



MAPBIOMAS

Projeto de Mapeamento Anual
da Cobertura e Uso do Solo

Nota Metodológica

V# 1.0

(em construção)

23.11.2015

O **MapBiomass** é uma iniciativa do SEEG/OC (Sistema de Estimativas de Emissões de Gases de Efeito Estufa do Observatório do Clima) e é produzido por uma **rede colaborativa de co-criadores** formado por ONGs, universidades e empresas organizados por biomas e temas transversais.

Coordenação nos Biomas

- Amazônia – Instituto do Homem e do Meio Ambiente da Amazônia (IMAZON)
- Caatinga – Universidade Estadual de Feira de Santana (UEFS) e Associação Plantas do Nordeste (APNE)
- Cerrado – Instituto de Proteção Ambiental da Amazônia (IPAM)
- Mata Atlântica – Fundação SOS Mata Atlântica e ArcPlan
- Pampa – Universidade Federal do Rio Grande do Sul (UFRGS)
- Pantanal – Instituto SOS Pantanal e ArcPlan

Coordenação Temas Transversais

- Pastagem- Universidade Federal de Goiás (LAPIG/UFG)
- Agricultura – Agrosatélite
- Zona Costeira - Instituto Tecnológico Vale

Parceiros de Tecnologia

- Google
- Ecostage
- Terras App
- ArcPlan

Financiamento

- Iniciativa Internacional de Clima e Florestas da Noruega (NICFI)
- Gordon & Betty Moore Foundation
- Instituto Arapyaú

Parceiros Institucionais

- Instituto Democracia e Sustentabilidade (IDS)
- Fundação AVINA

Coordenação Técnico Científica – Carlos Souza Jr.

Coordenação Geral - Tasso Azevedo

Processamento e Classificação de Imagens Landsat do Projeto MapBiomass

O projeto MapBiomass utiliza imagens Landsat (5, 7 e 8) para mapear e quantificar mudanças na cobertura do solo dos biomas brasileiros. O processamento das imagens Landsat é conduzido na plataforma Google Earth Engine, baseada em nuvens de computadores, e segue cinco etapas para o processamento digital das imagens, descritas abaixo (**Figura 1**). Essas etapas foram implementadas num programa Web denominado MapBiomass Workspace (**Figura 2**).

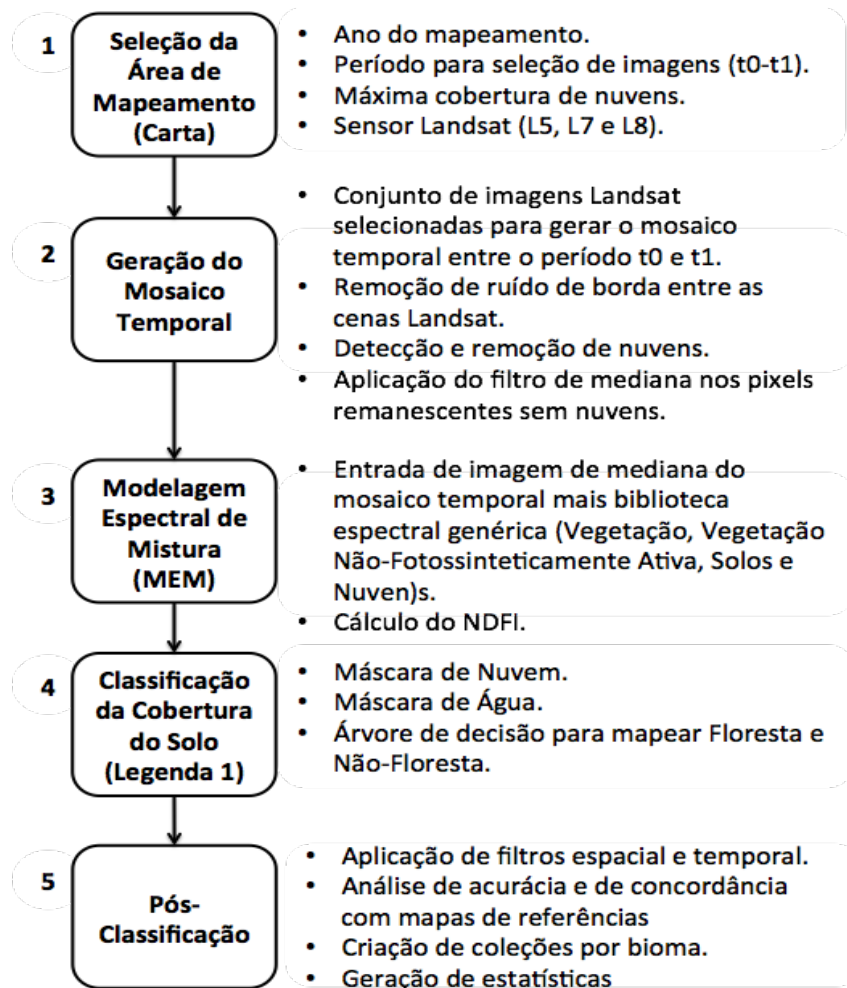


Figure 1. Etapas de processamento digital das imagens Landsat do MapBiomass para o mapeamento da cobertura do solo.

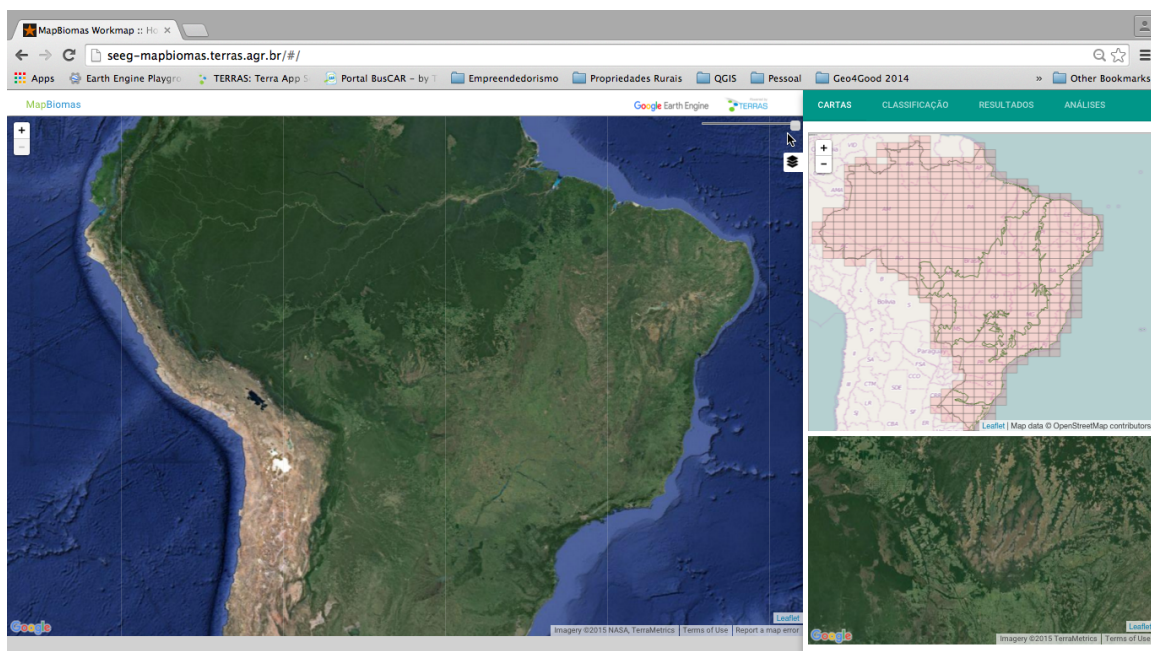


Figure 2. Interface do programa Web MapBiomias que implementa as etapas de processamento digital das imagens Landsat para o mapeamento da cobertura do solo.

Etapa 1 - Seleção da Área de Mapeamento

A unidade de mapeamento do MapBiomias é a Carta Internacional do Mundo ao Milionésimo na escala 1:250.000, cobrindo uma área de 1°30' de longitude por 1° de latitude (**Figura 3**). Um total de 558 cartas (ou folhas) cobre o território nacional. Para cada carta, são definidos os seguintes parâmetros das imagens Landsat:

- i) o ano de mapeamento (de 2008 a 2015 para essa fase do projeto);
- ii) o período no ano para seleção de imagens (t0 e t1 em dia/mês/ano) que possibilita o máximo contraste espectral entre as classes de florestas e não florestas;
- iii) a máxima cobertura de nuvens da imagem Landsat que será utilizada no processamento; e
- iv) o tipo de sensor (Landsat 5 [L5], Landsat [L7], e Landsat 8 [L8]). Esses parâmetros são obtidos do arquivo de metadados das coleções de imagens Landsat disponíveis na plataforma Earth Engine.

Utilizamos coleções de imagens de reflectância de superfície ortorretificadas (L1T) para L5 e L7, reflectância no topo da atmosfera para L8. As cartas que recobrem mais de um bioma são processadas separadamente, com os parâmetros dos seus respectivos biomas, e combinadas na etapa de pós-classificação (Etapa 5 da Figura 1). A distribuição das cartas por biomas é mostrada na Figura 3.

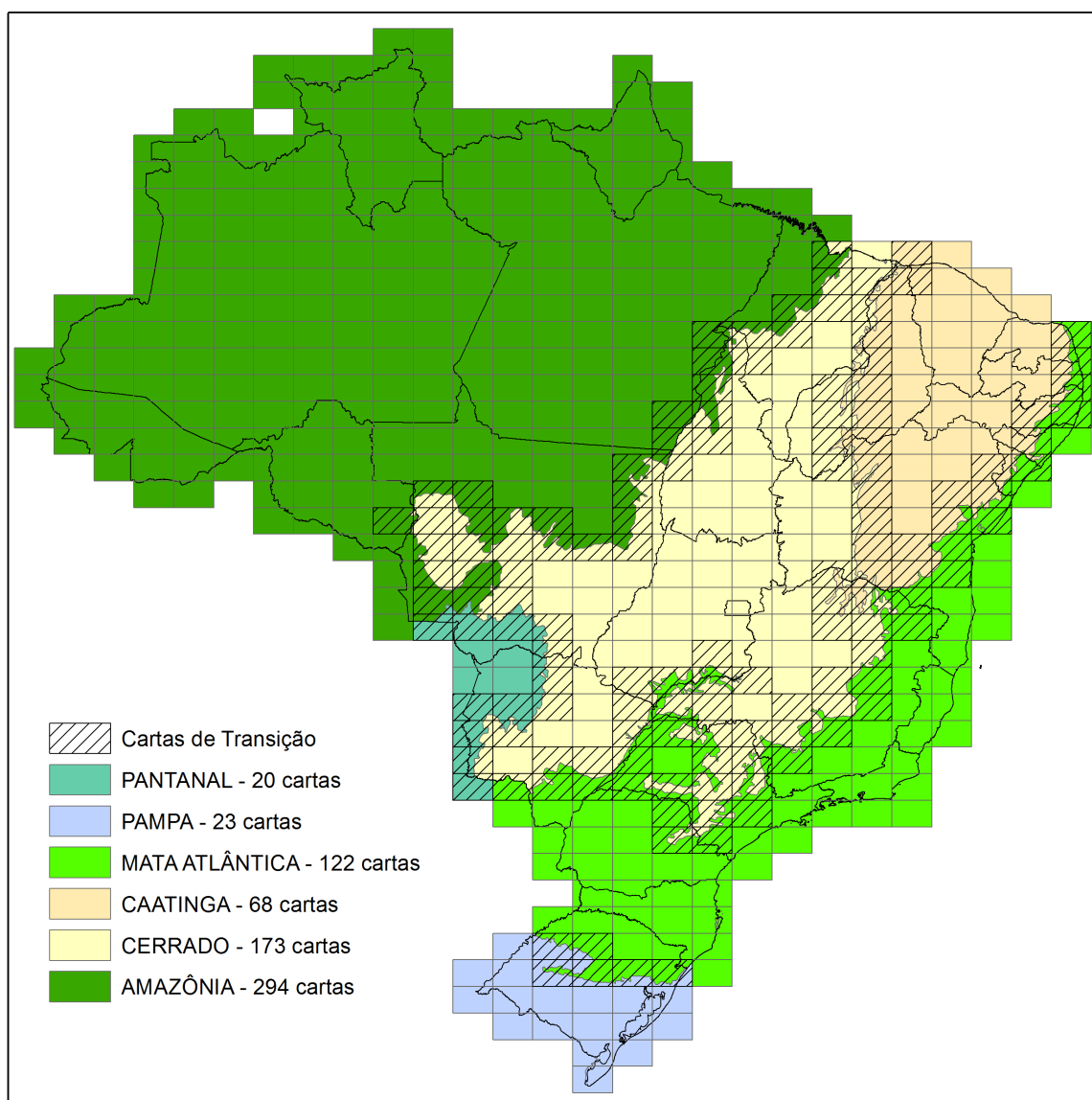


Figure 3. Distribuição das cartas utilizadas no processamento das imagens Landsat no ambiente Workspace do MapBiomos.

Etapa 2 - Geração do Mosaico Temporal

A aplicação dos parâmetros descritos acima (período e cobertura de nuvens) permite obter um conjunto de cenas de imagens Landsat para um dado ano, no período definido, cobrindo a área da carta selecionada para o processamento. Esse conjunto de cenas Landsat é processado com o objetivo de gerar um mosaico [temporal] com a menor cobertura de nuvens para a carta, otimizando o período de maior contraste espectral das classes de cobertura do solo que serão mapeadas (ver Legenda).

As cenas Landsat selecionadas são processadas para gerar o mosaico temporal que cobre a área da carta. Inicialmente, aplica-se uma zona de entorno (*buffer*) ao longo da cena da imagem Landsat para remoção de eventuais ruídos de borda. São considerados ruídos os píxeis da cenas Landsat que não possuem dados em todas as bandas. O próximo passo é aplicar um algoritmo de detecção e remoção de nuvens, baseado na modelagem espectral de mistura (MEM; ver Etapa 3 para detalhes sobre esse método), e imagens da faixa termal do espectro electromagnético (Banda 6 para o L5 e L7, e Banda 10 para o L8; todas convertidas para a escala Celsius de temperatura).

Píxeis com fração de nuvem maior ou igual a 10% e temperatura menor ou igual a 22 graus Celsius são detectados inicialmente como nuvem. Uma região de entorno com 10 píxeis (300 metros, considerando um píxel Landsat com 30m) é gerada ao longo das áreas detectadas inicialmente como nuvem, utilizando um filtro matricial (*kernel*) circular. Esse procedimento permite detectar áreas com sombra gerada pelas nuvens originalmente detectadas. Nessas áreas o percentual de fração de nuvem no pixel diminui e, por essa razão, todos os píxeis, no entorno das nuvens, com valores maiores ou iguais a 7% de fração de nuvem são adicionados às áreas inicialmente classificadas como nuvem.

Esse processo gera as áreas com nuvens, para todas as cenas Landsat que compõem a área da carta, no ano e período selecionados. As áreas classificadas como nuvem são ignoradas no processamento subsequente através da geração de uma máscara aplicada a essas regiões. Por último, um filtro de mediana é aplicado para os píxeis não cobertos por nuvens nas imagens, para compor a imagem final da carta, denominada de mosaico temporal. Mantivemos o período de seleção de imagens menor possível (i.e., em torno de três meses) para reduzir mudanças significativas na cobertura florestal e aumentar a área da superfície observada a partir do processo de filtragem de nuvens descrito acima (Figura 4).



Figure 4. Exemplo de mosaico temporal aplicado a carta SG-22-Z-C no ano de 2009, para o período de 01/06/2009 (t0) a 31/10/2009 (t1), no bioma Mata Atlântica. As cenas selecionadas mostram a cobertura de nuvens da imagem de outubro/2009 (a); em (b), observa-se o resultado da detecção e aplicação da máscara de nuvem e da aplicação do filtro de mediana.

Etapa 3 – Modelagem Espectral de Mistura (MEM)

A técnica MEM permite estimar a composição do píxel em termos de componentes (assinaturas espectrais) puros. O MEM aplicado no MapBiomas utiliza uma biblioteca espectral definida por cinco componentes puros: Vegetação, Vegetação Não-Fotossinteticamente Ativa, Solo e Nuvem (Figura 5). Essa biblioteca espectral foi definida originalmente para o Bioma Amazônia¹.

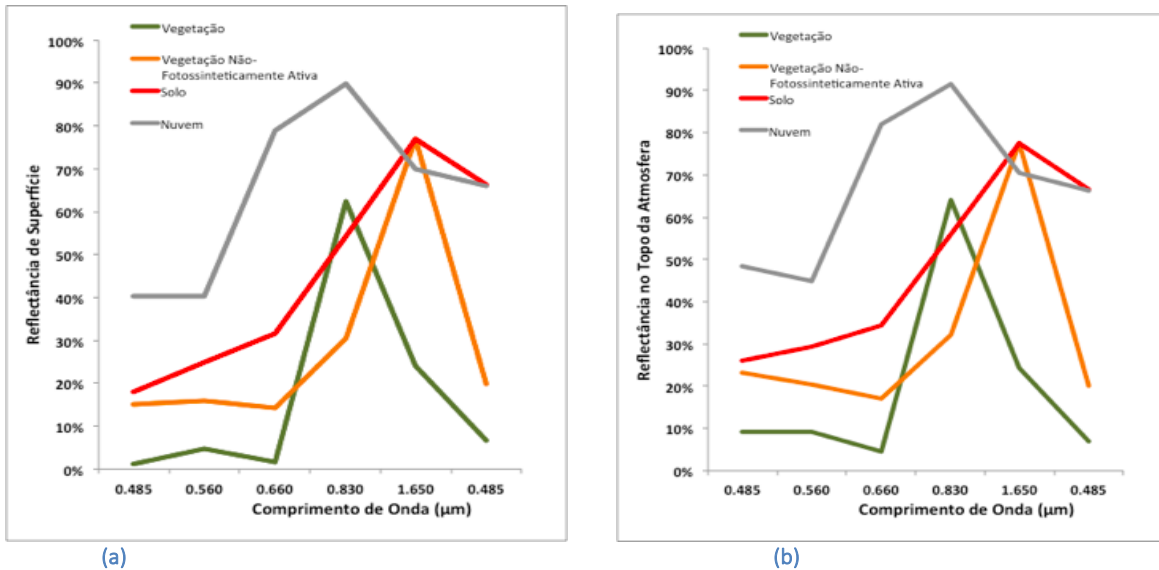


Figure 5. Curvas espectrais dos componentes puros utilizados no MME aplicados aos biomas brasileiros; (a) para Landsat 5 e 7 em reflectância de superfície, (b) para Landsat 8 para reflectância no topo da atmosfera.

A MME assume que a reflectância de cada píxel, em cada banda da imagem Landsat, resulta da combinação linear do produto do percentual do componente puro (i.e., proporção no píxel) pelo seu respectivo valor na banda, com base na equação:

$$R_b = F_i R_{i,b} + \epsilon_b \quad (1)$$

Para $\sum F_i = 1 \quad (2)$

onde R_b é a reflectância na band b , $R_{i,b}$ é a reflectância do componente puro i , na banda b , F_i a fração (ou proporção) do component puro i , cuja a soma é 1 (ou 100%), e ϵ_b é o erro residual de cada banda.

¹ Souza Jr., C. M., Roberts, D. and Cochrane, M. A. 2005. Combining Spectral and Spatial Information to Map Canopy Damages from Selective Logging and Forest Fires. **Remote Sensing of Environment**. 98. 329-343.

O MME é resolvido para estimar os valores F_i (i.e., a fração ou proporção do component puro), obtendo, dessa forma, imagens Fração para Vegetação, Vegetação Não-Fotossinteticamente Ativa, Solo e Nuvem, as quais foram utilizadas para o cálculo do índice espectral NDFI, obtido pela diferença normalizada da imagem fração Vegetação e da soma das imagens fração Vegetação Não-Fotossinteticamente Ativa e Solo, de acordo com a equação abaixo:

$$NDFI = \frac{Vegetação_s - (Vegetação NãoFotossintética Ativa + Solo)}{Vegetação_s + Vegetação NãoFotossintética Ativa + Solo}$$

onde,

$$Vegetação_s = \frac{Vegetação}{100 - Sombra}$$

e,

$$Sombra = 1 - \sum F_i$$

As imagens fração e o NDFI são utilizadas como parâmetros de entrada para classificação da cobertura do solo (ver Etapa 4 abaixo, e Legenda) e para gerar máscara de nuvem para gerar o mosaico temporal (Etapa 2).

Etapa 4 – Classificação da Cobertura do Solo

As imagens fração obtidas com o MME são utilizadas na classificação da cobertura do solo para as classes do nível 1 da Legenda de Classificação, incluindo: Floresta, Não-Floresta, Água e Nuvem. A classificação é conduzida seguindo três passos (Figura 1). Primeiro, aplicamos a máscara de nuvens descrita acima, para mapear áreas cobertas por nuvens que não puderam ser filtradas durante o processo de geração do mosaico temporal da carta. Segundo, aplicamos a máscara de água com base nos seguintes critérios: píxeis com fração de Sombra maior que 75%, e Vegetação menor que 10% e Solo menor que 5%, e que não esteja em áreas de sombreamento do relevo (detectamos a sombra de relevo, usando o modelo *Hill Shadow* e dados de iluminação de cada cena disponíveis em seus metadados). Por último, aplicamos uma árvore de decisão (definida empiricamente) para mapear as áreas de Floresta e Não-floresta. As máscaras de Nuvem e de Água são combinadas com o mapa resultante da árvore de decisão para gerar o mapa final (Figura 6).

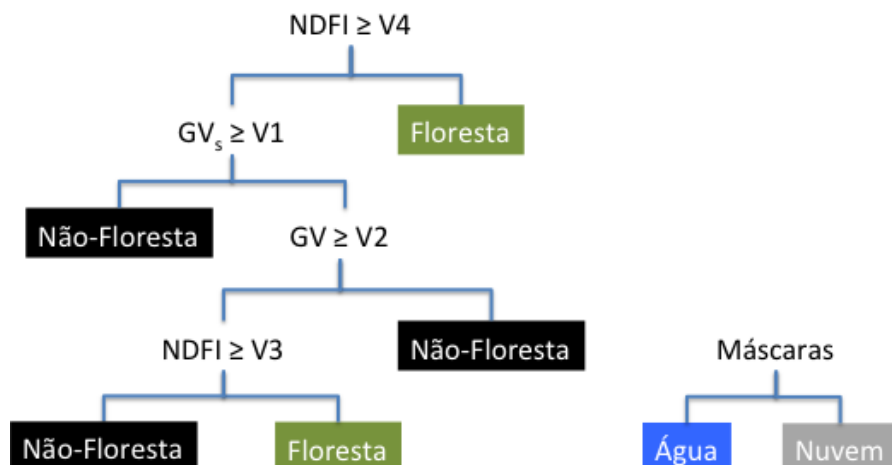


Figure 6. Árvore de decisão empírica aplicada a NDFI, GV e GV_s para mapear as classes de Floresta e Não-Floresta. Máscaras de Água e Nuvem são aplicadas para compor o mapa final. As variáveis (V1, V2, V3, V4) podem ser ajustadas para cada carta.

Etapa 5 – Pós-Classificação

Os mapas gerados na Etapa de Classificação da Cobertura do Solo passam por processos de filtragem espacial e temporal. O filtro espacial segmenta e indexa as classes Floresta, Não-Floresta e Água em regiões contíguas que são posteriormente identificadas e reclassificadas com base no seguinte critério: áreas menores ou iguais a meio hectare (i.e., aproximadamente 5 píxeis) são reclassificadas com base na maioria das classes vizinhas. Dessa forma, por exemplo, um segmento da classe Não-Floresta com até 5 píxeis é primeiramente identificado juntamente com píxeis vizinhos. Esse segmento é então reclassificado com o valor predominante da classe dos píxeis vizinhos (Figura 7). Esse processo é aplicado para todos os segmentos das classes selecionadas para a filtragem (i.e. Floresta, Não-Floresta e Água).

O próximo passo é utilizar informação temporal para identificar transições de classes não permitidas entre anos consecutivos. Por exemplo, se um píxel for classificado como Floresta num dado ano t_i (com $i = 2008, 2009, \dots, 2015$), e Não-Floresta no ano t_{i+1} e em t_{i+2} , esse píxel é reclassificado como Não-Floresta no ano t_i . Há várias regras de transição aplicadas pelo filtro temporal, listadas na Tabela 1.

Após a aplicação dos filtros espacial e temporal, os mapas de cobertura do solo resultantes são avaliados segundo dois critérios. Primeiro, aplicando técnicas estatísticas de análise de acurácia baseada em dados de referências coletados na Plataforma *Collect Earth*². Segundo, conduzimos análises de concordância espacial com mapas [de referência] dos biomas publicados por outras iniciativas. Os mapas de referência foram selecionados para cada bioma com base na disponibilidade. Uma grade de 500m x 500m

² Esse processo de análise de acurácia está em andamento.

foi criada para comparar a proporção da área de floresta classificada pelos mapas do MapBiomas e dos mapas de referências. Essas células da grade de análise são classificadas em três níveis de concordância: Alta (>80%), Média (80-60%) e Baixa (>60-5%).

Tabela 1. Regras de reclassificação aplicadas pelo filtro temporal.

Regras		Entrada (Ano)				Saída			
Gerais		t_i	t_{i+1}	t_{i+2}	t_{i+5}	t_i	t_{i+1}	t_{i+2}	t_{i+3}
RG-1		NF	F	F	F	F	F	F	F
RG-2		NF	F	NF	NF	NF	NF	NF	NF
RG-3		NF	SI	*	*	NF	NF	*	*
RG-4		NF	A	NF	NF	NF	NF	NF	NF
RG-5		F	SI	*	*	F	F	*	*
RG-6		F	A	F	F	F	F	F	F
RG-7		F	NF	F	F	F	F	F	F
RG-8		A	F	F	F	F	F	F	F
RG-9		A	NF	NF	NF	NF	NF	NF	NF
RG-10		A	*	A	A	A	A	A	A
RG-11		A	SI	*	*	A	A	*	*
RG-12		*	NF	SI	*	*	NF	NF	*
RG-13		*	F	SI	*	*	F	F	*
RG-14		*	A	SI	*	*	A	A	*
RG-15		*	*	NF	SI	*	*	NF	NF
RG-16		*	*	F	SI	*	*	F	F
RG-17		*	*	A	SI	*	*	A	A
Específicas		2012	2013	2014	2015	2012	2013	2014	2015
RE-1		NF	NF	F	NF	NF	NF	NF	NF
RE-2		NF	NF	A	NF	NF	NF	NF	NF
RE-3		F	F	NF	F	F	F	F	F
RE-4		F	F	A	F	F	F	F	F

NF = Não floresta; F = Floresta; A = Água; SI = Sem informação ou Nuvem; * = Qualquer classe; $i = \{2008, 2009, 2010, \dots, 2015\}$.

Por último, estatísticas de área das classes mapeadas são obtidas para Biomas, Estados, Municípios, e outras regiões de interesse que serão definidas futuramente, como bacias hidrográficas, áreas protegidas, bem como polígonos definidos por usuários.

Classificação das Áreas Convertidas (Agricultura, Pastagens e Florestas Plantadas)

À exemplo dos outros temas sendo mapeados no âmbito do projeto MapBiomas, as áreas de agricultura, pastagens e florestas plantadas do Brasil, em sua totalidade, estão sendo identificadas e mapeadas com base em imagens das séries Landsat 5, 7 e 8 (sensores TM, ETM+ e OLI, respectivamente), tipo L1T (i.e. corrigidos em relação ao terreno e com o maior nível de acurácia espacial possível), convertidas para valores de reflectância aparente (i.e. *top-of-atmosphere*), o que garante uma maior padronização e consistência radiométrica entre cenas / datas.

Todas as imagens (cenas / passagens) são avaliadas, píxel a píxel, quanto à ocorrência de nuvens, sendo descartadas todas as observações com alta probabilidade de contaminação por nuvens, i.e. todas as observações cujos valores indicados pelo *Google Earth Engine simple cloud score* sejam superiores à 40% e/ou com indicativo de presença de nuvens conforme o Landsat *band quality assurance*.

Após esta triagem, cada um dos aproximadamente 63 bilhões de píxeis compreendidos nas 380 cenas Landsat que recobrem o Brasil está sendo associado a uma série temporal, considerando, em dois anos consecutivos, as observações bimestrais de melhor qualidade. Esta abordagem, inédita e que constitui um produto para várias outras aplicações, tem por pressuposto fundamental de que a área de uso da terra no Brasil não sofre alterações significativas em um período de dois anos. Assim, e considerando a resolução temporal de 16 dias das imagens Landsat, cada pixel pode ser associado a até oito observações em um período de dois anos e intervalos de dois meses.

Tendo por referência o método do *maximum value composite* utilizado nos produtos MODIS de índice de vegetação, onde a melhor observação possível é aquela que corresponde ao maior valor do índice de vegetação da diferença normalizada (NDVI), cada um dos bilhões de píxeis distribuídos pelo Brasil é associado a um ciclo hidrológico completo para um determinado período (ex. 2007 - 2008), representado por seis valores NDVI bimestrais (i.e. janeiro-fevereiro, março-abril, maio-junho, julho-agosto, setembro-outubro, novembro-dezembro). E, uma vez definida a data de passagem do satélite que corresponde à melhor observação para cada um destes seis intervalos, outros dois índices e respectivas séries temporais para cada período considerado, são calculados: índice de stress hídrico da diferença normalizada (NDWI) e índice de celulose (CI). Ao todo, cada pixel é associado a um conjunto de 27 variáveis, i.e. 18 variáveis relacionadas às três séries temporais NDVI, NDWI e CI, além do desvio padrão e duas métricas de contraste sazonal para cada uma destas séries.

Com base nestas 27 variáveis e no algoritmo Random Forest, estamos identificando e classificando as áreas de agricultura, pastagem e floresta plantada para todo o território nacional. Random Forest é um método de aprendizado (*machine learning*) que combina vários modelos (decision trees) para resolver um único problema de predição (no nosso caso, decidir se um determinado pixel pode ou não ser atribuído a uma destas três classes). Por se tratar de um método de classificação supervisionada, o Random Forest necessita de amostras de treinamento, as quais estão sendo coletadas separadamente para cada uma das 380 cenas Landsat, tendo por referência a área de pastagem, agricultura e floresta plantada compiladas pelo LAPIG UFG e Agrosatélite.

Em cada uma das cenas Landsat, é coletado um conjunto de até 2000 pontos aleatórios, i.e., até 1000 pontos distribuídos em áreas previamente identificadas conforme os alvos de interesse (pastagem, agricultura ou floresta plantada) e até 1000 pontos em áreas supostamente associadas a outros tipos de alvos (p.ex. água), com base no qual o algoritmo Random Forest é “treinado” e executado. O resultado corresponde, para cada pixel da cena, à probabilidade de 0 a 100% do pixel em questão pertencer a uma das três classes de interesse. Todo este processo é repetido 10 vezes, obtendo-se assim a probabilidade média de cada pixel poder ser associado à classe de interesse, o que por sua vez, dependerá do limiar escolhido. Após uma análise de sensibilidade que levou em consideração o balanço entre erros de omissão e comissão em diferentes cenas, optou-se por um limiar de 60%, i.e. todo pixel cujo resultado da classificação é maior ou igual a 60% é atribuído, conforme o caso, a uma área de pastagem, agricultura ou floresta plantada. A acurácia desses mapas é estimada com método de validação-cruzada e registrada (ou relatada) separadamente.

Integração das classes naturais e convertidas

A integração dos mapas de Floresta e Não-Floresta com os mapas de Agricultura, Pastagem e Floresta Plantada segue regras de prevalência definidas na Figura 8. A aplicação dessas regras resulta no mapa final de classificação para o Nível 1 da Legenda de Classificação (ver Quadro da Legenda).

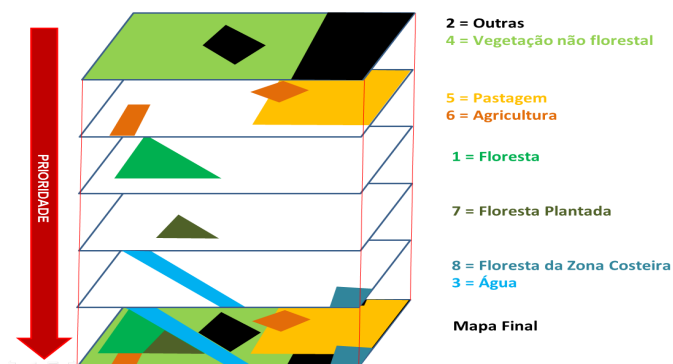


Figure 8. Prioridade para integração das classes.