An Evaluation of Protected Area Policies in the European Union

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

June 4, 2024

Abstract

The European Union designates 26% of its landmass as a protected area, limiting economic development to favor biodiversity. We use the staggered introduction of protected areas between 1985 and 2020 to study the selection of land for protection and the causal effect of protection on vegetation cover and nightlights. Protection did not affect these outcomes in any meaningful way across four decades, countries, protection cohorts, population density, or land, soil, and climate characteristics. We conclude that European conservation efforts lack ambition because policymakers protect land not threatened by development or choose weak protection levels on lands that face development pressure.

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

JEL codes: Q23, Q24, Q57, R14

*Earle Grupp: Stuart Weitzman School of Design, University of Pennsylvania; Mishra: Wharton School, University of Pennsylvania; Reynaert: Toulouse School of Economics and CEPR; van Benthem: Wharton School, University of Pennsylvania and NBER. We thank Adam Streff and Danial Syed for excellent research assistance. Thanks to seminar participants at the AERE 2023 Summer Conference, EAERE 2023 Summer Conference, Georgia Tech, Imperial College London, LSE/Imperial/King's Workshop in Environmental Economics, SUNY Albany, Toulouse School of Economics, University of Bristol, University of California at Berkeley, University of Cambridge, University of Delaware, University of Gothenburg, University of Mannheim, University of Oxford, 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, the Wharton Dean's Research Fund, 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 and if protection successfully limits economic activity. 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? Does protection impose substantial restrictions on land use?

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.

Advanced economies are of critical importance for meeting global protection targets yet exhibit different land-use dynamics from heavily-studied tropical forest regions, where illegal deforestation and forest fires are rampant (Burgess et al. 2012; Burgess, Costa, and Olken 2023; Balboni, Burgess, and Olken 2023). While the global south experiences deforestation and sees agriculture expanding, Europe and other northern areas afforest and see cropland abandonment Winkler, Fuchs, Rounsevell, et al. 2021. Europe greened substantially over the last 100 years. Forests have expanded by more than 30% since 1900, an area the size of Portugal (Eurostat 2021). However, our satellite data show that this trend has slowed down since 1980; in recent decades there are also many areas where nature is under pressure and vegetation cover deteriorates. European Environment Agency 2020 reports agriculture and urbanization as the most critical pressures for species and habitats. The diverse land-use changes across the EU raise the question of whether protection efforts have been effective in safeguarding nature under pressure, or contributed to expanding natural areas where land use pressures have lessened.

We evaluate the causal effect of Europe's protected area policies on an important dimension of biodiversity—vegetation cover—and on human presence in protected areas—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, soil and cli-

mate 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, land, and soil 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 areas 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 (Chaisemartin and D'Haultfœuille 2020; Sun and Abraham 2021; Goodman-Bacon 2021). 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, soil, and climate attributes, and the degree of local economic pressures from agriculture and forestry. Using our expansive data, we use the 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 statistically precise, 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 such as initial greenness, population density, or measures of agricultural land productivity; the zero effect is stable and pervasive across a wide range of CATEs. Sixth, we 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.

These findings are new and important. We argue that the zero effect can be explained by a combination of siting decisions of protected areas and the level of stringency of these protections. Our results suggest EU policymakers select sites that are never threatened in our sample. While protection might be effective against future development threats, our analysis uncovers no evidence of such threats in areas protected for several decades. These are sites with little economic opportunity cost. Such siting is perhaps motivated by "easy" green-glow benefits 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). We further document that protection also happens on land that is likely under threat of development or that could see substantial long-run greening effects from strict protection. These typically agricultural areas often face only weak protection regimes that do not put strong limits on economic activity, hence also partly explaining the zero treatment effect.

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 and to opt for stricter enforcement regimes.

There is a large literature on protected-area policies in tropical forests in developing countries (e.g., Andam et al. 2008; Sims 2010; Pfaff et al. 2015; Sims and Alix-Garcia 2017; Souza-Rodrigues 2019; Assunção et al. 2022; Keles, Pfaff, and Mascia 2022; Cheng, Sims, and Yi 2023; Rico-Straffon et al. 2023). The literature has generally found a modest impact on forest cover, but estimates are highly context-specific. Stronger positive 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).

Evidence on the effectiveness of land protection in advanced economies has been lacking.[1] Providing such evidence from a large part of the world with different land-use and enforcement dynamics is critical for understanding how to equalize the marginal costs of land protection across countries and the normative debate about "who should do more." There are global studies of the effect of land protection that include Europe (Joppa and Pfaff 2009; Joppa and Pfaff 2010; Abman 2018; Maxwell et al. 2020; Wolf et al. 2021). Our study is unique in that our data spans a time period of four decades at high frequency and has fine geographic resolution so that we can study the staggered introduction of land-protection policies using recent econometric advances that yield unbiased estimates in such settings and allow for unprecedented opportunities to study how treatment effects evolve over event time and cohorts. Prior studies of land protection—both global and national—have lacked one or more of these elements, typically covering relatively short time periods (about a decade or less), leveraging cross-sectional or limited panel variation through standard matching estimators, and/or lacking a causal research design.

1. Auffhammer et al. (2021)'s study of protection on land-market impacts in the US is a rare exception.

A rapidly-growing literature studies the economics of conservation. Harstad and Mideksa 2017 and Harstad 2023 study conservation decisions by competing and subsequent policymakers. Hsiao 2021 studies the interaction between trade and weak conservation regulation in Indonesia. Aronoff and Rafey 2023 and Aspelund and Russo 2024 study the mechanism design of conservation markets in the US; contrary to the EU, the US relies on explicit market mechanisms for wetland protection and the Conservation Reserve Program; Druckenmiller and Taylor 2022 show that wetland protection has economic benefits. Our paper contributes to this literature by focusing on the environmental outcomes of a very large regulation-based conservation program in Europe, showcasing that weak regulation and site selection in areas with low opportunity costs are also an issue in advanced economies. Potentially-ambitious international targets ratified by the EU might contribute little to conservation when local implementation suffers from poor siting and limited restrictions.

The rest of the paper proceeds as follows. Section 2 provides details about the EU's land-protection policies. In Section 3, we offer a conceptual framework to think about site selection and strictness of protection from a policymaker's perspective. Section 4 describes the data. In Section 5, we provide descriptive evidence that significant parts of the EU are at risk of development, and that protection-area siting and protection strictness are often poorly targeted. We outline our methodology in Section 6 and present the estimation results in Section 7. Section 8 concludes.

2 Protected-area policy in the European Union

The EU specifies biodiversity strategies for each decade that translate the ratification of international agreements into specific EU goals.[2] The 2010 and 2020 strategies progressively increased the scope of the EU's protected area policy, which will continue into the next decade. Europe's Green Deal and Nature Restoration Law aims to protect 30% of its land by 2030 (European Parliament 2023) in line with the ratification of COP 15.[3] The member states adhere to these targets by assigning areas under the Natura 2000 policy (European Union 2009). This policy combines two earlier EU directives: the habitat directive (The Council of the European Communities 1992) and the birds directive (The Council of the European Communities 1997). The directives describe a list of species and habitats requiring conservation measures to ensure survival. The list is split into annexes based on the extent to which species and habitats are threatened. The directives list a set of restrictions for each annex, such as a restriction on land use to preserve the habitat of endangered species. The directive requests member states to take measures to maintain the animal population's size and habitat's territorial presence while considering economic requirements. Every six years, all EU member states report on the state of listed species and habitats. We report the findings of these reports below.

2. See, for example, the communication on the 2020 strategy in European Commission 2011.

3. This Nature Restoration Law passed the EU Parliament and is currently contested at the EU Council level. Many member states oppose the 2030 measure because far more agricultural restrictions are embedded in the proposed legislation.

The directive requires countries to submit a standardized report on their protected areas to the European Commission, following International Union for Conservation of Nature (IUCN) guidelines. The Commission evaluates member state proposals and may amend them. This could lead to a back-and-forth between the Commission and member states over revisions to protection proposals. Member states then ratify their plans into national laws that specify the legal status of protected areas.

While the EU directives describe species and habitats needing protection, member states are responsible for translating these guidelines into actual policy. First and foremost, member states decide on the siting of protected areas. In line with the EU list of species and habitats, they identify territorial regions contributing to improving listed species and habitats. The member states then draw the boundaries of the areas in cooperation with ecologists, local communities, and farmers, and specify restrictions on different economic activities. The member states choose the siting and draft national laws and regulations that implement and enforce the policy. The Commission oversees the member state plans and gives feedback but has no direct regulatory or enforcement power. This legal setting leads to many potential local variations in the policy's siting, restrictions, and enforcement.

When studying the siting policies of member states more closely, it is striking that the procedure allows for a great deal of input from local communities and farmers. For example, the French procedure is detailed in Hassan Souheil and Douillet 2011 and is titled "Dialogue for Natura 2000." The French state contacts the prefect of a region if their experts in the Ministry of the Environment identify a region where protection should be implemented. The local prefect then establishes a steering committee ("comite de pilotage") to bring together all possible stakeholders to formulate the objectives of the protected areas and the siting. The report lists a dozen involved official structures (such as government bodies, hydrological agencies, and sports organizations) and two dozen stakeholders (ranging from hunters and farmers to tourists and scientists) as possible steering committee members. For each validation of a Natura 2000 area the collective procedure allows for participation and negotiation with local stakeholders. Participation is key; the document advises that "each stakeholder, each inhabitant, is legitimate to be involved closely or remotely in the Natura 2000 process simply because of their link to the area concerned. Leaving no one behind is a good way to avoid local stakeholders feeling a lack of consideration and thus to limit opposition" (Hassan Souheil and Douillet 2011). The regional steering committee validates a formal document (DOCOB) that is passed up to the French government, which in turn includes it in the national protected area plans submitted to the EU Commission.

Most member states' procedures incorporate local participation in the siting decisions. The Netherlands delegates the governance of protected areas to provinces in collaboration with local stakeholders. Each siting decision is publicly available and revised after a six-week period of public feedback.[4] Similarly, the UK Department for the Environment, Food and Rural Affairs

4. See https://www.natura2000.nl/werkwijze/aanwijzing-natura-2000-gebieden-0.

(Defra) oversees the UK's siting decisions after Natural England (Defra's executive body) identifies a possible area for siting. Natural England carries out a public consultation, to give everyone who might be affected by the designation or who has relevant scientific information an opportunity to comment. This includes landowners and occupiers, local planning authorities, other agencies and interested organizations. The results of the consultation are reported back to Defra, which may request Natural England to try and resolve any remaining objections to the designation or provide more scientific information to support the proposal.[5] Participation of local stakeholders seems to be the key unifying element of the procedure across EU member states.

Because the actual implementation is delegated to lower government bodies, the member states report many types of protected areas to the EU Commission. The policy thereby 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.[6] Furthermore, the policy encompasses protected areas beyond merely aiming to improve biodiversity. Some are protected landscapes and cultural heritage, and others serve recreational goals such as ecotourism.

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." The evaluation is based on expert opinion and member state surveys from the recurring six-year reporting duty. However, none of the member states collects high-quality quantitative data that allows causal evaluation. Our study aims to provide large-scale, long-term causal evidence for the effectiveness of the EU's land-protection policy.[7]

The EU evaluation states that the proportion of species listed in the birds directive with poor and bad status increased by 7% so that 81% of habitat assessments have a poor or bad conservation status, and only 15% of habitats have a good status. The main pressures that deteriorate the status of habitats are agricultural activities and urbanization. Agriculture is reported to affect habitats through changes in grassland management, landscape fragmentation, land-use conversion, and drainage. Member states reported more than twenty thousand areas under severe pressure, confirming that the protection policy encompasses land where nature could potentially expand.

5. See https://consult.defra.gov.uk/natural-england/crouch-roach-estuaries/supporting_documents/European%20leaflet%20Natura%202000.pdf.

6. Unfortunately, comprehensive ownership data across European countries are not available to us.

7. We use the term EU, but the data include the 27 member states as well as Albania, Bosnia, Montenegro, Macedonia, Switzerland, and Serbia. We exclude Iceland, Malta, Norway, and Liechtenstein due to missing data issues.

The report ends with actions that could improve the performance of the network. The first recommendation is to improve site selection: "Inefficient site selection has been linked to politically-motivated selection and giving low priority to conservation objectives compared with economic objectives." Furthermore, the report states that management and monitoring could be more effective and that authorities should prioritize ecological performance. The poor performance of the Natura 2000 network led the EU to formulate a more ambitious policy in the Nature Restoration Law of the EU Green Deal.

3 The policymaker's objective function

Site selection of protected areas is a crucial responsibility of EU member states. We formalize the policymaker's decision to protect land with some minimal notation. Consider a choice over protecting grids $i \in \{1, 2, \ldots, I\}$. Treated areas are protected at some time $t = g(i)$. Control areas are never protected. Land has economic use value v_{it} and ecological value e_{it} that captures the biodiversity, carbon capture, and other ecosystem services. We assume that e_{it} is decreasing in v_{it}—the more a plot develops economically, the lower its ecological value becomes. Land not used for human economic activity, with $v_{it} = 0$, obtains its maximum potential ecological value. We allow v_{it} and e_{it} to be correlated across i because of agglomeration effects in economic activity or ecology.

The total value from a non-protected unit of land accruing to a policymaker is the untreated potential outcome:

$$V_{it}(0) = \sum_{t}^{\infty} \delta^t [v_{it} + e_{it}(v_{it})],$$

where 0 refers to the absence of any land-use restriction and δ to the discount rate.

Protection limits the economic value to $v_{it}^1 \leq v_{it}$, where 1 refers to protected status.[8] A strict wilderness protection policy forbids all economic activity such that $v_{it}^1 = 0$ for all $t \geq g(i)$. Because e_{it} is decreasing in v_{it}, a lower level of v_{it}^1 results in more ecological gains. If v_{it}^1 is only slightly below v_{it}, protection imposes few economic restrictions; this might be the case for many of the EU's species management areas (see Appendix C.1). Protection comes with an administration, monitoring, and enforcement cost of c_{it}. Additionally, there might be political benefits u_{it} from protecting *any* plot of land, regardless of the ecological gains. These might come from a variety of reasons. The policymaker might benefit from compliance with an international agreement that specifies an area-based target, or there might be political green-glow benefits from implementing an environmentally-friendly policy (even if its impact is limited).

8. We exclude cases in which land use before protection is inefficient, and protection can increase both v and e.

The value of a protected unit of land at $t = g(i)$ equals:

$$V_{ig}(1) = \sum_{t=g}^{\infty} \delta^{t-g}[v_{it}^1 + e_{it}(v_{it}^1) - c_{it} + u_{it}].$$

The difference in the value of a plot i that is protected at $t = g$ relative to the no-protection counterfactual equals:

$$V_{ig}(1) - V_{ig}(0) = \sum_{t=g}^{\infty} \delta^{t-g}[v_{it}^1 - v_{it} + e_{it}(v_{it}^1) - e_{it}(v_{it}) - c_{it} + u_{it}], \tag{1}$$

which shows how land protection is a trade-off between losses from restricted economic activity, ecological gains, implementation costs, and political benefits.

A social planner would select sites and set a level of stringency to maximize its objective function by balancing economic value, ecological value, and protection costs, with no consideration of political green-glow benefits (hence, $u_{it} = 0$).[9] Protection must, therefore, positively affect ecological value—it must change economic production in the area relative to the status quo and foster natural vegetation cover and species survival; otherwise, it reduces total surplus. The planner would not engage in the protection of land that is never under any economic pressure ($v_{it}^1 = v_{it} \forall t$) as long as $c_{it} > 0$; such efforts would be considered a wasteful allocation of resources. Land at risk of future economic development may warrant protection.

Notice that even in areas where there is greening, a protection policy might still have meaningful effects as long as there are areas where the social planner can gain by restricting v_{it}. In cases where v_{it} declines over time, a social planner would still choose to allocate protection efforts (i.e., choose a path of v_{it}^1 that declines faster than v_{it}) to areas where the policy helps increase biodiversity most while balancing costs. The policy would accelerate a rebound of nature beyond what we would see without protection. This is relevant in Europe, where we see reforesting in some areas yet increasing pressure in other areas. Therefore, a protection policy can accelerate greening and reforestation in certain areas, and prevent loss of vegetation in other areas.

In reality, EU member states' politicians are not necessarily following the social planner's objective. They may prioritize protection differently than the supranational EU policymaker that sets the protection targets. Or they may choose inefficient levels of restrictions in response to lobbying from local stakeholders. For example, politicians might find inframarginal protection attractive as long as $u_{it} > c_{it}$. Moreover, politicians might undervalue ecological benefits, as benefits accrue to societal actors beyond their constituents, further raising the attractiveness of inframarginal protection. Or they may emphasize current local economic activity, especially when the economic burden on local land owners and farmers are immediate, and the ecological benefits are far less tangible. In such situations, we might expect a zero treatment effect caused by selecting sites or restrictions

9. Aspelund and Russo 2024 study mechanism design to select plots when the policymaker has imperfect information about c and v.

that are too lax to produce ecological benefits.

Different policy priorities will change the order in which land is protected and the resulting time path of treatment effects. Suppose a limited number of grids can be protected in each period, 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.[10] If instead a policymaker only cares about ecological benefits and not the economic costs of protection, land with the highest ecological gains from protection will be protected first, resulting in large treatment effects initially, but lower treatment effects in the future as the most ecologically-beneficial areas have already been protected. In contrast, if a policymaker cares only about the economic costs, green glow, and meeting an area-based protection target at minimum economic cost, we expect protection to occur first on land with the lowest economic opportunity cost—typically land not at risk of development. This would result in near-zero treatment effects initially, but larger treatment effects over time as the policymaker runs out of 'cheap' land to protect. Our rich panel data will allow us to causally estimate the time path of treatment effects across the order in which areas are protected.

4 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.[11] 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. We use NDVI because we can construct the measure with early Landsat data, allowing us to obtain a panel that is four

10. The optimal path of protection depends on the correlation of v_{it}, e_{it}, u_{it}, and c_{it} across i. See Weitzman 1998 and Metrick and Weitzman 1998 and the subsequent literature for theoretical models of optimal siting.

11. 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.

decades long. Alternative measures such as Vegetation Cover Fields (VCF) are only available since 2000.[12] We limit the sample to plots with an NDVI value of at least 40 at one occurrence, which includes plots with sparse vegetation, and only excludes urban grids, bare soil, and rocky landscapes; greener landscapes have NDVI values above 40 and include cropland, grassland, forests and suburban areas with a mix of buildings and vegetation.[13]

Previous research has shown that 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). The finding that NDVI is an excellent predictor for bird diversity is replicated in our study area for French (Bonthoux et al. 2018) and Mediterranean landscapes (Ribeiro et al. 2019). NDVI has shown to be effective in measuring the treatment effect of policies targeting land use across different sensors (Lassiter 2022).

Furthermore, we think it is helpful to focus on continuous measures. First, it allows us to capture gradual changes in vegetation that do not necessarily lead to land use re-categorization. When there is gradual vegetation growth, new construction such as housing or large-scale solar panel installation, or agricultural expansion, a plot's NDVI score decreases even when the discrete land-use category might be unaffected. 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 and that forests are much greener than agriculture. For example, in 2010, French forests had an average NDVI of 75 and croplands 50. Additionally, the continuous measure avoids classification errors that plague categorical land use classifiers (see Torchiana et al. 2023; Alix-Garcia and Millimet 2022). Continuous measures are critical because our data encompass various protected areas. In 93% of the protected areas some economic exploitation of the land is allowed for and we might expect some of these areas not to have discrete land-use changes but gradual greening.

Nightlights. We rely on Li et al. (2020) for a 1992-2018 one square kilometer panel of remotely-sensed nightlights. 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 and reduced traffic, which could reduce 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). We do not aim to interpret nightlights as a GDP/economic development measure. Still, we think it is useful as a measure complementary to NDVI, capturing the degree of human presence in the area.

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

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

13. See Appendix D and the discussion in Section 7.4 for details and evidence that our results are robust to the chosen NDVI threshold. Note that our sample selection procedure allows for the inclusion of plots that have NDVI below 40 before treatment and then green over time; a plot needs to cross the 40 threshold only once to be included.

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.12 shows land-use shares between 1900 and 2010, and Appendix Table A.16 reports the EU's land-use transition matrix for that period.

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 data allow for more direct measures of biodiversity. However, many count data suffer from one or all of the following issues: nonrandom location of counts, short panels with recent coverage only, a low number of species, model-based projections across space instead of raw data, 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.[14]

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

5 Descriptive evidence from 100 years of land-use data

In this section, we provide four empirical facts regarding land use in the European Union.

Fact 1. Large parts of Europe have experienced substantial land-use transition in the last 40 years. 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; white areas have only modest changes in greenness. The map reveals a substantial greening in certain parts of Europe, most pronounced in southeastern Europe and Scandinavia. NDVI decreases in the east of France, the west of Germany, and the Baltics. This land seems to be under pressure of economic development, and stringent protection may prevent vegetation loss. There is also lots of land in the EU that did not experience much change in greenness.

Panels (c) and (d) replicate panels (a) and (b) but only for areas that had received protection by 2019. Many protected areas have high greenness levels in 2019. Most of them experienced increases in NDVI between 1985-2019, although there are also many protected areas with small positive or negative NDVI changes (the white areas in panel (d)). However, comparing (b) and (d), many are

14. Promising improvements in more direct measures of biodiversity include several databases of animal tracking data, such as the Global Biodiversity Information Facility, Movebank, the PanEuropean Common Bird Monitoring Scheme, eBird, eButterfly, 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, https://www.e-butterfly.org/ and https://www.ebcc.info/.

also situated in areas that are greening regardless of land protection. The changes in greenness over time in protected areas vs. non-protected areas that are similar on observables, together with the staggered implementation of protection, form the primary identifying variation we exploit in this paper.

Fact 2. Economically valuable land continues to leave natural use. Table 2 demonstrates that economically valuable land has experienced deforestation in the EU. In the 2010 cross-section, compared to land that stayed forested, deforested land is more populated, more urbanized (higher nightlights), and more agriculturally suitable (longer growing seasons and potassium-rich soil). The table also underscores that the EU has protected land at risk of deforestation at equal rates with continuing forests. This balance table thus suggests that land protection does occur on land at threat of development.

Fact 3. Land protection does not uniquely occur on land at threat of development. We explore protected area siting visually in Figure 6. The figure indicates the probability of land protection since 1985, binning land by its starting greenness and population density measured in 1985. Then, for example, among all land which had 1985 population density less than 30 and greenness above 80, 22% by area is protected after 1985. Protection is more targeted towards land with higher greenness. However, beyond this, there appears to be little intentional siting with respect to population density. Densely populated land—which is correlated with higher deforestation pressure (fact 2)—is not more protected than low-pressure land.[15]

Fact 4. Land protection tends to be stricter on land which is less under pressure. We assess whether the stringency of protection correlates with land pressure. Figure 7 describes the likelihood that land protection is in IUCN Categories Ia, Ib, or II by greenness and population density. These three categories impose strict restrictions on human interaction: only scientific, ecological, and recreational use are permitted (though in some cases, an exception is made for use by indigenous peoples), while other categories allow for some economic use of the area.[16] Then, for example, among protected areas founded after 1985 which had 1985 population density less than 30 and greenness above 80, 26% are strict. The figure reveals a stark pattern: *post-1985* strict land protection occurs on low-density or already-green land as of 1985. While high land pressure areas have been protected in the EU, such areas are far more likely to be species management areas or other types of protection that do not substantively restrict human activity.

In summary, the European Union has been experiencing a major land transition over the last 40 years (Fact 1). Significant parts of the EU are still under threat of development by agriculture or urbanization (Fact 2). Despite such land pressure, protected areas are not targeted uniquely at land threatened by development (Fact 3). Land protection is less stringent on developed land with

15. Appendix Table A.12 presents the results from a linear probability model of land protection. We find that more economically valuable land, as measured by nightlight luminosity or by the presence of cropland or settled land use prior to protection, has a lower probability of protection.

16. The criteria for these three categories explicitly reference "pristinity" of nature: others do not. For more discussion of IUCN categories, see Appendix Table A.4. Adding areas of Type III (natural monuments) to the "strict" designation does not alter the overall pattern of results.

a high population density (Fact 4). The descriptive results support that land protection could be useful to preserve or strengthen greenness in the EU; they also suggest that siting often happens in already green, low-populated areas and that the EU applies weaker restrictions in at-risk areas. These may inhibit positive treatment effects on vegetation growth.

6 Methodology

As in Section 3, 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 or during $\mathbb{T}$.[17] 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 $D_{it} \in \{0, 1\}$ as equal to 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:

$$\tau_{igt} = \mathbb{E}[Y_{i,t} - Y_{i,g-1}|G, D_{it} = 1] - \mathbb{E}[Y_{i,t} - Y_{i,g-1}|G, D_{it} = 0], \tag{2}$$

comparing the difference in outcomes of grid i between periods t and $g - 1$ in treatment cohort G with the 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.

6.1 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 (2).[18] 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.

17. Because less than a quarter of land is treated at the end of our sample, we select control units from the remaining grids. These are the 'never-treated units' in our sample.

18. We use a difference-in-differences approach instead of a regression discontinuity design as in Turner, Haughwout, and van der Klaauw 2014. Many protected areas border grids with different land uses, causing discrete jumps in NDVI around protected areas.

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.[19] 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. Furthermore, 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 that 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 $g-1$ and t on plausibly exogenous observables, separately for every cohort group G:

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

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.

19. 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 with cohort-specific matching. For completeness, we also report results from standard two-way fixed effects estimation in Appendix Section C.3.

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 $\mathbb{T}$ as the set of time periods in our study sample; we refer to the last year of the sample as T. We compute country-specific treatment effects by averaging over the treated time periods and cohorts:

$$\theta_c = \sum_{g \in \mathcal{G}} w_{cg} \sum_{t>g}^{T} w_e \theta_{cgt} = \sum_{g \in \mathcal{G}} \frac{N_{cg}}{\sum_{g \in \mathcal{G}} N_{cg}} \sum_{t>g}^{T} \frac{1}{T - g + 1} \theta_{cgt}, \tag{3}$$

where N_{cg} is the number of observations in country c that were treated in foundation year g, and w are weights. 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 as 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:

$$\theta_{gt}^{EU} = \sum_{c \in \mathcal{C}} w_{cgt} \theta_{cgt} = \sum_{c \in \mathcal{C}} \frac{N_{cgt}}{\sum_{c \in \mathcal{C}} N_{cgt}} \theta_{cgt}. \tag{4}$$

Generally, such a weighted sum requires estimating the weights w_{cgt}. In our setting, we observe the true weights on observations because our data cover all land 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.[20]

We aggregate 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:

$$\theta_{e}^{EU} = \sum_{g \in \mathcal{G}} w_g \theta_{ge}^{EU}, \tag{5}$$

where w_g equals the fraction of cohorts observed at event time e. Finally, aggregating over event times gives a measure of the average effect of being treated at time g:

$$\theta_{g}^{EU} = \frac{1}{T - g + 1} \sum_{e=0}^{T-g} \theta_{ge}^{EU}, \tag{6}$$

20. Our standard errors for the EU-level effects are likely smaller relative to a confidence band that would incorporate covariance in country-level treatment effects. However, EU member states independently decide which land to protect, and we do not observe coordination between member states in practice. The choice of aggregation technique is also motivated by the computational cost of the doubly-robust estimator.

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.[21] 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 an equivalent aggregation at the country level to examine heterogeneity in the results across countries. We aggregate within-country and across event time to obtain θ_{cg}.

6.2 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 pre-protection greenness, population density, varying soil characteristics, or weather conditions impacting vegetation regeneration ability. For example, 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. (2024) 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 $\tau_x = \mathbb{E}[\tau|X = x]$ for covariate 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 treatment 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.

21. 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.

7 Results

7.1 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 0.03 with a standard error of 0.05 for the change in the NDVI index that ranges between 0 and 100.[22] We cannot reject a 0 effect of protection on vegetation greenness, and the confidence interval is tight around zero: the EU-wide average treatment effect estimate is a precise zero. The figure also displays the estimates from two naive, and biased, two-way fixed effects estimators (with and without matching).[23] Their EU-wide ATE of 0.4 is similarly small. In Appendix D, we conduct several robustness checks. By looking at the first differences in greenness as an outcome, we confirm that the null result and pre-treatment parallel trend are robust to changes in functional-form assumption (Roth and Sant'Anna 2023).[24] We also show results are unchanged when we lower the NDVI sample selection threshold from 40 to 30.

There are some differences across countries (bottom panel). However, the size of all estimated treatment effects in absolute terms is typically small—θ_c varies between -2 and 2. In 14 out of 31 of the countries, the treatment effect is less than a 0.5 points change in the NDVI index, which is small relative to the average NDVI in our sample. Save the Netherlands, Serbia, and Slovenia—every country has a point estimate less than 1 in absolute value.[25] In the distribution, nearly half of the countries have negative treatment effects and nearly half of the treatment effects are positive. The treatment effects for the five largest countries by surface area (Poland, Germany, France, Sweden, and Spain) are all less than 0.5 NDVI points. Three of these five countries have negative treatment effects. Altogether, Figure 2 shows some meaningful treatment effects in a handful of mostly smaller EU countries, but generally large positive treatment effects are rare, and the Europe-wide average treatment effect is small by any standard.

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

23. 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: $Y_{it} = \beta D_{it} + \lambda_i + \lambda_t + \epsilon_{it}$. We report the dynamic estimates from the two-way fixed effects model in Appendix C.3.

24. We also report results from a spatial first-difference specification. Because our results indicate a zero treatment effect of protection on greenness, we are not concerned about spillovers. Leite Mariante and Salazar Restrepo 2024 present a model to identify the general equilibrium effect of conservation in the Brazilian context where conservation has positive treatment effects on forests. Robalino, Pfaff, and Villalobos 2017 show empirically that protected areas in Costa Rica facing greater threats of deforestation show greater leakage.

25. The treatment effect of protected areas in the Netherlands is the only point estimate above two. However, the Netherlands has only three cohorts. Most countries in the data have many more (see Appendix Table A.11 for details). As our identifying variation comes from staggered timing, fewer cohorts introduce a loss of precision. For comparison, Macedonia also only has three cohorts, and has standard errors of a similar order of magnitude.

7.2 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 greenness is consistently close to zero—almost all confidence intervals are contained within $[-1, 1]$ up to 20 years post-protection. Even thirty years after treatment, we find no evidence of a positive protection effect. The lack of a pre-trend is reassuring and suggests the matched controls are appropriate comparisons for the treated areas.[26] This does not imply that vegetation greenness has been constant over time. Certain parts of Europe have been greening, but treatment and control plots in those areas 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.

7.3 Other outcomes

Having established a mostly-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 human activity in areas set aside for protection for at least two decades (left panel). In addition, the

26. There are some pre-treatment coefficients that are significantly different from zero, but these are close to zero and would likely not imply large changes to the post-treatment confidence intervals (Rambachan and Roth 2023).

effect of land protection is close to zero regardless of the year of first protection. This corroborates our NDVI results (right panel).

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 4), 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.

7.4 Further treatment-effect heterogeneity

As discussed in Section 6.2, 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.2). The sample average treatment effect on NDVI is -0.21 (0.03).

Our data rejects the existence of any meaningful heterogeneity. While our random forest detects statistically-significant heterogeneity in the test of Chernozhukov et al. (2024), economically meaningful heterogeneity is limited. The CATE in the bottom quartile of the CATE distribution is only 0.7 (1.2) NDVI units smaller than in the top quartile. We thus precisely estimate limited heterogeneity around zero across the whole covariate distribution, not just for average land.

In Figures 5a and 5b, we highlight two important dimensions on which we can reject meaningful heterogeneity. These figures plot conditional average treatment effects at ventiles (percentiles in increments of 5%) of key covariates. On the left, Figure 5a shows more sparsely (or barely) vegetated areas do not green faster due to land protection than greener areas. Conditional on a wide range of greenness at the start of the sample, average treatment effects are not significantly positive.

Our theoretical framework suggests that land pressure, which we measure by population density, is an important mediator for treatment effects. Therefore, on the right, we repeat the exercise for

conditional average treatment effects in population density in Figure 5b. We find no evidence that densely populated land greens as a consequence of protection. The difference in the conditional average treatment effect between high vs. low population density areas is on the order of 0.5 NDVI points. This implies that even less green, higher population density lands that likely face stronger economic development pressure—such as areas around urban centers—are hardly re-greening faster after protection relative to similar, unprotected areas (and the average treatment is approximately zero).

In the appendix, we also find no significant variation in treatment effects when considering covariates linked to high-quality agricultural land. The left panel of Appendix Figure A.7 shows CATEs along the distribution of the soil phosphorus content. Significant soil phosphorus deposits can aid the development of "roots and shoots" and often also mark prior agricultural fertilizer use. The findings show no notable difference in greening patterns between areas with different soil phosphorus content. Thus, protecting observably more valuable agricultural land does not lead to measurably richer vegetation. The right panel shows the CATE by growing season length in 1985. Longer growing seasons indicate climatic conditions that are more conducive to growing common field crops. The results do not suggest meaningful treatment effect heterogeneity along this dimension of agricultural productivity, either.

Finally, we explore the conditional average treatment effect as a function of the treatment "intensity", the strictness of the land protection regime. We calculate the conditional average treatment effect in each IUCN category (see our prior discussion of Figure 7) in Table 4. We find that the estimated conditional average treatment effect is strictly decreasing in the strictness of protection. Estimates are significantly discernible: IUCN category VI sustainable resource areas see a more negative treatment effect (e.g., green more slowly than comparable control areas) than category I strict nature reserves. Economically speaking, the magnitudes are small: the strictest protected areas deterred 0.1 NDVI points worth of greenness loss relative to the least strict areas.

Regarding the interpretation of our results, our estimates support a zero effect throughout the distribution of protection. CATEs across other covariates, which we would expect to capture salient dimensions of land pressure, such as distances to shorelines, all trend less than those shown in Appendix Figure A.7.

7.5 Discussion of the results

There are several potential explanations for the zero results. First, many protected plots are infra-marginal, and hence, there is no change in land use before and after protection. Second, protected areas are often located in areas with substantial development threat, but in those areas policymakers choose weak levels of protection. Third, protected plots were under threat of *future* development rather than contemporaneous economic activity. Fourth, authorities do not enforce the restrictions of economic activity.

The third and fourth explanations seem 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. Furthermore, 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 small effects in the vast majority of EU countries, and there is no clear pattern suggesting weaker effects in countries with a history of limited enforcement. While we cannot directly measure enforcement, we test for correlation between the country-level ATEs and an index of apparent corruption from www.transparency.org. The Pearson's correlation coefficient is 0.06.[27] We thus find no clear evidence of a relationship between a measure of state enforcement capacity and land protection outcomes.

We conclude that the first two explanations are the most likely: the zero effect can be explained by a combination of targeting land not at risk of economic development and choosing weak protection levels for land in areas that face development pressure. Our descriptive evidence in Section 5 confirms that land protection does not uniquely occur on land at threat of development, and that protection on land at pressure of development is likely to be the least stringent type of land protection. The latter aligns with the multi-stakeholder process by which new protected areas are added—consensus will be easier to reach when economic opportunity costs are low.

When interpreting our results, it is important to mention that NDVI has important strengths as well as limitations. As discussed in Section 4, NDVI has been found to correlate well with measures of biodiversity beyond vegetation cover, such as bird diversity. That said, NDVI will not be able to detect other types of biodiversity improvements. For example, species management areas may improve insect counts, which may not always correlate with vegetation cover. However, our overall zero result is quite pervasive—across almost all countries, time periods, areas of different greenness and land use types—and provides compelling evidence that land protection has not generally been effective, even if we cannot rule out potential improvements in certain biodiversity measures on particular types of land. One would need much higher-resolution data on species counts over long periods of time to test the causal effects on biodiversity beyond vegetation.

8 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, population density, land, soil, or climate characteristics. Control plots show equal greening trends as treated plots. This finding is new and important, as it helps evaluate whether the siting decisions and stringency of protected areas

27. Further, a linear regression of the average treatment effect on the log of this index, measured from 0 to 100, gives a coefficient of 0.33 (0.68).

are preventing land development and ecosystem degradation. It also provides the first large-scale evidence of the effects of land protection in advanced economies, which are critically important to meet global protection targets but exhibit very different land-use and enforcement dynamics than in tropical forest countries, which is where the literature has focused until now.

So far, Europe has expanded land protection with limited opportunity costs. For the most part, protection has happened on unthreatened areas or, where it has targeted land that faces development pressure, it has lacked the necessary stringency. There is a strong focus—both in the EU and globally—on achieving area-based targets, such as 30% of land mass under a protected designation, but it is the specific lands and their counterfactual outcomes in the absence of those designations that determine the true policy impact. 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 or has opted for lax protection in places with significant economic pressures. One interpretation is that politicians have focused on protecting land with low opportunity costs. Yet a more optimistic interpretation is that there is still significant scope for expanding the protection of European land at a low economic cost.

References

Abman, Ryan. 2018. "Rule of Law and Avoided Deforestation from Protected Areas." *Ecological Economics* 146:282–289.

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.

Aronoff, Daniel, and Will Rafey. 2023. "Conservation Priorities and Environmental Offsets: Markets for Florida Wetlands." NBER Working Paper No. 31495.

Aspelund, Karl M., and Anna Russo. 2024. "Additionality and Asymmetric Information in Environmental Markets: Evidence from Conservation Auctions." Working Paper.

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." *Review of Economic Studies* 90 (4): 1608–1641.

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." Working Paper, https://conference.nber.org/conf_papers/f157199.pdf.

Balboni, Clare, Robin Burgess, and Benjamin A. Olken. 2023. "The Origins and Control of Forest Fires in the Tropics." Working Paper, https://www.robinburgess.com/s/Balboni_etal_2020_The-Origins-and-Control-of-Forest-Fires-in-the-Tropics.pdf.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *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.

Burgess, Robin, Francisco J.M. Costa, and Benjamin A. Olken. 2023. "National Borders and the Conservation of Nature." Working Paper, https://www.robinburgess.com/s/Burgess_Costa_Olken_2019_The-Brazilian-Amazons-Double-Reversal-of-Fortune.pdf.

Burgess, Robin, Matthew Hansen, Benjamin A. Olken, Peter Potapov, and Stefanie Sieber. 2012. "The Political Economy of Deforestation in the Tropics." *Quarterly Journal of Economics* 127 (4): 1707–1754.

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.

Chaisemartin, Clément de, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110, no. 9 (September): 2964–96.

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. 2024. "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments." Forthcoming, *Econometrica.*

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

Druckenmiller, Hannah, and Charles A. Taylor. 2022. "Wetlands, Flooding, and the Clean Water Act." *American Economic Review* 112 (4): 1334–1363.

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

European Commission. 2011. *Our life insurance, our natural capital: an EU biodiversity strategy to 2020.* Document 52011DC0244, Accessed: 2024-05-22. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52011DC0244.

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

European Parliament. 2023. *Report on the proposal for a regulation of the European Parliament and of the Council on nature restoration.* Document A9-0220/2023, Accessed: 2024-05-22. https://www.europarl.europa.eu/doceo/document/A-9-2023-0220_EN.html.

European Union. 2009. *Directive 2009/147/EC of the European Parliament and of the Council of 30 November 2009 on the conservation of wild birds.* Accessed: 2024-05-22. https://eur-lex.europa.eu/eli/dir/2009/147/oj.

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.

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.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." Themed Issue: Treatment Effect 1, *Journal of Econometrics* 225 (2): 254–277.

Harstad, Bård. 2023. "The Conservation Multiplier." *Journal of Political Economy* 131 (7): 1731–1771.

Harstad, Bård, and Torben K. Mideksa. 2017. "Conservation Contracts and Political Regimes." *Review of Economic Studies* 84 (4): 1708–1734.

Hassan Souheil, Danielle Boivin, Laurent Germain, and Robert Douillet. 2011. "Guide Néthodologique d'Elaboration des Documents d'Objectifs Natura 2000," no. 82, https://www.natura2000.fr/documentation/references-bibliographiques/guide-methodologique-elaboration-documents-objectifs.

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.

Hsiao, Allan. 2021. "Coordination and Commitment in International Climate Action: Evidence from Palm Oil." Working Paper, Department of Economics, MIT, http://allanhsiao.com/files/Hsiao_palmoil.pdf.

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.

Joppa, Lucas N., and Alexander Pfaff. 2009. "High and Far: Biases in the Location of Protected Areas." *PLOS ONE* 4 (12): 1–6.

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.

Lassiter, Allison. 2022. "Identifying Causal Changes in Landscape Greenness with Very High-Resolution Airborne Multispectral Imagery and a Panel Data Model." *Urban Forestry & Urban Greening* 67:127380.

Leite Mariante, Gabriel, and Veronica Salazar Restrepo. 2024. "Does Conservation Work in General Equilibrium?" Working Paper, https://vsalazarr.github.io/jobmarketpaper/.

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.

Metrick, Andrew, and Martin L. Weitzman. 1998. "Conflicts and Choices in Biodiversity Preservation." *Journal of Economic Perspectives* 12 (3): 21–34.

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): e0129460.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends." *Review of Economic Studies* 90 (5): 2555–2591.

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).

Rico-Straffon, Jimena, Zhenhua Wang, Stephanie Panlasigui, Colby J. Loucks, Jennifer Swenson, and Alexander Pfaff. 2023. "Forest Concessions and Eco-Certification in the Peruvian Amazon: Deforestation Impacts of Logging Rights and Restrictions." *Journal of Environmental Economics and Management* 118:102780.

Robalino, Juan A., Alexander Pfaff, and Laura Villalobos. 2017. "Heterogeneous Local Spillovers from Protected Areas in Costa Rica." *Journal of the Association of Environmental and Resource Economists* 4 (3): 795–820.

Roth, Jonathan, and Pedro H.C. Sant'Anna. 2023. "When Is Parallel Trends Sensitive to Functional Form?" *Econometrica* 91 (2): 737–747. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA19402.

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.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." Themed Issue: Treatment Effect 1, *Journal of Econometrics* 225 (2): 175–199.

The Council of the European Communities. 1992. *Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora.* Accessed: 2024-05-22. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31992L0043.

———. 1997. *Council Directive 79/409/EEC of 2 April 1979 on the conservation of wild birds.* Accessed: 2024-05-22. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A31979L0409.

Torchiana, Adrian L., Ted Rosenbaum, Paul T. Scott, and Eduardo Souza-Rodrigues. 2023. "Improving Estimates of Transitions from Satellite Data: A Hidden Markov Model Approach." Forthcoming, *The Review of Economics and Statistics,* https://www.ptscott.com/papers/hmm_error_correction.pdf.

Turner, Matthew A., Andrew Haughwout, and Wilbert van der Klaauw. 2014. "Land Use Regulation and Welfare." *Econometrica* 82 (4): 1341–1403.

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.

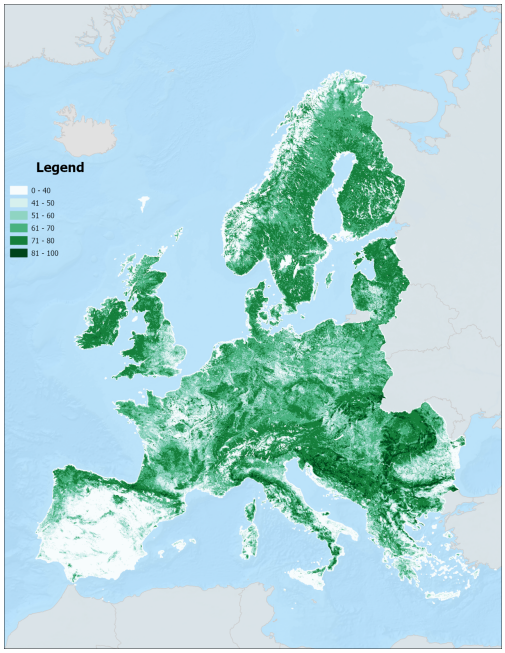
Weitzman, Martin L. 1998. "The Noah's Ark Problem." *Econometrica* 66 (6): 1279–1298.

Winkler, Klaus, Richard Fuchs, Mark Rounsevell, et al. 2021. "Global land use changes are four times greater than previously estimated." *Nature Communications* 12:2501. https://doi.org/10.1038/s41467-021-22702-2.

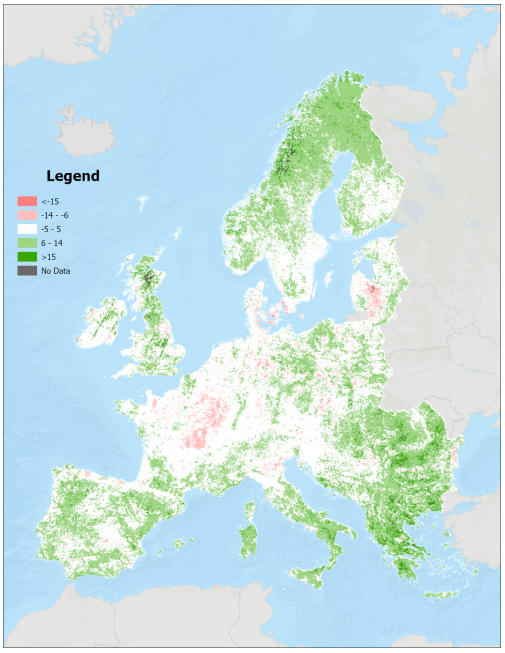
Wolf, Christopher, Taal Levi, William J. Ripple, Diego A. Zarrate-Charry, and Matthew G. Betts. 2021. "A Forest Loss Report Card for the World's Protected Areas." *Nature Ecology & Evolution* 5:520–529.

Figures and tables

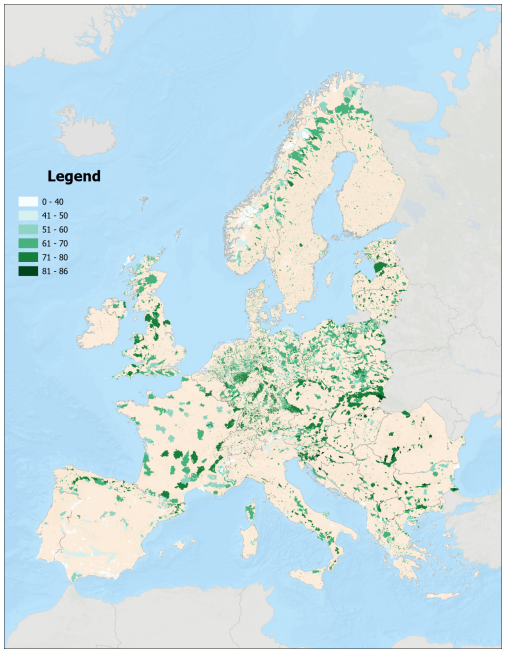
Figure 1: NDVI changes inside and outside protected areas



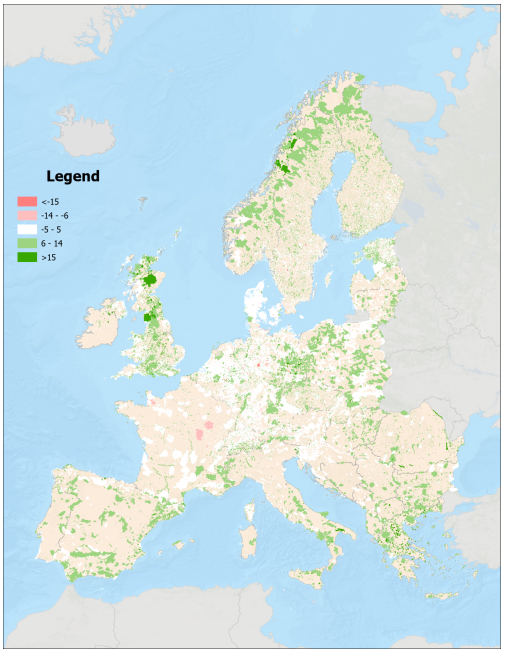
(a) Grid-level NDVI, end of sample



(b) Grid-level NDVI changes across sample



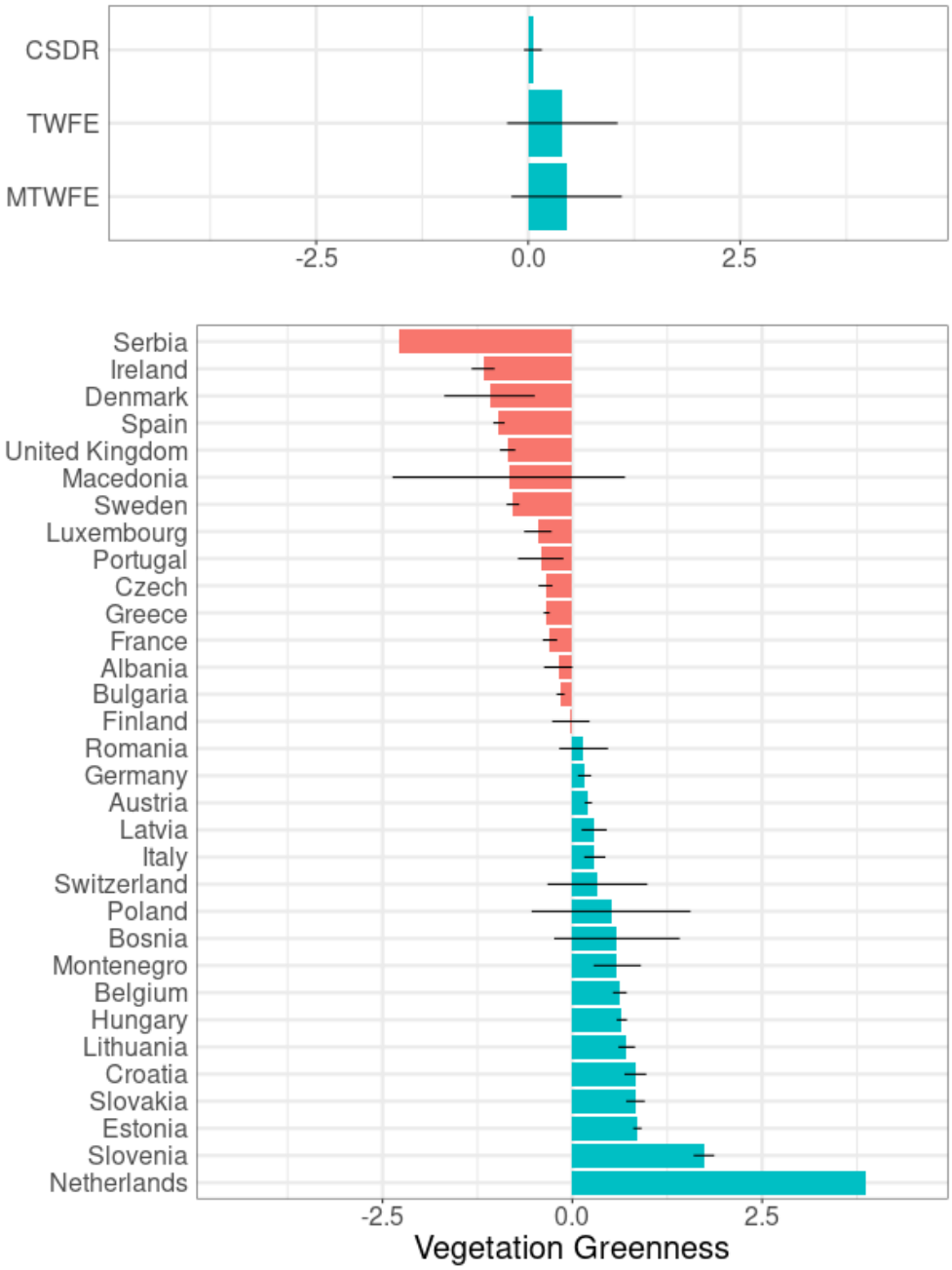
(c) CDDA-level NDVI, end of sample



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

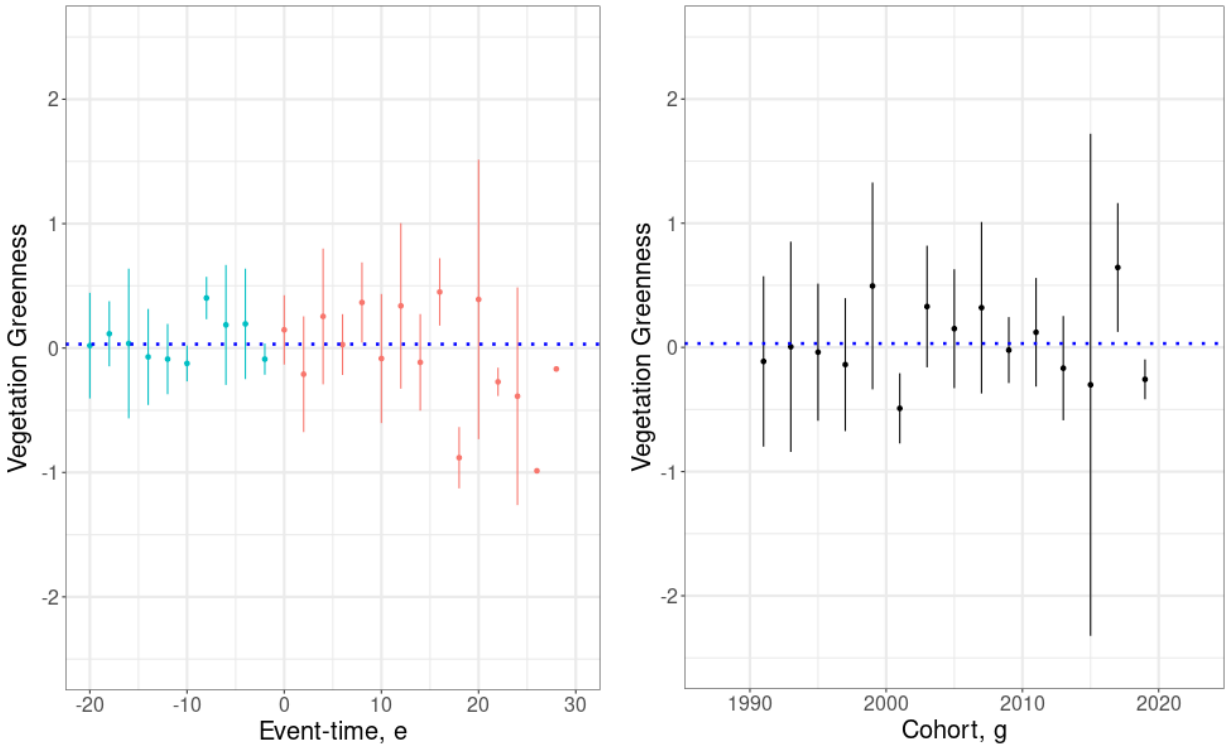
NOTES: Panel (a) plots the average NDVI across 2015-2019 for each grid. Panel (b) plots the change in average NDVI between 2015-2019 and 1985-1989. Panel (c) plots the average NDVI within each CDDA for 2015-2019. Panel (d) plots the change in average NDVI within each CDDA between 2015-2019 and 1985-1989.

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



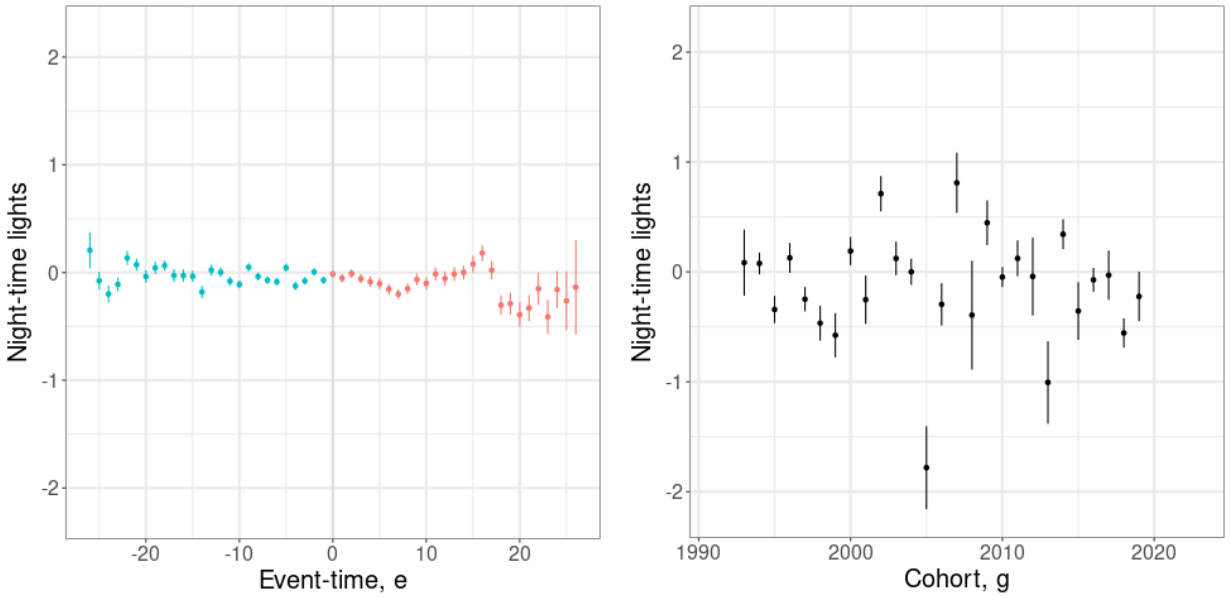
NOTES: Top panel reports treatment effects estimated through three distinct estimators: the Callaway and Sant'Anna doubly-robust procedure (CSDR), two-way fixed effects (TWFE), and matched two-way fixed effects (MTWFE). Bottom panel reports CSDR treatment effects θ_c aggregated by country with bootstrapped confidence intervals. Blue bars indicate positive ATEs; red bars indicate negative ATEs. Horizontal black lines indicate 95% confidence intervals (omitted for Serbia and the Netherlands). 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



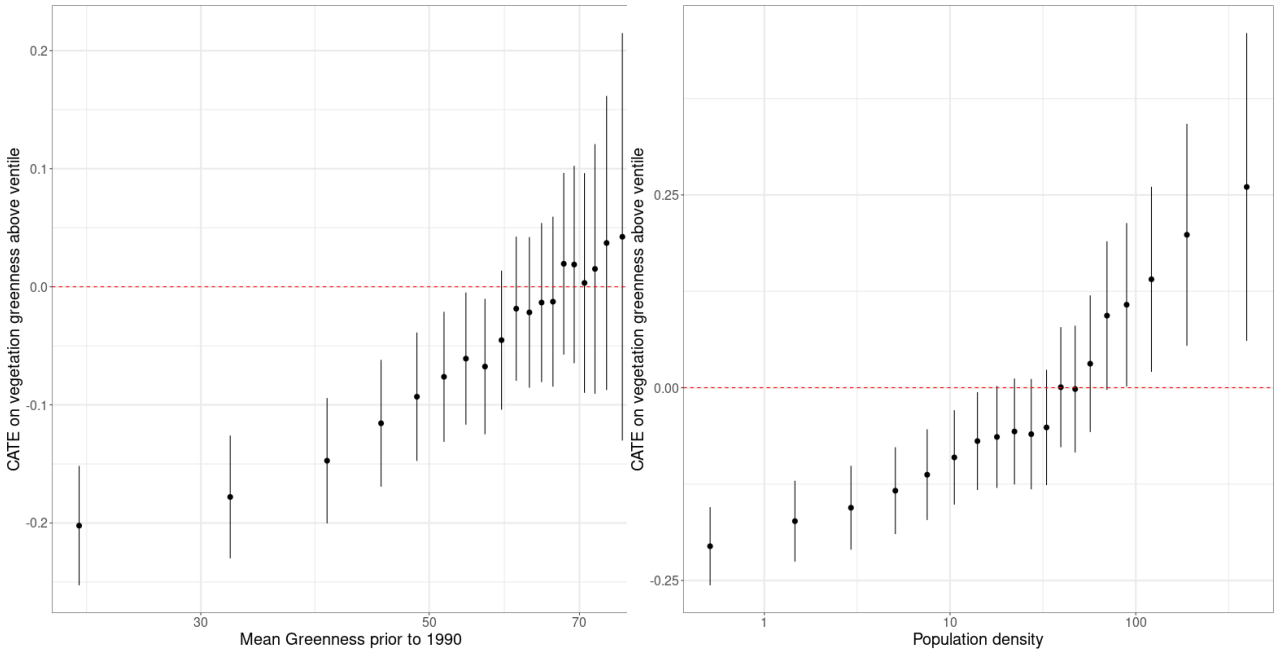
NOTES: Treatment effects on greenness aggregated by event-study period θ_e^{EU} in the left panel, and by cohort θ_g^{EU} in the right panel. Sample includes plots which have greenness above 40 at least once in the sample period with non-missing matching variables. Greenness varies from 0 to 100. Both 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. Later event-times with confidence intervals wider than $[-2, 2]$ are shown as points for visual clarity.

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_e^{EU} in the left panel, and by cohort θ_g^{EU} in the right panel. Night-time lights vary from 0 to 68, with all event-time estimates being smaller than 0.5 units of luminosity. Both panels use the Callaway and Sant'Anna doubly robust estimator with bootstrapped 95% confidence bands. Appendix Tables A.13 and A.14 present the regression results.

Figure 5: Distribution of the conditional average treatment effect for initial greenness and population density

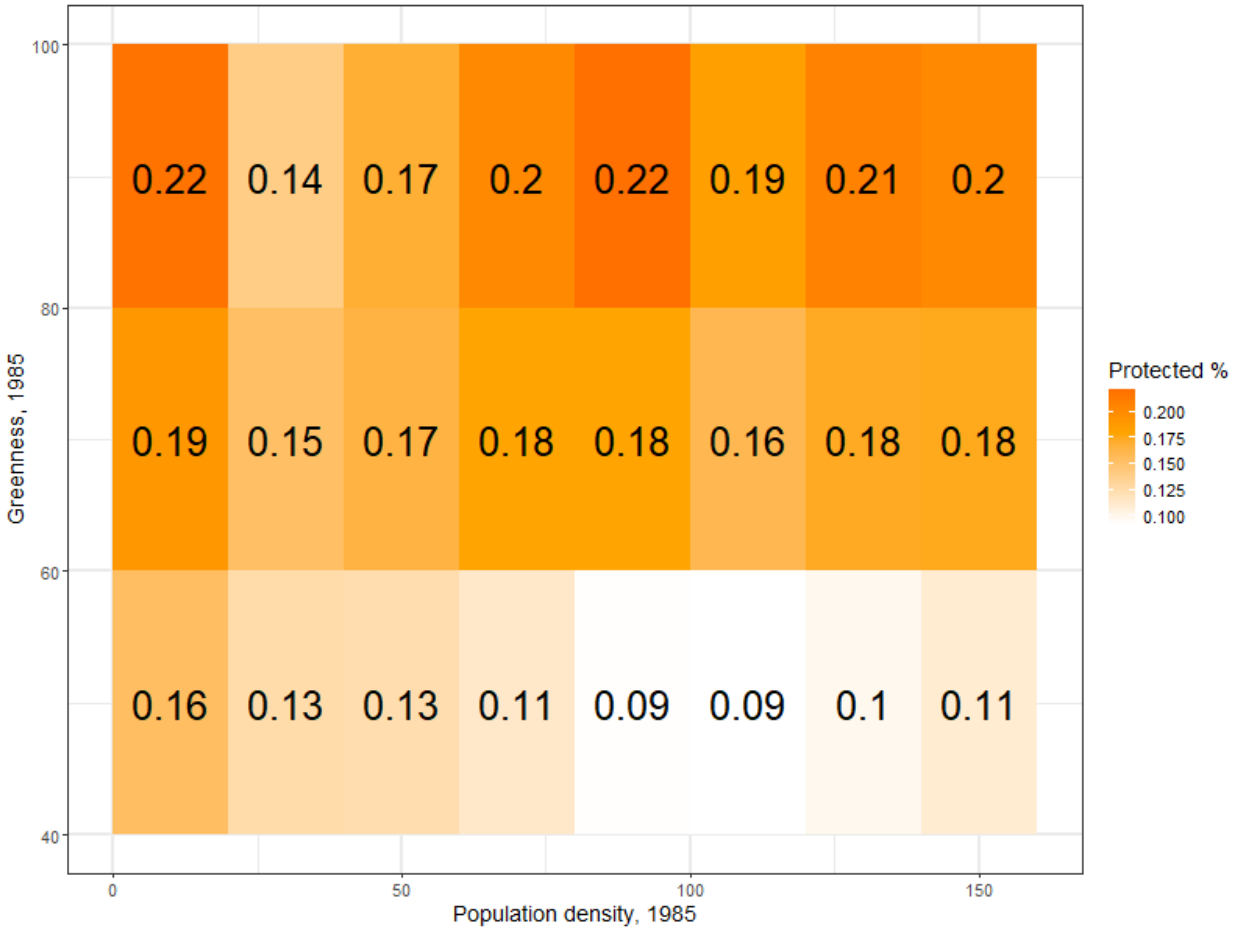


(a) Starting average NDVI 1985-1999

(b) Population density in 2000

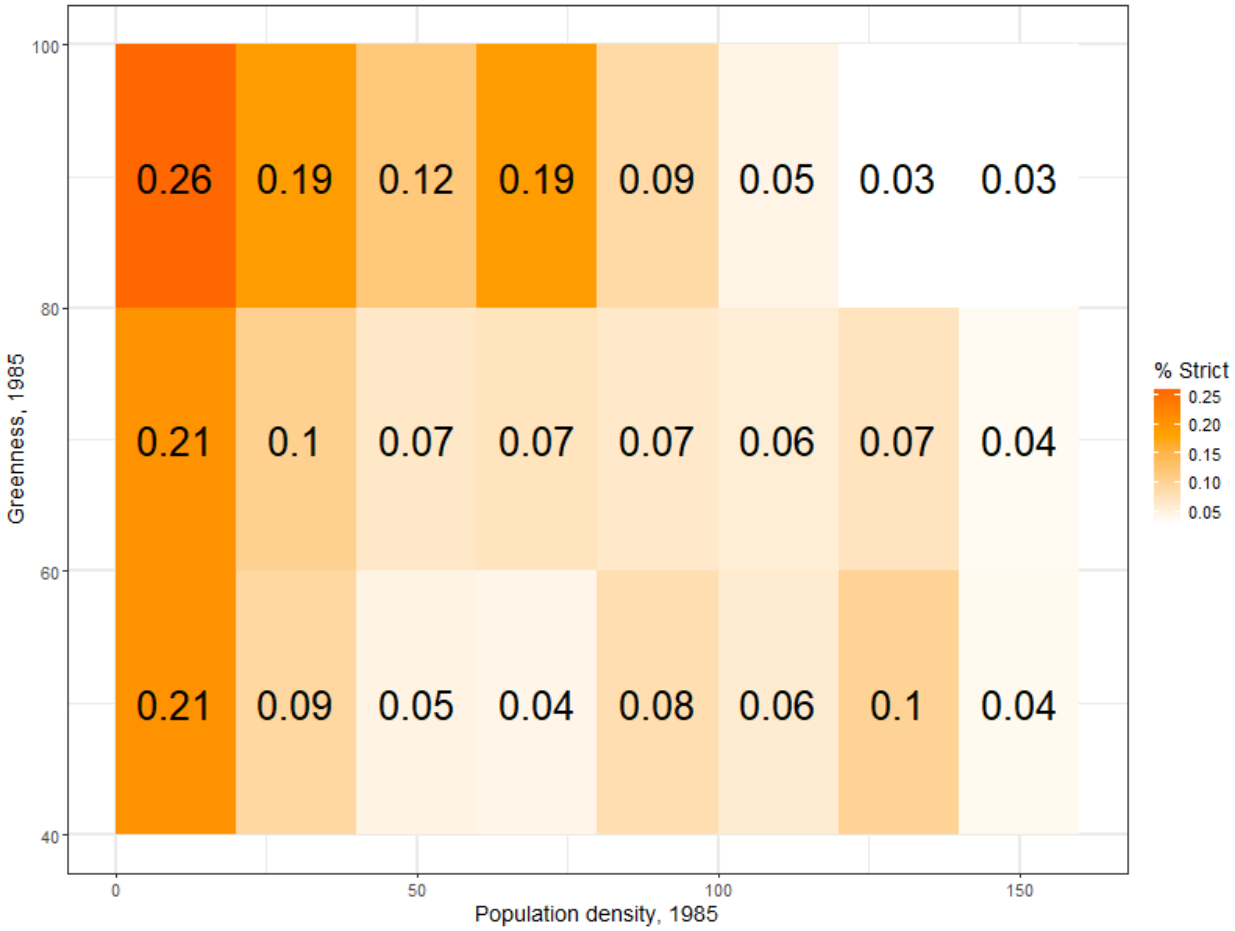
NOTES: Both panels plot estimates of the conditional average treatment effect across 20 ventiles of a key explanatory variable. Bar at ventile $q \in \{0.05, 0.1, \ldots, 0.95\}$ corresponds to the CATE of plot i conditional on a realization of X above that quantile, e.g., $\mathbb{E}[\tau_i | X > x(q)]$. Red dotted line highlights a 0 CATE. Confidence bands are 95% confidence intervals constructed through a doubly-robust procedure. Population density in people per square kilometer from Gridded Population of the World data; CATE computed using a cross-section in 2000.

Figure 6: Share of protected areas in population density and starting greenness bins



NOTES: Population density from the Global Human Settlement Layer and greenness from Landsat 5. Both are divided into increments of 20, from $[0, 160]$ and $[40, 100]$, respectively. Resolution of a grid cell is 300 $\times$ 300 meters. Grid cells are considered protected if any non-zero share of the grid cell protected. Greenness-by-population density cells with fewer than 100 grid cells are omitted. This removes areas with greenness higher than 90. Sample includes all land which had a greenness of at least 40 at some point in the sample.

Figure 7: Share of protected areas by starting greenness and population density which are strict



NOTES: Figure is constructed exactly following Figure 6. Entries are populated with the area-weighted average strictness of protected areas in each cell. Strictness is calculated using the International Union for the Conservation of Nature (IUCN) classification, with Ia, Ib, and II areas mapping to 1, and III or below mapping to 0. Ia, Ib, and II areas place the strictest limitations on human activity (only scientific, ecological, and recreational use, respectively). Thus, stricter areas have high (more red) values of the strictness index.

Table 1: Comparison of discrete land use transitions in protected areas and never-protected 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 the base year of 1980 in 2010. We tabulate transitions by treatment status: treated units are treated in 1970-80 and have at least 50% of landmass in a protected area. Controls are matched on pre-1970 observables. Transitions are defined based on the HILDA land cover data, which classifies land into one of six land categories in each decade from 1900 to 2010. We omit a small percentage of areas which are classified as 'NA'.

Table 2: Average of key variables among land classified as forest in 1900 by whether that land was deforested, 1900-2010

	Not deforested, 1900-2010		Deforested, 1900-2010	
	Mean	Std. dev.	Mean	Std. dev.
Nightlights in 2010	6.3	9.3	10.8	12.1
Percent of grid protected	24.2	45.9	28.7	46.0
Population density, 2000	46.7	143.4	103.8	273.8
Crop suitability	6.2	1.3	6.3	1.3
Forest suitability	3.5	1.5	4.0	1.4
Grassland suitability	5.6	1.4	5.6	1.6
Slope steepness	2.3	3.1	2.0	3.0
Solar radiance	10.7	2.2	11.4	2.0
Precipitation	767.3	266.9	791.2	261.4
Potassium	147.1	55.9	175.9	71.5
Nitrogen	2.2	1.0	2.3	1.0
Growing season length	222.5	67.0	256.0	57.1

NOTES: Table presents balance of several key variables against an indicator based on the discrete land-use classifications provided by the HILDA data. Not deforested indicates 1 km $\times$ 1 km grid cells which were coded as forest (333) in 1900 and were still coded as forest (333) in 2010. Deforested indicates areas which were cropland (333) in 1900 but were coded as any other category in 2010. Percent of grid protected indicates the percent of the 1 square kilometer grid which contains protected areas, regardless of their designation year and CDDA designation. Means of time-varying variables are calculated in a specific cross-section, as indicated in the table. Units for all variables are indicated in Appendix Table A.3.

Table 3: Percent of grid cells with given starting land use in 1900 and protected after 1900 broken out by final land use in 2010

Land use 1900 / Land use 2010		Cropland	Forest	Grassland	Settlement
Cropland	% row	59.7	15.3	20.2	4.8
Forest	% row	0.6	87.3	11.2	0.9
Grassland	% row	8.8	27.6	62.0	1.5
Settlement	% row	5.9	6.7	6.9	80.5

NOTES: Percentages indicate the fraction of grid cells which were protected after 1900 across the European Union at a 1 km resolution. Rows indicate land use 1900 and columns indicate land use in 2010. Percentages add up to 1 in the rows, indicating the percent of land which was in a given land use type in 1900 and protected between 1900-2010 which ended up in the column's land use type in 2010. Omits "other" and "water" categories: 0% of any presented land use in 1900 which transitioned to omitted categories in 2010 was protected.

Table 4: Conditional average treatment effect in each IUCN category

IUCN category	CATE estimate (Std. dev)
Strict nature reserves and wilderness areas	−0.096 (0.097)
National parks	−0.129 (0.026)
Natural monuments	−0.210 (0.025)
Species mangement areas	−0.197 (0.028)
Protected landscapes	−0.142 (0.031)
Sustainable resource use areas	−0.223 (0.025)

NOTES: Table summarizes conditional average treatment effects as a function of protected area strictness. CATEs are estimated using the Wager and Athey (2018) random forest estimator on a 5% sample of the EU, clustered at the grid-by-country-by-cohort level. Sample restricted to areas with overlap in propensity scores with controls, using a doubly-robust specification.

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 PAs	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 (Cornes et al. 2018)	European climate surface observations	Link to source
[7]	ESDAC (Ballabio et al. 2019)	Soils and topography	Link to source
[8]	ESDAC (Günther et al. 2014)	Climate-physical regions	Link to source
[9]	SEDAC (Warszawski et al. 2017)	Gridded population data	Link to source
[10]	ECMWF	Annual total precipitation	Link to source
[11]	HILDA (Winkler et al. 2020)	Discrete land-use data	Link to source
[12]	World Clim (Fick and Hijmans 2017)	Solar radiation and precipitation	Link to source
[13]	Li et al. 2020a, Li et al. 2020b	Harmonized global night time lights	Link to source
[14]	BioTIME (Dornelas et al. 2018)	Species count data	Link to source
[15]	ESDAC (Panagos et al. 2015)	Slope steepness and elevation	Link to source
[16]	ESDAC (Tóth and Hermann 2016)	Soil suitability	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.

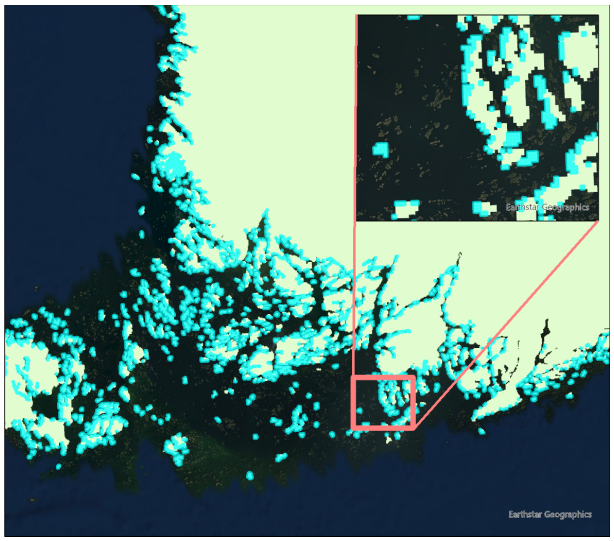
A.2 Bio-geographical regions and climate zones

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	WorldClim
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).

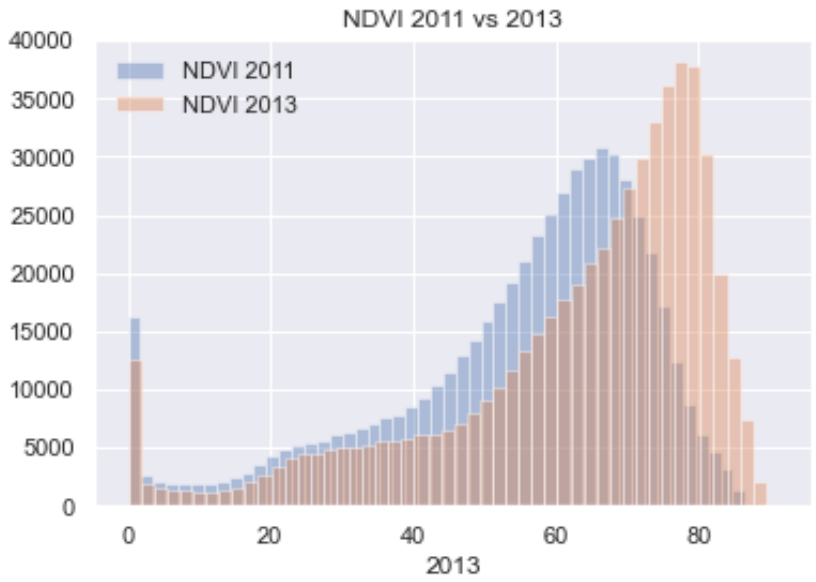
B Econometrics

B.1 Estimation details: Staggered difference-in-differences design

We use the 'did' package in the R programming language.[3] 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

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

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.

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 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 and 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.

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 Appendix 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.2 Estimation details: Conditional average treatment effects (CATE)

We use the CATE estimator provided in the R package 'grf', generalized random forest.[4]. 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. (2024) 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

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

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. (2024) 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' command, subsetting to the portion of data of interest.

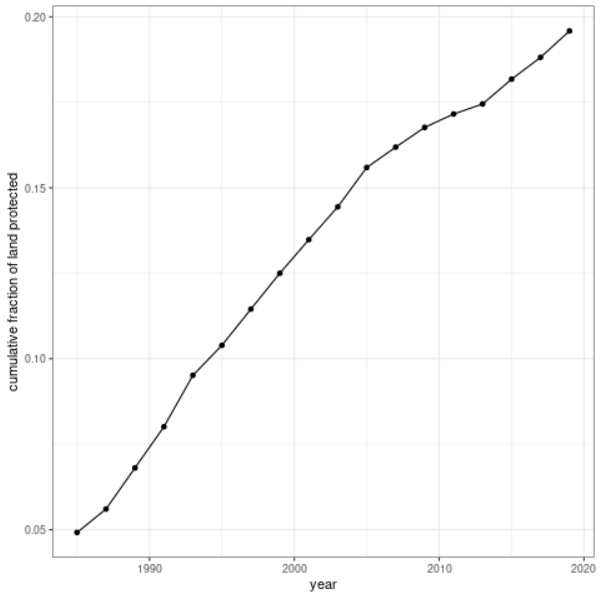
C Additional tables and figures

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 was protected (as of 2023, this percentage had increased to 26%). 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 a large 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

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

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 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.

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

Code	IUCN category	Percent	Mean greenness
Ia, Ib	Strict nature reserve or wilderness area	7.02	70.80
II	National park	0.55	56.32
III	Natural monument	4.35	63.20
IV	Species management area	46.58	66.52
V	Protected landscape	14.02	62.49
VI	Protected area with sustainable use of natural resources	1.39	63.18
	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.

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

	Never-treated		Treated	
	N	Mean (Std. dev.)	N	Mean (Std. dev.)
Greenness	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.

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 starting 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.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

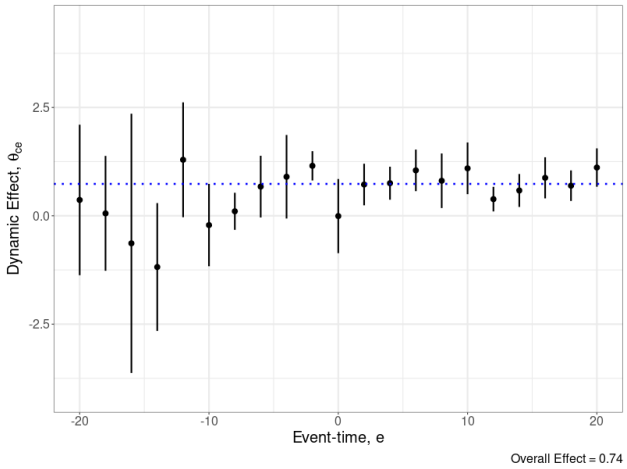
NOTES: Treated data are aggregated across all cohorts. The sample selection procedure for this table is the same as the rest of the paper. Plots which were protected before 1990 are trimmed to ensure we can compare several periods' worth of pre-trends. Plots which never attain a greenness value above 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=3,819,636)		Treated (N=64,146)		Raw		IPW	
	Mean	Std. dev.	Mean	Std. dev.	Diff. in means	Std. error	Diff. in means	Std. error
Greenness	54.63	15.47	66.71	12.45	12.08	0.05	0.61	0.06
Population density	106.52	387.65	45.00	113.31	-61.52	0.40	4.40	0.73
High-value farmland fraction	0.13	0.29	0.32	0.38	0.19	0.001	0.13	0.001
Rain	26.83	6.42	33.62	7.37	6.79	0.029	0.13	0.028
Heating degree days	1740.86	653.12	2380.81	792.37	639.95	3.11	-33.42	2.96
Growing season length	295.80	40.81	256.75	42.69	-39.05	0.17	-1.23	0.16
Soil suitability index	3.65	0.75	3.42	0.89	-0.23	0.003	-0.17	0.003
Slope angle	2.00	1.42	2.77	1.66	0.77	0.01	0.26	0.01
Slope steepness	1.48	2.39	2.51	3.08	1.03	0.01	0.59	0.02
Elevation	291.32	334.07	741.13	569.75	449.81	2.24	-19.13	1.47
Solar radiance	12324.24	1150.89	12911.23	1242.07	586.99	4.87	163.13	5.39

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, and a cubic 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.

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



NOTES: Plots estimates and confidence bands for θ_{ce} for Poland. The Poland-wide treatment effect is 0.74 (0.15), indicated by the dashed blue line. Confidence bands are based on a bootstrapping procedure discussed in Appendix B.

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; a few less-represented or smaller countries such as the Netherlands or Serbia are less precise. The single well-powered exception is Slovenia, robustly greater than one unit of greenness is gained on average due to protection on around 0.2% of the EU's total protected landmass.

We also 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 $\hat{\theta}_{ce}$ 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.

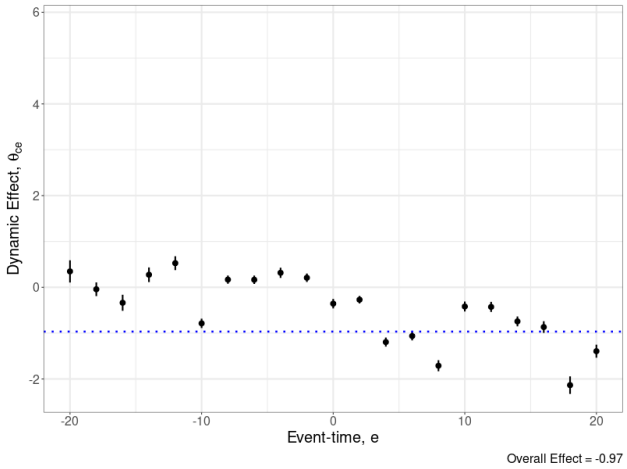
We provide a second example of parallel trends. Spain is in a different bio-geographical region and climate zone than Poland. It is also less densely vegetated. In Figure A.5, the trend up to 10 years prior to protection is flat and close to 0. If anything, the pre-trends suggest that the aggregate treatment effect would be higher by 0.03 NDVI points due to a deviation from parallel trends. We see that up to 20 years after protection, treatment effects hover around zero, and are negative on average. Note that both this warmer and sunnier climate and the densely-forested

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

	ATT estimate (Std. dev)	Non-missing, matched grid cells
CSDR	0.03 (0.05)	75,371,365
MTWFE	0.45 (0.32)	75,371,365
TWFE	0.41 (0.32)	75,371,365
Albania	−0.19 (0.10)	292,586
Austria	0.21 (0.02)	720,443
Belgium	0.62 (0.05)	312,941
Bosnia	0.59 (0.42)	3,562
Bulgaria	−0.15 (0.03)	1,078,545
Croatia	0.83 (0.07)	495,051
Czech	−0.35 (0.05)	709,934
Denmark	−1.09 (0.30)	412,549
Estonia	0.86 (0.03)	327,007
Finland	−0.02 (0.13)	4,350,516
France	−0.30 (0.05)	8,106,644
Germany	0.16 (0.04)	16,442,388
Greece	−0.34 (0.02)	905,270
Hungary	0.65 (0.03)	868,039
Ireland	−1.18 (0.08)	563,515
Italy	0.30 (0.07)	2,968,431
Latvia	0.29 (0.08)	631,576
Lithuania	0.72 (0.06)	668,860
Luxembourg	−0.45 (0.09)	26,683
Macedonia	−0.84 (0.78)	1,377
Montenegro	0.59 (0.16)	142,101
Netherlands	3.86 (1.55)	346,906
Poland	0.74 (0.15)	13,179,276
Portugal	−0.42 (0.15)	602,637
Romania	0.32 (0.09)	2,036,628
Serbia	−2.27 (1.61)	3,516
Slovakia	0.83 (0.06)	474,177
Slovenia	1.73 (0.07)	191,942
Spain	−0.97 (0.04)	7,504,290
Sweden	−0.78 (0.04)	9,826,314
Switzerland	0.33 (0.34)	4,047
United Kingdom	−0.85 (0.05)	1,173,614

NOTES: Average treatment effect θ_c of conservation on vegetation greenness (an index varying between 0 and 100) in Equation (3) estimated within each country in the European Union using data from 1985-2019 on a biannual basis. Top three rows report the Callaway and Sant'anna, doubly-robust estimator (CSDR), matched two-way fixed effects (MTWFE), and two-way fixed effects without matching (TWFE), 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 $\theta_c \leq 0$. "Non-missing, matched grid cells" represents the number of unique grid cells that were identified with non-missing matching variables across foundation years between 1991 and 2019.

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



NOTES: Plots estimates and confidence bands for θ_{ce} for Spain. The Spain-wide treatment effect is -0.97 (0.04), indicated by the dashed blue line. Confidence bands are based on a bootstrapping procedure discussed in Appendix B.

Poland demonstrate similar conclusions: very little positive 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 dynamic treatment 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. Because of the variation in treatment timing, very early and very late treatment bins have fewer observations, meaning that the estimated dynamic treatment effect is often noisier (e.g., in Table A.9, there are fewer than 1 million treated parcels outside of $[-26, 26]$, and more than 1 million elsewhere). We trim the plot window to $[-10, 20]$ for visual clarity as standard errors grow large outside this window.

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

	Mean (Std. dev.)	Num. treated grids
-34	−0.01 (0.00)	287,090
-32	0.14 (0.38)	543,676
-30	0.97 (0.20)	798,242
-28	−0.34 (0.77)	909,608
-26	−0.62 (0.10)	1,053,996
-24	−0.64 (0.19)	1,272,564
-22	−0.18 (0.09)	1,497,380
-20	0.02 (0.22)	1,969,748
-18	0.07 (0.13)	2,322,092
-16	0.03 (0.26)	2,682,074
-14	0.02 (0.18)	2,987,024
-12	−0.12 (0.10)	3,365,991
-10	−0.09 (0.07)	3,630,037
-8	0.41 (0.09)	4,183,559
-6	0.21 (0.25)	4,669,607
-4	0.25 (0.20)	4,669,184
-2	−0.12 (0.07)	4,664,460
0	0.21 (0.14)	4,667,137
2	−0.13 (0.24)	4,380,016
4	0.27 (0.29)	4,122,379
6	0.10 (0.12)	3,867,114
8	0.39 (0.17)	3,753,652
10	−0.06 (0.25)	3,611,444
12	0.39 (0.27)	3,391,546
14	−0.14 (0.19)	3,166,191
16	0.46 (0.13)	2,696,723
18	−0.86 (0.12)	2,346,079
20	0.54 (0.66)	1,986,763
22	−0.27 (0.05)	1,681,891
24	−0.43 (0.40)	1,302,535
26	−1.31 (1.35)	1,039,621
28	−0.18 (0.00)	486,102

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 $\theta_c(g,t)$. 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. Number of treated grids refers to the number of unique 300 square meter grid cells in our sample which were protected at a time such that $e = t - g$ was observed (e.g., for $e = -34$, this will be the number of CDDAs founded in 2019, the last year of the data). To convert to area in square kilometers, multiply the number of grid cells by 0.09.

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

	Mean (Std. dev.)	Num. treated grids
1991	−0.09 (0.31)	7,541,726
1993	0.02 (0.45)	5,994,594
1995	0.00 (0.31)	6,162,884
1997	−0.07 (0.28)	5,585,980
1999	0.52 (0.46)	6,957,316
2001	−0.46 (0.12)	5,000,460
2003	0.32 (0.23)	3,499,033
2005	0.18 (0.23)	5,359,575
2007	0.30 (0.35)	2,469,672
2009	−0.02 (0.15)	2,083,330
2011	0.12 (0.22)	1,409,957
2013	−0.17 (0.21)	1,353,424
2015	−0.30 (1.03)	2,901,151
2017	0.64 (0.26)	3,791,307
2019	−0.26 (0.08)	4,167,597

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 $\theta_c(g,t)$. 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).

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 treatment effects again center around zero, 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 presents cohort-level effects aggregated to the country level (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 (6), on cohorts $g - 1991$. we difference out 1991 so changes are interpreted as the effect of being treated two 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 positive trends is in Albania (0.17). Albania's trend suggests that on average, plots treated in the last cohorts in 2019 experienced a treatment effect 2.4 units higher than those treated in 1991. 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.

Two-way fixed effects

For completeness, we report the dynamic effects estimated by a classic matched, two-way fixed effects estimator. The estimating equation includes grid i and time t fixed effects:

$$Y_{it} = \beta D_{it} + \lambda_i + \lambda_t + \epsilon_{it} \tag{A.1}$$

Unlike the Callaway and Sant'Anna estimator, the classic matched two-way fixed effects estimator uses a single matching function rather than cohort-specific matching functions. Thus, the TWFE matching cannot address between-cohort variation in selection.

Figure A.6 plots the event-study coefficients. We make two remarks. First, the two-way fixed effects estimates indicate a long-run pre-trend compared to the Callaway and Sant'Anna methodology. This pre-trend is addressed in the Callaway and Sant'Anna methodology through (1) cohort-specific matching and (2) avoiding forbidden comparisons (Goodman-Bacon 2021). In our setting, absent a treatment effect after protection, we believe (1) is the primary channel through which the conventional TWFE estimator introduces bias. Second, the two-way fixed effects estimates at best indicate EU-wide greening of less than 0.45 NDVI points on average can be attributed to protection. Of course, this number is difficult to causally interpret in the presence of a pre-trend and statistical imprecision.

C.4 Treatment effect estimates: nightlights

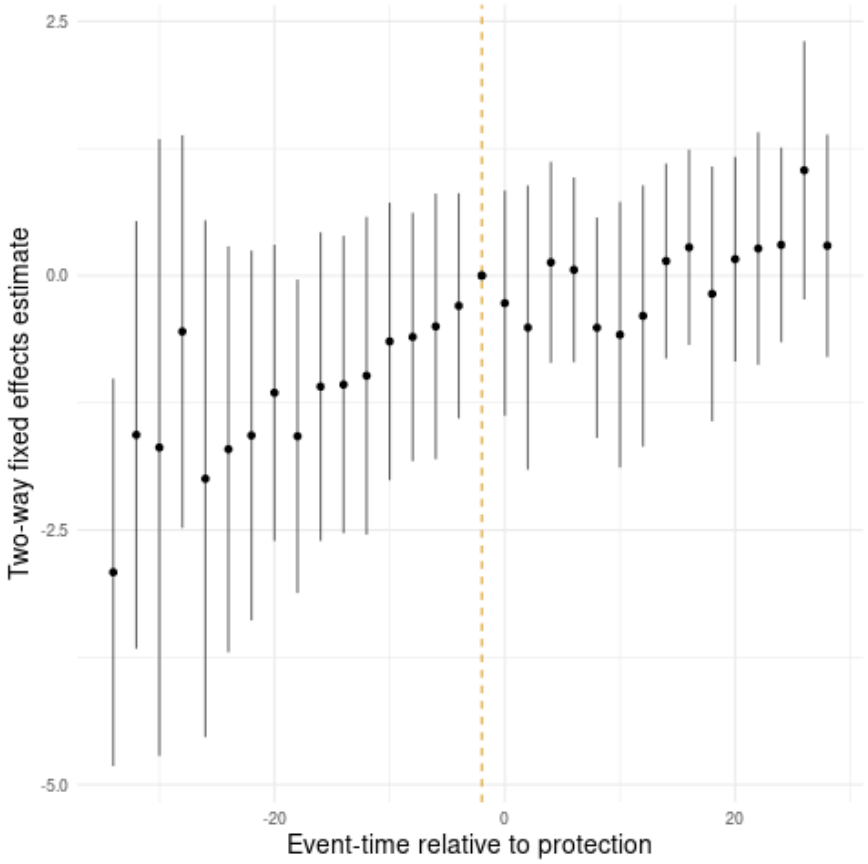
Nightlight outcomes are measured on a luminosity scale ranging from 0 to 68. Across the EU in 2021, the unconditional average luminosity is 9.9. However, 30% of our sample has 0 luminosity (at a 1-kilometer grid cell resolution). A 0 requires neither indoor nor outdoor electricity: these 0s correspond 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. In this paper, we do not interpret nightlights as a measure of economics activity; we are interested in its direct measurement of human presence.

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

	Number of cohorts	Trend: Estimate (Std. dev.)	Split difference: Estimate (Std. dev.)
Albania	10	0.17 (0.0172)	1.22 (1.54)
Austria	15	0.00 (0.0018)	−0.72 (0.80)
Belgium	14	−0.05 (0.0005)	−0.93 (0.60)
Bulgaria	13	0.13 (0.0031)	1.16 (1.02)
Croatia	12	−0.02 (0.0014)	−1.75 (1.43)
Czech	14	−0.01 (0.0011)	−0.76 (0.94)
Denmark	11	−0.04 (0.0082)	1.30 (2.14)
Estonia	14	0.11 (0.0028)	1.92 (0.81)
Finland	15	0.00 (0.0002)	0.03 (0.25)
France	15	0.01 (0.0005)	−0.39 (0.55)
Germany	15	0.01 (0.0008)	0.01 (0.59)
Greece	11	0.05 (0.0010)	0.70 (0.53)
Hungary	13	0.02 (0.0040)	0.89 (1.23)
Ireland	7	−0.07 (0.0239)	−0.63 (2.37)
Italy	10	0.05 (0.0057)	0.59 (0.86)
Latvia	8	−0.20 (0.0113)	−1.52 (1.91)
Lithuania	7	−0.08 (0.0011)	−1.42 (0.69)
Luxembourg	7	−0.06 (0.0010)	−0.63 (0.80)
Montenegro	5	−0.05 (0.0456)	1.92 (3.07)
Poland	15	−0.04 (0.0004)	−0.65 (0.44)
Portugal	12	−0.08 (0.0017)	−1.65 (0.84)
Romania	11	0.00 (0.0007)	−0.34 (0.98)
Slovakia	14	0.01 (0.0019)	−0.15 (1.01)
Slovenia	14	−0.01 (0.0008)	−0.19 (0.51)
Spain	15	0.02 (0.0012)	0.01 (0.63)
Sweden	15	−0.04 (0.0012)	−0.61 (0.59)
Switzerland	9	0.04 (0.0043)	0.09 (1.94)
UK	15	0.07 (0.0032)	1.97 (1.01)

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} on 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. Countries require at least 3 cohorts' worth of data to be included: Bosnia, the Netherlands, Macedonia, and Serbia are excluded.

Figure A.6: Dynamic estimates from matched two-way fixed effects estimation



NOTES: Figure plots estimates from Equation (A.1). Standard errors are clustered at the unit-by-foundation year level to be comparable with Callaway and Sant'Anna results. The dashed orange line indicates the omitted pre-period, $t - g = -2$. Units are matched on a cubic polynomial in greenness from 1985-1989, weather variables from 1985, and time-invariant controls (the same matching function was used in the Callaway and Sant'Anna exercise).

We provide some details regarding land protection given this data. Appendix Table A.12 constructs a linear probability model of land protection during our sample years, 1992-2019. We use the linear probability model to lend some simple interpretation to our coefficients, though the actual propensity score is generated via a logistic regression. The outcome is an indicator variable indicating any protection after 1992, creating a cross-sectional propensity score model. Column (1) focuses only on the land use recorded in the most recent decade (1990). Protection tends to occur at higher rates among land which starts out as forest, grassland, or "other," with settled and agricultural areas having statistically significantly fewer protected areas founded after 1992. Columns (2) and (3) instead focus on key (logged) covariates, many of which may correlate with underlying land use. From column (2), an apparent contradiction emerges: settlement and cropland are less likely to get protected than "natural" land use, but nightlights and growing season length seem to increase the probability of protection. Composition effects explain the contradiction: e.g., nightlights, pooled across Europe, will reflect not only comparisons between rural and urban areas within countries but also comparisons between the more developed west and the less "luminous" east. Thus, column (3) re-runs the exercise in (2) with a country fixed effect. Notably, these two variables change sign, suggesting that holding fixed the protection regime, more economically valuable land generates a lower probability of protection.

Next, 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.13 indicates the average treatment effects from the dynamic aggregation θ_e^{EU}, corresponding to the left panel of Figure 4. Standard errors are bootstrapped. At event-time 18, the estimates drop slightly, but this effect is inconsistent over the remaining periods, rising to -0.38 in event time 20 yet dropping to -0.07 by the last event time. This suggests 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.14 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 persistently-positive treatment effects, although there are some larger positive and negative estimates. Compositionally, we find no heterogeneity in treatment effects across any observables in later exercises, making these two cases outliers. Indeed, one is nearly as likely to find a positive effect on nightlights as a negative one. There is no discernible pattern in selection that indicates any systematic variation in these nightlights effects, either.

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

Table A.12: Linear probability model predicting land protection after 1992

	(1)	(2)	(3)
Constant	0.1318***	0.1602***	
	(0.0004)	(0.0050)	
Settlement in 1990	0.0047***		
	(0.0009)		
Cropland in 1990	-0.0377***		
	(0.0005)		
Forest in 1990	0.0305***		
	(0.0005)		
Grassland in 1990	0.0167***		
	(0.0005)		
Other in 1990	0.0821***		
	(0.0016)		
Water in 1990	0.0648***		
	(0.0012)		
Nighlight luminosity, 1992		0.0035***	-0.0009***
		(0.0002)	(0.0002)
Elevation		-0.0054***	0.0050***
		(0.0002)	(0.0003)
Distance to shore		0.0110***	-0.0171***
		(0.0002)	(0.0002)
Solar radiance		-0.2013***	-0.0686***
		(0.0016)	(0.0035)
Growing season length in 1992		0.0637***	-0.0105***
		(0.0011)	(0.0020)
R^2	0.00732	0.01289	0.10225
Observations	5,123,083	2,285,588	2,285,588
Country fixed effects			✓

NOTES: Land protection after 1992 defined as at least 50% of a 1 square kilometer grid cell being protected by the end of the study period in 2019. Standard errors indicate heteroskedasticity-robust errors. Land-use covariates come from the HILDA landcover dataset. Covariates are expressed in logs, so all coefficients are interpretable as percent changes.

Table A.13: Dynamic treatment effects on nightlight luminosity across the EU

	Mean (Std. dev)	Num. treated grids		Mean (Std. dev)	Num. treated grids
-26	0.22 (0.09)	53,798	1	−0.05 (0.02)	676,311
-25	−0.08 (0.05)	83,481	2	−0.01 (0.02)	640,168
-24	−0.20 (0.04)	95,401	3	−0.05 (0.02)	610,006
-23	−0.11 (0.03)	127,163	4	−0.08 (0.03)	597,909
-22	0.14 (0.03)	148,223	5	−0.10 (0.03)	565,977
-21	0.07 (0.03)	159,538	6	−0.15 (0.03)	544,846
-20	−0.04 (0.03)	169,934	7	−0.20 (0.03)	533,222
-19	0.04 (0.03)	185,662	8	−0.14 (0.02)	507,927
-18	0.06 (0.03)	201,941	9	−0.06 (0.03)	492,294
-17	−0.02 (0.03)	241,016	10	−0.09 (0.03)	475,973
-16	−0.03 (0.03)	255,909	11	−0.01 (0.03)	452,004
-15	−0.03 (0.03)	272,727	12	−0.05 (0.03)	436,424
-14	−0.18 (0.03)	299,729	13	0.00 (0.03)	419,560
-13	0.03 (0.03)	326,587	14	0.01 (0.04)	392,458
-12	0.00 (0.02)	376,016	15	0.09 (0.04)	366,014
-11	−0.08 (0.02)	410,306	16	0.19 (0.04)	317,202
-10	−0.11 (0.02)	442,975	17	0.03 (0.05)	282,830
-9	0.05 (0.02)	469,374	18	−0.30 (0.06)	250,867
-8	−0.03 (0.02)	504,864	19	−0.28 (0.06)	224,830
-7	−0.07 (0.02)	508,328	20	−0.38 (0.07)	189,581
-6	−0.09 (0.02)	566,392	21	−0.32 (0.08)	170,944
-5	0.04 (0.02)	604,843	22	−0.14 (0.08)	128,089
-4	−0.13 (0.02)	627,151	23	−0.39 (0.09)	89,553
-3	−0.08 (0.02)	659,784	24	−0.14 (0.11)	67,312
-2	0.01 (0.02)	676,419	25	−0.23 (0.18)	34,345
-1	−0.07 (0.02)	694,049	26	−0.07 (0.24)	17,223

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 Appendix 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.

Table A.14: Cohort-level treatment effects on nightlight luminosity across the EU

	Mean (Std. dev)	Num. treated grids
1993	0.09 (0.14)	482,244
1994	0.09 (0.05)	480,592
1995	−0.33 (0.06)	925,014
1996	0.14 (0.07)	623,146
1997	−0.25 (0.07)	1,077,606
1998	−0.45 (0.08)	1,200,292
1999	−0.56 (0.10)	521,881
2000	0.19 (0.06)	991,106
2001	−0.22 (0.13)	733,487
2002	0.71 (0.09)	909,630
2003	0.13 (0.09)	957,664
2004	0.01 (0.07)	1,381,841
2005	−1.79 (0.17)	746,899
2006	−0.29 (0.09)	758,410
2007	0.79 (0.15)	463,849
2008	−0.41 (0.24)	417,004
2009	0.47 (0.11)	1,032,970
2010	−0.05 (0.05)	455,812
2011	0.12 (0.08)	440,088
2012	−0.01 (0.17)	290,957
2013	−1.03 (0.20)	316,812
2014	0.34 (0.07)	589,656
2015	−0.34 (0.13)	889,336
2016	−0.06 (0.06)	333,308
2017	−0.04 (0.10)	831,124
2018	−0.54 (0.06)	1,016,820
2019	−0.20 (0.12)	485,128

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.

Athey (2018), which amounts to a random forest implementation of Chernozhukov et al. (2024). The results are shown in Table A.15. The test reports two coefficients in a regression. Observed greenness Y is 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 predicted by the random forest is not driving deviations from the average treatment effect. Our results indicate our measured CATE is accurate, with a coefficient of 25.32 (2.49). Thus, our random forest has picked up on statistically significant deviations from the average treatment effect.

Table A.15: Test of the random forest model calibration

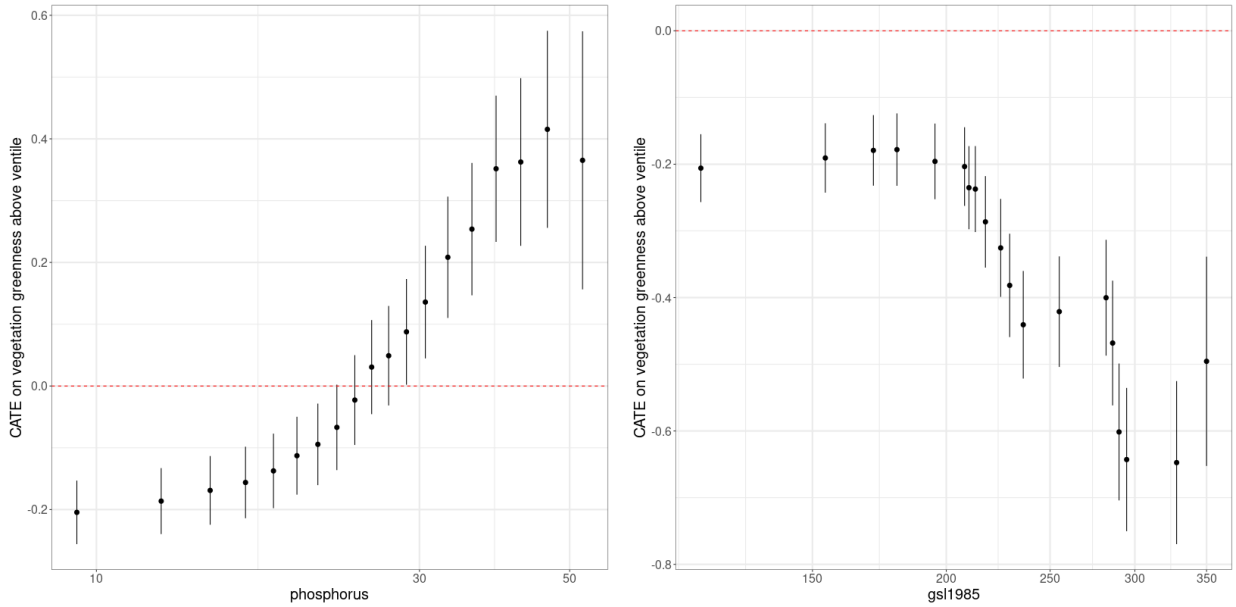
	Coefficient	Standard error
Mean forest prediction	1.08	(0.10)
Differential forest prediction	25.32	(2.49)

NOTES: Table presents the test in Wager and Athey (2018) for heterogeneous treatment effects. 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 in Figure A.7.

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, this 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 to a valuable counterfactual land value. We present the phosphorus CATE plot in Figure A.7 (left

Figure A.7: Distribution of the conditional average treatment effect for phosphorus content in soil and growing season length in 1985



NOTES: Both panels plot estimates of the conditional average treatment effect across 20 ventiles of a key explanatory variable. Bar at ventile $q \in \{0.05, 0.1, \ldots, 0.95\}$ corresponds to the CATE of plot i conditional on a realization of X above that quantile, e.g., $\mathbb{E}[\tau_i | X > x(q)]$. Red dotted line highlights a 0 CATE. Confidence bands are 95% confidence intervals constructed through a doubly-robust procedure.

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 of less than half of an NDVI point larger than the bottom 25%. This result is inconsistent with a pattern of high agricultural value soil returning to natural use.

As a second example, we illustrate the conditional average treatment effect as a function of the growing season length in 1985 in the right panel of Figure A.7. Longer growing seasons indicate weather conditions which are more conducive to growing common fieldcrops (though not specialty crops or horticulture). Plots in the top 25% of the sample with respect to their 1985 growing season length have a conditional average treatment effect of -0.5, which is statistically more negative than the full-sample ATE of -0.2. Thus, more agriculturally suitable land for field crops was less likely to experience beneficial effects of land protection, though the difference is small. Among all of our variables, those presented in the main text and appendix had the largest positive trends in treatment effects. None of them indicate a CATE larger than 1.

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 due to land protection. Here, we test a first difference. There are two advantages to this robustness check. First, as vegetation growth tends to be a slow process, there is an advantage to testing for a change in the first differences of vegetation rather than the levels. Where in levels it can take a very long time to convert to biodiversity-rich, green forest from a low-greenness land cover like grassland, in first differences a clear change in growth rates should be easier to detect earlier in the process of vegetation regrowth. Second, the first-differences specification allows us to check that our levels functional form does not drive the null result. By taking a first difference, we are removing the effect of any time-invariant contributors to greenness. This approach is thus robust to arbitrary time-invariant heterogeneity, at the cost of correctly specifying determinants of changes in greenness. 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).

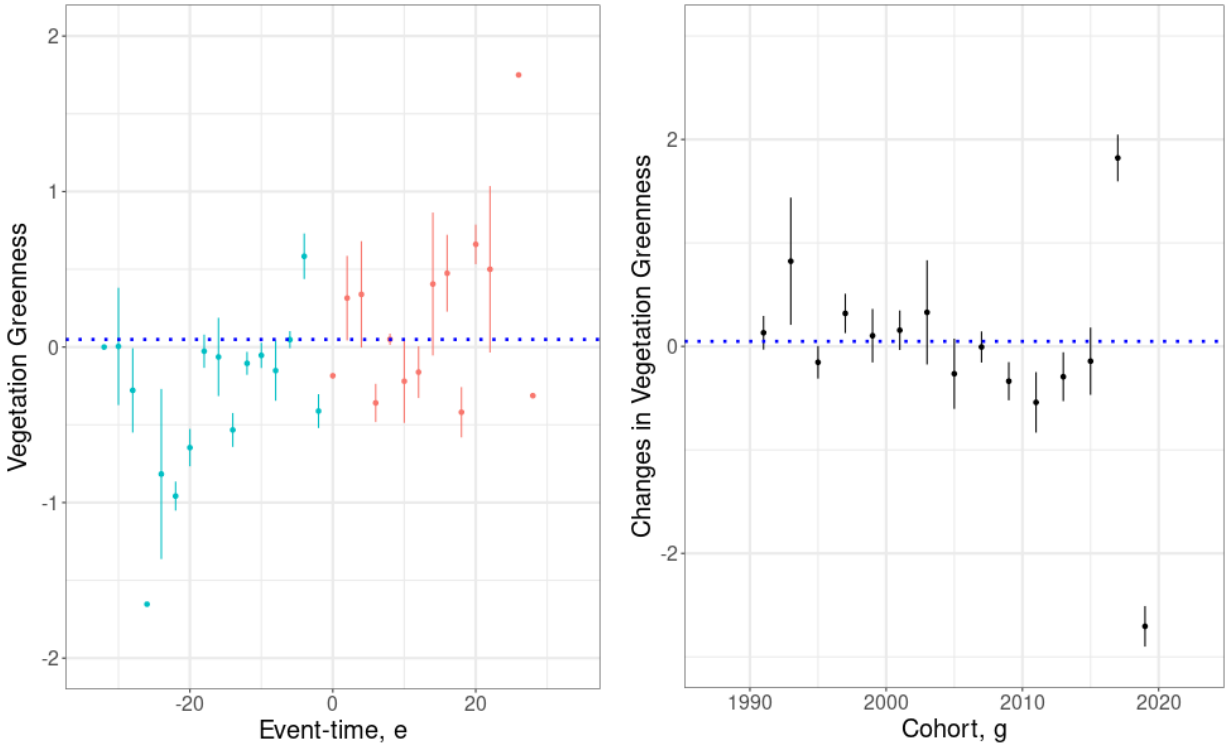
We report the EU-wide aggregated treatment effects in Figure A.8. The left panel illustrates the dynamic treatment effects on changes in vegetation greenness, θ_t^{EU}. We can see weak evidence of a positive trend in the slope of greenness for the full range of event times. There is no evidence of a 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. However, we see no persistent positive kink in first differences: indeed, we see about as many negative as positive treatment effects after treatment. The right panel of Figure A.8 illustrates the cohort-level treatment effects. Aside from an outlier in 2017—which only has one post-treatment period—we see very weak evidence for a negative trend in treatment effects.

Collectively, our first-difference results emphasize the main zero treatment effect finding. There is evidence for neither a notch nor a kink in vegetation growth driven by protection.

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 have meaningful vegetation cover. 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.2) 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 zero result.

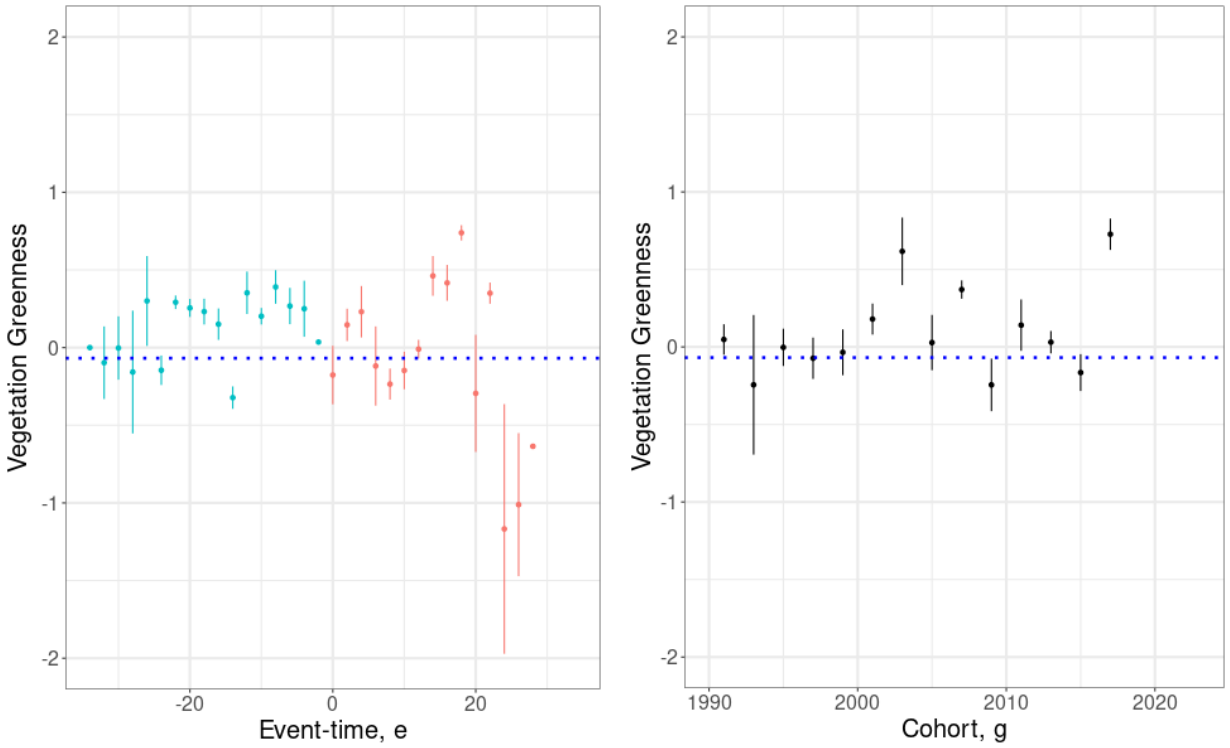
Our overall average treatment effect for an NDVI threshold of 30 is again a precise 0: $-0.07(0.02)$. The difference between this estimate and our baseline estimate is materially quite small, if precise. The direction of the change in estimates suggests that, if anything, protected land which never attained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

Figure A.8: Treatment effects on 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_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. Sample trims first calendar year, 1985, due to first differencing: we match on values from 1987 and 1989 for all cohorts treated in or after 1991. Both panels use the Callaway and Sant'Anna doubly robust estimator with bootstrapped confidence bands.

Figure A.9: Treatment effects on NDVI: expanded sample definition



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 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. Dotted blue lines indicate the EU-wide ATE of 0.05 (0.02).

When plotting EU-wide dynamic treatment effects θ_e^{EU} in the left panel of Figure A.9, we find little evidence of aggregate greening. We obtain precise balance in the 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 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 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.9, 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. 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. The advantage of SFD is to eliminate spatially correlated unobservables.[5] It ensures that difficult-to-measure but spatially correlated unobservables such as agricultural yields or the prevalence of certain vegetation species are not confounding our results. In particular, if agricultural yields are a critical control for greenness, our method does not have a perfect measure for yields and thus may not completely specify the matching function nor the outcome regression. By spatially differencing the data, we compare areas which are nearby to one another, thus indirectly controlling for these unobservable drivers. 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:

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

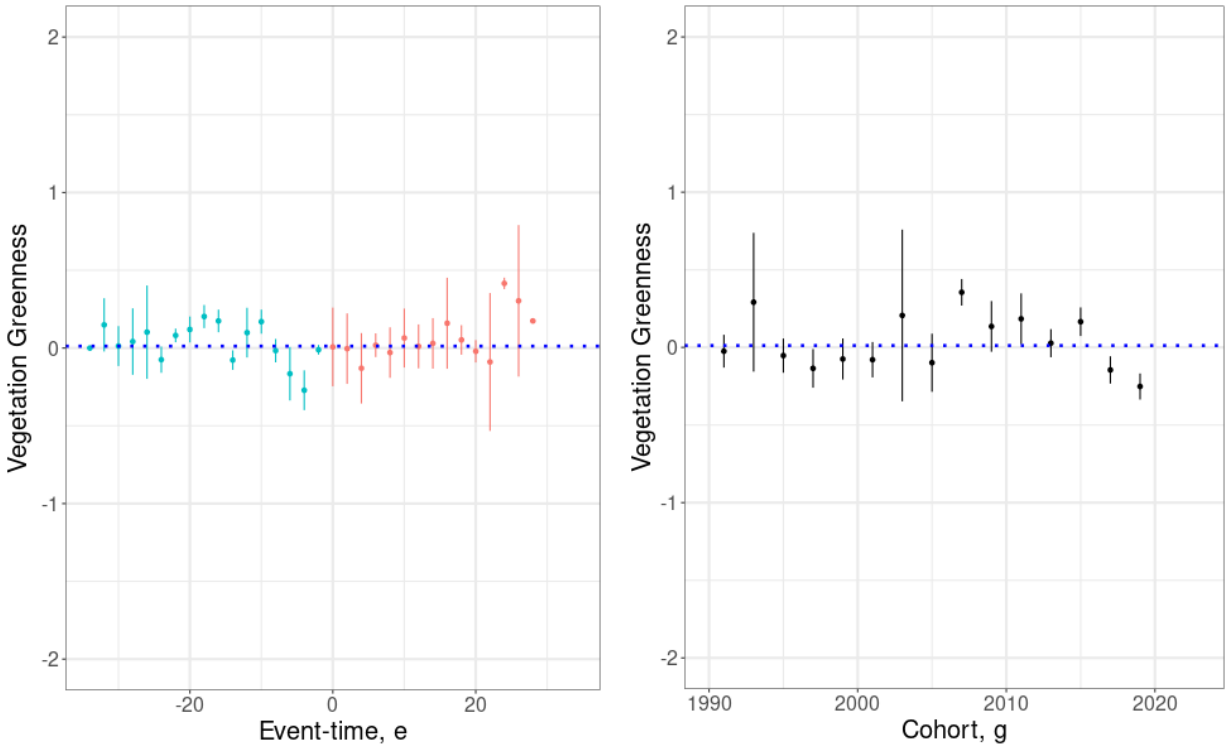
In principle, we could re-estimate SFD for many axes, such as a vertical or diagonal axis. When we estimate treatment effects, the SFD method estimates the difference-in-differences on the spatially-differenced outcome variable. Intuitively, a difference-in-differences estimator with SFD data captures the relative change in ΔY_{xyt}^{LAT} 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

5. It does not explicitly treat violations of SUTVA, where control plots nearby CDDAs may be affected by protection. A conventional method for testing for such spatial spillovers is a "donut" difference-in-differences design in which the researcher discards potentially contaminated units nearest a protected area and recomputes treatment effects. In our application, the donut approach is prohibitively expensive from a computational perspective: it requires calculating an individual buffer for each of our over 100,000 treated areas, many of which may have overlapping buffers. Moreover, violations of SUTVA should bias our results away from zero, as protection may push economic activity to the area just outside of the CDDA boundary; despite that, we find zero treatment effects of protection.

neighbors, relative to matched controls.

When calculating the spatial first difference across the x-dimension, we obtain an EU-wide treatment effect of 0.01 (0.02). Figure A.10 plots the event-study (left panel) and cohort-level aggregations (right panel). The left panel demonstrates a precise zero 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 $e = -2$. Cohort effects show no trend through 2015, with a slightly lower treatment effect in the 2017 and 2019 cohorts. Overall, the SFD estimates are an even more precise zero than our main specification.

Figure A.10: Treatment effects on NDVI: spatial first differences



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. 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 ATT of 0.01 (0.02).

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, 39 are within 5 kilometers of a protected area listed in the CDDA system. The remaining 19 studies are too few to serve as credible controls 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 a causal design.

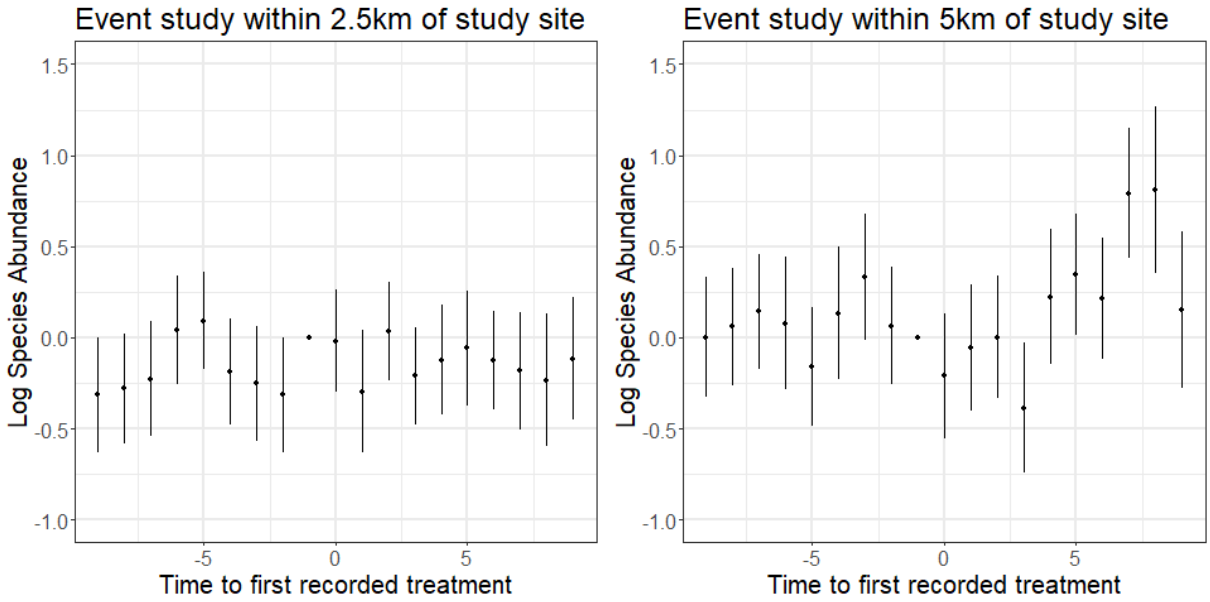
Our econometric specification considers a study i, species s, and year t. We construct a buffer of distance b kilometers around the study area, the neighborhood N_i^b. We assign treatment of the study area according to the minimum foundation year g of overlapping CDDAs. That is, the treatment indicator 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:

$$Y_{ist} = \sum_{e^b=-10}^{10} \beta_{e^b}^b D_{it}^b + \lambda_i + \lambda_s + \epsilon_{ist} \tag{A.2}$$

In Figure A.11 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 zero 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 that the biodiversity data lack spatial breadth, making valid inference difficult.

Figure A.11: Event-study coefficients measuring species abundance with respect to nearby protected area foundation



NOTES: Estimates show the effect of founding the first protected area within 2.5 km (left) and 5 km (right) of a BioTIME study site. Estimates are a pure event study. The left panel suggests that the first foundation is associated with a 0.1% decrease over 10 years post-protection; the right panel suggests this effect is closer to 0.2%.

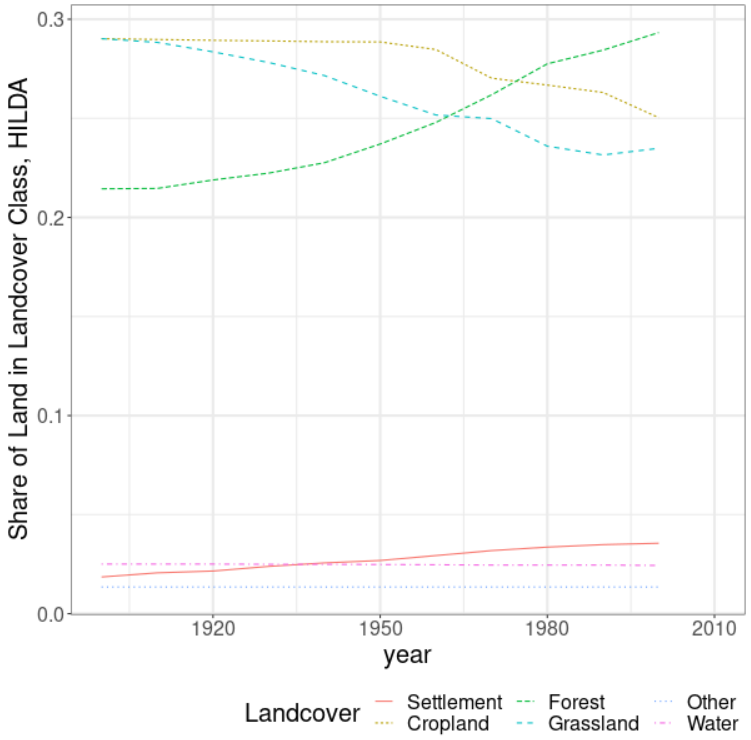
E.2 Discrete land-use data

The HILDA dataset provides over 100 years of land-cover maps at a 1 kilometer grid cell resolution. As we discuss in Section 4, 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 as mountains or barren land surface), settlement, and water. In Figure A.12 we plot the share of each land use in the HILDA data. The data is at a decadal frequency since 1900. The data confirms a century-long growth in forest cover across the EU. HILDA reports a forest share in the EU of around 22% in 1900 and a 29% forest share in 2010.

In Table A.16, 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

Figure A.12: Land share in the EU by decade and land use category



NOTES: The HILDA landcover data classify land into one of six land areas in each decade from 1900 to 2010. We omit a small percentage of areas which are classified as NA or missing values.

Table A.16: 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 or missing values.

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 became more forested over this long time horizon: 10% of cropland in 1900 appears to be forest in 2010, and 27% of grassland in 1900 is forest by 2010. This outweighs the 9% conversion from forests to grasslands. Settled areas also transition to forest—some subset of land returned to natural use.

References for Appendix

Ballabio, Cristiano, Emanuele Lugato, Oihane Fernández-Ugalde, Alberto Orgiazzi, Arwyn Jones, Pasquale Borrelli, Luca Montanarella, and Panos Panagos. 2019. "Mapping LUCAS Topsoil Chemical Properties at European Scale Using Gaussian Process Regression." *Geoderma* 355:113912.

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. 2024. "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments." Forthcoming, *Econometrica.*

Cornes, Richard C., Gerard van der Schrier, Else J.M. van den Besselaar, and Philip D. Jones. 2018. "An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets." *Journal of Geophysical Research: Atmospheres* 123 (17): 9391–9409.

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.

Dornelas, Maria, Laura H. Antao, Faye Moyes, Amanda E. Bates, Anne E. Magurran, Dušan Adam, Asem A. Akhmetzhanova, Ward Appeltans, Jose Manuel Arcos, Haley Arnold, et al. 2018. "BioTIME: A Database of Biodiversity Time Series for the Anthropocene." *Global Ecology and Biogeography* 27 (7): 760–786.

Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas." *International journal of climatology* 37 (12): 4302–4315.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." Themed Issue: Treatment Effect 1, *Journal of Econometrics* 225 (2): 254–277.

Günther, Andreas, Miet Van Den Eeckhaut, Jean-Philippe Malet, Paola Reichenbach, and Javier Hervás. 2014. "Climate-Physiographically Differentiated Pan-European Landslide Susceptibility Assessment Using Spatial Multi-Criteria Evaluation and Transnational Landslide Information." *Geomorphology* 224:69–85.

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. 2020a. "A Harmonized Global Nighttime Light Dataset 1992–2018." *Scientific Data* 7:168.

———. 2020b. "Harmonization of DMSP and VIIRS Nighttime Light Data from 1992-2021 at the Global Scale." *Scientific Data* 7:168.

Panagos et al., Panos. 2015. "A New European Slope Length and Steepness Factor (LS-Factor) for Modeling Soil Erosion by Water." *Geosciences* 5 (2): 117–126.

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., Emanuel Dutra, Pedro Viterbo, and Alon 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.

Tóth, Gergely, and Tamas Hermann. 2016. "European Map of Soil Suitability to Provide a Platform for Most Human Activities (EU28)," https://data.jrc.ec.europa.eu/dataset/jrc-esdac-42.

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.

Warszawski, L., K. Frieler, V. Huber, F. Piontek, O. Serdeczny, X. Zhang, Q. Tang, M. Pan, Y. Tang, Q. Tang, et al. 2017. "Gridded Population of the World, Version 4 (GPWv4): Population Density." *Atlas of Environmental Risks Facing China Under Climate Change,* 228.

Winkler, Karina, Richard Fuchs, Mark D.A. Rounsevell, and Martin Herold. 2020. "HILDA+ Global Land Use Change between 1960 and 2019." *Pangaea* 921846.