ACCURACY ASSESSMENT - Appendix 15

Collection 10

Version 1

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

Team
Vinicius Vieira Mesquita
Jairo Matos da Rocha
Leandro Parente
Luís Bauman
Tharles Andrade

August, 2025

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 10. 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	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. In addition, two sampling designs were used: one for the general, broad classes and a separate scheme for the categories that are of rare occurrence at the landscape scale.

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://mapbiomas.org/en/mapas-de-referencia?cama_set_language=en).

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.

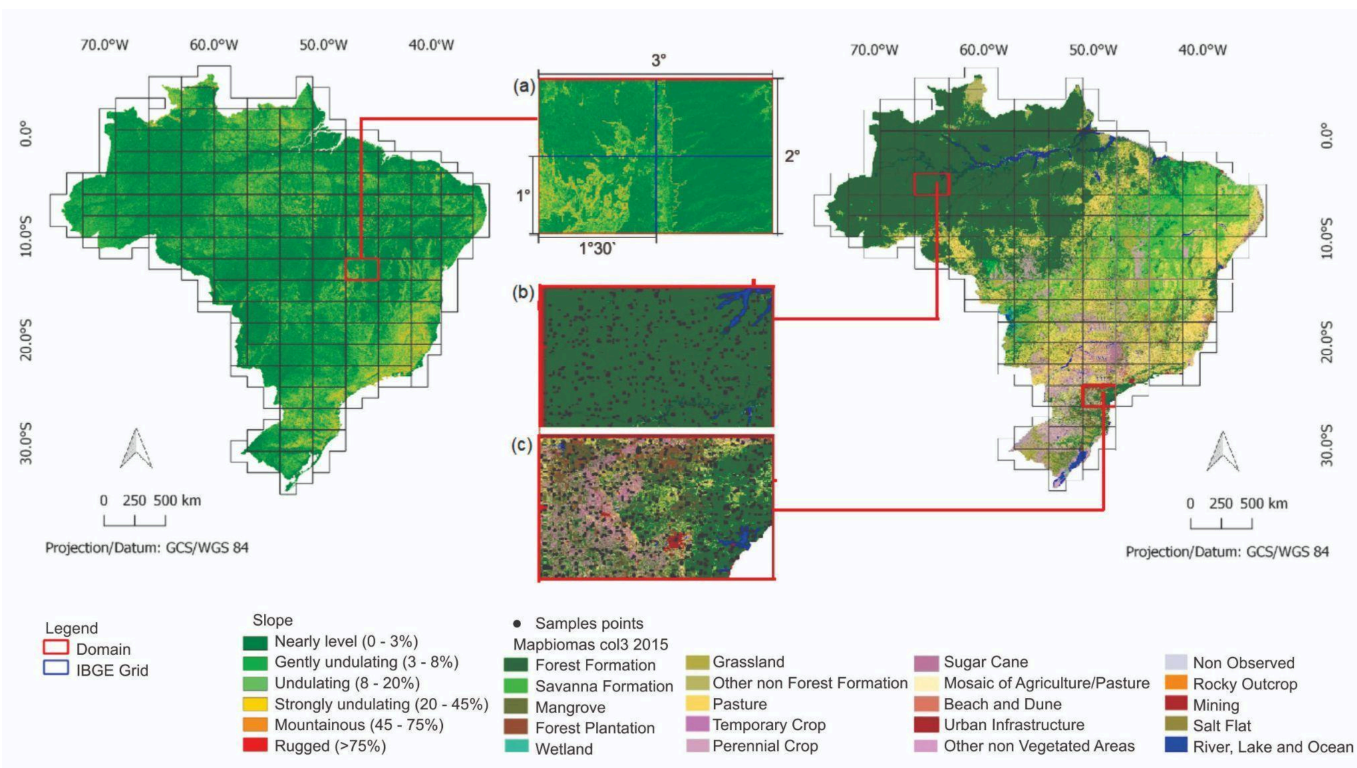


Figure 1. Slope categories used in sampling design for validation with independent points and group-chart size used as a subset in the sampling design for validation with independent points.

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:

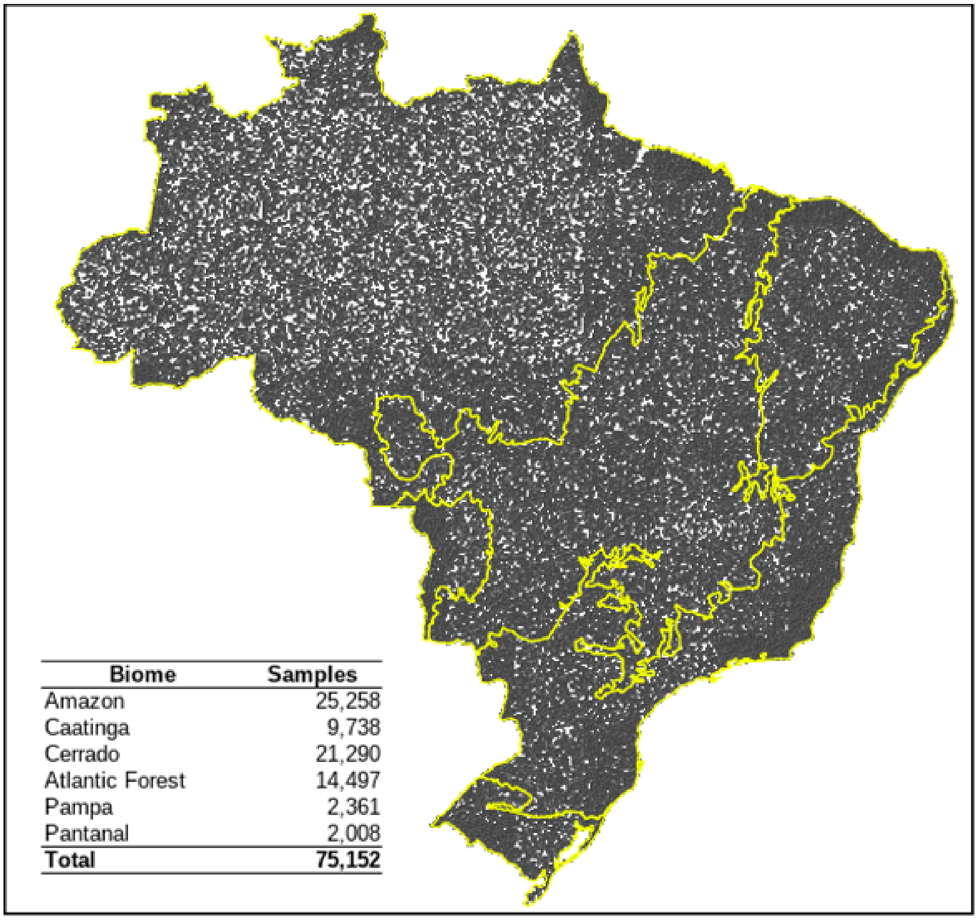
$$n_c = \frac{Nz^2 s_{max}^2}{(N-1)E^2 + z^2 s_{max}^2} > \frac{Nz^2 s_{max}^2}{NE^2 + z^2 s_{max}^2} = \frac{\left(\sum_h W_h s_h^2\right)^2}{\left(\frac{E}{z}\right)^2 + \frac{1}{N}\sum_h W_h s_h^2} = n_{PSS} , \qquad (1)$$

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 $s_{max}^2 = max_i\, p_i(1 - p_i)$ 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_h^2 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 Sampling design of Rare Classes

In the main accuracy assessment (performed for the general classes), some classes did not have enough samples to estimate accuracy metrics. These were then defined as “rare classes”, and are: Beach and Dune, Mining, Mangrove, Aquaculture, Hypersaline Tidal Flat, and Rocky Outcrop (Figure 3).

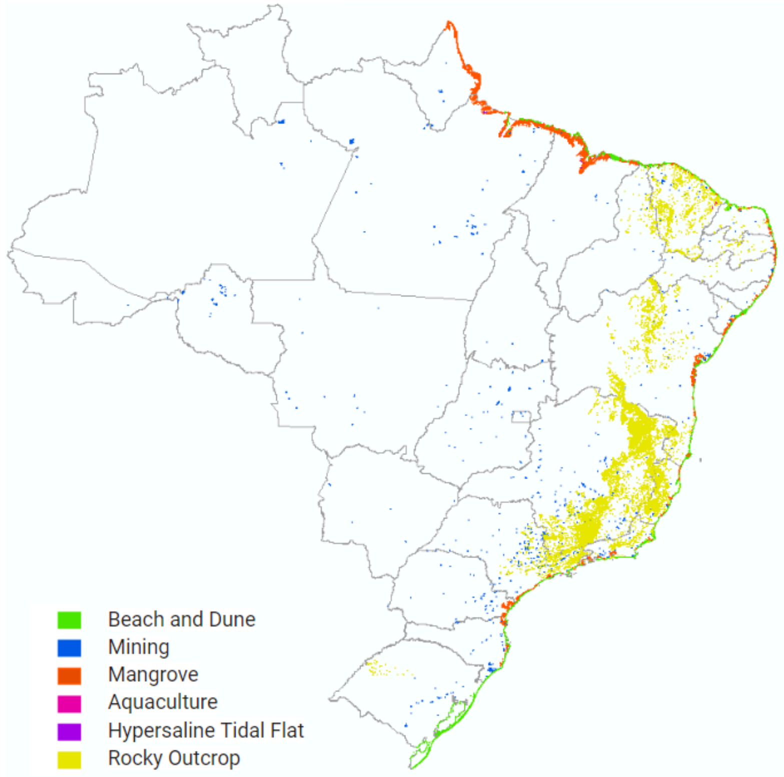


Figure 3. Map of rare classes in the MapBiomas Land Cover and Land Use maps.

To estimate accuracy metrics for these rare classes, we did a stratified random sampling with disproportionate allocation within new spatial units. These new units were created based on the sum of the mapped area for each rare class from 1985 to 2018 (i.e. accumulated maps), using MapBiomas Collection 7. In addition, we created buffer zones for each accumulated class map based on twice the size of each polygon (Figure 4). The buffer zones of the coastal classes (i.e. Mangrove, Beach and Dunes, Hypersaline Tidal Flat, Aquaculture) were merged into a single map given their spatial proximity (Table 2, buffer). The other two buffers correspond to the accumulated areas of Mining and Rocky Outcrop.

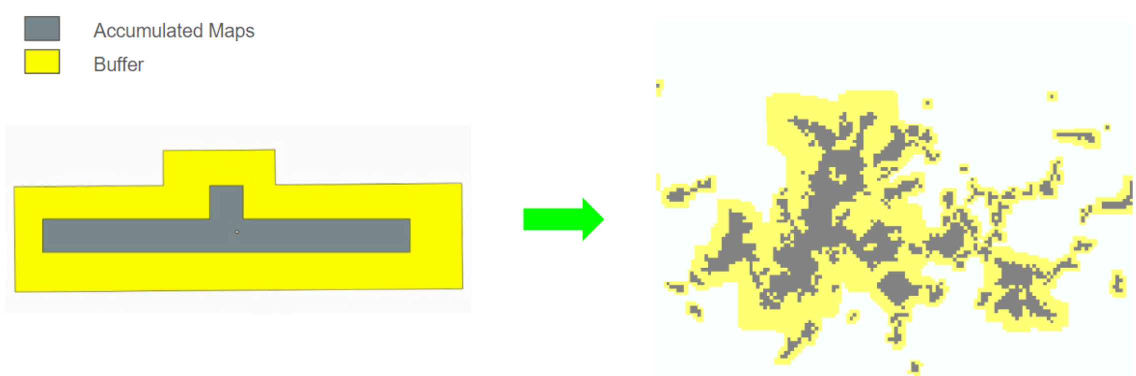


Figure 4. Spatial units used in the sampling design of the rare classes.

The total number of samples n_{rc} for the sampling scheme of rare classes was calculated using the formula (Cochran, 1977) :

$$n_{rc} = \frac{\sum_h \frac{W_h^2 s_h^2}{w_h}}{\left(\frac{E}{z}\right)^2 + \frac{1}{N}\sum_h W_h s_h^2},$$

where n_{rc} is the total sample size; W_h is the proportion of stratum *h* within the population and w_h is the proportion of the sample allocated to stratum h; $s_h^2 = p(1 - p)$ is the variance in the strata h such that the proportion p is what we want to estimate and $p(1 - p)$ was considered equal to 0.25, which is the maximum possible variability in the binary case; E is the expected maximum margin of error; z is the quantile of the standardized normal distribution corresponding to the confidence level $1 - \alpha$; N is the total number of points. In this way, the total sample amount of $n_{rc} = 3,600$ was acquired and distributed by stratum as shown in Table 2.

Table 2. Totals and distribution of the number of samples within each spatial unit.

Rare Classes	Samples Accumulated Map	Samples Buffers
Mangrove	450	600
Hypersaline Tidal Flat	450	
Beach and Dune	450	
Aquaculture	450	
Mining	450	150
Rocky Outcrop	450	150
TOTAL	2700	900
	3600	

2.2.3 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 5), 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).

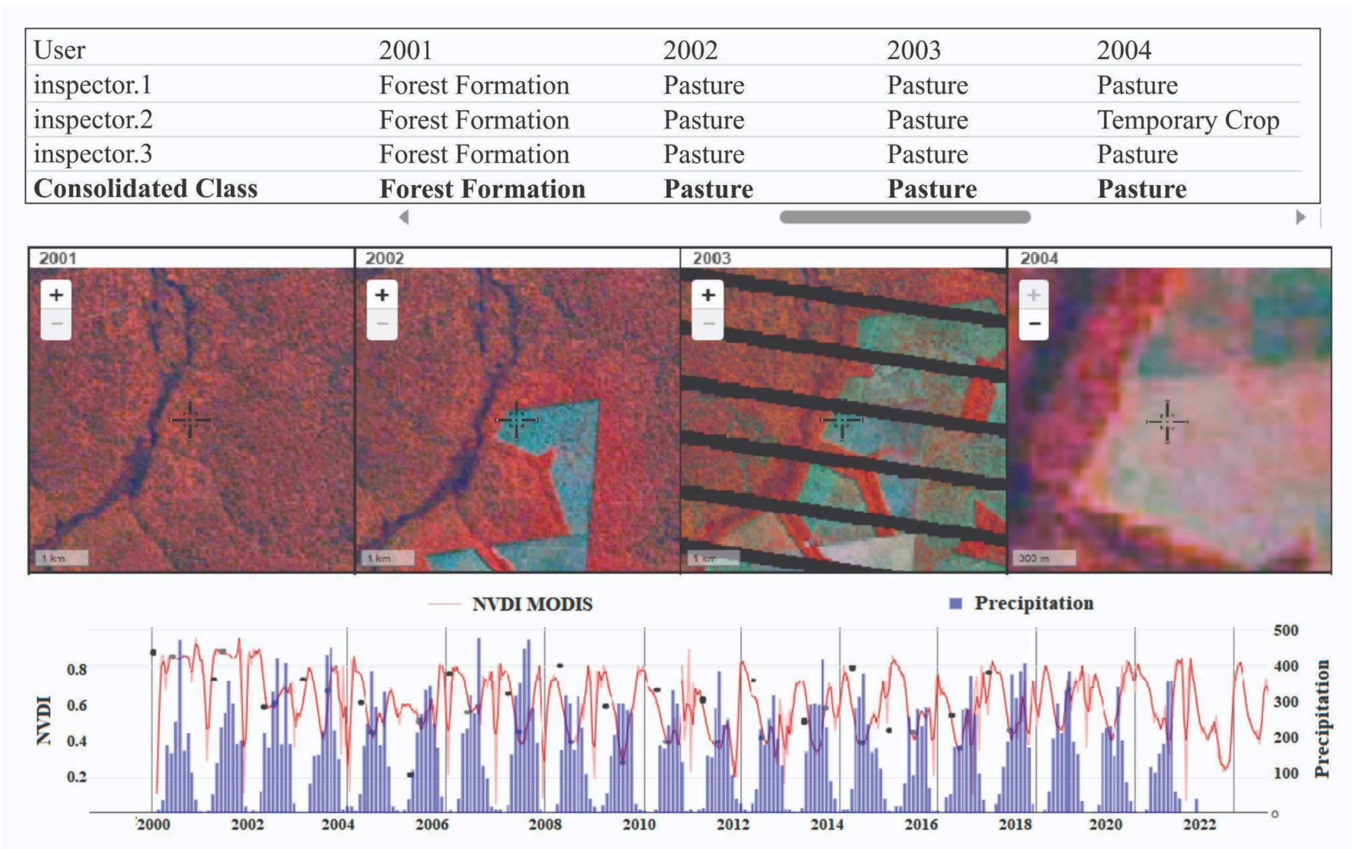


Figure 5. Temporal Visual Inspection (TVI)

2.2.4 Accuracy metrics

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 Collection 10

The global, producer's and user's accuracy for each level of the LCLU classes of the Collection 10 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 10 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. Finally, at Level 3, the global accuracy was 90.2%, with 6.7% allocation disagreement and 3.1% area disagreement. The global accuracy was stable over the mapped period, varying across biomes from 84.1% to 97.8% in Level 1.

4. 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.