









Replanting unproductive palm oil with smallholder plantations can help achieve Sustainable Development Goals in Sumatra, Indonesia

Ariadna Fosch ^{1,2,3,4✉}, Guilherme Ferraz de Arruda ^{1,3}, Alberto Aleta ^{2,4}, Adrià Descals^{5,6}, David Gaveau ^{7,8}, Courtney Morgans ⁹, Truly Santika¹⁰, Matthew J. Struebig ⁹, Erik Meijaard ^{9,11,12✉} & Yamir Moreno ^{1,2,3,4✉}

Oil palm (*Elaeis guineensis*) is a controversial crop. To assess its sustainability, we analysed the contribution of different types of plantations (smallholder, industrial and unproductive) towards meeting six Sustainable Development Goals. Using spatial econometric methods and data from 25,067 villages in Sumatra, Indonesia, we revealed that unproductive plantations are associated with more cases of malnutrition, worsened school access, more air pollution and increased criminality. We also proposed a strategy for sustainable palm oil expansion based on replanting unproductive plantations with either industrial or smallholder palm oil. Smallholder replanting was beneficial for five Goals (Zero poverty, Good health, Quality Education, Environmental preservation and Crime reduction), while the same intervention only improved two Goals in the industrial case (Zero poverty and Quality Education). Our appraisal is relevant to policymakers aiming towards the 2030 Agenda, organisations planning oil palm expansion, and retailers or consumers concerned about the sustainability of oil consumption.

¹CENTA Institute, Torino, Italy. ²Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Zaragoza, Spain. ³Fondazione ISI, Torino, Italy. ⁴Department of Theoretical Physics, University of Zaragoza, Zaragoza, Spain. ⁵CREAF, Center for Ecological and Forestry Applications, 08193 Cerdanyola del Vallès, Catalonia, Spain. ⁶CSIC, Consejo Superior de Investigaciones Científicas, Unitat d'Ecologia Global CREAM-CSIC-UAB, 08193 Bellaterra, Catalonia, Spain. ⁷TheTreeMap, Bagadou Bas, 46600 Martel, France. ⁸Jeffrey Sachs Center on Sustainable Development, Sunway University, 5, Jalan Universiti, Bandar Sunway, 47500 Petaling Jaya, Selangor, Malaysia. ⁹Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Canterbury, Kent CT2 7NR, UK. ¹⁰Natural Resources Institute (NRI), University of Greenwich, Chatham Maritime ME4 4TB, UK. ¹¹Borneo Futures, Bandar Seri Begawan, Brunei. ¹²The University of Queensland, School of Biological Sciences, Brisbane, QLD, Australia. ✉email: arifosch@gmail.com; emeijaard@borneofutures.org; yamir.moreno@gmail.com

In 2015, all United Nations Member States adopted the 2030 Agenda for Sustainable Development, an urgent call for action to all countries to achieve sustainable development by 2030¹. The Sustainable Development Goals (SDGs) are at the core of the initiative - 17 guiding principles encompassing economic growth, environmental protection, global peace, access to healthcare, basic education and others that should be followed by all countries to aim towards a more sustainable future. Since its adoption, many public and private organisations have also relied on the SDGs to evaluate the sustainability of their activities. Nonetheless, this analysis can be challenging, especially when the activities evaluated can have complex outcomes for many different development dimensions. Palm oil production in tropical countries is a typical example of such complex interactions between different SDGs².

The palm oil industry is frequently criticised due to environmental and social impacts³⁻⁵. Specifically, for contributing to air and water pollution⁶, biodiversity loss⁷ and a decline in social well-being in neighbouring communities due to land use conflict and labour exploitation⁸. However, palm oil production is also an important vehicle for economic growth and poverty alleviation with recent large-scale impact evaluations emphasising the trade-offs associated with the industry's influence on the SDGs^{7,9}. As global demand for vegetable oils increases from 201 to 307 million tonnes (Mt) by 2050^{4,10}, identifying the role of palm oil production in achieving the multifaceted elements of the SDGs will become crucial to guide policymakers towards more sustainable production systems, given its relatively low land requirements compared to other oil crops⁴.

As the largest worldwide palm oil producer, Indonesia provides the ideal case study to evaluate the contribution of palm oil to the SDGs^{11,12}. The Indonesian economy has thrived through palm oil production, with 17% of Indonesia's agricultural gross domestic product in 2014 depending on this crop^{13,14}. However, its expansion has been associated for many years with deforestation, ecological damage and a mix of positive and negative social outcomes^{3-5,9,11,15}. The Indonesian scenario is even more complex when considering the heterogeneity between the industrial and smallholder palm oil production profiles. Industrial plantations tend to be large, with thousands of hectares of land dedicated to the monoculture of palm oil. Smallholder plantations, on the other hand, are typically smaller (<25 hectares according to government definition - although wealthy individuals can sometimes own several hundred hectares). They also tend to exhibit a more heterogeneous spatial pattern, where palm oil is often mixed with other crops or land dedicated to other uses^{16,17}. The differences between both production profiles could also entail discrepancies in their ecological footprint¹⁸ and the socio-economic characteristics of the villages where they are located.

In this study, we evaluate the differences between smallholder and large-holder industrial plantations through the perspective of the SDGs. The analysis relies on spatial econometric models to reveal correlations between the socio-economic characteristics of the villages in the region of Sumatra (Indonesia) and their land dedicated to different palm oil production profiles (active industrial plantations, active smallholder plantations and two types of unproductive plantations). To characterise the contribution of each production system to the SDGs, we explore the differences in seven SDG indicators if the unproductive (damaged) plantations in Sumatra were replanted through a smallholder or an industrial profile. Our analysis reveals how successfully replanting damaged plantations with smallholder palm oil exhibits more beneficial outcomes toward most SDGs than its industrial equivalent. However, smallholder palm oil production is also responsible for most unproductive (damaged) plantations (72% of the total damaged land has smallholder

tenure), which indicates the need for careful planning and additional policies to ensure the viability of the replanting strategy. Our appraisal could be relevant for policymakers and stakeholders aiming to complete the 2030 Agenda in the context of palm oil production in Indonesia.

Results

Socio-economic differences between palm oil production systems. Palm oil production can play a major role in the socio-economic development of the villages in which is produced. However, it is still unclear how these impacts could vary depending on the production profile used for its production. To this end, we combined two datasets describing, respectively, the socio-economic characteristics of the villages of Sumatra (Indonesia) and their total area dedicated to palm oil plantations, classified by their type of ownership (industrial/smallholder) and production status (young, mature and unproductive damaged or unproductive but possibly replanted)^{16,17}. See Methods section for a more detailed description.

We aggregated the palm oil data into four production profiles: active industrial palm oil, active smallholder palm oil, unproductive damaged plantations and unproductive but possibly replanted plantations (see Methods for more details). Then, we estimated the socio-economic characteristics significantly associated with each profile through two types of spatial regression analysis. The Spatial Durbin Model (SDM)¹⁹, a global spatial dependency model and the Spatial Durbin Error Model (SDEM)¹⁹, a local-dependency variant. Both models produced broadly consistent results (see Supplementary Tables 5-8).

Unproductive (damaged) plantations are significantly associated with more cases of malnutrition, crime, air pollution and a higher distance to primary education centres (see Fig. 1). These socio-economic characteristics are detrimental to SDG 2 (Zero hunger), SDG 4 (Quality education), SDG 15 (Life on land) and SDG 16 (World peace). Damaged plantations are also positively correlated with more healthcare facilities and more types of credit facilities, two features that are beneficial respectively for SDG 3 (Good Health and well-being) and SDG 8 (Economic growth). Nevertheless, these positive associations are not unique to unproductive (damaged) land. They were also observed in the case of active smallholder plantations, which could indicate that such infrastructures are built during the early stages of plantation development and remain after their failure. These characteristics differ from the ones observed for unproductive (but possibly replanted) areas. In that case, we observe fewer negative effects overall and slight beneficial outcomes for poverty reduction.

Villages with industrial or smallholder plantations share some common characteristics compared to non-oil palm producing villages: they are less associated with natural disasters and tend to have fewer households receiving *Surat Keterangan Tidak Mampu* (SKTM) poverty letters. These letters are issued to families below the national poverty level to facilitate access to essential resources such as fuel, healthcare or education²⁰. Prior studies have already reported a positive influence of palm oil production on poverty alleviation^{9,11}. In our case, these beneficial economic outcomes are only observed for villages with active plantations from both profiles and for possibly replanted areas, which highlights the negative role of damaged plantations in sustainable development.

Additionally, we found that several characteristics distinguish villages with smallholder and industrial plantations. Those with a smallholder profile are associated with more credit facilities and cases of malnutrition, while in the case of industrial land, the reduced distance to primary schools is one of the main differential traits. The diversity observed in both production profiles evidences the need to evaluate their sustainability

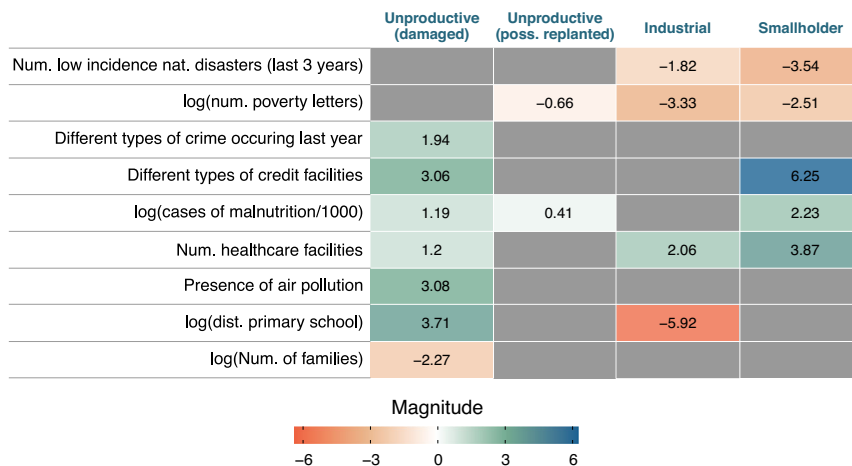


Fig. 1 Significant total spatial impacts for different palm oil production profiles. The profiles evaluated are: unproductive (damaged) plantations, unproductive (but possibly replanted) plantations, active industrial palm oil and active smallholder palm oil. Positive spatial impacts indicate a positive correlation between village land and the socio-economic characteristic involved, while negative spatial impacts indicate negative correlations (see Methods for a formal description of the spatial impacts metric). Grey panels indicate not statistically significant associations (p -value > 0.05). The spatial impacts were estimated using the SDM model trained as described in the Methods section, which resulted in a Nagelkerke’s $pseudo - R^2$ between 0.38 – 0.63 for the four models. Other quality metrics are available in Supplementary Tables 5-8.

separately. This is especially relevant for planning interventions over land use, as the same intervention could entail different sustainability outcomes depending on the village’s predominant production system.

Counterfactual analysis. With the global increase in vegetable oil demand and related environmental impacts of agricultural expansion, it is crucial to develop strategies to increase palm oil production while maintaining Indonesia’s efforts toward net zero deforestation²¹. With damaged palm oil plantations having multiple detrimental outcomes on sustainable development, replanting them could be a key strategy for sustainable palm oil expansion. Nevertheless, it is still unclear whether smallholder or industrial palm oil is better at maximising the beneficial outcomes for the SDGs. We, therefore, implemented a counterfactual analysis to estimate the change in seven SDG indicators if all unproductive (damaged) plantations were replanted through an industrial or smallholder profile. The seven indicators used for the counterfactual analysis span six different development goals (see Methods for the selection criteria). These are SDG 1 (Zero poverty): log(num. SKTM poverty letters), SDG 3 (Healthcare and Well-being): Num. healthcare facilities, SDG 4 (Quality Education): log(dist. primary school) and log(dist. junior-high school), SDG 9 (Industry, innovation and infrastructure): log(dist. nearest market), SDG 15 (Life on land): percentage of village area covered by forest, SDG 16 (Peace and Justice): number of different types of crime.

The counterfactual analysis followed the pipeline described in Fig. 2. The SDM model was trained seven times, using each SDG indicator as the dependent variable y and all the village characteristics (socio-economic and palm oil production features) as independent features X . The trained models were used to predict the expected change in the SDG indicators if all unproductive (damaged) land in Sumatra was replanted with either industrial or smallholder plantations.

The predictions obtained for each SDG indicator can be visualised as the counterfactual maps shown in Supplementary Fig. 2. These maps provide a detailed description of the location and magnitude of the changes produced in the SDGs, but they do not indicate how the intervention affects the well-being of local communities. For example, removing one hospital from a village

without other healthcare facilities might have a more drastic impact on the villager’s well-being than the same change in a village with other healthcare centres.

To contextualise the results in terms of the village’s well-being, we defined a deprivation threshold for each SDG indicator using the reference values proposed in the Indonesian Village Development Goals (*Indeks Pembangunan Desa*)²² (see Methods section for more details). Interventions that induce a change from a deprived state to a non-deprived state or vice-versa will be considered highly positive or highly negative contributors to the SDGs, as they drastically affect the well-being of the villages. If the intervention does not induce a change in the deprivation state, it will only be considered a moderate influence, regardless of the magnitude of the change observed. Finally, interventions where no changes in the SDG indicators are induced, are considered neutral interventions. Using these criteria, the counterfactual maps were translated into Fig. 3, which summarises the effects of both interventions across all villages.

Replanting unproductive (damaged) plantations is associated with a positive contribution towards SDG 1 for both production profiles. The smallholder strategy resulted in a slightly more beneficial outcome, as it is associated with highly positive effects for more than 27% of the villages where the counterfactual was applied. In the case of SDG 3, replanting unproductive land with industrial plantations was associated with a reduction in the number of hospitals, a negative influence on the SDG, while smallholder expansion is positively impacting the objective. This outcome must be further validated as the apparent lack of medical coverage in industrial plantations could be because some company-owned health facilities may not be included in the socio-economic data.

For SDG 4 both production profiles have mainly negative outcomes towards reducing the distance to junior-high schools. However, the industrial strategy seems to result in more beneficial effects towards improving the distance to primary school centres. Thus, in this case, promoting industrial plantations could be more beneficial overall for SDG 4, also because similar to health care, many companies provide schooling for children of their workforce²³.

Increasing industrial plantations is also associated with a small reduction in forest coverage, leading to a moderately negative

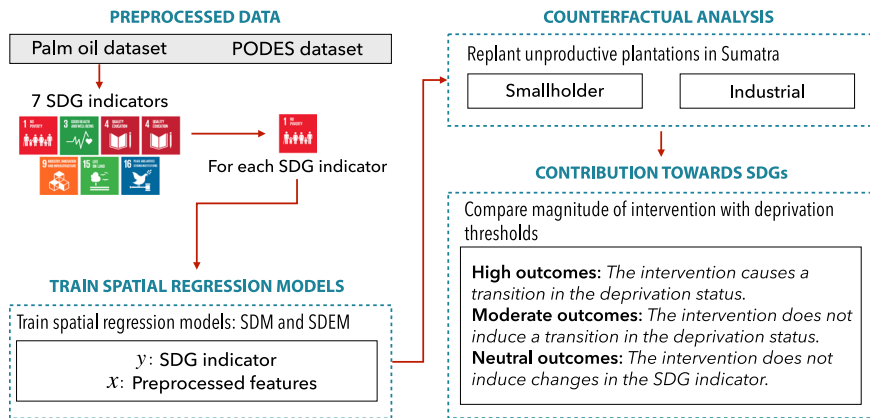


Fig. 2 Structure of the counterfactual analysis. A complete description of the pipeline can be found in the Methods section. The SDG icons are property of the United Nations, reprinted for informational purposes in accordance with ref. 67 (all rights reserved).

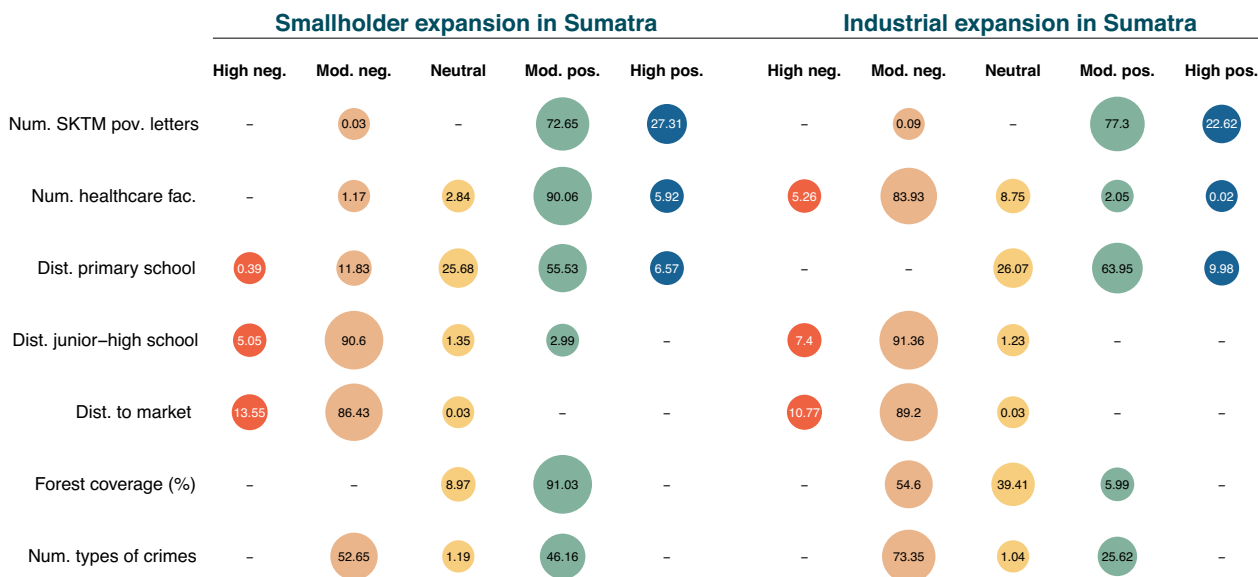


Fig. 3 Contribution towards the SDGs of replanting unproductive land with smallholder or industrial plantations. The values reported indicate the percentage of villages where the implementation of the counterfactual caused high (positive/negative), moderate (positive/negative), or neutral outcomes. Note that the percentage is computed relative to the number of villages affected by the counterfactual analysis. Interventions causing the transition from a deprived to a non-deprived village are regarded as highly positive interventions. In contrast, interventions with a beneficial outcome for the SDG indicator that do not entail a change in the village's deprivation state are only regarded as moderately positive. Similarly, interventions inducing the transition from the non-deprived to the deprived state are considered highly negative, while interventions with negative influences that do not cause major changes in well-being are only considered moderately negative. Interventions where no changes in the SDG indicator are produced are regarded as neutral.

outcome for SDG 15. Contrarily, replanting unproductive (damaged) land with smallholder plantations correlates with increased forest coverage levels, which is favourable for environmental protection. Though this outcome may seem counter-intuitive, it may result from the negative association between damaged land and SDG 15. Since palm oil replanting reduces the amount of damaged land, it is expected that a correlation analysis would interpret it as a positive outcome for SDG 15. However, this does not imply a causal relationship between both effects. Replanting unproductive (damaged) plantations is associated with negative outcomes for SDG9, for which both interventions induced an increase in the distance to the nearest market. This could result in oil palm development displacing local production and trade of goods usually supplied to local markets, such as fruits. Finally, for SDG 16, we identified that the interventions are associated with an increase in the number of different types of

crime or a decrease in crime variability depending on the village where it is implemented.

Overall, replanting damaged plantations was more beneficial when implemented through a smallholder profile than its industrial equivalent. Smallholder replanting was beneficial for SDGs 1, 3 and 15 and it exhibited mixed effects for SDGs 16 and 4 (positive effect for primary schools and negative for junior-high schools). Meanwhile, industrial replanting only presented synergistic associations with SDG 1 and one facet of SDG 4 (primary school). This last association was the only one more beneficial in the industrial replanting scenario than in the smallholder case. Interestingly, both counterfactual scenarios were not compatible with the progress of SDG 9 and SDG 4, in the case of the junior-high school indicator. This revealed the presence of a trade-off effect between the intervention proposed and the achievement of some development goals.

The findings demonstrate the intricate nature of the connections between an intervention and the SDGs. An identical intervention can yield positive outcomes for certain indicators while having negative effects on others, leading to various synergistic or trade-off relationships with the development goals.

Province analysis. To assess the scale dependency of the findings, we repeated the counterfactual analysis for each of the ten provinces in Sumatra individually (see Supplementary Figs. 3-12). We then estimated the production profile that yielded the most favourable and detrimental outcomes for each province and compared it to the results observed for the entire island.

The results presented in Table 1 confirmed the trends observed in the study for all of Sumatra. Replanting by smallholders was associated with the most beneficial outcomes for the SDGs across all provinces. The only exception was the case of SDG 4 (for the primary school indicator), where industrial replanting consistently outperformed smallholder replanting across the regions evaluated. This analysis also revealed consistent results in their trade-off effects, as replanting damaged plantations also entailed detrimental outcomes for SDG 9 across all provinces.

In general, the proposed counterfactual approach enabled us to differentiate between two seemingly similar interventions concerning palm oil usage. This tool could allow policy-makers to make more informed decisions to achieve sustainable palm oil development. However, it is crucial to recognise that the interplays described by the counterfactual evaluation are conditional on successfully transforming unproductive terrains into active plantations. Thus, guaranteeing the viability of the replanted plantations is of utmost importance for the sustainability of the strategy.

Sumatran villages tend to have more unproductive (damaged) plantations associated with smallholder tenure than with industrial palm oil (see Fig. 4). In terms of total area, 72.5% of the total unproductive damaged land is associated with the smallholder production profile (calculation in the Supplementary Methods). It is crucial to identify the mechanism behind the increased failure rate of smallholder plantations. Whether it is caused by sub-optimal soil and planting materials²⁴; or if indeed smallholder producers sometimes lack the capital or knowledge to make plantations thrive²⁴. This could reveal if replanting damaged plantations with a smallholder profile is indeed

sustainable by itself, or if other supplementary programs (e.g., educational, economic, etc.) should also be introduced to increase its effectiveness.








Discussion

The sustainability of palm oil production is frequently called into question, especially by commentators in high-income countries. This criticism focuses strongly on the environmental impacts of oil palm but it is less clear the extent to which negative impacts on the environment are balanced by positive impacts elsewhere. This is at the core of sustainable development considerations. The SDGs are intended to be targeted simultaneously, resulting in positive outcomes for all these objectives. In reality, trade-offs are, however, likely to occur. Here we have analysed the extent to which oil palm development in Indonesia, the world’s largest palm oil producer, results in synergies and trade-offs between different SDGs, and how this varies in different production profiles. This analysis provides a basis for a better understanding of the overall impact of oil palm on sustainable development and also a way to steer future developments that maximise positive outcomes and minimise negative ones.

Oil palm expansion, if indeed sought by the Indonesian government, should preferably occur on existing agricultural land, or in areas where previous attempts to cultivate oil palm have been ineffective²⁵. Replanting failed or poorly performing plantations with smallholder palm oil was associated with beneficial effects for poverty, healthcare provision, terrestrial environments, primary education access and in some villages crime variability reduction. This indicates a synergistic association between the proposed intervention and SDGs 1, 3, 4, 15 and 16. Meanwhile, industrial replanting only entailed beneficial outcomes for education and poverty (SDGs 4 and 1). These outcomes are quite consistent across all provinces and would indicate that smallholder production may be more beneficial for the SDGs than industrial plantations.

However, we also observed that 72,5% of unproductive (damaged) land in Sumatra was associated with smallholder tenure. The beneficial outcomes of the intervention proposed are conditional on transforming unproductive land into viable plantations. Thus, it is crucial to ensure the replanted areas do not fail again after the intervention. To this end, policymakers should invest in comprehending the mechanism behind the increased

Table 1 Summary of the palm oil production system with the most beneficial and detrimental effects in all provinces of Sumatra.

														
	Pov. letters		Healthcare fac.		Dist. primary sch.		Dist. high sch.		Dist. market		Forest Coverage		Types crime	
	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.
Aceh	S	I	S	I	I	S	S	I	-	I/S	S	I	S	I
Sum. Utara	S	I	S	I	I	S	S	I	-	I/S	S	I	S	I
Sum. Barat	I/S	-	S	I	I	S	S	I	-	I/S	S	I	S	I
Riau	I/S	-	S	I	I	S	S	I	-	S	S	I	S	I
Jambi	I	-	S	I	I	S	S	I	-	I	S	I	S	I
Sumatera Selatan	S	I	S	I	I	S	S	I	-	I/S	S	I	S	I
Bengkulu	S	I/S	S	I	I	S	S	I	-	I/S	S	I	S	I
Lampung	S	I	S	I	I	S	S	I	-	I/S	S	I	S	I
Kep. Bang. Belitung	I/S	-	S	I	I	S	S	I	-	I/S	S	I	S	I
Kep. Riau	I/S	-	I/S	I/S	-	-	S	I	-	I/S	S	I	S	I
Sumatra	S	I	S	I	I	S	S	I	-	S	S	I	S	I

“S” indicates that smallholder production has the largest (positive or negative) influence on the SDG indicators. “I” indicates that industrial production is the one with the greatest (positive or negative) influence on the SDG indicators. “I/S” indicates that both production systems exhibit the same positive or adverse effects. Missing values (“-”) indicate cases where no positive or negative associations are found. The geographical location of each province can be found in Supplementary Fig. 1. The SDG icons are property of the United Nations, reprinted for informational purposes in accordance with ref. 67 (all rights reserved).

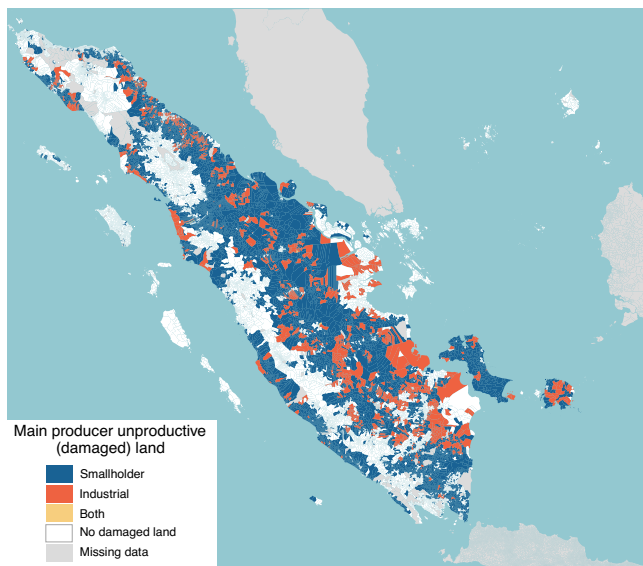


Fig. 4 Distribution of unproductive (damaged) land in Sumatra by type of plantation. Villages coloured according to the production profile responsible for most damaged land in each village. Blue for smallholder plantations, red for industrial plantations and yellow if the percentage of village area covered by unproductive (damaged) land is the same for both profiles (occurring in very few villages). Villages without damaged land are coloured in white, while missing data is coloured in grey.

failure rate of smallholder production to determine if it is linked to land quality or management issues.

Both mechanisms are often difficult to differentiate. Peatlands and hilly areas are sub-optimal for oil palm, the former because of low pH and low productivity (and high fertiliser needs), the latter because of potential lack of water and soil erosion^{26,27}. These are often marginal lands where aspiring farmers might try to start a plantation, but if the costs for generating good yields are too high they abandon it⁵. Complicating this is the fact that many smallholders opt for low-input low-output systems²⁸ that can generate similar profits to more intensively managed plantations but may have higher failure rates (i.e., the cost of walking away from a plantation is lower when investments have been low). Palm oil can also fail due to natural factors, like the presence of diseases, fire²⁹ or floods³⁰. In that case, it is necessary to comprehend why some landowners may not be rapidly re-purposing plantations after their failure. This could be linked to land speculation practices^{31,32}, contested land ownership, insufficient funding, or the lack of technical skills to re-purpose the plantation. Overall, comprehending the causes behind plantation failure could reveal the locations most suitable for the replanting strategy proposed and whether supplementary policies (educational programs or financial/technical support) are required to maximise its success chances.

Even if replanting damaged plantations with smallholder palm oil could aid sustainable development in Sumatra, these conclusions should not be directly extrapolated to other areas. The region of Sumatra is in a more advanced stage of oil palm development compared to other parts of Indonesia, such as Kalimantan and Papua¹⁶. Additionally, it is also the region with the largest amount of unproductive plantations, making it an ideal scenario to test the replanting strategy. If there is interest in assessing the sustainability of similar interventions in other Indonesian regions, or other countries, the whole counterfactual analysis should be repeated using adequate data.

Despite its caveats, our study provides valuable insights for government authorities and policy-makers involved in planning

oil palm and other monocultural plantation crops. We offer the possibility of evaluating simultaneously the impact of an intervention over seven different SDG indicators, revealing some of the existing synergies and trade-offs among them. With adequate data, this approach could be expanded to evaluate all 17 SDGs, providing a handy tool to evaluate the sustainability of palm oil production as a whole. Moreover, bringing together land cover, production scale and socio-economic indicators is an important step towards improving understanding of how agricultural development can have the greatest positive SDG impacts and avoid negatively affecting people, especially those in forest environments.

Our study also opens new lines of research in different fields. For example, it could be used as a hypothesis-generating tool to design on-site sociological studies. A similar approach combining data-driven models and experimental studies is starting to be used in SDG assessment³³. However, to the best of our knowledge, it has not yet been applied in the context of palm oil research. Complementing our study with household surveys could reveal the causal linkages behind the correlations identified in this study. It could also shed some light on the impacts of palm oil expansion on the villages' minority groups.

The model proposed could also be used to test other types of interventions in palm oil use. For example, there would be value in studying the social and environmental benefits of reforesting damaged plantations. This could help clarify if less intensive land uses (agroforestry cultivation) would be preferred in unproductive plantations with poor-quality soils. Another interesting counterfactual scenario could be assessing the impact of climate change on sustainable palm oil production. Tropical production areas are predicted to become drier and hotter³⁴, causing alterations in rainfall patterns and water stress. This makes oil palm more prone to pest and disease outbreaks, potentially reducing its yields. Other risks include sea-level rise in coastal production areas, especially on peat soils³⁵, and increased wildfires³⁶. These changes will likely influence interplays between palm oil production and the SDGs and should be modelled specifically to guide sustainable policies.

Other lines of research could be directed towards improving the quality of the data used in this analysis. The socio-economic data for 2018 was obtained from the Village Potential Survey (PODES) survey, an Indonesian survey of village well-being with a much higher spatial resolution than the one usually reported in the SDG Global Database (the United Nation's official country-wide SDG reports)³⁷. However, it only contains a small subset of variables that can be used as SDG indicators, only providing partial insight into each goal. For example, the use of the distance to education centres as a proxy indicator for SDG 4, does not capture other facets of this goal like the quality of the education provided. Improving the resolution of the socio-economic data could allow us to better characterise the SDGs from a multi-dimensional perspective.

Regarding the palm oil dataset, our study would significantly benefit from an improved ownership categorisation. The binary differentiation between smallholder and industrial plantations could be masking different sustainability outcomes between smallholders with different characteristics. This categorisation also disregards whether the owner resides in the same village where the plantation is located. However, we worked under the assumption that many of the costs and benefits of palm oil accrue irrespective of the owner's residence (e.g. ecological impacts). We also chose not to incorporate palm oil trade as a predictor in our analysis, despite its role in driving palm oil production through international and national demand. Estimating the influence of demand on individual villages would require extensive surveys to gather specific data, and we believe that palm oil demand does

not have a distinct impact on the SDG indicators at the village level beyond its influence on palm oil production itself.

Our study also provides some methodological improvements to ease the introduction to the field of spatial econometrics for researchers in the palm oil community. The use of spatial data in palm oil research has thrived in recent years^{3,16,17,26,38}. However, few studies have decided to follow spatial econometric approaches to evaluate the sustainability of palm oil plantations²⁶. Instead, the main focus has been on assessing the impacts of interventions retrospectively. This approach is very useful for describing the causal effects of an intervention but it does not take advantage of the spatial nature of the data^{3,7,39}. The use of spatial regression models offers the possibility for predictive evaluation, which could be a very valuable tool for policy-makers planning new interventions. To facilitate the use of spatial regression models in sustainability research we have improved the model selection pipeline described in ref. ⁴⁰, which was one of the most complete approaches towards model selection in spatial econometrics.

Overall, we hope this study can open the door to more research exploring the environmental and social interests of oil production. The current analysis and methodological approach also allow for future comparison between the SDG impacts of different vegetable oil crops, such as soybean, rapeseed/canola, coconut, peanut or sunflower. All these crops have different oil yields and thus land requirements and come with different sustainability profiles because of their different geographical location. The trade-offs and synergies between different crops and production profiles have never been analysed, making it impossible to compare their sustainability meaningfully. The current study thus provides a first methodological step towards such a global assessment of the sustainability of vegetable oil production.

Methods

Data. We relied on high-spatial-resolution data from 26,505 villages in Sumatra describing four different palm oil production situations, the area of forest and the socio-economic characteristics of each village. The land use data was generated by combining two satellite-derived datasets from refs. ^{17,41} and refs. ^{16,42}, which also specifies their type of ownership (smallholder or industrial). The approach used to classify the plantation's ownership relies on their differences in size, land homogeneity and infrastructure development and it is further described in the Supplementary Methods.

References ^{17,41} used radar and optical images (Sentinel-1 and Sentinel-2) to detect the area covered by closed-canopy oil palm plantations in 2019. These plantations are already productive, generally when oil palm is older than 3 years. On the other hand, refs. ^{16,42} used historical optical images (Landsat) to identify the year when the land was cleared to plant oil palm (between 2001 and 2019). This analysis considers all land cleared for palm oil production, including immature, damaged and failed plantations; and thus surpasses the estimated extent of the closed-canopy plantations. By combining the two datasets above with an additional analysis of recent replanting (from Sentinel-2 images) we generated the following four plantation classes: 1) mature oil palm corresponds to land cleared for oil palm before 2017 and detected as closed-canopy oil palm in 2019; 2) young oil palm is the land cleared for oil palm after 2017 (inclusive) and not detected as closed-canopy oil palm in 2019; unproductive oil palm corresponds to land cleared for oil palm before 2017 and is not closed canopy in 2019; 4) possibly replanted plantations is the land cleared for oil palm before 2017, is not closed-canopy in 2019, but is detected cleared between 2017 and 2019 if the normalised difference vegetation index from Sentinel-2 images presented a sharp decrease during 2017–2019. “Possibly replanted

plantations” can be interpreted as failed or damaged plantations established before 2017, cleared and replanted during 2017–2019. We specified separately the four production classes for each production profile (industrial and smallholder), resulting in the eight different types of palm oil plantations detailed in Supplementary Table 2.

To measure the forest area in 2019 at the village level we combined a map of the “primary” forest extent in 2000⁴³ and an annual forest loss map ref. ⁴⁴. The forest cover for the year 2000 was produced by including intact and degraded primary forest types described in ref. ⁴³. This map was updated to 2019 using the forest change maps in ref. ⁴⁴. Areas of forest in 2000 cleared between 2001 and 2018 were marked “cleared previously” and removed in the updated 2019 forest map.

For our regression analysis, the village area occupied by the eight plantation classes and the forest coverage was reported relative to the total village area. Absolute land units were only used for the estimation of the total percentage of unproductive land.

The socio-economic characteristics of the Indonesian villages were obtained from the PODES for 2018⁴¹. PODES is a national survey carried out by the Indonesian Central Bureau of Statistics every 2–4 years and includes more than 100 questions about the economic, social and environmental characteristics of Indonesian villages. We selected questions of most relevance to the SDGs and synthesised response variables into 27 features, 6 describing general village characteristics and 21 reflecting different indicators directly related to the SDGs. While the PODES survey was conducted in 2018, the oil palm datasets were created for 2019. The datasets have a one-year mismatch, but this mismatch is negligible because the oil palm plantation area does not change substantially from year to year.

SDG indicators. SDG indicators are quantifiable characteristics that reflect the progress towards achieving each development goal. Using the “Global indicator framework for the SDGs”⁴⁵, we identified seven features in our dataset associated with some objectives of the SDGs:

- Zero poverty: SDG 1, is dedicated to poverty eradication “End poverty in all its forms everywhere”⁴⁵. More precisely, objective 1.2 in ref. ⁴⁵ states, “By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions”. The closest indicator in the PODES dataset is the number of SKTM poverty letters. SKTM poverty letters have already been used as a proxy for poverty in other studies^{20,46,47}, and they reflect a poverty alleviation system created by the Indonesian government to fight against national poverty. SKTM letters are issued to families below the national poverty level to facilitate access to essential resources such as fuel, healthcare, or education²⁰. As such, they should provide a good proxy to determine the amount of poverty present in each village (see Supplementary Methods for a correlation analysis between poverty letters and income). The number of poverty letters is negatively associated with SDG 1 as the objective of this SDG is to reduce the number of poor families;
- Good health and well-being: objective 3.8 states the need to “Achieve universal health coverage”⁴⁵. Therefore, we have relied upon the number of healthcare facilities as an indicator for SDG 3. This indicator is positively associated with SDG 3 since increasing the number of hospitals is beneficial to the development goal;
- Quality education: SDG 4 considers access to primary and secondary education⁴⁵. In this case, two variables were

considered as SDG indicators, the distance to primary schools (km) and the distance to junior-high schools. Improving access to education facilities is crucial to ensure that all youth achieve literacy and numeracy, as proposed in objective 4.6 of the SDGs⁴⁵. Both indicators exhibit a negative association with the development goal, as the distance to education centres needs to be reduced to improve this objective;

- Industry, innovation and infrastructure: SDG 9 has been represented using the distance to the nearest market. Objective 9.1 states the need to promote the creation of resilient infrastructures that help to develop local communities and their economy⁴⁵. The most related indicator in the PODES dataset was the distance to the nearest market (km), which exhibits a negative association with the SDG it represents;
- Life on land: for SDG 15, which addresses environmental protection, we have relied on the percentage of village area covered by forest as the SDG indicator. This feature is directly proposed in the Global indicator framework of the SDGs as an indicator for objective 15.1 “Ensure the conservation, restoration and sustainable use of terrestrial and inland freshwater ecosystems and their services, in particular, forests”⁴⁵. The indicator is positively associated with the SDG;
- Peace and Justice: for Goal 16 “Promote peaceful and inclusive societies for sustainable development”⁴⁵, we have relied on the number of different types of crimes as the SDG indicator. This variable indicates the variety of crimes occurring in a village and was used as a proxy for criminality. Achieving world peace requires a reduction of criminality; thus our indicator is negatively associated with the SDG.

Note that the signs of the associations between the indicators and the SDGs play a major role in interpreting the positive and negative effects. Increasing one indicator does not always translate into beneficial outcomes for the SDG they represent.

The choice of SDG indicators was influenced by the methodological prerequisites of the spatial models utilised. The spatial econometric models used are designed to use continuous variables as their dependent feature. Consequently, we exclusively selected indicators of a similar nature as our SDG indicators.

Preprocessing. The dataset contains continuous, discrete, categorical and binary variables. Thus, the preprocessing differed for each type of feature. Categorical variables were transformed into binary “dummy” variables through a one-hot-encoding approach⁴⁸. In the process, one dummy variable for each categorical feature was removed to avoid perfect colinearity between the rest of the variables. Some other features were preprocessed to remove potential colinearities. The number of core natural disasters was subtracted from the total number of natural disasters to create an independent variable describing only low-incidence disasters. Also, the “village’s main source of income” was combined with the feature describing the “village’s main agricultural product” to avoid the colinearity between them. Given the non-normal distribution of some variables, we applied a log transformation to bring them closer to normality. For each feature x , the transformation $\log(x + 1)$ was applied. The discarded features are detailed in the Supplementary Methods.

Finally, 1,438 villages were removed from the analysis either because of missing data or due to inconsistencies between the number of families and the number of poverty letters issued. The final dataset (after filtering) contains data for 25,067 villages in Sumatra and is described in Supplementary Table 1.

Model selection. Through the use of the Global Moran’s I (MI)⁴⁹ we unveiled the presence of significant spatial clustering in all the features of our dataset and the residuals of the OLS regression, see Supplementary Table 3. Thereby, we relied on spatial regression models for the analysis. This methodology has been widely used in spatial econometrics, where it is common to observe that the geographical location of the samples can affect their characteristics^{50–52}.

To determine the preferred spatial regression model for the analysis, we followed the pipeline described in Fig. 5. This pipeline is based on the approach described in ref. ⁴⁰, but it was expanded to evaluate the Spatial Lagged X model (SLX) and the Manski model. The pipeline starts from the OLS model and evaluates the need to include spatial dependency in the dependent variable y or in the error term ε through the use of the Lagrange multiplier (LM)⁵³ and the Robust Lagrange Multiplier (RLM)⁵⁴. Both tests are based on the residuals of the OLS model and follow a chi-squared distribution with one degree of freedom⁴⁰. Suppose the hypothesis is rejected in favour of one type of dependency (p -value ≤ 0.05), the pipeline is continued only using spatial regression models with the preferred type of dependency, respectively the Spatial Autoregressive model (SAR) and the Spatial Error model (SEM), see ref. ¹⁹ for a formal description of these models. If both models are identified as significant, the pipeline is continued for both types of models.

The SAR and SEM models can be expanded by including spatial dependency in the independent features X , which results in the SDM and the SDEM. We relied on the Likelihood ratio (LR) to evaluate if this extension is needed. Using the LR it is possible to test if the SDM and SDEM can be simplified to their nested versions SAR and SEM, respectively. The LR was also used to evaluate if the SDM and SDEM models could be simplified into the Spatial Lagged X model (SLX)¹⁹, a model only considering spatial dependency over the independent features. A significant LR (p -value ≤ 0.05) indicates that the H_0 assuming that the complex model can be simplified to its nested version is rejected, thus the complex model is preferred. Contrarily if the test is not significant, the simplified version is preferred. If both models cannot be simplified they are both recommended for the sustainability analysis.

Finally, we evaluate if the SDM or SDEM models capture the whole spatial dependency in the data through the MI on the residuals of both regressions. If the residuals of the regressions are still spatially correlated, the preferred model is a model considering all types of spatial dependency, the Manski model⁵⁵. More details regarding the application of the pipeline to our dataset are available in the Supplementary Methods.

Spatial Durbin Model. The primary analysis was performed using the Spatial Durbin Model, which is defined as

$$y = \alpha I_N + X\beta + \rho Wy + WX\theta + \varepsilon, \quad (1)$$

for a model with N spatial units and K features. The term y is an $N \times 1$ vector describing the dependent variable for all the spatial units in the dataset. The αI_N term is a constant vector of dimensions $N \times 1$ that reflects the model’s intercept. The term $X\beta$ describes the dependency on the exogenous explanatory variables, where X is the $N \times K$ matrix describing independent features, and β is its $K \times 1$ coefficient vector. The spatial auto-regressive term ρWy introduces spatial dependency over y by assuming that W is the $N \times N$ weight matrix describing the connectivity pattern between spatial units and ρ is the scalar parameter reflecting the strength of spatial dependence. Similarly, the term $WX\theta$ describes the spatial dependency over the independent features, and the $K \times 1$ vector θ reflects its associated parameters. Finally, the $N \times 1$

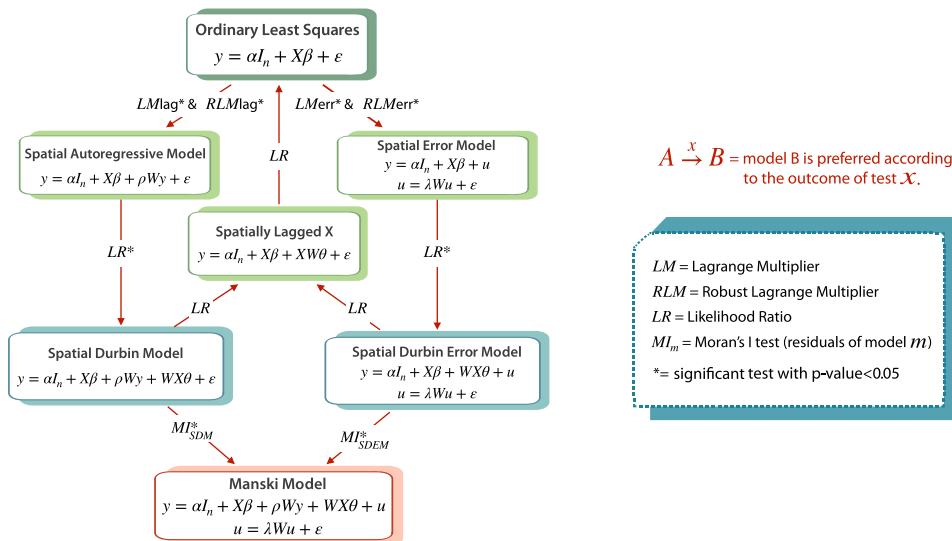


Fig. 5 Schematic representation of the model selection pipeline. The pipeline is based on the approach proposed by ref. 40, but it extends it by also evaluating the Spatially Lagged X and the Manski models. The pipeline relies on the LM, the LR and the MI to identify an adequate spatial regression model for each regression. The pipeline should be applied by starting from the OLS model and only progressing to the next model if the test meets the significance condition indicated on the arrow. The preferred model (or models) is identified when it is not possible to advance further in the pipeline.

disturbance vector ϵ contains independent, normally distributed terms with a vector mean zero ($N \times 1$), and constant variance, (σ^2)¹⁹. The SDM model is regarded as a global spatial dependency model, as the presence of the spatial lag term (ρWy) causes that changes applied in a specific location to propagate through the whole W matrix affecting the N -order neighbours recursively^{19,53}.

The spatial influence between samples (villages in this case) was defined through an adjacency matrix (A_{ij}) created using a queen contiguity criteria⁵³, in which villages are only considered neighbours if they share a common point along their borders. A_{ij} was row-normalised to build the weight matrix (W), defining the strength of the association between villages¹⁹. The spatial weights matrix was decomposed into 30 power of traces using Monte Carlo Markov Chain simulations to reduce the computational burden of the simulations¹⁹. Both models were implemented in R 4.2.1 using the package *spatialreg* (version 1.2-1)⁵⁶. Sparse Cholesky decomposition was used to compute the Jacobian of the W matrix (method 'Matrix' in the same package)⁵⁶.

Spatial impacts. The marginal contribution of each independent feature (x_k) in spatial regression models is described by an $N \times N$ matrix (where N is the number of samples), describing the expected change in the y of all locations when x_k is modified. To ease the interpretation of the results, the marginal contribution matrices can be transformed into the direct, indirect and total spatial impacts¹⁹. We report the average of the diagonal elements in the matrix as the direct impact, and it can be interpreted as the average change in y in a location when their own x_k is increased by 1 unit. The average for the off-diagonal effects can be interpreted as the influence of a change on their neighbouring locations, and it is reported as the indirect impact (or spatial spillover effect). Finally, summing both effects we obtain the total spatial impact and it describes the average influence of modifying x_k across all spatial units^{19,40}. We estimated the significance of the spatial impacts numerically using the regression coefficients and the covariance matrix to perform 10,000 estimations of the impact measures¹⁹. For our analysis, we only considered the features with statistically significant direct, indirect and total impacts (p -value ≤ 0.05) as significant associations.

We would like to remark that the associations identified by the spatial models should not be interpreted as causal outcomes but as the spatially averaged contribution of each feature x to estimating the dependent variable y .

Prediction. In-sample prediction in spatial regression models allows estimating the expected value of y (\hat{y}) when the independent variables are modified in the same locations where the model was trained. There are two possible predictors for spatial regression models the “trend” and the “trend-corrected” predictors, refer to ref. 57 for more detailed definitions. In models with spatial auto-correlation (ρWy), the trend-corrected predictor provides more accurate estimations than the “trend” model. Thereby, the counterfactual analysis for the SDM model was performed using the “trend-corrected” predictor while for the SDEM model the “trend” predictor was used instead.

Quality metrics. To validate the goodness of fit of the models, we evaluated several statistical metrics. First, we report two different pseudo- R^2 metrics, a pseudo- R^2 adjusted using Ezekiel’s correction⁵⁸ and the Nagelkerke’s (NK) pseudo- R^2 ⁵⁹. Pseudo- R^2 cannot be interpreted as an OLS R^2 , they can only be used as a model selection metric to compare models trained with the same dependent and independent variables. Other metrics for model selection are the Akaike and Bayesian Information Criteria (AIC and BIC respectively)^{60,61}. We also evaluated the model’s accuracy in describing the spatial dependency in the data. To this end, we estimated the MI ⁴⁹ on the residuals (ϵ) of the regressions. If the spatial regression model accurately characterises the spatial dependency between samples, we expect that the residuals are spatially uncorrelated and have a non-significant (p -value > 0.05) MI ⁴⁹. Finally, we relied on the Normalised Root Mean Squared Error (NRMSE) as a goodness-of-fit metric. The prediction error was estimated using the difference between the real value for the dependent variable (y) and the in-sample predicted variable (\hat{y}), using the “trend-corrected” predictor for the SDM model and the “trend” estimator for the SDEM⁵⁷. This measure was transformed

into the Root Mean Squared Error by

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}, \quad (2)$$

where N is the number of villages and $(y_i - \hat{y}_i)$ is the prediction error described before⁶². The RMSE was normalised by dividing it by the range of the dependent variable y ⁶³, as shown in

$$NRMSE = \frac{RMSE}{\max(y) - \min(y)}. \quad (3)$$

Sustainability of different production profiles. For this analysis, we aggregated the palm oil production data into four independent production profiles: active smallholder plantations, active industrial plantations, unproductive (damaged) plantations and unproductive (but possibly replanted) plantations. We considered an “active plantation” the sum of young and mature plantations for that production profile. Contrarily, the two “unproductive land” profiles were defined respectively as the sum of industrial and smallholder land for each class. Note that the four production profiles are considered independent because their sum does not reflect the totality of the village area, only the area dedicated to palm oil production.

We estimated the socio-economic characteristics significantly associated with each production profile by training four SDM models with the data for each production profile as the dependent variable y and all the socio-economic variables from PODES as independent features X . The complete list of exogenous variables used can be found in the Supplementary Methods. For this analysis, the dummy variables describing the “village’s main source of income” was excluded to avoid colinearity between villages reporting palm oil as primary income and villages with a high palm oil coverage.

The significant associations identified for the four regression modes were reported using the total spatial impacts. Additionally, several statistical quality metrics were reported for each model trained (refer to the Quality metrics section).

Counterfactual analysis. The counterfactual analysis was implemented following the structure described in Fig. 2. Seven regressions (SDM model) were trained using each SDG indicator as the dependent variable and the socio-economic and palm oil production features as the dependent variables (see Supplementary Methods for more details). To avoid perfect colinearity, when estimating SDG 15, variables describing village location, forest use and forest dependency were excluded from the analysis.

The counterfactual scenario proposed evaluates the influence on the SDGs of replanting all damaged plantations through a smallholder or an industrial profile. Thus, the percentage of land dedicated to unproductive (damaged) plantations (industrial + smallholder) was transferred to the young smallholder or young industrial categories, respectively.

The effect of the counterfactual was quantified by computing the difference between the dependent variable \hat{y} predicted after the intervention with the value predicted when the intervention is not applied. Before estimating the difference between both scenarios, the predictions obtained for each SDG indicator were bounded to reflect the characteristics of each type of variable. For all SDG indicators, the predictions were lower-bounded to zero, and only the percentage of forest coverage (indicator for SDG 15) was upper-bounded to 100%.

The effects produced by the counterfactual analysis were contextualised by determining their influence on the well-being of the villages affected. For each SDG indicator, a deprivation threshold was defined according to the values indicated by Village Development goals (*Indek Pembangunan Desa*)²². These values

define the minimum levels required to consider that the village is not deprived of that socio-economic characteristic. The following list describes the thresholds selected.

- SKTM poverty letters: villages with more than 10% of the population with SKTM poverty letters are considered deprived of this indicator;
- Number of healthcare facilities: villages without healthcare facilities are considered deprived of this indicator.
- Distance to primary school: villages without a primary school (distance to nearest primary school > 0 km) are considered deprived of this indicator;
- Distance to junior-high school: villages with more than 3 km to the closest junior-high school are considered deprived of this indicator;
- Distance to market: villages with more than 10 km to the nearest market are considered deprived of this indicator;
- Percentage forest coverage: villages without forest coverage (0% of the village covered by forest) are considered deprived of this indicator. Note that the Multidimensional Poverty Index does not consider Forest Coverage as an indicator; thus, the threshold was defined as strictly as possible;
- Number of different types of crime: Villages with more than 3 different types of crime are considered deprived of this indicator.

These thresholds were used to estimate the deprivation status for each village and SDG before and after the intervention of the counterfactual. Interventions causing a village to change from a deprived to a non-deprived state are classified as highly positive, while interventions causing the transition from a non-deprived to a deprived status are labelled as highly negative. Contrarily, interventions that have some impact on the SDG indicator but do not induce a transition between deprivation states are only regarded as moderately positive or negative interventions, as they modify the SDG indicator, but this is not translated into changes in village well-being. Finally, in villages where replanting damaged plantations did not change the SDG indicators, the intervention is considered to have a neutral effect.

Province analysis. For this analysis, Sumatra was divided into the ten provinces shown in Supplementary Fig. 1. The counterfactual analysis was repeated using the same trained models as for the previous analysis but assuming that only the unproductive plantations in the region evaluated are being replanted. The impact on village well-being of the regional interventions was evaluated using the thresholds described in the previous section.

The results for the counterfactual analysis in ten provinces are shown in Supplementary Figs. 3–12. However, we summarised them in a table describing the most beneficial and detrimental profile for each region (Table 1). The beneficial production profile for each region was estimated by selecting the interventions with the highest amount of villages with positive outcomes (highly pos. + moderate pos.) effects for each SDG indicator. The most negative outcome was selected accordingly. The production profile with the highest magnitude (positive/negative) is reported using the following abbreviations: “I” if industrial plantations or “S” for smallholder plantations, “I/S” if the percentage of positive and negative outcomes is equivalent for both production profiles, and “–” if no outcomes are observed.

Data availability

The palm oil plantation data was obtained by combining the datasets refs. 17 and 16, which can be freely accessed via the following databases refs. 41 and 42. The geometries used to generate the maps were extracted from the following sources: Indonesia

(obtained with the preprocessed socio-economic data), Malaysia⁶⁴, Thailand⁶⁵ and Singapore⁶⁶. A data usage license for the PODES data can be requested and purchased from the Indonesian Bureau of Statistics via <https://pst.bps.go.id>. Because of this, it is not possible to upload a minimal dataset on a public repository. However, we have uploaded the output from the trained regression models to the following public GitHub repository [AFosch/Figs_PalmOil_SDGs](https://github.com/AFosch/Figs_PalmOil_SDGs). This data can be used to obtain some of the figures shown in the manuscript.

Code availability

The code to generate the figures can be found in the following GitHub repository [AFosch/Figs_PalmOil_SDGs](https://github.com/AFosch/Figs_PalmOil_SDGs). The code used to train the regression models is available upon request.

Received: 7 March 2023; Accepted: 2 October 2023;

Published online: 16 October 2023

References

1. Cf. ODDS. *Transforming Our World: The 2030 Agenda For Sustainable Development* (United Nations, 2015).
2. Meijaard, E. & Sheil, D. The moral minefield of ethical oil palm and sustainable development. *Front. For. Glob. Change* <https://doi.org/10.3389/ffgc.2019.00022> (2019).
3. Carlson, K. M. et al. Effect of oil palm sustainability certification on deforestation and fire in indonesia. *Proc. Natl Acad. Sci. USA* **115**, 121–126 (2018).
4. Meijaard, E. et al. The environmental impacts of palm oil in context. *Nat. Plants* **6**, 1418–1426 (2020).
5. Jackson, T. A. Learning to love the world's most hated crop -review articles. *J. Oil Palm Res.* <https://doi.org/10.21894/jopr.2019.0046> (2019).
6. Ng, I. et al. The impacts of logging and palm oil on aquatic ecosystems and freshwater sources in southeast asia. *EnviroLab Asia* **1**, 3 (2017).
7. Santika, T. et al. Impact of palm oil sustainability certification on village well-being and poverty in indonesia. *Nat. Sustain.* **4**, 109–119 (2021).
8. Hasudungan, A. & Neilson, J. Process of land appropriation for large-scale oil palm development in west kalimantan, indonesia/ processos de aquisição de terra para a produção de óleo de palma em kapuas hulu, indonésia/proses pembebasan lahan untuk perkebunan kelapa sawit di kapuas hulu, kalimantan barat. *Rev. Nera* <https://doi.org/10.47946/rnera.v0i51.6542> (2020).
9. Lee, J. S. H., Miteva, D. A., Carlson, K. M., Heilmayr, R. & Saif, O. Does oil palm certification create trade-offs between environment and development in indonesia? *Environ. Res. Lett.* **15**, 124064 (2020).
10. Byerlee, D., Falcon, W. P. & Naylor, R. *The Tropical Oil Crop Revolution: Food, Feed, Fuel, And Forests* (Oxford University Press, 2017).
11. Santika, T. et al. Does oil palm agriculture help alleviate poverty? a multidimensional counterfactual assessment of oil palm development in indonesia. *World Dev.* **120**, 105–117 (2019).
12. Meijaard, E. et al. *Oil Palm And Biodiversity: A Situation Analysis By The Iucn Oil Palm Task Force*. Tech. Rep. (International Union for Conservation of Nature and Natural Resources (IUCN), 2018).
13. Pertanian, K. Analisis PDB Sektor Pertanian Tahun 2015. *Jakarta, Pusat Data dan Sistem Informasi Sekretariat Jenderal-Kementerian Pertanian Indonesia* (2015).
14. Pertanian, K. *Rencana Strategis Kementerian Pertanian Tahun 2015-2019*. (Kementerian Pertanian, 2015).
15. Miettinen, J. et al. Historical analysis and projection of oil palm plantation expansion on peatland in southeast asia. *ICCT White Paper* **17** (2012).
16. Gaveau, D. L. et al. Slowing deforestation in indonesia follows declining oil palm expansion and lower oil prices. *PLoS ONE* **17**, e0266178 (2022).
17. Descals, A. et al. High-resolution global map of smallholder and industrial closed-canopy oil palm plantations. *Earth Syst. Sci. Data Discuss.* **13**, 1–22 (2020).
18. Lee, J. S. H. et al. Environmental impacts of large-scale oil palm enterprises exceed that of smallholdings in indonesia: forest loss from sumatra's oil palm industry. *Conserv. Lett.* **7**, 25–33 (2014).
19. LeSage, J. & Pace, R. K. *An Introduction To Spatial Econometrics* (Chapman and Hall/CRC, 2009).
20. Krishna, V. V. & Kubitzka, C. Impact of oil palm expansion on the provision of private and community goods in rural indonesia. *Ecol. Econ.* **179**, 106829 (2021).
21. Government of Indonesia. REDD+: National strategy. (*REDD+Task Force, Jakarta, Indonesia*) (2012).
22. Badan Pusat Statistik. *Indeks pembangunan desa 2018*. Tech. Rep. (BPS - Statistics Indonesia, 2018).
23. Koh, L. P. & Wilcove, D. S. Cashing in palm oil for conservation. *Nature* **448**, 993–994 (2007).
24. Ogahara, Z., Jespersen, K., Theilade, I. & Nielsen, M. R. Review of smallholder palm oil sustainability reveals limited positive impacts and identifies key implementation and knowledge gaps. *Land Use Policy* **120**, 106258 (2022).
25. Smit, H. H. et al. Breaking the link between environmental degradation and oil palm expansion: a method for enabling sustainable oil palm expansion. *PLoS ONE* **8**, e68610 (2013).
26. Zhao, J. et al. Spatial patterns and drivers of smallholder oil palm expansion within peat swamp forests of riau, indonesia. *Environ. Res. Lett.* **17**, 044015 (2022).
27. Schoneveld, G. C., Ekowati, D., Andrianto, A. & Van Der Haar, S. Modeling peat-and forestland conversion by oil palm smallholders in indonesian borneo. *Environ. Res. Lett.* **14**, 014006 (2019).
28. Jelsma, I., Woittiez, L. S., Ollivier, J. & Dharmawan, A. H. Do wealthy farmers implement better agricultural practices? an assessment of implementation of good agricultural practices among different types of independent oil palm smallholders in riau, indonesia. *Agric.Syst.* **170**, 63–76 (2019).
29. Gaveau, D. L. et al. Major atmospheric emissions from peat fires in southeast asia during non-drought years: evidence from the 2013 sumatran fires. *Sci. Rep.* **4**, 1–7 (2014).
30. Abram, N. K. et al. Synergies for improving oil palm production and forest conservation in floodplain landscapes. *PLoS ONE* **9**, e95388 (2014).
31. McCarthy, J. F., Vel, J. A. & Afiff, S. Trajectories of land acquisition and enclosure: development schemes, virtual land grabs, and green acquisitions in indonesia's outer islands. *J. Peasant Stud.* **39**, 521–549 (2012).
32. Goh, C. S., Junginger, M., Potter, L., Faaij, A. & Wicke, B. Identifying key factors for mobilising under-utilised low carbon land resources: a case study on kalimantan. *Land Use Policy* **70**, 198–211 (2018).
33. Cronin, A. A. et al. Piloting water quality testing coupled with a national socioeconomic survey in yogyakarta province, indonesia, towards tracking of sustainable development goal 6. *Int. J. Hygiene Environ. Health* **220**, 1141–1151 (2017).
34. Contribution of AR6 Synthesis Report Core Writing Team, Working Group I, Working Group II, and Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Synthesis Report Of The IPCC Sixth Assessment Report (ar6)*. Tech. Rep. (Cambridge University Press, 2023).
35. Hein, L., Sumarga, E., Quiñones, M. & Suwarno, A. Effects of soil subsidence on plantation agriculture in indonesian peatlands. *Reg. Environ. Change* **22**, 121 (2022).
36. Trancoso, R. et al. Converting tropical forests to agriculture increases fire risk by fourfold. *Environ. Res. Lett.* **17**, 104019 (2022).
37. United Nations. *SDG Global Database*. <https://unstats.un.org/sdgs/dataportal> (2022). Last accessed: 28 September 23.
38. Descals, A. et al. Oil palm (*Elaeis guineensis*) mapping with details: smallholder versus industrial plantations and their extent in Riau, Sumatra. *Remote Sens.* **11**, 2590 (2019).
39. Santika, T. et al. Heterogeneous impacts of community forestry on forest conservation and poverty alleviation: evidence from indonesia. *People Nat.* **1**, 204–219 (2019).
40. Elhorst, J. P. Applied spatial econometrics: raising the bar. *Spat. Econ. Anal.* **5**, 9–28 (2010).
41. Descals, A. *High Resolution Global Industrial And Smallholder Oil Palm Map For 2019*. <https://doi.org/10.5281/zenodo.4473715> (2021). Last accessed: 28 September 23.
42. Gaveau, D. *Nusantara Atlas: Analyse Land Cover Change*. <https://nusantara-atlas.org> (2023). Last accessed: 28 September 23.
43. Margono, B. A., Potapov, P. V., Turubanova, S., Stolle, F. & Hansen, M. C. Primary forest cover loss in indonesia over 2000–2012. *Nat. Clim. Change* **4**, 730–735 (2014).
44. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
45. Assembly, U. G. Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development (United Nations Statistics Division: New York, NY, USA, 2017).
46. Morgans, C. L. et al. Evaluating the effectiveness of palm oil certification in delivering multiple sustainability objectives. *Environ. Res. Lett.* **13**, 064032 (2018).
47. Jagger, P. & Rana, P. Using publicly available social and spatial data to evaluate progress on REDD+ social safeguards in Indonesia. *Environ. Sci. Policy* **76**, 59–69 (2017).
48. Lantz, B. *Machine Learning With R: Expert Techniques For Predictive Modeling* (Packt publishing Ltd, 2019).
49. Moran, P. A. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
50. Pan, X., Guo, S., Li, M. & Song, J. The effect of technology infrastructure investment on technological innovation - a study based on spatial durbin model. *Technovation* **107**, 102315 (2021).

51. Takagi, D. & Shimada, T. A spatial regression analysis on the effect of neighborhood-level trust on cooperative behaviors: comparison with a multilevel regression analysis. *Front. Psychol.* **10**, 2799 (2019).
52. Jiang, L., Folmer, H., Ji, M. & Tang, J. Energy efficiency in the chinese provinces: a fixed effects stochastic frontier spatial durbin error panel analysis. *Ann. Reg. Sci.* **58**, 301–319 (2017).
53. Anselin, L. *Spatial Econometrics: Methods And Models*. Vol. 4 (Springer Science & Business Media, 1988).
54. Anselin, L., Bera, A. K., Florax, R. & Yoon, M. J. Simple diagnostic tests for spatial dependence. *Reg. Sci. Urban Econ.* **26**, 77–104 (1996).
55. Manski, C. F. Identification of endogenous social effects: the reflection problem. *Rev. Econ. Stud.* **60**, 531–542 (1993).
56. Bivand, R., Millo, G. & Piras, G. A review of software for spatial econometrics in R. *Mathematics* <https://doi.org/10.3390/math9111276> (2021).
57. Goulard, M., Laurent, T. & Thomas-Agnan, C. About predictions in spatial autoregressive models: optimal and almost optimal strategies. *Spat. Econ. Anal.* **12**, 304–325 (2017).
58. Ezekiel, M. *Methods of Correlation Analysis* (Wiley, 1930).
59. Nagelkerke, N. J. et al. A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692 (1991).
60. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
61. Wright, B. D. & Stone, M. H. *Best Test Design* (Mesa press, 1979).
62. Woodruff, S., Vitro, K. A. & BenDor, T. K. *Comprehensive Geographic Information Systems*. In Huang, B. (ed.), p. 236–257. <https://doi.org/10.1016/B978-0-12-409548-9.09655-X> (Elsevier, 2018).
63. Otto, S. A., Kadin, M., Casini, M., Torres, M. A. & Blenckner, T. A quantitative framework for selecting and validating food web indicators. *Ecol. Indic.* **84**, 619–631 (2018).
64. Princeton Library. Boundary, Malaysia, <https://maps.princeton.edu/catalog/stanford-bd012vf3991> (2015). Last accessed: 28 September 23.
65. Humanitarian Data Exchange. Thailand - Subnational Administrative Boundaries. <https://data.humdata.org/dataset/cod-ab-tha?> (2019). Last accessed: 28 September 23.
66. Berkeley Library. Boundary, Singapore. https://geodata.lib.berkeley.edu/?f%5Bdc_format_s%5D%5B%5D=Shapefile&f%5Bdc_spatial_sm%5D%5B%5D=Singapore (2015). Last accessed: 28 September 23.
67. United Nations Department of Global Communications. *Sustainable Development Goals: Guidelines For The Use Of the sdg Logo Including The Colour Wheel, And 17 Icons* (2019).

Acknowledgements

This paper was developed with the financial support of Soremartec SA and Soremartec Italia, Ferrero Group, in the frame of the Sustainable Nutrition Scientific Board. MJS and CLM were funded by a Leverhulme Trust Research Leadership Award. A.A. acknowledges support through the grant RYC2021-033226-I funded by MCIN/AEI/10.13039/501100011033 and the European Union “NextGenerationEU/PRTR”. Y.M. was partially supported by the Government of Aragon, Spain and “ERDF A way of making Europe” through grant E36-23R (FENOL), and by Ministerio de Ciencia e Innovación, Agencia Española de Investigación (MCIN/AEI/10.13039/501100011033) Grant No. PID2020-115800GB-I00. We acknowledge the use of the computational resources of COSNET Lab

at Institute BIFI, funded by Banco Santander and by the Government of Aragón. The authors would like to acknowledge Marco Pangallo for his insightful comments.

Author contributions

A.F.: conceptualisation, software, formal analysis, writing - original draft, writing - review and editing and visualisation. G.F.d.A.: software, conceptualisation, formal analysis, and writing - review and editing. A.A.: conceptualisation, writing - review and editing. A.D.: data curation, writing - review and editing. D.G.: data curation, writing - review and editing. C.M.: data curation, writing - review and editing. T.S.: data curation, writing - review and editing. M.J.S.: data curation, writing - review and editing. E.M.: conceptualisation, writing - original draft, writing - review and editing. Y.M.: conceptualisation, writing - review and editing, supervision, project administration, and funding acquisition.

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/s43247-023-01037-4>.

Correspondence and requests for materials should be addressed to Ariadna Fosch, Erik Meijaard or Yamir Moreno.

Peer review information *Communications Earth & Environment* thanks Maja Slingerland and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Fiona Tang, Joe Aslin and Aliénor Lavergne. A peer review file is available.

Reprints and permission information is available at <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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023