

Metsätehon raportti 182
14.1.2005

Rajoitettu jakelu

Koskitukki Oy
Metsäliitto Osuuskunta
Pölkky Oy
Stora Enso Oyj
UPM-Kymmene Oyj
Vapo Timber Oy

**Runkopankki ja k-MSN -menetelmä
puustotietojen ja pölkkyjakauman
ennustamisessa**

Tapio Räsänen
Ilkka Kuuramaa
Juha-Antti Sorsa

Runkopankki ja k-MSN -menetelmä puustotietojen ja pölkkyjakauman ennustamisessa

**Tapio Räsänen
Ilkka Kuuramaa
Juha-Antti Sorsa**

Metsätehon raportti 182
14.1.2005

Rajoitettu jakelu:

Koskitukki Oy, Metsäliitto Osuuskunta, Pölkky Oy, Stora Enso Oyj,
UPM-Kymmene Oyj ja Vapo Timber Oy

Asiasanat:

MSN, k-MSN, lähimmän naapurin menetelmä, puustotieto, runkopankki, leimikko, puunhankinta, puunhankintalogistiikka, hakkuukone, apteeraus, tietojärjestelmä

© Metsäteho Oy

Helsinki 2005

SISÄLLYS

TIIVISTELMÄ	4
1 TAUSTA JA TAVOITTEET	6
2 LASKENTAMENETELMÄN KUVAUS	7
2.1 Menetelmän yleiskuvaus.....	7
2.2 K lähimmän naapurin MSN-menetelmä.....	8
2.3 LAN MSN -menetelmä.....	9
3 HAKUMUUTTUIJEN VALINTAMENETELMÄT	10
3.1 Hakumuuttujien merkitys.....	10
3.2 Hakumuuttujien selvittäminen askeltavalla regressioanalyysillä	11
3.3 Hakumuuttujien soveltuvuuden arviointi.....	11
3.4 Hakumuuttujien subjektiivinen valinta.....	12
3.5 Hakumuuttujien valinnan automatisointi.....	12
4 AINEISTO	13
5 TESTAUKSET JA TULOKSET	15
5.1 Askeltavalla regressioanalyysillä poimitut selitettävät muuttujat....	15
5.2 Hakumuuttujat.....	16
5.2.1 Subjektiivinen valinta.....	16
5.2.2 Automaattinen valinta k-MSN ja LAN MSN -menetelmillä.....	17
5.3 Menetelmän testaukset.....	18
5.3.1 Testatut hakumuuttuja-, aineisto- ja menetelmäparametriyhdistelmät.....	18
5.3.2 Ennusteiden hyvyuden kriteerit.....	19
5.3.3 Hakumuuttujien virheen vaikutus.....	20
5.4 Testausten tulokset.....	20
5.4.1 Tilavuus- ja tukkiprosenttiennusteet.....	20
5.4.2 Pölkkyjakaumaennusteet	26
5.4.3 Kokonais- ja tukkitilavuuksien ennustevirheet sumatasolla	30
5.4.4 Hakumuuttujien virheen vaikutus.....	31
6 PÄÄTELMÄT	32
6.1 Tärkeimmät hakumuuttujat.....	32
6.2 Hakumuuttujien valintamenetelmät käytännön sovelluksissa	33
6.3 K-MSN –menetelmän ennustekyky tulosten perusteella.....	34
7 MENETELMÄN SOVELLUKSET	38
8 LOPUKSI	40
KIRJALLISUUS	41

LIITE

TIIVISTELMÄ

Puustotietojen kuvaukseen ja ennustamiseen sekä katkontavaihtoehtojen tarkasteluun on Metsätehossa yhteistyössä Joensuun yliopiston kanssa kehitetty hakkuukoneiden runkokohtaiseen mitattuun tietoon perustuvaa tietovarastoa (runkopankkia) sekä sen käyttöä varten laskentamenetelmiä. Puustotunnusten ja puutavaralajeittaisten pölkkyjakaumien ennustamiseen on kehitetty ns. k lähimmän naapurin MSN-menetelmää, jolle on olemassa paljon sovelluksia mm. metsätieteissä. K-MSN –menetelmän käyttö perustuu ennustettavaa kohdetta – metsikköä tai leimikkoa – mahdollisimman paljon puuston määrän ja rakenteen perusteella vastaavien kohteiden valintaan ja niiden perusteella muodostettuihin puustotunnus- ja pölkkyjakaumaestimaatteihin. Ennustettavasta kohteesta ja sen puustosta mitattuja tai arvioituja tunnuksia käytetään hakumuuttujatietoina vastinkohteiden valinnassa.

Projektin tavoitteena oli kehittää ja testata vaihtoehtoisia MSN-laskentamenetelmiä sekä käyttää testauksessa Metsätehon runkopankkiaineistoa. Tutkimusaineisto käsitti n. 400 hakkuukohdetta. Joensuun yliopiston osuuden tavoitteena oli selvittää erityisesti sopivimpien hakumuuttujien valintamenetelmiä sekä erilaisen hakumuuttujatiedon merkitystä ennusteiden tarkkuuteen. MSN-sovelluksen toimintojen, tiedostorajapintojen ja käyttöliittymän kehittäminen oli tavoitteena niin, että sovellus olisi käyttökelpoinen sekä Metsätehon tutkimustarpeisiin että mahdollisimman joustavasti myös yritysten tietojärjestelmiin liitettäväksi.

Tutkimuksessa testattiin k-MSN –menetelmää käyttäen perushakumuuttujiksi valittuja muuttujia niitä vaihdellen ja kokeillen erilaisia laskentaparametri- ja aineistorajoitusvaihtoehtoja. Menetelmän tarkkuutta, luotettavuutta ja käytettävyyttä leimikon puustotunnusten ja pölkkyjakauman ennustamisessa tarkasteltiin eri estimaattien keskimääräisten tunnuslukujen sekä niiden hajonnan ja vaihtelun osalta.

Testausten perusteella kokonais- ja tukkitilavuusestimaatit pystytään k-MSN –menetelmällä muodostamaan harhattomasti: parhaalla testausyhdistelmällä tilavuusennusteen keskimääräinen virhe oli kuusella 0,0 % ja männyllä 0,6 %. Leimikoista hieman alle 50 %:ssa tilavuusestimaatin virhe oli alle ± 10 %. Männyllä tulokset olivat kokonaisuutena heikommät, mikä johtunee runkopankin melko pienistä mäntyleimikkomääristä. Kuusen kokonais- ja tukkitilavuuksien ennustevirheiden kehitystä tarkasteltiin myös sumatason leimikkjoukon simuloinnissa, jolla pyrittiin kuvaamaan jonkin alueen hetkellistä leimikkovarantotilannetta ja korjuun suunnittelua. Osoitautui, että keskimääräinen tilavuusvirhe supistui alle 4 %:n tason jo hyvin pienillä, n. 1 000 m³:n puumäärillä ja läheni 0 %:a puumäärän kasvaessa.

Puuston rakenteen ennustamiskykyä tarkasteltiin tukkiprosenttiestimaattien sekä pölkkyjakaumien ennusteasteiden ja jakauma-asteiden avulla. Tukkiisuuden estimointi onnistuu menetelmällä muutaman prosenttiyksikön tarkkuudella, mutta harvennusleimikoissa hajonta oli melko suurta ja hyvin

järeillä leimikoilla tukkiosuus aliarvioitui hieman. Pölkkyjakaumien ennusteasteet kuvaavat puutavaralajien ennustettujen pölkkyjakaumien samankaltaisuutta leimikon todelliseen pölkkyjakaumaan. Ennusteasteet olivat kohtuullisen hyvät kuusella (keskimäärin 84–88 % puutavaralajista riippuen), mutta männyllä vähän heikommät (74–85 %). Jakauma-asteen avulla puolestaan voidaan kuvata, kuinka hyvin k-MSN –menetelmällä muodostetut pölkkyjakaumaennusteet vastaavat runkopankin apterauksen simuloinnissa käytettyjä tavoitejakaumia. Testauksissa saadut ennustettujen pölkkyjakaumien jakauma-asteet olivat varsin korkeita: kuusella keskimäärin 93 % ja männyn tyvitukilla 87 % ja muulla tukilla 91 %.

Puusto- ja pölkkyjakaumaestimaattien tarkkuuteen ja luotettavuuteen vaikuttavat suuresti hakumuuttujina käytettyjen tietojen virheet. Puustoa kuvaavat keskitunnukset ja pinta-ala usein vain arvioidaan ja niissä on aina eriasteista epävarmuutta ja hajontaa, vaikka ne mitattaisiinkin. Jos puustotiedot eivät ole systemaattisesti harhaisia, vaan satunnaisesti tietyn vaihteluvälin rajoissa olevia, ei ennusteidenkaan oleteta olevan systemaattisesti virheellisiä. Tutkimuksessa kokeiltiin satunnaisesti eri maksimivirhetasoilla vaihtelevien hakumuuttuja-arvojen vaikutusta tilavuustunnusten keskimääräiseen ennustevirheeseen. Tämä tehtiin vain tilavuusennusteen mukaan parhailla testausyhdistelmillä. Tarkasteluissa 10 %:n virhetaso ei vielä vaikuttanut merkittävästi ennustevirheisiin, mutta suuremmalla 25 %:n virhetasolla oli jo selvä estimaattien keskimääräistä virhettä ja leimikkotason hajontaa lisäävä vaikutus. Samankaltainen vaikutus on oletettavaa myös muiden tunnusten estimaattien kohdalla, mutta sitä ei erityisesti testattu.

Runkopankin ja k-MSN –menetelmän voitiin osoittaa pystyvän muodostamaan melko luotettavasti keskeiset puustotunnus- ja pölkkyjakaumaestimaatit, kun aineistoa on riittävästi. Testauksissa jäi kuitenkin vielä osoittamatta, olisivatko ennusteet parantuneet niillä kohteilla, joilla vastinleimikkoaineistoa oli niukasti (harvennusleimikot ja osa mäntyvaltaisista leimikoista), mikäli runkopankkiaineistoa olisi ollut enemmän. Ennusteiden tarkkuuteen ja luotettavuuteen vaikuttavat erityisesti tietovarastossa olevien hakkuukohteiden määrä ja edustavuus, puuston määrää ja rakennetta sekä leimikkoa muutoin kuvaavat hakumuuttajat ja niiden systemaattinen tai satunnainen virhe, eräät laskentatekijöihin liittyvät valinnat sekä oleellisena tekijänä myös runkopankin katkonnan simuloinneissa käytetyt puutavaralajit ja niiden läpimitta- ja pituusluokkien määrä. Laatutekijöitä ja runkovikoja huomioivaa katkonnan simulointia ei aineistolle nykyisellä simulaattorilla voitu tehdä.

K-MSN –menetelmän tärkeimpiä käyttökohteita ovat seuraavat alueet:

- Leimikon tai useiden leimikoiden muodostaman varannon puustotietojen muodostaminen puunhankinnan operatiivista suunnittelua varten
- Leimikkovarantojen puutavaralaji- ja katkontavaihtoehtojen vertailut
- Puustotietoarviot puukauppaa varten (puutavaralajikertymät eri katkontavaihtoehtojen ja leimikon hinnoitteluvaihtoehtojen vertailut)

Hyödyntämistä varten puunhankinnan tietojärjestelmiin olisi rakennettava runkopankkitietokannat ja niiden muodostamiseen tarvittavat datansiirto- ja käsittelysovellukset, apteeraussimulaattori sekä itse k-MSN –menetelmän sovellus. Metsätehossa laadittiin sovellus, jossa pohjana on käytetty Joensuun yliopistossa tehtyjä Fortran-ohjelmia. Sitä ajetaan MASI-apteeraussimulaattorin käyttöliittymän kautta. Sovellus on mahdollista tuotteistaa erillisenä työnä ja liittää moduulina yritysten tietojärjestelmiin.

1 TAUSTA JA TAVOITTEET

Puunhankinnan suunnittelussa ja ohjauksessa tarvitaan entistä tarkempaa ennakkotietoa yksittäisten leimikoiden tai leimikkovarantojen puustoista ja puutavaralajikertymistä. Tietoa tarvitaan etenkin puutavaran toimitusmahdollisuuksien arviointiin leimikoittain sekä toiminta-alueittain ja kausittain. Tietoa käytetään tehtaiden tuotannosuunnitteluun, puutavaralaji- ja katkontavaihtoehtojen valintaan ja testaukseen sekä korjuu- ja kuljetussuunnitelmiin tekoon. Puutavaran ominaisuuksille asetettavat tuote- ja asiakasvaatimukset aiheuttavat tarvetta tarkastella eri katkontavaihtoehtojen toteutusmahdollisuuksia ja vaikutuksia puutavaralajikertymien rakenteisiin ja kokonaisvolyyymiin. Tuotteiden raaka-ainevaatimukset aiheuttavat myös tarvetta kuvata puuston laatuominaisuuksia leimikkotasolla. Toiminnan suunnittelussa riittää usein aluetason ja toimitusjakson olemassa olevien varantojen tarkastelu, mutta suunnittelutarpeesta riippuen myös tulevia puunhankintamahdollisuuksia on voitava arvioida laskennallisesti. Puutavaralajikohtaisten kertymien arviointi on perusta lähes kaikelle puunhankintalogistiikan suunnittelulle. Puun ostossa ja käyttöarvon mukaisessa hinnoittelussa tarvitaan leimikoiden puutavaralajikertymien ennakkointia.

Puustotietojen kuvaukseen ja ennustamiseen sekä katkontavaihtoehtojen tarkasteluun on Metsätehossa yhteistyössä Joensuun yliopiston kanssa kehitetty hakkuukoneiden runkokohtaiseen mitattuun tietoon perustuvan tietovaraston, runkopankin, prototyyppejä sekä sen käyttöä varten laskentamenetelmiä ja -sovelluksia. Niitä ovat tietovaraston muodostamisen ohjelmat ja sovellukset, puustotietojen ennustamiseen kehitetty MSN-menetelmä (Most Similar Neighbour) sekä MASI-apteeraussimulaattori. Tietovaraston koostamista ja MSN-menetelmää on kehitetty muutamissa erillisissä projekteissa. Niissä saadut tulokset ovat olleet lupaavia. Menetelmän laskentaperiaatteita, algoritmeja ja ohjelmakoodia ei ole kuitenkaan ennen tätä projektia voitu testata kunnolla runkopankkiaineistoilla, vaihtoehtoisilla vastinleimikkomäärillä eikä sisällöltään ja tarkkuudeltaan erilaisella hakumuuttujatiedolla.

Tämän projektin tavoitteena oli kehittää ja testata vaihtoehtoisia menetelmiä sekä käyttää testauksessa uutta runkopankkiaineistoa. Tavoitteena oli MSN-sovelluksen toimintojen, tiedostorajapintojen ja käyttöliittymän kehittäminen niin, että se soveltuu sekä Metsätehon tutkimustarpeisiin että mahdollisimman joustavasti myös yritysten tietojärjestelmiin liitettäväksi. Työ tehtiin yhdessä Joensuun yliopiston *Hakutieto runkopankin käyttösovelluksessa* -hankkeen kanssa. Joensuun yliopiston hankkeen tavoitteina olivat:

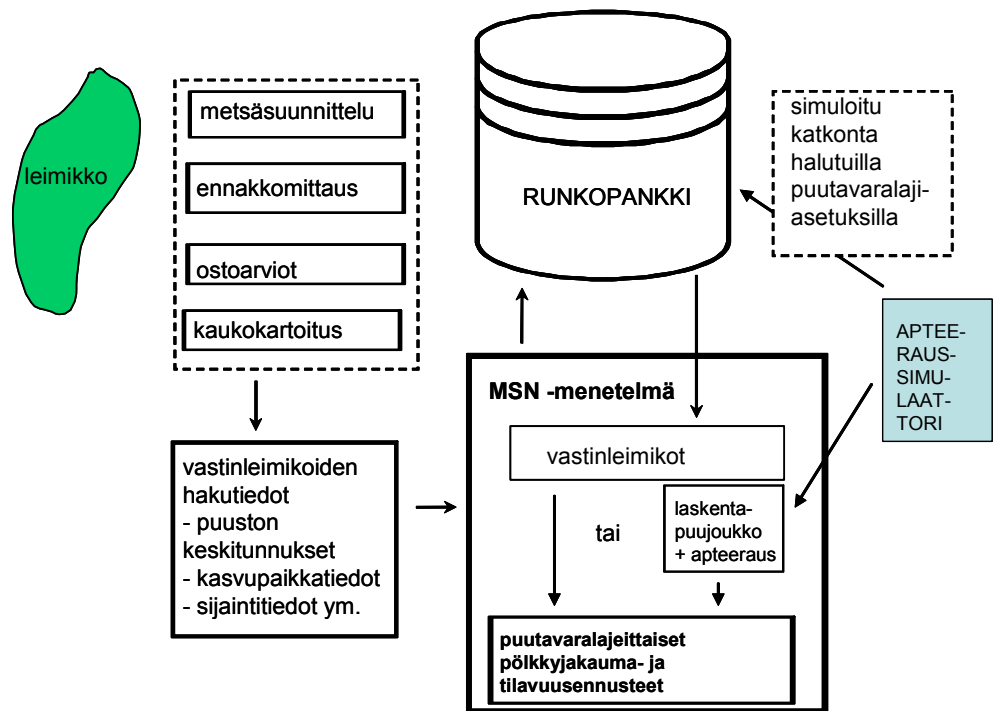
- 1) selvittää runkopankkia hyödyntävien ei-parametristen tukkijakauman ennustamismenetelmien tärkeimmät hakumuuttujat
- 2) k-MSN ja LAN MSN –menetelmien vertailu runkopankkia hyödyntävänä tukkijakauman ennustamismenetelminä
- 3) kehittää hakumuuttujien valinnan automatisointia
- 4) selvittää runkopankkia hyödyntävien k-MSN ja LAN MSN –menetelmien tuottamien tukkijakaumaennusteiden herkkyys hakumuuttujien virheelle sekä leimikon järeyden ja hakkuutavan vaihteluille.

Joensuun yliopiston hankkeessa tarkoituksena oli testauksin selvittää, miten MSN-menetelmää olisi paras soveltaa erilaisiin käyttötilanteisiin (leimikkotyypit, runkopankin laajuus). Menetelmässä voidaan käyttää joko kiinteää tai vaihtuvaa vastinkohdemäärää, jota siis myös testattiin. Tietosisällöltään ja luotettavuudeltaan erilaisen hakutiedon (sopivien hakumuuttujien valinta) käyttöä analysoitiin.

2 LASKENTAMENETELMÄN KUVAUS

2.1 Menetelmän yleiskuvaus

Puustotietojen ennustamiseen kehitetty ei-parametrinen lähimmän naapurin MSN-menetelmä perustuu ennustettavaa tai kuvattavaa leimikkoa tai metsikkökuviota mahdollisimman hyvin vastaavien tietovarastossa (esimerkiksi runkopankissa) olevien hakkuukohteiden valintaan kohteesta käytettävissä olevien perustietojen ja keskeisten puustotunnustietojen tai -arvioiden mukaan (hakumuuttujatiedot). Ennustettavalle leimikolle muodostetaan joko valituiksi tulleiden vastinkohteiden rungoista valittu laskentapuujoukko, joka voidaan apteraussimulaattorilla edelleen katkoa pölkyiksi, tai vaihtoehtoisesti muodostetaan suoraan pölkykohtainen katkontaennuste. Jälkimmäinen tapa edellyttää, että kaikki tietovaraston leimikot apteerataan samoilla puutavaralajimäärityksillä (läpimitta- ja pituusluokat) valmiiksi ja tallennetaan saadut runkojen simulointitulokset tietovarastoon. Ennustetta muodostettaessa näin simuloitua pölkyä poimitaan vastinleimikoista mukaan ennustettavan leimikon pölkykertymiin. Jälkimmäistä menettelyä on sovellettu pääasiallisesti menetelmän kehittämistyössä, sillä jo ensimmäiset analyysit osoittivat sen antavan parempia ja luotettavampia tuloksia (kuva 1).



Kuva 1. Leimikon puustotietojen ja pölkkyjakaumien ennustaminen runkopankkidatan ja MSN-menetelmän avulla.

Ennusteen muodostamisessa erilaista lähtötietoa käyttäen haetaan ennustettavaa leimikkoa parhaiten vastaavia kohteita runkopankkitietokannasta. Vastinkohteiden määrä voi olla vakio (esim. 5 tai 10), jolloin kyseessä on ns. k-MSN –menetelmä (tarkemmin kappaleessa 2.2). On kehitetty ja koekeltu myös LAN MSN –menetelmää (Locally Adaptable Neighbourhood), jossa vastinkohteiden lukumäärä on vaihtuva.

2.2 K lähimmän naapurin MSN -menetelmä

Ei-parametrinen k lähimmän naapurin MSN-menetelmä (k-MSN) on muokattu yhdistämällä k lähimmän naapurin menetelmä sekä Moeur ja Stage:n (1995) kehittämä "*Most Similar Neighbour Inference*" (MSN) –menetelmä. Ei-parametrisellä k lähimmän naapurin menetelmällä saadut ennusteet perustuvat referenssiaineistosta, naapurustosta, poimitujen vastinhavaintojen, naapurien k kappaletta, painotettuihin keskiarvoihin. Menetelmää sovellettaessa on otettava huomioon käytettävä etäisyysfunktio, painotusfunktio ja lopputuloksen laskennassa käytettävien naapurien lukumäärä k. Etäisyysfunktioilla saadaan etäisyys ennustettavan kohteen ja referenssiaineiston havaintojen välille. Etäisyysfunktioon lisättävällä muuttujakohtaisella painotuksella voidaan painottaa niitä muuttujia, jotka tehtyjen tutkimusten perusteella kertovat parhaiten samanlaisuudesta. Naapurien lukumäärällä voidaan vaikuttaa ennusteiden laatuun. Yleisesti voidaan sanoa, että pienellä naapurien määrällä (1–5) pystytään kuvaamaan paremmin luonnollista vaihtelua, ja suuremmalla (5–10) ennusteet keskiarvoistuvat (Malinen ym. 2001, Malinen 2003b).

Tässä tutkimuksessa käytettiin puustotietojen ennustamiseen k lähimmän naapurin menetelmää, jossa ennustettavan kohteen ja referenssiaineiston havaintojen väliset etäisyydet lasketaan MSN-menetelmällä. Toisin sanoen tutkimuksessa käytetyssä k-MSN –menetelmässä etäisyydet laskettiin käyttäen MSN-menetelmän etäisyysfunktioita, ja naapurien lukumäärää k. Runkopankin k kappaletta samankaltaisia leimikoita ovat kohdeleimikon ominaisuuksien ennustamisen perustana, jolloin näistä kohdeleimikon ominaisuuksista johdetaan varsinaiset puusto- ja pölkkyjakaumaennusteet. Seuraavassa on tiivistettynä k-MSN –menetelmä Malisen (2003a) mukaan. Yksityiskohtaisemmin menetelmä on kuvattu Joensuun yliopiston hankkeissa syntyneissä tutkimusraporteissa (Kuuramaa 2004 ja Malinen 2003).

K-MSN –menetelmän laskentavaiheet

- 1) Hakumuuttujien eli selittäjien valinta. Hakumuuttujien tulee kuvata mahdollisimman hyvin selitettäviä ominaisuuksia.
- 2) Ennustettavia ominaisuuksia kuvaavien muuttujien määrittäminen.
- 3) Etäisyysfunktioon tarvittavien hakumuuttujien ja selitettävien muuttujien välisten kanonisten korrelaatioiden laskeminen.
- 4) Painotusmatriisin laskeminen.
- 5) Laskennassa käytettävän naapureiden lukumäärän valinta.
- 6) K-MSN –menetelmän suorittaminen. Algoritmi valitsee referenssiaineistosta samankaltaisimmat naapurit sekä minimoii etäisyysfunktion neliötä.
- 7) Estimaattien muodostaminen lähimpien naapureiden painotetuista keskiarvoista.

2.3 LAN MSN -menetelmä

Tutkimuksessa tarkasteltiin myös Malisen (2003a) kehittämää ei-parametristä LAN (*Locally Adaptable Neighbourhood*) MSN –menetelmää leimikon tukkijakauman ennustamisessa. LAN MSN –menetelmässä naapurustosta haetaan sellaisia naapurikombinaatioita, joiden yhteinen hakumuuttujien keskiarvojen vektori on lähimpänä kohdeaineiston hakumuuttujien vektoria. Jokainen mahdollinen naapurikombinaatio muodostetaan ja niiden ennustettavan muuttujan keskiarvot lasketaan.

Jokaisen naapurikombinaation hakumuuttujan keskiarvoa verrataan kohdeaineiston hakumuuttujan arvoon käyttämällä MSN-etäisyysfunktioita. Naapurikombinaatiota, jolla hakumuuttujan keskiarvo on lähimpänä kohdeaineiston hakumuuttujan arvoa, käytetään estimoitavan muuttujan laskemiseen. Laskenta-ajan pitämiseksi mielekkäänä on naapuruston kokoa, josta optimaalinen naapurikombinaatio haetaan, rajoitettu 15 naapuriin. Vertailtavan naapuruston koon (n kappaletta naapureita) kasvaessa tietokoneen käyttämä laskenta-aika kasvaa 2^n .

Tutkimuksessa suoritettiin k-MSN ja LAN MSN –menetelmien vertailu, joka jakautui kahteen osaan hakumuuttujien valinnan tavoin. Ensimmäiseksi suoritettiin samoilla hakumuuttujilla laskettujen tukkijakaumien ennusteiden hyvyyden vertailu, jolloin ennusteiden laskennassa hakumuuttajat valittiin subjektiivisella menetelmällä. Toiseksi näiden kahden menetelmän vertailua suoritettiin hakumuuttujien valinnan automatisoinnin yhteydessä.

3 HAKUMUUTTUJIIEN VALINTAMENETELMÄT

3.1 Hakumuuttujien merkitys

MSN-menetelmän käytössä on tärkeää löytää runkopankista ennustettavaa kohdeleimikkoa vastaavat, mahdollisimman samankaltaiset vastinleimikot. Tähän tarvitaan hakumuuttujia, joiden ominaisuudet selittävät mahdollisimman hyvin leimikon puustotietoa ja sen vaihtelua. Tiettyjen hakumuuttujien avulla haettujen leimikoiden tulisi olla puustotiedoiltaan samankaltaisia kohdeleimikon puustotiedon kanssa. Aiemmin Malinen ym. (2001) ovat selvittäneet hakumuuttujien merkitystä saataviin ennusteisiin. Tutkimuksessa hakumuuttujina käytettiin hakkuukohdetta sekä puustoa kuvailevia tavanomaisesti mitattuja muuttujia, jolloin tulokset olivat paremmat kuin käytettäessä ainoastaan hakkuukohdetta tai puustoa kuvailevia muuttujia. On tarpeen määrittää sellaiset hakumuuttujat, jotka ovat helposti ja kustannustehokkaasti määritettävissä sekä joilla haetut naapurit antavat riittävän tarkkoja ennusteita. Soveliimmat hakumuuttujat tulisi pystyä valitsemaan jo olemassa olevista lähteistä, joita ovat muun muassa kuvioittaiset metsäsuunnitelmatiedot sekä leimikonsuunnittelun yhteydessä tuotettu informaatio. Tulevaisuudessa kaukokartoitusmenetelmillä (numeeriset ilmakuvat, korkearesoluutioiset satelliittikuvat ja laserskannaus) voitaneen saada riittävän luotettavia puustotunnusestimaatteja esimerkiksi puulajeittaisista runkoluvuista sekä latvusprofiileista johdetuista keskiläpimitoista.

Hakumuuttujien valinta tulisi olla runkopankin käytännön sovelluksessa mahdollisimman vaivatonta sekä riittävän hyviin leimikkotason ennusteisiin johtavaa. Aiemmissa ei-parametrisiä menetelmiä koskevissa tutkimuksissa on hakumuuttujat määritelty manuaalisesti. Joensuun yliopiston hankkeessa käytettiin askeltavaa regressioanalyysiä aineiston kuvaamiseen, muuttujien välisten yhteyksien selvittämiseen sekä osittain hakumuuttujien valintaan. Huomiota on kiinnitetty hakumuuttujien valintaan sekä etenkin valinnan automatisointiin ja sen kehittämiseen. Hakumuuttujien valinnan automatisoinnilla pyritään sisällyttämään valinta ei-parametriseen ennustamismenetelmään, jotta erillinen valintamenetelmä voitaisiin sivuuttaa. Hakumuuttujien valinnan automatisoinnin perustana olivat optimointialgoritmit, joilla parhaimman ennusteen antavat muuttujat poimitaan. Nämä algoritmit sisällytettiin käytettyihin ei-parametrisiin ennustamismenetelmiin, jolloin valinta oli osa ennusteiden laskentaprosessia.

3.2 Hakumuuttujien selvittäminen askeltavalla regressio-analyysillä

Soveltava parametrinen menetelmä hyvien selittäjien poimimiseen laajasta muuttujajoukosta on askeltava regressioanalyysi, joka päätettiin ottaa myös tässä tutkimuksessa yhdeksi hakumuuttujien valintatyökaluksi. Tämän lisäksi askeltavaa regressioanalyysiä käytettiin välineenä, jolla pyrittiin hahmottamaan aineistoa sekä analysoimaan muuttujien merkitsevyyttä selittäjinä ja selitettävinä. Saadulla regressiomallilla ei itsessään pyritty ennustamaan runkotietoja, vaan regressioanalyysin esiin tuomat muuttujat ovat k-MSN-laskennassa selittäjien perusjoukko. Askeltava regressioanalyysi tehtiin puulajeittain erikseen jokaiselle selitettävälle muuttujalle muunnokset mukaan lukien.

3.3 Hakumuuttujien soveltuvuuden arviointi

Löydettyjen hakumuuttujien soveltuvuutta voidaan arvioida erilaisin menetelmin ja kriteerein. Joensuun yliopiston hankkeessa soveltuvuutta arvioitiin saatujen pölkkyjakaumaennusteiden hyvyksien perusteella. Ennusteiden hyvyttä arvioitiin pölkkyjakauman ennusteasteen avulla (kaava 1). Tunnus on yhteneväinen yleisesti käytetyn katkonnan onnistumista kuvaavan jakauma-asteen kanssa (jakauma-aste -termiä käytetään Kuuramaan graduutyössä). Tunnus kertoo, kuinka monta prosenttia kohdeleimikon pölkyistä menetelmä sijoitti oikeaan pituus-läpimittaluokkayhdistelmään. Tarkasteluissa leimikon ”todellinen” pölkkyjakauma ei ole alkuperäinen hakkuukoneen katkontatuloks, vaan simuloinneilla tuotettu pölkkyjakauma (ks. kpl 4 Aineisto). Ennusteastetta laskettaessa ei verrata ennustetta siis simuloinnissa käytettyjen puutavaralajien tavoitematriisien tavoitearvoihin, vaan simuloinneilla tuotettuihin pölkkyjakaumiin. Ennusteasteen laskennassa arvoa painotettiin läpimittaluokkien pölkky määrillä.

$$Pölkkyjakauman_ennusteaste = 1 - \frac{\sum_{i=1}^n |todellinen - ennuste|}{2} \quad (1)$$

Hakumuuttujien soveltuvuuden selvittämisessä ei tarkasteltu puutavaralajien tilavuusennusteiden virheitä. Se olisi saattanut olla tarpeen tehdä, jotta voitaisiin varmistaa menetelmän ennustekyky myös puuston määrää kuvaavien tunnusten osalta. Pölkkyjakauman ennusteaste kuvaa lähinnä menetelmän soveltuvuutta puuston rakenteen ennustamiseen (puulajisuhteet, puuston koko ja järeys). Metsätehon tekemissä menetelmän testauksissa tarkasteltiin kuitenkin myös puuston määrän ennustamista pölkkyjakauman ennusteasteen lisäksi.

3.4 Hakumuuttujien subjektiivinen valinta

K-MSN –laskennassa selittäjien merkitsevyys saattaa poiketa regressioanalyysin osoittamasta. Tämän vuoksi ei tyydytty askeltavan regressioanalyysin osoittamiin hakumuuttujiin, vaan laskennan yhteydessä hakumuuttajat valittiin maksimoimalla saaduista pölkkyjakaumaennusteista laskettuja ennusteasteita. Jotta niitä voitiin vertailla, päätettiin laskennassa käytettäväksi $k:n$ arvoksi 10, jota käytettiin laskennassa läpi koko tutkimuksen.

Ennusteasteita kasvatettaessa laskettiin pölkkyjakaumaennusteet sekä näille ennusteasteet käyttäen askeltavalla regressioanalyysillä saadun hakumuuttujien perusjoukon lisäksi jokaista jäljelle jäänyttä mahdollista hakumuuttujaa ja/tai sen muunnosta yksitellen. Askeltavalla regressioanalyysillä saatuun hakumuuttujien perusjoukkoon lisättiin yksitellen jokainen hakumuuttuja, jonka jälkeen laskettiin pölkkyjakaumaennusteet ja näille ennusteasteet. Tämän jälkeen parhaimmat ennusteasteet antaneisiin hakumuuttujiin lisättiin jälleen jokainen hakumuuttuja yksitellen ja näillä uusilla hakumuuttujilla laskettiin pölkkyjakaumaennusteet ja niille ennusteasteet. Tämä silmukka suoritettiin kuusi kertaa, jonka jälkeen hakumuuttujien perusjoukon ulkopuolisista hakumuuttujista poistettiin yksitellen jokainen ja laskettiin pölkkyjakaumaennusteet ja niille ennusteasteet. Jos ne kasvoivat, jätettiin hakumuuttuja pois. Tällä tavalla saatiin pölkkyjakaumaennusteita parannettua ja hakumuuttujien määrän kasvua rajoitettua.

3.5 Hakumuuttujien valinnan automatisointi

Kehitettyssä hakumuuttujien valinnan automatisointimenetelmässä laskettiin ensimmäiseksi joko k -MSN tai LAN MSN –menetelmällä pölkkyjakaumaennusteet sekä näille ennusteasteet käyttäen yksinomaan aineistossa olevia muuttamattomia muuttujia hakumuuttujina sekä selitettävinä. Jotta tuloksia voitiin vertailla aiemmin esitetyn hakumuuttujien valintamenetelmän antamien tulosten kanssa, käytettiin myös tämän menetelmän k -MSN –laskennassa $k:n$ arvoa 10. Ensimmäisten k -MSN/LAN MSN –pölkkyjakaumaennusteiden ja ennusteasteiden jälkeen lisättiin yksitellen hakumuuttujaksi jokaisesta vähintään välimatka-asteikollisesta muuttujasta neliömuunnos, jonka jälkeen laskettiin uudet ennusteet. Jos ennusteasteet kasvoivat, jätettiin neliömuunnos hakumuuttujaksi. Tämän jälkeen lisättiin seuraavan hakumuuttujan neliömuunnos ja laskettiin uudet ennusteet. Samalla tavalla tarkasteltiin myös mainituista hakumuuttujista neliöjuuri-, käänneiluku- ja luonnollinen logaritmuunnos. Seuraavaksi jätettiin jokainen aineiston hakumuuttuja ja sen muunnokset yksitellen pois ja laskettiin jälleen uudet pölkkyjakaumaennusteet sekä ennusteasteet. Jos hakumuuttujan pois jättäminen ei pienentänyt ennusteasteita, jätettiin se pois. Tämän jälkeen aloitettiin edellä mainittujen muunnosten lisääminen jälleen yksitellen samoin periaattein kuin aiemmin, kuitenkin niin, että sama muunnos voitiin valita myös toisen kerran. Optimointialgoritmi voi näin painottaa jotakin muunnosta. Muunnosten poimimista jälleen hakumuuttujiksi pidettiin laskennan toisena silmukkana. Näiden silmukoiden määrää rajoitettiin kolmeen, koska kolmannen silmukan jälkeen ennusteasteiden muutokset olivat hyvin pieniä.

Lisäksi optimointialgoritmi pyrittiin pitämään laskennallisesti mahdollisimman kevyenä.

Ennusteasteen lisäksi hakumuuttujien automaattisella valinnalla saatuja pölkkyjakaumaennusteita arvioitiin tarkastelemalla puutavaralajiosuuksia sekä tilavuusennusteiden harhattomuutta, ja näin päästiin arvioimaan hakumuuttujien automaattista valintaa. Puutavaralajiosuuksia tarkasteltiin vertailemalla todellisia ja ennustettuja tukkiprosentteja. Ennustettujen tukkijakaumien tilavuuksien harhattomuutta tarkasteltiin leimikoiden tilavuuksien kumulatiivisen virheen avulla. Näiden tarkastelujen yhteydessä suoritettiin myös k-MSN ja LAN MSN –menetelmien vertailua.

4 AINEISTO

Menetelmän kehitystyössä käytettiin Metsätehossa koottua runkopankkia, jonka muodostivat 485 hakkuukohteelta tallennetut männyn, kuusen ja koivun runko- ja katkontatiedot (hakkuukoneen stm- ja prd-tiedostot) sekä kohteita kuvailevat tiedot. Näistä muodostettiin 419 kohteen tutkimusaineisto poistamalla virheelliset ja puutteelliset hakkuukohteet. Aineisto sisälsi taulukossa 1 esitetyt muuttujat. Lisäksi tutkimusaineistoa muokattiin askeltavaa regressioanalyysiä varten muodostamalla jokaisesta puustoa kuvailevasta muuttujasta neliö-, neliöjuuri-, käänteisluku- ja logaritmuunnokset. Näitä aineistoja käytettiin Joensuun yliopiston hankkeessa. Metsätehossa tehtävien testiajojen perusjoukkona oli alkuperäinen 485 kohteen aineisto, mutta osaa analyyseja varten siitä poistettiin 49 pinta-alaltaan todennäköisesti virheellistä leimikkoa.

Koska useimmat tutkimusaineiston hakkuukohteet olivat heterogeenisiä puulajijakauman suhteen, tarkasteltiin leimikoita puulajeittain. Männyn runkotietoja tutkimusaineistossa oli 404 kohteelta, kuusen 401 ja koivun 390 kohteelta. Kohteista oli päätehakkuita 63 prosenttia, harvennushakkuita 19,3 prosenttia, siemen- tai suojuspuuhakkuita 13 prosenttia ja siemen- tai suojuspuiden poistohakkuita 4,7 prosenttia. Taulukossa 2 on esitetty keskeisiä koko tutkimusaineiston kattavia tunnuksia. Runkopankin havaintoyksiköiden tunnusten laaja vaihtelu selittyy hakkuukohteiden puulajijakauman heterogeenisyydellä.

TAULUKKO 1 Tutkimusaineiston havaintoyksiköiden muuttujat.

Puustoa kuvailevat muuttujat	puulajiosuus: puulajin käyttöosan tilavuusosuus (plpros) pohjapinta-ala (ppa) ppa-mediaanipuun tilavuus (V_{gm}) ppa-mediaanipuun rinnankorkeusläpimitta (d_{gm}) ppa-mediaanipuun pituus (h_{gm}) puulajin käyttöosan hehtaarikohtainen tilavuus (til/ha) puulajin keskijäreys (keskij) Näslundin pituuskäyräyhtälön parametri a (Näslund a) Näslundin pituuskäyräyhtälön parametri b (Näslund b) ppa:n prosenttiosuuspisteet 0, 20, 40, 60, 80 ja 100 (prpist0, prpist20, prpist40, prpist60, prpist80, prpist100) (<i>runkolukusarjan ne $d_{1,3}$ -arvot, joissa leimikon ko. puulajin pohjapinta-alakerätyä saavuttaa esim. 20 % rajan, 0=minlpm, 100=maxlpm</i>) runkojen lukumäärä puulajeittain (rlkm)
Hakkuukohtetta kuvailevat muuttujat	puulaji (pl) p-koordinaatti (pkoor) i-koordinaatti (ikoor) kohteen pinta-ala (A) hakkuutapa (ht) kohteen pääpuulaji (pääpl) metsätyyppi (mt) puuston jaksoisuus (jakso) ikä

K-MSN sekä LAN MSN –laskennassa hakkuutapa, kohteen pääpuulaji, metsätyyppi ja puuston jaksoisuus muutettiin dummy-muuttujiksi. Lisäksi laskentaan mukaan otettavien kohteiden runkojen lukumäärän tuli olla vähintään kymmenen eli aineiston havaintoyksiköt, joiden runkojen lukumäärä oli vähemmän kuin kymmenen, jätettiin laskennan ulkopuolelle. Tällöin mäntykohteita oli 371, kuusikohteita 387 ja koivukohteita 354.

TAULUKKO 2 Tutkimusaineiston tärkeimpiä tunnuslukuja puulajeittain.

	Mänty, $n=404$	Kuusi, $n=401$	Koivu, $n=390$
Pinta-ala (ha)			
minimi	0,2	0,2	0,2
keskiarvo	2,4	2,4	2,4
maksimi	20	20	20
Ikä (a)			
minimi	30	30	30
keskiarvo	90	90	89
maksimi	160	160	155
Ppa/ha (m^2/ha)			
minimi	0,003	0,004	0,004
keskiarvo	5,7	12,1	1,4
maksimi	47,4	53,4	53,8
Keskiläpimitta, d_{gm} (cm)			
minimi	10,2	11,6	8,6
keskiarvo	28	25,2	19,5
maksimi	46,6	40,9	40,2
Keskipituus h_{gm} (m)			
minimi	9,6	8,5	9,8
keskiarvo	22,4	20,2	19,3
maksimi	32	30	29,7

Koska menetelmäkehityksessä ennustetulosten arviointi perustui pitkälti leimikon todellisten ja ennustettujen tukkijakaumien vertailuun, täytyi tutkimusaineiston rungot katkoa puutavaralajeiksi yhtenäisellä tavalla. Alkuperäisiä puutavaralajeja ja pölkytystietoja ei siis ollut mielekäästä käyttää. Muutoinkaan eri lähteistä ja ajankohdilta kootun runkopankkiaineiston alkuperäistä pölkytystä ei voida käyttää ennusteen teossa erilaisten puutavaralajien mittojen ja jakaumatavoitteiden vuoksi. Tutkimusaineiston runkojen apteeraus tehtiin MASI-simulaattorilla. Kuitupuun jakaumien tarkastelu jätettiin tutkimuksen ulkopuolelle, mutta kuitupuu oli mukana pölkytysten simuloinneissa. Männyn pölkytys tehtiin käyttäen kolmea puutavaralajia, jotka olivat tyvitukki, tukki ja kuitu. Kuusen osalta tutkimuksessa käytettiin kahta eri pölkytystä, jotka on eritelty myöhemmin merkinnöillä kuusi¹ ja kuusi². Kuusen¹ pölkytyksessä käytettiin kolmea puutavaralajia; tyvitukkia, tukkia ja kuitua. Vastaavasti kuusen² pölkytys koostui kahdesta puutavaralajista, tukista ja kuidusta. Tavaralajeille oli määritelty 7–9 pituusluokkaa (kuusen tyvitukilla 1 pituusluokka, sorvitukki) ja 11–15 läpimittaluokkaa. Pohjana luokitukselle oli todellinen erään osakasyhtiön puutavaralajimäärittely, jota hieman yksinkertaistettiin. Koivun pölkytys simuloitiin käyttäen kahta puutavaralajia, tukkia ja kuitua. Joensuun hankkeessa tehtiin koivulle samat tarkastelut kuin muillekin puulajeille, mutta Metsätehon testauksissa koivua ei käsitelty.

5 TESTAUKSET JA TULOKSET

5.1 Askeltavalla regressioanalyysillä poimitut selitettävät muuttujat

Selitettävänä k-MSN ja LAN MSN –laskennassa käytettiin muuttujia tai niiden muunnoksia, joita selitettäessä selitysasteet nousivat korkeimmiksi regressioanalyysissä (taulukko 3) pois lukien k-MSN ja LAN MSN –laskennat, joissa hakumuuttujat valittiin automaattista menetelmää käyttäen. Tällöin selitettävänä käytettiin alkuperäisiä muuntamattomia muuttujia. Hakumuuttujien subjektiivista valintaa käytettäessä suoritettiin vertailu, jossa k-MSN –laskennassa käytettiin yksitellen kaikkia selitettävien muuttujien muunnoksia. Tämä vertailu osoitti, että askeltavan regressioanalyysin osoittamat selitettävien muuttujien muunnokset johtivat myös k-MSN –laskennassa korkeimpiin pölkyjakauman ennusteasteisiin.

TAULUKKO 3 Regressioanalyysillä poimitut selitettävät muuttujat, joita käytettiin subjektiivisesti valittujen hakumuuttujien yhteydessä k-MSN ja LAN MSN –laskennoissa.

Mänty	Kuusi ¹ ja ²	Koivu
(Näslund a) ²	(Näslund a) ²	Näslund a
ln(Näslund b)	Näslund b	Näslund b
1/prpist0	1/prpist0	1/prpist0
1/prpist20	1/prpist20	1/prpist20
prpist40	prpist40	prpist40
prpist60	prpist60	prpist60
1/prpist80	1/prpist80	1/prpist80
1/prpist100	1/prpist100	1/prpist100
ln(V)	ln(V)	ln(V)
ln(rlkm)	ln(rlkm)	ln(rlkm)
ln(keskij)	ln(keskij)	ln(keskij)

5.2 Hakumuuttujat

5.2.1 Subjektiivinen valinta

Aluksi k-MSN ja LAN MSN –laskennassa pyrittiin hakumuuttujina käyttämään aineistosta askeltavalla regressioanalyysillä poimittuja muuttujia ja niiden muunnoksia, mutta tällöin hakumuuttujien määrä olisi ollut erittäin suuri. Askeltavan regressioanalyysin mukaista hakumuuttujien määrää pienennettiin, jolloin eräänlaisena hakumuuttujien perusjoukkona käytettiin askeltavalla regressioanalyysillä saatuja muuntamattomia muuttujia. Tämä hakumuuttujien perusjoukko on esitetty taulukossa 4.

TAULUKKO 4 K-MSN ja LAN MSN –laskennassa käytetty askeltavalla regressioanalyysillä poimittu hakumuuttujien perusjoukko.

Mänty	Kuusi ^{1 ja 2}	Koivu
d _{gm} plpros ppa A	d _{gm} plpros ppa A	d _{gm} V _{gm} ppa A

K-MSN ja LAN MSN –ennusteiden laskennassa hakumuuttujiksi valikoituivat aiemmin kuvatulla subjektiivisella hakumuuttujien valintamenetelmällä taulukossa 5 esitetyt muuttujat tai niiden muunnokset, regressioanalyysillä saatujen hakumuuttujien perusjoukon lisäksi.

TAULUKKO 5 Hakumuuttujien subjektiivisella valinnalla valikoituneet k-MSN ja LAN MSN –laskennan muuttujat.

Mänty	Kuusi ^{1 ja 2}	Koivu
plpros ² h _{gm} ² 1/plpros ppa ² pääpl	ln(plpros) ln(ppa) ppa ^{0,5} ikä jakso pääpl	plpros ^{0,5} ppa ² ht

5.2.2 Automaattinen valinta k-MSN ja LAN MSN –menetelmillä

Seuraavaksi on lueteltu taulukossa 6 ne hakumuuttujat, jotka optimointialgoritmi poimi automaattisessa valinnassa ja joita k-MSN –menetelmä käytti lopullisten tulosten laskennassa. Muuttujan jäljessä oleva luku (sulkeissa) ilmaisee, kuinka monta kertaa muuttuja on poimittu mukaan.

TAULUKKO 6 Automaattisella valinnalla saadut k-MSN –laskennan hakumuuttujat.

Mänty	Kuusi ¹	Kuusi ²	Koivu
ht(3)	ht(2)	ht(3)	ht(2)
pääpl	pääpl(2)	pääpl(2)	pääpl(2)
mt	mt	mt(2)	mt(2)
jakso(2)	jakso(3)	pkoor	jakso(2)
pkoor	pkoor	ikoor	ikoor
ikoor	ikoor	A	A
A	A	plpros	plpros
ikä	plpros	ikä	ikä
h_{gm}	ikä	ppa	ppa
$ikoor^2$	ppa	V_{gm}	V_{gm}
$plpros^2(3)$	d_{gm}	d_{gm}	d_{gm}
d_{gm}^2	$plpros^{0,5}$	h_{gm}	h_{gm}
$pkoor^{0,5}$	$ikä^{0,5}$	$pkoor^2$	$pkoor^2$
$ikä^{0,5}$	$V_{gm}^{0,5}$	$A^{0,5}$	A^2
$ppa^{0,5}$	ppa^2	$plpros^{0,5}$	ppa^2
$d_{gm}^{0,5}$	1/ikoor	$ikä^{0,5}$	$pkoor^{0,5}$
$ikä^2$	1/pkoor	$V_{gm}^{0,5}$	1/ikoor
h_{gm}^2		$d_{gm}^{0,5}$	1/plpros
1/plpros		1/ikoor	
		$ikoor^2$	

Tutkimuksessa selvitettiin myös hakumuuttujien automaattista valintaa, joka perustuu LAN MSN –pölkkyjakauman ennusteista laskettujen jakaumasteiden maksimointiin. Taulukossa 7 on lueteltu ne hakumuuttujat, jotka optimointialgoritmi poimi automaattisessa valinnassa ja joita LAN MSN –menetelmä käytti tulosten laskennassa. Muuttujan jäljessä oleva luku (sulkeissa) ilmaisee, kuinka monta kertaa muuttuja on poimittu mukaan.

TAULUKKO 7 Automaattisella valinnalla saadut LAN MSN –laskennan hakumuuttujat.

Mänty	Kuusi ¹	Kuusi ²	Koivu
ht(2)	ht(2)	ht(3)	ht(3)
pääpl	pääpl	pääpl(2)	pääpl(2)
mt(2)	mt	mt(2)	mt(2)
pkoor	jakso(2)	jakso(2)	jakso(2)
ikoor	pkoor	pkoor	pkoor
A	ikoor	ikoor	ikoor
plpros	A	A	A
ikä	plpros	plpros	plpros
ppa	ikä	ikä	ikä
V _{gm}	ppa	ppa	ppa
d _{gm}	V _{gm}	V _{gm}	V _{gm}
h _{gm}	d _{gm}	d _{gm}	d _{gm}
A ²	h _{gm}	h _{gm}	h _{gm}
ppa ²	pkoor ²	A ² (2)	pkoor ²
d _{gm} ²	A ²	plpros ²	A ²
1/ikoor	plpros ² (2)	pkoor ^{0.5}	ppa ²
1/pkoor	ppa ²	ikoor ^{0.5}	ikoor ^{0.5}
1/V _{gm}	V _{gm} ²	1/A	ppa ^{0.5}
	d _{gm} ²		1/ikoor
	h _{gm} ^{0.5}		1/ikä
	1/plpros		pkoor ^{0.5}
	1/V _{gm}		plpros ²
	1/ppa		
	1/pkoor		

5.3 Menetelmän testaukset

5.3.1 Testatut hakumuuttuja-, aineisto- ja menetelmäparametrijhdistelmät

K-MSN –menetelmää testattiin Metsätehossa Joensuun hankkeessa saatujen tulosten perusteella. Testauksissa päädyttiin käyttämään niitä hakumuuttujia, jotka olivat tulosten perusteella määritetty ns. perushakumuuttujiksi (taulukot 4 ja 5), mutta ilman muuttujien muunnoksia. Testauksissa käytetyt hakumuuttujat on esitetty taulukossa 8.

TAULUKKO 8 Testauksissa käytetyt hakumuuttujat (ks. muuttujien selitykset taulukosta 1).

Mänty	Kuusi ^{1 ja 2}	Koivu
d _{gm}	d _{gm}	d _{gm}
h _{gm}	plpros	V _{gm}
plpros	ppa	plpros
ppa	A	ppa
A	ikä	A
pääpl	jakso	ht
	pääpl	

Testaukset tehtiin vain kuuselle ja männylle. LAN MSN –menetelmää ei testattu. Testauksiin muodostettiin testausyhdistelmät seuraavin perustein:

- kuusisorvitukki mukana tai jätettiin pois (kuusi¹ tai kuusi² –pölkkytys)
- puulajikohtaiset perushakumuuttajat muuten aina mukana, mutta osassa testausyhdistelmistä kokeiltiin pinta-alan jättämistä pois
- aineistossa kohteiden runkomäärärajaukset
 - 30, 50 ja 100 kpl runkoja / ko. puulaji
- pinta-alavirheellisiksi tulkittujen kohteiden poisto aineistosta osassa testausyhdistelmiä
- skaalaustekijänä pinta-ala tai puulajin edustama osuus leimikon pinta-alasta
- k=10 pääsääntöisesti, mutta kuusen parhaalla testausyhdistelmällä kokeiltiin k=5
- perushakumuuttajien lisäksi hakumuuttujina ppa:n prosenttiosuus piste 25 tai 75 tai molemmat (läpimittajakauman ala- ja yläkvartiilit)

Eri tavoin muodostettuja testausyhdistelmiä oli kuusella 15 ja männyllä 14. Testausyhdistelmät on esitetty liitteessä.

5.3.2 Ennusteiden hyvyyden kriteerit

K-MSN –menetelmän ennustekykyä arvioitiin vertailemalla testausyhdistelmiä keskenään käyttäen seuraavia tarkastelukriteereitä (kaikki tarkastelut tehtiin puulajikohtaisina):

1. Kokonais- ja tukkitilavuuden keskimääräiset ennustevirheet. Leimikon ennustetun ja todellisen kokonais- tai tukkitilavuuden eron suhde todelliseen tilavuuteen. Testausyhdistelmän keskimääräinen ennustevirhe laskettu leimikoiden todellisilla tilavuuksilla painotettuna.
2. Puulajin kokonais- ja tukkitilavuuden ennustevirheeltään alle $\pm 10\%$, alle $\pm 20\%$ ja alle $\pm 30\%$ olevien leimikoiden osuus kaikista testausyhdistelmän leimikoista.
3. Tukkiprosentin ennustetarkkuus.
4. Tukkipuutavaralajien pölkkyjakaumien ennusteasteet ja jakaumaasteet. Ennustetun pölkkyjakauman pituus-/läpimittaluokittaisia pölkkymääriä verrattiin kohteen todellisena pidettävän (simuloimalla muodostetun) pölkkyjakauman vastaaviin. Näin muodostettiin puutavaralajin pölkkyjakauman ennusteasteet kaikille leimikoille. Testausyhdistelmän keskimääräisen ennusteasteen laskennassa painotus tehtiin leimikoiden todellisilla ko. puutavaralajin tilavuuksilla. Jakauma-aste laskettiin puutavaralajeille läpimittaluokittain verraten pituusluokkien suhteellisia pölkkyosuuksia simuloidussa apterauksessa olleisiin apt-tiedoston vastaavien läpimitta-/pituusluokkien jakaumatavoitteisiin. Testausyhdistelmän keskimääräinen jakaumaaste koko aineistossa laskettiin painottaen leimikoiden todellisilla ko. puutavaralajin tilavuuksilla.
5. Kokonais- ja tukkitilavuuksien ennustevirheet sumatasolla tarkastellen. Kaikkien ennustettujen leimikoiden puumäärä summattiin 100 simuloinnissa sijoittaen leimikot satunnaisesti eri järjestykseen. Näin kumuloituvaa suman ennustettua kokonais- tai tukkitilavuutta verrattiin vastaavaan suman todelliseen tilavuuteen 100 m³:n luokissa. Keskimääräinen ennustevirhe laskettiin simulointien keskiarvona vastaaville 100 m³:n luokille.

5.3.3 Hakumuuttujien virheen vaikutus

Hakumuuttujina käytettävät tiedot vaihtelevat aina käytännössä tarkkuudeltaan, ja niiden hajonta voi olla keskenään erisuuruista tiedosta ja sen mittaus- tai hankintatavasta riippuen. Tutkimuksessa haluttiin herkkyysanalyysin avulla selvittää, miten hakumuuttujien satunnaiset virheet vaikuttavat ennustetuloksiin. Oletuksena oli, että minkään puustotunnuksen arvioinnissa tai mittauksessa ei ole olemassa systemaattista virhettä, vaan virheet ovat enemmänkin satunnaisia ja mittaajasta, kohteesta tai mittausolosuhteista johtuvia, mutta vaihteluväli on silti asetetuissa rajoissa.

Analyysit tehtiin kahdella eri virhetasolla siten, että kaikkien perushakumuuttujien (keskiläpimitta, puulajiosuus, pohjapinta-ala ja pinta-ala) lähtötietojen annettiin vaihdella satunnaisesti $\pm 10\%$:n ja $\pm 25\%$:n välillä. Laskennat tehtiin vain kuusen testausyhdistelmille 7 (paras tilavuusennuste) sekä 12. Virhesimuloinnit ja ennusteet tehtiin viiteen kertaan, joista tuloksiin laskettiin keskiarvot.

5.4 Testausten tulokset

5.4.1 Tilavuus- ja tukkiprosenttiennusteet

Kokonaistilavuuden ja tukkitilavuuden keskimääräiset ennustevirheet testausaineistossa eri testausyhdistelmillä on esitetty taulukossa 9. Keskimääräiset ennustevirheet on laskettu leimikon puulajikohtaisilla tilavuuksilla painotettuina. Ennustevirhe = $(\text{ennustettu tilavuus} - \text{todellinen tilavuus}) / \text{todellinen tilavuus} * 100$.

Kuusella kaikilla testausyhdistelmillä saavutettiin tyydyttävä tai hyvä ennustetarkkuus: ennustevirhe oli sekä tukkitilavuuden että kokonaistilavuuden osalta alle 5 %. Siksi on oletettavaa, että itse ennustemenetelmässä, sen algoritmissa tai sovelluksen ohjelmakoodissa ei ole mitään systemaattista virhelähdettä. Kuusella aineiston määrä ennusteen tekemiseen oli myös riittävä. Parhaimmilla testausyhdistelmillä saavutettiin alle 1 %:n keskimääräinen ennustevirhetaso.

TAULUKKO 9 Kokonaistilavuuden ja tukkitilavuuden keskimääräiset ennustevirheet k-MSN –menetelmän erilaisilla testausyhdistelmillä. Testausaineistona 467 kuusta ja 473 mäntyä sisältänyttä leimikkoa.

Testausyhdistelmä	Tilavuusennusteen virhe, %			
	Kuusi		Mänty	
	Tukki	Kokonaistilavuus	Tukki	Kokonaistilavuus
1	3,5	3,3	1,1	0,6
2	3,8	3,5	1,7	1,3
3	3,1	2,8	-0,5	-1,0
4	-0,9	0,1	12,1	11,4
5	0,4	1,4	18,0	17,2
6	3,1	3,8	2,6	2,5
7	0,3	0,0	12,5	11,6
8	1,7	1,4	18,9	17,8
9	4,9	4,8	4,6	4,0
10	1,9	1,6	13,4	12,7
11	1,8	1,5	15,0	15,0
12	1,3	1,1	11,1	10,4
13	0,7	0,9	11,0	10,4
14	1,6	1,6	12,3	11,8
15	1,2	1,5		

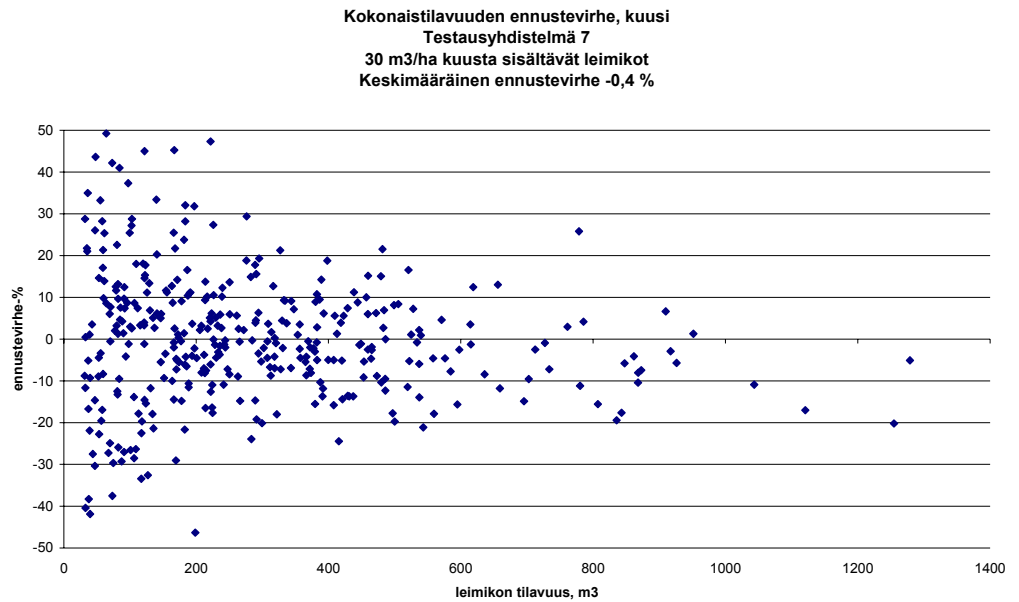
Männyllä ei saatu yhtä hyviä tuloksia. Parhailla testausyhdistelmillä kyllä päästiin kuusen ennustevirheen tasolle, mutta useimmilla keskimääräinen ennustevirhe ja myös sen hajonta oli merkittävästi suurempi. Suurimpana syynä tähän lienee ollut mäntykohteiden heterogeenisuus sekä se, että kohdeaineistossa mänty esiintyi useammin sivupuulajina kuin kuusi. Runkomäärältään vähän mäntyä sisältävien kohteiden rajaaminen pois vastinleimikoiden haussa antoikin testausyhdistelmistä parhaan tuloksen. Testauksissa näytti 30 runkoa/ha vähimmäisvaatimus riittävän, sillä jo runkoluvun rajaaminen suuremmaksi heikensi tulosta (vastinleimikkojoukko supistui). Aineiston rajaamisen merkitystä nimenomaan sivupuulajien ominaisuuksien ennustamisessa ei kuitenkaan pystytty analysoimaan tyhjentävästi. Käytännön sovelluksessa aineiston rajaaminen voitaisiin tehdä puulajiosuuksien ja mahdollisesti hakkuutavan mukaan, jolloin vastinleimikkojoukosta karsittaisiin pois jo ennen varsinaista vastinkohdehakua ne leimikot, jotka eivät vastaa puulajiosuuksiltaan ennustettavaa kohdetta.

Keskimääräinen ennustevirhe ei vielä kuvaa ennustemenetelmän luotettavuutta ja käytettävyyttä, vaan myös ennustevirheen hajonnalla on merkitystä. Erityisen tärkeää tämä on käytettäessä menetelmää yksittäisen leimikon puustonkuvaustarpeisiin. Leimikoiden ennustevirheen jakaumaa tarkasteltiin laskemalla niiden leimikoiden osuudet aineistosta, joissa kokonaistilavuuden ennustevirhe oli alle $\pm 10\%$, $\pm 20\%$ tai $\pm 30\%$ (taulukko 10). Parhaimmilla testausyhdistelmillä yli 50 %:lla leimikoista kuusen kokonaistilavuuden virhe oli alle 10 %. Alle 20 % ennustevirheluokkaan kuului 75–80 % kohteista. Männylläkin parhailla testausyhdistelmillä päästiin lähes samaan tasoon, mutta monilla yhdistelmillä hajonta oli suurempi.

Leimikon koon eli ennustettavan puulajin tilavuuskertymän vaikutusta ennustevirheen suuruuteen tarkasteltiin testausyhdistelmällä, joka antoi kuusen kokonaistilavuuden suhteen parhaan tuloksen (kuva 2). Suhteellinen ennustevirhe luonnollisesti pieneni ennustettavan puumäärän kasvaessa. Tarkastelusta jätettiin pois alle 30 m³ kuusta (kuusi sivupuulaji) sisältäneet leimikot, joilla ennustevirheprosentit saattoivat olla hyvinkin suuria. Yli 100 m³ kuusta sisältäneistä leimikoista 58 % oli ennustettu <10 %:n ennustevirheen tarkkuudella ja 87 %:lla ennustevirhe oli <20 %.

TAULUKKO 10 Kokonais- ja tukkitilavuuden ennustevirhetasojen 10, 20 ja 30 % alittavien leimikoiden osuus testiaineistossa k-MSN -menetelmän erilaisilla testausyhdistelmillä.

Testaus- yhdistelmä	Ennustevirhetasojen 10, 20 ja 30 % alittavien leimikoiden osuus testiaineistossa, %					
	- kokonaistilavuus - tukkitilavuus					
	Kuusi			Mänty		
	<±10%	<±20%	<±30%	<±10%	<±20%	<±30%
1	45,0	69,5	79,8	48,2	73,5	82,3
	38,5	65,1	77,2	42,4	67,7	78,0
2	46,9	71,6	82,3	45,3	72,7	85,3
	39,9	65,1	79,1	38,5	67,6	79,9
3	54,9	79,1	86,3	47,8	77,5	86,8
	45,7	70,7	82,9	40,1	68,1	79,1
4	44,3	68,5	78,8	22,0	42,3	55,4
	38,3	61,7	75,3	16,7	33,6	50,5
5	48,0	70,2	79,2	31,1	48,6	62,8
	37,9	60,6	75,4	18,6	36,4	53,1
6	41,5	63,6	72,6	35,1	55,4	66,2
	37,7	59,1	70,4	23,0	43,6	59,0
7	44,5	68,5	78,8	21,6	41,9	55,2
	39,8	64,9	76,7	20,3	39,3	53,9
8	48,0	70,2	79,2	29,6	47,8	62,6
	42,2	67,2	77,1	26,2	46,3	58,4
9	45,8	67,7	75,8	36,7	55,5	65,9
	38,5	60,1	72,0	33,4	52,0	63,8
10	43,9	71,5	81,2	23,3	44,5	59,5
	37,5	64,6	76,5	23,8	40,9	56,2
11	51,8	73,6	82,4	26,6	48,5	63,1
	43,9	69,4	81,2	25,6	47,1	60,7
12	47,0	72,7	81,7	24,0	43,5	60,5
	41,1	69,8	78,6	24,2	41,2	57,4
13	38,7	63,7	79,6	22,6	46,1	61,4
	36,6	60,8	76,7	22,1	42,8	56,9
14	42,8	64,8	77,7	21,9	41,9	60,2
	37,3	58,7	73,9	22,4	41,6	57,9
15	34,4	63,7	77,4			
	32,5	58,9	73,9			



Kuva 2. Leimikon koon eli ennustettavan puulajin kokonaistilavuuden vaikutus ennustevirheen suuruuteen.

Hakkuutapakohtainen tarkastelu osoitti, että harvennushakkuiden poistuman tilavuutta ei pystytty ennustamaan yhtä luotettavasti kuin päätehakkuuleimikoiden tilavuuksia. Kuusen parhaalla testausyhdistelmällä harvennushakkuiden keskimääräinen leimikon puulajin kokonaistilavuudella painotettu tilavuusvirhe oli 13,4 % (ennusteet yliarvioita). Hehtaariohtaiseksi lasketuna ennustevirhe oli keskimäärin 3,6 m³/ha (painottamaton) ja sen keskihajonta 12,4 m³/ha. Männyllä keskimääräinen ennustevirhe oli kuitenkin pienempi (0,3 % ja -1,7 m³/ha), mutta sen keskihajonta kuusta suurempi (19,3 m³/ha). Ennustevirheen vaihteluväli oli molemmilla puulajeilla melko suuri.

Harvennushakkuiden poistuman tilavuuden heikompi ennustettavuus johtuu todennäköisesti pienestä harvennusleimikoiden määrästä runkopankkiaineistossa. Niitä oli yhteensä vain 94, joista kuusivaltaisia 52 ja loput mäntyvaltaisia. Osassa leimikoita oli hakattu molempia puulajeja. Järeiden mukaan leimikot lisäksi vaihtelivat ensiharvennusleimikoista poistumaltaan hyvinkin järeisiin myöhempisiin harvennuksiin. Hakkuutapa ei valikoitunut k-MSN –menetelmän perushakumuuttujiin männyllä ja kuusella, vaan vastinleimikoiden valinta perustui puhtaasti puuston määrää ja kokoa kuvaaviin tunnuksiin. Ennustettavien harvennuskohteiden vastinleimikoista vajaa puolet oli hakkuutavaltaan harvennusta (kuusella keskimäärin 4,3 ja männyllä 4,2 vastinkohdetta kymmenestä) ja loput edustivat muita hakkuutapoja.

Harvennushakkuissa poistuman tilavuuden ennustamista vaikeuttaa myös harvennuksen voimakkuuden vaihtelu. Menetelmän ennustekyvyn parantamista voidaan kokeilla rajaamalla vastinleimikoiden ulkopuolelle muut kuin harvennusleimikot. Lisäksi tieto siitä, onko kohteella jo aiemmin tehty harvennusta tai onko puusto kaksijaksoinen (puustotunnukset näissä tapauksissa jaksoittain määritettyinä), saattaisi parantaa ennusteen tarkkuutta. Tietoja voitaisiin käyttää joko aineiston rajauksissa tai hakumuuttujana. Näiden merkityksen selvittäminen edellyttäisi kuitenkin lisäaineistojen hankintaa.

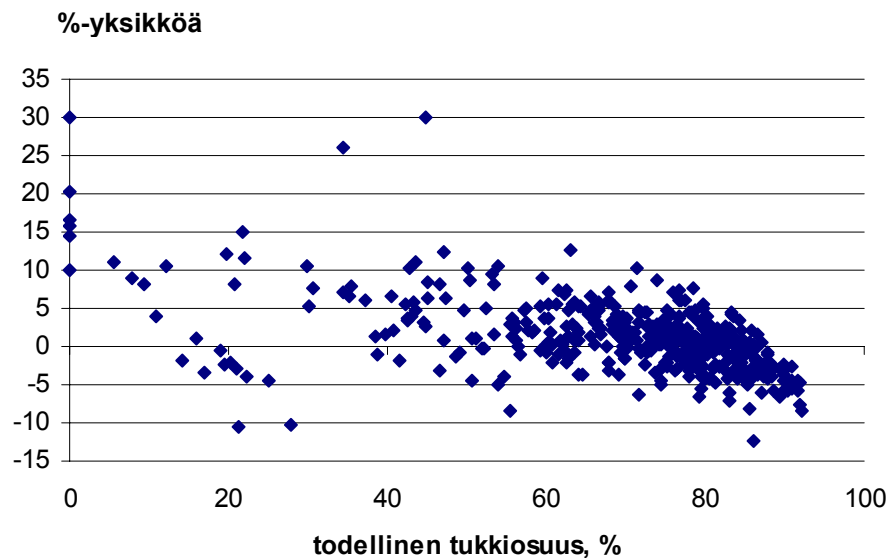
K-MSN –menetelmä näyttäisi pystyvän melko hyvin ennustamaan leimikoiden järeysrakenteen ja tukkiosuuden. Sitä tarkasteltiin kaikilla testausyhdistelmillä laskemalla yhdistelmän keskimääräiset leimikon puulajin kokonaistilavuudella painotetut ennustetut ja todelliset (katkonnan simuloinnin mukaiset) tukkiprosentit. Ne olivat lähellä toisiaan useimmilla testausyhdistelmillä ja enimmilläänkin niiden ero oli n. 1 %-yksikkö (taulukko 11).

Lähemmin tarkasteltiin sekä kuusella että männyllä niitä testausyhdistelmiä, jotka olivat antaneet kokonaistilavuuden ennustamisessa parhaan tuloksen (taulukko 12). Ennustetun ja todellisen tukkiprosentin ero oli n. 50 %:lla leimikoista enintään $\pm 2,5$ %-yksikköä ja 80–85 %:lla enintään ± 5 %-yksikköä. Tukkiprosentti näyttäisi olevan vaikeimmin ennustettavissa tukkiosuudeltaan pienillä leimikoilla, joissa todellinen tukkiosuus on alle 50 %. Leimikoissa, joissa tukkiosuus on suuri (yli 85 %), ennuste näyttäisi antavan systemaattisia aliarvioita (kuva 3).

Simuloidussa katkonnassa ei oteta huomioon todellisuudessa esiintyviä tukkiosuuteen vaikuttavia vikoja ja laatutekijöitä, esimerkiksi lahoa, lenkoutta ja mutkia. Siksi testauksissa saadut ennusteet antavatkin todennäköisesti liian hyviä tuloksia. Menetelmää sovellettaessa käytäntöön olisi simulaattoris- sa oltava toiminto, joka generoi runkoihin vikaisuuksia niiden käytännössä tai tutkimuksissa todettujen osuuksien mukaisesti ja tekee katkonnan runkopankkileimikoiden todellisia tukkiosuuksia vastaavaksi. Runkopankkiaineis- toissa olisi kuitenkin sitä varten oltava puuston laatua ja vikaisuuksia kuvaavaa tietoa. Tukkiosuuden katkontaan vaikuttavat myös käytettävät tukki- pituudet, joiden vaikutukset toteutuneeseen tukkiosuuteen pitää pystyä erot- tamaan laatutekijöistä.

TAULUKKO 11 Ennustetut ja todelliset (leimikoiden simuloitun katkonnän mukaiset) tukkiprosentit aineistossa eri testausyhdistelmillä.

Testausyhdistelmä	Leimikon kokonaistilavuudella painotetut tukkiprosentit keskimäärin eri testausyhdistelmillä			
	Kuusi		Mänty	
	Ennustettu	Todellinen	Ennustettu	Todellinen
1	76,7	76,9	82,1	82,5
2	76,8	76,9	81,7	82,0
3	76,8	77,0	77,3	78,4
4	77,2	77,4	82,7	83,2
5	77,2	77,4	82,7	83,2
6	77,2	77,3	82,7	82,7
7	76,7	76,9	82,4	83,1
8	76,7	77,0	82,4	83,1
9	76,9	77,1	83,3	83,8
10	76,9	77,1	83,3	83,8
11	76,9	77,2	83,3	83,3
12	76,9	77,1	83,3	83,7
13	76,9	76,8	83,3	83,7
14	76,9	77,0	83,3	83,6
15	76,9	76,7		



Kuva 3. Ennustetun ja todellisen tukkiprosentin ero suhteessa todelliseen tukkiprosenttiin. Kuusi, testausyhdistelmä 7.

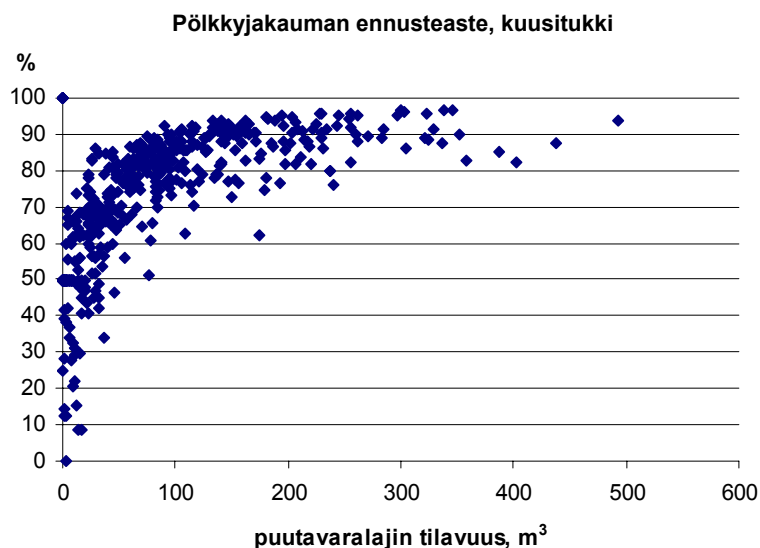
TAULUKKO 12 Ennustettujen ja todellisten tukkiprosenttien erojen tunnuslukuja parhaan kokonaistilavuuden ennusteen antamalla testausyhdistelmällä.

		Kuusi (testausyhdistelmä 7)	Mänty (testausyhdistelmä 1)
Ennustetun ja todellisen tukki-%:n ero (%-yksikköä)	Puulajin kokonaistilavuudella painotettu eron keskiarvo	0,3	0,7
	Painottamaton eron keskiarvo	1,1	1,0
	Keskihajonta	4,8	4,2
	Minimi	-12,3	-13,8
	Maksimi	30,0	28,0
Ennustetun ja todellisen tukki-%:n eron jakauma leimikoissa (% leimikoista)	< ±1 %-yksikköä	21,0	28,0
	< ± 3 %-yksikköä	58,0	69,5
	< ± 5 %-yksikköä	80,9	84,1

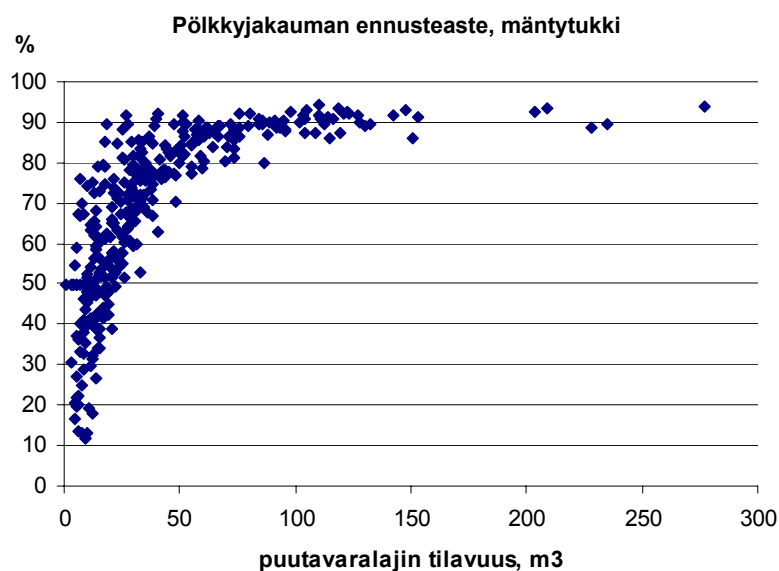
5.4.2 Pölkkyjakaumaennusteet

Puutavaralajien pölkkyjakaumien ennusteasteiden laskennassa verrattiin ennustetun pölkkyjakauman läpimitta-/pituusluokittaisia pölkkyjä kohteen todellisen (simuloimalla muodostetun) pölkkyjakauman vastaaviin. Testausyhdistelmän keskimääräinen ennusteaste koko aineistossa laskettiin painottamalla leimikoiden todellisilla ko. puutavaralajin tilavuuksilla. Tarkastelut tehtiin vain testausyhdistelmistä, jotka antoivat parhaan kokonaistilavuuden ennusteen (testausyhdistelmä 7 kuusella ja 1 männyllä). Kuusella mukana oli sorvitukki sekä männyllä tavaralajivaihtoehto, jossa mukana oli tyvitukki, jolla oli useita pituusluokkia ja toinen tukkipuutavaralaji (samoin useita pituusluokkia).

Kuusitukin keskimääräinen ennusteaste oli 84 %, sorvitukin 87 % ja kuitupuun 88 %. Männyllä ennusteasteet olivat keskimäärin kuusta huonommat: tyvitukilla 74 %, tukilla 81 % ja kuitupuulla 85 %. Alle 75 %:n ennusteasteita oli kuusitukilla 45 %:lla ja sorvitukilla 27 %:lla leimikoista. Mäntytyvitukilla ja mäntytyukilla osuudet olivat vastaavasti suuremmat. Puutavaralajin määrä leimikolla vaikutti ennusteasteeseen: mitä enemmän puuta, sitä korkeampi oli ennusteaste (kuvat 4 ja 5). Mäntytyukilla näkyi selvästi pienikertymäisten leimikoiden matala ennusteaste. Puutavaralajin tilavuusennusteen virheellä oli lievää korrelaatiota pölkkyjakauman ennusteasteen kanssa.



Kuva 4. Kuusitukkijakauman ennusteasteet leimikoilla tavaralajin tilavuuskertymän mukaan.



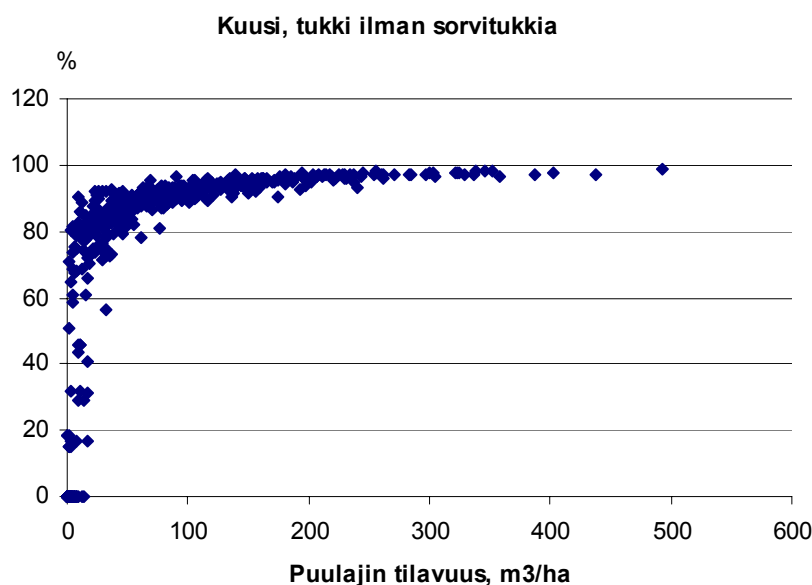
Kuva 5. Mäntytukkijakauman ennusteasteet leimikoilla tavaralajin tilavuuskertymän mukaan.

Jakauma-asteella kuvataan sitä, kuinka hyvin katkonnassa on onnistuttu toteuttamaan puutavaralajin tai sen kunkin läpimittaluokan tavoitteellisia pituusluokkaosuuksia. K-MSN –menetelmän yhteydessä sillä voidaan kuvata vastaavasti, kuinka hyvin ennustetut pölkkyjakaumat vastaavat apterauksen simuloinnissa käytettyjä tavoitejakaumia. Siihen vaikuttavat puutavaralajien määrä, niiden pituusluokkien määrä ja itse puustoennusteen osuvuus. Tunnuksen etuna on sen yksinkertaisuus ja havainnollisuus verrattaessa kahta erilaista pölkkyjakaumaa. Toisaalta vähäisellä pölkkyjen määrällä useaa pituus-läpimittaluokkaa kohden saattaa jakauma-aste antaa virheellisen kuvan, koska muutama pölkky ei voi edustaa useaan pituus-läpimittaluokkaan jaettua tukkijakaumaa (Malinen 2003a).

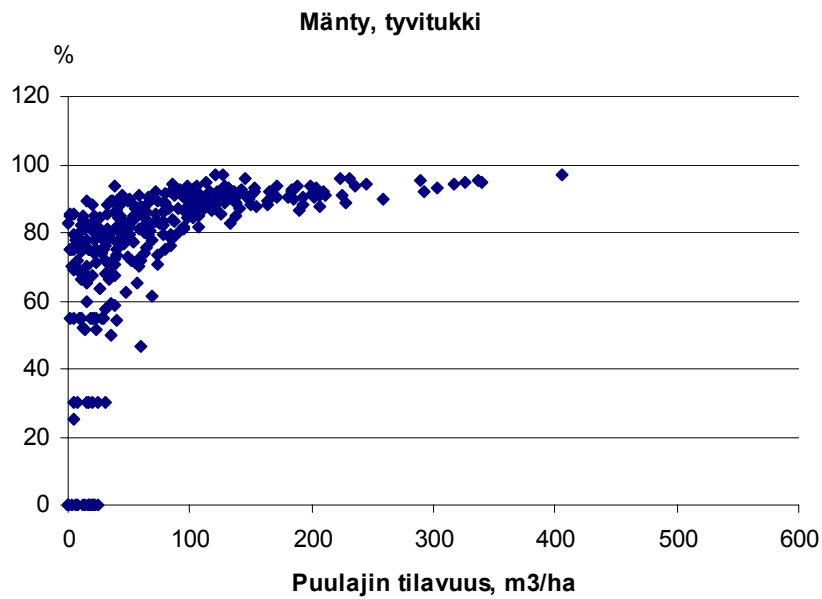
Parhailla testausyhdistelmillä tarkasteltiin ennustettujen pölkkyjakaumien jakauma-asteita eli puutavaralajien läpimittaluokittain laskettuja pituusluokkien suhteellisia pölkkyosuuksia verrattuina runkopankkiaineiston simuloitussa apterauksessa käytetyn apt-tiedoston läpimitta-/pituusluokkien jakaumatavoitteisiin. Leimikon puutavaralajin jakauma-astetta laskettaessa painotus tehtiin pölkkymäärien mukaan ja testausyhdistelmän keskimääräinen jakauma-aste koko aineistossa laskettiin painottamalla leimikoiden todellisilla ko. puutavaralajin tilavuuksilla.

Leimikon hehtaarikohtaisella puutavaralajin tilavuudella painotettu tukin jakauma-aste oli kuusella keskimäärin 93 %. Testausyhdistelmä, jossa katkonnassa ei ollut käytetty sorvitukkia, mutta muutoin testausparametrit olivat samat, antoi jakauma-asteeksi 90 %. Männyllä jakauma-asteet olivat hieman huonommat: tyvitukilla 87 % ja muulla tukilla 91 %. Pienillä puutavaralajin tilavuuksilla jakauma-aste saattoi olla huono. Kuitenkin esimerkiksi kuusella alle 75 %:n jakauma-asteita oli vain 22 %:lla leimikoista ja siitäkin puolet oli tapauksia, joissa kohteen ko. tukkipuutavaralajin todellinen tilavuus oli alle 10 m³/ha. Jakauma-asteet nousivat varsin nopeasti tavaralajin tilavuuden kasvaessa ja esim. yli 100 m³/ha tilavuuksilla ne olivat pääsääntöisesti 85–100 %:n tasoa (kuvat 6, 7a ja 7b).

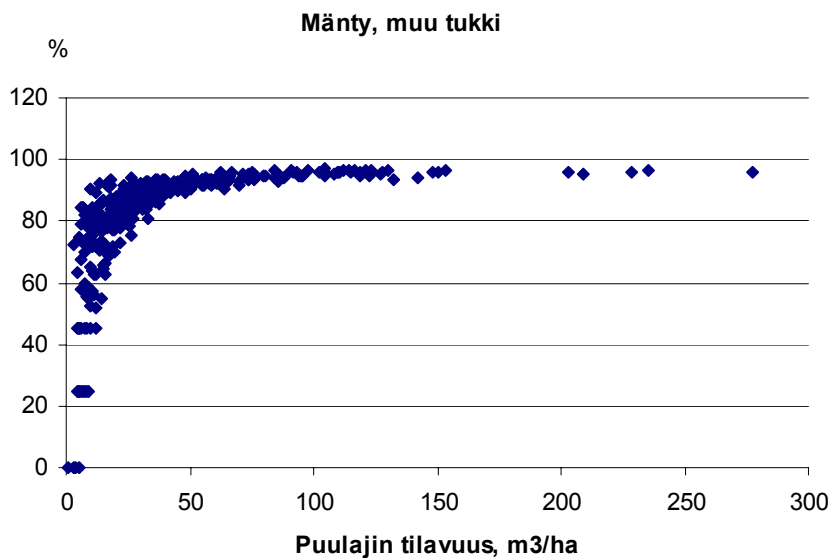
Ennustettujen pölkkyjakaumien jakauma-asteiden ero todelliseen (simuloinnin mukaiseen) oli keskimäärin -3,6 %-yksikköä kuusella, -8,8 %-yksikköä männyn tyvitukilla ja -2,8 %-yksikköä muulla tukilla. Muutamissa tapauksissa ennusteen jakauma-aste oli hieman korkeampi kuin leimikon todellinen jakauma-aste, mutta pääsääntöisesti aina hieman heikompi.



Kuva 6. Kuusitukin jakauma-asteet leimikoilla tavaralajin hehtaarikohtaisen tilavuuden suhteen parhaan kokonaistilavuuden ennusteen mukaisella testausyhdistelmällä. Katkonnassa tavaralajivaihtoehto, jossa sorvitukki on mukana.



Kuva 7a. Männyn tyvitukin jakauma-asteet leimikoilla tavaralajin hehtaarikohtaisen tilavuuden suhteen parhaan kokonaistilavuuden ennusteen mukaisella testausyhdistelmällä. Katkonassa tavaralajivaihtoehto, jossa tyvitukin lisäksi muu tukki on mukana.

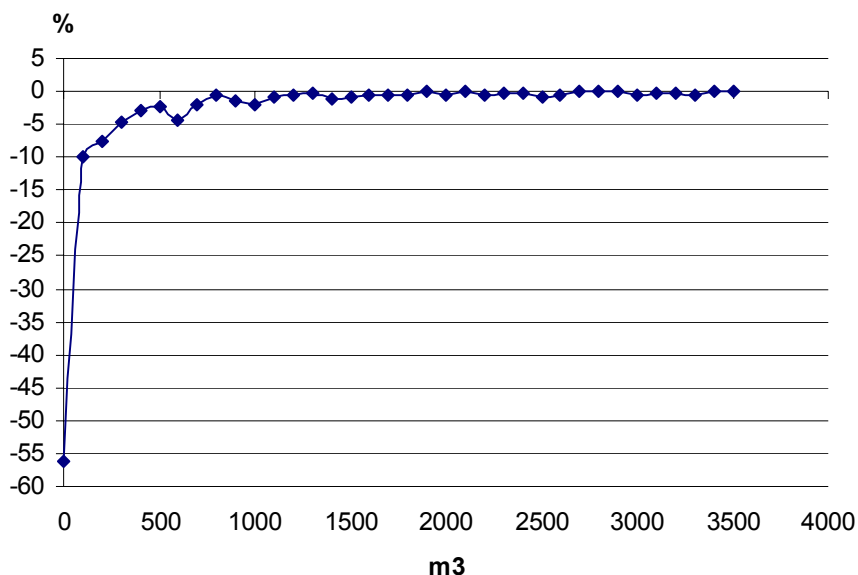


Kuva 7b. Männyn muun tukin jakauma-asteet.

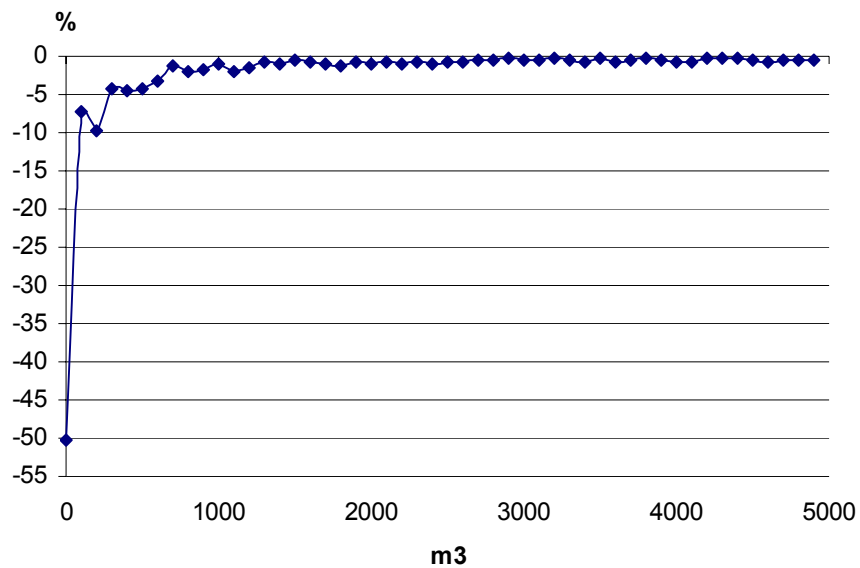
5.4.3 Kokonais- ja tukkitilavuuksien ennustevirheet sumatasolla

K-MSN –menetelmän käytettävyyttä puunhankinnan aluetasoisessa suunnittelussa ja korjuuseen tulevien varantojen puumäärän ennustamisessa kokeiltiin simuloimalla leimikoiden kuusen kokonais- ja tukkitilavuusennusteita ja laskemalla, kuinka ennustevirhe kehittyi puumäärän kumuloituessa (leimikoiden määrän mukaan). Simuloinnit tehtiin 100 kertaa sijoittaen leimikot satunnaiseen järjestykseen ja kumuloimalla todellista ja ennustettua tilavuutta. Simulointituloksista laskettiin keskiarvot 100 m³:n tilavuusluokkavälein. Simuloinnit tehtiin vain testausyhdistelmän 7 tilavuusennusteille.

Ennustevirheen keskimääräinen kehittyminen puumäärän suhteen on esitetty kuvissa 8 ja 9. Ennustevirheet näyttäisivät tasaantuvan melko nopeasti tilavuuden kasvaessa ja olevan alle 4 %:n tasolla jo n. 1 000 m³:n puumäärällä. Käytännön tilanteissa yksittäisten leimikoiden ennustevirheiden voidaan olettaa kumoavan toisensa siten, että varantojen puumääriä ennustettaessa jo melko pienillä puumäärillä ennustevirhe lähestyy nollaa.



Kuva 8. Kuusen kokonaistilavuuden ennustevirheen kehittyminen sumatason puumäärän ennustamisessa leimikoiden tilavuuden kumuloituessa. Ennustevirhe laskettu 100 simuloinnin keskiarvona.



Kuva 9. Kuusen tukkitilavuuden ennustevirheen kehittyminen simulointien keskiarvona.

5.4.4 Hakumuuttujien virheen vaikutus

Hakumuuttujien satunnaisten virheiden vaikutusta ennusteiden tarkkuuteen testattiin kuusen kahdella testausyhdistelmällä (ks. 5.3.3). Kokonais- ja tukkitilavuuden ennustevirheiden sekä pölkkyjakauman ennusteasteiden muutoksia tarkasteltiin tässä testauksessa. Vertailut tehtiin k-MSN –menetelmän perustestauksiin, joissa hakumuuttujatietona oli käytetty leimikoiden todellisia arvoja (runkopankkidatasta lasketut puustotunnukset). Koska virhettä ei asetettu laskennassa systemaattiseksi, ei ennusteiden keskimääräinen tarkkuus heikentynyt vielä merkittävästi virhetason kasvaessa. Sen sijaan leimikkotasolla hajonta kasvoi, mikä näkyy esimerkiksi kokonaistilavuuden alle 10 %:n ennustevirheen alittavien leimikoiden osuuden selvänä pientymisenä, kun lähtötietojen annettiin vaihdella ± 25 %:n välillä: osuus putosi 45 %:sta 26 %:iin. Pölkkyjakauman ennusteasteen muutos oli varsin lievä. Testausyhdistelmällä 7 keskimääräinen ennusteaste (93 %) ei muuttunut lainkaan virhetason kasvaessa 25 %:iin. Toisella testausyhdistelmällä se putosi 97 %:sta 96 %:iin. Joensuun yliopiston hankkeessa tehtiin samanlaisia herkkyyksanalyyskejä ja tulos oli samanlainen: pölkkyjakauman keskimääräinen ennusteaste ei heikentynyt lähtötiedon virheen kasvaessa. Oletettavasti kuitenkin myös tällä tunnuksella leimikoittainen hajonta kasvaa, mutta tätä ei erikseen tarkasteltu.

TAULUKKO 13 Hakumuuttujien virheiden vaikutukset ennusteiden tarkkuuteen (kuusi, viiden simuloinnin keskiarvot).

Testaus-yhdistelmä	Haku-muuttuja-tiedon virhetaso	Kokonais-tilavuuden ennuste- virhe, %	Tukki-tilavuuden ennuste- virhe, %	Kokonaistilavuuden en- nustevirheen 10, 20 ja 30 % alittavien leimikoiden osuus, %		
				< ±10 %	< ±20 %	< ±30 %
7	0 %	0,00	0,26	45	69	79
	±10 %	-0,04	-0,34	42	68	81
	±25 %	0,50	-1,39	26	44	61
12	0 %	1,05	1,30	47	73	82
	±10 %	1,00	0,82	45	70	84
	±25 %	2,50	0,49	24	45	61

6 PÄÄTELMÄT

6.1 Tärkeimmät hakumuuttujat

Tutkimuksen keskeisenä tavoitteena oli selvittää k-MSN –menetelmän tärkeimmät hakumuuttujat. Hakumuuttujajoukkoon valittiin sellaisia muuttujia, jotka voidaan helposti ja kustannustehokkaasti määrittää käytännössä. Toisaalta pyrittiin varmistamaan muuttujien monipuolisuus, koska aiemmissa tutkimuksissa (Malinen ym. 2001 ja Malinen 2003b) on todettu keskipuun tunnusten määrittävän puuston kokojakauman sijainnin ja muiden tunnusten jakauman muodon ja hajonnan. Hakumuuttujien valinnassa käytettiin kahta menetelmää, subjektiivista sekä automaattista valintamenetelmää. Subjektii-visen hakumuuttujien valinnan mukaan taulukoiden 4 ja 5 muuttujat ovat tärkeitä käytettäessä pölkkyjakauman ennustamiseen k-MSN –menetelmää. Lisäksi tutkimuksessa kaikista jatkuvista muuttujista muodostetut muunnokset osoittautuivat merkitseviksi, joten esitettyjen muuttujien muunnokset luetaan tärkeiden muuttujien joukkoon.

Käytetyistä hakumuuttujista kaikki, pois lukien pohjapinta-alamediaanipuun tilavuus (V_{gm}) löytyvät Solmu-metsäsuunnittelujärjestelmästä. Käytännön sovelluksen kannalta V_{gm} :n puuttuminen ei ole este, koska se on varsin vai-vatonta ennustaa yleisesti tunnetuilla malleilla. Hakumuuttujien perustuessa malleilla saatuihin ennusteisiin voidaan kuitenkin k-MSN ja LAN MSN -ennusteiden virheiden olettaa lisääntyvän, koska malleihin sisältyy aina vir-hettä. Tässä tutkimuksessa käytetyistä hakumuuttujista vain V_{gm} ja puula-jiosuus (puulajin käyttöosan tilavuuden prosenttiosuus, pl_{pros}) ovat sellaisia muuttujia, jotka voivat sisältää mallin aiheuttamaa virhettä, jos ne poimitaan kuviotietojärjestelmästä. On syytä huomioida, että tämän tutkimuksen tulosten valossa puulajiosuus on varsin merkittävä muuttuja, mutta vastaavasti V_{gm} esiintyy hakumuuttujien subjektiivisen valinnan tapauksessa ainoastaan koivun ennusteissa.

Tutkimuksessa käytetyn aineiston hakukuohdetta kuvailevat muuttujat ovat ihmistyönä määritettyjä. Näistä etenkin pinta-alan on todettu sisältävän

usein virhettä. Leimikon pinta-alan summittainenkin määrittäminen voi olla vaikeaa. Leimikko voidaan rajata toisin kuin metsäsuunnittelutiedoissa kuvioiden rajat ovat, jolloin pinta-ala arvioidaan tai tallennetaan kuviotietojen mukaisena. Koska pinta-ala runkopankin muuttujana sisältää virhettä, voi sillä olla vaikutusta saatuihin ennusteisiin. Tämän lisäksi pinta-alan virhe heijastuu välittömästi muihin hehtaariohtaisiin muuttujiin, jolloin useat runkopankin muuttujat sisältävät virhettä. Pinta-alan merkitys yhtenä valikoituneista hakumuuttujista on sinänsä vielä jonkin verran epäselvää, sillä kovin hyvää loogista selitystä sille ei ole. Sen poisjättäminen kuitenkin heikensi tuloksia. Pinta-alaa voidaan käyttää k-MSN –laskennassa myös skaalaustekijänä, jolloin pinta-alatiedon tarkkuus vaikuttaa myös sitä kautta saataviin estimaatteihin.

Esitettyä automaattista hakumuuttujien valintamenetelmää käytettäessä tärkeitä hakumuuttujia löytyi huomattavasti enemmän. Tästä suuresta valittujen hakumuuttujien joukosta ei ole mielekästä esittää joitakin tärkeitä muuttujia, koska kehitetty automaattinen hakumuuttujien valintamenetelmä käytti kaikilla puulajeilla lähes kaikkia käytössä olleista muuttujista. Huomionarvoista kuitenkin on se, että automaattinen valinta painotti hakumuuttujien valinnassa hakkuukohdetta kuvailevia muuttujia.

Testauksissa käytettyihin puulajeittaisiin perushakumuuttujayhdistelmiin päädyttiin siis hakumuuttujien valintamenetelmiä koskevien tulosten perusteella. Analyysissä hakumuuttujien subjektiivisen muuttujavalinnan kriteerinä käytettiin pelkästään pölkkyjakauman ennusteastetta, jota pyrittiin maksimoimaan. Se on hyvä kriteeri ennustettavan puuston rakennetta kuvattaessa, mutta hakumuuttujien valinnassa olisi käytettävä kriteerinä myös tärkeimpien puuston määrää kuvaavien tunnusten ennustevirheiden minimoimista. Hakkuutapa ei ollut kuusen ja männyn perushakumuuttujien joukossa. Sen merkitys voikin olla vähäinen puuston koko- ja määrämuuttujiin verrattuna. Aineiston niukkuuden vuoksi ei testattu sen merkitystä vastinkohdejoukkoa rajaavana muuttujana. Olisikin mielenkiintoista jatkossa selvittää, voidaanko esimerkiksi harvennushakkuukohteiden ennusteita parantaa rajaamalla hakuvaruutta vain harvennushakkuuleimikoihin.

6.2 Hakumuuttujien valintamenetelmät käytännön sovel-luksissa

Tutkimusta aloitettaessa päätettiin yhtenä hakumuuttujien valintatyökaluna käyttää askeltavaa regressioanalyysiä, jolla pystytään poimimaan laajasta muuttujajoukosta hyviä selittäjiä. Tulosten perusteella voidaan kuitenkin sanoa, ettei sen merkitys hakumuuttujien valintatyökaluna ole suuri. Hakumuuttujien valinta subjektiivisesti, käyttäen askeltavaa regressioanalyysiä, on työlästä ja käyttäjältä ammattitaitoa vaativaa, ja siksi kokeiltiin myös hakumuuttujien automaattista valintaa.

Hakumuuttujien automaattisessa valinnassa lähtökohtana on, että muuttujia ei valita subjektiivisesti, vaan ennustamismenetelmät hakevat parhaan mahdollisen tuloksen antavat muuttujat automaattisesti ennusteen laskennan

yhteydessä. Hakumuuttujien valinnan automatisoinnilla pyritään vähentämään runkopankin käyttäjän valintatilanteita. Tämä näyttäisikin onnistuvan ennusteiden hyvyyden kärsimättä. Lisäksi runkopankin käyttäjän valintatilanteiden karsimisella ja suuremmalla automatisoinnilla pystytään vähentämään inhimillisten tekijöiden vaikutusta ennusteisiin, ja siten lisäämään ennusteiden tasalaatuisuutta.

Hakumuuttujien subjektiivisella valinnalla saatiin kuitenkin yhtä hyviä ja osin parempia tuloksia kuin automaattisella valinnalla hakumuuttujien määrän pysyessä huomattavasti pienempänä. Hakumuuttujien suuri määrä epäselkeyttä k-MSN ja LAN MSN –menetelmien etäisyyksien välisiä eroja sekä lisää laskentakapasiteetin tarvetta. Ulotteisuuden kirouksen (*curse of dimensionality*) välttämisen eli hakumuuttujien pienemmän määrän sekä laskentakapasiteetin pienemmän tarpeen vuoksi subjektiivisessa valinnassa esitetty hakumuuttujien optimointialgoritmi ilman regressioanalyysiä saattaisi olla tarpeellista automatisoida. Hakumuuttujien valinnan automatisoinnin heikkoudeksi voidaan lukea suurehko muuttujien tarve jo lähtötilanteessa. Ei siis ole itsestään selvää, että automaattista hakumuuttujien valintaa käytettäessä voitaisiin mitata kohdekuviolta vain muutama ennalta määritellyt keskeinen muuttuja. Esitetyn hakumuuttujien automaattisen valinnan avulla päästään sitä parempiin ennusteisiin mitä suurempi joukko hakumuuttujia alussa on, mutta samalla lisääntyvät laskentakapasiteetin tarve sekä mahdollisesti ulotteisuuden kirous.

Hakumuuttujien määrittäminen ja valinta ovat keskeinen osa koko k-MSN –menetelmän käyttöä. Tässä tutkimuksessa ei saatu selvää näyttöä siitä, että muuttujien automaattinen valinta olisi antanut parempia ennustetuloksia kuin muuttujien subjektiivinen valinta. Hakumuuttujien valinnassa on otettava huomioon normaaleissa menetelmän käyttötilanteissa kustannustehokkaasti käytettävissä olevat tunnuksat. Tärkeimmät puuston määrää, järeyttä, laatua ja puulajisuhteita kuvaavat muuttujat olisi voitava valita. Menetelmän testaukset on tähän saakka tehty varsin pienillä runkopankin leimikkoaineistoilla ja voi olla, että kaikkiin erilaisiin leimikkotyyppeihin sopivia optimaalimpia hakumuuttujajyhdistelmiä ei ole vielä pystytty osoittamaan. Mahdollisissa suuremmilla aineistoilla tehtävissä jatkotestauksissa hieman eri muuttujat tai niiden muunnokset voivat osoittautua merkitsevämmiksi kuin tässä tutkimuksessa löydettyt.

6.3 K-MSN –menetelmän ennustekyky tulosten perusteella

Tutkimuksessa testattiin k-MSN –menetelmää käyttäen perushakumuuttujiksi valittuja muuttujia ja kokeillen erilaisia laskentaparametri- ja aineistorajoitusvaihtoehtoja. Metsätehon testaukset tehtiin vain k-MSN –menetelmällä, sillä Joensuun yliopiston hankkeessa oli todettu, että LAN MSN –menetelmä ei poikennut tulosten osalta merkittävästi k-MSN:stä. Joissakin tapauksissa sillä saatetaan saada kuvattua ominaisuuksia, joita k-MSN:llä ei pystytä. LAN MSN:n suurimpana heikkoutena tutkimuksen perusteella voidaan kuitenkin pitää suurempaa laskentakapasiteetin tarvetta.

K-MSN –menetelmän tarkkuutta, luotettavuutta ja käytettävyyttä leimikon puustotunnusten ja pölkkyjakauman ennustamisessa tarkasteltiin eri estimaattien keskimääräisten tunnuslukujen sekä niiden hajonnan ja vaihtelun valossa. Menetelmä ja sen algoritmi perustuvat toistaiseksi puulajikohtaisesti tehtävään laskentaan. Siksi myös tarkastelut tehtiin puulajikohtaisina.

Puuston tilavuuden ennustaminen mahdollisimman luotettavasti on peruslähtökohta menetelmän soveltamisen kannalta. Tilanteissa, joissa metsikkökuviolta ovat tiedossa puulajeittaiset ajantasaiset pohjapinta-ala- ja keskipituustiedot, voidaan kokonaistilavuudet laskea jo niiden perusteella riittävän tarkasti. K-MSN –menetelmän käyttö ei tuo ennusteisiin mitään lisäarvoa. Kokonaistilavuuden estimaateilla sinänsä on kuitenkin melko vähän käyttöä puunhankinnan suunnittelussa, ainakaan leimikkotasolla, vaan tarvitaan tukki- ja kuitupuu- tai läpimittaositteiden tilavuusestimaatteja. Usein puuston mittaustietoihin sisältyy virhettä tai niitä ei ole kasvumalleilla päivitetty. Tällöin tilavuusestimaatit voivat olla useiden vastinleimikoiden mukaan keskiarvoistuneita ja siten ainakin suuremmassa laskentapopulaatiossa keskimäärin harhattomampia. Tämä ei välttämättä pidä enää paikkaansa, jos runkopankissa on runsaasti samankaltaisia leimikoita, jotka valikoituvat vastinleimikoiksi hakumuuttujien arvojen mukaan. Tätä oletusta ei ole kuitenkaan pystytty nykyaineistoilla testaamaan.

Testausten perusteella kokonais- ja tukkitilavuusestimaatit pystytään k-MSN –menetelmällä muodostamaan harhattomasti: parhaalla testausyhdistelmällä tilavuusennusteen keskimääräinen virhe oli kuusella 0,0 % ja männyllä 0,6 %. Kuusella useimmilla muillakin testausyhdistelmillä tilavuusvirhetaso oli samaa luokkaa, mutta männyllä erot olivat monilla yhdistelmillä suurempia. Leimikoista hieman alle 50 %:ssa tilavuusestimaatin virhe oli alle ± 10 %. Tulosta voi pitää kohtuullisen hyvänä nykykäytäntöjen tilavuuden arviointeihin verrattuna. On kuitenkin eri asia millaista tilavuusvirhetasoa pidetään riittävänä ja tavoiteltavana (esim. 5 %). Tukkitilavuuden ennustamisvirhe oli keskimäärin hieman suurempi (kuusella 0,3 % ja männyllä 1,1 %) ja n. 40 % leimikoista alitti ± 10 %:n tukkitilavuuden virheprosentin tason. Ennustettavan puulajin tilavuuden määrä eli leimikon koko vaikutti virheprosentteihin. Hajonta oli suhteellisesti suurinta pienillä tilavuuksilla.

Hakkuutavoittain tarkasteltuna tilavuusestimaattien virhetasossa ja hajonnassa oli kuitenkin eroa. Harvennusleimikoiden tilavuuksia ei pystytty estimoimaan yhtä luotettavasti kuin päätehakkuuleimikoiden. Suurimpana syytä tähän voidaan pitää harvennusleimikoiden pientä määrää aineistossa. Jatkotestauksissa vastinleimikoiden rajausta hakkuutavan mukaan ja uusia harvennusleimikoiden puuston rakennetta kuvaavia muuttujia olisi syytä kokeilla. Nyt harvennuskohteiden vastinleimikoista yli puolet oli hakkuutavaltaan jotain muuta.

Puuston rakenteen ennustamiskykyä tarkasteltiin tukkiprosenttiestimaattien sekä pölkkyjakaumien ennusteasteiden ja jakauma-asteiden avulla. Tukkiisuuden estimointi onnistuu menetelmällä muutaman prosenttiyksikön tarkkuudella lukuun ottamatta leimikoita, joissa tukkiisuus on pieni (harvennusleimikot); niissä tukkiisuuden ennustamisvirheen hajonta oli suurempi. Leimikoissa, joissa tukkiisuus on korkea, näyttäisi estimaateissa olevan lievää systemaattista aliarviota. Tarkastelut eivät kuitenkaan vastaa aivan todellisia ennustetilanteita, sillä runkopankin pölkkytystä simuloitaessa ei ole voitu ottaa huomioon satunnaisia runkojen laatu- ja vikaisuustekijöitä, jotka pienentävät tukkiisuutta.

Pölkkyjakaumien ennusteasteet kuvaavat puutavaralajien ennustettujen pölkkyjakaumien samankaltaisuutta leimikon todelliseen pölkkyjakaumaan. Ennusteasteen laskennassa verrataan läpimitta- ja pituusluokkayhdistelmien pölkkymääriä ennustetun ja todellisen pölkkyjakauman välillä. Ennusteaste kuvaa siten puuston määrän lisäksi järeyden ja runkomuodon ennustamiskykyä. Testausten mukaan ennusteasteet olivat kohtuullisen hyvät ainakin kuusella (keskimäärin 84–88 % puutavaralajista riippuen). Männyllä ennusteasteet olivat vähän heikommät (74–85 %). Samat syyt vaikuttivat oletettavasti siihenkin kuin muiden tunnusten estimointien hyvytyteen. Pääsyyinä lienee aineistojen niukkuus ja siitä johtuva epäoptimaalinen vastinleimikoiden valinta.

Jakauma-asteen avulla voidaan kuvata, kuinka hyvin pölkkyjakaumaennusteet vastaavat apterauksen simuloinnissa käytettyjä tavoitejakaumia. On huomattava, että runkopankin katkonnan simuloinneissa saatavat jakaumaasteet ovat yleensä hyvin korkeita (aineistossa kuusella keskimäärin 96,5 % sekä männyllä 95,8 % tyvitukilla ja 93,7 % tukilla), eivätkä siten ole vertailtavissa käytännön katkonnan jakauma-asteisiin. Niihin vaikuttavat hakkuukoneen mittaustarkkuus ja runkomuodon ennustamiskyky, runkojen laatuviat sekä kuljettajan katkontatekniikka. Simuloinnin jakauma-aste on lähinnä leimikon puuston ja sen runkomuodon mukainen teoreettinen maksimi, joka voi toteutua vain, mikäli laatu- ja tekijöiden vuoksi ei poiketa katkontaehdotuksista ja mittaustarkkuuteen vaikuttavat tekijät ovat kunnossa.

Testauksissa saadut ennustettujen pölkkyjakaumien jakaumaasteet olivat myös varsin korkeita, mutta luonnollisesti kuitenkin pienempiä kuin katkonnan simuloinneissa lasketut ns. todelliset jakaumaasteet. Kuusella ennusteen jakauma-aste oli keskimäärin 93 % ja -3,6 %-yksikköä huonompi kuin todellinen jakauma-aste. Männyllä jakaumaasteet olivat pienemmät: tyvitukilla 87 % ja muulla tukilla 91 %. Pölkkyjakaumaennusteiden jakaumaasteet ovat sitä korkeampia mitä suurempi puumäärältään ja järeämpi ennustettava leimikko on. Tällöin on suuremmat mahdollisuudet saada läpimitta-/pituusluokkayhdistelmiin tavoitejakaumien mukaiset suhteelliset pölkkyosuudet, mikäli vain vastinleimikoiden valinta on puuston määrän ja järeyden suhteen osunut kohdalleen. Lisäksi jakauma-asteeseen vaikuttavat etenkin katkonnan simulointiin asetetut pituusluokkien määrät ja niiden jakaumatavoitearvot. Jakauma-astetta onkin paras tarkastella lähinnä apukriteerinä muiden perusteiden lisäksi, kun k-MSN –menetelmän ennustekykä arvioidaan.

K-MSN –menetelmän ennustekyvyn luotettavuutta ja menetelmän käytettävyyttä voidaan arvioida myös ennusteiden käyttötarpeen näkökulmasta. Yksittäisen metsikön puustotunnusten tarkkuus- ja luotettavuusvaatimukset poikkeavat laajemman leimikkjoukon (suma- tai varantotaso) vaatimuksista. Tehtäessä ennuste esimerkiksi leimikon hinnoittelua varten olisi ennusteen syytä olla oletetun vaihteluvälin rajoissa. Puunhankinnan suunnittelussa toimitaan kuitenkin suuremmilla leimikko- ja puumäärillä, jolloin ennusteita olisi voitava tehdä luotettavasti leimikkjoukolle, joka määräytyy toimialueen, suunnittelujakson, korjuukelpoisuuden ja mahdollisesti hakkuutavan ja puuston perustunnusten mukaan. Leimikkjoukolle voidaan kokeilla erilaisia katkontavaihtoehtoja ja vertailla niitä keskenään.

Testauksissa tarkasteltiin suppeasti kokonais- ja tukkitilavuuksien ennustevirheiden kehitystä sumatasolla tehdyssä leimikkjoukon simuloinnissa. Osoittautui, että keskimääräinen tilavuusvirhe supistui alle 4 %:n tason jo hyvin pienillä, n. 1 000 m³:n puumäärillä. Samantapaiset analyysit olisi hyvä tehdä myös muista estimaateista. Puunhankinnan suunnittelua palveleviin johtopäätöksiin päästään kuitenkin vain, jos tarkastelut tehtäisiin käytännössä saatavien hakumuuttujatietojen mukaan ja katkonnan simuloinneissa voitaisiin ottaa puuston laatutekijät huomioon.

K-MSN –menetelmällä saatavien puusto- ja pölkkyjakaumaestimaattien tarkkuuteen ja luotettavuuteen vaikuttavat suuresti hakumuuttujina käytettyjen tietojen virheet. Puustoa kuvaavat keskitunnukset ja mm. pinta-ala ovat aina joko mittauksin tai arvioinnein saatuja ja osa niistä voi olla malleilla johdettuja. Siksi lähtötietoihin sisältyy aina epävarmuutta ja hajontaa, joiden suuruudesta on oltava hyvä käsitys erilaisia sovellustarpeita analysoidessa.

Peruspuustotunnusten arviointivirheiden suhteellista keskivirhettä on tutkittu melko paljon lähinnä kuvioittaisessa arvioinnissa. Pohjapinta-alan suhteellinen keskivirhe on ollut niissä 13–32 %, pohjapinta-alan mediaanipuun keskiläpimitan 11–20 % ja keskipituuden 8–7 % (Heikkinen 2002). Muidenkin puustotunnusten arviointivirheitä on selvitetty mm. Heikkisen tutkimuksessa, jonka mukaan esimerkiksi puusto-ositteen maksimiläpimitan arvioinnin suhteellinen keskivirhe oli 18 %. Joidenkin tunnusten arviointivirhe oli siinä hyvinkin suuri, mm. minimiläpimitan 72 % ja runkoluvun 81 %. Tutkimuksessa selvitettiin myös mittajien ja mitattavien puustojen merkitystä arviointivirheisiin eri puustotunnuksilla. Mittaajasta johtuvan satunnaisvaihtelun todettiin olevan merkitsevää mm. pohjapinta-alan, keskiläpimitan ja minimiläpimitan mittauksissa. Keskiläpimitan mittauksessa todettiin olevan systemaattista aliarviota. Metsätehon erillisessä selvityksessä päädyttiin kutakuinkin samantasoiisiin keskimääräisiin arviointivirheisiin. Puustotunnusten silmävaraisella arvioinnilla ja kevyellä ennakkomittauksella ei ollut tarkkuuden suhteen suurta eroa, mutta mittauksissa hajonta oli arviointia pienempää (Räsänen ja Imponen 2004).

Hakumuuttujien harhan ja hajonnan vaikutusta estimaatteihin voidaan testata simuloimalla hakumuuttujien lähtötietoja virheellisiksi esimerkiksi oletetun suhteellisen keskivirheen puitteissa. Aiemmin on todettu, että k-MSN- ja LAN-MSN –menetelmät sietävät molemmat hyvin hakumuuttujien virhettä,

kun tarkasteltavana tunnuksena on ollut jakauma-aste (Kuuramaa 2004). Systemaattisesti harhaisten lähtöarvojen voidaan olettaa tuottavan myös harhaisia estimaatteja. Jos taas lähtöarvot eivät ole systemaattisesti harhaisia, vaan satunnaisesti tietyn vaihteluvälin rajoissa olevia, ei ennusteidenkaan voida olettaa olevan systemaattisesti virheellisiä. Vastinleimikkoaineiston määrällä ja edustavuudella on tällöinkin merkitystä. Tutkimuksessa kokeiltiin tällaista satunnaisesti eri maksimivirhetasoilla vaihtelevien hakumuuttaja-arvojen vaikutusta tilavuustunnusten keskimääräiseen ennustevirheeseen parhailla testausyhdistelmillä. Tarkasteluissa 10 %:n virhetaso ei vielä vaikuttanut merkittävästi ennustevirheisiin, mutta suuremmalla 25 %:n virhetasolla oli jo selvä estimaattien keskimääräistä virhettä ja leimikkotason hajontaa lisäävä vaikutus. Käytännön sovelluksissa ja käytettävissä olevilla lähtötiedoilla (metsäsuunnitelmat ja leimikoiden puusto-arviot) eri perushakumuuttajien virheet ovat näillä tasoilla, harvemmin ainakaan systemaattisesti sitä suurempia.

Lopuksi yhteenvetona voidaan todeta seuraavien tekijöiden vaikuttavan arvioiden mukaan eniten k-MSN –menetelmällä saatavien puustotunnusten ja pölkkyjakaumien estimaattien tarkkuuteen ja luotettavuuteen:

1. Runkopankin määrä ja edustavuus (hakkuutavoiltaan, puulajisuhteiltaan, puumäärältään ja järeydeltään erilaisia leimikoita).
2. Puuston määrää ja rakennetta sekä leimikkoa muutoin kuvaavat hakumuuttajat ja niiden valinta (valinnan optimointi)
3. Hakumuuttajien sisältämä systemaattinen tai satunnainen virhe
4. Vastinleimikoiden määrä eli k:n arvo menetelmässä
5. Skaalaustekijänä käytettävän muuttujan valinta
6. Runkopankin pölkkytyksen simuloinneissa käytetyt puutavaralajit ja niiden läpimitta- ja pituusluokkien määrä
7. Runkojen laatutekijöiden mukainen apterauksen simulointi (mikäli riittävän luotettavat perusteet löydettyissä)

7 MENETELMÄN SOVELLUKSET

Runkopankkia ja sitä hyödyntävää k-MSN –menetelmää ryhdyttiin kehittämään ensisijaisesti leimikoiden puustoestimaattien ennustamiseksi niin, että voidaan myös muodostaa halutuilla puutavaralaji- ja katkontavaihtoehdoilla toteutuvat pölkkyjakaumat. Mahdollisia menetelmän käyttöalueita voivat olla seuraavat:

1. Leimikon tai useiden leimikoiden muodostaman varannon puustotietojen muodostaminen puunhankinnan operatiiviseen suunnitteluun
 - korjuuohjelman optimointi ja varastotason määrittäminen
2. Puutavaralaji- ja katkontavaihtoehtojen vertailut
 - vaikutukset puutavaralajijakaumiin varantotasolla
3. Puustotiedot puukauppaa varten
 - puutavaralajikertymät eri katkontavaihtoehdoilla
 - leimikon hinnoitteluvaihtoehdot ja uudet hinnoittelukäytännöt
4. Omien metsien hakkuiden suunnittelu
 - tarkennetut puustotiedot ja simuloitu katkontatulos kuviotietojen puuston keskitunnuksista

Edellä kuvattuja käyttöalueita varten on yhtiöissä nykyisin olemassa omat toimintamallinsa, tietojärjestelmänsä ja sovelluksensa. Runkopankin ja k-MSN –menetelmän käyttö niissä edellyttäisi muutoksia niihin ja kokonaan uusien sovelluksien rakentamista. Keskeistä olisi tällöin itse runkopankkitietovaraston muodostaminen ja siihen tuotettavan hakkuukonetiedon hankinta- ja käsittelymenettelyt. Niitä on kuvattu tarkemmin Metsätehon raportissa 180 (Runkopankin käyttösovellukset). Toinen tärkeä sovellus on katkonnan simulointeja varten tarvittava apteeraussimulaattori. Metsätehon MASI-simulaattori on prototyyppi, joka on suunniteltu nimenomaan runkopankkia tai hakkuukoneen stm-tiedostoja käyttäväksi sovellukseksi. Simulaattoria käytetään k-MSN –menetelmän yhteydessä runkopankin leimikoiden eri pölkkysvaihtoehtojen tuottamiseksi. Runkojen pölkkytiedot tallennetaan tietokantaan.

Kolmanneksi tarvitaan k-MSN –menetelmän sovellus, jolla itse leimikon puustotietojen ja pölkkyjakaumien ennusteet tehdään. Menetelmästä on laadittu Metsätehossa MSN-sovellus, jossa pohjana on käytetty Joensuun yliopistossa tehtyjä Fortran-ohjelmia. Sovellusta ajetaan MASI-simulaattorin käyttöliittymän kautta. Sovelluksen käyttö sekä sen syötteiden ja tulosteiden anto tapahtuu kuvan 10 mukaisen lomakkeen kautta.

Kuva 10. K-MSN –sovelluksen käyttöliittymä MASI-simulaattorissa.

Käyttöliittymän kautta on helppo tehdä k-MSN –hakuja erilaisilla ennakkotiedoilla. Ennakkotietoon voidaan haluttaessa sisällyttää satunnaista virhettä, syöttämällä ennakkotietoa seuraavaan kenttään virheprosentti. Syötetyt ennakkotiedot on myös mahdollista tallentaa käyttäjän nimeämään tiedostoon ja tarvittaessa ladata ne taas käyttöön. Lomakkeen yläosassa on alavetovalikko, josta käyttäjä valitsee apt-tiedoston, jolla runkopankki on apteerattu ja jota käytetään haussa ja ennustetun pölkytyksen muodostamisessa. Haku voidaan suorittaa joko yksittäiselle ennustettavalle kohteelle tai niin, että käytetään kaikkia runkopankin kohteita "ennustettavina" kohteina ja suoritetaan haku kullekin erikseen. Jälkimmäisessä tapauksessa ennakkotietoja ei eksplisiittisesti anneta, vaan käyttöliittymässä vain merkitään mitä ennakkotietoja on tarkoitus käyttää. Tulosteet-välilehdellä käyttäjä antaa halutut tulostiedostojen nimet. Tulostusvaihtoehtoina ovat prd-tiedosto ja Excel-yhteensopivia tekstitiedostoja.

K-MSN –käyttöliittymä on puhtaasti tutkimustarpeita ja sovelluksen testausta varten suunniteltu. Se helpottaa tutkijan työtä poistamalla hitaan ja virheellisiin syötteiden tiedostotasaisen käsittelyn. Jos k-MSN –sovellusta halutaan tuotteistaa osakkaiden käyttöön, suunnitellaan sen käyttöliittymä ja ominaisuudet aina tapauskohtaisesti.

8 LOPUKSI

Runkopankin ja k-MSN –menetelmän käyttömahdollisuuksien selvittämistä on tehty nyt muutamissa hankkeissa varsin kattavasti. Tutkimus on ollut perusluonteista, mutta siinä on tavoiteltu samalla menetelmän soveltamista käytäntöön. Tulokset ovat osoittautuneet lupaaviksi, mutta vielä on kuitenkin jäänyt epäselväksi, kuinka menetelmä voisi toimia laajemmilla runkopankkiaineistoilla. Lisätestauksille olisi tarvetta, mutta niitä kannattaisi tehdä vasta merkittävästi nykyistä suuremmilla aineistoilla. Niihin päästäneen vasta kun jokin puunhankintaorganisaatio aloittaa laajamittaisen runkopankkidataan hankinnan. Toinen vaihtoehto voisi olla simuloidun leimikkoaineiston käyttö. Sellaista voitaisiin tuottaa nykyisen runkopankin pohjalta, mutta siinä olisi otettava huomioon alueelliset metsikkörakenteiden ja hakkuutapojen vaihtelut. Erilaisten vikojen ja rungon laatutekijöiden mukainen katkenta olisi voitava toteuttaa apterauksen simuloinnissa.

Menetelmän soveltamisen kannalta tärkeä tutkimuskohde ovat myös uusien metsikön puustoa kuvaavien hakumuuttujatietojen tietolähteet, joista kiinnostavimpia ovat kaukokartoitusmenetelmiin perustuvat inventointimenetelmät. Tärkeimmäksi ja käyttösovelluksiltaan potentiaalisimmaksi on arvioitu korkearesoluutioisia numeerisia ilmakuvia hyödyntävät yksittäisten puiden automaattiseen tulkintaan perustuvat menetelmät. Niiden osalta olisi kehitettävä menetelmiä latvusten ominaisuuksia kuvaavien estimaattien ja niistä malleilla johdettujen puustotunnusten muodostamiseksi.

Myös ei-parametrisille lähimmän naapurin menetelmille vaihtoehtoisia laskentamenetelmiä olisi mielenkiintoista selvittää. Näitä olisivat etenkin erilaiset data-analyysimenetelmät (data mining) ja neurolaskennan tekniikat,

joilla pyritään analysoimaan suuria tietoaaineistoja automaattisten menetelmien avulla. Kiinnostavia menetelmiä ovat esimerkiksi erilaiset klusterointialgoritmit, pääkomponenttianalyysi sekä SOM-laskentamenetelmät (Self-Organizing Maps). Menetelmäosaamista alueella on olemassa useissa korkeakouluissa ja tutkimuslaitoksissa, joiden kanssa verkostoitumalla kehityshankkeita voitaisiin käynnistää. Referenssejä ja valmiita sovelluksia analysoimalla olisi mahdollista selvittää, onko samankaltaisiin datankuvaus- ja ennustamistarpeisiin esimerkiksi teollisen tuotannon, logistiikan tai kaupan alueilla sovellettu jo aiemmin tehokkaita menetelmiä.

KIRJALLISUUS

- Heikkinen, E.** 2002. Puustotunnusten maastoarvioinnin luotettavuus ja ajanmenekki. Metsäsuunnittelun ja –ekonomian pro gradu –tutkielma. Metsätieteellinen tiedekunta, Joensuun yliopisto. 52 s. + liitteet.
- Kuuramaa, I.** 2004. Runkopankkia hyödyntävien ei-parametrinen tukki-jakauman ennustamismenetelmien hakumuuttujat. Metsä ja –puuteknologian pro gradu –tutkielma. Metsätieteellinen tiedekunta, Joensuun yliopisto. 39 s. + liitteet.
- Malinen, J.** 2003a. Locally Adaptable Non-parametric Methods for Estimating Stand Characteristics for Wood Procurement Planning. *Silva Fennica* 37(1): 109–120.
- Malinen, J.** 2003b. Prediction of characteristics of marked stand and metrics for similarity of log distribution for wood procurement management. D. Sc. (Agr. And For.) thesis. Faculty of Forestry, University of Joensuu.
- Malinen, J., Maltamo, M., Harstela, P.** 2001. Application of Most Similar Neighbour Inference for Estimating Marked Stand Characteristics Using Harvester and Inventory Generated Stem Databases. *International Journal of Forest Engineering*. 12(2): 33–41.
- Räsänen, T. ja Imponen, V.** 2004. Runkopankin käyttösovellukset. Metsätehon raportti 180. 33 s. + liitteet.
- Räsänen, T., Imponen, V., Lindroos, J., Malinen, J. ja Sorsa, J-A.** 2000. Runkopankki puunhankinnan ohjauksen välineenä. Metsätehon raportti 94. 40 s. + liitteet.
- Räsänen, T. ja Lukkarinen, E.** 1998. Puuston ennakkotiedon hankintamenetelmät ja käyttö. Metsätehon raportti 43. 45 s.

K-MSN –menetelmän testauksissa käytetyt hakumuuttujat, menetelmän piirteet, laskentaparametrit sekä aineistojen valinnat ja rajaukset.

Testausyhdistelmän numero	Kuusi															Mänty													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1	2	3	4	5	6	7	8	9	10	11	12	13	14
perushakumuuttujat (taulukot 4 ja 5)	x	x	x	x			x		x	x			x	x	x	x	x	x	x			x		x	x		x	x	x
perushakumuuttujat ilman pinta-alaa					x	x		x			x	x							x	x		x			x				
kuusisorvitukki / mäntytyvitukki mukana	x	x	x				x	x	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x	x	x	
kuusisorvitukki / mäntytyvitukki ei mukana				x	x	x												x	x	x									
ei pinta-alavirheellisten poistoa	x	x	x	x	x	x	x	x								x	x	x	x	x	x	x							
pinta-alavirheelliset kohteet poistettu									x	x	x	x	x	x	x									x	x	x	x	x	x
skaalaustekijänä pinta-ala	x	x	x			x			x							x	x	x			x			x					
skaalaustekijänä puulajin pinta-alaosuus				x	x		x	x		x	x	x	x	x	x				x	x		x	x		x	x	x	x	x
runkomäärä >= 30 kpl / ha	x															x													
runkomäärä >= 50 kpl / ha		x															x												
runkomäärä >= 100 kpl / ha			x															x											
perushakumuuttujien lisäksi ppa-kertymän 25 %:n läpimitta													x		x												x		x
perushakumuuttujien lisäksi ppa-kertymän 75 %:n läpimitta														x	x													x	x
k-MSN:ssä k = 5												x																	