

Versioon 4.0
Kalev Pärna 13.02.2023



Statistilise metsainventuuri (SMI) traktivõrgu analüüs ja arvutusmeetoditega seotud uuring

Tartu Ülikooli matemaatika ja statistika instituudi (MSI) lõpparuanne

Tartu Ülikool

Reg. kood 74001073

Ülikooli 18, Tartu 50090

E-post: info@ut.ee

Kodulehekülg: www.ut.ee

Versioon 4.0
Kalev Pärna 13.02.2023



Versiooni nr.	Kuupäev	Muudatus	Autor
1	16.12.2022	Lõpparuanne	Kalev Pärna, TÜ MSI
2	29.01.2023	Lõpparuande täiendatud versioon	Kalev Pärna, TÜ MSI
3	10.02.2023	Lõpparuande redigeeritud versioon	Kalev Pärna, TÜ MSI
4	13.02.2023	Lõpparuande turvaline versioon	Kalev Pärna, TÜ MSI

Sisukord

1.	Sissejuhatus ja tulemuste kokkuvõte	5
	Tulemuste lühikokkuvõte	6
2.	Traktivõrgu analüüs.....	9
2.1.	Lühiülevaade praegusest traktisüsteemist	9
2.2.	Alaliste traktide paigutus.....	9
2.2.1.	Konfidentsiaalsus.....	9
2.3.	Ajutiste traktide paigutus	10
2.4.	Kluppimisraadiuse optimaalsuse analüüs	11
2.4.1.	Kluppimisraadiuste 10m ja 7m võrdlev analüüs.....	11
2.4.2.	Mahuhinnangute standardvea graafikud (sõltuvus kluppimisraadiusest)	13
2.4.3.	Proovitükkide arvu seos kluppimisraadiusega	16
2.4.4.	Kahe erineva kluppimisraadiuse kasutamine	17
2.4.5.	Kokkuvõte kluppimisraadiuse analüüsist.....	18
3.	Ruumilised korrelatsioonid	19
3.1.	Proovitükkide vahelised korrelatsioonid	19
3.2.	Traktidevaheline ruumiline korrelatsioon.....	21
3.3.	Ruumiliste korrelatsioonide integreerimine hindamisprotsessi	24
4.	Hindamine ja veaarvutus	25
4.1.	Pindalade ja keskmiste hindamine ja veaarvutus	25
4.1.1.	Pindala hindamine	25
4.1.2.	Pindala hinnangu veaarvutus	26
4.1.3.	Soovituslik veaarvutuse skeem	27
4.2.	Kogutagavara hindamine ja veaarvutus	29
4.3.	Soovitusliku veaarvutuse skeemi valideerimine	29
4.4.	Hindamine ja veaarvutus korrapärase ruudustiku abil	38
4.4.1.	Ruudustiku loomine.....	38
4.4.2.	Pindalade hindamine ja veaarvutus ruudustiku kaudu	40
4.4.3.	Keskmise hektaritagavara hindamine ruudustiku kaudu	42
4.5.	Väikeste osakogumite hindamine ja veaarvutus.....	44
4.5.1.	Andmete ettevalmistamine väikeste osakogumite hindamiseks	44
	Näide 1.	47
	Näide 2.	49

4.5.2.	Analüüs ja kokkuvõte.....	51
5.	Mõõtmiste subjektiivsuse analüüs.....	52
5.1.	Subjektiivsustegur üksikpuu mõõtmisel.....	52
5.1.1.	Üksikpuu rinnasdiameetri mõõtmine (kluppimine).....	52
5.1.2.	Üksikpuu kõrguse mõõtmine.....	52
5.2.	Subjektiivsustegur puistu vanuse hindamisel.....	53
5.2.1.	Puistu vanuse hindamise täpsusest.....	53
5.2.2.	SMI tööühmade vahelised erinevused vanuse hindamisel.....	53
5.3.	Subjektiivsuse analüüs raieküpsuse määramisel.....	54
5.3.1.	Täiendavad vaated tööühmade taustatunnuste jaotuste kohta.....	56
5.3.2.	Logistiline regressioon raieküpsuse mõjurite uurimiseks.....	57
6.	Mõningad mudelitega seotud probleemid.....	62
6.1.	Statistiliste mudelite ahelkasutuse ohtudest.....	62
6.1.1.	Süsteematilised vead (nihked).....	62
6.1.2.	Prognoosi täpsus kahaneb.....	62
6.1.3.	Simulatsioon.....	63
6.2.	Mahumudeliga $M(A, H100)$ seotud probleeme.....	65
7.	Kaugseire kasutusvõimalusi SMI-s.....	66
7.1.	Kaugseire andmete kasutamine ajutiste traktide paigutamiseks.....	66
7.2.	Kaugseire andmete kasutamine hinnangute täpsustamiseks järelstratifikatsiooniga.....	66
8.	Täiendavad analüüsid ja arvutuseksperimendid.....	68
8.1.	Mahuhinnangute sõltuvus proovitüki keskkoha valikust.....	68
8.2.	SMI simuleerimine pindalade hindamismeetodite testimiseks.....	75
8.2.1.	Statistilise metsainventuuri simuleerimine.....	75
8.2.2.	Simuleerimisel kasutatavad tunnused.....	75
8.2.3.	Simulatsiooni tulemused ja analüüs.....	77
8.2.4.	Simulatsiooni R-kood.....	79
9.	Kasutatud kirjanduse loetelu.....	89

1. Sissejuhatus ja tulemuste kokkuvõte

Lõpparuanne käsitleb TÜ MSI projektimeeskonna poolt saadud tulemusi projekti ülesannete¹ täitmisel ajaperioodil 01.05.2022-16.12.2022. Lõpparuanne on loetav iseseisva dokumendina, sisaldades ka vahearuaande olulisi tulemusi (kohati täiendatud kujul).

Projekti raames oli pearõhk pandud SMI-s kasutatava statistilise meetodika korrektsuse uurimisele. Seejuures sai silmas peetud asjaolu, et lisaks matemaatilistele aspektidele on olemas ka mitmed praktilised aspektid, millest sõltub kogu süsteemi toimimine ja millega tuleb analüüsis arvestada.

Projektimeeskonna käsutuses olid SMI seni kogutud andmed ja ametlikud tööjuhendid (Arvutusmeetodiline juhend ning Välitööde juhend). Töö käigus konsulteeriti SMI spetsialistidega Keskkonnaagentuurist, kusjuures ühe päeva jooksul osaleti ka SMI välitöödes. Teadusliku kirjanduse kaudu tutvuti teiste riikide teadlaste tööga riikliku metsinventeerimise vallas, seda eriti statistilise külje pealt. Projektimeeskonnal oli võimalus töö käigus kohtuda ning konsulteerida ka maailmas tipptasemel olevate Rootsi ja Soome metsastatistika ekspertidega (Bo Ranney, Juha Heikkinen, Kari Korhonen, Annika Kangas jt). Nendes riikides on statistikud püsivalt kaasatud metsainventuuri metodoloogia arendusse. Üldine tähelepanek rahvusvahelisest kogemusest on see, et ei ole olemas ühte ja ainsat õiget statistilise metsainventuuri süsteemi (National Forest Inventory, lüh. NFI) – igas riigis on välja kujunenud oma süsteem, mida töö käigus arendatakse pidevalt edasi, võttes arvesse senist praktilist kogemust, oma teadusliku uurimistöö tulemusi ja uusi tehnilisi - eeskätt kaugseire - võimalusi (eri riikide NFI-de kohta vt. (Tomppo et al. 2010)).

On tähtis rõhutada, et kvaliteetsed maapealsed mõõtmised on vajalikud ka siis, kui SMI-s võetakse ulatuslikumalt kasutusele kaugseire andmed. Põhjus on selles, et kaugseire andmete kasutamiseks on esmalt vaja need kalibreerida ehk siduda tegelike maapealsete mõõtmistulemustega. Juhul kui maapealsed mõõtmised on ebatäpsed, kandub see üle ka kaugseire andmete põhjal saadud hinnangutesse.

¹ **Projekti** *Statistilise metsainventuuri (SMI) traktivõrgu analüüs ja arvutusmeetoditega seotud uuring lähteülesanded (väljavõte):*

- 1) Hinnang SMI-s kasutatavale traktivõrgule (traktide arv, trakti suurus, traktide ja proovitükkide paigutus, representatiivsus sh piirangutega/piiranguteta metsa osakaal). Töö käigus tuleb hinnata traktivõrgu suurust ning hinnata, kas traktivõrgu tihendamisel või traktide paigutuse muutmisel saaks saadavate hinnangute varieeruvust oluliselt vähendada. Hinnata, kas SMI pindalahinnangute juhusliku vea määramise meetodika on korrektne.
- 2) SMI empiiriliste mudelite (kännust rinnasdiameetri tuletamine, kõrguskõvera mudel, tagavaramudelid jne) kasutamisega kaasnevate arvutuste ning nendega seotud määramatuste arvutamise analüüs sh millised on olnud arvutusmeetodite ja R-koodi paranduste mõjud.
- 3) Välimõõtmisi teostavate gruppide subjektiivsusfaktori mõju olemasolu testimine ja suuruse hindamine.
- 4) Vajadusel näidata võimalused meetodika täpsustamiseks.

Tulemuste lühikokkuvõte

Valikudisain

1. SMI praegune andmekogumise süsteem (süsteemaatiline klastervalik) on detailides välja töötatud Rootsi spetsialistide poolt 1980-tel aastatel, mistõttu see tugineb põhjalikule eelanalüüsile ja vastab tollaegsele maailmatasemele.
2. SMI valikudisain on üldjoontes otstarbekas ja võimaldab saada statistilised hinnangud sellistele olulistele näitajatele nagu metsamaa pindala, keskmine hektaritagavara ja metsa kogutagavara.
3. Meie poolt läbi viidud arvutuste põhjal võiks kaaluda SMI valikudisaini lihtsustamist, ühtlustades alaliste ja ajutiste tagavara-proovitükkide suurused (kõik raadiusega $r=7m$, $10m$ vmt). See lihtsustaks mõnevõrra ka hindamist ja veaarvutust.

Välimõõtmised

4. Mõõtemeeskondade paratamatult väikese arvu (3) tõttu on SMI tundlik iga üksiku tööühiku subjektiivsete iseärasuste suhtes. (Võrdluseks: Rootsis ja Soomes on mõõtemeeskondade arv 20-25.) Projekti käigus tuvastati olulised erinevused mõõtemeeskondade vahel näiteks puistu vanuse hindamisel ja sellega seotuna ka raieküpsuse hindamisel.
5. Subjektiivsusest tulenevate hinnangunihete vastu aitab mõõtemeeskondade koolitus, nende töö hoolikas administreerimine, veel ulatuslikum järelkontroll ja tagasiside ning balanseeritud kasutus alaliste ja ajutiste proovitükkide lõikes.
6. Küsimusi tekitab klupitud puude märgendamine värvilaikudega, mis töötab vastu traktiasukohtade salajas hoidmise põhimõttele.

Hindamine

7. Üldjoontes SMI punkthinnangute tuletamise meetodid vastavad statistilise hindamise tuntud põhimõtetele.
8. Küsitav on mahtude prognoosimine kasvukoha proovitükkidele mahumudeli M (A, H100) abil, sest arvutused näitavad, et nii toimides üldhinnangu täpsus ei parane võrreldes juhuga, kus mahu üldhinnang tuletatakse üksnes tagavaraproovitükkide tegelike mahtude baasil. Kasvukohaproovitükkidele prognoositud mahud segavad ka veaarvutust, tekitades vea alahinnangu.
9. Mahumudeli M(A, H100) kalibreerimine kogu SMI mõõteajaloo järgi sisaldab teatud ohtu, kuna nii saadud mudel on väga inertne ja ei reageeri vajaliku kiirusega looduslike taustatunnuste (nt kliima) muutustele viimase 5-10 a jooksul. Seepärast võiks liiga vanad andmed, näiteks üle 10 a, mudeli kalibreerimisest jätta välja. Antud mudeliga prognoositud mahud muude tunnuste (näiteks maakondade) lõikes ei pruugi olla usaldusväärsed.

Määramatuse hindamine (veaarvutus)

10. Uuring kinnitas varasemat oletust, et veaarvutuse praegune meetodika on ülemäära lihtsustatud, kuna see ei arvesta SMI tegeliku andmete kogumise süsteemiga (valikudisainiga), milleks on süstemaatiline klastervalik.
11. Süstemaatilise klastervaliku korral on oluline, et statistika oleks traktipõhine: proovitükkide andmed tuleb esmalt agregeerida trakti tasemele ja edasine statistika tehakse juba traktiandmete põhjal.
12. Veaarvutuses on pakutud välja uus lahenduskeem, mis vastab süstemaatilise klastervaliku disainile ning on võimeline arvesse võtma ka ruumilisi korrelatsioone (p. 4.1.3). Meetod on kauaaegselt kasutusel Soomes ning projekti käigus on seda edukalt testitud nii tegelike SMI andmete kui ka simuleeritud andmete peal.
13. Uus veaarvutuse meetod hindab viga mõnevõrra üle, kuid teatav konservatiivsus on veahinnangu puhul parem kui liigne optimism. Samas on võimalik hinnangute täpsustamise (ehk hinnanguvea vähendamise) teel edasi minna, kaasates lisainfot kaugseire andmete näol.

SMI edasine areng

14. SMI hinnangute täpsuse tõstmine traktide arvu suurendamise teel on üks võimalus, kuid peab silmas pidama ligikaudset reeglit, et täpsuse suurendamine 2 korda vajab ca 4 korda tihedamat traktivõrku, mis seab omad piirid.
15. Tunduvalt perspektiivsem ja samas ka odavam võimalus hinnangute täpsuse tõstmiseks on kaugseire andmete (satelliidipildid ja lidari andmed) kasutuselevõtt. Antud projekti raames tehti näiteks paljulubav katse kaugseire info kasutamiseks pindalahinnangute täpsustamisel järelstratifikatsioonide teel ning saavutati hinnanguvea vähenemine 20-30% (vt. p. 8.2.3). See arengusuund nõuab täiendavaid uuringuid, mis arvestaks ka Põhjamaade kogemusi.
16. Põhjalikumat analüüsi vajab väikeste osakogumite (nt 41-50a kuusikud Harjumaal) hindamise meetodika. Esialgu soovitame siin sarnaselt koondhinnangutega kasutada üksnes tagavaraproovitükkide andmeid, mis tagab summeeruvusnõude täidetuse ja võimaldab adekvaatset veaarvutust. Samas oleks vaja uurida, kuidas paremini ära kasutada kasvukohaproovitükkides sisalduvat infot väikeste osakogumite kohta. Ka kaugseire andmete (ja muu lisainfo) abil on võimalik väikeste piirkondade hindamist oluliselt toetada. Selleks on vajalikud lisauuringud, et välja selgitada iga sihttunnuse jaoks kõige kõrgema prognoosivõimega kaugseire muutujad ning optimaalsed meetodid nende kasutamiseks.
17. Üks perspektiivne arengusuund kaugseire kasutamiseks on ka ajutiste traktide paigutuse uus meetodika, mis juurutati Rootsis 2018.a. Autorite väiteil on kaugseire andmete kasutamine traktide optimaalseks paigutamiseks efektiivsem kui nende hilisem kasutamine mahu ja teiste sihttunnuste hindamisel (lähemalt vt. p. 7.1). Leiame, et taoline valikudisain väärib lähemat analüüsi, mis selgitaks välja selle rakendatavuse Eestis ja mis arvestaks ka Soome teadlaste poolt läbi viidud võrdlevat uurimust.
18. Arvestades metsamajanduse olulist rolli Eesti majanduses tervikuna, tuleks ka edaspidi pöörata suuremat tähelepanu SMI meetodilisele läbitöötlusele. Põhjamaades on statistikute



kaasamine tagatud organisatoorselt kas ülikooli (SLU Rootsis) või ametkondliku uurimisinstituudi (LUKE Soomes) näol. Mõlemal juhul on tegemist metsandusega püsivalt tegelevate statistikateadlaste rühmaga. TÜ matemaatika ja statistika instituut omab piisavat erialast kompetentsi, et koostöös metsateadlastega lahendada ülal kirjeldatud uurimisprobleeme.

Lisamärkused

19. SMI tulemused on näiliselt täpsed, kuid tulemuste kasutaja peab alati arvestama hinnangute määramatustega (võimalike vigadega). Eeldusel, et veahinnang on adekvaatne, asub hinnatava suuruse tegelik väärtus 95% juhtudel tabelis näidatud vea piires, kuid 5% juhtudest on tegelik väärtus veapiiridest väljas.
20. On paratamatu, et harvaesinevate nähtuste nagu näiteks raiemahu hinnangute suhtelised (protsentuaalsed) mõõtmisvead on suured.
21. Detailsema pildi töö tulemustest saab, kui lugeda kokkuvõtteid aruande üksikosade lõpus.

2. Traktivõrgu analüüs

2.1. Lühiülevaade praegusest traktisüsteemist

On teada, et matemaatilises mõttes parim proovitükkide paigutus on võrdkülgsetest kolmnurkadest moodustatud korrapärane võre. Praktikas tuleb aga arvestada ka mõõtmiskuludega ja seetõttu on mõistlik koguda proovitükid klastritesse ehk traktidesse. Praegu SMI-s kasutatav trakt kujutab endast ruutu, mille igal küljel on neli proovitükki, neist pooled on tagavaraproovitükid (kus toimub puude klappimine) ning pooled kasvukohaproovitükid (klappimist ei toimu, hinnatakse visuaalselt teisi näitajaid). Ühe trakti mõõtmine on paras päevanorm ühele mõõtegrupile. Traktid ise on paigutatud geograafiliselt korrapärase, kogu riiki katva võrguna. Statistika terminites on tegemist *süsteemaatilise klastervalimiga*. Selline traktisüsteem rajati Eestis 1990-ndate aastate lõpus Rootsi samalaadse süsteemi (NFS – National Forest Survey) eeskujul, mis kätkeb endas 100-aastast kogemust ja rootsi teadlaste pikaajalist uurimistööd (vt. näit. Fridman jt. 2014, Ranney jt. 1987). Ka Soomes on kasutusel sarnane kogu riiki kattev korrapärane traktivõrgustik.

SMI traktid jagunevad alalisteks ja ajutisteks. Alaliste traktide asukohad on muutumatud, ajutiste traktide asukohad aga vahelduvad igal aastal, kusjuures mõlemat liiki traktid paiknevad korrapärase mustri järgi. SMI toimib 5-aastaste tsüklikena ning alalistest traktidest mõõdetakse igal aastal üks viiendik. Näiteks aastal 2022 mõõdeti kokku 368 trakti, millest täpselt pooled olid alalised ja pooled ajutised. Mõõtmisi teostas kolm mõõtemeeskonda.

2.2. Alaliste traktide paigutus

Alalised traktid paiknevad suhteliselt korrapärase võrena üle kogu riigi, millega välditakse proovitükkide liigseid kuhjumisi ja samas ka suuri proovitükkide poolt katmata alasid. Selline geograafiliselt ühtlane katvus (ehk süsteemaatiline klastervalik) on statistika mõttes on tunduvalt efektiivsem kui lihtne juhuslik valik (Cochran 1977). Võimalikke edasiarendusi vaagides tuleks märkida, et põhimõtteliselt oleks võimalik üldhinnangute täpsust tõsta kihtvaliku teel, mis arvestab valikudisainis looduslikku klasterdust homogeenseteks piirkondadeks, milleks on suured homogeenised metsaalad, soostunud alad, põlluharimise piirkonnad jne. Kihtvaliku korral tehakse igas homogeenises piirkonnas (kihis) statistikat eraldi ja tulemused summeeritakse (Traat jt 1997). Täiendavat analüüsi vajab siinjuures küsimus, kas kihtvalikuga kaasnev tõus hinnangute täpsuses õigustab samaaegselt suurenevat keerukust valikudisainis ja hilisemates arvutustes. (Võrdluseks: Rootsis ja Soomes on riik jaotatud alt üles 5-6 piirkonnaks, mis on looduslikult erinevad ja millest igaühes kasutatakse mõnevõrra erinevat valimidisaini ning mis kokkuvõttes tähendab kihtvaliku kasutamist).

2.2.1. Konfidentsiaalsus

Põhjamaade NFI juhtide üksmeelne arvamus, millega on kerge nõustuda, on see, et alaliste traktide koordinaadid peaksid olema salajased, mis tagaks proovitükkide tavapärase

majandamise. Kui proovitükke majandatakse muust metsamaast ja kõlvikustest erinevalt, siis põhjustaks see nihkeid SMI hinnangutes. (Rootsi eksperdid toovad näite, kus metsa majandaja ja SMI haldaja oli üks ja seesama asutus, kusjuures SMI huvi oli säilitada proovitükke, aga majandaja huvi oli raiuda. Tulemuseks oli, et ühel hetkel oli mets alaliste proovitükkide ümbert ära raiutud, aga proovitükid ise puutumata.)

Samas, SMI alaliste traktide paigutuse korrapära kasutades on nende asukohti mitte eriti raske tuvastada ning väidetavalt on seda ka juba tehtud. Sellele aitab omamoodi kaasa ka praegune SMI meetodika, mille kohaselt toimub klupitud puude märgistamine värvilaikudega. Nii on proovitükke looduses võrdlemisi kerge märgata ning see võib mõjutada metsaomaniku käitumist. Artiklis (B. Ranneby 1987) on juhitud tähelepanu samale vastuolule Rootsi NSF varasema korralduse juures.

Kui alaliste traktide asukohtade avalikuks tulek muutub reaalseks probleemiks, nii et see takistab adekvaatse metsastatistika tegemist, siis on vaja kaaluda kogu SMI süsteemi ümberkujundamist. Üheks variandiks on sel juhul üleminek ainult ajutiste proovitükkide kasutamisele. Variandi puuduseks on alaliste proovitükkide pealt tehtavate kordusmõõtmiste rea katkemine ja sellega kaasnev hinnangute efektiivsuse langus, seda eriti juurdekasvu osas. (Hinnangu efektiivsuse langus tähendab sama täpsuse saavutamist suurema proovitükkide arvuga st. suuremate kuludega.)

2.3. Ajutiste traktide paigutus

Ajutised traktid moodustavad hetkel enam-vähem samasuguse regulaarse võrgustiku nagu alalised traktid. Seejuures igal aastal toimub eelmise aasta ajutise traktivõrgustiku terviklik nihutamine konstantse nihkevektori võrra. Järelikult on ka ajutiste traktide puhul tegemist süstemaatilise klastervalikuga, mis tagab oluliselt täpsemad hinnangud võrreldes lihtsa juhuvalikuga. Põhimõtteliselt saab ajutiste proovitükkide paigutust veelgi efektiivsemaks muuta kihtvalimi kasutamise teel ehk arvestades looduslikult homogeensete piirkondade olemasolu. Selle idee elluviimisel vääriks tähelepanu ka sellised ajutiste proovitükkide paigutamise meetodid, mis kasutavad kaugseire andmeid. Üht sellist meetodit, mis suhteliselt hiljuti (2018) juurutati Rootsis, on lähemalt tutvustatud allpool (vt. p.7.1).

Kokkuvõte: Traktide jagunemine alalisteks ja ajutisteks traktideks vastab Põhjamaades väljakujunenud süsteemile ning on otstarbekas. Alaliste proovitükkide baasil saab efektiivselt arvutada metsa juurdekasvu Samuti saab sealsete mudelpuude kordusmõõtmiste käigus väärtuslikku informatsiooni, mille põhjal on võimalik koostada mitmesuguseid metsakasvu mudeleid. Jooksva mahu hindamiseks on optimaalne kasutada nii alalisi kui ajutisi proovitükke. Teiselt poolt, ajutiste traktide paralleelne kasutamine maandab riski, et alalised proovitükid oma regulaarse paiknemisega ei ole mõne olulise tunnuse mõttes representatiivsed. Samuti pole ajutised traktid tundlikud riskidele, mis on seotud alaliste traktide asukohaandmete ilmsikstulekuga ja sellele järgneva potentsiaalselt kallutatud majandamisega. Ajutiste proovitükkide piisav osakaal tagab võimaluse testida alalistel proovitükkidel tekkivaid

kõrvalekaldeid tavapärasest majandamisest. Alaliste ja ajutiste proovitükkide praegune osakaal (50%:50%) võimaldab hinnata nii metsa juurdekasvu kui ka jooksvat metsamaa pindala ja mahtu. (Märkusena lisame, et Soomes on lõppeva NFI korral see proportsioon 80:20 ja järgmises planeeritakse 60:40, kuid meist küllaltki erinev on kogu sealne NFI meetodika.)

2.4. Kluppimisraadiuse optimaalsuse analüüs

Kluppimine on töömahukas protseduur ning kluppimisraadius peab seepärast olema võimalikult väike, kuid samas piisavalt suur, et saavutada soovitud täpsusega hinnanguid. Seni on SMI kasutanud Rootsi NSF eeskujul alalistel proovitükkidel kluppimisraadiust 10m ja ajutistel 7m. Kahe raadiuse kasutamine tuleneb Rootsi teadlaste analüüsist (Ranneby jt 1987), kus näidati, et tüvemahu hindamiseks piisab raadiusest 7m, kuid juurdekasvu hindamiseks (alalised traktid) on vaja suuremat raadiust. Kuna kluppimisraadius on SMI oluline parameeter, siis uurime seda probleemi lähemalt. Järgnevas püüame olemasolevate SMI andmete põhjal välja selgitada, kuidas konkreetselt muutub hinnangute täpsus, kui tagavaraproovitüki raadiust vähendada.

2.4.1. Kluppimisraadiuste 10m ja 7m võrdlev analüüs

Selgitame esmalt välja, kui suur on täpsuse kadu, kui klupitud puude mahud arvutada mitte esialgse 10-meetrise raadiusega ringi baasil, vaid 7-meetrilise raadiusega kontsentrilise ringi baasil jättes puud kaugusega 7-10 m arvutustest välja. Analüüsis, mis põhineb aastate 2017-2021 andmetel, kasutame ainult klupitud puid (st mitte peenpuid, sest viimaste korral ei ole kaugus keskpunktist teada). Oma analüüsis loeme klupitud puudele arvutatud mahud õigeks, kuigi need on arvutatud mudelite põhjal. Igaks juhuks on mahuarvutustest välja jäetud liiga suure (kaugemal kui 10 m) kaugusega puud.

```
# andmete sisselugemine
library(dplyr)
aasta1=2017;aasta2=2021
tmp_prt=prt%>%filter(aasta>=aasta1,aasta<=aasta2,abs(pindala-314)<1,psees=="J")
tmp_prtosa=tmp_prt%>%select(aproovitykk_id)%>%inner_join(prtosa)
## Joining, by = "aproovitykk_id"
tmp_klupp=tmp_prt%>%select(aproovitykk_id)%>%inner_join(klupp)%>%filter(kaugus<=100)
## Joining, by = "aproovitykk_id"
tmp_klupp2=tmp_klupp%>%filter(kaugus<=70)
```

Järgnevalt arvutame erinevate puugruppide (näiteks elusad ja seisvad - ES jne) mahud proovitükkide kaupa ja teisendame saadud mahud võrreldavateks suurusteks (tagavara ha kohta) nii 10-meetrise kui ka 7-meetrise raadiuse korral.

```
grupid=unique(tmp_klupp$arv_grp)
```

```
taielik=tmp_prt%>%select(aproovitykk_id)%>%inner_join(tibble(arv_grp=grupid),by=character())
gruppide_kaupa1=tmp_klupp%>%group_by(aproovitykk_id,arv_grp)%>%summarize(maht=sum(arv_tagavara,
na.rm=TRUE)/(pi*10^2)*10000)%>%right_join(taielik)
## `summarise()` has grouped output by 'aproovitykk_id'. You can override using
## the `.groups` argument.
## Joining, by = c("aproovitykk_id", "arv_grp")
gruppide_kaupa2=tmp_klupp2%>%group_by(aproovitykk_id,arv_grp)%>%summarize(maht=sum(arv_tagavara
,na.rm=TRUE)/(pi*7^2)*10000)%>%right_join(taielik)
## `summarise()` has grouped output by 'aproovitykk_id'. You can override using
## the `.groups` argument.
## Joining, by = c("aproovitykk_id", "arv_grp")
erinevused=gruppide_kaupa1%>%
full_join(gruppide_kaupa2,by=c("aproovitykk_id","arv_grp"))%>%
mutate(maht.x=coalesce(maht.x,0),maht.y=coalesce(maht.y,0))%>%
mutate(vahe=maht.x-maht.y)
```

Arvutame **kirjeldavad statistikud** nii 10m kui ka 7m kluppimisraadiuse korral: keskmised, standardhälbed (keskmise standardhälve puhul tuleb arvestada, et proovitükke oli 7435), vahe standardhälbed ja korrelatsioonid.

```
erinevused%>%group_by(arv_grp)%>%summarize(nihe=mean(vahe),kesk.x=mean(maht.x),kesk.y=mean(maht.
y),std.x=sd(maht.x),std.y=sd(maht.y),stdhalve_vahe=sd(vahe),r_xy=cor(maht.x,maht.y))
## # A tibble: 10 x 8
```

arv_grp	nihe	kesk10	kesk7	std10	std7	Std_vahe	r_xy
1	0	0	0	0	0	0	NA
2 ER	0,081	2,57	2,49	27,5	28,3	8,94	0,949
3 ES	-1,32	103,89	105,21	147,76	158,00	54,3	0,939
4 KL	-0,346	0,904	1,25	4,57	6,3	3,5	0,839
5 KP	-0,233	0,285	0,519	1,82	3,3	1,78	0,919
6 SL	-0,155	5,58	5,74	18,2	22,9	12,6	0,838
7 SR	0,0216	0,184	0,163	3,15	3,42	2,61	0,688
8 SS	-0,0865	3,23	3,31	11,1	14,3	8,52	0,804
9 VR	0	0	0	0	0,00	0	NA
10 XX	0	0	0	0	0	0	NA

Tabelist nähtub, et kluppimisraadiuse vähendamisel 10m -> 7m keskmise hektarimahu hinnang oluliselt ei muutunud (kõige suuremas puude grupis ES toimus hektaritagavara suurenemine 103,89 -> 105,21-ni ehk ligikaudu 1,3%) . Samas ka hinnangu täpsus vähenes ainult veidi, ca 7%, vastavalt keskmise hektaritagavara standardhälve suurenemisele 147,76 -> 158,00.

Järgnevalt viime samasuguse analüüsi läbi **puuliikide kaupa**.

```

grupid=c("MA","KU","KS","HB","LV","LM","XX")
taielik=tmp_prt%>%select(aproovitykk_id)%>%inner_join(tibble(puuliik2=grupid),by=character())
liikide_kaupa1=tmp_klupp%>%filter(arv_grp=='ES')%>%
mutate(puuliik2=ifelse(puuliik%in%grupid,puuliik,"XX"))%>%group_by(aproovitykk_id,puuliik2)%>%summa
rize(maht=sum(arv_tagavara,na.rm=TRUE)/(pi*10^2)*10000)%>%right_join(taielik)
## `summarise()` has grouped output by 'aproovitykk_id'. You can override using
## the `.groups` argument.
## Joining, by = c("aproovitykk_id", "puuliik2")
liikide_kaupa2=tmp_klupp2%>%filter(arv_grp=='ES')%>%mutate(puuliik2=ifelse(puuliik%in%grupid,puuliik,"
XX"))%>%group_by(aproovitykk_id,puuliik2)%>%summarize(maht=sum(arv_tagavara,na.rm=TRUE)/(pi*7^2)
*10000)%>%right_join(taielik)
## `summarise()` has grouped output by 'aproovitykk_id'. You can override using
## the `.groups` argument.
## Joining, by = c("aproovitykk_id", "puuliik2")
erinevused_liigid=liikide_kaupa1%>%
full_join(liikide_kaupa2,by=c("aproovitykk_id","puuliik2"))%>%
mutate(maht.x=coalesce(maht.x,0),maht.y=coalesce(maht.y,0))%>%
mutate(vahe=maht.x-maht.y)
erinevused_liigid%>%group_by(puuliik2)%>%summarize(nihe=mean(vahe),kesk.x=mean(maht.x),kesk.y=mea
n(maht.y),std.x=sd(maht.x),std.y=sd(maht.y),stdhalve_vahe=sd(vahe),r_xy=cor(maht.x,maht.y))
## # A tibble: 7 x 8
##   puuliik2   nihe   kesk10  kesk7   std10   std7   std_vahe   r_xy
##   <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 HB      0.246   7.08   6.83   42.4   46.2   23.6   0.861
## 2 KS     -0.434  22.9   23.3   51.4   59.0   25.0   0.907
## 3 KU     -0.293  27.1   27.4   64.9   70.9   27.0   0.925
## 4 LM     -0.060   4.97   5.03   28.0   30.2   10.0   0.943
## 5 LV     -0.269   6.35   6.62   30.3   35.1   11.8   0.945
## 6 MA     -0.133  30.9   31.0   78.8   84.9   28.1   0.944
## 7 XX     -0.217   4.41   4.63   20.0   25.9   13.6   0.857

```

Tabelist näeme, et raadiuse vähendamine 10m pealt 7m peale suurendab üksikute puuliikide löikes arvatud hektaritagavara standardviga kuni 15%.

2.4.2. Mahuhinnangute standardvea graafikud (sõltuvus kluppimisraadiusest)

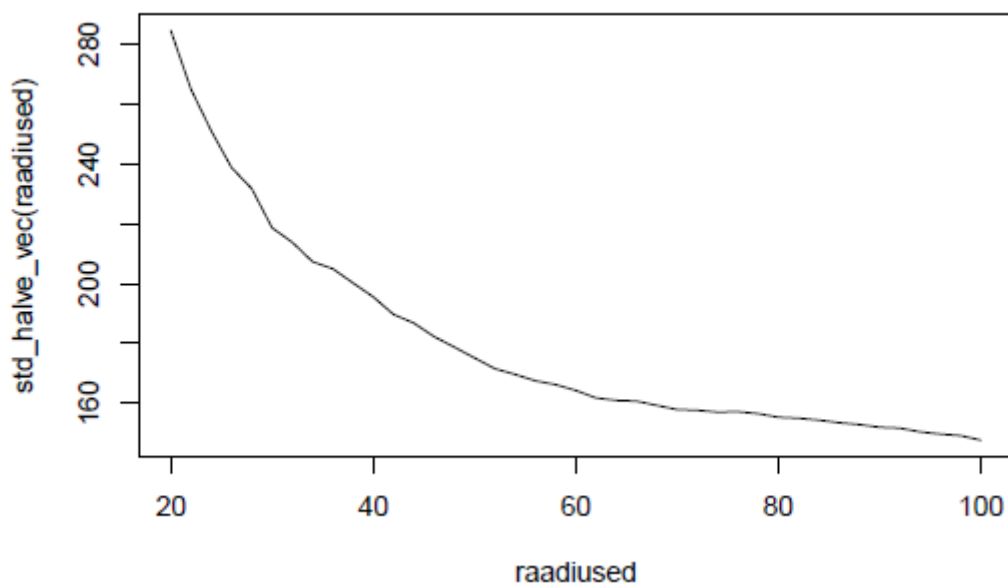
Et saada paremat ülevaadet mahuhinnangu standardhälbe sõltuvuse kohta kluppimisraadiusest, teeme ülalkirjeldatud analüüsi läbi laiemas raadiuste piirkonnas $r = 2, \dots, 10$ m ning koostame seejärel sõltuvuse kohta graafikud. Kõigepealt vaatleme suurimat puude gruppi ES (elus ja seisvad).

```

std_halve=function(r){
return(tmp_klupp%>%filter(arv_grp=='ES',kaugus<=r)%>%group_by(aproovitykk_id)%>%summarize(maht=s
um(arv_tagavara,na.rm=TRUE)/(pi*(r/10)^2)*10000)%>%right_join(select(tmp_prt,aproovitykk_id),by="apro
ovitykk_id")%>%mutate(maht=coalesce(maht,0))%>%pull(maht)%>%sd())

```

```
}  
std_halve_vec=Vectorize(std_halve)  
raadiused=seq(20,100,by=2)  
plot(raadiused,std_halve_vec(raadiused),type='l')
```

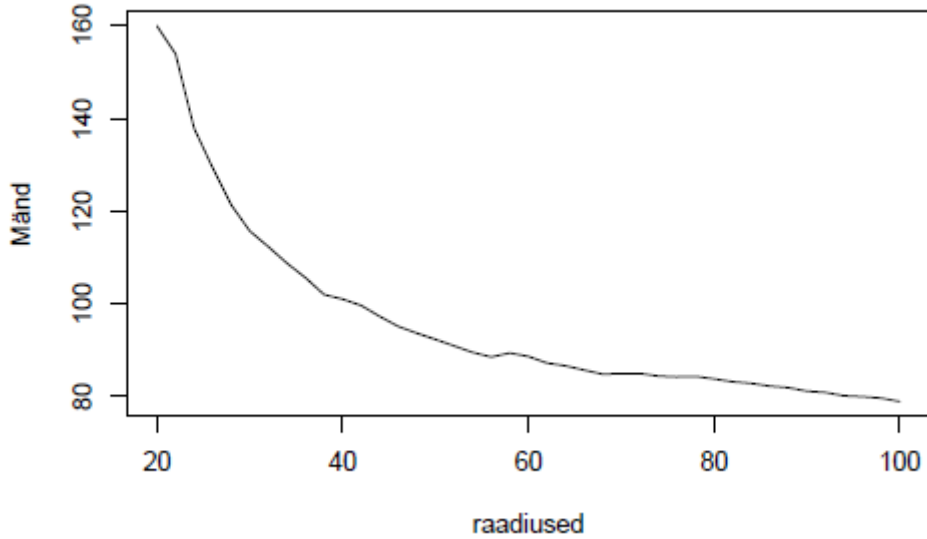


Ülaltoodud graafiku parempoolse, laugema osa põhjal võib silmaga hinnata, et piirkonnas 6-10m klappimisraadiuse vähendamisel 1 m võrra väheneb mahuhinnangu täpsus ca 2,5% . Edasine raadiuse vähendamine alla 6m toob aga kaasa juba märksa kiirema täpsuse kao. Igal juhul on ülaltoodud graafik heaks objektiivseks aluseks optimaalse klappimisraadiuse valimisel. Muidugi on vaja seejuures arvesse võtta ka mitmeid praktilise aspekte (aeg, kulud jne).

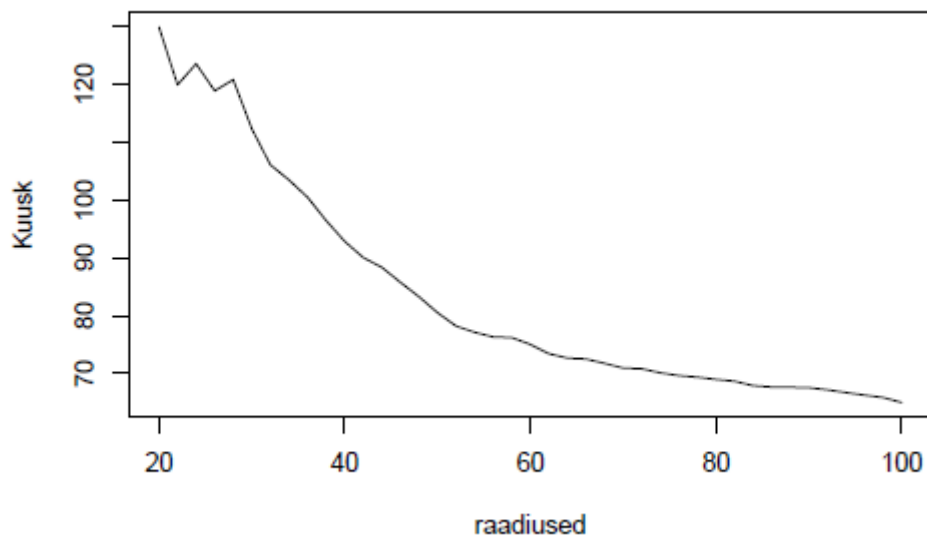
Järgnevalt esitame samasugused graafikud **puuliikide** kaupa.

```
std_halve2=function(r,puul){  
  return(tmp_klupp%>%filter(arv_grp=='ES',puuliik==puul,kaugus<=r)%>%group_by(aproovitykk_id)%>%sum  
  marize(maht=sum(arv_tagavara,na.rm=TRUE)/(pi*(r/10)^2)*10000)%>%right_join(select(tmp_prt,aproovitykk  
  _id),by="aproovitykk_id")%>%mutate(maht=coalesce(maht,0))%>%pull(maht)%>%sd()  
}  
std_halve2_vec=Vectorize(std_halve2,'r')  
raadiused=seq(20,100,by=2)
```

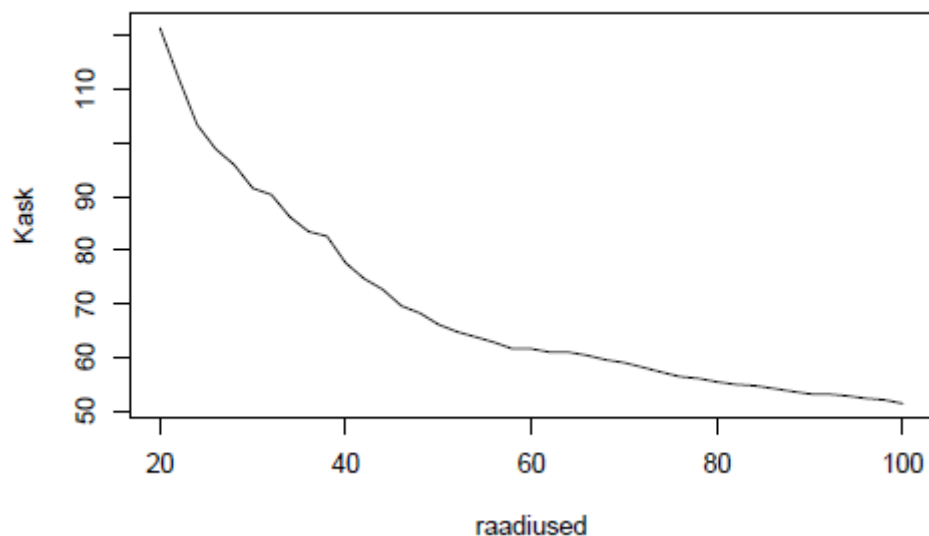
```
plot(raadiused,std_halve2_vec(raadiused,'MA'),type='l',ylab='Mänd')
```



```
plot(raadiused,std_halve2_vec(raadiused,'KU'),type='l',ylab='Kuusk')
```



```
plot(raadiused,std_halve2_vec(raadiused,'KS'),type='l',ylab='Kask')
```

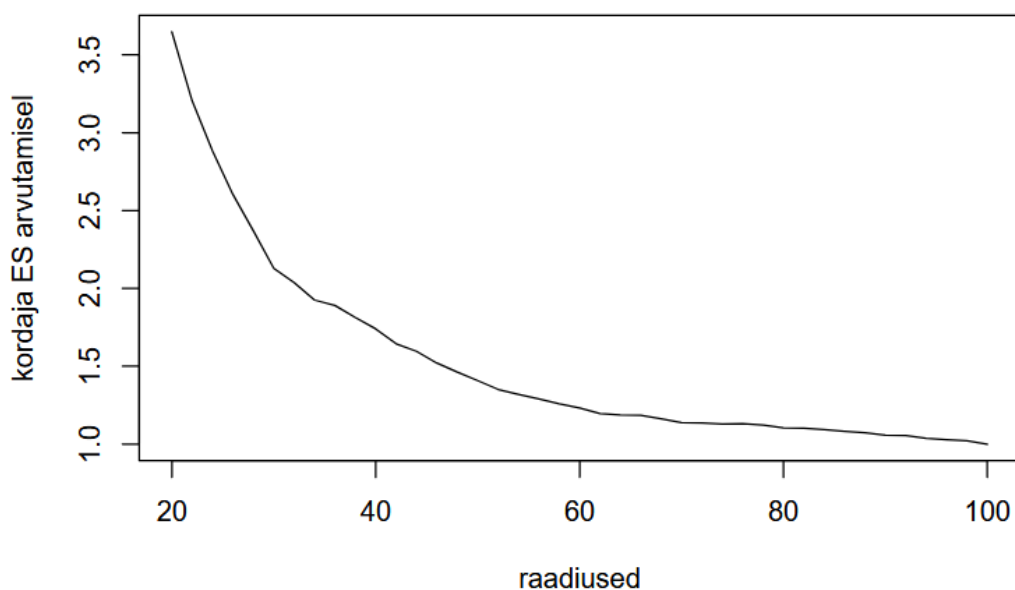


Kolm viimast graafikut näitavad, et üksikute puuliikide lõikes on tagavarahinnangute varieeruvused väiksemad (oodatavalt) võrreldes olukorraga, kus kõiki puuliike vaadeldakse koos, kuid hinnangu standardvea sõltuvus klappimisraadiusest on samasuguse kujuga.

2.4.3. Proovitükkide arvu seos klappimisraadiusega

Eelnevate kõverate baasil saab hinnata, kui palju tuleks väiksema klappimisraadiuse puhul suurendada sõltumatute proovitükkide arvu, et saada sama täpsust kui 10-meetrise klappimisraadiuse korral. See arvestus põhineb teadmisel, et sõltumatute vaatluste korral keskmise standardviga on pöördvõrdeline ruutjuurega vaatluste arvust.

Järgeval joonisel vertikaalteljel olev suurus „kordaja ES arvutamisel“ näitabki seda, mitu korda on antud raadiuse r korral vaja suurendada proovitükkide arvu, tagamaks sama täpsust kui 10 m raadiuse korral.



Jooniselt nähtub, et kui me tahaks nt 5m klappimisraadiusega saavutada sama täpsust kui 10m raadiusega, siis on meil vaja umbes 1,5 korda rohkem (sõltumatuid) proovitükke. Samas ühe proovitüki klappimise töömaht 5m korral on umbes 4 korda väiksem kui 10m korral ja 2 korda väiksem kui 7m korral. Seda arvesse võttes võib öelda, et kui näiteks ajutistel traktidel klappida kõiki (nii tagavara- kui kasvukoha-) proovitükke raadiuses 5m, siis klappimispindala (ja töömaht) jääks kokkuvõttes samaks, aga täpsus peaks olema tunduvalt suurem.

Eelnev arutlus põhineb eeldusel, et proovitükid iga üksiku trakti piires on sõltumatud vaatlused ehk teineteisest piisavalt kaugel, et sõltumatust saavutada. Kui aga need 5m proovitükid paigutada lihtsalt tihedama sammuga praeguste traktide külgedele, siis hakkavad täpsusele vastu töötama ruumilised korrelatsioonid. Tulemuseks on see, et teatud piirist alates toob proovitükkide arvu suurendamine kaasa ka hinnanguvea suurenemise. Seda efekti on teoreetiliselt kirjeldatud artiklis (Ranneby 1987) ning empiiriliselt kontrollitud SMI andmete põhjal käesoleva projekti raames.

2.4.4. Kahe erineva klappimisraadiuse kasutamine

Teatavasti SMI kasutab alalistel ja ajutistel proovitükkidel erinevaid klappimisraadiusi (vastavalt 10m ja 7m). Seejuures võib iga proovitükk olla jagatud kuni neljaks osaks vastavalt sellele, mitu erinevat kõlvikut või puistut tema sisse jääb ning igal osal on oma suhteline pindala (tunnuse nimi *pindosak*), mis kokku annavad 1. Jagamata proovitüki korral on tunnuse *pindosak* väärtus 1. Edasistes pindala ja mahuarvutuses kasutataksegi ainult tunnust *pindosak*, mistõttu informatsioon proovitüki raadiuse kohta jääb arvestamata. Tekib loomulik küsimus, kas suurema proovitüki pealt saadud hinnanguid ei peaks edasistes arvutustes võtma suurema

kaaluga? Näiteks Rootsi NFI hindamismetoodika seda teebki, tuletades kõik hinnangud eraldi 7m proovitükkidelt ja 10m proovitükkidelt ning leides seejärel kahe tulemuse kaalutud keskmise, kusjuures kaalud valitakse selliselt, et lõpptulemuse standardhälve oleks vähim (Fridman et al. 2014, Appendices A-C). Tuleb tõdeda, et arvutusvalemid muutuvad sel juhul mõnevõrra keerukateks ja tülikateks. Arvutused oleksid tunduvalt lihtsamad, kui kõigi proovitükkide raadiused oleksid võrdsed. Arvestades ka eespool toodud numbrilist analüüsi, kus selgus, et klappimisraadiuse vähendamine 10m -> 7m ei suurenda hinnangu viga eriti palju, võiks kaaluda valikudisaini lihtsustamist ühtlase klappimisraadiuse suunas.

2.4.5. Kokkuvõtte klappimisraadiuse analüüsis

1) Analüüs näitas, et klappimisraadiuse vähendamisel väheneb ka tulemuste täpsus (nagu on loogiline eeldada). Seejuures raadiuse vähendamine 10m pealt 7m peale suurendas mahuhinnangu standardhälvet üksnes mõõdukalt (tervikuna ligikaudu 7%, puuliigiti kuni 15%). Sellise täpsuse kao saab vajadusel kompenseerida proovitükkide arvu suurendamisega ~14% võrra. Analooorse arvutuse saab teha ka siis, kui on kaalumisel 7m klappimisraadiuse asendamine 5m raadiusega. Taoliste muudatuste sisseviimise otstarbekuse hindamiseks on kindlasti vaja infot ka välimõõtmiste kulukuse kohta.

2) Saadud tulemust on põhimõtteliselt võimalik kasutada ka erineva suurusega proovitükkidelt saadavate mahuhinnangute kombineerimisel ühtseks mahuhinnanguks. Näiteks, kuna ajutistel proovitükkidel on klappimisraadius 7m ja alalistel proovitükkidel 10m, siis teoreetiliselt täpseima (vähima dispersiooniga) koguhinnangu saame juhul, kui kasutame kaalutud keskmist, kus kaalud on pöördvõrdelised vastavate dispersioonidega.

3) Üldisem soovitus on kaaluda valikudisaini lihtsustamist, ühtlustades kõigi - nii alaliste kui ka ajutiste - tagavaraproovitükkide suurused. Näiteks Soomes on kompromissina kasutusel ühtne raadius 9m. Ühetaoline klappimisraadius lihtsustaks mõnevõrra ka veaarvutust, sest kaoks ära (vähemalt teoreetiline) vajadus arvestada pindala ja keskmise mahu lõpphinnangu saamisel proovitükkide erineva suurusega.

3. Ruumilised korrelatsioonid

Metsastatistikas on oluline silmas pidada ka ruumilisi korrelatsioone, mis tähendab seda, et geograafiliselt lähestikku paiknevad üksused (puud, proovitükid) omavad suhteliselt lähedasi väärtusi ka neid üksusi kirjeldavate tunnuste osas (näiteks mahtude mõttes), samas kui teineteisest kaugel paiknevad üksused erinevad teineteisest enam. Ruumilisi korrelatsioone saab väljendada korrelatsioonikordajate ja variogrammide abil. Mitmesugused ruumilisi seoseid iseloomustavad variogrammid on toodud allpool. Otstarbekas on teha vahet traktisisesel korrelatsioonil ja traktidevahelisel (globaalsel) korrelatsioonil, millest viimane tuleneb looduse suuremastaabilisest klasterdusest.

3.1. Proovitükkide vahelised korrelatsioonid

Alustuseks arvutame samasse trakti kuuluvate proovitükkide mahuhinnangute korrelatsioonikordajad.

```
tmp_cor=tmp_klupp%>%filter(arv_grp=='ES',kaugus<=100)%>%group_by(aproovitykk_id)%>%summarize(maht=sum(arv_tagavara,na.rm=TRUE)/(pi*(10)^2)*10000)%>%right_join(select(tmp_prt,aproovitykk_id),by="aproovitykk_id")%>%mutate(maht=coalesce(maht,0))
```

```
tabel_cor_leidmiseks=tmp_cor%>%inner_join(prt)%>%filter(prt_liik==1)%>%mutate(osanimi=paste(pkylg,pkaugus))
```

```
## Joining, by = "aproovitykk_id"
```

```
tabel_tmp=tabel_cor_leidmiseks%>%tidyr::pivot_wider(id_cols="atrakt_id",names_from="osanimi",values_from="maht",values_fill=0)
```

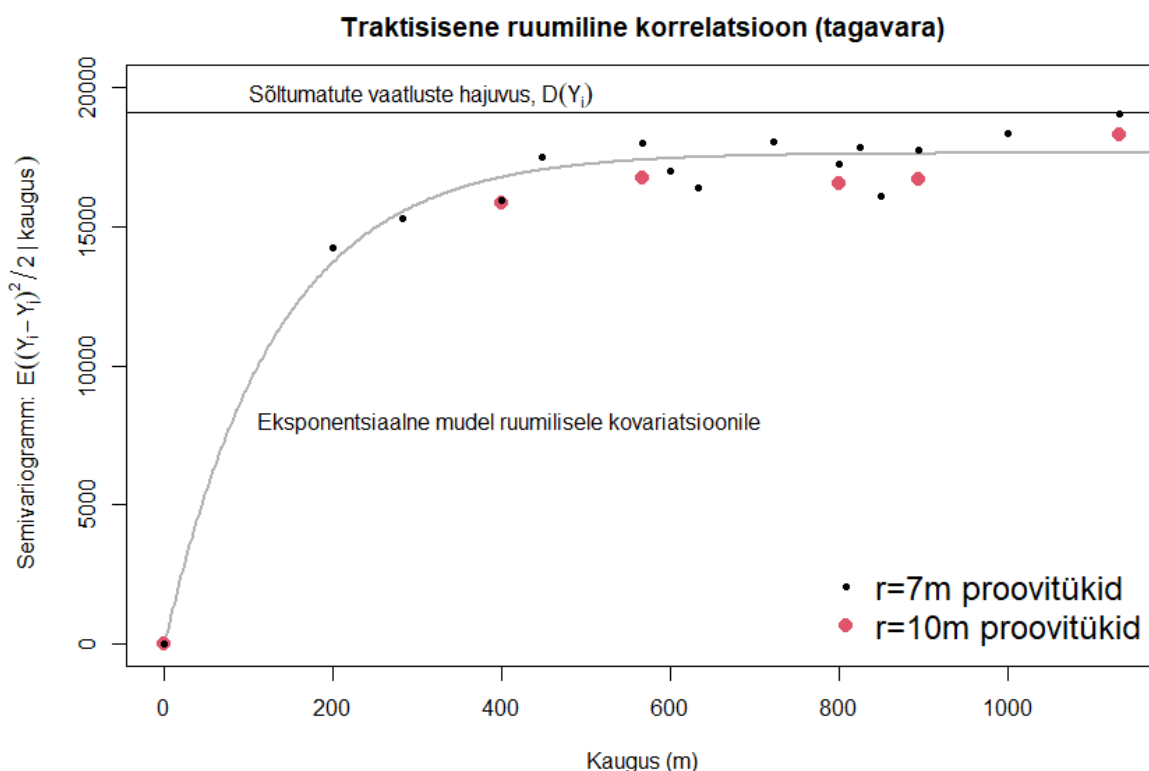
```
cor(tabel_tmp%>%select(-atrakt_id))
```

	N04	N08	E04	E08	S04	S08	W04	W08
N04	1,000	0,261	0,198	0,107	0,174	0,217	0,221	0,315
N08	0,261	1,000	0,217	0,142	0,188	0,136	0,188	0,203
E04	0,198	0,217	1,000	0,246	0,182	0,176	0,127	0,152
E08	0,107	0,142	0,246	1,000	0,241	0,192	0,182	0,118
S04	0,174	0,188	0,182	0,241	1,000	0,338	0,294	0,231
S08	0,217	0,136	0,176	0,192	0,338	1,000	0,237	0,211
W04	0,221	0,188	0,127	0,182	0,294	0,237	1,000	0,285
W08	0,315	0,203	0,152	0,118	0,231	0,211	0,285	1,000

Väga tugevaid korrelatsioone pole tabelis näha (suurim väärtus on 0,338), kusjuures iga proovitükk on enim korreleeritud oma lähima naabriga (rohelisega märgitud suured väärtused paiknevad vahetult peadiagonaali kõrval). Samuti on näha, et kauguse kasvades korrelatsioon reeglina väheneb. Toodud korrelatsioonide tõlgendamisel tuleb arvestada sellega, et need tulenevad paljuski traktikeskmiste vahelisest varieeruvusest ja vähemal määral traktisisesest proovitükkide varieeruvusest. Seda kinnitab asjaolu, et kui traktide vaheline mõju eemaldada

(seda saab teha traktikeskmise lahutamiseega proovitükkide näitudest), siis korrelatsioonid vähenevad järsult, jäädes piiridesse 0,05 ... 0,1 sõltuvalt proovitükkide omavahelisest kaugusest.

Järgnevalt esitame variogrammi, mis näitab seda, kuidas proovitükkide vahelise kauguse suurenedes kasvab tasapisi ka tagavarade erinevuse ruudu keskmine.



Graafiku esmane tõlgendus on järgmine:

- 1) teineteisest 400m kaugusel paiknevad proovitükid käituvad suuresti kui antud piirkonda iseloomustavad sõltumatud vaatlused – proovitükkide vaheliste vahemaade suurendamine märkimisväärset võitu hinnangute täpsuses ei tohiks juurde anda. Kui aga metsa takseeritakse 200m pikkuste vahemaade tagant, siis võib aimata väikest ruumilist sõltuvust. **Seega laias laastus käituvad proovitükid ühe trakti sees kui antud trakti keskmist hindavad sõltumatud mõõtmised.**
- 2) Joonisel kujutatud punktid (ja nende käitumist kirjeldava eksponentsiaalse ruumilise kovariatsiooni mudeli kõver) ei jõua päriselt välja tagavara dispersiooni iseloomustava horisontaalse sirgeni. Järelikult esinevad selged piirkondlikud eripärad (näiteks Saaremaal kasvavad metsad on selgelt teistsugused Võrumaa metsadest). Sestap ei tohiks iga proovitükki vaadelda eraldi sõltumatu vaatlusena vaid proovitükkidel tehtud mõõtmistulemused tuleks kas agregeerida trakti tasemele ja teha edasist analüüsi andmestiku abil, kus iga trakti kohta on üks rida/mõõtmine (või tuleks hinnangute

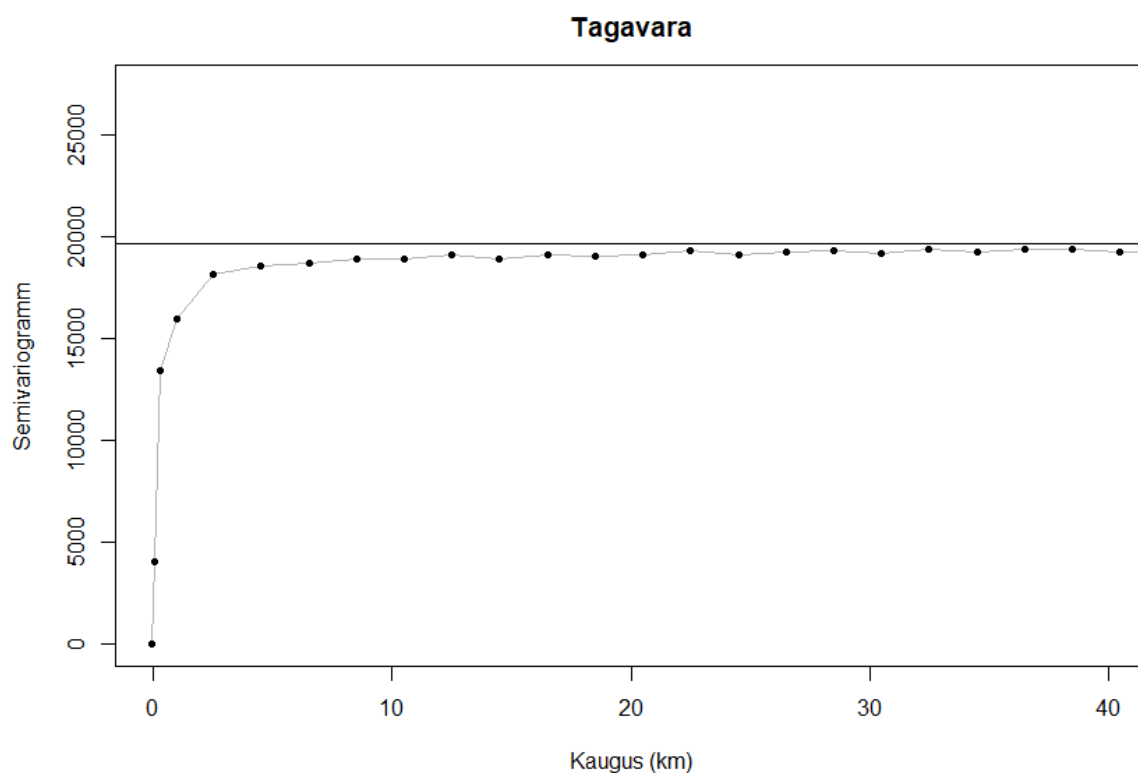
leidmiseks kasutada ruumiliste andmete/mitmetasandiliste andmete analüüsimiseks mõeldud statistilisi meetodeid).

Ülaltoodud variogrammi arvutamise **tehniline informatsioon**:

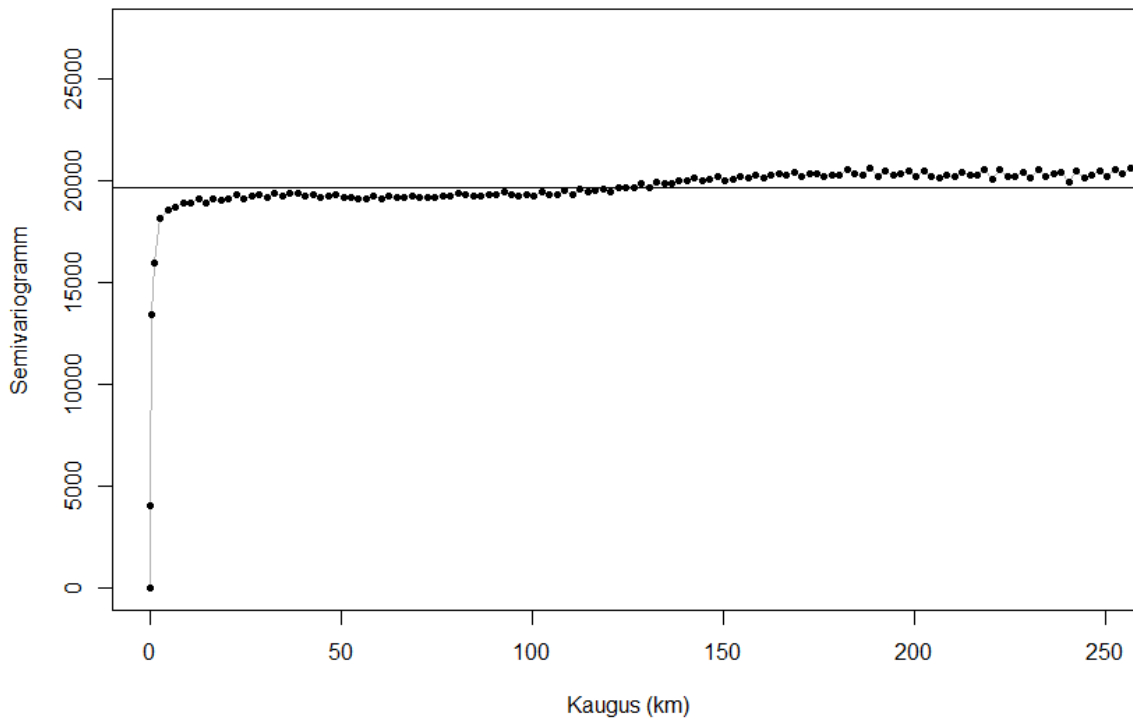
Esmalt on arvatud proovitüki tagavara, summeerides osatükkide tagavarade ja osatükkide osakaalude korrutised. Eemaldatud on proovitükid, mille puhul prtiik on 4 (ebatüüpilise pindalaga ja ka kohati muus mõttes teisiti käituvad proovitükid). Mudel on leitud vaid 7m proovitükke kasutades.

3.2. Traktidevaheline ruumiline korrelatsioon

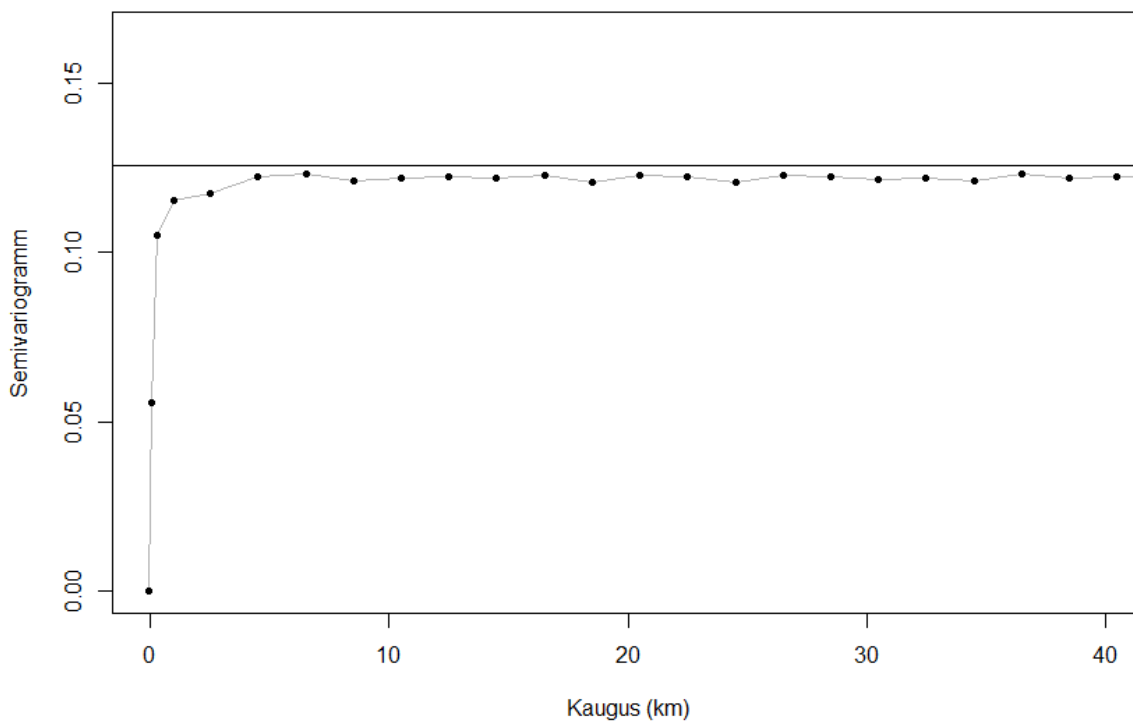
Järgnevalt vaatleme globaalseid (traktidevahelisi) ruumilisi korrelatsioone 1) tagavara osas ja 2) raieküpsuse osas. Vastavad variogrammid on arvatud esmalt keskpikkade distantside jaoks (0-40km) ning seejärel täispikkade distantside jaoks (0-250 km ja 0-400 km).

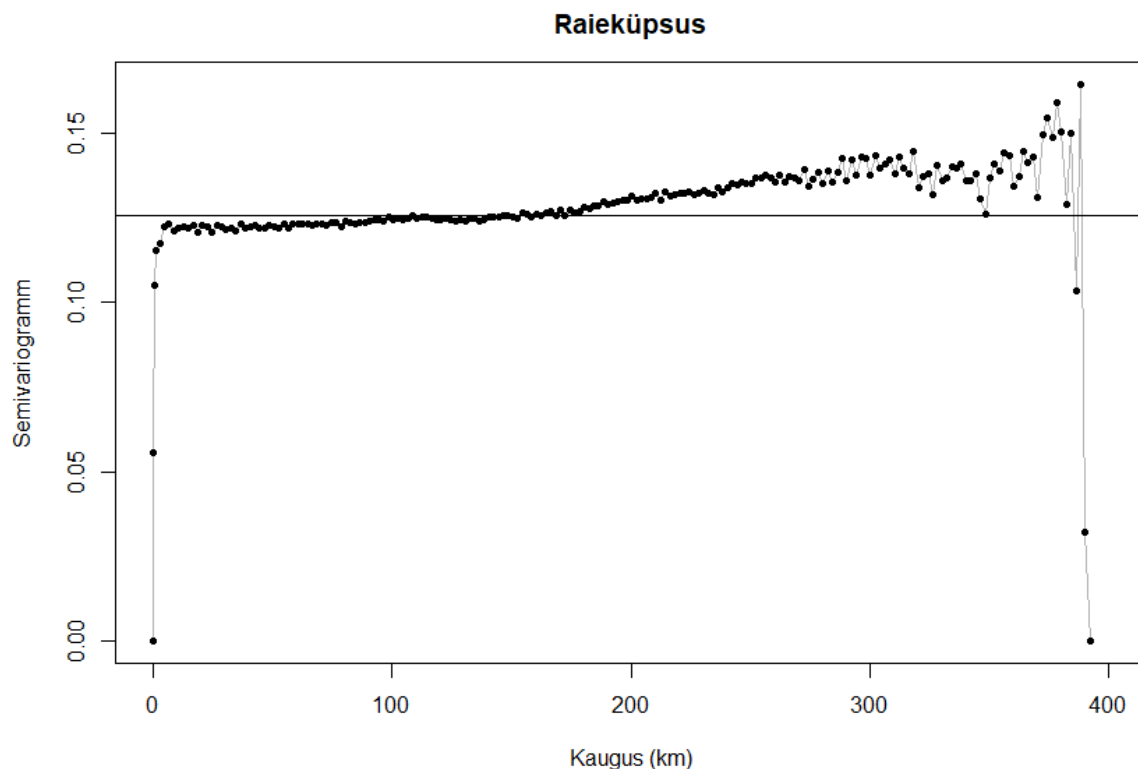


Tagavara



Raieküpsus





Selgitused nelja viimase variogrammi kohta:

Esimene punkt käib täpselt samas punktis tehtud mõõtmiste kohta. Kuna kordusmõõtmisi pole (vaatame viimase 5 aasta andmeid), siis on esimene punkti y-koordinaat alati hinnatud nulliks. Teine punkt kirjeldab proovitükke, mille vaheline kaugus on suurem nullist kuid väiksem kui 100m (0,1 km). Nii lähedaste proovitükkide olemasolu on veidi üllatav, aga mõned nii lähestikku paiknenud proovitükid siiski olid. Kolmas ja neljas punkt iseloomustavad sama trakti sisse jäävaid proovitükke (3. punkt: teineteisest 0,1...0,5 km kaugusel olevad proovitükid; 4. punkt teineteisest 0,5..1,5 km kaugusel olevad proovitükid). Edasised punktid kirjeldavad teineteisest 1,5...3,5km kaugusel olevaid proovitükke; 3,5...5,5km kaugusel olevaid proovitükke jne. Mingi muutus tundub esinevat ca 150 km kaugusel – kui proovitükid asuvad teineteisest rohkem kui 150km kaugusel, siis nad kipuvad olema teineteisest erinevamad kui kaks juhuslikult valitud proovitükki peaksid olema. See viitab teatud suuremastaabilisele looduslikule klasterdusele, mida tuleks arvesse võtta hindamisel ja/või valimi disainimisel.

Tehniline informatsioon: Kasutatud viimase 5 aasta andmeid; vaid neid trakte mille puhul psee="J" ja vaid neid proovitükke mille puhul on_sees="J". Hetkel on ühtemoodi analüüsi kaasatud nii 7m kui 10m raadiusega proovitükid.

Kokkuvõte ruumilistest korrelatsioonidest:

Ilmnes, et samasse trakti kuuluvate 16 proovitüki mahtude vahelised korrelatsioonid on mõõdukalt suured (maksimumväärtus 0,338). Need korrelatsioonid on suurel määral tingitud traktidevahelistest looduslikest erinevustest, sest traktisisesed lokaalsed ehk tinglikud

korrelatsioonid, kus trakti keskmise mõju on elimineeritud, on suhteliselt väikesed (arvestatav korrelatsioon ca 0,1 on üksnes 200 m piires). Ignoreerides neid tagasihoidlikke lokaalseid seoseid võib seega ühe trakti proovitükke lugeda laias laastus sõltumatuteks vaatlusteks selle trakti keskväärtuse kohta. Samas globaalsed ruumilised korrelatsioonid aga viitavad selgelt suurtele looduslikele klastritele ning nende arvestamine annab võimaluse veaarvutuse täpsutamiseks.

3.3. Ruumiliste korrelatsioonide integreerimine hindamisprotsessi

Ruumiliste korrelatsioonide integreerimine hindamisprotsessi toimub järgmiste põhimõtete kohaselt. Praegune SMI valikudisain (traktisüsteem) kujutab endast süstemaatilist klastervalikut, kus klastriteks on traktid. Klastervaliku korral on oluline, et vaatlusteks loetakse trakte, mitte aga trakti kuuluvaid proovitükke. Trakti kuuluvate proovitükkide mõõtmistulemused käivad antud trakti keskmise (mitte aga üldkeskmise) kohta ning nad agregeeritakse seda trakti iseloomustavateks suurusteks, näiteks metsamaa osakaal traktis, trakti keskmine hektaritagavara jne. Seejärel need traktidele arvatud hinnangud keskmistatakse ning niiviisi saadakse hinnang üldkeskmisele. Edasi on vaja lahendada kõige keerulisem ülesanne - leida üldhinnangu täpsus - ning siin on oluline arvesse võtta ka globaalseid ruumilisi korrelatsioone, mida tekitavad suured looduslikud klastrid (metsamassiivid, soolad jne). Ruumiliste korrelatsioonide arvesse võtmine võimaldab üldhinnangu täpsust suurendada võrreldes alternatiivse lihtsustatud variandiga, kus traktide tulemusi loetakse sõltumatuteks vaatlusteks üldkeskmise kohta. Kirjanduse põhjal võib öelda, et paljudes riikides korral toimubki NFI lõpphinnangute veaarvutus lihtsustatud korras ehk traktide sõltumatuse eelduse kohaselt, mis viib aga selgelt ülepaistatud veahinnanguteni (Räty et al. 2020). Üldhinnangu veaarvutuse põhimõtted on konkreetsemalt lahti kirjutatud allpool punktis 4.1.3.

4. Hindamine ja veaarvutus

Hindamine ja veaarvutus peavad olema kooskõlas valikudisainiga, milleks SMI korral on süstemaatiline klastervalik (Traat jt 1997, Cochran 1977). Teiselt poolt võib praegust SMI valikuskeemi käsitleda kui kihtvalikut, kus kihi all peame silmas ühel traktile vastavat geograafilist piirkonda (sellised piirkonnad saab määrata näiteks ruudukujulistena) ja igas kihis on tehtud üks klastervaatlus, milleks on trakt oma 16 proovitükiga. Keerulisem kui hindamine ise on tavaliselt hinnangu vea ehk määramatuse arvutamine. Alustuseks kirjeldame praegust pindalahinnangute süsteemi ning seejärel pakume välja edasiarenduse.

4.1. Pindalade ja keskmiste hindamine ja veaarvutus

4.1.1. Pindala hindamine

Hetkel leitakse pindala hinnang SMI arvutusmetoodilise juhendi kohaselt järgmiselt (tsiteering):

6.2. Pindalalised tulemused

Proovitükid ei ole kõik täielikult maismaal, seetõttu kasutatakse arvutustes proovitükkide arvu asemel maismaale jäävate osatükkide suhteliste pindalade kogusummat 100% väärtusena, kõik muud jaotamised on sõltuvalt sellest väärtusest. Näiteks, aastatel 2012–2016 mõõdeti 25 053 proovitükki, mille osatükkide suhteliste pindalade summa oli 25 050,825.

Mistahes meid huvitava objekti (nt metsamaa) pindala leitakse antud objekti osatükkide suhteliste pindalade summa jagatud kõikide osatükkide suhteliste pindalade summaga ning korrutades maakondade summaarse pindalaga:

$$P_x = \frac{SP_x}{SP_k} \cdot P = SP_x \cdot \frac{P}{SP_k}, \quad (4)$$

kus P_x on meid huvitava objekti pindala, SP_x on meid huvitava objekti suhteliste pindalade summa, SP_k on kõikide osatükkide suhteliste pindalade summa ja P on maakondade summaarne pindala.

Analüüs:

Valem (4) on meid huvitava objekti (nt metsa) kogupindala punkthinnanguna üldjoontes aktsepteeritav, kuna see tugineb tõenäosuse hindamise standardsele meetodile, mille kohaselt suhteline sagedus $\frac{SP_x}{SP_k}$ on tõenäosuse ehk osakaalu **nihketa hinnang**.

Võib diskuteerida, kas ajutiste proovitükkide (7m) ja alaliste proovitükkide (10m) pealt saadud hinnanguid saab lugeda võrdväärseteks, sest põhimõtteliselt on suurema pindalaga

proovitüki pealt saadud hinnang täpsem. Nagu juba eespool (p.2.4.4) mainitud, leitaksegi Rootsi NFI korral kaks üldpindala hinnangut - üks 7m ja teine 10m proovitükkide pealt - ning seejärel võetakse kahe hinnangu optimaalselt kaalutud keskmine (Fridman et al. 2014, Appendix B). Võrdluses Põhjamaadega võiks veel lisada, et Soome ja Norra NFI korral jälgitakse pindalade hindamisel üksnes seda, kuhu langeb proovitüki tšenter – kas metsamaale (tulemus 1) või mitte (tulemus 0) ja osaproovitükkide suhtelisi pindalaid ei võeta arvesse (Korhonen et al. 2021). Ilmselt tuleb siin kasuks nimetatud riikide suurus ja sellest johtuv palju suurem proovitükkide arv.

4.1.2. Pindala hinnangu veaarvutus

Vaatleme valemiga (4) leitava pindalahinnangu määramatuse ehk statistilise vea arvutust.

Praegu kasutatav pindala veahinnang on järgmine (tsiteering):

6.5.1. Pindalalised hinnangud

SMI hinnangute puhul vastavad pindalalised tunnused binoomjaotusele, kuna iga proovitüki puhul on küsimus, kas antud proovitükk on kindlas grupis (nt mets, põld vms) või mitte. Kuna proovitükk võib olla jaotatud, siis võib ühel proovitükil olla mitu kõlvikut. Selleks, et erinevate gruppide summad annaksid alati kokku 100%, arvutatakse vaatluste arvuks osatükkide suhteliste pindalade summa. ...

Standardsed binoomjaotuse usalduspiiride arvutamise funktsioonid erinevates tarkvarades eeldavad sisendina vaatluste arvuna täisarvulisi väärtusi. SMI puhul ei saa osatükkideks jaotamise tõttu kasutada täisarvulisi väärtusi.

Seetõttu arvutatakse tulemused ise järgnevalt:

- 1. kindlale tunnusele vastavate osatükkide suhteliste pindalade summa (nvalim) ja kõikide osatükkide suhteliste pindalade summa (nkogu);*
- 2. sündmuse esinemise tõenäosuse nihketa hinnang valemiga: $pp = (nvalim+1.92)/(nkogu+3.84)$;*
- 3. standard viga valemiga: $se = \sqrt{pp*(1-pp)/nkogu}$;*
- 4. sündmuse esinemise osakaal: $ke = nvalim/nkogu$;*
- 5. summaarne pindala: väärtus = $nvalim * 0,173518746$;*
- 6. usalduspiirid (up):*
 - a. $upmin = (pp - se * 1,96)/pp * väärtus$ ja*
 - b. $upmax = (pp + se * 1,96)/pp * väärtus$*
- 7. suhteline viga (%): $viga = (upmax-upmin)/väärtus/2*100$*

Analüüs:

Hetkel kasutatav pindalahinnangute veaarvutuse meetod põhineb binoomjaotusel s.t. eeldusel, et üksikutelt proovitükkidelt saadud hinnangud (0-1) vastavad sõltumatute sama jaotusega

juhuslike suuruste skeemile (Bernoulli skeem), kus katsete arv n on võrdne proovitükkide koguarvuga üle kõigi traktide. Jättes kõrvale pisidetailid (näiteks *pp* ja *ke* tarbetu eristamise suure valimimahu korral), oleks selline veaarvutus asjakohane juhul, kui kõik proovitükid oleksid visatud loodusesse üksteisest sõltumatult (lihtne juhuvalik). Süstemaatilise klastervalimi korral pole see meetod aga adekvaatne, kuna sama trakti proovitükkidel tehtud vaatlused (kokku 16 vaatlust) on positiivselt korreleeritud ning nad hindavad mitte üldkeskmist vaid antud trakti keskmist (lähemalt on seda analüüsitud p. 4.2., vt. ka joonis 1). Samuti on korreleeritud lähestikku asuvad traktid.

Ettepanek:

Lahenduseks pakume välja alternatiivse veaarvutuse süsteemi, mis arvestab SMI valikudisainiga ja ruumiliste korrelatsioonidega ning mida on pikaajaliselt kasutatud Soomes. Kõigepealt kirjeldame meetodit teoreetiliselt ning seejärel rakendame seda praktiliselt. Et sama meetodit saab kasutada ka mahtude hindamise juures, siis on all olevad tähistused üldisemat laadi.

4.1.3. Soovituslik veaarvutuse skeem

Nii metsamaa osakaalu kui keskmise hektaritagavara hinnang avalduvad suhtena

$$\hat{M} = \frac{\sum_i y_i}{\sum_i x_i}$$

kus **pindalade** hindamise korral:

- \hat{M} on meid huvitava objekti (nt metsamaa) osakaalu üldhinnang,
- y_i meid huvitava objekti osatükkide suhteliste pindalade summa traktis i ,
- x_i on kõikide osatükkide suhteliste pindalade summa traktis i ;

keskmise hektaritagavara hindamise korral:

- \hat{M} on keskmise hektaritagavara üldhinnang,
- y_i on trakti i kogumaht,
- x_i on trakti i metsamaa kogupindala.

1) Defineerime nüüd nn traktipõhised jäägid

$$z_i = y_i - \hat{M}x_i ,$$

kus y_i on uuritava suuruse (pindala või maht) tegelik väärtus traktil i ja $\hat{M}x_i$ võib tõlgendada sama suuruse oodatava väärtusena traktil i eeldusel, et metsamaa osakaal (või keskmine hektaritagavara) on igal pool ühesugune.

- 2) Üldhinnangu \hat{M} dispersiooni, mis vastab süstemaatilise klasterivalimi disainile, saab hinnata valemiga (Cochran 1977, pt. 6.3.)

$$V(\hat{M}) = \frac{V(\sum_i z_i)}{(\sum_i x_i)^2},$$

kus V tähistab dispersiooni.

- 3) Viimase valemi lugejat saab omakorda hinnata summaga $\sum_g T_g$, kus suurus T_g on lokaalse hajuvuse näitaja, mis arvutatakse lähinaabritest koosneva traktirühma g baasil. Näiteks korrapärase ruudukujulise traktivõrgu korral

$$T_g = \frac{(z_{i1(g)} - z_{i2(g)} - z_{i3(g)} + z_{i4(g)})^2}{4},$$

mis on arvutatud neljast naabertraktist koosneva grupi g alusel, mille paigutus looduses on ruudukujuline:

$$z_{i1(g)} \quad z_{i2(g)}$$

$$z_{i3(g)} \quad z_{i4(g)}$$

Kommentaari: Lokaalse hajuvuse mõõtu T_g tuleks siin tõlgendada kui kahe 'diagonaalhinnangu' $(z_{i1(g)} + z_{i4(g)})/2$ ja $(z_{i2(g)} + z_{i3(g)})/2$ (millest kumbki annab hinnangu sama ruudu keskpunktile) erinevuse mõõtu.

- 4) Kokkuvõttes avaldub osakaalu \hat{M} dispersioon valemiga (Heikkinen 2006, Korhonen et al 2021, sarnane valem ka Grafström ja Schelin 2014):

$$V(\hat{M}) = \frac{\sum_g T_g}{(\sum_i x_i)^2} \quad (1)$$

Soome NFI tulemustabelites raporteeritakse valimiveana (ingl.k. *sampling error*) ruutjuur dispersioonist ehk $\sqrt{V(\hat{M})}$.

Märkus 1: Kui traktisüsteem ei ole korrapärane nelinurkne võrgustik (nagu näiteks Eestis), siis lokaalse hajuvuse näitaja T_g arvutamiseks tuleb kasutada mõnevõrra erinevat skeemi, mis põhineb erinevustel ühelt poolt hinnangu z_i ning teiselt poolt tema n_i lähinaabri põhjal leitud keskmise hinnangu $\frac{1}{n_i} \sum_j z_{(i)j}$ vahel (Korhonen et al. 2021). Näiteks, kui traktid paiknevad ridadena (nagu Eestis), siis lähinaabriteks võib lugeda antud traktiga samal

traktireal paiknevat kahte naabertrakti ning sel juhul pakuvad huvi vahed $z_i - \frac{z_{i-1} + z_{i+1}}{2}$. (Viimast varianti, aga samuti klassikalist ruudukujulist varianti on meie poolt kasutatud allpool (vt. p. 4.2)).

Märkus 2: Siinkirjeldatud veaarvutuse skeemile viidates oleme allpool tihti kasutatud väljendit 'soome meetod' (kuigi meetodi teoreetiliste aluste kallal on töötanud ka Rootsi teadlased eesotsas Bertil Matérniga (Matérn 1960)).

4.2. Kogutagavara hindamine ja veaarvutus

Et kogutagavara hinnang avaldub keskmise hektaritagavara ja metsamaa pindala korrutisena, kus mõlemad tegurid on mõõdetud veaga, siis kogumahu hinnangu dispersiooni saab arvutada üksnes lähendmeetodiga, milleks on nn delta-meetod (Goodman 1960). Selle meetodi kohaselt avaldub kogutagavara dispersioon $V(T)$ järgmiselt:

$$V(T) = V(K_h \cdot P) \approx K_h^2 \cdot V(P) + P^2 \cdot V(K_h), \quad (2)$$

kus T on kogutagavara, K_h on keskmine hektaritagavara ja P on metsamaa pindala. Valem (2) eeldab küll, et hinnangud K_h ja P on statistiliselt sõltumatud, kuid selles pole antud juhul põhjust kahelda. Valem (2) on samaväärne järgmisega:

$$\frac{V(T)}{T^2} \approx \frac{V(P)}{P^2} + \frac{V(K_h)}{K_h^2}, \quad (3)$$

mille abil saab leida kogutagavara suhtelise vea, kui on teada lähtesuuruste K_h ja P suhtelised vead võrrandi paremal poolel.

4.3. Soovitusliku veaarvutuse skeemi valideerimine

Järgnevas analüüsimise eksperimentaalselt praegu kasutatavat metsamaa osakaalu veaarvutuse meetodit ja testime ka uut alternatiivset veaarvutuse meetodit. Uue veaarvutuse täiendav valideerimine, seekord simuleeritud andmete baasil, on esitatud punktis 8.2. Mõlemad arvutuslikud eksperimentid kinnitavad uue veaarvutuse meetodi paikapidavust.

1) Andmestiku moodustamine

```
# andmete sisselugemine
library(dplyr)
library(ggplot2)
aasta1=2017;aasta2=2021
tmp_prt=prt%>%filter(aasta>=aasta1,aasta<=aasta2,psees=="J")
tmp_prtosa=tmp_prt%>%dplyr::select(aproovitykk_id)%>%inner_join(prtosa)%>%filter(on_sees=="J")
## Joining, by = "aproovitykk_id"
```

```
#tmp_klupp=tmp_prt%>%dplyr::select(aprovitykk_id)%>%inner_join(klupp)%>%filter(kaugus<=100,arv_d13>=8)
tmp_trakt=trakt%>%filter(aasta>=aasta1,aasta<=aasta2)
```

2) Kõigepealt esitame **praeguse valemi abil** saadavad metsamaa osakaalu hinnangud:

```
mm_info=tmp_prtosa%>%mutate(mm_osak=if_else(maakategooria %in% c("M","MM"),pindosak,0))%>%select(pindosak,mm_osak)%>%summarize(nvalim=sum(mm_osak),nkogu=sum(pindosak))
osakaal1=mm_info$nvalim/mm_info$nkogu
pp=(mm_info$nvalim+1.92)/(mm_info$nkogu+3.84)
sd1=sqrt(pp*(1-pp)/mm_info$nkogu)
paste("Metsa osakaal:",osakaal1,"st_halbe hinnang:",sd1)
## [1] "Metsa osakaal: 0.534861097032177
st_halbe hinnang: 0.00298977581734018"
```

3) Metsa osakaalu täpsuse **alternatiivne** hindamine

Leiame kõigepealt metsamaa osakaalu igas traktis eraldi ning seejärel arvutame üldkeskmise ja selle statistilise vea. Selline lähenemine vastaks süstemaatilise klastervaliku disainile, mille korral traktisise informatsioon proovitükkidelt tuleb esmalt agregeerida trakti tasemele ning kogu edasine analüüs toimub traktide tasemel. (Siin tekivad väikesed tehnilised erinevused selles osas, et kuidas käsitleda välja jäävaid proovitükke ja nende osasid - kas need asendada nullidega või võtta hinnangute arvestamisel aluseks ainult sissejäävad osad. Vaatleme sarnaselt soomlastele juhtu, kus trakti poolt esindatava pindalaosa hinnangu saamiseks kasutame kõiki trakti osi, sisuliselt lugedes väljajäävate osade väärtuseks nulli.)

```
taiendatud_trakt=tmp_prtosa%>%mutate(mm_osak=if_else(maakategooria %in% c("M","MM"),pindosak,0))%>%
  group_by(atrakt_id)%>%summarize(mm_proc=sum(mm_osak)/16,pind_sees=sum(pindosak)/16)
osakaal2=sum(taiendatud_trakt$mm_proc)/sum(taiendatud_trakt$pind_sees)
sd2=sd(taiendatud_trakt$mm_proc)*sqrt(nrow(taiendatud_trakt))/sum(taiendatud_trakt$pind_sees)
paste("Metsa osakaal:",osakaal2,"st_halbe hinnang:",sd2)
## [1] "Metsa osakaal: 0.534861097032177
st_halbe hinnang: 0.00766688754788463"
```

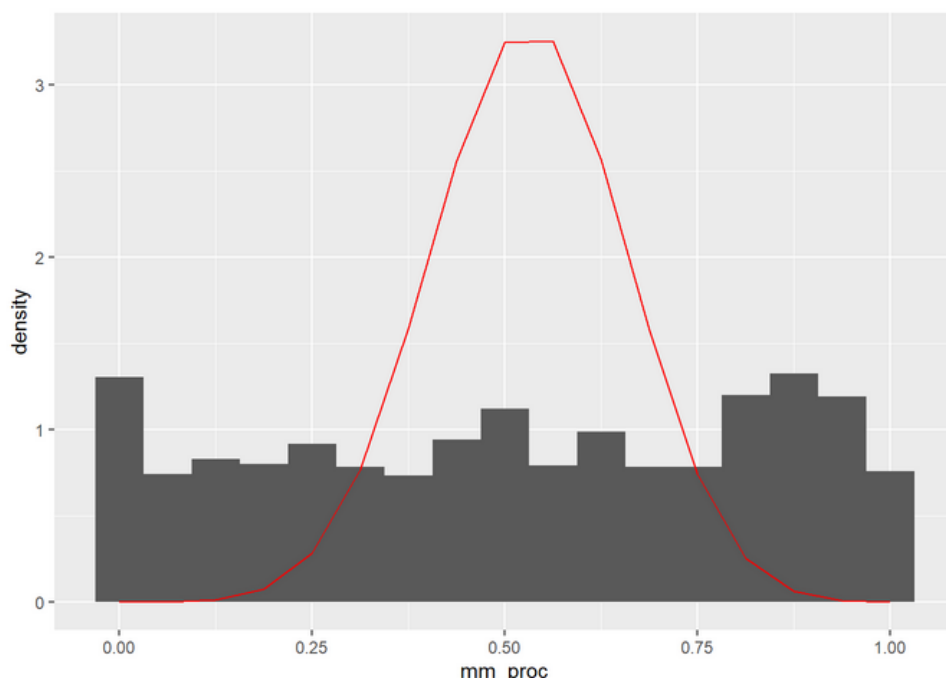
Arusaadavalt punkthinnang 0,534861 jäi samaks, kuid nagu näha, on lõpptulemuse määramatus ehk standardhälbe hinnang ca 2,5 korda suurem. Sõltumatute proovitükkide korral oleksid tulemused võrdsed (sest siis oleks iga trakti tulemus 4x täpsem, aga traktide arv on 16 korda väiksem kui proovitükkide arv, mis neutraliseerib üksikvaatluste 4x suurema täpsuse). Järelikult praeguse veaarvutuse eeldus, et kõik proovitükid on sõltumatud vaatlused, ei pea paika.

4) Traktisestest korrelatsioonide täiendav demonstratsioon

Me oleme juba näidanud, et märkimisväärsed korrelatsioonid on näiteks proovitükkide vahel sama trakti piires (p.3.1), kuid siinkohal saab seda demonstreerida veel selgemini. Nimelt, tõenäosusteooria ütleb, et kui kõik proovitükid oleksid sama jaotusega sõltumatud vaatlused (ja proovitükkide jagamist ei toimuks), siis peaks traktide kaupa summeeritud metsamaa osakaalud

käituma binoomjaotuse $B(16, p)$ kohaselt, mis alloleval joonisel on kujutatud punasega. Tegelik traktide kaupa summeeritud metsamaa osakaalude jaotus (joonisel tumedad tulbad) on aga pigem ühtlane jaotus, mis ei sarnane kuidagi binoomjaotusega. Selgituseks märgime, joonisel vasakul olev kõrgem tulp vastab põllumaale sattunud traktidele, paremal olev kõrgem ala vastab aga lausmetsa sattunud traktile, kus võib olla mõni üksik juhuslik mitte-metsa proovitükk.

```
ggplot(data=taiendatud_trakt)+geom_histogram(aes(mm_proc,after_stat(density)),bins=17)+geom_line(data=data.frame(x=seq(0,16,length.out=17)/16,y=dbinom(0:16,16,prob=0.53)*17),aes(x,y),color="red")
```



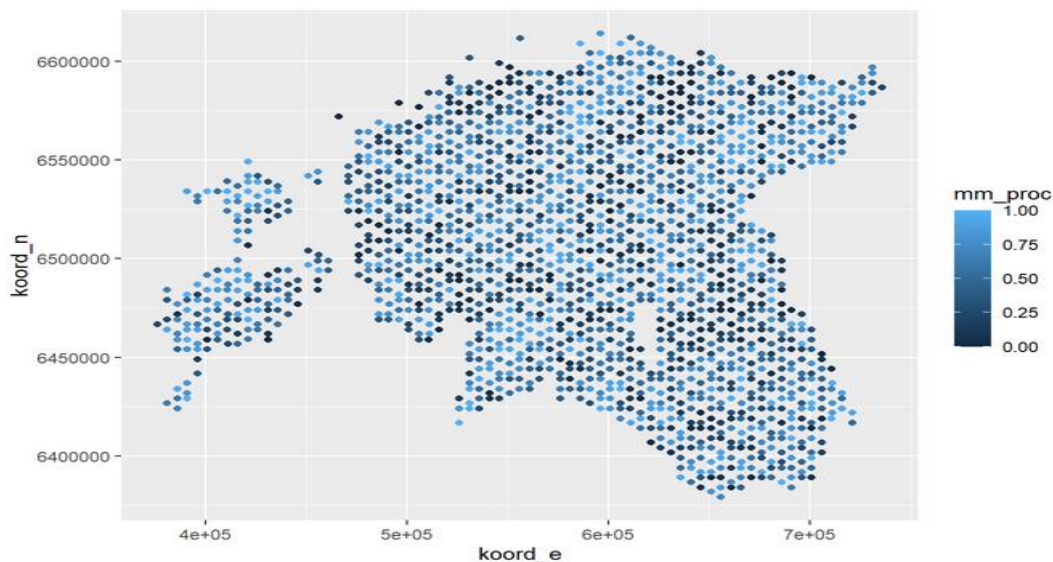
Joonis 1: Traktidele arvatud metsamaa osakaalu tegelik jaotus (tume histogramm) ja proovitükkide sõltumatuse eeldusel oodatav teoreetiline jaotus ehk binoomjaotus $B(16, p)$ (punane).

Kokkuvõttes ei saa (eriti trakti sees olevaid) proovitükke lugeda sõltumatuteks vaatlusteks. Traktisisesel vaatlusel on arvestatavad korrelatsioonid ehk 16 vaatlust mingil määral kopeerivad üksteist, kahandades niiviisi efektiivsete vaatluste arvu, mistõttu praegu kasutatav veahinnang (põhinedes kõikide proovitükkide sõltumatusel) alahindab üsna tugevasti võimalikku viga.

5) Loodusliku klasterduse väljaselgitamine metsasuse taseme järgi

Vaatleme joonisel traktide metsamaa osakaalude jaotust üle kogu territooriumi.

```
tr_koord=tmp_prt%>%filter(pkylg=="W",pkaugus=="08")%>%select(atrakt_id,koord_e,koord_n)  
ggplot(data=taiendatud_trakt%>%inner_join(tr_koord))+geom_point(aes(koord_e,koord_n,color=mm_proc))  
## Joining, by = "atrakt_id"
```



Pildi järgi võib öelda, et heledad ja tumedad punktid ei paikne läbisegi päris juhuslikult ning on märgatav teatav klasterdus ning keskvärtuse suhteliselt aeglane muutumine. **See annab võimaluse näidata metsmaa osakaalu hindamisel suuremat täpsust.** Selleks teisendame kõigepealt koordinaadid sobivateks.

6) Koordinaatide teisendamine sobivale kujule

```
theta=63.44297/180*pi
vec=c(cos(theta),sin(theta))
vec2=c(-sin(theta),cos(theta))
M=matrix(c(vec,vec2),byrow=TRUE,nrow=2)
tmp_prt_nurk=tmp_prt%>%filter(pkylg=="W",pkaugus=="08")%>%mutate(id=row_number())
katse=tmp_prt%>%pivot_wider(names_from = c(pkylg,pkaugus),values_from=c(koord_e,koord_n),id_cols =atrakt_id)%>%
  mutate(v1x=(koord_e_N_08-koord_e_W_08)/2,v1y=(koord_n_N_08-koord_n_W_08)/2,v2x=(koord_e_N_08-koord_e_E_08)/2,v2y=(koord_n_N_08-koord_n_E_08)/2)
vecs=katse %>%summarize(across(v1x:v2y,~mean(.x,na.rm=TRUE)))
v1=c(vecs$v1x[1],vecs$v1y[1])
v2=c(vecs$v2x[1],vecs$v2y[1])
teisendused=tibble(pkylg=rep(c("W","N","E","S"),each=4),pkaugus=rep(c("02","04","06","08"),times=4),
  muut_x=rep(c(v1[1],-v2[1],-v1[1],v2[1]),each=4)+rep(c(0.5,0,-0.5,-1),times=4)*rep(c(v2[1],v1[1],-v2[1],-v1[1]),
each=4),
  muut_y=rep(c(v1[2],-v2[2],-v1[2],v2[2]),each=4)+rep(c(0.5,0,-0.5,-1),times=4)*rep(c(v2[2],v1[2],-v2[2],-v1[2]),
each=4))
#katse=katse%>%mutate(koord_e=coalesce(koord_e,(koord_e_N_04+koord_e_S_04)/2),koord_n=coalesce(koord_n,(koord_n_N_04+koord_n_S_04)/2))
#katse=katse%>%mutate(koord_e=coalesce(koord_e,(koord_e_N_04+koord_e_S_04)/2),koord_n=coalesce(koord_n,(koord_n_N_04+koord_n_S_04)/2))
katse=tmp_prt%>%left_join(teisendused,by=c("pkylg","pkaugus"))%>%group_by(atrakt_id)%>%filter(row_number()==1)%>%
  ungroup()%>%mutate(koord_e=koord_e+muut_x,koord_n=koord_n+muut_y)%>%mutate(id=row_number())
```



```
#teisendatud=as.matrix(tmp_prt_nurk%>%dplyr::select(koord_e,koord_n))%*%t(M)
teisendatud=as.matrix(katse%>%dplyr::select(koord_e,koord_n))%*%t(M)
#head(teisendatud,50)
uusx=rep(NA,nrow(teisendatud))
uusy=rep(NA,nrow(teisendatud))
sammx=500
x=sort(teisendatud[,1])
ix=1
uusx[1]=1
for(i in 2:nrow(teisendatud)){
  if(x[i]-x[i-1]>sammx){ix=ix+1}
  uusx[i]=ix
}
colnames(teisendatud)=c("x","y")
Mx=as_tibble(cbind(x,uusx))
valikuks=as_tibble(teisendatud)%>%mutate(id=row_number())%>%left_join(Mx,by="x")
valikuks2=valikuks%>%group_by(uusx)%>%arrange(y)%>%mutate(uusy=row_number())%>%ungroup()
```

7) Metsasuse osakaalu veahinnangu arvutamine

Vaatleme hinnanguvea dispersiooni arvutust, mis arvestab ka ruumiliste korrelatsioonidega ehk geograafiliselt lähedaste väärtuste sarnasusega. Täpsemalt, vaatleme trakte, mis paiknevad enam-vähem ühel joonel (erinevad jooned vastavad erinevatele *uusx* väärtustele) ning vastavalt p. 4.1.3 kirjeldatud meetodikale (vt. ka sealset märkust 1.) vahede ruutusid $\left(z_i - \frac{z_{i-1}+z_{i+1}}{2}\right)^2$. Nagu valemist näha, võrreldakse trakti *i* näitu z_i temale eelneva naabri ja temale järgneva naabri näitude aritmeetilise keskmisega. (Kordame siinjuures p. 4.1.3 öeldut, et traktinäidu võrdlus sama joone peal olevate naabritega on ainult üks võimalikke viise lokaalse hajuvuse hindamiseks ning edaspidi oleme kasutanud ka teistsuguseid lähenemisi, kus naabrite määramine toimub ruumiliselt nt. ruudukujuliselt. Praktiline kogemus näitab, et lõpptulemus naabruse määramise meetodist oluliselt ei sõltu.)

```
andmestik=taiendatud_trakt%>%inner_join(katse,by="atrakt_id")%>%inner_join(valikuks2,by=c("id"))
disp_fun1=function(andmestik,osakaal){
  disp=0
  keskval=osakaal2
  xvals=sort(unique(andmestik$uusx))
  for(x in xvals){
    rida=andmestik%>%filter(uusx==UQ(x))%>%arrange(uusy)%>%mutate(mm_proc=mm_proc-pind_sees*osakaal)%>%pull(mm_proc)
    mitu=length(rida)
    if(mitu<3){
      disp=disp+sum(rida^2)
    } else {
      disp=disp+sum((rida[-c(1,2)]/2-rida[-c(1,mitu)]+rida[1:(mitu-2)]/2)^2)
      disp=disp+5/4*(rida[1])^2+1/4*(rida[2])^2+1/4*(rida[mitu-1])^2+5/4*(rida[mitu])^2
    }
  }
  return(sqrt(2*disp/3)/sum(andmestik$pind_sees))
}
disp_fun1(andmestik,osakaal2)
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
## [1] 0.006786566
```

Hinnang on veidi väiksem, kui eespool traktivaatluste sõltumatus eelduse põhjal saadud hinnang (0,00766688). See on märk ruumilistest korrelatsioonidest traktide vahel: kõrvuti paiknevad traktid kipuvad andma ka sarnasemaid metsasuse ja mahu hinnanguid. Kui andmetes ruumilisi korrelatsioone poleks, siis langenuks need kaks dispersiooni hinnangut ligilähedaselt kokku.

Illustreerimaks ruumiliste korrelatsioonide toimimist, leiame järjestikuste traktiväärtuste (mööda tüüpilisi paiknemisjooni) vahelise korrelatsiooni.

```
andmestik2=andmestik%>%select(mm_proc,uusx,uusy)
andmestik2.1=andmestik2%>%filter(uusy%%2==1)
andmestik2.2=andmestik2%>%filter(uusy%%2==0)
kokku=andmestik2.1%>%inner_join(andmestik2.2,by=c("uusx"))%>%filter(uusy.y==uusy.x+1)
cor(kokku$mm_proc.x,kokku$mm_proc.y)
```

```
## [1] 0.110903
```

See selgitab, miks regulaarse võrgu kasutamine aitab mõnevõrra täpsemat hinnangut saada - lokaalsed vead on keskmiselt veidi väiksemad kui trakti osakaalu erinevused üldkeskmisest.

8) Soovitusliku veahinnangu eksperimentaalne valideerimine

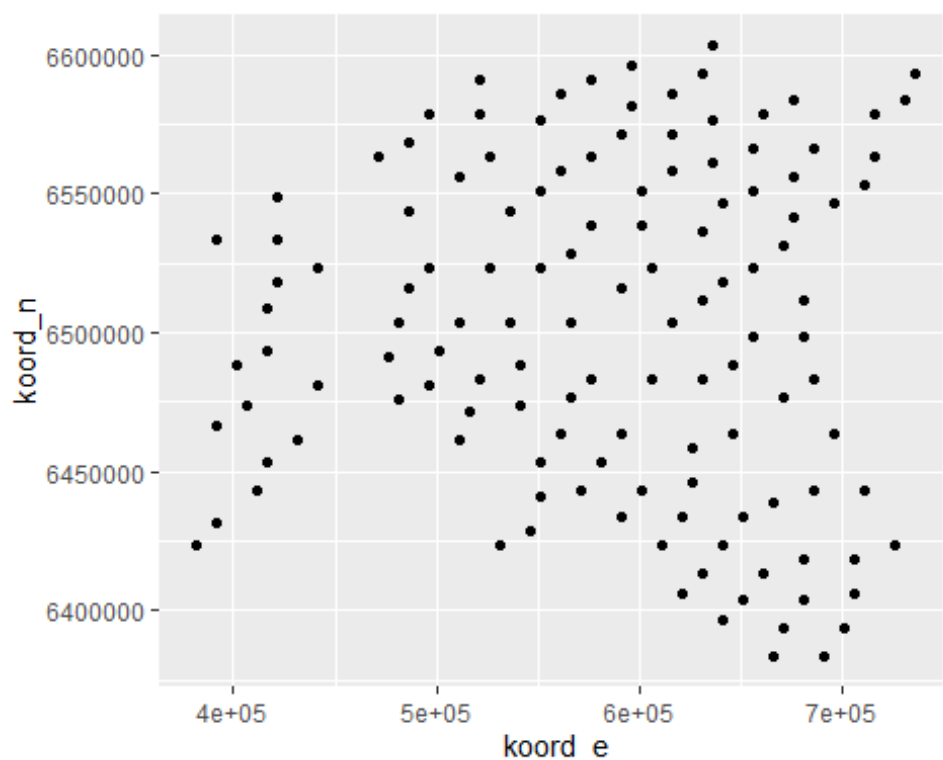
Vaatleme nüüd seda, kuidas uus standardhälbe hinnang on kooskõlas regulaarse võrgu erinevate nihete korral saadavate üldkeskmise hinnangute muutustega (tegelike vigadega). Selleks võrdleme omavahel 16 'hõrendatud' võrgu põhjal saadud hinnanguid.

Loome valimi moodustamise funktsiooni

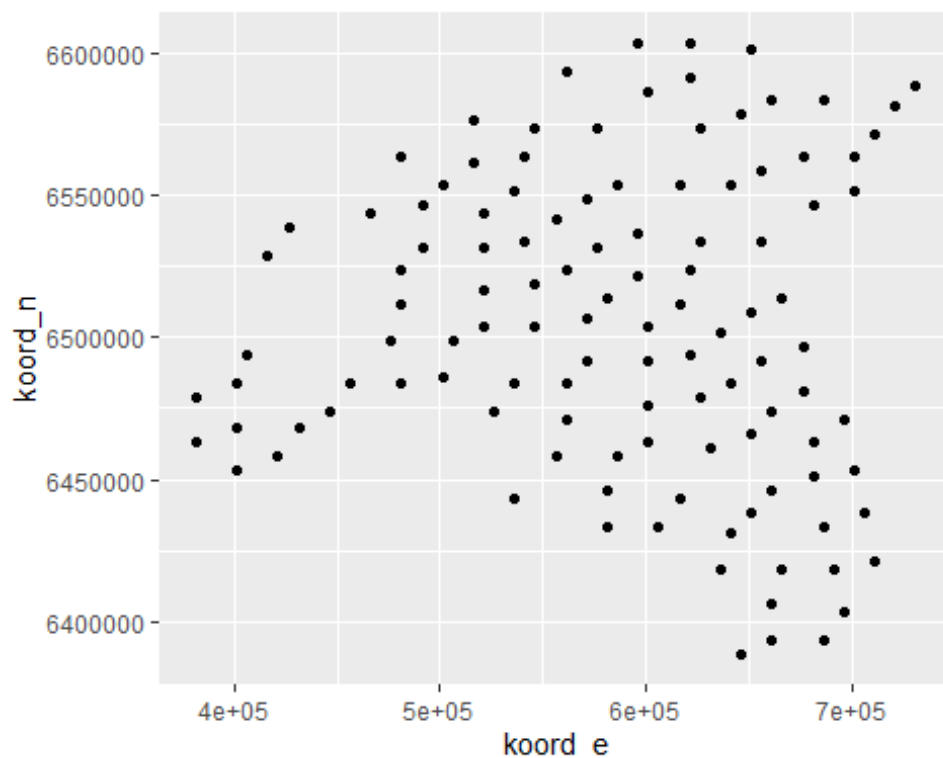
```
tee_valim=function(xstart,ystart,byval,valikuks=valikuks2){
  #byval - millise sammuga x ja y suunas võetakse
  #xstart ja ystart võivad omada väärtused 1,2,...,byval
  valim=data.frame(x=0,y=0,id=0)
  for(i in seq(xstart,max(valikuks$uusx),by=byval)){
    maxy=max(valikuks$uusy[valikuks$uusx==i])
    if(maxy>=ystart){
      for(j in seq(ystart,maxy,by=byval)){
        rida=valikuks%>%filter(uusx==i,uusy==j)%>%dplyr::select(x,y,id)
        valim=rbind(valim,rida)
      }
    }
  }
  valim=valim[-1,]
  return(valim)
}
```

Teostame arvutused, jaotades traktid 16-sse eri rühma. Jagamise aluseks on tähelepanek, et praegused traktid paiknevad paralleelsetel joonel. Valime iga neljanda joone ja sellel joonel iga neljanda trakti. Nii saame kokku 16 traktirühma ehk 'hõrendatud' võrku. Rühmad säilitavad enam-vähem regulaarsuse ning ükski trakt ei ole kahes erinevas rühmas. Näitena toome kahe selliselt 'hõrendatud' traktivõrgu paigutuse.

```
proov=tee_valim(1,1,4,andmestik)  
tmp_andmestik=andmestik%>%inner_join(proov,by=c("id"))  
ggplot(tmp_andmestik)+geom_point(aes(x=koord_e,y=koord_n))
```



```
proov=tee_valim(3,2,4,andmestik)  
tmp_andmestik=andmestik%>%inner_join(proov,by=c("id"))  
ggplot(tmp_andmestik)+geom_point(aes(x=koord_e,y=koord_n))
```



Arvutame nüüd iga 'hõrendatud' traktivõrgu korral (kokku 16 tarktivõrku) metsasuse üldkeskmise hinnangud koos vastavate standardhälbe hinnangutega, kasutades nii praegust SMI meetodit (st_orig) kui ka uut, alternatiivset meetodit.

```
samm=4
tulem=data.frame(valimi_nr=1:samm^2,i=NA,j=NA,kv=NA,sd_orig=NA,sd_uus=NA)
rida=0
for(i in 1:samm){for(j in 1:samm){
  rida=rida+1
  tulem$i[rida]=i
  tulem$j[rida]=j
  proov=tee_valim(i,j,samm,andmestik)
  tmp_andmestik=andmestik%>%inner_join(proov,by=c("id"))
  nvalim=sum(tmp_andmestik$mm_proc)*16
  nkogu=sum(tmp_andmestik$pind_sees)*16
  tulem$kv[rida]=nvalim/nkogu
  pp=(nvalim+1.92)/(nkogu+3.84)
  tulem$sd_orig[rida]=sqrt(pp*(1-pp)/nkogu)
  tulem$sd_uus[rida]=disp_fun1(tmp_andmestik,tulem$kv[rida])
}}
tulem
```

```
## valimi_nr i j kv sd_orig sd_uus
## 1 1 1 1 0.5250277 0.01143654 0.03028479
## 2 2 1 2 0.5398959 0.01157223 0.02684802
## 3 3 1 3 0.4765908 0.01202022 0.02635667
## 4 4 1 4 0.5800508 0.01215664 0.02921869
## 5 5 2 1 0.5683906 0.01168456 0.02712849
## 6 6 2 2 0.5538407 0.01187292 0.03038120
## 7 7 2 3 0.5404337 0.01235165 0.02811834
## 8 8 2 4 0.5493926 0.01251764 0.02954942
## 9 9 3 1 0.5163988 0.01174297 0.02752713
## 10 10 3 2 0.5643665 0.01172008 0.03041527
## 11 11 3 3 0.5462284 0.01213035 0.03007073
## 12 12 3 4 0.5253427 0.01249182 0.02600279
## 13 13 4 1 0.5299502 0.01153867 0.03057586
## 14 14 4 2 0.5026509 0.01165298 0.02813382
## 15 15 4 3 0.5004051 0.01202117 0.02484859
## 16 16 4 4 0.5440049 0.01240529 0.03013705
```

Erinevate võrgustike baasil saadud metsasuse hinnangud (16 vaatlust tabeli veerus kv) kõiguvad küllalt suurtes piirides: 0,47 kuni 0,58. Leiame hinnangute standardhälve:

```
sd(tulems$kv)
```

```
## [1] 0.02713944
```

Võrdluseks toome igas 'hõrendatud' võrgus eraldi hinnatud standardhälvete keskmise (tabeli veeru sd_uus keskmise):

```
mean(tulems$sd_uus)
```

```
## [1] 0.0284748
```

Nagu näha, on 16 võrgustiku pealt saadud üksikhinnangute tegelik varieeruvus 0,0284748 väga heas kooskõlas meie poolt pakutava alternatiivse veahindamise meetodiga 0,02713944. Samas ei haaku tegelik hinnangute varieeruvus kuigivõrd praegu kasutatava veahindamise valemiga.

Kokkuvõte:

1) Ülaltoodud analüüs on veenvalt näidanud, et kõigi proovitükkide hulka ei saa vaadelda ühe ja sama suuruse (metsamaa osakaalu) sõltumatute vaatlustena; proovitükkide ligikaudsest sõltumatusest saab rääkida üksnes traktide sees (tinglik sõltumatus), kuid rangemalt võttes tuleb ka siin arvestada mõningase ruumilise korrelatsiooniga. Sellest tulenevalt pole praegu kasutatav pindala veaarvutuse meetod korrektne ja vajab täpsustamist.

2) On välja pakutud alternatiivne pindala ja keskmise hektaritagavara veaarvutuse süsteem, mis tugineb traktipõhisel statistikal, arvestab ruumilisi korrelatsioone ja on kooskõlas klastervaliku disainiga. Teostatud arvutuseksperiment kinnitab uue traktipõhise veahindamise meetodi paikapidavust.

4.4. Hindamine ja veaarvutus korrapärase ruudustiku abil

Järgnevas pakume välja ühe täienduse eelpool kirjeldatud hindamise ja veaarvutuse skeemidele. Eesmärk on arvesse võtta asjaolu, et praegune traktivõrk oma vanade ja uute alaliste traktide ning ajutiste traktidega pole enam päris korrapärane võrgustik: traktid kohati rühmituvad ning samas on hõredamalt kaetud tsoone. Kuna üksteisele lähedal paiknevad traktid hindavad ligikaudu sama looduslikku keskkonda, siis oleks mõttekas need hinnangud esmalt agregeerida, kusjuures agregeerimise aluseks pakume välja korrapärase 5x5km ruudustiku. Sel juhul oleksid kõik geograafiliselt sama suured alad lõpphinnangus võrdselt esindatud.

Täpsemalt on idee selles, et jagada võrdlemisi ebaregulaarseks kujunenud traktivõrgustik ära korrapärase 5x5km ruutude vahel, misjärel igasse ruutu sattunud traktide näidud keskmistatakse (üldiselt tuleb igasse ruutu üks trakt) ning seejärel kasutatakse edasistes arvutustes juba ruutude näite. Ruudustiku sissetoomise otseseks ajendiks on Soomes kasutatav veaarvutuse skeem (p. 4.1.3), mis oma klassikalisel kujul eeldab traktide korrapärase ruudukujulist paigutust. Allolev kood sisaldab nii ruudustiku loomist kui ka ruudustiku põhjal hindamist.

4.4.1. Ruudustiku loomine

```
# andmete sisselugemine
library(tidyr)
library(dplyr)
library(ggplot2)
aasta1=2017;aasta2=2021
tmp_prt=prt%>%dplyr::filter(aasta>=aasta1,aasta<=aasta2,psees=="J")
tmp_prtosa=tmp_prt%>%dplyr::select(aproovitykk_id)%>%inner_join(prtosa)%>%filter(on_sees=="J")
## Joining, by = "aproovitykk_id"
#tmp_klupp=tmp_prt%>%dplyr::select(aproovitykk_id)%>%inner_join(klupp)%>%filter(kaugus<=100,arv_d13>=8)
tmp_trakt=trakt%>%filter(aasta>=aasta1,aasta<=aasta2)
```

Ruudukeste defineerimisel võib kasutada sobivalt pööratud koordinaadistikku. Arvutame kõigepealt traktikeskmiste koordinaadid ja siis need koordinaadid pööratud teljestikus.

```
theta=-0.27#63.44297/180*pi
vec=c(cos(theta),sin(theta))
vec2=c(-sin(theta),cos(theta))
M=matrix(c(vec,vec2),byrow=TRUE,nrow=2)
```

Trakti keskpunkti leidmiseks leiame esimese olemasolevate koordinaatidega seesoleva proovitüki ja liidame sellele vektori, mis peaks keskele nihutama.

```
#leiame andmetelt, millised on põhjaküljele ja idaküljele vastavad vektorid keskmiselt (ei ole päris vertikaalsed ja horisontaalsed)
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
#sisuliselt saaks väga sarnased keskpunkti koorinaadid, kui võtta nendeks vektoriteks (800,0) ja (0,800)
abi_nihe=tmp_prt%>%pivot_wider(names_from = c(pkylg,pkaugus),values_from=c(koord_e,koord_n),id_cols=
=atrakt_id)%>%

  mutate(v1x=(koord_e_N_08-koord_e_W_08)/2,v1y=(koord_n_N_08-koord_n_W_08)/2,v2x=(koord_e_N_08-
koord_e_E_08)/2,v2y=(koord_n_N_08-koord_n_E_08)/2)
vecs=abi_nihe %>%summarize(across(v1x:v2y,~mean(.x,na.rm=TRUE)))
v1=c(vecs$v1x[1],vecs$v1y[1])
v2=c(vecs$v2x[1],vecs$v2y[1])

#teisenduste tabel (mida peab liitma, et keskele saada)
teisendused=tibble(pkylg=rep(c("W","N","E","S"),each=4),pkaugus=rep(c("02","04","06","08"),times=4),
  muut_x=rep(c(v1[1],-v2[1],-v1[1],v2[1]),each=4)+rep(c(0.5,0,-0.5,-1),times=4)*rep(c(v2[1],v1[1],-
v2[1],-v1[1]),each=4),
  muut_y=rep(c(v1[2],-v2[2],-v1[2],v2[2]),each=4)+rep(c(0.5,0,-0.5,-1),times=4)*rep(c(v2[2],v1[2],-
v2[2],-v1[2]),each=4))

#lisame keskpunkti koorinaadid
trakt_keskpunktiga=tmp_prt%>%left_join(teisendused,by=c("pkylg","pkaugus"))%>%group_by(atrakt_id)%>%
filter(row_number()==1)%>%

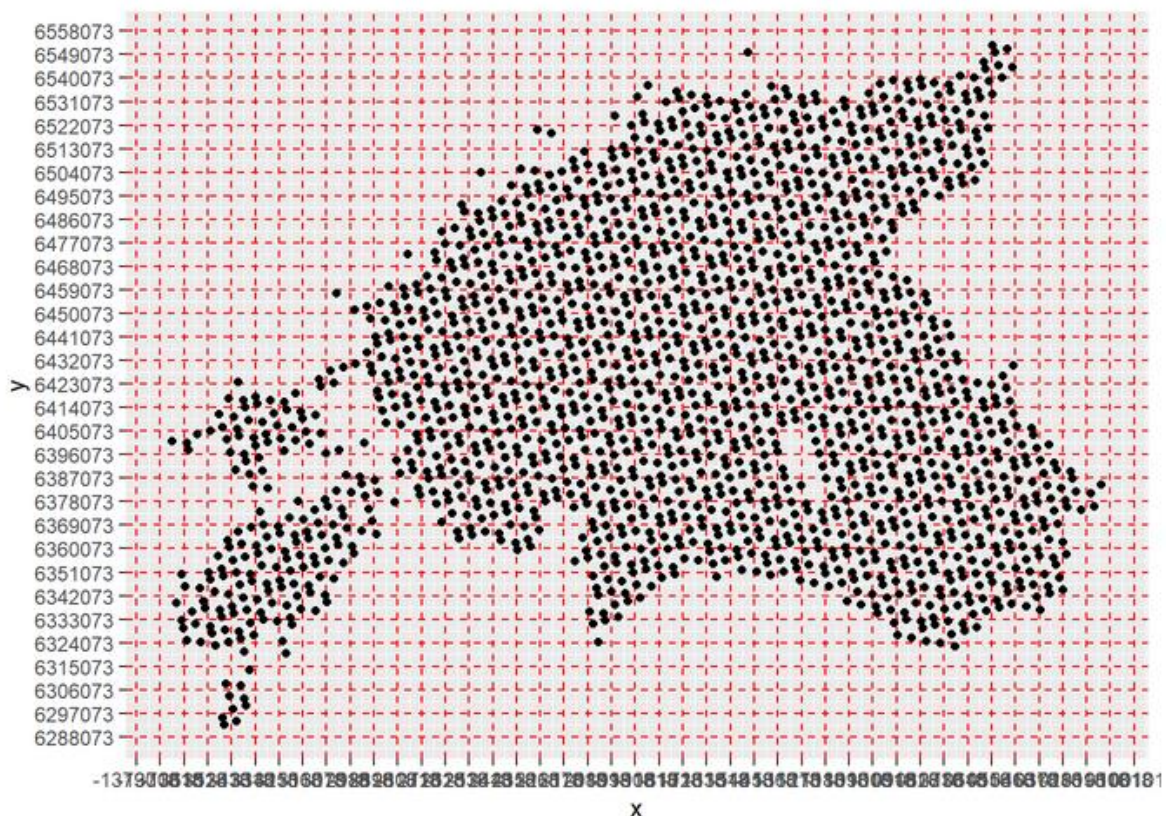
ungroup()%>%mutate(koord_e=koord_e+muut_x,koord_n=koord_n+muut_y)%>%mutate(id=row_number())%
>%select(atrakt_id,koord_e,koord_n,id)

#teisendatud=as.matrix(tmp_prt_nurk%>%dplyr::select(koord_e,koord_n))%*%t(M)
teisendatud=as.matrix(trakt_keskpunktiga%>%dplyr::select(koord_e,koord_n))%*%t(M)
colnames(teisendatud)=c("x","y")
trakt_teisendatud=trakt_keskpunktiga%>%dplyr::bind_cols(as_tibble(teisendatud))
```

Lisame nüüd traktile regulaarse võrgu info.

```
xlimits=range(trakt_teisendatud$x)
ylimits=range(trakt_teisendatud$y)
kyljepikkus=9000 #(meetrites)
xp=seq(xlimits[1]-2.5*kyljepikkus,xlimits[2]+2.5*kyljepikkus,by=kyljepikkus)
yp=seq(ylimits[1]-2.5*kyljepikkus,ylimits[2]+2.5*kyljepikkus,by=kyljepikkus)
i_x=as.numeric(cut(trakt_teisendatud$x,breaks=xp))
i_y=as.numeric(cut(trakt_teisendatud$y,breaks=yp))
trakt_piirkond=mutate(trakt_teisendatud,i=i_x,j=i_y)%>%select(atrakt_id,i,j)
koik_indeksid=tibble(i=rep(1:max(i_x),each=max(i_y)),j=rep(1:max(i_y),times=max(i_x)))
```

```
ggplot(trakt_teisendatud)+geom_point(aes(x,y))+scale_x_continuous(breaks=xp)+scale_y_continuous(breaks=y  
p)+ theme(panel.grid.major = element_line(color = "red",linetype = 2))
```



NB! Joonisel nähtav punane ruudustik vastab ruudu küljele 10 km.

4.4.2. Pindalade hindamine ja veaarvutus ruudustiku kaudu

Erinevate hinnangute arvutamiseks tuleks nüüd vajalikud suurused agregeerida kõigepealt trakti peale, seejärel leida hinnangud äsja moodustatud ruutude jaoks (keskmistades ruutu sattunud traktide näidud) ning siis saab leida ka veahinnanguid. Veaarvutuses on allpool välja pakutud kaks varianti (millest teine on Soomes kasutatav süsteem), mis annavad ligilähedaselt samad tulemused. Meetodid erinevad selle poolest, kuidas ühe trakti (ruudu) peal tekkivat hinnanguviga mõõta naaberruutude hinnangute abil.

```
taiendatud_trakt=tmp_prtosak>% mutate(mm_osak=if_else(maakategooria %in%  
c("M","MM"),pindosak,0))>%
```

```
group_by(atrakt_id)%>%summarize(mm_proc=sum(mm_osak)/16,pind_sees=sum(pindosak)/16)
```

```
piirkond=taiendatud_trakt>%inner_join(trakt_piirkond,by="atrakt_id")>%group_by(i,j)%>%summarize_at(  
vars(mm_proc,pind_sees),.funs=~mean(.x))>%ungroup()
```


Ruutudele väljaspool territooriumi omistame väärtused 0.

#laiendame nullidega

```
piirkond_laiendatud=koik_indeksid%>% left_join(piirkond,by=c("i","j"))%>% mutate(tinglik_kv=coalesce(mm_
proc,0),pind_sees=coalesce(pind_sees,0))%>% select(i,j,tinglik_kv,pind_sees)
```

Metsasuse osakaalu hinnang:

```
osakaal1=sum(piirkond_laiendatud$tinglik_kv)/sum(piirkond_laiendatud$pind_sees)
osakaal1
## [1] 0.5341887
```

Võrreldes eelneva, ilma ruudustikuta leitud hinnanguga (0.5348611) on tulemus pisut erinev, sest juhul, kui ühes ruudus on mitu trakti, esinevad need üldkeskmises väiksema kaaluga kui need traktid, mis on piirkonnas üksinda.

Leiame nüüd veahinnangu, kasutades korrapärast ruudustikku lokaalse hajuvuse hindamiseks kahel erineval viisil:

Meetod 1: Võrdleme joonisel olevat keskmist ruutu (hinnang h) ja tema nelja naaberruudu hinnangute keskmist ning arvutame 'lokaalsed vead' $h - (h_1+h_2+h_3+h_4)/4$ kõikide ruutude jaoks. Selliste lokaalsete vigade baasil arvutataksegi üldhinnangu viga. Vastava funktsiooni

	h3	
h1	h	h2
	h4	

arvuta_sd2 rakendamise tulemus on 0,0072.

```
arvuta_sd2=function(andmestik,kv){
  kohandatud=andmestik%>% mutate(hinnang=tinglik_kv-pind_sees*kv)%>% select(i,j,hinnang)
  kohandatud1=kohandatud%>% mutate(i=i-1,h1=hinnang)%>% select(i,j,h1)
  kohandatud2=kohandatud%>% mutate(i=i+1,h2=hinnang)%>% select(i,j,h2)
  kohandatud3=kohandatud%>% mutate(j=j-1,h3=hinnang)%>% select(i,j,h3)
  kohandatud4=kohandatud%>% mutate(j=j+1,h4=hinnang)%>% select(i,j,h4)
  vead=kohandatud%>%
    inner_join(kohandatud1,by=c("i","j"))%>%
    inner_join(kohandatud2,by=c("i","j"))%>%
    inner_join(kohandatud3,by=c("i","j"))%>%
    inner_join(kohandatud4,by=c("i","j"))%>%
    mutate(vead=hinnang-(h1+h2+h3+h4)/4)%>% pull(vead)
  return(sqrt(4*sum(vead^2)/5)/sum(andmestik$pind_sees))
}
arvuta_sd2(piirkond_laiendatud,osakaal1)
## [1] 0.007200813
```

Meetod 2: Alternatiivselt võrdleme joonisel oleva nelja ruudu hinnangute varieeruvust diagonaalide vahe $(h1-h2-h3+h4)/2$ abil (klassikaline Soome süsteem). Üldkeskmise veahinnang on nüüd 0,0070 (funktsiooni *arvuta_sd3* rakendamise tulemus).

h1	h2
h3	h4

```
arvuta_sd3=function(andmestik,kv){
  kohandatud=andmestik%>%mutate(hinnang=tinglik_kv-pind_sees*kv)%>%select(i,j,hinnang)
  kohandatud2=kohandatud%>%mutate(i=i+1,h2=hinnang)%>%select(i,j,h2)
  kohandatud3=kohandatud%>%mutate(i=i+1,j=j+1,h3=hinnang)%>%select(i,j,h3)
  kohandatud4=kohandatud%>%mutate(j=j+1,h4=hinnang)%>%select(i,j,h4)
  vead=kohandatud%>%
  inner_join(kohandatud2,by=c("i","j"))%>%
  inner_join(kohandatud3,by=c("i","j"))%>%
  inner_join(kohandatud4,by=c("i","j"))%>%
  mutate(vead=(hinnang-h2+h3-h4)/2)%>%pull(vead)
  return(sqrt(sum(vead^2))/sum(andmestik$pind_sees))
}
arvuta_sd3(piirkond_laiendatud,osakaal1)
## [1] 0.007020867
```

Näeme, et kahel eri meetodil saadud veahinnangud (0,0072 ja 0,0070) on väga lähedased, mistõttu mõlemad meetodid on võrdselt kasutatavad. Siiski, meetodi 2 kasuks räägib selle pikaajaline kasutuskogemus Soomes. Metsamaa osakaalu suhteline viga tuleb sel juhul: $0,0070 * 1,96 / 0,5342 = 2,57\%$.

4.4.3. Keskmise hektaritagavara hindamine ruudustiku kaudu

Sarnaselt pindalaga hindame ruudustiku abil ka keskmist hektaritagavara metsamaal.

1) Esialgu ignoreerime täielikult kasvukoha proovitükke ning teeme kõik arvutused tagavaraproovitükkide baasil, kasutades tegelikke (mitte mahumudeli abil saadud) mahtusid.

```
taiendatud_trakt=tmp_prtosa%>%inner_join(tmp_prt%>%select(prt_aid,pkaugus),by="prt_aid")%>%mutate(tagavara=if_els
e(maakategooria %in% c("M","MM"),arv_maht_es*pindosak,0),pindosak=if_else(maakategooria %in%
c("M","MM"),pindosak,0))%>%filter( pkaugus %in% c("04","08"))%>%
  group_by(atrakt_id)%>%summarize(tagavara=sum(tagavara)/8,pind_sees=sum(pindosak)/8)
piirkond=taiendatud_trakt%>%inner_join(trakt_piirkond,by="atrakt_id")%>%group_by(i,j)%>%summarize_at(vars(tagavar
a,pind_sees),.funs=~mean(.x))%>%ungroup()
#laiendame nullidega
piirkond_laiendatud=koik_indeksid%>%left_join(piirkond,by=c("i","j"))%>%mutate(tinglik_kv=coalesce(tagavara,0),pind_s
ees=coalesce(pind_sees,0))%>%select(i,j,tinglik_kv,pind_sees)
```

Keskmise hektaritagavara hinnang metsamaal (ainult tagavaraproovitükkide baasil):

Versioon 4.0
Kalev Pärna 13.02.2023



```
k_tagavara=sum(piirkond_laiendatud$tinglik_kv)/sum(piirkond_laiendatud$pind_sees)
k_tagavara
## [1] 208.5951
arvuta_sd2(piirkond_laiendatud,k_tagavara)
## [1] 2.565149
arvuta_sd3(piirkond_laiendatud,k_tagavara)
## [1] 2.54565
```

Näeme, et meetoditega 1 ja 2 saadud veahinnangud sd2 ja sd3 on jälle väga lähedased. Keskmise hektaritagavara suhteline viga tuleb klassikalise ruuduskeemi (meetod 2) abil arvutatuna seega $2,54565 * 1,96 / 208,5951 = 2,39\%$.

2) Teeme sama arvutuse, aga metsamaa pindala hindamisel arvestame ka kasvukoha proovitükkidelt saadavat infot (mahud võetakse endiselt ainult tagavaraproovitükkidelt).

```
taiendatud_trakt=tmp_prtosa%>%inner_join(tmp_prt%>%select(prt_aid,pkaugus),by="prt_aid")%>%mutate(tagavara=if_els
e((maakategooria %in% c("M","MM")) & (pkaugus %in%
c("04","08")),arv_maht_es*pindosak,0),pindosak=if_else(maakategooria %in% c("M","MM"),pindosak,0))%>%
group_by(atrakt_id)%>%summarize(tagavara=sum(tagavara)/8,pind_sees=sum(pindosak)/16)
piirkond=taiendatud_trakt%>%inner_join(trakt_piirkond,by="atrakt_id")%>%group_by(i,j)%>%summarize_at(vars(tagavar
a,pind_sees),.funs=~mean(.x))%>%ungroup()
#laiendame nullidega
piirkond_laiendatud=koik_indeksid%>%left_join(piirkond,by=c("i","j"))%>%mutate(tinglik_kv=coalesce(tagavara,0),pind_s
ees=coalesce(pind_sees,0))%>%select(i,j,tinglik_kv,pind_sees)
```

Keskmise hektaritagavara hinnang metsamaal (kui kasutatakse ka kasvukohaproovitükkide metsamaa infot):

```
k_tagavara=sum(piirkond_laiendatud$tinglik_kv)/sum(piirkond_laiendatud$pind_sees)
k_tagavara
## [1] 209.0796
arvuta_sd2(piirkond_laiendatud,k_tagavara)
## [1] 2.662394
arvuta_sd3(piirkond_laiendatud,k_tagavara)
## [1] 2.662913
```

Näeme, et kasvukohaproovitükkidelt saadav lisainfo metsasuse kohta jätab keskmise tagavarahinnangu praktiliselt samaks, kuid pisut suurendab hinnangu määramatust. Seega ei ole mõtet keskmise mahu hindamisel kasutada kasvukohaproovitükkide infot metsasuse kohta. Kuid kas ei oleks õige keskmise mahu hindamisel arvestada (lisaks tagavaraproovitükkidele) mõlemat - nii kasvukohaproovitükkide metsamaa pindalaid kui ka mahte? Ka siin on vastus eitav, sest tagavaraproovitükkide mahud pole tegelikud mahud vaid mahumudeli abil hinnatud väärtused, mis veaarvutuseks ei kõlba. Kokkuvõttes on keskmise hektaritagavara hindamine soovitatav teostada üksnes tagavaraproovitükkide alusel (tegelikud mahud, pindalad).

Kokkuvõte:

Hindamine ja veaarvutus korrapärase ruudustiku abil on praktiliselt realiseeritav, annab võrreldes traktipõhiste hinnangutega ligilähedaselt samad tulemused ning võimaldab veaarvutuses kasutada teistes riikides (sh Soomes) ammu järeleproovitud meetodit. Samas võib öelda, et erinevalt praegusest veaarvutusest, mis alahindab määramatust oluliselt, on uus veahinnang mõnevõrra konservatiivne ehk hindab määramatust pisut üle. Ühtlasi selgus ka, et keskmise hektaritagavara hindamisel ja veaarvutuses on mõistlik kasutada üksnes tagavaraproovitükkide pealt saadavaid pindalahinnanguid ja sealseid tegelikke mahtusid.

4.5. Väikeste osakogumite hindamine ja veaarvutus

Seni oleme vaadelnud metsamaa pindala ja metsa tagavara hindamist kogu riigi ulatuses, kuid samad hindamise ja veaarvutuse meetodid on rakendatavad ka väikeste osakogumite korral. **Väikese osakogumi** all mõistame ühe või mitme tunnuse järgi selekteeritud puistu tüüpi, näiteks 51-60 aastaste männikute kogum. Järgnevas demonstreerime kahe konkreetse näite varal väikeste osakogumite hindamist. Meie poolt soovitatud meetodid tagavad ka **summeeruvuse nõude** täidetuse, mis seisneb selles, et osakogumitele arvatud hinnangud summeeruvad üldhinnanguteks. See tähendab, et kui leida esmalt tagavara üldhinnang ning seejärel tagavara hinnangud väikestes osakogumites, mis on määratud näiteks vanuseklasside ja peapuuliikide järgi, siis viimaseid summeerides üle kõigi vanuseklasside ja peapuuliikide saame tulemuseks tagavara üldhinnangu. Summeeruvuse nõue on loomulik ning hästimõistetav loogiline tingimus ning seetõttu hinnangute üks soovitavaid omadusi.

4.5.1. Andmete ettevalmistamine väikeste osakogumite hindamiseks

Alljärgneva koodi algusosa, mille abil moodustatakse kogu Eestit kattev ruudustik, ühtib eespool toodud koodiga (p. 4.4.1), mida kasutati üldhinnangute leidmiseks.

Vaatleme hinnangute arvutamise loogikat juhul, kus andmed agregeeritakse kõigepealt traktide ja seejärel regulaarsete ruutude peale, mis katavad (teatava varuga) kogu Eesti pinna. Siin arvutustes oleks hea kasutada Eesti maismaa pindala teadaoleva suurusena

```
# andmete sisselugemine
```

```
library(tidyverse)
```

```
library(ggplot2)
```

```
aasta1=2016;aasta2=2020
```

```
tmp_prt=prt%>%dplyr::filter(aasta>=aasta1,aasta<=aasta2,psees=="J")
```

```
tmp_prtosa=tmp_prt%>%dplyr::select(aproovitykk_id)%>%inner_join(prtosa,by="aproovitykk_id")%>%filter(on_sees=="J")
```

```
#tmp_klupp=tmp_prt%>%dplyr::select(aproovitykk_id)%>%inner_join(klupp)%>%filter(kaugus<=100,arv_d13>=8)
```

```
tmp_trakt=trakt%>%filter(aasta>=aasta1,aasta<=aasta2)
```

Ruudukete defineerimisel võib kasutada sobivalt pööratud koordinaadistikku. Arvutame kõigepealt traktikeskmiste koordinaadid ja siis need koordinaadid pööratud teljestikus.

```
theta=0#-0.27  
vec=c(cos(theta),sin(theta))  
vec2=c(-sin(theta),cos(theta))  
M=matrix(c(vec,vec2),byrow=TRUE,nrow=2)
```

Trakti keskpunkti leidmiseks leiame esimese olemasolevate koordinaatidega seesoleva proovitüki ja liidame sellele vektori, mis peaks keskele nihutama.

```
#leiame andmetelt, millised on põhjaküljele ja idaküljele vastavad vektorid keskmiselt (ei ole päris vertikaalsed ja horisontaalsed)  
#sisuliselt saaks väga sarnased keskpunkti koordinaadid, kui võtta nendeks vektoriteks (800,0) ja (0,800)  
abi_nihe=tmp_prt%>%pivot_wider(names_from=c("pkylg","pkaugus"),values_from=c(koord_e,koord_n),id_cols=atrakt_id)%>%  
  mutate(v1x=(koord_e_N_08-koord_e_W_08)/2,v1y=(koord_n_N_08-koord_n_W_08)/2,v2x=(koord_e_N_08-koord_e_E_08)/2,v2y=(koord_n_N_08-koord_n_E_08)/2)  
vecs=abi_nihe %>%summarize(across(v1x:v2y,~mean(.x,na.rm=TRUE)))  
v1=c(vecs$v1x[1],vecs$v1y[1])  
v2=c(vecs$v2x[1],vecs$v2y[1])  
#teisenduste tabel (mida peab liitma, et keskele saada)  
teisendused=tibble(pkylg=rep(c("W","N","E","S"),each=4),pkaugus=rep(c("02","04","06","08"),times=4),  
  muut_x=rep(c(v1[1],-v2[1],-v1[1],v2[1]),each=4)+rep(c(0.5,0,-0.5,-1),times=4)*rep(c(v2[1],v1[1],-v2[1],-v1[1]),  
  each=4),  
  muut_y=rep(c(v1[2],-v2[2],-v1[2],v2[2]),each=4)+rep(c(0.5,0,-0.5,-1),times=4)*rep(c(v2[2],v1[2],-v2[2],-v1[2]),  
  each=4))  
#lisame keskpunkti koordinaadid  
trakt_keskpunktiga=tmp_prt%>%left_join(teisendused,by=c("pkylg","pkaugus"))%>%group_by(atrakt_id)%>%filter(row_number()==1)%>%  
  ungroup()%>%mutate(koord_e=koord_e+muut_x,koord_n=koord_n+muut_y)%>%mutate(id=row_number())%>%select(atrakt_id,koord_e,koord_n,id)  
#teisendatud=as.matrix(tmp_prt_nurk%>%dplyr::select(koord_e,koord_n))%*%t(M)  
teisendatud=as.matrix(trakt_keskpunktiga%>%dplyr::select(koord_e,koord_n))%*%t(M)  
colnames(teisendatud)=c("x","y")  
trakt_teisendatud=trakt_keskpunktiga%>%dplyr::bind_cols(as_tibble(teisendatud))
```

Lisame nüüd traktile regulaarse võrgu info.

```
xlimits=range(trakt_teisendatud$x)  
ylimits=range(trakt_teisendatud$y)  
kyljepikkus=5000 #(meetrites)  
pind_tykk=kyljepikkus^2/10^4 #hektarites  
xp=seq(xlimits[1]-12.5*kyljepikkus,xlimits[2]+12.5*kyljepikkus,by=kyljepikkus)  
yp=seq(ylimits[1]-12.5*kyljepikkus,ylimits[2]+12.5*kyljepikkus,by=kyljepikkus)  
i_x=as.numeric(cut(trakt_teisendatud$x,breaks=xp))  
i_y=as.numeric(cut(trakt_teisendatud$y,breaks=yp))  
trakt_piirkond=mutate(trakt_teisendatud,i=i_x,j=i_y)%>%select(atrakt_id,i,j)  
koi_k_indeksid=tibble(i=rep(1:max(i_x),each=max(i_y)),j=rep(1:max(i_y),times=max(i_x)))  
#joonis, kus on kõik regulaarse võrgu ruudud koos traktidega  
#ggplot(trakt_teisendatud)+geom_point(aes(x,y))+scale_x_continuous(breaks=xp)+scale_y_continuous(breaks=yp)+ theme(panel.grid.major=element_line(color="red",linetype=2))
```

Veahinnangute arvutamisel kasutame nn Soome skeemi regulaarsel võrgul teostatud kihtvalimi korral:

$$D\left[\sum_{i,j} X_{i,j}\right] = E\left[\left(X_{i,j} - X_{i,j+1} - X_{i+1,j} + X_{i+1,j+1}\right)^2/4\right],$$

kus $X_{i,j}$ on huvipakkuva suuruse (nt ruudus oleva metsamaa pindala või teatud tüüpi metsa maht vms) jaoks saadud hinnang ruudus olevate traktide abil. Enamasti aga on suurus X võrdne nulliga väljaspool mingile tingimusele A vastavat piirkonda (nt maismaa, Harjumaa

maismaa, Harjumaa metsamaa jms). Olgu S_A tingimusele A vastava kogupindala hinnang (või teadaolev väärtus).

```
#pind_eeesti= 4533902.3#hektarites
pind_eeesti= 4346800 #sellest on suured järved välja jäetud, maakondade summaarne pindala
arvuta_sd3=function(andmestik,kv){
  kohandatud=andmestik%>%mutate(hinnang=tinglik_kv-pind_sees*kv)%>%select(i,j,hinnang)
  kohandatud2=kohandatud%>%mutate(i=i+1,h2=hinnang)%>%select(i,j,h2)
  kohandatud3=kohandatud%>%mutate(i=i+1,j=j+1,h3=hinnang)%>%select(i,j,h3)
  kohandatud4=kohandatud%>%mutate(j=j+1,h4=hinnang)%>%select(i,j,h4)
  vead=kohandatud%>%
  inner_join(kohandatud2,by=c("i","j"))%>%
  inner_join(kohandatud3,by=c("i","j"))%>%
  inner_join(kohandatud4,by=c("i","j"))%>%
  mutate(vead=(hinnang-h2+h3-h4)/2)%>%pull(vead)
  return(list(sd=sqrt(sum(vead^2)),sd_tinglik=sqrt(sum(vead^2))/sum(andmestik$pind_sees)))
}
arvuta_tingimustel=function(cond_A,cond_y,y_string,samm,pind_A,max_per_trakt=16){
  #cond_A on tmp_prtoa jaoks arvessetulevate proovitükkide tingimus
  #cond_y on tingimus, mis määrab, kas summeeritav suurus nullitakse või mitte
  #y_string annab summeeritava väärtuse veerunime stringina
  #samm on selleks, et saaks hõrendatud võrgul korduvaid arvutusi teha
  #pind_A on tingimusele A vastava Eesti osa teadaolev pindala
  #max_per_trakt annab, mitu proovitükki trakti kohta maksimaalselt arvesse läheb
  taiendatud_trakt=tmp_prtoa%>%mutate(mm_osak=if_else(cond_y,get(y_string),0),pind_osak=if_else(cond_A,pindosak,
0))%>%
  group_by(atrakt_id)%>%summarize(mm_proc=sum(mm_osak)/max_per_trakt,pind_sees=sum(pind_osak)/max_per_trakt)
  piirkond=taiendatud_trakt%>%inner_join(trakt_piirkond,by="atrakt_id")%>%group_by(i,j)%>%summarize_at(vars(mm_p
roc,pind_sees),.funs=~mean(.x))%>%ungroup()
#laiendame nullidega
  piirkond_laiendatud=koik_indeksid%>%left_join(piirkond,by=c("i","j"))%>%mutate(tinglik_kv=coalesce(mm_proc,0),pin
d_sees=coalesce(pind_sees,0))%>%select(i,j,tinglik_kv,pind_sees)
#arvutame hõrendatud võrguga
  mitu=samm^2
  tulem=tibble(i=rep(NA,mitu),j=rep(NA,mitu),mu=rep(NA,mitu),sd=rep(NA,mitu))#,sd2=rep(NA,mitu))
  rida=1
  for(ii in 1:samm){
    for(jj in 1:samm){
      andmestik=piirkond_laiendatud%>%filter((i-1)%%samm==ii-1,(j-1)%%samm==jj-1)%>%mutate(i=i/%samm+1,j=j%/
%samm+1)
      osakaal=sum(andmestik$tinglik_kv)/sum(andmestik$pind_sees)
      tulem$[rida]=ii
      tulem$[rida]=jj
      tulem$mu[rida]=osakaal*pind_A
      sd3_tul=arvuta_sd3(andmestik,osakaal)
      #tulem$sd[rida]=sd3_tul$sd*pind_tykk*samm^2
      tulem$sd[rida]=sd3_tul$sd_tinglik*pind_A
      rida=rida+1
    }
  }
  return(tulem)
}
```

Suhtelised vead on allpool leitud standardhälbe ja väärtuse suhtena. Kui suhtelise vea all mõista 95% usaldusintervalli laiust, siis tuleb need korrutada veel vastava kordajaga (normaaljaotuse eeldusel arvuga 1.96).

Näide 1.

51-60a majandusmetsade (majandkategoria==0) kuusikute hindamine.

Pindala arvutused

Pindala koos veaga, kasutusel **kõik** andmed (tagavara+kasvukoha proovitükid).

```
tingimus=(tmp_prtosa$arv_epl %in% "KU") & (tmp_prtosa$vanus>=51)& (tmp_prtosa$vanus<=60) &
(!is.na(tmp_prtosa$qomand_kood)) & (tmp_prtosa$majandkategoria %in% c(0,1))
pindl_koik=arvuta_tingimustel(cond_A=!is.na(tmp_prtosa$atrakt_id),cond_y=tingimus,y_string="pindosak",samm=1,pind
_A=pind_eesti,max_per_trakt=16) %>% select(mu,sd) %>% mutate(suht_viga=sd/mu)
pindl_koik
```

```
## # A tibble: 1 × 3
##   mu sd suht_viga
##   <dbl> <dbl> <dbl>
## 1 35038. 2623. 0.0749
```

Pindala koos veaga, ainult **tagavara** proovitükkide andmed

```
tingimus=(tmp_prtosa$arv_epl %in% "KU") & (tmp_prtosa$vanus>=51)& (tmp_prtosa$vanus<=60) &
(!is.na(tmp_prtosa$qomand_kood)) & (tmp_prtosa$majandkategoria %in% c(0,1))&(tmp_prtosa$prtlilik==1)
pindl_tagavara=arvuta_tingimustel(cond_A=!is.na(tmp_prtosa$atrakt_id) &
tmp_prtosa$prtlilik==1,cond_y=tingimus,y_string="pindosak",samm=1,pind_A=pind_eesti,max_per_trakt=8) %>%
select(mu,sd)%>%mutate(suht_viga=sd/mu)
pindl_tagavara
```

```
## # A tibble: 1 × 3
##   mu sd suht_viga
##   <dbl> <dbl> <dbl>
## 1 37357. 3422. 0.0916
```

Tagavara arvutused

Defineerime lisaväljad mahuarvutuseks.

```
tmp_prtosa=tmp_prtosa
%>%mutate(maht1=pindosak*coalesce(arv_maht_es,0),maht2=pindosak*coalesce(teor_tagavara,0))
```

Keskmine hektaritagavara, ainult **tagavara** proovitükkide mõõtmisandmete pealt

```
tingimus=(tmp_prtosa$arv_epl %in% "KU") & (tmp_prtosa$vanus>=51)& (tmp_prtosa$vanus<=60) &
(!is.na(tmp_prtosa$qomand_kood)) & (tmp_prtosa$majandkategoria %in% c(0,1))&(tmp_prtosa$prtlilik==1)
kesk_tagavara1=arvuta_tingimustel(cond_A=tingimus,cond_y=tingimus,y_string="maht1",samm=1,pind_A=1,max_per_tra
kt=8) %>% select(mu,sd)%>% mutate(suht_viga=sd/mu)
kesk_tagavara1
```

```
## # A tibble: 1 × 3
##   mu sd suht_viga
##   <dbl> <dbl> <dbl>
## 1 319. 13.8 0.0435
```

Keskmine hektaritagavara, prognoositud mahtude pealt arvutatud tulemus (siin viga on alahinnatud!)

See arvutus on tehtud lihtsalt võrdluseks, veaarvutus ei ole siin korrektne.

```
tingimus=(tmp_prtosa$arv_epl %in% "KU") & (tmp_prtosa$vanus>=51)& (tmp_prtosa$vanus<=60) &
(!is.na(tmp_prtosa$qomand_kood)) & (tmp_prtosa$majandkategoria %in% c(0,1))&(tmp_prtosa$prtlilik<3)
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
arvuta_tingimustel(cond_A=tingimus,cond_y=tingimus,y_string="maht2",samm=1,pind_A=1,max_per_trakt=16) %>%  
select(mu,sd)%>% mutate(suht_viga=sd/mu)
```

```
## # A tibble: 1 × 3  
##   mu   sd   suht_viga  
##   <dbl> <dbl>   <dbl>  
## 1 302. 4.38  0.0145
```

Kogumahu arvutamine

- Kogumahu arvutamine nn otsemeetodil (ei kasuta eraldi pindala hinnangut)

```
kogumaht1_otse=arvuta_tingimustel(cond_A=!is.na(tmp_prtosa$atrakt_id)&tmp_prtosa$prtliik==1,cond_y=(tmp_prtosa$  
rv_epl %in% "KU") & (tmp_prtosa$vanus>=51) & (tmp_prtosa$vanus<=60) & (!is.na(tmp_prtosa$qomand_kood)) &  
(tmp_prtosa$majandkategoria %in%  
c(0,1))&tmp_prtosa$prtliik==1,y_string="maht1",samm=1,pind_A=pind_eesti,max_per_trakt=8) %>% select(mu,sd)%>%  
mutate(suht_viga=sd/mu)  
kogumaht1_otse
```

```
## # A tibble: 1 × 3  
##   mu   sd   suht_viga  
##   <dbl> <dbl>   <dbl>  
## 1 11902428. 1212883.  0.102
```

- Kogumahu arvutamine hektaritagavara ja pindalahinnangu abil

Kogumahtu saab arvutada ka korrutisena s.t. kogumaht= keskmine hektaritagavara * pindala, kusjuures vea arvutamisel suhtelised vead liituvad (eeldusel, et suhtelised vead keskmise hektaritagavara ja pindala arvutamisel on sõltumatud).

Esiteks vaatleme juhtu, kus pindala arvutame ainult **tagavara** proovitükkide pealt.

```
maht=kesk_tagavara1$mu*pind1_tagavara$mu  
sd=maht*sqrt(kesk_tagavara1$suht_viga^2+pind1_tagavara$suht_viga^2)  
kogumaht1_kaudne_tagavara=tibble(mu=maht, sd=sd)%>% mutate(suht_viga=sd/mu)  
kogumaht1_kaudne_tagavara
```

```
## # A tibble: 1 × 3  
##   mu   sd   suht_viga  
##   <dbl> <dbl>   <dbl>  
## 1 11902428. 1206715.  0.101
```

Samas, kui kasutada pindala täpsemat hinnangut (**tagavara+kasvukoha** proovitükid), saame sellise tulemuse:

```
maht=kesk_tagavara1$mu*pind1_koik$mu  
sd=maht*sqrt(kesk_tagavara1$suht_viga^2+pind1_koik$suht_viga^2)  
kogumaht1_kaudne_tagavara=tibble(mu=maht, sd=sd)%>% mutate(suht_viga=sd/mu)  
kogumaht1_kaudne_tagavara
```

```
## # A tibble: 1 × 3  
##   mu   sd   suht_viga  
##   <dbl> <dbl>   <dbl>  
## 1 11163416. 966293.  0.0866
```

Näite 1 tulemusi omavahel võrreldes näeme, et kui pindala hinnangusse kaasata ka kasvukoha proovitükid, siis pindala- ja mahuhinnangute täpsus mõnevõrra paraneb.

Näide 2.

Esimese boniteedi range kaitsega männikute hindamine.

Pindala arvutused

Pindala koos veaga, kasutusel **kõik** andmed (tagavara+kasvukoha proovitükid).

```
tingimus=(tmp_prtosa$arv_epl %in% "MA") & (tmp_prtosa$majandkategooria==2)&
(tmp_prtosa$maakategooria%in% c("M","MM")) & (tmp_prtosa$inv_bon %in% "1")
pindl_koik=arvuta_tingimustel(cond_A=!is.na(tmp_prtosa$atrakt_id),cond_y=tingimus,y_string="pindosak",samm=1,pind
_A=pind_eesti,max_per_trakt=16) %>% select(mu,sd) %>% mutate(suht_viga=sd/mu)
pindl_koik
```

```
## # A tibble: 1 × 3
##   mu sd   suht_viga
##   <dbl> <dbl> <dbl>
## 1 28268. 3023. 0.107
```

Pindala koos veaga, ainult **tagavara** proovitükkide andmed.

```
tingimus=(tmp_prtosa$arv_epl %in% "MA") & (tmp_prtosa$majandkategooria==2)&
(tmp_prtosa$maakategooria%in% c("M","MM")) & (tmp_prtosa$inv_bon %in% "1")&(tmp_prtosa$prtlilik==1)
pindl_tagavara=arvuta_tingimustel(cond_A=!is.na(tmp_prtosa$atrakt_id) &
tmp_prtosa$prtlilik==1,cond_y=tingimus,y_string = "pindosak",samm=1,pind_A=pind_eesti,max_per_trakt=8) %>%
select(mu,sd)%>%mutate(suht_viga=sd/mu)
pindl_tagavara
```

```
## # A tibble: 1 × 3
##   mu sd   suht_viga
##   <dbl> <dbl> <dbl>
## 1 29924. 3811. 0.127
```

Tagavara arvutused

Defineerime lisaväljad mahuarvutuseks

```
tmp_prtosa=tmp_prtosa %>%mutate(maht1=pindosak*coalesce(arv_maht_es,0),maht2=pindosak*coalesce(teor_tagavara,0))
```

Keskmine hektaritagavara, **tagavara** proovitükkide mõõtmisandmetelt

```
tingimus=(tmp_prtosa$arv_epl %in% "MA") & (tmp_prtosa$majandkategooria==2)&
(tmp_prtosa$maakategooria%in% c("M","MM")) & (tmp_prtosa$inv_bon %in% "1")&(tmp_prtosa$prtlilik==1)
kesk_tagavara1=arvuta_tingimustel(cond_A=tingimus,cond_y=tingimus,y_string="maht1",samm=1,pind_A=1,max_per_trak
t=8) %>% select(mu,sd)%>% mutate(suht_viga=sd/mu)
kesk_tagavara1
```

```
## # A tibble: 1 × 3
##   mu sd   suht_viga
##   <dbl> <dbl> <dbl>
## 1 367. 19.0 0.0517
```

Keskmine hektaritagavara, mahumudeliga prognoositud mahtudelt arvatud tulemus (NB!
Siin viga on alahinnatud!)

See arvutus on tehtud lihtsalt võrdluseks, veaarvutus on ebakorrektn.

```
tingimus=(tmp_prtosa$arv_epl %in% "MA") & (tmp_prtosa$majandkategooria==2)&
(tmp_prtosa$maakategooria%in% c("M","MM")) & (tmp_prtosa$inv_bon %in% "1")
arvuta_tingimustel(cond_A=tingimus,cond_y=tingimus,y_string="maht2",samm=1,pind_A=1,max_per_trakt=16) %>% sele
ct(mu,sd)%>% mutate(suht_viga=sd/mu)
```

```
### A tibble: 1 × 3  
##   mu sd   suht_viga  
##   <dbl> <dbl>   <dbl>  
## 1 355. 10.4 0.0293
```

Kogumahu arvutused

- Kogumahu arvutamine nn otsemeetodil (ei kasuta eraldi pindala hinnangut)

```
tingimus=(tmp_prtosa$arv_epl %in% "MA") & (tmp_prtosa$majandkategoria==2)&  
(tmp_prtosa$maakategoria%in% c("M","MM")) & (tmp_prtosa$inv_bon %in% "1")  
kogumaht1_otse=arvuta_tingimustel(cond_A=lis.na(tmp_prtosa$satrakt_id)&tmp_prtosa$sprtliik==1,cond_y=tingimus&tmp_  
prtosa$sprtliik==1,y_string="maht1",samm=1,pind_A=pind_eesti,max_per_trakt=8) %>% select(mu,sd)%>% mutate(suht_vi  
ga=sd/mu)  
kogumaht1_otse  
  
### A tibble: 1 × 3  
##   mu sd   suht_viga  
##   <dbl> <dbl>   <dbl>  
## 1 10992516. 1475982. 0.134
```

- Kogumahu arvutamine hektaritagavara ja pindalahinnangu abil

Kogumahtu saab arvutada ka korrutisena: kogumaht = keskmine hektaritagavara * pindala, kusjuures vea arvutamisel suhtelised vead liituvad (eeldusel, et suhtelised vead keskmise hektaritagavara ja pindala arvutamisel on sõltumatud).

Esiteks vaatleme juhtu, kus pindala arvutame ainult **tagavara** proovitükkide pealt.

```
maht=kesk_tagavara1$mu*pind1_tagavara$mu  
sd=maht*sqrt(kesk_tagavara1$suht_viga^2+pind1_tagavara$suht_viga^2)  
kogumaht1_kaudne_tagavara=tibble(mu=maht, sd=sd)%>% mutate(suht_viga=sd/mu)  
kogumaht1_kaudne_tagavara  
  
### A tibble: 1 × 3  
##   mu sd   suht_viga  
##   <dbl> <dbl>   <dbl>  
## 1 10992516. 1510819. 0.137
```

Samas, kui kasutada pindala täpsemat hinnangut (**tagavara+kasvukoha** proovitükid), saame sellise tulemuse:

```
maht=kesk_tagavara1$mu*pind1_koik$mu  
sd=maht*sqrt(kesk_tagavara1$suht_viga^2+pind1_koik$suht_viga^2)  
kogumaht1_kaudne_tagavara=tibble(mu=maht, sd=sd)%>% mutate(suht_viga=sd/mu)  
kogumaht1_kaudne_tagavara  
  
### A tibble: 1 × 3  
##   mu sd   suht_viga  
##   <dbl> <dbl>   <dbl>  
## 1 10384132. 1233453. 0.119
```

Keskmine vanus

```
tmp_prtosa=tmp_prtosa%>%mutate(keskvanus=vanus*pindosak)  
arvuta_tingimustel(cond_A=tingimus,cond_y=tingimus,y_string="keskvanus",samm=1,pind_A=1,max_per_trakt=16) %>%  
select(mu,sd)%>% mutate(suht_viga=sd/mu)
```

```
### A tibble: 1 × 3  
##   mu sd   suht_viga
```

<dbl> <dbl> <dbl>
1 77.0 2.94 0.0382

Näite 2 tulemusi omavahel võrreldes näeme, et kui pindala hinnangusse kaasata ka kasvukoha proovitükid, siis pindala- ja mahuhinnangute täpsus mõnevõrra paraneb.

4.5.2. Analüüs ja kokkuvõte

Ülaltoodud kahest näitest võib teha järelduse, et väikeste osakogumite hindamisel saame mõnevõrra täpsema tulemuse, kui koos tagavara proovitükkidega kasutada ka kasvukoha proovitükkide lisainfot pindalade kohta. Näiliselt on see vastuolus meie senise järeldusega, et kogupindalade ja kogumahtude hindamisel tuleks kasutada üksnes tagavara proovitükkide andmeid (vt. p. 4.4 kokkuvõtet). Vastuolu seletus peitub ilmselt selles, et väikeste osakogumite korral on valimi maht (=traktide arv) nii väike, nii et globaalne looduslik varieeruvus ei tule andmetes enam esile ning rohkem hakkab kaasa rääkima trakti sees olevate vaatluste arv. Kasvukoha proovitükkide kaasamine arvutustesse vähendab hinnangute suhtelist viga ca 10-20%.

Teisest küljest, kasvukoha proovitükkide kaasamine pindalade hinnangutesse ning samal ajal keskmise hektaritagavara hindamine ainult tagavara proovitükkide pealt viib summeeruvusnõude rikkumiseni, mis on ebasoovitav. Seepärast näeme väikeste osakogumite hindamisele kolme võimalikku lahendusvarianti, millest igaüks tagab ka summeeruvusnõude täidetuse.

- 1) Operatiivne lahendus: sarnaselt üldhinnangutega (p.4.4) **kasutada ka väikeste osakogumite korral pindalade ja mahtude hindamisel üksnes tagavara proovitükkide andmeid.**

Ülaltoodud analüüsi valguses tähendaks see mõningast järeleandmist hinnangute täpsuses, kuid kuna väikeste osakogumite korral on suhtelised vead nagunii suured, siis kasvukohaproovitükkide pindalaandmete ignoreerimisega kaasnev hinnanguvea kasv 10-20% ei tohiks tekitada lisaprobleeme.

- 2) Alternatiivne lahendus, mis kaasab hinnangutesse ka kasvukoha proovitükkide andmed, seisneb **praeguse mahtude prognoosimise skeemi asendamises mõne täiuslikuma süsteemiga.**

Praeguse mahumudeli M(A, H100) rakendamise oluliseks puuduseks on see, et saadavate mahuprognoside pealt arvatud vead on tugevasti alahinnatud. Selleks, et saada adekvaatseid veahinnanguid, tuleks mudeliga saadud mahuprognosile liita ka juhuslik viga. Raskus on selles, et lisatava juhusliku vea dispersioon sõltub prognoositud väärtusest, nii nagu näiteks suure puu mahuprognosiga on reeglina suurem kui väikese puu mahuprognosiga. Antud teema väärib ja vajab edasist uurimist. Esmahinnangul näib olevat otstarbekas lähemalt uurida nn mitmese imputeerimise meetodit kasvukoha proovitükkidele mahtude tekitamiseks.

- 3) Perspektiivne lahendus on **kaugseire andmete kaasamine väikeste osakogumite hindamisse.** Kaugseire andmed (ALS, satelliidipildid) katavad kogu territooriumi, mis lubab sobivalt välja töötatud prognoosimudelite abil prognoosida ka harva esinevaid sündmusi.

5. Mõõtmiste subjektiivsuse analüüs

Antud osa eesmärk on uurida subjektiivse faktori olemasolu ja selle mõju suurust välimõõtmistel. Täpsemalt, 1) kas välimõõtmistes osalevate töögruppide vahel eksisteerivad süstemaatilised erinevused mõõtmistulemustes, 2) milliste tunnuste mõõtmisel need esinevad ja 3) kui suured on erinevused.

Need küsimused on SMI planeerimisel olulised, kuna üksnes kolme mõõtegrupiga töötav süsteem on gruppidevaheliste erinevuste suhtes väga sensitiivne. Võrdluseks võib mainida, et Rootsis ja Soomes kasutatakse enam kui 20 mõõtemeeskonda, mis on üksiku meeskonna tööiseärasuste suhtes palju vähem tundlik.

Ekspertide üldine arvamus on, et mõõtmiste subjektiivsust esineb vähem seal, kus mõõdetakse tehnilise instrumendiga (nt diameetri, kõrguse, kauguse, asimuudi mõõtmine), ja rohkem seal, kus mõõtmine tähendab silma järgi hindamist (nt vanus, arenguklass).

Järgnevas on toetunud nii Rootsi ja Soome metsateadlaste samateemalistele uurimustele kui ka meie endi SMI empiiriliste andmete põhjal tehtud analüüsidele. Lähemalt on uuritud SMI mõõtemeeskondade erinevusi raieküpsuse hindamisel.

5.1. Subjektiivsustegur üksikpuu mõõtmisel

5.1.1. Üksikpuu rinnasdiameetri mõõtmine (klappimine)

Klappimise korral on tegemist rinnasdiameetri instrumentaalse mõõtmisega, kus tulemus fikseeritakse 1 mm täpsusega. Mõõtmistulemus sõltub ühelt poolt mõõtmise kõrgusest, mis teoreetiliselt peaks olema 1,3m juurekaelast, kuid tegelikult varieerub mõnevõrra, ja teiselt poolt ka mõõtesuunast, kuna puu tüve ristlõige ei ole ideaalne ring. Artiklis (Ranneby et al. 1987) on erinevatele uuringutele viidates välja toodud, et kokkuvõttes on üksikpuu klappimistulemuse standardhälve suurusjärku 4-5mm. Sellest 1-2mm moodustab mõõtjate (klappijate) vaheline standardhälve, mida võib nimetada subjektiivsusfaktori mõju suuruseks klappimisel. Ühtlasi on leitud, et vähima standardhälbe mõõtjate vahel tagab see, kui rinnasdiameetri mõõtekõrguse määramiseks kasutada standardset 1,3 m pikkust varrast, pisut suuremad erinevused tekivad mõõtja riietusel oleva kõrgusmärgi kasutamiseega.

5.1.2. Üksikpuu kõrguse mõõtmine

Kõrgusmõõtjaga tehtud üksikpuude kordumõõtmiste standardhälve on leitud olevat ligikaudu 0,5m (Ranneby 1987; Kujala 1979, Päivinen & Yli-Kojola 1983). Samade allikate põhjal on mõõtetehnikute vaheline standardhälve (ehk subjektiivsusfaktori mõju suurus) kõrguse mõõtmisel ca 0,2m.

5.2. Subjektiivsustegur puistu vanuse hindamisel

5.2.1. Puistu vanuse hindamise täpsusest

Puistu vanus on oluline sisend nii SMI mahu arvutamise mudelis kui ka juurdekasvu arvutamise mudelis. SMI tulemuste raporteerimine toimub sageli just vanuseklasside lõikes, mistõttu raportite lugeja peab arvestama ka vanuseklassi määramisel tehtava veaga.

Puistu vanus määratakse silmamõõdulise hinnangu ja ka puursüdamikelt loetud info järgi, millega kaasneb suhteliselt suur hinnanguviga. Soome teadlaste poolt (Haara, Korhonen 2004) läbi viidud ulatuslikus katses võrreldi puistu vanuse visuaalseid hinnanguid samadel puistutel objektiivsete mõõtmiste teel leitud väärtustega. Hinnangute täpsust mõõdeti keskmise ruutvea (RMSE) abil, mille väärtuseks tuli kokkuvõttes 15,9 aastat (suhteline RMSE 28,2%), kusjuures RMSE väärtus olenes puistu karakteristikutest, näiteks enamuspuid liigist. Samas artiklis viidatud kolmes varasemas uurimuses kõikus vanuse hinnangu standardviga RMSE piirides 8-21 aastat. Nähtavasti võib selliseid vahet hinnanguid oletada ka meie SMI korral. (Toodud arvud on kooskõlas meie ekspertide poolt väljendatud arvamusega, et puistu vanuse hindamisel eksimine 10 aastat on normaalne.) Puistu vanuse hinnangu suur hajuvus tekitab mitmeid probleeme. Näiteks kasvõi SMI raporteerimistulemuste tõlgendamisel tuleb arvestada sellega, et tabelites esinevad vanuseklassid (nt 51-60a jne) sisaldavad mingis osas puistusid ka naaberklassidest, moodustades nendest teatava segu.

5.2.2. SMI tööühmade vahelised erinevused vanuse hindamisel

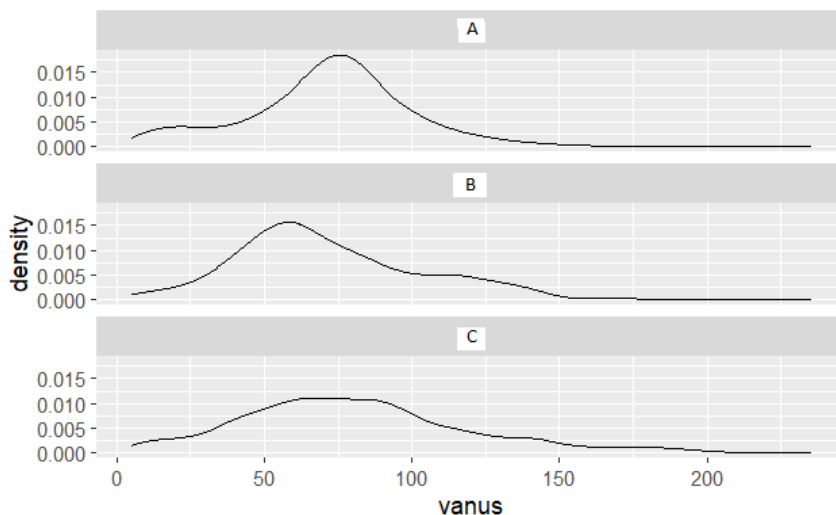
Järgnevas vaatleme lähemalt, kuidas SMI tööühmad hindavad puistu vanust oma poolt mõõdetud ajutistel proovitükkidel (võrreldavuse tagamiseks pole alalisi proovitükke analüüsi kaasatud, kuna üks tööühm tegutseb ainult ajutistel traktidel).

Allolevas tabelis on toodud kõikvõimalikud 'vanuse' väärtused, mis on märgitud tööühmade poolt ajutistel proovitükkidel (2017-2021, peapuuliik=mänd)

Tööühm-A		5	7	8	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105	110	120	125	130	140	150	160	170	180	190	200										
vanus		5	7	8	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105	110	120	125	130	140	150	160	170	180	190	200										
sagedus		4	3	1	17	31	22	23	23	14	26	20	57	25	89	34	181	35	208	26	92	14	75	2	39	29	1	14	8	5	1	1	2	1	1										
Tööühm-B		5	9	10	15	20	25	30	31	35	40	42	43	45	47	48	50	52	55	57	58	60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135	140	145	170					
vanus		5	9	10	15	20	25	30	31	35	40	42	43	45	47	48	50	52	55	57	58	60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135	140	145	170					
sagedus		1	1	1	2	2	3	3	1	6	8	2	1	8	1	3	9	3	18	1	1	13	21	7	10	13	7	6	6	6	1	8	5	5	3	4	3	3	1	1					
Tööühm-C		6	7	8	9	10	11	12	13	14	15	16	18	20	22	24	25	27	28	29	30	31	32	35	36	37	38	40	42	43	44	45	46	47	48	50									
vanus		6	7	8	9	10	11	12	13	14	15	16	18	20	22	24	25	27	28	29	30	31	32	35	36	37	38	40	42	43	44	45	46	47	48	50									
sagedus		2	2	1	1	6	2	1	3	2	5	1	1	4	3	2	3	4	2	2	5	1	2	6	1	4	2	30	2	2	1	12	1	5	3	18									
Tööühm-D		51	52	53	54	55	57	58	59	60	62	63	64	65	67	68	69	70	71	72	73	74	75	77	78	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	100
vanus		51	52	53	54	55	57	58	59	60	62	63	64	65	67	68	69	70	71	72	73	74	75	77	78	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	100
sagedus		1	2	3	3	25	7	3	2	25	3	6	1	27	5	5	2	28	1	2	2	2	25	8	1	23	2	2	3	5	25	3	3	3	1	31	1	2	2	1	27	1	1	2	27
Tööühm-E		102	104	105	106	107	110	112	113	115	117	120	125	130	135	137	140	145	150	155	160	163	165	170	173	175	180	182	185	190	195	200	235												
vanus		102	104	105	106	107	110	112	113	115	117	120	125	130	135	137	140	145	150	155	160	163	165	170	173	175	180	182	185	190	195	200	235												
sagedus		1	1	8	1	1	15	3	1	15	1	15	11	6	7	1	18	7	5	2	4	1	4	3	1	2	5	1	3	3	1	2	2												

Tabelis torkab silma, et enam eelistatakse 5 või 10-ga jaguvaid väärtusi. See kehtib eriti tööühma A korral, kus see on peaaegu kindel reegel ja kus täiendavaks omapäraks on vanuste 70 ja 80 erakordselt kõrge sagedus. Tööühmal C esineb küllalt palju ka vahepealseid väärtusi. Pole võimalik kinnitada, kas tööühma C näiliselt täpsemad hinnangud on ka tegelikult

täpsemad võrreldes tööruhaga A ja B. Igal juhul ilmneb vanuse hindamise juures gruppide „tööstiilide“ märgatav erinevus, mis peegeldub ka alljärgnevatel graafikutel, kus on esitatud puistute vanuseline jaotus tööruhade lõikes (tegemist on ülaltoodud sageduste põhjal koostatud tihedusgraafikutega).



Kokkuvõttes, ilmnenud subjektiivsed iseärasused puistu vanuse hindamisel ja selle võimalik mõju raieküpsuse määramisele ja muude näitajate prognoosimisele väärivad edasist analüüsi, mille juurde nüüd asumegi.

5.3. Subjektiivsuse analüüs raieküpsuse määramisel

Järgnevas esitame kolme tööruhma (A, B ja C) võrdleva analüüsi eesmärgiga tuvastada subjektiivse teguri olemasolu raieküpsuse määramisel. Allolev analüüs on ajendatud suurest erinevusest raieküpsuse metsa pindala kahe hinnangu vahel, millest üks on saadud alaliste ja teine ajutiste proovitükkide takserimisandmete põhjal (vastavalt 699 t. ha ja 591 t. ha). Kuna alaliste ja ajutiste proovitükkide mõõtmine toimub erinevate töögruppide poolt, siis on otstarbekas kontrollida hüpoteesi, et see erinevus on tingitud töögruppide vahelistest erinevustest metsa takserimisel ehk teisiti öeldes subjektiivsest faktorist. Analüüsis püütakse järk-järgult kõrvaldada teiste faktorite mõjud, mis samuti võivad tekitada erinevusi alaliste ja ajutiste proovitükkide raieküpsuse vahel.

Märkus: Terminit **'raieküpsus'** mõistetakse siinjuures „SMI välitööde juhendis“ toodud tähenduses, mille kohaselt SMI välitöötaja hindab raieküpsust metsa majandamise eeskirja (<https://www.riigiteataja.ee/akt/106042021008>) raievanuste tabelist lähtuvalt. Seega mets loetakse SMI mõttes raieküpsuks, kui metsa vanus ületab vastavale boniteedile ja puuliigile kehtestatud vanusepiiri. (Praktikas ei ole see aga vanus, mille järgi tegelikult saab raiuda, kuna metsa majandamise eeskirja kohaselt tuleb lisaks arvestada ka metsa koosseisu ja diameetrit.)

Töögruppide rakendamine alaliste ja ajutiste proovitükkide vahel on järgmine:

	Alalised	Ajutised
Töögrupp A	-	+
Töögrupp B	+	+
Töögrupp C	+	+

Ajutistest proovitükkidest 2/3 katab töörühm A, ülejäänud 1/3 jaotub töörühmade B ja C vahel. Alalised proovitükid jaotuvad enamvähem võrdselt töörühmade B ja C vahel. Seega eksisteerib võimalus, et ajutiste proovitükkide pealt saadud madalam hinnang tuleneb eeskätt sellest, et töörühm A hindab puistu raieküpsust tunduvalt konservatiivsemalt kui töörühmad B ja C. Püüame seda uurida lähemalt.

Vaatleme aastate 2017-2021 üldist statistikat ja raieküpsuse statistikat proovitükkide liikide järgi.

Raieküpsus on määratud tingimusega $prtosasarenguklass=7$.

tr_tyyp	pind	metsapind	raiekyps	raiekyps osakaal
Alaline vana 1	2191994,0	1083448,4	371597,7	0,1695
Ajutine 2	2170196,0	1040078,6	300549,0	0,1385
Alaline uus 3	1111427,0	551849,7	170426,2	0,1533

Näeme, et ajutistel proovitükkidel on raieküpsust määratud tunduvalt väiksemas määral.

Vaatame nüüd erinevust trakti tüübi ja töörühma kaupa.

	tr_tyyp	toogrupp	raiekyps	pind	osakaal
Alalised vanad	1	B	247574,01	1495406,04	0,1656
	1	C	123407,98	692843,50	0,1781
	1	D	615,75	3744,78	0,1644
Ajutised	2	A	182281,83	1454755,08	0,1253
	2	B	25156,77	158546,76	0,1587
	2	C	93110,43	556894,41	0,1672
Alalised uued	3	B	117057,96	773052,84	0,1514
	3	C	52432,00	334149,20	0,1569
	3	D	936,19	4225,44	0,2216

Näeme, et ajutiste proovitükkide väike raieküpsuse määr seondub eeskätt töörühmaga A. (Väga lähedased tulemused saime ka siis, kui arvestasime, et raieküpsuse hindamisel kasutatakse sama ulatusega (20m) infot kõikidel proovitükkidel ning eeldame, et osa rakendub ka suuremale piirkonnale proportsionaalselt. Sel juhul pind tähistab tingimustele vastavate proovitükkide arvu).

5.3.1. Täiendavad vaated töörühmade taustatunnuste jaotuste kohta

Uurime, kas töörühma A madal raieküpsuse määr võib olla tingitud mõnest muust objektiivsest tegurist, näiteks töörühma A poolt takseeritud proovitükkide teistsugusest metsakooslusest, väiksemast diameetrist vmt. Kõigepealt esitame tabeli erinevate töörühmade poolt hinnatud metsamaa tüüpi proovitükkide kohta (kuna mõnest läks ainult osa arvesse, siis võib olla murdarvuline).

toogrupp	pind
A	4309.443
B	5155.082
C	4072.881
D	19.000

Järgnevalt teostame analüüsi peapuuliikide kaupa. Iga töörühma korral on toodud jaotus peapuuliikide järgi (tulbad pind_A, pind_B, pind_C, pind_D annavad ülalt alla kokku 100%). Tulbad raiek_A, ..., raiek_D näitavad töörühmade A - D raieküpsuse määra antud peapuuliigi korral. Viimased tulbad diam_A, ..., diam_D näitavad keskmist diameetrit antud töörühma ja peapuuliigi korral.

Peapuuliik	pind_A	pind_B	pind_C	pind_D	raiek_A	raiek_B	raiek_C	raiek_D	diam_A	diam_B	diam_C	diam_D
HB	6,2%	6,2%	6,3%	NA	50,5%	49,7%	64,5%	NA	192,5	202,5	223,7	NA
KP	0,0%	0,0%	NA	NA	0,0%	0,0%	NA	NA	400,0	290,0	NA	NA
KS	34,1%	30,2%	24,6%	36,8%	25,0%	32,0%	28,8%	28,6%	164,3	172,1	163,9	201,4
KU	18,9%	17,5%	15,4%	27,3%	25,1%	28,9%	34,2%	38,6%	196,8	204,1	218,9	253,3
LH	0,0%	0,0%	0,2%	NA	0,0%	100,0%	0,0%	NA	390,0	340,0	162,1	NA
LM	2,7%	3,9%	5,5%	5,3%	26,1%	35,7%	33,0%	100,0%	179,2	181,8	169,2	270,0
LV	10,5%	10,1%	8,1%	14,8%	45,8%	45,9%	36,2%	35,5%	116,3	121,5	112,9	140,0
MA	26,6%	30,2%	38,1%	10,5%	18,8%	28,7%	29,2%	100,0%	222,2	233,8	237,0	290,0
PN	0,1%	0,2%	0,2%	NA	33,3%	10,3%	0,0%	NA	286,7	230,9	240,2	NA
PP	0,0%	0,1%	0,1%	NA	100,0%	66,7%	32,3%	NA	260,0	236,7	160,0	NA
RE	0,3%	0,4%	0,3%	5,3%	13,3%	29,9%	44,8%	0,0%	136,0	148,1	147,1	120,0
SA	0,2%	0,5%	0,2%	NA	31,2%	14,7%	20,9%	NA	301,0	220,2	235,7	NA
TA	0,1%	0,3%	0,6%	NA	0,0%	27,1%	33,3%	NA	156,7	276,5	332,3	NA
VA	0,1%	0,1%	0,2%	NA	0,0%	0,0%	0,0%	NA	180,0	168,3	185,0	NA
JA	NA	0,0%	0,1%	NA	NA	0,0%	66,7%	NA	NA	240,0	343,3	NA
PI	NA	0,0%	0,0%	NA	NA	69,3%	70,9%	NA	NA	130,0	160,0	NA
TL	NA	0,0%	NA	NA	NA	0,0%	NA	NA	NA	120,0	NA	NA
TM	NA	0,1%	0,0%	NA	NA	100,0%	72,2%	NA	NA	128,3	96,0	NA
KOKKU	100 %	100 %	100 %	100 %								

Tabelist nähtub esmalt, et rühma A korral on raieküpsete metsade osakaal kase, kuuse ja männimetsade puhul tunduvalt väiksem, kui teistel rühmadel (HB ja LV korral on grupp A teistega võrreldav). Samas aga on töörühmal A ka keskmine diameeter kuuse ja männi (ja osaliselt ka kase korral) väiksem kui teistel rühmadel. See jätab võimaluse, et töörühma A madalam raieküpsuse määr on peamiselt tingitud väiksemast diameetrist nende poolt mõõdetud proovitükkidel. Järgnevalt uurime võimalike tegurite rolli komplekselt, kasutades logistilist regressiooni.

5.3.2. Logistiline regressioon raieküpsuse mõjurite uurimiseks

Proovime mudeldamise teel kindlaks teha, kas töörühmal kui tunnusel on oluline mõju raieküpsuse määramisele. Kõigepealt vaatleme juhtu, kus raieküpsuse tõenäosust

prognoositakse peapuuliigi, trakti tüübi, 'onsaar' tunnuse ja kolme põhilise puuliigi (MA, KU, KS) ning tööruhma interaktsiooni abil. Alljärgnevates mudelites traktitüüp 1 ja töögrupp A määravad baastaseme, millega teisi võrreldakse.

```
raiekypsus_andmed <- prtosa%>%
  filter(aasta>a1,aasta<a2)%>%
  left_join(trakt,by="atrakt_id"%>%
  filter(on_sees=="J",pindosak==1,maakategooria=="M")%>%
  mutate(y=1*(arenguklass==7),toogrupp=factor(toogrupp),tr_tyyp=factor(tr_tyyp))
m1=glm(y~I(peapuuliik=="MA")*toogrupp+I(peapuuliik=="KU")*toogrupp+I(peapuuliik=="KS")*toogrupp+I(
peapuuliik=="LV")+I(peapuuliik=="HB")+I(peapuuliik=="LM")+tr_tyyp+onsaar,family=binomial(),data=raiek
ypsus_andmed)
summary(m1)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.117960	0.193964	-5.764	8.23e-09 ***
I(peapuuliik == "MA")TRUE	-0.248909	0.198843	-1.252	0.210648
Toogrupp_B	-0.002815	0.109466	-0.026	0.979482
Toogrupp_C	0.031620	0.107285	0.295	0.768198
Toogrupp_D	0.564536	1.022166	0.552	0.580747
I(peapuuliik == "KU")TRUE	0.118430	0.201214	0.589	0.556146
I(peapuuliik == "KS")TRUE	0.128975	0.193377	0.667	0.504796
I(peapuuliik == "LV")TRUE	0.888783	0.180163	4.933	8.09e-07 ***
I(peapuuliik == "HB")TRUE	1.359483	0.184041	7.387	1.50e-13 ***
I(peapuuliik == "LM")TRUE	0.428478	0.194143	2.207	0.027313 *
tr_tyyp2	-0.109500	0.061036	-1.794	0.072812 .
tr_tyyp3	-0.122680	0.057363	-2.139	0.032462 *
onsaar	0.077562	0.056810	1.365	0.172161
I(peapuuliik == "MA")TRUE:toogrupp_B	0.502960	0.135778	3.704	0.000212 ***
I(peapuuliik == "MA")TRUE:toogrupp_C	0.492162	0.139857	3.519	0.000433 ***
I(peapuuliik == "MA")TRUE:toogrupp_D	12.491066	139.2809	0.090	0.928540
Toogrupp_B:I(peapuuliik == "KU")TRUE	0.188660	0.147793	1.277	0.201774
Toogrupp_C:I(peapuuliik == "KU")TRUE	0.414785	0.156810	2.645	0.008166 **
Toogrupp_D:I(peapuuliik == "KU")TRUE	0.078233	1.371997	0.057	0.954528
Toogrupp_B:I(peapuuliik == "KS")TRUE	0.294282	0.126944	2.318	0.020438 *
Toogrupp_C:I(peapuuliik == "KS")TRUE	0.107052	0.138894	0.771	0.440858
Toogrupp_D:I(peapuuliik == "KS")TRUE	-0.440057	1.321301	-0.333	0.739098

Näha on, et **männi** korral on raieküpsuse tõenäosus rühmadel B ja C oluliselt suurem kui rühmal A (baastase). **Kuuse** korral on raieküpsuse tõenäosus oluliselt suurem rühmal C (võrreldes rühmaga A) ja **kase** korral rühmal B. Kõik see on hästi kooskõlas ülaltoodud tabeliga.

Vaatleme männimetsa raieküpsuse prognoosimist lähemalt, võttes arvesse ka diameetri.

Versioon 4.0
Kalev Pärna 13.02.2023



```
m2=glm(y~tr_tyyp+toogrupp*inv_diameeter,family=binomial(),data=filter(raiekypsus_andmed,peapuuliik=="MA",toogrupp!=D))
```

```
summary(m2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-8.490896	0.651920	-13.024	< 2e-16 ***
tr_tyyp2	-0.412479	0.120454	-3.424	0.000616 ***
tr_tyyp3	-0.200651	0.122978	-1.632	0.102764
toogrupp_B	2.377890	0.742112	3.204	0.001354 **
toogrupp_C	3.146126	0.720656	4.366	1.27e-05 ***
inv_diameeter	0.028803	0.002307	12.485	< 2e-16 ***
toogrupp_B:inv_diameeter	-0.008408	0.002632	-3.194	0.001402 **
toogrupp_C:inv_diameeter	-0.010831	0.002567	-4.219	2.45e-05 ***

Siit nähtub, et rühmadel B ja C on konstantne liige suurem (toogrupp_B ja toogrupp_C kordajad on positiivsed). See ütleb, et **ka pärast keskmise diameetri erinevusest tingitud mõju eemaldamist on rühma A raieküpsuse määr oluliselt väiksem kui rühmadel B ja C**. See tulemus tõendab tööühma kui subjektiivse faktori mõju olulisust raieküpsuse hindamisel.

Edasi nähtub, et tööühmadel B ja C on inv_diameetri kordajad väiksemad võrreldes rühmaga A (baastase). See tähendab, et **töögrupil A raieküpsuse otsus sõltub diameetrist enam kui teistel tööühmadel**. Täpsemalt, töögrupil A on väikeste diameetrite korral raieküpsuse määr väiksem ja mingist diameetrist alates suurem. Seega rühmade B ja C puhul on mõnevõrra suurem võimalus (võrreldes rühmaga A), et raieküpseks loetakse ka väiksema diameetriga mände.

Järgnevalt lisame logistilise regressiooni mudelisse ka maakonna ja majanduskategooria tunnused, et näha nende rolli raieküpsuse prognoosimisel (majanduskategooria: 0- tulundusmets; 1- kaitsemets; 2- rangekaitse).

```
glm(formula = y ~ tr_tyyp + toogrupp * inv_diameeter + factor(majandkategooria) +  
factor(qmaakond), family = binomial(), data = filter(raiekypsus_andmed,  
peapuuliik="MA", toogrupp != D))
```

Coefficients	Estimate	Std. Error	z value	Pr(> z)			
(Intercept)	-9,525825	0,705225	13,508	2,00E-16	***		
tr_tyyp2	-0,339607	0,132155	-2,570	0,010177	*		
tr_tyyp3	-0,226240	0,130511	-1,733	0,083008	.		
Toogrupp_B	2,774427	0,774309	3,583	0,00034	***		
Toogrupp_C	3,521559	0,767512	4,588	4,47E-06	***		
inv_diameeter	0,031546	0,002455	12,851	2,00E-16	***		
factor(majandkategoria)1	0,498161	0,129262	3,854	0,000116	***		
factor(majandkategoria)2	1,291870	0,106361	12,146	2,00E-16	***		
factor(qmaakond)0039 Hiiu	-0,689607	0,248464	-2,775	0,005512	**		
factor(qmaakond)0044 Ida-Viru	-0,458524	0,199745	-2,296	0,021702	*		
factor(qmaakond)0049 Jõgeva	0,373693	0,289470	1,291	0,196719			
factor(qmaakond)0051 Järva	0,273569	0,228729	1,196	0,231680			
factor(qmaakond)0057 Lääne	-1,124712	0,289362	-3,887	0,000102	***		
factor(qmaakond)0059 Lääne-Viru	0,263521	0,212663	1,239	0,215290			
factor(qmaakond)0065 Põlva	-0,202148	0,270460	-0,747	0,454808			
factor(qmaakond)0067 Pärnu	-0,791066	0,200358	-3,948	7,87E-05	***		
factor(qmaakond)0070 Rapla	-0,366678	0,231345	-1,585	0,112969			
factor(qmaakond)0074 Saare	-1,112286	0,201570	-5,518	3,43E-08	***		
factor(qmaakond)0078 Tartu	-0,412814	0,288740	-1,430	0,152800			
factor(qmaakond)0082 Valga	-0,271776	0,266527	-1,020	0,307873			
factor(qmaakond)0084 Viljandi	-0,093480	0,260819	-0,358	0,720036			
factor(qmaakond)0086 Võru	-0,052327	0,219286	-0,239	0,811397			
Toogrupp_B:inv_diameeter	-0,009163	0,002761	-3,319	0,000904	***		
Toogrupp_C:inv_diameeter	-0,011598	0,002722	-4,261	2,04E-05	***		
Null deviance:	4651.6	on	4048	degrees	of	freedom	
Residual deviance:	3071.4	on	4025	degrees	of	freedom	
AIC:							3119.4

Viimasena toodud logistilise regressiooni analüüsist nähtub, et majanduskategooriate 1 ja 2 kordajad on positiivsed, mis vastab kaitsealuste metsade oodatult suuremale raieküpsusele võrreldes tulundusmetsaga (mudeli baastase). Ka maakondade vahel on raieküpsuse osas tõendatud erinevusi, millele osundavad Lääne, Pärnu ja Saare maakondade suured negatiivsed kordajad (mürgendiga ***) võrreldes baastasemega ehk Harju maakonnaga. Samas on siiski näha, et majanduskategooria ja maakonnatunnuse lisamine mudelisse ei seleta ära tööruhmade erinevusi raieküpsuse hindamisel, sest endiselt on tööruhmade B ja C kordajad (vastavalt 2,77 ja 3,42) oluliselt suuremad tööruhmast A (baasnivoo).

Kokkuvõte subjektiivsusteguri mõju uurimisest:

Mõõtmiste subjektiivsusteguri uurimisel leiti, et kõigi mõõtmiste juures on subjektiivsusel teatav roll olemas. Instrumentaalsete mõõtmiste korral on subjektiivsusfaktor suhteliselt väike ja seda saab hõlpsasti kvantifitseerida. Hinnanguliste mõõtmiste korral (näiteks puistu vanuse või raieküpsuse määramisel) on subjektiivsusfaktor märkimisväärne ja seda saab neutraliseerida hooaja-eelse **koolitusega**, süstemaatilise **järelokrolli ja tagasisidega** ning mõõtemeeskondade **balansseeritud kasutamise**ga alalistel ja ajutistel proovitükkidel. Järelokroll ja tagasiside annavad tööühmadele võimaluse oma tööd kriitiliselt hinnata ja vältida subjektiivsete kõrvalekallete juurdumist oma hinnangutes. Paljude välisriikide metsainventuuri juures on järelokrolli määrana nimetatud 5% proovitükkidest (Tomppo et al. 2010).

Eesti eripära on territooriumi väiksusest tulenev mõõtemeeskondade paratamatult väike arv (käesoleval ajal 3) võrreldes suuremate naaberriikidega nagu Soome ja Rootsi, kus on kasutusel üle 20 välimõõtmisi teostava meeskonna. Väikese töögruppide arvu korral on aga lõpptulemus tundlik iga üksiku töögrupi 'käekirja' suhtes. Seetõttu on vaja iga üksiku mõõtemeeskonna tööd pidevalt monitoorida (järelmõõtmised), et süstemaatilised mõõtmiserinevused saaksid kiiresti avastatud ja tagasisidestatud meeskondadele. Kirjanduses ei ole andmeid selle kohta, et mõõtemeeskondi oleks individuaalselt kalibreeritud süstemaatiliste mõõteerinevuste kompenseerimiseks, mistõttu seda varianti on raske soovitada.

6. Mõningad mudelitega seotud probleemid

6.1. Statistiliste mudelite ahelkasutuse ohtudest

Alljärgnevalt on kirjeldatud mõningaid ohtusid, mis võivad kaasneda statistiliste mudelite järjestikuse rakendamisega, kus ühest mudelist saadud prognoose kasutatakse teise statistilise mudeli sisendina. SMI näitel on sellise olukorraga tegemist siis, kui kännu diameetri d_0 põhjal prognoositakse rinnasdiameeter d_{13} (selleks on lineaarne mudel M1) ja saadud prognoosi \hat{d}_{13} kasutatakse kõrguse h arvutamise mudeli sisendina (mittelineaarne mudel M2). Näitame, et niiviisi saadud kõrguse prognoos \hat{h} on mittelineaarne mudeli M2 korral süstemaatilise nihkega s.t. prognooside keskmine ei lange kokku tegeliku keskmisega: $E(\hat{h}) \neq Eh$. Õigem oleks kasutada otsemudelit M3, mis arvutab kõrguse otse kännu diameetrist ja millega ei teki süstemaatilist nihet. Esitatud väiteid illustreerime paari simulatsiooni abil.

6.1.1. Süstemaatilised vead (nihked)

Korrektelt hinnatud statistiliste mudelite prognoosid on enamasti nihketa – üksikutel juhtudel võib prognoos olla kas liiga suur või liiga väike, aga paljude prognooside keskmine on ligikaudu samasuur kui tegelike väärtuste keskmine. Kui aga kasutatava mudeli puhul leitakse prognoos argumenttunnuse **mittelineaarne** funktsiooni abil, siis mudelite järjestikusel rakendamisel tekib sisse nihe – prognooside keskmine ei kirjelda enam uuritava tunnuse keskmist. Miks selline nihe tekib? Seletame nihke tekkimist ühe teoreetilise näite abil. Kui teame ringi diameetrit d , siis saame leida ring pindala valemiga $S = \pi d^2/4$. Oletame aga, et ringi diameetrit on mõõdetud (või prognoositud) veaga, $d_{\text{mõõdetud}} = d + \varepsilon$, kus ε kirjeldab mõõtmisviga (või tunnuse väärtuse prognoosimisel tehtavat viga). Isegi kui mõõtmistel pole tehtud süstemaatilist viga, $E(\varepsilon) = 0$, on mõõdetud diameetrite põhjal arvutatud ringide pindalad $S_{\text{mõõdetud}} = \pi d_{\text{mõõdetud}}^2/4$ keskmiselt suuremad kui tegelikud pindalad:

$$\begin{aligned} E(S_{\text{mõõdetud}}) &= E(\pi d_{\text{mõõdetud}}^2/4) = \pi/4 E(d_{\text{mõõdetud}}^2) = \pi/4 E((d + \varepsilon)^2) = \pi/4 E(d^2 + 2d\varepsilon + \varepsilon^2) \\ &= \\ &= \pi/4 E(d^2 + 2d\varepsilon + \varepsilon^2) = \pi d^2/4 + \pi d/4 E(\varepsilon) + \pi/4 E(\varepsilon^2) = S + 0 + \pi \sigma^2/4, \end{aligned}$$

kus σ^2 tähistab mõõtmisvigade ε (ehk prognoosivigade) dispersiooni. Kuna dispersioon on alati positiivne, siis on mõõdetud/prognoositud diameetrite pealt arvutatud ringi pindalad süstemaatiliselt suuremad tegelikest pindaladest (prognoositud diameetrite pealt arvutatud pindalad on keskmiselt suuremad tegelikest pindaladest). Mida suuremad on mõõtmisvead/prognoosivead, seda suurem on ka tekkiv nihe.

6.1.2. Prognoosi täpsus kahaneb

Mudelite järjestikusel kasutamisel saadud prognoosid on sageli ebatäpsemad võrreldes prognoosidega, mis on saadud otse meid huvitavat tunnust prognoosides. Otse – ühe mudeli abil – leitud prognoosid on täpsemad nii seetõttu, et välditakse süstemaatilisi vigu (nihkeid), kui ka selle pärast, et mitme mudeli eraldi hindamisel saavad parameetrite hindamisel tehtud

vead akumul eeruda ja teineteist võimendama hakata. Lisaks kaotame potentsiaalselt mudelite järjestikusel rakendamisel informatsiooni, mis pole otseselt vajalik vahepealsete tunnuste prognoosimiseks, kuid millest oleks kasu lõppeesmärgi prognoosimisel. Täpsuse kadu kirjeldab ka peatüki lõppu lisatud simulatsioon.

6.1.3. Simulatsioon

Järgnevalt illustreerime esitatud väiteid simulatsiooni abil. Simuleerimine olukorda, kus kännu diameetri abil prognoositakse puu diameetrit rinna kõrgusel (mudel M1) ja seejärel kasutatakse prognoositud rinnasdiameetreid puu kõrguse prognoosimisel (mudel M2). Näite aluseks on mändide jaoks hinnatud mudelite 1 ja 2 parameetrid. Antud simulatsiooni korral genereerime andmed täpselt hinnatud mudelite järgi (seega garanteerime, et kasutatud mudelite kuju on õige – mida me pärisandmete puhul kunagi ei tea). Samuti genereerime kõik vaatlused sõltumatutena – seega ei pea me muretsema muidu andmeanalüüsi kummitada võivate ruumiliste korrelatsioonide pärast. Lihtsuse mõttes piirdume praegu ka ühe kindla boniteediga metsadega ($h_{100}=40$).

Vaatlusandmete genereerimine:

Vaatluste arv – hästi suur, lihtsustamaks süstemaatiliste vigade märkamist.

```
n=10000000.
```

Juurekaela läbimõõtude $d0_sim$ genereerimine:

```
set.seed(1)  
d0_sim=runif(n, 15, 50)
```

Rinnasdiameetrite $d13_sim$ genereerimine männi rinnasdiameetri mudeli järgi:

```
d13_sim=-1.069+0.811*d0_sim + rnorm(n, 0, sd=1.6)
```

Alljärgnevalt genereerime männi kõrgusmudeli järgi männipuude kõrgused h_sim :

```
h100=40;  
a1=7.9747; a2=1.1471; a3=-4.6135; b1=2.3337; c1=4.3775; c2=0.0922  
h_sim=1.3+(a1+a2*h100)*(d13_sim/(d13_sim+b1))^(c1+c2*h100)+rnorm(n)
```

Prognooside arvutus:

Kuna hetkel teame täpselt, kuidas andmed on genereeritud, siis võime kasutada ka täpseid mudeleid prognooside leidmisel, kuid nagu peagi selgub, ei ole ka täpseid mudeleid teades lihtne saada juurekaela läbimõõtu teades adekvaatset hinnangut puude kõrgustele, kui prognoosimisel rakendame järjest erinevaid statistilisi mudeleid.

Leiame esmalt juurekaela diameetri $d0_sim$ järgi prognoosi rinnasdiameetrile, $d13_prog$:

```
d13_prog = -1.069+0.811*d0_sim
```

ja seejärel kasutame prognoositud diameetreid $d13_prog$ mändide kõrguste prognoosimiseks:

Versioon 4.0
Kalev Pärna 13.02.2023



$$h_prog = 1.3+(a1+a2*h100)*(d13_prog/(d13_prog+b1))^(c1+c2*h100)$$

Esimene tähelepanek: saadud prognoosid on **nihkega** (keskmiselt on prognoosid 0,048 võrra liiga suured!)

```
t.test(h_prog-h_sim)
```

95 percent confidence interval:

```
0.04724359 0.04933389
```

sample estimates:

```
mean of x
```

```
0.04828874
```

Võrdlus otsemudeliga:

Võrdluseks hindame samade andmete pealt suhteliselt lihtsa otsemudeli juurekaela diameetri põhjal puude kõrguste prognoosimiseks. Antud mudel on üsna lihtne (ja seda saaks ilmselt parandada), kuid see võimaldab siiski üsna adekvaatselt kirjeldada tegelikku seost:

```
library(splines)
```

```
model_otse <- lm(h_sim~bs(d0_sim,6))
```

Leiame prognoosid lihtsat otsemudelit kasutades:

```
h_prog_otse = predict(model_otse, data.frame(d0_sim))
```

Saadud prognoosid on **nihketa** (keskmiselt õiged):

```
t.test(h_prog_otse-h_sim)
```

95 percent confidence interval:

```
-0.001045062 0.001045062
```

sample estimates:

```
mean of x
```

```
-8.14842e-14
```

Samuti on lihtsa otsemudeli prognoosid mõnevõrra **täpsemad** kui kahe täpselt teadaoleva matemaatilise mudeli järjest kasutamisel saadud prognoosid:

MSE(mudelite järjestikune kasutamine) - MSE(prognoosime otse):

```
mean((h_prog-h_sim)**2)-mean((h_prog_otse-h_sim)**2)
```

```
[1] 0.002825165
```

Kokkuvõte:

Mittelineaarsete mudelite ahelkasutus tekitab reeglina süstemaatilisi vigu, mistõttu võimalusel tuleks kasutada otsemudeleid, mis töötavad nihketa ning annavad ka väiksema prognoosivea.

6.2. Mahumudeliga M(A, H100) seotud probleeme

- 1) Mahumudeli M(A, H100) kalibreerimine kogu SMI mõõteajaloo järgi sisaldab teatud ohtu, kuna nii saadud mudel on väga inertne ja ei reageeri vajaliku kiirusega looduslike taustatunnuste (nt kliima) muutustele viimase 5-10 a jooksul. Seepärast võiks liiga vanad andmed, näiteks üle 10 a, mudeli kalibreerimisest jätta välja. Antud mudeliga prognoositud mahud muude tunnuste (näiteks maakondade) lõikes ei pruugi olla usaldusväärsed.
- 2) Küsitav on mahtude imputeerimine kasvukoha proovitükkidele mahumudeli M (A, H100) abil, sest arvutused näitavad, et nii toimides üldhinnangu täpsus ei parane võrreldes juhuga, kus üldhinnang tuletatakse üksnes tagavaraproovitükkide tegelike mahtude baasil. Kasvukohaproovitükkidele prognoositud mahud segavad ka veaarvutust, mis eeldab tegelikke andmeid.
- 3) Tasuks uurida alternatiivseid võimalusi kasvukoha proovitükkidel mahtude prognoosimiseks, kuna see võimaldaks kaasata kasvukoha proovitükkidel olevat infot (kasvõi näiteks pindalade kohta) ja samas tagaks summeruvusnõude täidetuse. Esmapilgul näib reaalne kandidaat sellisele prognoosimudelile olevat iteratiivne imputeerimine, mis prognoosib puuduvaid mahtusid korduvalt, kusjuures igal sammul kasutatakse ära eelmisel sammul tekkinud andmestikku.

7. Kaugseire kasutusvõimalusi SMI-s

7.1. Kaugseire andmete kasutamine ajutiste traktide paigutamiseks

Perspektiivne on kaugseirele tugineva ajutise traktivõrgustiku moodustamine. Üks selline meetod võeti kasutusele Rootsis alates 2018, asendamaks senist ajutiste proovitükkide valikustrateegiat. Meetodi töötasid välja ja katsetasid Rootsi Põllumajandusülikooli teadlased (Grafström et al 2017). Uue ajutiste proovitükkide valikustrateegia kõige olulisem erinevus traditsioonilisest strateegiast on see, et valimi disaini kaasatakse kaugseire andmetes olevad abimuutujad (geograafilised koordinaadid, kõrgus, prognoositud puu kõrgus, prognoositud maht), mis korreleeruvad maapealsete mõõtmistulemustega. Valim moodustatakse nii, et see oleks võimalikult heas kooskõlas abimuutujate jaotusega, mis on täpselt teada. Autorid väidavad, et kaugseire andmete kasutamine traktide paigutamiseks on efektiivsem kui nende hilisem kasutamine mahu ja teiste sihttunnuste hindamisel, samuti lihtsustab see analüüsi ja kokkuvõtete tegemist.

Leiame, et ülalkirjeldatud kaugseirel põhinev valikudisain vajab lähemat uurimist ning samas tuleks arvesse võtta ka Soome teadlaste poolt läbi viidud võrdlevat uurimust, kus seda meetodit võrreldakse süstemaatilise valikudisainiga ja lihtsa juhuvalikuga (Räty et al. 2020). Leitakse, et ühe tunnuse (nt keskmise tagavara) suhtes optimeeritud paigutus ei pruugi olla parim, kui soovitakse täpseid hinnanguid mitme tunnuse jaoks üheaegselt.

7.2. Kaugseire andmete kasutamine hinnangute täpsustamiseks järelstratifikatsiooniteel

Riiklikud metsainventuurid (NFI-d) on loodud selleks, et anda täpset teavet metsaressursside kohta riiklikul ja piirkondlikul tasandil, kuid nõudlus sellise teabe järele on olemas ka väiksemas ruumilises mastaabis. Väikese ala hindamise hõlbustamiseks saab kasutada lisaandmeid, näiteks lidariandmeid (ALS) ja satelliidipilte. Tavaliselt sel juhul kasutatav k-lähima naabri meetod (k-NN) pakub mudelipõhist hinnangut väikeste alade jaoks, kuid keskmise ruutvea nihketa hinnangut on sel juhul võimatu leida.

Järelstratifikatsioon (post-stratification - PS) on alternatiivne lähenemisviis abiteabe kasutamiseks, mis samas võimaldab ka disainipõhist dispersiooni hindamist. Järelstratifikatsiooniteel meetodi idee seisneb selles, et pärast maapealsete andmete kogumist võrreldakse saadud andmete jaotust mõne olulise tunnuse lõikes sellesama tunnuse tegeliku jaotusega, mis on teada (näiteks kaugseire andmete alusel). Seejärel võetakse appi kaalud, mille abil ala- või ülesindatud klasside andmed viiakse vastavusse tegeliku jaotusega. Soome teadlased on selgitanud välja piirid, kui väikese ala jaoks on hetke tehniliste võimaluste ning PS koostöös veel võimalik tuletada piisava täpsusega pindala ja mahu hinnanguid (Haakana jt 2019).

Version 4.0
Kalev Pärna 13.02.2023



Järelstratifikatsioonide efektiivsus on allpool demonstreeritud ka meie poolt läbi viidud simuleerimiskogemuses, kus pindalainangute täpsustamiseks kasutasime kaugseire (ALS) andmeid, mille tulemusel hinnangu täpsus kasvas 20-30%.

8. Täiendavad analüüsid ja arvutuseksperimendid

Siia on koondatud projekti käigus läbi viidud täiendavate analüüside ja arvutuslike eksperimentide kirjeldused, millele on mitmel puhul viidatud juba eespool. Punktis 8.2 toodud SMI simulatsioon annab hea ülevaate erinevate hindamiseetodite täpsusest.

8.1. Mahuhinnangute sõltuvus proovitüki keskkoha valikust

Probleemi selgitus

Eespool nägime, et kui tagavaraproovitükkide raadius 10m asendada raadiusega 7m (või ajutistel proovitükkidel 7m asendada 5m-ga), siis mahuhinnangu täpsus eriti palju ei kannata, mistõttu võib kaaluda klappimisraadiuse vähendamist. Kuid koos raadiuse vähendamisega hakkab suurenema hinnangu sõltuvus proovitüki keskkoha valikust, eeskätt seetõttu, et keskkohat ei tohi jääda puu tüve sisse. Seetõttu pakub huvi uurida, millistes piirides võib muutuda mahuhinnang, kui proovitüki tsentrit nihutada esialgselt tsentrist eemale kuni meie poolt lubatud kauguseni d.

1) Klappimine raadiusega 5m

Uurime esmalt seda, kui palju sõltuvad 5m raadiusega proovitükkidelt arvutatud mahuhinnangud proovitüki keskkoha valikust. Katsetamisel loeme klupitud puude arvutatud mahud õigeaks ja vaatleme ainult arvutatud tulemuste erinevust ning seda aastate 2017-2021 andmetel. Igaks juhuks jätame mahuarvutustest välja liiga suure (kaugemal kui 10 m) kaugusega puud 10-meetrise proovitüki originaalandmetest ning vaatleme samuti ainult elus ja seisvaid (ES) puid, mille diameeter on vähemalt 8cm.

```
# andmete sisselugemine
library(dplyr)

aasta1=2017;aasta2=2021
tmp_prt=prt%>%filter(aasta>=aasta1,aasta<=aasta2,abs(pindala-314)<1,psees=="J")
tmp_prtosa=tmp_prt%>%select(aprovitykk_id)%>%inner_join(prtosa)
## Joining, by = "aproovitykk_id"
tmp_klupp=tmp_prt%>%select(aprovitykk_id)%>%inner_join(klupp)%>%filter(kaugus<=100,arv_d13>=8,arv_
_grp=="ES",!is.na(asimuut))
## Joining, by = "aproovitykk_id"
tmp_klupp=tmp_klupp%>%mutate(x=cos(asimuut/180*pi)*kaugus,y=sin(asimuut/180*pi)*kaugus)

tmp_traktid=distinct(select(tmp_klupp,aproovitykk_id))

uued_min_kaugused=function(uusxy,kaugused0,alampiir=0.9,raadius=50){
  #uusxy on vektor, kaugused0 sisaldava x ja y koordinaate esialgse koord systeemi suhtes
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
min_kaugus_uus=kaugused0%>%mutate(kaugus_uus=sqrt((x-uusxy[1])^2+(y-
uusxy[2])^2))%>%group_by(aproovitykk_id)%>%mutate(minkaugus=min(kaugus_uus-
arv_d13/20))%>%ungroup()%>%filter(round(kaugus_uus,digits=3)<=raadius)%>%group_by(aproovitykk_id,m
inkaugus)%>%summarize(maht=sum(arv_tagavara,na.rm=TRUE)/(pi*(raadius/10)^2)*10000,groups="drop_las
t")%>%right_join(tmp_traktid,by
="aproovitykk_id")%>%mutate(maht=if_else(is.na(maht),0,maht))%>%filter(coalesce(minkaugus,alampiir)>=ala
mpiir)
return(min_kaugus_uus)
}
```

Tekitame andmestiku, kus iga esialgse (10m) proovitüki asemel on 5m proovitükid, mille tsentrid on esialgsest tsentrist kuni 4.5 meetri kaugusel (sammuga 0.5 meetrit).

```
max_kaugus_keskelt=45 #detsimeetrites
samm=5
raadius=50
sobivad_keskkohad=uued_min_kaugused(c(0,0),tmp_klupp,raadius=raadius)%>%mutate(x=0,y=0)
for(x1 in seq(-max_kaugus_keskelt,max_kaugus_keskelt,by=samm)){
  for(y1 in seq(-max_kaugus_keskelt,max_kaugus_keskelt,by=samm)){
    if(x1!=0 | y1!=0 & x1^2+y1^2<=max_kaugus_keskelt^2){

sobivad_keskkohad=rbind(sobivad_keskkohad,uued_min_kaugused(c(x1,y1),tmp_klupp)%>%mutate(x=x1,y=y
1))
  }
}
print(x1)
}
## [1] -45
## [1] -40
## [1] -35
## [1] -30
## [1] -25
## [1] -20
## [1] -15
## [1] -10
## [1] -5
## [1] 0
## [1] 5
## [1] 10
## [1] 15
## [1] 20
## [1] 25
## [1] 30
## [1] 35
## [1] 40
## [1] 45
ekstreemsed=sobivad_keskkohad%>%group_by(aproovitykk_id)%>%summarize(minmaht=min(maht),maxmah
t=max(maht))
mean(ekstreemsed$minmaht)
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
## [1] 42.97231
mean(ekstreemsed$maxmaht)
## [1] 391.2221
ekstreemsed
```

aproovitykk_id	minmaht	maxmaht
111365	0.000000	84.212063
111367	0.000000	1023.646397
111369	0.000000	717.483216
111371	120.919560	1495.674493
111373	121.683503	612.682869
111375	84.148402	611.205911
111377	117.736461	920.068360
111379	224.459399	849.874664
111383	0.000000	52.139159
111385	89.012177	417.304261

...

1-10 of 4,022 rows

Nagu näha, varieeruvad arvutatud mahud üsna suures ulatuses ning ka keskmine maht on väga tugevalt sõltuv sellest, millised keskpunktid on valitud.

```
tegelik=
tmp_klupp%>%group_by(aproovitykk_id)%>%summarize(maht=sum(arv_tagavara,na.rm=TRUE)/(pi*10^2)*10000)
cor(tegelik$maht,ekstreemsed$minmaht)
## [1] 0.6456868
cor(tegelik$maht,ekstreemsed$maxmaht)
## [1] 0.9272702
mean(tegelik$maht)
## [1] 191.4706
```

2) Kluppimine raadiusega 7m

Järgnevalt uurime seda, kui palju sõltuvad 5m raadiusega proovitükkidelt arvutatud mahuhinnangud proovitüki keskkoha valikust.

```
# andmete sisselugemine

aasta1=2017;aasta2=2021
tmp_prt=prt%>%filter(aasta>=aasta1,aasta<=aasta2,abs(pindala-314)<1,psees=="J")
tmp_prtosa=tmp_prt%>%select(aproovitykk_id)%>%inner_join(prtosa)
## Joining, by = "aproovitykk_id"
tmp_klupp=tmp_prt%>%select(aproovitykk_id)%>%inner_join(klupp)%>%filter(kaugus<=100,arv_d13>=8,arv_grp=="ES",!is.na(asimuut))
## Joining, by = "aproovitykk_id"
tmp_klupp=tmp_klupp%>%mutate(x=cos(asimuut/180*pi)*kaugus,y=sin(asimuut/180*pi)*kaugus)
```

```
tmp_traktid=distinct(select(tmp_klupp,aproovitykk_id))
```

```
uued_min_kaugused=function(uusxy,kaugused0,alampiir=0.9,raadius=50){
  #uusxy on vektor, kaugused0 sisaldava x ja y koordinaate esialgse koord systeemi suhtes
  min_kaugus_uus=kaugused0%>%mutate(kaugus_uus=sqrt((x-uusxy[1])^2+(y-
uusxy[2])^2))%>%group_by(aproovitykk_id)%>%mutate(minkaugus=min(kaugus_uus-
arv_d13/20))%>%ungroup()%>%filter(round(kaugus_uus,digits=3)<=raadius)%>%group_by(aproovitykk_id,minkaugus)%
>%summarize(maht=sum(arv_tagavara,na.rm=TRUE)/(pi*(raadius/10)^2)*10000,.groups="drop_last")%>%right_join(tmp_
traktid,by
"aproovitykk_id")%>%mutate(maht=if_else(is.na(maht),0,maht))%>%filter(coalesce(minkaugus,alampiir)>=alampiir)
return(min_kaugus_uus)
}
```

Tekitame andmestiku, kus iga esialgse (10m) proovitüki asemel on 7m proovitükid, mille tsentrid on esialgsest tsentrist kuni 3 meetri kaugusel (sammuga 0.5 meetrit).

```
max_kaugus_keskelt=30 #detsimeetrites
samm=5
raadius=70
sobivad_keskkohad=uued_min_kaugused(c(0,0),tmp_klupp,raadius=raadius)%>%mutate(x=0,y=0)
for(x1 in seq(-max_kaugus_keskelt,max_kaugus_keskelt,by=samm)){
  for(y1 in seq(-max_kaugus_keskelt,max_kaugus_keskelt,by=samm)){
    if((x1!=0 | y1!=0) & x1^2+y1^2<=max_kaugus_keskelt^2){
      sobivad_keskkohad=rbind(sobivad_keskkohad,uued_min_kaugused(c(x1,y1),tmp_klupp,raadius=raadius)%>%
mutate(x=x1,y=y1))
    }
  }
  print(x1)
}
## [1] -30
## [1] -25
## [1] -20
## [1] -15
## [1] -10
## [1] -5
## [1] 0
## [1] 5
## [1] 10
## [1] 15
## [1] 20
## [1] 25
## [1] 30
ekstreemsed=sobivad_keskkohad%>%group_by(aproovitykk_id)%>%summarize(minmaht=min(maht),maxmah
t=max(maht))
mean(ekstreemsed$minmaht)
## [1] 116.9795
mean(ekstreemsed$maxmaht)
## [1] 277.0706
ekstreemsed
```

Versioon 4.0
Kalev Pärna 13.02.2023



aproovitykk_id	minmaht	maxmaht
111365	12.98	47.29
111367	490.34	522.26
111369	284.58	575.64
111371	415.26	1104.49
111373	225.54	532.64
111375	228.32	497.27
111377	327.35	659.79
111379	396.79	630.29
111383	12.97	45.51
111385	205.56	346.98

...
1-10 of 4,022 rows

Nagu näha, varieeruvad arvutatud mahud üsna suures ulatuses ning ka keskmine maht on tugevasti sõltuv sellest, millised keskpunktid on valitud.

```
tegelik=  
tmp_klupp%>%group_by(aproovitykk_id)%>%summarize(maht=sum(arv_tagavara,na.rm=TRUE)/(pi*10^2)*1  
0000)  
cor(tegelik$maht,ekstreemsed$minmaht)  
## [1] 0.8580213  
cor(tegelik$maht,ekstreemsed$maxmaht)  
## [1] 0.9682662  
mean(tegelik$maht)  
## [1] 191.4706
```

Eraldame illustreerimise mõttes välja mõned **ekstreemsed proovitükid**, kus sõltuvus keskkoha valikust on eriti suur.

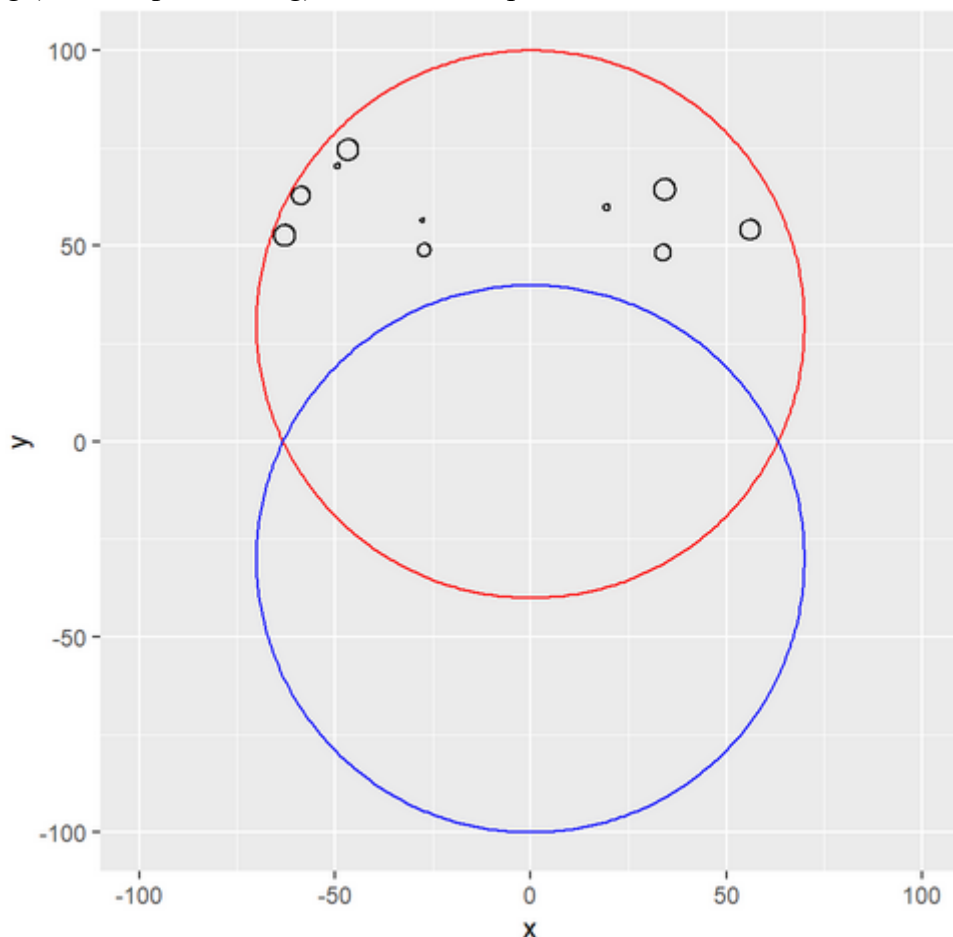
```
ekstreemsed %>% filter(maxmaht-minmaht>1000)
```

aproovitykk_id	minmaht	maxmaht
<int>	<dbl>	<dbl>
115713	391.00	1593.64
117515	0.00	1289.11
126950	336.76	1443.25

3 rows

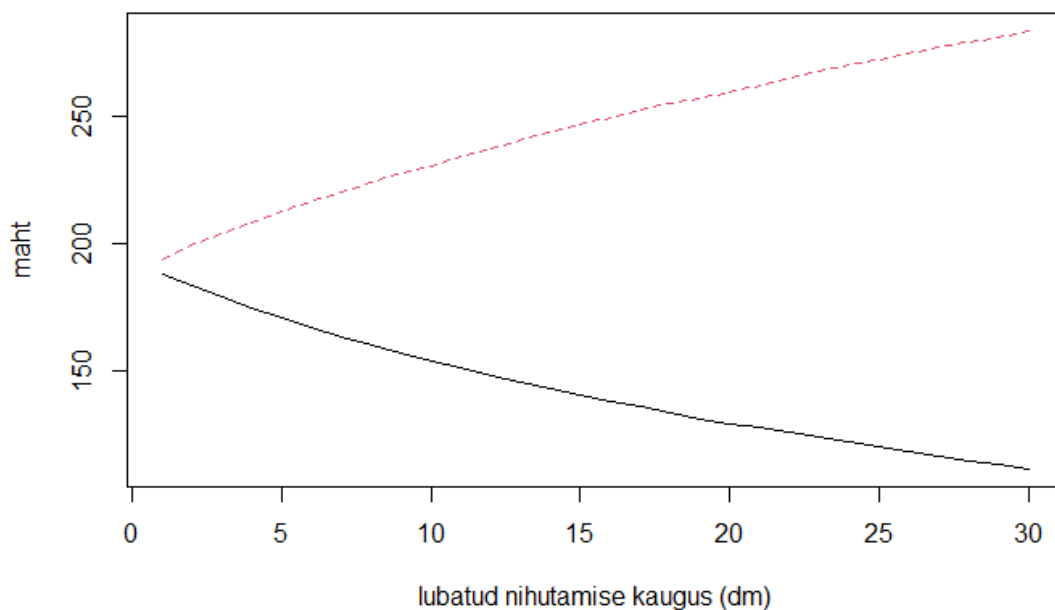
```
ggplot(tmp_klupp%>%filter(aproovitykk_id==117515))+geom_circle(aes(x0=x,y0=y,r=arv_d13/20))+xlim(-  
100,100)+ylim(-  
100,100)+coord_fixed()+geom_circle(aes(x0=0,y0=30,r=70),color="red")+geom_circle(aes(x0=0,y0=-  
30,r=70),color="blue")
```


Alloleval joonisel on näide 10m proovitükist (id 117515), kuhu on võimalik maha märkida 7m raadiusega ring, mis ei sisalda ühtki puud (alumine sinine ring), samas kui teine 7m raadiusega ring (ülemine punane ring) sisaldab kõiki puid.



3) Klappimine raadiusega d

Järgnev graafik näitab kokkuvõtvalt, milline on kõikvõimalike mahuhinnangute ulatus (alumine kõver=min ja ülemine kõver=max), kui tööühmal on vabadus nihutada proovitüki keskkohat originaaltsentrist eemale maksimaalselt x -teljel märgitud kauguse võrra. Näiteks 'lubatud nihutamise kaugus (dm)' väärtus 15 tähendab, et vaatleme kõikvõimalikke proovitükke (vähendatud raadiusega 7m), mille keskkohat pole originaaltsentrist kaugemal kui 15dm, ning leiame mahuhinnangute min ja max üle kõigi selliste proovitükkide. On näha, et lubatud nihutamise kauguse suurenedes min väheneb ja max suureneb, sest proovitükkide hulk, üle mille min ja max arvutatakse, kasvab koos lubatud nihkega.



Kokkuvõte

Ilmneb, et vähendatud raadiusega klappimisel (10 asemel 7m või 5m) on proovitüki keskpunkti valik olulise tähtsusega - erinevad keskpunktid võivad tekitada väga erinevaid mahuhinnanguid. Seega on siin võimalus subjektiivse faktori osatähtsuse tõusuks; näiteks keskkoha süstemaatiline liigutamine hõredama (tihedama) puistu suunas võib anda tugevalt alahinnatud (üle hinnatud) tagavara. Suure klappimisraadiuse korral on tsentri tahtlik kallutamine keerulisem ja subjektiivsuse mõju on seetõttu piiratum. Vähendatud raadiusega klappimisele üleminekul tuleks subjektiivsuse vähendamise eesmärgil välitööde juhendis tsentri valiku reegel täpsemalt fikseerida.

8.2. SMI simuleerimine pindalade hindamismeetodite testimiseks

Simulatsiooni lühikirjeldus: Antud simulatsiooni eesmärk on testida erinevaid hindamise ja veaarvutuse meetodeid. Oletatakse, et maismaa iga 4x4m piksli kohta on teada, kas tegu on metsamaaga või mitte (Metsaregistri andmete järgi). Kasutades selliseid simuleeritud andmeid, mille struktuur sarnaneb Eesti metsade tegelikule seisukorrale ning kus 'tõde' ehk 'tegelikud' väärtused on täpselt teada, testime pindala veaarvutuse erinevaid variante. Simulatsiooni tulemused on esitatud allpool asuvas koondtabelis, kus on lühidalt kirjeldatud ka kasutusel olnud hindamise ja veaarvutuse meetodid.

8.2.1. Statistilise metsainventuuri simuleerimine

Statistiliste uuringute planeerimisel ja kasutatava meetodika töökindluse hindamisel võib kasutada uuringu simuleerimist. Võetakse ette uurimisküsimus, mille puhul on õige vastus (tõde) uurijate eelnevalt teada. Ettenähtud uuringuplaani jälgides - ettenähtud paikades asuvatel traktidel ja proovitükkidel – tehakse virtuaalseid ehk simuleeritud mõõtmisi. Saadud mõõtmistulemusi kasutades hinnatakse huvipakkuva näitaja (või näitajate) väärtused ning kirjeldatakse hinnangute täpsust. Kuna sellisel juhul on tegelik hinnatav väärtus kogu Eesti jaoks teada – siis on võimalik hinnanguid ja tegelikku väärtust võrreldes saada aimu statistiliste hinnangute täpsusest ja ka sellest, kas etteantud arvutuseeskirjade järgi leitud hinnangu täpsused kirjeldavad adekvaatselt tegelikult esinevaid hindamisvigu. Antud simulatsiooni käigus eeldame, et igal proovitükil on võimalik eksimatult ja ilma mõõtmisveeta võimalik määrata meid huvitava tunnuse (kas tegemist on metsamaaga või kui kõrge on mets vms) väärtus. Seega ei ürita antud simulatsioon kirjeldada hindajate subjektiivsusest/kehvast ettevalmistusest või mõõtmiste keerulisusest tekkida võivaid eksimusi. Antud simulatsiooni eesmärgiks on hinnata vaid kasutatud statistilise metodoloogia pädevust.

Uuringu simuleerimisel tuleb kasutada tunnuseid, mille väärtus on kogu Eesti jaoks teada (tõde peab olema teada) ja mille käitumine – näiteks mõõtmistevahelised ruumilised korrelatsioonid – on sarnane tegelikult huvipakkuvatele tunnustele.

8.2.2. Simuleerimisel kasutatavad tunnused

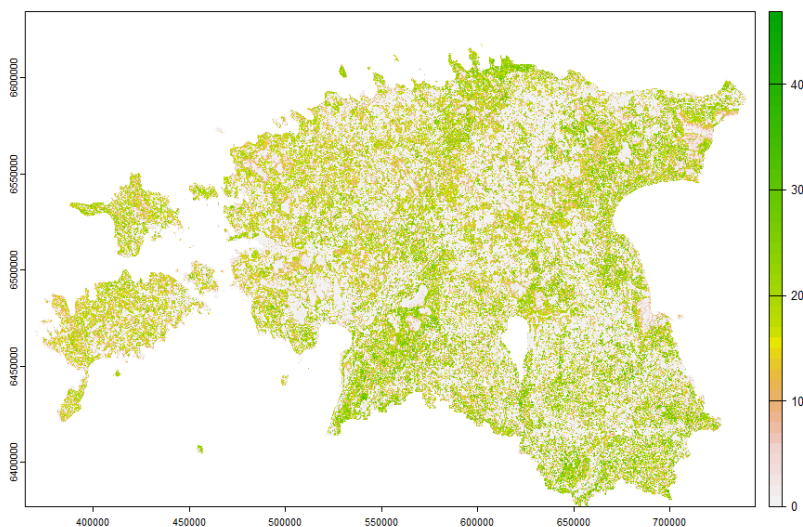
Üheks SMI eesmärgiks on hinnata pindalaid – kui suur osa Eestist on kaetud metsaga, kui suur osa metsamaast on kaetud teatud tüüpi metsaga (näiteks männi enamasuge metsa või raieküpse metsaga kaetud metsamaa suurus). Simulatsioonis hinnatakse SMI katseplaani abil metsaregistrisse 9. novembril 2022 kantud metsamaade pindala (vt. ka Joonis 2. SMI proovitükkidel pannakse simuleeritud uuringu käigus kirja, kas antud proovitükk (proovitüki keskkoh) sattub mõnda metsaregistrisse kantud eraldisse või mitte. Kasutades SMI arvutusvalemiteid, hinnatakse metsamaa kogupindala kõigi maakondade jaoks ja arvutatakse

saadud hinnangute täpsus. Saadud hinnanguid ja nende täpsuse kirjeldusi võrdleme tegelike andmetega – metsaregistrisse kantud eraldiste kogupindalaga vastavas maakonnas.



Joonis 2. Metsaregistris eraldistena arvel olevad alad 9.novembri 2022 seisuga

Lisaks pindaladele hinnatakse statistilise metsainventeerimise abil ka arvuliste tunnuste keskmiseid teatud tüüpi metsamaade jaoks – näiteks metsa hektaritagava (tm/ha), keskmist juurdekasvu vms. Taoliste hinnangute täpsuse uurimiseks kasutame maa-ameti poolt koostatud taimkatte kõrgusmudelit (CHM; Maa-amet; mõõtmised 2020 kevad – 2021 suvi). Maa-ameti poolt koostatud 4 x 4m täpsusega taimkatte kõrgusmudel võimaldab leida taimkatte tegelikku keskmist kõrgust kõigi maakondade metsaalade jaoks.



Joonis 3. Maa-ameti taimkatte kõrgusmudeli andmed

Taimkatte kõrgusandmete kasutamine tagavara hindamisel tugineb loomulikule eeldusele, et taimkatte kõrguste ruumilised korrelatsioonid on sarnased hektaritagavarade ruumilistele korrelatsioonidele (kõrgus on tagavaraga tugevas korrelatsioonis).

8.2.3. Simulatsiooni tulemused ja analüüs

Tabel. Metsamaa pindala hindamise ja veaarvutuse erinevad variandid

	Tegelik viga MAPE (1)	Tegelik viga RMSE (km ²) (2)	Keskmine hinnatud standardviga (km ²) (3)	Standardvea hinnang/ tegelik viga (4) = (3)/(2)	95% usaldusintervalli tegelik katvus (5)
1. Kõik proovitükid, veaarvutus SMI arvutusjuhendis lk 15-16 kirjeldatud viisil	0,79%	204,0	130,1	64%	77%
2. Vaid tagavara proovitükid, veaarvutus SMI arvutusjuhendis lk 15-16 kirjeldatud viisil	0,82%	212,1	184,0	87%	92%
3. Traktikeskmiste kaalutud keskmine – trakte vaadeldakse sõltumatutena	0,79%	204,0	342,0	168%	100%
4. Traktikeskmiste kaalutud keskmine, veaarvutus nelja naabertrakti diagonaalide põhjal (nn 'soome' meetod)	0,79%	204,0	292,9	144%	100%
5. Välise lisainfo (ALS) kasutamine, järelstratifikatsioon, (Haakana et al. 2019a) + 'soome' veahinnang	0,56%	149,4	184,5	123%	99%
Sama mis rida 4., kuid 4 korda rohkem trakte	0,25%	67,4	89,1	132%	98%
Sama mis rida 5., kuid 4 korda rohkem trakte (lisainfo kasutus + järelstratifikatsioon)	0,20%	54,9	68,4	125%	98%

Tabeli veergudes (1) ja (2) on veahinnagute tegelike vigade kaks paralleelset mõõdikut MAPE ja RMSE. Tabeli veerus (3) on keskmine hinnanguviga, mis on leitud suure hulga simulatsioonide keskmisena, kus igal üksikjuhul on kasutatud antud tabelireale vastavat

vearvutuse meetodit. Nii on tabeli 1. reas kasutatud SMI praegust veaarvutuse meetodit ning tulemuseks on tugev vea alahinnang, 130 km² tegeliku 204km² asemel (nende kahe suhe on 64%, mis on toodud veerus (4)). Veerg (5) näitab, mitmel protsendil kõigist simulatsioonidest leitud 95%-line usaldusintervall katab tegeliku pindala (ideaaljuhul on tegelik katvus samuti 95%). Reas 1 on usaldusintervalli katvus üksnes 77%, mis tuleneb vea tugevast alahinnangust.

Tabeli 2. reas jäetakse metsamaa pindala hindamisest katseliselt välja kasvukohaproovitükid ning näeme, et hinnangute täpsus selle all peaaegu ei kannata ning ka veaarvutus SMI meetodil muutub adekvaatsemaks, kuigi endiselt on tegu mõningase alahindamisega.

Alates tabeli reast 3 minnakse üle traktipõhisele hindamisele ja veaarvutusele. Kuna traktikeskmiste kaalutud keskmine (kus trakti kaaluks on maismaaproovitükide arv selles traktis) on sama mis kõigi proovitükide peale saadav keskmine, siis punkthinnangud ja sellest tulenevalt ka MAPE ja RMSE on samad mis reas 1. Keskmine veahinnang on siin leitud traktide sõltumatus eeldusel, mistõttu tulemuseks on suur ülehinnang (342 vs 204). Selline ülehinnang on oodatav, sest me ei arvestanud ruumilisi seoseid traktide vahel.

Tabeli reas 4 on erinevalt eelmisest reast veaarvutuses kasutatud nn 'soome meetodit' (p.4.1.3 valem (1)), mis püüab veadispersiooni vähendada ruumiliste korrelatsioonide arvesse võtmisega. Selle tulemusena keskmine veahinnang paraneb märgatavalt (enne 342, nüüd 292,9), kuid siiski on tegemist konservatiivse hinnanguga.

Olulise pöörde pindala hindamise täpsuses toob lisainformatsiooni kasutamine järelstratifitseerimise kaudu (tabeli rida 5). Lisainfoks on antud juhul võetud ALS taimkatte kõrgusandmed, mis on jaotatud 4 klassi. Rakendades artiklis (Haakana jt 2019 a) kasutatud järelstartifitseerimise skeemi, saadi märgatavalt täpsemad pindalahinnangud (RMSE paranemine 204 -> 149 ehk 27%) ja ka soome veaarvutuse skeem töötas siin tunduvalt efektiivsemalt (üle hinnang nüüd vaid 23%).

Tabeli viimased kaks rida vastavad juhule, kus traktivõrku on mõlemas suunas tihendatud 2 korda, trakte kokku on seega 4 korda rohkem. Näha on ootamatult suur paranemine hinnangute täpsuses ning endiselt suudab järelstartifitseerimine alandada standardviga ca 20%.

Kokkuvõte simulatsioonist:

Antud simulatsioon näitab, et

- 1) SMI praegune veaarvutus hindab pindala määramatust ehk pindalahinnangu viga tugevalt alla ning seega kinnitab meie varasemat hinnangut praegusele meetodile,
- 2) realistlikuma veahinnangu annab 'soome meetod', mis arvestab ruumiliste korrelatsioonidega, kusjuures tegu on konservatiivse veahinnanguga, mis hindab viga mõnevõrra üle,
- 3) pindalahinnangute edasisel parandamisel omab suurt potentsiaali lisainformatsiooni (näiteks kaugseire info või kaardinfo) kasutamine järelstratifitseerimise teel. Läbiviidud simulatsioonikatsed näitas, et see meetod on realselt teostatav.

Versioon 4.0
Kalev Pärna 13.02.2023



8.2.4. Simulatsiooni R-kood

```
# Töökataloogi seadmine
tookataloog = "xxxxxx"
setwd(tookataloog)

if(!require(terra)){
  install.packages("terra")
  library(terra)
}

if(!require(data.table)){
  install.packages("data.table")
  library(data.table)
}

library("RSQLite")

# Kombineerime neljal erineval ajaperioodil tehtud lidari mõõtmistest ühe
# ühise kogu Eestit katva kaardilehe. Tuleks jooksutada vaid üks kord -
# edaspidi kaudu faili taimkate_koos.tif

taimkate1 <- rast("CHM_2021_suvi.tif")
taimkate2 <- rast("CHM_2021_kevad.tif")
taimkate3 <- rast("CHM_2020_suvi.tif")
taimkate4 <- rast("CHM_2020_kevad.tif")

taimkate=merge(taimkate1, taimkate2)
taimkate=merge(taimkate, taimkate3)
taimkate=merge(taimkate, taimkate4)

plot(taimkate)

writeRaster(taimkate, paste(tookataloog, "taimkate_koos", sep=""), format="GTiff")

# Loeme sisse taimkatte kõrgusandmed
taimkate <- rast("taimkate_koos.tif")

# Metsaala panek ühte kihti üle kogu Eesti. Metsaregister eeldatakse olevat alamkataloogis metsregister.

nimekiri=list.dirs(paste(tookataloog, "metsaregister/", sep=""), full.names=FALSE)
nimekiri=nimekiri[nimekiri!=""]

mkond=nimekiri[1]
mets_raster =rast(paste(tookataloog, "raster", mkond, ".tif", sep=""))

for (z in 2:n_maakondi){
  mkond=nimekiri[z]
  print(mkond)
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
mets_raster_lisa=rast(paste(tookataloog, "raster", mkond, ".tif", sep=""))
mets_raster=merge(mets_raster, mets_raster_lisa)
}
plot(mets_raster)

writeRaster(mets_raster, "metsala_koos.tif")

mets_raster=rast("metsala_koos.tif")
plot(mets_raster)

taimkate=rast("taimkate_koos.tif")

# Välise lisainformatsiooni kasutamine stratifitseerimise abil.
# Järelstratifitseerimiseks kasutame taimkatte kõrgust. Kihid: -Inf..1m; 1m...3m; 3m..10m; 10m..Inf.
# Järelstratifitseerimiseks peame teadma kõigi kihtide tegelikke pindalasisid. Arvutame need:

# Teeme stratifitseerimise taimkatte kõrguse järgi
taimkatteklass <- classify(taimkate, cbind(c(-Inf, 1, 3, 10), c(1,3,10, Inf), c(1,2,3,4)))
sagedustabel = freq(taimkatteklass)
sagedustabel

# Eesti pindala, ilma siseveteta, m2, ligikaudu (eeldades, et kõik pikslid on sama suured - mida nad pole)
eesti_pindala=sum(sagedustabel)*16

# Ah - näitab kõigi kihtide pindalasisid ruutmeetrites (ühe piksli suurus on 4m x 4m = 16 m2):
Ah = sagedustabel$count*16

tegelik_metsamaa_pindala=global(mets_raster, "sum", weights=!is.na(taimkate), na.rm=TRUE)[[1]]*16

sammXsuunas=5000
sammYsuunas=5000

# Simulatsioonide arv:
nsim=10000

# Praegu SMIs kasutatavad hinnangud salvestatakse siia
pindala_hinnang1=rep(NA, nsim)
se1=rep(NA, nsim)
piir_al_1=rep(NA, nsim)
piir_yl_1=rep(NA, nsim)

# SMIs kasutatavad hinnangud, aga vaid tagavaraproovitükke kasutades - tulemused salvestatakse siia:
pindala_hinnang2=rep(NA, nsim)
se2=rep(NA, nsim)
piir_al_2=rep(NA, nsim)
piir_yl_2=rep(NA, nsim)

# Traktide keskmisi sõltumatutena käsitledes saadud hinnangud salvestatakse siia
pindala_hinnang3=rep(NA, nsim)
se3=rep(NA, nsim)
piir_al_3=rep(NA, nsim)
piir_yl_3=rep(NA, nsim)
```


Versioon 4.0
Kalev Pärna 13.02.2023



```
# Traktid sõltumatud, kaalutud keskmine (kaalud kasutatud proovitykkide arvu järgi)
pindala_hinnang4=rep(NA, nsim)
se4=rep(NA, nsim)
piir_al_4=rep(NA, nsim)
piir_yl_4=rep(NA, nsim)

# Järelstratifikatsioon + soomlaste veahinnang
pindala_hinnang5=rep(NA, nsim)
se5=rep(NA, nsim)
piir_al_5=rep(NA, nsim)
piir_yl_5=rep(NA, nsim)

# Tavaline kaalutud keskmine - Soomlaste veavalem
pindala_hinnang6=rep(NA, nsim)
se6=rep(NA, nsim)
piir_al_6=rep(NA, nsim)
piir_yl_6=rep(NA, nsim)

se6a=rep(NA, nsim)
piir_al_6a=rep(NA, nsim)
piir_yl_6a=rep(NA, nsim)

# Igaks juhaks salvestame ka võrgustiku nihked, mida kasutatakse:
nihevekX=rep(NA, nsim)
nihevekY=rep(NA, nsim)

# Simulatsioon
for (i_sim in 1:nsim){

print(paste(i_sim, "/", nsim))

# Genereerime juhuslikud nihked:
niheX=runif(1, 0, sammXsuunas)
niheY=runif(1, 0, sammYsuunas)
nihevekX[i_sim]=niheX
nihevekY[i_sim]=niheY

# Tekitame regulaarse võrgustiku:
# max(prt$koord_e)
algX=seq(ext(taimkate)[1]-niheX-sammXsuunas, ext(taimkate)[2]+sammXsuunas, sammXsuunas)
algY=seq(ext(taimkate)[3]-niheY-sammYsuunas, ext(taimkate)[4]+sammYsuunas, sammYsuunas)

# Kui suur on meie võrgustik:
nX=length(algX)
nY=length(algY)

xvek=rep(NA, nX*nY*16)
yvek=rep(NA, nX*nY*16)
tyyp=rep(NA, nX*nY*16)
trakt_aid = rep(NA, nX*nY*16)

# Tsükkel üle kõigi traktide - tsükli sees tekitame proovitükid
# tyyp=1 on tagavara proovitükk, tyyp=2 - kasvukoha proovitykk.
nr=1
```

```
for (i in 1:nY){
  for (j in 1:nX){
    trakt_aid[nr:(nr+15)]=paste("X_",j,"_Y_",i,sep="")

    xvek[ nr ] = algX[j]
    yvek[ nr ] = algY[i]
    tyyp[ nr ] = 1

    xvek[ nr+1 ] = algX[j]+200
    yvek[ nr+1 ] = algY[i]
    tyyp[ nr+1 ] = 2

    xvek[ nr+2 ] = algX[j]+400
    yvek[ nr+2 ] = algY[i]
    tyyp[ nr+2 ] = 1

    xvek[ nr+3 ] = algX[j]+600
    yvek[ nr+3 ] = algY[i]
    tyyp[ nr+3 ] = 2

    xvek[ nr+4 ] = algX[j]+800
    yvek[ nr+4 ] = algY[i]
    tyyp[ nr+4 ] = 1

    xvek[ nr+5 ] = algX[j]+800
    yvek[ nr+5 ] = algY[i]+200
    tyyp[ nr+5 ] = 2

    xvek[ nr+6 ] = algX[j]+800
    yvek[ nr+6 ] = algY[i]+400
    tyyp[ nr+6 ] = 1

    xvek[ nr+7 ] = algX[j]+800
    yvek[ nr+7 ] = algY[i]+600
    tyyp[ nr+7 ] = 2

    xvek[ nr+8 ] = algX[j]+800
    yvek[ nr+8 ] = algY[i]+800
    tyyp[ nr+8 ] = 1

    xvek[ nr+9 ] = algX[j]+600
    yvek[ nr+9 ] = algY[i]+800
    tyyp[ nr+9 ] = 2

    xvek[ nr+10 ] = algX[j]+400
    yvek[ nr+10 ] = algY[i]+800
    tyyp[ nr+10 ] = 1

    xvek[ nr+11 ] = algX[j]+200
    yvek[ nr+11 ] = algY[i]+800
    tyyp[ nr+11 ] = 2

    xvek[ nr+12 ] = algX[j]
    yvek[ nr+12 ] = algY[i]+800
    tyyp[ nr+12 ] = 1
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
xvek[ nr+13 ]=algX[j]
yvek[ nr+13 ]=algY[i]+600
tyyp[nr+13] = 2

xvek[ nr+14 ]=algX[j]
yvek[ nr+14 ]=algY[i]+400
tyyp[nr+14] = 1

xvek[ nr+15 ]=algX[j]
yvek[ nr+15 ]=algY[i]+200
tyyp[nr+15] = 2

nr=nr+16
}
}

# Loeme kõigi traktide jaoks proovitükkide mõõtmistulemused kaardilt.
val_h <-extract(x=taimkate, y=cbind(lon=xvek, lat=yvek))$layer
val_mets <- extract(x=mets_raster, y=cbind(lon=xvek, lat=yvek))$layer
val_mets[is.na(val_mets)]=0

# "Valiidseteks" proovitükkideks loeme vaid neid, mis sattuvad maismaale ja Eesti riigi territooriumile. Eestile kuuluva
maismaala määrame
# selle järgi, et taimkatte kõrgusmudelil on olemas reaalne väärtus (praktildas võib esineda segadus sellega, et mõnede linnades
paiknevate alade -
# äkki parkimisplatsid või midagi muud sellist - jaoks pole ka taimkatte kõrguse mõõtmisi tehtud.
ind= (!is.na(val_h))
# ind_tagavara = (prt_koos$aasta>=2017 & prt_koos$pees=="J" & prt_koos$prt_liik==1)
# Tagavara proovitükkide indeks.
ind_tagavara = (!is.na(val_h) & tyyp==1)

# Pindala hindamisel kasutatav proovitükkide koguarv
nn=sum(ind)
nn_tagavara=sum(ind_tagavara)

# Leiame koondnumbrid iga trakti jaoks

prt_sim = data.frame(trakt_aid,val_h, val_mets, tyyp, xvek, yvek)[ind,]
prt_sim_DT=data.table(prt_sim)

prt_sim_DT$ind_tagavara=ind_tagavara[ind]

yks_per_trakt_sim = prt_sim_DT[, .(proovitykke=.N, tagavara=sum(ind_tagavara==1), metsas=sum(val_mets>0),
  n1=sum(val_h<=1), n2=sum(val_h>1 & val_h<=3), n3=sum(val_h>3 & val_h<=10), n4=sum(val_h>10),
  n1m=sum(val_h<=1 & val_mets>0), n2m=sum(val_h>1 & val_h<=3 & val_mets>0),
  n3m=sum(val_h>3 & val_h<=10 & val_mets>0), n4m=sum(val_h>10 & val_mets>0),
  y=mean(val_h), mets_y=mean(val_mets*val_h), koord_n_tipp=min(yvek), koord_e_tipp=min(xvek),
  y_tykke=sum(!is.na(val_h)) ), by=trakt_aid]

# Määrame ruudustiku (sammXsuunas m x sammYsuunas m) kus iga trakt on antud ruudu unikaalne esindaja.
vorm_sim <- rast(xmin=min(yks_per_trakt_sim$koord_e_tipp)-sammXsuunas/2,
xmax=max(yks_per_trakt_sim$koord_e_tipp)+sammXsuunas/2,
```

Versioon 4.0
Kalev Pärna 13.02.2023



```

      ymin=min(yks_per_trakt_sim$koord_n_tipp)-sammYsuunas/2,
      ymax=max(yks_per_trakt_sim$koord_n_tipp)+sammYsuunas/2, res=c(sammXsuunas, sammYsuunas))

# Unikaalsuse kontroll:
#   plot(rasterize(cbind(yks_per_trakt_sim$koord_e_tipp,      yks_per_trakt_sim$koord_n_tipp),      vorm_sim,
value=yks_per_trakt_sim$proovitykke, fun=length))

# Hindame pindala SMI arvutusjuhendis lk 14-16 kirjeldatud viisil
# -----

nvalim=sum(val_mets[ind])
ke=nvalim/nn
pp=(nvalim+1.92)/(nn+3.84)
se=sqrt(pp*(1-pp)/nn)
sum_pindala=(ke*eesti_pindala)

piir_al_1[i_sim]=((pp-1.96*se)/pp*sum_pindala)
piir_yl_1[i_sim]=((pp+1.96*se)/pp*sum_pindala)

pindala_hinnang1[i_sim]=sum_pindala
se1[i_sim]=se*eesti_pindala

# Hindame pindala, SMI arvutusjuhendi järgi aga ainult tagavara proovitükke kasutades
# -----

nvalim_tagavara=sum(val_mets[ind_tagavara])
ke_tagavara=nvalim_tagavara/nn_tagavara
pp_tagavara=(nvalim_tagavara+1.92)/(nn_tagavara+3.84)
se_tagavara=sqrt(pp_tagavara*(1-pp_tagavara)/nn_tagavara)
sum_pindala_tagavara=(ke_tagavara*eesti_pindala)

piir_al_2[i_sim]=((pp_tagavara-1.96*se_tagavara)/pp_tagavara*sum_pindala_tagavara)
piir_yl_2[i_sim]=((pp_tagavara+1.96*se_tagavara)/pp_tagavara*sum_pindala_tagavara)

pindala_hinnang2[i_sim]=sum_pindala_tagavara
se2[i_sim]=se_tagavara*eesti_pindala

# Käsitleme trakte sõltumatutena, enne analüüsi agregeeritakse trakti andmed kokku - hinnanguks on traktide keskmiste
keskmise
# -----

keskmineosakaal_per_trakt = yks_per_trakt_sim$metsas/yks_per_trakt_sim$proovitykke

abi=t.test(keskmineosakaal_per_trakt)

piir_al_3[i_sim]=abi$conf.int[1]*eesti_pindala
piir_yl_3[i_sim]=abi$conf.int[2]*eesti_pindala

pindala_hinnang3[i_sim]= abi$estimate*eesti_pindala
se3[i_sim]=abi$stderr*eesti_pindala

# Käsitleme trakte sõltumatutena - hinnanguks on traktide keskmiste kaalutud keskmine (vaid pooleldi Eestis olevaid trakte
võtame väiksema kaaluga)
# -----
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
kaalud=yks_per_trakt_sim$proovitykke/16
kaalud=kaalud/sum(kaalud)
osakaalu_hinnang = sum(keskmineosakaal_per_trakt*kaalud)
pindala_hinnang4[i_sim]= osakaalu_hinnang*eesti_pindala

# Sõltumatuse korral
#  $D(w_1 Y_1 + \dots w_n Y_n) = w_1^2 D(Y_1) + \dots w_n^2 D(Y_n) = (w_1^2 + \dots w_n^2) * D(Y)$ 

se_kordaja = sqrt(sum(kaalud**2))
se4[i_sim] = se_kordaja*sd(keskmineosakaal_per_trakt)*eesti_pindala

piir_al_4[i_sim]=osakaalu_hinnang*eesti_pindala-1.96*se4[i_sim]
piir_yl_4[i_sim]=osakaalu_hinnang*eesti_pindala+1.96*se4[i_sim]

# Välise lisainformatsiooni kasutamine - poststratifikatsioon lidari abil.
# vaata kirjeldust: Annals of Forest Science (2019) 76: 9; https://doi.org/10.1007/s13595-018-0795-6
# -----

# Kihtide tegelikud pindalad on Ah
abi_sagedused=table(cut(val_h, c(-Inf,1,3,10, Inf)), val_mets)
metsa_osakaalud_kihis = abi_sagedused[,2]/rowSums(abi_sagedused)
pindala_hinnang5[i_sim]=Ah%%metsa_osakaalud_kihis
z_ch=cbind(yks_per_trakt_sim$n1m, yks_per_trakt_sim$n2m, yks_per_trakt_sim$n3m, yks_per_trakt_sim$n4m)-
        cbind(metsa_osakaalud_kihis[1]*yks_per_trakt_sim$n1,
              metsa_osakaalud_kihis[2]*yks_per_trakt_sim$n2,
              metsa_osakaalud_kihis[3]*yks_per_trakt_sim$n3,
              metsa_osakaalud_kihis[4]*yks_per_trakt_sim$n4)

abikaart1=rasterize(cbind(yks_per_trakt_sim$koord_e_tipp, yks_per_trakt_sim$koord_n_tipp), vorm_sim, value=z_ch[,1],
fun=sum, background=0)
abikaart2=rasterize(cbind(yks_per_trakt_sim$koord_e_tipp, yks_per_trakt_sim$koord_n_tipp), vorm_sim, value=z_ch[,2],
fun=sum, background=0)
abikaart3=rasterize(cbind(yks_per_trakt_sim$koord_e_tipp, yks_per_trakt_sim$koord_n_tipp), vorm_sim, value=z_ch[,3],
fun=sum, background=0)
abikaart4=rasterize(cbind(yks_per_trakt_sim$koord_e_tipp, yks_per_trakt_sim$koord_n_tipp), vorm_sim, value=z_ch[,4],
fun=sum, background=0)

vead1=matrix(nrow=dim(abikaart1)[1]-1, ncol=dim(abikaart1)[2]-1)
for (i in 1:(dim(abikaart1)[1]-1)){
  x11=unlist(abikaart1[i,])[-dim(abikaart1)[2]];
  x22=unlist(abikaart1[i+1,])[-1];
  x12=unlist(abikaart1[i,])[-1];
  x21=unlist(abikaart1[i+1,])[-dim(abikaart1)[2]];
  vead1[i,]=((x11+x22)/2-(x12+x21)/2)
}
vead2=matrix(nrow=dim(abikaart2)[1]-1, ncol=dim(abikaart2)[2]-1)
for (i in 1:(dim(abikaart2)[1]-1)){
  x11=unlist(abikaart2[i,])[-dim(abikaart2)[2]];
  x22=unlist(abikaart2[i+1,])[-1];
  x12=unlist(abikaart2[i,])[-1];
  x21=unlist(abikaart2[i+1,])[-dim(abikaart2)[2]];
  vead2[i,]=((x11+x22)/2-(x12+x21)/2)
}
vead3=matrix(nrow=dim(abikaart3)[1]-1, ncol=dim(abikaart3)[2]-1)
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
for (i in 1:(dim(abikaart3)[1]-1)){
  x11=unlist(abikaart3[i,])[-dim(abikaart3)[2]];
  x22=unlist(abikaart3[i+1,])[-1];
  x12=unlist(abikaart3[i,])[-1];
  x21=unlist(abikaart3[i+1,])[-dim(abikaart3)[2]];
  vead3[i,]=((x11+x22)/2-(x12+x21)/2)
}
vead4=matrix(nrow=dim(abikaart4)[1]-1, ncol=dim(abikaart4)[2]-1)
for (i in 1:(dim(abikaart4)[1]-1)){
  x11=unlist(abikaart4[i,])[-dim(abikaart4)[2]];
  x22=unlist(abikaart4[i+1,])[-1];
  x12=unlist(abikaart4[i,])[-1];
  x21=unlist(abikaart4[i+1,])[-dim(abikaart4)[2]];
  vead4[i,]=((x11+x22)/2-(x12+x21)/2)
}

veamat=data.frame(as.vector(vead1), as.vector(vead2), as.vector(vead3), as.vector(vead4))
abif=function(a1, a2) {
Ah[a1]*Ah[a2]*sum(veamat[,a1]*veamat[,a2])/(rowSums(abi_sagedused)[a1]*rowSums(abi_sagedused)[a2])

se5[i_sim] = sqrt(abif(1,1)+abif(1,2)+abif(1,3)+abif(1,4)+abif(2,1)+abif(2,2)+abif(2,3)+abif(2,4)+
  abif(3,1)+abif(3,2)+abif(3,3)+abif(3,4)+abif(4,1)+abif(4,2)+abif(4,3)+abif(4,4))

piir_al_5[i_sim]=pindala_hinnang5[i_sim]-1.96*se5[i_sim]
piir_yl_5[i_sim]=pindala_hinnang5[i_sim]+1.96*se5[i_sim]

# Vigade hindamisel vaatame erinevusi lokaalsest keskmisest - diagonaalides leitud hinnangute erinevus teineteisest
# -----

metsa_osakaalud_kihis1=sum(yks_per_trakt_sim$metsas)/sum(yks_per_trakt_sim$proovitykke)
z_ch_1 = yks_per_trakt_sim$metsas-metsa_osakaalud_kihis1*yks_per_trakt_sim$proovitykke

# Paigutame vead 5km x 5km ruudustikku
abikaart=rasterize(cbind(yks_per_trakt_sim$koord_e_tipp, yks_per_trakt_sim$koord_n_tipp), vorm_sim, value=z_ch_1,
fun=sum, background=0)
# plot(abikaart)

# Hindame prognoosimisel tekkivat viga
vead=matrix(nrow=dim(abikaart)[1]-1, ncol=dim(abikaart)[2]-1)
for (i in 1:(dim(abikaart)[1]-1)){
  x11=unlist(abikaart[i,])[-dim(abikaart)[2]];
  x22=unlist(abikaart[i+1,])[-1];
  x12=unlist(abikaart[i,])[-1];
  x21=unlist(abikaart[i+1,])[-dim(abikaart)[2]];
  vead[i,]=((x11+x22)/2-(x12+x21)/2)
}

se6[i_sim] = eesti_pindala*sqrt(sum(vead^2)/(sum(yks_per_trakt_sim$proovitykke)^2))

piir_al_6[i_sim]=osakaalu_hinnang*eesti_pindala-1.96*se6[i_sim]
piir_yl_6[i_sim]=osakaalu_hinnang*eesti_pindala+1.96*se6[i_sim]

if ((i_sim%%1000)==0){
  print("MAPE:")
}
```

```
print(mean(abs(pindala_hinnang1[1:i_sim]-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala))
print(mean(abs(pindala_hinnang2[1:i_sim]-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala))
print(mean(abs(pindala_hinnang3[1:i_sim]-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala))
print(mean(abs(pindala_hinnang4[1:i_sim]-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala))
print(mean(abs(pindala_hinnang5[1:i_sim]-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala))
print(mean(abs(pindala_hinnang4[1:i_sim]-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala))

disp1=mean(((pindala_hinnang1[1:i_sim]-tegelik_metsamaa_pindala)/1000000)**2)
disp2=mean(((pindala_hinnang2[1:i_sim]-tegelik_metsamaa_pindala)/1000000)**2)
disp3=mean(((pindala_hinnang3[1:i_sim]-tegelik_metsamaa_pindala)/1000000)**2)
disp4=mean(((pindala_hinnang4[1:i_sim]-tegelik_metsamaa_pindala)/1000000)**2)
disp5=mean(((pindala_hinnang5[1:i_sim]-tegelik_metsamaa_pindala)/1000000)**2)
disp6=mean(((pindala_hinnang4[1:i_sim]-tegelik_metsamaa_pindala)/1000000)**2)

print("se üle/alahinnang")
print( mean(se1[1:i_sim]/1000000) /sqrt(disp1) )
print( mean(se2[1:i_sim]/1000000) /sqrt(disp2) )
print( mean(se3[1:i_sim]/1000000) /sqrt(disp3) )
print( mean(se4[1:i_sim]/1000000) /sqrt(disp4) )
print( mean(se5[1:i_sim]/1000000) /sqrt(disp5) )
print( mean(se6[1:i_sim]/1000000) /sqrt(disp6) )
}
}

#MAPE väärtused erinevate meetodite jaoks
mean(abs(pindala_hinnang1-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala)
mean(abs(pindala_hinnang2-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala)
mean(abs(pindala_hinnang3-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala)
mean(abs(pindala_hinnang4-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala)
mean(abs(pindala_hinnang5-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala)
mean(abs(pindala_hinnang4-tegelik_metsamaa_pindala)/tegelik_metsamaa_pindala)
# Kuna 6. meetod kasutab sama keskmise hinnangut mis 4. meetod, siis MAPE arvutamisel kasutame 4. hinnangut ka 6. meetodi jaoks

# Usalduspiiride tegelikud katvused:
mean( piir_al_1< tegelik_metsamaa_pindala & piir_yl_1>tegelik_metsamaa_pindala)
mean( piir_al_2< tegelik_metsamaa_pindala & piir_yl_2>tegelik_metsamaa_pindala)
mean( piir_al_3< tegelik_metsamaa_pindala & piir_yl_3>tegelik_metsamaa_pindala)
mean( piir_al_4< tegelik_metsamaa_pindala & piir_yl_4>tegelik_metsamaa_pindala)
mean( piir_al_5< tegelik_metsamaa_pindala & piir_yl_5>tegelik_metsamaa_pindala)
mean( piir_al_6< tegelik_metsamaa_pindala & piir_yl_6>tegelik_metsamaa_pindala)

# Standardvigade üle- või alahinnangute suurused.
# Esmalt leiame tegelikud vea suurused (pindalad siin mõõdetud km^2)
disp1=mean(((pindala_hinnang1-tegelik_metsamaa_pindala)/1000000)**2)
disp2=mean(((pindala_hinnang2-tegelik_metsamaa_pindala)/1000000)**2)
disp3=mean(((pindala_hinnang3-tegelik_metsamaa_pindala)/1000000)**2)
disp4=mean(((pindala_hinnang4-tegelik_metsamaa_pindala)/1000000)**2)
disp5=mean(((pindala_hinnang5-tegelik_metsamaa_pindala)/1000000)**2)
disp6=mean(((pindala_hinnang4-tegelik_metsamaa_pindala)/1000000)**2)
```

Versioon 4.0
Kalev Pärna 13.02.2023



```
# Tegelik vea suuruse hinnangud (km2)
```

```
sqrt(displ)  
sqrt(displ2)  
sqrt(displ3)  
sqrt(displ4)  
sqrt(displ5)  
sqrt(displ6)
```

```
# Keskmised standardvea hinnangud
```

```
h_displ = mean(se1/1000000)  
h_displ2 = mean(se2/1000000)  
h_displ3 = mean(se3/1000000)  
h_displ4 = mean(se4/1000000)  
h_displ5 = mean(se5/1000000)  
h_displ6 = mean(se6/1000000)
```

```
h_displ  
h_displ2  
h_displ3  
h_displ4  
h_displ5  
h_displ6
```

```
# Standardvea üle-või alahinnangu suurused
```

```
mean(se1/1000000) /sqrt(displ)  
mean(se2/1000000) /sqrt(displ2)  
mean(se3/1000000) /sqrt(displ3)  
mean(se4/1000000) /sqrt(displ4)  
mean(se5/1000000) /sqrt(displ5)  
  
mean(se6/1000000) /sqrt(displ6)
```


9. Kasutatud kirjanduse loetelu

1. Cochran W. G. (1977) Sampling techniques, 3rd edition. Wiley, New York. 428 p.
2. Fridman J., Holm S., Nilsson M., Nilsson P., Ringvall A.H., Ståhl G. (2014). Adapting National Forest Inventories to changing requirements – the case of the Swedish National Forest Inventory at the turn of the 20th century. *Silva Fennica*, vol. **48** no.3 article id 1095. <https://doi.org/10.14214/sf.1095>
3. Goodman I. A. (1960). On the exact variance of products. *J. Am. Statist. Assoc.* **55**, 708-713.
4. Grafström A., Zhao X., Nylander M., and Petersson H. (2017). A new sampling strategy for forest inventories applied to the temporary clusters of the Swedish NFI. *Canadian Journal of Forest Research*, 47(9)
5. Grafström A., Schelin L. (2014) How to select representative samples. *Scand J Stat* 41(2):277–290. <https://doi.org/10.1111/sjos.12016>
6. Haara, A. & Korhonen, K.T. (2004). Kuvioittaisen arvioinnin luotettavuus. *Metsätieteen aikakauskirja*, 4/2004: 489–508.
7. Haakana H., Heikkinen J., Katila M., Kangas A. (2019 a). Efficiency of post-stratification for a large-scale forest inventory—case Finnish NFI. *Annals of Forest Science*, vol.**76**, Article number 9. <https://doi.org/10.1007/s13595-018-0795-6>
8. Haakana H., Heikkinen J., Katila M., Kangas A. (2019 b). Precision of exogenous post-stratification in small-area estimation based on a continuous national forest inventory. *Canadian Journal of Forest Research*. Vol. **50**, 4, <https://doi.org/10.1139/cjfr-2019-0139>
9. Heikkinen J. (2006). Assessment of uncertainty in spatially systematic sampling.- In: A.Kangas and M. Maltamo (eds.), *Forest Inventory - Methodology and Applications*. Springer, pp.155-176.
10. Kangas A. & Maltamo M. (Eds.) (2006). *Forest Inventory. Methodology and Applications* (2006). Springer.
11. Korhonen K. T., Ahola A., Heikkinen J., Henttonen H. M., Hotanen J.-P., Ihalainen A., Melin M., Pitkänen J., Rätty M., Sirviö M., Strandström M. (2021). Forests of Finland 2014–2018 and their development 1921–2018. *Silva Fennica* vol. 55 no. 5 article id 10662. <https://doi.org/10.14214/sf.10662>
12. Kujala M. (1979). Mättnogrannheten vid riksskogstaxeringen. *Finnish For. Res. Inst.*
13. Lindeberg J.W. (1924) Calculating the standard error in the strip-survey results (in German). *Acta Forest Fenn* 25(5): article 7080. <https://doi.org/10.14214/aff.7080>
14. Matérn B. (1947) Methods of estimating the accuracy of line and sample plot surveys. *Meddelanden från Statens Skogsforskningsinstitut* 36(1):138

15. Matérn B. (1960) Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden från Statens Skogsforskningsinstitut* 49(5): 144
16. Mäkisara K.; Katila M., Peräsaari, J. (2019) The Multi-Source National Forest Inventory of Finland – methods and results 2015. LUKE.
<http://urn.fi/URN:ISBN:978-952-326-711-417>.
17. Persson H.J., Olofsson K., Holmgren J. (2022). Two-phase forest inventory using very-high-resolution laser scanning. *Remote Sensing of Environment*, v.271, 2022, 112909.
18. Ranney B., Cruse T., Hägglund B., Jonasson H., Swärd J. (1987). Designing a new national forest survey for Sweden. *Studia Forestalia Suecica* 177. 29 p.
19. Rätty M., Kuronen M., Myllymäki M., Kangas A., Mäkisara K. & Heikkinen, J. (2020). Comparison of the local pivotal method and systematic sampling for national forest inventories. *Forest Ecosystems*. Vol. 7, Article number: 54.
20. Sims A. (2022) *Principles of National Forest Inventory Methods. Theory, Practice, and Examples from Estonia*. Springer.
21. Statistiline metsainventeerimine. Arvutusmetoodiline juhend. 2018.
22. Tomppo E., Gschwantner Th., Lawrence M., McRoberts R.E. (eds.). (2010). In: *National forest inventories: pathways for common reporting*. Springer. 612 p.
<http://dx.doi.org/10.1007/978-90-481-3233-1>.
23. Tomppo E., Heikkinen J., Henttonen H.M., Ihalainen A., Katila M., Mäkelä H., Tuomainen T., Vainikainen N. (2011). Designing and Conducting a Forest Inventory – case: 9th National Forest Inventory of Finland. *Managing Forest Ecosystems* 22. Springer, 305 p. DOI: 10.1007/978-94-007-1652-0.
24. Traat I., Inno J. (1997). *Tõenäosuslik valikuuring*. TÜ Kirjastus, 211 lk.

Versioon 4.0
Kalev Pärna 13.02.2023



Aruande koostaja: **Kalev Pärna**

Aruande valmimisse panustasid:

Raul Kangro, Märt Möls (analüüsid ja arvutuseksperimendid)

Imbi Traat, Meelis Käärrik (konsultatsioonid)

13.02.2023

Kalev Pärna

TÜ MSI emeriitprofessor

Projekti vastutav täitja