Open Dumps and the Global Trade in Garbage

Matthew Gordon and Anna Papp*

Click Here for Latest Version

Abstract

Globally, a large fraction of solid waste is disposed in open dumps, generating harmful externalities. Research on this topic is limited by a lack of data, however. We address this gap by developing methods to construct a globally representative time-series of open dumps, starting with a small, unrepresentative sample of crowdsourced observations. We use these observations to train a potentially biased machine learning model that can predict dumps using satellite imagery. We then employ active learning techniques to select and verify an approximately optimal representative subset of predictions. This method gives unbiased estimates of dump prevalence and results in significant efficiency gains relative to standard methods, even in the presence of model miscalibration. We use our data to study whether internationally traded waste ends up in open dumps. Our results show a dramatic increase in open-air landfills globally after China banned imports of plastic waste in 2018. The increase is concentrated in low-income countries that saw increased imports, suggesting that trade plays a role in overwhelming local waste management systems and the subsequent leakage of waste into the environment.

JEL: C45, C81, Q53, O17

Keywords: Solid Waste, Satellite Data, Machine Learning, Trade

*Gordon: Paris School of Economics (matthew.gordon@psemail.eu); Papp: MIT (annapapp@mit.edu). We are grateful for excellent research assistance from Marion Chadal, Mark Viti, and Francesca Cruz. Jim Puckett of the Basel Action Network facilitated data collection efforts, and Caleb Kruse of Earthrise shared data and expertise with us. We are grateful for feedback on this project from Marian Chertow, Eli Fenichel, Luke Sanford, Marten Ovaere, Pierre Biscaye, Bastien Michel, and seminar participants at CREST, MIT, the Tweeds Workshop, the Interdisciplinary PhD Workshop in Sustainable Development, and the SEA Annual Conference. Many individuals and organizations contributed to the Atlas of Plastic Waste - a complete list is available from: https://github.com/atlas-of-plastic-waste/atlas-of-plastic-waste. This document is an output from the research initiative 'Structural Transformation and Economic Growth' (STEG), a programme funded by the Foreign, Commonwealth & Development Office (FCDO), contract reference STEG_LOA_704_Gordon. The views expressed are not necessarily those of FCDO. We are grateful for additional support from the Minderoo Foundation, the International Growth Centre, and the Agence Nationale de la Recherche.

1 Introduction

What are the impacts of the global trade in waste? Since Larry Summers argued, "the economic logic behind dumping a load of toxic waste in the lowest wage country is impeccable" in a controversial 1991 World Bank memo, cross-country flows of plastic scrap have increased seven-fold. Despite Summers' assertion, economic theory does not give an unambiguous prediction as to the welfare impacts of this trade. This will depend on the magnitude of the externalities generated by waste imports relative to the private benefits (Chichilnisky 1994; Copeland et al. 2021). Yet studying these tradeoffs has been difficult, as we lack the most basic descriptive facts about what fraction of waste imports are recycled, and what fraction are discarded in environmentally harmful ways.

An estimated 90% of domestic waste is disposed in open dumps or burned in low-income countries. Open dumps have substantial environmental and social implications. The open burning of waste releases pollutants dangerous to human health, and this practice may be a significant contributor to poor air quality in developing countries.[1] Plastic leakage to marine ecosystems is also widely documented, with possible adverse effects on the productivity and sustainability of ecosystem services such as fishing, recreation, and pest control (Rochman et al. 2016). However, the scale and incidence of these externalities remains understudied due to the lack of data on open dumping. Existing research attempts to study mismanaged waste using rough assumptions about per capita waste generation and improper disposal (e.g., Jambeck et al. 2015; Meijer et al. 2021). These modeled estimates make it difficult to study causal relationships between changes in policies, waste management, and environmental and social outcomes.[2]

In this paper, we devise a method to collect the data required to begin to study the scale and impacts of open dumping. Our pipeline consists of three steps: first we collect 'seed' data: non-representative crowdsourced examples of open dumps. We then use these examples to train a potentially biased machine learning model that predicts dumps using satellite data. Finally, we use the model's predictions to select and verify an approximately optimally chosen subsample of predictions. These verified predictions form the basis of a globally representative, time-series dataset that allows us to precisely estimate the prevalence of open dumps across regions and over time. Our sampling method is effective and results in unbiased estimates of descriptive statistics and the causal effects of policies, even though our

[1] Examples of these pollutants include black carbon, a major contributor to particulate matter, and polycyclic aromatic hydrocarbons (Pathak et al. 2023).

[2] Law et al. (2020) is one of the few papers to try to estimate the contribution of trade to mismanaged waste; however, they simply assume that 25-75% of imports are improperly managed in countries with high levels of assumed mismanagement.

machine learning model's predictions may be biased due to the non-representative training data. We derive simple formulas to characterize the efficiency gains of our optimal sampling method in the presence of model miscalibration.

We then use our data to generate a set of novel descriptive facts about the global prevalence of open dumps. We document an environmental Kuznets curve — open dumps are rare in very low-income countries, peak in middle income countries, and decline slightly from the peak in high-income countries. We also document that dumps are disproportionately found in areas with high population density in low and middle-income countries, whereas in high-income countries dumps tend to be in slightly less dense areas. We also find that fires in open dumps are common: 10% of dumps in low and middle-income countries experience a fire at least once a year.

Next, we use our data to investigate whether the international trade in waste has a causal effect on open dumping. Our empirical strategy exploits a shock to the global trade network caused by a sudden ban in Chinese waste imports in 2018. China previously accounted for 50-90% of global imports of waste, depending on the category. After the ban, shippers diverted a significant fraction of this waste to countries across Southeast Asia and the rest of the world. We use event-study and difference-in-difference regressions to study this shock, finding that a 10% increase in net-imports of waste increases the prevalence of open dumps by 1.1-4.5%. These effects are concentrated in low-income countries that are net importers. Exporting countries see no increase in open dumps following a decrease in exports, suggesting that these countries have access to alternative disposal technologies, such as recycling and sanitary landfills.

Our data collection approach contributes to the literature on using machine learning and satellite data for policy evaluation (see Donaldson and Storeygard 2016 and Jain 2020 for reviews). An emerging literature has begun to document the potential biases that can result from using machine learned proxies for economic or environmental variables in causal inference. Several recent papers propose methods that can correct for non-classical measurement error using a sample of observations with 'ground-truth' data, where true labels are known with certainty (Fowlie et al. 2019; Alix-Garcia and Millimet 2023; Ratledge et al. 2021; Proctor et al. 2023; Angelopoulos et al. 2023; Rambachan et al. 2025; Sanford et al. 2025; Kluger et al. 2025; Carlson and Dell 2025).

These approaches take the set of labeled data as given, and rely on the assumption that either the labeled set is a representative sample of the broader population, or that the labeling process can be modeled as conditional on observables. Our approach is fundamentally

different in that we study how to optimally choose which data to label when starting with selected non-representative data. Our resulting procedure makes estimates unbiased and maximizes the precision of the resulting estimates for a given sampling budget. These methods are widely applicable to other hard-to-study land uses, and more generally, the can be used in any setting where data collection is expensive and true positives are rare in the population of interest, but can be predicted with reasonable accuracy.

The first step of our pipeline is collecting non-representative 'seed' data. For this, we rely on crowdsourcing. We collaborated with an international network of NGOs, researchers, and activists to solicit an initial set of open-air waste sites. This initiative, which we called "The Atlas of Plastic Waste", has yielded over 270 submissions across 24 countries. We supplement our crowdsourced data collection effort with existing datasets on open-air waste sites, which are highly constrained in either geographic scope or temporal coverage. We verify the extent and location of all sites using high-resolution satellite imagery.

In the second step, we train a machine learning model on these examples that can identify open dumps using radar and multi-spectral satellite imagery. These methods build on recent work using satellite imagery to identify plastic waste in more limited temporal and geographic settings (Kruse et al. 2023). We train an XGBoost model that differentiates open dumps from other land uses with high accuracy. The area under the receiver-operator curve (AUC) in our validation sample is 0.96 out of 1, and balanced accuracy is 90%, demonstrating that our model can find a distinctive spectral signature associated with open dumps in the seed data.

Since our training data is heavily selected, and dumps are a rare land use class, even small biases in our model's predictions when deployed out-of-sample could dramatically impact inferences drawn about the prevalence of open dumps. Thus, rather than use our model's predictions directly to generate statistics, we use them to inform an approximately optimal sampling and verification procedure that aims to minimize the variance of estimates of dump prevalence and change over time. Our procedure relies on Neyman (1934) optimal sampling methods, where we oversample the predictions that our model is most uncertain about. For sampled observations, we verify the existence of dumps using high-resolution satellite images. This gives us globally representative, time-series estimates of open dump prevalence.

These sampling techniques are common in the active sampling literature in computer science and remote sensing where the goal is to maximize model accuracy (Tuia et al. 2009; Stehman 2012; Zrnic and Candès 2024), but they have not been widely applied to study policy questions in the social sciences. We derive simple formulas to characterize the efficiency of

our method, relative to uniform random sampling, in the presence of model miscalibration. Efficiency gains depend on the distribution of uncertainty in the data, relative to the distribution of miscalibration. We find that, when estimating changes in open dumps over time, a perfectly calibrated model could have given a sample with a variance 96% lower than a uniform random sample of the same size. Our ex-post realized efficiency gains, accounting for model miscalibration are 87% relative to uniform random sampling.

In addition to the data collection and descriptive statistics, we contribute to long standing debates in economics about the efficiency and distributional implications of trade in the presence of environmental externalities. We show clear evidence of the pollution havens hypothesis – the idea that environmental regulations in one country can result in the migration of polluting industries to less-regulated jurisdictions (see Copeland et al. 2022 for a recent review of this hypothesis and associated literature).

Prior work has demonstrated that industry location is influenced by environmental regulations in the context of lead-acid battery recycling (Tanaka et al. 2022), and the Clean Air Act in the US (Hanna and Oliva 2015). A few studies have examined the pollution havens hypothesis using country-level data on the trade in plastic scrap. Kellenberg (2015) shows that the relative stringency of environmental regulations is associated with trade in waste - with materials flowing from jurisdictions with stricter regulations to more unregulated locales. Thakur (2022) uses a gravity model to show how the China waste ban increased waste flows to low-income countries. With the exception of Tanaka et al. (2022), however, none of these studies can directly measure changes in externalities associated with these trade flows or location decisions. Our work is the first to provide empirical evidence for the pollution havens hypothesis on a global scale.

Our empirical strategy relies on an unanticipated policy shock — the 2018 China National Sword policy — more commonly known as the China waste ban, which banned or restricted the imports of 24 categories of materials, including plastic scrap, unsorted paper, and some types of metal and electronic wastes. Aggregate trade statistics show that while total global waste trade flows fell, several low and middle-income countries saw sharp increases in waste imports after the ban. This natural experiment allows us to identify the effects of changes in flows of waste traded across borders on the prevalence of open dumps. While some previous work has studied the impacts of the China waste ban on air pollution in China and landfilled waste in the US (Guo et al. 2023; Sigman and Strow 2024), our novel data allows us to study an understudied outcome at a global scale.

We find that the global prevalence of open dumps increases by 50% in the year of the China

waste ban, despite an overall decrease in trade flows. This suggests that either a large quantity of imported waste is discarded in environmentally harmful ways, or that waste imports overwhelm local waste management and recycling facilities. In either case, these trade flows contribute to the leakage of waste into the natural environment and may pose a threat to human health and ecological functioning. Event study and difference-in-difference analyses estimate an elasticity of waste imports to open dumps of .11 to .45, depending on the specification. Heterogeneity analyses show that the increase in dumps is highest in low-income countries, where fires are more common and dumps tend to be found in areas with higher population density. We also find that the increase in dumps is concentrated in countries that see their imports rise. In contrast, in countries where exports fall, we see no change in open dumps, likely because the alternatives to exporting waste are recycling and sanitary landfill disposal, rather than open dumping.

The rest of the paper proceeds as follows – in Section 2, we describe relevant background, including existing research on open dumps, and the context of the Chinese waste import ban. In Section 3, we describe the data used for our analyses, including our crowdsourced landfills, and the satellite data used for training the machine learning model. Section 4 describes how we build our machine learning model and the approximately optimal active learning procedure to generate our globally representative time-series data. Section 5 describes our descriptive results on global time trends, the environmental Kuznets curve, population density, and fires. Section 6 describes our analysis of the China waste import ban. Finally, Section 7 discusses and concludes.

2 Background: Solid Waste and Trade

Solid waste generation is expected to increase by more than 70% by 2050 as the world continues to get richer and more urbanized (Kaza et al. 2018). This trend will pose challenges for cities in low-income countries, where waste management is already one of the largest line items in local budgets, comparable to expenditures on policing and education (Hoornweg et al. 2013). Yet an estimated 90% of solid waste is disposed in unregulated open-air dumps or burned.

Open dumps generate severe externalities. Runoff from dumps contaminates water and has contributed to the increase in ocean plastics which disrupts ecosystems and fisheries. Dumps attract vermin which spread disease. The decomposition of organic waste generates methane emissions which contribute to global warming, and the open-burning of waste generates air pollution, including particulate matter and black carbon (Pathak et al. 2023).

Existing research has argued that the mismanagement of solid waste is mostly an issue in low-income countries (Jambeck et al. 2015; Lebreton et al. 2017; Kaza et al. 2018). In the absence of data on mismanaged waste, these papers use simple models of waste emissions that depend on population, GDP, and the estimated fraction of mismanaged waste. For example, Meijer et al. (2021), finds that most ocean plastics come from a relatively small number of river systems, mostly located in low-income countries. For the most part, these models ignore waste traded across international borders. In trying to estimate the United States' contribution to ocean plastic pollution, Law et al. (2020) write "no quantitative estimates exist of the proportion of material exported for recycling that is ultimately discarded as waste."

Yet these trade flows are significant. Prior to 2018, around 50% of plastic waste collected for recycling was traded internationally, and China accounted for close to 50% of imports of these flows, including 70% of US exports and 95% of EU exports (Kellenberg 2015). Importers use some of this waste in informal recycling processes, though anecdotally a large share of the imports may be disposed of in environmentally damaging ways (Greenpeace 2021). As the quote above highlights, no good estimates exist on how much of this material is used productively or discarded.

Complaints about contamination levels in this shipments and the externalities associated with processing the waste led to a series of progressively stricter restrictions. In February of 2013, China announced the Green Fence policy, an increase in inspections of shipments of waste. Shipments that exceeded prescribed contamination levels could lead to the revocation of import licenses and the forced return shipment of the container.

In July 2017, the Chinese government announced the National Sword policy, which went into effect in January 2018, allowing very little time for anticipation. The policy banned or restricted imports on 24 categories of scrap material, including plastic and paper waste, and lowered the contamination thresholds on other materials to levels that were generally considered unachievable by waste exporters in the US and Europe. At the same time, the Chinese government cracked down on the domestic informal recycling industry. Guo et al. (2023) find that air pollution went down after the ban in Chinese cities that had a large recycling industry, but due to a lack of data on disposal, we don't know how much of this is due to a decline in productive recycling activities versus the open burning of discarded scraps.

Appendix Figure A2 shows that Chinese imports went to nearly zero almost immediately after the import ban went into effect. The loss of this market disrupted the global recycling

market and led to pileups of scrap in the US and the diversion of trade to other countries in the region. Appendix Figure A2 shows a spike in imports in Indonesia, Malaysia, and Turkey around the time of the ban. Thailand, and the Philippines also saw large increases in imports during this time, though total global trade flows decreased due to the loss of the Chinese market. Much of the material that would have otherwise been exported ended up in landfills in the US and Europe (Sigman and Strow 2024).

In Section 6, we will use this policy change to estimate the elasticity of open dumps to imported waste. This research is important for policy, current international discussions under the Basel Convention have debated changing the classification of plastic waste, which would severely restrict or ban trade across borders.

3 Data

Our goal is to estimate the prevalence of open dumps across countries and over time. We define an open dump as any aggregation of waste covering an area greater than $10m^2$, where waste is directly exposed to the surrounding environment. Several countries have statutory definitions of open dumps that refer to waste disposal sites that are not in accordance with regulations.[3] Since regulations vary across countries, we adopt a definition that is consistent globally and can be measured with our data. To be clear, we measure the surface area of dumps, rather than the volume of waste disposed, as it is difficult to measure the height or density of waste piles using satellite data. Despite this limitation, the surface area of open dumps is an important proxy for environmental impacts for a few reasons. First of all, the surface area of a dump determines how much waste is directly exposed to the environment. Secondly, many informal dumps have low density, without capital intensive construction equipment, it is difficult to pile waste very high or compress it to a significant degree. Still, to the extent that the intensive margin is important, our ability to identify open dumps from the air is a crucial first step in identifying locations for further study.

Our size requirement excludes small and transitory piles of litter that cannot be observed from satellite imagery. We also do not observe buried waste which leads us to exclude most sanitary landfills, but we include the open-air working faces of some sanitary landfills that have similar characteristics as open dumps. We also include discard pits next to recycling facilities, for example the ones that commonly accompany paper mills in Indonesia. We

[3]For example in the Philippines an open dump is defined as, "A disposal area wherein the solid wastes are indiscriminately thrown or disposed of without due planning and consideration for environmental and Health standards" (Congress of the Philippines 2001).

exclude construction debris and automobile junkyards, which both have different characteristics from the dumps in our data, however we will not otherwise be able to distinguish the material composition or the density of the waste, which we leave for future research.

In order to measure dump prevalence, we draw on techniques from machine learning known as active statistical inference (Zrnic and Candès 2024; Hamilton et al. 2025). Intuitively, our method starts with a non-representative sample of crowdsourced landfills. We use these sites to train a machine learning model that can predict the probability of a landfill from satellite data. Since the model's predictions are not very accurate and may be biased, we do not use them directly as measurements. Instead, we use them to inform a optimal sampling strategy (Neyman 1934) that results in a representative dataset of high-quality labeled observations, designed to minimize uncertainty. We use this data to obtain unbiased estimates of landfill prevalence. We start by describing the crowdsourced data and other inputs to our machine learning model.

3.1 Machine Learning Model Inputs

3.1.1 Unrepresentative Landfills

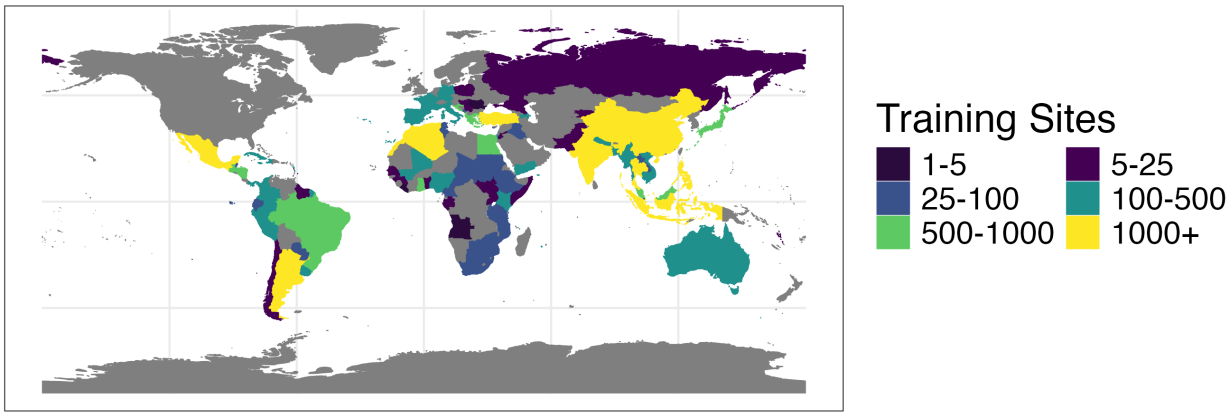
We gather an initial set of verified open-air landfills from various sources, including our crowdsourced data collection effort. We partnered with an international NGO, Basel Action Network (BAN), to create a web portal that allows individuals worldwide to submit the locations of open dump sites and contribute to a training data set[4]. We received more than 270 submissions from dozens of individuals and organizations across 24 countries.

We complement this effort using several other publicly available datasets on landfills and informal dumps that cover particular regions of the world or periods of time (Kruse et al. 2023; D-Waste 2014; Greenpeace 2018; 2021; Climate Trace 2023). In particular, we use nearly 1,000 additional sites from Earthrise's Global Plastic Watch, which uses satellite imagery to find dumps globally, though their data is only verified for 2021[5]. As discussed above, we don't use training data from sanitary landfills in the US or other wealthy countries.

In order to verify the existence of these waste sites and to determine the spatial and temporal extents of the sites, we look up historical high-resolution satellite imagery in Google Earth Pro for each site. Figures 2, 3, and A1 show several examples of open-air waste sites in this

[4]https://www.ban.org/plastic-waste-transparency-project-hub/the-atlas-of-plastic-waste.
[5]Our complete training dataset and list of sources are available from: https://github.com/atlas-of-plastic-waste/atlas-of-plastic-waste.

Figure 1: Distribution of Training Data by Country



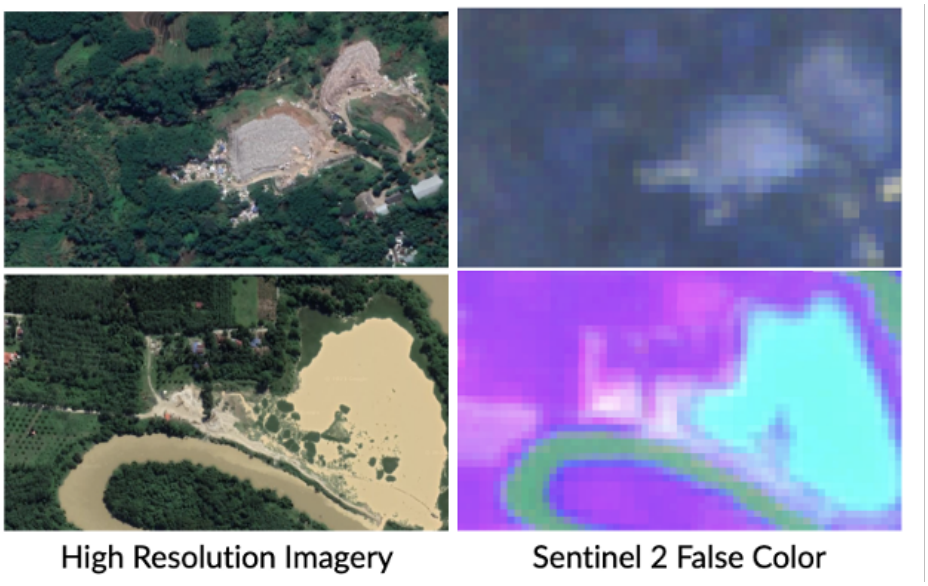
Notes: The distribution of training sites (both waste sites and control sites) by country based on crowdsourced and other data collection.

imagery, which are typically easy to identify. For each site, we look at the first available satellite image between 2017 and 2022, and we draw a polygon around the extent of the site if it exists in that year. At the same time, we also create polygons for nearby areas that are not waste sites but cover other diverse land-use categories (e.g., agricultural land, urban areas, forests).

Given the paucity of initial training data, we followed an iterative process to collect additional data. We would train an experimental model on what data we had on hand, and then use the model to make a set of out of sample predictions. We would then verify some subset of these predictions to assess accuracy, and incorporate new waste and control sites discovered into the next round of model training. These experiments were not always systematic as we tried a variety of modeling and sampling strategies, however it did result in a good amount of valuable training data, especially control sites representing difficult cases that the model would struggle to correctly classify. The most common of these were mines, quarries, and construction sites, where piles of rubble or stones can appear similar to piles of trash.

After several early iterations, we had in total over 36,000 validated site-year observations across 60 countries and over 18,000 control sites. Figure 1 shows the distribution of these training sites worldwide.

Figure 2: Comparison of Google Earth High-Resolution Imagery and Sentinel-2 Imagery



Notes: Figure shows high-resolution historical imagery from Google Earth Pro and Sentinel-2 false-color imagery. The figure demonstrates that high-resolution imagery can generally effectively distinguish between different land uses.

3.1.2 Satellite Data for Sites

Our primary sources of data are Sentinel-2 and Sentinel-1 satellite imagery. The European Space Agency operates both of these satellites. Sentinel-2 is an optical satellite constellation with a spatial resolution of 10m and a revisit frequency of 3-5 days[6]. Sentinel-1 is a radar satellite that provides all-weather and day-and-night coverage. The multi-spectral coverage of Sentinel-2 allows us to capture reflectance in the visible and near-infrared wavelengths, which have been previously used to differentiate between plastics and other materials (Biermann et al. 2020; Kruse et al. 2023; Cózar et al. 2024). Radar imagery of Sentinel-1 collects information on the "texture" and "smoothness" of land cover, which helps differentiate between waste sites and other objects made of similar materials (e.g., plastic roofs).

For each pixel located within our sites, we download satellite data using Google Earth Engine for each year from 2017-2021. We keep cloud-free, non-water pixels across all images.[7] This results in over 7.7 million 10m pixels. We use each of the 13 bands of Sentinel-2, many derived indices from these bands, and all Sentinel-1 bands collected in interferometric wide

[6]We begin our satellite analysis in 2017 when Sentinel 2-B was launched ensuring consistent data coverage.
[7]We use Google's Cloud Score+ and define cloudy pixels as those with a cloud score probability higher than 65% (Pasquarella et al. 2023). We use Dynamic World to mask out water pixels (Brown et al. 2022).

swath mode and derived indices. For a complete list of satellite bands and derived indices, see Appendix Section A.1. For each band and/or derived index, we take the median, mean, minimum, and maximum for each year. We also utilize focal statistics, averaging indices and bands in a 3 km radius around each pixel to include information on nearby areas' shapes and characteristics.

In addition to the satellite data, we use the Dynamic World predictions of major land use categories that Google derives for Sentinel-2 (2022). We also calculate the population density within 5km and the distance to the nearest port for each pixel (Earth Science Data Systems, NASA 2024; National Geospatial-Intelligence Agency 2019).

3.2 Other Data

Here we detail other data sources, not used in training the machine learning model, but in other parts of the analysis. We use the BACI trade database to collect data on the value and mass of waste traded between countries (Gaulier and Zignago 2010). We collect data on products banned by the National Sword Policy, including waste plastics, paper, textiles, and minerals (vanadium slag).[8] BACI reanalyzes UN Comtrade data to attempt to correct for reporter reliability. This is important for our use case, since some waste may be improperly categorized, especially in countries with import restrictions. The data source clearly shows the decline in imports from China after the waste ban, and a corresponding increase in many other countries (see Figure A2).

We also use the NASA VIIRS Active Fire Detection product to analyze fires in open dumps (Schroeder et al. 2025). This data is available at 375m resolution globally and uses satellite based sensors to detect fires. We overlay VIIRS fire detections with our verified dumps to estimate the prevalence of global garbage fires.

4 Dump Detection and Active Learning

4.1 Machine Learning Model

We train an XGBoost model using the training sites and satellite data described above (Chen and Guestrin 2016). XGBoost models are useful for their relative computational efficiency

[8]China banned imports the following: waste minerals/vanadium slag (HS codes 2619 and 2620), waste plastics (HS code 3915), waste paper (HS code 4707), and waste textiles (HS codes 5103, 5104, 5202, 5505, 6310). See, for example, https://www.laregionalagency.us/comprehensive-list-of-banned-materials/.

and their robustness to overfitting, given a limited number of hyperparameters available for tuning compared with neural networks. XGBoost builds an ensemble of tree-based models, similar to random forest, however it builds the trees sequentially using gradient descent, so that each tree tries to correct the errors of the previous trees.

For model development, we split our sites into training, testing, and validation sites. We split our sites into 500 global clusters using a k-means algorithm based on their coordinates. We set aside 20% of our clusters for calculating out-of-sample accuracy metrics, and use the other 80% for choosing hyperparameters. This isolates similar dumps that are geographically close together so that the out-of-sample accuracy metrics we report are not based on the model having seen pixels from within the same dump, or even the same country in many cases.

During training, we assign control pixels a relative weight of 99.6. This reflected a reasonable a priori guess about the relative prevalence of waste pixels we expected to encounter when deploying our model out of the training sample, which we calibrated based on early experiments. We choose our model hyper parameters to maximize the area under the receiver-operator curve (AUC) in the training data, a metric that is robust to heavily unbalanced classes. Our model builds an ensemble of 100 decision trees, limiting tree depth to 25 levels, balancing computational complexity and accuracy in the training data.[9]

Our model has an AUC of .96 out of the maximum possible value of 1 (Appendix Figure A4). Setting a classification threshold that maximizes accuracy, we calculate balanced accuracy at 90%, precision at 92%, and recall at 85%. Figure 3 shows an image of a dump in the high-resolution Google Earth Pro imagery and our associated model predictions for the years 2017 and 2021. The model puts elevated probability on the pixels corresponding to the open faces of the dump. Furthermore the model seems to capture changes in the shape and size of the dump over time, in the first row we see several small patches of trash, and in the second, the dump has been consolidated into a single larger patch, and the model accurately captures these changes.

We analyze which features in our data seem to be most helpful in detecting dumps by ranking them by their sum-of-split scores. This metric calculates the total reduction in the model's

[9]Additional hyperparameters include `min_examples = 27, forest_extraction='DART', l1/l2_regularization = 5.0/3.0, split_axis = 'AXIS_ALIGNED', growing_strategy = 'LOCAL', shrinkage = 0.02, and num_candidate_attributes_ratio = 0.1` using the Tensorflow Decision Forests library in Python. We choose these hyperparameters to maximize AUC, which is a metric that incentives correctly ranking observations by the probability that they are a dump. It may lead to miscalibrated probabilities, especially for unlikely cases, however. We thus apply a sigmoid transformation to the resulting predictions: $p_i = a \log(\tilde{p_i}/(1-\tilde{p_i}))$ where $\tilde{p_i}$ are the original predictions, and we set the coefficient $a = 3.15$ to minimize log loss.

loss contributed by that feature across all trees in the ensemble. Population, distance to the nearest port, and the Dynamic World labels (Brown et al. 2022) are the most important features. The Dynamic World labels are based on Sentinel 2 imagery and are a reasonable summary of the spectral information contained in a given pixel, so it makes sense that they would be important. Beyond that, the Sentinel 1 radar derived indices rank highly, as do the near infrared bands and red-edge, though there is a long tail of additional features that are important. These findings are broadly consistent with previous studies of the spectral signatures of landfills (Biermann et al. 2020; Kruse et al. 2023; Cózar et al. 2024), however given the large number of features that seem to contribute to the model, and the fact that in model tuning we find that relatively deep trees improve accuracy, it appears that our model is relying on relatively complex combinations of many features, rather than finding one or a few proxy variables that are highly predictive of landfills.

While our accuracy metrics clearly show that our model is able to find a signal in the satellite record that is associated with open dumps, we don't think it is accurate enough to use the predictions uncritically as a proxy for the ground truth. Given that open dumps are a very rare land use class, small biases or changes in the model's accuracy between time periods or across locations could overwhelm our ability to detect real changes on the ground. Thus we will not use our model's predictions directly in analysis, but instead we use them to inform an optimal sampling and verification strategy.

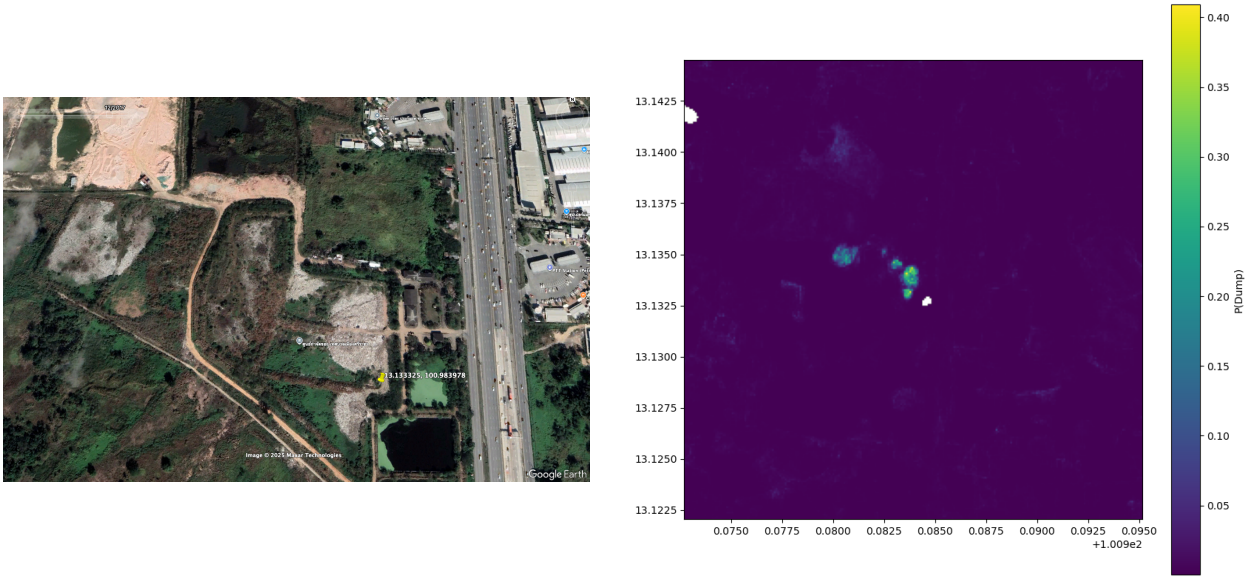
4.2 Active Sampling and Approximately Optimal Verification

Define D_i as a binary indicator for whether pixel i contains an open dump. Our model outputs p_i: the predicted probability that a pixel is a dump, based on its associated satellite data. We can define the model's prediction errors as $\nu_i = p_i - D_i$. A naïve estimate of the prevalence of open dumps (μ) in any subset of N pixels could be:

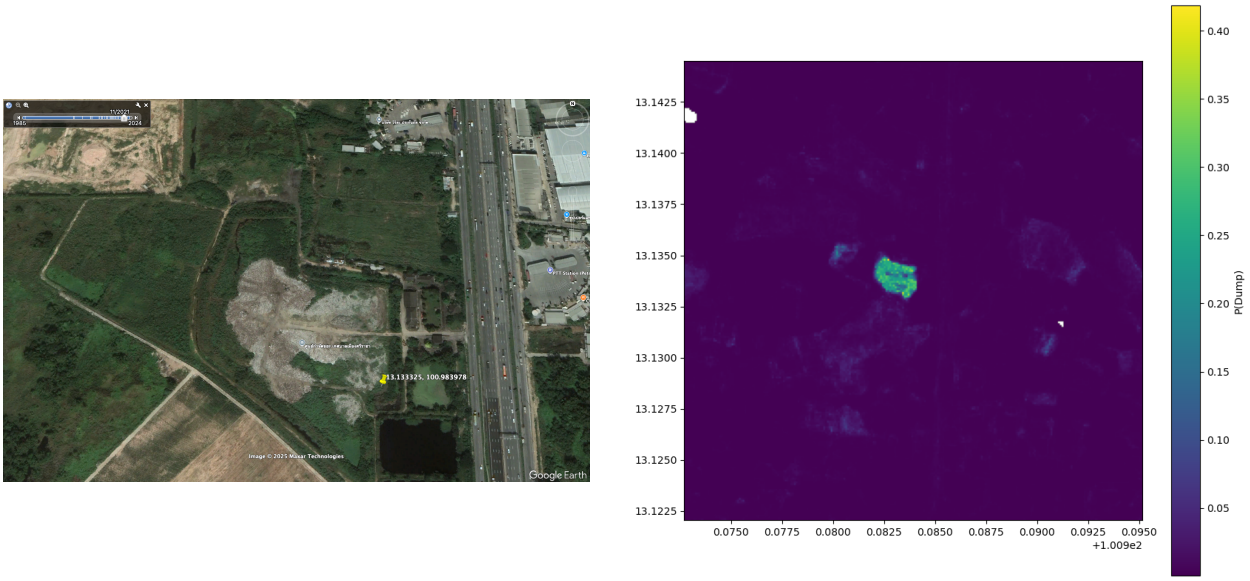
$$\widehat{\mu}^{naïve} = \frac{1}{N} \sum_{i \in N} p_i. \tag{1}$$

Abusing notation we use N to denote both the set of pixels and the size of the set. Given our definition of measurement error, this estimator will equal μ in expectation if $\frac{1}{N} \sum_{i \in N} \mathbb{E}(\nu_i) = 0$. This is unlikely — a large and growing literature documents non-classical measurement error in machine learned proxies for environmental and social science variables of interest (Jain 2020; Proctor et al. 2023; Angelopoulos et al. 2023; Kluger et al. 2025; Sanford et al. 2025; Rambachan et al. 2025), and we have exacerbated the likelihood of biases by

Figure 3: Illustration of Machine Learning Model for Waste Site in Chon Buri, Thailand



(a) 2017



(b) 2021

Notes: Figure shows satellite imagery and machine learning predictions (probability of each pixel being part of a waste site) for a small area in Chon Buri, Thailand.

using non-representative training data. For these reasons, we do not use the predictions of our model directly to estimate the prevalence of open-air landfills. The accuracy of our model changes substantially over time and across different regions, making it impossible to determine if an estimated change in landfill area is due to changes on the ground or differential prediction accuracy, which could result from something as simple as year-to-year changes in average cloud cover.

The prediction-powered inference literature has developed methods to correct for these potential biases (Angelopoulos et al. 2023; Kluger et al. 2025; Lu et al. 2025). This approach relies on a representative subset of labeled or ground-truth observations, in which D_i is known.

Let S denote the labeled set, and $S < N$ equal the number of labeled observations. The 'predict-then-debias' (PTD) estimator of μ can be defined as:

$$\widehat{\mu}^{ptd} = \frac{1}{N}\sum_{i \in N} p_i - \frac{1}{S}\sum_{i \in S} \nu_i. \tag{2}$$

In words, one can estimate the average bias of the subset in the labeled sample, and then use that estimate to correct for bias in the naive estimator. The PTD estimator is consistent as long as the labeled sample is representative. This clearly does not hold for our training data. We could construct a labeled sample by randomly verifying a subset of predictions using the higher-resolution satellite imagery from Google Earth discussed in Section 3. This verification procedure is more accurate than the satellite predictions, but it is costly and time-consuming. Furthermore, constructing this sample through random sampling would result in an estimate that is unbiased, but highly uncertain. Since waste sites are a tiny fraction of a percent of total land cover in most areas, the sample size of labeled pixels required to obtain a reasonably precise estimate of the prevalence of landfills would be enormous.

An alternative approach in the literature is to model the probability that a pixel is labeled as a function of observables, and then re-weight the labeled set to more closely match the population (Carlson and Dell 2025; Rambachan et al. 2025). This approach relies on a conditional independence assumption, which is too strong in our setting, given that we use our model to make predictions in out-of-sample locations where we do not have a good sense of the distribution of observables.

Instead, we use our machine learning model predictions to inform an active sampling and optimal verification approach (Neyman 1934; Stehman 2012; Bai et al. 2024; Zrnic and

Candès 2024; Hamilton et al. 2025). Define our estimator as:

$$\widehat{\mu} = \frac{1}{N}\sum_{i \in N} p_i - \frac{1}{N}\sum_{i \in N} w_i s_i \nu_i, \tag{3}$$

where, s_i is an indicator for labeled sample inclusion, and $w_i = 1/\pi_i$ are inverse probability weights with π_i representing labeled sample inclusion probabilities: $P(s_i = 1)$.

Given a limited budget for sampling, we can improve the efficiency of the PTD estimator by choosing π_i according to:

$$\min_{\pi_i} Var(\widehat{\mu}) \quad \text{such that} \quad \sum_i \pi_i = \bar{S}. \tag{4}$$

The constraint in this setup is an expected sample size constraint. In finite samples, S may exceed $\bar{S}$. If marginal labeling costs are small, this may not matter much, otherwise a risk averse researcher should scale $\bar{S}$ appropriately relative to their budget. Since the first term in equation 3 is a constant that is observed for the whole population, assuming a fixed model and training sample, if we assume independence across observations, then

$$Var(\widehat{\mu}) = \frac{1}{N^2}\sum_{i \in N} \frac{1-\pi_i}{\pi_i}\mathbb{E}(\nu_i^2).$$

Allow $p_i^* = \mathbb{E}[D_i|k_i]$ denote the population expectation of D_i conditional on the satellite data and other predictors k_i, so that $\mathbb{E}(p_i^* - D_i) = 0$. These p_i^* may differ from our model's estimated p_i, because our model only estimates p_i from the joint distribution of the predictors and the D_i in a selected, non-random subset of the data. Then $\mathbb{E}(\nu_i^2) = p_i^*(1-p_i^*) + (p_i - p_i^*)^2$. The first term reflects the irreducible uncertainty of pixel i, after conditioning on observables, and the second term reflects the increase in variance resulting from our model's bias.

Taking first order conditions of equation 4, and solving for π_i gives the familiar Neyman (1934) optimal sampling allocation rule, in that we should set $\pi_i \propto \sqrt{\mathbb{E}(\nu_i^2)}$. Since we lack ex-ante information on p_i^*, One way to approximate this optimal sampling rule is to simply plug in the model predicted probabilities p_i, which gives the sampling rule $\pi_i \propto \sqrt{p_i(1-p_i)}$. Of course, if our model is biased, this won't exactly minimize equation 4, but if p_i is close to p_i^*, it will get close.

This approximately optimal sampling rule results in sampling probabilities peaking for the pixels that the model is most uncertain about, with p_i close to 0.5, and shrinking as p_i approaches 0 or 1. Intuitively, this rule is effective in our setting because the vast majority of pixels have a very low probability of containing open dumps. While random uniform sampling

would result in expending a lot of effort verifying pixels that are farms or forests, which look very different from dumps, optimal sampling expends relatively more effort verifying pixels that are close to the decision threshold[10].

To demonstrate the utility of this approach, consider the following example based on a $1° \times 1°$ scene covering an area in north-east Argentina, near the border of Uruguay shown in Figure 4. The inset shows a small open dump near the border. The right panel shows the histogram of model-predicted probabilities for this scene — the y-axis is in log scale. The vast majority of pixels have predicted probabilities very close to zero[11]. This reflects the fact that it is quite easy for the model to distinguish some land-use classes, like forests and crops, from open dumps, due to their much different spectral signatures. A much fewer number of pixels have probabilities even exceeding 10%.

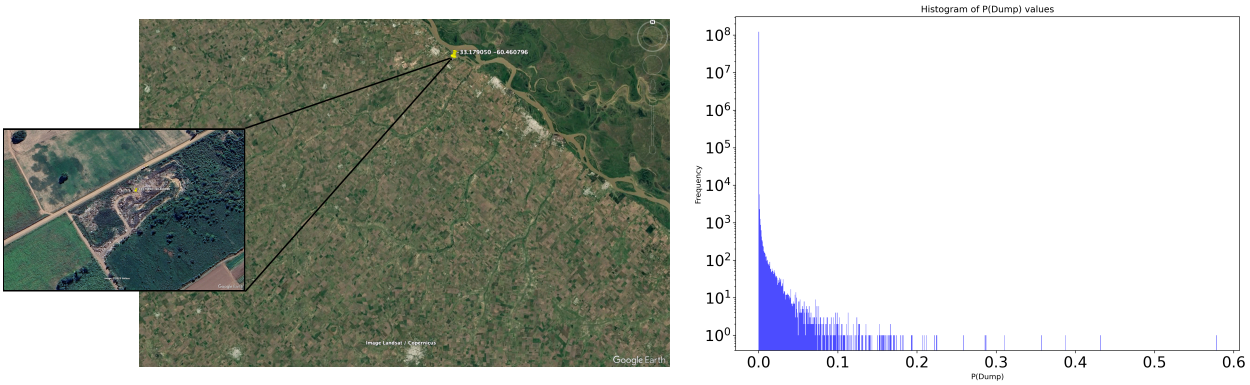
Neyman (1934) optimal sampling still ensures that we will sample a reasonably large number of the low-probability pixels due to their prevalence. Sampling proportional to $\sqrt{p_i(1-p_i)}$ means that, if probabilities were evenly distributed, we would sample about 50 pixels with a probability of 0.5, for every 1 pixel with a probability of 10^-4. However if there are $10,000$ more pixels with $p_i \approx 10^-4$ than $p_i \approx 0.5$, we will sample 200 times more of the low probability pixels than the mid probability pixels. Note that prediction errors for these low probability pixels could be quite important for our overall estimate. Thus it is important we sample a reasonably large number of them to learn about whether our model is badly miscalibrated for these pixels. Still, in this stylized example, we reduce our sampling effort on these low probability pixels by a factor of 50.

Another advantage of this technique is that, even if our machine learning model predictions are badly miscalibrated, this will not bias our resulting estimates of the mean. Ultimately, $\hat{\mu}$ is estimated based on a verified sample of pixels that is representative of the population of pixels. Poor predictions simply mean that the variance of $\hat{\mu}$ is larger than anticipated, and a larger sample size would be required for the same level of precision. Similarly, if our model has differential accuracy across different regions or time periods, this will result in differential uncertainty in our estimates of dump prevalence, or heteroskedasticity in other words, but it won't result in biased comparisons of means or trends.

[10] Common land-use types that our model frequently confuses with dumps include construction sites, quarries, empty urban lots, and mines.

[11] No pixels have probabilities exactly equal to zero or one, except that we exclude areas labeled as surface water by Dynamic World (Brown et al. 2022). Water is typically very easy for models to distinguish from other land use types.

Figure 4: A Sample Tile and Histogram of Pixel Probabilities



Notes: A sample tile near Rosario, Argentina. Inset image shows a small open dump. Histogram shows model predicted probability distribution for all pixels in the scene (y-axis in logs).

4.2.1 Efficiency Gains with Miscalibration

We can perform some simple calculations based on the scene in Figure 4 to estimate the precision gains associated with our sampling strategy relative to a uniform random sample. Define $\sigma_i^* = \sqrt{p_i^*(1-p_i^*)}$, $\sigma_i = \sqrt{p_i(1-p_i)}$, and $\mathcal{E}_i = \sqrt{p_i^*(1-p_i^*) + (p_i - p_i^*)^2}$. Plugging in our approximately optimal π_i, the variance of $\widehat{\mu}$ is:

$$Var(\widehat{\mu}) = \frac{1}{N^2 S} \sum_{i \in N} \frac{\mathbb{E}(\nu_i^2)}{\sigma_i} \sum_{i \in N} \sigma_i - \frac{1}{N^2} \sum_{i \in N} \mathbb{E}(\nu_i^2). \tag{5}$$

We can calculate the design effect, or the relative change in variance from following our sampling strategy, relative to the predict-then-debias (PTD) estimator using uniform random sampling as:

$$D_{eff} = \frac{Var(\widehat{\mu})}{Var(\widehat{\mu}^{ptd})} = \frac{\left(\sum_{i \in N} \frac{\mathcal{E}_i^2}{\sigma_i}\right)\left(\sum_{i \in N} \sigma_i\right) - S \sum_{i \in N} \mathcal{E}_i^2}{(N-S) \sum_{i \in N} \mathcal{E}_i^2}. \tag{6}$$

In the realistic case where $N >> S$, we can approximate the design effect ignoring the second terms in the numerator and denominator. Define $R_i = \frac{\mathcal{E}_i}{\sigma_i}$ as a measure of model miscalibration, noting that if $p_i = p_i^*$, then $R_i = 1$, and R_i is increasing as the difference

between p_i and p_i^* increases. Then the design effect can be rearranged to show:

$$D_{eff} \approx \frac{1}{N} \frac{\left(\sum_{i \in N} \mathcal{E}_i\right)^2}{\sum_{i \in N} \mathcal{E}_i^2} \frac{\sum_{i \in N} R_i^2 \frac{\sigma_i}{\sum_{i \in N} \sigma_i}}{\left(\sum_{i \in N} R_i \frac{\sigma_i}{\sum_{i \in N} \sigma_i}\right)^2} \tag{7}$$
$$= \frac{1 + CV_{\sigma R}^2}{1 + CV_{\mathcal{E}}^2},$$

where $CV_{\mathcal{E}}$ is the coefficient of variation of $\mathcal{E}_i$, and $CV_{\sigma R}$ is the coefficient of variation of R_i weighted by σ_i.

To gain some intuition for this formula, first consider the case where the model is well-calibrated, so $p_i = p_i^*$. In this case, $R_i = 1$ for all i, so $CV_{\sigma R} = 0$, and the numerator simplifies to 1. This case represents the optimal lower bound — there is no miscalibration, so the gains from sampling depend on the denominator, which characterizes variation in the amount of uncertainty in the data. Using D_{eff}^* to denote this optimal lower bound, the entire formula simplifies to:

$$D_{eff}^* \approx \frac{1}{N} \frac{\left(\sum_{i \in N} \sigma_i\right)^2}{\sum_{i \in N} \sigma_i^2}. \tag{8}$$

By Cauchy-Schwarz, this quantity will always be less than 1. We can calculate this theoretical quantity based on the distribution of p_i in the Argentinian scene described above, and we find $D_{eff}^* = 0.0012$. In other words, if our model was perfectly calibrated, the active sampling method would reduce the variance of $\widehat{\mu}$ relative to $\widehat{\mu}^{ptd}$ by 99.88% for the same sample size.

On the other hand, given that our model is imperfect, the correlation between the miscalibration R_i, and the sampling weight σ_i, becomes important. We can rewrite $1 + CV_{\sigma R}^2$ as:

$$1 + CV_{\sigma R}^2 = \frac{\sum_{i \in N} \frac{\mathcal{E}_i^2}{\sigma_i} \sum_{i \in N} \sigma_i}{\left(\sum_{i \in N} \mathcal{E}_i\right)^2}, \tag{9}$$

In an extreme case where our predictions were random noise, so $\mathcal{E}_i$ and σ_i are independent, this becomes:

$$1 + CV_{\sigma R}^2 = \frac{N \sum_{i \in N} \mathcal{E}_i^2}{\left(\sum_{i \in N} \mathcal{E}_i\right)^2} \frac{1}{N^2} \sum_{i \in N} \frac{1}{\sigma_i} \sum_{i \in N} \sigma_i \tag{10}$$
$$\approx (1 + CV_{\mathcal{E}}^2)(1 + CV_{\sigma}^2),$$

where the second line comes from taking a second order Taylor approximation of $\frac{1}{\sigma_i}$. In this case, $D_{eff} \approx 1/D_{eff}^*$. Now our highly variable weights blow up the variance, and we do much worse than a random sample for the same sample size. An intermediate case is if the model

just predicts a constant for all i, then our method reduces to uniform random sampling and $D_{eff} = 1$.

Thus, ultimately the efficiency gains are a function of both model calibration and the distribution of uncertainty in the population. This is an empirical question to which we will return in Section 4.3.

4.2.2 Estimating Changes

This sampling and verification method allows us to estimate the prevalence of informal landfills in levels precisely. Many interesting research questions depend on changes in the prevalence of landfills over time, however. We can easily adapt our approach to precisely estimate changes, however, by simply sampling with $\pi_i \propto \sqrt{p_i^{chg}(1 - p_i^{chg})}$, where p_i^{chg} is the probability that pixel i has changed from a dump to non-dump or vice-versa.

Though more sophisticated machine learning techniques can estimate p_i^{chg} directly given a time-series stack of satellite images and corresponding training data, these models tend to be computational and memory intensive. More pressingly, our training data is not temporally balanced, and many observations only correspond to the observation of a dump in a particular year. For these reasons, we use a simpler model that is trained to estimate p_i in levels based on the images from a particular year. We then make independent predictions for each year of interest, giving a vector of probabilities: $\mathbf{p}_i = [p_{i,2017}, ..., p_{i,2021}]$.

Any approach using this vector of predictions to estimate the probability that a pixel has changed needs to make an assumption about serial correlation in the model's errors. We assume that pixel-specific model errors are strongly correlated. In particular, we make a simple assumption: if $\min(\mathbf{p}_i)$ is a dump, then all of the p_{it} are dumps, and if $\max(\mathbf{p}_i)$ is not a dump, then all the p_{it} are not dumps. This implies a simple way to calculate p_i^{chg} as: $p_i^{chg} = \max(\mathbf{p}_i) - \min(\mathbf{p}_i)$. This is clearly not exactly correct, but as above, miscalibrated estimates of p_i^{chg} will only matter for the variance of our results.

Since levels and changes are both of interest, we dedicate half of our sampling budget to estimating each. We use the predicted probabilities from our machine learning model for 2018 as the p_i and the difference between the minimum and maximum probabilities for 2017-2021 as the p_i^{chg}. Once we decide to sample a given pixel, it is relatively cheap to view the complete set of historical satellite imagery available for that pixel in Google Earth Pro. Therefore, even though we only have satellite imagery (and therefore machine learning predictions) for 2017-2021, we extend our sample temporally by performing verifications for

2011-2023.

4.3 Model Deployment and Verifications

After completing model and hyperparameter testing, we train a final model using all pixels from all sites in both the training and validation sets so that our final model can learn from all of the examples we have. We then deploy this model globally out-of-sample to make predictions using the Vertex AI platform integrated with Google Earth Engine. Afterwards we verify an approximately optimal subset of these predictions.

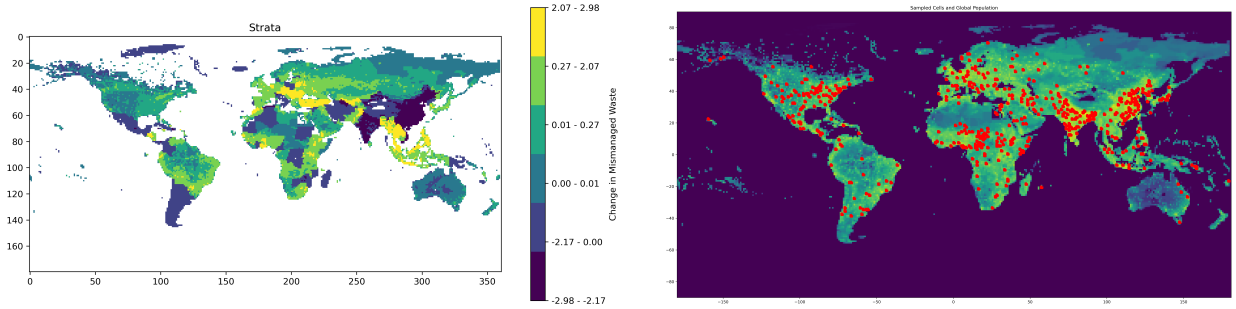
The economics of our deployment strategy may be interesting to some readers. Since Sentinel-2 has a spatial resolution of 10 meters, making predictions for the entire land area of Earth for multiple years is expensive. Downloading satellite data, preprocessing features, and then using our model to make predictions for a set of pixels can be done as part of a single 'task' that scales well within contiguous areas, but is subject to memory constraints on Google's servers. Thus we define our task to run on contiguous $1° \times 1°$ tiles (approximately 111 km x 111 km at the equator), which are the largest tiles that we can consistently run without hitting constraints.

Running this sequence of operations takes on average about 1 hour per tile-year. Thus we choose 500 tiles, stratified based on a rough estimate of where we expect to see the biggest changes in our outcome variable of interest (open dumps). We calculate the per-capita change in waste imports from 2017 to 2019 at the country level, and multiply it by the estimated fraction of mismanaged waste at the country level from Kaza et al. (2018). We then allocate this to tiles within countries by multiplying by tile population. The left panel of Figure 5 shows six strata defined by this variable using the 5th, 25th, 50th, 75th, and 95th percentiles as thresholds. We then sample an equal number of tiles within each stratum with probabilities proportional to the tile population to focus our analysis on areas where the externalities of improper waste management are large due to the proximity to population centers. The right panel in Figure 5 shows our selected tiles in red. For all of our results we re-weight tiles by the inverse probability of selection so that results are representative of global land area.

In each tile, we use our model to make predictions for all pixels with at least one high-quality, cloud-free Sentinel-2 and Sentinel-1 image for each year between 2017 and 2023. We exclude water pixels, and for pixels with missing imagery we impute the tile median probability.

We then use these predictions to apply our approximately-optimal sampling strategy within

Figure 5: Strata and Tile Sample for Global Analysis



Notes: Figure illustrates global deployment on $1° \times 1°$ tiles. The left panel shows six strata based on estimated changes in mismanaged imports 2017-2019. Units are the inverse hyperbolic sine of change in the dollar value of mismanaged imports. The right panel shows selected tiles in red.

tiles. For each chosen pixel, we ask two experts trained in visual interpretation of satellite imagery to independently ascertain whether the pixel is an open dump using the high-resolution Google Earth Pro historical imagery for all years from 2011-2023. For years where no high resolution imagery exists, we perform linear, pixel-specific interpolation using values from surrounding years. In 98.4% of cases the experts agree. In the remaining 1.6% of cases, we have one of the authors break the tie.

This verification process is also time consuming and expensive. We set our expected sampling budget to 15,000 pixels globally, which comes out to an expected 30 pixels in each of our 500 tiles. Thus our results are highly uncertain within tiles, but become more precise at higher levels of aggregation.

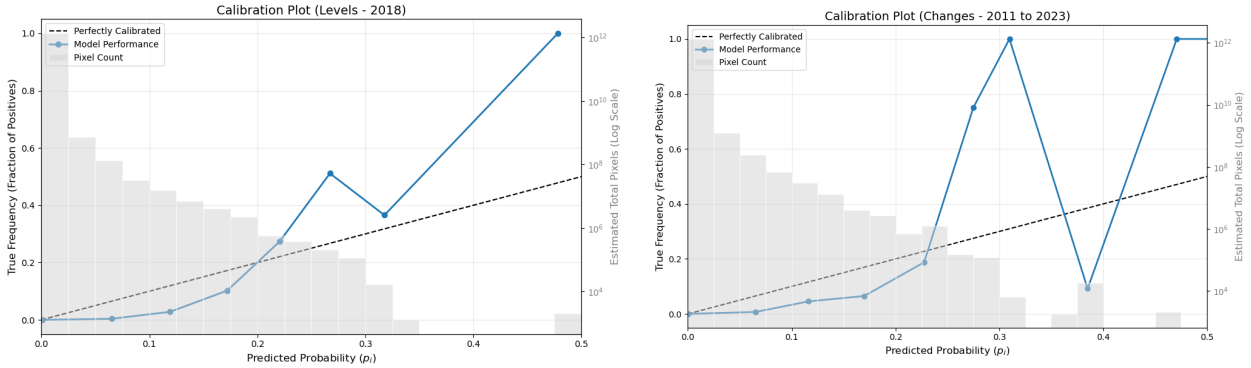
4.3.1 Model Calibration

Figure 6 shows model calibration plots. The x-axis shows the model predicted probability and the left y-axes show the fraction of pixels in that probability bin that were verified as containing an open dump. If our probabilities were perfectly calibrated, the blue line would equal the 45 degree line. Clearly this is not exactly the case - at low probabilities, we somewhat overestimate the probability of dumps, and at relatively higher probabilities ($> 20\%$) our model is too conservative. Still, it is clear that our model's predictions contain information.

A similar, but noisier story holds for our predicted changes, as seen in the right panel, though some of the variation at higher predicted probabilities is due to the relatively small number of pixels sampled at those probabilities. In both plots, the grey bars show a histogram of

the number of pixels at each predicted probability estimated for the entire world, with the axes in log scale on the right.

Figure 6: Model Calibration



Notes: Figure shows calibration plots in blue for p_i and p_i^{chg} against dashed black 45 degree line. Model estimated probabilities are on the x-axis and relative open dump frequencies are on the left y-axis in black. Grey histograms and right y-axes show the estimated number of pixels at each estimated p_i.

We can use this information to calculate the ex-post efficiency of our approximately optimal sampling strategy, by estimating p_i^* as a function of p_i, and then calculating our realized design effect. We estimate p_i^* as a function of p_i using a LOESS regression, and plug it in to equation 7. We find a design effect of 0.41, indicating that we will be able to estimate mean dump prevalence with a variance 60% smaller than if we had used uniform random sampling. In other words, we have increased our effective sample size by $1/0.41 = 2.44$, as if we had sampled 36,585 pixels randomly instead of our 15,000 strategically chosen pixels. We can also calculate D^*_{eff}, the lower bound of the design effect that we would have obtained with a perfectly calibrated model, and we find a value of 0.14, indicating that miscalibration reduced the value of our sampling strategy.

When it comes to estimating changes efficiently, we do even better. We estimate an ex-post design effect of 0.17, relative to an optimal lower bound of 0.06. This gives an effective sample size when estimating changes of more than 88,000 pixels.

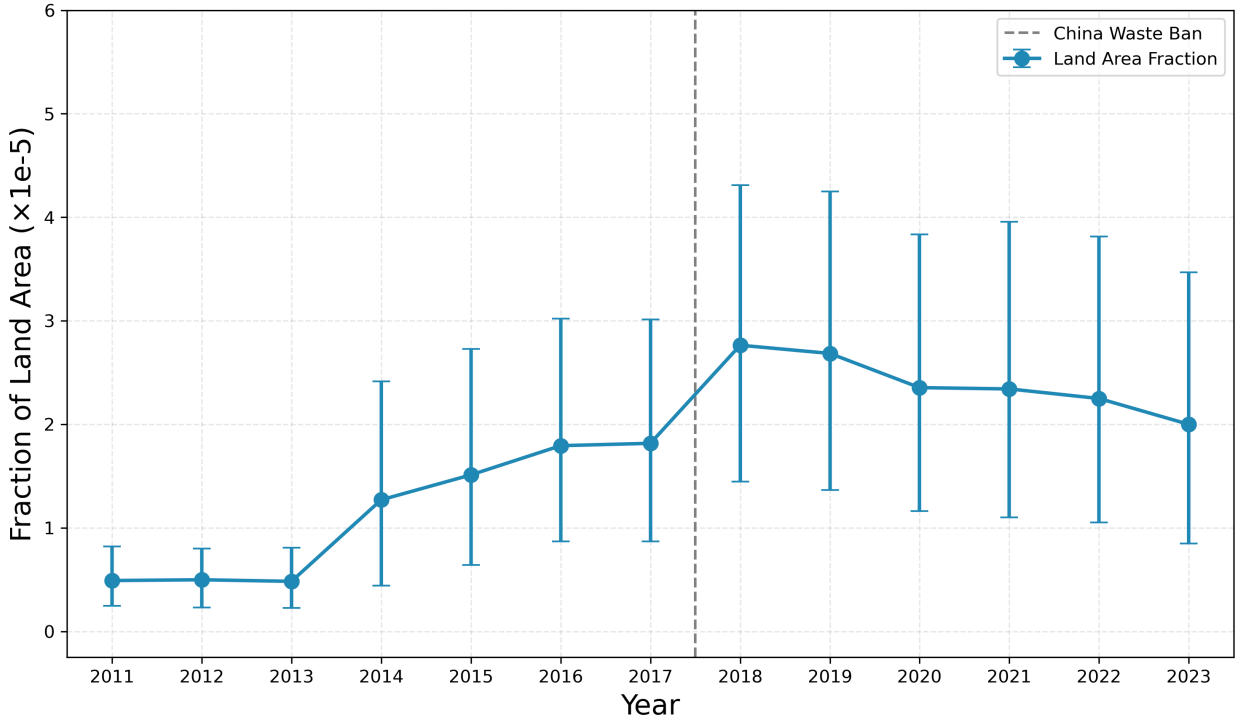
5 Descriptive Facts

Our method allows us to generate a novel set of descriptive facts about the distribution of open dumps globally. We estimate the fraction of global land area covered with open dumps in 2018 was 2.76×10^{-5} (95% CI: $[0.9, 4.3] \times 10^{-5}$). This corresponds to an area of more

than 4,000 square kilometers, larger the US state of Rhode Island. As a point of reference, this is about 0.2% of estimated urban land surface area in 2015.[12]

Figure 7 shows the time trend of our global estimate. There was a large jump between 2017 and 2018, corresponding to the year of the Chinese waste import ban. Our point estimates indicate a 50% increase in global dump area. While it is difficult to claim causality based on a single time series, it is strongly suggestive that the ban lead to a global increase in dumps, despite the decrease in international trade flows. We also find a similar increase between 2013 and 2014, corresponding to the year of the Green Fence policy, when China increased inspections on waste importers.

Figure 7: Global Results: Share of Land Area Covered by Waste Sites



Notes: Estimates based on 15,000 verified pixels selected using approximately optimal sampling technique. Bootstrapped 95% confidence intervals.

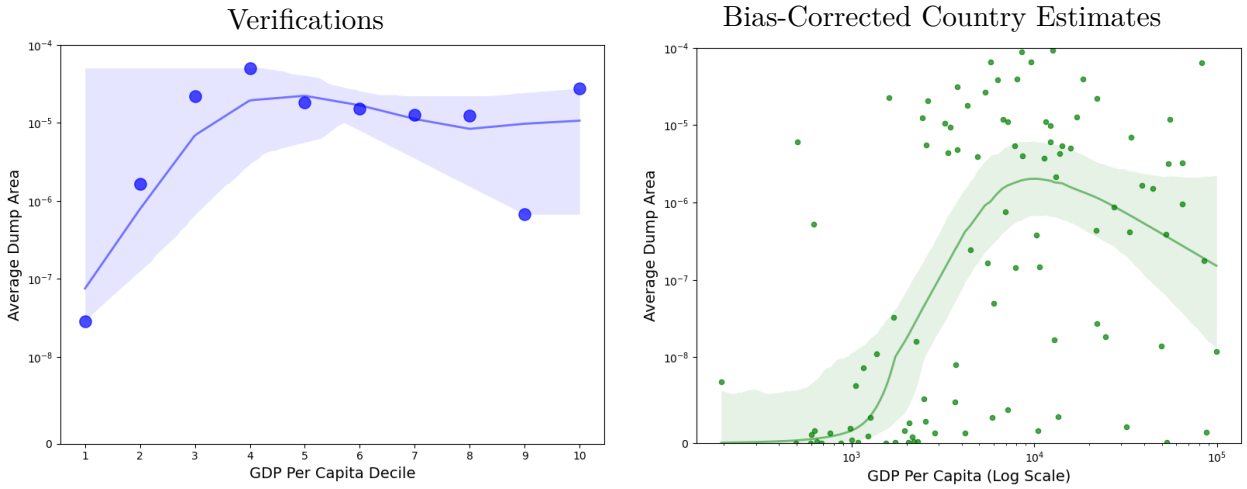
These increases are not obvious. If Chinese recyclers discarded a large fraction of the imports in open dumps, then we would expect dumps in China to go down after the ban, especially if the alternative is for waste to end up in sanitary landfills in high-income countries. Instead the pattern suggests that the waste was diverted to countries that had worse waste

[12]Estimates of urban land area from CIESIN (Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 3).

management practices than China. We explore this further in Section 6.

Figure 8 shows the prevalence of open dumps plotted against GDP per capita. The plots clearly show the existence of an 'Environmental Kuznets Curve' (EKC) in both the verified points (left) and the bias-corrected country level estimates (right) (Grossman and Krueger 1995). Low income countries have the lowest prevalence, presumably because they generate less waste. Dump prevalence sharply increases for middle income countries, and then declines a bit for high-income countries, though the right side of the EKC is fairly flat. This decline could reflect increased usage of sanitary landfills and other waste disposal techniques that are likely to be less environmentally harmful.

Figure 8: Dump Prevalence vs GDP Per Capita

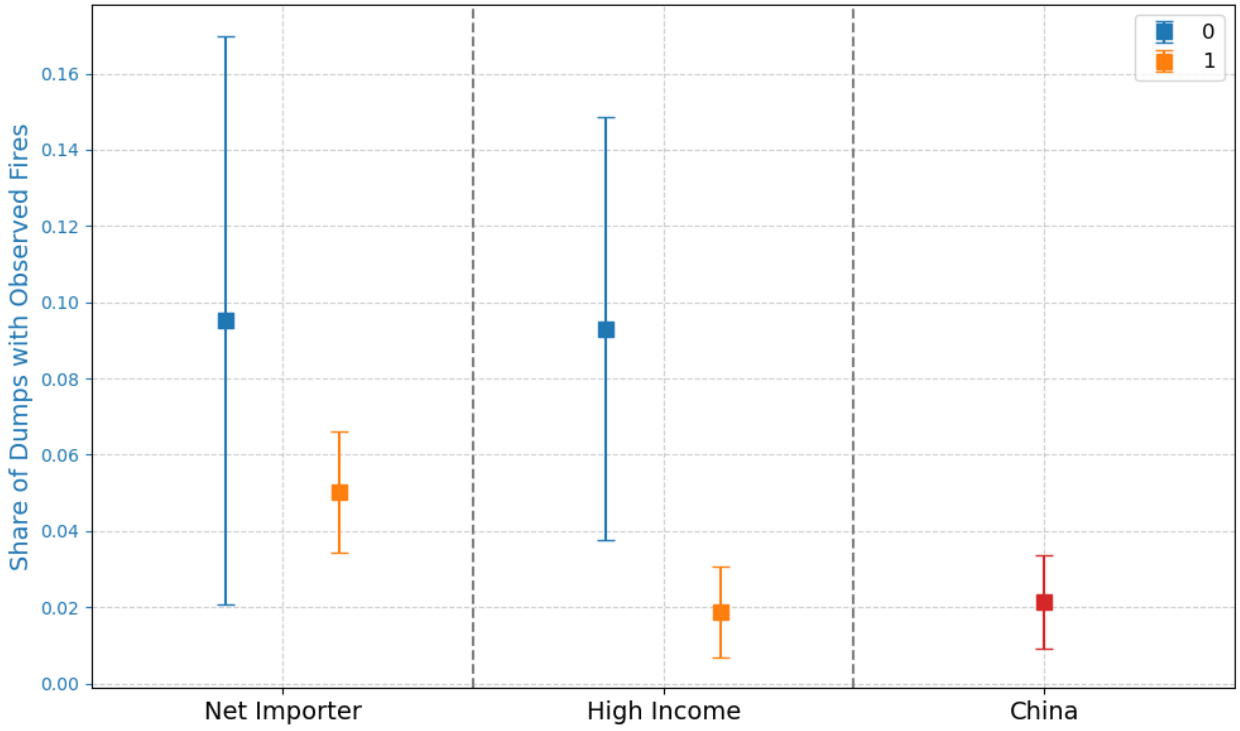


Notes: Estimates of dump prevalence vs country GDP, based on 15,000 verified pixels selected using approximately optimal sampling technique. Left panel shows average prevalence using only verified pixels, calculating means within country-GDP per capita deciles. Right panel shows debiased country-level estimates estimated using equation 3. Plots show LOESS smoothed lines and bootstrapped 95% confidence intervals with 1000 replicates.

We also find more evidence of heterogeneity in waste management practices across countries. Figure 9 shows the prevalence of fires in open dumps in different types of countries. Strikingly, nearly 10% of dumps in low-income countries have at least one fire detected in a given year. In high-income countries, the number is 2%. In China, the figure is also close to 2%, averaging across years. The figure also shows the comparison between waste net-importing countries and other countries — it seems that dump-fires are less frequent in importing countries, but the estimates are noisy.

Lastly, Figure 10 shows the prevalence of open dumps with respect to population density. In all countries, dumps are concentrated in areas at the 80-100th percentile of population

Figure 9: Garbage Fire Prevalence

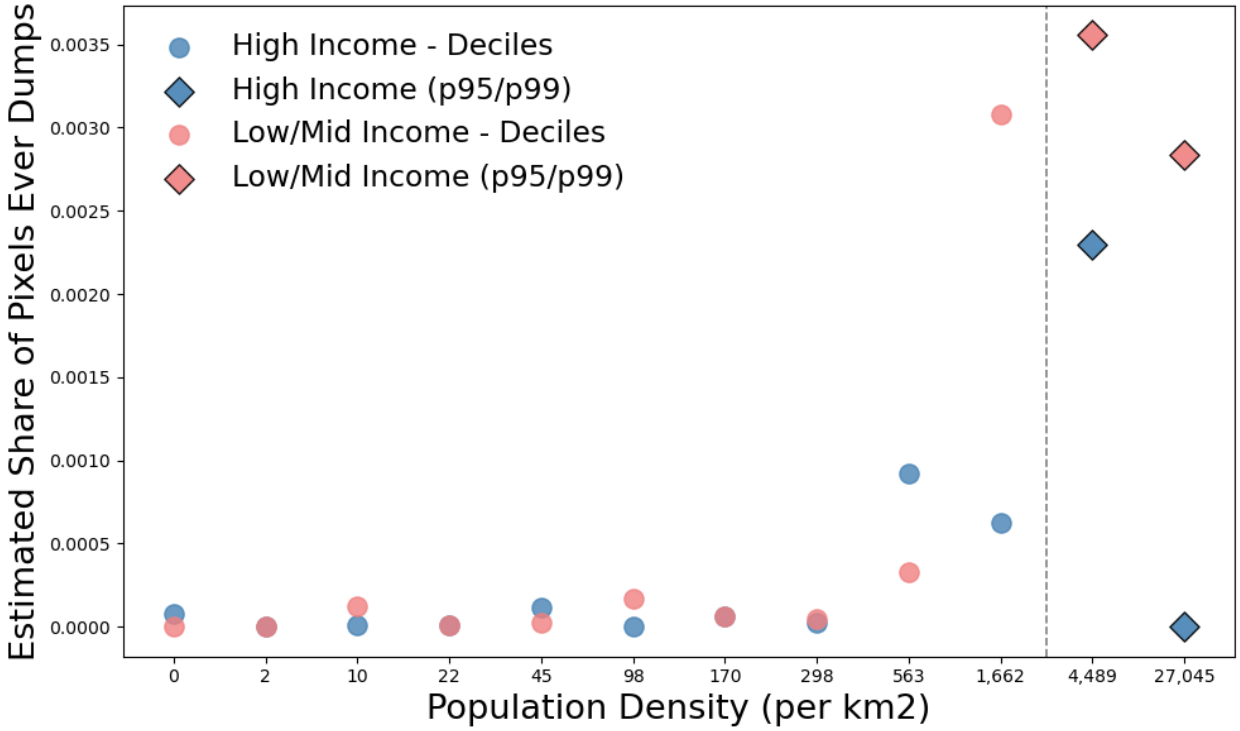


Notes: Estimates of dump fire prevalence by net importer status, country income classification, and for China, based on 15,000 verified pixels selected using approximately optimal sampling technique. Dump fires defined as the intersection of at least one VIIRS fire detection during the year with a verified dump. Net importer defined as imports greater than exports using Gaulier and Zignago (2010). Income uses World Bank classification threshold of $13,935 GDP per capita.

density. Waste is generated in population centers and is costly to transport, facts which are sufficient to explain this pattern. Given the externalities associated with open dumps, however, optimal policy might involve reducing human exposure to open dumps by locating dumps in less-dense areas. We see some evidence of this occurring in high-income countries represented by the dots in blue. In these countries, waste is relatively less prevalent at the 90-100th percentile of population density than at the 80-90th percentile. In fact we see almost no waste in the very densest cities of rich countries at the 99th percentile, where a density of 27,045 people per square kilometer approximates the population density of Manhattan.

In contrast, dumps are nearly monotonically increasing in population density in low and middle income countries, where we see the highest levels of dumps at the 95th and 99th percentile of population density.

Figure 10: Dump Prevalence vs Population Density



Notes: Point estimates of dump prevalence by population density decile, plus 95th and 99th percentiles, based on 15,000 verified pixels selected using approximately optimal sampling technique. Blue points show estimates for high income countries, and red points show combined low and middle income country estimates (World Bank classification threshold of $13,935 GDP per capita). Population density data is based on a 1km grid from Earth Science Data Systems, NASA (2024).

6 Dumps and Trade

We use the natural experiment induced by the Chinese waste import ban to estimate how much imported waste is mismanaged. On one hand, if imported waste is fully used in recycling processes, we might not see any increase in open dumps. On the other hand, if the raw material used in recycling is heavily contaminated and a large fraction is discarded, we should see dumps increase in countries that saw their imports increase.

Our primary measure of open dumps is the area estimated from manual verifications, as described in the previous sections. This approach provides a longer time series (2012-2023) than the debiased machine learning predictions, which only cover 2017 to 2021, though we show robustness of our estimates to incorporating these predictions in the years in which they are available. As discussed in Section 2, we only measure the surface area of dumps, so we can only detect extensive margin increases in waste mismanagement. To the extent that imported waste is piled up or compressed in landfills, it will bias downward our estimates of the fraction of mismanaged waste.

We define waste imports and exports as the trade in products banned by China's National Sword Policy, including waste plastics, textiles, paper, and minerals. We use CEPII-BACI derived from UN Comtrade data to collect bilateral trade flow quantities and values for 2012-2023 (Gaulier and Zignago 2010).[13] We then aggregate our verified open dump observations to the country level to merge with the country-level trade data. These country-level estimates have considerable uncertainty, and the uncertainty varies across countries due to a differing number of verified observations in different countries. Thus we will estimate all of our regressions using variance-weighted least squares, where we use the inverse of the empirically estimated country-specific variance as weights to increase precision.[14] Our final sample includes estimates of dump prevalence in 117 countries, however we only verify greater than zero dump pixels in all years for 50 of these countries.

As a first exercise, we regress net waste imports in each year on dump prevalence, controlling for country and year fixed effects. The results are found in Appendix Table A1, using log dump area in column one, for the countries with verified dumps, and dump area in levels in column two for the complete sample. In both cases, we find a positive association between

[13] Values are available for most countries in all years. To create a complete panel of trade *quantities*, we impute missing country-year quantity observations by applying the country-specific average value-to-quantity ratio from years with available data.

[14] For some countries with a limited number of verifications we observe no open dumps, which would result in an empirical variance estimate of 0. We correct for this by estimating an upper 95% confidence interval for dump prevalence using $U = 1 - .05^{1/n}$. We then estimate the country variance as $U(1-U)/n$ (Clopper and Pearson 1934).

the changes in the quantities of imported waste and changes in the area of open dumps, though the relationship in the complete sample is only significant at 90%. Specifically, a 10% increase in waste imports in a given year is correlated with a 2.5% to 4.5% increase in open dump area for the same year depending on the specification.

This simple panel fixed effect specification identifies the causal effect of imports on dump prevalence, conditional on the assumption that there are no time-varying trends that correlate with both changes in imports and changes in dump area. To analyze the plausibility of this assumption, we leverage changes in imports around the timing of China's National Sword Policy as a quasi-exogenous shock. We estimate a continuous two-way fixed effects model:

$$\log D_{it} = \beta \Delta \text{WasteImports}_{i,2016-19} \times \text{Post-2016} + \gamma_i + \delta_t + e_{it} \tag{11}$$

and its corresponding event-study specification:

$$\log D_{it} = \sum_{t=2011}^{2023} \beta_t \Delta \text{WasteImports}_{i,2016-19} \times T_t + \gamma_i + \epsilon_{it} \tag{12}$$

where D_{it} is the estimated open dump area in country i and year t. The continuous treatment variable, $\Delta\text{WasteImports}_{i,2016-19}$, captures a country's change in waste trade exposure and is calculated as the difference in net waste imports (imports minus exports) between 2019 and 2016. In equation 11, we interact this time-invariant measure with Post-2016, an indicator variable for years after the policy's 2017 announcement. The model includes country (γ_i) and year (δ_t) fixed effects. Equation 12, interacts our continuous 'treatment' variable with dummies for each year T_t. This allows us to inspect for pre-trends that could give evidence of confounding variables associated with changes in dumps in the countries that saw a large import shock, and also see how the shock persists over time.

Table 1 presents the results of estimating equation 11. Similar to the panel results, we see a highly significant and positive effect of the waste-ban induced import shock on the prevalence of open dumps in the years 2017-2023. Similar to above, column (1) shows results in logs for countries with non-zero open dump area in all years, and column (2) shows the full sample in levels. The elasticity estimates imply a 1.1 to 2.3% increase in dumps for a 10% increase in imports. These estimates are similar, though slightly smaller than we find in the panel regressions.

For the coefficient of interest β to identify the causal elasticity of dumps with respect to imports, $\Delta\text{WasteImports}_{i,2016-19}$ must not be correlated with unobserved drivers of time trends in open dumps (Callaway et al. 2021). This is plausible in our setting — not only

Table 1: Effect of Change in Net Imports on Open Dump Area

	Dependent variable:	
	Log(Dump Area Fraction)	Dump Area Fraction (x Million)
	(1)	(2)
Δ Net Import Quantity × Post-2016	0.0003**	0.0005***
	(0.0001)	(0.0002)
	p = 0.046	p = 0.007
Elasticity	0.111	0.234
Number of Countries	50	117
Observations	501	1,534

Note: Table shows results of regression accoring to equation 11 for 2012-2023. Column (1) shows log results for countries with non-zero open dump area, while Column (2) shows level results. All columns are weighted by the variance the open dump area. Import quantities in 1000s of tons. SEs clustered by country. *p< 0.1; **p< 0.05; ***p< 0.01.

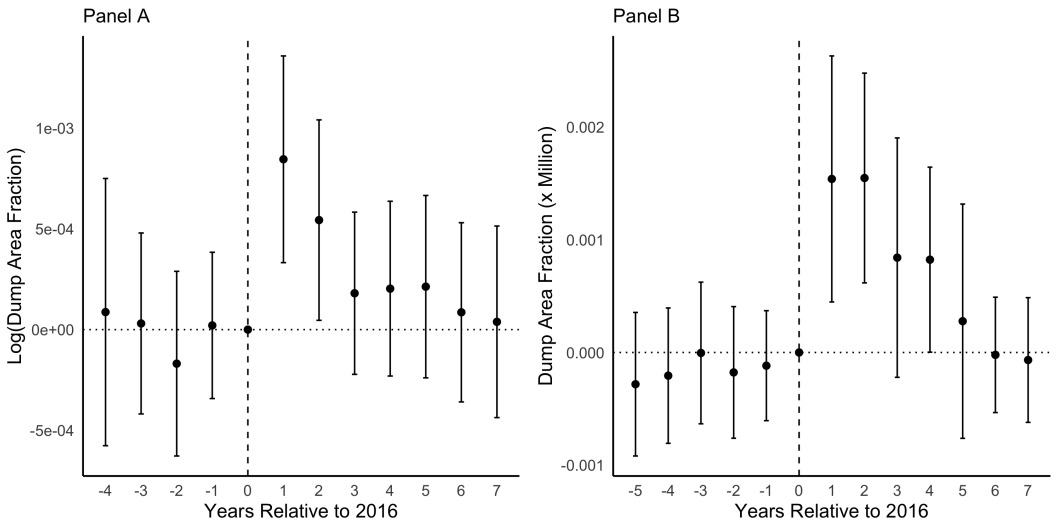
was the timing of the ban unanticipated, but the magnitude of the import shock was mostly determined by the structure of pre-ban international trade patterns. Since China has a significant trade surplus with the US in manufactured goods, shipping from the US to China is cross-subsidized by the need for container ships to make a round trip, thus making it affordable to ship recycled materials from the US to China. Since these ships continued to make the round trip journey after the waste ban, the most convenient destinations for the recycled goods were the places that were on the way back to China, or only required a short detour.

Furthermore, while it is still plausible that the geographic locations of these countries, or their positions in the international trade network are correlated with unobservable, time-varying determinants of changes in landfill prevalence, we should expect most of these determinants to evolve slowly over time. Similarly, the prevalence and extent of open dumps changes slowly of time. In contrast, the suddenness and size of the import shock should allow us to see changes relative to any background trends.

Figure 11 provides evidence for the plausibility of this argument by plotting the coefficients of the dynamic event-study estimated from equation 12 in logs (Panel A) and levels (Panel B). We see very little in the way of pre-trends, showing that the waste shock was orthogonal to trends determining the evolution of open dumps in the years before the waste ban. The

estimated coefficient then jumps in 2017 and 2018, the years of the announcement and implementation of the ban. Interestingly, the coefficient losses significance by 2019, and the point estimates approach zero by six years following the ban.

Figure 11: Effect of Change in Net Imports on Open Dump Area, Event Study Version



Notes: Figure shows estimates of the coefficients of the event-study regression (equation 12) for 2012-2023. The reference year is 2016. Panel A shows the coefficients on changes in imports (000s of tons) from 2016-2019 with log dump area as the dependent variable for countries with non-zero open dump area in all years. Panel B shows results for dump prevalence in levels as the dependent variable for all countries. Import quantities in 1000s of tons. SEs clustered by country. Bars show 95% confidence intervals.

These dynamic effects could be evidence of transition frictions, it is possible that new importing countries could not handle the large inflows of waste immediately, but after several years, they were able to set up the required infrastructure for recycling and proper disposal. An alternative explanation is that the countries responded with their own regulations on waste imports. This was the case for several countries. Malaysia banned waste imports shortly after China, and several other countries, including Vietnam and Indonesia passed their own bans in the following years. In this case, the 2016-2019 import shock would not be predictive of waste imports during the later years of the study period.

These findings are robust to several alternative specifications, including unweighted versions of the regressions, averaging pre- and post-policy imports to calculate the change in net imports, and using alternative categories of waste imports. We show the estimated coefficients of these alternative specifications in Appendix Figure A5.

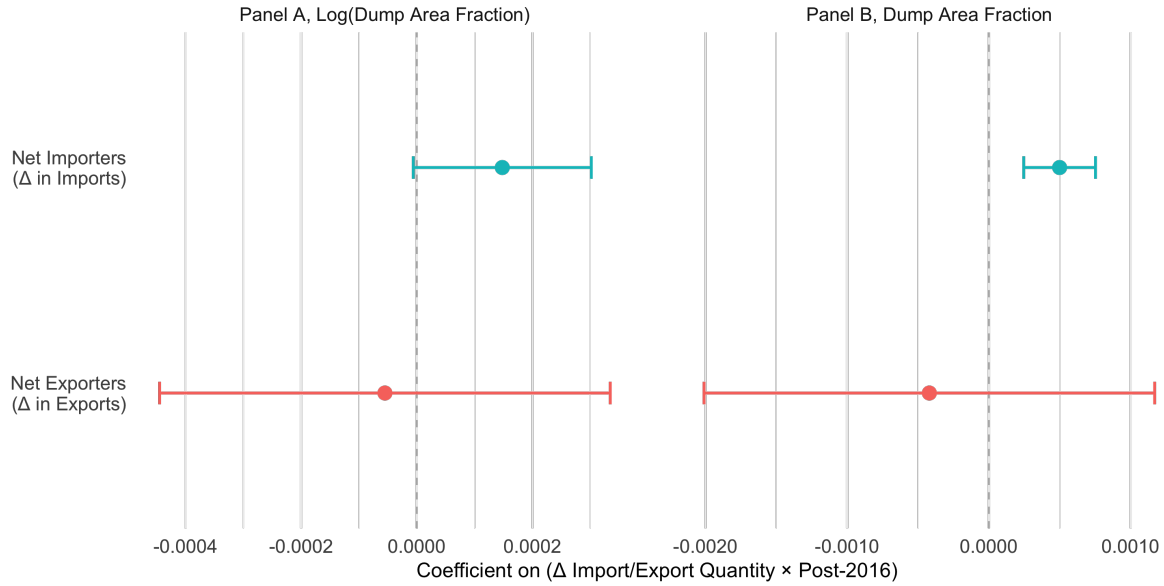
6.1 Heterogeneity

Next, we examine how the impact of waste trade changes on open dump area varies across different countries, regions, and contexts.

6.1.1 Net Importers and Net Exporters

First, we classify countries as net importers (imports > exports) or net exporters based on their average trade flow quantities from 2012 to 2023. Figure A3 shows a map of which countries fall in each category. The map is also colored by the changes in imports and exports respectively from 2016-2019. Net importing countries with notable increases in waste imports during this period include Vietnam, India, Indonesia, Turkey, Thailand, and Malaysia. Net exporting countries with large decreases in waste exports include Ghana, the United States, Japan, France, the United Kingdom, Australia, and Canada.

Figure 12: Effect of Change in Imports and Exports on Open Dump Area, by Trade Classification



Notes: Figure shows results of regression according to equation 11 for 2012-2023, separately for net importers and net exporters. The main independent variables are changes in imports for net importers and changes in exports for net exporters. Panel A shows log results for countries with non-zero open dump area, while Panel B shows level results for all countries. Import and export quantities in 1000s of tons. SEs clustered by country. Bars show 95% confidence intervals.

We rerun regression 11 separately for net importers and net exporters. For net importers (exporters), the main independent variable is the change in waste imports (exports). Figure

12 shows that the main results are driven by net *importing* countries. The blue lines in Figure 12 show that increases in waste imports are correlated with increases in open dump area in net importer countries. For net exporting countries, decreases in exports are also correlated with increases in open dump area, as the negative coefficient in red shows, but these results are very imprecisely estimated.

Notably, the net importing countries tend to be low and middle income countries with less developed waste management infrastructure, while exporting countries (with the exception of Ghana) tend to be high-income countries. If these countries have alternative means of disposing waste, such as recycling, incineration, or deposition in sanitary landfills, this could explain the results that we observe.

Table 2: Effect of Change in Net Imports on Open Dump Area, by GDP Per Capita

	Dependent variable:	
	Log(Dump Area Fraction)	Dump Area Fraction (x Million)
	(1)	(2)
Δ Net Import Quantity × Post-2016	0.0003**	0.001***
	(0.0001)	(0.0002)
	p = 0.044	p = 0.004
× High Income	−0.0004**	−0.0002
	(0.0002)	(0.001)
	p = 0.015	p = 0.732
Observations	501	1,521

Note: Table shows results of regression accoring to equation 11 for 2012-2023, including an additional interaction term for high-income (>$13,935 GDP per capita) countries. Column (1) shows log results for countries with non-zero open dump area, while Column (2) shows level results. All columns are weighted by the variance the open dump area. Import quantities in 1000s of tons. SEs clustered by country. *p< 0.1; **p< 0.05; ***p< 0.01.

6.1.2 GDP Per Capita

We next look at heterogeneity by country income. We use the World Bank's threshold of a GDP per capita of $13,935 to define high-income countries and include an additional interaction term in equation 11 with dummy variable variable for high-income countries.

Table 2 shows that the effects of imports on dump prevalence are only positive and significant in lower-income countries. In fact, the coefficient on high-income countries is negative, and highly significant in the specification where the dependent variable is in logs.

7 Conclusion

In this paper, we develop a methodology to generate globally representative time-series estimates of the prevalence of open dumps. We use crowdsourced, non-representative training data to train a biased machine learning model that predicts dumps based on satellite imagery. We then use active sampling methods to choose an optimal subset of predictions to verify using high-resolution imagery. We choose these predictions to make our estimates of open dump prevalence as precise as possible, and we come up with a simple formula to characterize the efficiency gains unlocked by our method, relative to uniform sampling, in the presence of model miscalibration. Our methods can be widely applied to other types of untracked and difficult-to-detect phenomena.

We use our data on dumps to present new descriptive facts about the global distribution of dumps, as well as causal evidence on the relationship between waste imports and open dumps. We demonstrate the existence of an environmental Kuznets curve for open dumps. We also show that dumps are concentrated in places with high population density, especially in low-income countries. A high fraction of dumps (10%) experience frequent fires — at least one a year.

Turning to our analysis of trade flows, we see a large increase in dumps globally in the year that China banned imports of several categories of waste, including plastics, paper, minerals, and textiles. The increase is concentrated in low-income countries, and countries that are net-importers of waste. This provides evidence that a lot of exported recycled materials are ultimately discarded in poorly managed dumps in low-income countries.

There is plenty of room for further research on this topic. Our data and methods should be useful to more precisely measure fires, air pollution, and fishing impacts associated with informal waste sites, which will lead to an improved understanding of plastic waste impacts on human health and ecosystem services. These are urgent research questions, given the long lifetimes of plastic and the rapid growth rate of plastic production. More work needs to be done on understanding the economic impacts of dumps, and the recycling jobs associated with imported waste as well. Dumps may provide jobs to informal workers, but they also have a high opportunity cost to the extent that they are located in dense and rapidly growing

cities.

Our work is also relevant for multilateral environmental and trade policy. In 2019, 187 countries signed an amendment limiting exports of plastic waste to developing countries, but several major economies, including the US, did not participate (United States Environmental Protection Agency 2024). Our results on the impacts of these international trade flows can inform future negotiations.

References

Alix-Garcia, Jennifer, and Daniel L. Millimet. 2023. "Remotely Incorrect? Accounting for Nonclassical Measurement Error in Satellite Data on Deforestation." *Journal of the Association of Environmental and Resource Economists* 10 (5): 1335–1367. https://doi.org/10.1086/723723.

Angelopoulos, Anastasios N., Stephen Bates, Clara Fannjiang, Michael I. Jordan, and Tijana Zrnic. 2023. *Prediction-Powered Inference.* arXiv: 2301.09633 [stat.ML]. https://arxiv.org/abs/2301.09633.

Bai, Yanbing, Xinyi Wu, Lai Xu, Jihan Pei, Erick Mas, and Shunichi Koshimura. 2024. *Towards Efficient Disaster Response via Cost-effective Unbiased Class Rate Estimation through Neyman Allocation Stratified Sampling Active Learning.* Accessed December 12, 2025. http://arxiv.org/abs/2405.17734.

Biermann, Lauren, Daniel Clewley, Victor Martinez-Vicente, and Konstantinos Topouzelis. 2020. "Finding Plastic Patches in Coastal Waters using Optical Satellite Data." *Scientific Reports* 10 (1): 5364. https://doi.org/10.1038/s41598-020-62298-z.

Brown, Christopher F., Steven P. Brumby, Brookie Guzder-Williams, Tanya Birch, Samantha Brooks Hyde, Joseph Mazzariello, Wanda Czerwinski, et al. 2022. "Dynamic World, Near real-time global 10 m land use land cover mapping." *Scientific Data* 9 (1): 251. Accessed December 12, 2025. https://www.nature.com/articles/s41597-022-01307-4.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2021. *Difference-in-Differences with a Continuous Treatment.* https://arxiv.org/abs/2107.02637.

Carlson, Jacob, and Melissa Dell. 2025. *A Unifying Framework for Robust and Efficient Inference with Unstructured Data.* Accessed May 15, 2025. http://arxiv.org/abs/2505.00282.

Center for International Earth Science Information Network. 2021. (Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 3).

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 785–794. KDD '16. ACM. http://dx.doi.org/10.1145/2939672.2939785.

Chichilnisky, Graciela. 1994. "North-South Trade and the Global Environment." *The American Economic Review* 84 (4): 851–874. Accessed December 27, 2019. https://www.jstor.org/stable/2118034.

Climate Trace. 2023. *Emissions Map.* Accessed September 28, 2023. climatetrace.org.

Clopper, C. J., and E. S. Pearson. 1934. "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial." *Biometrika* 26 (4): 404–413. Accessed January 15, 2026. https://www.jstor.org/stable/2331986.

Congress of the Philippines. 2001. *R.A. 9003.* Accessed December 12, 2025. https://lawphil.net/statutes/repacts/ra2001/ra_9003_2001.html.

Copeland, Brian R., Joseph S. Shapiro, and M. Scott Taylor. 2022. "Chapter 2 - Globalization and the environment." In *Handbook of International Economics,* edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 5:61–146. Handbook of International Economics: International Trade, Volume 5. Elsevier. Accessed March 17, 2023. https://www.sciencedirect.com/science/article/pii/S1573440422000028.

Copeland, Brian R., Joseph S. Shapiro, and M. Scott Taylor. 2021. *Globalization and the Environment.* Working Paper 28797. National Bureau of Economic Research. Accessed November 8, 2021. https://www.nber.org/papers/w28797.

Cózar, Andrés, Manuel Arias, Giuseppe Suaria, Josué Viejo, Stefano Aliani, Aristeidis Koutroulis, James Delaney, et al. 2024. "Proof of concept for a new sensor to monitor marine litter from space." *Nature Communications* 15 (1): 4637. Accessed December 12, 2025. https://www.nature.com/articles/s41467-024-48674-7.

D-Waste. 2014. *Waste Atlas.* Accessed March 17, 2023. http://www.atlas.d-waste.com/.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171–198. Accessed January 15, 2026. https://www.aeaweb.org/articles?id=10.1257/jep.30.4.171.

Earth Science Data Systems, NASA. 2024. *Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11 | NASA Earthdata.* https://www.earthdata.nasa.gov/data/catalog/sedac-ciesin-sedac-gpwv4-popcount-r11-4.11.

Fowlie, Meredith, Edward Rubin, and Reed Walker. 2019. "Bringing Satellite-Based Air Quality Estimates Down to Earth." *AEA Papers and Proceedings* 109:283–288. Accessed September 12, 2022. https://www.aeaweb.org/articles?id=10.1257/pandp.20191064.

Gaulier, Guillaume, and Soledad Zignago. 2010. "BACI: International Trade Database at the Product-Level. The 1994-2007 Version." *CEPII Working Paper,* https://www.cepii.fr/CEPII/fr/publications/wp/abstract.asp?NoDoc=2726.

Greenpeace. 2018. *THE RECYCLING MYTH: Malaysia and the Broken Global Recycling System.* Accessed March 17, 2023. https://www.greenpeace.org/malaysia/publication/1233/the-recycling-myth.

———. 2021. *Investigation finds plastic from the UK and Germany illegally dumped in Turkey.* Accessed March 17, 2023. https://www.greenpeace.org/international/press-release/47759/investigation-finds-plastic-from-the-uk-and-germany-illegally-dumped-in-turkey.

Grossman, Gene M., and Alan B. Krueger. 1995. "Economic Growth and the Environment." *The Quarterly Journal of Economics* 110 (2): 353–377. Accessed December 27, 2019. https://academic.oup.com/qje/article/110/2/353/1826336.

Guo, Liang, W. D. Walls, and Xiaoli Zheng. 2023. "Waste Import Bans and Environmental Quality: Evidence from China's Electronic Waste Disposal Towns." *Environmental and Resource Economics* 85 (1): 65–108. Accessed May 21, 2025. https://doi.org/10.1007/s10640-022-00756-0.

Hamilton, Max, Jinlin Lai, Wenlong Zhao, Subhransu Maji, and Daniel Sheldon. 2025. *Active Measurement: Efficient Estimation at Scale.* Accessed December 12, 2025. http://arxiv.org/abs/2507.01372.

Hanna, Rema, and Paulina Oliva. 2015. "The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City." *Journal of Public Economics* 122:68–79. Accessed December 27, 2019. https://linkinghub.elsevier.com/retrieve/pii/S0047272714002096.

Hoornweg, Daniel, Perinaz Bhada-Tata, and Chris Kennedy. 2013. "Environment: Waste production must peak this century." *Nature News* 502 (7473): 615. Accessed December 20, 2018. http://www.nature.com/news/environment-waste-production-must-peak-this-century-1.14032.

Jain, Meha. 2020. "The Benefits and Pitfalls of Using Satellite Data for Causal Inference." *Review of Environmental Economics and Policy* 14 (1): 157–169. Accessed October 30, 2020. https://academic.oup.com/reep/article/14/1/157/5735430.

Jambeck, Jenna R., Roland Geyer, Chris Wilcox, Theodore R. Siegler, Miriam Perryman, Anthony Andrady, Ramani Narayan, and Kara Lavender Law. 2015. "Plastic waste inputs from land into the ocean." *Science* 347 (6223): 768–771. Accessed February 4, 2020. https://science.sciencemag.org/content/347/6223/768.

Kaza, Silpa, Lisa C. Yao, Perinaz Bhada-Tata, and Frank Van Woerden. 2018. *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050.* World Bank. Book. https://hdl.handle.net/10986/30317.

Kellenberg, Derek. 2015. "The Economics of the International Trade of Waste." *Annual Review of Resource Economics* 7 (1): 109–125. Accessed February 4, 2020. https://doi.org/10.1146/annurev-resource-100913-012639.

Kluger, Dan M., Kerri Lu, Tijana Zrnic, Sherrie Wang, and Stephen Bates. 2025. *Prediction-Powered Inference with Imputed Covariates and Nonuniform Sampling.* Accessed May 15, 2025. http://arxiv.org/abs/2501.18577.

Kruse, Caleb, Edward Boyda, Sully Chen, Krishna Karra, Tristan Bou-Nahra, Dan Hammer, Jennifer Mathis, Taylor Maddalene, Jenna Jambeck, and Fabien Laurier. 2023. "Satellite monitoring of terrestrial plastic waste." *PLOS ONE* 18 (1): 1–20. https://doi.org/10.1371/journal.pone.0278997.

Law, Kara Lavender, Natalie Starr, Theodore R. Siegler, Jenna R. Jambeck, Nicholas J. Mallos, and George H. Leonard. 2020. "The United States' contribution of plastic waste to land and ocean." *Science Advances* 6 (44): eabd0288. Accessed March 17, 2023. https://www.science.org/doi/10.1126/sciadv.abd0288.

Lebreton, Laurent C. M., Joost van der Zwet, Jan-Willem Damsteeg, Boyan Slat, Anthony Andrady, and Julia Reisser. 2017. "River plastic emissions to the world's oceans." *Nature Communications* 8 (1): 15611. Accessed November 5, 2021. https://www.nature.com/articles/ncomms15611.

Lu, Kerri, Dan M. Kluger, Stephen Bates, and Sherrie Wang. 2025. *Regression coefficient estimation from remote sensing maps.* Accessed May 15, 2025. http://arxiv.org/abs/2407.13659.

Meijer, Lourens J. J., Tim van Emmerik, Ruud van der Ent, Christian Schmidt, and Laurent Lebreton. 2021. "More than 1000 rivers account for 80% of global riverine plastic emissions into the ocean." *Science Advances* 7 (18): eaaz5803. Accessed March 17, 2023. https://www.science.org/doi/10.1126/sciadv.aaz5803.

National Geospatial-Intelligence Agency. 2019. *World Port Index.* Accessed May 6, 2020. https://msi.nga.mil/Publications/WPI.

Neyman, Jerzy. 1934. "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection." *Journal of the Royal Statistical Society* 97 (4): 558. Accessed March 17, 2023. https://www.jstor.org/stable/10.2307/2342192?origin=crossref.

Pasquarella, Valerie J., Christopher F. Brown, Wanda Czerwinski, and William J. Rucklidge. 2023. "Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning." In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 2125–2135. Accessed December 12, 2025. https://ieeexplore.ieee.org/document/10208818.

Pathak, Gauri, Mark Nichter, Anita Hardon, Eileen Moyer, Aarti Latkar, Joseph Simbaya, Diana Pakasi, Efenita Taqueban, and Jessica Love. 2023. "Plastic pollution and the open burning of plastic wastes." *Global Environmental Change* 80:102648. https://www.sciencedirect.com/science/article/pii/S0959378023000146.

Proctor, Jonathan, Tamma Carleton, and Sandy Sum. 2023. "Parameter Recovery Using Remotely Sensed Variables." *NBER Working Paper.*

Rambachan, Ashesh, Rahul Singh, and Davide Viviano. 2025. *Program Evaluation with Remotely Sensed Outcomes.* Accessed May 15, 2025. http://arxiv.org/abs/2411.10959.

Ratledge, Nathan, Gabriel Cadamuro, Brandon De la Cuesta, Matthieu Stigler, and Marshall Burke. 2021. *Using Satellite Imagery and Machine Learning to Estimate the Livelihood Impact of Electricity Access.* Working Paper. Accessed September 12, 2022. https://www.nber.org/papers/w29237.

Rochman, Chelsea M., Mark Anthony Browne, A. J. Underwood, Jan A. van Franeker, Richard C. Thompson, and Linda A. Amaral-Zettler. 2016. "The ecological impacts of marine debris: unraveling the demonstrated evidence from what is perceived." *Ecology* 97 (2): 302–312. eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/14-2070.1. https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/14-2070.1.

Sanford, Luke C., Megan Ayers, Matthew Gordon, and Eliana Stone. 2025. *Adversarial Debiasing for Unbiased Parameter Recovery.* Accessed May 21, 2025. http://arxiv.org/abs/2502.12323.

Schroeder, Wilfrid, Louis Giglio, and Joanne Hall. 2025. "Collection 2 Visible Infrared Imaging Radiometer Suite (VIIRS) 375-m Active Fire Product User's Guide Version 1.2."

Sigman, Hilary, and Rachel Strow. 2024. "China's Waste Import Ban and US Solid Waste Management: Effects of the Loss of a Waste Haven." *Journal of the Association of Environmental and Resource Economists* 11 (6): 1527–1557. Accessed May 21, 2025. https://www.journals.uchicago.edu/doi/10.1086/729899.

Stehman, Stephen V. 2012. "Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change." *Remote Sensing Letters* 3 (2): 111–120. Accessed December 12, 2025. https://doi.org/10.1080/01431161.2010.541950.

Tanaka, Shinsuke, Kensuke Teshima, and Eric Verhoogen. 2022. "North-South Displacement Effects of Environmental Regulation: The Case of Battery Recycling." *American Economic Review: Insights* 4 (3): 271–288. Accessed March 17, 2023. https://www.aeaweb.org/articles?id=10.1257/aeri.20210201.

Thakur, Prakrati. 2022. "Welfare Effects of International Trade in Waste." *SSRN Electronic Journal,* accessed March 17, 2023. https://www.ssrn.com/abstract=4216639.

Tuia, Devis, FrÉdÉric Ratle, Fabio Pacifici, Mikhail F. Kanevski, and William J. Emery. 2009. "Active Learning Methods for Remote Sensing Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 47 (7): 2218–2232.

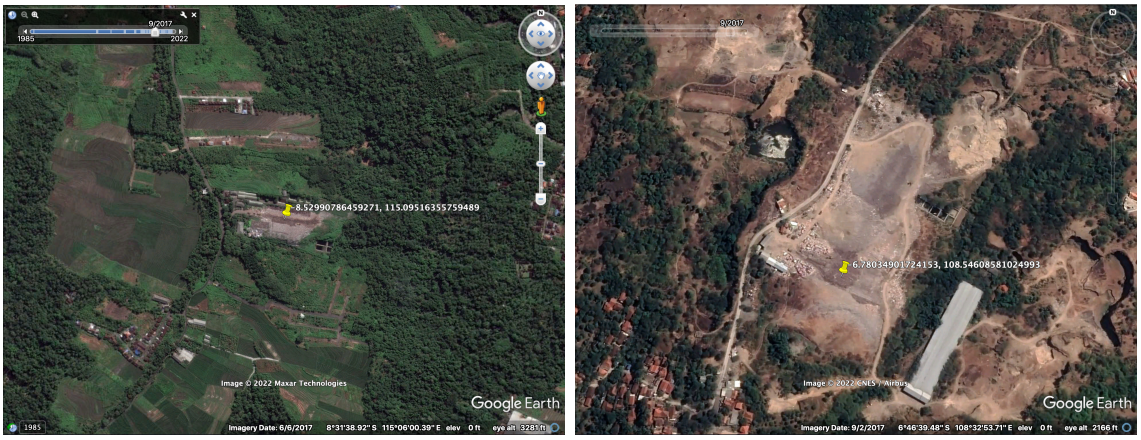
United States Environmental Protection Agency. 2024. "New International Requirements for the Export and Import of Plastic Recyclables and Waste." Accessed March 25, 2025. https://www.epa.gov/

hwgenerators/new-international-requirements-export-and-import-plastic-recyclables-and-waste#:~:text=In%20May%202019%2C%20187%20countries,country%20and%20any%20transit%20countries..

Zrnic, Tijana, and Emmanuel J. Candès. 2024. *Active Statistical Inference.* Accessed May 15, 2025. http://arxiv.org/abs/2403.03208.

A Additional Data Details

Figure A1: High-Resolution Satellite Data Verification Example



Notes: Figures show two examples of waste sites in Google Earth high-resolution satellite imagery.

A.1 Satellite Data Details

We collect the following satellite and other data for each pixel in our sample.

- Sentinel-1: All bands as well as DpRVIVV, VHVVR, VVVHR indices.
- Sentinel-2: All bands (B1-B12) as well as NDVI, BSI, NDWI, UI, MNDVI, MNDWI, CSI, DBSI, NDSWIR, NormNIR, NormG, NormR, SWIR1SWIR2, NIRNIR2, RE1RE2, RE2RE3, BG, BR, and GR.
- Other Data:
  - Distance to ports in meters
  - Population for 2010, 2015, and 2020 very close and close to each pixel
  - Probability of complete coverage by water, trees, grass, flooded vegetation, crops, shrub and scrub, built, bare, and snow and ice land uses from Dynamic World

Sentinel-1 Indices

- **DpRVIVV (Dual-Polarized Radar Vegetation Index VV):** $\text{DpRVIVV} = \frac{4\times\text{VH}}{\text{VV}+\text{VH}}$. A radar-based vegetation index that uses the VH and VV polarizations to estimate vegetation density.
- **VHVVR (VH/VV Ratio):** $\text{VHVVR} = \frac{\text{VH}}{\text{VV}}$. Ratio of cross-polarized (VH) to co-polarized (VV) backscatter, sensitive to vegetation structure and volume scattering.
- **VVVHR (VV/VH Ratio):** $\text{VVVHR} = \frac{\text{VV}}{\text{VH}}$. Inverse ratio of cross-polarized to co-polarized backscatter, providing complementary information to VHVVR.

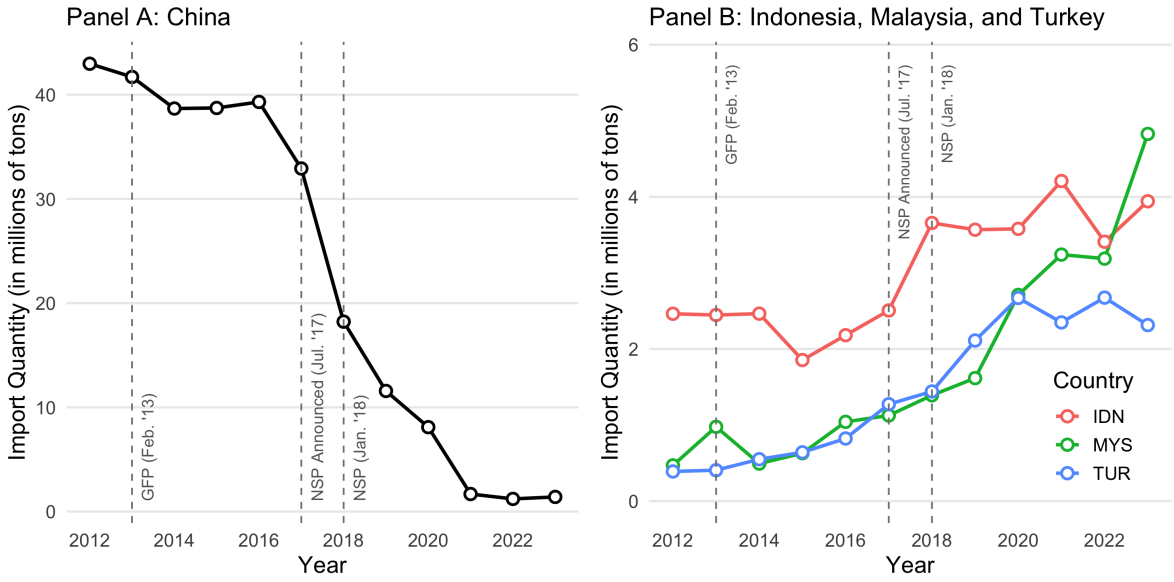
Sentinel-2 Indices

- **NDVI (Normalized Difference Vegetation Index):** $\text{NDVI} = \frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{Red}} = \frac{\text{B8}-\text{B4}}{\text{B8}+\text{B4}}$. Quantifies vegetation by measuring the difference between near-infrared (strongly reflected by vegetation) and red light (absorbed by vegetation).
- **BSI (Bare Soil Index):** $\text{BSI} = \frac{(\text{SWIR1}+\text{Red})-(\text{NIR}+\text{Blue})}{(\text{SWIR1}+\text{Red})+(\text{NIR}+\text{Blue})} = \frac{(\text{B11}+\text{B4})-(\text{B8}+\text{B2})}{(\text{B11}+\text{B4})+(\text{B8}+\text{B2})}$. Designed to identify bare soil areas and differentiate them from vegetation and water.
- **NDWI (Normalized Difference Water Index):** $\text{NDWI} = \frac{\text{Green}-\text{NIR}}{\text{Green}+\text{NIR}} = \frac{\text{B3}-\text{B8}}{\text{B3}+\text{B8}}$. Maximizes reflectance of water by using green wavelengths and minimizes low reflectance of NIR by water features.
- **UI (Urban Index):** $\text{UI} = \frac{\text{SWIR2}-\text{NIR}}{\text{SWIR2}+\text{NIR}} = \frac{\text{B12}-\text{B8}}{\text{B12}+\text{B8}}$. Highlights urban areas by exploiting the high reflectance of built-up areas in SWIR2 compared to NIR.
- **MNDVI (Modified NDVI):** $\text{MNDVI} = \frac{\text{NIR}-\text{SWIR2}}{\text{NIR}+\text{SWIR2}} = \frac{\text{B8}-\text{B12}}{\text{B8}+\text{B12}}$. A variant of NDVI that uses SWIR2 instead of red, providing different sensitivity to vegetation properties.
- **MNDWI (Modified Normalized Difference Water Index):** $\text{MNDWI} = \frac{\text{Green}-\text{SWIR1}}{\text{Green}+\text{SWIR1}} = \frac{\text{B3}-\text{B11}}{\text{B3}+\text{B11}}$. Enhances open water features while suppressing noise from built-up land, vegetation, and soil.
- **CSI (Char Soil Index):** $\text{CSI} = \frac{\text{NIR}}{\text{SWIR2}} = \frac{\text{B8}}{\text{B12}}$. Ratio that helps identify charred or burned soils and can detect recent fire events.
- **DBSI (Dry Bareness Index):** $\text{DBSI} = \frac{\text{SWIR1}-\text{Green}}{\text{SWIR1}+\text{Green}} - \frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{Red}} = \frac{\text{B11}-\text{B3}}{\text{B11}+\text{B3}} - \frac{\text{B8}-\text{B4}}{\text{B8}+\text{B4}}$. Designed to detect dry barren areas by combining a moisture-sensitive index with NDVI.
- **NDSWIR (Normalized Difference SWIR):** $\text{NDSWIR} = \frac{\text{NIR}-\text{SWIR1}}{\text{NIR}+\text{SWIR1}} = \frac{\text{B8}-\text{B11}}{\text{B8}+\text{B11}}$. Normalized difference between NIR and SWIR1, sensitive to vegetation moisture content.
- **NormNIR (Normalized NIR):** $\text{NormNIR} = \frac{\text{NIR}}{\text{NIR}+\text{Green}+\text{Red}} = \frac{\text{B8}}{\text{B8}+\text{B3}+\text{B4}}$. Proportion of NIR reflectance in the total reflectance of visible and near-infrared bands.

- **NormG (Normalized Green):** $\text{NormG} = \frac{\text{Green}}{\text{NIR+Green+Red}} = \frac{\text{B3}}{\text{B8+B3+B4}}$. Proportion of green reflectance in the total reflectance of visible and near-infrared bands.
- **NormR (Normalized Red):** $\text{NormR} = \frac{\text{Red}}{\text{NIR+Green+Red}} = \frac{\text{B4}}{\text{B8+B3+B4}}$. Proportion of red reflectance in the total reflectance of visible and near-infrared bands.
- **SWIR1SWIR2 (SWIR1/SWIR2 Ratio):** $\text{SWIR1SWIR2} = \frac{\text{SWIR1}}{\text{SWIR2}} = \frac{\text{B11}}{\text{B12}}$. Ratio between two SWIR bands that can highlight certain mineral compositions and soil properties.
- **NIRNIR2 (NIR/NIR2 Ratio):** $\text{NIRNIR2} = \frac{\text{NIR}}{\text{NIR2}} = \frac{\text{B8}}{\text{B8A}}$. Ratio between two near-infrared bands that can provide subtle information about vegetation structure.
- **RE1RE2 (Red Edge 1/Red Edge 2 Ratio):** $\text{RE1RE2} = \frac{\text{RE1}}{\text{RE2}} = \frac{\text{B5}}{\text{B6}}$. Ratio between two red edge bands sensitive to chlorophyll content and vegetation stress.
- **RE2RE3 (Red Edge 2/Red Edge 3 Ratio):** $\text{RE2RE3} = \frac{\text{RE2}}{\text{RE3}} = \frac{\text{B6}}{\text{B7}}$. Ratio between two red edge bands that can detect subtle changes in vegetation health.
- **BG (Blue/Green Ratio):** $\text{BG} = \frac{\text{Blue}}{\text{Green}} = \frac{\text{B2}}{\text{B3}}$. Ratio between blue and green reflectance, useful for water quality assessment and algal bloom detection.
- **BR (Blue/Red Ratio):** $\text{BR} = \frac{\text{Blue}}{\text{Red}} = \frac{\text{B2}}{\text{B4}}$. Ratio between blue and red reflectance, sensitive to atmospheric scattering and certain water properties.
- **GR (Green/Red Ratio):** $\text{GR} = \frac{\text{Green}}{\text{Red}} = \frac{\text{B3}}{\text{B4}}$. Ratio between green and red reflectance, related to chlorophyll absorption and vegetation properties.

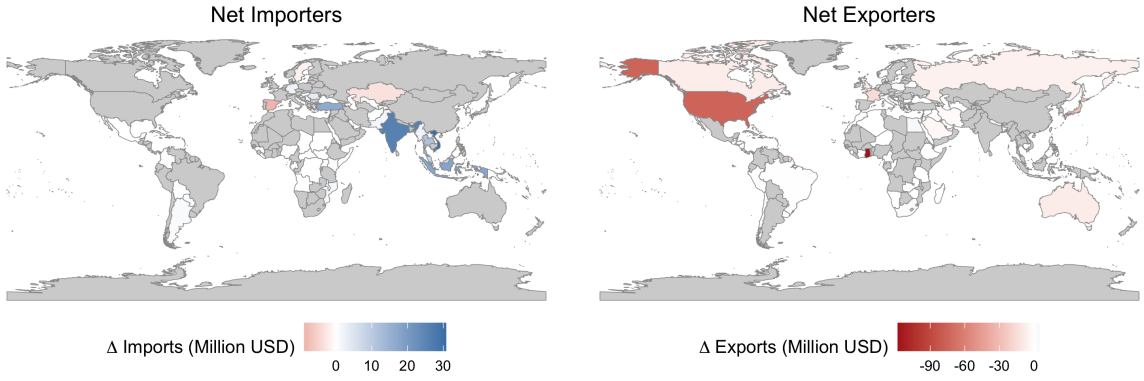
B Additional Background

Figure A2: Plastic Scrap Imports to China and Indonesia



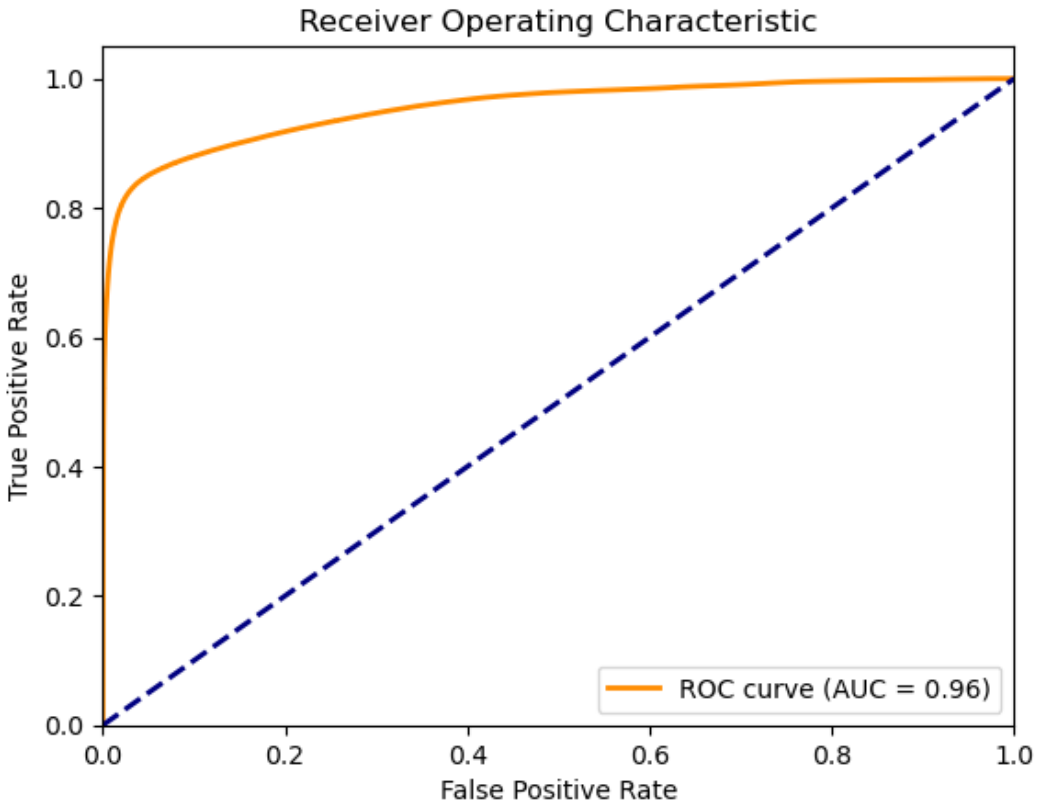
Notes: Figures show imports of categories regulated by the National Sword Policy in China from the UN Comtrade database for China (left panel) and Indonesia, Malaysia, and Turkey (right panel). Both panels also show the Green Fence Policy (February 2013), the announcement of the National Sword Policy (July 2017), and the effective date of the National Sword Policy (January 2018).

Figure A3: Net Importers and Exporters, 2012-2023



Notes: Figures show countries classified as net importers (imports > exports, left panel) and net exporters (right panel) based on their average trade quantities from 2012 to 2023. Shading shows changes in imports before (2016) and after (2019) China's National Sword Policy.

Figure A4: Machine Learning Model Receiver-Operator Curve (ROC)



Notes: The receiver-operator curve (ROC) for our final model described in the text. This curve traces the tradeoff between increasing true positive and false positive rates as the classification threshold increases from the bottom left to the top right. The AUC is the total area under this curve and ranges between 0.5 and 1.

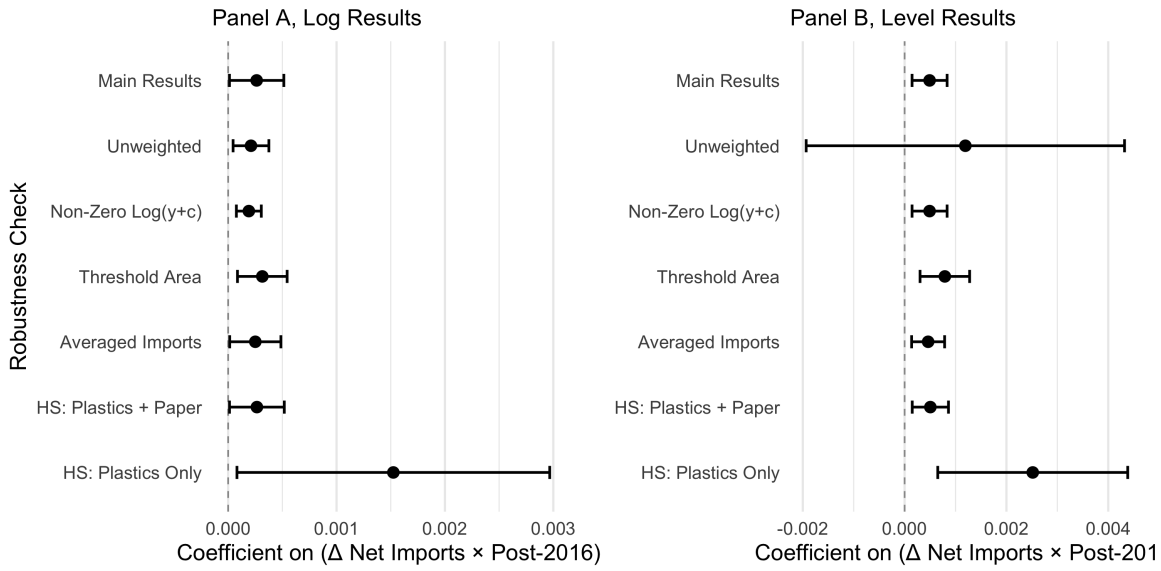
C Additional Results

Table A1: Waste Imports and Open Dump Area, 2012-2023 (All Countries, Excluding China)

	Dependent variable:	
	Log(Dump Area Fraction)	Dump Area Fraction (x Billions)
	(1)	(2)
Log(Import Quantity)	0.454***	
	(0.083)	
	p = 0.00001	
Import Quantity		0.0003*
		(0.0002)
		p = 0.094
Elasticity	0.454	0.249
Number of Countries	50	117
Observations	501	1,416

Note: Table shows results of a simple panel regression of open dump area on waste import quantities for 2012-2023. Column (1) shows log results for countries with non-zero open dump area, while Column (2) shows level results. All columns are weighted by the variance the open dump area. Import quantities in 1000s of tons. SEs clustered by country. *p< 0.1; **p< 0.05; ***p< 0.01.

Figure A5: Effect of Change in Net Imports on Open Dump Area, Robustness



Notes: Figure shows results of regression according to equation 11 for 2012-2013. Import quantities in 1000s of tons. SEs clustered by country. Bars show 95% confidence intervals.