IIASA International Institute for Applied Systems Analysis www.iiasa.ac.at

Lessons learned from building a training data set for land cover mapping at 10m

esa

WorldCover

Myroslava Lesiv, Daniele Zanaga, Ruben Van De Kerchove, Nandika Tsendbazar, Martin Herold and Steffen Fritz

vito remote sensing

WAGENINGEN UNIVERSITY & RESEARCH

living planet symposium | BONN 23–27 May 2022

TAKING THE PULSE OF OUR PLANET FROM SPACE

EUMETSAT ECMWF EUSPA

Background

- High quality training data is a critical input for land cover/ land use mapping
- New requirements:
  - Very high resolution mapping
  - More thematic details
  - Change detections
- Different sources of training data available:
  - on-ground observations and
  - visually interpreted very high resolution images.
  - existing land cover/land use maps
  - automatic generation

Copernicus Global Land Service
Providing bio-geophysical products of global land surface
Copernicus
Europe's eyes on Earth

RELEASE OF THE 10 M WORLDCOVER MAP

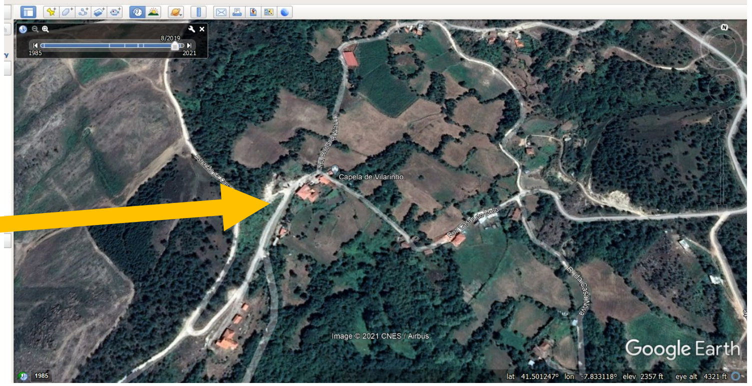
Esri Land Cover
Sentinel-2 10-Meter Land Use/Land Cover

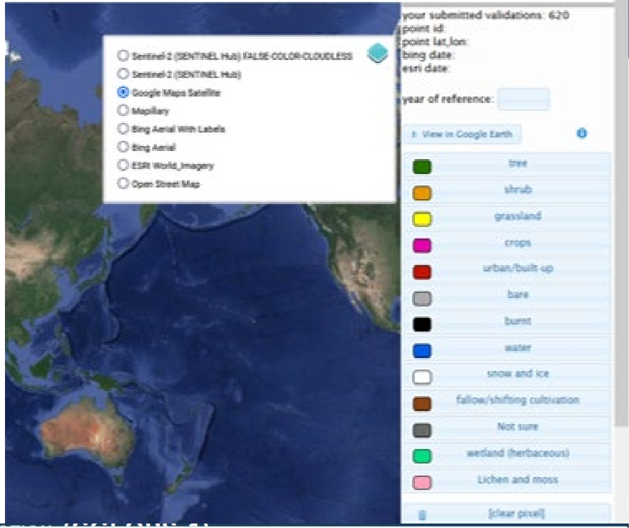
Challenges

- Unknown quality
  - Geolocation errors
  - Thematic errors
  - Timestamp
  - Not clear definitions
- Spatial distribution of data – overfitting issue
- Translation from one legend to another
- Translation of point observations into pixels
- Lack of data
- Access to data

Geo-Wiki tool box

- Very high resolution (VHR) imagery from Google maps, Microsoft Bing, and ESRI
- Google Earth VHR historical images
- Planet time series of images
- Sentinel-2 time-series in False color
- Street level images from Google and Mappilary
- NDVI time series derived from Landsat, Proba-V and MODIS data


1985
8/2019
2021
Capela de Vilarinho
Image © 2021 CNES / Airbus
Google Earth
lat 41.501247° lon -7.833118° elev 2357 ft eye alt 4321 ft


○ Sentinel-2 (SENTINEL Hub) FALSE-COLOR-CLOUDLESS
○ Sentinel-2 (SENTINEL Hub)
● Google Maps Satellite
○ Mapillary
○ Bing Aerial With Labels
○ Bing Aerial
○ ESRI World_Imagery
○ Open Street Map

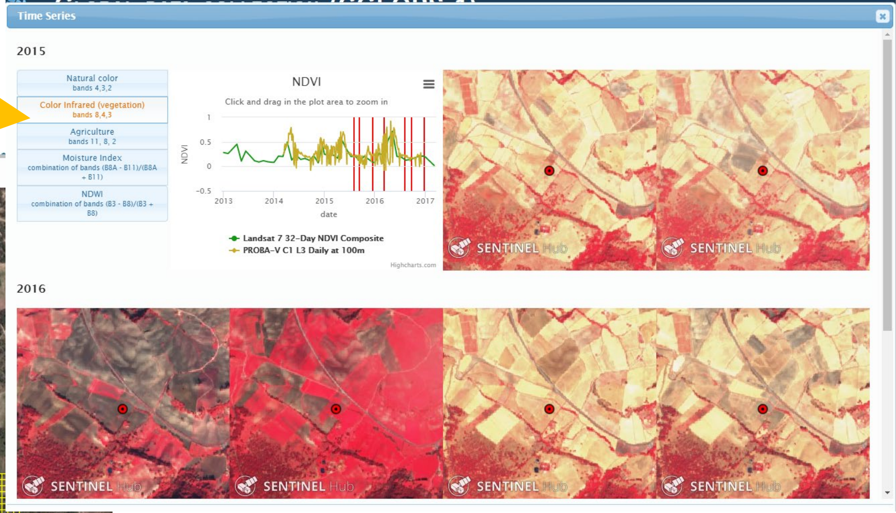
your submitted validations: 620
point id:
point lat_lon:
bing date:
esri date:

year of reference:

View in Google Earth

🟩	tree
🟧	shrub
🟨	grassland
🟪	crops
🟥	urban/built-up
⬜	bare
⬛	burnt
🟦	water
⬜	snow and ice
🟫	fallow/shifting cultivation
⬛	Not sure
🟩	wetland (herbaceous)
🟪	Lichen and moss

[clear pixel]


Time Series

2015

Natural color
bands 4,3,2

Color Infrared (vegetation)
bands 8,4,3

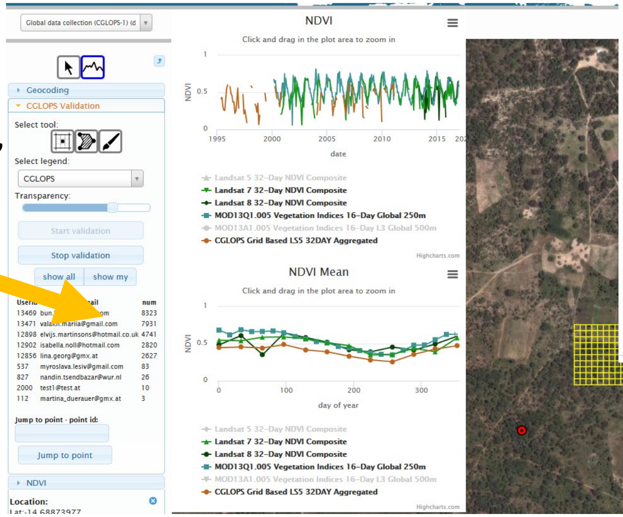
Agriculture
bands 11, 8, 2

Moisture Index
combination of bands (B8A - B11)/(B8A + B11)

NDWI
combination of bands (B3 - B8)/(B3 + B8)

NDVI
Click and drag in the plot area to zoom in
NDVI
1
0.5
0
-0.5
2013
2014
2015
2016
2017
date
Landsat 7 32-Day NDVI Composite
PROBA-V C1 L3 Daily at 100m
Highcharts.com
SENTINEL Hub

2016


Global data collection (CGLOPS-1) (d

Geocoding

CGLOPS Validation

Select tool:

Select legend: CGLOPS

Transparency:

Start validation

Stop validation

show all | show my

Useri	ail	num
13469	bun.[illegible]om	8323
13471	vala[illegible].mariia@gmail.com	7931
12898	elvijs.martinsons@hotmail.co.uk	4741
12902	isabella.noll@hotmail.com	2820
12856	lina.georg@gmx.at	2627
537	myroslava.lesiv@gmail.com	83
827	nandin.tsendbazar@wur.nl	26
2000	test1@test.at	10
112	martina_duerauer@gmx.at	3

Jump to point - point id:

Jump to point

NDVI

Location:
Lat:-14.68873977

NDVI
Click and drag in the plot area to zoom in
NDVI
1
0.5
0
1995
2000
2005
2010
2015
202
date
Landsat 5 32-Day NDVI Composite
Landsat 7 32-Day NDVI Composite
Landsat 8 32-Day NDVI Composite
MOD13Q1.005 Vegetation Indices 16-Day Global 250m
MOD13A1.005 Vegetation Indices 16-Day L3 Global 500m
CGLOPS Grid Based LS5 32DAY Aggregated
Highcharts.com
NDVI Mean
Click and drag in the plot area to zoom in
NDVI
1
0.5
0
0
100
200
300
day of year
Landsat 5 32-Day NDVI Composite
Landsat 7 32-Day NDVI Composite
Landsat 8 32-Day NDVI Composite
MOD13Q1.005 Vegetation Indices 16-Day Global 250m
MOD13A1.005 Vegetation Indices 16-Day L3 Global 500m
CGLOPS Grid Based LS5 32DAY Aggregated
Highcharts.com

Funded by the Horizon 2020 Framework Programme of the European Union

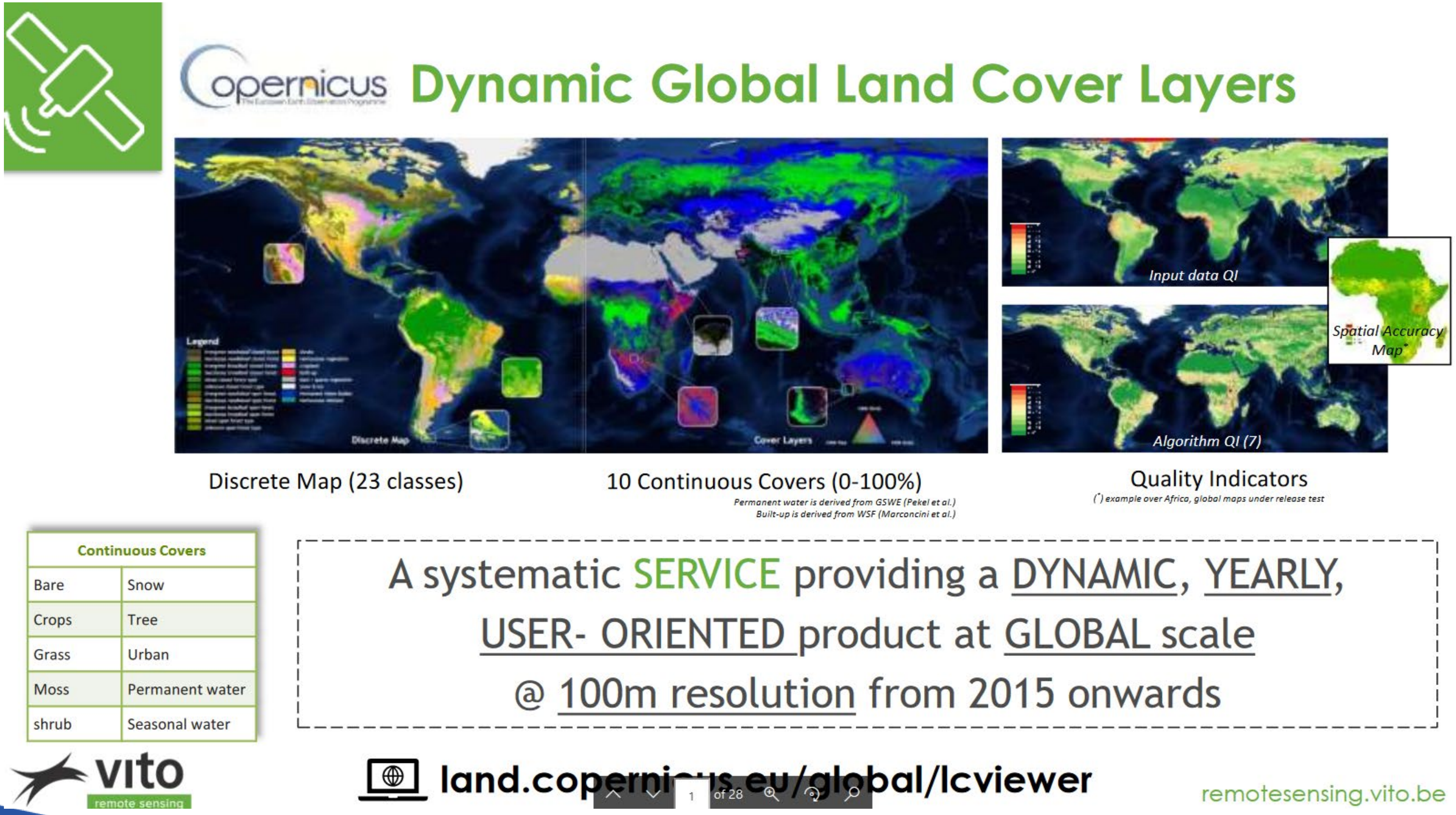
Concept of a multipurpose data set

Global Land Cover at 100m (JRC)

copernicus
Europe's eyes on Earth

World Cover at 10m (ESA)

Copernicus Global Land Cover 2015-2019 PROBA-V 100m



Copernicus Dynamic Global Land Cover Layers

Discrete Map (23 classes)

10 Continuous Covers (0-100%)

Permanent water is derived from GSWE (Pekel et al.)
Built-up is derived from WSF (Marconcini et al.)

Quality Indicators

(*) example over Africa, global maps under release test

Continuous Covers	
Bare	Snow
Crops	Tree
Grass	Urban
Moss	Permanent water
shrub	Seasonal water

A systematic SERVICE providing a DYNAMIC, YEARLY,
USER- ORIENTED product at GLOBAL scale
@ 100m resolution from 2015 onwards

vito remote sensing

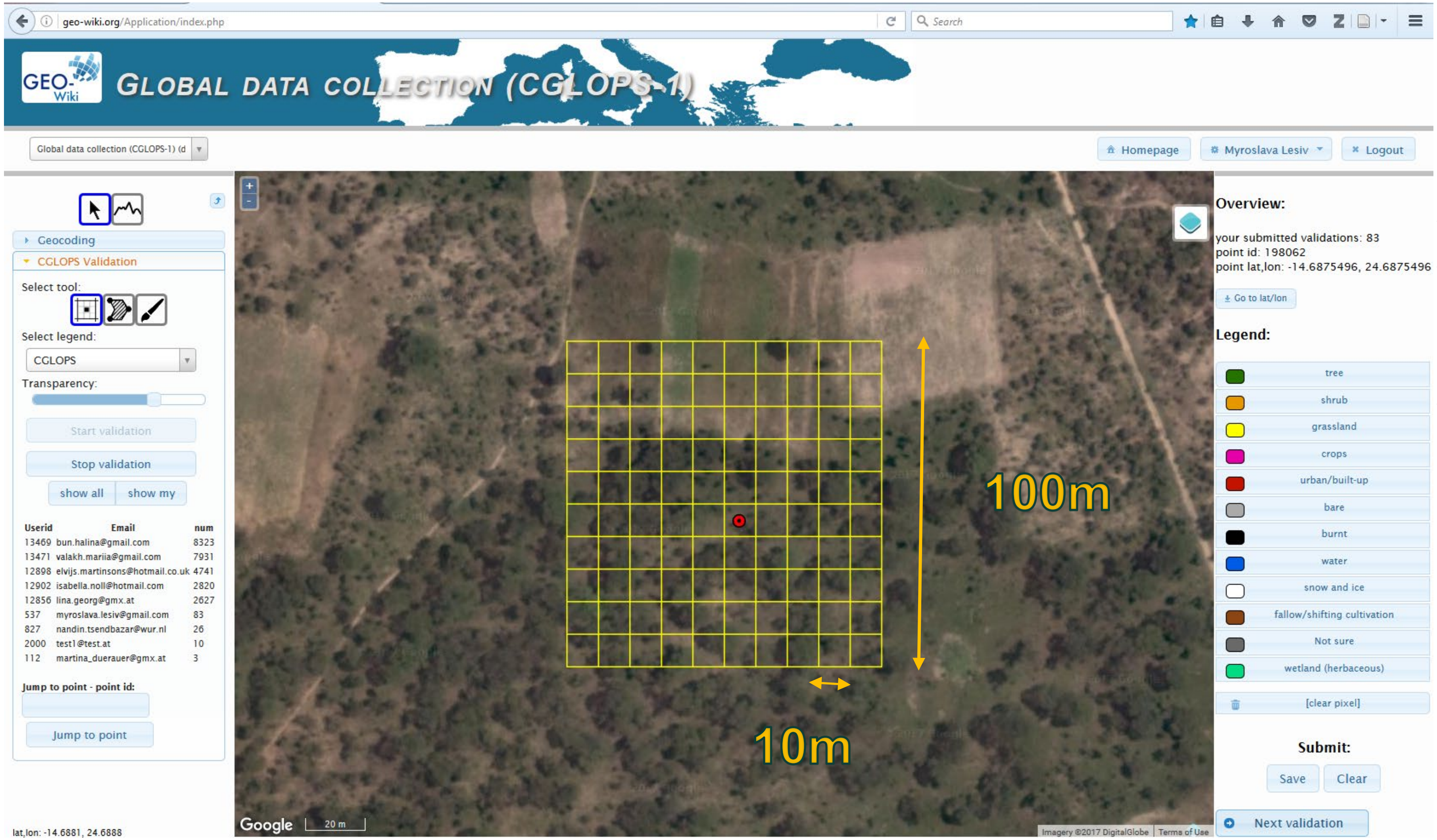
land.copernicus.eu/global/lcviewer

remotesensing.vito.be

25th May. 2022

Lesiv et al., 2022 LPS

Geo-wiki app



Userid	Email	num
13469	bun.halina@gmail.com	8323
13471	valakh.mariia@gmail.com	7931
12898	elvijs.martinsons@hotmail.co.uk	4741
12902	isabella.noll@hotmail.com	2820
12856	lina.georg@gmx.at	2627
537	myroslava.lesiv@gmail.com	83
827	nandin.tsendbazar@wur.nl	26
2000	test1@test.at	10
112	martina_duerauer@gmx.at	3

lat,lon: -14.6881, 24.6888

Overview:

your submitted validations: 83
point id: 198062
point lat,lon: -14.6875496, 24.6875496

Legend:

- 🟩 tree
- 🟧 shrub
- 🟨 grassland
- 🟪 crops
- 🟥 urban/built-up
- ⬜ bare
- ⬛ burnt
- 🟦 water
- ⬜ snow and ice
- 🟫 fallow/shifting cultivation
- ⬛ Not sure
- 🟩 wetland (herbaceous)

- Fractions at 100m
- Easy translation to discrete land cover classes
- Training data at 10m

Data collection workflow

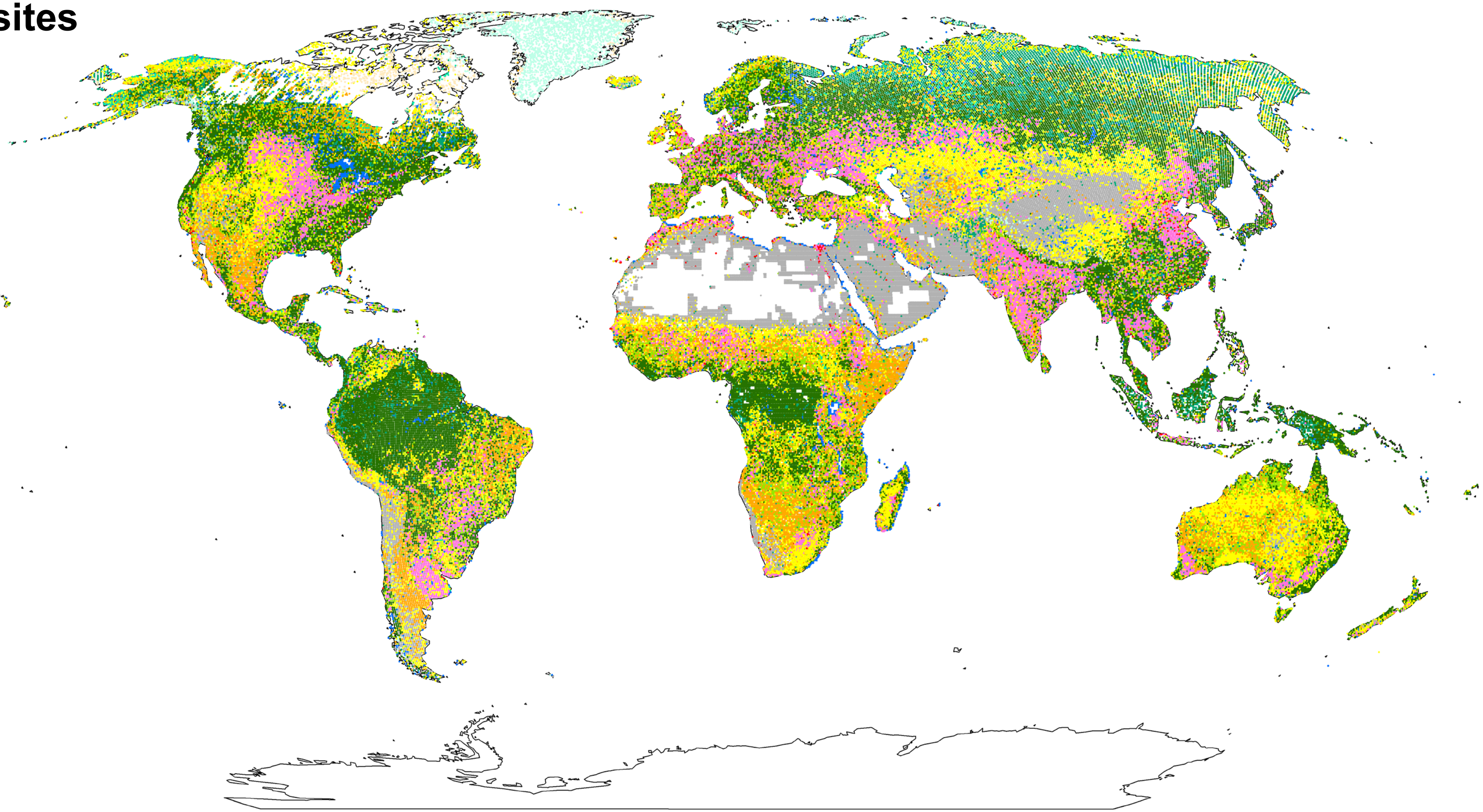
- Initial Geo-wiki training
  - Interface, tools, per class examples
- Weekly online seminars to check quality
  - discuss difficult locations
  - randomly revisit some classifications
    - Target <5 % of mistakes
- Comparison with regional products
  - Revise disagreeing locations
- Removing land cover class outliers based on spectral information
  - homogeneous pixels

Distribution of reference data 2015

~180 000 sample sites
by 20 experts

Sampling design:

- systematic
- uncertainty hotspots



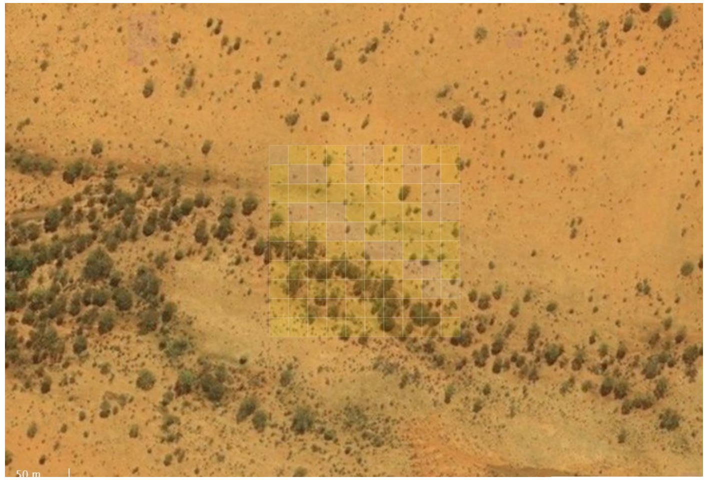
ESA World Cover 2020/2021 at 10m Sentinel 1 and Sentinel 2

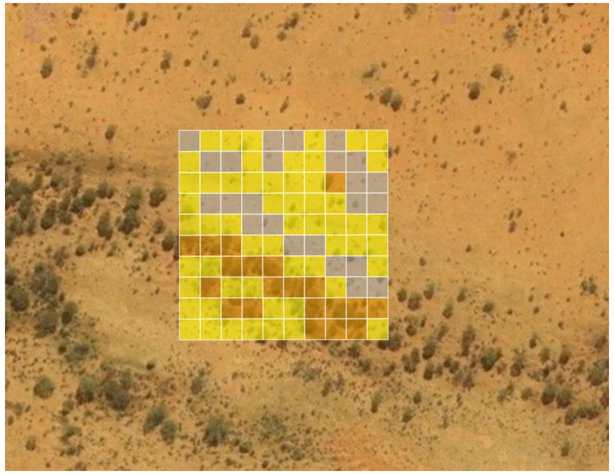
180 K (at 100m) pixels ~ 18 Millions (at 10m)

Issues:

- geolocation errors of the underlying images used for visual interpretations
- land cover/land use changes that happened after 2015
- Correct fractions at 100m ~misclassifications at 10m

Landscape in Australia





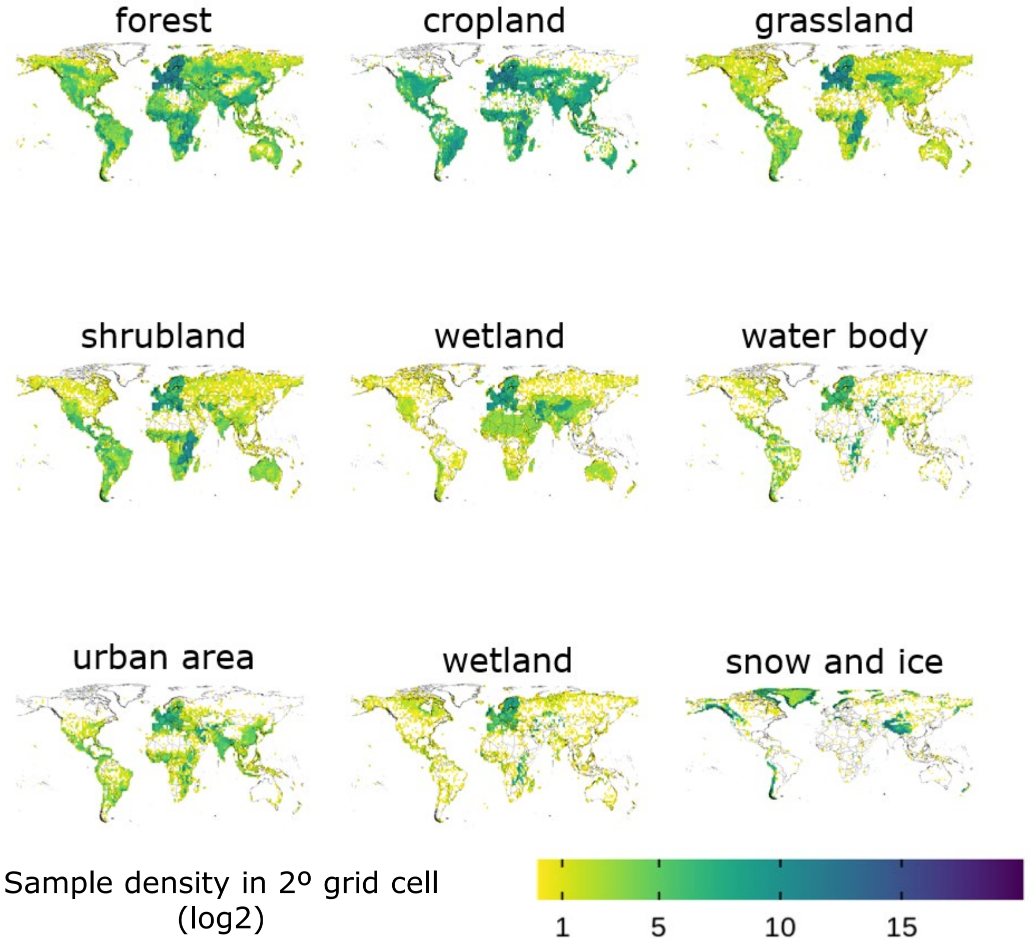
Training data optimization

- Geolocation errors of labels?
    - subset only those pixels that are surrounded by pixels with the same label
- Land cover changes ? –
    - subset sample sites of potential changes by running BFAST model and revising these sites
    - Set of rules based on spectral information
- Misclassifications at 10m resolution?
    - Set of rules based on spectral information

Lessons learned

- Having subpixel information about land cover is important for better defining classes at pixel level
- Homogeneous areas - it would be more convenient to label patches/segments rather than pixels
- There are always uncertainties associated with human labeling therefore additional data filtering is needed
  - E.g. taking into account spectral information
- The presented data will be made of open access

Collection of existing reference data sets


forest

cropland

grassland

shrubland

wetland

water body

urban area

wetland

snow and ice

~ 7 million samples

years: 1951-2020

spatial units: 10-5000 m

Sources:
LUCAS – Land use and land cover survey
GLIMS
Ramsar
GHS Urban Center Database
Global Croplands
JECAM
PRdataGO
…

caterina.barrasso@idiv.de | carsten.meyer@idiv.de

Macroecology and Society
A research group of iDiv

Thank you!

Myroslava Lesiv

International Institute for
Applied Systems Analysis

lesiv@iiasa.ac.at