United Nations
Statistical Commission
55th Session
ECONOMY • ENVIRONMENT • POPULATION • SOCIETY • TRADE

High-level virtual events

OPEN MEETING OF THE COMMITTEE OF EXPERTS ON FOOD SECURITY, AGRICULTURAL AND RURAL STATISTICS (UN-CEAG)

Achievements (2020-23) and Programme of work (2024-27)

20 February 2024
9:00 am – 11:30 am EST
Online

United Nations | Department of Economic and Social Affairs

Introduction

Susana Patricia Pérez Cadena
(National Institute of Statistics and Geography, Mexico)

Election of new UN-CEAG Chair

Susana Patricia Pérez Cadena

(National Institute of Statistics and Geography, Mexico)

SESSION 1

Guidelines on Processing Food Consumption Data from Household Consumption and Expenditure Surveys – for endorsement at the 55th session of the UN Statistical Commission

Astrid Mathiassen (Statistics Norway)

Nathalie Troubat (SPC)

Outline of this presentation

- ❑ About the guidelines
  - Why?
  - How?
  - What?
- ❑ Content of the guidelines
- ❑ Results of country consultation

About the guidelines

Why?

- Food data collected in Household Consumption and Expenditure Surveys (HCES) provide core information for poverty, food security or other economic analysis.
- BUT
  - data collected is comprehensive and complex to process and,
  - users, based on their needs or interests quite often follow different approaches when preparing the data for analysis.
- When data from the same survey is processed independently for different uses, it quite often leads to inconsistent results, it is inefficient and costly.
- At the 5th meeting of the IAEG-AG (ex UNCEAG) StatNorway proposed to produce "new guidelines on how to prepare consumption data from household budget surveys that could be used at the same time for producing poverty and food security statistics" → proposal was endorsed and creation of a task team

How?

- Joint SBS/SPC/WB/FAO collaboration with NSOs and other regional organizations such as COMESA
- A long walk… that started in 2019

What?

- The guidelines provide a set of standard recommendations to follow when processing food data from HCES
- The primary goal of these guidelines is to assist data owners in following one standard process when preparing their data for the main users and proposing a unique dataset on **quantity, dietary energy and monetary value** for every food item consumed by the household, from every source of consumption, to be used for further analysis.
- It is a step-by-step guide
- The guidelines are based on the IAEG-AG 2018 guidelines on how to better capture food consumption data and harmonise survey design worldwide, to derive global monitoring indicators that can be compared over time and between countries.
- One survey design = one survey processing = comparability over time

⚠ The new guidelines do not aim to substitute well-functioning national systems or approaches already established by NSOs for their food data processing.

Content of the guidelines

Structure of the guidelines

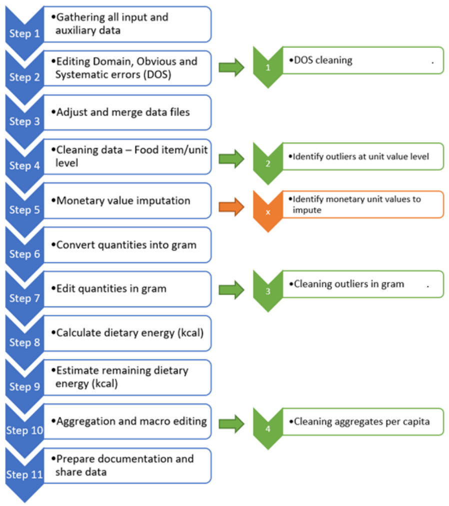
- INTRODUCTION
  - Background
  - About the guidelines for processing food consumption data
  - Outline of the guidelines
  - Recommendations
- PART 1: REVIEWING MODE OF DATA COLLECTION AND CLEANING
  - Food consumption data collection modules
  - Approaches to cleaning food data
- PART 2: THE STEP-BY-STEP PROCESS
- ANNEXES
- REFERENCES
- GLOSSARY & TERMINOLOGY

Recommendations

- ✓ Ensure that the processing of food consumption data from HCES is done in a single process and accommodates all the main users
- ✓ Data processing must result in a dataset that can be used for all the main analyses, and statistics can be disaggregated for all relevant populations
- ✓ Respect consistency in data processing between surveys
- ✓ Create one Nutrient Conversion Table (of good quality) for the food items captured in the survey in cooperation with nutrition experts and ensure consistency in its use
- ✓ Document all steps and all the decisions made throughout the process
- ✓ Share practices

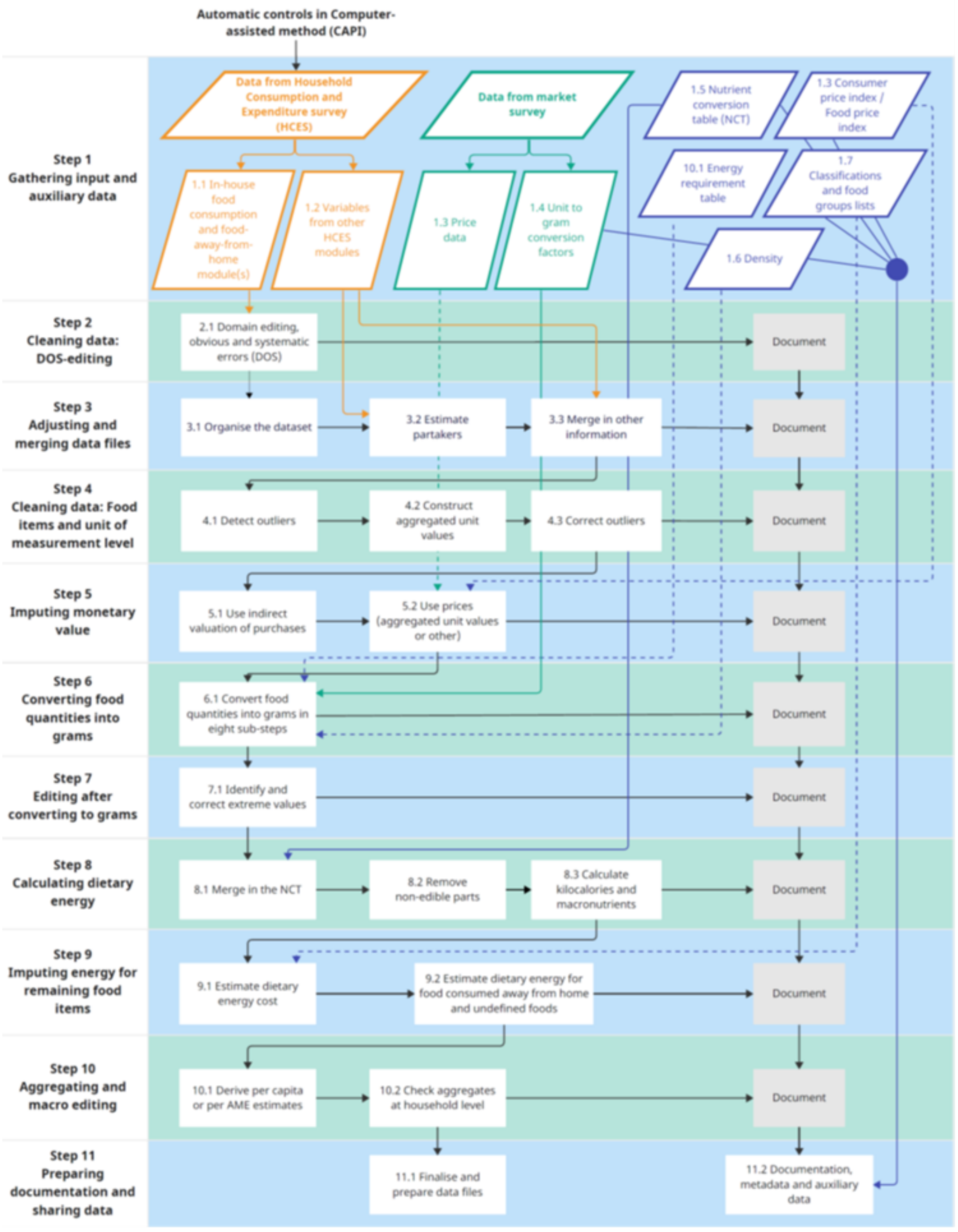
Section 1 – reviewing mode of data collection and data cleaning

- The guidelines assume that the food data collected in the HCES follows the WB/FAO guidelines on “Food Data Collection in Household Consumption and Expenditure Surveys: Guidelines for Low- and Middle-Income Countries” prepared by the IAEG-AG (ex UN-CEAG)
- The guidelines also involves a four stage of data cleaning
  - The first stage of data cleaning refers to ‘domain, obvious and systematic’ (DOS) error editing. This is where the first and most basic editing is done
  - The second stage comes before the quantities are transformed into grams
  - The third stage comes after all the food quantities have been transformed into grams.
  - The fourth stage consists of a final check of the aggregate distributions of dietary energy and monetary value once expressed in per-capita terms



Section 2 – Step by step process

- 11 step process
  - step 1: gathering input and auxiliary data
  - step 2: data cleaning: DOS- editing
  - step 3: adjusting and merging data files
  - step 4: cleaning data: food items and unit of measurement level
  - step 5: imputing monetary values
  - step 6: converting food quantities into grams
  - step 7: editing after converting to grams
  - step 8: calculating dietary energy
  - step 9: imputing dietary energy for remaining food items
  - step 10: aggregating and macro editing
  - step 11: preparing documentation and sharing data



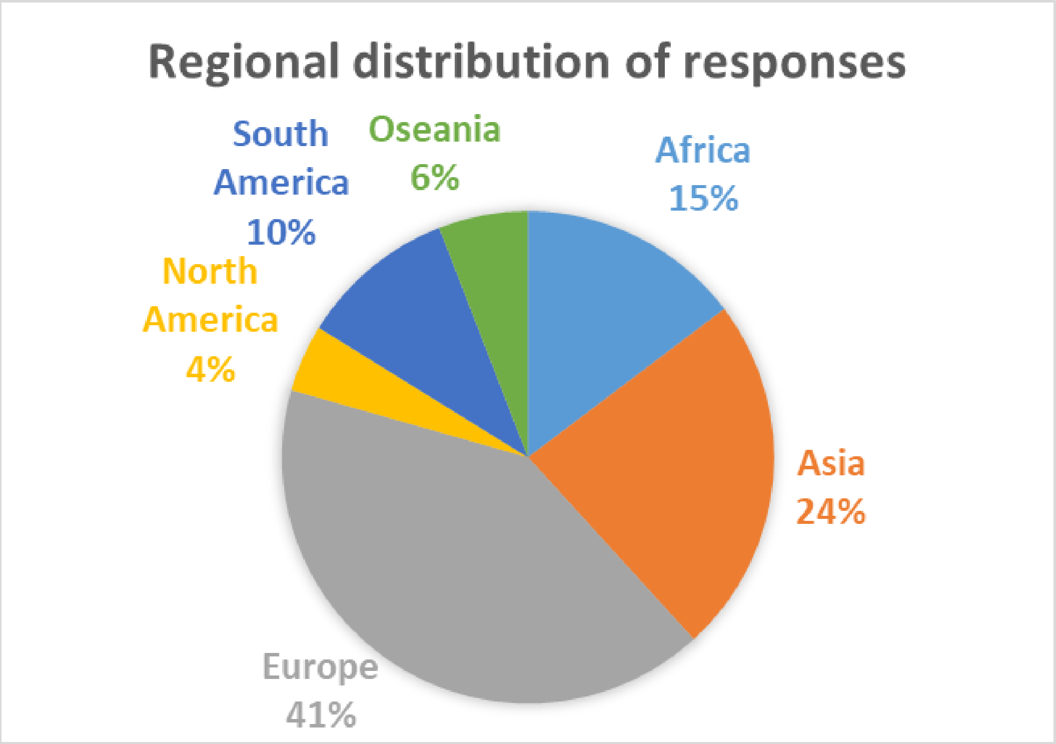
Results of the country consultation

Country consultation – October 2023

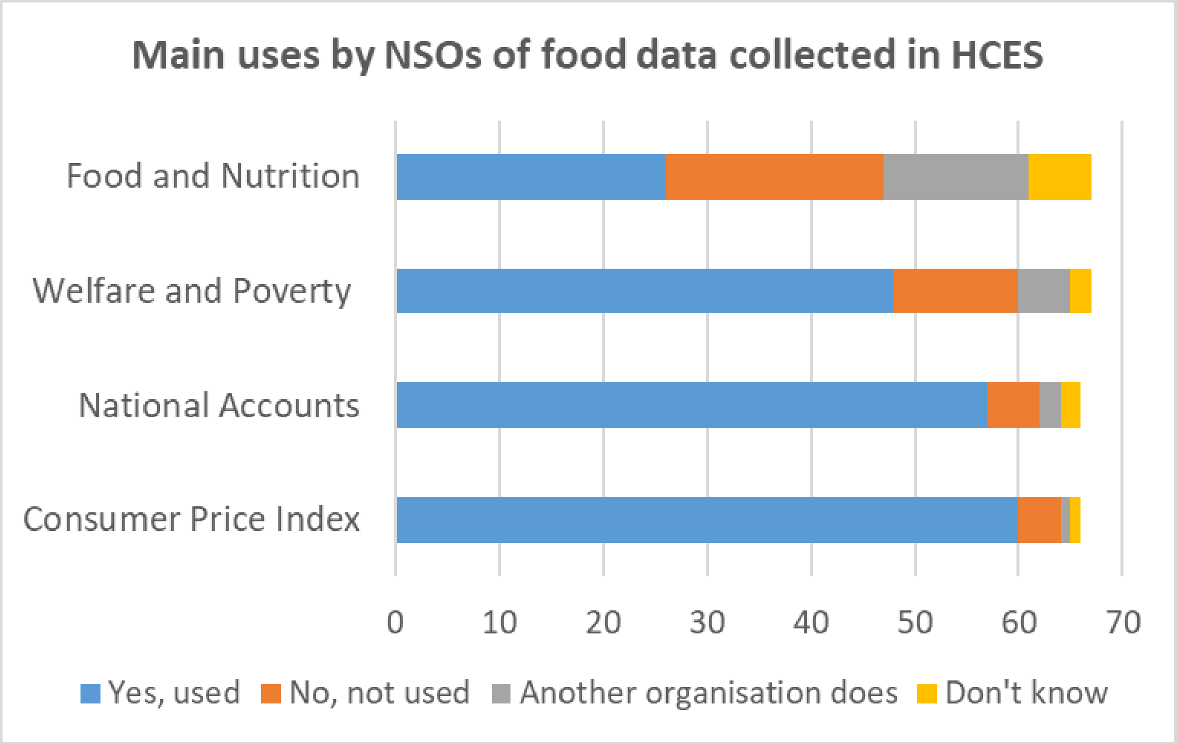
- A revised version of the Guidelines including comments received from UNCEAG experts was circulated to around 100 National Statistical Offices from low to high income countries
- Out of the 24 NSOs who responded, 16 acknowledged the Guidelines and provided no substantive comment
- The comments received during the global consultation were incorporated in the document when relevant. A response to each comment received was provided in document "The review process, comments received and responses on the guidelines for processing food consumption data from Household Consumption and expenditure surveys"
- NSOs were also requested to fill in an online questionnaire to collect additional information on the information collected in their survey and on the potential usefulness of the guidelines
- 70 NSOs filled the questionnaire (2 countries did not allow to share results so only 68 valid answers)

Survey results

68 countries responded –
Mainly from Europe and Asia

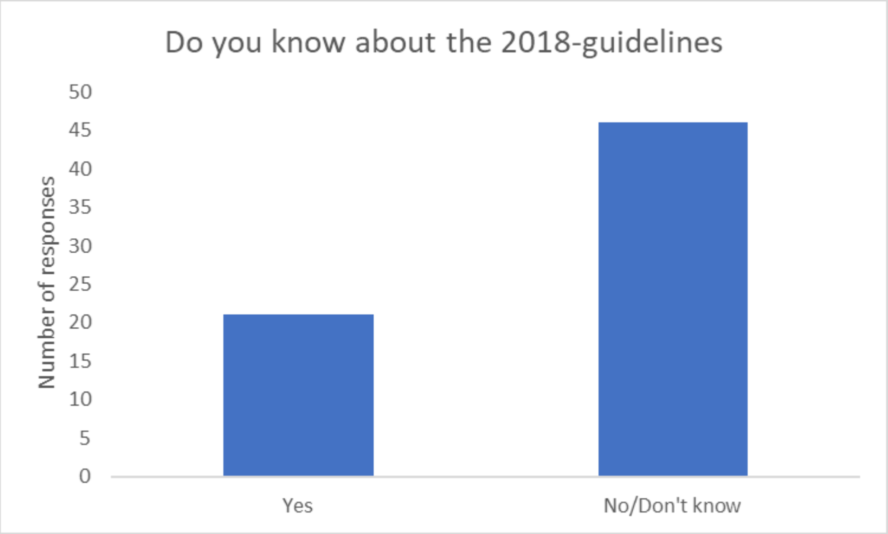


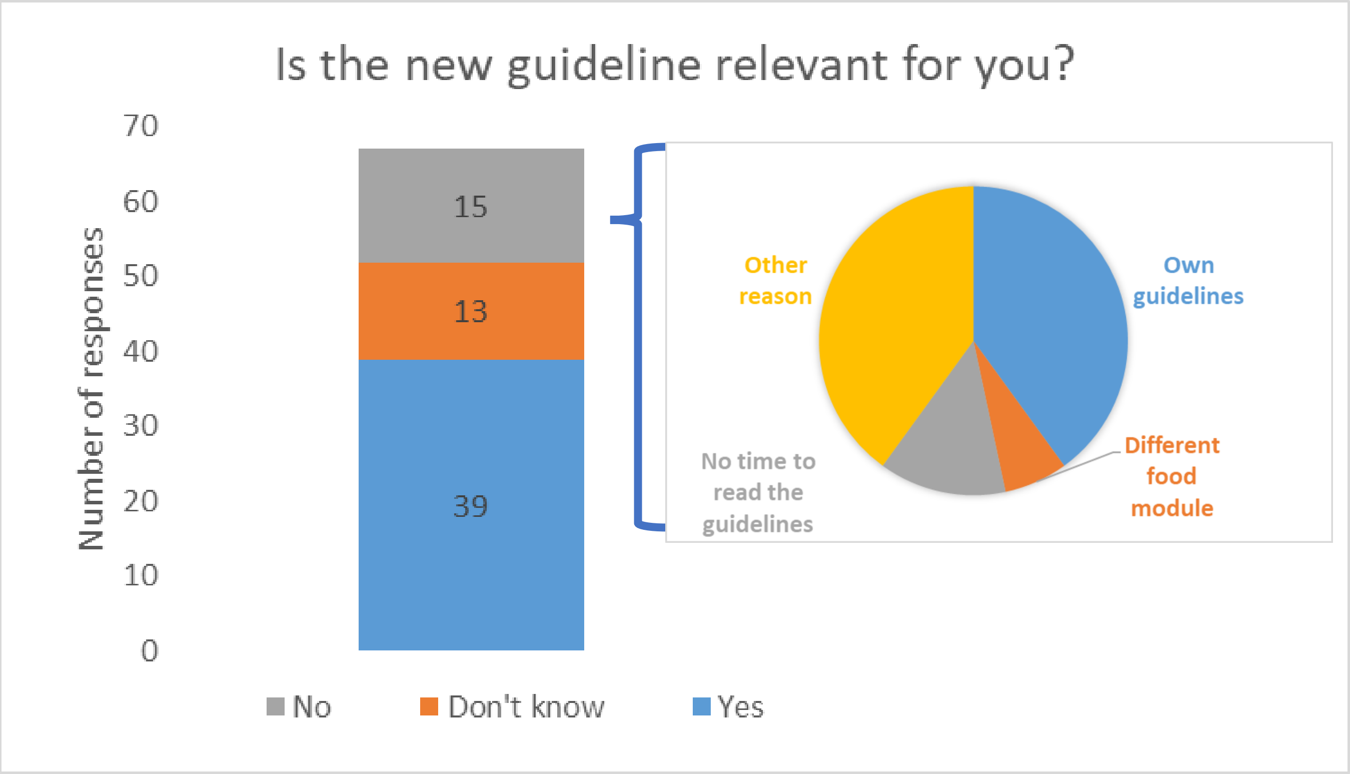
Most of NSOs used food data collected in their HCES to inform CPI, National accounts and poverty analysis



Impact of the guidelines

Two NSOs out of three are not aware of the guidelines on “Food data collection in Household Consumption and Expenditure Surveys” - endorsed by UNSC in 2018.





39 countries out of 68 believe the guidelines on food data processing will be relevant for their future work

Acknowledgment

- The drafting team included Nathalie Troubat (SPC); Elizabeth Foster (WB); Ana Moltedo (FAO ESS); Astrid Mathiassen (SSB); and Ellen Cathrine Kiøsterud (SSB)
- Experts who provided inputs and comments at several stages of the writing process included Alberto Zezza (WB), Kevin McGee (WB), Josefine Durazo (WB), Michael Sharp (University of Wollongong – former SPC staff), Andrea Borlizzi (SPC) and Edward Joy (London School of Hygiene & Tropical Medicine), as well as all participants in a workshop held in Rome in October 2022.
- The team also relied on inputs from other contributors, including Carlo Cafiero (FAO ESS); Abdul Sattar (FAO ESS), Haoyi Chen (UNSC); Owen Siyoto (COMESA); and Bridget Holmes (FAO ESN).
- All NSOs that provided comments about the guidelines during the global consultation.
- Assistance from Gaelle le Gall (SPC), Thorsdalen Bjørn (SSB), Tina Matilla (SPC) and Ann Horwarth (SPC) for editing
- The guidelines were funded by: Norwegian Agency for Development Cooperation (Norad), through its Agriculture for Development programme; the World Bank, through the Statistical Innovation and Capacity Building in the Pacific Islands (PACTSTAT) project; and the Australian Centre for International Agricultural Research, through its Pacific food system project (project number FIS2018/155), for direct financial contributions.

Thank you

https://www.fao.org/about/ce-on-food-security-agricultural-rural-statistics/en/

SESSION 2

National Quality Assurance Frameworks for key food and agricultural data – available tools

Kadarmanto Kadarmanto (BPS – Indonesia)

Marcello D'Orazio (FAO)

Members of the Task Team

Task Team Members	
Name	**Country/Institution**
Mr. Kadarmanto (**Chairperson**)	Indonesia, BPS-Statistics Indonesia
Mr. Sanjay Kumar Pant	India, Ministry of Statistics and Programme Implementation (MOSPI)
Ms. Susana Patricia Pérez Cadena Mr. José Luis Hernández Rodríguez Mr. Alan Martinez Hernández	Mexico, Instituto Nacional de Estadística y Geografía (INEGI)
Ms. Shadia Abu-Alzain	Palestine, Palestinian Central Bureau of Statistics (PCBS)
Mr. Babacar Ndir	Senegal, l'Agence Nationale de la Statistique et de la Démographie (ANSD)
Mr. Itani Magwaba	South Africa, Statistics South Africa (Stats-SA)
Mr. Rolando Ocampo Mr. Andres Gutierrez Rojas	UN-ECLAC
Mr. Pietro Gennari	FAO

Background and objectives

Objective:

Develop a framework for improving the quality of food and agriculture statistics

Background & references:

- UN NQAF Manual (2019) & self-assessment checklist

  Generic with four levels: (A) managing the statistical system, (B) managing the institutional environment (C), managing statistical processes, (D) managing statistical outputs

- IMF DQAF (2012) & data module of Reports on the Observance of Standards and Codes (ROSCs)

  Specific for economic statistics, three levels: the statistical institution, the statistical processes, the statistical product

Proposed approach

Instead of having explicit dataset-specific frameworks, as in IMF DQAF, we decided to develop tailored self-assessment checklists:

- crops and livestock statistics
- statistics on producers' prices of agriculture commodities
- statistics on land used for agriculture purposes

Covering three levels:

B) the statistical institution (partially, only resources)

C) the statistical processes (main phases of GSBPM)

D) the statistical outputs (Relevance; Accuracy & reliability; Timeliness & punctuality; Accessibility & clarity; Coherence & comparability)

The assessment with the checklist

Assessment-type questions on the **implementation of guidelines/best practices** (international standards; FAO manuals, etc.):

- Full implementation => score 1
- Partial implementation => score 0.5
- Not implemented => score 0
- *NA (Not Applicable) => no score assigned*

Elementary scores are aggregated (mapping provided) so as to align with IMF DQAF Reports on the Observance of Standards and Codes (ROSC) that adopts a **four-point rating scale**:

“**O**” = Practice Observed (0.8 < av_score <= 1)
“**LO**” = Practice Largely Observed (0.5 < av_score <= 0.8)
“**LNO**” = Practice Largely Not Observed (0.2 < av_score <= 0.5)
“**NO**” = Practice Not Observed (av_score <= 0.2)
NA = Not Applicable

The final report

Level	Item (from above, as relevant)	Outcome
Institutional framework	Assuring Adequacy of resources in producing the considered statistical outputs	
Statistical Process for producing the investigated statistical outputs	Design	
	Data collection	
	Data treatment	
	Data processing	
Quality of the investigated statistical outputs	Relevance	
	Accuracy and Reliability	
	Timeliness and Punctuality	
	Accessibility and Clarity (and metadata)	
	Comparability and Coherence	

Each applicable row has to be filled in with one of the outcomes ("O", "LO", "LNO", "NO") of the assessment

The assessment with the checklist: tools

1. The checklist (*MS Word doc*)

1. Scoring Mechanism (*MS Word doc*):
    1. Questions and associated scoring
    2. Mapping between questions and levels/items of the reporting template
    3. Template of the final report

2. Scores aggregation (*MS Excel spreadsheet*)

    From single scores to aggregated scores, needed for the report

The assessment with the checklist: how-to

1. Compile the checklist
2. Assign the score to each answered question, according to the scoring mechanism
3. Put the score into the MS Excel spreadsheet; it will provide automatically the final summary score for each applicable entry in the final report
4. Compile the final report:
    1. For outcomes "LNO" and "NO" the Assessor has to describe the actions to be undertaken to improve the relevant statistics (possibly indicating the priorities in implementation).

The assessment with the checklist: when

The national statistical agency has **already in place a regular statistical production process** (survey, etc.) in the considered sub-domain:

1. Assessment of an **ongoing process** (survey, etc.) to discover major strengths and weaknesses (that need improvement actions)
2. Assessment before the **re-design of the process** (survey, etc.) to understand where to concentrate the efforts to improve the quality of the

The national statistical agency does **NOT have in place a regular statistical production process** (survey, etc.) in the considered sub-domain:

3. Assessment before setting up a **new process** to understand which are the best practices, manual/guidelines, classifications, etc. to be implemented

SESSION 3

Improving the Use of Earth Observation (EO) data for agriculture statistics – Recent achievements and proposed activities for the programme of work 2024-27

Eduardo Vázquez Andrade (INEGI-Mexico)

Lorenzo De Simone (FAO)

Talip Kilic (World Bank)

Objective of the Joint Task Team

The Joint Task Team on the Earth Observation data for Agricultural Statistics created under the umbrella of both the UN-CEAG and the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD), supports countries through the provision of methods, tools and training on the use of EO data for estimating crop acreage and crop yield and producing thematic crop maps.

Recent Achievements

The joint task team shares experience and technical advice on the key components of earth observation analysis protocols.

In 2022-2023 the Task Team has continued its work on a series of use cases through collaborations with countries:

- In **Senegal,** An experimental protocol was tested to combine In situ data with EO, to reduce the coefficient of variation and to increase the accuracy in the final crop type map.

2018 Results

	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%

2021 Results

Expressed in number of pixels		Field survey						
		Non crop	Maize	Millet	Groundnut	UA	Contaminations (%)	Omissions (%)
Crop type map	Non crop	2169	84	95	58	90.15	20.49	9.85
	Maize	0	596	17	11	95.51	17.34	4.49
	Millet	378	19	2742	14	86.96	6.10	13.04
	Groundnut	181	22	66	3210	92.27	2.52	7.73
	PA	79.5	82.7	93.9	97.5			

Overall Accuracy: 90.2%

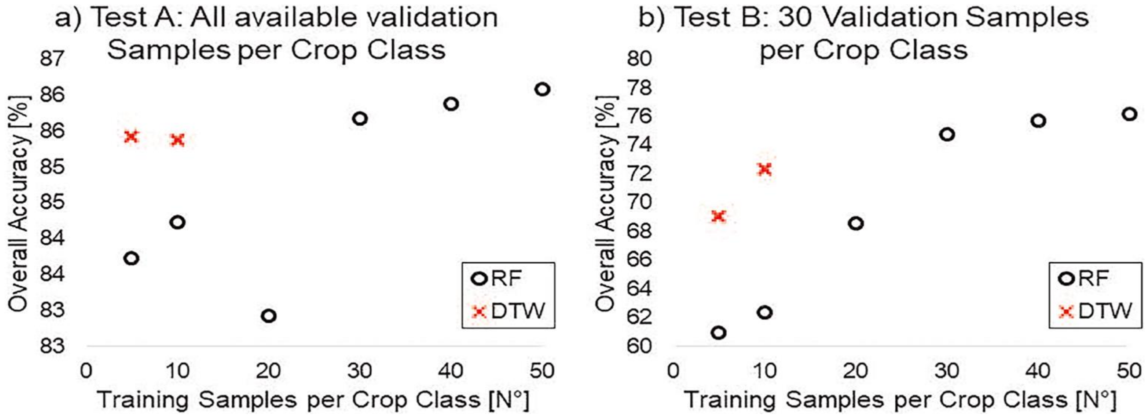
Recent Achievements

- In **Mali** a work that is still in its experimental phase is on going to produce crop type maps and to extract acreage. The results will allow to formulate recommendations to the NSO.

- In **Rwanda** a Pilot project was implemented to produce a National wall to wall map of crop fileds boundaries. The results are promising.

Recent Achievements

- In **Uzbekistan**, FAO carried out a comparative analysis of the performance of Random Forest Vs Dynamic Type Warping (DTW) algorithms in cases of in situ data scarcity.



Recent Achievements

- In **Ecuador** and in **Cameroon** a Pilot project was conducted, to integrate EO data with process based crop growth modeling SALUS (System Approach to Land Use Sustainability) for the forecasting of crop yield for Rice and maize in Ecuador.

Recent Achievements

- In **Mexico**, INEGI is using Landsat and Sentinel2 with Machine learning algorithms to obtain the cropland map of the whole country, this is an ongoing project.

Recent Achievements

In terms of outreach and capacity development efforts, and collaboration with Regional and Global Hubs in 2022-2023 the Task Team had:

- Participated in January and February 2023 in advisory meetings of the Brazilian Regional Hub to strengthen the organization and functionality of the center.
- Participated in the Webinar on EO for Agricultural statistics with the National Bureau of Statistics (NBS) of China in March 2023.
- And in the 9th International Conference on Agricultural Statistics (ICAS) in May 2023, sharing main achievements in training, data sharing and applications of EO in different countries.
- Participated in the 4th International Seminar on EO for Agricultural statistics with the National Bureau of Statistics (NBS) of China in November 2023.
- Members of the joint task team will also participate as members in the International Expert Committee on remote sensing for agriculture statistics created to advise China's Global Big Data Hub and build synergies on relevant activities.
- Participated in International Webinar on EO for Agricultural statistics (EO-STAT) with Countries of UN- Big Data Regional Hub for Africa in December 2023.


UN Big Data Regional Hub for Africa

Webinar Report: International Webinar on Earth Observations for Agricultural Statistics (EO-STAT)

Prepared by

UN Big Data and Data Science Regional Hub for Africa

https://ecastats.uneca.org/regionalhub

January 2024

Programme of Work 2024-27

Programme of Work 2024-27 Joint Task Team on Use of EO for Agricultural Statistics (UNCEAG-UNCEBD)		
Sub Task Team	***ID***	***Activity***
Methods	**1**	**Enhancement of classification algorythms crop crop type mapping**
	1.1	*Artificial intellingece*
	1.2	*Pattern matching*
	1.3	*Yield forecasting*
	2	**Integration of Big Data Sources**
	2.1	*Combined use of Optical and Radar data*
	2.2	*Crowdsorcing*
Data collection	**3**	**Improving quality of maps**
	3.1	*Best practices for ealuating map accuracy and area estimation using EO data*
	3.2	*Use of very high resolution images/drones*
	4	**New Topics**
	4.1	*Early Warning*
	4.3	*Biodiversity*
	5	**In situ data**
	5.1	*Optimized field survey design and protocols*
	5.2	*Crop pheno-spectral data (signatures) as additonal information for processing*
Data Sharing	**6**	**Governance**
	6.1	*Confidentiality*
	6.2	*Platforms*
	6.3	*Standards*
	7	***Global and Regional Hubs collaboration***
	7.1	*Global Hub China*
	7.2	*Regional Hub Indonesia*
	7.3	*Regional Hub Rwanda*
	7.4	*Regional Hub Brazil*

Using EO data to optimally design a field survey and support the field protocol – EOSTAT Zimbabwe

Establishing the area frame and strata

Created a regular grid of 98237 blocks each measuring of 2 x 2 Km per side (4sq Km)

- **Definition of strata** – Agricultural Intensity

Use of the ESA World Cover 10m 2020.

Extraction of the cropland mask within Zimbabwe

Determining the sample size

- Cochran's formula was used to compute the total number of samples

$$n = \frac{(\sum W_i S_i)^2}{\left[S(\hat{O})\right]^2 + (1/N)\sum W_i S_i^2} \approx \left(\frac{\sum W_i S_i}{S(\hat{O})}\right)^2$$

- where N= number of population units in the area of interest, S(Ộ) is the standard error of the estimated overall accuracy that we would like to achieve, Wi is the mapped proportion of area of class i, Si is the standard deviation of stratum i and Ui represents the User accuracy of class i
- Because N is large (over 2397 million pixels in this Zimbabwe), the second term in the denominator of Eq. (2) can be ignored. We specify a target standard error for overall accuracy of 0.01. From past crop type mapping experience in similar agro ecological zones, we know that errors of commission are relatively common for the mixed crops. Consequently, we conjecture a user's accuracies for stratum (≥ 0.70) based on previous crop mapping exercises.

The resulting sample size for Zimbabwe is n = 1836, where samples correspond to Secondary Sampling Units (PSU).

Table 2 provides the information computed for each strata following Cochran's formula.

Stratum	Cropping Intensity	Number of blocks ($4km^2$ each)	Proportion of total area	Expected User Accuracy	Standard Deviation	Total Number of Samples defined as PSU
1	2 - 5.5	11291	0.188393706	0.7	0.458258	1836.424687
2	5.5 - 16.2	22644	0.377821901	0.7	0.458258	
3	16.2 - 29.7	14742	0.245974672	0.7	0.458258	
4	29.7 - 48.5	8390	0.139989655	0.9	0.3	
5	48.5 - 99.9	2866	0.047820066	0.9	0.3	

Allocation of samples per class

Following recommendations from Stehman, 2012, we allocated equal number of SSU to each stratum.

In the first stage of sample allocation, **25** blocks were randomly selected from each stratum as Primary Sampling Units (PSU).

In the second stage of sample allocation, 15 Secondary Sampling Units (SSU) were randomly allocated in each PSU resulting in a total of 1875.

Field protocol

District	PSU ID	PSU per district	SSUs per district
Team 1		**30**	**450**
Beitbridge	PSU 29	1	15
Bikita	PSU 10, PSU 11 & PSU 28	3	45
Chipinge	PSU 22 to PSU 26	5	75
Chiredzi	PSU 14 to PSU 17; PSU 19 to PSU 21 & PSU 27	8	120
Gwanda	PSU 3	1	15
Insiza	PSU 4 & PSU 5	2	30
Masvingo	PSU 7 to PSU 9	3	45
Matobo	PSU 1 to PSU 2	2	30
Mwenezi	PSU 18 to PSU 30	2	30
Zaka	PSU 12 to PSU 13	2	30
Zvishavane	PSU 6	1	15

Field apps

Path Finder to find the optimal route to the PSU assigned to each team

Survey 123 to collect the data in the field

Locus app for windshield data collection while travelling by car from one PSU to another PSU

How Can Large-Scale Surveys Meet Training Data Requirements for Satellite-Based Crop Type Mapping: Cross-Country Evidence from Sub-Saharan Africa

***Shruti Jain*[1,3], Talip Kilic[2], Abera Muhamed[3], Siobhan Murray[2], Vivek Sakhrani[3]**

1. University of Oxford
2. Development Data Group, World Bank Group
3. Atlas AI PBC

Background

- Role of agriculture in rural livelihoods
  - Byerlee et al. 2007, Davis, et al. 2017
- Need for accurate, crop-specific measures of area under cultivation, production and yields – not only at the national-level but with enhanced within-country disaggregation
- Surge in availability of high-resolution satellite imagery
  - Still need data to train and validate the underlying models

Key takeaways from the literature

- Training data affect the quality and spatial resolution of satellite-based estimates (Lobell et al. 2019, 2020)
- Existing research has largely been at sub-national levels, with heterogeneity in the type of and approach to training data collection as part of surveys
- Large-scale surveys *can* address training data needs of earth observation applications on crop area mapping and crop yield estimation in lower-income countries, but…
  - No clear recommendations on survey methods and fieldwork protocols to generate the right training data

Research objectives

- Address operationally-relevant research questions related to crop type mapping in smallholder production systems, leveraging existing georeferenced national survey data, Sentinel-2 + ancillary geovariables + ML
  - How much **training data** do we need to achieve acceptable performance of a crop classification algorithm?
  - How does the **approach to georeferencing** plot locations in surveys impact algorithmic performance?
  - How do the type of satellite data, earth observation variables, choice of ML model, and exclusion of plots under specific area thresholds affect algorithmic performance?

- Build on our related research (Azzari et al., 2021) and expand the scope in terms of
  - **Countries** – Mali, in addition to Malawi and Ethiopia
  - **Crops** – Not only maize but also sorghum, millet, rice, teff, wheat, and barley
  - **Experimentation** (i.e. relative impacts of choice of ML model and inclusion of ancillary geovars)

- Create 10-m resolution crop type maps for Mali, Malawi, and Ethiopia and disseminate via the World Bank Development Data Hub

Survey data

- Georeferenced plot-level survey data stem from nationally-representative household surveys that were implemented by the Malawi National Statistical Office, the Ethiopian Statistical Service of Ethiopia and the Ministry of Agriculture of Mali under the World Bank LSMS-ISA Initiative

  - **Mali** Enquête Agricole de Conjoncture Intégrée aux Conditions de Vie des Ménages (EACI) – Rounds 2017 and 2018
    - Cross-sectional samples
    - Reference season: 2017 or 2018, depending on the year
    - Plot-level georeferenced information: Single plot corner point + plot boundaries
  - **Malawi** Integrated Household Panel Survey (IHPS) 2019
    - Longitudinal sample, dating back to 2010
    - Reference season: 2018/19
    - Plot-level georeferenced information: Single plot corner point + plot boundaries
  - **Malawi** Fifth Integrated Household Survey (IHS5) 2019/20
    - Cross-sectional sample
    - Reference season: 2017/18 or 2018/19
    - Plot-level georeferenced information: Single plot corner point + plot boundaries
  - **Ethiopia** Socioeconomic Survey (ESS) 2018/19
    - Baseline for a new longitudinal sample
    - Reference season: 2018 *meher* season
    - Plot-level georeferenced information: Single plot corner point

Input features from EO sources

- Satellite observations from Sentinel 2 (optical features) and Sentinel 1 (SAR features), both at 10 m resolution.
- We also added topography and seasonal weather metrics.

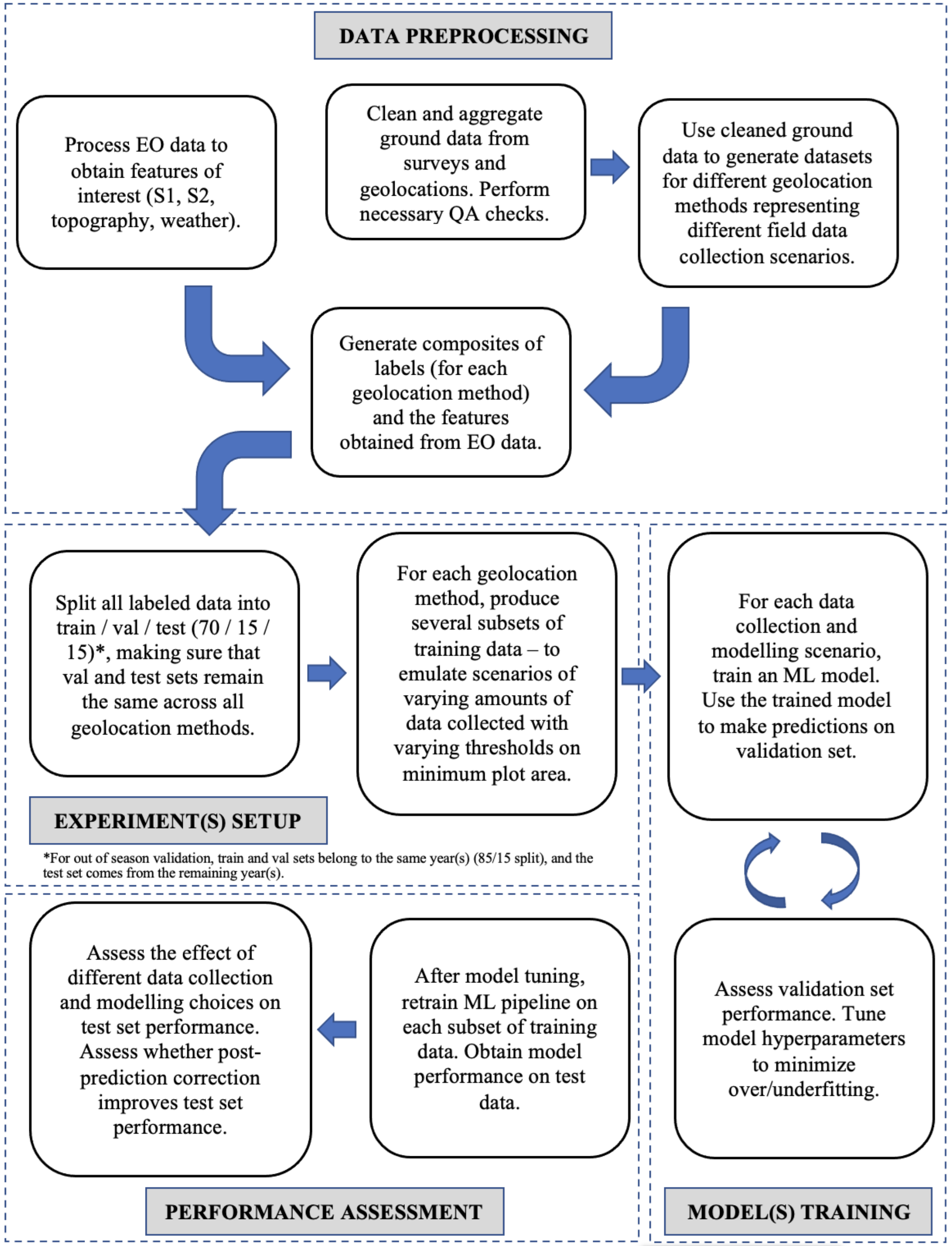
Feature	Explanation	Data Source
Elevation	Obtained using GEE's inbuilt *terrain* algorithm that uses an elevation raster to generate slope and aspect bands	Shuttle Radar Topography Mission (30 meter resolution)
Slope		
Aspect (direction of slope)		
Average temperature	Mean daily temperature during growing season	MODIS LST corrected using CHIRTS
Total precipitation	Total precipitation during growing season	CHIRPS

Band / Index	Name	Central wavelength / Index formula	Satellite
VV	Vertically polarized backscatter	5.5465763 cm	Sentinel-1
VH	Horizontally polarized backscatter	5.5465763 cm	Sentinel-1
RATIO	Ratio		Sentinel-1
DIFF	Difference		Sentinel-1
AEROS	Aerosol	443 nm	Sentinel-2
BLUE	Blue	490 nm	Sentinel-2
GREEN	Green	560 nm	Sentinel-2
RED	Red	665 nm	Sentinel-2
RDED1	Red Edge 1	705 nm	Sentinel-2
RDED2	Red Edge 2	740 nm	Sentinel-2
RDED3	Red Edge 3	783 nm	Sentinel-2
NIR	Near Infrared	842 nm	Sentinel-2
RDED4	Red Edge 4	865 nm	Sentinel-2
VAPOR	Water Vapor	940 nm	Sentinel-2
CIRRU	Cirrus	1375 nm	Sentinel-2
SWIR1	Short-wave Infrared 1	1610 nm	Sentinel-2
SWIR2	Short-wave Infrared 2	2190 nm	Sentinel-2

Modelling scenarios

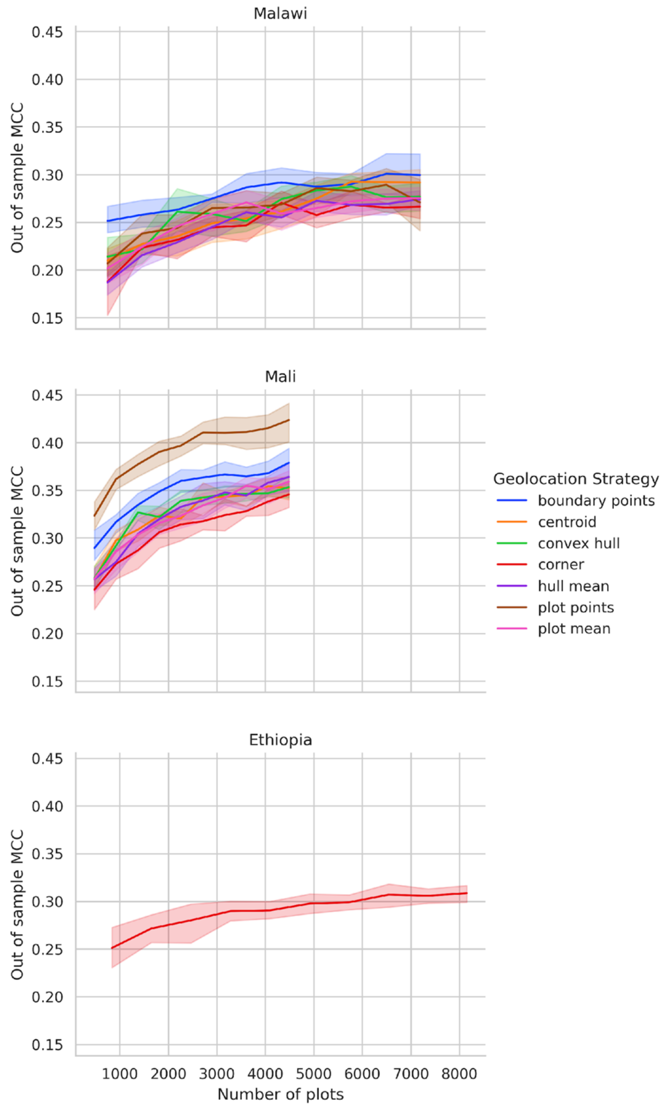
Crop type experiment	**Malawi**	**Mali**	**Ethiopia**
CT Experiment 1: Field polygons vs. points	corner, centroid, boundary points, convex hull points, convex hull mean, plot points, and plot mean	corner, centroid, boundary points, convex hull points, convex hull mean, plot points, and plot mean	n/a (corner only)
CT Experiment 2: Reducing field sample size	10 data subsets - 10% to 100% subsets of training data, at an increment of 10% points	10 data subsets - 10% to 100% subsets of training data, at an increment of 10% points	10 data subsets - 10% to 100% subsets of training data, at an increment of 10% points
CT Experiment 3: Applying area thresholds on training data	0, 0.05, 0.1, 0.15, and 0.2 ha	0, 0.1, 0.2, and 0.3 ha	n/a
CT Experiment 4: Sentinel-1 (SAR) and Sentinel-2 (Optical)	Optical only, SAR only, both	Optical only, SAR only, both	Optical only, SAR only, both
CT Experiment 5: Scope of geospatial variables	Including and excluding weather variables (temperature and precipitation)	Including and excluding weather variables (temperature and precipitation)	Including and excluding weather variables (temperature and precipitation)
CT Experiment 6: Testing different ML models.	RF, 1-D CNN	RF, 1-D CNN	RF, 1-D CNN
Total scenarios	**RF:** 7 (CT1) x 10 (CT2) x 5 (CT3) x 3 (CT4) x 2 (CT6) **CNN (CT5):** 7 (CT1) x 10 (CT2)	**RF:** 7 (CT1) x 10 (CT2) x 4 (CT3) x 3 (CT4) x 2 (CT6) **CNN (CT5):** 7 (CT1) x 10 (CT2)	**RF:** 10 (CT2) x 3 (CT4) x 2 (CT6) **CNN (CT5):** 10 (CT2)

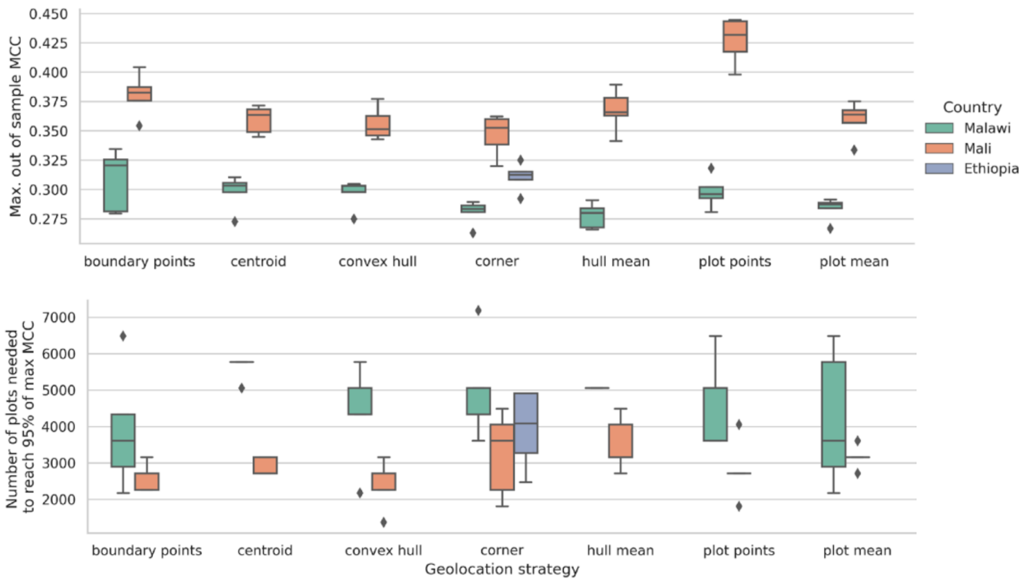
Modelling Workflow



Results on the role of geolocation methods and sample size

- Marginal improvements in MCC rapidly diminish after ~2,000 plots are available for training
- Models trained with "Plot Points" consistently outperformed all in Mali, whereas in Malawi, the best performing model was trained with "Convex Hull", followed by "Plot Points."
- 95% of maximum MCC was attained with 4,000-5,000 plots in Malawi (~60% of total sample); 2,500-3,000 plots in Mali (~65% of total sample); and with ~4,000 plots in Ethiopia (~50% of total sample).
- The "centroid" method does slightly better than the "corner point" method in Malawi and Mali.



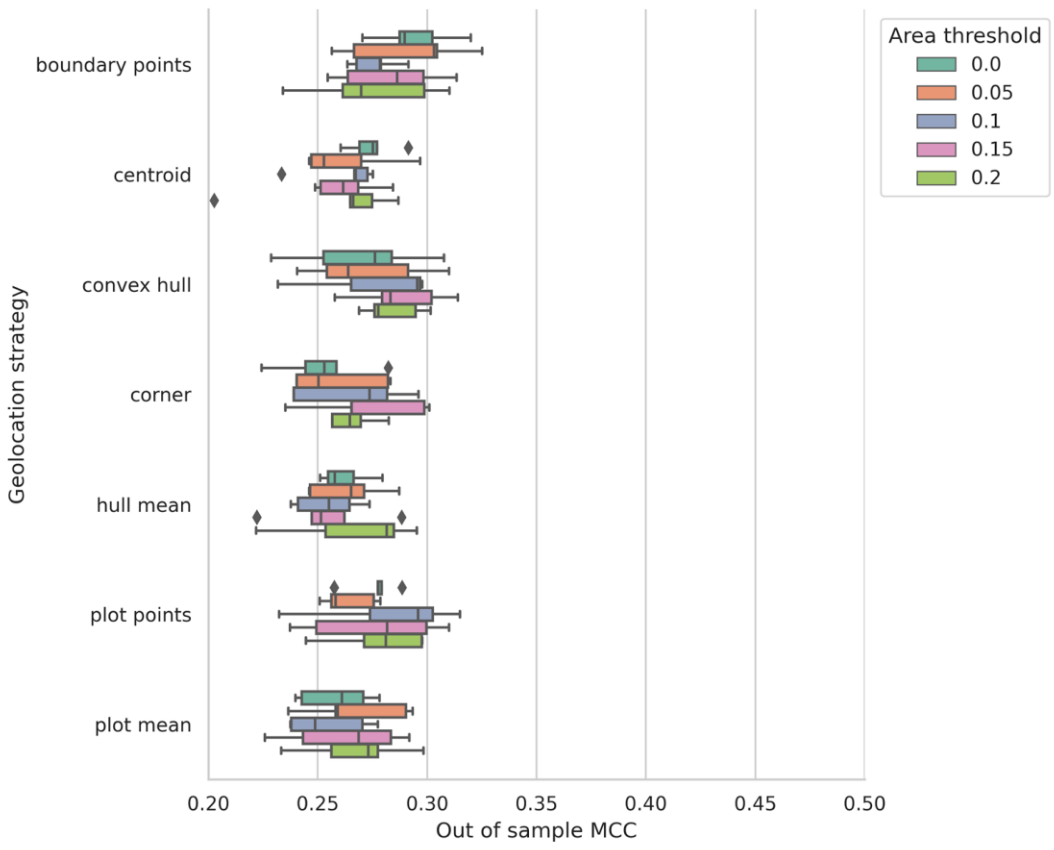


$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

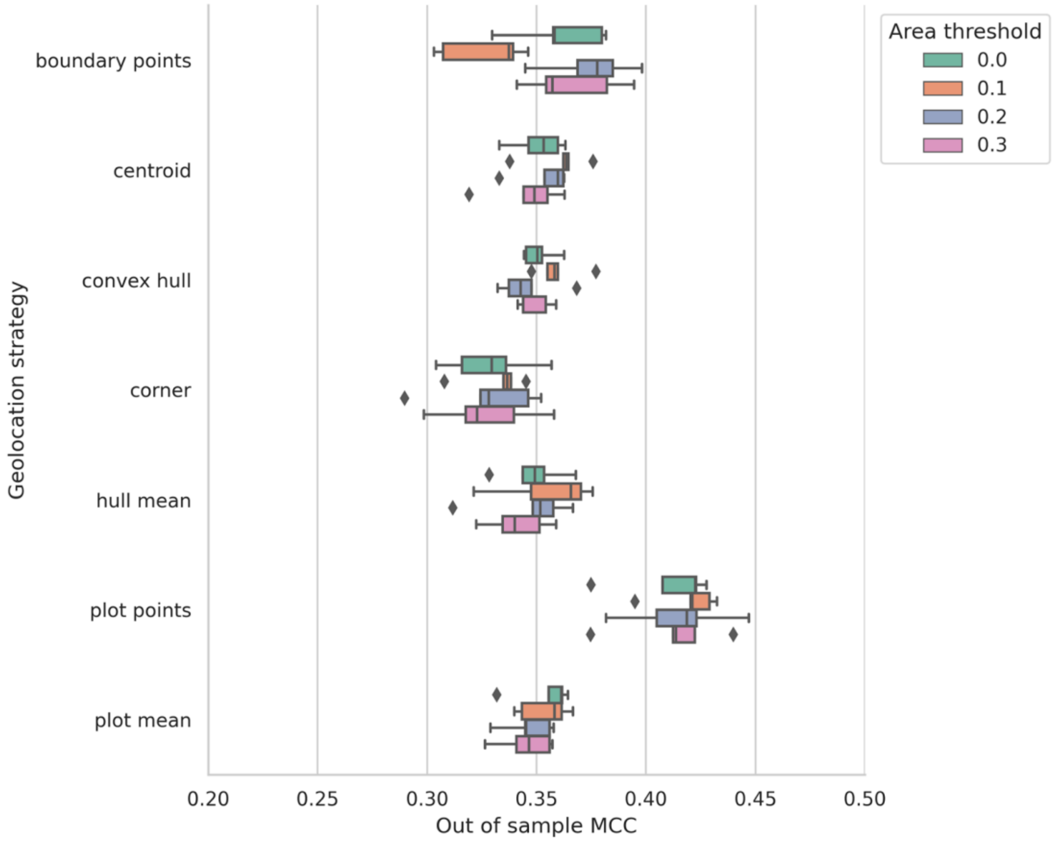
Results on the role of plot size

- Limiting training data by excluding plots under specific area thresholds has little to no impact on model performance.
- Some exceptions in Malawi, especially for the threshold of 0.05 ha, but insight still holds true for best performing models.

Malawi

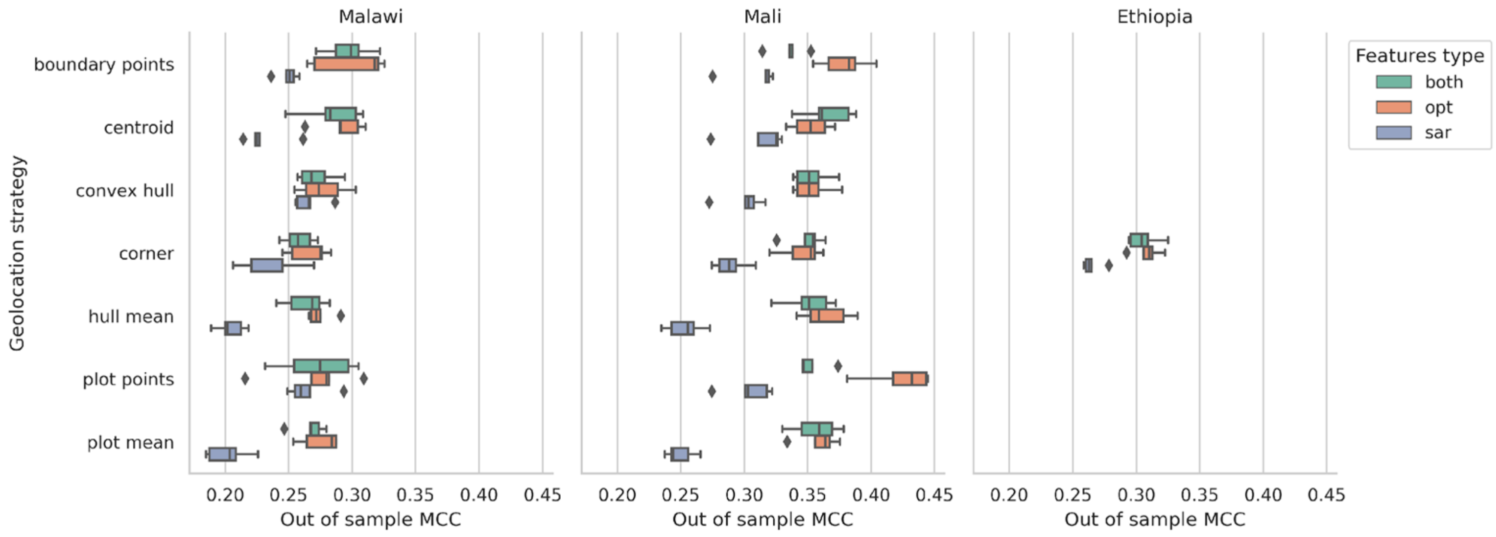


Mali



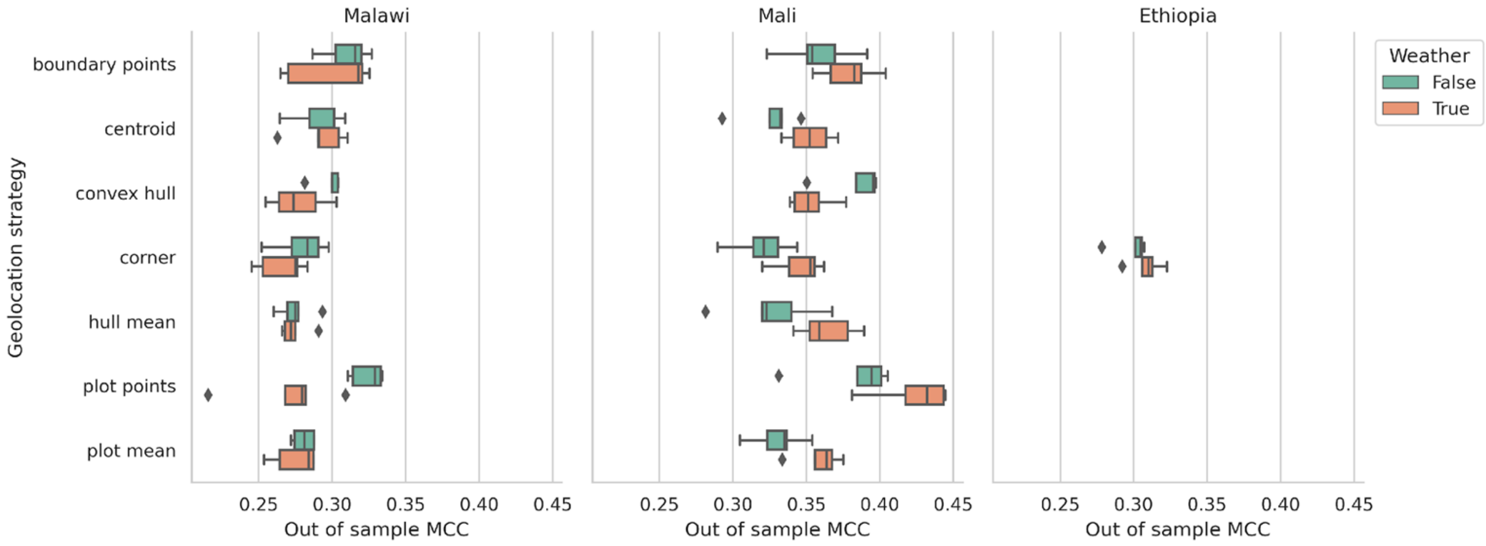
Results on the role of satellite data type

- SAR alone generally lower performance.
- Optical alone highest.
- Little to no gain in predicting power when combined.
- Exception: “centroid” approach in Mali



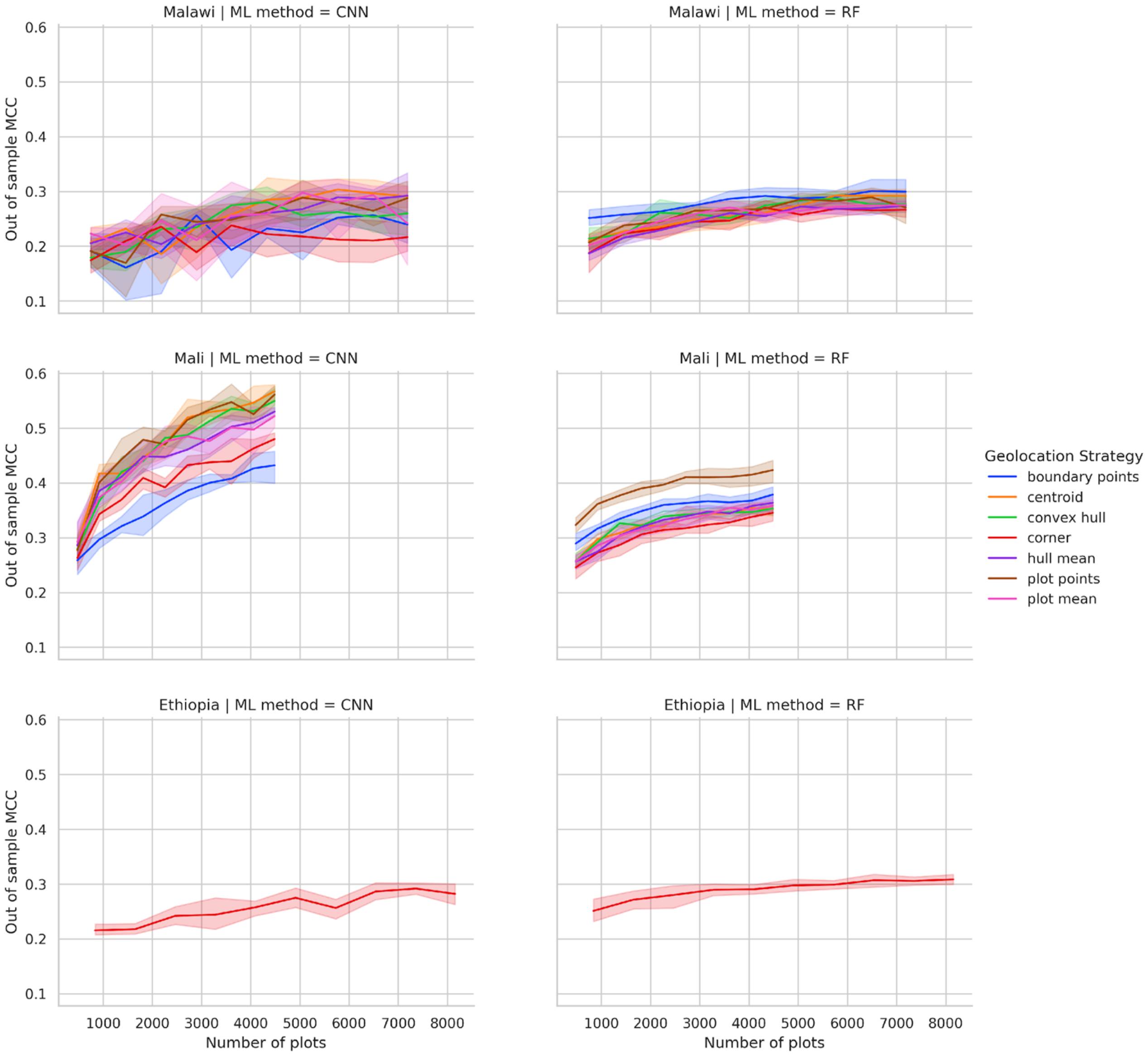
Results on the role of ancillary geospatial variables

- Inclusion of weather generally improves performance.
- Exception: “convex hull” approach in Mali, and “plot points” approach in Malawi.



Results on the role of ML modelling

- CNNs did not show substantial gains in Malawi and Ethiopia.
- However, gains in Mali were substantial, likely due to full boundaries + bigger plots.
- In Mali, higher learning rates were observed with increasing sample sizes, when using CNNs as compared to Random Forests.
- However, with smaller amounts of data (<1,000 fields), in the case of all three countries, Random Forests outperformed CNNs, likely due to CNNs overfitting on small amounts of training data



Small differences, large consequences

Small differences in model performance may lead to large differences in estimated areas. There is value in achieving small performance gains with better training data.

Malawi maize area as obtained by seven different classification models, and area misclassified as maize/ non-maize under each classification model as compared to “Convex Hull” (the best performing model)

Classification Model	Out of Sample MCC	Total Maize Area - 2018/19 Rainy Season (million ha)	Difference in Out of Sample MCC as compared to "Convex Hull"	Total Area with Disagreement as Compared to “Convex Hull” (million ha)
Boundary points	0.31	1.60	-0.01	1.29
Centroid	0.29	1.76	-0.03	1.11
Convex hull	0.32	2.73		
Corner	0.29	1.50	-0.03	1.69
Hull mean	0.27	1.43	-0.05	1.40
Plot points	0.31	2.53	-0.01	0.86
Plot mean	0.28	1.44	-0.04	1.40
Mean across models	0.30	1.86		

Mali sorghum area as obtained by seven different classification models, and area misclassified as sorghum/ non-sorghum under each classification model as compared to “Plot Points” (the best performing model)

Classification Model	Out of Sample MCC	Total Sorghum Area - 2018 Rainy Season (million ha)	Difference in Out of Sample MCC as compared to "Plot Points"	Total Area with Disagreement as Compared to “Plot Points” (million ha)
Boundary points	0.33	2.87	-0.06	2.08
Centroid	0.30	1.33	-0.09	2.13
Convex hull	0.36	2.87	-0.03	1.29
Corner	0.26	1.45	-0.13	2.19
Hull mean	0.30	1.34	-0.09	2.17
Plot points	0.39	2.95		
Plot mean	0.31	1.39	-0.08	2.15
Mean across models	0.32	2.03		

Conclusions

- Collecting a **complete plot boundary** is preferable to competing approaches to georeferencing plot locations in large-scale household surveys. This is particularly true if collection capacity is limited to fewer locations.
- Seemingly-small erosion in classification performance under less preferable approaches to georeferencing plot locations **results in large differences in total crop area estimated** - by as much as 50%.
- **Georeferencing the complete set of plot corners is a second-best strategy**, can approximate full plot boundaries and can in turn train models with comparable performance.
- Classification performance peaks with **~50-65% of the training data** under preferred approaches to georeferencing plot locations.
- **If only a single GPS point** can be collected, that location should be **near the plot centroid** rather than at the plot corner.
- **No plot observations should be excluded** from model training based on a minimum plot area threshold.
- **Optical features alone** can provide sufficient signal to maximize prediction quality.
- **Inclusion of weather** is generally beneficial to model performance.
- **CNNs** can provide performance gains over Random forest models, especially at larger sample sizes (>1,000 fields) and in systems with “larger” fields

Questions and contacts

- DeSimone, Lorenzo (OCS)
  - Lorenzo.DeSimone@fao.org

- Talip Kilic (World Bank)
  - tkilic@worldbank.org

- Eduardo Vázquez, INEGI(Mexico)
  - eduardo.vazquez@inegi.org.mx

Thank you

https://www.fao.org/about/ce-on-food-security-agricultural-rural-statistics/en/

SESSION 4

Improving Food security and nutrition statistics: definition, minimum set of core data and guidance for prioritization of FSN data at national level

José Rosero Moncayo (FAO)

The CFS Process

- ✓ The Committee on Global Food Security (CFS) is the main inclusive international and intergovernmental platform to ensure food security and nutrition for all
- ✓ Aims to provided actionable data policy recommendations addressed to Governments, International Organizations, Civil Society, Private Sector, and Donors
- ✓ In 2019, Data was selected as a stream of work
- ✓ It was a remarkable exercise in which an international policy-level forum has reached consensus on concrete recommendations to strengthen data systems
- ✓ The process involved more that 100 hours of formal negotiations
- ✓ Ended with the endorsement of CFS recommendations on: **“Strengthening Collection and Use of FSN Data and Related Analysis Tools to Improve Decision-Making in Support of the Progressive Realization of the Right to Adequate Food in the context of National Food Security”**

The recommendations are divided in five areas addressing different challenges:

1. CREATE GREATER AWARENESS AND FSN DATA USE IN DECISION-MAKING

- Establish multisectoral and multistakeholder FSN mechanisms
- Promote dialogues and cooperation
- Develop guidelines
- Use existing data and promote interoperability

2. INCREASE AND OPTIMIZE INVESTMENT

- Increase, sustain, and coordinate investment on FSN data
- Elaborate national plans to define priorities, integrated in the NSDS
- Outline a minimum set of core FSN data
- Identify data gaps and needs

3. DEVELOP CAPACITIES

- Invest in building the capacities of statisticians, data experts and social scientists
- Modernize infrastructures
- Expand training opportunities
- Invest in innovation
- Reduce language barriers

4. COLLABORATION ON HARMONIZATION AND SHARING OF FSN DATA

- Promote the harmonization, coherence, and interoperability of FSN data
- Consider FSN statistics as a domain within the UNSC
- Treat FSN data as open as possible but as closed as necessary to serve the public good
- Increase collaboration on access and sharing of data

5. STRENGTHEN FSN DATA GOVERNANCE FRAMEWORKS

- Include FSN data within a broader national data governance
- Discuss the development of FSN data principles
- Private Sector is asked to share FSN data and analytics for the public good

Why is the connection of this recommendations to the Statistical Commission?.

- ✓Creation of a new data domain on food security and nutrition statistics under the aegis of the UNSC
  - ✓Forum for discussion of methodologies for measuring food security and nutrition and the underlying data needed.
  - ✓Forum to standardize methodologies
  - ✓Forum to promote the use of food security and nutrition statistics
  - ✓Will imply the report to the Commission of the state and progress of this domain
- ✓From operational point of view, it will imply the creation of a UN Committee of experts on food security and nutrition statistics, or the adoption of this responsibility by an already existing Committee (UN-CEAG).
  - ✓The work of this work will be the basis for reporting to the Commission

Why is the connection of this recommendations to the Statistical Commission?.

- ✓For the next two years, two lines of work are consider priority under the aegis of this Group:
  - ✓Definition of food security and nutrition data
  - ✓Development of guidelines which includes:
    - ✓ The definition of a minimum set of core data on FSN,
    - ✓ References to recommended methodologies, data sources

Outputs and way forward

Outputs:

- ❖ Working paper on a minimum set of core FSN data and indicators
- ❖ Finalized and validated definition and minimum set of core data and indicators
- ❖ Development of guidelines, methodological, metadata documents, and training material
- ❖ Countries outreach, advocacy, communication and training

Way Forward:

- ✓ Draft report on the tentative guidelines (FAO/WHO/UNICEF)
- ✓ Review of the proposed core set of FSN indicators (Committee of experts)
- ✓ Validation meeting of the proposed core set with FSN experts
- ✓ Finalization of guidelines and training material

Tentative timeframe

		2024			2025			
Activities	**Deliverables**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Methodological work	Working paper							
Expert group meeting	Meeting report							
Peer review	Peer review findings							
Draft guidelines	First draft							
Validation meeting	Meeting report							
Editorial review and layout	Edited guidelines							
Finalized guidelines	Final version							
Training material	Presentations/supporting material							
UNSC discussion	Endorsement							
Publication of guidelines	Document published							
Advocacy and training	Regional Workshops							

Thank you

https://www.fao.org/about/ce-on-food-security-agricultural-rural-statistics/en/

World Programme of the Census of Agriculture: Development of FAO Guidelines for the WCA 2030

Jairo Castaño (FAO)

Background

- FAO is the UN agency responsible for providing census methodological guidelines to countries through the World Programme for the Census of Agriculture (WCA) .
- The current round is the WCA 2020, which ends in 2025.
- FAO is preparing the new WCA 2030 guidelines, for the period 2026-2035.
- This work started with a review of lessons, consultation with key experts, member countries (in regional meetings), FAO technical divisions.
- The consultation guided on possible areas of revision and improvement for the preparation of the new WCA 2030.
- A Concept Note was prepared, followed by a first draft of the guidelines.

Findings on the implementation of WCA 2020 – 1/3

- **An online survey sent to all member countries in 2022 showed that:**

1. The classical census (50% of the countries) and combined census with admin registers (32%, mainly in the EU) are the most popular methodologies. The modular approach (19%) is mainly used in Africa.
2. Complete enumeration is the main type of enumeration (82% of the countries).
3. The main source of census frame is the last population census (46% of the countries), followed by the last agricultural census (41%) and administrative registers (18%).
4. CAPI (66%) and CAWI (34%), and Post (28%) overtook PAPI (26%) as the main data collection mode. Telephone interviewing (CATI, 22%) is gaining ground.
5. A growing number of countries rely on the use of technologies such as GIS and interactive online databases to disseminate results.

Findings on WCA 2020 – 2/3

- **In terms of coverage of census themes, the online survey showed that:**

1. Some themes were infrequently covered:
    - Intra-household decisions (gender role)
    - Food Insecurity Experience Scale (FIES)
    - GHG emissions
    - Fisheries.
2. These themes are optional in the WCA, meaning they are not compulsory.
3. FIES and GHG emissions could be removed as they are now covered by other sources (e.g. SDG indicators)

Findings on WCA 2020 – 3/3

Some practices of concern:

- Census questionnaires continue to be overloaded in many countries.
- Items that are not structural (e.g. production) and belong to sample surveys are forced into the census.
- Some countries exclude juridical holdings (e.g. enterprises, cooperatives, government agencies) from their censuses.
- Geo-referencing is not widely used for the location of holdings.
- In some countries, censuses are sample-based (mainly in Africa), providing structural data on farms, but not fulfilling key objectives: data for <u>small</u> admin units, <u>benchmark</u> data, complete <u>frames</u>, and measurement of <u>rare events</u>.
- In some countries, there are significant delays in publication of the census results or not adequate dissemination.

Areas of work in the WCA 2030 – 1/2

The new guidelines will stress on:

- Key role of the census: providing structural data on agriculture and the foundation for the system of agriculture statistical surveys.
- Censuses provide what sample surveys cannot: data for small admin units, benchmark data, complete frames, and measurement of rare events (unusual crops or livestock).
- Census questionnaire should therefore focus on structural items.
- For subsistence farms, to recommend a smaller set of census items.
- Census excluding juridical holdings is an incomplete census.
- Encourage dissemination of results and the use of geo-referencing to improve the presentation of results and enable integration with other datasets, GIS.
- Add a new item on vertical farming (important in landless holdings).
- Explain the (limited) contribution of the census to some SDG indicators.

Areas of work in the WCA 2030 – 2/2

The new guidelines will also stress on :

- Structural items needed for small administrative areas, benchmark, frames and to measure rare events must be collected by complete enumeration.
- Items that do not meet the above conditions could be collected by sample enumeration (e.g. a module).
- If a ‘census’ is just sample-based, it would be deemed a farm structure survey.
- Some items could be considered structural:
  - Use of machinery and tools;
  - Non-residential buildings (e.g. for livestock or poultry);
  - Use of organic practices;
  - Use of technologies (e.g. automated steering, sprayers, drones, robotic milkers).

Process and development of WCA 2030 1/2

Activity	Period
Methodological review	
Methodological review of ag-census in Asia-Pacific, Africa, Americas, Europe	Done
Prioritize recommendations and lessons learnt	Done
Consultations	
Internal consultations at FAO	Ongoing
External consultations: countries, experts, int'l organizations	Ongoing
International meetings	
Discussion in regional meetings (e.g. Latin America-*done*, Africa-*done*, Asia)	Ongoing
Writing	
Concept note and draft work plan (based on review and consultations)	April 2023
Annotated outline for WCA 2030 and technical roles assigned	May 2023
Preparation of 1st draft of the Guidelines (based on concept note)	Dec 2023
Preparation of 2nd draft of the Guidelines (after consulting FAO divisions)	Mar 2024

Process and development of WCA 2030 2/2

Activity	Period
Validation	
Internal presentation and discussion at FAO (Draft 2)	June 2024
International Expert Review/UN-CEAG at FAO HQ hybrid (Draft 3)	Nov 2024
Global consultation (Draft 4)	Early 2025
Adoption by governing bodies: UNSC, FAO (Final draft)	Early 2026
Publication and dissemination	
Revision, editing, layout and publication	Q2 2026
Translation into other languages	2026 onwards
Dissemination seminars	2026 onwards

Suggested role of the UN-CEAG

- The Committee of Experts may wish to establish a Workstream to guide the preparation of the WCA 2030 guidelines.
- The Workstream could review the drafts of the guidelines (before and after the global consultation in 2024-2025).
- The endorsement of the guidelines by UNSC's is expected in Feb 2026 at its 57th session.
- We invite expression of interest of members wishing to participate Workstream, ideally from all world's regions.

SESSION 5

Additional topics of interest in the areas of Food security, agriculture and rural statistics for consideration in the UN-CEAG programme of work 2024-27

Thank you

https://www.fao.org/about/ce-on-food-security-agricultural-rural-statistics/en/