Precision ecology for targeted conservation action

Rebecca Spake[1,2], Eleanor E. Jackson[1,3], James M. Bullock[4], Emma Gardner[4], Elizabeth Tipton[5], Matthew Grainger[6], C. Patrick Doncaster[7]

1 School of Biological Sciences, University of Reading, Reading, RG6 6AS, UK

2 School of Geography and Environmental Sciences, University of Southampton, SO17 1BJ, UK

3 Department of Biology, University of Oxford, Oxford, UK.

4 UK Centre for Ecology and Hydrology, Benson Lane, Wallingford OX10 8BB, UK

5 Department of Statistics and Data Science, Northwestern University, U.S

6 Norwegian Institute for Nature Research, Postbox 5685 Torgarden, 7485 Trondheim, Norway

7 School of Biological Sciences, University of Southampton, SO17 1BJ, UK

Corresponding author: Rebecca Spake R.Spake@soton.ac.uk

Abstract

Addressing the coupled threats of catastrophic climate change and biodiversity loss requires implementation of conservation and restoration actions globally. However, on-the-ground action is hindered by context dependency: the ubiquitous challenge that implementation outcomes vary from place to place due to complex dependencies among social and ecological drivers. Policymakers and practitioners recognise the need to tailor solutions to contexts, and target actions to places where they will work effectively. To provide information for decision making, applied ecologists can learn from medicine and marketing, which aim to provide healthcare tailored to individual patients, and advertisements targeting individual tastes. These disciplines exploit big data and rapidly developing computational advances to predict treatment effects for individual units. Here we argue why and how ecological disciplines can begin to capitalise on these rich advances, to equip ecologists with a potentially powerful toolkit for applying big data to site-specific interventions, allowing effective conservation over large extents. We review approaches that hold promise for applied ecology, identify hurdles that must be overcome, and propose a roadmap for establishing the conditions that will permit adoption of precision ecology.

Keywords: Conditional average treatment effect, context dependence, heterogeneous treatment effects, individual treatment effect, propensity score, uplift modelling

Introduction

Environmental sustainability transformations are needed on a global scale[1]. Implementing ecological transformations happens at the local scale, by restoring, conserving, and managing individual ecosystems. Decision-makers are faced with difficult choices of how and where to invest limited resources. Scientists, policymakers and practitioners have long-recognised the futility of silver bullet, 'one-size-fits-all' strategies[2], the need to embrace complexity and context[3], the desirability of targeting actions to contexts where they will be most effective, and of developing place-based action plans tailored to individual sites[4]. For example, international tree planting initiatives often emphasise the need to plant 'the right tree in the right place' to achieve Net Zero[5], because the wrong trees planted in the wrong place may fail to establish, or worse, lead to net carbon emissions that persist for decades[6].

Effective targeting is challenging to achieve in practice because nature is complex. Applied ecologists aiming to develop guidance for policymakers must deal with 'context dependence': the unavoidable geographical and temporal variability in ecological responses to restoration, conservation, and management actions (hereon 'treatments'), according to local social-environmental conditions. For example, impacts of organic farming on biodiversity depend on surrounding landscape structure[7], while relationships between biodiversity and ecosystem function depend on global-change drivers such as drought[8]. Successful targeting and tailoring of treatments requires predictions of treatment effects on biodiversity and ecosystem functions, accounting for the environmental characteristics of these sites.

Applied ecologists do not currently generate causal predictions of treatment effects conditional on their site-level conditions, tending instead to focus either on making causal estimates of average treatment effects across all sites in a sample, or on site-level estimates of effects that are not conditioned on site-level covariates (e.g., by estimating random effects for sampled sites in multilevel models). However, if treatment effects vary in both magnitude and direction across all sampling units, average treatment effects are not actionable, nor are site-level estimates unlinked to site conditions. In human-centred disciplines such as medicine, behavioural sciences, and marketing, practitioners ask "what works, for whom, and under what conditions?"[9,10]. In questions of applied ecology, policymakers and practitioners ask, "what works, where does it work, and under what conditions?", or "where should I intervene in order to have greatest effect for least resource expenditure", for example: "Will *this* treatment improve carbon sequestration in *this* woodland, or, *which* woodland sites would sequester the greatest amount of carbon, if treated in this way?"[11].

Crucially, such lines of questioning ask about treatment effects for specific units, not about estimates of the average treatment effect across all units in a population[12], nor about unit-specific predictions of outcomes under a single treatment level (i.e., without intervention). Human-centred disciplines have begun to transform unit-specific prediction of treatment effects into thriving areas of research, notably in precision medicine[13] and marketing[14], to predict treatment effects on the health of patients, and advertising effects on the purchasing behaviour of potential customers. Given the urgent need to achieve nature recovery, here we argue that the time is ripe to adapt this approach to applied ecology, where it can serve to narrow the enduring gap between scientific knowledge and policy needs (Table 1). Unit-specific treatment effects can be used to tailor a medical drug therapy for particular patients, or to personalise advertising for targeted consumers. For applied ecology, each sampling unit is not a person but a location relevant to a particular grain ('site' from hereon), such as a forest stand, lake, grassland, or arable field. Unit-specific treatment effects could target restoration actions such as tree planting or assisted regeneration to sites where, for example, the most carbon would be sequestered, or where the treatment would contribute most to biodiversity conservation. Unit-specific treatment effects would also help to forecast how particular types of ecosystems in different areas are likely to respond to disturbances, such as fire and pest outbreaks, as they increase in frequency and intensity with global environmental change[15]. Several hurdles remain to be overcome, both general and ecology-specific, to optimise methodologies for predicting unit-specific treatment effects. In this Perspective, we review approaches that hold promise for applied ecology, identify issues and caveats that must be confronted prior to their adoption, and propose a roadmap for establishing the prospects for precision ecology.

Table 1. **Prediction of unit-specific treatment effects are currently a focus of investigation in medicine and marketing, and hold great potential for applied ecology.** Here we compare two examples from medicine[16] and marketing[17] with two potential applications from aquatic and terrestrial applied ecology.

	Medicine	Marketing	Potential application in ecology	
Context and reason for unit-specific treatment effect prediction	It is important to understand the heterogeneity in survival benefits of ventricular assistance devices in order to improve the current transplant priority allocation scheme[16].	Advertising resources are wasted on individuals who will never buy a product, or who will buy regardless of advertising. Targeting advertisements to individual consumers will increase return on investment[17].	Vegetation buffer zones are costly and vary in their effectiveness at maintaining aquatic biodiversity in catchments with surface water pollution. Where they are ineffective, alternate measures should be implemented.	Maintaining or increasing soil carbon by restoration could have many benefits for climate change mitigation, adaptation, and biodiversity conservation, yet restoration effects are highly variable. Areas where restoration actions result in large losses should be avoided.
Sampling units in sample	Patient awaiting heart donation	Potential customer	Lake	Forest stand
Data source	The United Network for Organ Sharing (UNOS) dataset	Marketing dataset obtained from Kevin Hillstrom's MineThatData blog[18]	National or citizen-science monitoring schemes Nutrient loading data.	Remote sensing National or citizen-science monitoring schemes
Reference sample	Patients	Customers	Lakes distributed across broad geographic extent	Forest stands distributed across a heterogeneous covariate landscape
Outcome	5-year survival probability of patients awaiting a heart donor	Amount of money spent purchasing products	Eutrophic status or presence of indicator taxa	Soil organic carbon tonnes/ha
Treatment (*X*)	Fitted with a Left Ventricular Assistance Device	Receives email from an internet-based retailer campaign	Planting of vegetation buffer zones	Restoration by stand thinning
Example covariates (*Z*) that influence treatment outcomes or assignment	Age; sex; comorbidities (e.g., diabetes); body mass index.	Amount of money spent in the previous year; date of last purchase; rural, suburban, urban zip code.	Size; depth; composition and configuration of catchment-level land uses; topography; climate; atmospheric deposition of pollutants.	Soil type and drainage; stand density; climate; topography; previous management; stand density; proportion of basal area contributed by different species.
Target	Current patients requiring transplant	Individual customers from reference sample	Currently untreated individual lakes in sample and out of sample	Currently untreated individual stands in sample and out of sample

Treatment effects: from estimating population averages to predicting individual, site-specific outcomes

Questions about treatment effects require causal-inference methods that have roots in the ‘potential outcomes framework’, in which treatment effects are derived from counterfactual comparisons of outcomes that would result from alternative treatment levels (labelled treatment and control). Suppose we are interested in the causal effect of forest thinning on soil carbon in the *i*th forest stand (Figure 1, top). In this example, the causal treatment can take only two values: an unthinned control ($Xi = 0$, where X corresponds to the treatment) or a thinned treatment ($Xi = 1$). When the plot is a control, its soil carbon outcome is $Y_i^{X=0}$. When the same plot is treated, its outcome is $Y_i^{X=1}$. Both $Y_i^{X=0}$ *and* $Y_i^{X=1}$ are called potential outcomes because either one is potentially observable. The difference between these potential outcomes is the plot-level causal effect of thinning, i.e., the individual treatment effect for plot *i.*

The Fundamental Problem of Causal Inference[19] is that we can never observe both potential outcomes, nor the associated treatment effect, for any individual unit (here a plot). To address this missing data problem, the potential outcomes framework typically considers how to estimate average effects. In a randomised controlled trial (RCT), because of random assignment, we can get an unbiased estimate of the average of these individual treatment effects, called the **Average Treatment Effect (ATE).** Thus, RCTs often focus on estimating the ATE, and are widely used when testing agricultural intervention effectiveness[20], and in fields including medicine, political science[21] and ecology[22]. However, because the sample in the RCT is never truly a random sample of any specific target population, generalising this ATE to other units outside of the RCT is much harder. That is, while the ATE estimate is unbiased for the units within the sample, it may be a biased estimate of the ATE for a broader population of units outside of the sample.

This generalisability problem arises when treatment effects vary across individuals. Unit-specific effects can differ in both magnitude and sign from the population ATE (Figure 1). The distribution of a treatment effect across sampling units is shifted and shaped by baseline differences and variability in the direction and magnitude of treatment effects across individual sampling units. Accordingly, several scientific fields attempt to estimate **Conditional Average Treatment Effects (CATEs),** the expected treatment effect for a sampling unit, conditional on the average covariate profile for the unit’s subgroup[13], i.e., the treatment effect among individuals with the same vector of covariates values. As an estimate, CATE relies on each treatment level-by-population subgroup having sufficient cases to reliably estimate the effect, which is often unachievable. The CATE nevertheless holds special interest for researchers because it enables an understanding of how treatment effects vary depending on the observed characteristics of each sampling unit, allowing treatments to be targeted effectively to units.

The simplest methods for estimating CATEs include stratification into subgroups or fitting models containing statistical interactions[23]. If the number of covariates is small, and all covariates are discrete (e.g., tree species, the presence of a ditch) stratification can be effective. But if covariates are numerous and continuous (e.g., rainfall, forest age), they are often either made discrete and estimated using stratification, or are (parametrically) interacted with the treatment effect in pre-specified regression models, e.g., generalised linear (multilevel) models. Multi-level models (and meta-regression) can estimate effects for clusters of sampling units within a sample, yet unless the sample is a random sample from the target population, this approach does not work for predicting effects in the target. For interactions, the researcher must make choices and assumptions about which covariates to include, their functional forms, and how to specify treatment-by-covariate interactions[24]. Often, researchers using these approaches focus on hypothesis testing, not prediction, though these same methods could be used for predictive purposes. If so, they would be contingent upon the model selected (by the analyst) and the parametric assumptions of that model.

CATEs only partly address the generalisability problem. They provide a means to estimate subgroup ATEs within the sample included in the RCT. But for policymaking and decision making – e.g., which medical intervention to use, which advertisement should be used for different individuals, or which forests to target for restoration actions – the goal is not one of estimation within the sample, but instead one of predicting unit-specific treatment effects (often referred to as individual treatment effects: henceforth 'ITEs') *outside the sample*. The prediction of treatment effects for individual units – what is called 'causal prediction' – is a relatively new field and combines causal inference approaches with predictive methods. Causal prediction is a harder problem than either the ATE or CATE prediction of outcomes alone, because it is impossible to observe the individual treatment effects that are desired. Nonetheless, flexible methods of predicting ITEs are under rapid development in fields including precision medicine and marketing, which use large datasets and non-parametric models that make no assumptions about the parametric form of the relationship between treatment effects and covariates. Among the nonparametric approaches available, the machine learning toolbox for ITE prediction is expanding rapidly[13,25,26].

Here we limit our discussion to binary treatments (e.g., intervention and control), noting that multiple treatment arms and continuous treatments are conceptually possible within the potential outcomes framework. Nevertheless, methods for binary treatments remain the most developed[27], with less emphasis on unit-specific prediction for continuous treatment effects from observational data[28]. In applied ecology, continuous treatments might include thinning intensity, fire intensity, riparian buffer width, etc.

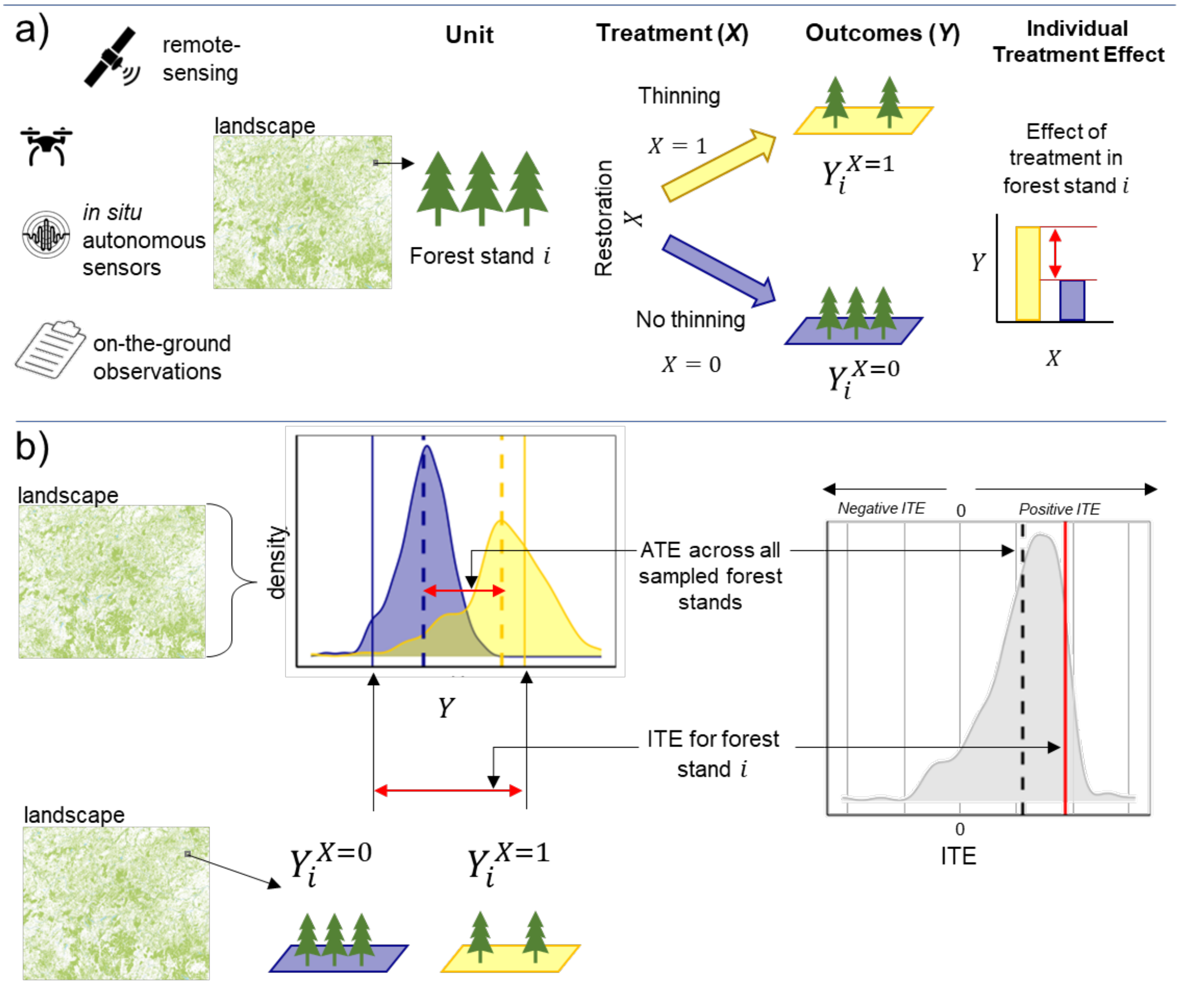


Fig. 1. Individual treatment effects can differ in magnitude and sign from the average treatment effect. Here shown is a hypothetical field experiment, in which a binary treatment *X* (forest thinning vs no thinning) is applied to forest stands across a heterogeneous covariate landscape. a) Data from multiple sources characterise forest stands across a broad geographic extent (left). For a specific stand *i*, we can consider the potential outcomes (*Y*) without ($Y_i^{X=0}$; blue) and with restoration thinning treatments ($Y_i^{X=1}$; yellow). Outcomes might relate to soil organic carbon, biodiversity, or probability of invasive species establishing. The difference between them corresponds to that stand's unit-specific, individual treatment effect (ITE). b) The distributions of outcome (*Y*) show the potential outcomes for all forest stands with (*X* = 0; blue) or without (*X* = 1; yellow) restorative thinning, and the distributions of unit-specific ITEs for all forest stands, demonstrating heterogeneity in treatment effects (grey). The coloured vertical dashed lines show the treatment geometric means for each treatment group; the difference between them represents the average treatment effect (ATE). The continuous lines show the potential outcomes after being treated (yellow) or not treated (blue) for one particular forest stand *i*, and the resulting unit-specific treatment effect. Here, the ITE for forest stand *i* (red line) is larger in magnitude than the ATE (dashed black line). Importantly, there are unit-specific ITE that are opposite in sign to the ATE.

Assumptions required for predicting individual treatment effects

ITE prediction is possible from a dataset of either design-based or observational origin, with covariates Z that present a potential source of confounding variation. In observational studies, Z needs to include all covariates related to the outcome Y and/or the assignment of treatment X.

ITE prediction rests on the same three strong assumptions as those of ATE estimation. 'Unconfoundedness' (or selection on observables) assumes there are no unobserved confounding variables that determine treatment allocation. 'Positivity' (or common support) assumes that every unit has a non-zero probability of being in either treatment group. If not met, differences in covariate overlap between the treatment groups may create regions in the relevant covariate space without appropriate comparators, i.e., where only treated, or only control, units are present. While positivity is implicitly fulfilled by a randomised design, observational studies may not have common support. Positivity can be tested directly with observational data, and induced by weighting on propensity score: the probability of a unit being assigned to a particular treatment level given a set of observed covariates. Finally, the 'Stable Unit Treatment Value Assumption' (SUTVA) is met when there is no interference amongst units: one unit's response to treatment is unaffected by other units' assignments. While SUTVA might seem restrictive, ecological studies that aim to infer causation are typically designed with SUTVA in mind[29,30]. For example, researchers might enforce a particular distance or lag between units in recognition of spatial spillover and temporal carryover effects, when units are influenced by neighbouring or prior treatments, respectively, or they may aggregate smaller units into larger units (e.g. quadrats into sites, streams into catchments).

For a binary outcome variable, the unit-specific treatment effect can be interpreted as the predicted difference in probability of an outcome for an individual with covariate values z_i under two different treatment conditions. In practice, there may not be sufficient cases with $Z = z_i$ under both treatments to reliably predict unit-specific treatment effects. This brings us to the need for a predictive model to smooth over the gaps in the observed set of all z_i across both treatments.

Approaches to ITE prediction: meta-learner algorithms

A range of ITE prediction approaches has been developed to exploit heterogeneous covariate data for diverse data types including RCTs and observational studies[31,32]. These prediction approaches are referred to as 'meta-learner' algorithms (see Box 1)[33]. They can apply any supervised learning or regression method (random forests, Bayesian additive regression trees, neural networks, etc[33–35]), although most often use machine learning. Here we describe four meta-learners: S-, T- X- and R-learners[33] (Box 1). These algorithms differ in how they handle treatment assignment *X*, and how they adjust for biases inherent to observational studies.

Meta-learners can be classified into the more simple 'conditional mean regression methods' of S- and T-learners, and the more complex 'pseudo-outcome' methods[24,34,36] of X-, R- and DR-learners. S- and T-learners do not account for selection biases; they predict ITEs indirectly, by first predicting the potential outcomes ($Y_i^{X=1}$ and $Y_i^{X=0}$) separately, and then taking the difference between these response surfaces. In other words, conditional mean regression methods rely on estimating conditional mean functions *Y* only. Several studies in medicine, marketing and statistics have used simulation to evaluate the relative performances of different meta-learners under various sampling and data conditions[25,33,36,37]. For conditional mean regression methods, S-learners perform poorly when treatment and control groups have very different covariate distributions causing positivity violations. Moreover, since machine learning models may regularise to omit predictors with little influence, S-learners can bias small-magnitude treatment effects to zero[33]. T-learner models should generally perform better than S-learners when the treatment effect is small or when the effects of covariates on outcomes differ between control and treatment groups[36]. However, a bias–variance trade-off can arise when predicting treatment effects based on two separate outcome models (T-learner); the larger sampling variance induced by data splitting may lead to more misclassifications of binary outcomes than for the single-model S-learner[38].

The more advanced pseudo-outcome methods including X-, R- and DR-learners predict ITEs directly using combined models with functions that attempt to account for selection bias[13]. These meta-learners involve more steps, and they incorporate information from the propensity score in order to increase statistical efficiency[25]. Pseudo-outcome methods model intermediate ITE predictions ('pseudo-outcomes') as a function of covariates, and thus can remove some of the bias induced by regularisation and overfitting compared to the S-learner and the T-learner. X-learners tend to perform better than conditional mean regression methods in the presence of unbalanced treatment group sizes and sparsity in areas of covariate space[33]. The pseudo-outcome of the DR-estimator is 'doubly-robust', and it predicts

ITEs well provided either of the two outcome models is correctly specified[24]. DR-learners perform poorly if an important confounder is omitted, or if there are near violations of the overlap assumption. R-learners are less sensitive to extreme propensity scores[36,39].

Box 1. Meta-learners for predicting individual treatment effects

Prediction of ITEs is possible when ecologists have site-level data on ecosystem outcomes of interest (*Y*), related to biodiversity and ecosystem functioning (e.g., species occurrences, lake water quality), information on treatments *(X)* that sites have been subjected to, and other environmental covariates (*Z*) that also predict those outcomes. The table (Box 1 Figure 1a) shows an example dataset targeting the effect of forest restoration (thinning) on soil organic carbon (SOC) 10 years after treatment (*Y,* tonnes/ha) as a function of covariates *Z* comprising mean annual rainfall (mm), temperature (˚C), pre-treatment SOC, initial soil carbon (tonnes/ha), canopy and slope (%).

Meta-learners are model-agnostic algorithms that decompose the task of ITE prediction for binary treatments into multiple sub-regression problems that can be solved using any modelling method, such as supervised learning or regression. Of the several learners that have been developed, here we describe four: S-, T-, X- and R-learners.

S-learners (single-model learners)

S-learners[40,41] (Box 1 Figure 1b) are the simplest algorithms, similar to those currently used in ecological modelling to predict outcome *Y.* S-learners predict ITEs indirectly by training a single outcome model, M_s, to predict outcomes *Y* as a function of covariates *Z*, handling treatment *X* like any other covariate in vector *Z*. The same model is used to predict outcomes for individual sampling units *i*, forcing control (*X*=0) and treatment (*X*=1) conditions. For site *i*, the ITE is the difference in predictions between the treatment and control, while holding all other covariate values fixed for the individual site in question.

T-learners (two-model learners)

T-learners[32,42,43] (Box 1 Figure 1c), or 'two-model learners' predict ITE indirectly by training two outcome models, M_1 and M_0 to predict outcomes *Y* separately for treatment and control datasets, respectively. Both models are used to predict outcomes for individual sampling units *i,* with the ITE as the difference between these predicted outcomes.

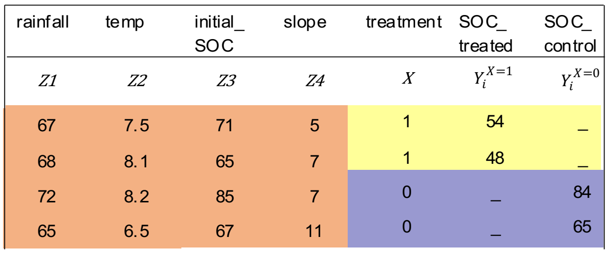
X-learners (cross-learners)

X-learners[33] (Box 1 Figure 1d), predict ITEs directly, and are designed for observational studies where positivity assumptions might otherwise be violated. Like T-learners, two outcome models are initially fitted (M_1 and M_0) to predict outcomes *Y* separately for treatment and control datasets, respectively. A propensity score model (M_{ps}) is also fitted, predicting the treatment probability (X=1) given *Z*. These outcome models and propensity score models are often referred to as 'nuisance functions' in the machine learning literature. Intermediate treatment effects are imputed from M_0 and M_1, using *Y* and *Z* for treated and control datasets respectively (hence the crossing over, Box 1 Figure 1d). A second pair of models is fitted to predict these intermediate treatment effects. Finally, the predicted treatment effects are adjusted by the propensity scores to predict ITEs. The adjustment puts more weight on treatment effects that have been estimated more precisely, i.e. the ones coming from the larger treated or control sample, respectively.

R-learners (residualisation learners)

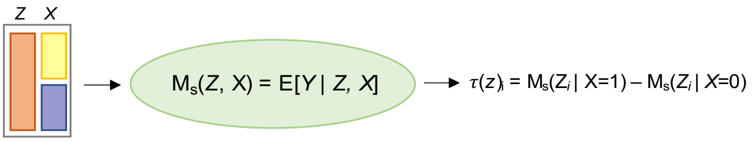
Like the X-learner, the R-learner approach[39] (Box 1 Figure 1e) is also designed for observational studies with selection bias, and so begins by fitting nuisance functions before training a model to predict ITEs directly. The 'R' denotes the residualisation approach, and also recognises the foundational work by Robinson (1988)[44]. The R-learner first trains two models: a single outcome model, M_r, to predict outcomes *Y* as a function of covariates *Z* (excluding the treatment indicator *X*), and a propensity score model M_{ps}. It then residualises the outcome *Y* and treatment *X* by the predictions of the M_r and M_{ps}, respectively, to construct a modified outcome (φ_i). In the second step, the R-learner trains a modified outcome model on the covariates *Z*, weighted by the squared residualised treatment to predict ITEs[45].

a) Example input dataset for meta-learner models

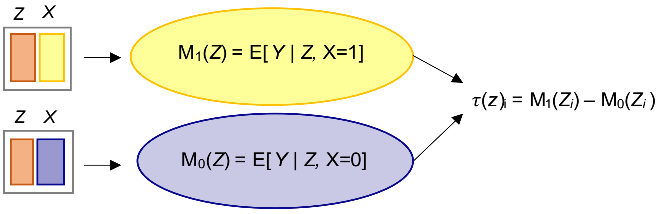


rainfall	temp	initial_SOC	slope	treatment	SOC_treated	SOC_control
Z1	*Z2*	*Z3*	*Z4*	*X*	$Y_i^{X=1}$	$Y_i^{X=0}$
67	7.5	71	5	1	54	_
68	8.1	65	7	1	48	_
72	8.2	85	7	0	_	84
65	6.5	67	11	0	_	65

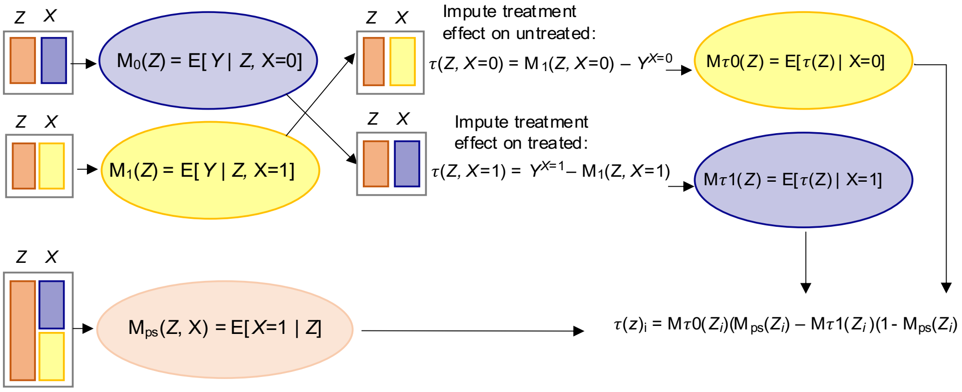
b) S-learner model



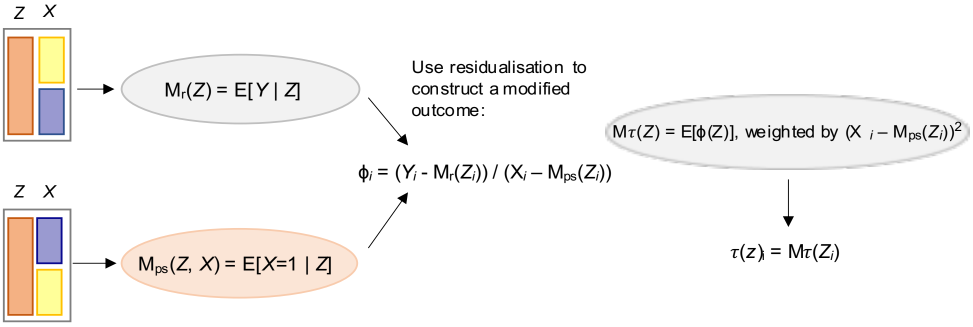
c) T-learner model



d) X-learner model



e) R-learner model



Box 1 Figure 1. Alternative meta-learner models. An example dataset (a) used to train different types of meta-learner models (b-e).

Unit-specific treatment effects enable tailoring and targeting of interventions to specific locations

A shift in applied ecology’s focus to ITE predictions would provide options for improving the efficacy of conservation, restoration, and management interventions. ITEs could support decision-making at two scales. Firstly, at the site level (e.g., a forest stand within a protected area), predicting treatment effects for *at least two different treatments* (i.e., management actions), specific to the covariate profile of the site would enable selection of the treatment that yields the greatest desired effect. Secondly, across broad extents (e.g. forest stands across a nation), predictions of unit-specific treatment effects for *multiple sites* would allow policy-makers with a fixed budget to prioritise the delivery of treatments to sites with the largest ITEs for a given treatment (i.e., the top *N*th percentile of unit-specific treatment effects). Finally, if unit-specific predictions for a given treatment are available for multiple *outcome* variables (e.g., carbon storage, tree health, biodiversity), predictions could help to quantify the trade-offs among these outcomes[4].

ITE prediction could facilitate targeted action by classifying units according to the direction and magnitude of the ITE. For example, uplift modelling, a family of techniques used in marketing, seeks to predict ‘uplift’, the incremental impact of a treatment (a marketing action) on an individual's behaviour. Uplift modelling classifies individual units (customers) according to a binary treatment (whether shown an advert or not) and a binary outcome (whether they purchase a product or not). Advertising resources are wasted on *Lost Causes* and *Sure Things* and should be targeted to *Persuadables*: individuals that would buy only if shown an advert; *Do Not Disturbs* should be avoided (Figure 2A).

The same principles translate to applied ecology for identifying currently unmanaged units that should or should not be targeted for a specific management intervention. For example, a clearcutting and restocking intervention has an outcome of change in total forest soil carbon that varies spatially, according to a range of climatic, soil and topographic covariates. If the objective of management is to increase soil carbon, action should be targeted to stands where clearcutting/restocking increases soil carbon (*Persuadables*; Figure 2B).

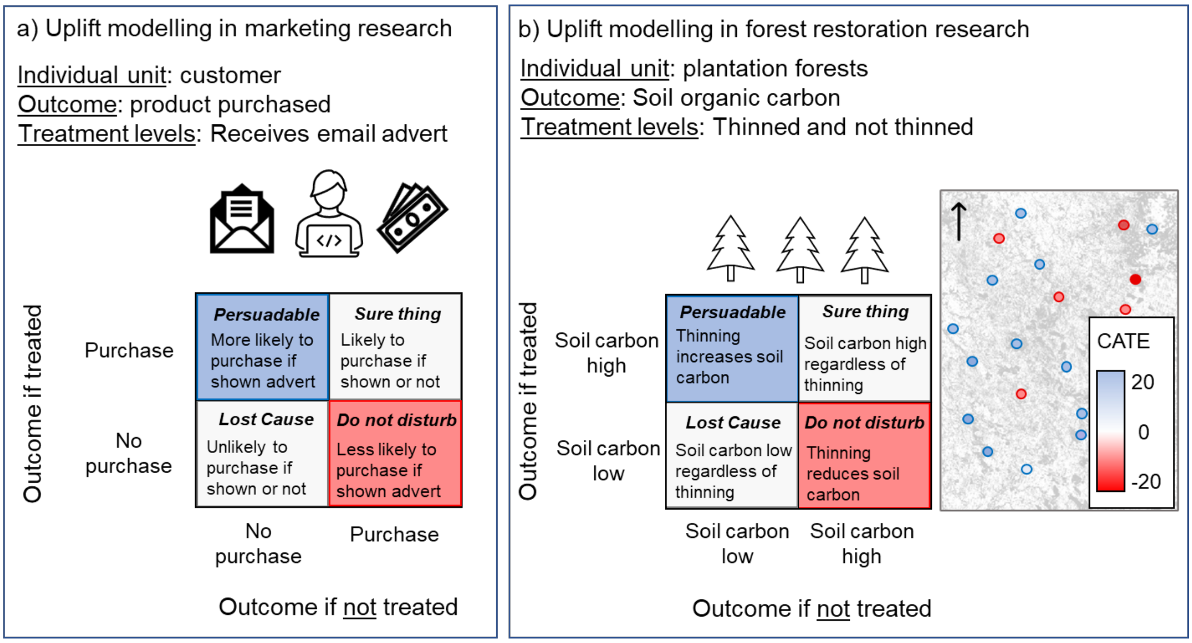


Figure 2. The principles of uplift modelling, used in marketing, apply also to ecology to identify currently unmanaged units that should or should not be targeted for a specific management intervention. a) In marketing, uplift modelling classifies individual customers according to differences in purchasing behaviour after being shown or not shown an advertisement. It is most cost effective to target advertisements at individual customers who are *Persuadable* - those individuals that would buy a product only if shown an advert. b) Uplift modelling applied to total soil carbon (tonnes/ha) in forest stands. Stands with lower soil carbon following thinning treatment should not be disturbed, while *Persuadable* stands that have greater soil carbon following treatment should be targeted. The map shows hypothetical individual treatment effects (difference in tonnes C/ha) predicted for untreated forest stands within the sample.

A roadmap to establishing prospects for precision ecology

Several generic issues remain outstanding for widening applications of ITE prediction within medicine and marketing and across other disciplines. These relate particularly to risks of biases in sample selection and confounding of cause with correlation. As for all statistical analysis, they require a causal understanding of mechanisms underpinning heterogeneity in treatment effects[46]. For ITE prediction, the covariates that matter, and therefore the ones to test with sufficient power to detect effects, are those that contribute most to explaining variation in treatment effects[47]. The problem of how to identify them *a priori* may be best addressed again through theoretical understanding of mechanisms, as well as evidence from past empirical research. It is well established that causal models should not condition on 'post-treatment variables that vary as a function of treatment (see[48]). Other general challenges include expansion to nonbinary treatments, which is an ongoing enterprise in precision medicine.

Ecology-specific issues for ITE prediction, of defining the observational unit for ITE, and risks of carryover and spillover effects amongst sample plots, have only recently been explored in other precision disciplines[49,50], and they remain as hurdles to the adoption of ITE prediction for applications to precision ecology. All are well-known issues of spatial analysis for conventional ATE estimation, however, and the same principles of data-collection design will apply to ITE. Where sampling units such as lakes or forest stands vary in size, the spatial extent of the sampling unit may need to be included in the predictive model as a conditioning covariate. A key ITE-specific challenge is the potential for heterogeneous treatment implementation. For example, for studies quantifying effects of field-margin presence on pollinator abundance, field-margin size, quality and maintenance might vary among fields, and might be systematically confounded with environmental covariates (e.g., farm economic size), yielding biased ITEs. Implementation heterogeneity is also problematic in medical drug trials, for example when patients do not comply with treatment recommendations.

For all of these general and ecology-specific challenges, simulations offer opportunities to explore costs and benefits of alternative sampling strategies, within an idealised environment[33,36,51]. Synthetic datasets with real covariates, for which both potential outcomes are simulated and hence known, are ideal for investigating how ITE prediction accuracy is influenced by sampling and modelling parameters[51]. Importantly, simulation studies have found that meta-learners vary in their predictive accuracy under different data conditions, and no one learner works uniformly best.[36] In the same way, ecologists can adopt a 'virtual ecologist' approach[52] as follows. First, ecologists can use process-based models to generate spatially-explicit virtual representations of ecosystems and their assignment to treatments across heterogeneous landscapes. A range of process-based models are increasingly being used to model how species and ecosystems are likely to respond to environmental changes and to potential management options[53–55]. Crucially, both potential outcomes of a treatment variable of interest can be simulated, to overcome the fundamental problem of causal inference and generate known ITEs for each virtually sampleable unit. Second, such virtual landscapes can be sampled using designs common to ecology (e.g., a citizen-science project or national monitoring programme) to generate datasets. Third, the resulting datasets can be subjected to different modelling approaches (e.g., meta-learner algorithms) to predict unit-specific treatment effects. Fourth, researchers can measure and compare the accuracy, precision and utility of ATE and ITE predictions by evaluating predictions against 'true' (i.e., simulated) treatment effects.

Within this virtual ecologist approach, researchers can compare performances of the different meta-learners for different sampling processes, and they can systematically vary response detectability by environmental variables (e.g., of birds in shrubland vs. open grass[56]). The unique advantage of simulation, which no empirical approach can better, is that the researcher has full knowledge of the sampled population, against which to quantify sampling biases in estimation or prediction by statistical analysis[52]. The hurdles to simulation principally concern the level of achievable realism in representing empirical landscapes or sampling processes.

We identify the following four key questions that will determine the future prospects for ITE prediction in applied ecological research.

1. What types of study question can precision ecology answer?

Precision ecology applies to any question about changes in a parameter of interest across a suite of heterogeneous covariate influences, be they observed across time or space, and it provides context-dependent answers. ITE prediction of intervention effects attempts to answer the what, how, where and when questions asked by conservation managers and policy makers, which conventional ATE estimation cannot adequately address for anything but homogeneous environments. Stakeholders in management- or policy-relevant outcomes should be involved in setting research questions.

As with ATE estimation, ITE prediction requires observations from individual units for treatments, outcomes, and covariates that influence treatment assignments or outcomes. ITE prediction, however, requires more parameters than ATE. For example, X- and R-learners must learn ‘nuisance functions’ including those that estimate propensity scores and intermediate treatment effects (Box 1). ITE prediction is correspondingly more data-hungry, potentially demanding larger training datasets and more covariates, although two-model meta-learners (e.g., T- and X-learners) reduce training sample sizes to their treatment groups. A growing number of high-resolution and large-scale datasets on biodiversity and ecosystem functioning are becoming available from multiple sources, including *in-situ* and remote sensors, eDNA, citizen scientists and monitoring networks. The rising availability and integration of data products will increasingly facilitate the parameterisation of ITE prediction models, by supplying conditional covariates related to topography, climate, and historical land uses. In addition, remote-sensing technology has expanding applications to data on treatment effects. For example, satellite data can be used to characterise the incidence of forest loss through clearance, fire, pest outbreaks etc, as well as outcomes related to forest biodiversity, productivity and condition[57].

2. What design of sampling strategy will yield unbiased individual treatment effects?

Data-collection designs that plan for ITE prediction at the outset will depend on the exact question and target population. It is important to note that sampling methods that optimise statistical power for the ATE do not necessarily optimise predictive validity of the ITEs of interest – especially when the interest is in ITE prediction for units with relatively unique covariate profiles[58]. Sampling designs for causal prediction should be informed by a ‘theory of treatment effect heterogeneity’[47], wherein researchers identify important subgroups and hypotheses regarding moderating covariates *a priori*, rather than *post hoc*. Hypothetically important moderators could then be used to stratify populations for sampling, and could aid in identifying how existing monitoring programmes might need strategic augmenting to maximise ITE prediction[47].

3. What biases and caveats are introduced by analysing pre-existing data?

The rapidly expanding range of big data available from monitoring programmes, citizen science initiatives and remote and near-Earth sensing, which typically cover heterogeneous landscapes of covariates, open up new opportunities for ITE prediction. Such datasets, however, are inherently susceptible to different forms of sample-selection bias. Patterns in the availability of biodiversity data, for example, are affected by the original motivations for, and constraints on, data-collection and reporting[56]. The meta-learner algorithms introduced in Box 1 offer some of the most promising available approaches to smoothing over the resulting gaps in observed data that raise vulnerabilities to sample-selection bias or underpowered analysis. The challenge is to decide which algorithm and type of model works best for which context, and the virtual ecologist has an important role to play here[52]. A model might be vulnerable to overfitting (mistakenly fitting sample-specific noise as if it were signal[59]), or an algorithm might overperform on simulated data due to specificities of the data-generating process, requiring comparison of ITE accuracy from the crossing of alternative data-generating mechanisms[51], sampling designs and model decisions. Knowledge of how sampling and modelling characteristics influence predictive accuracy will arm ecologists with an informed understanding of when precision approaches can and cannot work under what conditions, their underpinning assumptions, and when to anticipate biases.

Although spatial data now exist that permit quantification of spatial composition, pattern, and position of landscape features from local to continental scales,[60] researchers must consider whether they capture habitat features most relevant to the effects of interest[54]. Given the measurement error inherent to many ecological covariates and outcomes, for example depending on species' detection probabilities and accuracy of land cover maps, precision analyses will need to develop methods for translating measurement error into ITE prediction uncertainty.

4. How can precision outputs best inform managers and policy makers?

ITEs offer intuitively appealing predictions with directly actionable site-level predictions. Applied ecologists have scope to develop innovative and imaginative visualisations[61], maximising the actionability of predictions for diverse audiences of policymakers, practitioners, and other scientists. Mapping uncertainties in the treatment effect, or in covariate measurements or their interactions, could usefully identify future data-collection needs, or sites that might warrant alternative interventions such as adaptive management. Stakeholders should be involved in co-design of tools for visualising statistical outputs.

Conclusion

Medicine and marketing have pioneered unit-specific prediction, motivated by the potentially catastrophic consequences of ineffectual decisions from ATEs for a medical patient's survival and for a production company's returns on investment. As climate change and biodiversity loss drive applied ecology towards crisis management, national-scale policies on nature also need to be actionable at local scales. Ecology needs to explore the full scope and limits of its potential for thinking globally, acting locally[1]. Applied ecologists must capitalise wisely on the deluge of open-source data, to inform local-level management decisions about global-scale problems, armed with the knowledge of which approaches can and cannot work under what conditions, the necessary assumptions that must hold, and when biases will most likely be encountered. The process of predicting unit-specific treatment effects forces the analyst to evaluate numerous assumptions about the validity of predicting treatment effects at specific sites, and in turn, it marries internal with external validity. In so doing, it contributes to narrowing the science-policy gap. Moving on from estimating sample-average effects to predicting unit-specific treatment effects will be key to effectively supporting restoration and management of Earth's species, habitats, and ecosystems.

Acknowledgements

R.S. and E.E.J. were funded by Natural Environment Research Council Project ‘Transferable Ecology for a changing world (TREE)’ (NERC reference: NE/X009998/1). E.G. acknowledges funding from a NERC Research Programme Fellowship provided through the UKRI Landscape Decisions Programme (NE/V007831/1). J.M.B was funded by NERC-funded project ‘Trustworthy and Accountable Decision-Support Frameworks for Biodiversity - A Virtual Labs based Approach’ (NE/X002233/1). R.S. thanks I.J. Dahabreh and L. Evans for informative discussion of concepts.

Contributions

R.S. conceived the idea and developed a first draft with C.P.D. Contributions from E.E.J., J.M.B., E.G., E.T., M.G. substantially developed ideas and content in manuscript iterations.

Competing interests

The authors declare no competing interests.

References

1. IPBES. Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. doi.org/10.5281/zenodo.3831673. (2019).
2. Vira, B. & Adams, W. M. Ecosystem services and conservation strategy: beware the silver bullet. Conservation Letters **2**, 158–162 (2009).
3. Frischmann, B. M. Two enduring lessons from Elinor Ostrom. Journal of Institutional Economics **9**, 387–406 (2013).
4. Spake, R. et al. An analytical framework for spatially targeted management of natural capital. Nat Sustain **2**, 90–97 (2019).
5. Bateman, I. J. et al. A review of planting principles to identify the right place for the right tree for 'net zero plus' woodlands: Applying a place-based natural capital framework for sustainable, efficient and equitable (SEE) decisions. People and Nature **5**, 271–301 (2023).
6. Matthews, K. B. et al. Not seeing the carbon for the trees? Why area-based targets for establishing new woodlands can limit or underplay their climate change mitigation benefits. Land Use Policy **97**, 104690 (2020).
7. Winqvist, C., Ahnström, J. & Bengtsson, J. Effects of organic farming on biodiversity and ecosystem services: taking landscape complexity into account. Annals of the New York Academy of Sciences **1249**, 191–203 (2012).
8. Hong, P. et al. Biodiversity promotes ecosystem functioning despite environmental change. Ecology Letters **25**, 555–569 (2022).
9. Tipton, E. Beyond Generalization of the ATE: Designing Randomized Trials to Understand Treatment Effect Heterogeneity. Journal of the Royal Statistical Society Series A: Statistics in Society **184**, 504–521 (2021).
10. Public Health England. A Brief Introduction to Realist Evaluation. (2021).
11. Shackelford, G. E. et al. Dynamic meta-analysis: a method of using global evidence for local decision making. BMC Biol **19**, 33 (2021).

12. Lamont, A. et al. Identification of predicted individual treatment effects in randomized clinical trials. Stat Methods Med Res **27**, 142–157 (2018).
13. Curth, A., Peck, R. W., McKinney, E., Weatherall, J. & van der Schaar, M. Using Machine Learning to Individualize Treatment Effect Estimation: Challenges and Opportunities. Clinical Pharmacology & Therapeutics **115**, 710–719 (2024).
14. Zhang, W., Li, J. & Liu, L. A Unified Survey of Treatment Effect Heterogeneity Modelling and Uplift Modelling. ACM Comput. Surv. **54**, 1–36 (2022).
15. Seidl, R. et al. Forest disturbances under climate change. Nature Clim Change **7**, 395–402 (2017).
16. Alaa, A. M. Bayesian Inference of Individualized Treatment Effects using Multi-task Gaussian Processes.
17. Baier, D. & Stöcker, B. Profit uplift modeling for direct marketing campaigns: approaches and applications for online shops. J Bus Econ **92**, 645–673 (2022).
18. Hillstrom, K. The minethatdata e-mail analytics and data mining challenge. MineThatData https://blog.minethatdata.com/2008/03/minethatdata-e-mail-analytics-and-data.html (2008).
19. Holland, P. W. Statistics and Causal Inference. Journal of the American Statistical Association **81**, 945–960 (1986).
20. Fisher, R. The Design of Experiments. (Macmillan).
21. Blair, G. et al. Community policing does not build citizen trust in police or reduce crime in the Global South. Science **374**, eabd3446 (2021).
22. Pynegar, E. L., Gibbons, J. M., Asquith, N. M. & Jones, J. P. G. What role should randomized control trials play in providing the evidence base for conservation? Oryx **55**, 235–244 (2021).
23. Ruberg, S. J., Chen, L. & Wang, Y. The mean does not mean as much anymore: finding sub-groups for tailored therapeutics. Clin Trials **7**, 574–583 (2010).
24. Salditt, M., Eckes, T. & Nestler, S. A Tutorial Introduction to Heterogeneous Treatment Effect Estimation with Meta-learners | Administration and Policy in Mental Health and Mental Health Services Research. Administration and Policy in Mental Health and Mental Health Services Research **51**, 650–673 (2024).

25. Curth, A. Nonparametric Estimation of Heterogeneous Treatment Effects: From Theory to Learning Algorithms.
26. Wager, S. & Athey, S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. Journal of the American Statistical Association **113**, 1228–1242 (2018).
27. Bica, I. & Jordon, J. Estimating the Effects of Continuous-valued Interventions using Generative Adversarial Networks.
28. Kreif, N., Grieve, R., Díaz, I. & Harrison, D. Evaluation of the Effect of a Continuous Treatment: A Machine Learning Approach with an Application to Treatment for Traumatic Brain Injury. Health Economics **24**, 1213–1228 (2015).
29. Larsen, A. E., Meng, K. & Kendall, B. E. Causal analysis in control–impact ecological studies with observational data. Methods Ecol Evol **10**, 924–934 (2019).
30. Kimmel, K., Dee, L. E., Avolio, M. L. & Ferraro, P. J. Causal assumptions and causal inference in ecological experiments. Trends in Ecology & Evolution **36**, 1141–1152 (2021).
31. Shalit, U., Johansson, F. D. & Sontag, D. Estimating individual treatment effect: generalization bounds and algorithms. Preprint at http://arxiv.org/abs/1606.03976 (2017).
32. Lu, M., Sadiq, S., Feaster, D. J. & Ishwaran, H. Estimating Individual Treatment Effect in Observational Data Using Random Forest Methods. Journal of Computational and Graphical Statistics **27**, 209–219 (2018).
33. Künzel, S. R., Sekhon, J. S., Bickel, P. J. & Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. Proc. Natl. Acad. Sci. U.S.A. **116**, 4156–4165 (2019).
34. Caron, A., Baio, G. & Manolopoulou, I. Estimating Individual Treatment Effects using Non-Parametric Regression Models: a Review. Journal of the Royal Statistical Society Series A: Statistics in Society **185**, 1115–1149 (2022).
35. Dorie, V., Hill, J., Shalit, U., Scott, M. & Cervone, D. Automated versus Do-It-Yourself Methods for Causal Inference. (2024).
36. Okasa, G. Meta-Learners for Estimation of Causal Effects: Finite Sample Cross-Fit Performance.
37. Alaa, A. M. Limits of Estimating Heterogeneous Treatment Effects: Guidelines for Practical Algorithm Design.

38. Fernandez-Lorıa, C. & Provost, F. Causal Classification: Treatment Effect Estimation vs. Outcome Prediction.

39. Nie, X. & Wager, S. Quasi-oracle estimation of heterogeneous treatment effects. Biometrika **108**, 299–319 (2021).

40. Foster, J. C., Taylor, J. M. G. & Ruberg, S. J. Subgroup identification from randomized clinical trial data. Statistics in Medicine **30**, 2867–2880 (2011).

41. Hill, J. L. Bayesian Nonparametric Modeling for Causal Inference. Journal of Computational and Graphical Statistics **20**, 217–240 (2011).

42. Athey, S. & Imbens, G. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences **113**, 7353–7360 (2016).

43. Some methods for heterogeneous treatment effect estimation in high dimensions - Powers - 2018 - Statistics in Medicine - Wiley Online Library. https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.7623?casa_token=ai3GV8ZWNzwAAAAA:Cw2If9t5Mjf2hjOfjn3WHc2eB4rvbMiCoajVTWL929PkHAYlfTmNMshgbo_l6XsTq6d2rRZWK4A.

44. Robinson, P. M. Root-N-Consistent Semiparametric Regression. Econometrica **56**, 931–954 (1988).

45. Kennedy, E. H. Towards optimal doubly robust estimation of heterogeneous causal effects. Electronic Journal of Statistics **17**, 3008–3049 (2023).

46. McElreath, R. Statistical Rethinking: A Bayesian Course with Examples in R and STAN. (2020).

47. Tipton, E., Yeager, D. S., Iachan, R. & Schneider, B. Designing Probability Samples to Study Treatment Effect Heterogeneity. in Experimental Methods in Survey Research (eds. Lavrakas, P. et al.) 435–456 (Wiley, 2019). doi:10.1002/9781119083771.ch22.

48. Montgomery, J. M., Nyhan, B. & Torres, M. Replication Data for: How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It. 15369, 26998, 196880, 53422, 49642, 52375, 20125, 3243, 21892, 486920, 47044, 59963, 48706 Harvard Dataverse https://doi.org/10.7910/DVN/EZSJ1S (2018).

49. Zhang, Y. & Imai, K. Individualized Policy Evaluation and Learning under Clustered Network Interference. Preprint at https://doi.org/10.48550/arXiv.2311.02467 (2024).

50. Viviano, D. Policy Targeting under Network Interference. The Review of Economic Studies **92**, 1257–1292 (2025).

51. Curth, A., Svensson, D. & Weatherall, J. Really Doing Great at Estimating CATE? A Critical Look at ML Benchmarking Practices in Treatment Effect Estimation.

52. Zurell, D. et al. The virtual ecologist approach: simulating data and observers. Oikos **119**, 622–635 (2010).

53. Bocedi, G. et al. RangeShifter 2.0: an extended and enhanced platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. Ecography **44**, 1453–1462 (2021).

54. Gardner, E. et al. A family of process-based models to simulate landscape use by multiple taxa. Landsc Ecol **39**, 102 (2024).

55. Tree regeneration in models of forest dynamics: A key priority for further research - Díaz-Yáñez - 2024 - Ecosphere - Wiley Online Library. https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecs2.4807.

56. Bowler, D. E. et al. Treating gaps and biases in biodiversity data as a missing data problem. Biological Reviews **n/a**,.

57. Massey, R., Berner, L. T., Foster, A. C., Goetz, S. J. & Vepakomma, U. Remote Sensing Tools for Monitoring Forests and Tracking Their Dynamics. in Boreal Forests in the Face of Climate Change: Sustainable Management (eds. Girona, M. M., Morin, H., Gauthier, S. & Bergeron, Y.) 637–655 (Springer International Publishing, Cham, 2023). doi:10.1007/978-3-031-15988-6_26.

58. Tipton, E. & Hartman, E. Generalizability and Transportability. in Handbook of Matching and Weighting Adjustments for Causal Inference (Chapman and Hall/CRC, 2023).

59. Yarkoni, T. & Westfall, J. Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect Psychol Sci **12**, 1100–1122 (2017).

60. Jones, K. B. et al. Informing landscape planning and design for sustaining ecosystem services from existing spatial patterns and knowledge. Landscape Ecology **28**, 1175–1192 (2013).

61. Hullman, J. & Diakopoulos, N. Visualization Rhetoric: Framing Effects in Narrative Visualization. IEEE Trans. Visual. Comput. Graphics **17**, 2231–2240 (2011).