

SILVA FENNICA

1985 Vol. 19 N:o 3

Sisällys <i>Contents</i>	Metsien inventoinnin tilastolliset menetelmät	211
	<i>Statistical methods in forest inventories</i>	211
	ELJAS POHTILA & TAPANI POHJOLA: Maan kunnostus männyn viljelyssä Lapissa	245
	<i>Summary: Soil preparation in reforestation of Scots pine in Lapland</i>	270
	JUHA LAPPI & MARKKU SIITONEN: A utility model for timber production based on different interest rates for loans and savings	271
	<i>Summary: Säästö- ja lainakorkoon perustuva puuntuotannon hyötymalli</i>	280
	PEKKA HELLE & JYRKI MUONA: Invertebrate numbers in edges between clear-fellings and mature forests in northern Finland	281
	<i>Seloste: Selkärangattomien runsaussuhteista avohakkuualueiden ja metsien reunalueilla Pohjois-Suomessa</i>	294
	MATTI KÄRKKÄINEN & MARKKU HALINEN: Mäntysahatukien minimivaatimusten täsmäntäminen	295
	<i>Summary: Reappraisal of minimum requirements of pine sawlogs</i>	323
	EEVA-LIISA JUKOLA-SULONEN & MAIJA SALEMAA: A comparison of different sampling methods of quantitative vegetation analysis	325
	<i>Seloste: Kvantitatiivisessa kasvillisuusanalyysissä käytettyjen näytteenottomenetelmien vertailu</i>	337

Silva Fennica

A QUARTERLY JOURNAL FOR FOREST SCIENCE

PUBLISHER: THE SOCIETY OF FORESTRY IN FINLAND

OFFICE: Unioninkatu 40 B, SF-00170 HELSINKI 17, Finland

EDITOR: SEPPO OJA

EDITORIAL BOARD:

KAUKO HAHTOLA (Chairman), RIIKKO HAARLAA, TIINA HEINONEN, JOUKO HÄMÄLÄINEN, VEIKKO KOSKI, SIMO POSO and MARKKU KANNINEN (Secretary)

Silva Fennica is published quarterly. It is sequel to the Series, vols. 1 (1926) – 120 (1966). Its annual subscription price is 180 Finnish marks. The Society of Forestry in Finland also publishes *Acta Forestalia Fennica*. This series appears at irregular intervals since the year 1913 (vol. 1).

Orders for back issues of the Society, and exchange inquiries can be addressed to the office. The subscriptions should be addressed to: Academic Bookstore, P.O. Box 128, SF-00101 Helsinki 10, Finland.

Silva Fennica

NELJÄNNESVUOSITTAIN ILMESTYVÄ METSÄTIETEELLINEN AIKA-
KAUSKIRJA

JULKAISIJA: SUOMEN METSÄTIETEELLINEN SEURA

TOIMISTO: Unioninkatu 40 B, 00170 Helsinki 17

VASTAAVA TOIMITTAJA:

SEPPO OJA

TOIMITUSKUNTA:

KAUKO HAHTOLA (Puheenjohtaja), RIIKKO HAARLAA, TIINA HEINONEN, JOUKO HÄMÄLÄINEN, VEIKKO KOSKI, SIMO POSO ja MARKKU KANNINEN (Sihteeri)

Silva Fennica, joka vuosina 1926–66 ilmestyi sarjajulkaisuna (niteet 1–120), on vuoden 1967 alusta lähtien neljännesvuosittain ilmestyvä aikakauskirja. Suomen Metsätieteellinen Seura julkaisee myös *Acta Forestalia Fennica*-sarjaa vuodesta 1913 (nide 1) lähtien.

Tilauksia ja julkaisuja koskevat tiedustelut osoitetaan seuran toimistolle. *Silva Fennican* tilaushinta on 120 mk kotimaassa, ulkomailla 180 mk.

METSIIEN INVENTOINNIN TILASTOLLISET MENETELMÄT

STATISTICAL METHODS IN FOREST INVENTORIES

Saapunut toimitukselle 10. 4. 1985

Suomen tilastoseura ja Suomen metsätieteellinen seura järjesti 17. 10. 1984 keskustelutilaisuuden. Siellä pidetyt esitelmät käsitteivät Suomessa käytettyjen metsien inventoinnin tilastollisten menetelmien historiaa ja matemaattista perustaa sekä spatiaalisen tilastotieteen ja Bayesiläisen lähestymistavan tarjoamia uusia soveltamismahdollisuuksia metsien inventoinnissa.

SISÄLLYS

<i>Kuusela, K.</i> Metsävarojen inventointi – kansainvälistä vertailua	213	<i>Salminen, S.</i> Metsien inventointimenetelmän tilastomatemattinen perusta	226
Summary: Inventory of Forest Resources: an International Comparison	218	Summary: The Mathematic-Statistical Foundation of the Forest Inventories	232
<i>Seppälä, R.</i> Metsien inventointi ja otantamenetelmien kehitys	218	<i>Penttinen, A.</i> Metsänarvioimissysteemi spatiaalisen tilastotieteen näkökulmasta	233
Summary: Forest Inventories and the Development of Sampling Methods	219	Summary: Spatial Statistics in Forest Inventories	235
<i>Kilki, P.</i> Metsien inventointi osana metsätietojärjestelmää	220	<i>Pekkonen, T.</i> Suuraluetta koskevan keskimääräistiedon hyödyntäminen metsikkötunnusten estimoinnissa	236
Summary: Forest Inventory as a Part of a Forest Information System	223	Summary: Utilization of Inventory Data in Estimating of Stand Characteristics	240
<i>Roiko-Jokela, P.</i> Metsätilastojen käyttäjän tietotarpeet	224	<i>Mellin, I.</i> Lineaarinen regressiomalli ja prioritiedon käyttö	240
Summary: Forest Statistics and Users' Requirements	226	Summary: Linear Regression Model and the Use of Prior Information	243

Kirjoittajat

Kuusela Kullervo, MMT, professori, Metsäntutkimuslaitos, Metsänarvioimisen tutkimusosasto

Seppälä Risto, VTT, professori, Metsäntutkimuslaitos, Matemaattinen osasto

Kilki Pekka, MMT, apulaisprofessori, Joensuun yliopisto, Metsätieteellinen tiedekunta

Roiko-Jokela Pentti, MMK, Dr.Sc.Techn., metsäneuvos, Metsähallitus

Salminen Sakari, MH, Metsäntutkimuslaitos, Metsänarvioimisen tutkimusosasto

Penttinen Antti, FT, Jyväskylän Yliopisto, Sovelletun matematiikan osasto

Pekkonen Timo, VTL, Metsäntutkimuslaitos, Matemaattinen osasto

Mellin Ilkka, VTL, Helsingin Yliopisto, Tilastotieteen laitos

METSÄVAROJEN INVENTOINTI – KANSAINVÄLISTÄ VERTAILUA

KULLERVO KUUSELA

Otosta käytävien metsän inventointien pääasiallinen tavoite on ollut saada luotettavat ja tuoreet tiedot suuralueiden metsävaroista puuntuotannon suunnittelua ja metsien käytön kontrollia varten. Tähän tarkoitukseen ovat taloussuunnitelmainventointien tulosten yhdistelmät olleet ja ovat yhä liian epätarkkoja sekä osaksi vanhentuneita.

Yhdensuuntaiset kaistat tai yhdensuuntaiset linjat sekä niillä mitatut koealat olivat otosyksiköitä ensimmäisissä Pohjoismaiden inventoinneissa. Nykyiset otosyksiköt ovat lohkoittain maastoon sijoitettuja linja – koealayhdistelmiä.

Useimmissa muissa maissa inventoinnit alkoivat myöhemmin kuin Pohjoismaissa ja niissä käytettiin alusta alkaen ilmavalokuvia ja maastomittauksia. Suomen inventoinnissa on Lapissa käytetty 1970-luvun alusta alkaen kuvatulokinnan ja maastomittauksen kaksivaiheista otosta.

Metsävaroja inventoidaan joko pelkkää maastoarviointia tai ilmakuvatulkinnan ja maastomittausten yhdistelmää käyttäen. Maastoarviointi tehdään pääasiassa kolmella periaatteella erilaisella menetelmällä:

1. Metsiköittäinen ja koko alueen kattava, yleensä mitauksilla tarkistettu silmävarainen arviointi.
2. Jokaisen, tavallisesti jonkin vähimmäisläpimitan täytävän puun rinnankorkeusläpimitan mittaaminen ja koepuiden ottaminen rungon tilavuuden ja puutavara-
ralajien arvioimiseksi.
3. Tilastomatemattinen otos.

Ensimmäistä käytetään arvioitaessa metsävaroja metsälöiden ja hoitoalueiden taloussuunnitelman valmistamista varten. Niiden tuloksia yhdistelemällä saadaan tietoja suurten alueiden metsävaroista. Jälkimmäiseen tarkoitukseen tulokset ovat kuitenkin osittain vanhentuneita eivätkä ne koostu läheskään aina samansisältöisistä lähtötiedoista. Silmävaraisissa arvioissa on suhteellisen suuri systemaattisten virheiden riski. Niiden yhdistelmien tulokset yleensä aliarvioivat puustoa ja sen kasvua.

Kaikkien puiden luku ja mittaus tulevat kustannustensa vuoksi kysymykseen vain poikkeuksellisen arvokkaan metsän olosuhteissa, esim. sovellettaessa keski-eurooppalaista kontrollimenetelmää metsätalouden suunnittelussa. Suomessa sitä käytetään arvioitaessa myytävää puustoa sekä perinnönjoon yhteydessä.

Harhattomia, puuntuotannon suunnittelua varten riittävän tarkkoja ja kohtuullisen hintaisia tuloksia suuralueen metsävaroista saadaan tilastomatemattisella otoksella. Tämä aiheutuu metsän pinta-alan suuruudesta, yksityisten puiden erittäin suuresta lukumäärästä ja suhteellisen vähäisestä arvosta sekä jalan kuljettavasta, usein vaikeasta maastosta.

Otoksen mallin kehittymisen ymmärtämiseksi on tarpeen todeta, että malli on riippuvainen maastosta, kasviyhdyksunnista, kulkuoloista, lähinnä tiestöstä, asianomaisen yhteiskunnan tiedollis-teknisestä tasosta ja myös metsänarvioinnin traditioista. Ihmisellä on taipumus pitäytyä kerran valitsemaansa menetelmään ja menetelmän suuret muutokset tapahtuvat tavallisesti toiminnasta vas-

tuullisen henkilön vaihtuessa.

Otokseen perustuvan metsänarvioinnin suunnittelu ja kokeileminen alkoivat kuluvan vuosisadan alussa Ruotsissa ja vähän myöhemmin Suomessa ja Norjassa. Koska tilastomatematiikkaa ja otoksen teoriaa oli tuolloin selvitetty vielä suhteellisen vähän perustutkimuksena, suunnittelu oli kokeilua tavoitteena löytää menetelmä, jolla saataisiin riittävän tarkkoja ja luotettavia tuloksia tekemällä havaintoja ja mittauksia vain osasta metsää.

Kokeilut aloitettiin kaista-arviointina. Metsän läpi ajettiin yhdensuuntaisia linjoja kompassilla, linjan pituus mitattiin ja linjaan tukeutuen rajoitettiin muutaman metrin levyisiä kaistoja. Kaistalla kasvatavat puut luettiin ja niistä mitattiin rinnankorkeusläpimitta sekä rekisteröitiin puulaji ja mahdollisesti puuluokka sen mukaan, mitä puutavaralajeja runko sisälsi. Luetuista puista otettiin osa koepuiksi, joista tehtiin tarvittavat mittaukset ja havainnot puuston runkotilavuuden, kasvun ja puutavaralajien arvioimiseksi. Otoksen yksiköitä ovat yhdensuuntaiset ja tasaväliset kaistat.

Vaikka kaistoilla luettaviksi ja mitattaviksi tulevat puut käsittävätkin vain murto-osan alueen koko puustosta, niin kaista-arvioinnin kustannukset ovat suhteellisen korkeat. Tilastollisen tarkkuuden kannalta malli on heikko sen vuoksi, että kaistalta joudutaan mittaamaan paljon samanlaista puustoa ja keskenään samanlaisia puita, so. kaistan osien ns. naapurikorrelaatio on suuri. Tästä seuraa, että osalla puustoa on pieni informaatioarvo, ja että joudutaan tekemään mittauksia, jotka parantavat suhteellisen vähän arvioiden tilastollista tarkkuutta.

Työn ja kustannusten vähentämiseksi kaistasta erotettiin määrävällein suorakaiteen muotoisia koealoja ja puita luettiin ja mitattiin vain koealoilta. Nopeasti huomattiin, että ympyrän muotoinen koeala, jonka keskipiste on ennalta määrätty linjalla, on rajoitettavissa vähemmällä työllä ja kustannuksilla kuin suorakaiteen tai neliön muotoinen koeala. Näin syntyi linjoittainen ympyräkoeala-arviointi, joka on ollut useita vuosikymmeniä yleisin menetelmä Pohjoismaissa, ja jota on käytetty myös muualla.

Linjoittaisessa koeala-arvioinnissa sen alkuperäisessä muodossa linjat ovat yhden-suuntaisia, linjaluku tai linjojen etäisyys toisistaan suhteessa alueen kokoon sellainen, et-

tä metsäalueen vaihtelu huomioon ottaen arvioiden tilastollinen tarkkuus on halutun suuruinen, ja koealojen väli linjalla on yleensä lyhyempi kuin linjojen etäisyys toisistaan. Koealojen koko on suhteellisen suuri, 0,05–0,10 ha silloin, kun halutaan yksityistä metsikköä kuvaavia tietoja, tai noin 0,01–0,03 ha silloin, kun halutaan ensisijassa metsikköluokkien keskiarvoja.

Linjalla erotetaan maan ja metsikön tunteen perusteella metsikkökuvioita, jotka ovat puuntuotannon käsittely-yksiköitä. Jokaiselta kuviolta arvioidaan maan ja puuston luokitustunnuksia, metsikön metsänhoidollista laatua, rekisteröidään metsiköiden aikaisempi käsittely sekä ehdotetaan tulevaisuudessa tarpeellisia puuntuotannon toimenpiteitä. Linjan ja metsikkökuvion rajojen leikkauskohdat rekisteröidään ja metsikköluokkien pinta-alan arviot lasketaan kuvioille sattuvien linjanpituuksien perusteella. Metsikköluokkien puustotunnukset saadaan keskiarvoina koealamittauksista.

Kukin linja kuvioineen on otoksen yksikkö. Tilastollista tarkkuutta laskettaessa linja voidaan pätkiä osayksiköiksi. Pienillä alueilla kuten yksityisillä metsälöillä vaihtelun ja tilastollisen tarkkuuden tunnuksia lasketaan myös käsittelemällä koealoja yksikköinä. Esim. puuston keskitilavuuden keskivirhe voidaan arvioida koealojen lukumäärän ja koealametsiköiden runkotilavuuden vaihtelun perusteella.

Suomen metsävarojen inventointeja edeltänyt tärkein koetyö oli Werner Cajanuksen suunnittelema ja toteuttama Sahalahden ja Kuhmalahden metsien arviointi 1912. Yrjö Ilvessalon johtamat neljä ensimmäistä valtakunnan metsien inventointia vuosina 1921–24, 1936–38, 1951–53 ja 1960–63 tehtiin linjoittaisena koeala-arviointina. Lounaasta koilliseen mitattujen linjojen väli oli esim. 3. inventoinnissa pääosassa maata Suomussalmen korkeudelle asti 13 km, sitten 16 km ja kauimpana pohjoisessa 20 km. Varsinaisten ns. täyskoealojen väli linjalla oli 1 km. Niiden puolivälilläin sattuvilla väliskoealoilla mitattiin tukkipuusto.

Kolmessa ensimmäisessä inventoinnissa käytettiin silmävaraista arviointia tavalla, joka on suomalainen kehitelmä, ja jota ei ole käytetty kokeita lukuunottamatta tietyvästi muualla suuralueiden arvioinnissa. Jokaisella linjan leikkaamalla metsikkökuvioilla arvioi-

tiin silmävaraisesti puiden pituuden mittauksia ja aputaulukoita tukena käyttäen puuston runkotilavuus ja kasvu hehtaaria kohti sekä tilavuuden ja kasvun puulajisuhteet. Samat tunnukset arvioitiin silmävaraisesti 0,10 hehtaarin suuruisella täyskoealalla, jossa ne myös mitattiin. Silmävaraisesti arvioitujen ja mitattujen tunnusten regression perusteella laskettiin jokaiselle silmävaraisen arvioinnin suorittajalle korjaustekijä, jolla silmävaraiset arvioinnit korjattiin.

Menetelmä oli sopiva ja halpa erityisesti 1920- ja 1930-luvulla, jolloin teitä oli vähän, auto ei ollut vielä yleinen kulkuväline ja laajoilla alueilla oli täysin tietttömiä metsiä.

Yhdistetyn silmävaraisen arvioinnin ja koeala-arvioinnin heikkous on, että siinä mahdolliset systemaattiset virheet ovat hyvin vaikeasti todettavia ja korjattavia. Arvioimisolosuhteet linjakuvioilla ja koealalla ovat niin erilaiset, että niillä voi olla vaikutus arvioon. Koealalla tehtävään arvioon voi myös vaikuttaa tietoisuus siitä, että tulos tulee mittauksella tarkistetuksi. Näistä syistä koealan silmävaraisen arvioinnin ja mittaustuloksen regressio ei ehkä korjaakaan linjakuvioilla syntyviä systemaattisia virheitä. Valtakunnan metsien 4. inventoinnissa ja sen jälkeen onkin käytetty pelkästään koealamittauksia puuston tunnusten arvioimisessa.

Yhtäjaksoisten linjojen käyttämisestä luovuttiin ensiksi Ruotsissa ja 1960-luvulla myös Suomessa. Tilalle tuli lohkoittainen linja- ja koeala-arviointi. Pääasiallinen syy muutokseen oli yleisten olosuhteiden muuttuminen. Liikkuminen autoistui, tiet lisääntyivät niin yleisessä liikenteessä kuin puutavaran kuljetuksessa ja elintason noustessa vähenivät kyky ja halu raskaaseen metsässä liikkumiseen ja elämiseen. Jos yrittää kuvitella, että valtakunnan metsien inventoinnissa työryhmän pitäisi tällä hetkellä kävellä työtä tehden pitkin kartalle merkittyä linjaa läpi maan, kantaa mukanaan työvälineet ja henkilökohtaiset varusteet, asua yöt joko mukana kannetuissa teltoissa tai hakemalla yösjä ja ruoka linjan varteen sattuvista taloista, tajuaa nopeasti, kuinka vaikeata olisi saada työhön tekijöitä, ja jos heitä saisikin, niin palkat ja muut kustannukset olisivat erittäin korkeat. Yhtäjaksoinen linja-arviointi oli mahdollinen 2. maailmansodan jälkeen ja vielä 1960-luvun alussa. 1980-luvun alussa se on historiaa.

Yhtäjaksoisten linjojen käyttäminen valta-

kunnallisissa metsävarojen arvioinneissa on rajoittunut lähes yksinomaan Pohjoismaihin. Osaksi tämä johtuu siitä, että Pohjoismaissa inventoinnit aloitettiin aikaisemmin kuin muualla ja ennen ilmavalokuvauksen käyttöön tuloa. Pohjoismaiden metsäolot ja haja-asutettu metsä ovat myös poikkeuksellisen edulliset linja-arvioinnin toteuttamisen kannalta. Monissa muissa Euroopan maissa, joissa maasto-olojen perusteella sitä olisi voitu käyttää, metsä- ja puutaloudella ei ole ollut niin suurta kansantaloudellista merkitystä, että niissä olisi tehty otokseen perustuvia inventointeja eikä useimmissa maissa tehdä niitä vielä tälläkään hetkellä. Menetelmän pitkää käyttöä Pohjoismaissa selittää myös traditioon pitäytyminen.

Lohkoittainen linja- ja koeala-arviointi otettiin käyttöön ensin Ruotsissa ja sen jälkeen Suomessa 5. inventoinnissa 1963–1970. Tietyn suuruiselle lohkolle tehtävä linjan ja sille sijoitettujen koealojen muodostama otosyksikkö syntyi aluksi siten, että yhtäjaksoinen linja katkottiin pätkiksi ja niistä tehtiin neliöitä, joiden linjapituudesta ja koealojen lukumäärästä tuli keskimääräisissä metsäolosuhteissa yhden päivän työurakka. Euroopan maista Itävallassa metsät on inventoitu lohkoittaisella koeala-arvioinnilla kaksi kertaa 10 vuoden kierrolla.

Linjaneliöt koealoineen merkitään kartoille, työryhmä asuu kiinteässä majapaikassa, kulkee autolla joko kohtaan, jossa tie leikkaa linjan tai niin lähelle otosyksikköä kuin mahdollista ja kävelee linjalle, tekee arvioimistyön ja palattuaan aloituskohtaan palaa autolle ja majapaikkaan. Suomessa maa jaettiin ruutuihin, joiden koko Etelä-Suomessa oli 8 × 8 km, ja jokaisen ruudun keskusta sijoitettiin linjaneliö koealoineen. Maan pohjoispuoliskossa otosyksikön linjakuvio muutettiin puolineliöksi, jonka linjapituus on hieman lyhyempi kuin neliön, mutta maastoa enemmän peittäessään antaa likimain saman tilastollisen tarkkuuden kuin neliö.

Neliön muotoisen otosyksikön etuna on, että sitä kiertäessään ryhmä tekee koko ajan havaintoja ja mittauksia. Puolineliöllä joudutaan kävelemään enemmän sellaista matkaa, jonka aikana ei tehdä havaintoja eikä mittauksia. Tämän korvaa lyhyempi linja ja pienempi määrä koealoja. Jos puolineliön koko ja koealojen lukumäärä ovat oikeassa suhteessa neliön kokoon ja koealojen lukumää-

rään, niin kummallakin saatu likimain sama tilastollinen tarkkuus selittyy siten, että puolineliön metsiköiden tunnusten naapurikorrelaatio on pienempi kuin sama tunnus neliötä käytettäessä. Ja jos vielä kokonaistyöaika puolineliötä kohti on lyhyempi kuin neliötä kohti, on puolineliö edullisempi kuin neliö otoksen mallina.

Esitetystä on jo ilmennyt, että keskeisin tehtävä metsävarojen inventointia suunniteltaessa on löytää sellainen otoksen malli, jolla saadaan haluttu tilastollinen tarkkuus ja joka samalla minimoi kustannukset. Kustannukset ja otoksen malli ovat riippuvaisia maaston ja metsän vaihtelusta ja niiden kuljettavuudesta, teiden tiheydestä ja laadusta sekä saatavissa olevan työvoiman laadusta. Työn seuranta ja tarkistusmittaukset on myös otettava huomioon, sillä otoksen yksikön on oltava sellainen, että se löytyy ilman kohtuuttomia kustannuksia. Tilastomatemaattisesti edullisimman mallin mittaaminen Suomen olosuhteissa varsinkin suurilla alueilla vaatii niin paljon ihmistyötä ja informaatiota antamatonta kävelyä metsässä, että se ei ole taloudellisesti edullisin.

Yhtäjaksoiseen linja-arviointiin, jossa koealojen väli linjalla on paljon lyhyempi kuin linjojen etäisyys toisistaan, liittyy tilastomatemaattinen heikkous, joka aiheutuu koealametsiköiden suuresta naapurikorrelaatiosta. Toinen puoli samaa asiaa on, että linjojen väliin jää suuri pinta-ala, josta ei saada mitään tietoa. Tilastollinen virhe voi suureta ennalta arvaamattomaksi, jos maaston ja puuston pitkittäiset kuviot sattuvat linjojen suuntaisiksi. Tästä syystä linjojen suunnaksi on Suomessa suositeltu lounaasta koilliseen. Suunta on kohtisuorassa maaston kuvioiden yleisimpään pituussuuntaan. Tämäkään varokeino ei auta niissä tapauksissa, joissa uudistushakkuissa tehdään puustokuvioita, joiden pituussuunta poikkeaa maaston kuvioiden pituussuunnasta.

Tilastollisesti edullisin koealojen verkko on sellainen, jossa koealojen etäisyys toisistaan on yhtä suuri, tai jos ne sijoitetaan maastoon linjoittain, koealaväli ja linjaväli ovat yhtä suuret. Tällaisen mallin mittaamiseen tarvitaan Suomen olosuhteissa niin paljon kävelyä koealalta toiselle, että menetelmä tulee liian kalliiksi ainakin suuralueilla. Mallia on käytetty Keski-Euroopassa eräiksi suuriksi koekeiksi jääneissä inventoinneissa, esim.

1970–71 Baijerissa, missä olosuhteissa sitä puoltavia tekijöitä ovat tarkat kartat ja tiheä tieverkko.

Pohjoismaissa aloitetut valtakunnalliset metsävarojen inventoinnit herättivät ensiksi mielenkiintoa Yhdysvalloissa ja Kanadassa, missä otokseen perustuvia arviointeja alettiin pitää tarpeellisina 1930-luvulla. Linja- ja koeala-arviointi ei osoittautunut sikäläisen luonnon, maaston ja metsien alueellisen jakaantumisen olosuhteissa edulliseksi. Suunniteltaessa ensimmäisiä inventointeja ei 1930-luvulla enää voitu olla huomaamatta ilmalokuvauksen tarjoamia mahdollisuuksia.

Maastosta ja metsästä heijastuvan säteilyn rekisteröinnillä, tehtiinpä se lentokoneesta tai satelliitista, on ylivoimaisena etuna runsas ja monipuolinen informaatio, jonka saamiseen ei tarvita maastossa liikkumisen rasituksia ja kustannuksia. Etu on erityisen suuri laajoilla ja tiettömillä metsäalueilla, joiden metsävaroista on vähän tai ei ollenkaan tietoja. Ilmainformaatio onkin tällä hetkellä olennainen osa useimpia metsävarojen inventointeja maapallolla.

Pohjoismaiden valtakunnallisissa metsävarojen inventoinneissa on käytetty ilmalokuvia vain Suomessa ja meilläkin vasta 1970-luvulta alkaen Lapissa. Tosin ilmakuvia on käytetty kuvioittaisessa metsänarvioinnissa ja karttana 1940-luvulta alkaen, mutta osana otokseen perustuvia menetelmiä ne on sivuutettu tavalla, jota todennäköisesti monissa muissa maissa hämmästellään. Yksi ilmeinen syy tähän on maastoarviointeihin sitoutunut traditio. Toinen reaalin ja tälläkin hetkellä vaikuttava tekijä on järjestelmällisesti otettujen, tasalaatuisten kuvausten puuttuminen. Metsävarojen inventointeja varten tehtäviä omia kuvauksia ei ole katsottu tarpeelliseksi rahoittaa missään Pohjoismaassa. On myös totuttu niin tarkkoihin ja monipuolisiin metsävaratietoihin, että on kaihdettu riskiä heikentää niitä käyttämällä saatavissa olevia ilmakuvia.

Jos halutaan likimääräisarvioita tarkempia tietoja metsävaroista, on menetelmänä ilma-kuva- ja maastoarvioinnin yhdistelmä. Näitä menetelmiä on niin paljon ja niin erilaisia, ettei edes tärkeimpien luetteleminen ole tässä mahdollista. Suomen Lapissa käytetyn menetelmän pääkohtien esittäminen havainnollistaa ilma-kuva- ja maasto-otannan keskeisiä periaatteita.

Menetelmä on kaksivaiheinen otanta. Ensimmäisessä kuvatulkinnan vaiheessa systemaattisesti maaston kattavat stereokuvat muodostavat maastoinventoinnin lohkoja vastaavia yksiköitä. Kuville sijoitetaan verkko kuvatulkinntapisteitä, joiden kohdalla havaitaan ja mitataan kaikkia niitä maaston ja puuston tunnuksia, joista saadaan käyttökel-poista tietoa.

Toisessa vaiheessa mitataan pieneen osaan kuvatulkinntapisteitä koeala maastossa. Kuvatulkinalla saadaan arviot maa- ja metsikköluokkien pinta-aloista ja sen tuloksen perusteella voidaan maastokoealat sijoittaa keskitämällä niitä haluttujen tulosten kannalta tärkeimpiin maa- ja metsikköluokkiin. Kuvatulkinnan tulokset tarkistetaan ja varsinaiset metsän tunnuksat lasketaan maastohavainnoista.

Maastokoealojen sijoittamisessa on käytetty kahta menetelmää. Kuvatulkinntapisteitä ryhmitetään tärkeimpien tunnustensa puolesta keskenään homogeenisiin ryhmiin, joista otetaan maastomittaukseen otos suhteessa ryhmien metsätaloudelliseen merkitykseen ja niin paljon, että arvioille saadaan haluttu tilastollinen tarkkuus. Toinen vaihtoehto, jota käytettiin kuluneena kesänä päättyneessä 7. inventoinnissa, on mitata systemaattinen otos kuvatulkinntapisteistä. Ensimmäisellä vaihtoehdolla päästään paremmin kustannusten minimointiin. Jälkimmäisen etuja on tulosten laskeminen omana itsenäisenä arviona maastomittauksista. Kun näitä arvioita verrataan kuvatulkinnan ja maastomittauksen yhdistelmällä saatuihin tuloksiin, voidaan tarkistaa kuvatulkinnessa mahdollisesti syntyviä systemaattisia virheitä. Toinen etu on mahdollisuus jälkikäteen lisätä kuvatulkinnan pisteitä ja saada näin tarkentuvia tuloksia pienille osa-alueille ja eri omistajien metsille.

Kaksivaiheisen otannan etuja on ensinnäkin, että kuvatulkinnan pisteet ovat halpoja ja kuvatulkinnta voidaan tehdä kaikkina vuodenaikoina. Tiettyä tilastollista tarkkuutta silmällä pitäen kalliiden maastokoealojen tarve on suhteellisen pieni. Näiden syiden takia kaksivaiheisen otannan edullisuus paranee

suhteessa pelkkään maastoarviointiin sitä mukaa kuin ihmistyön kustannus nousee suhteessa ilmakuvien kustannukseen. Menetelmä on tällä hetkellä erillisen tutkimuksen kohteena. Likimääräinen arvio on, että Lapissa käytetyllä menetelmällä saadaan tietty tilastollinen tarkkuus kaksivaiheisella otannalla noin kolmannelle halvemmalla kuin pelkällä maastoarvioinnilla. Tässä ei ole mahdollista tarkastella inventoinnin kustannuksia lähemmin. Kahdeksan vuotta kestänyt 7. inventointi maksaa nykyrahassa noin 1 mk metsämaan hehtaarilta.

Menetelmiä vertailtaessa on tarpeen ottaa huomioon, että ilmakuvia käytettäessä ei saada yhtä luotettavia tuloksia kuin pelkällä maastoarvioinnilla sellaisista metsän tunnuksista, joista kuvatulkinalla ei saada tietoa tai tieto on epäluotettavaa. Esimerkkejä tällaisista tunnuksista on soiden ojituskelpoisuus, ojitusaluiden ja uudistusaluiden metsänhoidollinen tila, jne.

Yhdistämällä korkeakuvaus, matalakuvaus ja maastoarvioinnit saadaan kolmen vaiheen otanta. Vastaavasti voidaan yhdistää satelliitti-informaatio, ilmakuvaus ja maastomittaukset. Näitä menetelmiä käytetään arviointaessa ensimmäistä kertaa taloudellisen käytön ulkopuolella tähän asti olleita luonnonmetsiä. Ilma- ja satelliittikuvien tulkinasta on tulossa yhä tärkeämpi osa myös ympäristön muutosten seurantamenetelmiä.

Metsävarojen seuranta varten on pysyviä koealoja käytetty laajasti Yhdysvalloissa ja Kanadassa yli 40 vuoden ajan. Suomessa niitä on käytetty osakeyhtiöiden metsien inventoinneissa. Paraikaa suunnitellaan perustaa koko maahan pysyviä koealoja, joilla seurataan paitsi puustoa myös muita ympäristömuutoksia.

Metsävarojen inventoinnissa käytetään yleensä systemaattista otantaa. Samaan tilastolliseen tarkkuuteen pyrittäessä oikein suunniteltu ja käytetty systemaattisen otoksen malli on satunnaisotannan mallia halvempi. Systemaattisella otoksella saatujen arvioiden tilastollisen tarkkuuden tunnuksat ovat tosin likiarvoja, mutta riittävän tarkkoja metsätalouden tarpeita varten.

Kirjallisuus

- Ilvessalo, Y. 1927. Suomen metsät. Tulokset vuosina 1921–1924 suoritetusta valtakunnan metsien arvioimisesta. The forests of Suomi (Finland). Results of the general survey of the forests of the country carried out during the years 1921–1924. Commun. Inst. For. Fenn. 11: 1–421+192.
- 1956. Suomen metsät vuosista 1921–24 vuosiin 1951–1953. Kolmeen valtakunnan metsien inventointiin perustuva tutkimus. The forests of Finland from 1921–24 to 1951–53. A survey based on three National Forest Inventories. Commun. Inst. For. Fenn. 47 (1): 1–227.
- Kuusela, K. & Salminen, S. 1969. The 5th National Forest Inventory in Finland. General design, instructions for field work and data processing. Commun. Inst. For. Fenn. 69 (4): 1–72.
- National Forest Inventory. 1978. Papers of the Joint Meeting of IUFRO Groups: S4 02 Forest Resource Inventory, S4 04 Forest Management Planning and Managerial Economics. 18–26 iunie 1978 – București – România. 1–655 pages.

Silva Fenn. 19 (3)

METSIIEN INVENTOINTI JA OTANTAMENETELMIEN KEHITYS

RISTO SEPPÄLÄ

Metsien linja-arviointia voidaan Suomessa, samoin kuin Ruotsissa, pitää laaja-alaisen otannan soveltamisen tiennäyttäjänä. Ruotsin ”metsätalouden isän” kunnianimien saanutta Israel Adolf af Strömiä on pidettävä linja-arvioimisajatuksen luoja. Menetelmän perusteet, jotka kylläkin jonkin verran poikkeavat nykyisestä käytännöstä, hän esitti v. 1830, ja jo samalla vuosikymmenellä linja-arviointia kokeiltiin käytännössä. Suppeilla alueilla linja-arvioinnin käyttö Ruotsissa oli sangen yleistä 1800-luvun jälkipuoliskolla.

Suomessa menetelmää käytettiin ensi kerran v. 1885. Tällöin linjat edustivat viittä prosenttia arvioitavan alueen pinta-alasta. Tulosten tilastollisen täsmällisyyden laskemi-

Summary

Inventary of Forest Resources: an International Comparison

The principal objective of sampling inventories of forests has been to provide reliable and up-to-date information for planning and controlling the utilization of timber resources in large areas.

Parallel strips or parallel lines with plots were the sampling units in the first inventories in Finland, Sweden and Norway. These lay-outs have been developed into the current technique of line and plot units located tractwise in the terrain.

In most of the other countries, where the first inventories were carried out later than in the Nordic countries, the technique has been to combine information from aerial photographs and terrestrial measurements. There are several lay-outs of two-step sampling based on the interpretation of aerial photographs and ground plots. The latest technique employs remote sensing and computerized digital analysis of recorded radiation.

seen ei vielä tällöin ollut kehitetty menetelmiä, vaan saatujen arvojen luotettavuus selvitettiin tekemällä muutamilla alueilla tarkka, yksityiskohtainen puiden luku, jota sitten verrattiin linja-arviointiin. Siitä, minkälaisia tuloksia tämä vertailu antoi, ei ole säilynyt tietoa.

Ensimmäinen pätevästi suunniteltu ja toteutettu linja-arviointi maassamme oli Sahalahden ja Kuhmalahden pitäjien metsien arviointi v. 1912. Työn suunnitteli ja pani alulle Werner Cajanus. Vuonna 1908 tieteellisen uransa aloittanut ja v. 1914 väitöskirjan julkaissut Cajanus toi tilastotieteellisen tradition ja kvantitatiivisen tutkimusperinteen suomalaisen metsänarvioimistieteeseen. Ca-

janus oli tutustunut matemaattis-tilastollisiin menetelmiin oleskellessaan Lundin yliopiston tähtitieteen laitoksella.

Cajanus siirtyi kuitenkin pois tiedemiehen uralta v. 1918 ja kuoli seuraavana vuonna. Hänen tutkimustensa loppuun saattaminen jäi Yrjö Ilvessalon tehtäväksi.

Sahalahden ja Kuhmalahden metsien arviointi muodostui sekä otantamenetelmien soveltamisen että metsänarvioimistieteen kannalta pioneerityöksi Suomessa. Ensimmäisiä kertoja kiinnitettiin tällöin huomiota systemaattisen otannan yhteydessä esiintyvään alueellisesti vierekkäisten havaintojen autokorrelaatioon. Tässä ongelmassa Ilvessalo kääntyi matemaatikko J. W. Lindebergin puoleen ja saamiensa neuvojen sekä alunperin Cajanusen esittämien ajatusten perusteella kehitti graafiseen tasoitukseen pohjautuvan menetelmän systemaattisen muutoksen eliminoimiseksi.

Ilmeisesti se neuvonpito, joka Ilvessalolla oli Lindebergin kanssa Sahalahden ja Kuhmalahden metsien inventoinnin yhteydessä, sai Lindebergin kiinnostumaan linja-arvioinnin tarkkuuteen liittyvistä laskelmista. Kun valtakunnan metsien ensimmäistä arviointia ryhdyttiin v. 1921 toteuttamaan, alkoi Ilvessalon ja Lindebergin yhteistyö uudelleen tavoitteena pätevä keskivirhe-estimaattorin kehittäminen. Pohdiskelujensa tulokset Lindeberg julkaisi v. 1924 ilmestyneessä artikkelissa.

Tämä tutkimus merkitsi huomattavaa edistysaskelta systemaattisen otannan teoriassa. Lindebergin työ viritti myös pohjoismaisen tieteellisen kiistakirjoittelun. Norjalainen Langsaether esitti omassa kirjoituksessaan Lindebergin työtä kohtaan kritiikkiä, johon tämä vastasi v. 1927 uudella artikkelilla. Keskustelu olisi ehkä jatkunut, ellei Lindebergin kuolema v. 1932 olisi tullut väliin.

Vaikka Lindebergin vaikutus ajallisesti oli vain muutama vuosi ja hänen metsänarviointia koskevat oivalluksensa ovat pääosin vain yhden kirjoituksen varassa, tuntuu hänen vaikutuksensa maamme metsävarojen inventointityössä vielä nytkin. Sen yleismaailmallisen johtoaseman, joka Suomella Lindebergin, Ilvessalon ja Cajanusen ansiosta inventointimenetelmien kehittämisessä ja soveltamisessa oli 1920-luvulla, olemme kuitenkin menettäneet. Ruotsissa, Keski-Euroopassa ja varsinkin Yhdysvalloissa on tällä alalla tehty

erityisesti 1960-luvulta lähtien tutkimusta, jolla on ollut merkitystä myös otannan ja tilastotieteen kehitykselle yleensäkin.

Näyttää siltä, että ainakin Suomen kaltaisessa pienessä maassa kehitys etenee hyppäyksittäin. Usein nämä tasomuutokset liittyvät sukupolvien vaihtumiseen. Metsien inventoinnin menetelmäkehityksessä voi 1920-luvun jälkeen nähdä seuraavan hyppäyksen 1960-luvulla, ja tuntuu siltä, että nyt 1980-luvulla on taas tasosiirtymä ylöspäin odotettavissa. Tämä edellyttää kuitenkin, että metsätieteilijät ja tilastotieteilijät pystyvät toimimaan hedelmällisessä yhteistyössä Ilvessalon ja Lindebergin antaman esimerkin mukaisesti. Näin uskon meillä päästävän jälleen kansainvälisen kehityksen kärkeen.

Kirjallisuus

- Lindeberg, J. W. 1924. Über die Berechnung des Mittelfehlers des Resultates einer Liniertaxierung. Acta For. Fenn. 25.
- 1927. Zur Theorie der Liniertaxierung. Acta For. Fenn. 31.

Summary

Forest Inventories and the Development of Sampling Methods

The idea of the strip-survey of forest was first presented by Israel Adolf af Ström in Sweden in 1830. The method was first used in Finland in 1885 but there were no statistical means to estimate the precision of results. The first scientifically competent strip-survey in Finland was initiated by Werner Cajanus in Sahalahti and Kuhmalahdi communes in 1912. The work was completed by Yrjö Ilvessalo.

In conjunction with the Sahalahti and Kuhmalahdi survey, and later connected to the first national forest inventory, Ilvessalo approached J. W. Lindeberg in order to develop valid standard error estimators. As a result, considerable progress was achieved in the theory of systematic sampling. Lindeberg published his work on this topic in two articles in 1924 and 1927.

Although Lindeberg's work dates from the 1920s, it still has an influence on the forest inventory in Finland. Unfortunately, progress has not been very prominent in recent decades, but an increasing collaboration between statisticians and forest researchers will hopefully soon lead to new developments in inventory methods.

METSÄIN INVENTOINTI OSANA METSÄTIETOJÄRJESTELMÄÄ

PEKKA KILKKI

Metsien inventointi on osa metsätietojärjestelmää, jonka tehtävänä on hankkia, hallita ja muokata tietoa metsätaloudellisia päätöksiä varten. Metsän kasvupaikko- ja puita kuvaava havaintomatriisi on sopiva lopputuote mille tahansa metsänmittaustehtävälle. Tutkimuksen tehtävänä on muodostaa mittaustietojen ja metsää kuvaavan havaintomatriisin välille mahdollisimman tehokas laskentajärjestelmä. Tämän tavoitteen saavuttamiseksi on luotava metsätietojen hallintajärjestelmä, josta alkuperäiset aineistot, valmiit mallit ja laskenta-algoritmit ovat helposti saatavissa.

Johdanto

Metsätaloudessa on luonnon eli metsän osuus tuotantoprosessista huomattavasti suurempi kuin lähestulkoon missään muussa laajamittaisessa elinkeinossa. Kun metsä on toisaalta paljon monimuotoisempi ja tuotosyksikköä kohti laaja-alaisempi kuin esimerkiksi pelto, on metsän kasvupaikkoja ja puustoa koskevien tietojen hankintaan kehittynyt jo varhain metsänarvioimistieteenä tunnettu metsätieteen haara. Sen mittaukseen painotunutta osaa kutsutaan metsän mittaukseksi, ja jos ollaan kiinnostuneita nimenomaan metsäalueen mittauksesta, metsän inventoinniksi.

Vaikka metsien laajuus ja vaihtelevuus tekevätkin niistä hankalia mittaushaasteita, ei metsän inventointi ole likikään yhtä vaikeaa kuin esimerkiksi monien eläinpopulaatioiden arviointi. Mittauskohteet pysyvät paikallaan ja erityisesti kasvupaikoissa tapahtuvat muutokset ovat hitaita. Lisäksi pääosa puiden metsätaloudellisesti kiinnostavista tunnuksista saadaan selville verraten yksinkertaisilla ja halvoilla mittauksilla. Tästä syystä on inventointitutkimus voinut keskittyä suurten perusjoukkojen mukanaan tuomien otantaongelmien ja otannasta huolimatta varsin mitattavien numeroaineistojen käsittelyn aiheuttamien ongelmien ratkaisuun. Tämä on jo varhain johtanut tilastotieteen ja sittemmin automaattisen tietojenkäsittelyn menetelmien soveltamiseen.

Tiedon tarpeet

Metsien inventointi on osa metsätietojärjestelmää, jonka tehtävänä on hankkia, hallita ja muokata tietoa metsätaloudellisia päätöksiä varten. Päätöstilanteiden ääripäitä edustavat toisaalta puuston määräärvion tekeminen ja toisaalta metsän tuotanto-ohjelman valinta.

Inventaario

Termi metsän inventointi viittaa inventaarioon, jossa pyritään selvittämään metsäalueen kasvupaikkojen ja puuston määrä sekä niiden jakauma kiinnostaviin ositteisiin. Metsän inventoinnille inventaarion laatimiseksi on tyypillistä se, että ollaan kiinnostuneita vain muutamista kasvupaikoista ja/tai puustoa kuvaavista parametreista. Näiden parametrien estimaateille saatetaan kuitenkin asettaa varsin korkeat luotettavuusvaatimukset. Tyypillinen tapaus on puuerän kauppaa varten tehtävä inventointi, jossa ainoa kiinnostava parametri saattaa olla kokonaiskauppahinta, mutta se pyritään estimoimaan hyvin luotettavasti.

Suunnittelun lähtötiedot

Inventointitulosten toinen käyttöalue on metsätalouden suunnittelun lähtötietojen hankinnassa. Metsätalouden suunnittelun tehtävänä on laatia suunnittelun kohteena olevalle metsäalueelle joukko päätöksentekijää kiinnostavia tuotantovaihtoehtoja, joista päätöksentekijä valitsee itselleen sopivimman. Suunnittelun lähtötietoja varten mestästä joudutaan estimoimaan suuri määrä parametreja. Jo käsilaskennan ajan suunnittelulaskelmissa parametreja saattoi olla useita kymmeniä. Tietokoneita käytettäessä parametrien määrä nousee helposti satoihintuhansiin. Tällöin ei yhden parametrin luotettavuusvaatimus voi olla yhtä suuri kuin vain muutamia parametreja estimoitaessa.

Suomessa nykyisin käytettävässä Metsä-LAskelmassa – MELA – (Siitonen 1983) metsä kuvataan homogeenisina laskentayksikköinä. Laskentayksiköt vastaavat lähinnä metsiköitä ja niille hankitaan inventoinnilla sekä kasvupaikka- että puustotiedot. Puusto kuvataan puujoukolla, jossa jokaisesta puusta tunnetaan hehtaarikohtainen frekvenssi ja puulaji sekä joukko puun ikää, kokoa ja laatua kuvaavia muuttujia. Tavanomaisessa suunnittelutilanteessa laskentayksiköitä on noin 100 ja niistä jokaista kuvataan muutamalla kymmenellä puulla.

Tyypillinen otantaan perustuva koelainventointi ei ole aina paras mahdollinen menetelmä tuotettaessa suunnittelun tarvitsemia lähtötietoja. Koelat ovat usein niin pieniä, etteivät ne kuvaa riittävän luotettavasti metsän tilaa. Tässä suhteessa inventaarion vaatimukset ovat paljon pienemmät; inventaario ei periaatteessa aseta mitään vaatimuksia koelain koolle tai muodolle.

Inventointimenetelmänä on ongelmallisin silmävarainen kuvioittainen arviointi, jolla aluesuunnittelun lähtötiedot meillä pääasiassa hankitaan. Toistaiseksi ei kuitenkaan ole pystytty kehittämään sellaista objektiivisiin mittauksiin ja otantaan perustuvaa menetelmää, jolla silmävaraisen arvioinnin kuviokohtaisia tuloksia olisi kovin olennaisesti pystytty parantamaan, ilman että kustannukset olisivat nousseet kohtuuttomiksi.

Seuruu

Metsien kehityksen seurusta on Suomessa viime aikoina tullut yhä tärkeämpi inventointitehtävä. Tämä johtuu osoittain siitä, että meillä useimmat metsäalueet on jo kertaalleen inventoitu; esimerkiksi valtakunnan metsien inventointeja tehty kuuden vuosikymmenen ajan. Koko maan metsien tila tunnetaan täten verraten hyvin. Sen sijaan eivät tähän saakka käytetyt kertakoealat anna kovinkaan luotettavia tietoja metsässä tapahtuvista muutoksista. Tulevaisuudessa kiinnostavat erityisesti uusien viljelymetsien kehitys ja ilman saasteiden vaikutus metsien kasvuun. Näiden ilmiöiden seuruu edellyttää pysyviä tai puolipysyviä koelajoja.

Toinen merkittävä seuruinventointien käyttötapa liittyy metsätalouden alueelliseen suunnitteluun. Metsätalouden alueellisen suunnittelun laajetessa lähivuosina suurimpaan osaan yksityismetsiä – valtion ja yhtiöiden metsissä alueellinen suunnittelu on jo nykyisin kattava – saadaan ilmeisen suuria säästöjä inventointikustannuksissa ja parannetaan suunnittelun laatua pitämällä inventointitietoja jatkuvasti ajan tasalla. Tekniset mahdollisuudet tähän tarjoaa karttatietojen numeerinen käsittely. Myös satelliittikuvilla tulee olemaan merkittävä osuus metsätietojen seurussa.

Mallien aineistot

Jo silloin kun tavoitteena on metsävarojen inventaario tai suunnittelun lähtötiedot, joudutaan laatimaan malleja sellaisten kasvupaikka-, puu- ja puustotunnusten ennustamiseksi, joiden mittaus kaikista havaintoyksiköistä tulee liian kalliiksi. Näitä malleja on mahdollista hyödyntää myös kyseistä inventointia laajemmin. Erityisesti valtakunnan metsien inventoinnissa kerätystä koepuuaineistosta voidaan laatia moniin muihin tarkoituksiin käyttökelpoisia malleja: läpimitan jakaumamalleja, pituusmalleja, runkokäyrämalleja jne.

Metsätalouden suunnittelussa tehtävissä puuston kehitysvaihtoehtojen simuloinnissa tarvitaan puiden kasvua, syntymää ja kuolemaa kuvaavia malleja. Näitä malleja tuottavat hyvin monet metsäntutkimuksen alat. On

kuitenkin huomattava, että puiden pitkän kiertoajan ja osittain myös puiden vaihtelevien kasvuolojen takia ei puiden kehityksestä – päinvastoin kuin esimerkiksi yksivuotisten viljakasvien kohdalla – saada riittävästi kokeellista tietoa. Tästä syystä joudutaan mallit perustamaan inventoinneissa mitattuun tietoon. Myös tällöin on valtakunnan metsien inventoinnissa mitattu aineisto keskeisessä asemassa.

Mallien laadinnassa tarvitaan tulevaisuudessa yhä enemmän pysyviltä ja puolipysyviltä koealoilta mitattuja aineistoja. Esimerkiksi puiden syntymistä, karsiutumista ja kuolemista kuvaavien mallien laadinta on hyvin ongelmallista, ellei mittauksia toisteta samoista kohteista.

Metsätalouden suunnittelu

Mielenkiintoisimpia ja samalla tärkeimpiä ovat metsätalouden pitkän ajan suunnitteluun liittyvät inventointitehtävät, joiden yhteydessä tulevat kysymykseen likimain kaikki edellä luetellut inventointitulosten käyttötavat. Jotta inventointimenetelmä vastaisi mahdollisimman hyvin metsätalouden suunnittelun tarpeita, on lähtökohdaksi otettava suunnittelun asema metsätalouden ohjausjärjestelmässä. Inventointipainotteisesti voidaan ohjausjärjestelmässä erottaa seuraavat osat:

1. Otanta
2. Mittaus
3. Puuta kuvaavat mallit
 - a priori mallit
 - inventoinnin mittauksiin perustuvat mallit
4. Metsän alkutilanteen kuvaus
5. Puuston kehityksen simulointi
6. Tehokkaiden tuotantovaihtoehtojen haku
7. Parhaan tuotanto-ohjelman valinta
8. Toimenpideohjelman määrittäminen
9. Seuruu

Luettelon kukin otsikko edustaa itsessään monimutkaista systeemiä. Esimerkiksi otannassa voi olla useita vaiheita. Näin ollen on ilmeistä, että inventointimenetelmän valinta on vaikeaa, vaikka erityisesti kohtaan 7 liittyvät subjektiiviset tekijät voitaisiin ratkaista.

Inventointimenetelmän valinta on optimointitehtävä, joka voidaan periaatteessa

ratkaista esimerkiksi matemaattisella optimoinnilla. Tavoitteena voi olla esimerkiksi kustannusfunktion minimointi, kun kiinnostavien parametrien estimaattien maksimivirhevarianssit on annettu, tai painotetun virhevarianssin minimointi annettu inventointibudjetin rajoissa (ks. esim. Titus 1979). Valtakunnan metsien inventointimenetelmän valinnan kriteerinä voisi olla valtakunnallisten puuntuotanto-ohjelmien luotettavuus annettulla suunnittelubudjetilla.

Vaikka inventointimenetelmän valintaa voidaankin edellä kuvatuin tavoin formalisoida, joudutaan metsätalouden suunnittelussa käytettävä inventointimenetelmä ratkaistamaan useimmiten varsin likimääräisin perustein. Optimaalisen inventointimenetelmän hakemista realistisempaa on taata, että inventointi tuottaa kohtuullisen luotettavat estimaatit tarvittaville parametreille ja että menetelmä ei käytännössä johda liian monimutkaiseen mittauksiin ja suuriin mitausvirheisiin.

Ennakkoinformaation käyttö

Tilastotieteessä viime aikoina paljon tutkittu ennakkoinformaation hyväksikäyttö (esim. Pekkonen 1982) on metsien inventoinnissa ollut jo vanhastaan tapana. Esimerkiksi puiden tilavuudet on vain poikkeuksellisesti määritetty pelkästään inventoinnissa mitattujen tietojen perusteella; yleensä on käytetty apuna valmiita tilavuus- ja runkokäyrämalleja. Tämä ei kuitenkaan merkitse sitä, että ennakkoinformaatiota olisi käytetty tehokkaasti. Aivan liian paljon on mitattu tietoja, jotka olisi voitu saada aiemmista inventoinneista. Aina ei myöskään ole osattu yhdistää tehokkaasti ennakkoinformaatiota ja inventoinnissa mitattua tietoa. Yleensä on tyydytty joko – tai ratkaisuu.

Myös ennakkoinformaation käytössä on valtakunnan metsien inventoinnilla kerätyllä tiedolla keskeinen asema. Erityisesti koepuumittauksia tulisi hyödyntää entistä paremmin sekä muissa inventoinneissa että myöhemmissä valtakunnan metsien inventoinneissa. Puista mitattavien muuttujien välinen sekä näiden muuttujien ja puun ympäristömuuttujien yhteisvaihtelu on puun geneettisten ominaisuuksien suhteellisen suuren pysyvyy-

den takia varsin invarianttia. Koska koepuumittauksia voidaan täten hyödyntää yhtä inventointia laajemmin, on mittaukset pyrittävä tekemään nykyistä luotettavammin.

Erityisen tärkeää ennakkoinformaation käyttö on silloin, kun estimoitavana on lukuisia parametreja, kuten on laita suunnittelun lähtötietoja määritettäessä. Esimerkiksi kuvioittaisessa arvioinnissa voidaan ennakkoinformaation käytöllä vähentää pituusmittauksia ja parantaa tulosten luotettavuutta.

Laskentajärjestelmä

Metsän tarkin kuvaustapa on havaintomatriisi, jonka riveinä ovat puut ja sarakkeina kasvupaikkaa kuvaavat muuttujat sekä puun frekvenssi ja kaikki puun kiinnostavat muuttujat. Tämä havaintomatriisi on sopiva lopputuote mille tahansa metsänmittaustehävälle. Kun se on valmiina, on metsää kuvaavien parametrien laskenta yksinkertaista. Havaintomatriisi on lähtökohtana myös puuntuotantolaskelmille, esimerkiksi MELALLE.

Tutkimuksen tehtävänä on muodostaa mittaustietojen ja lopullisen havaintomatriisin välille mahdollisimman tehokas laskentajärjestelmä (Kilkki 1979). Mittausten ja olemassa olevien mallien avulla on metsä ensin kuvattava teoreettisella jakaumalla. Havaintomatriisiin riveiksi tulevat puittaiset havainnot poimitaan jakaumasta parhaaksi katsotulla otantamenetelmällä (ks. esim. Kleijnen 1974).

Havaintomatriisia kuvattaessa on otettava huomioon tietojen käyttötarkoitus ja varottava tiivistämisestä tietoja liian aikaisin summiksi ja keskiarvoiksi. Puuston keskipituus ja pohjapinta-ala saattavat antaa oikean mielikuvan harvennustarpeesta, mutta puuston raha-arvon määrittäminen edellyttää puuston kokojakauman tarkempaa tuntemista. Periaatteena on oltava, että muuttujan jakuma kuvataan sen odotusarvolla vain siinä tapauksessa, että odotusarvoa ei tarvita myöhemmin argumenttina epälineaarissa muunnoksessa.

Jotta laskentajärjestelmän käyttö olisi tehokasta, olisi luotava metsätietojen hallintajärjestelmä, josta alkuperäiset aineistot, valmiit mallit ja laskenta-algoritmit olisivat helposti saatavissa. Tietojärjestelmän runkona ovat valtakunnan metsien inventointitiedot. Tähän tietojärjestelmään tulisi liittää myös alueellisen suunnittelun tuottamat kuviotiedot ja muista tietorekistereistä saatavat tiedot.

Kirjallisuus

- Kilkki, P. 1979. Outline for a data processing system in forest mensuration. *Silva Fenn.* 13(4): 368–384.
- Kleijnen, J. P. C. 1974. *Statistical techniques in simulation.* Marcel Dekker, Inc. New York, 775 s.
- Pekkonen, T. 1982. Leimikon puuston arviointi regressioennustinta käyttäen. *Metsäntutkimuslaitoksen tiedonantoja.* 86.
- 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.
- Titus, J. 1979. Design of multivariate surveys for sampling wildland ecosystems. *University Microfilms International Ann Arbor, Michigan, U.S.A.*

Summary

Forest Inventory as a Part of a Forest Information System

The forest inventory is one part of an information system which acquires, controls, and processes information for forestry decisionmaking. The most important inventories are associated with the long term forestry planning. An observation matrix, in which the forest is described by trees, is the most appropriate result of any forest inventory. The matrix may be employed for calculating growing stock parameters and making growth simulations. Inventory research should develop efficient algorithms between the forest measurements and the observation matrix. To achieve this goal a data base management system for the original data and algorithms is needed.

METSÄTILASTOJEN KÄYTTÄJÄN TIETOTARPEET

PENTTI ROIKO-JOKELA

Perusongelma

Vaikka meillä on metsistämme käytettävissä tietoa enemmän kuin koskaan aikaisemmin on tuleva ja tapahtuva käynyt yhä yllätyksellisemmäksi. Metsiämme riepotellet myrskyt, taimikoiden ja kasvatuseksien hyönteis- ja sienituhot sekä ympäristösaasteiden aiheuttamat happamoitumisongelmat ovat esimerkkejä siitä, miten vaikeaa on ennakoida eri ilmiöiden todennäköisyyttä. Yhä useammin törmäämme juuri tällaisiin ennalta arvaamattomiin ongelmiin täysin odottamattomalta taholta. Mistä tämä johtuu, vaikka informaatiota ja tietoa on kerättyä ja tarjolla. Miksi me kaikki – asiantuntijoista aina tavallisiin ihmisiin – emme aina onnistu todellisuuden hallinnassa? Mistä johtuvat kaikki takaiskut ja epäonnistumiset? Miksi tilanteet liukuvat käsistämme? Onko metsämaailmamme käynyt yhä kompleksisemmäksi vai onko ihminen toiminnallaan sekoittanut syy- ja seuraussuhteiden verkoston vaikeasti selvitettäväksi vyyhdeksi.

Samantekevää miltä kannalta tarkastelemme nykyongelmia kysymyksessä on hyvin usein tiedon puute, joka johtuu tiedonhallinnan vajavaisuudesta. Tietoa on mutta se ei ole tarvittaessa saatavissa tai saavutettavissa. Tiedämme hyvin, että ratkaisut jäävät tällöin tekemättä ja tehtävät suorittamatta, jos välttämätön informaatio ja tieto puuttuvat. Joka tapauksessa ratkaisujen hyvyys ja suoritusten laadullinen aste riippuvat aina käytettävissä olevasta informaatiosta ja detaljitietoa hallitsevan kokonaissysteemin tuntemisesta.

Käytännön työssä toimivan suunnittelijan tehtävä ja tarkastelukulma on usein vastakkainen tutkijaan nähden. Kun tutkija porautuu yksityiskohtia analysoiden tiedon lähteille, suunnittelijan on luotava aina synteesi ja hallittava yksittäistiedon yhteysverkosto kokonaisuuteen. Ilman yhteyksiä saamme helposti väärän kuvan, joka johtaa meidät virheelliseen toimintaan.

Tavoite

Mitä me viime kädessä teemme ja miten toimimme, riippuu mitä suurimmassa määrin siitä, kuinka me tiedostamme ja ymmärrämme todellisuuden. Varmaa on, että tähän päivään saakka emme ole täysin ymmärtäneet esim. metsäekosysteemin kestävyttä. Ihmisen toimintojen kokonaisvaltainen suhde metsäluontoon on myös hämärä. Uutta ja monipuolista tietoa tarvitaan lisää. Tämä merkitsee etenemistä yhä syvemmälle detaljitietoon perinteisin tieteen keinoin. Tutkimustyössä on varauduttava lisääntyvään havaintojen keruun automaatioon, mistä tutkimus- satelliitin välittämä havaintomäärä (data) lienee huippuesimerkki.

Kun kerätyt ja metsästä kerättävissä olevat havainnot ovat nyt ja tulevaisuudessa vain murto-osaltaan mahdollista jalostaa metsätaloutta palvelevaksi informaatioksi ja tiedoksi, on mittausten ja havaintojen keruun lähtökohdaksi asetettava koko metsätietojärjestelmä. Metsällisen tiedon keruuseen – inventointiin – on lähdettävä päätöksenteon tiedontarpeesta eikä tiedon tarjonnasta. Näin havaintojen teko ja keruu on yhdistettävä kiinteäksi osaksi tiedonhallintaa. Tietokanta ja metsätietojärjestelmä ovat hyvä tiedontalioimisväline ja samalla kaukoputkeen verrattavissa oleva työväline suunnittelijoille kokonaisuuden hallitsemiseksi.

Meillä on jo paljon käyttökelpoista detaljitietoa tilastoissa, kuitenkin vähän kokonaisuuden kannalta tärkeitä sidoksia ja yhteyksiä. Tämä väittämä kuvastaa sitä kokonaisuutta, johon toisiinsa sidoksissa olevien ongelmien ratkaisussa olisi pyrittävä. Ratkaisut vaativat yhä laajempia kokonaisselvityksiä kuin mihin tutkimuksessa on tähän mennessä totuttu. Tämä merkitsee sitä, että on pystyttävä hallitsemaan yhä suurempia tietomääriä ja käyttämään niitä informaatiojärjestelmien tietopohjana.

Apua metsätietojärjestelmästä

Metsätietojärjestelmää suunnitellaan parhaillaan Suomessa. Sen tehtävänä on hankkia, hallita ja muokata tietoa metsätaloudellisia päätöksiä varten. Tästä syystä koko järjestelmä olisi luotava päätöksenteon näkökulmasta nähden. Käytännön toiminnassa esiin tulevat päätöksentekotilanteet, käytettävissä oleva ennakkotieto ja tiedonhankintamenetelmät luovat puitteet metsätietojärjestelmälle.

Metsätietojärjestelmä olisi alusta pitäen ymmärrettävä metsällistä ekosysteemiä, metsiä kuvaavaksi informaatiotietojärjestelmäksi tai -järjestelmäksi. Sen tietosysteemin yleisen mallin muodostus ja muutamien tietosysteemien toteuttaminen on parhaillaan vireillä. Edessäpäin ovat vielä tiedonkeruun ja mittausten yhtenäistäminen kaikkien metsätietojärjestelmän sidosryhmien kanssa. Samoin olisi luotava standardit metsätalouden suunnittelun tietojen käsittelyyn. Tämä vasta takaisi yleisen valmiuden tietosysteemien hyväksikäyttöön ja poistaisi turhan päällekkäisyyden.

Uudenlainen tietojenkäsittely antaa tiedonkeruun suunnittelulle uusia piirteitä. Tietovarastojen ja tilastojen kartoitus ja käyttö korostuvat. Kun suunnitteluprosessi alkaa päätöksentekoon liittyvästä tarveselvityksestä, tietojen tarve, sisältö ja käyttö kartoitetaan koko laajuudessaan. Tietovarastoista tarkistetaan, löytyykö haluttu tieto jo olevista tietokannoista, metsätietojärjestelmästä. Samoin voidaan tutkia vaihtoehtoiset tavat hankkia tietoa pitäytyen ensisijaisesti metsätietojärjestelmän käyttöön. Tällä on olennainen merkitys koko suunnitteluprosessin kustannus- ja hyötyselvitykselle. Tiedonkeruuseen lähdetään vain silloin, kun tieto puuttuu metsätietojärjestelmästä tai sen laatu – oikeellisuus, todenmukaisuus, tarkkuus – ei ole tyydyttävä.

Tiedonkeruusysteemin asema on kuitenkin keskeinen metsätietojärjestelmässä. Ilman tietoa järjestelmä on kuin kirjahylly ilman kirjoja. Voimavaroja olisikin valtakunnallisesti yhdistettävä kattavan metsätietojärjestelmän luomiseen. Vain näin menetellen on mahdollista suunnitella ja toimia tarkemmin ja pienipiirteisemmin nykyisilläkin resursseilla.

Asenteet ja pelot

Voimakkaasti laajeneva, tosiaikaan perustuva metsätietojärjestelmä saattaa kaikkine ekologisine ja ekonometrisine malleineen johtaa tämän tehokkaan apuvälineen väärinymmärtämiseen tai jopa väärinkäyttöön. Olenaisiin kohti on metsätietojärjestelmän ja todellisuuden välinen suhde. Tämän ymmärtäen tiedonkeruu ja suunnittelu eivät saa muodostua täysin automaattiseksi tapahtumaksi. Lähtödatan oikeellisuus, todenmukaisuus ja tarkkuus ovat koko järjestelmän perusta ja laadullinen mitta.

Samalla kun meidän on opittava tuntemaan tiedonhallintaan liittyvät mahdollisuudet, meidän on myös tunnettava niiden rajat. Pelko, että automaattinen tiedonkeruu johtaa kaaokseen on olemassa. Askel tavoitellusta ”tietopankista” tietojen kaatopaikalle on lyhyt. Uuden tiedon määrä ja laajuus myös hämmäntää. Tulevaisuuden mahdollisuuksia kaventavat ja kahlehtivat vanhat atk-asetteet. Ne ovat vielä tässä päivässä ja käytännön kentässä tällaiset:

- yhä enemmän oppimista, ilman ymmärrystä,
- yhä enemmän tietoa, mutta vailla viisautta,
- yhä enemmän tutkimusta, joka vain lisää oppimista kertomatta mihin se johtaa.

Kaiken uhalla meidän on kuitenkin yritettävä eteenpäin. Metsätietojärjestelmä tulisi rakentaa käyttäjiä palvelevaksi ei heitä hallitsevaksi. Avaimet ovat systeemianalyysissä ja tiedonhallintajärjestelmissä. Samoin tehtävään ja ongelmaan suuntautuva ratkaisutapa, jota tässäkin seminaarissa toteutetaan, on oikea. Se murtaa tiede- ja ammattikuntien rajoja.

Kehittämisenäkökohtia

Suunnittelun tavoitteeksi on asetettava tavonomaisten tilatavoitteiden lisäksi mahdollisten muutosprosessien hallinta. Tällöin epävarmuuksien, riskien ja mahdollisuuksien etsiminen vain lisää metsätietojärjestelmän tarvetta. Järjestelmä olisi kehitettävä valtakunnan metsien inventoinnin ja metsätilastojen pohjalta koko metsäluonnon ja ympäristön seuranta varten. Sen tulisi palvella aina käy-

tännön toiminnan tasolle ulottuvaa päätöksentekoa ja seurata mittaamalla harjoitetun metsätalouden tuloksellisuutta ja vaikutuksia.

Lieneekö meillä edes muuta tietä kuljettavana, minkä Cotta (1866) aikoinaan jo viitoitti metsänmittaajille:

"Meidän on kerättävä ja yhdistettävä monien kokemukset, monien vuosien ajalta, monilta seuduilta, mitä moninaisemmista olosuhteista voidaksemme johtaa niistä pääsäännöt, luoda perusteet ja oppia niitä muuntelemaan."

Summary

Forest Statistics and Users' Requirements

A forest information system is currently being planned in Finland. Its aim is the acquirement, management and preparation of information for decision-making in forestry. The focal point is the relation between the forest

Silva Fenn. 19(3)

METSIIEN INVENTOINTIMENETELMÄN TILASTOMATEMAATTINEN PERUSTA

SAKARI SALMINEN

Pohjoismaisissa valtakunnan metsien inventoinneissa on maastoaineisto lähes poikkeuksetta kerätty systemaattisena otoksena. Inventointitulosten keskivirheiden laskenta palautuu stokastisten prosessien teoriaan. Prosessin korrelaatiofunktiolla on keskeinen asema tulosten täsmällisyyden mittana johdettaessa. Koska korrelaatiofunktiota ei voida määrittää otoksesta, keskivirheiden estimointiin käytetään kertomiltaan vaaditut ehdot täyttäviä neliömuotoja. Inventointitulokset lasketaan yleensä suhteina. Lopulliset keskivirhearviot saadaan yhdistämällä neliömuodon käyttö tavanomaiseen suhde-estimaatin varianssin estimointiin.

Inventointimenetelmän tehokkuus riippuu merkittävästi lohkon koosta ja muodosta sekä koealojen lukumäärästä lohkolla. Käyttökelpoinen lohkomallin tehokkuuden mitta on koealakeskiarvon varianssi.

information system and reality. The correctness, relevance and accuracy of the initial data is the corner stone of the whole system and the standard of its quality.

As only a fraction of the potential information concerning forests can be employed by forestry the forest information system as a whole is the starting point for the making of observations. Forest inventories should be based on the demand for data for decision-making rather than on data supply considerations. The making of observations has to become a fixed part of data management. The data base and the forest information system are a means of filing data, and at the same time a tool for the planner.

It is necessary that, apart from the ordinary quantitative goals, forest management planning also controls possible processes of change. The search of uncertainties, risks and possibilities only adds to the demand for a forest information system. Decisions require more extensive overall clarifications than has been the case so far. This implies that one has to be able to control larger and larger quantities of information and to use it as a basis for information systems.

Johdanto

Pohjoismaisissa valtakunnan metsien inventoinneissa on erotettavissa selvät menetelmäkausensa. Alussa oli kaistaleinventointi ja yhdistetty linja- ja koealainventointi yhtenäisillä linjoilla. Sen jälkeen yhtenäiset linjat murrettiin lohkoiksi nimetyiksi otosyksiköiksi ja viimeisenä vaiheena on ollut pelkkä koealainventointi lohkoilla. Kansallisia erityispiirteitä on menetelmissä esiintynyt, mutta melkein poikkeuksetta on koko inventointien ajan maastoaineisto kerätty systemaattisena otoksena.

Systemaattisesta otoksesta lasketut arviot ovat metsänarviointien tapaisissa tehtävissä yleensä täsmällisempiä kuin vastaavan kokoisesta rajoittamattomasta satunnaisotoksesta lasketut arviot. Systemaattinen inventointilinjasto ja -lohkosto jakaantuu tasaisesti yli arviointialueen, mikä yksinkertaistaa tulosten laskentaa osa-alueille. Säännöllisen otosverkoston piirtäminen kartoille on helppoa.

Sattuman osuus systemaattisessa otoksessa on rajoitettu eikä sille koskaan pystyttäne kehittämään yleisiä otokseen perustuvia keskivirheen estimointikaavoja. Keskivirheen laskenta onkin ollut koko inventointien suoritusajan selvitysten kohteena. Suomessa alan kuuluisin edustaja on J. W. Lindeberg, joka kehitti laskentamenetelmänsä ensimmäisen inventointimme suoritusaikoihin 1920-luvulla. Viimeiset vuosikymmenet on ruotsalainen Bertil Matérn ollut alan johtava tutkija. Matérn'in työtä on jatkanut Bo Ranneby, jonka kirjoituksiin käsillä oleva tarkastelu lähinnä perustuu.

Keskivirheiden laskennan teoria

Tarkastellaan kaksiulotteisen avaruuden alueessa Q määriteltyä stokastista prosessia $\{z(x); x \in Q\}$. Keskimääräisarvon

$$\bar{z}(Q) = \frac{1}{|Q|} \int_Q z(x) dx$$

estimointia varten havainnoidaan $z(x)$ -arvot systemaattisesti alueelle Q sijoitetuissa pisteissä p_1, p_2, \dots, p_n . Keskimääräisarvon estimaattori on yksinkertainen keskiarvokaava

$$(1) \quad \bar{z}_n = \frac{1}{n} \sum_{i=1}^n z(p_i).$$

Määritellään havaintopisteen p_i varianssi lausekkeen

$$nE[(\bar{z}_n - \bar{z}(Q))^2]$$

raja-arvona, kun otostiheys pinta-alayksikköä kohti pidetään vakiona ja inventointialueen Q annetaan laajentua rajattomasti kaikkiin suuntiin. Merkitään raja-arvoa symbolilla σ_p^2 ja pidetään lauseketta σ_p^2/n arvion (1) täsmällisyyden mittana.

Oletetaan prosessi $\{z(x); x \in Q\}$ stationääriiseksi ja isotrooppiseksi. Oletuksesta seuraa, että

- kaikilla stokastisilla muuttujilla $z(x)$ on sama odotusarvo μ ja varianssi σ^2
- korrelaatio kahden mielivaltaisen stokastisen muuttujan $z(x)$ ja $z(y)$ välillä riippuu vain pisteiden x ja y välisestä etäisyydestä.

Kun $r(\cdot)$ valitaan korrelaatiofunktion merkiksi ja oletetaan havaintopisteiden p_i ($i = 1, 2, \dots$) muodostavan säännöllisen piste-verkon, voidaan σ_p^2 ilmaista kaavalla

$$\sigma_p^2 = \sigma^2 \left\{ \sum_i r(|p_i - p_i|) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r(\sqrt{x^2 + y^2}) dx dy \right\}.$$

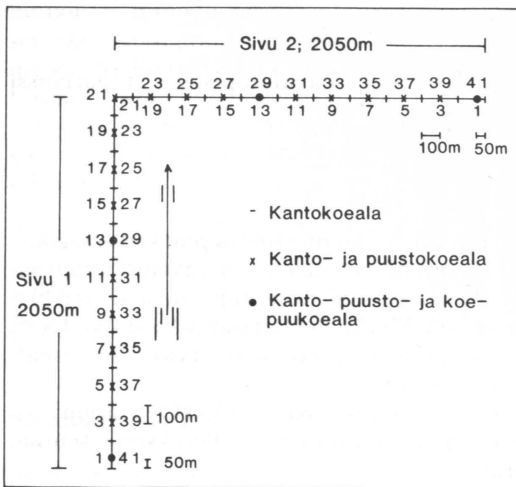
Oikean puoliskon jälkimmäistä termiä ei voida estimoida otoksen perusteella. Tavanomainen käytäntö on arvioida varianssi σ_p^2/n käyttäen neliömuotoa

$$T = \sum_i \sum_j c_{ij} z(p_i) z(p_j),$$

jonka kertoimet täyttävät ehdot

$$c_{ij} = c_{ji}, \sum_i \sum_j c_{ij} = 0 \text{ ja } \sum_i c_{ii} = 1.$$

Kuvatun mallikaavion havaintopistettä vastaa inventoinnissa lohkon koealojen muodostama ryvä, joka malliajattelussa on luonnollisinta paikallistaa lohkon kulmapisteeseen (kuva 1). Inventointialueen epäsäännöllisyys ja vedet ovat esimerkkejä syistä, joiden seurauksena lohkoilla on vaihtelevat määrät relevanttia tietoa sisältäviä koealoja. Tästä syystä inventointitulokset lasketaan yleensä suhde-estimaatteina.



Kuva 1. Valtakunnan metsien 7. inventoinnin lohkon kaaviokuva.

Tarkastellaan satunnaismuuttujaa U/V , kun $E(U) = 0$ ja $E(V) = \mu_v$. Keskineliöpoikkeaman $E(U/V)^2$ laskemiseksi muodostetaan funktion $f(u, v) = (u/v)^2$ Taylor-kehitemä pisteen $(u_0, v_0) = (0, \mu_v)$ ympäristössä. Oteetaan kehittämään mukaan termit, joiden kertoimet muodostuvat funktion ensimmäisistä ja toisista derivaatoista. Kehitemä supistuu muotoon

$$\left(\frac{U}{V}\right)^2 = \frac{1}{2} u^2 \cdot \frac{2}{\mu_v^2} + \text{jäännöstermi.}$$

Siirtymällä odotusarvoihin ja jättämällä jäännöstermin vaikutus huomioon ottamatta päädytään varianssin kaavaan

$$\text{Var} \left(\frac{U}{V}\right) \cong E \left[\left(\frac{U}{V}\right)^2\right] \cong \frac{\text{Var}(U)}{\mu_v^2}$$

Satunnaismuuttujan U/V keskivirhe kirjoituu muotoon

$$(2) \quad \varepsilon \left(\frac{U}{V}\right) \cong \frac{\varepsilon(U)}{\mu_v}$$

Siirtymällä inventoinnin termeihin tarkastellaan valitulla laskenta-alueella suhdetta $\Sigma x_i / \Sigma y_i$, jossa esimerkiksi x_i kuvaa metsämaalle ja y_i yleensä maalle sattuneiden koalakeskipisteiden lukumäärää loholla i . Prosentteiksi muutettuna suhde osoittaa alueen

metsämaasadanneksen. Keskivirheen määrittämiseksi oletetaan, että voidaan kirjoittaa

$$\Sigma y_i = k \Sigma x_i + \eta$$

jossa k on vakio ja η satunnaismuuttuja, jonka odotusarvo $E(\eta) = 0$.

Keskivirhekaavasta (2) saadaan

$$\varepsilon(\Sigma x_i / \Sigma y_i) = \varepsilon\left(\frac{\Sigma x_i}{\Sigma y_i} - k\right) = \varepsilon\left(\frac{\Sigma x_i - k \Sigma y_i}{\Sigma y_i}\right) \cong \frac{\varepsilon(\Sigma(x_i - k y_i))}{E(\Sigma y_i)}$$

Sekä k että $E(\Sigma y_i)$ ovat tuntemattomia. Valitaan niille luonnollisilta tuntuvat arvionsa $\Sigma x_i / \Sigma y_i$ ja Σy_i ja keskivirheen $\varepsilon(\Sigma(x_i - k y_i))$ estimoimiseksi muodostetaan jäännökset

$$(3) \quad z_i = x_i - \frac{\Sigma x_i}{\Sigma y_i} \cdot y_i$$

Varianssi

$$\text{Var} \left(\sum_{i=1}^n z_i\right) \cong n \cdot \sigma_p^2(z),$$

missä $\sigma_p^2(z)$ on z -jäännösprosessin varianssi havaintopistettä kohti. Kun $\sigma_p^2(z)$ estimoidaan neliömuodolla T , saadaan

$$\varepsilon\left(\frac{\Sigma x_i}{\Sigma y_i}\right) \cong \frac{\varepsilon(\Sigma z_i)}{\Sigma y_i} \cong \frac{\sqrt{nT}}{\Sigma y_i}$$

Keskivirheiden estimoinnissa käytetyn neliölohkoryhmittelyn perusajatus Suomen lohkoinventoinneissa on esitetty kuvassa 2. Kun kuvan lohkonumeroissa ensimmäisen numeron annetaan merkitä lohokoryhmää ja toisen lohkoa lohokoryhmän sisällä, saadaan kaavan (3) z_i -arvoille symmetrisiä erotuksia esittävä neliömuoto

$$T_v = \frac{1}{4} (z_{v1} - z_{v2} - z_{v3} + z_{v4})^2$$

Kaaviossa lohokoryhmä edustaa 9 lohkoa. Edustavuuskerroin on lohkojen lukumäärä jaettuna lohokoryhmien määrällä. Kun kerrointa edustaa symboli g , saadaan suhteen $\Sigma x_i / \Sigma y_i$ keskivirhearvioksi

$$\varepsilon\left(\frac{\Sigma x_i}{\Sigma y_i}\right) = \frac{\sqrt{g \Sigma T_v}}{\Sigma y_i}$$

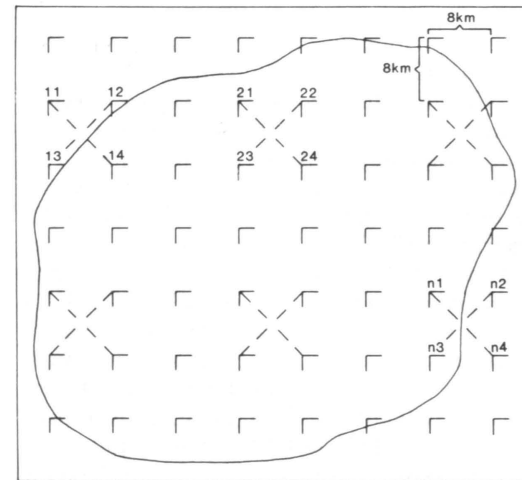
missä Σ on summaus yli käytettyjen lohokoryhmien. Alueen ulkopuolisten lohkojen ja

Taulukko 1. Keskeisiä metsävara-arvioita keskivirheineen (VMI7).

Alue	Maa-ala km ²	Metsämaa			
		Osuus alasta %	Suhteellinen keskivirhe (%)	Keskivilavuus m ² /ha	Suhteellinen keskivirhe (%)
Ahvenanmaan maakunta	1 481	42,9	10,4	129,8	7,1
Satakunnan pml ¹⁾	10 444	63,8	2,5	99,8	2,4
Itä-Savon pml	6 129	84,3	1,2	119,3	2,8
Keski-Suomen pml	15 310	83,0	1,0	99,0	1,9
Etelä-Suomi ²⁾	160 743	71,5	0,5	101,2	0,6

¹⁾ Pml = piirimetsälautakunta.

²⁾ Etelä-Suomi = Pohjois-Pohjanmaan ja Kainuun piirimetsälautakuntien alueiden eteläpuolinen osa maasta.

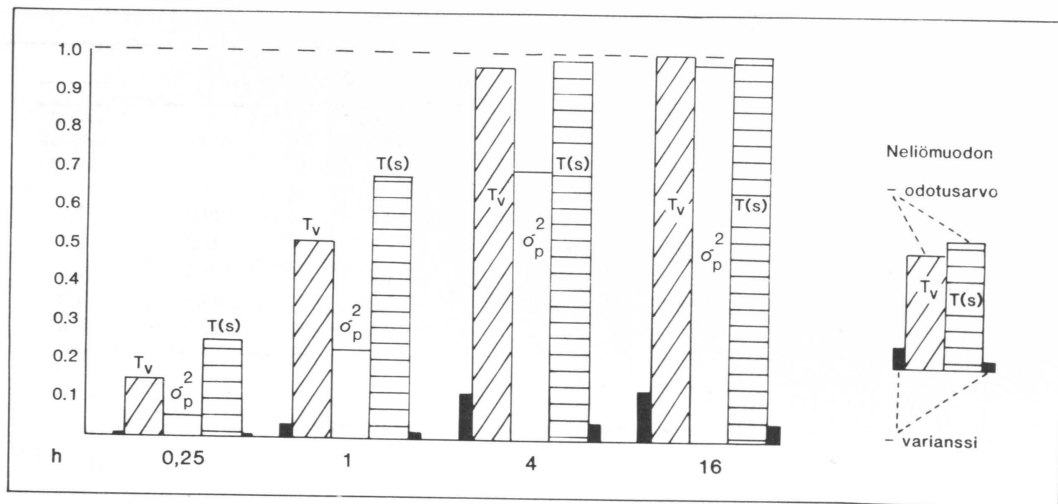


Kuva 2. Neliölohkoryhmien sijoittelun mallikaavio.

lohkonosien koaloille annetaan mitta-arvo = 0. Lohkoryhmää, jossa ainakin yhdellä loholla on nollasta poikkeava y_i -arvo, kutsutaan aktiiviseksi lohokoryhmäksi. Kaavion mukaisesti lohkon ei tarvitse kuulua mihinkään lohokoryhmään. Toisaalta lohko voi kuulua jokaiseen muodostettavissa olevaan neljään lohokoryhmään. Tämä onkin tavanomaisin vaihtoehto inventoinnin keskivirheen estimoinnissa. Taulukkoon 1 on koottu keskeisiä metsävara-arvioita keskivirheineen valtakunnan metsien 7. inventoinnista (VMI7).

Jos otosverkko ei ole niin säännöllinen kuin Suomen lohkoinventointialueella, voidaan inventointialue jakaa ositteisiin ja laskea keskivirhearviot stratifioitua otantaa käyttäen. Neliölohkoryhmittelyn ja stratifioidun otannan vertailuesimerkki on esitetty kuvassa 3. Esimerkissä prosessin satunnaismuuttujat ovat normaalisti jakautuneita ja lohokoryhmitys sisältää 16 yhteispisteetöntä lohokoryhmää. Lohkoryhmän edustavuuskerroin g on 4. Stratifioidun otannan ositteet on oletettu lohokoryhmien muodostamia neliöitä vastaaviksi.

Inventoinnissa käytetyllä neliölohkoryhmityksellä saadaan yleensä käyttökelpoisia tuloksia (kuva 3). Neliömuodon T_v odotusarvo esimerkiksi stratifioituun otantaan pohjautuvaan neliömuotoon verrattuna on alhainen, mutta kuitenkin yläarvio lohkoa kohti lasketulle varianssille σ_p^2 . Neliölohkomuotoa käytettäessä ei ole pelkoa systemaattisesta keskivirheen aliarvionnista. Kuitenkin tällä neliömuodolla on suurehko varianssi. Suurilla alueilla tämä ei ole oleellinen haitta. Pienillä alueilla, joilla lohokoryhmien määrä on vähäinen, on varovaisuussyistä edullista ottaa avuksi likimain vastaavat vaihteluominaisuudet omaava suuremmalle alueelle estimoitu keskivirhe. Jos ostostiheys on alueilla sama, saadaan pienemmän alueen keskivirhe yksinkertaisesti kertomalla suuremman alueen keskivirhe alueiden pinta-alasuhteen neliöjuurella.



Kuva 3. Neliölohkoryhmittelyyn (T_v) ja stratifioituun otantaan [$T(s)$] pohjautuvien neliömuotojen vertailu keskenään ja lohkoa kohti laskettuun varianssiin σ_p^2 . Korrelaatiofunktio on $r(|v|) = e^{-h|v|}$.

Aivan harkinnatta ei erilaisia neliömuotoja voida käyttää. Huolimaton käyttö voi johtaa keskivirheen aliarviointiin. Jos esimerkiksi ojitetulla suoalueella stratifioidussa otannassa ositteet valitaan ojen suuntaisiksi kapeiksi vyöhykkeiksi, päädytään todennäköisesti keskivirheen aliarviointiin.

Inventoinnin keskivirhetarkasteluissa prosessi $z(x)$ on oletettu isotrooppiseksi. Lohkoinventointien aineistoilla ei toistaiseksi ole tarkistettu tämän oletuksen todenperäisyyttä. Lohkot ovat kuitenkin niin suuria ja etäällä toisistaan, että suurta vaikutusta korrelaation mahdollisella vaihtelulla eri ilmansuuntiin ei tuloksiin ole. Ruotsissa on todettu, että esimerkiksi jokilaaksojen suunnassa lohkojen välinen korrelaatio ei merkittävästi poikkea muiden suuntien korrelaatioista. Meillä perinteisesti ajatellaan lounais-koillisuuntainen korrelaatio muita vähäisemmäksi. Tämä on peruste lohkon sivujen ja neliölohkoryhmän kattaman alueen sivujen pääilmansuuntien mukaiselle kululle.

Lohkon muoto ja koko

Keskivirheen laskentaa on edellä tarkasteltu vain lohkojen välisen varianssin valossa. Mitään huomiota ei ole kiinnitetty tiedon

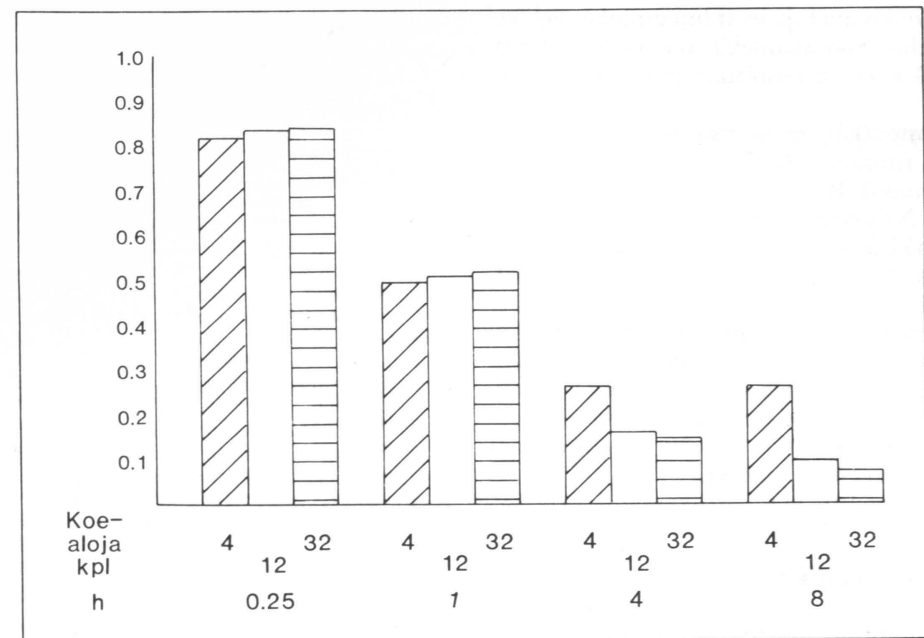
keruuseen lohkolla. Inventointilohkon tulisi keskimäärin olla päivän työtehtävä. Tämän rajaehdon vallitessa on löydettävä järkevä lohkon koko ja muoto sekä koealojen lukumäärä lohkolla. Lohkojen välinen etäisyys on aina niin suuri, että lohkon sijainnin vaikutus voidaan jättää huomiotta lohkotyyppiä määrättäessä. Kun lohkon muoto pidetään samana, on koealojen lohkokeskiarvon varianssi sellaisenaan sopiva vertailumitta. Kun lohkon koko ja muoto vaihtelee, on tarkastelussa lisäksi huomioitava erilainen linjanmittaukseen tai muuhun kävelyyhin kuuluva aika.

Tarkastellaan edelleen isotrooppista prosessia $\{z(x); x \in Q\}$ ja lohkolinjalle sijoitettuja koealakeskipisteitä $p_i, i = 1, 2, \dots, n$. Olkoon $c(|v|)$ prosessin kovarianssifunktio ja $V_{21} z(x) = \sigma^2$. Koealahavaintojen keskiarvolle \bar{z} saadaan varianssi

$$\text{Var}(\bar{z}) = \frac{1}{n^2} \sum_j \sum_j c(|p_i - p_j|).$$

Kun koealahavainnot ovat toisistaan riippumattomia, saa varianssi minimiarvonsa $\frac{\sigma^2}{n}$. Jos havaintojen välillä vallitsee vahva korrelaatio, varianssi on minimiarvoaan huomattavasti suurempi (kuva 4).

Kuva osoittaa, että voimakkaan korrelaation tapauksessa koealojen lukumäärän lisäys jopa kasvattaa varianssia. Haitta on vältettävä



Kuva 4. Koealakeskiarvon varianssi kovarianssin ja koealojen lukumäärän funktiona neliölohkolla, jonka sivun pituus on 1 km. Kovarianssifunktio on $c(|v|) = e^{-h|v|}$.

Taulukko 2. Kahden erilaisen neliölohkon koealakeskiarvon varianssi. Kovarianssifunktio on $c(|v|) = e^{-h|v|}$.

Lohkomalli	h	0,25	1	4	8
Lohkon sivu 1 km					
Koealoja 16 kpl		0,8338	0,5096	0,1502	0,0842
Lohkon sivu 3 km					
Koealoja 8 kpl		0,5887	0,2147	0,1256	0,1250

vissä laskemalla keskiarvo painotettuna keskiarvona

$$z_a = \sum_j a_j z(p_j).$$

Etu painotuksesta on vähäinen eikä sitä käytännön laskelmissa ole sovellettu.

Taulukossa 2 on esitetty kahden erilaisen neliölohkon koealahavaintojen keskiarvon varianssit kovarianssifunktion muuttuessa. Lohkon koon ja koealojen keskinäisen etäisyyden suhteen poikkeavat taulukon 2 lohkot selvästi toisistaan. Tästä huolimatta erilainen korrelaatio johtaa näilläkin lohkomalleilla hyvin erilaisiin ja jopa vastakkaisiin varians-

suhteisiin. Inventoinnissa havaitaan monia muuttujia, joilla kaikilla on erilainen kovarianssifunktio. Lisäksi linjan mittauksen vaatima työaika näillä lohkolinjamalleilla on erilainen. Esimerkki ilmentää sattuvasti lohkomallin optimointitehtävän monitahoisuuden.

Mitatuista inventointihavainnoista voidaan estimoida korrelaatiot eri muuttujille ja haluttaessa tasoittaa ne muotoon

$$\hat{r}(|v|) = ae^{-h_1|v|} + be^{-h_2|v|}$$

Korrelogrammit oletetaan positiivisiksi, laskeviksi ja konvekseiksi. Lisäksi niiltä oletetaan ehto $\hat{r}(0) = 1$. Käytännössä ajateltavissa

olevia lohkomuotoja ovat linjanpätkä, neliö ja puolinelio. Suorakaidekin on puolustettavissa, jos korrelaatio osoittautuu selvästi elliptiseksi.

Linjanpätkät ovat osoittautuneet vaatimansa runsaan ylimääräisen kävelyn takia tehottomiksi. Ruotsissa on päädytty neliölohkoihin. Neliöiden koko vaihtelee, mutta koealojen määrä pysyy lähes koko valtakunnassa samana. Koealojen väli vaihtelee lohkolinjalla välillä 200 – 600 m. Suomessa on päädytty puolinelioihin. Koealojen väli on 200 m. Kevennettyä pinta-alahavainnointia tehdään kuitenkin 100 metrin välein. Jos tätä kevennettyä havainnointia ei tehtäisi, olisivat pinta-alakeskivirhearviot 5 – 30 % suuremmat kuin nykyiset keskivirheet.

Keskivirheiden käyttö

Inventoinnin suorittajille keskivirhetietojen pääkäyttöalueet liittyvät uuden inventoinnin suunnitteluun sekä tarkasteluun, minkä kokoisille alueille erilaisia inventointitietoja voidaan laskea. Peräkkäisten inventointien keskivirhearviot testaavat havaittujen tuloseröjen merkitsevyyden. Käyttäjien mielenkiinto kohdistuu tulosten täsmällisyyssarvioihin yllättävän harvoin. Keskivirhearvioiden suuruudella ei myöskään ole oleellista painoarvoa keskusteluissa, jotka koskettelevat inventointitulosten uskottavuutta.

Kirjallisuus

- Kuusela, K. & Salminen, S. 1980. Ahvenanmaan maakunnan ja maan yhdeksän eteläisimmän piirimetsälautakunnan alueen metsävarat 1977–1979. Summary: Forest resources in the province of Ahvenanmaa and the nine northernmost Forestry Board Districts in Finland 1977–1979. *Folia For.* 446: 1–90.
- 1983. Metsävarat Etelä-Suomen kuuden pohjoisimman piirimetsälautakunnan alueella 1979–1982 sekä koko Etelä-Suomessa 1977–1982. Summary: Forest resources in the six northernmost Forestry Board Districts of South Finland, 1979–1982, and in the whole of South Finland, 1977–1982. *Folia For.* 568: 1–79.

- Lindeberg, J. W. 1926. Zur Theorie der Linientaxierung. *Acta For. Fenn.* 31 (6): 1–9.
- Matérn, B. 1947. Metoder att uppskatta noggrannheten vid linje- och provytetaxering. Resumé: Methods of estimating the accuracy of line and sample plot surveys. *Medd. från Statens Skogsforskn. Inst.* 36 (1): 1–138.
- 1960. Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Sammanfattning: Stokastiska modeller och deras tillämpning på några problem i skogstaxering och andra sampling undersökningar. Medd. från Statens Skogsforskn. Inst.* 49 (5): 1–144.
- Ranneby, B. 1979. Model studies of tract sizes in forest survey. *Forest Resource Inventories. Workshop Proceeding. Colorado State University; Fort Collins. July 23–26, 1979. Vol. I: 289–297.*
- 1981. Medelfelsformer till skattningar baserade på material från den 5:e rikskogstaxeringen. Sveriges lantbruksuniversitet. Avd. för skoglig biometri. *Rapport 21: 1–19.*
- 1981. Elliptisk korrelation. Sveriges lantbruksuniversitet. *Inst. för Skogstaxering. NYTAX 83; Rapport 6: 1–11.*

Summary

The Mathematic-Statistical Foundation of the Forest Inventories

In the Nordic countries, the field materials for the national forest inventories have usually been collected as a systematic sample. The calculation of the standard deviation for the inventory results is based on the theory of stochastic processes. The calculation method has been devised by B. Matérn and further developed by B. Ranneby.

The correlation function of the process is of primary importance for the derivation of the precision of the results. As the correlation function cannot be computed from the sample, some quadratic forms satisfying the conditions set by the coefficients are used for estimating the standard deviations. Inventory results are generally calculated as ratios. The final standard deviation estimates are obtained by combining the quadratic form method and the estimation of variance of the usual ratio-estimate.

The efficiency of an inventory method significantly depends upon the tract's size and form, and upon the numbers of sample plots in the tract. A practicable efficiency measure for the tract model is the variance of mean of the sample plots.

METSÄNARVIOINTISYSTEEMI SPATIAALISEN TILASTOTIETEEN NÄKÖKULMASTA

ANTTI PENTTINEN

Johdanto

Valtakunnan metsien inventoinnissa on kyse ensisijaisesti kokonaismäärän tai suhteen kuten keskitilavuuden arvioimisesta suuralueella. Spatiaalinen (tilassa tapahtuva) vaihtelu ei ole varsinaisesti inventoinnin kohde. Kuitenkin se on tekijä, joka vaikuttaa tulosten tarkkuuteen ja jota tarvitaan tarkkuuden arvioinnissa sekä otantakehikon valinnassa.

Tilastotieteen haaraa, joka tutkii spatiaalista vaihtelua ja sen huomioonottamista tilastollisissa analyyseissä sanotaan spatiaalseksi tilastotieteeksi. Metsäntutkimuksen tarpeet ovat osaltaan vaikuttaneet alan kehitykseen, mm. spatiaalisen analyysin ensimmäisiä monografeja on Bertil Matérnin väitöskirja (Matérn, 1960), johon perustuu nykyisen metsäinventointijärjestelmämme otantamenetelmä. Spatiaalisen vaihtelun tutkimus on laajentunut 70-luvun aikana, jolloin myös spatiaalisen tilastotieteen käsite on vakiintunut. Syitä tälle kehitykselle ovat mm. käytännön ongelmien synnyttämät tutkimustarpeet sekä alan tarvitsemien ATK-valmiuksien lisääntyminen.

Spatiaalisen vaihtelun huomioonottaminen

Kuviossa 1 on esitetty joitakin spatiaalisten muuttujien synnyttämiä kuvioita, jotka lienevät tuttuja metsäntutkijoille. Tällaisten metsäkuvioiden tutkimisessa voidaan käyttää stokastisia malleja (stokastisia prosesseja). Matérn (1960) käsittelee kuviotyypeille a) – d) soveltuvia malleja (kuvio 1).

Spatiaaliseen (auto)korrelaatiofunktioon tiivistyy ns. toisen kertaluvun piirre spatiaalisesta vaihtelusta. Miksi se on riittävä kuvaus inventoinnin tarpeisiin? Jos tutkittavan

muuttujan z arvoa karttapisteessä x merkitään $z(x)$:llä, niin tyypillisesti mielenkiinto on alueen Q keskiarvossa, ts. tyyppiä

$$\frac{1}{\text{ala}(Q)} \int_Q z(x) dx$$

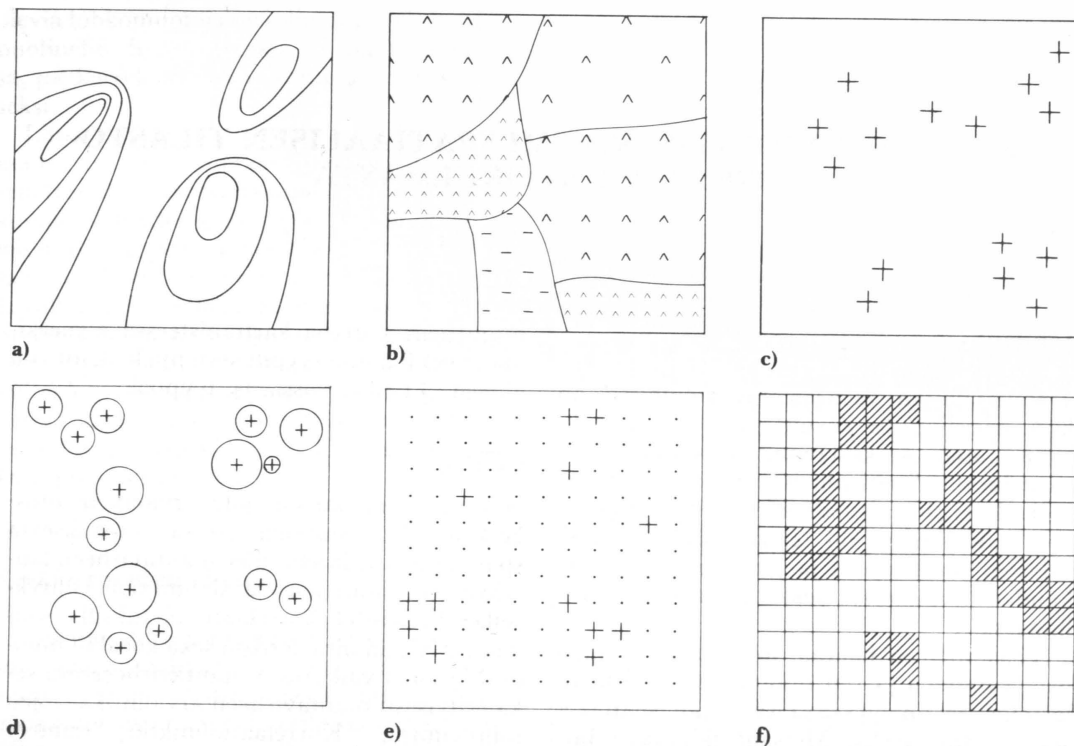
olevassa integraalissa, jota arvioidaan otoskeskiarvolla. Stationaarissa tapauksessa spatiaalinen vaihtelu tulee otantavirheen lausekkeeseen juuri korrelaatiofunktion välityksellä. Erilaisilla korrelaatio-oletuksilla voidaan selvittää mm. lohkon sekä koealan muodon ja koon vaikutusta otantavirheeseen, sekä erityisesti otantavirheen arviointikaavojen toimivuutta. Korrelaatiofunktio esiintyy myös painokertoimien välityksellä joissakin interpolointitehtävissä.

Spatiaalinen korrelaatiofunktio ei ole yksistään riittävä kuvaus spatiaaliselle vaihtelulle, jos tehtävänä on esimerkiksi kuvioiden simulointi. Silloin ovat lisäoletukset tarpeen.

Otantamenetelmään liittyviä kommentteja

Inventoinnissa sovellettava otantamenetelmä, ns. systemaattinen ryvästötanta, on pitkän kehitystyön tulos. Ryvästämällä (lohkojen muodostamisella) pyritään kustannussäästöön ja lohkojen systemaattisella sijoittamisella sekä käytännöllisyyteen että tarkkuuteen. Menetelmä on suuralueiden inventointimenetelmä, jossa ei painoteta muutoksen arviointia. Aika-aspektin korostaminen edellyttäisi pysyvien mitta-alojen mukaantuumista (vrt. Cruse ym., 1984).

Otantavirheen arvioinnin tarkastelun Salminen demonstroi suhde-estimoinnin yhteydessä. Ratkaisun yleisyyttä ja käyttökelpoisuutta kuvaa se, että kolme näennäisesti eri tilannetta palautuu samaan tarkasteluun. Ensinnäkin, z -muuttujasta halutaan esti-



Kuvio 1. Spatiaalisen muuttujan synnyttämiä kuvioita: a) jatkuva vaihtelu (korkeuskäyriä); b) mosaiikki (metsäkuviota); c) pistekuvio (puiden suhteelliset sijainnit); d) merkattu pistekuvio (puiden sijainnit ja peittävydet); e) lattiisikuvio (taimikko, jossa ristit merkkavat kuolleita taimia); f) lattiisimosaiikki (metsämaa satelliittikuvassa).

maatti (riittävän suuresta) osa-alueesta Q_0 . Tällöin on estimoitava muuttujien z : l_{Q_0} ja l_{Q_0} suhde, missä l_{Q_0} on alueen Q_0 indikaattori-funktio (ks. Matérn 1960, s. 120). Toiseksi, kohdemuuttujan z vaihtelu on suurempi kuin suhteen z/y vaihtelu (lohkoittain), missä y on apumuuttuja (esim. metsätyyppin karakterisointi). Jos y -muuttujan globaali keskiarvo tunnetaan, voidaan z -muuttujan keskiarvo määrätä suhteen avulla. Kolmas tapaus on ilmeinen: halutaan arvio kahden muuttujan z ja y suhteesta.

Neliölohkoryhmää esitetään käytettäväksi otantavirheen estimoinnissa. Todettakoon, että z -prosessin asymptoottisen, otospistettä kohti lasketun varianssin estimaattorilla

$$T_1 = \frac{1}{4} (z_{11} - z_{12} - z_{13} + z_{14})^2$$

on seuraavat ominaisuudet: Ensinnäkin, se on riippumaton keskiarvosta. Toiseksi, se on

harhaton (keskimäärin oikea) siinä tapauksessa, että z -prosessi on spatiaalisesti korreloimaton. On huomattava, että z -prosessi on joko havaittava muuttuja tai suhde-estimoinnin yhteydessä jäännösprosessi (selittämättä jäänyt vaihtelu). Spatiaalisen korrelaation läsnäollessa, mikä on reaalitilanne, voidaan ainoastaan arvioida suorituskykyä. Positiivisen korrelaation tapauksessa se osoittautuu varsin yleisesti konservatiiviseksi.

Inventointisysteemin suunnittelussa on oletettu eri ilmiansuunnat samanarvoiseksi (ns. isotrooppisuusoletus). Teoreettisesti ei ole ongelmallista laajentaa tarkastelua ns. elliptiseen tapaukseen, mikä johtaisi lohkojen toisenlaiseen sijoitteluun. Jos luode-kaakko-suuntaisen korrelaation oletetaan olevan muita suuntia voimakkaamman, niin pohjois-etelä-suunnassa olevan neliöhilan voidaan katsoa olevan robusti ratkaisu lohkojen sijoittamisessa, mikä ratkaisu vähentää ei-

isotrooppisuuden vaikutusta (vrt. Cruse ym. 1984, s. 23).

Tihennetyn havainnoinnin käyttöönotto lohkojen sisällä ei ole välttämättä tarkkuutta lisäävä keino, mikäli spatiaalinen autokorrelaatio on voimakas ja positiivinen kuten metsätyyppiä karakterisoiivien muuttujien tapauksessa. Lohkon sisäistä keskiarvoa laskettaessa painotuksella voidaan keskiarvon keskivirhettä tosin pienentää, mutta yksinkertaisempi vaihtoehto on valita näytealat siten, että ne sisältävät reaalista informaatiota. Ruotsin kokemukset eroavat lisähavainnoinnin suhteen, ks. Matérn ja Ranney (1983), s. 5.

Metsänhoitotoimien aiheuttama spatiaalinen vaihtelu

Metsän inventointimenetelmän kehityksen aikana on metsää kuvaaviin muuttujiin tullut uusi vaihtelunlähde: ihmisen toiminta kohdistuu vaihtelevan kokosiin ja muotoisiin osa-alueisiin kokonaisuutena. Tämä muuttaa spatiaalisen vaihtelun luonnetta siten, että mosaiikkimaisuus korostuu entisestään. Osa-alueet, metsäkuviot, ovat sisäisesti homogeenisiä, mutta kuvioiden välinen vaihtelu voi olla suurta (vrt. kuvio 1, tyyppi b).

Millä lailla inventointisysteemi käyttää hyväksi informaatiota mosaiikkimaisuudesta? Lohkoarviointi ja siihen liittyvät otantavirheen arviot nojautuvat positiiviseen, etäisyyden mukaan vähenevään korralaatiofunktion, mikä lienee tilanne mosaiikkimaisessa vaihtelussa. Lisäksi pinta-alapohjainen estimointi ottaa jossain määrin huomioon mosaiikkimaisuutta.

Pidemmällekin voidaan mennä. Metsäkuvioiden mosaiikki sisältää informaatiota, joka on mahdollista havaita ilmasta (avaruudesta) käsin. Metsäkuviosta voidaan muodostaa karkeita karttoja esimerkiksi kaukokartoituksen avulla. Jos tätä lisäinformaatiota käytetään inventoinnissa, joudutaan kaksivaiheiseen havainnointiin. Ensimmäisessä vaiheessa hankitaan karkea kuva metsäkuviosta (spatiaalisesta vaihtelusta joidenkin relevanttien muuttujien suhteen). Toisessa vaiheessa voidaan edetä kahdella tavalla: joko käyttää ensimmäisen vaiheen tietoa näytealojen sijoittelussa (ositettu otanta) tai kytkeä kauko-

kartoituksella tuotetut karkeat apumuuttujat estimointiin (regressioestimointi). Jälkimmäisessä vaihtoehdossa maastohavainnot voidaan tulkita ensimmäisen vaiheen tiedon kalibroinniksi (vrt. Matérn ja Ranney, 1983).

Edellä kuvattu kaksivaiheinen arviointi ei ole pelkästään tekninen prosessi. Se tuottaa myös uusia spatiaalisen analyysin ongelmia ja on haaste tilastollisten menetelmien kehitystyölle. Esimerkkinä mainittakoon autokorrelaation huomioonottavien erottelu- ja luokittelumenetelmien kehittäminen, joita menetelmiä tarvitaan apumuuttujista tehtävien karttojen laatimisessa. Tämä ala on voimakkaassa kehitysvaiheessa.

Kirjallisuus

- Cruse, T., Hägglund, B., Jonasson, H., Ranney, & Svärd, J. 1984. Designing a new national forest survey for Sweden. – Käsikirjoitus.
 Matérn, B. 1960. Spatial variation. Medd. statens skogsforskn. inst. Bd. 49 (5).
 — & Ranney, B. 1983. Structure of the temporal and spatial variation in forests. Implications for the choice of sampling designs. – Esitelmä Hyytiälässä 6. 9. 1983.

Summary

Spatial Statistics in Forest Inventories

Spatial variation is not the main target in forest inventories. However, it can not be by-passed because it has an important role in the choice of sampling schemes and in error estimation. From the statisticians point of view the statistical principles of forest inventories belong to the field of spatial statistics.

In this paper brief notes are made concerning the role of spatial autocorrelation in forest inventories in general, the uses of ratio estimation, the estimation of sampling errors, isotropy assumptions and extra observations within the tracks.

Finally, the Finnish forests show a tendency towards random tessellations (mosaics), partially caused by the modern forest harvesting and planting practices. How does the present sampling plan take this into consideration? This problem and further possibilities are briefly discussed.

SUURALUETTA KOSKEVAN KESKIMÄÄRÄISTIEDON HYÖDYNTÄMINEN METSIKKÖTUNNUSTEN ESTIMOINNISSA

TIMO PEKKONEN

Metsien inventoinnissa törmätään usein tilanteeseen, jossa halutaan arvioida tietyn alueen puustotiedot, mutta käytettävissä on vain vähän otostietoa kyseiseltä alueelta. Normaali tapa ongelman ratkaisemiseksi on ollut käyttää tarkasteltavalle alueelle laajemman alueen tunnuksia, jotka on kyetty estimoimaan riittäväällä tarkkuudella tai yleisiä esimerkiksi valtakunnallisia keskiarvotunnuksia. Tällöin alueen oma otostieto jää lähes kokonaan hyödyntämättä.

Esitelmässä esitetään menetelmä, jossa otostieto yhdistetään otoksen ulkopuolisen tiedon ns. ennakkoinformaation kanssa, jolloin voidaan pienentää estimaattoreiden otantavarianssia. Menetelmä soveltuu erityisesti pieniin otoksiin, joista laskettavien estimaattoreiden keskivirheet saattavat olla suuria ennakkoinformaation sisältämään harhaan verrattuna.

Otantateoriasta

Tarkastellaan aluksi erilaisia estimaattoreiden luotettavuusmittoja. Koska kyseessä on päättely otoksesta perusjoukkoon, sovellettavaksi tulee otantateoria. Normaalisti otantateoriassa tarkastellaan perusjoukkoa kiinteänä ja samoin sen tunnuslukuja kiinteinä ei-satunnaisina. Satunnaisuus ja sitä kautta todennäköisyysjakaumat tarkasteluihin tulevat otantamenetelmästä eli otoksen poimintaan liittyvästä satunnaisuudesta. Estimaattoreiden odotusarvot ja hajontaluvut määritellään otantajakauman suhteen eli kaikista mahdollisista otoksista laskettujen estimaattien keskiarvoina ja hajontalukuina. Merkitään otantajakauman suhteen laskettuja tunnuslukuja seuraavasti:

E_p : Odotusarvo
 B_p : Harha
 MSE_p : Keskineliövirhe

Toisenlainen lähestymistapa otantateoriaan saadaan, kun oletetaan, että perusjoukko on syntynyt jonkin satunnaisprosessin tuloksena. Tällöin perusjoukon estimoitavat suureet ovat oleellisesti satunnaislukuja. Niiden todennäköisyysjakaumaa kuvaavaa mallia kutsutaan superpopulaatiomalliksi. Perus-

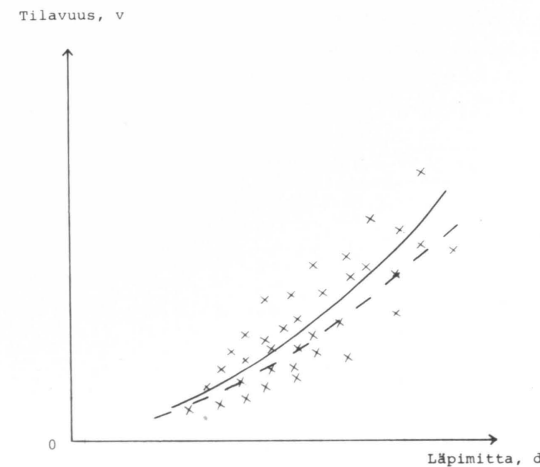
joukko ajatellaan poimituksi äärettömästä, postuloidun mallin mukaisesta superpopulaatiosta. Kiinnostus superpopulaatiomallista lähteviin otantateoreettisiin tarkasteluihin on tilastotieteessä lisääntynyt voimakkaasti viimeisten 10–15 vuoden aikana. Yksinkertainen esimerkki selventää lähestymistapaa.

Esimerkki 1: Tarkastellaan leimikon puulajien runkotilavuuden määrittämistä regressioestimoinnilla. Oletetaan, että puiden runkotilavuuden riippuvuutta läpimitasta kuvaa malli,

$$H: v_i = \beta_2 d_i^2 + \beta_3 d_i^3 + \beta_4 d_i^4 + \epsilon_i$$

jossa virhetermin ϵ_i odotusarvo on 0 ja varianssi σ^2 . Kuvassa 1 on hahmoteltu runkotilavuuden ja läpimitan välinen riippuvuus.

Kokonaistilavuuden estimointia varten estimoidaan ensin koepuista läpimitan ja tilavuuden riippuvuutta kuvaava regressiomalli. Tämän jälkeen lasketaan tilavuusestimaatit mallin mukaisesti ei-koepuille. Kokonaistilavuuden estimaattina käytetään koepuiden mitattujen tilavuuksien ja ei-koepuiden estimoitujen tilavuuksien summaa. Menettelyn mukaisen estimaattorin luotettavuustunnuksia määrittäessä on huomattava, että koepuiden tilavuudet tunnetaan tarkasti (kun mitausvirheet jätetään huomiotta). Toisin sano-



Kuva 1. Läpimitan ja runkotilavuuden välinen riippuvuus.

(—) Metsikkökohtainen tilavuusmalli.
 (---) Valtakunnallinen tilavuusmalli.

en epätarkkuutta estimaattoriin tulee vain ei-koepuiden tilavuuksien estimoinnista. Äärellisen perusjoukon korjaustermi tulee tällöin automaattisesti huomioitua, sillä mitä suurempaa osuutta kokonaistilavuudesta koepuut edustavat, sitä vähemmän estimaattoriin jää epätarkkuutta.

Ei-koepuiden tilavuuden luotettavuutta määrittäessä voidaan käyttää normaaleja regressioanalyysin teorian mukaisia kaavoja. Mikäli mallioletukset ovat riittävän oikeellisia, saadut luotettavuustunnuksukset ovat sopuisoinnissa otantajakauman suhteen määriteltujen tunnuksien kanssa, kuten myöhemmästä esimerkistä nähdään.

Normaalien otantajakauman suhteen määriteltujen tunnuslukujen rinnalle saadaan mallin suhteen määritellyt tunnusluvut. Merkitään niitä seuraavasti.

E_H : Mallin mukainen odotusarvo
 B_H : Mallin mukainen harha
 MSE_H : Mallin mukainen keskineliövirhe

Esimerkin valossa on selvää, että mallin mukainen keskineliövirhe riippuu otoksesta. Regressiomalli ja sitä kautta kokonaistilavuus saadaan luotettavammin estimoitua sellaisesta otoksesta, jossa koepuiden läpimitat ovat jakautuneet koko runkolukusarjan alueelle

kuin sellaisesta otoksesta, jossa läpimitat ovat keskittyneet kapealle läpimita-alueelle. Näin ollen mallin mukainen keskineliövirhe tarjoaa estimaattien luotettavuusmitan otoksen poimimisen jälkeen.

Yhteenvetona Käyttökelpoisista estimaattoreiden luotettavuusmitoista voidaan todeta

- MSP_p : — Metsikön tuottamaa satunnaisprosessia ei oleteta, vaan metsikkötunnuksia pidetään kiinteinä lukuina.
 — Mittaa estimaattorin luotettavuutta ennen otoksen poimimista.
 — Voidaan käyttää otantamenetelmää ja otoskokoa määrättäessä.
- MSE_H : — Oletetaan metsikkö satunnaisprosessin tuottamaksi, prosessia kuvaa superpopulaatiomalli.
 — Mittaa estimaattorin luotettavuutta otoksen poimimisen jälkeen.
 — Soveltuu estimaattien luotettavuuden raportointiin edellyttäen, että satunnaisprosessia kuvaava malli on riittävän oikea.
- $E_p(MSE_H)$: — Edellisen keskiarvo kaikista mahdollisista otoksista laskettuna.
 — Voidaan käyttää otantamenetelmää ja otoskokoa määrättäessä.

Otoksen ulkopuolisen informaation hyödyntäminen

Tarkastellaan otoksen ulkopuolisen informaation yhdistämistä otostietoon yksinkertaisen esimerkin valossa.

Esimerkki 2. Olkoon ongelmana edelleen puulajien kokonaistilavuuden määrittäminen. Oletetaan vahvasti yksinkertaistaen, että puiden runkotilavuudet noudattavat mallia

$$H: v_i = \mu + \epsilon_i,$$

jossa virhetermin ϵ_i odotusarvo on 0 ja varianssi σ^2 . Oletetaan edelleen, että esimerkiksi inventointiaineistojen perusteella on tiedossa että, kyseisen tyyppisissä metsiköissä keskitilavuus μ on keskimäärin μ_0 ja että keskitilavuuden varianssi on σ_0^2 . Tämä tieto, ns. ennakkoinformaatio voidaan yhdistää otostiedon kanssa käyttämällä metsikön keskitila-

vuuden μ estimaattorina otoskeskiarvon ja yleiskeskisarvon μ_0 painotettua keskiarvoa,

$$\hat{\mu} = \frac{1}{n + \kappa} (\sum v_i + \kappa \mu_0),$$

jossa \sum_s tarkoittaa summausta yli $n:n$ koe-puun. Estimaattorissa paino κ on tuntematon. Sen arvo voidaan määrätä vaatimukselta, että kokonaistilavuuden estimaattorin mallin mukaisen keskineliövirheen odotusarvo eri metsiköissä on mahdollisimman pieni, eli ratkaisemalla optimointitehtävä

$$\min_x E_G(\text{MSE}_H)$$

jossa odotusarvo E_G määritellään $\mu:n$ jakauman (odotusarvo μ_0 ja varianssi σ_0^2) suhteen. Optimointitehtävän ratkaisuksi saadaan (vrt. Pekkonen (1983)).

$$\kappa_0 = \frac{\sigma^2}{\sigma_0^2}$$

eli runkotilavuuden metsiköiden sisäisen ja välisen varianssin suhde. Optimaalista arvoa κ_0 käyttäen kokonaistilavuuden estimaattorin mallin mukaiseksi keskineliövirheeksi MSE_H saadaan

$$N^2 \left(1 - \frac{n}{N}\right) \left(1 + \frac{\kappa_0}{N}\right) \frac{\sigma^2}{n + \kappa_0}$$

Tarkastellaan kaavaa lähemmin. Ensimmäkin asettamalla ennakkoinformaation paino κ_0 nolaksi saadaan yksinkertaisen satunnaisotannan mukainen keskineliövirheen kaava, jossa perusjoukon varianssi S^2 on korvattu teoreettisella vastineellaan σ . Toiseksi kaavasta nähdään ennakkoinformaation käyttämisestä koituva hyöty. Se vastaa kutakuinkin κ kappaletta lisähavaintoa. Näin ollen ennakkoinformaation hyödyntäminen kannattaa, kun toisaalta metsiköiden välinen varianssi on pientä metsiköiden sisäiseen varianssiin verrattuna, jolloin optimaalinen $\kappa:n$ arvo on suuri ja toisaalta otoskoko on pieni, jolloin jokainen lisähavainto lisää merkittävästi estimaattorin luotettavuutta.

Tarkastellaan seuraavaksi miten ennakkoinformaation hyödyntäminen voidaan yleistää ensimmäisen esimerkin tilanteeseen, jossa tavoitteena oli regressioestimoinnilla määrätä puulajin kokonaisvaltaisuus. Edellisen esimerkin mukaista ennakkoinformaatio-

ta vastaisi tällöin se, että käytettävissä olisi otoksen ulkopuolista tietoa regressiomallin kertoimen jakaumasta. Jos näin on ja lisäksi ennakkoinformaatio voidaan ilmaista kertoimien jakaumalakisena ns. priorijakaumana, voidaan edetä ns. Bayesilaisen päättelyn mukaisesti.

Käytännössä Bayesilaista lähestymistapaa vaikeuttaa se, että käytettävissä ei ole ennakkotietoa regressiomallin kertoimista, vaan ennakkotieto on tilavuusyhtälöiden muodossa. Graafisesti esitettynä tilanne on kuvan 1 mukainen, otoshavaintojen lisäksi käytettävissä on esimerkiksi valtakunnallinen tilavuusyhtälö.

Tutkimuksessa Pekkonen (1983) tilavuusyhtälöitä on hyödynnetty teknisesti siten, että ensin on muodostettu tilavuusyhtälön mukainen teoreettinen aineisto, joka sitten sopivasti painottaen on yhdistetty otoshavaintoihin. Regressiokertoimet on määrätty näin muodostetusta yhdistetystä aineistosta. Menettely on analoginen esimerkin 2 menettelyn kanssa. Tarkastellaan menettelyä esimerkin 1 tilanteessa.

Esimerkki 1 (jatkoa). Merkitään otoshavaintoja

$$y = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, x = \begin{bmatrix} d_1^2 & d_1^3 & d_1^4 \\ \vdots \\ d_n^2 & d_n^3 & d_n^4 \end{bmatrix}$$

ja teoreettista aineistoa

$$r = \begin{bmatrix} f(d_{n+1}) \\ \vdots \\ f(d_{n+k}) \end{bmatrix}, R = \begin{bmatrix} d_{n+1}^2 & d_{n+1}^3 & d_{n+1}^4 \\ \vdots \\ d_{n+k}^2 & d_{n+k}^3 & d_{n+k}^4 \end{bmatrix}$$

jossa $f(d)$ on käytettävä tilavuusyhtälö. Yhdistämällä aineistot saadaan $\beta:n$ estimaattoriksi painotetun pns-menetelmän mukaisesti

$$\hat{\beta}_R = (X'WX + \kappa R'WR)^{-1} (X'Wy + \kappa R'WR)$$

jossa havaintokohtaisina painoina sekä otoksessa että teoreettisessa aineistossa on käytetty arvoja d_i^{-4} , ja kerroin κ määrää teoreettisen aineiston yleispainon. Arvolla $\kappa=0$ estimaattori supistuu normaaliksi otoksen mukaiseksi painotetun pns-menetelmän esti-

maattoriksi. Painon κ kasvaessa estimoitu tilavuusmalli lähenee tilavuusyhtälöä.

Estimaattorin $\hat{\beta}_R$ määrittämisessä on kaksi ongelmaa. Ensiksi, millaista läpimittajakautamaa käyttää teoreettisessa aineistossa ja toiseksi, mikä on painon κ optimaalinen arvo. Ongelmiin ei ole löytynyt analyttistä ratkaisua, vaan ratkaisuja joudutaan perustelemaan intuitiivisesti. Pekkonen (1983) on käyttänyt teoreettisessa aineistossa tasaista läpimittajakautamaa sekä määrännyt painon κ arvon suhteesta

$$\kappa_0 = \frac{\text{Ei-koepuiden kokonaistilavuuden estimaattorin varianssi mallin suhteen}}{\text{Tilavuusyhtälön mukaisen ei-koepuiden kokonaistilavuuden harhan neljän dototusarvo}}$$

Suhteessa osoittaja riippuu otoksesta eli siitä, millainen on koepuiden läpimittajakautama. Nimittäjässä oleva odotusarvo voidaan tulkita tilavuusyhtälön metsiköiden väliseksi varianssiksi.

Taulukossa 1 on esitetty simulointikokeiden tuloksia tilanteessa, jossa puulajin runkoja on ollut 1000 kpl ja runkopuun kokonaistilavuus 90 m³. Estimoidavana suurena on ollut puulajin runkopuun kokonaistilavuus. Ennakkoinformaationa on käytetty valtakunnallista rinnankorkeusläpimittaan perustuvaa tilavuusfunktiota, joka tarkasteltavassa puujoukossa on aliarvioinut runkotilavuutta keskimäärin 10 %. Otannassa koepuiden pöimintatodennäköisyydet ovat olleet verrannolliset rungon pohjapinta-alaan. Tulokset on laskettu 1000:sta toisistaan riippumattomasti pöimitusta otoksesta. Kokeet on selostettu tarkemmin tutkimuksessa Pekkonen (1983).

Taulukko 1. Kokonaistilavuuden estimaattoreiden vertailua teoreettisessa 1000:n rungon aineistossa. Tunnuksukset prosenttia kokonaistilavuudesta.

Otoskoko	Normaali estimaattori		Ennakkoinformaatiota hyödyntävä estimaattori	
	Harha	Keskivirhe	Harha	Keskivirhe
10	-0,3	10,6	-0,7	4,6
25	0,1	3,0	0,0	2,9
50	0,0	2,0	0,0	2,0
100	0,0	1,4	0,0	1,4
200	0,0	0,8	0,0	0,8

Yhteenveto

Esitetty menetelmä on tyypillisesti pieniin otoksiin soveltuva estimointimenetelmä. Otoksen ulkopuolisen tiedon kytkeminen otostiedon kanssa pienentää yleensä estimaattoreiden varianssia. Samalla estimaattoreihin kuitenkin tulee ennakkoinformaation suuntaista harhaa. Harhan suuruus riippuu ennakkoinformaation harhaisuudesta sekä siitä, millä painolla ennakkoinformaatiota hyödynnetään. Tilanne on aina analysoitava sovellutuskohtaisesti.

Estimointitilanteessa tulisi perusjoukko ja käytettävä ennakkoinformaatio mallittaa siten, että saataisiin kuva kummankin tiedon luotettavuudesta. Mikäli kyseessä on metsikötunnusten estimointi, tämä tarkoittaa, että tulisi selvittää estimoitavien tunnusten metsiköiden sisäisen ja metsiköiden välisen varianssikomponentin suuruus. Tällaiseen varianssin komponentteihin jakoon on vasta äskettäin kiinnitetty huomiota (vrt. Kilkki, 1983). Varianssikomponenttien estimoinnilla on vaikutuksensa myös käytännön otantasuunnitelmiin. Valtakunnallisia tai suuralueittaisia yhtälöitä määrättäessä tulisi otosyksiköitä sijoittaa nykyistä enemmän samoihin metsiköihin, jotta molemmat komponentit voitaisiin estimoida erikseen.

Yhteenvetona voidaan todeta, että ennakkoinformaatiota hyödyntävät menetelmät vaativat entistä tarkempaa ongelman analyysia, mutta samalla ne mahdollistavat havaintojen entistä tarkemman hyödyntämisen.

Kirjallisuus

- Kilkki, P. 1983. Sample trees in timber volume estimation. Acta For. Fenn. 182: 1-35.
 Pekkonen, T. 1984. Leimikon puuston tilavuuden arviointi regressioennustinta käyttäen. Metsäntutkimuslaitoksen tiedonantoja 86.

Summary

Utilization of inventory Data in Estimating of Stand Characteristics

In forest inventories the usual problem is that stand characteristics of a small area are to be estimated but there are only a few sample units available. The standard solution has been to use corresponding characteristics of

a larger area or general characteristics derived from the other sources. In such circumstances the sample units of the area have not been utilized efficiently.

The paper presents a method by which auxiliary information is combined with sample information with the aim of reducing the sampling variance of estimators. The method is most suitable for small samples, when the standard error of estimators can be large in comparison with the possible bias in auxiliary information.

Estimaattorin hyvyyden mittoja

Olkoot $\tilde{\beta}$ Jokin parametrivektorin β estimaattori. Määrittelemme nyt seuraavat estimaattorin $\tilde{\beta}$ hyvyyden mitat:

1° Estimaattorin $\tilde{\beta}$ kovarianssimatriisi on

$$\text{cov}(\tilde{\beta}) = E[(\tilde{\beta} - E(\tilde{\beta}))(\tilde{\beta} - E(\tilde{\beta}))'].$$

2° Estimaattorin $\tilde{\beta}$ kokonaiskeskineliövirhe on

$$\text{MSE}(\tilde{\beta}) = E[(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta)].$$

3° Estimaattorin $\tilde{\beta}$ prediktiiivinen keskineliövirhe on

$$\begin{aligned} \text{PMSE}(\tilde{\beta}) &= E[(\tilde{\beta} - \beta)'X'X(\tilde{\beta} - \beta)] \\ &= E[(\tilde{y} - X\tilde{\beta})'(\tilde{y} - X\tilde{\beta})], \\ &\text{missä } \tilde{y} = X\tilde{\beta}. \end{aligned}$$

Mittaa 1° käytetään harhattomien, mittoja 2° ja 3° harhaisten estimaattoreiden vertailuun. Mittaa 2° käytettäessä kiinnitetään huomio estimaattorin tarkkuuteen, kun taas 3° mittaa sovitteen \tilde{y} tarkkuutta. On syytä huomata, että hyvyyden mittojen arvot riippuvat parametriavaruuden pisteestä β .

Harhaisten estimaattorien yhteydessä käytettävät hyvyyden mittojen lausekkeet voidaan aina jakaa kahteen komponenttiin, joista ensimmäinen kuvaa estimaattorin varianssia ja toinen harhaa. Esimerkiksi

$$\text{MSE}(\tilde{\beta}) = \text{tr cov}(\tilde{\beta}) + \text{bias}(\tilde{\beta})'\text{bias}(\tilde{\beta}),$$

missä $\text{bias}(\tilde{\beta}) = \tilde{\beta} - E(\tilde{\beta})$. Jos $\tilde{\beta}$ on harhaton,

$$\text{MSE}(\tilde{\beta}) = \text{tr cov}(\tilde{\beta}).$$

Olkoot $\tilde{\beta}_1$ ja $\tilde{\beta}_2$ kaksi β :n estimaattoria. Hyvyyden mittojen 1° - 3° avulla voidaan määrittellä kolme erilaista paremmuuden käsitettä:

1° Estimaattori $\tilde{\beta}_1$ on tehokkaampi kuin $\tilde{\beta}_2$, jos matriisi $\text{cov}(\tilde{\beta}_2) - \text{cov}(\tilde{\beta}_1)$ on ei-negatiivisesti definitiini.

2° Estimaattori $\tilde{\beta}_1$ on kokonaiskeskineliövirheen mielessä parempi kuin $\tilde{\beta}_2$, jos $\text{MSE}(\tilde{\beta}_1) \leq \text{MSE}(\tilde{\beta}_2)$.

3° Estimaattori $\tilde{\beta}_1$ on prediktiiivisen keskineliövirheen mielessä parempi kuin $\tilde{\beta}_2$, jos $\text{PMSE}(\tilde{\beta}_1) \leq \text{PMSE}(\tilde{\beta}_2)$.

Johdannossa mainittu PNS-estimaattoria b koskeva optimaalisuustulos voidaan nyt pu-

kea seuraavaan muotoon: Olkoon $\tilde{\beta}$ jokin lineaarinen ja harhaton β :n estimaattori. Tällöin $\text{cov}(\tilde{\beta}) - \text{cov}(b)$ on ei-negatiivisesti definitiini matriisi.

Keskeinen perustelu harhaisten ja/tai epälineaaristen estimaattoreiden käytölle on ns. Steinin paradoksi: On olemassa β :n estimaattoreita, jotka ovat keskineliövirheen mielessä parempia kuin PNS-estimaattori b .

Prioritiedon liittäminen lineaariseen malliin

Käsitlemme tässä kolme erilaista tapaa liittää prioritietoa lineaariseen malliin. Näistä ensimmäinen käyttää hyväksen eksaktia prioritietoa ja johtaa rajoitettuun PNS-estimaattoriin. Toisessa ja kolmannessa tavassa prioritieto on luonteeltaan stokastista. Näitä vastaavat estimaattorit ovat ns. sekaestimaattori ja (eräs) Bayesin estimaattori.

On syytä huomata, että sekä rajoitettua PNS-estimaattoria että sekaestimaattoria saatta kannattaa käyttää, vaikka prioritieto olisi virheellistä. Tällaiseen tilanteeseen joudutaan esimerkiksi, kun prioritietona on jokin regressiokertoimien "spektrin" säännöllisyysoletus. Prioritiedon käytöllä on myös usein mahdollista vähentää multikollinearisuuden haittoja.

(i) Olkoon prioritieto esitettävissä eksakteina lineaarisina rajoituksina

$$(4) \quad r = R\beta + s,$$

missä R on $n \times p$ -matriisi. Vektori s edustaa rajoituksissa esiintyvää spesifointivirhettä. Kun rajoitukset (4) otetaan huomioon PNS-keinoa sovellettaessa, saadaan rajoitettu PNS-estimaattori

$$(5) \quad b_R = b + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - Rb).$$

Jos rajoitukset on spesifioitu oikein eli $s = 0$, b_R on harhaton ja tehokkaampi kuin tavallinen PNS-estimaattori b . Vaikka rajoitukset olisi spesifioitu väärin eli $s \neq 0$ saattaa b_R olla prediktiiivisen keskineliövirheen mielessä parempi kuin b .

LINEAARINEN REGRESSIOMALLI JA PRIORITIEDON KÄYTTÖ

ILKKA MELLIN

Johdanto

Lineaarista regressiomallia käytettäessä havaintoja koskeva otosinformaatio voidaan esittää muodossa

$$(1) \quad y = X\beta + \epsilon,$$

jossa y ja ϵ ovat n -vektoreita, X on $n \times p$ -matriisi ja β on p -vektori. Jäännöstermin ϵ käyttäytymisestä tehdään oletukset

$$(2) \quad E(\epsilon|X) = 0, \quad \text{cov}(\epsilon) = \sigma^2 I.$$

Parametrivektorin β pienimmän neliösumman estimaattori (PNS-estimaattori)

$$(3) \quad b = (X'X)^{-1}X'y$$

on paras lineaaristen ja harhattomien estimaattoreiden joukossa, kun paremmuutta mitataan estimaattorin kovarianssimatriisil-

la. Jos jäännöstermi on normaalisti jakautunut, PNS-estimaattori b on paras jopa kaikkien harhattomien estimaattoreiden joukossa. Vaikka PNS-estimaattori b on tällä tavalla optimaalinen, voidaan sitä parempia estimaattoreita muodostaa luopumalla harhattomuudesta tai lineaarisuudesta tai käyttämällä jotain toista hyvyyden mittaa. Tällaiset estimaattorit hyödyntävät yleensä β :n käyttäytymistä koskevaa prioritietoa.

Tässä esityksessä luodaan katsaus eri tapoihin liittää parametrivektoria β koskevaa prioritietoa lineaariseen malliin. Kappaleessa 1 esitetään estimaattorin hyvyyden mittoja, kappaleessa 2 esitetään prioritiedon liittämistapoja ja kappaleessa 3 käsitellään lyhyesti prioritiedon käytön tärkeintä kysymystä: mistä prioritieto saadaan? Yleisesityksiä prioritiedon käytöstä regressiomalleissa sisältävät Judge ym. 1980 sekä Vinod and Ullah 1981, jotka sisältävät myös hyvät lähdeluettelot.

(ii) Olkoon prioritieto esitettävissä stokastisina lineaarisina rajoituksina

$$(6) \quad r = R\beta + \delta,$$

missä R on $m \times p$ -matriisi, vektori s edustaa spesifiointivirhettä ja

$$(7) \quad E(\delta|R) = 0, \quad \text{cov}(\delta) = \sigma^2/k I.$$

Lisäksi oletetaan, että ε ja δ ovat riippumattomia. Parametri k kuvaa otos- ja prioritiedon tarkkuuden suhdetta. Tämän parametrin valintaan ei tässä puututa. Tulkitsemalla rajoitukset (7) lisähavainnoiksi, saadaan *sekaestimaattori*

$$(8) \quad b_R(k) = (X'X + kR'R)^{-1} (X'y + kR'r).$$

Sekaestimaattori $b_R(k)$ sisältää erikoistapauksenaan PNS-estimaattorin b ja rajoitetun PNS-estimaattorin b_R , sillä $b_R(0) = b$ ja $b_R(k) \rightarrow b_R$, kun $k \rightarrow \infty$. Sekaestimaattori antaa mahdollisuuden liittää lineaariset rajoitukset pehmeästi, eräänlaisina toivomuksina, malliin. Jos $s = 0$, niin $b_R(k)$ on harhaton ja tehokkaampi kuin PNS-estimaattori b . Vaikka $s \neq 0$, niin $b_R(k)$ saattaa olla predikttiivisen keskineliövirheen mielessä parempi kuin b .

(iii) Oletetaan, että jäännöstermi ε noudattaa normaalijakaumaa kohdassa (2) mainituin parametrein ja, että parametrivektorin β priorijakauma on myös normaalinen:

$$(9) \quad \beta \sim N_p(A\alpha, \sigma^2/k I).$$

Lisäksi oletetaan, että q -vektori α noudattaa epäinformatiivista priorijakaumaa. Tällöin β :n (eräs) *Bayesin estimaattori* on posterioijakauman odotusarvo

$$(10) \quad \bar{b}_A(k) = (X'X + k(I - A(A'A)^{-1}A'))^{-1}X'y.$$

Jos $A = 0$, jolloin

$$(9)' \quad \beta \sim N_p(0, \sigma^2/k I),$$

supistuu kaava (10) muotoon

$$(10)' \quad \bar{b}(k) = (X'X + kI)^{-1}X'y.$$

Estimaattori $\bar{b}(k)$ on ns. *harjaestimaattori*, joka on näin saanut Bayesilaisen tulkinnan. Harjaestimaattori voidaan myös johtaa liittämällä malliin epälineaarinen rajoitus

$$(11) \quad \beta'\beta = \text{vakio}.$$

Tällöin prioritieto koskee parametrivektorin pituutta.

Prioritiedon saanti

Prioritiedon lähteenä voivat esiintyä esimerkiksi seuraavat seikat:

- 1° subjektiiviset käsitykset,
- 2° mallin käyttöalueeseen liittyvä taustateoria, kuten talousteoria,
- 3° parametrivektorin käyttäytymistä koskevat toivomukset,
- 4° aikaisemmat tutkimukset.

Subjektiiivisten käsityksien mukaanotto johtaa Bayesin menetelmiin, mutta on myös syytä ottaa huomioon empiiristen Bayesin menetelmien antama mahdollisuus "estimoida" priorijakauma otoksesta.

Taustateorian implikoimat rajoitukset liitetään tavallisesti malliin eksakteina. Stokastisia rajoituksia on syytä käyttää varsinkin silloin, kun rajoituksiin ei sinänsä uskota, vaan ne halutaan liittää malliin toivomuksina. Tällaisia toivomuksia ovat esimerkiksi polynomirajoitukset äärellisen jakutuneen viipymän malleissa tai kausivaihtelumalleissa.

Tulokset aikaisemmista tutkimuksista, jossa on saatettu käyttää suurempaa otoskokoa tai tarkempia malleja, tuottavat usein käyttökelpoista prioritietoa. Tällöin prioritiedon käyttö mahdollista sen, että esimerkiksi jonkin ilmiön jatkuvassa seurannassa voidaan käyttää pientä otoskokoa tai yksinkertaista mallia estimoinnin tarkkuuden siitä kärsimättä.

Kirjallisuus

- Judge, G. G., Griffiths, W. E., Hill, R. C. & Lee, T.-C. 1980. *The Theory and Practice of Econometrics*. Wiley, New York.
- Vinod, H. D. & Ullah, H. 1981. *Recent Advances in Regression Methods*. Marcel Dekker, New York.

Summary

Linear Regression Model and the Use of Prior Information

It is well-known that the ordinary least squares estimator is the best linear unbiased estimator. However, it is possible to form better estimators if we give up linearity or unbiasedness and replace the variance with some other measure of superiority for estimators. Very often these estimators exploit some kind of prior information. In this paper some estimators combining sample and prior information are reviewed. The estimators considered are the restricted least squares estimator, the mixed estimator, a Bayesian estimator and the ridge estimator. Some measures of superiority and the question of obtaining prior information are also considered.