|  |
| --- |
| STANDARD OPERATING PROCEDURES FOR SAMPLED-BASED AREA ESTIMATION |
| Purpose: These Standard Operating Procedures (SPO) Notes acts as a set of general Cheat-sheets, which in conjunction with a more general posterior white paper, aim to be the main support backbone documents for national staff working in area reporting for REDD+. The SPO will be modified to fit the particular protocols of each country and ensure that they are repeatable. |
| Important Notes for Completing this DocumentUnder most headings in this template are instructions for completing each section; they are marked in this way: << instructions for completing the section >>. Please read through the instructions for details on minimum requirements.Important note: No sections are to be deleted from this document.Text contained within << >> provides information on how to complete that section and can be deleted once the section has been completed. |

Table of Contents

[Standard Operating Procedure 4 (Sop4): Sample-Based Area Estimation Analysis 3](#_Toc31088431)

[Section 1 Overview 3](#_Toc31088432)

[Section 2 Procedure 3](#_Toc31088433)

[Section 3 Quality management 5](#_Toc31088434)

|  |
| --- |
| Standard Operating Procedure 4 (Sop4): Sample-Based Area Estimation Analysis |
| Section 1 Overview |
| Purpose | Derive area estimates and their uncertainties through the combined use of reference data and maps (i.e., sample-based area estimation). Specifically, create an error matrix and construct estimators of area with confidence intervals; and calculate accuracy measures |
| Scope | Wall-to-wall, geospatial data supported by a reference sample |
| Responsibilities | Technical decision and policy makers in REDD+ countries and international agencies, multilateral and bilateral programmes, in particular for the estimation of activity data (i.e., land use/land cover change) for emissions/removals in REDD+ activities. |
| Prerequisites | * Map production (SOP0)
* Sampling design (SOP1)
* Response design (SOP2)
* Data collection (SOP3)
 |
| Requirements | * Interpreted reference data for all samples available
* Map data for all samples available
* Reference and map class labels are the same
* Area of all strata in map.
* Confidence level for the estimation of uncertainties

Important Note:  |
| Section 2 Procedure |
| The error matrix | The current SOP procedure focuses on the basis that a stratified random sampling has been performed. Yet, it is a general case that reveals the same estimators as those used when one has performed either systematic or simple random sampling (see end of subsection *Estimating areas and their uncertainty*). The aim of the stratified area estimation is to eliminate bias as far as can be judged and reduce uncertainties as far as practicable. To attain both goals, reference data must be compared with map information. The reference data (from a probability sampling design) adjust for estimated systematic classification errors and estimating uncertainty, making use of the previously seen (see SOP1) stratified estimators.The error matrix contrasts reference and map data and constitutes the main tool for the area analysis. First, one can construct a matrix where reference data (sample points class labels taken as ground truth) are represented in columns and map data (sample point class labels according to the available map) in rows. Let us assume three change classes in our map and reference data. Although the objective is generalizable to any area estimation, let us also assume that we aim to detect area change. Our *H=3* classes might be, in this case, *Forest* (stable), (Forest) *Loss* and *Non-Forest* (stable). Two basic steps in building error matrices are considered:1. First, an error matrix of sample counts of reference and map labels is built (Table SOP4.1). Along the diagonal, elements indicate matching labels (the map correctly describes the ground truth or reference), while off-diagonal elements in the tables below represent map omissions (top down direction) and commissions (left-right direction).
2. An extra column will calculate strata weights dividing the area of each class or stratum by the total area in the reporting land (i.e., country, province,..)

**Table SOP4.1. Error matrix of sample counts.**

|  | Reference data | Area map | Strata weight (*wh*) |
| --- | --- | --- | --- |
| Map data | Deforestation | Stable forest | Stable non-forest |
| High probability of deforestation  | *n11* | *n12* | *n13* | *n1.* | *a1.* |
| 40 m Buffer  | *n21* | *n22* | *n23* | *n2.* | *a2.* |
| Low probability of deforestation | *n31* | *n32* | *n33* | *n3.* | *a3.* |
| Stable forest | *n41* | *n42* | *n43* | *n4.* | *a4.* |
| Stable non-forest | *n51* | *n52* | *n53* | *n5.* | *a5.* |
| New deforestation | *n61* | *n62* | *n63* | *n6.* | *a6.* |
| Total  | *n.1* | *n.2* | *n.3* | *n* | *a* |

1. To translate sample counts into actual estimated area proportions per class, it is necessary to operate on the colored cells. For each cell area proportion is defined as

$$\hat{p}\_{hj}=w\_{h}∙ \frac{n\_{hj}}{n\_{h∙}}$$where h and j stand for row and column, respectively. Accuracy measurements, though not relevant for the current SOP, can also be calculated from dividing individual cell area proportions by the respective map (user’s accuracy) or reference (producer’s accuracy) class area proportions. **Table SOP4.2. Error matrix of area proportions.**

|  | Reference data | Total  | User’saccuracy (*Û*𝑖) |
| --- | --- | --- | --- |
| Map data | Deforestation | Stable forest | Stable non-forest |
| High probability of deforestation  | $\hat{p}$*11* | $\hat{p}$*12* | $\hat{p}$*13* | $\hat{p}$*1.*  | $\hat{p}$*11/*$\hat{p}$*1.* |
| 40 m Buffer  | $\hat{p}$*21* | $\hat{p}$*22* | $\hat{p}$*23* | $\hat{p}$*2.*  | $\hat{p}$*22/*$\hat{p}$*2.* |
| Low probability of deforestation | $\hat{p}$*31* | $\hat{p}$*32* | $\hat{p}$*33* | $\hat{p}$*3.*  | $\hat{p}$*33/*$\hat{p}$*3.* |
| Stable forest | $\hat{p}$*41* | $\hat{p}$*42* | $\hat{p}$*43* | $\hat{p}$*4.*  | $\hat{p}$*44/*$\hat{p}$*4.* |
| Stable non-forest | $\hat{p}$*51* | $\hat{p}$*52* | $\hat{p}$*53* | $\hat{p}$*5.*  | $\hat{p}$*55/*$\hat{p}$*5.* |
| New deforestation | $\hat{p}$*61* | $\hat{p}$*62* | $\hat{p}$*63* | $\hat{p}$*6.*  | $\hat{p}$*66/*$\hat{p}$*6.* |
| Total  | $\hat{p}$***.1*** | $\hat{p}$***.2*** | $\hat{p}$***.3*** | *1*  |  |
| Producer’s accuracy (*P𝑖*)  | $\hat{p}$*11/*$\hat{p}$*.1* | $\hat{p}\_{22}$*/*$\hat{p}$*.2* | $\hat{p}$*33/*$\hat{p}$*.3* |   | 𝑂𝑣𝑒𝑟𝑎𝑙𝑙 𝑎𝑐𝑐𝑢𝑟𝑎𝑐𝑦 (*Ô* )= $\hat{p}$*11 +* $\hat{p}$*22 +* $\hat{p}$*33* |

 |
| Estimating areas and their uncertainty | The mean estimator for the area of each class can be directly obtained from Table SOP4.2. While one would be tempted to use map class area proportions ($\hat{p}$*h.*), unbiased stratified estimators are provided using reference class area proportions ($\hat{p}$*.j*). As implied in Table SOP4.2, these follow:$$\hat{p}\_{∙j}=\sum\_{h=1}^{H}w\_{h}∙ \frac{n\_{hj}}{n\_{h∙}}=\sum\_{h=1}^{H}\hat{p}\_{hj}$$colored as orange in Table 2.Once the estimated reference class area proportions are obtained, the mean total area per class is calculated by multiplying them with the total reporting area *a*:$$\hat{A}\_{j}=\hat{p}\_{∙j}∙a$$Uncertainties in activity data were derived using non-parametric bootstrapping, where reference data points were re-sampled with replacement 100,000 times. For each permutation of reference data points, the bias-corrected area estimates were produced following the methods described in Olofsson et al. (2014). Uncertainty was estimated from the resulting distribution of area estimates. Although more complex to implement, bootstrapping has the advantages of not requiring any assumption about the shape of the probability distribution function of each land cover transition class, and avoids the generation of negative areas in rare classes where a probability distribution function crosses zero. The method was implemented in R, and the scripts used are available in the “*Mozambique ERPA 2018*” shared folder.The impact of using non-parametric bootstrapping to estimate uncertainties vs other methods was tested with a comparison of deforested areas derived from bootstrapping against sampling from a normal distribution with standard error calculated with the methods described in Olofsson *et al*. (2014). For the latter case two uncertainties were derived: one retaining any negative area estimates for rare transition classes, and another setting these to zero. The result (Figure 6) indicates that there is very little difference between any of the methods in either reference or monitoring periods, with the result that any chosen approach would produce equivalent emissions estimates.https://lh5.googleusercontent.com/i5GrCrYxHr4Xc3E-Xh2Z4FWkz1a76yzbcYeujUGCQx3aNwHLRUFAvFhxhEGAezcIjQnUmqMkwC-XQXlakDuoqmh2Ks7FxMKHIMMkUTFQvvem6UjBv2gzZe-8OGgIk0pqqSvd7-QJ **Figure SOP4.1:** Total activity data area estimates for monitoring period using normal distributions for each transition class (red), normal distributions with a minimum area of 0 ha (green), and non-parametric bootstrapping (blue). All three methods result in equivalent uncertainty estimates. |
| Section 3 Quality management |
| Error matrix | * The error matrix, particularly that for sample counts, is usually an effective tool to quickly detect the overall quality of the data collection process. High numbers in off-diagonal elements (commission, or mostly, omission errors), when compared to those in the diagonal, may reflect a poor number of samples overall, imply the need for a revision in data interpretation procedures or response design, or even reflect a poor map quality that may involve a new processing of the map. Often the people responsible may need to go back to review procedures. Although this may theoretically produce biases, it is often a necessary consequence derived by lack of expertise or knowledge in the specific area assessed.
* Report the error matrix in terms of estimated area proportions.
 |
| Area and uncertainty estimation | Set a maximum permissible uncertainty. |
| Data Flow Diagram |  |
| References | Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., Wulder, M. A. 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148:42–57. |

© 2021 by MRV Unit under the Fundo Nacional do Desenvolvimento Sustentável. Standard Operating Procedures for Sampled-Based Area Estimation is made available under a [Creative Commons By 3.0 IGO license](https://creativecommons.org/licenses/by/3.0/igo/): <https://creativecommons.org/license/by/3.0/igo/>

For more information, please visit:

* Mozambique MRV website: <https://www.fnds.gov.mz/mrv/>

**Please cite as**: Unidade de MRV-FNDS. 2020. Standard Operating Procedures for Sampled-Based Area Estimation: Sample-Based Analysis. Fundo Nacional do Desenvolvimento Sustentável (FNDS), Maputo, Mozambique.