



OPEN

## Quasi-experimental analysis of new mining developments as a driver of deforestation in Zambia

Jonathan Morley<sup>1✉</sup>, Graeme Buchanan<sup>2</sup>, Edward T. A. Mitchard<sup>1</sup> & Aidan Keane<sup>1</sup>

Mining is a vital part of the global, and many national, economies. Mining also has the potential to drive extensive land cover change, including deforestation, with impacts reaching far from the mine itself. Understanding the amount of deforestation associated with mining is important for conservationists, governments, mining companies, and consumers, yet accurate quantification is rare. We applied statistical matching, a quasi-experimental methodology, along with Bayesian hierarchical generalized linear models to assess the impact on deforestation of new mining developments in Zambia from 2000 to present. Zambia is a globally significant producer of minerals and mining contributes ~ 10% of its gross domestic product and ~ 77% of its exports. Despite extensive deforestation in mining impacted land, we found no evidence that any of the 22 mines we analysed increased deforestation compared with matched control sites. The extent forest lost was therefore no different than would likely have happened without the mines being present due to other drivers of deforestation in Zambia. This suggests previous assessments based on correlative methodologies may overestimate the deforestation impact of mining. However, mining can have a range of impacts on society, biodiversity, and the local environment that are not captured by our analysis.

The extraction of minerals and metals through mining is key to many aspects of the global economy and society<sup>1</sup>. For many low- and middle-income countries, mining makes up the majority of their national gross domestic product (GDP), exports, and government revenues<sup>2</sup>. However, mining also involves extensive land use and can alter socio-ecological systems with impacts ranging far from the mine itself<sup>3–5</sup>. As much as 49.9 million km<sup>2</sup> of land could be considered influenced by mining globally<sup>6</sup> and it is likely an understudied contributor to deforestation in mineral-rich countries<sup>7,8</sup>. Given the spatial overlap of mineral deposits and areas of high biodiversity<sup>5</sup>, and the predicted increase in demand for mined resources over coming decades, there is potentially a significant risk to biodiversity from mining<sup>6,9</sup>.

Of the wide range of impacts mining can have on nearby ecosystems, deforestation is perhaps the one most clearly measurable at large spatial and temporal scales<sup>4</sup>. Within the mining lease itself vegetation is cleared during mineral extraction, when waste rock is dumped, and for the creation of dams to store tailings (toxic by-products of mineral extraction)<sup>10</sup>. Secondary infrastructure such as processing plants, buildings, and access roads, along with timber requirements for the mine, lead to further forest clearance beyond a mining lease boundary<sup>11,12</sup>. Additional indirect impacts include influxes of workers and their families to previously sparsely populated areas who require infrastructure such as housing, timber for building, fuel, and charcoal production, as well as land for agriculture to supplement household incomes<sup>11–13</sup>. These secondary impacts can often be much larger than the direct initial clearing<sup>14</sup> and can happen either organically or as part of planned economic development corridors where mining and other developments are connected by linear infrastructure to open up new areas to global markets<sup>9,15</sup>. Mining developments can also cause the downgrading, downsizing, and degazetting of protected areas<sup>16</sup>, both as a consequence of a mining licence being issued and in response to increased pressure from population shifts<sup>9,13</sup>. An increase in forest cover is also possible in the proximity of a mine due to land management activities by the mining company as part of a restoration or compensation measures<sup>11,17</sup>. Understanding the deforestation footprint of mining developments is important as it represents impacts to biodiversity, through the loss of habitat<sup>3,18</sup>, the physical environment, through the loss of carbon and changes in ecosystem functions<sup>19,20</sup>, and people, through the loss of timber and non-timber forest products<sup>21,22</sup>.

Understanding the land cover change, especially deforestation, associated with mining is important for conservation stakeholders, governments, mining companies, and consumers higher up the supply chain interested in sustainability and ecological footprints<sup>11,23</sup>. However, the existing literature is limited<sup>3,23</sup> and most studies

<sup>1</sup>School of GeoSciences, University of Edinburgh, Edinburgh, UK. <sup>2</sup>RSPB Centre for Conservation Science, RSPB Scotland, Edinburgh, UK. ✉email: jonathan.morley@ed.ac.uk

are correlative, meaning that deforestation within or near mining developments is measured and it is assumed that observed deforestation is attributable to mining activity<sup>4</sup> and would not have occurred in the absence of mining. For example, in the Madre de Dios department of Peru, numerous studies have attributed high rates of deforestation to gold mining<sup>24–27</sup>. A recent study using Landsat derived land cover maps validated with very high-resolution imagery found 2,096 ha yr<sup>-1</sup> (10,482 ha total) of forest loss between 2013 and 2018 within mining concessions 51% of the total deforestation for the region<sup>24</sup>. Previous studies in the same region of Peru had produced similar estimates of 6,145 ha yr<sup>-1</sup> of forest loss between 2008 and 2012<sup>25</sup> and 6,600 ha yr<sup>-1</sup> between 2003 and 2009 which they attribute to new mining developments<sup>26</sup>. A study that utilised multiple remotely sensed datasets and covered a wider period (1985–2017) had an even greater estimate of 95,751 ha of deforestation, 64,586 ha of which happened between 2010 and 2017<sup>27</sup>. Elsewhere, a recent study using Landsat derived land cover maps estimated there had been 47,000 ha of deforestation across Ghana between 2005 and 2019 that can be attributed to the expansion of mining developments<sup>28</sup>. This follows previous, methodologically similar, studies which identified mining as the main driver of deforestation in the Asutifi North District<sup>29</sup> and the Ankobrah River<sup>30,31</sup>. The methodology these studies use gives a measure of the direct forest loss in the vicinity of new mining developments. While this is valuable, these studies do not have a comparison to counterfactuals (what would have happened without the mines' presence), and are therefore unable to estimate the net causal effect of mining on deforestation above the background rate<sup>32</sup>.

A few previous studies have used causal inference methodologies to estimate the effect of mining on deforestation separate from other drivers<sup>4</sup>. Butsic and colleagues used a panel instrumental variable methodology to estimate the effects of mining, armed conflicts, and protected areas on deforestation in the Democratic Republic of Congo<sup>33</sup>. They found mining had a significant detrimental effect on forest cover, increasing loss at a coarse spatial and temporal scale<sup>33</sup>. Another study of mining in India used propensity score matching at a very coarse spatial scale and found a positive association between mineral outputs at the district level and loss of forest cover with variation based on the type of minerals extracted<sup>34</sup>. In the Brazilian Amazon, one study used propensity score matching with post-matching regression to assess the impact of mining on deforestation in 1 km<sup>2</sup> pixels<sup>32</sup>. They found that mining had driven extensive deforestation in the region, with deforestation being significantly higher up to 70 km from a mining lease, and estimated that 11,670 km<sup>2</sup> of forest had been lost due to mining<sup>32</sup>. Another recent study undertook a pan-tropical analysis and found extensive forest loss directly associated with mining and a strong association between distance to mine and deforestation in the wider landscape for the majority of countries analysed<sup>35</sup>. However, the post-matching regression approach used in<sup>32</sup> and<sup>35</sup> do not account for the spatial structure of the data and neither study tested for residual autocorrelation, suggesting that the results are potentially vulnerable to type I errors<sup>36</sup> and may have inflated effect sizes<sup>37</sup>.

Here, we applied a quasi-experimental study design along with Bayesian hierarchical generalized linear spatiotemporal models to estimate the effect of new mining developments on deforestation in Zambia, a globally significant mineral exporter. Mining is key to the Zambian economy contributing 77% of all exports, 10% of GDP, and 29% of government revenues<sup>38</sup>. Zambia is the seventh largest copper producer globally and has significant industries for many other minerals and metals, including coal, precious and semi-precious stones (including 20% of the world's emeralds), and limestone<sup>38</sup>. Zambia is a forested country, with extensive Miombo and Mopane woodlands which are important for biodiversity conservation and the livelihoods of millions of people<sup>22,39</sup>. Previous analysis has concluded that mining has been a major driver of deforestation in the Copperbelt region of Zambia through both direct and indirect pathways<sup>13</sup>. These conclusions were however caveated by the difficulty in separating the effect of mining from other drivers<sup>13</sup>. We aim to build upon the work of Mwitwa and colleagues by applying a quasi-experimental methodology to isolate the effect of new large-scale mining developments from concurrent drivers of deforestation.

We identified when mining operations began in all currently issued large-scale mining leases to build a dataset of newly established mines from 2000 onwards (see Methods). We then analysed the extent of deforestation, estimated from the Hansen Global Forest Change dataset<sup>40</sup>, in new mining leases, and surrounding 25 km radius buffers. By including buffer areas, we aimed to capture secondary and indirect deforestation associated with mining that operate at a wider landscape scale. The buffer distance represented a conservative estimate of the range at which mines can affect nearby ecosystems<sup>6,11</sup>, and we tested the sensitivity of our results to this choice (See Methods). We used statistical matching methods<sup>41,42</sup>, with controls drawn from within large-scale exploration leases, to create a plausible counterfactual. We estimated the difference in deforestation between treated areas (i.e. leases and buffers) and matched control areas, for a period of five years after mine establishment, using statistical models to give an estimated average treatment effect. We analysed each mine individually, encompassing a range of different minerals and metals being extracted and mining processes. By analysing each mine individually our analysis can find any differences in forest lost due to different minerals and types of mining, such as a surface versus underground.

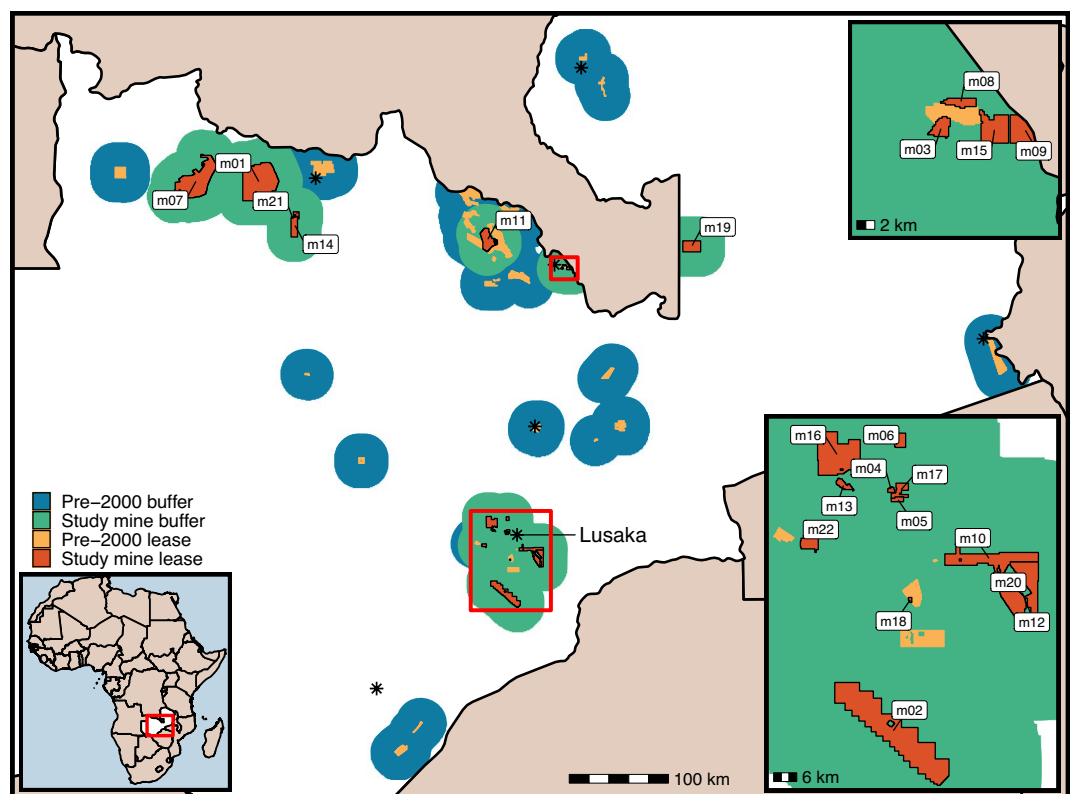
Our analysis looks at deforestation both within a mining lease and a buffer, so we measure both direct and indirect forest loss. This includes land cleared to extract minerals or store tailings and the secondary effects of activities such as new roads for access, processing facilities, influxes of new workers and related development activities. Deforestation is one of a wider range of possible impacts that mining can have and therefore our study should not be taken as a comprehensive assessment of the environmental impact of mining. We consider only large-scale mining leases, not small scale or unlicensed operations. Comparison to a counterfactual means we can estimate whether new large-scale mining developments are driving deforestation at the landscape scale compared to other drivers which operate at the similar scales such as agricultural expansion, charcoal production, or urbanisation<sup>13,21,43</sup>. Given the existing literature we hypothesised that mining developments would have substantially increased deforestation relative to a counterfactual of no mining.

## Results

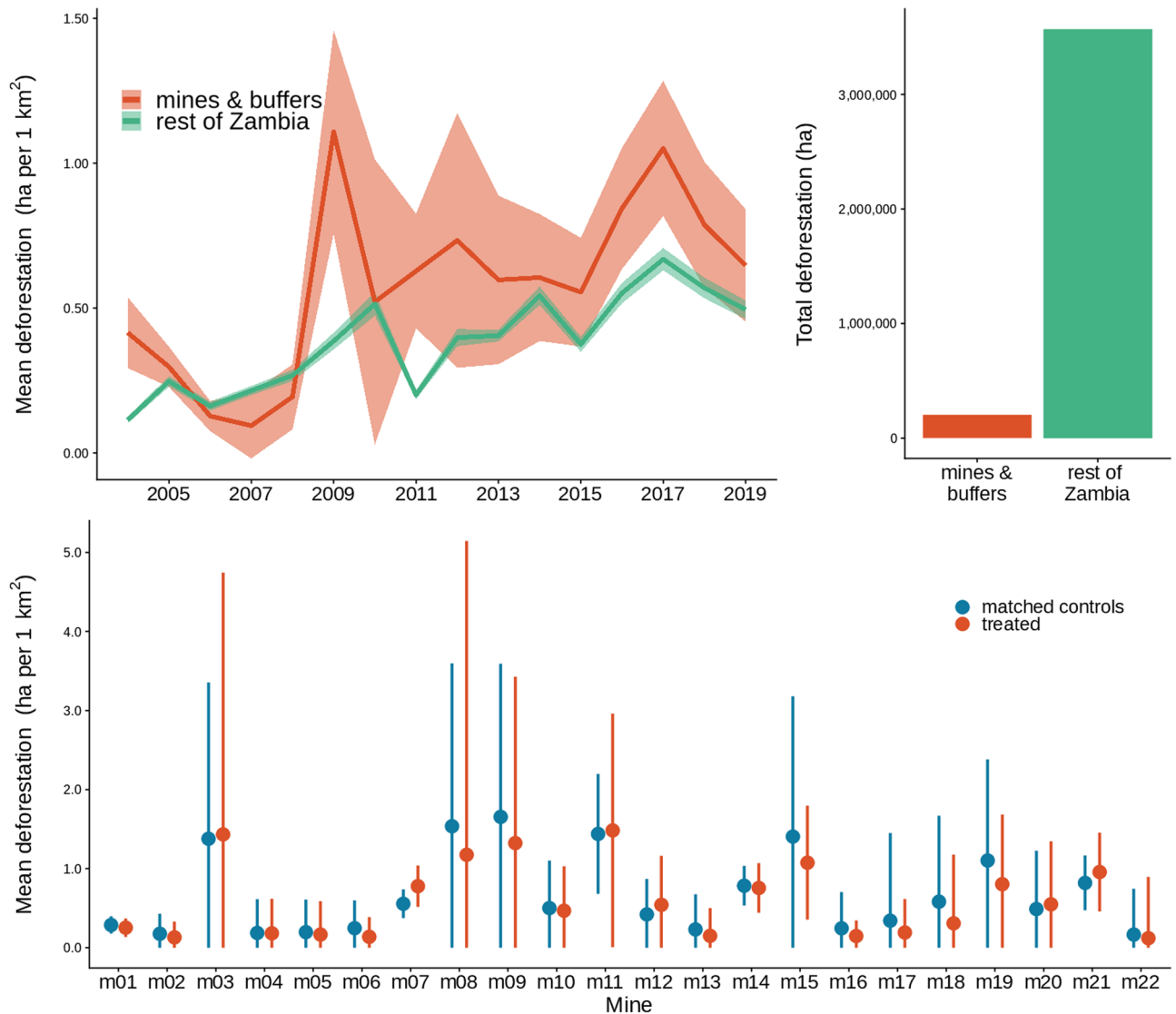
**Mines in Zambia.** In total 22 mining leases were established during our study period (2000–2019) with the first new mine during this period commencing operations in 2004 (Fig. 1). This represents 3422 km<sup>2</sup> of new land under a mining lease (Fig. 1), (Sup Mat Figure S1). As pre-existing leases covered an area of 2334 km<sup>2</sup>, the total area mined has more than doubled during the past 20 years (Sup Mat Figure S1). A further 77,868 km<sup>2</sup> of land is within the 25 km buffers (Fig. 1), (Sup Mat Figure S1). The new leases were for a range of minerals with limestone being the most common followed by copper, dolomite, cobalt, and gold (Sup Mat Table S2). All but a few of the new mines were open pit surface mining (Sup Mat Table S2). The growth in mined area reflects extensive growth in the amount of copper and gold mined within Zambia (Sup Mat Figure S2).

There is extensive deforestation in areas falling within new mining leases and 25 km buffers (Fig. 2). As forest pixels vary in canopy cover from 10 to 100% in Zambia, from sparse woodlands to dense forest, to ensure equivalence in terms of trees lost we used the area of tree canopy as a response variable. This means, effectively, that the clearance of a hectare with 100% canopy cover counts the same as ten hectares with 10% canopy cover. An estimated total of 201,143 ha of tree canopy area was lost, or 1,322,783 ha of total forest cover (treating all forest-covered pixels equally, regardless of canopy cover percentage, a method comparable with estimates produced by Global Forest Watch; see Methods), (Fig. 2), (Sup Mat Figure S3). There was also extensive deforestation in areas of Zambia further than 25 km from a mining lease, where a total of 3,571,373 ha of tree canopy area was cleared between 2004 and 2019 (25,679,539 ha of forest) (Fig. 2), (Sup Mat Figure S3). The rate of deforestation was on average slightly higher in new mining leases and their buffers when compared to the rest of Zambia (Fig. 2).

To estimate what the effect of mining has been on deforestation in Zambia, a plausible counterfactual (the extent of deforestation in the absence of a mine) was needed. To create a control dataset which represents this counterfactual, we used statistical matching (See Methods). A good quality match (where treatment and control are balanced in respect of key covariates) was achieved (Sup Mat Figure S4). Considering a period of five years after new mines were established, average rates of deforestation around our study mines are similar to those around matched controls (Fig. 2).



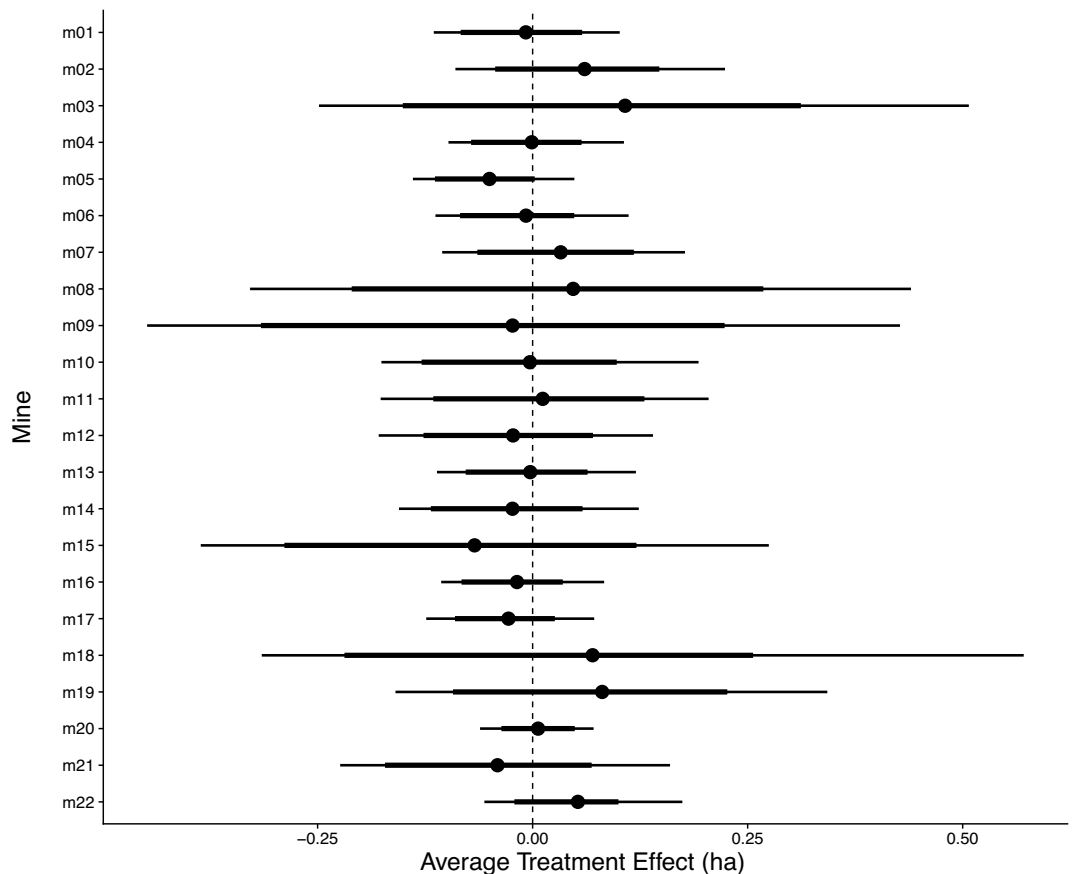
**Figure 1.** Map of large-scale mining leases in Zambia and 25 km radius buffers split by start date. Study mines leases (m01–m22) are all large-scale mining leases that started operations after 2000, the first being in 2004. Study Mine buffers are 25 km radius buffers drawn around these leases. Pre-2000 leases are large-scale mining leases which began operations before the year 2000, and pre-2000 buffers are 25 km radius buffers from these leases. Black stars are provincial capitals with the national capital Lusaka labelled. The deforestation impact of pre-2000 mines was not analysed, and the areas covered by their leases and buffers were excluded from consideration as possible controls. Where there is overlap the leases are shown, then study mine buffers, and then pre-2000 buffers. We analysed each study mine individually and considered any area falling within a lease or buffer as “treated”, ignoring overlapping buffers (i.e., treatment was coded as a binary variable). Some adjacent leases that were part of the same operation were merged. Full list of leases and details in (Sup Mat Table S1).



**Figure 2.** Top left panel shows the mean deforestation expressed as hectares lost per 1 km<sup>2</sup> pixel in Zambia estimated for all pixels which had greater than 10% tree coverage in 2000, comparing pixels within a mining lease that became active after 2000 and their 25 km buffers (red lines) and areas greater than 25 km from any mining lease (i.e., also excluding pre-existing mines and 25 km buffers; green lines). Lines are the mean and shaded areas are 1.96 times the standard error. Our deforestation metric is equivalent to tree canopy area lost for forested pixels. Top right panel shows the total deforestation in hectares for the period (2004–2019) for the same set of pixels. Bottom panel shows the mean deforestation within 1 km<sup>2</sup> pixels for five years after mine establishment for each mine (except mine 22 which was established only three years before 2019, the last year of deforestation data) comparing pixels within the mining lease and 25 km buffers (treated; red symbols) and matched controls (blue symbols). Points are the mean, lines are 1.96 times the standard error, with the lower bound limited to zero. These data were systematically sampled to 25% of the original data using a regular grid and filtered to only pixels which had greater than 10% tree coverage in 2000 prior to matching.

**The effect of mining on deforestation.** None of the 22 mines we analysed showed strong evidence of a treatment effect (i.e., a clear difference in tree canopy area lost between mine-affected areas and matched control areas) in the five years after mine establishment (Fig. 3). Although there were often wide credible intervals around effect estimates, the difference in deforestation for treated compared to matched control pixels was close to zero for most mines. No mine had 80% credible intervals that did not include zero (Fig. 3).

Although no overall effect was detected, the treatment effect for individual years indicates that there was elevated deforestation in treated pixels for some mines in some years. (Sup Mat Figure S5). Likely reflecting the initial large-scale clearing of vegetation for mining infrastructure, three mines (m01, m11, & m20) showed weak evidence of higher tree loss in treated pixels compared with matched control areas in the first year with 80%, but not 95%, credible intervals excluding zero (Sup Mat Figure S5). One of these mines (m20) also shows strong evidence of positive effect (i.e., greater deforestation in treated pixels) in year three, but this is balanced by strong evidence of negative effect (i.e., less deforestation in treated pixels) in years two, four, and five (Sup



**Figure 3.** Estimated average treatment effect of mine establishment across all 5 years for each mine (except mine 22 which was established only three years before 2019 the last year of deforestation data). The effect is the difference in the area of tree canopy lost in hectares between the treatment and control areas. Estimates are from spatiotemporal models with a zero-inflated negative binomial error structure implemented in a Bayesian modelling framework using R-INLA (see Methods). Points are estimates of the mean of the posterior distribution, thick lines are the 80% credible intervals (calculated as the highest posterior density interval) and thin lines the 95%.

Mat Figure S5). Annual estimates of loss from the Hansen dataset are, however, less reliable than the five-year averages as issues with cloud cover and seasonality can lead to loss being attributed to the wrong year<sup>44</sup>.

We produced results for five additional datasets as sensitivity tests, varying the type of matching approach used to create the control dataset and the buffer distance. This included (1) applying coarsened exact matching, (2) using the same set-up as the main result but allowing controls to be drawn from within and outwith exploration leases, (3) a no matching, random sampling approach, (4) applying the same method as the main result but with a 50 km buffer, and (5) the same method with a 10 km buffer. The results did not change substantially with these different methods. No mines showed evidence of higher deforestation for treated pixels compared to controls under any of the matching set-ups tested (Sup Mat Figure S6). One mine showed strong evidence of lower deforestation for treated pixels when coarsened exact matching was used, and three mines estimated lower deforestation for treated pixels when using a no matching random draw of controls (Sup Mat Figure S6). Using a buffer distance of 50 km or 10 km did not substantially alter our results (Sup Mat Figure S7). One mine with a 10 and one mine with a 50 km buffer showed weak evidence of higher deforestation for treated pixels with 80% credible intervals excluding zero (Sup Mat Figure S7). The majority of models showed no residual spatial autocorrelation (Sup Mat Figure S8). We also tested the ability of our response variable, based on the Hansen dataset, to predict deforestation measured by radar remote sensing<sup>45</sup>. We found a weak association with very little variation between treated and control pixels (Sup Mat Figure S9).

## Discussion

We found no evidence that the 22 mines we analysed had led to excess local deforestation compared to matched control sites in the period five years after they began operations. In the five years after a new mine is established, the average rate of deforestation is not any higher than plausible counterfactual scenarios, based on matching, where no mining takes place. This was true for all mines regardless of the mineral being extracted or the extraction method. Importantly this result does not mean that there is no deforestation due to mining activity. There is clearly extensive deforestation happening within mining leases and their 25 km buffers, much of which will

likely be due to direct mining activities and secondary processes, such as new roads, processing facilities, and influxes of workers, driven by the new mine being established. However, the extent of tree canopy area that is lost is no different than would likely have happened otherwise due to other drivers of deforestation in Zambia such as charcoal production, increased urbanisation, and conversion to agriculture<sup>13,21,43</sup>, all of which may be locally reduced due to mining activities. Zambia is a country with extensive deforestation<sup>21,39</sup> and therefore the most likely counterfactual scenario is one of high rates of forest loss. Comparing deforestation in areas impacted by mining to a high background rate, in areas which are similar, results in no overall increase in tree canopy area lost.

Our results differ from two previous studies of deforestation and mining that include Zambia. Previous analysis by Mwitwa and colleagues examined deforestation in the Copperbelt region of Zambia within 50 km and 200 km of mining developments combining this with interviews of key stakeholders<sup>13</sup>. They found extensive deforestation within mining developments and the buffer areas and identified a number of pathways, operating at multiple spatial scales, through which mining increased deforestation<sup>13</sup>. Their study did not consider counterfactuals and Mwitwa and colleagues themselves note that their methodology could not separate out deforestation due to mining from other co-occurring processes such as urbanisation and the associated increased resource demands<sup>13</sup>. Our descriptive results are in line with their findings but once we estimate the difference from a plausible counterfactual, we do not find any evidence that that mining has driven increased deforestation and without individual mines being present much, if not all, of the deforestation would have likely taken place regardless. Another recent pan-tropical study found a strong negative association between distance to a mine and total deforestation 2000–2019 in Zambia<sup>35</sup>. Our results likely differ because they use proximity to any mine which is a measure of a cumulative effect of all mining as opposed to individual new mines<sup>35</sup>. Additionally the majority of mines included in our analysis are missing from their dataset<sup>35</sup>. It is also important to note that they do not account for the spatial dependence within their data nor test for spatial autocorrelation<sup>35</sup> and therefore the effect they find may be inflated due to pseudo-replication.

Our results are likely to be context dependent and may not be generalizable to other regions where extensive mining takes place. In other regions mining has been found to be a driver of deforestation through secondary processes, such as new linear infrastructure, population shifts, and increases in illegal mining and logging<sup>46</sup>. We do not, however, find any evidence that secondary processes have caused excess deforestation. This may be because many of the new mines are surrounded by pre-existing mines, urban areas, and other intensive use land cover types. The effect of mining in these instances is likely very different compared to regions which are far from population centres with dense forest and low human population densities, as occurs in, for example, the Brazilian<sup>32</sup> or Peruvian<sup>25</sup> Amazon. The largest mine we examined, the Trident/Sentinel mine (Mine 7), was not near a major population centre and involved the building of several new roads, whole villages with houses, schools, and hospitals, and an airport. However this mine is unusual in that these secondary infrastructures were carefully planned and funded by the mining company along with the West Lunga Conservation Project<sup>47,48</sup>. These management, mitigation, and compensation measures likely reduced deforestation in the proximity of the mine, although it is important to note that nearby communities were still negatively impacted and measures which may reduce deforestation, such as restricting access to nearby forests, can have detrimental effects on food security<sup>48</sup>. Given these complexities we suggest that caution should be taken when generalizing studies such as this to other regions where patterns of historical land use and governance structures are different. Repetition of our approach or similar approaches that consider counterfactual rates of tree cover loss across multiple countries would contribute to our understanding of the impacts of mining on natural habitats.

Our study builds on previous studies from Zambia, but it is essential to stress that it focuses exclusively on deforestation. Mining developments can have a wide range of impacts on socio-ecological systems<sup>3,4</sup>. For example, contamination of water and soil with toxic chemicals is a significant problem in mining regions<sup>49–51</sup>. Tailings dams which store toxic by-products of mineral extraction processes in large reservoirs are a particular risk as they can fail causing pollution of soils and waterways which impacts land at some distance from the mine itself<sup>52</sup>. Mines can also impact socio-ecological systems, causing displacement of nearby communities and limiting their access to resources<sup>53</sup>. Our analysis cannot say anything about what effect the 22 newly established mines had on these outcomes. Therefore, we stress that any decisions about the issuing of new mining leases or other relevant policy decisions need to consider the full range of possible impacts.

Our analysis is limited in its ability to separate out the non-mine-specific effects and by the single level of treatment. Our methodology attempts to isolate the effect that individual mines have had on deforestation. However, it is not possible to completely separate out the broader, indirect effects that mining may have on other drivers of deforestation from the effect of specific mines being present or absent. For example, by contributing to the Zambian economy, mining is interconnected with processes such as increased urbanisation and higher resource requirements in certain regions. Additionally, we only considered a binary indicator of treatment, meaning we were not able to explore whether there were either additive or plateauing effects from a cell falling within the buffer areas of more than one mine. However, if either effect was substantial then our results would differ based on whether the mines were near other mines or not and we see no such pattern in our results.

Additional possible limitations, tested with sensitivity analysis, include the choice of buffer distance, the matching set up, and the use of the Hansen dataset as our response variable. The choice of 25 km buffers may be an underestimate of the range at which mining can impact nearby ecosystems. Other studies have assumed impacts to occur at much greater distances, for example within 50 km<sup>6,11</sup>, 100 km (within which significant effects were found up to 70 km)<sup>32</sup>, and 200 km<sup>13</sup>. However, the desire to capture the full extent of mine impacts must be balanced against the increased likelihood that deforestation is caused by other drivers the greater the distance from the mine, especially given Zambia's mosaic of different land cover types. We therefore chose 25 km as a compromise and neither our overall result nor the majority of results for each mine changed when using either a 10 km or 50 km buffer. Subjective choices in application of statistical matching have been shown to potentially affect results in impact evaluations<sup>54</sup>. Our main results are almost completely unchanged when using

coarsened exact matching, propensity score matching without the requirement that controls are drawn from exploration leases (our most restrictive condition), or a simple random draw of non-mine cells as controls. The random controls dataset showed some mines to have a negative treatment effect meaning that deforestation in these mining leases and their buffers, although similar to that under the counterfactual, was in fact lower when compared to random points in Zambia.

Our final sensitivity analysis examined whether our results were affected by the choice of forest cover data. The Hansen Forest Change dataset, which is derived from Landsat imagery, has been shown to underestimate the extent of forest loss in Zambia when compared with measures based on biomass estimated from radar remote sensing<sup>45</sup>. This is mainly due to its inability to see small scale losses and its sensitivity to seasonal variation in vegetation<sup>45</sup>. As the Hansen dataset underestimates the total change, it may also underestimate differences between mine impacted areas and our matched controls. Testing this, we found a weak but significant correlation between our response variable and an alternative data set based on radar remote sensing that covers a subset of our study period (2007–2010)<sup>45</sup>. There was, however, only a very small difference between treated and control pixels, with treated pixels having slightly more agreement between the two datasets. Therefore, it is unlikely that the use of the Hansen Forest Change data has substantially affected our overall result.

To conclude, our study finds no detectable effect of mines driving deforestation when compared to a matched counterfactual of no mining. This result could indicate that assessments of the deforestation impact of mining, which utilise purely descriptive or correlative methodologies, may overestimate the effect that mining has had. Statistical matching has been widely applied in other areas such as the effectiveness of protected areas<sup>41</sup> but rarely used for measuring mining impacts. Our results do not mean that mining has not led to deforestation but rather we found no evidence that it was different from what would likely have happened without a mine being present. Nor does it negate the threat present to nearby ecosystems from mining, given the range of other impacts possible. Large areas of Zambia are increasingly influenced by mining activity. The extensive growth in mined land in Zambia reflects a wider trend in sub-Saharan Africa where 58% of current mines were created between 2000 and 2018, driven by growing global demand for minerals and metals<sup>35</sup>. This highlights the connection between land use in mineral rich nations such as Zambia and demand for natural resources globally<sup>27,55,56</sup>. To assess the full impact of these new mines, and other development activities, it is important that evaluations of local impact implement robust causal inference methodologies that combine statistical matching with post-match modelling that accounts for the spatial temporal nature of the data, and that they are complemented by work drawing on a broader range of quantitative and qualitative impact evaluation approaches capable of capturing the indirect effects of mining.

## Materials and methods

**Mines in Zambia.** To measure the effect that new mining developments have had on deforestation, we examined 1 km<sup>2</sup> pixels within Large-scale Mining Leases (LMLs) that become active from the year 2000 onwards and a 25 km radius buffer. New mining developments in Zambia are regulated through the issuance of mining leases to companies. In the first instance, exploration leases are granted that allow companies to take samples in order to estimate the presence and quantities of the resource present<sup>38</sup>. If minerals and metals are found in sufficient quantities to be economically viable, companies then apply for a full exploitation lease which are differentiated based on the size of the mining operation<sup>38</sup>. We downloaded all Large-scale Mining Leases from the Zambian government's Mining Cadastre Map Portal (<https://portals.landfolio.com/zambia/>, accessed February 2020).

We assigned leases an 'active from' year through a multiple stage process. The date taken from the 'Date Granted' field in the lease dataset was first verified by visual inspection of high-resolution imagery using the time-lapse function in Google Earth pro. We looked for any evidence of the building of infrastructure or the clearing of land. If there was evidence of infrastructure being built or land cleared within two years of the granted date, this year was taken as the start date (Sup Mat Table S1). If there was no evidence to confirm this date, an internet search for industry news articles, company reports, and other documents was made. Any date suggested by these other sources was then confirmed by inspection of Google Earth imagery. If there was no evidence of activity or confirmation that the mine had not gone ahead, then the lease was marked as never active (Sup Mat Table S1). When there was uncertainty due to missing Google Earth imagery at important dates then, where possible, RGB cloud-free composites of Sentinel 2 images were inspected using Google Earth Engine. If after applying all steps the status of the mine was unclear, then it was not included in analysis and taken to be active before 2000 for the purpose of selecting controls. If multiple adjacent leases were owned by the same company and there was evidence of activity around the same time these were merged and treated as one (Sup Mat Table S1). There was a large difference for some mines between the granted date and when we found evidence of activity on site (Sup Mat Table S1).

We analysed individually all 22 mines that were found to be active after 2000, creating a dataset of treated and control units for each. Each lease was first buffered by 25 km then converted into a raster with 1 km<sup>2</sup> pixels such that any cell touching the buffered lease was considered 'treated'. We ignored cases where buffers overlap the buffers and leases of other mines. This means that in some instances pixels we considered 'treated' would have been previously influenced by mining by virtue of being within an existing mines buffer (see Fig. 1). All mines were treated the same regardless of the minerals being extracted or the methods of extraction, e.g., open pit versus underground shaft. Different types of mines will likely have different impacts on deforestation as they, for example, require different amounts of land to be cleared, mines of different sizes will have differing secondary impacts, and the mineral extracted will determine the amount of processing infrastructure required. By estimating the effects of each mine individually we allow for this variation.

To offer a reasonable counterfactual we considered controls drawn from Large-Scale Exploration Leases (LELs), taken from the government portal, as these are areas of land that could become treated (i.e., be converted in an active mine). All Large-Scale Mining Leases, both those included in our analysis and pre-2000 leases, and Small-scale Mining Leases buffered by 5 km, were clipped from the LELs polygons, to exclude them as potential controls. The resulting polygons were then converted to the 1 km<sup>2</sup> raster with any touching pixel considered a possible control. To reduce the dataset size, and the possible effect of spatial autocorrelation, all treatment and control rasters were then subsampled to 25% of the total in a regular pattern with only the bottom right of a 4 by 4-pixel square included. These rasters were then converted to data frames of treated units and possible controls, one per mine. The datasets were further filtered to remove pixels with an average tree cover below 10%. To reduce the distance between treatment and controls the datasets were also restricted to just treatments and controls within the same province (the largest administrative unit in Zambia), determined by the province the majority of the treated units were in<sup>57</sup>. Due to its small size Lusaka province was merged with the neighbouring Central province. The number of treated units varied greatly depending on the size of lease, ranging from 220 to 1654 (Sup Mat Table S3). Possible controls varied based on province, ranging from 1835 to 15,979 (Sup Mat Table S3). Datasets where possible controls were drawn both within and outwith exploration leases and with a buffer distance of 50 km and 10 km were also created for sensitivity analysis (Sup Mat Table S3).

**Deforestation in Zambia.** Our outcome variable of interest is deforestation expressed as the yearly area of tree canopy area lost in metres squared. Area lost and average tree cover were estimated for each 1 km<sup>2</sup> pixel for each year from the Hansen Global Forest Change dataset using Google Earth Engine<sup>40</sup>. Yearly tree cover was calculated by converting the ‘loss year’ layer to a band per year (2001–2019), with 1 assigned to every pixel that had forest in the year 2000, with these 1s subsequently changing to 0s in a year where there was a loss event in the ‘loss year’ layer. The Hansen dataset also has a ‘tree cover’ layer, giving the percentage tree cover of each pixel in the year 2000. Zambian forests range from very sparse savannas with as little as 10% canopy cover, through to dense forests of 100% canopy cover. The latter clearly has more trees, carbon, and likely biodiversity, so we chose to weight it more highly than forest loss of low-canopy-cover forest. These yearly bands were thus multiplied by the tree cover in 2000 layer to create a multiband image with tree cover each year at the 30 m × 30 m resolution. We then converted the yearly tree covers to a 0.5 ha resolution and areas below 10% coverage at 0.5 ha were masked out to meet the Zambian government’s definition of a forest<sup>58</sup>. Area lost was calculated by splitting the loss year layer into yearly bands, assigning the value 1 if there was loss and then converting to a 0.5 ha resolution by taking the weighted mean, giving the proportion of a 0.5 ha pixel lost each year. The 0.5 ha forest/non-forest mask for the year 2000 was then applied and the result weighted by the yearly tree cover at 0.5 ha. The resulting layers were then aggregated to the 1 km<sup>2</sup> resolution using the mean, weighted by the extent of overlap, giving yearly tree cover and yearly tree canopy area lost for forested areas in 1 km<sup>2</sup> pixels. The standard deviation of tree cover within a 1 km<sup>2</sup> pixels (with non-forested areas masked out) was also calculated.

By weighting our area lost estimates by the tree cover in 2000 layer we create a dataset of tree canopy area lost in metres squared each year. This differs from the methodology used by Global Forest Watch<sup>59</sup> and often reported by other organisations, e.g.,<sup>21,60</sup>. In these sources the area lost is not weighted by the extent of tree cover, meaning that a pixel with 30% tree cover being lost is treated as being the same as one with 70% tree cover. To make our results comparable with these sources we also created an unweighted dataset (which is equivalent to ‘forest area lost’, with forest defined as areas with at least 10% canopy cover over 0.5 ha) and include the relevant descriptive results in the supplementary material. One weakness of our approach is that it assumes that the tree cover in 2000 layer is a good measure of canopy cover in later years. As our method removes only pixels which appear in the loss year layer it will not capture changes in canopy cover due to regrowth and degradation not picked up as a loss by the Hansen dataset. Any changes in canopy cover due to degradation should eventually show up in the response once a threshold has been passed and regrowth will take several years to substantially impact tree canopy area at a 1 km<sup>2</sup> resolution. Although our canopy area estimates may be inaccurate in later years it is reasonable to assume that the patterns of high canopy and low canopy areas will be similar, and any inaccuracies will not be biased to either treated or control pixels. Examples for two mines showing the total deforestation in the mining leases and their buffers can be seen in the supplementary material (Sup Mat Fig. 10).

The Hansen Forest Change dataset has previously been shown to underestimate forest loss in miombo woodlands<sup>45</sup>. To test the sensitivity of our results to this limitation we compared our deforestation measure, for pixels included in the final analysis, to estimates derived from L-Band Radar satellites. Only a limited number of years of data (2007–2010) were available<sup>45</sup>. We applied the same methodology to get the weighted mean of the area deforested for 1 km<sup>2</sup> pixels. We compared total area lost in the two datasets for the period 2007–2010 in a linear model allowing the relationship to vary for treated and control pixels (Sup Mat Fig. 9).

**Statistical matching methods.** Quantitative impact evaluations estimate the effect a treatment or intervention (e.g., a mine being established) has on an outcome (deforestation) through comparison to a counterfactual, i.e., what would have been the outcome without the treatment/intervention<sup>61</sup>. This counterfactual does not exist and can only ever be inferred. Statistical matching is a pre-processing methodology to create a dataset of controls which represent a plausible counterfactual<sup>62</sup>. These controls are selected from a wider group of possible controls to be balanced in respect of key covariates to remove any bias that exists due to the non-random treatment assignment<sup>41</sup>. For example, mines are more likely to be established if they are near existing road networks and proximity to roads is a known predictor of deforestation. Therefore, to be plausible counterfactuals controls need to be drawn from similar distances to a road.

We assumed that deforestation was a function of the extent of forest present, ease of access, nearby populations, previous disturbance, and suitability of other land uses<sup>32,63</sup>. We assumed mine establishment was



determined by the opportunity cost of other land uses and the ease of access. The covariates used to represent these drivers were, tree cover<sup>40</sup>, distance to the nearest road<sup>64,65</sup>, elevation and slope<sup>66</sup>, population density<sup>67</sup>, evidence of recent burning<sup>68</sup>, agro-ecological zone<sup>69</sup>, and whether a pixel was within a protected area<sup>70</sup>. Following best practice, we also included a measure of pre-treatment outcome, i.e., whether there had been any deforestation in the previous two years. Details of the variables and their sources are available in the supplementary material (Sup Mat Table S4). The quantity of minerals present is obviously another key driver of whether an exploration lease is converted into a full mining lease, however, as these data are not publicly available, it was not possible to include it in our analysis. As the quantities of minerals present would not impact deforestation in the absence of mining this should not substantially bias our estimates of the treatment effect.

Our primary analysis was based on a dataset generated using nearest neighbour propensity score matching (PSM). This involves estimating the probability of assignment (the propensity score) based on a logistic regression which includes the matching confounders. Treated units are then matched to the control with the nearest propensity score value. We used 1–1 matching meaning that for each treated unit a single control was included and matched without replacement so controls can only be selected once. We applied a caliper distance, so controls were only matched to the treated unit if their propensity scores are within 0.25 standard deviations of each other<sup>42</sup>. If no control is within this distance, then the treated unit is dropped from the dataset. We also applied a common support restriction dropping both treatment and controls units with fall outside the area of overlapping propensity scores<sup>42</sup>. We used exact matching for the categorical variables of the protected area and agroecological zone, meaning controls were only selected if these values were the same as the treated unit. Each mine was matched individually using confounders from the relevant year. As the matching algorithm set up has been shown to alter the overall results in impact evaluations<sup>54</sup> we created three further datasets for sensitivity testing based on different matching approaches. The first used coarsened exact matching in which acceptable ranges of values are chosen for each confounder to group the data and then controls matched to treated units based on membership of the same groups for each confounder<sup>71,72</sup>. As many controls as match one or more treated units are included. The second used a similar matching set up as our main analysis but allowed controls to be drawn from both within and outwith exploration leases, relaxing our most restrictive matching condition. Quality of all matches was determined by examining standard mean difference plots (Sup Mat Figure S4). Our third dataset was unmatched. Instead, a random draw of controls equal to the number of treated units was selected from anywhere in Zambia (excluding all mine leases and 25 km radius buffers). After matching, the dataset size ranged between 426 and 3,306 for the main dataset and generally larger for other matching set ups (Sup Mat Table S1).

**Outcome analysis.** To ensure rigorous analysis of the difference in deforestation between treated and control units, while accounting for the spatial and temporal structure of the data, we applied statistical modelling after matching. We utilised hierarchical spatiotemporal Bayesian models in which the posterior distribution is approximated through an integrated nested Laplace approximation (INLA) using the ‘R-INLA’ package in R statistical software<sup>73,74</sup>. The INLA methodology approximates the marginal posterior distributions of model parameters and is several orders of magnitude more computationally efficient compared to approaches that use Markov Chain Monte Carlo algorithms<sup>75</sup>. This computational efficiency makes it an effective means of implementing complex spatiotemporal models that would otherwise not be feasible. We utilised continuous spatial models where a spatial effect follows a Matérn covariance that is based on the solution to a stochastic partial differential equation (SPDE)<sup>75</sup>.

We used a spatiotemporal extension of the SPDE models to analyse deforestation each year for five years after mine establishment (except mine 22 which only had three years of data). We used an ‘exchangeable’ temporal structure in our models<sup>76</sup>. In these models a spatial effect is estimated for each year with values drawn from the same distribution as other years but not necessarily correlated with the previous year<sup>76</sup>. We included an interaction between the year and treatment effects as a fixed effect so that the treatment effect (i.e. the difference between treatment and control) is estimated for each year, and included an independent and identically distributed Gaussian random effect for the pixel ID<sup>77</sup>. Deforestation was rounded to the nearest whole metre square and a zero-inflated negative binomial regression used. Covariates were selected by comparing deviance information criterion values for candidate models based on the matching confounders after removing any which were collinear. Final covariates were tree cover, the standard deviation of tree cover, slope, and whether a pixel was in a protected area or not (only used for relevant mines where some pixels were within a protected area).

We used INLA’s default priors for all parameters except the hyperparameters of the spatial random field. For the spatial element we used the penalised complexity priors<sup>76</sup> to control the extent of spatial smoothing with the range set to have a 95% probability of being greater than 50 km. For sigma, the prior varied between mines to ensure model convergence. Model fit was checked through posterior predictive checks and plots of simulated quantile residuals created with the ‘DHARMA’ package<sup>78</sup>. All direct neighbours were excluded by subsample, and we tested for residual autocorrelation at greater distances through the use of variograms. Reported results are the estimated difference in predicted loss between treated and control units for each mine, either for each year or aggregated across all years. These were created by sampling 4000 times, a number chosen to balance computing time and accuracy, from the posterior distribution using INLA’s inbuilt function.

### Data availability

All code used in the analysis can be found here ([https://github.com/joffy2/zam\\_mines\\_public.git](https://github.com/joffy2/zam_mines_public.git)) all the datasets used in analysis are publicly available, details and links are available in (Sup Mat Table S4).

Received: 13 June 2022; Accepted: 19 October 2022

Published online: 29 October 2022

## References

- Prior, T., Giurco, D., Mudd, G., Mason, L. & Behrisch, J. Resource depletion, peak minerals and the implications for sustainable resource management. *Glob. Environ. Chang.* **22**, 577–587 (2012).
- International Council on Mining and Metals. Role of mining in national economies Mining Contribution Index (MCI) 5th Edition. (2020).
- Sonter, L. J., Ali, S. H. & Watson, J. E. M. Mining and biodiversity: Key issues and research needs in conservation science. *Proc. R. Soc. B: Biol. Sci.* **285**, 20181926 (2018).
- Werner, T. T., Bebbington, A. & Gregory, G. Assessing impacts of mining: Recent contributions from GIS and remote sensing. *Extr. Ind. Soc.* **6**, 993–1012 (2019).
- Murguía, D. I., Bringezu, S. & Schaldach, R. Global direct pressures on biodiversity by large-scale metal mining: Spatial distribution and implications for conservation. *J. Environ. Manag.* **180**, 409–420 (2016).
- 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**, 6–11 (2020).
- González-González, A., Clerici, N. & Quesada, B. Growing mining contribution to Colombian deforestation. *Environmental Research Letters* **16**(6), 064046 (2021).
- Bebbington, A. J. *et al.* Resource extraction and infrastructure threaten forest cover and community rights. *Proc. Natl. Acad. Sci.* **115**, 13164–13173 (2018).
- Edwards, D. P. *et al.* Mining and the African environment. *Conserv. Lett.* **7**, 302–311 (2014).
- Werner, T. T. *et al.* Global-scale remote sensing of mine areas and analysis of factors explaining their extent. *Glob. Environ. Chang.* **60**, 102007 (2020).
- World Bank. Forest-Smart Mining: Large-scale mining on forests (LSM). <http://hdl.handle.net/10986/32025> (2019).
- Siqueira-Gay, J., Santos, D., Nascimento, W. R., Souza-Filho, P. W. M. & Sánchez, L. E. Investigating changes driving cumulative impacts on native vegetation in mining regions in the Northeastern Brazilian Amazon. *Environ. Manag.* **69**, 438–448 (2022).
- Mwitwa, J., German, L., Muimba-Kankolongo, A. & Puntodewo, A. Governance and sustainability challenges in landscapes shaped by mining: Mining-forestry linkages and impacts in the copper belt of Zambia and the DR Congo. *For. Policy Econ.* **25**, 19–30 (2012).
- Johnson, C. J., Venter, O., Ray, J. C. & Watson, J. E. M. Growth-inducing infrastructure represents transformative yet ignored keystone environmental decisions. *Conserv. Lett.* **13**, 1–7 (2020).
- Laurance, W. F., Sloan, S., Weng, L. & Sayer, J. A. Estimating the environmental costs of Africa's massive "development corridors". *Curr. Biol.* **25**, 3202–3208 (2015).
- Golden Kroner, R. E. *et al.* The uncertain future of protected lands and waters. *Science* **364**, 881–886 (2019).
- Devenish, K., Desbureaux, S., Willcock, S. & Jones, J. P. G. On track to achieve no net loss of forest at Madagascar's biggest mine. *Nat. Sustain.* <https://doi.org/10.1038/s41893-022-00850-7> (2022).
- Betts, M. G. *et al.* Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature* **547**, 441 (2017).
- Haddad, N. M. *et al.* Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* **1**, 1–9 (2015).
- Achard, F. *et al.* Determination of tropical deforestation rates and related carbon losses from 1990 to 2010. *Glob. Change Biol.* **20**, 2540–2554 (2014).
- Day, M., Gumbo, D., Moombe, K. B., Wijaya, A., Sunderland, T. Zambia country profile: Monitoring, reporting and verification for REDD+. (2014).
- Syampungani, S., Chirwa, P. W., Akinnifesi, F. K., Sileshi, G. & Ajayi, O. C. The Miombo woodlands at the cross roads: Potential threats, sustainable livelihoods, policy gaps and challenges. *Nat. Res. Forum* **33**, 150–159 (2009).
- House, C. The impact of mining on Forests: information needs for effective policy responses. *Energy Environ. Resour. Meet. Summ.* **44**, 1–10 (2015).
- Nicolau, A. P., Herndon, K., Flores-Anderson, A. & Griffin, R. A spatial pattern analysis of forest loss in the Madre de Dios region Peru. *Environ. Res. Lett.* **14**(12), 124045 (2019).
- Asner, G. P., Llactayo, W., Tupayachi, R. & Luna, E. R. Elevated rates of gold mining in the Amazon revealed through high-resolution monitoring. *Proc. Natl. Acad. Sci. USA.* **110**, 18454–18459 (2013).
- Swenson, J. J., Carter, C. E., Domec, J. C. & Delgado, C. I. Gold mining in the Peruvian Amazon: Global prices, deforestation, and mercury imports. *PLoS ONE* **6**(4), e18875 (2011).
- Espejo, J. C. *et al.* Deforestation and forest degradation due to gold mining in the Peruvian Amazon: A 34-year perspective. *Remote Sens.* **10**, 1–17 (2018).
- Barenblitt, A. *et al.* The large footprint of small-scale artisanal gold mining in Ghana. *Sci. Total Environ.* **781**, 146644 (2021).
- Gbedzi, D. D. *et al.* Impact of mining on land use land cover change and water quality in the Asutifi North District of Ghana West Africa. *Environ. Chall.* **6**, 100441 (2022).
- Obodai, J., Adjei, K. A., Odai, S. N. & Lumor, M. Land use/land cover dynamics using landsat data in a gold mining basin-the Ankobra, Ghana. *Remote Sens. Appl. Soc. Environ.* **13**, 247–256 (2019).
- Schueler, V., Kuemmerle, T. & Schröder, H. Impacts of surface gold mining on land use systems in Western Ghana. *Ambio* **40**, 528–539 (2011).
- Sonter, L. J. *et al.* Mining drives extensive deforestation in the Brazilian Amazon. *Nat. Commun.* **8**, 1013 (2017).
- Butsic, V., Baumann, M., Shortland, A., Walker, S. & Kuemmerle, T. Conservation and conflict in the democratic republic of Congo: The impacts of warfare, mining, and protected areas on deforestation. *Biol. Cons.* **191**, 266–273 (2015).
- Ranjan, R. Assessing the impact of mining on deforestation in India. *Resour. Policy* **60**, 23–35 (2019).
- Giljum, S. *et al.* A pantropical assessment of deforestation caused by industrial mining. *Proc. Natl. Acad. Sci.* **119**, e2118273119 (2022).
- Bini, L. M. *et al.* Coefficient shifts in geographical ecology: An empirical evaluation of spatial and non-spatial regression. *Ecography* **32**(2), 193–204. <https://doi.org/10.1111/j.1600-0587.2009.05717.x> (2009).
- Negret, P. J. *et al.* Effects of spatial autocorrelation and sampling design on estimates of protected area effectiveness. *Conservation Biology*, **34**, 1452–1462. <https://doi.org/10.1111/cobi.13522> (2020).
- Zambia EITI Council. Zambia Extractive Industries Transparency Initiative Zambia Report. (2019).
- Pelletier, J. *et al.* Carbon sink despite large deforestation in African tropical dry forests (miombo woodlands). *Environ. Res. Lett.* **13**(3), 09401 (2018).
- Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
- Schleicher, J. *et al.* Statistical matching for conservation science. *Conserv. Biol.* **00**, 1–12 (2019).
- Ho, D. E., Imai, K., King, G. & Stuart, E. A. MatchIt: Nonparametric preprocessing for parametric causal inference. *J. Stat. Softw.* **42**, 1–28 (2011).
- Handavu, F., Chirwa, P. W. C. & Syampungani, S. Socio-economic factors influencing land-use and land-cover changes in the miombo woodlands of the Copperbelt province in Zambia. *For. Policy Econ.* **100**, 75–94 (2019).
- Mitchard, E., Viergever, K., Morel, V., Tipper, R. Assessment of the accuracy of University of Maryland (Hansen et al.) Forest Loss Data in 2 ICF project areas—component of a project that tested an ICF indicator methodology. (2015).
- McNicol, I. M., Ryan, C. M., Mitchard, E. T. A. Carbon losses from deforestation and widespread degradation offset by extensive growth in African woodlands. *Nat. Commun.* **9**, (2018).

46. Sonter, L. J., Moran, C. J., Barrett, D. J. & Soares-Filho, B. S. Processes of land use change in mining regions. *J. Clean. Prod.* **84**, 494–501 (2014).
47. West Lunga Conservation Project. West Lunga Conservation Project. <https://www.westlunga.org/ourwork> (2021).
48. Swedwatch. Copper with a cost. [https://swedwatch.org/wp-content/uploads/2019/05/Copper-with-a-Cost-94\\_Zambia\\_191210.pdf](https://swedwatch.org/wp-content/uploads/2019/05/Copper-with-a-Cost-94_Zambia_191210.pdf) (2019).
49. Muhire, I. *et al.* The environmental impacts of mining on Gishwati Protected Reserve in Rwanda. *Environ. Monit. Assess.* **193**(9), 1–24 (2021).
50. Chanda, S. & Moono, W. The Effect of Mining on the Environment : A case Study of Kankoyo Township of Mufulira District of the Republic of Zambia (2017).
51. Muma, D., Besa, B., Manchisi, J. & Banda, W. Effects of mining operations on air and water quality in Mufulira district of Zambia: A case study of Kankoyo Township. *J. South Afr. Inst. Min. Metall.* **120**, 287–298 (2020).
52. Islam, K. & Murakami, S. Global-scale impact analysis of mine tailings dam failures: 1915–2020. *Glob. Environ. Chang.* **70**, 102361 (2021).
53. Takam Tiamgne, X., Kalaba, F. K. & Nyirenda, V. R. Household livelihood vulnerability to mining in Zambia's solwezi copper mining district. *Extr. Ind. Soc.* <https://doi.org/10.1016/j.exis.2021.101032> (2021).
54. Desbureaux, S. Subjective modeling choices and the robustness of impact evaluations in conservation science. *Conserv. Biol.* **35**(5), 1–12 (2021).
55. Ahmed, A. I., Bryant, R. G. & Edwards, D. P. Where are mines located in sub Saharan Africa and how have they expanded overtime?. *Land Degrad. Dev.* **32**, 112–122 (2021).
56. Nakajima, K. *et al.* Global distribution of material consumption: Nickel, copper, and iron. *Resour. Conserv. Recycl.* **133**, 369–374 (2018).
57. GADM. GADM database of Global Administrative Areas version 3.4. [www.gadm.org](http://www.gadm.org) (2018).
58. Government of Zambia. The Forests Act, 2015. (2015).
59. Global Forest Watch. Global Forest Watch. <https://www.globalforestwatch.org/> (2020).
60. Bond, L., Chambwera, M., Jones, B., Chundama, M. & Nhantumbo, I. REDD+in dryland forests Issues and prospects for pro-poor REDD in the miombo woodlands of southern Africa. (2010).
61. Ferraro, P. J., Sanchirico, J. N. & Smith, M. D. Causal inference in coupled human and natural systems. *Proc. Natl. Acad. Sci. USA.* **116**, 5311–5318 (2019).
62. Stuart, E. A. Matching methods for causal inference: A review and a look forward. *Stat. Sci.* **25**, 1–21 (2010).
63. Vinya, R., Syampungani, S., Kasumu, E. C. C., Monde, C., Kasubika, R. *Preliminary study on the drivers of deforestation and potential for REDD+ in Zambia. A consultancy report prepared for Forestry Department and FAO under the national UN-REDD+ Programme Ministry of Lands & Natural Resources* (2012) doi:<https://doi.org/10.1128/AAC.03728-14>.
64. National Imagery and Mapping Agency. Vector map level 0 (VMAP0). (2000).
65. Center for International Earth Science Information Network (CIESIN), Columbia University and Information Technology Outreach Services (ITOS) & University of Georgia. Global Roads Open Access Data Set, *Version 1 (gROADSv1)*. (2013).
66. Jarvis, A., Reuter, H. I., Nelson, A., Guevara, E. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database. (2008).
67. WorldPop ([www.worldpop.org](http://www.worldpop.org)) - School of Geography and Environmental Science - University of Southampton; Department of Geography and Geosciences - University of Louisville; Departement de Géographie-Université de Namur) and Center for International Earth Science Information Network (CIESIN) Columbia University. Global High Resolution Population Denominators Project-Funded by The Bill and Melinda Gates Foundation. doi:<https://doi.org/10.5258/SOTON/WP00675> (2018).
68. Giglio, L., Justice, C., Boschetti, L. & Roy, D. MCD64A1 MODIS/Terra+Aqua Burned Area Monthly L3 Global 500m SIN Grid V006. NASA EOSDIS Land Process. DAAC <https://doi.org/10.5067/MODIS/MCD64A1.006> (2015).
69. HarvestChoice, I. F. P. R. I. (IFPRI); Minnesota, U. of. CELL5M: A Multidisciplinary Geospatial Database for Africa South of the Sahara. DRAFT VERS. Harvard Dataverse. <https://doi.org/10.7910/DVN/G4TBLF>.
70. IUCN and UNEP-WCMC. The World Database on Protected Areas (WDPA). November 2018. [www.protectedplanet.net](http://www.protectedplanet.net) (2018).
71. Iacus, S. M., King, G. & Porro, G. cem : Software for coarsened exact matching. *J. Stat. Soft.* **30**, 1–27 (2009).
72. King, G. & Nielsen, R. Why propensity scores should not be used for matching. *Polit. Anal.* **27**, 435–454 (2019).
73. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71**, 319–392 (2009).
74. R Core Team. R: A language and environment for statistical computing. Preprint at (2019).
75. Lindgren, F. & Rue, H. Bayesian Spatial Modelling with R-INLA. *J. Stat. Softw.* **63**, 1–25 (2015).
76. Krainski, E. T. *et al.* *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA* (Chapman and Hall/CRC Press, 2020).
77. Gómez-Rubio, V. *Bayesian inference with INLA* (Chapman and Hall/CRC Press, 2020).
78. Hartig, F. DHARMA: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models (2022).

## Acknowledgements

JM is supported through the Natural Environment Research Council's E3 Doctoral Training Partnership (grant NE/L002558/1) in partnership with the Royal Society for the Protection of Birds.

## Author contributions

J.M. conceptualized the study, analysed the data, and wrote the manuscript. E.T.A.M., G.B., and A.K. conceptualized the study, provided supervision, and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-22762-4>.

**Correspondence** and requests for materials should be addressed to J.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022, corrected publication 2023