

**Luonnonvara- ja biotalouden tutkimus 119/2023**

# **Turvemaiden digitaalinen kartoitus ja turvepeltolohkojen tunnistaminen**

**Timo A. Räsänen, Maarit Middleton, Jonne Pohjankukka, Ville Mäkinen,  
Arttu Kivimäki, Timo P. Pitkänen, Jouni Lerssi, Matti Laatikainen,  
Tapio Väänänen, Jaakko Auri, Jaakko Heikkinen, Tapio Kananoja,  
Hanna Kekkonen, Janne Kivilompolo, Merja Myllys, Åke Möller,  
Maarit Nousiainen, Juha Oksanen, Eetu Puttonen, Aura Salmivaara,  
Jaakko Madetoja, Heikki Säävuori, Johanna Torppa ja Tapio Salo**



# **Turvemaiden digitaalinen kartoitus ja turvepeltolohkojen tunnistaminen**

**Timo A. Räsänen, Maarit Middleton, Jonne Pohjankukka, Ville Mäkinen,  
Arttu Kivimäki, Timo P. Pitkänen, Jouni Lerssi, Matti Laatikainen,  
Tapio Väänänen, Jaakko Auri, Jaakko Heikkinen, Tapio Kananoja,  
Hanna Kekkonen, Janne Kivilompolo, Merja Myllys, Åke Möller,  
Maarit Nousiainen, Juha Oksanen, Eetu Puttonen, Aura Salmivaara,  
Jaakko Madetoja, Heikki Säävuori, Johanna Torppa ja Tapio Salo**



**RUOKAVIRASTO**  
Livsmedelsverket • Finnish Food Authority

### Viittausohje:

Räsänen, T.A., Middleton, M., Pohjankukka, J., Mäkinen, V., Kivimäki, A., Pitkänen, T.P., Lerssi, J., Laatikainen, M., Väänänen, T., Auri, J., Heikkinen, J., Kananoja, T., Kekkonen, H., Kivilompolo, J., Mylly, M., Möller, Å., Nousiainen, M., Oksanen, J., Puttonen, E., Salmivaara, A., Madetoja, J., Säätuvuori, H. & Torppa J. & Salo, T. 2023. Turvemaiden digitaalinen kartoitus ja turvepeltolohkojen tunnistaminen. Luonnonvara- ja biotalouden tutkimus 119/2023. Luonnonvarakeskus. Helsinki. 76 s.

Timo Räsänen ORCID ID, <https://orcid.org/0000-0003-0839-3155>



ISBN 978-952-380-850-8 (Painettu)

ISBN 978-952-380-851-5 (Verkkójulkaisu)

ISSN 2342-7647 (Painettu)

ISSN 2342-7639 (Verkkójulkaisu)

URN <http://urn.fi/URN:ISBN:978-952-380-851-5>

Copyright: Luonnonvarakeskus (Luke)

Kirjoittajat: Timo A. Räsänen, Maarit Middleton, Jonne Pohjankukka, Ville Mäkinen, Arttu Kivimäki, Timo P. Pitkänen, Jouni Lerssi, Matti Laatikainen, Tapio Väänänen, Jaakko Auri, Jaakko Heikkinen, Tapio Kananoja, Hanna Kekkonen, Janne Kivilompolo, Merja Mylly, Åke Möller, Maarit Nousiainen, Juha Oksanen, Eetu Puttonen, Aura Salmivaara, Jaakko Madetoja, Heikki Säätuvuori, Johanna Torppa ja Tapio Salo

Julkaisija ja kustantaja: Luonnonvarakeskus (Luke), Helsinki 2023

Julkaisu vuosi: 2023

## Tiivistelmä

Timo A. Räsänen<sup>1</sup>, Maarit Middleton<sup>2</sup>, Jonne Pohjankukka<sup>1</sup>, Ville Mäkinen<sup>3</sup>, Arttu Kivimäki<sup>3</sup>, Timo P. Pitkänen<sup>1</sup>, Jouni Lerssi<sup>2</sup>, Matti Laatikainen<sup>2</sup>, Tapio Väänänen<sup>2</sup>, Jaakko Auri<sup>2</sup>, Jaakko Heikkinen<sup>1</sup>, Tapio Kananoja<sup>2</sup>, Hanna Kekkonen<sup>1</sup>, Janne Kivilompolo<sup>2</sup>, Merja Mylly<sup>1</sup>, Åke Möller<sup>4</sup>, Maarit Nousiainen<sup>2</sup>, Juha Oksanen<sup>3</sup>, Eetu Puttonen<sup>3</sup>, Aura Salmivaara<sup>1</sup>, Jaakko Madetoja, Tapio Salo<sup>1</sup>, Heikki Säävuori<sup>2</sup> ja Johanna Torppa<sup>2</sup>

<sup>1</sup> Luonnonvarakeskus (Luke)

<sup>2</sup> Geologian tutkimuskeskus (GTK)

<sup>3</sup> Maanmittauslaitos (MML), paikkatietokeskus

<sup>4</sup> Ruokavirasto

Maatalouden turvemaiden ilmasto- ja vesistö päästöjen vähentäminen edellyttää turvepelto-  
lohkojen tunnistamista, mutta maaperätieto ei ole ollut riittävän tarkkaa tähän tarkoitukseen.  
Raportissa esitellyn työn tavoitteena oli tuottaa tarkennettua paikkatietoa turvemaiden esiin-  
tymisestä ja paksuudesta turvepelto lohkojen tunnistamiseksi.

Uusi paikkatietoaineisto turvemaiden esiintymisestä ja paksuudesta luotiin hyödyntämällä ko-  
neoppimismallinnusta. Mallinnus tehtiin Random Forest -menetelmällä. Turpeen esiintymistä  
selittäviksi aineistoiksi valmisteltiin 117 kpl koko maan kattavia satelliitti- ja lentoalustoilta mi-  
tattuja kaukokartoitusaineistoja ja geologista paikkatietoaineistoa. Koneoppimismallin opet-  
tamista ja testausta varten koottiin 3,5 miljoonaa maaperähavaintoa, josta 70 % käytettiin  
mallin opetukseen ja 30 % mallin riippumattomaan testaukseen. Mallinnuksessa ennustettiin  
turvepaksuusluokkien  $\geq 10$  cm,  $\geq 30$  cm,  $\geq 40$  cm ja  $> 60$  cm esiintymistä 50 m  $\times$  50 m raste-  
riresoluutiassa ja ennusteet tuotettiin maankäyttömuodosta riippumatta kaikille maa-alueille.

Malliennusteiden tarkkuus oli korkea. Turvepaksuusluokat pystyttiin erottelemaan muista  
maalajeista ja turvepaksuusluokista 89–96 % tarkkuudella. Tarkkuudet olivat korkeimmillaan  
ohuissa turvepaksuusluokissa ja hieman heikompia paksuissa luokissa. Maatalousmailla vä-  
hintään 30 cm paksun turvemaan alaksi arvoitiin 273 000 ha, mikä on noin 11 % maatalous-  
maa-alasta. Tästä pinta-alasta 73 % turvekerros oli  $> 60$  cm. Saamamme arvio maatalousmai-  
den turvemaiden ( $\geq 30$  cm) pinta-alasta on 8 600 ha suurempi kuin mitä mittakaavaltaan  
1:200 000 maaperäkartasta voidaan arvioida. Peltolohkokohtainen tarkastelu osoitti, että  
turve-ennusteet mahdollistavat turvealan ja -paksuuden arvioimisen yksittäisillä peltoloh-  
koilla. Esimerkiksi turvepelto lohkot, joilla on vähintään 50 % alastaan  $\geq 30$  cm paksu turveker-  
ros, tunnistettiin yli 90 % tarkkuudella.

Uusi paikkatietoaineisto Turpeen paksuus 1.0/2023 tarkentaa aikaisempaa tietoa turvemaiden  
esiintymisestä ja paksuudesta koko maassa. Aineiston luokittelutarkkuus ja alueellinen erotte-  
lukyky ovat olemassa olevia maaperäkarta-aineistoja parempia ja sen avulla tunnistetaan ai-  
kaisemmin kartoittamattomia turvemaita. Yleistarkkuusmetriikat raportoidaan jokaiselle lu-  
okittelulle erikseen ja epävarmuuksien hajautuminen on esitetty Random Forest -puiden yksi-  
mielisyyden avulla rasterisolukohtaisesti. Uudet turve-ennusteet tuovat uusia mahdollisuuksia  
maaperään ja maankäyttöön liittyvien toimintojen suunnittelun, ohjaukseen ja vaikutusten ar-  
viointiin, sekä tutkimukseen.

**Asiasanat:** Turve, Histosol, kartoitus, kaukokartoitus, koneoppiminen, Random Forest

## Abstract

Timo A. Räsänen<sup>1</sup>, Maarit Middleton<sup>2</sup>, Jonne Pohjankukka<sup>1</sup>, Ville Mäkinen<sup>3</sup>, Arttu Kivimäki<sup>3</sup>, Timo P. Pitkänen<sup>1</sup>, Jouni Lerssi<sup>2</sup>, Matti Laatikainen<sup>2</sup>, Tapio Väänänen<sup>2</sup>, Jaakko Auri<sup>2</sup>, Jaakko Heikkinen<sup>1</sup>, Tapio Kananen<sup>2</sup>, Hanna Kekkonen<sup>1</sup>, Janne Kivilompolo<sup>2</sup>, Merja Mylly<sup>1</sup>, Åke Möller<sup>4</sup>, Maarit Nousiainen<sup>2</sup>, Juha Oksanen<sup>3</sup>, Eetu Puttonen<sup>3</sup>, Aura Salmivaara<sup>1</sup>, Jaakko Madetoja, Tapio Salo<sup>1</sup>, Heikki Säavuori<sup>2</sup>, Johanna Torppa<sup>2</sup>

<sup>1</sup> Natural Resources Institute Finland (Luke)

<sup>2</sup> Geological Survey of Finland (GTK)

<sup>3</sup> National Land Survey of Finland (NLS), Finnish Geospatial Research Institute (FGI)

<sup>4</sup> Finnish Food Authority

Arable lands on peat soils produce GHG emissions and impact surface water quality negatively. Mitigation of these effects requires identification of individual agricultural field parcels on peat soils. However, the accuracy and spatial detail of the existing spatial data of peat soil distribution has been limited. The objective of this work was to produce an advanced spatial dataset on the distribution of peat soils that allows identification of field parcels on peat soils.

New data on the peat soil distribution were produced by using machine learning modelling. Modelling was done using Random Forest method and 117 spatial datasets, including remote sensing data acquired on satellite and aircraft platforms and one geological GIS data source for predicting peat occurrence and depth. The model was trained and tested with 3.5 million soil observations of which 70 % were used for training and 30 % for independent testing. The occurrence of four peat depth classes  $\geq 10$  cm,  $\geq 30$  cm,  $\geq 40$  cm,  $> 60$  cm were predicted in 50 m  $\times$  50 m raster resolution. The predictions were produced for all land areas of Finland, enabling the future use of the data for various applications.

The accuracy of the predictions was high. The peat depth classes were classified and distinguished from other peat depth and soil textural classes with 89–96 % accuracy. The accuracy was the highest for the shallow peat depth classifications and slightly decreased for the deepest classifications. The total area of the  $\geq 30$  cm thick peat soils at agricultural lands was estimated to be 273,000 ha, which corresponds to 11 % of the total agricultural land area. Of this area, 73 % was estimated to be on  $> 60$  cm thick peat. The estimate of the total peat area ( $\geq 30$  cm) at arable lands is 8,600 ha larger than what can be estimated from the 1:200,000 Quaternary map. The field parcel scale evaluation showed that the new data enables the identification of peat soil extent and depth at individual field parcel level. For example, out of the field parcels with a peat cover of at least 50% of their area with a depth of  $\geq 30$  cm, over 90 % were identified correctly.

The new Peat thickness of Finland 1.0/2023 dataset improves the knowledge on peat soil distribution and depth across the entire country. The classification accuracy and the spatial detail are higher than in the existing country scale soil datasets, and previously unmapped peat soil areas are identified. Besides the classification specific accuracy measures, the spatial distribution of the agreement between Random Forest decision trees is presented to evaluate the uncertainties in the data. Altogether, the new spatial data provides improved opportunities for land use related planning, assessment, and research.

**Keywords:** Peat, Histosol, mapping, remote sensing, machine learning, Random Forest

Sisällys

<b>1. Johdanto .....</b>	<b>7</b>
<b>2. Aineistot ja menetelmät.....</b>	<b>9</b>
2.1. Turve.....	10
2.2. Mallinnusalue .....	12
2.3. Opetus- ja testausaineisto .....	12
2.3.1. Maaperähavainnot kivennäis- ja turvemailta.....	12
2.3.2. Kivennäismaapisteiden generointi .....	15
2.3.3. Turvepaksuushavaintojen korjaus peltomailla ja turvetuotantoalueilla .....	16
2.3.4. Opetus- ja testausaineistojen luominen .....	17
2.4. Kaukokartoitus- ja paikkatietoaineistot .....	18
2.4.1. Geofysikaaliset matalalentoaineistot.....	18
2.4.2. Korkeusmalliin perustuvat aineistot .....	19
2.4.3. Satelliittiaineistot.....	21
2.4.4. Geomorfologiset aineistot.....	23
2.5. Mallinnusmenetelmä.....	25
2.5.1. Random Forest (RF).....	25
2.5.2. Geneettinen algoritmi (GA) .....	26
2.5.3. Permutaatiotärkeys (PI) .....	26
2.5.4. Ositettu ristiinvalidointi .....	26
2.5.5. Metriikat.....	27
2.5.6. Mallinnusmenetelmien soveltaminen .....	28
2.6. Turpeen esiintyminen maatalousmailla .....	29
2.7. Turvepeltolohkojen tunnistaminen.....	29
<b>3. Tulokset.....</b>	<b>32</b>
3.1. Opetus ja testaus .....	32
3.2. Selittävät aineistot.....	33
3.3. Todennäköisyssennusteet.....	37
3.4. Binääriennusteet .....	39
3.5. Ennusteiden epävarmuus .....	43
3.6. Turpeen esiintyminen maatalousmailla .....	46
3.7. Turvepeltolohkojen tunnistaminen.....	48
<b>4. Tulosten tarkastelu .....</b>	<b>50</b>
4.1. Malliennusteet .....	50

4.2. Epävarmuus.....	51
4.3. Vertailu maaperäkarttojen turvetietoon .....	53
4.4. Turve-ennusteiden käyttö .....	57
4.5. Jatkotutkimus.....	60
<b>5. Johtopäätökset.....</b>	<b>62</b>
<b>Kiitokset .....</b>	<b>63</b>
<b>Viitteet.....</b>	<b>64</b>
<b>Liite .....</b>	<b>71</b>

# 1. Johdanto

Turvemaat muuttuvat usein ilmasto- ja vesistö päästöjen lähteiksi ihmistoiminnan myötä (IPCC 2014, Regina ym. 2019, Evans ym. 2021, Yli-Halla ym. 2022). Päästöjä voidaan vähentää ohjalla turvemaiden käyttöä kestäväan suuntaan, joka edellyttää luotettavaa paikkatietoa turvemaiden esiintymisestä ja paksuudesta. Maatalousmaiden osalta turvemaat olisi tunnistettava peltolohkotasolla.

Aikaisempi koko Suomen kattava paikkatieto turvemaista on perustunut Maastotietokanta (MML 2021) ja Maaperä 1:200 000 (GTK 2010) kartta-aineistoihin sekä jälkimmäisestä johdettuun Maannostietokanta 1:200 000-aineistoon (Lilja ym. 2017). Nämä aineistot ovat osoittaneet puutteellisiksi maaperätiedon tarkkuusvaatimuksien kasvaessa. Maastotietokannan suoalueiden kuvaus ei kata maatalousmaita eivätkä ne sisällä riittävän tarkkaa tietoa turvekerroksen paksuuden vaihtelusta. Maaperä 1:200 000 ja Maannostietokanta 1:200 000 sisältävät enemmän tietoa turvekerroksen paksuudesta, mutta näiden mittakaavatarkkuus ja maalajikuvioiden koko ovat riittämättömät turvepeltolohkojen tunnistamiseen. Peltolohkotasolla maalajitietoa on saatavilla laajasti näytteenottoon perustuvista viljavuusanalyysistä (esim. Eurofins Viljavuuspalvelu 2021), mutta näidenkin antama tieto on puutteellista turvepeltolohkojen tunnistamiseksi. Ne eivät kata kaikkia peltolohkoja, eivätkä kerro turvekerroksen paksuudesta tai turpeen alueellista jakautumisesta peltolohkojen sisällä, ja ne eivät ole avoimesti käytettävissä.

Digitaalinen maalajien kartoitus kehittyy jatkuvasti, ja kehitys erityisesti koneoppimismallin menetelmissä ja kaukokartoitus- ja paikkatietoaineistoissa on luonut uusia mahdollisuuksia maalajien ja maaperän ominaisuuksien kartoittamiseen (Minasny ym. 2019, 2023, Minasny & McBratney 2016). Turvemaiden kartoituksen osalta on käytetty erilaisia koneoppimismenetelmiä ja kaukokartoitus- ja paikkatietoaineistoista turvemaiden esiintymisen ja turvepaksuuden ennustamiseen (Minasny ym. 2019). Matalalennoilla tehdyt radiometriset gamma-säteilymittaukset ovat osoittautuneet tehokkaiksi turpeen esiintymistä ja paksuutta ennustettaessa (O'Leary ym. 2022, Gatis ym. 2019). Turvemaiden esiintymistä on ennustettu onnistuneesti myös korkean resoluution korkeusmalleihin (Ågren ym. 2021, 2022) ja satelliittihavaintoihin perustuvilla aineistoilla (Poggio ym. 2019).

Tässä raportissa esitetyssä työssä hyödynnetään viimeaikaista kehitystä ja tietoa turvemaiden kartoituksessa koneoppimismenetelmin ja pyritään luomaan entistä tarkempaa tietoa turvemaiden esiintymisestä ja paksuudesta koko Suomessa. Työn ensisijaisena tavoitteena oli tuottaa tarkennettua tietoa maatalousmaille, mutta työ laajennettiin kattamaan kaikki maa-alueet maankäytöstä riippumatta, sillä suurin osa olemassa olevista turvehavainnosta sijaitsi muilla kuin maatalousmailla ja muiden maankäyttömuotojen mukaan ottaminen oli menetelmällisesti suoraviivaista. Työn tarkemmat tavoitteet olivat:

1. Tuottaa uusi koko maan kattava turvema-aineisto, josta voidaan tunnistaa turvemaiden sijainti ja paksuus nykyistä tarkemmin.
2. Selvittää miten ja millä tarkkuudella uuden aineiston avulla voidaan tunnistaa turvepeltolohkoja koko maan tasolla.



Ensimmäisen tavoitteen saavuttamiseksi työssä hyödynnettiin Random Forest -koneoppimis-mallinnusta sekä laaja määrä maaperähavaintoja ja kaukokartoitus- ja paikkatietoaineistoja turvemaiden esiintymisen ja paksuuden ennustamiseksi. Ennusteet tuotettiin siten, että niillä voidaan tunnistaa turvealueet ja -paksuus Suomessa ja kansainvälisesti käytettyjen turvemaa-määritelmien mukaisesti. Toisen tavoitteen saavuttamiseksi uutta turveaineistoa tarkasteltiin peltomailla turvehavaintoja vasten ja luotiin suuntaviivoja aineiston jatkokäytölle vastaavan-laiseen käyttöön. Käytetyt aineistot, menetelmät ja tulokset on esitetty tässä raportissa. Työn aikana toteutettiin myös uusia maastokartoituksia maatalousmailla tarkoituksena parantaa mallinnuksessa havaittuja puutteita ja epävarmuuksia.

Raportin työ tehtiin Maatalousmaiden turvetieto (MaaTu) -hankkeessa osana Maa- ja metsä-talousministeriön rahoittamaa Hiilestä kiinni -tutkimus- ja innovaatio-ohjelmaa vuosina 2021–2023. Hanketta koordinoi Luonnonvarakeskus ja hankkeen muut toimijat olivat Geologian tut-kimuskeskus, Maanmittauslaitos ja Ruokavirasto. Hankkeessa on julkaistu aikaisemmin ra-portti *Turvepeltolohkojen määrittely ja tunnistaminen* (Räsänen ym. 2023), joka sisältää tietoa turvemaiden luokittelusta sekä turvepeltolohkojen tunnistamisesta ja ominaisuuksista.

## 2. Aineistot ja menetelmät

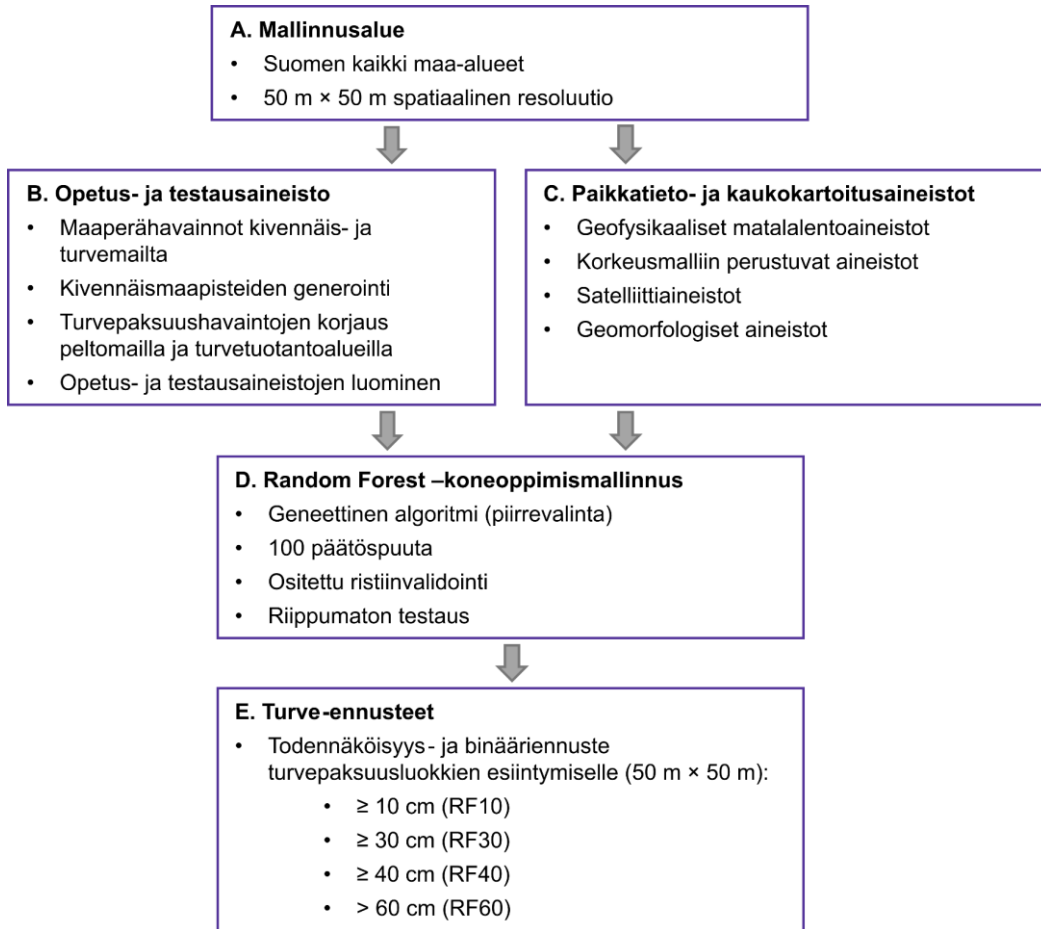
Turvemaiden esiintymistä ja turvekerroksen paksuutta mallinnettiin luokkamuuttujana Random Forest (RF) -koneoppimismenetelmällä. Mallinnuksessa ennustettiin turvepaksuusluokkien  $\geq 10$  cm,  $\geq 30$  cm,  $\geq 40$  cm ja  $> 60$  cm esiintymistä (Kuva 1). Mallinnus tehtiin kaikille maa-alueille 50 m  $\times$  50 m rasteriresoluutiassa.

Mallinnusta varten koottiin mallien opetusaineisto ja turvemaiden esiintymistä ja paksuutta selittävät paikkatietoaineistot. Opetusaineisto perustui pistemäisiin maaperähavaintoihin kivennäis- ja turvemailta, sekä työssä generoituihin kivennäismaapistetietoihin. Maaperähavainnot perustuivat vuosien saatossa maastossa tehtyihin maalajihavaintoihin ja turvekerroksen paksuusmittauksiin, mutta myös työn aikana tehtyihin maastotöihin. Generoituja kivennäismaapistetietoja tarvittiin tasapainottamaan aineistoa, sillä koottuja maaperähavaintoja oli enemmän turvemailloilla kuin kivennäismailla. Turvepaksuusmittauksien käytössä huomioitiin maatalousmailla tapahtuva turvekerroksen ohentuminen korjaamalla mittausajankohdan turvekerroksen paksuutta mitatulla keskimääräisellä turvekerroksen paksuuden ohenemisella.

Turvemaiden esiintymistä ja paksuutta mahdollisesti selittäviä kaukokartoitus- ja paikkatietoaineistoja kerättiin useista eri lähteistä ja valmisteltiin mallinnusta varten. Paikkatietoaineistot sisälsivät geofysikaalisia matalalentoaineistoja, korkean resoluution korkeusmalliin perustuvia aineistoja, satelliittiaineistoja ja geomorfologisia aineistoja. Kaikki kootut paikkatietoaineistot muunnettiin samaan 50 m  $\times$  50 m rasteriresoluutioon seuraten geofysikaalisten matalalentoaineistojen alueellista erottelukykyä. Radiometriset lentogeofysiikan aineistot on todettu tärkeimmiksi turpeen esiintymistä ja paksuutta kuvaaviksi aineistoiksi, mutta niiden alueellinen erottelukyky on samalla heikoin. Tämän vuoksi radiometrisen aineiston tietosisältöä ei haluttu muokata.

Mallinnuksessa kokeiltiin erilaisia mallinnusmenetelmiä. Mukana kokeilussa oli RF-koneoppimismenetelmän lisäksi Gaussian Process Regression (GPR) -koneoppimismenetelmä, sekä Geographically Weighted Regression (GWR), joka on spatiaaliseen regressioon perustuva tilastollinen menetelmä. Alustavien vertailuiden perusteella RF-menetelmä osoittautui tehokkaaksi menetelmäksi ja se tuotti hyvän lopputuloksen. Tämän vuoksi RF valittiin mallinnuksessa käytettäväksi menetelmäksi. GWR-menetelmä osoittautui raskaaksi aineistojen laajuuden vuoksi, ja GPR-menetelmän hyödyntäminen edellytti monimutkaisempaa mallinnusasetelmaa, jonka toteuttaminen ei ollut mahdollista hankkeen puitteissa. Mallinnus toteutettiin useaan kertaan ja mallinnuksen aineistoja ja mallinnusstrategiaa paranneltiin mallinnuskertojen välillä.

Turvepeltolohkojen tunnistamiseksi peruslohkoista valittiin testiaineisto, jonka lohkot luokiteltiin tunnetuiksi kivennäis- ja turvemaiksi opetuspisteiden perusteella. Opetuspisteiden kaltaisesti aineistoon lisättiin oletettuja kivennäismaan lohkoja tasapainottamaan turvelohkojen suurta osuutta. Samoilta lohkoille ennustettiin vastaavat luokat myös mallinnusaineistojen perusteella siten, että luokituksessa testattiin erilaisia mallinnustodennäköisyyden kynnyksarvoja. Tämän jälkeen mallinnusten soveltuvuutta turvepeltolohkojen tunnistamiseen arvioitiin vertaamalla mallinnus- ja opetuspisteluo- kkia toisiinsa.



**Kuva 1.** Turvemaiden esiintymisen ja turvepaksuuden mallinnusprosessi. Prosessin aineistot ja menetelmät on kuvat tarkemmin seuraavissa kappaleissa.

## 2.1. Turve

Turpeeksi määritellään tässä työssä orgaaninen materiaali, joka on syntynyt veden kyllästä-mässä, vähähappisessa, happamassa ja ravinneköyhässä ympäristössä ja se sisältää pääasi-assa osittain maatunutta orgaanista ainesta, kuten kasvien jäänteitä (Kuva 2) (IUSS Working Group WRB, 2022; Soil Survey Staff, 1999). Turpeen muodostumisen edellytyksenä on, että orgaanisen aineksen hajotus on rajoittunut hapen puutteen sekä mahdollisten viileiden ja kosteiden olosuhteiden vuoksi.



**Kuva 2.** Turveprofiili pellolla (kuva: Merja Mylly).

Turvemaan tai -maannoksen määrittelyssä työ ottaa huomioon Suomessa yleisesti käytetyn määritelmän ja kansainvälisesti käytetyn WRB-määritelmän, mutta ei ota kantaa siihen kumpaa määritelmää pitäisi käyttää. Suomessa turpeen määrittelyssä on käytetty orgaanisen aineen raja-arvoa  $\geq 40\%$ , ja turvemaita on luokiteltu ohuisiin turpeisiin turvekerroksen paksuuksilla 30–60 cm ja paksuihin turpeisiin paksuuksilla  $> 60$  cm (GTK 2010, Lilja ym. 2017). Kansainvälisesti käytetyssä WRB-määritelmässä turvemaat luetaan Histosol-maannoksen alle, jossa orgaanisen kerroksen orgaanisen hiilen määrän raja-arvo on  $\geq 20\%$ , orgaanisen kerroksen on oltava  $\geq 40$  cm tai  $\geq 10$  cm jos sen alla on kovaa materiaalia kuten kiveä, ja orgaanisen kerroksen on sijaittava maan ylimmässä 100 cm kerroksessa (IUSS Working Group WRB 2022). Orgaanisen aineen määrä 40 % vastaa kutakuinkin orgaanisen hiilen määrää 20 % (IUSS Working Group WRB 2022). Turvemaiden määrittelystä Suomessa ja kansainvälisesti on kirjoittanut laajemmin esimerkiksi Räsänen ym. (2023).

Työssä tuotetut turveaineistot huomioivat yllä mainittujen turveluokitusten turvepaksuuksien raja-arvot toteuttamalla mallinnus  $\geq 10$  cm,  $\geq 30$  cm,  $\geq 40$  cm ja  $> 60$  cm turvepaksuusluokille. Orgaanisen materiaalin ominaisuuksia tai orgaanisen aineen ja hiilen määrää ei ole ennustettu mallinnusprosessissa, mutta ne on pyritty huomioimaan mallinnuksessa käytettyjen maaperähavaintoaineistojen valmistelussa. Maaperähavainnoissa turve on tunnistettu suurelta osin aistinvaraisesti ja pienemmässä määrin orgaanisen aineen ja hiilen määritysten perusteella käyttäen yllä mainittuja raja-arvoja.

## **2.2. Mallinnusalue**

Turpeen esiintymistä ja paksuutta mallinnettiin koko Suomen maa-alueella. Mallinnus tehtiin 50 m × 50 m rasteriresoluutiolla. Rasterisolun koko ja asemointi määräytyivät keskeisten geofysikaalisten matalalentomittausten tarkkuuden mukaan. Mallinnusta varten luotiin ensin maamaski Maanmittauslaitoksen maastotietokannan (Maanmittauslaitos 2021) perustella, jonka tarkoituksena oli määrittää mallin avulla ennustettavat hilaruudut. Koska maastotietokannassa ei ole erillistä maaelementtiä, luotiin se poistamalla vesialueet (meri, järvet ja aluemuotoiset virtavedet) Suomen rajojen sisäpuolelle jäävästä alueesta. Tämän jälkeen mallinnuksessa käytettävä maamaski tehtiin valitsemalla ne rasterisolut, jotka osuivat kokonaan tai osittain maa-alueelle.

## **2.3. Opetus- ja testausaineisto**

### **2.3.1. Maaperähavainnot kivennäis- ja turvemailta**

Maaperähavainnot koostuvat GTK:n, Luke:n ja Euroopan komission yhteisen tutkimuskeskuksen (The Joint Research Centre JRC) maaperäaineistoista, jotka sisältävät turvekerroksen paksuusmäärytyksiä, sekä erilaisia havaintoja ja määrytyksiä pintamaasta. Turvekerroksen paksuus-havaintoja sisältäviä aineistoja olivat GTK:n turvetutkimusaineisto (GTK 2014) ja happamat sulfaattimaat (HASU) -aineisto (GTK 2018), sekä Luke:n Valtakunnan metsien inventointi (VMI) -aineisto (Korhonen ym., 2021) ja eri tutkimushankkeiden turveprofiiliaineistot. Pintamaahavaintoja sisältäviä aineistoja olivat GTK:n maaperän kartoitus- ja tutkimusaineistot (GTK 2017), JRC:n LUCAS-aineisto (Fernandez-Ugalde ym. 2022, Jones ym. 2020, Tóth ym. 2013), Luke:n peltomaiden kemiallisen tilan valtakunnallinen seurantatutkimus (VALSE) (e.g., Soinne ym. 2022) ja Luke:n tutkimushankkeiden pintamaa-aineistot. Maaperähavaintoaineistot on kuvattu seuraavissa kappaleissa ja Taulukko 1.



**Taulukko 1.** Mallinnuksessa käytetyt maaperähavaintoaineistot, eli mallien opetus- ja testausaineiston osa-aineistot.

Aineisto	Kuvaus	Havaintoa	Turvehavaintoa	Lähde
<b>Turvepaksuusaineistot</b>				
Turvetutkimusaineisto	Turvekerroksen paksuusmääryksiä turvekerroksen pohjaan saakka.	1 648 802	1 648 802	GTK (2014)
Valtakunnan metsien inventointi (VMI)	Turvekerroksen paksuuden määryksiä pääosin metsämailta. Max määryitys syvyys 400 cm.	163 909	60 479	Korhonen (2021)
Happamat sulfaattimaat (HASU) -aineisto	Havaintotiedot turve, multamaa, kivennäismaa. Turvepaksuus määritetty max. 200–300 cm.	23 301	6089	GTK (2018), Auri ym. (2022)
Tutkimushankkeiden turveprofiiliaineistot	Eri hankkeiden maaprofiilimääryksiä turvemailta.	71	71	Esim. Yli-Halla ym. (2022), Orgaanisten maiden ilmastopäästöjen hillintä nautakarjatilolla (OMAIHKA), Turvepeltojen hiilipäästöt kuriin innovatiivisella vesienhallinnalla (Vesihäisi)
<b>Pintamaa-aineistot</b>				
Maaperän kartoitus- ja tutkimusaineisto	Maaperä- ja kallioperäkartoituksen kivennäismaa- ja kalliopaljastumahavainnot	428 811	32 686	GTK (2017)
LUCAS	Aineistojen ja 2009, 2015 ja 2018 Pintamaanäytteitä kaikilta mailta, syvyys 0–20 cm.	1 979	356	Fernandez-Ugalde ym. (2022), Jones ym. (2020), Toth ym. (2013)
Peltomaiden kemiallisen tilan valtakunnallinen seuranta tutkimus (VALSE)	Aineistojen 2009 ja 2018 pintamaanäytteitä peltoilta, syvyys 0–20 cm.	757	69	Esim. Soinne ym. (2022)
Tutkimushankkeiden pintamaa-aineistot	Pintamaanäytteitä kaikilta mailta, syvyys 0–30 cm.	2231	28	Esim. Soinne ym. (2020), Urvas (1995), Maatietorekisteri, Jokioisten kartanon maanäytteet
<b>Generoidut aineistot</b>				
Generoidut kivennäismaapistet	Satunnaisesti generoidut kivennäismaahavaintopisteet	1 282 425	-	GTK, aineisto tuotettu osana tämän raportin työtä.

Turvetutkimusaineisto (GTK 2014) sisältää maastomittauksiin perustuvia turvekerroksen paksuustietoja. Aineisto on koottu GTK:n turvevarojen kartoituksessa vuosina 1945–2021 ollen mallinnuksessa suurin yksittäinen turvekerroksen paksuustietoja sisältävä aineisto. Se sisältää mitatun paksuustiedon koko turvekerroksen osalta turvemaan pinnasta turpeen ja kivennäismaan kontaktiin saakka, riippumatta turvekerroksen paksuudesta. Maalajimääritys on tehty maastossa visuaalisesti, eikä se perustu orgaanisen aineen tai hiilen määrän määrytyksiin. Mallinnusta varten kivennäis- ja turvemaalajimääritys tehtiin aineiston alkuperäisen luokittelun mukaisesti. Valtaosa turvetutkimusaineiston turvekartoituspisteistä sijaitsee ojittamattomilla ja metsäojitetuilla soilla, jonkin verran pisteitä on lisäksi viljelyssä olevilla tai hylätyillä turvepelloilla.

Happamat sulfaattimaat (HASU) -aineiston (GTK 2018) havainnot perustuvaa maastossa tehtyihin kairaushavaintoihin, joista on määritetty maalajit sekä maakerrospaksuudet 2–3 m syvyyteen saakka. Maalaji on määritetty aistinvaraisesti. Mallinnusta varten kivennäis- ja turve-maalajimäärittely tehtiin aineiston alkuperäisen luokittelun mukaisesti. Havainnot sijaitsevat pääosin muinaisen Litorina-meren alueella ja ne sisältävät havaintoja kivennäis- ja turvemailta, sekä maatalous- ja metsämaita.

VMI-aineisto (Korhonen ym. 2021) keskittyy metsävarojen inventointiin, mutta se sisältää myös inventointien yhteydessä tehtyjä turvekerroksen paksuusmittauksia. Mittaukset on tehty suorassilla joko kahden (suurin osa aineistoista) tai neljän (osa vanhimmista aineistoista) metrin enimmäissyvyyteen asti. Viimeisimmissä VMI-mittauksissa ympyräkoealan säteenä on ollut 9 m, jonka alueelta on objektiivisesti valittu neljä kairauspistettä. Näiden pisteiden perusteella on määritetty orgaanisen kerroksen laatu ja paksuus. Mittausten yksityiskohdissa on kuitenkin pieniä eroja mittausvuodesta riippuen.

Luke:n tutkimushankkeiden turveprofiiliaineistot koottiin Luke:n eri tutkimus hankkeissa kerätyistä maatalousmaiden turveprofiilihavainnoista (Esim. Yli-Halla et al. 2022; Orgaanisten maiden ilmastopäästöjen hillintä nautakarjatiloiilla OMAIHKA -projekti). Havainnot perustuivat kairaamalla tehtyyn näytteenottoon tai maakuopista tehtyyn tarkasteluun. Ne sisälsivät tiedon turvekerroksen paksuudesta ja turpeen ominaisuuksista eri syvyyksiltä (esim. orgaaninen hiili). Kaikki mittaukset eivät yllä turvekerroksen pohjaan saakka. Profiileja on analysoinut tarkemmin Räsänen ym. (2023).

Maaperän kartoitus- ja tutkimusaineistot (GTK 2017) koostuvat maaperäkartoituksen yhteydessä piikittämällä tehdystä pintamaahavainnosta (0–1 m) ja kairausten, koekuoppien sekä leikkausten kerroshavaintotiedosta. Kerroshavaintotiedoista on mukana ainoastaan ylimmän pintamaan kerrostieto. Lisäksi aineistossa on mukana kalliopaljastumahavaintoja.

Eurooppalainen LUCAS-aineisto (Fernandez-Ugalde ym. 2022, Jones ym. 2020, Tóth ym. 2013) sisältää näytteenottoon perustuvia havaintoja pintamaasta (0–20 cm). Vuosien 2009, 2015 ja 2018 aineistot yhdistettiin ja saman kohteen vanhemmat havainnot poistettiin. Pintamaahavainnot luokiteltiin mallinnusta varten kivennäis- ja turvemaiksi orgaanisen hiilen määrän määritysten perusteella. Luokittelussa käytettiin raja-arvoina orgaanisen hiilen määrää 20 %, joka vastaa kutakuinkin orgaanisen aineen määrää 40 % (IUSS Working Group WRB 2022). Aineiston havaintopisteet sijaitsevat metsä- ja maatalousmailla.

Peltomaiden kemiallisen tilan valtakunnallinen seurantatutkimuksen (VALSE) aineisto (esim. Soenne ym. 2022) sisältää näytteenottoon perustuvia havaintoja pintamaasta (0–15 cm). Vuosien 2009 ja 2018 aineistot yhdistettiin ja saman kohteen vanhemmat havainnot poistettiin. Pintamaahavainnot luokiteltiin mallinnusta varten kivennäis- tai turvemaiksi orgaanisen hiilen määrän määritysten perusteella. Luokittelussa käytettiin raja-arvoina orgaanisen hiilen määrää 20 %, joka vastaa kutakuinkin orgaanisen aineen määrää 40 % (IUSS Working Group WRB, 2022). Aineiston havaintopisteet sijaitsevat yksinomaan peltomailla.

Luke:n tutkimushankkeiden pintamaa-aineistot (esim. Soenne ym. 2020, Urvas 1995, Maatietorekisteri; Jokioisten kartanon maanäytteet) koostuivat eri tutkimushankkeissa tehdyistä pintamaahavainnoista ja -näytteenotoista (0–30 cm). Pintamaahavainnot luokiteltiin kivennäis- tai turvemaiksi orgaanisen aineen tai hiilen määrän perusteella. Luokittelussa käytettiin raja-arvoina orgaanisen aineen määrää 40 % (Haavisto 1983) ja orgaanisen hiilen määrää 20 %, jotka

vastaavat kutakuinkin toisiaan (IUSS Working Group WRB 2022). Aineiston havaintopisteet sijaitsevat pääosin peltomailla, mutta havaintoja on jonkin verran myös metsämailla.

Yllä lueteltujen aineistojen lisäksi GTK ja Luke tekivät vuosina 2022–2023 maastotöitä pelto-alueilla maaperähavaintoaineiston täydentämiseksi mallinnusta varten. Kaikkiaan kartoitettiin noin 7 000 ha peltoalaa, 560 yksittäistä peltolohkoa. Jokaiselle peltolohkolle tehtiin vaihteleva määrä mittauspisteitä (0,9 mittauspistettä/ha), joilta määritettiin maalaji visuaalisesti tarkastelemalla sekä mitattiin mahdollinen turvekerroksen paksuus venäläisellä laippakairalla tai maaperäpiikillä. Vuonna 2022 tehtiin lisäksi turvekerroksen paksuuden uusintamittauksia peltolohkoilla, joilta GTK tai Luke oli tehty aikaisemmin turvekerroksen paksuusmittauksia. Uusintamittauksilla selvitettiin mahdollista turvekerroksen paksuuden ohentumista (Räsänen ym. 2023). Vuosien 2022–2023 maastokartoituksissa tehtiin yhteensä 6 380 uutta mittauspistettä, joista turvetta esiintyi 34 %:lla pisteistä. 14 %:lla pisteistä maalajiksi määritettiin multa ja 51 % pisteistä oli kivennäismaapisteitä.

Kartoitettavien peltolohkojen valinnassa käytettiin hankkeessa valmisteltuja aineistoja, joilla kartoitettiin alueita, joilla oli havaintopuutteita tai mallinnustulos oli epävarma. Peltolohkot tunnistettiin Ruokaviraston peltolohkokisterin avulla. Kartoitettavat peltolohkot pyrittiin sijoittamaan mahdollisimman kattavasti eri puolille Suomea, mutta käytännössä maastotyölu-pien saaminen vaikutti kartoitettavien lohkojen sijaintiin.

### **2.3.2. Kivennäismaapisteiden generointi**

Varsinaisten maaperähavaintojen lisäksi luotiin keinotekoisia kivennäismaapisteitä (turpeen 0-syvyyspisteitä) maa-alueille, joilla ei ole turvetta tai sitä on alle 10 cm. Lisäopetusaineistoja luotiin, koska kootut aineisto olivat epätasaisesti jakautuneita turve- ja kivennäishavaintojen määrän ja sijainnin suhteen painottuen lähinnä turvemaille. Opetuspisteiden luonti on kuvattu seuraavissa kappaleissa.

Kivennäismaapisteiden luonnissa merkittävin koko Suomen kattava paikkatietoaineisto oli Maanmittauslaitoksen maastotietokanta, johon on tallennettuna maastokohteiden, muotojen ja infrastruktuurin tila. Maastotietokannan tietojen ylläpito on jatkuvaa ja sitä päivitetään perusteellisesti alueesta riippuen 3–10 vuoden väliajoin (Maanmittauslaitos 2023). Kivennäismaapisteiden luonnissa käytetty maastotietokanta oli vuoden 2012 versio (Haltik, Hallinnon tietotekniikkakeskus), joka ei ole enää ajantasainen rakennetun ympäristön ja peltojen suhteen. Tämän voidaan olettaa olevan etu turpeen 0-syvyyskohteita tunnistettaessa, koska ne alueet, jotka ovat esimerkiksi raivattu vuoden 2012 jälkeen pelloiksi, ovat kuvattuna vanhassa maastotietokannassa metsämaana, suona tai soistumana. Samoin myös alueet, jotka ovat myöhemmin metsittyneet, voivat olla vanhemmassa maastotietokannassa kuvattuna silloisen käyttötarkoituksen mukaan. Pelto- ja niittyalueet jätettiin kivennäismaa-aluepisteiden luonnista pois, koska näillä alueilla maanmuokkaaminen on muuttanut maanpinnan pientopografiaa eli tasoittanut sitä. Toisaalta luonnontilainen kasvillisuus on korvattu viljelykasveilla, joiden perusteella ei voida kovin hyvin päätellä maaperän laatua.

Kivennäismaa-alueiden löytämiseksi oletuksena (deduktiivinen päättely) oli, että sellaiset maastotietokannan alueet, jotka eivät ole vettä, suoalueita, soistumia, rakennettuja alueita (taajamat, rakennukset, täyttömaa-alueet, tiestö, rautatiet, johto- ja putkilinjat), maa-ainesten ottoalueita, peltoja ja niittyjä, olisivat tietyn suuruisen *puskurivyöhykkeen ulkopuolella* hyviä 0-pistealueita eli kivennäismaa-alueita. Todennäköiset kivennäismaa-alueet saatiin

maastotietokannan avulla esille sulkemalla epätodennäköiset alueet pois aineistosta. Jokaisesta edellä kuvatusta viivamaisesta tai aluekohdeluokasta puskurointivyöhykkeen luonnin jälkeen tehtiin binäärinen rasteri. Lopuksi nämä rasterit kerrottiin keskenään (esim. Tokola & Kalliovirta 2003). Näin saatiin binäärinen tulorasteri, jossa arvolla 1 olivat potentiaaliset kivennäismaa-alueet. Näillä alueilla turpeen esiintyminen oli epätodennäköistä. Arvon 0 saivat ei-toivotut alueet eli vesistöt, soistumat, suot, maa-ainesten ottoalueet, rakennetut alueet jne. Näille potentiaalisille kivennäismaa-alueille luotiin satunnaisesti alueen koko huomioiden edustava määrä 0- eli kivennäismaapisteitä. Näistä pisteistä lopuksi vielä suodatettiin pois sellaiset pisteet, jotka ovat jonkin maaperäkartta-aineiston (1:20 000, 1:50 000, 1:100 000 ja 1:200 000) kuviolla, joissa on lieju-, turve-, vesistö-, täytemaa- ja kartoittamattomia alueita tai soistumaa (1:200 000).

Maa-alueet, joille pisteitä generoitiin ovat käytännössä kivennäismaita, jotka eivät ole soistuneet tai niissä ei tapahdu soistumista. Lisäksi generoitavan pisteen kohdalla häiriöiden minimoimiseksi luotiin ei-haluttujen piirteiden (esimerkiksi sähkölinjat, putkilinjat, tiestö, rautatiet, rakennetut alueet ja vesistöt) puskurialue, jonka ulkopuolella generoitu piste sijoitettiin satunnaisesti.

Satunnaispisteet eivät kuitenkaan sijoittuneet rasterin polygonimuunnoksen jälkeen reuna-alueilla aina onnistuneesti, vaan osa pisteistä jäi alkuperäisen tulorasterin 1- alueen ulkopuolelle. Tämä johtui Raster to Polygon -muunnosvaiheessa tapahtuvasta polygonin reunaviivan yleistymisestä eli viiva pyrkii oikomaan ulospäin sisäänpäin kääntyviä reunaviivan mutkia. Tämän ilmiön takia vektoroitua tulospolygonia vielä puskuroitiin 25 m sisäänpäin, jotta polygonin reunaviiva hyvin todennäköisesti pysyy alkuperäisen rasteriarvon 1 alueen sisällä. Tämän supistetun tulospolygonin alueelle muodostettiin spatiaalisella satunnaisgeneraattorilla edustavasti yksittäisen polygonin alueeseen suhteutettu määrä pisteitä siten, että pienin pistemäärä oli 2 (pienimmät alle 2 hehtaarin kuviot) ja suurin pistemäärä 1205 (pinta-alaltaan suurin kuvio) polygonikuviota kohden. Pienen pistejoukon sijaintia muokattiin manuaalisesti visuaaliseen tarkasteluun perustuen. Kivennäismaapisteiden generointiprosessi on kuvattu tarkemmin raportin liitteessä.

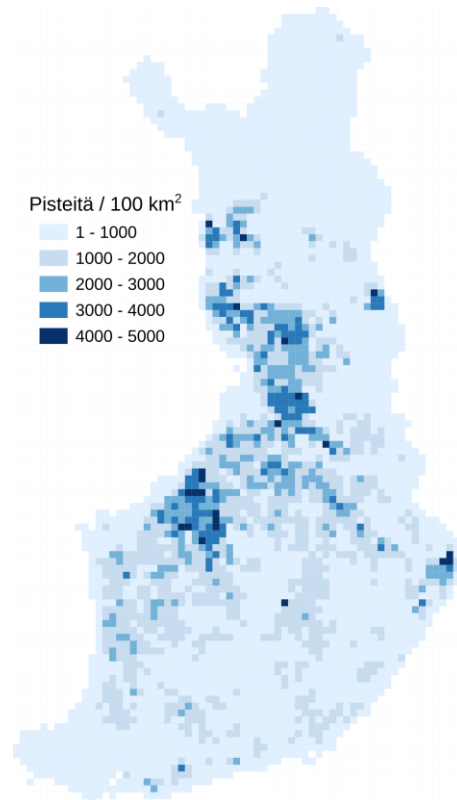
### **2.3.3. Turvepaksuushavaintojen korjaus peltomailla ja turvetuotantoalueilla**

Turveltojen turvekerroksen on havaittu ohenevan turpeen hajoamisen ja painumisen myötä. Tämän vuoksi turveltojen peltojen mittausajankohdan turvekerroksen paksuustietoa oli korjattava. Kesällä 2022 tehdyssä seurantatutkimuksessa tehtiin turvekerroksen paksuuden uusintamittauksia 35 pellolla keskimäärin 20 vuotta ensimmäisen mittauksen jälkeen. Tulokset osoittivat, että turvekerros on ohentunut keskimäärin 1,2 cm vuodessa (keskihajonta 0,6 cm; (Räsänen ym. 2023). Koneoppimismallinnuksessa käytetyt, peltoalueilla sijaitsevat turvekerroksen paksuusmittaustiedot olivat pääosin vuosilta 1991–2020 ja niitä korjattiin uusintamittauksissa mitatun keskimääräisen turvekerroksen ohentuman (1,2 cm/vuosi) perusteella.

Turvetuotantoalueilla sijaitsevia turvehavaintoja ei käytetty mallinnuksessa, sillä niiden paksuustiedot olivat usein vanhentuneita tai epäluottavia, eikä niitä voitu korjata samoin kuin peltomaiden havaintoja. Turvetuotantoalueet tunnistettiin Maanmittauslaitoksen maastotietokannan vuosien 2005–2021 versioiden avulla (MML 2021).

### 2.3.4. Opetus- ja testausaineistojen luominen

Eri maaperähavaintoaineistojen maalaji- ja turvepaksuustiedot yhdistettiin yhdeksi opetusaineistoksi. Yhdistetty aineisto sisälsi tiedon kunkin havaintopisteen maalajista kivennäis- ja turvemaaluokitellun tarkkuudella, tiedon mitatusta turvekerroksen paksuudesta ja tiedon, oliko turvekerroksen mittausta tehty turvekerroksen pohjaan eli kivennäismaahan saakka, vai oliko turvekerroksen paksuus mitattu tiettyyn syvyyteen saakka, vaikkei turvekerroksen pohjaa ollut saavutettu. Koottu opetusaineisto sisälsi yhteensä 3 513 954 opetuspistettä, joista 1 703 595 pisteellä oli mitattu tai havaittu turvetta. Maatalousmailla opetuspisteitä oli 97 309 pistettä, joista 54 867 pisteellä oli mitattu tai havaittu turvetta. Pisteet ovat jakautuneet spatiaalisesti epätasaisesti painottuen alueille, missä on ollut energiaturpeen käyttötarpeita (Kuva 3).



**Kuva 3.** Opetus- ja testausaineistojen pisteiden spatiaalinen tiheysjakauma.

Koko havaintoaineisto jaettiin kahteen ositteeseen, opetusaineistoksi (70 %) jolla koneoppimismalli koulutettiin, ja testausaineistoksi (30 %), jota käytettiin mallin tarkkuuden riippumattomaan arviointiin. Ositteet pyrittiin luomaan olennaisilta ominaisuustiedoiltaan samankaltaisiksi. Lisäksi opetus- ja testiaineistojen välinen spatiaalinen autokorrelaatio pyrittiin minimoimaan. Spatiaalisesti autokorreloituneissa ositteissa pisteet sijaitsivat liian lähellä toisiaan ja sen vuoksi sisältäisivät keskinäistä yhteisvaihtelua. Spatiaalisen autokorrelaation vaikutuksen vuoksi pisteiden jako pelkän satunnaisen valinnan perusteella ei soveltunut ositukseen, vaan se tehtiin seuraavalla tavalla:



1. Ominaisuustietovertailua varten koko opetuspisteaineistolle laskettiin jakaumaosuudet neljälle turvesyvyysluokalle (0 cm, 1–29 cm, 30–60 cm ja yli 60 cm) ja mitattujen / generoitujen pisteiden osuuksille. Tämän lisäksi pisteiden itä- että pohjoiskoordinaatit järjestettiin suuruusjärjestykseen ja määriteltiin molemmille sellaiset peräkkäiset koordinaattirajat, joiden jokaisen väliin jäi 25 % osuus kaikista pisteistä.
2. Pisteiden valintaa varten luotiin säännöllinen 10 x 10 km ruudukko yli koko Suomen. Opetuspisteitä valittiin testausositteeseen satunnaisista ruuduista, kunnes valittujen ja jäljellä olevien pisteiden (opetusosite) osuudet olivat suhteessa 30/70. Testausositteesta poistettiin tämän jälkeen sellaiset valintaruutujen reunoilla olleet pisteet, joista 200 m säteellä oli yksikin opetusositteen piste. Valittu 200 m säde vastasi erillisen analyysin mukaan sellaista etäisyyttä, jonka jälkeen läheisten turvesyvyyksien keskinäinen korrelaatio oli heikentynyt lähes satunnaisten pisteiden vertailua vastaavalle tasolle. Reunapisteiden poistolla ei ollut merkittävää vaikutusta ositteiden jako-osuuksiin.
3. Valitun testausositteen jakaumaosuuksia (turvesyvyysluokat, mitattujen / generoitujen pisteiden osuudet sekä itä- ja pohjoiskoordinaattiluokat) verrattiin koko opetuspistejoukon vastaaviin osuuksiin laskemalla niiden neliöidyt erotukset. Nämä erotukset summattiin yhdeksi vertailuarvoksi kuvaamaan sitä, miten hyvin valitun testausositteen (ja samalla myös opetusositteen) turveominaisuudet ja maantieteellinen jakautuminen keskimäärin vastasivat koko opetuspistejoukkoa.
4. Opetus- ja testausositteiden valinta toistettiin 1000 kertaa, ja lopulliset aineistot koostettiin vertailuarvoltaan pienimmän otoksen mukaisesti. Tämän perusteella opetusositteeseen valittiin 2 457 612 ja testausositteeseen 1 028 788 pistettä, ja valintaruutujen reunoilla olleet 27 554 pistettä jäivät aineistoista pois.

## 2.4. Kaukokartoitus- ja paikkatietoaineistot

### 2.4.1. Geofysikaaliset matalalentoaineistot

GTK teki koko maan kattavaa geofysikaalista lentomittauskartoitusta ns. matalalentomittauksena vuosina 1972–2007. Lentokorkeutena oli 30–40 metriä ja lento-linjojen väli oli pääasiassa 200 metriä. Lentolinjat kulkevat pohjoisesta etelään tai idästä länteen geologisen pääsuuntaukseen perustuen, ja mittauspisteiden väli lentolinjoilla oli 6–50 metriä. Mitatut geofysikaaliset suureet ovat maan magneettikenttä, maankamaran sähkömagneettisen kentän vaste ja luonnon gammasäteily (Airo, 2005). Lentomittausaineistot on interpoloitu linjamuotoisesta aineistosta rasteritiedostoiksi 50 metrin resoluutioon.

Mallinnuksessa käytettiin mitattua naparedusoitua magneettivuon tiheyttä, sähkömagneettisten mittausten reaali- ja imaginäärikomponentteja ja näiden suhdetta. Gammasäteilymittauksista mallinnukseen valittiin torium- ( $^{232}\text{Th}$ ) ja kaliumsäteily ( $^{40}\text{K}$ ), näiden suhdetta ja toriumin, kaliumin ja uraanin ( $^{238}\text{U}$ ) säteilystä laskettua säteilyannosnopeutta (Taulukossa 4). Magneettivuon tiheys ja sen vaihtelu aiheutuu maankamaran (suurelta osin kallioperä) magnetoituneista mineraaleista ja kivilajeista. Sähkönjohtavuusmittausten reaalikomponentti kuvaa lähinnä hyviä sähkönjohteita, jotka ovat pääasiassa kallioperän aiheuttamia. Imaginaarikomponentti tuo puolestaan esille heikompia johteita, jotka usein liittyvät soihin, järviin ja kosteikkoihin. Gammasäteilymittauksien avulla voidaan arvioida kallioperän ja mineraalimaa-aineksen

gamma-säteilyn vaimenemista väliaineessa. Gammasäteilyn vaimeneminen on verrannollinen väliaineen tiheyteen ja turpeen tapauksessa tiheyttä kasvattaa sen sisältämä vesimäärä (Davisson 1968, Endrestøl 1980, Gatis ym. 2019, Siemon ym. 2020). Gammasäteilyaineistot on todettu kansainvälisissä (Minasny ym. 2019) ja Suomessa tehdyissä (Airo ym. 2014, Hyvönen ym. 2005, Virtanen 1997) tutkimuksissa yhdeksi tärkeimmistä aineistoista turvemaiden kaukokartoituksessa. Tässä tutkimuksessa käytettiin myös 1:10 milj. mittakaavaisen kallioperäkartan kivilajien (13 kpl) kalium-säteilytasolla korjattua kaliumkomponenttia sekä kaikkia Taulukko 4 mainittuja aerogeofysikaalisia aineistoja. Kallioperän kivilajien kalium-säteilytasokorjaus tehtiin laskemalla kivilajiyksiköiden keskimääräiset kalium-säteilytasot, joiden keskiarvon suhteen säteilytasot normeerattiin. Tämä paransi hieman aineiston merkittävyyttä luokittelussa.

Aerogeofysikaalisiin sähkömagneettisiin ja gammasäteily mittaustuloksiin perustuvaa soiden paksuustulkintaa on esitetty raportissa Lerssi ym. (2023). Raporttiin sisältyy myös turpeen paksuusluokittelu tunnettujen turvepeltojen osalta Pohjanmaan alueelta.

#### **2.4.2. Korkeusmalliin perustuvat aineistot**

Mallinnuksessa käytettiin kahta aineistoa, jotka perustuvat korkean resoluution laserkeilausaineistosta johdettuun korkeusmalliin (MML 2020). Nämä on esitelty seuraavissa kappaleissa ja Taulukko 4.

Ensimmäinen oli Depth-to-Water (DTW) indeksi (Murphy ym. 2007, 2008, 2009), joka ennustaa maaperän kosteutta korkeusmallin ja siitä laskettujen virtaamaverkkojen perusteella. DTW indeksin yksikkö on metri (m). Se indikoi maaperän kosteutta kohteen etäisyyden, korkeuseron ja kaltevuuden perusteella lähimpään virtaamaverkon uomapisteeseen nähden. Mallinnuksessa käytettiin 2 m × 2 m resoluutiossa laskettua DTW-indeksiä (Salmivaara 2020), jota täydennettiin koko maan kattavaksi laskemalla puuttuville alueille DTW-indeksit Whitebox-ohjelmalla (Lindsay 2016). Lisäksi laskettiin DTW-indeksi 2 ha:n kynnysarvolla saatavilla olevien kynnysarvoindeksien lisäksi. Kynnysarvoista 1 ha kuvaa kosteita olosuhteita esimerkiksi lumen sulannan jälkeen, 2 ha keskimääräisiä olosuhteita, 4 ha kuvaa normaalin loppukesän kuivempaa tilaa, ja 10 ha kuvaa selvästi keskimääräistä kuivempia olosuhteita. DTW-indeksin voidaan olettaa ennustavan turvemaiden esiintymistä, sillä turvemaat ovat syntyneet märissä tai vedenkyllästämissä olosuhteissa. DTW-indeksit yleistettiin mallinnuksessa käytettyyn 50 m resoluutioon laskemalla alkuperäisten pikseliarvojen keskiarvo.

Toinen korkeusmallista johdettu aineistojoukko kuvaa maanpinnan rosoisuutta, jota tarkasteltiin erilaisia tunnuslukuja käyttäen. Tavoitteena oli kvantifioida topografian tekstuuria ja kuvata maaston paikalliseen korkeusvaihteluun liittyviä piirteitä, joiden ajateltiin avustavan turvesyvyyksien ennustamisessa. Työn aluksi tunnistettiin artikkelilähteisiin perustuen 11 eri tunnuslukua (Taulukko 2), ja niiden hyödyllisyyden ja samankaltaisuuden tarkastelua varten valittiin Etelä-Pohjanmaalta Keski-Suomeen ulottuva testialue, jolla oli noin 300 000 mallinnuksessa käytettyä opetuspistettä. Tältä testialueelta laskettiin kaikkien 11 tunnusluvun arvot Whitebox-ohjelmalla (Lindsay 2016) ja saadut arvot yhdistettiin opetuspisteisiin. Osa tunnusluvuista oli laskettavissa vain tietyssä ikkunakoossa (3x3 tai 5x5), viitaten laskennassa käytetyn naapuruston kokoon keskuspikselin arvoa laskettaessa, ja muut laskettiin useassa koossa (3x3, 5x5, 11x11, 21x21, 31x31). Tämän jälkeen tunnuslukujen arvojakaumia tarkasteltiin sekä maalajeittain (kalliopaljastuma, kivennäismaa, multamaa, turve) että turvesyvyyksien mukaan

luokiteltuina. Jatkotarkasteluun valittiin kuusi sellaista tunnuslukua, joiden havaittiin erottavan tarkastellut luokat.

**Taulukko 2.** Topografian monimittakaavaisen tekstuurin tarkasteluun käytetyt tunnusluvut ja niiden laskentaan WhiteBox-ohjelmassa hyödynnetyt työkalut. Jatkotarkasteluun valittujen tunnuslukujen nimet on lihavoitu, ja mallinnukseen valituilta tunnusluvuilta on ilmoitettu ikkunakoot.

Tunnusluku	Tunnusluvun kuvaus	Mallinnuksen ikkunakoot	Lähde
Absoluuttinen korkeusero (difference from mean elevation)	Korkeusero tarkastellun pikseliarvon ja ikkunan keskiarvon välillä (m)	-	Ågren ym. (2021), Langlois ym. (2017), Richardson & Millard (2018)
Topografiaan suhteutettu korkeusero (deviation from mean elevation)	Keskihajonnalla normalisoitu (suhteutus paikalliseen topografiavaihteluun) korkeusero tarkastellun pikseliarvon ja ikkunan keskiarvon välillä	-	Ågren ym. (2021)
<b>Korkeusarvojen vaihtelevuus (standard deviation filter)</b>	Korkeusmallin pikseliarvojen keskihajonta	31x31	Ågren ym. (2021), Grohman ym. (2011)
<b>Pinta-alasuhde (surface area ratio)</b>	Pinta-alojen suhde korkeusmallia mukailevan ja täysin tasaisen (ilman korkeusvaihtelua olevan) pinnan välillä	3x3	Grohman ym. (2011)
<b>Rosoisuusindeksi (terrain ruggedness index)</b>	Neliöityjen korkeuserojen keskiarvon neliöjuuri, laskettuna tarkastellun pikseliarvon ja sen naapurien välillä	-	Ågren ym. (2021), Riley ym. (1999)
<b>Rinnekaltevuus (slope)</b>	Rinteen kaltevuus (asteina, 0–90)	-	Ågren ym. (2021), Poggio ym. (2019), Aitkenhead (2017)
Rinnesuunta (aspect)	Rinteen suunta ilmaistuna ”pohjoisuutena” eli aste-erona pohjoissuunnasta, olettaen että etelä-pohjoinen akseli on merkityksellisin ja itä-länsi suuntia ei erotella (asteina, 0–180)	-	Aitkenhead (2017)
<b>Rinnekaltevuuden vaihtelevuus (standard deviation of slope)</b>	Rinnekaltevuusarvojen keskihajonta	31x31	Ågren ym. (2021), Grohman ym. (2011)
<b>Rinnesuunnan vaihtelevuusindeksi (circular variance of aspect)</b>	Rinnevektorien perusteella laskettava suuntavaihtelua (0–1) kuvaava indeksiarvo, jolla mitattuina pienemmät arvot tarkoittavat vähäisempää rinnevaihtelua	5x5, 11x11, 31x31	Ågren ym. (2021), Grohman ym. (2011), Lindsay ym. (2019)
Rinteen vertikaalinen kaarevuus (profile curvature)	Rinteen kaarevuus mitattuna pystysuunnassa (korkeuskäyriä vastaan)	-	Grohman ym. (2011), Aitkenhead (2017)
Rinteen horisontaalinen kaarevuus (plan curvature)	Rinteen kaarevuus mitattuna sivusuunnassa (korkeuskäyriä seuraten)	-	Grohman ym. (2011), Aitkenhead (2017)

Seuraavaksi haettiin kuuden edellä valitun sekä niistä eri ikkunakoissa laskettujen tunnuslukujen joukosta keskenään heikoiten korreloivat tunnusluvut laskemalla Pearsonin korrelaatio-kertoimet. Parittainen tarkastelu aloitettiin suurimman korrelaation saaneesta tunnuslukuparista (mukaan lukien saman tunnusluvun eri ikkunakoot) ja kummankin hyvyys mitattiin opeuspisteiden turve- vs. ei-turvemaiden luokittelulla maalajitiedon mukaisesti logistiseen regressioon pohjautuen. Tarkasteltavasta parista tiputettiin aina pois se tunnusluku, jonka AIC-arvo oli heikompi. Tätä jatkettiin, kunnes minkään kahden tunnusluvun välillä ei ollut 0.75 ylittävää korrelaatiota. Näin saatiin valittua mallinnuksessa käytetyt tunnusluvut.

Edellä mainitut 2 m korkeusmallin pohjalta lasketut tunnusluvut laajennettiin kattamaan koko mallinnusalue, ja ne yleistettiin mallinnuksessa käytettyyn 50 m resoluutioon laskemalla alkuperäisten pikseliarvojen keskiarvo ja keskihajonta. Vaikka osa alkuperäisistä tunnusluvuista perustui keskihajontaan, ajateltiin yleistyksessä käytetyn keskihajonta-arvon tuovan silti mahdollista lisätietoa rosoisuuden vaihtelusta 50 m rasterisolun alueella.

### 2.4.3. Satelliittiaineistot

Maisematasolla turvemaiden esiintymisen kannalta tärkeitä tekijöitä ovat muun muassa kasvillisuus ja hydrogeologia (Buffam ym. 2010). Näitä tekijöitä voidaan havainnollistaa optisilla ja tutka-alueen satelliittikuvilla, joita hyödynnetään usein turvemaiden kartoituksessa niiden spatiaalisen kattavuuden ja saatavuuden vuoksi (Minasny ym. 2019). Optisista satelliittikuvista lasketut indeksit indikoivat kasvillisuuden tyyppiä ja biomassan määrää, mikä voi auttaa erottamaan kivennäis- ja turvemaan toisistaan. Vastaavasti tutka-alueen satelliittikuvat kertovat maaperän kosteudesta, jonka temporaalinen vaihtelu turvemailla on kivennäismaita vähäisempää.

Optisen alueen satelliittikuvien lähteenä käytettiin Euroopan avaruusjärjestö ESA:n Sentinel-2 -ohjelmaa (ESA, 2023b). Sentinel-2 -kuvien spatiaalinen resoluutio vaihtelee 10, 20 ja 60 metrin välillä kuvauskanavan mukaan. Temporaalinen resoluutio Suomen alueella on muutama päivä, mutta pilvisuus vaikeuttaa kuvien käyttöä. Yksittäiset vuodelta 2021 kerätyt satelliittikuvat prosessoitiin koko Suomen kattaviksi mosaiikkeiksi hyödyntäen sen2mosaic-työkalua (Bowers 2020) perustuen mahdollisimman pilvettömien pikselien arvoihin. Kanavakohtaisista heijastearvoista laskettiin kirjallisuuskatsauksen (mm. Poggio ym. 2019) perusteella valittuja indeksejä (Taulukko 3). Indeksimosaiikit tuotettiin 30 vuorokauden liukuvin aikaikkunoin huhtikuusta lokakuuhun alkaen jokaisen kuukauden 1. ja 15. päivä, jolloin jokaista indeksia kohden syntyi 13 mosaiikkia.

Tutka-alueen (SAR, Synthetic Aperture Radar) satelliittikuvien lähteenä käytettiin Euroopan avaruusjärjestö ESA:n Sentinel-1 -ohjelmaa (ESA 2023a). Muokatun aineiston lähteenä ovat Suomen ympäristökeskuksen ja Ilmatieteen laitoksen yhteistyössä tuottamat SAR-kuvamosaiikit 11 vuorokauden tarkasteluajanjaksoilta, alkaen jokaisen kuukauden 1., 11. ja 21. päivä (SYKE 2021). Mosaiikit kuvaavat pikseleiden gamma<sup>0</sup>-takaisinsironnan keskiarvoja ajanjaksolta sekä VV- että VH-polarisaatioilta. Datan aukkoisuuden vuoksi mosaiikeista laskettiin edelleen pikselikohtaiset keskiarvot viideltä peräkkäiseltä mosaiikilta, jolloin saatiin mallinnuskäyttöön neljä mosaiikkia kummastakin polarisaatiosta (Taulukko 3).

Polarisaatiomosaiikkien lisäksi alkuperäisistä mosaiikeista laskettiin polarisaatiosuhde kaavalla  $(VH-VV)/(VH+VV)$  seuraten Poggio ym. (2019) artikkelia. Polarisaatiosuhde laskettiin kullekin 11 vuorokauden alkuperäismosaiikille erikseen. Yhdentoista vuorokauden polarisaatiosuhde-mosaiikeista laskettiin edelleen viiden peräkkäisen mosaiikin keskiarvot, kuten VV- ja VH-polarisaatioiden suhteenkin tehtiin. Kaikki Sentinel-1 -aineistoista tuotetut mosaiikit ovat siten samasta lähtöaineistosta ja samoilta, noin kahden kuukauden aikaikkunoilta huhtikuun ja lokakuun 2021 väliltä (Taulukko 3).

Kaikki satelliittiaineistot sovitettiin mallinnusalueen 50 m x 50 m rasteriresoluutioon, mikä heikensi sekä optisten että tutka-aineistojen tarkkuutta. Optisten aineistojen alkuperäinen spatiaalinen resoluutio oli 10 metriä ja tutka-aineistojen 20 metriä. Sovitus tehtiin R-ohjelmistolla (v.4.3.0; R Core Team 2022) ja terra-kirjaston resample-funktiolla (Hijmans ym. 2023). Mallinnukseen valmistellut satelliittiaineistot on esitelty aineistokohtaisesti Taulukko 3 ja Taulukko 4.

**Taulukko 3.** Turpeen esiintymisen mallintamista varten Sentinel-1 (S1) ja Sentinel-2 (S2) havainnoista johdetut mosaikit.

Indeksi	Indeksin kuvaus	Lähde
Normalised Difference Moisture Index (NDMI), S2	NDMI on kosteusindeksi, joka reagoi kasvillisuuden kosteuspitoisuuteen.	Gao (1996). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Normalised Difference Vegetation Index (NDVI), S2	NDVI reagoi vihreän kasvillisuuden määrään.	Rouse ym. (1973) Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Kernel Normalised Difference Vegetation Index (kNDVI), S2	kNDVI on matemaattisesti paranneltu versio NDVI:stä	Camps-Valls ym. (2021). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Enhanced Vegetation Index 2 (EVI2), S2	EVI2 on kasvillisuusindeksi, jonka on todettu reagoivan muutoksiin maaperän kosteudessa	Jiang ym. (2008). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
NBR (Normalized Burn Ratio), S2	NBR (Normalized Burn Ratio) on paloindeksi, joka reagoi palaneisiin alueisiin maastossa.	Key ja Benson (1999). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Normalized Difference Snow Index (NDSI), S2	NDSI on lumi-indeksi ja reagoi lumeen.	Riggs ym. (1994). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Soil-Adjusted Vegetation Index (SAVI), S2	SAVI indeksi on kasvillisuusindeksi ja huomioi kasvillisuuden lisäksi myös maaperän vaikutuksen heijastevasteeseen.	Huete (1988). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Soil-color Index (SCI), S2	SCI on maaperän väri-indeksi. Käytetty aiemmin turvemaakartoituksessa, esim. Poggio ym. (2019).	Esim. Poggio ym. (2019) Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Soil Moisture (SM), S2	SM on maaperän kosteusindeksi. Käytetty aiemmin turvemaakartoituksessa, esim. Poggio ym. (2019).	Dupigny-Giroux ja Lewis (1999). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Tasseled-Cap Transformation Brightness (TCTB), S2	Tasseled-Cap Transformation on kolmesta eri osa-alueesta koostuva kokonaisuus: kirkkaus, vihreys ja kosteus (engl. brightness, greenness & wetness), joka suunniteltiin alkujaan Landsat-aineistoille (Kauth & Thomas 1976). Indeksien laskentakertoimet Sentinel-2 –aineistoille artikkelista Shi ja Xu (2019). Käytetty aiemmin turvemaakartoituksessa, esim. Poggio ym. (2019).	Kauth ja Thomas (1976), Shi ja Xu (2019). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Tasseled-Cap Transformation Greenness (TCTG), S2	Tasseled-Cap Transformation on kolmesta eri osa-alueesta koostuva kokonaisuus: kirkkaus, vihreys ja kosteus (engl. brightness, greenness & wetness), joka suunniteltiin alkujaan Landsat-aineistoille (Kauth & Thomas 1976). Indeksien laskentakertoimet Sentinel-2 –aineistoille artikkelista Shi & Xu (2019). Käytetty aiemmin turvemaakartoituksessa, esim. Poggio ym. (2019).	Kauth ja Thomas (1976), Shi ja Xu (2019). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
Tasseled-Cap Transformation Wetness (TCTW), S2	Tasseled-Cap Transformation on kolmesta eri osa-alueesta koostuva kokonaisuus: kirkkaus, vihreys ja kosteus (engl. brightness, greenness & wetness), joka suunniteltiin alkujaan Landsat-aineistoille (Kauth & Thomas 1976). Indeksien laskentakertoimet Sentinel-2 –aineistoille artikkelista Shi & Xu (2019). Käytetty aiemmin turvemaakartoituksessa, esim. Poggio ym. (2019).	Kauth ja Thomas (1976), Shi ja Xu (2019). Indeksimosaiikki tuotettu osana raportin työtä (FGI).
SAR VV-polarisaatio, S1	VV-polarisaatio tarkoittaa SAR-tutkan vertikaalisesti lähetettyä ja vastaanotettua takaisinsirontasignaalia. Käytetty aiemmin turvemaakartoituksessa, esim. Poggio ym. (2019).	SYKE (2021). Indeksimosaiikki tuotettu yhteistyössä SYKE:n ja FGI:n kanssa ja osana raportin työtä.
SAR VH-polarisaatio, S1	VH-polarisaatio tarkoittaa SAR-tutkan vertikaalisesti lähetettyä mutta horisontaalisesti vastaanotettua takaisinsirontasignaalia. Käytetty aiemmin turvemaakartoituksessa, esim. Poggio ym. (2019).	SYKE (2021). Indeksimosaiikki tuotettu yhteistyössä SYKE:n ja FGI:n kanssa ja osana raportin työtä.
SAR Polarisaatiosuhde, S1	Polarisaatiosuhde on laskettu kaavalla $(VH - VV)/(VH + VV)$ SY-KE:n alkupelejä, 11 vuorokauden mosaikeista. Käytetty aiemmin turvemaakartoituksessa, esim. Poggio ym. (2019).	SYKE (2021). Indeksimosaiikki tuotettu yhteistyössä SYKE:n ja FGI:n kanssa ja osana raportin työtä.



#### **2.4.4. Geomorfologiset aineistot**

Ylin ranta (sub- ja supra-akvaattiset alueet) -aineisto kuvaa viimeisimmän Veiksel -mannerjäätikön sulamisen yhteydessä ja sen jälkeen holoseeniaikana vedenpeittämiä ja veden pinnan yläpuolelle jääneitä alueita. Suomen maankamara on ollut laajalti vedenpeittäjä Itämeren eri kehitysvaiheiden aikana (Tikkanen & Oksanen 2002). Supra-akvaattiset alueet ovat vedenpinnan tason yläpuolella. Vastaavasti ylimmän rannan alapuoliset alueet ovat nk. sub-akvaattisia alueita eli veden peittämiä. Ylin ranta sijaitsee Pohjois-Suomessa noin 220 m mpy ja Kaakkois-Suomessa noin 100 m mpy (Tikkanen & Oksanen 2002).

Turpeen muodostuminen alkoi tapahtua ensin supra-akvaattisilla alueilla oleviin painanteisiin ja pohjamaaltaan huonosti vettä läpäiseville maille. Seuraavaksi turvetta alkoi kerrostua veden alta paljastuneille tasaisille tai allasmaisille sub-akvaattisille maille eli soistumiselle alueille (Hokkanen 2005, Mäkilä ym. 2013). Suomen vanhin turvekerros (ikä 11 140–11 570 v) on löydetty Ilomantsin Parkusuolta, joka sijaitsee 142 m mpy ja on Baltian jääjärven tason yläpuolella supra-akvaattisella alueella (Mäkilä ym. 2013).

Ylin ranta (sub-/supra-akvaattinen) aluerasteri on muodostettu mallinnusta varten ylimmän rannan tasoa (highest shoreline) kuvaavasta vektorimuotoisesta tasosta, jonka lähtötietoina on käytetty kirjallisuudesta ja maaperäkartoituksessa kerättyä rantapintahavaintoaineistoa (Åberg, 2013). Tästä havaintoaineistosta on interpoloitu ylimmän rannan taso, josta on leikattu Maanmittauslaitoksen 10 m pikselikoon korkeusmallilla ylimmän rannan alueet (GTK 2013, Åberg 2013). Perushilaan istutettu sub- ja supra-akvaattinen luokittelurasteri (rasterisolukoko 50 m) luotiin ylimmän rannan aineistosta ArcGIS Pro –paikkatieto-ohjelmalla.

**Taulukko 4.** Tiivistelmä kaukokartoitus- ja paikkatietoaineistoista, joilla turpeen esiintymistä ja paksuutta pyrittiin selittämään. Taulukko jatkuu seuraavalla sivulla.

Aineiston nimi	Aineiston kuvaus	Lähde
<b>Geofysikaaliset matalalentomittausaineistot</b>		
Kalium-säteily (skaalattu)	Gammasäteilymittauksissa havaittu Kalium-säteily, jossa huomioitu kallioperän alueellisesta vaihtelusta johtuva säteilytasojen vaihtelu.	GTK, Airo (2005). Skaalaus tehty osana tämän raportin työtä.
Torium-säteily	Gammasäteilymittauksissa havaittu Torium-säteily.	GTK, Airo (2005)
Kalium- ja Torium -säteilyn suhde	Gammasäteilymittauksissa havaitun Kalium- ja Torium-säteilyn suhde.	GTK, Airo (2005)
Säteilyannosnopeus	Gammasäteilymittauksissa havaituista Kalium-, Torium- ja uraanisäteilystä laskettu säteilyannosnopeus.	GTK, Airo (2005), IAEA (1991)
Sähkönjohtavuus (reaali)	Maankamaran sähkömagneettisen kentän mittauksen reaalikomponentti.	GTK, Airo (2005)
Sähkönjohtavuus (imaginääri)	Maankamaran sähkömagneettisen kentän mittauksen imaginäärikomponentti.	GTK, Airo (2005)
Sähkönjohtavuus (reaali-imaginääri -suhde)	Maankamaran sähkömagneettisen kentän mittauksen reaali- ja imaginäärikomponenttien suhde.	GTK, Airo (2005)
<b>Korkeusmalliin perustuvat aineistot</b>		
Depth-to-Water (DTW)	Maaperän kosteutta kuvaava korkeusmallista laskettu topografinen indeksi kynnyksiarvoilla: 1 ha, 2 ha, 4 ha ja 10 ha. Yleistetty 50 m pikselikokoon keskiarvoina.	LUKE, Salmivaara (2020). Täydennetty osana tämän raportin työtä.
Maanpinnan rosaisuus	Korkeusmallista laskettu maaston monimittakaavainen topografinen tekstuuri -aineisto. Sisältää neljä tekstuuriin liittyvää tunnuslukua, jotka on laskettu tietyillä ikkunakoilla ja yleistetty 50 m pikselikokoon keskiarvoina ja -hajontoina: 1 = Pinta-alasuhde (3x3, keskiarvo) 2 = Korkeusarvojen vaihtelevuus (31x31, keskiarvo) 3 = Rinnekaltevuuden vaihtelevuus (31x31, keskiarvo) 4-6 = Rinnesuunnan vaihtelevuus (5x5, 11x11, 31x31, keskiarvo) 7 = Pinta-alasuhde (3x3, keskihajonta) 8 = Korkeusarvojen vaihtelevuus (31x31, keskihajonta) 9 = Rinnekaltevuuden vaihtelevuus (31x31, keskihajonta) 10-12 = Rinnesuunnan vaihtelevuus (5x5, 11x11, 31x31, keskihajonta)	LUKE, Tuotettu osana tämän raportin työtä.
<b>Satelliittiaineistot</b>		
Normalised Difference Moisture Index (NDMI)	Kosteusindeksi, joka kuvaa kasvillisuuden kosteutta (Gao, 1996).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
Normalised Difference Vegetation Index (NDVI)	Kasvillisuusindeksi, joka kuvaa vihreän kasvillisuuden määrää (Rouse ym., 1973).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
Enhanced Vegetation Index 2 (EVI2)	Kasvillisuusindeksi, joka reagoi muutoksiin maaperän kosteudessa (Jiang ym., 2008).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
NBR (nNormalized Burn Ratio)	Paloindeksi, joka kuvaa palaneita alueita maastossa (Key & Benson N, 1999).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
Kernel Normalised Difference Vegetation Index (kNDVI)	Matemaattisesti paranneltu versio NDVI:stä (Camps-Valls ym., 2021).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
Normalized Difference Snow Index (NDSI)	Lumi-indeksi, joka reagoi lumeen (Riggs ym., 1994).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä

Aineiston nimi	Aineiston kuvaus	Lähde
<b>Satelliittiaineistot</b>		
Soil-Adjusted Vegetation Index (SAVI)	Kasvillisuusindeksi, joka huomioi kasvillisuuden lisäksi myös maaperän vaikutuksen heijastavasteeseen (Huete, 1988).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
Soil-color Index (SCI)	Maaperän väri-indeksi ja kuvaa maaperän väriä (esim. Poggio ym., 2019).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
Soil Moisture (SM)	Maaperän kosteusindeksi, joka kuvaa maaperän kosteutta (Dupigny-Giroux & Lewis, 1999)	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
Tasseled-Cap Transformation Brightness (TCTB)	Kirkkausindeksi, joka on yksi osa kolmen osa-alueen TCT- kokonaisuutta: kirkkaus, vihreys ja kosteus (Kauth & Thomas, 1976; Shi & Xu, 2019).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
Tasseled-Cap Transformation Greenness (TCTG)	Kasvillisuusindeksi, joka on yksi osa kolmen osa-alueen TCT- kokonaisuutta: kirkkaus, vihreys ja kosteus (Kauth & Thomas, 1976; Shi & Xu, 2019).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
Tasseled-Cap Transformation Wetness (TCTW)	Kosteusindeksi, joka on yksi osa kolmen osa-alueen TCT- kokonaisuutta: kirkkaus, vihreys ja kosteus (Kauth & Thomas, 1976; Shi & Xu, 2019).	FGI, ESA (2023b). Tuotettu osana tässä raportissa esitettyä työtä
SAR VV-polarisaatio	VV-polarisaatio tarkoittaa SAR-tutkan vertikaalisesti lähetettyä ja vastaanotettua takaisinsirontasignaalia. Kukin mosaikki sisältää pikselikohtaiset keskiarvot noin 7 viikon ajalta. Yhteensä neljä mosaikkia vuodelta 2021 huhtikuun alusta lokakuun loppuun. (SYKE, 2021)	FGI, SYKE (2021), ESA (2023a). Tuotettu osittain osana tässä raportissa esitettyä työtä.
SAR VH-polarisaatio	VH-polarisaatio tarkoittaa SAR-tutkan vertikaalisesti lähetettyä mutta horisontaalisesti vastaanotettua takaisinsirontasignaalia. Kukin mosaikki sisältää pikselikohtaiset keskiarvot noin 7 viikon ajalta. Yhteensä neljä mosaikkia vuodelta 2021 huhtikuun alusta lokakuun loppuun.	FGI, SYKE (2021), ESA (2023a). Tuotettu osittain osana tässä raportissa esitettyä työtä.
SAR-polarisaatio-suhde	SAR-tutkan polarisaatiosuhde kaavalla $(VH - VV) / (VH + VV)$ SYKEN alkuperäisistä 11 vuorokauden mosaikkeista laskettuna. Kukin mosaikki sisältää pikselikohtaiset keskiarvot noin 7 viikon ajalta. Yhteensä neljä mosaikkia vuodelta 2021 huhtikuun alusta lokakuun loppuun. (SYKE, 2021)	FGI, SYKE (2021), ESA (2023a). Tuotettu osittain osana tässä raportissa esitettyä työtä.
<b>Geomorfologiset aineistot</b>		
Ylin ranta (sub- ja supra-akvaattiset alueet)	Jääkauden jälkeen vedenpinnan ylä- ja alapuolelle jääneet alueet	GTK, Tikkanen ja Oksanen (2002), Åberg (2013)

## 2.5. Mallinnusmenetelmä

### 2.5.1. Random Forest (RF)

Random Forest (RF, suom. satunnainen metsä; Hastie ym. 2001) on koneoppimisessa ja digitaalisten maalajikarttojen tuotannossa yleisesti käytetty luokittelu- ja regressiomenetelmä (Wadoux ym. 2020). Se on myös usein käytetty koneoppimismenetelmä turvemaiden esiintymisen ennustamiseen (Minasny ym. 2019). RF-koneoppimismenetelmässä rakennetaan selittävien tekijöiden perusteella suuri joukko ns. päätöspuita, eli loogisia päätösketjuja selittävien tekijöiden arvojen perusteella. Kyseisten päätöspuiden loogiset päätösketjut rakennetaan niin, että havaittu opetusaineisto onnistutaan erottelemaan oikein vastaaviin luokitteluryhmiin (eng. class labels). Yksittäinen päätöspuu ylisovittuu herkästi opetusaineistoon. Tämä

tarkoittaa, että päätöspuu oppii näkemänsä datan ulkoa, mutta yleistyy uusiin, erilaisiin datapisteisiin huonosti. Random Forest -menetelmässä yhdistetään useita päätöspuita metsäksi. Metsän päätöspuut rakennetaan tekemällä satunnaisia otantoja alkuperäisestä datasta ja opettamalla ko. otannoilla päätöspuita. Verrattuna yksittäiseen päätöspuuhun, metsän herkkyys ylisovittamiselle on pienempi ja luokittelun tarkkuus parempi.

### **2.5.2. Geneettinen algoritmi (GA)**

Optimaalisten selittävien tekijöiden eli piirteiden valinta toteutettiin geneettisellä algoritmilla (GA, engl. genetic algorithm; Melanie 1998), jossa evolutiivissatunnaisella tavalla etsitään optimaalisia piirrekombinaatioita koneoppimismallille, kunnes haun konvergenssi saavutetaan. GA-menetelmän idea on, että hyvin pärjäävät piirrekombinaatiot ja näiden kombinaatiot "selviytyvät" seuraaville sukupolville iteraatio iteraatiolta, kunnes paras mahdollinen kombinaatio on löydetty. Piirrekombinaatioiden soveltuvuutta mitataan suorituskykyometriikoilla kuten yleisellä luokittelutarkkuudella, ja parhaiten soveltuvat piirrekombinaatiot sekä niiden yhdistelmät jatkavat seuraavalle sukupolvi-iteraatiolle suuremmalla todennäköisyydellä. Myös GA-menetelmässä on sopivasti huolehdittava ylisovituksesta. Menetelmän avulla voidaan löytää piirrekombinaatioita, jotka sopivat hyvin ainoastaan kalibrointidataan, mutta yleistyvät huonosti uuteen aineistoon, jos sukupolvi-iteraatioita on liian paljon.

### **2.5.3. Permutaatiotärkeys (PI)**

Permutaatiotärkeys (PI, engl. permutation feature importance; Breiman, 2001) on menetelmä selittävien tekijöiden merkittävyyden mittaamiseen mallinnustehtävässä. PI-menetelmässä mitataan koneoppimismallin suorituskyvyn muutosta, kun tutkittavaa selittävää tekijää, eli piirrettä, sekoitetaan samalla kun muut piirteet pidetään ennallaan. Sekoittaminen tarkoittaa kyseisen piirteen arvojen satunnaista permutointia. Tutkittava piirre tulkitaan tärkeäksi, jos sen sekoituksen seurauksena koneoppimismallin suorituskyky huononee selvästi. Vastaavasti, jos koneoppimismallin suorituskyky pysyy samana, kyseinen piirre tulkitaan ei-tärkeäksi piirteeksi. Piirteen sekoittaminen ja vastaava koneoppimismallin suorituskyvyn muutoksen mittaaminen suoritetaan useita kertoja Monte Carlo-tyyliin ja piirteen permutaatiotärkeydeksi ilmoitetaan suorituskykymittausten keskiarvo ja -hajonta.

### **2.5.4. Ositettu ristiinvalidointi**

Ositettu ristiinvalidointi (eng. stratified cross-validation, esim. Hastie ym. 2001) on menetelmä, jota käytetään koneoppimismallien suorituskyvyn arvioimiseen. Ositettu ristiinvalidointi auttaa välttämään tilanteita, joissa jokin harvinaisempi vastemuuttujaluokka jää kokonaan pois validointivaiheesta, mikä voi johtaa vääristyneisiin suorituskyvyn arvioihin. Tämä menetelmä on erityisen tärkeä, kun käsitellään luokitteluongelmia, joissa luokkien jakauma ei ole tasapainoinen. Normaalin ristiinvalidoinnin tapaan ositettu ristiinvalidointi toteuttaa useita opetusjoukko-validointijoukko osajakoja, mutta osajaoissa otetaan huomioon alkuperäisen aineiston vastemuuttujan luokkajakauma.

### 2.5.5. Metriikat

Koneoppimismallien suorituskykyä mitattiin useilla eri metriikoilla, jotka esitellään seuraavissa kappaleissa.

Yleinen tarkkuus (Overall accuracy), määritellään koneoppimismallin oikein luokiteltujen vastemuuttujahavaintojen suhteena kaikkien havaintojen lukumäärään. Yleinen tarkkuus on siis prosentuaalinen osuus validointipisteistä, jotka koneoppimismalli luokitteli oikein.

Tuottajan tarkkuus (Producer's accuracy) määritellään sinä prosentuaalisena osuutena luokan  $k$  todellisista havainnoista, jotka koneoppimismalli onnistui luokittelemaan oikein. Tuottajan tarkkuus lasketaan kaikille aineiston vastemuuttujaluokille erikseen. Toisin sanoen, annettuna luokka  $k$ , tuottajan tarkkuus mittaa, kuinka hyvin koneoppimismalli onnistui ennustamaan todelliset luokan  $k$  havainnot.

Käyttäjän tarkkuus (User's accuracy) määritellään sinä prosentuaalisena osuutena kaikista koneoppimismallin tekemistä luokan  $k$  ennustuksista, jotka koneoppimismalli luokitteli oikein. Tuottajan tarkkuuden tavoin, käyttäjän tarkkuus lasketaan jokaiselle vastemuuttujaluokalle  $k$  erikseen. Käyttäjän tarkkuus siis mittaa todennäköisyyttä sille, kuinka usein koneoppimismallin käyttäjä voi odottaa luokan  $k$  ennusteiden olevan oikein.

F1-score on binäärisen luokittelumallin mittari, jota käytetään arvioimaan mallin tarkkuutta ottamalla huomioon niin tarkkuus (precision) kuin palautus (recall) samanaikaisesti. F1-score tarjoaa tasapainoisen arvion mallin suorituskyvystä, kun on tarpeen ottaa huomioon väärin positiivisten ja väärin negatiivisten ennusteiden vaikutukset. F1-score saa arvonsa 0 ja 1 välistä, ja arvon 1 tapauksessa luokittelumalli ei tee yhtään väärää luokitusta.

Kappa-kerroin (Cohen's Kappa) on tilastollinen mittari, jota käytetään arvioimaan kahden eri "tarkastelijan" välistä yksimielisyyttä (arvojen 0 ja 1 välistä), kun nämä luokittelevat tai arvioivat jotain. Tarkastelijoilla tässä tarkoitetaan koneoppimismallin ennustuksia ja todellista mitattua havaintoaineistoa. Kappa-kerroin on suunniteltu ottamaan huomioon (normalisoimalla) sattuman vaikutuksen tarkastelijoiden yksimielisyyteen, mikä tekee siitä hyödyllisen mittarin erityisesti silloin, kun luokkien jakauma voi vaikuttaa arviointiin. Jos sattuman vaikutusta aineiston jakauman takia ei oteta huomioon, koneoppimismallin suorituskyky voi vaikuttaa harhallisesti korkeammalta kuin se todellisuudessa on.

ROC-käyrä on graafinen esitys binäärisen luokittelijan suorituskyvystä eri kynnyksarvoilla. Kynnyksarvo kuvaa sitä todennäköisyysarvoa, jonka ylittäessä koneoppimismallin ennuste luokitellaan positiiviseksi luokaksi. ROC-käyrä esittää väärin ja oikeiden positiivisten havaintoprosenttien (eng. false and true positive rate, FPR, TPR) kuvaajan. ROC-käyrän avulla voi tulkita, miten TPR-FPR-suhde kehittyy eri ennustustodennäköisyyksien kynnyksarvoilla. Area under curve (AUC) on arvo 0 ja 1 välistä, joka kuvaa ROC-käyrän alla olevaa pinta-alaa. AUC-arvo 0,5 kertoo luokittelun ennustavan luokan oikein 50 % tapauksista. AUC arvo kuvaa todennäköisyyttä sille, että binääriluokittelija osaa luokitella negatiivisen ja positiivisen luokkahavainnon oikein. Mitä isompi AUC, sitä suurempi TPR ja pienempi FPR luokittelijalla on kaikilla mahdollisilla kynnyksarvoilla. AUC on myös ekvivalentti concordance index (c-index) -metriikan kanssa, jota voidaan hyödyntää AUC arvon laskentaan.

### 2.5.6. Mallinnusmenetelmien soveltaminen

RF-mallinnuksessa suoritettiin ensimmäiseksi piirrevalinta (turpeen esiintymistä selittävät aineistot), piirrevalintamallien suorituskyvyn mittaaminen ja ennustemallin kalibrointi (hyperparametrien valinta). Piirrevalinta tehtiin GA-menetelmän avulla, jossa hyödynnettiin populaatiokokoa 100, 10 % elitismiä ja 1 % mutaation todennäköisyyttä. Piirrevalintamallien suorituskyvyn mittaaminen ja ennustemallin kalibrointi (hyperparametrien valinta) toteutettiin ositetun 3-jakoisen ristiinvalidoinnin avulla. Mallinnuksessa hyödynnettiin kiinnitettyä 100 päätöspuun yhdistelmää, joten mallille ei suoritettu päätöspuulukumäärän kalibrointia. Malli optimoitiin yleisen tarkkuuden perusteella ja optimaalisille malleille tehtiin suorituskyvyn mittaaminen hyödyntäen ositettua ristiinvalidointia ja kappaleessa 2.5.5 esitettyjä metriikoita. Lopullinen mallien suorituskyvyn arviointi toteutettiin riippumattomalla ns. "out-of-sample"-testijoukon avulla (Kappale 2.3.4), joka tarjoaa harhattoman estimaatin mallien suorituskyvystä uusissa ennustetilanteissa. Ennustemallien kalibroinnin ja piirrevalinnan jälkeen sovitettiin lopulliset ennustemallit koko kalibrointiaineistoa hyödyntäen ja toteutettiin 50 metrin spatiaalisen rasteriresoluution ennustus koko maan yli.

Mallinnus toistettiin neljä kertaa, jonka tuloksena syntyivät rasterimuotoiset todennäköisyysennusteet (0-1)  $\geq 10$  cm (RF10),  $\geq 30$  cm (RF30),  $\geq 40$  cm (RF40) ja  $> 60$  cm (RF60) paksuille turvekerroksille, sekä näistä todennäköisyyden raja-arvoilla luokitellut binääriennusteet (0/1). Todennäköisyysennusteiden rasterien solut on määritelty vastaavien RF-mallien kaikkien 100 päätöspuun keskiarvoisena todennäköisyytenä turpeen esiintyvyydelle. Ennustetodennäköisyyksien raja-arvot määritettiin laskennallisesti binääristereille ROC-käyrien taitekohtaan sinä raja-arvona, joka minimoi etäisyyden ROC-käyrän (TPR=1, FPR=0) -pisteeseen. Toisin sanoen, laskennallinen raja-arvo pyrkii samanaikaisesti maksimoimaan oikeat ennusteet ja minimoimaan väärät ennusteet (TPR- ja FPR-arvot) vastaavasti.

Mallinnus tuotti myös päätöspuiden yksimielisyyskartat kullekin ennustekartalle. Ne kertovat vastaavan RF-mallin maksimimäärän ennusteesta samaa mieltä olevista päätöspuista (100 kpl) kussakin ennustesolussa. Jos esimerkiksi tietyssä ennusteessa 100 päätöspuusta 40 ennusti luokkaa ei-turvetta (luokka 0) ja 60 taas turvetta (luokka 1), niin tällöin kyseisen ennusteen yksimielisyysarvo on 60. Tämän voidaan ajatella kuvaavan ennusteen epävarmuutta mallin näkökulmasta. Ennusteen epävarmuus on suurin pienillä yksimielisyyden arvoilla (pätöspuut eri mieltä ennusteesta) ja pienin yksimielisyyden suurilla arvoilla (pätöspuut samaa mieltä ennusteesta). Toisin sanoen, yksimielisyyden pienillä arvoilla RF-mallin päätöspuiden luokkaennusteet jakautuvat lähes tasaisesti histogrammissa tarkasteltuna (esim. 50–50), ja päinvastoin (esim. 0–100 tai 100–0). Pienet yksimielisyyden arvot vastaavat tilannetta, jossa mallin ennusteen entropia, eli epävarmuus on suurin.

Ennusteiden yksimielisyyden ja vastaavien ennustetarkkuuksien välistä korrelaatiota tutkittiin myös. Laskennallinen vertaus toteutettiin ristiinvalidointitulosten perusteella, jossa jokaisella yksimielisyysarvolla (50–100) verrattiin RF-mallin tekemiä vastaavien ennusteita oikeisiin havaintoihin. Eli esimerkiksi yksimielisyyden 50 tapauksessa, kerättiin kaikki ennusteet, joille RF-malli ennusteen yksimielisyys oli 50 ja verrattiin vastaavia ennusteita oikeisiin havaintoihin tarkkuusmittarin (eng. accuracy) avulla.

## 2.6. Turpeen esiintyminen maatalousmailla

Turpeen esiintymistä maatalousmailla tarkasteltiin käyttäen RF30 ( $\geq 30$  cm) ja RF60 ( $> 60$  cm) ennusteita, sillä ne vastaavat aiemmissa tarkasteluissa käytettyjä turvepaksuuksia, jotka ovat perustuneet Maaperä 1:200 000- (GTK 2010) ja siitä johdettuun Maannostietokanta-aineistoihin (Lilja ym. 2017). Maatalousmaa määriteltiin Ruokaviraston peltolohkokisterin peruslohkojen mukaan (Ruokavirasto, 2021), jossa peruslohkot on kuvattu vektorimuotoisina alueina. Tarkastelu toteutettiin laskemalla turpeen esiintymisen tilastoja peruslohko- ja maakuntatasoilla. Tulosten tarkastelu -osiossa RF30-binääriennustetta verrattiin myös Maaperä 1:200 000-aineistoon, jota varten vektorimuotoinen Maaperä 1:200 000-aineisto muunnettiin samaan rasteriresoluutioon RF30-binääriennusteen kanssa.

## 2.7. Turvepeltolohkojen tunnistaminen

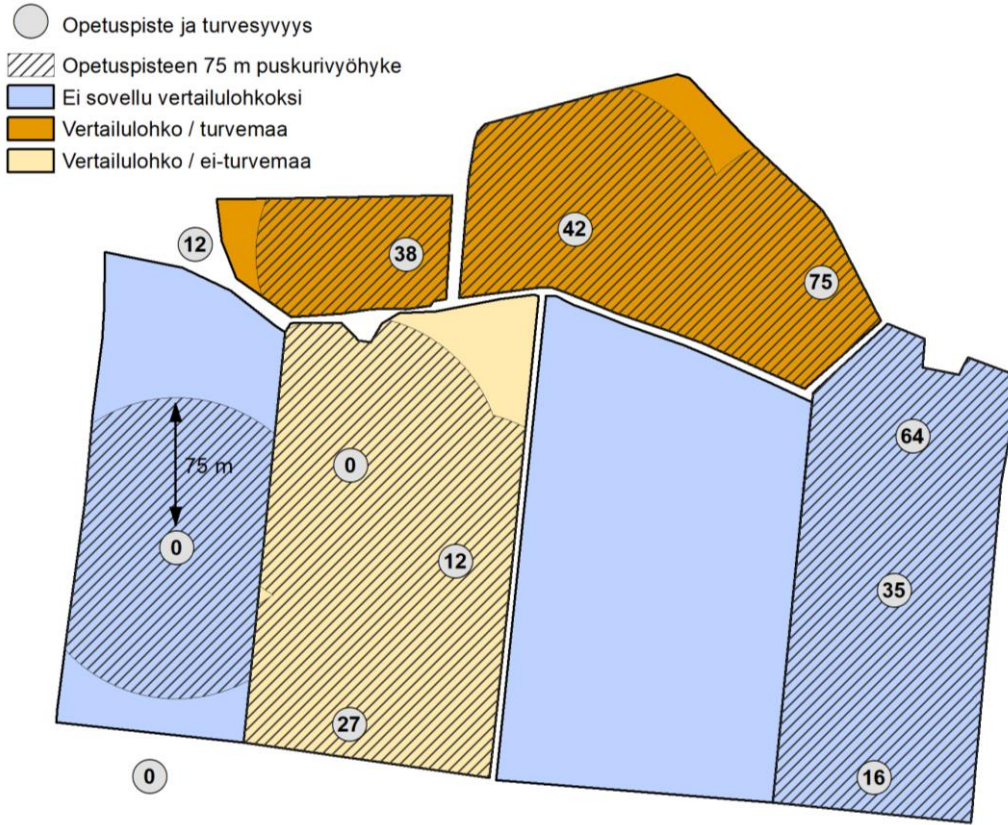
Mallinnustulosten soveltuvuutta ja tarkkuutta arvioitiin kokonaisten peltolohkojen tasolla. Tässä yhteydessä tarkasteltiin pääsääntöisesti sitä, millaisia tuloksia saadaan, jos rasterimuotoiset mallinnustulokset yleistetään aluemuotoisten kohteiden (peltolohkot) luokituksiksi. Mallinnustuloksista tarkastelussa hyödynnettiin RF30- ja RF60-todennäköisyysennusteita, joista ensimmäisen perusteella pyrittiin erottamaan turvelohkot ei-turvelohkoista, ja jälkimmäisen perusteella paksuturpeiset lohkot muista lohkoista. Turvepeltolohkona pidettiin sellaista lohkoa, jonka alasta vähintään puolet on 30 cm tai paksumpaa turvetta, ja paksuturpeisenä vastaavasti lohkoa, jonka alasta vähintään puolet on yli 60 cm syvää turvetta. Peltolohkot määriteltiin Ruokaviraston peltolohkokisterin peruslohkojen mukaisesti.

Tunnistustarkkuuden arviointia varten erotettiin ensin turvesyvyyksiltään tunnettu vertailulohkojen aineisto, joka perustui testausosittien (ks. Kappale 2.3.4) alueella oleviin peruslohkoihin (Ruokavirasto 2021). Vertailulohkojen turvesyvyyden luokitus tehtiin testausaineiston opetuspisteitä hyödyntäen. Tähän liittyi kuitenkin epävarmuus pistehavaintojen yleistettävyydestä aluemuotoisille kohteille. Lähtökohdaksi otettiin se, että ainoastaan tarkasteltavan peltolohkon sisällä olevia pistehavaintoja voitiin pitää lohkon kannalta edustavina, ja silloinkin yhden opetuspisteen tietosisältö ulottui vain rajoitetulle alueelle. Tarkoitukseen soveltuvien vertailulohkojen valinta peruslohkoista tehtiin seuraavien sääntöjen perusteella (Kuva 4):

1. **Pisteiden lukumäärä:** Tarkasteltavalla peltolohkolla oli vähintään yksi opetuspiste. Lohkon ulkopuoliset (naapurilohkoilla tai maatalousmaan ulkopuolella) olevat pisteet jätettiin huomioimatta, vaikka ne olisivat olleet lähelläkin, koska niiden soveltuvuutta kyseisen lohkon arviointiin ei voitu varmistaa.
2. **Pisteiden spatiaalinen kattavuus:** Lohkon sisällä olevista pisteistä luotu 75 m etäisyydelle ulottuva puskurivyöhyke kattoi vähintään kolme neljäsosaa lohkon pinta-alasta. Valittu 75 metrin säde arvioitiin erillisiin analyyseihin perustuen sellaiseksi etäisyydeksi, jonka sisällä turvesyvyyden oletettiin olevan luokituksen kannalta riittävän lähellä mitattua arvoa. Spatiaalisen kattavuuden ehto tarkoitti myös sitä, että lohkon koosta ja muodosta sekä opetuspisteiden sijainnista riippuen pisteitä tarvittiin usein enemmän kuin yksi.
3. **Pisteiden yhtenäinen tietosisältö:** Lohkon sisällä olevista opetuspisteistä vähintään kolme neljäsosaa kuului samaan tarkastelun kannalta yhtenäiseen syvyyskategoriaan, jonka perusteella lohkon luokitus päätettiin. Turvepeltolohkojen tapauksessa erotettiin ei-turvemaan ( $< 30$  cm turvetta) ja turvemaan ( $\geq 30$  cm) pisteet, ja paksuturpeisten



lohkojen tapauksessa ei-paksuturpeiset (0–60 cm) ja paksuturpeiset (> 60 cm) pisteet. Säännön vuoksi vähemmän kuin neljä opetuspistettä sisältävät lohkot huomioitiin, mikäli kaikki opetuspisteet kuuluivat samaan luokkaan. Sitä useampia pisteitä sisältävillä lohkoilla yksittäiset eri kategorian pisteet voitiin sallia.



**Kuva 4.** Esimerkki turvepeltolohkojen referenssiaineiston määrittelystä ja valituista lohkoista.

Koska opetuspisteiden jakauma painottui vahvasti turvemaille eikä kivennäismaapisteitä ollut generoitu maatalousmaille, saatiin sääntöjen mukaan valittua vain vähän ei-turvemaan lohkoja. Eri luokkien lohkoittaiset yhteispinta-alat kuitenkin pyrittiin luotettavaa tarkkuusarviointia varten saamaan sellaisiksi, että ne vastaisivat RF30- ja RF60-aineistojen perusteella koko maatalousmaalle arvioituja turvesyvyyksien osuuksia. Tämän vuoksi vertailulohkoihin lisättiin sellaisia peruslohkoja, joilla pääsääntöisesti ei ollut opetuspisteitä, mutta ne ympäröivän maanpeitteen ja maannostiedon pohjalta arvioituna olivat hyvin todennäköisesti ei-turve- maata. Lisättäviksi kelpaavien lohkojen valinta tehtiin seuraavia sääntöjä noudattaen:

- Lohkosta 500 m etäisyydellä (laskettuna lähimmän etäisyyden suhteen) ei ollut maastotietokannan (MML 2021) mukaista suo- tai soistuma- aluetta.
- Lohkosta 500 m etäisyydellä ei ollut maannostietokannan (Lilja ym. 2017) mukaista Histosol- eli turvemaannoksen aluetta.
- Lohkosta 250 m etäisyydellä ei ollut sellaista opetuspistettä, jossa oli havaittu turvetta (>0 cm).

Ei-turvemaan lohkoja lisättiin tämän jälkeen vertailulohkoihin satunnaisessa järjestyksessä siihen asti, että pinta-alajakauma turvelohkojen ja muiden lohkojen välillä vastasi koko maan arvioita.

Varsinainen tarkkuusarviointi toteutettiin siten, että RF30- ja RF60-todennäköisyysennusteiden binääriluokittelua testattiin useita eri todennäköisyyden raja-arvoja käyttäen. Kunkin tuloksen osalta laskettiin tämän jälkeen se, mitkä lohkot mallinnustuloksen mukaan ovat turve /ei-turvelohkoja (RF30), tai paksuturpeiksi/ei-paksuturpeisiä lohkoja (RF60). Rajana mallin mukaiselle luokitukselle turve- tai paksuturpeiseksi lohkoksi pidettiin turvemaan 50 % pinta-alaosuutta lohkolla. Mallin antamat pinta-alat laskettiin tarkasti osapikselien tasolla. Tämän jälkeen tuloksia tarkasteltiin oikeiden ja väärin positiivisten ennusteiden osuuksia (TPR ja FPR) hyödyntäen. Niillä ilmaistaan, kuinka suuri osa kaikista todellisista turvelohkoista löydettiin (TPR), ja miten suuri osuus ei-turvemaan lohkoista ennustettiin väärin turvelohkoiksi (FPR).

## 3. Tulokset

### 3.1. Opetus ja testaus

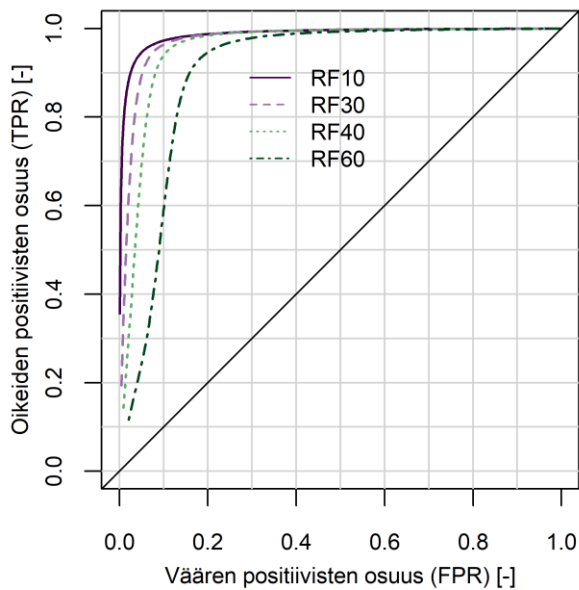
RF-mallin opetuksessa ja testauksessa lasketut yleinen tarkkuus, Kappa, AUC ja F1 -metriikat osoittavat, että mallit ovat onnistuneita, ennusteiden tarkkuus on korkea, ja opetus- ja testausmetriikat vastaavat toisiaan (Taulukko 5). Testauksen mukaan ennusteiden yleinen tarkkuus oli suurin  $\geq 10$  cm (RF10) paksujen turvekohteiden erottelussa kivennäismaakohteista (turvetta  $< 10$  cm) ja nämä kohteet eroteltiin tosistaan turve – ei-turve luokittelussa 96 % tarkkuudella (kappa 0,912). Turvepaksuusluokkien  $\geq 30$  cm (RF30) ja  $\geq 40$  cm (RF40) erottelussa yleinen tarkkuus oli myös korkea ja ne eroteltiin 93–94 % tarkkuudella (kappa 0,885–0,886). Turvepaksuus luokan  $> 60$  cm (RF60) erottelussa yleinen tarkkuus laski hieman ja ne pystyttiin luokittelemaan 89 % tarkkuudella (kappa 0,768).

Testaus osoittaa, että oikeiden positiivisten (TPR) ennusteiden osuus oli korkea ja väärin positiivisten (FPR) ennusteiden osuus pienehkö, mutta väärin positiivisten osuus kasvoi kohti paksumpia turvepaksuuksia (Taulukko 5). Havaituista turvemaista, joiden paksuus oli  $\geq 10$  cm,  $\geq 30$  cm ja  $\geq 40$  cm tunnistettiin oikein paksuusluokkiensa mukaisiksi kohteiksi 94–95 % tapauksista. Havaitut turvemaat, joiden paksuus oli  $> 60$  cm, tunnistettiin hieman heikommin ja ne tunnistettiin oikein paksuusluokkiensa mukaan 88 % tapauksista. Havaituista kivennäismaatapauksista ( $< 10$  cm) ennustettiin väärin turvemaiksi 4 %. Vastaavasti, turvepaksuusluokkiin  $\geq 30$  cm,  $\geq 40$  cm ja  $> 60$  cm kuulumattomia havaintokohteita ennustettiin väärin turvemaiksi 6–10 % tapauksista. RF10, RF30, RF40 ja RF60 -mallien oikeiden ja positiivisten ennusteiden suhteet on esitetty ROC-kuvaajissa Kuva 5.

**Taulukko 5.** RF10, RF30, RF40 ja RF60 -mallien opetus- ja testausmetriikka.

	RF10		RF30		RF40		RF60	
Turvekerroksen paksuusluokka [cm]	$\geq 10$		$\geq 30$		$\geq 40$		$> 60$	
	Opetus	Testaus	Opetus	Testaus	Opetus	Testaus	Opetus	Testaus
Yleinen tarkkuus [-]	0.947	0.956	0.934	0.943	0.917	0.929	0.859	0.891
Tuottajan tarkkuus 0*	0.942	0.960	0.923	0.938	0.903	0.920	0.854	0.896
Tuottajan tarkkuus 1*	0.953	0.952	0.947	0.949	0.936	0.942	0.868	0.882
Käyttäjän tarkkuus 0*	0.956	0.956	0.955	0.958	0.950	0.955	0.919	0.930
Käyttäjän tarkkuus 1*	0.938	0.956	0.910	0.926	0.878	0.897	0.773	0.829
Kappa [-]	0.894	0.912	0.867	0.885	0.832	0.856	0.704	0.768
AUC [-]	0.986	-	0.973	-	0.955	-	0.903	-
F1 [-]	0.947	-	0.934	-	0.917	-	0.861	-
Oikeiden positiivisten osuus [-]	0.954	0.952	0.947	0.949	0.936	0.942	0.868	0.882
Väärin positiivisten osuus [-]	0.060	0.040	0.077	0.062	0.097	0.080	0.146	0.104

\*0 = ei kuulu turvepaksuusluokkaan, 1 = kuuluu turvepaksuusluokkaan.



**Kuva 5.** Luokiteltujen RF10, RF30, RF40 ja RF60-ennusteiden ROC-kuvaajat.

### 3.2. Selittävät aineistot

RF10, RF30, RF40 ja RF60 -malleihin valikoitui laaja määrä aineistoja ennustamaan turpeen esiintymistä ja paksuutta, jotka on esitetty Taulukko 6, Taulukko 7, Taulukko 8 ja Taulukko 9. Parhaiksi selittäviksi aineistoiksi valikoituivat Kalium-säteily, korkeusarvojen vaihtelevuus, rinnen suunnan vaihtelevuus, DTW-indeksi 2 ha kynnyksarvolla ja satelliittiaineistoista NDSI ja SCI, NDVI ja NBR. Mukana oli myös muita Sentinel-2 aineistosta laskettuja indeksejä ja Sentinel-1 intensiteetti- ja polarisaatiosuhdemuuttujia, mutta optisen alueen aineistot olivat keskimäärin merkittävämpi kuin tutka-alueen aineistot.

Malleihin valikoidut aineistot vaihtelivat jonkin verran turvepaksuuden mukaan. DTW oli korkeassa roolissa RF10, RF30, RF40 -malleissa, mutta se ei valikoitunut mukaan RF60-malliin. Sen sijaan Torium-säteily nousi korkeaan rooliin RF60 mallissa. Satelliittiaineistojen rooli kasvoi myös paksumpia turveluokkia ennustettaessa. Erityisesti kNDVI ja TCTW:n nousivat korkeaan rooliin satelliittiaineistojen osalta RF60-mallissa. Satelliittiaineistojen järjestys vaihteli kuitenkin paljon eri mallien välillä. Ylin ranta-aineisto valikoitui myös RF10, RF30, RF40 -malleihin, mutta ei RF60-malliin.

**Taulukko 6.** RF10-malliin valikoituneet turpeen esiintymistä ( $\geq 10$  cm) selittävät aineistot ja näiden permutaatiotärkeys.

Nro.	Aineisto	Permutaatiotärkeys
1	Kalium-säteily	4.98
2	DTW	2.93
3	Rinnesuunnan vaihtelevuus (5×5, keskiarvo)	2.869
4	Korkeusarvojen vaihtelevuus (31×31, keskiarvo)	1.77
5	Rinnekaltevuuden vaihtelevuus (31×31, keskiarvo)	0.521
6	Pinta-alasuhde (3×3, keskihajonta)	0.492
7	Rinnesuunnan vaihtelevuus, (31×31, keskiarvo)	0.329
8	Korkeusarvojen vaihtelevuus (31×31, keskihajonta)	0.219
9	Rinnesuunnan vaihtelevuus (5×5, keskihajonta)	0.16
10	NDSI 20210601-20210630	0.147
11	SCI 20210615-20210715	0.141
12	NDSI 20210615-20210715	0.131
13	NDVI 20210615-20210715	0.111
14	NBR 20210401-20210430	0.105
15	Ylin ranta	0.081
16	SCI 20210715-20210815	0.08
17	EVI2 20210515-20210615	0.077
18	NDVI 20210501-20210531	0.077
19	SAR-polarisaatiosuhde 20210521-20210710	0.047
20	SAR VV -polarisaatio 20210901-20211031	0.045
21	SAR VV -polarisaatio 20210401-20210520	0.044
22	SCI 20210401-20210430	0.031
23	NDSI 20210815-20210915	0.03
24	SAR-polarisaatiosuhde 20210401-20210520	0.021
25	TCTB 20210901-20210930	0.017
26	Kalium- ja Torium-säteilyn suhde	0.003

**Taulukko 7.** RF30-malliin valikoituneet turpeen paksuutta  $\geq 30$  cm selittävät aineistot ja niiden permutaatiotärkeys.

<b>Nro.</b>	<b>Aineisto</b>	<b>Permutaatiotärkeys</b>
1	Kalium-säteily	4.549
2	DTW	1.942
3	Rinnesuunnan vaihtelevuus (5x5, keskiarvo)	1.865
4	Korkeusarvojen vaihtelevuus (31x31, keskiarvo)	1.257
5	Rinnekaltevuuden vaihtelevuus (31x31, keskiarvo)	0.327
6	Pinta-alasuhde (3x3, keskihajonta)	0.261
7	Korkeusarvojen vaihtelevuus (31x31, keskihajonta)	0.116
8	Sähkönjohtavuus (imaginääri)	0.097
9	NDSI 20210601-20210630	0.097
10	NDSI 20210615-20210715	0.088
11	Rinnesuunnan vaihtelevuus (31x31, keskiarvo)	0.085
12	SAVI 20210701-20210731	0.077
13	Ylin ranta	0.074
14	SCI 20210615-20210715	0.071
15	NBR 20210401-20210430	0.068
16	TCTW 20210415-20210515	0.06
17	kNDVI 20210601-20210630	0.06
18	NDVI 20210501-20210531	0.057
19	NBR 20210601-20210630	0.057
20	Pinta-alasuhde (3x3, keskiarvo)	0.048
21	NDSI 20210915-20211015	0.043
22	SAR VV-polarisaatio 20210711-20210831	0.038
23	SM 20210615-20210715	0.038
24	Rinnesuunnan vaihtelevuus (31x31, keskihajonta)	0.036
25	TCTW 20210501-20210531	0.033
26	NDVI 20210901-20210930	0.031
27	Rinnesuunnan vaihtelevuus (11x11, keskiarvo)	0.029
28	TCTW 20210815-20210915	0.024
29	SAR VV-polarisaatio 20210401-20210520	0.02
30	SCI 20210501-20210531	0.014
31	NDSI 20211001-20211031	0.013
32	EVI2 20210415-20210515	0.008
33	EVI2 20211001-20211031	0.004

**Taulukko 8.** RF40-malliin valikoituneet turpeen paksuutta  $\geq 40$  cm selittävät aineistot ja niiden permutaatiotärkeys.

<b>Nro.</b>	<b>Aineisto</b>	<b>Permutaatiotärkeys</b>
1	Kalium-säteily	5.562
2	Rinnesuunnan vaihtelevuus (5x5, keskiarvo)	2.272
3	DTW	1.732
4	Korkeusarvojen vaihtelevuus (31x31, keskiarvo)	1.639
5	Rinnekaltevuuden vaihtelevuus (31x31, keskiarvo)	0.891
6	NDVI 20210501-20210531	0.177
7	NBR 20210601-20210630	0.101
8	Korkeusarvojen vaihtelevuus (31x31, keskihajonta)	0.098
9	SCI 20210615-20210715	0.085
10	NBR 20210415-20210515	0.083
11	SCI 20210401-20210430	0.075
12	TCTW 20210401-20210430	0.063
13	TCTW 20210501-20210531	0.055
14	NDSI 20210601-20210630	0.052
15	NDSI 20210615-20210715	0.048
16	Ylin ranta	0.048
17	Rinnekaltevuuden vaihtelevuus (31x31, keskihajonta)	0.047
18	kNDVI 20210601-20210630	0.042
19	SAR VV-polarisaatio 20210521-20210710	0.039
20	NDSI 20210915-20211015	0.038
21	TCTW 20210901-20210930	0.036
22	NDSI 20210701-20210731	0.025
23	kNDVI 20210715-20210815	0.025
24	SCI 20210515-20210615	0.021
25	SAR VV-polarisaatio 20210711-20210831	0.020
26	TCTG 20210901-20210930	0.003
27	SM 20210615-20210715	<0.001

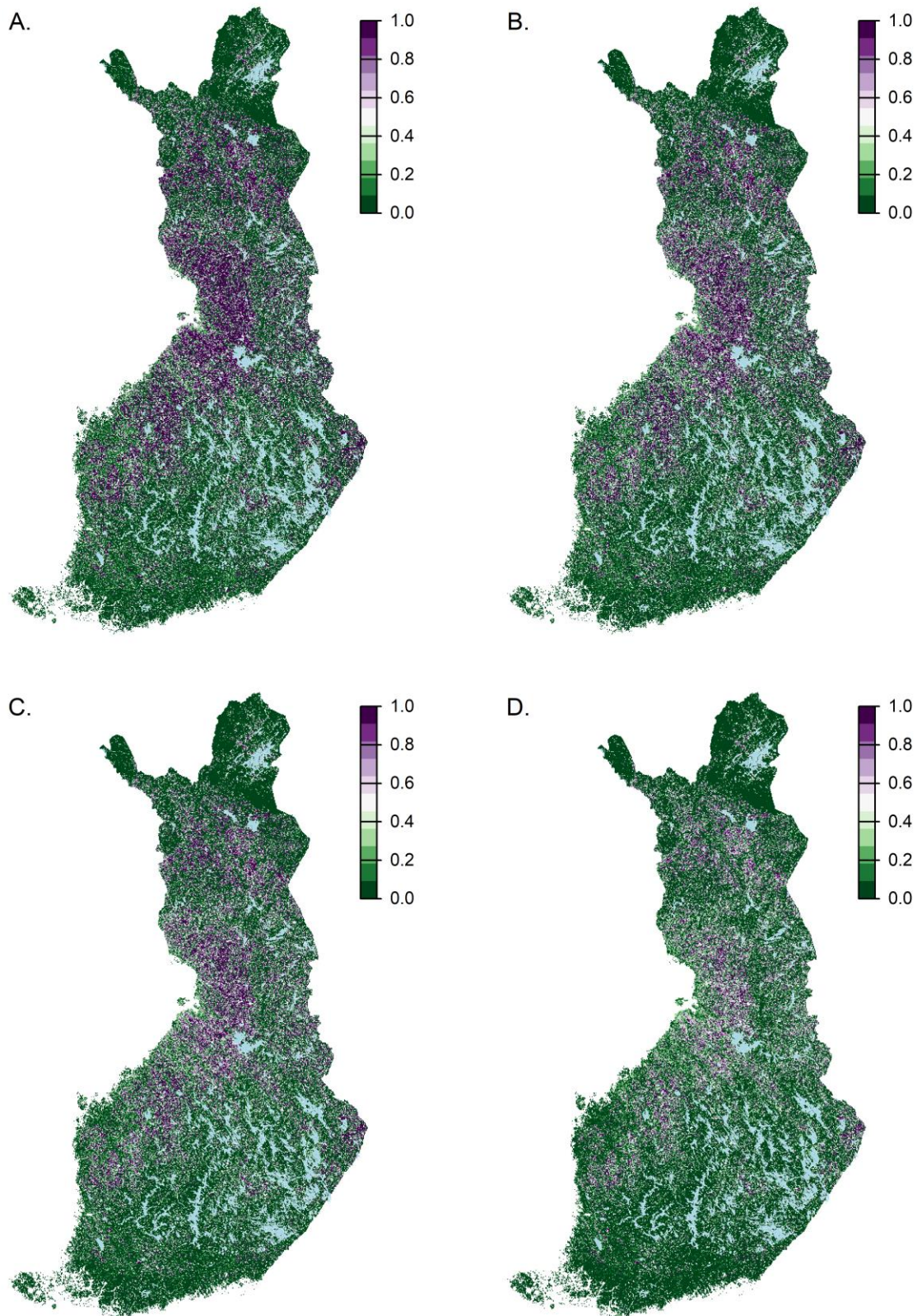


**Taulukko 9.** RF60-malliin valikoituneet turpeen paksuutta >60 cm selittävät aineistot ja näiden permutaatiotärkeys.

Nro.	Aineisto	Permutaatiotärkeys
1	Korkeusarvojen vaihtelevuus (31x31, keskiarvo)	4.832
2	Kalium-säteily	4.375
3	Thorium-säteily	3.500
4	Rinnesuunnan vaihtelevuus (11x11, keskiarvo)	0.565
5	Korkeusarvojen vaihtelevuus (31x31, keskihajonta)	0.371
6	SCI 20210401-20210430	0.260
7	Rinnekaltevuuden vaihtelevuus (31x31, keskihajonta)	0.223
8	kNDVI20210601-20210630	0.184
9	TCTW 20210401-20210430	0.146
10	TCTW 20210415-20210515	0.144
11	SCI 20210615-20210715	0.136
12	NDSI 20210701-20210731	0.126
13	EVI2 20210501-20210531	0.111
14	NDSI 20210715-20210815	0.109
15	NBR 20210601-20210630	0.093
16	TCTB 20210615-20210715	0.088
17	TCTW 20210501-20210531	0.088
18	TCTW 20210701-20210731	0.075
19	kNDVI 20210715-20210815	0.072
20	NDSI 20210615-20210715	0.068
21	NDSI 20210415-20210515	0.067
22	TCTB 20210601-20210630	0.059
23	NDSI 20210815-20210915	0.050
24	SAR-polarisaatiosuhde 20210711-20210831	0.044
25	TCTG 20210615-20210715	0.033
26	NDVI 20210801-20210831	0.022
27	TCTB 20210901-20210930	<0.001
28	SM 20210815-20210915	<0.001
29	SM 20210615-20210715	<0.001

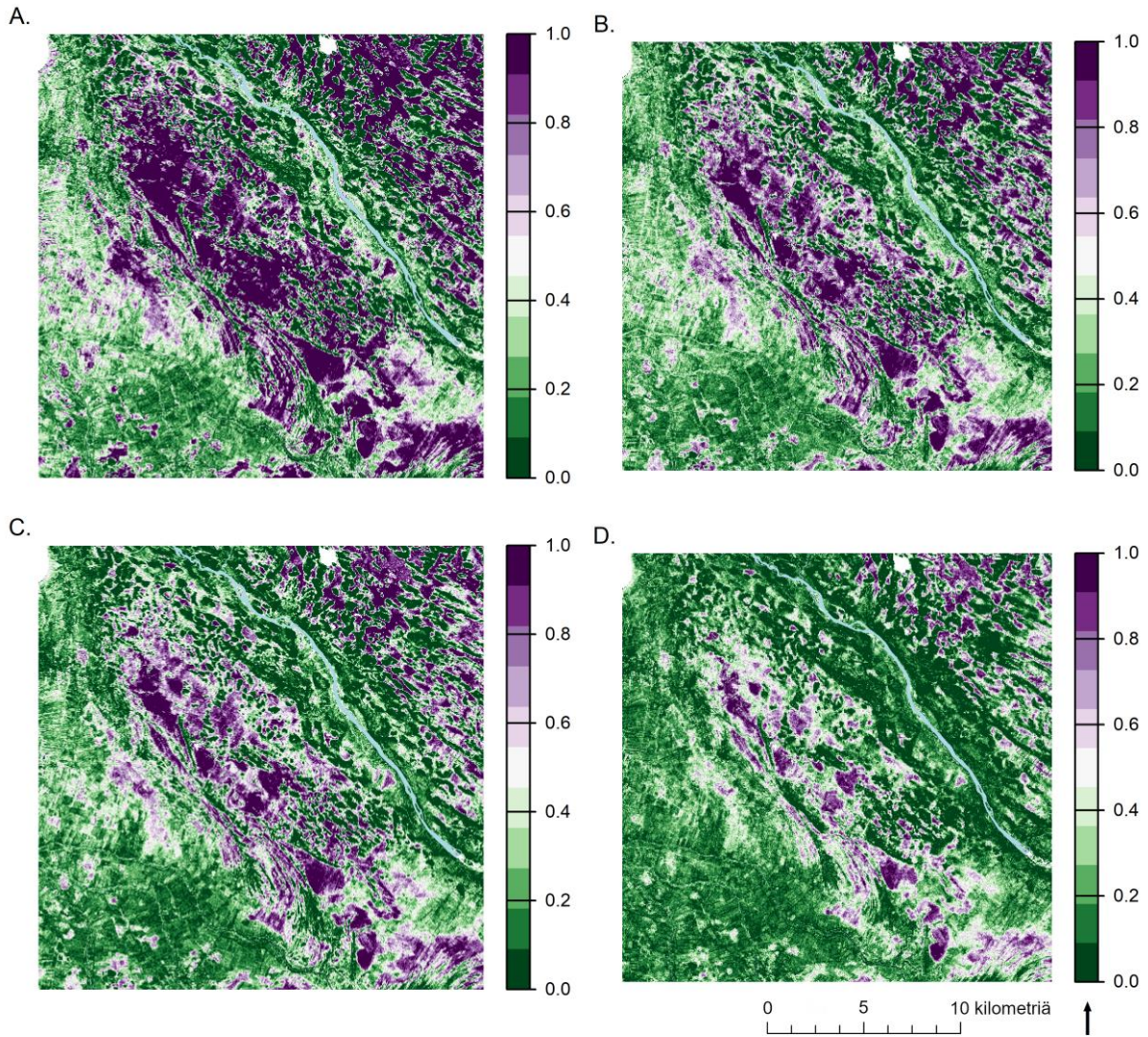
### 3.3. Todennäköisyysennusteet

RF10, RF30, RF40 ja RF60 -todennäköisyysennusteet, joka kuvaavat turvemaiden esiintymistä koko maassa todennäköisyyksillä 0-1 (0 = 0 % ja 1 = 100 %), on esitetty Kuva 6. Näiden mukaan laajimmat korkean todennäköisyyden alueet turpeen esiintymiselle ovat länsirannikolla ja Lapissa. Paksumpia turvekerroksia (erityisesti >60 cm) ennustettaessa korkean todennäköisyyden alueet pienenevät huomattavasti. Todennäköisyysennusteita on esitelty paikallisemmin Oulujoen varrelta (Kuva 7).



**Kuva 6.** RF10, RF30, RF40 ja RF60 -todennäköisyyssennusteet turvepaksuusluokkien A)  $\geq 10$  cm, B)  $\geq 30$  cm, C)  $\geq 40$  cm ja D)  $> 60$  cm esiintymiselle (0 = 0 %, 1 = 100 %).





**Kuva 7.** RF10, RF30, RF40 ja RF60 -todennäköisyyssennusteet turvepaksuusluokkien A)  $\geq 10$  cm, B)  $\geq 30$  cm, C)  $\geq 40$  cm, B)  $> 60$  cm esiintymiselle Oulujoenla (0 = 0 %, 1 = 100 %; 64.9 °P, 25.7 °I).

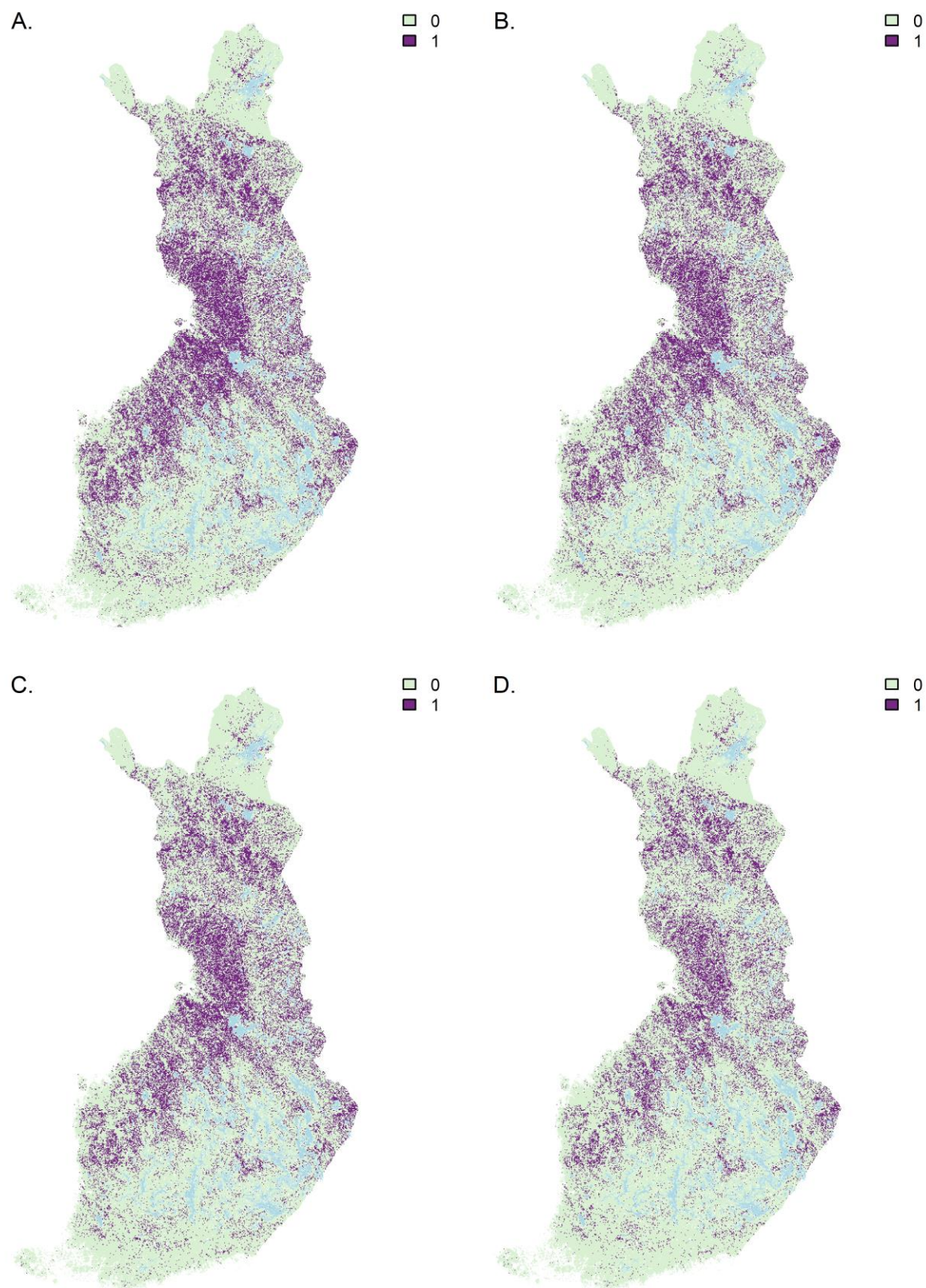
### 3.4. Binääriennusteet

RF10, RF30, RF40 ja RF60 -todennäköisyyssennusteiden luokitteluun binääriennusteiksi saadut todennäköisyyden raja-arvot vaihtelivat välillä 0,420–0,534 (Taulukko 10). Luokitellut koko maan kattavat RF10, RF30, RF40 ja RF60 -binääriennusteet, jotka kuvaavat turveluokan esiintymistä arvoilla 0 ja 1 (0=ei kuulu luokkaan, 1=kuuluu luokkaan) on esitetty Kuva 8 ja paikallisesti Oulujoen varrelta (Kuva 9).

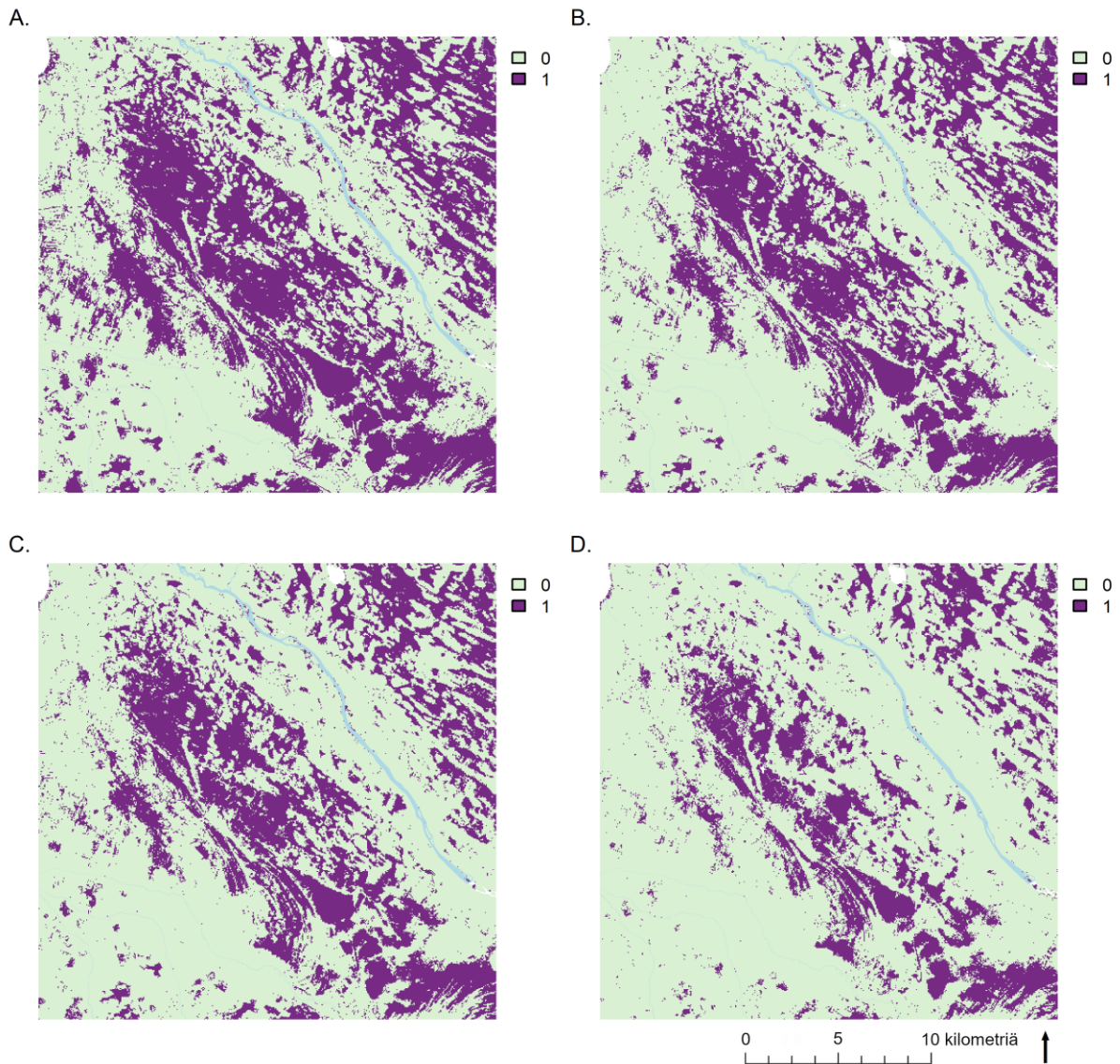
**Taulukko 10.** Todennäköisyyksien raja-arvot, joilla RF10, RF30, RF40 ja RF60 -todennäköisyyssennusteet luokiteltiin binääriennusteiksi.

Ennuste	Todennäköisyyden raja-arvo [-]
RF10	0.534
RF30	0.531
RF40	0.500
RF60	0.420



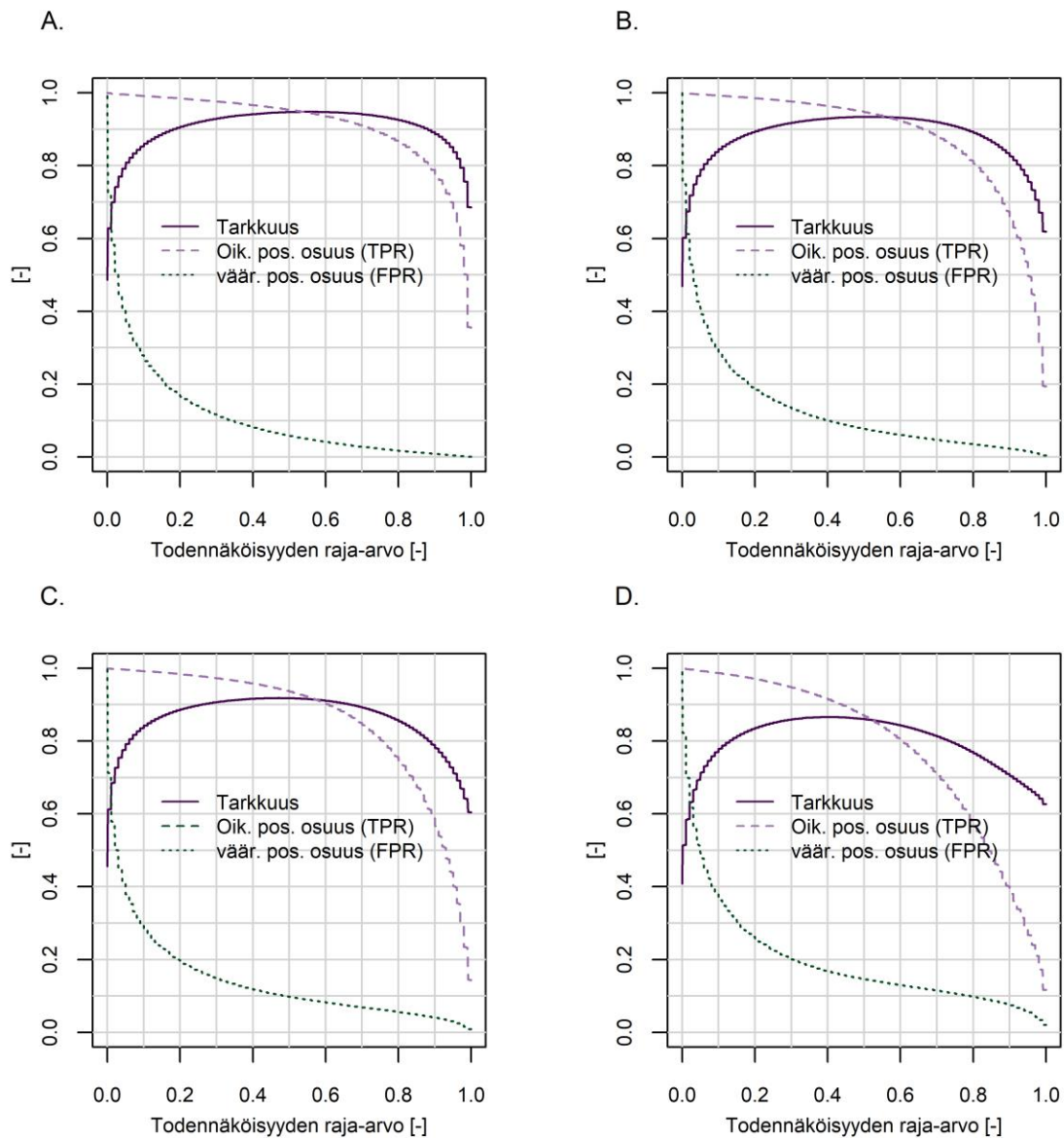


**Kuva 8.** RF10, RF30, RF40 ja RF60 -binääriennusteet turvepaksuusluokkien A)  $\geq 10$  cm, B)  $\geq 30$  cm, C)  $\geq 40$  cm ja D)  $> 60$  cm esiintymiselle (1=kuuluu luokkaan, 0=ei kuulu luokkaan).



**Kuva 9.** RF10, RF30, RF40 ja RF60 -binääriennusteet turvepaksuusluokkien A)  $\geq 10$  cm, B)  $\geq 30$  cm, C)  $\geq 40$  cm ja D)  $> 60$  cm esiintymiselle Oulujoella (1=kuuluu luokkaan, 0=ei kuulu luokkaan;  $64.9^\circ\text{P}$ ,  $25.7^\circ\text{I}$ ).

Todennäköisyssennusteiden luokittelussa käytettävän todennäköisyyden raja-arvon vaikutusta binääriennusteen yleiseen tarkkuuteen, sekä oikeiden ja väärin positiivisten ennusteiden osuuteen on tarkasteltu tarkemmin Kuva 10. RF10, RF30 ja RF40 ennusteiden osalta binääriennusteen tarkkuus voidaan pitää korkealla yli 0.9 todennäköisyyden raja-arvoilla 0.26–0.67, mutta RF60 ennusteen osalta korkeimpia tarkkuuksia saavutetaan kapeammalla valikoidulla raja-arvoilla. Pienemmillä raja-arvoilla oikeiden positiivisten ennusteiden määrä kasvaa, mutta samalla myös väärin positiivisten ennusteiden määrä kasvaa. Vastaavasti suuremmilla raja-arvoilla oikeiden positiivisten ennusteiden määrä pienenee, mutta samalla väärin positiivisten ennusteiden määrä pienenee.



**Kuva 10.** RF10 (A.), RF30 (B.), RF40 (C.) ja RF60 (D.) -todennäköisyyssennusteiden binääriluokittelussa käytettävän todennäköisyyden raja-arvon vaikutus yleiseen tarkkuuteen, sekä oikeiden (TPR) ja värien positiivisten (FPR) ennusteiden osuuteen.

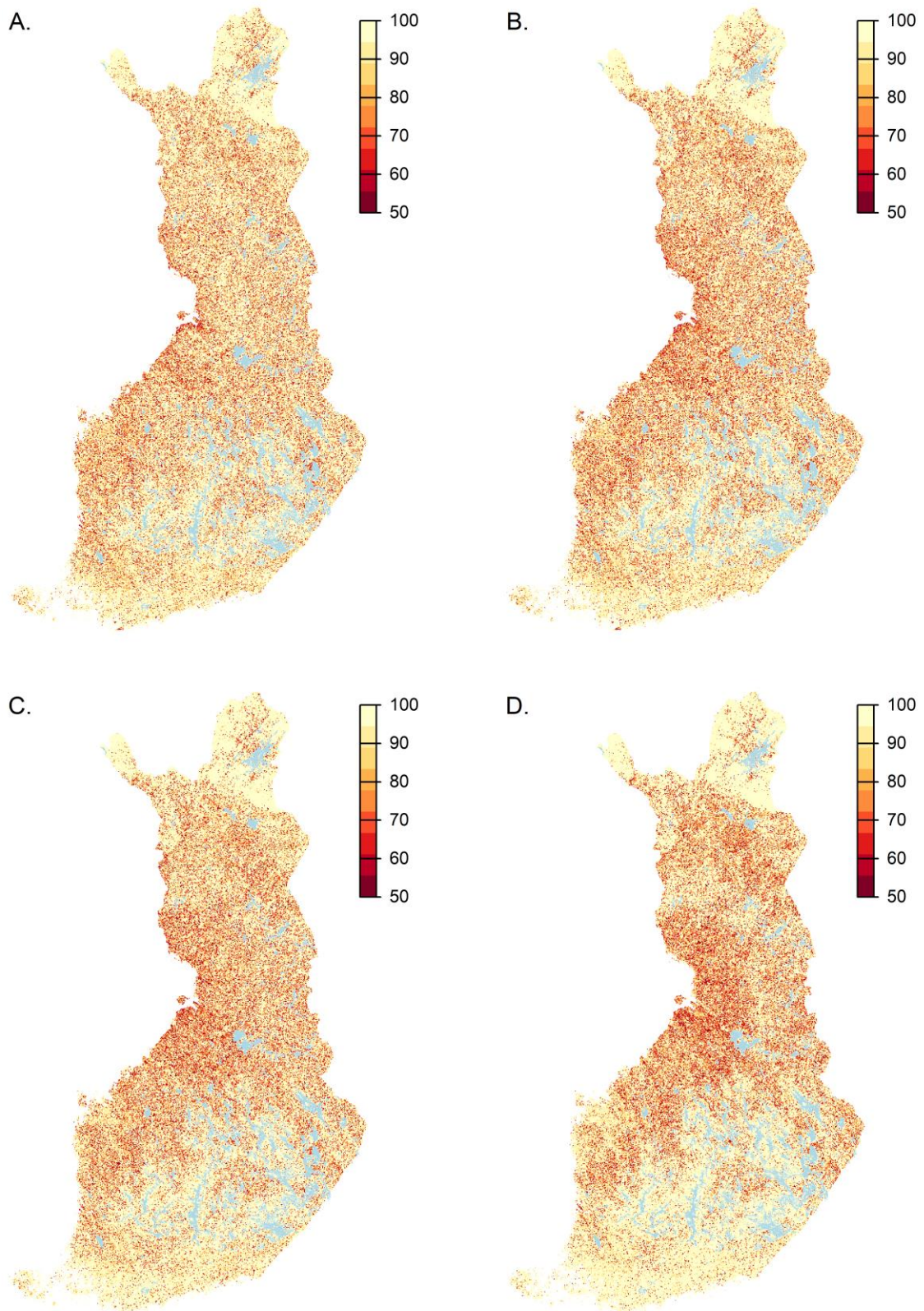
### 3.5. Ennusteiden epävarmuus

RF10, RF30, RF40 ja RF60-binäärisennusteiden päätöspuiden yksimielisyys ennusteista (ennusteiden varmuus mallin näkökulmasta) on keskimäärin 88–89 kaikilla maa-alueilla ja ne painottuvat jakauman yläpäähän (Taulukko 11). Puolet yksimielisyysarvoista on  $\geq 95$ . Päätöspuidenyksimielisyys ennusteiden turvepaksuusluokkien mukaisilla turvealueilla on hieman alhaisempi verrattuna kaikkiin maa-alueisiin, ja yksimielisyyden arvot ovat myös hieman matalampia paksumpien turveluokkien kohdalla. Esimerkiksi, RF60-binäärisennusteen kohdalla keskimääräinen yksimielisyys  $>60$  cm luokassa on 72 (Taulukko 11). Yksimielisyysarvojen alueellinen jakautuminen koko Suomen tasolla on esitetty (Kuva 11). Yksimielisyyden arvot vaihtelevat kuitenkin paikallisesti paljonkin, mistä on esitetty esimerkki (Kuva 12).

**Taulukko 11.** RF10, RF30, RF40 ja RF60 -binäärisennusteiden päätöspuiden yksimielisyys kaikilla maa-alueilla ja erikseen ennustettujen turvepaksuusluokkien  $\geq 10$  cm,  $\geq 30$  cm,  $\geq 40$  cm,  $>60$  cm mukaisilla turvealueilla.

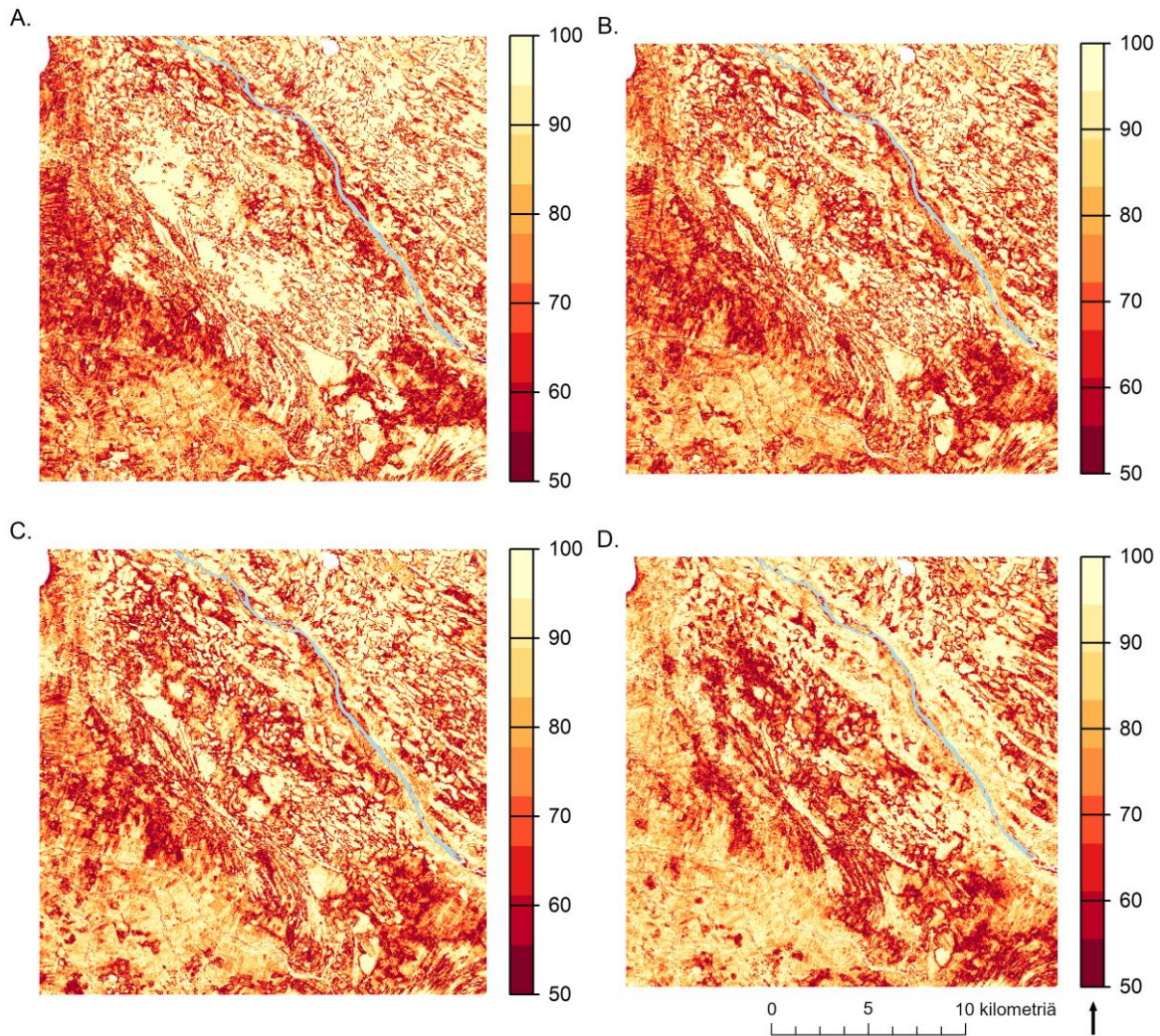
Keskiarvo / persentiilit	Kaikki maa-alueet				Turvealueet			
	Keskiarvo	25.	50.	75.	Keskiarvo	25.	50.	75.
RF10	89	82	95	99	86	76	92	99
RF30	88	80	95	99	83	71	86	95
RF40	88	80	95	99	79	66	81	93
RF60	89	81	96	99	72	57	71	85





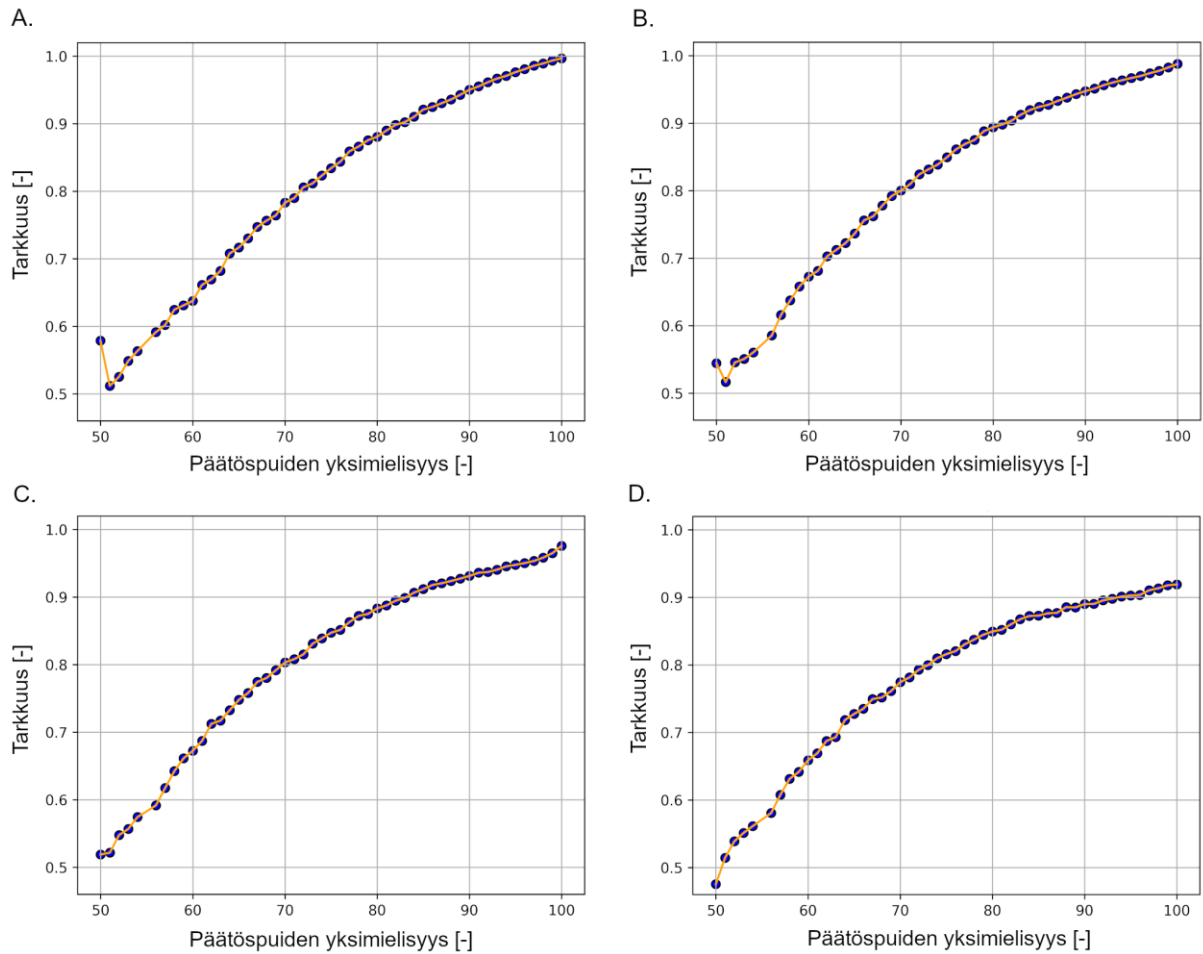
**Kuva 11.** RF10, RF30, RF40 ja RF60 -mallien päätöspuiden yksimielisyys A)  $\geq 10$  cm, B)  $\geq 30$  cm, C)  $\geq 40$  cm ja D)  $> 60$  cm turvepaksuusennusteille. Päätöspuiden yksimielisyys kuvaa ennusteen varmuutta RF-mallin näkökulmasta (100=kaikki päätöspuut yksimielisiä, 50=puolet päätöspuista yksimielisiä).





**Kuva 12.** RF10; RF30, RF40 ja RF60 -mallien päätospuiden yksimielisyys A)  $\geq 10$  cm, B)  $\geq 30$  cm, C)  $\geq 40$  cm ja D)  $> 60$  cm turpeen paksuusennusteille Oulujoella (100=kaikki päätöspuut yksimielisiä, 50=puolet päätöspuista yksimielisiä). Päätöspuiden yksimielisyys kuvaa ennusteen epävarmuutta RF-mallin näkökulmasta.

RF10, RF30, RF40 ja RF60 -mallien binääriennusteen päätöspuiden yksimielisyyden ja tarkkuuden välistä yhteyttä tutkittiin suorittamalla yksinkertaiset korrelaatioanalyysit hyödyntämällä Pearsonin ja Spearmanin korrelaatiomittoja. Kuva 13 esittää ristiinvalidoinnilla mitatun RF-mallien tarkkuuden ja ennusteen päätöspuiden yksimielisyyden välistä yhteyttä kuvaajassa. Kuvaajissa havaittiin merkittävät korrelaatioarvot Pearson's  $\geq 0.96$  ja Spearman's  $= 1.0$ . Toisin sanoen päätöspuiden yksimielisyys on vahva indikaattori malliennusteen tarkkuudesta. Poikkeava ilmiö erityisesti käyrien RF10 ja RF30 vasemmassa päässä selittyy niin, että mallin maksimiepävarmuuden (yksimielisyys=50) tapauksessa ennustetaan lähes aina maksimiluokkaa (eli luokkaa 0), joka on yli 50 % koko aineiston jakaumasta.



**Kuva 13.** RF10 (A.), RF30 (B.), RF40 (C.) ja RF60 (D.) -mallien ennusteen tarkkuuden ja päätöspuiden yksimielisyyden (epävarmuuden) välinen yhteys. RF-mallien epävarmuus on isoin pienillä yksimielisyyden arvoilla (pätöspuut eri mieltä ennusteesta) ja pienin yksimielisyyden isoilla arvoilla (pätöspuut samaa mieltä ennusteesta).

### 3.6. Turpeen esiintyminen maatalousmailla

RF30-binääriennusteen mukaan  $\geq 30$  cm paksua turvetta esiintyy maatalousmailla yhteensä 273 214 ha, eli noin 11 % koko maan maatalousmaa-alasta (Taulukko 12). Suurista maatalousalueista turvetta esiintyy maatalousmailla yleisimmin Pohjois- ja Etelä-Pohjanmaalla, joiden yhteenlaskettu maatalousmaiden turveala on 44 % koko maan maatalousmaiden turvealasta. Suurista maatalousalueista vähiten turvetta esiintyy Varsinais-Suomessa ja Uudellamaalla, joiden yhteenlaskettu maatalousmaiden turveala on 2 % koko maan turvealasta.

RF30-binääriennusteen mukaan rajatuista  $\geq 30$  cm paksuista turvealuista 30–60 cm paksujen turvealueiden osuus on 27 % koko maan tasolla. Paksumpien kuin 60 cm turvealueiden osuus on 73 % (Taulukko 12). Nämä osuudet vaihtelevat kuitenkin maakunnittain. Suurin  $>60$  cm paksujen turvealueiden osuus on Lapissa ja Kainuussa (83 %), ja pienin Kanta-Hämeessä (59 %) ja Pohjanmaalla (64 %).

**Taulukko 12.** RF30 ja RF60 -binääriennusteiden turvemaiden pinta-alat maatalousmailla, sekä 30–60 cm ja >60 cm paksujen turvealojen suhteelliset osuudet  $\geq 30$  cm paksuista turvealueista maakunnittain. Maatalousmaa määritetty Ruokaviraston peltolohkorekisterin peruslohkoista.

	Maakunta	Maatalous- maa-ala [ha]	$\geq 30$ cm turveala [ha]	$\geq 30$ cm turveala [%]	>60 cm turveala [ha]	>60 cm turveala [%]	30–60 cm turvealan osuus $\geq 30$ cm turvealasta [%]	>60 cm turvealan osuus $\geq 30$ cm turvealasta [%]
1	Ahvenanmaan maakunta	22 092	300	1	262	1	13	87
2	Etelä-Karjala	58 852	4 091	7	2824	5	31	69
3	Etelä-Pohjanmaa	258 212	36 960	14	24974	10	32	68
4	Etelä-Savo	90 021	7 203	8	5671	6	21	79
5	Kainuu	33 529	9 722	29	8056	24	17	83
6	Kanta-Häme	109 122	4 152	4	2447	2	41	59
7	Keski-Pohjanmaa	60 332	19 218	32	14290	24	26	74
8	Keski-Suomi	109 666	9 449	9	7252	7	23	77
9	Kymenlaakso	89 054	1 528	2	1045	1	32	68
10	Lappi	57 178	26 757	47	22174	39	17	83
11	Pirkanmaa	175 836	6 687	4	4562	3	32	68
12	Pohjanmaa	143 802	11 652	8	7476	5	36	64
13	Pohjois-Karjala	98 997	16 311	16	11286	11	31	69
14	Pohjois-Pohjanmaa	258 566	81 914	32	60690	23	26	74
15	Pohjois-Savo	165 035	18 222	11	13018	8	29	71
16	Päijät-Häme	88 137	1 572	2	1143	1	27	73
17	Satakunta	148 166	12 292	8	9467	6	23	77
18	Uusimaa	190 147	1 934	1	1304	1	33	67
19	Varsinais-Suomi	301 589	3 254	1	2432	1	25	75
	YHTEENSÄ	2 458 332	273 214	11	200372	8	27	73

RF30 ja RF60 -binääriennusteiden mukaan  $\geq 30$  cm ja >60 cm paksua turvetta esiintyy usein vain osalla peltolohkoa (Taulukko 13). Esimerkiksi  $\geq 30$  cm paksua turvetta esiintyy 100 % peltolohkon alasta 19 395 peltolohkolla. Peltolohkoja, joilla kyseisen paksuista turvetta esiintyy vähintään 30 % lohkon alasta, on 115 729 kpl. Vastaavasti >60 cm paksua turvetta esiintyy 100 % peltolohkon alasta 10 087 peltolohkolla. Peltolohkoja, joilla kyseisen paksuista turvetta esiintyy vähintään 30 % lohkon alasta, on 78 059 kpl.

**Taulukko 13.** RF30 ( $\geq 30$  cm) ja RF60 ( $> 60$  cm) -binääriennusteiden mukainen turpeen suhteellinen ala peltolohkoilla. Peltolohkot on määritetty Ruokaviraston peruslohkojen mukaan, joiden kokonaismäärä 1 093 924 kpl ja kokonaisala 2 458 332 ha.

Turvealueen suhteellinen ala peltolohkolla	Peltolohkoja [kpl]		Peltolohkojen ala [ha]		Turveala peltolohkoilla [ha]	
	$\geq 30$ cm	$> 60$ cm	$\geq 30$ cm	$> 60$ cm	$\geq 30$ cm	$> 60$ cm
100 %	19 395	10 087	46 247	26 585	46 247	26 585
$\geq 90$ %	38 624	22 305	115 253	75 222	112 662	73 307
$\geq 80$ %	50 265	30 640	150 306	103 104	142 566	97 042
$\geq 70$ %	61 161	38 466	181 598	127 622	166 042	115 433
$\geq 60$ %	72 234	46 582	211 892	151 870	185 725	131 189
$\geq 50$ %	84 274	55 268	243 976	177 182	203 360	145 096
$\geq 40$ %	98 490	65 340	281 276	206 032	220 093	158 055
$\geq 30$ %	115 729	78 059	326 234	241 849	235 722	170 501
$\geq 20$ %	139 364	95 019	388 177	290 543	251 007	182 543
$\geq 10$ %	177 355	122 154	491 886	370 979	265 987	194 166

### 3.7. Turvepeltolohkojen tunnistaminen

Turvepeltolohkojen tunnistamisessa koostettiin opetuspisteiden avulla turvesyvyyksiltään tunnettu vertailulohkojen aineisto, jota verrattiin RF30- ja RF60-mallien tuottamiin tuloksiin.

RF30-mallin arviointia varten vertailulohkojen aineistoon saatiin opetuspisteiden perusteella valittua 1262 peltolohkoa (3 570 ha), joista 596 (2 293 ha) oli turvelohkoja ( $\geq 30$  cm). Oletettuja ei-turvemaan lohkoja lisättiin lohkoaineistoon tämän jälkeen niin, että turvelohkojen pinta-alaosuus kaikista mukaan otetuista lohkoista laski 11,1 %:iin (Taulukko 12). Vertailulohkoaineiston kooksi saatiin näin 8 984 peltolohkoa (20 664 ha).

Paksuturpeisten lohkojen vertailuaineistoon saatiin opetuspisteiden perusteella 1 212 lohkoa (3 397 ha), josta 439 (1 791 ha) oli paksuturpeisia ( $> 60$  cm). Kokonaismäärän ero edellä kuvattuun turvelohkojen vertailuaineistoon tuli niiden lohkojen pois jättämisestä, joissa oli sekä paksu- ( $> 60$  cm) että ohutturpeisiä (0–60 cm) opetuspisteitä niin ettei kumpikaan saavuttanut kolmen neljäsosan enemmistöä. Ei-turvemaan oletettuja lohkoja lisättiin tämän jälkeen aineistoon, kunnes paksuturpeisten lohkojen ala saavutti 8,2 % osuuden. Vertailulohkoaineisto kattoi tällöin 9 479 lohkoa (21 849 ha).

Taulukko 14 ja Taulukko 15 esittävät RF30- ja RF60-mallien tunnistustarkkuuksia turvelohkojen (RF30) ja paksuturpeisten lohkojen (RF60) osalta. Tarkastelu on tehty uudelleenluokittelemalla (turve/ei-turve) todennäköisyysennusteet 0.05 yksikön välein tasoilla 0.40–0.90, ja määrittämällä mallin mukaan vähintään 50 % peitto-osuudella olevat peltolohkot turve- tai paksuturpeisiksi lohkoiksi. Mukaan on otettu myös havaintopistetasolla optimoidut luokitusrajat (Taulukko 10) merkittyinä \* symbolilla. Taulukoissa esitetään oikeiden positiivisten (TPR) ja väärin positiivisten (FPR) mallinnustulosten osuudet sekä lohkojen lukumäärän että niiden pinta-alan suhteen. Lukumäärä ilmaisee suhdetta tarkasteltujen lohkojen kokonaismäärään, eikä huomioi mahdollisia eroja oikein tai väärin luokittuneiden lohkojen pinta-aloissa. Vastavasti pinta-alaan suhteutetuissa tuloksissa vain kokonaispinta-alat merkitsevät, riippumatta siitä monestako lohkoista vertailtavat pinta-alat koostuvat. Suhteutus pinta-alaan tasoittaa myös eroja sen suhteen, että opetuspisteiden avulla valitut turve- ja paksuturpeiset lohkot ovat keskimäärin suurempia kuin ei-turvelohkot.

**Taulukko 14.** RF30-ennusteen tunnistustarkkuus turvelohkoille todennäköisyysennusteen luokittelussa käytetyn todennäköisyyden raja-arvon suhteen.

Todennäköisyyden raja-arvo	TPR / lukumäärä (%)	TPR / pinta-ala (%)	FPR / lukumäärä (%)	FPR / pinta-ala (%)
0.40	92,45	96,56	2,43	2,49
0.45	90,60	95,52	1,66	1,96
0.50	87,75	94,00	1,21	1,60
0.531 *	86,58	93,6	0,95	1,39
0.55	85,74	93,04	0,92	1,36
0.60	81,38	89,85	0,68	1,09
0.65	75,84	87,08	0,46	0,83
0.70	67,28	79,20	0,32	0,67
0.75	56,21	69,97	0,18	0,25
0.80	42,95	54,19	0,08	0,05
0.85	23,99	28,84	0,04	0,01
0.90	8,22	8,92	0,00	0,00

\*Havaintopistetasolla optimoitu todennäköisyyden raja-arvo (Taulukko 10).

**Taulukko 15.** RF60-ennusteen tunnistustarkkuus paksuturpeisille lohkoille todennäköisyysennusteen luokittelussa käytetyn todennäköisyyden raja-arvon suhteen.

Todennäköisyyden raja-arvo	TPR / lukumäärä (%)	TPR / pinta-ala (%)	FPR / lukumäärä (%)	FPR / pinta-ala (%)
0.40	84,97	93,26	1,39	1,90
0.42 *	82,23	91,77	1,20	1,69
0.45	79,27	90,37	1,02	1,23
0.50	72,67	84,04	0,71	0,99
0.55	65,83	78,39	0,46	0,54
0.60	56,26	68,35	0,33	0,42
0.65	42,37	55,76	0,17	0,20
0.70	30,52	41,30	0,14	0,17
0.75	18,68	23,22	0,08	0,08
0.80	10,48	12,11	0,03	0,02
0.85	4,33	4,15	0,01	0,01
0.90	0,68	1,53	0	0,00

\*Havaintopistetasolla optimoitu todennäköisyyden raja-arvo (Taulukko 10).

Tunnistustarkkuus lohkojen pinta-alaan suhteutettuna vastaa melko hyvin yksittäisiin opetus-pisteisiin perustuvaa tarkkuutta. Parhaiksi valituilla todennäköisyyden raja-arvoilla tunnistetaan yli 90 % turvepaksuudeltaan sekä turvelohkoista ( $\geq 30$  cm) ja että paksuturpeisista lohkoista ( $> 60$  cm). Lukumäärään suhteutettuna turve- ja paksuturpeisiä lohkoja löydetään hie-man pienempi osuus, mutta myös virhetunnistukset ovat pääsääntöisesti harvinaisempia. Mallinnus luokittelee pienet vertailuaineistojen turvelohkot isompia helpommin ei-turvelohkoiksi, mikä vähentää niiden vaikutusta pinta-alapohjaiseen TPR-lukemaan. Useat mallien väärin turvelohkoiksi tunnistamista ei-turvelohkoista puolestaan ovat opetuspisteiden perusteella valittuja lohkoja, jotka sijaitsevat turvealueiden lähellä mutta ne on vertailuaineistoissa luokiteltu ei-turvemaaksi tai ei-paksuturpeisiksi. Niiden pinta-ala on keskimääräistä lohkokoko-koa suurempi, mikä nostaa pinta-alapohjaista FPR-lukemaa lukumäärään suhteutettua enemmän. Vertailulohkoihin lisätyistä oletetuista ei-turvelohkoista vain hyvin pieni osuus tulee mallinnettua turvelohkoina.

## 4. Tulosten tarkastelu

### 4.1. Malliennusteet

RF-ennusteiden yleiset tarkkuudet ovat kokonaisuudessaan hyviä. Alueet, joilla esiintyy turvetta  $\geq 10$  cm eroteltiin muista aluista 96 % tarkkuudella. Turvepaksuusluokat  $\geq 30$  cm ja  $\geq 40$  cm tunnistettiin myös erittäin tarkasti noin 93–94 % tarkkuudella, mutta näiden ennustamisessa väärin positiivisten ennusteiden määrä kasvoi hieman verrattuna turpeen esiintymisalueiden ( $\geq 10$  cm) ennustamiseen. Paksujen turvemaiden ( $\geq 60$  cm) tunnistamisessa yleinen tarkkuus heikkeni hieman; ne tunnistettiin 89 % tarkkuudella. Väärien positiivisten ennusteiden määrä kasvoi verrattuna muihin ennusteisiin. Huomionarvoista on, että mallien opetukseen käytettävän aineiston diskretoitu (0/1) turvesyvyysjakauma, eli histogrammi, oli epätasaisempi suuremmilla turvesyvyyden raja-arvoilla ( $\geq 10, 30, 40, 60$  cm), jossa 0-luokka (ei-turvetta) oli aina enemmistöluokkana. Toisin sanoen, turveaineiston diskretoinnin raja-arvon kasvaessa, yhä enemmän turvesyvyyden havaintopisteitä luokittiin 0-luokkaan (ei-turve) kuin 1-luokkaan (turve). Tämä osaltaan selittää RF10-mallin parempaa suorituskkyä verrattuna RF60-malliin; koneoppimismallien on hankalampaa löytää hyvää päätösrajaa luokittelutilanteissa, kun luokkien välinen lukumääräero kasvaa. Yleisesti ottaen, isoissa opetusaineiston jakaumien epätasapainotilanteissa suurimman esiintyvyyden luokat helposti dominoivat mallien sovitusta.

Saavutetut yleiset ennustetarkkuudet vastaavat kirjallisuudessa esiintyvien vastaavien tutkimuksien ennustetarkkuuksia (Minasny ym. 2019). Esimerkiksi O'Leary ym. (2022) mallinsivat turvemaiden esiintymistä (ei-turve ja turve luokat) Skotlannin keskiosissa Artificial Neural Network (ANN) – koneoppimismenetelmällä ja radiometrisillä matalalentoaineistoilla tarkkuuksilla 95–96 %. Ågren ym. (2022) mallinsivat turvemaiden esiintymistä ( $\geq 30$  cm,  $\geq 40$  cm ja  $\geq 50$  cm) Ruotsissa lineaarisilla ja epälineaarilla malleilla ja korkearesoluutioisella ( $2 \text{ m} \times 2 \text{ m}$ ) topografisella maaperän kosteusaineistolla tarkkuuksilla 89–92 %. Poggio ym. (2019) mallinsivat turvemaiden esiintymistä (turve-, turpeinen- ja mineraalimaa) Skotlannissa lineaarisella Generalized Additive Model- menetelmällä sekä Random Forest and Convolutional Neural Networks -koneoppimismenetelmillä ja Sentinel1 ja -2 -aineistoilla tarkkuuksilla 65–90 %. Aitkenhead (2017) mallinsi turvemaiden esiintymistä (ei-turve, turve) Skotlannissa Backpropagation Neural Network -menetelmällä ja laajalla määrällä (91 kpl) turpeen esiintymistä ennustavia aineistoja tarkkuudella 86,4 %. Tässä raportissa tehty työ eroaa monista aikaisemmista tutkimuksista havaintoaineiston ja mallinnusalueen laajuuden osalta. Aikaisemmissa tutkimuksissa ennustettiin myös useimmiten turpeen esiintymistä ja harvemmin turvekerroksen paksuutta ja sen vaihtelua.

Geneettisen algoritmin RF-mallinnukseen valikoimat turpeen esiintymistä selittävät kaukokartoitus- ja paikkatietoaineistot vastaavat myös teoreettista käsitystä turvemaiden esiintyvistä selittävästä tekijöistä. Esimerkiksi Kalium-säteilyn ennustusvoiman oletettiin olevan korkea (Minasny ym. 2019, Airo ym. 2014, Hyvönen ym. 2005, Virtanen 1997) ja se valikoitui kaikissa malleissa yhdeksi tärkeimmistä turvemaiden esiintymistä selittävästä tekijöistä. Paksuja turvemaita ( $> 60$  cm) ennustettaessa Thorium-säteilyaineisto nousi myös keskeiseen rooliin. Veden tiedetään vaimentavan gamma-säteilyä ja se onkin todettu kirjallisuudessa turvemaiden esiintymistä ja paksuutta selittäväksi tekijäksi (Minasny ym. 2019). Maaperän kosteutta kuvaava topografinen DTW-indeksi oli myös keskeinen turvemaiden esiintymistä selittävä tekijä. Tämä vastaa myös Ruotsissa tehtyjen tutkimusten tuloksia, joissa turvemaita ennustettiin onnistuneesti korkeusmallista johdettujen maaperän kosteusaineistojen avulla (Ågren ym. 2022).



DTW:n merkitys kuitenkin laski  $>60$  cm paksuja turpeita ennustettaessa, mikä kertoo siitä, että puhtaasti korkeusmalliin pohjautuva DTW selittää hyvin turvemaiden esiintymistä, mutta ei niiden paksuutta. Turvemaiden topografian variaatio on usein pienempi kuin muiden maa-lajien. Tämä näkyi myös siinä, että topografista vaihtelua kuvaavat aineistot olivat korkeassa roolissa ennustettaessa erityisesti ohuiden turvemaiden olemassaoloa. Satelliittiaineistoista tärkeimmäksi osoittautuivat Sentinel-2 aineistosta lasketut NDSI ja SCI, NDVI ja NBR indeksit, mutta useat muutkin Sentinel-2 indeksit ja Sentinel-1 tutkakuva-aineistot olivat tärkeässä roolissa. NDSI:n korkea rooli malleissa oli yllättävää, sillä se on indeksi, joka reagoi ensisijaisesti lumeen (Riggs ym. 1994). SCI, joka kuvaa maaperän väriä, on myös aiemmissa tutkimuk-sissa osoittautunut hyväksi muuttujaksi turvemaiden kaukokartoituksessa, esimerkiksi Skot-lannissa (Poggio ym. 2019). NDVI kuvaa kasvillisuuden määrää ja on yleisesti käytetty indeksi kuivuuden kartoittamisessa (Alahacoon & Edirisinghe 2022). On oletettavaa, että indeksi rea-goin suoalueiden erilaiseen kasvillisuuteen ja kosteustilaan. NBR kuvaa palaneita alueita (Key & Benson N 1999) ja sen mukaan valikoituminen oli myös yllättävää. TCTW-indeksi rooli oli myös mielenkiintoinen, sillä se ei tullut valituksi mukaan  $\geq 10$  cm turvepaksuutta ennustetta-essa. Se valikoitu piirteeksi  $\geq 30$  cm,  $\geq 40$  cm ja  $> 60$  cm luokkia ennustettaessa ja sen merkitys painottui paksuimpia turpeen paksuusluokkia ennustettaessa. TCTW- kuvaa maaperän kos-teusolosuhteita (Shi & Xu 2019). Selittävien aineistojen keskinäistä tärkeysjärjestystä tarkastel-taessa on kuitenkin huomioitava, että geneettinen algoritmi on stokastinen menetelmä. Tämä vuoksi muuttujien järjestys vaihtelee eri piirrevalintakierrosten välillä. Näin käy erityisesti nii-den aineistojen kohdalla, joidenka suhteelliset tärkeydet ovat lähellä toisiaan, kuten on satel-liittiaineistojen tapauksessa.

Ennusteiden alueellinen kattavuus on 99.5 % Suomen maa-alasta. Luokkaennusteiden katta-vuus vaihtelee hieman mallinnusten välillä. Vaihtelu johtuu kullekin mallille erikseen tehdystä piirrevalinnasta ja selittäviksi tekijöiksi valittujen muuttujien joukosta. Näin tapahtuu, koska puuttuvat arvot selittämissä tekijöissä periytyvät puuttuviksi arvoiksi malliennusteeseen. En-nusteita ei ole laadittu Suomen rajojen ulkopuolelle eikä vesistöalueille. Selittämissä tekijöissä esiintyvien aineistoaukkojen vuoksi kaikissa ennusteissa on mukana maa-alueita, joille ei ole ennustetta. Alueet keskittyvät kaistaleiksi Suomen itärajalle sekä ulkosaaristoon niin etelä- kuin länsirannikolla. Myös maan sisäosissa ja suurten sisävesistöjen saarissa voi olla aukkoja ennusteessa. Mallintamattomien pikseleiden osuus maa-aluepikseleistä vaihtelee 0,41–0,48 % välillä: eniten mallintamattomia maa-alueita on RF30-ennusteessa ja vähiten RF60-ennus-teessa.

## 4.2. Epävarmuus

### Random Forest

RF-mallien päätöspuiden yksimielisyyssarvot ovat painottuneet lähelle arvoa 100 (Taulukko 11, Kuva 11 ja Kuva 12). RF10- ja RF30-mallien tapauksissa arvojen alueellisessa jakaumissa ei tehty merkittäviä havaintoja lukuun ottamatta Pohjois-Lapin erityisen korkeat arvot. Pohjois-Lapissa opetuspisteiden määrä oli matalin. RF60-mallin tapauksessa Pohjanmaan alue näyt-täisi erottuvan hieman matalimpina yksimielisyyssarvoina. Ennusteiden yksimielisyyssarvot voi-vat vaihdella suuresti lyhyenkin matkan sisällä. Yksimielisyyssarvolla ja ennusteen yleisellä tark-kuusarvolla näyttäisi olevan vahva epälineaarinen riippuvuus siten, että suurempi yksimieli-syyssarvo vastaa korkeampaa tarkkuutta (Kuva 13). Yksimielisyyssarvo on näin ollen vahva indikaat-tori ennusteiden epävarmuudelle.

## **Opetus- ja testausaineisto**

Mallinnuksessa käytetty opetus- ja testausaineisto oli erittäin laaja maaperähavaintomäärätään, mutta aineistoon liittyy myös epävarmuuksia. Maaperähavainnot oli kerätty lähtökohtaisesti muihin tarkoituksiin kuin koko maan kattava systemaattisen mallinnuksen tarpeisiin. Tästä johtuen maaperähavainnot eivät olleet tasaisesti jakautuneita koko Suomen alueelle vaan painoutuivat alueille, joilla esiintyy runsaasti soita ja turpeen energiakäytölle on painetta. Lisäksi maastohavainnot ovat usein keskittyneet yksittäisiin suoaltoiin. Mallin opettamisen kannalta tämä ei tuota optimaalista tulosta, vaan mallin oppiminen saattaa painottua alueellisiin olosuhteisiin. Maaperähavainnot painoutuivat myös turvemaille ja kivennäismaahavainnointoja on opetus- ja testausaineistossa tämän vuoksi suhteellisesti vähemmän. Havaintojakaumaa tasapainotettiin generoimalla turpeen 0-syvyys eli kivennäismaapisteitä. Peltomaiden turvehavainnointoja oli myös korjattava turvekerroksen hajoamisen vuoksi. Korjaus tehtiin käyttämällä keskimääräistä vuosittaista turpeen ohenemista, mikä myös aiheuttaa aineistoon epävarmuutta. Lisäksi maaperähavainnointoja oli peltomailla suhteellisen vähän noin 97 309 kpl. 3,5 miljoonasta havaintomäärästä. Yhdessä nämä tekijät tuottavat oman epävarmuutensa mallin opetukseen ja testaukseen.

## **Geofysikaaliset matalalentoaineistot**

Geofysikaalisten matalalentoaineistojen spatiaalinen erottelukyky 50 m × 50 m määritteli tuotettujen turve-ennusteiden erottelukyvyn, mikä näkyy ennusteissa spatiaalisena epätarkkuutena turvemaaksi ennustettujen alueiden rajoista. Karkea resoluutio saattoi myös aiheuttaa rantavyöhykkeillä yksittäisten rasterisolujen kohdalla virheellisen turve-ennusteen vesistön puolelta tulleen mittausignaalin vuoksi. Mallinnuksessa käytettyjen muiden aineistojen puolesta mallinnus olisi voitu tehdä tarkemmalla resoluutiolla. Esimerkiksi, käytetyt satelliittiaineistojen resoluutio oli 10 m × 10 m ja 20 m × 20 m. Korkeusmallista johdettujen indeksien alkuperäinen resoluutio oli 2 m × 2 m.

Geofysikaalisten matalalentoaineistojen mittausajankohta aiheuttaa myös epätarkkuutta peltomaiden turve-ennusteissa, koska peltojen turvekerros ohenee viljelykäytössä keskimäärin 1,2 cm vuodessa (Räsänen ym. 2023). Geofysikaalisten matalalentoaineistot mitattiin vuosina 1972–2007, mikä tarkoittaa, että mittaukset eivät vastaa mallinnusajankohdan turpeen paksuuden tilannetta. Peltomaiden turvekerroksen oheneminen huomioitiin mallinnuksessa käytetyissä opetus- ja testausaineistoissa, mutta tätä ei pystytty huomioimaan geofysikaalisissa matalalentoaineistoissa, mikä aiheuttaa ylimääräistä epälineaarisuutta ja siitä johtuvaa epävarmuutta mallinnukseen.

## **Satelliittiaineistot**

Käytettävien aineistojen valintaan vaikuttivat ensisijaisesti datan kattavuus, avoin saatavuus, maksuttomuus, sekä mahdollisuus hyödyntää valmiita työkaluja aineistojen prosessoinnissa. Kaikki projektissa käytetyt ja tuotetut satelliittikaukokartoitusaineistot kattavat koko maan. Sentinel-1 on C-kanavan tutkasatelliittiaineistoa. Lyhyen aallonpituuden vuoksi (n 5 cm) aineiston heikkoutena on signaalin heikko tunkeutuminen maaperään. Koko maan kattavat pidemmän aallonpituuden, kuten L-kanavan aineistot, olisivat projektin toteutushetkellä olleet maksullisia.

Satelliittiaineistojen tarkkuutta alkuperäisistä vähensi sovitus projektin hilaruutuihin 50 metrin spatiaalisella resoluutiolla. Aineistojen alkuperäiset resoluutiot olivat 10 metriä optisille aineistoille ja 20 metriä tutka-alueen aineistoille.

Optisista Sentinel-2-satelliittiaineistoista johdetut indeksit valittiin kirjallisuuskatsauksen perusteella. Valittujen indeksien lisäksi selittävinä muuttujina olisi voinut käyttää muita indeksejä sekä kuvauskanavakohtaisia heijasteita. Optisten aineistojen muuntaminen pilvettömiksi mosaiikkeiksi pohjautui aineistojen mukana jaettaviin automaattisesti tuotettuihin pikseliluokitukseen, joiden perusteella pilvisiksi luokiteltuja pikseleitä ei huomioitu mosaiikin muodostamisessa. Automaattisen pikseliluokituksen mahdolliset virheet ovat voineet johtaa väärin pikselien valintaan mosaiikkeihin ja siten vaikuttaa indeksien arvoihin, mutta näiden merkitys mallinnustuloksiin on kokonaisuutena vähäinen.

Tutka-alueen (SAR, Synthetic Aperture Radar) satelliittiaineistojen pohjana käytettiin Suomen ympäristökeskuksen ja Ilmatieteen laitoksen tuottamia Sentinel-1-mosaiikkeja. Alkuperäisten mosaiikkien aikaikkunat kattoivat 11 vuorokauden jaksot, mutta mosaiikkien aineisto oli aukkoista, minkä vuoksi niiden käyttäminen sellaisenaan mallinnuksessa ei onnistunut. Alkuperäisten mosaiikkien arvoja keskiarvoistamalla muodostettiin noin 7 viikon aikaikkunat kattavat mosaiikit, mikä vähensi aineiston temporaalista vaihtelua.

Satelliittiaineistoista johdetut kunkin ajankohdan indeksimosaiikit laskettiin kerralla koko maalle. Mosaikointi ei tällöin huomioi eri maantieteellisten alueiden kasvukausien eroavaisuuksia ja ajankohtia. Tämä koskee erityisesti kasvukauden alun ja lopun ajankohtia.

### 4.3. Vertailu maaperäkarttojen turvetietoon

RF30-luokkanennusteen kykyä tunnistaa turvemaita verrattiin Maaperä 1:20 000 (GTK 2015) ja 1:200 000 (GTK 2010) Maaperä 1:20 000-aineiston kattaman alueen kaikilla maa-alueilla. Maaperä 1:200 000 aineiston maalajikuvioiden geometriatiedot ovat samat kuin Maannostietokannan (Lilja ym. 2017) maannostietokannalla, joten vertailu kattaa myös maannostietokannan. On huomioitava, että kaikki maaperäkarta-aineistot sisältävät epävarmuutta, jota on Maaperä 1:20 000 osalta kvantifioineet esim. Sunila ja Krisp (2005). Tässä työssä vertailua varten testausaineistosta erotettiin myös Maaperä-aineistoista riippumaton havaintopisteiden joukko, jota ei käytetty maaperäkarttojen teossa eikä RF-mallinnuksessa. Tämä tarkoitti testiaineistoksi eristettyä ositteen vuodesta 2009 alkaen tehtyjä maaperähavaintoja ja generoituja kivennäismaapisteitä, joita oli yhteensä 165 708 kpl (13 % turvehavaintoja) vertailua varten.

Vertailu osoittaa, että pistemäisissä kohteissa RF30-luokkaennuste ja Maaperä 1:20 000 ovat yleisen tarkkuuden osalta hyvin samanlaisia. Maaperä 1:200 000 on hieman näitä epätarkempi ja sen positiivisten ennusteiden määrää hieman pienempi (Taulukko 16). RF30-luokkaennusteen ja Maaperä 1:20 000 tasapainotettu yleinen luokittelutarkkuus on 92 %, kun Maaperä 1:200 000 tasapainotettu yleinen tarkkuus on 89 %. Maaperä 1:200 000 positiivisten ennusteiden osuus on 6–7 prosenttipistettä pienempi kuin RF30-luokkaennuste ja Maaperä 1:20 000-aineistoilla. Kaikkien aineistojen väärin positiivisten ennusteiden osuus on alle 3 %.

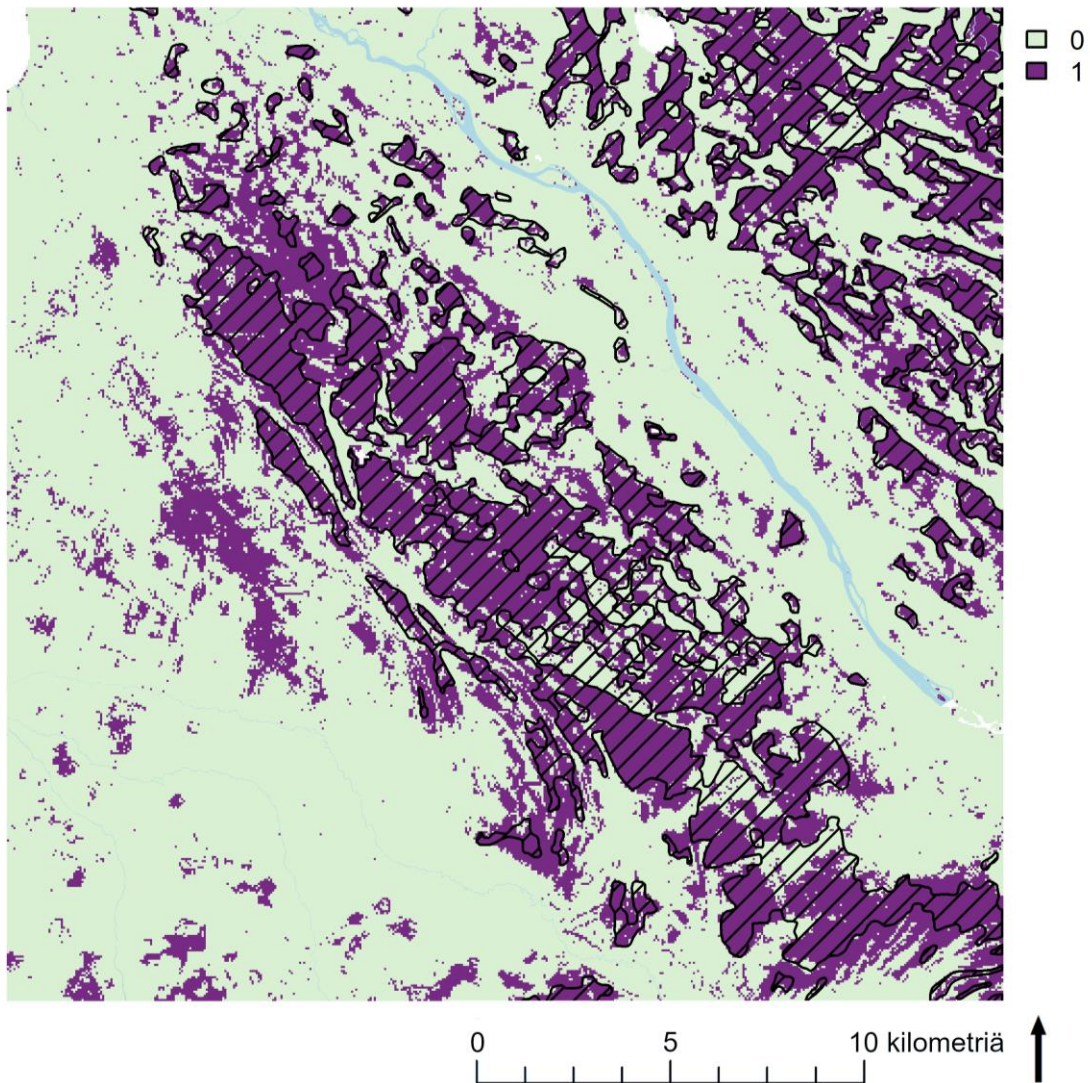
**Taulukko 16.** RF30-luokkaennusteen, Maaperä 1:20 000 (GTK 2015) ja Maaperä 1:200 000 (GTK 2010) -aineistojen maankäytöstä riippumattoman tarkkuus turvemaiden tunnistamisessa Maaperä 1:20 000 aineiston kattamalla alueella. Otantaan valittiin vuodesta 2009 jälkeen kerätyt maastohavainnot (n= 165 708 kpl), joita käytettiin aineistojen testaamiseen.

	<b>Maaperä 1:20 000</b>	<b>Maaperä 1:200 000</b>	<b>RF30</b>
Turvepaksuus	≥40 cm	≥30 cm	≥30 cm
Esiintyminen	0.117	0.126	0.126
Yleinen tarkkuus	0.967*	0.952*	0.961*
Tasapainotettu yleinen tarkkuus	0.923*	0.889*	0.924*
Kappa	0.842	0.781	0.827
TPR	0.866	0.806	0.874
FPR	0.019	0.027	0.027

\*Havaintoaineisto on epätasapainoinen (turvehavainnot 13 %), joten tasapainotettu yleinen tarkkuus on parempi metriikka kuin yleinen tarkkuus.

Huolimatta samankaltaisista tarkkuuksista RF30-luokkaanennusteen, 1:20 000 (GTK 2015) ja 1:200 000 (GTK 2010) aineistoissa, niiden välillä on myös eroja. Maaperä 1:20 000 ja 1:200 000 ovat vektoriaineistoja ja niiden pienimmän maalajikuvion koko on 2–4 ha/6,25 ha (Haavisto-Hyvärinen & Kutvonen 2007; Väänänen ym. 2010). RF30-luokkaennuste on puolestaan rasteriaineisto, jonka pienimmän maalajikuvion koko on 0,25 ha (50 m × 50 m). Maaperä 1:20 000 ja 1:200 000 perustuvat ilmakuvatulkintaan, lentogeofysikaalisen aineiston tulkintaan, korkeusmallitulkintaan ja maastohavainnointiin (Haavisto-Hyvärinen & Kutvonen 2007; Väänänen ym. 2010). Vektorimuotoiset maaperäkartta-aineistoissa maalajien rajat esitetään terävinä kartoittaneen geologin tulkitsemina rajoina. RF30-luokkaennuste puolestaan perustuu systemaattiseen numeeriseen monimuuttujamallinnukseen kaukokartoitus-, korkeusmalli-, lentogeofysikaalisten ja geologisia paikkatietoaineistoja käyttäen. Syntyneessä rasteriaineistona esitetyssä aineistossa luokkarajat kuvaavat paremmin luonnossa esiintyvä maalajien sumeaa vaihtelua, kuten Sunila ym. (2004) ovat myös esittäneet. Näin ollen RF30-luokkaennuste kuvaa turvealueiden rajoja pienipiirteisemmin sekä sisältää pienempiä turvealueita kuin Maaperä 1:20 000 ja 1:200 000.

Karttamuotoinen vertailu RF30-luokkaennusteen ja koko maan kattavan Maaperä 1:200 000-aineiston (GTK 2010) turvealueiden välillä (turvepaksuus ≥30 cm) osoittaa yleisesti yhteneväisyyttä aineistojen välillä, mutta aineistojen välillä havaitaan myös eroja (Kuva 14). Yksittäisten turvealueiden rajaus eroaa paikoin aineistojen välillä ja RF30 ennustaa laajojakin turvealueita, joita ei ole Maaperä 1:200 000-aineistossa. Cramér:n V tilastollinen assosiaatio -analyysi (Cramér 1946) osoittaa myös, että kaikki kolme aineistoa ovat keskenään erilaisia (Taulukko 17). RF30-luokkaennuste eroaa tilastollisesti Maaperä 1:20 000 ja 1:200 000-aineistoista saman suuntaisesti, mutta hieman enemmän Maaperä 1:200 000-aineistosta. Maaperä 1:20 000 ja 1:200 000-aineistot eroavat myös toisistaan, mutta eivät yhtä paljon kuin RF30-luokkaennusteesta.



**Kuva 14.** RF30-binääriennusteen ja Maaperä 1:200 000 (GTK, 2010) turvealueiden ( $\geq 30$  cm) vertailua Oulujoella ( $64.9^{\circ}\text{P}$ ,  $25.7^{\circ}\text{I}$ ). Maaperä 1:200 000-aineiston turvealueet kuvattu raidoitetuilla alueilla.

**Taulukko 17.** Cramér:n V -assosiaatiomatriisi (Cramér 1946) RF30-binääriennusteelle, Maaperä 1:20 000 (GTK 2015) ja 1:200 000 (GTK 2010) -aineistoille Maaperä 1:20 000-aineiston alueen kaikilla maa-alueilla (0=ei assosiaatiota, 1=täydellinen assosiaatio).

	<b>Maaperä 1:20 000</b>	<b>Maaperä 1:200 000</b>	<b>RF30</b>
Maaperä 1:20 000	-	0.749	0.644
Maaperä 1:200 000	-	-	0.609

RF30-binääriennusteen ja Maaperä 1:200 000 (GTK 2010) -aineiston vertailu koko maan maatalousmailla osoittaa, että maatalousmaiden turveala ( $\geq 30$  cm) on RF30-ennusteessa 8 637 ha (3 %) suurempi kuin Maaperä 1:200 000-aineistossa ja turvealat eroavat toisistaan maakuntatasolla (Taulukko 18). Suurimmat maakuntakohtaiset erot aineistojen välillä on Pohjois-Pohjanmaalla, jossa RF30-luokkaennusteen mukaan maatalousmailla on 12 247 ha (18 %) enemmän turvetta kuin Maaperä 1:200 000-aineiston mukaan, ja Etelä-Pohjanmaalla, jossa RF30-binääriennusteen mukaan maatalousmailla on -6 793 ha (16 %) vähemmän turvetta kuin Maaperä 1:200 000-aineiston mukaan. Koko maan tasolla RF30-luokkaennuste ( $\geq 30$  cm) ja Maaperä 1:200 000 (GTK 2010) -aineisto ennustavat turvetta samoille 50 m  $\times$  50 m rasterisoluille yhteensä 171 483 ha alueella ja ovat erimielisiä turpeen sijainnista yhteensä 93 095-101 731 ha alueella.

**Taulukko 18.** RF30-binääriennusteen perusteella arvioitu maatalousmaiden turveala ( $\geq 30$  cm) maakunnittain ja sen vertailu Maaperä 1:200 000 (GTK 2010) -aineiston maatalousmaiden turvealoihin ( $\geq 30$  cm). Maatalousmaa on määritelty Ruokaviraston peltolohkokisterin peruslohkojen mukaan. Yhteinen turveala tarkoittaa turve-esiintymien alaa, josta RF30-binääriennuste ja Maaperä 1:200 000 ovat samanmielisiä.

	Maakunta	Maatalousmaa-ala [ha]	RF30 turveala [ha]	Maaperä 1:200 000 turveala [ha]	Maaperä 1:200 000 ja RF30 yhteinen turveala [ha]	RF30 ja Maaperä 1:200 000 ero turvealoissa [ha]
1	Ahvenanmaan maakunta	22 092	300	201	57	99
2	Etelä-Karjala	58 852	4 091	6 994	2 964	-2 904
3	Etelä-Pohjanmaa	258 212	36 960	43 753	25 383	-6 793
4	Etelä-Savo	90 021	7 203	6 443	3 600	760
5	Kainuu	33 529	9 722	7 919	6 626	1 803
6	Kanta-Häme	109 122	4 152	6 421	2 344	-2 269
7	Keski-Pohjanmaa	60 332	19 218	17 585	13 786	1 634
8	Keski-Suomi	109 666	9 449	9 060	5 177	389
9	Kymenlaakso	89 054	1 528	3 604	1 090	-2 077
10	Lappi	57 178	26 757	18 066	15 822	8 690
11	Pirkanmaa	175 836	6 687	10 855	4 308	-4 168
12	Pohjanmaa	143 802	11 652	10 405	6 542	1 248
13	Pohjois-Karjala	98 997	16 311	11 820	8 677	4 492
14	Pohjois-Pohjanmaa	258 566	81 914	69 667	54 781	12 247
15	Pohjois-Savo	165 035	18 222	16 673	10 225	1 549
16	Päijät-Häme	88 137	1 572	3 354	982	-1 782
17	Satakunta	148 166	12 292	12 645	6 493	-353
18	Uusimaa	190 147	1 934	3 911	996	-1 977
19	Varsinais-Suomi	301 589	3 254	5 205	1 634	-1 952
	<b>YHTEENSÄ</b>	<b>2 458 332</b>	<b>273 214</b>	<b>264 578</b>	<b>171 483</b>	<b>8 637</b>



RF60-binääriennusteen ja Maaperä 1:200 000 (GTK 2010) -aineiston vertailu koko maan maatalousmailla puolestaan osoittaa, että maatalousmailla olevan paksun turpeen (>60 cm) ala on RF60-ennusteessa 27 954 ha (16 %) suurempi kuin Maaperä 1:200 000-aineistossa. Turvealat eroavat toisistaan myös maakuntatasolla niin, että erityisesti Pohjois-Pohjanmaalla ja Lapissa paksuja turvemaita esiintyy RF-ennusteen perusteella enemmän (Taulukko 19).

**Taulukko 19.** RF60 -binääriennusteen perusteella arvioitu maatalosumaiden paksun turpeen (>60 cm) ala maakunnittain ja sen vertailu Maaperä 1:200 000 (GTK, 2010) -aineiston maatalosumaiden paksujen turpeiden aloihin (>60 cm). Maatalousmaa on määritelty Ruokaviraston peltolohkorekisterin peruslohkojen mukaan. Yhteinen turveala tarkoittaa turve-esiintymien alaa, josta RF60-binääriennuste ja Maaperä 1:200 000 ovat samanmielisiä.

	Maakunta	Maatalous- maa-ala [ha]	RF60 turveala [ha]	Maaperä 1:200 000 turveala [ha]	Maaperä 1:200 000 ja RF60 yhteinen turveala [ha]	RF60 ja Maaperä 1:200 000 ero turvealoissa [ha]
1	Ahvenanmaan maakunta	22 092	262	63	20	199
2	Etelä-Karjala	58 852	2 824	6 051	2 092	-3 228
3	Etelä-Pohjanmaa	258 212	24 974	22 290	15 030	2 684
4	Etelä-Savo	90 021	5 671	4 834	2 667	838
5	Kainuu	33 529	8 056	6 562	5 476	1 494
6	Kanta-Häme	109 122	2 447	4 835	1 277	-2 388
7	Keski-Pohjanmaa	60 332	14 290	10 800	9 064	3 490
8	Keski-Suomi	109 666	7 252	6 988	3 956	264
9	Kymenlaakso	89 054	1 045	2 592	715	-1 547
10	Lappi	57 178	22 174	14 055	11 852	8 119
11	Pirkanmaa	175 836	4 562	7 764	2 893	-3 202
12	Pohjanmaa	143 802	7 476	4 882	3 437	2 594
13	Pohjois-Karjala	98 997	11 286	8 396	6 023	2 890
14	Pohjois-Pohjanmaa	258 566	60 690	43 832	36 192	16 858
15	Pohjois-Savo	165 035	13 018	12 735	7 600	283
16	Päijät-Häme	88 137	1 143	2 808	675	-1 664
17	Satakunta	148 166	9 467	7 242	4 288	2 226
18	Uusimaa	190 147	1 304	2 668	649	-1 363
19	Varsinais-Suomi	301 589	2 432	3 023	997	-591
	<b>YHTEENSÄ</b>	<b>2 458 332</b>	<b>200 372</b>	<b>172 418</b>	<b>114 899</b>	<b>27 954</b>

#### 4.4. Turve-ennusteiden käyttö

Turve-ennusteet tuotettiin ensisijaisesti turvealueiden ja turvepaksuuden tunnistamiseen maatalousmailla, Ennusteet tuotettiin kuitenkin maankäytöstä riippumatta ja aineistot soveltuvat käytettäväksi yhtä hyvin myös muiden maankäyttömuotojen alueilla. Keskeisin ero ennusteiden käytössä maatalousmailla ja muilla mailla on se, että viljelykäytössä olevien ja kuivatettujen peltojen turvekerros ohenee ajan kuluessa (keskimäärin 1,2 cm vuodessa; Räsänen ym., 2023), mikä aiheuttaa turve-ennusteiden vanhenemista pitkällä aikavälillä. Turve-ennusteet eivät anna luotettavaa tietoa turpeen paksuudesta alueilla, joissa turvekerros on ohentunut tai lisääntynyt merkittävästi. Tästä esimerkkinä ovat turpeenottoalueet. Ennusteet myös

sisältävät keskimääräistä enemmän epävarmuutta vesistöjen välittömässä läheisyydessä, sillä vesistön osuminen ennustepikselille on saattanut joissakin tilanteissa johtaa virheelliseen positiiviseen turvemaennusteeseen. Turvetta kuitenkin esiintyy yleisesti vesistöjen läheisyydessä.

Turve-ennusteet koostuvat RF10, RF30, RF40 ja RF60 -ennusteista, jotka kuvaavat nimiensä mukaisesti Random Forest (RF) mallien ennusteita turvepaksuusluokkien  $\geq 10$  cm,  $\geq 30$  cm,  $\geq 40$  cm ja  $> 60$  cm esiintymisestä 50 m  $\times$  50 m ennustepikselin tarkkuudella koko Suomen maa-alueilla. Kuhunkin ennusteeseen liittyy neljä rasterimuotoista tiedostoa (GeoTIFF) ja taulukkomuotoinen tietokenttä (Taulukko 20). Ennusteiden koordinaattijärjestelmä on ETRS-TM35FIN.

**Taulukko 20.** RF10 (turvepaksuusluokka:  $\geq 10$  cm), RF30 ( $\geq 30$  cm), RF30 ( $\geq 40$  cm), RF30 ( $> 60$  cm) -ennustekohtaiset tiedostot.

Aineisto	Muoto	Kuvaus	Tiedoston koko (noin)
RF-todennäköisyysennuste	Rasteri (GeoTIFF)	Turveluokan esiintymisen todennäköisyys pikselikohtaisilla arvoilla 0-1 (0 = 0 %, 1 = 100 %), pikselikoko 50 m $\times$ 50 m, ETRS-TM35FIN	210 Mt
RF-binääriennuste	Rasteri (GeoTIFF)	Todennäköisyysennusteesta luokiteltu turveluokan esiintyminen pikselikohtaisilla arvoilla 0 ja 1 (0=ei kuulu luokkaan, 1=kuuluu luokkaan), luokiteltu todennäköisyysennusteesta Taulukko 10 todennäköisyyden raja-arvoilla, pikselikoko 50 m $\times$ 50 m, ETRS-TM35FIN	50 Mt
RF-päätöspuiden yksimielisyys	Rasteri (GeoTIFF)	Random Forest-päätöspuiden yksimielisyys binääriennusteesta kussakin pikselissä arvoilla 50-100 (50 = puolet yksimielisiä, 100= kaikki yksimielisiä), pikselikoko 50 m $\times$ 50 m, ETRS-TM35FIN	120 Mt
RF-ennusteen kattavuus	Rasteri (GeoTIFF)	Ennusteen alueellinen kattavuus arvoilla 0 ja 1 (0= ei ennustetta ja 1 = ennuste), pikselikoko 50 m $\times$ 50 m, ETRS-TM35FIN	10 Mt
Metadatiiviite	Dokumentti (pdf)	Metadatiia täydentävä tieto, jossa kuvattuna mm. kunkin luokittelun virhematriisit ja ROC käyrät.	<1 Mt
ROC	Taulukko (csv)	Binääriennusteen luokittelussa käytettävän todennäköisyyden raja-arvon vaikutus oikeiden positiivisten ja väriiden positiivisten ennusteiden määrään. Taulukon avulla voidaan ennustetodennäköisyys-rasterista luoda uusia binääriluokitteluja, jos käyttäjä haluaa esim. minimoida väriiden positiivisten määrä eli ottaa vähemmän riskiä pienentämällä turveluokan laajuutta kartalla.	<1 Mt

Ensimmäinen tiedosto on RF-todennäköisyysennusterasteri, joka kuvaa turveluokan esiintymisen todennäköisyyttä arvoilla 0-1 (0=0 % ja 1=100 %) (Kuva 6 ja Kuva 7). Tätä jatkuvaa muuttujaa voidaan käyttää, kun käyttäjä haluaa itse luokitella ennusteet haluamallaan todennäköisyyden raja-arvolla. Aineistojen käyttäjä voi esimerkiksi haluta minimoida I- tai II-tyyppin virheitä eli vääriä positiivisia tai vääriä negatiivisia ennusteita. Tällöin pystytään minimoida kivennäis- tai multamaan virheellistä luokittumista turvemaaksi tai turvemaan virheellistä luokittumista kivennäis- tai multamaaksi. RF-todennäköisyysennusteiden luokittelu voidaan tehdä valitsemalla todennäköisyyden raja-arvo analysoimalla ROC-käyrää (Kuva 5, metadatiiviite taulukko) ja luokittelemalla todennäköisyysennuste luokkaennusteeksi valitun raja-arvon avulla. Raja-arvon valinnan vaikutusta ennustetarkkuuksiin voi tarkastella myös Kuva 10 avulla.

Toinen rasteritiedosto on RF-binääriennuste. Tämä diskreetti muuttuja kuvaa turveluokan esiintymistä arvoilla 0 ja 1 (0=ei kuulu luokkaan ja 1=kuuluu luokkaan) (Kuva 8 ja Kuva 9). Binääriennuste on todennäköisyyden raja-arvolla luokiteltu todennäköisyysennuste ja raja-arvo on valittu siten, että kokonaistarkkuus on suuri ja I- tai II-tyyppin virheet, eli väärät positiiviset ja negatiiviset ennusteet ovat tasapainossa (Taulukko 10). Raja-arvo on määritetty kohtaan, joka minimoi etäisyyden ROC-käyrän (TPR=1, FPR=0) -pisteeseen (ks. Kuva 3). Luokkaennusterasterit ovat niin sanottuja yleisennusteita. Niitä voidaan käyttää samaan tapaan kuin muitakin maalajikarttoja turvemaiden tunnistamiseen.

Kolmas tiedosto on RF-päätöspuiden yksimielisyysrasteri ja se kuvaa ennusteen epävarmuutta mallin näkökulmasta kussakin ennustepikselissä (Kuva 11 ja Kuva 12). Regressiopuita oli yhteensä sata ja ne ovat kaikki yksimielisiä rasterin arvolla 100. Käyttäjä voi arvioida luokkaennusteiden epävarmuutta yhdistämällä yksimielisyysrasteri Kuva 13 tietoihin, joka kuvaa yksimielisyyden ja ennustetarkkuuden yhteyttä. Ennusteiden epävarmuus on hyvin paikallista ja vaihtelee lyhyenkin matkan sisällä.

Neljäs rasteri tiedosto on diskreetti RF-kattavuusrasteri, joka kuvaa alueita, joille on olemassa ennuste arvolla 1 ja joille ei ole ennustetta arvolla 0. Turve-ennusteiden alueellinen kattavuus on erittäin hyvä, yli 99,5 % maa-alasta. Tietoaukot vaihtelevat ennustekohtaisesti.

RF10-ennuste kuvaa turpeen paksuusluokan  $\geq 10$  cm esiintymistä ja yleisesti erottaa turvealueet kivennäis- ja multamaista. RF30, RF40 ja RF60 -ennusteet kuvaavat turvepaksuusluokkien  $\geq 30$  cm,  $\geq 40$  cm ja  $> 60$  cm esiintymistä kussakin rasterisolussa. Turvepaksuusluokat  $\geq 30$  cm ja  $> 60$  cm vastaavat Maaperä 1:200 000 -aineiston turvepaksuusluokkia (GTK 2010) ja  $\geq 10$  cm ja  $\geq 40$  cm vastaavat kansainvälisesti käytetyn Histosol -maannoksen määrittelyssä käytettyjä turvepaksuuden raja-arvoja (IUSS Working Group WRB 2022). Eli RF-ennusteilla voidaan tunnistaa turvealueiden ja Histosol-maannoksen esiintymistä. Ennusteet eivät kuitenkaan erottele turvemaita, joiden pinnalla on ohut kerros turpeeksi luokituttamatonta maa-ainesta kuten multamaata.

RF-ennusteet ovat itsenäisiä ja toisistaan riippumattomia. Toisin sanoen syvemmän turveluokan (esim. RF30) ennusteen laajuutta ei ole rajoitettu matalamman turveluokan RF10 tuloksen perusteella. Tämä saattaa johtaa ristiriitaisuuksiin erityisesti toisiaan lähellä olevien paksuusluokkien ennusteiden välillä. Esimerkiksi RF40-ennuste voi ennustaa  $\geq 40$  cm paksun turveluokan esiintymistä, vaikka RF30 ei ennustaisikaan  $\geq 30$  cm paksun turveluokan esiintymistä samassa kohteessa.

Ennusteita voidaan käyttää rajattujen alueiden, kuten peltolohkojen ja metsäkuvioiden turpeen paksuusluokkien tunnistamiseen ja pinta-alojen laskemiseen. Tällöin aluekohtaiset pinta-alat voivat perustua valmiiseen binääriennusteeseen, jossa luokituksen epävarmuus voidaan tarvittaessa huomioida päätöspuiden yksimielisyyttä tarkastelemalla. Toinen vaihtoehto on käyttää todennäköisyysennustetta, jonka uudelleenluokittelulla voidaan tuottaa uusia binääriaineistoja halutuilla todennäköisyysrajoilla. Tätä lähestymistapaa on hyödynnetty turvepeltolohkojen esimerkissä (kappale 3.7), joka havainnollistaa todennäköisyysrajojen vaikutusta turve- ja paksuturpeisten lohkojen tunnistamiseen. Tulosten perusteella voidaan arvioida sopivinta raja-arvoa tietyn tavoitteen saavuttamiselle – jos esimerkiksi halutaan tunnistaa vähintään 90 % turvelohkojen pinta-alasta ja minimoida virheet (väärät positiiviset tunnistukset), voidaan RF30-todennäköisyysennusteesta erotella turvepikselit noin raja-arvolla 0.6 ja

valita turvepeltolohkot 50 % vähimmäispeitolla, jolloin virhetunnistukset jäävät noin yhteen prosenttiin koko turvelohkojen alasta (Taulukko 14).

Kun rasterimuotoisten mallinnusaineistojen perusteella tarkastellaan aluekohteiden ominaisuuksia, on huomioitava, että aineistojen erottelukyky (50 m x 50 m) on melko karkea pienialaisten tai muodoiltaan rikkonaisten kohteiden analyysiin. Haasteeksi muodostuu tällöin suuri reunapikselien osuus, jolloin osa pikselistä on tarkasteltavan alueen sisällä ja osa ulkona. Aineistojen yhdistämisessä on suositeltavaa käyttää menetelmiä, jotka huomioivat reunapikselien vaikutuksen suhteessa niiden alueen sisälle jäävään pinta-alaan. Tämä ei kuitenkaan poista epävarmuutta liittyen siihen, että reunoilla olevat pikselit on mallinnettu osittain tarkastellun alueen ulkopuolisia aineistoja käyttäen, eikä pikselin sisäistä vaihtelua tunneta.

Turve-ennusteiden käyttö on käyttäjän vastuulla. Ennusteiden tuottajat eivät vastaa ennusteiden käytöstä johtuvista mahdollisista vahingoista.

## 4.5. Jatkotutkimus

Tuotettuja turve-ennusteita on päivitettävä erityisesti peltomaiden osalta turvekerroksen ohenemisen vuoksi. Päivitystarpeen ajankohtaa ei ole tässä työssä arvioitu, mutta sitä voidaan arvioida pelloilla mitatun keskimääräisen turvekerroksen ohenemisvauhdin (1,2 cm vuodessa) perusteella (Räsänen ym. 2023). Turve-ennusteiden päivittämisen pullonkaulaksi tulevaisuudessa muodostuu geofysikaalisten matalalentoaineistojen tietosisällön vanheneminen peltoilla. Ajan myötä tapahtuva turpeen oheneminen, ojituksen kuivatusvaikutus ja lannoitus vaikuttavat radiometriseen säteilyyn epälineaarisesti. Nämä ilmiöt kasvattavat ennusteiden epävarmuuksia tässä työssä, mutta tulevaisuudessa ilmiöiden vaikutus on vielä merkittävämpi. Uuden radiometrisen ja sitä täydentävien lentogeofysikaalisten aineistojen kerääminen on pitkällä aikavälillä keskeistä, sillä geofysikaaliset matalalentoaineistot ovat tämän työn tulosten perusteella tärkeimpiä aineistoja turpeen paksuuden ennustamisessa. Sähkömagneettiset matalalentoaineistot ovat taas keskeisiä syvien turvekerrostumien paksuuden mallintamisessa (Laatikainen ym. 2011).

Käytetyistä 200 m linjavälillä mitatuista geofysikaalisista matalalentoaineistoista 50 m x 50 m pikselikokoon interpoloitu rasteriaineisto ei myöskään vastaa monien muiden nykyisten kaukokartoitusaineistojen tarkkuutta. Tihentämällä lentolinjaväliä voitaisiin myös parantaa tulevaisuuden turve-ennusteiden spatiaalista resoluutiota. Tämä vaatisi uuden matalalento-ohjelman toteuttamisen sähkömagneettisen ja radiometrinen mittauksen osalta.

Turve-ennusteiden päivitys edellyttää myös riittävää ajankohtaista peltomailta kerättyä turvehavaintoaineistoa, jota vasten mallit voidaan luotettavasti opettaa ja testata. Esimerkiksi, nykyisellään peltomailta tehtyjä turvekerroksen paksuusmittauksia on suhteellisen vähän koko maan kattavan mallinnuksen tarpeisiin nähden. Maastossa tehtävissä havaintoaineistojen keräämisessä olisi otettava huomioon havaintojen jakautuminen spatiaalisesti ja mahdollisesti muiden tekijöiden mukaan, sekä vältettävä uusien ja vanhojen havaintojen välistä spatiaalista autokorrelaatiota. Uusien havaintojen kerääminen on kuitenkin työlästä, sillä se vaatii maastotyön lisäksi maanomistajien luvan.

Eri mallinnusmenetelmien ja -asetelmien laajempi testaaminen, kuin mitä tässä työssä tehtiin, mahdollisesti tarjoaisi tarkempaa tai erilaista tietoa turpeen esiintymisestä ja turvekerroksen paksuudesta. Mahdollisia jatkotutkimusaiheita ovat turpeen paksuuden mallintaminen

jatkuvana muuttujana luokkamuuttujan sijasta ja suurempien turvesyvyyksien ennustaminen kuin mitä tämän raportin työssä tehtiin. Mallintaminen voitaisiin myös toteuttaa kahdessa vaiheessa siten, että ensimmäiseksi tunnistetaan turve- ja ei-turvemaan raja ja sen jälkeen turvealtaiden turvekerroksen paksuus mallinnettisiin parhaiten soveltuvilla menetelmällä. Mallinnus monimutkaisemmilla menetelmillä ja pienemmässä resoluutiossa, kuin mitä tässä työssä käytettiin, vaatii kuitenkin merkittävästi enemmän laskentaresursseja.

Tässä mallinnustyössä ei valituista mallinnusmenetelmistä johtuen kunnolla hyödynnetty satelliittiaineistojen aikasarjoja. Vaikka selittävinä muuttujina käytetyt satelliittiaineistot koostuivatkin useiden aikapisteiden havainnoista, mallit tulksivat niitä toisistaan erillisinä huomioimatta aineiston temporaalista ulottuvuutta. Tästä johtuen esimerkiksi pikselikohtaiset trendit kosteus- ja kasvillisuusindeksien kehityksessä turve- ja kivennäismaa-alueiden välillä tarkasteluajanjakson aikana jäivät huomioimatta. Satelliittipohjaisten aikasarjojen tueksi olisi voitu käyttää myös meteorologista aineistoa havainnoimaan kuivien ja kosteiden ajanjaksojen aiheuttamia eroja spektrisissä vasteissa turvesyvyysluokkien välillä. Tässä työssä tuotettua aineistoa voidaan kuitenkin myöhemmin käyttää tällaiseen tarkasteluun.

Tässä tutkimuksessa ei ennustettu multamaiden esiintymistä ja laajuutta, koska työn tavoitteena oli tunnistaa turvemaat yhtenevästi IPPC:n käyttämän määritelmän mukaisesti. IPPC noudattaa World Reference Base for Soil Resources (WRB) -luokittelua, jossa turvemaat luetaan Histosol-maannosluokkaan, eivätkä multamaat sisälly tähän maannosluokkaan. Maatalousmailla multamaat ovat usein entistä turvemaata, jossa turve on pitkälle hajonnutta ja siihen on saattanut sekoittua kyntökerroksessa muita maalajeja. Orgaanisen aineksen hajoamisen vuoksi turvemaat ajan myötä muuttuvat multamaiksi. Tämän ilmiön monitorointi peltoilla on mahdollista optisten satelliittiaineistojen avulla. Esimerkiksi Keski-Euroopassa tehdyissä tutkimuksissa optisen Sentinel-2 perusteella tehdyn koneoppimismallinnuksen on todettu tuottavan kohtuullisen tarkkoja orgaanisen hiilen ennusteita (nRMSE 8,7 %; (Castaldi ym. 2019). Menetelmä vaatii kuitenkin useiden vuosien ajan jatkuvaa satelliittikuvantamista, jotta maanmuokkaamisen jälkeinen paljas maapinta saadaan kuvattua peltolohkoilta. Menetelmän rajoitteena optisen alueen sähkömagneettisen säteilyn tunkeutumattomuus maahan. Tulkinta kertoo maan pinnan orgaanisen hiilen tai aineksen määrästä, mutta ei syvempien kerrosten tilanteesta.

Muita potentiaalisia käyttöjä aineistolle ovat esimerkiksi maa- ja metsätalousmaiden kasvuolosuhteiden tarkempi tarkastelu. Maaperän koostumus vaikuttaa muun muassa ravinteisyyteen ja hydrologiaan, mikä voi aiheuttaa eroja muilta osin samanlaisissa olosuhteissa sijaitsevien maa- ja metsätalousmaiden välillä.

## 5. Johtopäätökset

Uudet turve-ennusteet tunnistavat turvealueet korkealla tarkkuudella (89–96 %) ja ne parantavat aikaisempaa koko maan kattavaa tietoa turvemaiden esiintymisestä ja turvekerroksen paksuudesta 60 cm syvyyteen saakka. Verrattuna aikaisempiin koko maan kattaviin maaperäkarttoihin ennusteiden tarkkuus ja alueellinen erottelukyky ovat parempia ja ne tunnistavat aikaisemmin kartoittamattomia turvealueita. Turve-ennusteet mahdollistavat turvemaiden tunnistamisen kotimaisten ja kansainvälisten määritelmien mukaisesti ja ne kattavat 95,5 % koko Suomen maa-alasta. Ennusteet ja niihin liittyvä metadata on koottu Turpeen paksuus 1.0/2023-aineistoksi Geologian tutkimuskeskuksen tietovarantoon.

Uusien turve-ennusteiden perusteella laskettuna maatalousmaiden  $\geq 30$  cm paksujen turvemaiden pinta-ala koko maassa on 273 000 ha. Turvemaat kattavat 11 % maatalousmaa-alasta. Tästä alasta 73 % oli  $>60$  cm paksua turvetta. Turveala-arvio on 8 600 ha suurempi verrattuna koko maan kattavien maaperäkarttojen turvetietoihin ja turpeen alueellinen esiintyminen myös eroaa jonkin verran aikaisemmin esitetystä. Suurimmat muutokset olivat Pohjois-Pohjanmaalla, jossa maatalousmaan turveala kasvoi, sekä Etelä-Pohjanmaalla, jossa turveala pieneni. Peltolohkotasolla turve-ennusteet myös osoittivat, että turpeen esiintyminen peltolohkoilla on hyvin vaihtelevaa ja turvetta esiintyy usein vain osalla lohkoa. Uuden turpeen paksuus -aineiston avulla myös turvepeltolohkot tunnistettiin tarkasti.

Turve-ennusteisiin liittyy kuitenkin epävarmuuksia, kuten kaikkiin maaperäkarttoihin. Epävarmuuksien hallintaa varten tuotettiin tilastotietoa ennusteiden epävarmuudesta sekä epävarmuuden alueellista jakaantumista kuvaava paikkatietoaineisto. Käyttäjällä on lisäksi mahdollisuus tulkita ennusteita eri tarkoituksiin siten, että I. tai II. tyyppin virheitä, eli väärin positiivisten tai negatiivisten ennusteiden määrää, voidaan minimoida. Aineiston käytössä on myös huomioitava turvekerroksen oheneminen viljelykäytössä (keskimäärin 1,2 cm vuodessa), minkä vuoksi aineisto vanhenee ja edellyttää siten päivitystä ja havaintojen keruun uusimista pitkällä aikavälillä.

Turvemaiden digitaalisen kartoittamisen jatkokehittämismahdollisuudet sisältävät erilaisten mallinnusmenetelmien ja -strategioiden testaamista, paksumpien turvekerroksien ennustamisen testaamista, sekä satelliittiaineistojen monipuolisempaa hyödyntämistä. Paksumpien turvekerrosten ennustamisessa maaperään syvemmälle ulottuvat mittausaineistot, kuten gammasäteily- ja sähkömagneettiset mittaukset ovat merkittävimmissä roolissa.

Turpeen paksuus 1.0/2023 aineisto luo uusia mahdollisuuksia maaperään ja maankäyttöön liittyvien toimintojen suunnitteluun, ohjaukseen ja vaikutusten arviointiin, sekä tutkimukseen koko maan kaikilla maa-alueilla.



## Kiitokset

Työn rahoittivat Maa- ja Metsätalousministeriö osana Hiilestä kiinni -tutkimus- ja innovaatio-ohjelmaa (VN/16790/2021) sekä osallistuvat organisaatiot Luonnonvarakeskus, Geologian tutkimuskeskus, Paikkatietokeskus ja Ruokavirasto. Työn toteutuksessa käytettiin CSC – Tieteen tietotekniikan keskus Oy:n ja Geoportti tutkimusinfrastruktuurin (Avoin paikkatiedon tutkimusinfrastruktuuri, urn:nbn:fi:research-infras-2016072513) laskentapalveluja. Geoportti tutkimusinfrastruktuuria rahoittaa rahoittaa Suomen Akatemia, CSC – Tieteen tietotekniikan keskus Oy sekä Geoportti konsortion jäsenet.

## Viitteet

- Airo, M. (ed.) 2005. Aerogeophysics in Finland 1972–2004: Methods, System Characteristics and Applications. Special Paper 39. 197 p. Geological Survey of Finland. [https://tupa.gtk.fi/julkaisu/specialpaper/sp\\_039.pdf](https://tupa.gtk.fi/julkaisu/specialpaper/sp_039.pdf)
- Airo, M., Hyvönen, E., Lerssi, J., Leväniemi, H., & Ruotsalainen, A. 2014. Tips and tools for the application of GTK's airborne geophysical data. Report of Investigation 215. 33 p. Geological Survey of Finland. [http://tupa.gtk.fi/julkaisu/tutkimusraportti/tr\\_215.pdf](http://tupa.gtk.fi/julkaisu/tutkimusraportti/tr_215.pdf)
- Aitkenhead, M.J. 2017. Mapping peat in Scotland with remote sensing and site characteristics. *European Journal of Soil Science* 68(1): 28–38. <https://doi.org/10.1111/ejss.12393>
- Alahacoon, N. & Edirisinghe, M. 2022. A comprehensive assessment of remote sensing and traditional based drought monitoring indices at global and regional scale. *Geomatics, Natural Hazards and Risk* 13(1): 762–799. <https://doi.org/10.1080/19475705.2022.204-4394>
- Auri, J., Mattbäck, S., Boman, A., Liwata-Kenttälä, P., Räisänen, J. & Hirvasniemi, H. 2022. Happamien sulfaattimaiden yleiskartta: Loppu-raportti. GTK:n työraportti 25. 16 s. Geologian tutkimuskeskus.
- Bowers, S. 2020. sen2mosaic—Tools to aid in the production of large-scale cloud-free seasonal mosaic products from Sentinel-2 data. [Software]. <https://github.com/smfm-project/sen2mosaic>
- Breiman, L. 2001. Random Forests. *Machine Learning* 45(1): 5–32. <https://doi.org/10.1023/A:1010933404324>
- Buffam, I., Carpenter, S.R., Yeck, W., Hanson, P.C. & Turner, M.G. 2010. Filling holes in regional carbon budgets: Predicting peat depth in a north temperate lake district. *Journal of Geophysical Research: Biogeosciences*, 115(G1). <https://doi.org/10.1029/2009JG001034>
- Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M.D., Muñoz-Marí, J., García-Haro, F.J., Guanter, L., Jung, M., Gamon, J.A., Reichstein, M. & Running, S.W. 2021. A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances* 7(9): eabc7447. <https://doi.org/10.1126/sciadv.abc7447>
- Castaldi, F., Chabrillat, S. & van Wesemael, B. 2019. Sampling Strategies for Soil Property Mapping Using Multispectral Sentinel-2 and Hyperspectral EnMAP Satellite Data. *Remote Sensing* 11(3): Article 3. <https://doi.org/10.3390/rs11030309>
- Cramér, H. 1946. The two-dimensional case. *Teoksessa Mathematical Methods of Statistics*. Princeton University Press.
- Davisson, C.M. 1968. Interaction of gamma-radiation with matter. *Teoksessa Siegbahn, K. (ed.). Alpha-, Beta- and Gamma-ray Spectroscopy 1st Edition*. pp. 37–78.
- Dupigny-Giroux, L.-A. & Lewis, J.E. 1999. A Moisture Index for Surface Characterization over a Semiarid Area. *65(8): 937–945*.

- Endrestøl, G.O. 1980. Principle and method for measurement of snow water equivalent by detection of natural gamma radiation / Principe et méthode pour la mesure de l'hauteur d'eau équivalente par détection du rayonnement gamma naturel. *Hydrological Sciences Bulletin*, 25(1): 77–83. <https://doi.org/10.1080/02626668009491906>
- ESA 2023a. *Sentinel-1*. <https://sentinel.esa.int/web/sentinel/missions/sentinel-1>,
- ESA 2023b. *Sentinel-2*. European Space Agency. <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>
- Eurofins Viljavuuspalvelu 2021. Eurofins Viljavuuspalvelu Oy. <https://www.eurofins.fi/agro/>
- Evans, C.D., Peacock, M., Baird, A.J., Artz, R.R.E., Burden, A., Callaghan, N., Chapman, P.J., Cooper, H.M., Coyle, M., Craig, E., Cumming, A., Dixon, S., Gauci, V., Grayson, R.P., Helfter, C., Heppell, C.M., Holden, J., Jones, D.L., Kaduk, J., ... Morrison, R. 2021. Overriding water table control on managed peatland greenhouse gas emissions. *Nature* 1–5. <https://doi.org/10.1038/s41586-021-03523-1>
- Fernandez-Ugalde, O., Scarpa, S., Orgiazzi, A., Panagos, P., van Liedekerke, M., Maréchal, A. & Jones, A. 2022. LUCAS 2018 Soil Module (JRC Technical Reports). European Commission, Joint Research Centre (JRC). <http://doi:10.2760/215013>
- Gao, B. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3): 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Gatis, N., Luscombe, D.J., Carless, D., Parry, L.E., Fyfe, R.M., Harrod, T.R., Brazier, R.E., & Anderson, K. 2019. Mapping upland peat depth using airborne radiometric and lidar survey data. *Geoderma* 335: 78–87. <https://doi.org/10.1016/j.geoderma.2018.07.041>
- Grohmann, C.H., Smith, M.J. & Riccomini, C. 2011. Multiscale Analysis of Topographic Surface Roughness in the Midland Valley, Scotland. *IEEE Transactions on Geoscience and Remote Sensing*, 49(4): 1200–1213. <https://doi.org/10.1109/TGRS.2010.2053546>
- GTK 2010. Maaperä 1:200 000 (maalajit). Geologian tutkimuskeskus. [https://tupa.gtk.fi/paikkatieto/meta/maapera\\_200k.html](https://tupa.gtk.fi/paikkatieto/meta/maapera_200k.html)
- GTK 2013. Muinaisrantojen havainnot -metatieto. Geologian tutkimuskeskus. [https://tupa.gtk.fi/paikkatieto/meta/ancient\\_shorelines.html](https://tupa.gtk.fi/paikkatieto/meta/ancient_shorelines.html). Viitattu 22.11.2023.
- GTK 2014. Tutkitut turvealueet. Geologian tutkimuskeskus. [https://tupa.gtk.fi/paikkatieto/meta/tutkitut\\_turvealueet.html](https://tupa.gtk.fi/paikkatieto/meta/tutkitut_turvealueet.html)
- GTK 2015. Maaperä 1:20 000/1:50 000. Geologian tutkimuskeskus. [https://tupa.gtk.fi/paikkatieto/meta/maapera\\_20\\_50k.html](https://tupa.gtk.fi/paikkatieto/meta/maapera_20_50k.html)
- GTK 2017. Maaperän kerrostiedot. Geologian tutkimuskeskus. [https://tupa.gtk.fi/paikkatieto/meta/maaperan\\_kerrostiedot.html](https://tupa.gtk.fi/paikkatieto/meta/maaperan_kerrostiedot.html)
- GTK 2018. Happamat sulfaattimaat 1:250 000. Geologian tutkimuskeskus. [https://tupa.gtk.fi/paikkatieto/meta/happamat\\_sulfaattimaat\\_250k.html](https://tupa.gtk.fi/paikkatieto/meta/happamat_sulfaattimaat_250k.html)

- Haavisto-Hyvärinen, M. & Kutvonen, H. 2007. Maaperäkartan käyttöopas 61 s. Geologian tutkimuskeskus.
- Hastie, T., Tibshirani, R. & Friedman, J. 2001. *The Elements of Statistical Learning*. Springer New York Inc.
- Hijmans, R.J., Bivand, R., Forner, K., Ooms, J. & Pebesm, E. 2023. Terra: Spatial Data Analysis. R package version 1.7-29. The Comprehensive R Archive Network. <https://CRAN.R-project.org/package=terra>
- Hokkanen, K. 2005. Lounais-Hämeen muinaisranta-analyysit ja muinaisrantojen visualisointi. 7 s. Tietopalvelut, Geologian tutkimuskeskus.
- Huete, A.R. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3): 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Hyvönen, E., Turunen, P., Vanhanen, E., Arkimaa, H. & Sutinen, R. 2005. Airborne gamma-ray surveys in Finland. In: Airo, M.-L. (ed.). *Aerogeophysics in Finland 1972-2004; Methods, system characteristics and application*. Special Paper 39: 103–118. Geological Survey of Finland. [https://tupa.gtk.fi/julkaisu/specialpaper/sp\\_039.pdf](https://tupa.gtk.fi/julkaisu/specialpaper/sp_039.pdf)
- IAEA 1991. *Airborne Gamma Ray Spectrometer Surveying 323*. Technical Reports Series. 97 p. International Atomic Energy Agency (IAEA).
- IPCC 2014. 2013 Supplement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Wetlands. Hiraishi, T., Krug, T., Tanabe, K., Srivastava, N., Baasansuren, J., Fukuda, M. & Troxler, T.G. (eds.). IPCC.
- IUSS Working Group WRB 2022. *World Reference Base for Soil Resources*. International soil classification system for naming soils and creating legends for soil maps. 4th Edition. 234 p. International Union of Soil Sciences (IUSS).
- Jiang, Z., Huete, A.R., Didan, K. & Miura, T. 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment* 112(10): 3833–3845. <https://doi.org/10.1016/j.rse.2008.06.006>
- Jones, A., Fernandez-Ugalde, O. & Scarpa, S. 2020. *LUCAS 2015 Topsoil Survey, Presentation of dataset and results*. JRC Technical Reports. 83 p. European Commission Joint Research Centre (JRC). <http://doi:10.2760/616084>
- Kauth, R. & Thomas, G. 1976. The tasseled Cap—A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT. *LARS Symposia* 159. [https://docs.lib.purdue.edu/lars\\_symp/159/](https://docs.lib.purdue.edu/lars_symp/159/). [https://docs.lib.purdue.edu/lars\\_symp/159/](https://docs.lib.purdue.edu/lars_symp/159/)
- Key, C. & Benson N. 1999. Measuring and remote sensing of burn severity: The CBI and NBR. Poster abstract. *Proceedings Joint Fire Science Conference and Workshop Vol. II*: 284.
- Korhonen, K.T., Ahola, A., Heikkinen, J., Henttonen, H.M., Hotanen, J.-P., Ihalainen, A., Melin, M., Pitkänen, J., Rätty, M., Sirviö, M. & Strandström, M. 2021. Forests of Finland 2014–2018 and their development 1921–2018. *Silva Fennica*, 55(5). <https://www.silvafennica.fi/article/10662>

- Laatikainen, M., Leino, J., Lerssi, J., Torppa, J. & Turunen, J. 2011. Turvetutkimusten menetelmäkehitystarkastelu. Turvetutkimusraportti 414. Geologian tutkimuskeskus.
- Langlois, M.N., Richardson, M.C. & Price, J.S. 2017. Delineation of peatland lagg boundaries from airborne LiDAR. *Journal of Geophysical Research: Biogeosciences*, 122(9): 2191–2205. <https://doi.org/10.1002/2017JG003835>
- Lerssi, J., Säävuori, H. & Nousiainen, M. 2023. Turpeen paksuuden arviointi geofysikaalisten lentomittausten perusteella. GTK:n tutkimustyöraportti 74/2023.
- Lilja, H., Uusitalo, R., Yli-Halla, M., Nevalainen, R., Väänänen, R., Tamminen, P. & Tuhtar, J. 2017. Suomen maannostietokanta: Käyttöopas versio 1.1 (Finnish Soil Database: Manual, version 1.1) Luonnonvara- ja biotalouden tutkimus 6/2017. 68 s. Luonnonvarakeskus. Helsinki.
- Lindsay, J.B., Francioni, A. & Cockburn, J.M.H. 2019. LiDAR DEM Smoothing and the Preservation of Drainage Features. *Remote Sensing*, 11(16): Article 16. <https://doi.org/10.3390/rs11161926>
- Melanie, M. 1998. *An Introduction to Genetic Algorithms*. MIT Press.
- Minasny, B., Adetsu, D.V., Aitkenhead, M., Artz, R.R.E., Baggaley, N., Barthelmes, A., Beucher, A., Caron, J., Conchedda, G., Connolly, J., Deragon, R., Evans, C., Fadnes, K., Fiantis, D., Gagkas, Z., Gilet, L., Gimona, A., Glatzel, S., Greve, M. H., ... Zak, D. 2023. Mapping and monitoring peatland conditions from global to field scale. *Biogeochemistry*. <https://doi.org/10.1007/s10533-023-01084-1>
- Minasny, B., Berglund, Ö., Connolly, J., Hedley, C., de Vries, F., Gimona, A., Kempen, B., Kidd, D., Lilja, H., Malone, B., McBratney, A., Roudier, P., O'Rourke, S., Rudiyanto, Padarian, J., Poggio, L., ten Caten, A., Thompson, D., Tuve, C. & Widyatmanti, W. 2019. Digital mapping of peatlands – A critical review. *Earth-Science Reviews* 196: 102870. <https://doi.org/10.1016/j.earscirev.2019.05.014>
- Minasny, B. & McBratney, A.B. 2016. Digital soil mapping: A brief history and some lessons. *Geoderma* 264: 301–311. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- MML 2020. Elevation model 2 m | Maanmittauslaitos. <https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/expert-users/product-descriptions/elevation-model-2-m>
- MML 2021. Maastotietokanta. Maanmittauslaitos. <https://www.maanmittauslaitos.fi/kartat-ja-paikkatieto/asiantuntevalle-kayttajalle/tuotekuvaukset/maastotietokanta-0>
- Murphy, P.N.C., Ogilvie, J. & Arp, P. 2009. Topographic modelling of soil moisture conditions: A comparison and verification of two models. *European Journal of Soil Science*, 60(1): 94–109. <https://doi.org/10.1111/j.1365-2389.2008.01094.x>
- Murphy, P. N.C., Ogilvie, J., Connor, K. & Arp, P.A. 2007. Mapping wetlands: A comparison of two different approaches for New Brunswick, Canada. *Wetlands*, 27(4): 846–854. [https://doi.org/10.1672/0277-5212\(2007\)27\[846:MWACOT\]2.0.CO;2](https://doi.org/10.1672/0277-5212(2007)27[846:MWACOT]2.0.CO;2)

- Murphy, P.N.C., Ogilvie, J., Meng, F.-R. & Arp, P. 2008. Stream network modelling using lidar and photogrammetric digital elevation models: A comparison and field verification. *Hydrological Processes* 22(12): 1747–1754. <https://doi.org/10.1002/hyp.6770>
- Mäkilä, M., Säävuori, H., Kuznetsov, O. & Grundstöm, A. 2013. Soiden ikä ja kehitys Suomessa Turvetutkimusraportti 443. 41 s. Geologian tutkimuskeskus.
- O’Leary, D., Brown, C. & Daly, E. 2022. Digital soil mapping of peatland using airborne radiometric data and supervised machine learning – Implication for the assessment of carbon stock. *Geoderma* 428: 116086. <https://doi.org/10.1016/j.geoderma.2022.116086>
- Poggio, L., Lassauce, A. & Gimona, A. 2019. Modelling the extent of northern peat soil and its uncertainty with Sentinel: Scotland as example of highly cloudy region. *Geoderma*, 346: 63–74. <https://doi.org/10.1016/j.geoderma.2019.03.017>
- R Core Team 2022. R: A Language and Environment for Statistical Computing (<https://www.R-project.org/>). R Foundation for Statistical Computing. <https://www.R-project.org/>
- Regina, K., Heikkinen, J. & Maljanen, M. 2019. Greenhouse Gas Fluxes of Agricultural Soils in Finland. In: Shurpali, N., Agarwal, A.K. & Srivastava, V. (Toim.). *Greenhouse Gas Emissions: Challenges, Technologies and Solutions*. pp. 7–22. Springer. [https://doi.org/10.1007/978-981-13-3272-2\\_2](https://doi.org/10.1007/978-981-13-3272-2_2)
- Richardson, M. & Millard, K. 2018. Geomorphic and Biophysical Characterization of Wetland Ecosystems with Airborne LiDAR: Concepts, Methods, and a Case Study. In: *High Spatial Resolution Remote Sensing*. 39 p. CRC Press.
- Riggs, G.A., Hall, D.K. & Salomonson, V.V. 1994. A snow index for the Landsat Thematic Mapper and Moderate Resolution Imaging Spectroradiometer. *Proceedings of IGARSS '94 - 1994 IEEE International Geoscience and Remote Sensing Symposium 4: 1942–1944* vsk.4. <https://doi.org/10.1109/IGARSS.1994.399618>
- Riley, S.J., DeGloria, S.D. & Elliot, R. 1999. A Terrain Ruggedness Index That Quantifies Topographic Heterogeneity. *Intermountain Journal of Sciences* 5: 23–27.
- Rouse, J.W., Haas, R.H., Schell, J.A. & Deering, D.W. 1973. Monitoring Vegetation Systems in the Great Plains with ERTS (Earth Resources Technology Satellite). *Proceedings of 3rd Earth Resources Technology Satellite Symposium, Greenbelt 10-14 December*. SP-351: 309–317.
- Räsänen, T.A., Myllys, M., Kekkonen, H., Salo, T., Pitkänen, T., Laatikainen, M., Laine-Petäjäkangas, A., Väänänen, T., Palmu, J.-P., Kivimäki, A. & Oksanen, J. 2023. Turvepeltolohkojen määrittely ja tunnistaminen: Maatalousmaiden turvetieto (MaaTu) -hankkeen raportti. *Luonnonvara- ja biotalouden tutkimus 58/2023*. Luonnonvarakeskus. Helsinki. 40 s
- Salmivaara, A. 2020. Cartographic Depth-to-Water (DTW) index map, 2m. CSC – IT Center for Science Ltd. <https://etsin.fairdata.fi/dataset/6169811e-1f16-452b-ad6d-efec78fc51f2>.
- Shi, T. & Xu, H. 2019. Derivation of Tasseled Cap Transformation Coefficients for Sentinel-2 MSI At-Sensor Reflectance Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(10): 4038–4048. <https://doi.org/10.1109/JSTARS.2019.2938388>

- Siemon, B., Ibs-von Seht, M. & Frank, S. 2020. Airborne Electromagnetic and Radiometric Peat Thickness Mapping of a Bog in Northwest Germany (Ahlen-Falkenberger Moor). *Remote Sensing* 12(2): Article 2. <https://doi.org/10.3390/rs12020203>
- Soil Survey Staff 1999. *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys*. 2nd edition. U.S. Department of Agriculture Handbook 436. p. 886. Natural Resources Conservation Service.
- Soinne, H., Keskinen, R., Rätty, M., Kanerva, S., Turtola, E., Kaseva, J., Nuutinen, V., Simojoki, S. & Salo, T. 2020. Soil organic carbon and clay content as deciding factors for net nitrogen mineralization and cereal yields in boreal mineral soils. *European Journal of Soil Science* 72(4): 1497–1512. <http://DOI: 10.1111/ejss.13003>
- Soinne, H., Kurkilahti, M., Heikkinen, J., Euroala, M., Uusitalo, R., Nuutinen, V. & Keskinen, R. 2022. Decadal trends in soil and grain microelement concentrations indicate mainly favourable development in Finland. *Journal of Plant Nutrition and Soil Science* 185(5): 578–588. <https://doi.org/10.1002/jpln.202200141>
- Sunila, R. & Krisp, J.M. 2005. Determining the Level of certainty in Finnish Soil Maps. Conference Proceeding. 4th International Symposium on Spatial Data Quality (ISSDQ 2005), 25–26 August 2005, Beijing, China.
- Sunila, R., Laine, E. & Kremenova, O. 2004. Fuzzy model and kriging for imprecise soil polygon boundaries. Conference proceedings. 1st International Conference on Advances in Mineral Resources Management and Environmental Geotechnology, Hania, Greece.
- SYKE 2021. Sentinel-1 SAR-kuvamosaiikki (S1sar). <https://ckan.ymparisto.fi/dataset/sentinel-1-sar-image-mosaic-s1sar-sentinel-1-sar-kuvamosaiikki-s1sar>
- Tikkanen, M. & Oksanen, J. 2002. Late Weichselian and Holocene shore displacement history of the Baltic Sea in Finland. *Fennia* 180: 9–20.
- Tokola, T. & Kalliovirta, J. 2003. Paikkatietoanalyysi. Helsingin yliopiston metsävarojen käytön laitoksen julkaisuja 34. s. 185. Helsingin yliopisto.
- Tóth, G., Jones, A. & Montanarella, L. 2013. LUCAS Topsoil Survey. Methodology, data and results. JRC Technical Reports. doi: 10.2788/97922; Scientific and Technical Research series.
- Urvas, L. 1995. Viljelymaan ravinne- ja raskasmetallipitoisuuksien seuranta. *Tiedote* 15/95. 77 s. Maatalouden tutkimuskeskus.
- Virtanen, K. 1997. Lentomittausten gamma-aineisto – Menetelmä arvioida turvekerrostumien paksuutta. *Suo ja turve* 397(97): 8-11+1.
- Väänänen, T., Nevalainen, R., Liwata, P., Virkki, H. & Putkinen, H. 2010. Maaperän yleiskartointu: Loppuraportti. 22 s. Geologian tutkimuskeskus.
- Wadoux, A.M.J.-C., Minasny, B. & McBratney, A.B. 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews* 210: 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>



- Yli-Halla, M., Lötjönen, T., Kekkonen, J., Virtanen, S., Marttila, H., Liimatainen, M., Saari, M., Mikkola, J., Suomela, R. & Joki-Tokola, E. 2022. Thickness of peat influences the leaching of substances and greenhouse gas emissions from a cultivated organic soil. *Science of The Total Environment* 806: 150499. <https://doi.org/10.1016/j.scitotenv.-2021.150499>
- Åberg, A. 2013. Itämeren ylin ranta Suomessa. Pro gradu -tutkielma. Helsingin yliopisto. Geotieteiden ja maantieteen laitos. Geologian osasto. 84 s.
- Ågren, A.M., Hasselquist, E.M., Stendahl, J., Nilsson, M.B. & Paul, S.S. 2022. Delineating the distribution of mineral and peat soils at the landscape scale in northern boreal regions. *SOIL* 8(2): 733–749. <https://doi.org/10.5194/soil-8-733-2022>
- Ågren, A.M., Larson, J., Paul, S.S., Laudon, H. & Lidberg, W. 2021. Use of multiple LIDAR-derived digital terrain indices and machine learning for high-resolution national-scale soil moisture mapping of the Swedish forest landscape. *Geoderma*, 404: 115280. <https://doi.org/10.1016/j.geoderma.2021.115280>

## Liite

### Kivennäismaapisteiden generointi

Kivennäismaapisteitä (generoituja, 0 – eli negatiivisia pisteitä) luotiin lisää, koska mallinnusta varten varsinaisia havaintopisteitä mineraalimailta ei ollut saatavissa tasaisesti jakautuneena koko Suomen alueelta. Tällaisia generoituja kivennäismaapisteitä tarvittiin noin 1300 000 kappaletta. Kivennäismaapisteitä generoitiin ESRI:n ArcGIS Pro –paikkatieto-ohjelmistolla. Mallinnusta varten luotiin generoituja pisteitä maa-alueille, joilla ei ole turvetta tai sitä on alle 10 cm. Tällaiset maa-alueet ovat käytännössä kivennäismaita, jotka eivät vielä ole soistuneet tai niissä ei tapahdu soistumista. Lisäksi 0-pisteen tulee olla sellainen, että sen kohdalla ei voida katsoa olevan merkittävässä määrin tekijöitä, jotka vaikuttaisivat käytettävien opetusaineistojen laatuun. Esimerkiksi sähkölinjat vaikuttavat sähkömagneettisiin mittauksiin, teiden rungot mahdollisesti radiometriseen säteilyaineistoon tai kaikki ihmisen tuottama muutos fotogrammetriin tai lidar -aineistoihin. Näiden häiriöiden minimoimiseksi luotiin ei-haluttujen piirteiden ympärille puskurialue, jonka ulkopuolelle 0-piste voisi sijaita.

Maanmittauslaitoksen maastotietokanta (Maanmittauslaitos, 2023) on koko Suomen kattava vektorimuotoinen paikkatietoaineisto, johon on tallennettuna maastokohteiden, muotojen ja ihmisen tekemien rakennelmien tila. Kivennäismaapisteiden generoinnissa käytetty maastotietokanta oli vuoden 2012 Maastotietokanta (Haltik, Hallinnon tietotekniikkakeskus), joka ei ole enää ajantasainen rakennetun ympäristön ja peltojen suhteen. Tämän voidaan olettaa olevan etu kivennäismaita etsittäessä, koska ne alueet, jotka ovat esimerkiksi raivattu vuoden 2012 jälkeen pelloiksi ovat vielä tässä vanhassa maastotietokannassa metsämaana, suona tai soistumana. Samoin myös alueet, jotka ovat myöhemmin metsittyneet voivat olla vanhemmassa maastotietokannassa kuvattuna silloisen käyttötarkoituksen mukaan. Pelto- ja niittyalueet jätettiin kivennäismaa-aluepisteiden luonnissa pois siitä syystä, että näillä alueilla maanmuokkaaminen on muuttanut maanpinnan pientopografiaa ts. tasoittanut sitä ja toisaalta luonnontilainen kasvillisuus on korvattu viljelykasveilla, joiden perusteella ei voida kovin hyvin päätellä maaperän laatua.

Kivennäismaa-alueiden löytämiseksi oletuksena (deduktiivinen päättely) oli, että sellaiset maastotietokannan alueet, jotka eivät ole vettä, suoalueita, soistumia, rakennettuja alueita (taajamat, rakennukset, täyttömaa-alueet, tiestö, rautatiet, johto- ja putkilinjat), maa-ainesten ottoalueita, peltoja ja niittyjä, olisivat tietyn suuruisen puskurivyöhykkeen ulkopuolella hyviä 0-pistealueita eli kivennäismaa-alueita. Jokaisesta edellä kuvatusta viivamaisesta tai aluekohdeluokasta bufferoinnin jälkeen luotiin binäärinen rasteri ja lopuksi nämä rasterit kertomalla keskenään (esim. Tokola ja Kalliovirta, 2003 sivulla 124) saatiin binäärinen tulorasteri, jossa arvolla 1 olivat potentiaaliset kivennäismaa-alueet eli näillä alueilla turpeen esiintyminen oli epätodennäköistä ja arvolla 0 ei-toivotut alueet (vesistöt, soistumat, suot, maa-ainesten ottoalueet, rakennetut alueet jne.).

Maastotietokannan aineistoina käytettiin mittakaavan yli 1:15 000 aineistoja ja niiden kohteiden ympärille luotiin aineistokohtaisesti 25 m tai 50 m puskurivyöhyke ArcGIS Pro -paikkatieto-ohjelman Buffer -työkalulla. Taulukossa 1 on esitetty maastoanalyysiin valitut aineistot ja bufferivyöhykkeiden laajuus. Esimerkiksi alle 2 m levyisen ojan molemmille puolille on luotu 25 m leveä vyöhyke, jolle ei satunnaispistettä voida muodostaa. Bufferoinnin

koordinaattijärjestelmänä ja myöhemmin kaikissa muissa työvaiheissa käytettiin EU-REF\_FIN\_TM35FIN -koordinaattijärjestelmää.

Bufferoidut kohdeluokat olivat vektorimuotoisia polygoniaineistoja ja ennen rasterianalyysiä ne piti muuntaa rasterimuotoisiksi.

### Taulukko 1. Käytetyt maastotietokannan aineistot

Maastotietokannan > 1:15000 aineisto/kohdeluokka	Bufferialueen etäisyys kohteesta (m)	Valintalause kohdeluokan Properties Definition Query -ikkunassa
Taajamat	50	ei kyselyä
Rakennetut maastoalueet	50	KOHDELUO <> 'Retkeilyalue'
Liikenneväylät	50	KOHDELUO <> 'Ajopolku' And KOHDELUO <> 'Ajopolku (SILTA)' And KOHDELUO <> 'Lautta' And KOHDELUO <> 'Lossi' And KOHDELUO <> 'Pitkospuut' And KOHDELUO <> 'Pitkospuut (SILTA)' And KOHDELUO <> 'Polku' And KOHDELUO <> 'Polku (SILTA)' And KOHDELUO <> 'Talvitie' And KOHDELUO <> 'Talvitie (SILTA)'
Rakennukset	50	ei kyselyä
Virtavedet alle 2 m	25	ei kyselyä
Virtavedet 2–5 m	25	ei kyselyä
Johtoyhteydet	50	ei kyselyä
Rautatiet	50	ei siltoja, tunneleita, laivaväyliä, venereittejä
Vakavedet	25	ei kyselyjä
Kallioalueet ja maa-ainesten ottoalueet	25	KOHDELUO = 'Avoin vesijättö' Or KOHDELUO = 'Louhos' Or KOHDELUO = 'Maa-aineksen ottoalue, eloperäinen aines' Or KOHDELUO = 'Maa-aineksen ottoalue, hieno kivennäisaines' Or KOHDELUO = 'Maa-aineksenottoalue, karkea kivennäisaines' Or KOHDELUO = 'Maatuva vesialue'
Suoalueet	25	ei kyselyä
Pellot ja niityt	25	ei kyselyä

Bufferoituun polygonitason ominaisuustietotauluun lisättiin Zerovalue -kenttä (tyyppi Short integer) ja kenttään laskettiin kaikille bufferialueille arvo nolla (0). Tämän jälkeen bufferitaso muunnettiin rasteritasoksi, jossa rasterisolut bufferialueella sai arvon 0 ja muilla alueilla No-data.

## Aineiston uudelleen luokittelu

Muodostettavan nollapisteen pitää sijaita alueella, jossa ei ole minkään muodostetun bufferi-alueen aiheuttamaa rajoitetta. Tästä johtuen bufferirasterit luokiteltiin uudelleen siten, että 0-arvolla olevat alueet säilyivät nolla -arvolla ja nodata -alueet saivat arvon 1. Uudelleen luokitellussa rasterissa arvolla 1 olevat alueet ovat kunkin rajoitetason soveltuvat alueet nollapisteen sijainnille (Kuva 1)



**Kuva 1.** Bufferirasterin uudelleen luokittelu siten, että Nodata -alueet saavat arvon 1 (vihreät alueet). Nämä arvoltaan 1 olevat alueet ovat sopivia 0-pisteille. Maastotietokanta © Maanmittauslaitos.

## Laskentaketju: Nollarasterin luonti ja vektorointi

Eri rajoitetyyppien rastereista muodostettiin ArcGIS Pro ModelBuilderilla rasterianalysilaskentaketju, jossa 0/1-arvolla olevat rasterit ensin kerrottiin kaikki keskenään (esim. Tokola ja Kalliovirta, 2003 sivulla 124). Laskutoimituksen lopputuloksena kaikki solut, joissa on jossakin rasterissa arvo 0, saavat arvon 0. Arvon 1 saavat vain ne solut, joissa kaikkien eri kohderasterien solun arvo on ollut 1. Tulorasteri on siis edelleen 0/1 -rasteri, mutta 1-arvoiset eli satunnaispisteille soveltuvat alueet ovat suodattuneet kertolaskuketjun ansiosta.

Laskentaketjussa seuraavaksi tulorasteri muunnettiin polygonimuotoon ja syntyneestä polygonista valittiin gridcode = 1-arvoiset alueet. Valinta tallennettiin uudeksi polygoniksi ja leikattiin Suomen hallinnollisella aluerajalla. Tämän jälkeen Multipart -polygonit muunnettiin vielä SinglePart -muotoon eli kaikki polygonit ovat aineistoissa erillisiä alueita.

Seuraavana polygonin ominaisuustietotauluun luotiin Yli2Ha - kenttä, johon annettiin ketjun seuraavissa osissa arvo 1 yli 2 ha (20 000 m<sup>2</sup>) alueille. Nämä alueet valittiin ja niistä tehtiin uusi polygonitaso.

### **0-piste alueiden standardointi pinta-alaan perustuen**

Nollapisteille soveltuvien alueiden valinnan jälkeen näille yksittäisille polygoneille piti luoda satunnaiseen sijaintiin perustuen tasapainoisesti 0-pisteitä noin 1,3 miljoonaa kappaletta. Yksittäisiä polygoneja syntyi yhteensä 522 697 kappaletta, joten tarvittavia pisteitä tulisi olla noin 2,49 kertaisesti. Käytännössä alueiden koko vaihteli välillä 0,02 km<sup>2</sup>-358 km<sup>2</sup>, joten ei olisi kovin mielekäästä, että kaikkiin polygoneihin luotaisiin sama määrä satunnaispisteitä.

Ratkaisuna tähän ongelmaan käytettiin pinta-alan mukaan tehtävää standardointia Z-Score -menetelmällä. ArcGIS Prossa on Standardize Field – niminen työkalu, jolla voidaan tehdä tasoitus esimerkiksi kuvion pinta-alaeroihin perustuen (ArcGIS Pro Online Help, Standardize Field). Z-Score menetelmä mittaa yksittäisen arvon (yksittäisen polygonin pinta-ala) ja kaikkien arvojen keskiarvon (kaikkien polygonien pinta-alojen keskiarvo) eroa käyttäen keskipoikkeamaa (standard deviation). Menetelmä soveltuu yksittäisen arvon merkittävyyden arviointiin suhteessa koko arvojoukon jakaumaan eli miten pieni tai suuri joku yksittäinen arvo on suhteessa koko joukon arvoihin. Z-Score -menetelmä toimii parhaiten normaalijakautuneeseen aineistoon ja tässä mielessä menetelmä ei ole suositeltava tähän polygoniaineistoon sen pinta-alan suhteen vinon jakauman vuoksi. Aineistossa suurin osa polygoneista on alle 2 hehtaaria.

Huolimatta aineiston vinosta jakaumasta laskettiin alueiden pinta-aloihin perustuen Z-Score arvot. Z-Score arvoista valittiin ne, jotka olivat pienempiä tai yhtä suuria kuin 1,5 ja annettiin näille alueille Int\_Z\_Score\_area -kenttään arvo 1. Loput Z-Score arvot muunnettiin integer -tyyppiseksi, joka tekee desimaaliluvusta normaalin pyöristysäännön mukaan kokonaisluvun eli esimerkiksi luku 2,49 pyöristetään 2 :ksi, mutta luku 2,5 3:ksi.

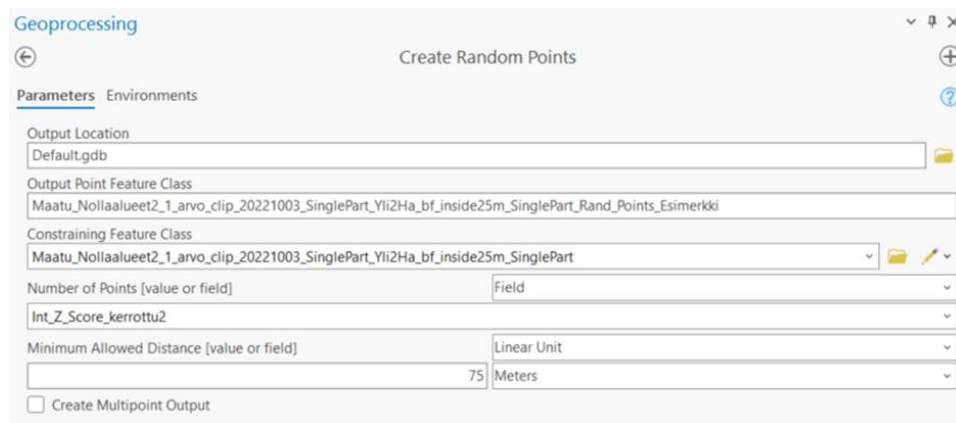
Testi 1:

1. Valittiin Shape\_Area\_Z\_Score -kentän arvot alle 1,5. Laskettiin kokonaislukuarvoiksi 1 Int\_Z\_Score\_Area -kenttään. Laskettiin Int\_Z\_Score\_kerrottu -kenttään  $\text{Int\_Z\_Score\_Area} \times 2 = 2$ . Eli kaikkiaan Int\_Z\_Score\_kerrottu -kentän lukuarvojen summa on 1 038 522 eli tämän verran tulisi satunnaispisteitä Z-Score -arvoltaan alle 1,5 polygoneihin. Kuhunkin polygoniin tulisi 2 pistettä.
2. Kertomalla Z-Score -arvoltaan vähintään 1,5 olevien polygonien Int\_Z\_Score\_Area -arvo lukuarvolla 14 saatiin näille alueille kaikkiaan 261 814 pistettä. Kuhunkin polygoniin tulee pisteitä polygonin pinta-alaan suhteutettuna 28 kappaleesta aina 3332 pisteeseen.
3. Kohtien 1. ja 2. kokonaispistemäärä on 1 300 336 kappaletta.

Polygonitason avulla tehdyt satunnaispisteet eivät kuitenkaan sijoittuneet reuna-alueilla onnistuneesti, vaan ne tippuivat paikoin alkuperäisen 0-rasterin ulkopuolelle. Tämä johtui Raster to Polygon -muunnosvaiheessa tapahtuvasta polygonin reunaviivan yleistymisestä eli viiva pyrkii oikomaan ulospäin sisäänpäin olevia kurveja. Tästä ilmiöstä johtuen polygonia bufferoitiin sisäänpäin (inside) 25 m, jotta polygonin reunaviiva hyvin todennäköisesti pysyy alkuperäisen rasteriarvon 1 alueen sisällä. Sisään bufferointi pilkkoo kuitenkin kapeikkokuvia useampaan osaan ja luo uusia polygoneja. Alkuperäisessä kohdeluokassa oli 522 697 kappaletta polygoneja ja sisäänpäin bufferoidussa 628 625 eli uusia polygoneja syntyi 75 928 kappaletta.

## Satunnaispisteiden luonti nollapolygonien luokittelun avulla

Satunnaispisteitä luotiin sisään bufferoidun (kutistetun) polygoni -kohdeluokan avulla. Yhteensä luotiin satunnaisesti sijoittuneita opetuspisteitä 1 249 620 kappaletta näihin pieniin alle 2 ha alueisiin (624 810 kpl). Vähentämällä tämä lukuarvo lukuarvosta 1 300 000 saadaan selville, että loppuihin polygoneihin pitää luoda yhteensä vähintään  $(1\,300\,000 - 1\,249\,620) = 50\,380$  pistettä. Koska loput alueet saavat kentässä Int\_Z\_Score\_Area2 arvoja välillä 2–241, haettiin kerroin, jolla vähintään tuo 50380 ylittyy kokeellisesti. Lopulta käytettiin kerrointa 5, jolloin näihin isompiin polygoneihin syntyy 105 350 kappaletta. Yhteensä satunnaispisteitä muodostuu laskennallisesti  $1\,249\,620$  (pienet alueet) +  $105\,350$  (isot alueet) =  $1\,354\,970$  kappaletta. Käytännössä pisteitä muodostuu vähemmän, koska satunnaispistegeneraattorissa annettiin vähimmäisetäisyys (75 m) yksittäisen alueen sisällä olevien pisteiden välillä. Tästä johdettujen pienissä alueissa putoaa toinen piste pois, koska se ei mahdu tuon vähimmäisetäisyyden sisälle. Työkalu antaa tästä tällöin ilmoituksen: The specified number of points could not be created in all cases due to restrictions from the minimum allowed distance. Satunnaispisteiden luontiprosessin tuloksen ohjelma teki 1 302 907 pistettä. Kuvassa 2 on esitetty satunnaispisteiden luontiin käytetyt muuttujien arvot.



**Kuva 2.** Esimerkki satunnaispisteiden luomisesta rajoittavan polygonitason kenttätietojen avulla. Kentässä Int\_Z\_Score\_kerrottu2 arvot ovat välillä 2–1 205 eli pieniin kuvioihin työkalu pyrkii luomaan 2 pistettä ja suurimpaan 1 205 pistettä.

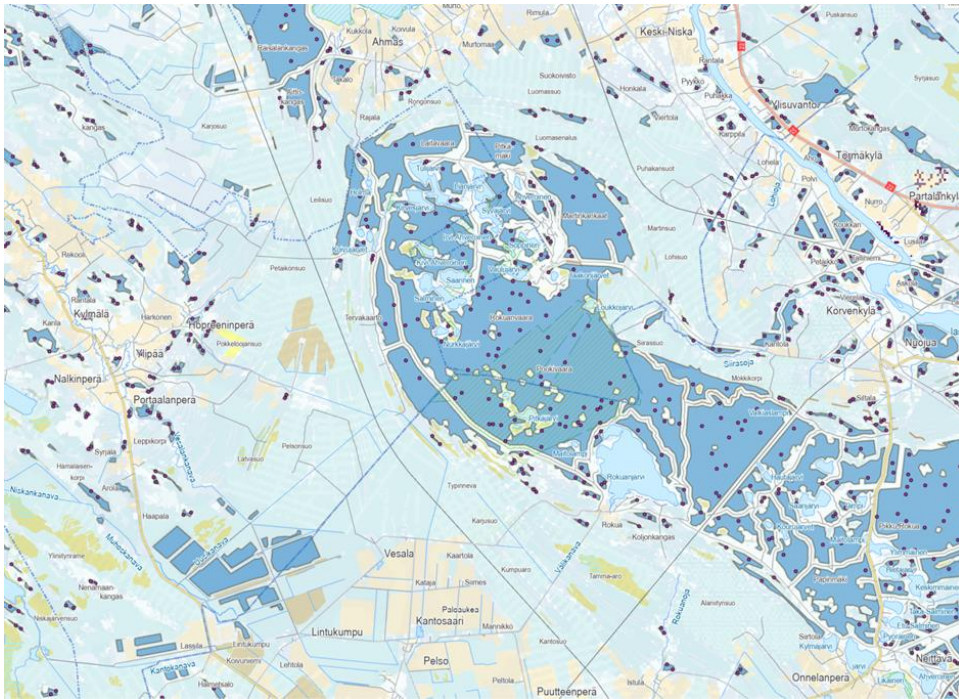
## Satunnaispisteiden tarkistus

Satunnaispisteaineisto luotiin Maanmittauslaitoksen maastotietokannan avulla sellaisille alueille, joilla ei ole soistumaa, suota tai vettä ja jotka ovat sopivan puskurietäisyyden päässä mahdollisesta häiriötä aiheuttavasta tekijästä (tiestö, rautatiet, rakennukset ym.). Maastotietokanta ei erottele turvekankaita kivennäismaa-alueista ja mahdollisesti myös muita eloperäisiä maita saattaa sijaita satunnaispisteiden paikalla. Tällaisia pisteitä etsittiin olemassa olevien maaperäkartojen avulla, erityisesti 1:20 000/1:50 000 mittakaavaisia aineistoja käyttäen. Lisäksi poistettiin tai siirrettiin lieju-, turve, vesistö, täytemaa- ja kartoittamattomien alueiden pisteet maaperäkarta-aineistojen ominaisuustietotauluihin tehtyjen hakujen avulla.

Valituilla alueilla valittiin spatiaalisella haulla satunnaispisteet, jotka osuivat näille alueille ja tämän jälkeen valitut tietueet poistettiin generoidusta kivennäismaapisteaineistosta. Valintojen ja osittain visuaalisen tarkastelun jälkeen lopulta mukaan otettiin 1 282 425 pistettä (poistoon meni noin 20 000 pistettä). Tähän kohdeluokkaan tehtiin MaaTu -projektin mukaiset



attribuuttikentät ja lopuksi kohdeluokka muunnettiin ArcView -shapefile muotoon. Kuvassa 3 on ote generoiduista ja tarkistetuista mineraalimaapisteistä Rokuanvaaran ympäristöstä.



**Kuva 3.** Generoituja mineraalimaapisteitä Rokuanvaaran ympäristössä.





**Löydät meidät  
verkosta**

**luke.fi**

