COLLOQUIUM PAPER

Causal inference in coupled human and natural systems

Paul J. Ferraro[a,b], **James N. Sanchirico**[c,d], **and Martin D. Smith**[e,f,1]

[a]Carey Business School, Johns Hopkins University, Baltimore, MD 21202; [b]Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, MD 21205; [c]Department of Environmental Science and Policy, University of California, Davis, CA 95616; [d]Resources for the Future, Washington, DC 20036; [e]Nicholas School of the Environment, Duke University, Durham, NC 27708; and [f]Department of Economics, Duke University, Durham, NC 27708

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved July 26, 2018 (received for review March 31, 2018)

Coupled human and natural systems (CHANS) are complex, dynamic, interconnected systems with feedback across social and environmental dimensions. This feedback leads to formidable challenges for causal inference. Two significant challenges involve assumptions about excludability and the absence of interference. These two assumptions have been largely unexplored in the CHANS literature, but when either is violated, causal inferences from observable data are difficult to interpret. To explore their plausibility, structural knowledge of the system is requisite, as is an explicit recognition that most causal variables in CHANS affect a coupled pairing of environmental and human elements. In a large CHANS literature that evaluates marine protected areas, nearly 200 studies attempt to make causal claims, but few address the excludability assumption. To examine the relevance of interference in CHANS, we develop a stylized simulation of a marine CHANS with shocks that can represent policy interventions, ecological disturbances, and technological disasters. Human and capital mobility in CHANS is both a cause of interference, which biases inferences about causal effects, and a moderator of the causal effects themselves. No perfect solutions exist for satisfying excludability and interference assumptions in CHANS. To elucidate causal relationships in CHANS, multiple approaches will be needed for a given causal question, with the aim of identifying sources of bias in each approach and then triangulating on credible inferences. Within CHANS research, and sustainability science more generally, the path to accumulating an evidence base on causal relationships requires skills and knowledge from many disciplines and effective academic-practitioner collaborations.

social-ecological systems | marine protected areas | quasiexperiment | bioeconomic | spatial dynamics

SUSTAINABILITY SCIENCE

Coupled human and natural systems (CHANS) are complex, dynamic, interconnected systems that have important feedbacks across social and environmental dimensions (1). Researchers have applied CHANS and precursor bioeconomic models to a wide range of human–environment interactions, including ocean fisheries, lake fisheries, human disease spread, wildfires, coastal landscape change, deforestation, invasive species spread, desertification, eutrophication, large mammal extinctions, and collapse of whole civilizations (1–11). In the science of CHANS, related social-ecological systems (SESs), and sustainability more broadly, scholars aim to understand and identify causal relationships (3, 12, 13). In CHANS, a "cause" is an attribute of the system that could change (i.e., be manipulated by humans or nature) and a "causal effect" is the difference between outcomes (results) that are experienced when the attribute is held at one value rather than another value.

To make causal claims, CHANS and SES scholars tend to take one of two approaches: (*i*) predictive inference, which fits models of deterministic or stochastic system dynamics to observations and judges success by goodness-of-fit criteria or quality of reconstruction ("Are the data consistent with the model?"), and (*ii*) causal inference, which exploits experimental or quasi-experimental variation in one or more variables to isolate causal relationships and judges success by the credibility of untestable assumptions about the data-generating process ("Are there plausible rival explanations for the estimated relationships between causes and effects?"). Because scholars often use similar statistical techniques (e.g., regression), conceptual confusion about the difference between predictive and causal approaches can lead to methodological confusion about the best ways to approach empirical analyses in CHANS, SESs, and sustainability science generally (*SI Appendix*).

Suppose, for example, that a researcher is interested in understanding whether changes in the value of X (e.g., road density) cause changes in the value of Y (e.g., species richness), and by how much. In predictive inference approaches, one assumes that if model A, which includes X, explains more of the variation in Y than model B, which excludes X, model A is preferred. Intuitively, predictive inference approaches are assumed to shed light on causal relationships through the following logic: If an estimated (fitted) model successfully predicts observations out of sample or is consistent with a theorized dynamic process, the model is likely to reflect true causal relationships underlying the system (14). Predictive inference approaches typically shed light on causal relationships only when strong, often unstated, identifying assumptions about model structures are invoked.

In contrast, the goal of causal inference approaches is to make the identifying assumptions transparent and credible in a particular study design. For example, the credibility of a causal inference in a randomized controlled trial that experimentally manipulates X is not judged on whether variation in X explains a lot of the variation in Y [although the precision of the estimate of X's effect on Y (i.e., the amount of statistical uncertainty) is an important attribute of the design, and the magnitude of variation in Y attributable to X may have important policy implications]. Instead, credibility is judged by whether one can plausibly argue that the experimental design isolates a true causal relationship (i.e., whether one can plausibly rule out rival explanations for the estimated relationship between X and Y) (15).

Judgment about the plausibility of rival explanations, or the ability of predictive approaches to elucidate causal relationships, relies on the plausibility of two assumptions:

i) Excludability: the assumption that factors driving variation in the treatment variable have no causal link to the outcome variable except through their effects on variation in the treatment (the "treatment" is the causal variable of interest). For example, if we could randomly manipulate fish stocks (a treatment) in marine patches, excludability implies that the manipulation (the driver of variation in treatment) only

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Economics, Environment, and Sustainable Development," held January 17–18, 2018, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/economics-environment-and.

Author contributions: P.J.F., J.N.S., and M.D.S. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: marsmith@duke.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1805563115/-/DCSupplemental.

Published online August 20, 2018.

affects economic returns to fishers (an outcome) through the manipulation's effect on the fish stocks.

ii) No interference [also known as stable unit treatment value assumption or "no spillovers among units" assumption]: the assumption that the outcome for a unit when exposed to a particular value of a treatment does not depend on the value of the treatment in any other units. For example, if we randomly manipulate fish stocks in marine patches (the units), no interference implies that the causal effect of a specific change in fish stock in one patch is unaffected by whether or not the fish stock has changed in any other patch.

In nonexperimental settings, like those found in CHANS, the plausibility of these two assumptions is suspect. The strong coupling in social and ecological systems implies that the most common ways of addressing confounding variables in the CHANS literature are unlikely to satisfy the excludability assumption. Moreover, the same coupled nature implies that interference among study units is a much more widespread, and complex, problem in CHANS research than previously appreciated.

Although concerns about these assumptions are not new and neither assumption can be verified empirically, we demonstrate how one can explore their plausibility in a given CHANS context through a combination of theory, field knowledge, and indirect empirical tests. We also conduct a deeper exploration of the ways in which the no-interference assumption can be violated in CHANS, how to think about the plausibility of violations in a specific context, and how to address potential violations should they likely exist in that context.

To improve the quality of the empirical evidence base within CHANS, scholars must undertake more deliberate, transparent efforts to ascertain the plausibility of these two key assumptions, and to explore the implications of potential violations. To illustrate our ideas, we analyze a marine CHANS and focus on estimating the effect on fish stocks from marine protected areas (MPAs) and from hypoxic events. An MPA is a possible positive shock to the CHANS, whereas hypoxia is a possible negative shock. Although we use a marine CHANS for illustration, the ideas and claims outlined in the following sections are broadly generalizable to all CHANS and highlight the challenges for CHANS researchers.

Excludability

Imagine that a CHAN marine system experiences a shock, S, such as an MPA placed on a seascape. An MPA could causally affect ecological and social outcomes, Y, through a variety of ecological and economic mechanisms, M (16) (Fig. 1). One could ask, "By how much do MPAs change, on average, the level of fish stocks in or around the MPAs?" This causal question implies a thought experiment: What is the difference in expected fish stocks in the presence of MPAs and, in the same locations, expected fish stocks in the absence of those MPAs ("absence" implies the counterfactual regulatory regimes that would have existed had the MPAs not been established)?

The first quantity, the stock with MPAs, will be observable after MPAs are placed on the seascape. In contrast, the second quantity, the counterfactual stock in the absence of MPAs, will not be observable. This counterfactual stock could be estimated by first taking seascape zones in which the effects of an MPA would be entirely contained, and then randomly assigning MPAs to some of these zones. Randomization can satisfy the excludability assumption: The procedure by which some zones are exposed to MPAs has no causal link to the fish stocks except through its effect on the probability that a zone will be exposed to an MPA. Thus, fish stocks in zones without MPAs provide an estimate of the counterfactual fish stocks in the MPA zones.

Randomization, represented by variable Z, ensures that confounding variables O and U are absent and the excludability assumption is satisfied (Fig. 1). Of course, even in an experiment, many other variables affect fish stocks, such as temperature (represented by P in Fig. 1). They are important for predictive

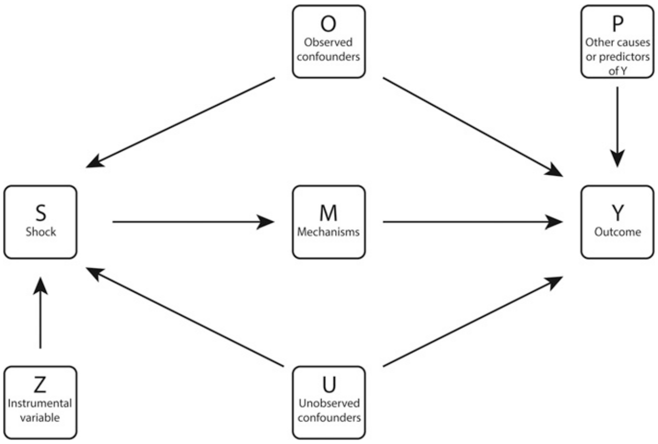

Fig. 1. Directed causal graph that depicts the causal relationship of S on Y and confounding variables that can mask or mimic the causal relationship. Single-headed arrows represent a causal path between two variables. Four empirical designs can be used to estimate the causal effect of S on Y (15): (*i*) Exert experimental control on S; (*ii*) condition on observable confounders (O), and assume that unobserved confounders (U) do not exist or are time-invariant and that their effects are eliminated via panel data designs; (*iii*) exploit nonexperimental variation in S that comes from a source (Z) that has no path to Y except through S; or (*iv*) identify an isolated and exhaustive set of mechanisms (M). Estimating the individual mechanism causal effects on Y requires additional design features (16, 46). Each design requires different assumptions for causal inference and, depending on those assumptions, may identify a different causal estimand.

modeling or improving the precision of the estimated causal effect, but they are not relevant in creating an unbiased estimator of the causal effect of S on Y. Randomization of S is sufficient.

In CHANS, however, random assignment of shocks is rare. Instead, observed and unobserved factors that systematically drive spatial and temporal variation in shocks also affect the outcomes. One cannot assume that O and U in Fig. 1 are absent, and failure to eliminate their influence in a causal study implies violation of the excludability assumption. This violation creates hidden bias. In the social sciences, the source of hidden bias is often termed unobservable heterogeneity and the problem it creates is called endogeneity.

To eliminate hidden bias, nonexperimental designs aim to replicate, conceptually, the idealized experimental design. In other words, high-quality, nonexperimental designs aim, via design and statistical methods, to eliminate the confounding effects of O and U, at least for a subgroup of the population. The theory for how to eliminate these confounding effects is summarized in Fig. 1 and has been described elsewhere (15, 17, 18). However, the implications of the excludability assumption for appropriate use of these designs in CHANS have not been well described.

We focus on two important implications: (*i*) Credible causal inference requires structural knowledge of the CHANS, particularly about the sources of variation in the causal variable (also know as the mechanisms through which treatment is assigned), and (*ii*) units exposed to shocks in CHANS, so-called "treated units," are best viewed as coupled pairings of environmental and human elements of the CHANS.

Structural Knowledge of the Treatment Assignment Mechanisms. Empirical designs that attempt to mimic randomized experiments are often referred to as "reduced-form" designs. In contrast to "structural" designs, reduced-form designs do not attempt to model the theoretical, functional relationship between the shock and the outcome, or between the shock, outcome, and confounding variables. Nevertheless, structural knowledge is essential for estimating causal relationships in reduced-form designs.

An essential piece of structural knowledge is information about the "treatment assignment mechanism" or "selection mechanism," the process by which some units in the system came to be exposed to particular values of the causal variable and other units were not. This knowledge helps to identify candidate variables for O, U, and Z (Fig. 1), which then points to the appropriate data and empirical design for estimating a causal relationship between S and Y. Thus, in CHANS studies aimed at causal inference, we would expect to find clear descriptions of the treatment assignment mechanism (e.g., why do we see protected areas where we see them, why are some places subject to more human interference than other places, what explains the spatial and temporal variation in species richness or invasive species?) However, these descriptions are rare outside of a few CHANS contexts [e.g., terrestrial protected areas (19)].

For example, we reviewed one of the larger empirical literatures in the marine CHANS context: the impact of MPAs on ecological and social outcomes (*SI Appendix*). Among nearly 200 MPA impact studies, less than 10% characterized the sources of spatial and temporal variation in MPA assignment. Without this information, one cannot evaluate the credibility of the studies' causal inferences because it is impossible to evaluate the plausibility of the excludability assumption. The same problem also makes it impossible to interpret the more than two dozen meta-analyses on MPA impacts. Meta-analyses cannot address hidden biases in the original studies, no matter how numerous the studies.

Even without a clear explanation of the treatment assignment mechanism, we would expect that most MPA studies would identify and control for many potential observable confounding variables through conditioning strategies (e.g., regression or matching estimators), and attempt to control for unobservable, but fixed (time-invariant), confounders through panel designs that exploit repeated observations before and after MPA establishment (20). However, fewer than half of the studies identify and control for observable confounding factors that affect both the outcomes and where, and when, MPAs are established, and most of these studies control for only one or two variables. Studies typically ensure that outcome indicators are measured in the same way in control and treated observations, but these efforts do not address underlying differences between the MPA and the control, or in the before period in a before-after design. Furthermore, only 6% of the studies have repeated observations on outcomes before MPA establishment (*SI Appendix*).

Viewing nonexperimental designs as attempts to replicate idealized experimental designs also yields expectations about studies of impacts on fisheries outside the MPAs. The impacts can arise via ecological dispersal mechanisms. Suppose, for example, that the protection covers a biological source patch and the unprotected patches are sink patches. The ideal randomized experiment would take pairs of interconnected sinks and sources, and then randomize some of the sources to be protected. This design benchmark implies that, in any nonexperimental study, plausibility of the excludability assumption requires a clear characterization of source/sink dynamics before protection. It is not sufficient for the affected and unaffected sinks to be similar at baseline; their sources should also be similar. However, not a single published study makes any assertion, or provides any data, about source/sink dynamics before protection, thereby making it difficult to evaluate the credibility of the causal inferences made.

Treated Units in CHANS Are Coupled Pairings of Environmental and Human Elements. Ecologists often view the treated units in CHANS as clusters of biophysical attributes (e.g., a patch or a population in a landscape or seascape). In reality, treatments are assigned jointly to biophysical clusters and human clusters, where the human clusters may be households, businesses, vessels, or communities. In the marine context, for example, a hypoxic event in a fishery affects both the ecological attributes of the system and the behaviors and welfare of the humans in the system. In other words, treatments (causes) are not assigned to the environmental attributes of the CHANS separate from the human attributes.

This coupled system perspective implies that behavioral confounding variables are just as important as biophysical confounding variables, whether the cause being studied is anthropogenic or not. Thus, the results of ecological studies in CHANS that fail to consider human sources of bias in their designs are unlikely to have clear causal interpretations (e.g., sardine and anchovy landings in ref. 14). Consider the phase diagram of the dynamic path toward equilibrium for a CHANS in which humans harvest a fish stock under open access (Fig. 2), and imagine the government has established an MPA. A standard approach to estimating the MPA's effect on the stock would be to match the MPA site to a comparison site (or sites) without an MPA. One would then contrast the change in stock at the MPA to the change in stock at the comparison site, a design known as a before-after-control-impact (BACI) design or a difference-in-differences design.

Imagine researchers pick a comparison site that, in the period before the MPA establishment, has identical ecological conditions to the MPA site, including the initial stock. MPA establishment has been hypothesized to be influenced by fishing history (21). If researchers ignore this nonrandom assignment mechanism, they ignore the fact that sites have had different exploitation histories as embodied by different levels of fishing effort before MPA establishment. Thus, even when the preprotection stocks and population dynamics are the same at both sites, the postprotection stocks are going to be different in the absence of any causal effect of the MPA. The same problem would arise if one were to match only on effort. The implication is that, in most CHANS, conditioning on both ecological and human confounding variables will be critical (or at least using two pretreatment time periods to check for parallel trends, which would imply that the systems may be subject to similar dynamic paths).

In our review of MPA impact studies, the joint consideration of ecological and human confounders is nearly absent, even in the more sophisticated BACI designs. The BACI studies tend to seek comparison sites that are similar, at baseline, in terms of habitat type, species richness, biomass, and other environmental variables (e.g., temperature) (*SI Appendix*). Greater baseline similarity is assumed to make the key BACI design assumption more plausible (22). The observable before-after trend in the outcome variable at the comparison sites represents the

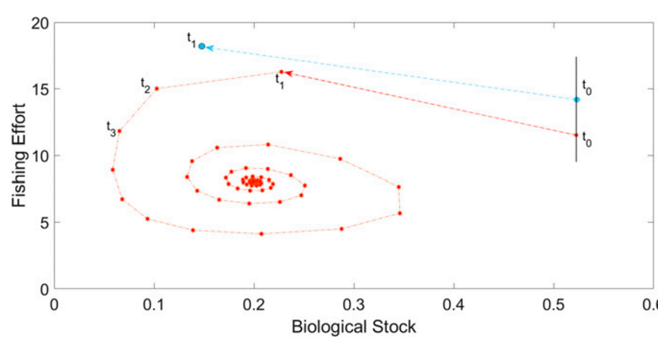

Fig. 2. Treated units are coupled pairings of ecological and social attributes of the CHANS. The phase diagram (red) shows how stock and effort of a potentially treated patch adjust over time (t) in a marine CHANS in the absence of a shock (the true counterfactual) (8). The stars are individual data points in the coupled pairing. To estimate the effect of a shock (e.g., hypoxia, creating a protected area), analysts compare a shocked site (red star at t_0) with a site that has a similar baseline stock level but is not shocked (blue circle at t_0). The excludability assumption implies that the dynamic path of postshock stock in the comparison site represents the counterfactual stock path in the shocked site had it not been shocked (e.g., at t_1). Failure to condition on baseline social conditions, in the form of fishing effort, results in a violation of the excludability assumption: Even in the absence of protection, the stock in the shocked site could be higher (as depicted) or lower than the stock in the site without a shock (compare the circle path with the star path).

unobservable counterfactual trend at the protected sites in the absence of protection. Human dimensions, however, are typically ignored: the harvest technology, the scale of the exploited resource, market types and proximity, proximity to fishing ports and processing infrastructure, employment in other fisheries or outside of fisheries, the management regime or institutional setting, and other sociocultural factors. More than 80% of the BACI studies fail to mention anything about the preprotection human dimensions (*SI Appendix*). In addition to directly addressing preprotection human dimensions, BACI designs with repeated observations before the MPA establishment could make the excludability assumption more plausible by showing parallel trends at MPA and comparison sites before MPA establishment.

In general, causal inference requires that controls and treated units behave similarly in the absence of treatment, and parallel trends are consistent with this requirement but are not a sufficient condition. In CHANS that are premised on the tight coupling both within and across the human and natural systems, the potential for parallel trends to be satisfied but excludability violated introduces caution in interpreting causal inference results. Given that initial conditions are critical for dynamic trajectories in CHANS over time, showing similar levels and having more observations to assess trends could help in this regard because initial conditions affect short-run dynamics (*SI Appendix*, Fig. S1). Assessing whether the excludability assumption is valid in CHANS should involve a combination of theory, empirical analysis of outcome variables pretreatment, and empirical knowledge about human and natural system components as well as the institutional environment (*SI Appendix*, Fig. S1).

No Interference

Like excludability, the no-interference assumption is implicit in both experimental and observational designs. The no-interference assumption implies that the outcome for unit i with or without treatment is only conditional on whether unit i has been treated or not; it does not depend on whether other units have been treated or not (technically, potential outcomes are stable no matter what the treatment assignment vector looks like). So, for example, if marine zone A is protected and, in response, fishers move to a nearby unprotected zone B to fish, zone B's fishing activity in the absence of protection depends on whether or not zone A is protected. That type of spillover effect is a violation of the no-interference assumption: an example of interference among units. Interference among treated units is also a possibility, but we will focus on treated-to-untreated interference in our simulations.

Interference in CHANS is generated by four mechanisms: (*i*) Treatment moves treated humans (e.g., fishers) into untreated zones; (*ii*) treatment changes behaviors of untreated humans in untreated zones; (*iii*) treatment induces ecological change outside of the treated zone through, for example, dispersal of biomass or changes in predator–prey relationships; and (*iv*) treatment generates market-mediated spillovers from treated to untreated zones. In the modeling below, we do not examine potential interference due to market-mediated spillovers directly and argue that they are subsumed in (*i*) and (*ii*). If treatment changes a price in a way that constitutes interference, it must create an incentive that moves treated humans, changes behaviors of untreated humans in the untreated zones, or both.

To evaluate how interference can hinder the performance of empirical designs aimed at causal inference in CHANS, we develop a spatial-dynamic bioeconomic simulation model of a fishery. A fishery bioeconomic model illustrates the key components of a CHANS: natural resource dynamics, human dynamics, and feedbacks between the two (1). This model also includes the central features of SES research: natural resource units, human users, governance regimes, and feedbacks between the human and natural subsystems that are mediated by governance (13). Our model emphasizes the tight coupling between human and natural systems: The natural system (biological stocks) are observable through behaviors in the human system (fishing).

The additional layer of spatial dynamics allows us to explore the consequences for the CHANS when one zone within the system is affected (treated) by an ecological shock such as hypoxia, a technological disaster such as an oil spill, or a policy intervention such as the creation of an MPA (6). The ecosystem is divided into discrete patches that correspond to fishing zones (23–25), and fishers make discrete decisions about whether to fish and, if so, which zone to fish in (6, 7, 21, 25). The model generates simulated data on fishing effort and catches from which fish stocks can be estimated (7). In all of our simulations, the excludability assumption is satisfied.

We analyze two natural (ecological) system types: (*i*) a closed system with own-patch population dynamics in each patch and (*ii*) a source-sink system with own-patch population dynamics and dispersal from a source patch to a sink patch. The model is a discrete-time version of previously published models (21, 23) (*SI Appendix*). The signs and magnitudes of the dispersal parameters determine the direction and flow of biomass over space (23). When all dispersal parameters are equal to zero, the system is closed. In the source-sink system, the source-to-sink dispersal parameter (emigration) and the sink-from-source dispersal parameter (immigration) are nonzero.

In the human system, each fisher decides on each choice occasion whether or not to fish and, if choosing to fish, picks one of the zones (*SI Appendix*). The economic model is sufficiently generic to apply to a range of fishery CHANS. Similar models have been applied to a broad range of economic scales and of species and ecological conditions (7, 26–33). Statistical generalizations are often used in empirical applications (28, 30, 31), but the model is sufficiently complex to produce simulated data that are consistent with the basic choice structure faced by fishers (6, 7, 25).

The human and natural systems are coupled through the link between fishing effort and biological stock (8) (Fig. 3 and *SI Appendix*). When stocks are lower, expected harvests and corresponding expected revenues in a patch are lower. Lower revenues discourage effort, making fishing in other patches, or not fishing, relatively more attractive. This mechanism affects the underlying state variables (e.g., fish stocks) in each location such that all dynamic trajectories of the CHANS are altered.

Before introducing the role of ecological dispersal, we simulate a shock to the closed system that highlights the role of human mobility on interference. Specifically, we first simulate the system without a shock to produce true counterfactual paths of the biological stocks. We then simulate the system again, but after some period of time, we decrease the stock and carrying capacity in the treated patch by 50%, allowing carrying capacity to stay at 50% of its original level for the remainder of the simulation (i.e., for all posttreatment periods) (Fig. 3, *Top*). A 50% negative shock could reflect a deleterious ecological disturbance such as hypoxia or an oil spill, but it could also reflect creating a marine reserve closed to fishing in half of the patch when there is no compensating ecological spillover. The marine reserve interpretation is consistent with two studies of large-scale marine reserves that compare changes in fishing, before and after reserve formation, in multiple zones: Each treated zone is a broad area that contains a reserve, and each control zone does not contain a reserve (34, 35). To simulate a positive shock, we follow the same process, but increase stock and carrying capacity by 50% in the treated patch in the posttreatment period. In both cases, the other two patches without a shock are candidate controls.

We then introduce the role of ecological dispersal by simulating a shock in a four-patch source-sink system (Fig. 3, *Bottom*). We first simulate the system without a shock to produce the true counterfactual paths of the four biological stocks. We then simulate the effect of creating a marine reserve in source patch 4. The sink for this source patch, patch 3, is the treated patch because it receives the biomass immigration from the source patch that is closed to fishing. Patch 1 is a sink patch whose source (patch 2) is not closed to fishing, and thus could serve as a control. In both systems, the simulations generate an actual (with-shock) and counterfactual (without-shock) time series of

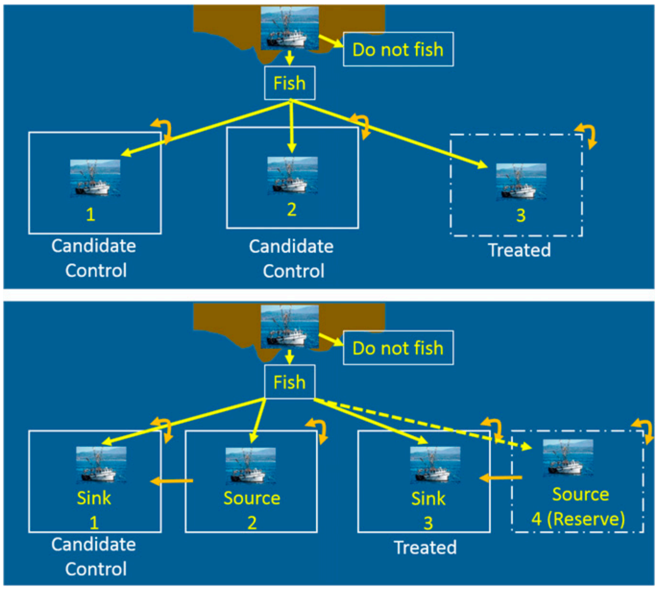

Fig. 3. Schematic of the coupled human-natural system with and without biological dispersal. (*Top*) In the ecologically closed system, there are three patches that correspond to fishing zones. Each patch has population dynamics that operate within the zone and do not disperse to other zones (ecological dispersal). The zones are connected by economic decisions about whether and where to fish (economic dispersal). When patch 3 is treated with a shock, ecological dispersal is not affected, and patches 1 and 2 are candidate control sites. (*Bottom*) In an ecologically interconnected system, there are four patches, which represent two pairs of source and sink patches. Each patch has population dynamics that operate within the zone and disperse across zones (from patch 2 to 1 and from patch 4 to 3). When a source (patch 4) is closed to fishing, its paired sink (patch 3) is treated because it receives dispersal from the closed source. The sink patch 1 is a candidate control because its corresponding source (patch 2) is not closed to fishing.

biological stocks. The difference between the two is the true treatment effect of the shock. The simulation also produces a panel dataset on catches and fishing effort. These panel data mimic data that are typically available in commercial fisheries and used to estimate the effects of marine reserves and other shocks. We can use these data to estimate the treatment effect of a shock in our systems. We can then compare the estimated effect with the true effect to assess the performance of designs that depend on observational data in the system to estimate the true treatment effect.

We explore how the difference between the true and estimated effects (in other words, the effect of interference) is moderated by human and capital mobility. In our simulated CHAN, human (labor) and capital mobility are the same: Fishing labor travels along with the fishing capital (the vessels). In other CHANS, the two forms of mobility may operate separately. We define mobility in the model as the responsiveness of fishers to fishing revenues relative to distance (in other words, the marginal rate of substitution between expected revenue and travel cost) (*SI Appendix*). If mobility is high, a small difference in expected revenues across patches will trigger vessel movements, whereas if mobility is low, a large difference in expected revenues is required to incentivize vessel movements (6, 23). In CHANS more generally, human and capital mobility is typically moderated not only by distance and mode of transport but also by institutional structures, such as property rights and regulations, and cultural norms, such as proclivities to migrate.

Interference in the Ecologically Closed System. In an ecologically closed system (Fig. 3, *Top*), human mobility is both a cause of interference in estimating the treatment effect and a moderator of the treatment effect itself (Fig. 4 and *SI Appendix*, Figs. S2 and S3). When mobility is low, interference is low: The negative shock to the treated patch has minimal effect on the untreated, candidate control patches. No interference implies that the actual stock of a candidate control tracks its own counterfactual path (Fig. 4, *Top* and *SI Appendix*, Fig. S2*A*). This path is also the counterfactual path of the treated patch's stock. As a result, the estimated treatment effect is similar to the true treatment effect (Fig. 4, *Bottom* and *SI Appendix*, Fig. S3, *Bottom*). As mobility increases, the negative shock to the treated patch has more pronounced effects on untreated patches, and the actual stock of a candidate control decreases relative to its counterfactual path and the counterfactual path of the treated stock (Fig. 4, *Middle* and *SI Appendix*, Fig. S2 *B–D*). This divergence leads to growing

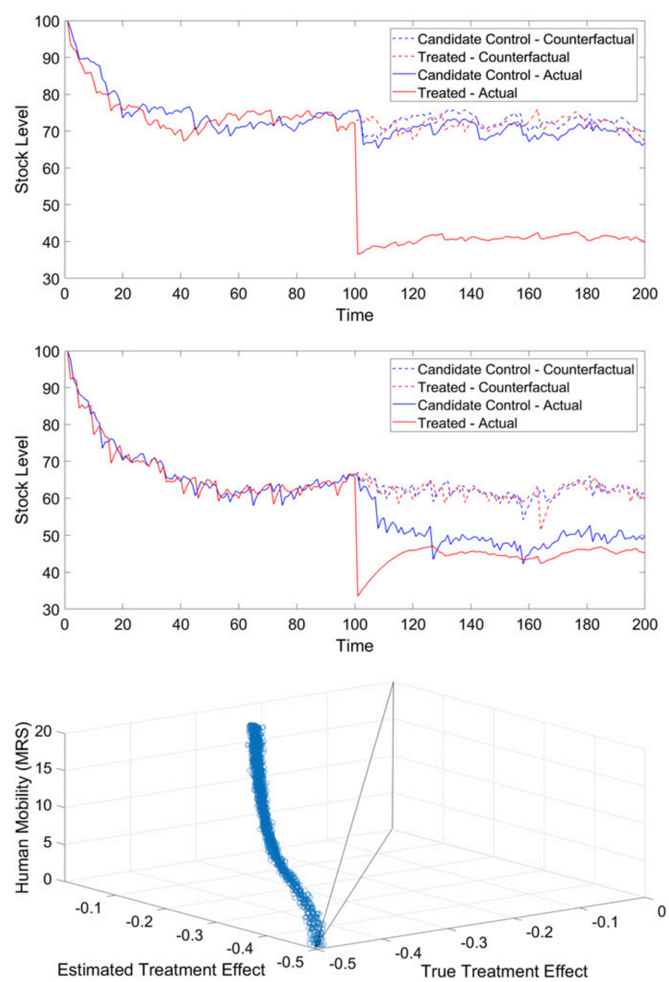

Fig. 4. Coupled human-natural system with a 50% negative shock in one zone and a candidate control zone within the same system. Treatment occurs in period 100. The degree of human mobility is the marginal rate of substitution of fishing revenue for travel cost (MRS). The counterfactual candidate control stock tracks the counterfactual treated stock. (*Top*) With low human mobility, the actual candidate control stock tracks the counterfactual treated stock, allowing identification of the treatment effect from actual data using a BACI design (difference-in-differences). (*Middle*) With high mobility, the actual candidate control stock decreases relative to the counterfactual treated stock, biasing downward the estimated treatment effects. The true treatment effect is also lower in the *Top* compared with the *Middle* because mobility transmits some of the actual effect of treatment to candidate control zones. (*Bottom*) As mobility increases, the true treatment effect decreases in magnitude and the bias toward zero in the estimated treatment effect increases. For an unbiased estimator, the scatter would follow the 45° plane depicted.

bias in the estimation of treatment effects as mobility increases (Fig. 4, *Bottom* and *SI Appendix*, Fig. S3, *Bottom*).

Mobility causes interference through its effect on candidate control stocks in the CHANS. A negative shock causes the treated patch to become less preferred as a fishing choice. With high mobility, fishers are more responsive to expected revenues in other locations, and so more fishers choose to fish in a candidate control patch. This extra fishing decreases the stocks in the candidate control patches relative to what they would have been had there been no shock (i.e., their counterfactual stock). Thus, the candidate control patch no longer represents the counterfactual stock of the treated patch. Dispersal of fishers over space in response to expected revenues is consistent with a large empirical literature on spatial behavior in fisheries, although the extent of responsiveness varies across empirical settings (26–31, 33).

The way in which mobility affects interference also moderates the magnitude of the treatment effect (Fig. 4 and *SI Appendix*, Figs. S2 and S3). With a negative shock, higher mobility results in a faster recovery of the stock (Fig. 4, *Top* and *Middle*), which means a smaller negative treatment effect from the shock (*SI Appendix*, Fig. S3, *Top*).

Human mobility thus creates two interconnected challenges for analyzing shocks to CHANS. First, as mobility increases, the size of the treatment effect on the treated zone decreases such that analysts are trying to detect an effect empirically that is increasingly subtle. Second, as mobility increases, the bias associated with interference increases because greater mobility means more contamination of the candidate control zones. In essence, the outcomes are more integrated across treated zones and candidate control zones when more of the shock is transmitted from the treated zone to candidate control zones. Visually, the actual and counterfactual paths compress as mobility increases (*SI Appendix*, Fig. S2). By depicting fishing effort without any shocks, we see another view of how this integration unfolds: With low mobility, fishing effort is roughly constant (low volatility) in each fishing zone, whereas with high mobility, it is highly reactive and bounces from one extreme to another with high mobility (*SI Appendix*, Fig. S4). This process is similar to how trading volume contributes to market integration (36).

The process of integration also points to the dynamics of the treatment effect. In the high-mobility system (Fig. 4, *Middle*), comparing the candidate control with the actual treated fishery in the period immediately after treatment would yield a larger estimated treatment effect with less bias. As time passes, the treated stock rises and the candidate control stock declines, introducing more bias in the estimation procedure. Dynamics of the treatment effect are also present in a simpler setting like the one in Fig. 2. Here, potential bias is more of a problem in the short-run dynamics, and the bias dissipates in the steady state (*SI Appendix*, Fig. S1). Together, these examples highlight the importance of when the treatment effect is measured and how theory and ancillary empirical information can assist causal inference.

The results for a positive shock in a closed ecological system are similar (*SI Appendix*, Figs. S5 and S6). As mobility increases, the system becomes more integrated (*SI Appendix*, Fig. S5) and the magnitude of the true treatment effect decreases, while bias from interference in the estimation increases (*SI Appendix*, Fig. S6). Despite the important potential role of human mobility in empirical analyses, fewer than 10% of the studies in our review of MPA impact studies mention human mobility as a potential problem.

Interference in the Ecologically Interconnected System. When the ecosystem has source-sink dispersal, human mobility is again both a cause of interference and a moderator of the treatment effect. In contrast to the closed system, low mobility in the source-sink system leads to more complex and nonmonotonic effects. Specifically, with very low mobility, the estimated treatment effects are biased upward (Fig. 5, *Top* and *SI Appendix*, Figs. S7*A* and S8). The mechanism reflects an interplay of economic and ecological gradients (24). Closing down the source patch leads to some economic dispersal to its paired sink patch (the treated patch), but also to some economic dispersal to the ecologically unconnected source and sink patches. The ecologically unconnected sink patch is the candidate control (Fig. 3, *Bottom*), so its stock slightly decreases relative to its counterfactual stock. As a result, the actual sink stock of the candidate control diverges from the counterfactual stock of the treated sink, creating a positive bias in the estimated treatment effects (Fig. 5, *Top*).

At slightly higher mobility, more economic dispersal takes place, but the overall system is more exploited because the fleet is more responsive to revenues, and the magnitude of the treatment effect is larger (*SI Appendix*, Figs. S6*B* and S7). This dampens the upward bias in the estimated treatment effects. It also reflects the well-understood theoretical result that a marine

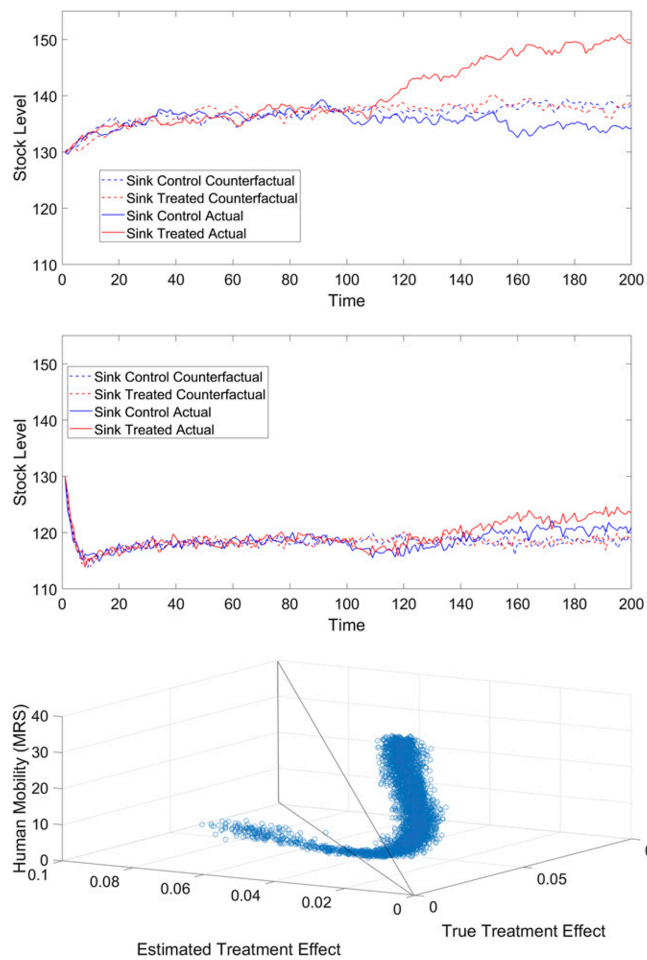

Fig. 5. Coupled human-natural system with two paired source and sink patches and a marine reserve placed in a source patch. Treatment occurs in period 100. The degree of human mobility is the marginal rate of substitution of fishing revenue for travel cost (MRS). The counterfactual candidate control stock tracks the counterfactual treated stock (*Top* and *Middle*, red and blue dashes). (*Top*) With low human mobility, the actual candidate control stock is below the counterfactual treated stock, introducing upward bias in the estimated treatment effects from actual data using a BACI design (difference-in-differences). (*Middle*) With high mobility, the true treatment effect is smaller, and the actual candidate control stock is above the counterfactual treated stock, such that the estimated treatment effects are biased downward. (*Top* and *Middle*) With higher mobility, the overall exploitation in the system is higher and the pretreatment stocks are lower. (*Bottom*) As mobility increases, the true treatment effect increases and then decreases in magnitude. For an unbiased estimator, the scatter would follow the 45° plane depicted. The cluster bends toward and then crosses the 45° plane. At low mobility, the estimated treatment effects are biased upward, and at high mobility, they are biased downward.

reserve can generate a win-win effect for the fishery when the ecological dynamics are source-sink dispersal and the baseline exploitation level is high (21, 24). To generate spillover benefits from closing a patch to fishing, the patch had to be heavily exploited before the closure; otherwise, the increased dispersal to the sink patch would be small.

At high mobility, the system becomes more integrated (i.e., stock paths are compressed) (*SI Appendix*, Fig. S7 *C* and *D*). This integration produces two countervailing effects: (*i*) The pre-reserve stocks are lower and can benefit more from the reserve (win-win effect), and (*ii*) the higher mobility level transmits more of the gains to the remaining open zones, including the candidate control zones (Fig. 5, *Middle* and *SI Appendix*, Fig. S7 *C* and *D*). The latter effect exacerbates interference to create a downward bias in the estimated treatment effects (Fig. 5, *Middle* and *SI Appendix*, Fig. S8, *Bottom*), and it reduces the true treatment effects (Fig. 5, *Bottom* and *SI Appendix*, Fig. S8, *Top*). The relationship between human mobility and system integration reflects a broad theme in CHANS research: As economic globalization progresses and the importance of global environmental change increases, couplings between human and natural systems tighten (12).

Conclusions

Careful consideration of excludability and interference is essential for making credible inferences about causal relationships in CHANS. Understanding these relationships is critical for creating an evidence base about the impacts of policy interventions, technological disasters, and ecological disturbances in CHANS. While CHANS are broader than marine systems and MPAs, we use the review of the large empirical MPA literature to illustrate that these issues are underappreciated. In fact, we found only one MPA study that addresses them both (37). Although these issues are widely omitted, one might legitimately question whether ignoring them matters for science and policy.

The terrestrial protected areas evaluation literature demonstrates that excludability violations are real problems, rather than abstract conceptual concerns. In this literature, a well-known pattern of treatment assignment exists: Protection is assigned to habitat that is less productive for alternative economic uses and near communities with low labor and capital mobility (19). This assignment pattern implies that not much land-use change would have occurred in the absence of protection, and thus terrestrial protected areas may have only modest effects on land-use change (38). These effects, however, will be overstated in studies that fail to address the ecological and economic factors that drive the nonrandom assignment. In one study (39), for example, the estimated treatment effect is between one-third and four-fifths smaller when the design directly addresses the nonrandom assignment compared with a simpler BACI design that is common in the ecology literature. To our knowledge, such comparisons have not been published in the literature on other CHANS. Without such comparisons, we cannot determine if the failure to take seriously the excludability assumption, as well as the potential for interference, is creating substantial bias in other CHANS studies. However, we can conclude that this failure makes the level and nature of MPA effects and other causal relationships in CHANS unclear.

Addressing nonrandom assignment requires structural knowledge of the CHANS, specifically the factors that affect variation in the exposure of units to the causal variable. A description of these factors is often called a description of "selection" or the treatment assignment mechanism. Such descriptions are almost entirely absent in CHANS studies. Moreover, in the absence of panel data with repeated observations before and after shocks, causal inference in CHANS will be a serious challenge. Panel data provide some scope for eliminating the confounding effects of fixed, unobservable attributes and time-varying observable attributes of the CHANS (20). Nevertheless, fully satisfying the excludability assumption in a nonexperimental CHANS design will be difficult. Thus, studies should also be complemented by sensitivity analyses or bounding approaches (partial identification) on inferences, which allow one to explore the implications of excludability violations (16).

Like violations of the excludability assumption, violations of the no-interference assumption are also likely to be real problems in CHANS. For example, in the US Gulf of Mexico, shrimp fishing vessels are highly mobile; large vessels routinely range from Louisiana waters to Texas waters. Empirical estimation of that mobility (6) implies substantial potential for interference in that system. Our simulation results predict that in such contexts, a negative shock on fish stocks will be masked by interference in typical causal inference designs. Consistent with this prediction, an evaluation of the impacts of hypoxia in the Gulf of Mexico using a BACI design like the one in Fig. 3, *Top*, could not detect any effect of the hypoxic event on landings (6).

Globalization, with its concomitant human and capital mobility, will exacerbate interference. Vessels in the global market for tuna and Alaskan groundfish, for example, are highly mobile, ranging over vast areas of the ocean (26, 28, 29). Moreover, large-vessel fishing effort is relatively unresponsive to fuel costs (40), suggesting high mobility. Other forces in the economy, such as declining transport costs and globalization of the seafood trade, reinforce the effects of high mobility (41, 42). In fact, a recent study that claims "oceanic isolation" moderates the impact of MPAs in a positive direction may simply be revealing that designs using isolated MPAs as the treated units are less subject to interference (43).

To address interference in CHANS, scholars have several options (details are provided in *SI Appendix*): (*i*) Acknowledge interference exists in the design and redefine the estimand to include interference; (*ii*) use structural knowledge of the CHANS to estimate the likely bias from interference and bound the true treatment effect from above or below; (*iii*) use experimental or quasiexperimental saturation designs to detect and estimate an interference function [if interference is not interesting from a policy perspective, one can use the estimated function to strip the effects of interference from the causal estimates; otherwise, one can use the estimated function to create a structural model for simulations (scenario projections under different treatment assignment permutations)]; and (*iv*) assume the population can be partitioned into groups or clusters, with the interference limited to units within the same cluster. This latter approach could be as complicated as graph-cutting methods from the network literature or as simple as choosing comparison sites far away from treated sites.

Like excludability, interference also has a connection to the nonrandom assignment of treatments. When the treatment assignment is negatively correlated with the causes of interference, we expect lower levels of interference. As noted above, terrestrial protected areas are often sited in zones with low mobility. Thus, we would predict that interference is unlikely to be a problem. The literature appears to confirm this prediction. To our knowledge, no terrestrial protected area studies that have tried to detect interference have found strong evidence of it. Thus, were MPAs assigned in the same way that terrestrial protected areas are assigned, interference in MPA studies would be less problematic than, for example, studies of the causal effects of hypoxic events.

No perfect solutions exist for satisfying the excludability and no-interference assumptions. To elucidate causal relationships in CHANS, multiple approaches will be needed, with the aim of identifying sources of bias in each approach for a given causal question and then triangulating on credible inferences (44). For example, a researcher might use structural (e.g., mechanistic modeling, such as empirical bioeconomic modeling or inverse modeling in ecology) and reduced-form predictive inference (e.g., time-series analysis, machine learning) approaches to aid in causal inference in CHANS.

Structural modeling can contribute to causal inference in at least five ways. First, a structural model, like an empirical bioeconomic model, can generate theoretically grounded *ex ante* predictions about how the system will respond to treatment (25). Modeling anticipated outcomes is important in CHANS because these systems involve feedbacks and nonlinearities that are difficult to synthesize without the presence of a formal model. Second, a structural model can reveal how sensitive outcome variables are to underlying parameters of the bioeconomic system, such as a fish stock's intrinsic growth rate (*SI Appendix*, Fig. S1). Using the model in this way can help researchers to select candidate controls that are consistent with the untestable excludability assumption. Third, a structural model can provide an alternative *ex post* estimate of the treatment effect, albeit one that is based on a number of mechanistic assumptions (45). The alternative estimate could be compared with an estimate from a BACI design to triangulate, or the structural model estimate might be the only way to estimate a treatment effect because a viable control is lacking. Fourth, a structural model can help to diagnose the severity of interference in a design. In the Gulf of Mexico hypoxia case, an empirical bioeconomic model was able to determine how much human mobility was necessary to create severe bias in a BACI design. A separate estimate of fishing fleet behavior then demonstrated that actual fleet mobility far exceeded this threshold. Fifth, structural modeling can be used to generate indirect tests of causal hypotheses using time series data. In fact, when there is high interference in reduced-form causal models due to high human mobility, indirect tests may be more powerful because high mobility leads to tighter mechanistic coupling, and thus stronger predictions about the behavior of time series. In the Gulf of Mexico hypoxia case, null findings from a BACI design attributed to high mobility prompted the use of a market counterfactual to isolate the hypoxia signal in price data. To generate and test predictions about the time series of shrimp prices, the analysis combined structural information about seafood markets with structural information about the ecological effects of hypoxia (6).

The CHANS causality literature is characterized by two opposing features: the challenge of causal inference in CHANS is formidable, yet the most widespread approaches to causal inference in the literature are rudimentary. This striking juxtaposition is a serious obstacle to advancing the science of CHANS and to accumulating an evidence base on which the effective management of CHANS is possible. As with all obstacles in sustainability science, overcoming it will require skills and knowledge from many disciplines and effective academic-practitioner collaborations.

ACKNOWLEDGMENTS. This research was supported by the National Science Foundation Grant CNH 1715638.

1. Liu J, et al. (2007) Complexity of coupled human and natural systems. *Science* 317:1513–1516.
2. Brander JA, Taylor MS (1998) The simple economics of Easter Island: A Ricardo-Malthus model of renewable resource use. *Am Econ Rev* 88:119–138.
3. Carpenter SR, et al. (2009) Science for managing ecosystem services: Beyond the Millennium Ecosystem Assessment. *Proc Natl Acad Sci USA* 106:1305–1312.
4. Fenichel EP, et al. (2011) Adaptive human behavior in epidemiological models. *Proc Natl Acad Sci USA* 108:6306–6311.
5. Horan RD, Fenichel EP, Drury KL, Lodge DM (2011) Managing ecological thresholds in coupled environmental-human systems. *Proc Natl Acad Sci USA* 108:7333–7338.
6. Smith MD, et al. (2017) Seafood prices reveal impacts of a major ecological disturbance. *Proc Natl Acad Sci USA* 114:1512–1517.
7. Smith MD, Sanchirico JN, Wilen JE (2009) The economics of spatial-dynamic processes: Applications to renewable resources. *J Environ Econ Manage* 57:104–121.
8. Smith VL (1969) On models of commercial fishing. *J Polit Econ* 77:181–198.
9. Smith VL (1975) The primitive hunter culture, Pleistocene extinction, and the rise of agriculture. *J Polit Econ* 83:727–755.
10. Turner BL, 2nd, et al. (2003) Illustrating the coupled human-environment system for vulnerability analysis: Three case studies. *Proc Natl Acad Sci USA* 100:8080–8085.
11. Werner B, Mcnamara DE (2007) Dynamics of coupled human-landscape systems. *Geomorphology* 91:393–407.
12. Liu J, et al. (2007) Coupled human and natural systems. *Ambio* 36:639–649.
13. Ostrom E (2009) A general framework for analyzing sustainability of social-ecological systems. *Science* 325:419–422.
14. Sugihara G, et al. (2012) Detecting causality in complex ecosystems. *Science* 338:496–500.
15. Ferraro PJ, Hanauer MM (2014) Advances in measuring the environmental and social impacts of environmental programs. *Annu Rev Environ Resour* 39:495–517.
16. Ferraro PJ, Hanauer MM (2015) Through what mechanisms do protected areas affect environmental and social outcomes? *Philos Trans R Soc Lond B Biol Sci* 370:20140267.
17. Imbens GW, Rubin DB (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge Univ Press, New York).
18. Morgan SL, Winship C (2014) *Counterfactuals and Causal Inference* (Cambridge Univ Press, New York).
19. Ferraro PJ, Hanauer MM, Sims KR (2011) Conditions associated with protected area success in conservation and poverty reduction. *Proc Natl Acad Sci USA* 108:13913–13918.
20. Ferraro PJ, Miranda JJ (2017) Panel data designs and estimators as substitutes for randomized controlled trials in the evaluation of public programs. *J Assoc Environ Resour Econ* 4:281–317.
21. Smith MD, Lynham J, Sanchirico JN, Wilson JA (2010) Political economy of marine reserves: Understanding the role of opportunity costs. *Proc Natl Acad Sci USA* 107:18300–18305.
22. Osenberg CW, Bolker BM, White J-SS, St Mary C, Shima JS (2006) Statistical issues and study design in ecological restorations: Lessons learned from marine reserves. *Foundations of Restoration Ecology*, eds Falk DA, Palmer MA, Zedler JB (Island Press, Washington, DC), pp 280–302.
23. Sanchirico JN, Wilen JE (1999) Bioeconomics of spatial exploitation in a patchy environment. *J Environ Econ Manage* 37:129–150.
24. Sanchirico JN, Wilen JE (2001) A bioeconomic model of marine reserve creation. *J Environ Econ Manage* 42:257–276.
25. Smith MD, Wilen JE (2003) Economic impacts of marine reserves: The importance of spatial behavior. *J Environ Econ Manage* 46:183–206.
26. Abbott JK, Wilen J (2011) Dissecting the tragedy: A spatial model of behavior in the commons. *J Environ Econ Manage* 62:386–401.
27. Eales J, Wilen JE (1986) An examination of fishing location choice in the pink shrimp fishery. *Mar Resour Econ* 2:331–351.
28. Haynie A, Layton DF (2010) An expected profit model for monetizing fishing location choices. *J Environ Econ Manage* 59:165–176.
29. Hicks RL, Schnier KE (2008) Eco-labeling and dolphin avoidance: A dynamic model of tuna fishing in the Eastern Tropical Pacific. *J Environ Econ Manage* 56:103–116.
30. Holland D, Sutinen JG (2000) Location choice in the New England trawl fisheries: Old habits die hard. *Land Econ* 76:133–149.
31. Smith MD (2002) Two econometric approaches for predicting the spatial behavior of renewable resource harvesters. *Land Econ* 78:522–538.
32. Stafford TM (2018) Accounting for outside options in discrete choice models: An application to commercial fishing effort. *J Environ Econ Manage* 88:159–179.
33. Zhang J, Smith MD (2011) Heterogeneous response to marine reserve formation: A sorting model approach. *Environ Resour Econ* 49:311–325.
34. Fletcher WJ, Kearney RE, Wise BS, Nash WJ (2015) Large-scale expansion of no-take closures within the Great Barrier Reef has not enhanced fishery production. *Ecol Appl* 25:1187–1196.
35. Smith MD, Zhang J, Coleman FC (2006) Effectiveness of marine reserves for large-scale fisheries management. *Can J Fish Aquat Sci* 63:153–164.
36. Bekaert G, Harvey CR (1995) Time-varying world market integration. *J Finance* 50:403–444.
37. Reimer MN, Haynie AC (2018) Mechanisms matter for evaluating the economic impacts of marine reserves. *J Environ Econ Manage* 88:427–446.
38. Ferraro PJ, Pressey RL (2015) Measuring the difference made by conservation initiatives: Protected areas and their environmental and social impacts. *Philos Trans R Soc Lond B Biol Sci* 370:20140270.
39. Andam KS, Ferraro PJ, Pfaff A, Sanchez-Azofeifa GA, Robalino JA (2008) Measuring the effectiveness of protected area networks in reducing deforestation. *Proc Natl Acad Sci USA* 105:16089–16094.
40. Kroodsma DA, et al. (2018) Tracking the global footprint of fisheries. *Science* 359:904–908.
41. Asche F, Smith MD (2018) Induced innovation in fisheries and aquaculture. *Food Policy* 76:1–7.
42. Hummels D (2007) Transportation costs and international trade in the second era of globalization. *J Econ Perspect* 21:131–154.
43. Edgar GJ, et al. (2014) Global conservation outcomes depend on marine protected areas with five key features. *Nature* 506:216–220.
44. Munafo MR, Smith GD (2018) Robust research needs many lines of evidence. *Nature* 553:399–401.
45. Smith MD, Zhang J, Coleman FC (2008) Econometric modeling of fisheries with complex life histories: Avoiding biological management failures. *J Environ Econ Manage* 55:265–280.
46. Ferraro PJ, Hanauer MM (2014) Quantifying causal mechanisms to determine how protected areas affect poverty through changes in ecosystem services and infrastructure. *Proc Natl Acad Sci USA* 111:4332–4337.