

1 Mapping Mining Areas in the Tropics from 2016–2024

2 Philipp Sepin¹, Lukas Vashold¹, and Nikolas Kuschnig^{1,*}

3 ¹Vienna University of Economics and Business

4 *Corresponding author: Nikolas Kuschnig (nikolas.kuschnig@wu.ac.at)

5 ABSTRACT

6 Mining provides crucial materials for the global economy and the climate transition, but can have severe environmental and social impacts. Current analyses of these impacts are limited by a lack of data on mining activity, particularly in the regions most affected. In this paper, we present a novel panel dataset mapping mining sites along the tropical belt from 2016 to 2024. Our approach uses a machine learning model, trained on over 25,000 mining polygons from the literature, to automatically segment mining areas in high-resolution (< 5 m) satellite imagery. The dataset maps over 145,000 mining polygons covering an average area of 65,000 km² annually, with an accuracy of 87.7% and precision of 84.1%. Our approach allows for accurate, precise, and consistent delineation, and can be scaled to new locations and periods. The dataset enables detailed analyses of local environmental, social, and economic impacts of mining in regions where conventional data is scarce or incomplete.

7 Transition minerals play a crucial role in climate action, and are necessary for the switch towards
8 cleaner production, storage, and distribution of energy.¹ Globally, mining operations are expanding to meet
9 the growing demand for raw materials, often encroaching upon vulnerable regions.² Projections suggest
10 that this expansion will accelerate drastically in pursuit of the Paris Agreement and subsequent climate
11 conferences.³ Yearly extraction of critical minerals is projected to increase by 150–450% depending on
12 the mineral, with a cumulative total of material extracted reaching 1.8–3.5 billion tons by 2050.⁴

13 Understanding the impacts of increasing mineral extraction is crucial, but relies on comprehensive
14 data that is often lacking. On the one hand, mining is linked to several adverse environmental and
15 social effects, including deforestation, loss of biodiversity, soil erosion, water pollution, air contamination,
16 corruption, and violent conflicts.^{5–14} On the other hand, mining can present economic opportunities, having
17 been shown to increase wealth levels, asset ownership, and incomes as well as related socioeconomic
18 indicators,^{15–18} although the effects on local and regional development are ambiguous,^{19–22} and may
19 vary by mineral type and target market.²³ Yet, if managed successfully, the demand for minerals could
20 positively affect economic development and help facilitate the delivery of the sustainable development
21 goals (SDGs).²⁴

22 The effective management of mining impacts necessitates detailed information on the location, areal
23 extent, and activity over time. Despite previous efforts,^{25–29} information on the status and especially the
24 development of mining over time remains scarce,³⁰ and comprehensive analyses are impeded.

25 In this paper, we introduce a panel dataset of mining areas within the tropical belt. The dataset spans
26 from 2016 to 2024, and consists of mining polygons that are automatically delineated using state-of-the-art
27 machine learning (ML) methods. Our methodology employs a transformer-based segmentation model,³¹
28 trained on an extensive dataset of mining polygons from the existing literature.^{26,28} We apply this model
29 to delineate polygons of known mine sites using frequent, high-resolution satellite imagery from Planet,
30 which is provided under Norway's International Climate and Forest Initiative (NICFI).

31 This approach provides a comprehensive dataset of mining polygons, which tracks yearly footprint

32 changes of previously identified mining sites. This data enables large-scale analyses of the various impacts
33 of mineral extraction — particularly in regions where such data has historically been scarce. The nature of
34 our approach allows for consistent and accurate predictions that can be easily extended across different
35 locations and temporal ranges.

36 Results and Discussion

37 Our dataset covers mining sites in the tropical belt and tracks them at annual frequency from 2016 until
38 2024. Mining sites and their development over time are captured by a total of over 147,000 mining
39 polygons that cover an average area of 66,400 km² over the observed period. [Figure 1](#) illustrates the spatial
40 coverage of the dataset, and presents the change in the area of monitored mining sites.

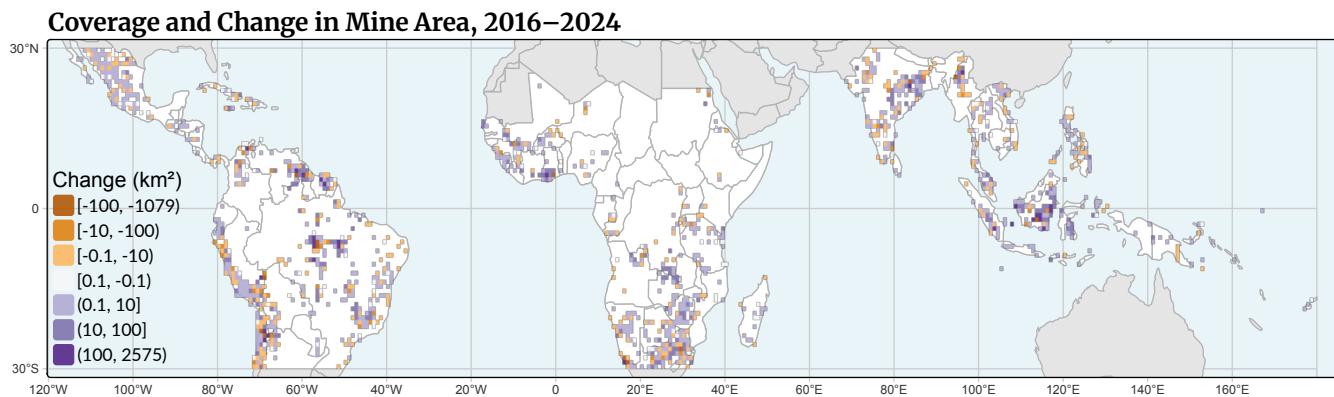


Figure 1. Change in delineated mining areas over 2016–2024. Colored regions indicate the coverage of our dataset (i.e., the tropical belt), while (non-white) pixels indicate the change in the area of tracked mining sites (km²).

41 In this section, we describe insights that can be obtained from our dataset as well as benefits from
42 our modeling approach, and discuss their usage, limitations, and areas for future work. We begin by
43 focusing on the temporal dimension of our dataset, which allows researchers to track mining sites and their
44 impacts over time. We illustrate how this information can be used to monitor the activity, expansion, and
45 — potentially — renaturalization of individual mining sites. Afterwards, we explore the development of
46 mining areas at aggregated levels. Then, we focus on our modelling approach. Segmentation models can
47 distill a consensus from heterogeneous training data, generating consistent polygons, and lower the cost of
48 delineating polygons considerably. Finally, we discuss the limitations of this study and the introduced
49 dataset, as well as directions for future research.

50 Tracking a mine site

51 [Figure 2](#) shows the development of the Toka Tindung mining site from 2016 to 2024 based on annual
52 satellite imagery as well as the corresponding annual model delineations. The mining site is located in
53 the Indonesian province of North Sulawesi, approximately 35 kilometers east of the provincial capital,
54 Manado. Commercial production began in 2011, and production and processing capacity have been added
55 since, making Toka Tindung one of the largest gold mining operations in Southeast Asia today. Over the
56 period from 2016–2024, multiple upgrades of the processing plant more than quadrupled the project's
57 throughput capacity to four million tons per year.³²

58 The increase in capacity is clearly reflected in [Figure 2](#), and captured by our predictions. The expansion
59 of the main pits of Toka Tindung has been accompanied by the addition of infrastructure, such as water

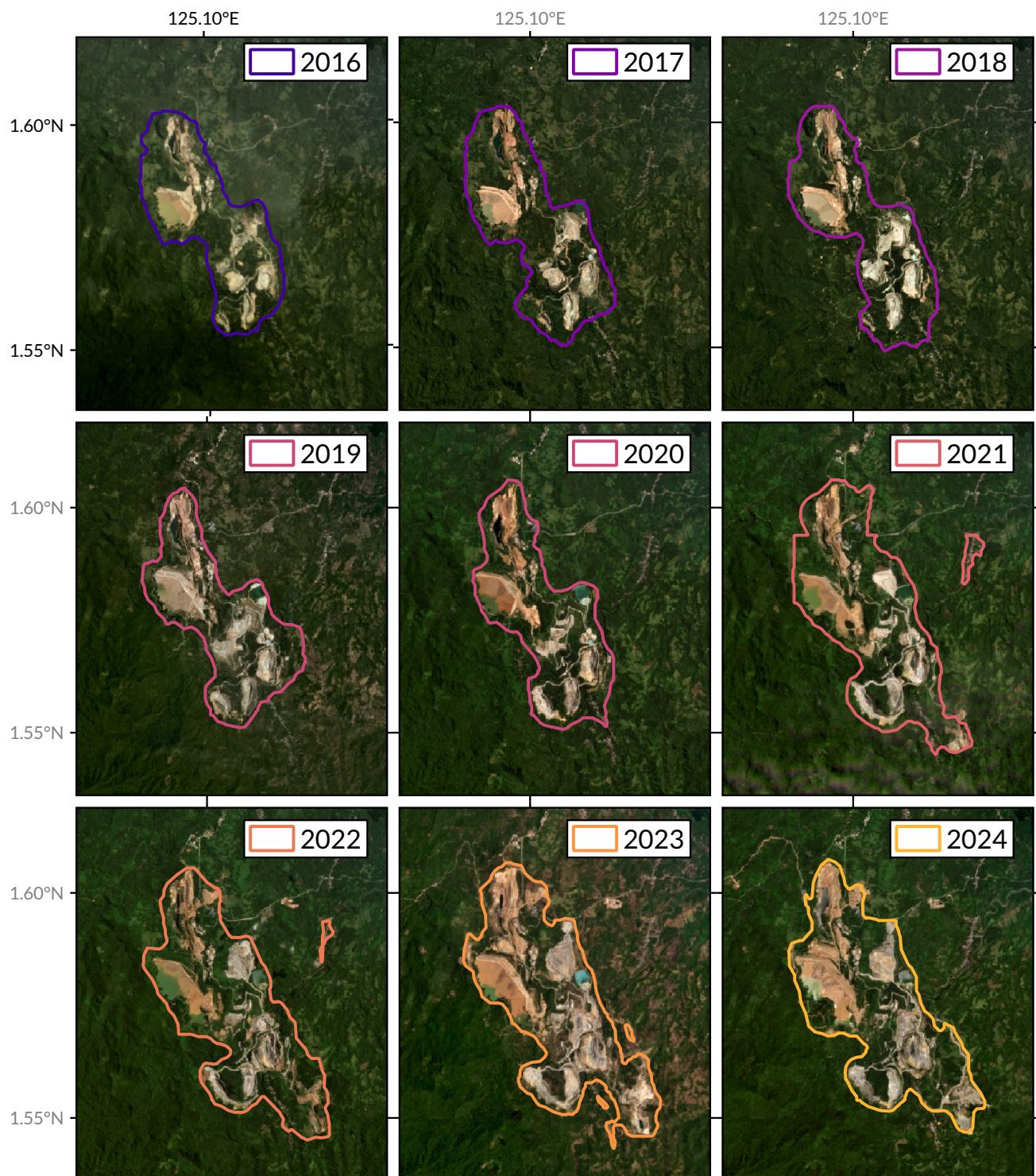


Figure 2. Annual predictions (2016–2024) of the extent of the Toka Tindung gold mining project in Indonesia ($1^{\circ}35'N$ $125^{\circ}06'E$) are plotted over the corresponding annual satellite image (Planet/NICFI).

storage facilities. Over 2016–2024, the Toka pit in the north and the Kopra, Blambangan, and Araren pits in the south have expanded and grown closer to each other and expanded considerably. Later developments were concentrated in the southern parts of the mine site, including the Araren pit in the bottom-right of the images.

In the existing mining polygon datasets that we use to train our model,^{26,28} none of these developments are reflected (see [Figure S2](#) of the Supplementary Information for a visualization). The southern pits are missing from one of the sources,²⁶ indicating the use of satellite imagery from before 2019. The other source²⁸ delineates these elements individually, resulting in five polygons. By contrast, our predictions generally result in one large polygon, stemming from the broad nature of the ground truth used for training. Overall, our mining polygons consistently trace the expansion of Toka Tindung, better reflecting its actual state in a given year.

Changes of mining sites over time

Mining is prevalent throughout the tropics, but is well known to be highly clustered in a few areas.^{26,28} Similarly, it is well known that mineral extraction is booming,⁴ but less is known about the changing footprint of mining areas. Below, we first show how our dataset reveals substantial changes in the location and extent of mining clusters in the period from 2016–2024. Then, we use the dataset to assess the development of mining sites over time.

Changing footprint of mining hotspots

In Latin America, Brazil and Chile feature the largest, and rapidly growing, areal footprints of mining sites (see [Table 1](#)). As visible in [Figure 1](#), the Amazon rainforest has been a particular hotspot for increasing footprints; especially along the Brazilian ‘arc of deforestation’ at the southern and eastern edges of the Amazon, and its northern edges, including the border region between Guyana and Venezuela, and Suriname. Traditional mining regions along the Chilean Andes show alternating patches of increases and decreases in the size of mining sites. Other countries in the Americas, such as Colombia and Mexico, show considerable, but more dispersed, growth in the footprint of mining sites.

| Country | Mining Area (km ²) | | | | | | | | |
|--------------|--------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 13,266 | 12,425 | 12,376 | 13,761 | 15,281 | 16,859 | 15,746 | 17,497 | 17,801 |
| Indonesia | 13,266 | 12,425 | 12,376 | 13,761 | 15,281 | 16,859 | 15,746 | 17,497 | 17,801 |
| Brazil | 7,829 | 8,785 | 8,002 | 8,023 | 8,548 | 8,215 | 9,764 | 9,342 | 9,360 |
| Chile | 6,293 | 5,748 | 5,746 | 7,161 | 7,269 | 6,434 | 7,561 | 7,176 | 7,218 |
| South Africa | 7,469 | 6,624 | 5,901 | 6,706 | 6,324 | 6,044 | 6,434 | 6,459 | 6,558 |
| India | 3,502 | 3,800 | 3,590 | 4,140 | 4,136 | 3,959 | 4,322 | 4,367 | 4,114 |

Table 1. Mining polygon area in the five countries with the largest mining footprint in the dataset over time. A complete version with all countries with a footprint of over 100 km² in the dataset is available as [Table S1](#) in the Supplementary Information.

In the western parts of Africa, pronounced growth of areal footprints can be observed in traditional gold mining areas. Ghana is a primary hotspot, but mining sites in eastern Guinea, Sierra Leone, and Burkina Faso have also expanded consistently throughout the observation period. In the south, mining sites have expanded overall, but there has been a slight shift from sites in South Africa, where the overall mining polygon area has been stagnant, to Namibia, Zimbabwe, and Botswana. The Congo river basin, where mining operations have expanded rapidly,³³ is only partially captured by our dataset. Mining sites

in the south are covered and growing strongly, while the often smaller and informal mining sites in the rest of the country are poorly covered by our sources and, hence, dataset.

In Asia, the islands of Borneo and Sulawesi in Indonesia, eastern India, and northern Myanmar are particular hotspots for growing mining sites. Other areas such as northern Thailand, Laos, and the Philippines also show expansion of mining areas, albeit at a more moderate pace. In contrast, parts of the Malay peninsula, southern Myanmar, and southern India show a declining aerial extent of mining operations over the observation period.

Trends in the area of mining sites

Mineral extraction is expanding globally,⁴ and Figure 3 shows how this coincides with increasing total area and number of mining polygons in our covered locations. Focusing on the period from 2017 to 2023,¹ the total area of mining polygons has increased by roughly 24% in the tropics, extending over about 74,100 km² in 2023. In this period, the number of unique mining polygons in covered locations also increased by 23% to slightly more than 19,000 in the year 2023, while their average size remained stagnant.

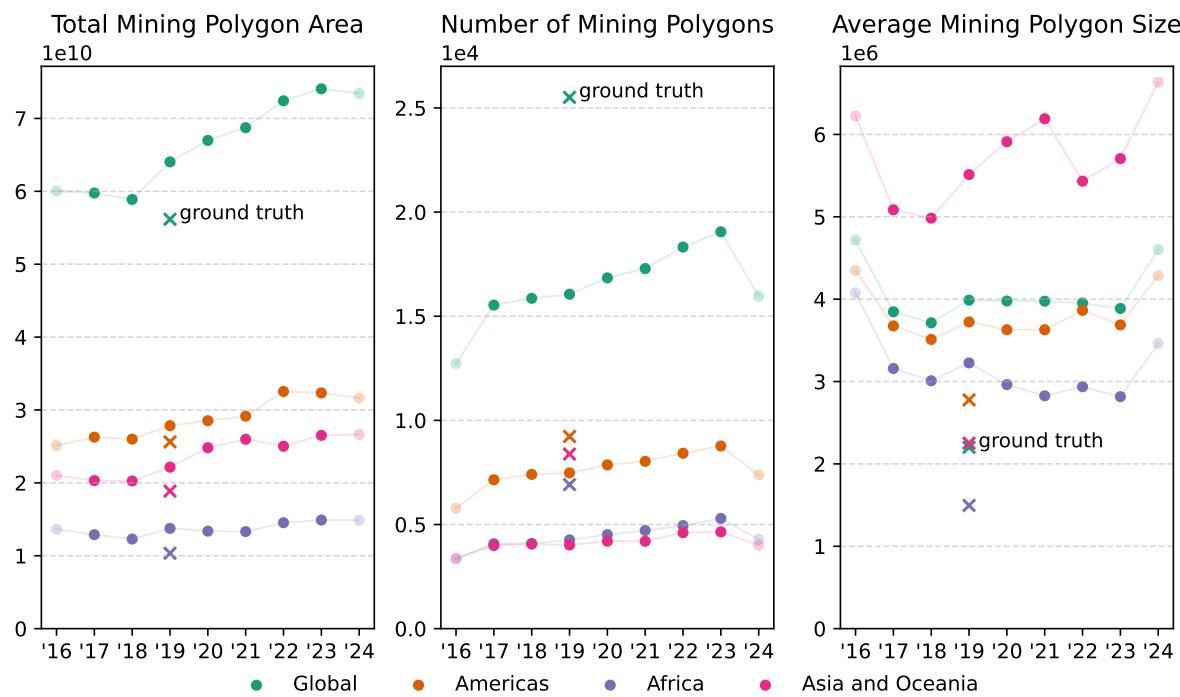


Figure 3. Summary of the predictions for the full dataset, and separated into regions — the Americas, Africa, and Asia and Oceania. The left panel shows the increase in area (in 10,000 km²), the center panel the number of individual mining polygons (in thousands), and the right panel shows the average size of these mining polygons (in km²). The cross (×) indicates values for the ground truth.

These dynamics are similar on the regional level.² The largest absolute increase in mining polygon area in covered sites occurred in Asia and Oceania, with a total increase of 6,200 km², or more than 31%, from 2017 to 2023. In terms of total area, the Americas host the most expansive mining sites in 2023,

¹The post-processing of predictions results in more conservative predictions for the initial and final years (2016 and 2024).

²Figure S1 in the Supplementary Information further decomposes the regional graphs in Figure 3 to show the three countries within them that have the largest share of total mining area.

107 at over 32,600 km², increasing by 23% since 2017. In Africa, the increase in total mining polygon area
108 amounted to 16%, expanding from 12,900 km² in 2017 to 14,900 km² in 2023.

109 In Asia and Oceania, Indonesia stands out as the country with the largest mining polygon area. From
110 2017–2023, we observe considerable growth of Indonesian mining sites at 40%, making it also the country
111 with the largest absolute increase in mining polygon area globally. Two events that relate to this trend are a
112 Chinese ban of coal imports from Australia in December 2020, and a 2020 Mining Law Amendment that
113 relaxed environmental requirements.³⁴ Notably, the annual mining polygon area in Indonesia is relatively
114 noisy, as the country houses some of the largest single mining sites. Other notable countries in the region
115 are India, which showed stable growth, and Myanmar. There, the area of mining polygons increased until
116 2020, and began to drop from 2021 onward, coinciding with the ongoing ‘Spring Revolution’.

117 In the Americas, Brazil is the country with the most expansive mining polygons. Significant growth
118 occurred from 2019–2022, during the administration of Jair Bolsonaro. Most of this expansion took place
119 in and along the fringes of the Amazon rainforest (cf., [Figure 1](#)), corroborating early concerns about
120 potential repercussions of legislative changes and increases in anti-environmental rhetoric.³⁵ Chile, which
121 is only partially covered by our dataset, comes a close second. Mining polygons in the country are, on
122 average, larger than in other countries in the region. From 2017 to 2023, the area of mining polygons in
123 the country increased by about 25%.

124 In Africa, the country with the highest share of mining polygons and area is, despite it being only
125 covered partially, South Africa. The country has been experiencing stagnating mineral production since
126 the early 2010s,³⁶ which is reflected by flat figures for the change of area and the number of mining
127 polygons in our dataset. Ghana drove most of the regional growth, with steady expansion that correlates
128 with rising international gold prices. Gold accounts for around 95% of the country’s mineral revenues, and
129 much of it is produced in small-scale artisanal mines.²⁵ For Africa, [Figure 3](#) shows a substantial increase
130 in the number of mining polygons, and moderate area increases which may be indicative of artisanal
131 mining. However, artisanal mining sites, which are widespread across Africa,³⁷ are underrepresented in
132 our dataset.

133 **Usage, limitations, and outlook**

134 The introduced dataset contributes a time dimension to mining polygons at known locations. This enables
135 various applications and extensions, while maintaining certain limitations that we discuss next.

136 **Applications** The temporal dimension provides accurate polygons over time and enables multiple
137 analytical applications. First, it allows for more nuanced analyses of the direct and indirect impacts
138 of mines on their surrounding environments, such as forest loss or vegetation impacts.^{9,14} Second, it
139 facilitates monitoring of mining sites, which can inform re-naturalization projects for abandoned mines.
140 Third, changes in the area of mining sites can serve as a proxy for on-site activity. This enables more
141 accurate assessments of mining-related impacts where data is incomplete or scarce,³ and facilitates
142 analyses in settings where no conventional data is available, e.g., due to conflict.

143 **Advantages** Our automated approach offers several distinct advantages. The segmentation process is
144 internally consistent and can be adjusted to meet specific requirements, such as balancing the cost of false
145 positives versus false negatives. The method also enables efficient scalability in both spatial and temporal
146 dimensions. As new satellite images become available, we can easily expand the temporal scope of the
147 dataset. Additionally, we can incorporate more frequent satellite imagery to assess intra-term changes

³For example, in analyses of societal impacts like conflict and corruption,⁶ environmental effects such as air and water pollution,¹² and their resulting socioeconomic consequences.

148 in mining footprints. This could, e.g., help reveal transitory substitution of agricultural activities due to
149 meteorological shocks in subsistence economies. The spatial scope can be extended by incorporating new
150 mining locations — in regions beyond the tropics, or ones that are only sparsely covered by our sources
151 (e.g., in the Congo basin).

152 **Current limitations**

153 The dataset and approach presented here have noteworthy limitations.

154 First, polygons of individual mining sites should be used with caution. While the model performs
155 well on average and post-processing reduces the noise in our predictions, the results are not free of errors.
156 This is particularly problematic (a) where mining sites are imprecisely delineated in the ground truth (e.g.,
157 artisanal mining sites in Western Africa or the Amazon), (b) in areas where the environment provides
158 limited contrast (e.g., along the western coast of South America). While our predictions compare favorably
159 with the manually delineated datasets used to train the model (see the Methods section), we recommend
160 using any of these datasets at more aggregated levels.

161 Second, our approach relies on numerous high-resolution satellite images covering vast areas at high
162 frequency. This paper was possible due to the Planet/NICFI program, which provided free satellite
163 imagery of the tropical belt for research purposes. After the program's expiration in January 2025, the
164 reproducibility and extensibility of the presented dataset is impeded. Future work and extensions will
165 require access to other sources of satellite imagery. These are currently prohibitively expensive (in the
166 case of high-resolution sensors), or require sacrifices in terms of the resolution and frequency of available
167 imagery (in the case of Sentinel-2).

168 Third, our dataset focuses on tracking known mining sites over time, rather than identifying newly
169 established sites far from existing ones. Our source data on mining locations and, consequently, our
170 predictions, primarily cover industrial mines and smaller mines in their vicinity, potentially under-
171 representing regions where illegal or informal mining predominates. As a result, our dataset cannot
172 provide a complete overview of the development of tropical mining areas. Enhancing the coverage and
173 delineation of artisanal mining sites, as well as obtaining more comprehensive coverage statistics and
174 validation results, will require additional data.

175 **Methods**

176 In this section, we describe the data and methods used to produce the dataset, and our validation strategy.

177 **Data**

178 We begin by describing the satellite imagery used to delineate polygons, as well as the mining polygons
179 that constitute our ground truth.

180 **Satellite imagery**

181 We use high-resolution (< 5 m per pixel) satellite imagery from Planet's PlanetScope constellation of
182 imaging satellites, made available for non-commercial use by NICFI. The imagery covers approximately
183 45 million square kilometers along the tropical belt, ± 30 degrees of latitude. This includes several
184 important mineral extracting countries such as Indonesia, the Democratic Republic of the Congo (DRC),
185 and Brazil (see [Figure 1](#)), but excludes, e.g., Australia, the US, Russia, or China, and only partially covers
186 Chile, Argentina, or South Africa.

187 The dataset spans from December 2015 through December 2024. Imagery is available as biannual
188 composites (December–May, June–November) until May 2020, as a quarterly composite for June–August
189 2020, and as monthly composites thereafter. Images are provided as so-called 'quads', which are compiled

190 from individual daily ‘scenes’ that are captured daily by the constellation of approximately 130 satellites.
191 An illustration of two monthly quads is provided in [Figure S3](#) of the Appendix. This frequent capture rate
192 of imagery results in composite images with relatively high clarity.

193 For our annual predictions, we use imagery from the second half of the year (June–November). For
194 2016–2019, we use the provided quads composites as is, while we select the composites with optimal
195 quality (in terms of minimum cloud coverage) for 2020–2024. While quality metadata only exists for
196 individual scenes and thus represents worst-case estimates, we observe over 96% clarity on average across
197 our dataset (see [Table S2](#) in the Supplementary Information for details).

198 **Mining polygons**

199 Our study relies on mining polygons from two comprehensive datasets to (a) establish a ground truth to
200 train our model and (b) provide mining locations for prediction.

201 The first dataset, initially presented in Maus et al. (2020)³⁸ and updated in Maus et al. (2022),²⁶ covers
202 both industrial and artisanal mines. The dataset contains 44,929 mining polygons covering approximately
203 101,583 km² globally. The original dataset combined undated imagery from Google Satellite, Microsoft
204 Bing Imagery, and Sentinel-2, the updated data is based on Sentinel-2 imagery (10 m per pixel) from 2019.
205 Starting from the SNL Metals and Mining database locations, the authors examined 10 km buffers around
206 known mines to account for expansive operations and nearby (artisanal) mining sites.

207 The second dataset is an extended version of Tang et al. (2021)³⁹, updated by Tang and Werner
208 (2023),²⁸ and contains 74,548 mining polygons, covering roughly 66,000 km² globally. The authors
209 combined various sources for mine locations, including the SNL database, as well as satellite imagery.
210 Imagery sources include Sentinel-2, Landsat, and high-resolution imagery available in Google Earth Pro
211 (< 1 m per pixel), although no single source was used consistently, and no timestamps are available.

212 These two datasets have considerable overlap in terms of mine locations, but differ in their segmentation
213 approaches. The second dataset²⁸ precisely contours individual mine features, such as waste rock dumps,
214 pits, ponds, tailings dams, and infrastructure. By contrast, Maus et al.²⁶ attempt to cover mining sites in
215 general, often including areas that cannot be strictly classified as mining areas themselves.⁴

216 **Ground truth**

217 Our ground truth data combines these two state-of-the-art mining polygon datasets, with each presenting
218 distinct challenges and advantages. The primary limitation in both datasets is the lack of exact timestamps.
219 We use 2019 as the reference year, but polygons may have been delineated using earlier satellite im-
220 agery. Additional imprecision may stem from heterogeneous satellite imagery, differences in delineation
221 approaches, and human factors, such as individual interpretation styles and fatigue effects.

222 To maximize the available information, we take the union of both datasets as our ground truth.⁵ The
223 union preserves polygons from both sources, and maintains outer boundaries where they intersect. The
224 resulting ground truth inherits the broader notion of mining polygons from the first dataset, while retaining
225 precise boundary information from the second one. While this approach sacrifices some granularity in
226 distinguishing individual mining features, it creates a more robust foundation for automatic delineation
227 and enables scalable analyses by incorporating the strengths of both source datasets.

⁴The Supplementary Information provides examples in [Figure S2](#), showing polygons from both datasets overlaid on Planet/NICFI imagery from 2019 at four locations.

⁵Experiments using either dataset individually yielded inferior results. Tang and Werner’s²⁸ precise delineation proved difficult to reproduce consistently, while using only Maus et al.’s²⁶ dataset produced slightly worse segmentation results than our combined approach.

228 **Segmentation model**

229 For the delineation of mining areas over time, we use a segmentation model that is based on state-of-the-art
230 transformer architectures from the ML literature.^{40,41} The process of obtaining mining polygons consists
231 of four steps: (1) preprocessing satellite imagery to generate datasets for training and prediction, (2)
232 training the model using our ground truth, (3) predicting mining site delineations for 2016–2024 and
233 mapping them back to their geospatial locations, and (4) postprocessing the predictions to reduce noise
234 and ensure temporal continuity.

235 **Preprocessing** Training the segmentation model requires two sources of data: satellite images and
236 corresponding segmentation masks that indicate mining areas. The model is trained on satellite imagery
237 from 2019, and segmentation masks are only generated for this reference year. The images, however, are
238 also used for prediction, and need to be generated for each year and polygon. We generate a square image
239 with twice the maximum side length of the respective polygon to accommodate potential growth over time,
240 and downscale them to 512×512 pixels if necessary. The resulting data is split into train/validation/test
241 samples using an 89/10/1 ratio, resulting in approximately 18,300 training samples, 2,100 validation
242 samples, and 200 samples for testing. The test samples were manually selected (from random samples) to
243 ensure they contain only precise high-quality delineations, avoiding noise in the ground truth.

244 **Training** For the segmentation task, we utilize the SegFormer model.³¹ SegFormer has demonstrated
245 strong performance on benchmark data sets,^{42,43} and outperforms older architectures like U-Net,^{44,45}
246 which has previously been used for delineating mining areas in Sub-Saharan Africa.²⁷ The model employs
247 a transformer architecture,⁴⁰ which provides a broad receptive field and enables parallel processing.⁴⁶
248 This allows transformer-based models to efficiently analyze images with global context, leading them to
249 regularly outperform convolutional neural networks like U-Net at segmentation.⁴⁵

250 We specifically employ SegFormer-B5,³¹ the largest model variant. The encoder undergoes pre-
251 training on the Imagenet-1k dataset,⁴⁷ before fine-tuning the whole model on our mining dataset for about
252 20 epochs.⁶ Due to the class imbalance, we use the *mean Intersection over Union* (mIoU) to assess the
253 model’s predictive performance. The model achieves an mIoU of 65.49% and a mean accuracy of 95.67%
254 on the test set when applying an optimized probability threshold $p \geq 0.625$ at the model’s probability
255 output. For a selected threshold of $p \geq 0.55$, chosen to match the ground truth in 2019, the model receives
256 an mIoU of 63.85% and a mean accuracy of 93.21%. For a rough comparison, a U-Net model trained
257 with similar parameters achieved an mIoU of 56.69% and a mean accuracy of 93.35%.

258 **Prediction** The prediction process, illustrated in Figure 4, generates mining polygons for all locations
259 for 2016–2024. The model outputs a probability for each pixel that indicates its confidence in the presence
260 of mining. To classify a certain pixel as part of a mine, we use a probability threshold that can be adjusted
261 to tweak, e.g., the false positive and false negative rates of the predictions. Lower thresholds result in more
262 and larger predicted mining polygons, and thus higher recall, while higher thresholds lead to fewer, but
263 more precise predictions.

264 To illustrate the flexibility that this parameter affords, we follow two strategies in setting it. First,
265 we consider multiple thresholds and manually optimize to $p \geq 0.55$ to emulate the ground truth in the
266 ‘Reference validation’ section below. This facilitates comparison, and highlights the model’s ability to
267 interpolate between high-recall and high-precision approaches. As an alternative, we set the parameter by
268 optimizing the mIoU over the validation set, reaching an mIoU of 65.49%. The corresponding threshold

⁶We train on two NVIDIA A30 GPUs, employing class-balanced loss to address the imbalance between mining and non-mining areas. Additionally, we apply data augmentation techniques, including random horizontal and vertical flipping, random cropping, random resizing, and online hard example mining,⁴⁸ to focus on challenging cases.

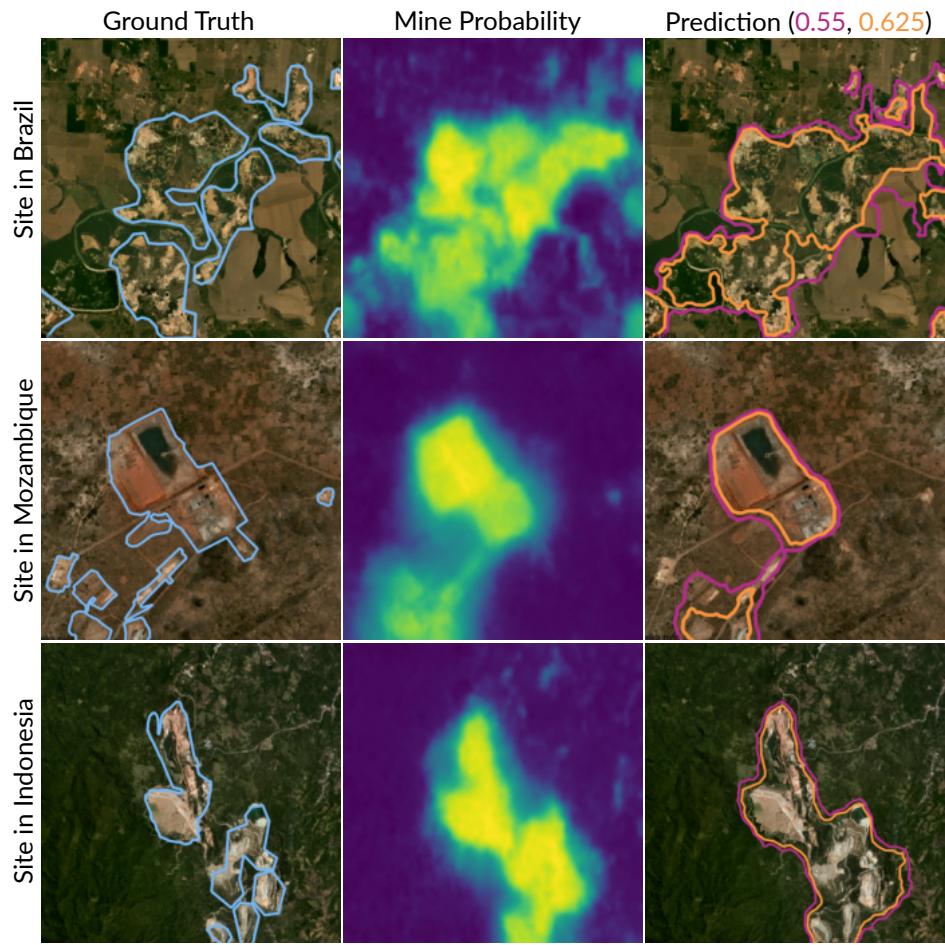


Figure 4. Prediction process for three mining sites in Brazil (near Peixoto de Azevedo), Mozambique (Balama graphite mine), and Indonesia (Toka Tindung, see [Figure 2](#)) in 2019. The left column shows ground truth polygons, the center column shows the model's probabilistic prediction, and the right column shows the predicted polygons for thresholds of $p \geq 0.55$ and $p \geq 0.625$.

269 of $p \geq 0.625$ means that only high-confidence pixels are classified as mines. The resulting dataset exceeds
270 the precision and specificity of both the ground truth and its individual components in segmenting mines,
271 but sacrifices recall.

272 To ease comparison with our source datasets, we use $p \geq 0.55$ as the default threshold, where not
273 mentioned otherwise. After creating the binary pixel-level predictions, we apply a contour finding
274 algorithm⁴⁹ to transform these segmentation masks to polygons, and map the polygons back to their
275 geospatial coordinates using the previously calculated bounding boxes.

276 **Postprocessing** We implement postprocessing to ensure temporal continuity and to reduce noise in
277 the predictions. We remove any polygons that do not have an intersecting polygon in the previous *or*
278 subsequent year. This allows us to borrow information across years and reflects the assumption that mines
279 (and their corresponding polygons) do not emerge and disappear within a single year. Hence, the years
280 2016 and 2024, which only have a single ‘neighboring’ year to compare to, feature a lower *number* of
281 polygons while the impact on mining areas is limited.⁷ Lastly, we merge multiple, intersecting predictions
282 that may stem from locations with overlapping bounding boxes, and apply hole-filling to remove fragments
283 inside the predictions.

284 Reference validation

285 To assess our model in relation to existing datasets in the literature and fine-tune probability thresholds for
286 comparability, we use a set of labeled validation points. We rely on the set of validation points from Maus
287 et al.,²⁶ which were generated within a 10 km buffer around the mines.⁸ Of their 1,220 original validation
288 points, 450 are covered by Planet/NICFI. We consider a random subset of 200 points and manually label
289 them as either mine or non-mine for each year using Planet/NICFI imagery. Based on these points, we
290 can systematically compare (1) our model’s predictions, (2) the ground truth, and (3) its two constituting
291 datasets.

292 **Sampling strategy** Before assessing the results, we note two drawbacks related to the sampling strategy
293 used to generate the validation points. The locations are not sampled randomly, but stratified across
294 the mine and non-mine categories, as segmented by Maus et al.²⁶ As a result, (likely) non-mine labels
295 are spread out across their (large) area of interest, while mine labels from within their polygons are
296 overrepresented. Inherent to this strategy, the labels offer almost no meaningful variation over time. Points
297 from the vicinity of mining polygons, which may develop into mining areas over time, are undersampled,
298 and 181 of the 200 points do not change their label over time. We address this by reporting averages of the
299 yearly validation results. Nevertheless, the validation exercise must be seen in the context of the Maus et
300 al.²⁶ dataset, from which it was sampled.

301 **Validation results** Our predictions achieve an average accuracy of 87.7% and an average Matthews
302 Correlation Coefficient (MCC) of 71.7% for a probability threshold of $p \geq 0.55$ on the validation points,
303 indicating performance that is comparable to the source data. For the $p \geq 0.625$ threshold, the average
304 accuracy is 83.3%, and the average MCC is 61.5%. By comparison, the ground truth achieves an average
305 accuracy of 88.1%, and MCC of 77.5%; the polygons of Maus et al.²⁶ achieve values of 88.6% and 78.3%,

⁷To investigate potential bias from this postprocessing step, we experimented with applying a 100 m buffer to the predicted polygons before the intersection, assuming that new mines may emerge or disappear near existing ones. We found that this buffer had little impact on the result. This suggests that the postprocessing step primarily removes isolated polygons, which may represent false positives and can be considered as noise.

⁸In their validation strategy, Tang and Werner²⁸ use 200 validation polygons that (i) are not readily available and (ii) have limited coverage in our area of interest.

306 respectively, although these values benefit from the stratified sampling of validation points. Meanwhile,
 307 the polygons of Tang and Werner²⁸ achieve an average accuracy of 86.8%, and MCC of 69.7%.

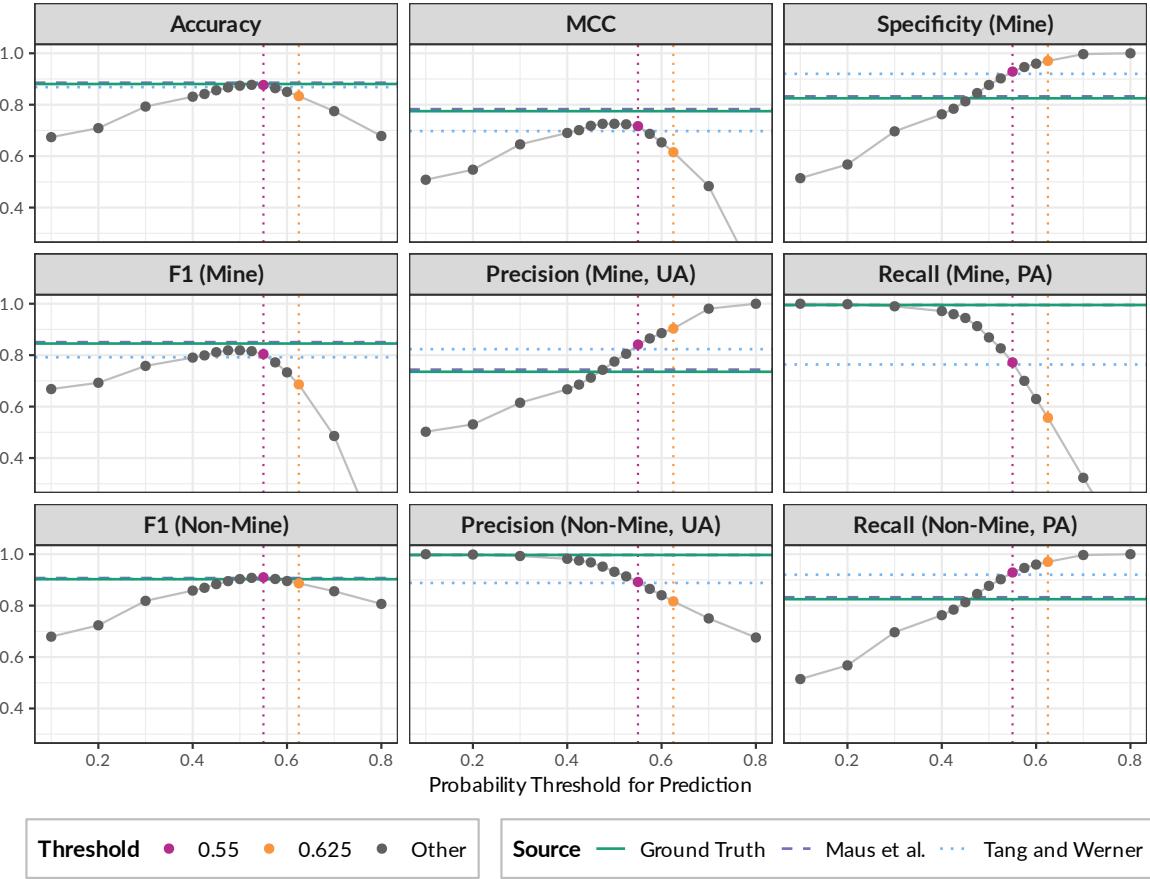


Figure 5. Reference validation results for our predictions (dots), the ground truth (solid line), and its components (dashed and dotted lines). Our predictions are evaluated at different probability thresholds for segmentation (along the horizontal axis). Two selected thresholds are highlighted: $p \geq 0.55$, the default value due to its comparability to the ground truth, and $p \geq 0.625$, which achieves the optimal mIoU. Due to the lack of variability over time, the values are averaged over the years; a variant of the plot with year-specific results is provided in Figure S4 in the Supplementary Information.

308 Further validation results, including ones for a range of probability thresholds, are visualized in
 309 [Figure 5](#). The scores of our selected threshold of $p \geq 0.55$, lie between the two sources for our ground
 310 truth, demonstrating how our method can combine their liberal and precise delineation styles. This
 311 indicates the model’s ability to effectively integrate information from both sources, identify a shared
 312 definition of mining polygons, and accurately extrapolate it to nine different timestamps.

313 What is notable is that our model can go beyond the performance of either dataset used to train it.
 314 Due to the imbalance between non-mine and mine areas, we particularly value precision (user’s accuracy,
 315 UA) in segmenting mines over their recall (producer’s accuracy, PA), although the opposite would also be
 316 possible. The selected threshold of $p \geq 0.55$, already achieves an average precision of 84.1%, exceeding
 317 the ground truth (73.5%), and its components (74.4% and 82.3%), while the higher threshold of $p \geq 0.625$
 318 even reaches 90.3%. As a result, predictions feature considerably lower false positive rates at 7.1% and
 319 3.0%, than the ground truth (17.5%), Maus et al.²⁶ (16.8%), or Tang and Werner²⁸ (8.0%). This highlights

320 the adaptability of our modeling approach, which can be fine-tuned to the task at hand.

321 **Usage notes**

322 The dataset of mining polygons is available in multiple formats to ensure accessibility. An interactive
323 map that allows the visualization of yearly predictions against (Sentinel-2) satellite imagery is available at
324 [layers.at](#). The full dataset, which includes the hand-validated test set, and validation points, can be obtained
325 from [owncloud.wu.ac.at](#), and will be made available openly on Zenodo. The dataset is licensed under the
326 Open Database License (ODbL, available at [opendatacommons.org](#)). Any rights in individual contents of
327 the database are licensed under the Database Contents License (DbCL, available at [opendatacommons.org](#)).
328 The dataset and validation points are provided in the GeoPackage (GPKG) format, and contain information
329 on the polygons themselves, the countries where they are located in, the years in which they are segmented,
330 and the areas of polygons.

331 **Code availability**

332 All code used to produce the dataset and results of this paper is available under the GNU General
333 Public License v3.0 (GPLv3, available at [gnu.org](#)) at [github.com](#). The repository includes comprehensive
334 documentation and usage instructions. Further, the training and validation datasets are available at
335 [kaggle.com](#). Scripts were written in Python and R, with geospatial processing heavily utilizing the
336 ‘GeoPandas’,⁵⁰ ‘Shapely’,⁵¹ and ‘sf’⁵² packages. The model was implemented using ‘PyTorch’⁵³ and
337 ‘MMSegmentation’.⁵⁴

References

1. Shukla, P. R. *et al.* *Climate Change 2022: Mitigation of Climate Change* (Cambridge University Press, Cambridge, 2022).
2. Luckeneder, S., Giljum, S., Schaffartzik, A., Maus, V. & Tost, M. Surge in global metal mining threatens vulnerable ecosystems. *Glob. Environ. Chang.* **69**, 102303, [10.1016/j.gloenvcha.2021.102303](https://doi.org/10.1016/j.gloenvcha.2021.102303) (2021).
3. International Energy Agency (IEA). Energy technology perspectives 2023. Tech. Rep., IEA (2023).
4. Hund, K., La Porta, D., Fabregas, T. P., Laing, T. & Drexhage, J. Minerals for climate action: The mineral intensity of the clean energy transition. Tech. Rep., World Bank Group (2023).
5. Kossoff, D. *et al.* Mine tailings dams: Characteristics, failure, environmental impacts, and remediation. *Appl. Geochem.* **51**, 229–245, [10.1016/j.apgeochem.2014.09.010](https://doi.org/10.1016/j.apgeochem.2014.09.010) (2014).
6. Berman, N., Couttenier, M., Rohner, D. & Thoenig, M. This mine is mine! How minerals fuel conflicts in Africa. *Am. Econ. Rev.* **107**, 1564–1610, [10.1257/aer.20150774](https://doi.org/10.1257/aer.20150774) (2017).
7. Sonter, L. J., Dade, M. C., Watson, J. E. M. & Valenta, R. K. Renewable energy production will exacerbate mining threats to biodiversity. *Nat. Commun.* **11**, 1–6, [10.1038/s41467-020-17928-5](https://doi.org/10.1038/s41467-020-17928-5) (2020).
8. Liu, Y. *et al.* A review of water pollution arising from agriculture and mining activities in central asia: Facts, causes and effects. *Environ. Pollut.* **291**, 118209, [10.1016/j.envpol.2021.118209](https://doi.org/10.1016/j.envpol.2021.118209) (2021).
9. Giljum, S. *et al.* A pantropical assessment of deforestation caused by industrial mining. *Proc. Natl. Acad. Sci.* **119**, e2118273119, [10.1073/pnas.2118273119](https://doi.org/10.1073/pnas.2118273119) (2022).

10. Villén-Pérez, S., Anaya-Valenzuela, L., Conrado da Cruz, D. & Fearnside, P. M. Mining threatens isolated indigenous peoples in the Brazilian Amazon. *Glob. Environ. Chang.* **72**, 102398, [10.1016/j.gloenvcha.2021.102398](https://doi.org/10.1016/j.gloenvcha.2021.102398) (2022).
11. Berman, N., Couttenier, M. & Girard, V. Mineral resources and the salience of ethnic identities. *Econ. J.* **133**, 1705–1737, [10.1093/ej/uead018](https://doi.org/10.1093/ej/uead018) (2023).
12. Macklin, M. G. *et al.* Impacts of metal mining on river systems: A global assessment. *Science* **381**, 1345–1350, [10.1126/science.adg6704](https://doi.org/10.1126/science.adg6704) (2023).
13. Mwelwa, S., Chungu, D., Tailoka, F., Beesigamukama, D. & Tanga, C. Biotransfer of heavy metals along the soil-plant-edible insect-human food chain in Africa. *Sci. The Total. Environ.* **881**, 163150, [10.1016/j.scitotenv.2023.163150](https://doi.org/10.1016/j.scitotenv.2023.163150) (2023).
14. Vashold, L., Pirich, G., Heinze, M. & Kuschnig, N. Mines–rivers–yields: Downstream mining impacts on agriculture in Africa. *WU Work. Pap.* [10.57938/7598a00e-3d24-4691-823e-aeb76032ae44](https://doi.org/10.57938/7598a00e-3d24-4691-823e-aeb76032ae44) (2024).
15. Aragón, F. M. & Rud, J. P. Polluting industries and agricultural productivity: Evidence from mining in Ghana. *The Econ. J.* **126**, 19802011, [10.1111/eco.12244](https://doi.org/10.1111/eco.12244) (2015).
16. von der Goltz, J. & Barnwal, P. Mines: The local wealth and health effects of mineral mining in developing countries. *J. Dev. Econ.* **139**, 1–16, [10.1016/j.jdeveco.2018.05.005](https://doi.org/10.1016/j.jdeveco.2018.05.005) (2019).
17. Bazillier, R. & Girard, V. The gold digger and the machine. evidence on the distributive effect of the artisanal and industrial gold rushes in Burkina Faso. *J. Dev. Econ.* **143**, 102411, [10.1016/j.jdeveco.2019.102411](https://doi.org/10.1016/j.jdeveco.2019.102411) (2020).
18. Ofosu, G., Dittmann, A., Sarpong, D. & Botchie, D. Socio-economic and environmental implications of artisanal and small-scale mining (asm) on agriculture and livelihoods. *Environ. Sci. & Policy* **106**, 210–220, [10.1016/j.envsci.2020.02.005](https://doi.org/10.1016/j.envsci.2020.02.005) (2020).
19. Bebbington, A., Hinojosa, L., Bebbington, D. H., Burneo, M. L. & Warnaars, X. Contention and ambiguity: Mining and the possibilities of development. *Dev. Chang.* **39**, 887–914, [10.1111/j.1467-7660.2008.00517.x](https://doi.org/10.1111/j.1467-7660.2008.00517.x) (2008).
20. van der Ploeg, F. Natural resources: Curse or blessing? *J. Econ. Lit.* **49**, 366–420, [10.1257/jel.49.2.366](https://doi.org/10.1257/jel.49.2.366) (2011).
21. Venables, A. J. Using Natural Resources for Development: Why Has It Proven So Difficult? *J. Econ. Perspectives* **30**, 161–84, [10.1257/jep.30.1.161](https://doi.org/10.1257/jep.30.1.161) (2016).
22. Badeeb, R. A., Lean, H. H. & Clark, J. The evolution of the natural resource curse thesis: A critical literature survey. *Resour. Policy* **51**, 123–134, [10.1016/j.resourpol.2016.10.015](https://doi.org/10.1016/j.resourpol.2016.10.015) (2017).
23. Franks, D. M., Keenan, J. & Hailu, D. Mineral security essential to achieving the Sustainable Development Goals. *Nat. Sustain.* **6**, 21–27, [10.1038/s41893-022-00967-9](https://doi.org/10.1038/s41893-022-00967-9) (2023).
24. Ali, S. H. *et al.* Mineral supply for sustainable development requires resource governance. *Nature* **543**, 367–372, [10.1038/nature21359](https://doi.org/10.1038/nature21359) (2017).
25. Barenblitt, A. *et al.* The large footprint of small-scale artisanal gold mining in Ghana. *Sci. The Total. Environ.* **781**, 146644, [10.1016/j.scitotenv.2021.146644](https://doi.org/10.1016/j.scitotenv.2021.146644) (2021).
26. Maus, V. *et al.* An update on global mining land use. *Sci. data* **9**, 1–11, [10.1038/s41597-022-01547-4](https://doi.org/10.1038/s41597-022-01547-4) (2022).
27. Provenzano, S. & Bull, H. The local economic impact of mineral mining in africa: Evidence from four decades of satellite imagery. *ArXiv e-prints* [10.48550/arXiv.2111.05783](https://arxiv.org/abs/2111.05783) (2021).

28. Tang, L. & Werner, T. T. Global mining footprint mapped from high-resolution satellite imagery. *Commun. Earth & Environ.* **4**, 134, [10.1038/s43247-023-00805-6](https://doi.org/10.1038/s43247-023-00805-6) (2023).
29. Quash, Y., Kross, A. & Jaeger, J. A. G. Assessing the impact of gold mining on forest cover in the Surinamese Amazon from 1997 to 2019: A semi-automated satellite-based approach. *Ecol. Informatics* **80**, 102442, [10.1016/j.ecoinf.2023.102442](https://doi.org/10.1016/j.ecoinf.2023.102442) (2024).
30. Maus, V. & Werner, T. T. Impacts for half of the world's mining areas are undocumented. *Nature* **625**, 26–29, [10.1038/d41586-023-04090-3](https://doi.org/10.1038/d41586-023-04090-3) (2024).
31. Xie, E. *et al.* Segformer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **34**, 12077–12090, [10.48550/arXiv.2105.15203](https://doi.org/10.48550/arXiv.2105.15203) (2021).
32. Archi Indonesia. Public expose 2024. https://www.idx.co.id/StaticData/NewsAndAnnouncement/ANNOUNCEMENTSTOCK/From_EREP/202406/fdfde002ad_a4aef307c1.pdf (2024).
33. Verweijen, J., Schouten, P., Simpson, F. O. & Pascal, C. Z. Conservation, conflict and semi-industrial mining: the case of eastern DRC. *Analysis Policy Briefs* (2022).
34. Werner, T. T. *et al.* Patterns of infringement, risk, and impact driven by coal mining permits in Indonesia. *Ambio* **53**, 242–256, [10.1007/s13280-023-01944-y](https://doi.org/10.1007/s13280-023-01944-y) (2024).
35. Ferrante, L. & Fearnside, P. M. Brazil threatens indigenous lands. *Science* **368**, 481–482, [10.1126/science.abb6327](https://doi.org/10.1126/science.abb6327) (2020).
36. Statistics South Africa. P2401 mining: Production and sales (2024).
37. Hilson, G. & McQuilken, J. Four decades of support for artisanal and small-scale mining in sub-Saharan Africa: A critical review. *Extr. Ind. Soc.* **1**, 104–118, [10.1016/j.jexis.2014.01.002](https://doi.org/10.1016/j.jexis.2014.01.002) (2014).
38. Maus, V. *et al.* A global-scale data set of mining areas. *Sci. Data* **7**, 1–13, [10.1038/s41597-020-00624-w](https://doi.org/10.1038/s41597-020-00624-w) (2020).
39. Tang, L., Werner, T. T., Heping, X., Jingsong, Y. & Zeming, S. A global-scale spatial assessment and geodatabase of mine areas. *Glob. Planet. Chang.* **204**, 103578, [10.1016/j.gloplacha.2021.103578](https://doi.org/10.1016/j.gloplacha.2021.103578) (2021).
40. Vaswani, A. *et al.* Attention is all you need. *Conf. on Neural Inf. Process. Syst.* **30**, [10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762) (2017).
41. Li, X. *et al.* Transformer-based visual segmentation: A survey. *ArXiv e-prints* [10.48550/arXiv.2304.09854](https://doi.org/10.48550/arXiv.2304.09854) (2023).
42. Papers with Code. Semantic segmentation on ADE20K. <https://paperswithcode.com/sota/semantic-segmentation-on-ade20k> (2025).
43. Papers with Code. Semantic segmentation on Cityscapes. <https://paperswithcode.com/sota/semantic-segmentation-on-cityscapes> (2025).
44. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241, [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28) (Springer, Cham, Switzerland, 2015).
45. Sourget, T., Hasany, S. N., Mériauudeau, F. & Petitjean, C. Can SegFormer be a true competitor to U-Net for medical image segmentation? In *Medical Image Understanding and Analysis*, 111–118, [10.1007/978-3-031-48593-0_8](https://doi.org/10.1007/978-3-031-48593-0_8) (Springer, Cham, Switzerland, 2023).

46. Luo, W., Li, Y., Urtasun, R. & Zemel, R. Understanding the effective receptive field in deep convolutional neural networks. In *Conference on Neural Information Processing Systems*, 4905–4913, [10.48550/arXiv.1701.04128](https://arxiv.org/abs/1701.04128) (Curran Associates Inc., 2016).
47. Russakovsky, O. *et al.* ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis. (IJCV)* **115**, 211–252, [10.1007/s11263-015-0816-y](https://doi.org/10.1007/s11263-015-0816-y) (2015).
48. Shrivastava, A., Gupta, A. & Girshick, R. Training region-based object detectors with online hard example mining. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27–30, [10.1109/CVPR.2016.89](https://doi.org/10.1109/CVPR.2016.89) (IEEE, 2016).
49. Suzuki, S. & Be, K. Topological structural analysis of digitized binary images by border following. *Comput. Vision, Graph. Image Process.* **30**, 32–46, [10.1016/0734-189X\(85\)90016-7](https://doi.org/10.1016/0734-189X(85)90016-7) (1985).
50. Jordahl, K. *et al.* Geopandas 1.0.1, [10.5281/zenodo.2585848](https://doi.org/10.5281/zenodo.2585848) (2024).
51. Gillies, S. & contributors, S. Shapely: A Python package for manipulation and analysis of planar geometric objects. <https://shapely.readthedocs.io/en/stable/index.html> (2024).
52. Pebesma, E. & Bivand, R. *Spatial Data Science: With applications in R* (Chapman and Hall/CRC, 2023).
53. Paszke, A. *et al.* PyTorch: An imperative style, high-performance deep learning library. In *Conference on Neural Information Processing Systems*, 8026–8037, [10.48550/arXiv.1912.01703](https://arxiv.org/abs/1912.01703) (Curran Associates Inc., 2019).
54. MMSegmentation. OpenMMLab Semantic Segmentation Toolbox and Benchmark. <https://github.com/open-mmlab/mmsegmentation> (2020).

Acknowledgements

The authors gratefully acknowledge financial support from the Austrian National Bank (OeNB anniversary fund, project No. 18799) and the City of Vienna (Hochschuljubiläumsfonds, project No. H-457973/2023). Computations were performed on the WUCluster, a facility supported by the Core Services of the Vienna University of Economics and Business.

Author contributions statement

N.K and P.S conceived the study. P.S implemented the method. All authors wrote and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Supplementary Information

Here we provide some additional figures and tables that allow for more detailed insights in the produced dataset as well as the underlying ground truth data and satellite imagery. [Figure S1](#) decomposes the global version [Figure 3](#) in the manuscript to depict regional summaries of the predictions over time. It shows the development of the total area and number of mining polygons for the three countries within each region that had the highest share on overall mining polygon area separately as well as for the remaining countries aggregated.

[Table S1](#) similarly provides information on the total mining polygon area for countries with a large mining footprint (defined as exceeding 100 km^2 at least once during the observation period). It extends [Table 1](#) in the manuscript and confirms the general trend of increasing total mining polygon area over the years also within most countries.

[Figure S2](#) shows satellite imagery from 2019 for four mine sites across the covered area overlayed with polygons from the two datasets that constitute the ground truth used for training the segmentation model. It shows that the rather broadly defined delineations of Maus et al. (2022)²⁶ often encompass the more precisely delineated ones from Tang and Werner (2023)²⁸, which also sometimes pick up additional mining features.

[Figure S3](#) and [Table S2](#) provide some additional information about the results from the steps undertaken to select the images with highest clarity (i.e., lowest cloud and heavy haze coverage) from a 6-month window (June–November) for each mining site individually. [Table S2](#) shows that the clarity of selected yearly satellite images for the period 2020–2024 is on average comparable or even exceeds the one for period 2016–2019, where we take the biannual composites as is. [Figure S3](#) provides an example for the importance of this procedure to select highest-clarity images, showing two composite images from October and December 2024 for the area surrounding the Toka Tindung mining site.

[Figure S4](#) shows some of the validation metrics discussed in the Methods section of the paper over time for two different probability thresholds. It shows limited variability in most of these metrics over time, which mostly stems from a lack of variation in the labels of the validation points. Yet, it can be discerned that our modeling approach achieves performance comparable to the ground truth and its constituting datasets over the year. Differences in metrics across the selected probability thresholds over the also emulate the ones shown in [Figure 5](#).

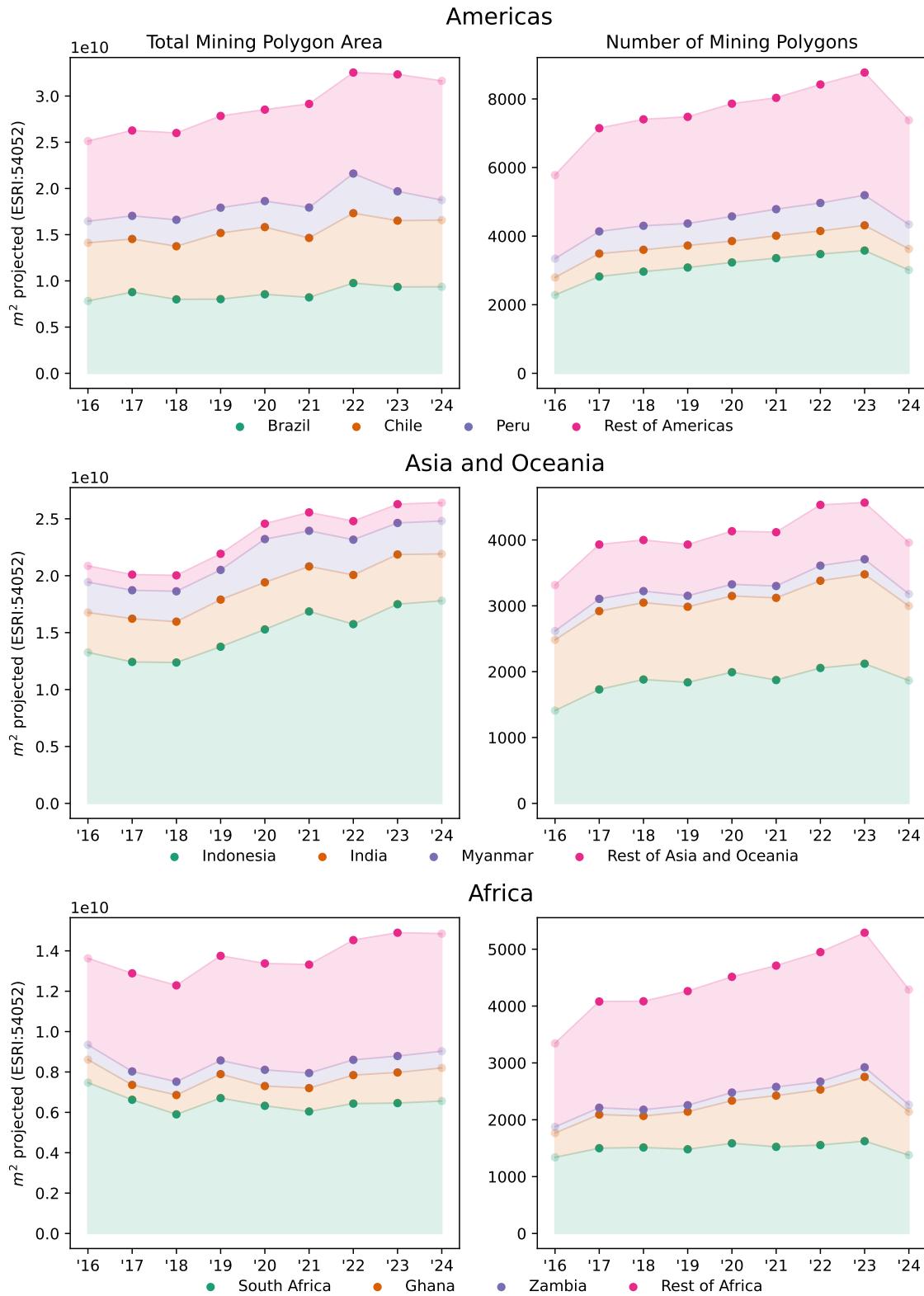


Figure S1. Regional summaries of the predictions over time; the left column shows total mining areas, and the right the number of mine polygons. Note, in particular, the strong increases in area for Chile and Peru in 2022 (top row), the sharp decrease in area following years of steady expansion for Indonesia in 2022 (middle row), and the steady expansion in the number of mining polygons coupled with only modest increases in area for Ghana (bottom row).

| Country | Total Mining Polygon Area (km ²) | | | | | | | | |
|------------------|--|--------|--------|--------|--------|--------|--------|--------|--------|
| | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
| Indonesia | 13,266 | 12,425 | 12,376 | 13,761 | 15,281 | 16,859 | 15,746 | 17,497 | 17,801 |
| Brazil | 7,829 | 8,785 | 8,002 | 8,023 | 8,548 | 8,215 | 9,764 | 9,342 | 9,360 |
| Chile | 6,293 | 5,748 | 5,746 | 7,161 | 7,269 | 6,434 | 7,561 | 7,176 | 7,218 |
| South Africa | 7,469 | 6,624 | 5,901 | 6,706 | 6,324 | 6,044 | 6,434 | 6,459 | 6,558 |
| India | 3,502 | 3,800 | 3,590 | 4,140 | 4,136 | 3,959 | 4,322 | 4,367 | 4,114 |
| Myanmar | 2,670 | 2,497 | 2,667 | 2,607 | 3,800 | 3,129 | 3,099 | 2,769 | 2,890 |
| Bolivia | 557 | 787 | 806 | 1,025 | 910 | 1,103 | 1,108 | 2,035 | 2,695 |
| Guyana | 2,082 | 2,245 | 2,368 | 2,246 | 2,380 | 3,213 | 2,942 | 3,197 | 2,570 |
| Suriname | 1,906 | 1,824 | 1,902 | 1,986 | 1,976 | 2,111 | 2,119 | 2,382 | 2,430 |
| Peru | 2,328 | 2,502 | 2,860 | 2,735 | 2,816 | 3,287 | 4,286 | 3,172 | 2,165 |
| Venezuela | 1,183 | 1,354 | 1,285 | 1,341 | 1,414 | 1,442 | 1,310 | 1,528 | 1,688 |
| Ghana | 1,143 | 735 | 957 | 1,189 | 979 | 1,158 | 1,411 | 1,516 | 1,641 |
| Namibia | 596 | 1,037 | 925 | 1,217 | 1,135 | 1,043 | 1,503 | 1,381 | 1,246 |
| Colombia | 961 | 908 | 885 | 1,013 | 927 | 1,067 | 1,083 | 1,117 | 1,083 |
| Mexico | 812 | 847 | 864 | 957 | 931 | 959 | 1,060 | 1,023 | 998 |
| Argentina | 816 | 916 | 898 | 933 | 931 | 884 | 816 | 875 | 963 |
| Zambia | 732 | 668 | 660 | 677 | 803 | 745 | 757 | 814 | 830 |
| Dem. Rep. Congo | 615 | 622 | 629 | 680 | 732 | 692 | 732 | 794 | 806 |
| Zimbabwe | 505 | 458 | 484 | 472 | 508 | 585 | 607 | 697 | 674 |
| Angola | 500 | 440 | 421 | 442 | 487 | 540 | 479 | 543 | 562 |
| Vietnam | 478 | 458 | 443 | 446 | 419 | 519 | 536 | 516 | 523 |
| Philippines | 372 | 370 | 407 | 415 | 395 | 428 | 456 | 489 | 430 |
| Botswana | 297 | 343 | 366 | 353 | 351 | 354 | 334 | 379 | 360 |
| Malaysia | 268 | 273 | 270 | 257 | 237 | 300 | 320 | 331 | 331 |
| Mali | 252 | 219 | 232 | 276 | 251 | 295 | 358 | 355 | 317 |
| Guinea | 214 | 239 | 213 | 263 | 242 | 290 | 278 | 294 | 236 |
| France | 193 | 227 | 257 | 218 | 189 | 216 | 247 | 242 | 233 |
| Burkina Faso | 188 | 223 | 190 | 206 | 184 | 191 | 203 | 204 | 227 |
| Thailand | 200 | 182 | 196 | 224 | 204 | 231 | 210 | 198 | 221 |
| Mozambique | 188 | 177 | 197 | 120 | 211 | 224 | 210 | 219 | 218 |
| Tanzania | 212 | 199 | 171 | 208 | 190 | 212 | 204 | 236 | 212 |
| Papua New Guinea | 150 | 197 | 219 | 220 | 240 | 407 | 221 | 218 | 199 |
| Cuba | 147 | 124 | 138 | 137 | 133 | 166 | 180 | 172 | 158 |
| Senegal | 104 | 127 | 143 | 128 | 134 | 133 | 130 | 127 | 149 |
| Niger | 142 | 181 | 171 | 178 | 195 | 138 | 207 | 138 | 143 |
| Sierra Leone | 93 | 113 | 123 | 132 | 143 | 134 | 139 | 152 | 136 |
| Ecuador | 62 | 70 | 74 | 92 | 118 | 73 | 117 | 119 | 121 |
| Côte d'Ivoire | 60 | 74 | 96 | 99 | 100 | 113 | 129 | 132 | 111 |

Table S1. Total mining polygon area in the countries with a large mining footprint ($\geq 100 \text{ km}^2$) in the dataset over time.

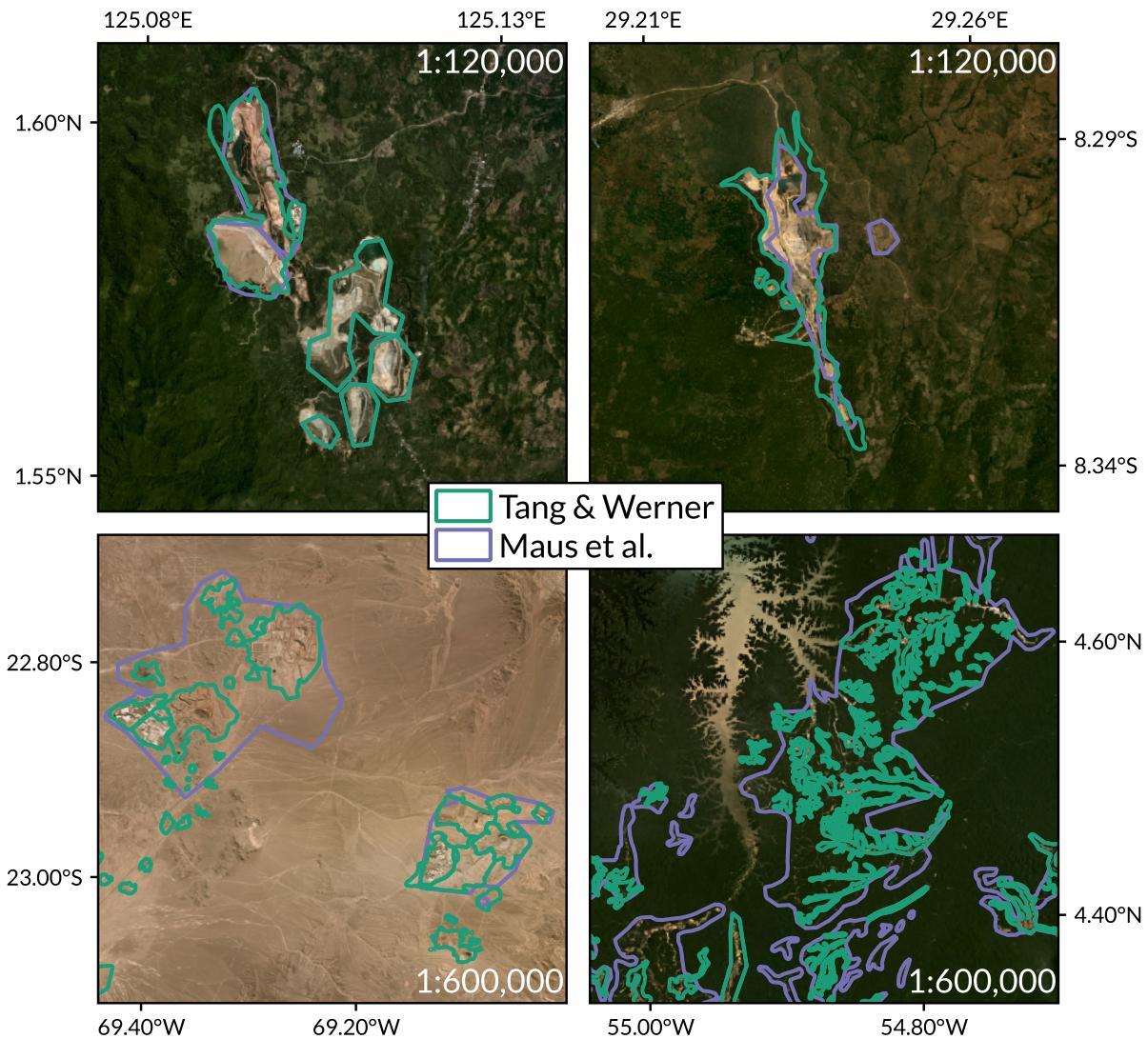


Figure S2. Four examples of the source polygons for our ground truth.^{26,28} The mine sites are located in Indonesia (Toka Tindung, from Figure 2), the DRC (Kapulo Copper Mine), Chile (Minera Sierra Gorda), and Suriname (south of the Brokopondo Reservoir) against 2019 Planet/NICFI imagery. Note the different zoom level of the top and bottom rows.

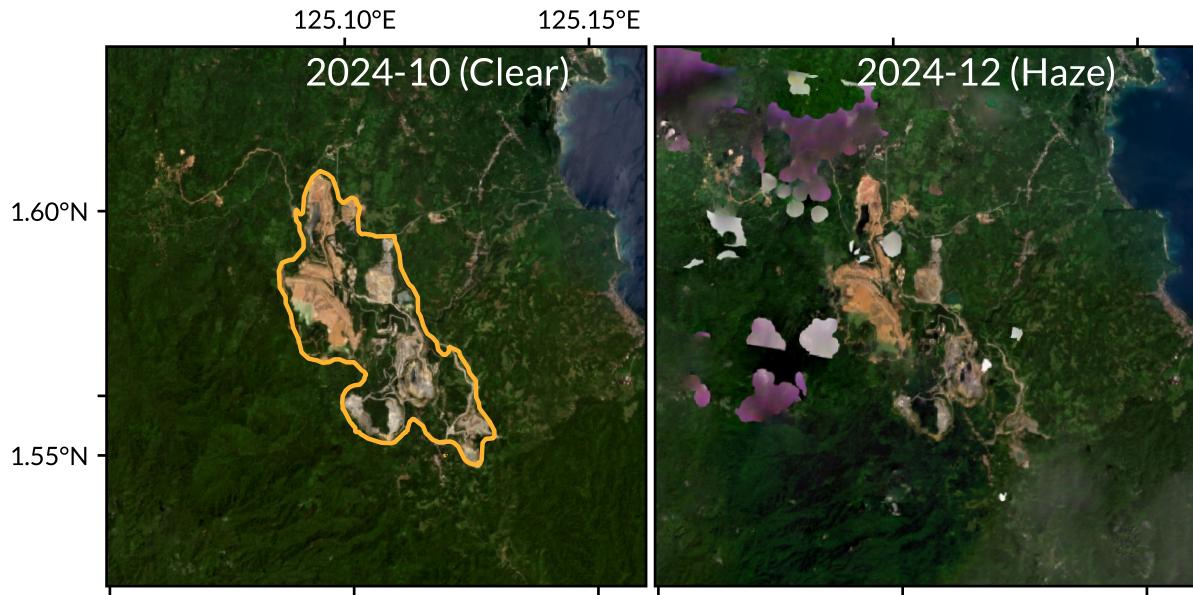


Figure S3. Two satellite images (Planet/NICFI) of the Toka Tindung mine (from [Figure 2](#)) for 2024. The left image is a composite from October 2024 and features high clarity (99.9%), while the latest image from December 2024 has clearly visible artifacts, reducing its clarity. Our approach selects the highest-clarity image from a six-month window, leading to high-quality images throughout.

| Year | Mean | SD | Percentiles | | | | |
|------|-------|---------|-------------|-------|-------|-------|------|
| | | | 1st | 5th | 10th | 50th | 90th |
| 2016 | 94.29 | (12.32) | 38.00 | 67.00 | 81.36 | 99.64 | 100 |
| 2017 | 95.24 | (10.21) | 46.94 | 72.83 | 87.00 | 99.96 | 100 |
| 2018 | 94.55 | (10.13) | 57.19 | 67.54 | 80.52 | 99.33 | 100 |
| 2019 | 97.05 | (8.15) | 63.43 | 73.79 | 89.48 | 100.0 | 100 |
| 2020 | 97.44 | (7.28) | 65.46 | 77.64 | 91.00 | 100.0 | 100 |
| 2021 | 97.48 | (7.57) | 62.73 | 78.07 | 94.25 | 100.0 | 100 |
| 2022 | 95.61 | (11.25) | 53.70 | 60.80 | 84.11 | 100.0 | 100 |
| 2023 | 95.80 | (10.77) | 53.74 | 65.16 | 82.20 | 100.0 | 100 |
| 2024 | 96.01 | (8.97) | 66.79 | 72.14 | 79.16 | 100.0 | 100 |

Table S2. Summary statistics for the visibility of selected yearly satellite images. Note that the visibility metadata is only available on a per-scene basis; we average scene-level characteristics to compute the relevant quad-level information.

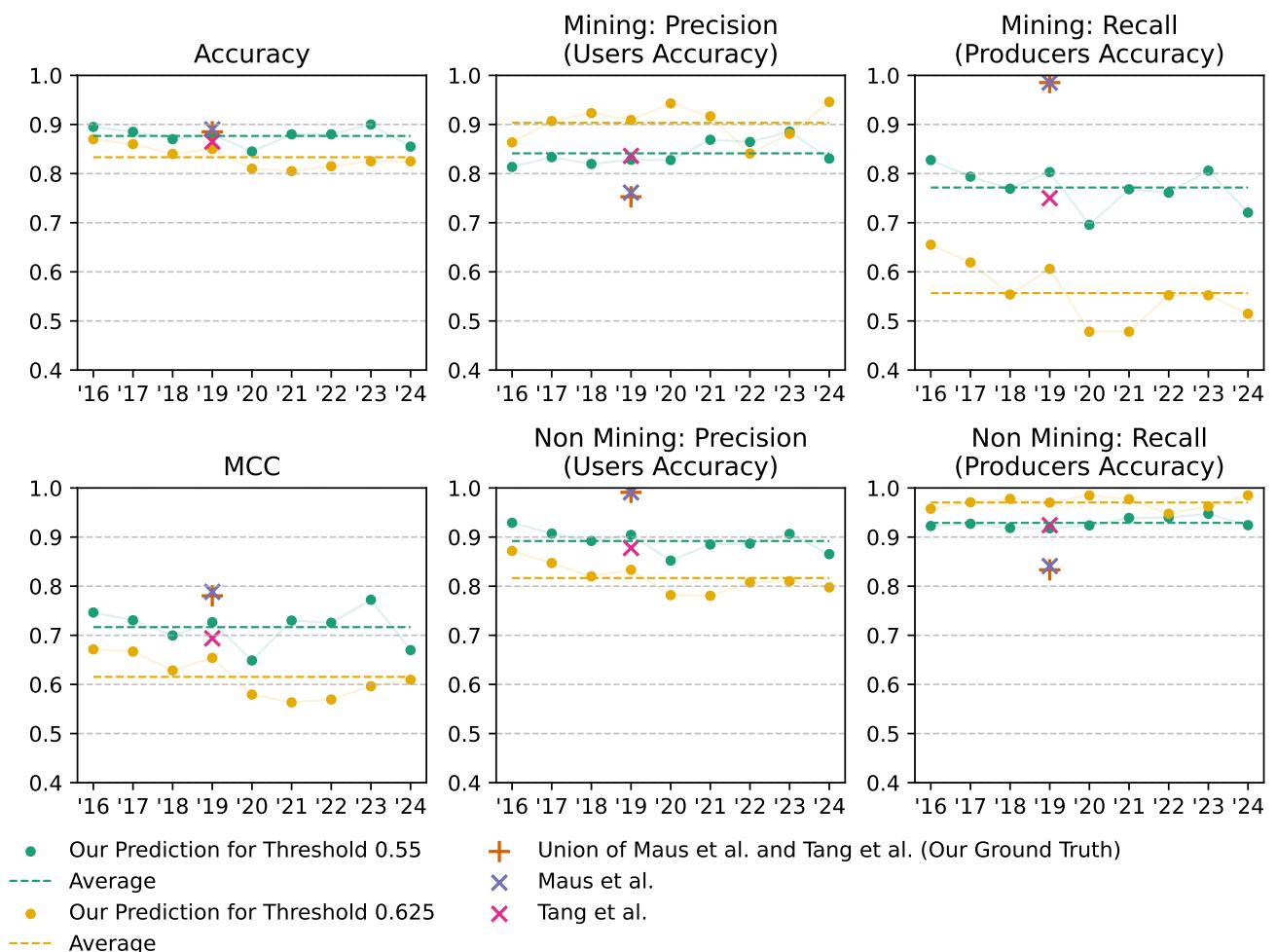


Figure S4. Reference point validation for individual years.