Open Dumps and the Global Trade in Garbage

Matthew Gordon, Anna Papp, and Monica Shandal*

Click Here for Latest Version

Abstract

A large fraction of the world's solid waste is disposed of in open dumps, generating serious environmental and health externalities. These effects are difficult to study as data on waste are scarce where mismanagement is worst. We develop a way to detect open dumps, portable to other rare land uses, that uses crowdsourced data, satellite imagery, machine learning, and active sampling. We combine these methods to create the first globally representative human-verified time series data on open dumps. The data show that open dumps are concentrated in densely populated areas of low- and middle-income countries, they are frequently on fire, and they expand substantially in response to increases in waste imports. Analysis of Java, Indonesia, where we observe nearly the universe of open dumps, shows that higher waste imports following China's 2018 import ban generated just over $12,000 in consumer surplus per premature death caused by the increase in garbage-fire pollution.

JEL: C45, C81, Q53, O17

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

*Gordon: Paris School of Economics and École Normale Supérieure (matthew.gordon@psemail.eu); Papp: University of California, Santa Barbara (apapp@bren.ucsb.edu); Shandal: University of California, Santa Cruz (mshandal@ucsc.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, ENS de Lyon, the WCERE, the Tweeds Workshop, the IPWSD, 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. Refine.ink helped with manuscript preparation.

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 post-consumer plastic and paper have more than doubled.[1] 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 whether waste imports are recycled or discarded in environmentally harmful ways.

An estimated 90% of domestic waste is disposed of in open dumps or burned in low-income countries (Kaza et al. 2018). Open dumps cause substantial environmental and social harm. 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.[2] Dumped waste is also a well-documented source of plastic leakage to marine ecosystems, 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 remain 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 (Jambeck et al. 2015; Law et al. 2020; 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.

In this paper, we devise an approach to collect the data required to begin to study the scale and impacts of open dumping. We start by collecting a crowdsourced sample of non-representative open dumps. We use these examples to train a machine learning model that predicts dumps using satellite data. Finally, we use the model's predictions to select and manually verify an approximately optimally chosen subsample of predictions. These verified predictions form the basis of a global time-series dataset of open dump prevalence. This approach allows us to gather globally representative data, even when the machine learning

[1] Calculated from BACI trade data (Gaulier and Zignago 2010). We sum trade across all country pairs in the two ban categories used in our subsequent analyses —plastic waste and scrap (HS 3915) and unsorted recovered paper (HS 4707.90)—in 1995 and 2024. Global trade in these categories grew from 8.2 to 18.2 million metric tons, a 2.2-fold increase.

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

model's predictions are arbitrarily biased or miscalibrated, though better model accuracy leads to more precise estimation. Our data shows that open-dumps tend to locate in dense cities in low and middle-income countries, and that they are frequently on fire, generating substantial amounts of air pollution. We analyze the welfare impacts of a shock to waste imports, exploiting a natural experiment resulting from the 2018 China waste import ban. In Indonesia, we estimate that waste imports generate about $12,000 in consumer surplus for every premature death resulting from the air pollution.

The first step of our pipeline is collecting non-representative 'seed' data. For this, we partnered with an international network of NGOs, researchers, and activists to crowdsource an initial set of open-air waste sites. This initiative, which we called "The Atlas of Plastic Waste", yielded over 270 submissions across 24 countries. We supplement our crowdsourced data collection effort with existing public repositories containing examples of open dumps. We verify the extent and location of all sites using high-resolution satellite imagery.

Next, we use these examples to train a machine learning model that can identify open dumps using 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 a gradient-boosted tree model that differentiates open dumps from other land uses with high accuracy. The area under the receiver-operating characteristics curve (AUC) in our validation sample is 0.96 out of 1, and the balanced accuracy is 90%, demonstrating that our model can find a distinctive spectral signature associated with open dumps.

Since our training data are heavily selected, and dumps are a rare land use class, even small biases in our model's predictions when deployed out-of-sample could have an outsized relative impact on inferences drawn about the prevalence of open dumps. Thus, rather than use our model's predictions directly, we use them to inform a sampling and verification procedure that aims to minimize the variance of estimates of dump prevalence and changes over time. This procedure relies on optimal sampling methods that devote relatively more sampling effort to observations that our model is most uncertain about (Neyman 1934; Zrnic and Candès 2024). For selected observations, we manually verify the existence of dumps using high-resolution satellite images. This results in globally representative, time-series data of open dump prevalence.

We use this 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 (LMICs), whereas in high-income countries dumps tend to be in slightly less dense areas. We also find that fires in open dumps are common: 12.5% of dumps in LMICs experience a fire at least once a year. The latter two facts—frequent fires and dense nearby populations—highlight some of the public health concerns about open dumping, and raise the question about the welfare consequences of shipping waste to LMICs.

We then investigate whether waste imports have a causal effect on open dumping in receiving countries. Our empirical strategy exploits a shock to the global trade network caused by China's sudden ban on waste imports, announced in 2017 and enforced in 2018. China previously accounted for 50-90% of global waste imports, depending on the category. After the ban, shippers diverted a large fraction of this waste to countries in Southeast Asia and the rest of the world. We use a shift-share instrumental variables strategy that exploits this shock, instrumenting observed changes in each country's waste imports with the diversion it would have received had every exporter reallocated its pre-2018 China export volumes to its other trade partners. We find an increase of 0.111-0.179 m^2 open dump area per kilogram of additional net imports, an elasticity of 1.2-1.5 that implies a substantial share of imported waste either ends up in open dumps or overwhelms domestic waste and recycling facilities. These results are robust to a range of alternative specifications, and placebo tests on commodities and policy timing show precise null effects.

Finally, we look at three smaller geographies—the island of Java in Indonesia, the US state of California, and three states in Southeast Brazil—where more exhaustive search methods become feasible to attempt to catalog the universe of dumps. In all three locations, we find dumps that are not present in administrative or other data sources. This more exhaustive data collection also allows for the more precise identification of the externalities associated with dumps. In Java, we find that the surge in waste imports after the China waste ban increased $PM_{2.5}$ by nearly 6% close to dumps resulting from an increase in garbage fires. We can rule out large increases in pollution close to dumps in California—if anything we find a modest decrease. A back-of-the-envelope calculation for Indonesia estimates that the ratio of consumer surplus gained from these imports amounts to $12,672 for each premature death caused by the resulting air pollution. This ratio is far lower than most estimates of the value of a statistical life, casting doubt on the efficiency of the international trade in waste.

Our data collection approach builds on the growing literature on using machine learning and satellite data for policy evaluation (see Donaldson and Storeygard 2016 and Jain 2020 for reviews). Previous work has documented the potential biases that can result from using machine learned proxies for economic or environmental variables in causal inference (Fowlie

et al. 2019; Alix-Garcia and Millimet 2023; Gordon et al. 2026). 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 (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 a function of observables. Our approach, which is informed by the active statistical learning literature in computer science (Zrnic and Candès 2024) is fundamentally different in that we study how to choose which data to label when starting with selected non-representative data. While these sampling techniques are common in the active sampling literature in computer science and remote sensing (Tuia et al. 2009; Stehman 2012; Zrnic and Candès 2024; Gordon et al. 2026), they have not been widely applied to study policy questions in the social sciences. The method results in unbiased and precise estimates of outcomes of interest. These methods are widely applicable to other hard-to-study land uses, and more generally, they 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.

We derive a simple way to estimate the ex-post 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. Using the quasi-natural experiment of China's import ban on various waste categories, 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 waste trade. Kellenberg (2015) shows that the relative stringency of environmental regulations is associated with trade flows—materials move 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. A few papers have looked at the impacts of the China waste ban on air pollution in China (Guo et al. 2023; Sigman and Strow 2024), but our novel data collection effort allows us to examine this understudied outcome worldwide and provide evidence for the pollution havens hypothesis on a global scale. In Indonesia we find that this influx of waste led to more fires in open dumps and a large increase in pollution, and simple calculations suggest that the value of these waste imports is low relative to the health costs of the resulting pollution.

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 cross-country analysis of the China waste import ban. Section 7 conducts more detailed analyses on the universe of dumps in Java, California, and Southeast Brazil, and estimates the consumer surplus and lives lost from the import shock in Indonesia. Finally, Section 8 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). Despite these expenditures, an estimated 90% of solid waste in low-income countries is disposed of in unregulated open-air dumps or burned, though the authors of this fact caution that the underlying data, "should be considered with a degree of caution due to undefined words or phrases, inconsistent or omitted units, or estimates made without basis"

(Kaza et al. 2018).

Open dumps generate severe externalities. Runoff from dumps may lead to surface and ground water contamination. This may also contribute to the increase in ocean plastics which disrupts ecosystems and fisheries. Dumps may also attract disease-spreading vermin and other pests. 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), find 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 (Brooks et al. 2018; Ritchie 2022). 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 quotes above emphasize, no good estimates exist on how much of this material is used productively or discarded.

Complaints about contamination levels in these 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 scrap and unsorted paper, and it 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 do not know whether this is due to a decline in productive recycling activities or the open burning of discarded scraps.

Appendix Figure A1 shows that Chinese imports of plastic and paper scrap – HS 3915 and 4707.90, the two main banned categories we focus on due to their likelihood of ending up in open dumps – started to dramatically decline as soon as the ban was announced, and they fell to nearly zero soon after implementation.[3] 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 A1 shows an increase 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 use this policy change to estimate the elasticity of open dumps to imported waste. This research is relevant 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 Dump Definitions and Data

Our goal is to estimate the prevalence of open dumps across locations and over time. Several countries have statutory definitions of open dumps that refer to waste disposal sites that are not in accordance with regulations.[4] Since regulations vary across countries, we adopt a definition that is consistent globally, can be measured with our data, and captures the key externalities associated with improper waste disposal. We define an open dump as any aggregation of waste covering an area of at least 100m^2 (10 meters by 10 meters, the spatial resolution of the satellite data we use), where waste is directly exposed to the surrounding environment.

To be clear, we measure the surface area of dumps, rather than the volume of waste disposed,

[3]The increased inspections of the 2013 Green Fence Policy did not meaningfully alter these imports.

[4]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).

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, the surface area of a dump determines how much waste is directly exposed to the environment. Additionally, 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, which are a source of frequent fires and plastic leakage into the environment. We exclude construction debris and automobile junkyards, which both have different characteristics from the dumps in our data, however we are not otherwise able to distinguish the material composition or the density of the waste, which we leave for future research. Clearly this definition excludes some forms of waste disposal that generate externalities, however it is still useful to identify the most egregious instances of mismanaged waste and concentrated externalities.

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

3.1 Machine Learning Model Inputs

Unrepresentative Training Data: We initially gathered a set of verified open dumps from a variety of sources, including our own crowdsourced data collection effort. To gather submissions, we enlisted the assistance of an international NGO, the 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[5]. 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). In particular, we include nearly 1,000 additional unique sites from Earthrise's Global Plastic Watch, which also used satellite imagery to find dumps in a number of countries[6].

In order to verify the existence of the dumps and to determine the spatial and temporal extents of the sites, we rely on the extensive library of historical high-resolution satellite imagery in Google Earth Pro. Figures A4, 1, and A2 show examples of open-air dumps in this imagery, which are typically easy to identify. For each site, we look at the first available satellite image for each year between 2017 and 2022, and we draw a polygon around the extent of any exposed waste visible 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 trained a small model on the set of example sites, and then we used the model to make out of sample predictions (see below for details on model training and data used). We then verified 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 valuable training data, especially control sites representing difficult cases that early versions of the model struggled 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 iterations, we have in total nearly 37,000 validated site-year observations across 93 countries and more than 84,000 control site-year observations. Figure A3 shows the distribution of our training data worldwide.

Satellite Data: Our primary sources of data are Sentinel-2 and Sentinel-1 satellite imagery. The European Space Agency has operated both of these satellites with consistent coverage since 2017[7]. Sentinel-2 is an optical satellite constellation with a spatial resolution of 10m and a revisit frequency of 3-5 days. Sentinel-1 is a radar satellite that provides all-weather

[5]Submission form available at: https://www.ban.org/plastic-waste-transparency-project-hub/the-atlas-of-plastic-waste.
[6]Our complete training dataset and list of sources are available from: https://github.com/atlas-of-plastic-waste/atlas-of-plastic-waste.
[7]We begin our satellite analysis in 2017 when Sentinel 2-B was launched.

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 or greenhouses).

For each pixel located within our training data polygons, we download satellite data using Google Earth Engine for each year from 2017–2021, keeping cloud-free, non-water pixels across all images.[8] This results in over 7.7 million 10m pixels.

The features used to train the model include each of the 13 bands of Sentinel-2, many derived indices from these bands, and all Sentinel-1 bands collected in interferometric wide swath mode and derived indices. For a complete list of satellite bands and derived indices, see Appendix Tables A1 and A2. For each band and/or derived index, we take the median, mean, minimum, and maximum value across all the images from each year. We also utilize focal statistics, taking the mean and standard deviation of each band and index in a 3 pixel radius around each pixel to include information on nearby areas' shapes and characteristics. This allows the model to use information on surrounding pixels in classification, much like a convolutional neural network, although we use a model that is simpler and less computationally intensive to train.

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.

Trade Flows: 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).[9] Our main treatment measure uses waste plastics (HS 3915) and unsorted

[8] We use Google's Cloud Score+ and define clear pixels as those with a clear-sky score of at least 65% (Pasquarella et al. 2023). We use Dynamic World to mask out water pixels (Brown et al. 2022).

[9] China banned imports of the following: waste minerals/vanadium slag (HS codes 2619 and 2620), waste plastics (HS code 3915), waste paper (HS code 4707.90), and waste textiles (HS codes 5103, 5104, 5202,

waste paper (HS 4707.90) as these codes map best to the ban, are most likely to end up in open dumps, and account for over 84% of China's pre-ban imported tonnage in banned headings. 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 and while we acknowledge that BACI's reconciliation methods may not address all systematic HS-code misclassifications, it is a step in the right direction. The BACI trade data clearly show a decline in imports in China after the waste ban, and a corresponding increase in many other countries (Appendix Figure A1).

Pollution, Fires, and Weather: We measure ambient fine particulate matter ($PM_{2.5}$) using the satellite-derived surface concentration estimates of Donkelaar et al. (2019), which combine satellite aerosol optical depth retrievals with chemical transport modeling and ground-monitor calibration on a 0.01° (∼1km) global grid. This allows for estimates of particulate matter concentrations in regions where ground monitors are sparse, which includes most of the world. We use the NASA VIIRS Active Fire Detection product to detect fires in open dumps (Schroeder et al. 2025). This product 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. Finally, we use monthly mean wind fields (10m u- and v-components) from the ERA5 reanalysis, available on a 0.25° grid from the Copernicus Climate Data Store, to measure the prevailing wind direction around dump sites (Hersbach et al. 2020).

Administrative Data on Dumps: In Brazil and California, we compare our dump detections to existing administrative datasets. For Brazil, we use the registry of landfills and dumps from the Sistema Nacional de Informações sobre Saneamento (SNIS) for 2017–2022. For the United States, we use the EPA's Landfill Methane Outreach Program (LMOP) database, which reports the location and area of municipal solid waste landfills. These comparisons allow us to assess the coverage of official data and to identify dumps missed by both sources.

GDP and Population: We use World Bank data on GDP per capita (current US$) and country-level population density (World Bank 2024). We classify countries as high-income using the World Bank threshold of $13,935 GDP per capita. We measure population density at the pixel level using the Gridded Population of the World (GPW v4.11) dataset, which provides population estimates on a 1km grid for 2010, 2015, and 2020 (Earth Science Data Systems, NASA 2024).

5505, 6310). See https://www.laregionalagency.us/comprehensive-list-of-banned-materials/.

4 Dump Detection and Active Learning

4.1 Machine Learning Model

We train a gradient boosted tree model using the training sites and satellite data described above (Chen and Guestrin 2016; Guillame-Bert et al. 2023). Gradient boosted tree models are useful for their relative computational efficiency and their robustness to overfitting, given a limited number of hyperparameters available for tuning compared with neural networks. The algorithm 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.

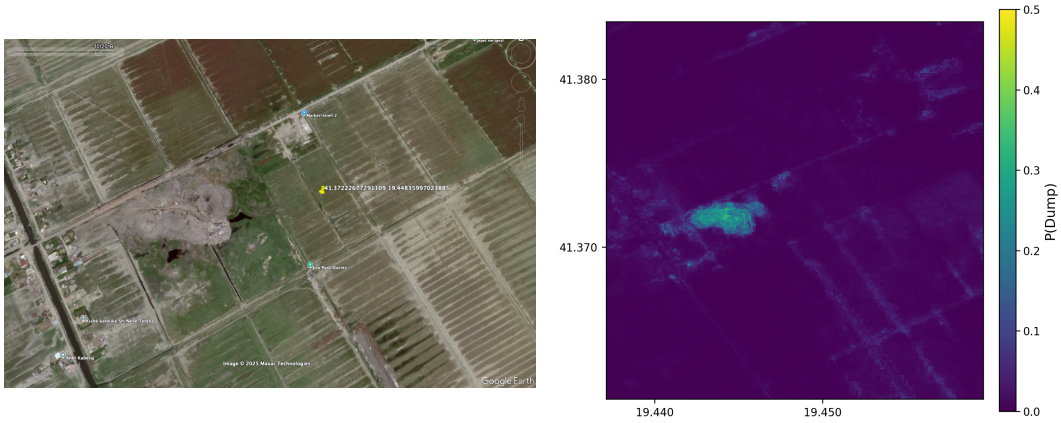
For model development, we split our sites into training, testing, and validation sites. We split our sites into 500 global clusters using a k-means algorithm based on their coordinates. We set aside 20% of our clusters for calculating out-of-sample accuracy metrics, and use the other 80% for choosing hyperparameters. This restricts the model from learning about 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 region or country in many cases.

During training, we assign control pixels a relative weight of 332. This reflects a 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 our early iterative experiments. We choose our model hyperparameters to maximize the cross-validated area under the receiver-operating characteristic curve (AUC) in the training data, a metric that is robust to heavily unbalanced classes. The final model builds an ensemble of 100 decision trees with a maximum depth of 25 splits.[10]

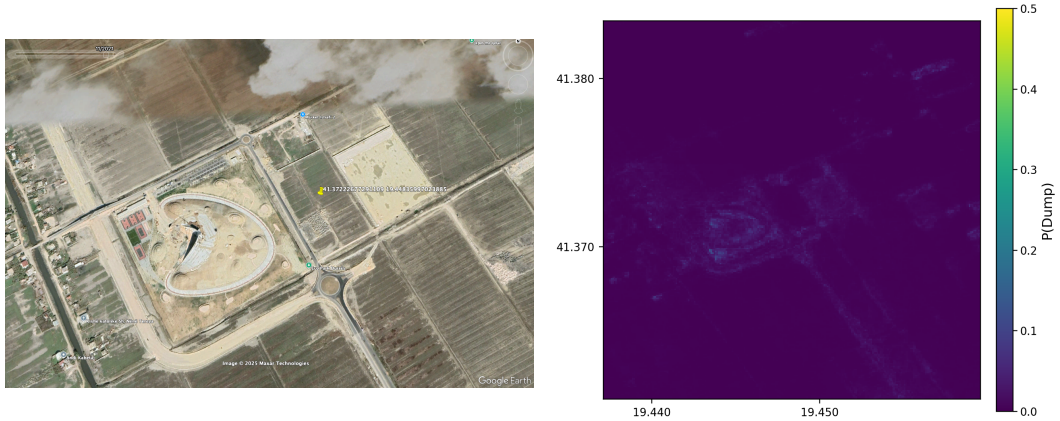
Assessed on the hold-out data, the final model has an AUC of 0.96 out of the maximum possible value of 1. Setting a classification threshold that maximizes accuracy, we calculate balanced accuracy at 90%, precision at 92%, and recall at 85%.

[10]Hyperparameter values tried and chosen are shown in Appendix Table A3. AUC is a metric that incentivizes correctly ranking observations by the probability that they are a dump, but it may lead to miscalibrated probabilities, especially for unlikely cases. We thus apply a sigmoid recalibration to the resulting predictions: $p_i = (1 + \exp(-a \log(\tilde{p}_i/(1-\tilde{p}_i))))^{-1}$ where $\tilde{p}_i$ are the original predictions, and we set the coefficient $a = 3.15$ to minimize log loss in the hold-out sample. This transformation rescales probabilities without changing their relative ranking.

Figure 1: Example Site in Durrës, Albania



(a) 2018



(b) 2021

Notes: Figure shows Google Earth Pro historical satellite imagery (image dates October 2018 and November 2021) and machine learning predictions (probability of each pixel being part of a waste site) for a small dump site near Durrës, Albania that was converted into an eco-park in 2020.

Figure 1 shows high-resolution Google Earth Pro imagery of a dump near Durrës, Albania alongside our model predictions for the years 2018 and 2021. In 2018, the model puts elevated probability on the pixels corresponding to the exposed waste at the site. The site was subsequently capped and redeveloped as a park, and by 2021 exposed waste is no longer visible in the imagery and the model's predicted probabilities decline accordingly. This suggests that the model captures not only the locations of dumps, but also changes in the amount of exposed waste at a site over time.

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 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 do not 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 Approximately Optimal Active Sampling

We now describe our active sampling and verification. Define D_i as an indicator for whether pixel i contains an open dump. Our model outputs $p(k_i)$: the probability that a pixel is a dump, conditional on its associated satellite data features (k_i). We can define the model's prediction errors as $\nu_i = p(k_i) - D_i$. A naive estimate of the prevalence of open dumps (μ) in any subset of N pixels could be:

$$\widehat{\mu}^{naive} = \frac{1}{N} \sum_{i=1}^{N} p(k_i). \tag{1}$$

Given our definition of measurement error, this estimator will equal μ in expectation if $\mathbb{E}(\nu_i) = 0$. 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; Gordon et al. 2026). We have exacerbated the likelihood of biases by using non-representative training data. For these reasons, we do

not use the naive estimator. 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.

A growing literature in statistics and computer science has developed methods to address this problem (Angelopoulos et al. 2023; Kluger et al. 2025; Lu et al. 2025; Carlson and Dell 2025). Methods emerging from different disciplines adopted slightly different names and approaches, including *Prediction Powered Inference* and *Design-Based Supervised Learning*, but the fundamental idea is the same. All methods rely on a representative subset of labeled or ground-truth observations, in which D_i is known, to correct the biases in the model predictions.

Let $s_i \in \{0, 1\}$ be an indicator for whether a ground-truth label exists for observation i, and π_i be the probability that pixel i is labeled. An Augmented Inverse Probability Weighted (AIPW) missing-data estimator of μ is:

$$\widehat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \left[p(k_i) + \frac{s_i}{\pi_i} \left(D_i - p(k_i) \right) \right]. \tag{2}$$

In words, we estimate the average model bias in the labeled sample, and then use that to correct for bias in the naive estimator. This correction can be conducted within a country, within a group of countries, or conditional on a continuous covariate. The estimator is doubly-robust; it will deliver consistent estimates of μ if either $p(k_i)$ is correctly specified, or if π_i reflects true label probabilities, which holds trivially when observations are labeled randomly with known probabilities. Alternatively, π_i can be estimated as a function of the satellite features.

In our setting, we do not want to rely on either of these functions being correctly specified, however. The selection process into our training data is unknown, and trying to estimate π as a function of the satellite features would risk omitting key variables. Instead we collect new ground-truth observations using the higher-resolution satellite imagery from Google Earth discussed in Section 3. We do this in such a way that we know the resulting verifications will be representative of the broader population by design.

Expert verifications are more accurate than the model's predictions, but the process is costly and time-consuming. Constructing a verification sample through random sampling would result in an estimate that is unbiased at the cost of a larger variance. Assuming independent

labels with fixed probabilities, the variance of our estimator is:

$$\text{Var}(\hat{\mu}) = \frac{1}{N}\left[\text{Var}(D_i) + \mathbb{E}\left[\frac{1-\pi_i}{\pi_i}\left(D_i - p(k_i)\right)^2\right]\right]. \tag{3}$$

Since N is large, the first term will be small, and the variance is dominated by the second term. This second term will be large when the labeled sample is small, since it is decreasing in π_i. Furthermore, waste sites are a tiny fraction of a percent of total land cover in most areas. This means that the sample size of labeled pixels required to obtain a reasonably precise estimate of the prevalence of landfills would be enormous, since samples of a few hundred, or even a few thousand sites chosen at random will likely contain few or even zero dumps.

Instead, we use our machine learning model predictions to inform an active sampling and optimal verification approach. This method has its precursors in two-stage 'proxy-Neyman' sampling techniques (Neyman 1934; Cochran 1977), and it has more recently been adapted by the computer science and remote sensing communities where it is called *Active Statistical Inference* (Stehman 2012; Bai et al. 2024; Zrnic and Candès 2024; Hamilton et al. 2025). The key idea can be seen directly in equation 3—the variance of $\hat{\mu}$ depends on the correlation between a function of the sampling probabilities and the squared prediction errors.

More precisely, given a limited budget for sampling, we can improve the efficiency of the estimator by choosing π_i according to:

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

The constraint in this setup is an expected sample size constraint. In practice, realized sample sizes may exceed S. If marginal labeling costs are small, this may not matter much, otherwise a risk averse researcher should scale S appropriately relative to their budget.

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 $\pi_i \propto \sqrt{\mathbb{E}(\nu_i^2)}$.

Allow $p_i^* = \mathbb{E}[D_i|k_i]$ to 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 $p(k_i)$, 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 $\mathbb{E}(\nu_i^2) = p_i^*(1-p_i^*) + (p(k_i) - p_i^*)^2$. The first term reflects the irreducible uncertainty of pixel i, after conditioning on satellite features, and the second term reflects the increase in variance resulting from our

model's bias.

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, which gives the sampling rule $\pi_i \propto \sqrt{p(k_i)(1-p(k_i))}$. Of course, if our model is biased, this will not exactly minimize equation 4, but if $p(k_i)$ is close to p_i^*, it will get close. This strategy is known as proxy-Neyman sampling.

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 towards zero as $p(k_i)$ 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 harder to classify[11].

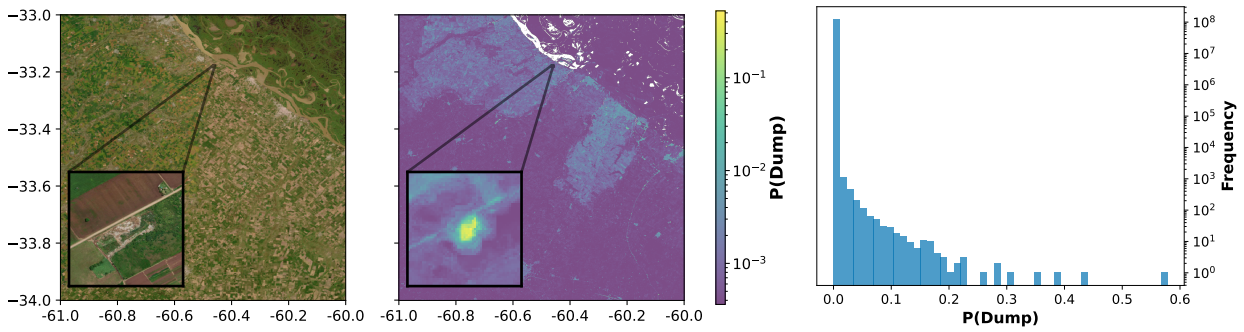
To demonstrate the utility of this approach, consider the following example based on a $1° \times 1°$ scene covering an area in north-east Argentina shown in Figure 2. The inset shows a small open dump near the border. The right panel shows the histogram of model-predicted probabilities for this scene — the y-axis is in log scale. The vast majority of pixels have predicted probabilities very close to zero[12]. 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. Far fewer number of pixels have probabilities even exceeding 10%.

Our sampling strategy still ensures that we will sample a reasonably large number of the low-probability pixels due to their prevalence. Sampling proportional to $\sqrt{p(k_i)(1-p(k_i))}$ 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 $p(k_i) \approx 10^{-4}$ than $p(k_i) \approx 0.5$, we will sample 200 times more of the low probability pixels than the mid probability pixels. 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.

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

[12] Crucially, 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 2: A Sample Tile and Histogram of Pixel Probabilities



Notes: A sample tile near Rosario, Argentina. Left panel: High resolution satellite imagery. Inset shows a small open dump. Middle panel: model predictions for the same region. Right panel: Histogram showing the model-predicted probability distribution for all pixels in the scene (log-scale y-axis).

As discussed above, even if our machine learning model predictions are badly miscalibrated, $\hat{\mu}$ remains an unbiased estimate of dump prevalence. Poor predictions simply mean that the variance of $\hat{\mu}$ would be larger than that obtained under alternative sampling designs. 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.

4.2.1 Efficiency Gains with Miscalibration

We can perform some calculations to estimate the precision gains associated with our sampling strategy relative to a uniform random sample. Define $\sigma_i^* = \sqrt{p_i^*(1-p_i^*)}$, $\sigma_i = \sqrt{p(k_i)(1-p(k_i))}$, and $\sigma_{\nu_i} = \sqrt{\mathbb{E}(\nu_i^2)} = \sqrt{p_i^*(1-p_i^*) + (p(k_i)-p_i^*)^2}$. Plugging in our approximately optimal π_i, the variance of the design-dependent portion of $\hat{\mu}$ becomes:

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

We can calculate the design effect, or the relative variance from following our sampling strategy, compared to using uniform random sampling (π_i is constant c) as:

$$D_{eff} = \frac{Var(\hat{\mu}|\pi_i \propto \sigma_i)}{Var(\hat{\mu}|\pi_i = c)} = \frac{\left(\sum_{i=1}^{N} \frac{\sigma_{\nu_i}^2}{\sigma_i}\right)\left(\sum_{i=1}^{N} \sigma_i\right) - S\sum_{i=1}^{N} \sigma_{\nu_i}^2}{(N-S)\sum_{i=1}^{N} \sigma_{\nu_i}^2}. \tag{6}$$

Since $N >> S$, we can approximate the design effect ignoring the second terms in the numerator and denominator. Define $R_i = \frac{\sigma_{\nu_i}}{\sigma_i}$ as a measure of model miscalibration, noting

that if $p(k_i) = p_i^*$, then $R_i = 1$, and R_i increases as the difference between $p(k_i)$ and p_i^* increases. Then the design effect can be rearranged to show:

$$D_{eff} \approx \frac{1}{N} \frac{\left(\sum_{i=1}^N \sigma_{\nu_i}\right)^2}{\sum_{i=1}^N \sigma_{\nu_i}^2} \frac{\sum_{i=1}^N R_i^2 \frac{\sigma_i}{\sum_{i=1}^N \sigma_i}}{\left(\sum_{i=1}^N R_i \frac{\sigma_i}{\sum_{i=1}^N \sigma_i}\right)^2} = \frac{1 + CV_{\sigma R}^2}{1 + CV_{\sigma_\nu}^2},$$

where CV_{σ_ν} is the coefficient of variation of σ_{ν_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 $p(k_i) = p_i^*$. 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:

$$D_{eff}^* \approx \frac{1}{N} \frac{\left(\sum_{i=1}^N \sigma_i\right)^2}{\sum_{i=1}^N \sigma_i^2}. \tag{7}$$

By Cauchy-Schwarz, this quantity will always be ≤ 1. We can calculate this theoretical quantity based on the distribution of $p(k_i)$ in the Argentinian scene described above, and we find $D_{eff}^* = 0.0012$. In other words, if our model were perfectly calibrated, the active sampling method would reduce the variance of $\widehat{\mu}$ relative to $\widehat{\mu}^{unif}$ 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:

$$1 + CV_{\sigma R}^2 = \frac{\sum_{i=1}^N \frac{\sigma_{\nu_i}^2}{\sigma_i} \sum_{i=1}^N \sigma_i}{\left(\sum_{i=1}^N \sigma_{\nu_i}\right)^2}, \tag{8}$$

In an extreme case where our predictions were random noise, so σ_{ν_i} and σ_i are independent, this becomes:

$$1 + CV_{\sigma R}^2 = \frac{N \sum_{i=1}^N \sigma_{\nu_i}^2}{\left(\sum_{i=1}^N \sigma_{\nu_i}\right)^2} \frac{1}{N^2} \sum_{i=1}^N \frac{1}{\sigma_i} \sum_{i=1}^N \sigma_i \approx (1 + CV_{\sigma_\nu}^2)(1 + CV_\sigma^2),$$

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 computationally and memory intensive. More pressingly, our training data are 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 the pixel is a dump in its lowest probability year, then it is a dump in all years, and if the pixel is not a dump in its highest probability year, then it was never a dump. 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 the Sentinel satellite data (and therefore machine learning predictions) only

exist for 2017–2021, we extend our sample temporally by performing verifications in the high-resolution historical imagery from 2011 to 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 according to the methods described above.

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 consistently run without errors.

Running this sequence of operations takes an average of 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 plastic 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. Appendix Figure A5 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. Our selected tiles are overlaid 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 unless otherwise stated.

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 2021. We exclude water pixels, and for pixels with missing imagery we impute the tile median probability.

We then use these predictions to apply the proxy-Neyman sampling strategy within 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 select 13,941 pixels for verification—on average less than 30 pixels in each of our 500 tiles. Thus our estimates are highly uncertain within tiles, which span more than 10,000 square kilometers near the equator, but become more precise at higher levels of aggregation.

4.3.1 Accuracy, Bias, and Calibration

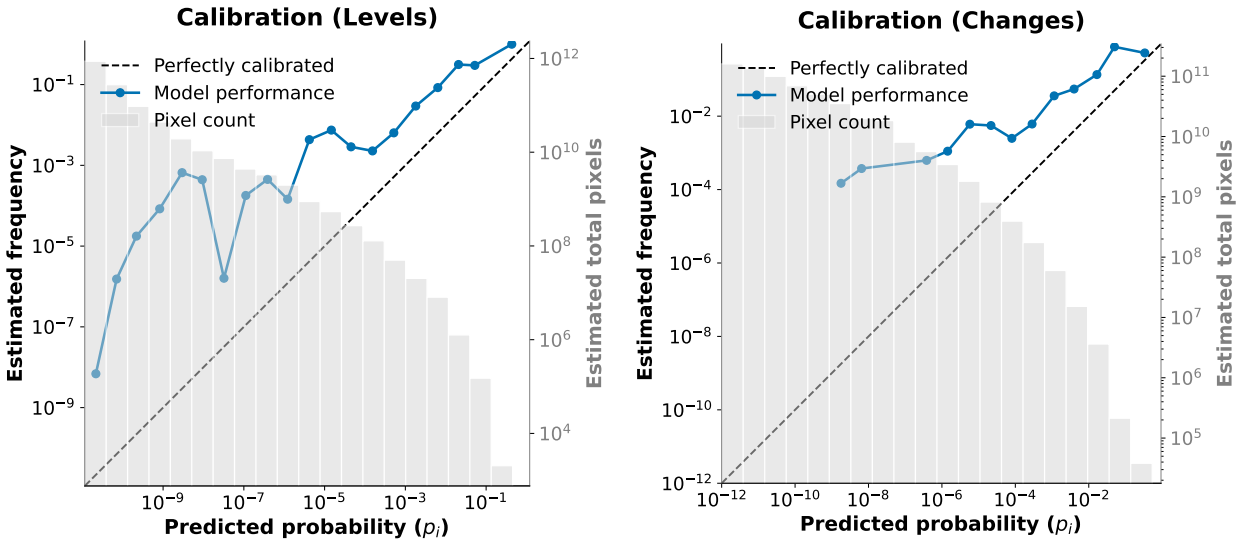
We can re-assess the model's classification performance in the verification sample, which measures accuracy at the point of deployment rather than in the training hold-out. For 2018 dump levels, the model achieves an AUC of 0.98 when we weight by inverse sampling probabilities so that the metric reflects performance across the full population of land pixels. At the population-optimal classification threshold, weighted balanced accuracy is 95%. Performance is even stronger for changes: the weighted AUC is 0.99, with balanced accuracy of 97%. These strong results give reassuring evidence that our model remains accurate outside of the selected training data.

We do see some evidence of biases, however. Appendix Table A4 shows the results of regressions of prediction errors (ν_i) on year dummies, world region dummies, a quadratic of GDP, and log population density. Relative to 2017, our model underpredicts dumps in 2018 and 2019—the year of the waste ban—meaning that the naive method would underestimate an increase in dumps that occurred during those years. Our model tends to predict too many dumps in Europe, Central Asia, and Sub Saharan Africa compared to East Asia & the Pacific, and too few dumps in areas of high population density. The magnitudes of these coefficients are on the same order as our estimated prevalence of dumps, meaning that the differential bias between regions—such as the gap between Sub-Saharan Africa and East Asia & the Pacific—can be just as large as the actual dump prevalence.

What is crucial for our sampling strategy, however, is not necessarily accuracy or biases, but calibration. Figure 3 shows model calibration plots. The x-axis shows model predicted probabilities and the left y-axes show the estimated frequency of pixels in that probability

bin that were verified as containing an open dump (on the left) or changing status at least once (on the right), using the inverse probability weights. If our probabilities were perfectly calibrated, the blue line would equal the 45 degree line. Clearly this is not exactly the case - at very low probabilities, we underestimate the probability of dumps. Still, it is clear that our model's predictions contain useful information. A similar story holds for our predicted changes, as seen in the right panel. 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 3: 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 prevalence and change 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. All axes are in log scale.

We can use this information to estimate the ex-post efficiency of our proxy-Neyman strategy, by estimating p_i^* as a function of $p(k_i)$, and then calculating the design effect using the formula above. We estimate p_i^* as a function of $p(k_i)$ using a LOESS regression, and plug it in to equation 4.2.1. We report the design effect separately for each of our two objectives. For estimating levels, we find a design effect of 0.32, indicating that we can estimate mean dump prevalence with a variance 68% smaller than if we had used uniform random sampling. In other words, we have increased our effective sample size by more than a factor of 3. 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 0.09, indicating that miscalibration substantially reduced the value of our sampling strategy. A perfectly calibrated model would

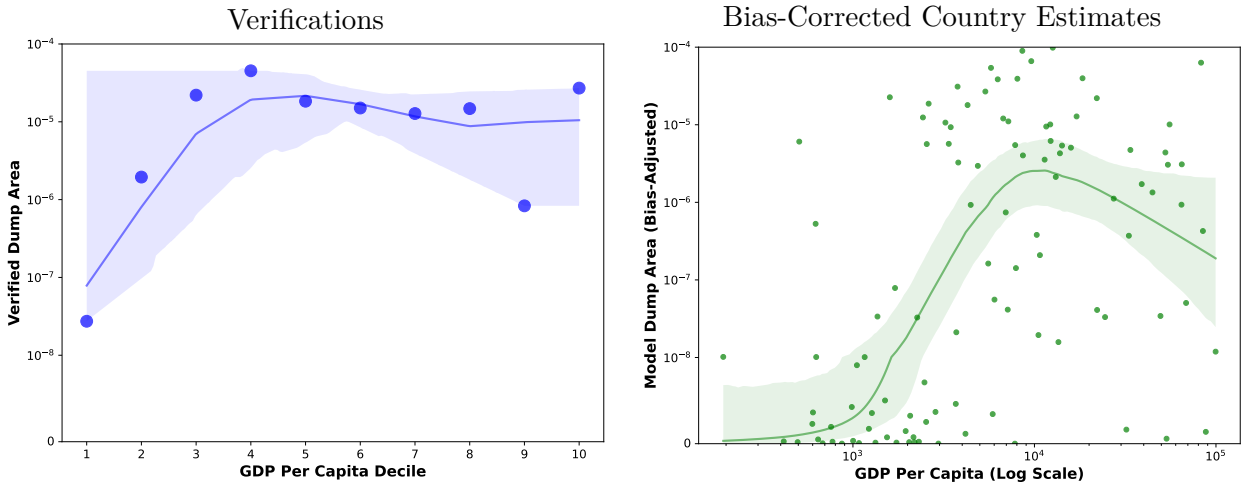
have been able to achieve a nearly 10x effective sample size.

When it comes to estimating changes efficiently, we do even better. We estimate an ex-post design effect of 0.13, relative to an optimal lower bound of 0.03. This means that we have effectively increased our ability to detect changes over time by 7-8x relative to a uniform random sample.

5 Global Dump Descriptives

This increased statistical power allows us to generate a novel set of descriptive facts about the global distribution of open dumps based on a relatively small sample of labeled pixels. We estimate the fraction of global land area covered with open dumps in 2018 was 2.76×10^{-5} (95% CI: $[1.5, 4.3] \times 10^{-5}$). This corresponds to an area of more than 4,000 square kilometers, about 50% larger than the US state of Rhode Island. For another point of reference, this is about 0.2% of estimated urban land surface area in 2015.[13]

Figure 4: Dump Prevalence vs GDP Per Capita

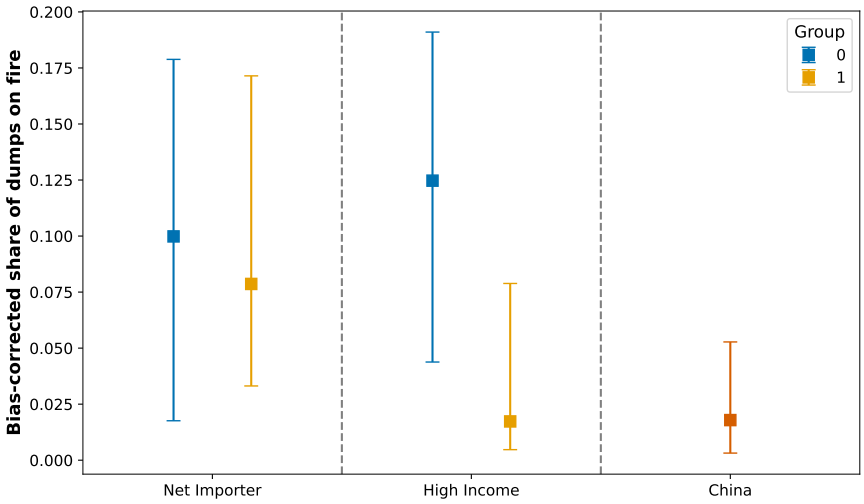


Notes: Estimates of dump prevalence vs country GDP. Left panel shows average prevalence using only verified pixels, calculating means within country-GDP per capita deciles. Right panel shows AIPW debiased country-level estimates estimated using equation 2. Plots show LOESS smoothed lines and 95% confidence intervals from a stratified tile-cluster bootstrap with 1000 replicates.

Figure 4 shows the prevalence of open dumps plotted against GDP per capita. The plots clearly show an 'Environmental Kuznets Curve' (EKC) in both the verified points (left) and the AIPW bias-corrected country level estimates (right) (Grossman and Krueger 1995). Low

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

Figure 5: Garbage Fire Prevalence, Bias-Corrected



Notes: Bias-corrected (AIPW) estimates of the share of dumps with fires by net importer status, country income classification, and for China. Dump fires defined as the intersection of at least one VIIRS fire detection during the year with a dump. Net importer defined as waste imports greater than waste exports using BACI data and definitions in Section 3. Income uses World Bank classification threshold of $13,935 GDP per capita. China is excluded from the net-importer and income groups and shown separately. Error bars show asymmetric 95% confidence intervals from a stratified tile-cluster bootstrap.

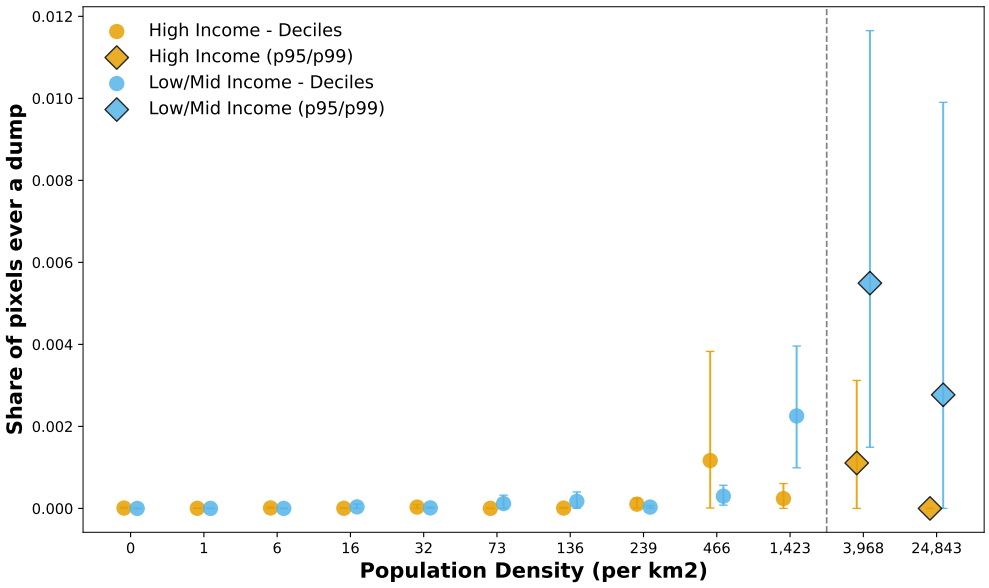
income countries have the lowest prevalence, presumably because they generate less waste, though it is also plausible that littering is more prevalent, and this more dispersed disposal method is harder for us to detect in the satellite record. Dump prevalence sharply increases for middle-income countries, 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 both less visible to satellites and likely to be less environmentally harmful.

Figure 5 shows more evidence of heterogeneity in waste management practices across countries: the prevalence of fires in open dumps. Fires may be set intentionally, but they also occur naturally due to the heat generated by decomposing organic waste. Strikingly, 12.5% of dumps in low- and middle-income countries have at least one fire detected in a given year. In high-income countries, the number is closer to 2%. In China, the share of open dumps ever on fire is also close to 2%, averaged across years. Figure 5 also shows a comparison between waste net-importing countries and other countries, though we do not see much difference along this dimension.

Figure 6 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 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 in high-income countries represented by the dots in orange. 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 24,843 people per square kilometer approximates the population density of Manhattan.

Figure 6: Dump Prevalence vs Population Density



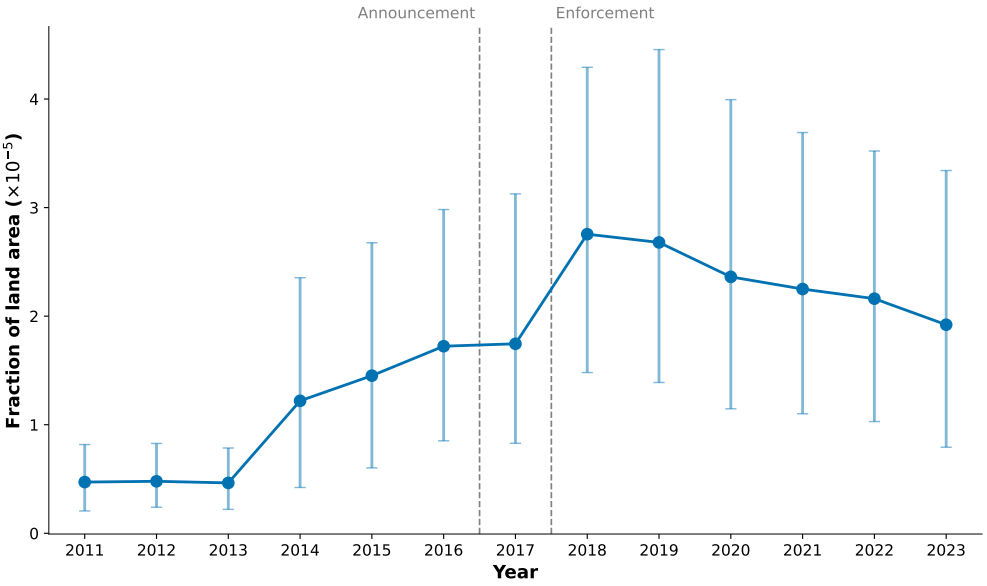
Notes: Point estimates of the share of pixels ever containing a dump, by population density decile plus the 95th and 99th percentiles, based on 13,941 verified pixels selected using approximately optimal sampling technique. Orange points show estimates for high-income countries, and blue points show combined low- and middle-income country estimates (World Bank classification threshold of $13,935 GDP per capita). Whiskers show 95% confidence intervals from a stratified tile-cluster bootstrap. Population density data are based on a 1km grid from Earth Science Data Systems, NASA (2024).

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. Given that these dumps are also more likely to be on fire, these facts highlight the importance of waste-mismanagement as a crucial public-health issue in low- and middle-income countries.

Finally, Figure 7 shows the time trend of our global estimates from the design-weighted

annual means of the verified pixels.[14] A large jump between 2017 and 2018, corresponding to the year of the Chinese waste import ban, is immediately visible. Point estimates indicate a 58% increase in global dump area, though confidence intervals are large. While we should not claim causality based on eyeballing a single time-series figure, these results strongly suggest that the ban led to a global increase in dump-area, despite the decrease in international trade flows. The other notable jump in the time-series occurs between 2013 and 2014, which corresponds to the year of the Green Fence Policy, when China increased inspections on waste imports (see Section 2).

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



Notes: Estimates based on 13,941 verified pixels selected using approximately optimal sampling technique. 95% confidence intervals from a stratified tile-cluster bootstrap. Dashed lines show China's waste ban announcement (2017) and enforcement (2018).

These increases are not a priori obvious. If Chinese waste importers 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, in Appendix Figure A6, we see dumps increase, even within China itself after the ban. This striking pattern suggests that the recyclers that were shut down after the

[14]These time series figures use only the verified pixels, not the AIPW estimator, since model predictions are only available from 2017–2021. Each annual point is an inverse-probability-weighted mean of the verified labels, with each pixel weighted by the inverse of its inclusion probability across both sampling stages, so the series — and the percentage changes we report from it — are globally representative despite the nonuniform sampling design. Confidence intervals are from a stratified tile-cluster bootstrap that resamples tiles within strata.

National Sword Policy was enacted may have been processing Chinese domestic waste as well as foreign imports. After the ban, this led to an increase in dumps within China, as domestic waste was no longer recycled. These results are in tension with previous work showing that reduced waste imports led to lower air pollution in Chinese recycling cities (Guo et al. 2023; Sigman and Strow 2024). These findings may be reconciled if the pollution generated from waste imports was due to recycling activities and not open-burning. It is also worth highlighting that if the increase in dumps came outside of Chinese recycling centers, then this spillover effect would contaminate regressions using other Chinese cities as a control group.

The left panel of Appendix Figure A6 shows that the increase in dumps after 2018 was not confined to China. This suggests that some of the waste imports previously destined for China may have ended up in the open dumps of other countries. We explore this further in the following section.

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 data are uniquely suited to study this question. Using our method outlined above, we measure open dumping consistently and we observe outcomes in the same units on both sides of the ban-induced reallocation of trade rather than relying on administrative waste statistics that are missing or incomparable precisely where mismanagement is most likely.

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 (2011–2023) 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 analyze our dump measurements at two levels of aggregation, although the treatment is at the country level for both. First, we aggregate

the verified pixels to a country-by-year panel of verified dump area[15], and merge it with the trade data. At the pixel level, we impose no aggregation at all. The outcome is each verified pixel's dump status in each year and every pixel is matched to the country it is located in. Our final sample includes estimates of dump prevalence in 116 countries (we exclude China), however we only verify greater than zero dump pixels in all years for 50 of these countries.

6.1 Empirical Strategy

We first describe how we measure exposure to the ban, and then the models we estimate. As described above, we define waste imports and exports as the trade in products banned by China's National Sword Policy. Our main measure covers plastic waste and scrap (HS 3915) plus unsorted waste paper (HS 4707.90), the two categories at the center of the 2017 ban, which together made up over 84% of the tonnage China imported across all banned categories in 2016. We broaden this to every HS code banned in robustness checks.[16] We use CEPII-BACI derived from UN Comtrade data to collect bilateral trade flow quantities and values for 2012–2023 (Gaulier and Zignago 2010).[17] Our treatment is each country's change in *net* waste imports (imports minus exports) around the ban.[18] Appendix Figure A7 maps the distribution of the treatment.

We leverage changes in imports around the timing of China's National Sword Policy as a quasi-exogenous shock in several ways. First, because the treatment varies at the country level, we estimate a continuous two-way fixed effects model on the country panel:

$$\text{DumpArea}_{it} = \beta \Delta \text{WasteImports}_{i,2016-19} \times \text{Post-2016} + \gamma_i + \delta_t + e_{it} \tag{9}$$

and its corresponding event-study specification:

$$\text{DumpArea}_{it} = \sum_{t=2011, t\neq 2016}^{2023} \beta_t \Delta \text{WasteImports}_{i,2016-19} \times T_t + \gamma_i + \delta_t + \epsilon_{it} \tag{10}$$

where DumpArea_{it} is our satellite-based estimate of open dump area (km^2, from the verified pixels as described above) in country i and year t. The continuous treatment variable,

[15] This is the weighted prevalence estimate of Section 4.3 scaled by the land area of each country

[16] This adds textile and fiber waste, HS 4707 in full, 5103–5104, 5202, 5505, and 6310.

[17] Values are available for most countries in all years. To create a complete panel of trade *quantities*, we impute missing quantity observations by applying the exporter-specific average quantity-to-value ratio within HS heading and year.

[18] Concretely, $\Delta\text{WasteImports}_{i,2016-19} = (\text{Imports}_{i,2019} - \text{Exports}_{i,2019}) - (\text{Imports}_{i,2016} - \text{Exports}_{i,2016})$ in the main trade category, measured in thousands of tonnes in the regressions. The variable is time-invariant within country and is set to missing for countries without BACI trade data.

$\Delta\text{WasteImports}_{i,2016-19}$, captures a country's change in waste trade exposure and is calculated as the difference in net waste imports between 2019 and 2016, as described above. In equation 9, we interact this time-invariant measure with Post-2016, an indicator variable for years including and after the policy's 2017 announcement. The model includes country (γ_i) and year (δ_t) fixed effects. Equation 10 interacts our continuous 'treatment' variable with dummies for each year T_t and omits 2016 as the reference category. 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. Standard errors are clustered by country throughout.

We estimate the same designs on the un-aggregated pixel panel. The outcome is verified pixel p's dump status in year t scaled by its country's land area, so that coefficients are in the same units as the country-level model; the treatment is still the pixel's country-level import shock; and we replace country fixed effects with sampling-tile fixed effects. Each pixel enters with its sampling weight, and standard errors remain clustered at the country level, the level of treatment. The two aggregations answer different questions. The country model weights every country equally and asks whether countries hit by larger import shocks saw their dumps grow more. The pixel model instead weights the world's land representatively and asks how much additional dump area appeared on average when the ban redistributed waste across countries in the pattern that it did.

For the coefficient of interest β to identify the causal level change of dump area with respect to net import quantities, $\Delta\text{WasteImports}_{i,2016-19}$ 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 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 (Kellenberg 2010). 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. Open dump area, by contrast, is a slow-moving stock. The suddenness and scale of the ban therefore work in our favor as a discrete, externally timed jump in imports is easy to distinguish from the smooth background trends that country and year fixed effects absorb.

However, it is still plausible that the geographic locations or positions in the international trade network of these countries are correlated with unobservable, time-varying determinants

of changes in landfill prevalence. Countries with weak environmental enforcement, for example, may both attract diverted waste and see dumps expand for domestic reasons. We therefore also implement a shift-share instrumental-variables version of the model above that uses only the component of the import change predicted by the pre-ban structure of the trade network. For each exporting country j we measure the flow of banned waste it sent to China before the ban, and we allocate that soon-to-be-displaced flow across destinations in proportion to j's pre-ban exports to destinations other than China:

$$z_i = \sum_c \sum_j s^c_{ji}\, g^c_j, \tag{11}$$

where g^c_j is exporter j's average 2015–2016 exports of waste category c to China (i.e., the shift) and s^c_{ji} is destination i's share of j's 2015–2016 exports of c to non-China destinations (i.e., the shares), constructed separately for each HS code and summed. The instrument z_i is thus the diversion *predicted* for country i had every exporter redirected its stranded China volumes along its pre-existing non-China routes. In the IV specification we instrument $\Delta\text{WasteImports}_{i,2016-19} \times \text{Post-2016}$ with $z_i \times \text{Post-2016}$. The design has a single aggregate shock, so identification rests on the exogeneity of the pre-ban *shares* (Goldsmith-Pinkham et al. 2020): countries' positions in the pre-2017 waste-trade network must be unrelated to counterfactual *trends* in open dumping. We provide evidence on this assumption below.

6.2 Results

Table 1 shows our main estimates, with our country-level aggregation in Panel A and our pixel-level version in Panel B. Column (1) shows the TWFE version, while Column (2) shows the shift-share instrumental variables version. In both panels the coefficients can be interpreted as m^2 of additional dump area per kg of additional net waste imports. The OLS estimates are positive at both levels but noisier, reaching conventional significance at the pixel level. The shift-share IV estimates are of slightly larger magnitude and are significant at both aggregations — 0.111 m^2 per kg (95% CI [0.007, 0.215]) at the country level and 0.179 m^2 per kg ([0.001, 0.357]) at the pixel level, with first-stage F-statistics of 40.3 and 15.6, respectively. The implied elasticities with respect to a country's import change are large. Evaluated at sample means relative to pre-ban dump area the elasticity is 1.2 at the country level and 1.5 at the pixel level.

Figure 8 plots the event study version of Column (2) in Table 1, while the first panel of Appendix Figure A8 shows the event study version of the OLS. The dynamics line up with the ban's timeline as estimates are flat and near zero in every pre-ban year and turn positive

within the first two years after the 2017 announcement.

Table 1: Effects of Change in Net Imports on Open Dump Area (Pre/Post Ban)

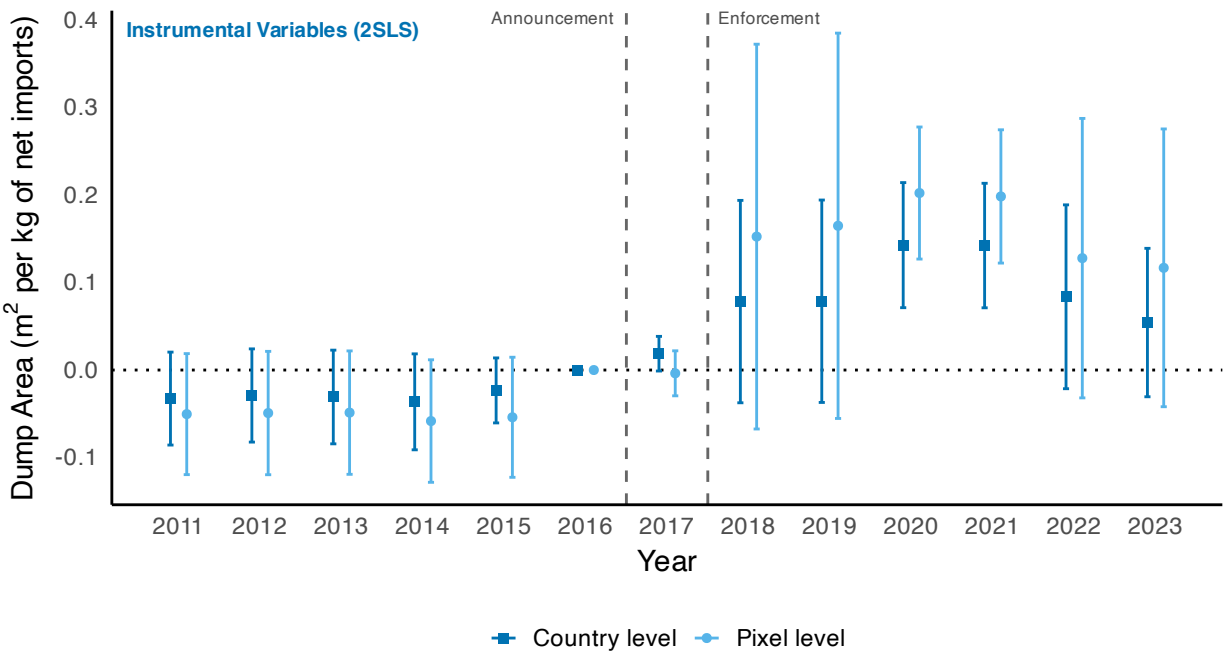
	Dependent variable: Dump area (km^2)	
	OLS	IV
	(1)	(2)
Panel A: Country-Level		
Δ Net Imports (kt)	0.061	0.111**
$\times$ Post-2016	(0.049)	(0.052)
	[0.212]	[0.037]
Implied elasticity	0.646	1.175
First stage		0.801*** (0.126)
First-stage F		40.3
Observations	1,508	1,508
Countries	116	116
Panel B: Pixel-Level		
Δ Net Imports (kt)	0.174*	0.179**
$\times$ Post-2016	(0.092)	(0.091)
	[0.059]	[0.049]
Implied elasticity	1.445	1.486
First stage		1.079*** (0.273)
First-stage F		15.6
Observations	159,744	159,744
Countries	116	116

Notes: Table shows results of estimating equation 9. Coefficients in both panels are m^2 of dump area per kg of increase in net imports (numerically identical to km^2 per kt). Treatment is change in net imports of plastics and unsorted paper (HS 3915 + 4707.90) $\times$ post (post = year $>$ 2016); instrument is predicted diversion from pre-ban bilateral trade shares (per-HS-code shift-share, equation 11). Panel A: country $\times$ year panel, outcome = verified dump area. Panel B: verified-pixel $\times$ year panel, outcome = pixel dump status $\times$ country land area, weighted by inverse sampling probabilities. Implied elasticity: coefficient $\times \bar{X}/\bar{Y}$, with $\bar{X}$ the estimation-sample mean net-import change and $\bar{Y}$ the pre-ban (year $\leq$ 2016) mean outcome — the percent increase in dump area relative to its pre-ban mean implied by a one-percent larger rise in net imports. First stage: realized change $\times$ post on predicted diversion $\times$ post (coefficient, SE, and cluster-robust F). Full reduced-form and first-stage estimates are shown in Appendix Table A5. SEs clustered by country in parentheses, p-values in brackets; China excluded. *p$<$ 0.1; **p$<$ 0.05; ***p$<$ 0.01.

Are these large magnitudes realistic? Although we measure only the surface area of dumps,

our estimates imply a depth once combined with typical waste densities. Importantly, these simple calculations assume that *all* of the imported waste is dumped. While this is unlikely to be true, it is nevertheless a helpful exercise in interpreting the magnitude of the coefficient we estimate. We can estimate depth $= T/(\beta \times \rho)$, where T is the number of years the extra waste has been flowing, β is the regression coefficient (m^2 per kg of annual net imports), and ρ is the density of waste (kg per m^3). With $T = 3.5$ years (the average post-ban horizon in our sample) and $\rho = 175$ kg per m^3 for loose plastic and paper waste, the country-level point estimate implies a layer 18 cm deep, and the ends of the confidence interval bracket 9 cm to 2.8 m. These depths correspond to different kinds of dumping that our back-of-the-envelope calculation cannot separate, but none is implausible. A 9–18 cm layer is consistent with the informal roadside dumping typical of the main receiving countries, while depths of a couple of meters point to larger, more consolidated sites, again with the caveat that these numbers assume all imported quantities are dumped.

Figure 8: Effect of Change in Net Imports on Open Dump Area, Event Study Version



Notes: Figure shows year-by-year IV estimates of equation 10 where each Δ net imports × year interaction is instrumented by predicted diversion (z_i, equation 11) × year, with 2016 as the omitted reference year. Darker squares show the country-level panel (country and year fixed effects, unweighted), while lighter circles show the pixel-level panel (tile and year fixed effects, weighted based on sampling probabilities). Coefficients are m^2 of dump area per kg of net imports; dashed vertical lines mark the ban's announcement (July 2017) and enforcement (January 2018). Bars are 95% confidence intervals from standard errors clustered by country. OLS, reduced-form, and first-stage versions of the same figure appear in Appendix Figure A8.

We now discuss evidence supporting the main identifying assumption that pre-ban trade

network position of countries is unrelated to their counterfactual trends in dumping. First, in Appendix Table A6 we note that while the instrument correlates with some baseline country characteristics (which are absorbed by our country fixed effects), it is uncorrelated with the pre-ban trend in dump area in the bivariate specification. Second, the pre-ban coefficients in Figure 8 and Appendix Figure A8 are all insignificant and near zero.[19] Third, the first stage of the IV is relatively strong (cluster-robust F of 40.3 and 15.6) and its Rotemberg weights are well-behaved (nearly all positive and concentrated in interpretable corridors as shown in Appendix Table A7). Fourth, we run two placebo exercises. The first is a commodity placebo that rebuilds the shift-share instrument using primary-form plastics (HS 3901–3914), an upstream commodity China never banned. This "predicted diversion" captures generic trade-network position with no actual shock attached. When we include this placebo instrument in the reduced form regressions, the real instrument retains its effect while the placebo is close to zero (Figure A10).[20] Our second placebo is a timing one that moves the date of the China ban. We re-estimate the design on pre-ban years only (2011–2016) with a counterfactual ban in 2014 and reassuringly find a pooled coefficient near zero and a flat year-by-year series (Figure A11). Overall, these four tests show that exposure is uncorrelated with pre-ban dump trends, no effect is visible before the ban, none appears at a placebo date, and none flows through a placebo commodity.

Finally, we test the robustness of our findings. Appendix Figure A9 re-estimates the IV under eleven alternatives: broadening the trade category to all banned waste headings; a yearly-dose version in which annual net imports are instrumented by the time-varying predicted diversion; replacing year with land-area-quartile-by-year fixed effects; restricting to the 50 countries with verified dumps in every year; substituting the debiased machine-learning area for the verified area (2017–2021); equalizing pixel design weights within countries; dating the ban at its 2018 enforcement rather than its announcement; a pooled-network version of the instrument; adding baseline GDP per capita, population density, and land area $\times$ post controls; comparing countries only within baseline-dump-area quartiles interacted with year; and rebuilding the instrument without the largest Rotemberg-weight cell (US waste paper). The point estimates are stable across every specification and nearly all remain statistically significant at the 95% level, and in the few where the confidence interval widens, it widens around an essentially unchanged coefficient.

To summarize, our confidence intervals are wide but the core finding of an increase in

[19] Joint Wald tests that all pre-ban interactions equal zero do not reject (cluster-robust $p = 0.103$ for the country-level IV).

[20] We run the reduced form version of the regression as the two exposures are highly correlated and a placebo regression alone would simply proxy the real effect.

dumping area in response to increases in net imports is stable in magnitude across two aggregation levels, two estimators, and the eleven alternative specifications of Appendix Figure A9, while both placebo exercises find no effect where none should exist.

7 Case Studies: Java, California, Southeast Brazil

The analyses above are globally representative and well-suited to analyzing the cross-country determinants of open dumps. They are not well powered to be able to detect the local externalities associated with dumps, however. To more precisely study the disamenities of open dumps, we choose a select subset of sub-national geographies and use our model to find as many dumps as we can.

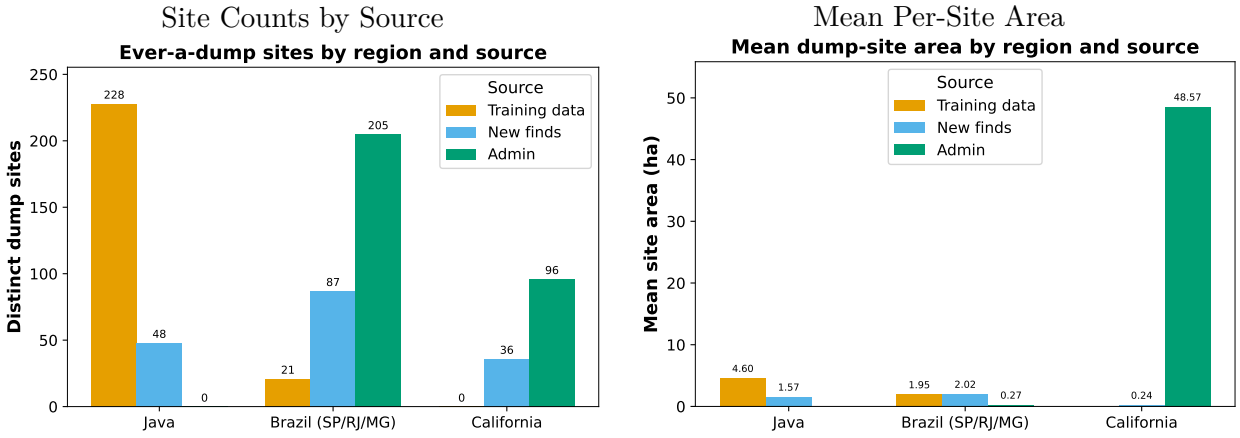
We select three regions based on a combination of factors. First, we select the island of Java in Indonesia. Java is not only one of the most densely populated regions on earth, but it was also a major recipient of diverted scrap materials after the 2018 China import ban, due to its convenient location along the US-China trade route (see discussion in previous section). Furthermore, Indonesia was a major source of training data for our model, as early attempts to use satellite data to detect dumps focused on Indonesia and found a large number of informal sites that we incorporated into our training pipeline (Kruse et al. 2023). For this reason, we expect our model to perform well there.

Secondly, we choose the state of California in the United States. While Indonesia is a middle-income importer of waste, California is a high-income exporter. Analysis of the dumps in that state allows us to see whether reduced exports after the China waste ban increased landfill-related pollution in the US. Furthermore, we can compare and contrast the open dumps that our model finds to the list of regulated landfills tracked by the EPA Landfill Methane Outreach Program (LMOP).

Lastly, we look at three densely populated states in Southeastern Brazil: Rio de Janeiro, São Paulo, and Minas Gerais. Brazil is not a noteworthy importer or exporter of waste, and saw only a very small increase in imports after 2018 due to its distance from major trade routes. Brazil is one of the few countries, however, that attempts to administratively track open dumps through the National Sanitation Information System (SNIS). This allows us to compare our model predictions to known open dumps in a country where we did not have much training data.

We follow a similar procedure in each region. First we use our existing model to predict the probability of a dump for every pixel in the region. Second, we set a region-specific

Figure 9: Case Study: Dump Sites by Source — Known Sites and Model New Finds



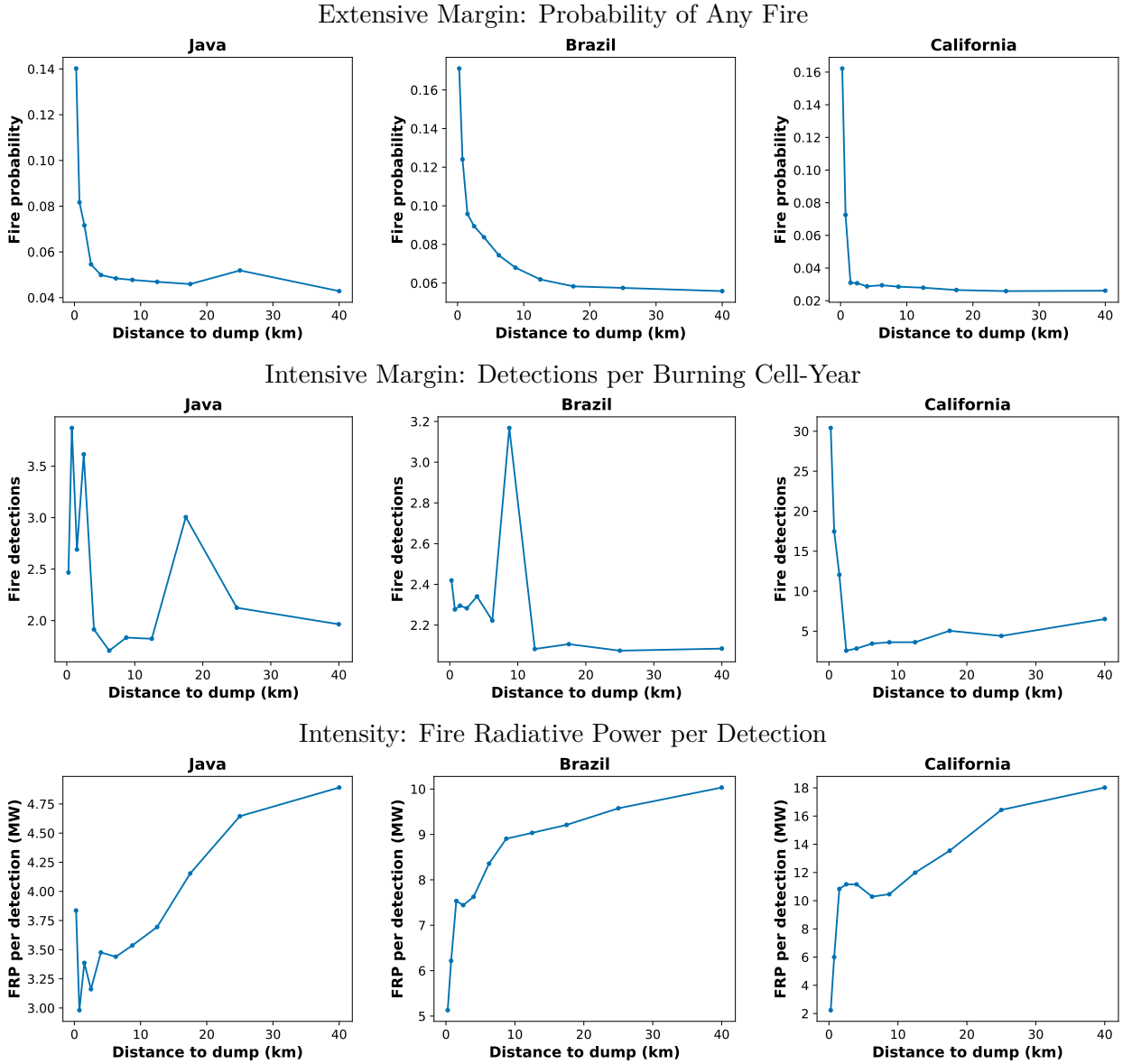
Notes: Dump sites by source. Admin includes SNIS registered dumps (Brazil), and EPA LMOP landfills (California). Training data described in Section 3. New finds are expert-verified in high resolution satellite imagery after reviewing model predictions, greater than 100m from known open dumps.

threshold probability, and use a clustering algorithm (DBSCAN) to draw a polygon around pixels with p_i greater than or equal to the threshold. Then we have expert annotators review all of these polygons and mark which ones were open dumps. The thresholds were set to give a manageable number of polygons to review in each region. In Java, we reviewed 348 polygons and found 48 new open dumps that were not previously in our training data. In Brazil, we reviewed 1,091 potential sites and found 87 open dumps. In California, we reviewed 936 potential sites and found 36 dumps not cataloged in the EPA LMOP data.

This exercise shows that, even when official data on open dumping exists, our model is able to locate dumps that are not known to authorities. Undoubtedly we missed some dumps in each region, however the relatively low success rates obtained by our verification efforts suggest that further lowering our thresholds in order to find more dumps would have resulted in many more potential polygons to investigate, the vast majority of which would be false positives. Figure 9 shows the number of sites in each region by source on the left, and the mean site area by source on the right. When we sum the area of all found dumps from any source, and compare it to the area of the region, we can compare these estimates to our estimates of the global prevalence of dumps from our representative global data in the previous sections. This is done in Appendix Figure A12. Our estimate for Java is somewhat higher than the global estimate, plausibly because of the high population density of the island. On the other hand our estimate for the three Brazilian states is somewhat lower—possibly because our model is less accurate in Brazil, and thus there are more dumps that we have yet to find. California is not directly comparable, since it includes the sanitary landfills in the EPA LMOP data

that our model is not designed to capture. Notably Figure 9 shows that these landfills are much larger than the typical dumps in Java and southeastern Brazil.

Figure 10: Case Study: Garbage Fires by Distance to Dump



Notes: VIIRS active-fire detections in grid cells around verified dump sites, by distance to the nearest dump (pooled 2012-2023). Second row plots the expected number of detections in pixels with at least one detection, and the bottom row plots mean fire radiative power per detection (MW).

Next, we overlap the VIIRS fire detections with our known dumps in each region. The first row of Figure 10 plots the probability of at least one fire detection in a given year as a function of distance to the nearest dump. In all three regions, we see elevated detections in the pixels closest to verified dumps, with 14-16% of dump pixels showing at least one fire detection

in any given year, consistent with the estimates from the global analysis in Figure 5. This probability drops off sharply as distance from the dump increases. In California, these fires are mostly methane flaring from regulated landfills, whereas in Java and Brazil, they are more likely to be smoldering trash fires. The second row of the figure shows the expected number of fire detections conditional on greater than one detection—the particularly high numbers for California support the hypothesis of regular methane flaring. The relatively low intensity of these fires displayed in the third row of the figure supports the interpretation of smoldering fires, whereas the higher-intensity less-frequent fires in the more distant pixels include more powerful wildfires.

An alternative possibility is that, since dumps tend to be in urban areas, this pattern is just picking up a variety of urban heat sources. Appendix Figure A13 shows fire detections as a function of distance to other sites of interest including hospitals and markets (OpenStreetMap 2026). While detections tend to be high relatively close to these places, reflecting a variety of urban heat sources, neither of these types of landmarks shows the same elevated rate of fires in the very closest pixels, monotonically decreasing with distance, as do the dumps.

7.1 Pollution from Waste Imports

These fires, especially the smoldering, unmanaged fires in the open dumps of Brazil and Indonesia, generate a lot of particulate pollution. As discussed in Section 5, these fires may occur naturally from the heat generated by the decomposition of organic waste. However, poorly managed plastic and paper waste is highly flammable, and may catch fire, whether intentionally or not. This raises the question of whether waste imports contribute to these garbage fires, thus raising pollution levels around open dumps in importing countries.

In this section, we use the China waste ban as an exogenous shock to study whether waste imports lead to increased fires in open dumps and increased particulate pollution. We use a basic event study framework, estimating the following specification:

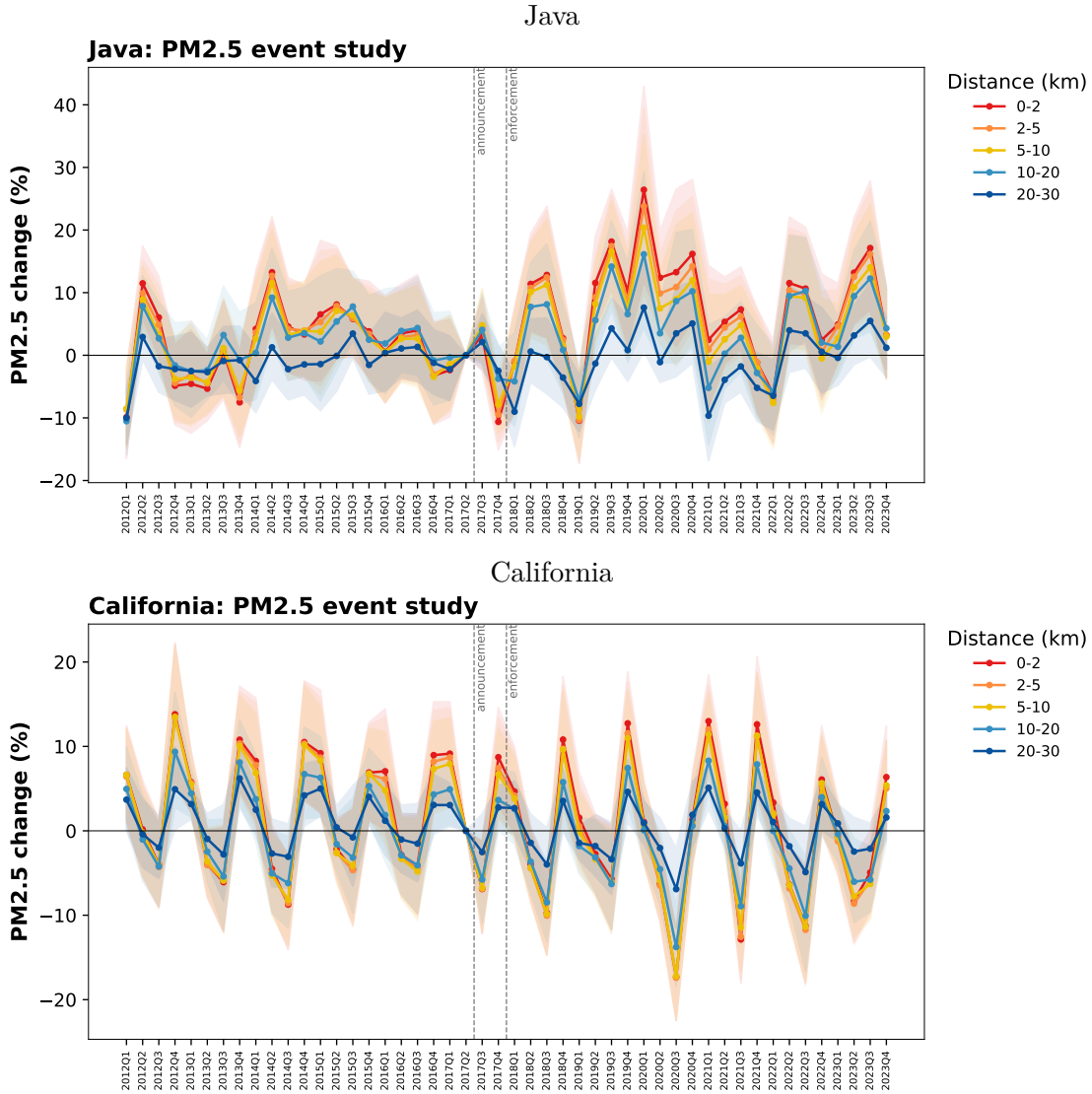
$$\log PM_{it} = \sum_{d} \sum_{q \neq q_0} \beta_{d,q} \left(Ring_i^d \times \mathbb{1}[q(t) = q] \right) + \alpha_i + \lambda_{a(i),t} + \varepsilon_{it}, \tag{12}$$

where i indexes grid cells, t indexes months, and cells are assigned to distance rings $d \in \{0\text{–}2, 2\text{–}5, 5\text{–}10, 10\text{–}20, 20\text{–}30\}$ km from the nearest dump, with the 30–40 km ring as the omitted reference. Pixels greater than 40 km from any dump are dropped from the analysis.

The specification includes cell fixed effects (α_i) to absorb time-invariant differences in pollution across space, and district by month fixed effects ($\lambda_{a(i),t}$) to absorb time varying shocks

common to all pixels in a district. Thus the coefficients $\beta_{d,q}$ show the difference in the change in pollution relative to 2017Q2 for pixels at a distance d from the nearest dump, relative to the change in pollution for pixels between 30 and 40 km away from the nearest dump, in time period q. The key identifying assumption is therefore that in the absence of the 2018 China waste ban, the pollution—distance-to-dump gradient would have remained similar to what it was before the policy was announced.

Figure 11: PM2.5 Around Dumps and the China Ban: Event Studies



Notes: Event-study estimates from Equation 12 on the monthly cell-level panel (2012–2023), with ring-by-quarter coefficients relative to the omitted 30–40km ring and 2017Q2 (the last pre-announcement quarter) as the reference period. All specifications include cell and district-by-month fixed effects, with standard errors clustered by district. Vertical lines mark China's WTO notification (July 2017) and the ban taking effect (January 2018).

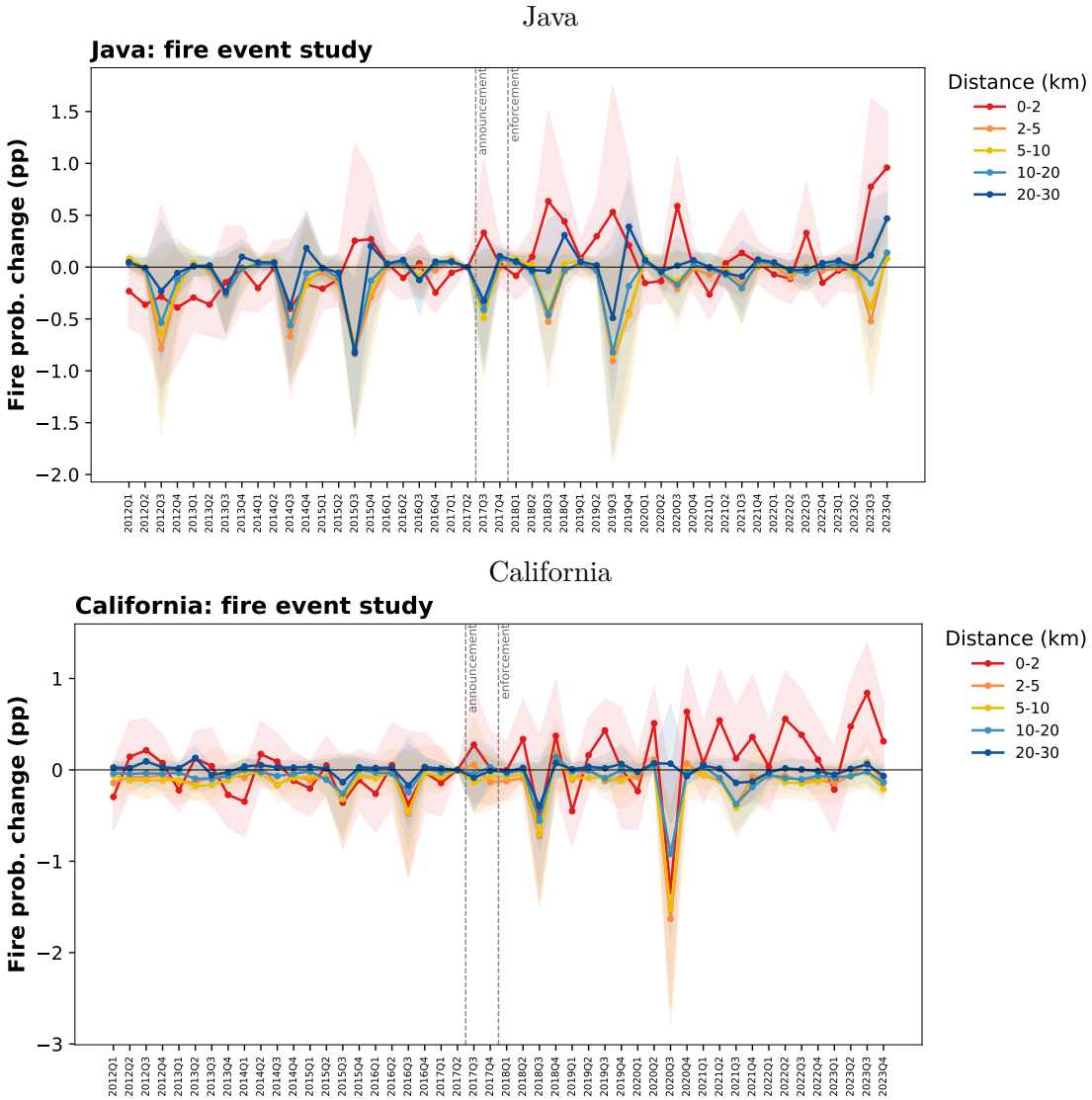
Figure 11 plots the coefficients of these event studies for Java and California. Vertical lines mark China's announcement of the National Sword Policy (July 2017) and the beginning of the enforcement of the import ban (January 2018). In Java we see reasonably flat pre-trends in different distance buckets before the announcement. After 2018, we see consistently elevated $PM_{2.5}$, especially in the distance rings closest to dumps.

In California, however, we see a clear seasonal pattern in particulate matter that varies with distance to the nearest dump. The seasonal pattern does not appear to be significantly altered after 2018, however, to test this more precisely, we estimate versions of the same regression with pixel by calendar month fixed effects. We also estimate a version aggregated to annual temporal resolution to smooth out the seasonality. Appendix Figure A14 shows the results of both of these regressions. The seasonal fixed effects and annual aggregation both smooth the pre-trends in California. In both cases, we see no significant increase in pollution close to dumps—in fact if anything, the opposite is true, with a particularly sharp decrease in 2020, though this effect is confounded with the effects of COVID and a particularly bad wildfire season. In both of these supplemental specifications, the increase in particulate matter in Java is still clearly apparent.

Appendix Tables A8 (for Java) and A9 (for California) pool time periods and estimate a single coefficient for each bin in the post-enforcement period, relative to changes in far-from-dump pixels relative to the pre-announcement period, thus excluding the anticipatory period between July 2017 and January 2018, when the ban had been announced but not yet enforced. In the main specification in Column (1), we see a 6.2% increase in pollution in the 0-2 km bin for Java, which fades to a 2.6% increase in the 10-20 km bin. In California, estimated effects are negative and significant across all distance bins, ruling out large increases in pollution as a result of the increased landfilling that occurred after 2018. These estimates are robust across a range of specifications, including cell by month fixed effects in Column (2), aggregating to annual resolution in Column (3), and dropping border pixels in Column (4). Dropping 2020 attenuates the estimates a bit in both geographies, especially in California.

Figure 12 shows the same event studies with a dummy for a VIIRS fire detection as the dependent variable. This time in both Java and California, we see elevated fires after 2018, but only in the pixels closest to dumps. Appendix Figure A15 shows the time-pooled coefficients for each distance bin for the post-enforcement period. In Java the coefficient on the 0-2 km distance bin is positive and significant, giving evidence for a clear mechanism for the increased pollution. In California, the point estimate is positive, but not significant. Since the detected fires in California are most likely methane flaring, which does not generate large amounts of particulate pollution, this explains the contrast with the results for Java.

Figure 12: Garbage Fires Around Dumps and the China Ban: Event Studies



Notes: Same design as Figure 11, with an indicator for at least one VIIRS active-fire detection in the cell-month as the outcome.

7.2 Waste Trade and Welfare

Taken together, these results suggest that much of the diverted waste was disposed of in open dumps where it caught fire and generated large amounts of pollution. Yet the trade data shows that Indonesian importers paid a positive price for these scrap imports suggesting some amount of private benefits. Therefore, we conduct a simple, back-of-the-envelope calculation to compare the magnitude of these private benefits to the external costs of the pollution generated from improper disposal.

To calculate the benefits to importers, we simply compute the change in consumer surplus in each banned waste category directly from the observed pre- and post-ban import prices and quantities in the BACI trade data. We assume that China's exit shifted the supply of scrap toward the remaining importers, so Indonesia's change between its pre-ban (2012–2016) and post-ban (2018–2023) average prices and quantities traces its import-demand curve[21]. Assuming linear demand, we can calculate consumer surplus as the area of the trapezoid between these two points (see Figure A17). While the linear demand assumption is conservative, since constant elasticity demand curves would give lower estimates of consumer surplus, two other caveats are worth noting. First of all, this exercise assumes that Indonesia's demand curve remained constant over the period. This is a reasonable approximation—imports spiked around the time of the ban, but do not show much of a time trend otherwise (see Appendix Figure A16 for plastic imports). Secondly, it assumes the quality of the imported waste was not changed by the ban. Scrap quality can vary (e.g. contamination levels), but assuming constant quality is probably a reasonable approximation.

Making these assumptions allows us to calculate increases in consumer surplus of $21 million for plastic scrap (HS 3915), $34 million for paper waste (HS 4707), $3 million for textiles (HS 5103, 5104, 5202, 5505, 6310), and $11 million for slag (HS 2619 and 2620) covering the major categories of materials banned by China. This gives a headline estimate of $69.2 million for all categories, or $55.5 million if we just count paper and plastic, which are the materials most likely to end up in dumps (since slag is more valuable for industry and more easily recyclable).

We compare this consumer surplus to one particular external cost, the air pollution close to dumps generated from the diverted imports, noting that water pollution, pests, and other nuisances are neglected in this analysis. To measure air pollution precisely, we estimate the

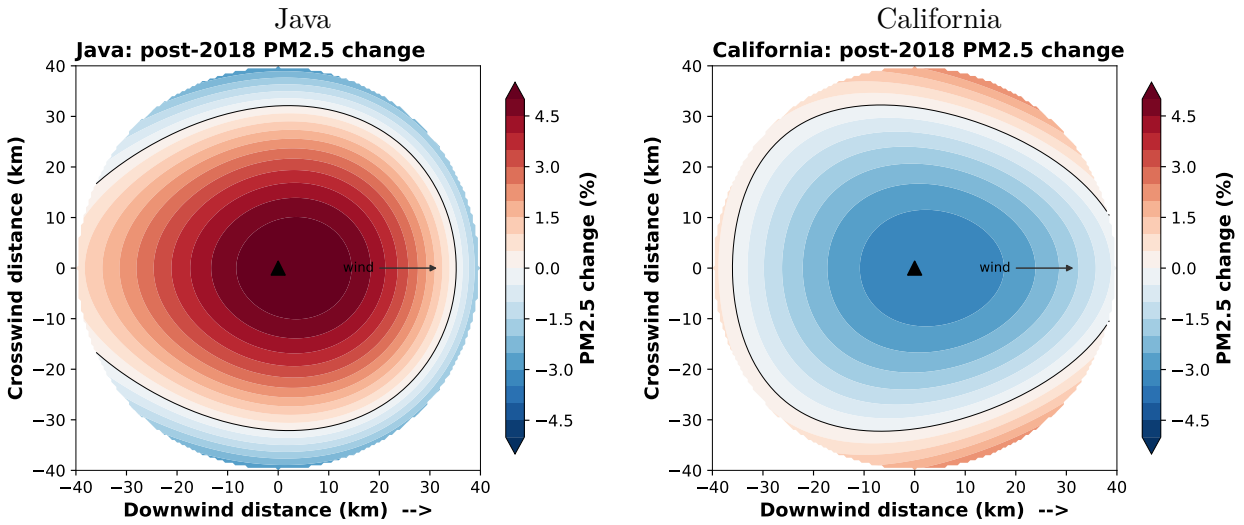
[21] For prices we use annual unit prices from the BACI data, converted to Indonesian rupiah, and then deflated to constant 2016 prices using exchange rates and Indonesian CPI from the World Bank. Prices do not reflect a PPP adjustment since these are tradable goods, but we convert back to 2016 USD for ease of interpretation.

average monthly pollution plume induced by the diverted waste imports as a flexible function of distance and wind direction, following Schlenker and Walker (2016). We use the following specification:

$$\log PM_{it} = f(d_i, w_{it}) \times P_t + \alpha_i + \lambda_{a(i),t} + \varepsilon_{it}. \tag{13}$$

In this regression, P_t denotes dummies for the announcement period (post July 2017) and the enforcement period (post January 2018). We allow the plume in each period to change flexibly as a function of distance to the nearest dump d_i, and monthly prevailing wind direction relative to the nearest dump, w_{it}. We use a cubic polynomial of downwind distance and a quadratic polynomial of cross-wind distance for $f(d_i, w_{it})$. Figure 13 plots the predicted change in pollution for the post-enforcement period as a function of distance and direction to the nearest dump for Java and California relative to far upwind pixels. In Java, we see a clear increase in pollution close to dumps, whereas in California, we see a modest decrease if anything.

Figure 13: Wind-Relative PM2.5 Plumes Around Dumps



Notes: Enforcement-period surface f_E from Equation 13, plotted as the percent change in PM2.5 relative to the upwind far-field (30–40km) mean. The horizontal axis is downwind distance from the nearest dump given the month's ERA5 wind direction; the vertical axis is crosswind distance.

To translate the estimated pollution changes into health impacts, we overlay the fitted plume surfaces on gridded population data. For each cell within 40km of a dump, we evaluate the enforcement-period plume surface at the cell's relative position in each month, and subtract off our estimated change (which is defined relative to changes at the far upwind pixels) to construct a counterfactual pollution path.

We then combine this exposure change with the gridded population data and a concentration-response function from Burnett et al. (2018) to estimate excess mortality. We construct confidence intervals by conducting 1000 Monte Carlo draws jointly from the plume coefficient covariance matrix and the sampling distribution of the CRF parameter. Several interpretive cautions apply. First, as stated previously, the plume identifies pollution changes near dumps relative to far upwind pixels. If the waste import shock changed pollution levels in these far upwind pixels, that change is not identified in our analysis. This is possible, for example if recycling factories generate pollution that is not co-located with dumps, we will underestimate mortality impacts. Secondly, overlaying these plumes on population data abstracts from post-ban migration or additional defensive actions that households may take in response to the increase in pollution. Lastly, these plumes are average, both over time and across dumps. If the pollution effect varies significantly with respect to population, that could affect our conclusions.

Conditional on these caveats, we estimate an additional 4,381 annual premature deaths in Java (95% CI: [1,836, 7,295]) as a result of the post 2018 waste import shock. Dividing our estimated Indonesian consumer surplus from paper and plastic imports by this number gives us an estimate of the welfare gains per life lost of $12,672 (95% CI: [$7,375, $30,245]). This goes up to $15,790 if slag and textiles are included. These numbers are orders of magnitude lower than standard estimates of the value of a statistical life.

These numbers do not reflect a complete welfare analysis, since they do not reflect producer surplus in exporting countries, and furthermore, the mortality estimates only reflect a subset of Indonesia, whereas the consumer surplus is at the country level. Still, they should be relevant for Indonesian policy makers considering whether to restrict waste imports. Indeed, if we conduct an analogous exercise to measure the lost producer surplus corresponding to the collapse in demand for US waste exports after the China import ban, we estimate $190.4 million in lost welfare in the US, although this number is not directly comparable to the Indonesian figure, since the US sends waste to multiple export destinations. Still, it highlights the distributional issues arising from this trade.

8 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 learning methods to choose an approximately optimal subset of

predictions to verify using high-resolution imagery. We choose these predictions to make our estimates of open dump prevalence and changes as precise as possible, and we come up with a simple formula to characterize the ex-post efficiency gains, relative to uniform sampling, in the presence of model miscalibration. These methods could be 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 show 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 in low- and middle-income countries (12.5%) 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 waste imports. Event studies and a shift-share IV show that this increase occurs in countries that received an influx of imports diverted by the China waste ban. This provides evidence that a meaningful share of exported recycled materials are ultimately discarded in poorly managed dumps in low-income countries. We also find that the import shock increased $PM_{2.5}$ levels close to dumps, and show evidence that increased garbage fires are the mechanism behind this pollution. A back-of-the-envelope welfare calculation for Indonesia estimates that the ratio of consumer surplus from increased imports to lives lost from the resulting pollution is $12,672, orders of magnitude lower than standard VSL estimates.

There is plenty of room for further research on this topic. Our data and methods should be useful to more precisely measure other externalities associated with informal waste sites, which will lead to an improved understanding of waste impacts on human health and ecosystem services. These are urgent research questions, given the long lifetimes of plastic scrap and the rapid growth rate of waste 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.

Brooks, Amy L., Shunli Wang, and Jenna R. Jambeck. 2018. "The Chinese import ban and its impact on global plastic waste trade." *Science Advances* 4 (6): eaat0131. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.aat0131. https://www.science.org/doi/abs/10.1126/sciadv.aat0131.

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.

Burnett, Richard, Hong Chen, Mieczysław Szyszkowicz, Neal Fann, Bryan Hubbell, C. Arden Pope, Joshua S. Apte, et al. 2018. "Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter." *Proceedings of the National Academy of Sciences* 115 (38): 9592–9597.

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.

Cochran, William G. 1977. *Sampling Techniques.* 3rd. New York: John Wiley & Sons.

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.

Donkelaar, Aaron van, Randall V. Martin, Chi Li, and Richard T. Burnett. 2019. "Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors." *Environmental Science & Technology* 53 (5): 2595–2611. Accessed February 4, 2020. https://doi.org/10.1021/acs.est.8b06392.

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.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110 (8): 2586–2624. Accessed December 23, 2021. https://www.aeaweb.org/articles?id=10.1257/aer.20181047.

Gordon, Matthew, Eliana Stone, Megan Ayers, and Luke Sanford. 2026. "Debiasing Estimates of Global Forest Cover Loss." *AEA Papers and Proceedings.*

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.

Guillame-Bert, Mathieu, Sebastian Bruch, Richard Stotz, and Jan Pfeifer. 2023. "Yggdrasil Decision Forests: A Fast and Extensible Decision Forests Library." In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* 4068–4077.

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.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. 2020. "The ERA5 Global Reanalysis." *Quarterly Journal of the Royal Meteorological Society* 146 (730): 1999–2049.

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. 2010. "Consumer Waste, Backhauling, and Pollution Havens." *Journal of Applied Economics* 13 (2): 283–304.

———. 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.

OpenStreetMap. 2026. "OpenStreetMap." Accessed July 9, 2026. https://www.openstreetmap.org.

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.

Ritchie, Hannah. 2022. "Ocean plastics: How much do rich countries contribute by shipping their waste overseas?" *Our World in Data.*

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.

Schlenker, Wolfram, and W. Reed Walker. 2016. "Airports, Air Pollution, and Contemporaneous Health." *The Review of Economic Studies* 83 (2): 768–809. Accessed May 6, 2020. https://academic.oup.com/restud/article/83/2/768/2461206.

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.

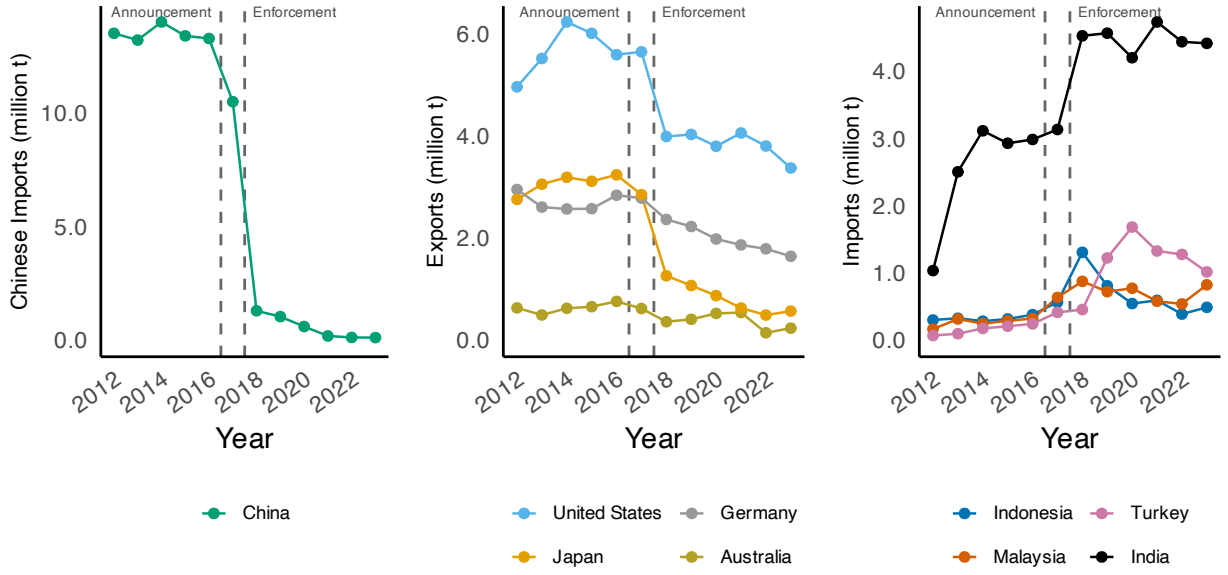
United States Environmental Protection Agency. 2024. "New International Requirements for the Export and Import of Plastic Recyclables and Waste." Accessed March 25, 2025. https://www.epa.gov/hwgenerators/new-international-requirements-export-and-import-plastic-recyclables-and-waste#:~:text=In%20May%202019%2C%20187%20countries,country%20and%20any%20transit%20countries..

World Bank. 2024. *World Development Indicators.* Technical report. The World Bank Group. https://databank.worldbank.org/source/world-development-indicators.

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

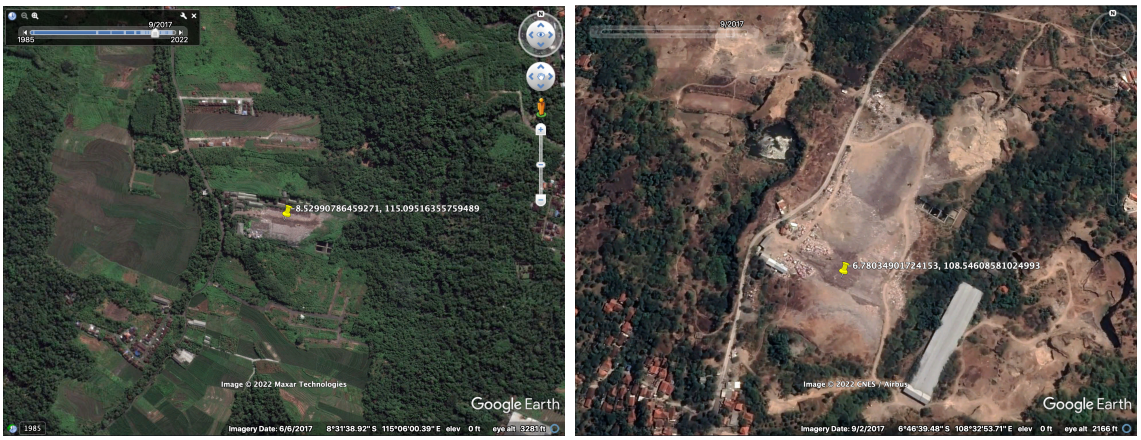
A Additional Figures

Figure A1: Plastic and Paper Scrap Imports and Exports Over Time



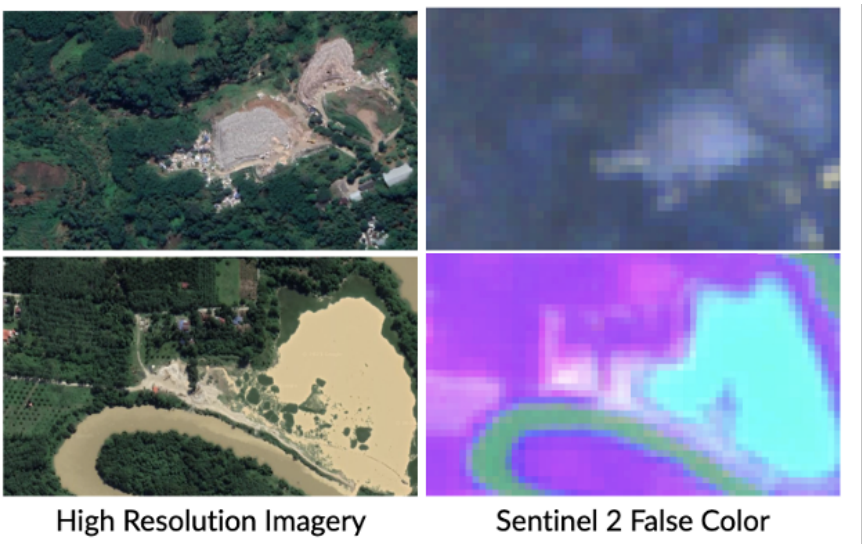
Notes: Figure shows trade in the two major categories of banned waste under the National Sword Policy: plastic waste and scrap (HS 3915) plus unsorted waste paper (HS 4707.90), which together account for over 84% of China's 2016 (pre-policy) imports in banned categories. Units are millions of tonnes. Data are from CEPII-BACI, 2012–2023. Left panel shows China's imports; center panel shows exports of four large pre-ban senders (United States, Japan, Germany, Australia); right panel shows imports of four major receiving countries (Indonesia, Malaysia, Turkey, India). We note that almost all of India's imports are in unsorted waste paper, not plastic waste. Dashed vertical lines mark the announcement of the National Sword Policy (July 2017) and its effective date (January 2018).

Figure A2: High-Resolution Satellite Data Verification Examples



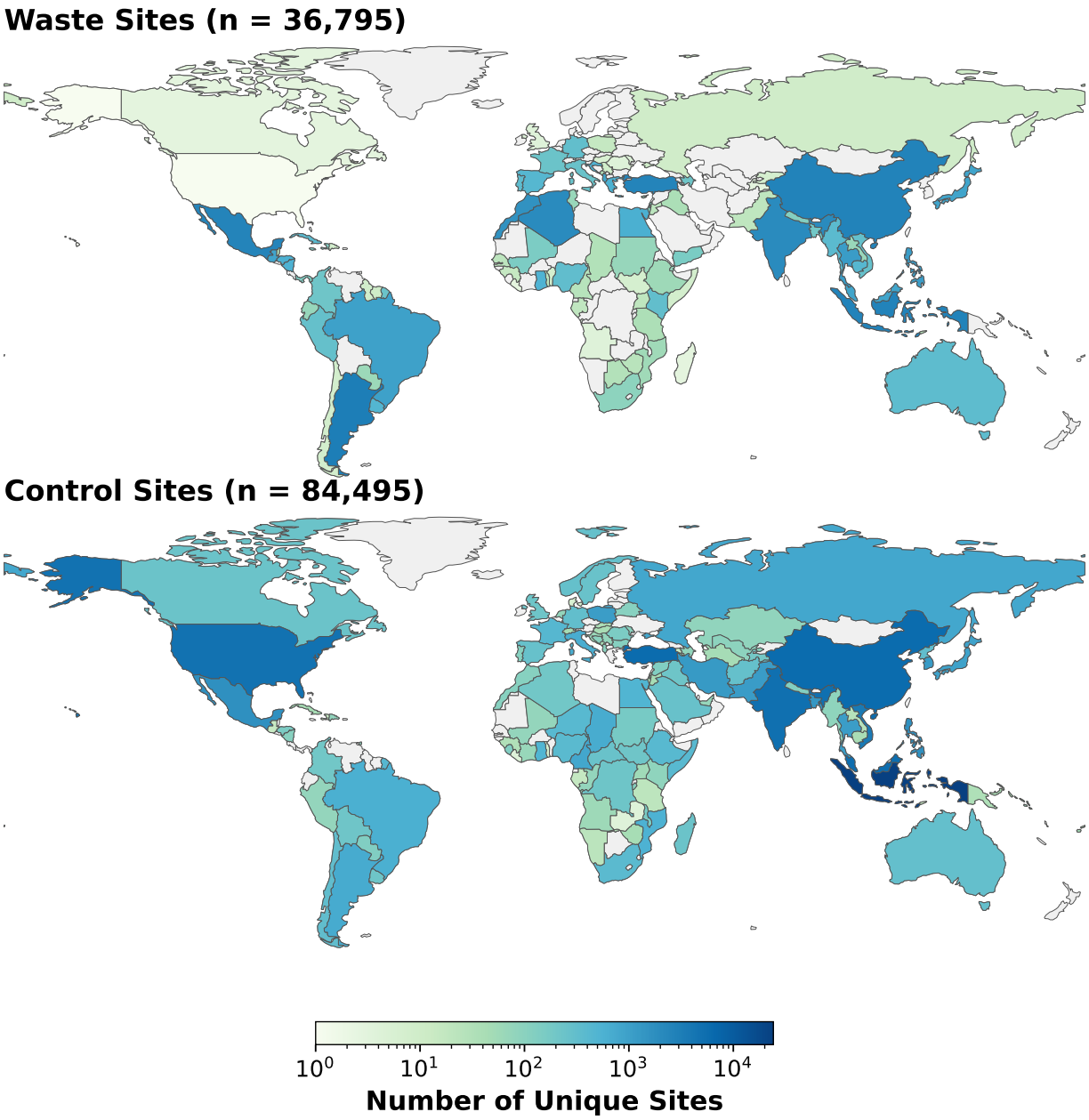
Notes: Figures show two examples of waste sites in Google Earth high-resolution satellite imagery.

Figure A4: Comparison of Google Earth High-Resolution Imagery and Sentinel-2 Imagery



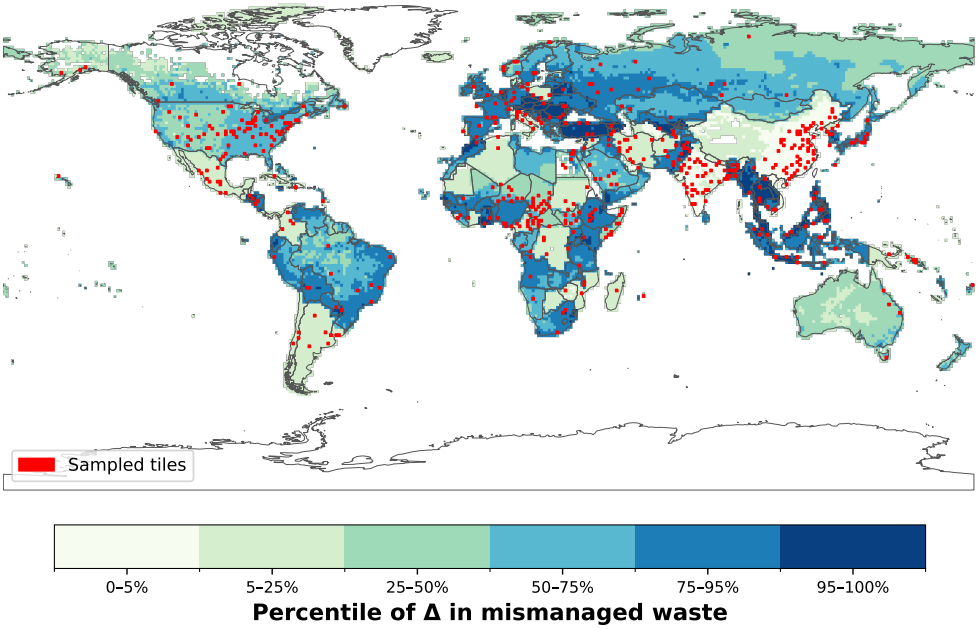
Notes: Figure shows high-resolution historical imagery from Google Earth Pro (left) and the same sites in Sentinel-2 false-color imagery (right). Dumps are able to be easily verified by human annotators using the high-resolution imagery. Models are trained on the right imagery, which includes radar and infrared information that is useful in classifying dumps.

Figure A3: Distribution of Training Data by Country



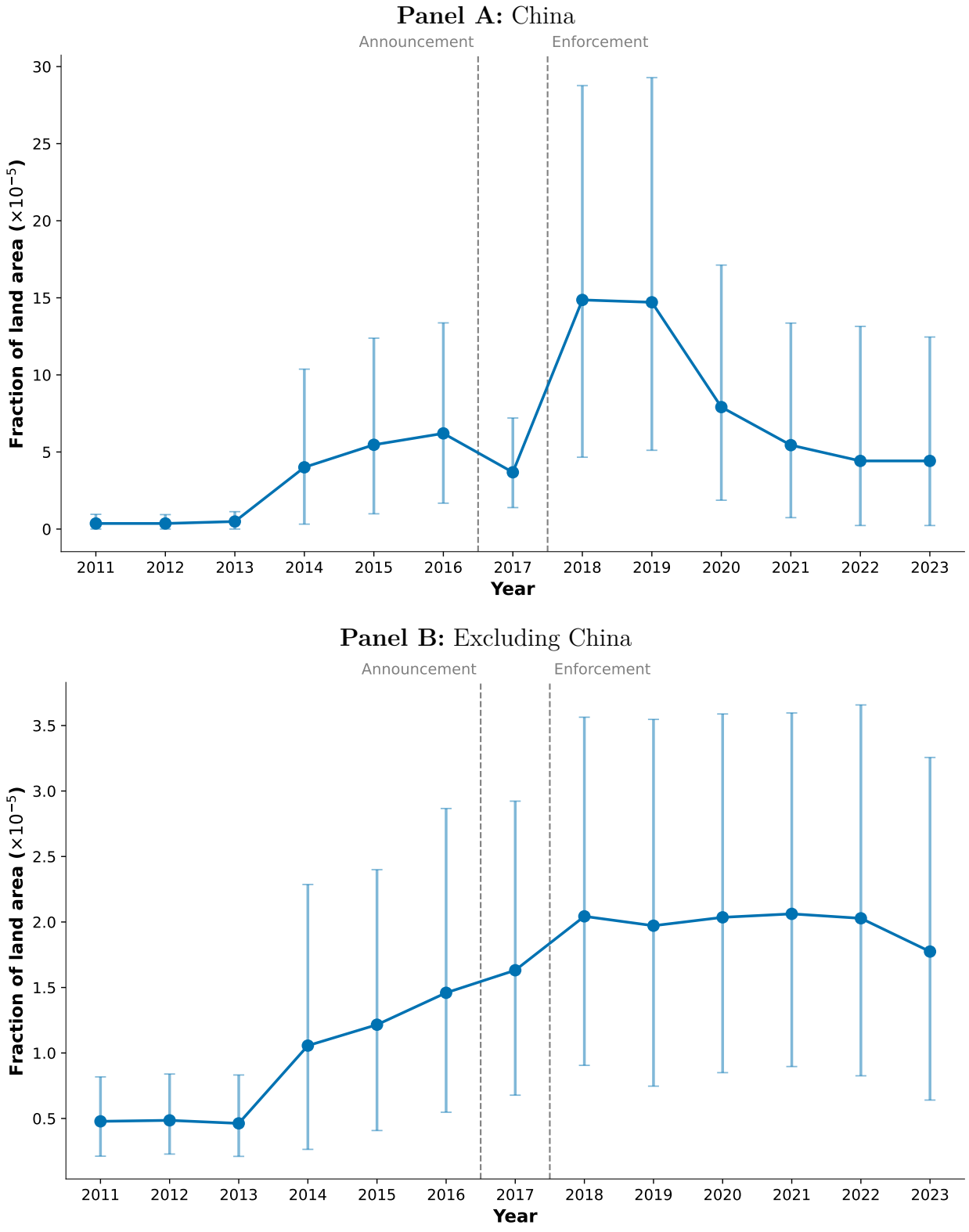
Notes: The distribution of training sites (both waste sites and control sites) by country based on crowdsourced and other data collection. Counts reflect site-year observations, not necessarily unique sites.

Figure A5: Strata and Tile Sample for Global Analysis



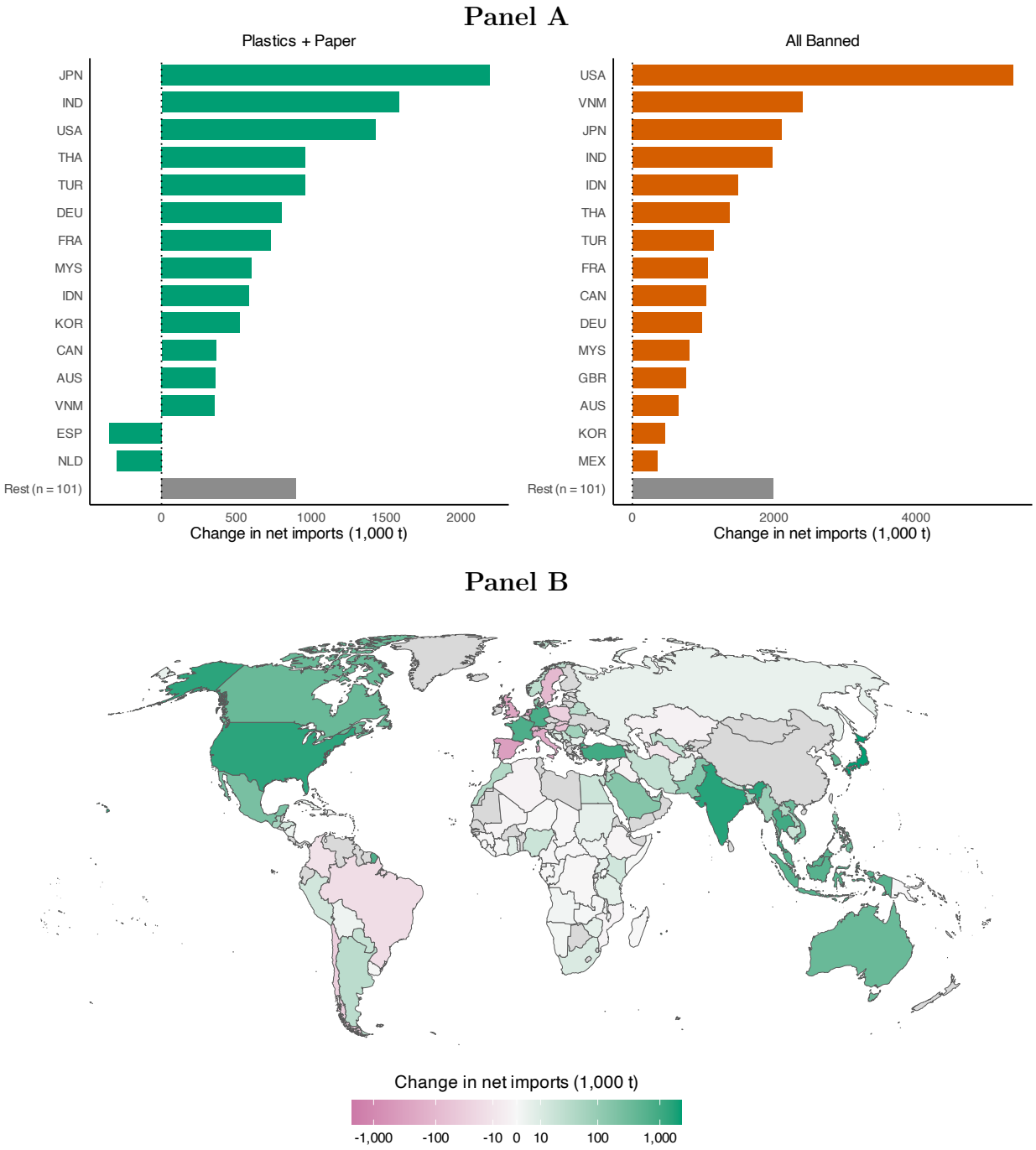
Notes: Figure illustrates global deployment on $1° \times 1°$ tiles. Colors show the six strata based on estimated changes in mismanaged imports 2017–2019, with the selected tiles overlaid in red.

Figure A6: Share of Land Area Covered by Waste Sites: China vs. Rest of World



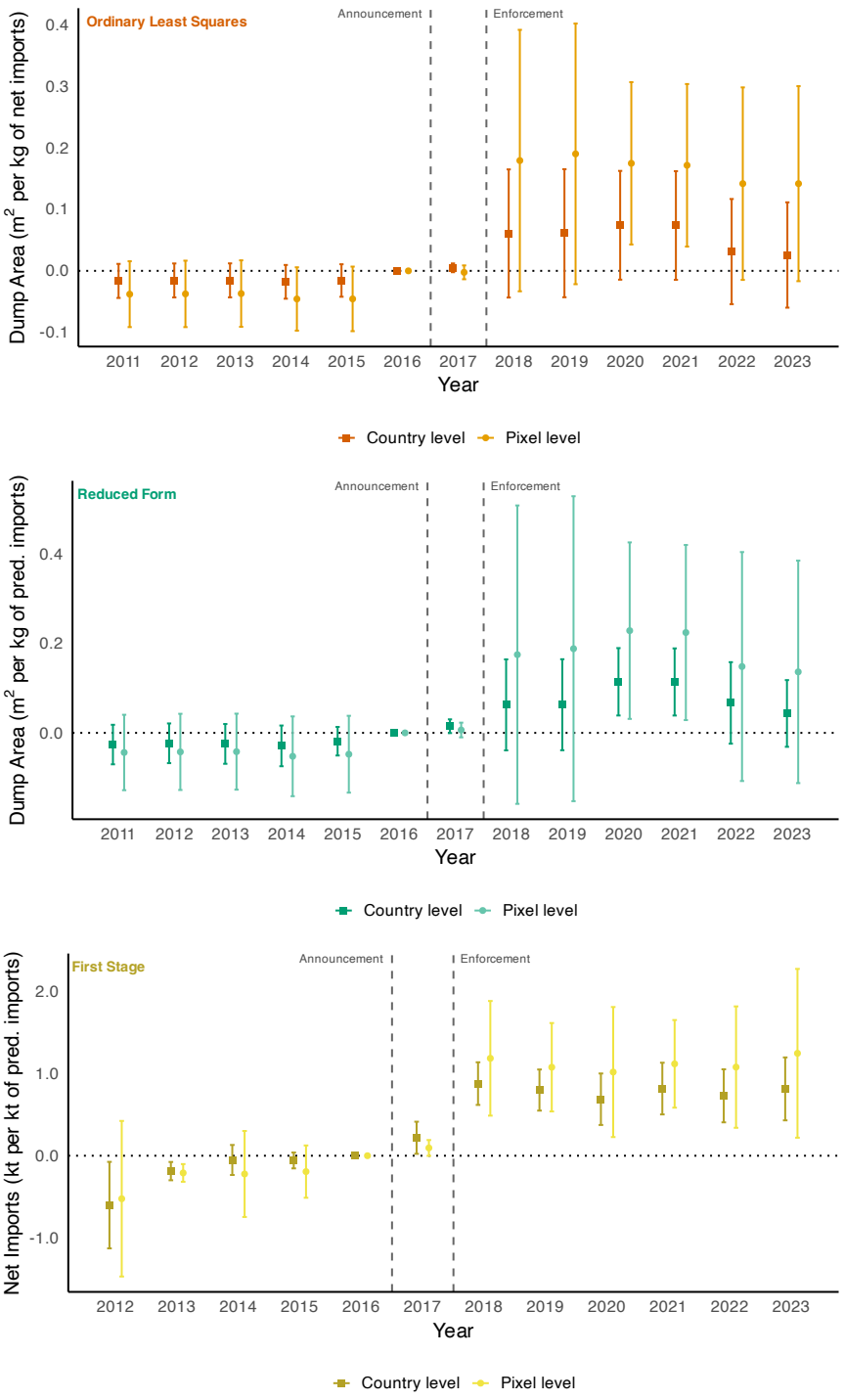
Notes: Estimates based on verified pixels, split into pixels inside and outside China. 95% confidence intervals from a stratified tile-cluster bootstrap. Dashed line marks the China waste import ban (January 2018).

Figure A7: Change in Net Imports Following China's National Sword Policy



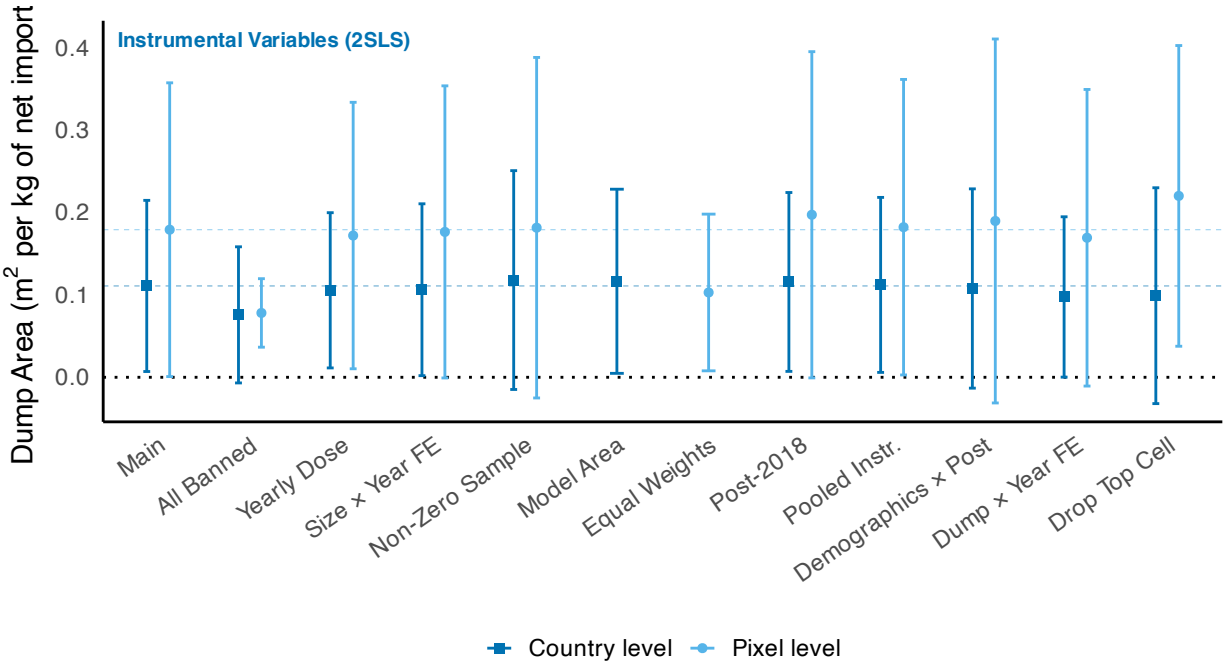
Notes: Figure shows the treatment variable in our analysis of trade and open dumps. Changes in net imports are defined as the change in imports minus exports between 2019 and 2016. Panel A shows the fifteen countries with the largest absolute changes and the pooled remainder for the main categories (left) and all categories (right) of banned materials. Panel B maps the change in net imports for all countries (excluding China) with available BACI trade data.

Figure A8: Effect of Change in Net Imports on Open Dump Area, Event Study Version



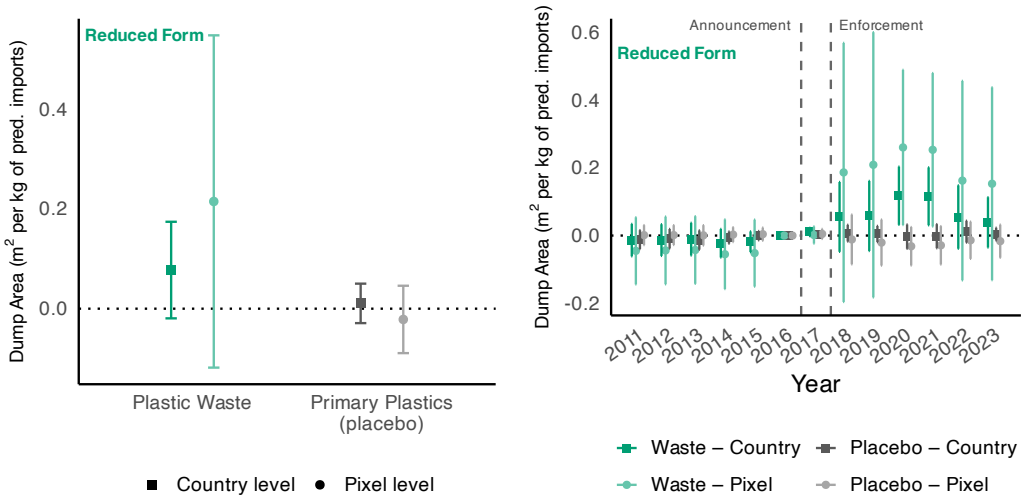
Notes: Figure shows alternative versions of the main instrumental variables event study figure. Top panel shows OLS (equation 10; m^2 per kg of realized net imports); middle panel shows the reduced form (dump area on predicted diversion, in m^2 per kg of *predicted* diversion); bottom panel shows the first stage of yearly net imports (thousands of tonnes) on predicted diversion × year. The first-stage panel begins in 2012, the first year of BACI trade data. In each panel, darker squares are the country level and lighter circles the pixel level, with 95% confidence intervals.

Figure A9: Effect of Change in Net Imports on Open Dump Area, Robustness



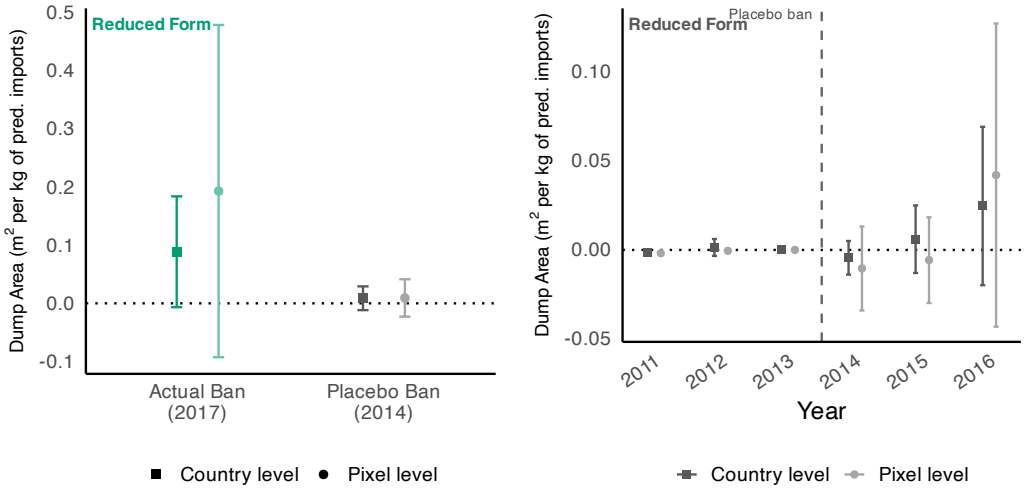
Notes: Figure shows robustness of the main IV estimates of Table 1, Column (2) at the country level (darker squares) and pixel level (lighter circles) across alternative specifications. Dashed horizontal lines mark the two main estimates. From left, the specifications are the following: Main (the Table 1 specification); All Banned (every waste heading in the ban's catalogue); Yearly Dose (annual net imports instrumented by time-varying predicted diversion); Size × Year FE (land-area-quartile × year fixed effects); Non-Zero Sample (countries with verified dumps in every year); Model Area (debiased machine-learning dump area as the outcome, 2017–2021, country level only); Equal Weights (pixel design weights renormalized within country × year, pixel level only); Post-2018 (ban dated at enforcement); Pooled Instr. (instrument built on the pooled rather than per-HS-code network); Demographics × Post (baseline log GDP per capita, log population density, and log land area interacted with post); Dump × Year FE (baseline-dump-area-quartile × year fixed effects); and Drop Top Cell (instrument rebuilt excluding US waste paper, the largest Rotemberg-weight cell). Bars are 95% confidence intervals, standard errors clustered by country.

Figure A10: Effect of Change in Net Imports on Open Dump Area, Shift-Share Placebo Analysis



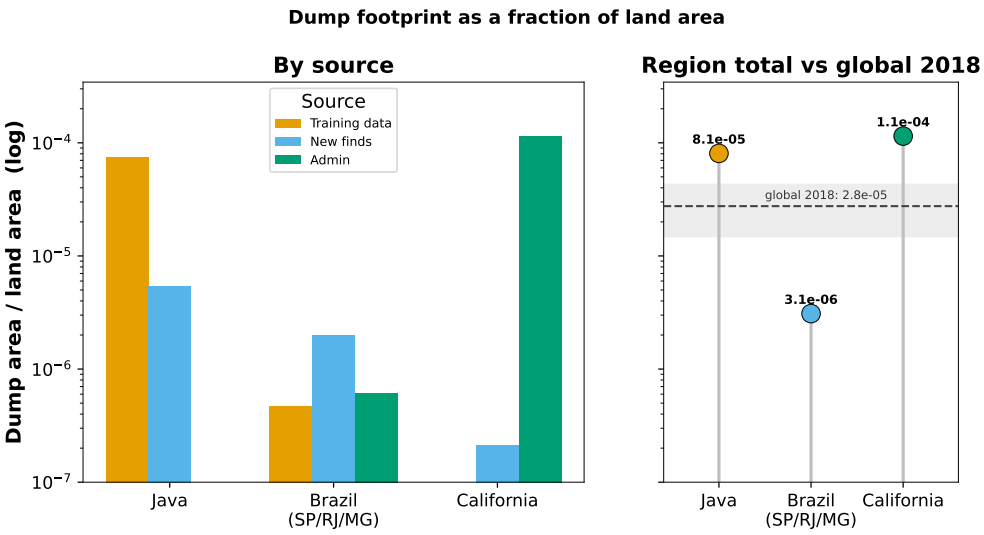
Notes: Figure shows the commodity placebo, where the same shift-share instrument construction is applied to primary-form plastics (HS 3901–3914), a commodity never banned by China. Left panel shows pooled pre/post coefficients for the waste and placebo exposures at each level. Right panel shows the year-by-year version, with both exposures interacted with event year in a single regression; dashed lines mark the ban's announcement and enforcement. Coefficients are in m^2 per kg of predicted diversion; 95% confidence intervals, standard errors clustered by country.

Figure A11: Effect of Change in Net Imports on Open Dump Area, Placebo Ban Analysis



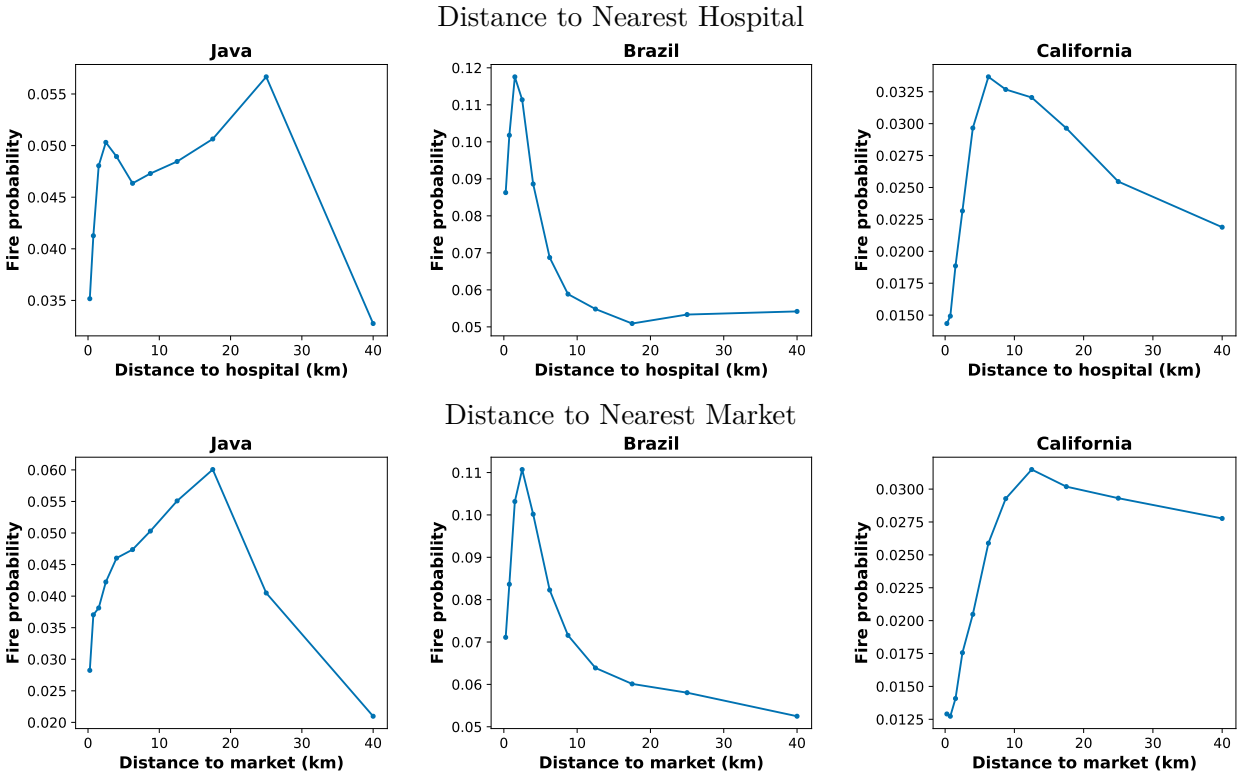
Notes: Figure shows the timing placebo, where the reduced form of the main design is re-estimated on pre-ban years only (2011–2016), with a counterfactual placebo ban dated between 2013 and 2014. The left panel shows the pooled reduced-form coefficient on predicted diversion at the actual ban in green, next to the same coefficient at the placebo ban (grey, pre-ban sample). The right panel shows the year-by-year reduced form around the placebo ban on the pre-ban sample, with 2013 as the reference year. Coefficients are in m^2 per kg of predicted diversion; 95% confidence intervals, standard errors clustered by country.

Figure A12: Case Study: Dump Footprint as a Fraction of Land Area



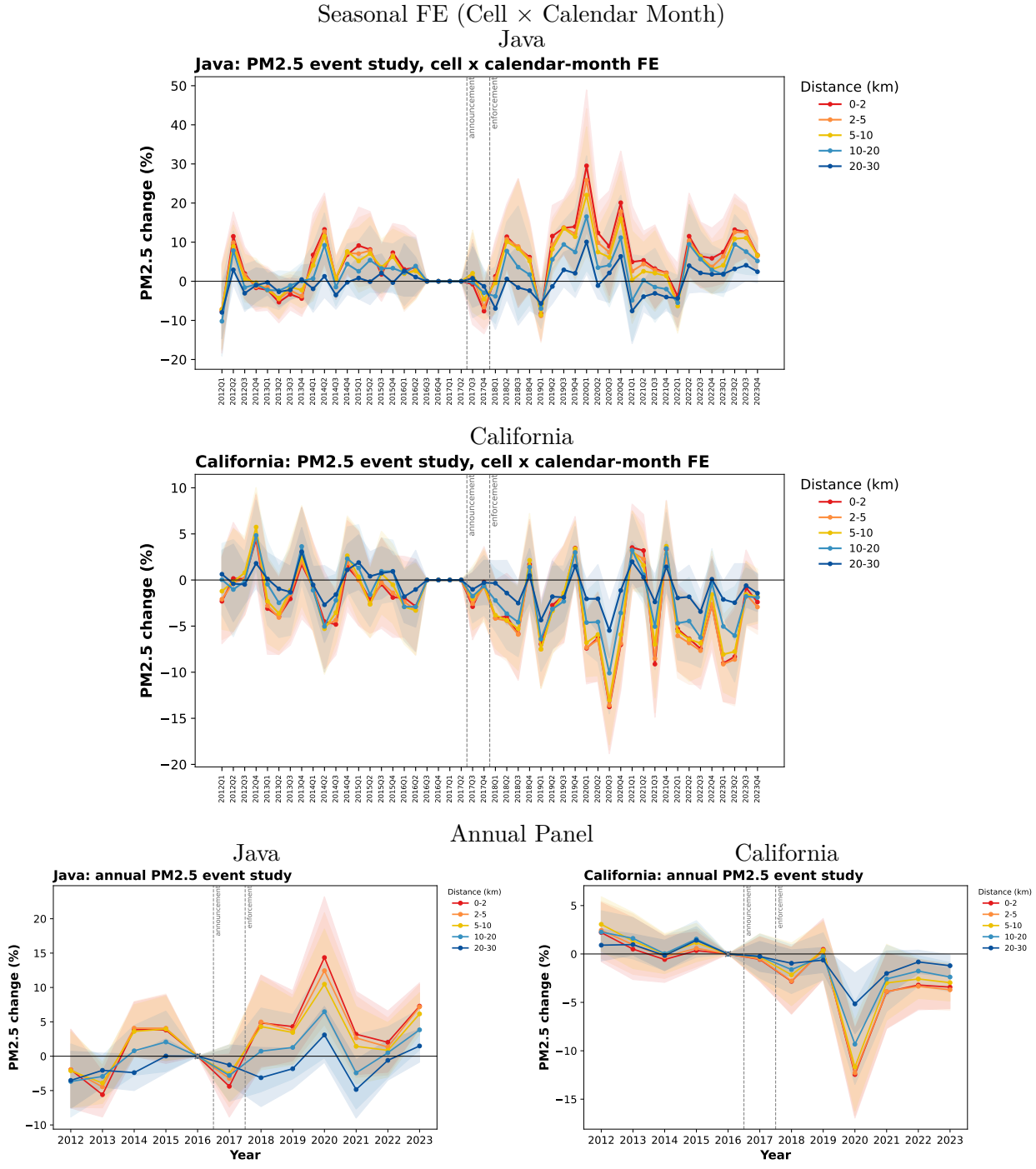
Notes: Total dump area from all sources as a share of each case-study region's land area, by source (left, log scale) and in total (right). The dashed line and shaded band show our globally representative 2018 estimate of verified dump prevalence and its 95% bootstrap confidence interval from Section 5. Note that the California estimates include sanitary landfills which are not comparable to open dumps.

Figure A13: Case Study: Fire Probability by Distance to Placebo Landmarks



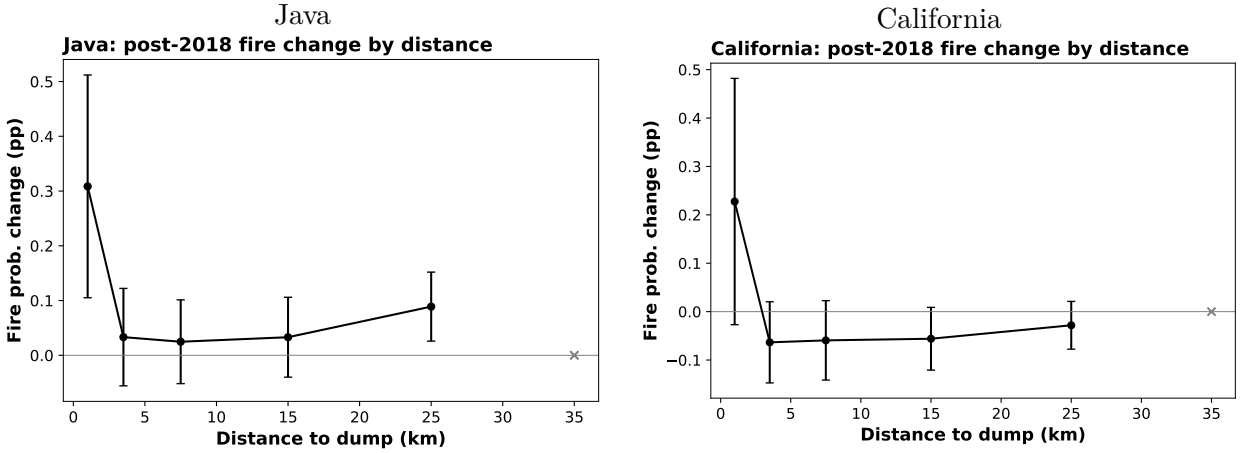
Notes: Probability of at least one VIIRS fire detection per cell-year (pooled 2012–2023), by distance to the nearest hospital (top row) or marketplace (bottom row). Hospital and marketplace locations are from OpenStreetMap (OpenStreetMap 2026), retrieved via the Overpass API in July 2026.

Figure A14: PM2.5 Event Studies: Seasonal Fixed Effects and Annual Panels



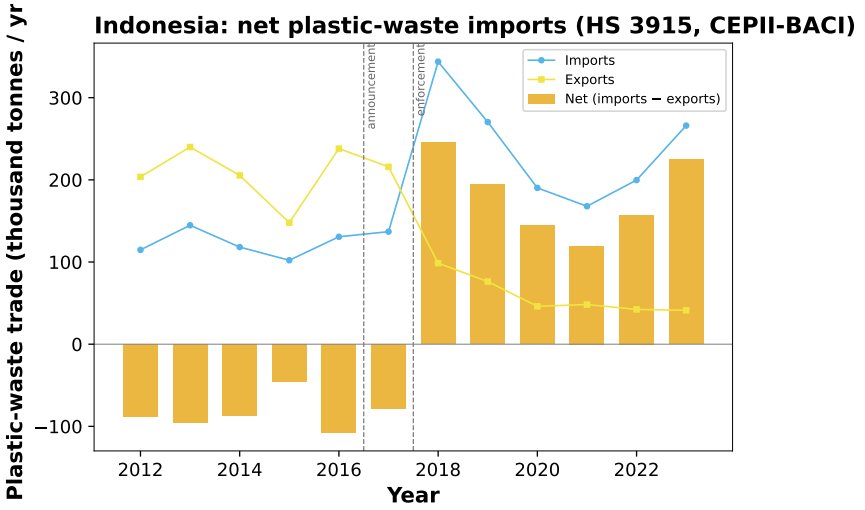
Notes: Alternative specifications of the event studies in Figure 11. The top row replaces the cell fixed effect with cell-by-calendar-month, absorbing each cell's seasonal cycle; the event study then uses the last pre-announcement occurrence of each calendar quarter (2016Q3–2017Q2) as references. The bottom row collapses the panel to cell-by-year with cell and district-by-year fixed effects, $Announce = 2017$, $Enforce =$ 2018 onward, and 2016 as the reference year.

Figure A15: Garbage Fires Near vs. Far From Dumps



Notes: Near–far difference-in-differences with an indicator for at least one VIIRS fire detection in the cell-month as the outcome, same design as Figure 11. California fire detections are contaminated by wildfires.

Figure A16: Indonesia: Net Imports of Plastic Waste Around the China Ban



Notes: Annual net imports (imports minus exports, tonnes) of plastic waste and scrap (HS 3915) for Indonesia, 2012–2023, from CEPII-BACI (Gaulier and Zignago 2010). Vertical lines mark China's WTO notification of the National Sword Policy (July 2017) and the ban taking effect (January 2018).

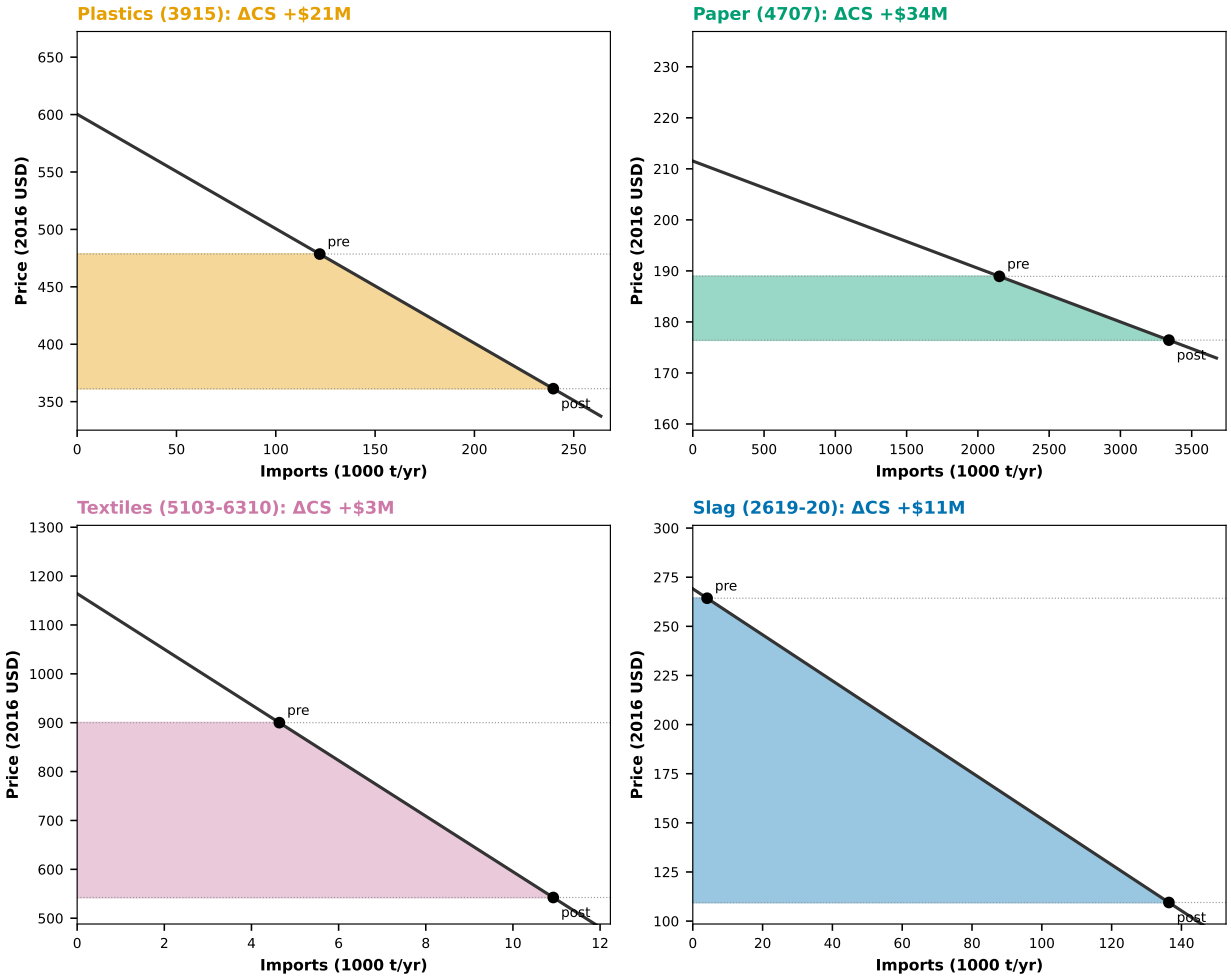


Figure A17: Indonesia consumer-surplus gain from the supply shock induced by the 2018 China waste-import ban, by category of banned waste. Each panel reads the surplus change off the observed pre (2012–16) → post (2018–23) price and quantity changes. Points mark pre/post (quantity, price); the solid line is the revealed linear demand; the shaded area is the realized surplus change. Constant 2016 prices deflated using Indonesian CPI converted to 2016 USD.

B Additional Tables

Table A1: Model Input Data

Source	Features
Sentinel-1	VV and VH backscatter (interferometric wide swath mode, averaged over ascending and descending orbits), plus the three radar indices in Table A2, Panel A.
Sentinel-2	All 13 top-of-atmosphere bands (B1–B12, including B8A), plus the 19 spectral indices in Table A2, Panel B, and the median annual Cloud Score+ value.
Focal statistics	Mean and standard deviation of each Sentinel-1 and Sentinel-2 band and index in a 3-pixel window.
Ports	Distance to the nearest port.
Population	Population density for 2010, 2015, and 2020, at the pixel and averaged over a 9 km $\times$ 9 km neighborhood (Earth Science Data Systems, NASA 2024).
Dynamic World	Probability of complete coverage by water, trees, grass, flooded vegetation, crops, shrub and scrub, built, bare, and snow and ice land uses.

Table A2: Sentinel-1 and Sentinel-2 Indices

Index	Formula	Description
Panel A: Sentinel-1		
DpRVIVV	$\frac{4\times VH}{VV+VH}$	Dual-polarized radar vegetation index; vegetation density.
VHVVR	$\frac{VH}{VV}$	Cross- to co-polarized ratio; vegetation structure and volume scattering.
VVVHR	$\frac{VV}{VH}$	Inverse of VHVVR; complementary surface information.
Panel B: Sentinel-2		
NDVI	$\frac{B8-B4}{B8+B4}$	Normalized difference vegetation index; vegetation greenness.
BSI	$\frac{(B11+B4)-(B8+B2)}{(B11+B4)+(B8+B2)}$	Bare soil index; bare soil vs. vegetation and water.
NDWI	$\frac{B3-B8}{B3+B8}$	Normalized difference water index; open water.
UI	$\frac{B12-B8}{B12+B8}$	Urban index; built-up areas.
MNDVI	$\frac{B8-B12}{B8+B12}$	Modified NDVI (SWIR2 in place of red).
MNDWI	$\frac{B3-B11}{B3+B11}$	Modified NDWI; water, suppressing built-up noise.
CSI	$\frac{B8}{B12}$	Char soil index; charred or burned soils.
DBSI	$\frac{B11-B3}{B11+B3} - \frac{B8-B4}{B8+B4}$	Dry bareness index; dry barren areas.
NDSWIR	$\frac{B8-B11}{B8+B11}$	Normalized difference NIR–SWIR1; vegetation moisture.
NormNIR	$\frac{B8}{B8+B3+B4}$	NIR share of visible–NIR reflectance.
NormG	$\frac{B3}{B8+B3+B4}$	Green share of visible–NIR reflectance.
NormR	$\frac{B4}{B8+B3+B4}$	Red share of visible–NIR reflectance.
SWIR1SWIR2	$\frac{B11}{B12}$	SWIR ratio; mineral and soil composition.
NIRNIR2	$\frac{B8}{B8A}$	NIR ratio; vegetation structure.
RE1RE2	$\frac{B5}{B6}$	Red-edge ratio; chlorophyll content and vegetation stress.
RE2RE3	$\frac{B6}{B7}$	Red-edge ratio; vegetation health.
BG	$\frac{B2}{B3}$	Blue/green ratio; water quality and algal blooms.
BR	$\frac{B2}{B4}$	Blue/red ratio; atmospheric scattering and water properties.
GR	$\frac{B3}{B4}$	Green/red ratio; chlorophyll absorption.

Notes: Sentinel-2 band designations: B2 blue, B3 green, B4 red, B5–B7 red edge 1–3, B8 near-infrared (NIR), B8A narrow NIR, B11 shortwave infrared 1 (SWIR1), B12 SWIR2. Sentinel-1 VV and VH denote co-polarized and cross-polarized backscatter.

Table A3: Classifier Hyperparameters: Search Space and Selected Values

Hyperparameter	Values considered	Selected
num_trees	50, 100, 150, 200	100
max_depth	10, 15, 20, 25, 30	25
min_examples	3, 6, 27	27
shrinkage	0.02, 0.05	0.02
l1_regularization	3.0, 5.0, 7.0	5.0
l2_regularization	3.0, 5.0, 7.0	3.0
num_candidate_attributes_ratio	0.1, 0.2, 0.3	0.1
split_axis	AXIS_ALIGNED, SPARSE_OBLIQUE	AXIS_ALIGNED
growing_strategy	LOCAL, BEST_FIRST_GLOBAL	LOCAL
forest_extraction	DART	DART

Notes: Hyperparameters of the gradient-boosted-trees classifier (TensorFlow Decision Forests), selected by random search (30 trials per fold) within each of three group k-fold cross-validation splits, with training sites grouped by cluster and validation AUC as the objective. Conditional parameters attached to non-selected options (`max_num_nodes` under BEST_FIRST_GLOBAL; the `sparse_oblique_*` settings under SPARSE_OBLIQUE) are omitted. `forest_extraction` was fixed at DART rather than tuned. Remaining parameters were left at library defaults.

Table A4: Correlates of Model Prediction Bias

	Dependent variable: prediction error ($\nu_i \times 100$)			
	Year	World region	Log GDP	Log pop. density
	(1)	(2)	(3)	(4)
Year 2018	−0.0012*			
	(0.0006)			
Year 2019	−0.0010*			
	(0.0006)			
Year 2020	−0.0007			
	(0.0005)			
Year 2021	−0.0006			
	(0.0005)			
Europe & Central Asia		0.0036**		
		(0.0018)		
Latin America & Caribbean		0.0013		
		(0.0028)		
Middle East & North Africa		−0.0046		
		(0.0082)		
North America		0.0003		
		(0.0036)		
South Asia		−0.0035		
		(0.0052)		
Sub-Saharan Africa		0.0040**		
		(0.0018)		
Log GDP per capita			−0.0040	
			(0.0066)	
Log GDP per capita^2			0.0002	
			(0.0004)	
Log population density				−0.0025***
				(0.0008)
Observations	68,226	68,226	68,226	68,226

Notes: Each column shows a separate regression of prediction errors ($p_i - D_i$) x 100 on the indicated covariates. Observations are pixel × year, 2017–2021. Regressions are weighted using inverse probability sampling weights. Omitted categories are 2017 in Column 1 and East Asia & Pacific in Column 2. Standard errors are clustered by tile. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table A5: Reduced-Form and First-Stage Estimates (Pre/Post Ban)

Outcome	Reduced Form Dump area (km^2) (1)	First Stage Δ Net Imp. (kt) $\times$ Post (2)
Panel A: Country-Level		
Predicted Δ (kt)	0.089*	0.801***
$\times$ Post-2016	(0.048)	(0.126)
	[0.068]	[0.000]
First-stage F		40.3
Observations	1,508	1,508
Countries	116	116
Panel B: Pixel-Level		
Predicted Δ (kt)	0.193	1.079***
$\times$ Post-2016	(0.146)	(0.273)
	[0.185]	[0.000]
First-stage F		15.6
Observations	159,744	159,744
Countries	116	116

Notes: Table shows the two components of the IV in Table 1. Reduced form (column 1): dump area (km^2) on predicted diversion $\times$ post — coefficients are m^2 per kg of predicted diversion (numerically identical to km^2 per kt). First stage (column 2): realized change in net imports $\times$ post (kt) on predicted diversion $\times$ post; F is the cluster-robust first-stage F. Predicted diversion = pre-ban bilateral trade shares $\times$ China's import decline (per-HS-code shift-share); post = year $>$ 2016. Panel A: country $\times$ year panel, outcome = verified dump area, country + year FE, unweighted. Panel B: verified-pixel $\times$ year panel, outcome = pixel dump status $\times$ country land area, tile + year FE, weighted by inverse sampling probabilities. SEs clustered by country in parentheses, p-values in brackets; China excluded. *p$<$ 0.1; **p$<$ 0.05; ***p$<$ 0.01.

Table A6: Shift-Share Exposure and Pre-Ban Country Characteristics

	Standardized exposure	
	Bivariate	Joint
	(1)	(2)
Pre-ban dump-area trend (km^2/yr)	0.079	-0.368*
	(0.092)	(0.193)
Pre-ban dump area (km^2)	0.463	0.603**
	(0.308)	(0.282)
Pre-ban net imports (kt/yr)	0.088	0.091
	(0.327)	(0.159)
Log GDP per capita	0.173**	0.230***
	(0.077)	(0.075)
Log population density	0.266**	0.433***
	(0.121)	(0.107)
Log land area	0.171*	0.350***
	(0.102)	(0.097)
Joint Wald p		0.000
Countries	116	116

Notes: Cross-country regressions of the shift-share exposure (predicted diversion, standardized) on pre-ban country characteristics (standardized): one at a time (column 1, the correlation) and jointly (column 2). Pre-ban = 2011–2016. Dump-area trend = per-country OLS slope of verified dump area on year over the pre-ban years. Correlation with pre-ban levels is absorbed by the country fixed effects in the main specifications; the identifying assumption concerns the trend row. Heteroskedasticity-robust SEs in parentheses. *p< 0.1; **p< 0.05; ***p< 0.01.

Table A7: Rotemberg Weights: Which Origins Drive the Shift-Share IV

Origin	HS code	$\hat{g}_k$ (kt/yr)	$\hat{\alpha}_k$	$\hat{\beta}_k$
USA	470790	1945.3	0.338	0.134
HKG	3915	1636.2	0.115	-0.021
CAN	470790	411.1	0.096	0.335
JPN	470790	1147.7	0.094	-0.027
GBR	470790	1054.0	0.076	0.065
AUS	470790	215.4	0.031	0.034
MEX	3915	156.2	0.028	0.366
NLD	470790	236.0	0.025	-0.004
USA	3915	775.8	0.024	0.083
CAN	3915	161.3	0.024	0.343
Cells (origin × code)				186
Sum of positive weights				1.006
Sum of negative weights (share of cells)				-0.006 (0.167)
Top-5 share of positive weights				0.714

Notes: Rotemberg-weight decomposition of the country-level IV (Goldsmith-Pinkham, Sorkin and Swift 2020): $\hat{\beta}_{IV} = \sum_k \hat{\alpha}_k \hat{\beta}_k$ over instrument cells k = (origin × HS code), where $\hat{g}_k$ is the cell's pre-ban exports to China, $\hat{\alpha}_k$ its weight, and $\hat{\beta}_k$ the just-identified IV using only that cell's destination shares. Weights sum to one; the 10 largest in absolute value shown. The pixel-level IV uses the same country-level instrument.

Table A8: Case Study: Near–Far DiD Across Specifications — Java, Indonesia

Ring (km)	Monthly, cell FE	Monthly, cell × moy FE	Annual, two-period	Border-trimmed	Drop 2020
0-2	+6.15	+6.09	+5.97	+6.32	+4.43
	[+3.48, +8.88]	[+3.28, +8.99]	[+3.19, +8.84]	[+3.56, +9.16]	[+2.33, +6.57]
2-5	+5.15	+5.10	+5.03	+5.34	+3.68
	[+2.68, +7.68]	[+2.49, +7.78]	[+2.41, +7.72]	[+2.77, +7.97]	[+1.72, +5.68]
5-10	+4.20	+4.13	+4.13	+4.39	+2.94
	[+1.87, +6.57]	[+1.67, +6.66]	[+1.61, +6.70]	[+1.96, +6.87]	[+1.09, +4.82]
10-20	+2.61	+2.49	+2.49	+2.79	+1.62
	[+0.31, +4.96]	[+0.06, +4.98]	[+0.03, +5.01]	[+0.37, +5.26]	[-0.24, +3.50]
20-30	+0.61	+0.52	+0.60	+0.74	-0.23
	[-1.20, +2.45]	[-1.41, +2.48]	[-1.33, +2.57]	[-1.19, +2.71]	[-1.83, +1.39]
Cadence	monthly	monthly	annual	monthly	monthly
Cell FE	cell	cell × moy	cell	cell	cell
Cells	106,633	106,633	106,633	104,597	106,633
Clusters (adm-2)	125	125	125	124	125

Notes: Difference-in-differences estimates of the effect of China's waste import ban on $PM_{2.5}$ by distance ring in the post-enforcement period, relative to the pre-announcement period, across five specifications. Effects are in percent, $100(e^{\beta} - 1)$, relative to the omitted 30–40km ring, with 95% confidence intervals in brackets (standard errors clustered by district). Monthly specifications include cell and district-by-year-month fixed effects; the second column replaces the cell fixed effect with cell-by-calendar-month fixed effect. The annual specification collapses the panel to cell-by-year with cell and district-by-year fixed effects. Border-trimmed drops cells that are close to region boundaries that might be contaminated by out-of-region dumps. Drop 2020 excludes calendar year 2020 from the post-enforcement period.

Table A9: Case Study: Near–Far DiD Across Specifications — California

Ring (km)	Monthly, cell FE	Monthly, cell × moy FE	Annual, two-period	Border-trimmed	Drop 2020
0-2	-3.17	-3.13	-4.78	-3.31	-2.22
	[-4.74, -1.58]	[-4.75, -1.48]	[-7.12, -2.38]	[-4.89, -1.71]	[-3.69, -0.73]
2-5	-3.50	-3.46	-5.05	-3.64	-2.62
	[-4.94, -2.03]	[-4.96, -1.94]	[-7.33, -2.71]	[-5.09, -2.17]	[-3.94, -1.27]
5-10	-3.28	-3.25	-4.80	-3.44	-2.43
	[-4.71, -1.82]	[-4.73, -1.74]	[-7.11, -2.43]	[-4.87, -1.99]	[-3.71, -1.13]
10-20	-2.70	-2.66	-4.07	-2.87	-2.12
	[-3.98, -1.39]	[-4.00, -1.31]	[-6.13, -1.97]	[-4.15, -1.56]	[-3.28, -0.95]
20-30	-1.43	-1.42	-2.42	-1.56	-1.17
	[-2.31, -0.54]	[-2.33, -0.49]	[-3.84, -0.98]	[-2.45, -0.66]	[-1.97, -0.36]
Cadence	monthly	monthly	annual	monthly	monthly
Cell FE	cell	cell × moy	cell	cell	cell
Cells	203,158	203,158	203,158	198,128	203,158
Clusters (adm-2)	56	56	56	55	56

Notes: Same specifications as Table A8, for California.