Annex 12.1. Estimation of Uncertainty of Activity Data

Mapped areas of land cover (or change in land cover) produced from the satellite interpretation can yield biased estimates of change because of image classification errors. Therefore, we need to estimate the level of accuracy of the land cover change map by comparing the map to a sample of reference observations, incorporating the reference observations into a sample-based estimation (SBE) of Activity Data. Area estimates and accuracy are inferred by analyzing the sample. In this analysis, we assess the accuracy of land cover change from 2006 to 2016 for three broad land cover classes, namely primary forest (PI), degraded primary forest (DP) and non-forest (NF). We then use those Reference Data and compare to the original forest cover maps of cover change, to assess accuracy. The assessment of the accuracy of the land cover map and the resulting sample-based estimation of area was based on Olofsson *et al.* (2014), substituting a modified variance estimator to account for the use of post stratification rather than pre stratification (Olofsson 2019, pers. com.).

Method for the Assessment of Uncertainty of Area Changes

The assessment of land cover changes uncertainty uses reference data that was generated by a collaboration project between Indonesia's Ministry of Environment and Forestry (KLHK) and its National Institute of Aeronautics and Space (LAPAN) together with researchers from the University of Maryland's Global Land Analysis and Discovery Group (GLAD) and the World Resources Institute (WRI) Indonesia. A set of 10,000 30x30m blocks corresponding to time series of Landsat satellite image pixels was selected from throughout the country using a simple random sampling technique. The goal of the project was to support efforts to sustainably manage Indonesia's forest resources by providing nation-wide estimates of primary forest loss in Indonesia over a 26-year period (1991-2016), together with data on land use trajectories following primary forest clearing.

The reference data used for visual interpretation includes annual cloud-free Landsat data composites from 1985-2016. The University of Maryland GLAD group and LAPAN provided composites for the years 1985-2016. LAPAN provided composites for 2015 and 2016 as well as very high resolution true color composites for each pixel based on data collected by the SPOT 6 and 7 satellites from 2013-2016. Graphs of minimum and maximum NDVI values based on cloud-screened Landsat observations were also provided for the years 1985-2014. Very high resolution Google Earth imagery was used where available, primarily to confirm interpretations towards the close of the study period together with SPOT data (Figure 1).

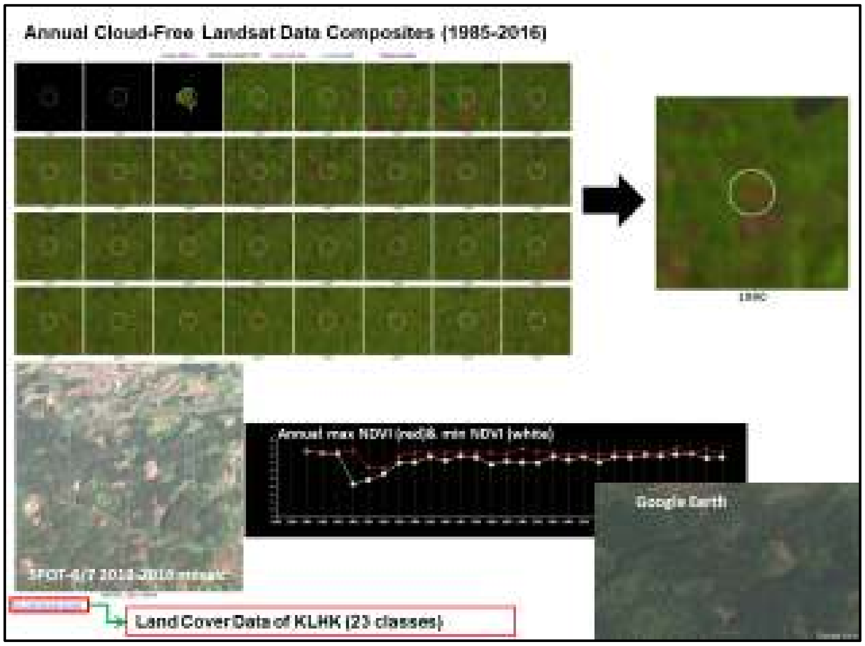

Figure 1 Data Reference

Each sample pixel was reviewed by at least two independent analysts. The first analyst assigned a 1990 land cover class – primary intact, primary degraded or non-primary forest – to each pixel. For primary forest pixels, analysts also evaluated pixels to determine whether any forest disturbance or land cover change events took place, recording the year and type of each change event and their confidence in their interpretations for all recorded changes. Pixels experiencing additional change events were assigned to a "more than three changes" category. A second analyst later reviewed these interpretations, revising the dataset as needed. This was done to control for biases among analysts and reduce the incidence of interpretation errors. The resulting dataset was again compared to existing map products, including KLHK forest/non-forest maps (1990, 2000, and 2012); primary forest maps created by the University of Maryland (2000 and 2001); tree cover maps created by LAPAN as part of the INCAS project (2000 and 2012); and the University of Maryland percent tree cover (2000) and tree cover loss (2001-2016) products. Pixels found to be inconsistent with the above map products, pixels marked as low confidence in one or more confidence category, and all pixels experiencing primary forest clearing during the study period, were reviewed again by the UMD analyst (Figure 2).

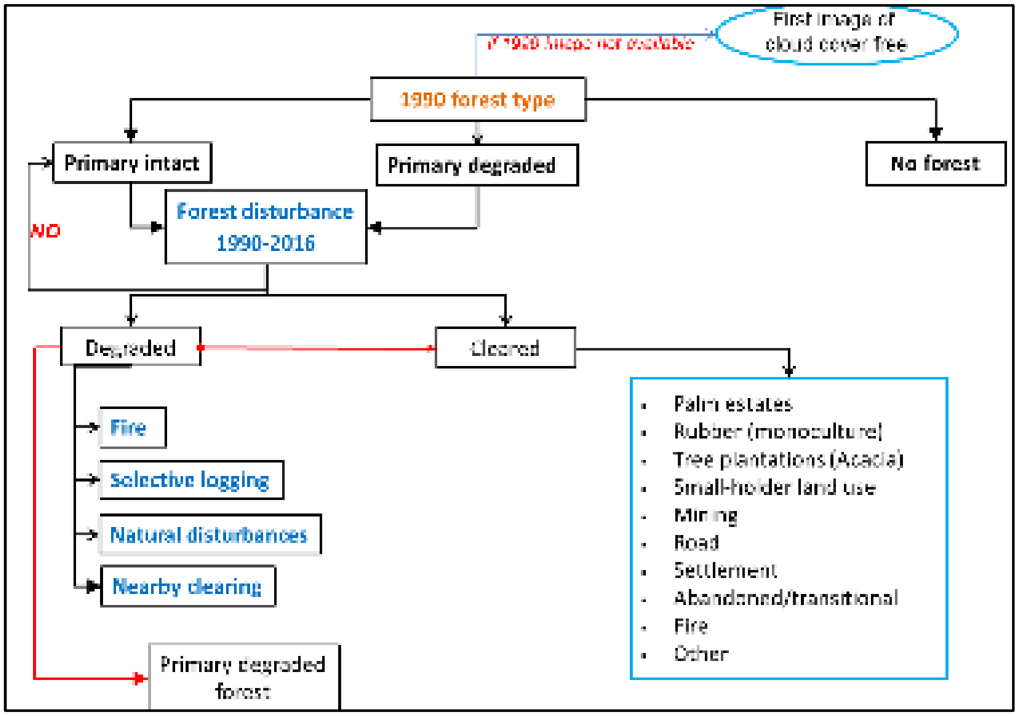

Figure 2 Flowchart of Sample Land Cover Change Interpretation

East Kalimantan has 639 sample units that clipped from 10,000 units of national data, covering three groups of land cover categories, i.e. primary forest (PI), degraded primary forest (DP) and non-forest (NF). Error matrix of the samples was developed in which rows represents map category and column represents reference as illustrated in Table 1.

Table 1 Error matrix of sample counts, nij

		Reference (sample)			
		PI	PD	NF	Grand Total
Map	PI	n_{11}	n_{12}	n_{13}	n_{1}
	PD	n_{21}	n_{22}	n_{23}	n_{2}
	NH	n_{31}	n_{32}	n_{33}	n_{3}
	Grand Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$

Suppose the objective is to assess the accuracy of a map with q categories and to estimate the area of a particular map category. A sample of assessment units (e.g., pixels) is selected by simple random, stratified random (with the map classes as strata), or systematic sampling. A sample error matrix is constructed where the map categories (i=1,2,...,q) are represented by rows and the reference categories (j = 1, 2, ..., q) by columns (Table 1). Note that in some presentations of an error matrix (e.g. Card, 1982), the rows and columns are reversed. The basic principles of the methods outlined below still apply to such situations but accommodation for the switch in row and column contents is required

Table 1 illustrates the common practice of reporting the error matrix in terms of sample counts. A more informative presentation of the error matrix is in terms of the unbiased estimator of the proportion of area in cell *i,j* of the error matrix

$$\hat{p}_{ij} = W_i \frac{n_{ij}}{n_{i\cdot}} \quad \text{...} \ (1)$$

where the total area of the map is A_{tot}, the mapped area of category i is $A_{m,i}$ (subscript m denotes "mapped"), and the proportion of the area mapped as category i is $W_i = A_{m,i} \div A_{tot}$. The error matrix in terms of estimated area proportions is shown in Table 2. An advantage of the presentation given in Table 2 is that accuracy and area estimates can be computed directly from the error matrix.

Table 2 Error matrix of estimated area proportions, p^ij

		Reference (samples)			
		PI	PD	NF	Grand Total
Map	PI	p_{11}	p_{12}	p_{13}	$p_{1.}$
	PD	p_{21}	p_{22}	p_{23}	$p_{2.}$
	NH	p_{31}	p_{32}	p_{33}	$p_{3.}$
	Grand Total	$p_{.1}$	$p_{.2}$	$p_{.3}$	$p_{..}$

Because of classification error, the mapped area proportions given by $A_{m,i} \div A_{tot}$ are usually biased when the objective is to estimate the true proportion of area of category i as determined from the reference classification. Instead of obtaining the area directly from the map classification, an area estimator can be based on the reference classification of each sample unit. The area proportions for each reference-defined category j are estimated from the column totals ($\hat{p}_{.j}$) in Table 2. An unbiased poststratified estimator of the total proportion of area $\hat{u}$ (based on the reference classification) of poststratum *i* = 1…q and sample units y_{ij} , j = 1 … n is (Cochran 1977, Eq. 5.1):

$$\hat{u} = \sum_{i=1}^{q} W_i \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

(1)

Lohr (1999, Eq 4.22) provides an approximation for the variance estimator for the poststratified estimate,

$$\hat{V}(\hat{u}) \approx \sum_{i=1}^{q} W_i \frac{s_i^2}{n} \qquad (2)$$

which can be combined with Cochran's expression for stratum variance (Cochran 1977, Eq. 3.5) to derive an approximate poststratified variance estimator (Olofsson 2019, pers. comm..)

$$\hat{V}(\hat{u}) \approx \frac{1}{n} \sum_{i=1}^{q} W_i \frac{n_{ij}\left(1-\frac{n_{ij}}{n_i}\right)}{n_i - 1} \qquad (3)$$

A 95% confidence interval for $\hat{u}$ is ± 1.96 $\sqrt{(\hat{V}(\hat{u}))}$, for n > 30, and the margin of error is (confidence interval / $\hat{u}$). Finally, the estimated total area in hectares for each stratum is ($\hat{u}$) x (total population area of 12,844,679 ha).

Result of the Assessment

Based on the assessment to the data between 2006 and 2016, the dynamic changes of land cover, i.e. from forest to non-forest, primary to secondary forest, non-forest to forest, stable forest (forest that remain unchanged between 2006 and 2016) and stable non-forest is found as presented in Table 3. Total area of deforestation, degradation, forest gain, stable forest and stable non-forest is presented in Table 4. From number of 639 sample units, the number of sample for each land cover categories is given in Table 5. The result of the estimation of the area proportion based on the number of sample units is given in Table 6.

Table 3 Dynamic change of land cover and its relation to REDD category activities

	Primary forests	Secondary forests	Non-forest
Primary forests	Stable forest	Degradation	Deforestation
Secondary forests	Stable forest	Stable forest	Deforestation
Non-forests	Forest gain	Forest gain	Stable non forest

Table 4 Change of land covers between 2006 and 2016 based on National map

Class of Land Cover Change	Total area in 2016 (ha)
Deforestation (Forest change to Non Forest)	701,685
Forest Degradation (Primary Forest change to Secondary Forest)	93,979
Forest Gain (Non Forest to Forest)	372,712
Stable Forest	6,525,057
Stable Non Forest	5,151,246
Total	12,844,679

Table 5 Error matrix (nij) from comparison of map to reference data

		Sample/Reference								
		Defore station	Forest Degradation	Forest Gain	Stable Forest	Stable Non-Forest	Total (N)	N-1	Map Area (Ha)	W_i
Map	Deforestation	18	2		5	12	37	36	701,685	0.055
	Forest Degradation				2		2	1	93,979	0.007
	Forest Gain	3			6	13	22	21	372,712	0.029
	Stable Forest	20	11		271	27	329	328	6,525,057	0.508
	Stable Non-Forest	17	1		19	212	249	248	5,151,246	0.401
	Total	58	14	0	303	264	639	638	12,844,679	1.000

Table 6 Estimated area proportion (p^ij)

		Sample/Reference								
		Defore station	Forest Degradation	Forest Gain	Stable Forest	Stable Non-Forest	Total (N)	N-1	Map Area (Ha)	W_i
Map	Deforestation	0.027	0.003	0.000	0.007	0.018	0.055		701,685	0.055
	Forest Degradation	0.000	0.000	0.000	0.007	0.000	0.007		93,979	0.007
	Forest Gain	0.004	0.000	0.000	0.008	0.017	0.029		372,712	0.029
	Stable Forest	0.031	0.017	0.000	0.418	0.042	0.508		6,525,057	0.508
	Stable Non-Forest	0.027	0.002	0.000	0.031	0.341	0.401		5,151,246	0.401
	Total	0.089	0.022	0.000	0.472	0.418	1.000		12,844,679	1.000
User accuracy (Ui)		0.486	0.054	0.000	0.135	0.324	1.000			
Producer accuracy (Pi)		0.299	0.137	0.000	0.016	0.042				
Overall accuracy (O)		0.786								

Table 7 shows the estimation of mean area, standard error, 95% confidence interval, and margin of error for each of class using equations (1) and (3).

Table 7 Sample based estimates of the land cover changes, 2006-2016

Stratum	Adjusted area (ha)	SE for the estimated area (ha)	CI (95%)	U (%)
Deforestation	1,140,536	136,641	267,815	23.48
Forest Degradation	276,780	73,544	144,146	52.08
Forest Gain	0	0	0	0
Stable Forest	6,058,260	172,288	337,684	5.57
Stable Non-Forest	5,369,103	167,743	328,776	6.12
Total	12,844,679	542,649.09	1,063,592.22	8.3

Referring to Table 7, the total area being deforested between 2006-2016 should be adjusted from 701,685 ha to 1,140,536 ha. Thus there is an in increase in the estimate of deforestation rate by about 62% from the original. Similarly the sample based estimate of forest degradation has increased from 93,979 ha to 276,780 ha, resulting in an increase of degradation rate by about 195% from the original. It is likely that at least some of this bias is a result of using a minimum mapping unit of 6.25 ha for creating the cover maps. This size unit will miss instances of deforestation or degradation when the area of the parcel is significantly smaller than 6.25 ha, resulting in underestimation of estimates of deforestation and degradation when those processes are increasing across the landscape.

For purposes of estimating emissions from deforestation and degradation, we need to stratify the total estimates of area into the original forest classes, to generate Activity Data for combining with Emission Factors. We allocated the difference in total area of deforestation and degradation (i.e. the adjustment) in simple equal proportional manner back to each of the estimates of land cover change (2006 to 2009, 2009 to 2011, 2011 to 2012, etc). . We chose this approach, as there is no other formal guidance for how to do this.

The result of adjustment compare to the original one can be seen in Tables 8, 9, 10, and 11. It can be seen that about 94% of total deforestation and 99% of total degradation from 2006-2016 occurs in the dry land secondary forest (code 2002), thus differences in error rate across starting forest types should not have significant impacts

Table 8 Area of deforestation from 2006-2016 before sample based estimation

	2016																Grand Total
	Row Labels	**0**	**2006**	**2007**	**2010**	**2012**	**2014**	**5001**	**20071**	**20091**	**20092**	**20094**	**20121**	**20122**	**20141**	**50011**	
2006	2001	0		3,367	978		2,454	4			817						7,620
	2002	0	51,608	230,501	225,146	168	91,016	29	860	474	41,763			134	17,061	17	658,777
	2004	0		54	43		194	23	242		7	130					693
	2005				1,216		118										1,334
	20041	2	3	28	2,541	40	2,165	59	4,252	10	739	4,415	18	3	183		14,458
	20051			71	5,829	352	4,308		4,603	203	392	2,238			39		18,034
	Grand Total	**2**	**51,611**	**234,021**	**235,753**	**560**	**100,254**	**115**	**9,956**	**686**	**43,718**	**6,782**	**18**	**136**	**17,284**	**17**	**700,917**

Table 9 Area of deforestation from 2006-2016 after sample based estimation

	2016																Grand Total
	Code	**-**	**2,006**	**2,007**	**2,010**	**2,012**	**2,014**	**5,001**	**20,071**	**20,091**	**20,092**	**20,094**	**20,121**	**20,122**	**20,141**	**50,011**	
2006	2001	0	-	5,479	1,591	-	3,993	6	-	-	1,330	-	-	-	-	-	12,399
	2002	0	83,977	375,073	366,359	274	148,102	46	1,399	771	67,957	-	-	218	27,763	28	1,071,966
	2004	0	-	88	70	-	315	38	394	-	12	211	-	-	-	-	1,128
	2005	-	-	-	1,979	-	192	-	-	-	-	-	-	-	-	-	2,171
	20041	4	5	45	4,135	65	3,522	97	6,919	16	1,202	7,183	30	4	298	-	23,526
	20051	-	-	116	9,485	573	7,009	-	7,489	330	637	3,642	-	-	64	-	29,345
	Grand Total	**4**	**83,982**	**380,800**	**383,619**	**912**	**163,135**	**187**	**16,201**	**1,117**	**71,139**	**11,036**	**30**	**222**	**28,124**	**28**	1,140,536

Table 10 Area of forest degradation from 2006-2016 before sample based estimation

	2016				
2006	**Code**	**2002**	**20041**	**20051**	**Grand Total**
	2001	92,538.62			92,538.62
	2004		575.52		575.52
	2005			792.19	792.19
	Grand Total	**92,538.62**	**575.52**	**792.19**	**93,906.32**

Table 11 Area of forest degradation from 2006-2016 after sample based estimation

	2016				
2006	**Code**	**2002**	**20041**	**20051**	**Grand Total**
	2001	272,748.48			272,748.48
	2004	-	1,696.28		1,696.28
	2005	-		2,334.91	2,334.91
	Grand Total	**92,538.62**	**575.52**	**792.19**	276,779.67