

FORM 3: DATA COLLECTION

Purpose and scope of the data collected	The data collection covers the Ghana Cocoa Forest REDD+ Programme (GCFRP) area and is used for reference level estimates reported in the Emissions Reductions Program Document (ERPD) to the Forest Carbon Partnership Facility (FCPF).
Version	1
Date	14/06/2021

Individuals involved

Name	Email	Institution	Role for data collection
Yusif Sitobu Abdullai	sitobuyusif14@gmail.com	Centre for Remote Sensing and Geographic Information Services	Sample interpretation
Foster Mensah	fkmawusi@gmail.com	CERSGIS	Sample interpretation
Ebenezer Kofi Baidoo	baidook32@gmail.com	ENVIRONMENTAL PROTECTION AGENCY	Sample interpretation
Christopher Ihejirika	chinihe123@gmail.com	Resource Management Support Center Forestry Commission, Kumasi	Sample interpretation
Justice Ankomah-Baffoe	ankoba.just@gmail.com	CSIR-Soil Research institute - Accra	Sample interpretation
Emmanuel Donkor	emmanueldonkor484@gmail.com	Forestry Commission	Sample interpretation
Richmond Konadu Amoah	Sarfoabredu3@gmail.com	Ministry of Food and Agriculture	Sample interpretation
Kofi Boateng Agyenim	bkofi646@gmail.com	Forestry Commission	Sample interpretation
Agyemang Afua Birago	nanaagyemangworship@gmail.com	Forestry Commission	Sample interpretation
ERIC OSEI	osieeric81@gmail.com	Forestry Commission	Sample interpretation
Jacob Amoako	jacobamoako2012@gmail.com	Forestry Commission	Sample interpretation
Nutefe Kwesi Dra	kwnutefe@gmail.com	Environmental Protection Agency	Sample interpretation

Nathanael Nii-Odai Laryea	niiodailaryea@gmail.com	Ministry of Food and Agriculture	Sample interpretation
Ebenezer Kwanin	ekwanin@gmail.com	Forestry Commission, RMSC	Sample interpretation
Senyo Yao Gakpo	senyogakpo@gmail.com	Ministry of Food and Agriculture	Sample interpretation
Mawuli Kwaku Gbekor	mgbekor@gmail.com	Environmental Protection Agency	Sample interpretation
Ernest Foli	efoli@hotmail.com	Forestry Research Institute	Sample interpretation
Frank Kwadwo Owusu	frankkwadwoowusu@gmail.com	Forestry Commission	Sample interpretation
Lawrence Akpalu	lakpalu@gmail.com	Forestry Commission, RMSC	Sample interpretation
Prince Boama	boamaprince@gmail.com	Forestry Commission, RMSC	Sample interpretation
Raymond Sakyi	rksakyi@yahoo.com	Forestry Commission	Sample interpretation
Tessia Boateng	tessiaboat@gmail.com	Forestry Commission	Sample interpretation
William Osei-Wusu	williamkay88@gmail.com	Forestry Commission, RMSC	Sample interpretation
Yakubu Mohammed	myakubu89@hotmail.com	Forestry Commission, RMSC	Sample interpretation
Baidoo			

Sample unit allocation to interpreters

The samples collected in the GCFRP area

Interpreter name	Number of sample units
A-Baffoe	45
agyemangafuabirago	13
Baidoo	625
CHRISTOPHER IHEJIRIKA	877
DONKOR EMMANUEL	55
Eben Baidoo	25
Ebenezer Kwanin	1139
Enerst	90
ERIC OSEI	28
Ernest	167
fowusu	898

Jacob Amoako	227
Kofi Boateng	28
kofibaffoe64	75
Lawrence	238
Mawuli	32
Prince Boama	462
Raymond Sakyi	271
remiremi	90
RICHMOND KONADU AMOAH	11
senyo yao gakpo	23
Tessia Boateng	1071
William	997
Yakubu Mohammed	18
Yelena	185
Yusif Abdullai	26

Cross validation

The duplicate samples were chosen by selecting samples that were marked as low confidence, had incomplete information and randomly selected samples. In total 598 sample were duplicated and assessed by two different interpreters. The duplicate sample that was used in the reference interpretation was selected by using the sample with high confidence and the most recently assessed sample.

Duplicate interpretation	
Interpreter name	Number of duplicates assessed
Baidoo	40
Christopher Ihejirika	32
Ebenezer Kwanin	156
Ernest	17
fowusu	187
Jacob Amoako	9
kofibaffoe64	8
Lawrence	15
Prince Boama	62
Raymond Sakyi	14
Tessia Boateng	205
William	148
Yakubu Mohammed	303
Grand Total	1196

Per class agreement and disagreement for duplicate samples

	Deforestation	Degraded forest	Forest gain	Stable forest	Stable non-forest	Grand Total
A	9	4	6	167	412	598
Agreement	2			140	348	490
Disagreement	7	4	6	27	64	108
B	7	4	8	194	385	598
Agreement	2			140	348	490
Disagreement	5	4	8	54	37	108
Grand Total	16	8	14	361	797	1196
Overall Interpreters' Agreement						
	82%					

Matrix of duplicate samples

		Interpreter B					
		Deforestation	Degraded forest	Forest gain	Stable forest	Stable non-forest	Grand Total
Interpreter A	Deforestation	2	0	0	0	7	9
	Degraded forest	0	0	0	2	2	4
	Forest gain	0	0	0	3	3	6
	Stable forest	0	2	0	140	25	167
	Stable non-forest	5	2	8	49	348	412
	Grand Total	7	4	8	194	385	598



Collect Earth output description

When exporting the sample data from Collect Earth, the database can be exported in CSV format. The CSV output includes the information collected for each sample in numerous columns. The description of each column from the Collect Earth database is provided here.

Column name	Description
id	Unique identifier for each sample. It can be matched with the ID from the intensification file.
round	Identifier for the number of times the sample was assessed. If a sample was assessed more than one time the most recent sample is used. The max value that can be entered is 10 1 = first time the sample was assessed 2= second time the sample was assessed, etc
round_label	Labels for the assessment phases, linked to the round column Ordinary = Round 1

	Re-assessment = Round 2 QA/QC = Round 3 4 = Round 4 5 = Round 5 6 = Round 6 7 = Round 7 8 = Round 8 9 = Round 9 10 = Round 10
location_srs	Spatial reference system for the coordinates
location_x	X coordinate
location_y	Y coordinate
operator	Name of person that assessed the sample
validation_comment	Hidden value – always blank
land_use_category	Land use category in 2019
land_use_category_label	Land use category label in 2019
land_use_subcategory	Land use change class between 2000-2019
land_use_subcategory_label	Land use change class label between 2000-2019
land_use_category_has_changed	Whether there has been a land use change identified TRUE = a land use change was identified FALSE = no land use change was identified
land_use_subcategory_year_of_change_know	Hidden value – always blank
land_use_subcategory_year_of_change	Year of change
land_use_subcategory_year_of_change_label	Year of change
aoi_classification	Country name
aoi_classification_label	Country name
land_use_subdivision	Land use subdivision in 2019
land_use_subdivision_label	Land use subdivision label in 2019
land_use_subdivision_change	Whether there has been a land use subdivision change identified TRUE = a land use subdivision change was identified FALSE = no land use change subdivision was identified
land_use_assessment_confidence	Confidence in the land use and land use change assessment TRUE = operator was confident in assessment

	FALSE = low confidence, the sample is rechecked by another operator
vhri_source	Primary high resolution source used for the assessment
vhri_source_label	Primary high resolution source used for the assessment
initital_aoi_classification	Country name
initital_aoi_classification_label	Country name
land_use_initial_subdivision	The land use subdivision before a land use change or a land use subdivision change
land_use_initial_subdivision_label	The land use subdivision label before a land use change or a land use subdivision change
land_use_subdivision_year_of_change	Land use subdivision year of change
land_use_subdivision_year_of_change_label	Land use subdivision year of change
second_lu_change	<p>Whether a second land use change occurred</p> <p>Has the land use changed more than once since 2000?</p> <p>For example this might happen if there is a transition from Forest in 2000 to Grassland in 2005 that becomes settlement in 2010.</p> <p>TRUE = a second land use change occurred</p> <p>FALSE = only one land use change occurred</p> <p>Blank = no land use change</p>
second_lu_conversion	<p>Land use change class of the second land use change</p> <p>This is the oldest possible land use change. In the example Forest to Grassland to Settlement, it would be Grassland</p>
second_lu_conversion_label	<p>Land use change class label of the second land use change</p> <p>This is the oldest possible land use change. In the example Forest to Grassland to Settlement, it would be Grassland</p>
plot_distribution	Distribution of the land use features within the plot
plot_distribution_label	Distribution label of the land use features within the plot
second_lu_conversion_year	<p>Year of the second (oldest) land use change</p> <p>If the dynamic is Forest -> 2005 Grassland ->2010 Settlement, then here you would set 2005.</p>
second_lu_conversion_year_label	Year of the second (oldest) land use change
oldest_aoi_classification	Country name or Global
oldest_aoi_classification_label	Country name or Global
second_lu_subdivision	Land use subdivision of the second (oldest) land use change
second_lu_subdivision_label	Land use subdivision label of the second (oldest) land use change

actively_saved	<p>Whether the plot was completed and saved in Collect Earth</p> <p>When a sample is assessed and completed in Collect Earth a green check mark will appear in Google Earth next to the sample.</p> <p>TRUE = the sample was assessed and successfully saved </p> <p>FALSE = the sample was not successfully saved in Collect Earth </p>
actively_saved_on_year	Year the sample was assessed
actively_saved_on_month	Month the sample was assessed
actively_saved_on_day	Day the sample was assessed
plot_file	Name of the input CSV with the samples that were assessed
elevation	Elevation from NASA SRTM Digital Elevation 30m
slope	Slope from NASA SRTM Digital Elevation 30m
aspect	Aspect from NASA SRTM Digital Elevation 30m
calculated_elevation_range	Codes for calculated elevation ranges
calculated_elevation_range_label	Calculated elevation ranges, binned in 100 m intervals
calculated_aspect	Calculated aspect (N,E,S,W)
calculated_aspect_label	Calculated aspect (N,E,S,W)
calculated_slope	<p>Calculated slope</p> <p>Flat = 0 – 5 degrees</p> <p>Slight = 6 – 15 degrees</p> <p>Steep = 16 – 30 degrees</p> <p>Very steep = 31 – 45 degrees</p> <p>Extreme = 46 – 60 degrees</p> <p>Fall = 61 – 90 degrees</p>
calculated_slope_label	Calculated slope
vhri_year	Year of the latest image from Google Earth
vhri_year_label	Year of the latest image from Google Earth
alu_2000_subcategory	Calculated land use subcategory in 2000
alu_2000_category	Calculated land use category in 2000
alu_2000_category_label	Calculated land use category label in 2000
alu_2000_subdivision	Calculated land use subdivision in 2000
alu_2000_subdivision_label	Calculated land use subdivision label in 2000
primary_disturbance	Primary disturbance in the sample
primary_disturbance_label	Primary disturbance label

primary_disturbance_year.1.	Year of primary disturbance Multiple years can be selected
primary_disturbance_year_label.1.	Year of primary disturbance Multiple years can be selected
primary_disturbance_year.2.	Year of primary disturbance Multiple years can be selected
primary_disturbance_year_label.2.	Year of primary disturbance Multiple years can be selected
primary_disturbance_year.3.	Year of primary disturbance Multiple years can be selected
primary_disturbance_year_label.3.	Year of primary disturbance Multiple years can be selected
primary_disturbance_year.4.	Year of primary disturbance Multiple years can be selected
primary_disturbance_year_label.4.	Year of primary disturbance Multiple years can be selected
primary_disturbance_year.5.	Year of primary disturbance Multiple years can be selected
primary_disturbance_year_label.5.	Year of primary disturbance Multiple years can be selected
secondary_disturbance	Secondary disturbance
secondary_disturbance_label	Secondary disturbance label
tertiary_disturbance	Tertiary disturbance
tertiary_disturbance_label	Tertiary disturbance label
climate	Climate codes from IPCC https://esdac.jrc.ec.europa.eu/projects/RenewableEnergy/
climate_label	Climate zones from IPCC
soil	Soil type codes from IPCC https://esdac.jrc.ec.europa.eu/projects/RenewableEnergy/
soil_label	Soil type from IPCC
gez	IPCC Global Ecologic Zone codes https://forest.jrc.ec.europa.eu/en/activities/lulucf/ipcc-classifications/
gez_label	IPCC Global Ecologic Zone labels

country	Country name from FAO GAUL: Global Administrative Unit Layers 2015
province	Province name from FAO GAUL: Global Administrative Unit Layers 2015
district	District name from FAO GAUL: Global Administrative Unit Layers 2015
land_useF_lu	"F" in all columns
land_useF_lu_label	"Forest" in all columns
land_useF_coverage	Forest land use percentage coverage in 2019
land_useF_coverage_label	Number of control points in sample with forest land use in 2019
land_useF_percentage	Forest land use percentage coverage in 2019
land_useC_lu	"C" in all columns
land_useC_lu_label	"Cropland" in all columns
land_useC_coverage	Cropland land use percentage coverage in 2019
land_useC_coverage_label	Number of control points in sample with Cropland use in 2019
land_useC_percentage	Cropland land use percentage coverage in 2019
land_useG_lu	"G" in all columns
land_useG_lu_label	"Grassland" in all columns
land_useG_coverage	Grassland land use percentage coverage in 2019
land_useG_coverage_label	Number of control points in sample with Grassland use in 2019
land_useG_percentage	Grassland land use percentage coverage in 2019
land_useS_lu	"S" in all columns
land_useS_lu_label	"Settlement" in all columns
land_useS_coverage	Settlement land use percentage coverage in 2019
land_useS_coverage_label	Number of control points in sample with Settlement use in 2019
land_useS_percentage	Settlement land use percentage coverage in 2019
land_useW_lu	"W" in all columns
land_useW_lu_label	"Wetlands" in all columns
land_useW_coverage	Wetlands land use percentage coverage in 2019
land_useW_coverage_label	Number of control points in sample with Wetlands land use in 2019
land_useW_percentage	Wetlands percentage coverage in 2019
land_useO_lu	"O" in all columns
land_useO_lu_label	"Otherlands" in all columns

land_useO_coverage	Otherlands land use percentage coverage in 2019
land_useO_coverage_label	Number of control points in sample with Otherlands land use in 2019
land_useO_percentage	Otherlands land use percentage coverage in 2019
dryland_category	Dryland category extracted from Drylands_UNCCD_CSV_july2014
biome	Biome from RESOLVE Ecoregions 2017
ecoregion	Ecoregion from RESOLVE Ecoregions 2017
land_prod	Land productivity code from UNCCD
land_prod_label	Land productivity label from UNCCD
grassland_mgmt	Grassland management in grassland land use
grassland_mgmt_label	Grassland management in grassland land use
watershed_lvl0	No value – always blank
watershed_lvl1	No value – always blank
project	No value – always blank
project_label	No value – always blank
comments	Comment left by operator
grid_type	How the samples are distributed Systematic samples are distributed in a consistent manner using the SIGRID grid
grid_type_label	How the samples are distributed Systematic samples are distributed in a consistent manner using the SIGRID grid
pre_disturbance_treecover	Percentage tree cover before the forest disturbance is recorded
pre_disturbance_treecover_label	Number of control points in the sample with tree cover before the forest disturbance is recorded
post_disturbance_treecover	Percentage tree cover after the forest disturbance is recorded
post_disturbance_treecover_label	Number of control points in the sample with tree cover after the forest disturbance is recorded
ecozone	Ghana ecozones
pre_deforestation_tree_cover	Percentage tree cover before deforestation is recorded
pre_deforestation_tree_cover_label	Number of control points in the sample with tree cover before deforestation is recorded
forestmask	Forest mask value
forestmaskcode	Forest mask value
forest_vegetation_zones	Combination of forest mask and ecozones

forest_vegetation_zones_label	Combination of forest mask and ecozones
game_reserve	Game reserve name
forest_reserve	Forest reserve name
vegcode.y	Ghana ecozones code
vegzone	Ghana ecozones
all_reserves	Calculated based on game_reserve and forest_reserve Presence (1) or absence (0) of forest or game reserves
forestreserves_binary	Calculated based on forest_reserve Presence (1) or absence (0) of forest reserves
gamereserves_binary	Calculated based on game_reserve Presence (1) or absence (0) of game reserves
Disturbance	Calculated based on primary_disturbance Presence (1) or absence (0) of disturbances
Disturbance on FF	Calculated based on primary_disturbance and land_use_subcategory Presence (1) or absence (0) of disturbances in forest remaining forest
land_use_REDD_2005	Calculated based on land_use_category_has_changed and land_use_subcategory_year_of_change and land_use_category Land use in 2005 (Forest/non-forest) Fo = Forest area in 2005 NF = Non-forest in 2005
land_use_2005_redd_label	Calculated based on land_use_category_has_changed and land_use_subcategory_year_of_change and land_use_category Land use in 2005 (Forest/non-forest) Forest = Forest area in 2005 Non-Forest = Non-forest in 2005
closed_open_degr	Calculated based on disturbance_ff_binary pre_disturbance_treecover Disturbance on open or closed forests 1=degradation in open forest 2= degradation in closed forest 0=no degradation
fm_eco_area_ha_combined	Area of stratum
fm_eco_label	Stratification class name

stratum	Sampling intensity – spacing in km between samples
closed_open_def	Calculated based on pre_deforestation_tree_cover Deforestation on Closed/Open Forest 1=deforestation in open forest 2= deforestation in closed forest 0=no deforestation
CODE	Concatenation of fm_eco_label stratum and CODE2
CODE2	FF = forest remaining forest, 0 = non-forest remaining non-forest, FN = forest to non-forest,