Ecology Letters

ECOLOGY LETTERS

SYNTHESIS

Foundations and Future Directions for Causal Inference in Ecological Research

Katherine Siegel[1,2] | Laura E. Dee[3]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado-Boulder, Boulder, Colorado, USA | [2]Department of Geography, University of Colorado-Boulder, Boulder, Colorado, USA | [3]Department of Ecology & Evolutionary Biology, University of Colorado-Boulder, Boulder, Colorado, USA

Correspondence: Katherine Siegel (katherine.j.siegel@colorado.edu)

Received: 24 July 2024 | **Revised:** 6 December 2024 | **Accepted:** 10 December 2024

Editor: Jonathan M. Chase

Funding: This work was supported by National Science Foundation Faculty Early Career Development Program (CAREER #2340606), NASA BioSCape (#80NSSC 22K0796), Cooperative Programs for the Advancement of Earth System Science, National Oceanic and Atmospheric Administration (N).

Keywords: big data | causal analysis | counterfactual | observational data | potential outcomes framework | statistical ecology | structural causal model | study design | synthesis science

ABSTRACT

Ecology often seeks to answer causal questions, and while ecologists have a rich history of experimental approaches, novel observational data streams and the need to apply insights across naturally occurring conditions pose opportunities and challenges. Other fields have developed causal inference approaches that can enhance and expand our ability to answer ecological causal questions using observational or experimental data. However, the lack of comprehensive resources applying causal inference to ecological settings and jargon from multiple disciplines creates barriers. We introduce approaches for causal inference, discussing the main frameworks for counterfactual causal inference, how causal inference differs from other research aims and key challenges; the application of causal inference in experimental and quasi-experimental study designs; appropriate interpretation of the results of causal inference approaches given their assumptions and biases; foundational papers; and the data requirements and trade-offs between internal and external validity posed by different designs. We highlight that these designs generally prioritise internal validity over generalisability. Finally, we identify opportunities and considerations for ecologists to further integrate causal inference with synthesis science and meta-analysis and expand the spatiotemporal scales at which causal inference is possible. We advocate for ecology as a field to collectively define best practices for causal inference.

1 | Introduction

Questions about causal relationships are common in ecology: we seek to understand the effect of biodiversity on ecosystem functioning (Tilman et al. 2001; Tilman, Isbell, and Cowles 2014), the impacts of climate change and disturbance regimes on ecosystems (García Criado et al. 2020; Halofsky, Peterson, and Harvey 2020), the effects of anthropogenic activities on animal behaviour (Gaynor et al. 2018), the effects of different abiotic variables on plant productivity across ecosystem types (Smith et al. 2024) and the effectiveness of restoration and conservation (Geldmann et al. 2019; Suding 2011). These are fundamentally causal questions: they seek to isolate and estimate the effect of a causal variable on an outcome (Box 1) and rule out alternative explanations for the estimated effects (Table 1).

Identifying and quantifying causal relationships, however, pose challenges in complex ecological systems. Many factors impact an outcome of interest, and confounding variables—which affect both the causal variable and the outcome—can bias estimates of

BOX 1 | Key terms in causal inference.

As different disciplines have contributed to the development of causal inference, the field has accumulated a dizzying array of jargon. These specialised terms pose barriers to ecologists seeking to engage with the literature. We provide definitions for some key terms, with an extended glossary in Appendix S1.

- *Average treatment effect (ATE)*: the average difference in the outcome variables between the treated and control populations (Figure 1).
- *Bias*: the difference between the estimated effect and the true value of the effect.
- *Collider*: a variable that is affected by both the treatment and the outcome. Conditioning on a collider can lead to incorrect estimates of the direction of the effect.
- *Complier*: a sample unit that received the treatment to which it was assigned: a unit that was assigned to the treated group and received the treatment, or a unit that was assigned to the untreated group and did not receive the treatment.
- *Conditioning*: an approach to isolating the effect of the treatment on the outcome of interest by considering the values of all other variables in a model given a certain value of the variable on which the model is conditioned. Also referred to as 'adjusting'.
- *Confounder*: a variable that affects both the treatment and the outcome. Failing to account for confounding variables biases estimates of the treatment effect.
- *Control*: the untreated units in an experiment or quasi-experiment.
- *Counterfactual*: well-defined alternative(s) to what we observe in the world.
- *Endogeneity*: correlation between the treatment variable and the error term, arising due to omitted confounding variables, reverse causality, simultaneity, or measurement error in the explanatory variable.
- *Estimand*: the effect of the treatment compared to the control for a specific population (e.g., average treatment effect, average treatment effect of the treated, local average treatment effect and conditional average treatment effect (Supporting Information)).
- *Estimator*: a statistical approach to estimating the value of a model parameter.
- *Exogeneity*: the condition in which the treatment variable is not correlated with or causally influenced by other model parameters.
- *Local average treatment effect (LATE)*: the treatment effect for units that were assigned to the treated group and did in fact receive the treatment, ignoring the effect of non-compliance.
- *Measurement error*: the difference between the true and recorded/observed value of a variable. Measurement error in the treatment variable biases the estimates, while measurement error in the outcome variable adds noise to the model without biasing the estimates.
- *Mediator*: a variable that lies on the causal pathway between the treatment and the outcome.
- *Moderator*: a variable that affects the magnitude of the causal effect, often implemented in statistical analyses and regression as an interaction term.
- *Omitted variable bias*: bias in estimates of the treatment effect that occurs when study designs do not account for confounding variables.
- *Panel data*: data collected for the same sample units over multiple time periods (i.e., longitudinal data).
- *Outcome*: the value of the response variable.
- *Quasi-experiments*: study designs that assess causal relationships in the absence of randomisation, using variation in units' exposure to treatment(s).
- *Random assignment*: an approach to treatment assignment in which all units have an equal probability of receiving the treatment, regardless of underlying characteristics. Randomisation ensures that there are no systematic differences between the treated and control units, allowing for an unbiased estimation of the treatment effect.
- *Reverse causality*: the outcome variable affects the treatment, rather than the treatment affecting the outcome.
- *Selection bias*: when the units that are exposed to the treatment are not randomly selected, there may be systematic differences between the treated and control samples, biasing the estimate of the treatment effect.
- *Simultaneity*: the treatment affects the outcome, and the outcome affects the treatment.
- *Stable unit treatment value assumption (SUTVA)*: the assumption that there is no interference in the system (the treatment status of one unit cannot influence the outcome of another unit) and that for each unit, there are not different versions of each treatment level or hidden variation in the treatment.
- *Treatment*: a potential manipulation by humans or nature. Causal inference focuses on treatments/causes where we could hypothetically imagine an ideal controlled experiment with randomised treatment assignment.

TABLE 1 | Matching distinct research aims with methods. Ecological studies that use observational data can have different aims, which require different methodological techniques. In addition to estimating causal relationships, we are often interested in description and prediction, or the ability to estimate outcomes outside of the observed data. Description, causal inference and prediction ask fundamentally different questions and require different methods (Hernán, Hsu, and Healy 2019). For instance, some methods ecologists use to assess the performance of their models are appropriate for predictive aims but not causal analysis (Addicott et al. 2022; Arif and MacNeil 2022a; Pichler and Hartig 2023). Here, we demonstrate the different data needs and methods required to answer descriptive, predictive and causal questions in ecology (table adapted from Hernán, Hsu, and Healy (2019) and Laubach et al. (2021)).

	Description	Prediction	Causal analysis
Urban ecology (Locke et al. 2021)			
Question	How does historical redlining relate to current patterns of tree canopy cover?	Can historical redlining predict current tree canopy cover?	What is the effect of historical redlining on current patterns of tree canopy cover?
Data	• Redlining polygons • Current tree cover	• Redlining polygons • Current tree cover	• Redlining polygons • Current tree cover • Current neighbourhood-level socioeconomic characteristics and zoning • Spatial data on tree-planting efforts
Methods	Summary statistics on tree cover in redlined vs. non-redlined neighbourhoods	Regression analysis predicting tree cover as a function of presence of historical redlining	Regression discontinuity design comparing tree cover at boundaries of redlined vs. non-redlined neighbourhoods
Invasion ecology (Knapp and Matthews 2000)			
Question	What are the population trends of introduced fish species and endemic amphibians in alpine lakes?	Which lakes are likely to provide suitable habitat for both introduced fish and endemic amphibians?	Does the increase in populations of introduced fish species cause a decline in endemic amphibian populations in alpine lakes?
Data	• Population of introduced fish species over time • Population of endemic amphibians over time	• Presence/absence or abundance data for introduced fish species • Presence/absence or abundance data for endemic amphibians • Lake-level data on nutrient levels, elevation, surface area, maximum depth, substrate composition, solar radiation input and isolation from other lakes	• Population of introduced fish species over time • Population of endemic amphibians over time • Lake-level data on nutrient levels, elevation, surface area, maximum depth, substrate composition, solar radiation input and isolation from other lakes
Methods	Summary trends over time for both taxa	Species distribution models	Difference-in-differences comparing population trends in lakes with and without introduced fish species, before and after their introduction
Protected areas (Xu et al. 2022)			
Question	Do protected forests have different land surface temperatures than unprotected forests?	How is climate change likely to change land surface temperature in protected and unprotected forests?	Do protected areas buffer against climate change impacts on land surface temperature?

(Continues)

TABLE 1 | (Continued)

	Description	Prediction	Causal analysis
Data	• Protected area polygons • Land cover maps • Land surface temperature data	• Protected area polygons • Land cover maps • Downscaled climate projections	• Protected area polygons • Land cover maps • Land surface temperature data • Site-level data on elevation, topographic roughness, distance to roads, distance to cities and forest type
Methods	Compare the average land surface temperatures of protected and unprotected forests	Use process-based models to project land surface temperature in forests under different climate change scenarios	Matching protected and unprotected forests, then regression analysis to estimate the effect of protection on land surface temperature

causal relationships. For example, precipitation is a confounding variable when estimating the effect of plant species richness on grassland productivity, by affecting both richness and productivity (Dee et al. 2023). Failure to account for precipitation in our model would lead to incorrect conclusions about the significance, magnitude and/or direction of the effect of species richness on productivity. Confounding variables occur frequently in ecological systems: as researchers, we may be aware of and able to measure some but not all of them (e.g., we may lack data on some confounding variables, or our model may be misspecified, causing us to omit a confounder). This creates challenges for understanding causal relationships in ecology.

To answer causal questions, ecologists have traditionally used randomised experiments or pseudo-experiments (Christie et al. 2019). However, many ecological questions face logistical and ethical challenges to experimentation, such as the inability to replicate natural disturbances or ethical issues regarding the manipulation of endangered or non-native species. Furthermore, experiments can be imperfect and do not always meet the assumptions required for causal inference: unexpected, non-random processes may pose challenges for their causal interpretation (Arif and Massey 2023; Kimmel et al. 2021). Other fields facing similar barriers, including public health and economics, have extended the foundations underlying experimental design to develop frameworks for inferring causal relationships from observational data (Greenstone and Gayer 2009; Little and Rubin 2000). These frameworks include statistical approaches for overcoming the challenges posed by experimental and observational data, emphasising clear articulation of the assumptions required for causal interpretations of estimated effects (Hernán and Robins 2016). While the conservation impact evaluation field has embraced these approaches, particularly to assess the effectiveness of protected areas (Ferraro and Pattanayak 2006; Jones and Shreedhar 2024), causal inference approaches are less widely adopted in ecology. Encouragingly, recent reviews have provided introductions to causal inference geared towards ecologists (Butsic et al. 2017; Larsen, Meng, and Kendall 2019), and ecological studies have increasingly applied quasi-experimental approaches (Box 1; Dee et al. 2023; Ramsey et al. 2019; Wu et al. 2023) and used causal graphs (Arif and MacNeil 2023; Grace et al. 2016; Shipley 1999) in empirical settings.

These approaches to causal inference can improve our ability to investigate causal relationships using both experimental and observational data. Stronger integration of causal inference into ecology can enable new insights by (1) strengthening experimental design and clarifying the assumptions required for deriving causal inference from experiments (Kimmel et al. 2021) and (2) advancing rigorous assessment of causal relationships from observational data (Butsic et al. 2017; Larsen, Meng, and Kendall 2019). These approaches can enable ecologists to leverage novel data streams from remote sensing, long-term monitoring, or citizen/community science to test ecological theory in natural, non-experimental ecosystems (Dee et al. 2016, 2023; Larsen and Noack 2020) and ask ecological questions at management-relevant spatial and temporal scales at which randomised controlled experiments are not possible (Ratcliffe et al. 2022, 2024; Siegel et al. 2022a, 2022b; Simler-Williamson and Germino 2022). This integration has not yet reached its full potential, as applying these approaches appropriately requires an in-depth understanding of the assumptions, strengths and limitations of causal inference.

As ecologists, we face significant jargon and disciplinary barriers to adopting causal inference approaches, despite the recent proliferation of applications to ecology and open-source software tools. Quasi-experimental approaches to causal inference are not part of most graduate curricula in ecology, and experimental design courses may not equip students with tools to interpret their results when the assumptions underlying randomised experiments are violated. Exploring causal inference using texts from multiple other disciplines (e.g., Angrist and Pischke 2008, 2015; Cunningham 2021), ecologists may struggle to find intuitive, applicable examples. Different fields' jargon also creates obstacles (Box 1 provides a glossary). For example, other fields use 'panel data' to describe what an ecologist might call 'longitudinal data' and 'fixed effects' has a different—nearly opposite—meaning in ecology than in econometrics (Byrnes and Dee 2024). These barriers raise the risk of misusing methods and missed opportunities to advance basic and applied ecology. The growth of machine learning highlights the urgency of clarifying best practices in the field of causal inference, as these popular methods may not be the best approach to answering causal questions (Pichler and Hartig 2023).

To help ecologists overcome these barriers, we provide an accessible translation of causal inference study designs by building intuition around the assumptions, strengths and limitations of different approaches. We present the underlying frameworks of causal inference; the assumptions upon which causal inference—both from experimental and observational approaches—rest and how our interpretation of 'arguably causal' results should reflect the assumptions underlying the approaches we use; and applications to ecological research. We highlight that studies are not simply 'causal' or 'not causal': there is a spectrum based on the strength of assumptions given the study design, data context and research question (Kimmel et al. 2021). Throughout, we introduce readers to foundational texts. For additional, self-guided study, we provide a curated reading list and reproducible demonstrations of individual causal inference approaches (Supporting Information), drawing on our experiences teaching a graduate-level causal inference course for ecologists (Box 2). Building from previous introductions (e.g., Arif and MacNeil (2022b), Butsic et al. (2017), Fick et al. (2021), Grace (2021) and Ramsey et al. (2019)), we discuss how strengths of causal inference approaches in terms of reducing bias—internal validity—can be weaknesses in terms of generalisability and emphasise that these approaches require substantial amounts of data to detect effects. To increase generalisability, we discuss potential integrations of causal inference with synthesis science and meta-analysis and highlight how the use of new data streams (e.g., from remote sensing) can increase both the scale of inference and sample sizes for causal inference. We end with a forward-looking view for the field to collectively define best practices for causal inference in ecology.

2 | Causal Inference Frameworks

Causal analysis—including experiments and quasi-experiments—must contend with the fundamental problem that we can only observe one state of the world (Hernan 2004). We cannot directly observe how a change (e.g., a treatment, exposure, or altered condition) affects the same individual unit (e.g., person, plant, place) under both treatment and control conditions simultaneously (Holland 1986). In other words, we cannot directly observe the counterfactual: if a given unit received the treatment, we cannot observe the alternative scenario in which that same unit did not (Box 1). To address this, two complementary frameworks for causal inference have emerged: the potential outcomes (PO) framework (Rubin 1972) and the structural causal model (SCM) (Pearl 2009). In both, and throughout this paper, we define a treatment as a potential manipulation or 'intervention' by humans or nature. Treatments can be binary (e.g., species presence/absence), categorical (e.g., ecosystem type), or continuous (e.g., precipitation levels). Treatments may be the result of active manipulation by humans (e.g., species introductions) or nature (e.g., beavers' transformation of hydrology) or a characteristic of a system (e.g., edaphic gradients) (Holland 1986).

The PO framework defines a causal effect based on a set of potential outcomes that could be observed in alternative states of the world (Rubin 1972, 2005): the causal effect is the difference in potential outcomes across two states of nature (Figure 1). The unobserved potential outcomes are counterfactuals (Morgan and Winship 2014). Counterfactuals, or well-defined alternatives to the outcomes that we observe in the world, are central to causal inference (Ferraro 2009). Different approaches are used to construct a counterfactual, all of which—including experiments, where control groups are often the counterfactual—require assumptions (Kimmel et al. 2021).

The other dominant causal inference framework is the SCM (Pearl 2009, 2010), which is related and complementary to the PO framework (Malinsky, Shpitser, and Richardson 2019; Pearl 2009; Richardson and Robins 2013). The SCM framework combines counterfactual causality from PO with graphical model approaches (Spirtes, Glymour, and Scheines 2001), generalising structural models more common in ecology, with roots in path analysis (Wright 1921). Recent reviews introduce the SCM to ecologists (Arif and MacNeil 2023; Laubach et al. 2021). Briefly, the SCM uses directed acyclic graphs (DAGs) to quantify the effects of interventions (Pearl 2009). Drawing on domain knowledge, previous research and ecological theory, DAGs are causal diagrams that map causal relationships among variables as directional arrows or paths in a graph (Figure 2). DAGs make transparent our assumptions about the relationships in our study system (Pearl 2009). DAGs include all known potential confounding variables (Box 1; Arif and MacNeil 2023)—whether or not they are observed in our data—and can clarify variables that fall on the causal path (mediators) or that create other sources of bias (e.g., colliders, Box 1) (Figure 2). DAGs thus provide a useful starting point for clarifying and articulating assumptions about causal relationships based on prior knowledge (Figure 3a) and for thinking through the spatial and temporal scales of the dynamics and variables of interest. We recommend drawing a DAG before performing an analysis and ideally before data collection. Arif and MacNeil (2023) provide guidance for ecologists on developing a DAG and testing its consistency with the underlying data, including R code.

BOX 2 | Teaching causal inference.

Formal coursework can increase ecologists' understanding of causal inference. To contribute to the development of causal inference curricula for ecologists, we developed and taught a graduate-level course on causal inference for ecology in the spring of 2023 in the Department of Ecology & Evolutionary Biology at the University of Colorado-Boulder, USA. The course attracted participants from diverse fields, including PhD students and postdoctoral scholars in ecology, evolutionary biology, microbiology, geography and environmental studies, as well as project scientists from academic research groups and a government agency. More than 90% of the course participants are now integrating causal inference methods into their dissertations, as side projects, or in their work in government agencies.

Course participants had different levels of statistical training, ranging from undergraduate-level statistics to extensive previous coursework in graduate-level biostatistics and econometrics. There was a similar diversity in experience and comfort with programming in R, the software language the course used. To meet the needs of this student body, we emphasised developing an intuitive understanding of the methods we taught, rather than stressing the underlying mathematics. For those with more technical training in statistics, we also provided key references for deeper dives into the math underlying these methods.

Our overall objectives were for students to gain an understanding of the main frameworks for counterfactual causal inference and how causal inference differs from other empirical research aims; familiarity with how causal inference is applied in experimental and quasi-experimental study designs; and experience reading the published literature with a critical eye towards appropriate use of methods for identifying causal relationships. Specifically, students learned to (a) summarise key threats to causal inference and identify these threats when evaluating study designs; (b) apply causal inference methods to real-world research questions and datasets; (c) identify the most appropriate study design(s) and methodology in non-experimental settings based on the available data and research question; (d) implement these designs and methods using R; and (e) appropriately interpret the results and their potential biases; and (f) communicate clearly about these methods, their results and their assumptions.

The course consisted of lectures introducing key topics and methods, demonstrations of how to implement quasi-experimental methods in R using simulated and real datasets, student-led discussions of publications that used different approaches and semester-long individual projects. Students demonstrated their understanding of the applications of different causal inference approaches, the underlying assumptions and the strengths and limitations of different methods through their projects: students identified a causal question, developed a DAG and revised it based on feedback, compiled the necessary data, conducted a preliminary analysis using a quasi-experimental method introduced in class and interpreted the results in the context of the method's underlying assumptions (Figure 3). The projects gave the students an opportunity to apply causal inference to their own research areas, with a focus on understanding the underlying intuition and learning the mechanics of applying causal inference to real-world problems. They also gained experience in providing feedback on each other's analyses, practicing skills required for peer review.

Students readily adopted DAGs, but many struggled to align their datasets with quasi-experimental designs. They often found matching and weighting to be the most intuitive approaches, even though these methods make the strongest assumptions. They also gravitated towards these methods due to data constraints (e.g., a single time period of data with no clear discontinuity or instrumental variable). Students were familiar with randomised experiments but not the approaches available when an experiment does not go to plan. Students' uncertainty in determining which quasi-experimental method best fit their research question and data motivated us to create Figure 3b.

We provide a curated reading list from our course (Appendix S2) as a resource for those interested in developing similar courses or using the reading list to structure their own self-guided learning. In our experience, a course on causal inference in ecology is useful for students familiar with ecological statistics and experimental design, but fundamental concepts of causal inference—such as underlying assumptions and issues with confounding—could be incorporated throughout research methods and study design curricula for ecologists.

In our experience, existing textbooks may not be well-suited as stand-alone texts for causal inference. Many textbook examples focus on binary treatments, while ecologists often encounter continuous or categorical treatments. This can create misconceptions about the applicability of quasi-experimental methods to ecological contexts (Box 3). Some textbooks pose issues for educators seeking to foster an inclusive and just classroom, as they may simplify complex social issues (e.g., positioning gender as a binary treatment).

3 | Challenges for Causal Inference

As previously introduced, the frequent occurrence of confounding variables makes causal analysis difficult in ecological systems. Confounders pose challenges for experimental and quasi-experimental approaches to causal inference: failure to account for confounding variables can bias estimates of the treatment effect (i.e., the estimated effect will differ from the true effect) because if confounders are omitted from the model, the model error will be correlated with the treatment (Figure 2). This phenomenon, where the treatment variable is correlated with the error term, is called endogeneity (conversely, if the treatment term is not correlated with the error term, then it is exogenous). Endogeneity can arise from other causes, like reverse or bidirectional causality or measurement error (Box 1) in the explanatory variable, but the challenge of confounding variables is especially pertinent in ecology. When confounding variables are not accounted for and thus cause bias in the estimator (Box 1), this is called omitted variable bias. Notably, omitted variable bias is an issue regardless of the sample size:

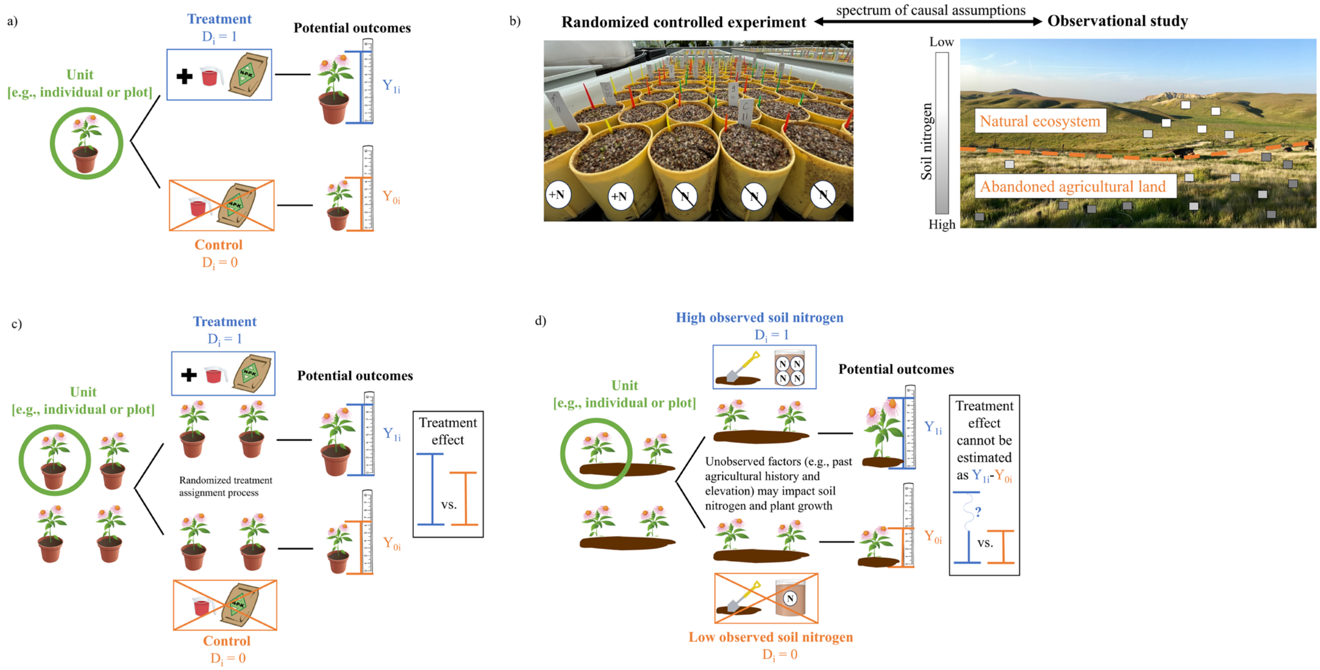

FIGURE 1 | The fundamental problem of causal inference poses a challenge for experimental and observational studies. (a) We cannot observe the outcomes of different treatment scenarios—receiving the treatment ($D_i = 1$) *and* not receiving the treatment ($D_i = 0$)—for a single unit (Splawa-Neyman 1923). In this example assessing nitrogen's effect on plant growth, because of the fundamental problem of causal inference, we can only observe the outcomes Y_{1i} when $D_i = 1$ and Y_{0i} when $D_i = 0$. The individual treatment effect is Y_{1i} -Y_{0i}, which is the causal effect of the treatment for unit *i*. (b) Different approaches to causal inference range in the strength of the assumptions they make to estimate causal effects, from randomised controlled experiments (which make the weakest assumptions) to purely observational studies (which make stronger assumptions). (c) Randomisation of treatment assignment ensures there is no systematic relationship between treatment assignment and underlying characteristics of the unit that could otherwise affect the outcome, allowing for estimation of the treatment (or causal) effect as the average difference between the outcomes for the different treatments. (d) Observational data, lacking randomisation, poses challenges for causal inference. In this example, the sample plots vary in their background characteristics (e.g., past land use, elevation), which affect soil nitrogen and plant growth, complicating our ability to estimate the potential outcomes. Icons from Saxby, Hawkey, and Anderson (2024). Photo credits: N. Emery and K. Siegel.

increasing the sample size does not reduce the bias in the estimate. Thus, confounding variables threaten causal inference. Note that we discuss regression-based approaches to estimating causal effects, but other approaches exist (Pearl 2010).

DAGs can help identify whether we have measured or unmeasured confounding variables (so-called 'back-door paths' that introduce endogeneity and lead to spurious correlations and bias) (Rohrer 2018). To satisfy Pearl's 'back-door criterion', we can use DAGs to identify which confounding variables to control for so that the effect of our causal variable of interest is conditionally independent (or *d*-separated) given this control (Arif and MacNeil 2022b; Pearl 2009). The back-door criterion must be completed for each pathway of interest to interpret the results causally.

As nonparametric causal graphs, DAGs encode our assumptions about causal relationships in a system to help guide the choice of variables to include or not when estimating their effects (e.g., in regression analyses). On their own, however, they do not quantify or estimate the magnitude of causal effects. For causal estimation, we next describe statistical designs for causal inference that fall along a spectrum from those that require the weakest assumptions for causal interpretation to approaches that require much stronger assumptions.

4 | Experimental Designs

The counterfactual model of causality described above was at the heart of Fisher's randomised controlled experiments (Fisher 1935). Randomised controlled experiments, or randomised control trials (RCTs), compare treated units to control units: control units serve as the counterfactual. Randomisation—or random treatment assignment—ensures that every unit has the same probability of receiving the treatment and therefore that there is no systematic relationship between the outcomes and observed or unobserved confounding variables (Figure 1b). Randomisation makes the treatment independent of confounders (Figure 1c), and the expected potential outcome for the control units is the same as the expected potential outcome for the entire population. Thus, random assignment makes two or more comparable groups. With perfect randomisation, groups should be identical on average prior to the treatment because every unit has an equal probability of being treated. In an ideal randomised controlled experiment, the effect of confounding variables is eliminated and the key assumptions of causal inference are met (Kimmel et al. 2021). Then, we can compare the differences-in-means of treatment groups to estimate an average treatment effect (Box 1) of the population in the experiment.

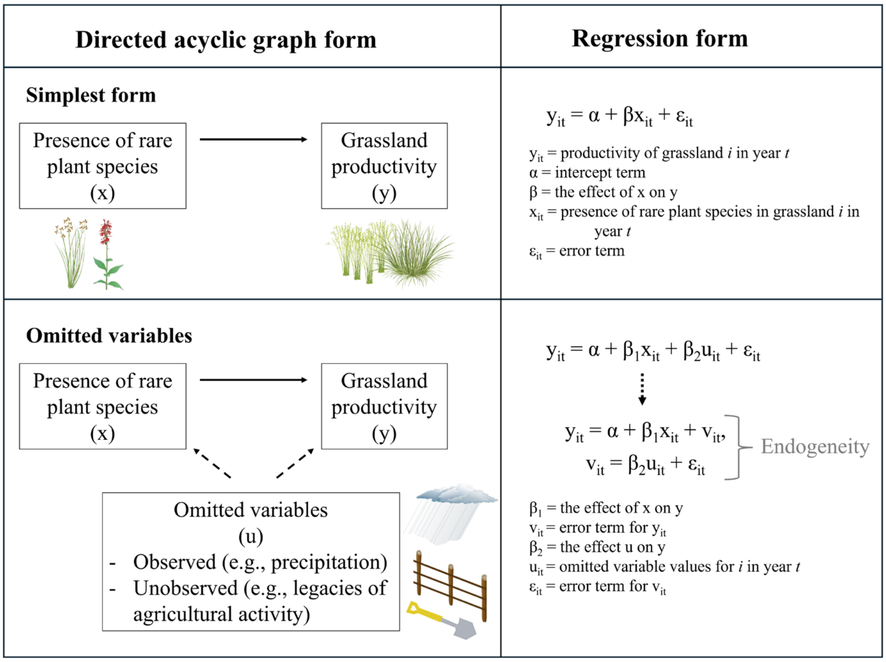

FIGURE 2 | When confounding variables are not accounted for, endogeneity occurs: The treatment term is correlated with the error term, yielding biased estimates of the treatment effect. We demonstrate this issue using directed acyclic graphs (DAGs) to visualise a hypothesised causal effect of the presence of rare plant species on grassland productivity. We show the regression equations corresponding with each DAG to demonstrate how omission of observed or unobserved confounding variables (e.g., precipitation, historical land use) leads to biased estimates of the treatment effect. We overcome challenges to endogeneity by conditioning on all confounding variables: This is equivalent to applying the back-door criterion (i.e., blocking all back-door paths). This can be challenging: All paths must be specified and correct, and all confounding variables must be controlled for and measured without error (Huntington-Klein 2022). Icons from Saxby, Hawkey, and Anderson (2024).

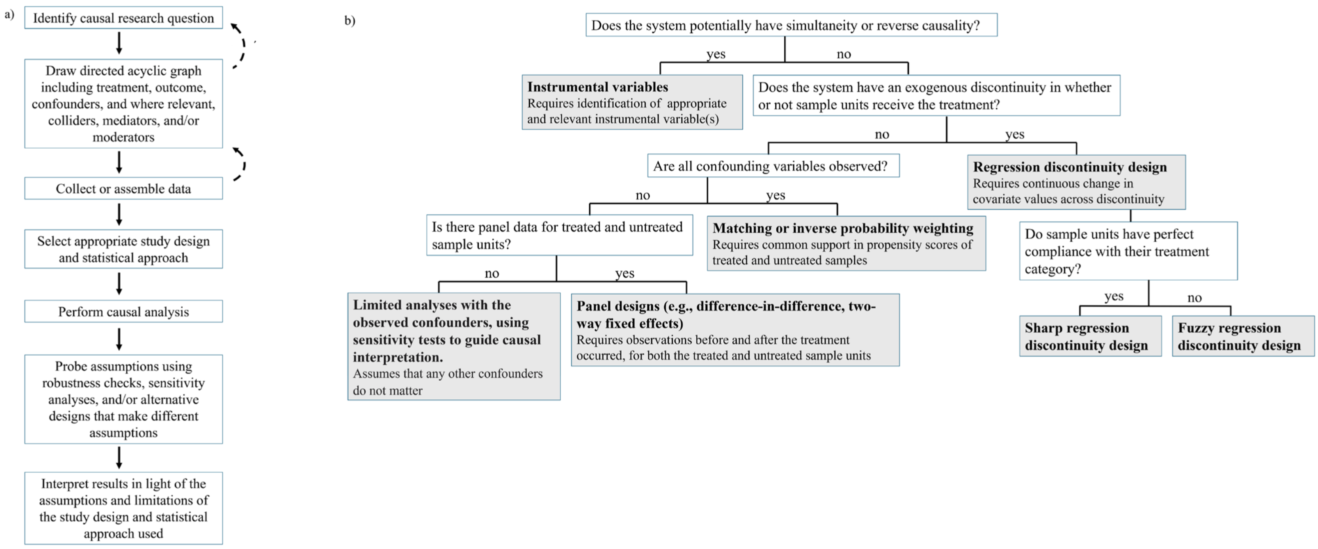

FIGURE 3 | (a) A workflow for causal inference in ecology. Dashed arrows indicate steps that may require iteration. For example, the process of drawing a DAG may lead us to modify our research question by clarifying the outcome we believe, based on prior knowledge, would actually be impacted by the treatment we have identified. Similarly, the process of collecting or assembling data may change our DAG by forcing us to use proxies for important confounders. (b) A decision tree for choosing a study design and statistical approach for causal inference. The research question, domain knowledge about the study system (e.g., understanding whether there are issues of simultaneity or compliance) and properties of the available data (e.g., presence of panel data) all shape the decision about which, if any, approaches will be appropriate and feasible. For each method in a grey box, we note additional key assumptions or requirements that must be met.

Randomised controlled experiments require the fewest and weakest assumptions for causal inference (Fernainy et al. 2024). However, even in experiments, several key assumptions must be met for potential outcomes—and thus counterfactuals—to be well-defined. First, experiments assume that the treatment T does not affect the outcome Y except through its effect on X, the cause being studied (the 'excludability' assumption): the treatment is solely responsible for the different outcomes observed, and there are no confounding variables. In addition, experiments must satisfy the stable unit treatment value assumption (SUTVA), an assumption common to all causal inference approaches we discuss. SUTVA has two key components: no interference (a unit's outcome is only conditional on whether it received treatment) and no multiple versions of the treatment (there is only a single, well-defined version of each treatment level). Finally, experiments assume that there is no 'non-compliance': units have or maintain the treatment they were assigned (e.g., in a seed addition experiment, the planted seeds emerge, no other species invade and no species fail to emerge). However, these assumptions can be challenging to meet; thus, experiments can deviate from perfect randomisation and compliance, highlighting the need to engage explicitly with causal thinking when interpreting the results of experiments (Kimmel et al. 2021). Furthermore, while randomised controlled experiments are viewed as the gold standard for causal inference in terms of internal validity (or the extent to which a study accurately estimates a causal relationship within a study population), generalising from experiments and creating experiments that replicate the conditions and scales of processes found in nature pose challenges.

5 | Quasi-Experimental Designs

Without randomised treatment assignment and experimental control, quasi-experimental designs can facilitate causal inference but require more assumptions—many of which are inherently untestable—to be met (Imbens 2024). These approaches require careful probing and justification of their assumptions based on system-specific knowledge to support the interpretation of arguably causal relationships. Quasi-experiments can be used at any spatial and temporal scale, while randomised, controlled ecological experiments in the field and lab are mostly restricted to smaller scales. Quasi-experiments often use specific data structures, such as cross-sectional and panel data. Cross-sectional data are observations from multiple units at a single point in time, facilitating comparison of treatment effects across individuals. Panel data are observations of multiple units across multiple time points.

Quasi-experimental approaches to causal inference must often contend with selection bias resulting from non-random treatment assignment. For example, we may be interested in the effect of land-based nutrient pollution on kelp cover (Krumhansl et al. 2016; Figure 4a). To answer this question, we might relate remotely sensed water quality data to long-term kelp cover monitoring data. However, distance to human settlements and the coast is likely an important confounding factor: it affects the amount of land-based pollution to which a kelp forest is exposed (the treatment) and it also affects fishing pressure on predators of sea urchins, which in turn affect kelp cover (Ling et al. 2009). If we simply compared kelp forests with high vs. low levels of nutrient pollution, we might attribute observed differences in kelp cover to pollution without accounting for the confounding effect of remoteness on fishing pressure. This example demonstrates selection bias: kelp forests exposed to the treatment are systematically different from untreated kelp forests in ways that affect the outcome. Kelp forests with the highest nutrient pollution are likely closer to coastal areas with high human population densities and thus also subject to higher

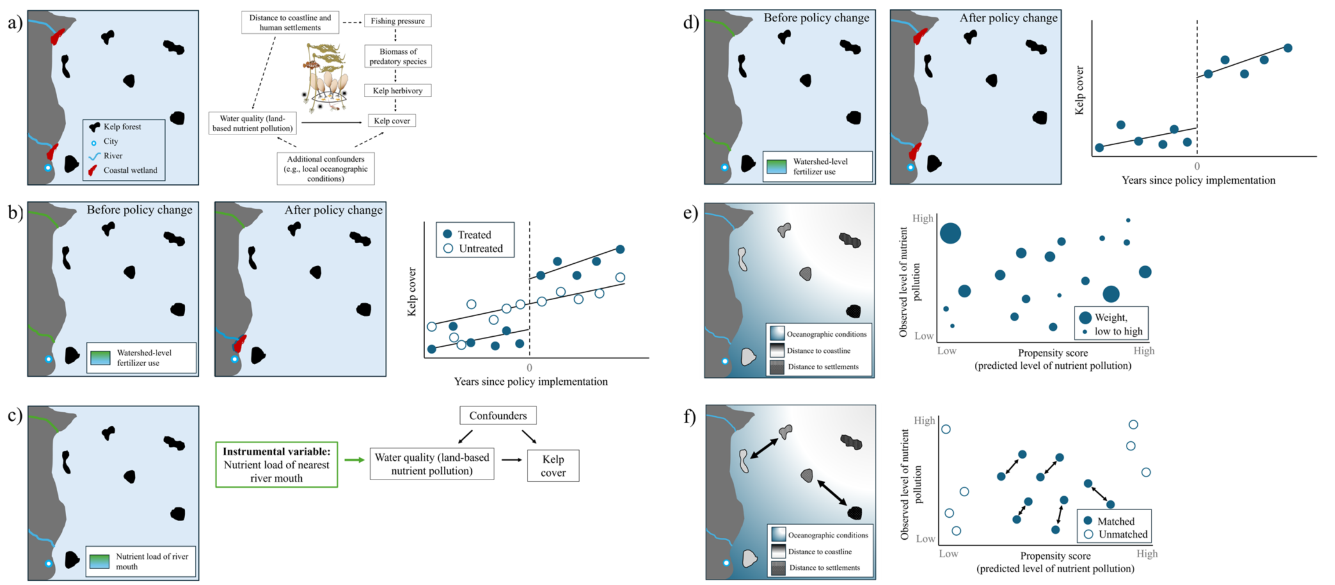

FIGURE 4 | Illustrations of quasi-experimental methods. (a) A DAG illustrates assumed causal relationships and confounders for a hypothetical study of land-based nutrient pollution's impact on kelp cover. (b) Difference-in-differences compares treated and untreated units before and after treatment implementation (here, a policy improving water quality discharged by rivers by reducing fertiliser use and restoring wetlands). (c) Instrumental variables isolate treatment effects through variables that impact the treatment but only influence the outcome through their relationship with the treatment. Here, green outlining indicates the instrumental variable (nearest river mouth's nutrient load). (d) Regression discontinuity designs compare units on either side of interventions (here, implementation of the policy from 4b). (e) Inverse probability of treatment weighting uses propensity scores to weight units based on the likelihood that their treatment status is the status predicted by their observable confounders. (f) Matching uses propensity scores to identify treated and untreated units with comparable confounding variables.

fishing pressure, while kelp forests with minimal pollution are far from the coast with less accessible fishing grounds (Witman and Lamb 2018). Selection bias stems from non-random treatment assignment: the units exposed to the treatment we wish to study are not randomly selected, which introduces confounding. When study designs fail to account for selection bias, the estimated difference in the mean outcomes for the treated and untreated groups actually represents the average causal effect *plus* the effect of selection bias.

Confounding variables may be observable (i.e., factors that the investigator has identified as potential confounders *and* measured) or unobservable (i.e., factors that are known and not measured, or unknown). It may not be possible to measure unobservable variables: the study site may lack historical records (Butsic et al. 2017), data may not be publicly available, or collecting these data may be prohibitively expensive. Different quasi-experimental methods take different approaches to dealing with, and make different assumptions about, the presence and importance of observable and unobservable confounders. And while quasi-experimental designs require more assumptions than randomised experiments to derive arguably causal findings, these approaches have the benefit of observing natural conditions (rather than conditions manipulated by the researcher) and enable analyses at broader scales.

We review the main approaches to quasi-experimental causal inference, categorising them according to whether or not they condition on unobservable confounders in addition to observables. We discuss each approach's assumptions, strengths, limitations and data requirements (Figure 3b). Butsic et al. (2017) and Larsen, Meng, and Kendall (2019) provide further introductions to these approaches. Across all these approaches, we recommend, as a first step, drawing a DAG—based on knowledge of the study system and ecological theory—with the treatment, outcome and all potential confounders, mediators and moderators, when they are relevant to the research question (Box 1).

5.1 | Conditioning on Observable and Unobservable Confounders

Among quasi-experimental approaches to causal inference, approaches that condition on both observable and unobservable confounders require the weakest assumptions for causal interpretation by relaxing the assumption that we have observed all confounders (Figure 4b–d). These approaches can yield arguably causal interpretations even if we cannot measure or do not know all confounding variables in our system, or if we have drawn an incorrect DAG and thus do not know the true data-generating process. We briefly review the core ideas and assumptions, applications to ecology and recent trends for these approaches, focusing on difference-in-difference designs, panel regressions, instrumental variables and regression discontinuity designs. In Table 2, we highlight more recent extensions for these designs.

We start with difference-in-difference (DiD) designs—similar conceptually to before-after control-impact (BACI) and thus familiar to ecologists (Green 1979; Stewart-Oaten and Bence 2001)—which compare the differences in control and treated groups before and after an intervention or exposure (reviewed, with extensions, in Wauchope Hannah et al. (2021). This approach compares the differences between the (*treated group after – treated group before*) – (*untreated group after – untreated group before*) to estimate how much more the treated group changed as compared to how much the untreated group changed (Figure 3b; Figure 4b). To create a counterfactual, difference-in-difference relies on the untestable assumption that the trends in time for these groups would be the same (or parallel) without the treatment. While most textbook examples consider binary treatments, difference-in-difference also applies to continuous treatments or treatments of different intensities (Callaway, Goodman-Bacon, and Sant'Anna 2024). This field is rapidly evolving, with emerging methodological extensions for cases where the parallel trends assumption is violated and effects are not homogenous (reviewed in Roth et al. (2023)).

Similarly, panel regressions, or 'within' estimators, make comparisons within groups, such as sites, individuals, or time periods. Panel regression controls for fixed differences across units and time-specific effects, or variables that affect all sites in a unit of time (Wooldridge 2010; Figure 3b). Time-invariant characteristics of sites can be confounding (e.g., more remote kelp forests also tend to be less impacted by land-based pollution (Figure 4)); these across-site differences are 'between variation'. To control for these differences, panel approaches use 'fixed effects'—dummy variables for each group to control for time-invariant, confounding differences across groups, whether or not the confounding variables are observed. Here, a fixed effect has a different meaning than its use in mixed effect and hierarchical modelling, which instead considers a fixed effect to be a parameter that does not vary by group (Bolker et al. 2009). With this approach, we can track how, within a location, kelp cover changes through time in response to other variables that change through time, like sea surface temperature. Thus, we can compare sites to themselves at different treatment levels (e.g., levels of nutrient pollution) observed at different points in time as the counterfactual (Dee et al. 2023).

These approaches differ from, and make weaker assumptions for causal identification than, mixed effects models using random effects (Byrnes and Dee 2024) or conditioning on observable confounding variables alone (Dee et al. 2023). The downside is that panel approaches 'throw out' the between variation (both confounding and otherwise) and require large panel datasets because they estimate a coefficient for each group and time (Angrist and Pischke 2008; Wooldridge 2010). Nested sampling designs can exploit cross-sectional data with multiple plots sampled across multiple sites and retain between-group variation (reviewed in Byrnes and Dee (2024) and Wooldridge (2010)). These approaches are increasingly used in ecology (e.g., Dudney et al. (2021), Ratcliffe et al. (2022), Suskiewicz et al. (2024)) and are straightforward to implement in R (Bergé 2018): see Dee et al. (2023) and Byrnes and Dee (2024) for tutorials.

Instrumental variables (IV) regression can eliminate sources of bias from all forms of confounding variables (including the time-varying confounding variables missed in DiD), measurement error, reverse causality and simultaneity (Box 1; Figure 3b). IV regression uses a third variable (an 'instrument', Z) that is related to the treatment, X, but not to the outcome, Y, except through its

TABLE 2 | A summary of some additional topics in causal inference that could inform ecological research, including recent methodological advances.

Topic	Description	Key readings
Experimental design and techniques to deal with imperfect experiments		
Challenges for experimental design and interpretation: non-compliance and attrition	Issues that arise when units do not receive the treatment they were assigned to or when units initially included in the sample are lost or otherwise not included in the analysis	Gerber and Green 2012*
Challenges for experimental and observational design and interpretation: interference and spillovers between units	SUTVA assumes no interference between units, but in ecological settings, there may be interactions and spillovers between units	Ogburn and VanderWeele 2014; Tchetgen and VanderWeele 2012* Ferraro, Sanchirico, and Smith 2018; Reich et al. 2021‡
Understanding mechanisms		
Mediation analysis and experimental design for mediation	Methods for assessing the direct effect of a treatment on a response and the indirect effect of the treatment, which is due to a mediator on the causal pathway from treatment to outcome	VanderWeele 2015; Pirlott and MacKinnon 2016* Huberman et al. 2020‡
Moderators, heterogeneous treatment effects and conditional average treatment effects	Methods for assessing when and how different units may respond differently to the treatment (e.g., when the effect of one variable on the response differs depending on the level of another variable, or moderator)	Athey and Imbens 2015*; Wager and Athey 2018 Ferraro and Hanauer 2014; Miller 2020; VanderWeele 2015*
Sensitivity analyses		
Partial identification	An approach to causal inference that uses weaker (and more plausible) assumptions to estimate the upper and lower bounds of a causal effect	Arriagada et al. 2012; Hazzah et al. 2014; McConnachie et al. 2016‡
Sensitivity analyses and placebo designs	Methods for testing the robustness of estimated causal effects to violations of underlying assumptions	Eggers, Tuñón, and Dafoe 2023; VanderWeele and Ding 2017; Cattaneo and Titiunik 2022; Liu, Kuramoto, and Stuart 2013*
Generalisability, reproducibility and transportability of effects		
Meta-analysis and generalisability	Approaches to generalising results from studies with varying levels of external validity	Spake et al. 2022; Nakagawa et al. 2023; Spake et al. 2023*
Replication and pre-registration	Improving reproducibility (e.g., through defining research questions and approaches a priori)	Nosek et al. 2018; Strømland 2019; Filazzola and Cahill 2021; Kimmel, Avolio, and Ferraro 2023*
Extensions: emerging tools for quasi-experimental approaches		

(Continues)

TABLE 2 | (Continued)

Topic	Description	Key readings
Staggered treatments, heterogeneity and robust difference in difference	Extensions to difference-in-difference methods that can accommodate units that enter treatment at different times and relax assumptions above parallel trends	Callaway and Sant'Anna 2021; Goodman-Bacon 2021*
Synthetic control methods	Methods of developing control units that use weighted averages of all potential control units to develop counterfactuals that are as comparable to the treated units as possible	Abadie, Diamond, and Hainmueller 2011* Abadie and Gardeazabal 2003; Sills et al. 2015; West et al. 2020‡
Causal inference with measurement error	Challenges posed by measurement error; methods for accounting for measurement error	Alix-García and Millimet 2023‡
Time series and dynamic panel models	Methods that allow for time lags, feedback and changing relationships between variables over time	Arellano and Bond 1991*
Causal discovery		
Causal discovery using graphical models	Data-driven approaches to learning causal relationships from large datasets	Glymour, Zhang, and Spirtes 2019; Runge et al. 2019, 2023; Spirtes, Glymour, and Scheines 2001*
Machine learning		
Machine learning fusion with causal inference	Recent advances blend machine learning approaches with causal inference (e.g., causal forests for heterogeneous treatment effects)	Athey 2015; Athey and Imbens 2015; Athey 2017; Athey and Imbens 2019; Pichler and Hartig 2023*

Note: For each topic, we provide a brief description and some key readings for further self-guided study. Key readings include texts discussing the fundamentals of a method (denoted with a *) and texts that demonstrate an application of the particular method (denoted with a ‡).

effect on X (or at least, after controlling for other variables in the system) (Angrist and Krueger 2001; Imbens 2014; Figure 4c). An IV in a regression mimics what an experiment's randomisation process would do, where the randomly assigned treatment process is independent of Y. The IV must be strongly related to the treatment but not with the outcome (after controlling for other covariates). When these two assumptions are met, IV regression yields a local average treatment effect (Box 1). The challenge is finding a valid and relevant IV.

For example, MacDonald and Mordecai (2019) used IV regression to isolate the effects of deforestation on malaria transmission and *vice versa* in the Amazon; because their effects are simultaneous, isolating one from another is challenging with standard methods such as mixed effect models. They used dry season aerosol pollution as an IV to isolate the effects of annual deforestation on malaria transmission from the reverse relationship. For causal interpretation, a key, untestable assumption is that dry season aerosol pollution and deforestation are strongly related (because most deforestation occurs and cleared forests are burned in the dry season) but that dry season aerosol pollution does not directly affect annual malaria transmission after controlling for other factors.

The final quasi-experimental approach to controlling for observable and unobservable confounders we discuss is regression discontinuity design (RDD) (Figure 4d). RDD is an option when there is a spatial, temporal, or policy discontinuity that separates treated from untreated units (Hahn, Todd, and Van der Klaauw 2001; Imbens and Lemieux 2008); Figure 3b. In RDDs, the treated and untreated units are sorted according to their position relative to a threshold in the 'running' variable (which defines the location of the discontinuity): on one side of the threshold, all units receive the treatment, while all units on the other side are untreated. RDDs compare the outcomes of units located directly on either side of the threshold to estimate the treatment effect (Cattaneo, Idrobo, and Titiunik 2019).

RDDs assume that the location of the discontinuity is exogenous: all observable and unobservable confounding variables are constant or continuous on either side of the threshold, without jumps in their values. As a result, units located directly on either side of the threshold are very similar to one another (there are generally no units observed directly at the threshold). In ecological systems, it can be difficult to identify appropriate, exogenous discontinuities in the absence of policy changes and management interventions (Englander 2019), although temporal discontinuities (e.g., before and after a disturbance event) may meet the assumptions of RDDs (Grainger and Costello 2014). RDDs assume that in the absence of the treatment, the outcome would not change discontinuously at the threshold (Hahn, Todd, and Van der Klaauw 2001). To assess the validity of this assumption, RDDs require sufficient data on both sides of the threshold (Wuepper and Finger 2023). We also assume that all unobserved confounders are either correlated with the running variable or not discontinuous across the threshold. Generally, RDDs estimate the treatment effect using a narrow bandwidth of units on either side of the threshold to avoid making assumptions about the shape of the underlying regression functions.

RDDs also assume that the probability of treatment changes discontinuously at the threshold (Cattaneo, Idrobo, and Titiunik 2019). In a sharp RDD, we assume perfect compliance: all units above the threshold receive the treatment, while none of the units below the cutoff receive it (Figure 3b). We can relax this assumption and use fuzzy RDD, which merely assumes that the probability of treatment changes discontinuously at the threshold (Wuepper and Finger 2023). Fuzzy RDDs allow for treatment noncompliance: the value of the running variable is a predictor of whether a unit received the treatment but does not completely determine its treatment status. The value of the running variable relative to the threshold thus functions as an IV that affects the outcome solely through its effect on the likelihood of treatment. RDDs also assume that there is no endogenous sorting of units: units do not seek to be on one side of the threshold (Lee 2008). In ecological applications, endogenous sorting may occur where animal behaviour comes into play—for example, the landscape of fear shapes animal movement (Gaynor et al. 2019)—or where treatments cause spatial spillovers—for example, protected area establishment increases extractive activities directly outside reserve boundaries (Ewers and Rodrigues 2008).

A strength of RDD is that many of the underlying assumptions can be tested visually (Cattaneo and Titiunik 2022). We can test whether the discontinuity is exogenous by plotting the values of confounding variables across the threshold and checking for discontinuous change (Cattaneo, Idrobo, and Titiunik 2019). Plotting the data using placebo thresholds can reveal whether there are locations with similar treatment effects in the absence of a treatment discontinuity (Noack et al. 2022; Wuepper and Finger 2023). Density tests that check for increased sample unit density on one side of the threshold can test for endogenous sorting (McCrary 2008).

5.2 | Conditioning on Observable Confounders

Quasi-experimental designs that condition on observable confounders (Figure 3b) make the strong, untestable assumption that all important confounders are observable. Two such approaches are inverse probability of treatment weighting (hereafter, 'weighting') and matching (Figure 4e,f). Both use observable confounders to calculate propensity scores, or the probability of a unit receiving a treatment based on that unit's covariate values (Rosenbaum and Rubin 1983; Stuart 2010). In matching, we develop a set of control and treated units by identifying the control units with propensity scores closest to those of the treated units (Figure 4f). We discard untreated units that do not have similar propensity scores to treated units and *vice versa*, maintaining only units with sufficient overlap in their covariate values (i.e., common support). In weighting, each unit is weighted based on its propensity score such that treated units with high propensity scores and untreated units with low propensity scores have lower weights than other units (Figure 4e). Weighting retains all units. We use the weights in the subsequent regression model to estimate the treatment effect. In both matching and weighting, including the covariates used to calculate the propensity score in the subsequent regression model increases the robustness of the treatment effect estimate (Jones and Lewis 2015). Matching is more commonly used with a binary treatment, although continuous treatments are sometimes stratified, and there is ongoing development of approaches for continuous treatments

(Brown et al. 2021; Fong, Hazlett, and Imai 2018; Hirano and Imbens 2004).

Weighting and matching integrate well with regression methods and work with both panel and cross-sectional data. Estimated treatment effects are also less sensitive to mis-specified models (Butsic et al. 2017), and propensity scores reduce bias from measurement error in the covariates (Austin 2010). There are also simple diagnostics to assess the quality of matches, including comparison of standardised mean differences pre- and post-matching. Matching can also reveal where there is not sufficient common support to make plausible causal claims (Ho et al. 2011). Finally, if unobserved confounders are correlated with the observed confounders, these approaches can adjust for unobservables.

The essential assumption of weighting and matching is that selection bias is caused by observable confounders or unobserved confounders that are correlated with observed variables: they suffer from omitted variable bias when there are unobservable confounders (Simler-Williamson and Germino 2022). Compared to experimental designs and quasi-experimental approaches that condition on unobservables, weighting and matching require stronger assumptions for causal interpretation. Finally, matching has several distinct limitations: it relies on sufficient common support for treated and untreated units, and it reduces the variation in the dataset because units without quality matches are dropped. The results must be interpreted in the context of the reduced dataset: the estimated treatment effect is valid for the range of units included in the matched dataset, but it may not be appropriate to extrapolate the estimated effect to the full dataset.

In addition, and complementary to, these quasi-experimental designs rooted in the PO framework are approaches from the SCM framework. In SCM, conditioning on all confounding variables is equivalent to applying Pearl's back-door criterion (blocking all back-door paths). This makes two assumptions: that all paths are specified correctly, and that confounding variables are observable and measured (Huntington-Klein 2022). When this cannot be achieved, an alternative is the front-door criterion, which adjusts for a mediator that is uncorrelated with the confounding variables of concern (Bellemare, Bloem, and Wexler 2024; Pearl 1995, 2009). Controlling for an exogenous mediator blocks the effect of omitted confounding variables and isolates the effect of the causal variable of interest (Pearl and Mackenzie 2018). To implement the front door criterion, one first estimates the effect of the treatment on the mediator without confounders and then estimates the effect of the mediator on the outcome. These two effects are multiplied to get the total effect of the treatment on the outcome (Arif and MacNeil 2023). However, identifying situations in which the front-door criterion works is challenging, so it is less frequently used (Huntington-Klein 2022).

6 | Discussion

Causal questions are central to ecological understanding, and ecology has a rich tradition of experiments to address causal questions and estimate the magnitude of causal effects. In recent years, ecological literature reviewing or applying causal inference approaches that complement experimental approaches has exploded, highlighting a variety of approaches that can exploit new data streams to extend ecological understanding to broader spatial and temporal scales. However, making sense of how and when to apply these approaches and navigating the wide-ranging, rapidly evolving, technical and jargon-filled fields that causal inference spans still pose challenges. In response, we review key challenges for causal inference using experimental and observational data in ecology, quasi-experimental approaches to answering causal questions and the key assumptions underlying these approaches. Building on previous reviews (e.g., Butsic et al. (2017) and Larsen, Meng, and Kendall (2019)), we explicitly define quasi-experimental designs in terms of their treatment of unobservable confounding variables. We believe that this distinction is very important for ecologists, as approaches that select on both observable and unobservable variables require weaker assumptions for causal interpretation. We demonstrate how the use of PO and SCM frameworks can be complementary and provide a workflow for moving from a causal question, to a DAG, to the appropriate methodological approach, to the interpretation of results (Figure 3). We also provide resources for self-guided study, including reproducible code with accompanying data and a curated reading list (Supporting Information).

Causal inference is not as straightforward as following a recipe or implementing a pre-existing software package. Robust causal inference requires careful combination of pre-existing knowledge (formalised in DAGs), appropriate data, study design and interpretation of estimated effects in light of key assumptions. Adding nuance, approaches for causal inference pose trade-offs and require different assumptions, some of which may be more or less plausible in particular contexts (Grace 2024). In addition, the approaches reviewed here emphasise carefully estimating one causal effect at a time, rather than estimating all causal effects in a system at once, although causal inference can contribute to the goal of building system-level knowledge (Box 3). To navigate these important nuances, we synthesise some critical considerations: the spectrum of weak to strong assumptions required for causal interpretation of estimated effects, different designs' trade-offs between internal and external validity and data requirements for causal inference. We offer recommendations for overcoming these limitations and outline future research needs.

6.1 | Tradeoffs Between Internal and External Validity

Causal inference designs exist along a spectrum from true randomisation to purely observational; this spectrum reflects both the strength of assumptions needed for causal interpretation and trade-offs in internal and external validity. While internal validity refers to accurate estimates of causal relationships within a study population, external validity is the extent to which a study's results can be applied beyond the study sample (Spake et al. 2022). Quasi-experimental approaches and randomised controlled experiments prioritise internal validity: researchers rigorously eliminate sources of bias in their estimates of the treatment effect (Desjardins et al. 2021). Much like experiments with different treatments, different quasi-experimental designs

BOX 3 | Common misconceptions about causal inference.

We clarify some misconceptions about causal inference approaches and their applicability to ecology.

1. Causal inference is purely statistical and does not rely on domain knowledge. Researchers using statistical designs for causal inference do not apply these methods in a vacuum, but rather draw on ecological knowledge to shape their research questions, hypotheses, study design and, importantly, the interpretation and caveats surrounding the estimated effects and reported causal relationships. Drawing a DAG is a useful first step for formalising prior knowledge (Figure 3).

2. Quasi-experiments can only handle binary treatments. Quasi-experiments can accommodate continuous and multivalue treatments (e.g., the effect of fire severity categories on forest biomass). They can also estimate heterogeneous effects (Table 2).

3. Causal inference does not provide information on mechanisms. The causal inference approaches we review can examine causal mechanisms, either by including moderators (e.g., interaction terms in a regression) or mediators on the causal path, using mediation analyses (VanderWeele 2015; Huberman et al. 2020) (Table 2, Figure 4).

4. The structural causal model (SCM) and the potential outcomes (PO) framework are competing and non-overlapping frameworks. SCM and PO are complementary frameworks and have been unified and translated from one to another (Malinsky, Shpitser, and Richardson 2019; Richardson and Robins 2013). Both seek to achieve unbiased estimates and make assumptions for causal interpretation transparent.

5. Prior ecological knowledge can tell us the size and direction of a causal relationship and the bias associated with its estimate. Ecological systems are complex, and while prior knowledge allows us to form hypotheses about the direction and magnitude of causal relationships, our knowledge is limited. We may be incorrect in our assumptions about the size and direction of the bias in our estimates. Bias could mean a true effect is masked (i.e., appears to be zero in an analysis) or the estimated effect is a mirage (i.e., a spurious effect, when there is no true effect). Bias can also lead to the incorrect sign of an effect or assumed relationship. Approaches such as the 'useful approximation standard' (Grace 2024), which suggest that an estimated causal effect must simply be 'predominantly causal' (i.e., the causal component of the estimate is greater than the bias component), thus make very strong assumptions, because the size of the bias versus that of the true causal effect is unknowable. Under these strong assumptions, researchers may run the risk of allowing confirmation bias to guide the interpretation of their results.

6. Studies or estimated effects are either causal or not causal. Causal inference designs fall along a spectrum based on the strength of the assumptions they make. Applying a quasi-experimental, experimental, or structural modelling approach does not guarantee that the estimated effect reflects a true causal effect. Causal interpretation of empirical estimates relies on assumptions and domain knowledge about whether those assumptions are met. While causal inference methods attempt to reach unbiased estimates, complete lack of bias is almost always unachievable. However, we believe that unbiased estimates should still be the goal, as there is no rigorous and reproducible definition of a 'good enough' estimate (*sensu* Grace 2024). For instance, if a researcher could fully randomise their experimental treatment, they would not opt to only partially randomise it. Interpretation of effect sizes and relationships from causal analyses in ecology requires transparency about their assumptions and limitations. Use of robustness tests can help assess the strength of findings (Box 4).

7. Causal inference methods do not allow for generalisability. While causal inference approaches prioritise internal validity, statistical designs such as matching provide more general estimands, and causal inference approaches can be integrated with meta-analysis to generalise their findings.

8. Quasi-experiments do not seek to understand how a system works. The approaches we discuss seek to build up ecological understanding through estimations of each individual element and process in the system. Often, multiple analyses and DAGs are needed to advance understanding of multiple relationships within that system, as an individual DAG may have many assumed relationships but not identify all causal pathways (i.e., an individual DAG may not satisfy the backdoor criterion for all relationships in the system).

9. There is a silver bullet for causal understanding. There is no one-size-fits-all approach or formula to follow for causal inference. Instead, causal inference is a process that iteratively integrates prior knowledge, data and causal assumptions. The choice of approach is based on the best available knowledge, methods and data, which are all evolving as science progresses.

yield distinct estimands, with varying implications for external validity. On one end of the spectrum are ideal, randomised controlled experiments, which prioritise internal validity and require the weakest assumptions for causal interpretation of effect sizes. However, experiments may struggle with external validity, as the controlled conditions and specific populations involved can limit the generalisability of findings to broader, real-world contexts (Dee et al. 2023) or other forms of treatment (Wolkovich et al. 2012). Moving further along the spectrum, quasi-experiments and imperfect experiments also have trade-offs between internal and external validity (Kowalski 2023). Both RDD and IV estimate local average treatment effects (LATE) rather than the average treatment effect (ATE) (Box 1). RDDs estimate the LATE using units located directly on either side of the discontinuity (Baker and Lindeman 2024): it may not be appropriate to extrapolate this LATE to units located far from the discontinuity, although emerging methods allow researchers to assess RDDs' external validity (Wing and Bello-Gomez 2018; Wuepper and Finger 2023). Similarly, the estimated causal effects of IV designs only apply to compliers (the units that vary in response to the IV, Box 1; Imbens 2010).

However, we often seek generalisability, or external validity, to extend our findings beyond the units and spatiotemporal scale that we studied (Spake et al. 2022). Moving further along the spectrum to observational studies that condition only on observables, approaches like matching estimate average treatment effects or average treatment effects on the treated (Box 1) but make stronger assumptions about our ability to identify and include all confounders. Still, in matching, because we exclude unmatched units, we cannot assume that the estimated treatment effect would apply to units whose covariate values fall outside the area of common support (Crump et al. 2009; Stuart 2010). For example, Siegel et al. (2022a) use matching to estimate the effect of federal vs. private land ownership on wildfire probability in western US forests. However, because federal wilderness areas tend to be at higher elevations than private forests, the matched dataset includes relatively few wilderness units (< 7% of federal units in the matched dataset). It would thus be inappropriate to naively extrapolate their findings to high-elevation wilderness forests.

6.2 | Data Considerations

Available data also determine generalisability, the choice of causal inference design (and therefore internal validity) and the statistical power to detect an effect. Quasi-experimental designs have specific data requirements: their appropriateness will depend on both the research question and the data context. As noted previously, cross-sectional and panel data are common dataset structures in quasi-experimental approaches. Difference-in-difference and panel designs require panel data, while IV and RDD require some plausibly exogenous sources of variation. With only cross-sectional data, design options are more limited (e.g., matching or weighting), and it is harder to flexibly control for confounding variables, particularly unobserved variables.

Cross-sectional versus panel data also may reflect different effects and degrees of generalisability. Cross-sectional data captures a snapshot in time and can be used for space-for-time comparisons, which are critiqued in applications such as climate change ecology for issues with generalisability (Lovell et al. 2023). While cross-sectional data allow us to examine how a particular treatment (e.g., exposure to reduced precipitation) affects multiple units (e.g., grassland plots in different locations), panel data allow us to examine trends over time across the treated and untreated units and generalise to multiple time points. This facilitates the study of ecologically interesting questions such as time lags in treatment effects, the effects of varying levels of treatment exposure over time and interactions between the treatment and covariates over time. However, there may also be trade-offs in existing datasets in terms of spatial extent versus resampling through time. Furthermore, a reliance on panel data that includes the pre-treatment period is ecologically limiting, as we are less likely to have these data for processes such as climate change impacts, species introductions and unexpected disturbances. The realities of funding and data collection logistics may also restrict the availability of panel data.

Sample size is a related consideration; many quasi-experimental approaches require relatively large datasets for sufficient statistical power to detect effects. Thus, ecologists working with limited datasets from field-based observations may not have sufficient data to leverage causal inference methods or enough power to detect a treatment effect (Kimmel, Avolio, and Ferraro 2023; Lemoine et al. 2016). When there are interactions between the treatment and other covariates, the required sample size increases. New data streams can not only scale up ecological understanding and inferences when coupled with quasi-experimental approaches but also increase statistical power.

Synthesis and meta-analyses can help expand external validity by combining multiple internally valid studies covering a range of naturally occurring conditions (Spake et al. 2022). Meta-analysis is a common approach to quantitative synthesis in ecology, especially of experiments. However, when meta-analyses include original studies with biased estimands, they can yield biased estimates and inaccurate results. This limitation is true for observational designs and imperfect experiments (Kimmel et al. 2021). Further, the estimands may not be the same across studies, muddying quantitative comparisons. Similarly, if the original studies feeding into a meta-analysis focus on different subpopulations with heterogeneous treatment effects, it becomes difficult to combine and generalise the estimated effects (Spake et al. 2022). Study eligibility criteria can reduce the probability of including original studies with bias: we recommend that the field develop eligibility criteria based on the treatment of confounding variables and other sources of bias. We may need to develop other approaches to account for remaining endogeneity in the original studies (Mathur and VanderWeele 2022) and for comparisons when different estimands and subpopulations are involved.

More generally, synthesis science combines datasets from disparate sources and often seeks to disentangle causal relationships (Carpenter et al. 2009; Halpern et al. 2020). Quasi-experimental approaches can expand and accelerate synthesis science's contributions to ecological knowledge, but measurement error (the difference between the true vs. recorded value of a variable) presents a challenge. Synthesis approaches combine multiple data sources, each with their own sources of measurement error. When the extent and types of measurement error differ across studies, error and uncertainty can propagate through models using synthesised data.

Another opportunity is using large-scale datasets, such as time series derived from satellite imagery, combined with causal inference approaches. The volume of data from Earth observations, community science programs and other distributed surveys and monitoring networks is rapidly increasing, expanding observations of ecological systems at larger scales and the sample sizes available for data-hungry approaches. More observations of ecosystems under a wider variety of time points, conditions and scales will also increase the generalisability of inferences and enable us to test new theories that span different spatial and temporal scales of causal relationships.

A challenge posed by these expanding data streams, however, is mismatches in spatial and temporal resolution between the treatment, confounders and outcome. For example, remote sensing data can facilitate analysis at broader spatial scales but are often available at coarser resolutions, which can obscure understanding of highly localised processes (Alix-García and

BOX 4 | Best Practices for Causal Inference in Ecology.

To take advantage of causal inference approaches responsibly and effectively, some understanding of their underlying assumptions and the contexts in which one study design is more robust than another is needed. To interpret an estimated effect or correlation as causal, assumptions are always required, even in randomised experiments (Kimmel et al. 2021). Indeed, just using a randomised experiment or quasi-experimental approach does not guarantee that the interpretation holds causal meaning. While we believe best practices should emerge collectively, we suggest some (non-exhaustive) guidelines to move this process forward, based on our experiences teaching, applying and reviewing papers on causal inference using ecological data.

Data collection and pre-analysis planning. Best practices start before data collection, and study design enables more robust causal inference. An emerging focus is on best practices for reproducibility in ecology prior to data analysis (Kimmel, Avolio, and Ferraro 2023; Parker et al. 2016).

- Create a DAG based on domain knowledge and use it to guide data collection or assembly of existing data.
- Where possible, collect or assemble pre-treatment data to expand the study designs available for your analysis (e.g., difference-in-difference, temporal regression discontinuity designs, or other panel design) in both experimental and quasi-experimental settings.
- Use pre-registration or pre-analysis plans to define your study design in advance. This enhances reproducibility, clarifies assumptions and reduces the likelihood of p-hacking.
- Perform power analyses, particularly in field studies with low replication, datasets with small sample sizes, or when aiming to estimate interactions (moderator) effects or site-specific effects.

Study and statistical design. When possible, use designs that make weaker assumptions and triangulate results using complementary methods and sensitivity tests.

- Use DAGs to guide your analysis choices and include them in a pre-analysis plan.
- Where possible (based on the research question and data), use methods that make fewer and weaker assumptions about confounding variables (e.g., methods that condition on observable and unobservable confounders: IV, RDD, difference-in-difference, within-estimator and other panel designs), because even with extensive domain knowledge, ecological surprises are still possible.
- Use multiple, complementary designs to assess the robustness of estimates to different assumptions about observed and unobserved confounding (e.g., Dee et al. (2023)).
- Use sensitivity tests to assess the robustness of estimates to the presence of confounding variables that have not been controlled for (Andraczek et al. 2024; Liu, Kuramoto, and Stuart 2013).

Communicating assumptions. Clarity around assumptions allows others to understand the study's strengths and limitations and can help identify further avenues for research.

- Include your DAG(s) in published analyses and interpret your model results in relation to your DAG. This will allow others to understand and critique your causal model, and we expect published DAGs to also help the field identify commonly unobserved confounders that may require new data streams to address.
- Clearly discuss the assumptions of your study and statistical design, including in randomised experiments, and how they are met as well as caveats.
- Explain how your design accounts for observable and unobservable sources of confounding.

Interpreting results. Careful interpretation of results given your study design's limitations and assumptions increases transparency and credibility.

- Interpret your findings in the context of your study's assumptions and all forms of bias that may remain in your estimates. If your model does not consider unobserved confounders, note that your estimates of causal effects may include the effect of both the causal variable and unobserved confounders.
- Explicitly discuss the limitations of your analysis in terms of potential reasons the assumptions required for causal interpretation could be violated.
- Interpret analyses based on their external validity and avoid over-generalising your inferences.

Changing incentives and norms. As a community, a cultural shift that prioritises transparency and robustness in publishing would improve the credibility of causal inference. As reviewers, we recommend viewing explicit acknowledgement of a study's assumptions and limitations (e.g., transparency around issues with internal validity given potential violations of assumptions and limits to the generalisability of findings) as a strength, rather than a weakness or a justification for rejecting a study's findings. Current publishing incentives, which may dissuade transparency around assumptions and limitations, put the robustness and credibility of causal inference at risk. More transparency around these could help the science to build on itself (e.g., by collecting new data to overcome assumptions or developing new methods to relax them).

Millimet 2023; Jain 2020). If we were interested in the effect of artificial nighttime lights on large mammal behaviour, for example, we might have outcome data at fine spatio-temporal scales (e.g., multiple data points per hour, accurate within several meters) from radiocollars, treatment data at 30 × 30 meter resolutions in the form of daily nighttime light data (Román et al. 2018) and data on environmental and socioeconomic confounders at various resolutions. However, newer remote sensing techniques and products (such as LiDAR) can generate data at finer spatial and temporal resolutions over larger spatial extents, helping to overcome issues with scale mismatches. There is also a growing literature examining the unique challenges when using remote sensing data for causal inference, as these data are often derived using machine learning. This magnifies challenges for controlling for confounders and of measurement error. For instance, if confounders are included in the machine learning model that predicts the data, that can introduce bias. Analogously, errors from machine learning model predictions are a form of measurement error in subsequent causal models that can introduce bias (Alix-García and Millimet 2023; Gordon et al. 2023; Jain 2020; Proctor, Carleton, and Sum 2023). New methodological and conceptual advances are needed to reconcile these challenges and facilitate larger scale ecological causal understanding (Van Cleemput et al. 2024).

6.3 | Establishing Shared Best Practices for Causal Inference in Ecology

Further integration of these approaches into the research design and statistics curriculum for graduate students in ecology can help us harness the power of causal inference (Box 2). Many ecologists report a desire for more statistical training and a mismatch between their formal training and current best practices (Barraquand et al. 2014; Touchon and McCoy 2016). In our experience, graduate students are eager to learn new approaches. Through an emphasis on building students' intuitions regarding the strengths, limitations and underlying assumptions of causal inference designs, focused coursework can strengthen students' research design and statistical skills. As the use of causal inference in ecology becomes more popular, we also need careful, critical reviewers to evaluate and provide input into these studies. Fortunately, a growing body of ecological studies, applications and general resources can contribute to self-guided and course-based learning (Heiss 2022; Huntington-Klein 2022; Supporting Information). As more ecologists gain a working understanding of how causal inference can integrate with ecological research, we can develop a collective and evolving set of best practices as a field (Box 4).

Finally, this synthesis is not exhaustive. Table 2 summarises additional topics and references. While we focus on counterfactual-based causal inference approaches to build on ecology's rich history of experimentation, there are alternative notions of causality, such as causal detection (Munch, Rogers, and Sugihara 2023; Runge et al. 2023; Sugihara et al. 2012) or causal discovery (Spirtes, Glymour, and Scheines 2001). Lastly, the designs we present can also be estimated using structural equation modelling (Shipley 1999, 2009) and Bayesian approaches (Li, Ding, and Mealli 2023; Oganisian and Roy et al. 2021).

7 | Conclusion

With growing interest in and use of causal inference techniques, best practices—that are decided on and adopted by the field of ecology—are needed. This will allow us to effectively and constructively evaluate each other's work and build on it. Transparency is key, as causal analyses rely on assumptions at multiple stages, from study design to estimation of treatment effects and causal interpretation. By clearly stating and justifying our assumptions, we can create more credible estimates of causal effects and more reproducible results, enabling others to build on existing studies through improvements in data and methods. Ongoing and future improvements in estimation and identification (reviewed in Athey and Imbens (2019) and Roth et al. (2023))—which are rapidly evolving in diverse fields, including ecology—can potentially weaken the underlying assumptions required for causal interpretations (Roth et al. 2023). Transparency about underlying assumptions can also help readers interpret the estimand, determine whether they believe a causal interpretation is appropriate and understand the limits of a result's generalisability (Spake et al. 2022). Finally, transparency ensures that the approaches used are appropriate for the question at hand. Credible causal estimates will enable us to advance basic and applied ecology, informing ecological theory and ecosystem management at broad scales.

Author Contributions

K.S. and L.D. both conceived of the paper, wrote the paper and created the figures and RMarkdown tutorials.

Acknowledgements

L.D. acknowledges support from the US National Science Foundation NSF CAREER #2340606 and NASA BioSCape #80NSSC 22K0796. K.S. acknowledges support from a NOAA Climate and Global Change Postdoctoral Fellowship. We thank Jon Chase, A. Simler-Williamson and two anonymous reviewers for their feedback. We thank the students in our course, Causal Inference in Ecological Data, in Spring 2023 at the University of Colorado-Boulder. We thank Andrew Heiss, Paul Ferraro and Van Butsic for inspiring teaching materials.

Data Availability Statement

Data used for the tutorials in the Supporting Information are available at (1) https://github.com/katherinesiegel/intro_causal_inf, along with accompanying code, and (2) the Open Science Framework at DOI: 10.17605/OSF.IO/3XVQG.

Peer Review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ele.70053.

References

Abadie, A., A. Diamond, and J. Hainmueller. 2011. "Synth: An R Package for Synthetic Control Methods in Comparative Case Studies." *Journal of Statistical Software* 42: 1–17.

Abadie, A., and J. Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93: 113–132.

Addicott, E. T., E. P. Fenichel, M. A. Bradford, M. L. Pinsky, and S. A. Wood. 2022. "Toward an Improved Understanding of Causation in the Ecological Sciences." *Frontiers in Ecology and the Environment* 20: 474–480.

Alix-García, J., and D. L. Millimet. 2023. "Remotely Incorrect? Accounting for Nonclassical Measurement Error in Satellite Data on Deforestation." *Journal of the Association of Environmental and Resource Economists* 10: 1335–1367.

Andraczek, K., L. E. Dee, A. Weigelt, et al. 2024. "Weak Reciprocal Relationships Between Productivity and Plant Biodiversity in Managed Grasslands." *Journal of Ecology* 112: 2359–2373.

Angrist, J. D., and A. B. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15: 69–85.

Angrist, J. D., and J.-S. Pischke. 2008. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Angrist, J. D., and J.-S. Pischke. 2015. *Mastering 'Metrics: The Path From Cause to Effect*. Princeton, NJ: Princeton University Press.

Arellano, M., and S. Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58: 277.

Arif, S., and M. A. MacNeil. 2022a. "Predictive Models Aren't for Causal Inference." *Ecology Letters* 25: 1741–1745.

Arif, S., and M. A. MacNeil. 2022b. "Utilizing causal diagrams across quasi-experimental approaches." *Ecosphere* 13: e4009.

Arif, S., and M. A. MacNeil. 2023. "Applying the Structural Causal Model Framework for Observational Causal Inference in Ecology." *Ecological Monographs* 93: e1554.

Arif, S., and M. D. B. Massey. 2023. "Reducing Bias in Experimental Ecology Through Directed Acyclic Graphs." *Ecology and Evolution* 13: e9947.

Arriagada, R. A., P. J. Ferraro, E. O. Sills, S. K. Pattanayak, and S. Cordero-Sancho. 2012. "Do Payments for Environmental Services Affect Forest Cover? A Farm-Level Evaluation From Costa Rica." *Land Economics* 88: 382–399.

Athey, S. 2015. "Machine Learning and Causal Inference for Policy Evaluation." In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 5–6. New York, NY, USA: ACM.

Athey, S. 2017. "Beyond Prediction: Using Big Data for Policy Problems." *Science (1979)* 355: 483–485.

Athey, S., and G. W. Imbens. 2015. "Machine Learning for Estimating Heterogeneous Casual Effects." (No. 3350).

Athey, S., and G. W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annu Rev Econom* 11: 685–725.

Austin, P. C. 2010. "The Performance of Different Propensity-Score Methods for Estimating Differences in Proportions (Risk Differences or Absolute Risk Reductions) in Observational Studies." *Statistics in Medicine* 29: 2137–2148.

Baker, S. G., and K. S. Lindeman. 2024. "Multiple Discoveries in Causal Inference: LATE for the Party." *Chance* 37: 21–25.

Barraquand, F., T. H. G. Ezard, P. S. Jørgensen, et al. 2014. "Lack of Quantitative Training Among Early-Career Ecologists: A Survey of the Problem and Potential Solutions." *PeerJ* 2: e285.

Bellemare, M. F., J. R. Bloem, and N. Wexler. 2024. "The Paper of How: Estimating Treatment Effects Using the Front-Door Criterion." *Oxford Bulletin of Economics and Statistics* 86: 951–993.

Bergé, L. 2018. Efficient Estimation of Maximum Likelihood Models With Multiple Fixed-Effects: The R Package FENmlm CREA Discussion Papers.

Bolker, B. M., M. E. Brooks, C. J. Clark, et al. 2009. "Generalized Linear Mixed Models: A Practical Guide for Ecology and Evolution." *Trends in Ecology & Evolution* 24: 127–135.

Brown, D. W., T. J. Greene, M. D. Swartz, A. V. Wilkinson, and S. M. DeSantis. 2021. "Propensity Score Stratification Methods for Continuous Treatments." *Statistical Medicine* 40: 1189–1203.

Butsic, V., D. J. Lewis, V. C. Radeloff, M. Baumann, and T. Kuemmerle. 2017. "Quasi-Experimental Methods Enable Stronger Inferences From Observational Data in Ecology." *Basic and Applied Ecology* 19: 1–10.

Byrnes, J. E. K., and L. E. Dee. 2024. "Causal Inference With Observational Data and Unobserved Confounding Variables." *bioRxiv*.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna. 2024. *Difference-In-Differences With a Continuous Treatment*. Cambridge, MA: National Bureau of Economic Research.

Callaway, B., and P. H. C. Sant'Anna. 2021. "Difference-In-Differences With Multiple Time Periods." *Journal of Econometrics* 225: 200–230.

Carpenter, S. R., E. V. Armbrust, P. W. Arzberger, et al. 2009. "Accelerate Synthesis in Ecology and Environmental Sciences." *Bioscience* 59: 699–701.

Cattaneo, M. D., N. Idrobo, and R. Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs*. Cambridge, UK: Cambridge University Press.

Cattaneo, M. D., and R. Titiunik. 2022. "Regression discontinuity designs." *Annu Rev Econom* 14: 821–851.

Christie, A. P., T. Amano, P. A. Martin, G. E. Shackelford, B. I. Simmons, and W. J. Sutherland. 2019. "Simple Study Designs in Ecology Produce Inaccurate Estimates of Biodiversity Responses." *Journal of Applied Ecology* 56: 2742–2754.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik. 2009. "Dealing With Limited Overlap in Estimation of Average Treatment Effects." *Biometrika* 96: 187–199.

Cunningham, S. 2021. *Causal Inference: The Mixtape*. New Haven, CT: Yale University Press.

Dee, L. E., P. J. Ferraro, C. N. Severen, et al. 2023. "Clarifying the Effect of Biodiversity on Productivity in Natural Ecosystems With Longitudinal Data and Methods for Causal Inference." *Nature Communications* 14: 2607.

Dee, L. E., S. J. Miller, L. E. Peavey, et al. 2016. "Functional Diversity of Catch Mitigates Negative Effects of Temperature Variability on Fisheries Yields." *Proceedings of the Royal Society B: Biological Sciences* 283: 20161435.

Desjardins, E., J. Kurtz, N. Kranke, A. Lindeza, and S. H. Richter. 2021. "Beyond Standardization: Improving External Validity and Reproducibility in Experimental Evolution." *Bioscience* 71: 543–552.

Dudney, J., C. E. Willing, A. J. Das, A. M. Latimer, J. C. B. Nesmith, and J. J. Battles. 2021. "Nonlinear Shifts in Infectious Rust Disease due to Climate Change." *Nature Communications* 12: 5102.

Eggers, A. C., G. Tuñón, and A. Dafoe. 2023. "Placebo Tests for Causal Inference." *American Journal of Political Science* 68: 1106–1121.

Englander, G. 2019. "Property Rights and the Protection of Global Marine Resources." *Nature Sustainability* 2: 981–987.

Ewers, R. M., and A. S. L. Rodrigues. 2008. "Estimates of Reserve Effectiveness Are Confounded by Leakage." *Trends in Ecology & Evolution* 23: 113–116.

Fernainy, P., A. A. Cohen, E. Murray, E. Losina, F. Lamontagne, and N. Sourial. 2024. "Rethinking the Pros and Cons of Randomized Controlled Trials and Observational Studies in the Era of Big Data and Advanced Methods: A Panel Discussion." *BMC Proceedings* 18: 1.

Ferraro, P. J. 2009. "Counterfactual Thinking and Impact Evaluation in Environmental Policy." In *Environmental Program and Policy Evaluation: Addressing Methodological Challenges*, edited by M.

Birnbaum and P. Mickwitz, vol. 2009, 75–84. San Francisco, CA: Jossey-Bass.

Ferraro, P. J., and M. M. Hanauer. 2014. "Advances in Measuring the Environmental and Social Impacts of Environmental Programs." *Annual Review of Environment and Resources* 39: 495–517.

Ferraro, P. J., and S. K. Pattanayak. 2006. "Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments." *PLoS Biology* 4: 482–488.

Ferraro, P. J., J. N. Sanchirico, and M. D. Smith. 2018. "Causal Inference in Coupled Human and Natural Systems." *Proceedings of the National Academy of Sciences* 116: 5311–5318.

Fick, S. E., T. W. Nauman, C. C. Brungard, and M. C. Duniway. 2021. "Evaluating Natural Experiments in Ecology: Using Synthetic Controls in Assessments of Remotely Sensed Land Treatments." *Ecological Applications* 31: e02264.

Filazzola, A., and J. F. Cahill. 2021. "Replication in Field Ecology: Identifying Challenges and Proposing Solutions." *Methods in Ecology and Evolution* 12: 1780–1792.

Fisher, R. A. 1935. *The Design of Experiments*. Edinburgh, Scotland: Oliver & Boyd.

Fong, C., C. Hazlett, and K. Imai. 2018. "Covariate Balancing Propensity Score for a Continuous Treatment: Application to the Efficacy of Political Advertisements." *Annals of Applied Statistics* 12: 156–177.

García Criado, M., I. H. Myers-Smith, A. D. Bjorkman, C. E. R. Lehmann, and N. Stevens. 2020. "Woody Plant Encroachment Intensifies Under Climate Change Across Tundra and Savanna Biomes." *Global Ecology and Biogeography* 29: 925–943.

Gaynor, K. M., J. S. Brown, A. D. Middleton, M. E. Power, and J. S. Brashares. 2019. "Landscapes of Fear: Spatial Patterns of Risk Perception and Response." *Trends in Ecology & Evolution* 34: 355–368.

Gaynor, K. M., C. E. Hojnowski, N. H. Carter, and J. S. Brashares. 2018. "The Influence of Human Disturbance on Wildlife Nocturnality." *Science (1979)* 360: 1232–1235.

Geldmann, J., A. Manica, N. D. Burgess, L. Coad, and A. Balmford. 2019. "A Global-Level Assessment of the Effectiveness of Protected Areas at Resisting Anthropogenic Pressures." *Proceedings of the National Academy of Science* 116: 1–7.

Gerber, A. S., and D. P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W.W. Norton & Company.

Glymour, C., K. Zhang, and P. Spirtes. 2019. "Review of Causal Discovery Methods Based on Graphical Models." *Frontiers in Genetics* 10: 524.

Goodman-Bacon, A. 2021. "Difference-In-Differences With Variation in Treatment Timing." *Journal of Econometrics* 225: 254–277.

Gordon, M., M. Ayers, E. Stone, and L. C. Sanford. 2023. "Remote Control: Debiasing Remote Sensing Predictions for Causal Inference." In *ICLR 2023 Workshop on Tackling Climate Change With Machine Learning*. Climate Change AI.

Grace, J. B. 2021. "Instrumental Variable Methods in Structural Equation Models." *Methods in Ecology and Evolution* 12: 1148–1157.

Grace, J. B. 2024. "An Integrative Paradigm for Building Causal Knowledge." *Ecological Monographs* 94: e1628.

Grace, J. B., T. M. Anderson, E. W. Seabloom, et al. 2016. "Integrative Modelling Reveals Mechanisms Linking Productivity and Plant Species Richness." *Nature* 529: 390–393.

Grainger, C. A., and C. J. Costello. 2014. "Capitalizing Property Rights Insecurity in Natural Resource Assets." *Journal of Environmental Economics and Management* 67: 224–240.

Green, R. H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. New York: Wiley.

Greenstone, M., and T. Gayer. 2009. "Quasi-experimental and experimental approaches to environmental economics." *Journal of Environmental Economics and Management* 57: 21–44.

Hahn, J., P. Todd, and W. Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design." *Econometrica* 69: 201–209.

Halofsky, J. E., D. L. Peterson, and B. J. Harvey. 2020. "Changing Wildfire, Changing Forests: The Effects of Climate Change on Fire Regimes and Vegetation in the Pacific Northwest, USA." *Fire Ecology* 16: 4.

Halpern, B. S., E. Berlow, R. Williams, et al. 2020. "Ecological Synthesis and Its Role in Advancing Knowledge." *Bioscience* 70: 1005–1014.

Hazzah, L., S. Dolrenry, L. Naughton, et al. 2014. "Efficacy of Two Lion Conservation Programs in Maasailand, Kenya." *Conservation Biology* 28: 851–860.

Heiss, A. 2022. "Program Evaluation for Public Service." Accessed June 12, 2024. https://evalf22.classes.andrewheiss.com/.

Hernan, M. A. 2004. "A Definition of Causal Effect for Epidemiological Research." *J Epidemiol Community Health (1978)* 58: 265–271.

Hernán, M. A., J. Hsu, and B. Healy. 2019. "A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks." *Chance* 32: 42–49.

Hernán, M. A., and J. M. Robins. 2016. "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available." *American Journal of Epidemiology* 183: 758–764.

Hirano, K., and G. W. Imbens. 2004. "The Propensity Score With Continuous Treatments." In *Applied Bayesian Modeling and Causal Inference From Incomplete-Data Perspectives*, edited by X.-L. Meng and A. Gelman, 73–84. New York, NY: Wiley Series in Probability and Statistics.

Ho, D. E., K. Imai, G. King, and E. A. Stuart. 2011. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software* 42: 1–28.

Holland, P. W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81: 945–960.

Huberman, D. B., B. J. Reich, K. Pacifici, and J. A. Collazo. 2020. "Estimating the Drivers of Species Distributions With Opportunistic Data Using Mediation Analysis." *Ecosphere* 11: e03165.

Huntington-Klein, N. 2022. *The Effect: An Introduction to Research Design and Causality*. New York, NY: Chapman & Hall.

Imbens, G. W. 2010. "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)." *Journal of Economic Literature* 48: 399–423.

Imbens, G. W. 2014. "Instrumental variables: An econometrician's perspective." *Statistical Science* 29: 323–358.

Imbens, G. W. 2024. "Causal Inference in the Social Sciences." *Annual Review of Statistics and Its Application* 11: 123–152.

Imbens, G. W., and T. Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics* 142: 615–635.

Jain, M. 2020. "The Benefits and Pitfalls of Using Satellite Data for Causal Inference." *Review of Environmental Economics and Policy* 14: 157–169.

Jones, J. P. G., and G. Shreedhar. 2024. "The Causal Revolution in Biodiversity Conservation." *Nature Human Behavior* 8: 1236–1239.

Jones, K. W., and D. J. Lewis. 2015. "Estimating the Counterfactual Impact of Conservation Programs on Land Cover Outcomes: The Role of Matching and Panel Regression Techniques." *PLoS One* 10: 1–22.

Kimmel, K., M. L. Avolio, and P. J. Ferraro. 2023. "Empirical Evidence of Widespread Exaggeration Bias and Selective Reporting in Ecology." *Nature Ecology & Evolution* 7: 1525–1536.

Kimmel, K., L. E. Dee, M. L. Avolio, and P. J. Ferraro. 2021. "Causal Assumptions and Causal Inference in Ecological Experiments." *Trends in Ecology & Evolution* 36: 1141–1152.

Knapp, R. A., and K. R. Matthews. 2000. "Non-native Fish Introductions and the Decline of the Mountain Yellow-Legged Frog From Within Protected Areas." *Conservation Biology* 14: 428–438.

Kowalski, A. E. 2023. "How to Examine External Validity Within an Experiment." *Journal of Economics and Management Strategy* 32: 491–509.

Krumhansl, K. A., D. K. Okamoto, A. Rassweiler, et al. 2016. "Global Patterns of Kelp Forest Change Over the Past Half-Century." *Proceedings of the National Academy of Sciences* 113: 13785–13790.

Larsen, A. E., K. Meng, and B. E. Kendall. 2019. "Causal Analysis in Control–Impact Ecological Studies With Observational Data." *Methods in Ecology and Evolution* 10: 924–934.

Larsen, A. E., and F. Noack. 2020. "Impact of Local and Landscape Complexity on the Stability of Field-Level Pest Control." *Nature Sustainability* 4: 120–128.

Laubach, Z. M., E. J. Murray, K. L. Hoke, R. J. Safran, and W. Perng. 2021. "A biologist's Guide to Model Selection and Causal Inference." *Proceedings of the Royal Society B: Biological Sciences* 288: 2020–2815.

Lee, D. S. 2008. "Randomized Experiments From Non-random Selection in U.S. House Elections." *Journal of Econometrics* 142: 675–697.

Lemoine, N. P., A. Hoffman, A. J. Felton, et al. 2016. "Underappreciated Problems of Low Replication in Ecological Field Studies." *Ecology* 97: 2554–2561.

Li, F., P. Ding, and F. Mealli. 2023. "Bayesian Causal Inference: A Critical Review. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences." 381.

Ling, S. D., C. R. Johnson, S. D. Frusher, and K. R. Ridgway. 2009. "Overfishing Reduces Resilience of Kelp Beds to Climate-Driven Catastrophic Phase Shift." *Proceedings of the National Academy of Sciences* 106: 22341–22345.

Little, R. J., and D. B. Rubin. 2000. "Causal Effects in Clinical and Epidemiological Studies via Potential Outcomes: Concepts and Analytical Approaches." *Annual Review of Public Health* 21: 121–145.

Liu, W., S. J. Kuramoto, and E. A. Stuart. 2013. "An Introduction to Sensitivity Analysis for Unobserved Confounding in Nonexperimental Prevention Research." *Prevention Science* 14: 570–580.

Locke, D. H., B. Hall, J. M. Grove, et al. 2021. "Residential Housing Segregation and Urban Tree Canopy in 37 US Cities." *Npj Urban Sustainability* 1: 15.

Lovell, R. S. L., S. Collins, S. H. Martin, A. L. Pigot, and A. B. Phillimore. 2023. "Space-For-Time Substitutions in Climate Change Ecology and Evolution." *Biological Reviews* 98: 2243–2270.

MacDonald, A. J., and E. A. Mordecai. 2019. "Amazon Deforestation Drives Malaria Transmission, and Malaria Burden Reduces Forest Clearing." *Proceedings of the National Academy of Sciences of the United States of America* 116: 22212–22218.

Malinsky, D., I. Shpitser, and T. Richardson. 2019. "A Potential Outcomes Calculus for Identifying Conditional Path-Specific Effects." In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*. Naha, Japan: PMLR.

Mathur, M. B., and T. J. VanderWeele. 2022. "Methods to Address Confounding and Other Biases in Meta-Analyses: Review and Recommendations." *Annual Review of Public Health* 43: 19–35.

McConnachie, M. M., C. Romero, P. J. Ferraro, and B. W. van Wilgen. 2016. "Improving Credibility and Transparency of Conservation Impact Evaluations Through the Partial Identification Approach." *Conservation Biology* 30: 371–381.

McCrary, J. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142: 698–714.

Miller, S. 2020. "Causal Forest Estimation of Heterogeneous and Time-Varying Environmental Policy Effects." *Journal of Environmental Economics and Management* 103: 102337. https://doi.org/10.1016/j.jeem.2020.102337.

Morgan, S. L., and C. Winship. 2014. *Counterfactuals and Causal Inference*. Cambridge, UK: Cambridge University Press.

Munch, S. B., T. L. Rogers, and G. Sugihara. 2023. "Recent Developments in Empirical Dynamic Modelling." *Methods in Ecology and Evolution* 14: 732–745.

Nakagawa, S., Y. Yang, E. L. Macartney, R. Spake, and M. Lagisz. 2023. "Quantitative Evidence Synthesis: A Practical Guide on Meta-Analysis, Meta-Regression, and Publication Bias Tests for Environmental Sciences." *Environmental Evidence* 12: 8.

Noack, F., A. Larsen, J. Kamp, and C. Levers. 2022. "A bird's Eye View of Farm Size and Biodiversity: The Ecological Legacy of the Iron Curtain." *American Journal of Agricultural Economics* 104: 1460–1484.

Nosek, B. A., C. R. Ebersole, A. C. DeHaven, and D. T. Mellor. 2018. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences* 115: 2600–2606.

Oganisian, A., and J. A. Roy. 2021. "A Practical Introduction to Bayesian Estimation of Causal Effects: Parametric and Nonparametric Approaches." *Statistics in Medicine* 40: 518–551.

Ogburn, E. L., and T. J. VanderWeele. 2014. "Causal Diagrams for Interference." *Statistical Science* 29: 559–578.

Parker, T. H., W. Forstmeier, J. Koricheva, et al. 2016. "Transparency in Ecology and Evolution: Real Problems, Real Solutions." *Trends in Ecology & Evolution* 31: 711–719.

Pearl, J. 1995. "Causal Diagrams for Empirical Research." *Biometrika* 82: 669–688.

Pearl, J. 2009. *Causality*. Cambridge, UK: Cambridge University Press.

Pearl, J. 2010. "An Introduction to Causal Inference." *International Journal of Biostatistics* 6: 7.

Pearl, J., and D. Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. USA: Basic Books, Inc.

Pichler, M., and F. Hartig. 2023. "Machine Learning and Deep Learning—A Review for Ecologists." *Methods in Ecology and Evolution* 14: 994–1016.

Pirlott, A. G., and D. P. MacKinnon. 2016. "Design Approaches to Experimental Mediation." *Journal of Experimental Social Psychology* 66: 29–38.

Proctor, J., T. Carleton, and S. Sum. 2023. *Parameter Recovery Using Remotely Sensed Variables*. Cambridge, MA: National Bureau of Economic Research.

Ramsey, D. S. L., D. M. Forsyth, E. Wright, M. McKay, and I. Westbrooke. 2019. "Using Propensity Scores for Causal Inference in Ecology: Options, Considerations, and a Case Study." *Methods in Ecology and Evolution* 10: 320–331.

Ratcliffe, H., M. Ahlering, D. Carlson, S. Vacek, A. Allstadt, and L. E. Dee. 2022. "Invasive Species Do Not Exploit Early Growing Seasons in Burned Tallgrass Prairies." *Ecological Applications* 32: e2641.

Ratcliffe, H., A. Kendig, S. Vacek, D. Carlson, M. Ahlering, and L. E. Dee. 2024. "Extreme Precipitation Promotes Invasion in Managed Grasslands." *Ecology* 105: e4190.

Reich, B. J., S. Yang, Y. Guan, A. B. Giffin, M. J. Miller, and A. Rappold. 2021. "A Review of Spatial Causal Inference Methods for Environmental and Epidemiological Applications." *International Statistical Review* 89: 605–634.

Richardson, T. S., and J. M. Robins. 2013. *Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical*

Approaches to Causality (No. 128). Seattle, WA: Center for Statistics and the Social Sciences, University of Washington.

Rohrer, J. M. 2018. "Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data." *Advances in Methods and Practices in Psychological Science* 1: 27–42.

Román, M. O., Z. Wang, Q. Sun, et al. 2018. "NASA's Black Marble Nighttime Lights Product Suite." *Remote Sensing of Environment* 210: 113–143.

Rosenbaum, P. R., and D. B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41–55.

Roth, J., P. H. C. Sant'Anna, A. Bilinski, and J. Poe. 2023. "What's Trending in Difference-In-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics* 235: 2218–2244.

Rubin, D. B. 1972. "Estimating Causal Effects of Treatments in Experimental and Observational Studies." *ETS Research Bulletin Series* 1972: i–31.

Rubin, D. B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100: 322–331.

Runge, J., A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls. 2023. "Causal Inference for Time Series." *Nature Reviews Earth and Environment* 4: 487–505.

Runge, J., P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic. 2019. "Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets." *Science Advances* 5: eaau4996.

Saxby, T., J. Hawkey, and S. Anderson. 2024. "Integration and Application Network Media Library."

Shipley, B. 1999. "Testing Causal Explanations in Organismal Biology: Causation, Correlation and Structural Equation Modelling." *Oikos* 86: 374–382.

Shipley, B. 2009. "Confirmatory Path Analysis in a Generalized Multilevel Context." *Ecology* 90: 363–368.

Siegel, K. J., L. Larsen, C. Stephens, W. Stewart, and V. Butsic. 2022a. "Quantifying Drivers of Change in Social Ecological Systems: Land Management Impacts Wildfire Probability in Forests of the Western US." *Regional Environmental Change* 22: 98.

Siegel, K. J., L. Macaulay, M. Shapero, et al. 2022b. "Impacts of Livestock Grazing on the Probability of Burning in Wildfires Vary by Region and Vegetation Type in California." *Journal of Environmental Management* 322: 116092.

Sills, E. O., D. Herrera, A. J. Kirkpatrick, et al. 2015. "Estimating the Impacts of Local Policy Innovation: The Synthetic Control Method Applied to Tropical Deforestation." *PLoS One* 10: e0132590.

Simler-Williamson, A. B., and M. J. Germino. 2022. "Statistical Considerations of Nonrandom Treatment Applications Reveal Region-Wide Benefits of Widespread Post-Fire Restoration Action." *Nature Communications* 13: 3472.

Smith, M. D., K. D. Wilkins, M. C. Holdrege, et al. 2024. "Extreme Drought Impacts Have Been Underestimated in Grasslands and Shrublands Globally." *Proceedings of the National Academy of Sciences* 121: e2309881120.

Spake, R., D. E. Bowler, C. T. Callaghan, et al. 2023. "Understanding 'It Depends' in Ecology: A Guide to Hypothesising, Visualising and Interpreting Statistical Interactions." *Biological Reviews* 98: 983–1002.

Spake, R., R. E. O'Dea, S. Nakagawa, et al. 2022. "Improving Quantitative Synthesis to Achieve Generality in Ecology." *Nature Ecology & Evolution* 6: 1818–1828.

Spirtes, P., C. Glymour, and R. Scheines. 2001. *Causation, Prediction, and Search*. 2nd ed. Cambridge, MA: MIT Press.

Splawa-Neyman, J. 1923. "On the Application of Probability Theory to Agricultural Experiments: Essay on Principles. Section 9." *Annals of Agricultural Sciences*: 1–51.

Stewart-Oaten, A., and J. R. Bence. 2001. "Temporal and Spatial Variation in Environmental Impact Assessment." *Ecological Monographs* 71: 305–339.

Strømland, E. 2019. "Preregistration and Reproducibility." *Journal of Economic Psychology* 75: 102143.

Stuart, E. A. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25: 1–21.

Suding, K. N. 2011. "Toward an Era of Restoration in Ecology: Successes, Failures, and Opportunities Ahead." *Annual Review of Ecology, Evolution, and Systematics* 42: 465–487.

Sugihara, G., R. May, H. Ye, et al. 2012. "Detecting Causality in Complex Ecosystems." *Science (1979)* 338: 496–500.

Suskiewicz, T. S., J. E. K. Byrnes, R. S. Steneck, R. Russell, C. J. Wilson, and D. B. Rasher. 2024. "Ocean Warming Undermines the Recovery Resilience of New England Kelp Forests Following a Fishery-Induced Trophic Cascade." *Ecology* 105: e4334.

Tchetgen, E. J. T., and T. J. VanderWeele. 2012. "On Causal Inference in the Presence of Interference." *Statistical Methods in Medical Research* 21: 55–75.

Tilman, D., F. Isbell, and J. M. Cowles. 2014. "Biodiversity and Ecosystem Functioning." *Annual Review of Ecology, Evolution, and Systematics* 45: 471–493.

Tilman, D., P. B. Reich, J. Knops, D. Wedin, T. Mielke, and C. Lehman. 2001. "Diversity and Productivity in a Long-Term Grassland Experiment." *Science (1979)* 294: 843–845.

Touchon, J. C., and M. W. McCoy. 2016. "The Mismatch Between Current Statistical Practice and Doctoral Training in Ecology." *Ecosphere* 7: e01394.

Van Cleemput, E., P. B. Adler, K. N. Suding, A. J. Rebelo, B. Poulter, and L. E. Dee. 2024. "Scaling-Up Ecological Understanding With Remote Sensing and Causal Inference." *Trends in Ecology & Evolution*.

VanderWeele, T. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford, UK: Oxford University Press.

VanderWeele, T. J., and P. Ding. 2017. "Sensitivity Analysis in Observational Research: Introducing the E-Value." *Annals of Internal Medicine* 167: 268–274.

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

Wauchope Hannah, S., T. Amano, J. Geldmann, et al. 2021. "Evaluating Impact Using Time-Series Data." *Trends in Ecology & Evolution* 36: 196–205.

West, T. A. P., J. Börner, E. O. Sills, and A. Kontoleon. 2020. "Overstated Carbon Emission Reductions From Voluntary REDD+ Projects in the Brazilian Amazon." *Proceedings of the National Academy of Sciences of the United States of America* 117: 24188–24194.

Wing, C., and R. A. Bello-Gomez. 2018. "Regression Discontinuity and Beyond." *American Journal of Evaluation* 39: 91–108.

Witman, J. D., and R. W. Lamb. 2018. "Persistent Differences Between Coastal and Offshore Kelp Forest Communities in a Warming Gulf of Maine." *PLoS One* 13: e0189388.

Wolkovich, E. M., B. I. Cook, J. M. Allen, et al. 2012. "Warming Experiments Underpredict Plant Phenological Responses to Climate Change." *Nature* 485: 494–497.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Wright, S. 1921. "Correlation and Causation." *Journal of Agricultural Research* 20: 557–585.

Wu, X., E. Sverdrup, M. D. Mastrandrea, M. W. Wara, and S. Wager. 2023. "Low-Intensity Fires Mitigate the Risk of High-Intensity Wildfires in California's Forests." *Science Advances* 9: eadi4123.

Wuepper, D., and R. Finger. 2023. "Regression discontinuity designs in agricultural and environmental economics." *European Review of Agricultural Economics* 50: 1–28.

Xu, X., A. Huang, E. Belle, P. De Frenne, and G. Jia. 2022. "Protected Areas Provide Thermal Buffer Against Climate Change." *Science Advances* 8: eabo0119.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.