

Land cover classification in Poland and Norway using Sentinel-2 data – constraints and challenges

Adam Waśniewski, Agata Hościło
Centre of Applied Geomatics, IGiK
Linda Aune-Lundberg
NIBIO

Łódź 04.04.2024

Comparison of land cover classification approach

Study area and land cover classes

Objective:

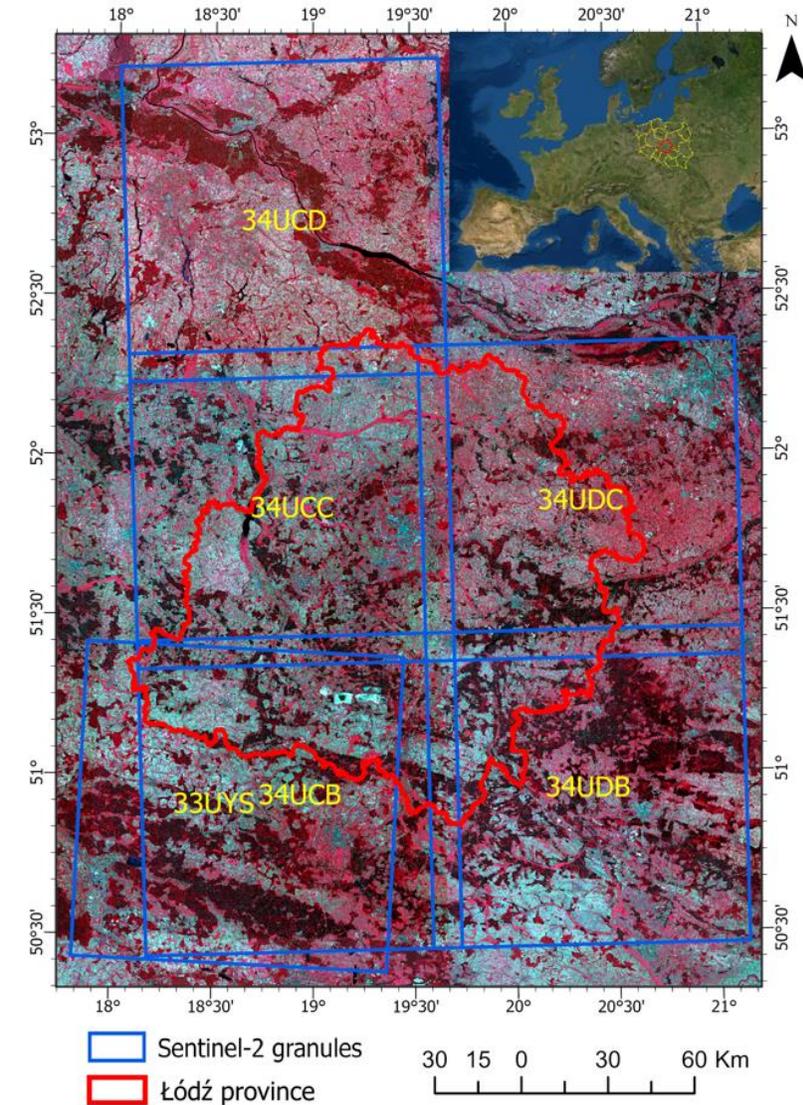
To develop methodology for land cover classification using the advance non-parametric machine learning algorithms, based on a time series of Sentinel-2 imagery.

Study area:

Poland – Łódź province

Land cover classes:

- Sealed surfaces
- Woodland broadleaved
- Woodland coniferous
- Shrubs
- Permanent herbaceous
- Periodically herbaceous
- Mosses
- Non vegetated (bare soil)
- Water



Data:

- Time series of Sentinel-2, from April to September for 2018 and 2020.

Resampling 20m spatial resolution bands to resolution of 10 m, 60 m bands were omitted in the classification.

Stacking all selected Sentinel-2 images, - layer stacking.

Reference data

- Points were created randomly on reference data:
 - BDOT10k - topographic database (scale 1:10 000),
 - Land cover classes from Agency for Restructuring and Modernisation of Agriculture (ARMA),
 - BDL – Forest Data Bank.
- Analysis of spectral histograms
- Reference samples: 60% training and 40% testing samples



Random Forest (RF) classifications was carried out in Python environment.

- RF parameters:
 - n_estimators = 100
 - criterion = 'entropy'
 - random_state = 42

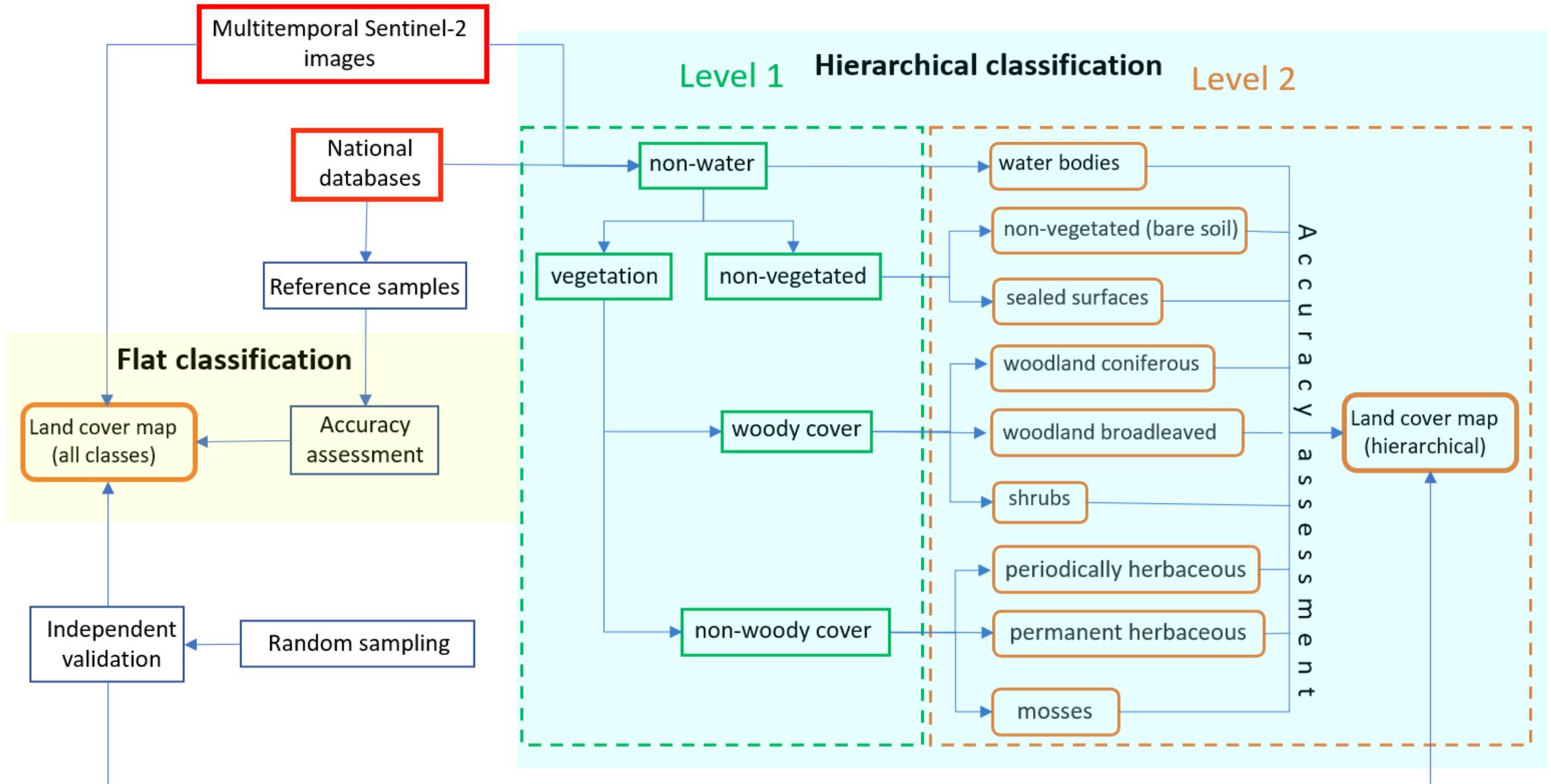


Classification was iteratively repeated 100 times.

Filtering the final map using Region Group and Nibble tools in ArcGIS Pro.



Classification scheme



Classification accuracy for all used S2 granules

HIERARCHICAL CLASSIFICATION		Overall accuracy	Kappa coefficient	F1 score
level 1	non-water/water bodies	0.99 - 1.00	0.93 - 0.99	0.96 - 1.00
	vegetation/non-vegetation	0.97 - 0.98	0.70 - 0.79	0.85 - 0.90
	woody cover/non-woody cover	0.95 - 0.99	0.86 - 0.97	0.92 - 0.99
level 2	sealed surfaces ,non-vegetated (bare soil)	0.92 - 0.97	0.56 - 0.85	0.78 - 0.92
	woodland broadleaved, woodland coniferous, shrubs	0.94 - 0.99	0.86 - 0.97	0.88 - 0.99
	permanent herbaceous, periodically herbaceous, mosses	0.93 - 0.99	0.68 - 0.79	0.77 - 0.87
FLAT classification		0.89 - 0.93	0.82 - 0.89	0.74 - 0.81

Flat classification (FLAT)

Land Cover Classes	UA	PA	F1 score
Sealed surfaces	0.63 - 0.82	0.79 - 0.82	0.72 - 0.83
Woodland broadleaved	0.77 - 0.89	0.76 - 0.83	0.77 - 0.84
Woodland coniferous	0.94 - 0.99	0.92 - 0.98	0.94 - 0.98
Shrubs	0.15 - 0.74	0.38 - 0.77	0.25 - 0.76
Permanent herbaceous	0.65 - 0.80	0.73 - 0.81	0.69 - 0.80
Periodically herbaceous	0.94 - 0.96	0.90 - 0.94	0.92 - 0.95
Mosses	0.32 - 0.67	0.55 - 0.79	0.40 - 0.73
Non-vegetated (bare soil)	0.18 - 0.76	0.50 - 0.89	0.26 - 0.80
Water bodies	0.90 - 0.99	0.92 - 0.99	0.92 - 0.99

Average accuracy

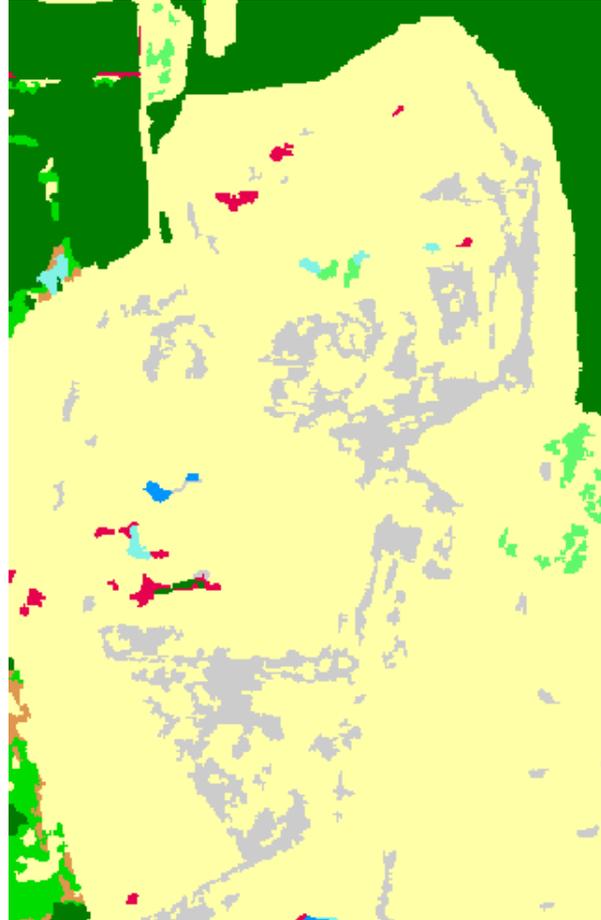
	FLAT		HIERARCHICAL	
	UA	PA	UA	PA
Sealed surfaces	0.79	0.74	0.77	0.92
Woodland broadleaved	0.80	0.74	0.71	0.88
Woodland coniferous	0.76	0.88	0.91	1.00
Shrubs	0.65	0.70	0.71	0.78
Permanent herbaceous	0.66	0.92	0.84	0.84
Periodically herbaceous	0.78	0.80	0.81	0.70
Mosses	0.82	0.64	1.00	0.74
Non-vegetated (bare soil)	0.97	0.66	0.92	0.68
Water bodies	0.96	0.98	0.96	0.98

Hierarchical vs flat land cover classification

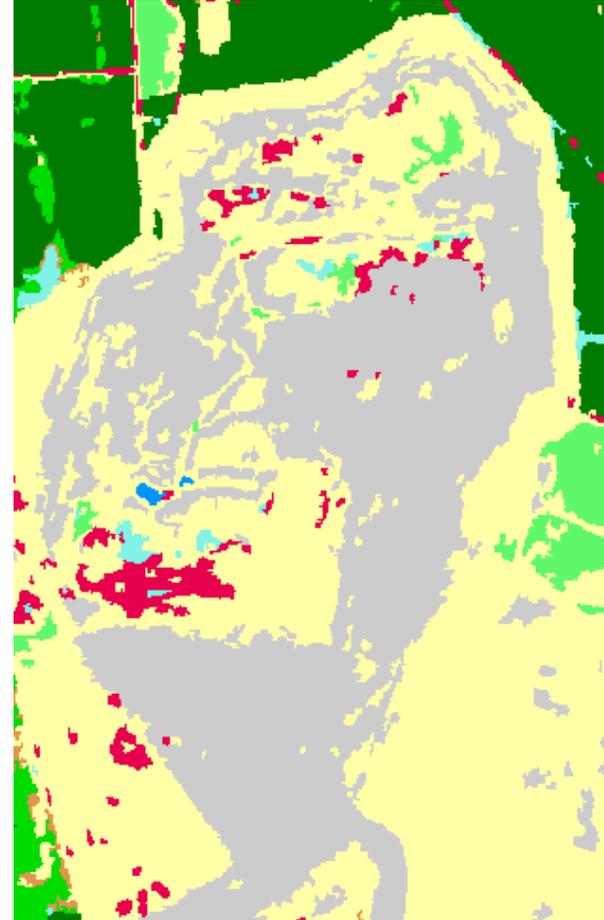
RGB



FLAT



HCL



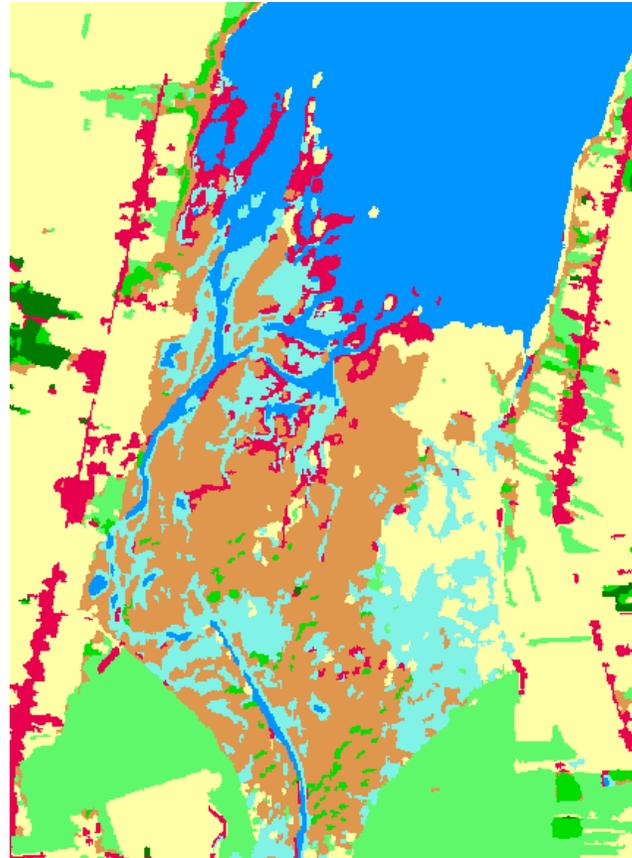
- sealed surfaces
- woodland broadleaved
- woodland coniferous
- shrubs
- permanent herbaceous
- periodically herbaceous
- wetlands and mosses
- non-vegetated
- water bodies

Hierarchical vs flat land cover classification

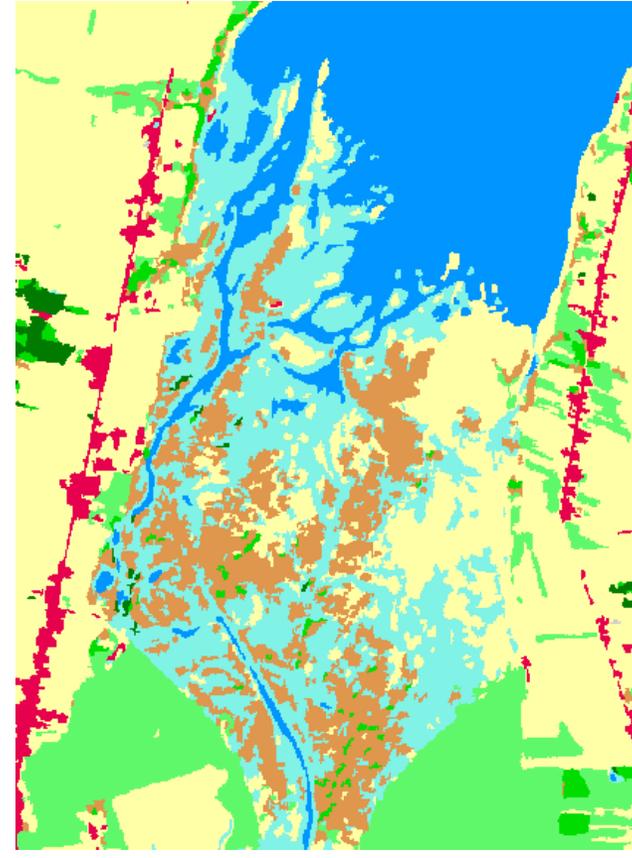
RGB



FLAT



HCL



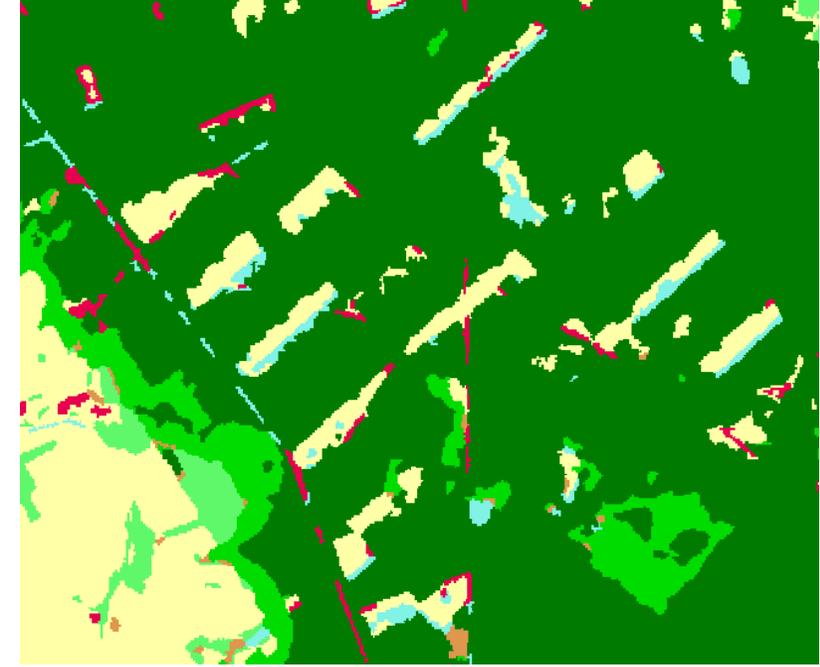
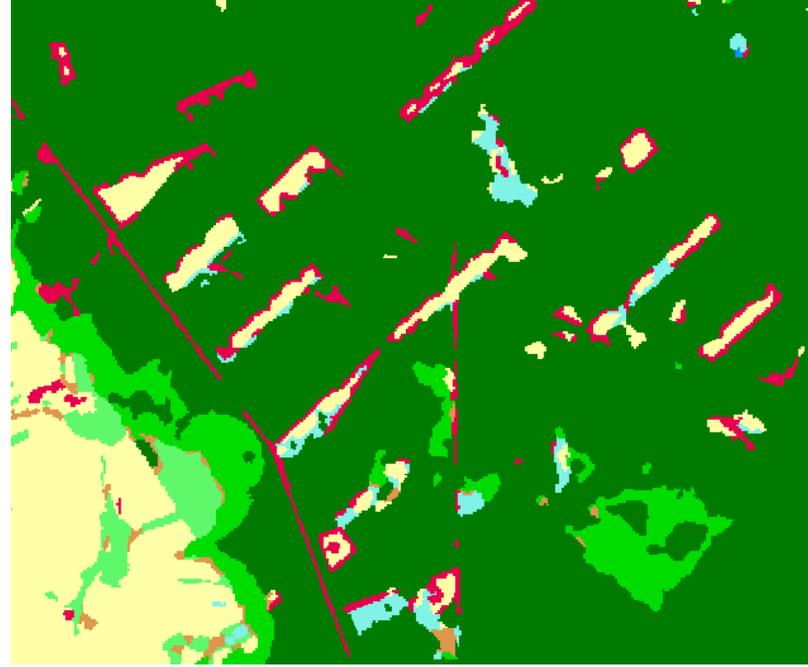
-  sealed surfaces
-  woodland broadleaved
-  woodland coniferous
-  shrubs
-  permanent herbaceous
-  periodically herbaceous
-  wetlands and mosses
-  non-vegetated
-  water bodies

Hierarchical vs flat land cover classification

RGB

FLAT

HCL



- sealed surfaces
- woodland broadleaved
- woodland coniferous
- shrubs
- permanent herbaceous
- periodically herbaceous
- wetlands and mosses
- non-vegetated
- water bodies

Hierarchical classification gives better visual results and higher accuracy than flat classification for the most problematic classes:

- the largest increase of UA were noticed for shrubs and non-vegetated classes, around 40 percentage points,
- UA and PA for wetlands and mosses increase from 61% and 80% to 73% and 87% respectively.

In hierarchical classification, over the densely built-up areas the building's shadows are misclassified as wetlands and mosses.

Hierarchical classification gives additional results: water extent, forest cover or impervious areas from level one.

In flat classification, the sealed surfaces were generally underestimated, the overestimated were visible in forest shadows and along the rivers.

Hierarchical classification is more time consuming and more laborious.

Obtaining satisfying results requires repeating the processes and experimenting with parameters.



Article

Can a Hierarchical Classification of Sentinel-2 Data Improve Land Cover Mapping?

Adam Waśniewski ^{1,2,*}, Agata Hościło ¹ and Milena Chmielewska ¹



40th EARSeL Symposium 2021
European Remote Sensing-New Solutions for Science and Practice

2nd EARSeL UAS Workshop
9th EARSeL Workshop on Remote Sensing of Coastal Zone
EO Education Workshop

Can a Hierarchical Classification of Sentinel-2 Data Improve the Land Cover Mapping?

esa THE EUROPEAN SPACE AGENCY

07 – 10 June 2021
© European Association of Remote Sensing Laboratories, University of Warsaw

FACULTY OF GEOGRAPHY AND REGIONAL STUDIES
UNIVERSITY OF WARSAW

Photo: Alleja Folbrier



living planet symposium BONN 2022

Comparison of hierarchical and flat land cover classification of Sentinel-2 data

esa

Impact of reference samples on land cover classification accuracy.

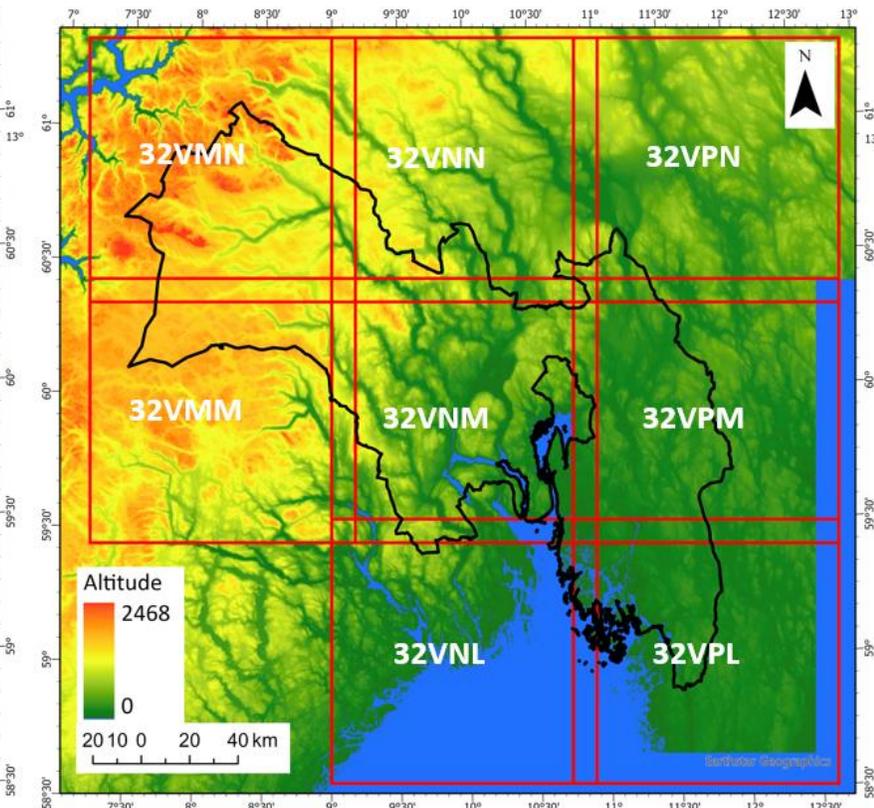
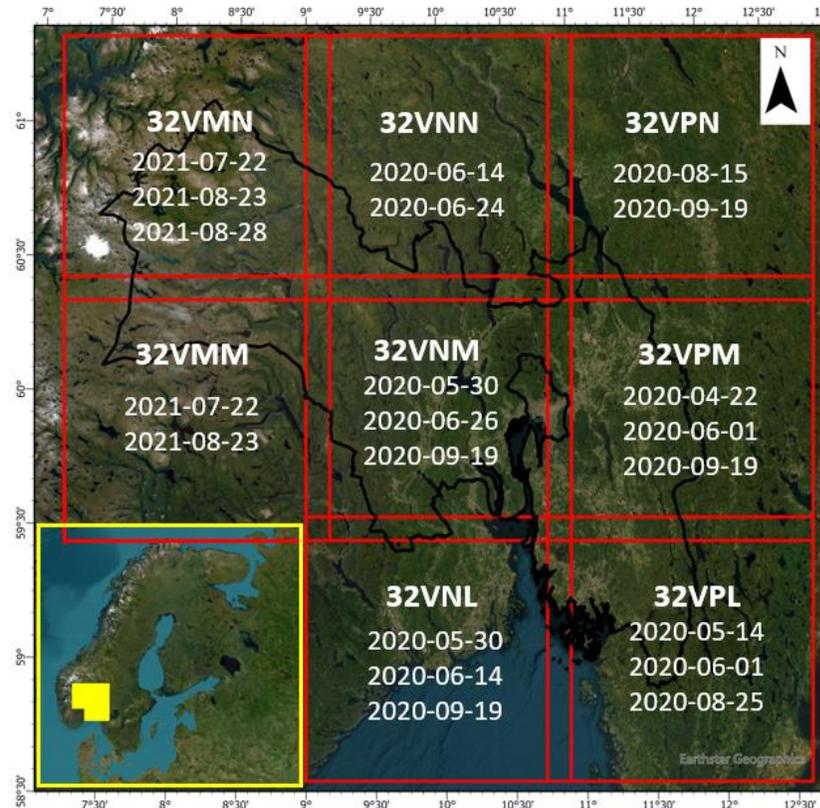
Study area and land cover classes

Objectives:

- to develop land cover map,
- to assess the impact of reference data on the classification result,
- to assess the impact of the Digital Elevation Model on the accuracy of classification.

The land cover classes:

- sealed surfaces,
- woodland coniferous,
- woodlands broadleaved,
- low vegetation,
- permanent herbaceous,
- periodically herbaceous,
- mosses,
- non- and sparse vegetation,
- water,
- snow and ice.



AR50 is the Norwegian medium-resolution land resource dataset, which covers the whole of mainland Norway, scale of 1:50 000. The primary classes are **built-up areas, agriculture, forest, peat bogs, open areas, glaciers, fresh water, oceans and not mapped area**. The minimum mapping unit for the primary classes in AR50 is **1.5 hectares**.

AR5 is a more detailed version of the AR50 database and includes land cover information below the tree line.

Reference points:

- points created randomly,
- an internal buffer of 10 m and 20 m for water was applied,
- proportions: 2 points per km²,
- minimum of 20 m distance between points,
- minimum of 200 points for one class on one granule,
- points divided into: 60% training and 40% verification.

ARTYPE/arealtype (land type)

Value	Land type
10	Built-up area
20	Agriculture
30	Forest
50	Open land
60	Peat bog
70	Glacier and permanent snow
81	Fresh water
82	Ocean
99	Not mapped

ARSKOGBON/skogbonitet (forest site class)

Value	Forest site class
18	High/particularly high productivity
13	Medium productivity
12	Low productivity
11	Non-productive
98	Not relevant
99	Not recorded

ARJORDBR/jordbruk (agricultural)

Value	Arable land
24	Fully and surface cultivated land
25	Pasture-land
98	Not relevant
99	Not recorded

ARVEGET/vegetasjonsdekke (open land)

Value	Open land
51	Not vegetated
52	Sparse vegetation
53	Lichen
54	Intermediate vegetation
55	Vigorous vegetation
98	Not relevant
99	Not recorded

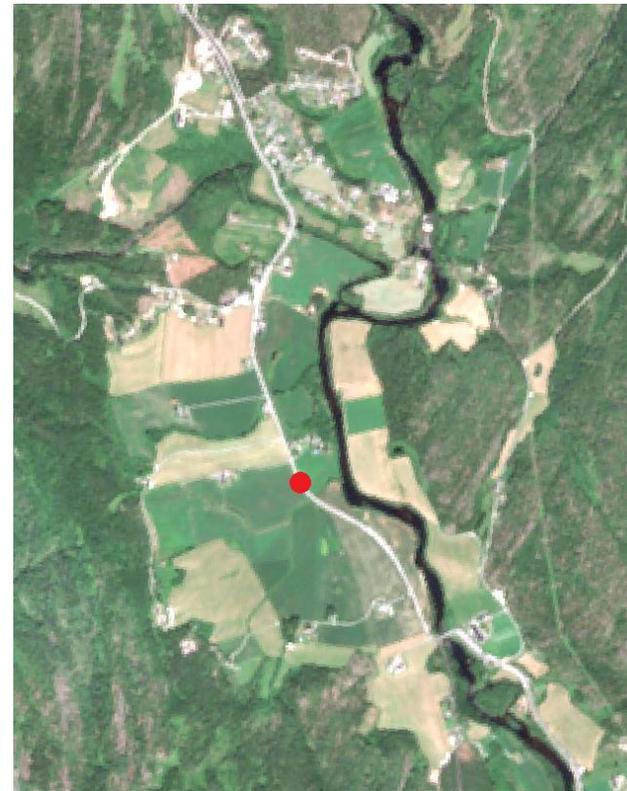
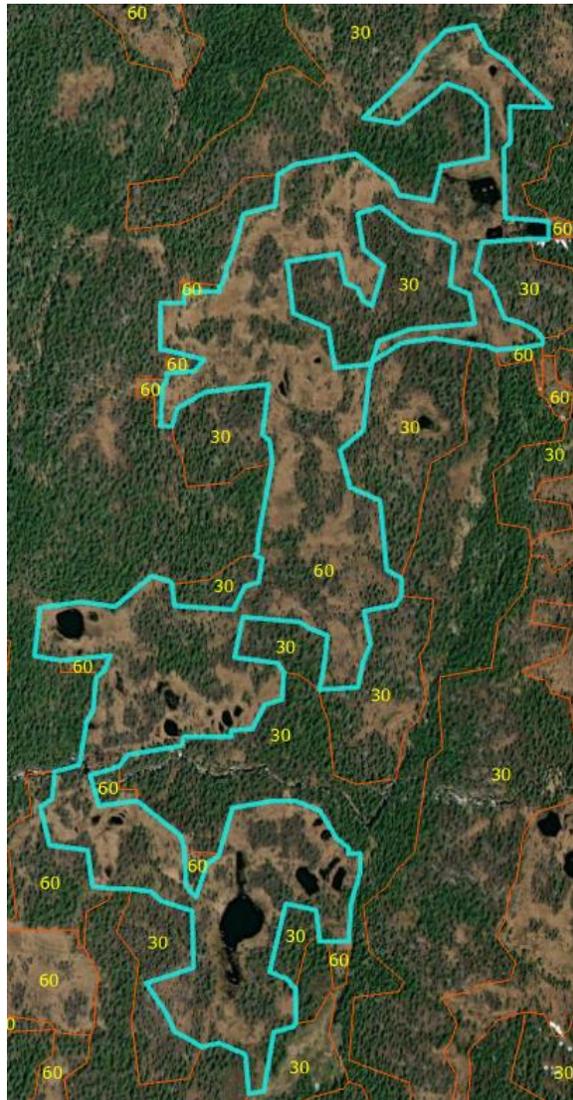
ARTRESLAG/treslag (tree type)

Value	Tree type
31	Coniferous forest
32	Deciduous forest
33	Mixed forest
39	Not forested
98	Not relevant
99	Not recorded

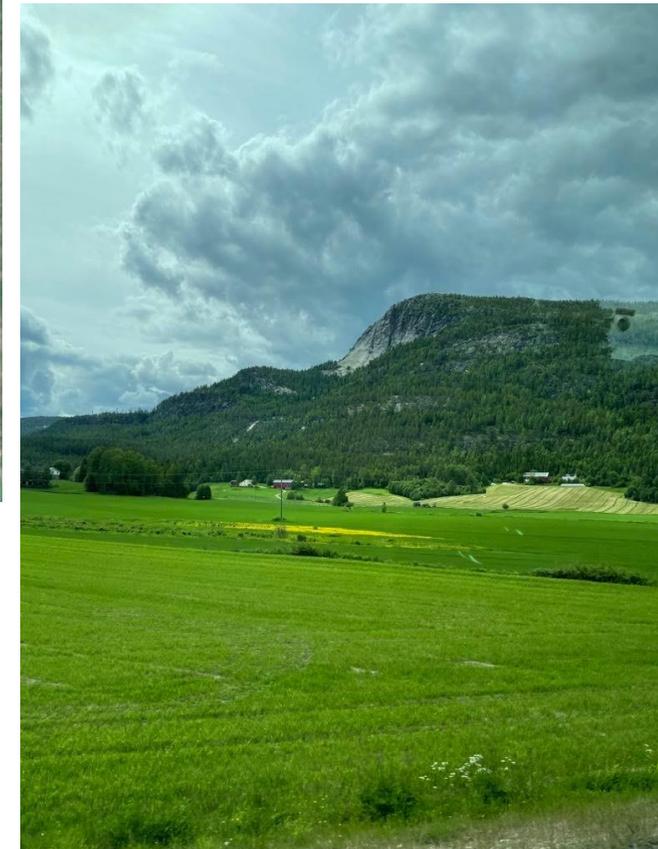
ARDYRKING/dyrkbarijord (arable land)

Value	Arable land
81	Non-cultivable soil
82	Cultivable soil
98	Not relevant
99	Not recorded

Mosses



Permanent herbaceous



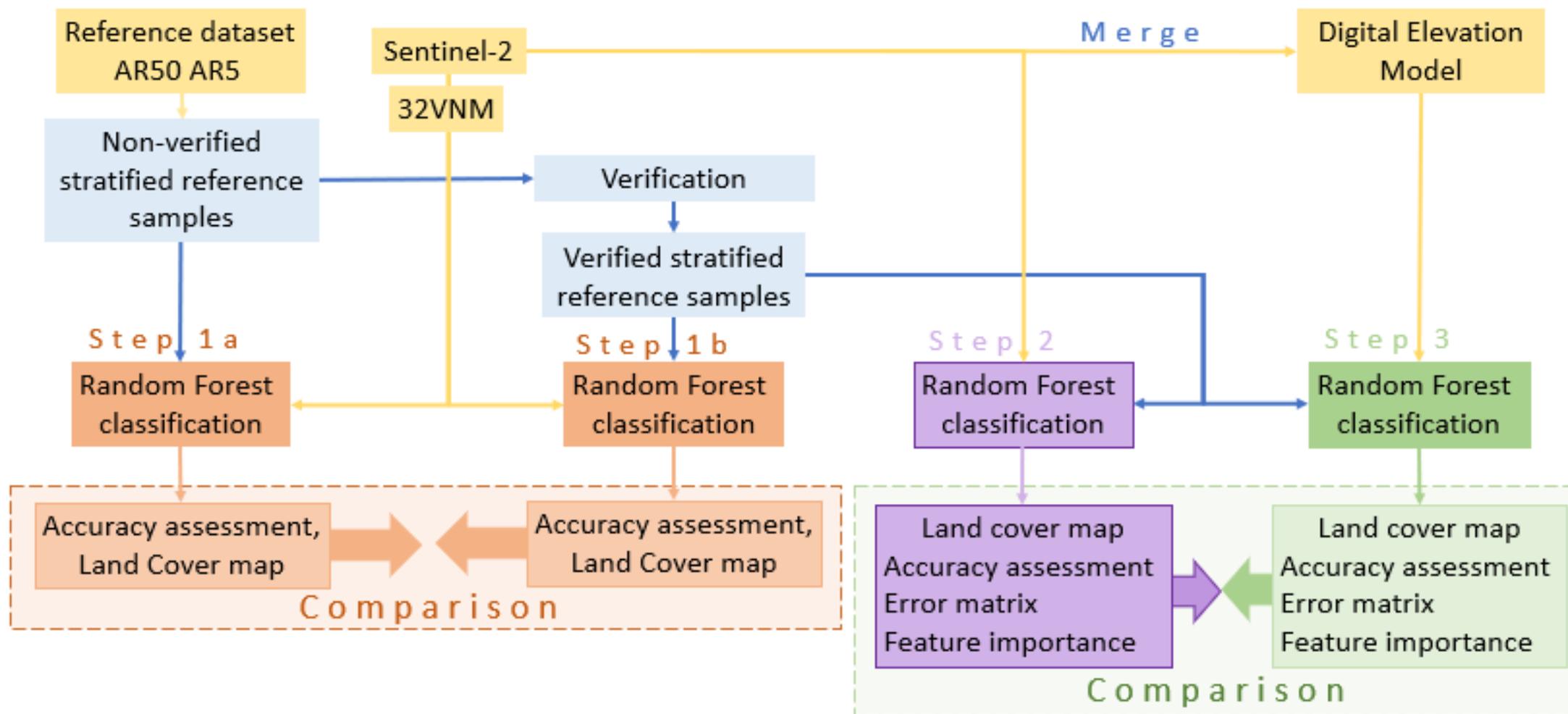
Land cover classification:

- Data – set of cloud free Sentinel-2A and -2B images captured between end of April and September 2018 and 2020.
- Clarification was carried out using the Random Forest machine learning algorithm.
- The analyses were carried out in the Python programming environment.
- The stability of the classification model was assessed using iterative accuracy assessment (iteration 100 times).
- Analyses were performed in cloud computing environment Amazon Web Services (AWS).
- After preliminary analyses, a Digital Elevation Model was included in the classification.



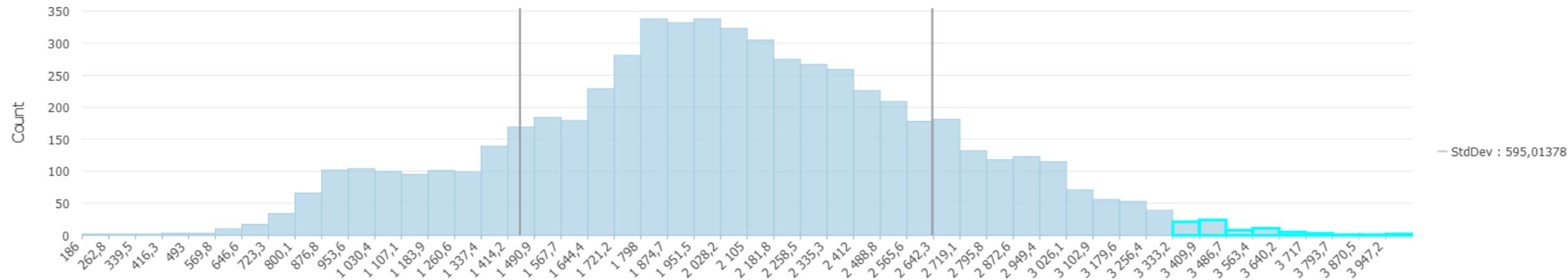
Digital Elevation Model:

- Shuttle Radar Topography Mission data (SRTM) is available from 56°S to 60°N.
- The freely available Digital Elevation Model was used for continental Norway with a spatial resolution of 10 meters, provided at geonorge.no.



Histogram analysis – locating and removing points

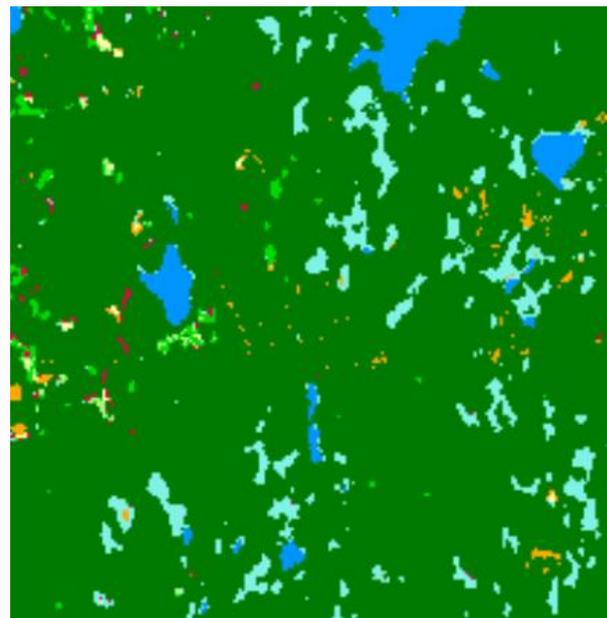
Distribution of s8



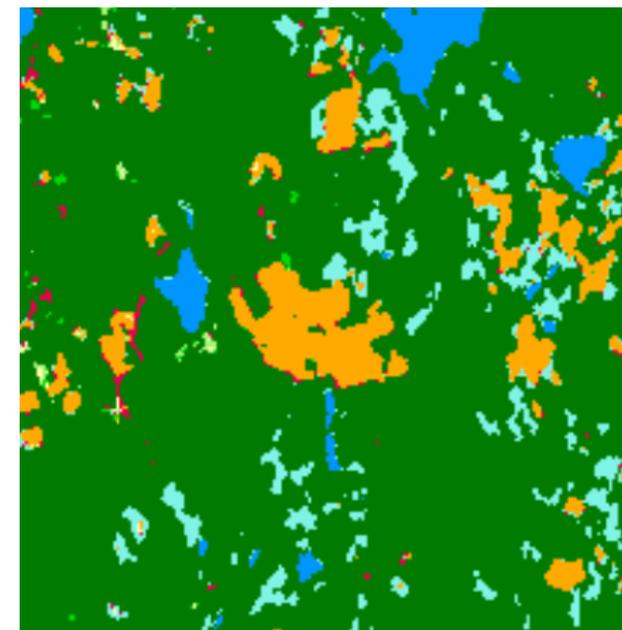
Locating points on clear-cuts



Classification result before removing points

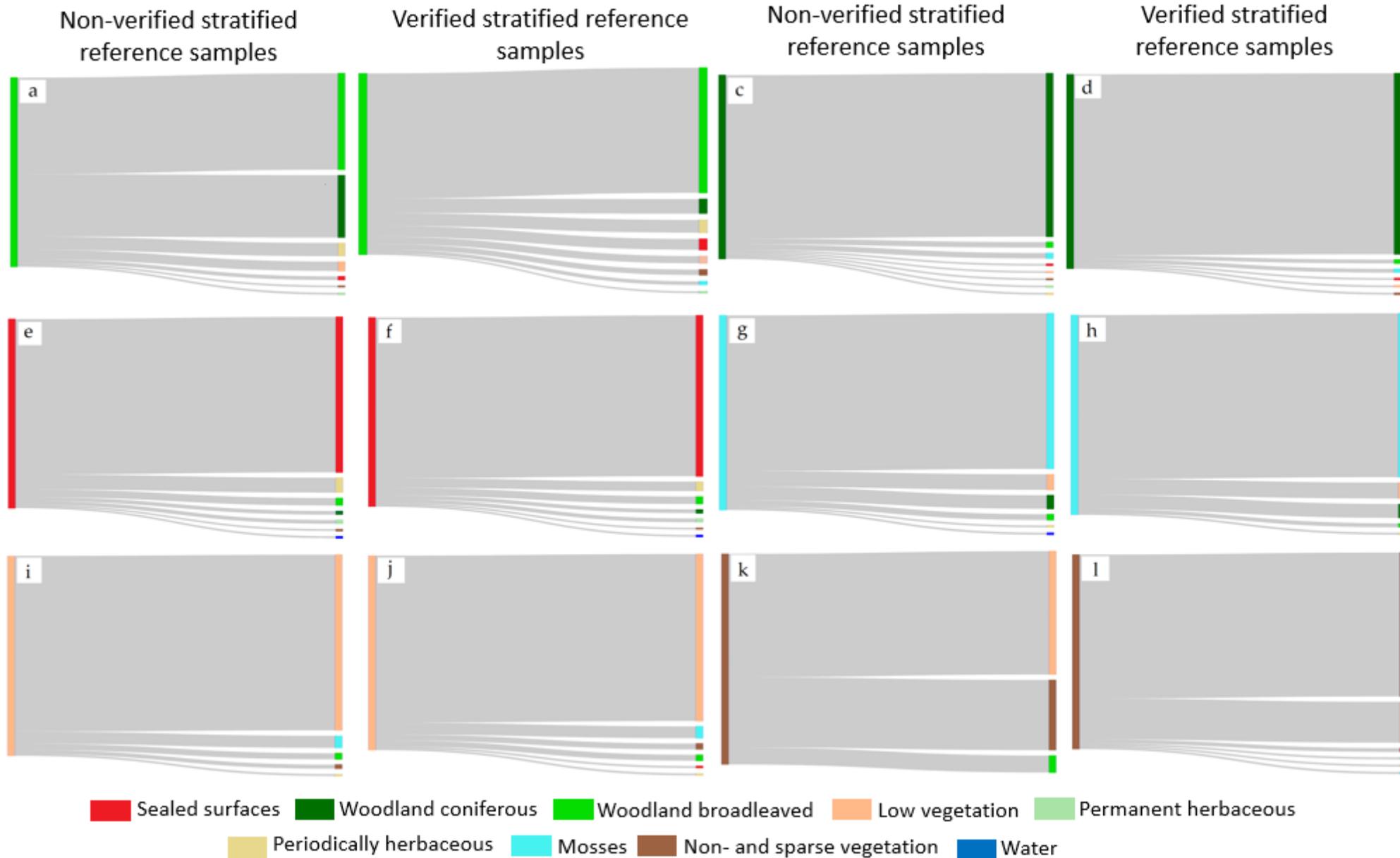


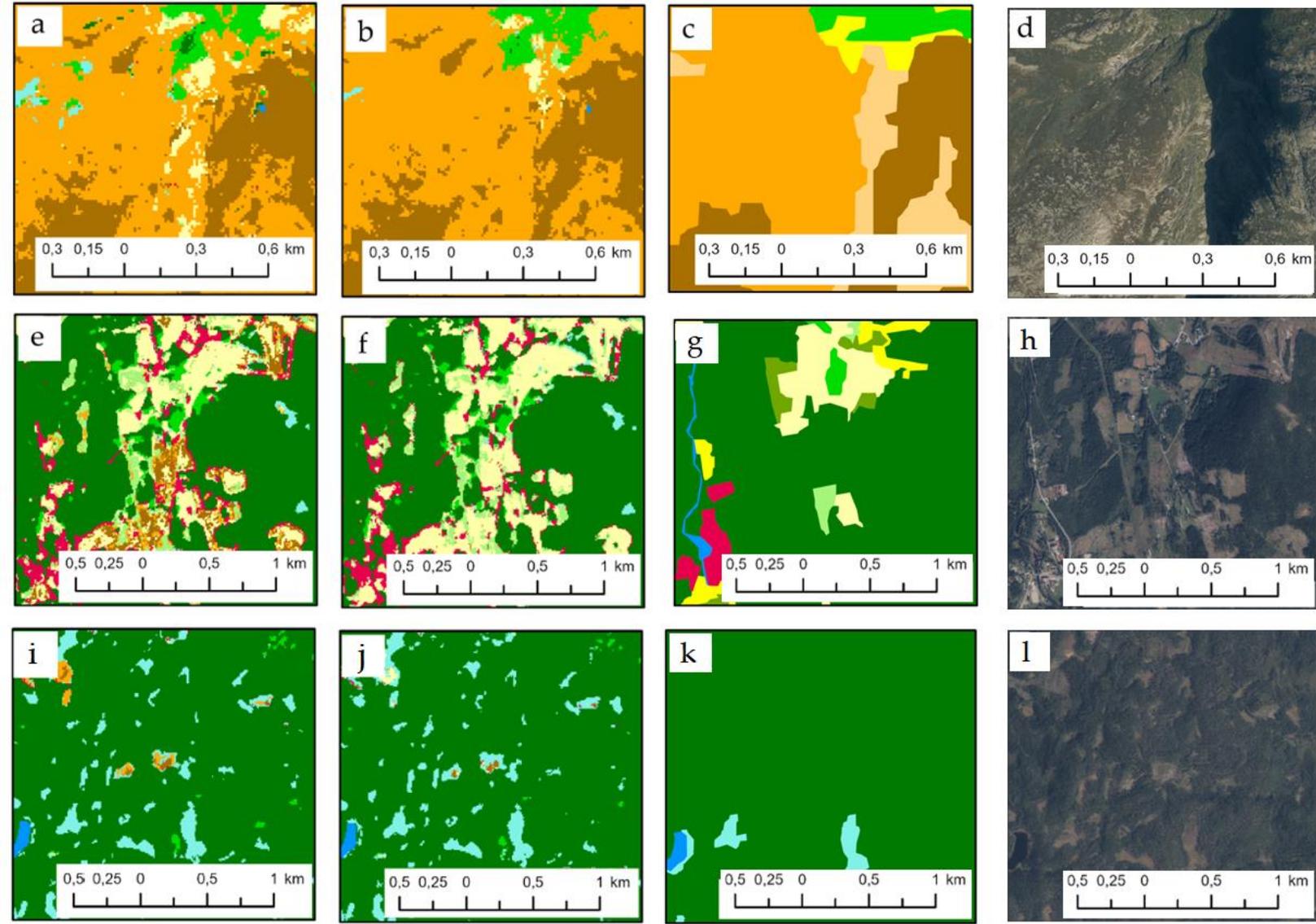
Classification results after removing points



- Viken county
- Sealed surfaces
- Woodland coniferous
- Woodland broadleaved
- Low vegetation
- Permanent herbaceous
- Periodically herbaceous
- Mosses
- Non- sparse vegetation
- Water
- Snow and ice

Impact of reference samples





a, e, i – Non-verified stratified reference samples (step 1a)

b, f, j – Verified stratified reference samples (step 1b)

c, g, k - AR50

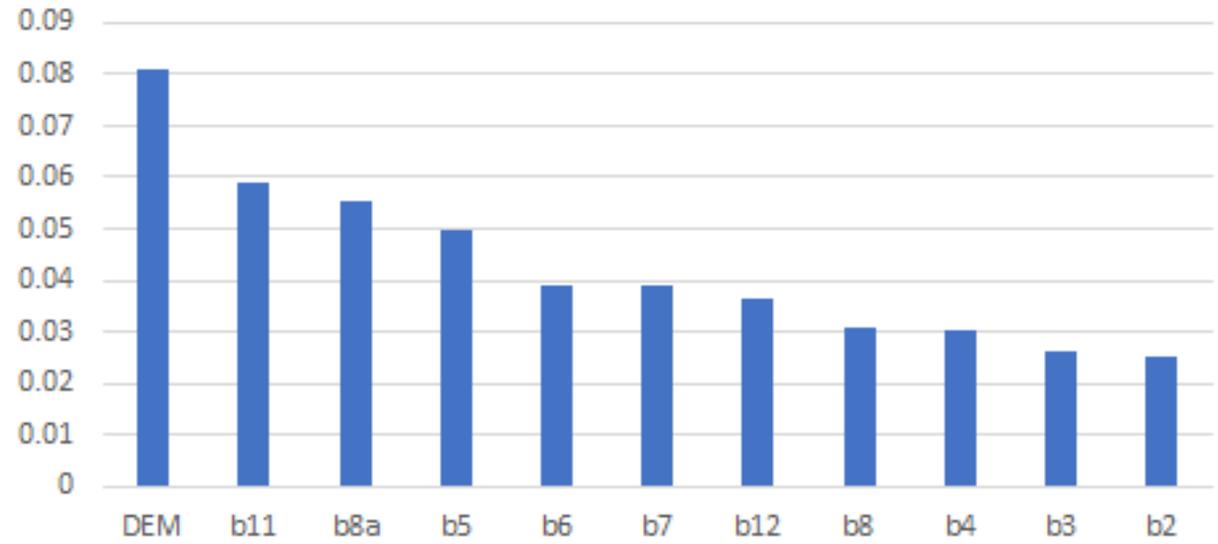
d, h, l - Ortophoto © Norge digitalt

AR50 legend

- | | | |
|---|---|---|
|  Built-up area |  Intermediate vegetation |  Other open land |
|  Coniferous forest |  Vigorous vegetation |  Peat bog |
|  Deciduous forest |  Pasture-land |  Lichen |
|  Mixed forest |  Fully and surface cultivated land |  Sparse vegetation |
| | |  Fresh water |

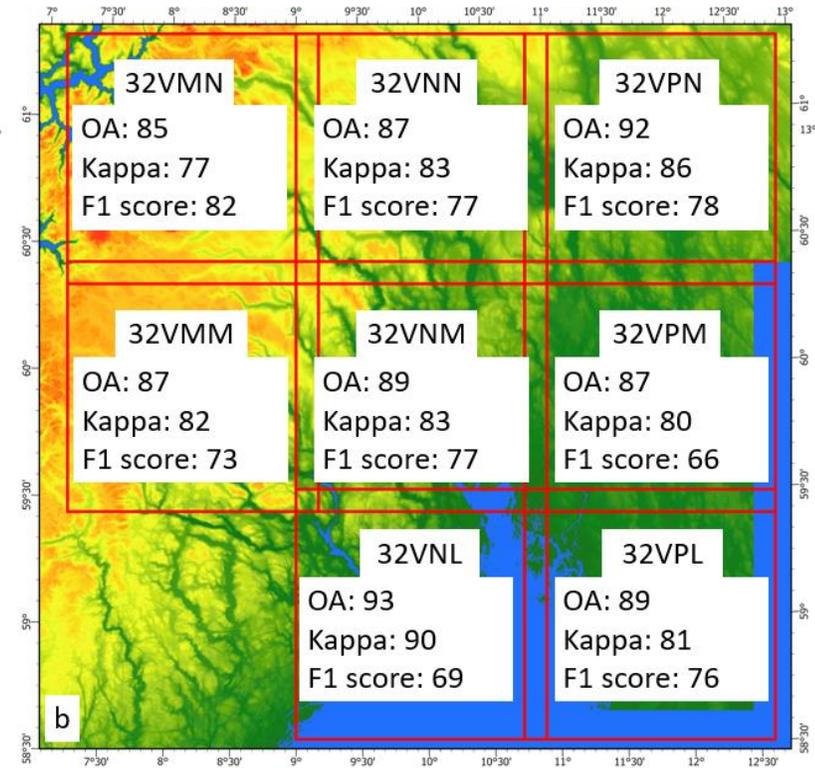
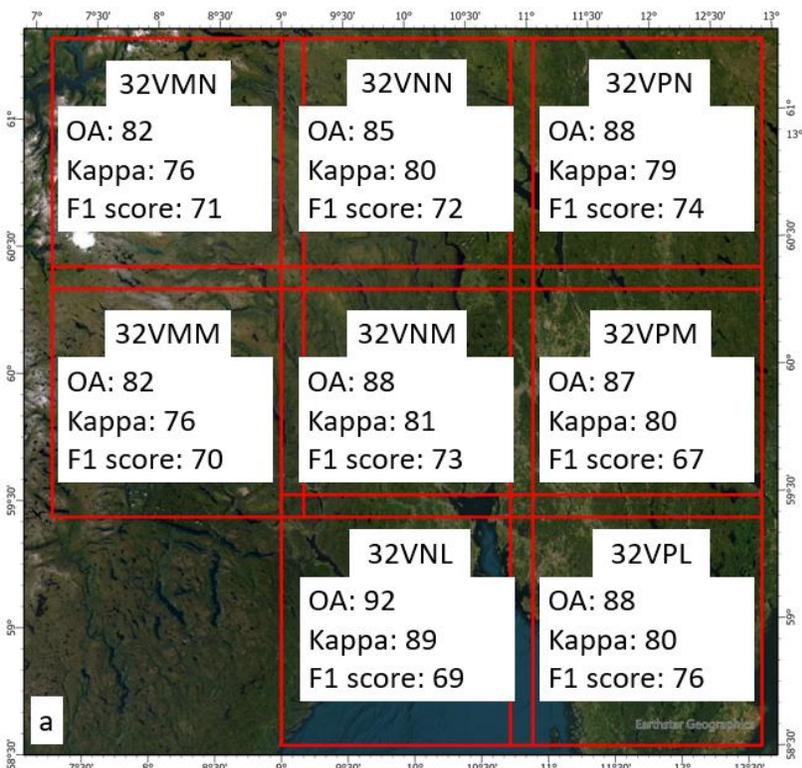
Classification accuracy

Variable importance



Sentinel-2

Sentinel-2 + DEM



- The smaller the scale of the reference data, the less accurately the reference points will be located.
- The rate of generalization of reference data affects the accuracy of classification.
- It is important to know the land characteristics and land cover class definitions in the reference databases.
- During Sentinel-2 data selection the phenology should be taken into consideration.
- Heterogeneous classes are more difficult to classify.
- Eliminating mislocated points increases the overall accuracy of the classification.
- DEM increases the accuracy of the classification and is the most informative variable used in classification.
- By adding DEM the accuracy for classes located at higher altitudes increased from 4 to 8 percentage points.

I Kongres Geoinformacyjny

/ X Ogólnopolskie Sympozjum Geoinformacyjne /

"Współczesne technologie geoinformacyjne w modelowaniu przestrzeni"

Kraków, 25-27.X.2023



online
Ogólnopolska Konferencja
Fotointerpretacji
Teledetekcji
28 - 30 listopada 2022

Organizatorzy: PAN, CBK, Wydział Geodezji i Kartografii POLITECHNIKA WARSZAWSKA

Patronat: PLSA, Minister Edukacji i Nauki, Ministerstwo Rozwoju i Technologii

Remote Sensing Applications: Society and Environment 32 (2023) 101035



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase



EARSeL Symposium 2023
3-6 July 2023, Bucharest

The impact of selection of reference samples and DEM on the accuracy of land cover classification based on Sentinel-2 data

Adam Waśniewski ^{a, b, *}, Agata Hościło ^a, Linda Aune-Lundberg ^c

^a Institute of Geodesy and Cartography, Centre of Applied Geomatics, Poland

^b Department of Geoinformatics, Cartography and Remote Sensing, Faculty of Geography and Regional Studies, University of Warsaw, Poland

^c Norwegian Institute of Bioeconomy Research, Division of Survey and Statistics, Norway

Thank you for your attention!

adam.wasniewski@igik.edu.pl

