

Voluntary national conservation payments retain biodiversity-relevant forest structure

Dario Schulz

`dario.schulz@efi.int`

European Forest Institute <https://orcid.org/0000-0001-5983-1456>

Johanna Kangas

University of Helsinki

Sven Wunder

European Forest Institute (EFI) <https://orcid.org/0000-0002-9422-0260>

Article

Keywords: market-based instrument, conservation, impact evaluation, payments for ecosystem services, land purchases, additionality

Posted Date: October 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10814387/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Voluntary national conservation payments retain biodiversity-relevant forest structure

Abstract: Despite increases in forest cover and conservation efforts, biodiversity and ecosystem services in European forests face continuous threats. Market-based incentives for private landowners, such as payments for environmental services, aim to address these, but few rigorous impact evaluations exist to scrutinize effectiveness. In Europe, where degradation from intensive management rather than deforestation is the main driver of biodiversity loss, such quantitative studies are particularly scarce. We provide quasi-experimental evidence on the effectiveness of METSO, a large-scale, long-standing Finnish set-aside program involving diverse public payments for forest biodiversity. We combine a unique dataset of continuously enrolled areas across Finland during 2008-2025 (90,238 ha) with annual remote-sensing indicators. We causally identify additionality for interconnected outcome variables, combining spatial coarsened exact matching with stacked difference-in-differences regression. Results show significant conservation-positive impacts of METSO, including avoided clear-cutting (effect size: 0.25) and improved stand structure (e.g., canopy extent, wood volume, old-growth features) compared to the counterfactual. We characterize effect heterogeneity, sensitivity to spatial spillovers, carbon co-benefits and cost-effectiveness of intervention arms. By linking mandatory harvest notifications to remotely sensed stand-structure trajectories, we trace mechanisms from behavioral responses to biodiversity-relevant habitat conditions, thus methodologically extending impact evaluation of forest conservation beyond standard canopy-cover changes.

Keywords: market-based instrument, conservation, impact evaluation, payments for ecosystem services, land purchases, additionality

26 1 Introduction

27 Forests cover 35% of Europe's land area, providing key ecosystem services and sustaining
28 biodiversity (1, 2). Despite long-run increasing forest cover trends in Europe, the heritage of
29 intensive wood-focused management (with predominantly even-aged, few-species stands),
30 combined with climate change and invasive species all contribute to a continued deterioration
31 of natural forest habitats and decline in forest biodiversity (3, 4). Ongoing regulatory efforts
32 have arguably helped to reduce losses and exceptionally restore forest biodiversity,¹ but
33 insufficiently so. About 60% of forests in Europe, including ecologically valuable areas, are
34 privately owned (9). Hence, complementary policy instruments to regulation and protected
35 areas are required to incentivize voluntary conservation on private land. Notably, market-based
36 instruments such as payments for environmental services (PES) (10) can create tangible
37 conservation incentives for private forest owners.

38 PES schemes vary much in their effectiveness to promote forest conservation worldwide,
39 including leaving knowledge gaps regarding their heterogeneous effects across contexts, scales,
40 scheme designs, and outcome indicators (11). Enrolment strategies in a conservation program
41 may ex ante restrict additional conservation outcomes when non-random selection of
42 participants and sites causes adverse selection biases (12, 13). Only counterfactual-based
43 evaluation methods can thus reliably assess effectiveness, accounting for potential selection
44 biases. Impact evaluation studies of PES are thus increasingly common in tropical biodiversity
45 conservation, but much less so in the Global North, and especially Europe (14–16). Related
46 evidence on North American conservation easements and land-acquisition programs, which
47 permanently restrict or transfer development and harvesting rights, suggests potential for
48 reduced land conversion, although effectiveness strongly depends on baseline threats and
49 contract design (17–20).

50 Measuring conservation effectiveness across instruments remains a particular priority in
51 temperate and boreal forests, where biodiversity loss is driven less by frontier deforestation,
52 and more by intensive and recurrent wood-extractive management degrading structural habitat
53 features. Frequently employed remote-sensing indicators of forest cover or degradation can
54 potentially cover large areas, yet their reliability as biodiversity proxies remains contested –
55 while studies measuring direct biodiversity impacts are exceptional (21, 22). Thus, we rarely
56 learn directly about how alternative conservation instruments not merely keep trees standing
57 but also maintain structural ecological conditions (e.g. height, volume, age) benefitting
58 biodiversity, thus limiting the evidence needed for informed policymaking (23).

59 Here we provide causal evidence on the environmental impact of METSO², a large-scale, long-
60 established national set-aside PES scheme in Finland based on nationwide spatial data on

¹ Including the Natura 2000 protected area network (5), the EU 2020 Biodiversity Strategy (6), the Nature Restoration Law (7), and the EU Forest Strategy within the European Green Deal (8)

² METSO stands for "Etelä-Suomen metsien monimuotoisuuden toimintaohjelma" (eng.: Forest Biodiversity Program for Southern Finland). Despite the stronger focus on Southern Finland, our sample contains 9.8% of sites also in Lapland (North), and the national scope is also reflected in the targets of the new program period (2026-2040)

61 forests enrolled between 2008 and 2025. METSO operates with three main contract modalities:
62 temporary conservation contracts (10 or 20 years), permanent conservation contracts, and state
63 land purchases. Previous METSO research has focused on features such as conditionality,
64 permanence, targeting and social motivation (24), social acceptability (25), tradeoffs between
65 biodiversity and carbon benefits (26), and ecological targeting (27, 28), all of which plausibly
66 shape additionality. The METSO program is widely accepted by landowners and constitutes a
67 high-quality network of contractually protected forest areas by improving trust and targeting
68 conservation to ecologically valuable sites (27).

69 However, evidence on the program's overall causal environmental additionality is lacking. We
70 therefore evaluate METSO along an explicit theory of change (Figure E1): conservation
71 contracts and purchases should reduce landowners' harvest intentions, leading to avoided
72 canopy loss, retained stand structure, and with this improving habitat-suitability proxies,
73 conditional on adequate targeting, compliance, and limited leakage. We use self-reported
74 harvest intentions (*outputs*), annual and biennial remote-sensing indicators of canopy cover
75 (and loss), stand structure (*outcomes*), and habitat-suitability proxies (*impacts*) in a combined
76 matching plus stacked difference-in-difference framework to estimate the counterfactual
77 scenario, i.e., what would have happened to forests enrolled into METSO in its absence. The
78 annual cohort enrolments (2008-2025) allow us to construct internally valid sub-experiments
79 for each cohort while flexibly allowing sufficient time for changes in observable outcomes to
80 materialize, as well as assessing pre-treatment trends depending on the availability of temporal
81 outcome data.

82 2 Results

83 2.1 Main results

84 Our data covers approximately two thirds of all METSO areas, as information on the exact
85 location and starting date for the residual METSO sites is partially missing or inconsistent
86 (Figure S2). Among those, we observe 70,888 suitable control units for 78,992 treated stands
87 (covering 90,238 ha; 94.6% of units with known enrolment date) across entire Finland from
88 sequentially enrolled 2008 to 2025 cohorts. However, we focus on 2010-2018 cohorts in the
89 main analysis to ensure sufficient pre- and post-treatment observations across all outcomes (cf.
90 Figure E2). After filtering and cohort-wise matching, static baseline covariate balance between
91 treated and control units is improved for each cohort as well as among dynamic pre-treatment
92 outcome variables (Tables S2, S3, S5.1-5.36). We thus identify the trimmed average treatment
93 effect (TATT) over a fixed event-time window under plausible assumptions of common support
94 and parallel trends within matched cohort stacks.

95 Our results show that METSO conservation contracts contribute significantly to maintaining
96 structural forest characteristics via avoided logging (Table 1). This entails a sequential causal
97 logic (cf. our METSO theory of change, Figure E1). At the *intermediary output stage*, the
98 proportion of stands with clearcutting use declaration (in Finland, an obligatory pre-logging
99 step) in METSO enrolled sites Table 1 drop within three years after enrolment by 3.1
100 percentage points vis-à-vis the counterfactual. This exceeds the average annual post-treatment

use-declaration rates in the control group (2.57%, Table S3), suggesting that METSO offset virtually all logging intentions that would otherwise be expected among enrolled stands. Despite possible unplanned canopy loss from natural disturbances, the share of actual canopy-cover changes anticipated in use declarations is high (>80% recently), thus constituting a reliable predictor of actual harvest decisions (Figure E3): avoiding planned harvests (*output stage*) is the primary causal pathway to METSO-reduced annual canopy-cover loss (average: -1.4 percentage points (95% CI: 1.2–1.7) (*outcome stage*), slightly below the average canopy-loss rate in control sites across observed post-treatment time periods (1.56%; SD: 9.98%, Table S3). Applying our average estimated effect to total cumulative area enrolled by 2025 (90,238 ha) indicates that METSO currently contributes to avoiding $\approx 1,263$ ha of canopy cover loss annually.

<TABLE 1 HERE>

Avoided canopy-cover loss translates into improved structural stand features relevant for biodiversity at the *impact stage* (Table 1). Within six years, METSO sites have on average $1.04\text{m}^2\text{ha}^{-1}$ higher stand basal area, (95% CI: 0.84–1.24), $10.46\text{m}^3\text{ha}^{-1}$ higher total wood volume (95% CI: 8.59–12.33) than their counterfactual (Table 1; Table S4 provides additional outcome variables with similar patterns). This does not necessarily imply that METSO sites grew better; rather, control sites were more often thinned or harvested (cf. Table S3 for absolute values).

Finally, we also estimate the impact on two derived binary indicators of *habitat suitability* for hawks and woodpeckers of conservation priority (Table 1). We find that within six years, METSO increases the share of stands with at least one pixel to be classified as old-growth (based on tree species-specific thresholds) by 0.7 percentage points (95% CI: -0.5–1.9), though this TATT estimate was not significantly different from zero. This most likely reflects the relatively short evaluation horizon in our main specification; among the pre-2014 cohorts we find a significant increase of 2.07 percentage points (95% CI: 0.80 - 3.35) after ten years (Figure E4). Similarly, we find that METSO increases the probability for stands containing at least one high-spruce volume pixel (minimum $150\text{m}^3\text{ha}^{-1}$) by 2.1 percentage points; 95% CI: 1.6–2.6 (Table 1, row F).

Moving beyond average annual estimates, Figure 1 shows estimated pooled and cohort-wise dynamic group-time effects on output and outcome variables relative to the start of the intervention. Estimating event-time study effects separately for each cohort shows overall similar cohort effects and – importantly – largely parallel pre-treatment trends for each cohort and outcome. Most pre-treatment placebo estimates are close to zero, without clear trends, as supported by the Wald test statistics (cf. Table 1, columns 5-6), thus lending overall plausibility to our parallel-trend assumption. For clearcutting use declarations and canopy cover loss, placebos are shifted downwards, indicating a slightly lower use pressure among treated units and possible anticipation, but at an order of magnitude smaller than the post-treatment effect. Based on visual interpretation, METSO avoids use declarations and canopy cover loss within three years after contract establishment (Panels A, B), i.e. the valid duration of use declarations. In contrast, impacts on canopy extent and tree volume materialize immediately and consistently

every year from the start (Panels D, E). Notably, the pooled effects displayed are not average treatment effects, due to dynamically changing cohort composition, which also drives the larger aggregated estimates using all available cohorts and observations (Figure E4).

FIGURE 1 HERE

2.2 Heterogeneity analysis

Our analysis reveals statistically significant treatment effect heterogeneities across major tree species, regions, and intervention types, yet the identified patterns are mixed (Figure 2). Unsurprisingly, heterogeneity is more pronounced for impact-proxying forest structural outcomes than for use declarations (outputs) and tree loss (outcomes), although within similar magnitudes and same direction of impacts.

<FIGURE 2 HERE>

Based on visual inspection of the estimated TATTs, the clearest pattern is that impacts on use declarations, total wood volume, tree loss and extent are smaller in pine-dominated Region III (Northern Finland, cf. Figure S1) exhibiting lower land-use pressure but also poorer growing conditions. This is consistent with theory and previous evidence that lower pressure ex ante reduces additionality in PES (Wunder et al. 2020) and other conservation instruments (15), although the extent to which this lower impact reflects adverse selection or lower potential gains due to growth-restricting environmental conditions (also increasing rotation length and annual logging probabilities) remains unresolved.

Disaggregating impacts by major tree species groups (pine, spruce, broadleaf) reveals strong yet complex differences in outcomes. Pine-dominated stands benefit the least in terms of canopy cover and wood volume (Figure 2), consistent with pines being often located in less productive, slower-growing stands with longer rotation periods: pines are harvested less often, implying a lower counterfactual logging pressure that will depress impact estimates. Volume gains are driven by spruce; surpassing the 150 m³ ha⁻¹ spruce volume threshold also only occurs within spruce-dominated stands, so the statistically insignificant estimates for broadleaf and pine should rather be seen as placebo tests.

Finally, disaggregating impacts by conservation modality (temporary vs. permanent conservation contracts, land purchases) yields mixed results (Figure 3). Notably, these do not directly identify the causal effect of (quasi-randomly) assigning a particular intervention but rather represent differences in arm-specific effects relative to a common control pool. Ten-year temporary contracts perform slightly worse in most outcomes (smallest absolute effect), perhaps due to customized targeting and selection. Estimates for temporary 20-year contracts have consistently larger confidence intervals, due to the smaller sample sizes, but point estimates suggest similar effects on use declarations, tree loss and canopy cover compared with state purchases and permanent contracts. The intervention arms have distinct baseline characteristics (Table S7), so differences across arms should be interpreted as exploratory rather than causal effects of intervention modality.

2.3 Policy implications

Our quantitative impact results also translate into actionable recommendations. For instance, surpassing the designated spruce-volume threshold is an important predictor of goshawk nesting suitability (29). Multiplying our impact estimate on this habitat suitability proxy by METSO-enrolled stands in 2025 suggests additional conservation of 1,658 stands with optimal habitat (95% CI: 1,263–2,053 stands), compared to the hypothetical characteristics without the program. This example demonstrates the interpretability of forest structural indicators relevant for biodiversity; similar linkages to other species (groups) of interest can be developed via ecological modelling based on those indicators.

Our tentative carbon calculation (cf. S1 for method) suggests that METSO avoided the direct release of $\approx 1.5 \text{ MtCO}_2\text{e}$ between 2008 and 2025 via non-harvested wood products (cf. Figure 3). This estimate is based on our cohort-wise average counterfactual canopy loss rate with species-specific biomass expansion factors and a recursive decay model for long-lived products, restricted to the immediate land-use sector. At an assumed carbon price of $\text{€}70 \text{ tCO}_2\text{e}^{-1}$ this would correspond to a gross, undiscounted value of $\text{€}104.2$ million, or 24.5% of the total budget spent on METSO contracts between 2008 and 2023.

<FIGURE 3 HERE>

If, based on our heterogeneity analysis, ten-year temporary contracts look less effective than permanent ones and land purchases, are the former also less cost-efficient? Using the estimated annual tree loss per intervention arm, we calculate cost effectiveness as the ratio between aggregated undiscounted cohort-specific compensation costs and undiscounted cumulative cohort-specific avoided canopy loss (Figure 4). Costs per effective hectare-year fall as the evaluation horizon lengthens, as expected: intervention expenditure is assigned once to each cohort, whereas estimated avoided tree loss accumulates over multiple post-treatment years. Yet, similar to previous findings (30), temporary 10-year agreements consistently have the lower estimated cost per cumulative unit of avoided tree loss over the entire evaluation range, dropping from $\text{€}30,000$ after five years below $\text{€}10,000$ after ten years. In comparison, land purchases and permanent contracts drop from over $\text{€}60,000$ after five years to just above $\text{€}30,000$ after ten years. However, the estimated cost-effectiveness advantage of 10-year agreements depends on persistence: estimates from early cohorts provide a weak but imprecise signal of declining avoided canopy loss after contract expiration, although our currently available data does not yet allow for a firm conclusion (Figure E5). While our data suggests that 15-30% of temporary 10-year contracts had follow-up contracts on at least 90% of their area within three years of contract expiration, evidence from the Conservation Reserve Program, on the other hand, suggests that many temporarily protection can also last beyond contract periods (31).

<FIGURE 4 HERE>

2.4 Robustness checks

We report similar estimates for complementary outcomes using additional indicators from the same data sources and comparable indicators from alternative remote-sensing products (Table S4). To maximize sample retention for the MS-NFI outcomes, we used 2010-2018 cohorts in our main analysis, with the tradeoff of not being able to formally assess pre-treatment trends. We therefore also used alternative estimands: a) 2014-2018 cohorts to get and assess at least three pre-treatment MS-NFI outcomes, b) 2008-2013 cohorts to assess a ten-year post-window where this was available, and c) all cohorts and all available outcome years (the latter no longer constitutes an unbiased causal TATT because of changing cohort composition). We find overall similar effect estimates, with increasing magnitudes for longer time horizons (Figure E4). Our estimates are insensitive to potential spatial spillovers exposure adjustment onto neighboring parcels, which we assess by including spillover exposure, i.e. the number or area of treated stands within certain buffers (200m, 500m, 1000m, 5000m) as control variables. Across outcomes, we find that magnitude and significance of main impact estimates do not change substantially (Figure E6). Our results are robust to alternative assumed error structures, including clustering at parcel and year level, at CEM-bin and municipality level, and standard errors robust to spatial autocorrelation (32) (Figure E7). We also show that results are not driven by any single cohort (Figure E8). Lastly, we tested alternative staggered-event study estimators, robust to heterogeneous treatment effects (33, 34), alongside conventional two-way-fixed-effects. These estimators use the unweighted, unstacked panel, and therefore do not preserve the trimmed time horizons and weights underlying our main TATT. Nevertheless, their post-treatment estimates produce similar direction and magnitude as in our main specification, thus corroborating our findings (Figure E9). These specifications should not be interpreted as independent validation of pre-trends (some of which are clearly violated), reflecting a different sample, weighting and estimand. We test the sensitivity to potential violations in parallel trend assumption (cf. Table 1, columns 5 + 6) using “honest” difference-in-differences (35) finding that affected outcomes A and C remain significant even if the parallel trends violations in the post treatment period are 1.3 times stronger than in the largest deviation during the observed pre-treatment period, or if the slope of the pre-trend changes by up to 0.2% (Figure E10).

3 Discussion

We provide quasi-experimental evidence that a voluntary national incentive scheme for forest biodiversity generates sizeable, upscaled improvements in environmental outcomes, relative to its counterfactual. The empirically underpinned underlying behavioral mechanism is that avoided planned harvesting leads to retained forest structure and improved habitat proxies relevant for biodiversity. Combining submitted use declarations as proxies of self-reported harvest intentions with remote-sensing indicators of stand structure enables us to explicitly quantify subsequent stages along the causal chain depicted in our METSO theory of change (Figure E1), testing its key assumptions.

Our impact estimate of avoided forest-cover loss (1.4 percentage points) is substantial vis-a-vis both the Finnish annual mean canopy loss (~1%, incl. temporary clearcuts and land-use change) and our matched control sites across all observed pre-treatment periods (-1.56%).

Putting this into perspective, our effect size of 0.25 (main effect divided by pooled standard deviation of forest loss over all observed periods) is high compared to other PES initiatives (Wunder et al., 2020) and other conservation instruments (11, 15).

Obviously, policy contexts are not always comparable. In Finland, deforestation for agricultural conversion, as predominating in the tropics, is exceptional, and forest biodiversity is mostly threatened by selective logging and clearcutting for wood harvesting. Finnish policy targets are thus not only about forest quantity, but also the standing forest's environmental quality for biodiversity. This limits how much forest impact sizes can be compared across ecosystems. Yet, voluntary permanent conservation agreements yielded similarly high additionality in the United States (18).

Regarding external validity, our stacked study design for subsequently METSO-enrolled cohorts impedes us from estimating the average treatment effect on the treated (ATT). Instead, our estimand is the somewhat narrower trimmed average treatment effect (TATT) for stands with comparable control units and matching outcome trajectories. Across many landscape characteristics METSO sites included in the analysis resemble those excluded from this analysis due to lacking enrolment dates while containing more spruce-dominated forest (Table S6). Relative to the wider forest landscape, METSO sites tend to have denser and taller forest, greater spruce prevalence, and less prior harvesting (Table S6). Within those stands, our combined matching and stacked difference-in-differences strategy mitigates major biases, contributing to internal validity of our estimates. The set of spatially explicit control variables contextually captures environmental and socioeconomic heterogeneity, and the improved post-matching balance within each stack indicates comparability of treatment and control sites. The pre-treatment placebo tests provide further plausibility to our parallel trend assumption, while the TWFE estimation eliminates remaining time-invariant unobserved heterogeneities and time-variant shocks common to each cohort. We acknowledge that certain pre-treatment estimates display a parallel vertical shift (e.g. Figure 1, Panel A). While jointly significant under a Wald test, this baseline shift is an order of magnitude smaller than the subsequent post-treatment effect and does therefore not affect conclusions. In addition, we could have missed important owner-level characteristics or time-varying factors that vary across space, including labor shortages or subsidies given to some stands.

Our study exemplifies how novel remote sensing-based time-series on stand structure can be employed as indicator for habitat suitability, thereby providing partial causal evidence on the outcome of interest: biodiversity. We thereby contribute to the nascent literature on causal inference for biodiversity conservation (14). Using the spruce volume threshold, a key predictor of goshawk nesting sites, we provide an interpretable, policy-relevant estimate of how many stands continue to support goshawk nesting conditions due to METSO, although our findings cannot be interpreted as direct proof of species recovery.

Previous research linked spatiotemporal variation in canopy height to habitat heterogeneity and tree species diversity (36). Our findings point towards an important limitation in our difference-in-differences setting because partial harvesting (leaving retention trees) increases height variability in logged-over control stands (Figure E4). The identified signal from partial harvesting could be beneficial for generalist and pioneer species, while some forest specialists

rely on shade and other old-growth features (37, 38). Thus, different dimensions of biodiversity rely on different indicators. Recent advances in the field of remote sensing show promising results, particularly in the realm of airborne lidar scanning of deadwood, linked to faunal biodiversity (39, 40). High-spatial-resolution airborne laser scans are widely available across Finland for such an analysis in the future (41), despite its varying spatiotemporal coverage (cf. Figure S4). Overall, more research on the transferability and interpretability of changes in stand structure is needed to establish a reliable association between remotely sensed outcomes and the biodiversity dimensions of interest.

In addition to biodiversity as METSO's target, other outcomes could be of interest, including analyzing potential trade-offs and synergies. Previous METSO research indicates likely tradeoffs in the additionality and cost-effectiveness of biodiversity and carbon sequestration, yet also synergies with carbon storage (26, 42). Yet, tradeoffs and synergies may be heterogeneous across space; knowing their distribution can improve spatial targeting with large potential gains in policy effectiveness and cost-efficiency (43). Our tentative carbon calculation is thus only a first step towards rigorously assessing the long-term net carbon balance of METSO, including standing stocks' carbon sequestration and regrowth after harvesting.

To conclude, methodologically our study demonstrates the benefit of evaluating conservation interventions across multiple interrelated outputs, outcomes and impact proxies along an explicit theory of change. To hold national policies accountable for changes in ecological trajectories, especially in the Global North, we need to look also beyond habitat extent (e.g. canopy cover) towards nuanced indicators linked to specific dimensions of structural, functional, or even genetic biodiversity across spatiotemporal scales. An ecological theory of change of biodiversity outcome-impact relationships is likely non-linear, with feedback loops yet to be understood. We show here that outcomes vary across space and tree species, implying potential gains in intervention efficiency by effective spatial targeting. This may ultimately also improve the cost-effectiveness, credibility and public acceptance of biodiversity conservation programs.

4 Online materials and methods

We conduct a spatially explicit impact evaluation to causally identify how the METSO program has affected forest structure and biodiversity levels. We start conceptually with the theory of change we developed for the program, i.e. how we expect the intervention to work. We then describe the data we used; a full overview of all data sources is provided in Table S1, summary statistics of all variables are in Table S2 and Table S3. This is followed by our empirical approach, which consists of two steps: 1) Spatial covariate matching based on observed biophysical and socioeconomic variables; 2) Stacked difference-in-differences regression analysis with multiple periods and multiple treatment timings.

4.1 The METSO program

METSO is a state-funded voluntary conservation scheme in Finland focusing on forest biodiversity. It was first established in 2008 and prolonged in 2014 and is jointly coordinated by the Ministry of the Environment and Ministry of Agriculture and Forestry. Its

implementation is carried out by regional centers for development and infrastructure (ELY-centers) and the Finnish Forest Centre, Suomen metsäkeskus. Its goal is to permanently protect 96,000 ha of forest by 2025 and thereby contribute to reaching the 30% protected boreal forests envisioned by the Finnish Forest Strategy 2030. The program will be continued and has set new targets for the upcoming phase 2026-2040. Syrjänen et al. (44) describe the ecological criteria used to ensure cost-effective targeting of METSO sites. These criteria are tailored for each main forest habitat, focusing on structural features and habitat types crucial for biodiversity. Additionally, presence of threatened species and proximity to existing protected areas are key factors considered in the process. A mid-term review (27) found, using a spatial optimization tool that, overall, METSO effectively targeted sites with high biodiversity values and is well-integrated into the existing network of protected areas in terms of ecological connectivity. Yet, the authors show that spatial targeting of conservation sites could be further optimized to achieve a higher level of ecological quality, connectivity and cost-effectiveness. Discussions about spatial targeting based on commercial harvesting pressure are not prominent, although the programs stronger focus on southern Finland is aligned with higher pressure there.

4.2 Theoretical framework

We start by outlining our vision of the METSO conservation program's theory of change (ToC) along its causal pathway from inputs to impact (cf. Extended Data Figure E1). This ToC is informed by the legal text, secondary literature, and interviews with experts and implementers during in person workshops in Helsinki (June 2025) Moving from left to right, the ToC shows how and under what conditions the program can be expected to reach its ultimate goal, namely, to contribute to halting the decline of forest habitats and species and to stabilize a favorable environment for biodiversity

The legal framework consists of the Nature Conservation Act, the Forest Act, the Act on the Financing of Sustainable Forestry (2015-2023), and the Act on a Temporary Forestry Incentive Scheme (2024-ongoing). Intermediaries are primarily the ELY centers as well as the Finnish Forest Centre, Suomen metsäkeskus, but also forestry professionals including forest management associations and forestry companies have sometimes acted as intermediaries between landowners and ELY centers. Given a supportive legal and institutional framework, sustained funding, and intermediaries that can quantify ecosystem service values and associated opportunity costs to landowners, the program can implement conservation actions such as permanent and temporary conservation contracts and land purchases. There are thus two options for permanent conservation: either a permanent conservation contract is made (resembling a conservation easement), or the land is purchased by the state. In addition to annual state purchases, there were two waves of conservation of state-owned, commercially managed land in the years 2009 (10,000 ha) and 2014 (13,000 ha). There is also the option of temporary conservation contracts of 20 years, which makes up a small share. In addition, there is a ten-year contract modality which is administered by the Finnish Forest Centre (Suomen metsäkeskus). The 10-year agreements (called Environmental forestry subsidy agreements) are financed through the Kemera scheme (Act on the Financing of Sustainable Forestry), replaced by the Metka scheme in 2024 (Act on a Temporary Forestry Incentive Scheme). The 10-year agreements have slightly different selection criteria compared to other modalities as they target

areas surrounding or connected to Woodland Key Habitats (as defined in Section 10 of the Finnish Forest Act) in addition to sites meeting METSO selection criteria. In our main analysis we grouped all modalities, but also show estimates disaggregated by contract modality. If payment levels are adequate and landowners accept the scheme, these treatments are expected to result in landowner understanding and participation in the program as well as changed logging intentions as first-order outputs. This, in turn, will lead to avoided logging in enrolled areas and support for forest-based incomes via the compensation payments, if non-compliance is adequately monitored and sanctioned. An important condition for additionality at this stage is successful targeting of sites with a real logging threat. If the program enrolls areas without real logging pressure, i.e., areas that would not have been logged in the absence of program participation, this effect cannot be attributed to METSO and results in lower additionality (11). In the short term, the possibility of participating in the program may also incentivize forest owners that would like to participate but remain on the waiting list due to limited funding to withhold any logging activities in expectation of future compensation payments. This is because once logged, the area is no longer eligible for METSO. Finally, avoided logging is expected to maintain structural and functional habitat diversity and connectivity, which in turn contributes to halting biodiversity loss, and supporting other ecosystem services. Two critical conditions must be met for this impact to materialize. First, the targeting of enrolled areas must effectively identify ecologically relevant areas that support forest biodiversity. If non-sensitive areas are protected, the contribution to halting biodiversity loss may be reduced, even if the program effectively mitigates logging threats. Second, land use spillovers to non-enrolled sites, also referred to as leakage, must remain limited. If protection of some sites leads to more intensive logging in non-enrolled, but ecologically similarly valuable sites, the net positive effect can be offset (45).

4.3 Data

Our study region is the entire Finnish mainland, covered by 75% with forests (22.8 million hectares) and being the third biggest producer of round timber and sawn wood in Europe. The unit of analysis are single forest stands, which are a connected subarea of privately owned forest plots with homogenous ecological conditions including dominant tree species. We excluded very small stands of less than 0.01 ha (i.e., 100 square meters) because 1) the granularity of remote sensing products makes zonal statistics more sensitive to outliers, and 2) connectivity and other ecological processes relevant to biodiversity are likely to occur at larger scales. Furthermore, we excluded 510 stands larger than 100 ha, because these are outliers in terms of the area sizes enrolled in the METSO program (median size: 0.8 ha). From the original set of 12,383,342 forest stands (17.13 million hectares, not including large national protected areas in Lapland), we filtered out 7.15% for being state-owned, and 4.1% for being otherwise protected (e.g., Natura 2000), and 725 stands (<0.0001%) for being less than 0.01 ha or larger than 100 ha in size. The remaining forest stands used for the subsequent analysis cover 15.28 million hectares of forest across 2 million cadaster units (the level at which we cluster standard errors).

4.3.1 Regional analysis

Finland is characterized by a strong latitudinal gradient both in vegetation and human population density. We therefore conduct regional subgroup analysis, in which we distinguish

between five regions (I-V), which we grouped based on their land use pressure, bioclimate, and METSO conservation targets (cf. Figure S1). Group I is in the most southern bio-geographical regions in the hemi boreal and southern boreal zone. It is characterized by the highest land use pressures competing with conservation (i.e., agriculture, buildings), the highest land prices in Finland, and the lowest share of protected areas. Timber harvests have in some parts of it exceeded the maximum sustainable yield in recent years, and METSO conservation targets were not reached. Group II is in the southern and middle boreal zone. It is characterized by high land use pressure, but METSO conservation targets have been achieved nonetheless because of ample participation and supply of potential sites by landowners. Group III is in the middle and northern boreal zones. It is characterized by low land use pressure, the highest share of protected areas in Finland and an ample supply of potential conservation sites with low land costs, so METSO conservation targets have been met. These regions are not the main focus of METSO as the emphasis is in Southern Finland. Group IV is in the southern boreal zone but faces less land use pressures than in the most southern Finland. Nevertheless, METSO targets have not yet been met. Group V is in the middle boreal zone with moderate land use pressures. METSO targets have been met.

4.3.2 Treatment indicator

We used spatial data on private conservation areas enrolled under the METSO program as our main treatment indicator. The data consists of a harmonized shapefile showing the exact area under conservation along with the year it started. We compiled and harmonized this data from the Finnish Forest Centre³ and Metsähallitus Parks & Wildlife Finland (personal communication, December 2025). We received different versions with partially overlapping areas and contradictory contract modalities, which we processed as follows. First, we started by reading and cleaning data, i.e. parse as polygons, splitting temporary or permanent contract modality based on code, harmonizing variable names and adding information (e.g. fixed year, data provenance) where needed. For provenance, we included a unique identifier from the original source dataset and calculated the area of original shape. Second, within each dataset, shapes were intersected and only strictly non-overlapping parts pass. We flagged overlapping parts (but maintained them in the dataset). Third, we identified and flagged intersections across datasets, some of which we assume to be more reliable than others. We thus used the latest (2026) state purchase data as the Masterfile and subordinate older, less reliable versions. Where it overlaps partially with older versions, we built their intersection and did not consider old versions if they intersect with the new. We included non-intersecting parts of old versions but added a flag for as overlapping and set the year value to NA. Number, share, and area of overlaps were recorded. Fourth, all datasets were merged and again intersected to identify overlaps and flag them for exclusion, recording number, share, and area of overlaps. Fifth, a final quality flag was added for observations to be used in treatment assignment; all other are still exported but just used to exclude potential control units from the analysis, similar to non-METSO protected areas. We then exported the data and validated its coverage over time against official statistics from the Natural Resources Institute Finland (cf. Figure S2).

³ Available online: <https://www.metsakeskus.fi/en/node/946>

We considered cohorts between 2008 and 2025, although the inclusion of cohorts after 2020 was restricted by varying outcome data availability (cf. Table S1, Figure E2), and we had to exclude state purchases before 2014 due to missing information about the exact starting date. We differentiate five contract modalities, i.e., permanent conservation contracts, temporary conservation contracts (10 or 20 years), and state purchases (including annual purchases as well as the two bigger transfer waves in 2009 and 2014). We overlaid the forest stands with this set of METSO conservation areas and created a binary treatment indicator that is one when a stands centroid falls within a METSO area, and zero otherwise, along with the year when protected started. This resulted in a total of $N = 83,513$ treated stands.

4.3.3 Outcome indicators

Our aim is to provide evidence of impact along the stages depicted in our theory of change. For the output stage, we measure the self-reported intention to harvest by looking at submitted use declarations. In Finland, forest owners must indicate at least 15 days before any activity (e.g., thinning, clear-cutting etc.) on what area they want to perform the activity. These use-declarations remain active for three years, but there is no obligation to carry out the declared activity within that timeframe. For each year, we calculated the share of each stand for which any use-declaration was active. We also calculated the share of each stand for which a clear-cutting use declaration was active. Figure E3 documents that the share of actual canopy cover changes anticipated in use declarations is substantial ($>75\%$) and has slightly increased over the past years, indicating that self-reported use declarations are a useful proxy of intended behaviour.

For the outcome stage, we used a range of remote-sensing based indicators. First, we used annual tree extent, removal (a proxy of logging), and height maps available at 30 m spatial resolution between 2001 and 2021 (46). In addition, we used publicly available data on forest structure from the Multi-Source National Forest Inventory (MS-NFI) raster maps, which are available at different spatiotemporal resolutions. MS-NFI is based in satellite imagery along with forest inventories that were used to calibrate a biophysical stand growth model and improved over the years (47–51). Indicator variables used here are available at 20 m or 16m spatial resolution every two years between 2009 and 2023 for the entire country. We extracted the mean annual values per forest stand for the following variables: age, diameter, wood volume, basal area, total canopy cover, and broadleaf canopy cover.

We derived additional structural indicators because they are associated with the presence of specific species. Following species distribution modelling principles, like maximum entropy models, we link spatial feature variations to species presence. Previous research (52) identified certain birds of prey (European honey buzzard, northern goshawk, common buzzard) and woodpeckers (White-backed, lesser spotted, and Eurasian three-toed) as indicators of general forest biodiversity. This is because these species prefer specific forest attributes or utilize ecological niches created by their actions. Notably, the first four mentioned above are also Red List species, highlighting their conservation importance. Virkkala et al. (52) found that total and deciduous wood volume, along with stand height, significantly predicted the presence of these bird nests using a maximum entropy model. Similarly, Björklund et al. (29) demonstrated that spruce volume effectively predicts northern goshawk habitat suitability, estimating a threshold

of 150 m³ ha⁻¹ for optimal nesting suitability. Along similar lines of using thresholds, Forsius et al. (53) used tree-species-specific age thresholds to identify valuable old-growth forests for certain species. Therefore, we utilize the same relevant habitat suitability indicators: broadleaf canopy cover, total canopy cover, average stand age, and basal area. Additionally, we calculated two binary threshold-based indicators: spruce volume exceeding 150 m³ ha⁻¹, and the presence of old-growth forests. The spruce volume indicator is 1 if a stand is spruce dominated with at least one pixel above the volume threshold, and 0 otherwise. The old-growth indicator is 1 if at least one pixel in the stand exceeds the species-specific age threshold (broadleaf: 60 years, spruce: 100 years, pine: 120 years), mirroring the approach of Forsius et al. (53). We conjecture that at least some of our structural indicators are suitable proxies of habitat suitability, thereby approximating the final impact stage in our theory of change.

4.3.4 Contextual variables

We used a combination of raster maps and administrative shapefiles paired with regional statistics to describe the site conditions that we assumed to affect forest growth, biodiversity levels, and protection status. The rationale for inclusion follows the principles of causal inference (54) while our identification strategy is described in more detail in Section 2.3. Table S1 provides a description of all variables along with their resolution, coverage and source reference. For all forest parcels included in the analysis, we extracted the following information.

We calculated the distance to other protected areas such as wilderness areas, national parks, Natura 2000 areas, public protected land, nationally important landscapes such as old growth forests and peatlands, and other private conservation areas protected before 2008. Proximity and connectedness to such areas with high biodiversity plays on the one hand an important role in the dispersion of seeds and mobility of species. On the other hand, this connectedness increases the potential ecological value that was used as a targeting criterion for METSO contracts.

We extracted physical, chemical, and hydraulic soil properties, because these site conditions influence species compositions, tree growth and overall land productivity. While the former is particularly relevant for biodiversity levels, the latter determines opportunity costs and thereby the level of compensation payments, so it is also related to the decision of a forest owner to enroll a parcel. In addition, we extracted topographic information on altitude and slope, because these variables determine not only growth conditions, but also relate to accessibility and harvesting cost proxies.

As a proxy of transportation costs, we calculated Euclidean distances to roads, ports, sawmills, and bigger cities, because these variables may affect profitability of harvesting the parcels. We control for long term bioclimatic growth conditions, that affect overall species compositions and tree growth by using long term averages of temperature, precipitation, along with measures of their variability.

In addition to these time-invariant baseline characteristics, we calculate time-variant control variables with different spatial resolutions, driven by data availability. We control for annual stumpage prices of spruce and pine at the state level, filling missing data gaps with the national

average. Climatic variations in growth conditions are captured by 13 annual bioclimatic variables, that we calculated at the municipality level.

4.4 Empirical framework

Our empirical approach in essence combines matching and difference-in-differences regression analysis, similar to previous studies on forest conservation or agro-environmental schemes (55, 56) and is informed by principles of causal inference in observational settings (57). Our estimator combines cohort-specific donor-pool construction for control units, exact support restrictions, nearest-neighbor matching on pre-treatment outcome histories, and a stacked event-study regression (58).

4.4.1 Common support and donor pool restriction

First, we excluded all control units known to be protected in any way, e.g. Natura 2000 sites or contained in the METSO shapefiles but without exact enrolment year, or lacking covariate observations. Then, we ensure broad comparability and overlap in covariate space between treated and control units using coarsened exact matching (CEM) (59). Our data contains dozens of static environmental variables, many of which are highly correlated. Therefore, we start by reducing dimensionality using groupwise principal component analysis. We create principal components for soil texture, soil fertility, hydrology, bioclimate, and travel time that explain at least 50% of respective total variation. In addition to the first principal components of these groups, we use the stand area, elevation, slope, distance to protected areas, baseline tree cover and tree height in the year 2007 (before the first METSO cohort), and the area share harvested between 1984 and 2007 as a proxy of historical management. CEM was performed by splitting each of these variables into three bins, which is relatively coarse but ensured high retention of treated stands. CEM-bins also had to be strictly within the same ELY-region and dominant tree species.

Next, within each CEM-bin, we perform k-nearest-neighbor (here: $k=5$) matching using additional baseline covariates. This maintains the common support and co-distribution of key variables from CEM while further improving baseline covariate balance without compromising the treatment sample size. At this stage, we added secondary principal components for the above-mentioned categories, as well as distance to major roads, sawmills, lakes, rivers, broadleaf canopy cover in 2006, average stand height in 2006, and the slope in tree extent and tree height between 2001 and 2007. These additional baseline covariates are expected to reflect profitability, growth conditions and historical stand structure trajectory. We used the ordinary multivariate Euclidean distance without distance threshold.

There are also potential non-linear interactions between (un-)observed confounders that are better captured by non-parametric methods such as random forests and could further add to unconfoundedness (43). Overall, however, the specific choice of matching specification has been shown to not drastically change results in similar evaluations (60).

4.4.2 Pre-treatment outcome matching and construction of stacked panel

Despite this comprehensive set and balance improvement in baseline matching covariates, earlier tests suggested a potential violation of the parallel trend assumption (cf. Section 4.4.3).

Therefore, we decided to additionally match stands based on these pre-treatment outcomes to construct our final stacked dataset as follows. Let forest stands be indexed by $i = 1, \dots, N$, and let years be indexed by t . Let Y_{it} denote the outcome of interest, such as tree extent, use declarations, tree height, or wood volume. Let G_i denote the first year in which stand i receives the treatment of interest; never-treated stands are denoted by $G_i = \infty$. Treatment status is therefore $D_{it} = \mathbf{1}(t \geq G_i)$. The analysis is implemented cohort by cohort. A treatment cohort is defined by the first treatment year $g \in \mathcal{G}$, here $\mathcal{G} = \{2008, \dots, 2025\}$. For each cohort g , the treated set is $\mathcal{T}_g = \{i: G_i = g\}$. The eligible control set are the never treated, i.e. $\mathcal{C}_g = \{i: G_i = \infty\}$ and we allow controls to be reused across different cohort stacks.

Then, let S_i denote the CEM stratum of unit i . For a treated unit $i \in \mathcal{T}_g$, the admissible donor pool is $\mathcal{C}_{g,S_i} = \{j \in \mathcal{C}_g: S_j = S_i\}$. Thus, all matched controls are never-treated units from within the same CEM stratum, thereby preserving exact or coarsened balance on the hard support variables used in the CEM stage and avoiding contaminated (later treated) controls. For each unit i and cohort g , pre-treatment matching features are constructed over the cohort-relative pre-period $t \in \{g - P, \dots, g - 1\}$, where P denotes the length of the pre-treatment window (here, we use a window of six years). For each matching variable X_{it}^m , we calculate summaries such as the pre-period mean, the last observed value before treatment ($g - 1$), and the linear pre-period slope. The slope is calculated only from a minimum of two to six available observations, which allows the procedure to handle annual and biennial MSNFI variables with missing years.

For cohorts after 2015, static biodiversity potential variables (61) are added to the matching-feature vector, because they are based on 2015 satellite imagery and could therefore confound earlier cohorts. For each stand and each year (cohort) we thus compute the covariate vector Z_{ig} containing the outcome trajectory summaries for use declarations, forest cover, forest loss, and average tree height. For each treated unit $i \in \mathcal{T}_g$, controls are selected from the same cohort-specific and CEM-restricted donor pool $\mathcal{C}_{g,S_i}$. With one nearest neighbor, the matched control is

$$m_g(i) = \arg \min_{j \in \mathcal{C}_{g,S_i}} d(Z_{ig}, Z_{jg}),$$

where $d(\cdot, \cdot)$ is the standardized Euclidean distance over the usable matching features. Controls are matched with replacement. Thus, the same control unit may be used for multiple treated units within a cohort and may also appear in several cohort stacks. If a control is selected multiple times, its matching weight in cohort g is proportional to the number of times it is selected.

For each cohort g , a cohort-specific event-time panel is constructed over $t \in \{g - P, \dots, g + L\}$. Event time is defined as $\ell = t - g$. For each cohort stack, the stack-specific treatment indicator is $D_i^g = \mathbf{1}(G_i = g)$. The stacked dataset contains observations indexed by (i, g, t) , or equivalently (i, g, ℓ) . Because controls can be reused across stacks, a stand may appear multiple times. We therefore define a stack-specific stand identifier $u = (i, g)$, the stack identifier $s =$

g , and stand-specific weights ω_i equal to one for treated stands, and equal to the number of times used for control stands.

Filtering and matching reduced our dataset from 83,513 treated and over ten million untreated stands, with sizeable differences in observed baseline characteristics to 78,993 treated and 70,888 unique control stands with improved balance and overlap across baseline characteristics (Table S2). Stack wise post matching balances are reported in Tables 5.1-36.

4.4.3 Event-study estimating equation

In the METSO case, treatment did not occur at one point in time, but conservation contract enrolment is staggered across annual cohorts from 2008 to 2025, and its effects may vary both across cohorts and with time since enrolment. Such heterogeneity is plausible because selection criteria, contract specificities, timber-market conditions, and the opportunity cost of conservation changed over the study period, while avoided harvesting and forest growth may cause treatment effects to accumulate over time. We therefore avoid a conventional two-way fixed-effects event study, which can make “forbidden comparisons” using already-treated units as controls and combine cohort-time effects using potentially negative or otherwise difficult-to-interpret weights when treatment effects are heterogeneous (34, 62). There are heterogeneity-robust estimators, including Sun and Abraham (34), Callaway and Sant’Anna (33), and Borusyak, Jaravel, and Spiess (63), which avoid the principal TWFE contamination problem. These approaches rely on a (conditional) parallel trends assumption, which we found to be partially violated when using the post-CEM filtered dataset (Figure E9). Therefore, we performed to above-described matching on pre-treatment outcome trends and implemented a weighted stacked difference-in-differences design following Wing et al. (58). The design treats each enrolment cohort and its matched controls as a separate sub-experiment, restricts controls to stands untreated throughout the relevant event window, and preserves the cohort-specific matching procedure used to construct comparable counterfactuals. With appropriate cohort weights, the resulting estimand is a coherent trimmed aggregate treatment effect on the treated (TATT).

The stacked estimator is particularly suitable here because it directly accommodates cohort-specific matched comparison sets, balanced event windows, and repeated use of eligible controls while retaining a transparent regression-based fixed-effects structure. The specification includes stand-by-stack fixed effects, which absorb time-invariant owner and stand characteristics within each cohort-specific comparison; stack-by-year fixed effects, which absorb unrestricted calendar-year shocks common to each sub-experiment; and municipality-by-year fixed effects, which absorb local time-varying factors such as weather, policies, infrastructure, and timber-market conditions. Identification therefore comes from changes within stands relative to contemporaneous changes among matched control stands in the same cohort-specific comparison and municipality-year. The weighted stacked event-study regression can be written as

$$Y_{igt} = \alpha_{ig} + \delta_{gt} + u_{mt} + \sum_{\ell \in L, \ell \neq -1} \theta_{\ell} [\mathbf{1}(t - g = \ell) D_i^g] + X'_{igt} \beta + \varepsilon_{igt} \quad (\text{Eq. 1})$$

Here, α_{ig} is a unit-by-stack fixed effect; δ_{gt} is a stack-by-calendar-year fixed effect, interacted with year, u_{mt} is a municipality-year fixed effect; and θ_{ℓ} is the event-time treatment effect at

relative year ℓ ; event year -1 is omitted as the reference period and X is an optional vector of time-varying controls including regional stumpage prices for spruce and pine and annual bioclimatic indicators. Standard errors are clustered by the original stand identifier to account for repeated use of the same stand across cohort stacks. Each coefficient θ_ℓ estimates the average treated-control difference at event time ℓ , relative to event year -1 , among matched treated units and their matched controls. For $\ell < 0$, the coefficients are placebo or pre-trend diagnostics. For $\ell \geq 0$, the coefficients estimate dynamic post-treatment effects.

Importantly, in one specification we leave a flexible window to include as many cohorts as possible to visually inspect effect trajectories. However, this implies that event times have changing cohort compositions so the estimand is not the TATT of the full sample, but of the cohorts contributing to that event time. Despite this limitation in interpretability, we see the benefit of reporting this estimand because it allows to trace also longer-term effects for early cohorts.

To overcome the limited interpretability arising from changing cohort composition, Wing et al. (58) propose a trimmed aggregate ATT (TATT) that is based only on cohorts for which a common event window is fully observed. Thus, the estimand is a matched supported-population TATT. It is therefore narrower than the full-population METSO ATT, but it is more credible for causal interpretation. In our case, we chose distinct time windows, e.g. -1 to 5 in our main specification, and -5 to 5 years to assess pre-treatment trends, corresponding to early cohorts (2010-2018) because some outcomes are only observed until 2023. Under the identifying assumptions described below, we can thus estimate the average difference between treated and control units at event time ℓ as $\theta_\ell = \mathbb{E}[Y_{i,G_i+\ell}(1) - Y_{i,G_i+\ell}(0) \mid i \in \mathcal{T}^{\text{matched}}]$. Using the trimmed event window and corrective sample weights compatible with those described by Wing et al. (58), we calculate a covariance-aware weighted average $\hat{\tau}$ of the post-treatment coefficients as

$$\hat{\tau} = \sum_{\ell \in \mathcal{L}_+} a_\ell \hat{\theta}_\ell \quad (\text{Eq. 2})$$

using equal weights a that sum up to one across post-treatment event times. Equal weights are appropriate because each stack has the same event window and, through 1:1 matching with control-reuse weights, equal effective treated and control group sizes. Its standard error uses the full covariance matrix $\text{Var}(\hat{\tau}) = a' \widehat{\text{Var}}(\hat{\theta}) a$, thereby reflecting to underlying cohort sample sizes.

Next, we estimate heterogeneous treatment effects using interaction terms. Let $H_i \in \{1, \dots, J\}$ denote a mutually exclusive subgroup within the heterogeneity variable, such as tree species, METSO intervention arm, or a region-by-species cell. We create an internal treatment indicator for each subgroup $D_{ig}^{(h)} = D_{ig} \mathbf{1}(H_i = h)$ and then jointly estimate:

$$Y_{igt} = \alpha_{ig} + \delta_{gt} + u_{mt} + \sum_{h=1}^J \sum_{\ell \in \mathcal{L}, \ell \neq -1} \theta_{\ell h} [\mathbf{1}(t - g = \ell) D_{ig} \mathbf{1}(H_i = h)] + X'_{igt} \beta + \varepsilon_{igt}. \quad (\text{Eq. 3})$$

Each $\theta_{\ell h}$ is the event-time effect for subgroup h , relative to the shared pooled matched control group. We estimate joint Wald tests of equality across subgroup effects.

4.4.4 Assumptions and Robustness checks

Our design relies on four key assumptions typical for difference-in-differences, which's plausibility we assess as follows. First, a conditional parallel-trends assumption within matched cohort-specific stacks. Specifically, after conditioning on the CEM stratum and matching on cohort-relative pre-treatment outcome histories and covariates, matched controls are assumed to represent the untreated counterfactual trajectory of treated units. This is supported by matching on forest type, location, stand characteristics, and pretreatment outcome histories. There is no way to test this assumption directly, but if there is a parallel trend prior to treatment this can be seen as an indication that the relationship would continue to hold after treatment. We therefore calculate and visually assess pre-treatment placebo effects, that should not differ substantially from zero or exhibit temporal trends. For transparency, we conducted and report Wald tests for placebos (pre-event estimates) being jointly equal to zero, a linear placebo-trend being equal to zero, and effects (post-event estimates) being jointly equal to zero using an asymptotic chi-squared reference distribution. We visually assess whether estimated pretreatment differences are small relative to the post-treatment effects. The principal remaining threat is selection based on stand-specific time-varying information—such as an owner's changing conservation preferences or unobserved forest damage—which cannot be eliminated by matching or fixed effects.

Second, we assume no anticipation, i.e. that treatment does not affect outcomes before G_i . This can also be assessed by looking at the pre-treatment placebos. Regarding the timing of outcome responses, we further rely on our theory of change, in the sense that outputs, outcomes and impacts are consequential and cannot occur in the opposite order.

Third, we rely on never-treated units to ensure clean controls throughout the event window and that G_i correctly measures the onset of treatment. We conservatively excluded control stands that could have some form of protection status or different management, i.e. state owned/managed land, METSO areas for which we were unable to identify the exact enrolment date, other (non-METSO) protected areas or identified landscapes of ecological importance that fall under other management restrictions.

Fourth, we assume common support, i.e. that treated units have comparable controls within the same CEM stratum. We ensure design-based common support, i.e. by using CEM, and assess resulting balance in group-wise summary statistics of key confounders (Table S2) as well as the empirical cumulative distribution function plots (Figure S3), finding overall reduction in imbalance and broad common support. Furthermore, we assess baseline and pre-treatment outcome balance individually for each stack (Tables S5.1-36).

Fifth, we assume negligibility of spillovers (also known as the stable unit treatment value assumption), i.e. that treatment of one stand does not affect outcomes of other units. Spillover effects can occur for different reasons, for example via learning from neighbors or by changing input or output market price conditions or relaxing previous labor and input constraints (45). Thus, we hypothesize that spillovers may occur in the close surroundings of treated parcels. To assess the sensitivity of our results to potential spatial spillovers, we calculated different time-variant indicators of spillover exposure and included them as control variables in a secondary regression analysis. For each year, we calculated the area and share of treated units in buffer zones of 200, 500, 1.000 and 5.000 m as well as the Euclidean distance to the closest treated area. Such spillovers could also occur on control units across stacks from earlier treated units, so we cluster errors at stand level across stacks. In a robustness check, we also used varying assumed error structures, including clustering at municipality level and standard errors robust to spatial autocorrelation (32) (Supplementary Figure E7).

We used R (64) to prepare the data, and *fixest* (65) for the regression-based estimator, robustness checks and spillover analysis. We used OpenAI's ChatGPT (GPT-5.1, accessed throughout 2025-2026) to support code development, implementation testing, and to refine and document computational approaches.

4.5 Carbon calculation

We calculated the impact of METSO in terms of avoided direct carbon emissions from harvested wood products following the calculation in related previous studies (26). Conceptually, we first calculated the carbon stored in the stands of each cohort protected under METSO. We then multiplied the volume by the counterfactual harvesting rate, using the average national estimate for each tree-species to obtain the volume harvested and derived CO₂ emissions from short-lived and long-lived products. In the following, we describe this calculation and its assumptions in detail.

Regarding the scope, we did not consider the emissions at the site after the harvest, nor the future sink after 20+ years as our time scale was below 20 years in which time the impact is approximately zero. We also did not consider the (higher) growth potential of younger stands after a counterfactual harvest, because this would require growth modelling beyond the scope of this analysis. In other words, we focused on the volume of standing timber into short-and long-lived products. Finally, using timber, e.g., in energy production decreases the use of fossil fuels but using biomass has carbon impacts as well (66). We did not consider any substitution effects because we only considered the impacts to the LULUCF sector.

A great majority of Finnish forests are clear cut at the end of the rotation period. Therefore, we assumed clear cuts, no continuous cover management. After the clear cut, the stand remains a source of carbon emissions (from the soil) for at least 20 years, depending on the site type. Therefore, it depends on the time span of the calculation whether the emissions or the future sink potential should be considered. In our case, we consider a rather short time span of 2008-2021, i.e. the period for which we estimated counterfactual harvesting rates.

We estimated the average timber volume and stand age of the treated stands per region, tree species and annual treatment cohort. To do so, we aggregated all available post-treatment

observations of volume per ha and stand age for the respective cohort (available biannually between 2009 and 2023) using the area-weighted mean. We then multiplied the total area of all enrolled parcels in any given year with the volume per ha to get the total timber volume in cubic meters. Timber volumes in cubic meters were converted into biomass (tons). We used biomass expansion factors for pine, spruce and birch from Lehtonen et al. (2004). They are functions of stand age: $v(a + b \times e^{-0,01t})$ where v is timber volume, t is stand age and a and b are parameters determined for different tree species (67). Biomass was converted into carbon (a multiplier of 0,5) and then to carbon dioxide (a multiplier of 44/12).

In Finland, quite a large share of carbon stored in harvested timber is immediately or very shortly released to the atmosphere. The remaining share of the harvested timber is used as wood products where the carbon is stored (and slowly released from this storage). We assumed that 60% of the harvested timber is logs, and half of this is used as wood products, i.e., 30% of the carbon in timber is stored in the wood products (based on how harvested timber is processed in Finland). Thus, 70% of the carbon stored in harvested wood is released into the atmosphere. This includes short-lived products and production side streams. For the remaining 30% carbon stored in long-lasting wood products, we assumed an annual 2% release of emissions.

4.6 Cost effectiveness

We calculated cost effectiveness ratio as cohort–intervention-specific compensation costs divided by cumulative avoided canopy-cover loss, yielding costs per effective hectare-year. First, we estimated the annual avoided canopy cover loss share with the pooled heterogeneity model (Eq. S3) by treatment arms (permanent agreements, recurring state purchases, temporary 10- and 20-year agreements) using all available cohorts (2008-2025) and all observed within the full event time window (−6 to +17). The model also estimated effects for the 2009/2014 state transfers, but these were excluded from the cost-effectiveness calculation because no corresponding cost series was available. This model implies homogenous effects across cohorts of the same intervention arm, a simplifying assumption to improve precision and motivated by our previous analysis that suggested negligible cohort-heterogeneity (Figure 1, Figure E5, Figure E8). We then multiplied each intervention $\times$ cohort-specific event-time estimate of avoided canopy cover loss rate by the observed enrolled area and accumulated these effects from the first post-treatment year to each evaluation horizon. We did not discount environmental effects, nor did we adjust nominal compensation levels per cohort. Compensation costs and enrolled areas were treated as fixed per cohort and intervention. Compensation was calculated from annual national compensation per hectare multiplied by the observed enrolled area and was assigned once to each intervention-cohort. Uncertainty in cumulative effects was propagated using the joint covariance matrix of the event-study coefficients, while confidence intervals for the cost-effectiveness ratios were obtained from 10,000 joint multivariate-normal draws.

Finally, we divided the aggregated compensation per intervention arm and cohort by the cumulative avoided canopy cover loss for the contract modality per cohort. Since the denominator accumulates and increases over time, while the cost numerator remains fixed, the cost effectiveness estimates decrease over event time by construction, increasing cost effectiveness over time. However, only cohorts with complete follow-up to a given horizon

were included, meaning that cohort composition also changes across horizons. We therefore focus on overall trends and magnitudes rather than specific values of any given event time and begin reporting estimates at five years (corresponding to the minimum post-treatment time horizon considered in our main specification).

Acknowledgements and funding sources: This study was conducted within the INTERCEDE project, funded by the European Union’s Horizon Europe Research and Innovation Programme, GA 101135159. JK was funded by the OP Pohjola Research Foundation. We appreciate comments received during the 2025 EAERE conference in Bergen, Norway and workshops held in Helsinki (June 2025) and Stockholm (January 2026). We thank Gustavo Magalhães de Oliveira and Markku Ollikainen for their support and feedback on earlier drafts.

5 References

1. FOREST EUROPE, “State of Europe’s Forests 2020” (Ministerial Conference on the Protection of Forests in Europe - FOREST EUROPE, 2020).
2. B. Muys, *et al.*, “Forest Biodiversity in Europe” (European Forest Institute, 2022).
3. EEA, “The European environment: state and outlook 2020 : knowledge for transition to a sustainable Europe.” (European Environment Agency, 2019).
4. IPBES, “Summary for policymakers of the global assessment report on biodiversity and ecosystem services” (Zenodo, 2019).
5. Council of the European Union, Directive on the conservation of natural habitats and of wild fauna and flora. (1992).
6. EC, EU 2020 biodiversity strategy. (2011).
7. European Parliament, Council of the European Union, Nature restoration regulation. (2024).
8. EC, AGRI, New EU Forest Strategy for 2030. (2021).
9. P. Pulla, *et al.*, “Mapping the distribution of forest ownership in Europe” (European Forest Institute, 2013).
10. S. Wunder, Revisiting the concept of payments for environmental services. *Ecol. Econ.* **117**, 234–243 (2015).
11. S. Wunder, J. Börner, D. Ezzine-de-Blas, S. Feder, S. Pagiola, Payments for Environmental Services: Past Performance and Pending Potentials. *Annu. Rev. Resour. Econ.* **12**, 209–234 (2020).
12. P. Gasparinetti, S. Wunder, L. Bakker, F. D. Gomes, Adverse selection and severe design constraints in PES: A choice experiment in a deforestation hotspot of the Brazilian Cerrado. *Ecol. Econ.* **247**, 109001 (2026).

- 851 13. M. U. Persson, F. Alpizar, Conditional Cash Transfers and Payments for
852 Environmental Services—A Conceptual Framework for Explaining and Judging
853 Differences in Outcomes. *World Dev.* **43**, 124–137 (2013).
- 854 14. K. Baylis, A. Garcia, R. Heilmayr, Causal Inference for Biodiversity Conservation.
855 *Rev. Environ. Econ. Policy* **20**, 98–124 (2026).
- 856 15. J. Börner, D. Schulz, S. Wunder, A. Pfaff, The Effectiveness of Forest Conservation
857 Policies and Programs. *Annu. Rev. Resour. Econ.* **12**, 45–64 (2020).
- 858 16. A. Caruana, et al., Still Money for Nothing? Two Decades of Empirical Evaluation of
859 Conservation Investments. [Preprint] (2026). Available at:
860 <https://ecoevorxiv.org/repository/view/13264/> [Accessed 10 July 2026].
- 861 17. N. E. Kolarik, et al., Conservation easements target high quality lands but do not
862 increase their quality. *Biol. Conserv.* **308**, 111234 (2025).
- 863 18. C. Nolte, S. R. Meyer, K. R. E. Sims, J. R. Thompson, Voluntary, permanent land
864 protection reduces forest loss and development in a rural-urban landscape.
865 *Conserv. Lett.* **12**, e12649 (2019).
- 866 19. J. R. Thompson, et al., Do working forest easements work for conservation?
867 *Environ. Res. Lett.* **19**, 114033 (2024).
- 868 20. K. R. E. Sims, J. M. Alix-Garcia, Emerging design principles for environmental,
869 economic, and equity successes in land conservation. *Proc. Natl. Acad. Sci.* **123**,
870 e2421731122 (2026).
- 871 21. H. S. S. C. Sagar, et al., Carbon finance initiatives can provide biodiversity benefits.
872 *Conserv. Sci. Pract.* **8**, e70254 (2026).
- 873 22. J. A. Zwerts, et al., FSC-certified forest management benefits large mammals
874 compared to non-FSC. *Nature* **628**, 563–568 (2024).
- 875 23. A. Caruana, M. Muir, T. B. White, J. P. G. Jones, Lessons lost: Lack of requirements
876 for post-project evaluation and reporting is hindering evidence-based conservation.
877 *Conserv. Sci. Pract.* **6**, e13260 (2024).
- 878 24. K. Batpurev, et al., Private land conservation through voluntary biodiversity
879 conservation schemes: lessons from a payment for ecosystem services scheme in
880 Finland. *BioScience* b1af155 (2025). <https://doi.org/10.1093/biosci/b1af155>.
- 881 25. P. Horne, Forest owners' acceptance of incentive based policy instruments in forest
882 biodiversity conservation – a choice experiment based approach. *Silva Fenn.* **40**
883 (2006).
- 884 26. J. Kangas, M. Ollikainen, A PES scheme promoting forest biodiversity and carbon
885 sequestration. *For. Policy Econ.* **136**, 102692 (2022).

- 886 27. J. Hohti, *et al.*, “Ten years of METSO - Mid-term review of the first decade of the
887 Forest Biodiversity Action Plan for Southern Finland” (Ministry of the Environment,
888 2019).
- 889 28. J. Hohti, A. Lensu, M. Mönkkönen, J. S. Kotiaho, Evaluating ecological efficiency of
890 voluntary conservation policy in boreal forest conservation. *Conserv. Sci. Pract.*
891 e70049 (2025). <https://doi.org/10.1111/csp2.70049>.
- 892 29. H. Björklund, *et al.*, Predicting valuable forest habitats using an indicator species
893 for biodiversity. *Biol. Conserv.* **249**, 108682 (2020).
- 894 30. P. Shah, A. W. Ando, Permanent and Temporary Policy Incentives for Conservation
895 under Stochastic Returns from Competing Land Uses. *Am. J. Agric. Econ.* **98**, 1074–
896 1094 (2016).
- 897 31. M. J. Roberts, R. N. Lubowski, Enduring Impacts of Land Retirement Policies:
898 Evidence from the Conservation Reserve Program. *Land Econ.* **83**, 516–538 (2007).
- 899 32. T. G. Conley, GMM estimation with cross sectional dependence. *J. Econom.* **92**, 1–
900 45 (1999).
- 901 33. B. Callaway, P. H. C. Sant’Anna, Difference-in-Differences with multiple time
902 periods. *J. Econom.* **225**, 200–230 (2021).
- 903 34. L. Sun, S. Abraham, Estimating dynamic treatment effects in event studies with
904 heterogeneous treatment effects. *J. Econom.* **225**, 175–199 (2021).
- 905 35. A. Rambachan, J. Roth, A More Credible Approach to Parallel Trends. *Rev. Econ.*
906 *Stud.* **90**, 2555–2591 (2023).
- 907 36. M. Torresani, *et al.*, LiDAR GEDI derived tree canopy height heterogeneity reveals
908 patterns of biodiversity in forest ecosystems. *Ecol. Inform.* **76**, 102082 (2023).
- 909 37. M. Koivula, I. Vanha-Majamaa, Experimental evidence on biodiversity impacts of
910 variable retention forestry, prescribed burning, and deadwood manipulation in
911 Fennoscandia. *Ecol. Process.* **9**, 11 (2020).
- 912 38. L. F. Wagenaar, O. Olsson, M. Stjernman, H. G. Smith, A systematic meta-review:
913 The relationship between forest structures and biodiversity in deciduous forests.
914 *For. Ecol. Manag.* **596**, 123072 (2025).
- 915 39. E. Heinaro, T. Tanhuanpää, T. Yrttimaa, M. Holopainen, M. Vastaranta, Airborne
916 laser scanning reveals large tree trunks on forest floor. *For. Ecol. Manag.* **491**,
917 119225 (2021).
- 918 40. J. Toivonen, A. Kangas, M. Maltamo, M. Kukkonen, P. Packalen, Assessing
919 biodiversity using forest structure indicators based on airborne laser scanning data.
920 *For. Ecol. Manag.* **546**, 121376 (2023).
- 921 41. Finnish Forest Centre, Canopy model. Deposited 28 May 2020.

- 922 42. J. Kangas, M. Ollikainen, Reforming a pre-existing biodiversity conservation
923 scheme: Promoting climate co-benefits by a carbon payment. *Ambio* **52**, 1847–
924 1860 (2023).
- 925 43. D. Schulz, *et al.*, Trade-offs between grassland plant biodiversity and yields are
926 heterogenous across Germany. *Commun. Earth Environ.* **5**, 514 (2024).
- 927 44. K. Syrjänen, *et al.*, “Identification of forest ecosystems valuable in terms of
928 biodiversity. Scientific selection criteria of the Forest Biodiversity Programme for
929 Southern Finland (METSO) 2016–2025” (Ministry of the Environment & Ministry of
930 Agriculture and Forestry, 2016).
- 931 45. A. Pfaff, J. Robalino, Spillovers from Conservation Programs. *Annu. Rev. Resour.*
932 *Econ.* **9**, 299–315 (2017).
- 933 46. S. Turubanova, *et al.*, Tree canopy extent and height change in Europe, 2001–2021,
934 quantified using Landsat data archive. *Remote Sens. Environ.* **298**, 113797 (2023).
- 935 47. K. Mäkisara, M. Katila, J. Peräsaari, E. Tomppo, *The Multi-source National Forest*
936 *Inventory of Finland - methods and results 2013* (Natural Resources Institute
937 Finland (Luke), 2016).
- 938 48. K. Mäkisara, M. Katila, J. Peräsaari, *The Multi-source National Forest Inventory of*
939 *Finland - methods and results 2015* (Natural Resources Institute Finland (Luke),
940 2019).
- 941 49. K. Mäkisara, M. Katila, J. Peräsaari, *The Multi-Source National Forest Inventory of*
942 *Finland - methods and results 2017 and 2019* (Natural Resources Institute Finland,
943 2022).
- 944 50. E. Tomppo, M. Katila, K. Mäkisara, J. Peräsaari, *The Multi-source National Forest*
945 *Inventory of Finland - methods and results 2009* (Finnish Forest Research Institute,
946 2013).
- 947 51. E. Tomppo, M. Katila, K. Mäkisara, J. Peräsaari, *The Multi-source National Forest*
948 *Inventory of Finland - methods and results 2011* (Finnish Forest Research Institute,
949 2014).
- 950 52. R. Virkkala, *et al.*, Developing fine-grained nationwide predictions of valuable
951 forests using biodiversity indicator bird species. *Ecol. Appl.* **32**, e2505 (2022).
- 952 53. M. Forsius, *et al.*, Developing a spatially explicit modelling and evaluation
953 framework for integrated carbon sequestration and biodiversity conservation:
954 Application in southern Finland. *Sci. Total Environ.* **775**, 145847 (2021).
- 955 54. J. D. Angrist, J.-S. Pischke, *Mostly harmless econometrics: an empiricist’s*
956 *companion* (Princeton University Press, 2009).

- 957 55. R. Giudice, J. Börner, S. Wunder, E. Cisneros, Selection biases and spillovers from
958 collective conservation incentives in the Peruvian Amazon. *Environ. Res. Lett.* **14**,
959 045004 (2019).
- 960 56. R. Uehleke, M. Petrick, S. Hüttel, Evaluations of agri-environmental schemes based
961 on observational farm data: The importance of covariate selection. *Land Use Policy*
962 **114**, 105950 (2022).
- 963 57. A. Henningsen, *et al.*, Estimating Causal Effects with Observational Data:
964 Guidelines for Agricultural and Applied Economists. [Preprint] (2025). Available at:
965 <https://arxiv.org/abs/2508.02310> [Accessed 9 December 2025].
- 966 58. C. Wing, M. Yozwiak, A. Hollingsworth, S. Freedman, K. Simon, Designing
967 Difference-in-Difference Studies with Staggered Treatment Adoption: Key Concepts
968 and Practical Guidelines. *Annu. Rev. Public Health* **45**, 485–505 (2024).
- 969 59. S. M. Iacus, G. King, G. Porro, Causal Inference without Balance Checking:
970 Coarsened Exact Matching. *Polit. Anal.* **20**, 1–24 (2012).
- 971 60. A. Guizar-Coutiño, D. Coomes, T. Swinfield, J. P. G. Jones, Sensitivity of estimates of
972 the effectiveness of REDD+ projects to matching specifications and moving from
973 pixels to polygons as the unit of analysis. [Preprint] (2024). Available at:
974 <http://biorxiv.org/lookup/doi/10.1101/2024.05.22.595326> [Accessed 23 January
975 2025].
- 976 61. N. Mikkonen, N. Leikola, J. Lehtomäki, P. Halme, A. Moilanen, National high-
977 resolution conservation prioritisation of boreal forests. *For. Ecol. Manag.* **541**,
978 121079 (2023).
- 979 62. C. De Chaisemartin, X. D’Haultfœuille, Two-Way Fixed Effects Estimators with
980 Heterogeneous Treatment Effects. *Am. Econ. Rev.* **110**, 2964–2996 (2020).
- 981 63. K. Borusyak, X. Jaravel, J. Spiess, Revisiting Event-Study Designs: Robust and
982 Efficient Estimation. *Rev. Econ. Stud.* **91**, 3253–3285 (2024).
- 983 64. R Core Team, R: A Language and Environment for Statistical Computing. (2020).
- 984 65. L. Bergé, Efficient estimation of maximum likelihood models with multiple fixed-
985 effects: the R package FENmlm. *CREA Discuss. Pap.* (2018).
- 986 66. P. Leskinen, *et al.*, “Substitution effects of wood-based products in climate change
987 mitigation” (European Forest Institute, 2018).
- 988 67. A. Lehtonen, R. Mäkipää, J. Heikkinen, R. Sievänen, J. Liski, Biomass expansion
989 factors (BEFs) for Scots pine, Norway spruce and birch according to stand age for
990 boreal forests. *For. Ecol. Manag.* **188**, 211–224 (2004).

Table 1: Estimated effects of METSO forest conservation

(1) Dependent variable	Main sample: cohorts 2010–2018		Restricted sample: cohorts 2014–2018		
	(2) TATT [95% CI]	(3) TATT p	(4) TATT [95% CI]	(5) Joint p (pre)	(6) Trend p (pre)
(A) Clearcutting use declaration [%]	-3.1 [-3.4, -2.7]	<0.001	-3.1 [-3.6, -2.7]	0.004	<0.001
(B) Canopy cover loss (Hansen) [%]	-1.4 [-1.7, -1.2]	<0.001	-1.4 [-1.7, -1.2]	0.320	0.542
(C) Forest canopy cover [%]	3.3 [2.7, 3.9]	<0.001	3.3 [2.7, 4.0]	0.008	0.011
(D) Total wood volume [m ³ ha ⁻¹]	10.5 [8.6, 12.3]	<0.001	10.3 [8.3, 12.3]	0.763	0.481
(E) Tree specific age threshold	0.7 [-0.5, 1.9]	0.273	0.3 [-0.3, 1.0]	0.633	0.339
(F) Spruce volume above 150 m ³ ha ⁻¹	2.1 [1.6, 2.6]	<0.001	2.0 [1.3, 2.7]	0.784	0.594

Notes: TATT denotes the trimmed average treatment effect on the treated over five years after treatment. Joint p reports the p-value from a Wald test that all available pre-treatment coefficients are jointly zero. Trend p reports the p-value from a test of a linear pre-treatment trend. The main sample includes cohorts 2010–2018; the restricted sample includes cohorts 2014–2018 and provides longer pre-treatment support.

All-available event studies: pooled estimate and cohort-specific estimates

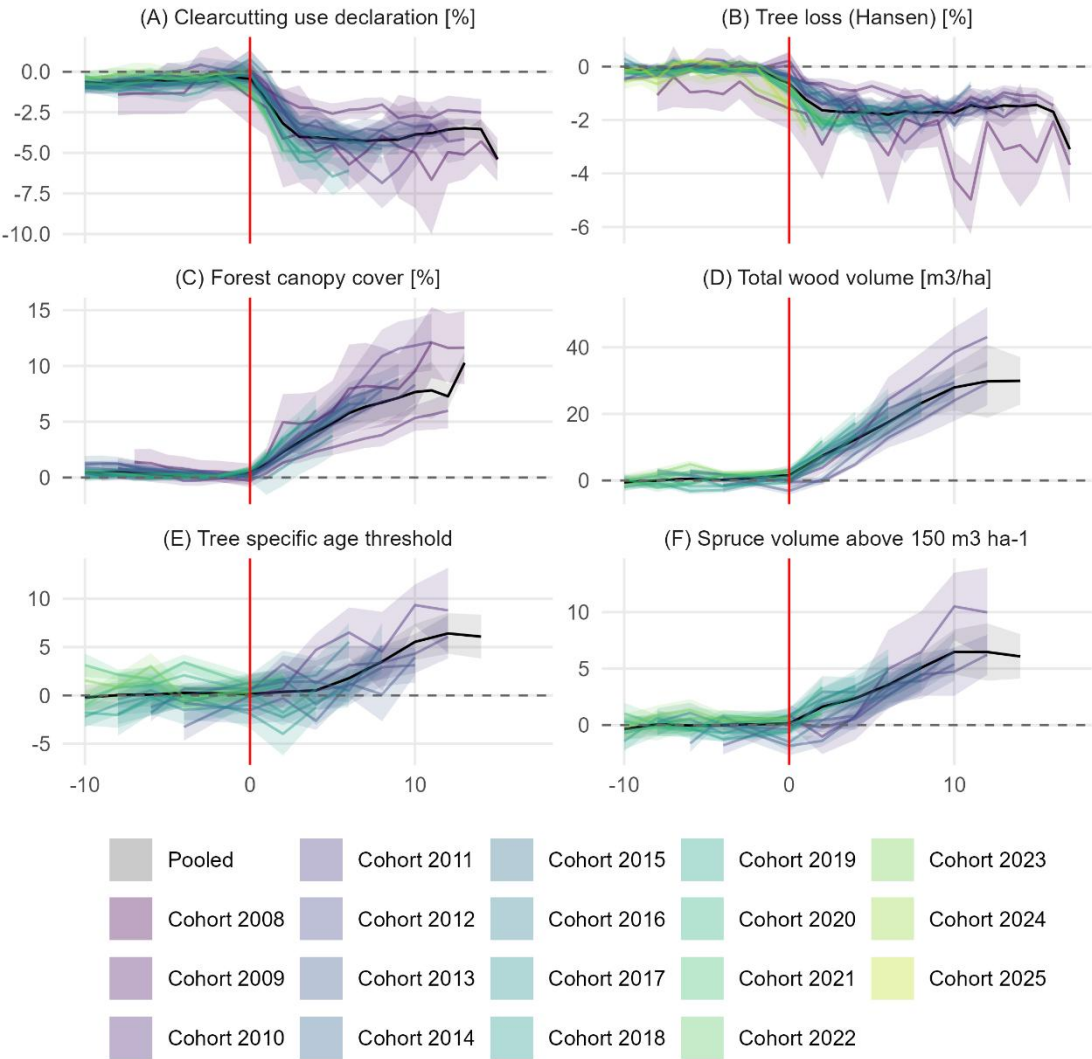


Figure 1: Cohort-wise and pooled event studies

Note: Estimates use all available observations to assess cohort-level pre-treatment trends and explore heterogeneity in post-treatment trajectories. Because cohort composition and temporal support vary across event time, the pooled estimate (black line) is descriptive and should not be interpreted causally. Shaded areas show 95% confidence intervals; the red line marks enrolment.

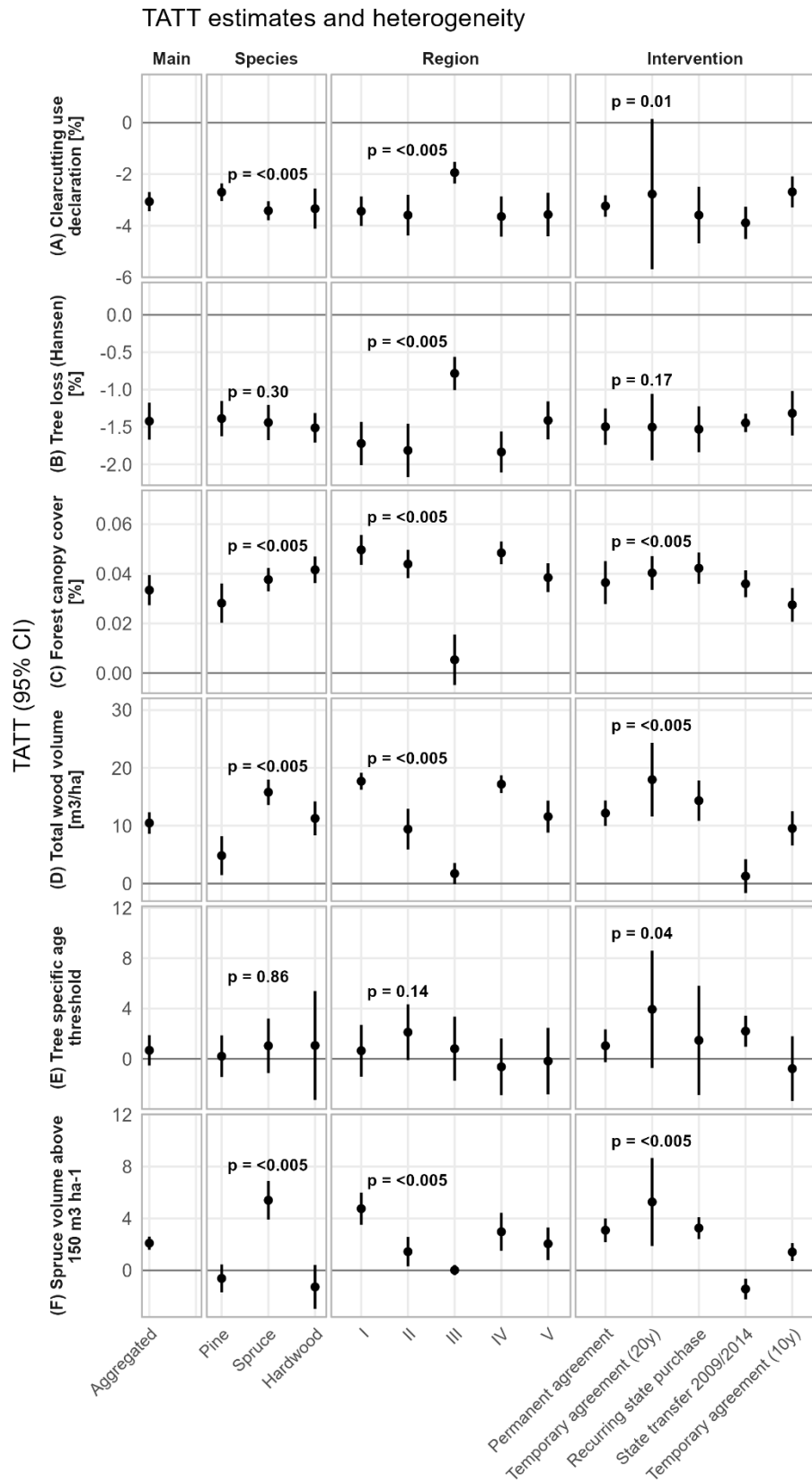


Figure 2: Heterogeneous impacts

Figure note: Five-year TATT by subgroup (Eq. S3).

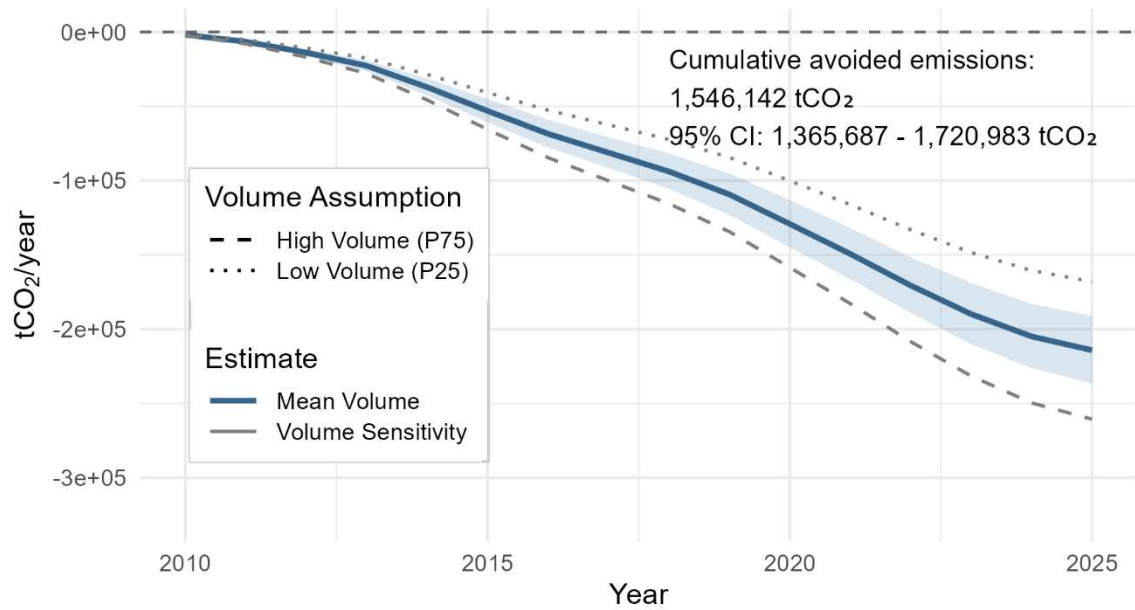


Figure 3: Net atmospheric CO₂ impact

Note: Blue solid line indicates CO₂ calculated from assumed mean stand volume on treated stands across all observed post-treatment periods with 95% confidence bands relating to the counterfactual forest loss rate. Grey estimates refer to 25th and 75th percentile of standing volume, showing sensitivity to assumed self-selection of stands to be harvested. Estimates exclude carbon uptake in standing or (counterfactually) replanted stands.

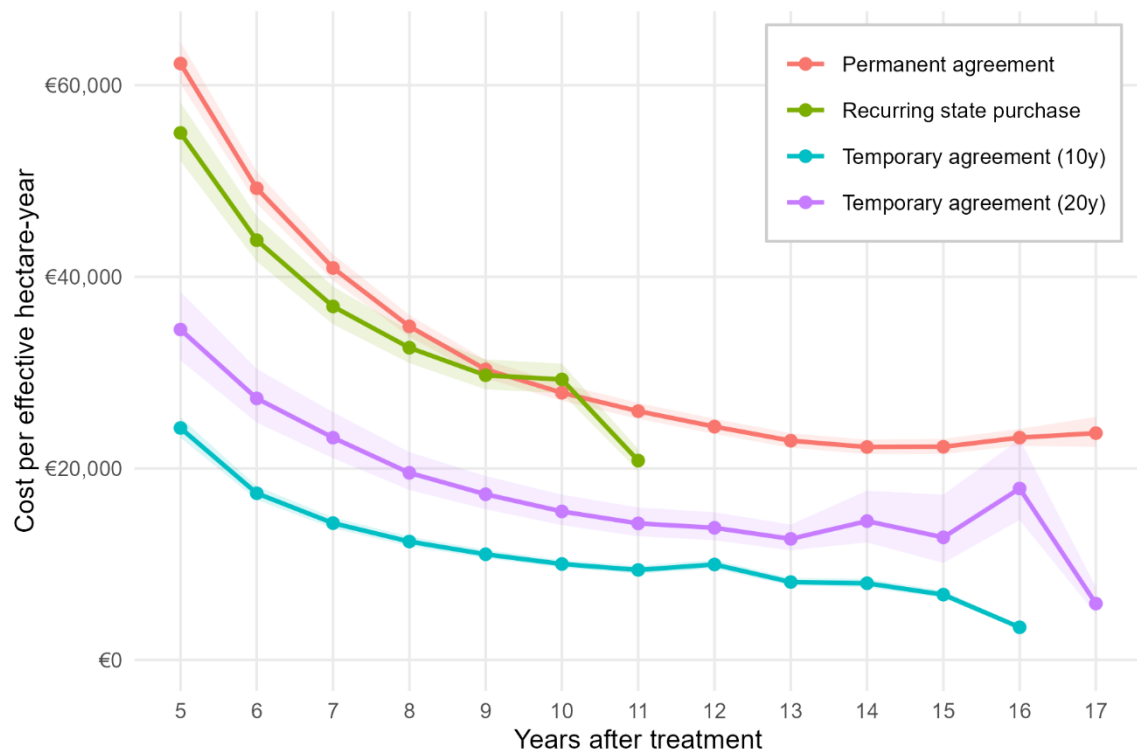
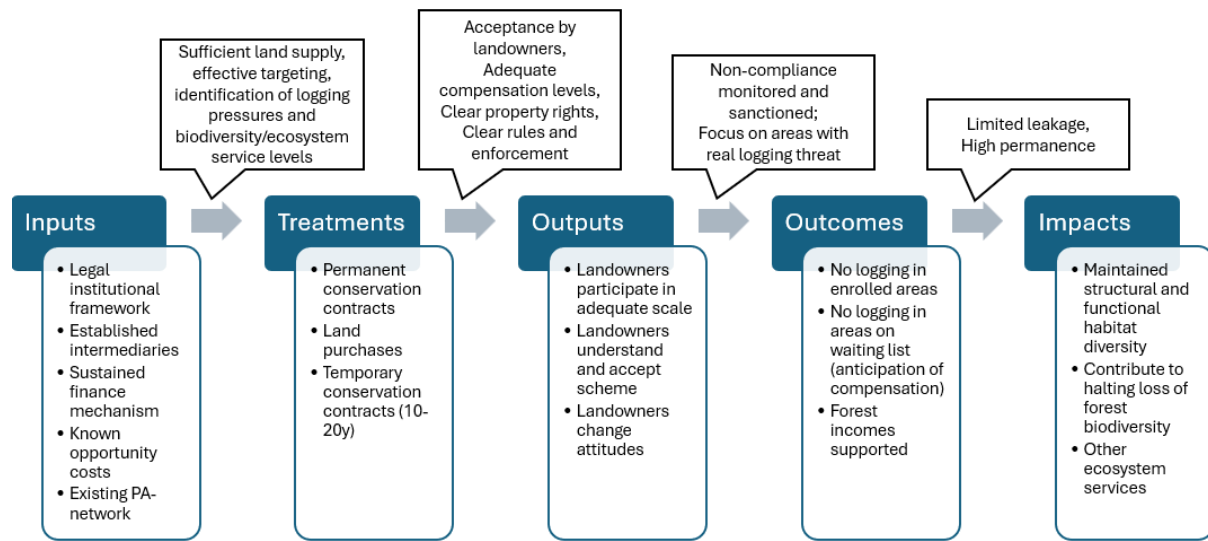
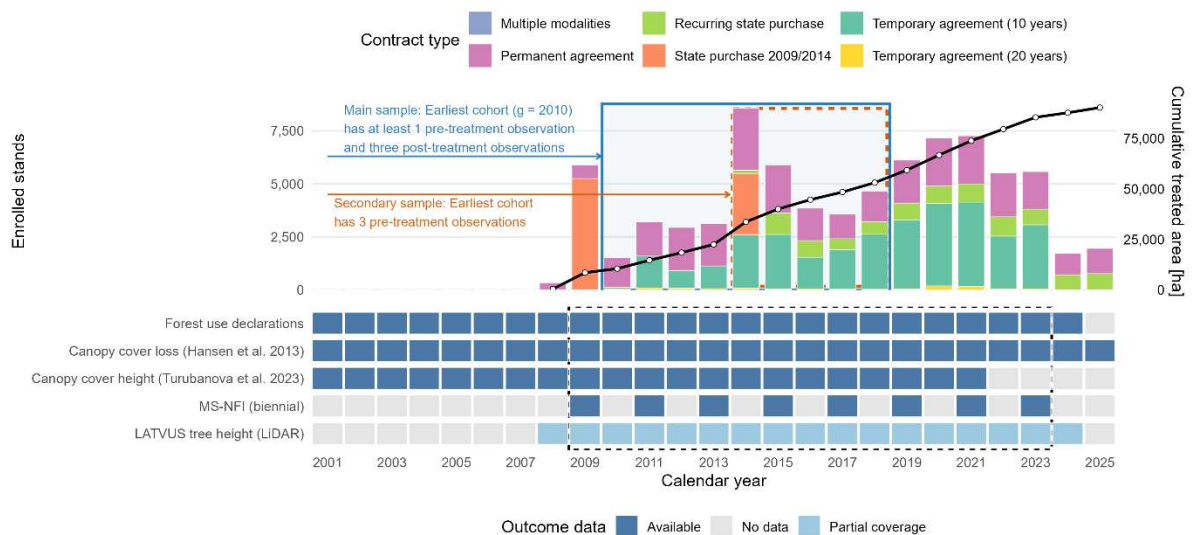


Figure 4: Cost per effective hectare over event time by intervention arm

Extended Data

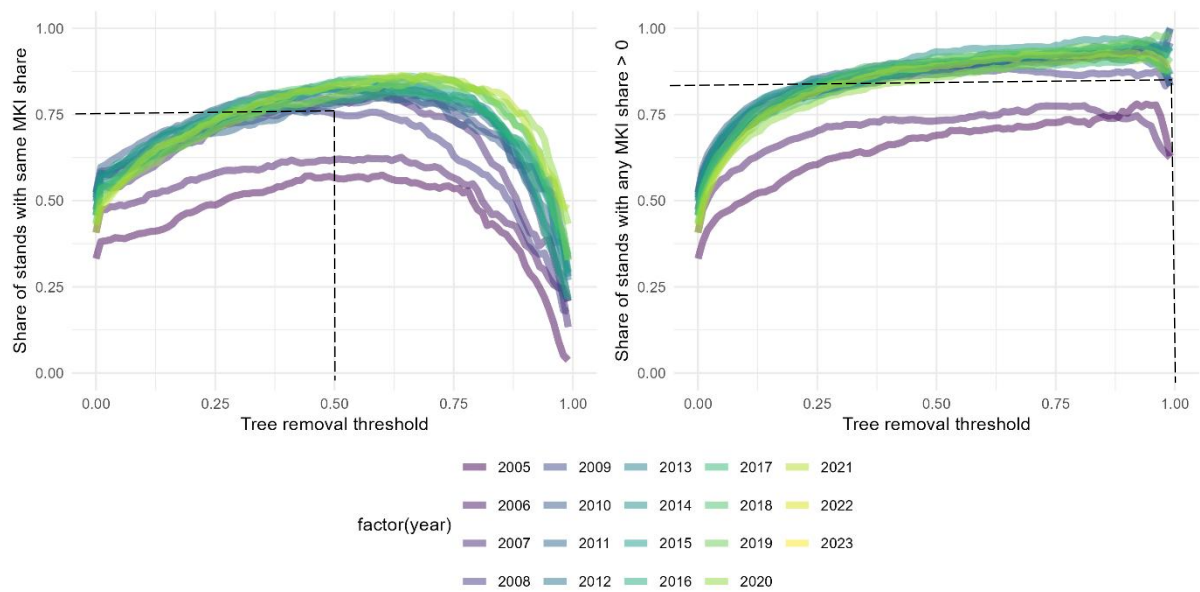


Extended Data Figure E1: METSO theory of change with assumptions on top



Extended Data Figure E2: Cohort timing and outcome data availability

Stacked bars show the number of stands enrolled annually by contract modality, while the black line reports cumulative treated area on the secondary axis. The solid blue outline identifies the main cohort sample (2010–2018), for which the earliest cohort has at least one pre-treatment and three post-treatment observations among all relevant datasets; the dashed orange outline identifies the restricted sample (2014–2018), which provides at least three pre-treatment observations. The lower panel shows calendar-year availability for each outcome dataset; alternating cells for MS-NFI reflect its biennial frequency, while light-blue cells indicate partial coverage. The dashed black outline marks the principal calendar-year observation window (2009–2023). Actual event-time support varies across outcomes and cohorts because datasets differ in temporal coverage.



Extended Data Figure E3: Intention to use vs. actual canopy cover change

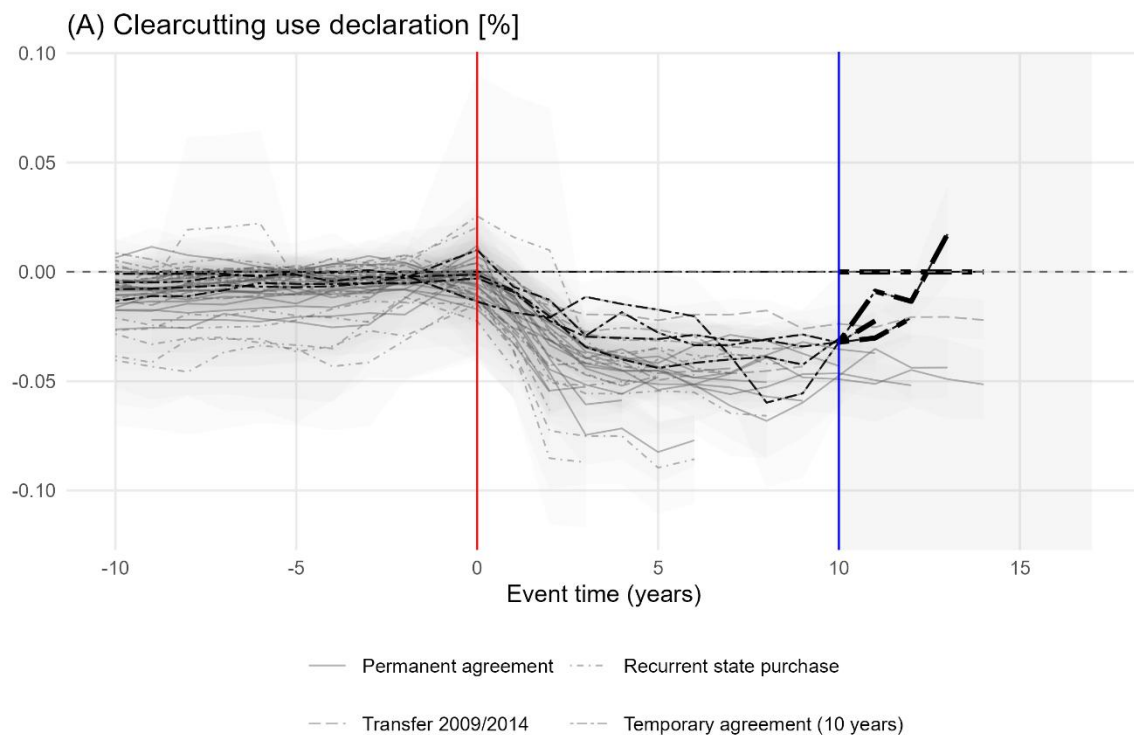
Note: Plots show the relationship between active use declarations and canopy cover change. The left plot indicates the share of stands harvested (Y-axis) that at the time of harvest had an active use declaration that covered at least the same share that changed from canopy to no canopy (X-axis). The right plot indicates the share of stands changing canopy cover that at the time of harvest had any active use declaration. Colors indicate years. Interpretation: left side (dashed lines): between 2008 and 2023, 75% of all stands that lost 50% of their canopy cover in a year had an active use declaration covering at least 50% of the stand area. Interpretation right side (dashed lines): Between 2008 and 2023, more than 80% of stands that lost 100% of their canopy cover had an active use declaration, although that may not have covered the entire stand. Overall, the share of actual canopy cover changes anticipated in use declarations is substantial and has increased over time.



● Diagnostic TATT: cohorts 2008–2014, 10-year post
 ● Diagnostic TATT: cohorts 2008–2024, all available
 ● Diagnostic TATT: cohorts 2014–2018
 ● Primary TATT: cohorts 2010–2018

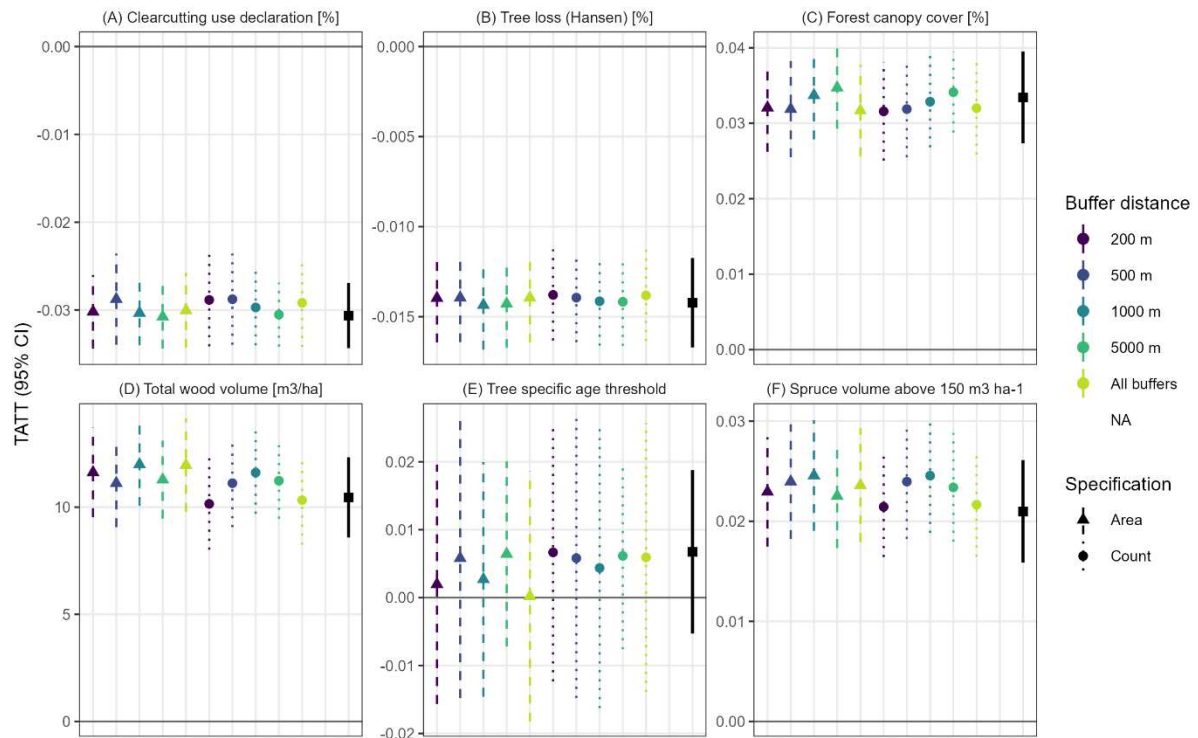
Extended Data Figure E4: Comparison of different estimands and cohort compositions

Note: Estimates using all available post-treatment observations are generally larger, particularly for forest-stock outcomes such as stand age, height and timber volume. This pattern is consistent with the cumulative consequences of avoided harvesting: whereas annual disturbance effects operate on contemporaneous rates, differences in forest stocks persist and grow as protected stands continue to mature while counterfactual stands are harvested. The long-horizon estimates should nevertheless be interpreted descriptively because cohort support declines with event time, so later coefficients are increasingly identified by a small number of early enrolment cohorts and receive equal weight in the aggregated TATT. We therefore retain the common six-year horizon as the primary estimand and use the all-available specification to characterize longer-run dynamics rather than as a directly comparable robustness estimate. See Supplementary Table S4 for parallel trends diagnostics.



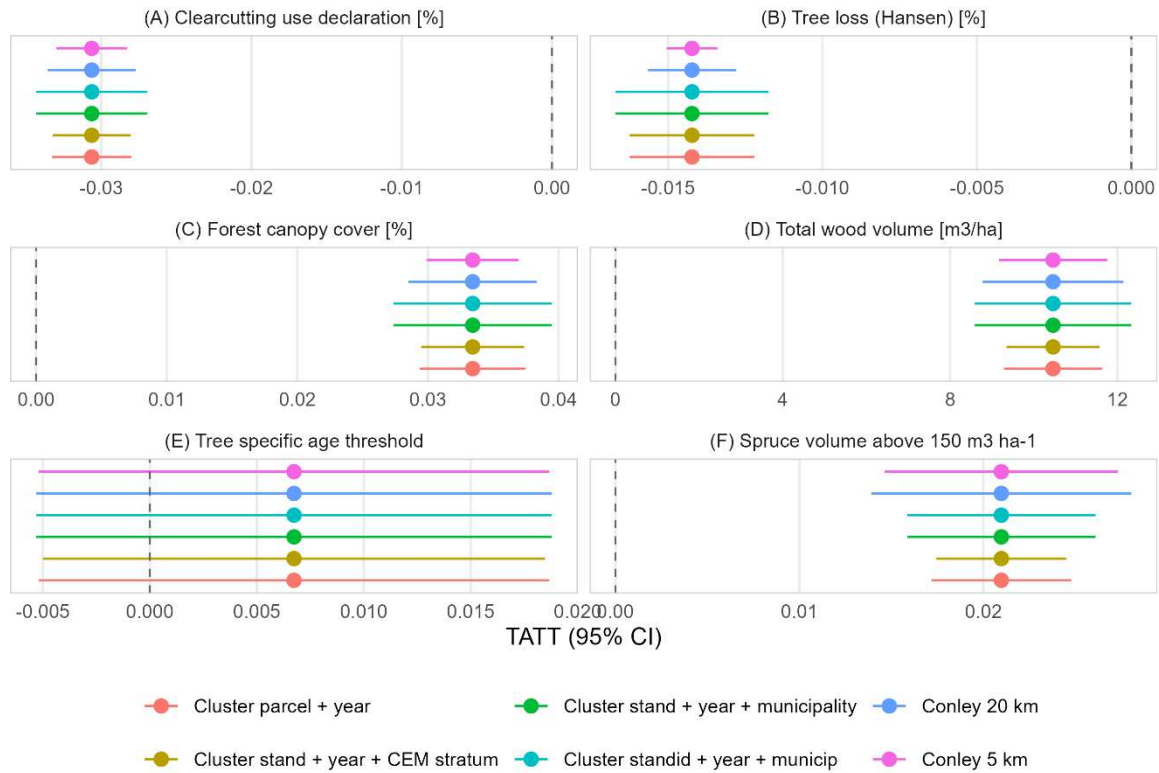
Extended Data Figure E5: Intervention arm by cohort over varying time horizons

Lines represent cohort-specific event-study estimates; shaded areas show 95% confidence intervals. The red line marks enrolment, while the blue line and grey area indicate expiration of 10-year agreements and the post-contract period. Ten-year agreement trajectories with post-expiration observations are highlighted in bold black. Their movement toward zero provides a weak indication of declining protection after expiration, but inference is limited by the few cohorts observed beyond ten years and substantial uncertainty. Furthermore, our available data suggests that 15-30% of temporary 10-year contracts established between 2010 and 2013 had follow-up contracts on at least 90% of their area within three years of contract expiration.



Extended Data Figure E6: Sensitivity analysis to spatial spillovers

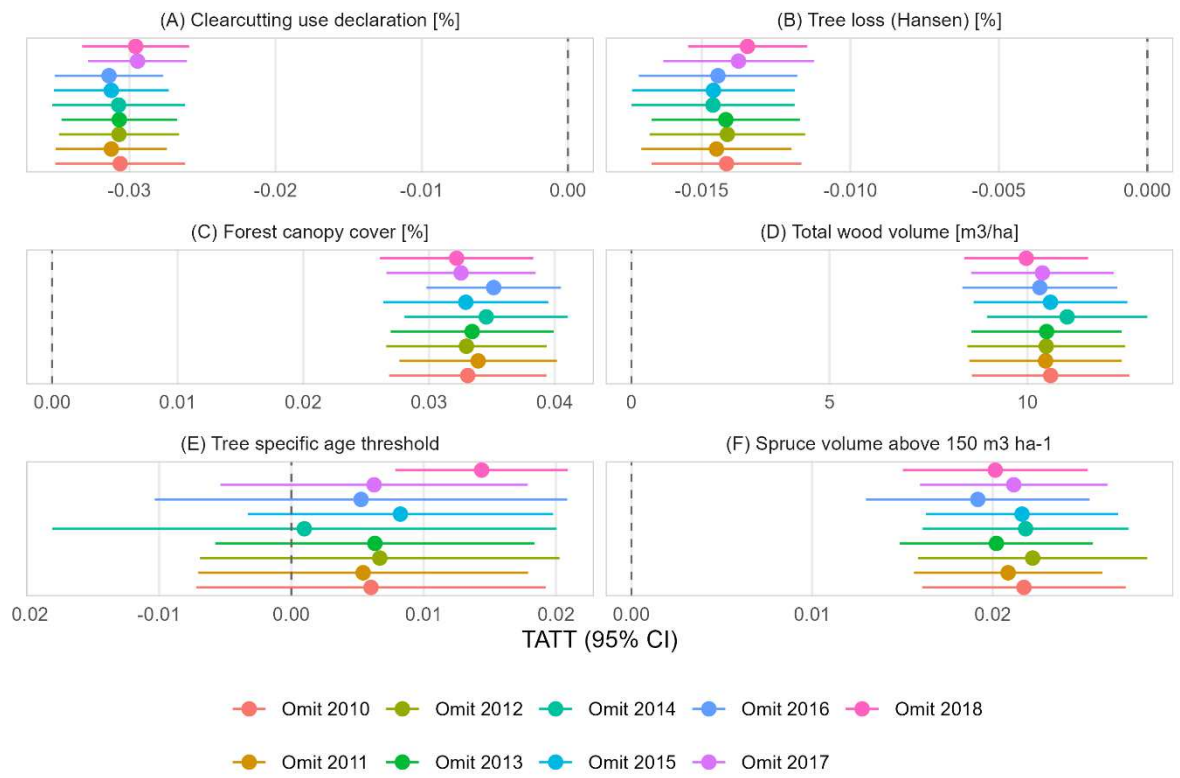
Note: to assess the sensitivity to spatial spillovers, we included a set of spillover exposure indicators as control variables. We calculated for each year the number (count, dot-estimates), and area (hectares, triangle estimates) of treated units within four buffer distances, 200, 500, 1000 and 5000 meters. We also performed a combined regression, including all spillover controls. We used Conley covariance structure (5km cutoff).



Extended Data Figure E7: Alternative assumed error-structures

Note: Panels show primary outcome variables. Estimates display different error structures, namely a simple two-way (parcel x year), adding municipality or CEM-bin for clustering, and Conley standard errors robust to spatial autocorrelation within a 5km and 20km buffer zone.

Leave-one-cohort-out robustness

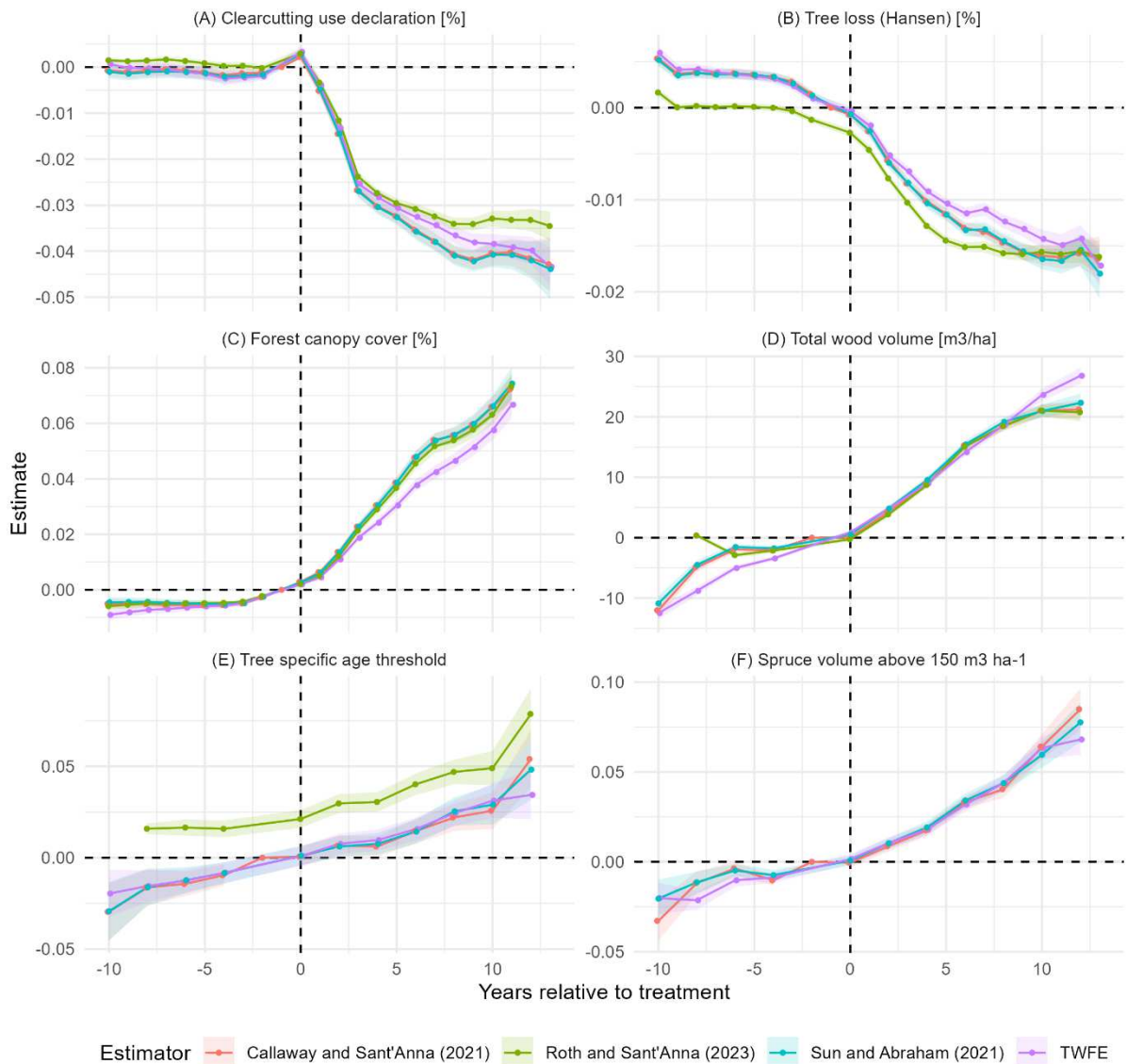


86

87

Extended Data Figure E8: Leave-one-cohort-out analysis

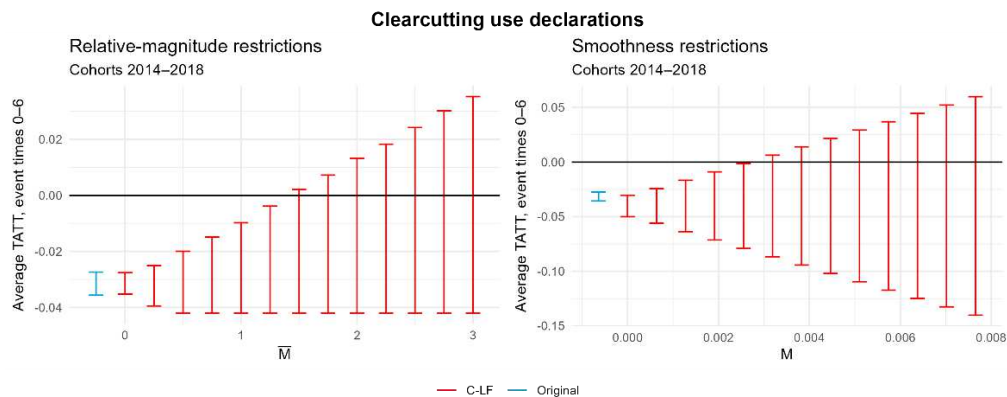
88



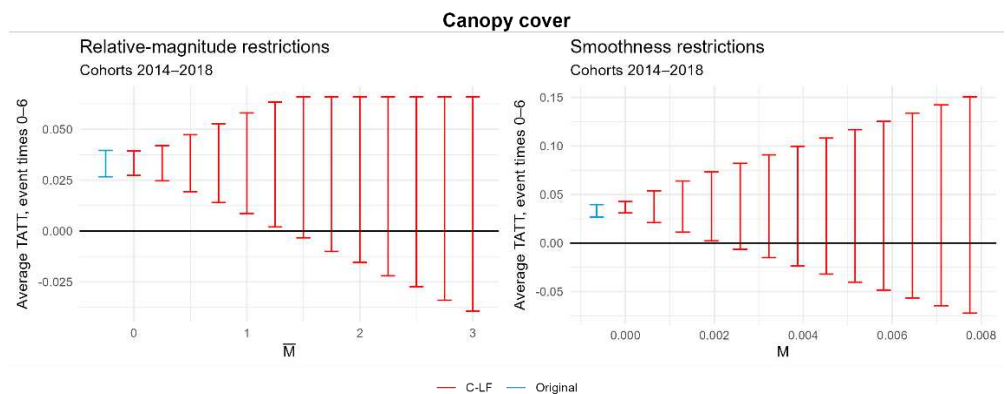
Extended Data Figure E9. Robustness to alternative staggered-adoption estimators.

Event-study estimates for the 2010–2018 cohorts are compared across alternative estimators: conventional TWFE, Sun and Abraham (2021), Callaway and Sant’Anna (2021), and Roth and Sant’Anna (2023). The comparison assesses sensitivity to treatment-effect heterogeneity and the problematic comparisons implicit in conventional TWFE; the Roth–Sant’Anna estimator additionally invokes stronger assumptions concerning quasi-random treatment timing. Unlike our primary Wing et al. (2024) TATT, these estimates use the unique unstacked panel rather than cohort-specific matched sub-experiments and therefore differs in comparison groups and aggregation weights. Stand and municipality-by-year fixed effects and stand-clustered standard errors are included where supported. Shaded areas denote 95% confidence intervals; the vertical dashed line marks enrolment. Estimates at the longest leads and lags draw on fewer cohorts.

A



C



101 Extended Data Figure E10: Honest difference-in-differences for selected outcomes

102 Note: The pre-treatment trend tests in Table 1, columns 5 and 6 indicate possible violations of the parallel trend
 103 assumption for outcomes A) Clearcutting use declaration [%] and C) Forest canopy cover [%]. Honest difference-
 104 in-differences (35) display the sensitivity of our estimates to potential violations of increasing magnitudes. Results
 105 indicate that estimates for outcomes A and C remain significant even if the parallel trends violations in the post
 106 treatment period are 1.3 times stronger than in the observed pre-treatment period (left side), or if the slope of the
 107 pre-trend changes by up to 0.2% (right side).

108

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NatEEMETS0suppinfv1.pdf](#)