

Lappeenrannan-Lahden teknillinen yliopisto LUT
LUT School of Engineering Science
Tietotekniikan koulutusohjelma

Kandidaatintyö

Jaakko Ketola

**PUUSTON INVENTOINTI LASERKEILAUSDATAA
KÄYTTÄEN**

Työn tarkastaja: TkT Virpi Junttila.

Työn ohjaajat: FT Tuomo Kauranne ja TkT Virpi Junttila.

TIIVISTELMÄ

Tekijä: Ketola, Jaakko

Nimi: **Puuston inventointi laserkeilausdataa käyttäen**

Osasto: Tietotekniikan osasto, Lappeenrannan-Lahden teknillinen yliopisto LUT

Vuosi: 2019

Paikka: Lappeenranta

Kandidaatintyö

31 sivua, 7 taulukkoa ja 5 kuvaa.

Tarkastaja: TkT Virpi Junttila

Hakusanat: puustoinventointi, laserkeilaus, kaukokartoitus

Puuston inventoinnilla tuotetaan tietoa metsävaroista, jota metsänomistajat ja tutkimuslaitokset hyödyntävät. Inventointia voidaan suorittaa kaukokartoitusmenetelmiä hyödyntäen. Kaukokartoituksen käytöllä pyritään kustannusten säästämiseen työvoimaa vaativasta maastotyöstä, ja inventointitulosten objektiivisuuteen ja tarkkuuteen.

Työn tavoitteena on toteuttaa puuston inventointisovellus, jolla pystytään estimoimaan puuston kannalta oleellisia tunnuslukuja laserkeilausdataa hyödyntäen. Laserkeilauksella tuotettua kaukokartoitusaineistoa käytetään inventoinnissa käyttäen estimaattoreina k similaarisimman naapurin menetelmää ja lineaarista harvaa bayesilaista regressiota. Näillä sovellukseen toteutetuilla menetelmillä saadaan tuotettua metsänomistajille perinteisiä menetelmiä tarkempia tuloksia, joiden validoinnissa käytetään ristiinvalidointia ja estimointien neliövirhettä.

Laserkeilausdata kerätään ilmasta käsin käyttämällä laservalon lentoajan mittaamiseen perustuvaa skanneria. Inventoitavat tunnusluvut johdetaan osin laskentamalleja hyödyntäen maastossa suoritetuista mittauksista.

Inventointitulosten validointi on tässä työssä suoritettu Matalansalon ja Juukan koetilojen maastossa mitatuilla koelajoukoilla ja niihin liittyvillä laserkeilausdatoilla. Perinteisellä maastoinventoinnilla esimerkiksi puuston kokonaistilavuuksien virhe voi olla metsätalouskuviotasolla yli 30% ja laserkeilausaineistoa käyttäen tilavuuksien virheeksi saatiin n. 20% koelatasolla, joka pienenee johdettaessa kokonaistulokset metsätalouskuviotasolle.

SISÄLLYSLUETTELO

| | | |
|----------|--|-----------|
| 1 | JOHDANTO | 3 |
| 1.1 | Tausta | 3 |
| 1.2 | Tavoitteet ja rajaukset | 4 |
| 1.3 | Työn rakenne | 5 |
| 2 | LASERKEILAUS JA SELITTÄVÄT MUUTTUJAT | 6 |
| 2.1 | Mittauslaitteisto | 6 |
| 2.1.1 | Laserkeilausdatan tallennus | 7 |
| 2.2 | Datan esikäsittely puustoinventointia varten | 7 |
| 2.2.1 | Digitaalinen maastomalli laserkeilausdatasta | 8 |
| 2.2.2 | Estimoinnin selittävät muuttujat | 9 |
| 3 | MAASTOMITTAUS JA SELITETTÄVÄT MUUTTUJAT | 10 |
| 3.1 | Koealojen otanta | 10 |
| 3.2 | Puuston korkeusmallit | 11 |
| 3.3 | Estimoinnin selitettävät muuttujat | 11 |
| 4 | ESTIMOINNIN MENETELMÄT | 13 |
| 4.1 | Muuttujien standardisointi | 13 |
| 4.2 | K similaarisimman naapurin regressio | 13 |
| 4.3 | Lineaarinen harva bayesilainen regressio | 15 |
| 5 | VIRHE- JA LUOTETTAVUUSARVIOINTI | 19 |
| 5.1 | Ristiinvalidointi | 19 |
| 5.2 | Virhemitat | 20 |
| 6 | INVENTOINTISOVELLUS | 21 |
| 7 | TUTKIMUSAINEISTOT JA TULOKSET | 24 |
| 7.1 | Matalansalon koetila | 24 |
| 7.2 | Juukan koealat | 26 |
| 8 | YHTEENVETO JA JOHTOPÄÄTÖKSET | 29 |
| | LÄHTEET | 30 |

SYMBOLI- JA LYHENNELUETTOLO

| | |
|--------------|---|
| ALS | Airborne Laser Scanner |
| DGPS | Differential Global Positioning System |
| DTM | Digital Terrain Model |
| GIS | Geographical Information System |
| GPS | Global Positioning System |
| KMSN | K-Most Similar Neighbour |
| KNN | K-Nearest Neighbour |
| LiDAR | Light Detection and Ranging |
| <i>RMSE</i> | Root Mean Square Error eli virheiden neliöiden keskiarvon juuri |
| <i>RMSE%</i> | Relative Root Mean Square Error eli suhteellinen RMSE |
| <i>bias</i> | Virheiden keskiarvo eli harha |
| <i>bias%</i> | Suhteellinen harha |

1 JOHDANTO

1.1 Tausta

Metsien tilasta tuotetaan tietoa inventoinneilla, jotka perustuvat otantaan ja mallien käyttöön. Menetelmän valintaan vaikuttaa menetelmän kustannukset ja tarkkuusvaatimukset. Metsän inventointeja tekevät niin yksityiset metsänomistajat, yritykset kuin julkiset tutkimuslaitokset (Kangas et al., 2006). Inventoinnin tuloksia hyödynnetään metsän käytön ja toimenpiteiden suunnitteluun. Valtakunnallisesti toistuvasti suoritetaan myös valtakunnan metsien inventointi (VMI), jossa kerätään tietoa maanlaajuisesti metsien tilasta ja metsävaroista (Maltamo et al., 2004).

Inventointia hyödynnetään metsien käytön suunnittelussa ja se on perustunut perinteisesti osin visuaaliseen arviointiin ja informaatioon, joka on kerätty paikanpäältä metsätalousalueilta eli metsätalouskuvioilta aiheuttaen sekä epätarkkuutta, että subjektiivisuutta (Haara et al., 2004). Metsätalouskuviot ovat perinteisesti muodostettu visuaalisesti ilmakuvia tarkastelemalla muodostaen niistä metsätyypin kannalta yhtenäisiä alueita. Tällöin kullakin metsätalouskuvion alueella on käyty paikan päällä mittaamassa tarvittavan suuri määrä koealoja, joiden tulokset on yhdistetty metsätalouskuvion tunnusluvuiksi. Usein yksittäiselle metsätalouskuviolle on mitattu vain muutamia koealoja työn käsityöluonteen vuoksi ja siten metsätalouskuviotasolla esimerkiksi puuston kokonaistilavuuden virheet ovat vaihdelleet mittaajakohtaisesti 10 ja 30% välillä (Haara et al., 2004).

Perinteinen maastoinventointi työvoimavaltaisena prosessina ei ole enää kilpailukykyinen menetelmä ja sen on korvannut viime vuosina kaukokartoitusmenetelmät. Kaukokartoitusmenetelmät perustuvat ilmakuvista, satelliittikuvista ja ilmasta suoritettavasta laserkeilausdatasta (ALS) saatavaan informaatioon. Useissa tutkimuksissa nämä kaukokartoitusmenetelmiin perustuvat puustoninventointimenetelmät ovat parantaneet inventoinnin tarkkuutta ja tehneet puustoninventoinnista perinteistä menetelmää kustannustehokkaampaa. Kaukokartoitusmenetelmiin perustuvissa lajikohtaisissa inventoinneissa on kuitenkin vielä parannettavaa. Ensimmäiset käytännön puustoninventoinnit laserkeilausdataa käyttäen ovat Norjasta vuodelta 2002 (Næsset et al., 2004).

Kaukokartoitusinventointia on tehty sekä yksinpuintulkintana, että alueellisena tulkintana. Yksinpuintulkinnassa etsitään ensin jokainen puu ja estimointi tehdään kullekin puulle erikseen esimerkiksi puun oletetulle alueelle osuneiden laserpulssien perusteella. Yksinpuintulkinta vaatii kuitenkin kaukokartoitusmenetelmiltä ilmakuvilta korkeampaa tarkkuutta ja laserkei-

lauksen osalta korkeampaa pulssitiheyttä (Hyypä et al., 1999). Aluepohjaisessa tulkinnassa hyödynnetään laserkeilauspulsseista laskettuja korkeus- ja tiheysjakaumia laskettavalla alueella. Alue voi olla koeala, metsätalouskuvio, metsätalouskuviota pienempi niin sanottu mikrokuvio, jossa laskentayksikön alueella olevan metsän oletetaan olevan homogeeninen (Heikkilä et al., 2010).

1.2 Tavoitteet ja rajaukset

Tämän työn tavoitteena on toteuttaa käytännön sovellus hyödyntäen kahta eri metsien inventoinnissa käytettyä menetelmää: k similaarisimman naapurin menetelmää (Sironen et al., 2001) ja lineaarista harvaa bayesilaista regressiota (Junttila et al., 2008). Näitä kahta menetelmää on käytetty puuston inventointiin liittyen, joten työ on rajattu niiden toteutukseen. Sovellukseen toteutetaan nämä molemmat menetelmät, jotta voidaan varmistaa kulloisellakin estimoitavalla aineistolla paremmin toimiva menetelmä.

Laserkeilausdatan käyttämistä metsätalouden tunnuslukujen estimointiin on tehty käyttäen k similaarisimman naapurin menetelmää ja lineaarista harvaa bayesilaista regressiota. Muita käytettyjä menetelmiä ovat lisäksi ainakin pienimmän neliösumman sovituksella lasketun lineaarisen regressio tai k lähimpien naapureiden menetelmä (Sironen et al., 2001).

Tässä työssä keskitytään alueelliseen tulkintaan eikä yksinpuintulkintaa ole toteutettu, vaikka käytettyjä menetelmiä voidaankin hyödyntää sekä yksinpuintulkinnassa, että alueellisessa tulkinnassa.

Työssä käytetyt selittävät muuttujat ovat tutkimuksissa esitettyjä, eikä näitä ja näiden laskentaa ole tässä muutettu tai muuttujia ei ole käsitelty esimerkiksi linearisoimalla.

Ilmakuvista saatavaa tietoa ei hyödynnetä tässä työssä, vaikka se yleisesti onkin käytössä ja myös työssä toteutettavan sovelluksen hyödynnettävissä. Ilmakuvien käyttö ei aiheuta juurikaan lisäkustannuksia inventoinnille, koska ilmakehän aineistoa on useimmiten saatavilla kuvattuna samassa yhteydessä laserkeilauksen kanssa. Metsätalouskuvioden muodostaminen pohjautuu ilmakuvien ja laserkeilausdatasta rasteroitujen kuvien hyödyntämiseen. Ilmakuvista voidaan laskea inventoinnille hyödyllisiä selittäviä muuttujia, joilla etenkin lajikohtaisten tunnuslukuestimaattien tarkkuutta voidaan parantaa.

1.3 Työn rakenne

Työn rakenteen on tarkoitus tukea laserkeilauskaukokartoituksella suoritettavan puuston inventoinnin prosessia ja niitä vaihteita, joilla tuotetaan ja käsitellään aineistoa ja lopulta näistä johdetaan puuston inventoinnin tulokset tai puuston inventoinnin laadun varmistamiseen hyödynnetyt validoinnin tulokset.

Työssä esitellään ensin ilmasta suorettavaa laserkeilausta yleisesti ja sillä tuotettavaa aineistoa. Lisäksi kuvataan aineiston tuottamiseen vaadittavaa esikäsitelyä ja varsinaisessa laskennassa käytettyjen muuttujien johtamista esikäsitellystä laserkeilausdatasta.

Seuraavaksi esitellään yleisesti maastomittauksia ja koealojen otantaa, sekä koealoilta laskettavia muuttujia. Näitä samoja mitattuja ja laskettuja muuttujia puuston inventoinnilla ja estimoinnilla on tarkoitus tuottaa metsätalouskuvioille tai muille laskentayksiköille.

Estimoinnista esitellään yleisesti käytössä olevat menetelmät, jotka toteutetaan sovelluksessa ja joita tässä työssä on tarkoitus hyödyntää. Näiden estimointimenetelmien ja puuston inventoinnin laadun arviointia varten esitetään sovelluksen hyödyntämät virhemitat, joista tuloksissa esitetään ristiinvalidointi ns. leave-one-out -validoinnilla saadut tulokset.

Lopuksi työssä kuvataan toteutetun sovelluksen ympäristöä, rakennetta ja tällä saadut tulokset kahdella eri tutkimuksessa käytetyn koetilan aineistoilla kahdella eri estimointimenetelmällä.

2 LASERKEILAUS JA SELITTÄVÄT MUUTTUJAT

Ilmasta suoritettava laserkeilauksessa (ALS) käytetään opto-mekaanisia skannereita, jotka lähettävät laser säteitä ja mittaavat takaisin heijastuvaa laser sädettä. Kaikille eri laserkeilauslaitteille on yhteistä, että ne mittaavat jollain keinolla etäisyyttä heijastumispisteeseen. Tyypillinen laserkeilauslaitteisto koostuu etäisyydsmittarista, opto-mekaanisesta skannerista ja prosessointiyksiköstä. (Wehr et al., 1999)

2.1 Mittauslaitteisto

Etäisyyden mittaamiseen on käytössä kaksi eri tapaa: pulssimodulointi ja jatkuvan aallon (CW) sinimodulointi (Wehr et al., 1999). Kummassakin modulointitavassa etäisyys lasketaan laser-säteen lentoajasta, joten erot ovat keinoissa tunnistaa paluupulssin vaihesiirtymä, joka rajoittaa suurinta mahdollista mittausetäisyyttä.

Opto-mekaanisina skannereina käytetään oskilloivaa peiliä, Palmer-skanneria ja kuituskanneria. Oskilloivassa peilissä mittauspisteitä tulee suoralta linjalta, joten mittauspisteet muodostavat ns. zigzag-kuvion. Palmer-skannerissa peiliä pyöritetään muodostaen ellipsin muotoisia kuvioita mittauspisteille maaston tasossa. Kuituskannerissa tekniikka on hyvin samankaltainen Palmer-skannerin kanssa, mutta optiikkana on valokuitunippu, jolloin lähetävä ja vastaanotettava optiikka ovat identtisiä ja pulssin lähetys ja vastaanotto pysyvät synkronoituina ja pulsseja voidaan lähettää suuremmalla tiheydellä mittausrésoluution ja suurimman mittausetäisyyden kärsimättä. (Wehr et al., 1999)

Laserin heijastumisen paluupulseja voidaan mitata diskreetisti tai jatkuvasti. Diskreettissä mittaamisessa tallennetaan yksittäisille pulsseille usein ensimmäinen paluupulssi ja viimeinen paluupulssi, usein myös välipulseja (Wehr et al., 1999). Jatkuvasti paluupulseja mittaavissa laitteistoissa voidaan tallentaa koko paluupulssi ajan funktiona, käytännössä tässäkin menetelmässä tallennetaan diskreetisti paluupulseja.

Tarkkojen mittauspisteiden saamiseksi mittauslaitteisto on kyettävä paikantamaan mahdollisimman tarkasti, joten paikannustarkkuutta parannetaan hyödyntämällä DGPS paikannusta (Monteiro et al., 2005), jossa sekä referenssimittauspisteessä (tukiasema), että mittauslaitteiston mukana kulkevassa laitteistossa tallennetaan jatkuvasti myös aikaleimat. Tukiasemassa ja lasermittauslaitteistolla kummallakin on GPS laitteisto, joka tallentaa jatkuvasti oman sijaintinsa määrittämiseen tarvittavaa dataa sekä GPS aikaa (Wehr et al., 1999). Tiedot yhdistetään

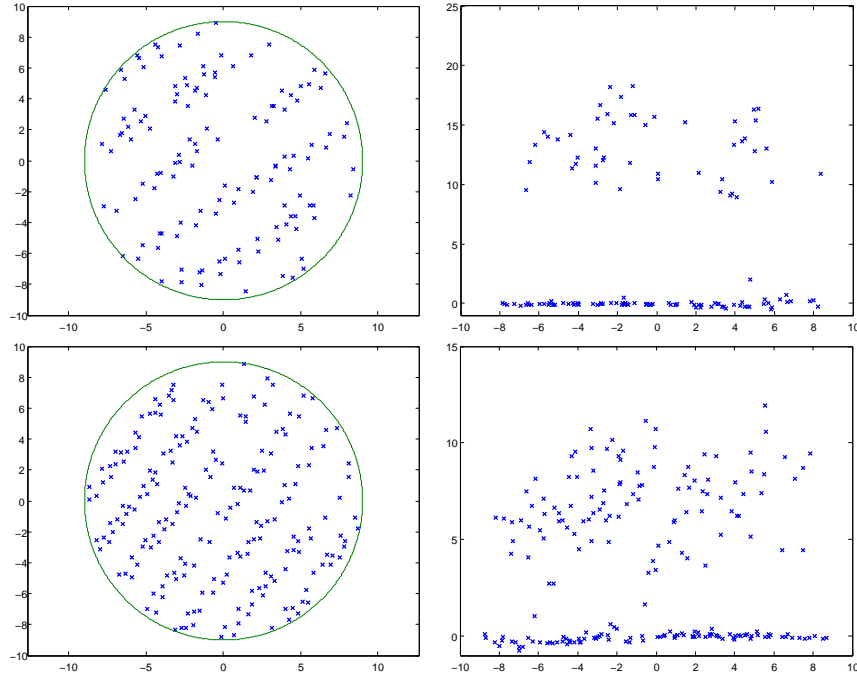
jälkeenpäin tai tukiasema voi lähettää tietoa paikannusvirheistä liikkuvalla laitteistolle, jotta laser-mittauspisteiden sijainnit saadaan mahdollisimman tarkoiksi.

2.1.1 Laserkeilausdatan tallennus

Laserkeilausdatan tallennukseen käytetään LAS-tiedostoformaattia (*LAS Specification Version 1.2*, 2010), joka on yleisesti käytössä eri datan tuottajilla ja käsittelijöillä. Saman yhteisen ja avoimen tiedostoformaatin käyttö mahdollistaa helpon yhteistoiminnan eri toimijoiden välillä. LAS-tiedostoformaatti on suunniteltu toimimaan tehokkaasti spatiaalisen pistejoukon tallennukseen. Kuitenkin edelleen käytetään jossain määrin datan tallennukseen ja siirtoon myös ASCII-tiedostoformaattia, jonka haittana on hidas käsittely, suuret käsiteltävät tiedostot ja mahdolliset epäselvyydet käytetystä referenssikoordinaatistosta.

2.2 Datan esikäsittely puustoinventointia varten

Laserkeilausdata LAS-tiedostoformaattissa on jo mittauslaitteiston osittain esikäsittelmää ja saatavilla pistejoukkona (x, y, z, i, n) (*LAS Specification Version 1.2*, 2010), jossa i on mitatun paluupulssin intensiteetti, n on kyseisen paluupulssin järjestysnumero ja z on korkeus merenpinnasta. Esikäsittelyssä muodostetaan digitaalinen maastomalli viimeisiä paluupulsseja hyödyntäen, joiden oletetaan suurelta osin olevan heijastuneita maanpinnasta eikä kasvustosta. Lopulta kaikista mittauspisteistä vähennetään tuotettu maastomalli, joten tämän esikäsittelyn jälkeen pisteet ovat (x, y, h, i, n) , jossa h on mittauspisteen korkeus maanpinnasta. Kuvassa 1 näkyy kahden eri koealan alueella olevat pistejoukot, joista on vähennetty maastomalli.



Kuva 1: Esikäsiteltyä laserkeilausdataa kahdelta eri koelalta ylhäältäpäin ja sivultapäin kuvattuna. Ylhäältä kuvattuna näkyy oskilloivan peilin zigzag-kuvio ja sivulta kuvattuna näkyy kasvustosta ja maasta heijastuneet laserpulsit. Yksikkönä metri.

2.2.1 Digitaalinen maastomalli laserkeilausdatasta

Digitaalinen maastomalli eli DTM voidaan yksinkertaisimmillaan luoda rasteroimalla laserkeilausdatan viimeisten paluupulssien lokaalit minimikorkeudet. Parempaan tulokseen päästään ns. vinouden balansointi -menetelmällä (Bartels et al., 2006). Tässä menetelmässä oletetaan mittausrvirheen olevan normaalijakautunutta, jolloin vinous on 0. Vinous määritellään seuraavasti:

$$\bar{z} = \frac{1}{N} \times \sum_{i=1}^N z_i, \quad (1)$$

$$\sigma_z = \sqrt{\frac{1}{N-1} \times \sum_{i=1}^N (z_i - \bar{z})^2}, \quad (2)$$

$$sk = \frac{1}{N \times \sigma_z^3} \times \sum_{i=1}^N (z_i - \bar{z})^3, \quad (3)$$

jossa N on alueellinen laserkeilausmittauspisteiden kappalemäärä, \bar{z} on samojen pisteiden korkeuksien keskiarvo, σ_z on samojen pisteiden keskihajonta ja sk on korkeusjakauman vinous. Lokaalille alueelle tulevat mittauspisteet luokitellaan kuuluviksi joko maastoon tai muuksi kohteeksi siten, että ylimmästä pisteestä lähtien pisteitä luokitellaan maastoon kuulumattomiksi

mittauspisteiksi, kunnes alueen pisteiden korkeuden jakauman vinous on tarvittavan lähellä arvoa 0 (Bartels et al., 2006). Pisteiden luokittelun jälkeen korkeusmalli voidaan tehdä esimerkiksi rasteroimalla lokaalit korkeuksien keskiarvot tai sovittamalla jokin korkeamman asteen tason yhtälö pistejoukkoon.

2.2.2 Estimoinnin selittävät muuttujat

Estimoinnissa käytetyt laserkeilausdatasta lasketut selittävät muuttujat on laskettu erilaisista mittauspisteen korkeus- ja intensiteettijakaumista olettaen niiden kuvaavan mahdollisimman hyvin kasvillisuuden korkeutta ja rakennetta. Valitut muuttujat ovat laskennallisesti helposti laskettavissa suoraan laserkeilausdatasta, josta on vähennetty digitaalinen maastomalli. Kaikille opetusjoukon koealoille ja estimointialueen estimointisoluille selittävät muuttujat on laskettu samalla tavalla.

| | |
|------------------------|--|
| $x_1 \cdots x_{10}$ | Lasketaan järjestetystä ensimmäisten paluupulssien korkeuksien kumulatiivisesta summasta. Muuttuja vastaa sen pulssin korkeutta, joka on lähinnä $r\%$ yhteenlasketusta summasta, $r = 10, 20, \dots, 100$. |
| $x_{11} \cdots x_{18}$ | Muuttuja vastaa viimeisen paluupulssin korkeusjakauman histogrammin persenttiä, jossa histogrammin keskikohdat ovat $3 \times r + 1.5$, $r = 0, \dots, 7$. |
| $x_{19,20,21}$ | Ensimmäisten paluupulssien intensiteettien kumulatiivisen histogrammin arvo. Histogrammin keskikohdat ovat $r + 0.5$, $r = 0, 1, 2$. |
| $x_{22,23,24}$ | Sama kuin edellinen, mutta viimeisillä paluupulssilla. |
| x_{25} | Ensimmäisten paluupulssien korkeuksien keskiarvo, jotka ylittävät 5 metriä. |
| x_{26} | Luonnollinen logaritmi ensimmäisten paluupulssien, joiden korkeus on alle 2 metriä, lukumäärän suhde ensimmäisten paluupulssien lukumäärään. |
| x_{27} | Keskiarvo kolmen korkeimman ensimmäisen paluupulssien korkeuksista. |

Taulukko 1: Estimoinnin selittävät muuttujat, jotka ovat alkujaan Næssetin esittämät

Taulukossa 1 on esitetty estimoinnin selittävien laserkeilausdataan perustuvien muuttujien laskenta (Næsset, 2002, Junttila et al., 2008). Kaikille muuttujille voidaan jälkeenpäin suorittaa erillinen linearisointi riippuen selitettävien ja kyseisen selittävän muuttujan suhteesta (Pahkinen et al., 1989). Tässä työssä ei kuitenkaan käsitellä näiden muuttujien mahdollista linearisointia.

3 MAASTOMITTAUS JA SELITETTÄVÄT MUUTTUJAT

Maastomittauksella tuotetaan estimointia varten opetusaineistoa, jolta tunnetaan estimoitavat selitettävät muuttujat mahdollisimman tarkasti. Mitattavat koealat ovat usein 9 metriä säteeltään olevia ympyränmuotoisia alueita, mutta myös pienempää koealakokoa käytetään esimerkiksi taimikoiden inventoinnissa. Perinteisesti koealoja on sijoitettu kullekin metsätalouskuviolle riippuen metsätalouskuvion pinta-alasta (Heikkilä et al., 2010). Kaukokartoitukseen perustuvassa inventoinnissa mitattavien koealojen lukumäärä voi olla huomattavasti pienempi, koska koealoja ei tarvita metsätalouskuvioittain vaan ennemminkin olemaan edustava otos inventoitavan alueen puustosta.

Mitattavien koealojen mahdollisimman tarkka paikannus on myös hyvin tärkeää, koska koealalle lasketaan selittävät muuttujat käyttäen ilmasta suoritettavaa laserkeilausta ja näiden aineistojen yhdistäminen on tehtävä mahdollisimman tarkasti. Koealojen keskipisteille tarkka paikannus on saavutettavissa vastaavalla menetelmällä kuten suoritetaan laserkeilaimen paikannus eli käyttämällä DGPS-paikannusta. Tällä menetelmällä päästään maastomittauksissa xy -tasossa noin 1 metrin paikannustarkkuuteen (Monteiro et al., 2005).

3.1 Koealojen otanta

Koealojen otannan tarkoituksena on valita sellaiset koealat, jotka edustavat mahdollisimman kattavasti inventoitavaa aluetta ja sen puustoa. Koealojen sijoittelussa liian lähekkäisten koealojen mittaaminen ei ole hyödyllistä, koska vierekkäisen ehkä jopa limittäiset alueet ovat usein keskenään liian samanlaisia eikä näin ollen tuo lisää informaatiota estimointiin. Erilaisia otantamenetelmiä ovat rajoittamaton satunnaisotanta, rajoittamaton systemaattinen otanta ja ositettu systemaattinen otanta (Kangas et al., 2006).

Rajoittamattomassa satunnaisotannassa koealat sijoitellaan inventoitavalle alueelle täysin satunnaisiin sijainteihin. Etuna satunnaisotannassa on yksinkertaisuus ja luotettavuuden arvioinnin helppous. Rajoittamattomassa systemaattisessa otannassa koealot sijoitellaan koko inventoitavalle alueelle tasavälein, jolloin koealojen sijoittelu ja mittaaminen on helpompaa. Ositettu systemaattinen otanta lienee tärkein kaukokartoitusinventoinnin kannalta, koska tässä menetelmässä käytetään jotain ennakkotietoa koealojen sijoitteluun. Sijoittelussa laserkeilaukseen pohjautuvan puustoinventoinnin kannalta on tärkeintä, että laserkeilausdatasta laskettavien selittävien muuttujien avaruus olisi mahdollisimman kattavasti edustettuna mitattavien koealojen joukossa. (Kangas et al., 2006)

Laserkeilausdatan ja muun kaukokartoitusaineiston käyttämistä sopivien koealojen etsimiseen on tutkittu esimerkiksi Duvemo et al., 2007. Otanta voidaan tällöin suunnitella esimerkiksi laskeamalla selittävät muuttujat laserkeilausdatasta rajoittamattoman systemaattisen otannan mukaisille mahdollisille koealoille ja valitsemalla tästä koealajoukosta sopivalla heuristiikalla ne koealat, jotka ovat halvimmalla mitattavissa ja kuvaavat selittävien muuttujien avaruutta mahdollisimman hyvin.

3.2 Puuston korkeusmallit

Maastomittauksessa koealoilla mitataan kaikkien läpimitaltaan yli 5 senttimetriä olevien puiden läpimitta, mittaus suoritetaan rinnankorkeudelta (1, 3 metriä). Kullakin koealalla mitataan näytepuilta myös pituudet, joita käyttäen sovitetaan lajikohtaiset läpimitta-pituus -mallit. Mallin perusteella lasketaan kaikille mitatuille puille mallinnetut pituudet. Usein käytettyjä läpimittakorkeus -malleja ovat Korf:n ja Näslund:n korkeusmallit, jotka voidaan sovittaa esimerkiksi pienimmän neliösumman sovituksella tai suurimman uskottavuuden estimoinnilla (Kangas et al., 2002, Mehtätalo, 2005).

Näslund:n korkeusmalli:

$$h(d) = \frac{d^\alpha}{(\beta_0 + \beta_1 \times d)^\alpha} + 1, 3, \quad (4)$$

jossa pituus h on läpimitan d funktio ja parametrit, joiden suhteen malli sovitetaan ovat α , β_0 ja β_1 .

Korf:n korkeusmalli:

$$h(d) = e^{A - B \times (d + \lambda)^{-C}}, \quad (5)$$

jossa pituus h on läpimitan d funktio ja parametrit, joiden suhteen malli sovitetaan ovat A , B , C ja λ , joka tulkitaan 1, 3 metrin korkeudelta mitatun läpimitan ja maan tasolla olevan läpimitan erotukseksi (Mehtätalo, 2005).

3.3 Estimoinnin selitettävät muuttujat

Selitettäviä muuttujia ovat metsätaloudellisesti tärkeät tunnusluvut, joita käytetään puuston arvon määrittämiseen, kehitysluokan määrittämiseen ja tarvittavien metsänhoidollisten toimenpiteiden suunnittelemiseen. Näitä yleisesti käytettäviä tunnuslukuja ovat puuston kokonaistilavuus, kokonaispohjapinta-ala, koealan mediaaniläpimittaisen puun pituus ja pohjapinta-ala, ja

kokonaisrunkoluku (Maltamo et al., 2004). Tunnusluvuista pohjapinta-ala ja tilavuus ovat laskettu mitatuista puiden läpimitoista ja kunkin puun mallinnetuista pituuksista.

| | | |
|-------|-----|---|
| y_1 | D | Koealalla olevien puiden mediaaniläpimitallisen puun läpimitta. Yksikkönä senttimetri. |
| y_2 | H | Koealalla olevien puiden mediaaniläpimitallisen puun pituus. Yksikkönä metri. |
| y_3 | G | Kaikkien koealan yli 5 senttimetriä läpimitaltaan olevien puiden yhteenlaskettu pohjapinta-ala. Tunnusluvun yksikkönä käytetään suhteellista lukua eli neliometriä/hehtaari |
| y_4 | V | Kaikkien koealan yli 5 senttimetriä läpimitaltaan olevien puiden yhteenlaskettu tilavuus. Yksikkönä suhteellinen tilavuus eli kuutiometriä/hehtaari. |
| y_5 | N | Runkolukuun lasketaan yli 5 senttimetriä läpimitaltaan olevien puiden lukumäärä. Yksikkönä kappaletta/hehtaari. |

Taulukko 2: Estimoinnin selittävät muuttujat

Kaikista näistä taulukon 2 tunnusluvuista voidaan kokonaistunnuslukujen lisäksi esittää lajikohtaiset tunnusluvut käytettäväksi estimoinnin selittävinä muuttujina. Suomen olosuhteissa estimoitavien lajien luokitteluun käytetään yleensä jakoa mänty-, kuusi- ja lehtipuihin (Laasenahto, 1982). Koealojen läpimittajakaumia ja korkeusjakaumia voidaan myös käyttää estimoinnin selittävinä muuttujina.

Selittävistä muuttujista pohjapinta-ala G lasketaan seuraavasti:

$$G = \sum_{i=1}^N \left(\pi \times \left(\frac{d_i}{2} \right)^2 \right) \times \frac{10000m^2}{A}, \quad (6)$$

jossa N on koealan puiden lukumäärä, d_i on i puun läpimitta ja A on koealan pinta-ala.

Tilavuus V lasketaan käyttäen runkokäyräyhtälöä, jotka ovat erikseen sovitettu laajempaan mitattuun aineistoon. Käytännössä runkokäyräyhtälöt ovat lajikohtaisesti sovitettuja. (Laasenahto, 1982)

$$V = \sum_{i=1}^N \left(\int_0^{h_i} \pi \times \left(\frac{d(h; h_i)}{2} \right)^2 dh \right) \times \frac{10000m^2}{A}, \quad (7)$$

jossa N on koealan puiden lukumäärä, h_i on i puun pituus, $d(h; h_i)$ on runkokäyräyhtälö, joka on h_i pituisen puun läpimitta korkeudella h ja A on koealan pinta-ala. Usein kuitenkin tilavuus lasketaan suoraan käyttäen erikseen sovitettua lajikohtaista tilavuusmallia, joka on rinnankorkeusläpimitan funktio.

4 ESTIMOINNIN MENETELMÄT

Puuston inventoinnissa estimointiongelman on estimoida selitettävät muuttujat selittävien muuttujien avulla. Käytettyjä estimaattoreita on k similaarisimman naapurin menetelmä eli kMSN ja harva bayesilainen regressio. K lähimmän naapurin menetelmä kNN on myös ollut käytössä (Sironen et al., 2001). Estimointi voidaan suorittaa alueellisesti tai yksinpuintulkintana, mutta estimoinnin periaate on molemmissa tapauksissa sama. Eroavaisuudet ovat tulosten yhdistämisessä ja käytettyjen selitettävien muuttujien määrittelyssä. Tässä työssä on keskitytty alueelliseen estimointiin. Estimaatit yhdistetään painotetulla keskiarvolla lopullisiksi tuloksiksi.

4.1 Muuttujien standardisointi

Sekä selitettävät, että selittävät muuttujat normalisoidaan tai standardisoidaan ja yhtenä normalisointimenetelmänä on siirtää data nollakeskiarvoiseksi ja jakaa se keskihajonnalla, josta nimitys standardisointi tai z-score standardisointi, jolloin etäisyysmittasuhteet eri muuttujien välillä saadaan vakioiksi.

Muuttujien standardisointi:

$$x_n = \frac{x_i - \bar{x}}{\sigma_x}. \quad (8)$$

Muuttujien palautus:

$$\hat{x}_i = \hat{x}_n \times \sigma_x + \bar{x}. \quad (9)$$

Yhtälöissä \bar{x} on x_i muuttujien keskiarvo ja σ_x on normaalihajonta. Tällöin muuttujien x_n varianssiksi tulee 1 ja keskiarvoksi 0. Muuttujien palautus on tarvetta tehdä pelkästään estimaateille \hat{x}_n .

4.2 K similaarisimman naapurin regressio

K lähimmän naapurin menetelmä (kNN) on monessa käyttötarkoituksessa esimerkiksi luokittelussa paljon käytetty menetelmä, jossa etsitään opetusjoukosta k kappaletta lähimpiä naapureita ja luokitellaan näyte naapureiden edustavimpaan luokkaan. Menetelmää voidaan käyttää myös regressiossa, jolloin estimoitava näyte on lähimpien naapureiden painotettu keskiarvo. Keskiarvon painotuksessa painokerroin on jokin etäisyyden funktio, esimerkiksi $w(d) = \frac{1}{d+1}$. (Altman,

1992)

K similaarisimman naapurin menetelmä (kMSN) on siis k lähimmän naapurin menetelmä, jossa etäisyysmittana käytetään Mahalanobis-etäisyyttä, jossa painokerroinmatriisi on johdettu selittävien ja selitettävien muuttujien kanonisten korrelaatioiden analyysistä. Moeur ja Stage esittivät similaarisimman naapurin menetelmän (MSN) (Moeur et al., 1995), johon k similaarisimman naapurin menetelmä on yleistys vastaavalla analogialla kuin k lähimmän naapurin menetelmä lähimmän naapurin menetelmälle (NN). Tämä similaarisimman naapurin menetelmän yleistys on esiintynyt etenkin metsätalouteen liittyvissä tutkimuksissa esimerkiksi (Sironen et al., 2001).

Similaarisimman naapurin menetelmässä etäisyysfunktio on muotoa:

$$(\hat{Y} - Y_i) \times W \times (\hat{Y} - Y_i), \quad (10)$$

jossa \hat{Y} on tuntemattoman näytteen selitettävien muuttujien rivivektori, Y_i on opetusjoukon näytteen i selitettävien (mitattujen) muuttujien rivivektori ja W on painokerroin matriisi, joka minimoi yhtälön kaikkien näytteiden $i = 1 \dots n$ suhteen \hat{Y} similaarisimmalle naapurille (Sironen et al., 2001).

MSN menetelmässä suhde estimoitavien selitettävien muuttujien ja opetusjoukon selitettävien muuttujien välillä korvataan selittävien muuttujien suhteella. Painokerroin matriisi tulee tällöin selitettävien ja selittävien muuttujien kanonisista korrelaatioista. Kanonisen korrelaation lineaarimuunnokset U_r ja V_r muodostetaan maksimoimaan niiden välinen korrelaatio löytämällä sopivat vektorit α_r ja γ_r ; $U_r = \alpha_r \times Y$ ja $V_r = \gamma_r \times X$, missä α_r on selittävien muuttujien kanoniset kertoimet ja γ_r on selitettävien muuttujien kanoniset kertoimet ($r = 1 \dots s$, s on mahdollisten U_r ja V_r parien lukumäärä). (Sironen et al., 2001)

Etäisyysfunktio on siten muotoa:

$$D_{ij}^2 = (X_i - X_j) \times \Gamma \Lambda^2 \Gamma' \times (X_i - X_j)', \quad (11)$$

missä

X_i on näytteen i selittävien muuttujien rivivektori, pituus p

X_j on näytteen j selittävien muuttujien rivivektori, pituus p

Γ on kanonisten kertoimien γ_r matriisi, koko $p \times s$

Λ^2 on kanonisten korrelaatioiden λ_r^2 diagonaalimatriisi, koko $s \times s$

Kanoniset korrelaatiot ja kertoimet lasketaan selittävien ja selitettävien muuttujien kovariansseista $\Sigma_{XX} = \text{cov}(X, X)$, $\Sigma_{XY} = \text{cov}(X, Y)$, $\Sigma_{YX} = \text{cov}(Y, X)$ ja $\Sigma_{YY} = \text{cov}(Y, Y)$.

$$\begin{aligned} (\Sigma_{YY}^{-\frac{1}{2}} \times \Sigma_{YX} \times \Sigma_{XX}^{-1} \times \Sigma_{XY} \times \Sigma_{YY}^{-\frac{1}{2}}) \times c &= \lambda^{\frac{1}{2}} \times c, \\ \Lambda_{ii} &= \lambda, \\ C_i &= c, \\ \Gamma &= (((\Sigma_{XX} \times \Sigma_{XY} \times \Sigma_{YY}) \times C)' \times \Sigma_{XX}^{-\frac{1}{2}})', \end{aligned} \tag{12}$$

missä matriisi Λ sisältää diagonaalilla ominaisarvojen $\lambda^{\frac{1}{2}}$ neliöt ja matriisi C ominaisvektorit c .

Etäisyysfunktion painokerroinmatriisia $\Gamma \Lambda^2 \Gamma'$ laskettaessa voidaan jättää pienimpien kanonisten kertoimien korrelaatiot kokonaan huomioimatta, jolloin s valitaan sen mukaan kuinka moni korrelaatio ylittää käytetyn kynnyksen (Sironen et al., 2001). Käytännössä pienimpien korrelaatioiden mukanaolosta on haittaa laskentatarkkuuden vuoksi.

Estimoitavan näytteen selitettävät muuttujat lasketaan similaarisimpien naapureiden selitettävien muuttujien painotetusta keskiarvosta. Similaarisimpien naapureiden regressiossa painokerroinfunktiona voidaan käyttää samaa kuin lähimpien naapureiden regressiossa.

$$\hat{Y}_j = \sum_{i=1}^k w_{ij} \times Y_i, \tag{13}$$

missä k on similaarisimpien naapureiden lukumäärä, Y_i on i similaarisimman naapurin selitettävät muuttujat ja w on jokin etäisyyden funktio (esimerkiksi $w = \frac{1}{D_{ij}+1}$).

4.3 Lineaarinen harva bayesilainen regressio

Lineaarinen regressiomalli kuvaa selitettävien muuttujien havaittuja arvoja selittävien muuttujien lineaarisen riippuvuuden avulla. Mallin lineaarisuus tarkoittaa sitä, että malli on lineaarinen selittävien muuttujien suhteen. (Daniel et al., 1975)

Lineaarinen regressiomalli on yleisesti muotoa:

$$Y = XW + \epsilon \tag{14}$$

$$\hat{Y} = XW \tag{15}$$

Mallin kertoimet voidaan ratkaista usealla eri tavalla. Yksinkertaisimmillaan painokertoimien

ratkaisemiseen käytetään pienimmän neliösumman menetelmää. Pienimmän neliösumman menetelmä on perinteinen menetelmä lineaarisen regressiomallin kertoimien laskemiseen. Menetelmässä minimoidaan havaintojen ja regressiosuoran välisen etäisyyden neliöiden summaa, jolloin saadaan virheiden neliösumman minimoivan regressiosuoran kertoimet (Daniel et al., 1975). Mikäli mallissa on suuri määrä selittäviä muuttujia, joista osa voi olla turhia, voi olla hyvä käyttää muuttujien valintaa tai jotain menetelmää, joka vähentää regressiossa käytettävien selittävien muuttujien määrää, esimerkiksi harvaa bayesilaista regressiota (Tipping, 2001).

Bayesilaisen päättelyn menetelmä perustuu Bayesin todennäköisyyskaavaan, josta voidaan optimoida halutut parametrit, joilla etsitty todennäköisyysjakauma toteutuu (Gelman et al., 2004). Menetelmä on laajalti hyväksytty ja todettu käyttökelpoiseksi moniin ongelmiin. Bayesin todennäköisyyskaavan mukaan todennäköisyys voidaan jakaa osiin (Gelman et al., 2004):

$$p(a|b)p(b) = p(b|a)p(a), \quad (16)$$

jossa $p(a)$ tarkoittaa tapahtuman a todennäköisyyttä ja $p(a|b)$ on a :n ehdollinen todennäköisyys, kun tapahtuma b on tapahtunut. Edellisestä kaavasta on yleensä vaikeaa määrittää havaintoaineiston todennäköisyys $p(b)$, koska se on integraali parametriavaruuden yli. Tästä tosin on etuna, että optimaalisia parametreja etsittäessä se todennäköisyys voidaan jättää huomioimatta - sen ollessa vakio kaikkialla parametriavaruudessa (Tipping, 2001).

$$p(a|b) = \frac{p(b|a)p(a)}{p(b)} \propto p(b|a)p(a) \quad (17)$$

Linearisessa regressiossa mallin lineaarinen aproksimaatio kohdevektorille selittävien muuttujien matriisista ϕ on:

$$t = \phi w + \epsilon, \quad (18)$$

jossa vektori $t = [t_{k,1}, \dots, t_{k,P}]^T$, jossa k on selitettävän muuttujan indeksi koealalla ja $\phi = [1_{P,1}, \phi_1, \dots, \phi_M]^T$ on $P \times M + 1$ matriisi sisältäen vakiosarakkeen ja mallissa käytetyt selittävät muuttujat koealalla. Matriisi $w = [w_{k,1}, \dots, w_{k,P}]^T$ on regressiomallissa käytettyjen painokertoimien matriisi ja mallin virhe on vektori $\epsilon = [\epsilon_{k,1}, \dots, \epsilon_{k,P}]^T$. Malli muodostetaan jokaiselle selitettävälle muuttujalle k erikseen riippumattomasti. (Junttila et al., 2008)

Linearisessa regressiomallissa olevan virhetermin ϵ oletetaan olevan normaalijakautunut.

$$\epsilon_i \sim N(0, \sigma^2) \quad (19)$$

Bayesilaisessa regressiossa pyritään maksimoimaan todennäköisyyttä, joka normaalijakaumassa kirjoitetaan tässä tapauksessa kirjoitetaan muotoon

$$p(t|w, \sigma^2) = \prod_{n=1}^P N(\phi_n w, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{P}{2}}} \exp -\frac{\|t - \phi w\|^2}{2\sigma^2}, \quad (20)$$

jossa todennäköisyyden keskiarvo on t :n estimaatti ja tuntematon hajonta on merkitty σ^2 . Suurin todennäköisyys saavutetaan täten minimoimalla $t - \phi w$ etäisyyksien neliösummaa.

Harvat mallit ja selittävien muuttujien vähentäminen on havaittu tehokkaaksi keinoksi välttää ylioppimista (Tipping, 2001). Michael E. Tipping on esittänyt algoritmin Bayesin uskottavuuden maksimointiin painokerrointen hyperparametrien avulla. Muita harvoja malleja hyödyntäviä menetelmiä ovat esimerkiksi tukivektorikone ja riippuvuusvektorikone. Tipping on kuitenkin osoittanut, että ylioppimisen välttämiseksi harva bayesilainen oppiminen on kuitenkin parempi kuin tukivektorikone ja riippuvuusvektorikone Tipping, 2001.

Maksimoitaessa uskottavuutta pelkästään parametrivektorin w suhteen on hyvin todennäköistä, että painokertoimien optimointi kärsii ylioppimisesta ja siten optimoitavaan todennäköisyysfunktioon lisätään painokertoimille priori-todennäköisyydet, joita hallitaan hyperparametreilla α (Tipping, 2001). Tällöin painokertoimet oletetaan nollakeskiarvoisiksi ja normaalijakautuneiksi. Hyperparametrit kontrolloivat painokertoimia, jos hyperparametri lähestyy ääretöntä, niin painokerroin lähestyy nollaa, jolloin sitä painokerrointa vastaava muuttuja voidaan poistaa mallista seuraavalle iteraatiolle ja täten harventaa regressiomallia eli vähentää käytettyjä muuttujia. (Junttila et al., 2008, Wipf et al., 2004)

$$p(w|\alpha) = \prod_{m=1}^M N(0, \alpha_m^{-1}). \quad (21)$$

Tästä saadaan esitettyä algoritmina iteratiivinen menetelmä II-tyypin uskottavuuden maksimoinnille.

1. Lasketaan standardisoidut arvot ϕ ja t opetusjoukosta.
2. Asetetaan alkuarvot:
 - (a) $\sigma^2 = 10$
 - (b) $\alpha_m = 1/(M + 1), m = 1, \dots, M + 1$
3. Iteroidaan kunnes konvergenssi löytyy:

- (a) $\Sigma^{-1} = A + \sigma^{-2} \phi^T \phi$, jossa A on hyperparametrien α diagonaalimatriisi.
- (b) $\mu = \sigma^{-2} * \Sigma \phi^T t$
- (c) $\alpha_m^{NEW} = \frac{\gamma_m}{\mu_m^2}$
- (d) $(\sigma^2)^{NEW} = \frac{\|t - \phi \mu\|^2}{P \cdot \Sigma_m \gamma_m}$

Kokeellisesti on havaittu, että riittävän pienillä hyperparametrien alkuarvauksilla ja riittävän suurella alkuvarianssilla, lopulliset arvot ovat alkuarvauksista riippumattomia. (Junttila et al., 2008)

5 VIRHE- JA LUOTETTAVUUSARVIOINTI

Estimaattorin validoinnissa tulee käyttää erillistä dataa opetukseen ja testaamiseen. Opetukseen käytettävää dataa käytetään regressiossa regressiomallin sovittamiseen ja k similaarisimpien naapureiden menetelmässä mahdollisten naapureiden joukkona. Testaamiseen käytettävä data on opetuksesta käytetystä datasta irrallinen ja siitä lasketuista estimaateista lasketaan estimointivirheet. Mikäli dataa on rajoitetusti ja halutaan mahdollisimman luotettavat virhemittat, estimointi on hyvä suorittaa käyttäen ristiinvalidointia virhemittojen kohinan minimoimiseksi. (Kohavi, 1995)

Puustoinventoinnille estimoinnin harha on tärkeää saada minimoitua, koska harhainen estimaattori tuottaa estimaatteihin systemaattista virhettä, joka säilyy ja korostuu lopullisia inventointituloksia laskettaessa pienempien estimointialueiden tuloksia yhdistämällä. Lopullisissa tuloksissa virheiden varianssi puolestaan pienenee.

5.1 Ristiinvalidointi

Ristiinvalidointi on yleinen käytäntö menetelmän hyvyyden arviointiin. Ideana on jakaa data erillisiin opetusjoukkoon ja testausjoukkoon useita kertoja, joista laskea yhteen estimointien virheet. Erilaisia ristiinvalidointimenetelmiä ovat esimerkiksi k -fold ja leave-one-out.

Ristiinvalidointi k -fold -menetelmällä

Jako opetusjoukkoon ja testausjoukkoon voidaan tehdä jakamalla data k kappaleeseen yhtä suurina joukkoja ja suorittamalla estimointi kullekin osajoukolle käyttäen aina opetukseen kaikkia muita osajoukkoja. Menetelmän laskennallinen vaativuus riippuu osajoukkojen lukumäärästä k . (Refaeilzadeh et al., 2009)

Ristiinvalidointi leave-one-out -menetelmällä

Laskennallisesti vaativampi menetelmä on leave-one-out -ristiinvalidointi, jossa opetus tehdään käyttäen kullekin datajoukon näytteelle erikseen käyttäen opetukseen koko datajoukkoa pois lukien estimoitava näyte. Leave-one-out -menetelmä on siis k -fold -menetelmän erikoistapaus, jossa k on opetusjoukon koko. (Refaeilzadeh et al., 2009) Puustoinventoinnissa virhearviointiin yleisesti vaaditaan käyttämään leave-one-out -ristiinvalidointia.

5.2 Virhemitat

Ristiinvalidoinnin tulosten tarkkuuden mittareina käytetään virheiden neliöiden keskiarvon juurta eli RMSE:tä ja estimointien harhaa (bias) eli virheiden keskiarvoa. Harhattomassa tilanteessa RMSE vastaa virheiden keskihajontaa. Näistä mittareista esitetään myös suhteellinen harha ja suhteellinen RMSE. (*Root-mean-square deviation*, 2010)

$$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n}, \quad (22)$$

$$bias\% = \frac{bias}{\bar{y}}, \quad (23)$$

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{n}, \quad (24)$$

$$RMSE\% = \frac{RMSE}{\bar{y}}, \quad (25)$$

$$var = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i - bias)^2. \quad (26)$$

Ylläolevissa kaavoissa n on näytteiden lukumäärä, \hat{y} on estimoitu tunnusluku, y on mitattu tunnusluku ja \bar{y} on mitattujen tunnuslukujen keskiarvo.

6 INVENTOINTISOVELLUS

Puustoinventointia varten toteutettiin sovellus, jolla lasketaan annetulle alueelle inventoinnissa käytettävät selittävät muuttujat ja estimoidaan selitettävät muuttujat käyttäen aiemmin mainittuja menetelmiä. Sovellus toteutettiin .net-kirjastona, jota voidaan käyttää omasta itsenäisestä ohjelmistaan tai ESRI ArcGIS -alustalle toteutetuista työkaluista. Laskennan kannalta kriittiset osat ovat toteutettu natiivina c++-ohjelmakoodina, johon toteutettiin rajapinta .net-koodia varten. Käyttöliittymät toteutettiin c#.net-ohjelmakoodina niin komentoriville kuin ESRI ArcGIS työkaluille. Ulkoisia natiivikoodista käytettyjä kirjastoja ovat boost, cvmlib, GDAL, OGR, TinyXml, MuParser, Optpp, SpatialIndex ja CryptoPP. Käytetyt kirjastot ovat kuvattuna taulukossa 3.

| | |
|--------------|---|
| boost | Boost kirjastot tarjoavat lukuisia apuvälineitä C++ -ohjelmointiin. Esimerkiksi hallitun muistinhallinnan, tiedostojärjestelmäapufunktiota, tietorakennesäiliöitä jne. |
| cvmlib | Cvmlib on kattava lineaarialgebra C++ -kirjasto, joka tarjoaa helppokäyttöisen rajapinnan BLAS ja LAPACK kirjastoihin. |
| GDAL | GDAL eli Geospatial Data Abstraction library, on C++ -kirjasto, joka tarjoaa samanlaisen abstraktin data mallin kaikille kirjaston tukemille geospaatialisille rasteriformaateille. |
| OGR | OGR-kirjasto sisältyy osittain GDAL-kirjastoon. OGR tarjoaa abstraktin datamallin kirjaston tukemille geospaatialisille vektoriformaateille. |
| TinyXml | TinyXml on pieni, helposti integroitavissa oleva C++ Xml-parseri ja xml-dokumentti datamalli. |
| MuParser | MuParser on helposti käytettävä ja tehokas matemaattisten yhtälöiden parseri. |
| Optpp | Optpp on helppokäyttöinen C++-kirjasto epälineaariseen optimointiin. |
| SpatialIndex | SpatialIndex -kirjasto tarjoaa helposti moneen tilanteeseen integroitavan spatiaalisen indeksoinnin sisältäen useita toteutuksia erilaisista puurakenteista. |
| CryptoPP | CryptoPP kirjasto tarjoaa hyvin korkean tason rajapinnan useimpiin kryptografisiin menetelmiin, joita voidaan hyödyntää käytettyjen parametrisointien salaamisessa. |

Taulukko 3: Sovelluksessa käytetyt ulkoiset kirjastot.

Toteutetuista kirjastoista ja sovelluksista on tehty erillisistä kaupallisista kirjastoista ja sovelluksista riippumattomia, jolla vältetään mahdolliset lisenssi-ongelmat ja käyttöoikeusongelmat. Esri ArcGIS -alusta puolestaan on kaupallinen, mutta sen ollessa monessa tapauksessa yleisesti käytetty paikkatietojärjestelmä eli GIS-alusta, joten sen päälle on myös hyödyllistä toteuttaa tarvittavat työkalut. Kuvassa 2 on kuvattuna sovellusympäristön kerrokset toteutettuine ja ulkoisine kirjastoineen.

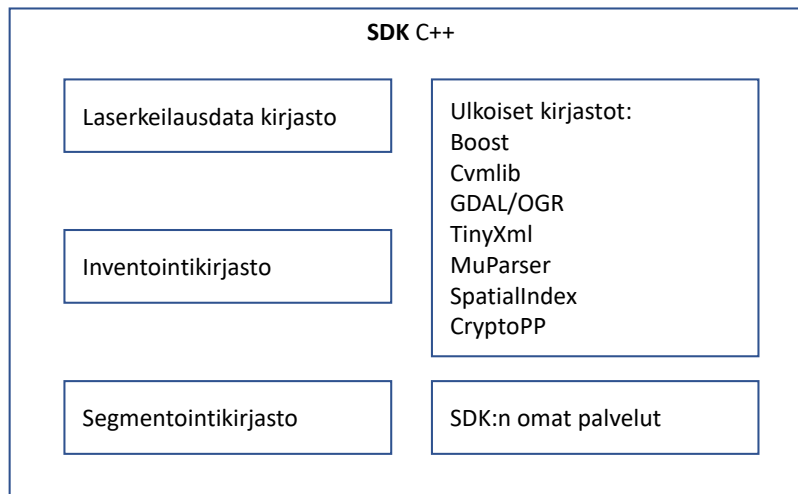
Käyttöliittymätaso:

Inventointityökalut C#.net

Sovellustaso:

Työkalukirjasto C#.net

Menetelmätaso:



Kuva 2: Inventointisovellusympäristön rakenne.

Toteutetuista inventointityökaluista tärkeimmät ovat: selittävien muuttujien laskeminen, inventoinnin validointi, ja inventoinnin laskeminen. Näiden lisäksi toteutettiin erinäisiä inventointiprojektin hallintaan käytettäviä työkaluja, joilla voidaan lisätä erilaisia aineistoja (koealoja, metsätalouskuvioita, parametreinteja) inventointiprojektiin.

Selittävien muuttujien laskeminen suoritetaan annetun inventointiprojektin koealoille ja metsätalouskuvioille, jolloin halutulla parametrien laskentamenetelmällä johdetaan kaikille koealoille ja metsätalouskuvioille vastaavat muuttujat esimerkiksi laserkeilausdatan pistepilvestä.

Inventoinnin validointi suorittaa inventoinnin hyödyntäen haluttua validointimenetelmää estimoimalla koealojen selitettäviä muuttujia parametroidin mukaisesti ja tuottamalla näistä tuloksista inventoinnin validointiraportin sisältäen tulosten statistiikat ja virheet.

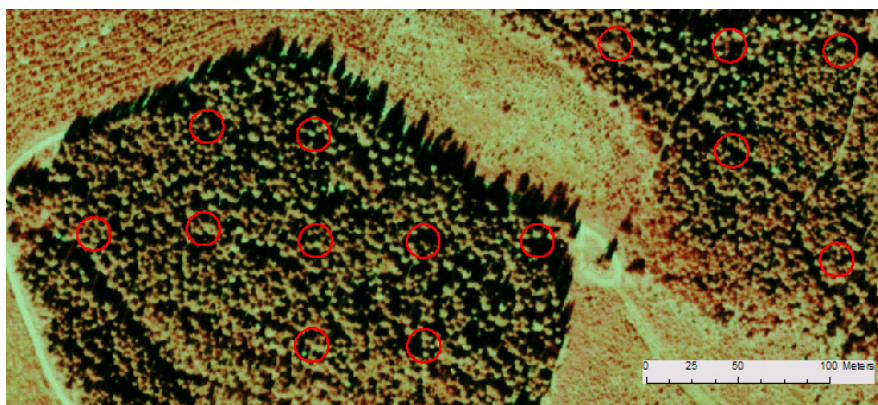
Inventoinnin laskemisessa tuotetaan estimaatit valittua menetelmää käyttäen halutun alu-

een metsätalouskuvioille. Metsätalouskuviot ovat sisäisesti jaettu ruudukkoon, jossa kunkin ruudun pinta-ala vastaa koealan pinta-alaa. Näille laskentaruuduille suoritetaan vastaava inventointi kuin validoinnissa. Lopulliset tulokset metsätalouskuvioille saadaan yhdistämällä metsätalouskuvion sisällä olevan laskentaruudukon estimaattien tulokset laskemalla painotettuja keskiarvoja mahdollisesti inventointiprojektin parametroidin mukaisesti.

Sovelluksen osat toteutettiin vaihdettaviksi rajapintojen kautta, jolloin uusia menetelmiä voidaan lisätä helposti esimerkiksi selitettävien muuttujien laskentaan tai inventoinnin estimaattoriksi. Inventointiprojekteissa käytettävät aineistot ovat usein tilallisesti suuria ja siksi sovellus pyrkii hyödyntämään spatiaalisia indeksointeja. Laserkeilausdata tai ilmakeilausaineisto inventointiprojektille voi olla satoja gigatavuja, ja inventoitavia metsätalouskuvioita pienempine yksikköineen voi myös olla usein satoja tuhansia.

7 TUTKIMUSAINEISTOT JA TULOKSET

Työssä käytetyissä Matalansalon ja Juukan koetilojen aineistoissa maastossa mitattujen koealojen keskipisteet ovat paikannettu GPS:ää käyttäen noin 1 metrin tarkkuudella. Kukin mitattu koeala on 9 metriä säteeltään oleva ympyrä, jonka alueelta on mitattu kaikkien rinnankorkeudelta (1, 3 metrin korkeudelta) läpimitaltaan yli 5 senttimetriä olevien puiden läpimitat. Kuvassa 3 on ilmakuva, jossa näkyy muutama Matalansalon koeala. Mitatuista puista on kirjattu puulaji ja rinnankorkeusläpimitta; joillekin koealan puille on lisäksi mitattu korkeus ja muiden puiden korkeudet on mallinnettu koeala- ja lajikohtaisesti sovittamalla Näslundin korkeusmalli koealan mitattujen puiden korkeuksiin. Metsätalouden kokonaistunnusluvut ja lajikohtaiset tunnusluvut on laskettu koealan mitatuista ja mallinnetuista puista.



Kuva 3: Ilmakuva muutamien Matalansalon koealan alueelta.

7.1 Matalansalon koetila

UPM-Kymmene Oyj:n Matalansalon koetila sijaitsee Pohjois-Savossa, Varkaudessa. Alueen pinta-ala on noin 1200 hehtaaria ja tila vastaa olosuhteiltaan ja kehitysluokkakajakaumaltaan tyyppillistä Järvi-Suomen metsätilaa. Alueen puusto on tosin tilan historian vuoksi keskimääräistä vähemmän hoidettua.

Alueelta on mitattu 474 koealaa vuonna 2004, jotka sijaitsevat 67 eri metsätalouskuvion alueella. Puusto tällä alueella on hyvin havupuuvaltaista; mäntyvaltaisten koealojen osuus oli 57%, kuusivaltaisia koealoja oli 34% ja lehtipuuvaltaisia koealoja oli 8%.

Alueen laserkeilaus on suoritettu lentokoneella 3.8.2004 laserkeilaukseen erikoistuneen Norjalaisen Blom Norkart Mapping AS:n toimesta. Mittauslaitteistona on käytetty Optech:n ALTM

2033 laserkeilainta. Laserkeilausdatasta on tallennettu ensimmäiset ja viimeiset paluupulssit, joita keskimäärin on 0,7 pulssia/m² eli aineisto on kohtalaisen harvaa. FM-kartta Oy on suorittanut laserkeilausdatalle esikäsitteilyn ja muodostanut aineistosta digitaalisen maastomallin.

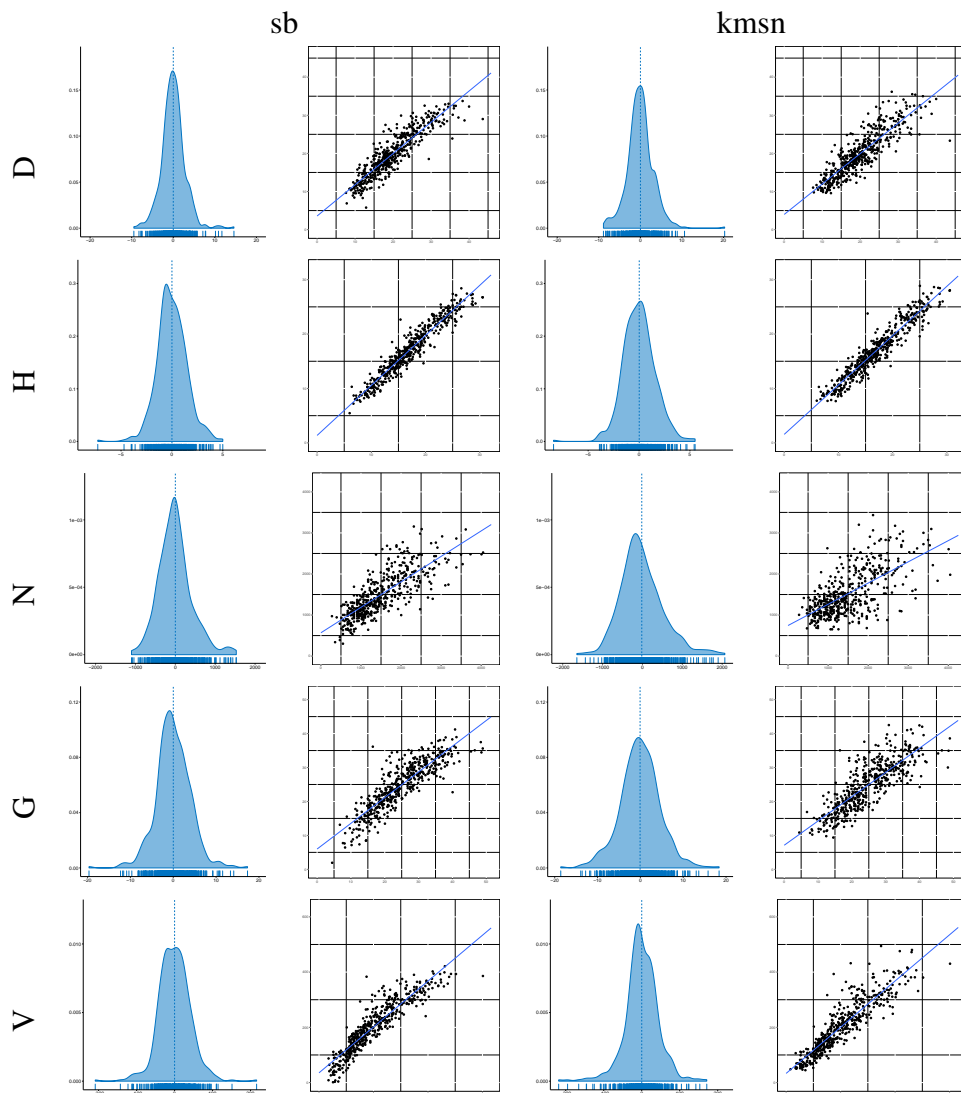
Matalansalon laserkeilausdatasta on laskettu esitetyt selittävät muuttujat, joita käyttäen on laskettu kumpaakin työssä esitettyä estimaattoria käyttäen koelakohtaiset estimaatit selitettävistä maastomittauksista lasketuista muuttujista. Taulukoissa 4 ja 5 esitetään inventoinnin validoinnin tulokset ja kuvassa 4 esitetään inventoinnin eri estimoitujen muuttujien virhejakaumat.

| | <i>Keskiarvo</i> | | <i>harha</i> | | <i>harha%</i> | |
|----------|------------------|---------|--------------|--------|---------------|--------|
| | sb | kmsn | sb | kmsn | sb | kmsn |
| <i>D</i> | 21,04 | 21,05 | 0,00 | 0,01 | 0,02% | 0,06% |
| <i>H</i> | 18,15 | 18,08 | 0,00 | 0,02 | 0,01% | 0,10% |
| <i>N</i> | 1506,55 | 1524,04 | 0,33 | -17,16 | 0,02% | -1,14% |
| <i>G</i> | 24,67 | 24,75 | 0,01 | -0,07 | 0,03% | -0,28% |
| <i>V</i> | 203,41 | 203,34 | -0,04 | 0,03 | -0,02% | 0,01% |

Taulukko 4: Matalansalon koelajojen estimointien keskiarvot ja harhat tunnusluville. Estimaattoreina sb (lineaarinen harva bayesilainen regressio) ja kmsn (k similaarisimman naapurin regressio).

| | <i>Hajonta</i> | | <i>RMSE</i> | | <i>RMSE%</i> | |
|----------|----------------|--------|-------------|--------|--------------|--------|
| | sb | kmsn | sb | kmsn | sb | kmsn |
| <i>D</i> | 5,80 | 5,98 | 2,77 | 3,11 | 13,11% | 14,72% |
| <i>H</i> | 4,78 | 4,83 | 1,46 | 1,56 | 8,03% | 8,61% |
| <i>N</i> | 557,81 | 549,19 | 424,91 | 534,00 | 28,20% | 35,44% |
| <i>G</i> | 7,03 | 7,04 | 4,00 | 4,65 | 16,21% | 18,84% |
| <i>V</i> | 95,24 | 95,78 | 41,76 | 44,25 | 20,53% | 21,76% |

Taulukko 5: Matalansalon koelajojen estimointien keskihajonnat ja keskineliövirheet (RMSE) tunnusluville. Estimaattoreina sb (lineaarinen harva bayesilainen regressio) ja kmsn (k similaarisimman naapurin regressio).



Kuva 4: Matalansalon koealojen selitettävien muuttujien estimointivirheiden jakaumat ja estimaattien hajontakuviot estimaattoreita sb ja kmsn käyttäen.

7.2 Juukan koealat

Juukan koetila sijaitsee Pohjois-Karjalassa ja siellä mittaukset on tehty Metsähallituksen metsätaloustalokuvioille. Käsitellyn alueen kokonaispinta-ala on noin 10000 hehtaaria ja tila vastaa olosuhteiltaan ja kehitysluokkajakaumaltaan tyypillistä Itä-Suomalaista metsätilaa. Alueen puusto on tosin tilan historian vuoksi keskimääräistä vähemmän hoidettua.

Alueelta on mitattu 506 koealaa peräkkäisinä kesinä 2005 ja 2006. Puusto Juukan koealojen alueella on hyvin havupuuvaltaista; mäntyvaltaisten koealojen osuus oli 73%, kuusivaltaisia koealoja oli 21% ja lehtipuuvaltaisia koealoja oli 6%.

Alueen laserkeilaus on suoritettu lentokoneella 13.7.2005 laserkeilaukseen erikoistuneen FM-kartta Oyn toimesta. Mittauslaitteistona on käytetty Optech:n ALTM 3100C laserkeilainta. Laserkeilausdatasta on käytetty ensimmäiset ja viimeiset paluupulssit, joita keskimäärin on 0,6 pulssia/m² eli aineisto on hieman harvempaa kuin Matalansalon laserkeilausdata. FM-kartta Oy on suorittanut laserkeilausdatalle esikäsittelyn ja muodostanut aineistosta digitaalisen maastomallin.

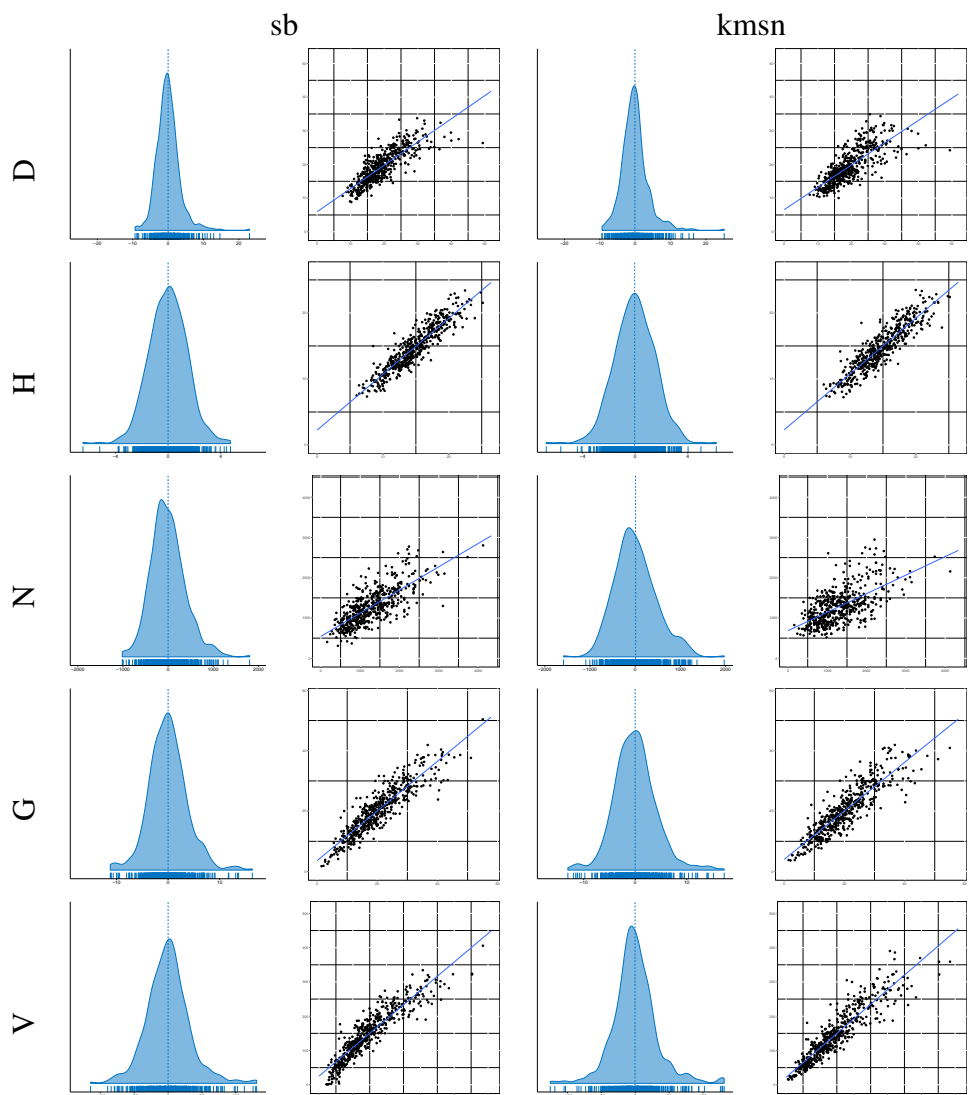
Juukan laserkeilausdatasta on laskettu esitetyt selittävät muuttujat, joita käyttäen on laskettu kumpaakin työssä esitettyä estimaattoria käyttäen koealakohtaiset estimaatit selitettävistä muuttujista. Taulukoissa 6 ja 7 esitetään inventoinnin validoinnin tulokset ja kuvassa 5 esitetään inventoinnin eri estimoitujen muuttujien virhejakaumat.

| | <i>Keskiarvo</i> | | <i>harha</i> | | <i>harha%</i> | |
|----------|------------------|---------|--------------|-------|---------------|--------|
| | sb | kmsn | sb | kmsn | sb | kmsn |
| <i>D</i> | 20,61 | 20,69 | 0,00 | -0,09 | -0,02% | -0,42% |
| <i>H</i> | 15,70 | 15,72 | 0,00 | -0,04 | -0,02% | -0,27% |
| <i>N</i> | 1284,97 | 1276,23 | -0,34 | 8,39 | -0,03% | 0,65% |
| <i>G</i> | 20,33 | 20,35 | 0,00 | -0,01 | 0,00% | -0,07% |
| <i>V</i> | 145,48 | 145,76 | -0,02 | -0,29 | -0,01% | -0,20% |

Taulukko 6: Juukan koealojen estimointien keskiarvot ja harhat tunnusluville. Estimaattoreina sb (lineaarinen harva bayesilainen regressio) ja kmsn (k similaarisimman naapurin regressio).

| | <i>Hajonta</i> | | <i>RMSE</i> | | <i>RMSE%</i> | |
|----------|----------------|--------|-------------|--------|--------------|--------|
| | sb | kmsn | sb | kmsn | sb | kmsn |
| <i>D</i> | 4,89 | 5,15 | 3,28 | 3,76 | 15,89% | 18,18% |
| <i>H</i> | 3,28 | 3,28 | 1,44 | 1,55 | 9,20% | 9,88% |
| <i>N</i> | 446,31 | 433,39 | 383,72 | 463,30 | 29,87% | 36,07% |
| <i>G</i> | 7,91 | 7,90 | 3,70 | 4,08 | 18,22% | 20,05% |
| <i>V</i> | 75,00 | 75,45 | 31,74 | 32,43 | 21,82% | 22,30% |

Taulukko 7: Juukan koealojen estimointien keskihajonnat ja keskineliövirheet (RMSE) tunnusluville. Estimaattoreina sb (lineaarinen harva bayesilainen regressio) ja kmsn (k similaarisimman naapurin regressio).



Kuva 5: Juukan koalojen selitettävien muuttujien estimointivirheiden jakaumat ja estimaattien hajontakuviot estimaattoreita sb ja kmsn käyttäen.

8 YHTEENVETO JA JOHTOPÄÄTÖKSET

Työssä on käyty läpi toteutettu sovellus ja siinä käytettävät estimointimenetelmät, joilla pystytään tuottamaan metsävaratiedon kannalta oleellisia tunnuslukuja kaukokartoituksella eli tässä työssä laserkeilausdatan perusteella käyttäen maastossa suoritettuja koealamittauksia opetusaineistona. Näitä tunnuslukuja voidaan hyödyntää metsän toimenpidesuunnittelussa, metsän kehitysluokan ja arvon arvioinnissa.

Tuloksista voidaan päätellä, että lineaarinen harva bayesilainen regressio toimii hyvin regressio-kertoimien sovittamiseen ja käytössä olevien selittävien muuttujien valitsemiseen estimointia varten. Estimointivirheiden osalta näissä kahdessa estimointimenetelmässä ei ole suurta eroa, mutta harva bayesilainen regressio tuottaa kuitenkin hieman harhattomampia estimaatteja, jolloin estimointituloksia yhdistettäessä suuremmalle alueelle estimoinnin kokonaisvirhe pienee.

K similaarisimman naapurin menetelmän etuna on kuitenkin eri estimoitavien muuttujien keskinäinen johdonmukaisuus, koska tässä menetelmässä kaikki estimoitavat muuttujat lasketaan kullekin estimointiyksilölle kerralla. Tässä työssä käytettyjen koetilojen aineistoilla ei ole kuitenkaan tulosten valossa varsinaista syytä hyödyntää tätä estimointimenetelmää. Tapauskohtaisesti tämäkin menetelmä voi olla käyttökelpoinen. Lineaarisen harvan bayesilaisen regressioon estimointituloksia voitaisiin yhdenmukaistaa jälkikäsittelemällä, mikäli tuloksissa on keskenään ristiriitaisia estimaatteja esimerkiksi pituuden ja pohjapinta-alan suhteen.

Molemmilla menetelmillä on kuitenkin näillä tässä työssä käytetyillä aineistoilla saavutettu hyväksyttävän pieniä estimointivirheitä. Tulokset ovat selvästi tarkempia kuin perinteisillä menetelmillä saatavat inventointitulokset.

Mittalaitteiden parantuminen mahdollistaa laserkeilauksen korkeammalla lentokorkeudella, jolloin samalla lentomatalla voidaan mitata suurempia pinta-aloja. Tämä laskee kaukokartoitusmenetelmien kustannuksia entisestään eikä perinteisille menetelmille ole mahdollisia poikkeustapauksia lukuun ottamatta käytännön perusteita.

LÄHTEET

- ALTMAN, N. S., 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*. Vol. 46, nro 3, s. 175.
- BARTELS, M.; WEI, Hong; MASON, D.C., 2006. DTM Generation from LIDAR Data using Skewness Balancing. Teoksessa: *DTM Generation from LIDAR Data using Skewness Balancing. 18th International Conference on Pattern Recognition (ICPR'06)*. IEEE.
- DANIEL, Wayne D.; TERRELL, James C., 1975. *Business Statistics: Basic Concepts and Methodology*. Boston: Houghton Mifflin Company.
- DUVEMO, Karl; BARTH, Andreas; WALLERMAN, Jörgen, 2007. Evaluating sample plot imputation techniques as input in forest management planning. *Canadian Journal of Forest Research*. Vol. 37, nro 11, s. 2069–2079.
- GELMAN, Andrew; CARLIN, John B.; STERN, Hal S.; RUBIN, Donald B., 2004. *Bayesian Data Analysis*. 2nd ed. Chapman ja Hall/CRC.
- HAARA, Arto; KORHONEN, Kari T., 2004. Kuvioittaisen arvioinnin luotettavuus. Vol. 4.
- HEIKKILÄ, J.; ÄRÖLÄ, E.; S., Kilpiäinen, 2010. *Kaukokartoitusperusteisen metsien inventoinnin koalojen maastotyöopas*. Metsäkeskus.
- HYYPPÄ, Juha; INKINEN, M., 1999. Detecting and estimating attributes for single trees using laser scanner. *Photogrammetric Journal of Finland*. Vol. 16, s. 27–42.
- JUNTTILA, Virpi; MALTAMO, Matti; KAURANNE, Tuomo, 2008. Sparse Bayesian Estimation of Forest Stand Characteristics from Airborne Laser Scanning. *Forest Science*. Vol. 54, nro 5, s. 543–552. ISSN 0015-749X.
- KANGAS, Annika; MALTAMO, Matti, 2002. Anticipating the variance of predicted stand volume and timber assortments with respect to stand characteristics and field measurements. *Silva Fennica*. Vol. 36, nro 4.
- KANGAS, Annika; MALTAMO, Matti, 2006. *Forest inventory: methodology and applications*. Dordrecht: Springer. ISBN 9781402043796.
- KOHAVI, Ron, 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Teoksessa: *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. Morgan Kaufmann, s. 1137–1143.
- LAASENAHO, Jouko, 1982. *Taper curve and volume functions for pine, spruce and birch*. Helsinki: Metsäntutkimuslaitos.
- LAS Specification Version 1.2*, 2010 [https://www.asprs.org/wp-content/uploads/2010/12/asprs_las_format_v12.pdf]. luettu: 23.1.2010.

- MALTAMO, Matti; KANGAS, Annika, 2004. *Metsän mittaus ja kartoitus*. Joensuu: Silva Carelica 40. Joensuun Yliopisto. ISBN 9526103009.
- MEHTÄTALO, Lauri, 2005. Height-Diameter Models for Scots Pine and Birch in Finland. *Silva Fennica*. Vol. 39.
- MOEUR, Melinda; STAGE, Albert R., 1995. Most Similar Neighbor: An Improved Sampling Inference Procedure for Natural Resource Planning. *Forest Science*. Vol. 41, s. 337–359.
- MONTEIRO, Luís Sardinha; MOORE, Terry; HILL, Chris, 2005. What is the accuracy of DGPS? *Journal of Navigation*. Vol. 58, nro 2, s. 207–225.
- NÆSSET, Erik, 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*. Vol. 80, s. 88–99.
- NÆSSET, Erik et al., 2004. Laser scanning of forest resources: the nordic experience. *Scandinavian Journal of Forest Research*. Vol. 19, nro 6, s. 482–499.
- PAHKINEN, Erkki; LEHTONEN, Risto, 1989. *Metsän mittaus ja kartoitus*. Helsinki: Gaudeamus. ISBN 9516624707.
- REFAEILZADEH, Payam; TANG, Lei; LIU, Huan, 2009. Cross-Validation. Teoksessa: *Cross-Validation. Encyclopedia of Database Systems*. Springer US, s. 532–538.
- Root-mean-square deviation*, 2010 [https://en.wikipedia.org/wiki/Root-mean-square_deviation]. luettu: 20.1.2010.
- SIRONEN, Susanna; KANGAS, Annika; MALTAMO, Matti; KANGAS, Jyrki, 2001. Estimating Individual Tree Growth with the k-Nearest Neighbour and k-Most Similar Neighbour Methods. *Silva Fennica*. Vol. 35.
- TIPPING, Michael E., 2001. Sparse Bayesian Learning and the Relevance Vector Machine. *J. Mach. Learn. Res.* Vol. 1, s. 211–244. ISSN 1532-4435.
- WEHR, Aloysius; LOHR, Uwe, 1999. Airborne laser scanning—an introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*. Vol. 54, nro 2-3, s. 68–82.
- WIPF, D.P.; RAO, B.D., 2004. Sparse Bayesian Learning for Basis Selection. *Trans. Sig. Proc.* Vol. 52, nro 8, s. 2153–2164. ISSN 1053-587X.