Supplementary Information for

### **Does formalizing artisanal gold mining mitigate environmental impacts? Deforestation evidence from the Peruvian Amazon**

Nora Álvarez-Berríos<sup>1,4</sup>, Jessica L'Roe<sup>2,4</sup> and Lisa Naughton-Treves<sup>3,4</sup>

<sup>1</sup>International Institute for Tropical Forestry, Río Piedras, Puerto Rico, 00926 Department of Geography, Middlebury College, Middlebury, USA, 05753 Department of Geography, University of Wisconsin, Madison, USA, 53706 Authors arranged alphabetically

Corresponding author: Nora Álvarez-Berríos Email: alvarez.nora@gmail.com

#### **Methods**

**Image Processing.** Landsat images (TM and ETM+) were acquired as surface reflectance climate data record products from the USGS Earth Resources Observation and Science Center Processing Architecture [\(http://espa.cr.usgs.gov\)](http://espa.cr.usgs.gov/). Scenes with 60% or less cloud coverage were chosen for the study region (path/row 2/69 and 3/69) for years 2001 to 2014. Imagery for 2002 and 2012 were not available due to cloud cover and sensor malfunction. Image processing steps on the acquired images were conducted in Python and included: (i) image reprojection to WGS84\_UTM zone19S, (ii) image normalization (1), (iii) image mosaicking, and (iv) study area extraction. Additionally, for each image date we derived spectral indices used to increase the separability of mined areas from other land cover types, including the Normalized Difference Vegetation Index (NDVI), Tasseled Cap indices (Brightness, Greenness, and Wetness), and the three principal components from a Principal Component Analysis (PCA) (2).

The Random Forest tree-based classifier in R was used to classify the processed images (3, 4). We collected reference samples for classifier training and model accuracy assessment by consulting high-resolution images available in Google Earth (e.g., Quickbird) and by characterizing the Landsat images based on fieldwork and experience on-site. Over 80,000 samples (pixels) were collected for mining areas, forests, agricultural/herbaceous fields, riverbanks, bare land, water, clouds and shadows. We used the per pixel probabilities to assign land cover classes based on the maximum class probability. Individual RF models were trained for each image date. The resultant RF models had OOBs (out-of-bag errors) accuracy estimates ranging from 0.94 to 0.99.

**Data Structure.** We used a hexagonal grid (5) with 25 ha hexagon cells to balance between a unit large enough to minimize spatial autocorrelation but small enough to meaningfully coincide with the scale of mining and titling (25 ha is a quarter of the size of the smallest mining concessions). All predictor variables were based on the status of the centroid of each hexagon. We used polygons rather than points as units so that we could examine annual clearing sizes (6). Figure S1 gives a sense of the variation across scales of the study site, the hexagon cell, and the area mined each year.

**Creating the Matched Data Subset.** To help account for the fact that parts of our study region outside of the mining corridor may be places where gold mining is infeasible, we did a pre-regression step using matching methods to create an appropriately trimmed and weighted spatial set of observations on which to run our temporally explicit models. We used Stata 16's **teffects nnmatch** module to generate nearest-neighbor matches outside the mining corridor for cells within the corridor. We sought exact matches for district and geology class, and closest match possible for distance to navigable rivers using the Mahalanobis distance

metric. We eliminated regions outside of common support and generated a new subsample that consisted of mining corridor cells and their best nearest neighbor. We eliminated the bottom (worst) 10% of matches. Control cells outside the corridor could be used multiple times as matches for 'treated' cells inside the mining corridor. These steps trimmed our dataset from 83,428 hexagon cells to the 37,152 cells that we use in our regression models.

**Event Study.** As a robustness check, we used an event study framework to examine mining deforestation as it related to the timing of initiating the titling process. We excluded extinguished titles and titles that were given prior to the study period. As in our other analyses, we used a panel model with fixed effects for each 25-ha cell and for each year, with standard errors clustered at the level of 5km hexagons. We used the **eventdd** package written for Stata, with time-to-event set in reference to the year of initiating the titling process, including 7 leads and 7 lags (given our 14-year study period) and a baseline set at 5 years prior to titling (7). See results in Figure S3.

**Other Model Robustness Checks.** We conducted robustness checks with district-level time trends, the complete sample of hexagons prior to matching, different thresholds for amount of mining that needed to be detected to switch on the binary indicator, and without excluding extinguished titles; these yielded substantively similar results. Presented models exclude 2001 when mining deforestation was erratically low, likely due to the remote sensing masking procedure removing areas mined before 2000; the main results are not affected when it is included.



**Fig. S1.** Study site, unit of analysis, and outcome variable: new mining-related deforestation each year.



**Fig. S2.** Mining titles issued during 1977 to 2014. As of 2014, none of the titles were more than 'provisional' and some (in gray) had been extinguished due to unpaid annual fees.



**Fig. S3.** Robustness check of title analysis using an event study framework. Probability of mining increases in the years following initiation of the titling process (at  $T=0$ ), and to some extent in the years immediately preceding, relative to a baseline of 5 years before titling. Dots are point estimates and blue lines are 95% confidence intervals.

# **Table S1.** Selected Formalization Events from 2002 to 2014 (*See sources below*)







## **Table S2.** Data Used in Models of Formalization Impacts.



# **Table S3.** List of Respondents, Roles and Affiliations (9 women, 38 men).



**Table Note**: The number of participants in focus groups is specified in parenthesis and bold font. Total number of participants in one-on-one interviews and group interviews = 47





a Some titles were extinguished during our study period because annual fees were not paid. We do not know the year in which titles were granted for these extinguished titles, thus we exclude them from our title-based analysis because we can't know whether they were active in a given year.

b We exclude instances of mining area less than .27 ha (3 pixels) because it is less certain that these are clearings related to mining vs. shifting river banks, crop fields, etc. Including these smaller clearings does not substantially change area totals – for example, overall mining in 2014 would increase from 38,399 to 38,630 ha.

c The dotted line highlights when the Mining Corridor was formally established (2010).



### **Table S5.** Model Results for Changes in Mining with Titling

Significance thresholds indicated with following symbols: †: p<0.1, \*: p<0.05, \*\*: p<0.01, \*\*\*: p<0.001

Model 1 is a linear probability model, Model 2 is a semi-elasticity model of ln(Area Cleared) conditional on any mining occurring, and Model 3 is the inverse hyperbolic sine of Area Cleared, unconditional on clearing.

All models use clustered standard errors based on 5km hexagons.

	Inverse Hyperbolic Sine of Mined Area	
	A	В
Post- $2010 \text{ x}$ <b>Outside Corridor</b>	$-0.11***$	$-0.19***$
Post- $2010 \times in a$ Protected Area		0.02
Post- $2010 \times in a$ <b>Buffer Zone</b>		$0.15*$
Post- $2010 \times in a$ Native Community		$0.16**$
Post- $2010 \times$ within 0.5km of Nav. River		$0.04*$
Post-2010	$0.28***$	$0.26***$
Constant	$-0.02$	0.00
Effects for each year	yes	yes
Cell fixed effects	yes	yes
$R^2$ - within	0.08	0.06
$R^2$ - between	0.06	0.02
rho	0.66	0.64
Num obs	520,128	520,128
Num groups / panels	37,152	37,152
Num. clusters for se's	344	344

**Table S6.** Corridor Analysis Robustness Check using Inverse Hyperbolic Sine of Mining

Significance thresholds indicated with following symbols: †: p<0.1, \*: p<0.05, \*\*: p<0.01, \*\*\*: p<0.001

This model generally accords with Models 1 and 2 in Table 1. Inverse-hyperbolic sine transformations can be interpreted conveniently as approximately the percent change in mined area corresponding to a step change in a predictor variable, unconditional on mining occurring (17).

After 2010, Model A shows an 11% decline in mined area outside the corridor relative to inside the corridor. This does not accord with the measured rise in proportion of mined area outside the corridor apparent in the descriptive statistics. The weight of a few significant and localized areas of expansion is imperfectly captured with the inverse hyperbolic sine transformation, whose distribution is still heavily skewed.

## **Table S7.** Illustrative Quotes.





**Table Note:** Interview transcriptions were codified via Dedoose (Version 8.0.35 (18)). To protect the informants' anonymity, we replaced names with a code.

#### **S1 References**

- 1. J. Graesser, N. Ramankutty, O. T. Coomes, Increasing expansion of large-scale crop production onto deforested land in sub-Andean South America. *Environ. Res. Lett.* **13** (2018).
- 2. P. A. Townsend, *et al.*, Changes in the extent of surface mining and reclamation in the Central Appalachians detected using a 1976–2006 Landsat time series. *Remote Sens. Environ.* **113**, 62–72 (2009).
- 3. A. Liaw, M. Wiener, Classification and regression by random forest. *R News* **2**, 18–22 (2002).
- 4. C. Dezécache, *et al.*, Gold-rush in a forested El Dorado: deforestation leakages and the need for regional cooperation. *Environ. Res. Lett.* **12**, 034013 (2017).
- 5. C. P. D. Birch, S. P. Oom, J. A. Beecham, Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecol. Modell.* **206**, 347–359 (2007).
- 6. A. Blackman, "Ex Post Evaluation of Forest Conservation Policies Using Remote Sensing Data: An Introduction and Practical Guide" (2012).
- 7. D. Clarke, K. Schythe, Implementing the panel event study (2020).
- 8. Defensoría del Pueblo, "Informe Defensorial N 167: Balance de la gestión estatal frente a la Minería Informal e Ilegal en el Perú, 2012–2014, supervisión a las entidades de alcance nacional y regional," (2014).
- 9. G. H. Damonte, The "Blind" State: Government Quest for Formalization and Conflict with Small‐Scale Miners in the Peruvian Amazon. *Antipode* **48**, 956–976 (2016).
- 10. D. Cortés-McPherson, Expansion of small-scale gold mining in Madre de Dios:'capital interests' and the emergence of a new elite of entrepreneurs in the Peruvian Amazon. *Extr. Ind. Soc.* **6**, 382–389 (2019).
- 11. ProNaturaleza, SER, "Línea del tiempo minería Observatorio de infraestructura e industrias extractivas" (2014).
- 12. C. J. Rubio, El impacto de la minería aurífera en el Departamento de Madre de Dios (Peru). *Obs. Medioambient.* **13**, 169–202 (2010).
- 13. J. J. Swenson, C. E. Carter, J.-C. Domec, C. I. Delgado, Gold mining in the Peruvian Amazon: global prices, deforestation, and mercury imports. *PLoS One* **6**, e18875 (2011).
- 14. M. Salo, *et al.*, Local perspectives on the formalization of artisanal and small-scale mining in the Madre de Dios gold fields, Peru. *Extr. Ind. Soc.* **3**, 1058–1066 (2016).
- 15. L. Valencia, *Políticas de pequeña minería y deforestación: el caso de Madre de Dios* (Sociedad Peruana de Derecho Ambiental, 2016).
- 16. C. A. Ipenza Peralta, *Manual para entender: la pequeña minería y la minería artesanal y los decretos legislativos vinculados a la minería ilegal* (Sociedad Peruana de Derecho Ambiental, 2012).
- 17. J. B. Burbidge, L. Magee, A. L. Robb, Alternative transformations to handle extreme values of the dependent variable. *J. Am. Stat. Assoc.* **83**, 123–127 (1988).
- 18. Socio Cultural Research Consultants LLC, Dedoose Version 8.0.35, web application for managing, analyzing, and presenting qualitative and mixed-method research data (2018).