Accuracy Assessment - Appendix

Collection 11

Version 1

General coordinator
Laerte Guimarães Ferreira Jr.
Technical coordinator
Ana Paula Matos

Team
Ana Paula Assunção
Amanda Rosa Falcão
Vinicius Vieira Mesquita
Leandro Parente
Tharles Andrade
Luís Bauman

August, 2026

Table of Contents

How to cite:

Ferreira Jr., L.; Matos, A.P., 2026, "MapBiomas Land Use and Land Cover - Algorithm Theoretical Basis Document (ATBD) - Collection 11 – Accuracy Assessment Appendix".

1. Overview

This document describes the methods, steps, and results of the accuracy assessment for the Land Cover and Land Use (LCLU) maps of the MapBiomas collection 11. This assessment was performed at the Landsat pixel level, with a spatial resolution of 30 meters. Our map accuracy was estimated by independent validation, using a set of LCLU point samples to generate accuracy statistics for our time-series maps. The sampling scheme was designed using a random sampling strategy stratified by the terrain slope and the topographic charts defined by the Brazilian Institute of Geography and Statistics-IBGE (Figure 1). A total of 85,152 samples were visually interpreted annually in a historical series based on a labeling criteria established for each LCLU previously defined Brazilian Biome and periodic training involving the team of interpreters.

The validation dataset evolved alongside the MapBiomas map collections (Table 1), which extends a new year to the time-series and/or adds new classes to the map legend. As accuracy was performed for each new collection, the class labels on the validation samples were revised and re-interpreted when necessary. .

Table 1. This overview presents a historical account of the evolution of the collections

Accuracy Collection	MapBiomas Collection	Times Series	Advances
1	3 and 4	1985 to 2018	Initial 20 classes
1.1	5 until 7	1985 to 2018	4 new classes
2	8	1985 to 2022	New classes and new years (2019 to 2022)
2.1	9	1985 to 2022	Revisions
3	10 and 11	1985 to 2024	Updates and Revisions

2. Validation Strategies

The validation strategy was based on two approaches: (i) comparative analysis with reference maps of specific biomes/regions and years, and (ii) accuracy assessment based on statistical techniques using independent sample points covering the entire extent of Brazil throughout the time-series.

2.1. Validation with Reference Maps

A set of reference maps were used (where available) to estimate a metric of spatial agreement with our maps, performed for each biome and by cross-cutting theme. More details are available in the Appendices and on the reference maps webpage (https://brasil.mapbiomas.org/en/mapas-de-referencia/).

2.2. Validation with Independent Points

2.2.1 Sampling design of General Classes

The sampling design was performed at a spatial unit based on the grouping of four IBGE topographic charts (“group charts”), where the sampling strategy was performed. We used a total of 127 grouped charts distributed across the country, where a stratified sampling scheme was performed for each group chart. The samples were stratified by slope using 6 slope classes defined by the Brazilian Agricultural and Research Organization - EMBRAPA.

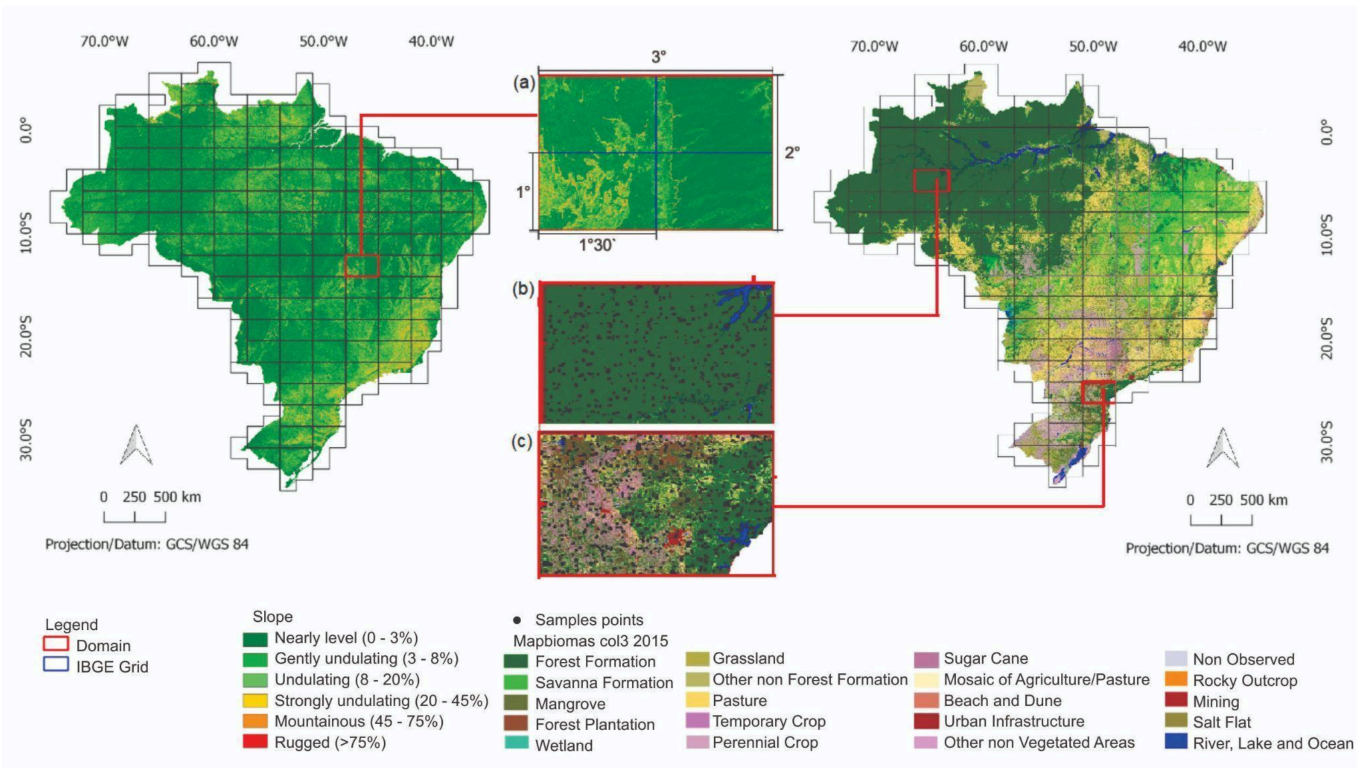


Figure 1. Slope categories used in sampling design for validation with independent points and group-chart size used as a subset in the sampling design for validation with independent points.

The total sample size was established to guarantee a maximum margin of error of 5% and confidence level of 95%, with a maximum error of 0.5% expected for the entire Brazilian territory. A minimum of 500 points were randomly distributed within each group chart, with extra points added as an increasing function relative to the variability and/or number of classes mapped within each spatial unit, maintaining the same level of confidence and margin of error. The MapBiomas Land Cover and Land Use map of Collection 3 for 2015 was used as a reference map to identify the number of classes available. The Bonferroni correction, which adjusts the quantile value of a normal distribution, was used to establish the sample size according to the number of classes, while the variability information was derived from the maximum variance of the land use and land cover classes.

The samples for each group-chart were randomly distributed using a proportion stratified sampling strategy n_{PSS} (Cochran, 1977). The sample size was approximated in an overestimated manner using the formula for simple random sample size n_c since we substituted the variability of each spatial unit with the maximum variability of the classes, as follows:

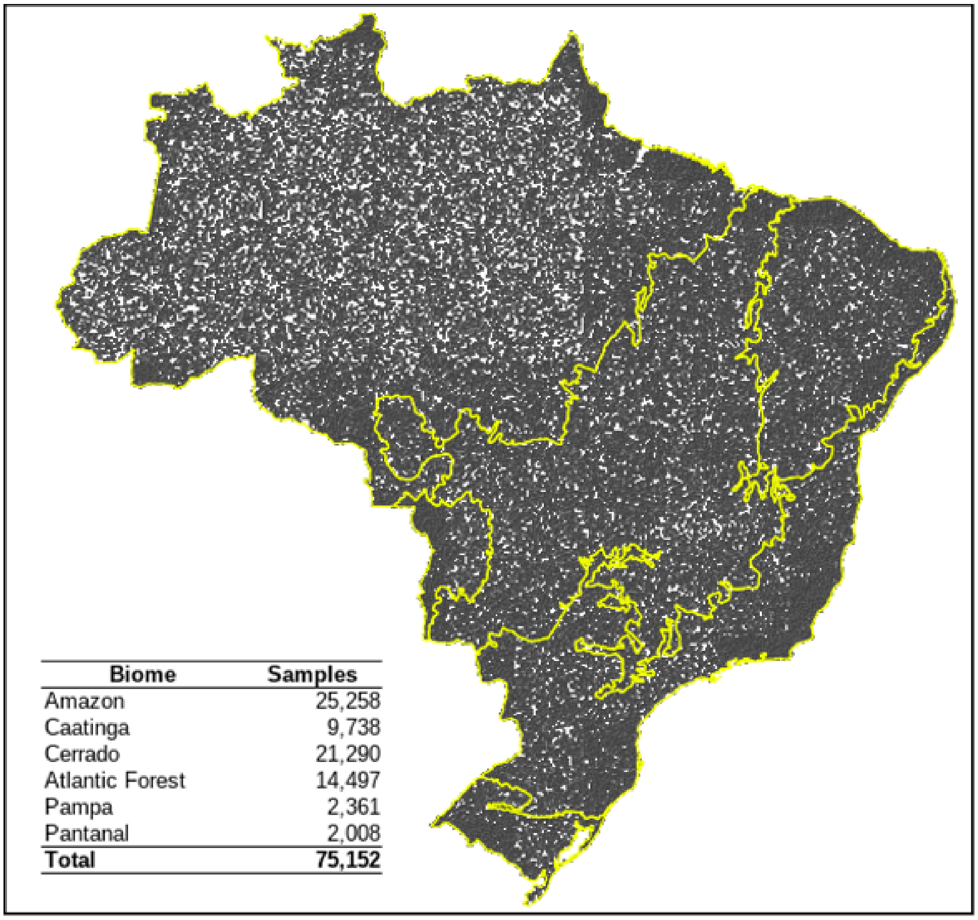
$$n_c = \frac{Nz^2 s^2_{max}}{(N-1)E^2 + z^2 s^2_{max}} > \frac{Nz^2 s^2_{max}}{NE^2 + z^2 s^2_{max}} = \frac{\left(\sum_h W_h s^2_h\right)^2}{\left(\frac{E}{z}\right)^2 + \frac{1}{N}\sum_h W_h s^2_h} = n_{PSS} \ , \qquad (1)$$

where n_c is the sample size in c-th grouped chart ; *N* is the total number of population; *E* is the expected maximum margin of error; *z* is the quantile of the standardized normal distribution corresponding to the adjusted confidence level $1 - \alpha_b$ calculated with the Bonferroni correction, where $\alpha_b = \alpha/2(k - 1)$ and $1 - \alpha$ is the desired confidence level and *k* is the number of land cover and land use classes; and $s^2_{max} = max_i\, p_i(1 - p_i)$ is the maximum variance among the *i*'s coverage within a given domain; W_h is the proportion of spatial unit *h* within the population and s^2_h is the variance of spatial unit *h*.

Thus, the final sample size n distributed throughout Brazil is given by

$$n = \sum_{c=1}^{127} max(n_c, 500) = 85{,}152 \ ,$$

while 10,000 samples out of this total were used as training samples for the Amazon biome. Thus, the accuracy assessment was done using a total of 75,152 samples per year (Figure 2). These samples were then randomly selected following a proportional stratified sampling scheme in each grouped chart, followed by a simple random sampling in each slope stratum. The number of samples were splitted by biome as follows: 35,258 points for the Amazon biome, 21,290 for the Cerrado biome, 9,738 for the Caatinga biome, 14,497 for the Atlantic Forest biome, 2,008 for the Pantanal biome, and 2,361 for the Pampa biome.



Biome	Samples
Amazon	25,258
Caatinga	9,738
Cerrado	21,290
Atlantic Forest	14,497
Pampa	2,361
Pantanal	2,008
Total	**75,152**

Figure 2. Independent random samples used in the accuracy assessment of the MapBiomas Land Cover and Land Use Map Collections.

2.2.2 Labeling Protocol

Three independent interpreters inspected and labeled each independent sample; in case of disagreement between interpreters, a senior interpreter decided the final LCLU class for the sample. This evaluation was based on the web platform Temporal Visual Inspection (TVI; figure 3), developed by the Remote Sensing and GIS Lab at the Federal University of Goiás (Lapig / UFG). The TVI platform allowed the assessment of all the classes mapped by MapBiomas since Collection 3.1 (https://mapbiomas.org/accuracy-statistics?cama_set_language=en). The classification was generated through the visual interpretation of satellite images acquired by the Landsat sensor series with low atmospheric interference (according to the CLOUD_COVER metadata) for the dry (i.e. June to October) and rainy (i.e. January to May) seasons. We used a false-color band composition as: NIR, SWIR1, RED , in addition to MODIS-NDVI times series, and high-resolution imagery from Google Earth (where available).

User	2021	2022	2023	2024
inspector.1	Forest Formation	Temporary Crop - BORDER	Temporary Crop - BORDER	Temporary Crop
inspector.2	Forest Formation	Temporary Crop - BORDER	Temporary Crop - BORDER	Temporary Crop
inspector.3	Forest Formation	Temporary Crop - BORDER	Temporary Crop - BORDER	Temporary Crop
Consolidated Class	**Forest Formation**	**Temporary Crop - BORDER**	**Temporary Crop - BORDER**	**Temporary Crop**

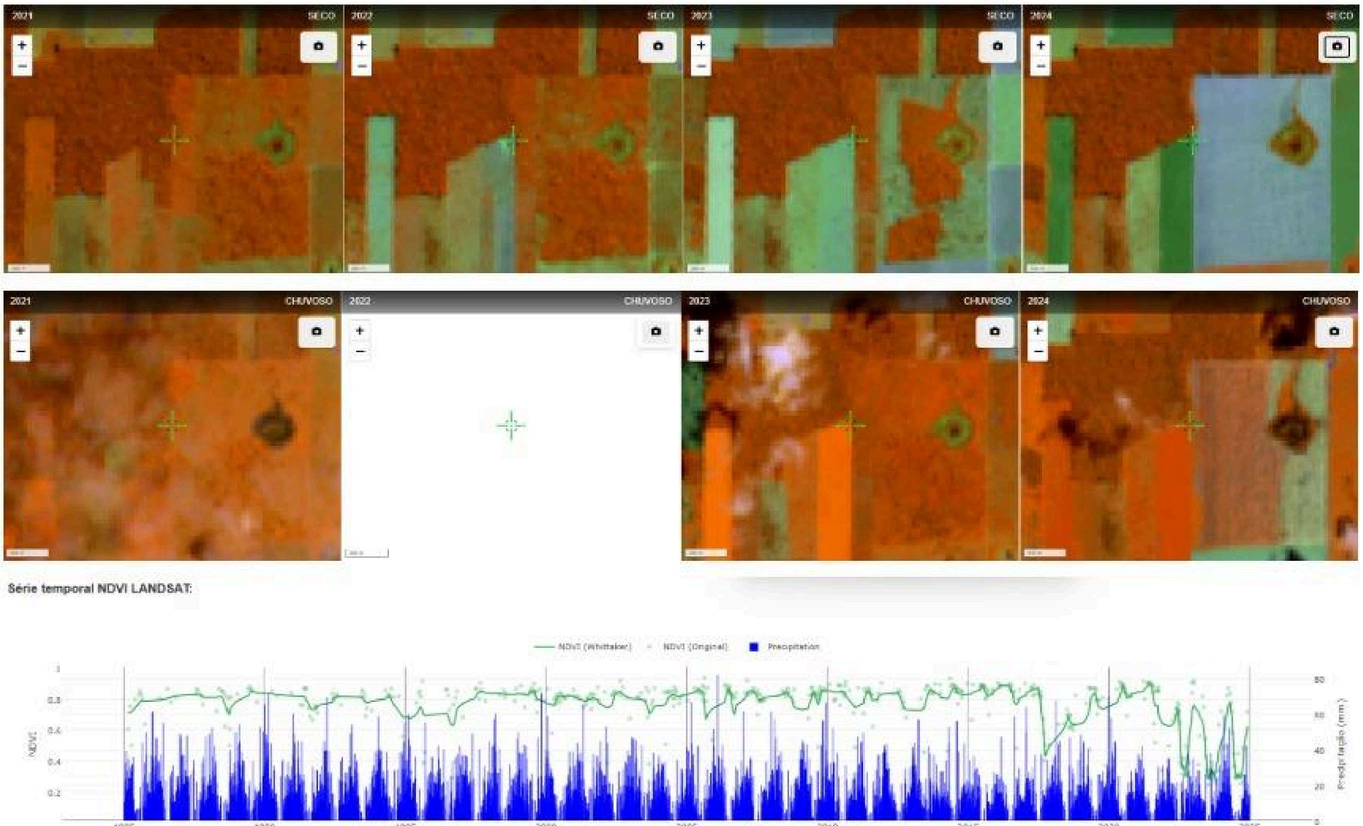


Figure 3. Temporal Visual Inspection (TVI) tool.

Considering the inherent limitations of visual interpretation of medium-spatial-resolution imagery, the original legend descriptions were adapted to prioritize characteristics that could be consistently identified in Landsat imagery and the auxiliary data sources used (Table 2). Descriptions based on highly specific vegetation types, floristic composition, or attributes that are difficult to distinguish through remote sensing were therefore simplified and reorganized primarily according to vegetation structure and physiognomy. The main criteria included vegetation height, canopy density and continuity, the presence and distribution of trees and shrubs, dominance of the herbaceous layer, soil exposure, presence of water, and spatial patterns associated with land use. Landscape context and temporal dynamics observed throughout the image time series were also considered, particularly to distinguish processes such as vegetation regeneration, pasture management, agricultural cycles, forest plantations, and seasonally or permanently flooded areas. This adaptation maintained correspondence with the original legend classes while establishing more objective criteria based on features that could be effectively observed during visual interpretation.

Table 2. Cross-reference of MapBiomas Collection 11 land use and land cover classes and their criteria for visual interpretation.

Level 1	Level 2	Level 3	Level 4	Biome	Brief description
Forest	Forest Formation			Amazon	Tall, dense vegetation with a predominantly arboreal structure, characterized by closed or nearly closed canopies. Areas showing signs of selective logging, fire, or degradation will be considered forest when they retain their arboreal structure and canopy cover. Regenerating areas with tall vegetation and dense cover are also included.
				Caatinga	Vegetation with a predominantly arboreal structure, relatively tall and dense growth, and a continuous canopy. This category includes vegetation types such as Wooded Steppe Savanna, Seasonal Semideciduous Forest, and Seasonal Deciduous Forest.
				Cerrado	Vegetation types with a predominantly arboreal structure, tall and dense growth, characterized by a continuous canopy or strongly connected tree crowns. Areas showing signs of fire or degradation processes will also be considered forest when they retain their arboreal structure and canopy cover. Regenerating areas with tall vegetation, high tree density, and continuous canopy cover may also be included.
				Atlantic Forest	Native vegetation with a predominantly arboreal structure, characterized by tall stature, high tree density, a closed or strongly connected canopy, and the presence of distinct vertical vegetation layers. Areas showing signs of degradation may also be included in this class when the forest structure remains predominant, particularly in terms of tree density and canopy cover.
				Pampa	Vegetation with a predominantly arboreal structure, characterized by high tree density and the formation of a continuous canopy. Areas with only sparse or isolated trees are not included in this class.
				Pantanal	Vegetation with a predominantly arboreal structure, characterized by tall trees, relatively dense cover, and the presence of shrubs in the understory. It may occur in environments subject to varying levels of flooding.
	Savanna Formation			Amazon	Open vegetation characterized by the presence of sparsely distributed trees and/or shrubs, without the formation of a continuous canopy. The herbaceous layer is always present and visible between woody individuals, while the tree and shrub components may vary in their level of development and density.
				Caatinga	Vegetation is predominantly composed of shrubs and low- to medium-height trees, with semi-continuous canopy cover and open spaces between the tree crowns.
				Cerrado	Vegetation is composed of low- to medium-height trees and shrubs distributed over a herbaceous layer, with discontinuous crowns that allow the ground vegetation or soil to be visible between woody individuals. This category includes different physiognomies of the Cerrado sensu stricto (Ribeiro & Walter, 2008).

		Atlantic Forest	Vegetation with sparsely distributed trees and shrubs, generally of low to medium height, forming an open or semi-continuous canopy cover.
		Pantanal	Vegetation is composed of sparsely distributed low-growing trees and shrubs over a continuous herbaceous matrix, with variations in the density of shrub and tree components that may form mosaics between more open and denser areas.
	Mangrove		Dense, evergreen forest vegetation associated with coastal mangrove ecosystems in humid environments influenced by tides. It has closed cover with a continuous canopy and a characteristic distribution along estuaries, channels, and river mouths.
	Floodable Forest	Amazon	Alluvial forest vegetation associated with watercourses, floodplains, riverbanks, and areas subject to periodic or permanent flooding. It has a predominantly arboreal structure and relatively dense canopy cover, maintaining these characteristics even in areas with the presence of water.
Herbaceous and Shrubby Vegetation	Wetland	Amazon	Predominantly herbaceous or low-growing vegetation associated with flat areas under fluvial and/or lacustrine influence, subject to periodic or seasonal flooding.
		Cerrado	Ecosystems predominantly composed of herbaceous vegetation, associated with wet areas or those influenced by rivers and/or lakes. In some areas, they may include scattered trees, shrubs, or palms, such as Cerrado Park and Veredas formations (Ribeiro & Walter, 2008), as long as the herbaceous matrix remains dominant.
		Atlantic Forest	Floodplain vegetation or grassland formations adapted to wet or waterlogged soils, under fluvial and/or lacustrine influence, characterized by the predominance of hygrophilous species and associated with river floodplains, lakes, and flood-prone areas. It may include emergent, submerged, or floating aquatic plants.
		Pampa	Predominantly herbaceous and hygrophilous vegetation associated with low-lying areas or terrain depressions with waterlogged soils, as well as shallow margins of lakes, banhados, and other wet areas. It may include emergent, submerged, or floating aquatic plants.
		Pantanal	Predominantly herbaceous vegetation, with emphasis on grasses, subject to temporary or permanent flooding according to natural flood pulses. It may present scattered shrubs, trees, or palms, forming mosaics within a wet grassland matrix. These waterlogged areas generally occur along the margins of temporary or permanent lagoons and may include emergent, submerged, or floating aquatic plants. Water surfaces are also included when their identification is hindered by the abundant presence of macrophytes and aquatic plants.
	Grassland	Amazon	Open vegetation, predominantly herbaceous, composed mainly of grasses and other low-growing plants. Trees and shrubs may occur sparsely without altering the predominance of the herbaceous cover.

		Caatinga	Predominantly herbaceous vegetation, which may include low and sparsely distributed shrubs. It also includes open areas subject to seasonal flooding, provided that herbaceous cover remains the dominant component.
		Cerrado	Open grassland vegetation dominated by grasses and other herbaceous species, with minimal presence or absence of woody species. It may occur on flat terrain, rocky areas, or sites with shallow soils.
		Atlantic Forest	Vegetation dominated by grasses, herbs, and small subshrubs. Trees and shrubs are absent or occur only as isolated individuals. Signs of grazing, patches of invasive species, or small areas of planted forage may be present, provided that the native grassland matrix remains predominant.
		Pampa	Vegetation dominated by grasses, herbs, and small subshrubs. Trees and shrubs are absent or occur only as isolated individuals. Signs of grazing or management may be present, provided that the native grassland matrix remains predominant.
		Pantanal	Predominantly open herbaceous vegetation, composed mainly of grasses, with occasional low and isolated shrubs. Signs of grazing or management may be present, provided that the native grassland matrix remains predominant.
	Hypersaline Tidal Flat		"Apicuns" or hypersaline tidal flats are formations often without tree vegetation, associated with a higher, hypersaline and less flooded area in the mangrove, generally in the transition between this area and the continent.
	Rocky Outcrop		Area with naturally exposed rock and little or no soil cover. Small patches of vegetation adapted to rocky substrates may be present, but the exposed rock surface remains predominant. It generally occurs on steep terrain.
	Herbaceous Sandbank Vegetation	Atlantic Forest	Low-growing herbaceous or shrubby vegetation established on sandy soils and coastal dunes.
		Caatinga	Low-growing vegetation, predominantly herbaceous, established on sandy soils or dunes in coastal areas.
		Cerrado	Coastal sandy plain ecosystems characterized by predominantly herbaceous and shrubby vegetation, with sparse shrub distribution.
		Pampa	Low-growing herbaceous or shrubby vegetation established on sandy soils and coastal dunes.
Farming	Pasture	Amazon	Area predominantly covered by planted (exotic) or managed grasses and associated with livestock production. It may include exposed soil, variations in vegetation vigor, isolated trees, shrubs, or signs of degradation. Temporary periods of regeneration, abandonment, or reduced management are still considered pasture when there is evidence of livestock use and the potential for this use to resume. Dense, tall, and persistent regeneration without recent signs of management should be assessed according to the observed vegetation structure.
		Caatinga	
		Cerrado	
		Atlantic Forest	
		Pantanal	

	Agriculture	Temporary Crop	Sugar cane	Areas cultivated with sugarcane, identified by the homogeneous pattern of fields, planting rows, and the harvest and regrowth cycle. Recently harvested areas or fields temporarily showing exposed soil remain in this class when the field context indicates continuous sugarcane cultivation.
			Other Temporary Crops	Areas occupied by short- or medium-cycle agricultural crops, typically organized into fields and planting rows. This class includes crop development, harvesting, soil preparation, and fallow periods between growing seasons, provided that the land-use history and surrounding context indicate continuity of agricultural use.
		Perennial Crop		Areas occupied by agricultural crops with long growing cycles that allow successive harvests without the need for replanting. These crops are typically arranged in rows or regular patterns and remain productive for several years. This class includes permanent tree and shrub crops.
	Forest Plantation			Homogeneous plantations of tree species established for commercial purposes, generally arranged in rows or regular stands. This class includes different stages of the production cycle, such as growth, harvesting, exposed soil after cutting, regrowth, site preparation, and fallow periods between rotations, when the context indicates continuity of forestry use.
Non Vegetated Area	Beach, Dune and Sand Spot			Predominantly natural sandy areas, generally light-colored and with little or no significant vegetation cover. This class includes beaches, dunes, and permanent areas of exposed sand.
	Urban Area	Urban Area		Areas with a high concentration of buildings, streets, paved surfaces, and other built-up or impervious surfaces. This class also includes open spaces and unvegetated land that are continuously integrated into the urban structure.
		Urban Vegetation		Vegetated areas within the urban fabric, such as parks, public squares, densely wooded areas, lawns, gardens, and large vegetated plots surrounded by buildings and urban infrastructure.
	Mining			Areas of industrial or artisanal mineral extraction (garimpos), characterized by extensive exposure of soil or rock, excavations, open pits, material stockpiles, dams, or other structures associated with mining activities.
	Other Non Vegetated Areas			Roads, yards, infrastructure located outside urban areas, photovoltaic power plants, permanent patches of exposed soil, erosion features, and other unvegetated surfaces that do not fit into the other classes. In the Amazon, this class may also include recently deforested areas that have not yet been converted to productive use and may subsequently undergo vegetation regeneration.
Water	River, Lake and Ocean			Natural or artificial surfaces covered by water, including rivers, lakes, dams, reservoirs, and other water bodies.
	Aquaculture			Artificial water bodies, generally small, regular in shape, and arranged in groups, associated with aquaculture or salt production. Artificial lakes, with a predominance of aquaculture and/or salt production activities.

Not Observed	Areas without adequate surface observation due to the presence of clouds, shadows, smoke, atmospheric interference, image errors, or the absence of valid data at the beginning or end of the time series.

2.2.3 Revision and Update Procedures

Collection maintenance is carried out through two complementary procedures: targeted revisions and time-series updates (Figure 4). Targeted revisions are conducted when errors are identified in specific years or classes of an existing collection. In these cases, the affected portions are reassessed and corrected by a specialist, resulting in a revised version of the collection without changing its temporal extent. For example, a collection originally released as Collection 1 becomes Collection 1.1 after targeted corrections are incorporated.

For time-series updates, the most recent revised version of the previous collection is used as the reference. The class assigned to the last available year is initially replicated to the newly added years and then reviewed by two additional interpreters. Based on image inspection, the interpreters assess whether the class observed in the new years remains consistent with that recorded at the end of the previous collection. When agreement is reached and no land-use or land-cover change is identified, the new years are incorporated into the time series while preserving the previously established classification. This approach takes advantage of the consistency of the consolidated collection and concentrates interpretation efforts on the newly added period.

When the interpretation of the new years does not confirm the persistence of the previously assigned class, the assessment is extended beyond the newly added period. In such cases, the entire corresponding time series is reassessed by three independent interpreters, allowing the timing and trajectory of the change to be evaluated and any inconsistencies throughout the historical series to be corrected. The revised results are then consolidated together with the new years, producing an updated collection with an extended temporal coverage. This new collection subsequently becomes the basis for future targeted revisions and time-series updates, establishing a continuous process of refinement and expansion of the historical series.

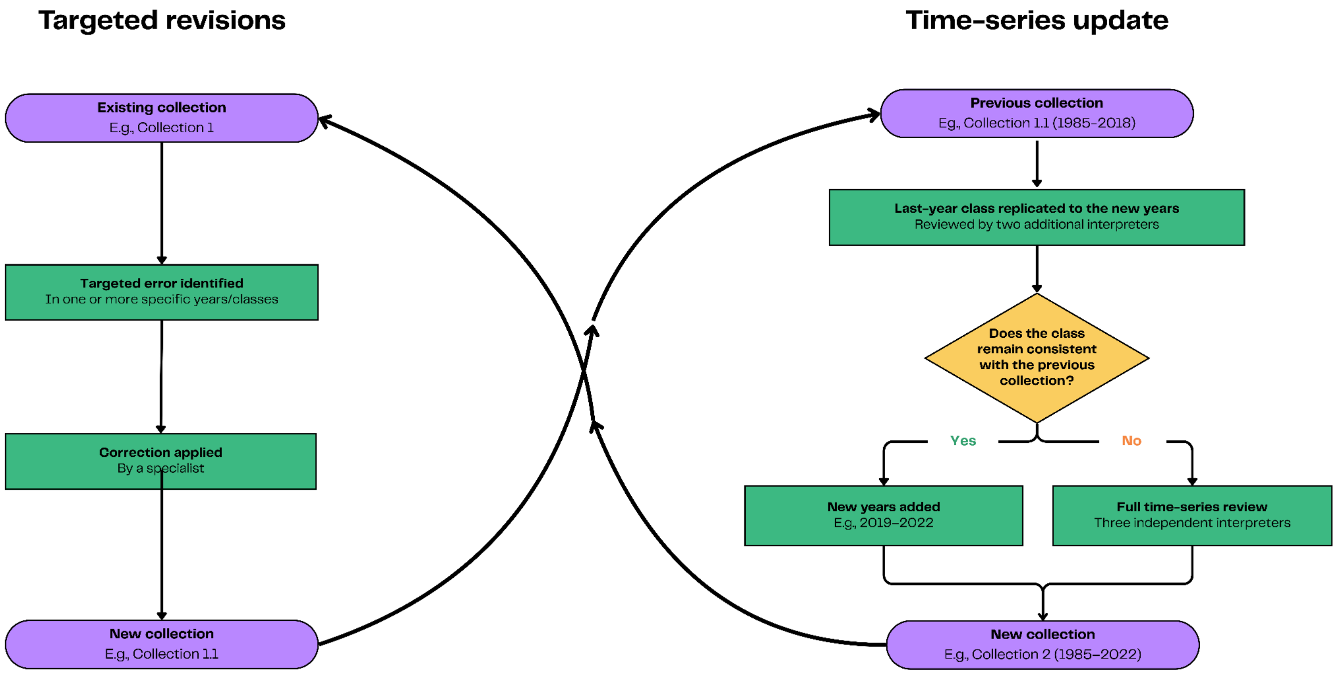


Figure 4. Workflow for maintaining and updating MapBiomas visual interpretation collections. Targeted revisions correct specific errors identified in an existing collection without changing its temporal extent, while time-series updates incorporate new years based on the most recent revised collection. When the class remains consistent with the previous collection, the new years are added directly; otherwise, the full time series is reassessed by three independent interpreters before generating a new updated collection.

2.2.4 Accuracy metrics

Before the accuracy assessment, all sample points located along class boundaries are removed because they are associated with greater interpretation uncertainty, and the remaining points are assigned weights according to the level of agreement among interpreters. Classes with an insufficient number of samples for a representative assessment are also excluded from the accuracy calculation. Legend harmonization rules are then applied. At the national level, Level 4 agricultural classes mapped are converted to their corresponding Level 3 classes: classes 20, 39, 40, 41, and 62 are grouped as Temporary Crops (19), while classes 35, 46, 47, and 48 are grouped as Perennial Crops (36). Wooded Sandbank Vegetation (49) is converted to Forest Formation (3), Herbaceous Sandbank Vegetation (50) to Grassland (12), and Floodable Forest (6) to Forest Formation (3). When the map indicates Mosaic of Uses (21), reference samples classified as Pasture (15), Temporary Crops (19 and 20), or Perennial Crops (36) are also considered Mosaic of Uses.

In addition to these general rules, biome-specific adjustments are applied to account for particular landscape characteristics and class interpretation. In the Pampa, Pasture (15) in the reference data is considered Mosaic of Uses (21), and Savanna Formation (4) is considered Grassland (12). In the Atlantic Forest, when the map indicates Grassland (12) and the reference indicates Pasture (15), the combination is considered an agreement as Grassland. In the Pantanal, specific combinations among Wetland (11), Grassland (12), classes associated with floodable areas (Water - 33), and Flooded Savanna (7) are treated as equivalent, allowing agreement among these classes according to the class indicated by the map. In the Amazon, Forest Formation (3) and Floodable Forest (6) are treated as equivalent. In the Cerrado, specific rules related to Herbaceous Sandbank Vegetation (50) and non-vegetated classes (25) are already accounted for by the general harmonization rules or by the exclusion of classes with low sample representation. In the Caatinga, Wetland (11) in the reference data is considered Grassland (12), and when the map indicates Herbaceous and Shrubby Formation (77), reference samples classified as Grassland (12), Savanna Formation (4), or originally as Wetland (11) may be considered in agreement with the class indicated by the map.

Following sample filtering, weighting, and the application of national and biome-specific legend harmonization rules, accuracy was assessed using the resulting reference dataset. We derived accuracy metrics described by Stehman et al. (2014) and Stehman & Fody (2019) using the error matrix to estimate global, user's, and producer's accuracies. Accuracy metrics are assessed for each year based on the cross-tabulation of the count of samples for the mapped and reference classes, as shown in Table 3. The frequencies N_{ij} represent the count of pixels classified as class i and evaluated as class i. The row marginal totals $N_{i\bullet}$ represent the total count of pixels mapped as class i, while the column marginal totals $N_{\bullet j}$ represents total number of pixels evaluated by technicians (i.e. reference) as class j.

Table 3: Populational error matrix (or confusion matrix) considering k classes.

Map/Reference	Class 1	⋯	Class k	Total
Class 1	N_{11}	⋯	N_{1k}	$N_{1\bullet}$
⋮	⋮	⋱	⋮	⋮
Class k	N_{k1}	⋯	N_{kk}	$N_{k\bullet}$
Total	$N_{\bullet 1}$	⋯	$N_{\bullet k}$	$N_{\bullet\bullet}$

The estimates of N_{ij} in Table 3 are obtained by

$$\hat{N}_{ij} = \sum_{s=1}^{n^*} w_s y_{sij},$$

where n^* is the total number of pixels used to estimate N_{ij}, w_s is the sampling weight of the pixel s, $y_{sij} = 1$ if the sample pixels classified as class i and evaluated as class j of the pixel s and $y_{sij} = 0$ otherwise. The weights w_s is adjusted according to missing data and the classification disagreement of the reference data, such that

$$w_s = w_s^h \times w_s^d \times w_s^p$$

where w_s^h is the sampling weight of the pixel within spatial unit h and is the inverse probability of a pixel being selected in the sample, which probability is given by $\frac{n_h}{N_h}$, $w_s^d \in \{1, 2, 3\}$ such that 1 represents the disagreement among the 3 reference evaluators, 2 the agreement of 2 reference evaluators, and 3 the agreement of all 3 reference evaluators and $w_s^h = \frac{N_h}{\sum_{s\in h} w_s^h \times w_s^d}$ is the adjustment for total population in the spatial unit h where N_h is the total population of pixels in the spatial unit h. The adjustment for total population is necessary to correct the population totals of pixel and area estimates, even with missing data, application of filters, and exclusion of samples, making it so that $\sum_{s=1}^{n^*} w_s$ matches the total population of pixels in the map N.

Estimates for the marginal totals $N_{i\bullet}$, $N_{\bullet j}$ and the overall total $N_{\bullet\bullet}$ are calculated by

$$\hat{N}_{\bullet j} = \sum_{i=1}^{k} \hat{N}_{ij}, \qquad \hat{N}_{i\bullet} = \sum_{j=1}^{k} \hat{N}_{ij} \qquad \text{and} \qquad \hat{N}_{\bullet\bullet} = \sum_{i=1}^{k} \sum_{j=1}^{k} \hat{N}_{ij}.$$

The proportion of pixels (or area) in each cell of Table 3, as well as the marginal proportion of rows and columns, are estimated by

$$\hat{p}_{ij} = \frac{\hat{N}_{ij}}{\hat{N}_{\bullet\bullet}}, \qquad \hat{p}_{i\bullet} = \frac{\hat{N}_{i\bullet}}{\hat{N}_{\bullet\bullet}} \qquad \text{and} \qquad \hat{p}_{\bullet j} = \frac{\hat{N}_{\bullet j}}{\hat{N}_{\bullet\bullet}}.$$

Thus, using the estimated proportions from Table 3, we obtain:

1) **User's accuracy:** fractions of mapped pixels relative to the total correctly classified pixels, for each class. User's accuracy is associated with commission error, which occurs when a pixel mapped as class *i* belongs to some other class. The user's accuracy for class *i* is estimated by $\hat{U}_i = \frac{\hat{p}_{ii}}{\hat{p}_{i\bullet}}$ and the commission error by $1 - U_i$. These metrics are individual accuracies assessed for each mapped class.
2) **Producer's accuracy:** fractions of pixels from each class correctly mapped by the classifier. The producer's accuracy is associated with the omission error, which occurs when we fail to map a pixel of class *j* correctly. The producer's accuracy for class *j* is estimated by $\hat{P}_j = \frac{\hat{p}_{jj}}{\hat{p}_{\bullet j}}$ and the omission error by $1 - P_j$. These metrics are associated with the sensitivity of the classifier, that is, the ability to correctly distinguish a certain class from others.

3) **Global or overall accuracy:** estimates the global/overall success rate of the classifier. The estimate is given by $\hat{G} = \sum_{i=1}^{k} \hat{p}_{ii}$, the sum of the main diagonal of the error matrix. The complement of accuracy, or the total error, is further decomposed into quantity disagreement and allocation disagreement (Pontius and Millones, 2011).

4) **Quantity disagreement:** it is the proportion of discrepancy between the amounts of each category between the classified map and the reference. Is defined by

$$Q = \frac{1}{2}\sum_{i=1}^{k} |\hat{p}_{\bullet i} - \hat{p}_{i\bullet}|.$$

5) **Allocation disagreement:** it is the proportion of spatial allocation disagreement of the categories between the classified map and the reference. Is defined by

$$A = \sum_{i=1}^{k} min\{\hat{p}_{\bullet i} - \hat{p}_{ii}, \hat{p}_{i\bullet} - \hat{p}_{ii}\}.$$

3. Accuracy assessment of MapBiomas comparison between Collections

A comparison among MapBiomas collections at Level 1 shows a progressive improvement in global accuracy over time for Brazil (Figure 5). Earlier collections presented lower accuracy, particularly Collection 3.1, while the most recent collections consistently achieved the highest values. Collections 10.1 and 11 showed the best overall performance, with global accuracy generally above 94% in the most recent years. This improvement was accompanied by a reduction in allocation disagreement across successive collections, indicating better spatial agreement between mapped and reference classes. Quantity disagreement remained relatively low in all collections, although its temporal behavior varied more markedly among versions. Overall, the results indicate that successive revisions and updates of the MapBiomas collections have increased classification agreement at the broadest legend level, especially by reducing allocation errors.

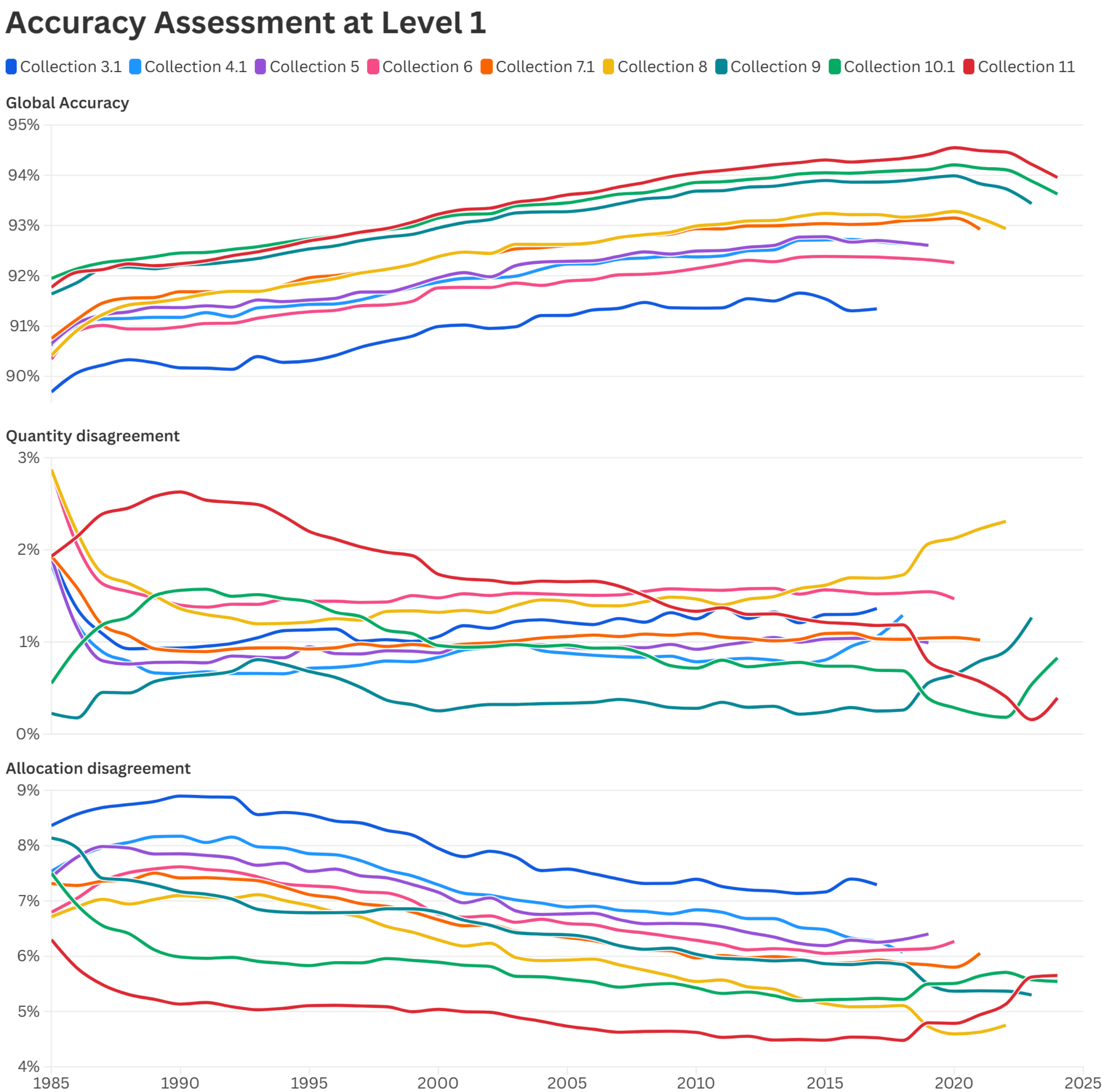


Figure 5. Comparison of Level 1 accuracy metrics across MapBiomas collections. The panels show global accuracy, quantity disagreement, and allocation disagreement over the years available for each collection.

At Level 2, global accuracy also improved progressively across MapBiomas collections, with the most recent versions showing the highest performance for Brazil (Figure 6). Collection 11 consistently presented the highest global accuracy, followed closely by Collections 10.1 and 9, whereas earlier collections generally showed lower values. Compared with Level 1, global accuracy was lower at Level 2 for all collections, as expected from the greater thematic detail introduced at this level. The same overall pattern of improvement across successive collections was maintained, however, indicating increasing consistency even with a more detailed legend.

Quantity disagreement at Level 2 was higher than at Level 1, reflecting greater differences in the estimated proportions of individual classes as thematic detail increased. Nevertheless, the most recent collections generally showed lower quantity disagreement than earlier versions. Allocation disagreement also decreased progressively across collections, with Collections 10.1 and 11 presenting the lowest values over much of the time series. Compared with Level 1, allocation disagreement was consistently higher at Level 2, indicating that distinguishing more detailed land-cover and land-use classes increases the difficulty of achieving exact spatial agreement. Overall, the comparison between the two legend levels shows that the successive MapBiomas collections improved accuracy at both levels, while the expected reduction in performance from Level 1 to Level 2 reflects the increased thematic complexity of the classification.

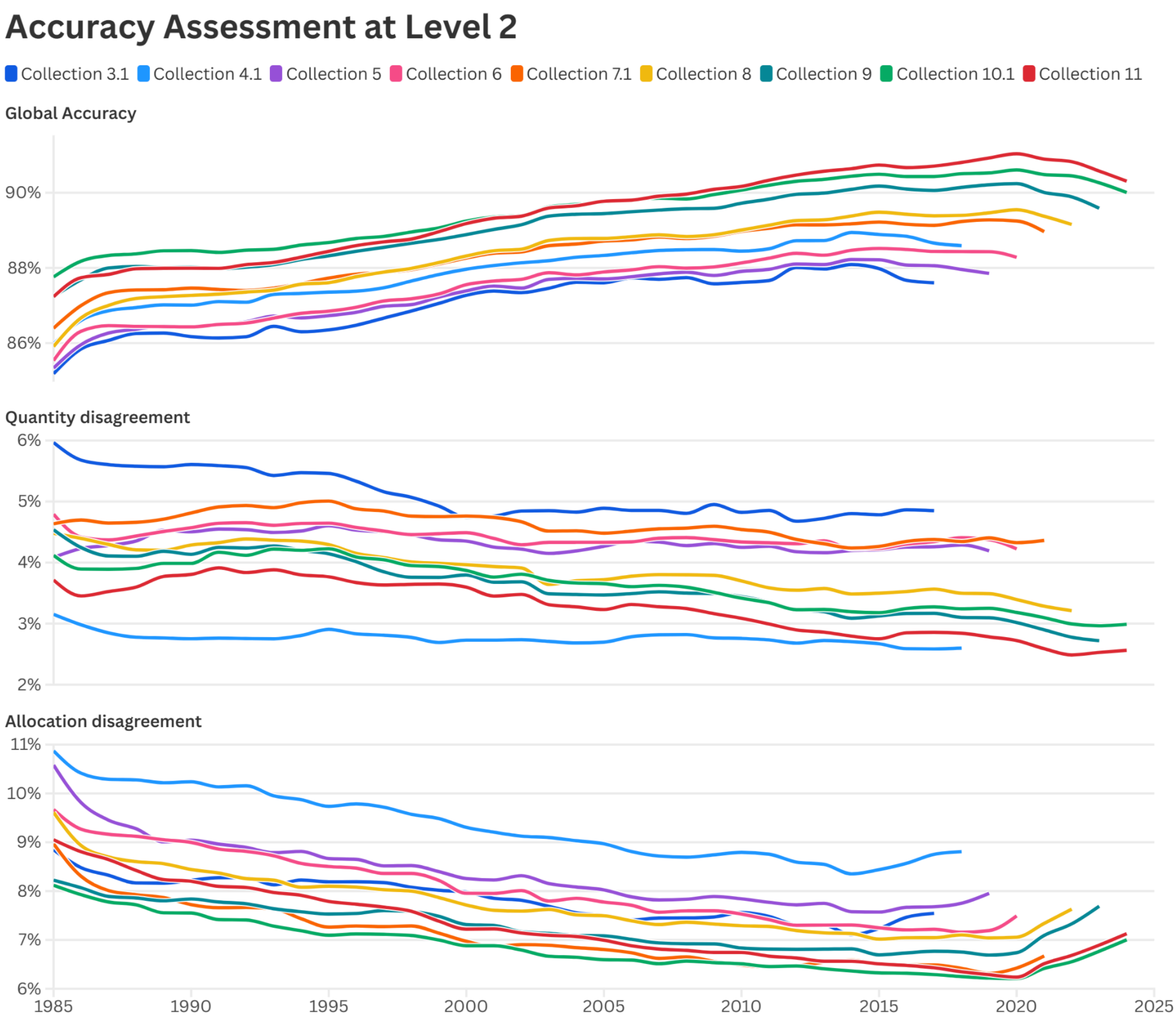


Figure 6. Comparison of Level 2 accuracy metrics across MapBiomas collections. The panels show global accuracy, quantity disagreement, and allocation disagreement over the years available for each collection.

4. Accuracy assessment of MapBiomas Collection 11

The global, producer’s and user’s accuracy for each level of the LCLU classes of the Collection 11 were calculated for each year, class, and biome (more details can be explored in the MapBiomas web platform: brasil.mapbiomas.org/en/analise-de-acuracia/) from 1985 to 2024. In Level 1 classes, the LCLU mapping product in the Collection 1 presented 93.4% mean global accuracy and 5.3 % allocation disagreement with 1.3% area disagreement. At legend level 1, the classes showed the following producer's and user's accuracy results: Forest (96.6% and 95.1%), Herbaceous and Shrubby Vegetation (80.2% and 75.9%), Farming (87.7% and 92.4%), Non vegetated area (86.3% and 89.8%), and Water (95.7% and 91.7%), respectively. At Level 2, the global accuracy was 90.2%, with 6.8% allocation disagreement and 3.0% area disagreement. The global accuracy was stable over the mapped period, varying across biomes from 84.1% to 97.8% in Level 1.

5. References

COCHRAN, W. G. Sampling techniques. 3. ed. New York: John Wiley & Sons, 1977.

PONTIUS JR, R. G.; MILLONES, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing, v. 32, n. 15, p. 4407–4429, 2011.

STEHMAN, S. V. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. International Journal of Remote Sensing, v. 35, p. 4923–4939, 2014. DOI: 10.1080/01431161.2014.930207.

STEHMAN, S. V.; FOODY, G. M. Key issues in rigorous accuracy assessment of land cover products. Remote Sensing of Environment, v. 231, p. 111199, 2019. DOI: 10.1016/j.rse.2019.05.018.