Mining – Appendix

Collection 8

Version 1

General coordinator
Pedro Walfir Martins e Souza Filho

Team
Alexandre Filho
Cesar Guerreiro Diniz
Luis Waldir Rodrigues Sadeck
Luiz Cortinhas Ferreira Neto
Maria Luize Silva Pinheiro
1 Overview

Today, Brazil is among the five largest producers of iron ore, niobium, bauxite, and manganese in the world (Bray, 2020), exporting a variety of mineral inputs, with a high level of purity and internationally recognized quality. Despite its low representativeness in area, mining expansion has been rising across the country, reaching ~350,000 hectares in 2021, a value ~7 times higher than what was reported in 1985 (~50,000 hectares).

In the new Collection 8, mining mapping carries the same method as in Collections 6 and 7. Still, it includes updates on the quantity and quality of the training samples and a larger grid size used in the processing steps of the mining recognition algorithm. The method still uses U-Net (Ronneberger et al., 2015), a CNN classifier based on Deep Learning. The stack of reference data now includes information from additional sources: CPRM (Brazilian Geological Service), AhkBrasilien (Brazil-Germany Chamber of Commerce and Industry), INPE (National Institute for Space Research), and ISA (Instituto Socioambiental) and AMW (Amazon Mining Watch). Details regarding the classification process are described below in Figure 1 and in the github: https://github.com/mapbiomas-brazil/mining.

![Processing diagram](image)

Figure 1 – Processing diagram. The steps related to image processing are in blue. The steps in green are related to the sample design. Classification procedures are in yellow. The accuracy assessment phase is in red. BQA states for Band Quality Assessment.
2 Landsat Mosaics

The classification of the cross-cutting theme “Mining” uses Landsat “Top of Atmosphere - TOA” mosaics, which differs from the “Surface Reflectance - SR” processing level used in the land use and land cover classification of the Brazilian biomes. The Landsat mosaics prepared for the mining mapping were cropped to comprise areas where mining sites are known to exist. These mosaics are the third generation of the methodology developed specifically for these cross-cutting themes.

The annual cloud-free mosaics were generated in Google Earth Engine (GEE) platform, with all raster data and sub-products derived from the United States Geological Survey (USGS) Landsat Collection 2 Tier 1 Top of Atmosphere (TOA) imagery, which includes Level-1 Precision Terrain (L1TP).

2.1 Temporal coverage

The annual cloud-free composites used in the mining classification are generated by calculating the median pixel value of all images available in the GEE image collection for January 1 to December 31 of each year.

2.2 Mosaic Subsets
2.2.1 Mining

For each year, Landsat Collection 2, Tier 1, TOA data were used to produce annual cloud-free composites of imagery acquired from January 1st to December 31st. The quality assessment (QA) band and a median filter remove clouds and shade from the imagery. QA values improve data integrity by indicating which pixels might be affected by artifacts or subject to cloud contamination. In addition, we use a GEE function that gets the median pixel value of an image stack (i.e., the entire image collection available for a predefined area and dates of interest). This function rejects values that are too bright (e.g., clouds) or too dark (e.g., shadows), returning the median pixel value (of all images available in our stack) in each band for each year of our time series. Then, the annual mosaics were clipped to grid polygons that are known to have mining activity according to our reference dataset and excluded large areas where these activities are not expected to occur.

2.2.2 Reference Data

Brazil, especially the Brazilian Amazon (BA), has many publicly available datasets, from geological surveys and change detection platforms to deforestation early-warning systems. Mining data availability is highly diverse in scale, type, and timeframe. Spatially explicit data may be found at a higher or lower resolution, with a greater or lesser degree of human intervention, for scientific or journalistic use, but out of which a great set of spatial

Table 1 – Reference data used in our products. References were visually analyzed and converted to bounding boxes.

<table>
<thead>
<tr>
<th>Class</th>
<th>References</th>
</tr>
</thead>
</table>

Reference data were visually analyzed and converted to bounding boxes (Figure 2), which were overlaid on grids that are used to process the deep-learning mining recognition algorithm in a parallel fashion. In Figure 2, in yellow, are the grids used in Collection 8. The number of searching grids has increased compared to the previous collection.

Figure 2 – Reference sites are in red. The yellow grids are present exclusively in Collection 8.
3 Classification

The semi-automatic classification of Landsat mosaics was performed entirely on local servers, based on a U-Net classifier, a deep learning model. Once the sample collection is finished, the U-net classification results in the pre-filtered classification product. The classified data is injected back into GEE, where spatial-temporal filters and visual inspection occur, Figure 3.

![Figure 3 - Mining Detection Earth Engine-TensorFlow pipeline. The process is structured in 4 steps. First (1), GEE generates the cloud-free composites and creates the initial training dataset. Second (2), the mosaics and training data are downloaded and stored locally. Three (3), initiate patch-wise training and classification. In the fourth step (4), the classified product is spatially and temporally filtered. The filtered product is visually and statistically inspected. Multiple iterations may be used until a satisfactory spatial and temporal quality is achieved.](image)

3.1 Classification scheme

For the supervised classification of the Landsat mosaics, we selected training samples (geometries) from the previously generated bounding boxes (grids). Like any supervised algorithm, our U-net-based approach depends on human-labeled training data, categorized as mining (Mi) and non-mining (N-Mi). Guided by the reference dataset, the mining and non-mining samples are visually delineated. It is essential to highlight that no differentiation was made between artisanal or industrial mining samples. Therefore, from this point on, every time mining samples or classes are mentioned, it includes both artisanal and industrial patterns. The dissociation between such patterns, *garimpo* or industrial, and the exploited main substance results from post-classification and visual analysis.

Once the sample collection and the U-net classification are done, the classified data is injected back into GEE, where spatial-temporal filters and visual inspection occur. This phase was undertaken to correct misclassified data and evaluate the need to collect (or not) additional training samples. Table 2 shows the parameters used in our classification model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier</td>
<td>U-Net</td>
</tr>
<tr>
<td>Tile-Size</td>
<td>256 x 256 px</td>
</tr>
<tr>
<td>Samples</td>
<td>8400</td>
</tr>
<tr>
<td>Attributes</td>
<td>Swir1, Nir1, Red, MNDWI, NDVI, and NDSI</td>
</tr>
<tr>
<td>Classes</td>
<td>2 (Mining and Not-Mining)</td>
</tr>
</tbody>
</table>
4 Mining Class, Mining Type, and Main Substances

Since Collection 6, the MapBiomas Platform counts with a specific mining related module. Thus, it is important to understand each mining-related product’s origin. In this sense, the pattern recognition of a mining site, regardless of its nature or main substance, is a task performed by the U-Net classifier in a binary fashion [mining (Mi) and non-mining (N-Mi)], as previously explained.

Once the mining class is noise-filtered, the final version is integrated as a layer in the MapBiomas LULC data, corresponding to the class ID “30”. Then, the mining raster data intersects with the CPRM-GeoSGB dataset, from which the attributes of substances (Gold, Iron, Silver, Copper…) and extraction type (garimpo or industrial) are extracted. Thus, the recognition of mining sites is U-Net-related. Still, categorizing its nature/type or main substances results from a spatial operation involving third-party references, as shown in Figure 4.

![Figure 4](image)

Figure 4 - The dots are the CPRM-GeoSGB dataset. The Yellow pixels represent the mining class. The recognition of the mine and mined substances are the resultant aggregation of both datasets.

The product published on the Mining Module aggregates both attributes in a three-digit identifier (“class_id”), resulting in the information in Table 3.
Table 3 - The product published on the Mining Module aggregates both attributes in a three-digit identifier ("class_id"). In Collection 8, the identifier 130 appears for the first time as a "class_id", due to the addition of a Zinc industrial extraction site.

<table>
<thead>
<tr>
<th>class_id</th>
<th>level_1</th>
<th>level_2</th>
<th>level_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>Metálicas</td>
</tr>
<tr>
<td>102</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.01 Ferro</td>
</tr>
<tr>
<td>103</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.02 Manganês</td>
</tr>
<tr>
<td>104</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.03 Níquel</td>
</tr>
<tr>
<td>105</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.04 Amianto</td>
</tr>
<tr>
<td>106</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.05 Molibdênio</td>
</tr>
<tr>
<td>107</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.06 Titânio</td>
</tr>
<tr>
<td>108</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.07 Cromo</td>
</tr>
<tr>
<td>109</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.08 Cobre</td>
</tr>
<tr>
<td>110</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.09 Alumínio</td>
</tr>
<tr>
<td>111</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.10 Magnésio</td>
</tr>
<tr>
<td>112</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.11 Bário</td>
</tr>
<tr>
<td>113</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.12 Niobio</td>
</tr>
<tr>
<td>114</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.13 Estanho</td>
</tr>
<tr>
<td>115</td>
<td>2. Industrial</td>
<td>2.2 Metálicas</td>
<td>2.2.14 Ouro</td>
</tr>
<tr>
<td>116</td>
<td>2. Industrial</td>
<td>2.3 Não Metálicas</td>
<td>Não Metálicas</td>
</tr>
<tr>
<td>117</td>
<td>2. Industrial</td>
<td>2.3 Não Metálicas</td>
<td>2.3.01 Minerais Classe 2</td>
</tr>
<tr>
<td>118</td>
<td>2. Industrial</td>
<td>2.3 Não Metálicas</td>
<td>2.3.02 Fluor</td>
</tr>
<tr>
<td>119</td>
<td>2. Industrial</td>
<td>2.3 Não Metálicas</td>
<td>2.3.03 Fósforo</td>
</tr>
<tr>
<td>120</td>
<td>2. Industrial</td>
<td>2.3 Não Metálicas</td>
<td>2.3.04 Gráfita</td>
</tr>
<tr>
<td>121</td>
<td>2. Industrial</td>
<td>2.3 Não Metálicas</td>
<td>2.3.05 Silício</td>
</tr>
<tr>
<td>122</td>
<td>2. Industrial</td>
<td>2.3 Não Metálicas</td>
<td>2.3.06 Calcário</td>
</tr>
<tr>
<td>123</td>
<td>2. Industrial</td>
<td>2.4 Pedras Preciosas &amp; Rochas Ornamentais</td>
<td>Pedras Preciosas &amp; Rochas Ornamentais</td>
</tr>
<tr>
<td>124</td>
<td>2. Industrial</td>
<td>2.4 Pedras Preciosas &amp; Rochas Ornamentais</td>
<td>Pedras Preciosas</td>
</tr>
<tr>
<td>125</td>
<td>2. Industrial</td>
<td>2.4 Pedras Preciosas &amp; Rochas Ornamentais</td>
<td>Rochas Ornamentais</td>
</tr>
<tr>
<td>126</td>
<td>2. Industrial</td>
<td>2.1 Energéticas</td>
<td>Energéticas</td>
</tr>
<tr>
<td>127</td>
<td>2. Industrial</td>
<td>2.1 Energéticas</td>
<td>2.1.01 Carvão mineral</td>
</tr>
<tr>
<td>128</td>
<td>2. Industrial</td>
<td>2.1 Energéticas</td>
<td>2.1.02 Urânio</td>
</tr>
<tr>
<td>129</td>
<td>2. Industrial</td>
<td>2.1 Energéticas</td>
<td>2.1.03 Gás natural e petróleo</td>
</tr>
<tr>
<td>214</td>
<td>1. Garimpo</td>
<td>1.1 Metálicas</td>
<td>1.1.02 Estanho</td>
</tr>
<tr>
<td>215</td>
<td>1. Garimpo</td>
<td>1.1 Metálicas</td>
<td>1.1.01 Ouro</td>
</tr>
<tr>
<td>216</td>
<td>1. Garimpo</td>
<td>1.2 Não Metálicas</td>
<td>Não Metálicas</td>
</tr>
<tr>
<td>217</td>
<td>1. Garimpo</td>
<td>1.2 Não Metálicas</td>
<td>1.2.01 Minerais Classe 2</td>
</tr>
<tr>
<td>223</td>
<td>1. Garimpo</td>
<td>1.3 Pedras Preciosas &amp; Rochas Ornamentais</td>
<td>Pedras Preciosas &amp; Rochas Ornamentais</td>
</tr>
<tr>
<td>224</td>
<td>1. Garimpo</td>
<td>1.3 Pedras Preciosas &amp; Rochas Ornamentais</td>
<td>1.3.01 Pedras preciosas</td>
</tr>
<tr>
<td>225</td>
<td>1. Garimpo</td>
<td>1.3 Pedras Preciosas &amp; Rochas Ornamentais</td>
<td>1.3.02 Rochas Ornamentais</td>
</tr>
</tbody>
</table>
5 Post-classification

Due to the pixel-based nature of the classification method and the very long temporal series, a chain of post-classification filters were applied to reduce the salt-and-pepper effect and add spatiotemporal consistency. The post-classification process includes the application of the following filters: gap-fill, temporal, spatial, and frequency.

5.1 Gap-Fill filter

The post-processing steps start by filling in possible no-data values. In a long time-series of severely cloud-affected regions such as tropical coastal zones, pixels with no-data values are expected to be present in median composite mosaics. The gap-fill filter replaces the no-data values (i.e., image “gaps”) with a classified pixel from the nearest date available. In this procedure, if no “future” valid class is available, the no-data value is replaced by the nearest previous valid class. Up to three prior years can be used to fill in persistent no-data pixels. Therefore, gaps should only exist if a given pixel has been permanently classified as no-data throughout the entire temporal series. A mask of “years” was built to keep track of pixel temporal origins, as shown in Figure 5.

![Figure 5 – Gap-filling filter mechanism. The next valid classification replaces existing no-data values. If no “future” valid position is available, then the no-data value is replaced by its previous valid classification based on up to a maximum of three (3) prior years. To keep track of pixel temporal origins, a “years” mask was built.](image)

5.2 Temporal filter

Next, we applied a temporal filter that uses sequential classifications in a 3-year unidirectional moving window to identify temporally non-permitted transitions. Based on a single generic rule (GR), the temporal filter inspects the central position of three consecutive years (“ternary”). It changes its value if it differs from the first and last years in the ternary,
which must have identical classes. The central year of the ternary is then reclassified to match its temporal neighbor class, as shown in Table 4.

Table 4 - The temporal filter inspects the central position for three consecutive years, and in cases of identical extremities, the center position is reclassified to match its neighbor. T1, T2, and T3 stand for positions one (1), two (2), and three (3), respectively. GR means “generic rule,” while Mi and N-Mi represent mining and non-mining pixels.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Input (Year)</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>GR</td>
<td>Mi</td>
<td>N-Mi</td>
</tr>
<tr>
<td>GR</td>
<td>N-Mi</td>
<td>Mi</td>
</tr>
</tbody>
</table>

5.3 Spatial filter

Then, a spatial filter was applied to avoid unwanted modifications on the edges of grouping pixels (clusters) by using the “connectedPixelCount” function. Native to the GEE platform, this function locates connected components (neighbors) that share the same pixel value. Thus, only pixels that do not share connections to a pre-defined number of identical neighbors are considered isolated, as shown in Figure 6. In this filter, at least ten connected pixels are needed to reach the minimum connection value. Consequently, the minimum mapping unit is directly affected by the spatial filter applied, which was defined as 10 pixels (~1 ha).

![Figure 6](image)

Figure 6 – The spatial filter removes pixels that do not share neighbors of identical value. The minimum connection value was 10 pixels.

5.4 Frequency filter

The last post-processing filter step is the frequency filter. This filter considers the frequency of a given class throughout the entire time-series. Thus, all class occurrences with less than 10% temporal persistence (3 years or fewer out of 37) are filtered out and incorporated into the non-class binary. This mechanism contributes to reducing the temporal oscillation in the classification, decreasing the number of false positives, and preserving consolidated classes.
5.5 Integration with biomes and cross-cutting themes

After applying the post-processing filters, we integrate the cross-cutting themes and the Biomes data into a single raster data. This integration is guided by a set of specific hierarchical prevalence rules (Table 5). The resulting output is a final land cover/land use map for each region of the MapBiomas project.

The top position classes in the prevalence rank are related to coastal ecosystems (such as mangroves, beaches, dunes, and sand spots; aquaculture) and anthropogenic land use (i.e., mining and urban infrastructure) present throughout the country (Table 5).

<table>
<thead>
<tr>
<th>Class</th>
<th>Pixel Value</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Beach, Dune, and Sand Spot</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>Mangrove</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Aquaculture/Salt-Culture</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td>Hypersaline Tidal Flat</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>Urban Infrastructure</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>Sugar Cane</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>Soybean</td>
<td>39</td>
<td>8</td>
</tr>
<tr>
<td>Rice</td>
<td>40</td>
<td>9</td>
</tr>
<tr>
<td>Other Temporary Crop</td>
<td>41</td>
<td>10</td>
</tr>
<tr>
<td>Perennial Crop</td>
<td>36</td>
<td>11</td>
</tr>
<tr>
<td>Coffee</td>
<td>46</td>
<td>12</td>
</tr>
<tr>
<td>Citrus</td>
<td>47</td>
<td>13</td>
</tr>
<tr>
<td>Other Perennial Crop</td>
<td>48</td>
<td>14</td>
</tr>
<tr>
<td>Temporary Crop</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>Forest Plantation</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Rocky Outcrop</td>
<td>29</td>
<td>17</td>
</tr>
<tr>
<td>Other Non-Vegetated Areas</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>River, Lake, and Ocean</td>
<td>33</td>
<td>19</td>
</tr>
<tr>
<td>Forest Formation</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>Savanna Formation</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>Wetland</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>Grassland Formation</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>Pasture</td>
<td>15</td>
<td>24</td>
</tr>
<tr>
<td>Mosaic of Uses</td>
<td>21</td>
<td>25</td>
</tr>
</tbody>
</table>
References


