

**Pro gradu -tutkielma**

**Kanahaukan pesäpaikkojen ennustaminen  
paikkatietoanalyysillä**

**Anssi Parkkinen**



**Jyväskylän yliopisto**

Bio- ja ympäristötieteiden laitos

Ympäristötiede ja -teknologia

14.6.2019

JYVÄSKYLÄN YLIOPISTO, Matemaattis-luonnontieteellinen tiedekunta  
Bio- ja ympäristötieteiden laitos  
Ympäristötiede ja -teknologia

Anssi Parkkinen: Kanahaukan pesäpaikkojen ennustaminen  
paikkatietoanalyysillä  
Pro gradu -tutkielma: 61 s., 6 liitettä (32 s.)  
Työn ohjaajat: Yliopistonlehtori Anssi Lensu, FT Heidi Björklund ja  
professori Markku Kuitunen  
Tarkastajat: Dosentti Elisa Vallius, yliopistonlehtori Anssi Lensu  
Kesäkuu 2019

---

Hakusanat: Havumetsävyöhyke, Maksimientropia-mallinnus, METSO,  
monimuotoisuuskohteita, Random Forest -luokittelu

Kanahaukka (*Accipiter gentilis*) on Suomessa silmälläpidettäväksi (NT) luokiteltu keskikokoinen päiväpetolintu, joka pesii varttuneissa metsissä. Kanahaukan valitsemat pesämetsät ovat todettuja monimuotoisuuskohteita. Kanahaukka onkin monella tapaa metsien avainlaji. Tämän takia oli tarpeen selvittää kanahaukan tyypillisen pesämetsän piirteitä mahdollisen elinympäristöjen suojelun kohdistamiseksi niille alueille, joilla saavutetaan monimuotoisuudelle suotuisin suojelutulos. Tässä tutkimuksessa tunnistettiin maksimientropiamallinnuksella kanahaukalle tärkeitä elinympäristömuuttujia ja paikannettiin lajille todennäköisesti sopivia metsiä maisematasolla Keski-Suomen alueella MaxEnt-ohjelmalla Luonnonvarakeskuksen valtakunnan metsien inventointiin liittyvän monilähdetietoaineiston perusteella. Lisäksi tutkittiin RandomForest-luokittelun avulla eroja pesäpisteiden ja satunnaispisteiden ympäristössä eri mittakaavoilla. Havaittiin, että parhaiten kanahaukan pesimisympäristöksi sopii sekapuustoinen metsä, jossa on kuusta yli 150 m<sup>3</sup>/ha. Maksimientropiamallinnuksen mukaan kanahaukan pesimisympäristöksi erinomaisesti tai hyvin sopivaa metsää oli 8 - 9 % Keski-Suomen maapinta-alasta. Maksimientropiamallinnus todettiin käytettävissä olleen aineiston puutteet huomioitaessa tämän työn analysointeihin paremmin soveltuvaksi menetelmäksi. Maksimientropiamallinnuksen tuottamia kanahaukalle parhaiten sopivia elinympäristömuuttujien arvoja ja todennäköisimpiä hyvän kanahaukkametsän sijainteja voidaan käyttää suunniteltaessa kanahaukan suojelutoimenpiteitä.

UNIVERSITY OF JYVÄSKYLÄ, Faculty of Mathematics and Science  
Department of Biological and Environmental Science  
Environmental Science and Technology

Anssi Parkkinen: Prediction of Goshawk nesting sites in Central Finland with spatial analysis  
MSc thesis: 61 p., 6 appendices (32 p.)  
Supervisors: Senior lecturer Anssi Lensu, PhD Heidi Björklund and Professor Markku Kuitunen  
Inspectors: University teacher Elisa Vallius, senior lecturer Anssi Lensu

June 2019

---

Key words: Biodiversity site, Boreal forests, the Finnish forest biodiversity program, Maximum entropy modeling, Random Forest classifier.

The Northern Goshawk (*Accipiter gentilis*, hereafter Goshawk) is a medium sized bird of prey species, that breeds in mature forests and is classified as nearly threatened (NT) in Finland. The Goshawk indicates forest biodiversity as many threatened species are present in sites that the Goshawk uses for nesting. Therefore, it is important to study forest characteristics typical for Goshawk's nesting sites in managed forests in order to be able to locate and to target conservation efforts to sites with potentially high biodiversity. In this study, data from National Forest Inventory (NFI) were analyzed with maximum entropy modelling of species' distributions using the MaxEnt program to identify environmental characteristics typical of the Goshawk's nest sites, and to locate forests likely suitable for Goshawk nesting within the region of Central Finland. In addition, nest site selection of the Goshawk was studied by detecting differences in habitats between the surroundings of Goshawk's nest sites and randomly selected forested locations with Random Forest classifier using forest patch inventories and other geographic data from multiple sources. According to my results, the Goshawk prefers forests with a minimum of 150 m<sup>3</sup>/ha of spruce mixed with some deciduous trees as a nesting habitat. Of the land area of the region of Central Finland, about 8 to 9 % was designated excellent or suitable for Goshawk nesting.

# SISÄLLYSLUETTELO

<b>1 JOHDANTO</b> .....	<b>1</b>
<b>2 TUTKIMUKSEN TAUSTA</b> .....	<b>4</b>
2.1 Kohdelaji .....	4
2.2 Kanahaukan elinympäristövaatimukset.....	5
2.3 Kanahaukan yhteydet muihin vanhan metsän lajeihin .....	6
<b>3 AINEISTO JA MENETELMÄT</b> .....	<b>7</b>
3.1 Tutkimuksessa käytetty aineisto ja järjestelyt .....	7
3.1.1 Tutkimusalue.....	7
3.1.2 Metsäkeskuksen metsävaratietoaineisto .....	7
3.1.3 Maanmittauslaitoksen maastotietokanta .....	8
3.1.4 Suomen ympäristökeskuksen (SYKE) suojelualueaineistot .....	8
3.1.5 Valtakunnan metsien inventoinnin monilähdetietoaineisto (MVMI) .	9
3.1.6 Pesätiedot.....	9
3.1.7 Kanahaukan pesimisympäristö .....	10
3.2 Menetelmät.....	10
3.2.1 Eri aineistojen käsittelyssä käytetyt menetelmät ja ohjelmat .....	10
3.2.2 Maksimientropiamallinnus .....	11
3.2.3 Random Forest-luokitteluanalyysi .....	17
<b>4 TULOKSET</b> .....	<b>28</b>
4.1 Rasteriaineiston solukoko maksimientropiamallissa .....	28
4.2 Haukkametsän piirteet MaxEnt-analyysin mukaan.....	31
4.3 Haukkametsän piirteet RandomForest-analyysin mukaan .....	40
<b>5 TULOSTEN TARKASTELO</b> .....	<b>53</b>
5.1 Kanahaukkametsän tunnistus maksimientropiamallinnuksella.....	53
5.2 Kanahaukkametsän tunnistus RandomForest-analyyseillä .....	55
5.3 Maksimientropiamallinnuksen ja RandomForest-analyysin vertailua.....	58
<b>6 JOHTOPÄÄTÖKSET</b> .....	<b>60</b>
<b>KIITOKSET</b> .....	<b>61</b>
<b>KIRJALLISUUS</b> .....	<b>62</b>
<b>LIITTEET</b> .....	<b>70</b>

<b>LIITE 1. AINEISTOLLE TEHDYT ESIKÄSITTELYT.....</b>	<b>70</b>
<b>LIITE 2. VMI-AINEISTON KASVUPAIKKA-LUOKAN LUOKKANIMET .....</b>	<b>79</b>
<b>LIITE 3. SOLUKOON 160 M TULOSKUVAAJIA .....</b>	<b>80</b>
<b>LIITE 4. CFOREST-AINEISTON TULOSKUVAAJAT .....</b>	<b>85</b>
<b>LIITE 5. SOLUKOON 80 M TULOSKUVAAJAT .....</b>	<b>92</b>
<b>LIITE 6. MEAN- JA MEDIAN- JAOTTELU MVMI-MUUTTUIJILLE .....</b>	<b>102</b>

## 1 JOHDANTO

Boreaalinen havumetsävyöhyke on pohjoisin suurista biomeista. Metsät ovat merkittävä ekosysteemipalveluiden tarjoaja, eli ne hyödyttävät ihmistä monin tavoin (Pohjanmies *ym.* 2017). Metsien puuston runsas hyödyntäminen aiheuttaa kuitenkin merkittävän uhkan muille metsissä eläville lajeille, jotka eivät yleensä tule toimeen ilman suojaavaa puustoa. Kun Pohjois-Euroopan uhanalaisista lajeista 37 % on metsälajeja (Kuuluvainen 2009), tulisi niiden edellyttämiä elinympäristöjä säilyttää.

Fennoskandia eroaa muusta boreaalisen vyöhykkeen alueesta siinä, että metsätalous on vaikuttanut alueen metsiin pidempään ja voimakkaammin kuin Siperiassa tai Pohjois-Amerikassa (Östlund *ym.* 1997, Pohjanmies *ym.* 2017). Toisen maailmansodan jälkeen metsätalouden käytännöille on ollut tyypillistä tasaikäisten metsien harvennuksille ja uudistushakkuille perustuva kasvatusmalli, joka korvasi aiemmin vallinneet poimintahakkuut. Viime aikoihin asti on vallinnut käsitys, että metsätalouden alaharvennus- ja uudistushakkuut mukailevat luonnollista suknessiota (Kuuluvainen 2009). Kuitenkin luonnontilaisissa metsissä metsäpalotkin vain harvoin aiheuttavat eri-ikäisrakenteisen metsän muuttumista tasaikäiseksi. Metsien rakennemuutoksen merkittävyys maisemamittakaavassa alkoi selvitä, kun elonkirjon vähenemiseen kiinnitettiin huomiota ja kun luonnontilaisten metsien rakennetta ja suknessiota alettiin tutkia 1990-luvun lopulla (Kuuluvainen 2009). Tällöin havaittiin, että luonnontilaisten metsien rakenne boreaalisella havumetsävyöhykkeellä poikkeaa huomattavasti talousmetsien rakenteesta (Kuuluvainen 2009). Vasta viime aikoina metsälain muutoksen myötä on tullut mahdolliseksi myös eri-ikäisrakenteisiin metsiin pohjautuva metsätalous, mutta sen osuus talousmetsissä on vielä pieni (Kuuluvainen *ym.* 2012).

Metsätalouk käytöstä on seurannut myös muita metsien rakenteellisia muutoksia. Metsäpalojen vähentyminen lähes olemattomiin on johtanut siihen, että suojelluissa metsissä paloja paremmin sietävän männyn osuus paloille herkempään kuuseen

verrattuna on vähentynyt (Kuuluvainen 2009). Lisäksi metsien ikärakenteessa on tapahtunut muutos: esimerkiksi Keski-Suomessa 21–60-vuotiaiden metsien osuus on noussut noin 23 %:sta 50 %:iin, kun taas 61–100-vuotiaiden metsien osuus on laskenut noin 50 %:sta 26 %:iin 1960-luvun lopulta vuoteen 2013 mennessä (Salminen 2015). Ruotsissa voidaan havaita vastaavanlainen muutos: 1910-luvulla vanhoja, yli 150-vuotiaita metsiä oli 83 % kaikista metsistä, kun 1980-luvulla vastaava luku yli 160-vuotiaille metsille oli enää alle 3 %. Selvimmin metsien rakenteen muutokset tulevatkin esiin maisematasolla vanhojen metsien ja vanhojen puiden määrän vähenemisen myötä (Kuuluvainen 2009).

Metsätalouden vaikutuksista vanhojen metsien lajeihin on useita tutkimuksia. Niissä on todettu, että jäkälien ja selkärangattomien määrä luonnontilaisissa metsissä on merkitsevästi suurempi kuin talousmetsissä, ja tätä kautta metsätalous vähentää lintujen ravinnonhankintamahdollisuuksia (Pettersson *ym.* 1995). Myös kääpien ja niillä elävien kovakuoriaisten määrä on vähäisempi talousmetsissä kuin luonnontilaisissa metsissä lahopuun puuttumisen vuoksi (Siitonen & Martikainen 1994, Selonen *ym.* 2005).

Pohjois-Euroopassa on tyypillistä, että suojelualueet ovat pieniä ja hajanaisesti sijoittuneita, joten vanhan metsän lajien kannalta olisi tärkeää, että suojelualueita lisättäisiin. Ruotsin ja Suomen suojelualueista valtaosa sijaitsee pohjoisessa vähätuottoisilla mailla, joilla on matala elonkirjo. Etelä-Suomen ja Etelä-Ruotsin viime aikoina perustetuille suojelualueille on leimallista ihmisvaikutuksen näkyminen, jopa nykyaikaisen metsätalouden jäljet (Kuuluvainen 2009). Suojelualueiden ohella täydentävä keino metsäluonnon elonkirjon ylläpitoon on tunnistaa ja huomioida monimuotoisuudelle tärkeitä kohteita talousmetsissä. Tällaisten kohteiden tunnistamisessa voidaan käyttää apuna keskeiselle lajille sopivia esiintymispaikkoja (Naumov *ym.* 2018), sillä keskeisen lajin tarpeiden ajatellaan kuvastavan monien muidenkin lajien tarpeita (Lambeck 1997).

Taloukskäytössä olevien metsien hallitseman maiseman rakennetta ja mahdollisia monimuotoisuuskohteita tutkittiin tässä pro gradu -työssä vanhan metsän lajin kautta. Keskeiseksi vanhan metsän lajiksi valikoitui kanahaukka, sillä se on metsien

avainlaji: sen pesäpaikan läheisyydessä on usein muitakin vanhan metsän lajeja (Burgas *ym.* 2014), ja jotkin lajit hakeutuvat erityisesti kanahaukan lähelle suojaan omilta saalistajiltaan (Mönkkönen *ym.* 2007, Pakkala 2012). Kanahaukka vaikuttaa muiden petolintulajien esiintymiseen suorasti ja epäsuorasti: kanahaukka voi saalistaa muita petolintuja (Selås 1997, Rutz & Bijlsma 2006), ja muut petolinnut voivat karttaa kanahaukan läheisyyttä saalistusuhan takia (Björklund *ym.* 2016). Kanahaukka luo myös pesäpaikkoja, sillä monet lajit pesivät kanahaukan rakentamissa pesissä (Meller *ym.* 2017). Kanahaukkaan kohdistetuilla suojelutoimilla voidaankin edistää myös muiden lajien suojelutilannetta.

Kanahaukka valitsee saalistusmaastonsa metsän rakenteen perusteella, ei niinkään saaliin esiintyvyyden mukaan (Beier & Drennan 1997). Tämän takia tässä tutkimuksessa käytettiin avaruudellisesti yksitulkintaisia mallinnusmenetelmiä, joiden avulla tunnistettiin kanahaukan pesäpaikkoja ja siten mahdollisia monimuotoisuuskohteita kuvaavia elinympäristöpiirteitä. Niiden perusteella mallinnettiin kanahaukan elinympäristövaatimusten toteutumista maisematasolla arviona metsämaiseman tilasta vanhan metsän lajien kannalta.

Tutkimuskysymyksiä olivat:

1. Mitkä ovat kanahaukan pesäpaikan valinnalle tärkeimpiä elinympäristömuuttujia eri mittakaavoilla? Missä määrin talousmetsämaisemassa on näiden muuttujien perusteella vanhan metsän lajille sopivaa elinympäristöä
2. Onko pesäpisteiden ja satunnaispisteiden ympäristöissä eroja elinympäristömuuttujien suhteen eri mittakaavoilla, mikä kertoisi kanahaukan elinympäristövalinnasta?
3. Saadaanko eri analyysimenetelmien ja aineistojen avulla samat pesäpaikkaa kuvaavat elinympäristömuuttujat esille, ja voidaanko päätellä, mikä menetelmä on soveltuvin pesäpaikan elinympäristön ominaisuuksien selvittämiseen?

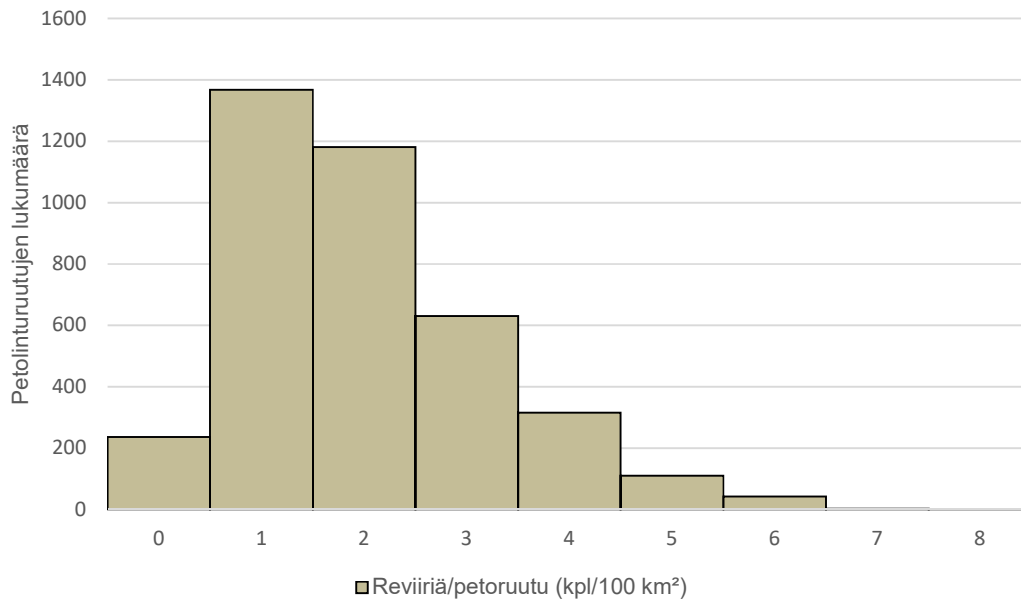


## 2 TUTKIMUKSEN TAUSTA

### 2.1 Kohdelaji

Kanahaukka on keskikokoinen päiväpetolintu, joka käyttää ravintonaan pääasiassa metsäkanalintuja, mutta myös muita lintuja ja nisäkkäitä (Linden & Wikman 1983, Tornberg 2001, Sulkava *ym.* 2014). Kanahaukka on aikuisena paikkalintu (Saurola *ym.* 2013). Kanahaukan reviirillä on yksi tai useampi vaihtopesäpuu, joita haukat käyttävät pesintään eri vuosina (Björklund *ym.* 2016). Kanahaukan pesimistapojen muutoksia pitkällä aikavälillä on tutkittu Länsi-Suomessa. Vaikka kanahaukan pesivän populaation määrä on pysytellyt kutakuinkin samankokoisena aikavälillä 1983–1996, yksittäisten pesien pitkäaikainen käyttö on vähentynyt ja lyhytaikaisten pesien käyttö lisääntynyt (Hakkarainen *ym.* 2004). Tämä johtunee hakkuiden lisääntymisestä (Hakkarainen *ym.* 2004). Kanahaukka välttää nuoria metsiä metsästäessään (Widén *ym.* 1987), ja hakkuista aiheutunut metsien rakennemuutos (Salminen 2015) on todennäköisesti vähentänyt kanahaukalle sopivia pesimis- ja metsästyksmaastoja.

Kanahaukan kannantiheys on tyypillisesti 1–2 reviiriä / 100 km<sup>2</sup> (kuva 1). Suomen kanahaukkapopulaatio on petolintuseurannan mukaan laskenut vuosina 1982–2017 n. 0,6 %:n vuosivauhdilla (Björklund *ym.* 2018). Kanahaukka onkin luokiteltu Suomessa silmälläpidettäväksi lajiksi (Lehikoinen *ym.* 2019).



Kuva 1. Keskimääräinen kanahaukkatiheys petolinturuuduilla vuosina 1982–2016, kpl/100 km<sup>2</sup> (Luonnontieteellinen keskusmuseo Luomus, julkaisematon). Saman petolinturuudun tiedot voivat esiintyä kuvassa monta kertaa ollessaan eri vuosien havaintoja.

## 2.2 Kanahaukan elinympäristövaatimukset

Useissa tutkimuksissa on todettu, että vanhan metsän ja kuusimetsän suuri osuus on tärkeää kanahaukan pesäpaikan valinnalle (Widén 1989, Widén 1997, Krüger 2002, Hakkarainen *ym.* 2004, Löhmus 2005, Selås *ym.* 2008). Suomessa on todettu, että kanahaukan elinympäristön ominaisuudet 2000 m säteellä pesästä selittävät parhaiten kanahaukan pesimismenestystä (Björklund *ym.* 2015). Huomattavaa oli se, että edes alueilla, joilla nuoren metsän osuus metsän pinta-alasta kasvoi, kanahaukan reviirin alueella nuoren metsän osuus ei kasvanut. Tämä viittaa siihen, että kanahaukka välttää nuoria metsiä. Kanahaukan on todettu hylkäävän pesäpaikkansa, mikäli latvuspeittävyys vähenee 30 % (Penteriani & Faivre 2001). Kyseisen tutkimuksen metsät olivat lähinnä pyökki- (*Fagus sylvatica*) ja tammimetsiä (*Quercus* spp.), ja metsänkäsittelytapana olivat toistuvat, 2–3 vuoden välein tehdyt harvennukset suojuspuuasentoon. Metsänkäsittelyt kohdistuivat noin 10 hehtaarin tai sitä pienempiin metsälöihin. Kun tutkittiin metsän ominaisuuksia hehtaarin alueella pesän ympärillä, tultiin tulokseen, että mikäli

hakuut oli suoritettu vähintään sadan metrin päässä pesästä, hakuilla ei todennäköisesti ollut vaikutusta kanahaukan pesintään (Penteriani & Faivre 2001). Italiassa tehdyssä tutkimuksessa havaittiin, että kanahaukkojen pesäpaikkojen ympäristössä muun muassa puuston rinnankorkeusläpimitta, pituus, latvuksen koko sekä puiden välinen etäisyys saivat suurempia arvoja kuin kontrollipisteissä (Penteriani & Faivre 1997). Tämä viittaa siihen, että kanahaukat suosivat pesäpaikkanaan järeää metsää. Pesimistiheydeksi Italiassa saatiin 5,03 paria/100 km<sup>2</sup>.

Kanahaukalle tyypillisen pesimisympäristön rakenteen selvittämiseksi on usein tutkittu elinympäristön koostumusta eri säteillä kanahaukan pesästä. Länsi-Suomessa tutkittiin elinympäristökoostumusta 250 m, 500 m, 1000 m, 2000 m ja 4000 m säteillä kanahaukan pesästä (Hakkarainen *ym.* 2004). Kanahaukan pesimisen onnistumisesta vuosien 1975 ja 1999 välillä Saksassa tehdyssä tutkimuksessa käytettiin säteitä 17,8 m ja 500 m useiden elinympäristömuuttujien osuuksien osalta sekä 200 m sädettä pysyvien häiriöiden osalta (Krüger 2002). New Yorkin yläköalueella Yhdysvaltain itäosassa käytettiin vain 43 m läpimittaista aluetta vertailtaessa kanahaukan pesimisympäristöä ja satunnaispisteiden ympäristöä. Italiassa ja Ranskassa n. 100 m etäisyys osoittautui rajaksi, jonka sisällä tehdyt hakuut alkoivat vaikuttaa pesinnän jatkumiseen (Penteriani & Faivre 2001).

### **2.3 Kanahaukan yhteydet muihin vanhan metsän lajeihin**

Kanahaukan pesäpaikka on merkki siitä, että alueella on todennäköisesti muitakin vanhaa metsää elinympäristönään käyttäviä lajeja, kuten uhanalaisia kääpiä (Burgas *ym.* 2014). Kanahaukka luo myös pesäpaikkoja, sillä monet lajit, kuten piekana, ampuhaukka, nuolihaukka, viirupöllö, lapinpöllö, varpushaukka ja hiirihaukka pesivät kanahaukan rakentamissa pesissä (Meller *ym.* 2017). Muut lajit, esimerkiksi liito-orava voivat myös hakeutua kanahaukan lähelle suojaan omilta saalistajiltaan (Mönkkönen *ym.* 2007, Pakkala 2012), sillä kanahaukka voi saalistaa myös muita petolintuja (Selås 1997, Björklund *ym.* 2016). Tällaista tilannetta, jossa keskenään kilpailevat pedot saalistavat ja syövät toisiaan tai toinen laji toista,

kutsutaan killansisäiseksi saalistukseksi (Polis *ym.* 1989, Polis & Holt 1992). Esimerkki killansisäisen saalistuksen epäsuorasta vaikutuksesta on se, että hiirihaukka karttaa kanahaukalle suotuisia alueita, joilla on runsaasti metsäkanalintuja, mahdollisesti välttääkseen killansisäisen saalistuksen riskiä (Björklund *ym.* 2016). Kanahaukka voikin vaikuttaa muiden petolintujen esiintyvyyteen alueella suoraan saalistuksen kautta tai epäsuorasti siten, että pienemmät lajit välttävät alueita, joilla kanahaukka esiintyy (Björklund *ym.* 2016).

### 3 AINEISTO JA MENETELMÄT

#### 3.1 Tutkimuksessa käytetty aineisto ja järjestelyt

##### 3.1.1 Tutkimusalue

Tutkimusalue oli Keski-Suomen maakunta, jonka pinta-ala on n. 20 000 km<sup>2</sup>. Keski-Suomen maapinta-alasta 87 % eli n. 14 460 km<sup>2</sup> on metsätalousmaata (Kaila & Ihalainen 2014). Tästä 97 % on käytettävissä puuntuotantoon täysimittaisesti tai rajoitetusti. Metsien puusto on pääosin alle 60-vuotiasta. Puuston tilavuus on kasvanut vuosien 1996 ja 2013 välillä 164,6 miljoonasta kuutiometristä 186,7 miljoonaan kuutiometriin. Keski-Suomessa kuusen tilavuus ajanjaksolla 1996–2013 on vähentynyt 67,8 miljoonasta kuutiometristä 65,3 miljoonaan kuutiometriin. Puustotilavuudesta 47 % on mäntyä (*Pinus sylvestris*), 35 % kuusta (*Picea abies*) ja 15 % koivua (*Betula* spp.) ja loput tilavuudesta on muita lehtipuita (Kaila & Ihalainen 2014).

##### 3.1.2 Metsäkeskuksen metsävaratietoaineisto

Suomen metsäkeskukselta (tästä eteenpäin Metsäkeskus) saatiin käyttöön yksityismetsien kuviokohtaista metsävaratietoaineistoa. Tässä tutkimuksessa käytin vuodelle 2017 laskettuja ennusteita. Metsävaratietoaineiston työssä käytettyjä kuviokohtaisia tietoja olivat maalaji, hakkuurajoite, kuivatusaste, maaperätyyppi, metsän pää- ja alaryhmä, sekä puuston osalta puun eri osien

biomassa, kehitysluokka, pääpuulaji, keski-ikä, pohjapinta-ala, runkoluku, keskiläpimitta ja keskikorkeus, tilavuus, tukki- ja kuitupuun osuus ja vuosittainen kasvu. Vastaavat puustotiedot olivat käytettävissä myös erikseen puulajeittain kuuselle, mänylle, hies- ja rauduskoivulle, haavalle ja lepälle. Metsäkeskuksen aineistossa oli runsaasti aukkoja (esim. peltoja, järviä ja muita kuin yksityisomistuksessa olevia metsiä). Tämän takia käytin täydentävänä aineistona Maanmittauslaitoksen ja Suomen ympäristökeskuksen aineistoja, jotka sisältävät tietoja Metsäkeskuksen aineistosta puuttuvista alueista. Nämä tiedot eivät kuitenkaan ole samanlaisia metsävaratietoja kuin Metsäkeskuksen aineistossa, vaan ne sisältävät tietoa muista aluetyypeistä. Lisäksi otin työhöni mukaan spatiaalisesti kattavampaan MVM-aineistoon perustuvan tarkastelun.

### 3.1.3 Maanmittauslaitoksen maastotietokanta

Täydentävänä elinympäristötietona käytin Maanmittauslaitoksen maastotietokantaa, joka on tarkin Suomessa valtakunnallisesti käytössä oleva paikkatietoaineisto. Sijaintitietojen tarkkuus ja karttaobjektien yksityiskohtaisuus vastaavat mittakaavaa 1:5 000 – 1:10 000. Maastotietokannan kohteet on jaettu 18 kohderyhmään, joiden alla on lukuisia eri kohdeluokkia. Tässä tutkimuksessa valitsin joitakin kohdeluokkia, jotka jaoin neljään tarkasteluyksikköön – rakennukset, maatalousmaa, suot ja vesistöt. Maastotietokanta on vektorialueaineisto, joka on ETRS89-TM35FIN-koordinaattijärjestelmässä.

### 3.1.4 Suomen ympäristökeskuksen (SYKE) suojelualueaineistot

Lisäksi täydensin Metsäkeskuksen aineistoa SYKE:n suojelualueaineistolla. Käytettyjä aineistoja olivat Natura 2000 -alueet ETRS89-TM35IN-koordinaattijärjestelmässä vuodelta 2012, Luonnonsuojelu- ja erämaa-alueet (ETRS89-TM35FIN, 2015) sekä luonnonsuojeluohjelmat (Kartastokoordinaattijärjestelmän yhtenäiskoordinaatisto, 2008). Kaikki aineistot ovat vektorialueaineistoja mittakaavassa 1:20 000.

### 3.1.5 Valtakunnan metsien inventoinnin monilähdetietoaineisto (MVMI)

Käytin työssäni myös Luonnonvarakeskuksen MVMI-aineistoa. MVMI-aineisto on yhdistelmä tietoa valtakunnan metsien inventoinnin maastomittauksista, satelliittikuvista (lähinnä Landsat-satelliittisarjan kuvaamia), korkeusmalleista, numeerisista peruskartoista sekä muista numeerisista tietolähteistä. MVMI-aineisto tallennetaan nykyään ETRS89-TM35FIN-koordinaattijärjestelmässä rasteri-aineistona, jonka solukoko on 16 m × 16 m (Mäkisara *ym.* 2019).

Tässä tutkimuksessa käytin MVMI-aineistoa, joka pohjautuu vuosina 2009–2016 kerättyyn valtakunnan metsien 11. ja 12. inventointien maastoaineistoon. Aineiston julkaisuvuosi on 2017. MVMI-aineisto sisältää metsän ominaisuustietoa seuraavista muuttujista: tilavuus, päätyyppi, pohjapinta-ala, ikä, kasvupaikka, keskiläpimitta, keskipituus, latvuspeittävyys sekä erikseen tiedot kuusen, männyn, koivun, lepän ja muun lehtipuun kokonaistilavuudesta sekä puun eri osien tilavuuksista.

### 3.1.6 Pesätiedot

Käytin tutkimuksessani kanahaukan asuttujen pesien aineistoa Keski-Suomesta. Aineiston keräsi vapaaehtoinen lintuharrastaja Tomi Hakkari vuosina 2015 ja 2016. Aineiston tietoja olivat pesätyyppi (risupesä, tekopesä tai risupesän tilalle tehty tekopesä) sekä pesän itä- ja pohjoiskoordinaatti kartastokoordinaattijärjestelmän yhtenäiskoordinaatiston koordinaatteina, jotka muunnettiin muussa aineistossa käytössä olevaan ETRS89-TM35FIN-koordinaattijärjestelmään. Tekopesät poistettiin, koska niiden sijainti ei ole linnun valitsema. Mikäli samalla reviirillä oli useampia vaihtopesiä, valittiin reviiriltä satunnaisesti yksi pesä valetoston välttämiseksi (Hurlbert 1984). Mikäli samalta pesältä oli havaintoja molemmilta vuosilta, vain vuoden 2016 havaintoa käytettiin valetoston välttämiseksi. Valitulla vuodella ei ollut väliä, sillä pesästä käytettiin vain sijaintitietoja. Jäljelle jäi tällöin 78 pesäpistettä.

### 3.1.7 Kanahaukan pesimisympäristö

Kanahaukan pesien ympäristöä tutkittiin myös luokittelumenetelmiä käyttäen vertaamalla pesien ympäristöä satunnaispisteiden ympäristöön. Loin Keski-Suomen alueelle 100 satunnaispistettä. Satunnaispisteiden luomisessa kriteerinä oli, että niiden tulee sijaita vähintään kahdeksan kilometrin etäisyydellä todellisista pesistä, jolloin analyysissä käytettävillä vyöhykkeillä ei tule päällekkäisyyksiä. Lisäksi satunnaispisteen edellytettiin sijaitsevan valtakunnan metsien inventointiaineiston (ks. 3.1.5) mukaan keski-ikänsä vähintään 40-vuotiaassa metsässä, ja sen neljän kilometrin säteisen puskurin piti sisältää vähintään 50 kpl metsäpikseliä (kooltaan 16 m × 16 m). Tällä pyrittiin varmistamaan satunnaispisteiden sijaitseminen metsäisessä ympäristössä siten, että osa niistä ei osunut esimerkiksi pieneen saareen tai pellon keskellä olevaan metsäsaarekkeeseen.

Vertailin pesä- ja satunnaispisteiden ympäristöjen eroja luomalla pisteiden ympärille eri säteisiä ympyrän muotoisia tarkastelualueita. Vertailussa käytin edellä esitetyistä säteistä johdettuja säteitä 200 m, 250 m, 500 m, 1000 m, 2000 m ja 4000 m. Metsäkeskuksen aineiston kuviokoko on niin suuri, että pienimmät kappaleessa 2.2 esitetyistä säteistä eivät kuvaisi tarkasti todellista tilannetta pesäpisteen ympäristössä, joten jätin ne pois tarkastelusta.

## 3.2 Menetelmät

### 3.2.1 Eri aineistojen käsittelyssä käytetyt menetelmät ja ohjelmat

Käytin tutkimuksessani kahta eri analysointimenetelmää -maksimientropiamallinnusta ja RandomForest-luokittelijaa. Maksimientropiamallinnusta käytin MVMI-aineistolle, joka on rasterimuotoista. RandomForest-analyysin yhteydessä käytin Metsäkeskuksen metsävaratietoaineistoa ja Maanmittauslaitoksen maastotietokannan ja SYKE:n suojelalueiden aineistoja, jotka ovat vektorimuotoisia aineistoja.

Aineiston käsittelyssä käytin ArcGIS Desktop v. 10.3.1 -ohjelmistoa (ESRI, Redlands, CA, USA), maksimientropiamallinnuksessa MaxEnt-ohjelman versiota 3.4.0 (Phillips *ym.* 2016) ja muissa analyyseissä R v 3.5.2 -tilasto-ohjelmaa (R Core Team 2018) sekä seuraavia paketteja (versionumerot sulkeissa) maptools (0.9–14), mclust (5.4.2), randomForest (4.6–14), gdata (2.18.0), mice (3.3.0), rgdal (1.3–6), varSelRF (0.7.8), caret (6.0–81), party (1.3–1), VIM (4.7.0) sekä Boruta (6.0.0).

### 3.2.2 Maksimientropiamallinnus

Maksimientropiamallinnus perustuu tiedossa oleviin esiintymispaikkatietoihin, ja mallinnus ei edellytä tietoa siitä, missä lajia ei varmasti esiinny (Phillips *ym.* 2006, Elith *ym.* 2011). Tämä on eduksi tapauksissa, joissa ei ole ehdottoman varmaa tietoa siitä, missä lajia ei esiinny, jolloin "ei esiinny" -tietoa edellyttävät mallit voisivat tuottaa virheellisiä päätelmiä (Elith *ym.* 2011). Pelkkää esiintymistietoa käyttävillä malleillakin on tyypillisesti rajoitteita: ei voida olla täysin varmoja siitä, että on havaittu tutkittavan lajin kaikki esiintymät alueella (Elith *ym.* 2011). Perustuen tiedossa olevien esiintymispisteiden kohdalla todettuihin ympäristömuuttujien arvoihin, esimerkiksi sää-, maaperä- tai kasvillisuustietoihin, maksimientropiamallinnuksessa ennustetaan todennäköisyyksiä sille, missä laji voisi esiintyä ja missä ei. Todettaessa elinympäristömuuttuja-aineistojen sisältävän lajille sopivaa elinympäristöä esiintymisen todennäköisyys muodostuu korkeammaksi kuin siellä, missä elinympäristö on soveltumatonta (Phillips *ym.* 2004). Kyseessä ei ole suoraan lajin esiintymistodennäköisyys kyseisessä rasterisolussa, sillä mallinnuksen tutkimusalueen pikseleille antamat todennäköisyydet koko tutkimusalueen yli tarkasteltuna summautuvat lukuun yksi. Näin ollen yksittäisen pikselin saama todennäköisyysarvo on tavallisesti hyvin pieni. Tuloksia havainnollistettaessa MaxEnt tekee todennäköisyyksille cloglog-muunnoksen kaavan 1 mukaan:

$$arvo_{cloglog} = 1 - e^{(-cr)}, \quad (1)$$

missä  $r$  on todennäköisyyden raaka-arvo ja  $c$  on MaxEnt-ohjelman tuottaman jakauman entropia-arvo muunnettuna eksponenttifunktiolla,  $c = e^H$  (Phillips



2017). Cloglog-muunnetut arvot skaalautuvat välille 0–1, ja huonosti kanahaukalle soveltuva metsä saa arvon  $\leq 0,46$ , tyypillinen kanahaukkametsä saa arvon 0,46–0,69 ja erinomainen kanahaukkametsä saa arvon 0,69–1,0 (Phillips 2017).

Maksimientropiamallinnuksessa käytetty MaxEnt-ohjelma edellyttää käsiteltävän ympäristötiedon olevan rasterimuotoista ja tutkimusalueen osalta aukotonta, joten käytin ympäristöaineistona MVMI-aineistoa. MaxEnt-ohjelmaan voidaan syöttää sekä luokiteltuja että jatkuvia ympäristömuuttujia (Phillips *ym.* 2006), mikä oli tarpeellinen ominaisuus, sillä MVMI-aineisto sisältää molempia.

Toinen edellytys on, että lajin esiintymistietojen keräämisessä ja valinnassa ei ole vinoumaa esimerkiksi valetoston tai virheellisten pesäpaikkatietojen muodossa (Elith *ym.* 2011). Tämän otin huomioon, kun jätin pois mallinnuksesta peräkkäisten vuosien havainnot samasta pesästä ja ihmisen valitsemiin paikkoihin tehdyt tekopesät (ks. 3.1.6). Yksi vinoumaa mahdollisesti aiheuttava tekijä, johon tätä tutkimusta tehdessäni en voinut vaikuttaa, oli kartoituksiin yleisesti liittyvä epävarmuus siitä, onko kaikkialle tutkimusalueelle kohdistettu yhtä suuri etsintätyö pesäpaikkoja kartoitettaessa (Elith *ym.* 2011).

Käsittelin pesäaineistoa ja MVMI-aineistoa ensin ArcGIS-ohjelmistolla (liite 1). Toin pesätiedot Excel-tiedostona ArcGISiin ja muunsin aineiston koordinaattijärjestelmän MVMI-aineiston käyttämään ETRS89-TM35FIN-koordinaattijärjestelmään. Käsittelin pesätiedot siten, että mukana oli enää laji- ja koordinaattisarakeet, jotka nimesin species-, east- ja north-sarakkeiksi, ja tallensin tiedoston csv-muotoon. Luokittelin MVMI-aineiston muuttujat eri kansioihin ja yhdistin kunkin muuttujan neljä eri karttalehtitiedostoa siten, että kutakin muuttujaa varten oli yksi yhdistetty tiedosto, joka sisälsi kaikki karttalehdet. Yleistin jatkuvien muuttujien arvot ArcGISin Aggregate-työkalulla sekä  $5 \times 5$  solun eli  $80 \text{ m} \times 80 \text{ m}$  solukoon että  $10 \times 10$  solun eli  $160 \text{ m} \times 160 \text{ m}$  solukoon tiedostoiksi laskemalla yleistettävälle alueelle joko muuttujan keskiarvo tai mediaani. Nämä yleistyksessä käytetyt kertoimet näkyvät esimerkiksi kuvan 9 muuttujanimissä. Luokitellut muuttujat yleistin Resample-työkalulla samoihin solukokoihin, koska sen työkalun yhteydessä voi käyttää menetelmänä luokittelumuuttujille paremmin

soveltuvaa MAJORITY-menetelmää. Käytin kahta eri solukokoa, koska halusin selvittää, onko yleispiirteisemmällä tarkastelumittakaavalla erotettavissa laajempia kanahaukan mahdollisia esiintymisalueita kuin pienipiirteisemmällä tarkastelumittakaavalla, ja toisaalta katoaako yleispiirteisemmässä tarkastelussa paljonkin lajille sopivia pienialaisia alueita. Rajasin yleistetyn aineiston Keski-Suomen maakunnan rajojen mukaan. MVMI-aineistossa oli alun perin 44 metsän rakennetta kuvaavaa ympäristömuuttujaa. Näistä karsittiin pois muuttujia, joiden merkitys kanahaukalle on tiettävästi vähäinen aiempien tutkimusten ja lajintuntemuksen perusteella (esim. juurten biomassa). Jäljelle jääneistä muuttujista moni oletettavasti korreloi keskenään (esim. kuusen eri osien tilavuuteen liittyvät muuttujat), mikä voi johtaa siihen, että maksimientropiamallinnuksen tulokset ovat epäluotettavia (Phillips 2017). Ympäristömuuttujien väliset Pearson-korrelaatiot tutkittiin käyttämällä muuttuja-arvoja pesien kohdalla (solukoko 80 m × 80 m) ja korkean korrelaation rajana pidettiin arvoa  $|r| \geq 0,7$  (Dormann *ym.* 2013). Korreloivista muuttujista valittiin ne, jotka ovat aiemmissa tutkimuksissa osoittautuneet lajin kannalta tärkeiksi, joiden merkitystä haluttiin tarkastella tai joiden perusteella tulokset ovat yleistettävissä muualle. Jäljelle jäi luokitelluista muuttujista kasvupaikkatyyppi sekä yhdeksän jatkuva-arvoista muuttujaa: latvuspeittävyys, lehtipuiden latvuspeittävyys, männyn tilavuus, muiden lehtipuiden tilavuus, muiden lehtipuiden tukkipuuosuuden tilavuus, koivun tilavuus, koivun kuitupuosuuden tilavuus, kuusen hukkapuuosuuden biomassa ja kuusen tilavuus. Kuusen tilavuuden tilalla kokeiltiin myös yleiskäyttöisempänä vaihtoehtona puuston kokonaistilavuutta. Se korreloi vahvasti kuusen tilavuuden kanssa, joten näitä kahta muuttujaa ei käytetty samassa mallissa.

Näin saadut csv-muotoiset pesätiedot ja eri solukokoihin yleistetyt ympäristömuuttujat ajoin MaxEnt-ohjelmassa neljänä eri ajona. MaxEnt asetettiin jakamaan pesäpisteaineisto kymmeneen osaan ristiinvalidointia varten. Kukaan osa jätettiin vuorollaan pois siten, että malli luotiin jäljellä olevien pesäpisteiden avulla. Pois jätettyä osaa käytettiin mallin testaukseen. Kun tehtiin kymmenen ajoa, saatiin laskettua mallin tuloksille vaihteluväli ja keskiarvon keskivirhe ristiinvalidoinnin avulla. Eri ajojen tulosten vaihteluväli on kuvattu tuloskuvaajissa sinisellä.

Ajon asetuksia tehdessä tulee ottaa huomioon, että mallinnus voi antaa epärealistisia todennäköisyyksiä sellaisille muuttujan arvoille, jotka ovat pienempiä tai suurempia kuin tutkimusalueelta löytyvät arvot (Phillips *ym.* 2006). Tämän välttämiseksi tulee käyttää niiden rajoittamista (engl. clamping) mallin muodostamisessa käytetyn aineiston muuttujien arvovälien sisälle, jolloin olemassa olevien arvojen ääripäiden ennustearvoja käytetään kaikille löytyviä ylimpiä arvoja suuremmille ja alimpia arvoja pienemmille arvoille muodostettaessa tuloksia.

Ajon aikana MaxEnt luo todennäköisyysjakauman tutkimusalueelle. Todennäköisyysjakauman luominen aloitetaan olettamalla tasaisesti alueelle levittäytynyt populaatio. Tätä jakaumaa muutetaan useita kertoja paremmaksi sovittamalla mallia asetettujen ympäristömuuttujien avulla paremmin todellista lajin levittäytymistä vastaavaksi. Jakauman hyvyysarvoa kuvaava suure on *gain*. Se lasketaan vähentämällä esiintymispisteiden keskimääräisistä sijoittumistodennäköisyyksistä vakio, joka aiheuttaa tasaisesti jakautuneen populaation tapauksessa *gain*in arvoksi nollan. Niinpä keskimääräiseen esiintymispisteeseen lasketun todennäköisyyden arvo voidaan laskea kaavalla

$$e^{gain} = x, \quad (2)$$

missä  $x$  kertoo, kuinka monta kertaa suuremman todennäköisyysarvon lajin esiintyminen saa esiintymispisteessä kuin satunnaisessa pisteessä, missä ei ole esiintymispistettä. Tämä todennäköisyysarvo on kuitenkin vain suhteellinen, ei absoluuttien esiintymisen todennäköisyys, sillä koko tutkimusalueen kaikkien pikseleiden yhteenlaskettu todennäköisyys on asetettu arvoon yksi (Phillips 2017).

Yhtenä mallin sopivuuden tarkasteluna saadaan ROC-käyrä (Phillips *ym.* 2006). ROC (Receiver Operator Characteristic) kuvaa ennusteen onnistumista kahden luokan asteikolla: "pisteessä on pesäpaikka / pisteessä ei ole pesäpaikkaa" (DeLeo 1993). ROC ei ole riippuvainen asetetuista kynnyсарvoista. Tällöin voidaan määritellä seuraavat toisensa poissulkevat lopputulokset:

1. luokittelija antaa positiivisen tuloksen ja todellinen luokka on positiivinen
2. luokittelija antaa positiivisen tuloksen ja todellinen luokka on negatiivinen
3. luokittelija antaa negatiivisen tuloksen ja todellinen luokka on positiivinen
4. luokittelija antaa negatiivisen tuloksen ja todellinen luokka on negatiivinen

Luokittelijan tarkkuutta voidaan mitata tarkastelemalla sen onnistumista luokittelussa seuraavasti:

$$\text{sensitiivisyys} = \frac{\text{oikein tunnistettujen positiivisten kohteiden lkm}}{\text{kaikkien positiivisten kohteiden lkm}} \quad (3)$$

Sensitiivisyys-suure tunnetaan myös nimellä "true positive rate" (Phillips *ym.* 2006).

$$\text{spesifisyys} = \frac{\text{oikein tunnistettujen negatiivisten kohteiden lkm}}{\text{kaikkien negatiivisten kohteiden lkm}} \quad (4)$$

Spesifisyys taas tunnetaan nimellä "true negative rate" (Phillips *ym.* 2006).

ROC-kuvaajassa y-akselilla on sensitiivisyys, x-akselilla taas on arvo 1–spesifisyys (Fielding & Bell 1997, Phillips *ym.* 2006).

AUC (Area Under the Curve) on suure, joka kertoo tämän tutkimuksen tapauksessa, kuinka usein esiintymispisteistä satunnaisesti valittu joukko saa suuremman arvon ROC-käyrän alle jäävässä pinta-alassa kuin taustapisteistä satunnaisesti valittu joukko. Erityyppisten ja eri aineistoilla toteutettujen tutkimusten luotettavina pidettävät AUC-arvot vaihtelevat, mutta yleisesti voidaan todeta, että AUC-arvoilla 0,9-1,0 malli on erinomainen, 0,8-0,9 hyvä, 0,7-0,8 kohtalainen, 0,6-0,7 heikko ja 0,5-0,6 epäonnistunut (Swets 1988). Mikäli AUC-arvo on 0,8, jää tällaisen esiintymispistejoukon muodostaman ROC-käyrän alle 80 %:ssa tapauksista suurempi pinta-ala kuin taustajoukosta poimitun satunnaisen joukon muodostaman ROC-käyrän alle (Fielding & Bell 1997). Voidaan myös sanoa, että AUC eli ROC-käyrän alle jäävä alue kertoo todennäköisyyden sille, että satunnainen negatiivinen kohde (jossa laji ei esiinny) ja satunnainen positiivinen kohde luokitellaan oikein (Phillips *ym.* 2006). Perinteisessä ROC-käyrän laskemisessa AUC saa arvon 0,5 - 1,0. AUC = 0,5 on tapaus, jossa mallin ennustamat arvot pitävät yhtä huonosti paikkansa kuin jos ne olisi arvattu

satunnaisesti, ja  $AUC = 1,0$  on tapaus, jossa ennustetut arvot ovat aina oikein (Fielding & Bell 1997). Suurin MaxEnt-ohjelmalla saavutettava AUC-arvo on kuitenkin alle 1 johtuen mallista puuttuvista varmoista negatiivisista "ei esiintymää" -tiedoista (Phillips *ym.* 2006). Kun tutkimuksessa ei ole negatiivisia kohteita, joiden avulla voitaisiin todentaa oikein tunnistettu negatiivinen kohde, vaikuttaisi siltä, että ROC-käyrää ja AUC-arvoa ei voida soveltaa. Negatiiviset kohteet voidaan kuitenkin korvata nimeämällä pseudonegatiivisiksi pisteiksi  $x_{\text{random}}$  kaikki ne tutkimusalueen pikselit, joissa ei ole tunnettua pesäpaikkaa. Vastaavasti nimetään positiivisiksi pisteiksi  $x_{\text{presence}}$  kaikki ne tutkimusalueen pikselit, joissa on tunnettu pesäpaikka (Phillips *ym.* 2006). Kun mallille nyt syötetään satunnaisotos  $x_{\text{presence}}$ -pisteistä ja tasaista tai muuta jakaumaa noudattava satunnaisotos  $x_{\text{random}}$ -pisteistä antamatta mallille tietoa kumman joukon pisteistä on kyse, voidaan  $x_{\text{presence}}$ -pisteet syöttää kaavaan 3 ja  $x_{\text{random}}$ -pisteet kaavaan 4 ROC-funktion luomista varten (Phillips *ym.* 2006). Kun todellisia negatiivisia pisteitä ei ole käytettävissä ja tilalle on otettu tutkimusalueelta satunnaisia pisteitä, AUC:n saavutettavissa oleva maksimi-arvo on

$$AUC_{\text{max}} = 1 - \frac{a}{2}, \quad (5)$$

missä  $a$  on lajin pesäpikseleiden osuus tutkimusalueen kaikista pikseleistä. AUC:n minimiarvo on tässäkin tapauksessa 0,5 (Phillips *ym.* 2006).

Lisäksi MaxEnt tekee jackknife-analyysin, jonka avulla voidaan arvioida muuttujien merkitystä mallissa. Jackknife-analyysin otannassa analyysiin mukaan valitusta muuttujajoukosta jätetään vuorotellen yksi muuttuja pois (McIntosh 2016). Jackknife-analyysi tehdään kolmella eri tavalla, jotka antavat arvot AUC:lle (liite 3), regularized training gainille (kuva 9) ja test gainille (liite 3). Jackknife-analyysi luo mallin moneen kertaan jättäen kullakin kerralla eri muuttujan pois mallista. Tämän jälkeen jackknife-analyysi tehdään kullekin muuttujalle erikseen siten, että maximum entropy -mallissa käytetään vain kyseistä muuttujaa. Lopuksi jackknife-analyysi sovittaa vielä mallin käyttäen kaikkia muuttujia (Phillips 2017).

### 3.2.3 Random Forest-luokitteluanalyysi

Tutkin Random Forest-luokittelumenetelmän avulla, onko pesäpisteiden ja satunnaispisteiden elinympäristöissä eroja eri mittakaavoilla (eri säteillä), ja mitkä ovat kanahaukan kannalta tärkeimmät eroja sisältävät muuttujat. Koska Random Forest-luokittelumenetelmiä on useita erilaisia, kutsun tässä tekstissä alkuperäistä Breimanin *ym.* kehittämää ja Liaw & Wienerin toteuttamalla randomForest-paketilla tehtävää luokitteluanalyysiä randomForest-luokitteluksi pienellä alkukirjaimella, ja yleisesti kaikille Random Forest-luokittelun eri muunnoksille yhteisistä asioista puhuessani käytän RandomForest-nimitystä. Käytin RandomForest-analyysissä Metsäkeskuksen metsävaratietoaineistoa täydennettynä Maanmittauslaitoksen maastotietokannasta ja SYKE:n suojelualueista muodostetulla lisäaineistolla. Esikäsittelin aineistot ArcGIS-ohjelmassa. Laskin pesäpisteiden ja satunnaispisteiden ympäristöstä eri etäisyyksillä (200 m – 4000 m, ks. 3.1.7) metsän rakennetta kuvaavien jatkuva-arvoisten muuttujien pinta-alaosuuksilla painotetut keskiarvot, joidenkin erityisalueiden, kuten pellot, vesistöt ja suojelualueet, pinta- alat, metsän laadun minimi- ja maksimi-arvot ja muiden luokka-asteikollisten muuttujien enemmistössä olevat arvot (taulukko 1).

Taulukko 1. RandomForest-analyysissä käytettyjen muuttujien nimet, lyhenteet ja niille lasketut suureet. Tähdellä merkityt muuttujat ovat olemassa eri puulajeille, tällöin niiden lyhenteen alussa on tuloksissa M=Mänty, K=Kuusi, R=Rauduskoivu, H=Hieskoivu, A=Haapa, L=Leppä. Muuttujista tuli laskea pesä- tai ylipäätään paikkakohtaisia arvoja, jolloin osassa tapauksista tuli valita pinta-alapainotettu keskiarvo, osassa pinta-alojen määrittäminen ja osassa tuli määrittää minimi- ja maksimiarvot tai valita kaikkein yleisimmin esiintyvä arvo.

<b>Muuttujan nimi</b>	<b>Lyhenne</b>	<b>Yhdistämistapa</b>
Puuston keski-ikä, vuotta*	MeanAge	pinta-ala-pain. keskiarvo
Puuston pohjapinta-ala, m <sup>2</sup> /ha*	BasalAr	pinta-ala-pain. keskiarvo
Runkoluku, kpl/ha*	StemCnt	pinta-ala-pain. keskiarvo
Keskiläpimitta, cm*	MeanDmt	pinta-ala-pain. keskiarvo
Keskipituus, m*	MenHght	pinta-ala-pain. keskiarvo
Tilavuus, m <sup>3</sup> /ha*	Volume	pinta-ala-pain. keskiarvo
Tukkipuun tilavuus, m <sup>3</sup> /ha*	SawLogVl	pinta-ala-pain. keskiarvo
Kuitupuun tilavuus, m <sup>3</sup> /ha*	PulpWdVl	pinta-ala-pain. keskiarvo
Vuosittainen tilavuuskasvu, m <sup>3</sup> /v*	VolmGrw	pinta-ala-pain. keskiarvo
Lehvästön biomassa, kg/ha*	LeafBms	pinta-ala-pain. keskiarvo
Hukkapuuosan biomassa, kg/ha*	BrnchBm	pinta-ala-pain. keskiarvo
Metsikön laatu (1 - 9)	StndQlt	minimi- ja maksimiarvo
Kasvupaikkatyyppi	FrtltyC	enemmistö
Pääpuulaji (01 - 28)	MnTrSpC	enemmistö
Pellot	36200	pinta-ala
Vesistöt	32600	pinta-ala
Rakennettu alue	4220	pinta-ala
Suot	35400	pinta-ala
Luonnonsuojeluohjelma-alue	lsohj	pinta-ala
Luonnonsuojelu- tai erämaa-alue	lsysa	pinta-ala
Natura 2000 -alue	natura	pinta-ala

Koska Metsäkeskuksen metsävaratietoaineistossa on mukana puulajikohtaiset tiedot, muuttujien lukumäärä muodostui varsin suureksi (taulukko 1). Tämä ei kuitenkaan haittaa; muuttujien määrää ei välttämättä tarvitse rajoittaa etukäteen esimerkiksi mahdollisen korrelaation perusteella, sillä RandomForest karsii muuttujia ja pyrkii määrittämään kaikki luokittelun eli tässä tapauksessa tarkasteltavan lajin kannalta tärkeimmät muuttujat. Karsimalla muuttujien määrää vain olennaisiin säästetään myöhempiin analyysihin käytettävää aikaa ja saadaan tarkennettua koneoppimisen avulla tuotettujen mallien tarkkuutta (Kursa & Rudnicki 2010).

RandomForest on *bagging*-menetelmä, jossa liitetään metsäksi useita kohtalaisen vähän systemaattista virhettä sisältäviä luokittelupuita ja jolle tehdään ristiinvalidointi ajon aikana. Luokittelupuissa on toisiinsa verrattuna satunnaisvaihtelua, mutta niiden kaikkien käyttäminen yhdessä vähentää lopputuloksen satunnaisvaihtelua, jolloin lopputuloksena saadaan yhdistetty luokittelija, jonka tulosten varianssi on pienempi kuin yksittäisten puiden (Hastie *ym.* 2009, s. 587–603).

RandomForest-menetelmässä poimitaan useita osajoukkoja muuttujista ja havainnoista siten, että alkuperäisestä otoksesta poimitaan satunnaisesti sen kanssa samankokoinen datajoukko, jossa on vain  $m$  selittävää muuttujaa mukana (yleensä  $m \ll p$ , missä  $p$  on selittäjien lukumäärä). Havaintojen poiminta tehdään palauttamalla poimittu havainto aina takaisin perusjoukkoon, jolloin osa havainnoista on mukana useampaan kertaan ja osa jää kokonaan pois. Tällaista uusio-otantamenetelmää kutsutaan *bootstrappingiksi*, ja se eroaa maksimientropiamallinnuksen selittäjäkohtaisessa tarkastelussa käytetystä selittävien muuttujien jackknife-otannasta siten, että tässä valitaan satunnaisesti sekä havaintoja eikä muuttujia, ja poistetun havainnon arvon tilalle tulee aina jokin muu todellinen havainto. Niitä havaintoja, joita ei poimittu mukaan tiettyihin otoksiin, kutsutaan Out-of-bag- eli OOB-näytteiksi. Menetelmään kuuluvassa ristiinvalidoinnissa OOB-näytteillä testataan ilman niitä laadittuja luokittelijoita.

Jokaista satunnaista datajoukkoa kohden RandomForest muodostaa yhden luokittelupuun. Jokainen luokittelupuu tarjoaa ehdotuksen, ”äänen”, luokasta, johon kukin havainto kuuluu. Lopuksi havainnon määritellään kuuluvan siihen luokkaan, mikä saa enemmistön äänistä (Hastie *ym.* 2009, s. 587–603). Tämän tutkimuksen tapauksessa luokat ovat siis ”on pesäpiste”/”ei ole pesäpiste”, ja RandomForest-ajon aikana muodostettavat 500 puuta äänestävät bootstrapping-otannalla valittuja ja tarvittaessa uusiakin havaintoja eri luokkiin.

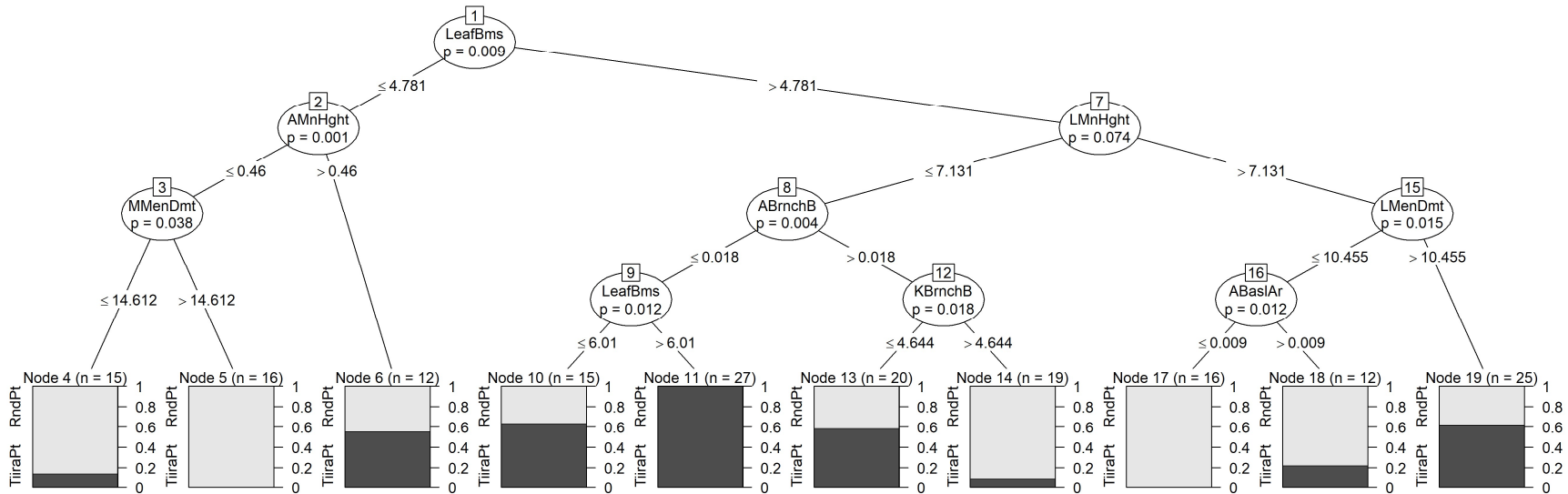
Luokittelupuita muodostettaessa muuttujien järjestys määrittyy sen mukaan, minkä muuttujien perusteella pesäpisteet ja satunnaispisteet saadaan parhaiten erottumaan toisistaan. Muuttujapuita muodostava algoritmi toimii testaamalla



nollahypoteesia (esimerkiksi: puuston tiheydellä ei ole eroa satunnaispisteissä ja tunnetuissa pesäpisteissä) kussakin satunnaisessa datajoukossa olevien muuttujien ja vasteen (on pesäpiste/ei ole pesäpiste) välillä. Sitten valitaan muuttuja, jolla on vahvin p-arvo. Tämän jälkeen havaintojoukko jaetaan kahtia valitun muuttujan sellaisen arvon mukaan, joka eniten parantaa pesäpisteiden ja satunnaispisteiden erottelua toisistaan. Sitten toistetaan muiden muuttujien kohdalle muuttujan valintaa ja kahtiajakoa, kunnes kaikki muuttujat on käyty läpi tai kunnes nollahypoteesia ei enää voida hylätä (kuva 2). Näin muuttujat saadaan tärkeysjärjestykseen pesäpisteiden ja satunnaispisteiden erottelijoina (Hothorn *ym.* 2018). Yksittäisen selittävän muuttujan merkittävyys koko RandomForest-luokittelijalle arvioidaan sen mukaan, kuinka hyvin se onnistuu saamaan eroa pesäpisteille ja satunnaispisteille, verrattaessa tulosta siihen, jos sen muuttujan arvot järjestetään uudelleen satunnaisesti. Tämän testaamiseksi RandomForest tekee z-testin, jonka tulos on yksi muuttujan muuntelun merkitsevyyttä kuvaava arvo (Strobl *ym.* 2009a). Z-testiä varten tarvitaan muuttujien merkitsevyysarvot, jotka lasketaan kaavalla 6 (Strobl *ym.* 2009b):

$$VI(X_j) = \frac{\sum_{t=1}^{ntree} VI^{(t)}(X_j)}{ntree}, \quad (6)$$

missä  $VI^{(t)}(X_j)$  on muuttujan  $X_j$  tarpeellisuus puussa  $t$  (tai 0, jos muuttuja ei ole kyseisessä puussa mukana) ja  $ntree$  on luokittelupuiden lukumäärä. Osoittajassa on siis muuttujan yhteenlaskettu tarpeellisuus kaikissa puissa, ja nimittäjässä luokittelupuiden lukumäärä. Yksittäiset tärkeyspisteet  $VI^{(t)}$  lasketaan  $ntree$  määrän luokittelupuiden bootstrap-näytteiden joukosta, jotka ovat tämän bootstrap-näytteen poimintatavan takia identtisesti jakautuneita ja riippumattomia, mikä mahdollistaa niiden keskivirheen estimoinnin kaavalla  $\frac{\sigma}{\sqrt{ntree}}$ .



Kuva 2. Muuttujapuun visuaalinen kuvaus yhdellä ajolla 4000 m tarkasteluetaisyudellä. Kuvassa näkyy ylipänä ensimmäisenä valittu muuttuja ja sen p-arvo. Lisäksi kuvan solmujen alapuolisissa viivoissa näkyy kyseisen puun muodostamisen aikana määritetyt pesäpisteiden ja satunnaispisteiden erottamisessa käytetyt arvot. Alimpana näkyvät palkit osoittavat pesäpisteiden (TiiraPt, tumma) ja satunnaispisteiden (RndPt, vaalea) osuudet siinä ryhmässä, joka on viimeisessä vaiheessa luokiteltu. Tästä kuvaajasta ei voi kuitenkaan päätellä lopullisia merkittävimpiä muuttujia tai lopullisia merkittäviä muuttujien arvoja, kun RandomForest-tyyppisessä luokittelupuiden metsässä oli tässä tapauksessa 500 luokittelupuuta. Lopullinen luokka-arvo saadaan eri luokittelupuiden antamia tuloksia yhdistelemällä, eli ottamalla yksittäisten puiden antamista tuloksista se, jota useimmiten ennustetaan. Kuvassa Node = luokittelupuun päätössolmu ja sulkeissa oleva  $n$  kertoo kyseiseen päätössolmuun päätyneiden havaintojen lukumäärän.

Standardoitu muuttujan tärkeys, z-testin testisuure  $z(X_j)$ , lasketaan tällöin kaavalla 7:

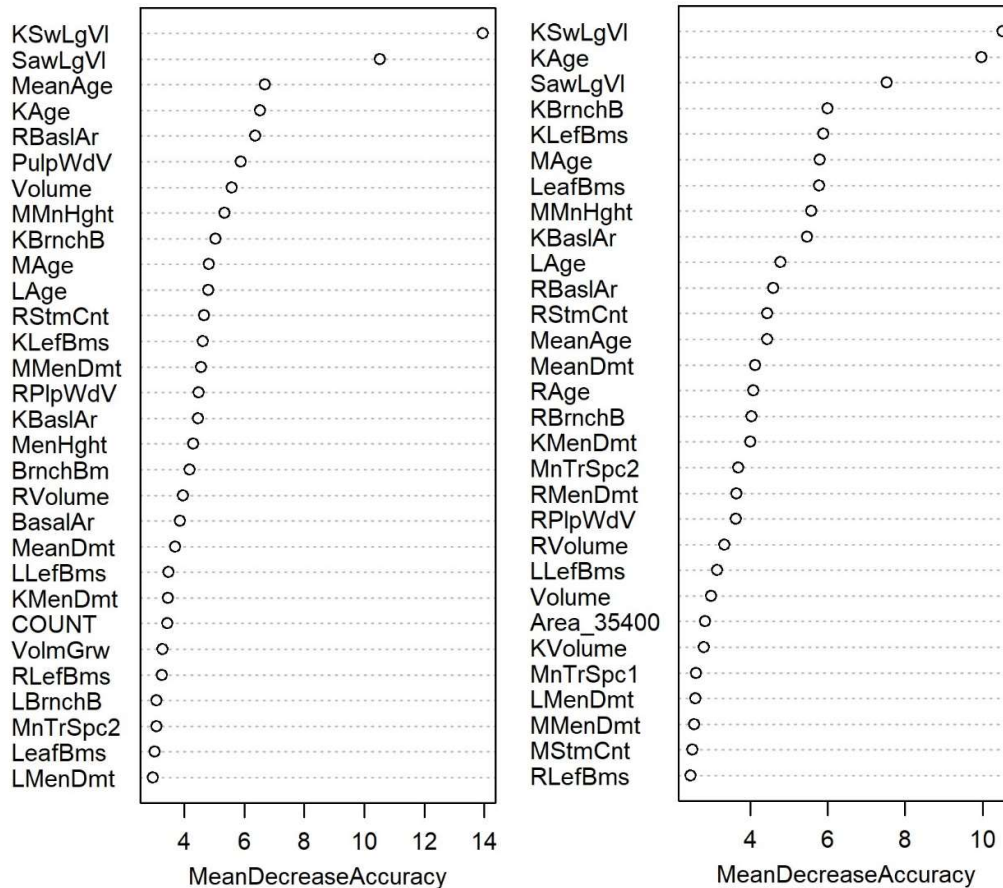
$$z(X_j) = \frac{VI(X_j)}{\frac{\sigma}{\sqrt{ntree}}}, \quad (7)$$

missä osoittajassa on kaavassa 6 laskettu alkuperäinen muuttujan merkitsevyysarvo, ja nimittäjässä on havaintojen keskihajonta jaettuna muuttujapuiden lukumäärän neliöjuurella.

Myöhemmin Kursu ja Rudnicki (2010) kuitenkin toteavat, että alkuperäisen randomForest-algoritmin z-testin testisuure ei ole  $N(0,1)$ -normaalijakautunut. Silti heidän kehittämässään Boruta-algoritminkin käytetään z-testiä, koska se voi tasata keskimääräistä tarkkuuden menetyksen määrää metsän luokittelupuiden välillä (Kursu & Rudnicki 2010).

Tavallisen randomForest-ajon tuloksena sain tarkasteluetäisyyskohtaisia muuttujien merkitsevyyskuvaajat (esimerkki kuvassa 3). Kuvaajissa tärkeimmät muuttujat ovat ylimpänä ja niiden x-akselin merkitsevyysasteikolla kuvattu MeanDecreaseAccuracy-arvo on suurin. Eli jos kyseinen muuttuja jätettäisiin aineistosta kokonaan pois, luokittelijan hyvyys heikkenisi keskimäärin x-akselilla esitetyn määrän verran kokeiltaessa useita tarkasteltavan muuttujan arvojen uudelleenjärjestämiä. Muuttujien järjestys kuvaajassa ja merkitsevyysasteikon skaala voivat muuttua ja käytännössä osin muuttuivatkin erilaisiksi eri tarkasteluetäisyyksillä (kuva 3).

Toisaalta myös OOB-näytteitä käytetään arvioitaessa muuttujien merkitsevyyttä: OOB-virhe-estimaatti on hyvin samankaltainen kuin n-kertaisen ristiinvalidoinnin virhe-estimaatti, joka saatiin maksimientropiamallinnuksessa. OOB-virhe-estimaatin avulla voidaan arvioida luokittelumallin hyvyttä. OOB-virhe-estimaatti muodostetaan tutkimalla, kuinka paljon virheitä valmis puu sisältää luokiteltaessa OOB-näytteet pesäpisteiksi tai satunnaispisteiksi (Hastie *ym.* 2009, s. 587–603).



Kuva 3. Perinteisen randomForest-analyysin tulokset 200 m (vasen) ja 250 m (oikea) säteellä tarkasteltuna. MeanDecreaseAccuracy = randomForest-ajojen keskimääräinen mallin tarkkuuden alenema, kun tarkasteltavan muuttujan arvot permutoidaan eli laitetaan uuteen satunnaiseen järjestykseen. Mitä suuremman MeanDecreaseAccuracy-arvon muuttuja saa, sitä enemmän sen pois jättäminen mallista heikentää mallin tarkkuutta. Kuvasta ei voi päätellä kanahaukan kannalta merkitseviä muuttujan arvoja.

RandomForest-analyysissä käytetään perinteisesti R-ohjelman randomForest-pakettia (Liaw & Wiener 2002). Perinteisessä randomForest-luokittelussa tapahtuu harvoin mallin ylisovittamista, mutta mallin muodostama harha on vain yhtä pieni kuin yksittäisen muuttujapuun harha (Hastie *ym.* 2009, s. 587–603). Lisäksi on näyttöä, että perinteisen randomForest-analyysin käyttäminen yliarvioi keskenään korreloivien muuttujien merkitsevyyttä (Strobl *ym.* 2009a).

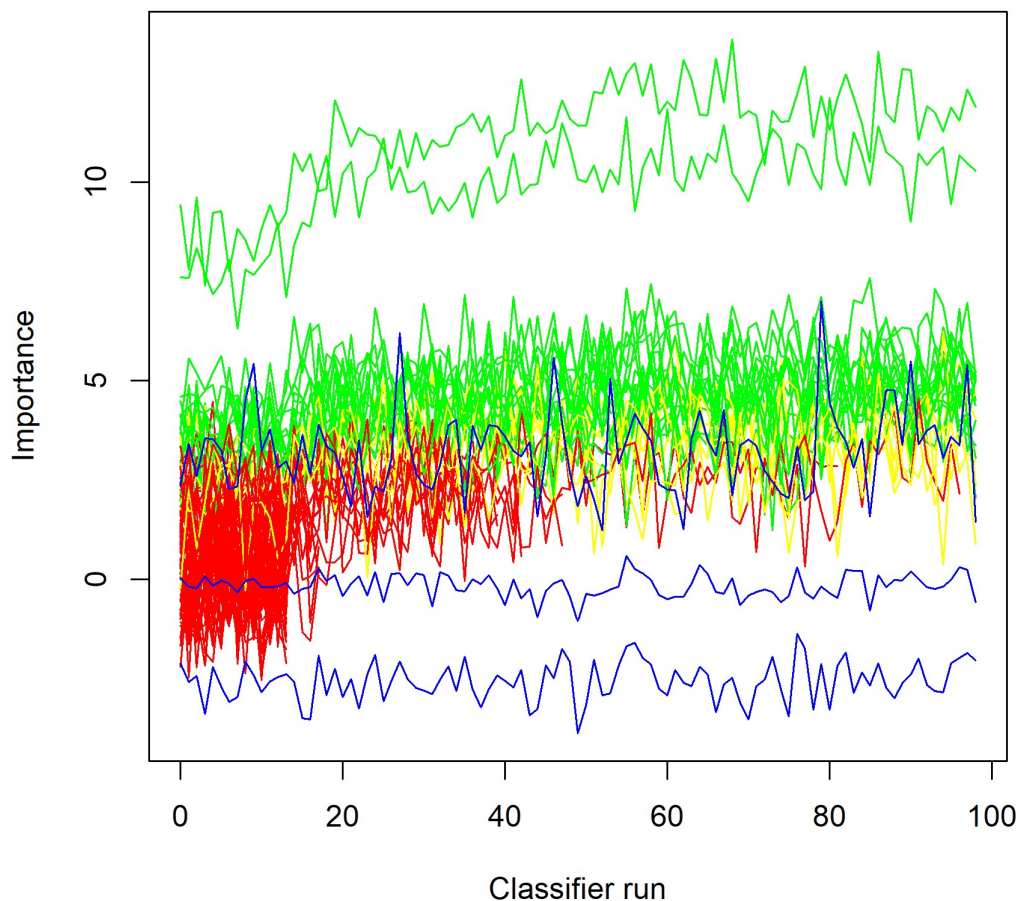
R:n party- paketin cForest-funktio ja Boruta-muuttujiervalintamenetelmä ovat uudempia RandomForest-luokittelun ja muuttujien valinnan muunnoksia. RandomForest-mallin harhaa voidaan mahdollisesti vähentää käyttämällä cForest-menetelmää, joka ei oletuksena käytä bootstrap-otantaa vaan poimii datan osajoukkoja palauttamatta havaintoja perusjoukkoon (Strobl *ym.* 2009a). cForest on randomForestin sovellus, jossa käytetään mallin opetukseen ehdollisia päättelypuita (Hothorn *ym.* 2018). Kun tavallisessa randomForest-analyysissä muuttujan merkitsevyysarviointi tehdään satunnaismuuntelun jälkeen, cForest-menetelmässä merkitsevyysarviointi tehdään tutkimalla muuttujan merkitsevyyden eroa ennen satunnaismuuntelua ja sen jälkeen. Z-testin tuloksen (kaava 7) sijaan cForest käyttää merkitsevyyden arvioinnissa alkuperäistä muuttujan tärkeysarvoa (kaava 6). Lisäksi ennen muuttujien merkitsevyyden arviointia tehdään selittävien muuttujien korrelaatioiden tutkiminen ja otetaan tämä huomioon arvioitaessa muuttujien merkitsevyyttä. Tätä kutsutaan ehdolliseksi merkitsevyydeksi. Tällöin esimerkiksi puuston keskipituuden ja puuston iän välinen korrelaatio otetaan huomioon, jolloin on odotettavissa, että kanahaukalle konkreettisesti merkitsevien muuttujien, kuten puuston keskipituuden tai tilavuuden merkitsevyys verrattuna puuston ikään korostuu. Muuten cForestin luokittelupuun rakentaminen tapahtuu samalla tavalla kuin perinteisessä randomForest-menetelmässäkin (Strobl *ym.* 2009a).

cForestin muodostama muuttujien tärkeyttä osoittava kuvaaja on histogrammimuotoinen (liite 4). cForestille ominaista on, että muuttujat, joiden tärkeys on pieni, saavat tärkeysarvoja, jotka vaihtelevat nollan molemmin puolin.

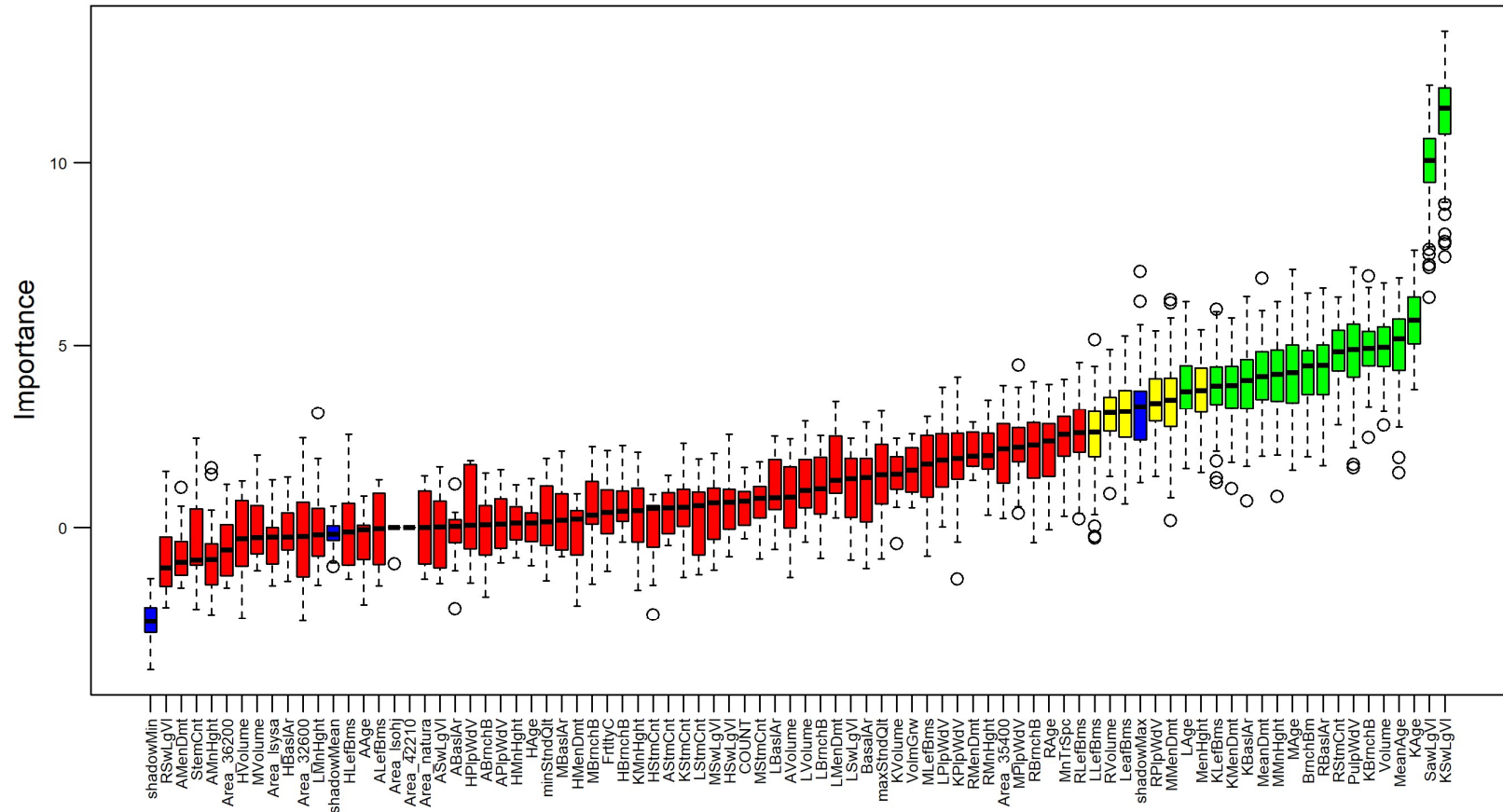
Boruta on muuttujanvalintamenetelmä, jonka avulla saadaan selville RandomForest-luokittelijan merkitsevimmät muuttujat. Boruta tekee randomForest-tyyppisiä luokitteluja ja niiden selittäjille jälkianalyysiä. Boruta tuottaa tyypillisesti tarkempia ja varmempia tuloksia selittäjien merkityksestä kuin perinteinen randomForest-analyysi. Oletuksena Boruta käyttää syötteenään randomForestin antamia muuttujan merkitsevyysarvoja. Kuten perinteinen randomForest-menetelmäkin, myös Boruta käyttää muuttujan merkitsevyyden

arvioinnissa z-testin tulosta (kaava 7). Boruta kuitenkin vertailee muuttujien onnistumista luokittelussa verrattuna varjomuuttujiin, joita luodaan jokaiselle selittävälle muuttujalle omansa. Varjomuuttujat luodaan sekoittamalla alkuperäisen muuttujan arvot satunnaiseen järjestykseen. Tämän myötä varjomuuttujat korreloivat tutkittavan vasteen (pesäpiste/satunnaispiste) kanssa vain sattumalta. Kun varjomuuttujat on lisätty muiden muuttujien joukkoon, ajetaan randomForest-analyysi ja saadaan muuttujakohtaiset z-testin tulokset. Tämän jälkeen verrataan parhaat pisteet saaneen varjomuuttujan ja kunkin varsinaisen muuttujan z-testin tuloksia kaksisuuntaisella yhtäsuuruustestillä. Kun muuttujan suoriutuminen luokittelusta alkaa jäädä säännöllisesti huonommaksi kuin parhaalla varjomuuttujalla, se jätetään pois luokittelusta ja luokittelua jatketaan jäljelle jääneillä muuttujilla (kuva 4). Mikäli muuttujan suoriutuminen on säännöllisesti parempaa kuin varjomuuttujalla, se määritellään merkitseväksi muuttujaksi. Boruta lopettaa luokitteluajot, kun vain merkitseviä muuttujia on jäljellä tai kun ennalta asetettu luokitteluajojen määrä on täyttynyt. Boruta tekee useita vertailuajoja, ja jokaiseen vertailuajoon varjomuuttujat luodaan uudestaan. Boruta-menetelmällä saadaan muuttujien muuntelulle normalisoitu tärkeysjärjestys (Kursa & Rudnicki 2018).

Lopulta Borutan tuloksista saadaan kuvaaja, jossa on muuttujien merkitsevyydet ja keskivirheet. Esimerkkikuvasta 5 nähdään yksilöityinä samat muuttujat kuin kuvassa 4, jonka kaksi ylintä käyrää ovat kuvan 5 oikeanpuoleisimmat muuttujat.



Kuva 4. Muuttujien (200 m säteellä pesä- tai satunnaispisteestä) tärkeys luokittelijana pesä- ja satunnaispisteiden erottelussa sadan luokitteluajon aikana. Sinisellä on kuvattu varjoluokittelijoista vain kolme: huonoin (alin), keskimäinen ja paras. Punaisella on merkitty huonoiten suoriutuneet varsinaiset muuttujat, keltaisella keskiverrot ja vihreällä merkitsevimmät muuttujat. Kuvaajasta nähdään, että kaksi muuttujaa (ylimmät vihreät käyrät) ovat selvästi merkitsevempiä kuin muut.



Kuva 5. Muuttujien tärkeys pesäpisteitä ja satunnaispisteitä erottelevana luokittelijana, kun tarkastellaan pesien ympäristöä 200 m säteellä. Parhaiten luokittelevat muuttujat ovat kuvaajassa oikealla. Siniset ovat huonoimman, keskimmäisen ja parhaan varjomuuttujan tulokset, punaiset huonoiten, keltaiset kohtalaisesti ja vihreät parhaiten luokittelevia muuttujia.



## 4 TULOKSET

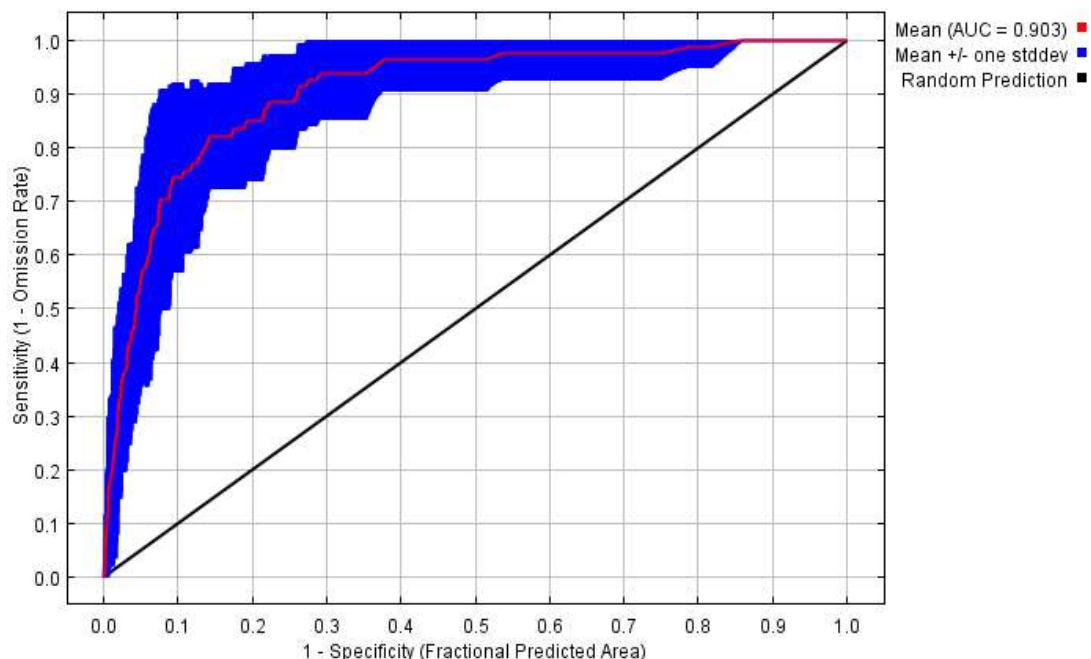
### 4.1 Rasteriaineiston solukoko maksimientropiamallissa

Alustava tarkastelu tehtiin kahdelle eri muuttujakokoonpanolle siten, että mukana olivat kaikki taulukon 1 muuttujat, paitsi toisesta kokoonpanosta oli jätetty pois kuusen tilavuus ja toisesta puuston kokonaistilavuus. Alustavan tarkastelun mukaan kuusen tilavuuden sisältävän maksimientropiamallin AUC-arvo oli suurempi kuin puuston kokonaistilavuuden sisältävän mallin AUC-arvo kummallakin solukoolla (liite 6). Koska kuusen tilavuus oli hieman parempi selittäjä, tässä tarkastellaan kuusen tilavuuden sisältävän mallin tuloksia puuston kokonaistilavuuden sisältävän mallin tulosten sijaan. Solukoolla 80 m tarkasteltuna kanahaukalle erinomaisesti sopivaa metsää oli 3,7 % tutkimusalueen maapinta-alasta, tyypillistä kanahaukkametsää oli 4,4% ja huonoa oli 91,9 %. Solukoolla 160 m vastaavat osuudet olivat: erinomainen 4,1 %, tyypillinen 4,9 % ja huono 91,0 % (taulukko 2).

Taulukko 2. Kanahaukalle sopivan elinympäristön osuus maapinta-alasta Keski-Suomessa maksimientropiamallin perusteella.

<b>Cloglog- muunnettu todennäköisyys</b>	<b>Soveltuvuus kanahaukalle</b>	<b>Osuus pinta-alasta (%), 80 m solukoko</b>	<b>Osuus pinta-alasta (%), 160 m solukoko</b>
0,69 - 1,00	Erinomainen	3,73	4,11
0,46 - 0,69	Tyypillinen	4,35	4,87
0,00 - 0,46	Huono	91,92	91,02

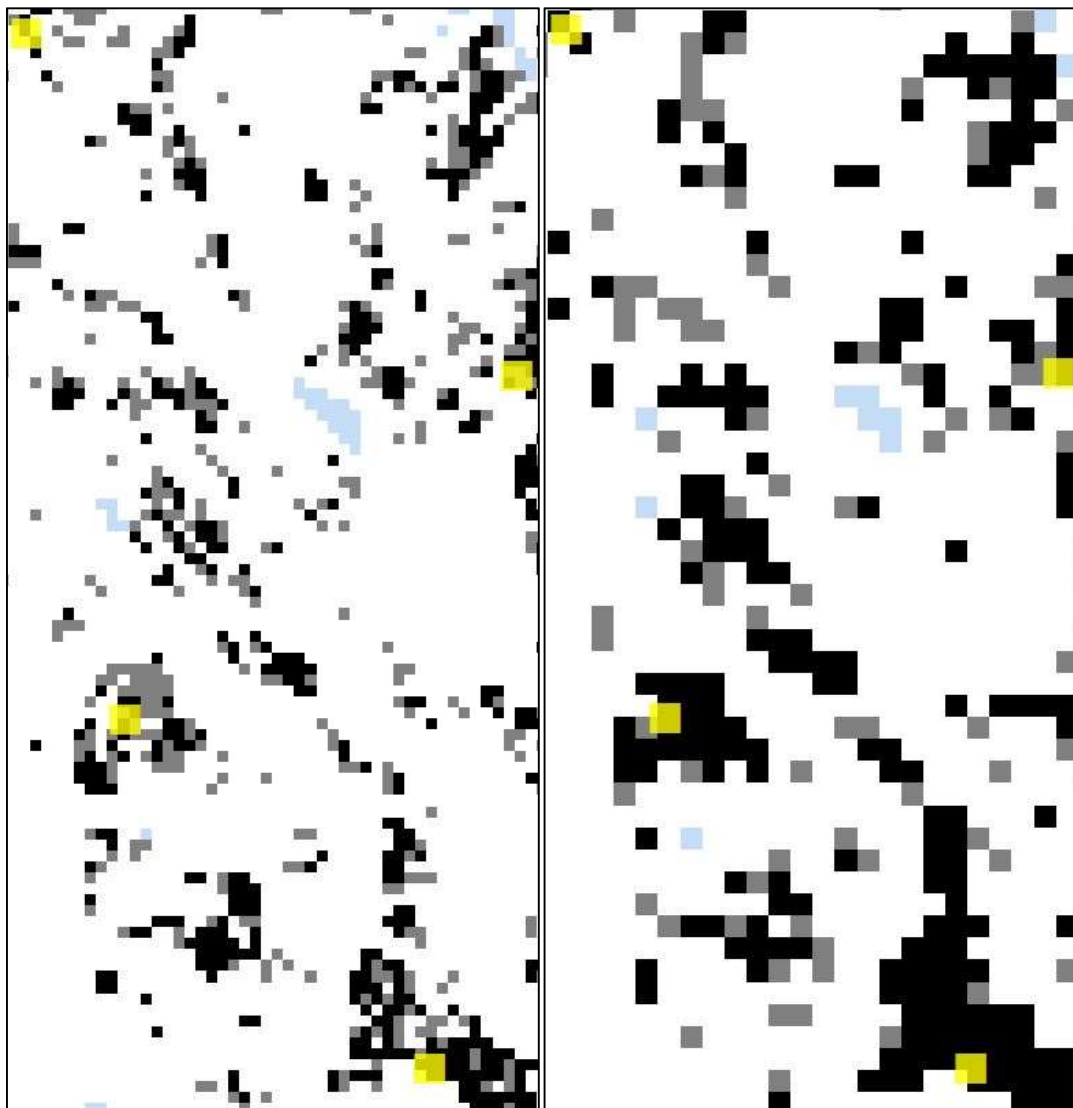
Solukoolla 80 m × 80 m tehdyn kuusen tilavuuden sisältävän mallin testiaineiston ROC-käyrän AUC-arvo oli 0,895 (liite 6) ja vastaavan 160 m × 160 m solukoonanalyysin testiaineiston ROC-käyrän AUC-arvo oli 0,903 (kuva 6). Täten 80 m × 80 m analyysin AUC-arvo oli hyvä ja 160 m × 160 m analyysin AUC-arvo oli erinomainen (Swets 1988). Voidaan siis todeta, että 160 m solukoolla malli selitti kanahaukalle soveltuvien elinympäristöjen esiintymistodennäköisyydet luotettavammin.



Kuva 6. ROC-käyrä solukoon  $160\text{ m} \times 160\text{ m}$  kuusen tilavuuden sisältävälle aineistolle. Mikäli malli suoriutuisi lajin esiintymispisteiden sijainnin ennustamisesta virheettömästi, olisi punainen käyrä suorakulmainen, ja se kulkisi pisteestä (0,0) pisteen (0,1) kautta pisteeseen (1,1). Musta viiva kuvaa tilannetta, jossa ennuste olisi täysin satunnainen. Sininen alue kuvaa eri MaxEnt-ajojen tulosten vaihteluväliä.

Tarkasteltaessa lähikuvaa eri solukokojen MaxEnt-ajojen myötä syntyneistä ennustekartoista (kuva 7) havaitaan, että solukoolla  $80\text{ m}$  kanahaukalle sopivia elinympäristöjä havaitaan hienopiirteisemmin – kaikki lajille todennäköiseksi merkityt paikat eivät kuitenkaan välttämättä ole kanahaukalle oikeasti soveltuvaa metsää pienialaisuutensa takia. Ennusteeseen tulee siten runsaammin vaihtelua, joka näkyy kuvassa ”kohinana”. Sen sijaan solukoon  $160\text{ m}$  ajolla saadut erinomaisesti sopivien elinympäristöjen alueet ovat suuria ja yhtenäisempiä (kuva 7). Lisäksi tällä suuremmalla solukoolla VMI-aineiston yksittäisten solujen mahdollisesti sisältämien epätarkkuuksien merkitys vähenee, joten kartta tuo varmemmin esille kanahaukalle erityisen hyvin soveltuvat alueet, ja siten myös mahdolliset elonkirjokohteet. Kuvasta 7 nähdään myös, että

molempien solukokojen mallit ennustavat tiedossa oleviin kanahaukan pesäpisteisiin hyvää tai erinomaista elinympäristöä.



Kuva 7. Kanahaukan pesäpaikan esiintymistodennäköisyys pienellä alueella Keski-Suomessa kuusen tilavuuden ja muut muuttajat (taulukko 3, ei kuitenkaan puuston kokonaistilavuus) sisältävän MaxEnt-analyysin mukaan. Vasemmalla solukoolla 80 m luotu kartta ja oikealla solukoolla 160 m luotu kartta. Musta = lajille erityisen hyvin sopiva elinympäristö (todennäköisyys sopivuudelle kanahaukan pesämetsäksi 0,69–1,0, ks. taulukko 2), harmaa = tyypillinen kanahaukkametsä (0,46–0,69), valkoinen = lajille huonosti sopiva elinympäristö (0,0–0,46), vaaleansininen = ei MVM-tietoa (vesistöt). Keltaisella kuvattuna neliöt, joiden sisällä tunnetut kanahaukan pesät sijaitsevat.

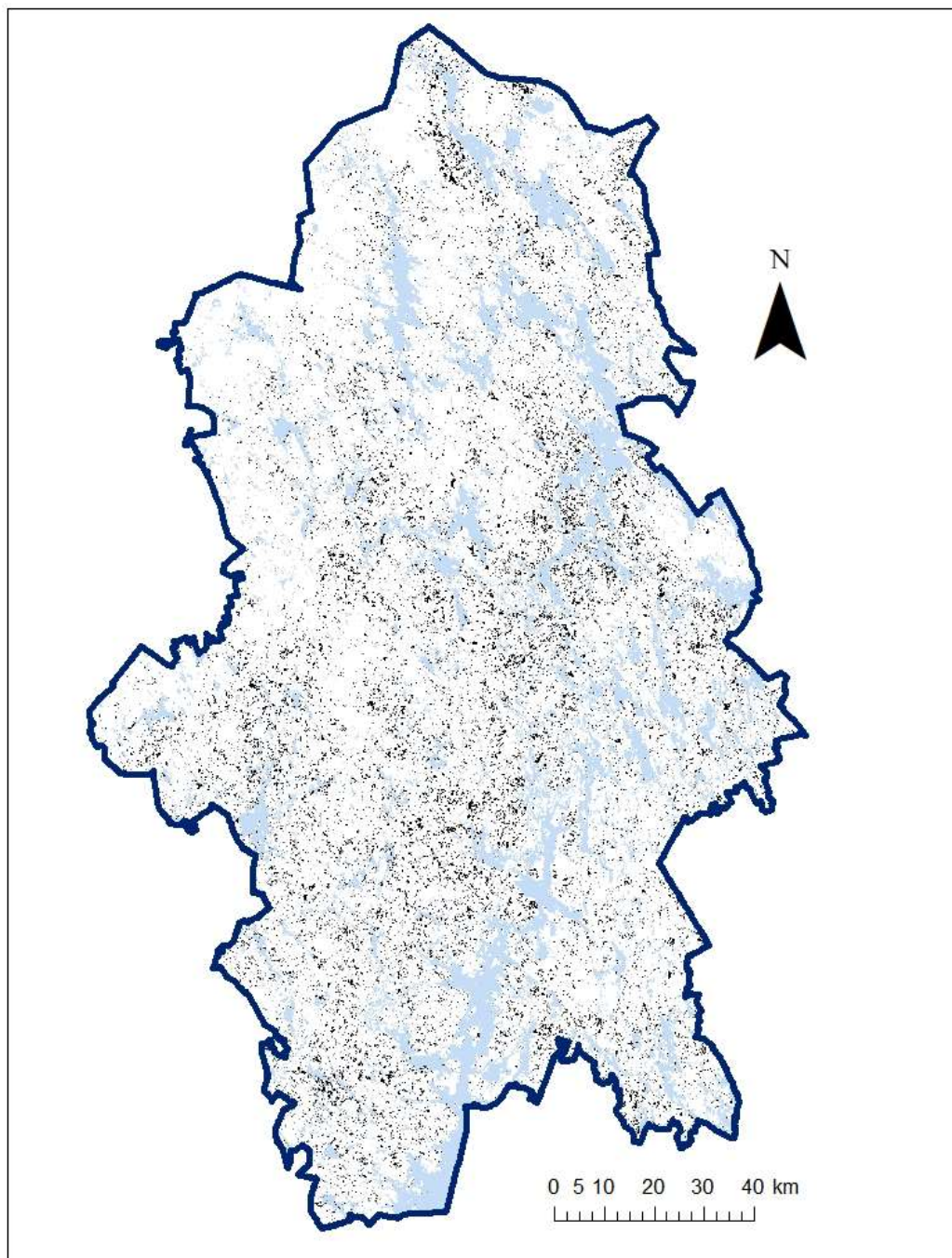
Pienemmällä solukoolla maisemasta paikantuu pienialaisiakin kanahaukalle soveltuvia kohteita, kun taas isommalla solukoolla paikannetaan varmemmin kanahaukalle todellisuudessa sopivat metsäalueet. Tarkastelen siten jatkossa vain kuusen tilavuuden 160 metrin solukoolla sisältävän mallin tuloksia.

Maksimientropiamallinnuksessa syntyy tuloksena kartta, jossa on esitetty todennäköisyys kanahaukalle soveltuvan elinympäristön esiintyvyydelle pikseleittäin (kuva 8). Kartassa kanahaukalle erinomaisesti soveltuvat pesäpaikat on kuvattu mustalla ja tyypillinen haukkametsä harmaalla.

#### **4.2 Haukkametsän piirteet MaxEnt-analyysin mukaan**

Maksimientropia-analyysin mukaan kanahaukan pesäpaikkojen kohdalla kuusen tilavuus sai systemaattisesti kohtalaisen suuria arvoja (kuva 10). Se oli selvästi vahvin kanahaukalle sopivan elinympäristön sijaintia yksinään selittävä tekijä (selittävien muuttujien nimet taulukossa 3, tulokset kuvassa 9). Jackknife-kuvaajista nähdään, että kuusen tilavuuden pois jättäminen heikensi mallia huomattavasti enemmän kuin minkään muun muuttujan pois jättäminen (kuva 9+liite 5). Kuusen tilavuus -muuttuja sisälsi siis eniten sellaista tietoa, jota muut muuttujat eivät pystyneet korvaamaan.

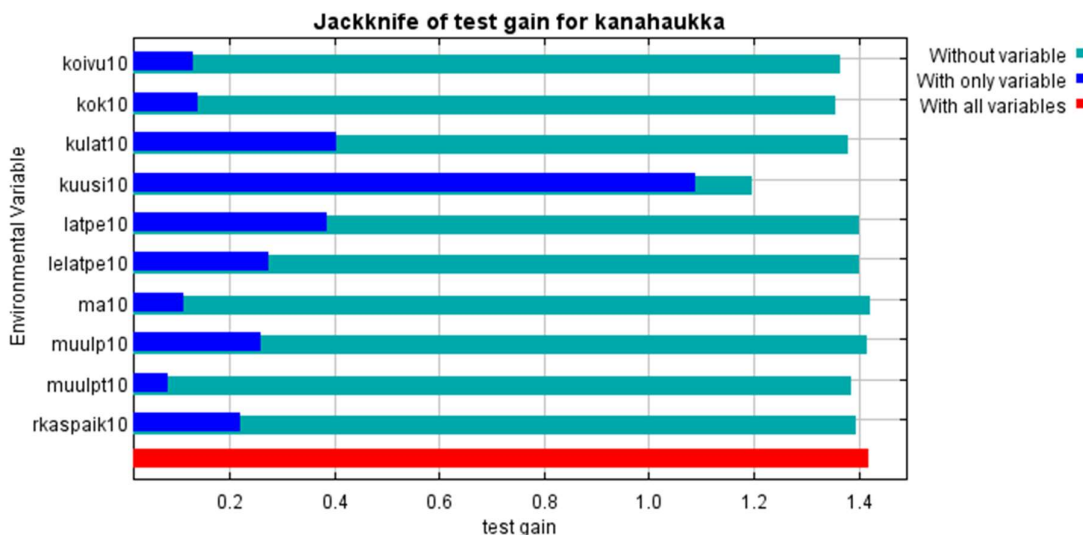
Seuraavaksi eniten mallia heikentäisivät koivun tilavuuden, koivukuidun tilavuuden ja muun lehtipuun tukkitilavuuden pois jättäminen (turkoosit pylväät, kuva 9). Nämä eivät kuitenkaan jackknife-analyysin mukaan ole kovin hyviä selittäjiä yksinään. Kuusen hukkapuuosuudenkin pois jättäminen heikentää mallia kohtalaisen paljon, ja se taas on yksinäänkin merkitsevä muuttuja. Mallin kannalta merkitsevien muuttujien löytyminen ei tarkoita, että näiden muuttujien kaikkein suurimmat arvot olisivat aina parhaita kanahaukalle, vaan että niille on löydettävissä ylipäätään joitakin sellaisia arvoja, jotka vaikuttavat kanahaukalle soveltuvan elinympäristön esiintymistodennäköisyyteen.



Kuva 8. Alueiden soveltuminen kanahaukan pesäpaikaksi Keski-Suomessa solukoolla 160 m. Musta = suuri todennäköisyys sopivalle elinympäristölle (4,11 % Keski-Suomen maapinta-alasta), harmaa = tyypillinen kanahaukkametsä (4,87 %), valkoinen = pieni todennäköisyys (91,02 %, taulukko 2), vaaleansininen = ei MVM-tietoa (vesistöt tai rak. alue), tummansininen = Keski-Suomen maakuntarajat (poimittu hallintoraja-aineistosta, © Maanmittauslaitos 2017).

Taulukko 3. Maksimientropiamallinnuksen muuttujista käytetyt lyhenteet. Hukkapuuosa on rungon tukiksi tai kuiduksi kelpaamaton osuus.

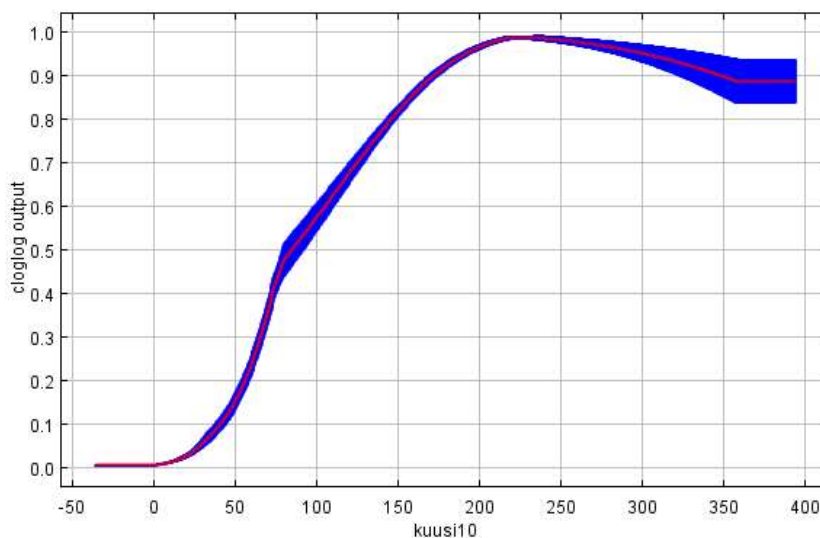
Lyhenne	Muuttujan nimi (yksikkö)
kuusi	Kuusen tilavuus (m <sup>3</sup> /ha)
tila	Puuston kokonaistilavuus (m <sup>3</sup> /ha)
lelatpe	Lehtipuiden latvuspeittävyys (%)
rkaspaik	Kasvupaikka (luokat 1–10)
muulpt	Muiden lehtipuiden tilavuus, tukkipuu (m <sup>3</sup> /ha)
kok	Koivun tilavuus, kuitupuu (m <sup>3</sup> /ha)
muulp	Muiden lehtipuiden tilavuus (m <sup>3</sup> /ha)
kulat	Kuusen hukkapuuosan tilavuus (10 kg/ha)
koivu	Koivun tilavuus (m <sup>3</sup> /ha)
ma	Männyn tilavuus (m <sup>3</sup> /ha)
latpe	Puuston latvuspeittävyys (%)



Kuva 9. MaxEnt-mallin muuttujien merkitsevyysarvioinnin tulokset 160 m × 160 m solukoolla kuusen tilavuuden sisältävällä mallilla. Muuttujien nimet on esitetty taulukossa 3. Punainen = mallin kokonaisselittävyys. Turkoosi = mallin selittävyys, mikäli kyseinen muuttuja jätetään pois analyysistä. Siten lyhyempi turkoosi pylväs kertoo, että muuttuja on mallissa tärkeä (esim. kuusi). Sininen = mallin selittävyys, mikäli käytetään vain kyseistä muuttujaa. Tämä osoittaa kuusen yksinäänkin olevan kohtuullisen hyvä selittäjä hukanpesien esiintyvyydelle.

Toisena tulosten havainnollistuksena saadaan muuttujakohtaisia vastekuvaajia. Ensimmäisessä vastekuvaajatyyppissä (kuva 10, kuva 11) mallissa on käytetty vain yhtä muuttujaa kerrallaan ja tarkastellaan, miten mallin ennuste muuttuu valitun muuttujan arvojen mukaan (Phillips 2017).

Kuvasta 10 nähdään, että kanahaukalle suotuisten olosuhteiden esiintymistodennäköisyys on vielä varsin pieni (n. 0,15), kun kuusen tilavuus on  $50 \text{ m}^3/\text{ha}$ , mutta esiintymistodennäköisyys kasvaa jyrkästi tilavuuden kasvaessa arvoon  $200 \text{ m}^3/\text{ha}$ . Järeä kuusimetsä on siis tyypillistä kanahaukan pesäpaikoille, sillä kuusen tilavuus  $150 \text{ m}^3/\text{ha}$  ja sitä suuremmat arvot ovat kanahaukalle suotuisimpia sopivan pesämetsän esiintymistodennäköisyyden ollessa tuolloin yli 80 % (kuva 10).

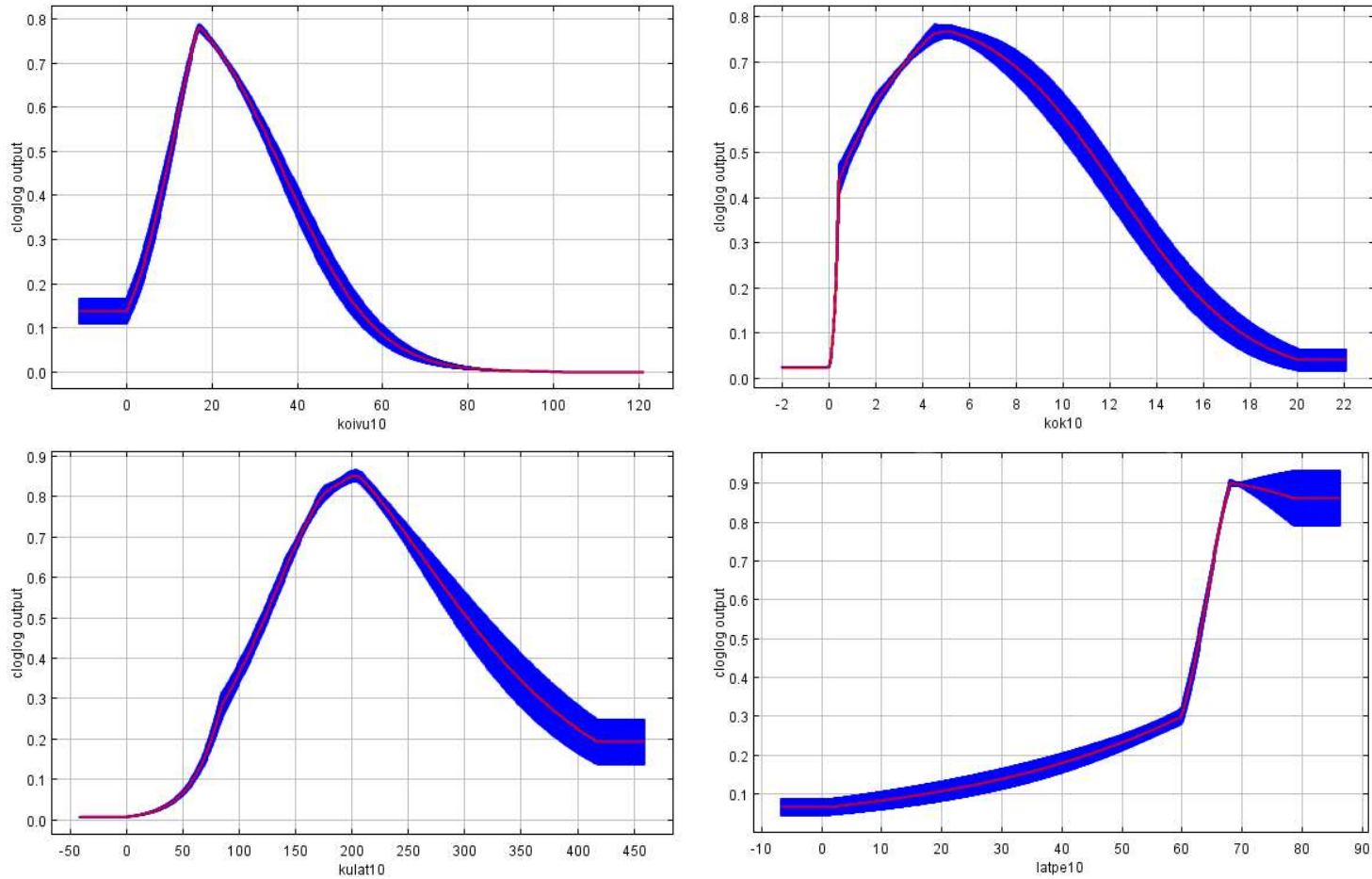


Kuva 10. Kanahaukalle suotuisan ympäristön todennäköisyyden keskiarvo (punainen viiva, y-akseli) kuusen tilavuuden ( $\text{m}^3/\text{ha}$ , x-akseli) suhteen solukoon  $160 \text{ m} \times 160 \text{ m}$  aineistolla, kun mallissa on mukana vain kuusen tilavuus. Sininen kuvaa arvojen keskihajontaa kymmenen toiston aikana. Elinympäristön sopivuus on korkea ( $>0,8$ ) tilavuudesta  $150 \text{ m}^3/\text{ha}$  alkaen.

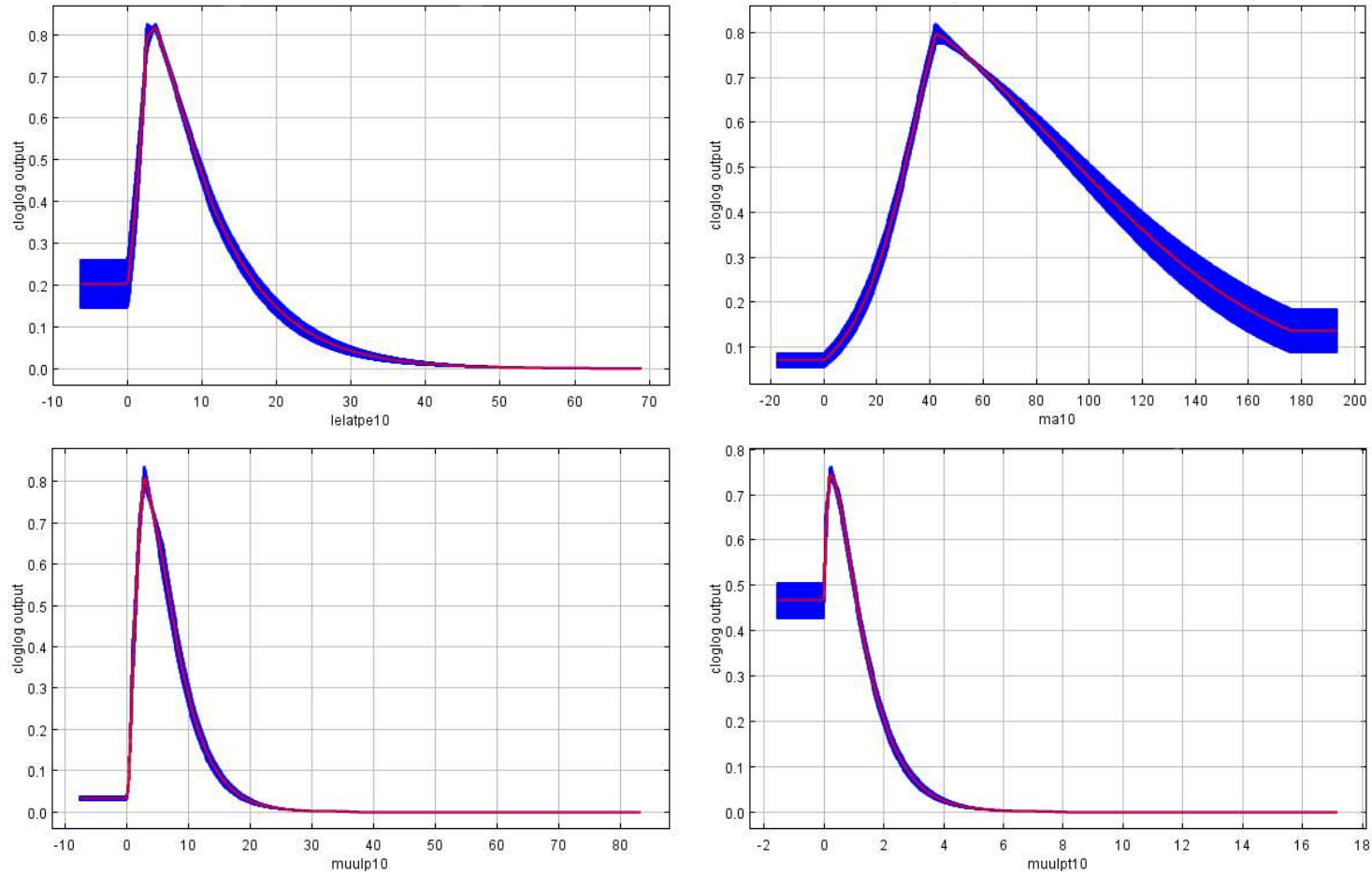
Kanahaukan pesän esiintymistodennäköisyys on suurin, kun kuusen hukkapuuosan biomassassa on noin  $200 \text{ m}^3/\text{ha}$  (kuva 11). Kuvasta 11 nähdään myös, että latvuspeittävyden ylittäessä 60 % lähtee kanahaukalle suotuisan metsän esiintymistodennäköisyys jyrkkään nousuun. Kanahaukalle suotuisan metsän esiintymistodennäköisyys on 0,9 latvuspeittävyden ollessa 68 %, jonka jälkeen esiintymistodennäköisyys alkaa loivasti laskea. Pesämetsäksi sopivimmilla paikoilla lehtipuun latvuspeittävyys on puolestaan alle 5 %, mutta lehtipuuta kuitenkin on sekapuustona, sillä lehtipuuston latvuspeittävyden ollessa suurempi

kuin nolla saa kanahaukalle sopivan pesimisympäristön todennäköisyys nolaa suurempia todennäköisyyksiä (kuva 12). Muun lehtipuun tilavuus -muuttujan suhteen saa kanahaukan pesäpaikalle sopivan metsän esiintymistodennäköisyys suurimman arvonsa, yli 0,8, muuttujan arvolla 3 m<sup>3</sup>/ha. Muun lehtipuun tukkitilavuuden suhteen esiintymistodennäköisyyden suurin arvo saavutetaan muuttujan arvolla 2 m<sup>3</sup>/ha. Näiden muuttujien kuvaajat ovat selvästi piikittyneitä, eli kaikista suurin esiintymistodennäköisyys saavutetaan hyvin kapealla muuttujien arvojen vaihteluvälillä (kuva 12). Lisäksi koivua oli sekapuustona kanahaukalle erinomaisesti sopivissa elinympäristöissä, sillä kanahaukalle sopivan elinympäristön esiintymistodennäköisyys saa suurimman arvonsa koivun tilavuuden suhteen arvolla 18 m<sup>3</sup>/ha ja koivukuidun suhteen arvolla 5 m<sup>3</sup>/ha. Männyn tilavuuden ylitettyä 40 m<sup>3</sup>/ha alkaa kanahaukan esiintymistodennäköisyys laskea (kuva 12). Mäntyä kuitenkin saa olla sekapuustona, sillä millään aineistosta löytyvällä männyn tilavuuden arvolla ei kanahaukan esiintymistodennäköisyys laske alle sen, mitä se on männyn tilavuuden ollessa nollassa. Kasvupaikka on käytetyn aineiston ainoa luokkamuuttuja. Kasvupaikkamuuttujan numeerisia arvoja vastaavat kasvupaikkojen nimet on kerrottu liitteessä 2. Sen luokista erityisesti luokka 3 eli tuoreet kankaat (Korhonen 2009) olivat kanahaukan pesäpaikkana tyypillisiä (kuva 13).

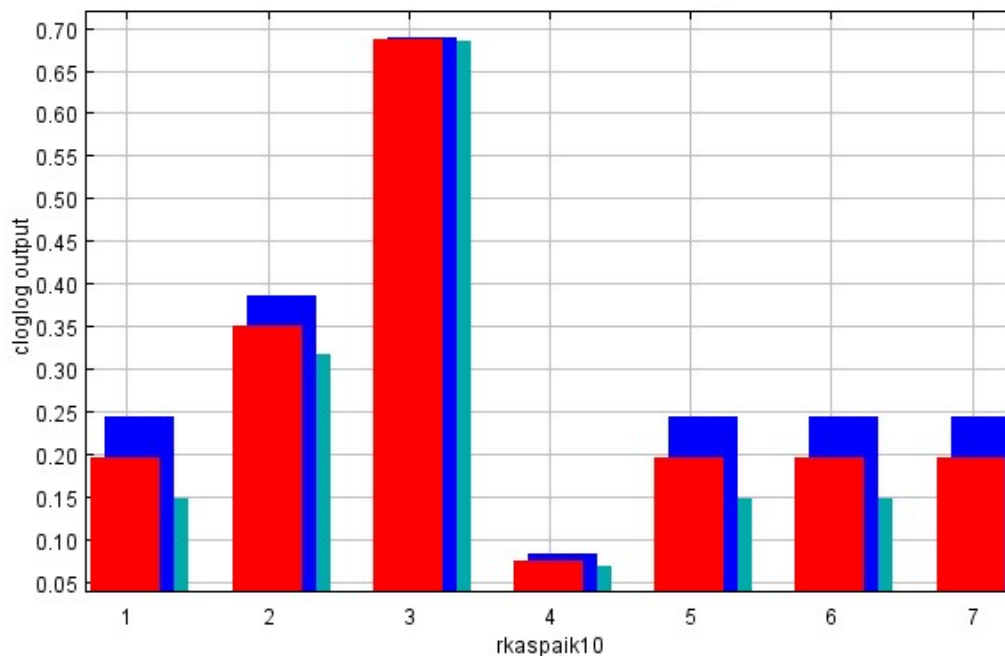




Kuva 11. Vastekuvaajat solukoolla  $160 \text{ m} \times 160 \text{ m}$ , kun kuvatun muuttujan arvoa muutetaan ilman, että muita muuttujia on mallissa mukana. Ylhäällä vasemmalta koivun tilavuus ja koivun kuitupuun tilavuus, alhaalla vasemmalta kuusen hukkapuosuuden tilavuus ja koko puuston latvuspeittävyys.



Kuva 12. Vastekuvaajat solukoolla  $160 \text{ m} \times 160 \text{ m}$ , kun kuvatus muuttujan arvoa muutetaan ilman, että muita muuttujia on mallissa mukana. Ylhäällä vasemmalta lehtipuuston latvuspeittävyys ja männyn tilavuus, alhaalla vasemmalta muun lehtipuun tilavuus ja muun lehtipuun tukkipuun tilavuus.



Kuva 13. Vastekuvaaja kasvupaikkatyyppi-muuttujalle solukoolla 160 m × 160 m. Kuvaaja esittää haukkametsän tapauksessa todettuja kasvupaikkatyyppin arvoja. Selitteet kasvupaikkatyyppin arvoille on esitetty liitteessä 2. Punaisella on kuvattu MaxEnt-ajojen myötä saatu clog-log- muunnetun muuttujan merkitsevyyden keskiarvo, tumman- ja vaaleansiniset pylväät ovat keskiarvo ± keskihajonta.

Kuvan 10 tyyppisistä kuvaajista voidaan myös päätellä jonkin verran muuttujien keskinäisistä suhteista. Kuvista 10 ja 12 nähdään, että vaikka kanahaukalle suotuisten olosuhteiden esiintymistodennäköisyys saavuttaa korkeimman arvonsa (n. 0,78) männyn tilavuuden suhteen männyn tilavuuden ollessa vasta n. 40 m<sup>3</sup>/ha, saavuttaa esiintymistodennäköisyys vielä korkeamman arvon (lähes 1,0) kuusen tilavuuden ollessa 220 m<sup>3</sup>/ha. Niinpä voidaan päätellä, että tyyppillisesti kanahaukalle erinomaisesti soveltuvalla alueella on runsaasti kuusta, mutta usein jonkin verran myös mäntyä. Samankaltainen päätelmä kanahaukalle soveltuvimmista metsistä saadaan latvuspeittävyden perusteella. Koska lehtipuiden latvuspeittävyden optimi on 5 % ja männyn tilavuuden optimiarvo on varsin matala, on suuri osa latvuspeitosta tällöin kuusta. Varsin yhtenäinen tuoreen kankaan kuusivaltainen järeä metsä, jossa on sekapuustona mäntyä ja lehtipuita, soveltuu siis tämän mallin perusteella parhaiten kanahaukan pesimisympäristöksi.

Toisessa vastekuvaajatyypissä (liite 3) tarkastellaan, miten mallin ennuste muuttuu, kun valitun muuttujan arvoa muutetaan ja muiden muuttujien arvot pidetään keskiarvossaan (Phillips 2017). Tällaisten kuvaajien tulkinnessa pitää muistaa, että ne ovat yksinkertaistettuja versioita mallista, sillä mallin luonnissa on voitu käyttää hyväksi useita muuttujia, kun kuvaajassa esitetään vain yksittäisen muuttujan muuntelun vaikutusta (Phillips 2017).

Edellä mainittujen kuvaajien lisäksi MaxEnt antaa myös numeeriset estimaatit muuttujien vaikutuksesta malliin. Taulukossa 4 sarake "Osuus mallista" kertoo, miten suuri osuus mallista on selitettävissä yksittäisellä muuttujalla, ja sarake "Mallin heikkeneminen" kertoo, miten paljon malli heikkenee, jos kyseisen muuttujan arvot sekoitetaan satunnaiseen järjestykseen (Phillips 2017). Taulukosta 4 nähdään, että kuusen tilavuus on merkitsevin yksittäinen muuttuja, ja lehtipuuston latvuspeittävyys on toiseksi merkitsevin. Lehtipuun latvuspeittävyys tai koivun tilavuuden pois jättäminen heikentäisivät mallia varsin paljon, seuraavaksi eniten kuusen tilavuuden pois jättämisen jälkeen, mutta varsinkin koivun tilavuuden osuus mallista on hyvin pieni verrattuna siihen, kuinka paljon sen pois jättäminen heikentäisi mallia. Taulukon 4 numeroarvoista ei voi kuitenkaan päätellä, ovatko muuttujan suuret vai pienet arvot kanahaukan kannalta merkitseviä. Nämä muuttujan arvot selviävät edellä esitetyistä kuvista 10–13.

Taulukko 4. Numeeriset estimaatit muuttujien vaikutuksesta malliin solukoolla 160 m kuusen tilavuuden sisältävällä mallilla. Taulukossa sarake ”osuus mallista” kertoo, miten suuri osuus mallista on selitettävissä kyseisellä muuttujalla, ja sarake ”mallin heikkeneminen” kertoo, miten paljon malli heikkenee, jos kyseiselle muuttujalle annetaan satunnaisia arvoja.

<b>Muuttuja</b>	<b>Osuus mallista (%)</b>	<b>Mallin heikkeneminen (%)</b>
kuusi	72,7	64,6
lelatpe	7,7	12,0
muulpt	6,7	3,4
kaspaik	5,4	1,5
kok	2,1	3,6
kulat	1,6	0,9
muulp	1,4	1,5
koivu	1,2	9,9
latpe	0,8	1,8
ma	0,5	0,6

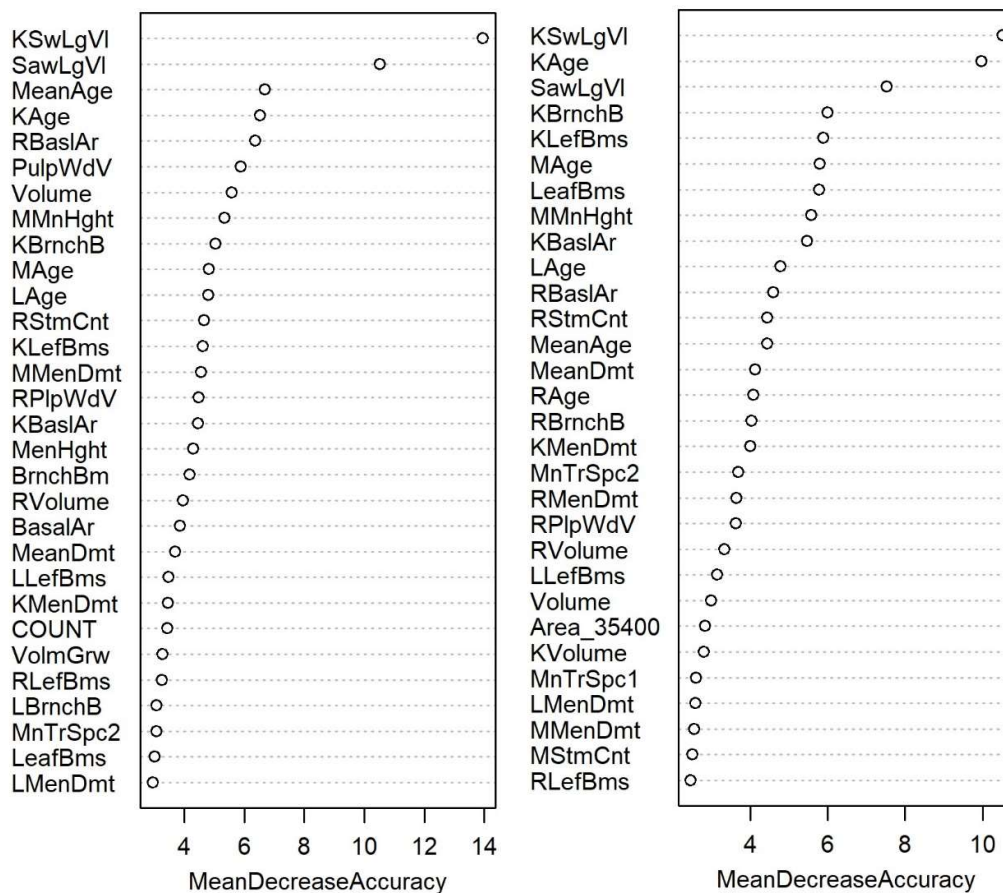
### 4.3 Haukkametsän piirteet RandomForest-analyysin mukaan

Tarkasteltaessa RandomForest analyysien tuloksia todettiin eri menetelmien antavan kohtalaisen yhtäläisiä tuloksia muuttujien merkitystä tarkasteltaessa. Kuusen tukkipuun tilavuus sekä koko puuston tukkipuun tilavuus olivat merkitsevimpien muuttujien joukossa useilla eri säteillä niin randomForest-, Boruta- kuin cForest-menetelmälläkin tarkasteltuna (kuvat 14–16, kuvat 18–23, liite 4). cForest-menetelmällä saadut tulokset on esitetty vain liitteessä johtuen tulosten vaikeatulkintaisuudesta (liite 4).

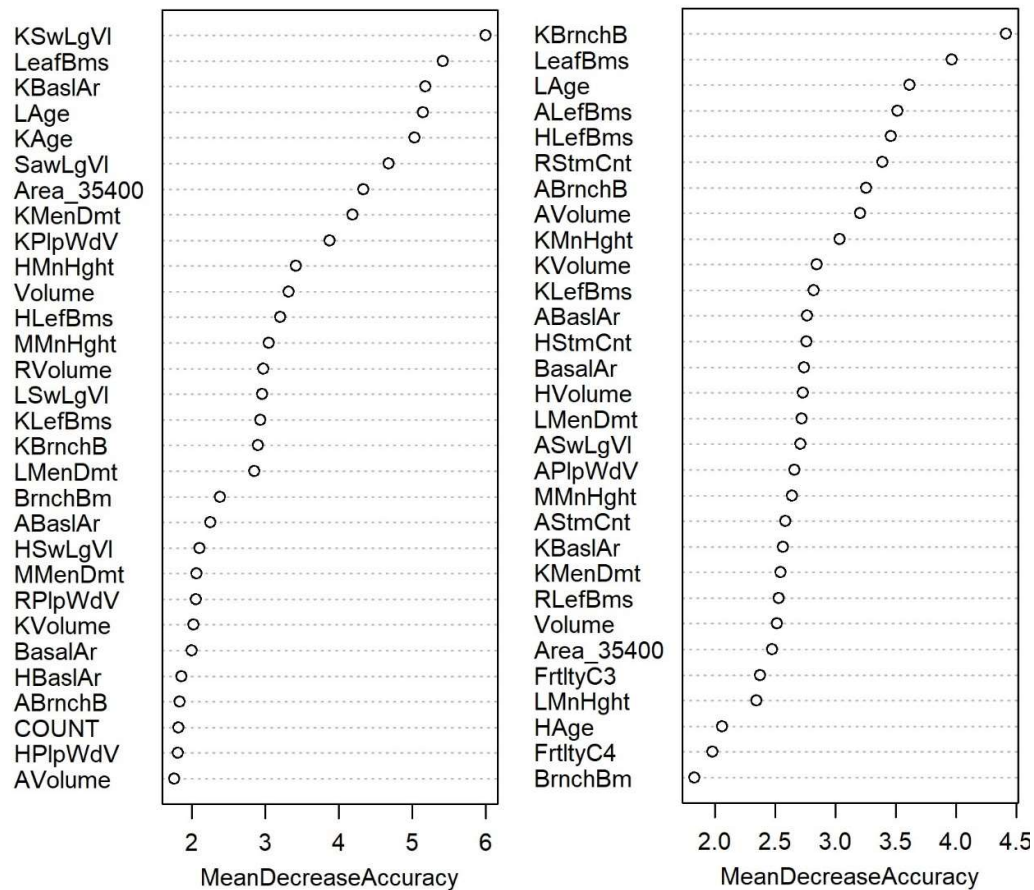
Pesä- ja satunnaispisteiden ympäristöjen eroja randomForest-analyysillä 200 m säteellä vertailtaessa huomataan, että kuusen tukkipuun tilavuus on merkitsevin yksittäinen muuttuja ja saa MeanDecreaseAccuracy-arvon (alempana MnDecAcc) 14 (kuva 14). Tämä on kaikista korkein millään säteellä saavutettu MnDecAcc-arvo. Luokittelumallin suorituskyky heikkenisi siis kaikkein eniten, jos se jätettäisiin mallista pois. Tukkipuun tilavuus on toiseksi merkitsevin muuttuja (MnDecAcc  $\approx$  10) muiden muuttujien saadessa arvoja MnDecAcc  $\leq$  6. Säteellä 250 m merkitsevimmät pesä- ja satunnaispisteitä erottelevat muuttujat, kuusen tukkipuun tilavuus ja kuusen ikä, saavuttavat molemmat noin MnDecAcc-arvon 10 (kuva 14).

Säteillä 500 m ja 1000 m merkitsevimmät muuttujat eivät erotu niin selkeästi muista muuttujista eivätkä saa niin suuria MeanDecreaseAccuracy-arvoja kuin pienemmillä säteillä. Kun tarkastellaan kuvaajaa 500 m säteellä, nähdään, että kuusen tukkipuun tilavuus on merkitsevin muuttuja (MnDecAcc = 6). Sen jälkeen merkitsevimpiä muuttujia ovat lehvästön biomassa ja kuusen pohjapinta-ala. Myös kuusen ikä ja lepän ikä saavat arvon MnDecAcc > 5. Vastaavasti 1000 m säteellä tarkasteltaessa merkitsevimmät muuttujat ovat kuusen oksien biomassa ja lehvästön biomassa MnDecAcc-arvoilla 4,5 ja 4. (kuva 15).

Männyn keskipituus ja haavan tukkipuun tilavuus ovat merkitsevimpien pesä- ja satunnaispisteitä erottelevien muuttujien joukossa 2000 m ja 4000 m säteillä tarkasteltuina. Säteellä 2000 m männyn keskipituuden tarkkuuden heikentymisarvo on 8, haavan runkoluvun noin 7,5, haavan tukkipuun tilavuuden 7 ja soiden pinta-alan 6. Tarkasteltaessa MnDecAcc-arvoja 4000 m säteellä männyn keskipituus saa arvon 10 ja haavan tukkipuun tilavuus arvon 6. Kuusen ja lepän iät saavat MnDecAcc-arvon 5,5 (kuva 16). Myös näillä suuremmilla säteillä merkitsevimmät muuttujat saavat suurempia MnDecAcc-arvoja kuin keskisäteiden merkitsevimmät muuttujat ja ne erottuvat selvemmin muista muuttujista.

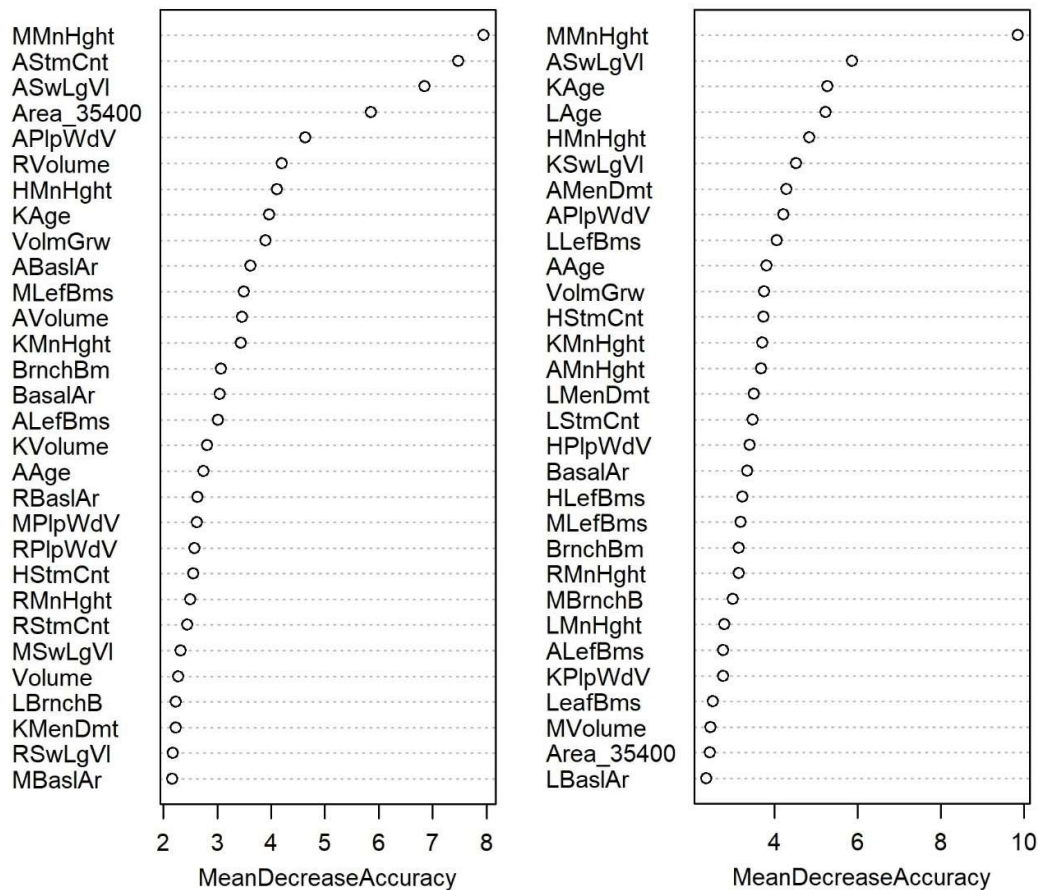


Kuva 14. Perinteisen randomForest-analyysin tulokset 200 m (vasen) ja 250 m (oikea) säteellä tarkasteltuna. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1. MeanDecreaseAccuracy = randomForest-ajojen keskimääräinen mallin tarkkuuden alenema, kun muuttujan alkuperäiset arvot järjestetään satunnaisesti uudelleen. Mitä suuremman MeanDecreaseAccuracy-arvon muuttuja saa, sitä enemmän sen pois jättäminen mallista heikentää mallin tarkkuutta. Kuvasta ei voi päätellä kanahaukalle parhaiten sopivan elinympäristön sisältämiä näiden selittävien muuttujien arvoja.



Kuva 15. Tulokset randomForest-analyysistä 500 m (vasen) ja 1000 m säteellä (oikea) tarkasteltuna. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1. Tulkinnat kuten kuvassa 14.





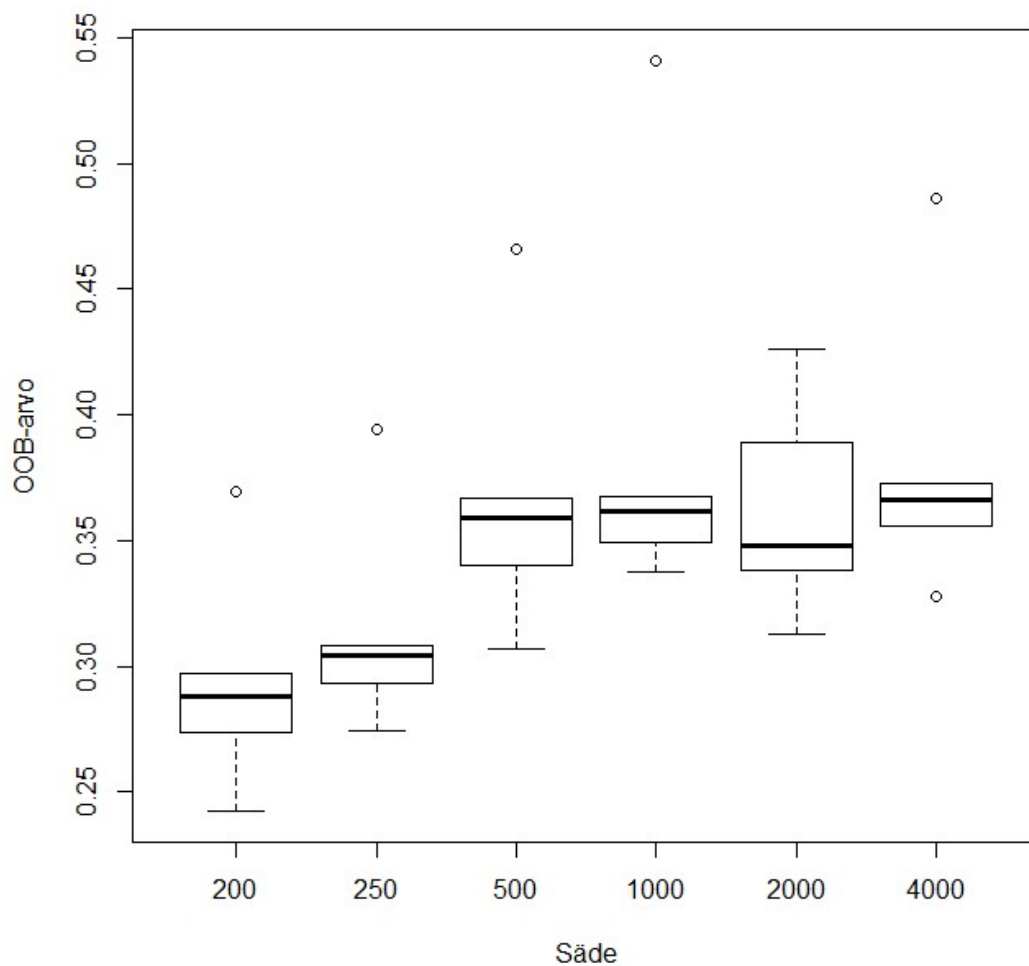
Kuva 16. Tulokset randomForest-analyysistä 2000 m (vasen) ja 4000 m (oikea) säteellä tarkasteltuna. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1. Tulkinnat kuten kuvassa 14.

Ajon jälkeen voitiin tutkia randomForest-luokittelun onnistumista eri säteillä tutkimalla niiden OOB-arvoja. OOB-arvoista nähdään, että 200 m ja 250 m säteillä luokittelu onnistui selvästi parhaiten (taulukko 5). OOB-arvojen keskiarvo 200 m säteelle oli 0,29 ja 250 m säteelle 0,31. Muiden säteiden OOB-arvot olivat välillä 0,34 – 0,37 (kuva 17).

Boruta-menetelmällä tehdyssä analyysissä 200 m säteellä selkeästi parhaiksi luokittelijoiksi nousivat kuusen tukkipuun tilavuus (tärkeys 12) ja tukkipuun tilavuus (tärkeys 10) (kuva 18). Kuusen ikä ja kuusen oksien biomassa (tärkeydet noin 5) olivat seuraavaksi parhaita luokittelijoita. Seuraavat kolme hyvää luokittelijaa olivat kuitupuun tilavuus, puuston keski-ikä ja puuston tilavuus.

Taulukko 5. OOB-arvojen keskiarvot eri tarkastelusäteillä tehdyissä randomForest-analyysissä. Mitä pienempi OOB-arvo, sen paremmin luokittelu onnistuu.

Säde	OOB:n keskiarvo
200	0,29
250	0,31
500	0,36
1000	0,36
2000	0,34
4000	0,37



Kuva 17. Kuvassa on esitetty randomForest-luokittelun onnistuminen eri tarkastelusäteillä 500 ajon aikana (x-akseli). Mitä pienempi OOB-arvo (y-akseli), sitä paremmin luokittelu onnistuu. Luku kertoo siis virheluokitteluiden suhteellisen osuuden luokiteltaessa havaintoja, jotka eivät ole olleet luokittelijan opettamisessa käytetyssä aineistossa mukana.

Boruta-menetelmällä tehdyssä analyysissä 250 m säteellä parhaiksi luokittelijoiksi nousivat kuusen tukkipuun tilavuus (tärkeys 10) sekä tukkipuun tilavuus (tärkeys 7). Kuusen ikä ja kuusen oksien biomassa (tärkeydet noin 7) olivat seuraavaksi parhaita luokittelijoita. Myös kuusen lehvästön biomassa, kuusen pohjapinta-ala, keski-ikä ja rauduskoivun pohjapinta-ala määritettiin hyväksi luokittelijoiksi (kuva 19).

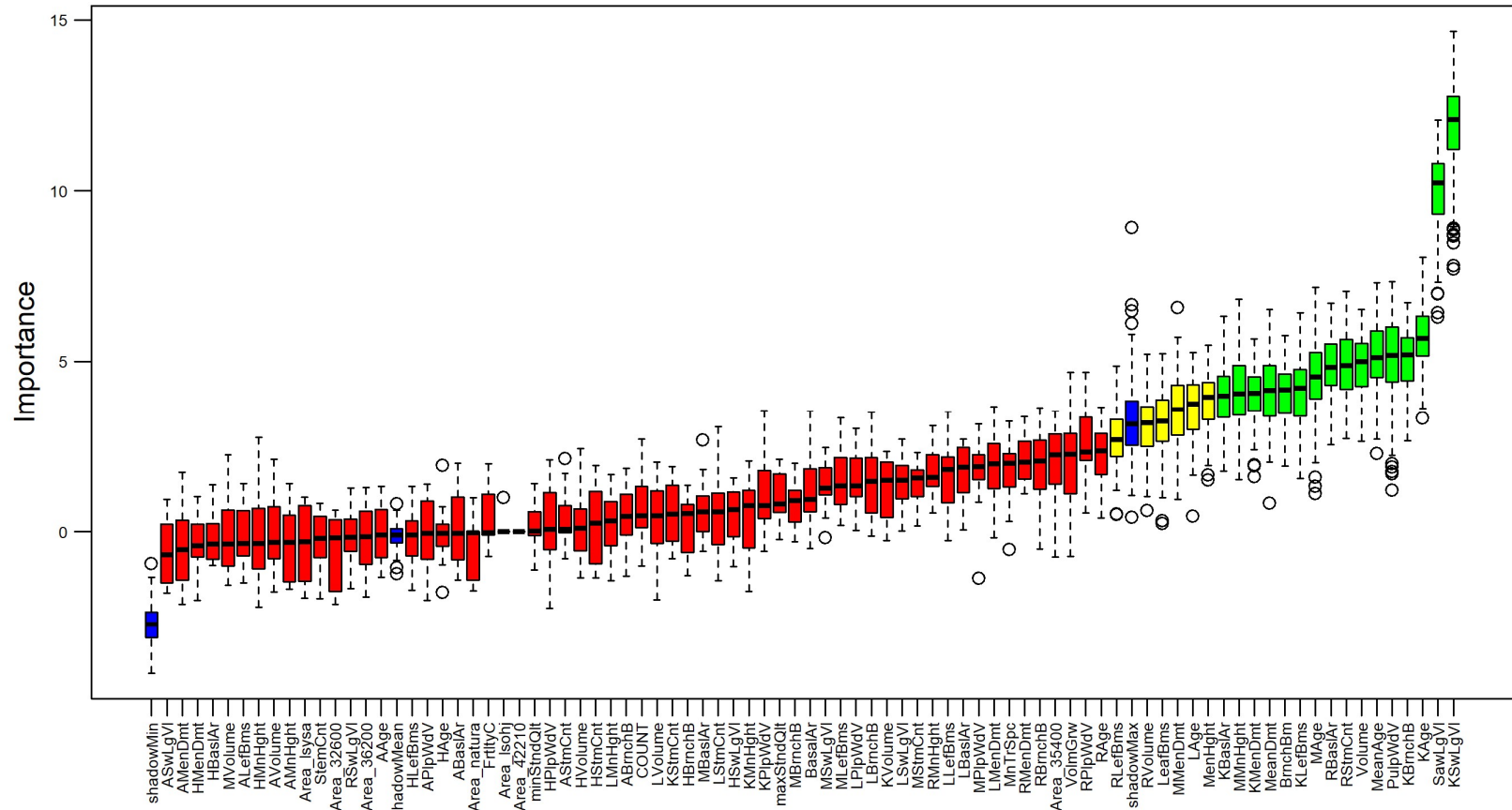
Boruta-analyysin tuloksissa 500 m säteellä parhaita luokittelijoita olivat kuusen ikä, kuusen tukkipuun tilavuus ja soiden pinta-ala (kaikilla tärkeys 7). Seuraavat kolme hyvää luokittelijaa olivat kuusen keskiläpimitta, tukkipuun tilavuus ja kuusen pohjapinta-ala (kuva 20).

Boruta-analyysin mukaan 1000 m säteellä viisi muuttujaa toimivat hyvinä luokittelijoina. Parhaiten luokittelivat männyn keskipituus ja soiden pinta-ala (tärkeys 6). Haavan runkoluvun, haavan tilavuuden ja kuusen iän tärkeys oli 5 (kuva 21).

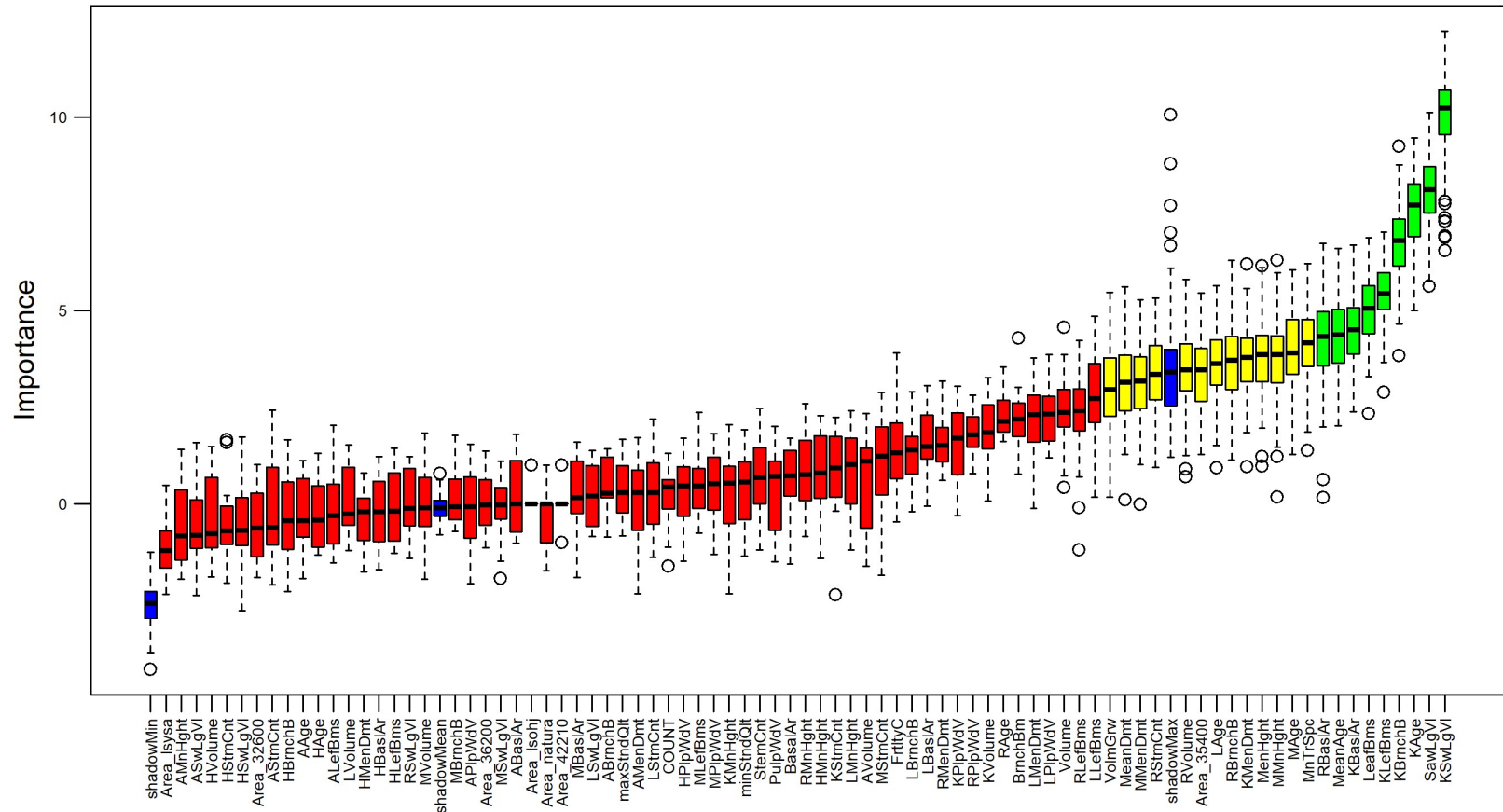
Edellä esitettyjen 500 m ja 1000 m säteiden parhaiten luokittelevat muuttujat eivät erottuneet niin selvästi muista muuttujista kuin muilla säteillä parhaina luokittelijoina esiin tulleet muuttujat. Ne eivät saaneet myöskään niin suuria tärkeysarvoja kuin tärkeimmät muuttujat muilla säteillä tarkasteltuna.

Parhaita luokittelijoita 2000 m säteellä olivat haavan runkoluku, männyn keskipituus ja soiden pinta-ala (tärkeys 8). Seuraavat kolme hyvää luokittelijaa olivat tukkipuun tilavuus, haavan tilavuus ja kuusen ikä (kuva 22).

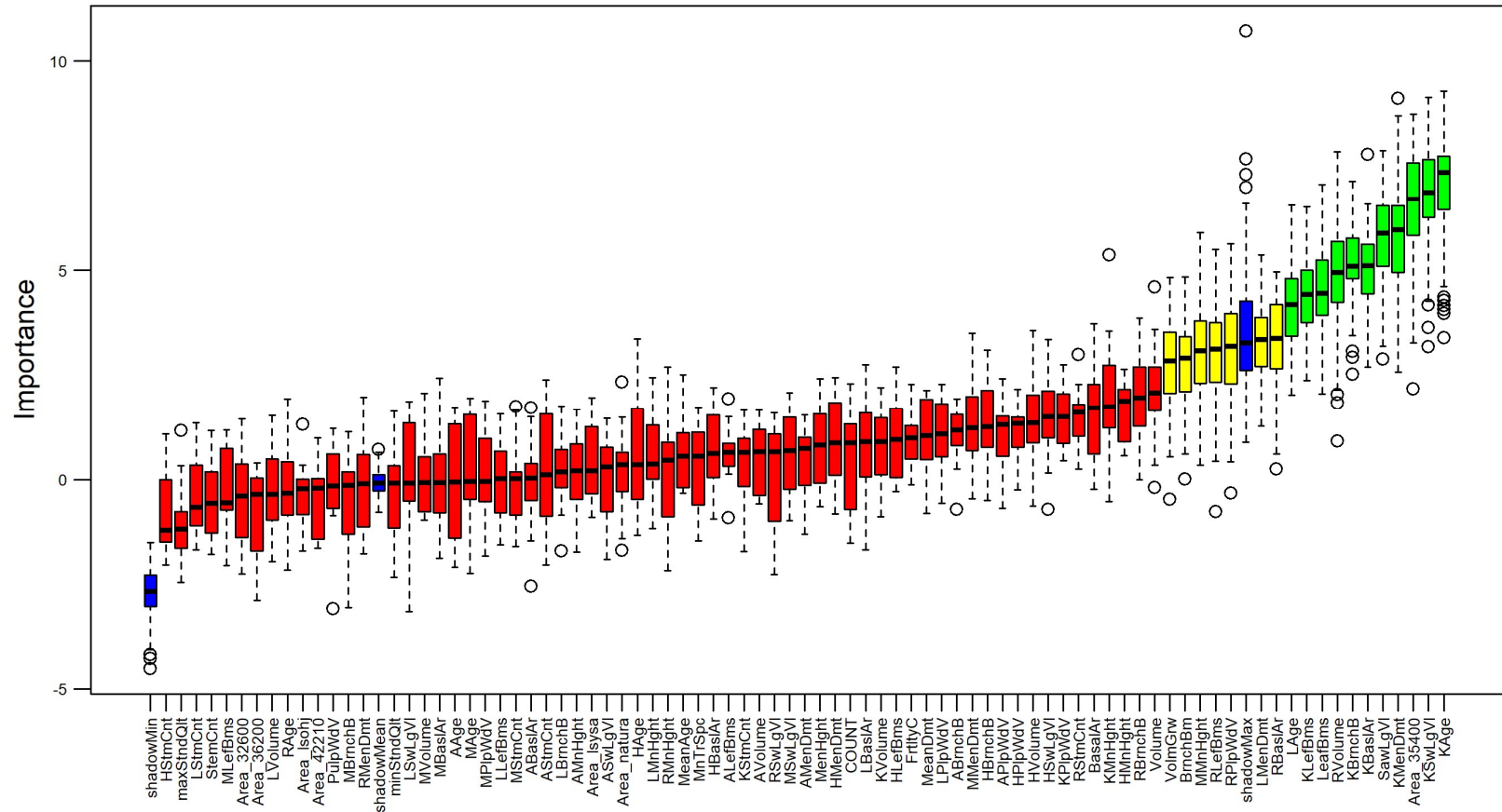
Vain kolme muuttujaa määritettiin hyväksi luokittelijoiksi 4000 m säteellä Boruta-analyysissä (kuva 23). Nämä olivat männyn keskipituus (tärkeys 7,5), haavan tukkipuun tilavuus (tärkeys 5) ja lepän ikä (tärkeys 4,5).



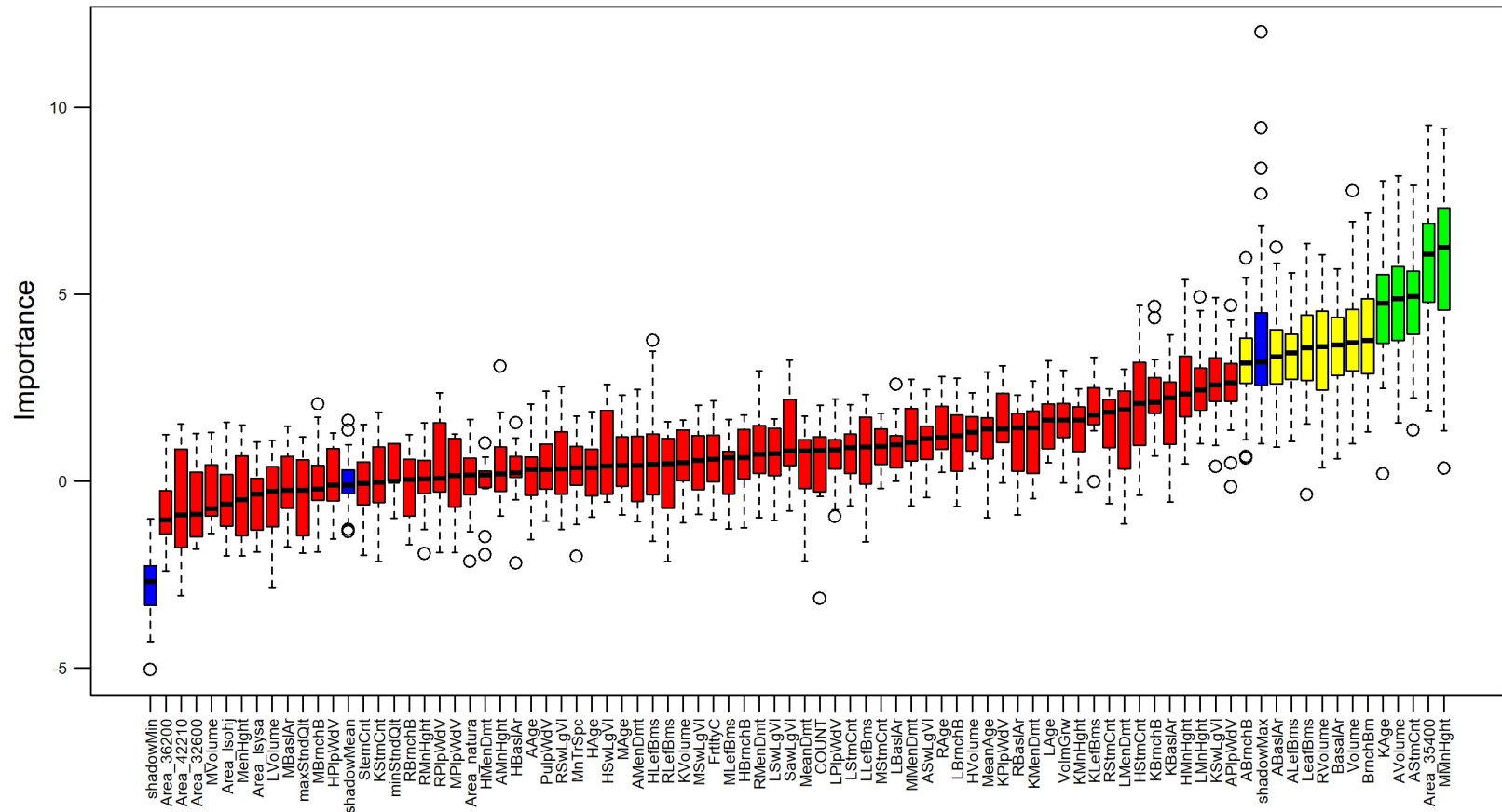
Kuva 18. Muuttujien tärkeys pesäpisteitä ja satunnaispisteitä erottelevana luokittelijana, kun tarkastellaan muuttujia 200 m tarkastelusäteellä. Parhaiten luokittelevat muuttujat ovat kuvaajassa oikealla. Siniset ovat varjomuuttujia, punaiset huonoiten, keltaiset kohtalaisesti ja vihreät parhaiten luokittelevia muuttujia. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1.



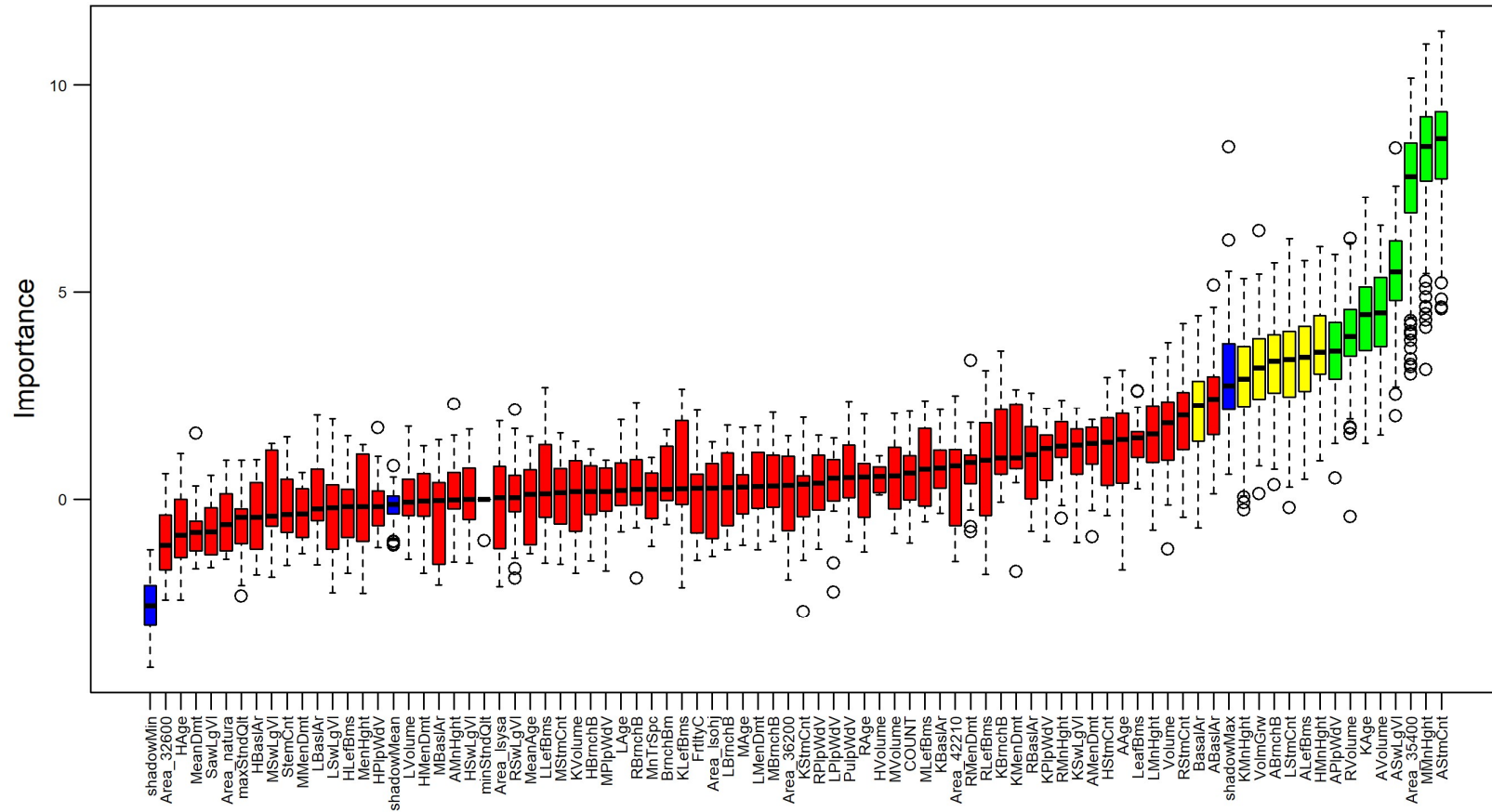
Kuva 19. Muuttujien tärkeys pesäpisteitä ja satunnaispisteitä erottelevana luokittelijana, kun tarkastellaan muuttujia 250 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1. Tulkinnat kuten kuvassa 18.



Kuva 20. Muuttujien tärkeys pesäpisteitä ja satunnaispisteitä erottelevana luokittelijana, kun tarkastellaan muuttujia 500 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1. Tulkinnat kuten kuvassa 18.

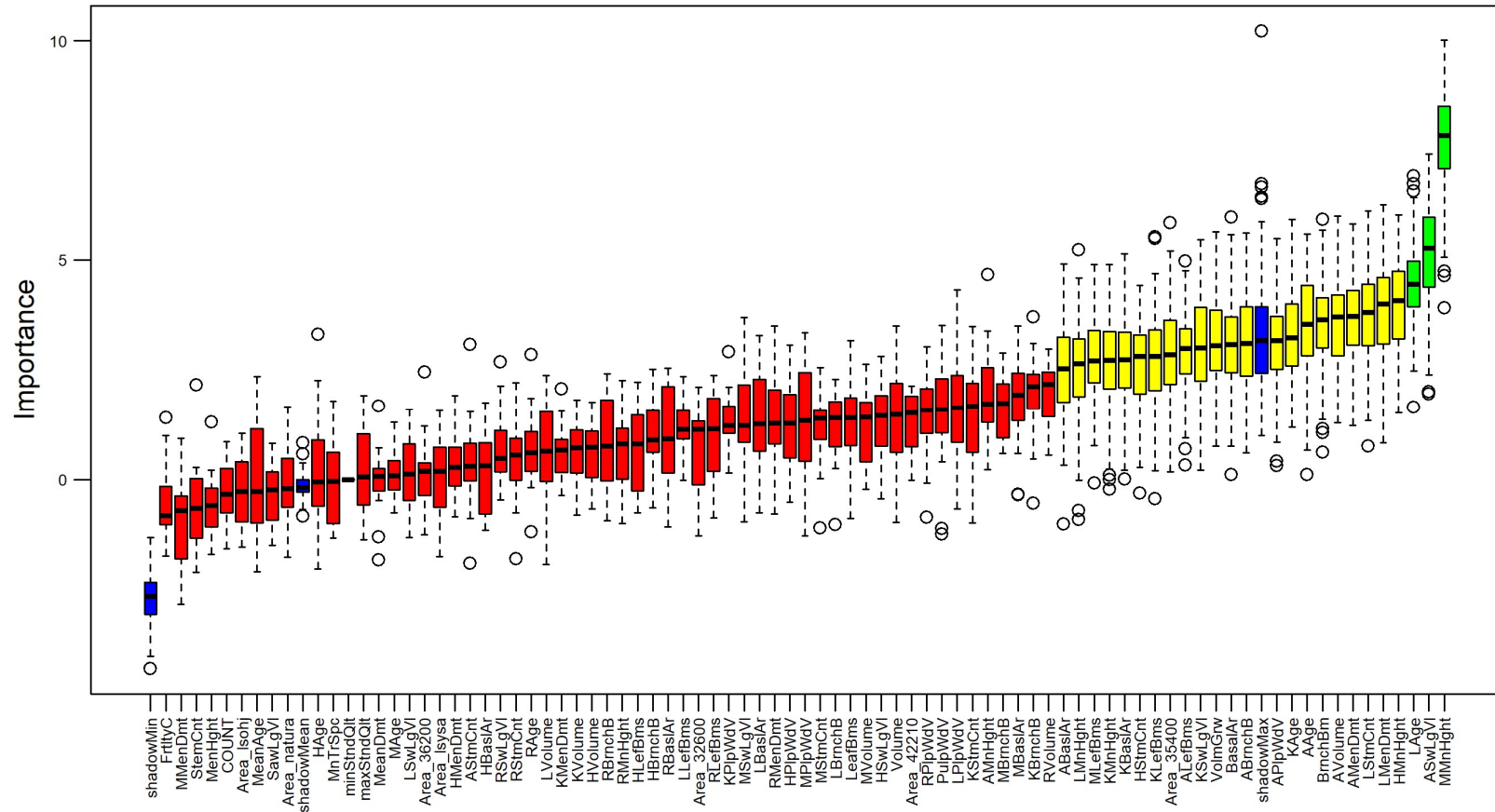


Kuva 21. Muuttujien tärkeys pesäpisteitä ja satunnaispisteitä erottelevana luokittelijana, kun tarkastellaan muuttujia 1000 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1. Tulkinnat kuten kuvassa 18.



Kuva 22. Muuttujien tärkeys pesäpisteitä ja satunnaispisteitä erottelevana luokittelijana, kun tarkastellaan muuttujia 2000 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1. Tulkinnat kuten kuvassa 18.





Kuva 23. Muuttujien tärkeys pesäpisteitä ja satunnaispisteitä erottelevana luokittelijana, kun tarkastellaan muuttujia 4000 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1. Tulkinnat kuten kuvassa 18.

## 5 TULOSTEN TARKASTELU

### 5.1 Kanahaukkametsän tunnistus maksimientropiamallinnuksella

Maksimientropiamallinnuksen mukaan kuusen tilavuus oli parempi selittävä tekijä kanahaukan pesäpaikalle kuin puuston kokonaistilavuus. Tämä nähdään kuusen tilavuuden sisältävän mallin paremmasta AUC-arvosta (liite 5). Lisäksi MVMI-aineistosta tehdyn MaxEnt-analyysin tulokset ovat usealla tulosten esittämistavalla elinympäristötekijöiden, kuten järeän kuusimetsän suuren osuuden, osalta linjassa aikaisempien Pohjois-Amerikassa, Skandinaviassa ja muualla Euroopassa tehtyjen tutkimusten tulosten kanssa (Reynolds *ym.* 1992, Widén 1997, Penteriani & Faivre 1997, Daw & DeStefano 2001, Penteriani 2002, Reich *ym.* 2004, Löhmus 2005, Tornberg *ym.* 2006, Selås *ym.* 2008). Tutkimukseni tulokset ovat siten samansuuntaisia laajalla maantieteellisellä alueella tehtyjen tutkimusten kanssa.

Solukoko 160 m oli parempi tarkastelumittakaava kuin 80 m, mikä näkyy 160 m solukoon mallin paremmista AUC-arvoista (liite 5). Aikaisemmissa tutkimuksissa on havaittu, että resoluution tarkentaminen ei välttämättä tuota parempia malleja, ja lisäksi se hidastaa laskentaa (Swets 1988, Palamakumbure *ym.* 2015, Arnone *ym.* 2016). Tämä tosin riippuu sekä tutkittavasta ilmiöstä että käytettävästä aineistosta. MaxEnt-mallin AUC-arvo 160 metrin solukoolla oli erinomainen, joten voidaan päätellä, että malli kuvaa erinomaisesti tässä mallinnuksessa käytettyjä kanahaukalle sopivia pesäpaikkoja. Joskus korkeat AUC-arvot voivat johtua keskenään korreloivien muuttujien seurauksena olevasta mallin ylisovittamisesta (Phillips *ym.* 2006), mutta tässä tutkimuksessa karsittiin keskenään korreloivia muuttujia etukäteen, joten tästä tuskin on kyse.

Kanahaukalle erinomaisesti sopivan elinympäristön osuus maapinta-alasta Keski-Suomessa oli pieni, 160 m solukoolla 4,11 %. Tämä on lajin kannalta huolestuttavaa, mutta ei sinänsä yllättävää, kun otetaan huomioon metsien ikärakenteen muuttuminen nuorempaan suuntaan (Kuuluvainen 2009). Kanahaukka on silmälläpidettävä laji, jolle uhkatekijöitä ovat vanhojen metsien väheneminen ja metsien käyttö eli toisin sanoen sopivan elinympäristön väheneminen (Lehikoinen

*ym.* 2019). Toisaalta mallin hyvä AUC-arvo ja kanahaukalle erinomaisesti sopivan elinympäristön pieni osuus viittaavat siihen, että malli onnistuu määrittämään kanahaukan elinympäristövaatimukset riittävän yksityiskohtaisesti ja siten tuskin yliarvioi kanahaukalle soveltuvien alueiden määrää. Erityisesti kanahaukalle parhaiten soveltuvat paikat erottuvat siten muusta ympäristöstä.

MaxEnt-analyysin tulosten muuttujakohtaisten jackknife-analyysien perusteella joidenkin muuttujien pois jättäminen heikensi mallin ennustuskykyä huomattavasti ja joillakin muuttujilla ei ollut siihen juurikaan merkitystä. Muuttujien merkitys mallille -tarkastelussa saatiin kuitenkin hieman erilaiset tulokset. Tämä voi osittain johtua jackknife-kuvaajan (kuva 9) ja muuttujan merkittävyystaulukon (taulukko 4) taustalla olevasta erilaisesta merkitsevyyservojen laskentatavasta tai muuttujien käsittelyjärjestyksen eroista (Elith *ym.* 2011, Phillips 2017). Tärkeiksi muuttujiksi nousivat kuitenkin molemmissa kuusen tilavuuden ohella puuston latvuspeittävyys, lehtipuuston latvuspeittävyys ja mäntytukan tilavuus. Puuston latvuspeittävyys kanahaukalle parhaiten sopivissa pesimisympäristöissä on tulosteni mukaan suuri. Tämä voi johtua latvusten pesälle tarjoamasta auringon- ja sateensuojasta sekä suojasta petoja, erityisesti huuhekajia, vastaan (Selås 1997, Kenward 2006, Selås *ym.* 2008).

Puuston suuren latvuspeittävyden ohella kanahaukan pesäpuun ympäristössä on todettu tärkeäksi suuri lentotila (runkojen etäisyys  $\times$  ensimmäisen oksan korkeus) (Penteriani & Faivre 1997). Tällöin tässäkin tutkimuksessa havaitsemani alhainen lehtipuuston latvuspeittävyys kanahaukalle erinomaisesti sopivissa elinympäristöissä selittyy lehtipuuston latvuksen muodolla, joka runsaasti lehtipuustoa sisältävässä metsässä rajoittaisi kanahaukan lentämistä. Kanahaukan pesäpuu on tyypillisesti järeä kuusi, toisinaan myös järeä mänty (Penteriani & Faivre 1997, Löhmus 2006), ja tällaisten löytäminen olisi vaikeaa lehtipuuvaltaisissa talousmetsissä. Samat syyt voivat selittää myös esille tulleet muiden koivuun ja muihin lehtipuihin liittyvien muuttujien pienet optimiarvot. Lehtipuita on kuitenkin sekapuustona ainakin jonkin verran parhaissa haukkametsäkohteissa. Kanahaukan tyypillinen metsästystapa on metsän sisällä tapahtuvat lyhyet lennot

oksalta toiselle ja väijyminen (Kenward 1982, Bloxton 2002). Niinpä suuri latvuspeittävyys takaa sopivien oksien löytymisen riittävän läheltä, ja tiheyden pysyessä kohtuullisena kanahaukka voi vielä saalistaa tehokkaasti. Toisaalta suotuisten olosuhteiden esiintymistodennäköisyyden lasku puuston tilavuuden ylittäessä  $375 \text{ m}^3/\text{ha}$  selittyy sillä, että liian tiheässä metsässä kanahaukka ei pysty lentämään hyvin.

Kanahaukalle sopivan metsän esiintymistodennäköisyys kasvaa jyrkästi mäntytukin tilavuuden kasvaessa nollostaa arvoon  $40 \text{ m}^3/\text{ha}$ , jonka jälkeen se alkaa laskea. Myös tämä voi liittyä metsästykseseen ja pesintään: järeät männyt tarjoavat mahdollisia pesäpaikkoja (Penteriani & Faivre 1997, Löhmus 2006), mutta männyn latvusmuoto myös mahdollistaa oksilla väijymisen ja nopean lennähtelyn puusta toiseen. Koska kuusi on kuitenkin kanahaukan suosima pesäpuu ja kuusimetsä suosituin pesimisympäristö, ei männyn tukkipuun tilavuus hehtaarilla nouse kovin suureksi kanahaukan pesintään parhaiten soveltuvissa paikoissa. Maksimientropiamallinnuksella saatiin tehokkaasti esille kanahaukan pesäpaikan esiintymisen kannalta tärkeimmät MVMi-aineiston muuttujat. Lisäksi MaxEnt-analyysin eduksi voidaan katsoa selkeästi kartalla esitetyt esiintymät kanahaukan kannalta otollisimmista pesämetsäalueista. Tulokset pätevät Keski-Suomen alueella kanahaukan osalta. Muille maantieteellisille alueille ja muille vanhan metsän lajeille yleistettäessä tulee täten noudattaa varovaisuutta (Phillips *ym.* 2004).

Ottaen kuitenkin huomioon, että jotkin lajit esiintyvät usein kanahaukan kanssa (Reynolds *ym.* 1992, Byholm *ym.* 2012), maksimientropiamallinnuksen esittämät kanahaukan pesäpaikaksi soveltuvat metsät ovat mahdollisesti myös muiden vanhan metsän lajien esiintymispaikkoja.

## 5.2 Kanahaukkametsän tunnistus RandomForest-analyyseillä

Perinteisen randomForest-analyysin tuloksuvaajista nähtiin, että suurimmat MeanDecreaseAccuracy-arvot saatiin pienimmillä (200 m ja 250 m) ja suurimmilla tarkastelusäteillä (2000 m ja 4000 m), keskimmäisten tarkastelusäteiden (500 m ja

1000 m) MeanDecreaseAccuracy-arvojen jäädessä pienemmiksi. Tämä voi johtua siitä, että kanahaukka on pesäpaikkansa valinnassa hyvin valikoiva, jolloin pesäpisteiden lähiympäristössä korostuvat tietyt muuttujat. Kauempana pesäpisteestä taas on hyvin selkeästi kanahaukan vain saalistamiseen käyttämää ympäristöä, ja saalistuksen kannalta tärkeimmät muuttujat tulevat esiin tässä mittakaavassa. Keskimatkan säteet taas voivat monien pesien tapauksessa olla eräänlaista pesäympäristön ja saalistusalueen vaihtumisvyöhykettä, jolloin sekä saalistukselle että pesinnälle tärkeät muuttujat vaikuttavat alueella hieman eivätkä yksittäiset muuttujat pääse erottumaan niin selvästi. Myös aikaisemmissa tutkimuksissa (Arnone *ym.* 2016) on havaittu, että eri tarkasteluresoluutioiden avulla luotujen mallien paremmuus ei välttämättä nouse tai laske suoraviivaisesti suurimmasta pienimpään. Myös Boruta-muuttujiervalintamenetelmän antamat tulokset eri mittakaavojen suhteen olivat samansuuntaiset: keskisäteillä muuttujien Importance-arvot olivat pienempiä kuin pienillä ja suurilla säteillä.

cForest-analyysin tuloskuvaajista (liite 4) nähdään, että kyseisellä menetelmällä epäolennaisten muuttujien merkitsevyysarvot saavat satunnaisia arvoja nollan molemmin puolin. Niinpä merkitsevyysarvot, jotka eivät poikkea nolasta enempää kuin suurimman negatiivisen arvon saavat merkitsevyysarvot, voidaan jättää huomiotta. Tätä suurempia arvoja voidaan pitää merkityksellisinä, tai ainakin niitä voidaan ottaa lähempään tarkasteluun (Strobl *ym.* 2009b). Tulosten selkeyden perusteella suosittelenkin käyttämään mieluummin ensisijaisesti Borutaa, sitten randomForestia kuin cForestia.

Kuten maksimientropia-analyysissä, myös randomForest- ja Boruta-analyysillä järeään ja vanhaan kuusimetsään liittyvät muuttujat olivat merkitsevimpiä muuttujia. Tämä on linjassa aikaisempien tutkimusten kanssa (ks. viitteet aliluvussa 5.1). Boruta -analyysissä ilmennyt kuusen oksien biomassan esiin nouseminen merkitsevänä muuttujana on niin ikään huomattu jo aikaisemmissa tutkimuksissa (Selås 1997, Kenward 2006, Selås *ym.* 2008). Sen tärkeys johtunee latvusten tarjoamasta suojasta sadetta, auringonpaahdetta ja saalistajia vastaan (Selås 1997, Kenward 2006, Selås *ym.* 2008). Maksimientropia-analyysissä ei ollut kuusen oksien

biomassan kanssa täysin vastaavaa muuttujaa, mutta kuusen hukkapuuosan biomassassa on lähellä sitä – se tosin sisältää oksien lisäksi kuitu- ja tukkipuuksi kelpaamattoman runkopuun osan. Myös kuusen hukkapuuosan biomassan suhteen tulokset olivat samansuuntaisia.

Männyn keskipituus oli merkitsevimpien muuttujien joukossa suuremmilla säteillä tarkasteltuna niin randomForest- kuin Boruta-analyysilläkin, ja tämä voi johtua kanahaukan metsästystavasta, jossa sopivan pituiset männyt tarjoavat hyviä väijymispaikkoja (Bloxton 2002). Myös järeät haavat voivat toimia hyvinä tähystyspaikkoina, ja mahdollisesti tämän takia myös ne nousivat esiin suurten säteiden randomForest-analyysissä.

Soiden osuus pinta-alasta nousi esiin sekä randomForest- että Boruta-analyysissä 500 m – 2000 m säteillä. Tämä on yhtenevä aikaisemmissa tutkimuksissa esille tulleiden tietojen kanssa, joiden mukaan soiden osuus on suurempi pesäpisteiden kuin satunnaispisteiden ympäristössä (Speiser & Bosakowski 1987). Tämä voi johtua soiden tarjoamasta monimuotoisemmasta saalistarjonnasta.

Mielenkiintoista oli, että rakennetun ympäristön osuuden suhteen kanahaukan pesäpisteillä tai satunnaispisteillä ei ollut merkitsevää eroa minkään Random Forest -analyysitavan mukaan. Tämä tarkoittaa sitä, että kanahaukka sietää jonkin verran ihmisvaikutusta elinalueellaan, kunhan elinympäristövaatimukset muuten täyttyvät, mikä on linjassa aikaisempien tutkimusten kanssa (Rutz 2006).

Luokittelu onnistui randomForest-analyysillä eri säteillä 29–37 %:n virheillä. Luokittelun onnistumisen suhteen säteiden paremmuusjärjestys parhaasta huonoimpaan on 200 m, 250 m, 2000 m, 500 m, 1000 m ja 4000 m. Tämä tukee jossain määrin MeanDecreaseAccuracy-arvoista tehtyä päätelmää eri muuttujien merkitysten hämärtyvyydestä 500 m ja 1000 m säteillä. Kaikista suurimman säteen merkitys voi olla kanahaukan elinympäristössä vähäisin, minkä takia se saa myös suurimman OOB-virhearvon. cForest-analyysin luokittelu onnistui 29–39 %:n virheosuuksilla (liite 4). cForest-analyysissä eri säteiden keskinäinen järjestys

poikkesi randomForest-analyysissä saadusta järjestyksestä, järjestyksen parhaasta huonoimpaan ollessa 250 m, 4000 m, 2000 m, 500 m, 1000 m ja 200 m.

### 5.3 Maksimientropiamallinnuksen ja RandomForest-analyysin vertailua

MaxEnt-ohjelmalla tehtävä maksimientropiamallinnus on toimiva ja tehokas mallinnusmenetelmä, kun käytettävissä on kattavaa rasterimuotoista elinympäristön ominaisuustietoa ja kun pesäpisteitä on käytettävissä riittävän kattavasti. MaxEnt-analyysiin ei tarvita tällöin kovinkaan monimutkaista elinympäristötiedon esikäsittelyä. Maksimientropiamallinnus MaxEnt-ohjelmalla antoi kunkin elinympäristömuuttujan optimiarvot kanahaukan pesämetsän kannalta, ja lisäksi se tuotti kartan, josta nähdään mahdolliset kanahaukan pesämetsäpaikat.

R-tilasto-ohjelmalla tehdyistä randomForest-, Boruta- ja cForest-analyyseistä saatiin vahvistusta MaxEnt-analyysin esiin tuomille muuttujille. Tässä tapauksessa käytettävissä ei kuitenkaan ollut yhtenäistä vektorimuotoista elinympäristötietoaineistoa Keski-Suomen alueelta. Tämä teki aineiston esikäsittelystä hyvin monimutkaista. Lisäksi tiedossa ei ollut kaikkien muuttujien kohdalta, ovatko aineistossa tietyillä kuvioilla nolla-arvoja saaneet tai tyhjiksi jääneet muuttujat jääneet mittaamatta, vai eikö mitään mitattavaa kyseisen muuttujan kohdalta ollut. Koska kyseessä on laaja alue, arvoja mittaamassa ja tallentamassa on ollut suuri joukko henkilöitä. Yhtenäiset mittaus- ja tallennuskäytännöt metsävaratietoaineistoja kootessa olisivat tärkeitä, jotta aineisto olisi mahdollisimman monikäyttöistä. RandomForest voisi olla toimivampi ja hyödyllisempi menetelmä, jos saataisiin myös muut kuin yksityismetsissä olevat alueet mukaan aineistoon ja näin vähennettyä aineiston aukkoisuutta.

Molemmat käytetyt menetelmät voivat olla epäluotettavia, mikäli käytettävissä ei ole riittävää määrää pesäpisteitä. Tällöin voi käydä niin, että kanahaukan kannalta tärkeät muuttujat eivät erotu esimerkiksi sen takia, koska sattumalta aineistossa ei ole niitä pesäpisteitä, jotka kuvaavat kaikista tyypillisintä kanahaukkametsää

tärkeiden muuttujien osalta, eikä eroa pesäpisteiden ja satunnais-/taustapisteen välille löydetä.

Metsäkeskuksen metsävaratietoaineisto sisälsi vain osan yksityisten metsänomistajien metsäpalstoista ja vaati erittäin paljon käsittelyä, jotta sitä voitiin käsitellä RandomForest-analyysillä. Metsäkeskuksen metsävaratietoaineisto sisälsi runsaan esikäsittelyn jälkeenkin spatiaalisia ja tiedollisia aukkoja, joita ei voinut käytössäni olleilla tiedoilla täydentää. Tällaisia spatiaalisia aukkoja on muun muassa valtion metsien sekä ainakin joidenkin kuntien omistamien metsien alueilla. Joidenkin muuttujien kohdalla oli puuttuvuutta. Kun taas maksimientropia-analyysiä varten tehdyt MVMI-aineiston käsittelyt olivat huomattavasti suoraviivaisempia. Kun maksimientropia-analyysin avulla saatiin myös tuotettua kartta potentiaalisista kanahaukka-alueista, voidaan todeta, että maksimientropia-analyysi on näistä menetelmistä ainakin näillä aineistoilla paremmin käyttöön soveltuva.

Lisäksi RandomForest-analyysiin vaikuttaa se, että satunnaispisteiden joukossa voi olla sellaisia pisteitä, jotka todellisuudessa ovat hyvinkin lähellä sellaisia kanahaukan pesäpisteitä, joita ei vaan ollut mukana käytettävissä olleessa pesäpisteaineistossa. Näiden syiden takia tämän aineiston RandomForest-analyysin tuloksia tulee käyttää lähinnä suuntaa-antavina. Myös mallinnuksen tarkkuutta kuvaavat OOB- ja AUC-arvot viittaavat RandomForest-analyysin olevan maksimientropiamallinnusta epätarkempi menetelmä tämän tutkimusasetelman tapauksessa.

MaxEnt ja RandomForest-analyyseissä tuli esiin samoja muuttujia merkitsevinä muuttujina. Tuloksia voi pitää luotettavina, kun RandomForest-analyysin aineiston puutteista huolimatta eri analyysimenetelmillä tuli samankaltaisia tuloksia. Lisäksi RandomForest-analyysissä merkitsevinä esille tulleet muuttujat, joita ei ollut mukana MaxEnt-analyysissä, ovat linjassa aikaisempien tutkimustulosten kanssa. Voidaan todeta, että RandomForest- ja maksimientropia-analyyseissä esille tulleita muuttujia ja niiden maksimientropia-analyysissä esille tulleita merkitseviä arvoja tulee ottaa huomioon suunniteltaessa vanhojen metsien lajien suojelua.



## 6 JOHTOPÄÄTÖKSET

Tutkittaessa kanahaukan pesäpaikkoja sisältävää aineistoa MaxEnt- ja Random Forest -menetelmillä todettiin vanhaan ja järeään kuusimetsään viittaavien muuttujien olevan kanahaukan pesäpaikan valinnan kannalta tärkeimpiä muuttujia. Koska tämä on linjassa aikaisempien tutkimusten kanssa ja koska MaxEnt-analyysin AUC-arvot ovat korkeita, voidaan tuloksia pitää luotettavina. Kanahaukan tunnetut pesäpisteet eroavat satunnaispisteistä randomForest-analyysin perusteella useiden selittävien muuttujien osalta. RandomForest-analyysillä saatiin tietoa merkittävistä elinympäristömuuttujista eri etäisyyksillä kanahaukan pesäpaikoista, ja niiden avulla voidaan päätellä pesäpaikan lähiympäristön ominaisuuksien lisäksi myös kanahaukan saalistusympäristölle tärkeitä ominaisuuksia. Koska randomForest-analyysissä käytössä ollut metsävara-aineisto sisälsi spatiaalisia ja muuttujakohtaisia puutteita, maksimientropiamallinnus on kuitenkin käytettävissä olleella aineistolla luotettavampi ja helppokäyttöisempi mallinnusmenetelmä. Lisäksi maksimientropiamallinnuksen eduksi voidaan katsoa siitä saatava kartta mahdollisista parhaiten kanahaukan pesäpaikaksi sopivista kohteista, jollaista ei randomForest-analyysillä nyt tässä tutkimuksessa saatu. Kanahaukalle erinomaisesti sopivan elinympäristön pieni osuus Keski-Suomessa kertoo kanahaukan suojelutoimien tärkeydestä.

## KIITOKSET

Haluan kiittää ohjaajiani kärsivällisestä opastuksesta. Heidiä kiitän rahoituksen hankkimisesta, projektin ja Pro gradu -työn suunnittelutyöstä, kohdelajiin liittyvän kirjallisuuden toimittamisesta, avusta gradun ohjaamisesta kirjoitusasun ja menetelmien selityksen osalta oikeaan suuntaan sekä säännöllisistä palavereista, jotka ovat auttaneet gradun saamisessa valmiiksi. Anssia haluan kiittää kovasta työstä ja suuresta avusta tilasto- ja paikkatietomenetelmien osalta, käyttämäni ohjelmien ja niihin liittyvän kirjallisuuden toimittamisesta, niiden käytössä opastamisesta ja huolellisesta gradun kommentointityöstä. Kiitän myös Ympäristöministeriötä rahoituksesta graduuni liittyen. Tämä tutkimus oli osa Luonnontieteellisen museon (Luomus) koordinoimaa metsäpetolintuprojektia, joka on osa Etelä-Suomen metsien monimuotoisuuden toimintaohjelma METSOa.

**KIRJALLISUUS**

Arnone E., Francipane A., Scarbaci A., Puglisi C. & Noto L.V. 2016. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. *Environmental Modelling & Software* 84: 467–481. doi: 10.1016/j.envsoft.2016.07.016.

Beier P. & Drennan J.E. 1997. Forest Structure and Prey Abundance in Foraging Areas of Northern Goshawks. *Ecol Appl* 7: 564–571. doi:FSAPAI]2.0.CO;2.

Björklund H., Saurola P. & Valkama J. 2018. Kolea kevät koetteli petolintuja. *Linnut-vuosikirja 2017*: 56–59.

Björklund H., Valkama J., Tomppo E. & Laaksonen T. 2015. Habitat Effects on the Breeding Performance of Three Forest-Dwelling Hawks. *PLOS ONE* 10: e0137877. doi:10.1371/journal.pone.0137877.

Björklund H., Santangeli A., Blanchet F., Huitu O., Lehtoranta H., Lindén H., Valkama J. & Laaksonen T. 2016. Intraguild predation and competition impacts on a subordinate predator. *Oecologia* 181: 257–269. doi:10.1007/s00442-015-3523-z.

Bloxtton T.D. 2002. *Prey abundance, space use, demography, and foraging habitat of northern goshawks in western Washington*. University of Washington, Seattle, WA USA.

Burgas D., Byholm P. & Parkkima T. 2014. Raptors as surrogates of biodiversity along a landscape gradient. *J Appl Ecol* 51: 786–794. doi:10.1111/1365-2664.12229.

Byholm P., Burgas D., Virtanen T. & Valkama J. 2012. Competitive exclusion within the predator community influences the distribution of a threatened prey species. *Ecology* 93: 1802–1808. doi:10.1890/12-0285.1.

Daw S.K. & DeStefano S. 2001. Forest Characteristics of Northern Goshawk Nest Stands and Post-Fledging Areas in Oregon. *The Journal of Wildlife Management* 65: 59–65. doi:10.2307/3803277.

DeLeo J.M. 1993. Receiver operating characteristic laboratory (ROCLAB): Software for developing decision strategies that account for uncertainty. Teoksessa: *Proceedings of (2nd) International Symposium on Uncertainty Modeling and Analysis* : 318–325.

Dormann C.F., Elith J., Bacher S., Buchmann C., Carl G., Carré G., Marquéz J.R.G., Gruber B., Lafourcade B. & Leitão P.J. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 27–46.

Elith J., Phillips S.J., Hastie T., Dudík M., Chee Y.E. & Yates C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43–57. doi:10.1111/j.1472-4642.2010.00725.x.

Fielding A.H. & Bell J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24: 38–49.

Hakkarainen H., Mykrä S., Kurki S., Tornberg R., Jungell S. & Nikula A. 2004. Long-term change in territory occupancy pattern of goshawks (*Accipiter gentilis*). *Écoscience* 11: 399–403. doi:10.1080/11956860.2004.11682848.

Hastie T., Tibshirani R. & Friedman J. 2009. *The elements of statistical learning* – 2<sup>nd</sup> Edition. Springer series in statistics, New York, NY, USA.

Hothorn T., Hornik K., Strobl C. & Zeileis A. 2018. *party: A Laboratory for Recursive Partytioning*. 1.3-1. <https://cran.r-project.org/web/packages/party/party.pdf>.

Hurlbert S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol Monogr* 54: 187–211.

Kaila A. & Ihalainen A. 2014. Metsävarat. Teoksessa: Peltola, A. (toim.): *Metsätilastollinen vuosikirja 2014*, Metsäntutkimuslaitos, Vantaa: 33–78.

Kenward R.E. 1982. Goshawk hunting behaviour, and range size as a function of food and habitat availability. *The Journal of Animal Ecology* 51: 69–80.

Kenward R.E. 2006. *The Goshawk*. T & AD Poyser.

Korhonen K.T. 2009. *VMI11 maastotyöohje 2009 Koko Suomi*. Metsäntutkimuslaitoksen erillisjulkaisut.

Krüger O. 2002. Analysis of nest occupancy and nest reproduction in two sympatric raptors: Common buzzard *Buteo buteo* and goshawk *Accipiter gentilis*. *Ecography* 25: 523–532. doi:10.1034/j.1600-0587.2002.250502.x.

Kursa M.B. & Rudnicki W.R. 2010. Feature selection with the Boruta package. *J Stat Softw* 36: 1–13.

Kursa M.B. & Rudnicki W.R. 2018. *Package 'Boruta'*. 6.0.0. <https://cran.r-project.org/web/packages/Boruta/Boruta.pdf>.

Kuuluvainen T. 2009. Forest Management and Biodiversity Conservation Based on Natural Ecosystem Dynamics in Northern Europe: The Complexity Challenge. *AMBIO* 38: 309–315. doi:10.1579/08-A-490.1.

Kuuluvainen T., Tahvonen O. & Aakala T. 2012. Even-Aged and Uneven-Aged Forest Management in Boreal Fennoscandia: A Review. *Ambio* 41: 720–737. doi:10.1007/s13280-012-0289-y.

Lambeck R.J. 1997. Focal Species: A Multi-Species Umbrella for Nature Conservation. *Conserv Biol* 11: 849–856. doi:10.1046/j.1523-1739.1997.96319.x.

Lehikoinen A., Jukarainen A., Mikkola-Roos M., Below A., Lehtiniemi T., Pessa J., Rajasärkkä A., Rintala J., Rusanen P., Sirkiä P., Tiainen J. & Valkama J. 2019. Linnut. Teoksessa: Hyvärinen E., Juslén A., Kempainen E., Uddström A. & Liukko U. (toim.), *Suomen lajien uhanalaisuus – Punainen kirja 2019*, Ympäristöministeriö, Helsinki: 560–570.

Liaw A. & Wiener M. 2002. Classification and regression by randomForest. *R news* 2: 18-22.

Linden H. & Wikman M. 1983. Goshawk Predation on Tetraonids: Availability of Prey and Diet of the Predator in the Breeding Season. *J Anim Ecol* 52: 953-968. doi:10.2307/4466.

Lõhmus A. 2005. Are timber harvesting and conservation of nest sites of forest-dwelling raptors always mutually exclusive? *Animal Conservation* 8: 443-450. doi:10.1017/S1367943005002349.

Lõhmus A. 2006. Nest-tree and nest-stand characteristics of forest-dwelling raptors in east-central Estonia: implications for forest management and conservation. *Proc.Estonian Acad.Sci.Biol.Ecol* 55: 31-50.

McIntosh A. 2016. *The jackknife estimation method*. arXiv preprint arXiv:1606.00497 .

Meller K., Björklund H., Saurola P. & Valkama J. 2017. Teoksessa:Petolintuvuosi 2016, pesimistulokset ja kannankehitykset. *Linnut-vuosikirja* 2016: 16-31.

Mäkisara K., Katila M. & Peräsaari J. 2019. *The Multi-Source National Forest Inventory of Finland – methods and results 2015*. <http://urn.fi/URN:ISBN:978-952-326-711-4>  
<http://urn.fi/URN:ISBN:978-952-326-712-1>.

Mönkkönen M., Husby M., Tornberg R., Helle P. & Thomson R.L. 2007. Predation as a landscape effect: The trading off by prey species between predation risks and protection benefits. *J Anim Ecol* 76: 619-629. doi:10.1111/j.1365-2656.2007.01233.x.

Naumov V., Manton M., Elbakidze M., Rendenieks Z., Priednieks J., Uhlianets S., Yamelynets T., Zhivotov A. & Angelstam P. 2018. How to reconcile wood production and biodiversity conservation? The Pan-European boreal forest history gradient as an “experiment”. *Journal of Environmental Management* 218: 1-13. doi:10.1016/j.envman.2018.03.095.

Pakkala T. 2012. *Spatial ecology of breeding birds in forest landscapes : an indicator species approach*. Finnish Society of Forest Science, Vantaa.

Palamakumbure D., Flentje P. & Stirling D. 2015. Consideration of optimal pixel resolution in deriving landslide susceptibility zoning within the Sydney Basin, New South Wales, Australia. *Comput Geosci* 82: 13–22. doi:10.1016/j.cageo.2015.05.002.

Penteriani V. 2002. Goshawk nesting habitat in Europe and North America: A review. *Ornis Fenn* 79: 149–163.

Penteriani V. & Faivre B. 1997. Breeding density and nest site selection in a Goshawk *Accipiter gentilis* population of the Central Apennines (Abruzzo, Italy). *Bird Study* 44: 136–145. doi:10.1080/00063659709461049.

Penteriani V. & Faivre B. 2001. Effects of harvesting timber stands on goshawk nesting in two European areas. *Biol Conserv* 101: 211–216. doi:10.1016/S0006-3207(01)00068-4.

Pettersson R.B., Ball J.P., Renhorn K., Esseen P. & Sjöberg K. 1995. Invertebrate communities in boreal forest canopies as influenced by forestry and lichens with implications for passerine birds. *Biological Conservation* 74: 57–63. <http://www.sciencedirect.com/science/article/pii/000632079500015V>.

Phillips S.J., Dudík M. & Schapire R. 2004. A maximum entropy approach to species distribution modeling. Teoksessa: *Proc. ICML '04* : 83. <http://dl.acm.org/citation.cfm?id=1015412>.

Phillips S.J., Anderson R.P. & Schapire R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259. <http://www.sciencedirect.com/science/article/pii/S030438000500267X>.

Phillips S.J., Dudík M. & Schapire R.E. 2016. *Maxent software for modeling species niches and distributions*. 3.4.0. [http://biodiversityinformatics.amnh.org/open\\_source/maxent](http://biodiversityinformatics.amnh.org/open_source/maxent).

Phillips S.J. 2017. *A Brief Tutorial on Maxent*.  
[http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/)  
 Luettu 05.05.2017.

Pohjanmies T., Triviño M., Le Tortorec E., Mazziotta A., Snäll T. & Mönkkönen M. 2017. Impacts of forestry on boreal forests: An ecosystem services perspective. *Ambio* 46: 743–755. doi:10.1007/s13280-017-0919-5.

Polis G.A. & Holt R.D. 1992. Intraguild predation: The dynamics of complex trophic interactions. *Trends Ecol Evol* 7: 151–154. doi:10.1016/0169-5347(92)90208-S [doi].

Polis G.A., Myers C.A. & Holt R.D. 1989. The ecology and evolution of intraguild predation: potential competitors that eat each other. *Annual review of ecology and systematics*.20 : 297–330.

Reich R.M., Joy S.M. & Reynolds R.T. 2004. Predicting the location of northern goshawk nests: modeling the spatial dependency between nest locations and forest structure. *Ecological Modelling* 176: 109–133. doi:10.1016/j.ecolmodel.2003.09.039.

Reynolds R.T., Graham R.T., Reiser M.H., Bassett R.L., Kennedy P.L., Boyce D.A., Goodwin G., Smith R. & Fisher E.L. 1992. Management recommendations for the northern goshawk in the southwestern United States. *RMRS-GTR-217*. doi: 10.2737/RM-GTR-217.

Rutz C. 2006. Home range size, habitat use, activity patterns and hunting behaviour of urban-breeding Northern Goshawks *Accipiter gentilis*. *Ardea* 94: 185–202.

Rutz C. & Bijlsma R.G. 2006. Food-limitation in a generalist predator. *Proceedings of the Royal Society B: Biological Sciences* 273: 2069–2076. doi:10.1098/rspb.2006.3507.

Salminen O. 2015. Keski-Suomen metsien tila ja hakkuumahdollisuudet. <http://jukuri.luke.fi/bitstream/handle/10024/531531/Keski-Suomen%20metsien%20tila%20ja%20hakkuumahdollisuudet.pdf?sequence=1&isAllowed=y> Luettu 17.3.2018.



Saurola P., Valkama J. & Velmala W. 2013. *The Finnish Bird Ringing Atlas*. Finnish Museum of Natural History Luomus and Ministry of Environment, Helsinki.

Selonen V.A.O., Ahlroth P. & Kotiaho J.S. 2005. Anthropogenic disturbance and diversity of species: polypores and polypore-associated beetles in forest, forest edge and clear-cut. *Scand J For Res* 20: 49–58. doi:10.1080/14004080510041002.

Selås V. 1997. Nest-site selection by four sympatric forest raptors in Southern Norway. *Journal of Raptor Research* 31: 16–25.

Selås V., Steen O.F. & Johnsen J.T. 2008. Goshawk breeding densities in relation to mature forest in southeastern Norway. *Forest Ecology and Management* 256: 446–451. doi:10.1016/j.foreco.2008.04.047.

Siitonen J. & Martikainen P. 1994. Occurrence of rare and threatened insects living on decaying *Populus Tremula*: A comparison between Finnish and Russian Karelia. *Scand J For Res* 9: 185–191. doi:10.1080/02827589409382830.

Speiser R. & Bosakowski T. 1987. Nest site selection by northern goshawks in northern New Jersey and southeastern New York. *Condor* 89: 387–394. doi:10.2307/1368492.

Strobl C., Hothorn T. & Zeileis A. 2009a. Party on! *R J* 1: 14–17.

Strobl C., Malley J. & Tutz G. 2009b. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 14: 323–348. doi:10.1037/a0016973.

Sulkava S., Lokki H. & Solonen T. 2014. Kanahaukan pesimäaikainen ravinto kaupunkiympäristössä ja maaseudulla Etelä-Suomessa. *Suomen Riista* 60: 20–30.

Swets J.A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285–1293. doi:10.1126/science.3287615.

Tornberg R. 2001. Pattern of goshawk *Accipiter gentilis* predation on four forest grouse species in northern Finland. *Wildl Biol* 7: 245–256.

Tornberg R., Korpimäki E. & Byholm P. 2006. Ecology of the Northern Goshawk in Fennoscandia. *Stud Avian Biol* : 141.

Widén P. 1997. How, and why, is the Goshawk (*Accipiter gentilis*) affected by modern forest management in Fennoscandia? *The journal of raptor research*. 31: 107–113.

Widén P. 1989. The hunting habitats of Goshawks *Accipiter gentilis* in boreal forests of central Sweden. *Ibis* 131: 205–213. doi:10.1111/j.1474-919X.1989.tb02763.x.

Widén P., Andrén H., Angelstam P. & Lindström E. 1987. The effect of prey vulnerability: goshawk predation and population fluctuations of small game. *Oikos* 49: 233–235. doi:10.2307/3566032.

Östlund L., Zackrisson O. & Axelsson A. 1997. The history and transformation of a Scandinavian boreal forest landscape since the 19th century. *Can J For Res* 27: 1198–1206. doi:10.1139/x97-070. .

## LIITTEET

### LIITE 1. AINEISTOLLE TEHDYT ESIKÄSITTELYT

1. Avattiin Excelissä Tiiratiedot.xlsx-tiedosto, josta valittiin Laji-sarakkeen mukaan kanahaukan pesäpaikat sisältävät rivit. Näistä valittiin rivityyppi-sarakkeen mukaiset havaintorivit, ja edelleen R/T/RT- sarakkeesta R- tai RT- tietoja sisältävät rivit, eli kanahaukan rakentamat risupesät tai alun perin risupesänä olleeseen pesäpaikkaan rakennetut tekopesät, jolloin saatiin kanahaukan valitsemat pesäpaikat. T- eli ihmisen valitsemaan paikkaan rakennetut tekopesät jätettiin pois. Samasta pesästä olevista useammista havainnoista valittiin aikaisin havainto. Pesistä, joiden Rev-ID- sarakkeessa on keskenään sama numero, valittiin pienimmän Random-pesä-sarakkeen arvon mukainen pesä, jolloin kultakin reviiiriltä tuli valituksi vain yksi pesä satunnaisesti. Tämä valinta tallennettiin uudeksi Tiiratiedot.xls-muotoiseksi tiedostoksi.
  2. ArcGisissä Add Data -> lisättiin VMI-taso, jonka koordinaattijärjestelmää projektissa käytetään.
  3. Ladattiin File -> Add Data-> Add XY-Data -toiminnolla tiiratiedot.xls
  4. Toiminnon View-> Data frame properties välilehden Coordinate System lisätoiminnolla Transformation määritettiin muunnos: GCS\_KKJ:stä GCS\_EUREF\_FIN-järjestelmään Using: KKJ\_to\_Euref\_Fin
  6. Analysis tools -> Proximity -> Multiple ring buffer, jossa määritettiin: Input features: Tiiratiedot, Output: tiirabuffer, distances: 20 m, 100 m, 200 m, 250 m, 500 m, 1000 m, 2000 m, 4000 m (Units: meters, Dissolve: None)
  7. Lajiteltiin VMI-tasot ominaisuuksittain kansioihin, esimerkiksi: lehtip\_latvuspeitto\_vmi1x\_1216\_M4.tif, lehtip\_latvuspeitto\_vmi1x\_1216\_N4.tif, lehtip\_latvuspeitto\_vmi1x\_1216\_P4.tif ja lehtip\_latvuspeitto\_vmi1x\_1216\_Q4.tif kansioon le\_latpei. Koska Keski-Suomen alueen kattavaa aineistoa varten tarvittiin kaikki tiedot neljältä eri karttalehdeltä (M4, N4, P4, Q4), kutakin ominaisuustietoa varten tuli neljä eri tiedostoa.
  8. Käytettiin ohjaaja Anssi Lensun luomaa työkalua Mosaic\_rasterit, jonka python-koodi on lopussa liitteenä. Käytettiin batch- eli eräajo-ominaisuutta. Input-workspace with VMI-rasters-kohtaan valittiin aina ominaisuuksittain nimetyt kansiot. Nyt esim. le\_latpei-kansion \*M4.tif-, \*N4.tif, \*P4.tif ja \*Q4.tif- tiedostot yhdistyivät tiedostoksi lehtip\_latvuspeitto\_vmi1x\_1216.tif.
  9. Suoritettiin Spatial analyst tools -> Generalization -> Aggregate eräajona. Cell factor 5, MEAN, Ignore No Data in calculations=true suurimmalle osalle eri ominaisuuksittain järjestetyissä kansioissa oleville yhdistetyille tiedostoille (liite 6), esim. lehtip\_latvuspeitto\_vmi1x\_1216.tif. Output raster (kauttamerkit kuvaavat tallennuskansioita): /Aggregate/5/a5lelatpe Tästä syntynyt tiedostot, joiden nimet ovat muotoa a5ma tai a5lelatpe.
- Luokitelluille muuttujille ajettiin vastaavasti Spatial analyst tools -> Generalization -> Aggregate eräajona. Cell factor 5, MEDIAN, Ignore No Data in calculations=true

Suoritettiin Spatial analyst tools -> Generalization -> Aggregate eräajona. Cell factor 10, MEAN, Ignore No Data in calculations=true osalle eri ominaisuuksittain järjestetyissä kansioissa oleville yhdistetyille tiedostoille, esim. lehtip\_latvuspeitto\_vmi1x\_1216.tif. Output raster: /Aggregate/10/a10kuusi. Tästä syntyi tiedostot, joiden nimet olivat muotoa a10ma tai a10lelatpe.

Osalle muuttujista ajettiin vastaavasti Spatial analyst tools -> Generalization -> Aggregate eräajona. Cell factor 10, MEDIAN, Ignore No Data in calculations=true MEDIAN ja MEAN-erottelu on kerrottu liitteessä 6.

10. Käytettiin eräajona Resamplea Fraluokka-, Kasvupaikka-, Maaluokka- ja Päätyyppi-VMI-muuttujille. Cell size X=80 ja Y=80. Saatiin vastaava solukoko kuin Aggregaten Cell factor 5. Input Raster Dataset= kasvupaikka\_vmi1x\_1216.tif (ja vastaavat fralu-, maalu- tai paatyy-alkuiset tiedostot) Resampling technique=MAJORITY, Snap Raster=ika\_vmi1x\_1216 (yhdistetty VMI-tiedosto), Output raster dataset=rekaspaik, remaaluok, repaatyyp

11. Käytettiin Resample-työkalua Fraluokka-, Kasvupaikka-, Maaluokka- ja Päätyyppi-VMI-muuttujille. Cell size X=160 ja Y=160 Saatiin vastaava solukoko kuin Aggregaten cell factor 10. Input Raster Dataset= kasvupaikka\_vmi1x\_1216.tif (ja vastaavat fralu-, maalu- tai paatyy-alkuiset tiedostot) Resampling technique=MAJORITY, Snap Raster=ika\_vmi1x\_1216 (yhdistetty VMI-tiedosto), Output raster dataset=rekaspaik, remaaluok, repaatyyp

12. Ladattiin Paituli-latauspalvelusta Hallintorajat vuodelta 2017 osoitteesta <https://avaa.tdata.fi/web/paituli/latauspalvelu> mittakaavassa 1:10 000.

13. Tuotiin maakuntarajat ArcGisiin: Add Data -> kansio, johon ladattiin hallintorajat/mml/hallintorajat\_10k/2017/SuomenMaakuntajako\_2017\_10k.shp

14. Erotettiin Keski-Suomi omaksi tasokseen: Avattiin maakuntarajojen ominaisuustietotaulukko -> valittiin Keski-Suomen sisältävä rivi -> suljettiin ominaisuustietotaulukko -> valittiin maakuntajaon shapefilestä oikean hiirenpainikkeen avulla avautuvasta valikosta Export Data, jossa valittiin tallennettavaksi Selected features, Use the same coordinate system as: the data frame ja output feature class= hallintorajat/keskisuomi.shp

15. Rajattiin Keski-Suomen maakuntarajan avulla Aggregatella ja Resamplella luoduista tiedostoista Keski-Suomen alueeseen rajoittuvat tasot. Useana eräajona, rajoitettu alle 30:n tason eriin, jotta muisti riittää: Extract by mask -> input raster a5ma (tai vastaavasti esim. a5lelatpe) -> Snap Raster=ika\_vmi1x\_1216 -> output raster /Extract/a5e/a5ema

16. Muokattiin Tiiratiedot.xls- tiedostoa siten, että poistettiin muut paitsi laji-, X-koord- ja Y-koord sarakkeet, ja nimettiin ne species-, eas-t ja north-sarakkeiksi. Nämä tallennettiin tiiratiedot.csv- tiedostoksi, jota MaxEnt- analyysiohjelma lukee.

17. Avattiin MaxEnt (Maximum Entropy Species Distribution Modeling, Version 3.4.1). Asetettiin Samples-valintalaatikkoon tiiratiedot.csv-tiedosto ja environmental layers -valintalaatikkoon kaikki kansion a5e-aggregate ja 80\*80-solukoon rajatut Resample-tiedostot. Laitettiin ruksit kohtiin Create response curves, Make pictures of predictions, Do jackknife to measure variable importance. Output format = Cloglog ja Output file type=asc. Settings-valikon asetukset pidettiin oletusarvoissa.

18. Tehtiin sama käsittely MaxEntillä a10e- ja 160\*160-solukoon rajatuille Resample-tiedostoille.

Nyt MaxEnt-käsittely oli valmis ja tulokset luettavissa tulosteista.

19. Laskettiin osuudet huonosta, tyypillisestä ja erinomaisesta kanahaukkametsästä MaxEnt-kartassa:

-Avattiin ArcGISissä uusi projekti, toin ensin yksittäisen VMI-tason, sekä MaxEntin 5-solukoon tuloskansioista kanahaukka\_avg.asc-tiedoston.

-Asetin kanahaukka\_avg.asc:n projektion samaksi kuin Data Framen projekti

-Data Management tools -> Raster -> Raster properties -> Calculate statistics. Input raster: kanahaukka\_avg.asc

-Reclassify -> Input raster:kanahaukka\_avg.asc. Reclass field: Value. Classify...: Natural Breaks, Classes:3, Break values(%): 0,46; 0,69; 1

-Avattiin Reclass.asc:n ominaisuustietotaulukko, otettiin count-sarakkeen tiedot exceliin ja kerrottiin kunkin luokan solujen lukumäärä (count) solun pinta-alalla 6400. Laskettu kunkin luokan osuus koko pinta-alasta.

20. Toistettiin kohta 19 solukoon 10 kanahaukka\_avg.asc-tiedostolle. Käytetty pinta-alan kertoimena lukua 25600.

Nyt MaxEnt-analyysiin liittyvät kohdat on suoritettu

Seuraavaksi alkoivat Random Forest -luokitteluun liittyvät työvaiheet:

21. Analysis tools -> Proximity -> Multiple ring buffer -> input features : Tiiratiedot -> Output :tiirabuffer -> distances: 20 m, 100 m, 200 m, 250 m, 500 m, 1000 m, 2000 m, 4000 m (Units: meters, Dissolve: None)

22. Analysis tools -> Proximity -> Multiple ring buffer -> input features : randompoints -> Output :randombuffer -> distances: 20 m, 100 m, 200 m, 250 m, 500 m, 1000 m, 2000 m, 4000 m (Units: meters, Dissolve: None)

23. Tuotiin Yhdistetty.shp ArcGisiin (Taso, jossa yhdistetty koko Keski-Suomen alueelta Metsäkeskuksen kunnittaiset metsävaratietoaineistot)

24. Avattiin Yhdistetty.shp-tason ominaisuustietotaulukko

25. Poistettiin tarpeettomia muuttujia ominaisuustietotaulukosta Delete field-toiminnolla:

realEst, AreDcrs, CmplSt, DataSrc, DtchnG, GrwtPDS,Idntfrs, StndBDD, StndInf, DwDate (1657 riviä, yhteensä taulukossa 560349 riviä, rivit vuosilta 1993-2016, volume 1-180, yli 100 arvoja 10 kpl. mn dmt 0-185), DwMnDmt, DwVolum, MDatSrc, KDatSrc,HDatSrc,LDatSrc, ADatSrc, RDatSrc. MSwLgPr, KSwLgPr, RSwLgPr, HSwLgPr,ASwLgPr, LSwLgPr

26. Avattiin ominaisuustietotaulukosta Field Calculator Dvlpnmn- muuttujalle, ja annettiin parametrit:

```
CalculateField Yhdistetty DvlpnmnC "Reclass (!DvlpnmnC!)" PYTHON_9.3 "def
Reclass (DvlpnmnC) :
\n if (DvlpnmnC == "A0") :
\n return 1
\n elif (DvlpnmnC ==
"S0") :
\n return 2
\n elif (DvlpnmnC == "T1") :
\n return 3
\n elif (DvlpnmnC ==
"T2") :
\n return 4
\n elif (DvlpnmnC == "Y1") :
\n return 5
\n elif (DvlpnmnC ==
"02") :
\n return 6
\n elif (DvlpnmnC == "03") :
\n return 7
\n elif (DvlpnmnC ==
"04") :
\n return 8
\n elif (DvlpnmnC == "05") :
\n return 9
\n else :
\n return 0"
```

27.

Parser : Python, Type: Number, Show Codeblock,

Pre-logic Script code:

```
def Reclash (Devlpmn) :
  if (Devlpmn == "A0") :
    return 1
  elif (Devlpmn == "S0") :
    return 2
  elif (Devlpmn == "T1") :
    return 3
  elif (Devlpmn == "T2") :
    return 4
  elif (Devlpmn == "Y1") :
    return 5
  elif (Devlpmn == "02") :
    return 6
  elif (Devlpmn == "03") :
    return 7
  elif (Devlpmn == "04") :
    return 8
  elif (Devlpmn == "05") :
    return 9
  else :
    return 0
```

Devlpmn =

Reclash (!Devlpmn!)

Tehty vastaava muunnos DvlpmnC -muuttujalle, korvattu Devlpmn koodissa DevlpmnC:llä.

28. Ajettiin Merge- toiminto Metsäkeskuksen kunnittain järjestetyille metsävaratietoaineistolle: Data Management tools -> General -> Merge (Input datasets kaikki kunnat, joissa on Tiira-tiedoston mukaan kanahaukan pesähavainnot, Output dataset: kskunnatkaikki)

29. Analysis tools -> Overlay -> Intersect (Input features: tiirabuffer, kskunnatkaikki ; Output feature class: tiirametsa ; JoinAttributes : ALL ; Output type: INPUT)

30. Analysis tools -> Overlay -> Intersect (Input features: randombuffer, kskunnatkaikki ; Output feature class: rndmetsa ; JoinAttributes : ALL ; Output type: INPUT)

31. Ajettiin R:ssä ohjaajani Anssi Lensun tekemällä koodilla komennot:

```
library (maptools)
library (mclust)
library (randomForest)
library (gdata)
library (mice)
library (VIM)
library (dplyr)
library (caret)
library (party)
library (varSelRF)
library (Boruta)
library (rgdal)
```

```

pamuut <- c("MeanAge", "BasalAr", "StemCnt", "MeanDmt", "MenHght", "Volum
e",
"SawLgVl", "PulpWdV", "VolmGrw", "LeafBms", "BrnchBm",
"MAge", "MBaslAr", "MMenDmt", "MMnHght", "MStmCnt", "MVVolume", "MSwLgVl"
,
"MPlpWdV", "MLefBms", "MBrnchB",
"KAge", "KBaslAr", "KMenDmt", "KMnHght", "KStmCnt", "KVVolume", "KSwLgVl"
,
"KPlpWdV", "KLefBms", "KBrnchB",
"RAge", "RBaslAr", "RMenDmt", "RMnHght", "RStmCnt", "RVVolume", "RSwLgVl"
,
"RPlpWdV", "RLefBms", "RBrnchB",
"HAge", "HBaslAr", "HMenDmt", "HMnHght", "HStmCnt", "HVVolume", "HSwLgVl"
,
"HPlpWdV", "HLefBms", "HBrnchB",
"AAge", "ABaslAr", "AMenDmt", "AMnHght", "AStmCnt", "AVVolume", "ASwLgVl"
,
"APlpWdV", "ALefBms", "ABrnchB",
"LAge", "LBaslAr", "LMenDmt", "LMnHght", "LStmCnt", "LVVolume", "LSwLgVl"
,
"LPPlpWdV", "LLefBms", "LBrnchB")
mimamuut <- c("StndQlt")
majmuut <- c("FrtltyC", "MnTrSpC")

n.na <- function(x) {
  sum(as.integer(is.na(x)))
}

na2zero <- function(x) {
  new.x <- ifelse(is.na(x), 0, x)
}

# dist-muuttujalle tulee kokeilla arvoja 200, 250, 500, 1000, 2000 ja 400
0
# poimien niitä tässä yksi kerrallaan ja sitten ajamalla alla olevat käs
kyt
for(dist in c(200, 250, 500, 1000, 2000, 4000)) {

print(paste("Current distance", dist))

for(prefix in c("r", "t")) {
  tied <- paste(prefix, as.character(dist), "_yhd.shp", sep="")
  if(!file.exists(tied)) next
  sptied <- readShapeSpatial(tied)
  src <- sptied@data
  if(prefix == "t") fldnam <- "PESA_ID"
  else fldnam <- "Id"
  tulokset <- data.frame(ID=tapply(src[,fldnam], src[,fldnam], head, n=1)
,
COUNT=tapply(src[,fldnam], src[,fldnam], length)
)
  for(muut in majmuut) {
    tap_res <- tapply(src[,muut], src[,fldnam], majorityVote)
    tulosvec <- tap_res[[1]]$majority
    for(i in 2:length(tap_res))
      tulosvec <- c(tulosvec, tap_res[[i]]$majority)
    tulos <- data.frame(f=tulosvec)
    names(tulos) <- muut
    if(!exists("tulokset")) tulokset <- tulos
    else tulokset <- cbind(tulokset, tulos)
  }
}

```

```

for(muut in mimamuut) {
  pesa.min <- tapply(src[,muut], src[,fldnam], min)
  pesa.max <- tapply(src[,muut], src[,fldnam], max)
  tulos <- data.frame(f1=pesa.min, f2=pesa.max)
  names(tulos) <- paste(c("min", "max"), muut, sep="")
  tulokset <- cbind(tulokset, tulos)
}
for(muut in pamuut) {
  locSrc <- src #[src[,muut] != 0, ]
  p.alat <- locSrc$Area
  sum.p.alat <- tapply(p.alat, locSrc[,fldnam], sum)
  painotettu <- p.alat * locSrc[,muut]
  pesa.pain <- tapply(painotettu, locSrc[,fldnam], sum)
  inds <- match(as.character(tulokset$ID), names(pesa.pain))
  tulos <- data.frame(f=c(pesa.pain / sum.p.alat, NA))
  inds <- ifelse(!is.na(inds), inds, length(tulos[,1]))
  utulos <- data.frame(tulos[inds,])
  names(utulos) <- muut
  tulokset <- cbind(tulokset, utulos)
}
for(luokka in c("36200", "32600", "42210", "35400", "lsohj", "lsysa",
               "lsvsa", "natura")) {
  tied <- paste(prefix, as.character(dist), "_", luokka, ".shp", sep="")
)
  if(!file.exists(tied)) next
  tinfo <- ogrInfo(tied)
  if(tinfo$nrows == 0) next
  leikk <- readShapeSpatial(tied, repair=T)
  lkm <- length(leikk[,1])
  Areas <- rep(0, lkm)
  for(kuv in 1:lkm) Areas[kuv] <- leikk@polygons[[kuv]]@area
  sum.p.alat <- tapply(Areas, leikk@data[,fldnam], sum)
  inds <- match(as.character(tulokset$ID), names(sum.p.alat))
  tulos <- data.frame(f=c(sum.p.alat, 0))
  inds <- ifelse(!is.na(inds), inds, length(tulos[,1]))
  utulos <- data.frame(tulos[inds,])
  names(utulos) <- paste("Area", luokka, sep="_")
  tulokset <- cbind(tulokset, utulos)
}
if(prefix == "t") mv("tulokset", "Tiira")
else mv("tulokset", "RndGen")
}
# Tähän päättyy datojen valmistelu ja tuloksena on kehykset Tiira ja RndG
en

# Tutkitaan puuttuvuuden määriä ja poistetaan vähintään 10 puuttuvaa
# arvoa sisältävät muuttujat sekä lasketaan puuttuvuudet uudelleen
Tiira.na <- unlist(lapply(Tiira, n.na))
RndGen.na <- unlist(lapply(RndGen, n.na))
Tdata <- Tiira[,Tiira.na < 10]
Rdata <- RndGen[,RndGen.na < 10]
Td.na <- unlist(lapply(Tdata, n.na))
Rd.na <- unlist(lapply(Rdata, n.na))

# Onko meillä datassa (vähäistä) puuttuvuutta muuttujissa
if(sum(Td.na, Rd.na) > 0) {
  # Jos on, havainnollistetaan jäljelle jäänyt puuttuvuus muuttujittain
  #print(md.pattern(Tdata))
  #print(md.pattern(Rdata))
  windows(title="Tiira-datan puuttuvuuskuvaajat")
  mplot_Td <- aggr(Tdata, col=c('navyblue','yellow'),

```



```

        numbers=TRUE, sortVars=TRUE, labels=names(Tdata),
        cex.axis=.7, gap=3, ylab=c("Missing data", "Pattern"))
windows(title="Random-pisteiden puuttuvuuskuvaajat")
mplot_Rd <- aggr(Rdata, col=c('navyblue','yellow'),
        numbers=TRUE, sortVars=TRUE, labels=names(Rdata),
        cex.axis=.7, gap=3, ylab=c("Missing data", "Pattern"))

# Seuraavana puuttuvien tietojen täydentäminen mice-paketilla
TdImp <- mice(Tdata, m=5, maxit=50, method='pmm', seed=300)
RdImp <- mice(Rdata, m=5, maxit=50, method='pmm', seed=300)

# Havainnollistetaan yksi imputoitu tulos (number 2 of 5)
windows(title="Tiiran sirontakuvaajien matriisi")
plot(complete(TdImp,2)[,Td.na > 0], col=ifelse(complete.cases(Tdata), 1
, 2))
windows(title="Satunnaispisteiden sirontakuvaajat")
plot(complete(RdImp,2)[,Rd.na > 0], col=ifelse(complete.cases(Rdata), 1
, 2))
} else {
# Tässä osuudessa oletuksena on, että datassa ei ole puuttuvuutta!!!
set.seed(7) # Varmistetaan johdonmukaiset tulokset (voit toki testata
# useita erilaisia siemenluvun arvoja)

# Täällä luokitellaan vain yksi data, jossa ei ollut puuttuvuutta
# Aineisto saadaan yhdistämällä Tdata ja Rdata (luokkatieto lisättävä)
opetus <- bind_rows(transform(Tdata, class=1), transform(Rdata, class=0
))

# Tarkistetaan vielä, puuttuuko jokin selittäjä toisesta datakehuksesta
# (siis Tiiran puolelta tai satunnaispisteiden puolelta)
op.na <- unlist(lapply(opetus, n.na))
print(noquote(paste("Puuttuvuutta muuttujassa: ", names(opetus)[op.na >
0],
        ", nollataan", sep="")))
muutt <- names(opetus)[op.na != 0]
for(mm in muutt) {
  opetus[,mm] <- na2zero(opetus[,mm])
}
opetus$class <- factor(opetus$class, levels=0:1,
        labels=c("RndPt", "TiiraPt"))
training <- opetus[, 2:length(opetus)]

# 2. Random Forest -luokittelija paketista randomForest
# Tässä on mukana muuttujien määrän / puu optimointi
rfCtrl <- trainControl(method="repeatedcv", number=10, repeats=3,
        search="grid")
mtry <- sqrt(ncol(training)-1)
tuneGrid <- expand.grid(.mtry=1:as.integer(round(2*mtry,0)))
rfFit <- train(class ~ ., data=training, method="rf", metric="Accuracy"
,
        tuneGrid=tuneGrid, tuneLength=15, trControl=rfCtrl,
        importance=TRUE)
print(summary(rfFit$finalModel$err.rate[,1]))
#windows(title="Random Forestin muuttujien määrän/puu valinta")
#plot(rfFit)

# Selvitetään luokittelun tärkeimmät selittävät muuttujat
# äskeisen mtry-parametrin optimoinnin tuloksena saadusta mallista
png(paste("D", dist, "_RF_varImp.png", sep=""), 1800, 1800, res=300)
varImpPlot(rfFit$finalModel, cex=0.8)

```

```

dev.off()

# 3. Kun meillä on useita korreloivia muuttujia (osa voimakkaastikin),
# tässä on vielä toteutus party-paketin cforest()-funktioilla
crfCtrl <- trainControl(method="repeatedcv", number=10, repeats=3,
                        classProbs=TRUE, summaryFunction=twoClassSummar
y)
mtry <- round(sqrt(ncol(training)-1),0)
crfFit <- train(class ~ ., data=training, method="cforest", metric="Acc
uracy",
                tuneGrid=expand.grid(.mtry=as.integer(mtry/2):(3*mtry))
,
                trControl=crfCtrl, controls=cforest_unbiased(ntree=500)
)
print(table(training$class, predict(crfFit$finalModel, OOB=TRUE)))
#windows(title="cForestin muuttujien lukumäärän / puu valinta")
#plot(crfFit)

# Selvitetään luokittelun tärkeimmät selittävät muuttujat
# äskeisen mtry-parametrin optimoinin tuloksena saadusta mallista
vi <- varimp(crfFit$finalModel, conditional=TRUE)
png(paste("D", dist, "_cF_varImp.png", sep=""), 2700, 1500, res=300)
par(mar=c(4,3,0.5,0.5))
bpres <- barplot(vi, cex.names=0.55, cex.axis=0.8, las=2)
axis(1, at=bpres[,1]+0.1, lwd=0, lwd.ticks=1, labels=F)
dev.off()

# 4. Erillinen muuttujien merkityksen tarkastelu varSelRF-paketilla
# Tämä käyttää tavallista randomForest-paketin luokittelijaa
set.seed(12)
vSelRes <- varSelRF(training[,-length(training)], Class=training$class,
                   ntree=500, ntreeIterat=300, mtryFactor=3,
                   vars.drop.frac = 0.2)
print(vSelRes$selec.history)
print(vSelRes)
# Yksi paras ratkaisu (OOB:n perusteella valittu):
print(vSelRes$selec.history$Vars.in.Forest[which.min(vSelRes$selec.hist
ory$OOB)])

# 5. Muuttujien merkityksen tarkastelu Boruta-paketilla
bor_out <- Boruta(class ~ ., data=training, doTrace=2)
boruta_sig <- names(bor_out$finalDecision[bor_out$finalDecision %in% c(
"Confirmed", "Tentative")])
print(boruta_sig)
png(paste("D", dist, "_Boruta_varImp.png", sep=""), 2700, 1500, res=300
)
par(mar=c(4,2.5,0.5,0.5), mgp=c(1.5,0.7,0))
plot(bor_out, cex.axis=.55, las=2, xlab="", main="",
      colCode=c(rgb(217/255, 95/255, 14/255), rgb(254/255, 196/255, 79/2
55),
              rgb(1, 247/255, 188/255), "black"))
dev.off()
png(paste("D", dist, "_Boruta_history.png", sep=""), 1800, 1800, res=30
0)
plotImpHistory(bor_out)
dev.off()
}

cf <- crfFit$finalModel
# Yhden puun kuvaaja (nätti versio)

```

```
pt <- party::prettytree(cf@ensemble[[1]], names(cf@data@get("input")))
pt
pt$weights <- rep(1, length(cf@responses@variables$.outcome))
pt_upd <- update_tree(pt, cf@data@get("input"))
nt <- new("BinaryTree")
nt@tree <- pt_upd
nt@data <- cf@data
nt@responses <- cf@responses
nt@cond_distr_response <- cf@cond_distr_response
nt@predict_response <- cf@predict_response
nt
png(paste("D", dist, "_PrettyTree.png", sep=""), 5400, 1800, res=300)
plot(nt)
dev.off()

# Siltä varalta, että jatketaan eri vaikutusalueella saman R-session
aikana
rm(Tiira); rm(RndGen)
```

## LIITE 2. VMI-AINEISTON KASVUPAIKKA-LUOKAN LUOKKANIMET

1 *Lehdot* sekä lehtomaiset ja lettosuot (luonnontilaiset, ojikko- ja muuttumasuot) ja lehtoturvekankaat.

2 *Lehtomaiset kankaat* ja ruohoiset suot sekä turvekankaat.

3 *Tuoreet kankaat* ja suursaraiset sekä mustikkaiset suot ja turvekankaat.

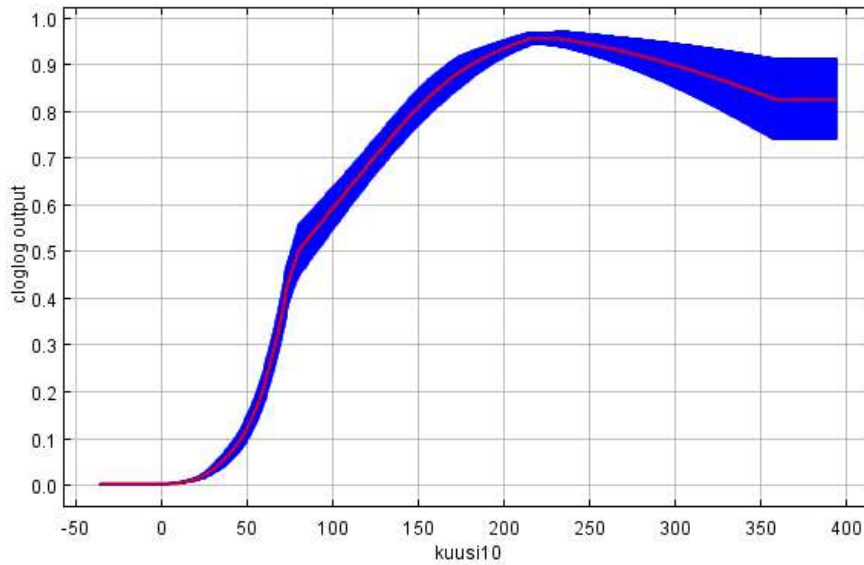
4 *Kuivahkot kankaat* ja piensaraiset sekä puolukkaiset suot ja turvekankaat.

5 *Kuivat kankaat* ja tupasvillaiset sekä isovarpuiset suot ja turvekankaat.

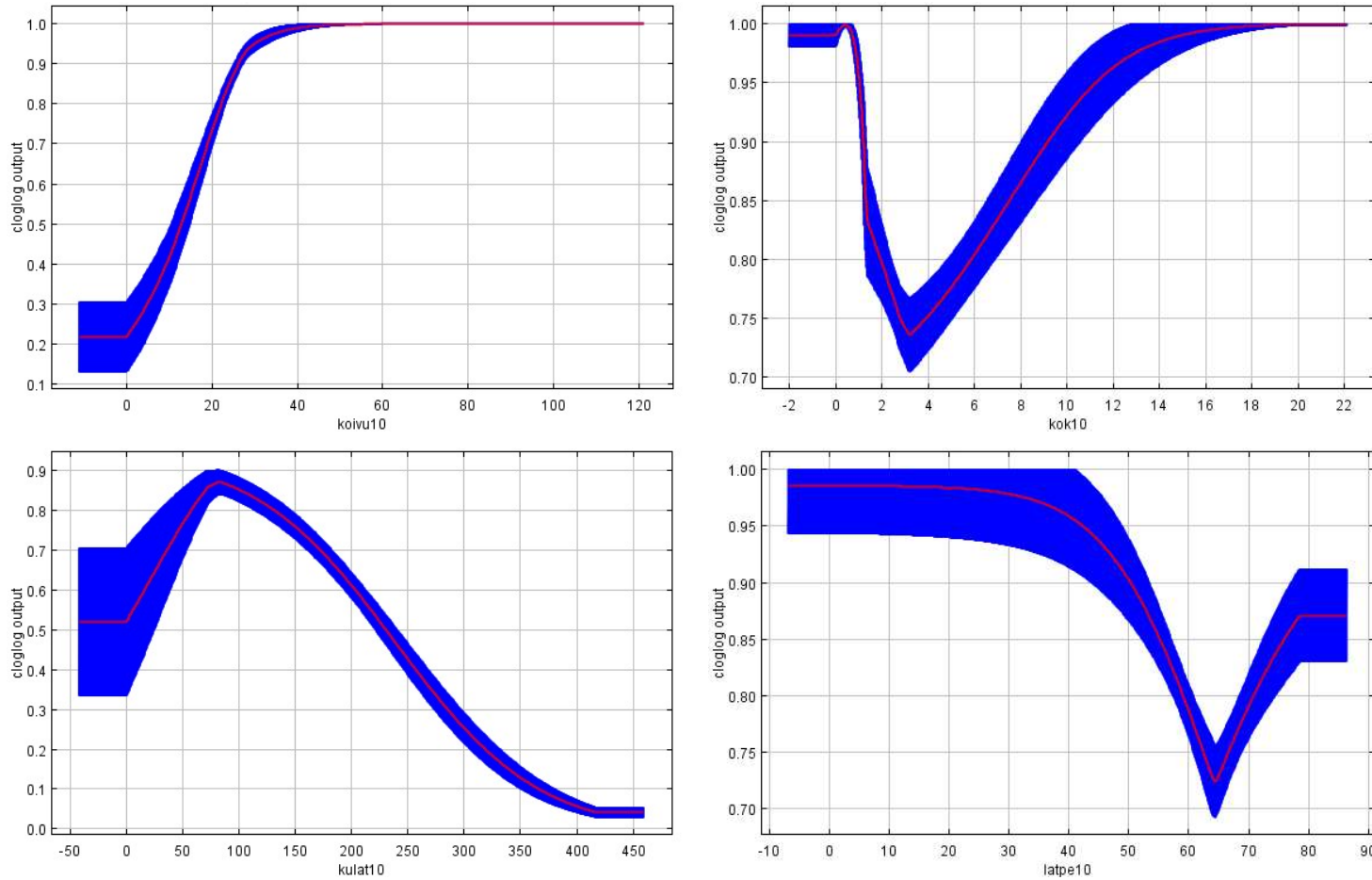
6 *Karukkokankaat* ja rahkaiset suot sekä turvekankaat.

7 *Kalliomaat* ja hietikot sekä vesijättömaat.

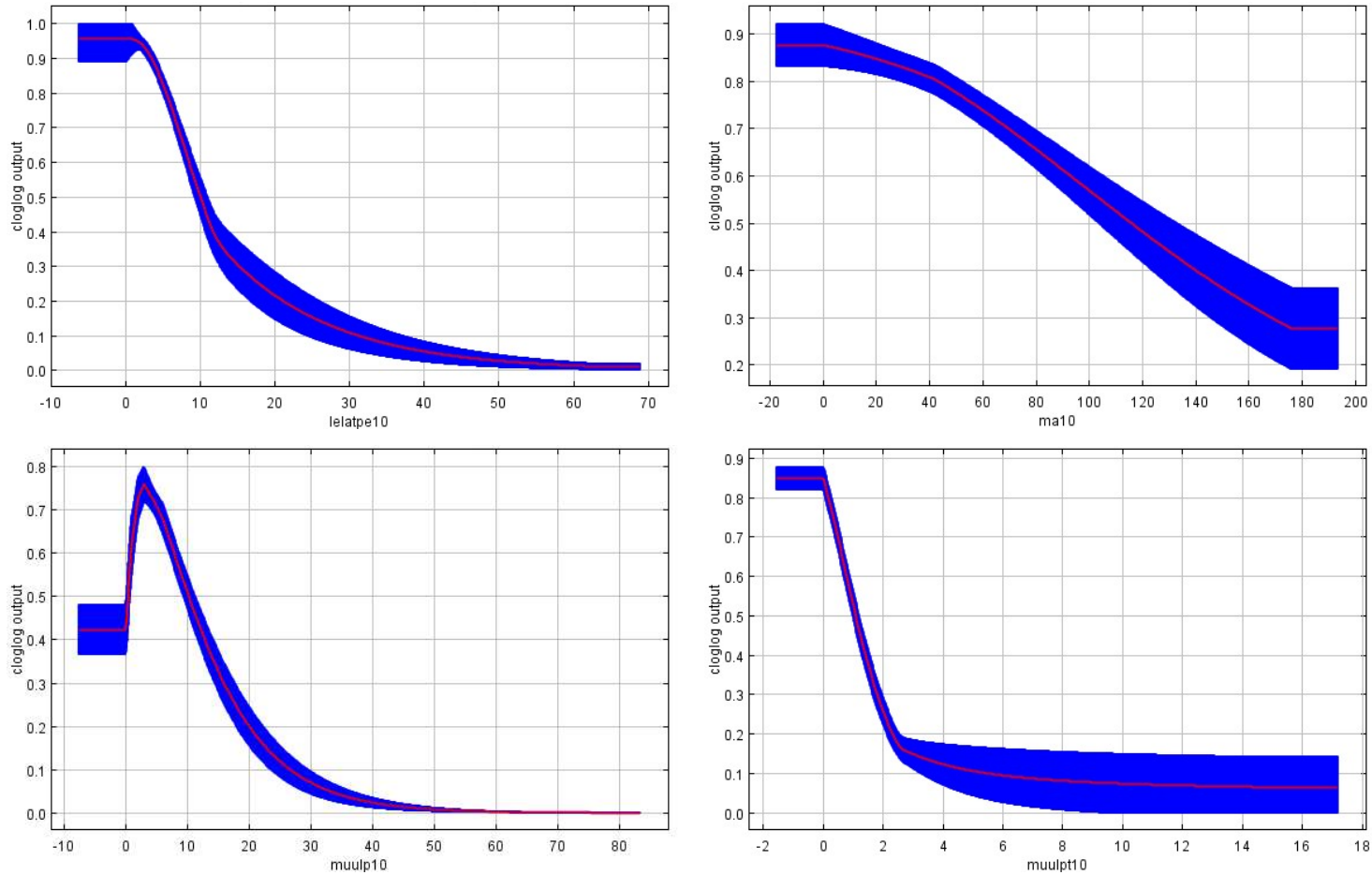
(Korhonen 2009)

**LIITE 3. SOLUKOON 160 M TULOSKUVAAJIA**

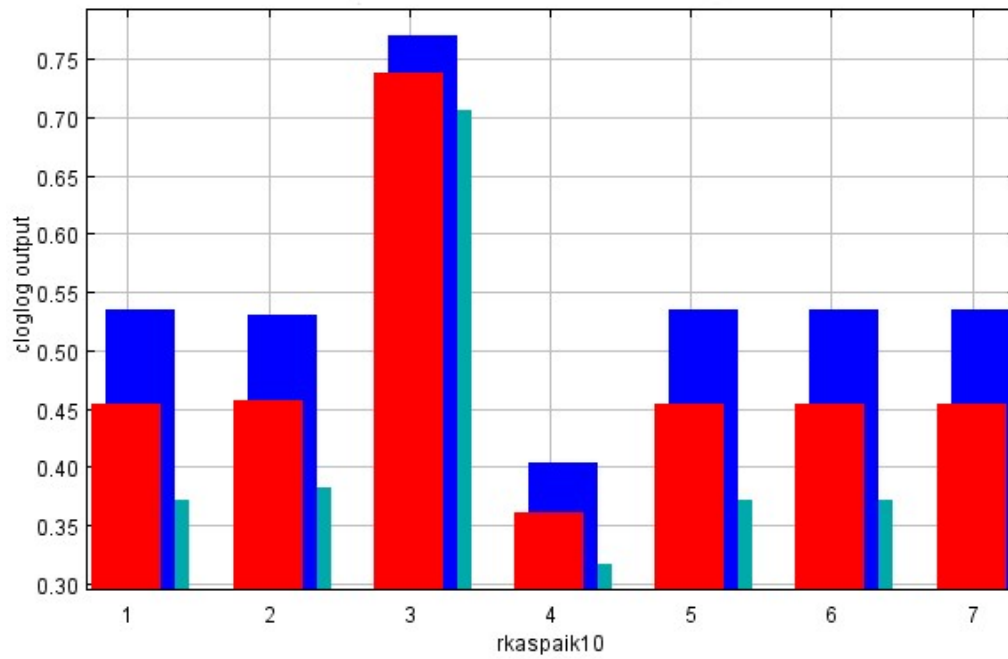
Kuva 1. Vastekuvaaja kuusen tilavuudelle solukoon 160 m × 160 m aineistolla, kun kuusen tilavuuden arvoa muutellaan ja muiden muuttujien arvot ovat keskiarvossaan. X-akselilla kuusitukin tilavuus kuutiometreinä hehtaarilla.



Kuva 2. Vastekuvaajat solukoolla  $160 \text{ m} \times 160 \text{ m}$ , kun kuvatus muuttujan arvoa muutetaan muiden muuttujien pysyessä keskiarvossaan. Ylhäällä vasemmalta koivun tilavuus ja koivun kuitupuun tilavuus, alhaalla vasemmalta kuusen hukkapuosuuden tilavuus ja koko puuston latvuspeittävyys.

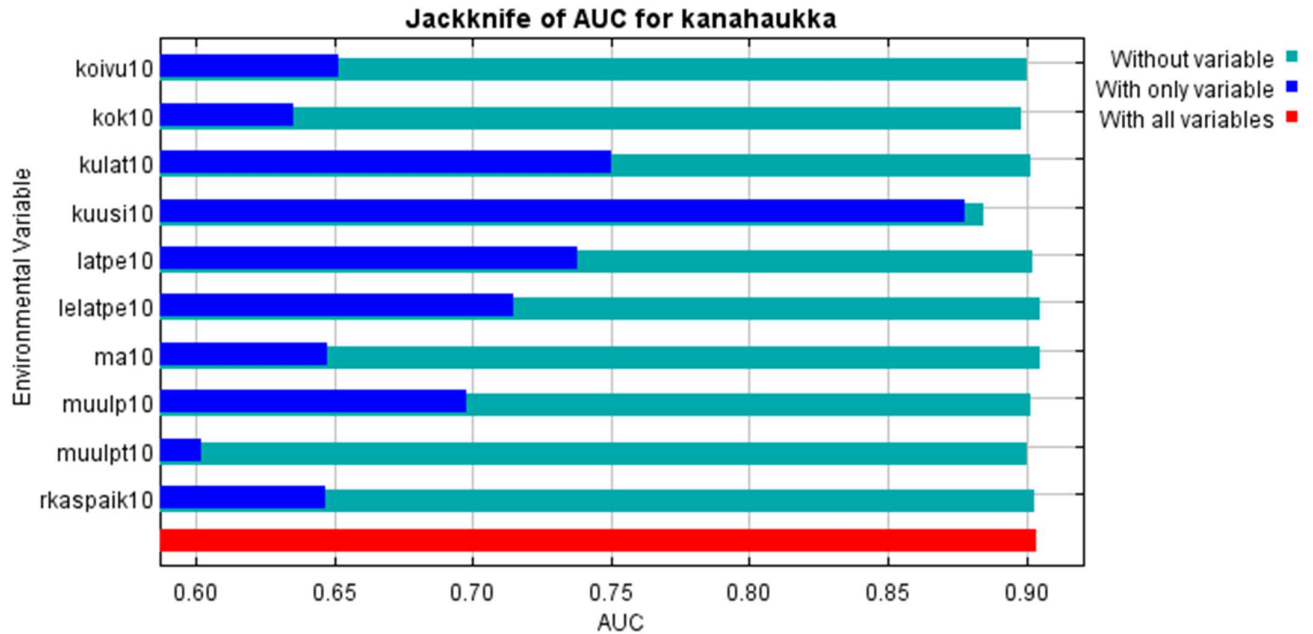


Kuva 3. Vastekuvaajat solukoolla  $160 \text{ m} \times 160 \text{ m}$ , kun kuvatus muuttujan arvoa muutetaan muiden muuttujien pysyessä keskiarvossaan. Ylhäällä vasemmalta lehtipuuston latvuspeittävyys ja männyn tilavuus, alhaalla vasemmalta muun lehtipuun tilavuus ja muun lehtipuun tukkipuun tilavuus.

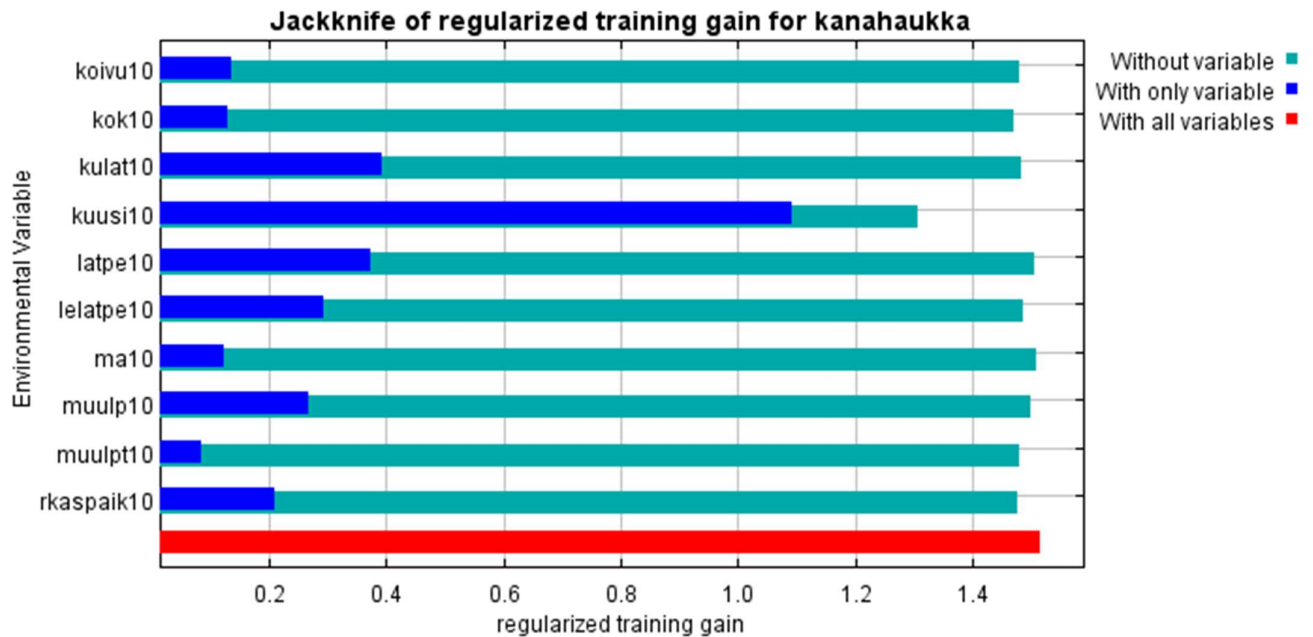


Kuva 4. Vastekuvaaja kasvupaikkatyyppi-muuttujalle solukoolla  $160 \text{ m} \times 160 \text{ m}$ , kun kuvatus muuttujan arvoa muutetaan muiden muuttujien arvojen pysyessä keskiarvossaan.



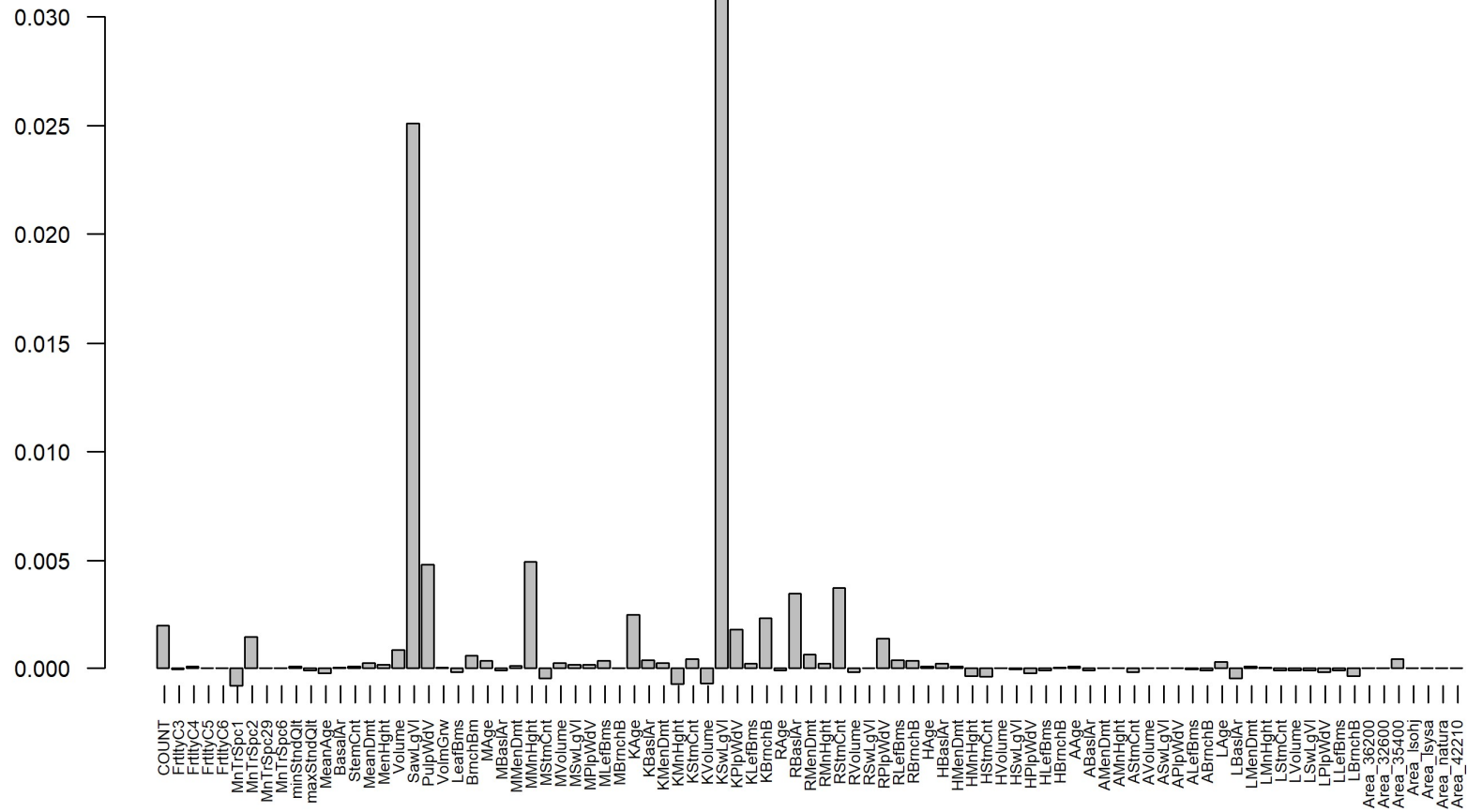


Kuva 5. MaxEntin-mallin muuttujien merkitsevyysarvioinnin tulokset AUC:lle 160 m × 160 m solukoolla kuusen tilavuuden sisältävällä mallilla. Selite kuten kohdan 4.2 kuvassa 9. Muuttujien nimet on esitetty kohdan 4.2 taulukossa 3.

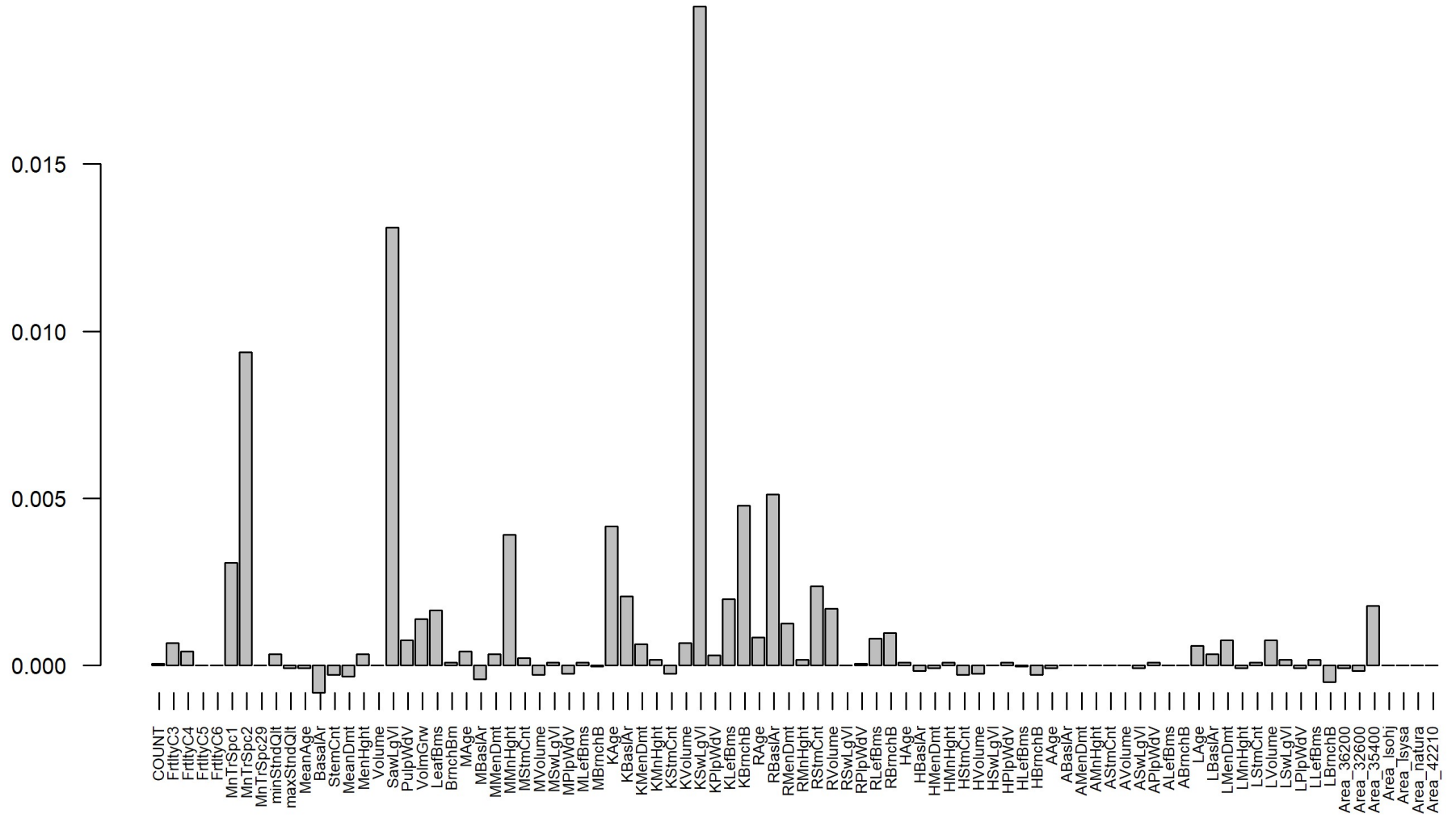


Kuva 6. MaxEntin-mallin muuttujien merkitsevyysarvioinnin tulokset regularized training gainille 160 m × 160 m solukoolla kuusen tilavuuden sisältävällä mallilla. Selite kuten kohdan 4.2 kuvassa 9. Muuttujien nimet on esitetty kohdan 4.2 taulukossa 3.

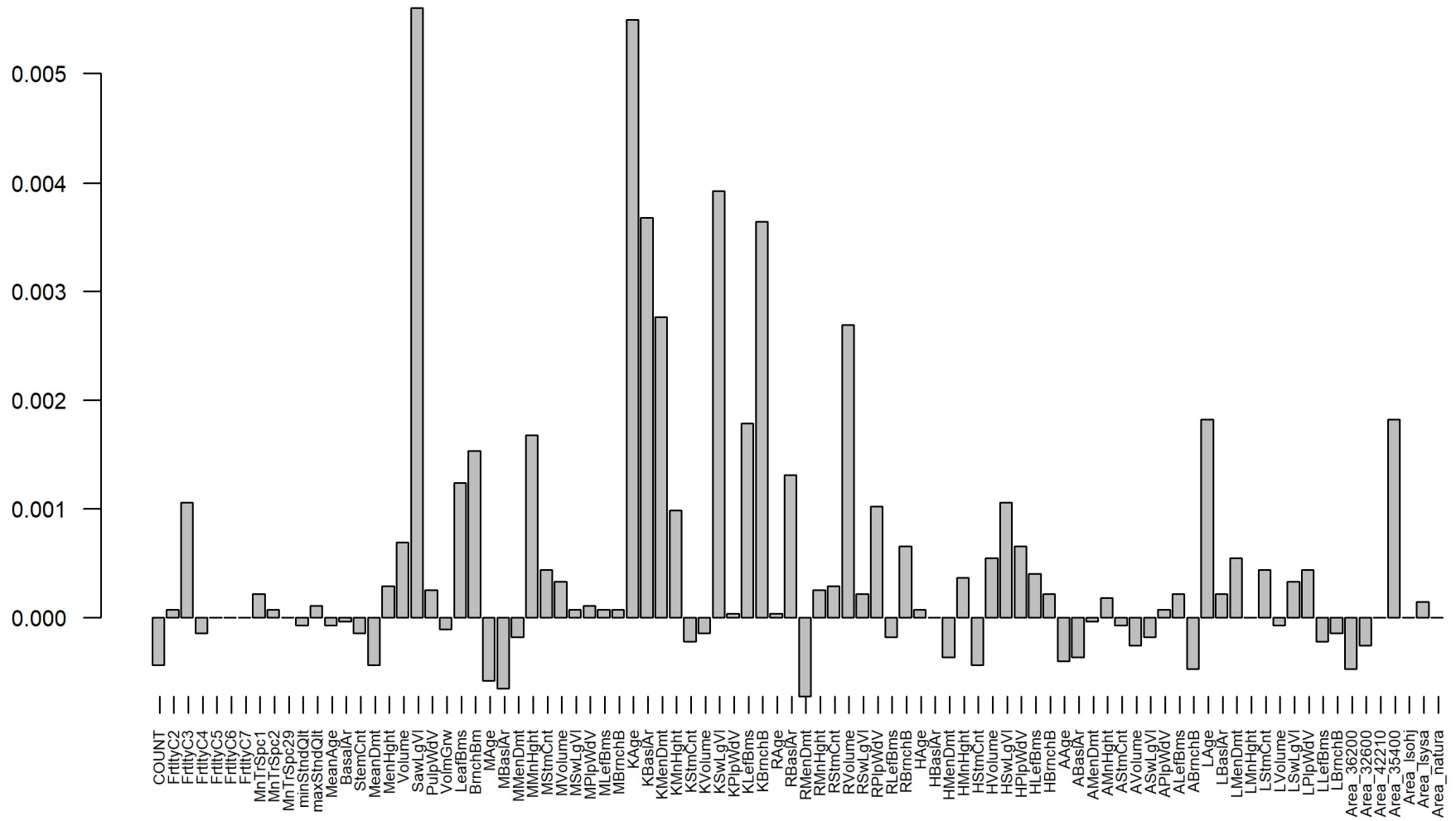
#### LIITE 4. CFORREST-AINEISTON TULOSKUVAAJAT



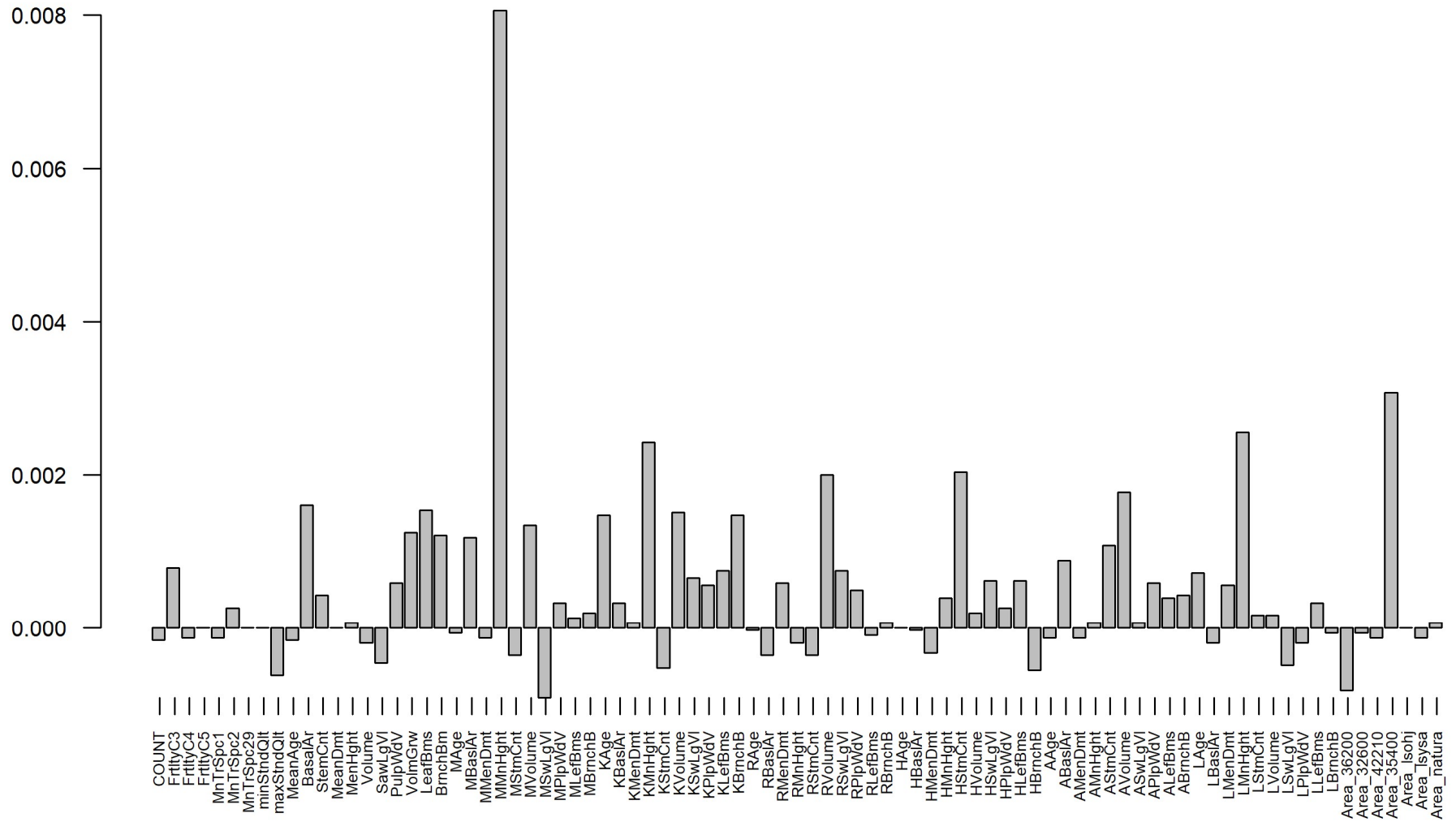
Kuva 1. cForest-analyysin tulostulokuvaaaja 200 m säteellä. Muuttujien nimet on esitetty osion 3.2.3. taulukossa 1.



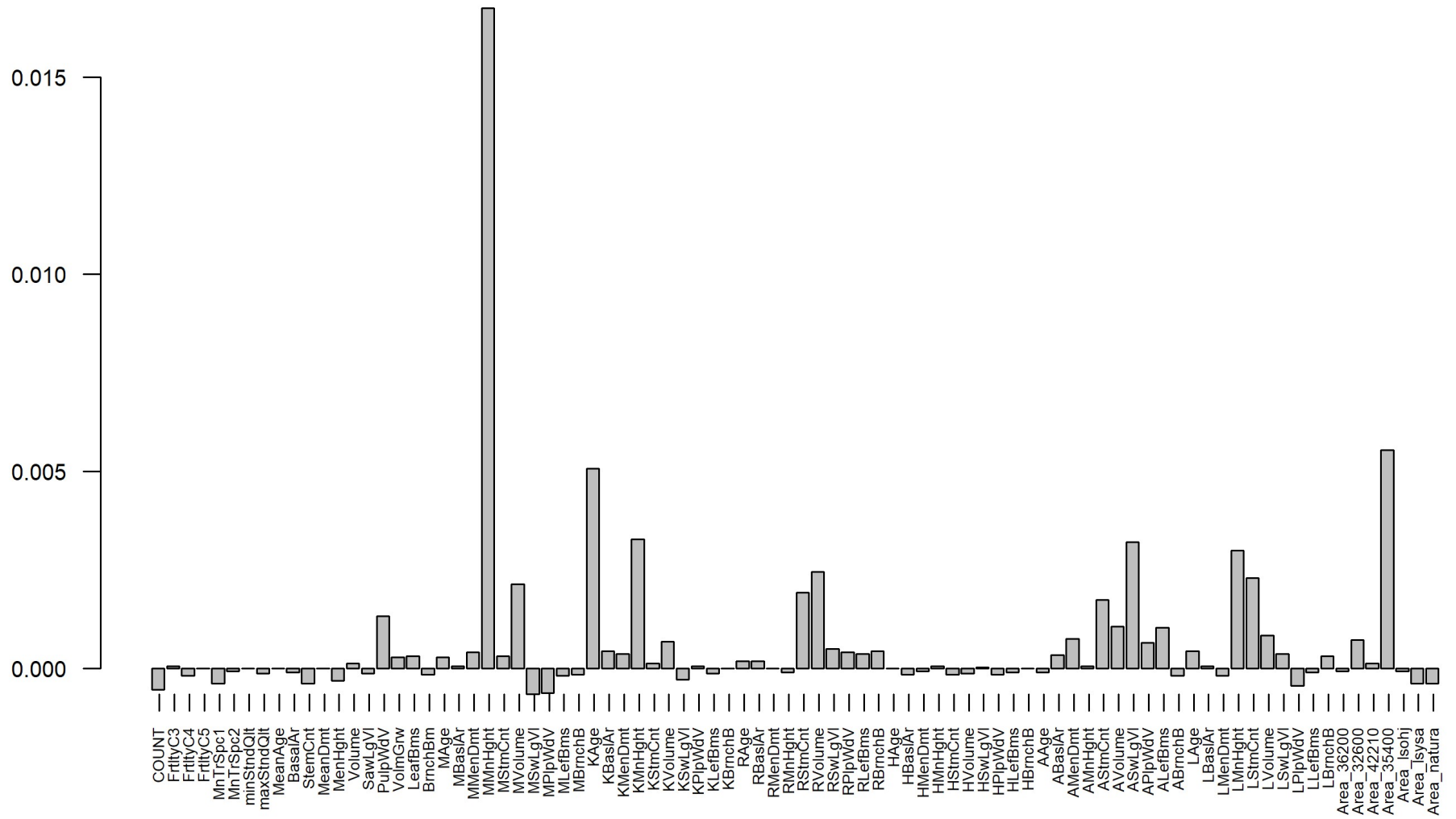
Kuva 2. cForest-analyysin tuloskuvaaja 250 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1.



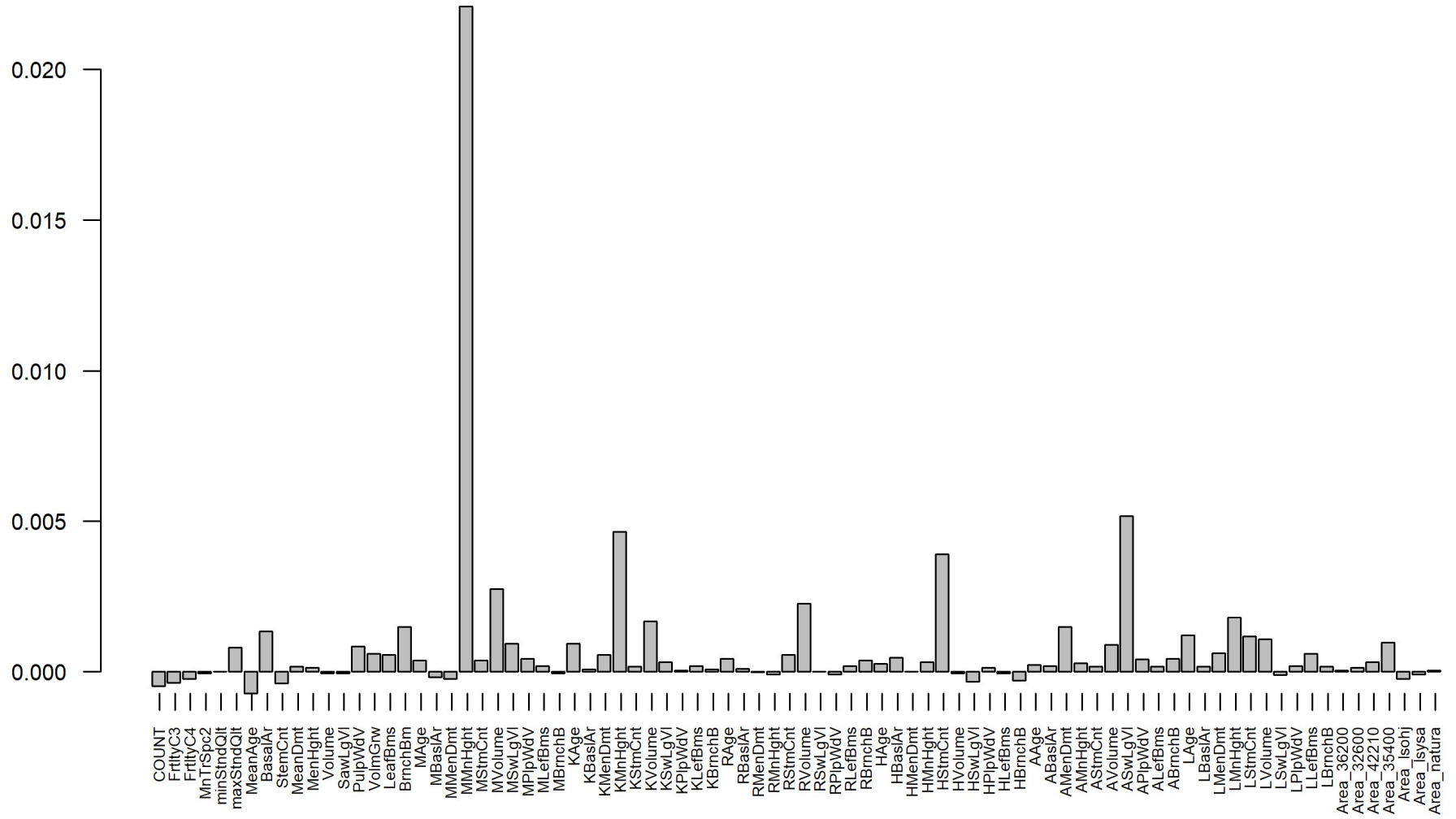
Kuva 3. cForest-analyysin tulostulostus 500 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1.



Kuva 4. cForest-analyysin tuloskuvaaja 1000 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1.



Kuva 5. cForest-analyysin tuloskuvaaja 2000 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1.

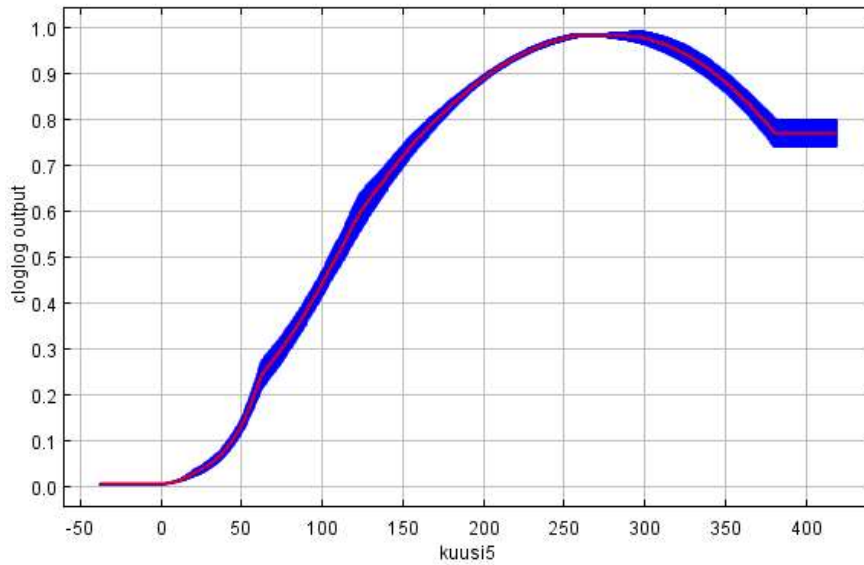


Kuva 6. cForest-analyysin tuloskuvaaja 4000 m säteellä. Muuttujien nimet on esitetty osion 3.2.3 taulukossa 1.

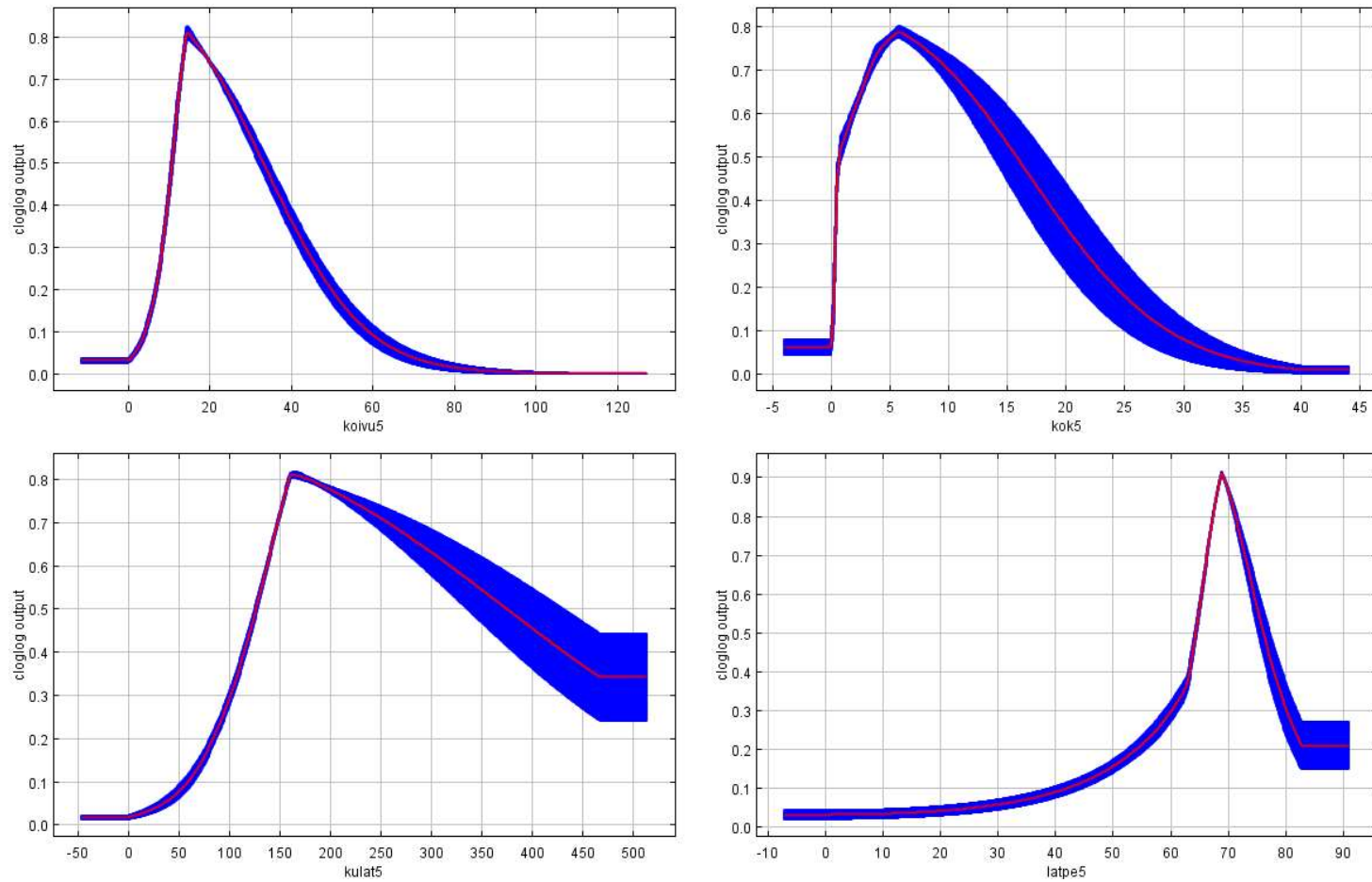
Taulukko 1. Virheiden osuus testattaessa luokittelun onnistumista satunnaispisteisiin ja pesäpisteisiin cForest-analyysillä. Mitä pienempi virheiden osuus, sen paremmin luokittelu onnistuu.

<b>Säde</b>	<b>Virheiden osuus (%)</b>
200	39,47
250	29,32
500	33,33
1000	33,73
2000	31,82
4000	31,07

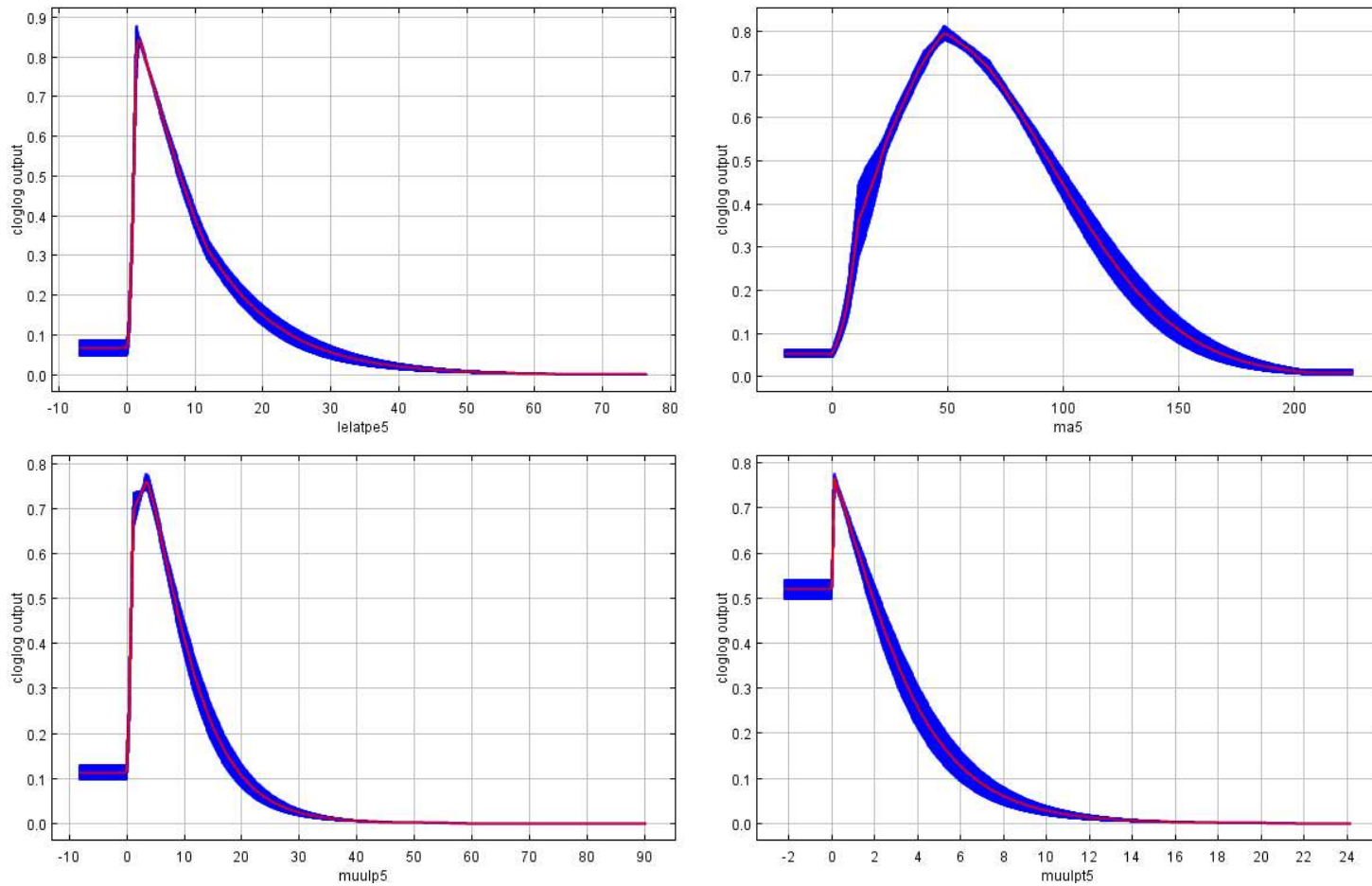


**LIITE 5. SOLUKOON 80 M TULOSKUVAAJAT**

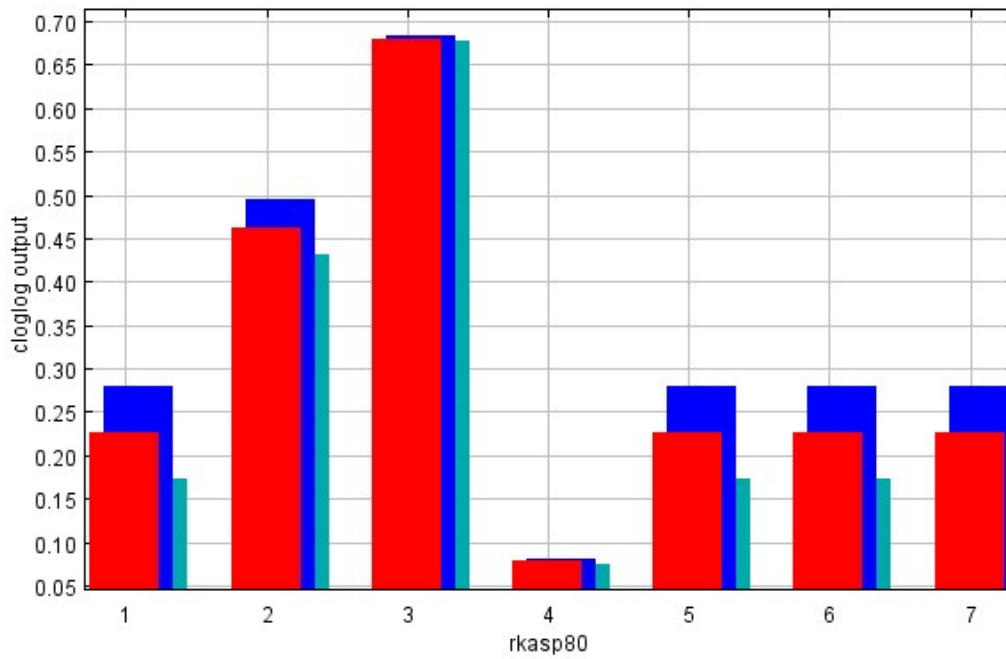
Kuva 1. Vastekuvaaja kuusen tilavuudelle solukoon  $80\text{ m} \times 80\text{ m}$  aineistolla, kun mallissa ei ole mukana muita muuttujia. X-akselilla kuusitukin tilavuus kuutiometreinä hehtaarilla.



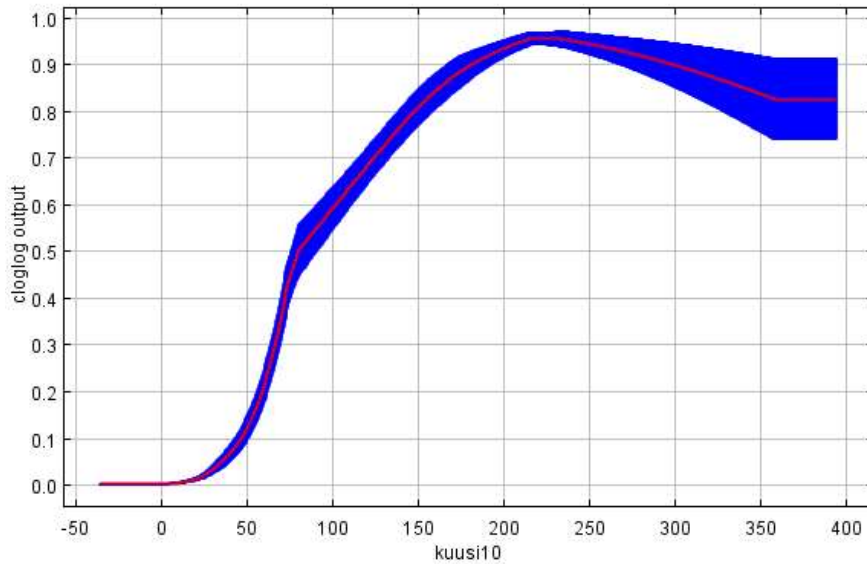
Kuva 2. Vastekuvaajat solukoolla  $80 \text{ m} \times 80 \text{ m}$ , kun kuvatun muuttujan arvoa ilman, että mallissa on mukana muita muuttujia. Ylhäällä vasemmalta koivun tilavuus ja koivun kuitupuun tilavuus, alhaalla vasemmalta kuusen hukkapuuosuuden tilavuus ja koko puuston latvuspeittävyys.



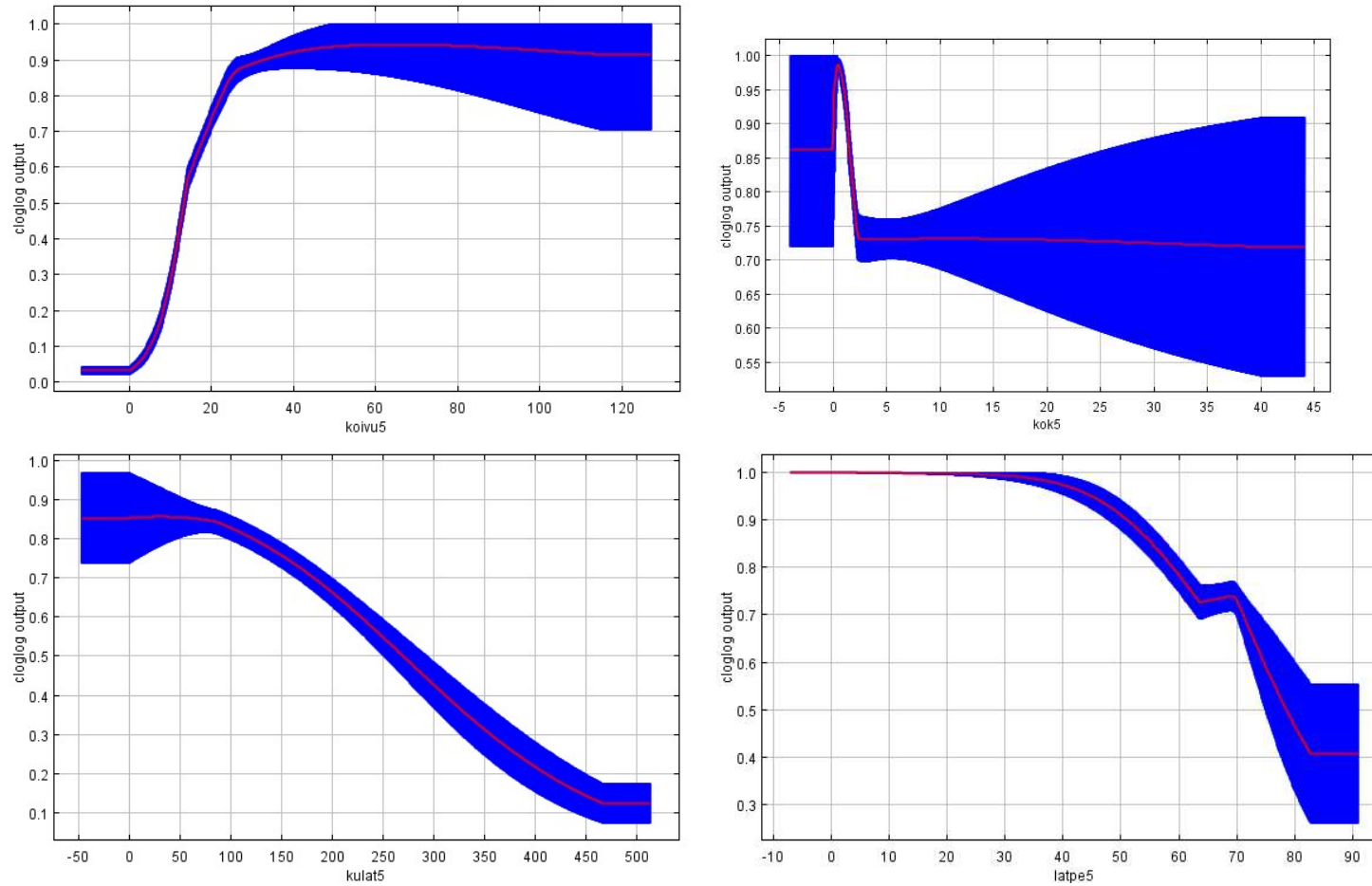
Kuva 3. Vastekuvaajat solukoolla  $80 \text{ m} \times 80 \text{ m}$ , kun kuvatun muuttujan arvoa muutetaan ilman, että mallissa on mukana muita muuttujia. Ylhäällä vasemmalta lehtipuuston latvuspeittävyys ja männyn tilavuus, alhaalla vasemmalta muun lehtipuun tilavuus ja muun lehtipuun tukkipuun tilavuus.



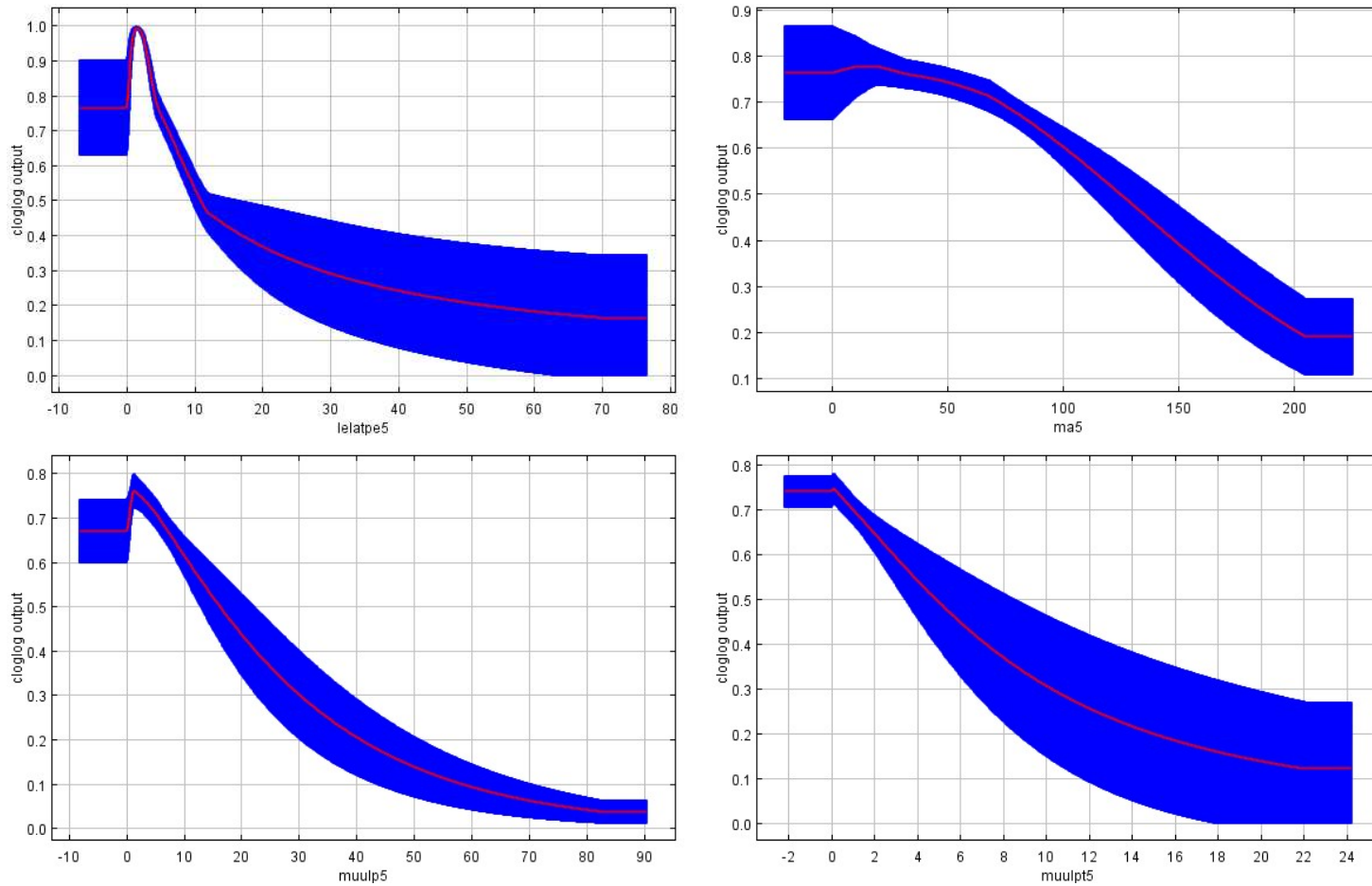
Kuva 4. Vastekuvaaja kasvupaikkatyyppi-muuttujalle solukoolla  $80 \text{ m} \times 80 \text{ m}$ , kun mallissa ei ole mukana muita muuttujia.



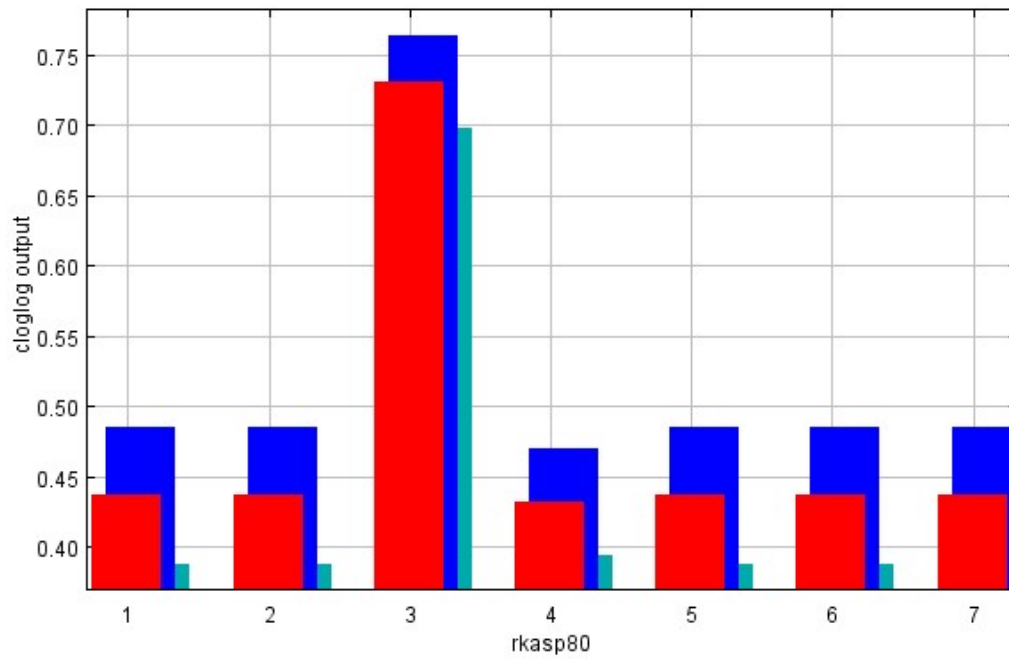
Kuva 5. Vastekuvaaja kuusen tilavuudelle solukoon  $80 \text{ m} \times 80 \text{ m}$  aineistolla, kun kuusen tilavuuden arvoa muutellaan ja muiden muuttujien arvot ovat keskiarvossaan. X-akselilla kuusitukin tilavuus kuutiometreinä hehtaarilla.



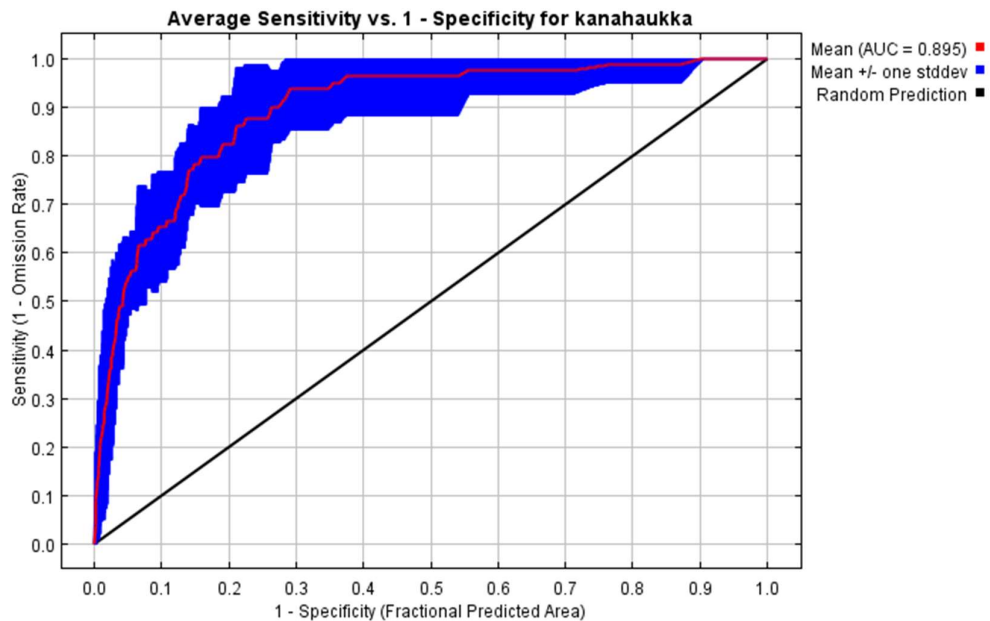
Kuva 6. Vastekuvaajat solukoolla  $80 \text{ m} \times 80 \text{ m}$ , kun kuvatus muuttujan arvoa muutetaan muiden muuttujien pysyessä keskiarvossaan. Ylhäällä vasemmalta koivun tilavuus ja koivun kuitupuun tilavuus, alhaalla vasemmalta kuusen hukkapuosuuden tilavuus ja koko puuston latvuspeittävyys.



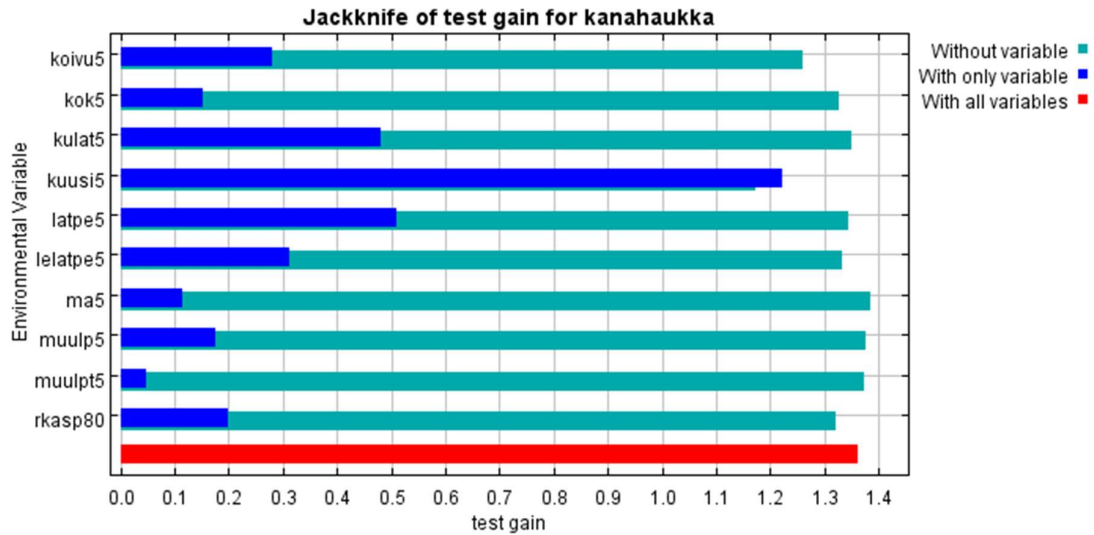
Kuva 7. Vastekuvaajat solukoolla  $80\text{ m} \times 80\text{ m}$ , kun kuvatus muuttujan arvoa muutetaan muiden muuttujien pysyessä keskiarvossaan. Ylhäällä vasemmalta lehtipuuston latvuspeittävyys ja männyn tilavuus, alhaalla vasemmalta muun lehtipuun tilavuus ja muun lehtipuun tukkipuun tilavuus.



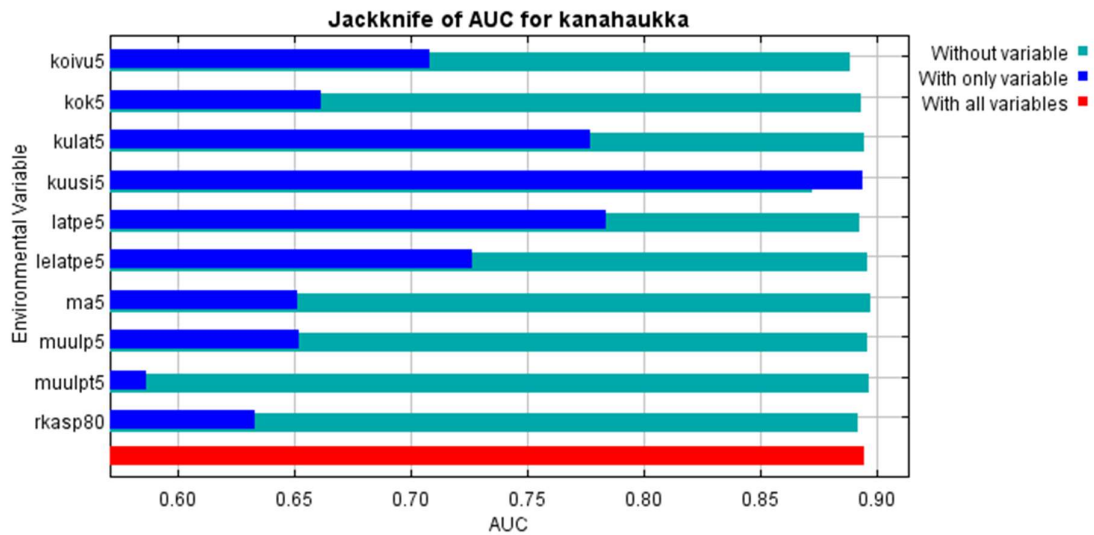
Kuva 8. Vastekuvaaja kasvupaikkatyyppi-muuttujalle solukoolla  $80\text{ m} \times 80\text{ m}$ , kun kuvatun muuttujan arvoa muutetaan muiden muuttujien arvojen pysyessä keskiarvossaan.



Kuva 9. ROC-käyrä solukoon 5 eli  $80\text{ m} \times 80\text{ m}$  kuusen tilavuuden sisältävälle aineistolle. Selite kuten kohdan 4.1 kuvassa 6.

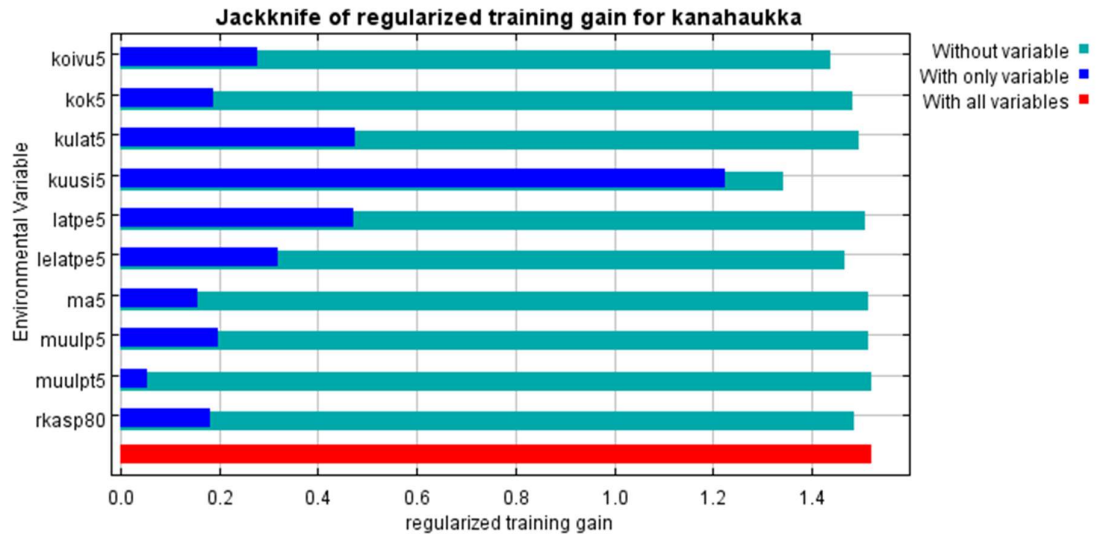


Kuva 10. MaxEntin-mallin muuttujien merkitsevyysarvioinnin tulokset 160 m × 160 m solukoolla kuusen tilavuuden sisältävällä mallilla. Selite kuten kohdan 4.2 kuvassa 9. Muuttujien nimet on esitetty osion 4.2 taulukossa 3.

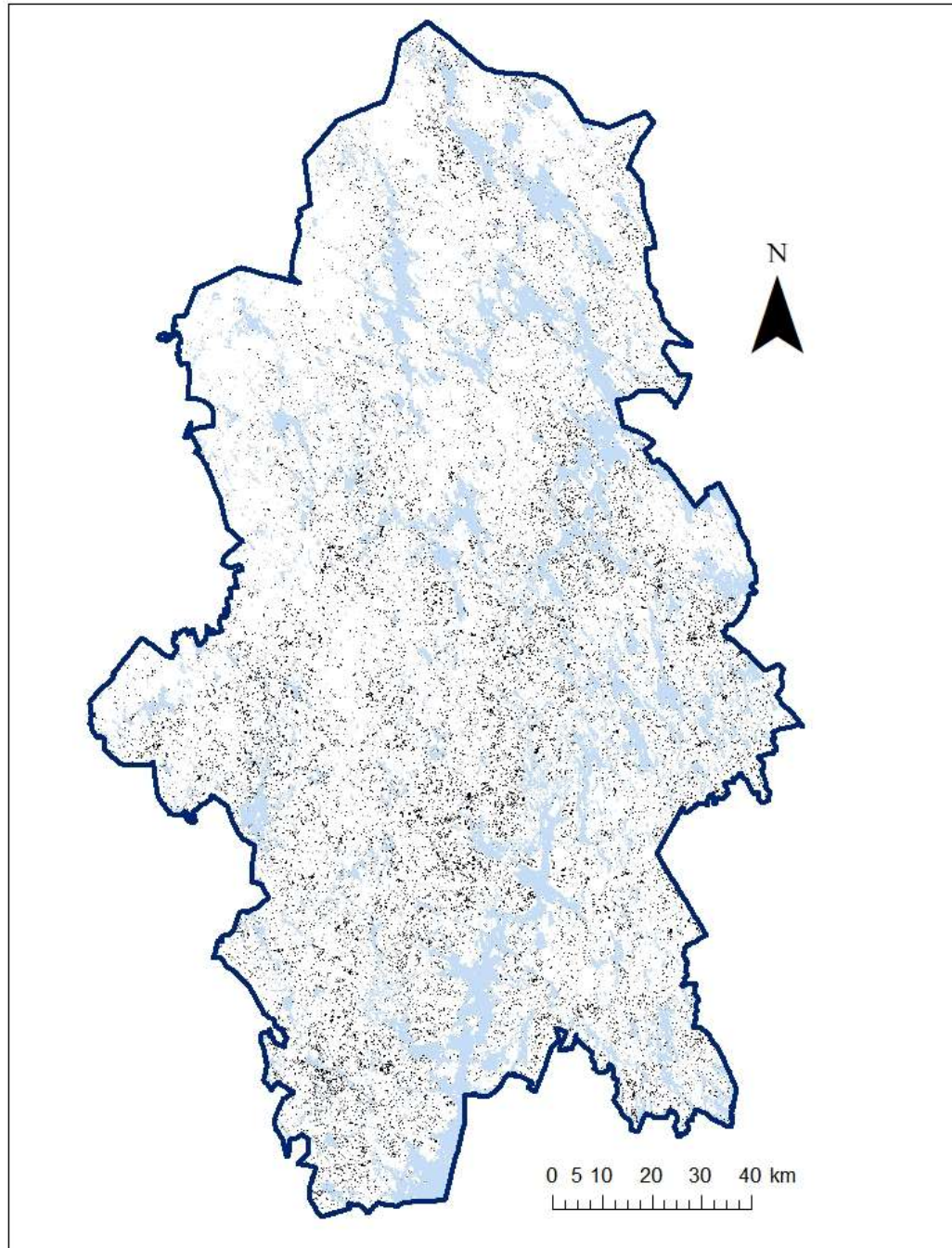


Kuva 11. MaxEntin-mallin muuttujien merkitsevyysarvioinnin tulokset AUC:lle 80 m × 80 m solukoolla kuusen tilavuuden sisältävällä mallilla. Selite kuten kohdan 4.2 kuvassa 9. Muuttujien nimet on esitetty osion 4.2 taulukossa 3.





Kuva 12. MaxEntin-mallin muuttujien merkitsevyyssarviointin tulokset AUC:lle 80 m × 80 m solukoolla kuusen tilavuuden sisältävällä mallilla. Selite kuten kohdan 4.2 kuvassa 9. Muuttujien nimet on esitetty osion 4.2 taulukossa 3.



Kuva 13. Alueiden soveltuminen kanahaukan pesäpaikaksi Keski-Suomessa esitettyinä 80 m × 80 m solukoolla. Musta = suuri todennäköisyys (0,69-1,0), harmaa = tavanomainen kanahaukkametsä (yli 0,46-0,69), valkoinen = pieni todennäköisyys (0,0-0,46), vaaleansininen = ei MVMI-tietoa (vesistöt), tummansininen = Keski-Suomen maakuntarajat (poimittu hallintoraja-aineistosta, © Maanmittauslaitos 2017). Tässä mallissa on käytetty kuusen tilavuutta.

## LIITE 6. MEAN- JA MEDIAN- JAOTTELU MVMI-MUUTTUJILLE

Liitteen 1 kohdan 9 työvaiheessa tehty jaottelu.

MEAN:

- Biomassa, kuusi, elävät oksat 2015 (10 kg/ha)
- Biomassa, kuusi, hukkapuuosa 2015 (10 kg/ha)
- Biomassa, kuusi, juuret,  $d > 1$  cm 2015 (10 kg/ha)
- Biomassa, kuusi, kanto 2015 (10 kg/ha)
- Biomassa, kuusi, kuolleet oksat 2015 (10 kg/ha)
- Biomassa, kuusi, kuorellinen runkopuu 2015 (10 kg/ha)
- Biomassa, kuusi, neulaset 2015 (10 kg/ha)
- Biomassa, lehtipuut, elävät oksat 2015 (10 kg/ha)
- Biomassa, lehtipuut, hukkapuuosa 2015 (10 kg/ha)
- Biomassa, lehtipuut, juuret,  $d > 1$  cm 2015 (10 kg/ha)
- Biomassa, lehtipuut, kanto 2015 (10 kg/ha)
- Biomassa, lehtipuut, kuolleet oksat 2015 (10 kg/ha)
- Biomassa, lehtipuut, kuorellinen runkopuu 2015 (10 kg/ha)
- Biomassa, lehtipuut, lehvästö 2015 (10 kg/ha)
- Biomassa, mänty, elävät oksat 2015 (10 kg/ha)
- Biomassa, mänty, hukkapuuosa 2015 (10 kg/ha)
- Biomassa, mänty, juuret,  $d > 1$  cm 2015 (10 kg/ha)
- Biomassa, mänty, kanto 2015 (10 kg/ha)
- Biomassa, mänty, kuolleet oksat 2015 (10 kg/ha)
- Biomassa, mänty, kuorellinen runkopuu 2015 (10 kg/ha)
- Biomassa, mänty, neulaset 2015 (10 kg/ha)
- Puuston ikä 2015 (vuosi)
- Puuston keskiläpimitta 2015 (cm)
- Puuston keskipituus 2015 (dm)
- Puuston latvuspeittävyys, koko puusto 2015 (%)
- Puuston latvuspeittävyys, lehtipuut 2015 (%)
- Puuston pohjapinta-ala 2015 (m<sup>2</sup>/ha)
- Tilavuus, koivu 2015 (m<sup>3</sup>/ha)
- Tilavuus, koivu kuitupuu 2015 (m<sup>3</sup>/ha)
- Tilavuus, koivu tukkipuu 2015 (m<sup>3</sup>/ha)
- Tilavuus, kuusi 2015 (m<sup>3</sup>/ha)
- Tilavuus, kuusi kuitupuu 2015 (m<sup>3</sup>/ha)
- Tilavuus, kuusi tukkipuu 2015 (m<sup>3</sup>/ha)
- Tilavuus, muu lehtipuu 2015 (m<sup>3</sup>/ha)
- Tilavuus, muu lehtipuu kuitupuu 2015 (m<sup>3</sup>/ha)
- Tilavuus, muu lehtipuu tukkipuu 2015 (m<sup>3</sup>/ha)
- Tilavuus, mänty 2015 (m<sup>3</sup>/ha)
- Tilavuus, mänty kuitupuu 2015 (m<sup>3</sup>/ha)
- Tilavuus, mänty tukkipuu 2015 (m<sup>3</sup>/ha)
- Tilavuus, puusto yhteensä 2015 (m<sup>3</sup>/ha)

MEDIAN:

- Kasvupaikan päätyyppi 2015 (1-4)
- Kasvupaikka 2015 (1-10)
- Maaluokka 2015 (1-3)
- Maaluokka FAO:n FRA-määritelmän mukaan 2015 (1-4)