CESifo WORKING PAPERS

10820
2023

December 2023

An Evaluation of Protected Area Policies in the European Union

Tristan Earle Grupp, Prakash Mishra, Mathias Reynaert, Arthur A. van Benthem

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10820*

An Evaluation of Protected Area Policies in the European Union

Abstract

The European Union designates 26% of its landmass as a protected area, limiting economic development to favor biodiversity. This paper uses the staggered introduction of protected-area policies between 1985 and 2020 to study the selection of land for protection and the causal effect of protection on vegetation cover and nightlights. Our results reveal protection did not affect the outcomes in any meaningful way across four decades, all countries, protection cohorts, and a wide range of land and climate attributes. We conclude that European conservation efforts lack ambition because policymakers select land for protection not threatened by development.

JEL-Codes: Q230, Q240, Q570, R140.

Keywords: land protection, conservation, biodiversity, deforestation, vegetation cover, nightlights, staggered difference-in-differences.

Tristan Earle Grupp
Stuart Weitzman School of Design
University of Pennsylvania
Philadelphia / PA / USA
tgrupp@design.upenn.edu

Prakash Mishra
Wharton School
University of Pennsylvania
Philadelphia / PA / USA
mishrap@wharton.upenn.edu

Mathias Reynaert
Toulouse School of Economics / France
mathias.reynaert@tse-fr.eu

Arthur A. van Benthem
Wharton School, University of Pennsylvania
Philadelphia / PA / USA
arthurv@wharton.upenn.edu

November 29, 2023
We thank Adam Streff and Danial Syed for excellent research assistance. Thanks to seminar participants at the AERE 2023 Summer Conference, EAERE 2013 Summer Conference, Georgia Tech, Imperial College London, LSE/Imperial/King's Workshop in Environmental Economics, Toulouse School of Economics, University of Bristol, University of California at Berkeley, University of Gothenburg, University of Mannheim, University of Pennsylvania, and the University of Wisconsin-Madison. We thank Robin Burgess, Eli Fenichel, Alex Pfaff, Andrew Plantinga, Santiago Saavedra, Ulrich Wagner, and Matthew Wibbenmeyer for helpful comments and suggestions. We gratefully acknowledge financial support from FORMAS grant number 2020-00371. Reynaert acknowledges funding by the European Union (ERC, SPACETIME, grant n° 101077168). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Reynaert acknowledges funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). van Benthem thanks Penn Global, the Kleinman Center for Energy Policy, the Mack Institute, and Analytics and Wharton for generous support.

1 Introduction

How effective are protected area policies at restoring vegetation cover, constraining economic activity, and improving biodiversity more broadly? These questions are central to assessing the pledge of 196 countries to protect 30% of the earth's land and waters by 2030. This '30x30 target', which was the main outcome of the COP 15 Kunming-Montreal global biodiversity conference and sometimes referred to as the 'Paris Agreement for Nature', was ratified in December 2022 (Einhorn 2022). How ambitious is this 30% target? This depends strongly on how policymakers select land to protect: Does protection occur on at-risk land with a high economic opportunity cost, or are sites chosen in areas with little risk of economic development?

We study these questions with a focus on the European Union (EU), the region that is closest to reaching the target. As of 2023, the EU has protected over 26% of its total landmass (Eurostat 2022). The EU's flagship land protection program, *Natura 2000*, is the largest coordinated land protection regime in the world (Oceana Europe 2022), and might offer insights into the effectiveness of the 30x30 target in other regions. Europe has greened substantially over the last 100 years. Forests have expanded by more than 30% since 1900, an area the size of Portugal, and by 10% since 1990 (Eurostat 2021). We ask: how much of the EU's greening did its land protection policy contribute?

Specifically, we evaluate the causal effect of Europe's protected area policies on an important dimension of biodiversity—vegetation cover—and on economic activity—measured by nightlights. We make three primary contributions. First, we provide unique estimates of the long-term impact of the world's largest land protection policy over a time span of four decades. Second, we use recent econometric advances that yield unbiased estimates in the presence of time-varying site selection in a staggered policy context. Because the EU has protected more than 100,000 areas over the course of many decades, we can estimate how the effect of protection varies across countries, across time since protection, and across earlier and later protected areas. We also estimate treatment-effect heterogeneity with unprecedented granularity across observable land and climate characteristics. Third, we provide evidence on the environmental impacts of land protection in advanced economies, which has been virtually non-existent to date.

We assemble a high-resolution remote sensing dataset that spans the entirety of the European Union from 1985. Our data include key outcome variables (vegetation cover, nightlights) at the 300x300 meter or 1x1 km level, treatment variables (location of protected areas and date of first protection), and a wide range of control variables that measure climate, weather, and land characteristics. Our continuous measure of vegetation greenness is of intrinsic interest as it reflects gradual changes in vegetation cover not captured by discrete land-use measures, and is also an imperfect yet reasonable indicator of other measures of biodiversity. We also collect alternative outcomes—discrete land use classes and species counts—for use in robustness analysis.

The gradual implementation of land protection in Europe allows us to employ a staggered

difference-in-differences design. Because of selection into treatment, treated and untreated land differ on several dimensions. Moreover, time-varying selection can cause treatment effects to vary by cohort (the calendar year in which land first gets protected) and event time (years relative to treatment). Typical two-way fixed effects estimators are biased in such settings. To overcome this challenge, we apply the doubly-robust estimator of Callaway and Sant'Anna (2021). This estimator combines cohort-specific inverse probability matching with an outcome regression adjustment to compare protected areas with observably similar unprotected land and control for time-varying differences in observable plot attributes.

In addition to time variation in treatment effects, previous literature on tropical forests—described below—has revealed that the effect of land protection is highly context-specific. This underscores the importance of testing for treatment-effect heterogeneity along many dimensions of our covariate space, such as land and climate attributes, and the degree of local economic pressures from agriculture and forestry. Using our expansive data, we use the recent non-parametric causal random forest method of Wager and Athey (2018) to estimate highly-granular conditional average treatment effects (CATEs).

The results are sobering. First, the Europe-wide ATE—aggregated across countries, cohorts, and event time—is statistically and economically close to zero. In none of the EU member states do we find evidence for a meaningful causal contribution of protection to vegetation cover. Second, up to three decades after treatment, event-time treatment effects indicate a zero effect of protection on vegetation cover. Third, we find no trend in cohort-level ATEs. Land protected later in time does not contribute more to vegetation cover than land protected early in our sample. Fourth, these three findings are identical for nightlight measures. Fifth, we find the treatment effects are not meaningfully heterogeneous across the covariate space; the zero effect is stable and pervasive across a wide range of CATEs. Sixth, we also show that our conclusions hold up when using discrete land use data. Neither does the limited species count data suggest a relationship between protection and biodiversity.

We argue that site selection likely explains the lack of effectiveness of the EU protection policy. Our results suggest EU policymakers successfully identify land that has never been threatened or has recently been greening. While protection might be effective against future development threats, our analysis has uncovered no evidence of such threats in areas protected for several decades. We also find no evidence for improved effectiveness in countries with a history of strong enforcement. Instead, our findings are consistent with policymakers selecting sites with little economic opportunity cost, perhaps motivated by either the private benefit of obtaining "easy" green credentials, or the focus on area-based targets such as the 30x30 Kunming-Montreal agreement, which may provide incentives to find the "cheapest" land and protect it first (Maxwell et al. 2020). Overall, we conclude that Europe's protected area policy lacks ambition. While current protection may safeguard against very long-run economic development pressures, there is a significant opportunity to focus new protection on currently-at-risk land where ecological benefits outweigh economic costs.

There is a large literature on protected-area policies in tropical forests in developing countries (e.g., Sims 2010; Pfaff et al. 2015; Souza-Rodrigues 2019; Assunção et al. 2022; Cheng, Sims, and Yi 2023). Evidence in advanced economies is lacking; Auffhammer et al. (2021)'s study of protection on land-market impacts in the U.S. is a rare exception. The literature has generally found a modest impact on forest cover, but estimates are highly context-specific. Stronger effects are observed in well-enforced areas experiencing economic development pressure (Börner et al. 2020; Assunção, Gandour, and Rocha 2023; Reynaert, Souza-Rodrigues, and Benthem 2023). Most studies estimate the effects of land protection over fairly short time frames of a decade or less (e.g., Sims and Alix-Garcia 2017; Keles, Pfaff, and Mascia 2022); our study spans four decades. Non-random site selection for protected areas has been a central concern in the literature. Many prior studies address this through a matching design (e.g., Andam et al. 2008; Joppa and Pfaff 2010; Geldmann et al. 2019; Maxwell et al. 2020); we build on this by applying recent econometric advances critical for mitigating bias in a staggered-adoption setting.

2 Policy

The Natura 2000 policy (2009/147/EC) combines two earlier EU directives: the habitat directive (92/43/EEC) and the birds directive (79/409/EEC). It requires countries to submit a standardized report on their protected areas to the European Commission, following International Union for Conservation of Nature (IUCN) guidelines. Europe's green deal and nature restoration law aim to protect 30% of its land by 2030 (A9-0220/2023/EP). The Commission oversees and may amend member states' proposals, while the member states handle the siting and local enforcement of the protected areas. The directive covers municipal, regional, and national protected areas. Some areas restrict all or most human activity (e.g., strict nature reserves and national parks; 7.6% of protected landmass), while others allow some industrial and agricultural activities (e.g., habitat or species management areas; 47% of protected landmass). We discuss the breakdown of these categories in Appendix C.1. Member states have different policies regarding protection and land ownership.[1]

Our analysis studies protected areas under the early EU directives or member state policies before 2009 and the complete EU Natura 2000 program after 2009. In 2020, the EU released an evaluation of the Natura 2000 network (European Environment Agency 2020). The report describes the difficulty of such an evaluation: "Measuring the ecological effectiveness of a network of protected areas is difficult, as baseline data are scarce and the data have many limitations, such as the lack of data enabling comparison of the conservation status of and trends in species and habitats inside and outside of the Natura 2000 network." Our study aims to provide large-scale, long-term causal evidence for the effectiveness of the EU's land protection policy.[2]

Site selection of protected areas is a crucial responsibility of EU member states. A social

1. Unfortunately, ownership data at the EU scale are not available to us.

2. We use the term EU, but the data includes the 27 member states and Albania, Bosnia, Montenegro, Macedonia, Switzerland, and Serbia. We exclude Malta, Iceland, and Liechtenstein due to missing data issues.

planner's site selection would prioritize protecting plots with the best balance between ecological benefits and economic costs (including the opportunity costs of foregone economic activity and the administration, monitoring, and enforcement costs). For protection to be meaningful, it must change economic production in the area relative to the status quo and foster natural vegetation cover and species survival. Protecting land that is never under any economic pressure would be considered a wasteful expenditure of resources; land at risk of future economic development may warrant protection. In reality, EU member states' politicians may prioritize protection differently. They might emphasize current local economic production, undervalue future biodiversity benefits, or benefit politically from the "green glow" of protecting *any* type of land, regardless of the ecological gains. In such cases, protection will occur first on land with the lowest economic opportunity cost, typically land not at risk of development. Different policy priorities will change the order in which land is protected and the resulting time path of treatment effects.[3]

3 Data

We collect six types of data to assemble two remote-sensing datasets spanning the entirety of the European Union between 1985-2019. The most granular dataset (to analyze vegetation cover) divides Europe into 117 million equal-sized grids of 300 by 300 meters. The second dataset (to analyze nightlights) has grids of one square kilometer. Our data delivers comparable and consistent measures across space and time (see Appendix A for details):

Policy rollout. For every protected area in the EU, we assign the date of its initial protection based on data from the Common Database on Designated Areas (CDDA), consolidating land protection policies across 39 European countries. Focusing solely on terrestrial protection and excluding marine reserves, our dataset includes details on 118,511 distinct areas that were protected between 1800 and 2019. We establish a grid cell as protected if any non-zero fraction of its land area falls under a conservation agreement.

Vegetation cover. We aggregate satellite images from the Landsat 5, 7, and 8 data to construct a continuous normalized difference vegetation index (NDVI) at a bi-annual frequency, with higher values on a 0-100 scale indicating denser and richer vegetation.[4] The remote sensing measures start in 1985. We use biennial aggregation to reduce missing data problems caused by cloud coverage and focus on the summer months when perennial vegetation is most visible. The continuous measure avoids classification errors that plague categorical land use classifiers (see Torchiana et al. 2020;

3. Conceptually, suppose a limited number of grids can be protected, and that protection is non-retractable. In that case, the problem of optimal selection is isomorphic to one of resource extraction. The planner's solution to a dynamic extraction program is to protect land with the highest benefit-cost ratios first. If instead a policymaker only cares about ecology and not the economic costs of protection, land with the highest ecological gains from protection will be protected first, resulting in large treatment effects. In contrast, if a policymaker cares only about the economic costs, we expect a policy that starts with land that is costless to protect, resulting in near-zero treatment effects.

4. We rescale NDVI indices to be between 0-100 instead of the [-1,1] range standard in the remote-sensing literature. We focus on the 0-1 range and scale the index to 0-100. We drop observations with NDVI less than 0, as this range corresponds to snow, water, and clouds.

Alix-Garcia and Millimet 2022). It also allows us to capture gradual changes in vegetation after protection that do not necessarily lead to land use re-categorization. We use NDVI because we can construct the measure with early Landsat data, allowing us to obtain a long panel. Recent measures such as Vegetation Cover Fields (VCF) are only available since 2000.[5] Furthermore, NDVI correlates equally well as other vegetation measures with the frequently-used biodiversity marker bird species richness (Nieto, Flombaum, and Garbulsky 2015; Hobi et al. 2017).[6] We limit the sample to plots with an NDVI value of at least 40 at one occurrence. This excludes urban grids, bare soil, and rocky landscapes.[7]

Figure 1 maps the levels and changes in NDVI in Europe. Panel (a) shows that NDVI indices range between 60 and 100 in many areas. Panel (b) plots changes in the NDVI index between 1985 and 2019. We document decreasing NDVI in red and increasing NDVI in green. The map reveals a substantial greening of Europe, most pronounced in Southeastern Europe and Scandinavia. NDVI decreases in the east of France, the west of Germany, and the Baltics. Panels (c) and (d) replicate panels (a) and (b) but only for areas that had received protection by 2019. Protected areas are very green in 2019; most of them experienced increases in NDVI between 1985-2019. The changes in greenness over time, together with the staggered implementation of protection, is the primary identifying variation we exploit in this paper.

Nightlights. We rely on Li et al. (2020) for a 1992-2018 one square kilometer panel of remotely-sensed nightlights. Nightlights have been used as a proxy for economic development and GDP in remote areas (Donaldson and Storeygard 2016) and urban/settled areas alike (Gibson et al. 2021). Here, our goal is to measure human presence on a granular scale. If protected areas limit economic activity, we expect outward migration from the area, which could reduce nightlights.

HILDA. We obtain discrete land-use data via the Historical Land Dynamics Assessment project, or HILDA, dating back to 1900 at a decadal frequency and a resolution of one square kilometer (Fuchs et al. 2015). HILDA classifies each grid as settlement, cropland, forest, grassland, other land, or water. The data is constructed by harmonizing historical land cover information such as national inventories, maps, and aerial photographs with remote-sensing data. HILDA allows us to investigate long-term trends in EU land use. We are mainly interested in forests, grassland (which includes pastures), and cropland. Appendix Figure A.11 shows land-use shares between 1900 and 2010, and Appendix Table A.15 reports the EU's land-use transition matrix for that period; both confirm the EU has greened substantially.

Species counts. We use the BioTIME dataset, which is the largest available aggregation of species count studies across space and time, see also Liang, Rudik, and Zou 2023. Species count

5. The Enhanced Vegetation Index (EVI) targets measurement of tropical forests not present in the EU.

6. The finding that NDVI is an excellent predictor for bird diversity is replicated in our study area for France (Bonthoux et al. 2018) and Mediterranean landscapes (Ribeiro et al. 2019).

7. Using discrete land-use data from HILDA (see below), we verify that forests, grassland and cropland typically have a (much) higher NDVI index than 40. For example, in 2010, French forests have an average NDVI of 75, grasslands 70, and croplands 50. See Appendix D for details and evidence that our results are robust to the chosen NDVI threshold.

data allow for more direct measures of biodiversity. However, many count data suffer from one or all of these issues: nonrandom location of counts, short panels with recent coverage only, a low number of species, model-based projections across space, and limited regional coverage. Each of these issues is problematic for the goal of our study: a comprehensive long-term EU-wide evaluation of protection. None of the species count data sources, including BioTIME, allow us to achieve a similar causal research design as with remote-sensing data.[8]

Control variables. Finally, we add data on bio-geographical regions from the European Environment Agency; climate zones, soil properties, and topography from European Soil Data Centre; precipitation from the European Centre for Medium-Range Weather Forecasts; and solar radiation from WorldClim. We merge this rich set of controls with the NDVI and nightlight data to control for each grid's natural vegetation growth propensity.

4 Methodology

The unit of observation is a grid, $i \in \{1, 2, \ldots, I\}$. We observe every grid i at a biennial frequency. Define periods $t \in \mathbb{T}$, where $\mathbb{T} = [1985, 2019]$ is the sample of years we have data for. Treated areas correspond to areas that are protected at some time $g(i) \in \mathbb{T}$, and control areas are assigned $g(i) = \infty$ to indicate that they are not treated before $\mathbb{T}$.[9] Define G as the group of all units treated at time g. We use subscript t for calendar time 1985 to 2019 and we define event time $e \equiv t - g$ as the number of years before or after treatment. We define the treatment indicator $D_{it} \in \{0, 1\}$. The treatment indicator equals one after plot i is protected: $D_{it} = \mathbb{1}[t \geq g(i)]$.

We aim to estimate the treatment effect of protection on the plot-level outcomes of vegetation cover and nightlights. We define the treatment effect of protection at time g on an outcome variable Y_{it} as:

$$\tau_{igt} = \mathbb{E}[Y_{i,t} - Y_{i,t-1}|G, D_{it} = 1] - \mathbb{E}[Y_{i,t} - Y_{i,t-1}|G, D_{it} = 0], \tag{1}$$

comparing the difference in outcomes of grid i in treatment cohort G with an unobserved difference in counterfactual outcomes should grid i not have been treated. We hypothesize that a protected area policy should have a positive treatment effect on vegetation cover and a negative treatment effect on nightlights.

8. Other species count data include Global Biodiversity Information Facility, Movebank, the PanEuropean Common Bird Monitoring Scheme, eBird, and the European Bird Census Council. See https://www.gbif.org/, https://www.movebank.org/cms/movebank-main, https://pecbms.info/, https://ebird.org/home, and https://www.ebcc.info/.

9. Because only 20% of land is treated at the end of our sample we select control units from the remaining 80% of grids. These are the 'never-treated units' in our sample.

4.1 Staggered difference-in-differences design

We apply the doubly-robust difference-in-differences estimator of Callaway and Sant'Anna (2021) to obtain country, event-time, and cohort-specific average treatment effect estimates θ_{cgt} of (1).[10] We estimate average treatment effects at the country level because EU member states have jurisdiction over the policy and differ in their site selection and enforcement. In addition, we might expect treatment effect heterogeneity across event time e. For example, when treated plots regenerate after protection, we expect vegetation cover to increase gradually. Finally, there is reason to expect cohort-specific treatment effects whenever site selection differs across cohorts g when more and more land gets protected, and the policymaker selects land with increasing opportunity costs of protection.

The doubly-robust estimator is a crucial improvement in our setting relative to standard two-way fixed effects methods for two reasons. First, we observe a staggered introduction of the European policy over a thirty-five-year window from 1985 to 2019, where time-varying site selection could cause cohort-specific treatment effects. In such settings, standard two-way fixed effects estimation is biased. To obtain unbiased estimates, the doubly-robust estimator allows for cohort-specific matching probabilities, outcome correction models, and treatment effects.[11] Cohort-specific matching addresses time-varying selection bias stemming from the correlation between unobserved plot-level economic activity and treatment status. The outcome correction controls for time-varying differences in observable weather-related plot attributes. This is essential because of the weather-induced variability in NDVI. Second, we can estimate specific dynamics for each cohort to explore the policy's potential heterogeneity across years since treatment and across cohorts.

We specify a model for the matching procedure and the outcome correction based on variables which drive vegetation growth. We rely on previous literature assessing land conditions and yields (Schlenker and Roberts 2009) to select yield-relevant variables for the inverse probability matching. Fixed land factors include a measure of soil suitability for agriculture, elevation, slope angle, a long-term average of rainfall, and solar radiance. Time-varying matching variables are rainfall, heating-degree days, and the length of the growing season. Additionally, we include an average greenness measure over the first three in-sample two-year (pre-treatment) periods to limit level differences between treatment and control areas. We also include the variance in rainfall and heating degree days over the three pre-periods. To test for parallel trends, we select plots treated at least three two-year periods after the first year of our outcome measures so that the first-treated cohort is in 1991.

The outcome regression adjustment linearly projects control units' change in outcomes between

10. We use a difference-in-differences approach instead of a regression discontinuity design. Many protected areas border grids with different land uses, causing discrete jumps in NDVI around protected areas.

11. TWFE with cohort-interacted treatment effects introduces bias because whenever there is site selection, the control group contains not-yet-treated units that differ on observables from the treated units. The estimation procedure of Callaway and Sant'Anna (2021) addresses this by cohort-specific matching. For completeness, we also report results from standard two-way fixed effects estimation.

$g-1$ and t on plausibly exogenous observables, separately for every cohort group G:

$$m_{gt}(X) = \mathbb{E}[Y_{igt} - Y_{i,g-1}|G, X, D_{it} = 0].$$

By subtracting the fitted values of $m_{gt}(X)$ from the outcome difference of treated units, the procedure corrects for the confounding effects of time-varying observable differences in control variables.

The procedure involves estimating the propensity scores, outcome correction, and dynamic average treatment effect by country-cohort. We estimate standard errors using the multiplier bootstrap procedure. Because multiple plots are assigned to treatment simultaneously, we cluster at the assignment level of the CDDA identifier (Bertrand, Duflo, and Mullainathan 2004) in addition to the default plot-level clustering used by the Callaway and Sant'anna procedure. We provide details on the estimation in Appendix B.

We aggregate the estimated treatment effects θ_{cgt} across different dimensions. First, we aggregate across countries, cohorts, and event time to construct a single treatment effect estimate across the European Union, θ_{EU}. Next, we aggregate across cohorts and event time to obtain a country-specific θ_c. Third, we aggregate across countries and cohorts to obtain θ_e to evaluate how treatment effects evolve dynamically in the years following protection. Finally, we aggregate across countries and event time to obtain cohort-country level estimates θ_g. Appendix B provides details on the computation of these aggregated average treatment effects.

4.2 Conditional average treatment estimator

In addition to differences in treatment effects across countries, event times, and protection years, we anticipate treatment effect heterogeneity due to factors such as varying soil, pre-protection greenness, or weather conditions impacting vegetation regeneration ability (e.g., previous agricultural land will green differently than land that was already forested). We use the conditional average treatment effect framework put forward by Chernozhukov et al. (2017) and test directly whether there is a statistically significant deviation from the average treatment effect across the covariate space.

While we could introduce additional heterogeneous treatment effects in the Callaway and Sant'Anna framework, we have a large set of covariates to consider. We therefore apply the honest random forest estimator of Wager and Athey 2018 to select statistically-relevant dimensions of heterogeneity, avoid multiple testing problems, and return a causally-valid conditional average treatment effect (CATE). Both Callaway and Sant'Anna (2021) and Wager and Athey (2018) are doubly-robust difference-in-differences methods as we apply them: the first is semiparametric and the second nonparametric. In contrast to current literature, where third differences are chosen by the researcher, a nonparametric method has the advantage of choosing salient dimensions of heterogeneity directly from the data. Thus, the data informs how to construct $\tau_x = \mathbb{E}[\tau|X = x]$. The estimator first constructs plot-level treatment effects by comparing treated units to control plots

with a similar propensity score. It uses a random forest as a nonparametric regression: it projects the plot-level treatment effects on a (potentially large) set of explanatory variables. The random forest finds maximally informative "splits" of the data by drawing separating hyperplanes in the covariate space. These splits determine conditional average effects as the maximal deviations from average treatment effects explained by covariates. This is relevant in estimating what variables make protection policy successful at fostering vegetation growth. We describe estimation details in Appendix B.

5 Results

5.1 Average treatment effects on NDVI by country

Figure 2 presents the EU-wide average treatment effect θ_{EU} on vegetation greenness aggregated across all countries, event times, and cohorts, along with the country-specific ATEs θ_c ordered from small to large (red and blue bars). Detailed results for Figure 2 and all subsequent figures are presented in Appendix C. The top panel summarizes θ_{EU} across different econometric specifications. The preferred EU-wide ATE via the doubly-robust Callaway and Sant'anna estimator is 0.08 with a standard error of 0.01 for the change in the NDVI index that ranges between 0 and 100.[12] In economic terms, this is a highly-precise zero effect of protection. The figure also displays the estimate from two naive, and biased, two-way fixed effects estimators (with and without matching).[13] Their EU-wide ATE of 0.5 is similarly small. In Appendix D, we conduct several robustness checks. By looking at the first differences in greenness as an outcome, we confirm the null result is not driven by a functional-form assumption. We also show results are unchanged when we lower the NDVI sample selection threshold from 40 to 30. Finally, we address spillovers between treated and control units in a spatial first-difference specification.

There are some differences across countries (bottom panel), but the size of all estimated treatment effects in absolute terms is still very small—θ_c varies between -2 and 2, and is not even statistically different from zero in many countries. Changes of less than 1 point in NDVI index are negligible relative to the large improvements in the index of 6-15 points in many areas of the EU between 1985 and 2019, as shown in Figure 1.

5.2 Average treatment effects on NDVI across event time and cohorts

Figure 3 (left panel) shows the ATE estimates θ_e^{EU} and associated 95% confidence intervals—aggregated across the entirety of the European Union—over event time e, with $e = 0$ indicating when protected areas were established. These dynamics reveal no post-protection upward trend in regeneration or protection benefits for treated land units over time. The effect of conservation on

12. Appendix C.2 shows balance tables with and without cohort-specific matching.

13. For the two-way fixed effects regression, we estimate a difference-in-differences model with grid- and time- fixed effects, λ_i and λ_t: $Y_{it} = \beta D_{it} + \lambda_i + \lambda_t + \epsilon_{it}$.

greenness is consistently close to zero—all confidence intervals are contained within $[-1, 1]$. Even thirty years after treatment, we find no evidence of a protection effect. The lack of a pre-trend is reassuring and suggests the matched controls are appropriate comparisons for the treated areas. This does not imply that vegetation greenness has been constant over time. Europe has been greening, but treatment and control plots have been greening in parallel, both before and after the establishment of land protection policies.

Even if the average treatment effect is zero, we investigate if there is a time trend in the treatment effects θ_g^{EU} by cohort of protection g. As discussed in Section 2, one might expect that land with low opportunity costs is protected first (small treatment effect). As time progresses and the land with the lowest opportunity cost has been protected, governments might focus on areas with higher opportunity costs (larger treatment effect). Figure 3 (right panel) shows the EU-wide cohort-level treatment-effect estimates. There is no time trend in the treatment effects across cohorts—the effect of land protection is close to zero regardless of when the land got protected.

To formally establish this result, we compute two test statistics applied at the country level in Appendix Table A.11. First, we test if there are statistically-significant linear trends in the country-cohort level estimates θ_{cg}. Second, for each country, we split the estimate θ_{cg} in an early-treated and late-treated group and test for a difference in the treatment effect size between the two groups. With these two tests, we find no statistical evidence for any meaningful positive trends across cohorts of protection at the individual country level.

5.3 Other outcomes

Having established a precise zero effect of the effect of land protection on NDVI across countries, event times, and cohorts, we now present evidence on the effect of protection on nightlights. One might hypothesize that protection reduces human activity relative to matched controls, which will manifest itself through decreased brightness of night-time light.

Figure 4 shows that—whether broken down by event time or by cohort—there is no evidence that land protection reduced nightlights. Nightlights vary from 0 to 68 units of luminosity, while treatment effects are generally within 0.5 units. While often statistically significant, the effects are not different from zero in an economically meaningful way. Our results suggest a zero impact on economic output in areas set aside for protection for at least two decades. In addition, the effect of land protection is close to zero regardless of the year of first protection. This corroborates our NDVI results.

It is conceivable that land protection increased biodiversity in ways neither measured by NDVI nor nightlights. The lack of comprehensive granular, consistent and reliable species count data does not enable us to test this hypothesis causally. Appendix E.1 contains some limited, non-causal, descriptive results using the BioTIME data. This event-study analysis shows no clear increase in species counts following CDDA establishments near the BioTIME study locations.

Finally, we confirm that our main results are consistent with evidence from discrete land-use

data. Despite the limitations of discrete classifications of land use (see Section 3), we use decadal data from HILDA in Table 1 to compare land-use transition probabilities between protected and never-protected areas. We limit the transitions to our study period, 1980 to 2010, to facilitate comparison with our previous empirical exercises. To focus on a consistent 30-year time window, we limit treated observations to areas protected in 1970-1980. Controls are matched on 1970 land-use and time-invariant observables: slope angle, slope steepness, solar radiation, long-run precipitation, and distance to a shoreline. The forest transition shares are very similar for protected and control units, which is consistent with our null results for continuous vegetation greenness. The largest discrepancy is a 5% larger transition of cropland into grassland in protected areas. However, the grassland-cropland classification is particularly difficult for discrete land-use measures and may be confounded by pasture.

5.4 Further treatment-effect heterogeneity

As discussed in Section 4.2, our null finding on average does not reject the existence of a (small) tail of high-impact conservation efforts. We explore whether there is meaningful underlying heterogeneity in treatment effects. We apply the Wager and Athey (2018) estimator on a random sample of the data (whose construction is discussed in Appendix B.3). The sample average treatment effect on NDVI is $-0.51(0.05)$.

Our data rejects the existence of any meaningful heterogeneity. While our random forest detects statistically-significant heterogeneity in the test of Chernozhukov et al. (2017), economically meaningful heterogeneity is limited. The CATE in the bottom quartile of the CATE distribution is only 0.7 (1.2) units smaller than in the top quartile. We precisely estimate limited heterogeneity around zero across the whole covariate distribution, not just for average land.

We emphasize our precise 0 persists along two key dimensions of heterogeneity. First, we find no significant variation in treatment effects when considering covariates linked to high-quality agricultural land. The left panel of Appendix Figure A.6 shows CATEs along the distribution of the soil phosphorus content, which is often fertilizer-driven in areas with intense agriculture. The findings show no notable difference in greening patterns between areas with different phosphorus content. Our null results are thus not driven by differences in NDVI changes between cropland and natural vegetation. Secondly, regions with lower initial greenness, like bare soil or sparsely-forested areas, may exhibit more greening potential post-protection than areas already green. In the right panel, we show that pre-period greenness has a CATE of -0.23 in the top quartile relative to a CATE of -0.75 in the bottom quartile: this is only one-half of an NDVI point difference on a scale of 0 to 100 and shows land with lower initial greenness gained less NDVI. The zero effect is thus not only there for preexisting forests but also for areas not yet forested at the beginning of our sample.

Regarding the interpretation of our results, this supports a precise zero throughout the distribution of protection. Other CATEs across measures of high-value farmland, distances to shorelines,

and population density, which we would expect to capture salient dimensions of land pressure, all trend less than those shown in Appendix Figure A.6. These results are consistent with limited evidence that protected land was under economic pressure.

5.5 Discussion of the results

There are several potential explanations for the zero results. First, protected plots are inframarginal, and hence, there is no change in land use before and after protection. Second, protected plots were under threat of *future* development rather than contemporaneous economic activity. Third, authorities do not enforce the restrictions of economic activity.

The second explanation seems implausible. Our long panel allows us to identify treatment effects up to 30 years after the first treatment in 1991. Even for the earliest cohorts, we do not see any positive treatment effects decades later. The third explanation is also unlikely. While enforcement of land use policy is often imperfect, such as in the Amazon rainforest (Keles, Pfaff, and Mascia 2022), this is less of a concern in the EU institutional context. If enforcement explained the small effects, treatment effects should vary between countries with varying degrees of institutional strength. However, our results show near-zero effects for all EU countries. We conclude that the first explanation is the most likely: Europe has not targeted land at risk of economic development.[14] Policymakers have selected land that would have seen equal increases in vegetation cover without protection.

6 Conclusions

Protecting a quarter of the EU's landmass has not led to a change in vegetation cover measured by NDVI or human activity measured by nightlights. We find no meaningful heterogeneity in treatment effects across event-time, protection cohorts, land, soil, or climate characteristics. Control plots show equal greening trends as treated plots. So far, Europe's greening has allowed land protection to expand with limited opportunity costs. For the most part, un-threatened areas have been protected, and land at risk of development may not have been. This—perhaps sobering—finding does not imply that protection may never restrict economic activity in the long run, but it suggests that Europe's current land-protection regime has not been well-targeted. One interpretation is that politicians have focused on protecting land with low economic development value. Yet a more optimistic interpretation is that there is still significant scope for expanding the protection of European land at a low economic cost. The administrative blueprint provided by the Natura 2000 network seems well-suited to form the basis for Europe's pursuit of such more ambitious biodiversity targets.

14. Appendix Figure A.12 shows that countries with a higher share of land considered nature (forest or grassland) protect a greater share of land. This association aligns with a story that inframarginal land drives current protection.

References

Alix-Garcia, Jennifer, and Daniel Millimet. 2022. "Remotely Incorrect? Accounting for Non-classical Measurement Error in Satellite Data on Deforestation." *Journal of the Association of Environmental and Resource Economists* 10 (5): 1335–1367.

Andam, Kwaw S., Paul J. Ferraro, Alexander Pfaff, G. Arturo Sanchez-Azofeifa, and Juan A. Robalino. 2008. "Measuring the Effectiveness of Protected Area Networks in Reducing Deforestation." *Proceedings of the National Academy of Sciences* 105 (42): 16089–16094.

Assunção, Juliano, Clarissa Gandour, and Romero Rocha. 2023. "DETER-ing Deforestation in the Amazon: Environmental Monitoring and Law Enforcement." *American Economic Journal: Applied Economics* 15 (2): 125–156.

Assunção, Juliano, Robert McMillan, Joshua Murphy, and Eduardo Souza-Rodrigues. 2022. "Optimal Environmental Targeting in the Amazon Rainforest." *The Review of Economic Studies,* rdac064.

Auffhammer, Maximilian, Eyal Frank, David McLaughlin, Beia Spiller, and David Sundig. 2021. *The Cost of Species Protection: The Land Market Impacts of the Endangered Species Act.* https://conference.nber.org/conf_papers/f157199.pdf.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119 (1): 249–275.

Bonthoux, Sebastien, Solenne Lefevre, Pierre-Alexis Herrault, and David Sheeren. 2018. "Spatial and Temporal Dependency of NDVI Satellite Imagery in Predicting Bird Diversity over France." *Remote Sensing* 10 (7): 1136.

Börner, Jan, Dario Schulz, Sven Wunder, and Alexander Pfaff. 2020. "The Effectiveness of Forest Conservation Policies and Programs." *Annual Review of Resource Economics* 12 (1): 45–64.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics,* Themed Issue: Treatment Effect 1, 225 (2): 200–230.

Cheng, Audrey, Katharine R.E. Sims, and Yuanyuan Yi. 2023. "Economic Development and Conservation Impacts of China's Nature Reserves." *Journal of Environmental Economics and Management* 121 (September): 102848.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Iván Fernández-Val. 2017. *Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments.* https://arxiv.org/abs/1712.04802.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171–198.

Einhorn, Catrin. 2022. "Nearly Every Country Signs on to a Sweeping Deal to Protect Nature." *The New York Times* (December).

European Environment Agency. 2020. *State of Nature in the EU.* Technical report 10/2020. European Environment Agency. https://www.eea.europa.eu/publications/state-of-nature-in-the-eu-2020.

Eurostat. 2021. *Forests, Forestry and Logging.* https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Forests,_forestry_and_logging.

———. 2022. *Protected Areas: Over a Quarter of EU Land.* https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20220521-1.

Fuchs, Richard, Peter H. Verburg, Jan G.P.W. Clevers, and Martin Herold. 2015. "The Potential of Old Maps and Encyclopaedias for Reconstructing Historic European Land Cover/Use Change." *Applied Geography* 59:43–55.

Geldmann, Jonas, Andrea Manica, Neil D. Burgess, Lauren Coad, and Andrew Balmford. 2019. "A Global-Level Assessment of the Effectiveness of Protected Areas at Resisting Anthropogenic Pressures." *Proceedings of the National Academy of Sciences* 116 (46): 23209–23215.

Gibson, John, Susan Olivia, Geua Boe-Gibson, and Chao Li. 2021. "Which Night Lights Data Should We Use in Economics, and Where?" *Journal of Development Economics* 149:102602.

Hobi, Martina L., Maxim Dubinin, Catherine H. Graham, Nicholas C. Coops, Murray K. Clayton, Anna M. Pidgeon, and Volker C. Radeloff. 2017. "A Comparison of Dynamic Habitat Indices Derived from Different MODIS Products as Predictors of Avian Species Richness." *Remote Sensing of Environment* 195:142–152.

Joppa, Lucas, and Alexander Pfaff. 2010. "Reassessing the Forest Impacts of Protection: The Challenge of Nonrandom Location and a Corrective Method." *Annals of the New York Academy of Sciences* 1185:135–149.

Keles, Derya, Alexander Pfaff, and Michael Mascia. 2022. "Does the Selective Erasure of Protected Areas Raise Deforestation in the Brazilian Amazon?" *Journal of the Association of Environmental and Resource Economists* 10 (4): 1121–1147.

Li, Xuecao, Yuyu Zhou, Min Zhao, and Xia Zhao. 2020. "A Harmonized Global Nighttime Light Dataset 1992–2018." *Scientific Data* 7:168.

Liang, Yuanning, Ivan Rudik, and Eric Zou. 2023. "The Environmental Effects of Economic Production: Evidence from Ecological Observations," https://static1.squarespace.com/static/56034c20e4b047f1e0c1bfca/t/647a7f669c2ddd18a1e933e7/1685749629464/LRZ_biodiversity_2023-06.pdf.

Maxwell, Sean L., Victor Cazalis, Nigel Dudley, Michael Hoffmann, Ana S.L. Rodrigues, Sue Stolton, Piero Visconti, et al. 2020. "Area-Based Conservation in the Twenty-First Century." *Nature* 586 (7828): 217–227.

Nieto, Sebastian, Pedro Flombaum, and Martin F. Garbulsky. 2015. "Can Temporal and Spatial NDVI Predict Regional Bird-Species Richness?" *Global Ecology and Conservation* 3:729–735.

Oceana Europe. 2022. *As EU Celebrates 30 years of Natura 2000, NGOs Call for These Areas to be Actually 'Protected' and for an EU-Wide Trawl Ban in Them.* https://europe.oceana.org/press-releases/eu-celebrates-30-years-natura-2000-ngos-call-these-areas-be-actually/.

Pfaff, Alexander, Juan Robalino, Catalina Sandoval, and Diego Herrera. 2015. "Protected Areas' Impacts on Brazilian Amazon Deforestation: Examining Conservation – Development Interactions to Inform Planning." *PLOS ONE* 370 (1681): 20140273.

Reynaert, Mathias, Eduardo Souza-Rodrigues, and Arthur A. van Benthem. 2023. "The Environmental Impacts of Protected Area Policy." Forthcoming, *Regional Science and Urban Economics,* https://souza-rodrigues.economics.utoronto.ca/wp-content/uploads/RSUE_land_protection-Paper.pdf.

Ribeiro, Ines, Vânia Proença, Pere Serra, Jorge Palma, Cristina Domingo-Marimon, Xavier Pons, and Tiago Domingos. 2019. "Remotely Sensed Indicators and Open-Access Biodiversity Data to Assess Bird Diversity Patterns in Mediterranean Rural Landscapes." *Scientific Reports* 9 (6826).

Schlenker, Wolfram, and Michael Roberts. 2009. "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change." *Proceedings of the National Academy of Sciences* 106 (37): 15594–15598.

Sims, Katharine R.E. 2010. "Conservation and Development: Evidence from Thai Protected Areas." *Journal of Environmental Economics and Management* 60 (2): 94–114.

Sims, Katharine R.E., and Jennifer M. Alix-Garcia. 2017. "Parks versus PES: Evaluating Direct and Incentive-Based Land Conservation in Mexico." *Journal of Environmental Economics and Management* 86:8–28.

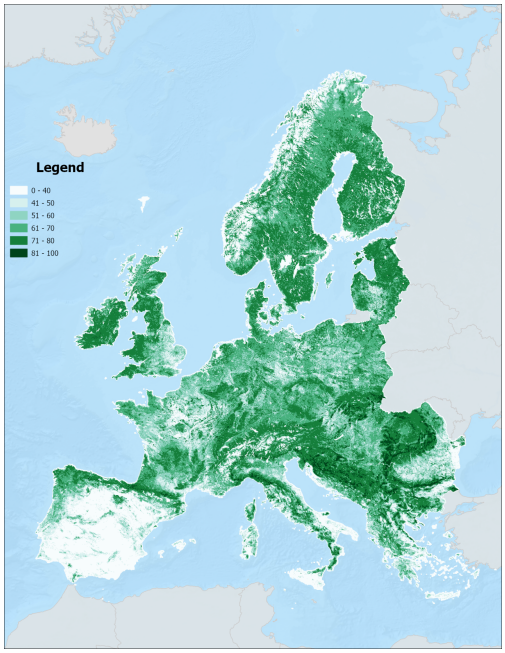
Souza-Rodrigues, Eduardo. 2019. "Deforestation in the Amazon: A Unified Framework for Estimation and Policy Analysis." *Review of Economic Studies* 86 (6): 2713–2744.

Torchiana, Adrian L, Ted Rosenbaum, Paul T. Scott, and Eduardo Souza-Rodrigues. 2020. "Improving Estimates of Transitions from Satellite Data: A Hidden Markov Model Approach," https://www.ptscott.com/papers/hmm_error_correction.pdf.

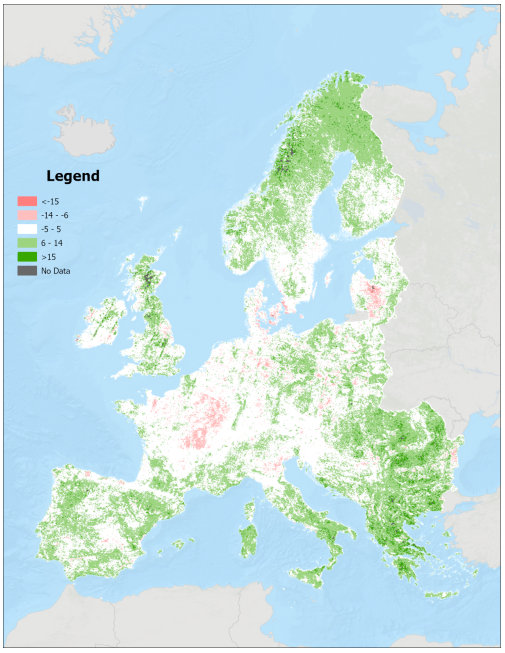
Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–1242.

Figures and Tables

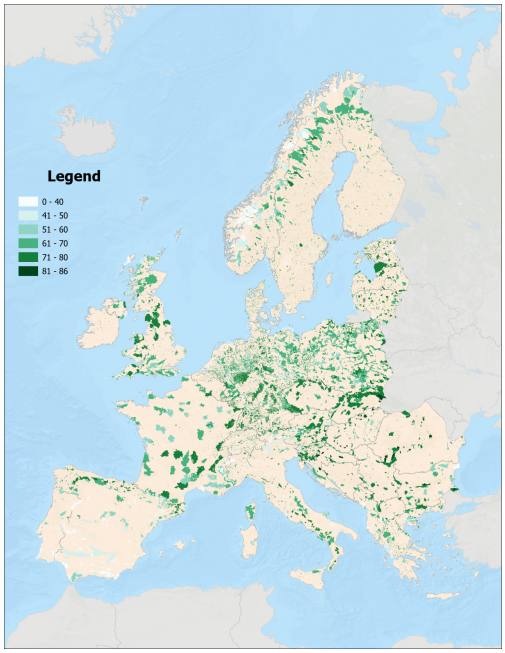
Figure 1: CDDA-level NDVI changes across sample



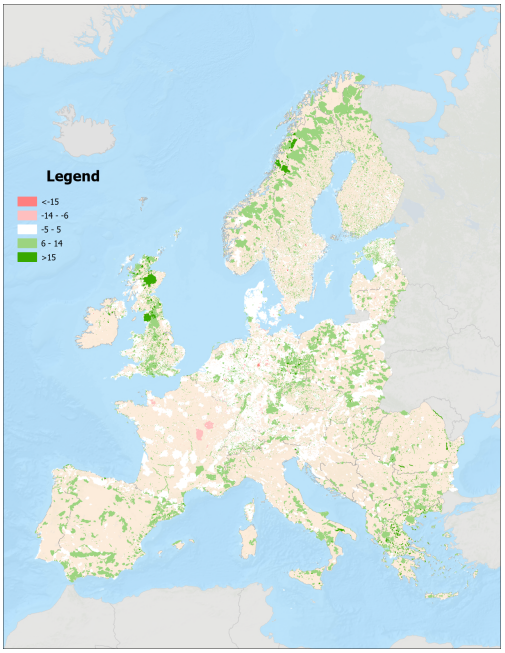
(a) Grid-level NDVI, end of sample



(b) Grid-level NDVI changes across sample



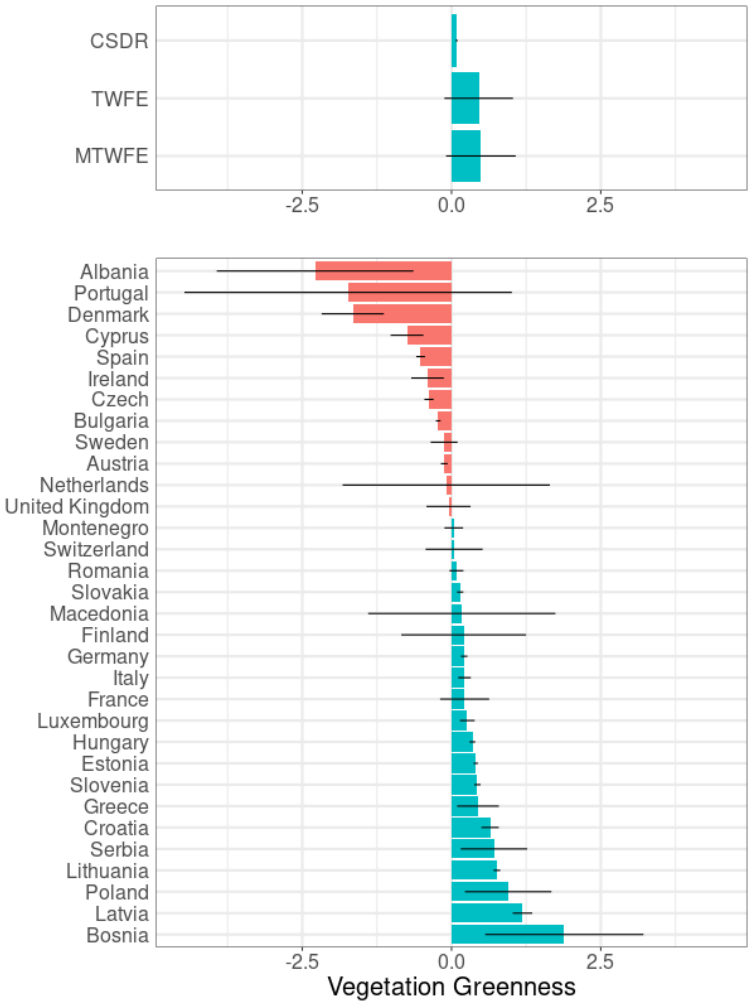
(c) CDDA-level NDVI, end of sample



(d) NDVI changes 2019-1985 within CDDA's

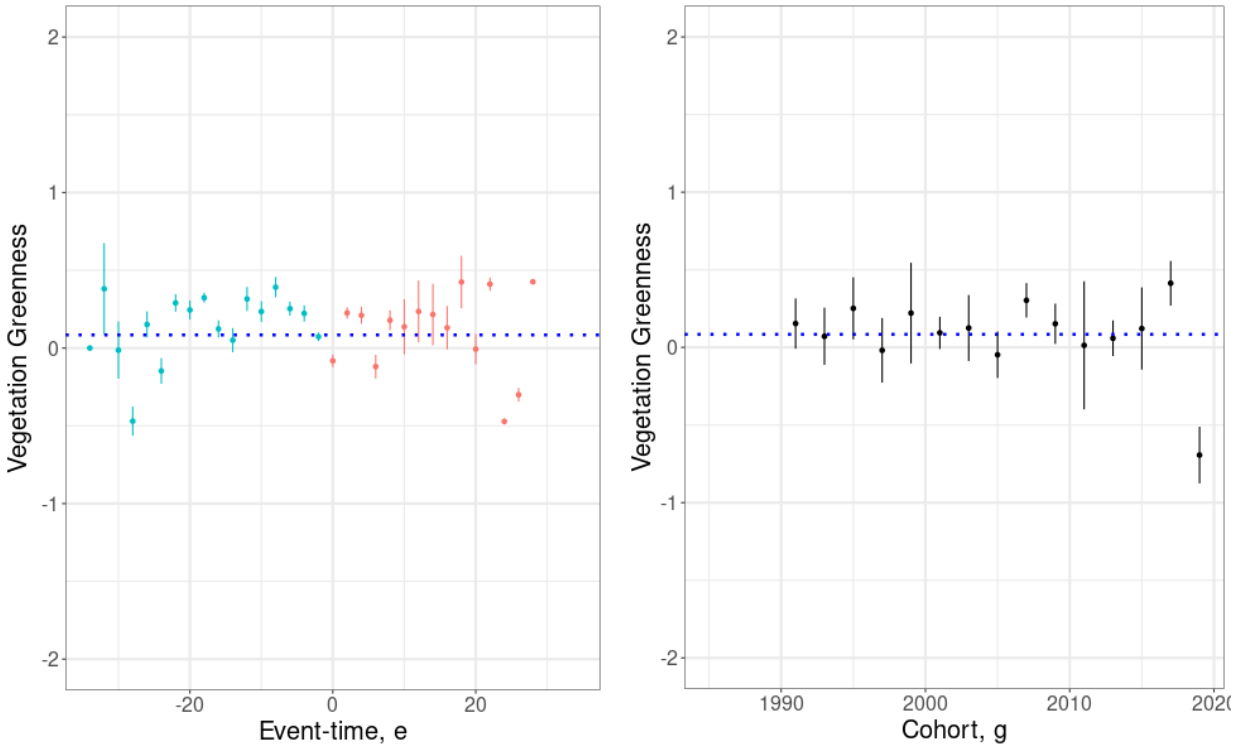
NOTES: Panel (a) plots the average NDVI across 2015-2019 for each grid. Panel (b) plots the change in average NDVI between 2015-2019 and 1985-1989. Panel (c) plots the average NDVI within each CDDA for 2015-2019. Panel (d) plots the change within each CDDA in average NDVI between 2015-2019 and 1985-1989.

Figure 2: Treatment effects on NDVI, EU-wide and by country



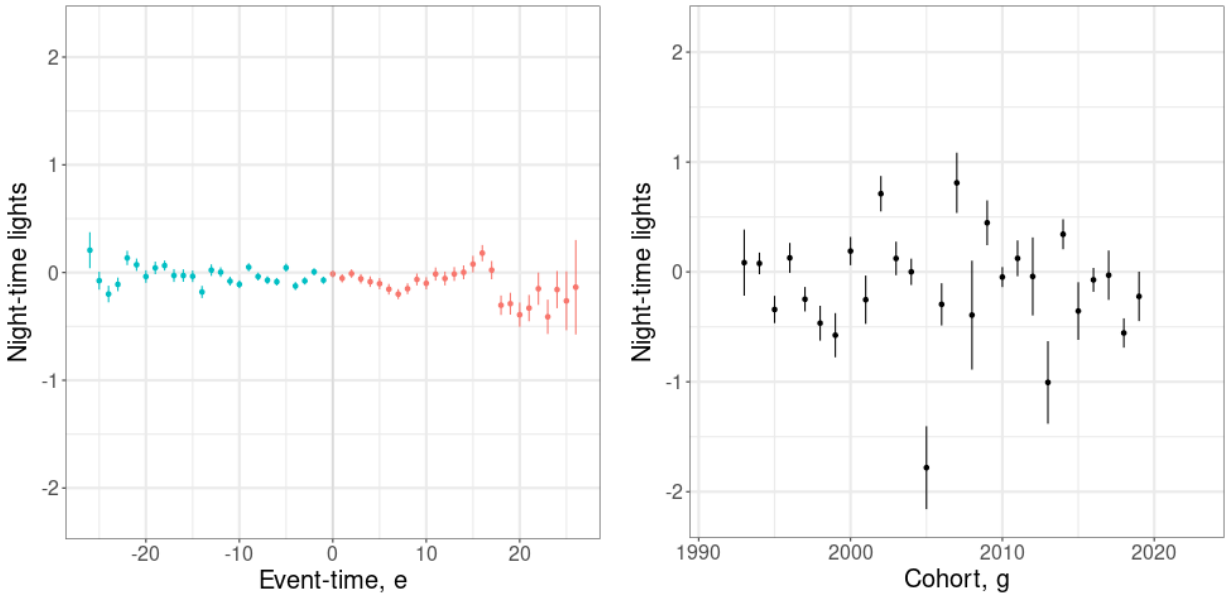
NOTES: Top panel reports treatment effects estimated through three distinct estimators: the Callaway & Sant'anna doubly-robust procedure (CSDR), two-way fixed effects (TWFE), and matched two-way fixed effects (MTWFE). Bottom panel reports CDSR treatment effects θ_c aggregated by country with bootstrapped confidence intervals. Blue bars indicate positive ATEs; red bars indicate negative ATEs. Horizontal black lines indicate 95% confidence intervals. Sample includes plots which have greenness above 40 at least once in the sample period with non-missing matching variables. Greenness varies from 0 to 100. Appendix Table A.8 presents the regression results.

Figure 3: Treatment effects on NDVI over event time and by cohort



NOTES: Treatment effects on greenness aggregated by event-study period θ_e^{EU} in the left panel, and by cohort θ_g^{EU} in the right panel. Sample includes plots which have greenness above 40 at least once in the sample period with non-missing matching variables. Greenness varies from 0 to 100. Both panels use the Callaway and Sant'anna doubly robust estimator with 95% bootstrapped confidence bands. Appendix Tables A.9 and A.10 present the regression results.

Figure 4: Treatment effects on night-time lights over event time and by cohort



NOTES: Treatment effects on night-time lights aggregated by event-study period θ_t^{EU} on the left panel, and by cohort θ_g^{EU} on the right. Night-time lights vary from 0 to 68, with all event-time estimates being smaller than 0.5 units of lumosity. Both panels use the Callaway and Sant'anna doubly robust estimator with bootstrapped 95% confidence bands. Appendix Tables A.12 and A.13 present the regression results.

Table 1: Comparison of discrete land use transitions in protected areas and never-treated areas, 1980-2010

	1980 / 2010		Cropland	Forest	Grassland	Other	Settlement	Water
Control	Cropland	% row	81.4	3.5	14.3	0.0	0.8	0.0
	Forest	% row	0.2	96.1	3.6	0.0	0.2	0.0
	Grassland	% row	3.5	10.8	85.2	0.0	0.4	0.0
	Other	% row	0.0	0.0	0.0	100.0	0.0	0.0
	Settlement	% row	0.7	0.2	0.0	0.0	99.1	0.0
	Water	% row	0.0	0.0	0.2	0.0	0.0	99.8
Treated	Cropland	% row	77.3	3.4	19.1	0.0	0.2	0.0
	Forest	% row	0.1	96.9	2.9	0.0	0.1	0.0
	Grassland	% row	1.5	9.5	88.8	0.0	0.2	0.0
	Other	% row	0.0	0.0	0.0	100.0	0.0	0.0
	Settlement	% row	0.2	0.3	0.2	0.0	99.3	0.0
	Water	% row	0.0	0.1	0.2	0.0	0.0	99.7
	Total (2010)	% row	18.3	43.4	30.4	2.1	3.4	2.4

NOTES: Table reports land use transitions relative to base year of 1980 in 2010. We tabulate transitions by treatment status: treated units are treated in 1970-80 have at least 50% of landmass in a protected area, and controls are matched on pre-1970 observables. Transitions are defined based on the HILDA land cover data, which classifies land into one of 6 land areas in each decade from 1900 to 2010. We omit a small percentage of areas which are classified as NA values.

Appendix

A Data

In this data appendix, we describe how we consolidate a variety of publicly available data sources to create country-level data files containing outcome and control variables. Table A.1 lists all data sources.

Table A.1: Data sources

	Name	Description	Data source
[1]	CDDA	Location and foundation date of protected areas	Link to source
[2]	Eurostat	National boundaries	Link to source
[3]	EEA	Shorelines of Europe	Link to source
[4]	Landsat 5, 7, 8	NDVI/Greenness	Link to source
[5]	EEA	Biogeographical regions of Europe	Link to source
[6]	EOBS	European climate surface observations	Link to source
[7]	ESDAC	Soils and topography	Link to source
[8]	ESDAC	Climate-physical regions	Link to source
[9]	SEDAC	Gridded population data	Link to source
[10]	ECMWF	Precipitation rasters	Link to source
[11]	HILDA	Discrete land-use data	Link to source
[12]	World Clim	Solar radiation	Link to source
[13]	Li et al. (2020)	Harmonized global night time lights	Link to source
[14]	BioTIME	Species count data	Link to source

A.1 Creation of grids

We define a unit of observation as a square grid cell. Grid cells divide geographic areas into evenly spaced areas with corners given by latitude and longitude coordinates. The grids are constant across time. For our vegetation greenness sample, the grid cell is 300 meters by 300 meters in resolution. We also generate a 1km by 1km grid for the nightlights analysis.

Grids are spatially joined with vector data that is spatially explicit (e.g., data that comes in the form of a shapefile or other geodatabase) using ArcGIS. Grids which intersect more than one geometry are assigned the characteristics of the geometry with the largest intersection.

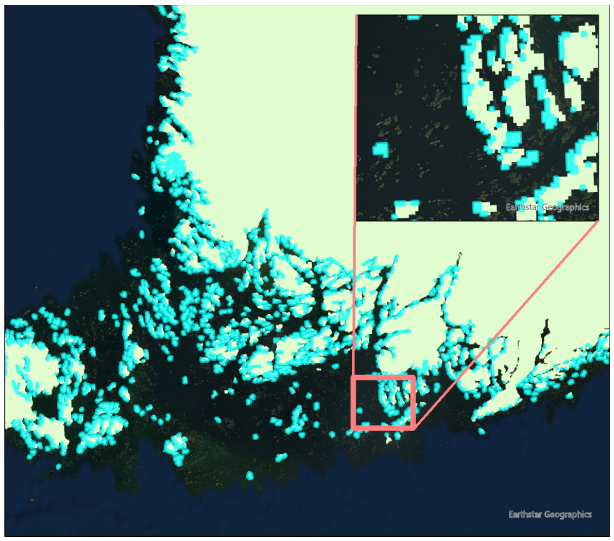
A.2 Bio-geographical regions and climate zones

We add bio-geographical regions and climate-physical zones from the European Environment Agency (EEA) and the European Soil Data Centre (ESDAC). Bio-geographical regions describe distributions and patterns of terrestrial life. The EEA data delineates these bio-geographical regions to

show distinct habitats across Europe. ESDAC produces climate-physical regions.[1] Climate-physical regions are essentially Köppen climate zones with adjustments for high mountains.

By nature, the 300m by 300m grids are not spatially fine enough to pick up on complex geographies. When merging with bio-geographical data and climate-physical regions, the grids miss craggy islands or coastlines of countries such as Norway and Finland. See Figure A.1 for an example. Of the 3,777,950 grids in Finland, fewer than 0.05% were null for bio-geographical regions and 0.5% were null for climate-physical regions. There are similar numbers for Norway. Simpler coastlines have fewer missings. For example, of Poland's 3,473,457 grids, only 114 are null for bio-geographical regions and 3,594 for climate-physical zones. Table A.2 summarizes the different bio-geographical regions and climate zones in Europe.

Figure A.1: Example of grids outside the climate-physical region boundary (Finland)



Notes: The grids highlighted in cyan are null for climate-physical zones. The country geometry for Finland is more complicated than the climate-physical zone geometry. It captures more islands and coast edges. Grids were generated over these more complex geometries. These grids do not intersect with the more simplified climate-physical regions, so therefore are null.

A.3 Protected sites

We add spatial data on the protected sites from the Common Database on Designated Areas (CDDA), an inventory of European protected areas for 38 nations. The database, maintained by the EEA, includes the location and foundation dates of protected areas established as early as 1800.

The foundation dates of 1,447 designated areas in France are missing. We manually search for the dates of the 127 largest CDDAs with missing foundation dates. We list the dates here with a URL source and notes. The remaining CDDAs with missing foundation dates did not return search

1. The climate-physical regions are based on an intersection of Köppen climate zones with NORDREGIO mountain classification deduced from GTOPO30 information.

Table A.2: Distribution of bio-geographical regions and climate zones in Europe

Bio-geographical region	Land coverage percentage
Continental	28
Mediterranean	18
Boreal	18
Atlantic	17
Alpine	13
Pannonian	3
Arctic	2
Steppic	1
Black Sea	0.2
Climate zone	**Land coverage percentage**
Cold climate, warm summer	29
Temperate climate without dry period	22
Arid/temperate climate	14
Cold climate, cold summer	13
Polar/cold climate	12
Arid/temperate climate, dry summer	9
Coastal area	2

results. We mainly retrieve the missing foundation dates from site management documents, the conservation pages of provinces, press articles of site establishment or purchase, on the "reserves naturelles" directory, on EEA site factsheets, and tourism pages for CDDAs with recreational and educational uses. The links provide examples to each of these different types of sources. Modern CDDAs may have been the result of multiple staggered land acquisitions rather than a single act of protection. We assign treatment year based on the largest additional acquisition. As a tiebreaker, we assign the average of the purchase dates. See the notes here for detail.[2]

Many of the protected area polygons provided by the EEA have topology errors. Self-intersections are common in the CDDA dataset. These are corrected in ArcGIS.

After these additions and corrections, we relate the CDDA information to the grids via a spatial join in ArcGIS. For each CDDA we add the foundation year and unique CDDA ID so that we can match non-spatial information of the CDDA below. We will also calculate the area of overlap of each CDDA with the grid(s) it covers. Some grids do not fall entirely within CDDAs. Such partial coverage of grids is important when determining which grids are protected vs. treated. We define any plot that overlaps a CDDA as treated.

2. One example is Les Pelouses de Blere. In 2003, 14.34 hectares were acquired by Le Conservatoire d'Espaces Naturels Centre-Val de Loire. In 2005, the municipality gave the conservatory 63.58 hectares to manage. The second land acquisition was more than four times greater than the original land acquisition. Because the second land acquisition was larger than the first, we chose the date of the second acquisition for the foundation year. Another site, Les Friches Des Parterres, was acquired "par le Conservatoire de 22.87 ha entre 1995 et 1999" For this site, we chose 1997 for the foundation date.

A.4 Raster-derived variables

Raster data employs a matrix-based structure, where each cell (or pixel) in the matrix stores a value representing a particular attribute (such as NDVI, elevation, or rainfall). To relate raster data to our grids, we use the exact_extract() function in the exact-extractr R package to efficiently relate raster data to polygons. Table A.3 gives a description of each of the raster-derived variables added to the country-level grids data. The table provides a brief description of each variable, the units of the data, the spatial resolution, the frequency of the time series for time-series variables (annual, biennial, every 5 years, etc), and the source of the data.

Table A.3: Table of variables which come from a raster format

Variable name	Units	Spatial resolution	Time step, years	Data reference year(s)	URL
Climate					
Long-run precipitation	mm/year	0.5 degrees		average, 1970-2000	NOAA
Growing season length	days	0.1 degree	1	1985-2019	Copernicus
Heating degree days	deg C	0.1 degree	1	1985-2019	Copernicus
Rainfall	meters	11 km	1	1985-2019	Google Earth Engine
Land characteristics data					
Topsoil potassium concentration	g/kg	500 meters		2019	ESDAC Website
Topsoil nitrogen concentration	mg/kg	500 meters		2019	ESDAC Website
Topsoil phosphorus concentration	mg/kg	500 meters		2019	ESDAC Website
Terrain measures					
Slope steepness	combined LS-factor	100 meters		2015	ESDAC Website
Elevation	meters	100 meters		2015	ESDAC website
Slope angle index	0-8	200 meters		2018	ESDAC Website
Soil suitability index	0-4	1000 meters		2016	ESDAC Website
Economic value measures					
High-value farmland indicator	binary	100 meters		2015	EEA Europa
Distance to shoreline	meters	300 meters		2017	EEA Europa
Population density	count per km^2	30 arc sec	5	1985-2019	SEDAC Website
Outcome measures					
Greenness, LANDSAT-5	index	300 meters	2	1985-2013	Google Earth Engine
Greenness, LANDSAT-8	percent	300 meters	2	2013-2019	Google Earth Engine

Here we provide further information on the raster data that we use in our study.

Climate data. We obtain both a long-term average precipitation and a time-varying measure of precipitation for each grid. We collect monthly long-term averages (long-term mean 1981-2010) of total rainfall from WorldClim. The month's values are an average precipitation for that month from 1981-2010. We average the monthly data to get a total annual figure.

We generate time-varying precipitation from the total precipitation band in the ERA5-Land Monthly Averaged - ECMWF Climate Reanalysis dataset. The climate measures in ERA-5 have a resolution of 11,132 meters. The total precipitation band is the depth of monthly precipitation in meters. ERA-5 total precipitation captures most precipitation but does not include fog, dew, or precipitation which evaporates before reaching the earth.

We use annual total precipitation, rather than summer precipitation that would coincide with our NDVI imagery, because of the importance of year-round rainfall for summer vegetation growth. Pasho et al. (2012) found that larger tree-ring width—an indication that the tree grew more during that year—is related to the amount of autumn and winter rain that had recharged the soils. Dannenberg, Wise, and Smith (2019) and Vieira, Nabais, and Campelo (2021) found that trees in the United States had decreased radial growth and higher mortality risk when winter and summer precipitation were lower. These two articles demonstrate the importance of winter rainfall in vegetation health and justify our use of a yearly total rainfall measure.

To obtain the time-varying climate data from "cold indices" on E-OBS indices, we select "annual" in growing season length and "annual" in heating degree days.

Land characteristics data. The chemical and physical land characteristics are sourced from ESDAC. ESDAC conducted a large survey with approximately 20,000 topsoil samples of soils in Europe to produce a coherent pan-European physical and chemical topsoil database, which can serve as baseline information for an EU wide harmonized soil monitoring. We extract nitrogen, potassium, and phosphorus levels from the LUCAS 2009/2012 topsoil database. We extract soil biomass productivity variables from the EEA 2006 classification. The soil suitability score was created in 2016. We also extract elevation, slope angle, and steepness of slopes from ESDAC.

Economic value measures. The Center for International Earth Science Information Network (CIESIN) of NASA's Socioeconomic Data and Applications Center (SEDAC) provides gridded population density rasters. CIESIN estimates population density every 5 years to be consistent with national censuses. These numbers are scaled to match UN country-level totals. The data is available at 30 arc sec (1km x 1km) spatial resolution, slightly coarser than the grids. We interpolate the population density data linearly across time.

We compute the distance of each grid to the closest shoreline from polyline data available from the EEA. We compute the the Euclidean Distance in ArcGIS at 300 meter pixel resolution and store the data as a raster.

The EEA has created a binary image of high nature value farmland (HNVF). This HNVF measure indicates the potential biodiversity value of existing farms. A value of 1 represents farmland

of high nature value and 0 represents low nature value farmland. We use the data as a proxy of the counterfactual value of agricultural land.

Solar radiation data is available from WorldClim for the period 1950-2020 and is based on ground-based on-site observations and a CERES global radiation satellite product. The data reports average monthly solar radiation levels. From this data, we compute an annual average solar radiation raster.

A.5 NDVI

A.5.1 Background

NDVI (normalized difference vegetation index) is an index with values between -1 and 1 representing the level of "greenness" of land cover. It is a well-established index for vegetation monitoring, indicated level of greenness, and plant health. Negative values of the index correspond to water. Low values (0.1 to 0.2) correspond to barren areas, settlements, snow and clouds. Values between 0.2 and 1 correspond to vegetation. NDVI is the ratio between the red (R) and near infrared (NIR) bands:

$$NDVI = (NIR - R)/(NIR + R)$$

We use Google Earth Engine (GEE) to transform Landsat surface reflectance imagery to NDVI and export the NDVI image. Landsat is a satellite imagery program for the entire earth. The early versions, Landsat 1-4, are very similar, are not of sufficient quality, and do not correspond to later versions. Landsat 5-8 have much higher resolution and contain the necessary visual information to capture NDVI consistently across each sensor. We, therefore, use the Landsat 5 data starting in 1985.

We use two-year periods for obtaining NDVI images. Annual data suffers greater missing data issues related to cloud cover. Because accurate NDVI measurement requires leaf-on conditions, we limit the sample of images to summer months. We center our search for images around July as it is the height of summer greenness (Peled et al. 2010; Van Oijen et al. 2014). Ideally, we would produce a single cloudless image using only a two week period—corresponding to the complete earth imaging time of 16 days of the Landsat satellites—in July. However, clouds are often present, requiring that the range of months to search for images be extended to either June through August or, in exceptional circumstances, May through September, until the percentage of missing images falls below 5 percent. We discuss the details of the procedure for producing a cloud-free mosaic in our data repository.

When we compute average NDVI by the discrete land use categories in HILDA, we find that forests are the greenest land use category with NDVI above 60, except for those in the Mediterranean. Forests in Germany and France have an average NDVI of 70 in 2010, and seven countries have an average NDVI above 70 in their land that is categorized as forest. Grassland has a somewhat lower NDVI that is mostly between 60 and 70, with a much lower value for the Mediterranean

countries. Cropland has the lowest NDVI, with values of 58 for Germany and 50 for France in 2010. NDVI variation thus matches our priors about land use but differences between biomes and climate regions are important. Forests in arid regions of Spain are characterized by shrubs and greater spacing of vegetation and have an NDVI of only 55 in 2010, while grassland in the United Kingdom in 2010 has an NDVI of 65. We estimate treatment effects separately by country and include climate, weather, and soil controls to ensure valid treatment-control pairs.

A.5.2 Landsat 7 scan line correction

The Landsat 7 satellite was launched on April 15, 1999. It collected quality images until May 31, 2003 when the Scan Line Corrector (SLC) in the Enhanced Thematic Mapper (ETM+) instrument failed. Landsat 7 images after this date are not usable for our analysis. The 1999-2003 Landsat 7 operation period overlaps with Landsat 5 (April 1, 1984 - June 5, 2013). During this time frame, we use Landsat 7 images and fill in any missing pixels with Landsat 5 data.

A.5.3 Landsat 8 OLI to Landsat 5 ETM+ spectral response correction

Landsat 8 and Landsat 5/7's sensors are largely comparable; they have the same spatial resolution and 16-day revisit time. However, the spectral response functions of the two sensors differ. Landsat 8's Operational Land Imager (OLI) is an improvement on Landsat 5/7's Enhanced Thematic Mapper Plus (ETM+). However, we need to correct the spectral response function of Landsat 8 to make NDVI directly comparable throughout the panel.

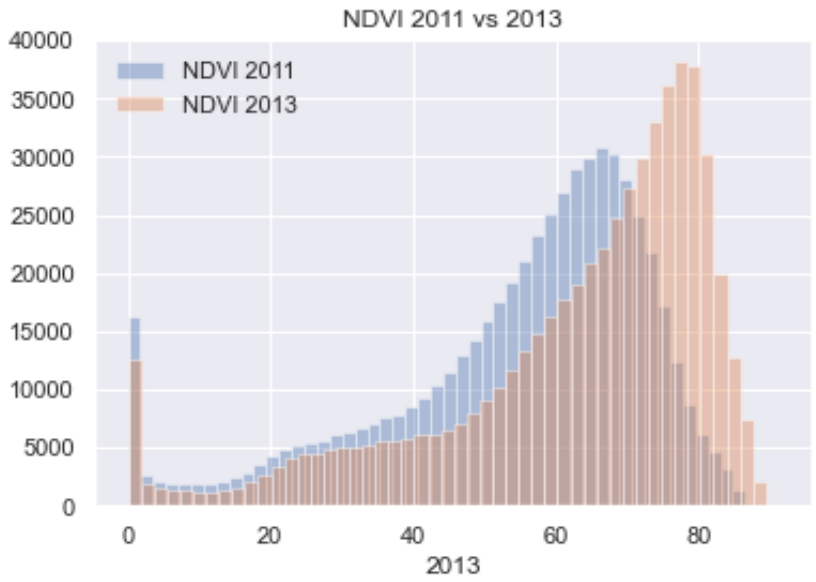
The differences in the spectral response functions of the two sensors lead to a "brighter" NDVI for Landsat 8 imagery than for Landsat 5. This is clear from the histogram of NDVI for all European grids in 2011 (Landsat 5 ETM+) and in 2013 (Landsat 8 OLI) in Figure A.2. There is a noticeable skewing towards higher values of the Landsat 8 NDVI in a brief period of just two years. Because the differences in NDVI in Figure A.2 appear to come largely from measurement, we use a correction to harmonize the NDVI measure across each satellite. Roy et al. 2016 provide the coefficients to apply to each band to harmonize Landsat 8 to Landsat 5. We chose to harmonize OLI to ETM+ because the thematic mapper makes up the majority of our NDVI imagery (1985 to 2011, 13 two-year images).

B Econometrics

B.1 Treatment effect aggregations

The estimated average treatment effects θ_{cgt} vary across countries, over event time $e = t - g$, and across treatment-assignment cohorts g. Define the set of countries as $\mathcal{C}$ and the set of cohorts as $\mathcal{G}$. Recall we defined $\mathbb{T} \in \{1992, \cdots, 2019\}$ as the set of time periods in our study sample; we will refer to the last year of the sample as T. The country-specific treatment effects reported in Figure

Figure A.2: Spectral response affecting NDVI 2011-2013



NOTES: The histogram shows a strong skewing towards higher NDVI values from 2011 to 2013 across all of Europe. This is not attributable to large scale land cover shifts; this is the result of the difference in the spectral response functions of ETM+ and OLI. Between 2011 and 2013, Landsat 5 was decomissioned and Landsat 8 became operational.

2 are computed by averaging over the treated time periods and cohorts:

$$\theta_c = \sum_{g \in \mathcal{G}} w_{cg} \sum_{t>g}^{T} w_e \theta_{cgt} = \sum_{g \in \mathcal{G}} \frac{N_{cg}}{\sum_{g \in \mathcal{G}} N_{cg}} \sum_{t>g}^{T} \frac{1}{T-g+1} \theta_{cgt}, \tag{A.1}$$

where N_{cg} is the number of observations in country c that were treated in foundation year g. We average treatment effects across all cohorts; within each cohort, we average over all treated periods for that cohort in our sample. Because our panel is balanced, every event time has an equal weight $w_e = \frac{1}{T-g+1}$ within its cohort.

We compute the overall treatment effect estimate across the European Union with a weighted average of the country-specific θ_{cgt} parameters where the weights depend on the number of treated observations for each tuple $\{c, g, t\}$. Define the number of plots i treated in cohort g and observed in period t within country c as N_{cgt}. Then, define the EU-wide θ_{gt}^{EU} as:

$$\theta_{gt}^{EU} = \sum_{c \in \mathcal{C}} w_{cgt} \theta_{cgt} = \sum_{c \in \mathcal{C}} \frac{N_{cgt}}{\sum_{c \in \mathcal{C}} N_{cgt}} \theta_{cgt}. \tag{A.2}$$

Generally, such a weighted sum requires estimating the weights w_{cgt}. In our setting, we observe the true weights on observations because we have the land census in the European Union. As a result, we directly adjust standard errors without calculating a separate standard error for the weights. We obtain the overall θ^{EU} as the arithmetic mean over cohorts and their treated periods.

The EU-wide aggregation is asymptotically valid when treatment effects are i.i.d. across countries. Violations of this aggregation where standard errors would need to be adjusted include cases where multiple countries create and implement regional conservation plans.[3] We do not observe such coordination in practice; EU member states independently decide which land to protect. The choice of aggregation technique is also motivated by the computational cost of the doubly-robust estimator.

Figure 3 aggregates θ_{gt}^{EU} over event time and cohorts. We first recast θ_{gt}^{EU} as θ_{ge}^{EU} by setting $e = t - g$. Aggregating over cohorts gives an event-study type estimator θ_{e}^{EU} indexed by event times:

$$\theta_{e}^{EU} = \sum_{g \in \mathcal{G}} w_g \theta_{ge}^{EU}, \tag{A.3}$$

where w_e equals the fraction of cohorts treated at event time e.

Finally, aggregating over event times gives a measure of the average effect of being treated at time g:

$$\theta_{g}^{EU} = \frac{1}{T - g + 1} \sum_{e=0}^{T-g} \theta_{ge}^{EU}, \tag{A.4}$$

where we weight each event time within cohort g the same, implying a balanced panel.

Event-study style aggregations represent multiple underlying mechanisms if the composition of cohorts represented in each θ_e is different.[4] To ensure that results are not driven by sample composition alone, we construct a series of panels requiring units to be in-sample for at least three periods before and after treatment. This functionally excludes the earliest treated and latest treated plots from the sample.

We also use equivalent aggregations at the country-level to test for selection mechanisms that vary across countries. We aggregate within-country and across event time to obtain θ_{cg} in Appendix C.3.

B.2 Estimation details: Staggered difference-in-differences design

We use the 'did' package in the R programming language.[5] We select all observations for which there are no missing variables and keep grids that cross the greenness threshold of NDVI 40 at least once in the sample. Plots that fully or partially overlap a CDDA are considered treated units for our results on NDVI. We construct our estimator in the syntax of the 'att_gt' command. We treat the data as a panel, using the rounded year of protection as the cohort definition. The command constructs

3. As a consequence of this assumption, our standard errors for the EU-level effects are likely smaller relative to a confidence band that incorporates covariance in country-level treatment effects.

4. Because we aggregate across cohorts, the sample is densest close to event time 0, with fewer units used to calculate effects in the earliest pre-periods and latest post-periods. For example, if a unit is treated in 2019, it has only one post-period (2019) but has 17 biennial pre-periods dating back to 1985. No other cohort is represented in the dynamic effect for the bin $t - g = -34$ as 2019 is the last in-sample cohort.

5. See: https://www.rdocumentation.org/packages/did/versions/2.1.2).

bootstrapped confidence intervals. We keep all computational defaults regarding bootstrap size and sampling procedures. Our estimator is calculated at the country level.

For large countries, the did package is too slow to produce estimates in a reasonable timeframe. Therefore, we conduct a subsampling procedure of the estimator, where we draw subsamples at the cohort-observation level. We do this for Finland, France, Germany, Italy, Poland, Romania, Spain, Sweden, and the United Kingdom. The subsampling procedure mirrors the bootstrap of the estimator by ensuring that every cohort and event time is represented in each subsample, so that the full set of θ_{gt} is estimable for each subsample. We subsample with replacement and define new plot identifiers (e.g., a new grid cell-level ID) to ensure the command runs with potential duplicates in the data. We sample 100 clustered 5% draws for each country, bootstrapping the entire matrix θ_{cgt}, θ_{ce}, θ_{cg}, and θ_c as well, obtaining bootstrapped standard errors. Because we already find the estimator to be slow with this bootstrapping scheme, we do not apply the Callaway & Sant'anna wild bootstrap to compute uniform confidence bands. These could, in principle, widen our standard errors as they account for covariance between dynamic treatment effects, but given the level of precision at which our treatment effects are computed, we do not anticipate these change the interpretation of our results as a precise 0.

Estimates for Cyprus, Malta, and Liechtenstein are not computed due to missing data in the time-varying weather patterns. There are significant missing shares of time-invariant variables for Switzerland, Luxembourg, and Montenegro. A subset of countries have a very high ($> 99\%$) share of one or more matching variables: these are matched on the remaining covariates. Countries lacking slope sleepness and soil suitability measures are Switzerland, Serbia, Bosnia, Albania, and Montenegro. Malta lacks rainfall data on 60% of its landmass. We report results for estimators that omit time-invariant variables from the matching procedure.

The 'did' package computes clustered standard error at default by clustering at the plot level, allowing for serial correlation in the error terms at the observation level. Additionally, we cluster at the protected area unit level (CDDA number) to account for potential spatial correlation. This two-way clustering makes standard errors robust to correlation in the time and cross-sectional (spatial) dimensions.

We maintain the same matching variables and technical specifications for our nightlights analysis. There are two key exceptions. Rather than matching on pre-period greenness, we match on pre-period nightlights. We require a single pre-period of 1992 as the nightlights data do not cover the same breadth of data as the greenness data. Because nightlight imagery is taken at night, it is less frequently missing than greenness and we obtain an annual panel. We therefore have treated cohorts from 1993 through 2019 (27 total potential values of g). Our nightlights data is also collected at a lower spatial resolution (1 km grid cells, see A), so we reconsider our treatment definition. As our second deviation from the greenness strategy, treated grid cells must overlap at least 50% (in area terms) with a protected area.

B.3 Estimation details: Conditional average treatment effects (CATE)

We use the CATE estimator provided in the R package 'grf', generalized random forest.[6]. Our data matches the package's description of a medium-sized dataset. We follow the package documentation to construct a suitable estimator for our dataset. Computationally, we target an accurate calibration that passes the test in Chernozhukov et al. (2017) and precisely separates the top and bottom quartiles of estimated CATEs. This latter point ensures heterogeneity is rejected with precision if no significant heterogeneity exists.

We apply the Wager and Athey (2018) estimator to a sample of plots. From each country, we draw a 5% sample clustered at the cohort level. We replicate the research design of our Callaway and Sant'Anna (2021) estimator and focus on plots that had a greenness measure above 40 at some point in the sample and omit plots protected before 1991. We estimate our CATEs on the full sample (a stacked design) rather than calculating cohort-time-specific treatment effects, as in the Callaway and Sant'anna estimator. Econometrically, estimating the stacked design remains unbiased as our prior results reject the existence of heterogeneity across cohort and event time.

We supply all potential exogenous variables to compute the forest. We include interaction terms: initial greenness and elevation, initial greenness and distance to shore, and initial greenness and climate zone. Exogenous variables are used for matching classification trees and heterogeneity regression trees, but each tree is trained on separate sub-samples ("honesty", in the language of Wager and Athey (2018)). We omit missing observations. Categorical variables are converted to indicator variables.

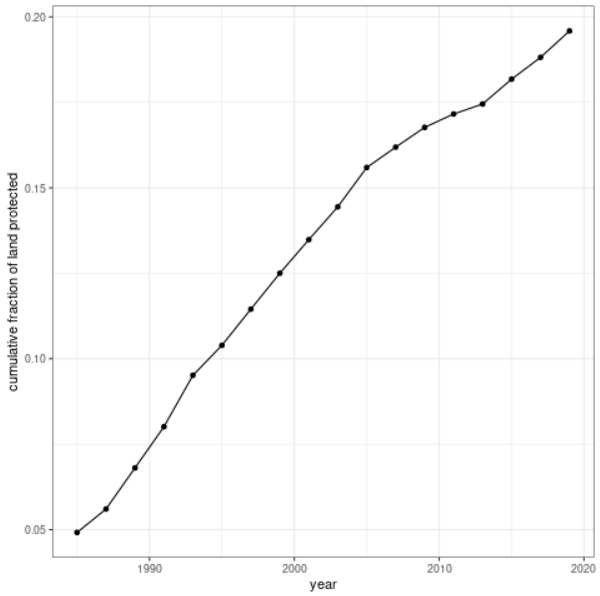
Finally, treatment is defined by an indicator equal to 1 as long as $t \geq g$, and 0 otherwise. With a matrix of observables, an outcome vector of greenness, and a treatment indicator, we split the data into a training and a test dataset. We retain 30% of the data for training, and the remainder for testing. Training data is used to calibrate the random forest. The algorithm statistically tests the calibrated random forest on a separate, withheld test dataset.

We change three computational parameters to calibrate the estimator. These are: (1) we increase the number of trees trained to 4000 to account for our large sample size, (2) we set a minimum node size of 10,000 to aggressively prune trees and ensure calibration is only on meaningful heterogeneity, and (3) we lower the sample fraction to 20% to make computation tractable. We find a higher minimum node size is critical to passing our heterogeneity tests. Meanwhile, the number of trees should be seen as a minimum—more trees do not improve calibration (we have tested up to 10,000). The sample fraction is largely dictated by computational constraints.

We report the results of the Chernozhukov et al. (2017) test of meaningful heterogeneity using the 'test_calibration' command. For all computed conditional average treatment effects, we ensure that they are estimated on the subsample of data with overlap (the 'target.sample' parameter) and we test both a doubly robust and an inverse propensity weight (option 'AIPW') approach. Reported values are doubly robust. CATEs are thus calculated using the 'average_treatment_effect'

6. See: grf-labs.github.io/grf/

Figure A.3: Fraction of landmass in the study area which was protected in or before the given year



NOTES: Data aggregates up from the 300 meter grid-cell level. Any 300 meter grid cell with non-zero overlap of a protected area is defined as treated. The y-axis thus reports the percent of grid cells which overlap any protected area founded on or after the year reported on the x-axis.

command, subsetting to the portion of data of interest.

C Additional Tables and Figures

C.1 Summarizing greenness and protection across the EU

Our large data collection effort gives us the unique opportunity to collate novel descriptive statistics, both about vegetation growth and conservation policy. Our data in Figure A.3 shows that in the study period, by 2020, around 20% of European landmass is protected. Further, relative to the start of our study period in 1985, this is a 300% increase in protected landmass. Our sample period thus captures the epoch with the largest percentage point increase in protected area growth.

There is variation in protection regimes. Some areas restrict all human access, while others allow industrial and agricultural activities. While we do not observe the realized level of enforcement, the IUCN categorizes protected areas by their relative strictness. For example, categories Ia and Ib indicate the strictest level of protection, where due to wildlife preservation human activity is strictly limited to either indigenous communities with prior claims to that land or approved research activities. Table A.4 shows that these strict reserves comprise around 7% of the treated areas, or around 5,600 unique protected areas. They are also on average the greenest classification among areas which have IUCN classifications. 47% of protected areas are habitat and species management areas, 14% are protected landscapes, 12% are nature reserves, national parks, and

Table A.4: Average greenness in 1989 by IUCN protected area category

IUCN category	Percent	Mean greenness
National park	0.55	56.32
Natural monument	4.35	63.20
Protected area with sustainable use of natural resources	1.39	63.18
Protected landscape	14.02	62.49
Species management area	46.58	66.52
Strict nature reserve or wilderness area	7.02	70.80
No IUCN	26.09	72.63

NOTES: Table summarizes the strictness of protection of CDDAs. Sample restricted to terrestrial protected areas with non-missing greenness. Excludes areas with missing (NA) IUCN data. Percentages as a percent of the total number of grid cells in this sample, where a grid cell is included in the IUCN category if it overlaps any CDDA.

natural monuments, and 26% lack an IUCN category. Species management areas (IUCN category IV) are designed to encourage the propagation of a particular species in the region but are not necessarily as limited in their economic uses. Notably, areas which are protected but are not assigned an IUCN category are on average greener than the areas with the strictest protection.

C.2 Propensity score weighting and balance tables

The main identification challenge we face is the nonrandom selection of land for treatment. While the aggregate picture in the EU suggests (Table A.5) that there may be overlap between treated and never-treated units, zooming into countries suggests significant imbalances. We demonstrate this imbalance in Table A.6 for the case of France (chosen because it is relatively large and has highly-heterogeneous protected areas). Protected areas are in less populated areas (lower population density), and steeper and less accessible areas (elevation and slope). To obtain balance, the algorithm of Callaway and Sant'Anna (2021) applies cohort-specific propensity score weighting. This is important in our setting as we expect that later-treated cohorts consist of different land than early-treated cohorts. In each cohort, the procedure develops a propensity weight based on variables that appear in the vegetation greenness production function: elevation, slope steepness, soil quality, solar radiance, growing season length, heating degree days, rainfall, and lagged greenness. In each cohort, the resulting estimator weighs control units by their similarity on vegetation-relevant observables to treated protected areas. Table A.7 shows an example of how the inverse propensity weighting shrinks the difference in means of variables between treated and untreated units in the year before treatment, in France and for the cohort treated in 2005.

C.3 Treatment effect estimates: NDVI

Here we report the results of the Callaway and Sant'Anna estimators in more detail. Table A.8 lists the average treatment effect aggregated over both event time and cohort for each country in

Table A.5: Balance table for the entire European Union

	Never-treated		Treated	
	N	Mean (Std. dev.)	N	Mean (Std. dev.)
NDVI	584,120,807	60.90 (13.46)	98,347,631	63.70 (12.52)
Growing season length	583,331,749	255.97 (28.13)	98,020,824	248.10 (29.47)
Heating degree days	583,331,749	2,779.79 (637.83)	98,020,824	3,004.24 (631.23)
High-value farmland fraction	569,877,764	0.14 (0.27)	95,463,575	0.18 (0.30)
Population density	583,331,749	99.36 (298.25)	98,020,824	105.58 (258.29)
Rainfall (mm)	583,331,749	0.02 (0.01)	98,020,824	0.03 (0.01)
Slope angle	584,120,807	2.06 (1.14)	98,347,631	2.45 (1.23)
Slope steepness	584,120,807	1.51 (1.70)	98,347,631	2.12 (1.95)
Soil suitability index	584,120,807	3.51 (0.79)	98,347,631	3.34 (0.81)
Solar radiance	584,120,807	11,350.22 (651.39)	98,347,631	11,090.44 (569.62)

NOTES: Tables compares observables for ever- vs. never-treated units. The table aggregates over areas with a greenness of at least 40 at one point in the sample period. We enforce that land has non-missing values of controls to be included in the table. Plots which were protected before 1990 are trimmed to ensure we can compare several periods' worth of pre-trends. Time-varying variables (heating degree days, rainfall, and growing season length) are averaged at their 1989 levels, the last pre-period before we analyze treatment effects.

Table A.6: Balance table for France

	Never-treated (N=4,552,137)		Treated (N=472,741)			
	Mean	Std. dev.	Mean	Std. dev.	Diff. in means	Std. error
Greenness	57.58	12.87	61.80	12.54	4.22	0.02
Population density	98.47	369.26	45.85	157.06	-52.62	0.15
High-value farmland fraction	0.12	0.29	0.26	0.38	0.14	0.00
Rain	25.33	6.73	28.08	8.23	2.75	0.01
Heating degree days	2291.87	609.08	2408.68	767.18	116.81	1.08
Growing season length	316.91	36.50	295.19	52.22	-21.72	0.07
Soil suitability index	3.67	0.73	3.34	0.89	-0.33	0.00
Slope angle	1.95	1.36	2.91	1.85	0.96	0.00
Slope steepness	1.39	2.23	2.80	3.27	1.41	0.00
Elevation	275.84	318.55	545.61	472.87	269.77	0.67
Solar radiance	12224.16	1117.75	13172.44	1139.29	948.28	1.57

NOTES: Treated data are aggregated across all cohorts. The sample selection procedure for this table is the same as the rest of the paper. Plots which were protected before 1990 are trimmed to ensure we can compare several periods' worth of pre-trends. Plots which never attain a greenness value above 40 in the entire sample are trimmed from the sample. Plots with a missing covariate are also omitted. Time-varying variables are captured in 1989, the last pre-period year in the sample. Standard errors computed assuming independent populations.

Table A.7: Balance table for France, specific to the cohort treated in 2005 after propensity-score matching

	Never-treated (N=1,080,841)		Treated (N=54,838)		Raw		IPW	
	Mean	Std. dev.	Mean	Std. dev.	Diff. in means	Std. error	Diff. in means	Std. error
Greenness	69.13	9.83	71.01	8.97	1.88	0.04	0.85	0.04
Population density	53.85	150.81	17.71	34.14	-36.14	0.015	-23.71	0.22
High-value farmland fraction	0.21	0.35	0.35	0.38	0.14	0.002	0.11	0.002
Rain	30.70	8.42	36.64	4.88	5.94	0.02	3.87	0.02
Heating degree days	2068.74	749.14	2555.52	743.92	486.78	3.09	232.00	2.78
Growing season length	277.11	46.70	246.20	41.69	-30.91	0.17	-19.75	0.17
Soil suitability index	3.57	0.76	3.40	0.85	-0.17	0.004	-0.19	0.004
Slope angle	2.39	1.64	3.03	1.57	0.64	0.012	0.66	0.01
Slope steepness	2.01	2.64	2.77	2.88	0.76	0.012	1.07	0.02
Elevation	512.96	461.92	899.64	512.99	386.68	2.15	196.19	1.90
Solar radiance	12542.72	1158.48	13144.22	1061.73	601.50	4.39	364.77	4.12

NOTES: Balance table calculated in first pre-treatment panel year, 2003. IPW = inverse probability weight matching. Matching variables were long-run precipitation, elevation, solar radiance, slope steepness, a soil suitability index, slope angle, rainfall, heating degree days, a second-order polynomial in greenness. These variables were taken in the pre-period. Additionally, we included a three-year average and variance of rainfall, greenness, and heating degree days between 1985-1989. The sample here is trimmed for balance: propensity scores lower than 0.01 and higher than 0.99 are removed.

the European Union. We report bootstrapped standard errors using the Callaway and Sant'Anna methodology. We also report the number of unique individual plot identifiers remaining in our data for each country. Some countries, such as Switzerland, are under-represented due to missing data. Others are under-represented because of a lack of sufficient overlap in some of the control variables. Most countries are very precisely estimated (less-represented or smaller countries such as Albania or Bosnia are less precise): protected areas on average experience a change in greenness of less than one. Exceptions tend to skew in the negative direction: protected areas became relatively less green, rather than more green. The single well-powered exception is Latvia, where one unit of greenness is gained on average due to protection.

We report examples of dynamic treatment effect plots θ_{ce} here. These dynamic effects are most useful as visual tests of the parallel trends assumption. It would be expositionally overwhelming to report visual evidence for all countries; instead, we show two examples (Poland and Spain) that are representative of an overall absence of (trends in) treatment effects.

We start with the results for Poland, a large country with lots of standing old-growth forest. Figure A.4 shows the estimated dynamic treatment effects $\hat{\theta}_{ct}$ for Poland. The treatment effect for the base period -2 is fairly close to 0, suggesting that we have decent claim to a conditional parallel trends assumption. We see that up to 20 years after protection, all treatment effects are less than 2 in absolute value.

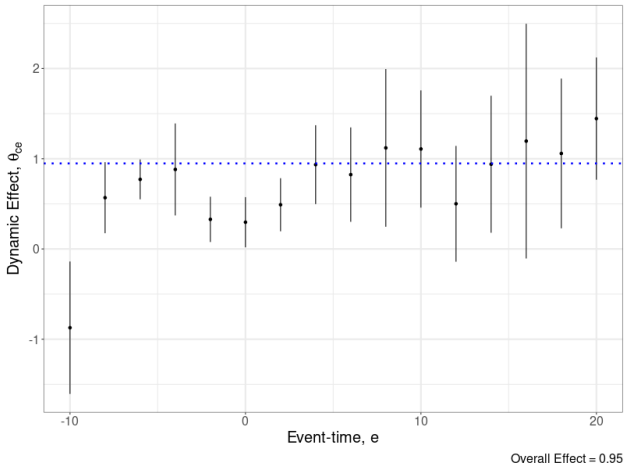
We provide a second example of parallel trends. Spain is in a different bio-geographical region and climate zone from Poland. It is also less densely vegetated. In Figure A.5, the treatment effect for the base period -2 is a precise 0, visually confirming that we have obtained conditional parallel trends for this base period. We see that up to 20 years after protection, all treatment effects are less

Table A.8: Estimated average treatment effect on vegetation greenness for each country, aggregated over cohort and event time

	ATT Estimate (SD)	Non-missing, matched grid cells
CSDR	0.08 (0.01)	81,728,332
MTWFE	0.49 (0.30)	81,728,332
TWFE	0.45 (0.29)	81,728,332
Albania	−2.29 (0.84)	633
Austria	−0.12 (0.03)	728,440
Bosnia	1.89 (0.68)	3,448
Bulgaria	−0.22 (0.02)	1,103,802
Croatia	0.65 (0.07)	522,525
Cyprus	−0.75 (0.14)	21,195
Czech	−0.38 (0.04)	750,218
Denmark	−1.66 (0.27)	401,326
Estonia	0.41 (0.02)	326,705
Finland	0.20 (0.53)	4,530,486
France	0.22 (0.21)	8,248,952
Germany	0.21 (0.03)	17,605,764
Greece	0.44 (0.18)	2,814,637
Hungary	0.35 (0.02)	927,980
Ireland	−0.40 (0.14)	342,171
Italy	0.22 (0.05)	3,165,318
Latvia	1.19 (0.08)	659,158
Lithuania	0.76 (0.03)	696,176
Luxembourg	0.26 (0.06)	19,817
Macedonia	0.17 (0.80)	1,173
Montenegro	0.04 (0.08)	140,316
Netherlands	−0.09 (0.89)	356,550
Poland	0.95 (0.37)	13,679,208
Portugal	−1.73 (1.40)	588,064
Romania	0.08 (0.06)	2,443,662
Serbia	0.71 (0.28)	3,241
Slovakia	0.15 (0.03)	494,283
Slovenia	0.43 (0.03)	196,302
Spain	−0.52 (0.04)	8,441,838
Sweden	−0.12 (0.12)	9,503,885
Switzerland	0.04 (0.24)	3,547
UK	−0.05 (0.19)	3,007,512

NOTES: Average treatment effect θ_c of conservation on vegetation greenness (an index varying between 0 and 100) in Equation (A.1) estimated within each country in the European Union using data from 1985-2019 on a biannual basis. Top three rows report the Callaway & Sant'anna, Doubly robust estimator (CSDR), Matched Two-way Fixed Effects (MTWFE), and Two-way Fixed Effects (TWFE, no matching), respectively. Observations are at a 300 meter resolution and are restricted to those plots which had a greenness value above 40 at some point in the sample period. To ensure adequate pre-period variation, treated units are limited to those units protected in or after 1991. Reported p-values test whether $\theta_c > 0$ relative to the null $\theta_c \leq 0$. n represents the number of unique grid cells were identified in the borders with non-missing matching variables across foundation years between 1991 and 2019.

Figure A.4: Plot of dynamic treatment effects on vegetation greenness for Poland



NOTES: Plots estimates and confidence bands for θ_{ce} for Poland. Poland-wide treatment effect was 0.95(0.37), indicated by the dashed blue line. Confidence bands are based on a bootstrapping procedure discussed in the Appendix B. Plot window trimmed to $[-10, 20]$ for visual clarity.

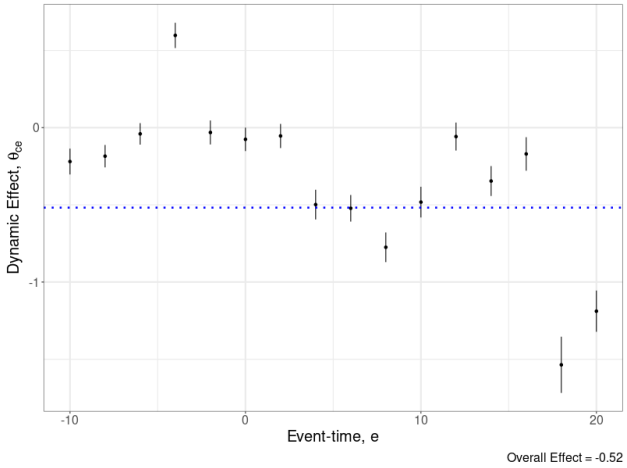
than 2 in absolute value. Note that both this warmer and sunnier climate and the densely-forested Poland demonstrate similar conclusions: very little impact on greenness up to 20 years after the date of land protection. This long time horizon shows neither a change in greenness nor evidence of a trend in greenness which might indicate more gradual vegetation growth. If anything, treatment effects are negative.

The overall EU effects, shown in the main text in Figure 3 (left panel), are presented numerically here in Table A.9. There is no meaningful trend in EU-wide dynamic treatment effects, rejecting that protection has led to long-term vegetation (re-)growth for the average protected area. We trim the plot window to $[-10, 20]$ for visual clarity as standard errors grow large outside this window.

We also report the overall EU cohort-level effects in Figure 3 (right panel) as Table A.10 here. There are 15 cohorts, each with several million observations contributing to estimation (the land area of the smallest cohort is about 129,000 sq. km.). The treatment effects again center around 0, with no discernible trend in treatment effects moving from early- to late-treated cohorts. Larger treatment effects generally appear in the later cohorts, where there are fewer post-treatment periods (e.g., we only observe one period of treatment for the 2019 cohort, which could be confounding treatment with a number of unobservables specific to 2019).

Finally, Table A.11 indicates summary statistics for the cohort-level effects aggregated to the country level in our sample (rather than at the EU level in Table A.10). We use these country-level aggregations to test for selection of protected areas over time (and to compare any potential selection across regimes). The first column indicates the number of cohorts for which treatment effects are calculated in the range 1991-2019. The second column reports a trendline from regressing θ_{cg}, the country-level analog of Equation (A.4), on cohorts $g - 1991$ (we difference out 1991 so changes are

Figure A.5: Plot of dynamic treatment effects on vegetation greenness for Spain



NOTES: Plots estimates and confidence bands for θ_{ce} for Spain. Spain-wide treatment effect was $-0.52(0.04)$, indicated by the dashed blue line. Confidence bands are based on a bootstrapping procedure discussed in the estimation appendix B. Plot window trimmed to $[-10, 20]$ for visual clarity.

Table A.9: Dynamic treatment effects on vegetation greenness across the EU

Event time	Mean (Std. dev.)	Event time	Mean (Std. dev.)
-34	0.00 (0.00)	-2	0.07 (0.01)
-32	0.38 (0.15)	0	−0.08 (0.02)
-30	−0.01 (0.09)	2	0.23 (0.02)
-28	−0.47 (0.05)	4	0.21 (0.03)
-26	0.15 (0.04)	6	−0.12 (0.04)
-24	−0.15 (0.04)	8	0.18 (0.03)
-22	0.29 (0.03)	10	0.14 (0.09)
-20	0.24 (0.03)	12	0.24 (0.10)
-18	0.32 (0.02)	14	0.22 (0.10)
-16	0.12 (0.03)	16	0.13 (0.07)
-14	0.05 (0.04)	18	0.42 (0.09)
-12	0.32 (0.04)	20	−0.01 (0.05)
-10	0.23 (0.03)	22	0.41 (0.02)
-8	0.39 (0.03)	24	−0.47 (0.01)
-6	0.25 (0.02)	26	−0.30 (0.02)
-4	0.22 (0.03)	28	0.43 (0.00)

NOTES: Dynamic treatment effect θ_t^{EU} of conservation on vegetation greenness (an index varying between 0 and 100) estimated for the entire European Union. Estimates are aggregated from country-level estimates of $\theta_c(g,t)$ as detailed in Appendix B. Underlying dataset spans 1985-2019 on a biannual basis. Observations are at a 300 meter resolution and are restricted to those plots which had a greenness value above 40 at some point in the sample period. The earliest included foundation year period is 1991, meaning the latest event-time in the sample is 28. Similarly, the last valid foundation year is 2019.

Table A.10: Cohort-level treatment effects on vegetation greenness across the EU

	Mean (Std. dev.)	N
1991	0.15 (0.08)	8,346,043
1993	0.07 (0.09)	6,399,360
1995	0.25 (0.10)	6,821,430
1997	−0.02 (0.11)	6,508,460
1999	0.22 (0.17)	7,505,035
2001	0.09 (0.05)	7,110,196
2003	0.13 (0.11)	4,309,810
2005	−0.05 (0.08)	6,484,608
2007	0.30 (0.06)	3,249,503
2009	0.15 (0.07)	3,352,802
2011	0.01 (0.21)	1,825,835
2013	0.06 (0.06)	1,442,024
2015	0.12 (0.14)	3,251,702
2017	0.41 (0.07)	3,589,112
2019	−0.69 (0.09)	4,456,613

NOTES: Cohort-level treatment effect θ_{cg}^{EU} of conservation on vegetation greenness (an index varying between 0 and 100) estimated for the entire European Union. Estimates are aggregated from country-level estimates of $\theta_c(g,t)$ as detailed in Appendix B. We sum across all available event-times. Underlying dataset spans 1985-2019 on a biannual basis. Observations are at a 300 meter resolution and are restricted to those plots which had a greenness value above 40 at some point in the sample period. Treated units are excluded if they had a foundation year earlier than 1991 to ensure at least 2 periods of parallel trends (3 data observations: 1985, 1987, and 1989).

interpreted as the effect of being treated 2 years later). Trends are economically small, suggesting that the change in treatment effect across successive cohorts varies by as little as 0.01% of maximum NDVI. The largest trends are in Ireland (0.51) and Denmark (-0.29). Ireland's trend suggests that on average, plots treated in the last cohorts in 2019 experienced a treatment effect 7.5 units higher than those treated in 1991—though it is estimated over 6 cohorts only. Denmark, on the other hand, suggests that the same difference in treatment timing results in treatment effects 4 units lower. The third column constitutes a less parametric approach to estimating trends. These "split-difference" estimators compare average treatment effects θ_{cg} in the back half of the study period to those in the first half. The estimators find similar results with less extrapolation involved. Overall, these results demonstrate that the selection of protected areas does not appear to manifest in significant trends in cohort-level effects of land protection on greenness.

C.4 Treatment effect estimates: nightlights

Nightlight outcomes are measured on a luminosity scale ranging from 0 to 68. Across the EU in 2021, the unconditional average luminosity is 9.9. However, 30% of our sample has 0 luminosity (at a 1-kilometer grid cell resolution. A 0 requires no indoor nor outdoor electricity: these 0s correspond to clear "undisturbed" land area). Conditional on having positive luminosity, the average luminosity is 14.3. Countries vary in average nightlights from very bright in Belgium (conditional on a positive luminosity, mean of 30.5) and the Netherlands (26.2) to relatively less so in Ireland (7.9) or Bulgaria (9.3). Classically, long-run changes in nightlights have been associated with long-run GDP growth. When regressing GDP growth on changes in nightlights, Henderson, Storeygard, and Weil (2012) report that a 1% change in nightlights is associated with a 0.3% increase in GDP growth over a 20-year panel.

Here, we provide tables that summarize the treatment effects depicted in Figure 4. We identify a grid cell as treated if it is at least 50% covered by a protected area. Table A.12 indicates the average treatment effects from the dynamic aggregation θ_t^{EU}, corresponding to the left panel of Figure 4. Standard errors indicate bootstrapped standard errors on treatment effects. Our estimates are statistically indistinguishable from 0 until event-time 18, at which point they indicate a 0.30 point drop in luminosity. This treatment effect is inconsistent over the remaining periods, rising to -0.39 in event time 20 dropping to -0.14 by the last event time, suggesting that if there is a drop in nightlights, it is (1) not sustained at its initial levels and (2) occurs 20 years after treatment, making it difficult to attribute to protection alone.

Table A.13 indicates the average treatment effects from the cohort-level aggregation θ_g^{EU}, corresponding to the right panel of Figure 4. As with the previous table, we find little evidence of treatment effects, which are significantly more negative than -0.5 with the exception of the 2005 cohort and the 2013 cohort. These later cohorts appear to have some larger drop in their nightlights. However, this drop in nightlights is not a persistent feature of the data. Compositionally, we find no heterogeneity in treatment effects across any observables in later exercises, making these

Table A.11: Testing for differences in treatment effects across cohorts

	Number of Cohorts	Trend: Estimate (Std. dev.)	Split difference: Estimate (Std. dev.)
Austria	15	0.02 (0.0027)	0.20 (0.74)
Bulgaria	12	0.14 (0.0043)	0.98 (1.13)
Croatia	10	−0.01 (0.0010)	−0.09 (0.59)
Czech	15	0.03 (0.0003)	0.11 (0.48)
Denmark	11	−0.29 (0.0059)	−4.82 (1.43)
Estonia	14	0.03 (0.0002)	0.61 (0.31)
Finland	15	−0.02 (0.0001)	−0.27 (0.21)
France	15	0.01 (0.0004)	0.16 (0.45)
Germany	15	−0.02 (0.0002)	−0.38 (0.40)
Greece	11	0.01 (0.0006)	−0.15 (0.50)
Hungary	14	−0.06 (0.0018)	−1.07 (0.70)
Ireland	6	0.51 (0.0946)	4.24 (4.76)
Italy	10	0.03 (0.0007)	0.25 (0.38)
Latvia	7	−0.09 (0.0037)	0.43 (1.19)
Lithuania	5	−0.10 (0.0009)	−0.78 (0.88)
Luxembourg	6	−0.05 (0.0008)	−0.68 (0.62)
Montenegro	4	0.02 (0.0004)	0.00 (0.22)
Poland	15	0.00 (0.0009)	−0.12 (0.37)
Portugal	11	−0.06 (0.0018)	−1.45 (1.20)
Romania	11	0.00 (0.0002)	−0.19 (0.28)
Slovakia	13	−0.05 (0.0004)	−0.61 (0.43)
Slovenia	14	−0.01 (0.0006)	0.19 (0.42)
Spain	15	0.01 (0.0008)	0.00 (0.50)
Sweden	15	0.02 (0.0003)	0.56 (0.30)
Switzerland	10	−0.03 (0.0010)	−0.21 (0.63)
UK	15	−0.03 (0.0005)	0.20 (0.72)

NOTES: Table reports the number of cohorts in each country (note that the maximum number here is 15 as our sample contains biannual cohorts from 1991-2019). The trend estimates a linear regression of the cohort level treatment effect θ_{cg} against the cohort itself $g - 1991$. The split difference estimates a difference of means in θ_{cg} in the later half of the treated cohorts relative to the first half of the treated cohorts.

Table A.12: Dynamic treatment effects on nightlight luminosity across the EU

Event time	Mean (Std. dev.)	Event time	Mean (Std. dev.)
-26	0.21 (0.09)	1	−0.05 (0.02)
-25	−0.08 (0.04)	2	−0.01 (0.02)
-24	−0.20 (0.04)	3	−0.06 (0.02)
-23	−0.11 (0.03)	4	−0.09 (0.03)
-22	0.14 (0.03)	5	−0.10 (0.03)
-21	0.07 (0.03)	6	−0.16 (0.03)
-20	−0.04 (0.03)	7	−0.20 (0.02)
-19	0.04 (0.03)	8	−0.15 (0.03)
-18	0.06 (0.02)	9	−0.06 (0.03)
-17	−0.03 (0.03)	10	−0.10 (0.03)
-16	−0.03 (0.03)	11	−0.01 (0.03)
-15	−0.03 (0.03)	12	−0.06 (0.03)
-14	−0.18 (0.03)	13	−0.01 (0.03)
-13	0.02 (0.03)	14	0.00 (0.03)
-12	0.00 (0.02)	15	0.08 (0.04)
-11	−0.08 (0.02)	16	0.18 (0.04)
-10	−0.11 (0.02)	17	0.02 (0.04)
-9	0.05 (0.02)	18	−0.30 (0.05)
-8	−0.04 (0.02)	19	−0.29 (0.05)
-7	−0.07 (0.02)	20	−0.39 (0.06)
-6	−0.09 (0.02)	21	−0.33 (0.06)
-5	0.04 (0.02)	22	−0.15 (0.08)
-4	−0.13 (0.02)	23	−0.41 (0.08)
-3	−0.08 (0.02)	24	−0.16 (0.09)
-2	0.01 (0.02)	25	−0.26 (0.14)
-1	−0.07 (0.02)	26	−0.14 (0.22)
0	−0.01 (0.02)		

NOTES: Dynamic treatment effect θ_t^{EU} of conservation on nightlight luminosity (an index varying between 0 and 68) estimated for the entire European Union. Treatment effect defined at the grid-cell observation level, 300 by 300 meters. Treatment requires grid cells overlap with a protected area over at least 50% of their area. Estimator is detailed in B. Underlying dataset spans 1992-2019 on an annual basis. The earliest included foundation year period is 1993, meaning the latest event-time in the sample is 26.

two cases outliers. Indeed, in this table, one is nearly as likely to find a positive effect on nightlights, a proxy for GDP, as a negative treatment effect. There is no discernible pattern in selection that indicates any systematic variation in these nightlights effects, either.

Table A.13: Cohort-level treatment effects on nightlight luminosity across the EU

Year	Mean (Std. dev.)
1993	0.08 (0.15)
1994	0.08 (0.05)
1995	−0.34 (0.06)
1996	0.13 (0.07)
1997	−0.25 (0.06)
1998	−0.47 (0.08)
1999	−0.58 (0.10)
2000	0.19 (0.07)
2001	−0.25 (0.11)
2002	0.71 (0.08)
2003	0.12 (0.08)
2004	0.00 (0.06)
2005	−1.78 (0.19)
2006	−0.30 (0.10)
2007	0.81 (0.14)
2008	−0.39 (0.25)
2009	0.45 (0.10)
2010	−0.05 (0.05)
2011	0.12 (0.08)
2012	−0.04 (0.18)
2013	−1.01 (0.19)
2014	0.34 (0.07)
2015	−0.36 (0.13)
2016	−0.07 (0.06)
2017	−0.03 (0.11)
2018	−0.56 (0.07)
2019	−0.22 (0.11)

NOTES: Cohort-level treatment effect θ_g^{EU} of conservation on nightlight luminosity (an index varying between 0 and 68) estimated for the entire European Union. Treatment effect defined at the grid-cell observation level, 300 by 300 meters. Treatment requires grid cells overlap with a protected area over at least 50% of their area. Estimator is detailed in B. Underlying dataset spans 1992-2019 on an annual basis. The earliest included foundation year period is 1993.

C.5 Heterogeneous treatment effects

Here we describe in more detail the heterogeneous treatment effects obtained via the random forest method of Wager and Athey (2018). We present results estimated on a random sample of the data as described in Appendix B. In the sample, the average treatment effect is -0.51 (0.05).

The first result is a test of heterogeneity in treatment effects. We use the test in Wager and Athey (2018), which amounts to a random forest implementation of Chernozhukov et al. (2017). The results are shown in Table A.14. The test reports two coefficients in a regression. Observed greenness Y are projected onto the average treatment effect estimated by the random forest and the conditional average treatment effect estimated by the random forest:

$$Y_{it} = \beta_0 ATE_{it} + \beta_1 (CATE_{it} - ATE_{it}) + \epsilon_{it}$$

Intuitively, the true average treatment effect should contribute a coefficient of exactly 1 as a one-unit increase in the average treatment effect drives a 1 unit increase in expected counterfactual outcomes. Thus, the test for the coefficient on the ATE is a two-sided test for whether the coefficient is statistically different from 1. We find an estimate of 1.08 (0.10). Similarly, the coefficient on the CATE should be at least 1: if the CATE changes by 1 unit, we should expect the outcome itself to change by at least this much if the CATE is meaningful. The test on the CATE is one-sided, determining if the coefficient is greater than 1: rejection implies that the CATE the random forest has predicted is not driving deviations from the average treatment effect. Our results indicate our measured CATE is accurate, with a coefficient of 25 (2.49). Thus, our random forest has picked up on statistically significant deviations from the average treatment effect.

Table A.14: Test of the random forest model calibration

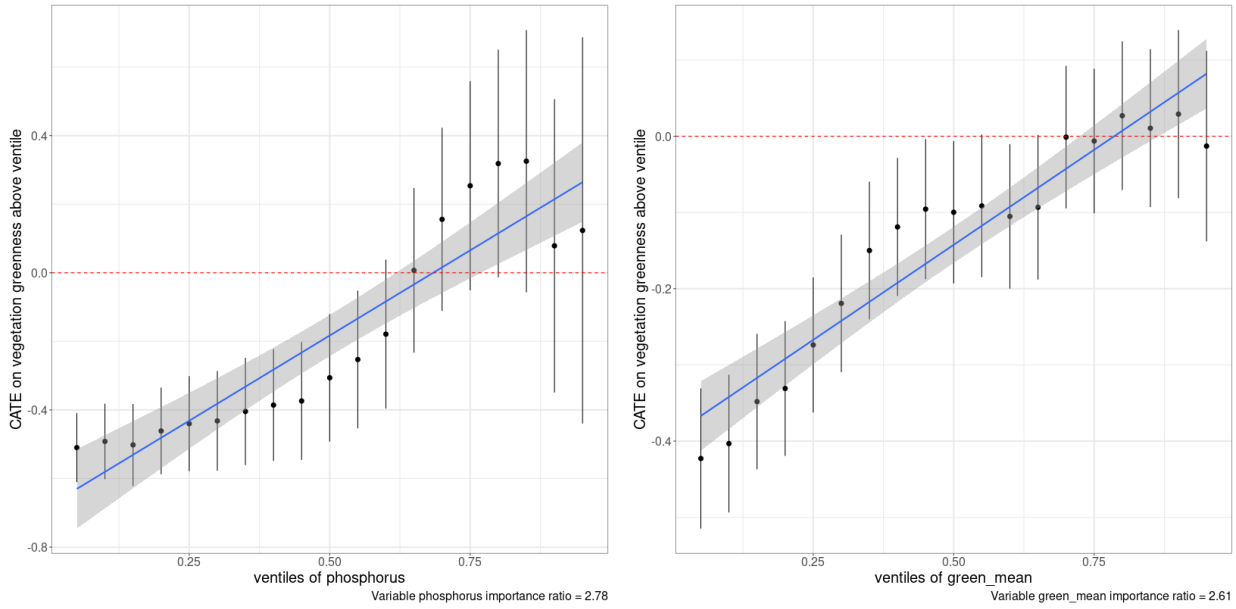
	Coefficient	Standard error
Mean forest prediction	1.08	(0.10)
Differential forest prediction	25.32	(2.49)

NOTES: Tests calibration of the Wager and Athey (2018) heterogeneous treatment effects model. The causal random forest is trained with 500 trees and a minimum node size of 10,000 using a 5% sample fraction rate. The first coefficient describes whether the model captures the mean forest prediction. As it is statistically indistinguishable from 1, the random forest appears to be fit well. The second coefficient describes whether the model is able to find heterogeneity in calibrated treatment effects. The coefficient is robustly greater than 1, confirming that we have found salient heterogeneity.

Despite the statistically significant heterogeneity in the data, treatment effects vary little in economic terms. The top quartile of treatment effects is noisily 0.71 (1.21) NDVI units above the lowest quartile of treatment effects. To further test the economic significance of the CATE measures, we plot conditional average treatment effects along the distribution of selected control variables X in Figure A.6.

We are interested, in particular, if variables associated with high-quality agricultural land may change the predicted CATE. If we see, for example, that high soil phosphorus content, indicating soil that is highly agriculturally productive, has a higher treatment effect, it suggests protection causes plots more suitable for agriculture to green more than plots less suitable for agriculture. Such a result is consistent with a subset of land protection having positive treatment effects due

Figure A.6: Plot of distribution of conditional average treatment effect for phosphorus content in soil (left) and mean pre-period greenness (right)



NOTES: The dependent variable on these plots indicates the quantile of a particular control, $q(x)$. The plotted treatment effect is a conditional average treatment effect (CATE) given by $\mathbb{E}[\tau|X > q(x)]$. Thus, each point corresponds to a smaller sample of the data. In the first bin, we see the ATE as this is $\mathbb{E}[\tau|X \geq \min(x)]$. In the right-most bins, we plot the CATE at the largest values of the controls. Confidence bands are 95% confidence intervals constructed through a doubly-robust procedure. Figures are overlaid with a trend line through the estimates and associated confidence band. Figures report variable importance ratio as reported by random forest: both have very high relative importance, meaning they are highly informative of changes in CATE.

to a valuable counterfactual land value. We present the phosphorus CATE plot in Figure A.6 (left panel). While there does appear to be a significant positive trend in the treatment effect, the top 25% of the data has a treatment effect barely discernible from 0 and less than half of an NDVI point on a scale of 100 units. This result is inconsistent with a pattern of high agricultural value soil returning to natural use.

While a number of our control variables do display trends, we find that magnitudes are small for any deviations from a 0 treatment effect. For example, the pre-period mean greenness in Figure A.6 (right panel) exhibits a statistically significant trend. Plots in the top 5% of the sample with respect to their pre-period greenness have a precise conditional average treatment effect of 0, which is statistically discernible from the sample-level ATE of -0.5. Yet, this result can again be interpreted as a "precise 0". Overall, we conclude from this exercise that our high-powered dataset has returned another set of heterogeneous treatment effects that are effectively zero.

D Robustness checks

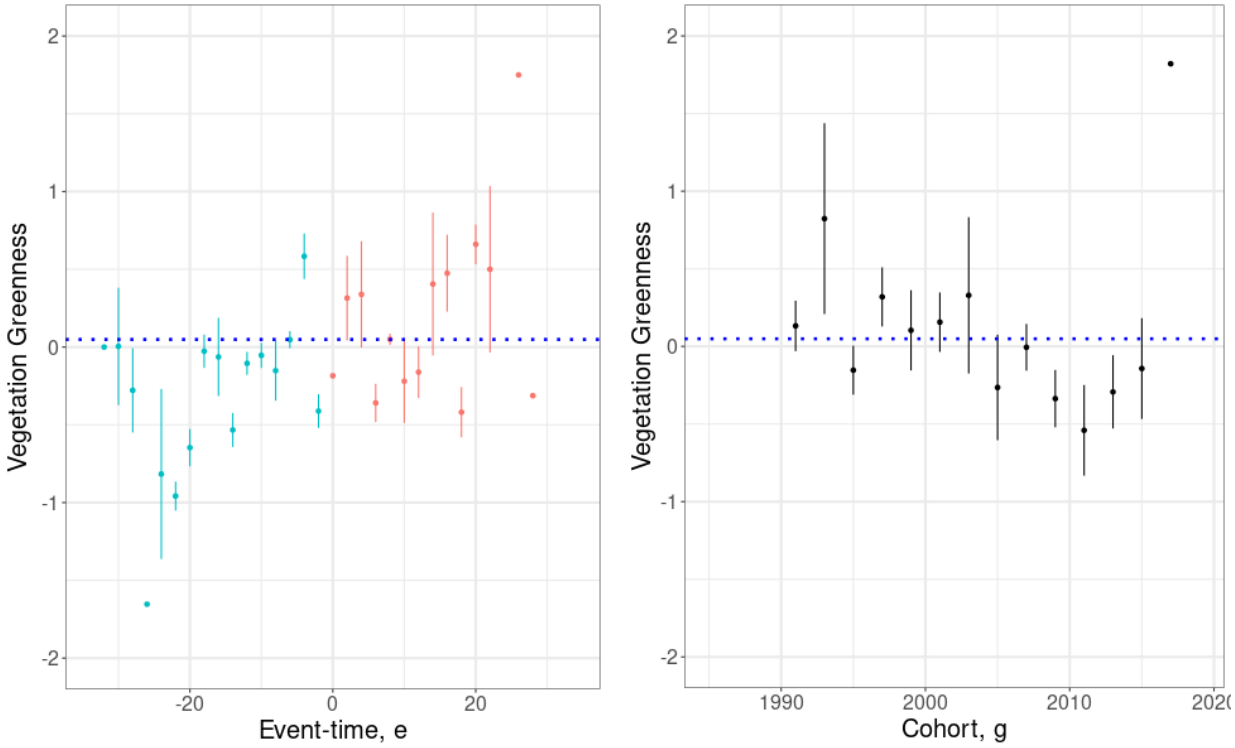
In this section we discuss various robustness checks of the doubly-robust difference-in-differences estimator discussed in Appendix B.

Functional form: first differences. Our main specification considers differences in levels of greenness at the time of protection. We test whether our null result is driven by this functional form assumption by testing the same specification in first differences of the outcome variable. We apply the exact same matching function with the exception of matching on pre-period average greenness in levels and in trends. The aggregate treatment effect θ^{EU} is 0.05 (0.02).

We report the EU-wide aggregated treatment effects in Figure A.7. The left panel illustrates the dynamic treatment effects on changes in vegetation greenness, θ_t^{EU}. We can see weak evidence of a positive trend in the slope of greenness for the full range of event times. There is no evidence of a kink in greenness at the time of treatment: slopes do not change around protection. While treatment effects in levels may take time to appear, we should see an immediate kink in first differences if there is indeed vegetation growth occurring post-treatment which was not present in the counterfactual. In other words, we should see an immediate increase in the slope of vegetation greenness indicating a divergence in vegetation growth rates between treated and untreated units. However, any positive treatment effects are equivalently supported by negative treatment effects in the post-treatment portion of the figure. The right panel of Figure A.7 illustrates the cohort-level treatment effects. Aside from a single positive effect for the 1993 cohort, we reject any trends in cohort-level treatment effects.

Collectively, our first-difference results does not suggest that the specification in levels finds a precise 0 purely due to a model misspecification. There is no evidence of a kink in vegetation growth driven by protection.

Figure A.7: Treatment effects on a first difference of vegetation greenness over event time and by cohort

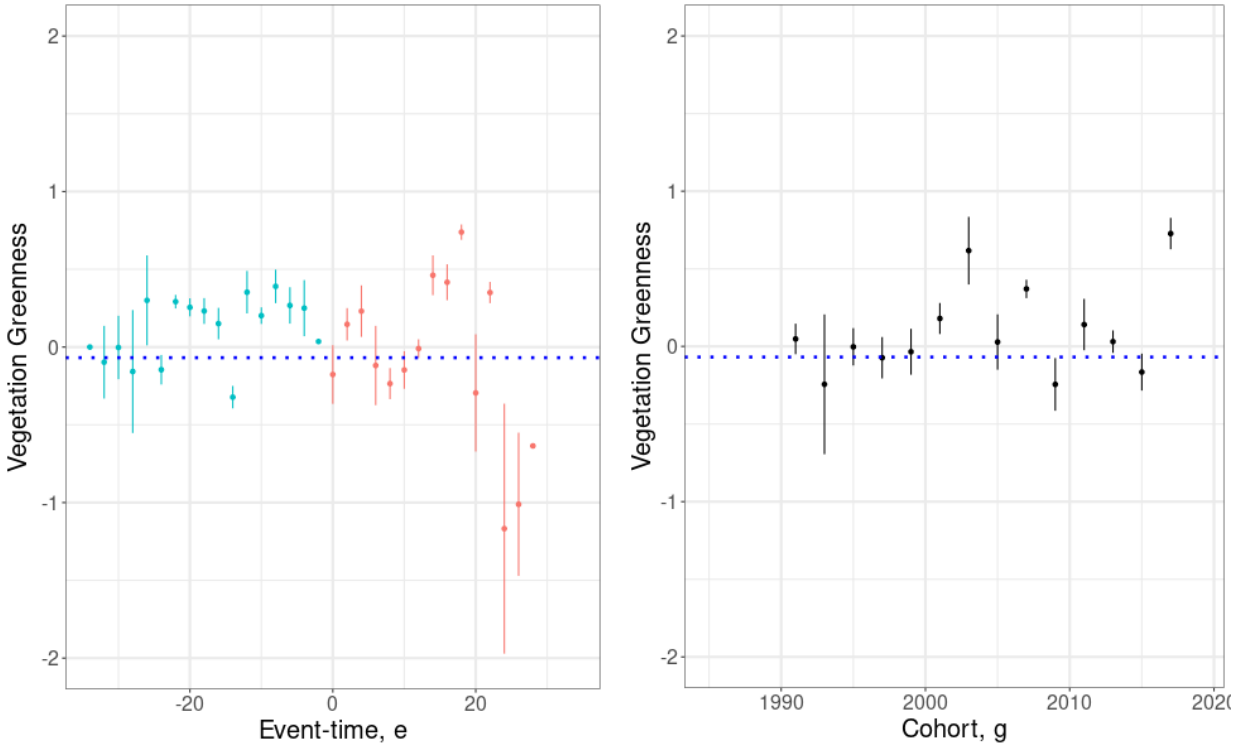


NOTES: Treatment effects on a first difference of greenness aggregated by event-study period θ_t^{EU} in the left panel, and by cohort θ_g^{EU} on the right. Sample includes plots which have greenness above 40 at least once in the sample period with non-missing matching variables. Sample trims first calendar year, 1985, due to first differencing: we match on values from 1987 and 1989 for all cohorts treated on or after 1991. Both panels use the Callaway and Sant'anna doubly robust estimator with bootstrapped confidence bands.

Sample definition: NDVI thresholds. Our main specification trims the total EU to a subset that meets a minimum NDVI threshold of 40 at least once in the period 1985-2019. We motivate this by noting that land below this NDVI threshold is unlikely ever to be forest or dense vegetation. The matching and outcome regression procedures model the suitability of land for vegetation growth. Here, we test the sensitivity of the results to the threshold of NDVI 40 by choosing a larger sample with a more permissive NDVI threshold of 30. Because our heterogeneity analysis (see Appendix B.3) finds no heterogeneity across initial NDVI for land above the NDVI threshold of 40, we are confident that restricting the sample to higher NDVI cutoffs would not change our precise 0.

Our overall average treatment effect for an NDVI threshold of 30 is again a precise 0: $-0.07(0.02)$. The difference between this estimate and our baseline estimate is materially quite small, if precise. The direction of the change in estimates suggests that, if anything, protected land which never attained an NDVI of 40 greened relatively less than land in our main sample.

Figure A.8: Treatment effects on NDVI: expanded sample definition



NOTES: Treatment effects on greenness aggregated by event-study period θ_t^{EU} on the left panel, and by cohort θ_g^{EU} on the right. Sample includes plots which have greenness above 30 (note: in our main specification, this number is 40) at least once in the sample period with non-missing matching variables. Both panels use the Callaway and Sant'anna doubly robust estimator with bootstrapped confidence bands.

When plotting EU-wide dynamic treatment effects θ_t^{EU} in the left panel of Figure A.8, we find little evidence of aggregate greening. We obtain precise balance in the key first pre-period with an

estimated difference of 0.04 (0.01). The overall dynamics are centered at zero for the first 12 years of treatment. After this, we encounter some minimal variation—first positive effects on the scale of 0.5 NDVI points, then a negative jump of -1 NDVI points. We make two observations about these later treatment effects. (1) These late-stage jumps in greenness are driven by a changing sample composition, and (2) the largest evidence for protection-driven greenness is the a positive treatment effect of 0.75 in period 19. Overall, the treatment effects are noisily negative rather than positive, again suggesting that protection does not contribute to greening.

In the right panel of Figure A.8, we plot treatment effects at the cohort-level θ_g^{EU}. These cohort effects largely resemble our baseline specification: there is no discernible trend which would indicate site selection across treated cohorts.

Spatial correlation: spatial first differences. If some of our control units are affected by treatment, our difference-in-differences estimate violates the stable unit treatment value (SUTVA) assumption. A conventional method for testing for such spillovers is a "donut" difference-in-differences design. In such a design, the researcher discards potentially contaminated units nearest a protected area and recomputes treatment effects. It thus calculates a treatment effect based only on non-local variation in greenness. We find the donut approach is prohibitively expensive from a computational perspective: it requires calculating an individual buffer for each of our 100,000 treated areas, many of which may have overlapping buffers and therefore no plausible donut controls.

Instead, we use a method of spatial first differences (SFD). This methodology takes a first difference of the data along a given spatial axis, thus creating comparisons between areas that are close by in space. We difference observations along the x-dimension so that areas with the same latitude are differenced against their neighbors. Introducing the index $i = (x, y)$ to identify a grid cell by the coordinates of the centroid of that grid cell, the outcome variable is:

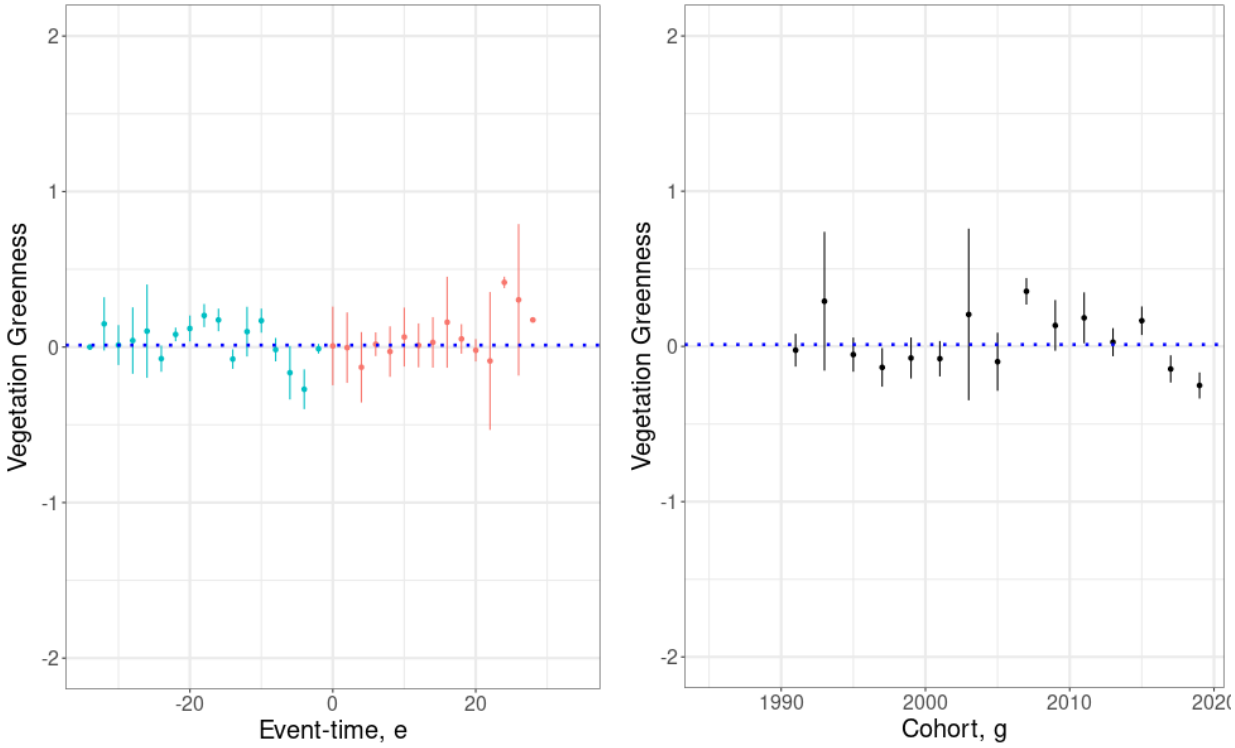
$$\Delta Y_{xyt}^{LAT} = Y_{xyt} - Y_{x-1,yt}$$

In principle, we could re-estimate SFD for many axes, such as a vertical or diagonal axis. When we estimate treatment effects, the SFD method estimates the difference-in-differences on the spatially differenced outcome variable. Intuitively, a difference-in-differences estimator with SFD data captures the relative change in $\Delta_{LAT} Y_{xyt}$ in treated and control units. Thus, rather than computing a direct comparison of greenness between treated units and distant controls through donuts, we focus on estimating the greenness differences between treated units and their nearest neighbors.

When calculating the spatial first difference across the x dimension, we obtain an EU-wide treatment effect of 0.01 (0.02). Figure A.9 plots the event-study (left panel) and cohort-level aggregations (right panel). The left panel demonstrates an exact 0 in the post-treatment periods. There is a slight dip in the pre-treatment coefficients, suggesting that treated areas were greening slightly less relative to neighbors than a comparable control unit in the lead-up to treatment, with

a magnitude of around −0.2 in the second-to-last pre-treatment period. We then attain a precise balance on the key pre-treatment period. Cohort effects show no trend through 2015, with a slightly lower treatment effect in the 2017 and 2019 cohorts. Overall, the SFD estimates an even more precise 0 than our main specification.

Figure A.9: Treatment effects on NDVI: spatial first differences



NOTES: Treatment effects on greenness aggregated by event-study period θ_t^{EU} on the left panel, and by cohort θ_g^{EU} on the right. Outcomes are spatially differenced in the x direction. Sample includes plots which have greenness above 40 at least once in the sample period with non-missing matching variables. Both panels use the Callaway and Sant'anna doubly robust estimator with bootstrapped confidence bands. Dotted blue lines indicate the EU-wide ATE of 0.01.

E Other Outcome Variables

E.1 Biodiversity outcomes from species counts

BioTIME data consist of a panel of animal species and vegetation flora biomass studies. Studies enter and exit the panel as they are conducted by biologists and ecologists. Records consist of a year, a species identifier, a study identifier, and either a count of species abundance or a measure of vegetation biomass. When focusing on the landmass of the European Union, there are a total of 58 studies, most of which focus on animal counts. We restrict our analysis in this section to species counts of animals as a measure of biodiversity.

Of the 58 BioTIME studies in the EU, 51 are within 25 kilometers of a protected area listed in

the CDDA system. Notably, there are only 7 BioTIME studies further than 25km from a CDDA. These seven studies are too few to serve as a valid control group to establish causal effects as in our core analysis of greenness and nightlights. Instead, we construct event-study estimators of the impact of nearby CDDA openings on measured abundance. As a result, rather than leveraging variation relative to control units, we only look at within-study variation to determine whether structural breaks appear around the foundation of protected areas. The potential selection of these few study sites and the lack of a valid control group in the BioTIME data limits the external validity of these results. We thus present our specification as descriptive evidence rather than causal designs.

Our econometric specification considers a study i, species s, and year t. We construct a buffer of distance b kilometers around the study area, the neighborhood N_i^b. We assign treatment of the study area according to the minimum foundation year g of overlapping CDDAs. That is, the treatment indicator is $D_{it}^b = \mathbb{1}[t \geq \min_{N_i^b} g]$ with event time $e^b = t - \min_{N_i^b} g$. Then, the event-study design amounts to the following regression, where λ is used to denote a fixed effect:

$$y_{ist} = \sum_{e^b=-10}^{10} \beta_{e^b}^b D_{it}^b + \lambda_i + \lambda_s + \epsilon_{ist} \tag{A.5}$$

In Figure A.10 we plot the event-study coefficients $\beta_{e^b}^b$ for two buffer distances: 2.5 km and 5 km. We estimate the regression in logs. Event-study coefficients are indistinguishable from 0 in the immediate time frame around the first foundation year in the vicinity. The average effect across post-periods was 0.03 (0.05) and 0.10 (0.06) at the 2.5 and 5 km buffers, respectively. Overall, we cannot reject the null of no structural break in these settings.

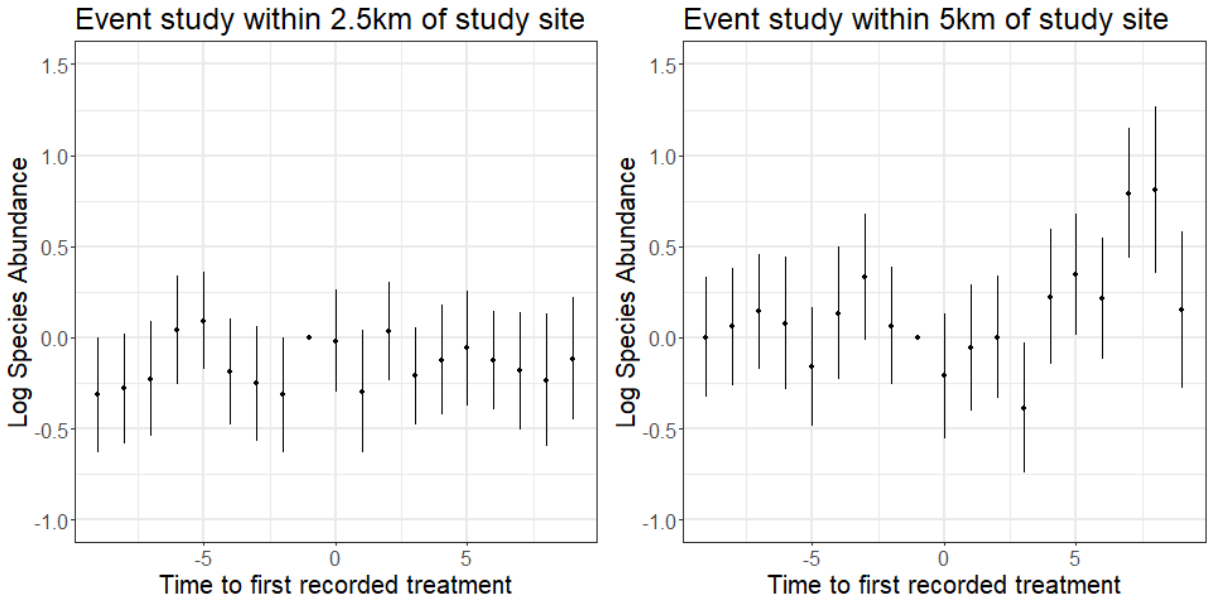
The BioTIME data is spatially concentrated in a few areas within Europe. This lack of spatial variation means many sites are close to one another, and close to many potential CDDAs. When expanding the treatment buffer from 2.5 km to 5 km, we move from 35 to 39 treated studies out of the original pool of 58. The jump in species counts represented is much larger, from 750 to 2,422 total species tabulated. Hence, there is a large density of species studies within the 5 km boundary. This indicates the biodiversity data lacks spatial breadth, making valid inference difficult.

E.2 Discrete land use data

The HILDA dataset provides over 100 years of land-cover maps at a 1 kilometer grid cell resolution. As we discuss in Section 3, it has severe limitations as a main outcome variable for measurement of biodiversity. However, HILDA does provide descriptive insights into long-run land-cover trends since 1900. HILDA omits a subset of countries which are included in our analysis sample. These are Albania, North Macedonia, Montenegro, Croatia, Bosnia and Herzegovina, Serbia, Norway, and Iceland. We omit HILDA reporting on several smaller nations which are not in our main analysis sample: Andorra, Monaco, Jersey, Guernsey, Isle of Man, and Faroes.

HILDA categorizes land into one of six discrete classes: cropland, forest, grassland, other (such

Figure A.10: Event-study coefficients measuring species abundance with respect to nearby protected area foundation

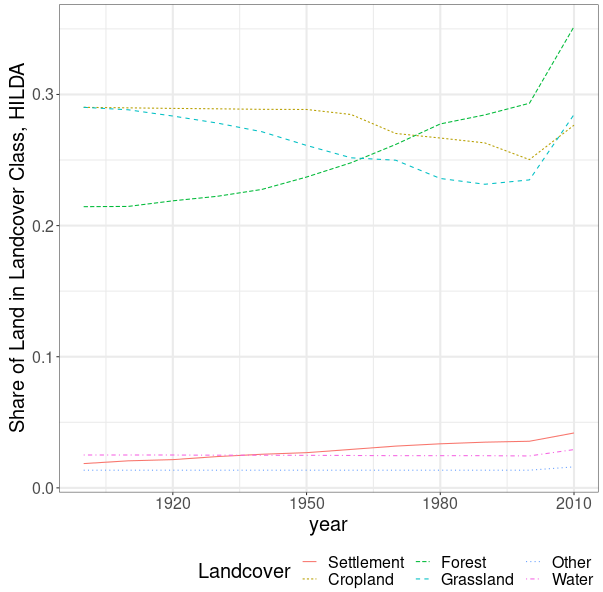


as mountains or barren land surface), settlement, and water. In Figure A.11 we plot the share of each land use in the HILDA data. The data is at a decadal frequency since 1900. The data confirms the century-long growth in vegetation across the EU. HILDA reports a forest share in the EU of just over 20% in 1900 and a 35% forest share in 2010.

In Table A.15, we report a transition matrix over the full breadth of the HILDA data. Looking at the diagonal entries, it is clear that cropland and grassland experienced the greatest land-cover shifts in percentage terms: 30% and 40% point of the 1900 landcover has transitioned to other land uses. The chief beneficiary of both appears to have been forest, though there is non-negligible transition between grassland and cropland. This latter transition can be both a true return of cropland to natural use (or vice versa), or can encompass measurement error as cropland and grassland are less readily discerned by classifiers, or finally can also indicate conversion between crop and pasture (which is not a dimension of biodiversity we are particularly interested in). However, overall, Europe is clearly greening: 10% of cropland in 1900 appears to be forest in 2010, and 27% of grassland in 1900 is forest by 2010. Settled areas also transition to forest—some subset of land returned to natural use.

We also report a cross-tabulation of land-cover data in the final year of the HILDA data, 2010, against one treatment definition of protection in Table A.16. Because HILDA has larger grid cell sizes (1 km) than in our core dataset, we report a more stringent treatment definition which identifies a grid cell as treated if it is at least 50% covered by a protected area. The cross-tabulation is not very sensitive to this decision: raising the proportion required to be considered protected decreases the total number of protected areas but does not affect the composition of those protected

Figure A.11: Land share in the EU by decade and land use category



NOTES: HILDA land classes report dominant classified land use in a 1 km grid. We omit land which is classified as NA or missing from this sample.

Table A.15: Discrete land use transitions for the entirety of Europe between 1900-2010

		Land-use transition probabilities					
Land use in 1900 / 2010		Cropland	Forest	Grassland	Other	Settlement	Water
Cropland	% row	68.5	9.6	17.7	0.0	4.1	0.0
Forest	% row	0.8	89.7	8.8	0.0	0.8	0.0
Grassland	% row	11.7	27.2	58.8	0.0	2.3	0.0
Other	% row	0.0	0.0	0.0	100.0	0.0	0.0
Settlement	% row	7.4	5.0	6.8	0.0	80.8	0.0
Water	% row	1.2	0.2	0.9	0.0	0.4	97.2
		Land-use shares					
Land use in 2010		Cropland	Forest	Grassland	Other	Settlement	Water
Total (2010)	% row	27.7	35.2	28.4	1.6	4.2	2.9

NOTES: Table reports land use transitions relative to base year of 1900 in 2010. Transitions are defined based on the HILDA landcover data, which classifies land into one of 6 land areas in each decade from 1900 to 2010. We omit a small percentage of areas which are classified as NA values.

Table A.16: Break-down of land use under protection in 2010

landuse		Less than 50% protected	At least 50% protected
Cropland	% row	91.0	9.0
Forest	% row	78.8	21.2
Grassland	% row	80.5	19.5
Other	% row	58.9	41.1
Settlement	% row	91.7	8.3
Water	% row	79.4	20.6

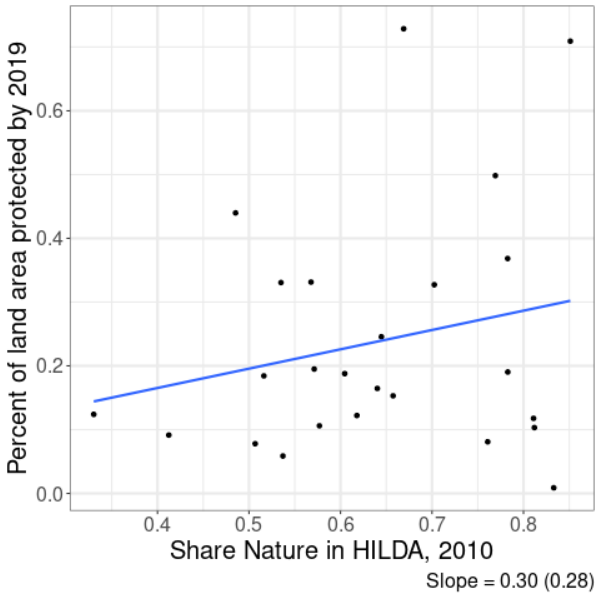
NOTES: We cross-tabulate the land cover reported in the last decade of the HILDA data (2010) according to whether the 1 kilometer grid cell is at least 50% treated at any time before 2023. Column headers indicate ever-treatment status, and row-headers indicate measured land cover in 2010.

areas (results available on request). The table shows that cropland and settlement are far less likely to be protected than forests, grassland, or water. Notably, the "other" category (which includes, for example, mountain ranges that are protected in larger clusters of land) is most likely to have a protected status.

Finally, we explore evidence supporting our hypothesis that protected land is a highly selected sample. Recall that the EU Directive does not provide country-specific quotas. Then, under a selection scheme that minimizes the economic opportunity costs of land set aside for protection, countries with higher land area in natural use are likely the countries with the highest share of land protected—large protected areas are likely remote and to a large extent at minimal risk of development. That is, the supply of natural land would be correlated with actual protection as policy-makers avoid protecting economically valuable cropland or settlement. Figure A.12 illustrates that natural land availability and protection are slightly positively associated. We find some evidence that countries with particularly high nature shares protect more land. A higher natural land share is associated with a higher share of total land protected.

However, we supply this evidence with two caveats. Our definition of the supply of natural land (forest and grassland) is incomplete. Thus, our measure of nature "budget" across countries is limited: we would ideally be able to include mountains, permafrost, wetlands, and other natural land uses that HILDA does not report. In addition, the share of available natural land is endogenous to previous protection choices.

Figure A.12: Natural land supply and protection



NOTES: Observations represent countries. The blue line is the line of best fit. The figure is produced with a 1-kilometer resolution grid. We compute the share of natural land in HILDA's 2010 dataset as the total land area classified as forest or grassland over a country's total land area.

References for Appendix

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics,* Themed Issue: Treatment Effect 1, 225 (2): 200–230.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Iván Fernández-Val. 2017. *Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments.* https://arxiv.org/abs/1712.04802.

Dannenberg, Matthew P., Erika K. Wise, and William K. Smith. 2019. "Reduced Tree Growth in the Semiarid United States Due to Asymmetric Responses to Intensifying Precipitation Extremes." *Science Advances* 5 (10): eaaw0667.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.

Li, Xuecao, Yuyu Zhou, Min Zhao, and Xia Zhao. 2020. "A Harmonized Global Nighttime Light Dataset 1992–2018." *Scientific Data* 7:168.

Pasho, Edmond, J. Julio Camarero, Martín de Luis, and Sergio M. Vicente-Serrano. 2012. "Factors Driving Growth Responses to Drought in Mediterranean Forests." *European Journal of Forest Research* 131:1797–1807.

Peled, E., E. Dutra, P. Viterbo, and A. Angert. 2010. "Technical Note: Comparing and Ranking Soil Drought Indices Performance over Europe, Through Remote-Sensing of Vegetation." *Hydrology and Earth System Sciences* 14 (2): 271–277.

Roy, D.P., V. Kovalskyy, H.K. Zhang, E.F. Vermote, L. Yan, S.S. Kumar, and A. Egorov. 2016. "Characterization of Landsat-7 to Landsat-8 Reflective Wavelength and Normalized Difference Vegetation Index Continuity." *Remote Sensing of Environment* 185:57–70.

Van Oijen, M., J. Balkovi, C. Beer, D. R. Cameron, P. Ciais, W. Cramer, T. Kato, et al. 2014. "Impact of Droughts on the Carbon Cycle in European Vegetation: A Probabilistic Risk Analysis Using Six vegetation Models." *Biogeosciences* 11 (22): 6357–6375.

Vieira, Joana, Cristina Nabais, and Filipe Campelo. 2021. "Extreme Growth Increments Reveal Local and Regional Climatic Signals in Two Pinus pinaster Populations." *Frontiers in Plant Science* 12:658777.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–1242.