Quantifying intermediary processes in ecology using causal mediation analyses

Hannah E. Correia*[1], Laura E. Dee[2], and Paul J. Ferraro[1,3]

[1]Department of Environmental Health and Engineering, Johns Hopkins University, Maryland, 21218, USA
[2]Ecology and Evolutionary Biology, University of Colorado Boulder, Colorado, 80309, USA
[3]Carey Business School, Johns Hopkins University, Maryland, 21202, USA

Abstract

Ecologists seek to understand the intermediary ecological processes through which changes in one attribute in a system affect other attributes. Yet, quantifying the causal effects of these mediating processes in ecological systems is challenging. Researchers must define what they mean by a "mediated effect", determine what assumptions are required to estimate mediation effects without bias, and assess whether these assumptions are credible for a study. To address these challenges, scholars in fields outside of ecology have made significant advances in mediation analysis over the past three decades. Here, we bring these advances to the attention of ecologists, for whom understanding mediating processes and deriving causal inferences are important for testing theory and developing resource management and conservation strategies. To illustrate both the challenges and the advances in quantifying mediation effects, we use a hypothetical ecological study. With this study, we show how common research designs used in ecology to detect and quantify mediation effects may have biases and how these biases can be addressed through alternative designs. Throughout the review, we highlight how causal claims rely on causal assumptions, and we illustrate how different designs or definitions of mediation effects can relax some of these assumptions. In contrast to statistical assumptions, causal assumptions are not verifiable from data, so we also describe procedures that researchers can use to assess the sensitivity of a study's results to potential violations of its causal assumptions. The advances in causal mediation analyses reviewed herein will provide ecological researchers with approaches to clearly communicate the causal assumptions necessary for valid inferences and examine potential violations to these assumptions, which will enable rigorous and reproducible explanations of intermediary processes in ecology.

Keywords and phrases: ecological mechanisms, causality, confounding, mediator, indirect effects.

*Corresponding author: H. E. Correia, 3400 N. Charles St., Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. E-mail: hcorrei2@jhu.edu

1 Introduction

Ecologists seek a causal understanding of ecosystem dynamics. A key part of this understanding is obtained by identifying and quantifying the intermediary ecological processes through which changes in one variable produce changes in other variables. The study of these intermediary processes, which we refer to as "mediators" but are sometimes also referred to as ecological "mechanisms" (Heger, 2022; Poliseli et al., 2022), involves decomposing overall effects into their constituent mediation effects. For example, scientists may be interested in the effect of drought on tree mortality and whether this effect is mediated by changes in carbohydrate reserves. Similarly, conservation scientists and practitioners may seek to understand whether and by how much changes in poaching may mediate the effect of protected areas on species abundance.

Although ecological studies frequently make implicit or explicit causal claims about mediation effects in experimental and observational studies, the substantial challenges in making such claims have not been clearly addressed in the ecology literature. To address these challenges, ecologists require a standardised language that can guide the development of appropriate empirical designs and articulate the causal assumptions necessary for drawing causal inferences about mediating variables. Although recent publications have introduced causal inference concepts to ecologists (Arif and MacNeil, 2023, 2022b; Grace and Irvine, 2020; Larsen et al., 2019; Ramsey et al., 2019; Ribas et al., 2021), they have not described the challenges in estimating mediation effects. Moreover, the literature has inadequately emphasised the importance of causal assumptions and has not clearly distinguished them from statistical assumptions (e.g., linearity, additivity, homoscedasticity, and normality). The mantra that credible causal inferences are not possible without explicit causal assumptions is one of the most important insights from the last three decades of advances in methods for causal inference (Rubin, 2006; Pearl, 2009; Shipley, 2000), and significant developments have been made in extending these methods for mediation analyses (MacKinnon, 2012; VanderWeele, 2015) of which ecologists can take advantage.

Here, we review important conceptual and methodological advances in mediation analysis that have been made in statistics, social science, biostatistics, and computer science but which have remained largely unadopted in ecology despite their potential for elucidating intermediary ecological processes. To introduce the terminology that is commonly used in the causal mediation literature, we use a hypothetical ecological study. We also use this study to describe how common designs in ecology for detecting or quantifying mediation effects may have biases, that is, systematic deviations between the estimated effect and the true underlying causal effect. We then show how these biases can be addressed through alternative experimental or statistical designs. Each design relies on different causal assumptions to make causal claims about the signs and magnitudes of mediation effects. Throughout our review, we focus on transparently describing these assumptions and discussing when they may be violated for ecological studies. At the end of our review, we demonstrate how these assumptions can be articulated and understood using the potential outcomes framework (Holland, 1986, 1988; Rubin, 2005), one of several analogous causal inference frameworks available for defining and estimating causal effects in experimental and observational studies (Dawid, 2000, 2021; Pearl, 2009; Rubin, 1974, 2006). The potential outcomes framework extends classical approaches to mediation analysis by providing a unifying and rigorous

structure that can be flexibly applied across ecological settings and data distributions.

2 Motivating example

We illustrate the concepts, methods, and challenges associated with studying mediators in ecological systems using a hypothetical example of an experimental study in which researchers aim to quantify how drought affects productivity in grassland ecosystems (e.g. Hoover et al. 2018; Pennisi 2022; Wilkins et al. 2022). The researchers hypothesise that one way that drought reduces productivity in grasslands is by changing soil moisture. In other words, they hypothesise that soil moisture is a mediator through which drought affects productivity in grasslands (Figure 1a). The researchers are not only interested in determining whether changes in soil moisture induced by drought lead to changes in productivity. They also want to quantify how much of the influence of drought on productivity comes from this change in soil moisture: "On average, about X% of the effect of a drought treatment on productivity arises from the effect of drought treatment on soil moisture." The researchers are aware that soil moisture may not be the sole mediator, but they choose soil moisture as the mediation effect to quantify in the study. Estimating the effects of multiple mediating variables within one study can bring additional challenges that are discussed briefly in Supplement S.5.

In the experiment, researchers randomly assign grassland plots to a rainfall exclusion treatment, which mimics drought conditions by preventing access to rainfall using overhead shelters (Figure 1b and 1c). Some time after random assignment of the treatment, the researchers measure soil moisture and productivity on each plot. Thus, the drought treatment is binary and soil moisture and productivity are continuous variables. We assume that the idealised experimental conditions for a randomised controlled trial are met (Cox 1958; Neyman et al. 1935; Rubin 1974, and reviewed in Kimmel et al. 2021). At the end of the experiment, the plots randomly assigned to the drought treatment are found to exhibit, on average, decreased productivity in comparison to the control plots (Figure 1b).

By using an illustrative example in which researchers randomly assign the treatment, we can focus on the key issues that arise in all study designs aimed at estimating the effects of mediators, whether the treatment is randomised or not. Although the drought treatment was randomised across plots, the mediator, soil moisture, was not. This is a common feature in experimental designs in ecology, because randomising intermediary ecological processes is challenging (see Section 5.1).

When estimating the causal effect of a mediator in a design that does not randomise the mediator, researchers face the same challenges that must be addressed in any observational design, particularly the challenge of eliminating the effects of other variables that influence both soil moisture and productivity, such as the influence of grazing by herbivores (Eldridge et al., 2017; Sitters and Olde Venterink, 2015; Veldhuis et al., 2014). Variables like herbivory and the challenges they pose for estimating the effects of mediation are described in more detail in Sections 3 to 5. Even if it were possible to conduct a follow-up experiment in which soil moisture was randomised, or both soil moisture and drought were randomised, drawing inferences about the mediation effects in the original experiment can be challenging (see Section 5.1).

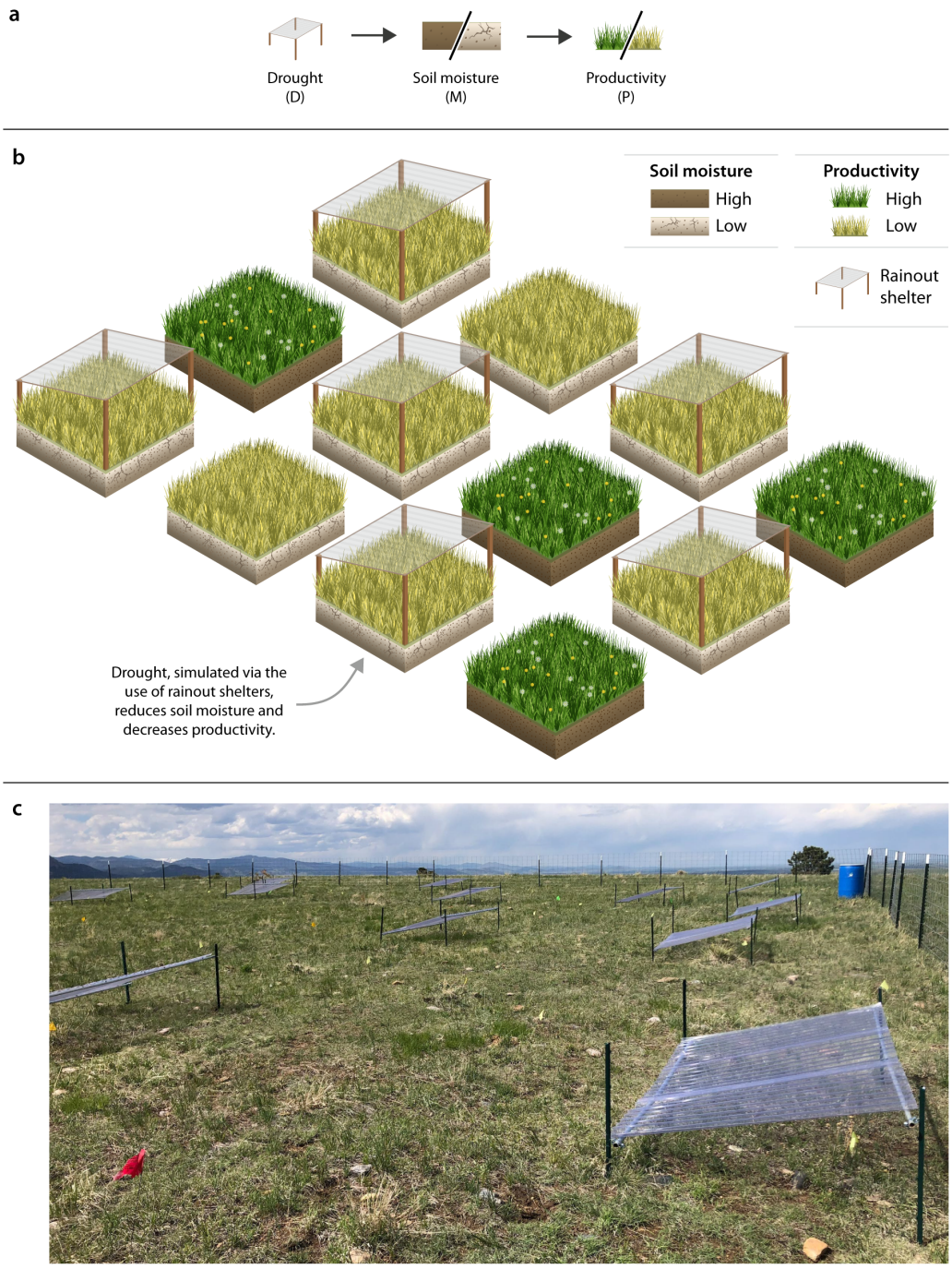


Figure 1: (a) The hypothesis for our hypothetical drought study expressed as a causal diagram in which arrows imply causal relationships between variables. For visual simplicity, the continuous variables soil moisture and productivity are represented as binary. (b) Results from the hypothetical experiment on 12 grassland plots, where 6 plots have been randomly assigned treatment with a rainout shelter. Rainout shelters reduce soil moisture by blocking precipitation, which in turn reduces plot productivity. (c) Photo of a drought experiment with rainout shelters in Boulder, Colorado, USA (Photo credit: Meghan Hayden).

3 Concepts

In this section, we introduce causal inference terminology that is used to distinguish the roles of key variables in a system, define the causal effects to be estimated, determine how to estimate these effects without bias, and communicate the results. Familiarity with this terminology is useful for articulating and verifying the assumptions required to make causal claims in a study.

3.1 Causal graphs

A tool often used by researchers to identify hypothesised causal relationships in a study, communicate the underlying assumptions required for causal inference, and obtain guidance on appropriate statistical analyses is a causal directed acyclic graph (DAG) (Digitale et al., 2022; Greenland et al., 1999; Pearl, 2000). In DAGs (e.g., Figure 1a), arrows between variables imply causal dependence between the variables but do not specify a functional relationship (i.e., they are 'non-parametric'). DAGs are directed, meaning that arrows defining causal relationships go in only one direction between two variables; there are no bidirectional arrows. The absence of an arrow between two variables implies that the researchers assume no causal relationship between the variables. Additionally, DAGs do not allow for feedback loops or paths of directed arrows that create a closed loop, hence they are "acyclic." Bidirectional and feedback relationships usually reflect unresolved temporally ordered effects or unmeasured common causes (Hernán and Robins, 2006; Murray and Kunicki, 2022; Pearce and Lawlor, 2017). A complete DAG includes all known or hypothesised common causes of any pair of variables represented in the causal diagram. Path diagrams of structural equation models (SEMs) are a special case of DAGs that include additional parametric and distributional assumptions (Kunicki et al., 2023; Pearl, 2000; Shipley, 2000).

3.2 Variables in mediation analysis

Before designing or conducting the hypothetical drought experiment, researchers may describe their hypothesis with a DAG to identify the relevant variables in the study. We begin with an incomplete DAG that does not yet include all common causes in the drought study (Figure 2a). In experimental designs, the manipulated causal variable is typically referred to as the **treatment** or exposure. In our hypothetical study, the treatment is drought, which is represented by two possible states: a treated state in which drought conditions are applied through rainout shelters and a control state in which no drought conditions are applied. This treatment is binary, but it could be discrete (e.g., "low", "medium", "high") or continuous (e.g., millimeters of precipitation). The treatment is randomised across units, which are plots in our example study (Figure 1b). The variable hypothesised to be causally affected by a change in the treatment is referred to as the **outcome**, which in the case of our example study is aboveground grassland productivity in a plot.

Since soil moisture is hypothesised to act as an intermediary between the treatment and outcome in the drought study, it is referred to as a **mediator**. A mediator is always on the **causal path**, that is, the path between a treatment and an outcome (indicated in red in Figure 2). The process through which the treatment's effect arises through one or multiple

mediators is called mediation, and the set of methodologies by which the magnitudes of the mediating effects are estimated is known as mediation analysis. In an ecological system, there can be multiple mediators by which a cause can influence an outcome, and multiple mediators can be on the same causal path (Figure 2b).

Mediators are often confused with **moderators**, which leads to misconceptions and misinterpretations in causal analyses (Ferraro and Hanauer, 2015; Holmbeck, 2019; Kraemer et al., 2008; Wu and Zumbo, 2008). Mediators and moderators play very different roles in the effect of a treatment on an outcome, and thus the distinction between the two is important for valid causal mediation analyses (Baron and Kenny, 1986; MacKinnon, 2011). Moderators do not lie on the causal path but instead affect or "moderate" the strength or direction of a causal effect. Moderators interact with treatments and mediators to alter their effects on the outcome, a phenomenon known as interaction or "effect modification". In our hypothetical study, soil type or texture in each plot may modify the effect of drought on productivity (Figure 2c). For example, drought may have a different effect on soil moisture in clay soil than in sandy soil, because clay soil can retain moisture for longer periods. The moderation of the effect of drought on soil moisture would thus modify the overall effect of drought on productivity across different soil types, creating heterogeneous treatment effects. If distinguishing the heterogeneous effects of drought on productivity for different soil types is of interest, moderator or subgroup analysis can be used (VanderWeele, 2012a; Wu and Zumbo, 2008). Moderator analysis can also be combined with mediation analysis (VanderWeele, 2012a, 2014; Wu and Zumbo, 2008). While we primarily focus on methods for directly estimating mediation effects, interactions created by moderators introduce heterogeneity that must be handled appropriately to estimate causal effects without bias (see Sections 4 and 6 and Supplement S.3).

Factors that influence at least two variables along the causal path are known as **confounders**, or sometimes "common causes". Confounding is a key concern for estimation of causal effects, as confounders induce dependence between treatment, mediator, and outcome that may not be due to true causal relationships. Confounders can therefore mask or mimic causal relationships among treatment, mediator, and outcome. Hence, failure to account for confounders leads to bias in the estimation of causal effects (Addicott et al., 2022). Consider the potential confounders W, K, and G in the hypothetical drought system (Figure 2d). Treatment-mediator confounders, such as topographic features or climate zones, influences both drought and soil moisture (W in Figure 2d). Treatment-outcome confounders, like temperature (K in Figure 2d) or air pollution, can affect grassland productivity as well as the frequency and duration of drought. Mediator-outcome confounders, such as historical grazing (G in Figure 2d), affect both soil moisture and productivity. Like moderators, confounders do not lie on the causal path.

The labels "treatment", "mediator", and "outcome" are context dependent. Drought, for example, could be viewed as a mediator if we consider an expanded version of our hypothetical drought study where the manipulated treatment is cloud seeding, which is hypothesised to influence grassland productivity through drought, soil moisture, and other processes (Figure 2e). While these labels may be somewhat artificial when describing an ecological system, adhering to causal terminology is helpful for clearly identifying key parts of a study and their respective roles when estimating causal effects. This nomenclature has not been used in a standardised manner in ecology and related fields like conservation science (e.g., Cinner et al.

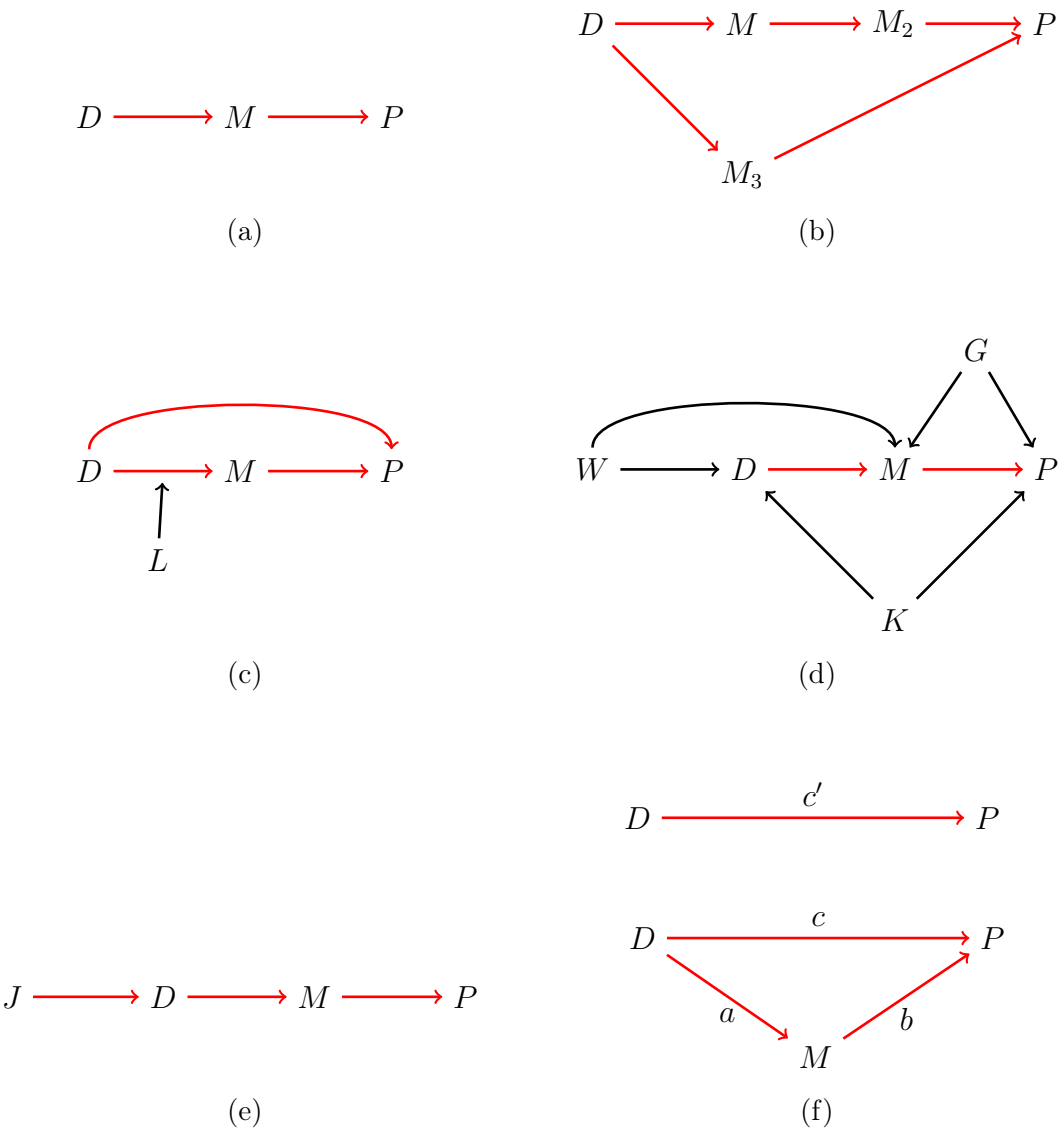


Figure 2: Causal diagrams of potential hypotheses for the hypothetical drought system with (a) only the treatment D, mediator M, and outcome P (an incomplete DAG); (b) multiple mediators between the treatment and the outcome; (c) a moderator L that interacts with the drought treatment to affect the relationship between treatment and outcome; (d) treatment-mediator confounder W, mediator-outcome confounder G, and treatment-outcome confounder K; (e) an alternative exposure of interest J that relabels the original treatment D as a mediator; and (f) the causal path of the incomplete DAG in panel (a) labelled to indicate the total effect c', direct effect c, and indirect effect composed of a and b. D = drought, M = soil moisture, M_2 = secondary mediator (e.g., photosynthesis), M_3 = alternative mediator (e.g., frequency of wildfires), P = productivity, L = soil type, W = topography, K = temperature, G = historical grazing, J = cloud seeding. Causal paths are in red.

2018), which makes it difficult to identify the roles of the variables under investigation in a study and the assumptions that researchers presume are met when estimating mediation effects, including which confounders are accounted for and which are not (Arif and MacNeil, 2023; Kimmel et al., 2021). Having identified the relevant components of a causal DAG that represents a study system, we next describe the effects to be estimated in mediation analysis.

3.3 Effects in mediation analysis

In mediation analysis, we are interested in breaking down the total effect of a treatment on an outcome into its constituent parts through one or more mediators in the system (Figure 2f). The overall effect of a treatment on an outcome is known as the **total effect**, which includes the effects of all conceivable mediators along all possible paths from the treatment to the outcome (path c' in Figure 2f). The total effect represents the change in the outcome when the treatment is changed from control to treated (if the treatment variable is binary), or when the treatment is changed by one unit value (in the case of a discrete or continuous treatment) while holding all other variables not on the causal path constant. The total effect provides no information on how the effect on the outcome is achieved through changes in mediators.

The effect of the treatment on the outcome that operates through an observed mediator is known as an **indirect effect**, which captures the magnitude of the relationship between the treatment and outcome that is attributable to the mediator. Hence, an indirect effect is sometimes referred to as a "mediated effect" (VanderWeele and Vansteelandt, 2014; MacKinnon et al., 2007). An indirect effect is influenced by both the magnitude and direction of the relationship between the treatment and mediator (path a in Figure 2f) and by the magnitude and direction of the relationship between the mediator and the outcome (path b in Figure 2f).

The causal effect of the treatment on the outcome that is not transmitted through the mediator of interest is referred to as the **direct effect** (path c in Figure 2f). The direct effect is not equivalent to an unmediated effect, although some texts refer to it as such. Indeed, there is no such thing as a truly unmediated causal effect (Le Poidevin, 2007; Mellor, 1995). The direct effect represents the effect through all other pathways from the treatment to the outcome that are not of interest or are unobservable to the researchers. We therefore think of the direct effect as the part of the total effect that does not pass through the mediator of interest. In many causal diagrams, the direct effect is not drawn but is implied (e.g., Figures 2a, 2b, 2d and 2e).

In our hypothetical drought study, the total effect of drought D on grassland productivity P represents the causal effect that would occur if we could change drought from the control state (no rainout shelter), $D = 0$, to the treated state (with rainout shelter), $D = 1$, in a grassland plot. Hence, the total effect for a given plot is often referred to as the individual treatment effect. In our example study, the total effect is hypothesised to be mediated, at least in part, by soil moisture M, which means that the total effect is composed of (1) the indirect effect, which is the effect that would occur if D were fixed at 1 and the value of soil moisture were changed from the value it takes when $D = 0$ to the value it takes when $D = 1$, and (2) the direct effect, which is the effect that would occur if D were changed from 0 to 1 but the value of soil moisture were held to the value it takes when $D = 0$. In Section 6, we

will discuss how mediation effects can be defined in other ways that may also be of interest to ecologists.

In the next section, we outline the causal and statistical assumptions necessary for estimating effects in mediation analysis without bias. In subsequent sections, we will examine approaches to quantifying effects in mediation analysis and discuss the conditions under which these approaches may or may not satisfy key causal assumptions.

4 Causal assumptions for estimating effects in mediation analyses

To draw causal inferences about effects in mediation analysis using experimental or observational data, researchers must make several causal and statistical assumptions. In this section, we describe the requisite causal assumptions and distinguish them from the statistical assumptions that are often addressed in ecological analyses of mediators. The causal assumptions typically required for causal mediation analyses are as follows:

Assumption A1 *No unmeasured treatment-outcome confounders, i.e., no unmeasured variables that influence both the treatment and the outcome.*

Assumption A2 *No unmeasured treatment-mediator confounders, i.e., no unmeasured variables that influence both the treatment and the mediator.*

Assumption A3 *No unmeasured mediator-outcome confounders, i.e., no unmeasured variables that influence both the mediator and the outcome.*

Assumption A4 *No mediator-outcome confounders (measured or unmeasured) that are influenced by the treatment.*

Assumption A5 *No interaction between the treatment and mediator.*

Assumption A6 *Mediation effect is not influenced by moderators.*

Assumption A7 *No hidden variation (multiple versions) of treatment or mediators.*

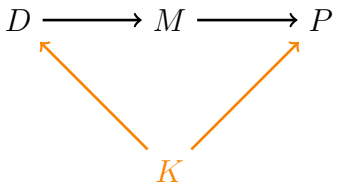
Assumption A8 *No interference among units.*

Assumption A9 *Treatment temporally precedes the mediator, and mediator temporally precedes the outcome (i.e., no reverse causality).*

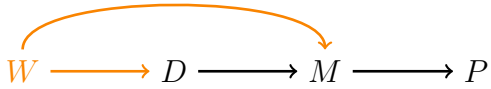
Assumptions A1 to A4 address confounding variables that can introduce bias in mediation analysis (Figure 3), while Assumptions A5 to A8 address other factors that can introduce bias and create challenges for interpretation of mediation effects.

As we describe in subsequent sections, researchers interested in quantifying mediation effects must find ways to satisfy these causal assumptions or relax them. For example, in an experimental design in which the treatment is randomised, such as in our hypothetical drought study, Assumptions A1 and A2 can be satisfied by statistical theory. In the absence

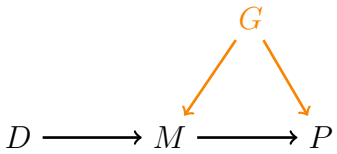
of treatment randomisation, we would need to explicitly account for all treatment-outcome and treatment-mediator confounders, increasing the challenge of estimating mediation effects without bias (Imai et al., 2010; Pearl, 2014; VanderWeele, 2015). However, randomisation of the treatment does not imply that Assumptions A3 or A4 are satisfied (for a detailed explanation of the intuition for this claim using our drought study, see Supplement S.2). To satisfy Assumptions A3 and A4 in an experiment in which the treatment was randomized, researchers must either measure the confounders or eliminate their effects through specific research designs or statistical techniques (Sections 5.1 to 5.4). If violations to Assumptions A3 or A4 are still suspected in a study, researchers should quantify how robust the estimated effects are to such violations (Section 5.5). If Assumptions A5 or A6 cannot be satisfied in a study, researchers may have to use a conceptual framework for causal inferences (Section 6) or change their definitions of indirect and direct mediation effects (Supplement S.6).



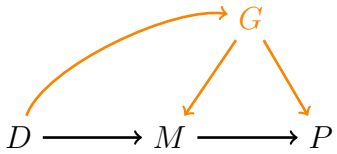
(a) Assumption 1: The treatment-outcome confounder K cannot be unmeasured.



(b) Assumption 2: The treatment-mediator confounder W cannot be unmeasured.



(c) Assumption 3: The mediator-outcome confounder G cannot be unmeasured.



(d) Assumption 4: The mediator-outcome confounder G cannot be influenced by the treatment D, regardless of whether G is measured or unmeasured.

Figure 3: Causal diagrams illustrating four causal assumptions related to confounding variables that could exist in our hypothetical drought study. Labels are as in Figure 2. The confounder addressed by each assumption is shown in orange.

Causal assumptions are distinct from statistical assumptions, which permit valid population-level statistical inferences from available sample data (Berry, 1993). Examples include assumptions about the distribution of the data, the forms of the relationships between measured variables, and properties of the residuals. We generalise these statistical assumptions as follows:

Assumption B1 *Model is correctly specified (e.g., parametric, non-parametric, additive).*

Assumption B2 *Model-specific assumptions are satisfied (e.g., linearity, constant variance, independent errors, normality).*

Statistical assumptions are theoretically valid with sufficiently large data, and much work has gone into developing methods to obtain valid inference in the presence of violations to many common statistical assumptions (Wilcox, 2010).

Unlike statistical assumptions, causal assumptions cannot be expressed using probability calculus, and they cannot be verified without extensive experimental controls, even with unlimited data, because these assumptions reflect conceptual beliefs about unobserved, and therefore unmeasured, variables (Pearl, 2001a; Stone, 1993). Thus, determining whether causal assumptions have been satisfied is subjective, and their plausibility in a specific context is ascertained by a mix of theory, field knowledge, and indirect tests.

Without an explicit description and justification of the causal assumptions on which a study relies, the scientific community cannot assess the credibility of any causal claims in the study. Researchers can quantify how robust their estimated effects are to violations of Assumptions A1 to A4 (Section 5.5). If researchers are unable to satisfy Assumptions A5 to A8, such as in studies with heterogeneous treatment effects and interactions between the treatment and mediator, alternative definitions of direct and indirect mediation effects are available using conceptual frameworks for causal inferences (see Section 6). But, just like statistical assumptions, causal assumptions cannot be ignored. In the next section, we presume Assumptions A1 and A2 are satisfied (e.g., via randomisation, as in our drought experiment example), and we explore ways in which ecologists can satisfy Assumptions A3 and A4 and assess the robustness of the estimated mediation effects to violations of these assumptions. In Section 6, we introduce the potential outcomes causal inference framework that can help ecologists address potential violations to Assumptions A5 to A8. Overcoming violations to temporal precedence is fundamentally difficult (Pearl and Verma, 1995); thus, we presume Assumption A9 can be met for all mediation analyses discussed herein.

5 Addressing mediator-outcome confounders

In studies where Assumptions A1 and A2 have been satisfied, researchers must also eliminate the effects of mediator-outcome confounders to estimate mediation effects without bias (James and Brett, 1984). For example, consider again our hypothetical drought study, but imagine that, prior to the experimental stage, some plots experienced heavy grazing by herbivores while other plots had little to no grazing activity (Figure 4). Suppose that plots with historically more grazing are also, on average, less productive and have less soil moisture in the current period, perhaps through soil compaction by grazers (Eldridge et al., 2017; Sitters and Olde Venterink, 2015; Veldhuis et al., 2014). The correlation of historical grazing with both soil moisture and productivity introduces bias into the estimation of the effect of drought on productivity and the effect of soil moisture on productivity (see Supplement S.2 for details). Thus, when the assumption of no unmeasured mediator-outcome confounding is violated, estimated mediation effects cannot be imbued with causal interpretations, even in experimental designs in which the treatment is randomised (Holland, 1988; MacKinnon, 2012; VanderWeele and Vansteelandt, 2009). The assumption of no unmeasured mediator-outcome confounding is typically not explicitly stated or interrogated in ecological studies, and this assumption is likely violated in practice.

In this section, we describe approaches that can address mediator-outcome confounders

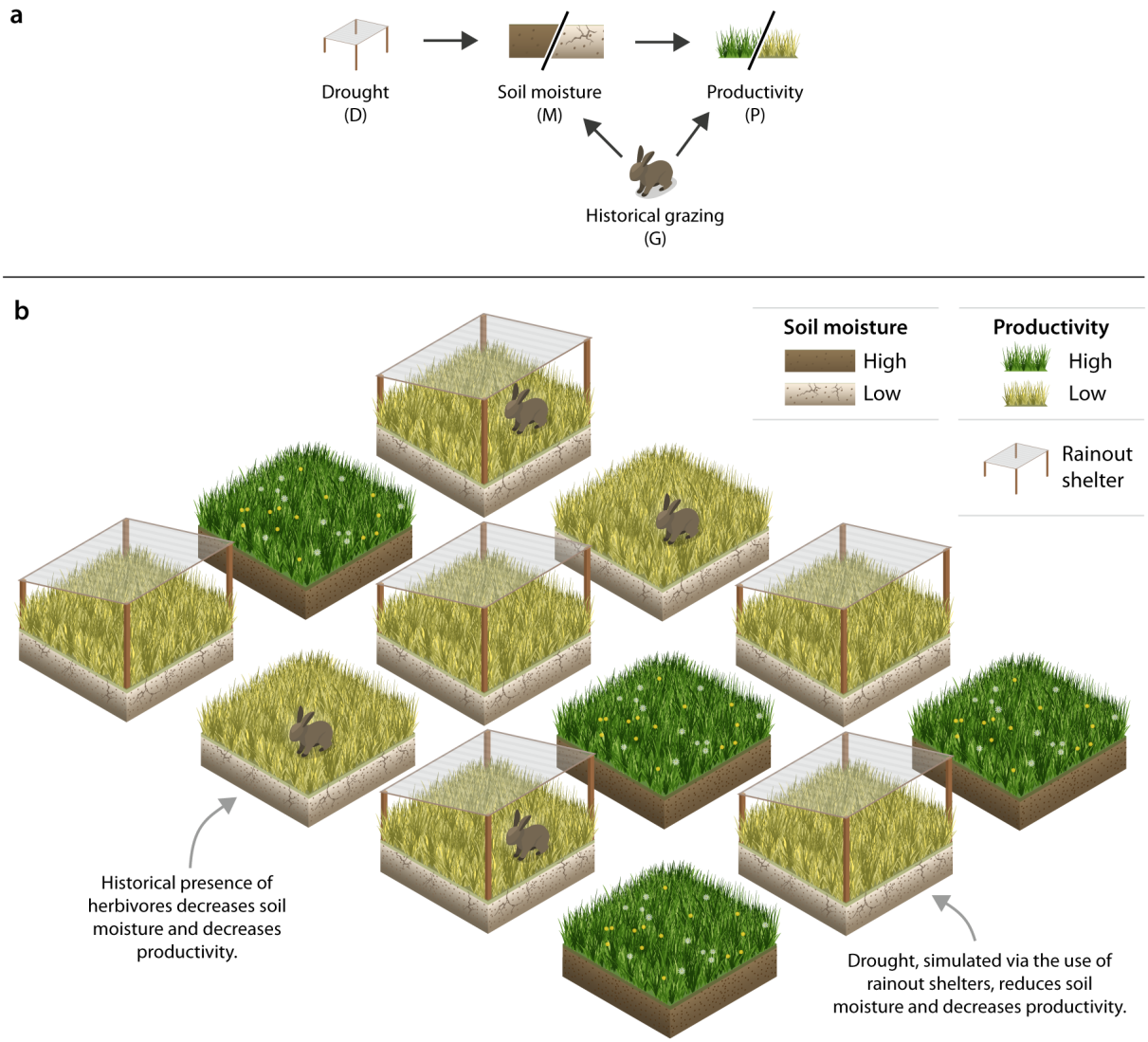


Figure 4: (a) A revised causal DAG of the hypothesis for our hypothetical drought study with the addition of a mediator-outcome confounder, historical grazing. For visual simplicity, the continuous variables soil moisture and productivity are represented as binary. The effect of historical grazing cannot be eliminated through randomisation of the drought treatment. (b) Results from the hypothetical experiment on 12 grassland plots, where 6 plots have been randomly assigned treatment with a rainout shelter. The historical presence of herbivores also reduces soil moisture through compaction of substrate and reduces productivity through grazing. Historical grazing is not manipulated or randomised, but it could be measured during the experimental phase: herbivores grazed on four of the plots, with no preference towards treated or control plots (as expected from randomisation of the rainout shelters).

and can be implemented using linear regression models. For each approach, we describe when and how it can mitigate the effects of mediator-outcome confounders and the challenges faced in implementing the approach.

5.1 Experimental manipulation of mediators

One way to eliminate the effects of mediator-outcome confounders is to randomise the mediator in an experimental design, i.e., a "manipulation-of-mediator" design (Carnevale et al., 1988; Pirlott and MacKinnon, 2016). Although these designs are less common in ecology, there are some examples of ecological experiments that randomised a suspected mediator. For instance, to quantify how productivity reduces plant species richness through shading, studies have manipulated ground light availability directly (Eskelinen et al., 2022; Hautier et al., 2009).

In manipulation-of-mediator approaches, direct manipulation of the mediator typically requires at least two experiments with staged manipulations of the treatment and mediator to separate the treatment's effect from the mediator's effect on the outcome (Imai et al., 2013; Pirlott and MacKinnon, 2016). For example, the double randomisation design splits the sample into two subsamples, randomises treatment assignment while measuring the mediator and outcome in the first subsample, then randomises assignment to different mediator values in the second subsample (Pirlott and MacKinnon, 2016). Other experimental designs that manipulate the mediator are also available, such as parallel designs, cross-over designs, and blockage and enhancement manipulation designs (Jacoby and Sassenberg, 2011; Pirlott and MacKinnon, 2016). These designs provide experimental design-based solutions for ecologists interested in quantifying mediating effects in a wide range of contexts.

While manipulation-of-mediator designs eliminate mediator-outcome confounders, other considerations must be addressed in these designs to be able to estimate mediation effects (Bullock et al., 2010). Choosing meaningful values for manipulating the mediator in a way that accurately represents natural changes in the mediator as caused by the treatment can prove difficult. Additionally, manipulating the mediator, if it is indeed a process or consequence of the treatment, requires either the treatment to be manipulated or the manipulation of another cause of the mediator to induce a change in the mediator. For example, in our hypothetical drought study, inducing values of soil moisture that occur when drought is present ($D = 1$) in plots that are assigned to the no-drought condition ($D = 0$) may be impossible without manipulating another causal factor, say Z, to induce changes in soil moisture.

Manipulation-of-mediator designs also create challenges for quantifying the effects of the treatment and mediator on an outcome. Experimental manipulation of a mediator can affect the outcome in ways that are undesirable for capturing the effect of treatment on outcome through the mediator (Bullock et al., 2010), leading to difficulty in separating the direct and indirect effects of treatment on the outcome (Imai et al., 2010). Returning to our hypothetical drought study, if Z is manipulated for drought-absent ($D = 0$) plots to obtain values of soil moisture that would occur in drought-treated ($D = 1$) plots without actually changing drought (D), then productivity under $D = 0$ is likely no longer being influenced by changes in D, producing misleading estimates of indirect effects through the mediator. Thus, directly manipulating the mediator may result in violations to the causal assumption of no multiple versions of the treatment (Assumption A7; Kimmel et al. 2021). It may therefore

be preferable to encourage or discourage experimental units to take on particular mediator values, resulting in imperfect manipulation of the mediator that can still be informative. Such designs include parallel encouragement designs and crossover encouragement designs (Imai et al., 2013; Pirlott and MacKinnon, 2016).

Even if researchers could address the quantification and interpretation challenges of manipulation-of-mediator designs, mediating variables in ecology are often ecological processes that are difficult to manipulate. For instance, carbohydrate reserves are a hypothesised mediator of drought's effect on tree mortality (Adams et al., 2017); and local adaptation and functional diversity are hypothesised mediators of biodiversity's effect on productivity in decomposers (Keiser et al., 2014). Carbohydrate reserves, local adaptation in decomposers, and decomposers' functional diversity are challenging ecological variables to directly manipulate. Thus, many ecological experiments are similar to our hypothetical drought experiment in which the mediator is not randomised but instead measured for each plot (i.e., a "measurement-of-mediator" designs, Spencer et al. 2005). In the next four subsections, we explore approaches to either eliminate the effects of mediator-outcome confounders in measurement-of-mediator designs or quantify the degree to which mediation effects would change if the effects of all mediator-outcome confounders have not been eliminated in a study.

5.2 Measured mediator-outcome confounders

In the absence of experimental manipulation of the mediator, a researcher must eliminate the effects of mediator-outcome confounders through other means. In our hypothetical drought study, we assume that historical grazing (G) is a mediator-outcome confounder that influences both soil moisture and productivity (Figure 4). If historical grazing had been measured for each of the plots, we would estimate the mediation effects using the following three equations:

$$P_i = \beta_0 + \beta_1 D_i + \varepsilon_{i1} \tag{1}$$

$$M_i = \theta_0 + \theta_1 D_i + \varepsilon_{i2} \ , \tag{2}$$

$$P_i = \delta_0 + \delta_1 D_i + \delta_2 M_i + \delta_3 G_i + \varepsilon_{i3} \ , \qquad i = 1, \ldots, n \ , \tag{3}$$

where D_i is the treatment assigned to plot i; P_i is the plot-level productivity; M_i is the plot-level soil moisture; G_i is the amount of historical grazing on plot i; β_0, θ_0, and δ_0 are intercepts; β_1, θ_1, δ_1, δ_2, and δ_3 are coefficients; and ε_{i1}, ε_{i2}, and ε_{i3} are plot-level error terms (e.g., ε_{i3} represents all other plot-level variation not accounted for by drought, soil moisture, or historical grazing). The average productivity of all plots under the no-drought control is represented by β_0, while β_1 represents the average change in productivity across all plots when going from the control state ($D = 0$) to the drought-treated state ($D = 1$).

Some mediation studies in ecology use only Equations (1) and (2) to estimate a dependence between the treatment and the outcome and between the treatment and the mediator, respectively. If the dependencies are statistically significant, the studies claim to have detected a mediator in the system (Borer et al., 2014; Cadotte, 2017; Fornara and Tilman, 2009; Liu et al., 2018; Oliveira et al., 2022; Tian et al., 2016). This "two-part estimation approach" has two important limitations: (1) the indirect effect cannot be quantified, i.e., researchers cannot estimate the proportion of the effect of drought on productivity that

is mediated by soil moisture; and (2) multiple conclusions can be drawn from the results, including a conclusion that the hypothesised mediator plays no mediating role at all (see Supplement S.1 for details).

By including historical grazing in a regression equation of productivity as a function of both the treatment and mediator (Equation (3)), we eliminate the part of the effect of soil moisture on productivity that is due to the correlation with historical grazing (Figure 5a). If we further assume that that no other mediator-outcome confounders exist (Assumption A3), then Equation (3) will produce estimates of both δ_1 and δ_2 without bias. If the estimated total effect of drought on productivity is negative ($\widehat{\beta}_1 < 0$) then drought reduces productivity on average across plots (Figure 5b). If the estimated effect of drought on soil moisture is negative ($\widehat{\theta}_1 < 0$) and the estimated effect of drought on productivity increases when both soil moisture and historical grazing are included in the model ($\widehat{\delta}_1 > \widehat{\beta}_1$), then drought reduces productivity by reducing soil moisture on average. In other words, after controlling for the mediator-outcome confounder (historical grazing), the negative effect of drought on productivity is smaller in magnitude (i.e., smaller in absolute value) when the effect of soil moisture on productivity is held constant. This procedure is characteristic of analyses using SEMs in ecology (e.g., Grace et al. 2016), although such analyses are not typically framed in these terms. To estimate the effect of drought on productivity through soil moisture using Equations (2) and (3), researchers can use the product method, in which the indirect effect is $\theta_1\delta_2$ (see Supplement S.2 for details and for indirect effects defined using the three-part procedure when the mediator and outcome are not continuous).

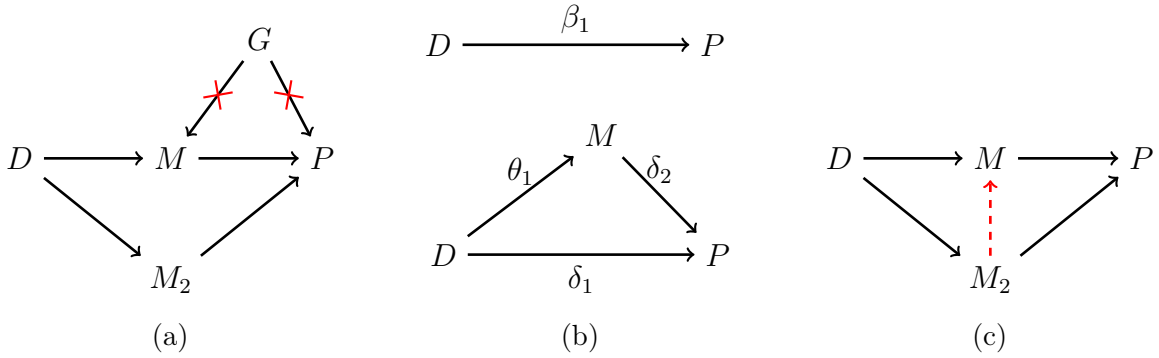


Figure 5: Mediation analysis of the hypothetical drought study is subject to bias arising from confounders. (a) If a mediator-outcome confounder exists, such as historical grazing G, and is measured in the study, bias from G can be eliminated by including the variable as in Equation (3).(b) The three-part procedure estimates four components of the relationship between D and P. (c) The procedure assumes no mediator-outcome confounders, but the effect of drought can operate through other mediators, such as M_2, in addition to soil moisture. However, M must not be affected by any other mediators; e.g., M_2 becomes a mediator-outcome confounder that is influenced by the treatment if the dashed red path exists (a violation of Assumption A4). Labels are as in Figure 2.

Regardless of how the indirect effect is quantified, the effect is only estimated without bias if all mediator-outcome confounders are accounted for (Figures 5a and 5c) and if the effect of soil moisture on productivity is homogeneous across different levels of drought, i.e., there

is no interaction between drought and soil moisture (Valeri and Vanderweele, 2013). For a detailed explanation of the bias that arises in the presence of heterogeneous effects using the hypothetical drought study, see Supplement S.3, but see Section 6 for options to relax Assumption A5. The assumption that a study's research design has controlled for all possible confounders is a strong assumption that is unstated in many mediation analyses (Bollen and Pearl, 2013; Grace et al., 2015; Kunicki et al., 2023; VanderWeele, 2012b; VanderWeele and Rothman, 2021).

In real ecological systems, there will likely be many mediator-outcome confounders, and identifying and measuring them all will be challenging. Additionally, many confounders (e.g., historical grazing patterns, weather, soil composition) are multi-dimensional, and measuring the relevant dimensions can be challenging. In the next three subsections, we describe approaches for mediation analysis that do not rely on measuring every potential mediator-outcome confounder in all their relevant dimensions.

5.3 Unmeasured mediator-outcome confounders: instrumental variable designs

Suppose that the mediator-outcome confounder historical grazing cannot be measured in our hypothetical drought experiment. Suppose also that there exists another variable that affects productivity only through its effect on soil moisture (V in Figure 6), i.e., there exist no other pathways from V to productivity except through the effect of V on soil moisture. When measured, this variable can be used as an instrumental variable to estimate the effect of the mediator without bias, even in the presence of mediator-outcome confounders. If V only affects productivity through its effect on soil moisture (Figure 6a), an untestable causal assumption known as the "exclusion restriction", we can replace Equations (2) and (3) with

$$M_i = \theta_0 + \theta_1 D_i + \theta_2 V_i + \varepsilon_{i2} \tag{4}$$

$$P_i = \delta_0 + \delta_1 D_i + \delta_2 \widehat{M}_i + \varepsilon_{i3} \tag{5}$$

where V_i is the instrumental variable measured at each plot i and $\widehat{M}_i$ is the fitted value of soil moisture estimated from Equation (4) (Chen et al., 2023; Dippel et al., 2020). As in Section 5.2, researchers can use the product method to estimate the indirect effect from Equations (4) and (5) as $\theta_1\delta_2$. If the exclusion restriction assumption is violated (Figure 6b), one cannot estimate the effect of soil moisture on productivity, δ_2, without bias using Equations (4) and (5).

Finding and measuring instrumental variables that do not violate the exclusion restriction is challenging in ecological systems (Grace, 2021; Kendall, 2015; Rinella et al., 2020), although, in some cases, the assumption can be made more plausible after eliminating the effects of measured confounders (Section 5.2). Furthermore, instrumental variable designs have interpretation challenges: unless the average effect of the mediator is constant across units, we can only estimate the indirect effect for a subgroup of plots (Angrist and Imbens, 1995; Frölich and Huber, 2017; Rudolph et al., 2021; Wang and Tchetgen Tchetgen, 2018).

Figure 6: Causal diagrams illustrating instrumental variable designs for mediation analysis. (a) In the presence of the unmeasured mediator-outcome confounder G, an instrumental variable V can be leveraged to estimate the effects of D on P that occur through M. (b) V is not a valid instrumental variable if it affects P through any other pathways (a violation of the exclusion restriction). Labels D, M, P, and G are as defined in Figure 2.

5.4 Unmeasured mediator-outcome confounders: longitudinal data designs

The effects of unmeasured mediator-outcome confounders can also be eliminated if clustered longitudinal data on soil moisture and productivity have been collected. By "clustered" longitudinal data, we mean data on productivity and soil moisture from $i = 1, \ldots, n$ plots clustered within multiple sites $s = 1, \ldots, S$ and measured across multiple time points $t = 1, \ldots, T$ both before and after the drought treatment is randomly assigned (Figure 7). In a randomised experiment, data from time points before random assignment of the treatment are not necessary to estimate the effect of drought on productivity without bias, but such data can be helpful for estimating the effects of a mediator by eliminating the effects of unobserved mediator-outcome confounders. The benefits of collecting such data for mediation studies in ecology will need to be balanced with the increased time and expense required for data collection.

Below, we describe two popular approaches for eliminating mediator-outcome confounding effects: multilevel modelling and autoregressive modelling designs. For a review of additional approaches to leveraging clustered longitudinal designs for causal inference, see Wooldridge (2010). As we will show, valid inference from clustered longitudinal data designs requires additional attention to correctly modelling the structure of the data (e.g., serial correlation of the errors).

5.4.1 Multilevel modelling approach

Ecologists often analyse clustered longitudinal data using a multilevel model structure, which captures the groupings of clustered data by specifying at least two levels of equations: (1) first level equations which model the observation-level data (e.g., productivity on each plot at each time period); and (2) higher-level equations, which include sets of equations for each cluster or grouping (e.g., productivity on each plot averaged over all time periods) (Gelman and Hill, 2006). Modelling clustered longitudinal data with the classical multilevel

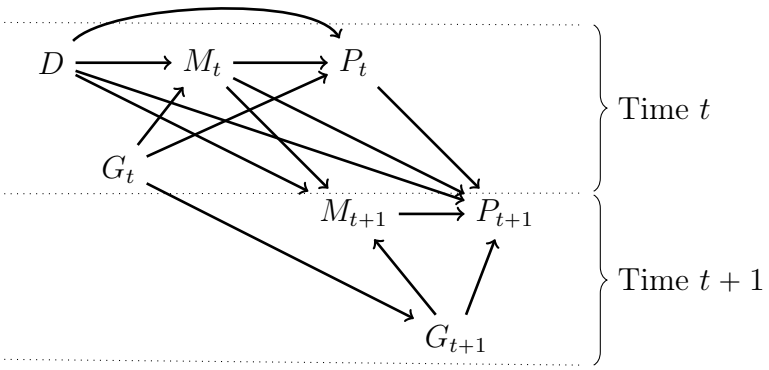


Figure 7: Causal diagram for a longitudinal version of the hypothetical drought study for plot i at site s. For simplicity, time is represented by two periods: t and $t+1$. The diagram can be extended to include all times $t = 1, \ldots, T$. G_t is the unmeasured mediator-outcome confounder at time t, and G_{t+1} is the same unmeasured mediator-outcome confounder at the next time point $t+1$. All other labels are as defined in Figure 2.

structure, which is often referred to as mixed effects modelling in ecology, includes error terms in each of the higher-level equations and allows researchers to quantify the variation within and among various groupings (Bolker et al., 2009). To use mixed effects modelling to estimate mediation effects without bias, researchers must assume that the unmeasured differences in the outcome among plots or among sites, including differences that arise from the effects of confounders, are uncorrelated with the model's predictors (i.e., the treatment and mediator) (Gelman, 2006; Seber and Lee, 2003). Even in ecological settings where the treatment is randomised, this assumption is likely violated. For a discussion on how bias arises in estimating mediation effects using mixed effects modelling for the hypothetical drought experiment, see Supplement S.4.

An alternative multilevel modelling approach can accommodate correlations between unmeasured differences among groups and predictors in the model. This approach, sometimes called the Mundlak regression approach (Mundlak, 1978) or multilevel modelling for causal inference (Gelman and Hill, 2006), adds group-averaged predictors from the observation-level equations as predictors in the higher-level equations (Gelman, 2006; Gelman and Hill, 2006). These group-averaged predictors remove the effect of unmeasured plot-level and site-level confounding variables that do not vary over time or change very slowly. As explained below, the clustered structure of the data can also be used to eliminate unmeasured confounders that vary over time (see also Byrnes and Dee, 2024).

To implement the multilevel approach for our hypothetical drought study, we include intercepts at the plot-level and the site-time group level to account for unmeasured confounding at both levels. We provide the full set of multilevel equations in Supplement S.4, but the primary difference between a traditional mixed effects modelling approach and a multilevel modelling approach for causal inference for our hypothetical drought study lies in the inclusion of plot-averaged and site-time-averaged soil moisture terms in the higher level equations. Recall that in the clustered longitudinal version of our drought study, a plot i is observed at multiple time points $t = 1, \ldots, T$. We represent an individual observation on plot i at time t as an observation h. Thus, for an observation h measured at time t

and belonging to plot i within site s, we describe the effect of drought and soil moisture on productivity as

$$P_h = \phi_{3,i[h]} + \mu_{3,st[h]} + \delta_1 D_h + \delta_2 M_h + \varepsilon_{3,h} \,, \quad i = 1, \ldots, n \,; s = 1, \ldots, S \,; t = 1, \ldots, T \,; h = 1, \ldots, nST \tag{6}$$

where each site is composed of n_s plots, for a total of $n = n_1 + n_2 + \cdots + n_s$ plots, and each plot is repeatedly measured over T time points; $i[h]$ is the plot i containing observation h; $st[h]$ is the site-time group containing h; P_h, D_h, and M_h are the productivity, drought, and soil moisture values measured for an observation h; δ_1 and δ_2 represent the effects of drought and soil moisture on productivity; $\phi_{3,i[h]}$ is the plot-level intercept; $\mu_{3,st[h]}$ is the site-time group-level intercept; and $\varepsilon_{3,h}$ is the error term.

To eliminate the effects of unmeasured mediator-outcome confounders, we must specify second-level equations for Equation (6) that include group-averaged soil moisture as predictors of the group-level intercepts. These equations are

$$\phi_{3,i} = \phi_{3\cdot} + \nu \overline{M}_i + \eta_{3,i} \tag{7}$$

$$\mu_{3,st} = \mu_{3\cdot} + \kappa \overline{M}_{st} + \eta_{3,st} \tag{8}$$

where $\phi_{3\cdot}$ is the average of the plot-varying intercepts $\phi_{3,i[h]}$; $\mu_{3\cdot}$ is the average of the site-time group-varying intercepts $\mu_{3,st[h]}$; ν is the coefficient for the predictor $\overline{M}_i$ representing plot-level averages of soil moisture; κ is the coefficient for the predictor $\overline{M}_{st}$ representing the site-time grouped means of soil moisture; $\eta_{3,i}$ is the plot-level error; and $\eta_{3,st}$ is the site-time group-level error. Researchers can again use the product method to estimate the indirect effect as $\theta_1 \delta_2$ (see Supplement S.4 for details).

Including plot-level effects $\phi_{i[h]}$, which are intercepts estimated for each plot i in site s where the plot-level differences over time are averaged, allows us to account for unmeasured differences between plots that do not change over time, such as differences associated with unmeasured mediator-outcome confounders that occur at the plot level. Likewise, including site-time group-level effects $\mu_{st[h]}$, which are intercepts estimated for each site-time group st where the differences across plots at each site and time point are averaged, allows us to account for unmeasured differences between sites that change over time, including differences associated with unmeasured mediator-outcome confounders that vary over time at the site level but do not vary across plots within the same site. Further, including the plot-level ($\overline{M}_i$) and site-time group-level ($\overline{M}_{st}$) averages of the mediator in the higher-level equations eliminates any potential correlation between soil moisture and the plot or site-time groupings. As long as plot-level, time-varying confounders (e.g., micro-climate) do not exist, or they are observed and included in the multilevel model, then η_i and η_{st} are not correlated with soil moisture, and the assumption of independence between the levels or groupings (i.e., plot and site-time) and the mediator in the model is not violated (Greenland and Robins, 1985; Robins et al., 2000; Roth and MacKinnon, 2012). Longitudinal data can also be used to control for unmeasured, plot-level confounding variables that vary over time, but we do not consider those methods here (Greenland and Robins, 1985; Roth and MacKinnon, 2012).

In addition to assuming that time-varying, plot-level confounding variables are observed or do not exist, the multilevel modelling approach also requires three additional assumptions:

(1) linearity and additivity of the effects, (2) the effects of the treatment and mediator do not change across groupings or over time, and (3) the outcome variable for the treated and control plots would have the same mean trend over time in the absence of treatment, conditional on ϕ_i and μ_{st} (called the parallel trend assumption; Imai and Kim 2021). These assumptions, particularly the parallel trends assumption, may not hold in long-term ecological experiments. More recent advances for multilevel models provide options for relaxing the assumptions of linearity (Imai and Kim, 2019), homogeneous treatment effects (de Chaisemartin and D'Haultfœuille, 2020), and parallel trends (Rüttenauer and Ludwig, 2023).

5.4.2 Autoregressive approach

The multilevel modelling approach described in Section 5.4.1 assumes that the mediator-outcome confounders are unchanging attributes of the system or time-varying site-level attributes. Alternative approaches to modelling clustered longitudinal data require alternative assumptions about the potential sources of confounding. For example, autoregressive models with fixed effects, sometimes called "dynamic panel models" in econometrics (Arellano and Bond, 1991; Blundell and Bond, 1998), can be used if the most likely sources of confounding are time-varying, plot-level attributes that are correlated with values of the outcome variable at previous time points (e.g., prior values of productivity affect current values of soil moisture). Autoregressive models with fixed effects can incorporate lagged effects and between-cluster effects over time, but like all approaches to mediation analysis, they rely on untestable causal assumptions (Bellemare et al., 2017). Some of these assumptions can be relaxed when these models are used within the SEM setting (Allison et al., 2017), but no autoregressive approach can address all potential sources of mediator-outcome confounders simultaneously.

5.5 Sensitivity analyses for unmeasured mediator-outcome confounders

The assumption of no unmeasured mediator-outcome confounders (Assumption A3) is not verifiable using data, but researchers can quantify their uncertainty over potential violations of the assumption by drawing on a range of recent advances to (1) explore how the results change after using multiple estimation approaches that rely on different causal assumptions about the nature of mediator-outcome confounders (e.g., compare the estimated mediation effects from an instrumental variable design with the estimates from a multilevel model); or (2) assess the degree to which the sign or magnitude of the estimated effects could change if the assumption of no unmeasured mediator-outcome confounders is violated. Sensitivity analyses explore how much the estimated mediation effects can change in the presence of a specific source of confounding (Ding and VanderWeele, 2016; Imai et al., 2010; Hong et al., 2018; Sullivan and VanderWeele, 2021; VanderWeele, 2010). In contrast, partial identification approaches estimate mediation effects under the least restrictive or weakest causal assumptions to obtain the widest bounds for each effect and then explore how the bounds shrink as the causal assumptions are strengthened (Flores and Flores-Lagunes, 2013; Huber, 2020; Miles et al., 2017; Richardson et al., 2014).

The assessment of the sensitivity of estimated mediation effects to potential violations in the causal assumptions is an important step in mediation analyses (MacKinnon and Pirlott,

2015; VanderWeele, 2015). Causal assumptions are almost certainly violated to some degree in most real-world systems. Rather than discard causal analyses altogether, every mediation study should be supplemented by analyses that assess the implications of potential violations to causal assumptions (Hafeman, 2011; Imai et al., 2010; MacKinnon and Pirlott, 2015; Tchetgen Tchetgen and Shpitser, 2012; VanderWeele and Ding, 2017). Such analyses allow researchers to evaluate their level of confidence for causal claims and provide avenues for addressing gaps in satisfying causal assumptions in future studies.

6 Addressing other causal assumptions: causal inference frameworks for mediation analysis

In this section, we introduce the potential outcomes causal inference framework, which researchers can use to define and estimate direct and indirect effects that systematically incorporate the complexities that we ignored in Section 5. These complexities include heterogeneous mediation effects and interference among units (i.e., violations of causal assumptions A5 to A8) as well as conditions such as nonlinearity (i.e., violations of statistical assumptions assumptions B1 and B2). The potential outcomes framework is one of several well-developed causal inference frameworks for mediation analysis and is commonly employed in epidemiology, behavioural sciences, econometrics, and public health. Causal inference frameworks provide clearly defined terminology for the roles that key variables play in an ecological system and supply a language to describe the relationships between these variables. Without a formal causal inference framework, the assumptions and interpretations of any analyses that aim to estimate causal effects from data are opaque and difficult to evaluate or reproduce (Ferraro and Hanauer, 2015). The potential outcomes framework allows researchers to define direct and indirect effects in the absence of any parametric assumptions about the data or specific functional forms that describe the relationships between variables, and it also allows researchers to decompose total effects into interpretable components under conditions in which some of the causal assumptions in Section 4 are not satisfied. For example, when mediation effects are heterogeneous because of treatment-mediator interactions or mediator-mediator interactions, the potential outcomes framework illustrates how one can decompose and separate the contributions of the interactions and the mediation to the total effect (see Supplement S.6 for details).

Using our hypothetical drought study, we introduce the potential outcomes notation for direct and indirect effects (also called "counterfactuals" notation). Recall that we are interested in measuring the effect of drought on productivity while considering the mediating effect of soil moisture. A plot can potentially be under the drought-treated condition, $D = 1$, or the no-drought control condition, $D = 0$. Imagine that researchers assigned a plot to the control condition and recorded the productivity after some time. At the same time in a parallel world in which all other conditions are identical, the same researchers assigned the same plot to the drought-treated condition instead and recorded the productivity. If they were able to monitor both worlds simultaneously, the researchers would have a measure of productivity for the same plot under both the control condition, which we can define as the plot's potential outcome P_0, and under the treated condition, which we can define as the plot's potential outcome P_1. The difference in productivity between the two potential states

of the same plot is the total effect (TE) of drought on productivity in that plot:

$$TE = P_1 - P_0 \ . \tag{9}$$

In the potential outcomes framework, the total effect can be decomposed into two components: one that represents the indirect effect of drought on productivity through soil moisture, and another that represents the effect of drought on productivity that goes through other mediators that are not the focus of our hypothetical drought study (Robins and Greenland, 1992; VanderWeele, 2014). Continuing with our parallel worlds thought experiment, we define two potential outcomes for the mediator: M_0 is the potential value that soil moisture would take in the plot's no-drought control condition ($D = 0$), while M_1 is the potential value that soil moisture would take in the same plot's drought-treated condition ($D = 1$). Thus, the plot has four potential outcomes: P_{1M_1}, P_{1M_0}, P_{0M_1}, and P_{0M_0} (e.g., P_{1M_0} is the plot's productivity in the drought-treated condition with soil moisture held to its values in the no-drought control condition).

The effect of drought on productivity through soil moisture is represented by the **total indirect effect** (TIE), which describes the amount by which productivity would change in a plot if drought were fixed at $D = 1$ and soil moisture changed from the value it would be at $D = 0$ to the value it would be at $D = 1$,

$$TIE = P_{1M_1} - P_{1M_0} \ . \tag{10}$$

The remaining effect of drought on productivity that does not go through soil moisture, the **pure direct effect** (PDE), describes how much productivity would change if drought were changed from $D = 0$ to $D = 1$ and soil moisture were kept at the value it would have been when $D = 0$ (i.e., M_0),

$$PDE = P_{1M_0} - P_{0M_0} \ . \tag{11}$$

Although we can imagine parallel worlds and define these effects in terms of potential outcomes, in our one world, we cannot observe the same plot under both the treated condition and the control condition simultaneously. This dilemma is known as the "fundamental problem" of causal inference (Holland, 1986). For a treated plot, we can observe only one of the potential outcomes – the potential outcome under the drought-treated condition ($P_{1M_1} = P_1$). We cannot observe the potential outcomes of the treated plot as it would be under control conditions (P_{1M_0}, P_{0M_1}, or P_{0M_0}). These are counterfactual potential outcomes (counter to fact). Similarly, for a control plot, we can only observe one potential outcome ($P_{0M_0} = P_0$). We cannot observe the counterfactual potential outcomes P_{0M_1}, P_{1M_0}, or P_{1M_1}. Thus, the individual plot-level causal effects in Equations (9) to (11) cannot be estimated.

While we cannot observe all potential outcomes for a plot in our drought experiment, we can combine the potential outcomes framework with statistical theory and assumptions to obtain from data a population-level approximation of our hypothetical parallel worlds (VanderWeele, 2015). When the treatment is completely randomised, the observed average productivity of the plots under the control condition provides an estimate of the population-level productivity had all plots been under the control condition, i.e., $E[P_0]$, where $E[\cdot]$ is the expectation operator. Similarly, the average productivity of the plots under the drought-treated condition provides an estimate of the population-level productivity had all plots

been under the drought-treated condition, i.e., $E[P_1]$. The difference between these two quantities provides us with an estimate of the average total effect of drought on productivity when changing from the control condition to the treated condition (sometimes called the "average treatment effect", ATE):

$$\begin{aligned} ATE = E[P_1 - P_0] &= E[P_1] - E[P_0] \\ &= E[P_{1M_1}] - E[P_{0M_0}] \,. \end{aligned} \tag{12}$$

We can also estimate two components of the ATE: the average pure direct effect ($E[P_{1M_1} - P_{1M_0}]$) and average total indirect effect ($E[P_{1M_0} - P_{0M_0}]$), where the ATE is the sum of the average PDE and the average TIE:

$$\begin{aligned} E[P_1 - P_0] &= E[P_{1M_1}] - E[P_{0M_0}] \\ &= (E[P_{1M_1}] - E[P_{1M_0}]) + (E[P_{1M_0}] - E[P_{0M_0}]) \\ &= E[P_{1M_1} - P_{1M_0}] + E[P_{1M_0} - P_{0M_0}] \,. \end{aligned} \tag{13}$$

In our drought study with its binary treatment, we could use Equations (1) to (3) to estimate the ATE, which would be equal to β_1, the average PDE, which would be equal to δ_1, and the average TIE, which would be equal to $\theta_1\delta_2$ (Figure 8), but only if Assumptions A1 to A9 were satisfied and the statistical assumptions of the regression estimators were satisfied (Assumptions B1 and B2). In many ecological systems, however, one or more of these assumptions may not be valid, and, in such cases, a conceptual framework like the potential outcomes framework is valuable for decomposing the total effect into interpretable components and suggesting appropriate estimation procedures.

Figure 8: Mediation effects defined using the potential outcomes framework and the three-part estimation procedure for the hypothetical drought study. The three-part procedure estimates four components of the relationship between D and P. If Assumptions A1 to A9 and Assumptions B1 to B2 are satisfied for regression estimators, then we can use Equations (1) to (3) to estimate the ATE and average PDE and TIE. The estimate of the ATE is β_1, shown in red. The estimate of the average PDE is δ_1, shown in teal. The estimate of the average TIE is $\theta_1\delta_2$, shown in orange. Labels D, M, and P are as in Figure 2.

The causal assumptions of no heterogeneous mediator effects (Assumptions A5 and A6) will be routinely violated in ecological systems. For example, in our drought experiment, the effect of soil moisture on productivity may be functionally different in the presence of drought than in the absence of drought, which would suggest an interaction between the treatment and mediator in violation of Assumption A5 (VanderWeele, 2009; VanderWeele and Robins,

2007). The estimation procedures in Sections 5.1 to 5.4 will not generate estimates of the direct and mediated effects of drought on productivity without bias when treatment-mediator interactions are present, even if both drought and soil moisture were randomised (Bullock et al., 2010; Glynn, 2012; Pearl, 2001b; for a detailed justification, see Supplement S.3). To address treatment-mediator interactions, direct and indirect effects estimators have been developed using traditional regression-based approaches, including SEM (MacKinnon et al., 2020; Rijnhart et al., 2017, 2021; VanderWeele and Vansteelandt, 2010), but these estimators are only valid under certain conditions (e.g., for continuous outcomes and continuous or binary mediators). The potential outcomes framework has been used to develop more general approaches that allow for treatment-mediator interactions and both continuous and non-continuous mediators and outcomes (e.g., Loh et al. 2022, 2020; Xue et al. 2022). For example, in the presence of treatment-mediator interactions, the total effect can be decomposed into four component effects instead of just a PDE and a TIE (VanderWeele, 2014; see Supplement S.6 for details). Moreover, in observational studies or randomised studies with non-compliance, other mediation effects not defined in traditional regression-based approaches may be more plausibly estimated with available data. Causal inference frameworks can help to clearly differentiate these mediation effects from others and suggest appropriate estimation strategies (e.g., Ferraro and Hanauer 2014; see also Supplement S.6 for other mediation effects of potential interest to ecologists).

A key advantage of causal inference frameworks is that they allow researchers to separate the definitions of the mediation effects from the estimation procedures for those effects (Pearl, 2001b; Robins and Greenland, 1992; VanderWeele, 2015). In that way, the relevant assumptions that must be invoked to estimate a particular effect can be transparently evaluated or, when those assumptions are not likely to hold, the study aims can be transparently redefined to focus on more plausible assumptions under which mediation effects can be estimated. For example, mediation effects obtained using the regression-based approaches in Sections 5.2 to 5.4 require assumptions of additivity and linearity. However, direct and indirect effects can be defined for more flexible semi- and non-parametric models. Bootstrapping can be used to nonparametrically estimate direct and indirect effects (Imai et al., 2010) and is particularly useful when the sample size is small or the distribution of the mediator or outcome is non-Gaussian. Semiparametric methods have also been used to estimate direct and indirect effects (Tchetgen Tchetgen, 2011; Tchetgen Tchetgen and Shpitser, 2012), and more recent work has extended these methods to settings with multiple mediators and confounding (Miles et al., 2020; Zhou, 2021). To accommodate nonlinear relationships and interactions between the treatment, mediator and outcome, kernel-based approaches can be used (Carter et al., 2020; Devick et al., 2022; Singh et al., 2022) and have also been applied in SEM settings (Shen et al., 2017). For data with non-Gaussian distributions or nonlinear relationships between treatment, mediator, and outcome variables, Bayesian nonparametric models have been shown to be effective for estimating direct and indirect causal effects (Kim et al., 2017, 2019; Linero and Antonelli, 2023). More recently, machine learning methods have been incorporated into mediation analyses with high-dimensional data to provide a data-driven approach for handling large sets of measured confounders (Farbmacher et al., 2022; Linero and Zhang, 2022; Xu et al., 2022).

The potential outcomes framework is not the only causal inference framework that ecologists could use. Several publications in ecology have promoted various methodologies or

frameworks for causal inference, such as SEMs (Grace, 2006; Grace et al., 2012), structural causal models (SCMs) (Arif and MacNeil, 2022a, 2023; Laubach et al., 2021), and the potential outcomes framework (Clough, 2012; Larsen et al., 2019; Ramsey et al., 2019). These approaches to causal inference, along with the decision theoretic approach to statistical causality (Dawid, 2000, 2003, 2021), are equivalent under identical causal assumptions. For example, SEMs can be expressed mathematically using the *do*-calculus of Pearl (2009) (Bollen, 1989; Mulaik, 2009) and have been shown to be equivalent to SCMs (Pearl, 2009, 2023), the potential outcomes framework (Hernán and Robins, 2006), and the decision theoretic approach to statistical causality (Dawid, 2015). Thus, SEM methodologies with which ecologists may be familiar can be used to estimate mediation effects if the required causal assumptions are transparently described and plausibly satisfied in the analysis (Bollen, 1989; Bollen and Pearl, 2013; Hernán and Robins, 2006; Mulaik, 2009; Pearl, 2009, 2023; VanderWeele, 2012b).

Regardless of the causal inference framework used, the focus of any mediation analysis should be on clearly articulating and satisfying causal assumptions, thereby reducing potential bias that arises from violations of these assumptions (Larsen et al., 2019). Including sensitivity analyses (Section 5.5) in mediation analyses to quantify potential bias from violations to causal assumptions also allows researchers to further assess the plausibility of causal claims made in studies of mediation.

7 Conclusion

Quantifying the effects of intermediary ecological processes is challenging and requires careful attention to study designs, including defining the causal effects to be estimated and explicitly describing the untestable causal assumptions on which causal inferences rely. Those definitions and descriptions allow ecologists to better identify and eliminate rival explanations for observed patterns in data and rigorously explore the implications of potential hidden biases.

Although ecological studies often describe and justify statistical assumptions, they have given less attention to describing and justifying causal assumptions (Section 4). The credibility of these causal assumptions determines the credibility of mediation studies in ecology, regardless of the causal inference framework used (Dawid, 2021; Pearl, 2000; Rubin, 2006).

In our review, we highlighted challenges in quantifying the effects of ecological mediators, but we do not view these challenges as insurmountable. Rather than view these challenges as reasons to avoid making inferences about ecological mediators, we instead view them as reasons for being transparent when making causal claims about mediation and for using more advanced techniques for estimating mediation effects.

To address these challenges and advance the empirical literature on ecological mediators, we described tools and a conceptual framework for causal inferences that emphasise transparency, and we described many of the steps that every empirical mediation study should include (summarised in Table 1). Although we have emphasised how methodological innovations in other fields can contribute to advances in ecology, we also believe that well-executed mediation analyses in ecology have the potential to contribute innovations to other fields. Ecologists' extensive experience in modelling heterogeneous spatial and temporal dynamics, decades of development of mechanistic theories of ecological processes, and vast collections

of field data provide unique opportunities to address challenges of causal inferences for mediation in observational settings and complex systems (Clough, 2012; Larsen et al., 2019; Laubach et al., 2021; Schlüter et al., 2023).

For researchers in ecology to make meaningful contributions to both methodological advancements and ecological theory through the study of mediators, they must carefully consider and explicitly state the causal and statistical assumptions they make when estimating the effects of intermediate ecological processes from data. Clearly communicating the assumptions necessary for valid inferences and examining potential violations to these assumptions are key for providing rigorous and reproducible mediation analyses that explain important intermediary processes in ecology.

Table 1: Essential steps in mediation analysis.

Steps		**Reference**
1.	Define the mediation effect(s) of interest using a conceptual framework for causal inferences.	Section 6
2.	Identify the likely confounding variables using theory and field knowledge, including all hypothesised treatment-outcome, treatment-mediator, and mediator outcome confounders.	Section 3
3.	Pre-register the mediation hypotheses, including how treatments mediators, and moderators will be measured.‡	Kimmel et al. 2023
4.	For each mediation effect of interest, develop a strategy for estimating the effect and mitigating the biases that confounding variables may introduce.	Section 5
5.	Select a mode of statistical inference that is appropriate for the data generating process.	
6.	Assess the presence of treatment-mediator interactions, i.e., heterogeneity.	Section 6
7.	Estimate mediation effects.	
8.	Perform sensitivity analyses of how the estimated effect(s) would change if assumptions A1 to A4 in Section 4 were violated.	Section 5.5
9.	Assess the likelihood that causal assumptions A5 to A8 in Section 4 are violated and discuss the implications of potential violations for the estimation procedures of the interpretation of the estimated effects.	

‡The set of treatments, mediators, and moderators should be kept small given the challenges of satisfying the assumptions in Section 4 for multiple treatments and mediators and the dangers of detecting spurious relationships through multiple comparisons (i.e., data mining).

Acknowledgements

We thank Jarrett Byrnes and his research group for their insightful feedback on the manuscript. We also thank Elisa Van Cleemput and Katherine Siegel for their comments on an early version of this work. Ferraro and Correia acknowledge support from the USDA's National Institute of Food and Agriculture (2019-67023-29854). Dee acknowledges support from an NSF CAREER Grant (2340606).

References

Adams, H. D., Zeppel, M. J. B., Anderegg, W. R. L., Hartmann, H., Landhäusser, S. M., Tissue, D. T., Huxman, T. E., Hudson, P. J., Franz, T. E., Allen, C. D., Anderegg, L. D. L., Barron-Gafford, G. A., Beerling, D. J., Breshears, D. D., Brodribb, T. J., et al. (2017). A multi-species synthesis of physiological mechanisms in drought-induced tree mortality. *Nature Ecology & Evolution*, 1(9):1285–1291.

Addicott, E. T., Fenichel, E. P., Bradford, M. A., Pinsky, M. L., and Wood, S. A. (2022). Toward an improved understanding of causation in the ecological sciences. *Frontiers in Ecology and the Environment*, 20(8):474–480.

Allison, P. D., Williams, R., and Moral-Benito, E. (2017). Maximum likelihood for cross-lagged panel models with fixed effects. *Socius*, 3:2378023117710578.

Angrist, J. D. and Imbens, G. W. (1995). Identification and estimation of local average treatment effects. Working Paper 118, National Bureau of Economic Research.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297.

Arif, S. and MacNeil, M. A. (2022a). Predictive models aren't for causal inference. *Ecology Letters*, 25(8):1741–1745.

Arif, S. and MacNeil, M. A. (2022b). Utilizing causal diagrams across quasi-experimental approaches. *Ecosphere*, 13(4):e4009.

Arif, S. and MacNeil, M. A. (2023). Applying the structural causal model framework for observational causal inference in ecology. *Ecological Monographs*, 93(1):e1554.

Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*, 51(6):1173–1182.

Bellemare, M. F., Masaki, T., and Pepinsky, T. B. (2017). Lagged explanatory variables and the estimation of causal effect. *The Journal of Politics*, 79(3):949–963.

Berry, W. (1993). *Understanding Regression Assumptions*. Quantitative Applications in the Social Sciences. SAGE Publications, Inc., Thousand Oaks, California.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3):127–135.

Bollen, K. and Pearl, J. (2013). Eight myths about causality and structural models. In Morgan, S., editor, *Handbook of Causal Analysis for Social Research*, chapter 15, pages 301–328. Springer.

Bollen, K. A. (1989). *Structural Equation with latent variables*. Wiley, New York, NY.

Borer, E. T., Seabloom, E. W., Gruner, D. S., Harpole, W. S., Hillebrand, H., Lind, E. M., Adler, P. B., Alberti, J., Anderson, T. M., Bakker, J. D., Biederman, L., Blumenthal, D., Brown, C. S., Brudvig, L. A., Buckley, Y. M., et al. (2014). Herbivores and nutrients control grassland plant diversity via light limitation. *Nature*, 508(7497):517–520.

Bullock, J. G., Green, D. P., and Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer. *Journal of Personality and Social Psychology*, 98(4):550–558.

Byrnes, J. E. K. and Dee, L. E. (2024). Causal inference with observational data and unobserved confounding variables. *bioRxiv*.

Cadotte, M. W. (2017). Functional traits explain ecosystem function through opposing mechanisms. *Ecology Letters*, 20(8):989–996.

Carnevale, P. J. D., Harris, K. L., Idaszak, J. R., Henry, R. A., Wittmer, J. M., and Conlon, D. E. (1988). Modeling mediator behavior in experimental games. In Tietz, R., Albers, W., and Selten, R., editors, *Bounded Rational Behavior in Experimental Games and Markets*, pages 160–169, Berlin, Heidelberg. Springer Berlin Heidelberg.

Carter, K. M., Lu, M., Jiang, H., and An, L. (2020). An information-based approach for mediation analysis on high-dimensional metagenomic data. *Frontiers in Genetics*, 11:1–11.

Chen, F., Hu, W., Cai, J., Chen, S., Si, A., Zhang, Y., and Liu, W. (2023). Instrumental variable-based high-dimensional mediation analysis with unmeasured confounders for survival data in the observational epigenetic study. *Frontiers in Genetics*, 14:1092489.

Cinner, J. E., Maire, E., Huchery, C., MacNeil, M. A., Graham, N. A. J., Mora, C., McClanahan, T. R., Barnes, M. L., Kittinger, J. N., Hicks, C. C., D'Agata, S., Hoey, A. S., Gurney, G. G., Feary, D. A., Williams, I. D., et al. (2018). Gravity of human impacts mediates coral reef conservation gains. *Proceedings of the National Academy of Sciences*, 115(27):E6116–E6125.

Clough, Y. (2012). A generalized approach to modeling and estimating indirect effects in ecology. *Ecology*, 93(8):1809–1815.

Cox, D. R. (1958). *Planning of experiments*. John Wiley & Sons, New York.

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424.

Dawid, A. P. (2003). Causal inference using influence diagrams: the problem of partial compliance (with discussion). In Green, P., Hjort, N., and Richardson, S., editors, *Highly Structured Stochastic Systems*, Oxford Statistical Science Series (0-19-961199-8) Series, pages 45–81. Oxford University Press, Oxford.

Dawid, A. P. (2015). Statistical causality from a decision-theoretic perspective. *Annual Review of Statistics and Its Application*, 2(1):273–303.

Dawid, P. (2021). Decision-theoretic foundations for statistical causality. *Journal of Causal Inference*, 9(1):39–77.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

Devick, K. L., Bobb, J. F., Mazumdar, M., Claus Henn, B., Bellinger, D. C., Christiani, D. C., Wright, R. O., Williams, P. L., Coull, B. A., and Valeri, L. (2022). Bayesian kernel machine regression-causal mediation analysis. *Statistics in Medicine*, 41(5):860–876.

Digitale, J. C., Martin, J. N., and Glymour, M. M. (2022). Tutorial on directed acyclic graphs. *Journal of Clinical Epidemiology*, 142:264–267.

Ding, P. and VanderWeele, T. J. (2016). Sensitivity analysis without assumptions. *Epidemiology*, 27(3):368–377.

Dippel, C., Gold, R., Heblich, S., and Pinto, R. (2020). Mediation analysis in iv settings with a single instrument. preprint. preprint; `https://christiandippel.com/IVmediate_.pdf`.

Eldridge, D. J., Delgado-Baquerizo, M., Travers, S. K., Val, J., and Oliver, I. (2017). Do grazing intensity and herbivore type affect soil health? insights from a semi-arid productivity gradient. *Journal of Applied Ecology*, 54(3):976–985.

Eskelinen, A., Harpole, W. S., Jessen, M.-T., Virtanen, R., and Hautier, Y. (2022). Light competition drives herbivore and nutrient effects on plant diversity. *Nature*, 611(7935):301–305.

Farbmacher, H., Huber, M., Lafférs, L., Langen, H., and Spindler, M. (2022). Causal mediation analysis with double machine learning. *The Econometrics Journal*, 25(2):277–300.

Ferraro, P. J. and Hanauer, M. M. (2014). Quantifying causal mechanisms to determine how protected areas affect poverty through changes in ecosystem services and infrastructure. *Proceedings of the National Academy of Sciences*, 111(11):4332–4337.

Ferraro, P. J. and Hanauer, M. M. (2015). Through what mechanisms do protected areas affect environmental and social outcomes? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1681):20140267.

Flores, C. A. and Flores-Lagunes, A. (2013). Partial identification of local average treatment effects with an invalid instrument. *Journal of Business & Economic Statistics*, 31(4):534–545.

Fornara, D. A. and Tilman, D. (2009). Ecological mechanisms associated with the positive diversity-productivity relationship in an n-limited grassland. *Ecology*, 90(2):408–418.

Frölich, M. and Huber, M. (2017). Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 79(5):1645–1666.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, 48(3):432–435.

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.

Glynn, A. N. (2012). The product and difference fallacies for indirect effects. *American Journal of Political Science*, 56(1):257–269.

Grace, J. B. (2006). *Structural Equation Modeling and Natural Systems*. Cambridge University Press.

Grace, J. B. (2021). Instrumental variable methods in structural equation models. *Methods in Ecology and Evolution*, 12(7):1148–1157.

Grace, J. B., Anderson, T. M., Seabloom, E. W., Borer, E. T., Adler, P. B., Harpole, W. S., Hautier, Y., Hillebrand, H., Lind, E. M., Pärtel, M., Bakker, J. D., Buckley, Y. M., Crawley, M. J., Damschen, E. I., Davies, K. F., et al. (2016). Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature*, 529(7586):390–393.

Grace, J. B. and Irvine, K. M. (2020). Scientist's guide to developing explanatory statistical models using causal analysis principles. *Ecology*, 101(4):e02962.

Grace, J. B., Scheiner, S. M., and Schoolmaster Jr., D. R. (2015). Structural equation modeling: Building and evaluating causal models. In Fox, G. A., Negrete-Yankelevich, S., and Sosa, V. J., editors, *Ecological statistics: contemporary theory and application*, chapter 8, pages 168–199. Oxford University Press, Oxford, UK.

Grace, J. B., Schoolmaster Jr., D. R., Guntenspergen, G. R., Little, A. M., Mitchell, B. R., Miller, K. M., and Schweiger, E. W. (2012). Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere*, 3(8):art73.

Greenland, S., Pearl, J., and Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, 10(1):37–48.

Greenland, S. and Robins, J. M. (1985). Confounding and Misclassification. *American Journal of Epidemiology*, 122(3):495–506.

Hafeman, D. M. (2011). Confounding of indirect effects: a sensitivity analysis exploring the range of bias due to a cause common to both the mediator and the outcome. *Am J Epidemiol*, 174(6):710–717.

Hautier, Y., Niklaus, P. A., and Hector, A. (2009). Competition for light causes plant biodiversity loss after eutrophication. *Science*, 324(5927):636–638.

Heger, T. (2022). What are ecological mechanisms? suggestions for a fine-grained description of causal mechanisms in invasion ecology. *Biology & Philosophy*, 37(2):9.

Hernán, M. A. and Robins, J. M. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology*, 17(4):360–372.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

Holland, P. W. (1988). Causal inference, path analysis and recursive structural equations models. *ETS Research Report Series*, 1988(1):i–50.

Holmbeck, G. N. (2019). Commentary: Mediation and Moderation: An Historical Progress Report. *Journal of Pediatric Psychology*, 44(7):816–818.

Hong, G., Qin, X., and Yang, F. (2018). Weighting-based sensitivity analysis in causal mediation studies. *Journal of Educational and Behavioral Statistics*, 43(1):32–56.

Hoover, D. L., Wilcox, K. R., and Young, K. E. (2018). Experimental droughts with rainout shelters: a methodological review. *Ecosphere*, 9(1):e02088.

Huber, M. (2020). Mediation analysis. In Zimmermann, K. F., editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–38. Springer International Publishing.

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, 15(4):309.

Imai, K. and Kim, I. S. (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science*, 63(2):467–490.

Imai, K. and Kim, I. S. (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, 29(3):405–415.

Imai, K., Tingley, D., and Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1):5–51.

Jacoby, J. and Sassenberg, K. (2011). Interactions do not only tell us when, but can also tell us how: Testing process hypotheses by interaction. *European Journal of Social Psychology*, 41(2):180–190.

James, L. R. and Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of applied psychology*, 69(2):307.

Keiser, A. D., Keiser, D. A., Strickland, M. S., and Bradford, M. A. (2014). Disentangling the mechanisms underlying functional differences among decomposer communities. *Journal of Ecology*, 102(3):603–609.

Kendall, B. E. (2015). A statistical symphony: instrumental variables reveal causality and control measurement error. In Fox, G. A., editor, *Ecological Statistics: Contemporary theory and application*, chapter 7, pages 149–167. Oxford University Press.

Kim, C., Daniels, M. J., Hogan, J. W., Choirat, C., and Zigler, C. M. (2019). Bayesian methods for multiple mediators: Relating principle stratification and causal mediation in the ananlysis of power plant emission controls. *Ann Appl Stat*, 13(3):1927–1956.

Kim, C., Daniels, M. J., Marcus, B. H., and Roy, J. A. (2017). A framework for bayesian nonparametric inference for causal effects of mediation. *Biometrics*, 73(2):401–409.

Kimmel, K., Avolio, M. L., and Ferraro, P. J. (2023). Empirical evidence of widespread exaggeration bias and selective reporting in ecology. *Nature Ecology & Evolution*, 7(9):1525–1536.

Kimmel, K., Dee, L. E., Avolio, M. L., and Ferraro, P. J. (2021). Causal assumptions and causal inference in ecological experiments. *Trends in Ecology & Evolution*, 36(12):1141–1152.

Kraemer, H. C., Kiernan, M., Essex, M., and Kupfer, D. J. (2008). How and why criteria defining moderators and mediators differ between the baron & kenny and macarthur approaches. *Health Psychol*, 27(2S):S101–8.

Kunicki, Z. J., Smith, M. L., and Murray, E. J. (2023). A primer on structural equation model diagrams and directed acyclic graphs: When and how to use each in psychological and epidemiological research. *Advances in Methods and Practices in Psychological Science*, 6(2):25152459231156085.

Larsen, A. E., Meng, K., and Kendall, B. E. (2019). Causal analysis in control–impact ecological studies with observational data. *Methods in Ecology and Evolution*, 10(7):924–934.

Laubach, Z. M., Murray, E. J., Hoke, K. L., Safran, R. J., and Perng, W. (2021). A biologist's guide to model selection and causal inference. *Proc. R. Soc. B.*, 288(1943):20202815.

Le Poidevin, R. (2007). Action at a distance. *Royal Institute of Philosophy Supplements*, 61:21–36.

Linero, A. R. and Antonelli, J. L. (2023). The how and why of bayesian nonparametric causal inference. *WIREs Computational Statistics*, 15(1):e1583.

Linero, A. R. and Zhang, Q. (2022). Mediation analysis using bayesian tree ensembles. *Psychological Methods*.

Liu, G., Wang, L., Jiang, L., Pan, X., Huang, Z., Dong, M., and Cornelissen, J. H. C. (2018). Specific leaf area predicts dryland litter decomposition via two mechanisms. *Journal of Ecology*, 106(1):218–229.

Loh, W. W., Moerkerke, B., Loeys, T., and Vansteelandt, S. (2020). Heterogeneous indirect effects for multiple mediators using interventional effect models. *Epidemiologic Methods*, 9(1):20200023.

Loh, W. W., Moerkerke, B., Loeys, T., and Vansteelandt, S. (2022). Disentangling indirect effects through multiple mediators without assuming any causal structure among the mediators. *Psychological Methods*, 27:982–999.

MacKinnon, D. P. (2011). Integrating mediators and moderators in research design. *Research on Social Work Practice*, 21(6):675–681. PMID: 22675239.

MacKinnon, D. P. (2012). *Introduction to statistical mediation analysis*. Routledge.

MacKinnon, D. P., Fairchild, A. J., and Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58(1):593–614. PMID: 16968208.

MacKinnon, D. P. and Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the m to y relation in mediation analysis. *Personality and Social Psychology Review*, 19(1):30–43. PMID: 25063043.

MacKinnon, D. P., Valente, M. J., and Gonzalez, O. (2020). The correspondence between causal and traditional mediation analysis: the link is the mediator by treatment interaction. *Prevention Science*, 21(2):147–157.

Mellor, D. H. (1995). *The facts of causation*. Routledge, United Kingdom.

Miles, C., Kanki, P., Meloni, S., and Tchetgen Tchetgen, E. (2017). On partial identification of the natural indirect effect. *Journal of Causal Inference*, 5(2):20160004.

Miles, C. H., Shpitser, I., Kanki, P., Meloni, S., and Tchetgen Tchetgen, E. J. (2020). On semiparametric estimation of a path-specific effect in the presence of mediator-outcome confounding. *Biometrika*, 107(1):159–172.

Mulaik, S. (2009). Causation. In *Linear Causal Modeling with Structural Equations*, Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences, chapter 3. CRC Press.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1):69–85.

Murray, E. J. and Kunicki, Z. (2022). As the wheel turns: Causal inference for feedback loops and bidirectional effects. preprint. preprint, `https://doi.org/10.31219/osf.io/9em5q`.

Neyman, J., Iwaszkiewicz, K., and Kolodziejczyk, S. (1935). Statistical problems in agricultural experimentation. *Supplement to the Journal of the Royal Statistical Society*, 2(2):107–180.

Oliveira, B. F., Moore, F. C., and Dong, X. (2022). Biodiversity mediates ecosystem sensitivity to climate variability. *Communications Biology*, 5(1):628.

Pearce, N. and Lawlor, D. A. (2017). Causal inference—so much more than statistics. *International Journal of Epidemiology*, 45(6):1895–1903.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Pearl, J. (2001a). Bayesianism and causality, or, why i am only a half-bayesian. In Corfield, D. and Williamson, J., editors, *Foundations of Bayesianism*, pages 19–36. Springer Netherlands, Dordrecht.

Pearl, J. (2001b). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, pages 411–420, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Pearl, J. (2009). *Causality*. Cambridge University Press, New York, NY, 2nd edition edition.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychol Methods*, 19(4):459–481.

Pearl, J. (2023). The causal foundations of structural equation modeling. In Hoyle, R. H., editor, *Handbook of Structural Equation Modeling*, chapter 3. Guilford Press, second edition.

Pearl, J. and Verma, T. S. (1995). A theory of inferred causation. In Prawitz, D., Skyrms, B., and Westerståhl, D., editors, *Logic, Methodology and Philosophy of Science IX*, volume 134 of *Studies in Logic and the Foundations of Mathematics*, pages 789–811. Elsevier.

Pennisi, E. (2022). Global drought experiment reveals the toll on plant growth. *Science*, 377(6609):909–910.

Pirlott, A. G. and MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology*, 66:29–38. Rigorous and Replicable Methods in Social Psychology.

Poliseli, L., Coutinho, J. G. E., Viana, B., Russo, F., and El-Hani, C. N. (2022). Philosophy of science in practice in ecological model building. *Biology & Philosophy*, 37(4):21.

Ramsey, D. S. L., Forsyth, D. M., Wright, E., McKay, M., and Westbrooke, I. (2019). Using propensity scores for causal inference in ecology: Options, considerations, and a case study. *Methods in Ecology and Evolution*, 10(3):320–331.

Ribas, L. G. S., Pressey, R. L., and Bini, L. M. (2021). Estimating counterfactuals for evaluation of ecological and conservation impact: an introduction to matching methods. *Biological Reviews*, 96(4):1186–1204.

Richardson, A., Hudgens, M. G., Gilbert, P. B., and Fine, J. P. (2014). Nonparametric bounds and sensitivity analysis of treatment effects. *Stat Sci*, 29(4):596–618.

Rijnhart, J. J., Twisk, J. W., Chinapaw, M. J., de Boer, M. R., and Heymans, M. W. (2017). Comparison of methods for the analysis of relatively simple mediation models. *Contemporary Clinical Trials Communications*, 7:130–135.

Rijnhart, J. J., Valente, M. J., MacKinnon, D. P., Twisk, J. W., and Heymans, M. W. (2021). The use of traditional and causal estimators for mediation models with a binary outcome and exposure-mediator interaction. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(3):345–355. PMID: 34239282.

Rinella, M. J., Strong, D. J., and Vermeire, L. T. (2020). Omitted variable bias in studies of plant interactions. *Ecology*, 101(6):e03020.

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155.

Robins, J. M., Hernán, M. Á., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560.

Roth, D. L. and MacKinnon, D. P. (2012). Mediation analysis with longitudinal data. In *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences.*, Multivariate application series., pages 181–216. Routledge/Taylor & Francis Group, New York, NY, US.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701.

Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100(469):322–331.

Rubin, D. B. (2006). *Matched Sampling for Causal Effects*. Cambridge University Press, Cambridge, England.

Rudolph, K. E., Sofrygin, O., and van der Laan, M. J. (2021). Complier stochastic direct effects: identification and robust estimation. *J Am Stat Assoc*, 116(535):1254–1264.

Rüttenauer, T. and Ludwig, V. (2023). Fixed effects individual slopes: Accounting and testing for heterogeneous effects in panel data or other multilevel models. *Sociological Methods & Research*, 52(1):43–84.

Schlüter, M., Brelsford, C., Ferraro, P. J., Orach, K., Qiu, M., and Smith, M. D. (2023). Unraveling complex causal processes that affect sustainability requires more integration between empirical and modeling approaches. *Proceedings of the National Academy of Sciences*, 120(41):e2215676120.

Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition.

Shen, Y., Baingana, B., and Giannakis, G. B. (2017). Kernel-based structural equation models for topology identification of directed networks. *IEEE Transactions on Signal Processing*, 65(10):2503–2516.

Shipley, B. (2000). A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(2):206–218.

Singh, R., Xu, L., and Gretton, A. (2022). Kernel methods for multistage causal inference: Mediation analysis and dynamic treatment effects. arXiv.

Sitters, J. and Olde Venterink, H. (2015). The need for a novel integrative theory on feedbacks between herbivores, plants and soil nutrient cycling. *Plant and Soil*, 396(1):421–426.

Spencer, S. J., Zanna, M. P., and Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of personality and social psychology*, 89(6):845.

Stone, R. (1993). The assumptions on which causal inferences rest. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(2):455–466.

Sullivan, A. J. and VanderWeele, T. J. (2021). Bias and sensitivity analysis for unmeasured confounders in linear structural equation models. *arXiv preprint arXiv:2103.05775*.

Tchetgen Tchetgen, E. J. (2011). On causal mediation analysis with a survival outcome. *The International Journal of Biostatistics*, 7(1):000010220215574679135 1.

Tchetgen Tchetgen, E. J. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Ann Stat*, 40(3):1816–1845.

Tian, Q., Liu, N., Bai, W., Li, L., Chen, J., Reich, P. B., Yu, Q., Guo, D., Smith, M. D., Knapp, A. K., Cheng, W., Lu, P., Gao, Y., Yang, A., Wang, T., et al. (2016). A novel soil manganese mechanism drives plant species loss with increased nitrogen deposition in a temperate steppe. *Ecology*, 97(1):65–74.

Valeri, L. and Vanderweele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with sas and spss macros. *Psychol Methods*, 18(2):137–150.

VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.

VanderWeele, T. and Rothman, K. (2021). Formal causal models. In Zinner, S., editor, *Modern Epidemiology*, chapter 3. Wolters Kluwer Health, fourth edition edition.

VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. *Epidemiology*, 20(6):863–871.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, 21(4):540–551.

VanderWeele, T. J. (2012a). Comments: Should principal stratification be used to study mediational processes? *Journal of Research on Educational Effectiveness*, 5(3):245–249. PMID: 25558296.

VanderWeele, T. J. (2012b). Invited Commentary: Structural Equation Models and Epidemiologic Analysis. *American Journal of Epidemiology*, 176(7):608–612.

VanderWeele, T. J. (2014). A unification of mediation and interaction: A 4-way decomposition. *Epidemiology*, 25(5):749–761.

VanderWeele, T. J. and Ding, P. (2017). Sensitivity analysis in observational research: Introducing the e-value. *Annals of Internal Medicine*, 167(4):268–274. PMID: 28693043.

VanderWeele, T. J. and Robins, J. M. (2007). Four types of effect modification: A classification based on directed acyclic graphs. *Epidemiology*, 18(5):561–568.

VanderWeele, T. J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, 2(4):457–468.

VanderWeele, T. J. and Vansteelandt, S. (2010). Odds Ratios for Mediation Analysis for a Dichotomous Outcome. *American Journal of Epidemiology*, 172(12):1339–1348.

VanderWeele, T. J. and Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiol Methods*, 2(1):95–115.

Veldhuis, M. P., Howison, R. A., Fokkema, R. W., Tielens, E., and Olff, H. (2014). A novel mechanism for grazing lawn formation: large herbivore-induced modification of the plant–soil water balance. *Journal of Ecology*, 102(6):1506–1517.

Wang, L. and Tchetgen Tchetgen, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J R Stat Soc Series B Stat Methodol*, 80(3):531–550.

Wilcox, R. R. (2010). *Fundamentals of modern statistical methods: Substantially improving power and accuracy*, volume 249. Springer.

Wilkins, K. D., Smith, M. D., Holdrege, M. C., Wilfahrt, P. A., Gherardi, L. A., Ohlert, T. J., Collins, S. L., Dukes, J. S., Knapp, A. K., Phillips, R. P., Sala, O. E., Tatarko, A., Felton, A., and International Drought Experiment Network (IDE) (2022). Impacts of intensified drought across space and time: Results from the international drought experiment. In *2022 ESA Annual Meeting: A Change is Gonna Come*, Montreal, Canada. Ecological Society of America.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data, second edition*. The MIT Press. MIT Press.

Wu, A. D. and Zumbo, B. D. (2008). Understanding and using mediators and moderators. *Social Indicators Research*, 87(3):367–392.

Xu, S., Liu, L., and Liu, Z. (2022). Deepmed: Semiparametric causal mediation analysis with debiased deep learning. arXiv.

Xue, F., Tang, X., Kim, G., Koenen, K. C., Martin, C. L., Galea, S., Wildman, D., Uddin, M., and Qu, A. (2022). Heterogeneous mediation analysis on epigenomic PTSD and traumatic stress in a predominantly African American cohort. *Journal of the American Statistical Association*, 117(540):1669–1683.

Zhou, X. (2021). Semiparametric Estimation for Causal Mediation Analysis with Multiple Causally Ordered Mediators. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(3):794–821.

Supplementary Material

S.1 Limitations of the two-part estimation approach for mediation analyses

Some ecological studies attempt to detect mediators in experiments by first manipulating the treatment and then estimating the dependence between the treatment and outcome and between the treatment and mediator. In our hypothetical study, this approach would be represented by two equations, where the effect of drought (D) on soil moisture (M) and the effect of drought on productivity (P) are estimated by

$$P_i = \beta_0 + \beta_1 D_i + \varepsilon_{i1} \tag{S1}$$

$$M_i = \theta_0 + \theta_1 D_i + \varepsilon_{i2} \ , \qquad i = 1, \ldots, n \ , \tag{S2}$$

where D_i is the treatment assigned to plot i, P_i is the plot-level productivity, M_i is the plot-level soil moisture, β_0 and θ_0 are intercepts, β_1 and θ_1 are coefficients, and ε_{i1} and ε_{i2} are plot-level error terms. The average productivity of all plots under the no-drought control is represented by β_0, while β_1 represents the average change in productivity across all plots when going from the control state ($D = 0$) to the drought-treated state ($D = 1$).

Complete randomisation of the drought treatment allows us to assume that the plot-level observations are independent and identically distributed and that the effects of any treatment-mediator and treatment-outcome confounders have been removed. Thus, ordinary least squares (OLS) estimation of Equation (S1) yields an unbiased estimator of β_1. Under complete randomisation, if the OLS-estimated coefficient $\widehat{\beta}_1 < 0$, the drought treatment reduces productivity on average across plots. Likewise, using OLS regression to estimate Equation (S2) yields an unbiased estimator of θ_1. If the estimated coefficient $\widehat{\theta}_1 < 0$, then, on average, the drought treatment induces a reduction in soil moisture across plots. If $\widehat{\beta}_1 < 0$ and $\widehat{\theta}_1 < 0$ and both are statistically significant, some studies may conclude that there is sufficient evidence to claim that the effect of drought on productivity is mediated by soil moisture (Figures S1a to S1c). However, the two-part estimation procedure does not quantify the indirect effect; that is, the proportion of the effect of drought on productivity that is mediated by soil moisture is not estimated (Figure S1d). Thus, other possible conclusions can also be drawn from the results of the two-part estimation approach, including a conclusion that the hypothesised mediator plays no mediating role at all (Figures S1e and S1f).

S.2 Effect of mediator-outcome confounders on mediation effects in randomised controlled trials

Here, we answer a question that many readers may have: why, exactly, is the three-part estimation procedure invalid for identifying and estimating the effect of drought on productivity through soil moisture when drought was randomised but there exist mediator-outcome confounders?

Consider our hypothetical drought experiment in which some plots experienced heavy grazing by herbivores (Figure 4). Because drought was randomised across plots, researchers may incorrectly believe that historical grazing (G), which is correlated with both soil moisture

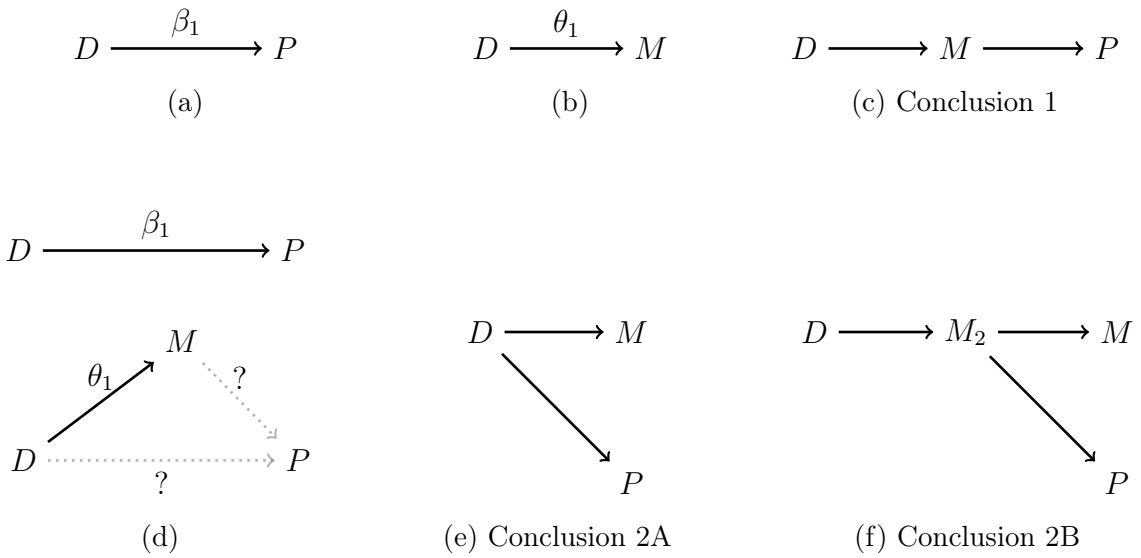


Figure S1: The two-part estimation procedure for the hypothetical drought study can result in multiple conclusions. A two-part estimation process in which (a) drought D is found to relate to productivity P, and (b) drought is found to be related to soil moisture M, leads to (c) the conclusion that drought influences productivity though soil moisture. However, the two-part procedure does not estimate the effects of soil moisture and drought on productivity (d), which means that alternative conclusions (e) and (f) are also possible from the evidence given by (a) and (b) alone. D = drought, M = soil moisture, M_2 = secondary mediator (e.g., photosynthesis), P = productivity.

and productivity, need not be added to Equation (3). Thus, the researcher would instead estimate the following three equations:

$$P_i = \beta_0 + \beta_1 D_i + \varepsilon_{i1} \tag{S1}$$

$$M_i = \theta_0 + \theta_1 D_i + \varepsilon_{i2} \ , \tag{S2}$$

$$P_i = \delta_0 + \delta_1 D_i + \delta_2 M_i + \varepsilon_{i3} \ , \qquad i = 1, \ldots, n \ . \tag{S3}$$

With randomisation of the drought treatment, the distribution of historical grazing across all plots is, on average, the same in the drought-treated plots as it is in the control plots. This property ensures that θ_1 in Equation (S2) is an unbiased estimator of the average effect of drought on soil moisture when changing D from 0 to 1, as detailed in Supplement S.1. In Equation (S2), we do not need to control for any other variables that may affect productivity – the variation in P resulting from those factors is included in the error term ε_{i2}. Of course, we still have sampling variability, represented by ε_{i2}, but modes of statistical inference (e.g., confidence intervals) have been developed to quantify the uncertainty that the differences in treatment and control plots have arisen by chance. However, sampling variability is different from bias: as the sample size grows, the sampling variability of the θ_1 estimates will converge around the true value of θ_1.

In contrast, randomisation of the treatment does not render Equation (S3) unbiased in the estimation of δ_1, nor is it unbiased in the estimation of δ_2. For estimation of δ_2, Equation (S3) is biased, because it does not control for historical grazing G, which is positively correlated

with both M and P; i.e., ε_{i3} is correlated with M. In contrast to the effect suggested by Figure 4, we suppose here that plots with historically more grazing are, on average, more productive and have more soil moisture, possibly through nutrient addition by grazers' waste (Sitters and Olde Venterink, 2015; Veldhuis et al., 2014). If plots that have been historically free of grazing are, on average, less productive and have less soil moisture, then the estimate of δ_2 includes both the effect of M on P and some of the effect of G on P. In other words, the estimate includes the unconfounded effect of soil moisture on productivity caused by drought, but also includes the effect of soil moisture confounded by historical grazing. Thus, the estimate of δ_2 is positively biased, because it is a weighted average of the uncounfounded and confounded effects of soil moisture.

Bias also enters the estimation of δ_1 – specifically, the estimate is also positively biased. The sign of the bias in estimating δ_1 is the same as the sign of the correlation between M and P in the absence of a randomised experiment, which is positive in our drought study. Recall that researchers declare mediation to be present if the estimated effect of drought on productivity gets less negative when controlling for M, i.e., $\widehat{\delta}_1 > \widehat{\beta}_1$ (see Supplement S.1). Also recall that ε_{i3} in Equation (S3) is positively correlated with M and P – if G increases, M increases and P increases. So, for estimation of δ_1, Equation (S3) will be upwardly biased. The direction of bias implies that we would detect mediation when soil moisture is not, in fact, a mediator at all (i.e., when there is no arrow from M to P in Figure 5 and the detection of mediation only reflects the non-causal correlation between M and P that comes from G). Thus, soil moisture will appear more influential on productivity than it is.

To illustrate the intuition behind these claims without referring to equations, consider a prediction made by a researcher for the hypothetical drought experiment: the drought treatment, on average, lowers soil moisture, and lower soil moisture, on average, reduces grassland productivity. In addition, the researcher predicts that plots with more historical grazing are more productive and have more soil moisture. Imagine we selected at random a drought-treated plot and a no-drought control plot from the field experiment and told the researcher only the treatment status of each plot. The researcher would anticipate that the control plot has higher average productivity, based on their initial experimental prediction. This prediction step is akin to Equation (S1), which is answering the question, "For a randomly selected plot from the study population, what is the expected effect of the drought treatment?"

Now, suppose that before revealing each plot's measured productivity, we tell the researcher that the two plots were randomly selected from a subgroup of plots that all had identical soil moisture levels. In light of the new information, the researcher is given the opportunity to revise their initial guess of which plot has higher measured productivity. They might wonder why the drought-treated plot had the same soil moisture as the no-drought control plot, despite the control plot not being exposed to drought. Based on the researcher's original predictions about the effect of historical grazing on soil moisture, one possible explanation for the control and treated plots to have identical soil moisture is that the treated group experienced more historical grazing. Greater historical grazing is associated with higher productivity, independent of soil moisture. Based on this insight, the researcher would update their first guess and instead predict that the drought-treated plot has higher productivity. This adjustment step is akin to using Equation (S3) in the presence of an unmeasured mediator-outcome confounder. In the case of the drought study, the

adjustment includes unmeasured differences in historical grazing across plots, making the effect of soil moisture appear more influential than it really is.

If the effects of all mediator-outcome confounders have been appropriately eliminated, researchers can estimate the magnitude of the effect of drought on productivity through soil moisture using the three-part procedure in one of two ways: by taking the difference between β_1 and δ_1, or by taking the product of θ_1 and δ_2. Both of these traditional regression-based approaches have been commonly applied and studied in many other scientific fields. The traditional regression approach to mediation analysis used in fields such as epidemiology and public health relies on Equations (S1) and (S3) and is known as the "difference method". With this method, the magnitude of the indirect effect of drought on productivity through soil moisture is $\beta_1 - \delta_1$, while the coefficient δ_1 represents the magnitude of the direct effect. The presence of mediation is thus determined if soil moisture explains some of the effect of drought on productivity, i.e., $|\beta_1 - \delta_1| > \epsilon$, $\epsilon > 0$. In contrast, the traditional regression approach to mediation used in social sciences and psychology is known as the "product method" (popularised by Baron and Kenny 1986) and uses Equations (S2) and (S3). With the product method, δ_1 again represents the direct effect, while the indirect effect is $\theta_1\delta_2$. If $|\theta_1\delta_2| > \epsilon$, $\epsilon > 0$, then mediation is determined to be present. The product method is typically how the direct and indirect effects from Equations (S2) and (S3) are represented in SEM (Muthén and Asparouhov, 2015). It should be noted, however, that the product and difference methods only coincide when the outcome and mediator are continuous and the regression equations are fit using OLS estimation, provided the statistical assumptions for OLS are satisfied. For a binary outcome that is not a rare event, the difference and product methods do not give identical results (Mackinnon and Dwyer, 1993; MacKinnon et al., 1995), and the estimates from both methods are not directly interpretable as indirect effects (VanderWeele and Vansteelandt, 2010; Valeri and Vanderweele, 2013). In such cases, the product method using log-linear models is typically preferred for binary outcomes (MacKinnon et al., 2007; Rijnhart et al., 2019, 2023).

S.3 Effect of heterogeneity on mediation effects estimated using traditional regression-based approaches

For the hypothetical drought study, suppose we fit the models

$$M_i = \omega_0 + \omega_{1i} D_i + \varepsilon_{2i} \tag{S4}$$

$$P_i = \alpha_0 + \alpha_{1i} D_i + \alpha_{2i} M_i + \varepsilon_{3i} \tag{S5}$$

instead of Equations (S2) and (S3), where the coefficients ω_{1i} and α_{1i} are allowed to vary for each plot i. If both drought and soil moisture are randomly assigned to plots (e.g., a manipulation-of-mediator design), the average effect of D on M is $\overline{\omega_1}$, and the average effect of M on P is $\overline{\alpha_2}$. If each of these effects are consistent across all plots, then using the product method of defining the indirect effect as $\overline{\omega_1}\overline{\alpha_2}$ would provide an unbiased estimator of the effect of drought on productivity through soil moisture. Conversely, suppose the effects of D on M and the effects of M on P are heterogeneous across plots. For one set of plots, the effect of D on M is negative, $\overline{\omega_1} < 0$, and the effect of M on P is also negative, $\overline{\alpha_2} < 0$. The average effect of D on P through M for this group of plots would be positive (Figure S2a).

For a different set of plots, the effect of D on M is small but positive, $\overline{\omega_1} > 0$, and the effect of M on P is also positive $\overline{\alpha_2} > 0$. The mediated effect of D on P through M for this different set of plots would again be positive (Figure S2b). If we averaged across all i plots, the indirect effect of D on P, $\overline{\omega_1}$, could be negative or zero, while the effect of M on P, $\overline{\alpha_2}$, could be negative, zero, or positive. Thus, the indirect effect of drought on productivity through soil moisture averaged across all plots could also be negative, zero, or positive, despite the indirect effect in both subsets of plots being positive.

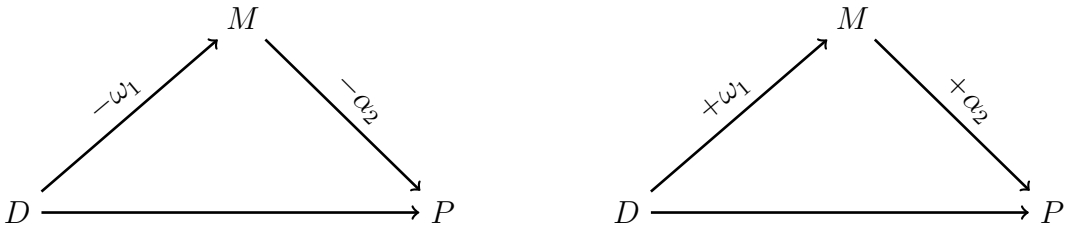


Figure S2: The effect of D on P through M can be identical in magnitude and size for two different plots where (a) the effects of D on M and M on P are negative or (b) the effects of D on M and M on P are positive. In both (a) and (b), the indirect effect of D on P through M is positive: $\omega_1\alpha_2$. Labels are as defined in Figure S1.

S.4 Multilevel models for clustered longitudinal data

A multilevel model typically captures distinct groupings of clustered data by specifying an observation-level equation with group-level intercepts in concert with higher-level equations that describe the group-level intercepts for each for each grouping of the data (Gelman and Hill, 2006). In mixed effects modelling, a variant of multilevel modelling commonly applied in ecology, error terms are included in the higher-level equations (Bolker et al., 2009). Modelling clustered longitudinal data with error terms in each of the higher-level equations allows researchers to quantify the variation within and among various groupings (Bolker et al., 2009) and has the benefit of partial pooling which reduces the effect of outlying groups on parameter estimation without eliminating their effect entirely.

To estimate mediation effects without bias using a mixed effects model for our hypothetical drought study, researchers must assume that the differences in productivity among plots or among sites are uncorrelated with other predictors in the model (Seber and Lee, 2003). This assumption is likely violated in many ecological settings, leading to estimates that are biased (Gelman, 2006). To see how the violation of this assumption could occur, let us consider the problem from the perspective of our drought study. In a mixed effects model, for an observation h that is measured at time t and belongs to plot i within site s,

we replace Equations (S1) to (S3) with

$$P_h = \phi_{1,i[h]} + \mu_{1,st[h]} + \beta_1 D_h + \varepsilon_{1,h} \tag{S6}$$

$$M_h = \phi_{2,i[h]} + \mu_{2,st[h]} + \theta_1 D_h + \varepsilon_{2,h} \tag{S7}$$

$$P_h = \phi_{3,i[h]} + \mu_{3,st[h]} + \delta_1 D_h + \delta_2 M_h + \varepsilon_{3,h} \ , \tag{S8}$$
$$i = 1, \ldots, n \ ; s = 1, \ldots, S \ ; t = 1, \ldots, T \ ; h = 1, \ldots, nST \ ,$$

where each site is composed of n_s plots, for a total of $n = n_1 + n_2 + \cdots + n_s$ plots, and each plot is repeatedly measured over T time points; $i[h]$ is the plot i containing observation h; $st[h]$ is the site-time group containing h; P_h, D_h, and M_h are the productivity, drought, and soil moisture values measured for an observation h; β_1 represents the overall effect of drought on productivity; θ_1 represents the effect of drought on soil moisture; δ_1 and δ_2 represent the effects of drought and soil moisture on productivity; $\phi_{1,i[h]}$, $\phi_{2,i[h]}$, $\phi_{3,i[h]}$ are plot-level intercepts; $\mu_{1,st[h]}$, $\mu_{2,st[h]}$, $\mu_{3,st[h]}$ are site-time group-level intercepts; and $\varepsilon_{1,h}$, $\varepsilon_{2,h}$, $\varepsilon_{3,h}$ are the error terms. Note that Equation (S8) was introduced in Section 5.4.1 as Equation (6).

For a mixed effects model, we must also specify higher-level equations that include group-averaged intercepts. These equations are

$$\phi_{1,i} = \phi_{1\cdot} + \eta_{1,i} \tag{S9}$$

$$\phi_{2,i} = \phi_{2\cdot} + \eta_{2,i} \tag{S10}$$

$$\phi_{3,i} = \phi_{3\cdot} + \eta_{3,i} \tag{S11}$$

$$\mu_{1,st} = \mu_{1\cdot} + \eta_{1,st} \tag{S12}$$

$$\mu_{2,st} = \mu_{2\cdot} + \eta_{2,st} \tag{S13}$$

$$\mu_{3,st} = \mu_{3\cdot} + \eta_{3,st} \ , \tag{S14}$$

where $\phi_{1\cdot}$, $\phi_{2\cdot}$, $\phi_{3\cdot}$ are the averages of the plot-varying intercepts $\phi_{1,i[h]}$, $\phi_{2,i[h]}$, $\phi_{3,i[h]}$, respectively; $\mu_{1\cdot}$, $\mu_{2\cdot}$, $\mu_{3\cdot}$ are the averages of the site-time group-varying intercepts $\mu_{1,st[h]}$, $\mu_{2,st[h]}$, $\mu_{3,st[h]}$, respectively; $\eta_{1,i}$, $\eta_{2,i}$, $\eta_{3,i}$ are plot-level errors; and $\eta_{1,st}$, $\eta_{2,st}$, $\eta_{3,st}$ are site-time group-level errors.

For simplicity when discussing how mediation effects can be estimated with bias when mediator-outcome confounding exists, we will focus on the effect of drought and soil moisture on productivity described by Equations (S8), (S11) and (S14).

In a large-scale regional or global set of drought experiments where one might expect to obtain clustered longitudinal data, some sites could be in regions with low soil moisture, resulting in the differences in productivity between those sites and others in the study to be correlated with soil moisture. This would result in a correlation between soil moisture and the site-time groupings which, if not explicitly modelled in Equation (S14), would be included in the error term $\eta_{3,st}$. Let us substitute Equation (S11) and Equation (S14) into Equation (S8), which gives us

$$P_h = \phi_{3\cdot} + \mu_{3\cdot} + \delta_1 D_h + \delta_2 M_h + \eta_{3,i} + \eta_{3,st} + \varepsilon_{3,h} \ . \tag{S15}$$

As $\eta_{3,i}$, $\eta_{3,st}$, and $\varepsilon_{3,h}$ are all error terms, we can combine them into a new error term e' and rewrite the model as

$$P_h = \phi_{3\cdot} + \mu_{3\cdot} + \delta_1 D_h + \delta_2 M_h + e' \ . \tag{S16}$$

Since $\eta_{3,st}$ is correlated with soil moisture, and $\eta_{3,st}$ is now part of the new error term, then e' is correlated with M_h, thus violating the assumption that the errors should be independent of predictors in the regression model.

One way around this issue is to instead allow for group-level effects where error terms are not estimated at the group-level (Gelman, 2006). The observation-level model describing the effect of drought and soil moisture on productivity would remain the same as in Equation (S8), but the second-level equations for $\phi_{3,i}$ and $\mu_{3,st}$ would instead be given as

$$\phi_{3,i} \sim N(\phi_{3\bullet}, \infty) \tag{S17}$$

$$\mu_{3,st} \sim N(\mu_{3\bullet}, \infty) \tag{S18}$$

where the infinite variances allow for maximum variation in the plot-level and site-time group-level effects from the data. This is equivalent to fitting separate regression models for each plot and each site-time grouping, where estimates that vary across groups are completely unpooled (Bafumi and Gelman, 2006; Gelman and Hill, 2006). The same effect could be achieved by using dummy variables for plot and site-time groupings (Bollen and Brand, 2010). The coefficient estimates will thus be unbiased even in the presence of unmodelled correlation between the differences among plots or among sites and soil moisture, such as in the presence of unmeasured mediator-outcome confounding (Fitzmaurice et al., 2012).

Unfortunately, fitting separate models requires a large number of parameters to fit separate intercepts for each plot and each site-time grouping. Instead, an alternative multilevel modelling approach described in Section 5.4.1 can accommodate the presence of correlation between differences among groups and predictors in the model without the need for separate models for each grouping. We would use the same observation-level models specified in Equations (S6) to (S8) above, however we must specify different higher level equations from those specified in mixed effects modelling to accommodate correlation introduced by mediator-outcome confounders.

To account for mediator-outcome confounders, we must specify second-level equations that include group-averaged soil moisture as predictors of the group-level intercepts. We use the same second-level equations for $\phi_{1,i}$, $\phi_{2,i}$, $\mu_{1,st}$, and $\mu_{1,st}$ as in Equations (S9), (S10), (S12) and (S13), but the second-level equation for $\phi_{3,i}$ would instead be specified with a plot-averaged soil moisture term, $\nu\overline{M}_i$, in Equation (S19) and the second-level equation for $\mu_{3,st}$ would be specified with a site-time group-averaged soil moisture term, $\kappa\overline{M}_{st}$, in Equation (S20), as we showed in Section 5.4.1 with Equations (7) and (8). The full set of second-level equations are

$$\phi_{1,i} = \phi_{1\bullet} + \eta_{1,i} \tag{S9}$$

$$\phi_{2,i} = \phi_{2\bullet} + \eta_{2,i} \tag{S10}$$

$$\phi_{3,i} = \phi_{3\bullet} + \nu\overline{M}_i + \eta_{3,i} \tag{S19}$$

$$\mu_{1,st} = \mu_{1\bullet} + \eta_{1,st} \tag{S12}$$

$$\mu_{2,st} = \mu_{2\bullet} + \eta_{2,st} \tag{S13}$$

$$\mu_{3,st} = \mu_{3\bullet} + \kappa\overline{M}_{st} + \eta_{3,st}\ , \tag{S20}$$

where ν is the coefficient for the predictor $\overline{M}_i$ representing plot-level averages of soil moisture and κ is the coefficient for the predictor $\overline{M}_{st}$ representing the site-time grouped means of soil moisture.

By including $\overline{M}_i$ in Equation (S19) and $\overline{M}_{st}$ in Equation (S20), we explicitly model any potential correlation between soil moisture and differences in productivity at the plot or site-time group levels. We do not need to include group-averaged terms for drought, since drought being randomised allows us to assume no unmodelled correlation between drought and the differences in productivity among plots or among sites. Thus we arrive at the formulations for obtaining unbiased estimators of mediation effects using multilevel models as given in Equations (6) to (8) in Section 5.4.1, and the indirect effect can be estimated as $\theta_1\delta_2$ using the product method. When observations are only collected for two points in time, multilevel modelling for causal inference is equivalent to a difference-in-differences analysis (Abadie, 2005; Wooldridge, 2021), which has been recommended for observational ecological studies (Butsic et al., 2017; Larsen et al., 2019). Multilevel models without group-level error terms and with group-averaged variables as predictors in the higher level regression equations (as in Section 5.4.1) can be estimated using SEMs (Allison, 2009; Andersen, 2022; Bollen and Brand, 2010).

S.5 Estimating effects for multiple mediator pathways

In the hypothetical drought study, we declared that the researchers were only interested in the mediating effect of soil moisture (Section 2). However, researchers may also be interested in additional mediators through which drought affects productivity. In the analyses outlined in previous sections, the effect of other mediators are lumped into the direct effect, which is interpreted as the effect of drought on productivity through mediators other than soil moisture. If we are interested in measuring the effect of drought on productivity through multiple mediating variables separately (Figure S3), we require additional causal assumptions to estimate effects for each mediator without bias.

To identify individual indirect effects for each of m mediators, which is a common objective in SEM analyses, one might presume that the traditional approach could be repeated for each mediator separately by replacing M with M_j, $j = 1, 2, \ldots, m$, in Equations (S2) and (S3) to estimate the effect of drought on productivity through M_j. This approach, however, requires at least three more causal assumptions. First, we must assume that there are no mediator-outcome confounders for each of the measured mediators. In other words, Assumption A3 must be satisfied for each measured mediator. Second, we must assume that there are no unmeasured mediator-mediator confounders, i.e., there are no common causes between two mediators that have not been accounted for in the regression equations (Grace et al., 2015; Loh et al., 2022; VanderWeele and Vansteelandt, 2014). If we have an unmeasured confounder U between two mediators M_1 and M_2 as in Figure S3a, U acts as an unmeasured confounder between M_1 and P through its effect on M_2, resulting in correlation between M_1 and P not due to the treatment D and biasing the coefficient estimates in Equation (S3). Similarly, U acts as an unmeasured confounder between M_2 and P through its effect on M_1, again producing bias. To satisfy Assumption A3 when using the instrumental variable approach described in Section 5.3, we must either assume that no other mediators are observed or obtain an instrumental variable for each mediator. Third, we must assume that the mediators are independent from each other, i.e., the values of one mediator do not depend on the presence or values of another mediator, which is to satisfy Assumption A4 for each mediator. If interdependencies between mediating variables exist (Figure S3b), then in-

dividual direct and indirect effects of multiple mediators cannot be estimated (VanderWeele and Vansteelandt, 2014).

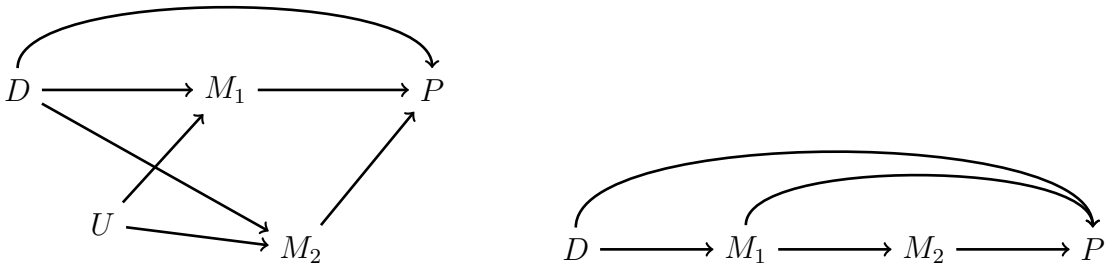


Figure S3: Additional dependencies among variables can introduce bias when estimating the effects of more than one mediator; for example, (a) an unmeasured common cause U of M_1 and M_2, or (b) a dependency of M_2 on M_1. Labels are as defined in Figure S1.

If the assumptions of no unmeasured mediator-mediator confounders or independence of the mediators are unlikely to hold, one could instead estimate the effect of drought on productivity through the entire set of mediators $\{M_1, M_2, \ldots, M_m\}$ jointly. Joint direct and indirect effects are defined for continuous outcomes with binary or continuous mediators and for binary outcomes with continuous mediators (VanderWeele and Vansteelandt, 2014). To estimate the joint direct and indirect effects when mediator-mediator interactions exist, one must make additional statistical assumptions, and when exposure-mediator interactions are present, the formulae become increasingly complicated (VanderWeele and Vansteelandt, 2014). Further, if the mediators are time-varying, estimating direct and indirect effects typically requires a different class of estimation procedures, different definitions of direct and indirect effects, and additional causal assumptions (MacKinnon, 2012; VanderWeele, 2015; VanderWeele and Tchetgen Tchetgen, 2017).

S.6 Decomposition of causal effects

We can decompose the total effect of drought on productivity derived from Equations (2) and (3) given in Section 5.2 into the direct and indirect effects. We assume that the drought treatment is binary and soil moisture and productivity are continuous variables, as we have done in Section 2. We also assume that there is no interaction between the drought treatment and the soil moisture mediator. The average effect of drought on productivity operating through the soil moisture mediator, called the average total indirect effect (TIE), is given by $\delta_2\theta_1$. More specifically, $\delta_2\theta_1$ describes the average change in productivity if drought was implemented on all plots ($D = 1$) but soil moisture changed from the value it would be under the no-drought control condition (M_0) to the value it would be under the drought-treated condition (M_1). The remaining effect of drought on productivity not operating through soil moisture, but possibly going through other mediated causal paths not explicitly denoted in the DAG, is described by the average pure direct effect (PDE) and is given by δ_1. That is, δ_1 describes the average amount by which productivity would change if drought were changed from control ($D = 0$) to treated ($D = 1$) on all plots but soil moisture remained at the level

it would have been under no drought conditions (M_0). Combining the average PDE and the average TIE gives us the average total effect $TE = \delta_1 + \delta_2\theta_1$.

In ecological studies, it is often more realistic to expect an interaction between the treatment and mediator. Indeed, a common recommendation for mediation analyses is to include interactions between the treatment and mediator if an interaction cannot be ruled out, since interactions are often difficult to detect with significance tests and not accounting for interactions can bias the estimates of direct and indirect effects (VanderWeele, 2015). If we wish to include an interaction between drought and soil moisture, an interaction term $\delta_4 D_i M_i$ is added to Equation (3). When defining direct and indirect effects that include treatment-mediator interactions, the potential outcomes framework provides clear intuition for where an interaction coefficient should appear. Thus, the average PDE and average TIE are given as

$$PDE = \delta_1 + \delta_4(\theta_0 + \theta_1) \tag{S21}$$

$$TIE = (\delta_2\theta_1 + \delta_4\theta_1) + \delta_4\theta_1 \ . \tag{S22}$$

As with a mediation analysis that does not include a treatment-mediator interaction, combining the average TIE and average PDE give us the average total effect:

$$TE = PDE + TIE = [\delta_1 + \delta_4(\theta_0 + \theta_1)] + [(\delta_2\theta_1 + \delta_4\theta_1) + \delta_4\theta_1] \ . \tag{S23}$$

In many ecological studies, the treatment variable may be continuous. Drought, for example, could be specified using one of several possible drought indices. For a continuous drought treatment with an interaction term between the treatment and mediator, we can instead define the average PDE and average TIE in terms of the difference between the treated and control drought values:

$$PDE = \delta_1(d - d^*) + \delta_4(\theta_0 + \theta_1 d^*)(d - d^*) \tag{S24}$$

$$TIE = (\delta_2\theta_1 + \delta_4\theta_1 d^*)(d - d^*) + \delta_4\theta_1(d - d^*)(d - d^*) \ , \tag{S25}$$

where d is the treated value of drought and d^* is the untreated value of drought. The total effect can be derived again as a combination of the average PDE and average TIE:

$$\begin{aligned} TE = {} & [\delta_1(d - d^*) + \delta_4(\theta_0 + \theta_1 d^*)(d - d^*)] \\ & + [(\delta_2\theta_1 + \delta_4\theta_1 d^*)(d - d^*) + \delta_4\theta_1(d - d^*)(d - d^*)] \ . \end{aligned} \tag{S26}$$

In some cases, it may be desirable to break down the direct and indirect effects to obtain further interpretations of mediation effects (Figure S4). We now describe these alternative mediation effects using our hypothetical drought study with the outcome productivity P influenced by a drought treatment D and soil moisture mediator M, but these can be generalized to any outcome Y with treatment A and mediator M.

Using the potential outcomes notation, the pure direct effect can be split into two parts: a controlled direct effect in which the mediator can be set to specific values not necessarily determined by the state of the drought treatment, and a reference interaction term (Figure S4). The **controlled direct effect** (CDE) captures the average amount by which productivity would change if drought were changed from $D = 0$ to $D = 1$ across all plots

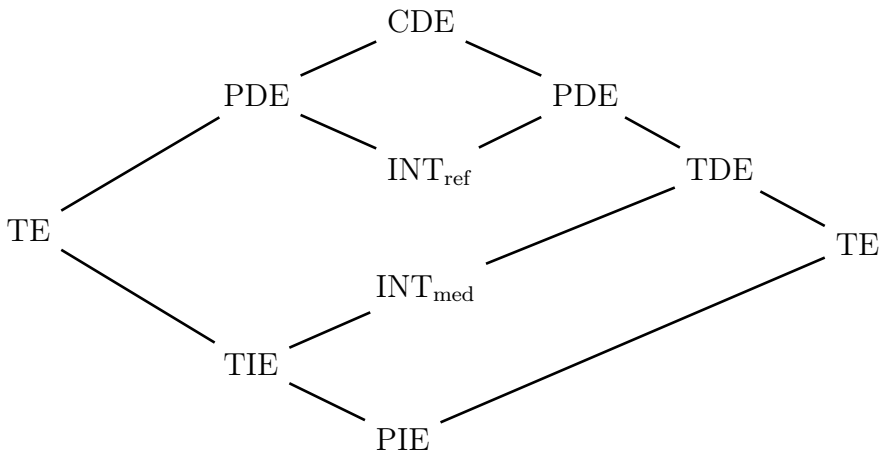


Figure S4: Two decompositions of mediation effects. Adapted from VanderWeele (2014).

and soil moisture were fixed at a specified level $M = m$ for all plots. The controlled direct effect is given by

$$CDE(m) = E[P_{1m} - P_{0m}] \,. \tag{S27}$$

We only need to satisfy Assumptions A3 and A5 to estimate the controlled direct effect. The average CDE of drought on productivity for all plots in the hypothetical drought experiment is the difference in the average productivity for treated and control plots if soil moisture were held (controlled) at a single level across all plots. For each possible soil moisture level that could be fixed across all plots, there is a different average controlled direct effect.

Why would an ecologist be interested in controlled direct effects? Let's say that an ecosystem manager wants to reduce the effects of drought on productivity and determines some values of soil moisture for which the controlled direct effect of drought on productivity is small and thus less of a management concern. The manager would then have the option of reducing the effect of drought on productivity by externally increasing the soil moisture, say, through a ground-level irrigation system, to the levels implied by the favourable controlled direct effect estimates.

The **reference interaction** (INT_{ref}) represents an additive interaction of the effect of drought and soil moisture on productivity that only occurs when soil moisture remains at the value it would be under the no-drought control condition (M_0). This interaction effect is given by

$$INT_{ref} = E[(P_{1M_1} - P_{1M_0} - P_{0M_1} + P_{0M_0})(M_0)] \,. \tag{S28}$$

If there exists no interaction between drought and soil moisture, the average $CDE(m) = PDE = \delta_1$ for our drought study represented by Equations (2) and (3). The equivalence of the average controlled direct effect and the average pure direct effect is generally true for all regression-based approaches without mediator-outcome interactions, i.e., no $\delta_4 D_i M_i$ term in Equation (3). If an interaction between the treatment and mediator is present, the $CDE(m)$ must be redefined to include δ_4 (VanderWeele and Vansteelandt, 2009).

The total indirect effect can also be separated into two components: a pure indirect effect in which the mediator changes while the treatment is fixed at $D = 0$ (instead of $D = 1$ as in

the TIE), and a mediated interaction term (Figure S4). The **pure indirect effect** (PIE) captures the amount by which productivity changes if M were changed from the level it was under the no-drought control condition (M_0) to the level it was under the drought-treated condition (M_1) while fixing drought to the control condition ($D = 0$),

$$PIE = E[P_{0M_1} - P_{0M_0}] \,. \tag{S29}$$

The **mediated interaction** (INT_{med}) represents the additive effect of both drought and soil moisture on productivity and the effect of the drought on soil moisture. The mediated interaction is given as

$$INT_{med} = E[(P_{1M_1} - P_{1M_0} - P_{0M_1} + P_{0M_0})(M_1 - M_0)] \,. \tag{S30}$$

Combining the mediated interaction with the pure direct effect gives us the **total direct effect** (TDE), which describes the amount by which productivity changes if drought were changed from $D = 0$ to $D = 1$ but soil moisture were fixed to the value it would be under the drought-treated condition (M_1):

$$TDE = INT_{med} + PDE = E[P_{1M_1} - P_{0M_1}] \,. \tag{S31}$$

Note that, in contrast to the TDE, the PDE fixes soil moisture to the value it would be under the no-drought control condition (M_0). Adding INT_{med} to PDE captures additional information about the effect of soil moisture under the drought treatment to give the TDE (Figure S4).

The decomposition of causal effects can be extended to cases of two or more mediators that can potentially interact with both the treatment and each other, but doing so requires the researcher to define more potential outcomes and more decomposable components of the total effect and to designate which contrasts among the many potential outcomes one wants to consider (e.g., Bellavia and Valeri, 2017). The researcher would also have to eliminate the effects of mediator-outcome confounders for all mediators in the analyses (as detailed in Supplement S.5).

References

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, 72(1):1–19.

Allison, P. D. (2009). Structural equation models with fixed effects. In *Fixed Effects Regression Models*, Quantitative Applications in the Social Sciences, chapter 6. SAGE Publications, Inc., Thousand Oaks; Thousand Oaks, California.

Andersen, H. K. (2022). A closer look at random and fixed effects panel regression in structural equation modeling using lavaan. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(3):476–486.

Bafumi, J. and Gelman, A. (2006). Fitting multilevel models when predictors and group effects correlate. Available at SSRN 1010095. Paper presented at the Annual Meeting of the Midwest Political Science Association. Chicago, IL. 20-23 April.

Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*, 51(6):1173–1182.

Bellavia, A. and Valeri, L. (2017). Decomposition of the Total Effect in the Presence of Multiple Mediators and Interactions. *American Journal of Epidemiology*, 187(6):1311–1318.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3):127–135.

Bollen, K. A. and Brand, J. E. (2010). A General Panel Model with Random and Fixed Effects: A Structural Equations Approach. *Social Forces*, 89(1):1–34.

Butsic, V., Lewis, D. J., Radeloff, V. C., Baumann, M., and Kuemmerle, T. (2017). Quasi-experimental methods enable stronger inferences from observational data in ecology. *Basic and Applied Ecology*, 19:1–10.

Fitzmaurice, G., Laird, N., and Ware, J. (2012). *Applied Longitudinal Analysis*. Wiley Series in Probability and Statistics. Wiley.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, 48(3):432–435.

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.

Grace, J. B., Scheiner, S. M., and Schoolmaster Jr., D. R. (2015). Structural equation modeling: Building and evaluating causal models. In Fox, G. A., Negrete-Yankelevich, S., and Sosa, V. J., editors, *Ecological statistics: contemporary theory and application*, chapter 8, pages 168–199. Oxford University Press, Oxford, UK.

Larsen, A. E., Meng, K., and Kendall, B. E. (2019). Causal analysis in control–impact ecological studies with observational data. *Methods in Ecology and Evolution*, 10(7):924–934.

Loh, W. W., Moerkerke, B., Loeys, T., and Vansteelandt, S. (2022). Disentangling indirect effects through multiple mediators without assuming any causal structure among the mediators. *Psychological Methods*, 27:982–999.

MacKinnon, D., Lockwood, C., Brown, C., Wang, W., and Hoffman, J. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials*, 4(5):499–513. PMID: 17942466.

MacKinnon, D. P. (2012). *Introduction to statistical mediation analysis*. Routledge.

Mackinnon, D. P. and Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17(2):144–158.

MacKinnon, D. P., Warsi, G., and Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behav Res*, 30(1):41.

Muthén, B. and Asparouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1):12–23.

Rijnhart, J. J. M., Twisk, J. W. R., Eekhout, I., and Heymans, M. W. (2019). Comparison of logistic-regression based methods for simple mediation analysis with a dichotomous outcome variable. *BMC Medical Research Methodology*, 19(1):19.

Rijnhart, J. J. M., Valente, M. J., Smyth, H. L., and MacKinnon, D. P. (2023). Statistical mediation analysis for models with a binary mediator and a binary outcome: the differences between causal and traditional mediation analysis. *Prevention Science*, 24(3):408–418.

Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition.

Sitters, J. and Olde Venterink, H. (2015). The need for a novel integrative theory on feedbacks between herbivores, plants and soil nutrient cycling. *Plant and Soil*, 396(1):421–426.

Valeri, L. and Vanderweele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with sas and spss macros. *Psychol Methods*, 18(2):137–150.

VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.

VanderWeele, T. J. and Tchetgen Tchetgen, E. J. (2017). Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 79(3):917–938.

VanderWeele, T. J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, 2(4):457–468.

VanderWeele, T. J. and Vansteelandt, S. (2010). Odds Ratios for Mediation Analysis for a Dichotomous Outcome. *American Journal of Epidemiology*, 172(12):1339–1348.

VanderWeele, T. J. and Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiol Methods*, 2(1):95–115.

Veldhuis, M. P., Howison, R. A., Fokkema, R. W., Tielens, E., and Olff, H. (2014). A novel mechanism for grazing lawn formation: large herbivore-induced modification of the plant–soil water balance. *Journal of Ecology*, 102(6):1506–1517.

Wooldridge, J. M. (2021). Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators. Available at SSRN 3906345.