Innovative methods for land cover/use mapping coping with limited availability of in-situ data

Lorenzo De Simone, PhD
Technical Adviser Geospatial
Office of the Chief Statistician,
FAO

EO-STAT

Iniciativa
Mano de la mano

Organización de las Naciones
Unidas para la Alimentación
y la Agricultura

Food and Agriculture Organization of the United Nations

CONTENTS

- Relevance of land cover and crop maps
- Global vs National LCLU Maps
- Challenges in the operational use of LCLU maps at the national level
- Innovative solutions to use less in-situ data and to make better use of it
- Alignment with the IGIF and the GSGF

RELEVANCE OF LAND COVER AND CROP TYPE MAPS

LAND COVER & LAND USE DATA ARE FUNDAMENTAL

Land cover and **land use** data have been included in the list of the <u>global fundamental geospatial data themes</u> by the Committee of Experts on Global Geospatial Information Management in 2018 (E/C.20/2018/7/Add.1).

UN-GGIM
United Nations Committee of Experts on Global Geospatial Information Management

Theme title:	**Land Cover and Land Use**

Description

Land cover represents the physical and biological cover of the Earth's surface. Land use is the current and future planned management, and modification of the natural environment for different human purposes or economic activities.

Why is this theme fundamental?

Land Cover data is required, for example, for developing land management policy, understanding spatial patterns of biodiversity and predicting effects of climate change. It may also help to forecast other phenomena, such as erosion or flooding. It is critical data in national assessments of biodiversity, conservation efforts, and water quality monitoring.

The use of the land informs land management impacts, especially on changes in natural resources, agriculture, conservation, and urban developments. Land cover and land use affect the greenhouse gases entering and leaving the atmosphere and provide opportunities to reduce climate change. It is required at a disaggregated level to allow local planning to manage and monitor land use at land parcel level.

Which sustainable development goals (SDGs) will it help to meet?

The theme plays a role in SDGs 1, 2, 3, 5, 6, 7, 8, 9, 11, 12, 13, 14 and 15.

Geospatial data features in more detail

Land Cover includes artificial surfaces, agricultural areas, forest, semi-natural areas, wetlands and waterbodies etc. Land Use in some ways describes the human activities and the consequences of such activities on the landscape.

Both Land Cover and Land Use are separated into different classes based on an agreed classification schema which is usually hierarchical. The data can be represented either as polygons or as a raster. It may also be found as attributes of a land parcel.

Possible sources of geospatial data

- Classified Earth observation (EO) data, potentially as a Data Cube;
- National datasets relating to environmental information and land parcels; and,
- International organisations, Regional United Nations Centre, different levels of public authorities (in particular municipalities) and the private sector.

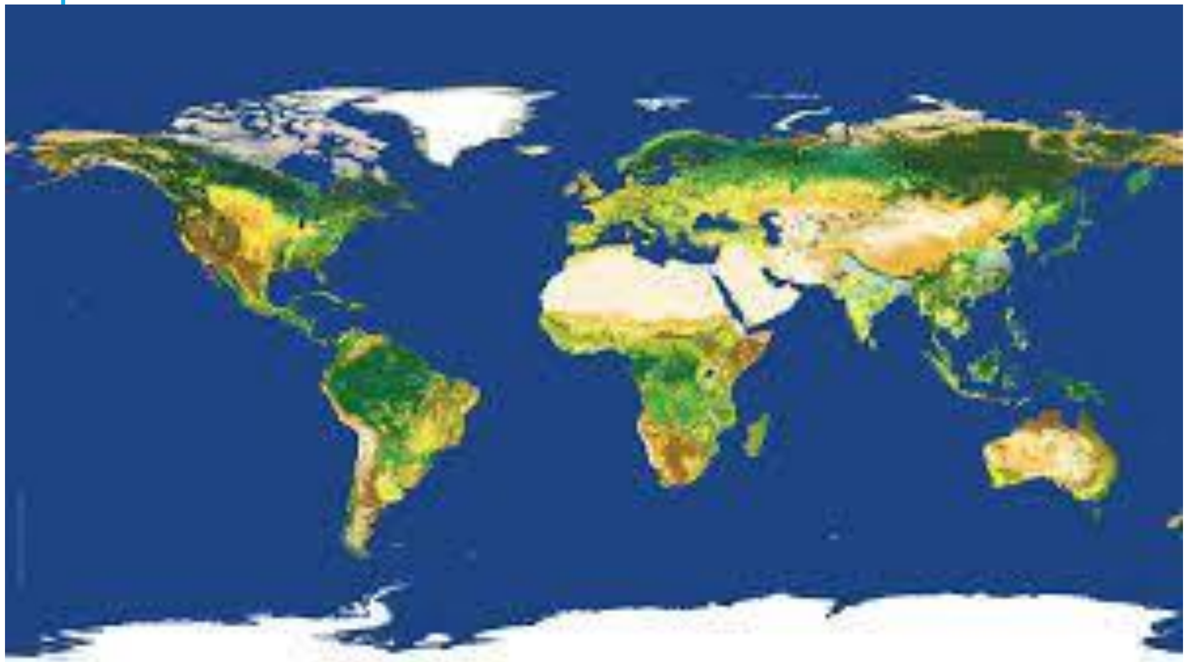
Existing geospatial data standards

Note: This is indicative. Other lists of standards exist and UN-GGIM will seek to work with thematic experts to develop a list of relevant data standards.

- ISO 19144-1:2009 – Geographic Information Classification system – Part 1 Classification system structure (last reviewed in and confirmed in 2015);
- ISO 19144-2:2012 - Part 2 - Land Cover Meta Language (LCML) (there are limitations on this standard);
- ISO 19115:2003 Geographic information – Metadata; and,
- INSPIRE data specification on Land Cover and on Land Use.

GLOBAL AND NATIONAL LAND COVER AND LAND USE MAPS

GLOBAL MAPS



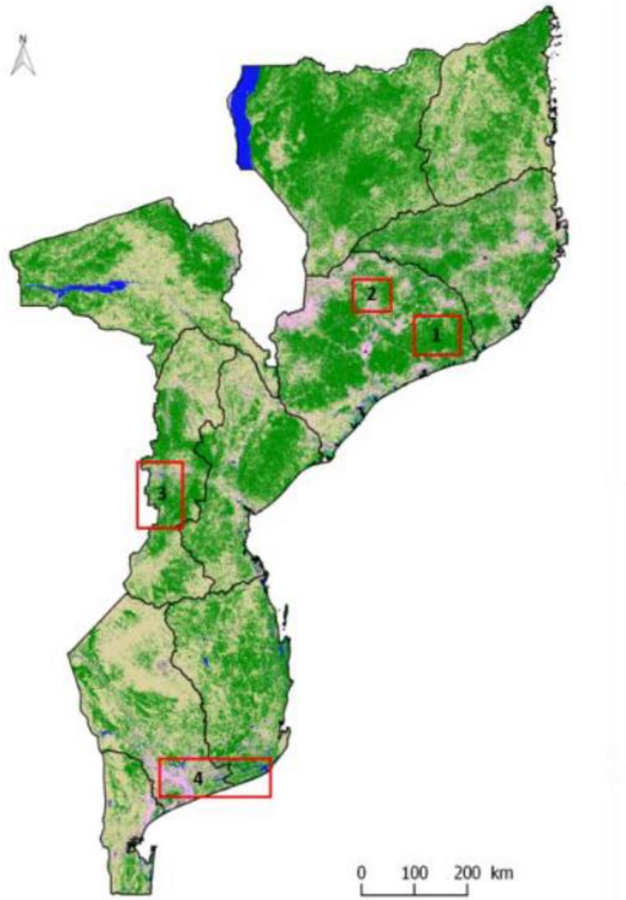
Pros:

- Wide availability: e.g. ESA WorldCover, ESA CCI, Copernicus, GlobeLand30, FROM-GLC30, etc.
- Available from 1992 through 2021
- Medium to high resolution (1km to 10m)
- Strong consistency at a global scale, but large deviations at the regional scale.
- Overall accuracy at the global level is from 72% to 80%.

Cons:

- Low accuracy at the regional scale and very low at the national level
- The legend may not satisfy national requirements
- Minor LC/LU classes are underestimated

NATIONAL LC MAPS



Mozambique, 2000-2005-2010-2016, Nitidae and CIRAD in the LAUREL project. Landsat, 30m.

National Maps

Pros

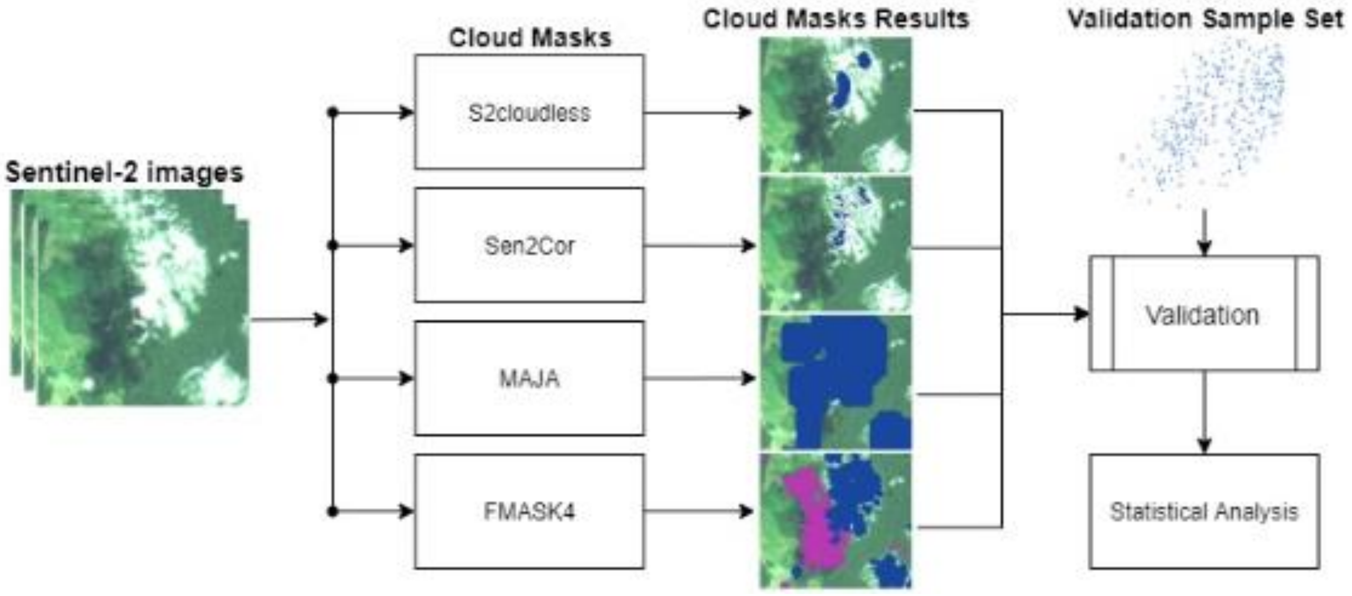
- High spatial accuracy and thematic accuracy
- Legend contains classes which satisfy national requirements

Cons

- Maps are often old and are not not kept up-to-date regularly
- Only few countries in the world produce their national land cover maps on a regular bases
- Uncertainty in accuracy measures
- Heterogeneity methods and data among different countries.
- Many of the maps are produced under the auspices of specific projects by third parties that come to an end.

MAIN CHALLENGES TO THE OPERATIONAL UPTAKE OF EO IN COUNTRIES FOR LCLU MAPPING

COMPLEXITY OF IMAGE PRE-PROCESSING



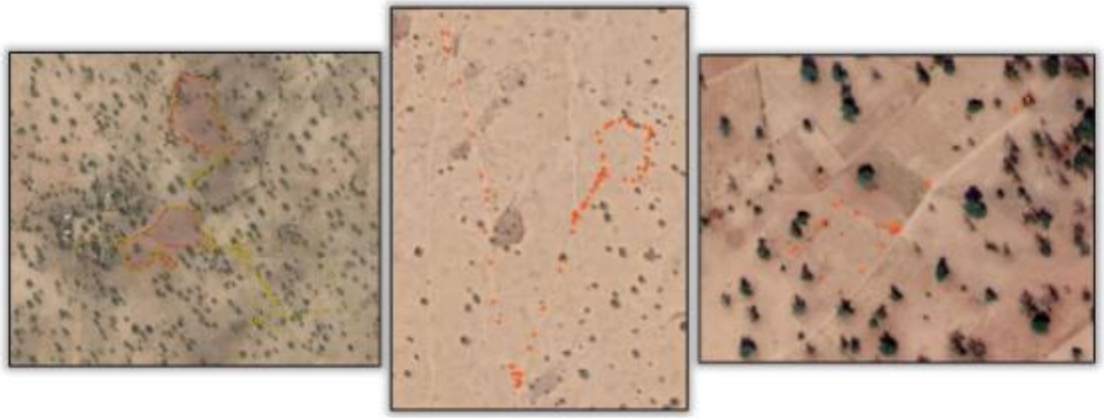
The complexity of image pre-processing (including image atmospheric correction and cloud masking), to more advanced temporal compositing and gap-filling which are required to derive analysis-ready data (ARD) also called data-cubes. Such operations are not trivial and require specialized expertise in Remote Sensing and big data handling.

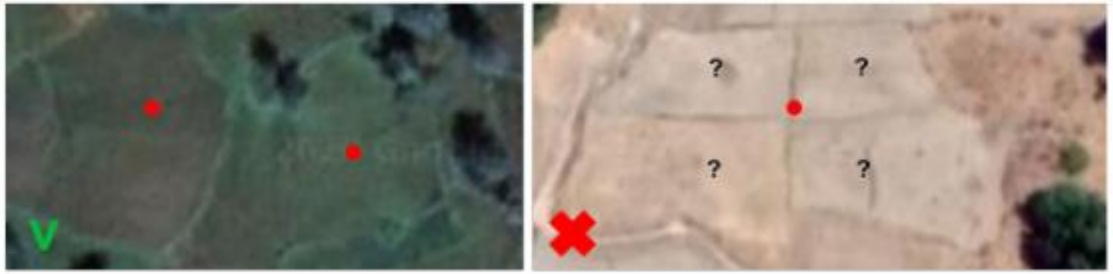
PROGRAMMING SKILLS REQUIRED

EO data platforms which provide user with access to EO data archives and analytical libraries, function through code editors. GEE uses Javascript, SEPAL and DE Africa Jupyter Notebooks (Python).

Many technical experts in NSOs and concerned line ministries who are very familiar with traditional GIS packages (QGIS, ArcGIS) may not be very familiar with scripting, and may prefer the use of graphic user interfaces.

LACK OF IN-SITU DATA OF ADEQUATE QUALITY





Top figure, example of GPS traces that contain more than parcels boundaries. Bottom figure, challenge to localize the surveyed parcels in the statistical database due to the fact that parcels are localized by geo-points (i.e. a single GPS coordinate)

In-situ data are essential for the automatic classification of satellite images into land cover or land use classes, as opposed to visual interpretation.

In-situ data of sufficient quality is rare to find. Common reasons for this are:

I) high costs of field survey campaigns

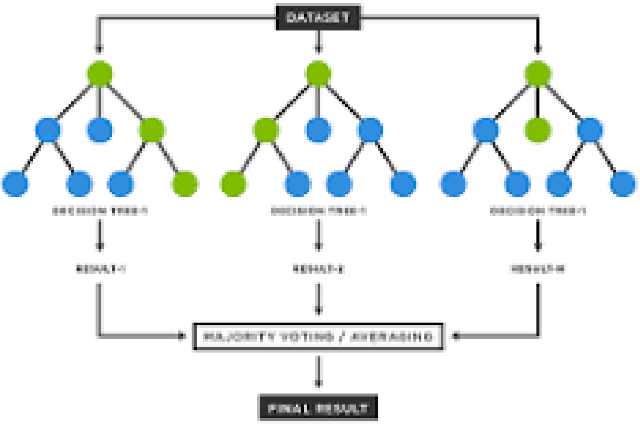
- Many data points to collect (e.g.12K for one country)
- Transportation cost, plus human resources
- Need to repeat survey for every reporting year

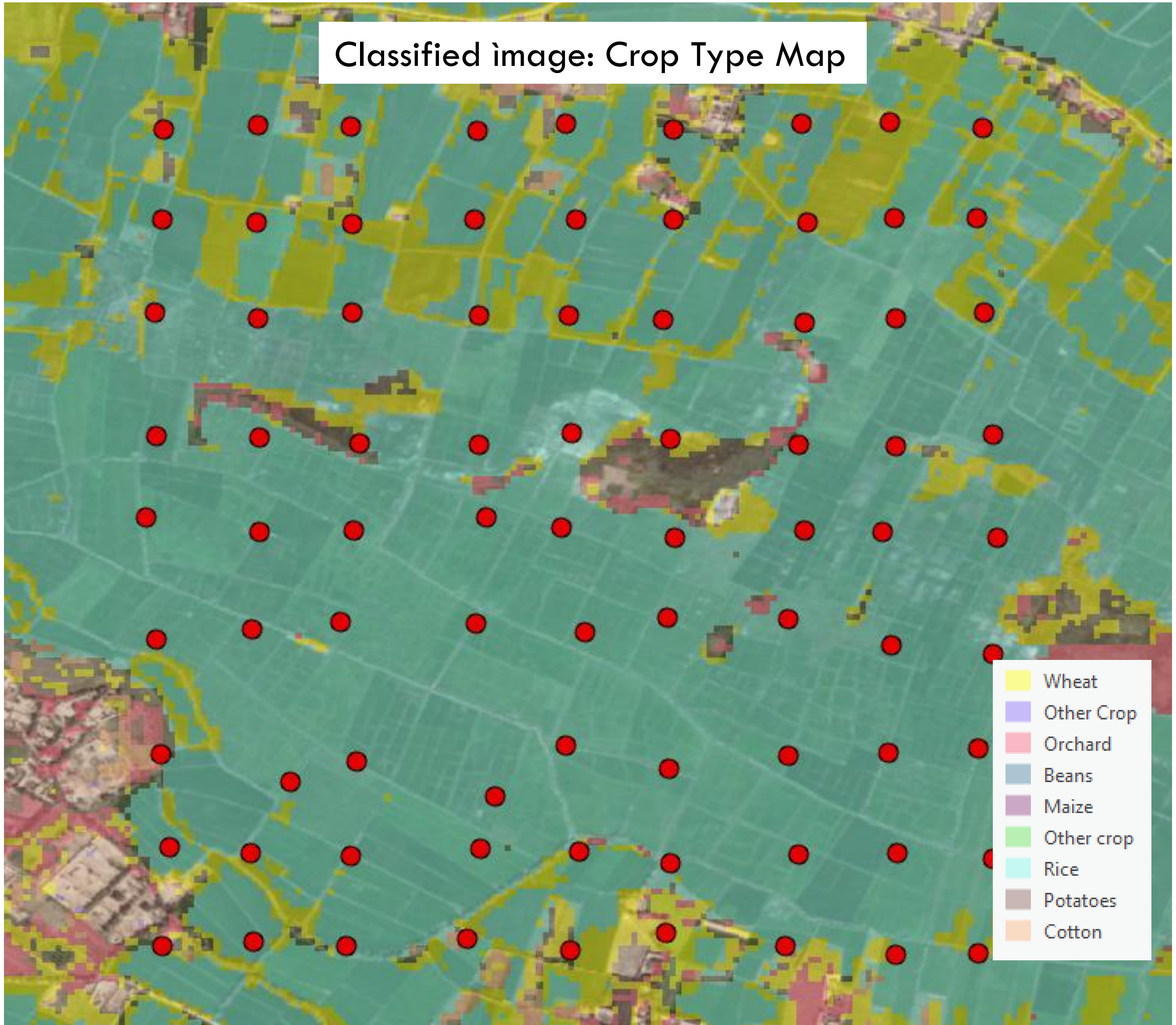
ii) suboptimal stratification of the samples

iii) suboptimal geo-referencing methods used in the field.

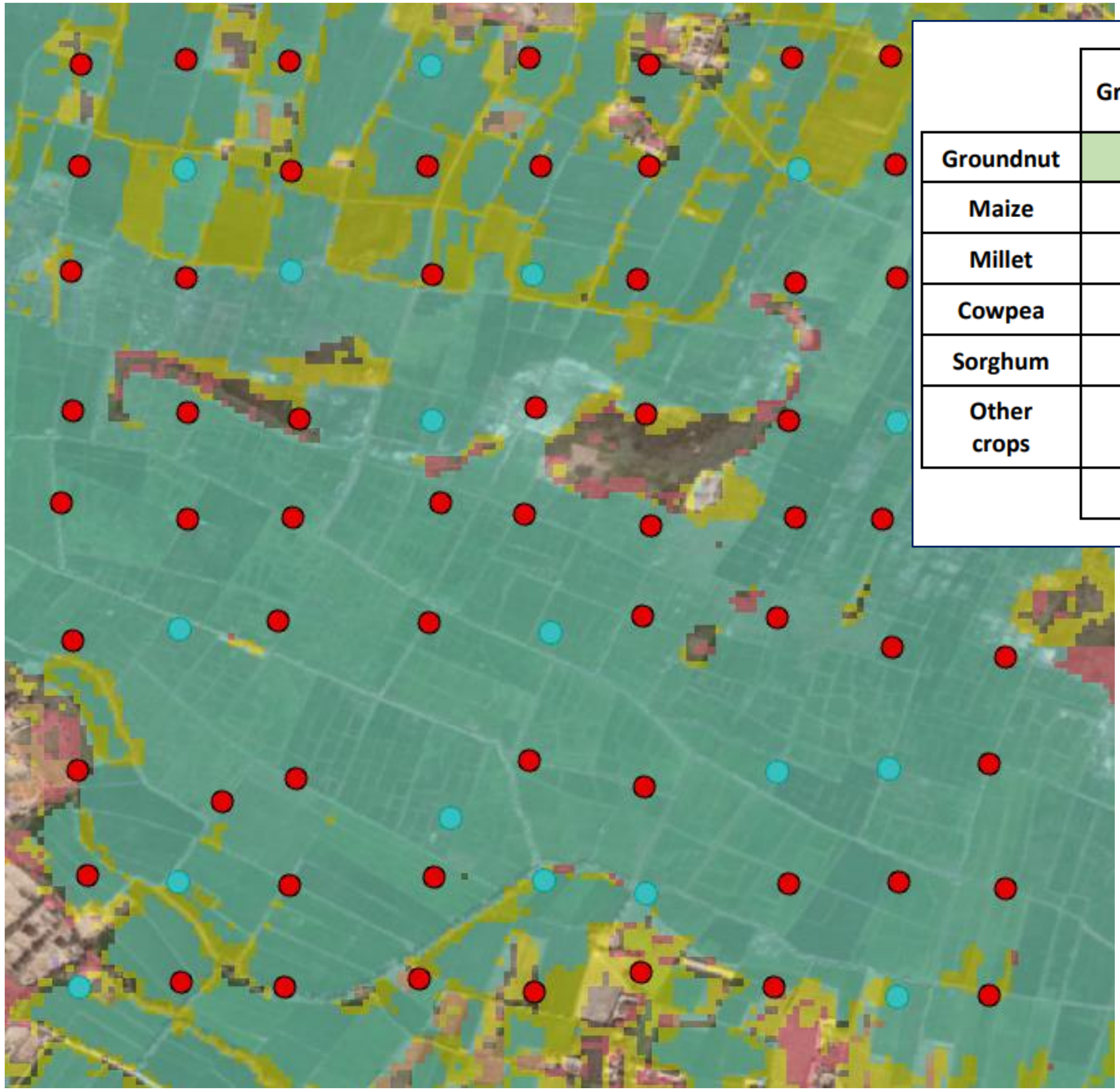
WHY IN-SITU DATA IS NEEDED - CALIBRATION

In-situ data is used for the calibration of classification algorithms (e.g. Random Forest) and for the validation of results, allowing for calculating the accuracy of map (OA, UA and PA) and estimate the standard error in area estimation and the confidence intervals based on 95% confidence.





WHY IN-SITU DATA IS NEEDED – VALIDATION



	Groundnut	Maize	Millet	Cowpea	Sorghum	Other crops	
Groundnut	13172	289	233	178	79	184	93%
Maize	578	1110	284	0	136	162	49%
Millet	631	600	6282	87	193	88	80%
Cowpea	329	19	81	1203	1	20	73%
Sorghum	106	651	162	0	590	42	38%
Other crops	959	46	239	257	104	2076	56%
	83%	41%	86%	70%	53%	81%	78%

Accuracy Statistics
Overall Accuracy
Producer Accuracy
User Accuracy
Kappa Statistics
F1-Statistics

INNOVATION TO OVERCOME THE CHALLENGE OF LACK OF IN-SITU DATA OF ADEQUATE QUALITY FOR EO USE

OVERCOMING THE SHORTAGE OF IN-SITU DATA BY ADDRESSING 2 KEY QUESTIONS?

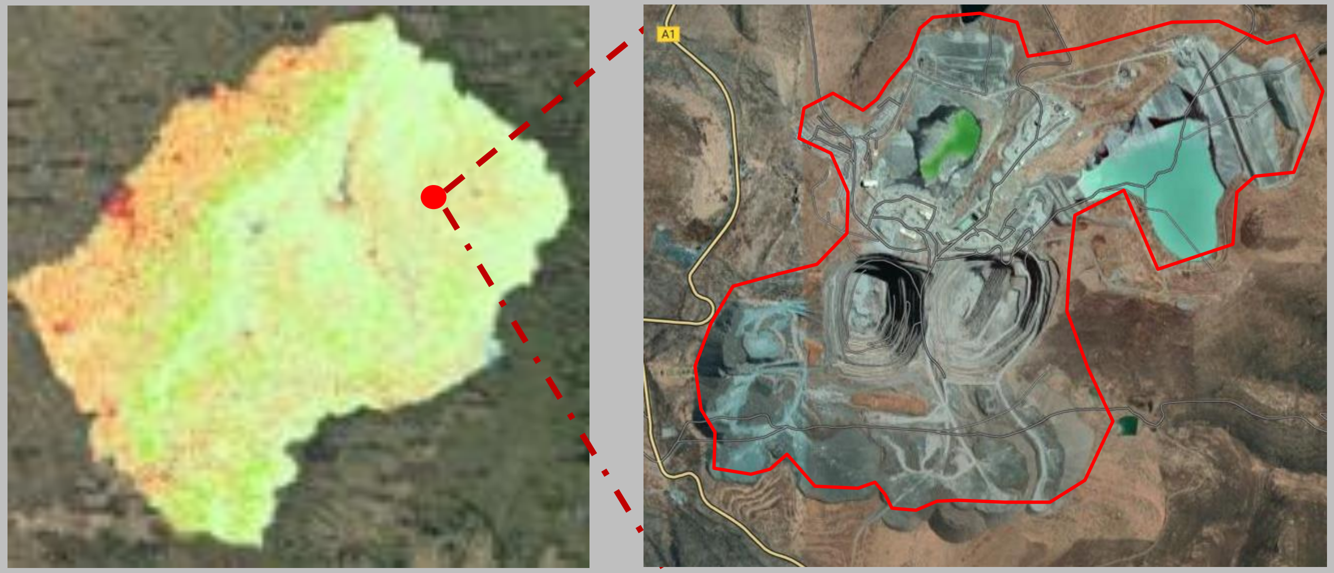
I. Can we reuse in-situ data?
II. Can we use data frugal classification algorithms?

CAN WE REUSE IN SITU DATA?

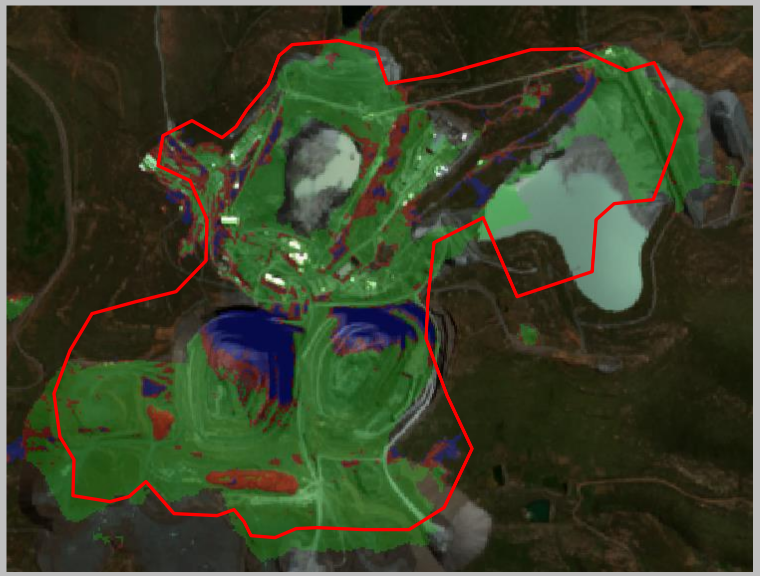
The question is if we have an in-situ dataset with high positional and attribute accuracy (we are lucky) for a given baseline year, can we re-use such a dataset to train and validate a classifier for a different year?

The issue in re-using the in-situ data for a different year is connected to the risk that the land cover or the land use at that specific location may have changed in the reference year, and therefore the in-situ data point would be outdated. Using outdated in-situ data to train the classifier would introduce bias. Similarly, using the outdated in-situ dataset for validation could hide commission errors.

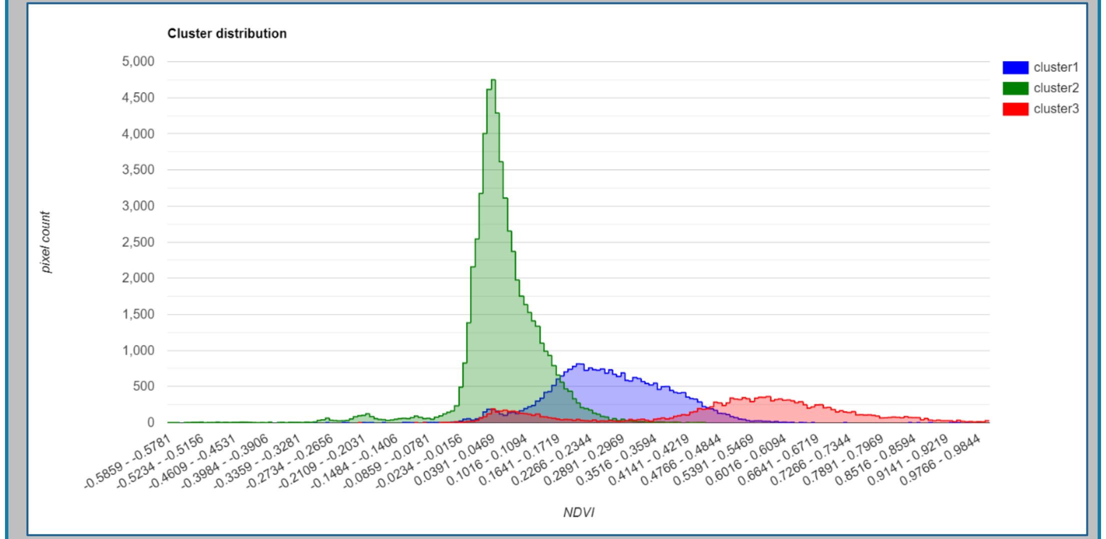
However, it is possible to use EO data (e.g NDVI) to assess the consistency of the spectral characteristics of a pixel in time, and therefore to judge whether the LC/LU at this location has changed. In this context, we can use a modified version of the method developed by Paris and Bruzzone based on K-Means clustering to artificially update an in-situ dataset.

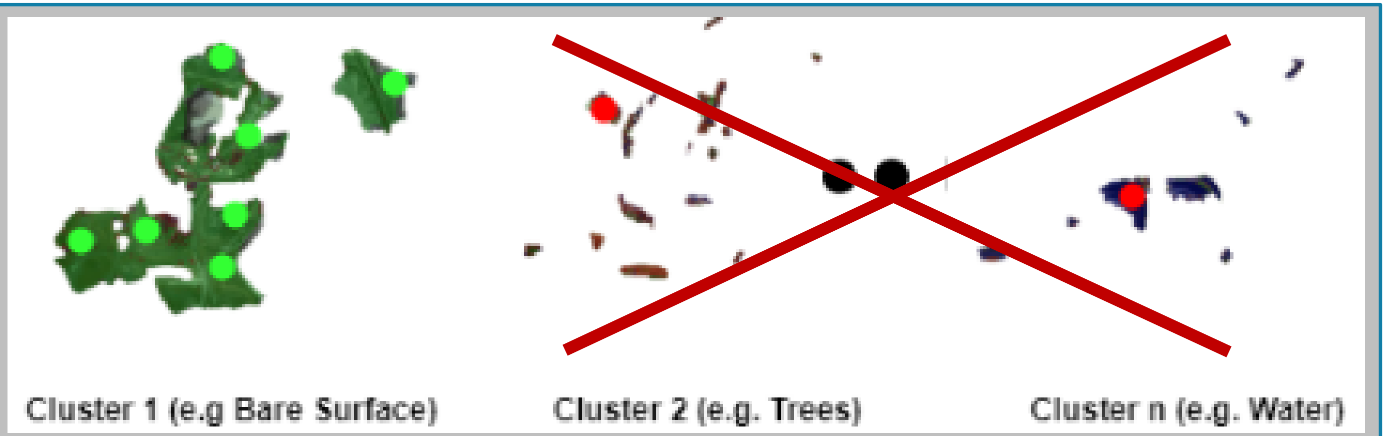


Step I: extraction of a land cover stratum from an existing LC baseline 2021. In this example, we extract only one feature for the 'Mining' class and overlay it on a true color image background.



Step II we apply the Kmeans on NDVI for our target year, 2019. Above we can clusters of pixels inside the target feature, and the distribution of pixels per cluster (below)

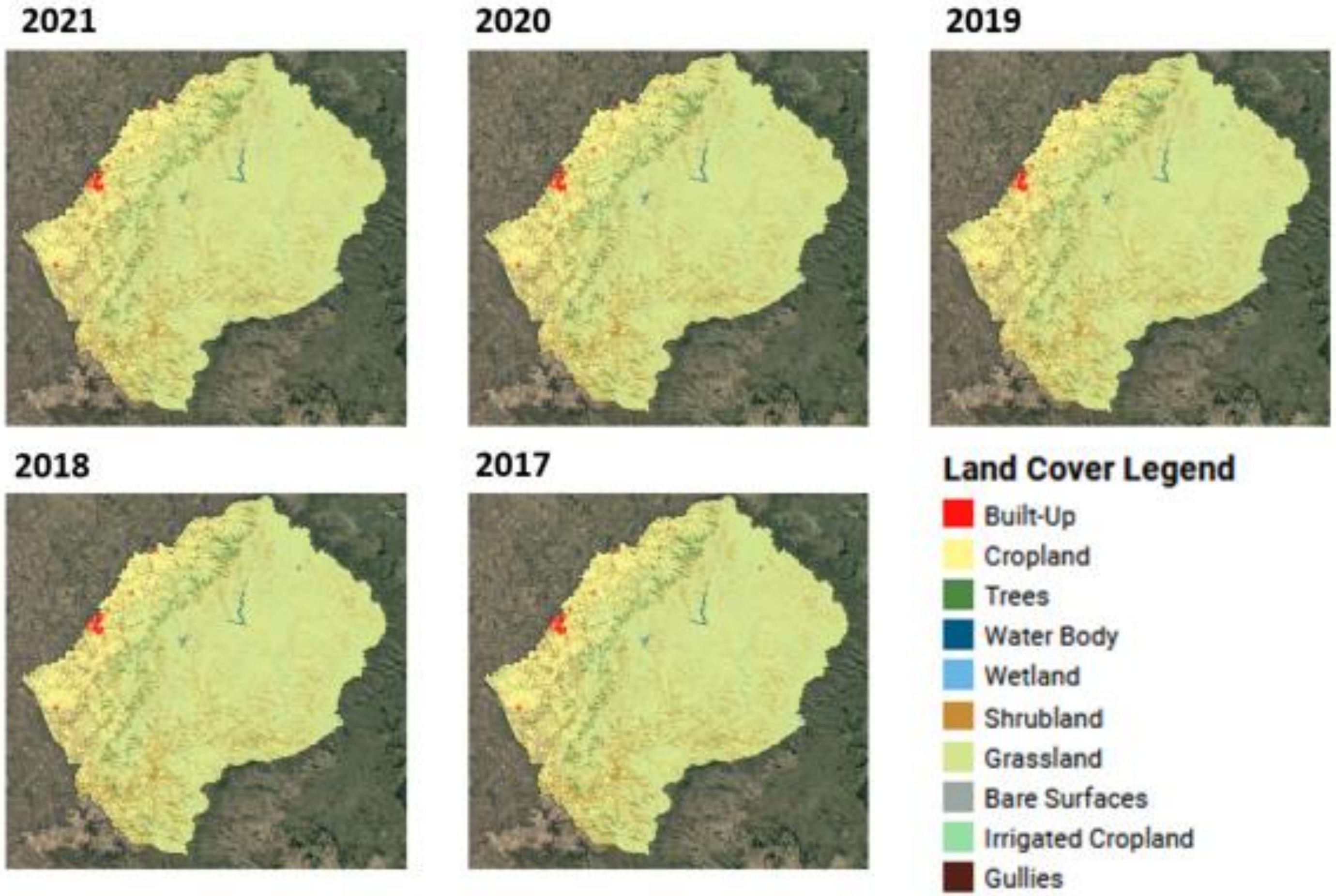


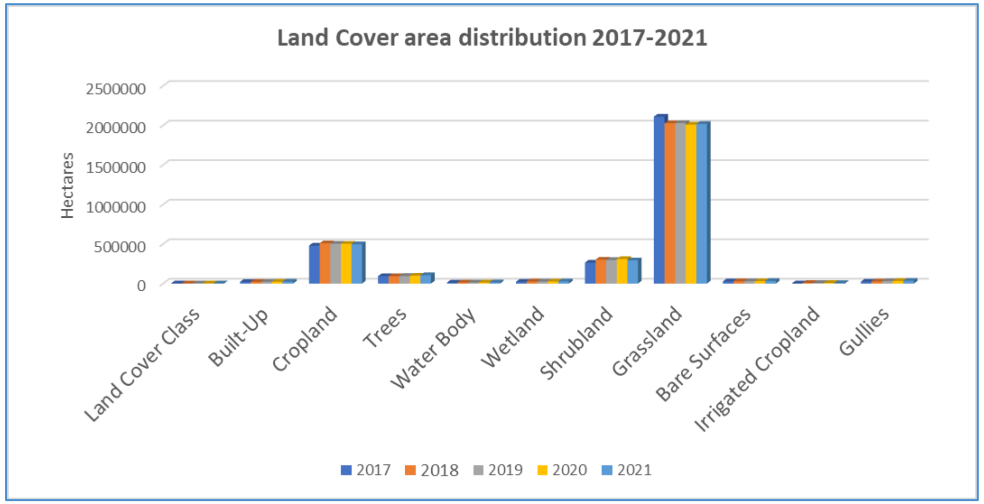


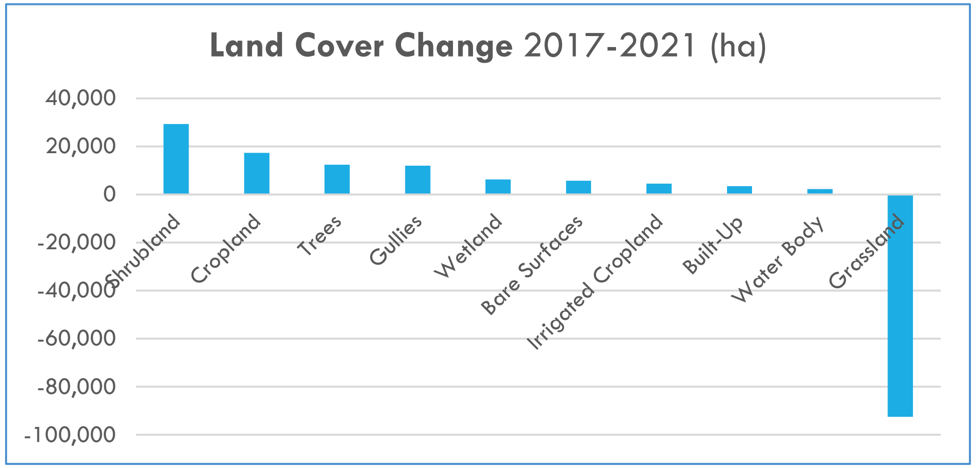
Step III: We select the pixels from the dominant cluster and we filter out minor clusters

LESOTHO

5 Land Cover Maps 2017-2021







Overall Accuracy: 87% to 84%
Spatial Resolution: 10 meters
Geographic scope: National
Reference year(s): 2017 - 2021
Crop classes: 10
HiH sharing status: published

Project funded by a EU-GIZ trust fund

remote sensing

MDPI

Article

Operational Use of EO Data for National Land Cover Official Statistics in Lesotho

Lorenzo De Simone [1,*], William Ouellette [2] and Pietro Gennari [1]

[1] Office of the Chief Statistician, Food and Agriculture Organization of United Nations, 00153 Rome, Italy; pietro.gennari@fao.org
[2] FAOLS, Maseru 7588, Lesotho; william.ouellette@fao.org
* Correspondence: lorenzo.desimone@fao.org

Abstract: The Food and Agriculture Organization of the United Nations (FAO) is building a land cover monitoring system in Lesotho in support of ReNOKA ('we are a river'), the national program for integrated catchment management led by the Government of Lesotho. The aim of the system is to deliver land cover products at a national level on an annual basis that can be used for global reporting of official land cover statistics and to inform appropriate land restoration policies. This paper presents an innovative methodology that has allowed the production of five standardized annual land cover maps (2017–2021) using only a single in situ dataset gathered in the field for the reference year, 2021. A total of 10 land cover classes are represented in the maps, including specific features, such as gullies, which are under close monitoring. The mapping approach developed includes the following: (i) the automatic generation of training and validation datasets for each reporting year from a single in situ dataset; (ii) the use of a Random Forest Classifier combined with postprocessing and harmonization steps to produce the five standardized annual land cover maps; (iii) the construction of confusion matrixes to assess the classification accuracy of the estimates and their stability over time to ensure estimates' consistency. Results show that the error-adjusted overall accuracy of the five maps ranges from 87% (2021) to 83% (2017). The aim of this work is to demonstrate a suitable solution for operational land cover mapping that can cope with the scarcity of in situ data, which is a common challenge in almost every developing country.

Keywords: supervised classification; automatic generation of training and validation data; Sentinel-2 temporal composites; Random Forest Classifier; land cover class accuracy stability

check for updates

Citation: De Simone, L.; Ouellette, W.; Gennari, P. Operational Use of EO Data for National Land Cover Official Statistics in Lesotho. *Remote Sens.* **2022**, *14*, 3294. https://doi.org/10.3390/rs14143294

Academic Editors: Conghe Song and Dino Ienco

Received: 1 June 2022
Accepted: 4 July 2022
Published: 8 July 2022

1. Introduction

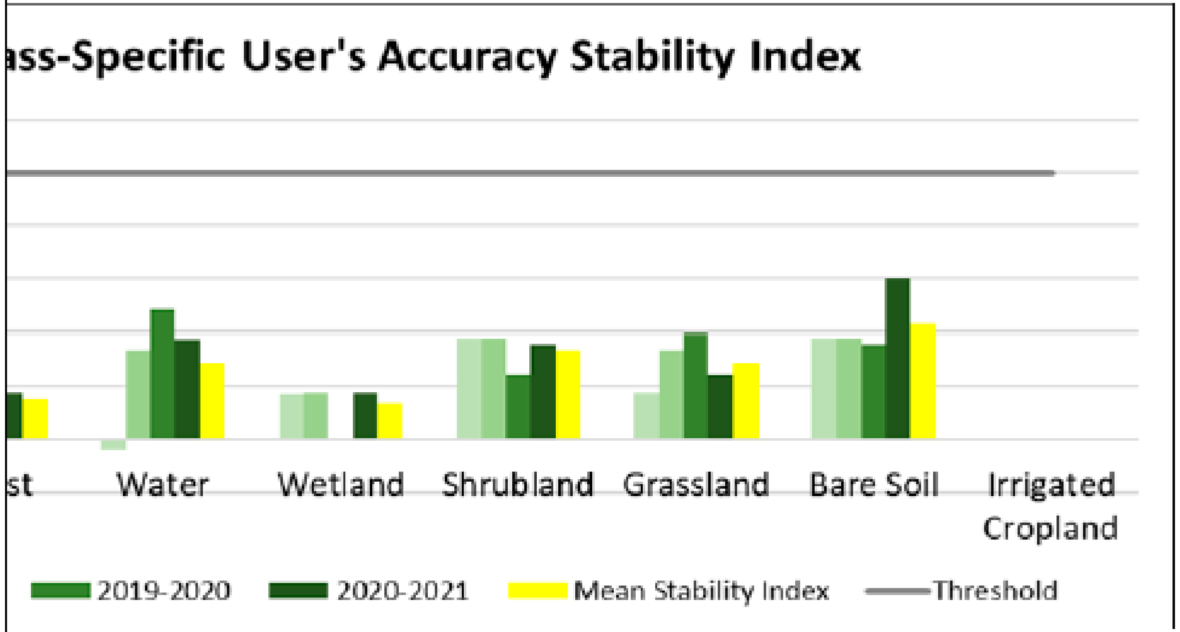
Land Cover (LC) maps can be used to extract key information for a series of national applications, such as environmental monitoring, identification of land degradation trends, spatial planning, and for a wide range of scientific research fields. However, continuous monitoring and reporting of land cover maps requires regular updating, the use of standardized methods, and the adoption of a robust validation framework ensuring that every estimate is accurate and consistent over time. Such land cover mapping solutions are very rare to find in countries due to the inherent technical and financial challenges found in both traditional and modern LC mapping methods.

The most traditional methods that have been typically used in the last two decades have been based, initially, on visual image interpretation and pixel (or object) classification, relying on the use of very high-resolution images (commercial satellite images and orthophotos), and subsequently, on the combination of Earth Observations and in situ data for calibration and validation of automatic classification models. Such solutions have been extensively used in the research community [1–4].

FAO adopted a visual interpretation approach in 2015 to deliver the first edition of the Lesotho Land Cover Atlas [5]. The methodology relied on a manual labeling of segmented

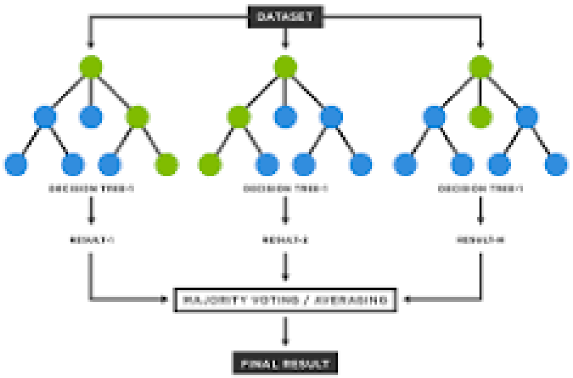
acy Assessment for each year

r	Wetland	Shrubland	Grassland	Bare Soil	Irrigated Cropland	User Accuracy
0	2	0	0	1	0	41.67
0	1	1	13	0	0	82.35
1	4	13	11	0	0	94.37
29	1	0	0	0	0	90.63
0	418	0	29	0	0	91.07
0	1	36	33	1	0	48.00
0	5	11	169	9	0	81.64
0	0	0	2	9	0	60.00
0	0	0	3	0	3	50.00
96.67	96.76	59.02	65.00	45.00	100.00	

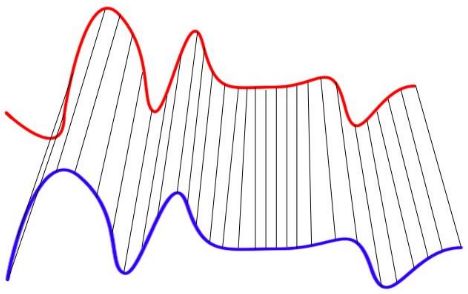


CAN WE USE LESS IN-SITU DATA?

- **Random Forest** (RF) is the de-facto mostly used supervised classifier. RF though necessitates of large amounts of training data to avoid overfitting, and this means thousands of data points.



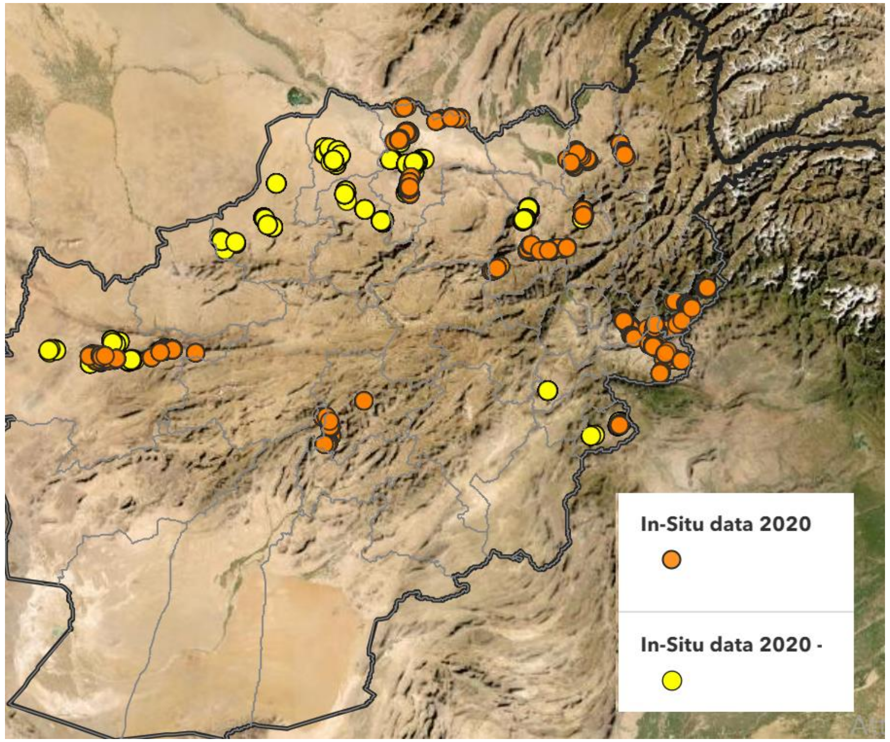
- FAO has explored the possibility to use a data frugal algorithm (**Dynamic Time Warping**), which instead works by assessing the similarities by pairs of time series data and allows also for coping with misalignment between these

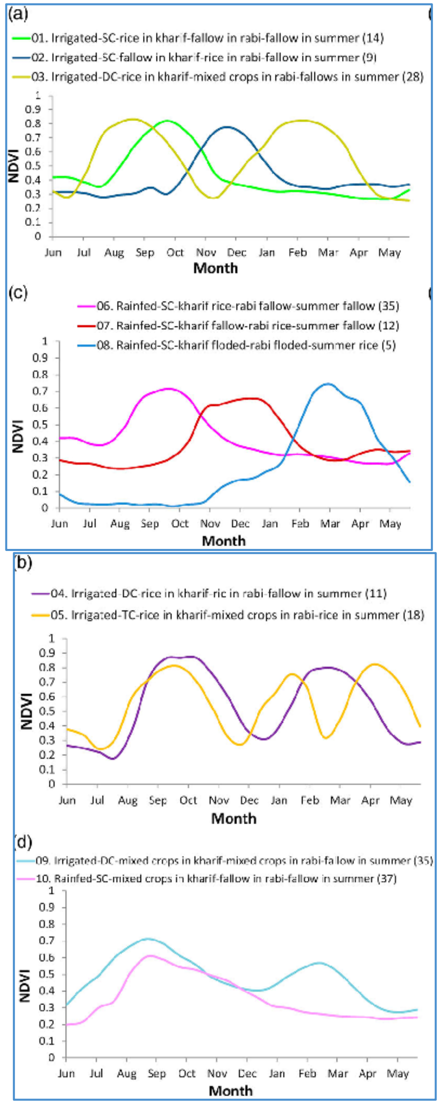


Dynamic Time Warping Matching

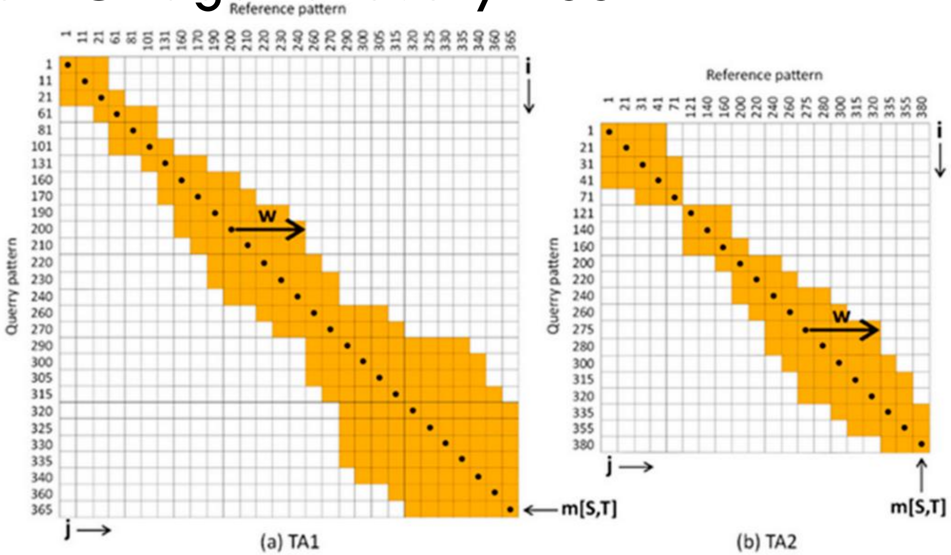
THREE STEPS STEPS OF DTW

Step 1 Define spectral signatures for different crops in the study area.

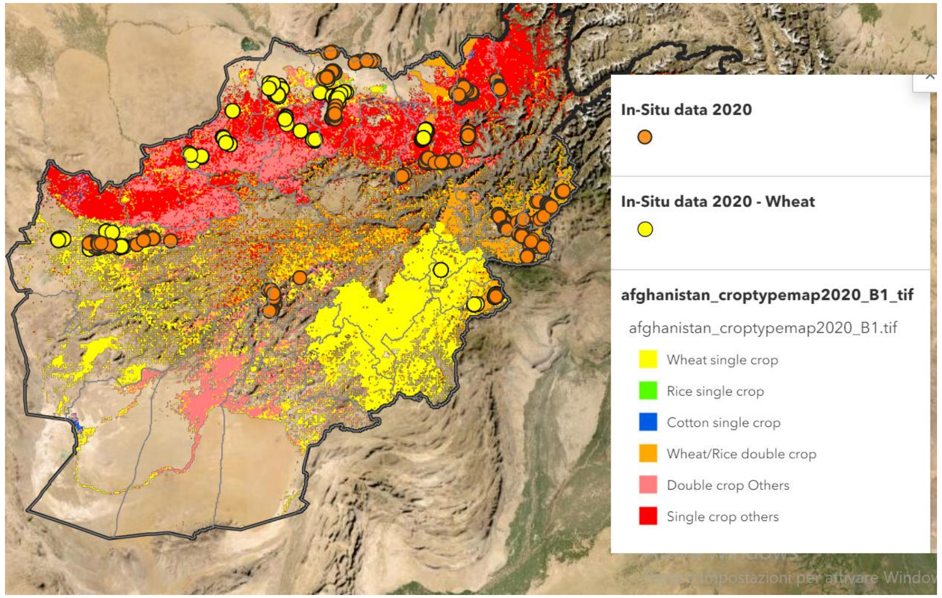


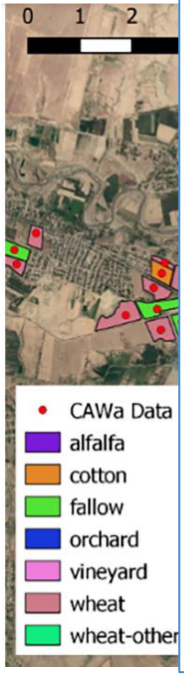


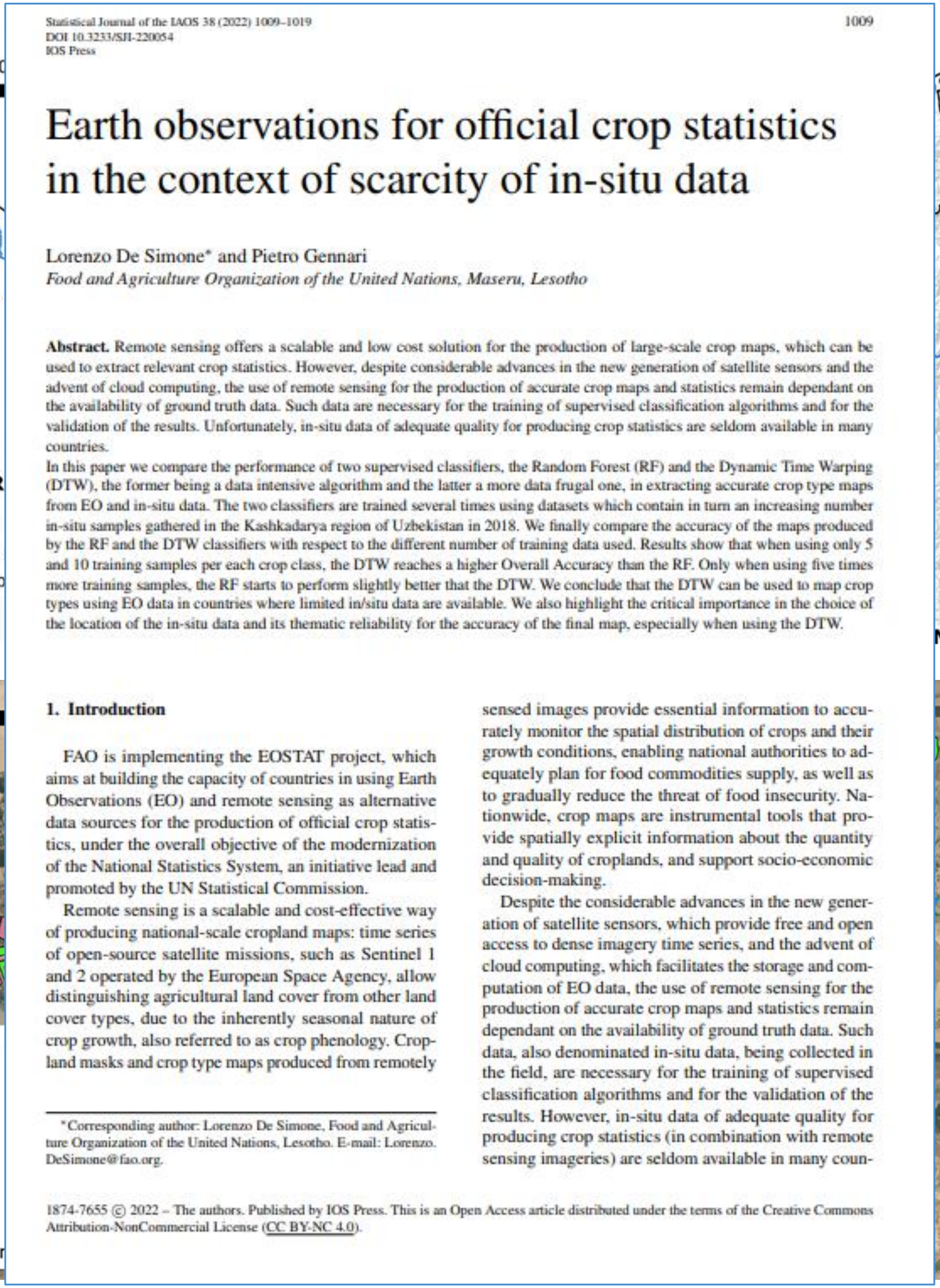
Step 2 Dissimilarity analysis of each pixels in the AOI against every RSS



Step 3 Classification of pixels based on match with least min dissimilarity







Statistical Journal of the IAOS 38 (2022) 1009–1019
DOI 10.3233/SJI-220054
IOS Press

Earth observations for official crop statistics in the context of scarcity of in-situ data

Lorenzo De Simone* and Pietro Gennari
Food and Agriculture Organization of the United Nations, Maseru, Lesotho

Abstract. Remote sensing offers a scalable and low cost solution for the production of large-scale crop maps, which can be used to extract relevant crop statistics. However, despite considerable advances in the new generation of satellite sensors and the advent of cloud computing, the use of remote sensing for the production of accurate crop maps and statistics remain dependant on the availability of ground truth data. Such data are necessary for the training of supervised classification algorithms and for the validation of the results. Unfortunately, in-situ data of adequate quality for producing crop statistics are seldom available in many countries.

In this paper we compare the performance of two supervised classifiers, the Random Forest (RF) and the Dynamic Time Warping (DTW), the former being a data intensive algorithm and the latter a more data frugal one, in extracting accurate crop type maps from EO and in-situ data. The two classifiers are trained several times using datasets which contain in turn an increasing number in-situ samples gathered in the Kashkadarya region of Uzbekistan in 2018. We finally compare the accuracy of the maps produced by the RF and the DTW classifiers with respect to the different number of training data used. Results show that when using only 5 and 10 training samples per each crop class, the DTW reaches a higher Overall Accuracy than the RF. Only when using five times more training samples, the RF starts to perform slightly better that the DTW. We conclude that the DTW can be used to map crop types using EO data in countries where limited in/situ data are available. We also highlight the critical importance in the choice of the location of the in-situ data and its thematic reliability for the accuracy of the final map, especially when using the DTW.

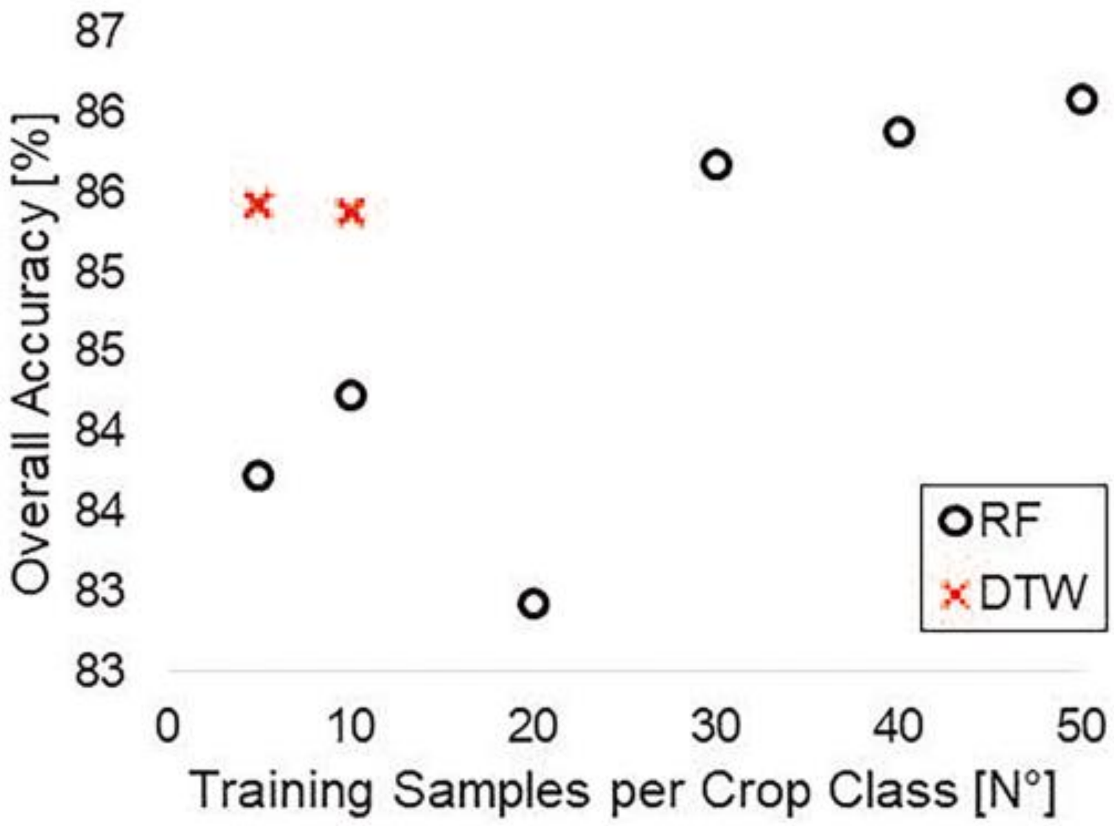
1. Introduction

FAO is implementing the EOSTAT project, which aims at building the capacity of countries in using Earth Observations (EO) and remote sensing as alternative data sources for the production of official crop statistics, under the overall objective of the modernization of the National Statistics System, an initiative lead and promoted by the UN Statistical Commission.

Remote sensing is a scalable and cost-effective way of producing national-scale cropland maps: time series of open-source satellite missions, such as Sentinel 1 and 2 operated by the European Space Agency, allow distinguishing agricultural land cover from other land cover types, due to the inherently seasonal nature of crop growth, also referred to as crop phenology. Cropland masks and crop type maps produced from remotely sensed images provide essential information to accurately monitor the spatial distribution of crops and their growth conditions, enabling national authorities to adequately plan for food commodities supply, as well as to gradually reduce the threat of food insecurity. Nationwide, crop maps are instrumental tools that provide spatially explicit information about the quantity and quality of croplands, and support socio-economic decision-making.

Despite the considerable advances in the new generation of satellite sensors, which provide free and open access to dense imagery time series, and the advent of cloud computing, which facilitates the storage and computation of EO data, the use of remote sensing for the production of accurate crop maps and statistics remain dependant on the availability of ground truth data. Such data, also denominated in-situ data, being collected in the field, are necessary for the training of supervised classification algorithms and for the validation of the results. However, in-situ data of adequate quality for producing crop statistics (in combination with remote sensing imageries) are seldom available in many coun-

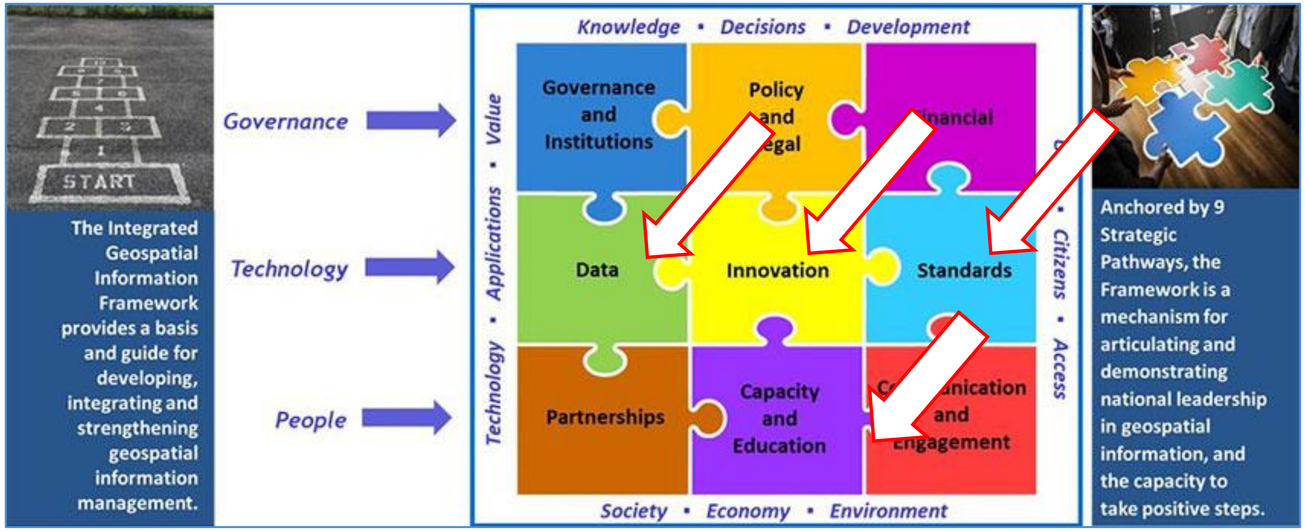
*Corresponding author: Lorenzo De Simone, Food and Agriculture Organization of the United Nations, Lesotho. E-mail: Lorenzo.DeSimone@fao.org.



Overall accuracy of the DTW and RF classifiers as a function of the number of training samples used.

ALIGNMENT WITH IGIF AND GSGF

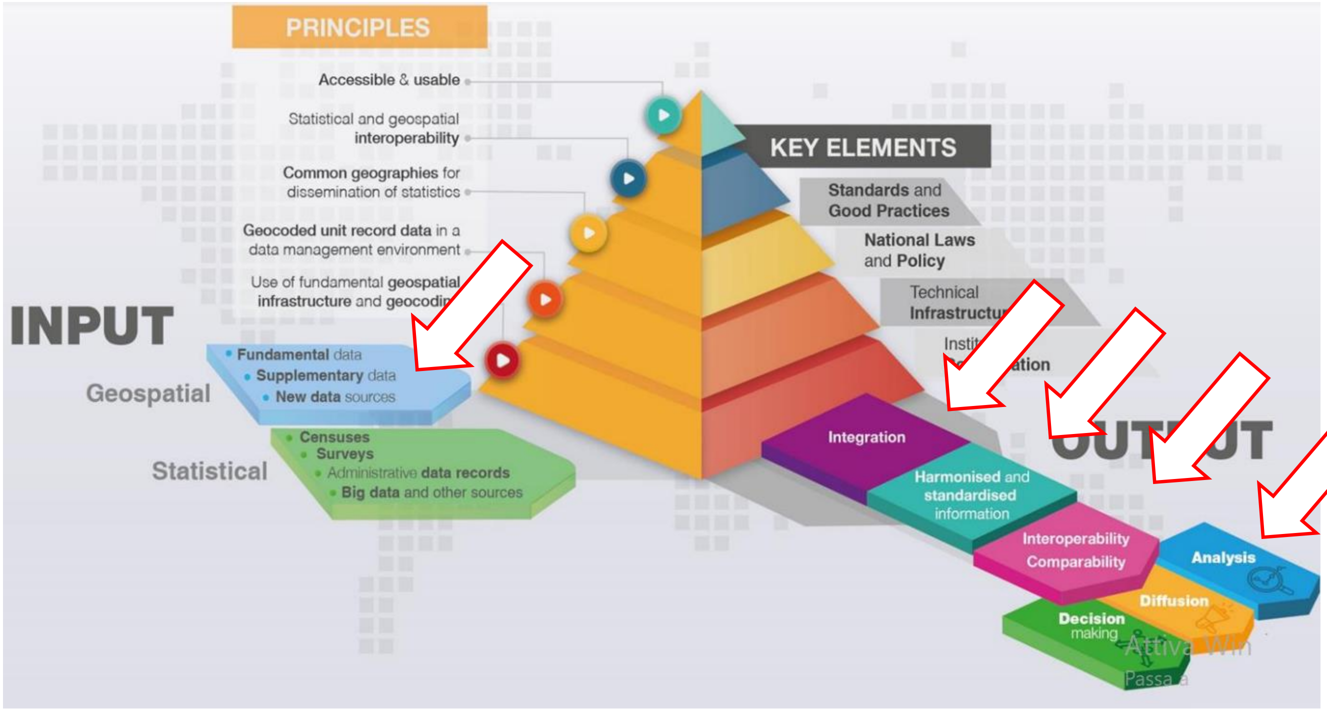
Integrated Geospatial Information Framework



FAO's work provides direct inputs to the Data, Innovation, Standards and Capacity and Education Pathways

Can support the implementation of the IGIF providing standardized methods and tools

Global Statistical Geospatial Framework



Direct contribution as:

1) Input: Fundamental Data
2) Output:
   I. Integration
   II. Standardization
   III. Interoperability
3) Analysis: land cover statistics, land cover change analysis etc.

THANK YOU

Standardization of EO Methods

Standardizatiomn of in situ data requirements

Definition of wuality criterias for LC and LU maps

Integration of standardized methods to produce fundamental geospatila data themes within the IGIF