



Annika Kangas



Jyrki Kangas

Annika Kangas ja Jyrki Kangas

Mallit, ennusteet ja simulointi metsätalouden laskentajärjestelmissä

Kangas, A. & Kangas, J. 1997. Mallit, ennusteet ja simulointi metsätalouden laskentajärjestelmissä. *Metsätieteen aikakauskirja – Folia Forestalia* 3/1997: 389–404.

Metsätalouden laskentajärjestelmät pohjautuvat yhä enemmän simuloituihin metsien kehityksen ennusteisiin, mutta näin saatujen kehitysenusteiden ja -skenaarioiden luotettavuutta ei ole riittävästi tarkasteltu. Tässä artikkelissa esitellään erilaisten puustotunnusten ennustamiseen liittyviä epävarmuuden lähteitä. Lähinnä tarkastellaan mallien jäännösvirheiden ja kertoimien sekä malleissa tarvittavien lähtötietojen virheiden vaikutuksia saataviin ennusteisiin. Lisäksi tarkastellaan esimerkkilaskelman avulla virheiden vaikutuksia metsäsuunnitelmaan. Puustotunnusten ennustamisen lisäksi metsäsuunnittelussa aiheuttavat epävarmuutta esimerkiksi päätöksentekijän tavoitteiden analysointi ja kuvaaminen. Ennustamiseen ja suunnitteluun liittyvät erilaiset epävarmuustekijät ja niiden yhteisvaikutukset tulisikin tuntee nykyistä paremmin. Epävarmuus pitäisi myös pystyä kuvaamaan päätöksentekijälle tämän ymmärtämässä muodossa.

Asiasanat: ennustaminen, epävarmuus, metsäsuunnittelu, päätöstuki

Kirjoittajien yhteystiedot: METLA / Kannuksen tutkimusasema, PL 44, 69101 Kannus.
Faksi 06-874 3201, sähköposti annika.kangas@metla.fi ja jyrki.kangas@metla.fi.
Hyväksytty 28.5.1997

1 Johdanto

1.1 Mallit metsätaloudessa

Malleja, useista malleista koostuvia simulointisysteemejä sekä niillä saatavia ennusteita on käytetty metsätalouden laskentajärjestelmissä apuvä-

lineenä aina tietokoneiden käytön yleistymisestä lähtien. Malleja käytetään metsätalouden laskentajärjestelmissä johdattaessa mitatuista tunnuksista halutut tiedot. Simulointia käytetään ennustettaessa metsien kehitystä erilaisilla toimenpideohjelmilla esimerkiksi optimointilaskelmia varten (esim. Siitonen 1983, Pukkala 1993). Simuloimalla myös päivitetään vanhat inventointitiedot ajantasaisiksi. Viime vuosina

simuloinnin soveltaminen mitä erilaisimpien ilmiöiden tarkasteluun myös metsäntutkimuksessa on suuresti lisääntynyt.

Mallien ja ennusteiden käyttö on perusteltua erityisesti silloin, kun niillä saatavan informaation hankinta muilla keinoin olisi kalliimpaa tai hankalampaa. Tietoja, joiden tarkka mittaaminen maastossa on liian kallista, ovat esimerkiksi pystyvuoston eri puutavaralajien tilavuudet. Joidenkin kiinnostavien tunnusten, kuten tulevan kasvun, mittaaminen on jopa mahdotonta.

Simuloinnilla saadut tulokset ovat tilastollisilla malleilla saatuja ennusteita ja sisältävät useita epävarmuustekijöitä. Lisäksi simulointi on aina ehdollista tehtyjen oletusten suhteen: elleivät lähtöoletukset pidä paikkansa, on myös simuloinnin tulosten paikkansapitävyyttä syytä epäillä. Usein kaikista ennustettavaan muuttujaan vaikuttavista ilmiöistä ei ole empiiristä tietoa, ja järjestelmässä hyödynnetään asiantuntijoiden mielipiteitä. Myös niistä aiheutuu ennusteisiin epävarmuutta (ks. myös Alho 1990). Simuloinnin tuloksiin liittyvä epävarmuus johtuukin sekä käytettyjen mallien että tehtyjen oletusten epävarmuudesta.

1.2 Epävarmuuden lähteet ennusteissa

Simuloimalla saatuihin ennusteisiin vaikuttavat monesta lähteestä peräisin olevat virheet. Näistä ilmeisin on käytettyjen mallien aiheuttama epävarmuus. Tilastollisella mallilla tehdyissä ennusteissa on epävarmuutta, joka syntyy neljästä päälähteestä: 1) mallin jäännösvirhe, 2) mallin kertoimien epävarmuus 3) mallin selittävässä muuttujissa olevat virheet ja 4) mallin spesifioinnista johtuvat virheet.

Jäännösvirheiden oletetaan yleensä olevan satunnaisia ja riippumattomia. Jäännösvirheet ovat metsäsovelluksissa kuitenkin yleensä sekä spatiaalisesti että temporaalisesti korreloituneita. Mallien kertoimissa on satunnaista estimointivirhettä, jota voidaan pienentää keräämällä lisää aineistoa. Kertoimien virheen satunnaisuus näkyy kuitenkin vain estimaattaessa mallin parametrit uudelleen uudesta aineistosta. Mallia sovellettaessa kertoimet pysyvät koko ajan samoina, jolloin estimointivirheet voidaan tulkita kertoimien harhaksi (Lappi 1993).

Estimointivirheiden lisäksi mallin kertoimissa voi olla laadinta-aineiston virheistä johtuvaa harhaa (esim. Fuller 1987, 1995). Malli voi olla myös väärän muotoinen tai jokin olennainen selittävä muuttuja voi puuttua. Tämä voi näkyä esimerkiksi siten, että mallin antamat ennusteet ovat pienille puille pääsääntöisesti yliarvioita ja suurille puille aliarvioita.

Lähtötietojen virheet ovat joko 1) mittausvirheitä, 2) otantavirheitä, 3) luokitusvirheitä tai 4) ennustamisvirheitä. Esimerkiksi läpimittojen ilmoittaminen luokittain aiheuttaa luokitusvirhettä. Ennustamisvirhe lähtötiedoissa taas johtuu siitä, että yhdellä mallilla saatua ennustetta käytetään toisessa mallissa lähtötietona. Muun muassa puun tilavuuden laskenta perustuu usein malliketjuun, jossa ensin ennustetaan puun pituus mallilla ja saatua pituusennustetta käytetään edelleen tilavuusmallin lähtötietona. Mittaus- ja luokitusvirheitä ja niiden vaikutuksia tilavuuden ennustamiseen ovat tutkineet mm. Laasasenaho (1973), Hyppönen ja Roiko-Jokela (1978), Kilkki (1983), Päivinen ym. (1992) ja Kangas (1996).

Mallin spesifioinnin epävarmuutta voidaan tarkastella mallin loogisia rakenteita sekä oletuksia arvioimalla tai eri malleilla saatuja ennusteita vertailemalla (Draper 1995, Salminen 1996). Jos kaikki käytettävissä olevat mallit antavat toisiaan tukevia tuloksia, malleja voidaan pitää rakenteellisesti perusteltuina ja tuloksia tältä osin luotettavina. On kuitenkin mahdollista, että kaikki käytettävissä olevat mallit perustuvat samanlaisiin virheellisiin oletuksiin, jolloin ennusteet ovat kaikesta huolimatta harhaisia. Tällaisia epäilyksiä on esitetty esimerkiksi kasvihuoneilmiön vaikutuksia ennustavista malleista (Loehle 1996).

Ennusteiden epävarmuus koostuu kahdesta komponentista: ennusteiden harhasta sekä satunnaisista ennustamisvirheistä. Satunnaiset mallien jäännösvirheet sekä lähtötiedon virheet aiheuttavat ennusteisiinkin satunnaista virhettä. Harhaa ennusteisiin tulee silloin, kun lähtötiedot ovat harhaiset tai kun malli on väärin muotoiltu. Kuitenkin myös satunnaiset virheet lähtötiedoissa aiheuttavat ennusteisiin harhaa, mikäli ennustettava muuttuja riippuu epälineaarisesti virheellisesti mitatusta muuttujasta (Kilkki 1979, Kangas 1996).

2 Ennustamisharha

2.1 Ennustamisharhan syntyminen

Oletetaan, että ennustamiseen käytetään mallia

$$Y = f(B, X) + e \quad (1)$$

missä X on selittävien muuttujien vektori, B on parametrivektori ja e on jäännösvirhe. Vaikka oletettaisiin, että funktio f tunnetaan ja että parametrivektori B on estimoitu virheettä, voivat ennusteet olla harhaisia, mikäli selittävissä muuttujissa on esimerkiksi mittausvirhettä. Usein selittäviä muuttujia X ei ole havaittu täsmällisesti, vaan havaitut arvot x sisältävät virhettä

$$x = X + u \quad (2)$$

missä u on lähtötiedon virheiden vektori. Tällöin ennusteet saadaan itse asiassa

$$\hat{Y} = f(B, X + u) \quad (3)$$

jolloin niihin sisältyy mittausvirheestä johtuva ennustamisvirhe $f(B, x) - f(B, X)$.

Mikäli funktio f on selittävien muuttujien X suhteen lineaarinen, ennusteet ovat harhattomia, jos mittausvirheetkin ovat harhattomia. Linearisessa tapauksessa virheen odotusarvo $E(f(B, x) - f(B, X))$ on siis nolla, jos $E(u) = 0$. Muussa tapauksessa ennuste on harhainen, vaikka satunnaisvirheet olisivatkin harhattomia.

Mikäli funktio f on epälineaarinen virheellisesti mitatun X :n suhteen, saadaan ennustamisharhan likiarvo Taylorin sarjakehitelmän avulla (ks. Rothatgi 1976, Kilkki 1979, Gertner 1991, Lappi 1993). Ennustusharhan odotusarvo voidaan esittää funktion osittaisderivaattojen, selittävien muuttujien X_j ($j = 1, \dots, p$) ja mittausvirheiden varianssien $\text{var}(u_j)$ avulla. Mikäli eri tunnusten mittausvirheet ovat korreloimattomia, saadaan ennustamisharha kaavalla (Lappi 1993)

$$E(f(x) - f(X)) \approx \frac{1}{2} \sum_{j=1}^p \frac{\partial^2 f}{\partial X_j^2} |X (\text{var}(u_j)) \quad (4)$$

Ennustamisharhan merkitys korostuu, kun ennustetaan metsien kasvua pitkälle ajanjaksolle. Kas-

vua ennustetaan yleisesti peräkkäisissä viiden vuoden jaksoissa. Tällöin kullakin jaksolla tehdyt ennustevirheet kertautuvat kaikissa myöhemmissä ennusteissa, koska kunkin jakson ennusteet perustuvat edellisten jaksosten tuloksiin. Siksi kasvunennusteet ovat sitä epätarkempia, mitä kaukaisempaa tulevaisuutta ne kuvaavat. (Salminen 1996, Kangas 1997a). Vastaava ennustamisharha syntyy myös silloin, kun mallien lähtötiedot on laskettu keskiarvona alkuperäisistä mitatuista arvoista (Moerur ja Ek 1981).

2.2. Ennustamisharhan korjaaminen

Teoreettisesti yksinkertaisin tapa korjata ennustamisharhaa on käyttää ns. Monte Carlo -simulointia. Tällöin funktion (1) arvo lasketaan useilla virheen u arvoilla ja Y :n odotusarvo estimoidaan saatujen funktion arvojen keskiarvolla

$$E(Y) \approx \frac{1}{m} \sum_{i=1}^m f(B, X + u_i) \quad (5)$$

Arvot poimitaan jakaumasta satunnaisesti. Otoskoko m voi vaihdella tilanteen mukaan vaikkapa kymmenestä jopa tuhansiin. Toinen tapa on poimia u :n arvoja jakaumasta tasavälein. Tällöin kutakin u :n arvoa painotetaan suhteessa sen frekvenssiin (Kilkki 1979). Tällöin kyseessä ei kuitenkaan ole Monte Carlo -simulointi, vaan funktion odotusarvon laskeminen integroimalla numeerisesti virheen jakauman yli (esim. Ripley 1987). Usein numeerinen integrointi on menetelmänä tehokkaampi kuin varsinainen Monte Carlo -simulointi.

Ennustamisharhaa voidaan korjata myös Taylorin sarjakehitelmän avulla. Tällöin korjaustermi lasketaan kaavan (4) avulla. Kaavassa tarvitaan todellisia muuttujan arvoja X , joita ei kuitenkaan yleensä ole tiedossa. Korjauksia voidaan silti soveltaa, kun kyseessä on mallivirhe: tällöin estimoitu arvo on populaation odotusarvo annetuilla selittävien muuttujien arvoilla ja todelliset arvot vaihtelevat tämän odotusarvon ympärillä (Lappi 1993).

Monte Carlo -menetelmän hyvä puoli on menetelmän soveltuvuus monimutkaisiinkin ongelmiin. Sitä voi käyttää myös silloin, kun käytettävät mallit eivät ole derivoituvia. Monte Carlo -menetelmä edellyttää virheiden yhteisjakauman tuntemista. Taylorin

sarjakehitelmää käytettäessä virheiden odotusarvojen sekä varianssin ja kovarianssin tunteminen riittää. Virheiden yhteisjakaumaa ei käytännössä yleensä tunneta. Tällöin Monte Carlo -menetelmää käytettäessä tehdäänkin usein normaalijakaumaoletus, jolloin menetelmien vaatimat lähtötiedot ovat identtiset. Taylorin sarjakehitelmän käyttö ei myöskään ole vailla ongelmia: menetelmän toimivuus riippuu mm. virheiden suuruudesta suhteessa mallin epälineaarisuuteen. (Kangas 1996).

2.3. Esimerkki harhan korjaamisesta

Esimerkkilaskelma kuvaa malliketjun vaikutusta puiden tilavuuden ennustamiseen (Kangas 1996). Puun tilavuuden laskemiseen käytettiin Laasasenahon (1982) kolmen selittäjän yhtälöä $v = f(d, d_6, h) + e_v$. Tilavuusennusteita, jotka saatiin mallilla käyttäen selittävinä muuttujina mitattua läpimittaa, pituutta ja yläläpimittaa, pidettiin esimerkissä todellisina tilavuuksina (ts. oletettiin, että $e_v = 0$). Ennustamisharha tässä esimerkissä aiheutui siitä, että tilavuusmallissa käytetty pituus ja/tai yläläpimitta ennustettiin mallilla. Käytetty pituusmalli on muotoa $\ln(h) = f(d, G, d/d_m, T) + e_{\ln(h)}$, missä G on metsikön pohjapinta-ala, d_m pohjapinta-alamediaanipuun läpimitta ja T on metsikön ikä. Yläläpimittamalli on muotoa $d_6 = f(d, G, d/d_m, T, h) + e_d$.

Yläläpimittamalli on lineaarinen, ja virhetermi e_d oletettiin normaalisti jakautuneeksi. Virhetermin varianssin estimaatti saatiin mallin estimoinnin yhteydessä. Tällöin Monte Carlo -menetelmässä otos poimitaan satunnaisesti normaalijakaumasta, jonka varianssi on $\text{var}(e_d)$. Tilavuuden odotusarvon estimaatti saadaan laskemalla kullakin virhetermin realisaatiolla tilavuus ja laskemalla näiden keskiarvo

$$\hat{v} = \frac{1}{m} \sum_{i=1}^m f(d, h, \hat{d}_6 + e_{di}) \quad (6)$$

Pituusmalli perustuu logaritmuunnokseen, jolloin voitiin olettaa, että logaritmisessa muodossa virhetermi on normaalisti jakautunut varianssilla $\text{var}(e_{\ln(h)})$. Tällöin normaalijakaumasta poimitut virhetermit $e_{\ln(h)i}$ voidaan lisätä pituuden logaritmin estimaattiin ennen pituuden estimaatin laskemista. Kun logaritminen virhetermi on normaalisti jakautunut, pituuden virhetermi $e_{hi} (= \exp(e_{\ln(h)i}))$ on arit-

meettisessä skaalassa multiplikatiivinen ja lognormaalisti jakautunut. Tällöin tilavuudet voidaan laskea

$$\hat{v} = \frac{1}{m} \sum_{i=1}^m f(d, \hat{h} \times e_{hi}, d_6) = \frac{1}{m} \sum_{i=1}^m f(d, \exp(\ln(\hat{h}) + e_{\ln(h)i}), d_6) \quad (7)$$

Monte Carlo -simuloinnilla tulee myös otettua huomioon epälineaarisesta muunnoksesta johtuva harha, joten erillistä harhakorjausta logaritmimalliin ei tarvita (Kilki 1979).

Keskimääräiset koepuiden tilavuudet Uusimaa-Hämeen, Pirkka-Hämeen sekä Itä-Hämeen metsäkeskusten alueella laskettiin kahdeksannen VMI:n koepuuaineistossa, käyttäen tilavuuden ennustamiseen mitattuja pituuden ja yläläpimitan arvoja ("todellinen tilavuus") sekä käyttäen edellä esitetyillä malleilla ennustettuja pituuksia ja yläläpimittoja. Ennusteita korjattiin sekä Monte Carlo -menetelmään että Taylorin sarjakehitelmään perustuen. Kun sekä yläläpimitta että pituus ennustettiin mallilla, ennustettua pituutta käytettiin yläläpimittamallin lähtötietona. Tästä aiheutuvaa lisävirhettä ei ole korjauksissa huomioitu (taulukko 1).

Suurimmillaan eli noin 2,5 dm³ harha oli, kun pituus ennustettiin mallilla ja yläläpimitta oli mitattu. Koska harha on keskimääräinen runkokohtainen harha, hehtaarikohtaiseksi harhaksi tulisi noin 1,25 m³/ha, jos runkoja olisi hehtaarilla 500 kappaletta. Koska pituusennusteiden varianssi kasvaa puun kasvaessa, myös harha on sitä suurempi, mitä suuremmasta puusta on kysymys. Kun sekä pituus että yläläpimitta on ennustettu mallilla, läpimital-

Taulukko 1. Alueen mäntyrunkojen keskitilavuus dm³ eri menetelmillä laskettuna. Tosi tilavuus on 534,42 dm³. Suluissa malliketjun käytöstä seuraava harha.

Menetelmä	Korjaamaton	Monte Carlo -korjaus	Taylorin sarjakehitelmä -korjaus
Tilavuus kun h ja d_6 ennustettu mallilla	533,06 (1,36)	535,70 (-1,28)	535,69 (-1,27)
Tilavuus, kun h ennustettu mallilla	531,86 (2,56)	533,54 (0,88)	533,63 (0,79)
Tilavuus, kun d_6 ennustettu mallilla	533,45 (0,97)	534,30 (0,12)	534,63 (-0,21)

taan yli 40 cm:n puiden ennusteet olivat yli 20 dm³ suurempia kuin ”todelliset tilavuudet”. Esimerkkiaineiston tapauksessa Monte Carlo -menetelmä sekä Taylorin sarjakehitelmä toimivat yhtä hyvin. Menetelmien paremmuus eri tilanteissa riippuukin lähinnä niiden vaatimista resursseista, taustaoletuksista ja -informaatiosta.

3 Simuloimalla saatujen ennusteiden luottamusväli

3.1 Empiirinen luottamusväli

Monimutkaisella laskentajärjestelmällä saaduille ennusteille on hankalaa antaa virhemarginaali, koska se muodostuu usean mallin virhemarginaalin yhteistuloksena. Virhemarginaali voidaan kuitenkin laskea empiirisesti, jos käytettävissä on soveltuva tarkistusaineisto. Empiirisenkin virhemarginaalin laskemiseen liittyy ongelmia, mikäli tarkistusaineistossa on mittausvirheitä: tällöin sekä mallien tulokset että vertailukohtana olevat mittaukset sisältävät satunnaista virhettä.

Tulevaisuuden kasvua ei voi mitata. Kasvuennusteiden luotettavuutta voidaan arvioida kuitenkin tekemällä kasvuennusteita jonkin jakson alussa mitattujen tietojen perusteella ja vertaamalla niitä kyseisen jakson kasvumittauksiin. Jos näin halutaan tehdä, tarvitaan kokeita, jotka on mitattu tarkasti ennustejakson alussa ja lopussa. Kokeet on kuitenkin yleensä perustettu jotakin erityistarkoitusta varten, eikä niihin sisälly riittävästi erilaisia

metsiköitä, jotta niiden perusteella voitaisiin antaa luotettava arvio ennustamisen luotettavuudesta kaikissa olosuhteissa. Lisäksi hyvin pitkäaikaisia kokeita, esimerkiksi viidenkymmenen vuoden mittaisia, on melko vähän.

Esimerkkilaskelman malliketjuilla saatuja tilauksia verrattiin empiiriseen aineistoon, jolloin saatiin niiden suhteelliset keskivirheet. Laasasenahon (1982) mallilla lasketut, mitattuihin tietoihin perustuvat tilavuudet oletettiin todellisiksi tilavuukiksi. Puukohtaisen tilavuuden keskivirhe määritettiin olettaen, että 1) tilavuus ennustetaan suoraan läpimitan ja metsikkötunnusten avulla (samat selittävät muuttujat kuin pituusmallissa), tai että 2) ensin ennustetaan pituus ja/tai yläläpimitta mallilla ja lasketaan tilavuus Laasasenahon mallilla (taulukko 2).

Tilavuus saatiin esimerkissä ennustettua luotettavimmin, jos pituus tai yläläpimitta oli tunnettu ja vain toinen ennustettiin mallilla. Seuraavaksi tarkin tulos saatiin ennustamalla tilavuutta suoraan läpimitan ja metsikkötunnusten avulla. Pituus- ja yläläpimittamallin sisällyttäminen laskentajärjestelmään ei siis tuonut lisäinformaatiota. Ellei malliketjuun tuoda uutta informaatiota, esimerkiksi kalibroimalla pituus- ja yläläpimittamalleja mittausten avulla, malliketju onkin vain tavallista monimutkaisempi malli, jossa tilavuuden ennusteet kuitenkin perustuvat pelkkään läpimittaan ja metsikkötunnuksiin. Tämä pätee myös yleisemmin malliketjuille. Myös Mowrerin (1989) tulosten mukaan mallien lisääntyvä monimutkaisuus vain heikensi ennusteiden luotettavuutta.

Malliketjun käyttöön voi kuitenkin olla muita syitä: jos kaikki tunnuksot tarvitaan joka tapauksessa, esimerkiksi tulevan kasvun simuloimista varten, malliketjun käyttö voi olla perusteltua. Malliketjua käyttäen saadut tunnuksot (esimerkiksi pituus ja tilavuus) ovat keskenään yhteensopivia. Mikäli kaikki tarvittavat tunnuksot laskettaisiin erillisillä malleilla läpimitan ja metsikkötunnusten perusteella, eri tunnusten väliset suhteet voisivat muodostua epäloogisiksi (Ranneby ja Svensson 1990).

Taulukko 2. Puukohtaisen tilavuuden suhteellinen keskivirhe eri menetelmillä.

Menetelmä	Suhteellinen keskivirhe
Tilavuus, kun h ja d_6 ennustettu mallilla	14,49 %
Tilavuus, kun h ennustettu mallilla	7,35 %
Tilavuus, kun d_6 ennustettu mallilla	7,62 %
Tilavuus suoraan läpimitan ja metsikkötunnusten perusteella	12,29 %

3.2 Ennusteiden luottamusväli mallien perusteella

Paitsi empiirisesti, luottamusväli voidaan tuottaa myös Taylorin sarjakehitelmän perusteella, jolloin puhutaan ns. varianssi-propagoinnista, tai simuloimalla, jolloin puhutaan yleensä Monte Carlo -menetelmästä (esim. Rubinstein 1981, Ripley 1987). Sovellettaessa Taylorin sarjakehitelmää varianssi-propagointiin käytetään yleensä ensimmäisen asteen sarjakehitelmää. Mikäli selittävien muuttujien virheet ovat korreloimattomia, tästä voidaan johtaa Y :n ennusteen varianssiksi

$$\text{var}(f(x) - Y) \approx \sum_{j=1}^p \left(\frac{\partial f}{\partial X_j} \Big|_X \right)^2 \text{var}(u_j) + \text{var}(e) \quad (8)$$

Taylorin sarjakehitelmää ovat kasvuennusteiden luotettavuuden arviointiin käyttäneet esim. Mowrer ja Frayer (1986), Gertner (1987) ja Mowrer (1990).

Harhan ja varianssin yhteisvaikutus, keskineliövirhe MSE , saadaan laskemalla yhteen varianssi ja harhan neliö. Jos harhakomponentteja on useita, ne voivat vaikuttaa eri suuntiin ja osittain kompensoida toisensa. Tällöin keskineliövirhettä ei saada laskemalla yhteen eri harhakomponenttien neliöt, vaan laskemalla harhakomponenttien muodostaman kokonaisharhan neliö (Lappi 1993, vrt. Kilkki 1983).

Monte Carlo -menetelmässä luotettavuusarviot tehdään ”arpomalla” kuhunkin yksittäiseen ennusteeseen virhetermi kyseisen mallin virheen jakaumasta. Kun yksittäisten puiden ennusteet summaa yhteen, saadaan yhdelle metsikölle yksi skenaario mahdollisesta tulevasta kehityksestä. Tällaisia skenaarioita tuotetaan tietokoneella satoja. Näistä sadoista skenaarioista muodostuu ennusteiden vaihtelualue. Monte Carlo -simuloinnilla tuotettuja ennusteiden luotettavuusarvioita ovat esittäneet mm. Gertner ja Dzialowy (1984), Mowrer ja Frayer (1986), Gertner (1987), Mowrer (1990), McRoberts (1996) sekä Kangas (1997a).

Taylorin sarjakehitelmällä saadut varianssiarviot ovat tehdyissä tutkimuksissa olleet yleensä pienempiä kuin Monte Carlo -menetelmällä saadut. Tämä voi johtua esimerkiksi siitä, että Taylorin sarjakehitelmä edellyttää virheiden olevan melko pieniä suhteessa mallin epälinearisuuteen. Pitkän aikavälin ennusteissa virheet voivat olla liian suuria

menetelmän toimivuutta ajatellen. Lisäksi varianssi-propagointi perustuu yleensä ensimmäisen asteen Taylorin sarjaan. Korkeampien termien mukaanottaminen voisi pienentää menetelmien välisiä eroja. (Mowrer 1990).

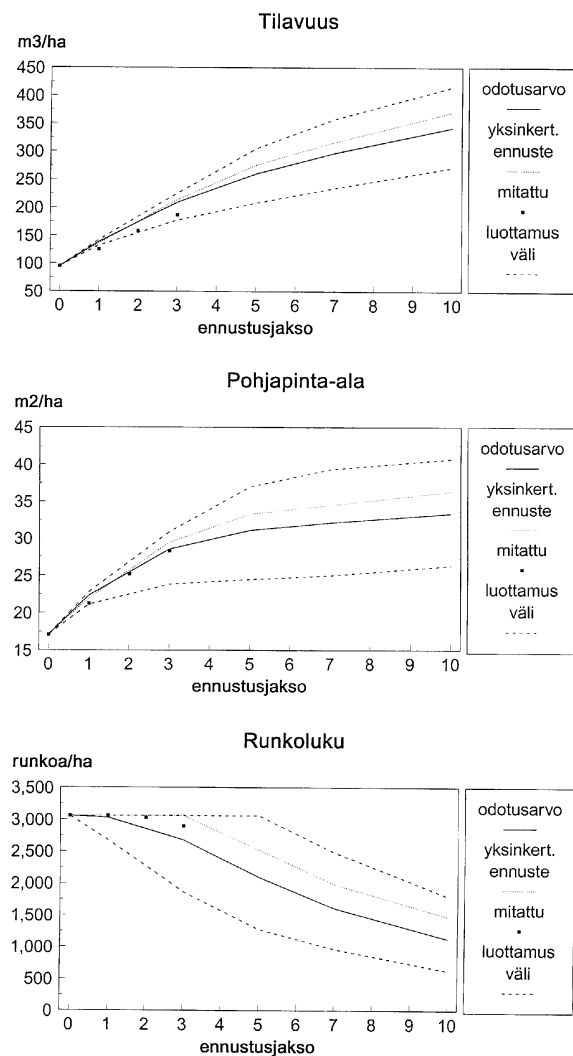
Monte Carlo -menetelmällä erisuuntaisten harhakomponenttien yhteisvaikutus saadaan suoraan (Kangas 1997a). Tosin Monte Carlo -menetelmä on yleensä laskennallisesti Taylorin sarjakehitelmää vaativampi. Toisaalta varianssi-propagoinnilla on helpompi ottaa huomioon esimerkiksi lähtötietojen ja mallien parametrien välisten korrelaatioiden vaikutukset (Mowrer 1990).

Tutkimusten, joissa on käytetty Monte Carlo -menetelmää, keskinäinen vertailu on vaikeaa. Osassa tutkimuksista on keskitytty parametrien estimointivirheisiin (Mowrer ja Frayer 1986, Mowrer 1990, McRoberts 1996), kun taas muissa tutkimuksissa on keskitytty jäännösvirheisiin (Gertner 1987, Kangas 1997a). Jotta saataisiin selvitettyä ennusteiden kokonaispävarmuus, pitäisi molempien komponenttien vaikutukset ottaa huomioon.

3.3 Esimerkki simuloidusta luotettavuusarviosta

Kangas (1997a) on tutkinut sekä kasvun ennusteiden ennustamisharhaa että ennustevarianssia simuloimalla. Tutkimuksessa käsiteltiin pelkästään mallin jäännösvirheen aiheuttamaa ennustevirhettä. Metsikön kasvatus tehtiin simulointisysteemillä, johon kuuluu läpimitan ja pituuden kasvumalli sekä latvussuhdemalli (Hynynen 1995a), valtapuiden pituuden kasvumalli (Vuokila ja Väliäho 1980), tilavuusmalli (Laasasenaho 1982), sekä metsikön maksimirunkolukua kuvaava malli (Hynynen 1993).

Läpimitan kasvu kuvattiin funktiolla $i_{d5} = f(d, cr, GL, G, H_{dom}, H_{100}) + e_{id}$, missä d on läpimita, cr on latvussuhde, H_{dom} on valtapituus, H_{100} valtapituusboniteetti, GL on kohdepuuta suurempien puiden pohjapinta-ala ja G on metsikön pohjapinta-ala. Pituuskasvu kuvattiin funktiolla $i_{h5} = f(d, D_{dom}, IH_{dom}) + e_{ih}$, missä D_{dom} on valtaläpimita, joka määritellään sadan paksuimman puun keskiläpimitäksi (Hynynen 1995a). Valtapituuden pituuskasvu, IH_{dom} , ennustettiin metsiköittäin funktiolla $IH_{dom} = f(H_{dom}, T)$, missä T on metsikön ikä (Vuokila ja Väliäho 1980).



Kuva 1. Metsikön tilavuuden (a), pohjapinta-alan (b) ja runkoluvun (c) yksinkertainen ennuste, ennusteiden odotusarvo, ennusteiden luottamusväli sekä metsikön mitattu kehitys.

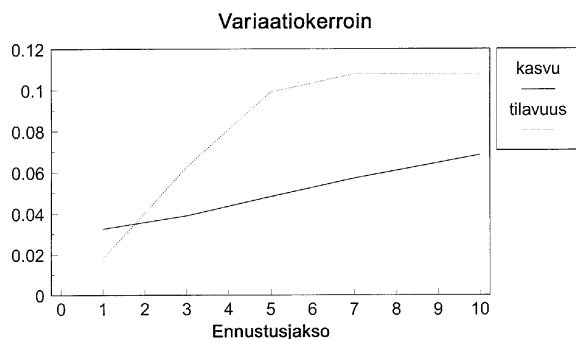
Luonnonpoistuma kuvattiin mallilla, joka kertoo maksimaalisen runkojen määrän keskiläpimitan ja valtapituusboniteetin funktiona (Hynynen 1993). Jos metsikön runkoluku ylittää tämän raja-arvon, puujoukosta poistetaan puita satunnaisesti siten, että runkoluku vastaa maksimirunkolukua. Poistettavat puut valittiin läpimittaluokittaisilla kuolemistodennäköisyyksillä, jotka laskettiin käytetystä aineistosta.

Kasvumallien lisäksi tarvittiin myös latvussuhdemalli $cr = f(G, Thin, H_{dom}, d, h) + e_{cr}$, missä G on metsikön pohjapinta-ala, h pituus ja $Thin$ harvennusvaikutusta kuvaava muuttuja (Hynynen 1995b). Tilavuusfunktiona käytettiin Laasasenahon kahden selittäjän funktiota $v = f(d, h) e_v$ (Laasasenaho 1982).

Simuloinnin tulokset laskettiin käyttäen lähtötietoja yhtä ns. HARKAS-aineiston koelaa (ks. Kangas 1997a). Simulointitulokset laskettiin kahdella tavalla. Ensiksi tulokset laskettiin siten, että jäännösvirheitä ei otettu mitenkään huomioon. Tällöin kunkin ennustuskauden tuloksia käytettiin sellaisenaan seuraavan kauden ennusteiden lähtötietoina (myöhemmin kutsutaan yksinkertaiseksi ennusteeksi). Lisäksi tulokset laskettiin siten, että aina kutakin mallia käytettäessä mallin jäännösvirheen jakaumasta generoitiin virheelle realisaatio, joka lisättiin ennusteeseen. Tällä tavoin laskettiin koko 50 vuoden ennustejakso 1000 kertaa. Näiden tuhannen realisaation keskiarvoa kullekin tarkasteltavalle tunnukselle pidetään ennusteen odotusarvona ja realisaatioiden välistä varianssia ennusteen varianssina. Ennusteiden 95 % luottamusväli puolestaan saatiin valitsemalla luottamusrajaksi se ennuste, jonka ylätai alapuolella oli 2,5 % simuloituista skenaarioista. (Kangas 1997a).

Metsikön tilavuuden, pohjapinta-alan sekä runkoluvun kehitys laskettiin yksinkertaisena ennusteenä ja kehityksen odotusarvo lisäksi 1000 skenaarion keskiarvona (kuva 1). Tässä laskelmassa metsikköä kasvatettiin käsittelemättömänä. Näiden kahden käyrän välinen ero kuvaa mallien epälineaarisuudesta sekä epävarmoista lähtötiedoista aiheutuva ennustamisharha. Simuloimalla saatu ennusteen luottamusväli sekä mitattu kehitys auttavat myös arvioimaan ennusteiden luotettavuutta (kuva 1). Sekä ennustamisharha että variaatiokertoimet (kuva 2) kasvavat ennustejakson pituuden myötä.

Saaduissa ennusteissa luonnonpoistumamallin aiheuttama epävarmuus vaikutti tuloksiin yksittäisistä malleista eniten. Luonnonpoistumamallin merkitys kuitenkin pieni, jos metsiköiden puustot harvennettiin. Ennusteharha ei katoa harvennusten myötä, mutta se voi käsitellyissä metsiköissä olla toisen suuntainen kuin käsittelemättömissä. Kehitysennusteet laskettiin käsittelemättömän vaihtoehdon lisäksi myös kahdelle eri alaharvennusvaihtoehdolle. Kaikissa vaihtoehdoissa metsikön kasvun odotus-



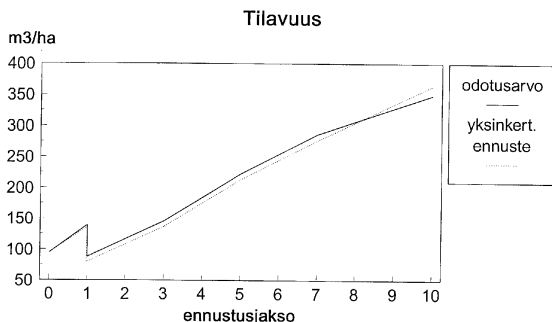
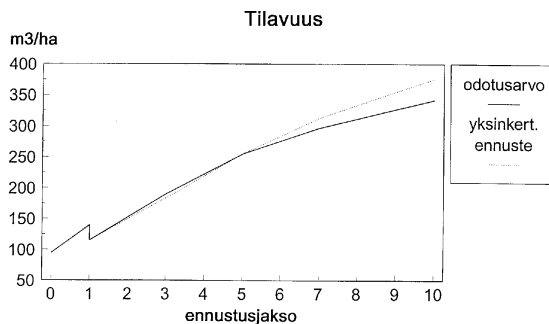
Kuva 2. Metsikön tilavuusennusteen ja tilavuuskasvun ennusteen variaatiokerroin.

arvo oli suurempi kuin yksinkertaisella ennusteella saatu. Kun luonnonpoistumaa ei ole, kasvun ennustusharha dominoi tuloksia (kuva 3).

Esimerkin laskelmassa tehtiin monia yksinkertaisia oletuksia. Käytetyt mallit oletettiin oikean muotoisiksi ja harhattomiksi. Mallien kertoimien estimointivirheitä ei otettu huomioon. Virheet oletettiin kaikissa malleissa normaalisti jakautuneiksi. Tämä virherakenne ei välttämättä ole kaikkien mallien osalta perusteltu. Kaikkien mallien virheet oletettiin autokorreloituneiksi peräkkäisillä ennustuskauilla, mutta vain pituuden ja läpimitan kasvumallien oletettiin olevan keskenään korreloituneita. Monissa tapauksissa virheet ovat myös metsikköittäin korreloituneita, mitä ei tässä esimerkissä myöskään ole huomioitu. Metsikkökohtainen korrelaatio kuitenkin lisää metsikköennusteiden epävarmuutta huomattavasti (Kangas 1997a). Esimerkin epävarmuusarviot voidaankin tulkita epävarmuuden alarajaksi.

4 Mittausvirheen vaikutus mallien kertoiimiin

Jos mallin laadinta-aineistossa on selittävässä muuttujissa virhettä, estimoidut mallin kertoimet ovat harhaiset. Jos mittausvirheet ovat riippumattomia ja additiivisia (kaava 2), yksinkertaisessa yhden selittäjän lineaarisessa mallissa kulmakerroin on harhainen nollaa kohti, eli estimoitu malli on liian



Kuva 3. Metsikön tilavuuden yksinkertainen ennuste sekä ennusteen odotusarvo, kun metsikön runkoluvusta on harvennettu 30 % (a) ja 60 % (b).

loiva (Fuller 1987). Tällöin estimoidun mallin kulmakerroin on muotoa

$$\hat{\beta}_1 = \frac{\sigma_{XY}}{\sigma_X^2 + \sigma_u^2} \quad (9)$$

Harhaton kulmakerroin voidaan estimoida jakamalla saatu kulmakerroin ns. luotettavuussuhteella (*reliability ratio*) λ

$$\lambda = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2} \quad (10)$$

missä σ_X^2 on todellisten arvojen varianssi ja σ_u^2 on mittausvirheiden varianssi. Mittausvirhe lisää mallin jäännösvarianssia ja vaikuttaa myös kertoimien merkitsevyyteen.

Mittausvirheen vaikutus ei ole aina samanlainen. Mikäli virhetermi ei noudata mallia (2), vaikutus on monimutkaisempi. Sama pätee myös silloin, kun mallissa on useita selittäjiä tai malli on epälineaarinen. Riippuen havaittujen ja todellisten selittäjien arvojen suhteista ja myös muista mallin muuttujis-

ta saattaa olla, että 1) mittausvirheiden todelliset vaikutukset eivät näy, 2) havaitussa aineistossa näkyy riippuvuuksia, joita ei ole tosi aineistossa tai jopa 3) mallin kertomien merkit voivat vaihtua. (Carroll ym. 1995).

Mikäli mallissa on useita selittäjiä, joista yhdessä on mittausvirhettä (X) ja muut on mitattu oikein (Z), mittausvirhe vaikuttaa kaikkien muuttujien kertoiimiin, elleivät X ja Z ole toisistaan riippumattomia (Carroll ym. 1995). Jos esimerkiksi puiden kasvua selitetään sekä valtapituusboniteetilla että pohjapinta-alalla, ja valtapituusboniteetissa on ennustamisvirhettä, on myös pohjapinta-alan kerroin harhainen, koska muuttujat ovat korreloituneita.

Mittausvirheitä voi esiintyä myös selitettävässä muuttujassa Y . Tällöin estimoidut kertoimet ovat harhattomia, mikäli havaittu y on harhaton ja mittausvirhe on riippumaton ja additiivinen. Mittausvirheet selittävässä muuttujassa kuitenkin kasvattavat mallin varianssia. (Carroll ym. 1995).

Mittausvirheet vaikeuttavat lähinnä mallien tulokintaa: jos halutaan saada selville tietyn tekijän merkitys mallissa, tulokset ovat harhaanjohtavia. Goelz ja Burk (1992, 1996) tutkivat mittausvirheiden vaikutuksia pituusboniteetikäyriin. He totesivat, että pituusboniteetti-indeksi tuli mittausvirheiden vuoksi aliarvioituksi hyvillä kasvupaikoilla ja yliarvioituksi huonoilla kasvupaikoilla. Lisäksi mittausvirheet pienensivät käyrien muotojen eroja eri kasvupaikkojen välillä.

Vaikka mallin kertoimet ovat harhaiset laadinta-aineiston virheiden vuoksi, ennusteet voivat silti olla harhattomia. Tällainen tilanne on, kun selittävien muuttujien jakauma mallin sovellusaineistossa on samanlainen kuin laadinta-aineistossakin (Lindley 1947). Mikäli mittausvirheet laadinta- ja sovellusaineistossa ovat eri tavoin jakautuneet, tai esimerkiksi jonkin selittävän muuttujan keskiarvo on sovellusaineistossa merkittävästi pienempi tai suurempi kuin laadinta-aineistossa, harhaiset kertoimet aiheuttavat ongelmia (Ganse ym. 1992).

Harhaiset kertoimet voivat siis olla ongelmallisia, kun mallia käytetään laadinta-aineistonsa ulkopuolelle. Ongelmia voi syntyä myös silloin, kun simulointitutkimuksilla haetaan yksittäisten kertomien vaikutuksia: esimerkiksi harvennusten vaikutusten arviointi simuloimalla perustuu lähinnä pohjapinta-alan tai jonkin muun kilpailua kuvaavan

tunnuksen vaikutuksen analysointiin. Jos tämän muuttujan kerroin on harhainen, voivat tehdyt johtopäätökset olla vääriä (Kangas 1997b).

5 Mallien, simuloinnin ja ennusteiden epävarmuuden vaikutus metsäsuunnitteluun

5.1 Epävarmuus, suunnittelu ja päätöksenteko

Metsäsuunnittelu koskee aina tulevaisuutta. Suunnittelun tehtävä on tuottaa päätöstukea: informaatiota metsien hoidon ja käytön vaihtoehdoista sekä niiden toteuttamisen todennäköisistä seuraamuksista. Tavoitteena on auttaa päätöksentekijää tekemään tavoitteidensa kannalta mahdollisimman hyvät valinnat, jotka samalla täyttävät esimerkiksi lakien ja asetusten asettamat vaatimukset. Tässä artikkelissa tarkastellaan lähinnä taktista ja strategista metsäsuunnittelua, joissa aikahorisontti ulottuu viidestä vuodesta kymmeneen vuosiin.

Suunnittelussa tuotetaan usein metsäsuunnitelma, joka sisältää suosituksen tarkasteltavan metsäalueen tuotanto-ohjelmaksi metsiköittäisine käsitteilyehdotuksineen. Keskeinen osa metsäsuunnittelua on tarkasteltavan metsäalueen tulevaisuuden tuotantomahdollisuuksien selvittäminen. Mallit ja simulointi ovat nykyisessä metsäsuunnittelussa toistuvasti käytettäviä tuotantomahdollisuuksien arvioinnin työkaluja. Niiden avulla ennustetaan, miten erilaiset metsänkäsittelyohjelmat vaikuttavat metsän kehitykseen ja metsän kuvaan sekä asetettujen tavoitteiden täyttymiseen. Simulointimallien ja -tekniikoiden käyttöönotto on monipuolistanut, syventänyt ja tehostanut olennaisesti metsäsuunnittelun laskelmia. Kuten edellä on selostettu, kaikki mallit ja simuloinnit sisältävät kuitenkin enemmän tai vähemmän epävarmuutta. (ks. myös Siitonen 1993).

Metsäsuunnittelun näkökulmasta ennusteiden epävarmuus on päätöstukeen liittyvä riskitekijä (Pukkala ja Kangas 1995). Mallien epävarmuuden takia pieleen menneet ennusteet voivat johtaa väärin johtopäätöksiin ja niiden perusteella tehtäviin

epäoptimaalisiin valintoihin. Virheiden vaikutus voi olla kahtalainen: valitaan ”väärä” toimenpidevaihtoehto, jolloin saavutettava tavoitefunktion arvo on pienempi kuin saavutettavissa ollut optimaalinen ratkaisu (esim. Bell 1982), tai optimaalisen ratkaisun arvo yliarvioidaan (esim. Bell 1985).

Virheet kehitysennusteissa eivät kuitenkaan aina johda väärin valintoihin: yksittäisen metsikön toimenpideohjelmaa valittaessa riittää, että vaihtoehtojen käsittelyjen järjestys on oikea. Kun tarkasteltavia muuttujia ja metsiköitä tulee lisää, on ennustevirheiden vaikutusten arvioiminen monimutkaisempaa. Erilaisten tunnusten ennustamisen luotettavuus on myös hyvin erilainen, mikä osaltaan vaikeuttaa tulosten tulkintaa. Esimerkiksi metsikön tilavuusennusteen luotettavuus voi olla moninkertainen lahoppuun määrän ennusteeseen nähden.

Epävarmuuden merkitystä suunnittelussa on tarkasteltu lähinnä lineaarisen optimoinnin tapauksessa. Voidaan osoittaa, että mikäli metsikön kehitysennusteissa on satunnaisia virheitä, lineaarisella optimoinnilla saatavat optimiratkaisut ovat optimistisesti harhaisia (Pickens ja Dress 1988, Pickens ym. 1991). Sama pätee myös silloin, kun tavoitefunktion arvoissa on satunnaista virhettä (Hobbs ja Hepenstal 1989). Ongelmia aiheuttavat myös lineaarisen optimoinnin rajoitteiden satunnaiset virheet (Itami 1974). Virheet johtavat myös siihen, että asetettuja rajoitteita ei suurella todennäköisyydellä saavuteta (esim. Pickens ja Dress 1988). Tämä johtuu siitä, että optimointialgoritmien luonteen vuoksi tietyn vaihtoehdon valintatodennäköisyys on sitä suurempi, mitä enemmän päätösmuuttujan arvoa yliarvioidaan. Sama ilmiö on havaittu myös käytännön suunnittelussa (Siitonen 1996), ja se koskee millä tahansa optimointitekniikalla tuotettuja ratkaisuja.

Laskelmien ja ennusteiden tulkinnan kannalta olisi tärkeää tuntea ennusteiden luotettavuudet. Optimointilaskelmien tulokset palvelevat harhaisinakin suunnittelua, mutta niiden tulkitseminen muihin tarkoituksiin on kyseenalaista. Luotettavuus pitäisi pystyä ilmoittamaan ymmärrettävästi myös päätöksentekijälle, joka päätöstukea hyödyntää, jotta hän ei kuvittelisi tulosten olevan varmuuden vallitessa laskettuja. Päätöksentekijää tulee vähintään varoittaa virheiden ja epävarmuuden mahdollisista seurauksista. Laskelmien epävarmuuksien tuntemi-

nen auttaa myös tiedostamaan, että suunnittelu ei tuota valmiita päätöksiä ja että vastuu niitä hyödyntäen tehtävistä valinnoista on inhimillisellä päätöksentekijällä.

Suunnittelun yhteydessä termit riski ja epävarmuus määritellään sen mukaan, mikä on käytössä oleva tietämys suunnitelmavaihtoehtoista – erityisesti siitä, mihin eri suunnitelmien toteuttaminen voi johtaa. Päätöksenteosta riskin vallitessa puhutaan silloin, kun kunkin suunnitelman toteuttaminen saattaa tuottaa erisuuruisia hyötyjä ja kunkin hyötyvaihtoehdon todennäköisyys tunnetaan. Epävarmuuden vallitessa päätös joudutaan tekemään, jos mahdollisten hyötyvaihtoehtojen todennäköisyyksiä ei tunneta. Käytännön suunnittelussa ja päätöksenteossa eivät eri suunnitelmavaihtoehtojen mahdollisten hyötyjen jakaumat ole varmuudella tiedossa ennusteiden epävarmuuden takia. Suunnittelu ja päätöksenteko joudutaankin aina tekemään epävarmuuden vallitessa. Tosin käytännössä laskelmat tehdään teknisesti ikäänkuin riskin vallitessa, jolloin todennäköisyysjakaumat oletetaan tunnetuiksi.

Päätöstukea epävarmuuden vallitessa monipuolistaa suunnitelmavaihtoehtojen hyötyjakaumien tuottaminen. Eri tavoin riskiin ja epävarmuuteen suhtautuvat päätöksentekijät nimittäin tekevät valintansa eri kriteerein. Esimerkki valaiskoon asiaa. Riskin karttaja valitsee metsänuudistamistavan, joka mahdollisimman suurella varmuudella johtaa tietty vähimmäisvaatimukset täyttävään taimikkoon. Riskin suosija taas valitsee uudistamistavan, joka nollasta poikkeavalla, pienelläkin todennäköisyydellä tuottaa erittäin tuottoisan taimikon, vaikka täydellisen epäonnistumisenkin mahdollisuus olisi huomattava. Riskineutraali taas valitsee sen uudistamistavan, jonka tuottama hyödyn odotusarvo on suurin. (Kangas 1992). Sama logiikka toimii myös metsäalueen suunnitelmavaihtoehtojen vertailussa.

5.2 Inhimillinen arvottaminen epävarmuuden lähteenä metsäsuunnittelussa

Jotta suunnitelmavaihtoehtoja voidaan vertailla numeerisesti päätöksentekijän tavoitteiden suhteen, myös tavoitteet ja preferenssit tulee kuvata numeerisesti. Tavoitteiden ja preferenssien mallinnuksen

epävarmuudet aiheuttavat epävarmuutta suunnittelulaskelmiin siinä kuin vaikkapa puuston kasvumallienkin epävarmuudet. Jo vertailu- ja optimointimenetelmien tavoitteiden kuvaukselle asettamat tekniset vaatimukset aiheuttavat virhettä tavoiteanalyysiin. Metsäsuunnittelun tutkimuksessa on viime aikoina panostettu erityisesti Suomessa tavoitteiden ja preferenssien mallinnustekniikoiden kehittämiseen.

Tavoiteanalyysissä laadittavien mallien oikeellisuutta ei kuitenkaan pystytä arvioimaan aivan samaan tapaan kuin esimerkiksi puuston kasvua kuvaavien mallien epävarmuutta. Puuston kasvu voidaan mitata ja mallin antamaa tulosta voidaan verrata mitattuun havaintoon. Tavoitteiden ja preferenssien mallinnuksessa tällainen vertailumahdollisuus puuttuu. Tavoitemallin laadinta tehdään jokaisessa suunnittelutilanteessa erikseen, eikä päätöksentekijän tavoitemallin vertaaminen muiden päätöksentekijöiden malleihin tai välttämättä edes saman päätöksentekijän aikaisempiin tavoitteiden mallinnuskertojen tuloksiin kerro mitään mallin oikeellisuudesta.

Tavoitemallin laatimiseksi välttämättömään tiedusteluun annettujen vastausten yhdenmukaisuutta ja siitä johdettuja tunnuslukuja voidaan käyttää erinä mallin epävarmuuden mittareina (Alho ym. 1996). Ongelmana on tällöin se, että vastaukset voivat olla täysin ristiriidattomat mutta samalla yhdenmukaisesti pielessä. On esimerkiksi mahdollista, että päätöksentekijä tarkoituksellisesti pyrkii keskenään ristiriidattomiin vastauksiin senkin kustannuksella, että ei koe kaikkia antamiaan vastauksia todenmukaisiksi. Ryhmäpäätösprosesseissa, kuten yhtymän metsien päätöksenteko, ja osallistavassa suunnittelussa, missä tavoitteita udellaan kansalaisilta ja kansalaisryhmiltä, voidaan lisäksi törmätä tietoiseen tavoitemallin vääristelyyn. Tulevaa neuvottelutilannetta ajatellen saattaa olla eduksi aluksi yliarvostaa itselle tärkeitä tavoitteita ja samalla taktikoida itselle tinkimisvaraa erimielisyyksien sovittamistilanteeseen. Aukotonta menetelmää tällaisen taktikoinnin paljastamiseen ei valitettavasti ole.

Eräs keino parantaa tavoiteanalyysin luotettavuutta on toteuttaa se vuorovaikutteisesti ja iteratiivisesti (Kangas ym. 1996). Vuorovaikutteisuus ei välttämättä johda laadittavan tavoitemallin täsmen-

tymiseen siinä mielessä, että lopullinen malli voitaisiin tulkita päätöksentekijän oikeiden tavoitteiden ja niiden tärkeyksien tarkaksi kuvaajaksi. Eikä täydellisen tavoitemallin laatiminen ole siinä edes päämääränä. Vuorovaikutteisessa suunnittelussa tavoitemallia käytetään teknisenä työkaluna haettaessa mahdollisimman hyvin päätöksentekijän näkemykset tyydyttävää suunnitelmaa: arvioitavana on pikemminkin tuotetun suunnitelman hyvyys eikä tavoitemallin täsmällisyys. Parhaaksi valittavaan suunnitelmaan johtava tavoitemalli ei välttämättä ole parhaiten päätöksentekijän tavoitteet kuvaava funktio.

Laskelmissa tarvittavan perustiedon puutteita voidaan yrittää paikata asiantuntemuksen mallinnuksella. Viime aikoina on panostettu erityisesti ekologisen asiantuntemuksen mallintamiseen suunnittelulaskelmia varten. Asiantuntemuksen hyödyntämisessä on yleisten mallinnukseen liittyvien epävarmuustekijöiden (luvut 2–4) lisäksi omat erityiset, muunmuassa asiantuntijoiden erimielisyyksistä ja pätevyyseroista johtuvat virhelähteensä (Alho ym. 1996). Asiantuntijan vastausten epävarmuuden analysointiin liittyvät paljolti samat ongelmat kuin päätöksentekijän tavoitteiden mallinnuksen epävarmuuden arviointiin. Kun hyödynnetään useita asiantuntijoita samassa tehtävässä, voidaan epävarmuuden analysoinnissa kuitenkin turvautua myös eri asiantuntijoiden antamien lausuntojen keskinäisten erojen tarkasteluun sen sijaan, että tarkasteltaisiin vain kunkin asiantuntijan vastausten sisäistä ristiriidattomuutta.

5.3 Esimerkki lähtötietojen epävarmuuden vaikutuksista suunnitteluun

Esimerkki lähtötietojen epävarmuuden merkityksestä suunnittelussa tehtiin siten, että muodostettiin kuvitteellinen kymmenen vaihtoehdon vaihtoehtoavaruus, jossa oli kaksi päätösmuuttujaa U_1 ja U_2 (taulukko 3). Kumpikin muuttujista oli esimerkin tapauksessa päätöksentekijälle yhtä tärkeä, ja kokonaishyöty oletettiin muodostuvan niiden painotetusta summasta

$$U = 0,5 U_1 + 0,5 U_2 \quad (11)$$

Taulukko 3. Päätösmuuttujien $U1$ ja $U2$ arvot kymmenessä kuvitteellisessa vaihtoehdossa.

Vaihtoehto	$U1$	$U2$	U
1	1,0	0,3	0,65
2	0,9	0,5	0,7
3	0,8	0,6	0,7
4	0,7	0,8	0,75
5	0,6	1,0	0,8
6	0,5	0,8	0,65
7	0,4	0,6	0,5
8	0,3	0,4	0,35
9	0,2	0,2	0,2
10	0,1	0,1	0,1

Tavoitefunktion maksimi saadaan siis $\max(U) = 0,5 \max(U1 + U2)$. Mikäli päätösmuuttujiin liittyy additiivinen virhe, havaittu hyöty päätösmuuttujasta $U1$ on $0,5 U1 + e1$. Tällöin tavoitefunktion maksimi on

$$\max(u) = 0,5 \max(U1 + e1 + U2 + e2) = \max(U + e) \quad (12)$$

missä $e = 0,5 e1 + 0,5 e2$. Tavoitefunktion maksimin odotusarvo on siis $E(\max(u)) = E(\max(U + e))$.

Mikäli esimerkin virheet ovat normaalisti jakautuneita, voidaan $\max(u)$:n odotusarvo johtaa analyttisesti normaalisti jakautuneen muuttujan maksimin jakauman perusteella. Voidaan osoittaa, että $E(\max(U + e))$ on aina vähintään yhtä suuri kuin $\max(U)$, joten maksimi on optimistisesti harhainen, kun päätösmuuttujien arvoihin liittyy epävarmuutta. Voidaan myös osoittaa, että harha kasvaa virheen e hajonnan ja vaihtoehtojen määrän kasvaessa (ks. Stuart ja Ord 1993 s. 491).

Usein virheiden jakauma on niin monimutkainen, että tarvitaan simulointia. Yleensä yhden metsikön vaihtoehtojen virheet ovat selvästi korreloituneita – jos metsikön tilavuus on yliarvioitu yhden vaihtoehdon tapauksessa, näin on yleensä myös muiden vaihtoehtojen tapauksessa. Mitä suurempi korrelaatio virheissä on, sitä pienempi on optimoinnin harha. Usein virheiden hajonta kasvaa tunnuksen arvon kasvaessa. Tällaisessa tapauksessa odotettavissa oleva harha on suurempi kuin additiivisten ja riippumattomien virheiden tapauksessa.

Esimerkin simuloinnissa päätösmuuttujiin $U1$ ja $U2$ generoitiin satunnaisia virheitä erilaisilla ha-

Taulukko 4. Arvioidun hyödyn odotusarvo erilaisilla päätösmuuttujien $U1$ ja $U2$ suhteellisilla virrehajonnoilla. Virheiden logaritmit oletettiin normaalisti jakautuneiksi ja niiden korrelaatioksi oletettiin 0,9. Päätösmuuttajat oletettiin toisistaan riippumattomiksi. Todellinen hyötyfunktion arvo on 0,8.

Virrehajonta	$U2$			
$U1$	0,05	0,1	0,3	0,5
0,05	0,801	0,802	0,827	0,881
0,1	0,802	0,805	0,820	0,865
0,3	0,823	0,826	0,860	0,913
0,5	0,892	0,895	0,903	0,970

jonnoilla. Generoitujen virheiden logaritmit oletettiin additiivisiksi ja normaalisti jakautuneiksi, jolloin päätösmuuttujien virheet ovat multiplikaatiivisia ja lognormaalisti jakautuneita. Virheet voidaan tällöin tulkita päätösmuuttujien suhteellisiksi virheiksi. Logaritmisten virheiden korrelaatioksi vaihtoehtojen välillä oletettiin 0,9, mutta päätösmuuttajat $U1$ ja $U2$ oletettiin toisistaan riippumattomiksi. Havaittu u :n arvo siis saatiin kussakin vaihtoehdossa $u = 0,5 U1 e1 + 0,5 U2 e2$. Simuloinnissa valittiin optimivaihtoehto näin saatujen virhettä sisältävien muuttujien perusteella.

Kullakin virrehajontojen yhdistelmällä optimivaihtoehto valittiin 1000 kertaa. Näille 1000 optimiarvolle laskettiin keskiarvo, joka tulkittiin arvioidun hyödyn odotusarvoksi. Lisäksi laskettiin, miten suuri oli optimaalisten valintojen osuus sekä odotusarvo tappiolle, joka syntyy kun valitaan todellisen optimivaihtoehdon sijasta jokin muu vaihtoehto.

Päätösmuuttujien virheillä oli selvä vaikutus hyötyfunktion odotusarvoon. Mitä suurempia virheet lähtötiedoissa olivat, sitä harhaisempi oli hyötyfunktion odotusarvo. Esimerkin tapauksessa optimoinnista johtuva harha oli pahimmillaan 21,2 % (taulukko 4). Myös tappion odotusarvo oli sitä suurempi, mitä suurempi oli virheiden hajonta (taulukko 5). Optimoitaessa epävarmoilla tiedoilla joudutaan siis pettymään kahdestakin syystä: optimivaihtoehdon arvo yliarvioidaan, eikä valittu vaihtoehto kenties tuota yhtä hyvää tulosta, kuin todellisuudessa paras vaihtoehto.

Taulukko 5. Niiden realisaatioiden osuus (%), joissa parhaaksi vaihtoehdoksi valittiin todellisuudessa paras vaihtoehto, eli vaihtoehto 5. Suluissa odotettavissa oleva, väärän vaihtoehdon valinnasta johtuva tappio hyötyfunktion arvossa.

Virrehajonta U1	U2			
	0,05	0,1	0,3	0,5
0,05	100 (0)	98,9 (0,001)	75,6 (0,017)	56,4 (0,034)
0,1	99,3 (0)	96,8 (0,002)	73,7 (0,018)	56,4 (0,037)
0,3	75,1 (0,022)	74,0 (0,022)	62,4 (0,032)	54,3 (0,042)
0,5	57,8 (0,039)	57,6 (0,040)	51,1 (0,045)	45,7 (0,052)

6 Lopuksi

Edellä esitellyissä tutkimuksissa on paneuduttu eräisiin metsätalouden laskentajärjestelmien epävarmuuden lähteisiin. Monet epävarmuuden lähteet on tähän mennessä tehdyissä tutkimuksissa jätetty kokonaan huomiotta. Tällainen tekijä on puuston kehityksen ennustamisen osalta esimerkiksi kesän sääolojen aiheuttama kasvun vuotuinen vaihtelu. Pitkän ajan ennusteissa myös erilaiset tuhot, puuden luontainen uudistuminen sekä taimien kuoleminen aiheuttavat epävarmuutta. Ongelmia voi aiheuttaa lisäksi ennusteiden alueellinen luotettavuus: keskimääräiset luotettavuusarviot eivät välttämättä päde pienille osa-alueille, kuten yksittäisille metsätiloille. Puuston kehityksen ennustamisen lisäksi metsäsuunnittelun laskelmiin tuovat epävarmuutta monet muut seikat, kuten hinta- ja kustannustietojen ennusteet.

Useimpien mainittujen epävarmuustekijöiden vaikutuksia voidaan arvioida tilastotieteen menetelmin. Laskelmiin liittyy kuitenkin aina myös sellaista epävarmuutta, jota ei voida ottaa huomioon. Esimerkiksi ennustettaessa tulevaa kehitystä menneiden tapahtumien perusteella joudutaan oletamaan muuttujien välisten suhteiden sekä olosuhteiden pysyvän ennallaan. Suunnitteluun epävarmuutta aiheuttavat ennakoimattomat metsänkäsittelysäästösten ja ohjeistojen muutokset, vaihtoehtojen ar-

vottamisessa käytettävän perustiedon puutteet sekä tavoiteanalyysin epävarmuudet. Suurten alueiden metsälaskelmien tuloksia tulkittaessa on lisäksi huomattava, että puuston kehitysennusteiden laadinnassa joudutaan oletamaan jokin metsänkäsittelymenetelmien valikoima ja määrällinen jakauma tarkasteltavalla alueella, sekä toimenpiteiden kohdentuminen tietynlaisiin metsiköihin. Esimerkiksi juuri metsänhoitosuosituksen muutokset, mutta myös – ja ennenkaikkea – metsänomistajien ja metsätalouden muiden päätöksentekijöiden valintojen arvaamattomuus aiheuttavat tällaisten oletusten paikkansapitämättömyyttä.

Ennusteissa ja suunnittelun eri vaiheissa olevasta epävarmuudesta huolimatta jokin valinta on tehtävä kussakin päätöstilanteessa. Tulevaisuuden tuotantomahdollisuuksia sekä päätösvaihtoehtoja ja niiden seuraamuksia kuvaavan informaation tuottaminen on keskeinen osa metsäsuunnittelua. Mallit, simulointi ja ennusteet ovat välttämättömiä numeerisen metsäsuunnittelun työkaluja. Metsäsuunnittelun tehtävä on tuottaa käytettävissä olevin keinoin ja saatavilla olevan tietämyksen perusteella mahdollisimman vankka ja todenmukainen päätöstuki. Päätöstuki on kuitenkin puutteellinen, ellei siinä esitettävien laskelmien luotettavuutta kyetä selvittämään ja – mikä tärkeää – ilmaisemaan sitä päätöksentekijälle ymmärrettävällä tavalla. Metsäsuunnittelun tutkimuksessa tulisi jatkossa pyrkiä kehittämään menetelmiä ja tekniikoita tarkastella suunnittelulaskelmien epävarmuuksia nykyistä perusteellisemmin.

Edellä esitetty yksinkertainen laskelma metsäsuunnitelmien vertailusta tilanteessa, missä suunnitelmavaihtoehtojen arvioinnissa kahden tarkastellun tavoitteen suhteen on epävarmuutta (mutta ei systemaattisia virheitä), osoittaa, että mallien ja simulointien epävarmuudella todella on merkitystä myös suunnittelussa eikä vain yksittäisten ennusteiden tulkinnassa. Arviointi- ja arvottomismallien epävarmuudet johtavat saatavan hyödyn yliarviointiin, jos hyötyennusteeksi valitaan optimointilaskelman mukainen hyödyn maksimi. Käytännössä tämä tarkoittaa sitä, että jos optimilaskelmassa esimerkiksi maksimoidaan puuston tilavuutta suunnittelukauden lopussa ja optimiratkaisun mukainen lopputilavuus tulkitaan tilavuuden kehitysennusteeksi, ennuste on yliarvio. Linearisessa optimoin-

nissa myös ratkaisuun vaikuttavien rajoitteiden suhteen on vastaavanlaisia ongelmia. Mallien epävarmuudet voivat lisäksi johtaa väärän suunnitelman arvioimiseen eniten hyötyä tuottavaksi, ellei epävarmuuksien vaikutuksia vertailulaskelmaan analysoida.

Metsäsuunnittelussa tulisi päätösvaihtoehtojen hyvyydet pystyä arvioimaan kaikkien päätöksentekijää kiinnostavien seikkojen suhteen. Niin yhteiskunnan kuin kansalaismielipiteen ja yhä useampien yksityismetsänomistajienkin näkökulmissa muut kuin puuntuotannolliset tavoitteet ovat saaneet aikaisempaa suuremman painoarvon. Viime aikoina onkin pyritty kehittämään mallinnustekniikoita ja malleja metsänkäsittelyvaihtoehtojen arvioimiseksi esimerkiksi keräilytuotteiden, maiseman kauneuden ja eri eliölajien ja -ryhmien elinympäristövaatimusten suhteen. Toistaiseksi muuta kuin puuston kehittymistä kuvaavat mallit sisältävät yleensä huomattavan paljon epävarmuutta. Mallinnumenetelmien edelleen kehittämisessä, mallien tuottamisessa ja mallien tulosten analysoinnissa riittääkin työsarkaa ja haasteita tutkijoille.

On kuitenkin muistettava, että kaikkia suunnittelun ja päätöstuen tarpeita ei voida tyydyttää numeerisilla malleilla ja simuloinnilla. Numeronmurskaus ja optimointilaskelmat ennusteineen ovat vain osa päätöstukea. Tärkeää olisi tutkia myös mallien ja simuloinnin käyttöä osana koko päätöksentekoprosessia sekä kehittää keinoja integroida numeeriset laskelmat ja ennusteet muuhun päätöstukeen kuvailevampine ja kvalitatiivisine tarkasteluineen. Suunnittelu voi olla tiedettäkin, mutta käytännön toteutuksessa tarvitaan myös aimo annos taidetta.

Kiitokset

Kirjoittajat haluavat kiittää professori Juha Alhoa ja VTT Juha Lappia lukuisista hyödyllisistä kommenteista ja korjausesityksistä.

Kirjallisuus

- Alho, J. 1990. Stochastic methods in population forecasting. *Int. J. Forecasting* 6: 521–530.
- , Kangas, J. & Kolehmainen, O. 1996. Uncertainty in the expert predictions of the ecological consequences of forest plans. *Applied Statistics* 45: 1–14.
- Bell, D.E. 1982. Regret in decision making under uncertainty. *Oper. Res.* 30(5): 961–981.
- 1985. Disappointment in decision making under uncertainty. *Oper. Res.* 33(1): 1–27.
- Carroll, R.J., Ruppert, D. & Stefanski, L.A. 1995. Measurement errors in nonlinear models. *Monographs on Statistics and Applied Probability* 63. Chapman & Hall. London.
- Draper, D. 1995. Assessment and propagation of model uncertainty. *J. R. Statist. Soc. Series B* 57(1): 45–97.
- Fuller, W.A. 1987. Measurement error models. John Wiley & Sons. New York, 440 s.
- 1995. Estimation in the presence of measurement error. *Int. Stat. Review* 63: 121–147.
- Gainse, R.A., Amemiya, Y. & Fuller, W.A. 1983. Prediction when both variables are subject to error, with application to earthquake magnitudes. *J. Am. Stat. Assoc.* 78: 761–765.
- Gertner, G. 1987. Approximating precision in simulation projections: an efficient alternative to Monte Carlo methods. *Forest Science* 33: 230–239.
- 1991. Prediction bias and response surface curvature. *Forest Science* 37: 755–765.
- & Dzialowy, P.J. 1984. Effects of measurement errors on an individual tree-based growth projection system. *Canadian Journal of Forest Research* 14: 311–316.
- Goelz, J.C.G. & Burk, T.E. 1992. Development of a well-behaved site index equation: jack pine in north central Ontario. *Canadian Journal of Forest Research* 22: 776–784.
- & Burk, T.E. 1996. Measurement error causes bias in site index equations. *Canadian Journal of Forest Research* 26: 1585–1593.
- Hobbs, B.F. & Hopenstal, A. 1989. Is optimization optimistically biased? *Water Resources Research* 25(2): 152–160.
- Hynynen, J. 1993. Self-thinning models for even-aged stands of *Pinus sylvestris*, *Picea abies* and *Betula pendula*. *Scandinavian Journal of Forest Research* 8: 326–336.
- 1995a. Predicting the growth response to thinning for Scots Pine stands using individual-tree growth models. *Silva Fennica* 29(3): 225–246.

- 1995b. Predicting tree crown ratio for unthinned and thinned Scots pine stands. *Canadian Journal of Forest Research* 25: 57–62.
- Hypönen, M. & Roiko-Jokela, P. 1978. On the accuracy and effectivity of measuring sample trees. *Folia Forestalia* 356. 25 s. ISBN 951-40-0344-6.
- Itami, H. 1974. Expected objective value of a stochastic linear program and the degree of uncertainty of parameters. *Manage. Sci.* 21(3): 291–301.
- Kangas, A. 1996. On the bias and variance of tree volume predictions. *Scandinavian Journal of Forest Research* 11: 281–290.
- 1997a. On the prediction bias and variance of long-term growth predictions. *For. Ecol. Manage. Painossa*.
- 1997b. Effect of errors-in-variables on coefficients of a growth model and on prediction of growth. *For. Ecol. Manage. Painossa*.
- Kangas, J. 1992. Metsikön uudistamisketjun valinta – monitavoitteiseen hyötyteoriaan perustuva päätösanalyysimalli. Summary: Choosing the regeneration chain in a forest stand: A decision model based on multi-attribute utility theory. Joensuun yliopiston luonnontieteellisiä julkaisuja 24. 230 s.
- , Pukkala, T. & Pykäläinen, J. 1996. Vuorovaikutteinen heuristinen optimointi yksityismetsien suunnittelussa. *Folia Forestalia – Metsätieteen aikakauskirja* 1996(3): 231–244.
- Kilki, P. 1979. Outline for a data processing system in forest mensuration. *Silva Fennica* 13: 368–384.
- 1983. Sample trees in timber volume estimation. *Acta Forestalia Fennica* 182. 35 s. ISBN 951-651-057-4.
- Laasasenaho, J. 1973. Unequal probability sampling by dbh cumulator. *Communicationes Instituti Forestalis Fenniae* 79(6).
- 1982. Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae* 108. 74 s. ISBN 951-40-0589-9.
- Lappi, J. 1993. Metsäbiometrian menetelmiä. *Silva Carelica* 24. University of Joensuu, 182 s. ISBN 951-708-157-X.
- Lindley, D.V. 1947. Regression lines and the linear functional relationship. *J. Royal Stat. Soc. (Suppl.)* 9: 218–244.
- Loehle, C. 1996. Do simulations predict unrealistic dieback? *Journal of Forestry*. s. 13–15.
- McRoberts, R.E. 1996. Monte Carlo simulations of non-linear size-age relationships. In: *Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium*. May 21–23, 1996, Fort Collins, Colorado. s. 659–666.
- Moeur, M. & Ek, A.R. 1981. Plot, stand and cover-type aggregation effects on projections with an individual based stand growth model. *Canadian Journal of Forest Research* 11: 309–315.
- Mowrer, H.T. 1989. The effect of forest simulation model complexity on estimate precision: Julkaisussa: Burkhardt, H.E. & Johann, K. (toim.). *Artificial intelligence and growth models for forest management decisions*. Proceedings of a meeting held in Vienna, Austria, September 18–22, 1989. Virginia Polytechnic Institute and State University, School of Forestry and Wildlife Resources, Blacksburg, Virginia, USA. Publication FWS-1-89: 110–118.
- 1990. Estimating components of propagated variance in growth simulation model projections. *Canadian Journal of Forest Research* 21(3): 379–386.
- & Frayer, W.E. 1986. Variance propagation in growth and yield projections. *Canadian Journal of Forest Research* 16: 1196–1200.
- Pickens, J.B. & Dress, P.E. 1988. Use of stochastic production coefficients in linear programming models: objective function distribution, feasibility, and dual activities. *Forest Science* 34(3): 574–591.
- , Hof, J.G. & Kent, B.M. 1991. Use of chance-constrained programming to account for stochastic variation in the A-matrix of large-scale linear programs. *A Forestry application. Ann. Oper. Res.* 31: 511–526.
- Pukkala, T. 1993. Metsäsuunnitteluohjelma MONSU. Ohjelmiston toiminta ja käyttö. *Moniste* 42 s.
- & Kangas, J. 1995. A method for integrating risk and attitude toward risk into forest planning. *Forest Science* 42(2): 198–205.
- Päivinen, R., Nousiainen, M. & Korhonen, K.T. 1992. Puutunnusten mittaamisen luotettavuus. *Folia Forestalia* 787. 18 s. ISBN 951-40-1197-X.
- Ranneby, B. & Svensson, S.A. 1990. From sample tree data to images of tree populations. Proceedings of the International IUFRO S. 4.02 and S. 6.04 Symposium, May 14–16, 1990 Birmensdorf, Switzerland. s. 102–117. ISBN 3-905620-06-5.
- Ripley, B.D. 1987. *Stochastic Simulation*. John Wiley & Sons. 237 s. ISBN 0-471-81884-4.
- Rohatgi, V.K. 1976. *An introduction to probability theory and mathematical statistics*. John Wiley & Sons. 684 s. ISBN 0-471-81884-4.
- Rubinstein, R.Y. 1981. *Simulation and the Monte Carlo Method*. John Wiley & Sons. 278 s. ISBN 0-471-08917-6.
- Salminen, H. 1996. Miten arvioidaan metsää kuvaavia malleja? Teoksessa: Hökkä, H., Salminen, H. & Varvola, M. (toim.) *Pohjoisten metsien kasvu – ennen, nyt ja tulevaisuudessa*. Metsäntutkimuspäivä Rovaniemellä 1996. Metsäntutkimuslaitoksen tiedonantoja 589. s. 47–58.

- Siitonen, M. 1983. A long-term forestry planning system based on data from the Finnish national forest inventory. Proceedings of the IUFRO subject group 4.02 meeting in Finland, September 5–9, 1983. University of Helsinki, Department of Forest mensuration and management. Research Notes 17: 195–207.
- 1993. Experiences in the use of forest management planning models. Tiivistelmä: Kokemuksia mallien käytöstä metsätalouden suunnittelussa. *Silva Fennica* 27(2): 167–178.
- 1996. MELA ja metsien kehityksen ennustaminen. Teoksessa: Hynynen, J. & Ojansuu, R. (toim.) Puuston kehityksen ennustaminen – MELA ja vaihtoehtoja. Tutkimusseminaari vantaalla 1996. Metsäntutkimuslaitoksen tiedonantoja 612. s. 7–19.
- Stuart, A. & Ord, K.J. 1993. Kendall's advanced theory of statistics. Volume 1. Distribution theory. Sixth edition. Edward Arnold. 676 s.
- Vuokila, Y. & Väliäho, H. 1980. Viljeltyjen havumetsiköiden kasvatusmallit. *Communicationes Instituti Forestalis Fenniae* 99(2). 271 s.

52 viitettä