

TARTU ÜLIKOOL
BIOLOOGIA- GEOGRAAFIATEADUSKOND
Geograafia Instituut

RIIGIMETSA TAKSEERANDMETE KASUTAMINE
EESTI METSADE KAUGSEIRES TEHISÕPPE
RAKENDUSE ABIL

Magistritöö kartograafias- ja geoinformaatikas

Tanel Tamm

Juhendaja
TÜ geograafia instituudi
geoinformaatika teadur
Kalle Remm

TARTU 2005

SISSEJUHATUS	3
1. KIRJANDUSE ÜLEVAADE	5
1.1. NÄIDISTEL PÕHINEV JÄRELDAMINE	5
1.2. LOKAALSTATISTIKUD.....	6
1.3. METSADE KAUGSEIRE	6
2. METOODIKA.....	14
2.1. UURIMISALA KIRJELDUS	14
2.2. ANDMETE PÄRITOLU JA ETTEVALMISTUS	14
2.2.1. Kaardiandmed	15
2.2.2. Pildiaandmed.....	16
2.2.3. Metsakorralduse takseerandmed	18
2.2.4. Tarkvara	21
2.3. TUNNUSTE VALIMINE TEHISÖPPEKS	22
2.4. PARIMAD TUNNUSED ORTOFOTODELT	23
2.5. PARIMAD TUNNUSED SATELLIITPILTIDELT JA MULLAKAARDILT.....	25
2.6. TUNNUSTE VÕRDLUS.....	26
2.7. PROGNOOSIKAARTIDE KOOSTAMINE	26
2.8. TÄPSUSHINNANGUD	27
3. TULEMUSED JA ARUTELU	30
3.1. TUNNUSTE EELVALIK	30
3.2. PARIMAD PROGNOOSID JA TUNNUSED ORTOFOTODELT	31
3.2.1. Parimad tunnused.....	32
3.2.2. Tunnuse komponentide võrdlus.....	33
3.3. PARIMAD PROGNOOSID JA TUNNUSED SATELLIIDIPILDILT JA MULLAKAARDILT	35
3.3.1. Parimad tunnused.....	36
3.3.2. Tunnuse komponentide võrdlus.....	37
3.4. PROGNOOSIKAARTIDE KOOSTAMINE	39
KOKKUVÕTE	44
VIITED	45
SUMMARY	48

Sissejuhatus

Informatsiooni olemasolu loodusressursside kohta on väga vajalik nende edukaks majandamiseks ja kaitsmiseks. Nimetatud informatsiooni hankimiseks kasutatakse tänapäeval väga tihti kaugseire ja GIS-i vahendeid. Vaatamata metsandussektori küllaltki olulisele rollile Eesti majanduses (22,5 % kogueksportist ja 6,5 % SKT-st) on 62 % metsamaal (erametsad, tagastamata ja erastamata metsad) seire puudulik (Aastaraamat Mets '01: 10.05.2005). Võttes Eesti metsamaa pindalaks kokku 2,3 mln ha (Aastaraamat Mets '01: 10.05.2005) ja eeldades, et üks takseerandmete koguja jõuab päevas keskmiselt koguda andmeid 30 ha kohta (Mäkelä & Pekkarinen 2004), kuluks sajal takseerijal puudulikkude andmetega metsamaa läbivaatamiseks 475 päeva. Viiesaja kroonise päevapalga juures oleksid kulutused töötasudele ~24 mln krooni. Seetõttu on vajadus metsade kirjeldamise efektiivsete ja majanduslikult otstarbekate meetodite järele suur.

Käesolev magistritöö käsitleb ühte võimalikku viisi metsamajandamise seisukohalt olulise informatsiooni hankimiseks. Uurimuses püütakse ära kasutada riigimetsade hindamiseks kogutavaid takseerandmeid metsamajanduslikult oluliste tunnuste prognoosimiseks kaugseire- ja kaardiandmete ning näidistel põhineva järeldamise süsteemi abil. Kuigi metsaparameetrite ruumilise jaotuse modelleerimiseks üldiste ja kõikjalkehtivate seaduspärade leidmine on problemaatiline, on seda kirjeldatud hulgaliste välivaatluste (näiteks (edaspidi N) riigimetsa takseerandmed) ja eksperimentidega. Just selliste probleemide lahendamiseks, kus üldiste seaduspärade asemel esineb palju üksikjuhtumite lahendusi, sobib hästi siinses uurimuses kasutatud näidistel põhineva järeldamise (*Case-Based Reasoning* — CBR) meetodika. Nimetatud meetodika toimib nii, et uues olukorras otsuse langetamiseks otsitakse sellele olukorrale kõige sarnasemat näidist või näidiseid, mille parameetrid on ära kirjeldatud. Uurimistöös loodud näidistel põhineva järeldamise süsteemis kasutatakse prognoosi andmiseks vajalike sarnaste näidiste ülesleidmiseks tehisoõpet.

Magistritööle on seatud järgmised eesmärgid.

- Sobiva meetodi leidmine tehisoõppele eelnevaks tunnuste valikuks.
- Leida kasutatavate lähteandmete ja meetodika abil saavutatavad parimad prognoosid arenguklassi, enamuspüüliigi, puistu juurdekasvu ja tagavara jaoks.

- Leida prognoosimisel kasutatavad parimad tunnused (lokaalstatistiku, andmekihi/kanali ja raadiuse kombinatsioon).
- Võrrelda raadiuste, indeksite ja andmekihtide indikaatorväärtusi.
- Metsaparameetrite prognoosikaartide koostamine, kasutades satelliitpilte ja mullakaarti sisaldavat näidistel põhineva järeldamise süsteemi.

1. Kirjanduse ülevaade

1.1. Näidistel põhinev järeldamine

Käesolevas uurimistöös kasutatud näidistel põhineva järeldamise meetodikat on defineeritud kui multidistsiplinaarset teadusharu, mis põhineb eelnenud kogemuste ärakasutamisel (Aha 1998). Ülevaate näidistel põhinevast järeldamisest, tehisõppest ning kasutusel olevatest mõistetest võib leida minu bakalaureusetööst ja minu juhendaja poolt publitseeritud artiklist (Tamm 2003, Remm 2004). Järgnevalt on esitatud lühike kokkuvõte CBR-süsteemidest ja tehisõppest.

CBR süsteemide puhul hoitakse teadmisi näidistebaasis. Eesmärgiks ei ole mudeli loomine, vaid antud probleemile võimalikult täpse lahenduse väljapakkumine. Äärmiselt raske on reeglitega täielikult ära kirjeldada kompleksseid muutujaid ja mitmeseid seoseid. Samas on aga tihti võimalik tuua üksikuid näiteid, kus on suudetud formaliseerida vähemalt teatud kindlale situatsioonile vastav probleem. Selliseid formaalselt lahendatud probleeme saab ära kasutada sarnaste olukordade ilmnemisel.

Mida sarnasemad on vaatlused treeningandmetes, seda usaldusväärsemateks ennustajateks ehk näidisteks neid üksteise suhtes peetakse. Oletust, et sarnaste juhtumitega kaasnevad sarnased tulemused, on püütud formaliseerida. Sellel eesmärgil on ühe võimaliku mudeli välja pakkunud Hüllermeier (2001). Ei ole olemas universaalset sarnasuse mõõtu nii pidevate kui nominaalsete muutujate jaoks. Ülevaate sarnasuse funktsioonidest on andnud Wilson ja Martinez (2000).

Näidistel põhinev järeldamine on leidnud rakendust paljudes valdkondades. Neid iseloomustavad teatud ühised tunnusjooned: reeglid ei ole ranged, üldiste järelduste asemel on suur hulk üksikjäreldusi ja uuringuid ning esinevad suured andmebaasid asetleidnud ja lahendatud juhtumite kohta.

Näidistel põhineva järeldamise süsteemides kasutatakse prognoosi andmiseks vajalike sarnaste näidiste ülesleidmiseks sageli tehisõpet. Tehisõppeks (*machine learning*) nimetatakse protsessi, millega funktsionaalüksus täiustab oma talitlust uute teadmiste ja/või oskuste omandamise või olemasolevate ümberkorraldamise kaudu.

Käesolevas töös on kasutatud CBR-rakenduste seast selgelt eristatavat tehisõppe meetodit, mida nimetatakse laiskõppeks (*lazy learning*). Laiskõppe on induktiivse

tehisõppe vorm. Induktiivne tehisõpe põhineb dünaamilise mälu teoorial (Schank 1982). Paralleelselt terminiga laiskõpe on kasutatud ka terminit laisk õpe. Laiskõppe algoritmide kasutamine CBR-süsteemide treenimismeetodina on väga sage. Laiskõppe algoritmid hoiavad teadmisi (treeningvaatluseid) alguses vormis. Probleemi lahendamise ajal ei toimu informatsiooni kombineerimist, ega mudelite jaoks teadmiste üldistamist ning vahepealsete tulemuste kasutamist.

1.2. Lokaalstatistikud

Lähiümbruse arvestamise vajalikkust on keskkonnamodelleerimises tõdenud näiteks Remm ja Luud (2003), metsade kaugseires Kilpeläinen ja Tokola (1998) ning Tuominen ja Pekkarinen (2005). Lokaalstatistikuid arvutatakse libiseva akna piires, mis liigub üle pildiandmete. Pildiandmetes esinevat tekstuuri saab jagada statistiliseks ja struktuurseks (Haralick 1979). Statistiline kujutise tekstuur kirjeldab väärtuste statistilist jaotust, struktuurne aga ruumilist jaotust, mis tuleneb väärtuste asukohast üksteise suhtes. Lihtsad statistilise tekstuuri näitajad on standardhälve ja dispersioon, mida arvutatakse teatud suurusega akna piires (Herold *et al.* 2004). Käesolevas töös on kasutatud lokaalselt arvutatud keskmist, üle keskmise väärtuse osakaalu, standardhälvet, variatsioonikoefitsienti ja moodi ning struktuuri kirjeldavatest statistikutest Morani I'd ja kauguse pöördväärtusega kaalutud Morani I'd.

Sageli kasutust leidvad struktuuri näitajad on kohalikult arvutatud variogrammid ja korrelogrammid (Muinonen *et al.* 2001). Nende meetodite eeliseks on see, et nad kirjeldavad nii kohalike väärtuste varieeruvust, kui suhtelist paiknemist. Nõrkuseks on aga see, et kasutatud vaatluste või pikslite arvu suurenedes suureneb eksponentsiaalselt arvutuste maht. Lokaalstatistikute arvutamisel tuleb eelnevalt ära otsustada, millise kuju ja suurusega aknas ning millise aknasisesega kaalujaotusega seda teha (Remm in press).

1.3. Metsade kaugseire

Aristoteles kirjutas kaheosalisest protsessist: kujutlusvõimet kasutav pärimine ja avastamine ning avastuse demonstreerimine, selgitamine ja tõestamine (Britannica 1989, viidatud Skidmore 2002 järgi). Avasta ja tõesta printsiip on teadusliku käsitluse aluseks ja seda kasutatakse universaalsena ka keskkonnamudelite loomisel GIS-s. Keskkonnamudelid simuleerivad looduslike protsesside funktsioneerimist. Nende

kasutamisel on peamiselt kaks eesmärki: parandada arusaamist ökosüsteemi toimimisest ja prognoosida keskkonnaningimusi ajas ja ruumis (Skidmore 2002).

Üldistatud tähenduses võib õppimine ja ka modelleerimine olla, kas induktiivne (järeldamine üksikult üldisele — teatud hulgast faktidest tehakse üldine otsus) või deduktiivne (üksiku tuletamine üldisest — üldistest teesidest/eeldustest luuakse konkreetne väide). Induktiivset järeldamist peetakse teadusliku meetodi lahutamatuks osaks, eriti loodus- ja sotsiaalteadustes. Tabel 1 annab ülevaate tänapäeval enimkasutatavatest keskkonnamodelleerimise meetoditest.

Tabel 1. Keskkonnateadustes ja GIS-is kasutatavate mudelite taksonoomia (Skidmore 2002)

		Mudelid toimimise loogika järgi		
		Deduktiivsed	Induktiivsed	
Mudelid arvutusmeetodi järgi	Deterministlikud	Empiirilised (empirical)	modifitseeritud induktiivsed mudelid (<i>modified inductive</i>) klassifitseerimine juhendatud klassifitseerijatega (mudeli invessioon) (<i>classification by supervised classifiers (model inversion)</i>)	statistilised mudelid juhendatud klassifitseerijate treenimine (näiteks (edaspidi N:) suurima tõepära algoritm) geostatistilised mudelid geneetilised algoritmid (<i>genetic algorithms</i>)
		Teadmistest tulenevad (knowledge)	ekspertsüsteemid	hägusad süsteemid (<i>fuzzy systems</i>)
		Protsessist tulenevad (process)	hüdrooloogilised mudelid ökoloogilised mudelid	mudeli koefitsientide induktiivne muutmine kohalikele oludele vastavaks
	Stohhastilised	Monte Carlo meetod	kunstlikud närvivõrgud (<i>Neural network classification</i>) Monte Carlo meetod	

Deterministlikud mudelid annavad teatud sisendväärtuste korral alati kindla väljundväärtuse. Stohhastiliste mudelite puhul võib samade sisendandmete korral väljundväärtus varieeruda iga kord kui mudel uuesti käivitatakse. See tuleneb neid mudelites sisalduvast juhuslikust komponendist.

Deterministlikest ja võib öelda ka, et kõigist keskkonnamodelleerimise mudelitest kõige levinumad on empiirilised mudelid, mida teatakse ka statistiliste või numbriliste

mudelite nime all. Need mudelid saadakse andmetest induktiivselt, tavaliselt statistiliste modelleerimise vahendite abil (N: regressioon). Kuna andmeid kogutakse enamasti lokaalselt, on need mudelid reeglina kohaspetsiifilised ja neid ei saa üldjuhul ekstrapoleerida ega rakendada väljaspool uuringuala.

Liikide leviku modelleerimise põhisuunad on kokku võtnud Guisan & Zimmermann 2000. Kõige tavalisemad meetodid on mitmene regressioon ja üldine lineaarne mudel (*GML* — *general linear model*). Käesolevas töös kasutatud metoodika võib liigitada stohhastiliste induktiivsete mudelite hulka.

Nagu eelpool mainitud, on tänapäeval üheks peamiseks keskkonnainformatsiooni allikaks kaugseire. Kaugseire kui teadussuuna kiirenev areng sai alguse 1950-ndatel ja 1960-ndatel aastatel. Termin kaugseire võeti esmalt kasutusele Michigani ülikoolis 1960-ndatel ja 1970-ndatel toimunud sümposiumitel. Formaalselt defineeriti kaugseire Ameerika Fotogrammeetria ja Kaugseire Ühingu poolt 1980-ndate alguses (*American Society for Photogrammetry and Remote Sensing – ASPRS*) ja see kõlab järgnevalt: salvestava seadme abil objekti või nähtuse kohta informatsiooni kogumine, olemata sellega füüsilises kontaktis (Colwell 1983, viidatud Jensen 2000 järgi). Sellise definitsiooni järgi ei ole määratletud kui kaugel on kaugel, mistõttu ka näiteks röntgenpiltide tegemist saab käsitleda kaugseirena.

Metsade kaugseire oli üks esimesi uurimisobjekte, millele hakati tähelepanu pöörama kaasaegsete taimkatte kaugseire meetodite väljaarendamisel. Eelkõige kosmiline kaugseire, kuid ka teised kaugseire liigid, on metsandusega seotud ülesannete lahendamisel leidnud kasutamist peamiselt järgnevate tüüpprobleemide korral (Nilson 1994):

- metsavarude kaardistamine ja hindamine, seda eelkõige seni väheuuritud või raskesti ligipääsetavates piirkondades;
- kontroll raiete üle — harvendus- ja lageraied;
- metsakahjustuste avastamine, kahjustuste suuruse hindamine ning inimõju hindamine;
- metsaseire muutmine pidevaks ja andmete integreerimine vastavatesse geoinfosüsteemidesse.

Pika ajalooga ja ka tänapäeval kasutamist leidvad on visuaalsed fotointerpreteerimise meetodid (Congalton *et al.* 2002), (Lonard *et al.* 2000). Kasutades metsavaru hindamisel puistukaupa visuaalset fotointerpreteerimist on suhteliseks

täpsuseks saadud 14–45 % ja kasutades satelliitpiltide interpreteerimist 26–70%. Tulemused sõltuvad tugevalt keskmisest puistu suurusest ja sellest, millist lisainformatsiooni kasutatakse (Kilpeläinen & Tokola 1998). Puistu all mõistetakse struktuuri, vanuse, tekkeviisi ja muude tunnuste poolest ühetaolist puude kogumit metsas (Masing 1992).

1990. aastate keskpaigas tõdeti, et metsa tunnuste leidmiseks satelliitkujutiste põhjal on võrdlemisi piiratud arv meetodeid. Neid meetodeid saab jagada empiirilisteks (N: spektraalsetel vegetatsiooni indeksitel põhinevad) või füüsikalisteks (N: metsa heleduse mudel) (Gemmel 1999).

Empiiriliste kaugseireandmete klassifitseerimine on jagatav kahte faasi: süsteemi õpetamine/treenimine kasutades induktiivset lähenemist ja klassifitseerimine kasutades deduktiivset lähenemist. Tänapäeval on empiirilistest meetoditest väga palju kasutamist leidnud *k*-NN (*k nearest neighbours*) ehk *k* lähima naabri meetod. Soomes põhineb koguni riiklik mitmeallikaga metsainverteerimise programm *k*-NN meetodil (Katila & Tomppo 2001). Meetodi arendamisega on tegeletud juba pikemat aega ja uuemad rakendused kasutavad lisaandmetena suurte alade metsatunnuste (tagavara liikide kaupa) varieeruvuse hinnanguid (lahutusega 1×1 km) ja optimaalsete kaalude leidmiseks geneetilist algoritmi (*genetic algorithm*). Tulemused näitavad, et sellise meetodiga saadakse oluliselt paremaid tulemusi, kui teiste senilevinud *k*-NN meetodi variantidega (Tomppo & Halme 2004).

Füüsikaliste meetodite aluseks on metsa heleduse mudel. Reeglina lihtsustatakse pikslite heledusarvutusi ja leitakse kõikide komponentide (päikesevalguses ja varjus olevad võrad ning tagapõhja komponendid) heledus signatuuride summana, mis on kaalutud vastavalt sensori poolt registreeritud komponentide suhteliste hulkadega (Nilson 1994). Puistu biofüüsikaliste tunnuste hindamiseks kasutatakse füüsikalise meetodi puhul metsa heledusmudeli inversiooni. Inversioonitehnikate eeliseks on see, et nad on rakendatavad kõigi kohtade ja tingimuste puhul. Inversioon sisaldab mudeli parameetrite häälestamist seni, kuni mudeli peegeldused kajastavad mõõdetud peegeldusi kõige paremini (Gemmel 1999).

Eestis on füüsikaline kaugseire suund tugevalt esindatud. Andres Kuuse ja Tiit Nilsoni poolt loodud mudel töötab spektraalses vahemikus 400–2500 nm ja kõikide päikese ning vaatenurkade korral (Kuusk & Nilson 2000). Kasutades eelpoolnimetatud puistu heleduse mudelit sai Mait Lang kõige paremaid seoseid metsa takseerparameetritega Landsat TM 3 ja 5 kanaliga (Lang 2001). Mudeli laialdasemat

kasutamist on piiranud see, et suure hulga mudeli sisendparameetrite määratlemine on praktikas väga raske ülesanne.

Peamised faktorid, mis tõenäoliselt vähendavad kaugseirel põhinevate meetodite täpsust, on järgmised (Tomppo *et al.* 2002, osaliselt muudetud):

- õpetusandmete geomeetiline täpsus. Leidub näiteid, kus proovitükkide asukoha täpsustamise protsess vähendas pikslipõhise tagavarahinnangu RMSE (*root mean square error* — ruutjuur ruutkeskmisest hälbest (ISI : 05.04.2005)) väärtust 36 % võrra (Halme & Tomppo 2001).
- Õpetusandmete eraldisesisene varieeruvus. On välja töötatud meetodeid eraldistesiseseid takseerandmete poolest homogeensete tükkide segmenteerimiseks (Pekkarinen 2002).
- Sensorist pärineva informatsiooni geomeetiline täpsus.
- Kaugseire ja maapinnal läbiviidud mõõtmiste ajaline kokkulangevus.
- Andmete radiomeetiline kvaliteet, oluliseks mõjutajaks atmosfäär.
- Topograafilised efektid, N: varjud.
- Signaali võimalik küllustumine biomassi keskmiste väärtuste juures.
- Taustaefektid alustaimestikust ja mullast.

Aktsepteeritavaks takseerandmete väljas kogumise veaks loetakse viga alla 20 %. 20 % veaga hinnanguid suudetakse kaugseire meetoditega anda, kui metsaeraldiste suurus on 30 ha. 1-3 ha suuruste eraldiste korral suudetakse anda vaid ligikaudseid hinnanguid (Mäkelä & Pekkarinen 2004).

Enimlevinud parameeter, mida püütakse kaugseire meetoditega modelleerida on puistu tagavara või ka üldisemalt biomass. Viimastel kümnenditel on teaduskirjanduses välja pakutud mitmeid spektroskoopilisi meetodeid taimekoosluste biomassi ja produktsiooni hindamiseks. Eriti intensiivselt on uuritud päikesevalguse poolt põhjustatud klorofüllü fluorestsentsi taimedes. Moya jt tutvustavad passiivse kaugseire instrumenti, mis mõõdab klorofüllü fluorestsentsi hapniku absorptsiooni vahemikus (760 nm), senini küll vaid kuni 50 m kauguselt (Moya *et al.* 2004).

Tavapäraste ühesuunaliste kaugseiremeetoditega jääb teadmata, kui palju lehepinda on varjus. Mitme vaatenurgalise (*multi-angular*) kaugseire vajalikkust demonstreerisid näitega kogu primaarproduktsiooni hindamisest Chen jt (Chen *et al.* 2003).

Riiklike metsatakseerandmeid kasutavates kaugseireuringutes on tihti jõutud järeldusele, et nende andmete kasutamine väikeste puistute metsamajanduslikul planeerimisel on problemaatiline (Kilpeläinen & Tokola 1998; Kull 1999). Mäkelä & Pekkarinen püüdsid kasutades k -NN meetodit, Landsat TM satelliitpilte ja puistupõhiseid metsainverteerimisandmeid hinnata tagavara ka liikide kaupa. Testiti mitmeid spektraalseid tunnuseid, kuid pikslite keskmine väärtus eraldise keskel andis kõige täpsemaid hinnanguid. Kogu tagavara hinnangu täpsuseks saadi 48 % RMSE. Kuuse puistu prognoosimise täpsushinnang 81 %, männi ja laialehiste liikide jaoks üle 100 %. Piksli tasemel on esitatud vead olnud 60-80 % kogutagavara puhul ja veelgi suuremad liikide puhul (Mäkelä & Pekkarinen 2004). Kuuse puistu tagavara prognoosiveaks 80 % RMSE sai ka veidi varasem uuring (Katila & Tomppo 2001).

Soomes on Hyyppä jt võrrelnud puistu atribuutide hindamiseks erinevaid kaugseire instrumente, kasutades modelleerimiseks kunstlike närvivõrkude ja mitmese regressiooni meetodeid. Õpetusandmetena kasutati metsainverteerimise andmeid (483 eraldist) ja ka põhjalikku välitööd valitud kohtades (140 eraldises). Kontrollmõõtmised näitasid, et tavalised metsainverteerimise andmed olid piisavalt täpsed uurimuse läbiviimiseks. Tagavara hindamisel saadi erinevate lähteandmete korral järgmised puistupõhised (*standwise*) treeningtäpsused R^2 järgi (kontrolltäpsused olid 3-15 % kehvemad) (Hyyppä *et al.* 2000):

- 0,68 — profüleeriv radar (*profiling radar*) (HUTSCAT) loodud looduslike objektide tagasihajuvuse (*backsckattering*) omaduste mõõtmiseks, ruumiline lahutusvõime 8m;
- 0,48 — ortofotod (lähis-infrapuna kanaliga), 0,85m;
- 0,44 — Spot XS, 20m;
- 0,35 — Spot PAN, 10m;
- 0,31 — Landsat TM, 30m.

Kaugseire hinnangute täpsust on püütud parandada spektraalsete ja tekstuuri statistikute kasutamisega (Tuominen ja Pekkarinen 2005). Mainitud uurimuses koguti õpetusaandmeid spetsiaalsete välitöödega. Spektraalsetel ja tekstuuri statistikutel põhinevaid tunnuseid arvutati ortofotolt välja originaal kanaliväärtustest, NDVI-st ja kanalite suhetest (lähis-infrapunane (edaspidi LIP)/punane, LIP/roheline ja punane/roheline). Parimate tunnuste väljavalimiseks kasutati korrelatsioonimaatriksit. Kolme (roheline, punane ja LIP) originaal kanali väärtuseid kasutava süsteemi korral

saadi tagavara täpsushinnanguks 78,2 % RMSE. Tunnuste lisamisel täpsus reeglina paranes ja 30 tunnust kasutava prognoosisüsteemi puhul oli see 57,8 %. Tagavarahinnanguga andsid kõige tugevama korrelatsiooni järgmised statistikud:

- originaalkanalite puhul AVG (keskmine) $-0,67$, S-TXR (blokkideks jagatud akna standardhälve) $0,67$ ja COR (Haralicki korrelatsioon) $-0,64$;
- NDVI puhul HOM (Haralicki kohalik homogeensus) $-0,72$, CON (Haralicki kontrast) $0,72$, ENT (Haralicki entroopia) $0,69$;
- kanali suhete puhul CON $0,72$, HOM $-0,71$, ENT $0,69$.

Rakendades puistupõhiseid statistikuid (RGB (R— punane, G — roheline, B — sinine) keskmine, punase värvi standardhälve) on saadud ka üsna paljulubava täpsusega tulemusi. Kasutades LIP kanaliga ortofotosid, k-lähima naabri ja variogrammide meetodit ning täpseid välitööde andmeid, saadi puistu põhiseks tagavara täpsushinnanguks 18 % – 27 % RMSE (Muinonen *et al.* 2001).

Metsade kaugseires on väga levinud ka uurimused, mis püüavad kahjustusi avastada ja mõõta. Suurtel kuivanud puutüvedel on metsa ökosüsteemis eriline ökoloogiline tähtsus bioloogilise mitmekesisuse säilitajatena. Stereoskoopilise aerofotode interpreteerimise meetodi puutüvede ülesleidmiseks on välja pakkunud Büttler & Schlaeper 2004. 67 % tüvesid, mille diameeter oli ≥ 25 cm, suudeti avastada.

Kasutades infra-puna kanaliga värvilisi ortofotosid on hinnatud ja kaardistatud viinamarjaistanduste füüsilisi parameetreid (N: lehtede tihedus). Algoritm oli võimeline eristama ja kaardistama viinamarjaridu ning vastavalt etteantud taimedevahelisele kaugusele leidma üksikute viinamarjavõsude kohalikke koordinaate (N: 2 rida ja 40 taim). Teades viinamarjavõsude kohalikke koordinaate ja omades infot nende füüsilise seisukorra kohta, saab selliseid andmeid istanduse edukaks majandamiseks ära kasutada (Hall *et al.* 2003).

On uuritud ka defoliatsiooni ehk lehtede väljalangemist. Uurimuses kasutati Landsat TM, mustvalgeid ortofotosid erinevatest aegadest ja Soome metsa invertteerimise (*Finnish national forest inventory*) vaatlusplatside (*field sample plots*) andmeid ja kaugusega kaalutud k -NN meetodit. Puistu tasemel oli kolme muutusklassi eristamise täpsus $K = 0,12$. Kõige paremad kanalid olid TM 4 ja 5 (Heikkilä *et al.* 2002).

Tormimurru avastamiseks on välja pakutud ortofotosid ja fotogrammeetrilist meetodit kasutav lahendus, millega suudeti puistu kõrgust hinnata, täpsusega $\pm 0,7$ m.

Leiti, et sellist meetodikat võiks rakendada abiootiliste kahjustuste hindamiseks ja sisendina tuulekahjustuste mudelitele (Miller *et al.* 2000).

Siinses uurimuses ei kasutata treeningandmeid mudeli või reeglite loomiseks, vaid näidistebaasi loomiseks. Prognoosi saamiseks võrreldakse kaardistatava üksuse (N: piksel) kirjeldavaid tunnuseid (pildi- ja kaardiandmed) näidistega ja sarnasuse alusel leitakse sellele väljundväärtus. Kasutatava meetodika oluline eripära on see, et tunnustele ja näidistele kaalude leidmiseks kasutatakse tehisõpet.

2. Metoodika

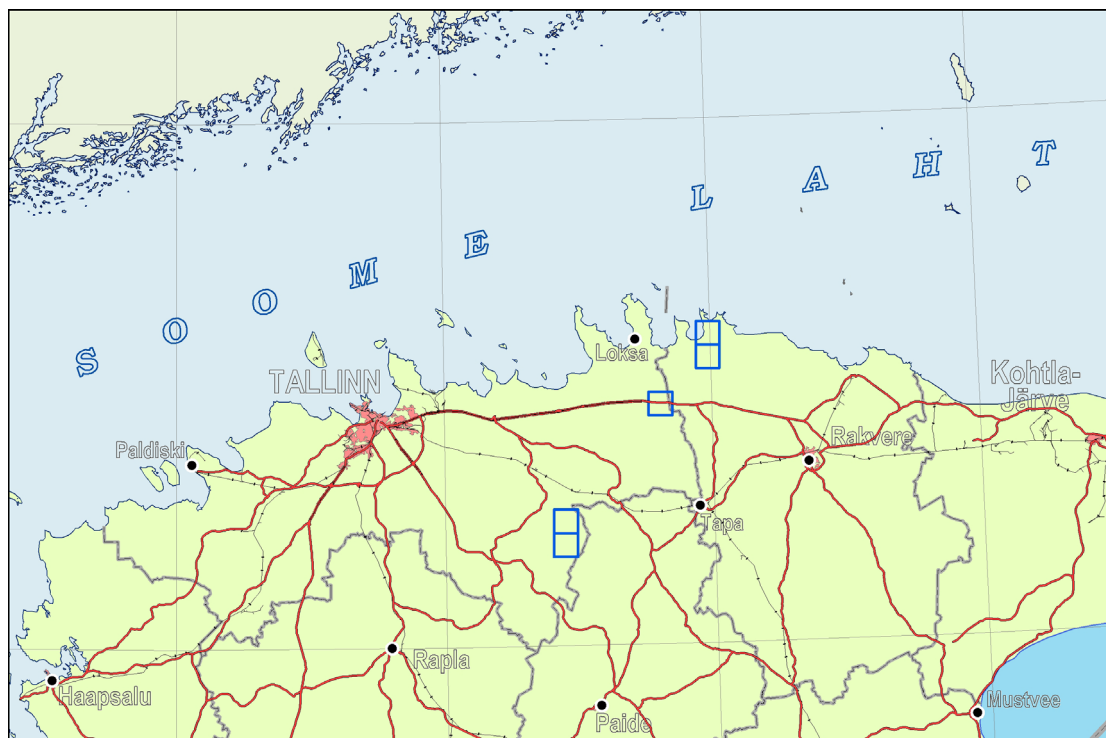
2.1. Uurimisala kirjeldus

Käesoleva uurimuse teema on tihedalt seotud metsa majanduslikult oluliste parameetritega. Metsa ja metsamaa tähendus on aja jooksul muutunud ja vaidlused nende terminite tähenduse üle kestavad praegugi. Eesti metsade olukorda kajastavates väljaannetes on metsamaana käsitletud maad, mis on suuteline tulundusmetsade normaalse raieringi keskmisena tootma vähemalt ühe tihumeetri tüvepuitu aastas hektari kohta ning seda maad ei kasutata olulisel määral metsandusega mitteseotud tulunduseesmärkidel. Metsana aga käsitletakse metsamaad, millel kasvavate puude keskmine kõrgus on vähemalt 1,3 m, puistu täius on ümardatult vähemalt 0,3 ja puistu tootlikus vastab vähemalt 5a boniteediklassile. Puistutes kõrgusega kuni 6 m rakendatakse täiuse asemel kõrgustäiust. 1999. aasta Metsaseadus sätestab metsa miinimumpindalaks lahustükis 0,5 ha (Eltermann, Kohava 2001).

Uurimistöös kasutatud ortofotod (põhikaardilehed numbritega 63681, 63683, 64901, 74011 ja 74013) paiknevad Vahe-Eesti metsase vööndi põhjaosas. Kaks esimest neist lõikavad Kõrvemaa maastikukaitseala läänepiiri, kolmas Lahemaa rahvuspargi lõunapiiri ja kaks viimast jäävad tervenisti Lahemaa rahvuspargi territooriumile (Joonis 1). Ka uurimistöös kasutatud satelliitpilt hõlmab Põhja-Eesti rannikupiirkonda (Joonis 3). Järgnevalt on CORINE 2000 alusel välja toodud uurimisalal enamlevinud maakattetüüpide osakaalud: okasmets 22 %, segamets 18 %, haritav maa 16 %, metsa-põõsastiku üleminekuala 11 %, karjamaa 8 % ja heitlehised lehtmetsad 7 %.

2.2. Andmete päritolu ja ettevalmistus

Uurimistöös on kasutatud värvilisi ortofotosid, Landsat 7 ETM+ satelliitpilte, 1:10000 mullakaarti ja põhikaarti. Riigimetsa takseerandmed tehisõppesüsteemi treenimiseks ja tulemuste kontrollimiseks on saadud Riigimetsa Majandamise Keskuselt (RMK).



□ ortofotode paiknemine

Joonis 1. Ortofotode asukohaskeem

2.2.1. Kaardiandmed

Kaardiandmeid on käesolevas töös kasutatud eelkõige lisaandmetena. Prognoosid põhinevad reeglina pildiaandmetel. Maapinna omaduste iseloomustamiseks kasutati argumenttunnustena 1:10000 mullakaarti, eesmärgiga saada ülevaadet, kui suurt mõju prognoositäpsusele see võib omada. Digitaalne kogu Eestit kattev mullakaart koos andmebaasiga koostati Maa-ameti tellimisel vahemikus juuli 1997 kuni märts 2001. Kaart on vektorformaadis ja esitatud Eesti Põhikaardi koordinaatsüsteemis (Lambert-EST). Kaardi koostamisel olid aluseks suurmajandite ja metskondade mullastiku kaardid. Varem 1: 5000 mõõtkavas koostatud kaardid vähendati ja generaliseeriti 1:10000 mõõtkavale vastavalt. Kaardi koostamisel kehtisid samad sisulised nõuded ja ka muldade kaardistamisüksuste nimestik, mis käsitsi tehtud töö puhul.

Eelpoolmainitud mullakaardil ja sellega kaasas käivas andmebaasis esineb väga palju formaalselt erinevaid mulla- ja lõimisekoode, mistõttu ei ole need modelleerimiseks kasutajasõbralikud. CBR-süsteemi treenimisel kasutati üldistatumat mullaklassifikatsiooni (Remm 2002).

Metsamaskina kasutati 1:10000 põhikaardi metsaklasse ja RMK takseerandmete eraldisi. Metsamaski kasutamine on vajalik lokaalstatistikute arvutamisel selleks, et

arvesse võtta vaid eraldise piires olevaid pikseleid. Sellel teemal on pikemalt kirjutatud peatükkides 2.4 ja 2.5

2.2.2. Pildiandmed

Uurimistöös kasutati värvilisi ortofotosid lahutusvõimega 1 m. Kahjuks ei ole Eestis infrapuna spektriala vajalikkust aeropildistamisel senini teadvustatud. Seetõttu on kättesaadaval ortofotodel puudu taimkatte kaardistamise seisukohalt väga oluline infrapuna spektriala kanal.

Algselt oli ortofotode lahutusvõime 0,4 m, kuid arvesse võttes arvutite võimsust ja uurimisobjekti ruumilist jaotust, otsustasin pikslite küljepikkust vähendada. Viisin läbi uuringu testimaks, kas lahutusvõime vähendamine kahandab oluliselt CBR-süsteemi prognoosivõimet. Kasutades veamaatriksite statistilise erinevuse testimise meetodikat (Congalton, Green 1999), leidsin, et keskeltläbi jääb 0,4 ja 1 m lahutusvõimega ortofotosid kasutavate CBR-süsteemide prognoosivõime samaks. Leidsin näiteid, kus pikslisuuruse suurendamine halvendas ja ka parandas treeningtäpsust. Seetõttu otsustasin jääda kasutama 1m lahutusvõimega ortofotosid.

Ortofotode lahutusvõime vähendamine viidi läbi kasutades Adobe Photoshopi tarkvara. Ortofotode ülesvõtteeag varieerub juunist kuni juulini 2002. aastal. Ortofotode kvaliteet on ebahühtlane. Joonis 2 kujutab ühte ortofotot läbivat erinevate kaadrite liitekohta. Samuti hakkavad väga paljudes kohtades silma valged udemed ja karvad.

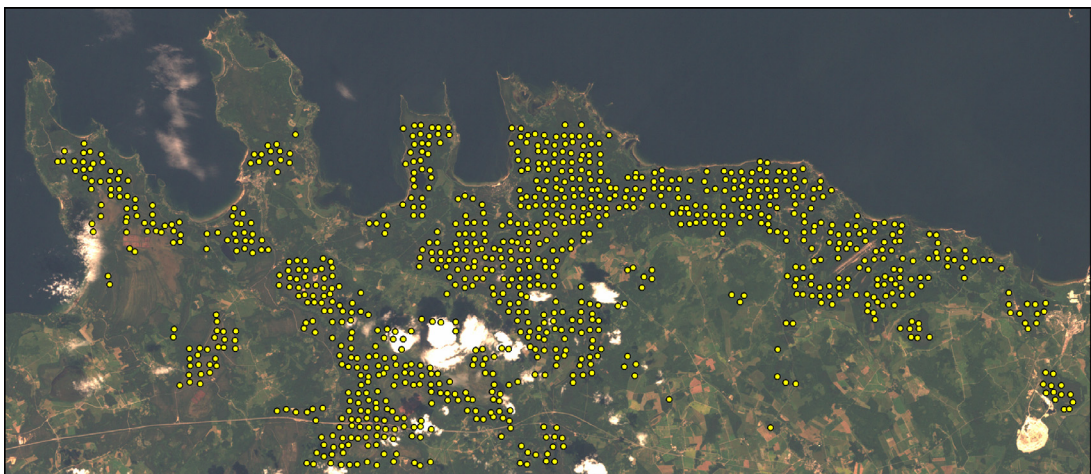
Teise pildiandmete liigina kasutati Landsat 7 Enhanced Thematic Mapperi satelliitpilte. Landsati programm on käigus aastast 1972 ja see on USA vanim maapinna vaatlusteks kasutatav satelliitsüsteem. Selle programmi puhul on tegemist kaugseire alaliigiga, mida nimetatakse multispektraalseks kaugseireks. Multispektraalset kaugseiret on defineeritud kui käsitletavalt objektilt või alalt peegeldunud, emiteeritud või hajunud energia registreerimist mitmes elektromagnetlainete lainelas.

Landsat 7 on maapinnast 705 km kõrgusel orbiidil 1999. aastast alates. Andmeid kogutakse korraga 185 km laiuselt alalt, enamikes lainelades resolutsiooniga 25 m × 25 m. Sama ala kohta maapinnal saab andmeid koguda 16-päevaste vahemike järel. Digitaalsed sensorid säilitavad andmeid kaheksast bitist koosnevate baitidena. Seetõttu on võimalik registreeritavate väärtuste vahemik 0-255 (Jensen 2000).



Joonis 2. Silmatorkav erinevate kaadrite liitejoon

Tabel 2 annab ülevaate Landsati lainealadest ja nende peamistest kasutusotstarvetest. Käesolevas töös kasutati Landsat 7 ETM+ satelliitpildi kaadri numbriga 187018, mille salvestuskuupäev on 06.07.2001 (Joonis 3). See satelliitpilt on saadud üleeuroopalise projekti “*Image 2000 — the Spatial Reference for Europe*” vahendusel. Satelliitpildid on ortorektifitseeritud ja projektsioonis L-EST 97. Satelliitpildid on läbinud geomeetrilise kvaliteedikontrolli, mille tulemusena on asukohatäpsuseks saavutatud $RMSE < 25$ m. Pilte ei ole atmosfääri mõju vähendamiseks kalibreeritud, kuid üks piltide valiku eeltingimus oli, et pilvedega kaetud pind peab olema alla 5 % (IMAGE 2000 : 17.04.2005).



Joonis 3. Kasutatud juhupunktide paiknemine Landsat ETM+ komposiitpildil

Tabel 2. Landsat TM lainelad ja nende kasutatavus maakatte kaardistamisel (Meiner 1999), osa andmeid lisatud

Lainela	Spektrivahemik	Mõned kasutusvaldkonnad
TM 1 (sinine)	0,45 – 0,52 μm	Maakasutuse, muld- ja taimkatte kaardistamine
TM 2 (roheline)	0,52 – 0,60 μm	Taimekoosluste elujõulisuse esiletõoja
TM 3 (punane)	0,63 – 0,69 μm	Taimekoosluste eristamine, muldade piiritlemine
TM 4 (lähis-infrapunane)	0,76 – 0,90 μm	Taimekoosluste biomassi näitaja, põlluviljade äratundmine, teraviljapõldude/mulla ning veekogude piiritlemine
TM 5 (kesk-infrapunane)	1,55 – 1,75 μm	Taimkatte niiskusesisalduse näitaja (kulu)
TM 6.1 ja 6.2 (soojus-infrapunane)	10,4 – 12,5 μm	Taimkatte stressinähtused, mulla niiskus. Erandina on selle resolutsioon 60 x 60 meetrit. Kanal 6.1 ja 6.2 on vastavalt madala ja kõrge radiomeetrilise lahutusvõimega.
TM 7 (kesk-infrapunane)	2,08 – 2,35 μm	Mapinnal esinevate mineraalide eristamine
TM 8 (pankromaatiline lainela) – kõigi värvuste suhtes tundlik	0,52 – 0,90 μm	Landsat 7 programmis on see lainela uuena juures. Resolutsioon: 15 x 15 m

2.2.3. Metsakorralduse takseerandmed

Takseerandmete kogumine on hinnanguline tegevus, mille käigus ekspert fikseerib välitööde käigus vajalikud andmed. Puistu hektaritagavara määratakse puistuelementide hektaritagavarade summana. Puistuelemendi hektaritagavara arvutatakse puistuelemendi kõrguse ning täiuse või rinnaspindala või puude arvu alusel. Looduses määratavad takseertunnused peavad olema määratud sellise täpsusega, mis võimaldaks kasvava metsa tagavara määrata nii, et 2/3 juhtudel oleks tagavara määramise viga alla 15 % ja ülejäänud juhtudel alla 20 % (Elektrooniline Riigi Teataja : 10.05.2005). Puistu tagavara juurdekasvu saab määrata puistu korduval mõõtmisel, nummerdatud või nummerdamata puudega proovitükil ning valemite abil. Puistu lõikepindala määramiseks kasutatakse relaskoopi. Puu vanust saab määrata mitmeti (Etverk 1980):

- kultuurpuistute korral saab vanuse teada dokumentatsioonist;
- aastarõngaste järgi värsketelt kändudelt või juurdekasvu puuriga võetud proovist;

- kuni keskeani männi aastaste kõrguskasvude loendamise teel (kasutatav ka kuusel);
- ekspertarvamusena, toetudes puu välistunnustele ja kasvukohale.

Eraldise kirjeldamine omaette üksusena on põhjendatud, kui eraldis vajab erinevat metsa majandamise viisi või selle takseertunnused erinevad naabereraldiste takseertunnustest vähemalt ühe takseertunnuse osas alljärgnevas ulatuses (Elektroniline Riigi Teataja : 10.05.2005):

- mõne puuliigi osatähtsus koosseisus erineb üle 20 %;
- enamuspuuliigi keskmine vanus erineb üle 10 aasta;
- puistuelemendi rinnaspindala, puude arv hektari kohta või täius erineb üle 20 %;
- puistuelemendi kõrgus erineb üle 10 %;
- puistuelemendi hektaritagavara erineb üle 20 %;
- enamuspuuliigi päritolu on erinev;
- eraldis kuulub metsakasvukohatüübilt erinevasse tüübirühma.

Takseerandmete kogumise hind Skandinaavia metsanduses on reeglina olnud umbes 10-15\$ hektari kohta. Andmete suhteline täpsus varieerub 11–33 %. Soomes on konventsionaalse puistupõhise metsainventeerimise veaks tagavara puhul 23 m³/ha (Hyypä *et al.* 2000). Vead tulenevad liiga vähestest mõõtmistest, subjektiivselt valitud kohtadest ja puistu struktuuri heterogeensusest (Kilpeläinen & Tokola 1998). Takseerandmete tootja keskmine töömaht päevas on 30 ha (Mäkelä & Pekkarinen 2004).

Empiiriliste ennustusmodelite usaldusväärsuse seisukohalt on väga oluline see, kui esinduslikud on õpetus- ja võrdlusandmed. Õpetusandmeid ehk võimalikke argumenttunnuseid koos funktsioontunnustega on vaja selleks, et saaks näidistel põhineva järeldamise meetodikaga õppimist läbi viia. Ajalised nihked välitööde, üldistatumalt õpetusandmete, ja erinevate argumenttunnusteks olevate andmete päritoluaaja vahel tekitavad müra iga ennustava süsteemi puhul. Käesolevas töös kasutatakse Loobu ja Paunküla metskondade takseerandmed, mis on kogutud 2001. ja 2002. aasta suvekuudel. Seega on ortofotode ja takseerandmete fikseerimise vahe üks kuni kaks (kaks ortofotot Paunküla metskonna territooriumil) aastat. Keskmine aastane juurdekasv Loobu metskonnas on 3,4 tihumeetrit hektari kohta aastas. Seega ei ole

aastane juurdekasv ja looduslik struktuurimuutus puistus veel nii märkimisväärne, et käesoleva uurimistöö tulemusi oluliselt mõjutada. Järsud muutused puistus esinevad (lage)raiate puhul.

Takseerandmetena fikseeritakse ~25 tunnust, millest käesoleva uurimistöö jaoks on huvipakkuvad enamuspuielik, arenguklass, esimese rinde tagavara ($tm\ ha^{-1}$) ja kõikide rinnete jooksev juurdekasv ($tm\ ha^{-1}\ a^{-1}$). Enamuspuieliigiks puistute korral on põhirinde puielik, mille tagavara on põhirindes suurim. Juhul kui kahe või enama puieliigi tagavara on võrdne, loetakse enamuspuieliigiks okaspuielik, kui okaspuielikide tagavarade summa on suurem lehtpuielikide tagavarade summast. Kui sel juhul on kahe okaspuieliigi tagavara võrdne, on enamuspuieliigiks majanduslikult väärtuslikum puielik. Sama kehtib ka lehtpuielikide kohta, kui lehtpuielikide tagavara on suurem okaspuielikide tagavarast. Enamuspuielik fikseeritakse metsa, põõsastike ja puudega soode puhul. Puistude arenguklassides¹ on üheks klassiks kokku võetud lage ala ja selgusetu ala, kuna need on sisuliselt väga kattuvad (OÜ Eesti Metsakorralduskeskus 2002).

Õpetusandmete saamiseks omistati eraldistena esitatud takseerandmed juhupunktidele. Juhupunktide genereerimisel lähtuti soovist kasutada juhupunkte ümbritsevat ala vähemalt 5 piksli raadiuses. Selleks, et lokaalstatistikuid oleks võimalik arvutada piisavalt suure raadiusega niimoodi, et kahe juhupunkti ümber arvatud raadiused omavahel ei kattuks, genereeriti juhupunktid ortofotode puhul vähemalt 100m ja satelliitpiltide puhul vähemalt 250m vahega. Selliste tingimuste saamiseks genereeriti punktid algselt korrapärase võrgustikuna, ortofotode puhul 200m ja satelliitpiltide puhul 500m vahemaaga. Seejärel lubati iga punkti x ja y koordinaadil juhuslikult varieeruda vastavalt ortofotodel $\pm 50m$ ja satelliitpiltidel $\pm 125m$. Tulemuseks saadud juhupunktide paiknemine võimaldab leida kattumatute raadiustega lokaalstatistikuid ortofotode puhul 50m ja satelliitpiltide puhul 125m raadiuses. Juhupunktidest arvatavate

¹ **Lage ja selgusetu ala** — ala, kus puistu täius on alla 0,3 ja kus kultiveeritud või loodusliku uuenemisega tekkinud peapuieliigiks sobivad taimed puuduvad või nende taimede arv ei ole metsa uuenemiseks piisav. Ka kultiveeritud või looduslikult uuenev ala, kus kasvab enam kui 1200 peapuieliigi taime hektari kohta, kuid mis ei vasta teistele uuenenud ala tingimustele. **Noorendik** — kultiveeritud või looduslikult uuenenud ala, kus kasvab enam kui 1200 ülepinnaliselt paiknevat vähemalt 0,8 m kõrgust peapuieliigi taime hektari kohta ja mille keskmine rinnasdiameeter on kuni 6 cm. **Latimets** — puistu täiusega üle 0,3, keskmise rinnasdiameetriga 6...12 cm ning keskmise vanusega alla ½ küpsusvanust. **4. Keskealine mets** — puistu täiusega üle 0,3 ja keskmise rinnasdiameetriga üle 12 cm, mis tõenäoliselt 10 aasta jooksul ei saavuta küpsuse kriteeriumitele vastavat keskmist vanust. Keskealiste metsade hulka kuuluvad ka puistud, mille vanus on üle ½ küpsusvanuse, kuid $d_{1.3} < 12\text{ cm}$. **Valmiv mets** — puistu täiusega üle 0,3, mis takseerimise hetkel ei vasta küpsuskriteeriumitele, kuid saavutab küpsusvanuse eeloleva 10 aasta jooksul. **Küps mets** — puistu täiusega üle 0,3, mille enamuspuieliigi vanus vastab küpsuse kriteeriumitele.

Arenguklasside määramisel kehtivad erandid hall-lepikutele nende kiire kasvu ja madala küpsusvanuse tõttu. Hall-lepikutes on noorendike ülempiiriks rinnasdiameeter 8 cm, latimetsi ei määrata ja 20 aastat on nii keskealise lepiku ülempiiriks kui valmiva lepiku alampiiriks

lokaalstatistikute raadiuste ülekatet tuleb vältida, sest sellisel juhul ei ole juhupunktid enam statistiliselt sõltumatud vaatlused.

Kokku viiele 5×5 km ortofotole, põhikaardilehtede numbritega 63681, 63683, 64901, 74011 ja 74013, genereeriti vastavalt eelpoolkirjeldatule 1846 juhupunkti. Takseerandmetena kasutati Loobu ja Paunküla metskondade kohta käivad andmeid. Juhupunktidele asukoha leidmisel ei seatud mingisuguseid reegleid eraldisesise paiknemise kohta. See tähendab, et juhupunkt võis sattuda näiteks vaid 2m kaugusele eraldise servast või ka täpselt eraldise keskele. Samuti ei seatud eraldiste kujule piiranguid. Eraldis võib olla väga kitsas, pikk ja väga mitmete sissesopistustega. Üks eraldis võis sisalda rohkem kui ühte juhupunkti. Juhupunktide sattumist eraldiste kitsastesse osadesse ja servaaladele ei püütud takistada seetõttu, et ka sellised üleminekulised alad, kus tihti võib esineda geomeetrilisi vigu ja muid segavaid faktoreid, näiteks varje, vajavad hea prognoosi saamiseks näidiseid. See tähendab, et ka metsaeraldiste servaalad tuleb ära kirjeldada. Eelpool mainitud seisukohta toetab ka näide kirjandusest, kus on leitud, et tänu puistusisesele looduslikule varieeruvusele võivad servapikslid erineda tuumpikslitest ja servapikslite väljajätmine võib anda tulemuseks nihkega/kõrvalekaldega hinnangu (Mäkelä & Pekkarinen 2004).

Satelliitpildi kaadrile 187018, ülesvõtmisajaga 06.07.2001 genereeriti vähemalt 250m vahega 969 juhupunkti. Takseerandmetena kasutati Loobu metskonna andmeid. Eelpoolkirjeldatud põhjustel ei pööratud ka nende punktide genereerimisel tähelepanu nende eraldisesisesele paiknemisele.

2.2.4. Tarkvara

Töös on kasutatud Tartu Ülikooli geograafia instituudis Kalle Remmi poolt loodud näidistel põhineva järeldamise süsteemi ja lokaalstatistikute arvutamise programmi. Kasutatud on CBR-süsteemi kahte erinevat versiooni: MLNN ja Pidevstuudium.

MLNN-i võib mõtteliselt jagada tehisõppe ja prognoosi osaks. Tehisõppe osas leiab programm argumenttunnustele ja näidistele kõige täpsemat prognoosi andvad tunnuste ja näidiste kaalud. Kasutades leitud kaale ja näidistebaasi, prognoosib prognoosi osa etteantud punktidele funktsioontunnuse väärtuse. Seda süsteemi võib kasutada erinevat tüüpi (nominaalsed ja binominaalsed, pidevad, osakaalud ja kompleksmuutujad, N: puistuvalem) funktsioontunnuste prognoosimiseks. Samuti

võivad argumenttunnused olla nii pidevad kui nominaalsed ja seda ka ühes näidistebaasis. Selle CBR-süsteemi kasutaja võib valida erinevate tehisõppe, veaarvutus- ja muutujate normalisatsioonimeetodite vahel.

Pidevstudium on MLNN-i ja lokaalstatistikute arvutamise programmi ühteliitev edasiarendus. See programm funktsioneerib andmebaasipõhiselt. Andmebaasis hoitakse andmeid lähteandmete asukoha kohta, neist arvutatud argumenttunnuseid, vaatlusvektoreid, tunnuste ja näidiste kaale ning reegleid tehisõppeks, prognoosimiseks ja lokaalstatistikute arvutamiseks. Pidevstudium on programm, kus tehisõppe protsess võib lõputult korduda. Võimalik on väga põhjalikult häälestada CBR-süsteemi loomist ja funktsioneerimist. Lokaalstatistikute programmiga on võimalik diskreetsetele tunnustele arvutada 12 statistikut ja pidevatele tunnustele 14 statistikut. Siinses uurimuses kasutati lokaalselt arvutatud keskmist, üle keskmise väärtuse osakaalu, standardhälvet, variatsioonikoefitsienti ja moodi ning struktuuri kirjeldavatest statistikute Morani I'd ja kauguse pöördväärtusega kaalutud Morani I'd.

Tehisõppe programmi funktsioneerimise, sealhulgas vaatlustevahelise sarnasuse hindamise, tunnustele ja näidistele kaalude leidmise ja prognoosimise kohta on täpsemalt võimalik lugeda minu bakalaureusetööst (Tamm 2003).

2.3. Tunnuste valimine tehisõppeks

Seletavate tunnuste hulk võib erinevate statistikute, raadiuse ja elektromagneetilise kiirguse vahemike kombinatsioonide tõttu olla väga suur (N: 6 andmekihti korda 5 raadiust korda 5 statistikut teeb argumenttunnuste arvuks 150). Seetõttu tuleks tehisõppele kuluva aja vähendamiseks kasutada eelvalikut. Selgitamaks välja head eelvaliku meetodit, võrreldi pidevate tunnuste puhul klasteranalüüsi ja samm-regressiooni ning nominaalsete tunnuste puhul klasteranalüüsi ja üldistatud lineaarseid mudeleid (*GLZ — generalized linear models*).

K-keskmise klasteranalüüsi (*k-means clustering*) algoritm võimaldab leida etteantud arvu omavahel võimalikult erinevaid klastreid. Programm alustab k suvalise klastriga ning objekte nende vahel liigutades püüab minimeerida klastritesisest ja maksimeerida klastritevahelist varieeruvust. Esialgsete klastrite väljavalimiseks kasutati meetodit, mis kõigepealt sorteerib kõik objektid (tunnused) vastavalt nende omavahelistele kaugustele ja seejärel valib esialgseteks klastritsentriteks objekte (tunnuseid) kindla intervalliga (STATISTICA 6).

Igast klastrist valiti välja tunnus, mis oli kõige lähemal vastava klatri keskmel. Sellist valikuprintsiipi järgides peaksid väljavalituks osutama sellised tunnused, mis kirjeldavad n-mõõtmelises tunnusruumis esinevast varieeruvusest kõige enam. Selleks, et vältida mõõtühikute (% , keskmine signaali tugevus (0-256)) mõju klastrite moodustumisele, standardiseeriti tunnused enne klasteranalüüsi kasutamist.

Regressioonimudeli saamisel kasutati tunnuste ükshaaval lisamise võimalust ehk samm-regressiooni. Nii võetakse mudelisse need tunnused, mis ei ole omavahel tugevalt korreleerunud ja mille lisamine mudeli täpsust kõige rohkem parandab. Eelvaliku tegemisel reastati tunnused determinatsioonikordaja R^2 alusel ja valiti välja 30 parimat tunnust. Üldistatud lineaarsete mudelite puhul kasutati tunnuse indikaatorväärtuse hindamiseks Wald-statistikut. Klasteranalüüsiks, regressioonimudeli ja üldistatud lineaarse mudeli leidmiseks kasutati programmi STATISTICA 6.

Sobiva eelvalikumeetodi leidmiseks, võrreldi diskreetsete tunnuste puhul veamaatrikseid, mis leiti iga tunnustekomplekti abil prognoositud tulemustest eraldi. Veamaatrikseid võrreldi vastavalt Congaltoni ja Greeni poolt kirjeldatud meetodikale (Congalton, Green 1999). Pidevate tunnuste puhul kasutati korrelatsioonikoefitsientide erinevuse testimise võimalus STATISTICA 6 programmis. Eelvaliku meetodite võrdlemisel kasutati viiel ortofotol paikneva 1846 juhupunkti andmeid.

2.4. Parimad tunnused ortofotodelt

Analüüsis kasutati 1846 takseerandmetega seotud juhupunkti. Ortofotodel on iga piksli kohta esitatud kolme kanali väärtused. Need väärtused on esitatud RGB-värvikoodina. On võimalik välja arvutada neile arvudele vastavaid väärtusi teistes värvisüsteemides, näiteks HSL-süsteemis. Selleks, et esitada iga piksli kohta andmeid, mis kombineeriksid RGB väärtusi kasutasin lokaalstatistikute programmis olemasolevat võimalust värvitooni (H — *hue*), küllastuse (S — *saturation*) ja heleduse (L — *lightness*) väärtuste väljaarvutamiseks.

Juba 1994 on leitud, et metsatakseerandmete ortofotode abil hindamisel on tekstuuri indeksitest rohkem kasu kui üksikute pikslite väärtustest. (Holopainen & Lukkarinen 1994, viidatud Muinonen *et al.* 2001 järgi). Lähima ümbruse arvestamise tähtsust on tõdenud ka Remm ja Luud 2003. Selline väide näib loogilisena, sest ortofotode üksikpikslite mõõtmel erinevad oluliselt eraldise mõõtmeltest ja nende väärtuste eraldisesisene varieeruvus võib olla suur. Teadmata on aga, kui suurt

lähiümbust tuleks parima prognoosi saamiseks arvestada. Seetõttu arvatati lokaalstatistikuid 10, 20 30, 40 ja 50 meetri raadiuses.

Lähiümbust iseloomustavatest näitajatest arvatati välja keskmine (lühendatult k), standardhälve ($stdv$), variatsioonikoefitsient (var), üle keskmise väärtuste osakaal ($ykvo$) ja külgnevate pikslite autokorrelatsiooni näitaja Morani I ($akor$).

Ortofotodelt lokaalstatistikute arvutamisel kasutati riigimetsa eraldiste maski. Eesmärgiks oli, et lokaalstatistikute arvutamine toimuks ainult selle eraldise piires, kuhu juhupunkt langes. Riigimetsa takseerandmetes on atribuutiivinfo esitatud eraldiste kaupa. Kuna ka metsamajandus käib valdavalt eraldistepõhiselt, võivad kaks kõrvuti asetsevat eraldist ka looduses üksteisest atribuutandmete poolest suurel määral erineda. Seetõttu peeti oluliseks arvestada järskude piiride olemasolu ka lokaalstatistikute arvutamisel.

Kokku arvatati eelpoolkirjeldatud kanali, raadiuse ja lokaalstatistiku kombinatsioonidena välja 150 tunnust, millest vastavalt eelvalikumeetodile jäeti tehiseppe läbiviimiseks alles 30. Enamuspuuliigi, arenguklassi, juurdekasvu ($tm\ ha^{-1}a^{-1}$) ja tagavara (tm/ha) prognoosimiseks viidi tehiseppe läbi programmiga MLNN.

Tehiseppe läbiviimiseks kasutati peamiselt programmis olevat 12. meetodit, mida saab vastavalt soovile mitmeid kordi järjest kasutada. Samas on see meetod põhiolemuselt väga sarnane meetodile 13, mille head omadused on tõestust leidnud (Remm 2004). Meetod 12 alustab tunnuste ükshaaval lisamisega kasutades selleks kõiki vaatlusi, seejärel jätkab tunnuste kaalumise kasutades vaatlusi suurenevates valimites ja lõpetab näidiste kaalumise. Tehiseppet viidi iga funktsioontunnuse jaoks läbi kuus korda, selleks, et vähendada juhusliku komponendi mõju.

Uurimuse käigus leiti, et tehiseppega leitud tunnuste kaalude keskmiste väärtuste kasutamine indikaatorväärtustena ei anna päris adekvaatset pilti tunnuse kasulikkusest. Näiteks võis üks tunnus tehiseppe käigus saada suure kaalu vaid ühe korra, mõni teine tunnus aga sai veidi madalamat kaalu korduvalt. Leides sellises olukorras tunnuste kaalude keskmised väärtused, selgub, et esimene tunnus on suurema indikaatorväärtusega kui teine. Tegelikuses peeti esimese tunnuse lisamist prognoosimisse vajalikuks vaid ühel korral, samal ajal oli teisest tunnusest mitmel korral kasu. Seetõttu kasutati indikaatorväärtuste väljaarvutamiseks tunnuse/komponendi kasutuskordadega kaalutud tunnuste kaalude keskmist. Iga üksiku tunnuse/komponendi indikaatorväärtusena käsitletakse sellisel juhul selle osakaalu kaalutud keskmise arvutamisel. Kuna ilmnes, et üldkeskmisest on suuremad reeglina

kümne argumenttunnuse indikaatorväärtused, esitatakse iga funktsioontunnuse kohta just 10 parimat tunnust.

2.5. Parimad tunnused satelliitpiltidelt ja mullakaardilt

Analüüsis kasutati 969 genereeritud juhupunktist 907, sest 62 juhul olid punktid langenud pilvedele või nende varjudele. Pilved ja nende varjud on võrreldes tavaliste maismaanähtustega väga eriliste spektraalsete signatuuridega, mistõttu ei ole mõistlik kasutada neid takseerandmete õpetusandmetena.

Satelliitpiltide üks piksel katab treeningandmetena kasutatud eraldiste keskmisest pindalast (3,4 ha) ~2 %. Kuna takseerandmed on esitatud eraldiste keskmistena, võib sellest järeldada, et üks piksli väärtus ei tohiks takseerandmete jaoks olla hea prognoosija. Võrdlusmomendi tekitamiseks võeti üksikpikslite väärtused siiski analüüsis arvesse. Lisaks arvutati lokaalstatistikuid 40, 80 ja 120 meetri raadiuses. Lokaalstatistikutest kasutati keskmist, standardhälvet, variatsioonikoeffitsienti, kauguse pöördväärtusega kaalutud Morani I ja mullakaardi puhul moodi.

Riigimetsa takseerandmete eraldiste väiksuse tõttu oli nende kasutamine metsamaskina satelliitpiltide puhul raskendatud. Juhupunkti ümbruse 120m raadiusesse jääb juba ~1,3 korda keskmise eraldise pindalast suurem pind. Seetõttu võeti lokaalstatistikute arvutamisel metsamaskina kasutusele põhikaardi põhialad: mets ja noormets (Eesti põhikaardi juhend 2002). Viidi läbi ka katsetus, kuidas mõjub prognoositäpsusele riigimetsa eraldiste maski kasutamine ja veenduti, et põhikaardi maski kasutamine on otstarbekam. Põhikaarti kasutati maskina ka prognoosimisel. See tähendab, et prognoosimisel otsitakse näidiseid vaid samale põhikaardi põhialale jäävate vaatluste hulgast.

Kokku arvutati välja 108 tunnust, millest klasteranalüüsi kasutades valiti tehisõppeks välja 30. Ilma eelvalikuta valiti tehisõppe jaoks välja mullakaardi tunnused ja satelliitpildi kanalite üksikpiksli väärtused. Nende tunnuste osalemist tehisõppes peeti vajalikuks eelkõige seetõttu, et saada paremat ülevaadet kui olulisteks prognoosijateks neid süsteem peab. Niisiis anti tehisõppele ette 40 tunnusest koosnev komplekt. Parimate tunnuste väljaselgitamine käis vastavalt eelpoolkirjeldatud indikaatorväärtuste leidmisele.

2.6. Tunnuste võrdlus

Tunnuse komponentideks nimetatakse käesolevas uurimistöös andmekihti/kanalit, raadiust ja lokaalstatistikut. Nende kolme komponendi kombinatsioonina moodustub üks tunnus. Näiteks 30m raadiuse keskmine lähi-infrapuna kanali väärtus. Andmekihtide/kanalite all on antud kontekstis mõistetud elektromagnetkiirguse vahemikke, mille väärtusi kasutatav sensor registreerib ja samuti RGB koodist arvutuslikult saadud HSL vahemikke ning kaardiandmeid. Komponentide indikaatorväärtused leiti vastavalt eelpoolkirjeldatule. Komponentide indikaatorväärtusi uuriti ortofotode ja satelliitpildi puhul eraldi. Seda eelkõige eesmärgiga kirjeldada võimalikult hästi ühe pildiandmete liigi omadusi. Indikaatorväärtusi leiti iga prognoositud funktsioontunnuse jaoks. Saamaks terviklikumat ja üldisemat ülevaadet ortofotode ja satelliitpiltide omadustest takseerandmete prognoosimisel, keskmistati iga üksiku funktsioontunnuse indikaatorväärtused ka üheks pildiandmestikku kirjeldavaks näitajaks.

2.7. Prognoosikaartide koostamine

Üheks uurimistöõ eesmärgiks oli metsaparameetrite prognoosikaartide koostamine. Prognoosikaartide koostamiseks kasutati satelliitpildidel ja mullakaardil põhinevat CBR-süsteemi. Kaardid (põhikaardileht nr 7402) koostati neljale puistu parameetrile: puistu tagavara, puistu tagavara juurdekasv, puistutüüp ja puistu vanuseklass. Kaks viimast tunnust tekitati takseerandmete ümberklassifitseerimisega, eesmärgiga muuta kaardistatavate tunnuste klassifikatsioon võimalikult lihtsaks.

Puuliigi osakaalu järgi eraldises loodi uus tunnus puistutüüp, mis koosneb kolmest klassist:

- okasmets (okaspuude osakaal >70 %);
- lehtmets (lehtpuude osakaal >70 %);
- segamets (kaks esimest tingimust pole täidetud).

Kuus puistute arenguklassi jaotati ümber uueks tunnuseks — puistu vanuseklass:

- küps ja valmiv (küps 10 aasta jooskul) mets;
- keskealine ja noor mets (arenguklassid: keskealine-, lati- ja noormets ning lage ja selgusetu ala).

2.8. Täpsushinnangud

Uurimistöös kasutatud metoodika treeningtäpsuse hindamine põhineb jäta-üks-välja-ristkontrollil (*leave-one-out cross-validation* (LOOC)). Iga treeningjuhtumi kohta arvutatakse prognoositav väärtus seda juhtumit ennast välja jättes. Kasutades niisugust ristkontrolli, saab hinnata prognoosiva meetodi täpsust treeningandmete piires. Selleks, et saada adekvaatset täpsushinnangut, tuleb koguda prognoosivast süsteemist täiesti sõltumatuid võrdlusandmeid. Sõltumatutel võrdlusandmetel põhinevat täpsushinnangut nimetatakse kontrolltäpsuseks. Selles uurimuses genereeriti kontrolltäpsuse saamiseks uued juhupunktid eraldistesse, mida ei kasutatud näidistel põhineva järeldamise süsteemi treenimisel.

Nominaalsete tunnuste kooskõla hindamiseks kasutasin K -analüüsi (Congalton, Green 1999; Tamm 2003). Meetod põhineb veamaatriksil ja on väga levinud kaugseire abil saadud kaartide ja hinnangute kooskõla leidmiseks. Kapa analüüs on diskreetne multivariatsiooniline tehnika, mida kasutatakse täpsusehinnangutes ja selleks, et statistiliselt määrata, kas üks veamaatriks on oluliselt erinev teisest. (Bishop *et al.* 1975 viidatud Congalton, Green 1999 järgi). Selle analüüsiga püütakse täpsushinnangu andmisel arvesse võtta klassifitseerimisülesandes esineda võivat juhuslikku komponenti. Termineid K (kapa) ja KHAT kasutatakse sünonüümidenä. K arvutatakse veamaatriksi jaoks ja see on mõõdupuuks, kui hästi klassifikatsioon vastab võrdlusandmetele. Kui $K > 0$ näitab see, et klassifikatsioon vastab võrdlusandmetele juhuslikust klassifikatsioonist paremini. K väärtus jääb -1 ja 1 vahele ja iseloomustab täpsust järgnevalt:

- $0,8$ — tugev vastavus;
- $0,4$ kuni $0,8$ — keskmine vastavus;
- $0,4$ kuni 0 — kehv vastavus;
- 0 — klasside juhusliku omistamise korral oodatav vastavus;
- <0 — väiksem kui juhusliku omistamise oodatav vastavus.

K leidmiseks kasutatakse valemit.

$$K = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} n_{+i}}{n^2 - \sum_{i=1}^k n_{i+} n_{+i}}, \quad (\text{valem 1})$$

kus:

n_{ii} — veemaatriksi lahter, sama numbriga rea ja veeru korral;

n_{i+} — rea summa;

n_{+i} — veeru summa.

K olulisust hinnatakse Z statistiku järgi. Z statistik näitab, kas klassifikatsioon on oluliselt parem, kui juhuslik tulemus. 95% usaldusväarsuse juures on kriitiline väärtus 1,96. Kui Z absoluutväärtus on suurem 1,96 on tulemus statistiliselt oluline ($p < 0,05$) ja võib tõdeda, et klassifikatsioon on parem kui juhuslik. Kapa analüüs tehti programmis MatCad. Usaldusväärse veahinnangu saamiseks tuleb kasutada piisavalt suurt valimit. Vajaliku valimimahu leidmiseks kasutasin valemit:

$$N = \frac{B * P * (1 - P)}{d^2}, \quad (\text{valem 2})$$

kus:

N — valimimaht;

B — X^2 (hii ruut) jaotuse väärtus, mille parameetriteks on 1 ja a/k ;

a — olulisuse tõenäosus, mille juures püütakse tulemusi kontrollida;

k — nominaalse tunnuse klasside arv;

P — klassi osakaal, mis on kõige lähemal 0,5-le;

d — lubatava vea suurus osakaaluna.

Kooskõla usaldusväärseks (olulisuse nivoo 0,05) hindamiseks sobivad valimimahud on ortofotode uuringu puhul järgnevad: tunnusel enamuspuuliik 653 ja arenguklass 695. Lubatud vea suuruseks võeti 0,05. Järelikult on nominaalsete tunnuste kooskõla hindamiseks vajaminev valimimaht 695. Satelliitpildi puhul olid vajalikud valimimahud järgmised: 573 enamuspuuliigi ja 659 arenguklassi puhul.

Pidevate tunnuste korral räägitakse prognoosi täpsusehinnangutest. Veaks loetakse siinkohal mitte valet klassifitseerimist, vaid tegeliku ja prognoositud väärtuse vahet. Tekkinud hälve võib olla nii positiivne kui negatiivne. Siinkohal tuleb silmas pidada, et ka siis, kui keskmine viga on null, võivad eksisteerida kõrvalekalded. Sellises olukorras positiivsed ja negatiivsed hälbed lihtsalt tasakaalustava üksteist ning keskmine prognoositud väärtus on tõene.

Vigade suurust on käesolevas töös iseloomustatud RMSE-ga. Kuigi RMSE leidmiseks võetakse ruutjuur keskmisest ruutveast, võib seda käsitleda keskmise veana iga prognoositud väärtuse jaoks, olenemata sellest, kas hälve on positiivne või negatiivne.

RMSE annab edasi vaid keskmise vea suurusjärku. Paljud hälvetest on suuremad ja paljud ka väiksemad kui RMSE. Vigade jaotust on vajalik teada näiteks RMSE 95 % usalduspiiride andmiseks. Statistikud on modelleerinud mitmeid vigade jaotusi, milledest aga tuntuim ja tähtsaim on normaaljaotus. Praktikas on paljud vigade jaotused just normaaljaotuse lähedased. Matemaatiliselt on näidatud, et normaaljaotusega vead tekivad siis, kui hälbe teket mõjutavad paljud juhuslikud faktorid ja nende faktorite mõjud kombineeruvad aditiivselt (Longley *et al.* 2001).

Teadustööde tulemused on näidanud, et suure lahutusvõimega andmete kasutamisel esinevad piksli tasemel suured RMSE väärtused. Ala suurenedes, mille kohta ruutkeskmisi vigu arvutatakse, usaldusväarsus kasvab. On täheldatud, et ristkontroll piksli tasemel annab prognoositäpsusele kehva hinnangu. (Tomppo *et al.* 2002) Täpsushinnang 1-ha puistu puhul on ~10 % parem kui üksikuid piksleid arvestaval hinnangul (Kilpeläinen & Tokola 1998). Arvestades eeltoodud tõdemusi on siinses uurimuses satelliitpildi andmetel põhineva prognoosisüsteemi kohta antud ka eraldisepõhised täpsushinnangud. Eraldisepõhise kontrolltäpsuse saamiseks valiti juhuslikult välja eraldised, mida ei kasutatud süsteemi treenimisel ning prognoositi iga eraldisesisesele pikslile väärtus. Seejärel leiti eraldisepõhine täpsushinnang arvutades eraldise piiridesse jäävate pikslite väärtuste mood või keskmine.

3. Tulemused ja arutelu

3.1. Tunnuste eelvalik

Sobiva eelvalikumeetodi väljaselgitamiseks viidi iga prognoositava tunnuse jaoks läbi kuus tehisõppeprotsessi. Tehisõpet viidi läbi mitu korda selleks, vähendada tehisõppeprotsessi juurde kuuluvate juhuslike otsustuste mõju. Eelvaliku meetodeid võrreldi keskmiste treeningtäpsuste järgi kasutades jäta-üks-välja-ristkontrolli (Tabel 3). Kõige suuremad erinevused kahe eelvalikumeetodi vahel esinesid puistu tagavara ja enamuspuiiigi puhul.

Erinevust kahe eelvalikumeetodi vahel suudeti olulisuse nivoo 0,05 juures tõestada tagavara ($p=0,008$) ja enamuspuiiigi ($p=0,047$) prognoosimise puhul. Eelpoolnimetatud kahel juhul andis klasteranalüüsi kasutamine eelvalikumeetodina halvema tulemuse. Arenguklassi prognoosimisel ei andnud klasteranalüüs üldistatud lineaarsete meetodite kasutamisest oluliselt kehvemaid tulemusi ($p=0,77$). Juurdekasvu prognoosimisel andis klasteranalüüsi kasutamine samm-regressioonist küll parema tulemuse, kuid erinevus ei olnud statistiliselt oluline ($p=0,23$). Kuna klasteranalüüsiga saadud tunnused on suhteliselt head kõigi funktsioontunnuste prognoosimiseks, väheneb võrreldes teiste eelvalikummeetodite kasutamisega arvutusteks kuluv aeg ja näidistel põhineva järeldamise süsteemi saab hoida lihtsamana. Seetõttu kasutasin argumenttunnuste eelvalikul klasteranalüüsi.

Tabel 3. Keskmised treeningtäpsused

	Tagavara, RMSE (%)	Juurdekasv, RMSE (%)	Arenguklass, K	Enamuspuiiik, K
Klasteranalüüs	34,01	37,80	0,49	0,63
Samm-regressioon	32,55	38,45	—	—
GLZ	—	—	0,50	0,66

Kirjandusest võib leida näiteid, kus eelvaliku tegemiseks on kasutatud korrelatsioonimaatriksit (Tuominen & Pekkarinen 2005). Korrelatsioonimaatriksit ei saa välja arvutada nominaalsete tunnuste jaoks ja 150 tunnuse puhul muutub maatriks väga

mahukaks ja selle kasutamine tülakaks. Samuti ei ole eelvaliku tegemiseks kindlasti vajalik teada kõikide argumenttunnuste omavahelisi korrelatsioone.

3.2. Parimad prognoosid ja tunnused ortofotodelt

Kokku 24 tehisõppeprotsessi tulemused on alljärgnevalt kokku võetud. Tabel 4 esitab nende hulgast parimad pikslipõhised treeningtäpsused ja samade parameetritega (maksimaalse sarnasuse määr; näidiste- ja tunnuste kaalud) ja näidistebaasiga saadud kontrolltäpsused (712 piksli põhjal). Täpsuse halvenemine prognoosimisel väljapoole uurimisala oli järgmine: arenguklass 53 %, enamuspuuliik 41 %, juurdekasv 21 % ja tagavara 23 %. Kirjandusest võib leida näiteid, kus kontrolltäpsused on treeningtäpsustest 3-15 % võrra halvemad (Hyypä *et al.* 2000). Täpsuse ulatuslik vähenemine viitab sellele, et näidistebaas võib olla liialt väike. Teisiti öeldes, näidistebaasis esinevate näidistega kirjeldatakse koguarieeruvusest ära liialt väike osa.

Võrreldes teiste uuringutega võib käesolevas uurimistöös saadud tulemusi pidada headeks. Näiteks on lähis-infrapuna kanaliga ortofotodega saadud tagavara hindamise puistupõhiseks (treening)täpsuseks $R^2 = 0,48$ (Hyypä *et al.* 2000). Paljudest avaldatud artiklitest jääb segaseks, kas esitatud täpsused on treeningtäpsused või mudeli ekstrapoleerimisel saadud kontrolltäpsused. Käesolevas uurimuses oli parim tagavara prognoosimise pikslipõhine treeningtäpsus $R^2 = 0,51$.

Tabel 4. Parimad pikslipõhised treeningtäpsused ja neile vastavad kontrolltäpsused

	Treeningtäpsus (K; RMSE (RMSE %))	Kontrolltäpsus (K; RMSE (RMSE %))
Arenguklass	0,51	0,24
Enamuspuuliik	0,64	0,38
Juurdekasv	1,54 tm ha ⁻¹ a ⁻¹ (38 %)	1,87 tm ha ⁻¹ a ⁻¹ (44 %)
Tagavara	62,02 tm ha ⁻¹ (33 %)	76,54 tm ha ⁻¹ (41 %)

Kasutades lokaalstatistikuid on saadud tagavara prognoosimise täpsushinnanguks 57,8 % RMSE (Tuominen & Pekkarinen 2005). Täpsete välitööde andmetega ja lähis-infrapuna kanaliga ortofotodega on saadud ka väga häid tulemusi, tagavara täpsushinnang 18-27 % RMSE (Muinonen *et al.* 2001). Oluline on, et siinses uurimuses kasutatud ortofotodel puudus lähis-infrapuna kanal, mis on taimkatte kaardistamise seisukohalt väga informatiivne. Samuti asjaolu, et täpsuse esitamine puistupõhiselt

üldjuhul parandab täpsust. Näiteks on täpsuse esitamine 1 ha puistu puhul parandanud täpsust ~10 % võrra, võrreldes pikslipõhise täpsusega (Kilpeläinen & Tokola 1998).

3.2.1. Parimad tunnused

Tabel 5 annab ülevaate tunnustest, mis andsid keskmiselt kõige paremaid prognoose: 30S_k, 30H_k, 30S_stdv. Vaid üks tunnus, 30S_k ehk küllastuse keskmine väärtus 30 meetri raadiuses, on esindatud kõigi nelja funktsioontunnuse 10 parema prognoosija hulgas. Kolmel korral on esindatud 30B_var, 30H_k, 30S_stdv. Tulemustest ilmnevad teatud seaduspärasused. Näiteks on tagavara puhul heaks indikaatoriks korduvalt peetud heleduse andmekihti, 30 meetrist raadiust ja lokaalstatistikutest keskmist. Kõiki funktsioontunnuseid koos vaadates on 30 m raadiuses arvutatud lokaalstatistikud andnud kõige paremaid tulemusi. Järgnevalt on analüüsitud tulemustes esinevaid seaduspärasusi tunnuse komponentide kaupa.

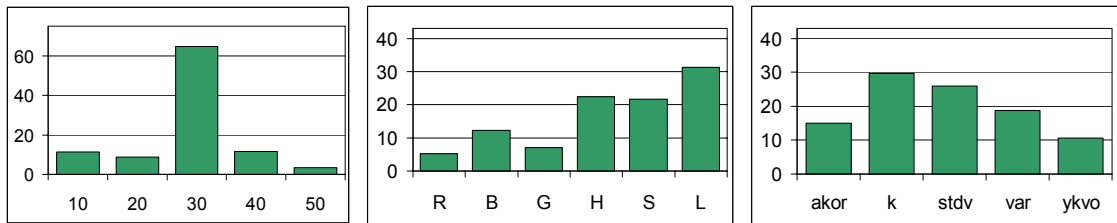
Tabel 5. Parimad argumenttunnused järjestatult keskmise ja kasutuskordadega kaalutud keskmise kaalu (veerg Kokku) järgi (tunnuse nimi koosneb raadiusest (m), andmekihist ja lokaalstatistikust; lühendid on seletatud meetoodika peatükis)

Arenguklass	Enamuspuuliik	Juurdekasv	Tagavara	Kokku
30S_k	30H_k	30H_ykvo	40L_k	30S_k
30H_stdv	30H_stdv	30H_k	20L_k	30H_k
30H_akor	30S_k	30S_var	30B_k	30S_stdv
30R_stdv	30G_stdv	30S_k	30S_k	30H_stdv
30G_stdv	30S_var	20L_akor	10L_stdv	30B_k
30S_stdv	30R_stdv	30H_var	30H_k	30R_stdv
10L_akor	40G_k	40B_stdv	30S_ykvo	30H_ykvo
50L_akor	40B_stdv	30B_var	50L_akor	30G_stdv
30B_var	40L_k	30S_stdv	30S_stdv	40S_k
10L_var	30B_var	20L_ykvo	20L_ykvo	40L_k

3.2.2. Tunnuse komponentide võrdlus

Kolme koostisosa (andmekiht, raadius ja lokaalstatistik), mille kombinatsioon üks argumenttunnus on, nimetatakse käesolevas uurimuses komponentideks. Tunnuse komponentide analüüsi eesmärgiga leida üldiseid seaduspärasusi, mida oleks kasulik teada tulevikus sarnaste uurimuste läbiviimisel.

Keskmistades tunnuse komponentide indikaatorväärtusi kõigi funktsioontunnuste prognoosimiseks osutus kõige selgemini eristatavaks seaduspäraks 30 meetri raadiuse suur indikaatorväärtus (Joonis 4). 30 m raadiuse suurt indikaatorväärtust võib seletada järgnevalt. Tulenevalt eraldiste keskmisest kujust ja pindalast (3,6 ha) satuvad juhupunktid tihti eraldiste servade lähedale ja tänu eraldiste maskile läheb suurte raadiuste korral rohkem andmete varieeruvust kaduma, kui väikeste raadiuste korral.

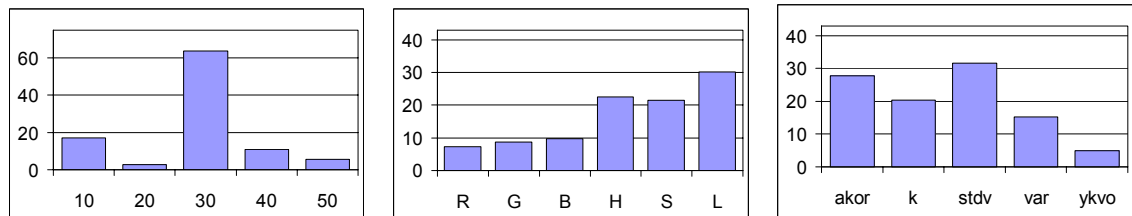


Joonis 4. Tunnuste komponentide võrdlus kõigi funktsioontunnuste alusel (lühendid on seletatud meetodika peatükis; horisontaalteljel komponentide väärtused, vertikaalteljel indikaatorväärtus (osakaal kaalutud keskmisest))

Arvutuslikult saadud ja RGB väärtusi sünteesivad HSL väärtused osutusid paremateks prognoosijateks ilmselt just sellel põhjusel, et kõik nad sisaldavad kombinatsiooni punasest, rohelisest ja sinisest värvist. Piksli väärtus ühes kanalis on juhuslikum ja sisaldab vähem informatsiooni kui sama piksli väärtus kolme kanali kombinatsioonina.

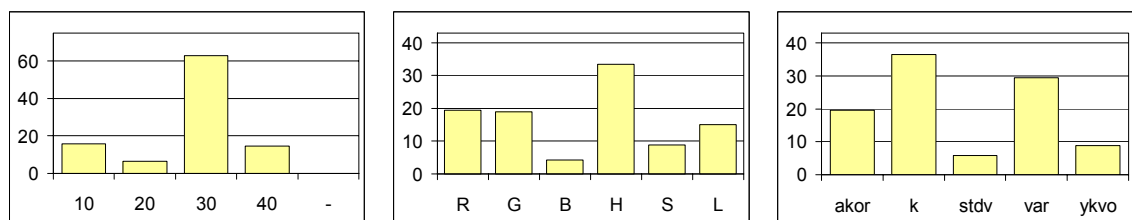
Lokaalstatistikutest osutus kõige paremaid prognoose andvaks keskmine pikslite väärtus. Ka kirjandusest esineb näiteid, kus ortofotode originaalkanaleid kasutades osutus kasutatud lokaalstatistikute hulgast kõige kasulikumaks keskmine (Tuominen & Pekkarinen 2005). Üks võimalik põhjendus, miks standardhälve oli variatsioonikoefitsiendist suurema indikaatorväärtusega on see, et variatsioonikoefitsiendi arvutamisel kasutatakse keskmist. Keskmine on aga paljudel juhtudel iseseisva tunnusena prognoosivasse komplekti võetud. Lähtudes eelnevast on standardhälbes sisalduv informatsioon antud kontekstis unikaalsem ja seetõttu väärtuslikum kui variatsioonikoefitsiendis.

Arenguklassi prognoosimisel oli kõige kasulikumaks ümbruse arvestamine 30 meetri raadiuses (Joonis 5). Andmekihtidest on selgelt parimad HSL. Lokaalstatistikute puhul on parim standardhälve. Häid prognoose on saadud ka tunnuste komplektidega, milles osaleb autokorrelatsiooni indeks Morani I.



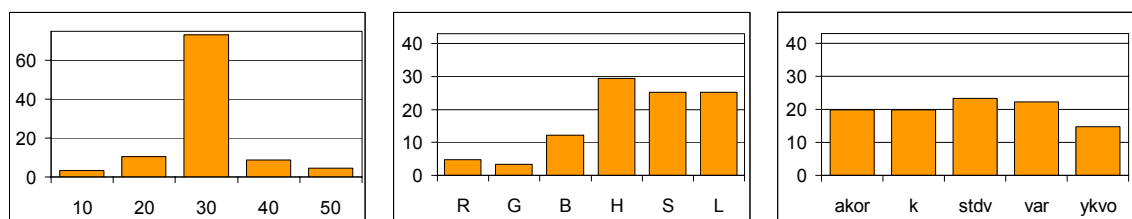
Joonis 5. Tunnuse komponentide võrdlus arenguklassi prognoosimisel (telgede seletus Joonis 4)

Veidi teistsugused seaduspärad ilmnevad enamuspüüliigi prognoosimisel (Joonis 6). Raadiuse 30 m indikaatorväärtuse ülekaal on ka siin väga suur. Enamuspüüliigi puhul on küllalt olulisteks peetud ka punase ja roheline kanali kasutamist, kuid parim on värvus. Lokaalstatistikutest andsid parimaid prognoose keskmine ja variatsioonikoefitsient. Üldjuhul on standardhälve ja variatsioonikoefitsient üksteist dubleerivad lokaalstatistikud ning väga harva valib tehisõpe prognoosimiseks välja mõlemad. Enamuspüüliigi puhul osutus paremaks variatsioonikoefitsiendi kasutamine.



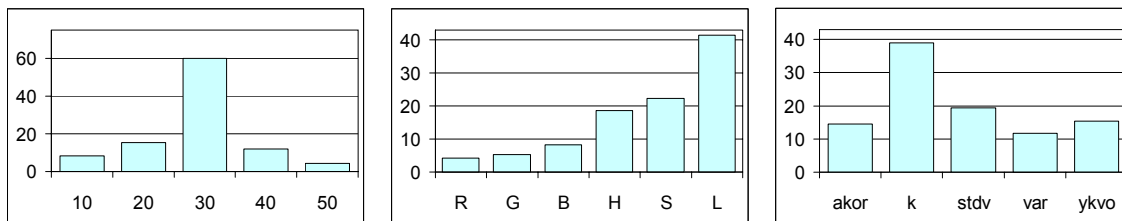
Joonis 6. Tunnuse komponentide võrdlus enamuspüüliigi prognoosimisel (telgede seletus Joonis 4)

Juurdekasvu prognoosimisel valis tehisõpe järjekordselt 30 m raadiuse teistest raadiustest oluliselt paremaks (Joonis 7). Andmekihtidest eelistatakse jällegi HSL vahemikke. Kõigi lokaalstatistikute kasutamine prognoosimisel on andud häid tulemusi.



Joonis 7. Tunnuse komponentide võrdlus juurdekasvu prognoosimisel (telgede seletus Joonis 4)

Ka tagavara prognoosimisel kehtib juba eelnevalt ilmnunud seaduspära, et ümber juhupunkti peetakse oluliseks arvestada 30 m raadiust (Joonis 8). Teistest oluliselt paremaks peetakse heleduse andmekihti ja lokaalstatistikutest keskmist.



Joonis 8. Tunnuse komponentide võrdlus tagavara prognoosimisel (telgede seletus Joonis 4)

3.3. Parimad prognoosid ja tunnused satelliidipildilt ja mullakaardilt

Satelliitpiltidelt takseerandmete prognoosimiseks kasutati Pidevstudiumi programmi. Kokku viidi läbi 679 tehisõppeprotsessi. Nendest parimad prognoosid leiti üles üldjuhul esimese 30-40 tehisõppe protsessi jooksul. Korduvate tehisõppeprotsesside läbiviimine oli siinkohal vajalik seetõttu, et igas õppeprotsessi võeti 40 võimalikust argumenttunnusest juhuslikult 15. Piisavalt paljude tehisõppeprotsesside korral saab hinnangu kõigi 40 tunnuse prognoosivõimest. 40 argumenttunnust ei võetud korruga tehisõppeprotsessi seetõttu, et vähemate tunnuste kasutamine kiirendab nii õppimist kui prognoosi ja reeglina piisab funktsioontunnuste prognoosimiseks kümnest tunnusest.

Tabel 6 esitab parimad funktsioontunnuste prognoosimisel saadud pikslipõhised treening- ja kontrolltäpsused (660 piksli põhjal) ja eraldisepõhised kontrolltäpsused (660 eraldise põhjal). Võrreldes mujal saadud tulemustega võib siinseid pidada headeks. Mäkelä ja Pekkarineni (2004) uurimuses on esitatud vead kogutagavara puhul piksli tasemel 60-80 % ja eraldise tasemel 48 % RMSE. Veidi parema tulemuse on spetsiaalset välitööd kasutades saanud Hyypä jt. (2000). Tagavara hindamise eraldisepõhiseks treeningtäpsuseks saadi $R^2 = 0,31$, mis vastab RMSE % väärtusele ~40 %.

Tabel 6. Parimad pikslipõhised treening- ja kontrolltäpsused ning eraldisepõhised kontrolltäpsused satelliidiandmeid ja mullakaarti kasutava prognoosi korral

	Pikslipõhine treeningtäpsus (K; RMSE (RMSE %))	Pikslipõhine kontrolltäpsus (K; RMSE (RMSE %))	Eraldisepõhine kontrolltäpsus (K; RMSE (RMSE %))
Arenguklass	0,50	0,17	0,23
Enamuspuuliik	0,66	0,31	0,41
Juurdekasv	1,61 tm ha ⁻¹ a ⁻¹ (42 %)	2,07 tm ha ⁻¹ a ⁻¹ (52 %)	1,98 tm ha ⁻¹ a ⁻¹ (49 %)
Tagavara	68,42 tm ha ⁻¹ (34 %)	87,00 tm ha ⁻¹ (42 %)	74,64 tm ha ⁻¹ (36 %)

Kuna käesolevas uurimuses kasutati riigimetsade takseerandmeid seadmata sealjuures miinimumnõudeid eraliste pindadele ja pööramata tähelepanu eraldiste kujule, siis osutavad saadud vastavused kasutatava meetodika suurele potentsiaalile. Treeningandmete aluseks olnud eraldiste keskmine pindala oli 3,4 ha. Mäkela ja Pekkarinen (2004) olid nii väikeste eraldiste kasutamise suhtes väga skeptilised ja arutlesid, et selliste treeningandmetega suudetakse anda vaid väga ligikaudseid hinnanguid.

3.3.1. Parimad tunnused

Võttes kokku parimad argumenttunnused, mida funktsioontunnuste prognoosimisel kasutati, leiti, et vaid üks tunnus on esindatud parimate tunnuste hulgas kõigi nelja funktsioontunnuse prognoosimisel, see on TM6.2 kanali standardhälve 80 meetri raadiuses (Tabel 7). Kolmel korral on parimate tunnuste hulgas esindatud TM5 üksikpiksli väärtus, TM6.1 standardhälve 40 m raadiuses ja TM7 kauguse pöördväärtusega kaalutud Morani I väärtus 120 m raadiuses.

Tabel 7. Parimad tunnused järjestatud vastavalt indikaatorväärtusele (wMI — kauguse pöördväärtusega kaalutud Morani I, ülejäänud lühendid on seletatud Metoodika peatükis)

Arenguklass	Enamuspuuliik	Juurdekasv	Tagavara	Kokku
TM5	TM4 kesk 80m	TM6.2 stdv 80m	TM6.2 stdv 80m	TM6.2 stdv 80m
TM6.1 stdv 40m	TM7	TM6.1 stdv 40m	TM6.1 stdv 40m	TM5
TM6.2 stdv 80m	TM5 kesk 80m	TM4 wMI 80m	TM3 wMI 40m	TM6.1 stdv 40m
TM7	TM3	TM1 var 40m	TM7 wMI 120m	TM4 kesk 80m
TM6.2 kesk 80m	TM4	TM5	TM6.2 wMI 120m	TM7
TM6.2	TM5	TM6.2 wMI 120m	TM4 wMI 80m	TM5 kesk 80m
TM6.1	TM6.2	TM5 wMI 40m	TM5 wMI 40m	TM7 wMI 120m
TM7 wMI 120m	TM3 stdv 120m	TM7 wMI 120m	TM4 var 80m	TM4
TM2 kesk 80m	TM6.2 stdv 80m	muld	TM4	TM4 wMI 80m
TM7 var 40m	TM6.2 kesk 80m	muld mood 40m	TM1 var 40m	TM6.2 wMI 120m

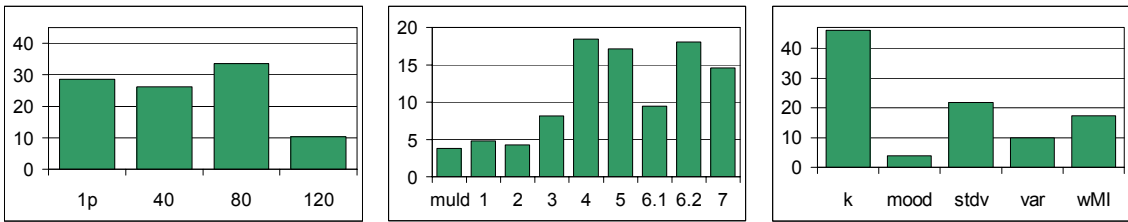
3.3.2. Tunnuse komponentide võrdlus

Järgnevalt on analüüsitud satelliidipildi tunnuste komponente eraldi. Üksikpikslite väärtusi mõõtmega 25×25 m on tehisõpe valinud paljudel juhtudel headeks prognoosijateks (Joonis 9). See võib tuleneda treeningandmete saamiseks kasutatud eraldiste väiksusest. Raadiustest on teistest oluliselt kehvema indikaatorväärtusega 120 m raadiuse kasutamine. Kuna satelliitpiltide puhul kasutati lokaalstatistikute väljaarvutamisel eraldistemaski asemel põhikaardimaski sisaldub neis nii suurte raadiuste korral ilmselt liialt palju naabereraldiste andmeid.

Parimate tunnuste seas on väga vähe silmale nähtava spektrivahemiku kanaleid (1, 2, 3). 1:10000 mullakaardi kasutamine on kokkuvõttes kõige väiksema indikaatorväärtusega. See võib tuleneda mullaeraldiste paiknevuste erinevusest võrreldes metsamajanduslike eraldistega. Kõige paremaks kanaliks osutus ootuspäraselt lähis-infrapuna kanal 4, järgnesid kõrge radiomeetrilise lahutusvõimega soojus-infrapuna kanal 6.2 ja keskinfrapunakanalid 5 ja 7. Inimestele paistavad taimelehed rohelistena, kuid laiemat elektromagnetlainete vahemikku registreeriva sensori jaoks on ülekaalus lähis-infrapuna vahemik. Just seetõttu on kasutades 4. kanalit võimalik taimkatte kohta saada kõige rohkem informatsiooni.

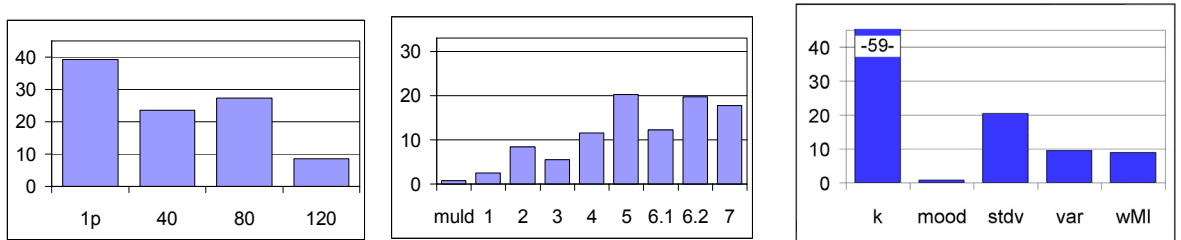
Parimaks lokaalstatistikuks on pikslite keskmine väärtus. Veidi üllatav on küllaltki suur kauguse pöördväärtusega kaalutud Morani I indikaatorväärtus, sest eelnevates

uuringutes ei ole autokorrelatsiooni indeksid nii head tulemust andnud. Autokorrelatsioon iseloomustab pildi laigulisust, mis on tihedalt seotud võrade ja varjude suurusega.



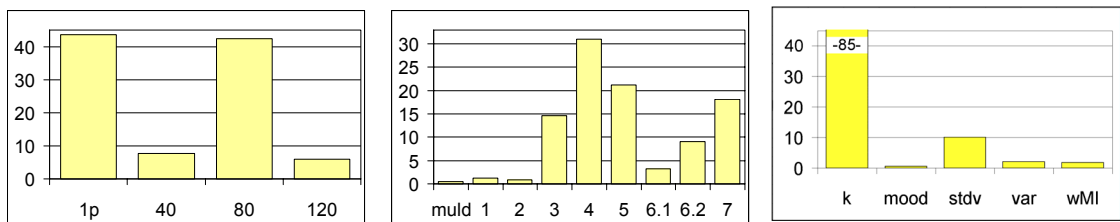
Joonis 9. Tunnuste komponentide indikaatorväärtused kõigi funktsioontunnuste keskmistena (lühendid on seletatud metoodika peatükis; horisontaalteljel komponentide väärtused, vertikaalteljel indikaatorväärtus (osakaal kaalutud keskmisest)); 1p — üksikpikslite väärtus, 1-7 Landsat ETM+ kanalid)

Järgnevalt on vaadeldu tunnuse komponentide indikaatorväärtusi iga funktsioontunnuse jaoks eraldi. Arenguklassi prognoosimisel osutusid kõige informatiivsemateks Landsat ETM+ üksikpikslite väärtused, kesk-infrapuna ja soojus-infrapuna kanalid ning lokaalstatistikute keskmine (Joonis 10).



Joonis 10. Tunnuse komponentide võrdlus arenguklassi prognoosimisel (telgede seletus Joonis 9)

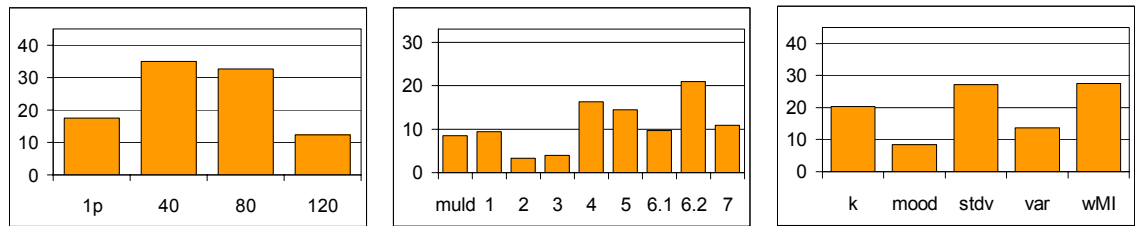
Enamuspuuliigi puhul olid võrdselt kasulikud üksikpikslite väärtused ja 80 raadiuses arvatud statistikud. Kanalitest olid parimad lähis-infrapuna ja kesk-infrapuna. Lokaalstatistikute ülekaalukalt keskmine (Joonis 11).



Joonis 11. Tunnuse komponentide võrdlus enamuspuuliigi prognoosimisel (telgede seletus Joonis 9)

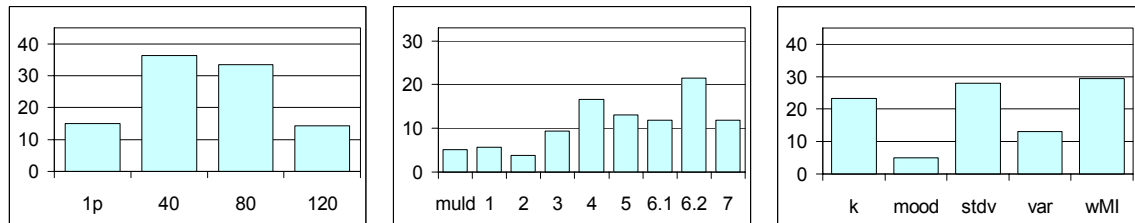
Juurdekasvu prognoosimisel olid raadiustest parimad 40 ja 80 m, soojus-infrapuna ja lähis-infrapuna ning lokaalstatistikute standardhälve ja kauguse pöördväärtusega

kaalitud Morani I. 1:10000 mullakaart leidis juurdekasvu prognoosimisel suuremat kasutust, kui enamuspuuliigi ja arenguklassi puhul (Joonis 12).



Joonis 12. Tunnuse komponentide võrdlus juurdekasvu prognoosimisel (telgede seletus Joonis 9)

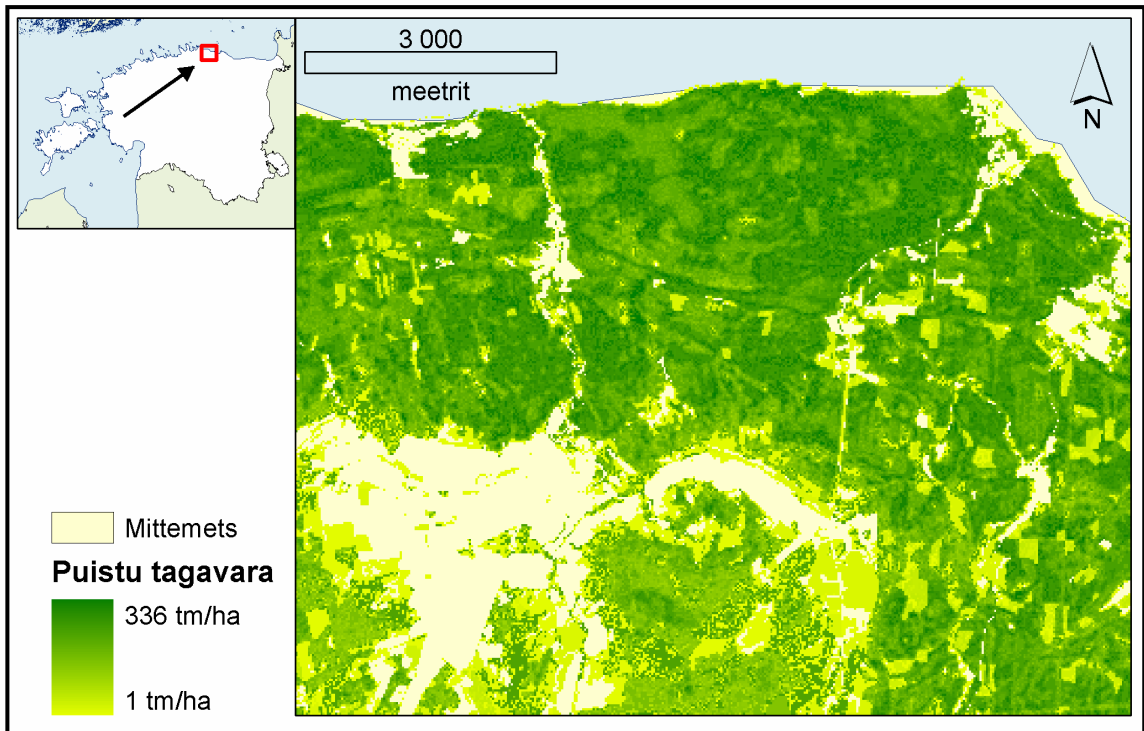
Tagavara prognoosimisel saadud tunnuste komponentide indikaatorväärtustes võib näha sarnasusi juurdekasvuga. Raadiustest on samuti parimad 40 ja 80 m, kanalitest soojus-infrapuna ja lähis-infrapuna ning lokaalstatistikutest kauguse pöördväärtusega kaalitud Morani I ja standardhälve (Joonis 13).



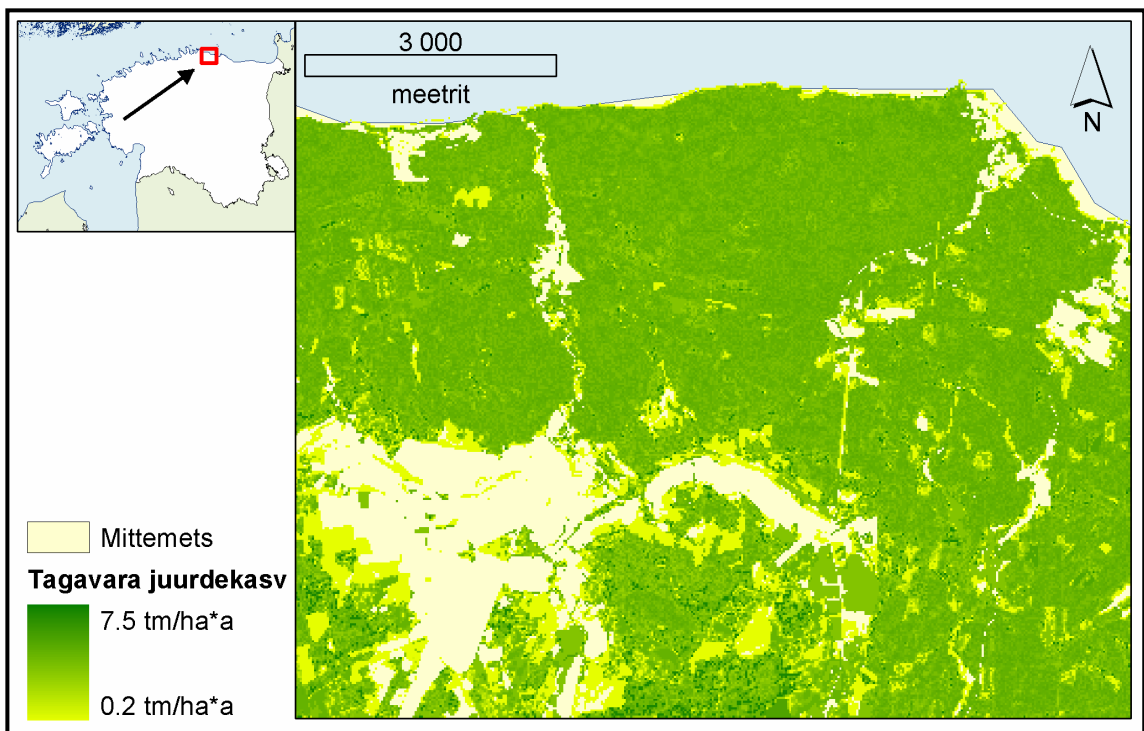
Joonis 13. Tunnuse komponentide võrdlus puistu tagavara prognoosimisel (telgede seletus Joonis 9)

3.4. Prognoosikaartide koostamine

Tagavara ja juurdekasvu prognoosikaartide (Joonis 14, Joonis 15) koostamisel kasutati satelliitpilte ja mullakaarti sisaldavat näidistel põhineva järeldamise süsteemi. Tabel 6 annab ülevaate nende kahe tunnuse prognoosimisel saavutatud parimatest täpsustest. Kaardistatud ala piires arvutati välja eraldisepõhine kontrolltäpsus: tagavara puhul 85,8 tm ha⁻¹ (39 %) ja juurdekasvu puhul 2,0 tm ha⁻¹ a⁻¹ (43 %) RMSE.



Joonis 14. Proгноositud puistu tagavara kaart



Joonis 15. Proгноositud puistu tagavara juurdekasvu kaart

Tunnuste puistutüüp ja puistu vanuseklass progноosimisel saadud treeningtäpsused olid paljulubavad (Tabel 8, Tabel 9). Treeningtäpsused on saadud kasutades takseerandmete progноosimiseks loodud näidistebaasi. Puistutüüpidest suudeti kõige paremini progноosida segametsa. Vanuseklassidest õnnestus kõige paremini progноosida keskealist ja noort metsa.

Tabel 8. Pikslipõhised treeningtäpsused puistutüübi prognoosimisel

	K	Z
Okasmets	0,67	17,1
Lehtmets	0,70	15,2
Segamets	0,72	7,5
Kokku	0,69	22,4

Tabel 9. Pikslipõhised treeningtäpsused puistu vanuseklassi prognoosimisel

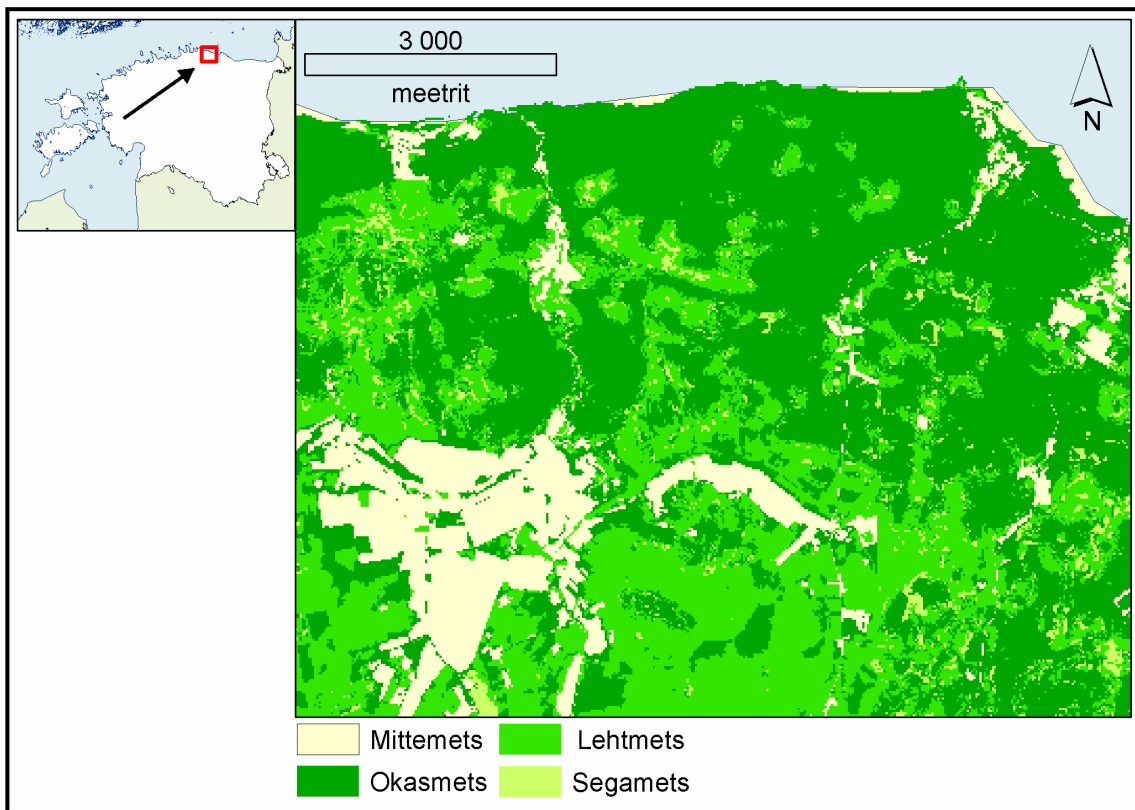
	K	Z
Küps ja valmiv mets	0,63	20,5
Keskealine ja noor mets	0,69	22,0
Kokku	0,66	26,1

Mõlema tunnuse jaoks koostati ka prognoosikaardid 10 × 10 km põhikaardilehe (7402) piires (Joonis 16, Joonis 17). Kasutades kõiki sellele kaardilehele jäävaid riigimetsaeraldiste andmeid arvutati välja pikslipõhine kontrolltäpsus (Tabel 10, Tabel 11). Täpsushinnangute 2-3 kordset langust võrreldes pikslipõhiste treeningtäpsustega saab seletada sellega, et näidistebaasis, mis koosnes 907 potentsiaalsest näidiseist ei esinenud piisavalt häid näidiseid selleks, et kirjeldada ära 49736 pikslit kaardil, mille põhjal kontrolltäpsus on esitatud. Treeningtäpsus on aga esitatud põhinedes jäta-üks-välja-ristkontrollil ja näidistebaasi 907 juhupunkti andmetel. Osa veast on põhjustatud ka sellest, et võrdlusandmetena kasutatud takseerandmed on esitatud eraldiseppõhiselt. See tähendab, et eraldiste siseselt ei esine takseerandmetes mingisugust varieeruvust. Looduses ja ka pildiandmetelt välja lugedes olukord selline kindlasti ei ole. Kogu eraldistesisene varieeruvus, mis põhimõtteliselt võis olla ka tegelikkusele vastav läks läbiviidud kontrolltäpsuse arvutamisel kirja veana.

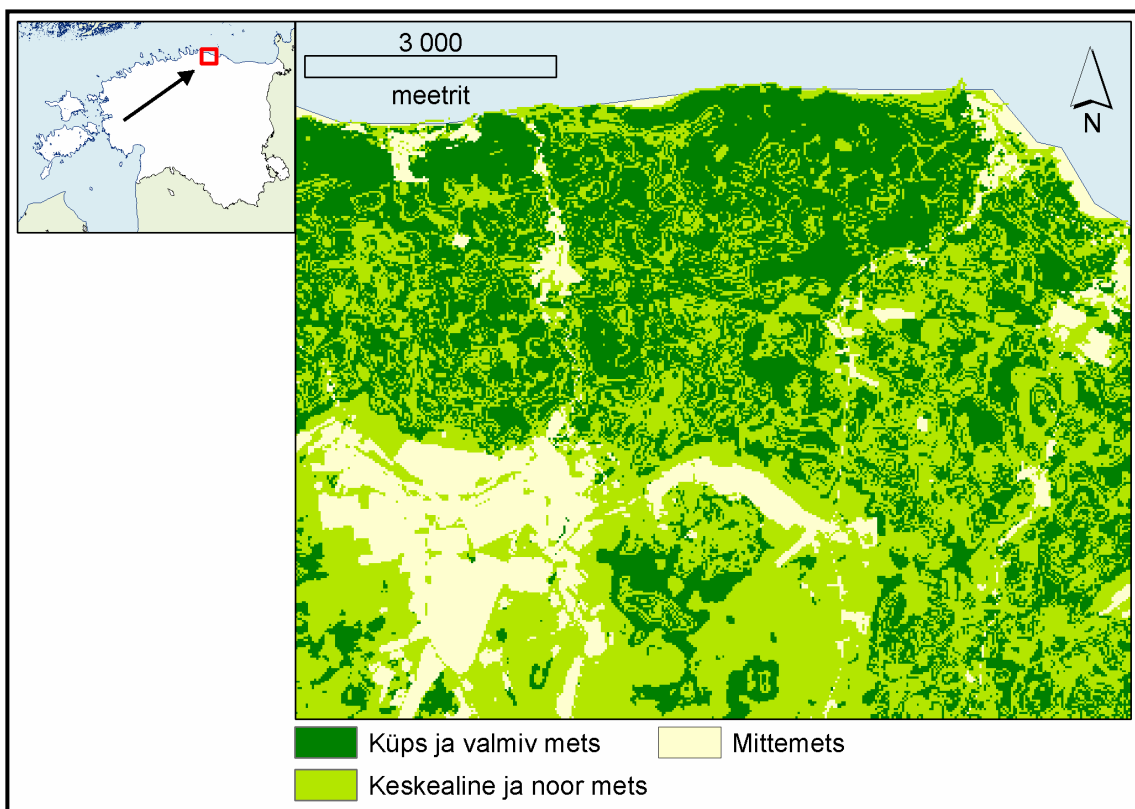
Pikslipõhiste täpsushinnangute puuduste leevendamiseks arvutati välja ka eraldistepõhine kontrolltäpsus (Tabel 10, Tabel 11). Ootuspäraselt on eraldiseppõhised täpsushinnangud veidi paremad, kui pikslipõhised.

Selleks, et teada saada, kas saadud prognoosikaardi ja metsaeraldiste kaardi kokkulangevus erineb oluliselt juhuslikust kasutati Monte-Carlo meetodil põhinevat toroidnihutuste programmi. Juhunihtustel põhineva programmi on loonud Kalle

Remm. Mõlema kaardi puhul erines kõikide klasside prognoosimine oluliselt juhuslikust.



Joonis 16. Prognoositud puistutüüpide kaart



Joonis 17. Prognoositud puistu vanuseklasside kaart

Tabel 10. Piksli- ja eraldisepõhised kontrolltäpsused puistutüübi prognoosimisel

	Pikslipõhine kontrolltäpsus		Eraldisepõhine kontrolltäpsus	
	<i>K</i>	<i>Z</i>	<i>K</i>	<i>Z</i>
Okasmets	0,44	89,56	0,40	4,13
Lehtmets	0,31	57,40	0,60	4,30
Segamets	0,13	12,04	-0,12	-3,40
Kokku	0,36	91,73	0,44	4,79

Tabel 11. Piksli- ja eraldisepõhised kontrolltäpsused puistu vanuseklassi prognoosimisel

	Pikslipõhine kontrolltäpsus		Eraldisepõhine kontrolltäpsus	
	<i>K</i>	<i>Z</i>	<i>K</i>	<i>Z</i>
Küps ja valmiv mets	0,16	41,50	0,22	3,01
Keskealine ja noor mets	0,22	42,03	0,36	3,16
Kokku	0,19	41,62	0,28	3,05

Kokkuvõte

Magistritöös kasutati riigimetsade takseerandmeid, ortofotosid, satelliitpilte, 1:10000 mullakaarti ja 1:10000 põhikaarti näidistel põhineva järeldamise süsteemi (*Case-Based Reasoning* — *CBR*) loomiseks. CBR-süsteeme ja tehisõpet on taimkatte kaardistamisel seni vähe kasutatud. Uurimistöö andis kinnitust riigimetsade takseerandmete kasutatavuse kohta õpetusandmetena pildi- ja kaardiandmete abil metsaparameetrite prognoosimiseks.

Uurimistöö tulemustena selgus, et tehisõppele eelnevaks tunnustevalikuks võib kasutada klasteranalüüsi, mis aitab hoida erinevate funktsioontunnuste prognoosiks kasutatavate tunnuste arvu võimalikult väiksena. Kasutatavate lähteandmete ja meetoodikaga saavutatavad parimad pikslipõhised kontrolltäpsused (712 piksli põhjal) olid ortofotode kasutamisel järgmised: puistu tagavara $76,54 \text{ tm ha}^{-1}$ (41 %) RMSE, juurdekasv $1,87 \text{ tm ha}^{-1} \text{ a}^{-1}$ (44 %) RMSE, arenguklass $K = 0,24$ ja enamuspuuliik $K = 0,38$. Sealjuures parimad tunnused olid küllastuse (*saturation*) keskmine väärtus 30 m raadiuses, värvuse (*hue*) keskmine väärtus 30 m raadiuses ja küllastuse standardhälve 30 m raadiuses. Üldistatumalt väljendades saadi häid prognoose kasutades 30 m suurust raadiust, andmekihina heleduse (*lightness*) väärtusi ja lokaalselt arvatud keskmist.

Satelliitpildi kasutamisel olid eraldisepõhised kontrolltäpsused (660 eraldise põhjal) järgmised: puistu tagavara $74,64 \text{ tm ha}^{-1}$ (36 %), juurdekasv $1,98 \text{ tm ha}^{-1} \text{ a}^{-1}$ (49 %), arenguklass $K = 0,23$ ja enamuspuuliik $K = 0,41$. Parimad tunnused olid soojus-infrapuna (TM6.2 — kõrge radiomeetiline lahutusvõime) väärtuse standardhälve 80 m raadiuses, kesk-infrapuna (TM5) üksikpikslite väärtused ja soojus-infrapuna (TM6.1 — madal radiomeetiline lahutusvõime) väärtuse standardhälve 40 m raadiuses. Parimaid prognoose saadi kombineerides 80 m suurust raadiust, lähis-infrapuna (TM4) ja soojus-infrapuna (TM6.2) kanali väärtusi ning lokaalselt arvatud keskmist.

Üheks uurimistöö tulemuseks on puistu tagavara (Joonis 14), tagavara juurdekasvu (Joonis 15), puistutüüpide (okasmets, lehtmets ja segamets) (Joonis 16) ja puistu vanuseklasside (küps ja valmiv mets ning keskealine ja noor mets) (Joonis 17) prognoosikaardid.

Tulevikus tuleks uurida, kuidas mõjutab prognoositäpsust see, kui vaatluspunktide (potentsiaalsete näidiste) asukoha eraldises või üldisemalt looduses valib välja ekspert ning NDVI (*normalized difference vegetation index*) ja kanalisuhete (N: TM6.2/TM4) kasutamine argumenttunnustena. Samuti tuleks uurida, kui palju tõstaks prognoositäpsust mahukamate ning iga prognoositava tunnuse jaoks esinduslike näidistebaaside loomine.

Viited

- Aha** D.W. 1998. The Omnipresence of Case-Base Reasoning in Science and Application. *Knowledge-Based Systems*, 261–273.
- Britannica**, E., 1989. The new Encyclopaedia Britannica. Chicago, Encyclopaedia Britannica, viidatud Skidmore, A. (Editor) 2002. Environmental Modelling with GIS and Remote Sensing. Taylor & Francis. kaudu.
- Bütler**, R., Schlaepfer, R. 2004. Spruce snag quantification by coupling colour infrared aerial photos and a GIS. *Forest Ecology and Management* 195, 325-339.
- Chen**, J. M., Liu, J., Leblanc, S. G., Lacaze, R., Roujean, J-L. 2003. Multi-angular optical remote sensing for assessing vegetation structure and carbon absorption. *Remote Sensing of Environment* 84, 516-525.
- Congalton**, R.G., Green, K. 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Lewis Publishers.
- Congalton**, R. G., Birch, K., Jones, R., Schriever, J. 2002. Evaluating remotely sensed techniques for mapping riparian vegetation. *Computers and Electronics in Agriculture* 37, 113-126.
- Colwell** R.N. (Ed) 1983. Manual of Remote Sensing 2nd Ed., Falls Church, VA: American Society of Photogrammetry, viidatud Jensen J.R. 2000. Remote Sensing of the Environment: an Earth Resource Perspective. Prentice Hall Series in Geographic Information Science kaudu.
- Eltermann**, V. (toimetaja), Kohava, P. (koostaja) 2001. Eesti metsad 2000. Metsavarude hinnang statistilisel valikmeetodil. OÜ Eesti Metsakorralduskeskus.
- Etverk**, I. 1980. Metsamajanduse teatmik. Tallinn, Valgus.
- Gemmel** F., Varjo J., Strandstrom M. 2001. Estimating Forest Cover in a Boreal Forest Test Site Using Thematic Mapper Data from Two Dates. *Remote Sensing of Environment*, 77, 197-211.
- Guisan**, A., Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147-186.
- Hall**, A., Louis, J., Lamb, D. 2003. Characterising and mapping vineyard canopy using high-spatial-resolution aerial multispectral images. *Computers & Geosciences* 29, 813-822.
- Halme**, M., Tomppo, E. 2001. Improving the accuracy of multisource forest inventory estimates by reducing plot location error — a multicriteria approach. *Remote Sensing of Environment* 78, 321-327.
- Haralick**, R.M. 1979 Statistical and structural approaches to texture. *Proceedings of the IEEE* 67(5), 786-804.
- Heikkilä**, J., Nevalainen, S., Tokola, T. 2002. Estimation defoliation in boreal coniferous forests by combining Landsat TM, aerial photographs and field data. *Forest Ecology and Management* 158, 9-23.
- Herold**, N. D., Haack B. N., Solomon, E. 2004. An evaluation of radar texture for land use/cover extraction in varied landscapes. *International Journal of Applied Earth Observation and Geoinformation* 5, 113-128.
- Holopainen**, M., Lukkarinen, E. 1994. Digitaalisten ilmakuvien käyttö metsävarojen inventoinnissa. English summary: digital aerial photographs in forest inventory. Department of Forest Resource Management, University of Helsinki (Publications 4), viidatud Muinonen, E., Maltamo, M., Hyppänen, H., Vainikainen, V. 2001. Forest stand characteristics estimation using a most similar neighbour approach and image spatial structure information. *Remote Sensing of Environment* 78, 223-228, kaudu
- Hüllermeier** E. 2001. Similarity-Based Inference as Evidential Reasoning. *International Journal of Approximate Reasoning* 26, 67–68
- Hyypä**, J., Hyypä, H., Inkinen, M., Engdahl, M., Linko, S., Zhy, Y. H. 2000. Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecology and Management* 128, 109-120.
- Jensen** J.R. 2000. Remote Sensing of the Environment: an Earth Resource Perspective. Prentice Hall Series in Geographic Information Science.

- Katila**, M., Tomppo, E. 2001. Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sensing of Environment* 76, 16-32.
- Kilpeläinen P.** & Tokola T. 1998. Gain to be achieved from stand delineation in LANDSAT TM image-based estimates of stand volume. *Forest Ecology and Management* 124. 105-111.
- Kull E.** 1999. Palumetsade takseerparameetrite hindamine kaugseire meetodidel. Magistritöö. Juhendaja: Urmas Peterson. Eesti Põllumajandusülikool.
- Kuusik**, A., Nilson, T. 2000. A directional multispectral forest reflectance model. *Remote Sensing of Environment* 72, 244-252.
- Lang**, M. 2001. Satelliitpildi ja metsandusliku andmebaasi sidumine metsa heleduse mudeli abil. Magistritöö. Juhendaja: geofüüsikadoktor Tiit Nilson. Eesti Põllumajandusülikool.
- Lonard**, R. I., Judd, F. W., Everit, J.H., Escobar, D.E., Davis, M. R., Crawford M. M., Desai, M. D. 2000. Evaluation of color-infrared photography for distinguishing annual changes in riparian forest vegetation of the lower Rio Grande in Texas. *Forest Ecology and Management* 128, 75-81.
- Longley**, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. 2001. Geographic Information Systems and Science. John Wiley & Sons, LTD.
- Meiner A.** (toimetaja) 1999. Eesti maakate. CORINE Land Cover projekti täitmine Eestis. Tallinn. 3-13.
- Masing V.** 1992. Ökoloogia leksikon. Loodusteaduslik oskussõnastik. Eesti Entsüklopeedia kirjastus.
- Miller**, D. R., Quine, C. P., Hadley, W. 2000. An investigation of the potential of digital photogrammetry to provide measurements of forest characteristics and abiotic damage. *Forest Ecology and Management* 135, 279-288.
- Moya**, I., Camenen, L., Evain, S., Goulas, Y., Cerovic, Z. G., Latouche, G., Flexas, J., Ounis, A. 2004. A new instrument for passive remote sensing I. Measurements of sunlight-induced chlorophyll fluorescence. *Remote Sensing of Environment* 91, 186-197.
- Mäkelä**, H., Pekkarinen, A. 2004. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. *Forest Ecology and Management* 196, 245-255.
- Nilson T.** 1994. Metsade kaugseire alused. Eesti Põllumajandusülikooli Metsakorralduse Instituut. Eesti Metsaamet. Tartu, 44-51; 83-112.
- Pekkarinen**, A. 2002. Image segment-based spectral features in the estimation of timber volume. *Remote Sensing of Environment* 82, 349-359.
- Remm**, K. 2002. Otepää looduspargi taimkatte kasvukohatüüpide kaart. *Eesti süsinikubilansi ökoloogiast ja ökonoomikast*. Toimetanud T. Frey. 62-76.
- Remm**, K & Luud, A 2003. Regression and point pattern models of moose distribution in relation to habitat distribution and human influence in Ida-Viru county, Estonia. *Journal for Nature Conservation* 11, 197-211.
- Remm K.** 2004. Case-based predictions for species and habitat mapping. *Ecological Modelling*, 177, 259-281.
- Schank R.C.** 1982. Dynamic Memory. A Theory of Reminding and Learning in Computers and People. Cambridge University Press.
- Skidmore**, A. (Editor) 2002. Environmental Modelling with GIS and Remote Sensing. Taylor & Francis.
- Tomppo**, E., Nilsson, M., Rosengren, M., Aalto, P., Kennedy, P. (2002). Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. *Remote Sensing of Environment*, 82, 156-171.
- Tomppo**, E., Halme, M. 2004. Using coarse scale forest variables as ancillary information and weighting of variables in *k*-NN estimation: a genetic algorithm approach. *Remote Sensing of Environment* 92, 1-20.
- Tuominen**, S., Pekkarinen, A. 2005. Performance of different spectral and textural aerial photograph features in multi-source forest inventory. *Remote Sensing of Environment* 94, 256-268.
- Wilson D.R.**, Martinez T.R. 2000. Reduction Techiques for Instance-Based Learning Algorithms. *Machine Learning*, 38, 257-286.

Interneti viited

Aastaraamat Mets '01; <http://www.envir.ee/metsandus/statistika/statistika.html> :10.05.2005

IMAGE 2000; <http://image2000.jrc.it/> :17.04.2005

ISI International Statistical Institute; <http://www.cbs.nl/isi/glossary> :05.04.2005

Elektrooniline Riigi Teataja; "Metsa korraldamise juhendi" kinnitamine. Vastu võetud keskkonnaministri 17. märtsi 1999. a määrusega nr 25 ([RTL 1999, 69, 902](#))
<https://www.riigiteataja.ee/ert/act.jsp?id=723067> :10.05.2005

Käsikirjalised viited

Eesti põhikaardi juhend 2002. Eesti põhikaardi 1:10000 digitaalkaardistuse juhend. Maa-ameti kartograafiabüroo. Tartu.

OÜ Eesti Metsakorralduskeskus 2002. Välitööde juhend.

STATISTICA Electronic Manual. STATISTICA 6.

Remm, K in press. Correlations between forest stand diversity and landscape pattern in Otepää NP, Estonia. *Journal for Nature Conservation*.

Tamm, T. 2003. Tehisõppe rakenduse katsetamine satelliitpiltide ja kaardiandmete järgi puistutüübi kaardistamiseks. Bakalaureusetöö keskkonnatehnoloogias. Juhendaja Kalle Remm. Käsikiri Geograafia Instituudi raamatukogus.

Using National Forest Inventory data in the remote sensing of Estonian forests with a machine learning approach

Summary

There have been quite few investigations that try to use case-based reasoning methodology in vegetation mapping. Locally computed image texture statistics, a case-based reasoning (CBR) system and a machine learning modules that was designed at the Institute of Geography Tartu University by Kalle Remm, were used for mapping forest attributes. There were three major goals to achieve: 1) to find a suitable method for feature preselection to be used before machine learning; 2) to predict forest attributes, including total stand volume, annual increase in stand volume, stand growth classes and dominant tree species in stands; 3) to retrieve best features that were used in forest attributes mapping; 4) to compare usefulness of different feature components (kernel size, type of explanatory variable, type of locally computed statistic) and 5) to produce maps of forest attributes. The results of the posed objectives were as follows: although in the comparison of former feature selection methods two statistically significant differences out of four comparisons were discovered (between cluster analysis vs generalized linear models (GLZ) and step-by-step correlation) it is suggested to use cluster analysis. The best control accuracy based on forest stands using satellite images was as follows: for stand volume 74.64 m³ ha⁻¹ (36 %) RMSE, annual increase in stand volume 1.98 m³ ha⁻¹ a⁻¹ (49 %), stand growth classes $K = 0.23$ ja dominant tree species in stands $K = 0.41$. Using orthophotos pixel based control accuracy was as follows: stand volume 76.54 m³ ha⁻¹ (41 %) RMSE, annual increase in stand volume 1.87 m³ ha⁻¹ a⁻¹ (44 %), stand growth classes $K = 0.24$ and dominant tree species in stands $K = 0.38$. For orthophotos mean saturation in the 30 m radius and for satellite images standard deviation of Landsat TM6.2 values in 80 m radius were the best features. The best feature components for orthophotos were 30 m kernel, lightness and mean and for satellite images 80 m kernel radius, near-infrared and mean. A map of stand volume, increase in stand volume, stand types (3 classes) and a map of stand age (2 classes) were computed. National Forest Inventory data showed its potential as training data for remote sensing applications and more extensive databases of carefully chosen training samples should be developed to increase the accuracy of predictions.