä matriisi, β on parametrivektori ja ε on satunnaisvaikutusten vektori.

Yhtälö (5) on stokastista ennakkoinformaatiota, missä r ja ϕ ovat satunnaisvaikutusten vektoreita ja \mathbf{R} on kiinteiden selittäjien matriisi.

Oletetaan, että

1. $\text{cov}(\varepsilon) = \sigma^2 \mathbf{I}$
2. $\text{cov}(\phi) = \kappa^{-1} \sigma^2 \mathbf{I}$
3. ε ja ϕ ovat riippumattomia.

Tällöin sekaestimaattori parametrivektorille β on:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X} + \kappa\mathbf{R}'\mathbf{R})^{-1} (\mathbf{X}'y + \kappa\mathbf{R}'r), \quad (6)$$

missä κ kuvastaa stokastisen ennakkoinformaation ja otosinformaation varianssien suhdetta.

Sekaestimaattorissa (6) ennakkoinformaation oletetaan olevan luonteeltaan stokastista (satunnaistekijänä ϕ). Julkaisussa II on tutkittu estimaattoria, jossa osa regressioyhtälön kertoimista on eliminoitu ennakkoinformaation avulla. Tässä lähestymistavassa eliminointuihin parametreihin liittyvä tieto oletetaan luonteeltaan ei-stokastiseksi eli tarkaksi (exact) (Theil & Goldberger 1961, sivut 72-73). Stokastista ja ei-stokastista informaatiota hyödyntävän estimaattorin kuvaamiseksi merkitään:

$$y = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon, \quad (7)$$

$$\mathbf{r} = \mathbf{R}_1\beta_1 + \mathbf{R}_2\beta_2 + \phi, \quad (8)$$

missä \mathbf{X}_1 ja \mathbf{R}_1 -matriisit sisältävät arvot sellaisille selittäville muuttujille, joiden parametrit estimoidaan vain ensimmäisen tason tiedolla (ei-stokastisella ennakkoinformaatiolla);

\mathbf{X}_2 ja \mathbf{R}_2 -matriisit sisältävät arvot sellaisille selittäville muuttujille, joiden parametri estimoidaan käyttäen sekä ensimmäisen että toisen tason tietoa (ensimmäisen tason tieto tulkitaan tältä osin stokastiseksi ennakkoinformaatioksi);

β_1 ja β_2 ovat parametrivektoreita ja

\mathbf{y} ja \mathbf{r} ovat selitettävän muuttujan arvot sisältäviä vektoreita ensimmäisen ja toisen tason aineistossa.

Olettaen, että β_1 on estimoitu virheettä ennakkoinformaation avulla, β_2 :n estimaattorin on (Korhonen 1992, II):

$$\hat{\beta}_2 = (\mathbf{R}_2'\mathbf{R}_2 + \kappa \mathbf{X}_2'\mathbf{X}_2)^{-1} (\mathbf{R}_2'\mathbf{u} + \kappa \mathbf{X}_2'\mathbf{v}), \quad (9)$$

missä

$$\mathbf{u} = \mathbf{r} - \mathbf{R}_1\beta_1,$$

$$\mathbf{v} = \mathbf{y} - \mathbf{X}_1\beta_1,$$

$\hat{\beta}_1$ = ennakkoinformaatiosta (tarkasti) estimoitu β_1 ja

κ kuvastaa ensimmäisen ja toisen tason tiedolle annettavien painojen suhdetta.

Korhonen (1992) sovelsi sekaestimaattoria (9) vanhan ja uuden koepuutiedon yhdistämiseen ylläpäimittämällin estimoisessa. Seitsemännen VMI:n koepuuaineistolla estimoitii puulajeittain ylläpäimittämällin kertoimet. Mallia sovellettiin kahdeksannen VMI:n koepuuaineistoon siten, että osa mallin parametreista estimoitii metsälautakunnittain uudelleen käyttäen sekä seitsemännen että kahdeksannen inventoinnin aineistoa. Tutkimus osoitti, että käyttämällä sekaestimointia koepuuaineistojen yhdistämiseen mallien luotettavuutta voidaan parantaa erityisesti kun koepuiden määrä on pieni (alle 100 havaintoa).

Osajulkaisussa II sekaestimointia on käytetty männyn tilavuusyhtälöiden estimoiseseen. Tutkimuksessa osoitetaan, että ennakkoinformaatiolla voidaan parantaa merkittävästi mallin luotettavuutta. Jos ennakkoinformaatio on eri populaatiosta kuin otosinformaatio, voi

sekaestimaattori olla hyvinkin harhainen. Harhaa voidaan pienentää huolellisella mallien tuotoilulla. Esimerkiksi tilavuusyhtälöiden laadinnassa on otettava huomioon maantieteellisen sijainnin vaikutus männyn runkomuotoon. Osajulkaisussa II tämä vaikutus mallitettiin toisen asteen trendipinnalla. Trendipinta estimoitiin seitsemännen VMI:n koko maan kattavasta aineistosta. Estimoitu trendipinta oletettiin luonteeltaan ei-satunnaiseksi (exact) ennakkoinformaatioksi ja tilavuusmallin muut kertoimet estimoitiin sekaestimoinnilla alueittain käyttäen sekä uuden että vanhan inventoinnin otosinformaatiota. Muiden kuin sijaintiin liittyvien selittäjien kohdalla ennakkoinformaatio oletettiin luonteeltaan satunnaiseksi. Tulosten mukaan toisen asteen trendipinnalla voitiin tyydyttävästi kuvata maantieteellisen sijainnin vaikutusta männyn runkomuotoon ja poistaa näin ennakkoinformaation harhaa.

Satunnaisparametrimalleja tai sekamalleja on käytetty menestyksellisesti alueellisesti luotettavien mallien estimointiin (vertaa local tariff tables) (Lappi 1986, Lappi & Bailey 1988, Ojansuu 1993, Gregoire & Schabenberger 1994, I). Satunnaisparametrimalli on yksinkertaisimmillaan varianssikomponenttimalli:

$$y_{ki} = \mu + b_k + e_{ki} \quad (10)$$

missä μ on kiinteä yleiskeskisarvo,

b_k on luokkakohtainen satunnaisvaikutus ja

e_{ki} satunnaisvirhe.

Malli (10) sopii tilanteisiin, joissa yleiskeskisarvo, luokittaisten keskiarvojen ja satunnaisvirheen varianssit voidaan estimoida suuresta edustavasta otoksesta. Kun nämä on tunnettu, sovellustilanteessa voidaan estimoida luokittaisia keskiarvoja ($\mu + b_k$) luotettavasti vähilläkin havainnoilla.

Yleinen esitys sekamallille on (Lappi 1993, sivu 133):

$$y_{ki} = \mathbf{x}_{ki}'\mathbf{b} + \mathbf{z}_{ki}'\mathbf{c}_k + e_{ki}, \quad (11)$$

missä indeksi k viittaa luokkaan (oletettu yksitasoinen luokitus),

indeksi i viittaa havaintoon,

\mathbf{x}_{ki} on kiinteitä parametreja vastaavien selittäjämuuttujien arvot luokan k havainnolle i ,

\mathbf{b} on kiinteiden selittäjien parametrivektori,

\mathbf{z}_{ki} on satunnaisparametreja vastaavien selittäjämuuttujien arvot luokan k havainnolle i ,

\mathbf{c}_k on satunnaisparametrien vektori, ja

e_{ki} satunnaisvirhe.

Seuraava esimerkki kuvaa sekamallien perusajatusta. Oletetaan, että yhtälön (11) satunnaisparametrit ovat metsikkökohtaisia. Jos selitettävän muuttujan (y) arvoja on mitattu muutamille metsikön k puille, voidaan metsikkökohtaiset satunnaisparametrit estimoida seuraavalla yhtälöllä (Lappi 1993, sivu 140).

$$\hat{\mathbf{c}}_k = [\mathbf{Z}_k'\mathbf{var}(\mathbf{e})^{-1}\mathbf{Z}_k + [\mathbf{var}(\mathbf{c})]^{-1}]^{-1} \mathbf{Z}_k'\mathbf{var}(\mathbf{e})^{-1}(\mathbf{y}_k - \mathbf{X}_k'\hat{\mathbf{b}}), \quad (12)$$

missä $\hat{\mathbf{b}}$ on kiinteitä parametreja vastaavien selittäjien parametrien GLS-estimaatit,

\mathbf{X}_k on kiinteitä parametreja vastaavien selittäjien arvot sisältävä matriisi,

\mathbf{Z}_k on satunnaisparametreja vastaavien selittäjien arvot sisältävä matriisi,

$\mathbf{var}(\mathbf{c})$ on satunnaisparametrien varianssi-kovarianssimatriisi ja

$\mathbf{var}(\mathbf{e})$ on satunnaisvirheiden varianssi-kovarianssimatriisi.

Kiinteät parametrit estimoidaan yleistetyn pienimmän neliösumman menetelmällä (GLS) käyttäen havaintojen varianssi-kovarianssimatriisina (Lappi 1993):

$$\mathbf{V} = \mathbf{var}(\mathbf{Zc} + \mathbf{e}).$$

Matriisi \mathbf{V} on lohkodeagonaalinen eli luokkien sisällä havainnot oletetaan korreloituneiksi, mutta luokkien välillä havaintojen oletetaan olevan korreloimattomia. Yhtälössä (12) ei oteta huomioon kiinteiden parametrien estimointivirhettä. Myös matriisin \mathbf{V} tilalla joudutaan käytännön sovelluksissa käyttämään estimaattia, jonka estimointivirhettä yhtälössä (12) ei ole otettu huomioon.

Estimaattoria (12) voidaan soveltaa, jos satunnaisparametrien kovarianssimatriisi ($\text{var}(\mathbf{c})$), $\hat{\mathbf{b}}$ -vektori ja satunnaisvirheen kovarianssimatriisi ($\text{var}(\mathbf{e})$) (tai edellisten estimaatit) tunnetaan. Tällöin metsikkökohtaiset satunnaisparametrit voidaan estimoida jo yhdellä havainnolla. Jos metsikkökohtaiset parametrit olisi oletettu kiinteiksi, niiden estimoimiseen tarvittaisiin vähintään yksi havainto kutakin parametria kohden. Lisäksi sekamalliestimaattori antaa vähillä havainnoilla stabiilimpia estimaatteja kuin esimerkiksi kiinteäparametriset regressiomallit, joita kalibroidaan kiinteillä korjauskertoimilla (Lappi 1993, sivu 140).

Yksi esimerkki sekamallien käytöstä on luvussa 2.1 mainitut Lapin (1986) ja osajulkaisun I runkokäyrämallit. Näissä malleissa rungon eri kohtien läpimittoja estimoidaan yhtälöillä, joissa on kiinteiden selittäjien lisäksi satunnainen metsikkötekijä ja satunnaisvirhe (satunnainen puukohtainen tekijä). Satunnaistekijöiden varianssit ja eri kulmien yhtälöiden satunnaistekijöiden väliset kovarianssit estimoitiiin laajasta edustavasta aineistosta. Näitä estimoituja varianssi-kovarianssimatriiseja hyödynnettiin sovellustilanteessa metsikkökohtaisen runkokäyrämallin estimoimisessa.

Osajulkaisussa I on esitetty tulokset runkokäyrämallin soveltamisesta VMI:ssa. Tulokset osoittavat, että sekamalleihin perustuvat runkokäyrämallit ovat tehokkaita metsikkökohtaisen runkomuodon ennustamisessa (kuvat 2, 3 ja 4 osajulkaisussa I). Tutkimuksessa ilmeni myös, että yhtälöt eivät ole riittävän joustavia, kun niissä on vain yksi metsikkökohtainen satunnaisparametri (vakio). Jotta malliin voitaisiin lisätä useita metsikkökohtaisia parametreja, olisi parametrien varianssi-kovarianssimatriisien estimoimiseksi oltava laaja ja edustava aineisto, jossa on useita havaintoja jokaisesta aineiston metsiköstä.

Lappi (1991) on esittänyt pituusmallin, jossa on useita satunnaisparametreja. Mallissa pituutta estimoivan yhtälön vakio ja läpimitta-selittäjään liittyvät kertoimet oli estimoitu metsikkökohtaisina satunnaismuuttujina. Mallien todettiin tarkentuvan metsikkökohtaisiksi selvästi paremmin kuin kiinteäparametristen regressiomallien, etenkin kun pituushavaintoja oli vähän (esimerkiksi alle 10 metsikköä kohden) (Lappi 1991). Lapin tutkimuksessa (1991) esitettiin myös menetelmä tilavuusyhtälöiden kalibroimiseksi metsiköittäin. Menetelmä perustui pituus- ja tilavuusmallin jäännösvirheiden kovarianssien hyödyntämiseen.

Kangas & Korhonen (1995) tutkivat semiparametristen mallien käyttöä koepututiedon yleistämisessä. Tutkimuksessa männyn tilavuutta estimoitiiin mallilla, jossa oli "tavanomainen"

parametrinen osa ja ei-parametrinen osa. Ei-parametrinen osa kuvasi maantieteellisen sijainnin vaikutusta parametrisen mallin residuaaliin. Tulokset osoittivat, että semiparametrisilla menetelmillä voidaan mallittaa maantieteellisen sijainnin vaikutusta männyn runkomuotoon. Parametrinen sekamalli antoi puun tilavuudelle luotettavampia estimaatteja kuin semiparametrinen malli. Sekamallin soveltaminen tutkimuksessa esitetyllä tavalla olisi kuitenkin vaatinut yhden koepuun mittaamista jokaiselta koealalta. Tämän vuoksi tutkimuksessa päädyttiin suosittelemaan semiparametrista mallia VMI:n tapaisissa sovelluksissa, joissa koealat ovat pieniä.

3. PÄÄTELMÄT

Suuralueen inventoinnin otanta-, mittaus- ja laskentamenetelmien valinnassa asetetaan yleensä suuri paino menetelmän harhattomuudelle. Kun inventoinnin tehtävä on seurata esimerkiksi puuvarojen määrän muutosta, vaatimus menetelmien harhattomuudelle on perusteltu. Kun inventointia käytetään strategisen suunnittelun lähtötiedon tuottamiseen, tulosten harhattomuuden suuri painoarvo on perusteltua, jos suunnitteen laskentamenetelmä on esimerkiksi perinteinen ikäluokkajakaumiin ja ikäluokittaisiin tilavuuksiin perustuva tavoitehakuulaskelma (Kuusela & Nyssönen 1962). Kun inventointiaineistoa käytetään metsien tulevan kehityksen simuloimiseen (esimerkiksi suunnitelaskelmissa), on annettava suuri paino myös puu- ja koelakkohtaisten estimaattien luotettavuudelle.

Ranneby & Svensson (1991) asettivat seuraavat vaatimukset koepuutunnusten yleistämisessä käytettävälle menetelmälle.

1. Eri laskentaositteiden hehtaariohittaiset tilavuusestimaatit ja kokonaistilavuusestimaatit ovat harhattomia.
2. Yksittäisten koelalojen estimaatit ovat harhattomia ja koelalojen välinen vaihtelu on todenmukainen.
3. Lukupuiden estimaatit ovat harhattomia ja koelalojen sisäinen koepuutunnusten vaihtelu on todenmukainen.
4. Eri koepuutunnusten välinen yhteisvaihtelu on puutasolla todenmukainen. Esimerkiksi puun tilavuuden ja tilavuuskasvun estimaateista johdettu suhteellinen tilavuuskasvu ei saa olla epärealistinen.
5. Yleistämismenetelmä on helppo toteuttaa.

Osajulkaisussa I esitetyt sekamallitekniikkaan perustuvat runkokäyrämallit eivät sellaisenaan täytä harhattomuusvaatimuksia. Mallien koelakkohtainen tarkentuminen edellyttäisi useampien metsikkökohtaisten satunnaisparametrien käyttöä. Mallitettujen läpimittojen yhteisvaihtelu on realistinen.

Osajulkaisuissa II ja III käytetyt kiinteäparametriset tilavuus- ja kasvumallit eivät ota huomioon aineiston jakautumista koeloihin. Puittaiset ja koeloihittaiset estimaatit ovat harhattomia. Estimaattien varianssi on kuitenkin parametrisilla malleilla aina pienempi kuin estimoitavan muuttujan todellinen varianssi. Tilavuuden ja tilavuuskasvun yhteisvaihtelua ei ole malleissa

otettu huomioon. Osajulkaisuissa II ja III käytetyt menetelmät soveltuvat siten tilanteisiin, joissa inventointitiedoista halutaan laskea harhattomia ja luotettavia tilastoja.

Osajulkaisuissa IV käytetty ei-parametrinen lähimmän naapurin menetelmä mahdollistaa estimoitavien muuttujien alkuperäisen vaihtelun säilyttämisen. Siten menetelmää voidaan suositella erityisesti tilanteisiin, joissa inventointiaineistoa käytetään lähtöaineistona puiden tulevan kehityksen simuloinnissa.

Satunnaisparametreja sisältävät mallit ja sekaestimointi hyödyntävät tehokkaasti ennakkoinformaatiota (I,II). Kaikissa tilanteissa ei ole saatavilla riittävän luotettavaa ennakkoinformaatiota. Esimerkiksi puun kasvun vuotuinen vaihtelu on niin suurta, ettei vanhojen inventointien koepuutietoja ole järkevää käyttää läpimitan kasvun yleistämisessä lukupuille (katso esimerkiksi Henttonen 1990). Tällöin esimerkiksi osajulkaisuissa III käytetyt tavanomaiset regressiomallit ovat perusteltuja.

Osajulkaisuissa II on oletettu, että maantieteellisen sijainnin vaikutus on estimoitu virheettä laajasta valtakunnallisesta aineistosta. Sijaintiin liittyvien selittäjien parametreja käsitellään tarkkana (exact) ennakkoinformaationa. Vaikka ennakkoinformaationa käytetty aineisto on suuri ja alueellisesti edustava, oletusta ei voi pitää totena. Saadut tulokset kuitenkin osoittavat, että maantieteelliseen sijaintiin liittyvät parametriestimaatit ovat riittävän luotettavia sovellettavaksi tutkimuksessa esitetyllä tavalla.

KIRJALLISUUS

- Altman, N.S. 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Estimation. *The American Statistician*. Vol 46. No. 3. Sivut 175-185.
- Burk, T.E. & Ek, A.R. 1982. Application of empirical Bayes/James-Stein procedures to simultaneous estimation problems in forest inventory. *Forest Science* 28:753-771.
- Cajanus, W. 1911. Puunrunnon muotoa koskevia tutkimusmetodeja. Referate: Über zahlenmäßige Darstellung der Stammformen der Waldbäume. *Suomen metsänhoitoyhdistyksen julkaisuja* 26(6&7):1-10.
- Cunia, T. 1986a. Error of forest inventory estimates: its main components. In: *Estimating Tree Biomass Regressions and Their Error. Proceedings of the Workshop on Tree Biomass Regression Functions and Their Contribution to the Error of Forest Inventory Estimates*. May 26-30, 1986. Syracuse, New York. Sivut 1-13.
- Cunia, T. 1986b. An optimization model to calculate the number of trees and plots. In: *Estimating Tree Biomass Regressions and Their Error. Proceedings of the Workshop on Tree Biomass Regression Functions and Their Contribution to the Error of Forest Inventory Estimates*. May 26-30, 1986. Syracuse, New York. Sivut 15-25.
- Cunia, T. 1986c. Construction of tree biomass tables by linear regression techniques. In: *Estimating Tree Biomass Regressions and Their Error. Proceedings of the Workshop on Tree Biomass Regression Functions and Their Contribution to the Error of Forest Inventory Estimates*. May 26-30, 1986. Syracuse, New York. Sivut 27-36.
- Cunia, T. 1986d. Use of dummy variables techniques in the estimation of biomass regressions. In: *Estimating Tree Biomass Regressions and Their Error. Proceedings of the Workshop on Tree Biomass Regression Functions and Their Contribution to the Error of Forest Inventory Estimates*. May 26-30, 1986. Syracuse, New York. Sivut 37-48.
- Cunia, T. 1991. Main objectives and basic characteristics of national forest inventories. In: *Forest inventories in Europe with special reference to statistical methods. Proceedings of the International IUFRO S.4.02 and S.6.04 Symposium, May 14-16.1990 Birmensdorf, Switzerland*. Published by Swiss Federal Institute for Forest, Snow and Landscape Research. CH-8903 Birmensdorf, Switzerland. Pages 27-33.
- Green, E.J. & Strawderman, W.E. 1985. The use of Bayes/empirical Bayes estimation in individual tree volume equation development. *Forest Science* 31(4):975-990.
- Gregoire, T.G. & Schabenberg, O. 1994. Fitting bole-volume equations to spatially correlated within-tree data. Manuscript. College of Forestry and Wildlife Resources. Virginia Polytechnic Institute and State University Blacksburg, VA. USA. 13 s.
- Henttonen, H. 1990. Kuusen rinnankorkeusläpimitan kasvun vaihtelu Etelä-Suomessa. Summary: Variation in the diameter growth of Norway spruce in Southern Finland. *Helsingin yliopisto. Metsänarvioimistieteen laitos. Tiedonantoja* N:o 25. Helsinki. 88 s.

- Holm, S., Hägglund, B. & Mårtensson, A. 1979. En metod för generalisering av Riksskogstaxeringens provträdsdata. Summary: A method for generalization of sample tree data from the Swedish National Forest Survey. Swedish University of Agricultural Sciences. Department of Forest Survey. Report No. 25. 94 s.
- Ilvessalo, Y. 1943. Suomen metsävarat ja metsien tila, II valtakunnan metsien arviointi. *Communicationes Instituti Forestalis Fenniae* 30. 446 s.
- Jonsson, B., Jacobsson, J. & Kallur, H. 1993. The Forest Management Package. Theory and application. *Studia Forestalia Suecica* No. 189. Swedish University of Agricultural Sciences. Uppsala, Sweden. 56 s.
- Kangas, A. 1991. Updated measurement data as prior information in forest inventory. *Silva Fennica* Vol. 25(3): 181-191.
- Kangas, A., Kangas, J., Korhonen, K.T., Maltamo, M. & Päivinen, R. 1990. Metsää kuvaavat mallit. *Silva Carelica* 17. Joensuun Yliopisto. Metsätieteellinen tiedekunta. 143 s.
- Kangas, A. & Päivinen, R. 1994. Metsän mittaus. *Silva Carelica* 27. Joensuun Yliopisto. Metsätieteellinen tiedekunta. 172 s.
- Kangas, A. & Korhonen, K.T. 1995. Generalizing sample tree information with semiparametric and parametric models. *Silva Fennica* [in print]. 14 pp.
- Kilkki, P. 1979. An outline for a data processing system in forest mensuration. *Silva Fennica* 13(4):368-384.
- Kilkki, P. 1983. Sample trees in timber volume estimation. *Acta Forestalia Fennica* 182. 35 s.
- Kilkki, P., Saramäki, M. & Varmola, M. 1978. A simultaneous equation model to determine taper curve. *Silva Fennica* 12(2):120-125.
- Kilkki, P. & Siitonen, M. 1975. Metsikön puuston simulointimenetelmä ja simuloituun aineistoon perustuvien puustotunnusmallien laskenta. Summary: Simulation of artificial stands and derivation of growing stock models from this material. *Acta Forestalia Fennica* 145. 33 s.
- Kilkki, P. & Varmola, M. 1981. Taper curve models for Scots pine and their application. *Acta Forestalia Fennica* 174. 60 s.
- Korhonen, K.T. 1992. Calibration of upper diameter model in large scale forest inventory. *Silva Fennica* 26(4):231-239.
- Kuusela, K. 1965. A method for estimating the volume and taper curve of tree stem and preparing volume functions and tables. *Communicationes Instituti Forestalis Fenniae* 60(2).
- Kuusela, K. & Nyysönen, A. 1962. Tavoitehakuulaskelma. Summary: The cutting buget for a desirable growing stock. *Acta Forestalia Fennica* 74(6). 36 s.

- Kuusela, K. & Salminen, S. 1991. Suomen metsävarat 1977-1984 ja niiden kehittyminen 1952-1980. *Acta Forestalia Fennica* 220. 84 s.
- Kujala, M. 1980. Runkopuun kuorellisen tilavuuskasvun laskentamenetelmä. Summary: A calculation method for measuring the volume growth over bark of stemwood. *Folia Forestalia* 441. 8 s.
- Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae* 108. 74 s.
- Lahtinen, A. & Laasasenaho, J. 1979. On the construction of taper curves by using spline functions. *Communicationes Instituti Forestalis Fenniae* 95(8). 63 s.
- Lappi, J. 1986. Mixed linear models for analyzing and predicting stem form variation of Scots pine. *Communicationes Instituti Forestalis Fenniae* 134. The Finnish Forest Research Institute. Helsinki, Finland. 69 s.
- Lappi, J. 1991. Calibration of height and volume equations with random parameters. *Forest Science* 37(3):781-801.
- Lappi, J. 1993. Metsäbiometrian menetelmiä. *Silva Carelica* 24. Joensuun Yliopisto. Metsätieteellinen tiedekunta. 190 s.
- Lappi, J. & Bailey, R.L. 1988. A height prediction model with random stand and tree parameters: an alternative to traditional site index methods. *Forest Science* 34:907-927.
- Liedes, M. & Manninen, P. 1975. Otantamenetelmät. Oy Gaudeamus Ab. Helsinki. 255 s.
- Loetsh, F., Zöhrer, F. & Haller, K.E. 1973. Forest Inventory. Volume II. BLV Verlagsgesellschaft. München. 469 s.
- Meng, C.H., Tang, S.Z. & Burk, T.E. 1990. A stochastic restrictions regression model approach to volume equation estimation. *Forest Science* 36(1):54-65.
- Ojansuu, R. 1993. Prediction of Scots pine increment using a multivariate variance component model. *Acta Forestalia Fennica* 239. 72 s.
- Pekkonen, T. 1983. Leimikon puuston tilavuuden arviointi regressioennustinta käyttäen. Metsäntutkimuslaitoksen tiedonantoja 86. 63 s.
- Pelz, D.R. 1991. National forest inventory systems in Europe. Teoksessa: Forest inventories in Europe with special reference to statistical methods. Proceedings of the International IUFRO S.4.02 and S.6.04 Symposium, May 14-16.1990 Birmensdorf, Switzerland. Published by Swiss Federal Institute for Forest, Snow and Landscape Research. CH-8903 Birmensdorf, Switzerland. Sivut 59-65.
- Poso, S. 1973. A method for combining photo and field sample in forest inventory. *Communicationes Instituti Forestalis Fenniae* 76(1). 133 s.

- Poso, S., Karlsson, M., Pekkonen, T. & Härmä, P. 1990. A system for combining data from remote sensing, maps and field measurements for forest planning purposes. University of Helsinki. Department of forest mensuration and management. Research Notes 23. 40 s.
- Päivinen, R. 1987. Metsän inventoinnin suunnittelumalli. Summary: A planning model for forest inventory. Joensuun Yliopiston luonnontieteellisiä julkaisuja 11. 179 s.
- Ranneby, B. & Svensson, S.A. 1991. From sample tree data to images of tree populations. Teoksessa: Forest inventories in Europe with special reference to statistical methods. Proceedings of the International IUFRO S.4.02 and S.6.04 Symposium, May 14-16.1990 Birmensdorf, Switzerland. Published by Swiss Federal Institute for Forest, Snow and Landscape Research. CH-8903 Birmensdorf, Switzerland. Sivut 27-33.
- Salminen, S. 1978. Increment calculations on the basis of relascope sampling in the Finnish national forest inventory. IUFRO Meeting in June 18.- 25. Bucharest. Konekirjoite. Metsäntutkimuslaitos, Helsinki. 7 s.
- Siitonen, M. 1993. Experiences in the use of forest management planning models. *Silva Fennica* 27(2):167-178.
- Sterba, H. 1990. Stem curves - a review of the literature. *Forest Products Abstracts* Vol. 3 No. 4:69-73.
- Svensson, S.A. 1988. Estimation of annual stem volume increment. Swedish University of Agricultural Sciences. Department of Forest Survey. Report 46. 140 s.
- Teräsvirta, T. 1981. Some results on improving the least squares estimation of linear models by mixed estimation. *Scandinavian Journal of Statistics* 8:33-38.
- Theil, H. & Goldberger, A.S. 1961. On pure and mixed statistical estimation in economics. *International Economic Review* 2:65-78.
- Tokola, T. 1990. Satelliittikuvan ja VMI-koealatiedon käyttö metsätalousalueen puuston inventoinnissa. Metsätalouden suunnittelun lisensiaattitutkimus. Joensuun Yliopisto. Metsätieteellinen tiedekunta. 56 s.
- Tomppo, E. 1992. Multi-source national forest inventory of Finland. In: Nyssönen, A., Poso, S. & Rautala, J. (ed.) 1992. Proceedings of Ilvessalo symposium on national forest inventories. The Finnish Forest Research Institute. Research Papers 444. Sivut 52-60.

LUETTELO OSAJULKAISUJEN VIRHEISTÄ

OSAJULKAISU I. "Sekamalliteknikalla laadittujen runkokäyrämallien käyttö metsäninventoinnissa"

Sivu 3, selite: $RMSE = \text{keskivirhe} = b^2 + sd^2$ tulisi olla: $RMSE = \text{keskivirhe} = \sqrt{(b^2 + sd^2)}$

Sivu 6, Kaava k_1

$= \sum^n / \sum^n \hat{h}$ pitäisi olla: $k_1 = \sum^n h / \sum^n \hat{h}$

OSAJULKAISU II. "Mixed estimation in calibration of volume functions of Scots pine".

Sivu 271, Yhtälö (1) pitäisi olla:

$y = X\beta + \varepsilon$, where $\varepsilon = \text{vector of residual}$

Sivu 271, Yhtälöt (3) ja (4) tulisi olla:

$r = R_1\beta_1 + R_2\beta_2 + \varepsilon$

$s = S_1\beta_1 + S_2\beta_2 + \varepsilon$

Sivu 272, Yhtälöiden (7) ja (8) merkinnät ovat väärinpäin.

Sivu 272, Yhtälö (7): d_2 pitäisi olla d^2 .

Sivu 272, Yhtälö (7): RDIST pitäisi olla:

$RDIST = \begin{cases} 0, & \text{if } DIST > 20 \\ 1/(DIST+0.2), & \text{otherwise} \end{cases}$

OSAJULKAISU III. "Calculation system for large-scale forest inventory".

Sivu 12, Yhtälö (1) tulisi olla:

$\ln(d_6) = a_0 + a_1*d^2 + a_2*h^2 + a_3*t + a_4*d/t + a_5*\ln(\hat{d}_{6L}) +$

$a_6*\ln(G) + a_7*YC + a_8*YC2 + a_9*XC + a_{10}*XC2 +$

$+ a_{11}*YC*XC$

Sivu 27, lähdeluettelosta puuttuvat:

Lappi, J. & Bailey, R. 1988. A height prediction model with random stand and tree parameters: an alternative to traditional site index methods. *Forest Science* 34:907-927.

Searle, S.R. 1971. *Linear models*. John Wiley, New York. 532 pp.

I

Kari T. Korhonen

SEKAMALLITEKNIKALLA LAADITTUJEN RUNKOKÄYRÄMALLIEN
KÄYTTÖ METSÄNINVENTOINNISSA

Using taper curve models based on mixed linear models
in forest inventory

Approved on 27.8.1991

SISÄLLYS

1. JOHDANTO	3
11. Tutkimuksen tausta	3
12. Tutkimustehtävä	4
2. AINEISTO	4
3. MENETELMÄT	5
31. Puun runkokäyrän mallittaminen	5
32. Koepuiden käytön tutkiminen	5
4. RUNKOKÄYRÄMALLIT MÄNNYLLE, KUUSELLE JA KOIVULLE	6
41. Lasketut parametrit ja kovarianssimatriisit	6
42. Parametrien ja kovarianssimatriisien tulkinta	6
43. Runkokäyrämallien käyttö	7
431. Pääkomponenttianalyysin soveltaminen	7
432. Peruskulmien läpimittojen estimointi	8
44. Harhattomuuskorjaukset	8
5. RUNKOKÄYRÄMALLIEN LUOTETTAVUUS	8
51. Estimaattien luotettavuuden kriteerit	8
52. Läpimittaestimaattien luotettavuus	9
53. Tilavuusestimaattien luotettavuus	9
6. PÄÄKOMPONENTTIEN ENNUSTAMINEN	13
7. TULOKSET	13
71. Puittaiset tilavuusestimaatit	13
72. Keskitilavuusestimaatit	14
8. TULOSTEN TARKASTELU	16
81. Runkokäyrämalleissa havaitut puutteet	16
82. Tutkimuksen suuntaaminen jatkossa	18
KIRJALLISUUS	19
SUMMARY	19
LIITTEET	23

Korhonen, K. T. 1991. Sekamallitekniikalla laadittujen runkokäyrämallien käyttö metsäninventoinnissa. Summary: Using taper curve models based on mixed linear models in forest inventory. *Folia Forestalia* 774. 27 p.

Tutkimuksessa esitetään männylle, kuuselle ja koivulle runkokäyrämallit, jotka koostuvat sekamallitekniikalla laadituista eri korkeuksien läpimittoja estimoivista yhtälöistä. Runkokäyrämallien antamien läpimitta- ja tilavuusestimaattien luotettavuutta testattiin olettaen puista mitatuksi eri tunnuksia. Mallien havaittiin antavan tutkituissa tilanteissa yhtä luotettavia tilavuusestimaatteja kuin nykyisin käytössä olevat mallit. Laadittujen mallien merkittävimmät edut ovat:

- 1) mallien käyttö ei edellytä puun pituuden tuntemista,
- 2) miltä tahansa korkeudelta mitattua läpimittaa voidaan käyttää hyväksi runkokäyrän estimoimisessa,
- 3) malleihin on sisäänrakennettuna menetelmä koepuista mitatun tiedon siirtämiseen lukupuille koelakokohaisesti ja
- 4) mallien antamien läpimittaestimaattien keskivirhe voidaan estimoida.

Laadittuja runkokäyrämalleja testattiin myös valtakunnan metsien inventoinnin koepuilla. Tutkimuksen tulos oli, että sekamallitekniikalla laadituilla runkokäyrämalleilla voidaan estimoida yksittäisten puiden tilavuudet selvästi luotettavammin kuin kiinteäparametrisella tilavuusmallilla. Runkokäyrämalleilla lasketut keskitilavuusestimaatit olivat kuitenkin useissa tapauksissa selvästi harhaisia.

Keywords: forest inventory, mixed linear models, taper curve models.
FDC 52--015.5

Author's address: Finnish Forest Research Institute, Joensuu Research Station, P.O. Box 68, SF-80101 Joensuu, Finland.

Taper curve models for pine, spruce and birch were estimated using mixed linear models. The taper curve models were tested assuming different dimensions measured. Results showed e.g. that when the diameter at breast height and height are measured the root mean square error of the volume estimate is about equal to the rmse obtained with volume functions. If height is measured only for few sample trees in some cases the volume estimates for tally trees are significantly biased.

ISBN 951-40-1162-7

ISSN 0015-5543

Helsinki 1991. Valtion painatuskeskus

1. Johdanto

11. Tutkimuksen tausta

Puun rungon tilavuus on hankalasti mitattava tunnus. Tämän vuoksi metsänarvioimisen keskeinen tutkimusongelma on ollut rungon tilavuutta estimoivien mallien laadinta. Etenkin suomalaisissa tutkimuksissa on rungon tilavuutta lähdetty usein mallittamaan runkokäyrämallien avulla. Kun puun runkokäyräestimaatti tunnetaan, voidaan rungon minkä tahansa osan tilavuus laskea analyyttisellä tai numeerisella integroinnilla.

Puun tilavuutta tai runkokäyrää kuvaavissa malleissa lähtökohtana käytetään yleisesti rinnankorkeusläpimittaa. Inventoinneissa sovelletaan yleisesti kaksivaiheisia otantamenetelmiä, joissa ensimmäisen vaiheen otoksella estimoidaan puuston runkolukusarja mittaamalla puiden rinnankorkeusläpimittoja. Toisen vaiheen otoksessa mitataan puista myös muita rungon dimensioita kuvaavia tunnuksia kuin rinnankorkeusläpimittaa. Näiksi niin sanotuiksi koepuutunnuksiksi ovat Suomessa vakiintuneet pituus ja läpimitta kuuden metrin korkeudella.

Inventointitulosten laskemiseksi tarvitaan malleja, joilla mitatuista tunnuksista voidaan esti-

moida puuston tilavuus puutavaralajeittain. Mallien on voitava käyttää hyväksi koepuista mitattavaa informaatiota puiden runkomuodosta. Mikäli mallit käyttävät myös ennakkoinformaatiota oikealla painolla, voidaan päästä luotettavampiin tuloksiin kuin käytettäessä joko pelkästään koepuumittaustietoa tai pelkästään mallien sisältämää ennakkoinformaatiota.

Suomen valtakunnan metsien inventoinnin (VMI) tulosten laskentamenetelmässä on painotettu koepuumittauksia enemmän kuin ennakkoinformaatiota. VMI:ssä lasketaan koepuille puutavaralajeittaiset tilavuudet Laasasenahon (1982) runkokäyrämalleilla. Muun muassa metsälautakunnan, puulajin, veroluokan ja puuluokan määrittämissä ositteissa lasketaan läpimittaluokittain tilavuustunnusten keskiarvot. Koepuille lasketut tilavuustunnukset siirretään ositteittain ja läpimittaluokittain lukupuille, joita summaamalla saadaan lasketuksi aluekohtaisia tilavuustunnuksia.

Pystymittatulosten (PMP) laskentajärjestelmässä on pyritty minimoimaan koepuumittausten määrää ennakkoinformaatiota käyttämällä. PMP-järjestelmissä leimikosta oletetaan mitatuksi joukko koepuita, joista tunnetaan rinnankorkeus-

Merkinnät – Notation

d = läpimitta rinnankorkeudella
 d = diameter at breast height

d_6 = läpimitta kuuden metrin korkeudella maasta tai kannosta
 d_6 = diameter at the height of six meters

h = puun pituus syntypisteestä latvaan
 h = height

v = rungon kuorellinen tilavuus syntypisteestä latvaan
 v = stem volume

$D(u)$ = läpimitta kulmassa u
 $D(u)$ = diameter at angle u

$d(u)$ = kulman u läpimitan luonnollinen logaritmi
 $d(u)$ = logarithmic diameter at angle u

b = estimoitu harha = todellisen ja estimoidun läpimitan tai tilavuuden erotuksen keskiarvo
 b = estimated bias = mean difference between real and estimated diameter or volume

sd = jäännösvirhe = todellisen ja estimoidun läpimitan tai tilavuuden erotuksen keskihajonta
 sd = residual error = the standard deviation of difference between real and estimated diameter of volume

sb = estimaatin residuaalin koelakohtainen komponentti
 sb = the plotwise component of the residual error

sw = estimaatin residuaalin puukohtainen komponentti
 sw = the treewise component of the residual error

RMSE = keskiarvo = $b^2 + sd^2$
RMSE = root mean square error = $b^2 + sd^2$

läpimitta ja pituus sekä mahdollisesti yläläpimitta. Koepuille lasketaan tilavuudet Laasasenahon (1982) runkokäyrämalleilla. Lukupuiden tilavuudet lasketaan polynomimallilla, jonka kertoimet ratkaistaan leimikon koepuualuekohtaisesti. Kertoimien laskennassa mitatut koepuut muodostavat kukin yhden havainnon. Lisäksi käytetään ennakkoinformaationa keinotekoisia havaintoja, jotka ovat kattava otos leimikon läpimittajakaumasta. "Keksityille" havainnoille lasketaan tilavuudet yleisillä tilavuusmalleilla. Ennakkoinformaatio ja osininformaation painojen suhde määräytyy puiden tilavuuden leimikoiden välisen ja sisäisen varianssin suhteen mukaan (Pekkonen 1983).

Tässä tutkimuksessa esitetään puuston tilavuuden laskentajärjestelmä, jossa ennako- ja otosinformaation yhdistämiseen liittyvät ongelmat on pyritty ratkaisemaan lineaarisiin sekamalleihin perustuvilla runkokäyrillä.

12. Tutkimustehtävä

Tutkimuksen tehtävä on tutkia runkokäyrämallien ja koepuiden käyttöä suuralueen inventoinnin tilavuustunnusten laskennassa. Tutkimushypoteesi on, että käyttämällä satunnaisparametrimalleja voidaan yksittäisen puun ja puujoukon

tilavuustunnukset estimoida luotettavasti vähemmillä mittaauksilla kuin käyttämällä kiinteäparametrisiä malleja. Hypoteesia testataan empiirisesti simuloimalla valtakunnan metsien inventoinnin aineistolla koepuutoksia ja vertaamalla erilaisia malleja käytävillä laskentamenetelmillä saatujen estimaattien luotettavuuksia.

Tutkimuksessa laaditaan tilavuuden laskentajärjestelmä, joka sisältää menetelmän olemassa olevien mallien ja inventoinnissa mitattavien koepuiden mittaustietojen yhdistämiseksi. Työn ensimmäisessä vaiheessa laaditaan runkokäyrämallit männylle, kuuselle ja koivulle. Toisessa vaiheessa tutkitaan laadittujen mallien soveltuvuutta inventointitulosten laskentaan. Runkokäyrämallien laadinnassa käytettävät menetelmät on kehittänyt Lappi (1986). Tässä työssä esitetään menetelmän pääpiirteet ja tutkitaan empiirisesti menetelmän toimivuutta.

Tämä tutkimusraportti on tiivistetty ja osin korjattu vedos Joensuun Yliopiston metsätieteellisen tiedekunnan metsätalouden suunnittelun lisensiaattitutkintoa varten tehdystä työstä. Erityisesti työn ohjaajana toiminut edesmennyt professori Pekka Kilkki sekä VTT Juha Lappi ja FK Jaakko Heinonen ovat ohjeillaan mahdollistaneet työn valmistumisen. Keskustelut MMK Risto Ojansuun sekä lisensiaattityön tarkistajien, MMT Risto Päivisen ja MMT Jouko Laasasenahon kanssa ovat olleet suureksi avuksi.

2. Aineisto

Tutkimuksessa laadittujen runkokäyrämallien aineistona käytettiin Metsäntutkimuslaitoksen metsänarvioimisen tutkimusosaston mittauttamia koepuita (Laasasenaho 1982). Vuosina 1968–1970 kerätty aineisto koostuu 2326 männystä, 1864 kuusesta ja 863 koivusta. Koska malleja laadittaessa analysoitiin koealojen sisäistä runkomuodon vaihtelua, aineistosta hylättiin ne koepuut, jotka olivat ainoat puulajinsa edustajat koealalla. Jäljelle jäi 2034 mäntyä, 1667 kuusta ja 602 koivua. Koealojen lukumäärät puulajeittain vastaavassa järjestyksessä olivat 593, 487 ja 200.

Koepuuaineisto on kerätty valtakunnan metsien viiden inventoinnin metsä- ja kitumaan koealoilta. Koealojen valinnassa on pyritty saamaan tasainen otos koko Suomesta. Koepuut on valittu relaskoopilla (kerroin 2). Kultakin koealalta on mitattu korkeintaan viisi puuta. (Laasasenaho 1982).

Aineistoon kuuluvista puista on mitattu läpimitat rinnankorkeudelta ja kuuden metrin korkeudelta sekä seuraavilta suhteellisilta korkeuksilta: 1, 2,5, 5, 7,5, 10, 15, 20, 30, 40, 50, 60, 70, 80 ja 90 prosenttia puun pituudesta.

Lisäksi puista tunnettiin muun muassa kannonkorkeus, pituus, latvusrajan korkeus ja ikä. Läpimitat on mitattu pystypuista kiipeämällä. Koepuille oli laskettu tilavuudet kannosta latvaan splinifunktioiden avulla (Laasasenaho 1982).

Koepuiden käyttöä tutkittiin paitsi edellä kuvatussa mallien laadinta-aineistossa myös valtakunnan metsien kahdeksannessa inventoinnissa (VMI8) kerätyllä aineistolla. VMI8:n aineistosta poimittiin ne koepuukoelat, jotka olivat Lounais-Suomen, Uudenmaan-Hämeen tai Pirikka-Hämeen metsälautakuntien alueella. Aineisto on kerätty vuosina 1986–1987. (Valtakunnan metsien... 1986).

VMI8:n koepuukoeloilta on kuviokohtaisina muuttujina mitattu joukko kasvupaikkaa ja puustoa kuvaavia tunnuksia. Relaskoopilla valituista koepuista on puukohtaisina tunnuksina mitattu muun muassa d , d_6 ja h (Valtakunnan metsien... 1986). Runkokäyrämallien käyttöä tutkittaessa aineistosta poistettiin katkenneet ja kuolleet sekä alle 3 m:n pituiset puut. Aineistoon jäi 3061 mäntyä 623 koealalla, 4088 kuusta 634 koealalla ja 1019 koivua 379 koealalla.

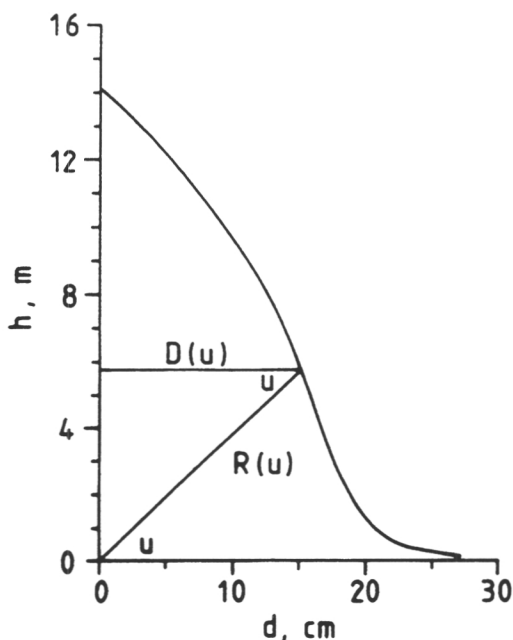
3. Menetelmät

31. Puun runkomuodon mallittaminen

Tutkimuksessa laadittiin runkokäyrämallit Lapin (1986) esittämällä menetelmällä. Menetelmässä puun runkomuoto kuvataan napakoordinaatistossa, jossa runkokäyrä ilmoitetaan tiettyjä kulmia vastaavina läpimittoina tai säteiden pituuksina (Kilki 1983, Lappi 1986). Napakoordinaatioesitystä on havainnollistettu kuvassa 1.

Puun runkokäyrä kuvattiin laatimalla 13 yhtälöä, joista 12 ennusti läpimitan eri kulmissa (kutsutaan myöhemmin peruskulmiksi) ja 13:s ennusti puun pituuden. Peruskulmina käytettiin seuraavia kulmia: 0,25, 0,7, 1,5, 3, 5, 8, 14, 21, 31, 41, 56 ja 72 astetta (Lappi 1986). Myöhemmin tässä kirjoituksessa 13:n peruskulman läpimitalla tarkoitetaan puun pituutta.

Puun läpimittojen oletettiin riippuvan puun koosta ja muodosta. Varianssien vakioimiseksi käytettiin läpimittoja ennustavissa yhtälöissä logaritmita läpimittaa ja pituutta ennustavassa yhtälössä pituuden logaritmia. Läpimittayhtälöiden selittäjiksi valittiin puun koko ja koon toinen potenssi. Puun koko määritettiin eri korkeuksilta mitattujen logaritmiten läpimittojen ja pituuden keskiarvona. (Lappi 1986).



Kuva 1. Puun runkomuodon kuvaaminen napakoordinaatistossa. Lähde: Lappi 1986.
Figure 1. Polar coordinate system (Lappi 1986).

Runkokäyrämallia laadittaessa oletettiin, että samassa metsikössä kasvavat puut muistuttavat runkomuodoltaan toisiaan. Tämän vuoksi läpimittayhtälöihin sisällytettiin satunnainen metsikköparametri, jolloin yhtälöt saatiin muotoon:

$$d_k(u) = a_0(u) + a_1(u)s_{ki} + a_2(u)s_{ki}^2 + v_k(u) + e_{ki}(u), \quad (31.1)$$

missä $d_k(u)$ = koealan k puun i peruskulman u läpimitan logaritmi,

s_{ki} = koealan k puun i koko,
 $a_0(u)$, $a_1(u)$ ja $a_2(u)$ ovat kertoimia,
 $v_k(u)$ on satunnainen koealavaikutus ja
 $e_{ki}(u)$ on satunnainen puuvaikutus.

Malliin liittyvät oletukset olivat:

- 1) v -termien odotusarvo on nolla ja varianssi vakio,
- 2) e -termien odotusarvo on nolla ja varianssi vakio sekä
- 3) v - ja e -termit eivät korreloi keskenään.

Yhtälössä (31.1) olevan satunnaisen metsikkötekijän avulla voidaan ottaa huomioon se, että läpimittayhtälöiden kiinteän osan residuaalit korreloivat samassa metsikössä kasvavien puiden kesken. Lisäksi eri kulmien v -termit korreloivat keskenään, samoin eri kulmien e -termit keskenään (Lappi 1986).

Runkokäyrämallin käyttämiseksi oli estimoitava kiinteät parametrit $a_0(u)$, $a_1(u)$ ja $a_2(u)$ ($u = 1, \dots, 13$); v -termien varianssit eri kulmissa; v -termien kovarianssit eri kulmien välillä; e -termin varianssit eri kulmissa sekä e -termien kovarianssit eri kulmien välillä. Kiinteät parametrit estimoitiin yleistetyllä pienimmän neliösumman menetelmällä (GLS – generalized least squares). v - ja e -termien varianssit ja kovarianssit estimoitiin Lapin (1986) esittämällä menetelmällä.

Läpimittayhtälöillä saadaan puun läpimitta 12:lla eri korkeudella sekä puun pituus. Näistä voidaan laskea puun minkä tahansa kohdan läpimitta tai minkä tahansa osan tilavuus splinifunktioiden avulla. Tässä vaiheessa sovellettiin Lahtisen & Laasasenahon (1979) tutkimuksiin perustuvia, erityisesti runkokäyrän tasoittamiseksi laadittuja splinifunktioita.

32. Koeuiden käytön tutkiminen

Tutkimuksessa selvitettiin koeuiden lukumäärän vaikutusta toisaalta puittaiten tilavuusestimaattien ja toisaalta suuralueen keskitilavuusestimaattien luotettavuuteen. Koeputietojen siirtämisessä lukuvoille vertailtiin tässä tutkimuksessa laadittuja runkokäyrämalleja ja suoraan tila-

vuutta ennustavia regressiomalleja. Vertailtavien menetelmien toimivuuden tutkimiseksi tehtiin VMI-aineistossa koepuuantotaja viidellä eri koepuumäärällä. Koepuiksi valittiin joka 15., 7., 5., 3. tai joka 2. puu. Kullakin koepuumäärällä otanta toistettiin 100 kertaa valiten koepuut satunnaisesti. Sadasta toistosta laskettiin lukupuiden tilavuusestimaattien keskineliövirhe sekä koko laskenta-alueen keskitilavuusestimaatin ja todellisen keskitilavuuden poikkeaman keskiarvo ja keskihajonta.

Laskennassa puiden todelliseksi tilavuudeksi oletettiin puusta mitattuja rinnankorkeus- ja yläläpimittoja sekä pituuksia käyttäen lasketut tilavuudet. Puittaisten tilavuusestimaattien ja keskitilavuusestimaattien luotettavuuksia laskettaessa pidettiin mukana myös koepuuna käytetyt puut, mutta olettaen tällöin koepuustakin tunnetuksi vain rinnankorkeusläpimitta.

Vertailumenetelmänä käytettyjä tilavuusmalleja laadittaessa laskettiin aluksi koepuiksi valittujen puiden tilavuudet koepuutunnusten avulla. Seuraavaksi laadittiin koepuista saadulla aineistolla mallit, joissa puun tilavuutta ennustettiin pelkän rinnankorkeusläpimitan ja puulajin avulla. Mallit laadittiin ositteittain. Kun mitattuja koepuita oli paljon, tehokkaimmaksi menettelyksi osoittautui koepuiden osittaminen metsälautakunnan ja veroluokan mukaan. Jos koepuita oli mitattu vähän, osittaminen pelkän metsälautakunnan mukaan oli tehokkainta.

Tilavuutta ennustavan mallin muodoksi valittiin polynomi:

$$v/d^2 = ad^2 + bd + c, \quad (32.1)$$

missä a ja b ovat estimoitavia kertoimia.

Mallin (32.1) parametrit estimoititiin joka simulointikierröksellä (eli jokaisesta otoksesta). Tilavuuden ennustaminen polynomilla (32.1) on erittäin epäluotettavaa puul-

le, jonka läpimitta on suurempi kuin polynomin kertoimien laskennassa käytettyjen suurimpien puiden läpimitat. Tästä syystä koepuita suuremmille puille ennustettiin ensin pituusmallilla pituus ja sitten Laasasenahon (1982) kahden tunnuksen tilavuusyhtälöllä tilavuus. Pituusmallina käytettiin yhtälöä (Henttonen 1989):

$$\ln(h - 1,3) = a_0 + a_1/(d+5) + a_2/(d+5)^2. \quad (32.2)$$

Yhtälön (32.2) kertoimet ($a_0 - a_2$) laskettiin puulajeittain käyttäen VMI8:ssa mitattuja kuuden eteläisimmän metsälautakunnan koepuita. Kertoimet on esitetty liitteessä 5. Estimoidaessa yhtälöllä (32.2) pituutta logarimuunnoksesta aiheutuva harha otettiin huomioon lisäämällä korjaustermi $s^2/2$, missä s^2 on mallin keskineliövirhe.

Sekä laadittu pituusmalli että Laasasenahon (1982) tilavuusmalli kalibroitiin ensin aineistoon kussakin inventointisimuloinnissa mitatuilla koepuilla. Pituusmallin kalibrointi tehtiin laskemalla koepuiden avulla puulajeittain korjaustermit:

$$k_1 = \bar{\Sigma} / \bar{\Sigma} \hat{h}$$

missä $\bar{\Sigma}$ tarkoittaa summausta yli koepuiden ja \hat{h} on mallilla (32.2) estimoitu pituus.

Korjaustermit laskettiin erikseen samoille ositteille, joita käytettiin tilavuusyhtälön (32.1) laskennassa. Korjaustermin laskennan jälkeen lukupuiksi oletettujen puiden pituudet saatiin yhtälöllä:

$$h = k_1 * \hat{h}, \quad (32.3)$$

missä \hat{h} on mallilla (32.2) saatu pituusestimaatti.

Laasasenahon (1982) tilavuusyhtälö kalibroitiin aineistoon samalla tavoin kuin pituusmalli.

4. Runkokäyrämallit männylle, kuuselle ja koivulle

41. Lasketut parametrit ja kovarianssimatriisit

Sekä männylle, kuuselle että koivulle eri kohtien läpimittoja ennustavana yhtälönä mallia (31.1). Parametrit laskettiin kullekin puulajille erikseen. Hies- ja rauduskoivua käsiteltiin laskennassa yhdessä. Taulukossa 1 on esitetty estimaatit kiinteille kertoimille a_0 , a_1 ja a_2 .

Mallin (31.1) satunnaisvaikutusten (v ja e) kovarianssit eri kulmien välillä ja varianssit eri kulmissa on esitetty korrelaatiomatriisien muodossa liitteissä 1, 2 ja 3. Matriiseissa diagonaaleilla ovat koela- tai puukohtaisen satun-

naistekijän keskihajonnat ja diagonaalien yläpuolella ovat satunnaistekijöiden korrelaatiot eri kulmien välillä.

42. Parametrien ja kovarianssimatriisien tulkinta

Kaikilla puulajeilla koon (s) kerroin on positiivinen kaikissa kulmissa. Koon toisen potenssin kerroin on positiivinen yhdeksänteen peruskulmaan saakka joten rungon tyviosissa läpimitta kasvaa kiihtyen koon kasvaessa.

Yhdeksännestä kulmasta alkaen koon toisen

Taulukko 1. Lämpimittayhtälön (31.1) kiinteät parametrit eri puulajeilla.
 Table 1. Fixed parameters of the diameter function (31.1) for pine, spruce and birch.

Kulma Angle	Mänty - Pine			Kuusi - Spruce			Koivu - Birch		
	a ₀	a ₁	a ₂	a ₀	a ₁	a ₂	a ₀	a ₁	a ₂
1	.526	.971	.023	.821	.757	.061	.478	1.000	.015
2	.515	.912	.032	.725	.755	.057	.486	.917	.027
3	.461	.869	.041	.616	.757	.055	.506	.793	.053
4	.301	.945	.025	.479	.821	.040	.438	.771	.060
5	.288	.914	.031	.490	.781	.048	.298	.872	.037
6	.283	.870	.040	.429	.803	.043	.200	.917	.028
7	.220	.874	.037	.285	.880	.026	.107	.950	.022
8	.123	.934	.019	.125	.968	.007	-.017	1.016	.009
9	-.036	1.045	-.011	-.073	1.089	-.021	-.124	1.086	-.010
10	-.094	1.101	-.033	-.230	1.189	-.044	-.176	1.158	-.035
11	-.146	1.166	-.058	-.391	1.301	-.068	-.141	1.214	-.064
12	-.126	1.202	-.076	-.467	1.395	-.090	-.018	1.244	-.087
13	-.062	1.238	-.089	-.534	1.542	-.123	.372	1.148	-.084

potenssin kerroin on negatiivinen kaikilla puulajeilla. Tämä merkitsee sitä, että kun koko kasvaa riittävän suureksi, läpimittaestimaatti alkaa pienentyä koon kasvaessa. Kulminaatiopiste, josta alkaen läpimittaestimaatit pienevät, voidaan laskea derivoimalla ko. läpimittayhtälö koon suhteen ja etsimällä derivaatan nollakohta. Esimerkiksi männyn kulman pituutta ennustava yhtälön derivaatan nollakohdaksi saadaan $s = 6,96$. Koska aineiston puissa koon vaihtelualue on 0,5:stä 3,5:een, ei kulminaatiopistettä todellisuudessa saavuteta koskaan.

Mallin jäännösvirheiden korrelaatiomatriisien (liitteet 1, 2 ja 3) diagonaalilla olevat jäännöshajonnat voidaan tulkita läpimittayhtälöiden suhteellisiksi keskivirheiksi, koska yhtälöt ovat logaritmisia.

Jäännösvirheiden korrelaatiomatriisit osoittavat myös, että eri kulmien satunnaisvirheet korreloivat voimakkaasti. Vierekkäisten kulmien välinen korrelaatio on yleensä positiivinen. Peruskulman 8 yläpuolisten kulmien yhtälöiden virheet korreloivat yleensä negatiivisesti kahdeksatta peruskulmaa alempien yhtälöiden satunnaisvirheiden kanssa.

43. Runkokäyrämallien käyttö

43.1. Pääkomponenttianalyysin soveltaminen

Runkokäyrämallia laadittaessa laskettiin koelakohtaisen satunnaisvektorin varianssi peruskulmissa. Liitteistä 1-3 nähdään, että eri kulmien varianssikomponentit korreloivat selvästi. Tämä antoi perusteen pääkomponenttianalyysin sovel-

tamiseen koelajojen välisen vaihtelun ilmaisemisessa.

Pääkomponentit laskettiin eri kulmien koelakohtaisen satunnaisvaikutuksen kovarianssimatriisista. Koska ko. matriisin aste on 13, siitä voidaan muodostaa 13 pääkomponenttia. Neljään ensimmäiseen pääkomponenttiin liittyvät ominaisvektorit ja -arvot on esitetty liitteessä 4.

Koska neljä ensimmäistä pääkomponenttia selittävät kaikilla puulajeilla yli 99 % eri kulmien koelakohtaisten satunnaisvaikutusten varianssien summasta, yhtälön (31.1) $v(u)$ -termit voidaan ilmaista kovarianssimatriisista laskettujen neljän ensimmäisen ominaisvektorin avulla (Lappi 1986). Tällöin yhtälö (31.1) saadaan muotoon:

$$d_{ki}(u) = a_0(u) + a_1(u)s_{ki} + a_2(u)s_{ki}^2 + \sum_{p=1}^4 q_p(u)c_p + e_{ki}(u), \quad u = 1, \dots, 13, \quad (43.1)$$

missä $q_p(u)$ = p:nnen ominaisvektorin kulman u alkio
 c_p = p:s pääkomponentti

Pääkomponenttien käytöstä saavutettava etu on, että sovellustilanteessa kunkin kulman koelavaikutus saadaan lasketuksi ominaisvektorien kertoimien avulla, kun ensin on estimoitu neljä pääkomponenttia. Ellei pääkomponentteja käytettäisi, jouduttaisiin ennustamaan kunkin kulman v -tekijä erikseen, mikä hidastaisi laskentaa, sillä laskenta edellyttäisi operaatioita 13×13 -ulotteisella satunnaisvaikutusten kovarianssimatriisilla.

Pääkomponenttianalyysin avulla voidaan keventää laskentaa. Lisäksi pääkomponentteja voi-

daan käyttää jännösvirheiden vaihtelun analysoimiseen. Ensimmäinen ominaisvektori on lautaunut voimakkaimmin rungon latva- ja tyviosien läpimitoille, joten ensimmäisen pääkomponentin arvojen vaihtelu kuvaa rungon solakkuuden vaihtelua (ks. Lappi 1986, s. 19). Toinen pääkomponentti kuvaa rungon keskiosan läpimittojen vaihtelua.

432. Peruskulmien läpimittojen ennustaminen

Läpimittoja ennustavassa yhtälössä (43.1) ovat selittäjinä puun koko ja neljä pääkomponenttia sekä näihin liittyvät ominaisvektorit. Sovellustilanteessa ominaisvektorit ja kertoimet a_0 , a_1 ja a_2 ovat tunnettuja, mutta puun kokoa ja pääkomponentteja ei tunneta. Ensimmäinen vaihe runkokäyrämallien käytössä on näiden tuntemattomien parametrien estimoiminen tai ennustaminen puusta mitattujen läpimittojen avulla. Koska pääkomponentit ovat koealakohtaisia, niiden arvot ennustetaan koealoittain.

Puiden kokoparametrien estimoimiseksi ja pääkomponenttien ennustamiseksi koealan puista mitatuista läpimitoista ja niitä estimoivista yhtälöistä muodostetaan yhtälöryhmä. Yhtälöryhmässä tunnettuja tekijöitä ovat mitatut läpimitat ja läpimittayhtälöiden kertoimet. Lisäksi tunnetaan koealakohtaista satunnaisvaikutusta kuvaavien pääkomponenttien varianssit (= ominaisvektorien ominaisarvot). Yhtälöryhmän tuntematto-

mina tekijöinä olevat kiinteät kokoparametrit ja satunnaiset pääkomponentit voidaan ratkaista yleisillä sekamallien estimoimisessa käytetyillä menetelmillä (ks. luku 31). Yhtälöryhmien muodostaminen on kuvattu yksityiskohtaisesti Lapin (1986) tutkimuksessa.

44. Harhattomuuskorjaukset

Läpimittayhtälöt ennustavat läpimittojen luonnollisia logaritmeja. Siirryttäessä aritmeettisiin yksiköihin on estimaatteihin tehtävä harhattomuuskorjaus. Jos läpimitan logaritminen ennuste oletetaan normaalijakautuneeksi, on harhaton läpimitan ennustin (Lappi 1986):

$$E(D(u)) = \text{Exp}[\hat{d}(u) + 1/2(\text{var}(d(u)) - \text{var}(\hat{d}(u)))]. \quad (44.1)$$

Käytettäessä runkokäyrämallia tilavuuden estimoimiseen on saatava harhattomat läpimitan toisen potenssin estimaatit. Nämä saadaan yhtälöllä (Lappi 1986):

$$E(D^2(u)) = \text{Exp}[2\hat{d} + (\text{var}(d(u)) - \text{var}(\hat{d}(u)))]. \quad (44.2)$$

Harhattomuuskorjauksia laskettaessa tarvitaan kulman todellisen ja ennustetun läpimitan varianssi. Todellisen läpimitan varianssi saadaan summaamalla läpimitan varianssikomponentit. Läpimitan ennusteen varianssi saadaan Lapin (1986, s. 62) esittämällä tavalla.

5. Runkokäyrämallien luotettavuus

51. Estimaattien luotettavuuden kriteerit

Runkokäyrämallien antamien läpimittaestimaattien luotettavuuden selvittämiseksi tutkittiin logaritmissen ja aritmeettisten estimaattien poikkeamia todellisista arvoista. Poikkeamat esitettiin keskiarvoina ja keskihajontoina muuttujas-
ta:

$$e = d(u) - \hat{d}(u),$$

missä $d(u)$ = todellinen logaritminen tai aritmeettinen läpimitta ja
 $\hat{d}(u)$ = estimoitu logaritminen tai aritmeettinen läpimitta.

Todellisen ja ennustetun läpimitan erotuksen varianssille laskettiin metsikkökohtainen komponentti (sb) ja puukohtainen komponentti (sd)

Searlen (1971, s. 474) esittämällä kaavoilla.

Runkokäyrämallin tulisi antaa mitattujen läpimittojen suhteen harhattomia tilavuusestimaatteja aritmeettisissa mittayksiköissä. Koska tässä tutkimuksessa laadituissa runkokäyrämalleissa selittäjänä ei ole läpimitta, vaan puun koko, eivät saadut tilavuusestimaatit ole kuitenkaan harhattomia mitattujen läpimittojen suhteen, vaan puun koon suhteen (Lappi 1986).

Koska läpimittayhtälöt on laadittu sekamallitekniikalla, eivät mallien luotettavuuden mittarina yleensä käytetyt suhteelliset virheet ($e_r = (v - \hat{v}) / \hat{v}$) anna oikeaa kuvaa mallin toimivuudesta. Oikea tapa laskea suhteelliset virheet on käyttää kaavaa:

$$e_r = (v - \hat{v}) / E(v),$$

missä $E(v)$ tarkoittaa tilavuuden odotusarvoa koon suhteen.

Koska $E(v)$ on koon määritelmästä johtuen kiinteässä yhteydessä puun todellisen tilavuuden kanssa (ks. Lappi 1986 s. 11), voidaan $E(v)$ korvata puun todellisella tilavuudella.

Tilavuusestimaattien luotettavuuden selvittämiseksi tulostettiin keskiarvot ja keskihajonnat sekä absoluuttisista että suhteellisista poikkeamista. Suhteelliset poikkeamat laskettiin käyttäen jakajana sekä todellista että estimoitua tilavuutta. Tuloksissa luotettavuudet on ilmaistu mainitulla kolmella kriteerillä seuraavassa järjestyksessä:

- 1) $(v - \hat{v}) / \hat{v}$
- 2) $(v - \hat{v}) / v$
- 3) $v - \hat{v}$

Kriteeri 1 kuvaa mallin harhattomuutta mitattujen tunnusten suhteen. Kriteeri 2 kuvaa mallin ominaisuuksia sekamalliteorian pohjalta eli kertoo mallin estimoimisen onnistumisesta. Myös tilavuusestimaattien jäännösvirheet jaettiin koeala- ja puukohtaiseen komponenttiin.

52. Läpimittaestimaattien luotettavuus

Läpimittaestimaattien virheitä tutkittiin seuraavissa tilanteissa:

- 1) puista tunnettiin vain d ,
- 2) puista tunnettiin d ja h .

Sekä logaritmisten että aritmeettisten läpimittaestimaattien luotettavuustunnukset on esitetty taulukoissa 2 ja 3. Läpimittaestimaattien harhat ovat useimmiten hyvin lähellä nolaa. Kun vain d on tunnettu, on männyn pituusestimaatin harha tosin 32 cm; koivulla ja kuusella 'pituusmallin' harha on huomattavasti pienempi.

Logaritmisten estimaattien jäännösvirheet voidaan tulkita aritmeettisten estimaattien suhteellisten jäännösvirheiden likiarvoiksi. Näin saadaan esimerkiksi pituusestimaattien jäännösvirheeksi männyllä 20,8, kuusella 15,5 ja koivulla 16,1 %. Puulajien välinen vertailu osoittaa, että d :n ollessa ainoa mitattu muuttuja pituusestimaattien ja rungon latvaosien läpimittaestimaattien jäännösvirheet ovat suurimmat männyllä ja pienimmät kuusella. Tyviosien läpimittaestimaattien jäännösvirheet ovat männyllä pienempiä kuin kuusella tai koivulla.

Taulukosta 2 havaitaan, että kun puusta tunnetaan vain d , rungon tyviosien läpimittaestimaattien puukohtainen virhekomponentti on suurempi kuin koealakohtainen. Latvaosien läpimittaesti-

maateilla suhde on päinvastainen. Tulos on seurausta siitä, että tyviosissa puun runkomuoto on epäsäännöllinen.

Pituuden tunteminen parantaa odotetusti rungon latvaosan läpimittaestimaattien luotettavuutta, mutta vaikuttaa vain vähän tyviosien läpimittaestimaattien luotettavuuteen.

53. Tilavuusestimaattien luotettavuus

Taulukossa 4 on runkokäyrämalleilla laskettujen tilavuuksien luotettavuus seuraavissa tilanteissa:

- 1) puista tunnetaan vain d ,
- 2) puista tunnetaan d ja h ,
- 3) puista tunnetaan d , d_c ja h sekä
- 4) puista tunnetaan d ja puulajeittain yhdestä koealan puusta lisäksi d_c ja h .

Luotettavuuden kriteereinä ovat keskiarvot, keskihajonnat ja koealojen väliset sekä sisäiset hajonnat muuttujille:

- 1) $(v - \hat{v}) / \hat{v}$,
- 2) $(v - \hat{v}) / v$ ja
- 3) $v - \hat{v}$

Kun vain d on tunnettu, on tilavuusestimaatin absoluuttinen harha etenkin männyllä huomattavan suuri. Suhteellisten harhojen suuruus riippuu siitä, käytetäänkö jakajana todellista vai estimoitua tilavuutta — kriteerin $(v - \hat{v})/v$ mukaan harhat ovat varsin vähäisiä, mutta kriteerin $(v - \hat{v})/\hat{v}$ mukaan selvästi suurempia. Kun puusta tunnetaan pituus, ovat harhat niin pieniä, ettei suhteellisten virheiden laskentatavalla ole kovin suurta merkitystä.

Kun ainoana mitattuna läpimittana on d , on tilavuusestimaatin koealakohtaisen satunnaisvirheen osuus jäännösvirheestä puulajista riippuen 67–80 %. Kun myös pituus on tunnettu, kasvaa jäännösvirheen puukohtaisen komponentin osuus. Ilmiö on seurausta siitä, että pelkästään d :n tunteminen kertoo paljon puun koosta. Tämän vuoksi huomattava osa läpimittojen puukohtaisesta vaihtelusta tulee selitetyksi pelkän d :n avulla. Kun d on kiinteä, pituus kuvaa lähinnä puun runkomuotoa, jonka vaihtelusta huomattava osa on metsiköiden välistä vaihtelua. Siksi pituuden mittaaminen vaikuttaa tilavuusestimaattien jäännösvirheen koealakohtaiseen komponenttiin enemmän kuin puukohtaiseen komponenttiin.

Taulukosta 4 nähdään, että yhden koepuun mittaaminen koealalta pienentää tilavuusestimaattien jäännösvirhettä 15–20 % verrattu-

na tilanteeseen, jossa koepuita ei ole mitattu. Jäännösvirheen pieneneminen on seurausta jäännösvirheen koelakohtaisen komponentin pie-

nenemisestä — koelakohtainen jäännösvirhe pienenee kaikilla puulajeilla noin 50 % yhden koepuun mittaamisen myötä.

Taulukko 2. Eri kulmien logaritmisten ja aritmeettisten läpimittaestimaattien luotettavuus, kun d mitattu. Logaritmiin estimaatteihin liittyvät tunnuksat on kerrottu 100:lla, aritmeettisissä estimaateissa yksikkönä on cm paitsi pituudella m (ks. Merkinnät s. 3).

Table 2. Reliability of logarithmic and arithmetic diameter estimates at different angles when d is measured. Biases and standard errors concerning logarithmic estimates has been multiplied by 100. For arithmetic diameters biases and standard errors are in centimeters except for height in meters (see Notation, page 3).

Kulma Angle	Logaritmiset - Logarithmic				Aritmeettiset - Arithmetic			
	b	sd	sb	sw	b	sd	sb	sw
Mänty - Pine								
1	-0.03	7.37	3.61	6.42	0.09	1.86	0.91	1.62
2	-0.03	5.56	3.01	4.68	0.05	1.11	0.55	0.97
3	-0.04	4.46	2.53	3.68	0.04	0.68	0.32	0.60
4	-0.03	3.26	1.87	2.67	0.04	0.40	0.18	0.36
5	-0.03	3.00	1.50	2.59	0.06	0.51	0.20	0.47
6	-0.04	3.24	1.21	3.01	0.08	0.64	0.21	0.60
7	-0.03	4.60	2.56	3.82	0.12	0.83	0.48	0.68
8	-0.02	6.76	4.90	4.66	0.15	1.09	0.79	0.76
9	-0.02	10.37	8.31	6.22	0.17	1.39	1.11	0.84
10	-0.02	13.62	11.42	7.42	0.19	1.46	1.23	0.79
11	-0.01	17.05	14.69	8.67	0.16	1.25	1.09	0.61
12	-0.01	19.19	16.61	9.62	0.09	0.75	0.66	0.36
13	-0.01	20.83	18.18	10.20	0.32	2.80	2.48	1.32
Kuusi - Spruce								
1	0.00	9.47	5.28	7.87	-0.01	2.28	1.06	2.02
2	0.00	6.24	3.40	5.24	0.00	1.22	0.54	1.09
3	0.00	4.61	2.40	3.94	0.00	0.73	0.31	0.66
4	0.00	3.18	1.64	2.72	0.00	0.41	0.15	0.38
5	0.00	2.81	1.21	2.54	0.02	0.46	0.09	0.45
6	0.00	3.20	1.49	2.83	0.03	0.60	0.24	0.55
7	0.00	4.73	3.20	3.51	0.05	0.83	0.54	0.64
8	0.00	6.83	5.21	4.42	0.06	1.07	0.79	0.71
9	0.00	9.33	7.42	5.66	0.07	1.19	0.94	0.73
10	0.00	11.20	8.92	6.78	0.07	1.16	0.93	0.69
11	0.00	13.13	10.31	8.14	0.05	0.94	0.75	0.56
12	0.00	14.49	11.12	9.29	0.03	0.57	0.45	0.35
13	0.00	15.51	11.88	9.98	0.09	2.12	1.67	1.30
Koivu - Birch								
1	-0.05	9.86	5.44	8.23	-0.09	2.28	1.13	1.98
2	-0.07	6.75	3.44	5.82	-0.05	1.25	0.56	1.12
3	-0.03	4.85	2.13	4.36	-0.02	0.76	0.25	0.71
4	-0.09	3.54	1.50	3.21	-0.04	0.55	0.09	0.54
5	-0.09	3.36	1.36	3.07	-0.03	0.69	0.27	0.64
6	-0.01	3.73	1.52	3.41	-0.01	0.79	0.32	0.72
7	0.03	4.77	1.81	4.42	0.02	0.92	0.32	0.87
8	0.26	6.05	2.85	5.34	0.06	1.04	0.47	0.92
9	0.32	8.12	4.70	6.64	0.07	1.11	0.60	0.93
10	-0.03	10.23	6.69	7.75	0.04	1.12	0.72	0.86
11	-0.19	12.72	9.47	8.50	0.04	0.95	0.70	0.65
12	-0.02	14.78	11.48	9.34	0.03	0.61	0.47	0.38
13	-0.21	16.14	13.05	9.54	0.15	2.50	2.02	1.48

Taulukko 3. Logaritmistien ja aritmeettisten läpimittaestimaattien luotettavuus, kun puusta on mitattu d ja h .

Table 3. Reliability of logarithmic and arithmetic diameter estimates when d and h are measured.

Kulma Angle	Logaritmiset – Logarithmic				Aritmeettiset – Arithmetic			
	b	sd	sb	sw	b	sd	sb	sw
<i>Mänty – Pine</i>								
1	0.18	7.29	3.28	6.51	-0.11	1.83	0.85	1.62
2	0.20	5.46	2.79	4.70	-0.05	1.11	0.54	0.97
3	0.21	4.25	2.19	3.65	-0.04	0.66	0.29	0.59
4	0.20	3.04	1.56	2.61	-0.03	0.40	0.17	0.36
5	0.17	2.85	1.21	2.58	-0.04	0.52	0.22	0.47
6	0.15	3.10	1.21	2.85	-0.03	0.64	0.27	0.58
7	0.11	3.74	1.76	3.30	-0.03	0.69	0.35	0.59
8	-0.01	4.01	2.01	3.47	-0.01	0.64	0.33	0.55
9	0.03	4.20	2.14	3.62	0.01	0.55	0.29	0.47
10	0.04	4.15	2.28	3.47	-0.01	0.43	0.24	0.36
11	0.02	3.37	1.96	2.74	-0.01	0.24	0.15	0.19
12	0.01	2.16	1.15	1.82	0.00	0.08	0.05	0.07
<i>Kuusi – Spruce</i>								
1	0.05	9.33	4.86	7.97	0.00	2.27	1.01	2.03
2	0.05	6.13	3.15	5.27	0.00	1.21	0.53	1.08
3	0.05	4.47	2.01	3.40	0.00	0.72	0.29	0.66
4	0.04	3.09	1.30	2.80	0.00	0.40	0.13	0.38
5	0.04	2.74	0.85	2.60	-0.02	0.43	0.05	0.43
6	0.04	2.84	0.84	2.72	-0.02	0.51	0.12	0.50
7	0.04	3.21	1.36	2.91	-0.02	0.55	0.22	0.50
8	0.04	3.67	1.90	3.14	-0.01	0.55	0.26	0.49
9	0.04	4.13	2.44	3.33	-0.02	0.49	0.26	0.42
10	0.03	4.21	2.60	3.32	-0.02	0.41	0.22	0.34
11	0.02	3.82	2.41	2.96	-0.01	0.25	0.14	0.21
12	0.01	2.76	1.87	2.03	0.00	0.10	0.06	0.08
<i>Koivu – Birch</i>								
1	-0.15	9.75	5.17	8.28	0.06	2.26	1.06	2.00
2	-0.16	6.75	3.36	5.85	0.05	1.25	0.55	1.13
3	-0.13	4.87	2.06	4.41	0.03	0.75	0.24	0.71
4	-0.20	3.50	1.37	3.23	0.04	0.54	0.04	0.54
5	-0.20	3.40	1.38	3.11	0.04	0.67	0.26	0.62
6	-0.13	3.58	1.47	3.27	0.03	0.75	0.31	0.68
7	-0.06	4.18	1.36	3.95	0.02	0.81	0.26	0.76
8	0.19	4.60	1.57	4.32	0.00	0.79	0.29	0.74
9	0.32	4.77	1.83	4.41	0.01	0.67	0.24	0.62
10	0.02	4.67	2.19	4.12	0.03	0.52	0.22	0.47
11	-0.07	4.25	2.44	3.48	0.02	0.32	0.17	0.27
12	-0.03	3.19	1.75	2.66	0.01	0.12	0.07	0.10

Taulukko 4. Tilavuusestimaattien suhteelliset ja absoluuttiset virheet, kun puusta on mitattu eri tunnuksia. Suhteelliset virheet ovat prosentteina, absoluuttiset kuutiodesimetreinä.

Table 4. Relative and absolute reliability of volume estimates with different measurements. Relative figures are in %, absolute in dm^3 .

		Mänty – Pine							
		b	sd	sb	sw	b	sd	sb	sw
		d mitattu – d measured				d, h mitattu – d, h measured			
$(v - \hat{v}) / \hat{v}$		3.12	19.75	16.12	11.42	0.85	7.29	3.63	6.33
$(v - \hat{v}) / v$		-0.72	20.38	16.39	12.14	0.32	7.26	3.59	6.32
$v - \hat{v}$		15.17	74.59	59.29	45.31	0.87	30.67	15.17	26.67
		d, d _s ja h mitattu d, d _s and h measured				d ja l koepuu mitattu d and l sample tree measured			
$(v - \hat{v}) / \hat{v}$		0.33	3.50	1.53	3.15	1.10	14.07	8.49	11.23
$(v - \hat{v}) / v$		0.45	3.56	1.53	3.21	-0.90	14.85	9.14	11.71
$v - \hat{v}$		1.34	18.91	8.96	16.65	2.32	51.01	26.69	43.48
		Kuusi – Spruce							
		b	sd	sb	sw	b	sd	sb	sw
		d mitattu – d measured				d, h mitattu – d, h measured			
$(v - \hat{v}) / \hat{v}$		2.51	18.70	14.35	12.00	0.53	7.40	3.97	6.25
$(v - \hat{v}) / v$		-0.96	19.31	14.87	12.33	-0.01	7.39	4.09	6.16
$v - \hat{v}$		6.64	70.90	47.54	52.63	0.38	28.82	9.31	27.28
		d, d _s ja h mitattu d, d _s and h measured				d ja l koepuu mitattu d and l sample tree measured			
$(v - \hat{v}) / \hat{v}$		-0.32	3.23	1.23	2.98	1.50	14.82	8.65	12.04
$(v - \hat{v}) / v$		-0.43	3.28	1.27	3.03	-0.66	15.03	8.55	12.36
$v - \hat{v}$		-1.59	17.32	4.18	16.81	-1.64	58.04	24.47	52.64
		Koivu – Birch							
		b	sd	sb	sw	b	sd	sb	sw
		d mitattu – d measured				d, h mitattu – d, h measured			
$(v - \hat{v}) / \hat{v}$		2.47	19.65	13.27	14.50	0.50	8.69	3.71	7.86
$(v - \hat{v}) / v$		-1.44	20.79	13.67	15.67	-0.31	9.53	3.86	8.72
$v - \hat{v}$		0.67	74.11	24.03	70.11	-5.84	52.60	21.96	47.81
		d, d _s ja h mitattu d, d _s and h measured				d ja l koepuu mitattu d and l sample tree measured			
$(v - \hat{v}) / \hat{v}$		-0.36	5.38	2.08	4.96	2.30	15.73	8.57	13.20
$(v - \hat{v}) / v$		-0.64	5.31	2.01	4.92	-0.38	17.77	8.40	15.66
$v - \hat{v}$		-3.03	1.25	11.29	18.01	-4.05	79.36	31.11	73.01

6. Pääkomponenttien ennustaminen

Tutkimuksessa laaditut runkokäyrämallit käytävät koepuumittauksia koelakohtaisen satunnaisvaikutuksen ennustamiseen. Kuten luvussa 431 on kerrottu, koelakohtainen satunnaisvaikutus kuvataan neljän pääkomponentin avulla. Jos koelalta ei ole mitattu yhtään koepuuta, ei pääkomponentteja voida ratkaista ja lukupuiden runkokäyrä joudutaan estimoimaan käyttäen selittäjänä ainoastaan rinnankorkeusläpimittaa.

Laadittujen runkokäyrämallien soveltuvuutta koepuutietojen siirtämiseen lukupuille tutkittiin VMI8:n koepuilla. Jos inventointitulosten laskennassa käytettävä menetelmä edellyttää, että jokaiselta koelalta on mitattu vähintään yksi koepuu kustakin koelalan puulajista, joudutaan suuralueen inventoinnissa mittaamaan hyvin suuri määrä koepuita. Tämän vuoksi käytettiin Kilkin & Lapin (1987) esittämää menetelmä, jonka avulla runkokäyrämallit voivat hyödyntää koepuista mitattua tietoa puun runkomuodosta myös sellaisilla koelaloilla, joilta ei ole mitattu yhtään koepuuta.

Menetelmässä lasketaan aluksi koelakohtaiset pääkomponentit sellaisille koelaloille, joilta on mitattu koepuita. Laskettujen pääkomponenttien ennustamiseksi laaditaan mallit, joissa käytetään selittäjinä koelakohtaisia muuttujia. Merkittäviksi selittäjiksi osoittautuivat muun muassa kasvupaikan boniteetti, lämpösumma, koelalan puiden keskiläpimitta ja pohjapinta-ala. Laadituilla pääkomponenttimalleilla voidaan ennustaa puun muoto myös sellaisilla koelaloilla, joilta ei ole mitattu koepuita. Koelaloilla, joilta on mi-

tattu koepuita, käytetään mallien antamia pääkomponenttimestimaatteja alkuestimaatteina. (Kilkki & Lappi 1987).

Ensimmäisessä vaiheessa laskettiin koelakoh- taiset pääkomponentit kaikkien laskenta-alueen VMI8:n koepuiden avulla. Eteenpäin askelta- valla regressioanalyysillä tutkittiin, mitkä muut- tujat ja muuttujien muunnokset selittävät kuta- kin neljää pääkomponenttia. Esimerkiksi ensim- mäistä pääkomponenttia kuvaavaksi malliksi saa- tiin männylle seuraava yhtälö.

$$c_1 = a_0 + a_1 \ln(\bar{d}_k) + a_2 DD^2 + a_3 \ln(PPA_m) + a_4 \ln(DD) + a_5 PPA_m + a_6 \ln(\bar{d}_m) + a_7 PPA_k^2 + a_8 PPA_k + a_9 T^2 + a_{10} T + a_{11} B_1 + a_{12} B_2 \quad (6.1)$$

- missä \bar{d}_k = koelalan puiden keskiläpimitta,
DD = lämpösumma,
PPA_m = koelalametsikön pohjapinta-ala,
PPA_k = koelalan pohjapinta-ala,
 \bar{d}_m = koelalan mäntyjen keskiläpimitta,
T = metsikön ikä,
B₁ = 1, jos metsikön veroluokka = IA ja 0 muulloin,
B₂ = 1, jos metsikön veroluokka = III ja 0 muulloin, sekä
a₀-a₁₂ ovat kertoimia.

Tutkittaessa koepuumittausten lukumäärän vai- kutusta tilavuusestimaattien luotettavuuteen las- kettiin valittujen pääkomponenttimallien para- metrit kullakin laskentakierroksella erikseen. Mallien laskenta-aineistona käytettiin ainoastaan ko. laskentakierroksella mitatuiksi oletettuja koe- puuta.

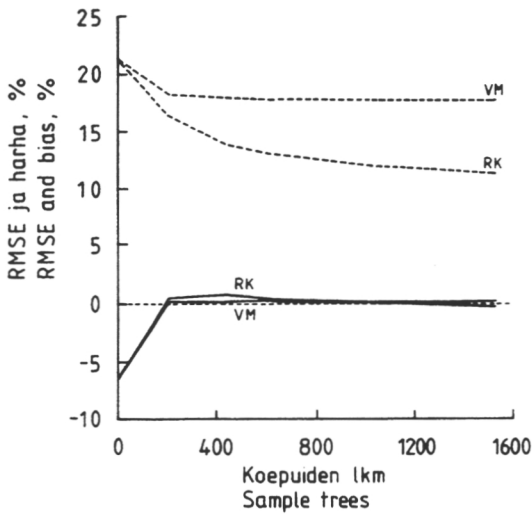
7. Tulokset

7.1. Puittaiset tilavuusestimaatit

Kuvissa 2, 3 ja 4 on esitetty puulajeittain luku- puiden tilavuusestimaatin suhteellinen keskivir- he (RMSE) ja harha eri koepuumäärillä, kun koepuumittauksia on hyödynnetty joko tässä tut- kimuksessa esitetyillä runkokäyrä- ja pääkom- ponenttimalleilla tai kiinteäkertoimisella tila- vuusmallilla (ks. luku 32). Kuvista havaitaan, että sekamallitekniikkaan perustuvilla runkokäy- rämalleilla lasketuilla estimaateilla on selvästi

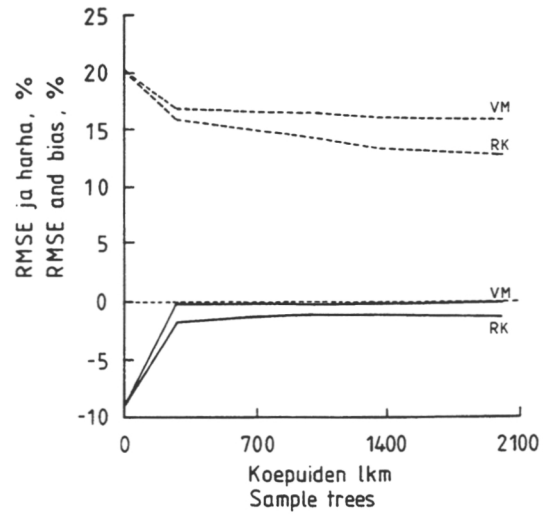
pienempi keskivirhe kuin pelkästään koepuiden avulla laadituilla tilavuusmalleilla lasketuilla esti- maateilla.

Männyn kohdalla molemmat menetelmät ant- tavat tutkituilla koepuumäärillä lähes harhatto- mia estimaatteja. Kuusen tapauksessa runkokäy- rämallin antamat tilavuusestimaatit ovat jonkin verran yliarvioita ja koivun tapauksessa aliarvi- oita, kun harhaa tarkastellaan suhteellisilla yksi- köillä. Kiinteäparametrisella tilavuusmallilla las- kettujen tilavuusestimaattien keskivirhe piene-



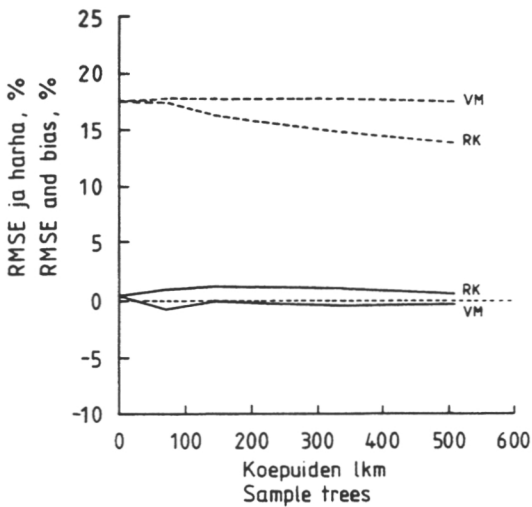
Kuva 2. Männyn puittaisten tilavuusestimaattien luotettavuus eri koepuumäärillä. Merkinnt: — = runkokäyrämalli ja --- = tilavuusmalli.

Figure 2. Reliability of tree-wise volume estimates with different number of sample trees. Notation: — = taper curve model and --- = volume function. Pine.



Kuva 3. Kuusen puittaisten tilavuusestimaattien luotettavuus eri koepuumäärillä. Merkinntä kuten kuvassa 2.

Figure 3. Reliability of tree-wise volume estimates with different number of sample trees. Notation as in Fig. 2. Spruce.



Kuva 4. Koivun puittaisten tilavuusestimaattien luotettavuus eri koepuumäärillä. Merkinntä kuten kuvassa 2.

Figure 4. Reliability of tree-wise volume estimates with different number of sample trees. Notation as in Fig. 2. Birch.

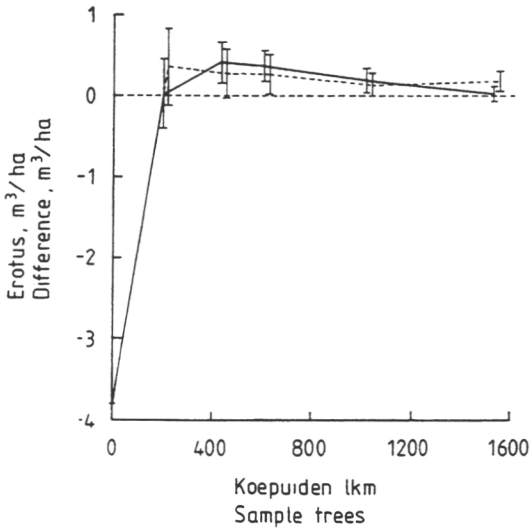
nee vain vähän koepuiden määrää lisättäessä. Keskivirheet olisivat voineet pienentyä enemmän, jos tilavuusmallien laadinnassa olisi käytetty pienempiä ositteita. Osittamisessa käytettävien muuttujien valinta osoittautui kuitenkin vaikeaksi; kokeiltaessa esimerkiksi latvuskerrosta osittamisperusteena keskivirheet kasvoivat.

72. Keskitilavuusestimaatit

Kuvissa 5, 6 ja 7 ovat puulajeittain sekä runkokäyrämalleilla että tilavuusmalleilla laskettujen laskenta-alueen keskitilavuusestimaattien ja todellisten keskitilavuuksien erotusten keskiarvot (= harhat) ja keskihajonnat eri koepuumäärillä. Männyn ja kuusen tapauksessa saadaan ilman koepuumittauksia selvä aliarvio puuston tilavuudelle molemmilla menetelmillä. Tämä on seurausta siitä, että laskenta-alueen puut ovat solakampia kuin puut keskimäärin valtakunnallisessa aineistossa, josta runkokäyrämallit on laadittu. Koivun tapauksessa ilman koepuumittauksia saatava keskitilavuusestimaatti koko laskenta-alueelle on hieman yliarvio.

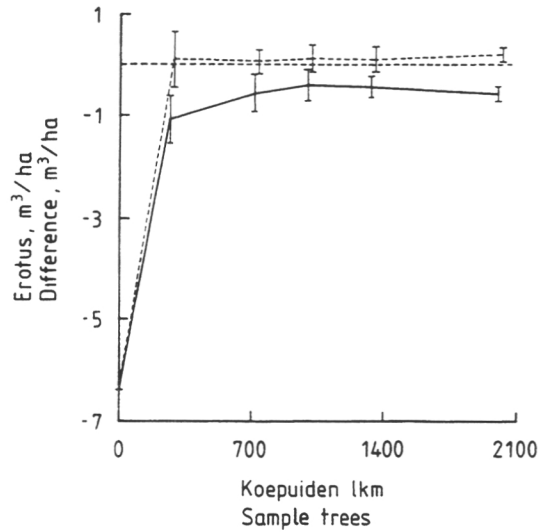
Männyllä runkokäyrämallia käyttämällä saadaan keskitilavuus pienimmällä ja suurimmalla tutkitulla koepuumäärällä luotettavammin kuin kiinteäparametrisellä tilavuusyhtälöllä. Koepuumäärän ollessa noin 350:sta 1100:een ovat runkokäyrämallien antamat keskitilavuusestimaatit kuitenkin selvästi harhaisia.

Kuusen kohdalla runkokäyrämalleilla saadut keskitilavuudet ovat aliarvioita kaikilla tutkituilla koepuumäärillä, mutta koepuiden määrää lisättäessä harha pienenee. Koivun kohdalla tulokset osoittavat, että ilman koepuumittauksia saataisiin ko. laskenta-alueelle keskitilavuusestimaatit luotettavammin kuin mittaamalla vähän koepuita. Kun koepuita mitataan riittävästi, myös



Kuva 5. Männyn keskitilavuusestimaatin poikkeama todellisesta keskitilavuudesta 100 simuloinnissa. Pystysuorat janat kuvaavat erotuksen keskiahjontaa. Merkinät kuten kuvassa 2.

Figure 5. Difference between estimated and true mean volume in 100 simulations. Vertical lines describe the standard deviation of differences. Notation as in Fig. 2. Pine.



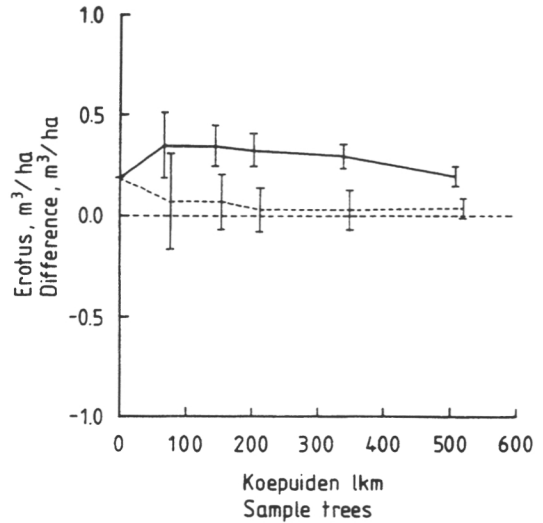
Kuva 6. Kuusen keskitilavuusestimaatin poikkeama todellisesta keskitilavuudesta 100 simuloinnissa. Pystysuorat janat kuvaavat erotuksen keskiahjontaa. Merkinät kuten kuvassa 2.

Figure 6. Difference between estimated and true mean volume in 100 simulations. Vertical lines describe the standard deviation of differences. Notation as in Fig. 2. Spruce.

koivun keskitilavuusestimaatin harha laskee pienemmäksi kuin harha ilman koepuumittauksia.

Kuvissa 5, 6 ja 7 näkyvä todellisen ja estimoidun keskitilavuuden keskimääräinen ero kuvaa osittain samaa asiaa kuin puittaisten tilavuusestimaattien harhat kuvissa 2, 3 ja 4. Täsmälleen samasta asiasta ko. kuvissa ei kuitenkaan ole kysymys, sillä puittaisten tilavuusestimaattien luotettavuutta tarkasteltiin suhteellisten virheiden avulla. Lisäksi keskitilavuuksia laskettaessa yksittäisen puun tilavuutta ja tilavuusestimaattia painotetaan puun pohjapinta-alan käänteisluvulla. Nämä kaksi asiaa selittävät esimerkiksi sen, että koivulla runkokäyrämallin puittaiset tilavuusestimaatit ovat aliarvioita (kuva 4) kun taas keskitilavuusestimaatit ovat yliarvioita (kuva 7).

Yhteenvetona voidaan todeta, että runkokäyrämalleilla saadut keskitilavuusestimaatit ovat yleensä jonkin verran harhaisempia, mutta keskivirheeltään pienempiä kuin kiinteällä tilavuusyhtälöllä saadut estimaatit. Koepuumäärän kasvussa keskivirheiden ero pienenee. Koska metsäninventoinnissa on tärkeää, että kerran tehty inventointi antaa estimoitaville tunnuksille mahdollisimman lähellä todellisia olevat arvot, on käytettävän estimaattorin valinnassa huomioitava sekä estimaattien varianssi että harha.



Kuva 7. Koivun keskitilavuusestimaatin poikkeama todellisesta keskitilavuudesta 100 simuloinnissa. Pystysuorat janat kuvaavat erotuksen keskiahjontaa. Merkinät kuten kuvassa 2.

Figure 7. Difference between estimated and true mean volume in 100 simulations. Vertical lines describe the standard deviation of differences. Notation as in Fig. 2. Birch.

8. Tulosten tarkastelu

81. Runkokäyrämalleissa havaitut puutteet

Yksi tutkimuksen tavoitteista oli laatia runkokäyrämallit, joilla voidaan yhdistää tehokkaasti inventoinnissa tehtävät koepuumittaukset ja mallien muodossa oleva ennakkoinformaatio. VMI8:n koepuilla tehdyt puittaiset tarkastelut (kuvat 2–4) viittaavat siihen, että tavoite on saavutettu. Toisaalta keskitilavuusestimaattien tarkastelu osoittaa, että tutkimuksella laadittu menetelmä antaa joissakin tapauksissa harhaisia tuloksia. Keskitilavuusestimaattien harha on ajoittain niin suuri, ettei laadittua menetelmää voida suositella inventointitulosten laskentaan.

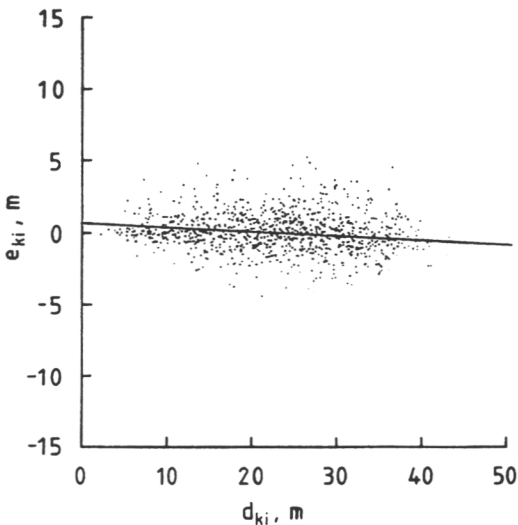
Kuvasta 5 havaitaan, että männyllä pelkän ennakkoinformaation (= runkokäyrämallin ja rinnankorkeusläpimitan) avulla lasketut keskitilavuusestimaatit ovat selviä aliarvioita. Sekamalliteorian mukaan koepuiden mittaamisen tulisi pienentää harhaa siten, että koepuiden määrää lisättäessä harha pieneneisi, mutta pysyisi jatkuvasti samanmerkkisenä. Koepuumäärää lisättäessä männyn tapauksessa keskitilavuusestimaatit muuttuvat aliarvioista yliarvioiksi. Seuraavassa ilmiölle esitetään selitys tarkastelemalla

yksinkertaisuuden vuoksi tilavuusestimaattien sijasta runkokäyrämallin antamia pituusestimaatteja.

Kuvassa 8 on esitetty lukupuiden pituusestimaattien residuaalit ($e = h - \hat{h}$) läpimitan funktiona, kun jokaiselle koelalle on estimoitu pääkomponentit käyttäen kaikista puista rinnankorkeus- ja yläläpimittaa sekä pituutta. Kuvasta näkyy, että pienille puille pituusestimaatit ovat aliarvioita ja suurille yliarvioita.

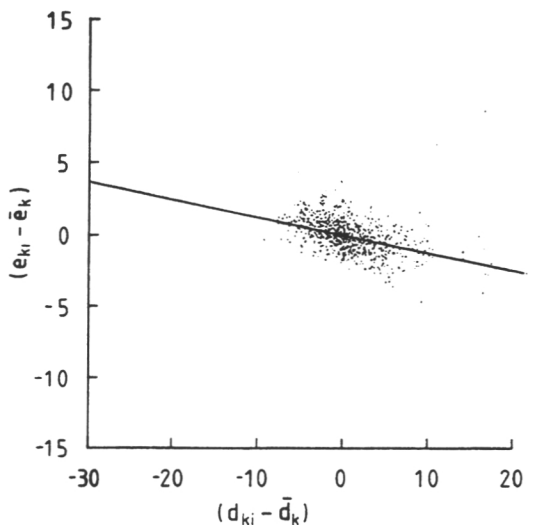
Kuvassa 9 pituusestimaattien poikkeamia on tarkasteltu koelaittain. Kuva on muokattu kuvasta 8 siten, että x-akselin muuttujana on puun läpimitan ja koelalan kaikkien puiden keskiläpimitan erotus ja y-akselin muuttujana on pituusestimaatin virheen ja koelalan pituusestimaattien virheen keskiarvon erotus. Kuva 9 osoittaa, että metsikkötasolla pituusestimaattien harha läpimitan suhteen on suuri.

Syy kuvan 9 muotoon on se, että runkokäyrämalleissa oleva pituusmalli ei ole oikean muotoinen yksittäiselle metsikölle. Muun muassa Hagberg (1939, ks. Loetsch ym. 1973) on todennut, että läpimitan ja pituuden välinen riippuvuus on eri muotoista yksittäisen metsikön kohdalla tar-



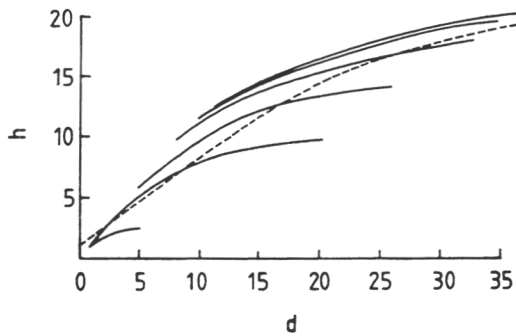
Kuva 8. Lukupuiden pituusestimaattien residuaali d :n funktiona männyllä, kun pääkomponentit on ennustettu koelalan kaikkien puiden avulla.

Figure 8. Residual of the height estimates of tally trees as a function of d . The random plot effect (v) is predicted using all trees on a plot as sample trees. Pine.



Kuva 9. Lukupuiden pituusestimaattien residuaalin ja koelalan residuaalin keskiarvon erotus puun läpimitan ja koelalan keskiläpimitan erotuksen funktiona.

Figure 9. Difference between residuals of height estimates and plotwise means of residuals as a function of difference between diameter and the mean diameter of a plot.



Kuva 10. Pituuskäyriä eri ikäisissä metsiköissä. Lähde: Loetsch ym. (1973).

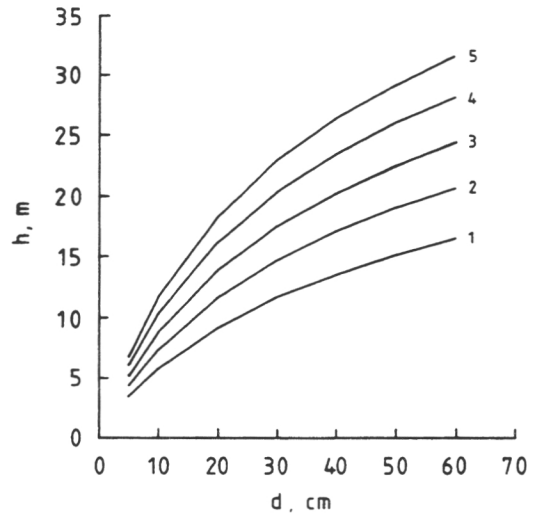
Figure 10. Height curves of different aged stands (Loetsch et al. 1973).

kasteltuna kuin useasta metsiköstä koostuvan aineiston avulla tarkasteltuna. Tätä havainnollistaa kuva 10.

Tässä tutkimuksessa laadittu runkokäyrämalli on tarkoitettu kalibroituun metsikkökohtaiseksi mallissa olevan satunnaisvaikutuksen (= pääkomponenttien) avulla. Koelakohtaisen satunnaisvaikutuksen lisääminen pituusestimaatin logaritmiin ei kuitenkaan taivuta runkokäyrämallin pituusyhtälöä metsikkökohtaiseen muotoon; metsikkökohtaisesti tarkasteltuna pituusmalli antaa satunnaisvaikutuksen lisäyksestä huolimatta yliarvioita isoille ja aliarvioita pienille puille. Jos jokaisesta metsiköstä on mitattu esimerkiksi vain yksi koepuu ja koepuiksi sattuu jokin pienimmistä puista, kalibroitu runkokäyrämalli suurimpien puiden suhteen väärin, koska metsikön pituuskäyrä ei ole samanmuotoinen kuin runkokäyrämallin sisältämä pituusyhtälö. Tästä aiheutuva yliarvio ei kumoudu täysin sillä, että jossakin muussa metsikössä koepuiksi sattuu jokin suurimmista puista, sillä pienen koepuun käytöstä aiheutuva yliarvio suurten puiden estimaateissa on absoluuttisissa yksiköissä suurempi kuin ison koepuun käytöstä aiheutuva aliarvio pienten puiden estimaateissa.

Männyn tuloksissa olevan harhan aiheuttajan varmistamiseksi laskettiin lukupuiden tilavuusestimaattien luotettavuus olettaen kultakin koelalalta mitatuksi koepuuna läpimitaltaan suurin puu. Tällöin saatiin lukupuiden pituus- ja tilavuusestimaateiksi lieviä aliarvioita. Myös tämä tulos vahvistaa sen, että runkokäyrämallin estimoina metsikön pituuskäyrä on liian jyrkkä.

Kuva 11 esittää männyllä runkokäyrämallin antamat pituusestimaatit d :n funktiona. Kuvassa oleva pituuskäyrän odotusarvo (käyrä 3) on saatu olettamalla koelakohtainen satunnaisvaikutus



Kuva 11. Männyn runkokäyrämallin antamia pituusestimaatteja pääkomponenttien eri arvoilla. Käyrien numeroita vastaavat pääkomponenttien arvot ovat: 1 = (-0.43, 0.03, 0.01, -0.01), 2 = (-0.24, 0.02, 0.01, -0.01), 3 = (0.0, 0.0, 0.0, 0.0), 4 = (0.26, -0.02, -0.01, 0.01) ja 5 = (0.62, -0.04, -0.01, 0.02).

Figure 11. Height estimates obtained by the taper curve model with different values of the random plot effect. Pine.

nollaksi. Muut pituuskäyrät on piirretty neljän pääkomponentin eri arvoilla. Pääkomponenttien arvot on saatu olettamalla metsiköstä mitatuksi yhdestä puusta pituus. Käyriä 1 ja 2 laskettaessa pituuskoepuu on oletettu keskimääräistä lyhyemmäksi ja käyriä 4 ja 5 laskettaessa keskimääräistä pitemmäksi. Kuva 11 havainnollistaa saman, mikä on nähtävissä yhtälön muodosta — koelakohtaiset pääkomponentit säätelevät käyrän tasoa, mutta eivät pysty taivuttamaan pituuskäyrää esimerkiksi sellaisiin muotoihin, joita on kuvassa 10.

Toinen syy pituusestimaattien harhaisuuteen läpimitan suhteen on se, että runkokäyrämallisä käytetään selittäjänä puun kokoa — ei rinnankorkeusläpimittaa. Tämän vuoksi runkokäyrämallin antamat estimaatit ovat harhattomia puun koon suhteen, mutta eivät ole harhattomia rinnankorkeusläpimitan suhteen (Lappi 1986). Lappi (1986) on todennut, että estimoituessa puun kokoa pelkän rinnankorkeusläpimitan avulla saadaan suurilla läpimitoilla yliarvioita koolle, mikä aiheuttaa myös tilavuus- tai pituusestimaattien yliarvioita suurille puille.

Lappi (1986) käytti männyn runkokäyrämalleissa puun suhteellista kokoa yhtenä selittäjänä. Myös tätä mallia testattiin VM18-aineistolla. Saadut keskitilavuusestimaatit olivat harhaisem-

pia kuin tässä tutkimuksessa laaditulla mallilla saadut.

Kuusen tapauksessa keskitilavuusestimaatin harha pienenee odotetusti koepuiden määrää lisättäessä. Koska kuusi on varjopuulaji, sen metsikkökohtainen pituuskäyrä on jyrkempi kuin männyn (Loetsch ym. 1973). Tämän vuoksi laadittu runkokäyrämalli tarkentuu koepuiden määrää lisättäessä odotetulla tavalla.

Koivulla koepuiden mittaaminen lisää keskitilavuusestimaattien harhaa aluksi (ks. kuvaa 7). Tämä on seurausta samasta ilmiöstä, joka on kuvattu männyn kohdalla — runkomuodon metsikkökohtainen kalibroiminen lisää suurien puiden harhaa eikä tämä harha kumoudu täysin pienten puiden vastakkaissuuntaisella harhalla. Kun mitattavien koepuiden määrää lisätään, harha pienenee hitaasti.

82. Tutkimuksen suuntaaminen jatkossa

Metsäinventoinnissa käytettävän puustotunnusten esimaattorin on annettava hyviä estimaatteja. Estimaattorin hyvyys määritetään estimaattien harhan ja keskivirheen yhdistelmänä; pelkkä estimaattorin harhattomuus ei ole riittävä mittari. Hyvän estimaattorin tulisi olla myös tarkentuva; koepuiden määrää lisättäessä estimaattien luotettavuuden pitää parantua. Sekamallitekniikalla laadittu estimaattori on metsikkö- ja puutasolla harhainen, mutta tarkentuva ja varianssiltaan pienempi kuin vastaava kiinteäparametrinen estimaattori.

Tutkimuksen tulokset osoittavat, että sekamallitekniikkaa käyttämällä voidaan pienentää tutkittujen estimaattorien keskivirhettä. Sekamallitekniikan soveltaminen tutkimuksessa esitetyllä tavalla johtaa kuitenkin selvästi harhaisiin tilavuusestimaatteihin.

Runkokäyrämallien laadinnassa yksi tavoite oli laatia mallit, jotka eivät ole sidottuja mihinkään tiettyyn mittaukseen. Tavoite saavutettiin

käyttämällä napakoordinaatioesitystä ja valitsemalla kiinteäksi selittäjäksi puun koko. Seuraus puun koon käyttämisestä selittäjänä oli, että malli ei ole harhaton puusta mitattujen tunnus-
ten suhteen. Sovellusten kannalta on kuitenkin usein keskeistä, että estimaatit ovat harhattomia mittausten suhteen. Esimerkiksi VMI:ssa puiden valinnassa käytetään relaskoopia, jolloin siirryttäessä puukohtaisista estimaateista hehtaarikohtaisiin estimaatteihin yksittäisiä puita on painotettava näiden poikkileikkausalan käänteisluvulla (Kuusela 1966). Jos yksittäisten puiden estimaatit eivät ole harhattomia läpimitan suhteen, voivat hehtaarikohtaiset estimaatit olla hyvinkin paljon harhaisia.

Rinnankorkeusläpimitta on vakiintunut puusta useimmin mitattavaksi tunnukseksi. Runkokäyrämallin käytettävyys ei juurikaan huononi-
si, jos malleissa olisi kiinteänä selittäjänä rinnankorkeusläpimitta. Rinnankorkeusläpimitan käyttämisellä saavutettaisiin toinenkin etu: runkokäyrämalliin voitaisiin tällöin lisätä satunnainen d:stä riippuva parametri. Tämä tekisi mallin entistä joustavammaksi, jolloin se todennäköisesti olisi mahdollista kalibroida metsikkökoh-
taiseksi entistä luotettavammin.

Runkokäyrämallien kalibroiminen on ongelmallista koealoilla, joilta ei ole mitattu koepuita. Tässä tutkimuksessa ongelma ratkaistiin Kilkin & Lapin (1987) esittämään tapaan regressiomalleilla. Malleissa selitettävänä muuttujina käytetään pääkomponenttimestimaatit sisältävät satunnaisvirhettä, jonka suuruus riippuu muun muassa koealalta mitattujen koepuiden määrästä. Koska toisaalta koepuiden määrä riippuu puuston tiheydestä ja toisaalta puun runkomuoto riippuu puuston tiheydestä, pääkomponenttimallit eivät välttämättä estimoisi harhattomasti puun runkomuotoa. Pääkomponenttien ennustamisessa tulee jatkossa selvittää Lapin (1990) esittämää menetelmää koealakohtaisten satunnaisparametrien ennustamiseksi.

Kirjallisuus

- Henttonen, H. 1989. Ikä- ja pituusmallit. Moniste. Metsäntutkimuslaitos, Helsinki. 14 s.
- Hicks, C.R. 1973. Fundamental concepts in the design of experiments. Holt, Rinehart and Winston, New York. 308 s.
- Kilkki, P. 1983. Sample trees in timber volume estimation. Seloste: Koepuut puuston tilavuuden estimoinnissa. Acta Forestalia Fennica 182. 35 s.
- & Lappi, J. 1987. Estimation of taper curve using stand variables and sample tree measurements. Scandinavian Journal of Forest Research 2:121–126.
- Kuusela, K. 1966. A basal area-mean tree method in forest inventory. Seloste: Pohjapinta-alakeskipuunemetelmä metsäinventoinnissa. Communicationes Instituti Forestalis Fenniae 61(2).
- Laasasenaho, J. 1982. Taper curve and volume functions for pine spruce and birch. Seloste: Männyn, kuusen ja koivun runkokäyrä- ja tilavuusyhtälöt. Communicationes Instituti Forestalis Fenniae 108. 74 s.
- Lahtinen, A. & Laasasenaho, J. 1979. On the construction of taper curves by using spline functions. Seloste: Runkokäyrän muodostaminen splinifunktiolla. Communicationes Instituti Forestalis Fenniae 95(8). 63 s.
- Lappi, J. 1986. Mixed linear models for analyzing and predicting stem form variation of Scots pine. Seloste: Männyn runkokuodon analysointi ja ennustaminen lineaaristen sekamallien avulla. Communicationes Instituti Forestalis Fenniae 134. 69 s.
- 1990. Statistical methods for changing and non-homogeneous forests. Proceedings of XIXth IUFRO World Congress. Division 6. Montreal, Canada. s. 115–122.
- Loetsch, F., Zöhrer, F. & Haller, K.E. 1973. Forest inventory. Vol. II. BLV Verlagsgesellschaft, München. 469 s.
- Pekkonen, T. 1983. Leimikon puuston tilavuuden arviointi regressioennustinta käyttäen. Metsäntutkimuslaitoksen tiedonantoja 86. 63 s.
- Searle, S.R. 1971. Linear models. John Wiley, New York. 532 s.
- Valtakunnan metsien 8. inventoinnin maastotyöohjeet. 1989. Moniste. Metsäntutkimuslaitos, Helsinki. 96 s.

Total of 13 references

Summary

Using taper curve models based on mixed linear models in forest inventory

Introduction

In forest inventory, characteristics describing the growing stock are nowadays estimated with methods based on treewise models. Stand volume tables and stand volume functions are used only when rough estimates are needed quickly. When treewise methods are used diameter distribution of a stand or a plot is necessary. The distribution can be measured or derived from theoretical models. When the diameter distribution is measured or estimated the total volume can be calculated from the sum of volumes estimated for single trees. (Kilkki 1983).

The volume of a tree can be estimated using volume functions or taper curve models. If taper curve models are used the volumes of different timber assortments can also be calculated. To estimate the volume with a taper curve model estimates of the squares of diameters at different heights are needed. Since taper curve models usually estimate the diameters at different heights the residual variances of the diameter estimates are required to calculate an unbiased volume estimate. (Lappi 1986).

In order to minimize the cost of an inventory a sample

of the growing stock is taken at two phases. First, so-called tally trees are measured to estimate the diameter distribution. Then, the height and possibly one upper diameter are measured from some of the tally trees to describe the form of the trees. The number of sample trees needed depends on:

- 1) the reliability of the result required,
- 2) the reliability and accuracy of the models, and
- 3) the efficiency of the method used to utilize sample tree measurements.

In this paper taper curve models for Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*) and birch (*Betula pendula* and *B. pubescens*) are presented. The goal of the study was to build models which make it possible to combine the sample tree measurements and prior information (models) efficiently. To achieve this the mixed linear model technique described by Lappi (1986) has been applied in the construction of the models.

The use of the taper curve models is studied using data from the 8th National Forest Inventory (NF18) of Finland. The reliability and accuracy of the volume estimates for

single trees and for the whole area are studied. The effect of the number of sample tree measurements is also studied.

Study material

The taper curve models were built using material collected by the Department of Forest Inventory and Yield of the Finnish Forest Research Institute. Data was collected in the period 1968–1970 (Laasasenaho 1982). The data consisted of 2034 pines on 593 plots, 1667 spruces on 487 plots and 602 birches on 200 plots.

The sample tree plots were selected from the plots of the fifth National Forest Inventory of Finland. Thus, the material represents all of Finland. Sample trees on each plot were selected using a relascope (factor 2). One to five trees were measured on each plot. (Laasasenaho 1982).

Diameters at breast height and at the height of six meters were recorded for every sample tree. Diameters at the relative heights of 1, 2.5, 5, 7.5, 10, 15, 20, 30, 40, 50, 60, 70, 80 and 90 % were also measured. Other tree characteristics measured included for example, stump height, total height, height of the living crown and tree age. The volumes of the trees were calculated using spline functions. (Laasasenaho 1982).

The reliability of the taper curve models was tested using both the above mentioned data and data measured at the NFI8. Only sample trees measured in south west Finland were selected from the NFI8 data. This data was collected in the years 1986–1987 (Valtakunnan metsien ... 1986).

Several characteristics describing the site and stand — for example, site class, stand age, basal area and height above sea level — were recorded on the NFI8 sample tree plots. Sample trees were selected with a relascope and d , d_6 and height measured. In this study dead and broken trees were rejected. Thus, the test data consisted of 3061 pines on 623 plots, 4088 spruces on 634 plots and 1019 birches on 379 plots.

Methods

Modelling the tapering of a stem

Taper curve models were built using the method described by Lappi (1986). In this method the so-called polar coordinate system is applied — tree dimensions are stated as diameters or rays at different angles (Sloboda 1977). Diameter can be expressed as a function of the angle and ray (see Fig. 1):

$$D(u) = \cos(u) * R(u),$$

or as a function of the angle and height:

$$D(u) = h(u)/\tan(u),$$

where u = the angle in degrees,

$h(u)$ = the height corresponding to the angle and

$R(u)$ = the ray corresponding to the angle.

The polar coordinate system is presented in Fig. 1.

The taper curve of a stem was described with 13 equations — 12 for estimating diameters at certain angles (later called knot angles) and one for estimating the height of a tree. Knot angles were: 0.25, 0.7, 1.5, 3, 5, 8, 14, 21, 31, 41, 56 and 72 degrees (Lappi 1986).

The diameter of a stem at a certain height depends on the size and form of the tree. In this study the size of a tree is determined as the mean of diameters at the previously-mentioned knot angles. Following function was selected to estimate the diameters at different angles.

$$d_{ki}(u) = a_0(u) + a_1(u)s_{ki} + a_2(u)s_{ki}^2 + v_k(u) + e_{ki}(u),$$

where $d_{ki}(u)$ = the logarithmic diameter at angle u of the i^{th} tree at the k^{th} plot,

s_{ki} = the size of the i^{th} tree at the k^{th} plot,

$a_0(u)$, $a_1(u)$ and $a_2(u)$ are parameters,

$v_k(u)$ is a random stand effect, and

$e_{ki}(u)$ a random tree effect.

The function above consists of two main components: a fixed and a random component. Random effects (the standwise v -terms and tree-wise e -terms) were added because the residuals of each diameter function are correlated within a stand and residuals of functions of different heights are correlated within a tree.

The fixed parameters (a_0 – a_2) were estimated using GLS (generalized least squares). The variances and covariances of the random effects (v and e) were estimated by methods described by Lappi (1986).

The diameter functions estimate the height of a tree and diameters at 12 different angles. The total volume or diameter at any height of the stem was derived from these estimates using cubic spline functions (Lahtinen & Laasasenaho 1979).

Study of the use of the taper curve models

One goal of the study was to study how these taper curve models based on mixed linear models can be used in forest inventory. Sample tree selections with a varying number of measured sample trees were simulated using the NFI8 sample tree data. In these simulations the 'true' total volume of the inventory area was obtained using all the measurements (d , d_6 , and h) of all the trees. The taper curve models built in this study were calibrated plotwise assuming that some of trees were measured as sample trees (= d , d_6 , and h measured) and the rest of the trees as tally trees (= only d measured).

Calibration of the taper curve model was studied assuming that every 15th, 7th, 5th, 3rd, and 2nd tree measured was a sample tree. Sample trees were selected at random from each plot. Each sample was repeated 100 times to study the effect of selection of sample trees.

The standard deviation and mean of differences of true and estimated volumes of tally trees were calculated and presented as a function of the number of sample trees. The reliability of the mean volume estimate (m³/ha) was also calculated.

Results obtained with the taper curve models were compared with those obtained using simple volume functions. In each inventory iteration, functions estimating the stem volume ($v = f(d)$) were estimated using data of sample tree measurements. To improve results sample tree data were divided into strata, e.g., by the site classes of the plots. Parameters of the volume function were estimated separately for each stratum.

Taper curve models for pine, spruce and birch

Parameter estimates for functions estimating the diameters at different angles are presented in Table 1. The standard deviations and correlations of the random effects (v- and e-terms) at different angles are presented in Appendices 1, 2 and 3. The standard deviations of the v- and e-terms can be interpreted as the relative plotwise and treewise standard error components of the diameter functions because the functions are logarithmic.

The size (s) of a tree is used as a fixed independent variable in the selected taper curve model. In applications, however, only the fixed parameters (a_0 , a_1 and a_2) and the variances of the random effects are known. The size of a tree and value of the plotwise random effect are not known. The size variable can be estimated if one diameter at any height of the stem is known. The plotwise random effect can be predicted if at least two diameters (or one diameter and the height) are known for at least one tree on a plot. The prediction process is based on general theory of mixed linear models and is presented in detail in Lappi (1986).

Reliability of the taper curve models

The reliability of the taper curve model was studied using biases and standard errors of the diameter and volume estimates when different variables were measured. Only absolute biases and standard errors are presented for the diameter estimates. Relative errors are also presented for the volume estimates. The relative errors are calculated using both the true and estimated volume as divisors. The standard error estimates are divided into plotwise and treewise components (see e.g. Searle 1971).

The reliability of the diameter estimates was studied

assuming a) only the diameter and b) the diameter and height to be known. The results are presented in Tables 2 and 3, respectively. Taper curve models are practically unbiased and the standard errors about equal to those obtained by Laasasenaho (1982).

The reliability of the volume estimates was studied in the following cases:

- 1) only d was measured,
- 2) d and h were measured,
- 3) d, d₆ and h were measured, and
- 4) only d was measured for all trees and d, d₆ and h were measured for one tree on each plot.

The standard errors and biases of the volume estimates are presented in Table 4. The volume estimates are practically unbiased when the height of the tree is known. The calibration of the taper curve model with one sample tree (case 4 in the list above) significantly reduces the standard errors. In this case however, the volume estimates are somewhat biased, if the bias is measured by the mean of relative errors using estimated volume as a divisor. This is due to the fact that the taper curve model is based on the size (s) of the tree. Thus, the model gives unbiased results in respect to the size, but not in respect to the measurements (see Lappi 1986).

Results

When the taper curve models built in this study were tested using NF18 data, both treewise volume estimates and mean volume estimates (m³/ha) were studied. Treewise root mean square errors and biases are presented in Figures 2, 3 and 4 for pine, spruce and birch, respectively. The results obtained with volume functions are also presented in these figures. The results show that taper curve models based on mixed linear models can be calibrated efficiently with fewer sample tree measurements than volume functions based on fixed regressors.

The results of the mean volume estimates (Figures 5, 6 and 7) show that taper curve models give biased estimates, especially for birch. The standard errors for the mean volume estimates obtained with taper curve models are lower than the standard errors for the volume functions. However, in some cases the biases are so high, that the taper curve models built in this study cannot be recommended.

Conclusions

The results presented in Chapter 6 show that the calibration of the taper curve model using few sample trees leads to biased mean volume estimates. The bias is caused by two factors.

- 1) The diameter models were built using the size of a

tree as a fixed regressor. Thus, the taper curve estimates are not unbiased in respect to the measured diameters. When the mean volume estimates are calculated, the volume estimates of single trees (selected by a relascope) are multiplied by a factor which is dependent on the basal area of the tree. Thus, if volume estimates of single trees are biased in respect to the diameter at breast height, the mean volume estimate can be remarkably biased.

- 2) The relative size of a tree is not taken into account in the diameter functions. Thus, for example, the height function is too 'steep': it underestimates the heights

of small trees and overestimates the heights of big trees. The random plotwise effect in the height function cannot eliminate this bias.

The results of this study indicate that mixed linear models form an efficient method for combining sample tree measurements and a priori information. The taper curve models presented in this study, however, need to be re-constructed to guarantee unbiased results. It is probable that the use of diameter at breast height as a fixed regressor instead of the size parameter would lead to better results.

Liite 1. Männyn läpimittayhtälöiden satunnaistekijöiden korrelaatiomatriisit. Diagonaaleilla olevat hajonnat on kerrottu 100:lla.

Appendix 1. Correlation matrices of plotwise and treewise random effects. The diagonal elements are standard deviations multiplied by 100. Pine.

Kulma Angle	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>Koelakohtainen tekijä – Plotwise random effect</i>													
1	7.56	0.98	0.94	0.91	0.91	0.88	0.80	0.37	-0.88	-0.93	-0.95	-0.95	-0.95
2		6.76	0.98	0.95	0.93	0.88	0.76	0.29	-0.92	-0.95	-0.96	-0.96	-0.95
3			6.68	0.99	0.96	0.92	0.80	0.31	-0.94	-0.98	-0.97	-0.97	-0.95
4				6.20	0.99	0.96	0.85	0.38	-0.94	-0.99	-0.98	-0.97	-0.96
5					6.06	0.99	0.90	0.48	-0.91	-0.98	-0.99	-0.98	-0.97
6						5.53	0.96	0.60	-0.85	-0.96	-0.98	-0.97	-0.97
7							3.89	0.81	-0.69	-0.86	-0.91	-0.91	-0.91
8								1.93	-0.11	-0.40	-0.50	-0.53	-0.54
9									3.14	0.95	0.90	0.87	0.85
10										6.03	0.99	0.97	0.95
11											9.26	0.99	0.98
12												11.32	0.99
13													12.99
<i>Puukohtainen tekijä – Treewise random effect</i>													
1	6.51	0.81	0.57	0.44	0.32	0.21	0.01	-0.25	-0.58	-0.66	-0.65	-0.62	-0.59
2		5.27	0.85	0.65	0.52	0.38	0.12	-0.25	-0.70	-0.80	-0.80	-0.76	-0.72
3			4.90	0.82	0.70	0.53	0.25	-0.15	-0.67	-0.82	-0.84	-0.81	-0.79
4				4.24	0.85	0.71	0.42	-0.01	-0.60	-0.79	-0.85	-0.84	-0.81
5					3.98	0.80	0.53	0.09	-0.52	-0.75	-0.81	-0.80	-0.78
6						3.63	0.65	0.25	-0.38	-0.63	-0.74	-0.75	-0.73
7							3.19	0.56	-0.03	-0.37	-0.53	-0.58	-0.59
8								2.77	0.45	0.12	-0.12	-0.22	-0.25
9									3.18	0.77	0.57	0.45	0.39
10										4.22	0.86	0.76	0.71
11											5.60	0.94	0.90
12												6.76	0.97
13													7.50

Liite 2. Kuusen läpimittayhtälöiden satunnaistekijöiden korrelaatiomatriisit. Diagonaaleilla olevat hajonnat on kerrottu 100:lla.

Appendix 2. Correlation matrices of plotwise and treewise random effects. The diagonal elements are standard deviations multiplied by 100. Spruce.

Kulma Angle	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>Koelakohtainen tekijä – Plotwise random effect</i>													
1	7.10	0.94	0.81	0.76	0.71	0.67	0.40	-0.66	-0.83	-0.83	-0.84	-0.85	-0.89
2		5.42	0.95	0.87	0.82	0.77	0.48	-0.72	-0.92	-0.92	-0.93	-0.93	-0.92
3			5.47	0.96	0.93	0.89	0.61	-0.69	-0.96	-0.97	-0.97	-0.96	-0.92
4				5.05	0.99	0.94	0.69	-0.61	-0.95	-0.98	-0.97	-0.96	-0.90
5					4.53	0.98	0.77	-0.51	-0.92	-0.96	-0.96	-0.95	-0.89
6						3.64	0.85	-0.38	-0.87	-0.93	-0.95	-0.94	-0.88
7							1.85	0.11	-0.52	-0.66	-0.73	-0.74	-0.69
8								1.52	0.79	0.67	0.58	0.53	0.48
9									3.41	0.98	0.95	0.91	0.85
10										4.88	0.99	0.96	0.90
11											6.34	0.99	0.94
12												7.30	0.97
13													8.34
<i>Puukohtainen tekijä – Treewise random effect</i>													
1	7.51	0.81	0.43	0.34	0.20	0.09	-0.16	-0.43	-0.62	-0.65	-0.61	-0.56	-0.52
2		5.17	0.80	0.55	0.39	0.25	-0.09	-0.48	-0.76	-0.81	-0.77	-0.71	-0.65
3			4.90	0.78	0.65	0.48	0.13	-0.32	-0.70	-0.80	-0.80	-0.77	-0.73
4				4.14	0.83	0.68	0.33	-0.15	-0.62	-0.77	-0.82	-0.81	-0.77
5					3.90	0.82	0.54	0.03	-0.48	-0.69	-0.78	-0.79	-0.76
6						3.46	0.67	0.22	-0.32	-0.56	-0.70	-0.72	-0.71
7							2.88	0.62	0.12	-0.19	-0.41	-0.50	-0.51
8								2.56	0.61	0.35	0.10	-0.04	-0.10
9									3.02	0.84	0.66	0.50	0.42
10										3.85	0.88	0.77	0.68
11											5.08	0.93	0.85
12												6.24	0.95
13													7.04

Liite 3. Koivun läpimittayhtälöiden satunnaistekijöiden korrelaatiomatriisit. Diagonaaleilla olevat hajonnat on kerrottu 100:lla.

Appendix 3. Correlation matrices of plotwise and treewise random effects. The diagonal elements are standard deviations multiplied by 100. Birch.

Kulma Angle	1	2	3	4	5	6	7	8	9	10	11	12	13
Koelakohtainen tekijä – Plotwise random effect													
1	6.99	0.96	0.81	0.75	0.66	0.66	0.57	0.30	-0.82	-0.87	-0.85	-0.85	-0.83
2		5.17	0.93	0.86	0.78	0.75	0.64	0.33	-0.89	-0.95	-0.93	-0.92	-0.88
3			4.55	0.97	0.90	0.87	0.77	0.43	-0.83	-0.97	-0.96	-0.94	-0.91
4				4.48	0.96	0.92	0.86	0.57	-0.73	-0.95	-0.96	-0.96	-0.94
5					3.91	0.97	0.94	0.67	-0.61	-0.89	-0.94	-0.94	-0.93
6						3.35	0.98	0.82	-0.48	-0.84	-0.93	-0.95	-0.94
7							2.60	0.84	-0.32	-0.74	-0.86	-0.89	-0.89
8								1.70	0.03	-0.39	-0.59	-0.65	-0.67
9									1.70	0.88	0.78	0.69	0.62
10										3.58	0.98	0.94	0.88
11											6.45	0.99	0.94
12												8.59	0.98
13													10.22
Puukohtainen tekijä – Treewise random effect													
1	7.76	0.86	0.46	0.29	0.16	0.07	-0.12	-0.36	-0.55	-0.60	-0.62	-0.59	-0.53
2		5.50	0.80	0.46	0.30	0.15	-0.07	-0.41	-0.67	-0.74	-0.74	-0.69	-0.64
3			5.06	0.69	0.52	0.35	0.08	-0.30	-0.64	-0.75	-0.75	-0.70	-0.67
4				4.45	0.75	0.63	0.33	-0.05	-0.51	-0.71	-0.75	-0.76	-0.74
5					3.92	0.72	0.53	0.13	-0.41	-0.65	-0.69	-0.72	-0.72
6						3.97	0.62	0.28	-0.28	-0.56	-0.62	-0.66	-0.65
7							3.34	0.53	-0.02	-0.29	-0.41	-0.46	-0.48
8								3.02	0.44	0.15	0.01	-0.11	-0.15
9									3.41	0.73	0.55	0.43	0.36
10										4.31	0.83	0.74	0.68
11											5.44	0.89	0.82
12												6.57	0.92
13													7.07

Liite 4. Koelakohtaisen satunnaistekijän kovarianssimatriisista lasketut ominaisvektorit ja -arvot.

Appendix 4. *Eigenvectors and eigenvalues calculated from the covariance matrix of plotwise random effect.*

Kulma Angle	1	2	3	4
Ominaisvektorit MÄNTY – Eigenvectors PINE				
1	0.279	0.311	-0.545	-0.514
2	0.252	0.420	-0.179	0.101
3	0.251	0.309	0.195	0.318
4	0.235	0.112	0.357	0.260
5	0.230	-0.077	0.321	-0.007
6	0.207	-0.276	0.234	-0.149
7	0.134	-0.438	0.031	-0.259
8	0.036	-0.423	-0.115	-0.116
9	-0.109	-0.267	-0.261	0.272
10	-0.227	-0.113	-0.331	0.344
11	-0.355	0.054	-0.160	0.272
12	-0.435	0.143	0.107	-0.100
13	-0.496	0.245	0.346	-0.422
Ominaisarvot (suluissa osuudet varianssista,%) <i>Eigenvalues (percentages of the variance in brackets)</i>				
0.0674 (96.2) 0.0015 (2.1) 0.0008 (1.2) 0.0002 (0.3)				
Ominaisvektorit KUUSI – Eigenvectors SPRUCE				
1	0.344	0.714	0.013	-0.448
2	0.281	0.280	0.223	0.325
3	0.287	-0.110	0.282	0.539
4	0.263	-0.268	0.174	0.036
5	0.231	-0.327	0.018	-0.158
6	0.180	-0.304	-0.115	-0.267
7	0.067	-0.208	-0.296	-0.202
8	-0.050	-0.094	-0.373	-0.085
9	-0.174	0.043	-0.367	0.136
10	-0.258	0.127	-0.268	0.271
11	-0.339	0.162	-0.071	0.235
12	-0.392	0.118	0.183	-0.056
13	-0.439	-0.134	0.596	-0.326
Ominaisarvot (suluissa osuudet varianssista,%) <i>Eigenvalues (percentages of the variance in brackets)</i>				
0.0341 (91.4) 0.0019 (5.0) 0.0008 (2.3) 0.0003 (0.7)				
Ominaisvektorit KOIVU – Eigenvectors SPRUCE				
1	0.326	0.687	-0.333	-0.189
2	0.258	0.366	0.100	0.132
3	0.231	0.020	0.409	0.351
4	0.228	-0.143	0.335	0.231
5	0.193	-0.260	0.192	-0.017
6	0.165	-0.258	-0.016	-0.236
7	0.119	-0.265	-0.096	-0.230
8	0.053	-0.226	-0.209	-0.314
9	-0.067	-0.186	-0.281	-0.054
10	-0.181	-0.102	-0.340	0.118
11	-0.338	0.012	-0.290	0.439
12	-0.454	0.113	0.052	0.288
13	-0.533	0.243	0.478	-0.519
Ominaisarvot (suluissa osuudet varianssista,%) <i>Eigenvalues (percentages of the variance in brackets)</i>				
0.0355 (91.5) 0.0021 (5.5) 0.0008 (2.0) 0.0003 (0.8)				

Liite 5. VMI8-aineistosta laaditut pituusmallit.
Appendix 5. Height models estimated from the NF18 data.

Malli – Model: $\ln(h - 1.3) = a_0 + a_1 * 1/(d+5) + a_2 * 1/(d+5)^2$

Kertoimet – Parameters:	a_0	a_1	a_2
Mänty – Pine	3,59	-24,67	-1,67
Kuusi – Spruce	3,71	-22,88	-24,75
Koivu – Birch	3,47	-16,36	-10,76

II

Mixed estimation in calibration of volume functions of Scots pine

Kari T. Korhonen

TIIVISTELMÄ: SEKAESTIMOINTI MÄNNYN TILAVUUSYHTÄLÖIDEN KALIBROINNISSA

Korhonen, K.T. 1993. Mixed estimation in calibration of volume functions of Scots pine. Tiivistelmä: Sekaestimointi männyn tilavuusyhtälöiden kalibroinnissa. *Silva Fennica* 27(4): 269–276.

Regression models for estimating stem volume of Scots pine were constructed using sample tree data measured in the 7th and 8th National Forest Inventory of Finland. Stem volume were regressed on diameter, basal area of growing stock, and geographic location. The results of the study show that using second order trend surface to describe the geographic variation of the residuals gives satisfactory results. Using mixed estimation for combining old and new sample tree data improves the efficiency of an inventory. The weight of the prior information must be low, because remarkable difference in stem form was found in the two inventories.

Tutkimuksessa laadittiin valtakunnan metsien 7:n ja 8:n inventoinnin koepuuaineistosta regressiomallit männyn tilavuuden estimoimiseksi. Työssä tarkasteltiin sekaestimointia kahden inventointiaineiston koepuumittausten yhdistämisessä. Tutkimus osoittaa, että tilavuusmallin alueellista jäännösvaihtelua voidaan kuvata tyydyttävästi toisen asteen trendipinnalla. Vanhan koepuuaineiston hyödyntämisellä voidaan tehostaa inventoinnissa käytettäviä malleja. Koska inventointien välissä näyttää tapahtuneen merkittävä muutos mäntyjen runkomuodossa, ennakkoinformaatiolle on annettava pieni paino. Lisätutkimuksia tarvitaan mallien kalibroimisessa pienille alueille ja puun runkomuodon muutoksista inventointien välillä.

Keywords: forest inventories, volume, models, estimation, *Pinus sylvestris*.
FDC 524

Author's address: The Finnish Forest Research Institute, Joensuu Research Station, P.O. Box 68, FIN-80101 Joensuu, Finland.

Accepted January 21, 1994

Notation

[recording units in brackets]

d	= diameter at breast height (1.3 meters from the ground level) [cm]
d_6	= diameter measured at the height of 6 meters [cm]
h	= height of a tree [dm]
v	= volume of a tree [dm ³]
G	= basal area of the growing stock [m ² /ha]
DIST	= distance from the coast (of Gulf Botnia or Finnish Gulf) [km]
RDIST	= 0, if DIST > 20 1 / DIST - 0.05, otherwise
Y	= p-coordinate of the plot (distance from the Equator) [km]
YC	= (Y - 6620) / 1000
X	= i-coordinate of the plot (distance from the Greenwich meridian) [km]
XC	= (X - 60) / 1000
ln(X)	is natural logarithm of variable X

1 Introduction

Two phase sampling is generally applied in forest inventories. At the first phase a large number of trees, so called tally trees, is measured for diameter and other easily measurable variables. At the second phase, a relatively small number of trees, so-called sample trees, are selected for further measurements. The goal of the measurements at the second phase is to estimate the relationship between the easily measurable tally tree variables (e.g. diameter) and the variables of final interest (e.g. stem volume) (see e.g. Cunia 1986, Kilkki 1979).

When inventories are repeated it is customary to select new sample trees in each inventory occasion. Both theoretical studies on methods

involved (e.g. Theil and Goldberger 1961, Johnston 1972, Teräsvirta 1981, Pekkonen 1983, Meng et al. 1990, Lappi 1991) and empirical results on the use of different methods in combining prior information with new measurements has been published (e.g. Pekkonen 1983, Green and Strawdeman 1985, Burk and Ek 1982, Meng et al. 1990, Korhonen 1992). However, there is only little knowledge of how applicable different methods are for example in the National Forest Inventory (NFI) of Finland.

The purpose of this study is to investigate mixed estimation in the use of sample tree data in the NFI of Finland. This study is limited to Scots pine (*Pinus sylvestris* L.).

2 Material and methods

2.1 Study material

Two data sets were used in the study. The first data set was the pine sample trees measured in the 7th National Forest Inventory (NFI7) of Finland. The second data set was the pine sample trees measured in the 8th National Forest Inventory (NFI8) in National Board Districts Etelä-Karjala and Pirkka-Häme (see Fig. 1). The sample tree data of NFI7 consists of 28575 pines measured during 1977–1983. The sample tree data of NFI8 consists of 1157 pines in Etelä-Karjala and 1226 pines in Pirkka-Häme. The

NFI8 data were measured in years 1986–1987.

From each sample tree in both data sets the following measured dimensions were used in this study (recording units in brackets):

- diameter at breast height [cm],
- height of the tree from ground level to top of the tree [dm], and
- upper diameter at the height of 6 meters from the ground [cm].

Sample trees were selected with a relascope (basal area factor 2). From each plot several charac-

teristics describing the site and growing stock were registered (see e.g. Valtakunnan metsien... 1988).

2.2 Construction of volume functions

For each sample tree the stem volume was estimated with volume functions of Laasasenaho (1982) using d , d_6 (for trees higher than 8 meters, only), and h as independent variables. Estimated volumes were used as 'true' values when volume functions for tally trees were established.

Models for estimating stem volume as a function of diameter and variables describing the growing stock and location were constructed at two stages. At the first stage the sample tree data of NFI7 were used for finding the form of the model and for calculating first-level estimates of the parameters. At the second stage second-level (localized) estimates of the model parameters were estimated using the data of NFI8.

At Stage 1 ordinary least squares technique was applied for choosing the independent variables and the form of the volume function. At Stage 2 the volume function was re-estimated using mixed estimation (Theil and Goldberger 1961, Johnston 1972). In this method the first-level information (sample tree data of NFI7) and second-level information (data of NFI8) are combined as follows. Let us assume a regression model:

$$y = X\beta \quad (1)$$

where

y = vector of dependent variables,
 X = matrice containing independent variables, and
 β = parameter vector.

Let us note the first-level information on y and X with r and R , respectively. Correspondingly, the second-level information on y and X is noted with s and S . The parameters of the model are estimated with Formula (2) using both data sets (Theil and Goldberger 1961, Teräsvirta 1981, Korhonen 1992).

$$\hat{\beta} = (S'S + kR'R)^{-1} (S's + kR'r) \quad (2)$$

where

k = weight of the prior information.

If some of the parameters of the model are estimated using only the first-level data and some of

the parameters are estimated with both data sets the model (1) must be re-written as follows.

$$r = R_1\beta_1 + R_2\beta_2 \quad (3)$$

$$s = S_1\beta_1 + S_2\beta_2 \quad (4)$$

where

R_1 and S_1 contain regressors whose parameters are estimated using the first-level information only,

R_2 and S_2 contain regressors whose final parameter estimates are obtained using both first- and second-level information, and

β_1 and β_2 are parameter vectors, respectively.

Then, Formula (5) gives the parameter estimates (Korhonen 1992).

$$\hat{\beta}_2 = (S_2'S_2 + k m / n R_2'R_2)^{-1} (S_2'u + k m / n R_2'v) \quad (5)$$

where

m = number of observations in the second-level data,

n = number of observations in the first-level data,

$u = s - S_1\hat{\beta}_1$,

$v = r - R_1\hat{\beta}_1$, and

$\hat{\beta}_1$ = first-level estimate for β_1 .

2.3 Analysis of residuals

The geographic variation of residuals was studied by plotting the residuals on a map. The mean value of relative residuals $(v - \hat{v} / \hat{v})$ were calculated for each cluster of sample plots in the NFI7 data. Clusterwise means were smoothed using moving averages -method with formula (6) (Ripley 1981, page 36).

$$\hat{Z}(x) = \sum_{i=1}^L W_i \cdot Z(x_i), \quad (6)$$

where

$\hat{Z}(x)$ = smoothed surface,

L = number of points (clusters),

$Z(x_i)$ = observed value at point x_i (mean of residuals for cluster i),

W_i = weight of observation i .

In this study,

$W_i = j_i / J_i$, for the cluster itself and all the 8 neighbouring clusters,

= 0, for all the other clusters,

j_i = number of sample trees at cluster i ,

J_i = number of sample trees at cluster i and all the 8 neighbouring clusters.

2.4 Sampling simulations

Use of mixed estimation in calibration of volume functions was studied by sampling simulations in the NFI8 data. In the simulations, height and upper diameter ('true' volume) was assumed to be measured from varying number of trees. These data were used to estimate volume functions for tally trees (trees with d as only measured dimension). Mixed estimation was compared with OLS in estimating the functions. Different values for weight parameter k (see Formula (5)) were tested. Selection of sample trees was done with systematic sampling with random

starting point. Sampling with each sample size was repeated 100 times varying the selected sample trees (see Korhonen 1992).

In each sampling simulation RMSE of mean volume estimates of the inventory area and RMSE of volume estimates for single trees were calculated. RMSE's of tree-wise volume estimates were calculated for relative errors $(= (v - \hat{v}) / \hat{v})$. It should be noticed that the mean volumes estimated in this study are the mean volumes of pines on those sample plots where at least one pine was measured. Therefore, results can not be interpreted as mean volumes of pines on forest land.

3 Volume functions

3.1 First-level estimates

Equations (7) and (8) were chosen to estimate the stem volume of pine.

$$v / d^2 = b_0 + b_1 \cdot d + b_2 \cdot d_2 + b_3 \cdot \ln(G) + b_4 \cdot \text{RDIST} + b_5 \cdot \text{YC} + b_6 \cdot \text{YC}^2 + b_7 \cdot \text{XC} + b_8 \cdot \text{XC}^2 + b_9 \cdot \text{YC} \cdot \text{XC} \quad (7)$$

$$v / d^2 = b_{10} + b_{11} \cdot d + b_{12} \cdot d^2 + b_{13} \cdot \ln(G) \quad (8)$$

where

b_0, \dots, b_{13} are parameters,

v = volume of a tree [dm³]

d = diameter at breast height (1.3 meters from the ground level) [cm],

G = basal area of the growing stock [m²/ha],

$\text{RDIST} = 0$, if $\text{DIST} > 20$

$1 / \text{DIST} - 0.05$, otherwise

DIST = distance from the coast (of Gulf Botnia or Finnish Gulf) [km],

$\text{YC} = (Y - 6620) / 1000$,

Y = p-coordinate of the plot (distance from the Equator) [km],

$\text{XC} = (X - 60) / 1000$,

X = i-coordinate of the plot (distance from the Greenwich meridian) [km].

The first-level estimates for the parameters were estimated using the NFI7 data with OLS-technique. Parameter estimates and their t-values are in Table 1.

In Model (8) the effect of geographic location on the volume is described using distance from the coast and a second order trend surface as

Table 1. First-level parameter estimates, t-values and RMSE for volume functions (7) and (8).

Variable	Function (7)		Function (8)	
	Parameter estimate	t-value	Parameter estimate	t-value
constant	-0.055623	9.72	-0.0000578	0.02
d	0.025963	119.41	0.0257744	114.38
d^2	-0.000284	60.31	-0.0002817	57.55
$\ln(G)$	0.052251	68.18	0.0591519	76.55
RDIST	-0.072047	19.58		
YC	0.024861	1.72		
YC^2	-0.066980	1.48		
XC	0.372593	14.60		
XC^2	-0.219698	6.21		
$\text{YC} \cdot \text{XC}$	-0.247642	8.24		
	RMSE = 0.093		RMSE = 0.097	

Table 2. Comparison of mean volumes estimated with different models. Notation: measured = measured volumes of trees were used; Model 7a = Model (7) with parameters estimated in NFI7 data from whole country were used; Model 7b = Model (7) with parameters estimated in NFI7 data only from the studied district were used; Model 8 = Model (8) with parameters estimated in NFI7 data from whole country.

District	Measured	Model 7a	Model 7b	Model 8
Etelä-Karjala	72.15	70.58	74.29	75.37
Pirkka-Häme	73.67	71.30	72.48	71.10

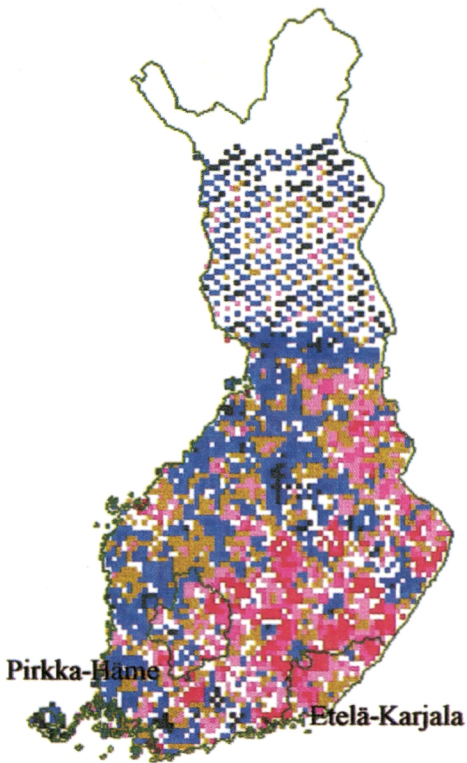


Fig. 1. Residuals of Model (7) in the NFI7 data. Blue colours indicate negative residuals and red colours positive residuals. Yellow colour indicates residuals close to zero. White colour is for missing values.

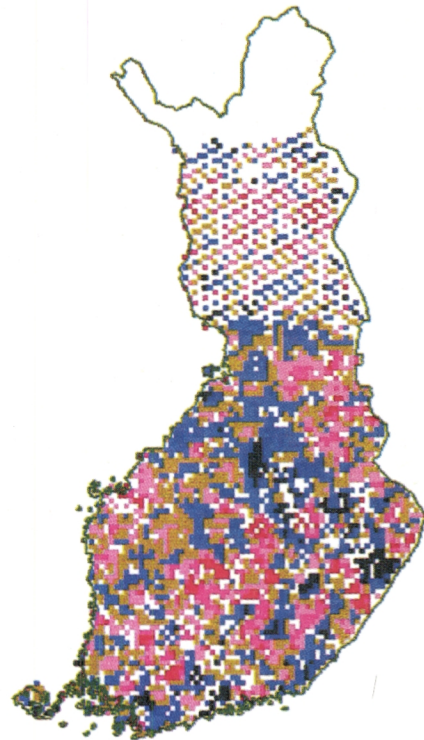


Fig. 2. Residuals of Model (8) in the NFI7 data. Blue colours indicate negative residuals and red colours positive residuals. Yellow colour indicates residuals close to zero. White colour is for missing values.

regressors. Both of these regressors are statistically significant. The usefulness of these variables was further studied by comparing the geographic variation of residuals of Models (7) and (8) (see Figs. 1 and 2). Residual maps show that using geographic variables as regressors clearly diminishes large scale spatial correlation of residuals. For Model (7) the residuals are mostly negative for west and north parts of Finland and mostly positive in east and south-east parts of the country. For Model (8) no clear differences between different parts of the country are found.

In order to study if there are differences in the stem form of NFI7 and NFI8 data mean volumes for Etelä-Karjala and Pirikka-Häme districts were calculated using Models (7) and (8) and NFI7 data. Mean volumes were calculated in 4 different ways:

1. Using measured volumes of trees (= volumes obtained using measured d and h).

2. Using Model (7) with parameters estimated in NFI7 data from whole country.
3. Using Model (7) with parameters estimated in NFI7 data from the studied district, only.
4. Using Model (8) with parameters estimated in NFI7 data from whole country.

Results in Table 2 show that both Models (7) and (8) are remarkably biased in NFI8 data if the parameters are estimated using NFI7 data. When Model (7) with parameters estimated for the whole country is used (7a in Table 2) for Etelä-Karjala district, the bias evidently consists of two components: differences as a function of location and differences as a function of time. When Model (8) is used the first component of bias is reduced and results are close to those obtained with Model (7) with parameters estimated for the study area (7b in Table 2). In Pirikka-Häme district the difference between Models (7) and (8) is negligible but the difference

Table 3. Parameter estimates for functions (7) and (8) in the NFI8 data from Etelä-Karjala and Pirkka-Häme districts.

	Function (7)		Function (8)	
	Etelä-Karjala	Pirkka-Häme	Etelä-Karjala	Pirkka-Häme
With prior information				
constant	-0.1707758	-0.1304988	-0.07471644	-0.0603038
d	0.0299622	0.0291630	0.02964140	0.0291997
d ²	-0.0003405	-0.0003033	-0.00033479	-0.0003018
ln(G)	0.0655989	0.0634001	0.07094529	0.0644148
RDIST	-0.0720467	-0.0720467		
YC	0.0248618	0.0248618		
YC ²	-0.0166980	-0.0166980		
XC	0.3725925	0.3725925		
XC ²	-0.2196978	-0.2196978		
YC · XC	-0.2476425	-0.2476425		
Without prior information				
constant			-0.0893162	-0.0695003
d			0.0300710	0.0298306
d ²			-0.0003416	-0.0003105
ln(G)			0.0743722	0.0648874

between the data sets from NFI7 and NFI8 is clear.

3.2 Second-level estimates

Second-level estimates for the parameters of Functions (7) and (8) were calculated using the NFI8 data measured in National Board districts Etelä-Karjala and Pirkka-Häme. Only constant and parameters of d , d^2 , and $\ln(G)$ were re-estimated. Parameters of Function (8) were estimated also without using NFI7 data as prior information. The parameter estimates are in Table 3.

3.3 Comparison of mixed estimation and OLS in sampling situation

Reliability of the volume functions and effect of sampling error were studied by sampling simulations using the the NFI8 data measured in National Board districts Etelä-Karjala and Pirkka-

Häme. Results obtained by calibration of Function (8) with mixed estimation were compared with results obtained by using Function (7) without prior information (= first-level parameter estimates). Different values for weight parameter k were tested. Results using values 0.1 and 0.5 for k are presented in Figs. 3–6.

Figs. 3 and 4 show the expected fact that mixed estimation improves reliability of mean volume estimates most effectively when the number of sample trees is low. In most cases it is more effective to use relatively small weight for prior information. When several hundreds of sample trees are measured the reliability of OLS estimator improves more than the reliability of mixed estimators. Figs. 3 and 4 indicate that if more than 400–500 sample trees are measured, the effect of sampling error (on OLS estimator) is smaller than the effect of the bias in the prior information (on the mixed estimators).

Comparison of RMSE's of treewise volume estimates (Figs. 5 and 6) shows that mixed estimators are superior to OLS in all studied cases. Weight 0.5 gives better results than weight 0.1.

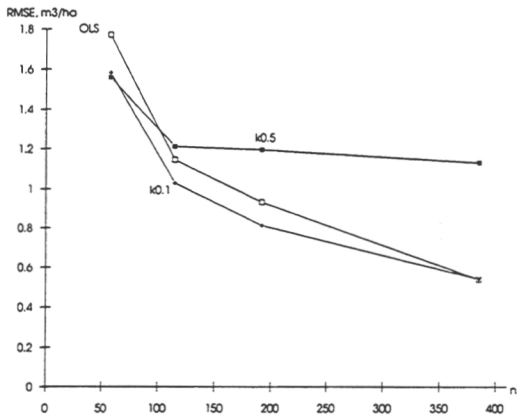


Fig. 3. Root mean square error (RMSE) of mean volume estimate of pines in Etelä-Karjala district as a function of number of sample trees. OLS = ordinary least squares estimator, k0.1 = mixed estimator with $k = 0.1$, k0.5 = mixed estimator with $k = 0.5$.

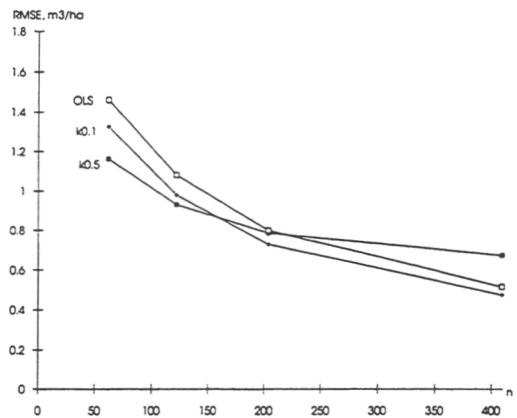


Fig. 4. Root mean square error (RMSE) of mean volume estimate of pines in Pirkka-Häme district as a function of number of sample trees. OLS = ordinary least squares estimator, k0.1 = mixed estimator with $k = 0.1$, k0.5 = mixed estimator with $k = 0.5$.

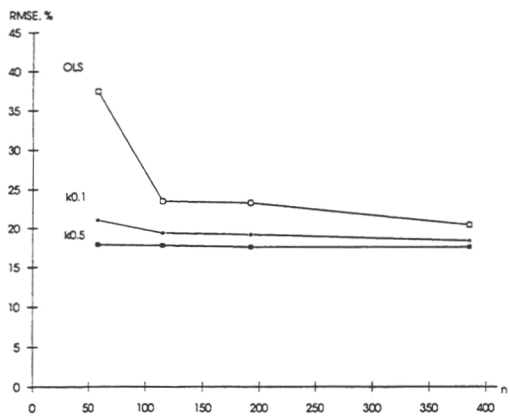


Fig. 5. Root mean square error (RMSE) of tree-wise volume estimates in Etelä-Karjala district as a function of number of sample trees. OLS = ordinary least squares estimator, k0.1 = mixed estimator with $k = 0.1$, k0.5 = mixed estimator with $k = 0.5$.

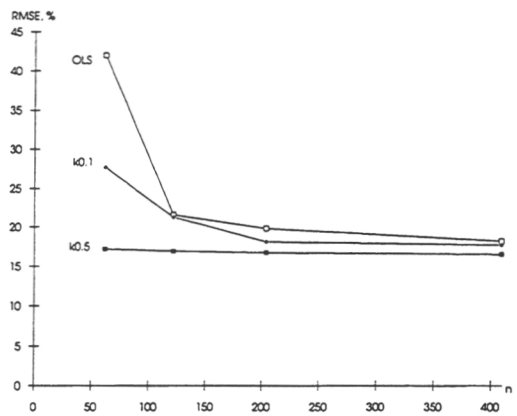


Fig. 6. Root mean square error (RMSE) of tree-wise volume estimates in Pirkka-Häme district as a function of number of sample trees. OLS = ordinary least squares estimator, k0.1 = mixed estimator with $k = 0.1$, k0.5 = mixed estimator with $k = 0.5$.

4 Conclusions

Comparison of Figs. 1 and 2 shows that using second order trend surface as a regressor improves the reliability of the model for different parts of the country. However, Fig. 2 shows that geographic variables can not totally remove the spatial correlation of residuals. Second order trend

surfaces can not handle small scale variation – “...trend surface analysis is best at showing broad features of the data” (Ripley 1981, page 35). Moving averages or kriging -methods would be useful in describing small scale geographic variation.

'Optimal' value for the weight parameter k depends on the number of sample trees and inventory area. If the bias of the prior information is high (as in district Etelä-Karjala) small weight is preferable. Tests done in this study showed that value 0.1 for k works well in most cases – reliability of mean volumes is good and RMSE of tree-wise estimates is not significantly higher than when using for example value 0.5 for k .

Naturally, it is possible to determine the optimal value for the weight of the prior information analytically, also (see e.g. Pekkonen 1983). Formula for optimal value of k would require deter-

mination of variance of residuals between sample tree areas. These sample tree areas are not fixed from one inventory to another. Furthermore, analytical formula for optimal k would work only if variation between sample tree areas would not change as a function of time. Therefore, empirical optimisation for k was regarded as a better choice.

Acknowledgements: The author wishes to thank Dr. Risto Ojansuu for his critical comments and suggestions to improve the manuscript.

References

- Cunia, T. 1986. Error of forest inventory estimates: its main main components. In: Estimating tree biomass regressions and their error. Proceedings of the Workshop on Tree Biomass Regression Functions and their Contribution to the Error of Forest Inventory Estimates, May 26–30, 1986, Syracuse, New York. p. 1–13.
- Green, E.J. & Strawderman, W.E. 1985. The use of Bayes/empirical Bayes estimation in individual tree volume equation development. *Forest Science* 31: 975–990.
- Burk, T.E. & Ek, A.R. 1982. Application of empirical Bayes/James Stein procedures to simultaneous estimation problems in forest inventory. *Forest Science* 28: 753–771.
- Johnston, J. 1972. *Econometric methods*. 2nd edition. McGraw-Hill Kogakusha, Ltd., Tokyo. 437 p.
- Kilkki, P. 1979. An outline for a data processing system in forest mensuration. *Silva Fennica* 13(4): 368–384.
- Korhonen, K.T. 1992. Calibration of upper diameter models in large scale forest inventory. Tiivistelmä: Yläläpimittamallien kalibrointi suuralueen metsäinventoinnissa. *Silva Fennica* 26(4): 231–239.
- Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae* 108. 74 p.
- Lappi, J. 1991. Calibration of height and volume equations with random parameters. *Forest Science* 37(3): 781–801.
- Meng, C.H., Tang, S.Z. & Burk, T.E. 1990. A stochastic restrictions regression model approach to volume equation estimation. *Forest Science* 36(1): 54–65.
- Pekkonen, T. 1983. Leimikon puuston tilavuuden arviointi regressioennustinta käyttäen. [Regression estimators in predicting the volume of the stand marked for cutting. In Finnish]. *Metsäntutkimuslaitoksen tiedonantoja* 86. 63 p.
- Ripley, B.D. 1981. *Spatial statistics*. John Wiley & Sons, New York. 252 p.
- Teräsvirta, T. 1981. Some results on improving the least squares estimation of linear models by mixed estimation. *Scandinavian Journal of Statistics* 8: 33–38.
- Theil, H. & Goldberger, A.S. 1961. On pure and mixed statistical estimation in economics. *International Econom. Rev.* 2: 65–78.
- Valtakunnan metsien 8. inventointi. Kenttätöön ohjeet. Pohjois-Karjalan versio. [Field instructions for the 8th National Forest Inventory of Finland. In Finnish.] The Finnish Forest Research Institute, Helsinki. 96 p.

Total of 14 references

III

Korhonen, Kari T. 1994. Calculation system for large-scale forest inventory. Metsäntutkimuslaitoksen tiedonantoja 505. The Finnish Forest Research Institute, Research Papers 505. 36 pp. ISBN 951-40-1371-9, ISSN 0358-4283.

Summary: This paper describes a calculation system for large-scale forest inventory. National Forest Inventory data from the Forestry Board District of Kainuu are used for demonstrating the calculation system. Mixed estimation is used for estimating localized volumes functions from the sample tree data measured in the inventory and data measured in a previous inventory. Ordinary least squares -technique is used for estimating growth models and volume functions by timber assortments. Estimated models and some basic statistics calculated for the test area are presented.

Keywords: forest inventory, mixed estimation, models, volume, growth

Author's address: Korhonen, Kari T. The Finnish Forest Research Institute, Joensuu Research Station, P.O.Box 68, FIN-80101 Joensuu, Finland.

Publisher: The Finnish Forest Research Institute, Department of Forest Resources, project 3001. Accepted by: Aarne Reunala, professor, 24.5.1994

Distributor: The Finnish Forest Research Institute, Joensuu Research Station, P.O.Box 68, FIN-80101 Joensuu, Finland.

ACKNOWLEDGEMENTS

Data used in this study were collected by the field groups of the project National Forest Inventory of Finland. Several persons working in the same project in the Finnish Forest Research Institute have done valuable work in further processing of the field data. Especially the work of Mr. Alpo Aarnio, Mr. Arto Ahola, Mr. Matti Kujala and Ms. Helena Mäkelä have made it possible to carry out this study. Discussions with people mentioned above and with Ms. Helena Henttonen, Mr. Antti Ihalainen, Mr. Juha Lappi, Mr. Sakari Salminen and Mr. Erkki Tomppo have helped in solving several problems met during the research work. Mr. Juha Lappi, Mr. Sakari Salminen and Mr. Erkki Tomppo have read the manuscript and given valuable critical comments. Ms. Joann von Weissenberg has revised the English text. The author wishes to thank all these people for their efforts.

Joensuu 1.6.1994 Kari T. Korhonen

CONTENTS

1. BACKGROUND	8
2. MATERIAL AND METHODS	9
2.1 Study material	9
2.2 Methods	11
3. DESCRIPTION OF SYSTEM	12
3.1 Processing sample tree data	12
3.1.1 Estimation of volumes	12
3.1.2 Estimation of volume increment	13
3.2 Estimation of volumes and increment for tally trees	15
3.2.1 Volume functions	15
3.2.2 Growth models	18
3.3 Generating reports	19
4. APPLICATION OF THE SYSTEM FOR KAINUU DISTRICT	22
4.1 Estimated models	22
4.2 Examples of calculated results for Kainuu district	23
5. DISCUSSION	25
APPENDICES	29

Concepts and Notation

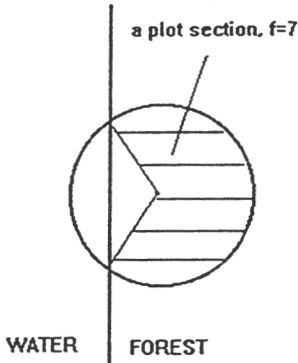
The following concepts related to the measurements of the field data are used in this paper.

sample point = a point where a relascope plot is measured

relascope plot = a set of concentric circles with each circle having a fixed radius for each diameter (the radius is a function of the cross sectional area of the tree)

restricted relascope plot = a relascope plot having a maximum radius; trees with distance from the sample point greater than the maximum radius are not tallied

plot section = section of a plot, when the plot is near a land-class boundary



plot factor = relative size of a plot; when a plot is located near a land-class boundary and the center point of the plot is on forestry land, the plot factor indicates the proportion of the whole plot circle made by the plot section on forestry land

plot stand = stand containing trees tallied on a relascope plot, usually a plot contains trees from only one stand; delineation of stands is based on the characteristics of the site and the growing stock

tally tree = a tree belonging to the (restricted) relascope plot

sample tree = a tally tree for which more detailed measurements are taken

timber tree = tally tree containing high enough timber quality for at least one saw log

non-timber tree = tally tree whose dimensions and/or quality are not enough for any saw logs

The following notations for the most common tree and stand variables are used in this paper.

d = diameter at breast height (1.3 meters from the ground), cm

h = height of a tree, m

d_6 = upper diameter of a tree measured at a height of 6 m from the ground, cm

v = stem volume of a tree from stump height to the top of the tree, dm^3

v_t = volume of the timber part of the stem, dm^3

v_w = volume of the cull part of the stem, dm^3

g = cross-sectional area of a stem at breast height, cm^2

i_d = diameter increment of the past five-year period at breast height, cm

i_h = height increment of the past five-year period, dm

b = thickness of bark (sum of measurements from two sides of the stem), mm

d_i = diameter inside the bark, cm

g_i = cross-sectional area inside the bark at breast height, cm^2

$r = v/g_i$

d_5 = diameter at the beginning of the past 5-year period, cm

h_5 = height at the beginning of the past 5-year period, m

t = age of a tree at breast height, a

g_{i5} = cross-sectional area inside the bark at breast height at the beginning of the past 5-year period, cm^2

v_5 = volume of a tree at the beginning of the past 5-year period [dm^3]

Vu = 'unit volume' of a tree ($Vu=v/g$)

SVu = 'seed volume' of a tree, ($SVu=v_j/g$)

V = volume sum of a plot (section), sum of the volumes of measured trees multiplied by the number of stems, m³/ha

F = plot factor, relative size of a plot or of a plot section

G = basal area of the growing stock, m²/ha

D = mean diameter (weighed with basal area) of the growing stock, cm

T = mean age of the dominant trees, a

S1 = 1, if the site class of the plot stand is I (i.e. IA or IB according to the Finnish tax-class system)
0, otherwise

S2 = 1, if the site class of the plot stand is 2 (i.e. II according to the Finnish tax-class system)
0, otherwise

S3 = 1, if the site class of the plot stand is 3 (i.e. III according to the Finnish tax-class system)
0, otherwise

SOIL = 1, if the soil type is mineral soil
0, otherwise

DD = effective temperature sum at the location of a plot = sum of daily mean temperatures exceeding 5 degrees Celsius, ° Celsius

DIST = distance from the coast (of the Gulf of Bothnia or the Gulf of Finland), km

RDIST = 0, if DIST > 20
1/DIST - 0.05, otherwise

Y = y-coordinate of the plot (distance from the Equator), km

YC = (Y-6620)/1000

X = x-coordinate of the plot (distance from the meridian of Greenwich, England), km

XC = (X-60)/1000

q = basal area factor (BAF)

ln(X) is the natural logarithm of variable X

1. BACKGROUND

In large-scale forest inventories sampling methods are used to obtain a representative sample of the population. In many cases it is not possible or rational to measure directly those variables that we are interested in. Therefore, mathematical methods (also methods other than simple summation) are needed to derive the statistics of final interest from the characteristics measured for the sample.

A complete calculation system of inventory results should include following components:

1. checking of field data,
2. derivation of volumes, volume growth etc. for sample trees using existing models and measured data,
3. generalization of volumes and other characteristics for tally trees,
4. summation of statistics for any chosen calculation stratum, and
5. estimation of reliability of the results.

The aim of this study was to develop a system of calculation for the National Forest Inventory of Finland (NFI). The system should be based on tested and documented methods and should cover estimation of areas of different strata (e.g. forest types), mean volume, growth and percentages of timber assortments. Estimations of future growth and cutting possibilities are, however, excluded from the system. Nor are procedures for detecting errors in the field data nor estimation of sampling error within the scope of this study.

The calculation system is tested using Kainuu district as a study area.

2. MATERIAL AND METHODS

2.1 Study material

Data from the 7th National Forest Inventory of Finland (NFI7) for the whole country and data from the 8th National Forest Inventory of Finland (NFI8) for Kainuu district (see Fig. 1) were used in this study. The NFI7 was carried out during 1977-1984 (Kuusela & Salminen 1991). The NFI8 data for Kainuu district were measured in 1992.

In both inventories the sampling method was systematic cluster sampling. In the NFI7 the distance between clusters was 8 km and each cluster consisted of 21 relascope plots. In the NFI8 in the Kainuu district the distance between clusters was 7 km and each cluster consisted of 15 restricted relascope plots.



Figure 1. Location of Kainuu district.

In both inventories, several variables describing the site and growing stock of the plot stand(s) were recorded. Tallied trees were selected with a relascope. In the NFI7 a relascope with a basal area factor of 2 was used. In the NFI8 in Kainuu district a restricted relascope plot with a basal area factor of 1.5 and a maximum radius of 12.45 m was used. In both sets of data the following variables were recorded for every tallied tree:

- tree species
- diameter
- quality class describing the quality of the stem, later called 'tally tree quality class'
- crown class.

In the NFI7 data, the living trees were divided into 3 quality classes: non-timber trees (based on the dimensions of the stem), non-timber trees (based on defects on the stem), and timber trees. In the NFI8 data a more detailed classification was used to describe, e.g. whether a timber tree is of good or poor quality.

In the NFI7, tally trees measured at four plots in every cluster were used as sample trees. In the NFI8 every 7th tallied tree was measured as a sample tree. In both sets of data the following variables were registered for sample trees:

- height
- age
- length and location of different timber assortments (=saw log quality classes A, B, and C; pulp wood; cull)
- diameter increment for the past 5-year period
- height increment for the past 5-year period (only for conifers).

In addition to above mentioned variables, diameter at six meters height and thickness of bark at breast height were registered for all sample trees in the NFI7 data and for a sub-sample of sample trees in the NFI8 data.

In the NFI8 data a variable called 'sample tree quality class' was also recorded. This variable describes the cruisers opinion of the quality of the stem after measurement of the sample tree. The 'tally tree quality class' describes the cruiser's opinion about the quality before the sample tree was measured. The cruiser may change his opinion about the quality during detailed

examination of the stem when measuring the characteristics of the sample tree.

2.2 Methods

The first phase in the calculations is to derive volume, growth and percentages of timber assortments for every sample tree measured. The volumes of the trees were calculated using volume functions of Laasasenaho (1982). Volumes of timber assortments were calculated using the taper curve models of Laasasenaho (1982) as a function of d , d_0 , and h . Volume increment of the sample trees was estimated according to the methods described by Salminen (1978) and Kujala (1980).

Regression analysis (Ordinary Least Squares, OLS) and mixed estimation were used to estimate the volume and increment of tally trees. Mixed estimation is widely used in problems requiring combination of two or more data sets (Teräsvirta 1981). Korhonen (1992, 1993) has shown that mixed estimation is efficient for combining sample tree data from two inventories.

SAS statistical software was used for studying the relationships between different measured variables in order to determine the correct form of the necessary models (SAS Institute Inc. 1989). The parameters of the models were estimated with Fortran-programs made by the author. IMSL-routines were used for matrix operations (such as inversion) in these programs (IMSL library... 1982). The reason for selecting Fortran-programs instead of available statistical software was that Fortran-programs makes it possible to simulate sampling and thus test the methods used in the calculation system. Fortran-programs were also used to derive the volumes of different timber assortments for sample trees measured.

Volume and growth of tally trees were estimated with Fortran-programs developed by the author. The treewise characteristics were summed up into statistics for the whole calculation area with SAS statistical software.

3. DESCRIPTION OF THE SYSTEM

3.1 Processing sample tree data

3.1.1 Estimation of volumes

In the NFI8 data, d and h were measured for every sample tree. Upper diameter, d_6 , however, was measured only for a sub-sample of the sample trees. The first phase in the calculation of volumes was to construct models for estimating the upper diameter of all sample trees (higher than 8 meters). Function (1) was applied as the model (Korhonen 1992) but was used in its full form only for pine and spruce. For other species, only variables d^2 , h^2 , d/t , and $\ln(\hat{d}_6)$ were significant regressors; the rest of the variables were excluded from the model.

$$\begin{aligned} \ln(d_6) = & a_0 + a_1*d^2 + a_3*h^2 + a_4*d/t + a_5*\ln(\hat{d}_6) + \\ & a_6*\ln(G) + a_7*YC + a_8*YC^2 + a_9*XC + a_{10}*XC^2 + \\ & + a_{11}*YC*XC, \end{aligned} \quad (1)$$

where d_6 = diameter measured at 6 meters height,

d = diameter at breast height,

h = height,

t = age of the tree at breast height,

\hat{d}_6 = upper diameter estimated using the taper curve model of Laasasenaho (1982) with d and h as independent variables,

G = basal area of the growing stock,

YC = relative y-coordinate of the plot (see 'Concepts and Notation'), and

XC = relative x-coordinate of the plot (see 'Concepts and Notation').

The parameters of the model (1) for each tree species were estimated using mixed estimation. In the first stage of the estimation process, first-level estimates of parameters were obtained using NFI7 data for the whole country. In the second stage, second-level estimates of parameters $a_0 - a_6$ were obtained using NFI8 data from Kainuu district (Korhonen 1992). Parameters related to the coordinates are not estimated in the second stage because the data measured in this stage are quite few and geographically not representative.

When all sample trees have measured values for variables d and h and all trees higher than 8 meters have a measured or estimated value for variable d_6 , the stem volume (from the estimated stump height to the top of the tree) can be estimated with the functions presented by Laasasenaho (1982).

Estimation of timber assortments is based on the dimensions of the stem (d , d_6 , h) and the measured lengths of different quality classes. These measurements were used to divide the stem into saw logs that fulfilled the dimension (top diameter and length) requirements. The stem was divided into saw logs by maximizing the value of the stem with a 'complete enumeration' method. In this method, all possible solutions (=combinations of saw logs of different lengths) are tested and the solution that gives the best value is chosen. This timber ruling was made with a Fortran-program in which the dimension requirements (minimum and maximum lengths and minimum diameters) and relative values of different assortments are optional parameters so that they can easily be changed.

3.1.2 Estimation of volume increment

The sample tree variables that are related to the estimation of growth are diameter and height increment during the past 5-year period and thickness of the bark. The bark is measured only for some of the sample trees. Therefore, a regression model was constructed for estimating the thickness of the bark. Function (2) was found as a suitable model for pine.

$$b = a_0 + a_1*d + a_2*h, \quad (2)$$

where b = thickness of bark.

For other species a logarithmic model (Function 3) was found to be necessary to solve the problem of heteroscedasticity.

$$\ln(b) = a_0 + a_1*\ln(d) + a_2*\ln(h). \quad (3)$$

For deciduous trees height increment was not measured. Therefore, this variable was estimated

using the tables of Ilvessalo (1948, see Kujala 1980), in which height, age, and crown class of the tree are independent variables.

When the bark models are estimated, each sample tree have measured or estimated values for variables related to volume growth: diameter and height increment and thickness of the bark. No characteristics are measured to directly describe the changes in stem form and in thickness of the bark during the past 5-year period. When calculating the volume increment these changes can be taken into account with the method presented by Salminen (1978) and Kujala (1980). In this method it is assumed that the change in v/g_i (ratio of volume and cross sectional area) during the past 5-year period can be estimated with help of a function ($v/g_i=f(h)$) estimated from the present v , g_i and h of the trees.

To describe the method of Salminen (1978) and Kujala (1980), let us note that:

$$r = v/g_i, \quad (4)$$

$r(h)$ = a function that estimates r as a function of height,

$$Vu = v/g \quad (5)$$

(Vu is the unit volume of a sample tree = volume of the tree divided by its cross sectional area),
and

$$SVu = v/g \quad (6)$$

(SVu is 'seed volume' of a relascope tree = the volume of the tree 5 years ago divided by its present cross sectional area)

Formula (6) for SVu can be further written as follows:

$$\begin{aligned}
 SVu &= g_{i5}/g * (v_j/g_{i5}) \\
 &\Leftrightarrow \\
 SVu &= g_{i5}/g * (v/g_i - (v/g_i - v_j/g_{i5})) \\
 &\Leftrightarrow \\
 SVu &= g_{i5}/g * v/g_i - g_{i5}/g * (v/g_i - v_j/g_{i5}) \\
 &\Leftrightarrow \\
 SVu &= g_{i5}/g * (g/g_i * v/g - (v/g_i - v_j/g_{i5})) \tag{7}
 \end{aligned}$$

Using notation Vu for v/g , $\hat{r}(h)$ for estimated v/g_i , and $\hat{r}(h_5)$ for estimated v_j/g_{i5} , Function (7) can be written:

$$SVu = g_{i5}/g * (g/g_i * Vu - (\hat{r}(h) - \hat{r}(h_5))). \tag{8}$$

After the 'seed volume' of a tree is estimated using Function (8) and areas in cross section and heights now and 5 years ago are measured, the volume 5 years ago can be calculated with Formula (9):

$$v_5 = g * SVu \tag{9}$$

3.2 Estimation of volumes and increment for tally trees

3.2.1 Volume functions

Functions for estimating the volume of the whole stem from stump to the top of the tree were constructed using sample trees measured in the NFI7 and the NFI8. The two data were combined using mixed estimation (Korhonen 1993). At the first stage of the construction of the volume functions, NFI7 data were used for determining the form of the models and for obtaining first-level estimates of the parameters. In a previous study (Korhonen 1993) Function 10 was shown to work well for pine.

$$v / d_2 = a_0 + a_1*d + a_2*d^2 + a_3*RDIST + a_4*\ln(G) + a_5*YC + a_6*YC^2 + a_7*XC + a_8*XC^2 + a_9*YC*XC, \quad (10)$$

where RDIST = relative distance from the sea coast (see 'Concepts and Notation').

Function (10) was also found to be satisfactory for spruce and birches. For other species, the variable RDIST was excluded from the model.

At the second stage in the construction of the volume functions, mixed estimation was used for obtaining second-level estimates of the parameters (Korhonen 1993). Only data from the calculation area (typically one district of the Central Board of Forestry in Finland) were used for the re-estimation. Because the number of sample trees measured was quite high for pine, spruce and birches, models were estimated separately for each site class. Only constant and parameters of variables d , d^2 , and $\ln(G)$ were estimated at the second stage; for other variables, first-level estimates were used (Korhonen 1993).

Regression models were also constructed for estimating the volumes of different timber assortments: timber and cull. The volume of pulp wood quality was estimated by subtracting the estimated volumes of timber and cull from the estimated stem volume. A function with a form of Equation (11) was found to be suitable.

$$v_t / d^2 = a_0 + a_1*\ln(d) + a_2*DD, \quad (11)$$

where v_t = volume of timber or cull and

DD = effective temperature sum at the location of the plot.

Because the form of the Function (11) for timber assortmentwise volumes differs markedly from the Function (10) for whole-stem volume, it was also necessary to estimate a model of a form similar to Equation (11) for stem volume. Otherwise, all errors due to the form of the functions would have been summed up in the volume estimate for pulp wood. In some cases this could even have led to negative estimates for pulp wood. Final estimates for, e.g. the saw log volume of a tally tree were then obtained with formula (12).

$$\hat{v}_t = \hat{v}_1 / \hat{v}_2 * \hat{v}_1, \quad (12)$$

where \hat{v}_t = final estimate for v_t ,

\hat{v}_1 = estimate for v_t obtained with the model for the form of Equation (11)

\hat{v}_1 = estimate for v obtained with the model for the form of Equation (10), and

\hat{v}_2 = estimate for v obtained with the model for the form of Equation (11).

Because logging rules have changed since the NFI7, it was not possible to use the NFI7 data as prior information. Timber assortmentwise models were estimated using OLS and sample tree data measured from the calculation area.

Naturally, the proportions of timber assortments vary markedly by tree class. Therefore, the sample tree data were grouped by tree classes as follows.

1 = non-timber tree

2 = a good saw log tree

3 = a poor saw log tree.

As mentioned in Section 2.1, in the field measurements two different codes for tree class were recorded for the sample trees. The first of these describes the quality 'at first glance', which refers to the way it is coded for tally trees (later called 'tally tree quality class'). The second one is coded according to more detailed sample tree measurements (later called 'sample tree quality class'). Usually, the codes match, but in some cases the cruiser may change the quality code when taking sample tree measurements. Thus, a sample tree with tally tree class code 'non-timber tree' can include saw log quality, and vice versa.

Only trees with the sample tree quality class code 'good saw log tree' or 'poor saw log tree' were used in estimating the above mentioned regression models for timber assortment volume. At the application stage tally trees were grouped into the respective quality classes according to the tally tree quality class. To avoid possible bias caused by differences in the two quality classifications an adjustment was made as follows. For all sample trees the mean of the measured timber volume and the mean of the estimated timber volume were calculated by tree species. The timber volume estimate of a single tree was then multiplied by the ratio of these

two means. A similar adjustment was made for the estimates of cull volume.

3.2.2 Growth models

Model (13) was selected as a basic model for estimating the volume increment of the past 5-year period. For some species not all the variables in the equation were significant regressors. In these cases, only significant variables were used (see Section 4.1).

$$\ln(i_v/d^2) = a_0 + a_1*d + a_2*\ln(d) + a_3*G + a_4*DD + a_5*D/T + a_6*S1 + a_7*S2 + a_8*S3 + a_9*SOIL \quad (13)$$

where S1, S2, S3 are dummy variables for different site classes, and SOIL is a dummy variable to separate mineral soils from peatlands.

When the growth models were estimated, trees growing on poorly productive land were separated from trees growing on forest land for two reasons:

1. the growth of trees growing on poorly productive land is markedly different from the growth of trees on forest land, and
2. most of the variables describing the growing stock are not measured on poorly productive land.

Equation (14) was used as a growth model for trees on poorly productive land.

$$\ln(i_v/d^2) = a_0 + a_1*d + a_2*\ln(d) \quad (14)$$

The natural logarithm of i_v/d^2 is used as an dependent variable in models (13) and (14). In applications, unbiased estimates for i_v/d^2 in the arithmetic scale are needed. The most common way to correct the bias due to the non-linear transformation is to add the term $1/2*MSE$ to the logarithmic estimate. This correction is based on the assumption that the residuals of the logarithmic model are normally distributed. If this does not hold true, following estimator can be justified according to Snowdon (1991):

$$\hat{y}_i = c * \exp(\hat{\mu}_i), \quad (15)$$

$$\text{where } c = \frac{\sum y_i}{\sum \exp(\hat{\mu}_i)},$$

y_i = measured value for observation i , and

$\hat{\mu}_i$ = estimated logarithmic value for observation i .

Estimator (15) was used in this study to correct the bias due to the non-linear transformation in growth models.

3.3 Generating reports

Using the models described in Section 3.3, the field data are transformed into a file, which is suitable for further processing with SAS statistical package. In this file, each plot stand has one record for characteristics describing the location of the plot and site and the growing stock of the stand; and one record for each measured tally tree. The record of a tally tree contains following data:

- tree species
- tally tree quality class
- d
- \hat{v}/d^2
- \hat{v}_w/d^2
- \hat{v}_n/d^2
- \hat{i}_w/d^2 .

(Volumes and and growth are divided by d^2 before they are stored in the file in order to decrease the errors due to rounding off).

When area estimates are calculated for different strata, e.g. site classes, the sample point is used as one observation. Area estimates are obtained with Formula (17) (Salminen 1993).

$$\hat{A}_i = m_i/M * \text{AREA}, \quad (16)$$

where \hat{A}_i = estimated area for stratum i,

m_i = number of sample points in stratum i,

M = total number of sample points on the land in the
calculation area, and

AREA = land area of the calculation area.

Volume sum and mean statistics for a calculation area or different strata of the area are compiled by summing the volumes and plot factors of the plots over the strata in question. The mean volume is calculated by dividing the volume sum by the sum of the plot factors. No mean volumes for single stands of a relascope plot are needed. In fact, it is impossible to calculate such mean volumes in the NFI data for stands that do not cover the whole relascope plot.

In the case of a restricted relascope plot with maximum radius of 12.45 m (as in NFI8 in Kainuu district) the volume sum of a plot (section) is calculated as follows.

$$V = \sum n_j * v_j, \quad (17)$$

where V = volume sum of the plot (section),

$$n_j = \begin{cases} q*4/\pi * 10000/d^2, & \text{if } d_j < 30.5 \\ 20.53579, & \text{otherwise} \end{cases}$$

q = basal area factor (BAF)

v_j = volume of a tally tree (dm^3)

\sum = summation over all trees measured on the plot.

It should be noted, that Formula (17) does not differentiate between whole plots and plots sections. The size of a plot is taken into account later when the means or sums are calculated for the calculation strata.

After the volume sums are calculated for each part plot, mean volume estimates for different strata of a inventory area are obtained with Formula (18).

$$\bar{V}_i = \frac{\sum V_i}{\sum F_i}, \quad (18)$$

where \bar{V}_i = estimated mean volume for stratum i,

V_i = volume sum of a plot (section) (see Formula 17), and

F_i = plot factor of the plot (section) i

Σ = summation over all plots in a calculation stratum of the inventory area.

4. APPLICATION OF THE SYSTEM FOR KAINUU DISTRICT

4.1 Estimated models

The upper diameter functions estimated using sample trees from Kainuu district are in Appendix 2. Separate models were estimated for following species: pine (*Pinus sylvestris*), spruce (*Picea abies*), white birch (*Betula pendula*), silver birch (*B. pubescens*), aspen (*Populus tremuloides*), alder (*Alnus incana* and *A. glutinosa*). As described earlier, NFI7 data were used as prior information for some of the parameters and only as information for other parameters (Korhonen 1992).

The bark models for sample trees of Kainuu district are in Appendix 3. Separate models for pine, birches (no difference between white and silver birch), aspen and alder were used. The models were estimated using NFI7 and NFI8 sample tree data from Kainuu district.

Volume functions estimated for different species for Kainuu district are in Appendix 4. For pine, spruce and birches the final parameter estimates were calculated separately for 4 site class groups. For aspen, alder and the group of other species the sample tree data were too few to distinguish between site classes. The models presented in Appendix 4 are not logical for small sized trees. Therefore, general volume functions were used for trees with $d < 3$ cm.

The regression models for timber assortmentwise volumes are presented in Appendix 5. The data for pine and spruce were divided into two site class groups. Separate models for timber and cull volume models were estimated for the two groups. The data for birches were too few to make a distinction between site classes. For aspen only few sample trees were coded as timber quality trees, and therefore no model for timber volume were estimated. At the application stage, timber volume model of birches were used for those aspens that were coded as timber quality trees. For other species, no timber trees were measured; models were estimated only for cull volume. The correction factors for timber assortmentwise volumes (see Section 3.2.1) are also given in Appendix 5.

The estimated models for the past 5-year volume growth of different species are in Appendix 6. Separate models for trees growing at forest land and poorly productive land were estimated. For forest land separate models for pine, spruce, birches and aspen were used. For poorly

productive land birches and aspen were combined to the group of other species. The correction terms (see Equation (16)) are also given in Appendix 6.

4.2 Examples of calculated statistics for Kainuu district

According to the statistics of the National Board of Survey, the total land area of Kainuu district is 2 156 690 ha. Estimated area of forest land is 1 664 015 ha. Figures 2 and 3 present two examples of estimated percentages of different strata. In Figure 2 is the distribution of forest land by dominant species. Figure 3 presents the age class distribution on the forest land in Kainuu district.

The estimated mean and total volumes and timber assortmentwise mean and total volumes by species on forest land and poorly productive land are given in Table 1. The growth statistics obtained using the calculation system are presented in Table 2.

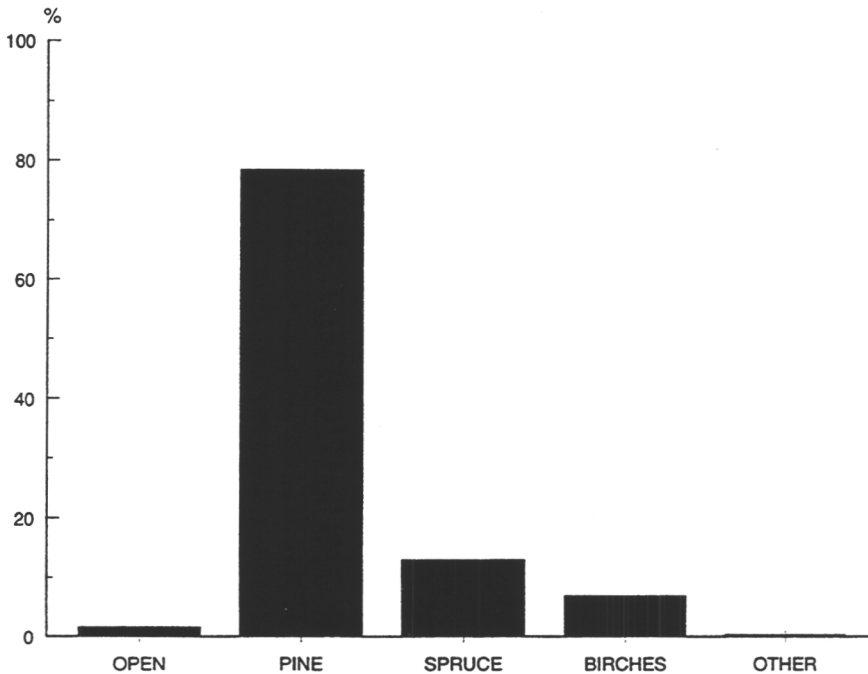


Figure 2. Dominance of tree species on forest land.

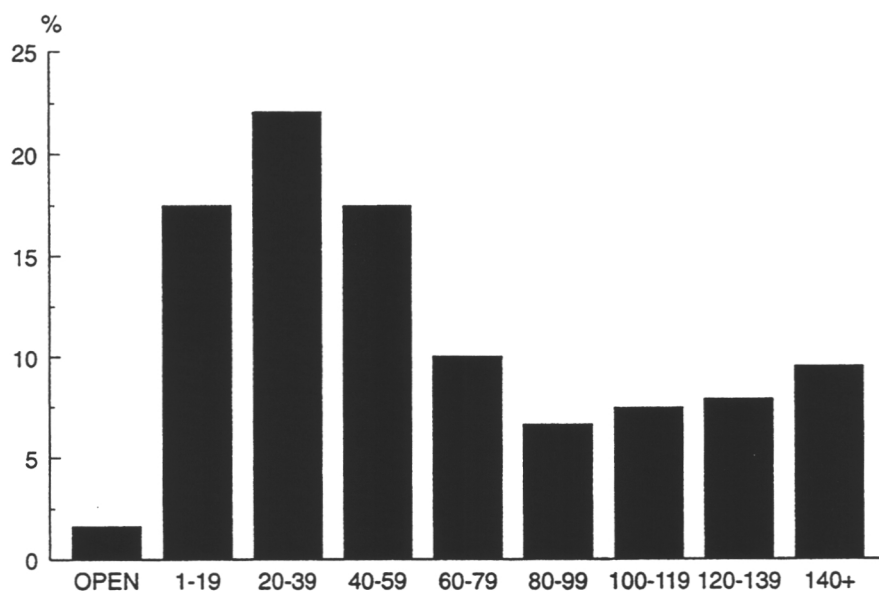


Figure 3. Age structure on forest land

Table 1. Mean and total values of stem volume and timber assortmentwise volumes by species for forest and poorly productive land.

	Mean volumes, m ³ /ha			Total volumes, 1000 m ³		
	saw log	pulp	total	saw log	pulp	total
Pine	13.3	21.9	37.7	25089	41288	71062
Spruce	6.1	9.8	16.9	11427	18427	31947
Birches	0.2	6.2	8.7	414	11763	16382
Others	0.0	0.7	1.1	29	1273	2009
All	19.8	44.8	64.4	36959	72750	121401

Table 2. Mean and total value estimates for growth by species for forest and poorly productive land.

	Pine	Spruce	Birches	Others	All
m ³ /ha	1.35	0.41	0.38	0.06	2.20
1000 m ³	2552	782	715	104	4153

5. DISCUSSION

The system of calculation presented in this paper differs from the present calculation system for NFI in the following ways:

1. Stem volume and the timber assortmentwise volumes are generalized for tally trees using regression models instead of classwise mean values.
2. Growth estimates are calculated for tally trees with a method similar to that for estimating the timber assortmentwise volumes. In the present calculation system growth is not generalized for tally trees (growth is estimated using growth percentages estimated from the sample tree data and diameter distributions estimated from the tally tree data).

The above-mentioned solutions make the system flexible. Area, volume and growth estimates can easily be calculated for any sub-class of the data. Plotwise volume and growth estimates can easily be used as a 'ground truth' for processing satellite images (Tomppo 1992).

Characteristics of 'final interest', such as timber assortmentwise volumes and volume growth, were used as dependent variables when the models were estimated from the sample tree data. Another possibility would have been to use variables describing the dimensions (d , d_6 , h , b , i_d , i_b) and quality (lengths of different quality classes) as independent variables. The models would then give estimates of all sample tree variables for every tally tree. These estimates could be used as independent variables for further estimations, e.g. of the volume or growth of tally trees. Various modifications of this method described by Kilkki (1979) are widely used in inventory systems. The method has one serious drawback, however: residual variance of the different models must be taken into account when, e.g., using estimated h , i_d , and i_b as independent variables in estimating volume increment (Kilkki 1979). The joint distributions of the errors of

the different models are difficult to estimate. Therefore, after some trials this approach was rejected.

As stated, in this system of calculation sample tree variables such as height and age are not generalized for tally trees. Therefore, the data generated for tally trees cannot be used to predict the future development of forests with simulation systems like the Finnish Mela-system (Siitonen 1983). Further studies are needed if, for example, the grid method (Holm et al. 1979) is applicable for generalizing the sample tree characteristics for tally trees in such a way that the results can be used as a basis both for calculating unbiased inventory statistics and for simulating the future development of the trees.

LITERATURE

- Holm, S., Hägglund, B. & Mårtensson, A. A method for generalization of sample tree data from the Swedish National Forest Survey. Swedish University of Agricultural Sciences. Department of Forest Survey. Report 25. 94 pp.
- IMSL library reference manual. Edition 9. 1982. IMSL Inc. Houston, Texas.
- Kilkki, P. 1979. Outline for a data processing system in forest mensuration. *Silva Fennica* 13(4):368-384.
- Korhonen, K.T. 1993. Mixed estimation in calibration of volume functions of Scots pine. *Silva Fennica* 27(4):269-276.
- Korhonen, K.T. 1992. Calibration of upper diameter models in large-scale forest inventory. *Silva Fennica* 26(4):231-239.
- Kujala, M. 1980. Runkopuun kuorellisen tilavuuskasvun laskentamenetelmä. Summary: A calculation method for measuring the volume growth over bark of stemwood. *Folia Forestalia* 441. 18 pp.
- Kuusela, K. & Salminen, S. 1991. Suomen metsävarat 1977-1984 ja niiden kehittyminen 1952-1980. Forest resources of Finland in 1977-1984 and their development in 1952-1980. *Acta Forestalia Fennica* 220. 84 pp.
- Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae* 108. 74 pp.
- Salminen, S. 1993. Eteläisimmän Suomen metsävarat 1986-1988. Summary: Forest resources of Southernmost Finland. *Folia Forestalia* 825. 111 pp.
- Salminen, S. 1978. Increment calculations on the basis of relascope sampling in the Finnish National Forest Inventory. Paper presented in IUFRO Meeting of June 18-25, 1978 in Bucharest, Romania. Subject Group S4.02. 7 p. The Finnish Forest Research Institute. Helsinki, Finland.
- SAS Institute Inc., SAS/STAT User's guide, Version 6, Fourth Edition, Volumes 1 and 2, Cary, NC: SAS Institute Inc., 1989.
- Siitonen, M. 1983. A long term forestry planning system based on data from the National Forest Inventory of Finland. In: Forest inventory for improved management. Proceedings of the IUFRO Subject Group 4.02 Meeting in Finland, September 5-9, 1983. University of Helsinki. Department of Forest Mensuration and Management. Research Notes 17:195-207.
- Snowdon, P. 1991. A ratio estimator for bias correction in logarithmic regression. *Canadian Journal of Forest Research* 21(5):720-724.
- Teräsvirta, T. 1981. Some results on improving the least squares estimation of linear models by mixed estimation. *Scandinavian Journal of Statistics* 8:33-38.

Tomppo, E. 1992. Multi-source National Forest Inventory of Finland. In: Proceedings of Ilvessalo Symposium on National Forest Inventories. Finland 17-21 August 1992. Nyysönen, A., Poso, S. & Rautala, J. (ed.). The Finnish Forest Research Institute. Research Papers 444:52-59.

Valtakunnan metsien 8. inventointi. Kenttätöohjeet, Kainuun ja Pohjois-Pohjanmaan veriso. 1992. [Field instructions for the field work of the 8th National Forest Inventory of Finland at Kainuu and Pohjois-Pohjanmaa districts. In Finnish.] Manuscript. 67 pp + appendices. The Finnish Forest Research Institute. Helsinki, Finland.

Valtakunnan metsien inventoinnin kenttätöohjeet. VMI7. 1978. [Field instructions for the field work of the 7th National Forest Inventory of Finland. In Finnish.] Manuscript. 59 pp + appendices. The Finnish Forest Research Institute. Helsinki, Finland.

Appendix 1. Functions for ratio of volume over bark and cross sectional area inside bark ($r = v/g_i$) (Kujala 1980).

$$r(h) = 0.39 + 0.39*h + 2/(h.1-3) + 0.77*\sqrt{(h-1.3)}$$

$$r(h) = 0.44 + 0.355*h + 2/(h.1-3) + 0.65*\sqrt{(h-1.3)}$$

$$r(h) = 0.48 + 0.48*h + 3.5/(h.1-3)$$

Appendix 2. Upper diameter models for Kainuu district.

Model:

$$\ln(d_6) = a_0 a_0 + a_1 * d_2 + a_2 * h_2 + a_3 * t + a_4 * d/t + a_5 * \ln(\hat{d}_{6L}) + \\ a_6 * \ln(G) + a_7 * YC + a_8 * YC^2 + a_9 * XC + a_{10} * XC^2 + \\ + a_{11} * YC * XC$$

Parameters:

	Pine	Spruce	White birch	Silver b.	Aspen	Alder
a ₀	-0.0721	0.0659	0.0602	0.2173	0.0191	-0.0829
a ₁	-4.684E-5	7.169E-5	-2.857E-6	1.278E-4	1.246E-5	1.148E-
a ₂	1.489E-4	-8.169E-5	8.577E-5	2.375E-4	1.324E-4	-1.100E-
a ₃	1.281E-4	-1.115E-4	-3.214E-4	-4.367E-4	-0.1201	2.168E-
a ₄	-0.1182	-0.0746	-0.1000	-0.3855	-	0.1436
a ₅	1.0081	0.9588	0.9534	0.9473	0.9874	0.9308
a ₆	-1.147E-3	9.742E-3	0.1471	-0.0463	-0.2063	0.0358
a ₇	0.0941	0.1476	-	-	-	-0.0261
a ₈	-2.964E-3	-0.2212	-0.0836	-0.0289	0.0764	0.2898
a ₉	0.0942	0.0421	0.2444	-0.1988	0.2153	-0.2811
a ₁₀	-5.546E-3	-0.0658	-0.1121	0.2827	-0.3818	0.5933
a ₁₁	-0.1076	0.0351	-0.2176	0.2297	0.1430	-0.6086

Correction terms for anti-logarithm transformation

.0034	0.0042	0.0049	0.0097	0.0086	0.0086
-------	--------	--------	--------	--------	--------

Appendix 3. Bark models for Kainuu district.

Pine

Model:

$$b = a_0 + a_1*d + a_2*h$$

Parameters:

Pine

a₀ 2.437408a₁ 1.070841a₂ -0.213063

Other species:

Model:

$$\ln(b) = a_0 + a_1*\ln(d) + a_2*\ln(h)$$

Parameters:

Spruce

Birches

Aspen

Alder

Others

a₀ 0.632961

0.127477

-0.058233

0.172857

-0.246579

a₁ 1.021860

1.037599

1.012449

0.807302

1.107772

a₂ -0.323278

-0.175557

Correction terms for anti-logarithm transformation

0.026473

0.052313

0.047472

0.068787

0.066722

Appendix 4. Volume models for Kainuu district

Model:

$$v/d2 = a_0 + a_1*d + a_2*d2 + a_3*RDIST +$$

$$a_4*\ln(G) + a_5*YC + a_6*YC2 + a_7*XC +$$

$$a_8*XC2 + a_9*XC*YC$$

Parameters:

Pine

	Site 1	Site 2	Site 3	Site 4
a ₀	-0.1757	-0.1249	-0.0712	0.0441
a ₁	0.0264	0.0238	0.0198	0.0146
a ₂	-2.825E-04	-2.164E-04	-1.377E-04	-4.935E-05
a ₃	0.0879	0.0761	0.0666	0.0353
a ₄	-0.0720	-0.0720	-0.0720	-0.0720
a ₅	0.0249	0.0249	0.0249	0.0249
a ₆	-0.0167	-0.0167	-0.0167	-0.0167
a ₇	0.3726	0.3726	0.3726	0.3726
a ₈	-0.2197	-0.2197	-0.2197	-0.2197
a ₉	-0.2476	-0.2476	-0.2476	-0.2476

Spruce

a ₀	-0.0893	3.115E-03	0.0237	0.1862
a ₁	0.0327	0.0286	0.0265	0.0155
a ₂	-3.967E-04	-3.216E-04	-3.720E-04	-3.144E-05
a ₃	0.0618	0.0405	0.0401	-4.315E-03
a ₄	-0.0557	-0.0557	-0.0557	-0.0557
a ₅	0.0392	0.0392	0.0392	0.0392
a ₆	-0.1380	-0.1380	-0.1380	-0.1380
a ₇	0.2755	0.2755	0.2755	0.2755
a ₈	-0.3008	-0.3008	-0.3008	-0.3008
a ₉	-0.2573	-0.2573	-0.2573	-0.2573

Birches

a ₀	0.0343	0.0449	0.1013	0.1421
a ₁	0.0194	0.0150	0.0130	0.0117
a ₂	-1.476E-04	-2.589E-05	-1.084E-05	9.621E-05
a ₃	0.0449	0.0477	0.2686	3.202E-03
a ₄	-0.0343	-0.0343	-0.0343	-0.0344
a ₅	0.0289	0.0289	0.0289	0.0289
a ₆	0.0191	0.0191	0.0191	0.0191
a ₇	0.5617	0.5617	0.5617	0.5617
a ₈	-0.3332	-0.3332	-0.3332	-0.3332
a ₉	-0.6914	-0.6914	-0.6914	-0.6914

Appendix 4 continues...

Aspen, all site classes

a_0 0.1148
 a_1 0.0247
 a_2 -3.010E-04
 a_3 -
 a_4 0.0208
 a_5 0.0169
 a_6 0.1352
 a_7 0.3626
 a_8 -0.0835
 a_9 -0.8237

Other species, all site classes

a_0 0.1893
 a_1 6.795E-03
 a_2 -1.286E-04
 a_3 -
 a_4 3.807E-03
 a_5 0.0793
 a_6 -0.2261
 a_7 0.2853
 a_8 -0.3899
 a_9 -1.1990

Appendix 5. Models for timber assortmentwise volumes for Kainuu district.

Model:

Timber volume

$$vt/(10*d2) = a_0 + a_1*\ln(d) + a_2*DD$$

Cull volume

$$vw/(10*d2) = a_0 + a_1*\ln(d) + a_2*DD$$

Total volume

$$vt/(10*d2) = a_0 + a_1*\ln(d) + a_2*DD$$

Parameters

Pine, non-timber trees

	Site 1 and Site 2			Site 2 and Site 3		
	Timber	Cull	Total	Timber	Cull	Total
a_0		-2.740E-3	-0.0182		-3.447E-3	-0.0180
a_1	0.0663		0.0233	0.0713		0.0223
a_2	5.418E-6		1.065E-04	3.961E-6		6.580E-5

Pine, good quality timber trees

	Site 1 and Site 2			Site 2 and Site 3		
	Timber	Cull	Total	Timber	Cull	Total
a_0	-0.0572	5.122E-4	-4.881E-3	-0.0633	8.366E-4	0.0168
a_1	0.0351	4.687E-3	0.0232	0.0357	-2.816E-3	0.0254
a_2	1.764E-04	2.539E-6	9.187E-5	1.553E-4	2.465E-6	8.462E-5

Pine, poor quality timber trees

	Site 1 and Site 2			Site 2 and Site 3		
	Timber	Cull	Total	Timber	Cull	Total
a_0	-0.0116	-9.316E-4	3.245E-3	-0.0325	-1.3096E-3	-0.0301
a_1	0.0146	0.0537	0.0199	0.0196	0.0559	0.0291
a_2	1.292E-4	6.058E-6	9.331E-5	8.893E-5	5.5309E-6	9.981E-5

Spruce, non-timber trees

	Site 1 and Site 2			Site 2 and Site 3		
	Timber	Cull	Total	Timber	Cull	Total
a_0		-2.460E-3	-0.0271		-3.640E-3	-0.0137
a_1	0.0654		0.0274	0.0796		0.0207
a_2	7.298E-6		7.031E-5	5.179E-6		6.134E-5

Spruce, good quality timber trees

	Site 1 and Site 2			Site 2 and Site 3		
	Timber	Cull	Total	Timber	Cull	Total
a_0	-0.1240	-3.500E-4	-0.0231	-0.1054	7.606E-4	-0.0169
a_1	0.0550	0.0242	0.0277	0.0476	-3.107E-3	0.0241
a_2	7.424E-5	9.236E-7	7.394E-5	7.313E-5	1.154E-6	7.060E-5

Spruce, poor quality timber trees

	Site 1 and Site 2			Site 2 and Site 3		
	Timber	Cull	Total	Timber	Cull	Total
a_0	-0.0404	-3.026E-3	4.054E-3	0.0342	-1.478E-3	0.0529
a_1	0.0231	0.1066	0.0185	-1.041E-3	0.0708	1.734E-3
a_2	1.335E-4	5.733E-6	6.860E-5	9.959E-5	7.944E-6	8.710E-5

Appendix 5 continues...

Birches, non-timber trees

	Site 1 and Site 2			Site 2 and Site 3		
	Timber	Cull	Total	Timber	Cull	Total
a ₀		-8.520E-4	-0.0206		-6.973E-4	-0.0159
a ₁		0.0574	0.0248		0.0570	0.0212
a ₂		2.476E-5	7.342E-5		3.493E-5	5.192E-5

Birches, good and poor quality timber trees

	Site 1 - 4		
	Timber	Cull	Total
a ₀	1.069E-3	1.091E-3	0.0490
a ₁	0.0108	-2.698E-3	3.270E-3
a ₂	1.036E-4	3.038E-6	6.943E-5

Aspen

	Site 1 - 4		
	Timber	Cull	Total
a ₀		1.049E-3	-0.0109
a ₁		0.1044	0.0214
a ₂		1.972E-4	5.124E-5

Alder

	Site 1 - 4		
	Timber	Cull	Total
a ₀		4.753E-3	0.0179
a ₁		0.0360	7.404E-3
a ₂		8.392E-5	5.452E-5

Other species

	Site 1 - 4		
	Timber	Cull	Total
a ₀		0.0189	0.0274
a ₁		-0.1023	-5.770E-4
a ₂		1.458E-4	1.264E-5

Correction terms

	Pine	Spruce	Birches	Aspen	Alder	Others
timber	0.952	0.939	0.781	0.681		
cull	0.994	0.975	0.963	0.989	0.985	0.949

Appendix 6. Volume growth models for Kainuu district.

A. Forest land

Model:

$$\ln(iv/d2) = a_0 + a_1*d + a_2*\ln(d) + a_3*G + a_4*DD + a_5*D/T + a_6*S0 + a_7*S1 + a_8*S2 + a_9*S3 + a_{10}*SOIL$$

Parameters:

	Pine	Spruce	Birches	Aspen	Others
a ₀	-4.3044	-6.3137	-5.5625	-1.6637	-1.5368
a ₁	-0.0526				
a ₂	0.4547	0.1150	-0.1783	-0.4314	-0.4879
a ₃	-0.0312	-0.0195	-0.0403		
a ₄	1.1839	2.8326	3.2296		
a ₅	1.8928	3.5339	2.5957		
a ₆	0.6923	0.4706	0.6429		
a ₇	0.6100	0.4047	0.2326		
a ₈	0.5005	0.4047	0.1427		
a ₉	0.3282	0.3782			
a ₁₀	-0.1210	-0.3810	0.2448		

Correction terms

Pine	1.086
Spruce	1.140
Birches	1.073
Aspen	1.045
Others	1.372

B. Poorly productive land

Model:

$$\ln(iv/d2) = a_0 + a_1*\ln(d)$$

Parameters:

	Pine	Spruce	Others
a ₀	-2.3182	-3.0830	-1.7185
a ₁	-0.3738	-0.3218	-0.6303

Correction terms

Pine	1.360
Spruce	1.004
Others	1.360

**APPLICATION OF NEAREST-NEIGHBOR REGRESSION FOR
GENERALISING SAMPLE TREE INFORMATION**

Kari T. Korhonen & Annika Kangas

Abstract

Nearest-neighbor regression was tested for generalising sample tree information in data from the national forest inventory of Finland. The following variables were found to be good regressors: stem diameter, mean diameter, density and age of growing stock and plot location. The nearest-neighbor estimator appears to maintain well the natural variation of the variables to be estimated. Reliable volume and height estimates can be obtained even when using only one nearest neighbor. Increasing the number of neighbors improves the accuracy of estimates.

Keywords: inventory, models, nonparametric regression, nearest-neighbor regression, volume.

1. INTRODUCTION

The forests of Finland are inventoried systematically to produce information for strategic planning. One aim of the national forest inventories (NFIs) is to monitor changes in the forest. For this purpose, basic statistics on the area, volume and growth of different strata of forests are estimated. Another aim of NFI is to forecast the future development of forests, mainly to estimate cutting possibilities.

Two-phase sampling has been found to be an effective inventory method for inventories of large areas. In the NFIs conducted in Finland the first-phase sample consists of a large number of trees (tally trees) for which diameter at breast height is measured and some other characteristics are registered. The second phase sample (sample trees) consists of trees measured more thoroughly. The characteristics measured of the sample trees include height, age, additional diameter at 6 m height, diameter and height increment, bark thickness and the lengths of different timber assortments.

The idea of measuring sample trees is to establish 'local tables' (formulated in the form of functions) for volume by timber assortments and volume increment. These functions are applied in estimating the volumes and increment of the tally trees. Once the volume and increment of each measured tree has been estimated, calculating parameters such as mean or total volume and growth for a given calculation area is a matter of simple summation (e.g. Korhonen 1994). Satellite images and other auxiliary data can be used for improving the reliability of estimates for small areas (Tomppo 1992). When NFI data are used to estimate forest parameters for small areas, the need for accurate 'local tariff tables' becomes more pronounced.

Inventory data can be used in simulating the future development of forests with the help of increment, mortality and felling models. In Finland, these models have been brought together in the MELA system. In the MELA system, the forest area is described by means of sample plots. The plot data include description of the site and

growing stock. The growing stock is described by means of trees for which diameter, height, age and some other characteristics are stored in the system. The development of stands is predicted via sample trees on the sample plots. The development of the trees is predicted according to the characteristics of each tree and its site. Increment is handled using diameter and height increment models. (Siitonen 1983).

When the NFI data are used as input data for the MELA system, several models are needed to estimate the sample tree characteristics for the tally trees. Height and age are the most important of these variables. Since estimated values are used as regressors in increment models, it is necessary for the estimates to be as accurate as possible.

When simulation purposes are considered, the aim in generalising the sample tree data for the tally trees is to create realistic images of forests with the incomplete tally tree and plot data as the basis. Ranneby & Svensson (1991) list requirements for methods that are used in generalising sample tree data for the tally trees:

"1. Estimates of, for example, total volume and volume per hectare for arbitrary groups, are unbiased.

2. Estimates for single plots are unbiased and the variation between plots is realistic.

3. Estimates for calibered trees are unbiased and the within-plot variation between calibered trees is realistic.

4. Estimates of different tree variables for each single tree harmonize. Consequently, estimates of volume and volume increment for a calibered tree should not result in an unrealistic estimate of volume increment percentage.

5. The generalization method is easy to carry out."

The purpose of this study is to test the application of the nearest-neighbor regression for generalising the sample tree data. Estimation of stem volume and sample tree characteristics were of main interest in this study.

2. MATERIAL

The primary data used in this study were the data on the pine sample trees of the 8th NFI (NFI8) conducted eastern Finland (see Figure 1). Only sample trees growing on site class II were included in the data. The data consist of 2063 pines measured on 375 plots. The pine sample trees from the 7th NFI (NFI7) of the same area were used as the additional data. These data consist of 2465 pines. The NFI7 data used in this study were collected in 1977-1980 and the NFI8 data in 1986-1990.

The following characteristics measured for the sample trees were used in this study:

- diameter
- height
- upper diameter measured at the height of six meters and
- age.



Figure 1. The study area.

The stem volumes were calculated using the measured dimensions and the volume functions of Laasasenaho (1982). The stem volume of the sample trees in the data of the 8th NFI varied from 0.8 dm³ to 3078.5 dm³ with the mean value being 310.8 dm³ and standard deviation 302.8 dm³.

Several variables describing the site and growing stock were registered for each plot in the NFI data. These variables included location, altitude, site class, soil type, basal area of growing stock, dominant tree species, mean diameter and age of growing stock, etc.

3. METHODS

The nearest-neighbor estimator may be presented as follows (modified from Altman 1992):

$$\hat{Y}(X_j) = \frac{1}{k} \times \sum_{i \in K(X_j)} Y_i \quad (1)$$

where y is the dependent variable,

X_j is the value of independent variable x (or a vector of values of several independent variables),

$K(X_j)$ is a neighborhood of the point defined by X_j , and

k is the number of neighbors (data points in $K(X_j)$).

Estimator (1) is a simple mean of the neighboring observations. The nearest-neighbor estimator uses a neighborhood consisting of a constant (k) number of observations. If the size of the neighborhood is of constant width and number of observations varies, the estimator is called the Kernel estimator (Härdle 1989, Altman 1992).

A measure of distance is needed to define the neighborhood of a given point. The Euclidean distance is a widely used measure of distance (see e.g. Tokola 1990, Tomppo 1992). In this study, the distance function was defined as:

$$D(X_j, X_i) = \sqrt{\sum_{l=1}^p a_l (x_{jl} - x_{il})^2} \quad (2)$$

where X_j is a vector of the values of the independent variables at point j (the point to be estimated),

X_i is a vector of the values of the independent variables for observation i in the data,

x_{jl} (x_{il}) is an element of the vector X_j (X_i),

a_l are parameters and

p is the number of independent variables.

The bias of the nearest neighbor estimator can be reduced using weighted averages (Altman 1992). Let us note:

$$D = D(X_j, X_i).$$

A nearest-neighbor estimator using weight function is in Formula (3) (Härlde 1989, Altman 1992).

$$\hat{Y}(X_j) = \frac{\sum_{i \in K(X_j)} W(D) Y_i}{\sum_{i \in K(X_j)} W(D)} \quad (3)$$

where $W(D)$ is the weight function.

In this study, Function (4) was used as the weight function (see e.g. Tomppo 1992).

$$W(D) = 1/(1+D). \quad (4)$$

The 'optimal' values of the parameters of the distance function (2) can be searched for using the cross validation method (Härdle 1989, Altman 1992). In this method, each observation is predicted with data excluding the observation itself. The estimator with the lowest mean of squared residuals is regarded to be the best. In this study, the standard deviation and mean of the residuals of volume estimates are presented for comparing different parameter values. Additionally, within-plot and between-plot standard deviations of height estimates are presented to demonstrate how realistic the variation of the estimates is.

The nearest-neighbor estimator were also tested with the NFI7 data. In these tests, the sample trees of the NFI7 were used as the data points (i.e. for "estimating the model") and the sample trees of NFI8 as the test material. The mean and standard deviation of the residuals of the volume estimates for the NFI8 sample trees were applied as the criteria for comparing the

different parameter values of the distance function (2). The between-plot and within-plot variance components of the height estimates were also estimated. The correlation of the height and age estimates were calculated to study how well the estimates of the different characteristics harmonised.

4. RESULTS

Following variables were found to predict the stem volume and sample tree characteristics (height and age):

- diameter (d)
- location (y and x-coordinates)
- mean age of growing stock (T)
- basal area of growing stock (G)
- mean diameter of growing stock (D_m).

The second power of diameter (d^2) was used as independent variable insted of diameter because the relation between d^2 and volume is more close to linear than the relation between d and volume (see Figures 2a and 2b).

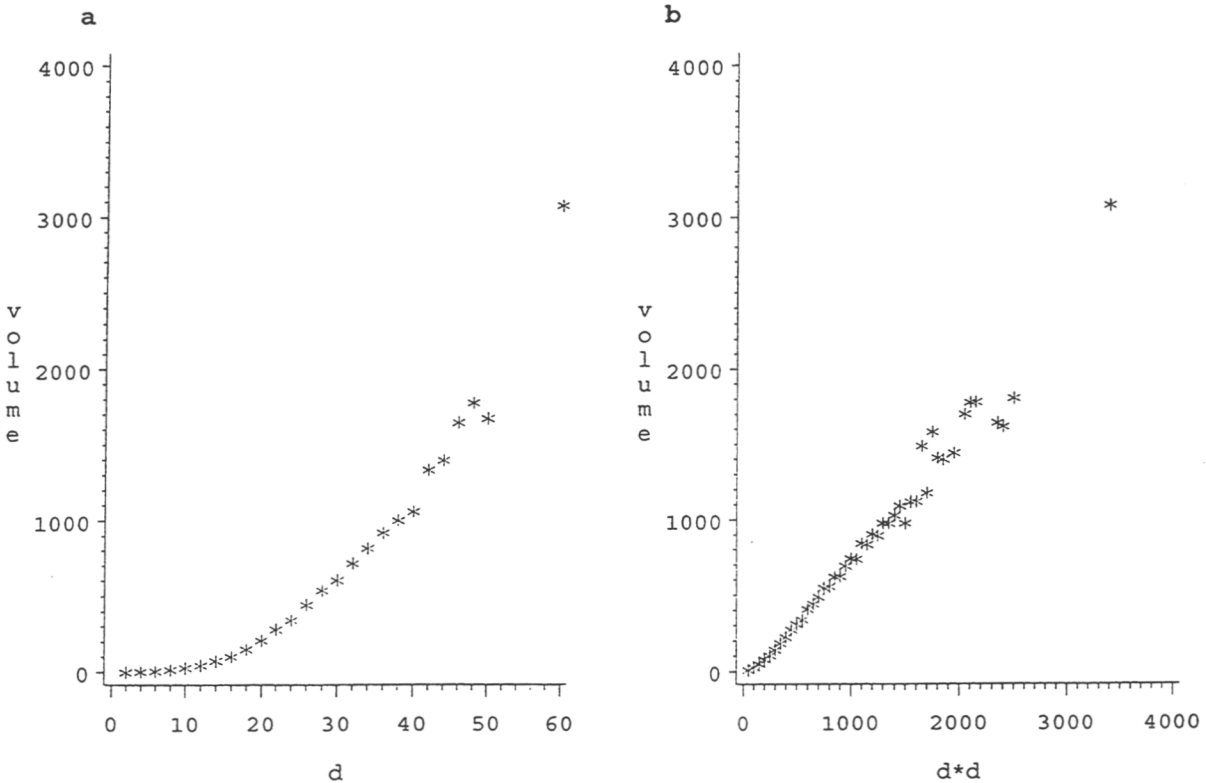


Figure 2. Volume vs. diameter (a) and volume vs squared diameter (b).

Numerous different combinations of the parameters of the distance function (Function 2) were tested. Different values of the k parameter (number of neighbors) were also tested. Good fit (i.e. low mean of residuals and low mean of squared residuals) of the model was obtained with the parameter combination presented in Table 1 (when $k=10$). Table 1 presents also the standard deviation of each independent variable. Each parameter value has been multiplied by the standard deviation of the independent variable to help in the interpretation of the results.

Table 1. Example of the values of the distance function parameters (a_1) when $k=10$. (std = standard deviation of the independent variable.)

1 regressor	a_1	std	$a_1 \cdot \text{std}$
1. d^2 , cm^2	0.567	441.19	250
2. x-coordinate, km	0.149	67.19	10
3. y-coordinate, km	0.053	94.22	5
4. age, a	1.812	33.11	60
5. basal area, m^2/ha	2.020	7.92	16
6. mean diameter, cm	1.937	8.26	16

The mean and standard deviation of the residuals of volume estimates were 0.11 dm^3 and 67.92 dm^6 , respectively, when the parameter values in Table 1 were applied.

Table 2 presents how changing the value of a_1 parameter for d^2 affects the results. The parameter values of the other independent variables are the same as in Table 1 and $k=10$.

Table 2. Reliability of volume estimates (mean and standard deviation of residuals) and between plot (std(plot)) and within plot (std(tree)) standard deviations of height estimates at different a_1 parameter values for d^2 .

a_1	mean of residuals	std. dev. of residuals	std(plot)	std(tree)
0.249	1.26	67.74	23.65	4.31
0.340	0.72	67.29	23.37	4.58
0.567	0.11	67.92	23.09	5.07
0.068	-0.01	68.76	22.99	5.24

Standard deviations of measured heights			26.82	4.96
---	--	--	-------	------

a

b

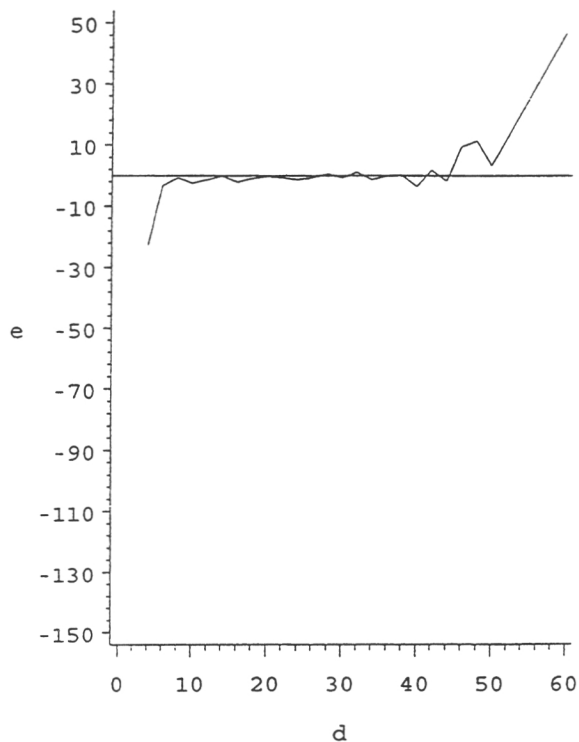
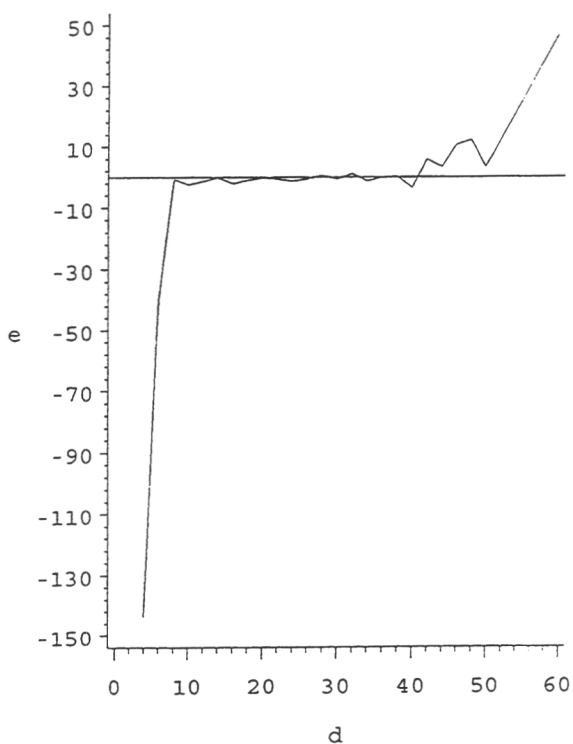


Figure 3. Relative bias (100 times mean of residuals divided by 'true' volume) in 2 cm diameter classes. **a)** $a_1 = 0.567$ for all trees and **b)** $a_1 = 2.267$ for small and large trees.

The plot of the residuals vs diameter showed that the residuals are biased. The mean of the residuals by 2 cm diameter classes (when a_1 for $d^2 = 0.567$ and rest of the parameters are as in Table 1) is presented in Figure 3a. The volumes were over-estimated for small trees and under-estimated for large trees. The results were markedly improved by increasing the a_1 parameter values of d^2 for the smallest and largest trees. The bias was decreased by setting, for example:

$a_1 = 2.267$, if $d < 6$ cm or if $d > 40$ cm,

$a_1 = 0.567$, otherwise

(see Figure 3b). With these values for a_1 and the values from Table 1 for the rest of the parameters, the mean of the residuals was -0.43 and the standard deviation was 67.91 dm³.

Table 3 presents how changing the k parameter (number of nearest neighbors) affects the results. The a_1 parameter value for d^2 is 0.567 or 2.267 (the latter for small and large trees) and the rest of the parameter values are the same as in Table 1.

Table 3. Reliability of volume estimates (mean and standard deviation of residuals) and between plot (std(plot)) and within plot (std(tree)) standard deviations of height estimates at different k parameter values.

k	mean of residuals	std. dev of residuals	std(plot)	std(tree)
1	0.01	76.98	26.42	7.39
5	0.16	67.79	23.71	5.30
10	-0.43	67.91	23.36	5.32
15	0.01	69.11	23.07	5.31
*Standard deviations of measured heights			26.82	4.96

The nearest neighbor estimator was tested further using the sample trees of the NFI7 data as the data points and the sample trees of the NFI8 as the test material. In these tests, both the data for estimating the model and the test data are samples of Finnish forests. However, the target population is not exactly the same

because of the different years of measurement.

Table 4 shows how changing the a_1 parameter values for d^2 affects the results in these tests. The number of nearest neighbors included (parameter k) was five. The other parameter values are the same as in Table 1.

Table 4. The reliability of the volume estimates (mean and standard deviation of residuals) and between-plot ($\text{var}(\text{plot})$) and within-plot ($\text{var}(\text{tree})$) standard deviations* of the height estimates at different a_1 parameter values for d^2 . The sample trees from NFI7 data provided the data for estimating the model.

a_1	mean of residuals	std. dev. of residuals	$\text{var}(\text{plot})$	$\text{var}(\text{tree})$
0.340	-8.20	70.11	24.10	5.30
0.227	-7.01	73.92	23.98	4.88
0.113	-3.48	77.15	24.62	3.58
0.068	-0.19	88.44	25.07	2.87
*Standard deviations of measured heights			26.82	4.96

Again, the volume estimates were found to be clearly biased in the different diameter classes. Using higher values of a_1 parameter of d^2 for small and large trees improved the results.

Table 5 shows the standard deviation and correlation of height and age estimates with the values described above for a_1 parameters and values 1, 5 and 10 for the k parameter. The sample trees from the NFI7 were used as data for "estimating the model".

Table 5. Correlation of the height and age estimates* for different values of the k parameter.

k	1	5	10
corr	0.78	0.84	0.86

*correlation of measured heights and ages was 0.73

5. CONCLUSIONS

The results in Table 2 show that changing the value of a_1 parameter of d^2 (within the range presented in Table 2) have only little effect on the standard deviation of the residuals of the volume estimates. Figures 3a and 3b show, however, that the value of the a_1 parameter of d^2 strongly affects the estimates for small and large trees. These trees are quite few in number in the data - this explains why the improved estimates for small and large trees have only negligible effect on the standard deviation of residuals calculated over the whole data. The scarcity of large and small trees (e.g. there is only one observation for diameter class 60 cm) explains also why a relatively high value of a_1 parameter of d^2 is necessary for large and small trees. The "distance" (in respect to the independent variables other than diameter) between the small (or large) trees is long. Therefore, in most cases the nearest neighbors of a small-sized tree are the middle-sized trees that are more frequent in the data. This causes over-estimation of the volumes for small trees (see Figure 3a). Increasing the weight of the diameter improves the results for small and large trees.

Another factor causing biased results for small and large trees is that the relation of d^2 and volume is linear only for middle-sized trees (Figure 2b). In Kernel estimation, Vieu (1991) recommends using varying window widths if the shape of the function describing the relation between the dependent and independent variables varies. To eliminate bias, narrow windows are used at locations where the regression curve is steep. To minimize variance, large windows are used at locations where the regression curve is less steep. This procedure was found to be effective in this study for nearest-neighbor estimation also.

The within plot variation of height estimates increases as the weight given for d^2 increases (Table 2). With a high weight given for diameter, even decreasing height curve estimates were observed when the plotwise height curve estimates were plotted. A relatively low value of a_1 parameter gives both reliable and realistic estimates of volume and height (Table 2).

The between-plot and within-plot standard deviations of height estimates were realistic in most of the studied cases (Table 2). However, when $k=1$, the within-plot standard deviation of height estimates is markedly higher than the standard deviation of measured heights (Table 3). The number of neighbors have only a negligible effect on the variance components when k is 5 or more.

When the NFI7 data were used for "estimating the model", the volume estimates were markedly biased if the weight of d^2 was high (Table 4). The bias could be decreased by giving less weight to d^2 (or more weight to the other independent variables). This result indicates that the stem form of pines has changed during the time between the two inventories. Changes in stem form have been reported in earlier studies also (see e.g. Korhonen 1993). Increasing the weight of standwise independent variables improves the results.

The correlation of height and age estimates is higher than the correlation of measured heights and ages (Table 5). Decreasing the k parameter value decreases the correlation.

The most serious drawback of the applied method is that unbiasedness of estimates is not automatically guaranteed. With parametric regression models, unbiasedness of estimates is in most (or at least in some) cases easy to obtain. When the nearest-neighbor regression is applied it is necessary to carefully study the estimates and to compare different parameters. The algorithms are simple but 'expensive' when the time consumption of the computations is considered.

The results of this paper indicate that nearest-neighbor regression can be recommended for generalisation of sample tree data for both the estimation of basic statistics and creation of realistic images of forests for simulation purposes.

Acknowledgements. The authors wish to thank prof. Juha Alho, Dr. Juha Lappi and Dr. Jussi Saramäki for critical comments and suggestions to improve the manuscript.

LITERATURE

- Altman, N.S. 1992. An Introduction to Kernel and Nearest-neighbor Nonparametric Regression. *The American Statistician* 46:175-185.
- Härdle, W. 1989. *Applied nonparametric regression*. Cambridge University Press. 323 pp.
- Korhonen, K.T. 1993. Mixed estimation in calibration of volume functions of Scots pine. *Silva Fennica* 27(4):269-276.
- Korhonen, K.T. 1994. Calculation system for large-scale forest inventory. Finnish Forest Research Institute Research Papers No. 505. Joensuu Research Station. 36 pp.
- Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. *Communications Instituti Forestalis Fenniae* 108. 74 pp.
- Ranneby, B. & Svensson, S.A. 1991. From sample tree data to images of tree populations. In: *Forest inventories in Europe with special reference to statistical methods*. Proceedings of the International IUFRO S. 4.02 and S. 604 Symposium, May 14-16, 1990. Swiss Federal Institute for Forest, Snow and Landscape Research. Birmensdorf, Switzerland.
- Siitonen, M. 1983. A long term forestry planning system based on data from the Finnish National Forest Inventory. In: *Forest Inventory for Improved Management*. Proceedings of the IUFRO Subject Group 4.02 Meeting in Finland, September 5-9, 1983. University of Helsinki, Department of forest mensuration and management. Research Notes No. 17:195-207.
- Tokola, T. 1990. *Satelliittikuvan ja VMI-koealatiedon käyttö metsätalouselueen puuston inventoinnissa*. [Use of satellite images and NFI data in timber inventory. In Finnish]. Licentiate of Science thesis. University of Joensuu, Finland. Faculty of forestry. 56 pp.
- Tomppo, E. 1992. Multi-source National Forest Inventory of Finland. In: *Proceedings of Ilvessalo Symposium on National Forest Inventories*. Finland 17-21 August 1992. Nyysönen, A., Poso, S. & Rautala, J. (ed.). The Finnish Forest Research Institute. Research Papers 444:52-59.
- Vieu, P. 1991. Nonparametric regression: optimal local bandwidth choice. *Journal of Royal Statistical Society. Series B* 53(2):453-464.

METSÄNTUTKIMUSLAITOS
Joensuun tutkimusasema

Käyntiosoite: Yliopistokatu 7
Postiosoite: PL 68, 80101 Joensuu
Puhelin: 973-151 4000
Telefax: 973-151 4567

ISSN 0358-4283
ISBN 951-40-1469-3