Author(s): Tuomas Rajala, Juha Heikkinen, Sophia Gogo, Joyce Ahimbisibwe, Geofrey Bakanga, Nurdin Chamuya, Javier Garcia Perez, Edward Kilawe, Shani Kiluvia, David Morales, Emmanuel Nzunda, Jared Otieno, Jonathan Sawaya, Lauri Vesa, Eliakimu Zahabu, Matieu Henry

Title: NAFORMA: National Forest Resources Monitoring and Assessment of United Republic of Tanzania Mainland – Sampling Design Options for 2nd Biophysical Inventory (NAFORMA II)

Year: 2022

Version: Publisher's version

Please cite the original version:

Rajala, T., Heikkinen, J., Gogo, S., Ahimbisibwe, J., Bakanga, G., Chamuya, N., Garcia Perez, J., Kilawe, E., Kiluvia, S., Morales, D., Nzunda, E., Otieno, J., Sawaya, J., Vesa, L., Zahabu, E., Henry, M. 2022. NAFORMA: National Forest Resources Monitoring and Assessment of United Republic of Tanzania Mainland – Sampling Design Options for 2nd Biophysical Inventory (NAFORMA II). Rome, FAO. https://doi.org/10.4060/cc0572en

Food and Agriculture Organization of the United Nations

NAFORMA:
National Forest Resources Monitoring and Assessment of Tanzania Mainland

Sampling design options for 2nd Biophysical Inventory (NAFORMA II)

TANZANIA FOREST SERVICES AGENCY
TFS
FOREST IS WEALTH

Luke
NATURAL RESOURCES INSTITUTE FINLAND

NAFORMA: National Forest Resources Monitoring and Assessment of Tanzania Mainland

Sampling Design Options for 2nd Biophysical Inventory (NAFORMA II)

By Tuomas Rajala, Juha Heikkinen, Sophia Gogo, Joyce Ahimbisibwe, Geofrey Bakanga, Nurdin Chamuya, Javier Garcia Perez, Edward Kilawe, Shani Kiluvia, David Morales, Emmanuel Nzunda, Jared Otieno, Jonathan Sawaya, Lauri Vesa, Eliakimu Zahabu, Matieu Henry

Food and Agriculture Organization of the United Nations

Rome 2022

Required citation:

Rajala, T., Heikkinen, J., Gogo, S., Ahimbisibwe, J., Bakanga, G., Chamuya, N., Garcia Perez, J., Kilawe, E., Kiluvia, S., Morales, D., Nzunda, E., Otieno, J., Sawaya, J., Vesa, L., Zahabu, E., Henry, M. 2022. *NAFORMA: National Forest Resources Monitoring and Assessment of United Republic of Tanzania Mainland – Sampling Design Options for 2nd Biophysical Inventory (NAFORMA II).* Rome, FAO. https://doi.org/10.4060/cc0572en

ISBN 978-92-5-136435-2

Contents

Figures

Tables

Acknowledgements

Special thanks to Tanzania Forest Services (TFS) Agency Conservation Commissioner Prof. Dos Santos A. Silayo for accepting and directing the formulation of concept note that eventually guided technical discussions and directives for the development the NAFORMA II sampling design. He further accepted and supported the formulation of a dedicated National Task Force team (Shani Khalid, Jared Otieno, Zainabu Bungwa, Dr. Elikana John, Prof. Emmanuel Nzunda and Prof. Eliakimu Zahabu) that worked very closely with Luke and FAO to achieve the desired NAFORMA II sampling design document.

We are grateful to TFS Forest Inventory team that participated in the testing of the use of new design to relocate selected NAFORMA I clusters in Morogoro Region.

We are so appreciative to FAO Country Office for providing technical experts who worked tirelessly with the team in the development process of the design and supported Land Cover Classification.

NAFORMA I Phase 1 data was kindly provided by Prof. Erkki Tomppo. We are also grateful to Prof. Annika Kangas and Dr. Kari T. Korhonen from Luke for their insightful advice when formulating the new design option.

Abbreviations and acronyms

Abbreviation/acronym	**Description**
CV	Coefficient of variation
DBH	Tree diameter at breast height
FAO	Food and Agriculture Organization of the United Nations
GPS	Global positioning system
ha	Hectare (10 000 square meters)
LC	Land cover
LPM	Local pivotal method
LULC	Land-use and land-cover
Luke	Natural Resources Institute Finland
NDVI	Normalized Difference Vegetation Index
NFI	National Forest Inventory
RS	Remote sensing
SEPAL	System for Earth Observation Data Access, Processing, and Analysis for Land Monitoring
TFS	Tanzania Forest Services Agency
URT	United Republic of Tanzania
USD	United States Dollar

Executive summary

Three options for the sampling design of the field plot clusters of NAFORMA II biophysical survey are compared in this report. Option 1 consists of re-measuring all NAFORMA I field sample plots (3 205 clusters) and Option 2 of re-measuring only those that were established as permanent (848 clusters). The recommended Option 3 is a compromise between these two "extreme" options: Re-measure a subset (1 405 clusters) of NAFORMA I field sample plots including (almost) all permanent clusters and a carefully selected set of other NAFORMA I field plot clusters to obtain a uniform sample within each TFS zone.

Design Option 3 has the following features:

- Sampling intensity is uniform within each TFS zone. This makes it simple to use the data. For example, mean volumes can be estimated by averages over the plots.
- The selected clusters are well-spread over the target population.
- The anticipated precision of land-class area and mean wood volume relative to sample size is nearly as good as that of NAFORMA I.
- All proposed clusters were measured in NAFORMA I, which enables precise estimation of change based on repeated measurements.

The costs and precision were anticipated by utilizing NAFORMA I field data, information about subsequent improvements in the road network, and changes in land-use using satellite imaging derived land-class maps.

1. Introduction

United Republic of Tanzania (URT) is one of the twelve mega-diverse countries of the world endowed with different natural ecosystems that harbor a massive wealth of biodiversity. United Republic of Tanzania's mainland covers a total area of 945,087 km^2, with 48.1 million ha of forests of which 93% is Central Zambezian Miombo Woodland (Ministry of Natural Resources and Tourism 2015). The country's diverse ecosystems range from semi-arid to tropical forest and encompass some of the most diverse landscapes in the world, providing a wide range of services that are of vital importance to the livelihoods and economies of the country. The country hosts parts of six Global Biodiversity Hotspots, with more than one-third of the total plant species on the continent and about 20% of the large mammal population.

Tanzania Forest Services Agency (TFS) is a semi-autonomous government agency established in 2010 to take over most of the operational roles and functions of the Forestry and Beekeeping Division of the Ministry of Natural Resources and Tourism. TFS is mandated to manage national forest and bee reserves, forests on general land and tree seed production. The Agency operates in seven zones namely: Central, Eastern, Lake, Northern, Southern Highlands, Southern and Western; and 169 districts of the URT mainland. In its 11 years of operation, TFS ensured that there is efficient and effective management of the national forest and bee reserves (Tanzania Forest Services Agency, 2020).

The first-ever comprehensive National Forest Inventory (NFI) in URT, popularly known as National Forestry Resources Monitoring and Assessment (NAFORMA; to be called NAFORMA I in this report), was created to address some of the issues raised in the National Forest Programme. NAFORMA has been a pre-request source of information for different users. TFS and stakeholders plan to conduct a second assessment (NAFORMA II) that will assess the current forest status and hence determine changes since NAFORMA I. NAFORMA II aims to address the growing demands of forest data users by applying the latest technological and cost-effective approaches. A need for several technical improvements has also been recognized, including stronger institutionalization, national ownership of the process to ensure its sustainability, better integration of ground and remote sensing information, improved data management, data sharing and information system capacities, and increased use of national forest information by national stakeholders including research and academia.

A two-phase stratified sampling design was developed for NAFORMA I (Tomppo *et al*., 2014) and data collection according to that design was implemented in 2009–2014 (Ministry of Natural Resources and Tourism 2015). The relatively high complexity of the sampling design is one of the bottlenecks hindering utilization of the collected information. The variable sampling intensity between fragmented strata needs to be counter-balanced by appropriate weighting of measurements. Determination of appropriate weights is particularly challenging, when computing results for smaller regions or sub-classes.

Comprehensive field measurements and time-use records that were collected in NAFORMA I can provide useful support for anticipating the costs and efficiency of alternative options for the design of a new assessment. The information has also been complemented by remote-sensing based estimates of land-cover changes and by up-to-date information on the road network.

In general, the aim of National Forest Inventories, such as NAFORMA, is to provide information at national and regional levels. In order to obtain useful results for smaller areas, such as forest reserves, NFI data can be supplemented by auxiliary information, such as digital maps and/or remote sensing ("multisource inventory"; e.g. Tomppo 2009). Another option for obtaining useful information for forest management planning is to supplement NFI data with additional field measurements at management units of particular interest.

The aim of this report is to propose and evaluate three options for the sampling design of biophysical measurements of NAFORMA II. The evaluation also includes estimates of the fieldwork costs and of the anticipated precision of the most important inventory estimates. Anticipated precision is presented for the estimates of the areas of primary land-use classes and wood volume for both national and regional results.

While the first two options follow the design of NAFORMA I, Option 3 was particularly aimed at simplifying the design and, consequently, the proper estimation methods. It is also expected to provide more stable framework for long-term monitoring, as well as spatially representative and balanced layout of field sample plots, which is also important when NFI data are used as the basis of multisource estimates. The report also briefly discusses how the proposed NAFORMA II sample can supplement local information needs. The preliminary work by Kangas *et al*. (2015) was taken into consideration when constructing Option 3.

2. Sampling design objectives

The main update objectives were identified in discussions between TFS, FAO and Luke. Significant simplifications were to be made to the NAFORMA I two-phase stratified sampling design, where each potential field plot cluster was designated to one of 18 different strata and sampling intensity varied between these strata. Calculation of statistics at regional and sub-regional levels was too complicated. Hence, the new design should be more flexible in terms of regional applications. The new design should give TFS a better understanding and ownership of the data collection process to improve knowledge transfer to national stakeholders and to have the possibility to design increased field data collections for specific forest management and conservation purposes.

The objectives also included updating of cost-per-cluster calculation utilizing information about improvements in the road network and about changes in land cover obtained using satellite images. The land-cover changes were also to be taken into account when evaluating the anticipated precision of the inventory estimates.

The new design was also expected to provide a statistically sound framework for long-term monitoring. While NAFORMA I sampling design was highly dependent on the current state of forests at the time of its planning, the proposed new design is temporally stable since it is based on regional rather than cluster-level stratification.

The proposed design suggests sample sizes that are in line with comments from TFS on their capacity to include new permanent plots. As the main purpose is a national level inventory, it is clear that the limited number of resources will not be sufficient for serving all sub-regional applications. Section 5 discusses how to intensify the data collection in specific locations.

3. Sampling design options

Three options for NAFORMA II sampling design are compared and discussed in this report:

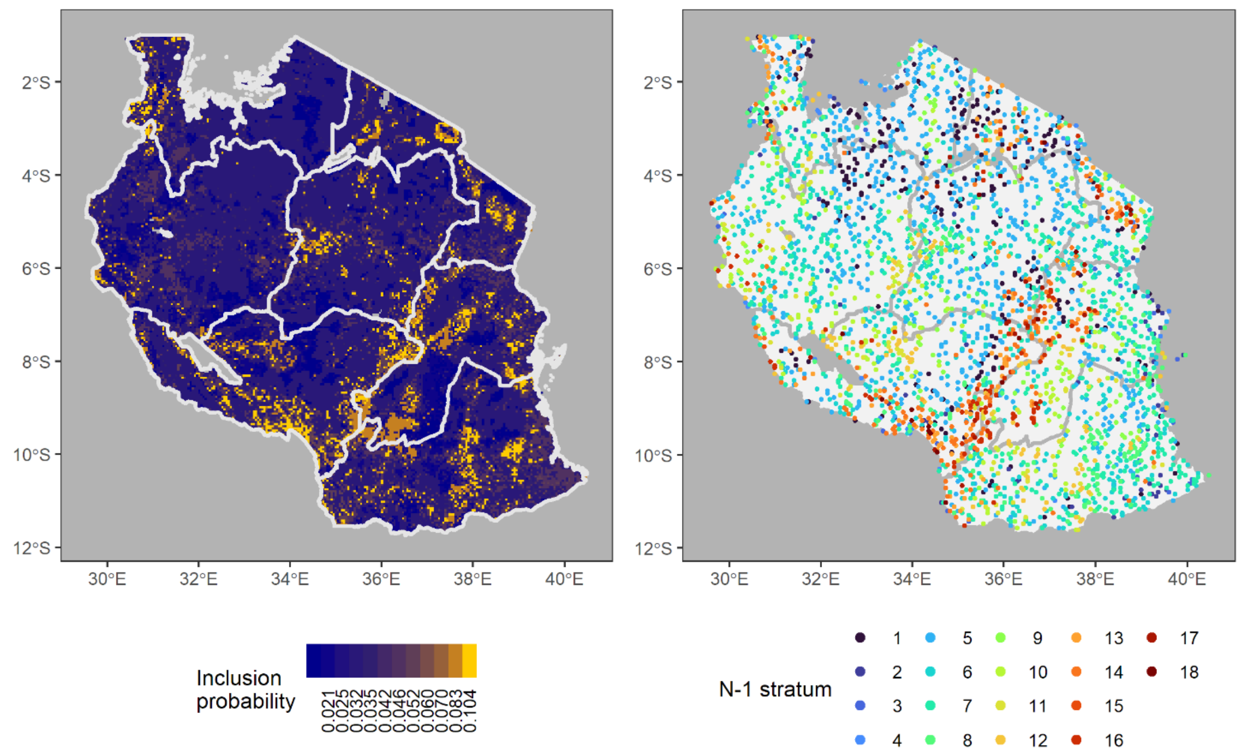
- re-use of NAFORMA I design (Option 1: “Business as usual”);
- re-measurement of NAFORMA I permanent clusters (Option 2: “Monitoring”);
- augmentation of Option 2 with some clusters that were established as temporary in NAFORMA I (Option 3: “Balanced”).

In Option 3, the augmentation was deviced so that the resulting set of clusters can be considered as a regionally stratified random sample. From such sample, it is statistically sound to compute regional inventory results as simple expansion estimators.

Option 1: “Business as usual”

Option 1 consists of re-measuring all 3 205 field plot clusters of NAFORMA I (Figure 3-1). The two-phase stratified sampling design that was used to select them is described in detail in publications by Tomppo *et al*. (2014) and Ministry of Natural Resources and Tourism (2015).

Figure 3-1 Cluster-specific sample inclusion probabilities in NAFORMA I scaled to match Option 3 cluster count; TFS zones marked with light gray (left). Locations of the 3 205 NAFORMA I field plot clusters, colored by stratum (right).



Sources: Administrative boundary: National Geospatial Information Authority Surveys and Mapping Division, Ministry of Lands and Human, Settlements Development P.O.Box 9201 Dar-es-Salaam, United Republic of Tanzania (the) http://www.ardhi.gov]. [Cited 30 June 2022]. Maps conform to UN. 2006. Map No. 3667, Rev. 6. https://www.un.org/geospatial/content/united-republic-tanzania

The land area of URT was covered with a regular grid in 5 km intervals to create so called Phase 1 locations for clusters. For each location, several auxiliary variables were gathered and each Phase 1 cluster was designated to one of 18 strata constructed on the basis of anticipated measurement time, predicted wood volume, and slope of the terrain. The Phase 2 clusters, to be measured in the field were selected as a stratified random sample from the Phase 1 clusters. The sampling density of Phase 1 clusters, as well as number of field plots in a cluster varied between strata, the latter ranging between six and ten plots per cluster (Table 3-1).

For the purposes of long-term monitoring, approximately 25% of the Phase 2 clusters were established as permanent, and measures were taken to facilitate accurate re-measurement of these clusters. The center location of each plot in a permanent cluster was recorded with high-precision GPS (Global positioning system). The center also was marked in the field with a permanent metal rod, forward and back bearings toward permanent objects such as big trees or rocks close to the center were recorded, and such objects were sprayed with red paint. Temporary clusters were not marked with metal rods, but high-precision GPS coordinates were recorded. In the field test, conducted by TFS during 2021, it was discovered that both types of clusters can be re-visited accurately using information gathered during NAFORMA I.

Table 3-1 Stratum-specific numbers of Phase 1 and Phase 2 clusters, relative areas of strata (Phase 1 %), sampling densities, proportions of permanent clusters and average realized work-times of cluster measurements.

Stratum	Phase 1 clusters	Phase 1 %	Phase 2 sampling %	Phase 2 measured clusters	Phase 2 permanent %	Average days per cluster
1	3 080	8.7	8.3	244	23.4	0.53
2	626	1.8	10.0	63	30.2	0.61
3	254	0.7	12.5	26	26.9	0.64
4	83	0.2	50.0	37	27.0	0.54
5	8 852	24.9	7.7	667	26.1	0.62
6	7 282	20.5	8.3	592	26.4	0.79
7	4 149	11.7	11.1	445	26.5	0.83
8	896	2.5	25.0	205	24.4	0.81
9	2 252	6.3	5.0	95	31.6	0.89
10	2 766	7.8	5.9	159	27.0	0.99
11	2 033	5.7	7.7	139	27.3	1.02
12	673	1.9	20.0	110	30.0	1.15
13	741	2.1	14.3	105	27.6	0.71
14	738	2.1	25.0	177	29.4	0.80
15	165	0.5	7.7	8	12.5	1.31
16	598	1.7	20.0	87	24.1	1.22
17	246	0.7	16.7	29	20.7	0.99
18	94	0.3	25.0	17	23.5	1.26
all	35 528	100.0	9.6	3 205	26.5	0.78

Source: FAO, 2022.

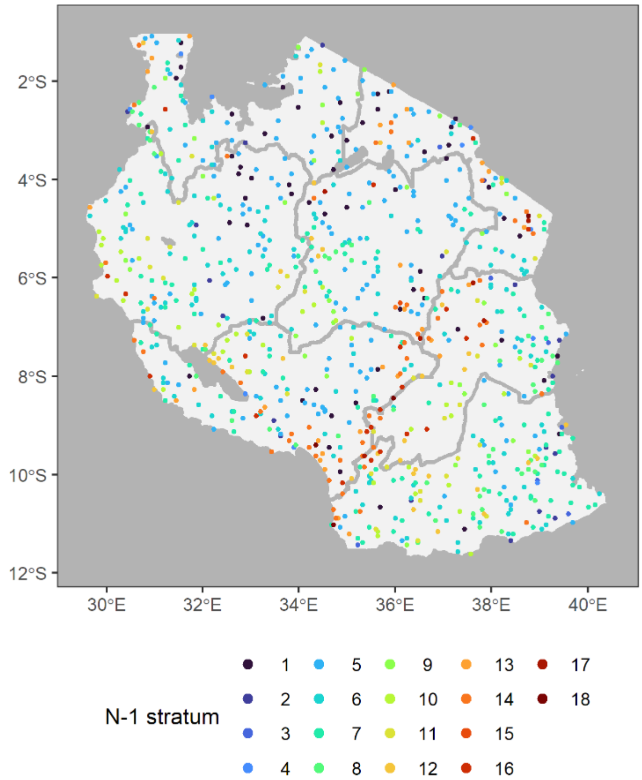
The strata of NAFORMA I are based on multi-dimensional, bio-physical properties of the landscape and are not generally regionally contiguous, which complicates calculations at regional and sub-regional level as the variable sampling density needs to be accounted for in order to avoid representation biases.

Option 2: "Monitoring"

Option 2 consists of re-measuring the 848 Phase 2 clusters of NAFORMA I that were established as permanent (Figure 3-2). This design inherits the statistical properties of the whole NAFORMA I field plot sample (Option 1), since the former is a 25% random sample from the latter.

Difficulties due to the stratification in sub-regional and class-specific estimation would increase. Owing to the reduction in sample size, some strata would have very few or no clusters at all, and the estimation is difficult already at TFS zone level. As illustrated in Figure 3-2, the presence of some strata is low in some zones.

Figure 3-2 Locations of the 848 NAFORMA I permanent Phase 2 clusters, colored by stratum. TFS zones with gray lines.



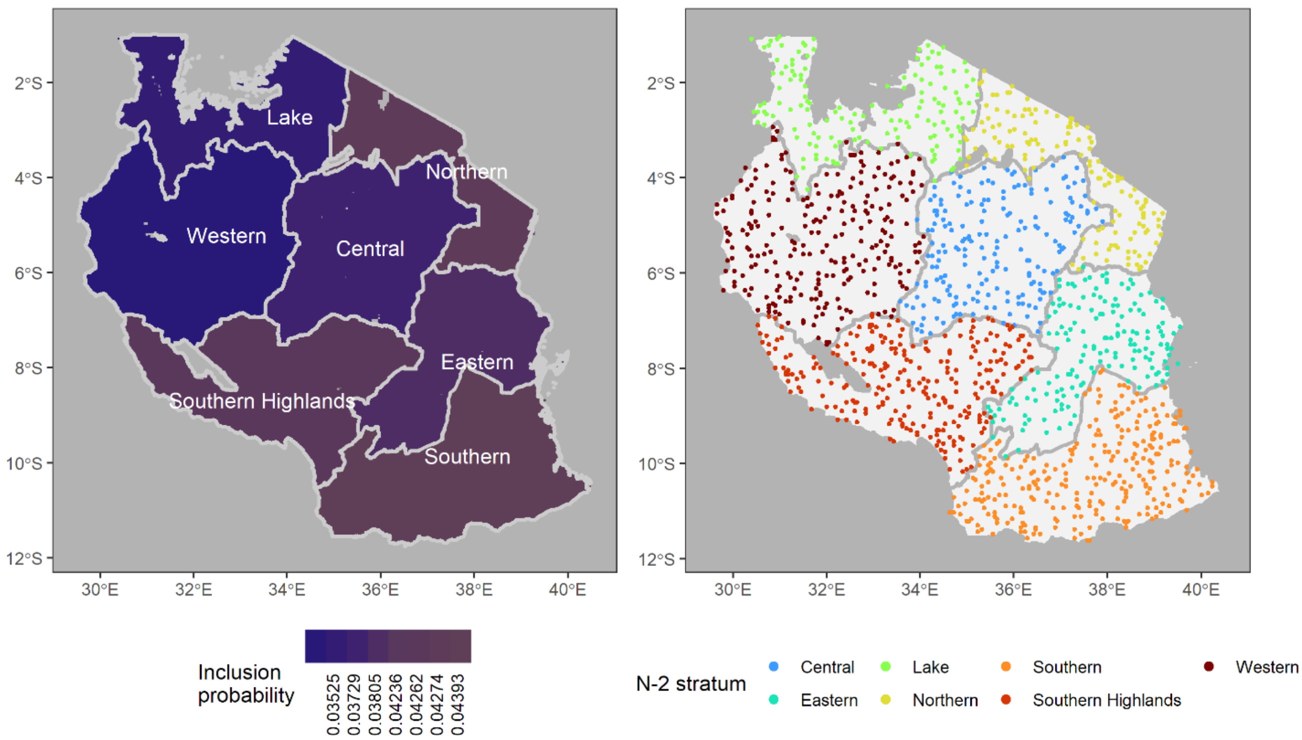
Sources: Administrative boundary: National Geospatial Information Authority Surveys and Mapping Division, Ministry of Lands and Human, Settlements Development P.O.Box 9201 Dar-es-Salaam, United Republic of Tanzania (the) http://www.ardhi.gov]. [Cited 30 June 2022]. Maps conform to UN. 2006. Map No. 3667, Rev. 6. https://www.un.org/geospatial/content/united-republic-tanzania

Option 3: "Balanced"

Option 3 results in a set of clusters that can be considered as a simple random sample within each TFS zone, and hence a regionally stratified sample. This is obtained by including (almost) all permanent clusters of NAFORMA I plus some carefully selected Phase 2 clusters that were established as temporary. Thus Option 3 is intermediate between the other two: A subset of Option 1 clusters is selected and (almost) all clusters of Option 2 are included. The sample size, 1 405 clusters, is roughly half of that in Option 1 and twice of that in Option 2. It was determined on the basis of comments received from TFS on their capacity to include new permanent plots.

It is not possible to counter-balance the NAFORMA I stratification for all possible sub-regional calculations due to the limited number of NAFORMA I Phase 2 clusters. However, the benefit of restricting the NAFORMA II design to that set of potential clusters (highly efficient change estimation based on repeated measurements), was considered more important. As the main aim of the NFI is national level assessment and the work is managed by TFS, Option 3 was based on regional split of URT into administrative zones used by TFS (Figure 3-3). It will render the sampling weights uniform within these seven zones.

Figure 3-3 Zone-specific sample inclusion probabilities of Phase 1 clusters in Option 3 (left) and the locations of 1 405 clusters of one sample generated according to the design, colored by the resulting new stratification (right).



Sources: Administrative boundary: National Geospatial Information Authority Surveys and Mapping Division, Ministry of Lands and Human, Settlements Development P.O.Box 9201 Dar-es-Salaam, United Republic of Tanzania (the) http://www.ardhi.gov]. [Cited 30 June 2022]. Maps conform to UN. 2006. Map No. 3667, Rev. 6. https://www.un.org/geospatial/content/united-republic-tanzania

Note that the sampling inclusion probabilities in Figure 3-3 correspond to sampling in the hypothetical Phase 2 situation of NAFORMA II: The proposed Option 3 will not sample new locations from the NAFORMA I Phase 1 grid, but will suggest re-measuring a specific set of clusters already present in NAFORMA I. The proposed sampling weights are determined so that they balance those of NAFORMA I. As a result, NAFORMA I stratification can be ignored in the future. This also holds in the estimation of changes, provided that it is based on changes in those clusters selected only according to Option 3 design, i.e., ignoring NAFORMA I measurements of the "left-out" clusters. More precise change estimators would result from utilizing the full NAFORMA I data, but that would require weighting according to stratum-specific inclusion probabilities of NAFORMA I.

In other words: Option 3 is, in essence, a stratified sampling design such that each TFS zone is one stratum and each potential sample plot cluster within one zone has the same probability density for inclusion in the field sample. To achieve this by choosing clusters to re-measure from NAFORMA I field clusters (Phase 2), the original sampling weights had to be balanced. This was implemented as follows. Within each zone, the permanent clusters were distributed as they were in NAFORMA I strata, call this distribution F_1. A distribution for sampling the additional clusters was then deviced so that the joint distribution F_2 of all clusters between the 18 strata of Tomppo *et al*. (2014) is as close as possible to the distribution of their area proportions within the zone (more details in Appendix A4). Then the NAFORMA I stratification can be ignored when computing statistical results, and unbiased estimates within each TFS zone are obtained by simple averages (expansion estimators) over the field plot clusters.

Finally, to get as close as possible to a properly balanced design within each zone, the number of clusters in some NAFORMA I strata should be smaller than then number of clusters established as permanent. We therefore recommend "relocating" a small number of permanent clusters: "Abandon" a few permanent clusters from some NAFORMA I strata and include, instead, clusters from some other NAFORMA I strata from the same TFS zone. If this is not done, some NAFORMA I strata will be overly represented within a zone, and NAFORMA I design can not be fully ignored in NAFORMA II based calculations. The number of relocations needed per zone is not large, between 3 and 10 clusters per zone (Appendix A1).

Within stratum and zone, the field plot clusters for NAFORMA II can then be sampled from the Phase 2 clusters of NAFORMA I. To sample the actual clusters and fix the final design, the use of a spatially balanced sampling technique, such as the local pivotal method (Grafström *et al*., 2012), is recommended to improve spatial homogeneity in the sample (Appendix A4). One benefit of a spatially spread and balanced sample is that it likely to be more representative of the environmental properties of the area at large.

4. Comparison of options

Design Option 3 was mainly motivated by the need to simplify the two-phase stratified sampling design of NAFORMA I. This was achieved by replacing the cluster-level stratification (18 spatially non-contiguous strata) with regional stratification (seven zones). Computation of regional inventory results from data collected according to Option 3 sampling design is straightforward. Furthermore, sub-regional computations do not suffer from empty strata, although the sample size may not be large enough for sufficiently precise estimation in small subregions (see Chapter 5).

Sampling design of NAFORMA I, including stratification based on cluster-level auxiliary data, was motivated by optimized cost-efficiency. Option 3 inherits some of its optimized properties:

- spatially representative systematic design of the underlying Phase 1 sample;
- clustering of sample plots for cost-efficient field work (mainly one cluster per work day with walking distances between plots within a cluster);
- less plots per cluster in clusters with greater expected time consumption per plot;
- zone-specific sampling densities proportional to those of NAFORMA I reflecting differences in mean wood volume and accessibility;

However, Option 3 cannot be expected to be quite as efficient as the other options, because it does not entail utilization of auxiliary information at cluster level. The anticipated loss in statistical efficiency (the price to be paid for the significant simplification) was measured in terms of statistical uncertainty, using coefficient of variation (CV) of the estimators of mean wood volume (m^3/ha) and the areal coverage (Million hectares, Mha) of the combined forest and woodland (Level 1 land use and land cover classes as described in Ministry of Natural Resources and Tourism, 2015). This was based on variance estimators presented by Tomppo *et al*. (2011, sections 3.5.1 and 3.5.2).

Since it was not possible to reconstruct the exactly same version of data and to exactly replicate the computations used in the NAFORMA I report (Ministry of Natural Resources and Tourism 2015), our estimates differ slightly from those in Table 4.1 of that report. However, this is not expected to impact comparison of the design options presented here, since the same data and methods were applied to each option. It should be noted that the CV only measures uncertainty due to sampling and, in particular, does not account for potential model errors in tree-level predictions in wood volume.

The cost estimates considered in this comparison only included the time cost of cluster measurement. In particular, fixed costs and the travel from cluster to cluster or base to cluster are not included. The estimates were based on the realized measurement times of NAFORMA I clusters, which were updated according to the changes in land cover (Appendix A5) and improvements in road network. The details are provided in Appendix A2.

Option 3 suggests measuring 1405 clusters, or 44% of the 3205 NAFORMA I clusters. The CV% is anticipated to increase from 1.0-1.2% to 1.3-1.8%, which is close to the expected from such a reduction in sample size under similarly efficient designs (Table 4-1). The cost scales as expected, with an anticipated 56% reduction. Note that since the cost calculation utilized here are rather simplistic and only related to the plot measurements, qualitative comparisons between the options are more meaningful than the absolute values themselves. The reduction in precision depends on the zone as the sampled cluster counts vary, but in most zones the CV% by Option 3 stays below 10%, and nowhere exceeds 15% (Table 4-2, Figure 4-1). Design option 3 can be supplemented by additional data collection to improve precision in regions of particular interest as briefly explained in Chapter 5.

The main benefit of Option 3 is further illustrated in

Figure 4-2. The Dodoma region covers ca. 30% of the Central zone. Not all strata are present for Options 1 and 2, but in Option 3 the sub-region is subsumed in a single stratum. The sample size of Option 3 is 40% of Option 1, and the CV% of mean wood volume is expected to increase only modestly from 10.2% to 12.5%.

Table 4-1 Comparison of the three design options in terms of anticipated precision (CV) in the estimates of the area of combined forest and woodland and mean wood volume over the whole country and the anticipated costs. Assumed cost/day is approximately 1 000 USD.

Option	Design	CV% of area	CV% of m3/ha	N. of clusters	N. of plots	Anticipated person-days	Anticipated costs, USD
1	Business as Usual	1.0	1.2	3 205	30 844	2 501	2 502 000
2	Monitoring	1.8	2.2	848	8 168	656	657 000
3	Balanced	1.3	1.8	1 405	13 725	1 095	1 096 000

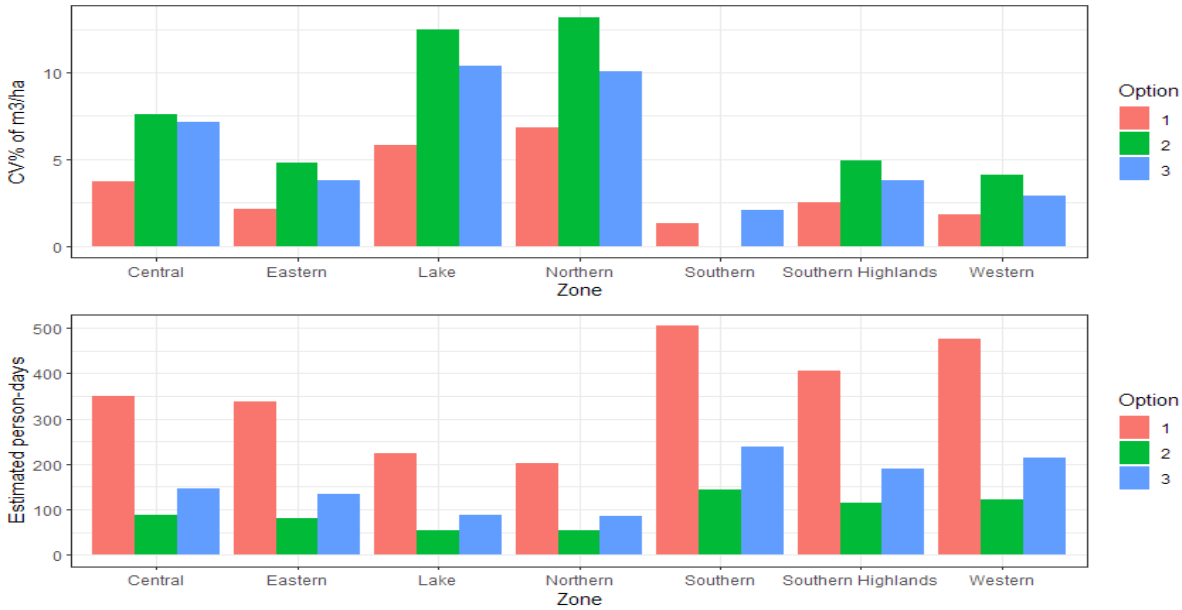
Source: FAO, 2022

Table 4-2 Zone-level comparison (as in Table 4-1).

TFS zone	Option	CV% of area	CV% of m3/ha	N. of clusters	N. of plots	Estimated person-days	Estimated cost, USD
Central	1	5.7	3.7	470	4 622	350	351 000
	2	11.5	7.6	115	1 122	89	89 000
	3	8.7	7.2	191	1 891	146	147 000
Eastern	1	6.2	2.1	399	3 755	338	339 000
	2	12.7	4.8	95	894	81	81 000
	3	9.0	3.8	159	1 525	133	134 000
Lake	1	8.1	5.8	360	3 469	225	226 000
	2	17.0	12.5	86	830	53	53 000
	3	12.6	10.4	143	1 400	87	88 000
Northern	1	7.8	6.8	304	2 871	202	202 000
	2	13.6	13.2	80	745	55	55 000
	3	11.1	10.0	131	1 264	85	86 000
Southern	1	4.1	1.3	583	5 703	505	505 000
	2	7.5		167	1 634	144	144 000
	3	5.1	2.1	278	2 743	238	239 000
S. Highlands	1	5.1	2.5	540	5 011	406	406 000
	2	8.7	4.9	153	1 443	114	114 000
	3	6.4	3.8	253	2 415	190	191 000
Western	1	4.1	1.8	549	5 413	476	477 000
	2	7.5	4.1	152	1 500	122	122 000
	3	6.1	2.9	250	2 483	213	214 000

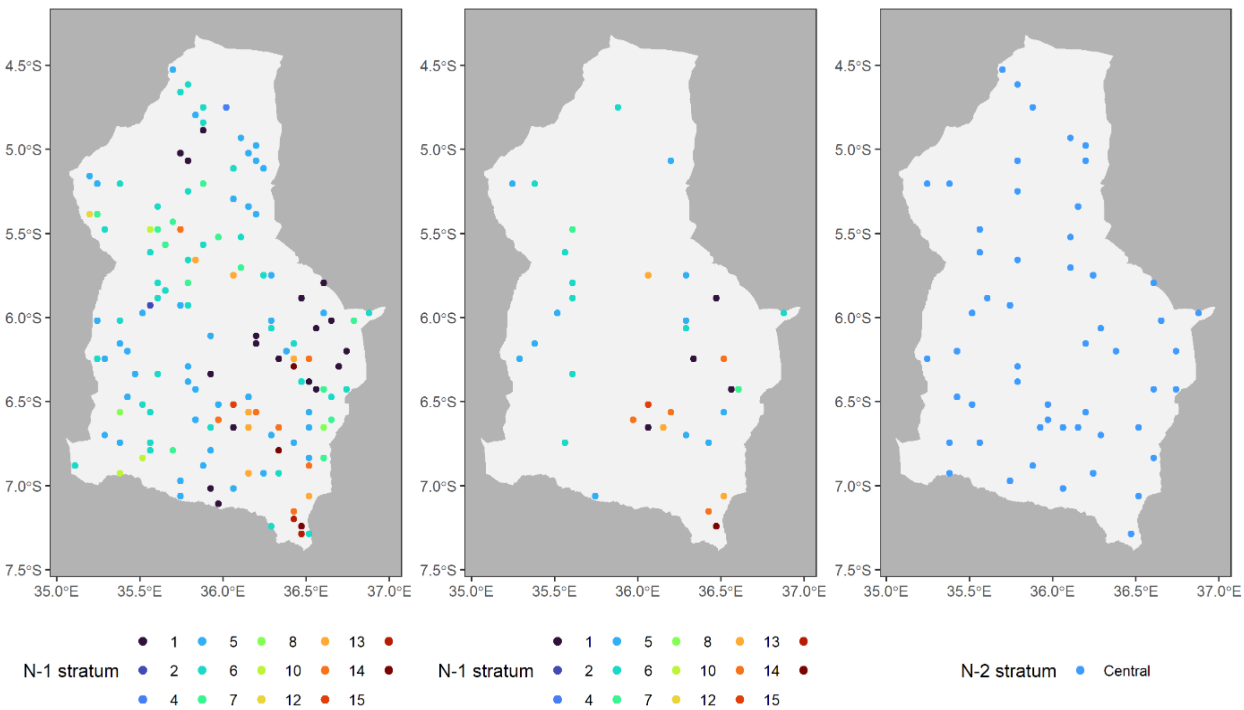
Source: FAO, 2022

Figure 4-1 Comparison of the three design options across TFS zones: Statistical efficiency in terms of CV% of mean wood volume (top) and anticipated cost in person-days (bottom).



Source: FAO, 2022

Figure 4-2 Field plot clusters in the Dodoma region selected according to the three sampling design options: Option 1 (left, 139 clusters), Option 2 (middle, 35 clusters), and Option 3 (right, 53 clusters).



Sources: Administrative boundary: National Geospatial Information Authority Surveys and Mapping Division, Ministry of Lands and Human, Settlements Development P.O.Box 9201 Dar-es-Salaam, United Republic of Tanzania (the) http://www.ardhi.gov [Cited 30 June 2022]. Maps conform to UN. 2006. Map No. 3667, Rev. 6. https://www.un.org/geospatial/content/united-republic-tanzania

5. Supplemental data collection

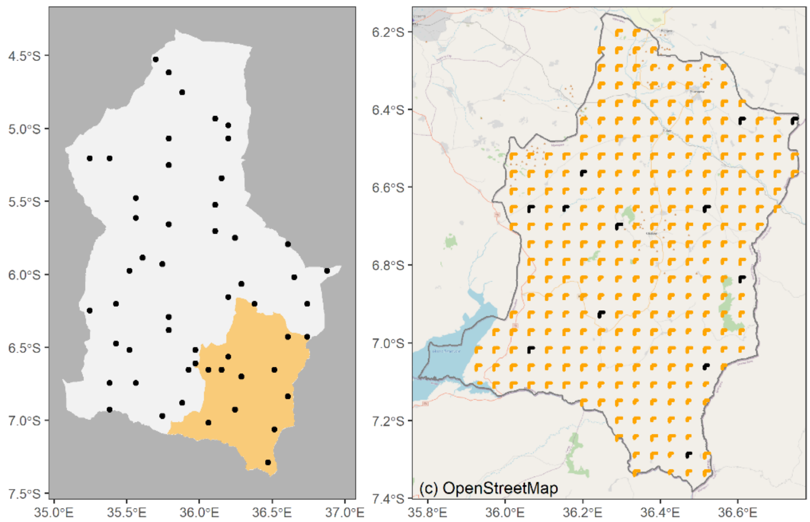
NFI generate reliable statistics and information mainly at national level based on level of sampling intensity required to represent the whole country. Due to higher costs associated with ground surveys, level of precision required, and time to cover the whole country, only a reconnaissance inventory can be obtained. More samples are recommended to be collected when detailed forest inventory is conducted on the specific forest areas or district.

Stakeholder's requirement from NAFORMA II to generate reliable statistics to district and forest management unit level will be implemented in separate and specific inventories conducted regularly depending on the information necessity.

According to our results, Option 3 is anticipated to allow for reasonably precise estimation of primary parameters, such as forest and woodland area and mean wood volume, at the regional level of TFS zones from the collected field data only. For smaller areal units of particular interest, multisource inventory methods may be used to improve precision or supplemental data collection can be based on Phase 1 clusters of NAFORMA I (see Figure 5-1, for an example). Although design option 3 allocates only 12 clusters to the Mpwapwa district, it contains a total of 290 NAFORMA I Phase 1 clusters.

The field sample collected according to Option 3 is balanced within each TFS zones, but not necessarily in smaller regions. It is therefore recommended that a similar balancing procedure is undertaken within the area of interest as was done in Option 3 at the level of TFS zones in case the NAFORMA II field data are to be supplemented by additional field plot clusters.

Figure 5-1 Option 3 clusters in the Dodoma region; Mpwapwa district highlighted (left), and NAFORMA I Phase 1 clusters in the Mpwapwa district (right, Option 3 clusters in black).



Sources: Administrative boundary: National Geospatial Information Authority Surveys and Mapping Division, Ministry of Lands and Human, Settlements Development P.O.Box 9201 Dar-es-Salaam, United Republic of Tanzania (the) http://www.ardhi.gov. [Cited 30 June 2022]. Maps conform to UN. 2006. Map No. 3667, Rev. 6. https://www.un.org/geospatial/content/united-republic-tanzania

6. Challenges and recommendations

There were difficulties in identifying the most complete NAFORMA I dataset, and it was unclear which version was used in the official report by the Ministry of Natural Resources and Tourism (2015). Partly due to these issues and partly due to the scarcity of details on the exact methods, there were difficulties in reproducing the statistical analysis of that report. We recommend improving data management and distribution protocols, especially versioning, and increasing reproducibility practices in processing and analysis stages.

The methodology implemented in NAFORMA I can be adapted and improved, for example by adding new variables and using the latest technology, such as using mobile data collection devices in the field, data storage and processing in the cloud application, and using hypsometer with transponder which enables automatic slope corrections for distances. Furthermore, it is recommended that NAFORMA II will improve the following aspects of NFI:

- Take into account the historical changes in the national forest definition in order to harmonize the data analysis and reporting consistently.
- Establish clear protocols for data flow between the NFI levels (i.e. operators).
- Reinforce links between remote sensing analysis, inventory planning and field data collecting.
- Use the field data for the forest cover/change detection analysis.
- Include forest degradation indicators into the list of attributes to be collected.
- Reduce the number of measurement errors through efficient training and supervision, and improve the QA/QC procedures.

NAFORMA data sharing protocols should be revised to reinforce links with the academia by allowing the research teams in URT and Tanzanian MSc/PhD overseas students to access the NFI data and to improve the quality of the variables collected. This will yield an increased number of topical publications, theses, and books that can emphasize the importance, complexity and uniqueness of the forest and woodlands.

A national forest monitoring system, such as NAFORMA, requires an efficient data management system that encompasses data entry, quality control, and archival and long-term accessibility of both collected data and associated metadata (FAO 2017, section 5.4.1). Data management starts with recording the data. The technology for recording will depend on the type of data collected (e.g. remote sensing, manual analogical data entry and digital data logging). Both raw field data and "clean" data need to be permanently stored and backed up. Ideally, a single, time-stamped and version-controlled current copy of the official data should be stored on a central server (with an exact copy on another server), rather than as multiple versions on various workstations. This approach facilitates data integrity, currency, and sharing. The documentation of the information management system should include descriptions of the data, the database structure, and the metadata.

High priority should be given to capacity building on data management skills of the staff. In addition, NAFORMA II shall:

- implement a detailed database structure and management protocol (including hardware and software requirements);
- choose data formats that will be in use for the foreseeable future and that permit interoperability, rather than developing and/or using custom-built or obscure formats;
- ensure integrity of the source database, especially if a part of the data is exported for analysis using different software;
- include metadata comprising the description of the various datasets (e.g. age of creation, location, data owner, access rights, etc.) in the data storage system, following international standards of metadata format to the extent possible;

If NAFORMA II data are collected according to the recommended Option 3, then the field plot **clusters** within each TFS zone can be considered as a random sample of all possible clusters within the zone. It is recommended that the inventory results based on NAFORMA II field data only (as opposed to multisource estimates, for example) are computed by zone with equal weight given to all clusters within the zone. Most clusters have ten sample plots, but some have less. Thus the plots from those clusters that have less than ten plots by design should be upweighted accordingly.

7. Conclusion

Tomppo *et al*. (2014) devised a cost-efficient sampling design for NAFORMA I by considering the anticipated wood volume and accessibility at the level of individual Phase 1 clusters and determining their Phase 2 sample inclusion probabilities accordingly. However, the price that was paid for the optimized efficiency was a considerable complexity in the design leading to the need of non-trivial weighting of field data to obtain unbiased estimators. As an alternative, Option 3 is now proposed as a design that includes most of the clusters established as permanent in NAFORMA I, but at the same time simplifies the design with the help of a moderate number of additional clusters chosen from those that were measured in the field in NAFORMA I but regarded as temporary ones. In Option 3, the sampling intensity is uniform within each TFS zone, which means that the zone-level results can be computed without any re-weighting.

In Option 3, sampling density varies slightly between TFS zones following the corresponding average differences in areas and wood volumes in NAFORMA I results. In that way, some of the efficiency of NAFORMA I design is inherited by Option 3: Zones which contained largest proportions of wood volumes are sampled more intensively, just like in NAFORMA I. According to the sampling simulations from NAFORMA I data, the country-level efficiency (in relation to sample size) of Option 3 is actually anticipated to be quite close to that of NAFORMA I.

The simplified design is also expected to offer a better basis for long term monitoring. If sub-sampling is necessary in future campaigns, simple random samples of the clusters of Option 3 can be drawn and simple estimators will remain unbiased. It should also be noted that the relevance of NAFORMA I stratification will diminish in the future as it was based in estimates of wood volume and site accessibility that were tied to a particular time point in the past.

Furthermore, a design optimized for maximal precision in the estimate of mean wood volume cannot be expected to be optimal for estimating land-use changes. Simple, equal-probability sampling designs provide a robust basis for dealing with multiple target variables (e.g., both volume and land use) that can vary in quite different manners. In the current sample survey theory, the trend seems to be towards utilizing auxiliary information rather in the estimation phase than as a basis of sampling design, utilizing the techniques of post-stratification and regression estimation (e.g., Särndal *et al*., 2003). This allows the use of different auxiliary variables for different target parameters.

A1. Relocation of permanent plots for better balancing

The original objective when designing Option 3 was to include all clusters that were established as permanent in NAFORMA I and restrict the sample size to 1 405 clusters. However, it was found that much better balance (equal inclusion probabilities within TFS zones) can be achieved by relocating some permanent clusters, and that the optimal balance was achieved with a relatively small number of relocated permanent clusters (Table A1-1; details in Appendix A4). The difference in balance is substantial. For example, relocation of three permanent clusters, or 4%, more than halves the least squares metric of imbalance.

Table A1-1 Number of clusters (n) in Options 1 (NAFORMA I) and 3 (balanced), the optimized least squares metric of balance (LS, 0=perfect balance; details in Appendix A4) with and without strict restriction of inlcuding all permanent clusters, the number of permanent clusters in NAFORMA I (n_{perm}), and the number of clusters that need to be relocated in the optimal solution (n_{reloc}).

TFS zone	Option 1		Option 3			n_{perm}	n_{reloc}
				LS			
	n	LS	n	all permanent	relocation allowed		
Central	470	0.21	191	0.11	0.03	115	8
Eastern	399	0.36	158	0.16	0.03	95	10
Lake	360	0.25	143	0.07	0.03	86	3
Northern	304	0.30	133	0.11	0.02	80	5
Southern	583	0.33	277	0.08	0.02	167	9
S. Highlands	540	0.40	254	0.07	0.05	153	7
Western	549	0.20	252	0.04	0.01	152	5

Source: FAO, 2022

A2. Cost estimates

Several data analysis steps were develop to anticipate the time cost of measuring each cluster as accurately as possible. The plot- and cluster-locations were taken from the NAFORMA I design. Reported timestamps from NAFORMA I were used to approximate, how long plots took to measure and what were the transition times and speeds within the clusters. These were regressed on land-use and land-cover (LULC) class and terrain slope information to build a statistical model. We calibrated the model using data from NAFORMA I and other data sources from dates close to the measurement time of NAFORMA I.

Land cover (LC) map for URT was estimated by combining NAFORMA I field reports with Landsat 8 satellite remote sensing mosaics (Appendix A5). The LC map was estimated for 2015 and 2020, and these maps were assumed to provide reasonable approximations of LULC classes for NAFORMA I and

NAFORMA II, respectively. A plot was assigned as “deforested”, if the LC had changed from forest or woodland (classes 1-2) to open areas (classes 5-8).

The LULC classes considered in the regression models were:

- 1: Forest
- 2a: Closed Woodland
- 2b: Other Woodland
- 3a: Bushland, thicket
- 3b: Other bushland
- 4: Grassland
- 5: Cultivated land
- 678: Combined class consisting of open lands, water bodies, and built-up areas.

A symmetric model between LULC classes was developed to predict walking time from vehicle to cluster and from plot to plot. For example, walking speed from a class 1 location to class 2b location was assumed to be equal to that from class 2b location to class 1 location.

Terrain slope map was calculated from the URT 30 meters Digital Elevation Model provided online at https://rcmrd.org/ by the Regional Center for Mapping of Resources for Development (RCMRD) (accessed 8.2.2022). For the walking speed modeling, the from-slope and to-slope were averaged to avoid asymmetric treatment of the from- and to-plot, as explained above for the LULC grouping.

Accessibility of a cluster was assessed by computing the distance of each plot from a usable road and using the smallest distance as a “cluster distance”. As an approximation of an usable road, we used the OpenStreetMap planet exports for URT dated 01-01-2014 and 16-02-2022, representing NAFORMA I and NAFORMA II, respectively. We considered Highway-key with values “motorway”, “trunk”, “primary”, “secondary”, “tertiary”, “track” and “road”.

For the plot measurement times, we regressed the NAFORMA I time with LULC, slope, and NAFORMA I stratum (to include possible extra information used in NAFORMA I design). To predict plot to plot transition times, we fitted a plot-to-plot walking speed model by regressing the computed average speed between two plots, given simply by time / distance, against the to-from LULC, to-from average slope, and stratum.

The above regression models were then used to predict the time it took to measure each plot in a NAFORMA II cluster and the transition times between plots. In addition, we used the transition speed model to predict the time it takes for the field team to traverse from a vehicle to the nearest plot in a cluster. The vehicle location predictor variables were set to LC=678 and slope=0.

Finally, the predictions of time components for each cluster in NAFORMA II designs were added together to get an estimate of the cluster’s labour requirement in hours. This was converted to labour-days by assuming an 8h working day and then further truncating the result between 0.5 and 3 labour-days.

A3. Supporting statistics for stratifying by TFS zones

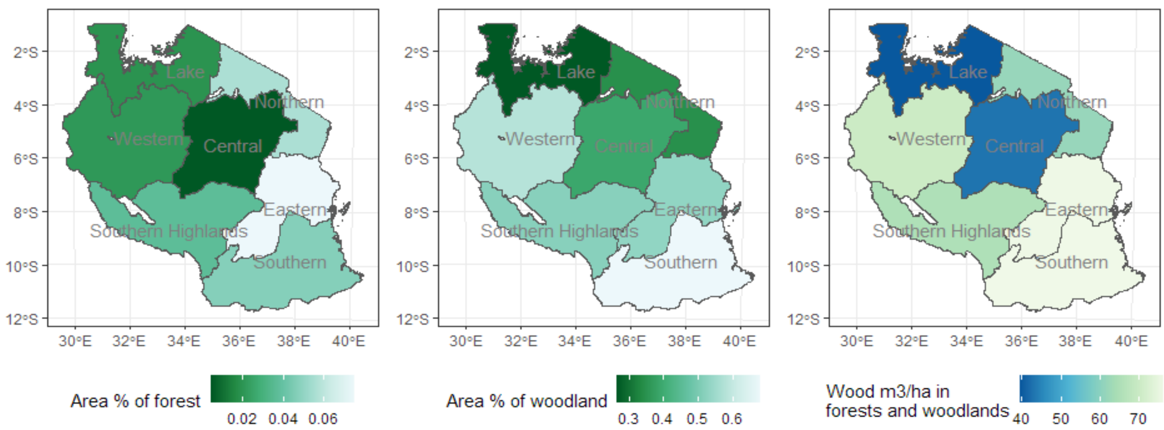
For the stratification to provide statistical efficiency over simple random sampling, the strata should differ in the parameters of interest. Table A3-1 provides some evidence of the TFS zones being, on average, diverse in terms of the amount of forest and woodland, and the mean wood volume. The values are also illustrated in Figure A3-1.

Table A3-1 NAFORMA I Phase 1 and 2 cluster counts and estimates of area proportions and and mean wood volume by TFS zones.

Zone	Phase 1	Phase 2	Phase 2/ Phase 1, %	Area % Forest	Area % Woodland	m^3/ha Forest	m^3/ha Forest & Woodland
Central	5 457	462	8.5	0.4	38.9	167.5	42.6
Eastern	4 164	388	9.3	8.4	53.0	116.4	75.7
Lake	4 126	337	8.2	1.9	26.8	44.1	39.0
Northern	3 083	294	9.5	5.6	34.4	196.2	61.1
Southern	5 859	575	9.8	4.7	68.0	117.9	76.0
S. Highlands	5 576	528	9.5	3.6	51.8	80.7	65.0
Western	6 981	547	7.8	2.0	57.9	99.1	69.9

Source: FAO, 2022

Figure A3-1 Proportion of forest and woodland, and mean wood volume in TFS zones.



Sources: Administrative boundary: National Geospatial Information Authority Surveys and Mapping Division, Ministry of Lands and Human, Settlements Development P.O.Box 9201 Dar-es-Salaam, United Republic of Tanzania (the) http://www.ardhi.gov. [Cited 30 June 2022]. Maps conform to UN. 2006. Map No. 3667, Rev. 6. https://www.un.org/geospatial/content/united-republic-tanzania

A4. Option 3 sampling details

Option 3 suggests sampling field plot clusters of NAFORMA I so that the NAFORMA I stratification is balanced into a new stratification given by uniform sampling within TFS zones. This requires carefully sampling specific numbers of clusters of NAFORMA I within each zone with different proportion per NAFORMA I stratum.

First, the target sample size (number of clusters) of the design was determined. According to TFS, a suitable addition would be approximately 83 new clusters per zone on average. To allow for some relocations of old permanent clusters, $7 \cdot 80 = 560$ clusters were distributed to the zones according to the proportions of clusters in zones in NAFORMA I (see Table A3-1). The numbers of clusters per zone are given at the bottom row of Table A4-1.

Within a zone, the calculations for balancing the NAFORMA I stratification were as follows: First, the NAFORMA I Phase 1 cluster counts per strata were computed and normalized, as these represent the area fractions, say f_0. This distribution was multiplied by the target count of clusters, say n_0, for the zone, providing us the target proportional distribution of clusters per strata, $f_0 \cdot n_0$. We then needed to find the closest distribution to $f_0 \cdot n_0$, or spread n_0 clusters into 18 strata, with the constraints 1) No stratum could have more clusters than in NAFORMA I, 2) each stratum should have at least 50% of NAFORMA I permanent clusters. We chose least squares metric so that the optimization was simplified into a quadratic optimization problem. The R-package *quadprog* was used for the optimizations. Table A4-1 provides the resulting balancing distributions of clusters within zone and strata based on the target total count of 1 405 clusters. Figure A4-1 illustrates the different distributions involved for the Southern zone.

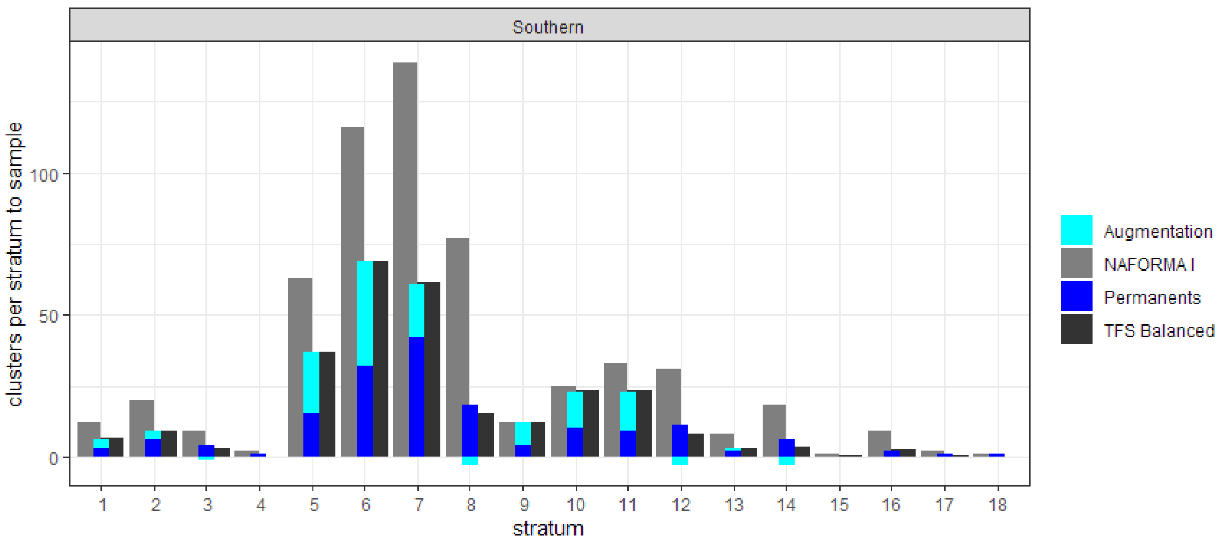
Within stratum and zone, the clusters needed on top of the permanent ones can then be sampled from the temporary clusters of NAFORMA I. A technique for sampling clusters that are well spaced over the landscape is the local pivotal method (LPM; Grafström *et al.*, 2012). LPM is a technique for sampling a target number of units from a population (here: NAFORMA I clusters) with given inclusion probabilities, and at the same time making sure the units are well spread out in space. Illustration of the Option 3 inclusion probabilities according to Table A4-1, and one possible selection of clusters in the Southern zone is given in Figure A4-2. Note that when a stratum was to lose some permanent clusters, the remaining were also sampled using LPM.

Table A4-1 Option 3 sampling of clusters per zone and stratum. Number of clusters to aim for per zone and stratum / number of those from NAFORMA I permanent clusters.

Stratum	Central	Eastern	Lake	Northern	Southern	Southern Highlands	Western
1	17 / 9	9 / 4	22 / 10	23 / 12	7 / 3	19 / 8	19 / 11
2	1 / 1	3 / 3	3 / 3	2 / 1	9 / 6	5 / 2	2 / 2
3	0 / 0	2 / 2	1 / 0	1 / 1	3 / 3	2 / 0	0 / 0
4	0 / 0	1 / 1	2 / 2	2 / 2	1 / 1	1 / 1	0 / 0
5	74 / 40	21 / 6	57 / 30	39 /19	37 / 15	56 / 31	56 / 33
6	41 / 19	31 / 14	18 / 13	19 /11	68 / 32	42 / 28	70 / 39
7	17 / 13	20 / 13	8 / 5	11 / 5	61 / 42	23 / 16	29 / 24
8	4 / 4	5 / 5	3 / 3	2 / 1	16 / 16	6 / 6	2 / 2
9	14 / 5	12 / 4	12 / 6	9 / 4	12 / 4	10 / 2	10 / 5
10	9 / 4	16 / 5	6 / 2	3 / 1	23 / 10	22 / 7	35 / 14
11	4 / 2	11 / 7	4 / 2	3 / 3	23 / 9	18 / 5	20 / 10
12	2 / 2	6 / 5	2 / 2	2 / 1	8 / 8	6 / 6	2 / 2
13	2 / 2	3 / 3	3 / 3	6 / 6	3 / 2	13 / 10	1 / 1
14	3 / 3	6 / 6	1 / 1	4 / 4	3 / 3	15 / 15	2 / 2
15	1 / 1	2 / 0	0 / 0	0 / 0	0 / 0	2 / 0	0 / 0
16	1 / 1	8 / 4	1 / 1	2 / 2	2 / 2	8 / 8	2 / 2
17	1 / 1	2 / 2	0 / 0	2 / 1	1 / 1	3 / 0	0 / 0
18	0 / 0	1 / 1	0 / 0	1 / 1	1 / 1	2 / 1	0 / 0
All	191/107	159/85	143/83	131/75	278/158	253 /146	250/147

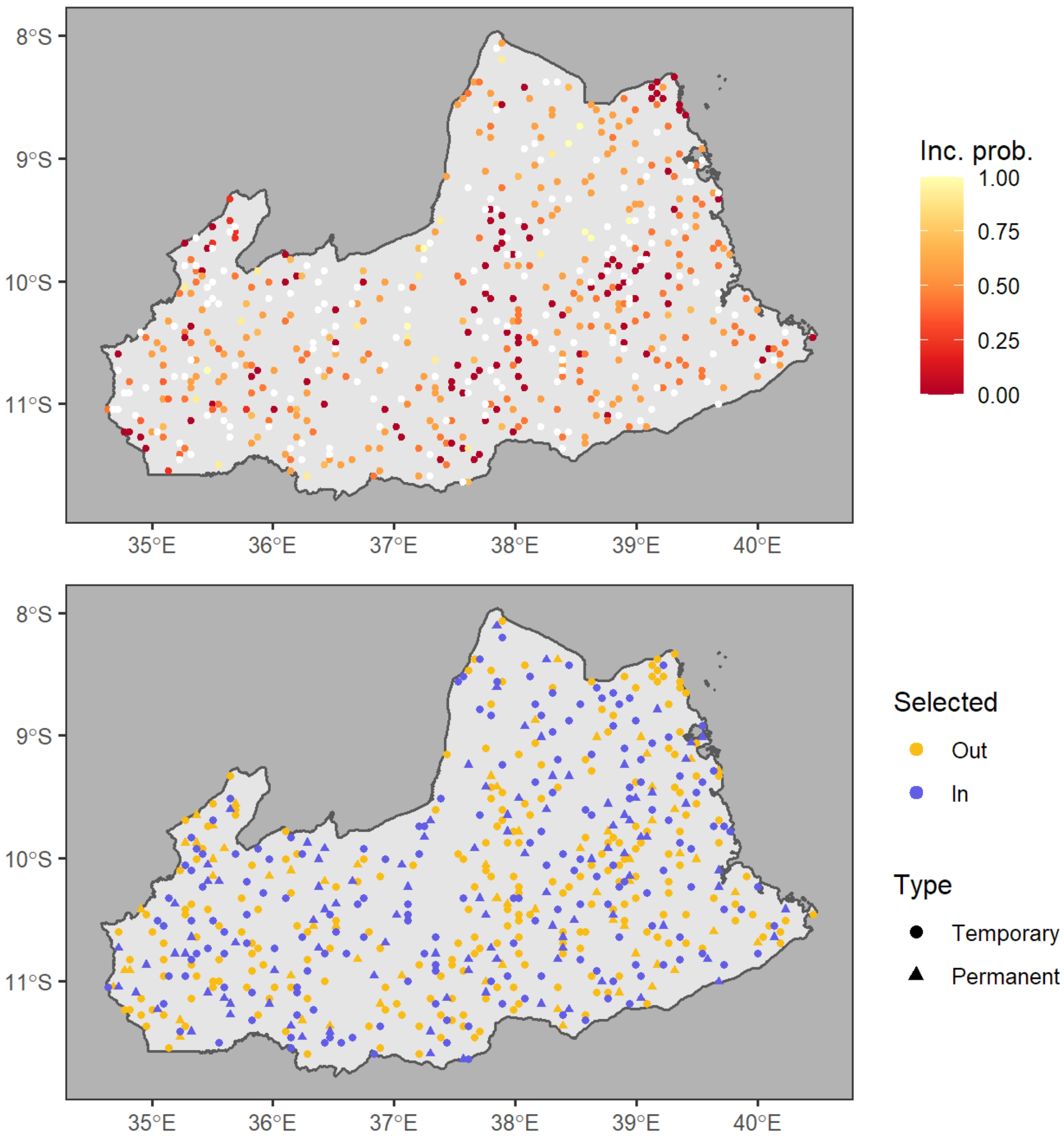
Source: FAO, 2022

Figure A4-1 Cluster count distributions over NAFORMA I strata in one TFS zone. The target total count is the number of permanent clusters + approximately 80. 'TFS Balanced' refers to the number of clusters needed to achive the targeted balance. Design option 3 includes additional clusters ('Augmentation') to be added on top of the permanent ones ('Permanents'). Note that for strata 3, 8, 12 and 14 the strategy requires relocating some permanent clusters.



Source: FAO, 2022

Figure A4-2 Southern TFS zone clusters of NAFORMA I. Above: Probabilities to be included in NAFORMA II according to design Option 3, calculated based on stratum membership and the optimal cluster distribution that balances the sample. Below: Example set of clusters to re-measure, sampled with the local pivotal method.



Sources: Administrative boundary: National Geospatial Information Authority Surveys and Mapping Division, Ministry of Lands and Human, Settlements Development P.O.Box 9201 Dar-es-Salaam, United Republic of Tanzania (the) http://www.ardhi.gov. [Cited 30 June 2022]. Maps conform to UN. 2006. Map No. 3667, Rev. 6. https://www.un.org/geospatial/content/united-republic-tanzania

A5. Land cover classification

URT Land Cover (LC) and Land Cover change is fundamental in the study of sampling design for National Forest inventory. Remote sensing (RS) data have been one of the most important data sources for studies of LC spatial and temporal changes. In fact, multi-temporal RS datasets allow to map and identify landscape changes, giving an effective effort to sustainable landscape planning and management. In particular, by means of the integration of RS and GIS techniques, it is possible to analyze and classify the changing pattern of LC during a long time period and, as a result, to understand the changes within the area of interest

There are both supervised (uses human guidance including training data) and unsupervised (no human guidance) classification methods. The random forest classifier in the System for Earth Observation Data Access, Processing, and Analysis for Land Monitoring (SEPAL) platform used here, uses training data and is thus a supervised classification method. The random forest algorithm creates numerous decision trees for each pixel. Each of these decision trees votes on how each pixel should be classified. The land cover class that receives the most votes is then assigned as the map class for that pixel. Random forests are efficient on large data and accurate when compared to other classification algorithms.

The analysis was conducted using the cloud-based geospatial processing and modelling tools available from the SEPAL platform of the Food and Agricultural Organization. SEPAL is a big-data processing platform that combines super-computing power, open-source geospatial data processing software and modern geospatial data infrastructures like Google's Earth Engine. SEPAL overcomes barriers of poor internet connections and low computing power or storage space on local computers and can also connect to and use data and outputs from FAO's free and open-source software tools Open FORIS.

The SEPAL platform is open source, free of charge and can be accessed at https://sepal.io.

Data and methods

The aim of this analysis was to undertake land cover classification and produce a land cover map of URT using the available training data collected during National Forest Monitoring and Assessment (NAFORMA I) from May 2010 to March 2013. As such, the land cover map was used here as a base to improve the sampling design. Using the Land cover layer greatly simplifies the stratification tasks (McRoberts *et al*., 2002)

Accessing, pre-processing and selection of satellite images are very complex tasks. The SEPAL platform was used for faster and simpler access to free image mosaics from cloud presence. It allows users to access powerful cloud-computing resources to query, access and process satellite data quickly and efficiently for creating advanced analyses.

2015 and 2020 Landsat 8 mosaic data was used for the Land Cover Map. In order to achieve cloud free imagery, a period from July to October (low rainfall season) was selected and then composited into a resulting single image that contains the 'best' pixels in terms of spectral information. The selected

2015 and 2020 Landsat8 Mosaic was then retrieved as a Google Engine Asset i.e., a geospatial dataset stored into a private user account that can be further used in scripts for classification or export.

The training plot data set collected during 2010 – 2013 NAFORMA I field survey was used to generate land cover of the selected area. Validation of training data quality is very important to assess the reliability of the training data before classification. The technique used for quality check was based on the detection of outlier pixels for each land cover class. Before removing outliers, training data was split according to the time data was collected i.e., 2010, 2011 and 2013. Then, outliers were removed based on NDVI data extracted from Google Earth Engine (GEE; Arvor *et al*., 2008) A total number of NDVI points of 8759, 33007, 15214 and 3483 was collected for each land cover for 2010, 2011,2012 and 2013 respectively and used for removing outliers using R. Results were exported as CSV files and used for classification in SEPAL.

Results

The Land Cover Map layouts generated from the Landsat8 Images data set for 2015 and 2020 are displayed in Figure A5-1 and Figure A5-2, respectively. According to the result the largest category was open woodland that accounts for 27% to 22% of the total land in 2015 and 2020, respectively (Table A5-1). Also the most substantial changes in area are observed for the open woodlands, followed by scattered croplands and open bushland.

Accuracy assessment

The aim of an accuracy assessment is to quantitatively assess how effectively the pixels were sampled into the correct land cover classes. The emphasis for accuracy assessment pixel selection was on areas that could be clearly identified on both Landsat 8 high resolution image and reference data (field data)

The results from the accuracy assessment showed an overall accuracy of 64%, User's accuracy ranged from 44.44% to 100.0% while producer's accuracy ranged from 28.57% to 99.2% among the land cover classes. Coastal bare soil, inland water, grassland scattered cultivation and mangrove forest were found to be reliable with user's accuracy ranging from 86% to 100%. Cultivated land with mixed tree cropping had low producers' accuracy (28%) due to misclassification.

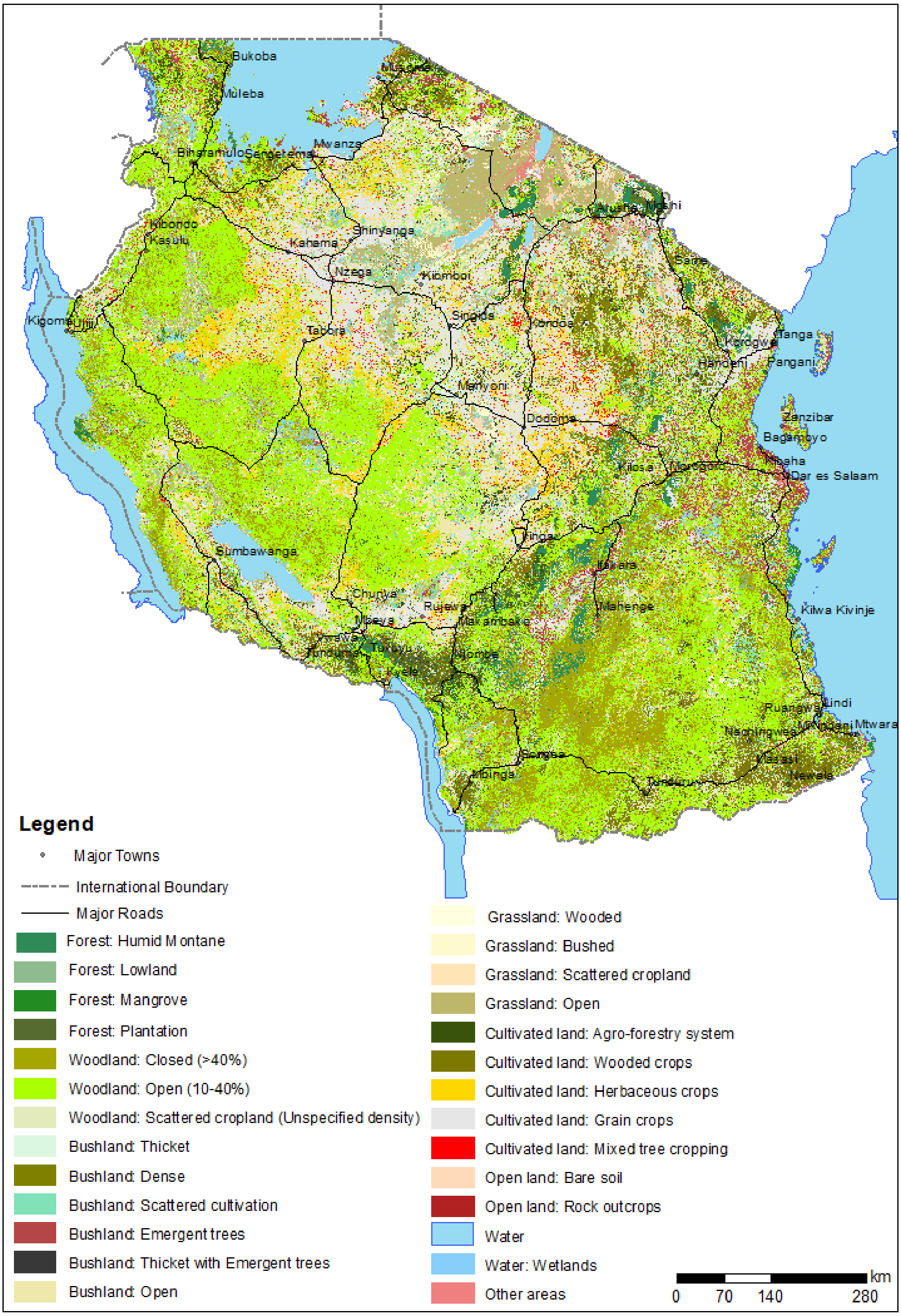
Conclusion

When performing a supervised classification, it is essential to acquire reliable ground truth data. Constraints on these data can be different depending on the classifier used. A validation step is necessary to ensure the reliability of the collected data and eventually determine an optimal subsample, which improves the classification result.

In this study, remote sensing and GIS were integrated for quantifying and understanding the LULC changes in URT from 2015 to 2020 with the aim of improving the sampling design. SEPAL (an open tool platform) and the Landsat 8 data set were used to determine the land cover change.

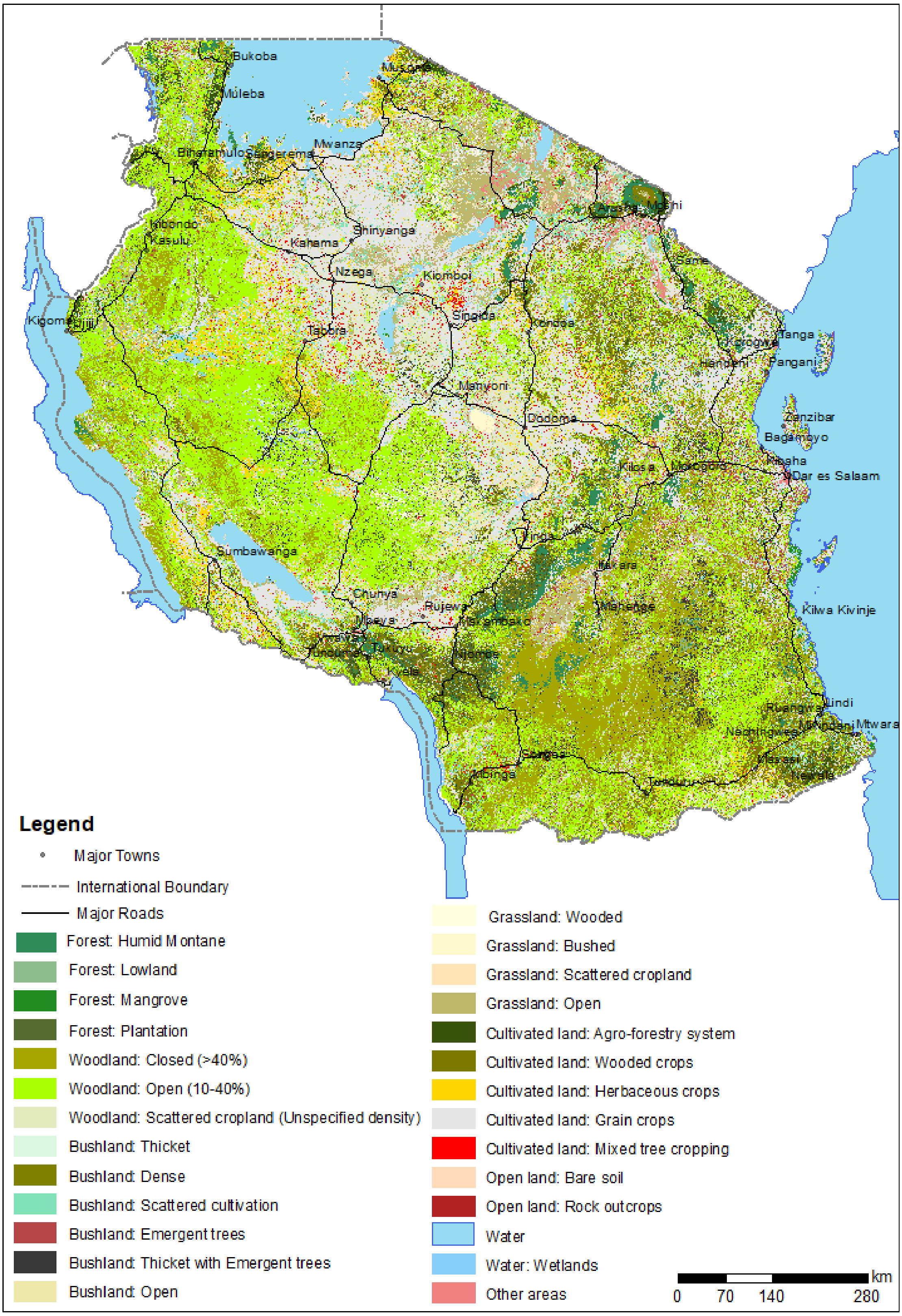
Accuracy assessment was performed using an error matrix. The study had an overall classification accuracy of 64%. The inconsistency of the map accuracy might depend on some factors such as spatial resolution, the homogeneity or heterogeneity of land surface (Ma *et al.,* 2017) or classification algorithms (Gómez *et al.,* 2016). Among the factors, the number of land cover classes is crucial: map accuracy can decrease when the number of classes (e.g., water, urban, forest) increases, because of more chances of misclassifications among classes. For this study 27 classes were identified and classified.

Figure A5-1 2015 Land Cover Map



Sources: Administrative boundary: National Geospatial Information Authority Surveys and Mapping Division, Ministry of Lands and Human, Settlements Development P.O.Box 9201 Dar-es-Salaam, United Republic of Tanzania (the) http://www.ardhi.gov. [Cited 30 June 2022]. Maps conform to UN. 2006. Map No. 3667, Rev. 6. https://www.un.org/geospatial/content/united-republic-tanzania

Figure A5-2 2020 Land Cover Map



Sources: Administrative boundary: National Geospatial Information Authority Surveys and Mapping Division, Ministry of Lands and Human, Settlements Development P.O.Box 9201 Dar-es-Salaam, United Republic of Tanzania (the) http://www.ardhi.gov. [Cited 30 June 2022]. Maps conform to UN. 2006. Map No. 3667, Rev. 6. https://www.un.org/geospatial/content/united-republic-tanzania

Table A5-1 Land Cover classes distribution, United Republic of Tanzania

Vegetation Type	**Area**			
	2015		**2020**	
	ha	**%**	**ha**	**%**
Humid Montane and Lowland Forest	1 741 426	2	1 669 691	2
Mangrove Forest	352 574	0	337 036	0
Plantation Forest	466 471	0	762 764	1
Woodland: Closed (>40%)	9 288 203	10	11 902 119	12
Woodland: Open (10-40%)	25 280 966	27	21 265 268	22
Woodland: Scattered cropland (Unspecified density)	4 648 557	5	3 042 707	3
Bushland: Thicket	689 582	1	1 857 781	2
Bushland: Dense	2 952 694	3	3 180 821	3
Bushland: Scattered cultivation	2 704 485	3	3 042 745	3
Bushland: Emergent trees	2 406 501	3	1 449 253	2
Bushland: Thicket with Emergent trees	928 049	1	1 637 824	2
Bushland: Open	7 006 451	7	3 869 661	4
Grassland: Wooded	2 761 909	3	838 662	1
Grassland: Bushed	457 577	0	270 072	0
Grassland: Scattered cropland	412 113	0	299 641	0
Grassland: Open	4 275 245	4	4 978 460	5
Cultivated land: Agro-forestry system	1 728 629	2	1 884 151	2
Cultivated land: Wooded crops	3 924 106	4	3 736 039	4
Cultivated land: Herbaceous crops	4 618 392	5	3 872 378	4
Cultivated land: Grain crops	9 396 828	10	17 074 785	18
Cultivated land: Mixed tree cropping	1 490 335	2	832 756	1
Open land: Bare soil	457 726	0	144 945	0
Open land: Coastal bare land	5641.47	0	1 821	0
Open land: Rock outcrops	296	0	1 981	0
Water: Inland water	5 942 318	6	6 374 403	7
Water: Wetlands	904 904	1	530 066	1
Other areas	357 759	0	437 730	0
Total	95 199 739	100	95295561	100

Source: FAO, 2022

A6. Testing re-measurement of NAFORMA I clusters

The first-ever comprehensive National Forest Inventory (NFI) in URT popularly known as National Forestry Resources Monitoring and Assessment (NAFORMA) was implemented from 2009 and finalized in 2014 (Ministry of Natural Resources and Tourism., 2015). The NAFORMA results were needed to support national policy processes for the enhancement of sustainable forest management and at the same time addressing issues related to Reducing Emissions from Deforestation and Forest Degradation (REDD+) and the Greenhouse Gas inventory as international reporting obligations. It was observed that different stakeholders, even outside the forest sector, have been using the NAFORMA results. It is ten years since the first assessment was accomplished, hence leading to a need to re-assess again the status of forest resources country-wide.

TFS in collaboration with the F are in the process of preparing a phase two of the National Forest Resources Monitoring and Assessment (NAFORMA II). The purpose of NAFORMA II is to update information of NAFORMA I by developing a new design which will improve both the sampling strategy and the methodological approach.

In the process, TFS was assigned to test the proposed approach of collecting the biophysical data in selected clusters. The fieldwork involved cluster and plots identification and re-location, and assessing the time taken to measure both a plot and a cluster. Five clusters were selected, out of which four are permanent ones and the remaining one is a temporary cluster. To execute properly the fieldwork all necessary procedures were taken into account. They included: special training for the field team (on both theoretical and practical aspects), report to the nearby village/ward/district offices to introduce the team to the local administration and conducting a reconnaissance survey.

Figure A6-1 Field practical training

NAFORMA I cluster relocation

In the field, the team employed a variety of methods to ensure it reached the intended cluster plots with precision. Among the methods employed are

- the use of NAFORMA filled forms of the selected clusters;
- where available, the use of people in that locality, who were involved in the NAFORMA I exercise;
- metal detector and the use of high precision GPS;

Through these means, the team re-located the cluster easily. For the permanent clusters where the plot centers had been marked with a permanent iron bar during NAFORMA I, the plot centers were easily re-located with the use of high precision GPS and the metal detector. For the temporary cluster; the use of NAFORMA I's field form information and applied local knowledge helped the team to reach and identify the cluster and its plots. The team also was able to identify the same natural and special features of the clusters and plots measured during the NAFORMA first phase. Those features include big trees, seasonal streams and tree species.

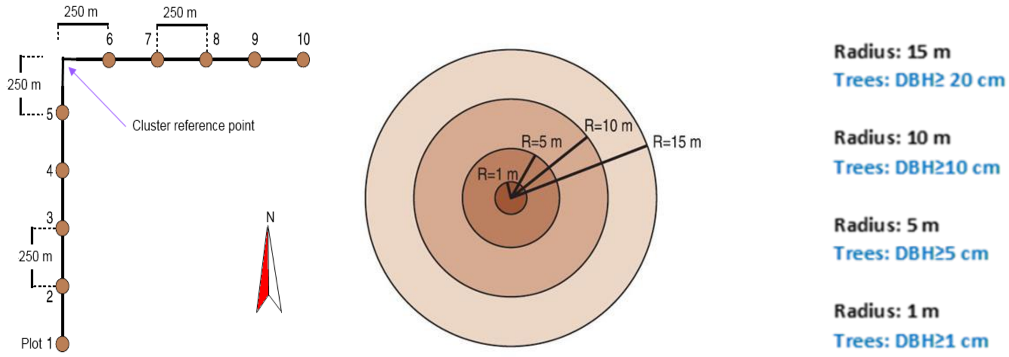
After plot relocation the cluster, coordinates of the center of each plot within were taken using high-precision GPS and recorded. Additional information of the plot, like forward and back bearings towards permanent objects such as big trees or rocks close to the center were also recorded, and such objects were sprayed with red paint.

Field measurement

Fieldwork adopted the plot design of NAFORMA I. That is, concentric circular plots of 1m, 5m, 10m and 15m radii were established as shown in Figure A6-2.

The field team ensured that the plot centre was mapped by referring, if available, the permanent features recorded during NAFORMA I. Tree parameters were then collected in clockwise fashion starting from the North direction. The use of a 1.3m DBH (tree diameter at breast height) stick was emphasized. Tree parameters recorded included tree number, stem number, species code, species name, DBH, health, tree origin, stump height, stump diameter, total height, bole height, and whether the tree was marked with paints plus other remarks. Dead wood and stumps with stump diameter > 10 cm within a plot radius of 15m were also recorded.

Figure A6-2 Cluster and plot design



Source: FAO, 2022

Data recording

During this testing, the team applied two means of data recording. Data recording was through android tablet and hardcopy field forms. The aim was to compare the methods and provide feedback that will improve the NAFORMA II data collection and recording approaches.

Figure A6-3 Data collection and recording

Observations

During cluster relocation, it was observed that the use of High precision GPS is inevitable as it facilitated the team to easily reach the plot centre compared to hand held Garmin GPS. Likewise, using NAFORMA I filled forms and people who participated in the NAFORMA I exercise was also contributed to facilitate the relocation.

During data recording, it was observed that less time is taken and it is more cost-effective when using tablets than when using hardcopy field forms.

Conclusion and recommendations

Conducting NAFORMA II as an assessment for monitoring forest resources status requires the accurate relocation of those plot samples established during the NAFORMA I that will be selected in NAFORMA II. TFS conducted field testing during 2021 and discovered that both types of clusters can be re-visited accurately using information gathered during NAFORMA I.

NAFORMA II data collection approaches should adhere to the recommendations provided in the full field report.

References

Arvor, D., Jonathan, M., Meirelles, P., & Dubreuil, V. 2008. *"*Detecting outliers and asserting consistency in agriculture ground truth information by using temporal VI data from Modis." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. https://www.isprs.org/proceedings/XXXVII/congress/7_pdf/5_WG-VII-5/53.pdf

FAO 2017. "Voluntary guidelines on national forest monitoring." http://www.fao.org/publications

Gómez, C., White, J. C. & Wulder, M. A. 2016. "Optical remotely sensed time series data for land cover classification: A review". ISPRS Journal of Photogrammetry and Remote Sensing, 116, 55–72. https://doi.org/10.1016/J.ISPRSJPRS.2016.03.008

Grafström, A., Lundström, N. L. P. & Schelin, L. 2012. "Spatially Balanced Sampling through the Pivotal Method". Biometrics, 68: 514-520. https://doi.org/10.1111/j.1541-0420.2011.01699.x

Kangas, A., Vainikainen, N. & Pekkarinen, A. 2015. "Options for monitoring permanent NAFORMA sample sites". FAO-Finland Forestry Programme: Sustainable Management in a Changing Climate' GCP/GLO/194/MUL. Unpublished.

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., & Liu, Y. 2017. "A review of supervised object-based land-cover image classification". *ISPRS Journal of Photogrammetry and Remote Sensing*, *130*, 277–293. https://doi.org/10.1016/J.ISPRSJPRS.2017.06.001

McRoberts, R. E., Wendt, D. G., Nelson, M. D., & Hansen, M. H. 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sensing of Environment, 81(1), 36–44. https://doi.org/10.1016/S0034-4257(01)00330-3

Ministry of Natural Resources and Tourism. 2015. "NAFORMA: Main Results." https://www.tfs.go.tz/uploads/NAFORMA_REPORT.pdf

Särndal, C.-E., Swensson, B. & Wretman, J. 2003. "Model Assisted Survey Sampling". Springer.

Tanzania Forest Services Agency. 2020. "Strategic Plan 2021/2022 – 2025/2026".

UN. 2006. Map No. 3667, Rev. 6. *Geospatial, location information for a better world.* [Cited 30 June 2022]. https://www.un.org/geospatial/content/united-republic-tanzania

Tomppo, E. 2009. "The Finnish multisource national forest inventory: small-area estimation and map production." https://www.fs.usda.gov/treesearch/pubs/17326

Tomppo, E., Heikkinen, J., Henttonen, H.M., Ihalainen, A., Katila, M., Mäkelä, H., Tuomainen, T., Vainikainen, N., 2011. "Designing and Conducting a Forest Inventory - Case: 9th National Forest Inventory of Finland, Managing Forest Ecosystems". Springer

Tomppo, E., Katila, M., Mäkisara, K., Peräsaari, J., Malimbwi, R., Chamuya, N., Otieno, J., Dalsgaard, S. & Leppänen, M. 2010. "A Report to the Food and Agriculture Organization of the United Nationsited Nations (FAO) in Support of Sampling Study for National Forestry Resources Monitoring and Assessment (NAFORMA) in United Republic of Tanzania,". http://www.suaire.sua.ac.tz/handle/123456789/1296

Tomppo, E., Malimbwi, R., Katila, M., Mäkisara, K., Henttonen, H. M., Chamuya, N:, Zahabu, E. & Otieno, J. 2014. "A Sampling Design for a Large Area Forest Inventory: Case United Republic of Tanzania." *Canadian Journal of Forest Research* 44: 931–48. https://doi.org/10.1139/cjfr-2013-0490.

Contact

FAO Representation in United Republic of Tanzania
https://www.fao.org/countryprofiles/index/en/?lang=en&ISO3=TZA

FAO Geospatial Unit at Land and Water Division (NSL)
https://www.fao.org/geospatial/en/

FAO Forestry Division
https://www.fao.org/forestry/en/

Food and Agriculture Organization of the United Nations