HRL verification report for tree cover density in Finland

I. Administrative part

HRL	Tree Cover Density
Country (and region, if regions are verified separately)	Finland
Institution carrying out the work	Natural Resources Institute Finland (Luke)
General overview of data quality done by	Hanna Huitu, Research Scientist, hanna.huitu@luke.fi Matti Katila, Research Scientist, matti.katila@luke.fi Sakari Tuominen, Senior Scientist, sakari.tuominen@luke.fi
Look-and-feel analysis done by	Hanna Huitu
Statistical verification done by	Hanna Huitu Matti Katila
In situ data used	National forest inventory (NFI) field plots, from systematic cluster sampling 2013, except for northern Lapland (see Fig. 2 for sampling regions) 2012 (n=9766). All land cover types (except sea) are represented in the data. Additional inspection carried out with NFI field plots 2014-2016, with data set covering national forestry land (n=13729)
	Finnish multisource-NFI thematic map of canopy cover 2015
	False Colour Aerial Photographs from National Land Survey, year 2015 (used as a WMS layer)
	Topographic database of National Land Survey
	Corine CLC2012 land cover map
Internal quality control done by (name, position and e-mail)	
Date and place of writing the report	8.2.2019

II. General overview of data quality

Results of the general overview of data quality (obligatory)	*High Resolution Layer for Tree Cover Density (2015) is a raster layer in 20 m resolution, with tree cover density values varying between 0-100 within the classified area.* *The data set subject to verification was a binary raster layer, with tree cover density in the given pixel being equal to or higher than 30 %.* *In order to improve the quality of assessment, this report includes inspection of the values in the original data layer as well.* ***Classification error matrices** together with TCD error analytics are placed to the end of part IV statistical verification (Table 1. and Table 2.)* ***Geometric accuracy:** Level was geometric accuracy was good. Based on overlaying this product with topographic data layers, the product was not found to contain shifts or other major problems of geometric accuracy.* ***Thematic accuracy:** In general, tree covers were detected reasonably well. Some areal patchiness was found. In some contexts the spatial detail of a landscape was too high to be captured with the 20 m pixel size (see part III look and feel).* ***Errors of commission and omission** are assessed in detail in parts III and IV.* ***Issues found in this verification:*** *i. Distribution of tree cover density values was not realistic to Finland (overly condensed around value 75 %)* *ii. High commission error (26 %) was present on all development stages of forestry land (see IV)* *iii. Areal distribution of tree cover density values showed patchiness that was not explained by our in situ (ground truth) data, and systematic over-estimation error that increased northward (see Fig.4 and Fig.5)* **II.i Distribution of tree cover density values** Distribution of tree cover density values in the high resolution layer does not agree with the canopy cover predictions for the plots in the national forest inventory (see histograms in Fig. 1). Especially the left tail of the distribution (values < 25 %) is almost non-existent for the high resolution layer.

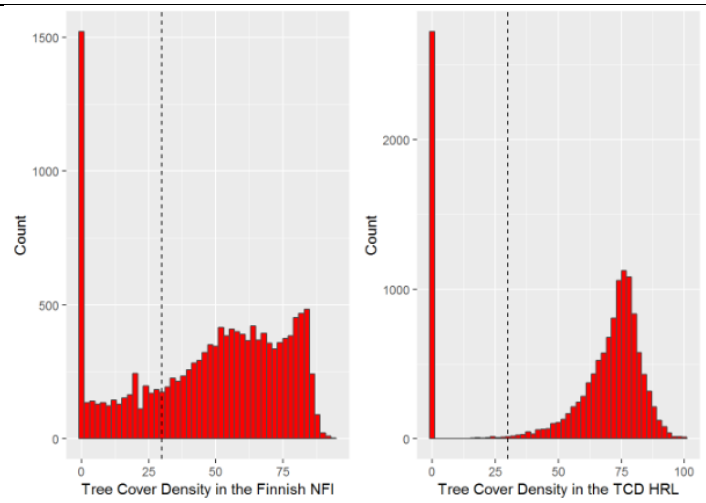

Fig.1 Histogram of the tree cover density value on forestry land, based on estimates for NFI field plots (left) and in the HRL layer (right). Dashed line denotes 30 % threshold used in statistical verification.

The canopy cover for the NFI11 and NFI12 field plots was estimated using the NFI10 field plots, for which the canopy cover had been visually assessed, see section IV. Measurement of canopy cover on the field was done visually -although supported by crown diameter tables- and there may be differences between the observers which can generate variation to the estimates. The magnitude of the error in the NFI canopy cover predictions we used as ground truth is not well known and conclusions must be drawn accordingly. However, it is clear that the HRL layer does not capture the northwards decline of tree cover density (assessed within tree species and development classes) which is clearly visible in ground truth data. Secondly, it practically does not contain very low values of tree cover density, which are visible in ground truth data, and commonly observed on-site. Third, the histogram peak (mode) location of the TCD layer at the 75 % tree cover lacks empirical evidence, especially as the mode value and shape of tree cover density distribution remains notably similar despite different forest ages and types. Thus, we conclude that the value distribution in the HRL layer is erroneous for Finland.

As the HRL verification is carried out on a thresholded, (CC>=30%) binary variable instead of the original (continuous) variable, this error is not fully captured by the verification results in tree cover density.

II.iii Patchiness and systematic errors in tree cover density values

Visual inspection of the TCD HRL layer revealed odd looking patches of size of a municipality or larger on the map, see an example in Fig. 3. The forest map data from the Multi-Source National Forest Inventory (MS-NFI) carried out by Natural Resource Institute Finland (http://kartta.luke.fi/) was compared to the TCD HRL product. The canopy cover raster layer from

	MS-NFI-2015 with a spatial resolution of 16*16 m2 was used. The MS-NFI is based on medium resolution satellite images, mainly Landsat series satellites, numerical map data and NFI field plot data. A new image was created by subtracting the MS-NFI-2015 from the HRL TCD image on the pixels with common thematic cover (MS-NFI-2015 – TCD). Tendency of the HRL layer to overestimate tree cover density values (particularly in the west and north of Finland) when compared to the national product is visible in the Fig.4. The difference image is presented scaled between -50 % to 50 % tree cover density using rainbow colors (blue-green-red) LUT, which means that the green color is approximately near zero difference and blue is overestimate and red is underestimate from the HRL TCD. An example of northward increasing overestimation is presented in Fig. 5, where tree cover density estimates for one forest development stage (Young thinning stand) are plotted as a function of geo-coordinates in north-south direction.

III. Look-and-feel (obligatory)

Stratum	Name of the stratum (see proposed strata in Tables 17-21)	Number of samples verified	
1	Urban vegetation (Trees in parks, cemeteries, etc.)	10	**Acceptable.** Open green areas (with little or no tree cover) were often classified correctly. Some cases found where open fields falsely classified as tree cover. Raster cell size (20 m) is large in context of small-scale urban landscape, posing challenge to the classification.
2	Trees in sport and recreation areas	10	**Acceptable**. Both open green areas and groups of trees were often classified correctly. Some cases where open fields had been erroneously classified as tree cover
3	Orchards, fruit trees	5	**Acceptable.** Trees in the orchards are often found correctly, some omission error in form of scattered, erroneously classified pixels. In Finland share of this stratum from all tree-covered strata is small.
4	Forest along rivers & lakes	5	**Acceptable**. Scattered pixels of both omission and commission errors found. Use of a lake mask layer (topographic map) recommended – however small-scale details in landscape cannot be captured with the given raster cell size.
5	Coastal forests	5	**Acceptable**. Some commission errors with coastal meadows, tree cover on rocky shores detected fairly well (however, seems that in the original data layer (with tree cover continuous 0-100 %), low < 30 % tree cover on rocks is not detected for inspected cases.
6	Agricultural areas with scattered small forest patches (if ≥ 0.5 ha)	10	**Acceptable**. 10/11 of the inspected ≥ 0.5 ha forest patches were found (had at least one pixel with forest cover). Scattered pixels of both omission and commission error were present however.
7	Non-tree woody vegetation (Transitional woodland-shrub, moors and heathland, sclerophyllous vegetation)	10	**Insufficient.** Checked based on in situ information on the existing tree cover. Several commission errors in cases with very low or zero tree canopy ground truth, mostly in the northern vegetation zones (Lapland inventory area).**See example in V for details**.
8	Wetland	10	**Insufficient.** Patchy, detailed textures of vegetation and water on wetlands may resemble tree canopies, and most error-prone locations were visually looked up and checked for commission errors. Some commission errors were found. **See example in V for details.** Use of a peatland mask (particularly for open mires) is recommended to support future production efforts.
A1	Peatland areas with scattered small forest patches (if ≥ 0.5 ha)	5	**Insufficient.** While most of the inspected ≥ 0.5 ha forest patches were found, the classification was not able to sufficiently distinguish them from the surrounding peatland area of no tree cover, but commissions were occurring. Use of a peatland mask is recommended to support future production efforts.
A2	Peat production areas	5	**Acceptable.** Peat production areas in use were correctly classified as not having tree cover, and forest re-growth on areas no longer in use was often detected correctly.
A3	Seedling stands	5	**Insufficient.** Inspected cases of young forest were partly classified erroneously as not having tree cover. Classification based on remote sensing material is difficult, so for this assessment in situ measurements were used.
A4	Power lines and local roads	10	**Acceptable.** Many sporadic omission and commission error pixels were found, and lines of infrastructure were typically not continuous. However, level of detail in landscape was so fine on this stratum, that the 20 m resolution in the HRL product cannot be expected to capture it effectively.
Overall evaluation			Acceptable
Comments			Typical locations of misclassification (and likely error prone locations) are presented above. Classification

	performance was weaker (insufficient/acceptable) on the nationally selected strata, and better on the recommended strata. Overall, tree cover density for majority of area is reasonably well predicted.
VERY IMPORTANT:* In case of critical findings and to allow traceability, *please, document errors, together with* justifications/explanations/meaningful examples & screenshots, *in section V of this document (see instructions in Ch. 6.3. in Guidelines)	

IV. Statistical verification[1]

For statistical verification of the HRL forest layers, there is an extensive field sample available based on systematic cluster sampling. Firstly a set of plots (i) from the 11th National Forest Inventory from year 2013 (NFI11) covering all the land use classes forest land, built-up, arable land, roads and power lines and inland waters (https://www.luke.fi/en/natural-resources/forest/forest-resources-and-forest-planning/forest-resources/). The northernmost Lapland was an exception, the field sample was selected based on double sampling with stratification and originated from the year 2012 (NFI11). Secondly, there was also more up-to-date NFI12 data (ii) from years 2014-2016, for which the crown cover was available only for the national forest land field plots.

The set (i) contained 9766 field plots selected for quantitative verification. All the field plots on land and inland water were included. In order to follow verification guidelines regarding minimum sampled patch size (section 5.3.), it was also required that minimum distance to the nearest stand boundary was 20 m on national forestry land and 12.5 m on non-forest land. The radius of the of the NFI 11 field plot is 12.52 m or 12.45 m in South Finland and North Finland, correspondingly.

The set (ii) contained 13729 field plots located on the forestry land (national definition, see Tomppo, E., Heikkinen, J., Henttonen, H.M., Ihalainen, A., Katila, M., Mäkelä, H.,Tuomainen, T. & Vainikainen, N. 2011. Designing and conducting a forest inventory - case: 9th National Forest Inventory of Finland. Springer, Managing Forest Ecosystems 21. 207 p.). Only the plots where the distance to the nearest stand boundary was at least 20 m were included to the set. The radius of the filed plots in the NFI12 is 9 m.

For both sets (I) and (ii), those field plots where a drastic change of land cover or a clear-cut of forest had occurred between the field measurement date and assumed image acquisition date (30.6.2015 was assumed for the HRL product) were removed using MS-NFI2015 satellite images, and land use change monitoring data from Greenhouse gas reporting project.

The canopy cover percentage was readily modeled for the field plots on the land use classes “forest”, “poorly productive forest” land and “unproductive land” (Mäkisara, K., Katila, M., Peräsaari, J. & Tomppo, E. 2016. The Multi-Source National Forest Inventory of Finland -methods and results 2013. Natural resources and bioeconomy studies 10/2016, Natural Resources Institute Finland. 215 p. http://urn.fi/URN:ISBN:978-952-326-186-0, sect. 3.2.1). Originally, the canopy cover was visually assessed for the NFI10 field plots (for a test of the accuracy of different assessment methods in the field see Korhonen L., Korhonen K.T., Rautiainen M., Stenberg P. 2006. Estimation of forest canopy cover: a comparison of field measurement techniques. Silva Fennica vol. 40 no. 4. https://doi.org/10.14214/sf.315), and predicted for NFI11 and NFI 12 forestry land field plots using NFI10 data. The canopy cover for deciduous trees was computed from the canopy cover according to the proportion on of deciduous trees in the field plot. In NFI11 trees were also tallied on field plots outside forestry land, and the canopy cover was predicted using a statistical model estimated using NFI10 field plots on forest land mineral soils. On the plots outside forestry area, the tree species dominance was defined based on basal area of the tree species tallied.

[1] not relevant for Grassland product, and also not relevant for permanent/temporary wet, and temporary water classes of WAW product

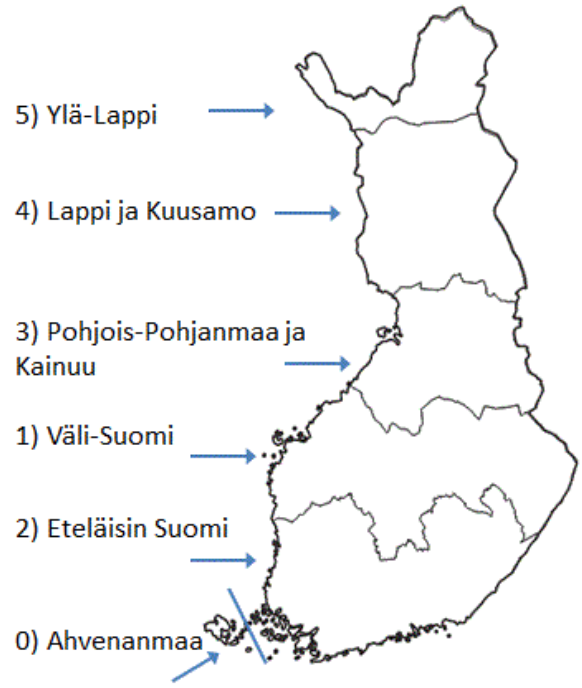

Fig. 2 Sampling regions for the Finnish National Forest Inventory.

Quantitative error estimates were reported according to the Copernicus verification guidelines version 1.4. Additionally, many of the estimates were presented for the sub-areas formed by the sampling regions used in Finnish NFI (Fig. 2).

As the NFI11 data set covered the whole country and was based on systematic sampling, it was found reasonable to calculate the omission and the commission errors solely based on the confusion matrices from NFI field plot points of set (i) between NFI canopy cover class vrs. TCD class from the HRL layer. The HRL tree cover density was validated also as a continuous variable, as canopy cover percentage was available for the NFI field plots: e.g. RMSE and BIAS of the HRL of the tree cover density estimates were calculated (See Table 2.).

Stratification	no stratification
Comment on stratification	Field measurements from the national forest inventory (NFI) were used as ground truth data in this verification. NFI is based on systematic cluster sampling over all land use classes and ownership types. Number of field plots per area decreases towards north. The country is divided into six inventory areas (Fig.2), and results are presented also for these sub-regions. In Finland, over 78 % of the land area is covered by forestry land, and tree cover is found also on other land use classes for the NFI field plot set (i). Due to sampling methodology and high prevalence of the class to be inspected, no stratification was used.
Number of random samples for finding omission errors	*3927* (NFI plots with CC >= 30 %)
Number of valid (applicable) samples for finding omission errors	*3927* samples by inventory regions: Ahvenanmaa 177 Väli-Suomi 1295 Eteläisin Suomi 1196 Pohjois-Pohjanmaa ja Kainuu 621 Lappi ja Kuusamo 393 Ylä-Lappi 245
Omission error (%)[1] with uncertainty	6.06 %, uncertainty 0.75 % omission error by inventory regions: Ahvenanmaa 9.0 % 4.22 % Väli-Suomi 5.6 % 1.26 % Eteläisin Suomi 8.8 % 1.60 % Pohjois-Pohjanmaa ja Kainuu 4.7 % 1.66 % Lappi ja Kuusamo 2.8 % 1.63 % Ylä-Lappi 1.6 % 1.59 %
Comment on omissions	OK. Errors of omission (area erroneously classified as having tree cover density <= 30 %) were not very common. Southern Finland and Ahvenanmaa sampling areas seemed more error prone than other areas. Producer's accuracy was 0.94 and the goal level set for the data was thus achieved in Finland.
Number of random samples for finding commission error	*4999*
Number of valid (applicable) samples for finding commission error	*4999* Ahvenanmaa 218 Väli-Suomi 1429 Eteläisin Suomi 1283 Pohjois-Pohjanmaa ja Kainuu 718 Lappi ja Kuusamo 663

	Ylä-Lappi 688
Commission error (%)[2] with uncertainty	26,2 % uncertainty 0.62 % Ahvenanmaa 26.1 % 5.83 % Väli-Suomi 14.5 % 1.82 % Eteläisin Suomi 15.0 % 1.95 % Pohjois-Pohjanmaa ja Kainuu 17.5 % 2.78 % Lappi ja Kuusamo 42.4 % 3.76 % Ylä-Lappi 65.0 % 3.56 %
Comment on commissions	Not OK. Error of commission (an area erroneously classified as having tree cover density value >=30 %) were very common. In approximately 35 % of the cases, ground truth value of tree crown cover was zero. Commission errors were most common in northern Finland, where forests in general are sparser than in more southern locations. The tree limit is reached in the north-ernmost sampling region (Ylä-Lappi), where commission error is really large and implies to overall poor classification accuracy
Overall evaluation	The error diagnostics of the tree cover density as continuous variable are presented in Table 2. below. The RMSE of the TCD, 25 % is larger than that reported for the MS-NFI2011 canopy cover layer (14-20 %) in metadata (http://kartta.metla.fi/). However, the MS-NFI2011 was validated using all the field plots within forestry land which significantly increases the error estimate compared to using only field plots at minimum 20 m from stand boundary in current validation. There is a significant overestimation (bias) of tree cover density (percentage units) compared to NFI field plot canopy cover. There is overestimation, 12 %, which increases towards north of Finland (Lappi ja Kuusamo, Ylä-Lappi sampling regions). This systematic error increases also the RMSE values.

[2] User's accuracy (%) = 1 – commission error (%)

Table 1 Classification error matrices for tree cover density

TRUE / ESTIMATED	TCD < 30 %	TCD>=30 %	TOTAL
TCD < 30 %	**4 529**	**238**	4 767
TCD>=30 %	**1 310**	**3 689**	4 999
TOTAL	5 839	3 927	9 766

TRUE / ESTIMATED	TCD < 30 %	TCD>=30 %	TOTAL
TCD < 30 %	**46.4 %**	**2.4 %**	48.8 %
TCD>=30 %	**13.4 %**	**37.8 %**	51.2 %
TOTAL	59.8 %	40.2 %	100.0 %

Table 2 Error diagnostics of the tree cover density as continuous variable, whole country and by inventory regions. The NFI mean, absolute and relative RMSE and bias (TCD – NFICC), the standard error of the bias, the standard deviation of the NFI field variable and R^2 coefficient (the proportion of the variation explained by the classification). (see Katila & Tomppo (2001) http://dx.doi.org/10.1016/S0034-4257(00)00188-7 Table 4 for more explanation).

Region	No. of plots	NFI mea	RMSE	RMSE (%)	BIAS	BIAS (%)	2*stdE BIAS	NFI st.dev.	R^2
Total	9766	24.2	24.9	102.6	12.30	50.76	0.44	26.4	0.11
Ahvenanmaa	606	19.0	21.6	113.6	6.67	35.06	1.67	26.3	0.32
Väli-Suomi	2759	27.9	23.4	83.8	10.86	38.94	0.79	28.7	0.34
E-Suomi	2948	24.1	21.6	89.5	7.92	32.82	0.74	28.6	0.43
Pohjois-Pohjanmaa ja Kainuu	1334	26.9	25.0	92.7	13.02	48.33	1.17	26.0	0.07
Lappi ja Kuusamo	986	23.3	31.0	133.1	21.67	93.13	1.41	19.4	0.0
Ylä-Lappi	1133	16.0	31.0	193.8	21.21	132.51	1.35	16.0	0.0

V. Documentation of errors and critical findings.

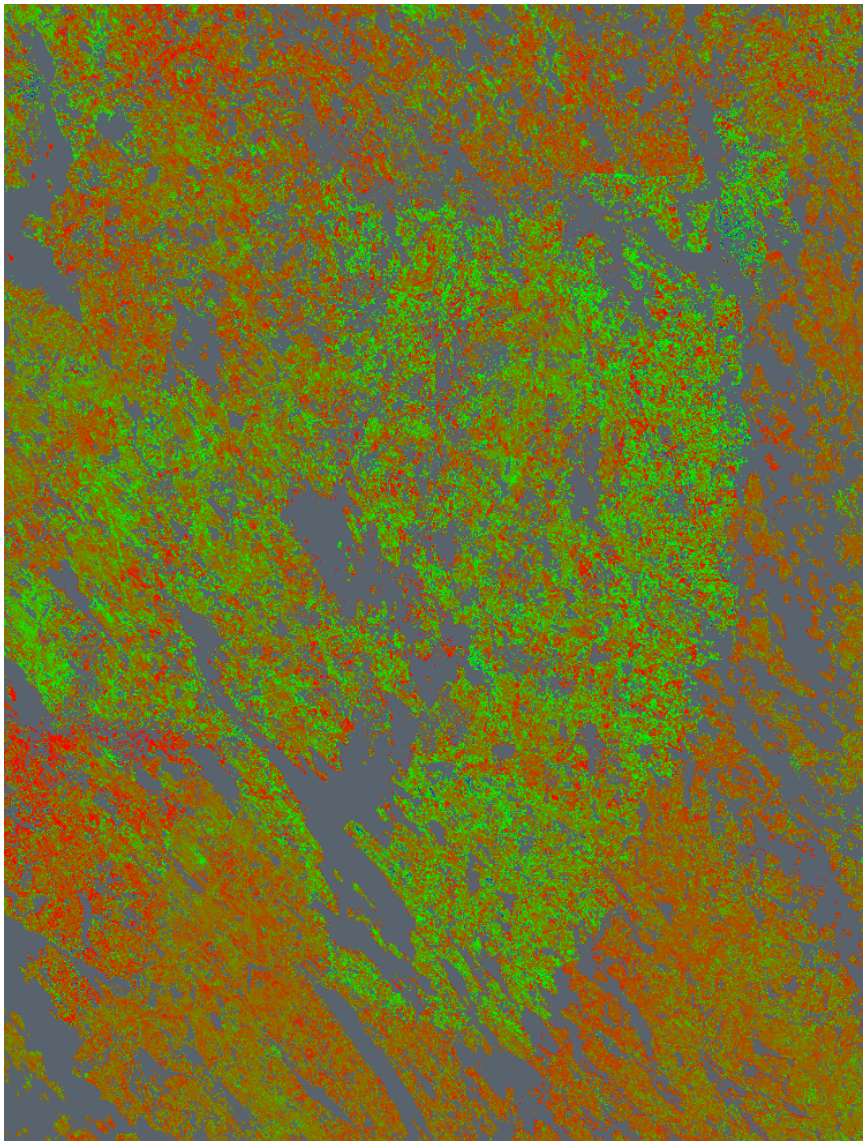

Fig. 3 A zoom of the HRL TCD from Central Finland, rainbow colors (blue-green-red) LUT scaled 1-100, 0 in grey color. Polygon 7053259, 514292; 6978083, 491357; 7015034, 446762; 7052303, 473201 in TM35-FIN geoy, geox coordinate pairs.
(II general overview of data quality)

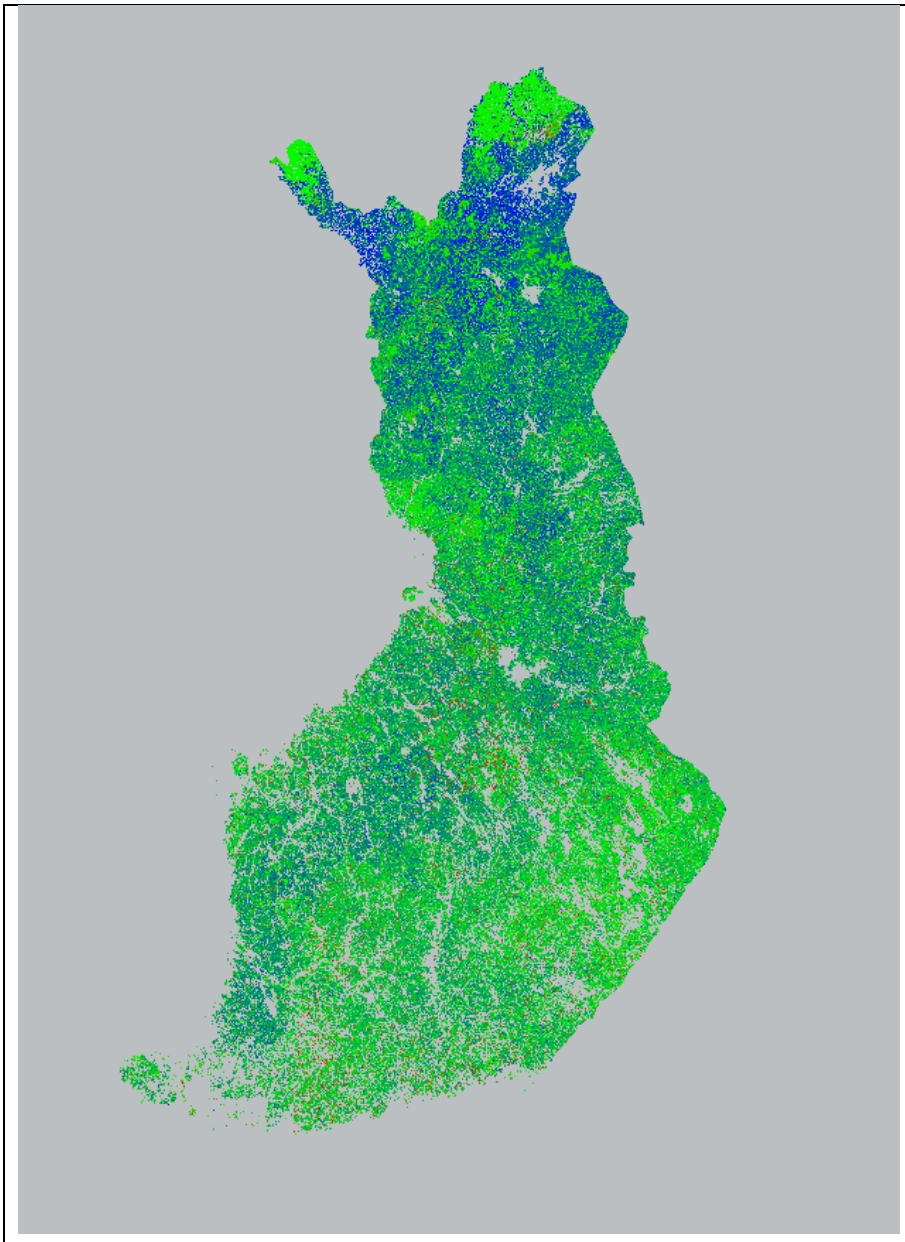

Fig. 4 A difference image between HRL TCD and the MS-NFI-2015 crown cover thematic map (national data) (MS-NFI-2015 – TCD). Rainbow colors (blue-green-red) LUT scaled between -50 – 50 values.
(II general overview of data quality)

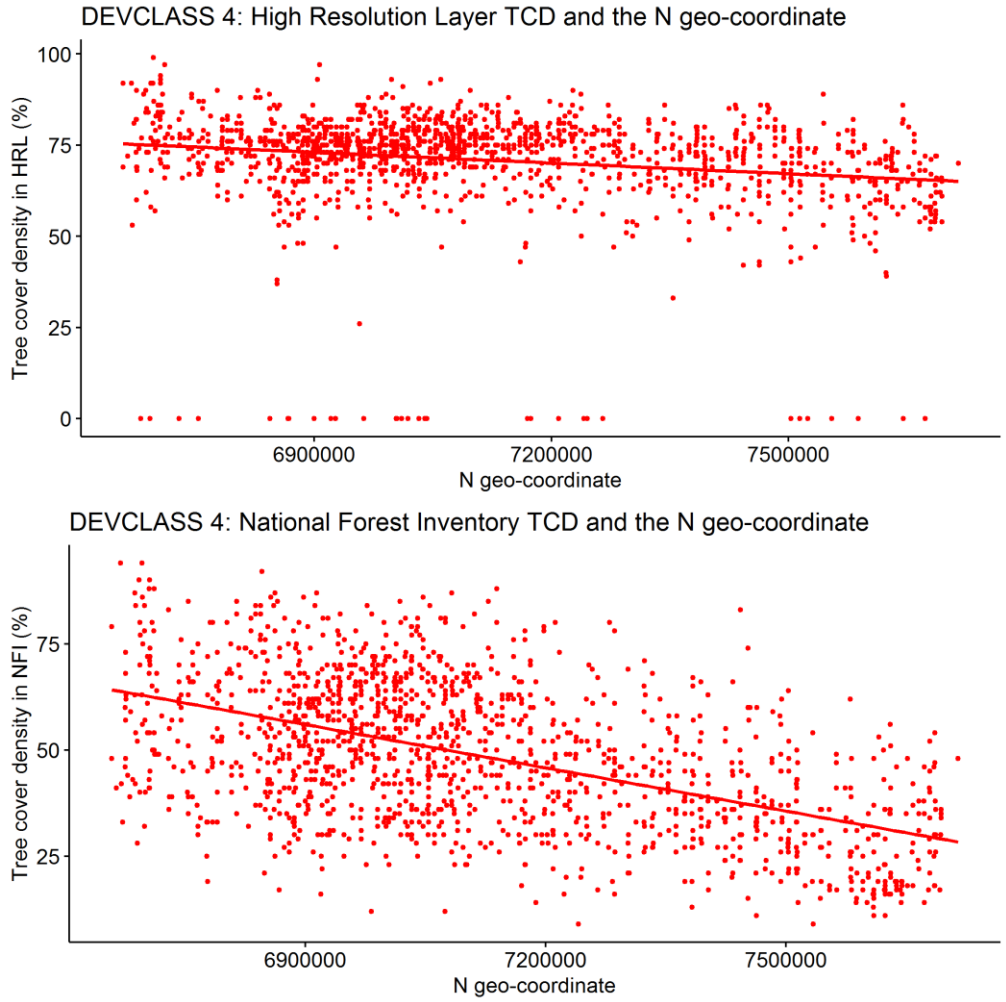

Fig. 5 Tree cover density (%) for the study plots in forest development class 4 (young thinning stand), shown as a function of geo-coordinates (km) north-south direction in ETRS-TM35-FIN coordinate system. Top graph for Forest HRL Tree cover density; bottom graph for results in the National Forest Inventory for the same set of locations.

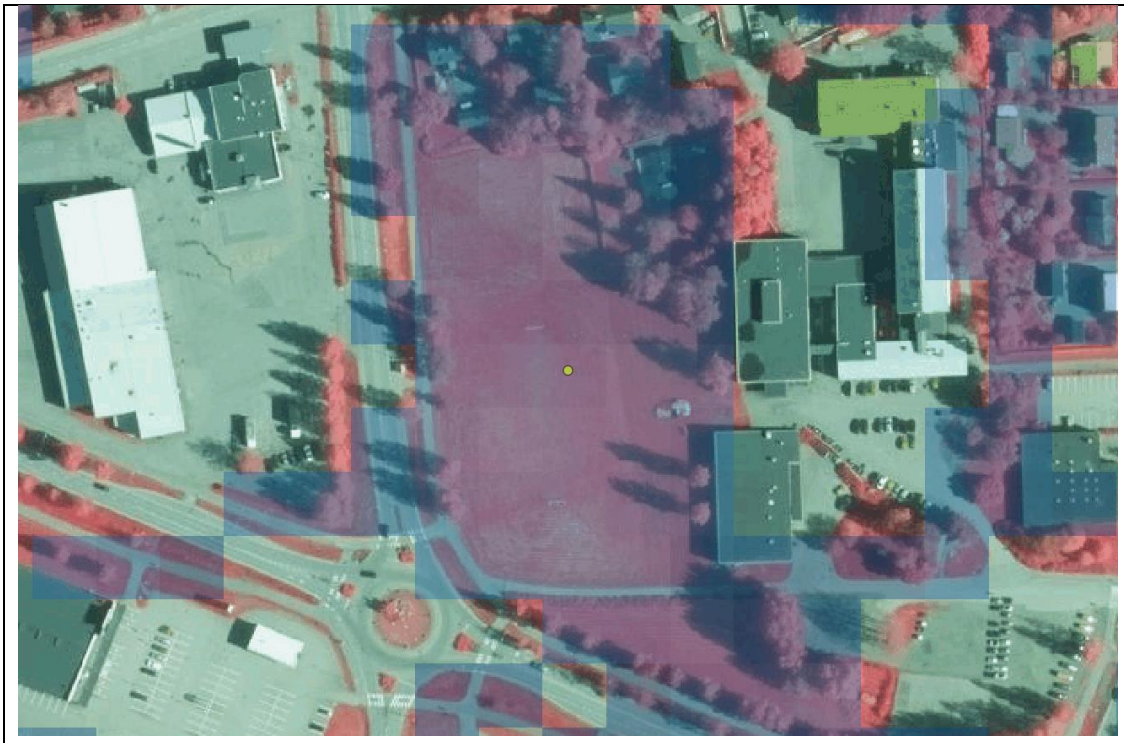

Fig. 6 Commission error on urban area, urban grass field classified as tree cover. TCD raster layer (blue) laid over aerial imagery (6800312, 250448).

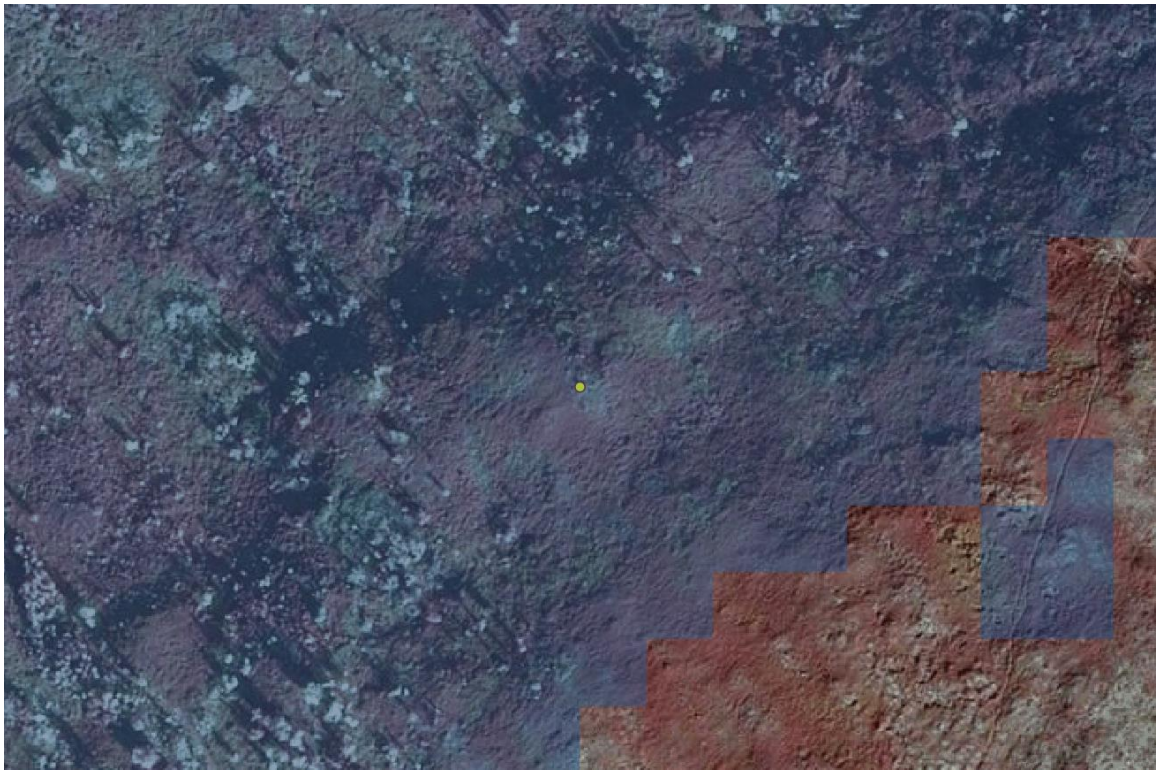

Fig. 7 Commission error for bushy non-tree vegetation in Northern Finland. Also the area where no tree tops are present was erroneously classified as tree cover. (7774815, 533900)

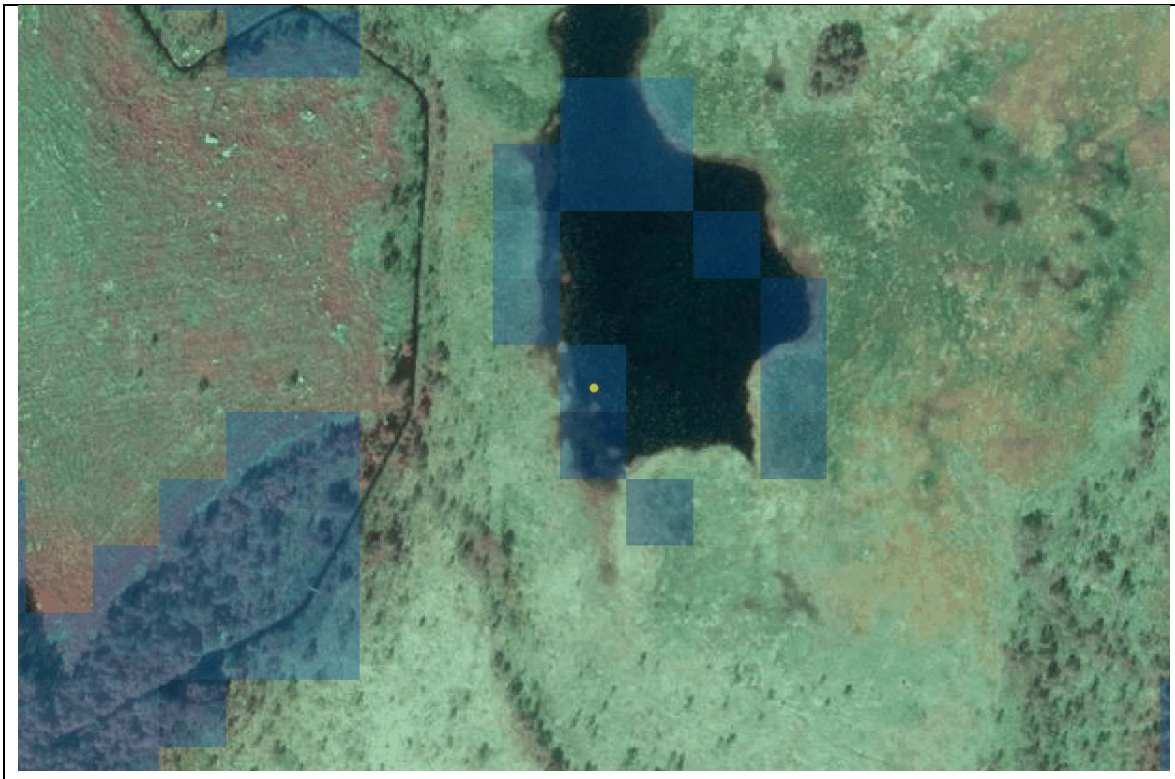

Fig. 8 Commission errors on peatland areas with water surfaces. Especially open peatland areas in Northern Finland experienced commission errors (6966667, 312030).

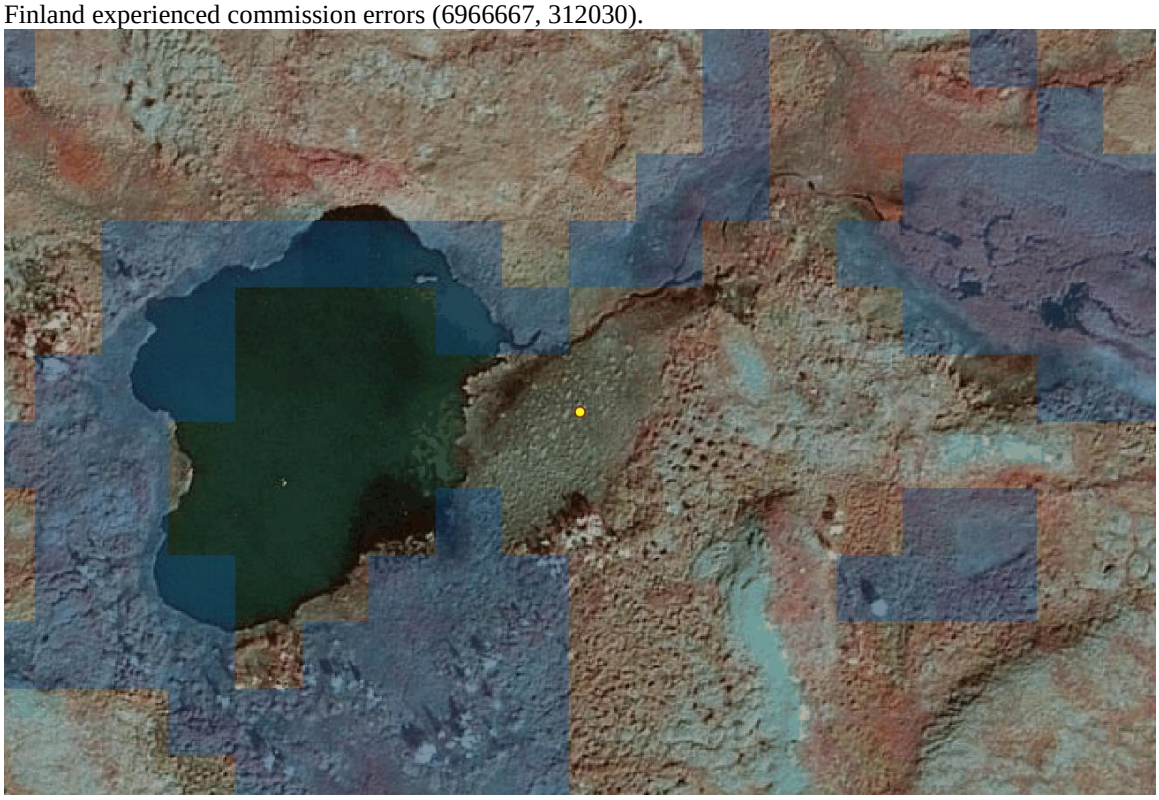

Fig. 9 Commission error on peatland areas (7772263,524703).

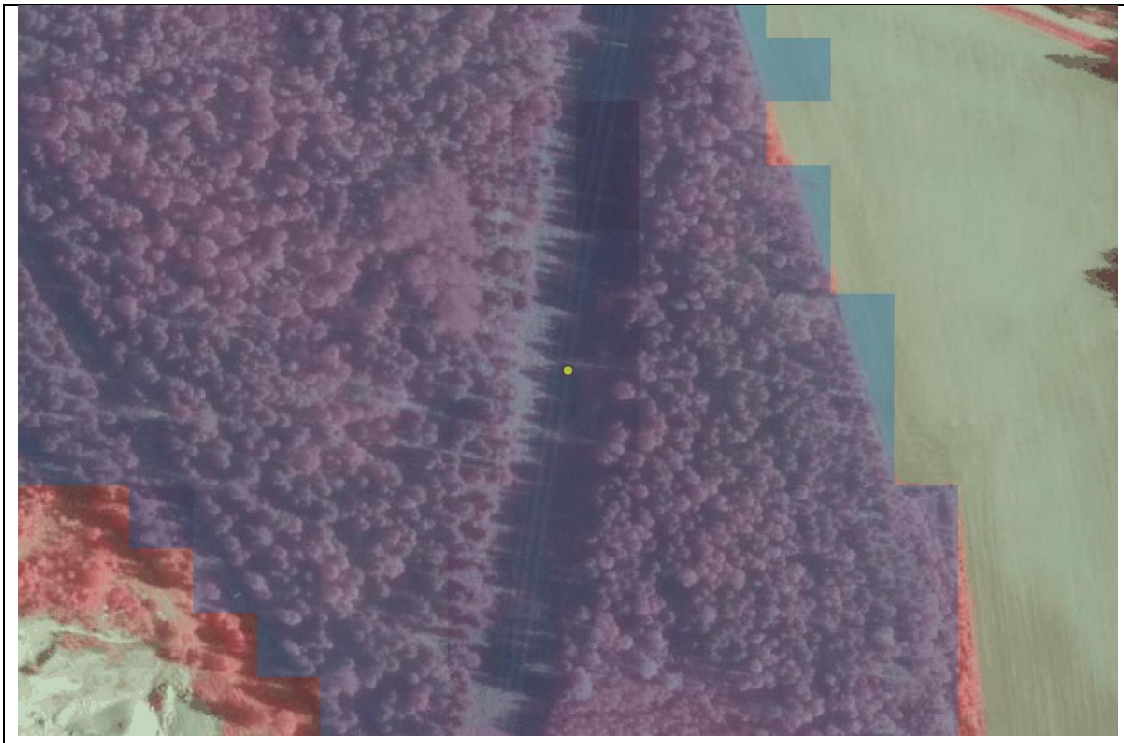

Fig. 10 Commission error on power lines in forest area. Here no visible tree crowns underneath the power line. In surroundings like this, ground under the power line grows seedlings or young forest due to natural regeneration, and is maintained by clearing as the trees grow too high. Thus, correct classification is challenging (6747256, 408378).

VI Documentation of software used for verification

Detailed information on the software type and exact version of software used for the validation
R version 3.4.4 + RStudio Version 1.1.442. (Base R + packages:dplyr,ggplot2,gridExtra,psych) SAS 9.4. QGis 3.0, ArcMap 10.3.1.