



Arto Haara

Arto Haara

Kasvuennusteiden luotettavuuden selvittäminen knn-menetelmällä ja monitavoiteoptimoinnilla

Haara, A. 2002. Kasvuennusteiden luotettavuuden selvittäminen knn-menetelmällä ja monitavoiteoptimoinnilla. *Metsätieteen aikakauskirja* 3/2002: 391–406.

Tässä tutkimuksessa testattiin k-lähimmän naapurin (knn) menetelmän ja monitavoiteoptimoinnin soveltuvuutta metsikön kasvuennusteiden luotettavuuden arviointiin. Tutkimuksen aineistona käytettiin Metsäntutkimuslaitoksen kangasmaiden kasvukoealoja. Puulajeista tarkasteltiin mäntyä, kuusta ja koivua. Koealoilta muodostettiin kuvioittaisen arvioinnin SOLMU- ja TASO-muotoiset tiedostot, joihin simuloitiin lisäksi kuvioittaisen arvioinnin mittausvirhettä. Luotettavuusarviot estimoitiin pohjapinta-alamediaanipuun läpimitalle ja pituudelle, pohjapinta-alalle ja tilavuudelle. Luotettavuusarviot estimoitiin puulajeittain ja metsikölle. Kuvion puustotunnusten luotettavuusarviot saatiin k-lähimmän naapurin vastaavien tunnusten kehitysennusteiden virheiden hajonnasta, joille oli ennustettu samantasoinen kasvu samanlaisilla puustotunnuksilla. Virhe saatiin kasvukoealalta mitatun ja MELAlla simuloitun kehityksen erona. Lähimpien naapureiden haussa käytetyt metsikkö- ja puustotunnukset ja tunnusten painot etsittiin monitavoiteoptimoinnilla. Menetelmää testattiin estimoimalla kunkin kuvion puustotunnuksien kehitysennusteiden luotettavuusarviot etsien lähimmät naapurit tutkimusaineistosta, josta poistettiin kulloinkin käsiteltävä kuvio. Menetelmän hyvyttä tarkasteltiin vertaamalla puustotunnusten kasvuennusteiden toteutuneita keskivirheitä ja harhaa knn-menetelmän tuottamiin keskivirheiden ja harhan ennusteisiin. Lisäksi hyvyttä arvioitiin tarkastelemalla keskivirheitä ja harhaa metsätyypeittäin ja pohjapinta-ala- ja ikäluokittain. Menetelmä toimi hyvin tutkimusaineistossa. Kuvioaineiston tietosisällöllä (SOLMU/TASO) ei ollut merkittävää vaikutusta tuloksiin. Menetelmää voidaan käyttää metsänomistajan päätöksenteon tukena ennusteiden luotettavuuden arvioinnissa. Luotettavuusarviot poistavat liian optimistiset kuvitelmat suunnittelulaskelmien paikkansa pitävyydestä. Menetelmää voidaan käyttää myös laskennallisessa ajantasaistuksessa päätettäessä, käytetäänkö kuviotiedoston ajantasaistusta kuvioittaisen inventoinnin tukena tai korvaamassa osittain uutta inventointitietoa.

Asiasanat: ennustaminen, luotettavuus, k-lähimmän naapurin menetelmä, kuvioittainen arviointi, monitavoiteoptimointi

Yhteystiedot: Metla, Joensuun tutkimuskeskus, PL 68, 80101 Joensuu

Sähköposti arto.haara@metla.fi

Hyväksytty 17.9.2002

I Johdanto

Monitavoitteisen metsäsuunnittelun lähtökohdaksi tarvitaan monipuolista ja paikkaan sidottua inventointitietoa. Kuvioittaista arviointia käytetään Suomessa yleisesti metsävaratietojen hankintamenetelmänä. Esimerkiksi alueelliset metsäkeskukset keräävät yksityismetsien suunnittelua varten metsävaratiedot kuvioittaisella arvioinnilla.

Kuvioittaisessa arvioinnissa suunnittelija kerää kultakin metsikkökuviolta kasvupaikka- ja puustotiedot. Puustotietojen keruussa tapahtui merkittävä muutos vuonna 1996 käyttöön otetun SOLMU-metsäsuunnittelujärjestelmän myötä. Tällöin kuvion puustotiedot siirryttiin keräämään puusto-ositteittain (SOLMU. Maastotyöopas... 1996). Kaikki alueellisissa ja tilakohtaisissa metsäsuunnittelussa kerättävät metsävaratiedot tallennetaan nykyisin metsäkeskusten Luotsi-luonnonvaratietokantaan (Oksanen-Peltola 1999). Aiemmassa TASO-järjestelmässä (TASO. Maastotyöopas... 1993) arvioitiin kullekin kuviolle yksi pohjapinta-alamediaanipuun, jonka tunnukset yleistettiin edelleen kullekin kuviolla olleelle puulajille. Pohjapinta-alan tai tilavuuden mukaiset puulajiosuudet merkittiin ylös, mikäli puulajin osuus kuviolla oli vähintään 10 prosenttia. Käyttöön otetussa SOLMU-järjestelmässä kullekin puusto-ositteelle arvioidaan oma pohjapinta-alamediaanipuunsa tunnuksineen. Koska SOLMU-tietosisältöisessä aineistossa puuston kuvaus on monipuolisempaa kuin TASO-tietosisältöisessä aineistossa, on TASO-tietosisältöinen aineisto muunnettava SOLMU-tietosisältöiseksi ennen sen siirtämistä Luotsi-tietokantaan. Muunnoksessa TASO-järjestelmän mukaiset metsikön keskipuun tunnukset yleistetään suoraan kullekin puusto-ositteelle keskipuuksi. Kokonaispohjapinta-ala ja runkoluku jaetaan kullekin puusto-ositteelle suoraan ositteen suhteellisen osuuden mukaan. Maastossa tehdyt puulaji-ositteiden jakovirheet ja metsikön keskipuun tunnusten yleistäminen puusto-ositteiden keskitunnuksiksi tuovat lisävirhettä sekametsien aineistoihin.

Kuvioittainen arviointi on subjektiivinen inventointimenetelmä. Kuvion puusto-ositteesta arvioidaan keskitunnuksia, ja arvioija käyttää useimmiten omaa harkintaansa koealojen sijoituksessa ja pohjapinta-alamediaanipuun valinnassa. Kuvioittaisen

arvioinnin tarkkuus vaihteleekin paljon arvioijan mukaan. Esimerkiksi Poso (1983) sai tärkeimpien kuviotunnusten keskivirheeksi kokeneilla inventoijilla 20 prosenttia. Kokemattomilla arvioijilla se saattoi nousta 30 prosenttiin.

Kuvioittaisen metsävaratiedon käytössä ollaan siirtymässä myös tietojen ajantasalla pitoon laskennallisella ajantasaistuksella. Ajantasaistus tuo uuden virhekomponentin jo ennestäänkin useita erillisiä virhelähteitä sisältävään metsävaratietojen laskentaan ja kasvu- ja tuotosennusteiden laadintaan. Laasasenahon ja Päivisen (1986) tutkimuksessa arvioidun ja tarkistetun puuston tilavuuksien erojen hajonta tilavuuksien keskiarvosta oli uudessa kuvioittaisessa arvioinnissa 15 prosenttia ja käsittelemättömillä, 4–7 vuotta ajantasaistetuilla kuvioilla 24 prosenttia keskiarvosta. Suutarla (1985) päivitti laskennallisesti 6–9 vuotta vanhat inventointitiedot ja sai runkotilavuusarvioiden erojen keskimääräiseksi hajonnaksi 17,3 prosenttia verrattuna tarkistusinventointiin, kun taas uudella inventoinnilla samalla alueella päästiin 14,4 prosentin hajontaan. Samassa tutkimuksessa ennustetut hakkuumäärät jäivät vanhan tiedon perusteella uutta inventointia pienemmiksi. Pilhjeran (1987) tutkimuksessa puustotunnusten arvioimisvirheet vaikuttivat kuvioittaisiin toimenpidevalintoihin, mutta metsänomistajan kokonaisyhteyteen näillä virheillä oli kuitenkin melko pieni vaikutus.

Metsien käytön ja hoidon suunnitteluvälineenä käytetään monimutkaisia, useista malleista koostuvia simulointisysteemejä (esim. Siitonen 1993, Siitonen ym. 1996, Pukkala 1998). Nämä ohjelmistot käyttävät yleensä lähtötietoina kuvioittaisessa arvioinnissa arvioituja metsikkö- ja puustotunnuksia, kuten keskiläpimittaa ja -pituutta, metsätyyppiä ja pohjapinta-alaa. Keskiläpimitan avulla ennustetaan runkolukusarja, jonka jälkeen voidaan hyödyntää puun läpimitan ja tilavuuden välistä relaatiota (Kilki 1984). Suunnittelussa käytettyjen ohjelmistojen ja tuotosmallien lähtötiedot tuotetaan useimmiten erilaisilla läpimittajakaumamalleilla (esim. Kilki ja Siitonen 1975, Pukkala 1994).

Simuloimalla kasvatettujen tunnusten luotettavuudelle on vaikea antaa ennustetta, koska monimutkaisten laskentajärjestelmien kautta tuleva virhe sisältää useita virhelähteitä. Ilmeisimpänä virhelähteenä on käytettyjen mallien aiheuttama

epävarmuus. Malleilla tehtyihin ennusteisiin liittyy epävarmuutta, joka syntyy neljästä päälähteestä (Kangas ja Kangas 1997): 1) mallin jäännösvirheestä, 2) mallin kertoimien epävarmuudesta, 3) mallin selittävissä muuttujissa olevista virheistä ja 4) mallin spesifioinnissa johtuvista virheistä. Myös metsäsovellusten ennustevirheiden spatiaalinen ja temporaalinen korreloituneisuus voi aiheuttaa ongelmia luotettavuuden arvioinnissa. Lähtötietojen virheet muodostuvat useasta eri virhelähteestä: 1) mittausvirheistä, 2) otantavirheistä, 3) luokitusvirheistä ja 4) ennustamisvirheistä. Lisäksi laskentajärjestelmissä hyödynnetään usein asiantuntijankemeyksiä, jotka aiheuttavat ennusteisiin edelleen epävarmuutta (Alho 1990).

Ennusteiden epävarmuutta voidaan selvittää empiirisesti, mikäli käytettävissä on soveltuva tarkistusaineisto. Tähän liittyy kuitenkin ongelmia. Mikäli ylipäänsä onnistutaan löytämään soveltuva tarkistusaineisto, on sen käyttö rajoitettua. Epävarmuuden arviot joudutaan laatimaan tietyllä ajanjaksolle, aineistolle ja alueelle. Tällöin ennusteiden soveltuvuus tulevalle kasvulle on epävarmaa. Yleensä joudutaankin käyttämään mallipohjaisia lähestymistapoja empiiristen menetelmien sijasta.

Ennusteiden luotettavuuden selvittämisessä on käytetty yleisesti kahta menetelmää: Monte Carlo -menetelmää (esim. Mäkelä 1988, McRoberts ym. 1994, Kangas 1997, Kangas 1999) ja varianssipropagointimenetelmää (esim. Gertner 1987, Kangas 1996). Molemmista menetelmistä monista eri virhelähteistä olevat virheet yhdistetään kokonaisvirheeksi. Monte Carlo -menetelmässä luotettavuusarviot tuotetaan generoimalla kullekin ennusteelle virhetermi kyseisen mallin virheen jakaumasta. Tämän ”lottoarvonnan” tuottamia ennustetuloksia tuotetaan useita satoja, joista muodostuu ennusteiden virhearvio. Varianssipropagointimenetelmässä sovelletaan yleensä ensimmäisen asteen Taylorin sarjakehitelmää. Molempien menetelmien ongelmana on huono yleistettävyyttä muuttuneissa olosuhteissa. Esimerkiksi jonkin mallin muuttuminen laskentajärjestelmässä tai uuden, tarkasteltavan aineiston erilaisuus voi edellyttää uusien laskelmien tekemistä. Lisäksi varianssipropagoinnin käyttö voi olla hankalaa monimutkaisissa laskentajärjestelmissä. Monte Carlo -menetelmä puolestaan edellyttää nykyäänkin paljon laskenta-aikaa ja tehoa. Lisäksi eri virhelähteit-

ten huomiointi voi tuottaa ongelmia esimerkiksi eri virheiden korreloituneisuuden takia.

Tässä tutkimuksessa käsitellään k-lähimmän naapurin menetelmän (k nearest neighbour classification) soveltamista luotettavuuden arviointiin. k-lähimmän naapurin menetelmässä (Hall 1985) kohdeyksikölle haetaan k kappaletta mahdollisimman lähellä tätä yksikköä olevia naapuristoyksiköitä isommasta joukosta yksiköitä (ns. referenssiaineistosta). Kunkin referenssiaineiston yksikön etäisyys kohdeyksiköstä lasketaan etäisyysfunktiolla, jossa on yhden tai useamman tunnuksen etäisyydet käsiteltävästä yksiköstä. Kohdeyksikön kiinnostaville tunnuksille voidaan antaa naapuruston tunnusten keskiarvot tai lähimpien naapureiden tunnusten arvoa voidaan painottaa esimerkiksi etäisyydellä. k-lähimmän naapurin menetelmällä (knn-menetelmä) voidaan estimoida valitulle kuviolle samanaikaisesti kaikki kiinnostavat tunnuksat referenssiaineistosta etsittyjen lähimpien naapureiden vastaavista tunnuksista (Härdle 1989, Altman 1992).

Tutkimuksen tavoitteena on tuottaa luotettavuusarviot kuvion puustotunnusten kehitysennusteille hyödyntämällä samankaltaisten kuvioiden tiedossa olevia kehitysennusteiden virheitä. Virheet saadaan kasvukoealalta mitatun ja MELAlla simuloitun kehityksen erona. Kuvion simuloitujen puustotunnusten arvoille tuotetaan luotettavuusarviot hakemalla knn-menetelmällä mahdollisimman samankaltaisia kuvioita, joiden virheiden hajonnasta estimoidaan luotettavuusarviot. knn-menetelmän etäisyysfunktion muuttujat ja muuttujien painot haetaan monitavoiteoptimoinnilla.

2 Aineisto ja menetelmät

2.1 Aineisto

Tutkimusaineistona käytettiin Metsäntutkimuslaitoksen kangasmaiden koemetsikköjen pysyviä kasvukoealoja ns. INKA-aineistoa (Gustavsen ym. 1988). Kasvukoealat sijaitsevat kattavasti Suomen alueella. Koemetsiköistä on mitattu kolmesta ympyräkoealasta koostuva ryvä, jonka yhteenlaskettujen puiden lukumäärä on vähintään 100 (Pohjois-Suomessa) tai 120 (Etelä-Suomessa). Koealojen kaikilta

Taulukko I. Tutkimusaineiston puusto-ositteiden ja kasvukoealoista muodostettujen metsiköiden keskitunnukset. Käytetyt muuttujat: n = lukumäärä, PPA = pohjapinta-ala, RL = runkoluku, V = kokonaistilavuus, V_{tukki} = tukkitilavuus, d_{gM} = pohjapinta-alamediaanipuun läpimitta, h_{gM} = pohjapinta-alamediaanipuun pituus.

		Mänty	Kuusi	Lehtipuut	Metsikkö
n	Lukumäärä	625	361	207	754
PPA	Minimi	1,0	1,0	1,0	1,1
	Maksimi	32,7	34,1	24,8	37,4
	Keskiarvo	10,7	12,3	4,3	16,1
	Keskihajonta	6,8	8,9	4,2	8,2
RL	Minimi	22	35	27	200
	Maksimi	4619	3186	7219	8212
	Keskiarvo	1046	884	485	1473
	Keskihajonta	847	632	820	850
V	Minimi	3,4	2,9	2,7	3,4
	Maksimi	310,2	378,7	249,2	378,7
	Keskiarvo	71,7	93,3	29,6	113,0
	Keskihajonta	60,1	82,4	36,8	82,5
V_{tukki}	Minimi	0,0	0,0	0,0	0,0
	Maksimi	203,6	317,5	133,6	317,5
	Keskiarvo	25,4	40,3	6,2	42,1
	Keskihajonta	36,0	55,1	15,6	54,9
d_{gM}	Minimi	3,7	3,0	2,4	
	Maksimi	40,8	36,2	40,3	
	Keskiarvo	17,0	17,3	16,3	
	Keskihajonta	7,3	6,0	6,9	
h_{gM}	Minimi	2,7	3,3	3,8	
	Maksimi	27,3	31,2	26,5	
	Keskiarvo	12,6	14,0	14,0	
	Keskihajonta	5,7	5,1	5,2	

puilta on merkitty ylös puulaji ja läpimitta, lisäksi osalta koealan puista on mitattu pituus. Koealat on mitattu kolme kertaa viiden vuoden välein. Koealoille tehdyt toimenpiteet on myös merkitty ylös. Kunkin metsikön koealat yhdistettiin kuvaamaan tätä metsikköä. Aineistosta rajattiin pois kaikki metsiköt, joiden kaikista koealoista ei löytynyt kolmea mittauskertaa. Lisäksi rajattiin pois metsiköt, joissa oli suoritettu avohakkuu. Käytettyjen metsiköiden puustotiedot on koottu taulukkoon 1.

2.2 Aineiston käsittely

Puuston keskitunnukset ja tilavuudet laskettiin ensimmäiseltä ja kolmannelta mittauskerralta INKA-aineiston koemetsiköistä. Empiirinen runkolukusarja muodostettiin kustakin metsiköstä yhdistämällä

metsikön koealojen puut runkolukusarjaksi. Kunkin puun edustama hehtaarikohtainen runkoluku saatiin jakamalla hehtaarin ala kolmen ympyräkoelan yhteenlasketulla pinta-alalla. Vain koepuilta oli mitattu pituudet. Lukupuiden puuttuvat pituudet ennustettiin kullekin metsikölle erikseen Näslundin pituusmallilla (Näslund 1936). Kullekin metsikölle estimoitiin metsikkökohtaiset pituusmallin parametrit pituuskoepuiden avulla. Kuhunkin pituusestimaattiin lisättiin vielä satunnaistekijä normaalijakaumasta, jonka hajonta saatiin Näslundin (1936) mallin keskivirheestä. Kullekin puulle laskettiin tilavuudet Laasasenahon (1982) tilavuus- ja runkokäyrämallilla. Kertomalla saadut tilavuudet puun edustamalla runkoluvulla ja summaamalla saadut tilavuudet yhteen saatiin metsiköiden kokonaistilavuudet. Metsiköiden nettokasvu saatiin eri mittauskertojen tilavuuksien erotuksesta. Empiirisen aineiston virheet

Taulukko 2. Puustotunnusten mittausvirheitä eri tutkimuksista (Purola 1983, Mähönen 1984, Laasasenaho ja Päivinen 1986, Pussinen 1992, Pigg 1994). Suhteelliset harhat ja keskihajonnat suluissa. Käytetyt muuttujat: d_{gM} = pohjapinta-alamediaanipuun läpimitta, PPA = pohjapinta-ala, h_{gM} = pohjapinta-alamediaanipuun pituus, RL = runkoluku.

		Tutkimus				
		Pigg	Purola	Pussinen	Mähönen	Laasasenaho ja Päivinen
d_{gM} (cm)	Harha	-0,8 (-4,8)			-1,0 (-4)	-1,0 (-4)
	Keskihajonta	2,5 (14,1)		3,4 (13,6)	2,4 (9)	2,2 (10)
PPA ($m^2 ha^{-1}$)	Harha	0,2 (1,2)	-2,9 (-11,2)		-2,88 (-12)	-0,1 (-1)
	Keskihajonta	2,8 (18,5)	2,3 (10,9)		4,68	3,3 (16)
h_{gM} (m)	Harha	-0,2 (-2)	0,29 (1,6)		-0,04 (0)	-0,6 (-3)
	Keskihajonta	1,7 (14,1)	1,5 (7,8)	18 (10,8)	1,6	2,1 (11)
RL ($r ha^{-1}$)	Harha		-220 (-12,8)			
	Keskihajonta	540	301 (20,1)			

muodostuivat koealamittausten mittausvirheistä ja yllä käytettyjen mallien virheistä.

Ensimmäisen mittauskerran empiirisestä aineistosta muodostettiin kuvioittaisen arvioinnin SOLMU- ja TASO-muotoiset kuviotiedostot, jotka koostuivat metsikön yleistiedot (mm. kasvupaikka, metsätyyppi, pinta-ala) sisältävästä metsikkötietueesta ja puuston tiedot sisältävästä puustotietueesta tai puusto-ositetietueista. Puustotietueessa koko metsikön puusto kuvattiin keskitunnuksilla ja eri puulajit tilavuus- tai pohjapinta-alaosuuden mukaan, kun taas puusto-ositetietueissa kuvion puusto kuvattiin puulajeittain ja -jaksoittain. Empiirisistä runkolukusarjoista laskettiin puusto-ositteittaiset pohjapinta-alamediaanipuun tunnuksat (läpimitta (d_{gM}), pituus (h_{gM}), ikä) ja puusto-ositteiden pohjapinta-alat kuvioaineiston muodostusta varten. Kuvioaineiston puusto-ositteista laskettujen keskitunnusten avulla muodostettiin SOLMU-tietosisältöiset puusto-ositetietueet. Koko kuviolle lasketun keskipuun ja puusto-ositteiden pohjapinta-alaosuuksien avulla muodostettiin kustakin kuvioista myös TASO-tietosisältöiset puustotiedot. Nämä TASO-tiedot muutettiin edelleen SOLMU-tietosisältöisiksi yleistämällä kuvion keskipuun tunnuksat kunkin puusto-ositteen keskipuun tunnuksiksi ja jakamalla kokonaispohjapinta-ala ja -runkoluku kunkin puusto-ositteen suhteellisella osuudella ositteiden pohjapinta-alan ja runkoluvun saamiseksi.

2.3 Virheiden simulointi kuvioaineistoon

SOLMU- ja TASO-muotoisiin kuvioaineistoihin simuloitiin myös kuvioittaisen arvioinnin virhettä. Kunkin puustotunnuksen virheiden suuruusluokat saatiin kirjallisuudesta (Purola 1983, Mähönen 1984, Laasasenaho ja Päivinen 1986, Pussinen 1992, Pigg 1994) (taulukko 2). Koska puusto-ositteen pohjapinta-alamediaanipuun tunnuksat (d_{gM} , h_{gM} , ikä) mitataan samasta puusta, tunnusten virheet korreloivat keskenään (esim. Ståhl 1992, Pigg 1994). Pohjapinta-alamediaanipuun tunnusten virheet simuloitiin multilog-normaalijakaumasta tämän riippuvuuden saamiseksi kuvioaineistoihin. Myös kuvion puusto-ositteiden pohjapinta-alojen virheiden katsottiin riippuvan toisistaan (esim. Pigg 1994). Niinpä pohjapinta-alankin virheet simuloitiin multilog-normaalijakaumasta. Sekä pohjapinta-alamediaanipuun että pohjapinta-alan tunnusten välisinä korrelaatioina käytettiin Haaran (2002) tutkimuksessa laskettuja korrelaatioita. Runkoluvun virhe simuloitiin log-normaalijakaumasta.

2.4 Kasvatussimuloinnit

Tutkimuksessa MELA-metsikkösimulaattoria (Siitonen ym. 2001) käytettiin kuvioiden kehityksen simulointiin. Muodostettuja kuvioaineistoja käytettiin

MELAn lähtöaineistoina. Kustakin puusto-ositteesta muodostettiin kuvio- ja puustotietojen avulla läpimittajakaumat Weibull-jakaumalla (Mykkänen 1986, Kilkki ym. 1989). Mikäli ositteen keskiläpimitta oli alle 5 cm, puusto-osite kuvattiin kuitenkin pituusjakaumalla. Pituusjakauman kuvauspuille esitettiin edelleen läpimitat. Teoreettisten jakaumien läpimittaluokkien kuvauspuita kasvatettiin yksittäisten puiden kasvumalleilla (Hynynen ym. 2002). Kuvioiden kehitystä simuloitiin kymmenen vuotta. Kuvauspuiden tilavuustunnukset laskettiin kaikissa tapauksissa Laasasenahon (1982) tilavuus- ja runkokäyrämalleilla. Metsikön tilavuuskasvu saatiin yhteenlaskemalla kuvauspuiden kasvut. Metsikön puiden kuoleminen kasvatusaikana mallitettiin kuolemistodennäköisyyksimalleilla, joita korjattiin tarvittaessa itseharvenemismalleilla (Hynynen ym. 2002). Kuvioille kasvatusaikana tehdyt toimenpiteet (esim. harvennukset) simuloitiin kuvion käsittelyksi.

2.5 knn-menetelmän ja monitavoiteoptimoinnin käyttö kasvatuksen luotettavuuden arvioinnissa

knn-menetelmällä kuviolle etsittiin kasvupaikka- ja puustotunnuksiltaan 10 lähintä naapuria, joiden tunnusten ajantasaisuuden virheistä johdettiin luotettavuusarviot kasvatetuille kuvion puustotunnuksille. Etäisyysfunktioilla laskettiin kohdekuvion i tunnusten itseisarvoetäisyys kunkin naapurikuvion j tunnuksista:

Etäisyys $_{ij}$ =

$$a_1|x_{1i} - x_{1j}| + a_2|x_{2i} - x_{2j}| + \dots + a_n|x_{ni} - x_{nj}| \quad (1)$$

a_1, a_2, \dots, a_n = etäisyysfunktion muuttujien painot

$x_{1i}, x_{2i}, \dots, x_{ni}$ = muuttujien standardoidut arvot kohdekuviolla i

$x_{1j}, x_{2j}, \dots, x_{nj}$ = muuttujien standardoidut arvot naapurikuviolla j

n = muuttujien lukumäärä

Etäisyysfunktiossa kokeillut muuttujat standardoitiin vähentämällä kustakin muuttujan arvosta muuttujan hajonta ja jakamalla erotus muuttujan keskiarvolla, jotta eri muuttujien arvot saatiin yhteismi-

tallisiksi. Lähimpien naapureiden etsintä toteutettiin erikseen koko kuviolle ja puulajeittain puusto-ositteille. Etäisyysfunktion muuttujina kokeiltiin kuviolle kasvupaikkatyyppiä, pohjapinta-alaa, kuvion pohjapinta-alamediaanipuun keskiläpimittaa ja -pituutta, ikää, puulajiositteiden lukumäärää ja maksimipuulajiositteen suuruutta. Puulajiositteiden tapauksessa etäisyysfunktion muuttujina kokeiltiin puulajin pohjapinta-alaosuutta, pohjapinta-alamediaanipuun läpimittaa, pituutta ja ikää, puulajin pohjapinta-alaa ja kasvupaikkatyyppiä.

Monitavoiteoptimoinnilla etsittiin knn-menetelmän etäisyysfunktion muuttujat ja niiden painot. Tähän käytettiin epälineaarista optimointialgoritmia (Hooke ja Jeeves 1961). Optimoinnin päätösmuuttujina olivat etäisyysfunktiossa kokeillut muuttujat. Optimoinnin tavoitemuuttujana minimoitiin kohdemetsikön ja naapureiden MELA-simulaattorilla ennustettujen kasvujen poikkeamaa. Tällöin kohdemetsikön virheet tulivat niistä metsiköistä, joille oli ennustettu samantasoinen kasvu samantaisilla puustotunnuksilla. Optimointiongelman päätösmuuttujien alkuarvoiksi valittiin neljästä sadasta satunnaisratkaisusta paras, ja optimointi aloitettiin siitä. Koska tavoitemuuttujan summa on epäkonvekksi tavoitemuuttuja, ei ollut varmaa että löydetty paikallinen optimi oli myös yleinen. Tämän vuoksi jokainen optimointiongelma ratkaistiin 10 kertaa käyttäen kullakin kerralla parhaan satunnaisratkaisun päätösmuuttujien arvoja. Paras optimiratkaisu kymmenestä valittiin globaaliksi optimiksi. Päätösmuuttujien arvojen etsintä lopetettiin, kun askel päätösmuuttujien painon muuttamiseen oli vähemmän kuin 0,1. Päätösmuuttujien painot rajoitettiin välille 0–10.

Monitavoiteoptimoinnissa käytettiin cross-validation -tekniikkaa, jossa kullakin optimointikerralla jokaiselle metsikölle etsittiin vuorollaan puusto- ja metsikkötunnuksiltaan k lähintä naapuria muista metsiköistä optimointikerran päätösmuuttujien arvoilla. Optimoinnit tehtiin erikseen SOLMU- ja TASO-muotoisille aineistoille ilman simuloituja virheitä ja virheiden kanssa. Lisäksi aineistot jaoteltiin optimointia varten myös sen mukaan, oliko niissä tehty ajantasaisuuskaudella toimenpiteitä vai ei. Optimoinnit tehtiin kymmenen vuoden kasvatuskaudelle.

2.6 Tulosten laskenta

Luotettavuusennusteiden hyvyttä tarkasteltiin estimoimalla kunkin kuvion tunnusten luotettavuusennuste optimiratkaisun antamalla etäisyysfunktioilla käyttäen kuvioaineiston muita kuvioita naapurikuvioina. Empiiristen puustotunnusten keskivirheitä (RMSE) ja harhaa verrattiin knn-menetelmällä saattuihin keskivirheiden ennusteisiin ja harhaan. Puustotunnusten empiirinen harha ja keskivirhe (RMSE) laskettiin seuraavasti:

$$\text{Harha} = \sum_{i=1}^N \frac{(k_i - \hat{k}_i)}{N} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (k_i - \hat{k}_i)^2}{N-1}} \quad (3)$$

k_i = tunnuksen k todellinen arvo puusto-ositteella/metsikössä i

\hat{k}_i = tunnuksen k ennustettu arvo puusto-ositteella/metsikössä i

N = havaintojen lukumäärä

Eri tunnusten suhteellinen harha ja RMSE laskettiin jakamalla absoluuttinen harha ja RMSE tunnusten ennusteiden keskiarvolla. Luotettavuusarvioiden paikkansa pitävyyttä tutkittiin myös luokittelemalla aineisto kasvupaikka-, pohjapinta-ala- ja ikäluokittain ja tarkastelemalla luokkien empiirisiä ja ennustettuja keskivirheitä sekä harhoja.

Luotettavuusennusteiden suhteellinen keskivirhe ja harha isommalle alueelle saatiin summaamalla kuvioille ennustetut suhteelliset keskivirheet ja harhat ja jakamalla saadut summat kuvioiden määrällä. Puustotunnusten luotettavuusennusteiden paikkansa pitävyyttä yksittäisillä kuvioilla tutkittiin generoimalla kullekin kuviolle puustotunnuksen estimoidusta keskivirheestä 95 prosentin luottamusväli ja laskemalla koko alueella kuinka suuri osuus kuvioiden puustotunnusten kehitysennusteiden virheistä sattui tälle luottamusvälille. Toteutunut virhe sattui luottamusvälille, mikäli sen itseisarvo oli pienempi kuin $1,96 \cdot$ keskivirheen ennuste. Puustotunnusten luotettavuusennusteiden paikkansa pitävyyttä yksittäisillä kuvioilla tutkittiin myös graafisesti.

3 Tulokset

Tutkimuksessa tarkasteltiin knn-menetelmällä saattujen kuviotunnusten ennusteiden luotettavuusarvioiden hyvyttä, kun etäisyysfunktion muuttujat ja muuttujien painot oli määritetty monitavoiteoptimoinnilla. Simuloimalla kasvatettujen puusto-ositteiden tunnusten luotettavuuden arvioinnissa etäisyysfunktioon valituiksi muuttujiksi tulivat kasvupaikka, pohjapinta-ala, keskiläpimitta, ikä ja puulajin pohjapinta-alaosuus koko metsikön pohjapinta-alasta. Simuloimalla kasvatettujen kuviotunnusten luotettavuuden arvioinnissa etäisyysfunktion muuttujiksi tulivat maksimipuulajiosuuden suuruus, kasvupaikka, pohjapinta-ala, ikä ja keskiläpimitta. Tutkimuksessa laskettiin empiiriset ja knn-menetelmällä ennustetut puustotunnusten ennusteiden keskivirheet ja harhat koko aineistolle SOLMU- ja TASO-muotoisissa aineistoissa. Tilavuuden ja pohjapinta-alan keskivirhe-ennusteet olivat melko harhattomia (taulukko 3). knn-menetelmä tuotti jonkin verran liian pieniä keskivirhe-estimaatteja kuvioilla, joissa oli tehty toimenpiteitä. Niissä tilavuuden keskivirhe oli keskimäärin 4,3 prosenttia liian pieni, ja suurimmillaan keskivirhe jäi 6,3 prosenttia liian pieneksi. Ennustetut pohjapinta-alan ja tilavuuden kehitysennusteiden harhat jäivät jonkin verran pienemmiksi kuin toteutuneet harhat. Kun kuvioaineistoihin simuloitiin virhettä keskivirheiden ja harhojen osalta ei muodostunut kovinkaan suurta eroa verrattuna virheettömiin aineistoihin. Myöskin pohjapinta-alamediaanipuun läpimitan ja pituden kehitysennusteiden keskivirheet ja empiiriset keskivirheet olivat lähellä toisiaan. Ennustetut keskivirheet olivat jonkin verran pienempiä kuin empiiriset keskivirheet (SOLMU: 0,11–1,73 %, TASO: 1,48–7,54 %) poikkeaman ollessa suurimmillaan TASO-aineistossa, johon oli simuloitu virhettä ja jossa oli tehty toimenpiteitä (7,54 %). Ennusteet aineistoissa, joihin oli simuloitu virhettä, olivat jonkin verran huonompia kuin virheettömissä aineistoissa. Myöskin ennustetut harhat seurasivat melko hyvin empiirisiä harhoja.

Aineistot (SOLMU ja TASO) luokiteltiin metsätyypeittäin kolmeen luokkaan (MT: tuoreet kankaat ja rehevämmät, VT: kuivahkot kankaat, CT: kuivat kankaat ja karummat). Näille luokille laskettiin kuvion pohjapinta-alan ja tilavuuden keskivirheet

Taulukko 3. Metsikön pohjapinta-alan ja tilavuuden empiiriset ja knn-menetelmällä ennustetut suhteelliset keskivirheet (RMSE) sekä harhat INKA-aineistossa.

Aineisto		RMSE, PPA (%)		RMSE, tilavuus (%)	
		Empiirinen	Ennustettu	Empiirinen	Ennustettu
SOLMU	– Ei simuloitu virheitä, ei toimenpiteitä	11,63	11,20	14,03	15,31
	– Ei simuloitu virheitä, tehty toimenpiteitä	26,23	24,69	29,88	26,95
	– Simuloitu virheitä, ei toimenpiteitä	19,03	19,28	22,72	23,24
	– Simuloitu virheitä, tehty toimenpiteitä	27,16	24,05	33,98	29,52
TASO	– Ei simuloitu virheitä, ei toimenpiteitä	11,06	11,10	15,93	16,14
	– Ei simuloitu virheitä, tehty toimenpiteitä	24,86	22,93	28,72	25,22
	– Simuloitu virheitä, ei toimenpiteitä	19,12	18,78	24,05	22,03
	– Simuloitu virheitä, tehty toimenpiteitä	29,89	23,21	30,94	24,63
		Harha, PPA (%)		Harha, tilavuus (%)	
		Empiirinen	Ennustettu	Empiirinen	Ennustettu
SOLMU	– Ei simuloitu virheitä, ei toimenpiteitä	2,80	1,95	2,91	2,86
	– Ei simuloitu virheitä, tehty toimenpiteitä	6,40	4,12	8,90	5,75
	– Simuloitu virheitä, ei toimenpiteitä	1,74	0,75	2,27	2,42
	– Simuloitu virheitä, tehty toimenpiteitä	8,45	5,74	12,64	8,92
TASO	– Ei simuloitu virheitä, ei toimenpiteitä	0,01	–0,41	–0,43	0,63
	– Ei simuloitu virheitä, tehty toimenpiteitä	2,99	–0,31	4,70	0,54
	– Simuloitu virheitä, ei toimenpiteitä	–1,63	–2,46	–2,43	–1,56
	– Simuloitu virheitä, tehty toimenpiteitä	–17,79	–9,08	–17,31	–8,20

Taulukko 4. Metsikön pohjapinta-alan empiiriset ja knn-menetelmällä ennustetut suhteelliset keskivirheet (RMSE) sekä harhat INKA-aineistossa metsätyypeittäin (MT: tuoreet kankaat ja rehevämät, VT: kuivahkot kankaat, CT: kuivat kankaat ja karummat).

Aineisto		RMSE, PPA(%)					
		MT		VT		CT	
		Empiirinen	Ennustettu	Empiirinen	Ennustettu	Empiirinen	Ennustettu
SOLMU	– Ei simuloitu virheitä, ei toimenpiteitä	9,44	9,11	12,89	12,01	20,31	15,38
	– Ei simuloitu virheitä, tehty toimenpiteitä	24,13	24,13	27,58	23,76	34,16	32,83
	– Simuloitu virheitä, ei toimenpiteitä	16,00	15,60	20,14	20,50	33,64	27,58
	– Simuloitu virheitä, tehty toimenpiteitä	24,94	24,51	27,28	22,59	44,57	29,62
TASO	– Ei simuloitu virheitä, ei toimenpiteitä	9,35	9,86	11,52	10,96	20,71	16,33
	– Ei simuloitu virheitä, tehty toimenpiteitä	21,92	20,75	27,05	23,10	34,13	33,66
	– Simuloitu virheitä, ei toimenpiteitä	18,22	17,54	18,17	18,99	28,35	22,48
	– Simuloitu virheitä, tehty toimenpiteitä	27,31	19,65	32,17	24,62	33,08	34,90
		Harha, PPA (%)					
		MT		VT		CT	
		Empiirinen	Ennustettu	Empiirinen	Ennustettu	Empiirinen	Ennustettu
SOLMU	– Ei simuloitu virheitä, ei toimenpiteitä	2,17	1,02	4,46	3,14	–2,99	0,22
	– Ei simuloitu virheitä, tehty toimenpiteitä	9,00	8,33	6,82	3,96	–18,79	–17,56
	– Simuloitu virheitä, ei toimenpiteitä	1,80	1,54	3,36	1,32	–8,54	–4,65
	– Simuloitu virheitä, tehty toimenpiteitä	11,73	9,45	8,05	4,46	–17,21	–7,13
TASO	– Ei simuloitu virheitä, ei toimenpiteitä	–1,00	–1,46	1,74	0,60	–3,36	–0,92
	– Ei simuloitu virheitä, tehty toimenpiteitä	5,19	2,61	3,52	0,40	–19,87	–19,95
	– Simuloitu virheitä, ei toimenpiteitä	–2,13	–3,65	–0,45	–1,50	–5,46	–2,21
	– Simuloitu virheitä, tehty toimenpiteitä	–17,00	–8,25	–18,91	–10,38	–16,21	–6,02

empiiriselle aineistolle ja keskivirhe-ennusteille 10 vuoden kasvatuksen jälkeen. Pohjapinta-alan keskivirhe-ennusteiden ja empiiristen keskivirheiden eroista huomattiin, että erot luokkien välillä eivät olleet ennusteissa yhtä isot kuin empiirisissä, min-kä vuoksi luokkien keskivirhe-ennusteet lähestyivät toisiaan (taulukko 4). Tämä johtui osaltaan siitä, että lähimpien naapureiden joukkoon tuli myös muilta metsätyypeiltä olevia kuvioita. Tämä korostui karuilla kasvupaikoilla, joista oli vähiten havaintoja, ja näin ollen myös vähiten naapuriehdokkaita. Pohjapinta-alan ennusteharhat seurasivat myös melko hyvin empiirisiä harhoja, tosin ne olivat jonkin verran pienempiä. Tilavuuden osalta ennustetut keskivirheet seurasivat myös hyvin empiirisiä keskivirheitä. Myös ennusteharhat olivat melko pitkälle samansuuntaisia kuin empiiriset harhat.

Luotettavuusennusteiden hyvyttä tarkasteltiin myös jakamalla eri kuvioaineistot ikäluokkiin ja laskemalla luokittaiset tilavuuden ja pohjapinta-alan keskivirheet ja harhat. Tilavuuden empiiriset keskivirheet ja harhat vaihtelivat huomattavasti ikäluokittain varsinkin virheettömillä aineistoilla, joissa ei oltu tehty toimenpiteitä (kuva 1). Virheel-lisillä aineistoilla vaihtelu tasoittui jonkin verran. knn-menetelmällä ennustetut keskivirheet poikke-sivat eniten empiirisistä niissä ikäluokissa, joissa keskivirheet olivat selvästi poikkeavimpia. Harhan ennusteet seurasivat melko hyvin toteutuneita harhoja. Myös pohjapinta-alan keskivirheet vaihtelivat ikäluokittain 10 vuoden kasvatuksen jälkeen suurimpien virheiden ollessa nuorimmissa ikäluokissa. Ikäluokittainen vaihtelu vähentyi aineistoissa, joihin oli simuloitu virhettä. Pohjapinta-alan keskivirheen ennusteet seurasivat melko hyvin empiirisiä keskivirheitä. Keskivirheiden ja myös harhojenkin osalta ennusteet pyrkivät tasoittamaan luokkien välisiä eroja sekä tilavuudella että pohjapinta-alalla.

INKA-aineistosta generoidut kuvioittaisen arvi-oinnin aineistot jaoteltiin myös pohjapinta-alaluokittain ja empiiriset ja knn-menetelmällä ennustetut tilavuuden keskivirheet ja harhat laskettiin näissä luokissa. Tilavuuden empiiriset keskivirheet olivat 10 vuoden kasvatuksen jälkeen suurimmillaan kuvioilla, joissa pohjapinta-ala oli pieni (kuva 2). Kes-kivirheen ja harhan ennusteet seurasivat luokittain melko hyvin vastaavia empiirisiä. Myös pohjapinta-alaositteilla, joilla olivat suurimmat empiiriset kes-

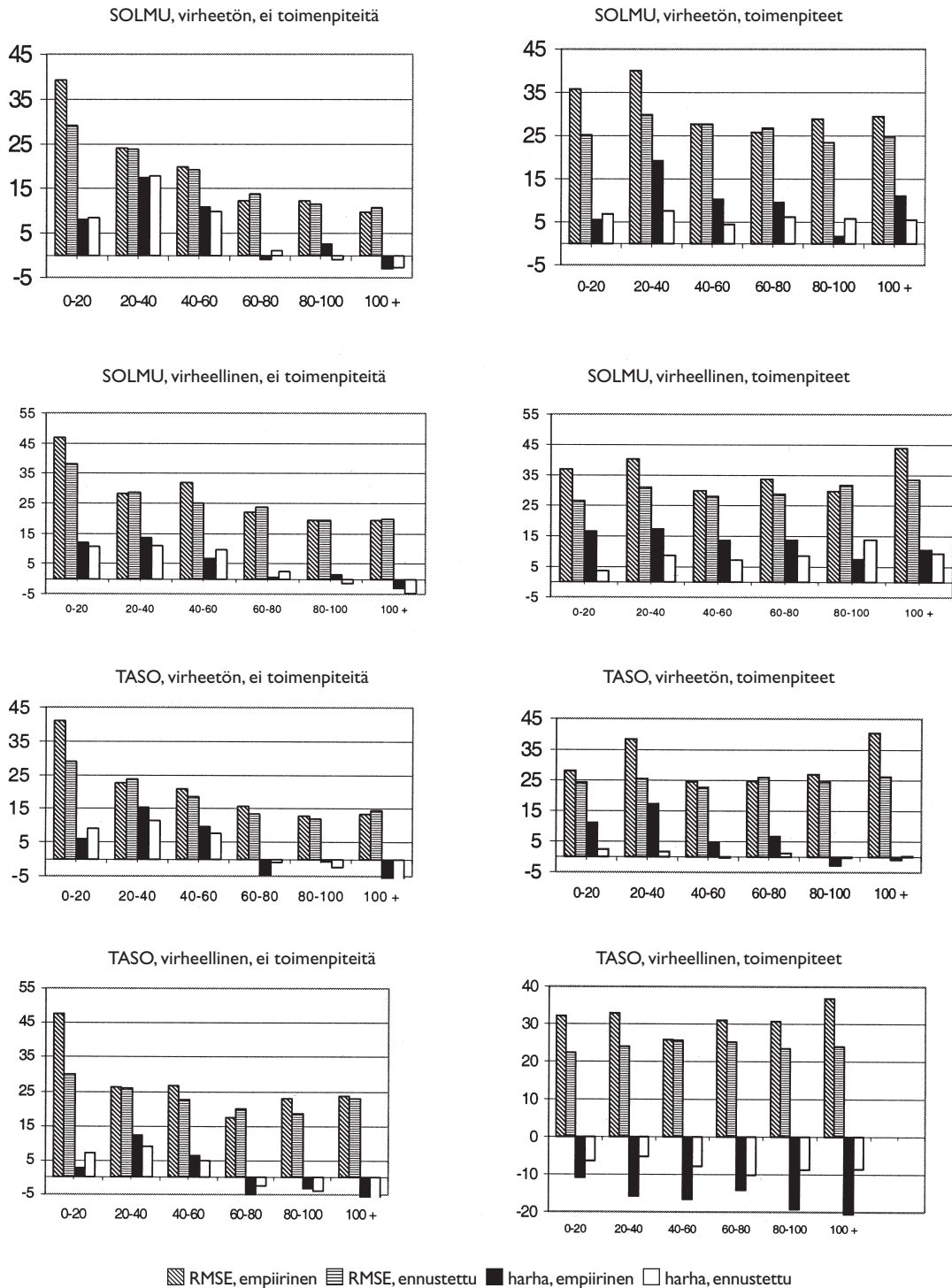
kivirheet ja harhat, ennusteet pienensivät luokkien välistä todellista vaihtelua.

Puustotunnusten empiirisiä ja ennustettuja keskivirheitä ja harhoja 10 vuoden kasvatuksen jälkeen tarkasteltiin myös puulajiositteittain. Männyllä pohjapintapinta-alan ja tilavuuden empiiriset ja knn-menetelmällä estimoidut keskivirheet ja harhat seurasivat melko hyvin toisiaan. Sen sijaan pohjapinta-alamediaanipuun läpimitan keskivirheen ennusteet jäivät jonkin verran pienemmiksi kuin empiiriset (taulukko 5). Pituuden ennustetut keskivirheet ja harhat seurasivat melko hyvin vastaavia empiirisiä.

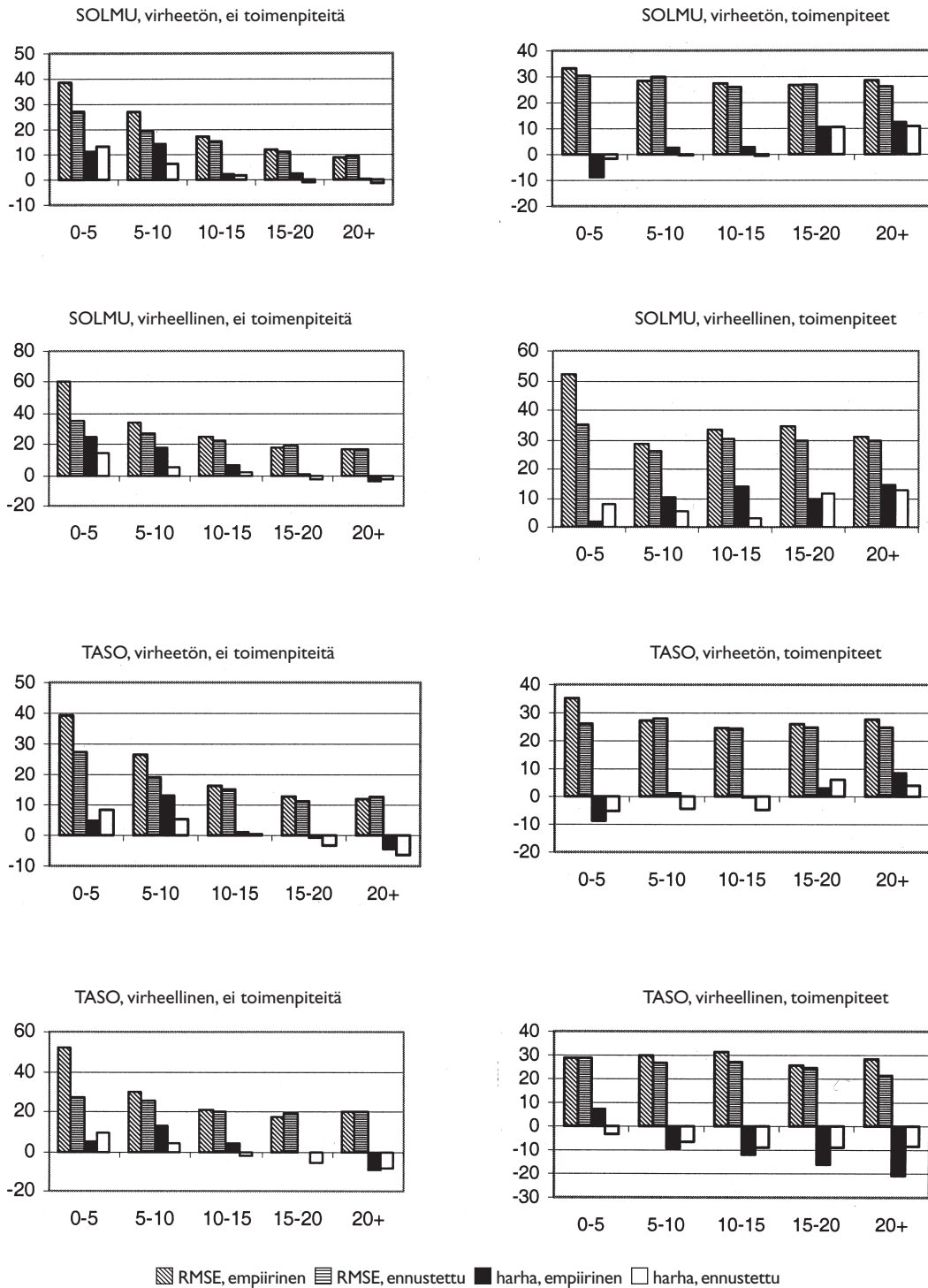
Kuusiositteessa pohjapinta-alan ja tilavuuden keskivirheet ennustettiin kymmenen vuoden kasvatuk-sen jälkeen jonkin verran yliarvioksi (taulukko 6). Keskivirhe- ja harhaennusteet olivat kuitenkin melko hyvin kohdallaan ainakin kuvioilla, joilla ei ollut tehty toimenpiteitä. Pohjapinta-alamediaanipuun läpimitan ja pituuden keskivirheen ja harhan ennusteet osuivat hyvin kohdalleen. Esimerkiksi läpimitan en-nustetun keskivirheen poikkeamat empiirisestä kes-kivirheestä vaihtelivat välillä $-2,98 \dots 1,62 \%$ (em-piirinen keskivirhe vaihteli välillä $9,0-26,78 \%$).

Koivulla ei tuloksia laskettu käsitellyille kuvioille aineiston pienuuden takia. Koivuositteen tilavuuden keskivirheen ennusteet jäivät jonkin verran empiiri-siä pienemmiksi, kun taas pohjapinta-alalle saatiin liian suuret keskivirheet virheel-lisillä aineistoilla. Pohjapinta-alan ja tilavuuden harhan ennusteet osui-vat hyvin kohdalleen. Suurimmat poikkeamat tuli-vat pohjapinta-alan ennusteille SOLMU-aineistossa, johon oli simuloitu virhettä (keskivirhe: empiirinen $31,73 \%$ ja ennustettu $37,45 \%$, harha: empiirinen $7,61$ ja ennustettu $16,99$). Koivun pohjapinta-ala-mediaanipuun läpimitan ja pituuden keskivirheen ja harhan ennusteet seurasivat melko hyvin vastaa-via empiirisiä.

Kussakin INKA-aineistosta tuotetussa kuvioai-neistossa tarkasteltiin myös yksittäisten kuvioiden kasvatettujen puustotunnusten empiiristen virheiden sattumista luottamusvälille, joka tuotettiin kuvioille estimoidusta keskivirheestä. Tilavuuden ja pohja-pinta-alan osalta kaikissa aineistoissa empiirisistä virheistä noin 90 prosenttia sisältyi virheen ennu-steesta generoidulle 95% luottamusvälille (tau-lukko 7). Em. asiaa tarkasteltiin myös graafisesti (kuvat 3–4). Kuvissa yhtenäinen viiva (+) kuvaa



Kuva 1. Kuvion tilavuuden empiiriset ja knn-menetelmällä ennustetut tilavuuden suhteelliset keskirvirheet (RMSE) ja harhat (y-akseli) ikäluokittain (x-akseli) INKA-aineistosta muodostetuissa kuvioaineistoissa.



Kuva 2. Kuvion tilavuuden empiiriset ja knn-menetelmällä ennustetut tilavuuden suhteelliset keskivirheet (RMSE) ja harhat (y-akseli) pohjapinta-alaluokittain (x-akseli) INKA-aineistosta muodostetuissa kuvioaineistoissa.

Taulukko 5. Männyn pohjapinta-alamediaanipuun läpimitan (d_{gM}) empiiriset (RMSE) ja knn-menetelmällä ennustetut suhteelliset keskivirheet (RMSE) ja harhat INKA-aineistossa.

Aineisto		RMSE, d_{gM} (%)		Harha, d_{gM} (%)	
		Empiirinen	Ennustettu	Empiirinen	Ennustettu
SOLMU	– Ei simuloitu virheitä, ei toimenpiteitä	11,09	6,22	–0,87	–1,37
	– Ei simuloitu virheitä, tehty toimenpiteitä	12,94	9,17	–0,95	–1,31
	– Simuloitu virheitä, ei toimenpiteitä	18,81	15,43	–0,62	–0,74
	– Simuloitu virheitä, tehty toimenpiteitä	17,82	13,08	0,73	–0,49
TASO	– Ei simuloitu virheitä, ei toimenpiteitä	18,23	11,78	6,42	4,67
	– Ei simuloitu virheitä, tehty toimenpiteitä	11,88	9,40	2,34	0,51
	– Simuloitu virheitä, ei toimenpiteitä	21,52	16,85	5,35	3,93
	– Simuloitu virheitä, tehty toimenpiteitä	18,86	16,28	1,14	0,49

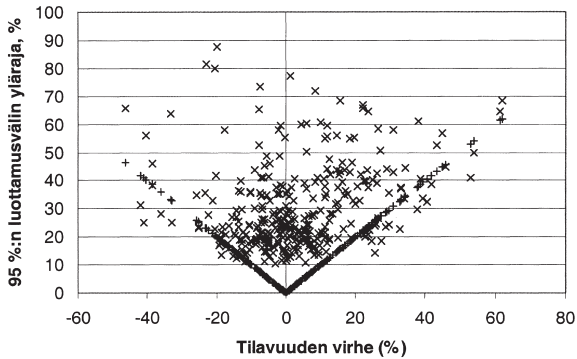
Taulukko 6. Kuusen pohjapinta-alan ja tilavuuden empiiriset ja knn-menetelmällä ennustetut suhteelliset keskivirheet (RMSE) ja harhat INKA-aineistossa.

Aineisto		RMSE, PPA (%)		RMSE, tilavuus (%)	
		Empiirinen	Ennustettu	Empiirinen	Ennustettu
SOLMU	– Ei simuloitu virheitä, ei toimenpiteitä	9,50	11,36	15,35	16,13
	– Ei simuloitu virheitä, tehty toimenpiteitä	19,17	20,81	22,84	25,12
	– Simuloitu virheitä, ei toimenpiteitä	22,11	21,33	28,02	28,29
	– Simuloitu virheitä, tehty toimenpiteitä	24,80	44,24	34,34	53,48
TASO	– Ei simuloitu virheitä, ei toimenpiteitä	12,24	17,80	18,96	24,25
	– Ei simuloitu virheitä, tehty toimenpiteitä	20,21	26,18	23,45	28,52
	– Simuloitu virheitä, ei toimenpiteitä	25,46	26,93	27,53	33,22
	– Simuloitu virheitä, tehty toimenpiteitä	28,91	28,76	28,09	29,40
		Harha, PPA (%)		Harha, tilavuus (%)	
		Empiirinen	Ennustettu	Empiirinen	Ennustettu
SOLMU	– Ei simuloitu virheitä, ei toimenpiteitä	0,68	1,47	1,50	1,01
	– Ei simuloitu virheitä, tehty toimenpiteitä	6,56	4,41	6,63	4,32
	– Simuloitu virheitä, ei toimenpiteitä	–2,91	–1,21	–2,97	–3,10
	– Simuloitu virheitä, tehty toimenpiteitä	11,82	19,52	12,03	19,86
TASO	– Ei simuloitu virheitä, ei toimenpiteitä	–0,73	–0,98	–3,63	–7,94
	– Ei simuloitu virheitä, tehty toimenpiteitä	6,13	5,69	4,89	2,77
	– Simuloitu virheitä, ei toimenpiteitä	–1,34	–5,58	–5,35	–12,26
	– Simuloitu virheitä, tehty toimenpiteitä	–15,05	–8,48	–14,11	–8,58

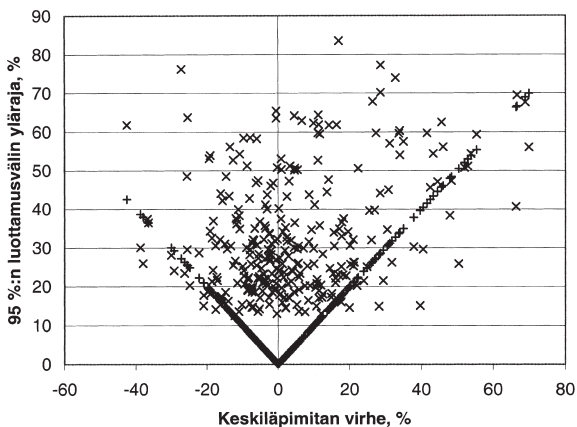
tilannetta, jossa havaittu virhe on 95 %:n luottamusvälin ylärajan suuruinen. Tällöin 95 prosenttia havainnoista pitäisi olla viivan yläpuolella ja 5 prosenttia sen alapuolella. Osalle suurista virheistä ennustettiin pieni keskivirhe-estimaatti. Nämä ovat ns. ongelmakuvioita. Toisaalta taas osalle kuvioista, joissa toteutuneet virheet olivat pieniä, keskivirhe-estimaatit olivat liian suuria.

4 Tulosten tarkastelu

Tässä tutkimuksessa testattiin knn-menetelmän ja monitavoiteoptimoinnin käyttöä kuvioittaisen arvioinnin puustotunnusten kehitysennusteiden luotettavuuden arvioinnissa. Luotettavuustunnusten hyvyttä tarkasteltiin vertaamalla toteutuneita kasvuennusteiden virheitä ja harhoja knn-mene-



Kuva 3. Virheettömän SOLMU-aineiston kuvioiden kokonaistilavuuden havaitut virheet 10 vuoden kasvatuksen jälkeen knn-menetelmällä kuviolle estimoidusta keskivirheestä johdetun 95 prosentin luottamusvälin ylärajan funktiona (x). Kuvaan on piirretty myös yhtenäinen viiva (+), jonka yläpuolella havaittu virhe on pienempi kuin luottamusvälin yläraja.



Kuva 4. Virheettömän TASO-kuvioaineiston mäntyositteen keskiläpimitan havaitut virheet 10 vuoden kasvatuksen jälkeen knn-menetelmällä kuviolle estimoidusta keskivirheestä johdetun 95 prosentin luottamusvälin ylärajan funktiona (x). Kuvaan on piirretty myös yhtenäinen viiva (+), jonka yläpuolella havaittu virhe on pienempi kuin luottamusvälin yläraja.

telmällä saatuihin virheisiin ja harhoihin. Lisäksi luotettavuustunnusten hyvyttä tarkasteltiin luokittelemalla metsiköt kasvupaikka-, pohjapinta-ala- ja ikäluokkiin ja laskemalla luokittaiset keskivirheet ja harhat.

Taulukko 7. Kuvion pohjapinta-alan ja kokonaistilavuuden kehitysnusteiden havaittujen virheiden osuus, jotka osuvat luotettavuusnusteesta generoidulle 95 prosentin luottamusvälille.

	PPA (%)	Tilavuus (%)
SOLMU, virheetön, ei toimenpiteitä	90,4	90,2
SOLMU, virheetön, toimenpiteet	89,3	90,8
SOLMU, virheellinen, ei toimenpiteitä	91,5	92,4
SOLMU, virheellinen, ei toimenpiteet	88,4	90,7
TASO, virheetön, ei toimenpiteitä	91,7	89,3
TASO, virheetön, toimenpiteet	90,1	91,6
TASO, virheellinen, ei toimenpiteitä	90,8	90,8
TASO, virheellinen, ei toimenpiteet	92,1	92,1

Käytetty menetelmä osoittautui lupaavaksi kuvioittaisen arvioinnin puustotunnusten kasvatuksen luotettavuuden arvioimisessa. Menetelmä antoi hyviä tuloksia alue- ja luokkatasolla sekä TASO- että SOLMU-tietosisältöisillä aineistoilla. Kun lähtöaineistoihin simuloitiin kuvioittaisen arvioinnin virhettä, tulokset eivät huonontuneet merkittävästi.

Vaikka TASO-muotoisia tietoja ei enää kerätä, on huomattava osa nykyisistä kuvioittaisen arvioinnin aineistoista vielä TASO-tiedoista muunnettuja. knn-menetelmä huomioi sekametsikköjen TASO-muunnoksesta aiheutuvan lisävirheen. TASO- ja SOLMU-muotoisten aineistojen tulosten pienet erot voivat selittyä osaksi sillä, että INKA-aineisto koostuu pääosin tasaikäisistä metsiköistä, joissa on yksi selvästi vallitseva puulaji. INKA-aineiston metsissä on kuitenkin usein yksi tai useampia vallittuja puusto-ositteita, joiden puustotiedot poikkeavat selvästi vallitsevasta puustosta. Männiköissä on yleensä alikasvoksena jonkin verran kuusia ja kuusikoissa puolestaan on kuusta huomattavasti kookkaampia mäntyjä ja koivuja. Empiiriset puulajeittaiset keskivirheet ovatkin selvästi suurempia TASO-aineistossa varsinkin pohjapinta-alamediaanipuun tunnusten osalta. Käytetty knn-menetelmä huomioi tämän hyvin. Saatuja tuloksia ei kuitenkaan voi suoraan yleistää aineistoihin, joissa on paljon sekametsiä ja eri-ikäisrakenteisia metsiä. Tämä pätee myös SOLMU-aineiston tarkastelun osalta.

Etäisyysfunktion muuttujien optimointi osoittautui tehokkaaksi tavaksi testata eri muuttujien tärkeyttä knn-menetelmän etäisyysfunktiossa. Etäisyysfunk-

tion valittavien muuttujien tulee olla yksiselitteisiä ja löytyä myös käytettävistä aineistoista, kuten tässä tutkimuksessa käytetyt tunnuksat löytyivät kuvioitaisen arvioinnin aineistosta.

Yksittäisten kuvioiden luotettavuusennusteiden tarkastelussa löytyi vähän kuvioita, joiden havaittu virhe ei mahtunut luotettavuusennusteesta tuotetulle 95 prosentin luottamusvälille. Tällaisten kuvioiden tarkempi kartoittaminen ja syiden etsiminen niiden huonohkoihin luotettavuusennustustesteisiin on jatkossa menetelmän kehityksen kannalta tärkeää. On kuitenkin mahdotonta antaa tarkkaa luotettavuusennustetta joka kuviolle.

Metsikön kehitysennusteiden luotettavuuden arvioinnissa on käytetty perinteisesti Monte Carlo-menetelmää ja varianssipropagointia (Kangas ja Kangas 1997). Vaikka malliperusteiset menetelmät antavatkin hyviä tuloksia luotettavuuden arvioinnissa (esim. Kangas 1999) on niiden käytäntöön soveltaminen jonkin verran työlästä. Ne on laskettu tietyssä tilanteessa (simulointijärjestelmän käytössä olevat mallit), tietylle alueelle ja tietylle ajanjaksolle. Niinpä olosuhteiden muuttuessa on tarve uusintalaskelmiin. knn-menetelmän etuna on sen hyvä muuntautumiskyky. Mikäli oletukset lähimpien naapureiden etsinnästä pitävät paikkaansa, ts. päätösmuuttujat ja niiden painotus on kohdallaan, on menetelmällä joustavaa laskea uudet luotettavuusarviot jonkin laskentajärjestelmässä tapahtuneen muutoksen, esim. kasvumallien uudistumisen myötä. Tämä edellyttää tosin, että referenssiaineistolla pystytään laskemaan uudet ajantasaistuksen virheet myös uudessa tilanteessa. Tällöin alkupe- räiset kuvioaineistot ja mahdolliset tarkistusinventointiaineistot on säilytettävä.

Käytettäessä knn-menetelmää kasvuennusteiden luotettavuuden arvioinnissa aineistoa ei tarvitse erikseen luokitella. Sen sijaan riippuvuus halutuista tunnuksista, kuten iästä, saadaan jatkuvaan muotoon. knn-menetelmä on joustavampi kuin parametrinen mallilähestymistapa, koska esimerkiksi hajonnan riippuvuutta tietyistä metsikkötunnuksista ei ole sidottu parametriseen malliin (esim. Kangas 1999). Toisaalta myös knn-menetelmässä pitäisi pystyä huomioimaan virheiden ajallinen ja spatiaalinen riippuvuus toisistaan esimerkiksi isojen alueiden ennusteita laskettaessa.

Tässä tutkimuksessa käytetty referenssiaineisto ei

sisältänyt aivan nuoria taimikoita, eri-ikäisrakenteisia eikä myöskään vanhoja metsiä. knn-menetelmässä pystytään tarvittaessa täydentämään referenssiaineistoa. Näin saadaan referenssiaineiston kattavuutta paremmaksi. Mikäli uusi referenssiaineisto ja testiaineisto poikkeavat rakenteeltaan huomattavasti aikaisemmasta voi kuitenkin olla paikallaan kokeilla parantaisiko uusi optimointi luotettavuuden arviointituloksia.

Myöskin referenssiaineiston keräämistä suppeammalta alueelta kuin tutkimusaineistossa, kuten esimerkiksi yhden metsäkeskuksen alueelta, voisi kokeilla parantamaan saatuja tuloksia. Tutkimuksessa ei kiinnitetty erityistä huomiota alueelliseen vaihteluun, vaikka tutkimusaineisto olikin hyvin laajalta alueelta. knn-menetelmä itsessään ottaa jonkin verran huomioon tätä vaihtelua, koska metsikkötunnuksat tietyllä alueella samankäisissä metsiköissä muistuttanevat toisiaan keskimääräistä enemmän kuin koko aineistossa samankaltaisen kasvun vuoksi. Tällöin samalla alueella kasvavat samantyyppiset metsiköt tulevat helposti valituiksi naapureiksi. Alueellisten erojen huomioimiseksi voi jatkossa kokeilla sijainnin sisällyttämistä (koordinaatit, etäisyys rannikosta, korkeus maanpinnasta jne.) päätösmuuttujaksi. Tällöin naapureita haettaisiin mahdollisimman pieneltä maantieteelliseltä alueelta.

Menetelmän jatkokehityksessä on keskityttävä yksittäisten kuvioiden luotettavuusennusteiden parantamiseen. Jatkossa voidaan esimerkiksi kokeilla knn-menetelmän naapureiden lukumäärän optimointia ja naapureiden painottamista etäisyyden mukaan. Tässä tutkimuksessa kokeiltiin kokonaistilavuuden ennusteen luotettavuusarvion parantamista virheetömässä TASO-aineistossa naapureiden lukumäärää vaihtamalla (5–15). Naapureiden lukumäärän muuttaminen ei kuitenkaan parantanut estimaattia verrattuna kymmeneen naapuriin. Jatkossa naapureiden lukumäärää voisi määräytyä referenssiaineiston ja kohdemetsikön ominaisuuksien mukaan. Kuvioitaisen arvioinnin tarkkuus vaihtelee suuresti mitaajan kokemuksen mukaan (Laasasenaho ja Päivinen 1986, Poso 1983). Jatkossa voitaisiinkin tutkia myös mitä vaikutusta on tilanteella, jossa yhdeksi päätösmuuttujaksi otettaisiin muuttuja, joka kuvaa mitaajan kokemusta.

Metsäsuunnittelun näkökulmasta ennusteiden epä-

varmuus on päätöstukeen liittyvä riskitekijä (Kangas 2001). Nykyisessä metsäsuunnitelmassa kertyvät ja odotettavissa olevat tulot ja menot esitetään tarkkoina lukuarvoina, jotka antavat liian optimistisen kuvan suunnitelman paikkansa pitävyydestä. Metsänomistajalle tulisi pystyä ilmoittamaan ennusteiden luotettavuus, jotta hän ei kuvittelisi tulosten olevan varmoja. Tutkimuksessa käytetty menetelmä on hyvinkin potentiaalinen päätöksentuen menetelmä. Menetelmällä saadaan kehitysennusteiden luotettavuusennuste halutuille tunnuksille, mikäli käytössä on kattava referenssiaineisto.

Tutkimuksessa käytetty menetelmä sopii myös hyvin apuvälineeksi kuvioittaisen arvioinnin metsävaratietojen hankinnassa yleistyvän laskennallisen ajantasaistuksen luotettavuuden selvittämisessä. Luotettavuustunnukset voivat antaa tukea suunnittelijalle päätöstilanteessa, jossa valitaan ajantasaistettavat kuviot ja kuviot joiden metsävaratiedot hankitaan uudella maastoinventoinnilla tai korvaavalla menetelmällä esim. jollakin kaukokartoitusmenetelmällä (esim. Anttila 2002). knn-menetelmän ideaalitalanteessa referenssiaineisto olisi ajantasaistettu samalta ajalta kuin kohdeaineisto. Tällöin kasvu olisi samalta ajanjaksolta ja inventointimenetelmä ja jopa mittajaikkin samat.

Kiitokset

Tutkimus on osa maa- ja metsätalousministeriön yhteistutkimusvaroin toteutettua Metsäsuunnittelun tietohuolto -tutkimushanketta. Käsikirjoitusta kommentoivat sen eri vaiheissa MMT Annika Kangas, MMT Kari, T. Korhonen, MMT Mikko Kurttila, MMT Matti Maltamo ja Ph.D. Tuula Nuutinen. MMT Jari Miina auttoi monitavoiteoptimoinnin integroinnissa osaksi knn-menetelmää. Lisäksi kaksi esitarkastajaa antoivat hyödyllisiä ja huolellisia käsikirjoitusta koskeneita kommentteja. Lämpimät kiitokseni kaikille yllämainituille.

Kirjallisuus

- Alho, J. 1990. Stochastic methods in population forecasting. *International Journal of Forecasting* 6: 521–530.
- Altman, N.S. 1992. The introduction to kernel and nearest neighbour nonparametric regression. *American Statistician* 46: 175–185.
- Anttila, P. 2002. Updating stand level inventory data applying growth models and visual interpretation of aerial photographs. *Silva Fennica* 36(2): 549–560.
- Gertner, G. 1987. Approximating precision in simulation projections: an efficient alternative to Monte Carlo methods. *Forest Science* 33: 239–244.
- Gustavsen, H.G., Roiko-Jokela, P. & Varmola, M. 1988. Kivennäismaiden talousmetsien pysyvät (INKA ja TINKA) kokeet. *Metsäntutkimuslaitoksen tiedonantoja* 292. 212 s.
- Haara, A. 2002. Comparing simulation methods for modelling the errors of the compartment inventory data. *Käsikirjoitus*.
- Hall, P. 1985. Near-neighbour estimation. *Julkaisussa: Kotz, S., Johnson, N.L. & Read, C.B. (toim.). Encyclopedia of statistical sciences* 6. Wiley, New York. s. 167–169.
- Hynynen, J. 1996. Puuston kehityksen ennustaminen MELA-järjestelmässä. *Julkaisussa: Hynynen, J. & Ojansuu, R. (toim.). Puuston kehityksen ennustaminen – MELA ja vaihtoehtoja. Metsäntutkimuslaitoksen tiedonantoja* 612. s. 21–37.
- , Ojansuu, R., Hökkä, H., Siipilehto, J., Salminen, H. & Haapala, P. 2002. Models for predicting stand development in MELA system. *Metsäntutkimuslaitoksen tiedonantoja* 835. 116 s.
- Hooke, R. & Jeeves, T.A. 1961. 'Direct search' solution of numerical and statistical problems. *Journal of the ACM* 8: 212–229.
- Härdle, W. 1989. *Applied nonparametric regression*. Cambridge University, Cambridge. 333 s.
- Kangas, A. 1996. On the bias and variance in tree volume predictions due to model and measurement errors. *Scandinavian Journal of Forest Research* 11: 281–290.
- 1997. On the prediction bias and variance in long-term growth projections. *Forest Ecology and Management* 96: 207–216.
- 1999. Methods for assessing uncertainty of growth and yield predictions. *Canadian Journal of Forest Research* 29: 1357–1364.
- & Kangas, J. 1997. Mallit, ennusteet ja simulointi metsätalouden laskentajärjestelmissä. *Metsätieteen aikakauskirja – Folia Forestalia* 3/1997: 389–404.
- Kangas, J. 2001. *Metsäsuunnittelun tutkimus- ja kehitys-*

- tämishaasteita. Julkaisussa: Kangas, J. & Kokko, A. (toim.). Metsän eri käyttömuotojen arvottaminen ja yhteensovittaminen. Metsän eri käyttömuotojen yhteensovittamisen tutkimusohjelman loppuraportti. Metsäntutkimuslaitoksen tiedonantoja 800. s. 309–314.
- Kilikki, P. 1984. Metsänmittausoppi. *Silva Carelica*. Joensuun yliopisto. 222 s.
- & Siitonen, M. 1975. Simulation of artificial stands and derivation of growing stock models from this material. *Acta Forestalia Fennica* 145. 33 s.
- , Maltamo, M., Mykkänen, R. & Päivinen, R. 1989. Use of the Weibull function in estimating the basal area diameter distribution. *Silva Fennica* 23: 311–318.
- Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. *Communications Instituti Forestalis Fenniae* 108. 74 s.
- & Päivinen, R. 1986. Kuvioittaisen arvioinnin tarkistamisesta. *Folia Forestalia* 664. 19 s.
- McRoberts, R.E., Hahn, J.T., Hefty, G.J. & Van Cleve, J.R. 1994. Variation in forest inventory field measurements. *Canadian Journal of Forest Research* 24: 1766–1770.
- Mykkänen, R. 1986. Weibull-funktion käyttö puuston läpimittajakauman estimoinnissa. Metsätalouden syventävien opintojen tutkielma. Joensuun yliopisto. 80 s.
- Mähönen, M. 1984. Kuvioittaisen arvioinnin luotettavuus. Metsänarvioimistieteen tutkielma maatalous- ja metsätieteiden kandidaatin tutkintoa varten. Helsingin yliopisto. 55 s.
- Mäkelä, A. 1988. Performance analysis of a process-based stand growth model using Monte Carlo techniques. *Scandinavian Journal of Forest Research* 3: 315–331.
- Näslund, M. 1936. The forest research institute's thinning experiment in pine forests. *Meddelanden från Statens Skogsförsöksanstalt* 29: 1–172.
- Ojansuu, R., Hynynen, J., Koivunen, J. & Luoma, P. 1991. Luonnonprosessit metsälaskelmassa (MELA) – Metsä 2000 -versio. Metsäntutkimuslaitoksen tiedonantoja 385. 59 s.
- Oksanen-Peltola, L. 1999. Metsäsuunnittelun tietotarpeet. Julkaisussa: Heikinheimo, M. (toim.). Metsäsuunnittelun tietohuolto. Metsäntutkimuslaitoksen julkaisuja 741. s. 3–15.
- Pigg, J. 1994. Keskiläpimitan ja puutavaralajijakauman sekä muiden puustotunnusten tarkkuus Metsähallituksen kuvioittaisessa arvioinnissa. Metsänarvioimistieteen pro gradu -työ. Helsingin yliopisto, metsävarojen käytön laitos. 92 s.
- Pilhjerta, K. 1987. Puustotunnusten arviointivirheiden vaikutus kuvioittaisiin toimenpidevalintoihin MELASSA. Pro gradu -työ. Joensuun yliopisto, metsätieteellinen tiedekunta. 68 s.
- Poso, S. 1983. Kuvioittaisen arvioimismenetelmän perusteita. *Silva Fennica* 17: 313–343.
- Pukkala, T. 1994. Metsäsuunnittelun perusteet. Gummerus kirjapaino Oy, Jyväskylä. 242 s.
- 1998. Monsu Metsäsuunnitteluohjelma. Ohjelmiston toiminta ja käyttö. Moniste. 68 s.
- Purola, M. 1983. Silmävaraisen arvioinnin tarkkuus metsätaitokilpailussa. Metsänarvioimistieteen pro gradu -työ. Helsingin yliopisto. 89 s.
- Pussinen, A. 1992. Ilmakuvat ja Landsat TM -satelliittikuva välialueiden kuvioittaisessa arvioinnissa. Pro gradu -työ. Joensuun yliopisto, metsätieteellinen tiedekunta. 48 s.
- Päivinen, R., Nousiainen, M. & Korhonen, K. 1992. Puustotunnusten mittaamisen luotettavuus. *Folia Forestalia* 787. 18 s.
- Siitonen, M. 1993. Experiences in the use of forest management planning models. Tiivistelmä: Kokemuksia mallien käytöstä metsäsuunnittelussa. *Silva Fennica* 27(2): 167–178.
- , Härkönen, K., Hirvelä, H., Jämsä, J., Kilpeläinen, H., Salminen, O. & Teuri, M. 1996. MELA handbook 1996 edition. Metsäntutkimuslaitoksen tiedonantoja 622. 452 s.
- , Anola-Pukkila, A., Haara, A., Härkönen, K., Redsvén, V., Salminen, O. & Suokas, A. (toim.). 2001. MELA handbook 2000 edition. The Finnish Forest Research Institute. 498 s.
- SOLMU. Maastotyöopas. 1996. Metsätalouden kehittämiskeskus Tapio, Helsinki. 1996. 80 s.
- Ståhl, G. 1992. En studie av kvalitet i skogliga avdelningsdata som insamlats med subjektiva inventeringsmetoder. Sveriges Landbruksuniversitet, Institutionen för biometri och skogsindelning, Rapport 24. 128 s.
- Suutarla, T. 1985. Kuvioittaisen inventoinnin päivitys ja sen luotettavuus. Metsänarvioimistieteen pro gradu -tutkielma. Helsingin yliopisto, metsävarojen käytön laitos. 47 s.
- TASO. Maastotyöopas. 1993. Metsäkeskus Tapio, Helsinki. 50 s.

43 viitettä