

Osa 1: Maakotkien esiintymisen painopistealueiden tunnistaminen Keski-Suomessa



Kirjoittaja: Pinja Pelkonen, Jyväskylän Yliopisto

Työn ohjaajat: Anssi Lensu ja Elisa Vallius, Jyväskylän Yliopisto

Projektin työryhmässä toimivat asiantuntijoina:

Markus Erkkilä, Keski-Suomen liitto

Reima Väливаara, Keski-Suomen liitto

Kari Lehtinen, Keski-Suomen ELY-keskus

Johanna Viljanen, Keski-Suomen ELY-keskus

Tomi Hakkari, Keski-Suomen ELY-keskus

Hannu Tikkanen, Metsähallitus

Ulla-Maija Humppi, Saarijärven kaupunki

Ossi Nokelainen, Keski-Suomen Lintutieteellinen Yhdistys

Kristiina Nyholm, Keski-Suomen Lintutieteellinen Yhdistys

Jouni Lamminmäki, Suomenselän Lintutieteellinen Yhdistys

Sisällysluettelo

1. Johdanto	1
2. Mallinnuksessa käytetyt ympäristö muuttuja aineistot	2
2.1 Pesäpaikka aineistot.....	2
2.2 Valtakunnan metsien inventoinnin monilähdetietoaineisto (MVMI)	2
2.3 Metsänkäyttöilmoitukset.....	2
2.4 Korkeusmalli ja maastotietokanta.....	3
2.5 Metsänmonimuotoisuus ja kasvillisuus aineistot.....	3
3. Aineiston käsittelyyn käytetyt ohjelmistot ja menetelmät	4
4. Maksimientropia mallinnus	5
4.1 MaxEnt mallinnukseen valitut muuttujat ja asetukset.....	5
5. Mallien vertailu	7
6. Parhaan mallin tulokset	9
7. Yhteenveto	17
LÄHTEET	18
LIITE 1	19
LIITE 2	24
LIITE 3	27
LIITE 4	29

1. Johdanto

Maakotka (*Aquila chrysaetos*) on Suomessa suojeltu ja vaarantunut suurikokoinen petolintu. Suomen kannan on arvioitu olevan noin 350–470 paria (Lintuatlas, 2010; Luontoportti, 2019), mutta suurin osa kannasta keskittyy Pohjois-Suomeen. Koska laji on ns. harvinainen, on tärkeää tutkia ja suojella maakotkakantaa myös Keski-Suomessa. Lajia pidetään myös hyvänä luonnonmukaisen ekosysteemin tilan ja ympäristönlaadun indikaattorina, koska se on huippusaalistaja ravintoketjussa (Stefănescu and Bălescu, 2019).

Koska maakotkat vaativat suuren ja monimuotoisen reviirin metsästyksen ja pesintään, ja laji kärsii usein tilan puutteesta. Projektin tavoitteena oli mallintaa maakotkien potentiaalista elinaluetta maksimientropiamallinnuksella (Osa 1). Tavoitteenamme oli myös saada lisätietoja maakotkan elinympäristön tärkeimmistä kriteereistä ja tutkia potentiaalisia alueita, jotka voisivat sopia lajille Keski-Suomessa.

Uusiutuvien energianlähteiden kasvavan tarpeen takia tuulivoimapuistot ovat yhä suosittumia ja tarvitsevat tilaa kaukana asutuksesta. On siis hyvä tietää mahdolliset maakotkan elinalueet, jotta rakennusprojektit näillä alueilla voidaan välttää tai harkita tarkemmin maakotkien kannalta, sillä lajin elinalueet ja tuulivoimala hankkeet törmäävät yhteen harmillisen usein samankaltaisten ympäristökriteerien takia. Tutkimme maakotkien potentiaalisia elinalueita Keski-Suomessa, jotta tulevaisuudessa mahdollisia yhteentörmäyksiä lajin ja tuulivoima projektien välillä voidaan välttää. Mallituksen tulosta voitaisiin myös hyödyntää esimerkiksi tekopesien pystytyksessä tai muussa suojelu toiminnassa.

2. Mallinnuksessa käytetyt ympäristömuuttuja aineistot

2.1 Pesäpaikka-aineistot

Maakotkan pesäpaikka-aineistot saatiin Keski-Suomen ELY-keskukselta tutkimusta varten. Tutkimuksen kohdealueena oli Keski-Suomen maakunta, mutta mallintamiseen hyödynnettiin myös maakunnan rajojen ulkopuolella sijaitsevia pesimistietoja, jotta mallin ennustavuus ja sijaintitietojen määrä olisi suurempi. Tutkielmassa päätettiin käyttää vain pesäpaikkoja, jotka olivat olleet maakotkan käytössä viimeisen 10 vuoden aikana. Historiallisia pesiä ei käytetty, jotta tutkimustulos olisi mahdollisimman ajankohtainen. Yhteensä pesätietoja oli 41 eri pesäpaikalta, joista osa sijaitsi Keski-Suomen ulkopuolella (tekopesiä ei otettu huomioon projektin 1 osassa). Käytetty aineisto sisälsi tietoja pesäpaikan koordinaateista ja vuosittaisesta seurannasta.

Spatiaalisen autokorrelaation välttämiseksi MaxEnt malli tehtiin 40 pesäpaikalla (kaikki) ja 21 pesäpaikalla (rajoitettu). Rajoitetussa aineistossa 5 kilometrin säteellä olevia vierekkäisiä pesäpaikkoja poistettiin satunnaisesti spatiaalisen autokorrelaation ja valetiston välttämiseksi. Rajoitetut pesäpaikat tarkistettiin myös poiston jälkeen, jotta suosituimmat ja ympäristömuuttujien aikajanaa muistuttavat pesät olisivat mukana mallinnuksessa.

2.2 Valtakunnan metsien inventoinnin monilähdetietoaineisto (MVMI)

Puustoon liittyvät ympäristömuuttuja-aineistot haettiin Luken avoimien aineistojen tiedostopalvelusta. MVMI-aineisto on koottu erilaisista lähteistä mm. satelliittikuvista (Landsat kuvista), korkeusmalleista ja maastokartoista. Kaikki aineistot ovat jo ETRS89-TM35FIN koordinaatistossa, ja ne ladattiin sivustolta GEOTIFF muodossa. Aineiston solukoko on 16 m × 16 m.

Tutkielmassa valittiin hyödyntää uusinta 2019 julkaistua aineistoa, joka pohjautuu 2013-2017 koealatietoon (11 ja 12 valtakunnan metsien inventointi) ja 2017-2018 satelliittikuviin. Tiedoston teemoista (45) päätettiin hyödyntää kuutta: männyn tilavuutta (kannon yläpuolelta latvaan), puuston ikää (painotettu keski-ikä koealalta), keskipituutta (pohjan pinta-alaa vastaavan pituuden mediaani), kasvupaikkaa (luokat 1-7: 1=lehto, 2=lehtomainen kangas, 3=tuore kangas, 4=kuivahko kangas, 5=kuiva kangas, 6=karukkokangas, 7=kalliomaat, hietikot, kitu- tai joutomaat), latvuspeittävyyttä (kuvaa latvuston osuutta % koealan alasta) ja männyn hukkapuuosan tilavuutta (kuvaa määrän tai laadun vuoksi kelpaamatonta puun rungon kuorellisen osan biomassaa). Nämä tasot valittiin, koska niiden ajateltiin olevan maakotkan pesänvalintaan, joko erittäin hyödyllisiä tai haitallisia selittäjiä.

2.3 Metsänkäyttöilmoitukset

Jotta MVMI-aineisto olisi ajantasainen käytettiin metsänkäyttöilmoituksia hyödyksi. Metsänomistajan on ilmoitettava 11 pv ennen hakkuutoimia metsäkeskukselle aikomuksistaan ja toteutettava tämä 3 vuoden kuluttua ilmoituksesta. Tämän takia

käytettiin 2017-2019 vuoden hakkuutoimia päivittämään metsän kuntoa 2020 vuoden tasolle, muuntamalla nämä alueet 0-arvolle kaikissa paitsi kasvupaikka aineistossa. Hakkuista valittiin käytettäväksi yleisimmät luokat: avohakkuu, ensiharvennus ja harvennushakkuu.

2.4 Korkeusmalli ja maastotietokanta

Maaston korkeus ja kaltevuus ovat mainittu tärkeinä maakotkan elinalueiden ympäristömuuttujina monissa eri tutkimuksissa (Stefănescu and Bălescu, 2019). Maanmittauslaitoksen avoimien aineistojen tiedostopalvelusta päätettiin käyttää 10 m korkeusmallia, josta voitiin laskea ArcMap -ohjelmiston avulla maaston kaltevuus ja rinteensuunta. Aineisto kattoi koko tutkimusalueen ja oli saatavana 10 m × 10 m resoluutiossa (tarkkuus 1,4 m). Aineisto oli myös valmiiksi ETRS89-TM35FIN koordinaatistossa.

Maanmittauslaitoksen maastotietokantaa hyödynnettiin myös tutkielmassa. Tämä on tarkin vektoraineisto Suomen maanpiirteistä (tarkkuus 1:5000-1:10000). Aineistossa on 18 eri kohdeluokkaa, joihin kuuluvat mm. vesistöt ja asutus. Tutkielmassa hyödynnettiin rakennusten sijaintia ja luokkia, josta voitiin laskea tiettyjen rakennusten etäisyys ja tiheyslaskelmia. Maastotietokanta aineistosta hyödynnettiin myös tiestön viiva -aineistoa, joista laskettiin alle 5 m levyisten ja yli 5 m levyisten teiden etäisyysrasteritasot.

2.5 Metsänmonimuotoisuus ja kasvillisuus aineistot

SYKE latauspalvelu LAPIOn aineistoja käytettiin myös täydennyksenä mallintamisessa. Lapiro palvelusta ladattiin maanpeitteestä kertovat Corine 2018 ja Zonation 2018 -tasot. Corine tuotetaan SYKE:ssä paikkatietoaineistojen ja satelliitti kuvien perusteella. Aineistossa on eri luokittelutasoja, joiden perusteella voidaan maanpeite luokitella tarkemmin (kolmas luokittelutaso: 44 luokkaa). Corine 2018 -tasoa hyödynnettiin avointen alueiden tunnistamiseen, koska nämä ovat tärkeitä saalistusalueita maakotkalle. Aineisto on saatavissa rasteri tai vektori muodossa (rasterisolukoko 20 m × 20 m).

Zonation 2018 -aineisto on metsä analyysin tulokartta, joka kuvaa metsän monimuotoisuutta 0-1 arvoilla. Karttatasosta on 12 eri versiota, tähän tutkimukseen päätettiin käyttää Keski-Suomen ELY-keskuksen tulokartta versiota 6 (solukoko 96 m × 96 m). Versio 6 luokittelu sisältää mm. alueen lahoppuupotentiaalin ja punaisen listan metsälaji havainnot.

SYKEN sivuilta päätettiin myös ladata Ranta10 -aineisto. Tästä kehitettiin myöhemmin järvien tiheystaso. Tämä aineisto pohjautuu Maanmittauslaitoksen rantaviiva-aineistoon, joka on osa maastotietokantaa. Aineisto oli valmiiksi koottuna vektorimuodossa (1:5 000-1:10 000).

3. Aineiston käsittely ohjelmistot ja menetelmät

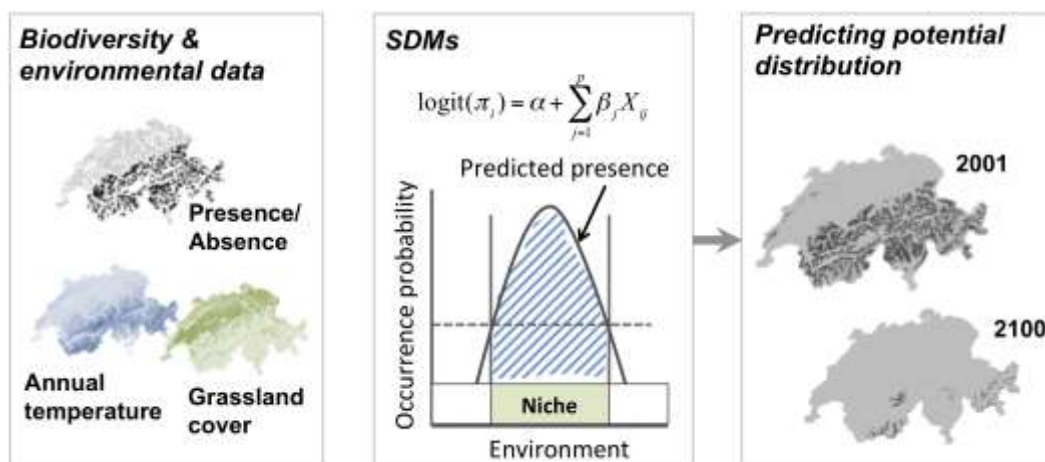
Tutkielmassa aineistojen käsittelyyn käytettiin ArcGIS ArcMap v. 10.6.1 -ohjelmistoa (ESRI, Redlands, CA, USA), jolla aineistoja rajattiin samaan kokoon ja muotoon. Myös osaa aineistoista käytettiin laskupohjana uusille tasoille.

VMI-aineistoja päivitettiin MKI-aineistoilla ja Aggregate -työkalua käytettiin aineistojen rasteri solukokojen muutokseen, eli aineiston yleistämiseen. Tämä tehtiin, jotta kaikki aineistot olisivat joko 160 m × 160 m tai 80 m × 80 m solukoon mukaisia. Korkeus-aineistosta laskettiin Slope -työkalulla maaston kaltevuus ja Aspect -työkalulla rinteensuunta. Maastotietokannasta käytettiin eri rakennusluokkia, joille laskettiin Kernel Density -työkalulla tiheys tasoja ja Euclidean Distance -työkalulla etäisyys tasoja. Zonation 2018 -aineistoa yleistettiin myös Resample -työkalulla, ja Corine 2018 -aineistosta eriteltiin sopivat maaluokat ja näille laskettiin myös tiheystasoja.

Kaikki tasot tallennettiin Raster to ASCII -työkalulla oikeaan muotoon. Näin aineistoa muokattiin sopivaksi maksimientropiamallinnukseen MaxEnt 3.4.0 -ohjelmistolla. Aineistoja käsiteltiin huomattavasti ennen mallintamisvaihetta, ja tarkat ohjeet eri käsittelyvaiheista löytyvät liitteestä 1. ArcMap -ohjelmistoa käytettiin myös korreloinnin ja pääkomponenttianalyysin ajamiseen, sekä MaxEnt mallin tulokarttojen tekoon.

4. Maksimientropia mallinnus

Maksimientropiamalli perustuu eniten hajautuneen, eli vähiten oletuksia todellisesta jakaumasta sisältävän mallin valintaan. Näin voidaan välttää datasta tehtyjä oletuksia, koska mallin ennustus perustuu mahdollisimman vähän puuttuviin tietoihin (Jaynes 1957; Leonard 2017). MaxEnt-mallin tavoite on siis arvioida lajin esiintymispaikkatietojen ja annettujen taustatietojen (ympäristömuuttujien) yhdenmukaisin jakautuminen tutkimusalueella (Ștefănescu ja Bălescu, 2019). Malli arvioi lajin esiintymispaikkoja muuttujien arvojen perusteella, ja kysyy, onko kohdealue sopiva kohdelajille (kuva 7). Tämä perustuu siihen, että soveltuvassa elinympäristössä lajin esiintymistodennäköisyys kasvaa (Philips ym. 2004).



Kuva 1. Species distribution modelling (SDM) eli lajin levinneisyyden mallinnus on alueen soveltuvuuden todennäköisyyden laskentaa laji havainnoilla ja ympäristömuuttujilla. SDM mallinnusta voidaan myös käyttää ennustamaan lajin esiintymisen kehittymistä alueella, tai toisen alueen esiintymistä. Kuva lähteestä Zurell, 2020.

Mallin tuloksena saadaan myös rasterikartta maksimientropian esiintymistodennäköisyyksistä, jotka summautuvat lukuun 1. Tulokset voidaan havainnollistaa cloglog muunnoksella MaxEnt -ohjelmistolla (Gibbsin todennäköisyysjakauma, asteikolla 0-1). Tämä käänös havainnollistaa mallinnustulosta kääntämällä sen ennustettuun todennäköisyyteen alueen soveltuvuudesta lajille.

4.1 MaxEnt mallinnukseen valitut muuttujat ja asetukset

Mallin oletusasetus käyttää 10 000 satunnaispistettä mallin luontiin. Koska tämän tutkielman kohdealue oli suuri, käytettiin mallintamisessa 20 000 satunnaispistettä. MaxEnt-mallinnus tehtiin useita kertoja, jotta voitiin valita parhaat selittävät muuttujat (eri mallinnusvaiheet ja tasot liitteessä 2). Mallin testausaineiston voi erotella monella tavalla, mutta mallinnuksessa päätettiin käyttää 5-kertaista ristiinvalidointia, jolloin malli ajetaan 5 kertaa, ja joka kerta 1/5 pesäaineistosta käytetään testausaineistona. Näin pesäpaikkoja ei

jouduta jakamaan kahteen osaan (mallin teko ja testaus aineistoiksi), vaan jokaista pistettä voidaan käyttää sekä testaukseen että mallinnukseen.

Lopuksi malleja vertailtiin toisiinsa AUC-arvoilla (Area under ROC-curve), TSS-arvoilla (True skill statistic) ja AICc-arvoilla (Akaiken informaatiokriteeri). MaxEnt 3.4.0 -ohjelma laskee ajojen AUC-keskiarvon automaattisesti, mutta tämän lisäksi päätettiin TSS-arvot laskea R-ohjelmistolla ja AICc-arvot Glover-Kapfer (2015) ohjeiden mukaan. Tämä tehtiin koska AUC-arvojen vaihteluväli oli hyvin kapea (tarkat tiedot arvoista ja laskukaavoista liitteessä 3). Mallien vertailua tehtiin myös Hannu Tikkasen tekemään maakotkamalliin, joka oli rakennettu BRT (boosted regression trees) menetelmällä, ja jonka kohdealueena oli koko Suomi.

5. Mallien vertailu

Kun malleja vertailtiin, osoittautuivat kaikki mallit AUC-arvoltaan hyviksi. Vertailun perusteella 1. mallintamiskierroksen mallit olivat heikoimpia ennustavuudeltaan. Mutta 2. ja 3. mallintamiskierroksen mallit olivat AUC-arvoiltaan hyvin samankaltaisia. Paras malli AUC-arvon perusteella olisi versio 3 80 m × 80 m malli, jossa käytettiin regularization multiplier arvoa 2,5 (oletus asetus=1 sijaan).

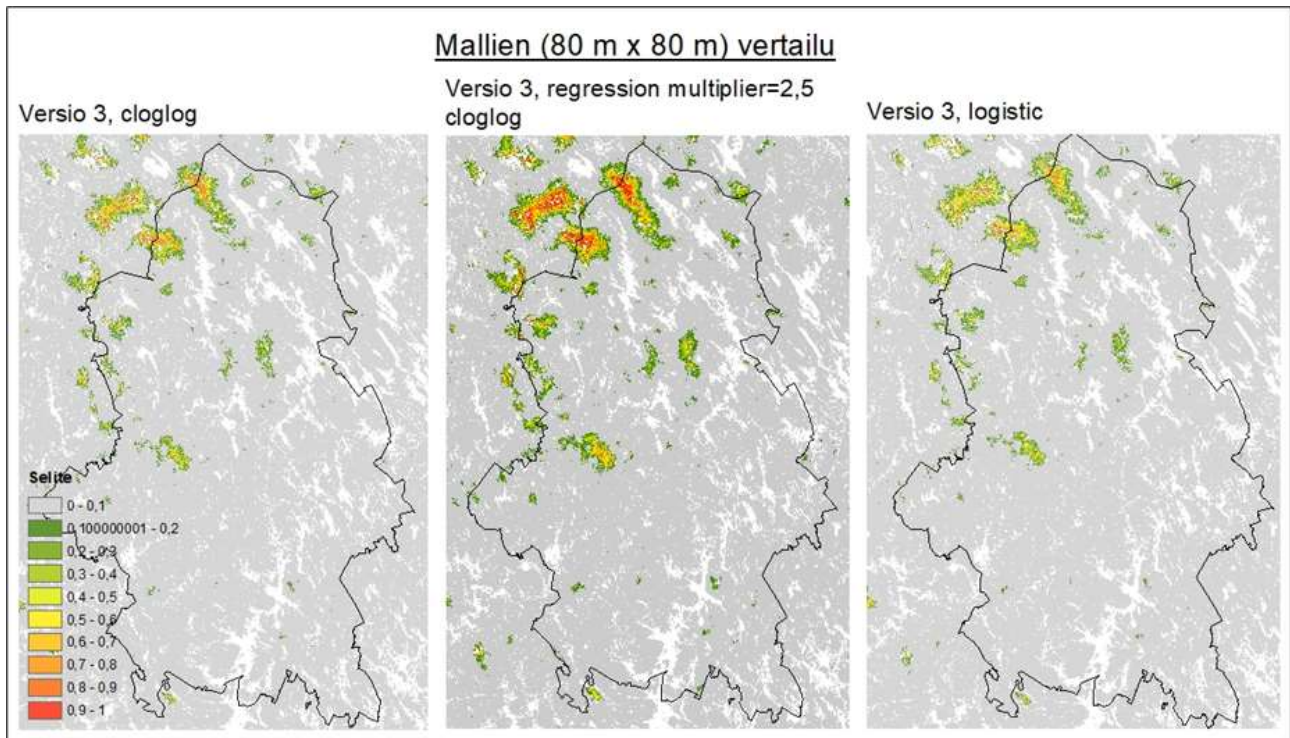
Koska AUC-arvoissa oli vain vähän eroavaisuuksia, päätettiin myös laskea TSS-arvojen ja AICc-arvojen vertailu. Myös kaikki toisen ja kolmannen mallintamiskierroksen mallien TSS-arvot olivat korkeita osoittaen hyvää ennustavuutta (>0,8). TSS-arvon perusteella parhaat mallit olisivat versio 2: 80 m × 80 m mallista ja versio 3: 160 m × 160 m mallista. Huomattavaa oli, että regularization multiplier:in nosto 2,5 ei lisännyt mallien TSS-arvoa.

AICc (Akaiken informaatiokriteeri) arvot eivät myöskään noudattaneet täysin AUC-arvojen järjestystä parhaista malleista. AICc-arvoissa pienin luku katsotaan paremmaksi, tässä malli on linjassa aineiston kanssa parhaiten. Vaikka malliversio 3: 80 m × 80 m ei ollut paras, oli sen AICc-arvo kuitenkin suhteellisen pieni. Kun otetaan huomioon suurempi pesämäärä, jolla malli rakennettiin. Nämä tiedot huomioon ottaen parhaaksi malliksi valikoitui malliversio 3 80 m × 80 m, joka oli laskettu kaikilla pesäpaikoilla.

Taulukko 1. Mallien vertailuun käytetyt arvot.

Malli	Ajojen testiaineistojen AUC-keskiarvo	Ajojen testiaineistojen TSS-keskiarvo	Ajojen AICc keskiarvo
v.1 80 x 80, k. pesät	0,963	0,832	409,2
v.1 80 x 80 r. pesät	0,943	0,667	238,5
v.1 160 x 160 k. pesät	0,968	0,862	388,9
v.1 160 x 160 r. pesät	0,943	0,724	242,1
v.2 80 x 80	0,973	0,880	387,5
v.2 160 x 160	0,971	0,860	395,3
v.3 80 x 80	0,977	0,832	330,4
v.3 160 x 160	0,975	0,873	358,2
v.3 80 x 80, r.m.=2,5	0,981	0,823	1000,4
v.3 160 x 160, r.m.=2,5	0,978	0,850	521,5
v.3 80 x 80 logistinen	0,977	0,857	330,4
v.3 160 x 160 logistinen	0,975	0,835	357,0

Mallien AUC- ja TSS-arvojen lisäksi tarkastettiin myös mallien karttaennusteet. Regression multiplier:in kasvattaminen näkyi selvästi mallin karttaennusteessa, koska se lisäsi hyvin sopivia alueita. Mutta kaikista ennusteista erottuivat kuitenkin samat alueet, hieman pienemmillä sopivuuksilla (kuva 2).



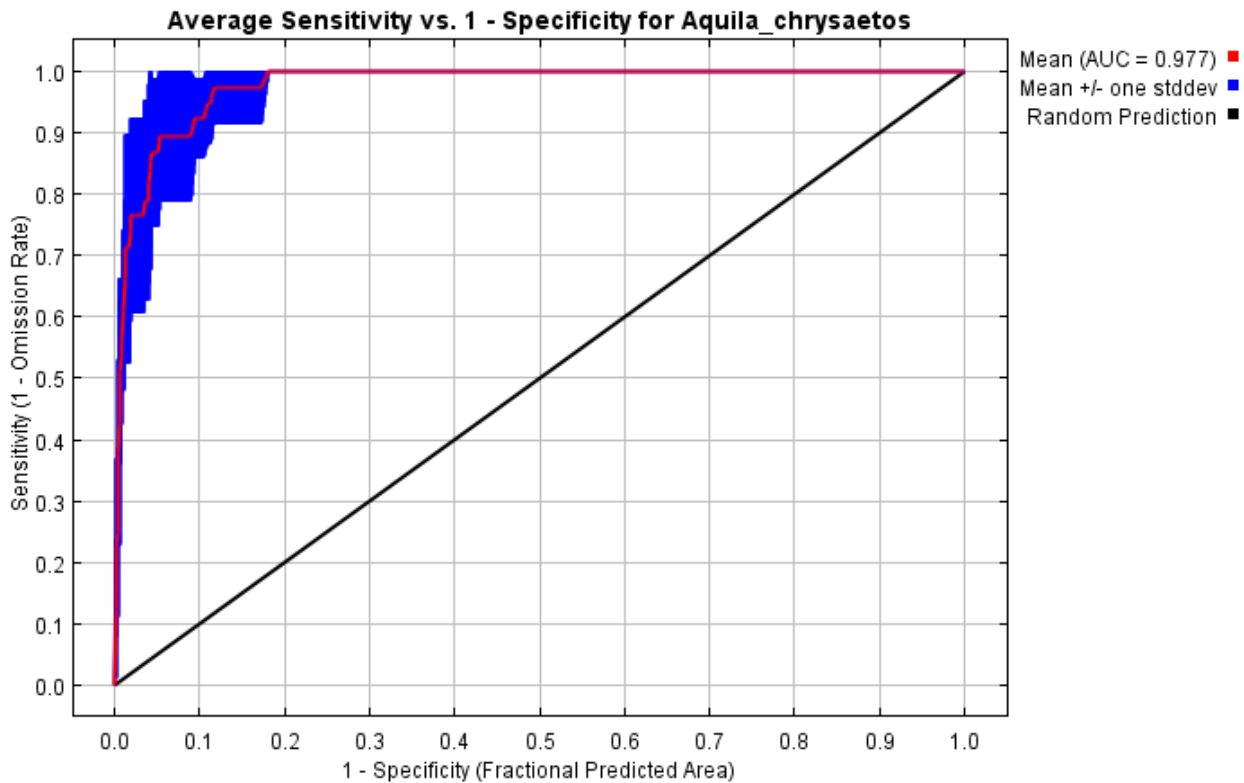
Kuva 2. Karttavertailu 3 mallin ennustamista elinympäristön soveltavuudesta maakotkalle. Punainen indikoi hyvää elinaluetta ja vihreä väri heikkoa soveltuvuutta. Harmaa väri kuvaa hyvin heikkoa tausta aluetta. (Kartta sisältää Maanmittauslaitoksen Kuntajakopohjaiset hallinnolliset aluejaot 10/2020 aineistoa, © Maanmittauslaitos 2021).

Parasta mallia päätettiin myös vertailla toiseen maakotkan elinympäristöä arvioineeseen malliin, joka oli luotu erityyppisellä mallinnustavalla ja aineistoilla. Vertailu toteutettiin Hannu Tikkasen antamaan BRT-malliin, johon oli otettu mukaan koko Suomen maakotka kanta. Mallintamistapojen erot tulivat selkeästi esiin heikosti sopivilla alueilla. Voidaan myös tulkita MaxEnt mallin sisältävän vähemmän kohinaa.

Vertailu osoitti, että MaxEnt menetelmä voi ylisovittaa mallia, jos muuttujia on liikaa. Tuloksena on malli, jossa rakentamiseen annettuja pesäpaikkoja ennustetaan hyvin, mutta tämä ei välttämättä päde uusille pesäpaikoille. MaxEnt metodilla tehty malli siis hyötyy, jos malliin voidaan laittaa suuri määrä lajihavaintoja, tarkkaan valitut ympäristömuuttujat, sekä mallin ylisovittamista voidaan vähentää regression multiplier:in lisäämisellä. Vaikka mallien välisiä eroja oli kohtalaisesti, kartalta erottuivat samat hyvät elinympäristöt ja tämä luo varmuutta mallin ennusteeseen.

6. Parhaan mallin tulokset

MaxEnt mallin v.3 80 m × 80 m ennustavuus oli erinomainen, joka ilmenee mallin ROC-käyrästä (Kuva 3).



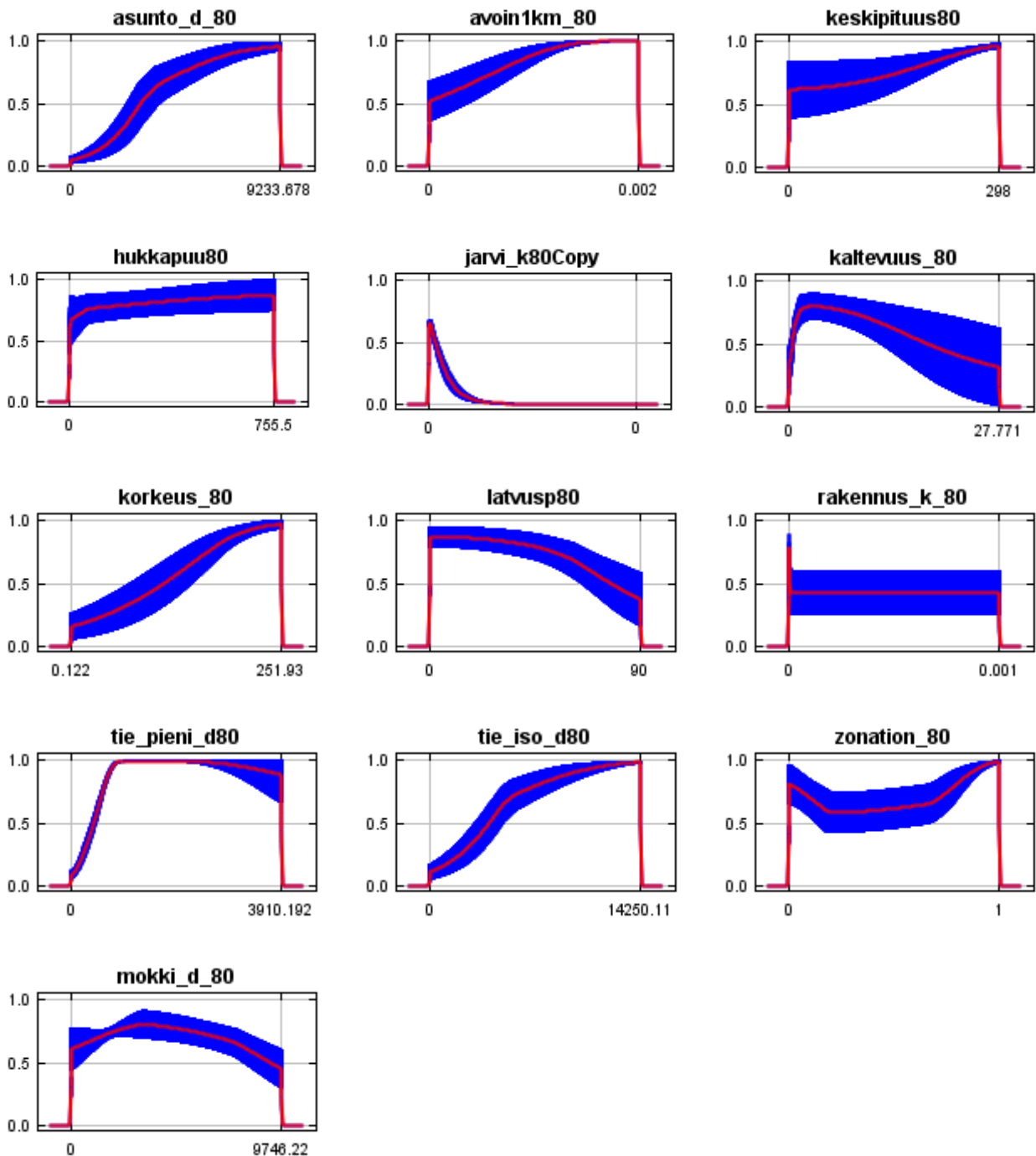
Kuva 3. Mallin v.3 80 m × 80 m, kaikilla pesäpaikoilla ROC-käyrä. Suorakulmainen käyrä kertoo erinomaisesta lajin pesäpaikkojen ennustamisesta. Musta viiva kuvaa satunnaista ennustavuutta ja sininen alue riistiinvalidoinnin ajojen vaihteluväliä.

Mallin tuloksista eriteltiin myös ympäristömuuttujien merkitystä mallin rakentamiseen ja pesäpaikkojen tunnistamiseen. Erityisesti kaikkien rakennusten tiheys, asuinrakennusten etäisyys ja isojen teiden etäisyys olivat mallissa merkittäviä muuttujia. Myös järvien tiheyden korvaaminen satunnaisluvuilla, aiheutti mallin huomattavaa heikkenemistä, vaikka sen osuus mallin rakentamisessa olikin hyvin pieni.

Taulukko 2. Ympäristömuuttujien osuus mallista.

Muuttuja	Muuttujan lyhenne	Osuus mallista (%)	Mallin heikkeneminen (%)
Rakennusten tiheys	rakennus_k_80	32,2	6,6
Asuntojen etäisyys	asunto_d_80	23,7	14,7
Suurteiden etäisyys	tie_iso_d80	13,3	18,4
Zonation 2018	zonation_80	12,4	6,7
Avoimen alueen tiheys	avoin1km_80	10	4,1
Pienteiden etäisyys	tie_pieni_d80	2,5	8
Maaston kaltevuus	kaltevuus_80	1,8	2,6
Maaston korkeus	korkeus_80	1,5	3
Järvien tiheys	jarvi_k80Copy	1	27,7
Mökkien etäisyys	mokki_d_80	0,7	3,4
Puuston latvuspeittävyys	latvusp80	0,4	1,6
Männyn hukkapuuosa	hukkapuu80	0,4	1,3
Puuston keskipituus	keskipituus80	0,1	1,8

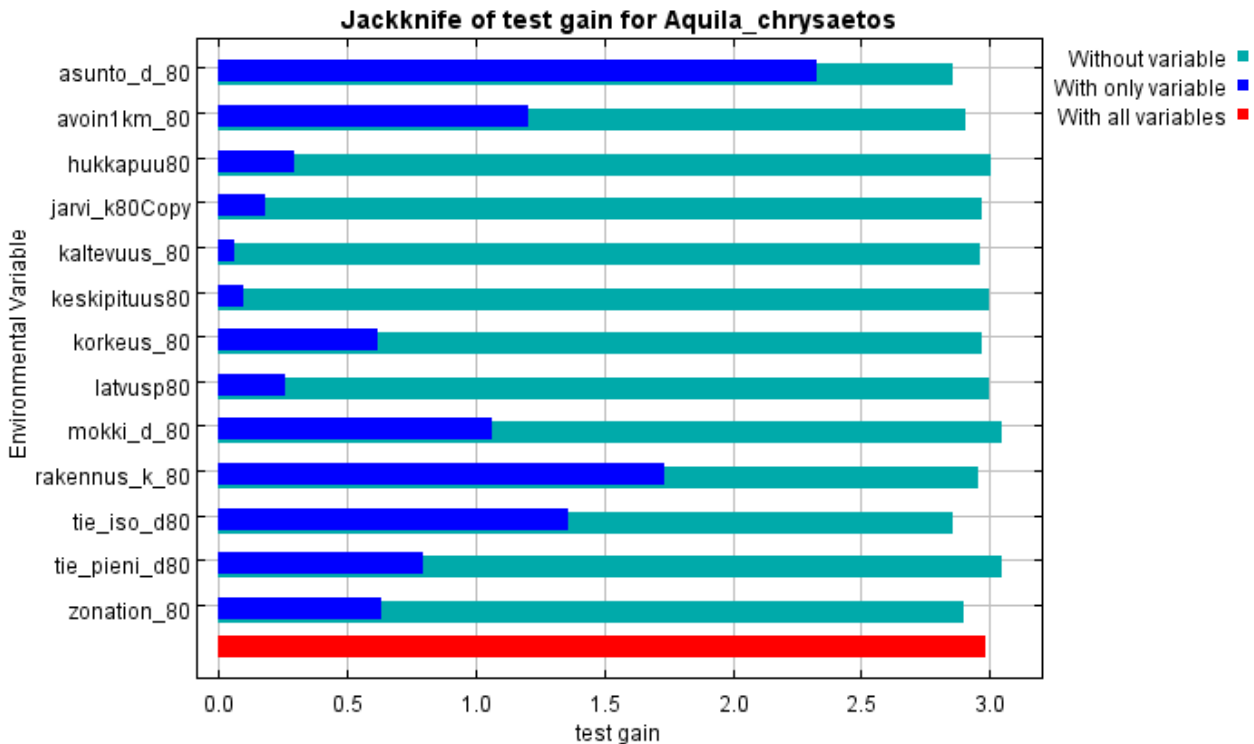
Paremmen kuvan eri muuttujien merkityksestä saa muuttujakohtaisista vastekuvaajista. Näistä voidaan tarkastella muuttujien arvojen vaikutusta ajojen keskiarvolliseen cloglog ennustettuun todennäköisyyteen. Muun muassa maaston korkeuden kasvu lisää pesintäpaikan todennäköisyyttä, kun taas lasvuspeittävyiden kasvu vähentää tätä (kuva 4). Ainoana tasaisena vastekäyränä näkyy olevan rakennusten tiheys, joka vaikuttaa alhaisilla arvoilla voimakkaasti todennäköisyyksien kasvuun, mutta suuremmilla arvoilla on hyvin tasainen n. 0,5 tasolla. Tämän takia rakennettiin myös malli versio ilman rakennustiheyttä, mutta mallin AUC-arvo olisi laskenut ja eroavaisuudet ennusteessa olivat hyvin pieniä, joten päätettiin rakennustiheystaso jättää malliin.



Kuva 4. Mallin ympäristömuuttujien vastekuvaajat. Punainen viiva kuvaa ajojen keskiarvoa ja sininen alue ajojen keskihajontaa. Vastekuujista nähdään muuttujan arvojen (x-akseli) vaikutus ennustettuun cloglog todennäköisyyteen (y-akseli). Vastekäyrien lyhenteet löytyvät taulukosta 2.

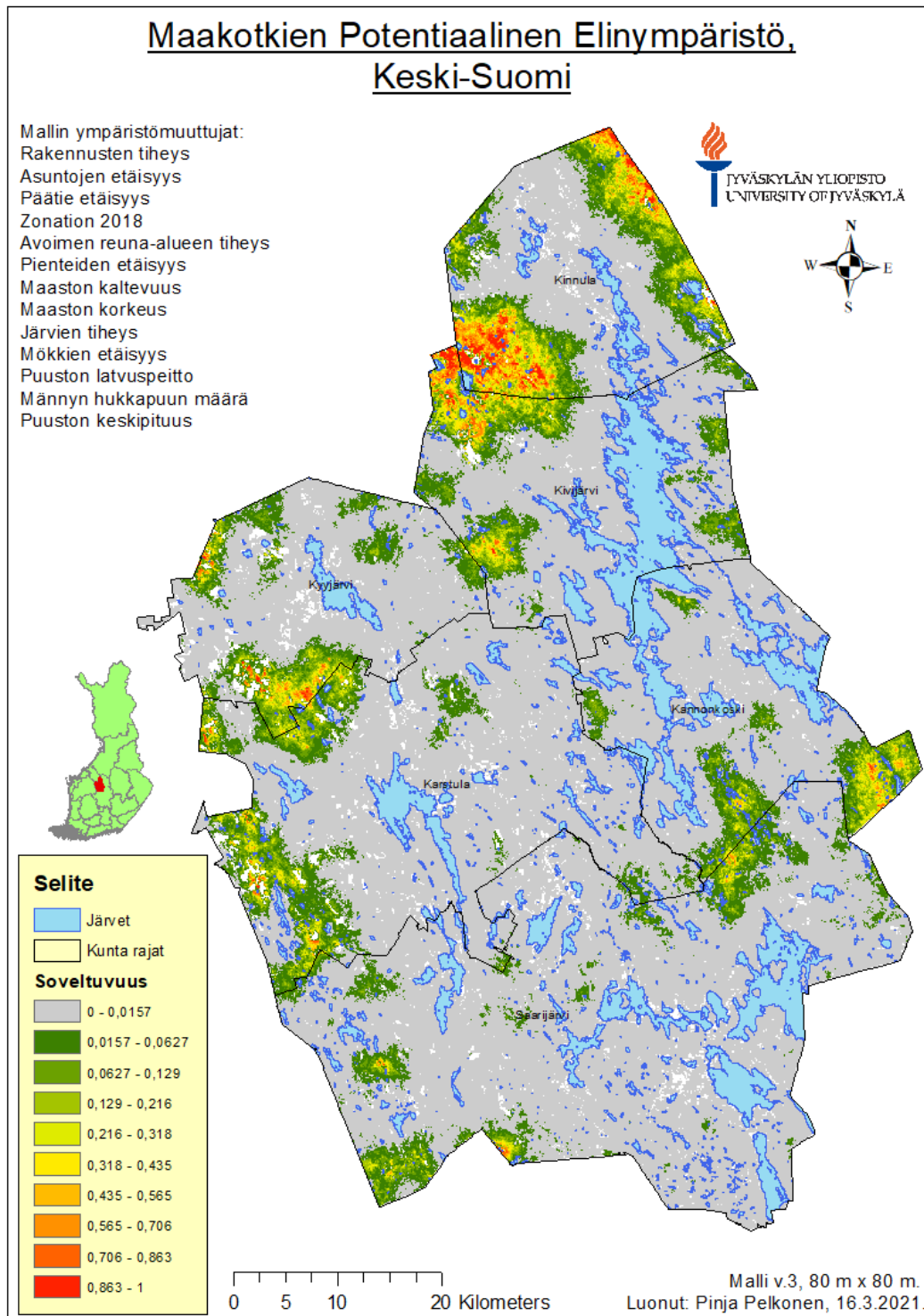
Mallin tuloksista voidaan myös tarkastella muuttujien merkitystä (test gain) jackknife testin kuvaajasta. Tästä nähdään miten muuttujien poisjätö heikentäisi mallin ennustavuutta (turkoosit pylväät), ja mitkä olisivat eniten selittäviä muuttujia yksinään mallissa (siniset pylväät). Koska malli sisältää monia ympäristömuuttujia ei yksittäisen tasojen poisto vaikuttaisi malliin paljon, mutta mitä lyhyempi turkoosi pylväs on sitä tärkeämpi muuttuja

mallissa on. Vähiten yksinään maakotkan pesäpaikkoja selittäisivät mallissa maaston kaltevuus ja puuston keskipituus (kuva 5).

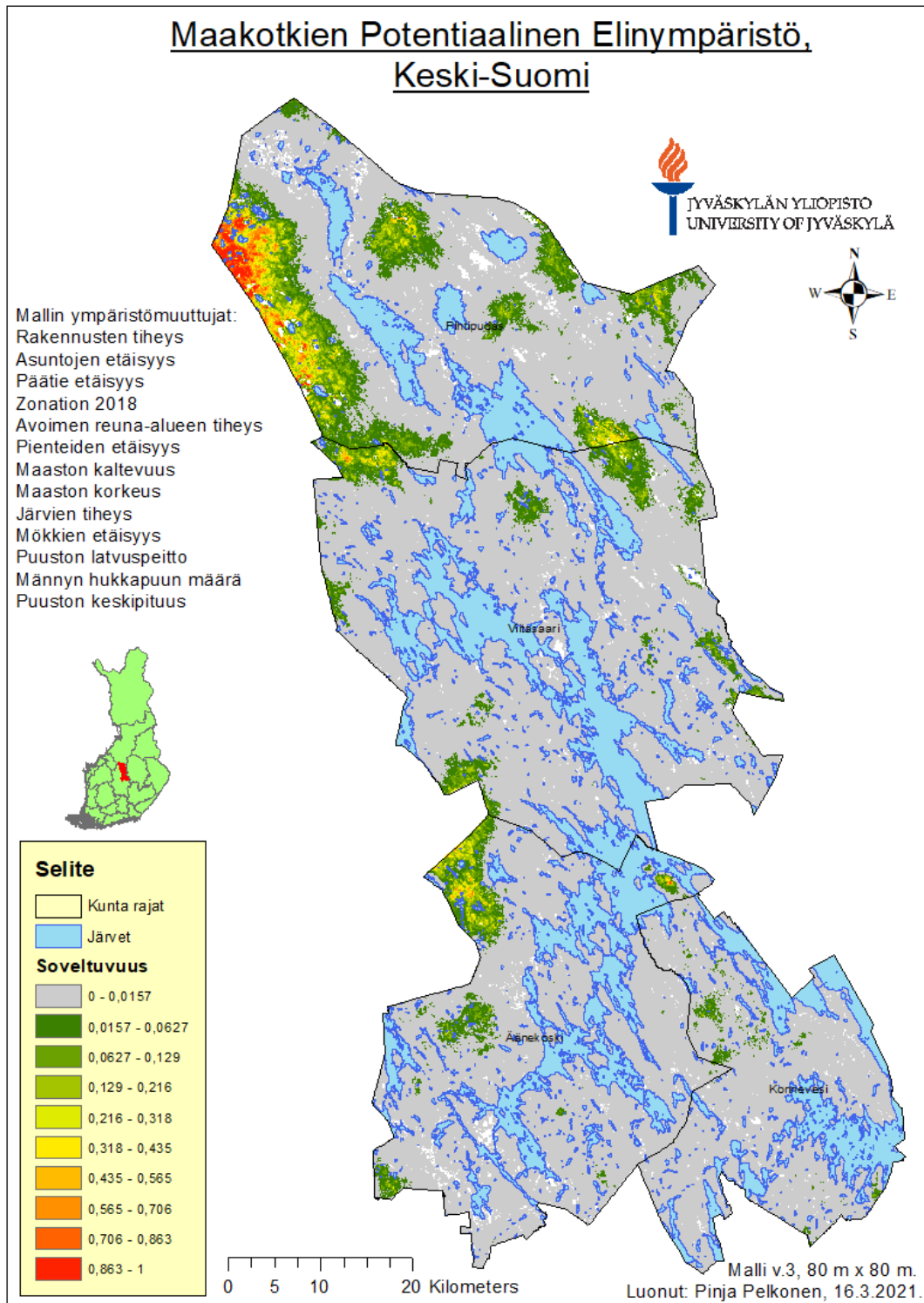


Kuva 5. Ympäristömuuttujien merkitsevyysarviointi parhaassa mallissa. Punainen = mallin kokonaiselittävyys, turkoosi = mallin selittävyys ilman kyseistä muuttujaa, sininen = mallin selittävyys vain kyseisellä muuttujalla. Muuttujien lyhenteiden nimet löytyvät taulukosta 1. Asuntojen etäisyys näkyy selittävän parhaiten maakotkapesä esiintyvyyttä.

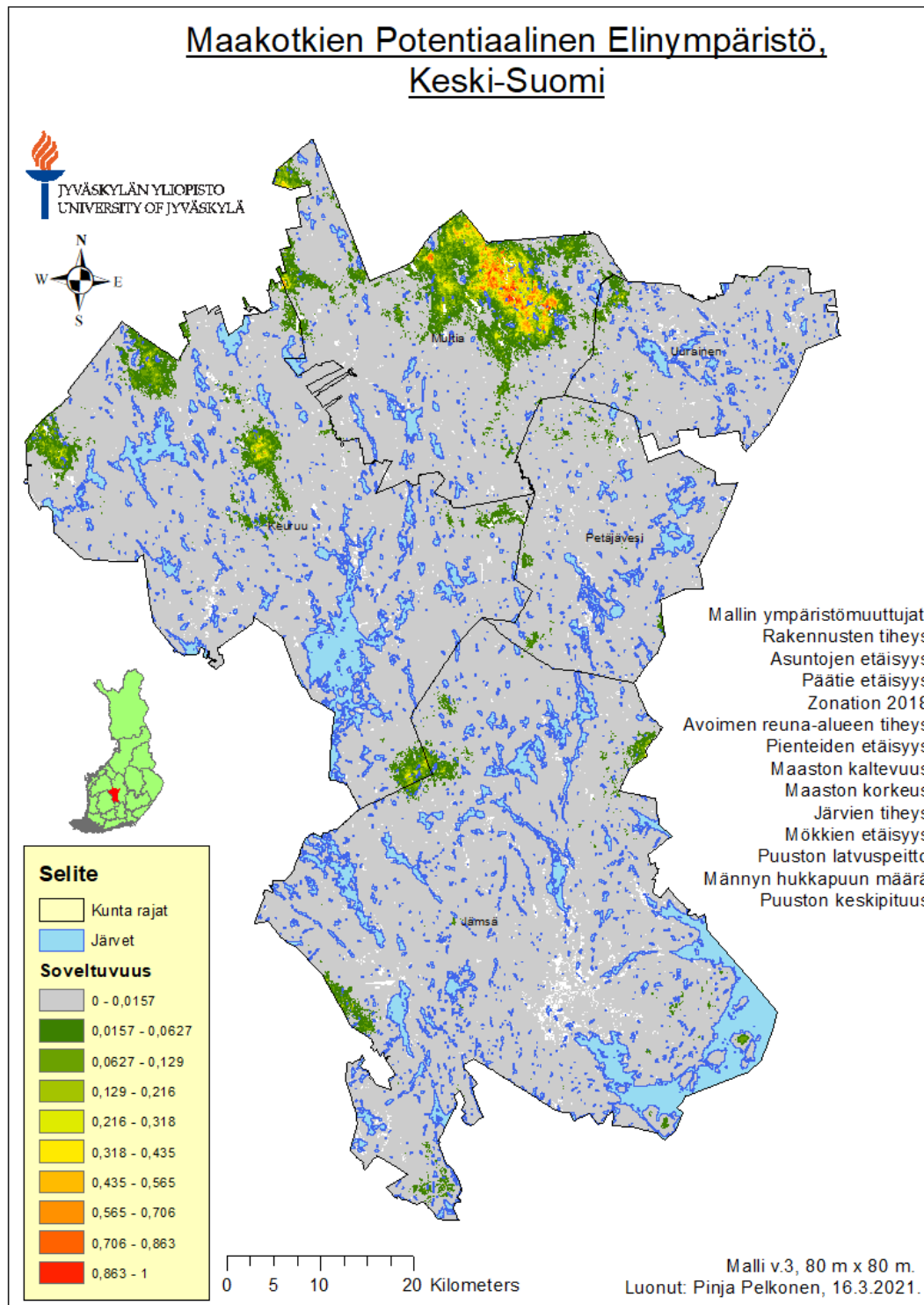
Tutkimuksen kannalta MaxEnt-ohjelman tuottama karttataso soveltuvuudesta on kenties tärkein tulos. Tästä voidaan identifioida alueita, jotka ovat maakotkan suojelun kannalta tärkeitä ja voisivat olla mahdollisia maakotkan pesintäalueita tulevaisuudessa. Kategoriallisessa todennäköisyys kartassa näkyvät mallin ennustamat erot jatkuvana arvona. Näin voidaan erottaa myös heikommin sopivia alueita vihreänä ja keltaisena (kuva 6-9). Kartta-aineistosta voidaan myös luoda eri asteisia tasoja. MaxEnt-ohjelmisto käyttää 1-0,69 arvoja optimaaliselle alueelle 0,69-0,46 arvoja tyypilliselle ja 0-0,46 arvoja huonosti sopivalle alueelle (Philips, 2017), mutta myös muita kynnyksiarvoja on käytetty kirjallisuudessa. Koska alue on laaja päätettiin Keski-Suomi jakaa 4 osaan karttoja varten (Koko kartta liitteessä 4). Kartoista käy ilmi että, hyvät ja suuri kokoiset elinympäristöt sijaitsevat Pihtiputaan, Kinnulan, Kivijärven, Kyyjärven, Multialan ja Äänekosken kunnissa. Soveltuvat elinympäristöt keskittyvät myös voimakkaasti Keski-Suomen luoteisosiin.



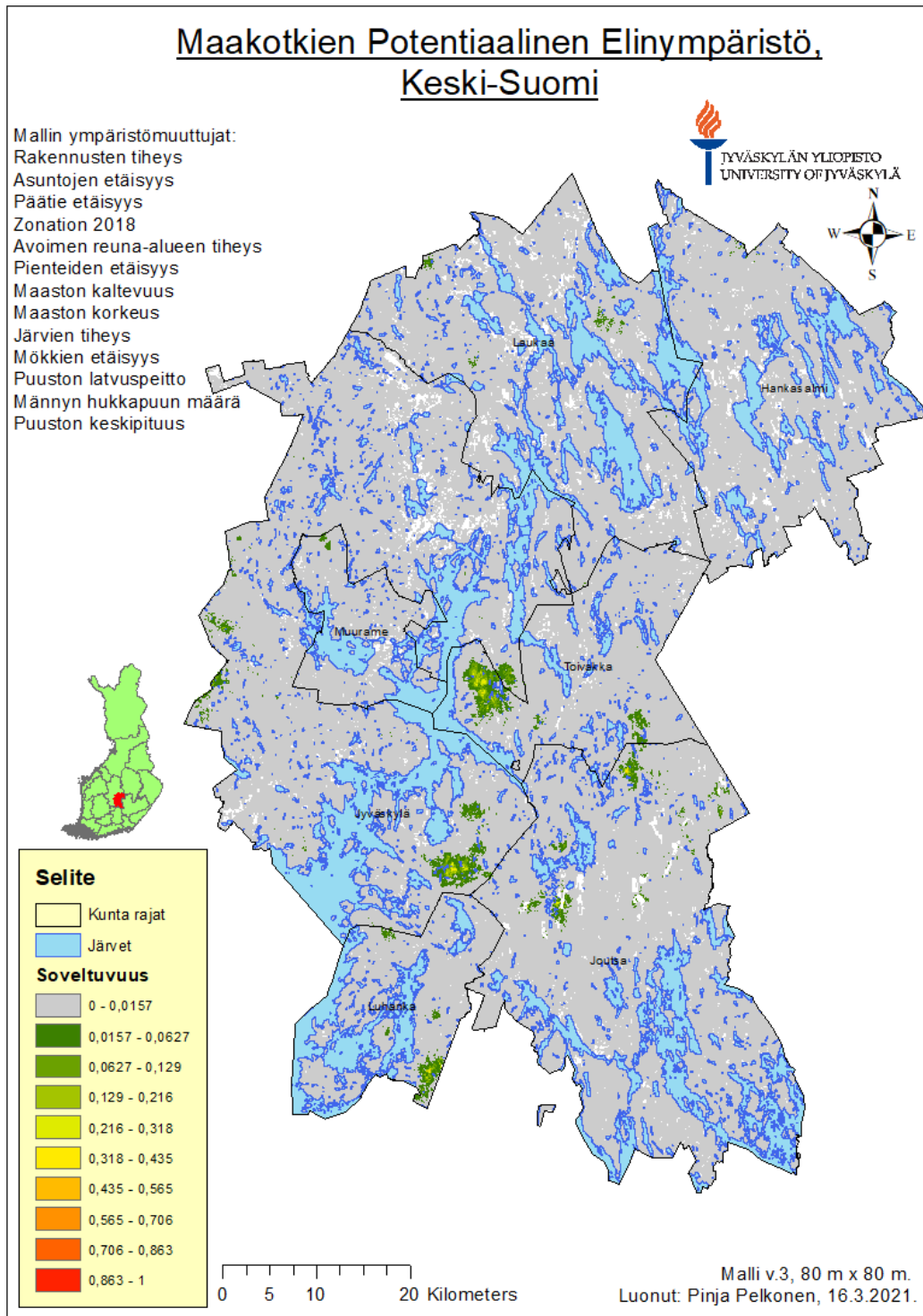
Kuva 6. Alueen ennustettu soveltuvuus maakotkan pesintäpaikaksi Keski-Suomen luoteisosissa. Punainen väri indikoi suurta soveltuvuutta, keltainen tyyppillistä ja vihreä heikkoa. Arvot muodostettiin jakamalla koko Keski-Suomen ennusteaineisto luonnollisella luokkajaolla (natural break, Jenks) 10 ryhmään. Harmaa tausta indikoi pientä ennustettua todennäköisyyttä alueen soveltuvuudesta. (Kartta sisältää Maanmittauslaitoksen Kuntajakopohjaiset hallinnolliset aluejaot 10/2020 aineistoa ja maastotietokannan rantaviiva-aineistoa, © Maanmittauslaitos 2021, sekä SYKE rantaviiva10-aineistoa 01/2021, © SYKE 2021).



Kuva 7. Alueen ennustettu soveltuvuus maakotkan pesintäpaikaksi Keski-Suomen koillisissa. Punainen väri indikoi suurta soveltuvuutta, keltainen tyypillistä ja vihreä heikkoa. Arvot muodostettiin jakamalla koko Keski-Suomen ennusteaineisto luonnollisella luokkajaolla (natural break, Jenks) 10 ryhmään. Harmaa tausta indikoi pientä ennustettua todennäköisyyttä alueen soveltuvuudesta. (Kartta sisältää Maanmittauslaitoksen Kuntajakopohjaiset hallinnolliset aluejaot 10/2020 aineistoa ja maastotietokannan rantaviiva-aineistoa, © Maanmittauslaitos 2021, sekä SYKE rantaviiva10-aineistoa 01/2021, © SYKE 2021).



Kuva 8. Alueen ennustettu soveltuvuus maakotkan pesintäpaikaksi Keski-Suomen lounaisosissa. Punainen väri indikoi suurta soveltuvuutta, keltainen tyypillistä ja vihreä heikkoa. Arvot muodostettiin jakamalla koko Keski-Suomen ennusteaineisto luonnollisella luokkajaolla (natural break, Jenks) 10 ryhmään. Harmaa tausta indikoi pientä ennustettua todennäköisyyttä alueen soveltuvuudesta. (Kartta sisältää Maanmittauslaitoksen Kuntajakopohjaiset hallinnolliset aluejaot 10/2020 aineistoa ja maastotietokannan rantaviiva-aineistoa, © Maanmittauslaitos 2021, sekä SYKE rantaviiva10-aineistoa 01/2021, © SYKE 2021).



Kuva 9. Alueen ennustettu soveltuvuus maakotkan pesintäpaikaksi Keski-Suomen kaakkois-osissa. Punainen väri indikoi suurta soveltuvuutta, keltainen tyypillistä ja vihreä heikkoa. Arvot muodostettiin jakamalla koko Keski-Suomen ennusteaineisto luonnollisella luokkajaolla (natural break, Jenks) 10 ryhmään. Harmaa tausta indikoi pientä ennustettua todennäköisyyttä alueen soveltuvuudesta. (Kartta sisältää Maanmittauslaitoksen Kuntajakopohjaiset hallinnolliset aluejaot 10/2020 aineistoa ja maastotietokannan rantaviiva-aineistoa, © Maanmittauslaitos 2021, sekä SYKE rantaviiva10-aineistoa 01/2021, © SYKE 2021).

7. Yhteenveto

Projektin aikana luotiin maksimientropiamalli maakotkan elinalueista Keski-Suomessa, jonka ennustavuus oli hyvä. Työn tuloksista voidaan havaita kriittisiä alueita lajin tulevaisuudelle Keski-Suomessa. Monet soveltuvista alueista olivat jo maakotkan käytössä, mutta hyviä elinalueita havaittiin myös asumattomilla alueilla, jonne laji voisi tulevaisuudessa levitä. Mallin mukaan lajille erityisen merkittäviä ympäristömuuttujia olivat rakennusten tiheys, asuntojen etäisyys, isojen teiden etäisyys ja zonation 2018 -taso. Tämä kertoo lajin herkkyydestä ihmistoimintaan, eikä ole yllättävä tulos mallinnuksessa. Laajimmat soveltuvat alueet lajille sijaittivat mallinnuksen mukaan Keski-Suomen luoteisosissa. Mutta myös joitakin lajille uusia alueita nousi mallinnuksen pohjalta esille, mm. Pyhä-häkin kansallispuiston lähetyvillä.

Tämän tuloksen pohjalta jatkettiin maakotkan potentiaalisten esiintymisalueiden tunnistus projektin toiseen osaan. Toisessa osassa tarkastellaan jokaista maakotka reviiriä erikseen, jotta saadaan tietoa maakotkien reviirien käytöstä. Tämä toteutetaan reviirin käytön mallintamisella Tikkanen ym. (2018) ohjeiden mukaan. Reviirinkäytön tarkastelun perusteella voitaisiin myös vähentää lajin yhteentörmäyksiä ihmistoiminnan kanssa.

LÄHTEET

- Glover-Kapfer, P. 2015. A training manual for habitat suitability and connectivity modeling using tigers (*Panthera tigris*) in Bhutan as example. doi: 10.13140/RG.2.2.34804.86409.
- Hoehler F. 2000. Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *J Clin Epidemiol*, 53(5), 499-503. doi: 10.1016/s0895-4356(99)00174-2. PMID: 10812322.
- Jaynes, E.T. 1957. Information Theory and Statistical Mechanics. *The Physical Review*, 106, 620-630. doi:10.1103/PhysRev.106.620.
- Leonard, L. 2017. Chapter One - Web-Based Behavioral Modeling for Continuous User Authentication (CUA). *Advances in Computers*, 105, 1-44.
- Parkkinen A. 2019. Kanahaukan pesäpaikkojen ennustaminen paikkatietoanalyysillä. (pro gradu -tutkielma, Jyväskylän yliopisto). Haettu osoitteesta <https://jyx.jyu.fi/handle/123456789/64736>.
- Phillips S.J. 2017. A Brief Tutorial on Maxent. http://biodiversityinformatics.amnh.org/open_source/maxent/.
- Phillips, S., Anderson, R. & Schapire, R. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.
- Phillips, S., Dudík, M. & Schapire, R. 2004. A Maximum Entropy Approach to Species Distribution Modeling. *Proceedings of the Twenty-First International Conference on Machine Learning*, 655-662.
- Simoes, M., Romero-Álvarez, D., Nuñez-Penichet, C., Jiménez, L. & Cobos, M. 2020. General Theory and Good Practices in Ecological Niche Modeling: A Basic Guide. *Biodiversity Informatics*. 15, 67-68. doi:10.17161/bi.v15i2.13376.
- Ștefănescu, D. & Bălescu, D. 2019. Predicting the distribution of Golden Eagle (*Aquila chrysaetos*) in Romania using the Maxent method. *North-Western Journal of Zoology*, 15(1), 67-74.
- Tikkanen H., Rytönen S., Karlin O-P., Ollila T., Pakanen V-M., Tuohimaa H., Orell M. 2018. Modelling golden eagle habitat selection and flight activity in their home ranges for safer wind farm planning. *Environmental Impact Assessment Review*, 71, 120-131. doi:10.1016/j.eiar.2018.04.006.
- Velasco, J. & González-Salazar, C. 2019. Akaike information criterion should not be a “test” of geographical prediction accuracy in ecological niche modelling. *Ecological Informatics*, 51, 25-32. doi:10.1016/j.ecoinf.2019.02.005.
- Zurell D. 2020. Introduction to species distribution modelling (SDM) in R. https://damariszurell.github.io/SDM-Intro/#23_Model_fitting.

LIITE 1

1 Pesätietojen käsittely

Pesätiedot käsiteltiin poistamalla ensin aineistosta tekopesät. Samassa pisteessä olevat pesät ja toisiaan läheiset pesät poistettiin spatiaalisen autokorrelaation välttämiseksi, ArcMap -ohjelman lisä työkalupaketilla SDM Toolbox v2.4, "1.Spatially Rarefy Occurrence Data for SDM's (reduce spatial autocorrelation)" -työkalulla käyttäen 5 km etäisyyttä maakotkan suuren territorion takia (Stefanesku ja Bailesku, 2019). Rajatuista pesistä kolme päätettiin vaihtaa niiden suuremman pesintä onnistumisen takia. Nämä kuitenkin katsottiin ensin läpi, jotta vaihto ei aiheuttanut läheisyyttä toisiin pesäpaikkoihin. Tämän jälkeen aineisto tallennettiin taulukkona ja yksinkertaistettiin Excel ohjelmalla, sisältämään vain koordinaatit east ja north sarakkeilla, ja lajitiedot species sarakkeessa. Lopuksi tiedot tallennettiin .csv muotoon. Pesäpaikkoja valikoitui 21 mallinnusta varten. Mallinnuksessa käytettiin vaihtoehtona myös kaikkia ei tekopesä pesäpaikkoja (yht. 41).

2 Puusto aineiston käsittely

1. Aineiston käsittely aloitettiin ladattujen laattojen yhdistämisellä "Mosaic to New Raster" työkalulla.
2. Tämän jälkeen voitiin määrittää koordinaatisto "Project Raster" työkalulla (käytettävä koordinaatisto kaikissa aineistoissa: EUREF_FIN_TM35FIN).
3. Tämän jälkeen laatta leikattiin "Clip, Data Management" työkalulla haluttuun alueeseen. Ennen "Clip" työkalun käyttöä määritellään kohderasterin "Extent" ja "Snap raster".
4. "Set Null" työkalulla vähennettiin puuttuvat arvot (32766), komennolla VALUE > 32760, saatiin muutettua pikselit NoData-arvoiksi (käsiteltävä aineisto False Raster kohtaan).

5. Metsänkayttöilmoitusten eliminointi (a-c)

- a. Aineisto ladattiin, yhdistettiin tasaiseksi laataksi "Merge" ja leikattiin "Clip Analysis" työkalun avulla kohdealueeseen.
- b. Ilmoitusten tyypit tarkastettiin ja tarkoituksena oli luoda aineisto, johon jätettiin vain 2017-2019 hakkuualueet luokista 2, 3 ja 5 monikulmioina. Myös aineiston koon takia turhat sarakkeet poistettiin attribute table:sta.

("cuttingarea" = 2) OR ("cuttingarea" = 3) OR ("cuttingarea" =5)

ja ("declarationarrivaldate" >= time '2017-01-01 00:00:00') AND ("declarationarrivaldate" >= time '2019-12-30 00:00:00')

- c. Nämä alueet voitiin sitten eliminoida VMI-aineistosta "Raster Calculator" työkalulla (kaikista paitsi kasvupaikasta). Käsky esim. Con(SetNull("MKI_rasteri"), VMI_rasteri) ja Con(IsNull(VMI_rasteri),0, VMI_rasteri). Environments>cell size määritys ja snap/extent piti olla valittuna.

6. "Aggregate" työkalua käytettiin, jotta saatiin yhtenäinen ja halutun kokoinen rasteri resoluutio. Metodiin voidaan käyttää joko solujen mediaania, keskiarvoa tai maksimi arvoa muuttujasta riippuen. Koska NoData arvot halutaan sisällyttää uusien solujen laskentaan, käytettiin "Aggregate" työkalua kun mahdollista. Laskuissa käytettiin puuston ikä = maksimi, puuston keskipituus = maksimi, latvuspeittävyys = keskiarvo, männyn tilavuus = keskiarvo, männyn hukkapuuosa = keskiarvo ja puuston kasvupaikka = mediaani.

7. Lopuksi käsitellyt tasot tallennettiin haluttuihin kansioihin ASCII muodossa valmiina seuraavaan vaiheeseen. SDM työkalupaketin "2a. Raster to Ascii (folder)" työkalulla. ja tasojen arvot tarkistettiin "Batch Calculate Statistics, Data Management" työkalulla.

(SDMtoolbox Version 2.4)

3 Korkeus aineiston käsittely

1. Aineiston käsittely aloitettiin ladattujen laattojen yhdistämisellä "Mosaic to New Raster" työkalulla.

2. Sen jälkeen laatta leikattiin "Clip" työkalulla haluttuun kohdealueeseen. Ennen "Clip" työkalun käyttöä määriteltiin kohderasterin "Extent" ja "Snap raster".

3. Mallista laskettiin kaltevuus "Slope" (z factor=1, planar metodilla ja tulos asteina) työkalulla.

4. Mallin rinteen suunta laskettiin "Aspect" työkalun avulla (z unit= meters, planar metodilla).

5. Tämän jälkeen voidaan määrittää tason koordinaatisto "Project Raster" työkalulla ja muuttaa tason solukoko halutuksi "Aggregate" työkalulla (solujen keskiarvoa käyttäen, paitsi rinteensuunta mediaania käyttäen).

HUOM! kun aineistoja vertailtiin ei korkeus aineisto tasautunut, joten käytettiin toista käsiteltyä tasoa, Environments>snap raster. Jotta tasojen rasteri soluseinät olisivat samoilla paikoilla.

6. Lopuksi käsitellyt tasot tallennettiin haluttuihin kansioihin ASCII muodossa valmiina seuraavaan vaiheeseen "Raster to ASCII" työkalulla ja tasojen arvot tarkistettiin "Calculate Statistics" työkalulla.

4 Metsän monimuotoisuus aineiston käsittely

1. Ensin aineiston koordinaatisto tarkistettiin ja laatta leikattiin "Clip" työkalulla haluttuun alueeseen. Ennen "Clip" työkalun käyttöä määritellään kohderasterin "Extent" ja "Snap raster".

2. Tämän jälkeen taso muunnettiin haluttuihin solukokoihin "Resample" työkalulla, jossa hyödynnetään toista jo käsiteltyä tasoa (Environments>snap raster ja Raster analysis>cell size) soluseinien kohdentamisessa. Tämän jälkeen havaittiin tasossa liikaa rivejä, joten se leikattiin uudelleen, toisella tasolla niin että maintain clipping extent oli valittuna.

3. Lopuksi käsitellyt tasot tallennettiin haluttuihin kansioihin ASCII muodossa valmiina seuraavaan vaiheeseen. "Raster to Ascii" työkalulla, ja tasojen arvot tarkistettiin "Calculate Statistics" työkalulla.

5 Asutus aineiston käsittely

1. Tiedostoista muodostettiin ensin yksi yhtenäinen taso, "Merge" työkalulla ja laatta leikattiin "Clip (Analysis)" työkalulla haluttuun alueeseen. Ennen "Clip" työkalun käyttöä määriteltiin kohderasterin "Extent".

2. Tämän jälkeen laskettiin etäisyys tasot "Euclidean distance" (planar metodilla) työkalulla. Tuloksena saadaan etäisyys kartta jokaiseen rakennukseen, käytössä oleviin asuinrakennuksiin (42210, 42211, 42212) ja lomarakennuksiin (42230, 42231, 42232).

HUOM! Tason attribute table:sta tallennettiin eri luokat omaksi tasoikseen ("LUOKKA" = 42210) OR ("LUOKKA"= 42211) OR ("LUOKKA"= 42212) ja laskettiin etäisyys tämän perusteella. Molempien työkalujen (2 ja 3) output cell size määritettiin 80 m ja 160 m kokoon.

3. Ydinestimointi tasot laskettiin ensin hahmottamalla talojen keskipisteet "Feature to Point" työkalulla ja sen jälkeen laskemalla ydinestimointi taso "Kernel Density" työkalulla, tämä tehtiin myös kaikille rakennuksille, asuinrakennuksille ja lomarakennuksille. Kernel density search radius:ksi määritettiin 2km, koska mm. keskimääräinen etäisyys ei tekopesien ja kaikkien rakennusten välillä oli n. 1319 m (min. 49 m ja max. 3251). 2 km search radius myös vähensi haja-asutuksen merkitystä aineistossa. Output cell size 80 tai 160.

3. Aineiston koordinaatisto tarkistettiin, sekä aineisto piti vielä kerran leikata "Clip, data management" työkalulla ylimääräisten alueiden poiston takia. Lopuksi käsitellyt tasot tallennettiin haluttuihin kansioihin ASCII muodossa valmiina seuraavaan vaiheeseen.

6 Avoin alue aineiston käsittely

1. Tason koordinaatisto tarvittaessa muutettiin "Project Raster" työkalulla ja laatta leikattiin "Clip" työkalulla haluttuun alueeseen. Ennen "Clip, Data Management" työkalun käyttöä määriteltiin kohderasterin "Extent" ja "Snap raster".

2. Seuraavaksi eroteltiin halutut alueet "Con" ja "Reclassify" työkalulla, jolla myös määritettiin alueille halutut arvot NoDataksi.

Raster calculator Con sääntö: ("VALUE" = 19) OR ("VALUE" = 20) OR ("VALUE" = 22) OR ("VALUE" = 21) OR ("VALUE" = 33) OR ("VALUE" = 34) OR ("VALUE" = 35) OR ("VALUE" = 36) OR ("VALUE" = 39) OR ("VALUE" = 41) OR ("VALUE" = 43) OR ("VALUE" = 44) OR ("VALUE" = 37)

Taulukko 1. Aukeisiin alueisiin sopivat luokat.

Corine määritelmä	Tason arvo	Tason uusi arvo
Laidunmaat	19	1
Luonnon laidunmaat	20	1
Puustoiset pelto- ja laidunmaat	22	1
Maataloustukijärjestelmän ulkopuoliset maatalousmaat	21	1
Harvapuustoiset alueet, cc<10%	33	1
Harvapuustoiset alueet, cc 10-30% kivennäismaalla	34	1
Harvapuustoiset alueet, cc 10-30%, turvemaalla	35	1
Harvapuustoiset alueet, cc 10-30%, kalliomaalla	36	1
Kalliomaat	39	1
Sisämaan kosteikot maalla	41	1
Avosuot	43	1
Harvapuustoiset alueet, sähkölinjan alla	37	1
Turvetuotantoalueet	44	1
MUUT	1..48	0

3. Tämän jälkeen rasteri aineisto muunnettiin pistemäiseksi vektoriksi "*Raster to Point*" työkalulla. Ja aineiston ydinestimointiin "*Kernel Density*" työkalulla (planar method).

Output cell size: 80 tai 160, Population field: None. Search radius: 1000 m ja 500 m.

Ștefănescu and Bălescu (2019) käytti 5 km etäisyyttä eri aluetyyppien tiheyden laskuetäisyytenä. Myös laskujen perusteella keskimääräinen etäisyys pesien ja avoimen alueen välillä oli n. 71 m (min. 4 m, max. 389 m). Tämän takia laskettiin ydinestimointi 1 km ja 500 m ytimen leveydellä vertailun vuoksi, mutta aineistoista päätettiin käyttää 1 km etäisyys ydinestimointi tasoa.

4. "*Majority Filter*" työkalulla pienten alueiden eliminoinnin jälkeen etäisyys taso laskettiin "*Euclidean distance*" työkalulla. Cell size 80 tai 160.

5. Koska avoimien alueiden läheisyys pesälle on hyväksi metsästyksen kannalta, mutta linnut eivät välttämättä pesiydy sinne, voidaan avomaiden keskiosat nollata "*Raster Calculator*" työkalulla Con() käskyllä. Tällöin malli antaa isoja arvoja vain avoimien alueiden rajoille, ei niiden keskiosille pesäpaikkana.

Con(SetNull("majority_avoin_alueet"), "kernel_tiheys_rasteri") ja tämän jälkeen

Con(IsNull("kernel_tiheys_rasteri"), 0, "kernel_tiheys_rasteri")

6. Aineiston koordinaatisto tarkistettiin, sekä aineisto piti vielä kerran leikata "*Clip, data management*" työkalulla ylimääräisten alueiden poiston takia. Lopuksi käsitellyt tasot tallennettiin haluttuihin kansioihin ASCII muodossa valmiina seuraavaan vaiheeseen.

7 Vesistö aineiston käsittely

1. Laatta leikattiin "*Clip Analysis*" työkalulla haluttuun alueeseen.
2. "*Polygon to Raster*" työkalulla muutettiin monikulmio muodot rasteri tasoksi.
3. "*Raster to point*" työkalulla rasteri alueet voitiin järjestää pisteiksi.
4. "*Near, Analysis*" työkalulla voitiin laskea matka pesien ja vesistöjen välillä, tätä käytettiin kernel tiheyden laskuetäisyyden määrittelyyn (ka. 981 m).
4. "*Kernel density*" työkalulla laskettiin tiheys tasot, search radius 1 km. Cell size 80 tai 160.
5. Aineiston koordinaatisto tarkistettiin, sekä aineisto piti vielä kerran leikata "*Clip, data management*" työkalulla ylimääräisten alueiden poiston takia. Viimeiseksi tasot tallennettiin "*Raster to ASCII*" työkalulla ja näiden arvot tarkistettiin "*Calculate Statistics*" työkalulla.

8 Tiestö aineiston käsittely

1. Aineisto yhdistettiin "*Merge*" työkalulla, ensin 3 osaan ja sitten yhteen suureen tasoon, jossa otettiin mukaan vain halutut sarakkeet.
2. Laatta leikattiin "*Clip Analysis*" työkalulla haluttuun alueeseen.
3. Attribute table:sta voitiin valita halutut luokat ja tallentaa nämä omiksi tasoiksi. Lisää tietoja maastotietokannan luokista:
<https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/old/maastotietokohteet.pdf>.

Isot tiet	Pienet tiet
12111	12131
12112	12132
12121	12141
12122	

4. "*Euclidean Distance*" työkalulla laskettiin etäisyys tasot. Cell size 80 tai 160. Extent: kohdelaue, Raster analysis>Mask>kohdealue.
5. Viimeisenä aineiston koordinaatisto tarkistettiin ja tasot tallennettiin "*Raster to ASCII*" työkalulla ja näiden arvot tarkistettiin "*Calculate Statistics*" työkalulla.

LIITE 2

Malliversiot 1.

Ennen maksimientropia mallinnusta tarkastettiin aineistojen korrelaatio ja pääkomponentti analyysi (PCA). Näiden tutkimusten tulokset osoittivat vahvaa korrelaatiota eri VMI-tasojen välillä, molemmissa rasterisolukoissa. Lisäksi pääkomponentti analyysin mukaan aineistoja oli liian useita käytössä, koska kaikki tasot eivät tuoneet lisää variaatiota aineistoon. Eivätkä kaikki ympäristömuuttujat näin lisäisi lopullisen mallin ennustavuutta.

Tämän takia testattiin vaihtoehtoisia tasoja malleilla, joissa vain nämä muuttujat olivat mukana. Malleja tehtiin VMI miinus MKI tasoilla (5 vaihtoehtoa), VMI miinus ForestWatch tasoilla (5 vaihtoehtoa), rakennus tasoilla (6 vaihtoehtoa) ja avoin alue tasoilla (2 vaihtoehtoa). VMI-aineistoa, josta oli vähennetty MKI hakkuu alueet päätettiin käyttää, koska VMI miinus MKI mallien AUC-keskiarvo oli $80 \text{ m} \times 80 \text{ m}$ mallissa korkeampi kuin VMI miinus ForestWatch mallissa. Täten VMI miinus MKI tasot olivat parempia selittäjiä mallintaessa maotkapesäpaikkoja. Aineistojen ennustavuuden erot olivat kuitenkin pieniä, sillä $160 \text{ m} \times 160 \text{ m}$ malleilla AUC-keskiarvot olivat samat. VMI miinus MKI tasoista parhaita selittäjiä olivat puiden latvuspeittävyys ja männyn hukkapuun määrä. Heikoimpia mallin selittäjiä olivat männyn tilavuus ja puuston keskipituus. Tämä nähtiin MaxEnt mallin tekemästä merkitsevyyden arvioinnista. Koska latvuspeittävyys korreloi vahvasti puiden keskipituuden ja puuston iän kanssa, $|r| \geq 0,7$, päätettiin keskipituus ja puuston ikä poistaa malliversioista 1. Myös männyn tilavuudella oli vähäinen osuus mallista, ja näin ollen sekin poistettiin malli versioista 1.

Asutus tasoista kuvaajien perusteella parhaita selittäjiä olivat asuntojen tiheys ja asuntojen etäisyys. Mutta koska asuntojen tiheys ja kaikkien rakennusten tiheys korreloivat vahvasti. Päätettiin lopullisessa mallissa käyttää kaikkien rakennusten tiheyttä ja vakituisesti asuttujen rakennusten etäisyyttä. Loma-asuntojen etäisyys tai tiheys ei ollut tärkeä selittäjä, esim. mökkien etäisyyden pois jättö ei $160 \text{ m} \times 160 \text{ m}$ mallissa olisi heikentänyt mallia. Tämän takia myös mökkien etäisyys ja tiheys muuttujat päätettiin jättää pois malli versioista 1.

Avoin alue tasoista 1 km etäisyydellä laskettu tiheys taso oli malleissa selvästi parempi selittäjä maotkan pesäpaikoista, kuin 500 m etäisyydellä luotu tiheys taso. Avoin 1 km tasoa päätettiin tämän takia käyttää.

Koska tasojen vaihtoehtoja oli paljon, päätettiin myös pääkomponentti analyysi tehdä uudestaan alustavilla muuttujilla valinnoilla ennen mallinnusta. Tämän tuloksena päätettiin myös zonation 2018 ja rinteensuunta ympäristömuuttujat karsia, koska ne eivät lisänneet tasojen variaatiota merkittävästi. Ensimmäisellä mallinnus kierroksella testattiin myös kaikkien pesäpaikkojen ja rajoitettujen pesäpaikkojen käyttöä. Koska kaikissa malleissa

kaikilla pesäpaikoilla luodut mallit saivat korkeamman AUC-arvon, päätettiin tätä käyttää vastaisuudessa.

Ympäristömuuttajat 1.version malleissa:

- Rakennusten tiheys
- Avoin reuna-alue tiheys
- Puuston kasvupaikka
- Puuston latvuspeitto
- Maaston kaltevuus
- Asuntojen etäisyys
- Maaston korkeus
- Järvien tiheys
- Männyn hukkapuuosa

Malli versiot 2

Vaikka 1 mallintamiskierroksen tulokset olivat hyviä, päätettiin testata malleja myös useammilla ympäristömuuttujilla. Tässä vaiheessa lisättiin myös järvien tiheys taso ympäristömuuttuja listaan. Kun malli oli ajettu kaikilla muuttujilla, tuloksista nähtiin, että rinteensuunta, puuston ikä ja männyn tilavuus tasojen poisto oli olennaista. Mutta seuraavaan malliin haluttiin kaikki tasot mallin perusteella sisällyttää myös mökkien etäisyys, puuston keskipituus ja zonation 2018 tasot. Lisäksi haluttiin poistaa puuston kasvupaikka taso, sen vähäisen vaihtelun takia (ei suurta vaikutusta ennustettuun cloglog arvoon).

Ympäristömuuttajat 2.versio malleissa:

- Rakennusten tiheys
- Avoin reuna-alue tiheys
- Puuston keskipituus
- Puuston latvuspeitto
- Maaston kaltevuus
- Mökkien etäisyys
- Asuntojen etäisyys
- Maaston korkeus
- Järvien tiheys
- Männyn hukkapuuosa
- Zonation 2018

Malli versiot 3

Ympäristömuuttuja listaa päivitettiin vielä vähän 2. mallintamiskierroksen perusteella. Muuttujalistaan lisättiin myös 2 tiestöstä kertovaa tasoa, ja nämä osoittautuivat hyviksi mallin selittäjiksi, joten ne päätettiin sisällyttää. Tässä vaiheessa testattiin myös mallien ennustavuutta logistisella mallinnus tuloksella ja regression multiplieri:n kasvattamisella (2,5; oletus arvon 1 sijaan). Regression multiplier avartaa mallia, niin että suurempi alue saa heikomman ennusteen, eikä malli takerru yksittäisten muuttujien arvoihin niin paljon.

Ympäristömuuttujat 3.versio malleissa:

- Rakennusten tiheys
- Avoin reuna-alue tiheys
- Suurteiden Etäisyys
- Puuston latvuspeitto
- Maaston kaltevuus
- Mökkien etäisyys
- Asuntojen etäisyys
- Maaston korkeus
- Järvien tiheys
- Männyn hukkapuuosa
- Zonation 2018
- Puuston keskipituus
- Pienteiden Etäisyys

LIITE 3

AUC-, TSS- ja AICc-arvojen lasku

MaxEnt-ohjelma tuottaa mallin ajojen jälkeen myös materiaalia, josta voidaan tarkastella mallin onnistumista. Kohdealueen sopivuus kartan lisäksi voidaan tarkastella mallin ennustuskykyä ROC-käyrillä (receiver operating characteristic curve), AUC arvoilla (ROC-käyrän alle jäävä alue) ja jackknife-analyysillä. Mallin testaus aineiston voi erotella monella tavalla, tutkielmassa päätin käyttää 5-kertaista ristiinvalidointia. Ohjelma laskee myös ajojen vaihteluvälin ja keskihajonnan automaattisesti.

ROC-käyrä tarkastelee mallin onnistumista, vertaillen oliko kartan pisteessä pesäpaikka vai ei, olematta riippuvainen kynnysarvoista (Parkkinen 2019). ROC-käyrä siis tunnistaa tyyppin I ja II virheitä. Käyrän herkkyyttä voidaan mitata sensitiivisyydellä i.e. true positive rate, jossa oikein tunnistetut positiiviset pisteet jaetaan kaikilla positiivisilla pisteillä. Tai spesifisyydellä i.e. true negative rate (Philips ym. 2006). ROC-käyrän y-akseli kuvastaa sensitiivisyyttä ja x-akseli spesifisyyttä. AUC alue kertoo ROC-käyrän alle jäävästä alueesta eli todennäköisyydestä, jossa satunnainen piste luokitellaan oikein. AUC-arvo vertailee, kuinka usein esiintymispisteiden satunnaisjoukko on suurempi, kuin ROC-käyrän alle jäävä pinta-ala verrattuna taustapisteistä valittuun joukkoon. Mitä lähempänä AUC-arvo on lukua 1, sitä luotettavampi malli on, ja jos AUC-arvo jää alle 0,6 mallia voidaan pitää huonona, sillä 0,5 AUC-arvo tarkoittaa mallin olevan yhtä luotettava kuin satunnainen arvaus kahden vaihtoehdon välillä. MaxEnt malli ei voi koskaan saavuttaa AUC-arvoa 1, koska siitä puuttuvat varmat negatiiviset osumat. Tämä voidaan kuitenkin korjata nimeämällä pisteet ilman pesäsijaintia X_{random} eli pseudonegatiivisiksi ja pesäsijainti pisteet X_{presence} eli positiivisiksi. Näin malliin voidaan syöttää satunnaisotos molemmista joukoista. Presence pisteet voidaan arvioida sensitiivisyydellä ja pseudonegatiiviset spesifisyydellä ROC funktiota varten. Mutta koska tunnettuja negatiivisia pisteitä ei ole, AUC maksimiarvo on kaavan 3 mukainen, jossa a kuvaa pesäpisteiden osuutta kaikista pikseleistä.

$$AUC_{\text{max}} = 1 - \frac{a}{2}, \quad (3)$$

Jokaisen ympäristömuuttujan merkitystä mallissa voidaan arvioida myös jackknife testin avulla. Testissä yksi muuttuja poistetaan mallista kerrallaan ja malli ajetaan uudelleen, tämän jälkeen malli ajetaan vain kyseisellä muuttujalla. Testi sovittaa viimeisenä kertana mallin kaikilla muuttujilla, Philips 2017. Nämä kolme tapaa antavat AUC-arvon mallin luontiaineiston hyvyysarvolle (regularized training gain) ja mallin testiaineiston hyvyysarvolle (test gain). Näistä testiaineiston hyvyysarvon jackknife tuloksia voidaan käyttää yleisesti mallin muuttujien merkitsevyyden analyysiin.

SDM malleja voidaan myös testata muilla tilastollisilla menetelmillä, riistiinvalidoinnin ja AUC-arvon lisäksi. Suosittuja menetelminä ovat TSS (True skill statistic) (Velasco ja González-Salazar, 2019) ja kappa koeffisientin lasku (Hoehler, 2000). AIC-arvoa (Akaiken informaatiokriteeri) voidaan myös käyttää eri mallien valintaan, mutta tätä menetelmää on

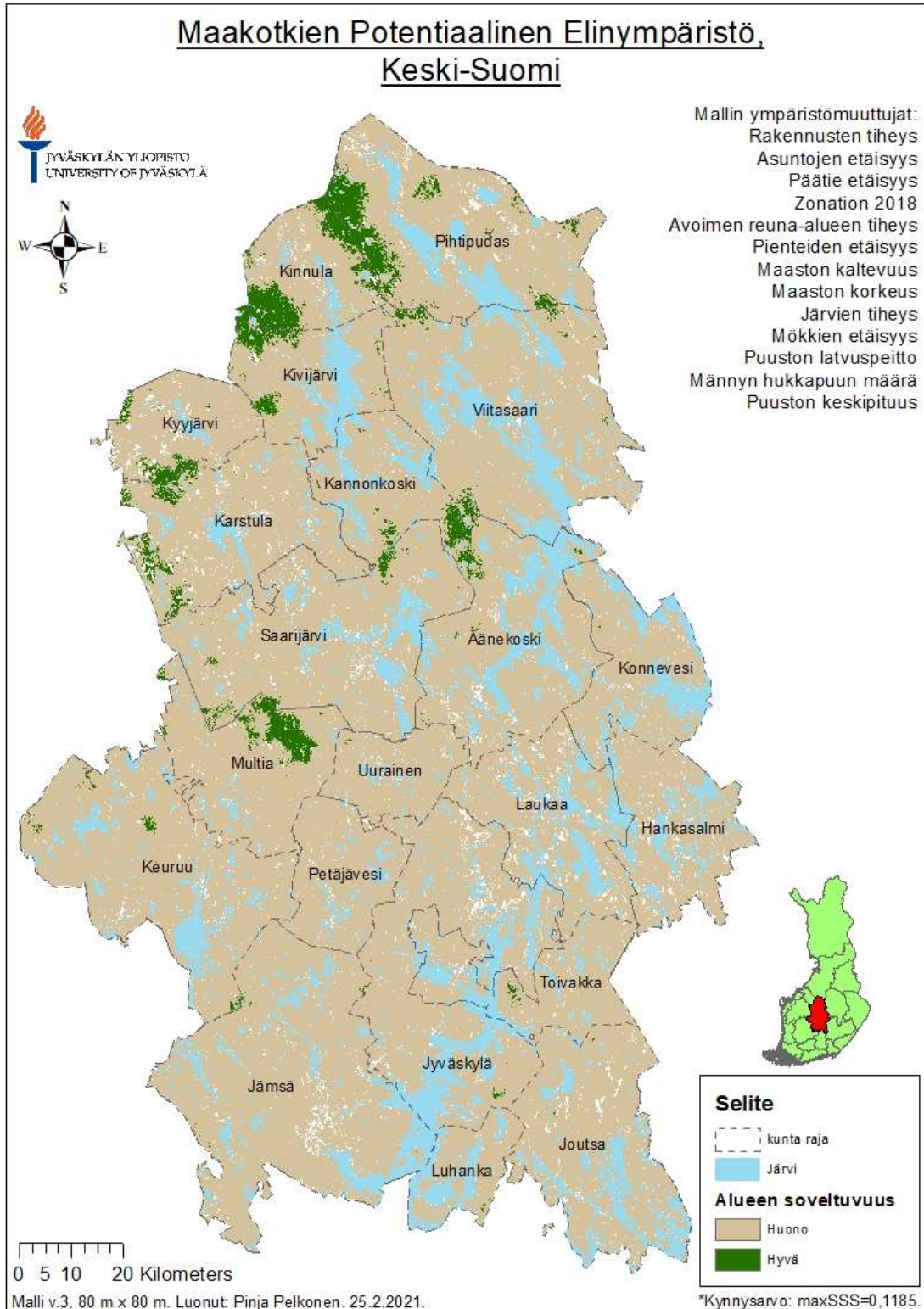
kritisoitu MaxEnt-mallien arvioinnissa (Velasco ja González-Salazar, 2019). Mutta myös hyödynnetty parhaan regularization multiplierin valinnassa (Glover-Kapfer, 2015). TSS-arvo lasketaan binäärisille tuloksille, koska siinä käytetään valittua kynnyksiarvoa mallin testiaineiston ja taustan onnistuneeseen luokitteluun. Näitä lukuja MaxEnt-ohjelma ei kuitenkaan laske automaattisesti, joten ne pitää laskea tulosten lisäksi erikseen.

Mm. Simoes ym. (2020) suositteli 3 eri kriteerin käyttöä mallien valitsemiseen. Ensimmäisenä ROC-käyrien analysoinnin, tämän jälkeen mallin ennustavuuden tarkastelun ja viimeisenä mallin monimutkaisuuden tarkastelun AICc-arvojen avulla. AICc-arvoa voidaan käyttää parhaan yksinkertaisen ja kompleksin mallin välimuodon löytämiseen, mutta se on huono indikaattori mallin maantieteellisen ennusteen tarkkuudesta. AICc-arvo laskettiin kaavan 4 mukaisesti. Jossa K kuvaa mallin parametrien määrää, n kuvaa mallin rakentamiseen käytettyjen pesäpaikkojen määrää ja logLikelihood käsittelemättömien (raw) pesäpaikka-arvojen luonnonlogaritmeista laskettua summaa.

$$(2 * K - 2 * \logLikelihood) + (2 * K) * (K+1) / (n - K - 1) \quad (4)$$

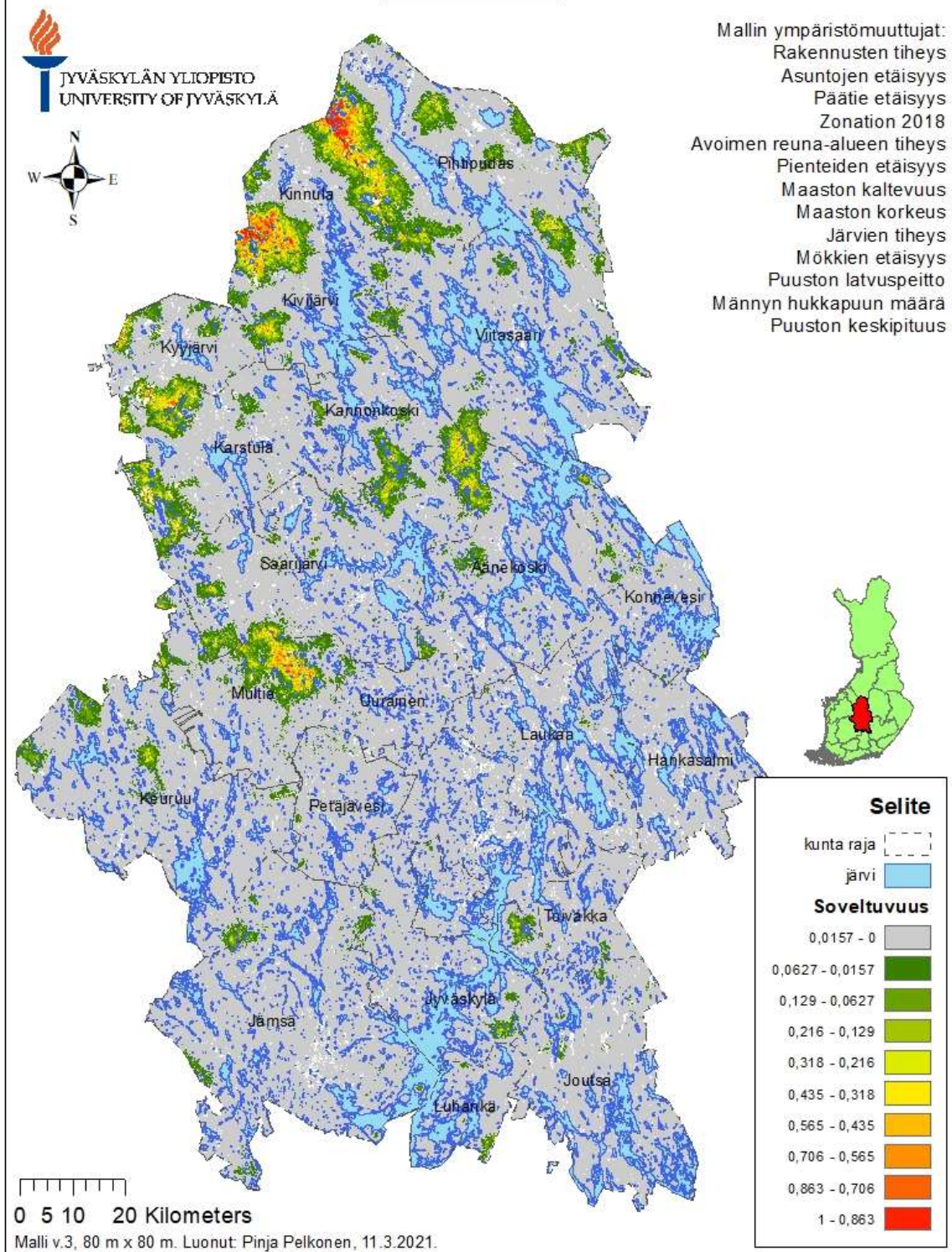
Koska pesäpaikkojen määrä oli rajallinen, oli AICc-arvo laskettava Glover-Kapfer, 2015 mukaan, vaikka R -ohjelmistolla maksimientropiamallien AIC-arvojen lasku on myös mahdollista.

LIITE 4



Binäärinen kartta lajille hyvin ja huonosti sopivista alueista. (Kartta sisältää Maanmittauslaitoksen Kuntajakopohjaiset hallinnolliset aluejaot 10/2020 aineistoa ja maastotietokannan rantaviiva-aineistoa, © Maanmittauslaitos 2021, sekä SYKE rantaviiva10-aineistoa 01/2021, © SYKE 2021).

Maakotkien Potentiaalinen Elinympäristö, Keski-Suomi



Koko Keski-Suomen kartta alueiden soveltuvuudesta, kartan teossa käytetty cubic convolution asetusta. (Kartta sisältää Maanmittauslaitoksen Kuntajakopohjaiset hallinnolliset aluejaot 10/2020 aineistoa ja maastotietokannan rantaviiva-aineistoa, © Maanmittauslaitos 2021, sekä SYKE rantaviiva10-aineistoa 01/2021, © SYKE 2021).