Article:
Law, E.A., Ferraro, P.J., Arcese, P. et al. (2017) Projecting the performance of conservation interventions. Biological Conservation, 215. pp. 142-151. ISSN: 0006-3207

https://doi.org/10.1016/j.biocon.2017.08.029

This is the peer reviewed version of the following article: Law, E. A., Ferraro, P. J., Arcese, P., Bryan, B. A., Davis, K., Gordon, A., Holden, M. H., Iacona, G., Marcos Martinez, R., McAlpine, C. A., Rhodes, J. R., Sze, J. S. and Wilson, K. A. (2017) Projecting the performance of conservation interventions. Biological Conservation, 215 142-151. doi:10.1016/j.biocon.2017.08.029.

Projecting the performance of conservation interventions

Authors:

Elizabeth A. Law[1,2,*], Paul Ferraro[3], Peter Arcese[4], Brett A. Bryan[5,6], Katrina Davis[1,2,7], Ascelin Gordon[8], Matthew H. Holden[1,2], Gwenllian Iacona[1,2], Raymundo Marcos Martinez[5], Clive McAlpine[1,9], Jonathan R. Rhodes[1,9], Jocelyne S. Sze[10], and Kerrie A. Wilson[1,2].

[1] ARC Centre of Excellence for Environmental Decisions, The University of Queensland, Australia

[2] School of Biological Sciences, The University of Queensland, Australia

[3] Bloomberg School of Public Health, Carey Business School, Whiting School of Engineering, Johns Hopkins University, Maryland

[4] Department of Forest and Conservation Sciences, the University of British Columbia, Canada

[5] CSIRO Land and Water, Waite Campus, Urrbrae, Australia

[6] School of Life and Environmental Sciences, Deakin University, Burwood, Australia

[7] School of Agricultural & Resource Economics, University of Western Australia, Australia

[8] School of Global, Urban and Social Studies, RMIT University, Australia

[9] School of Earth and Environmental Sciences, The University of Queensland, Australia

[10] Dept of Life Sciences (Silwood Park) Imperial College London, Buckhurst Road, Ascot SL5 7PY, United Kingdom

* Corresponding author:

Elizabeth A. Law

School of Biological Sciences, The University of Queensland, QLD, 4072, AUSTRALIA.

Ph: +1 250 278 4339

E: e.law@uq.edu.au

Abstract

Successful decision-making for environmental management requires evidence of the performance and efficacy of proposed conservation interventions. Projecting the future impacts of prospective conservation policies and programs is challenging due to a range of complex ecological, economic, social and ethical factors, and in particular the need to extrapolate models to novel contexts. Yet many extrapolation techniques currently employed are limited by unfounded assumptions of causality and a reliance on potentially biased inferences drawn from limited data. We show how these restrictions can be overcome by established and emerging techniques from causal inference, scenario analysis, systematic review, expert elicitation, and global sensitivity analysis. These technical advances provide avenues to untangle cause from correlation, evaluate and transfer models between contexts, characterize uncertainty, and address imperfect data. With more rigorous projections of prospective performance of interventions, scientists can deliver policy and program advice that is more scientifically credible.

Keywords (6): Causal inference, evidence-based policy, policy evaluation, prediction, projection, transportability.

Highlights

- Conservation policies and programs are hotly debated, with complex, uncertain impacts.
- To make informed decisions, reliable projections of the likely performance of interventions are required.
- Robust projections need to focus on model assumptions, bias and uncertainty.
- Clarifying causal assumptions will lead to better data and better use of data.

1. Introduction

Reliable evidence of future performance and efficacy of interventions is a critical component of successful decision-making for environmental management (Ferraro and Pattanayak, 2006; Rissman and Smail, 2015). Examples of such decision-making include achieving global protected area targets (Visconti et al., 2015), designing new national-level payments for ecosystem services programs (Bryan et al., 2014), and controlling invasive species (Firn et al., 2015; Martin et al., 2015). Yet determining future impacts of conservation interventions is challenged by a range of complex ecological, economic, social and ethical factors, as well as trade-offs between multiple objectives. Increasingly, scholars and practitioners are more systematically collating and synthesizing existing literature on past impacts for use as an evidence base in conservation (Sutherland et al., 2004). But making accurate inferences from this relies on the quality of this evidence base. Researchers and practitioners are also seeking to improve the quality of this evidence by conducting more robust assessments of past policy impacts through *retrospective* evaluations (Miteva et al., 2012; Pressey et al., 2015; Baylis et al., 2016). These retrospective evaluations typically use principles of causal inference (Box 1), which focuses on clarifying the assumptions needed to infer causal relationships from data, and on reducing the bias of impact estimates (Miteva et al., 2012; Meyfroidt, 2015; Pressey et al., 2015). This movement towards enhanced transparency and reduced bias is a response to the historical deficiencies of retrospective policy evaluations in conservation science (Ferraro and Hanauer, 2014; Meyfroidt, 2015; Baylis et al., 2016).

Yet when used to inform the design of conservation policies and interventions, retrospective evaluations only tell half the story: predictions of expected outcomes are also necessary. While 'improving future policy and interventions' is a commonly stated goal of retrospective analyses (Baylis et al., 2016), rigorous analysis of past outcomes alone is insufficient for this purpose. Evidence from past interventions can be highly context-specific (Pfaff and Robalino, 2012), and may not extrapolate to other times and areas (Sinclair et al., 2010; Dobrowski et al., 2011; Cook et al., 2014; Oliver and Roy, 2015). Such extrapolation is traditionally the domain of projection analyses: the use of modelling to project intervention impacts across time and space.

If, in developing projections, analysts ignore the new insights and methods of retrospective evaluations, the advice yielded by these projections will lack scientific credibility. Scientific credibility refers to the plausibility and technical accuracy of the science. Implicit and untested assumptions regarding causality limit the credibility of prospective policy analysis, as associations observed in the past may not hold in the future (Meyfroidt, 2015). Scientific credibility may also be limited if projections rely on potentially biased inferences from limited data (Miteva et al., 2012; Pressey et al., 2015), and which have an unclear treatment of uncertainty or poor interpretation of potentially biased results. These issues of untested assumptions, limited data, and imperfect use of this data are important for successful conservation decision-making: overestimation of benefits associated with proposed conservation interventions may lead to sub-optimal outcomes, whereas underestimation of benefits may result in more effective options being overlooked.

Here, we outline the relevance, benefits, and challenges of integrating into prospective evaluation of conservation interventions the principles of causal inference and associated principles of systematic literature review, expert elicitation, and scenario analysis. We discuss how these established and emerging techniques can be employed to (1) improve problem definition by clarifying causal assumptions, key variables, alternative scenarios, and using appropriate model frameworks, (2) improve model parameterization by identifying potential bias in data, and avoiding these where possible, and (3) improve model use and interpretation through analyses to understand model sensitivity and parameter or model uncertainty. These techniques are designed to encourage conservation scientists to use and interpret imperfect data more effectively, thereby delivering policy and program advice that is more scientifically credible, and, if heeded by decision-makers and acceptable to stakeholders, capable of delivering improved conservation outcomes.

2. Problem definition: clarifying causal assumptions

2.1 Characterizing key variables in a causal context

A key challenge in creating robust and transparent model projections of conservation interventions is to define the problem. How is the intervention expected to work within the environmental, social, and economic context? To answer this question, models that depict mechanism-based, causal relationships between interventions, processes and variables are developed, ideally explicitly and graphically (Pearl, 2009;

Margoluis et al., 2013)(Box 2). Causal relationships between key variables may be supported by a variety of evidence (Meyfroidt, 2015), or be based on hypotheses. While defining the 'treatment' and 'outcome' in a graphical model may appear trivial, there is a challenge in explicitly identifying treatments and outcomes that are relevant across a wide social and environmental spectrum (Meyfroidt, 2015; Pressey et al., 2015). Graphical models are useful, especially when sufficiently informative and detailed to enable elucidation of assumed causal impacts through potentially complex causal pathways (Firn et al., 2015), and characterize variables as confounding factors, mechanisms, and moderators (see Box 1; Box 2)(Ferraro and Pressey, 2015; Meyfroidt, 2015; Pressey et al., 2015).

2.2 Establishing valid baselines and alternative scenarios

Projections aim to determine potential future impact; that is, the difference between alternative future states, typically arising from a 'baseline' and alternative scenarios (Bryan et al., 2014; Bull et al., 2014; Oliver and Roy, 2015; Visconti et al., 2015). Future scenarios are hypotheses of how a system may operate under different conditions or assumptions; a set of functions and parameters that lead to potential future states. Baselines are commonly set as a continuation of current or historical conditions, or as a projection of the 'most likely' or 'business as usual' scenario (Bull et al., 2014). In prospective analyses, predicting impacts is more difficult than in retrospective analyses, as there is not yet a 'fact' for scenarios to run counter to: future scenarios cannot be directly observed. Therefore while retrospective analyses have an observable, factual case against which to compare constructed alternative scenarios to, in prospective analyses both alternative scenarios and baselines must be constructed through assumptions and narrative. Care needs to be taken not to construct 'straw man' arguments (i.e. impossible or highly improbable scenarios) and thereby give the false impression that a particularly positive or negative outcome is likely. This does not mean that more qualitative descriptions of 'futures' (e.g. Coreau et al., 2009) are not valuable, but rather emphasizes the need to transparently communicate the assumptions of each scenario: as variation in scenario definition can substantially change recommendations (Bull et al., 2014; Visconti et al., 2015), robust prospective evaluation requires clearly articulated, conscientious and defendable definitions of baselines and alternative scenarios (Pressey et al., 2015; Visconti et al., 2015). Ideally, projections should be analyzed over a set of scenarios that (to some extent) approximates the full set of plausible states of the modelled system, thereby accounting for relevant exogenous uncertainties, discontinuities and dynamics of the system being modeled (Bryant and Lempert, 2010; Kasprzyk et al., 2013; Kwakkel et al., 2013). The evaluation of these scenarios informs not only the bounds and mean impacts of specific treatments but also the regions in the parameter space where relevant outcomes could be achieved (Gerst et al., 2013; Lempert, 2013).

2.3 Choosing an appropriate model framework, given causal assumptions

Understanding and explicitly articulating the causal relationships that are implicit within a model framework helps to explain the key differences between different modeling approaches. Here, we illustrate this idea using the Species Area Relationship (Box 2) as an example of the causal assumptions underlying three types of models commonly used in making future projections: (1) exploratory models with many variables (i.e. ‘kitchen sink’ models), (2) ‘reductionist’ models, and (3) ‘all-cause’ models.

Exploratory (‘kitchen-sink’) models aim to identify associations between multiple variables and an outcome. Such models are useful for hypothesis generation, and are commonly used in simple multiple regression-type analyses. However, several assumptions often made by simple regression analyses and other correlation-based procedures limit the usefulness of these types of models in future projections. First, causal effects among the predictors are not required, and therefore describing correlates as drivers or determinants, and their coefficients as effects or impacts, represents an often untested causal assumption (Meyfroidt, 2015). Second, it is often implicitly assumed that there is no specification error (no incorrect functional forms, or missing predictors), which can bias impact estimators from regression analyses in potentially uncertain ways (Kline, 2015). Third, if models are parameterized based on correlation, rather than causation, there can be little *a priori* confidence that these relationships remain constant when projected (Oliver and Roy, 2015). This problem is demonstrated by the poor performance of spatial and temporal projections of some species distribution models based on bioclimatic correlates (Sinclair et al., 2010; Dobrowski et al., 2011). While some effort is usually made to select variables in exploratory studies based on a theoretical or empirical understanding of the system, causal pathways need to be made much more explicit within the design of the analysis to more robustly infer causality (Gelman and Hill, 2006; Pearl, 2009; Ferraro and Hanauer, 2014).

‘Reductionist’ models focus on reduced model complexity and are common in retrospective causal inference analyses (particularly for quasi-experiments to estimate ‘counterfactuals’) (Bollen and Pearl, 2013; Ferraro and Hanauer, 2014). A benefit is that they do not require the full model to be specified: the focus is on developing a reliable estimator of the effect of a specific cause, rather than estimating marginal impacts of all potential covariates (Ferraro and Pattanayak, 2006; Ferraro and Hanauer, 2014). Potential covariates are not ignored: the analysis focuses on controlling for covariates that affect both the outcome and exposure to the cause – in other words, confounding variables. To exert such control, this type of analysis will often ‘match’ samples from treated and untreated populations to balance confounding covariates, and thereby limit the bias they may have on the impact estimator (Jones and Lewis, 2015; Meyfroidt, 2015). Similarly, the construction of a ‘synthetic control’ makes this approach practical for assessing specific causal impacts of conservation interventions where there is only one ‘treatment’ sample (Sills et al., 2015). The emphasis on internal validity (the minimization of bias) in reductionist models means care must be taken when projecting these estimates of causal impacts to novel contexts: estimates are typically specific to certain sub-populations (Box 3), though understanding what factors moderate impacts can help refine projections across heterogeneous contexts (Ferraro et al., 2011; Ferraro et al., 2015).

‘All-cause’ models embrace the complexity of a larger graphical model framework, offering a practical compromise between regression and reductionist models, and include developments that increase their utility for making projections. All-cause models include structural equation models (Shipley, 2002; Lamb et al., 2014), Bayesian Networks (Martin et al., 2015; Pascual et al., 2016), and Structural Causal Models (Pearl, 2009; Runge et al., 2015). Structural equation models inherently describe a graphical model framework (Bollen and Pearl, 2013; Martin et al., 2015). This allows structural equation models to incorporate unobservable (latent) variables, tolerate uncertainty in the model predictors, and differentiate between direct and indirect effects of treatments (Bollen and Pearl, 2013). They therefore offer a useful option for developing elaborate causal models (Box 3). Bayesian Networks can offer similar benefits and more easily incorporate alternative types of information such as expert opinion (Pascual et al., 2016). Structural equation models and Bayesian Networks typically assume the correct model is specified, though Hyttinen et al. (2015) suggest methods for incorporating model uncertainty. Structural Causal Modeling extends and increases the utility of graphical models, and represents a growing area of research and theory (Pearl, 2009; Runge et al., 2015). For example, structural causal models can identify critical network nodes and interactions (Runge et al., 2015), and emerging theory on transportability of Structural Causal Models may facilitate more confident model transfer from well-studied to less well-studied species and populations (Pearl, 2009; Bareinboim and Pearl, 2013) (Box 3).

3. Parameterization: using better data

Biases are pervasive in empirical conservation research because this research is often conducted in contexts of strong personal motivations, extremely low rates of study replication, complex systems, and high intrinsic rates of variability (Iftekhar and Pannell, 2015). Causal inference, systematic literature reviews, and robust expert elicitation methods offer ways to identify and mitigate biases in data drawn from a wide variety of sources (Martin et al., 2012b; Cook et al., 2014; Martin et al., 2014; Martin et al., 2015; Pascual et al., 2016).

3.1 Identifying bias in observational and experimental data

Concepts of bias have been long discussed in ecology; for example it is recognized that even the idealized 'gold standard' of experimental design, randomized controlled trials, can also be subject to 'demonic' and 'non-demonic' biases (Hurlbert, 1984). Demonic bias derives from foreseeable causes and can impact an experiment if the design of the experiment does not adequately control for this during sample selection. Non-demonic bias is derived from chance events that occur while an experiment is in progress. A sample that is unrepresentative of the population of interest may often be a source of bias. This 'selection bias' may arise when selection occurs non-randomly due to certain sub-populations being specifically selected for treatment, self-selecting for treatment, or more susceptible to sample attrition, for example in pilot programs, prioritization, or voluntary participation. Dealing with bias is not about more advanced statistical methods, rather, it should focus on experimental design (Ferraro and Pressey, 2015; Jones and Lewis, 2015; Baylis et al., 2016), and conscientious interpretation of results to avoid confirmation bias and 'just-so' storytelling (Nuzzo, 2015). Confirmation bias describes a cognitive bias in which people selectively collate, interpret, present, and recall information that support their beliefs or hypotheses, and give disproportionately less consideration to alternative possibilities. Confirmation bias is particularly common in emotionally charged issues or when beliefs are entrenched. 'Just-so' storytelling is an *ad hoc* fallacy, a narrative explanation of facts made after the event, and therefore contemporarily unverifiable and unfalsifiable. These explanations are not necessarily wrong, rather they are hypotheses that require further assessment. An understanding of potential sources of bias and how causal inference methods (Ferraro and Pattanayak, 2006; Miteva et al., 2012; Fisher et al., 2014; Ferraro and Pressey, 2015) can address these issues is useful for researchers and practitioners designing experiments as well as researchers collating data from the published literature. These approaches facilitate the identification and treatment of potential bias, and appraisal of the rigour of experimental results.

3.2 Recognizing biases in collated data: robust systematic review and expert elicitation

Additional sources of bias become relevant when collating parameter values from published research. Several biases are common when drawing data from a single source, including bias towards parameters used by previous similar work or that have been highly cited, towards the most recent analyses, or to a parameter that favorably supports the researcher's position (i.e. 'confirmation bias') (Haddaway et al., 2015; Nuzzo, 2015). To avoid these biases, many researchers turn towards a literature review. However, bias can be inherent in the literature, as well as resulting from personal biases of the researcher selecting and interpreting the literature (Stocks et al., 2008; Martin et al., 2012a; Martin et al., 2012b; Haddaway et al., 2015; McKinnon et al., 2015; Nuzzo, 2015). Such problems are further compounded in expert elicitation, where bias may be present in the published evidence base (Stocks et al., 2008; Martin et al., 2012a), and in an experts' experience and translation of this evidence base (Iftekhar and Pannell, 2015; Nuzzo, 2015). Further, in expert elicitation there are substantial challenges in designing an elicitation procedure that is robust to biases (Martin et al., 2012b; Firn et al., 2015). Such biases may originate from the confidence of individual experts and the social dynamics of the expert group (Martin et al., 2012b), from personal preferences and perceptions (e.g. optimism, pessimism, or loss aversion), or from limitations on rationality, including framing effects, reference-point bias, or reliance on limited or available information (Iftekhar and Pannell, 2015). Substantial advances have been made in the field of systematic review methodology, providing guidelines on how literature can be comprehensively sampled, consistently evaluated, and evidence appropriately weighted in synthesis (Collaboration for Environmental Evidence, 2013; McKinnon et al., 2015). While a full systematic review for every parameter may not by warranted (Addison et al., 2013), it is relatively easy to integrate the principles of systematic review into workflows (Haddaway et al., 2015). Similarly, expert elicitation methods have been developed (Martin et al., 2012b), and are increasingly applied as modes to source information where data are lacking (Firn et al., 2015; Martin et al., 2015), as is often the case when developing novel conservation policies or interventions (McKinnon et al., 2015).

4. Interpretation: using data better

Biases may still be unavoidable even with greater attention to experimental design and analysis, systematic review procedures, and rigorous expert elicitation methods. For example, bias is likely in regional or global scale analyses, when data are not necessarily collected for the specific purpose of the evaluation (McKinnon et al., 2015). However, if data shortcomings are made transparent, improvements in model specification and interpretation may be possible. Model and data imperfections can influence the design of sensitivity and uncertainty analyses, inform model transportability to novel contexts, and indicate the usefulness of partial identification to explore the influence of assumptions on the results. In this section we outline key methods for dealing with data interpretation issues, including sensitivity analyses and partial identification.

4.1 Dealing with imperfect data and data uncertainty

Techniques for dealing with imperfect data and parameter uncertainty have centered on sensitivity and uncertainty analyses. Sensitivity analysis aims to characterize how variation in model inputs cause changes in model outputs (Saltelli and Annoni, 2010). It determines which model input parameters are most influential, and identifies where reducing model uncertainty might improve model performance. In practice, sensitivity analysis is often carried out by varying one parameter at a time from a given baseline parameterization, often within some specified variation (e.g. one standard deviation) from the mean parameter estimate. This 'one-at-time' approach can be misleading as most of the model input parameter space remains unexplored, and is particularly problematic when there are non-linear interactions between parameters (Saltelli and Annoni, 2010). Global sensitivity approaches, which vary multiple parameters simultaneously to account for possible interactions and nonlinear responses, are generally preferable (Saltelli and Annoni, 2010).

Uncertainty analysis aims to provide confidence bounds on a model output (or its probability density function). In practice, determining output uncertainty can be similar to a global sensitivity analysis, however the focus of uncertainly analysis is not on the extent to which parameters are causing changes in model output, but on how uncertainty in all model inputs propagates through the model and results in uncertainty in the output (Norton, 2015). In developing projections, an ideal rigorous uncertainty analysis would account for the full uncertainty in all model input parameters as well as structural uncertainty in the underlying model. Rigorous uncertainty analyses allow for defensible confidence intervals on model projections, in particular when modelling specific alternative scenarios, as the size of these confidence intervals will determine whether a model predicts a statistically significant impact. It also allows best and worst case outcomes to be identified, explicitly allowing levels of risk aversion to be incorporated into decisions made using the model projections.

Partial identification is an alternative or complementary method for dealing with uncertainty regarding assumptions (Box 4) (Manski, 2007). In retrospective analyses, this method systematically explores the implications of assumptions regarding the counterfactual on the range of impact estimates (the identification region) thereby addressing questions of uncertainty and potential bias that relate to these (Manski, 2007). For prospective analyses, partial identification can be particularly useful to give bounds on parameter estimates when there is uncertainty or controversy regarding potential impacts of interventions (McConnachie et al., 2015).

5. Synthesis and ways forward

To support the development of conservation interventions in complex environmental, social, economic, and ethical contexts, transparent, evidence-based models are critical. More transparent assumptions and more believable causal models engender greater confidence in the predictions of prospective evaluations, and these predictions will be more justifiable in the face of critique. This confidence in the robustness of the science is, of course, only one element contributing to the wider salience, legitimacy, and other forms of credibility of policy advice and of policies themselves (Cash et al. 2003; Clark et al. 2016; Posner et al. 2016), but it is an important element to maintain public trust in science. Poor data, inappropriate models, erroneous assumptions, and bias lead to advice that may systematically over or under-estimate the impacts of policies or programs. Techniques drawn from causal inference, scenario analysis, systematic literature review, and expert elicitation can help to recognize and reduce the inevitable bias and uncertainty in analysing the likely impacts of conservation interventions (Figure 1). Further, when models need to be extrapolated to novel contexts, emerging techniques of structural causal modelling (including transportability theory) and of partial identification could be integrated into projections of conservation policy and thereby enhance the robustness of results and their interpretation.

In modelling the projected impacts of conservation interventions, a more diverse array of tools and approaches is warranted. We acknowledge that the tools and approaches reviewed here may not all be necessary for every prospective modelling situation, or may not always be time or cost-effective in delivering better policy and program advice in every context. For example, in relatively simple, widely studied, and non-controversial contexts, lengthy and elaborate fine-scale projection models may not be required. However, even in these cases transparently clarifying model causal assumptions, considering potential bias in parameter data, and conducting simple uncertainty and sensitivity analysis may add little to no additional cost and result in more confidence in the robustness of resulting policy advice. In more complex, uncertain and controversial contexts, ignoring these advances in causal inference and associated techniques will ensure that the current deficiencies in prospective evaluations will remain. Broader recognition and uptake of these tools and approaches will help to develop more scientifically credible projections of impacts, and thereby, if heeded in policy development, better outcomes for conservation.

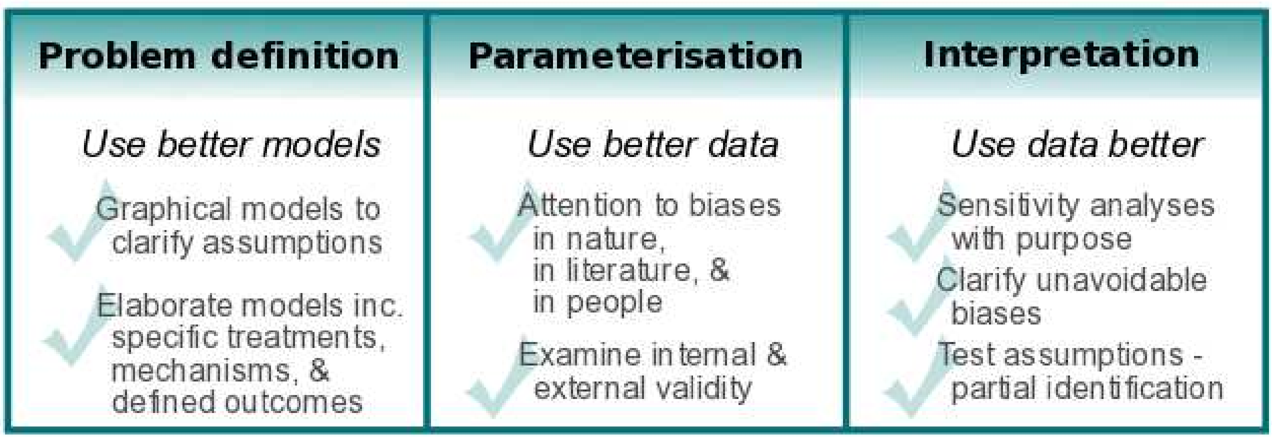

Figure 1: An overview of the methods available to enhance the quality of model projections.

Acknowledgements

Funding from the Australian Research Council (http://www.arc.gov.au/) is acknowledged, including Centre of Excellence (CE110001014)(EAL, KD, MHH, GI, CM, JR, KAW), Discovery (DP150101300)(EAL, KAW), and Future Fellowship (FT100100413)(KAW) programs. This manuscript is based on discussions from the interdisciplinary workshop "Causal inference in environmental decisions" organized in Brisbane, Australia, June 2015, attended by the authors and funded by the ARC Centre of Excellence for Environmental Decisions.

References

Addison, P.F., Rumpff, L., Bau, S.S., Carey, J.M., Chee, Y.E., Jarrad, F.C., et al. (2013). Practical solutions for making models indispensable in conservation decision‐making. *Diversity and Distributions* 19(5-6)**,** 490-502.

Arrhenius, O. (1921). Species and area. *Journal of Ecology* 9(1)**,** 95-99.

Bareinboim, E., and Pearl, J. (2013). A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference* 1(1)**,** 107-134.

Bareinboim, E., and Pearl, J. (Year). "Transportability from multiple environments with limited experiments: Completeness results", in: *Advances in Neural Information Processing Systems*), 280-288.

Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P.J., et al. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters* 9(1)**,** 58-64. doi: 10.1111/conl.12180.

Bollen, K.A., and Pearl, J. (2013). "Eight myths about causality and structural equation models," in *Handbook of Causal Analysis for Social Research*. Springer), 301-328.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A., Rylands, A.B., Konstant, W.R., et al. (2002). Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology* 16(4)**,** 909-923.

Bryan, B.A., Nolan, M., Harwood, T.D., Connor, J., Navarro-Garcia, J., King, D., et al. (2014). Supply of carbon sequestration and biodiversity services from Australia's agricultural land under global change. *Global Environmental Change* 28**,** 166-181.

Bryant, B.P., and Lempert, R.J. (2010). Thinking inside the box: a participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change* 77(1)**,** 34-49.

Bull, J., Gordon, A., Law, E., Suttle, K., and Milner‐Gulland, E. (2014). Importance of baseline specification in evaluating conservation interventions and achieving no net loss of biodiversity. *Conservation Biology* 28(3)**,** 799-809.

Collaboration for Environmental Evidence (2013). *Guidelines for Systematic Review and Evidence Synthesis in Environmental Management. Version 4.2.* [Online]. www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf: Environmental Evidence. [Accessed 2016].

Cook, C.N., Inayatullah, S., Burgman, M.A., Sutherland, W.J., and Wintle, B.A. (2014). Strategic foresight: how planning for the unpredictable can improve environmental decision-making. *Trends in Ecology & Evolution* 29(9)**,** 531-541.

Coreau, A., Pinay, G., Thompson, J.D., Cheptou, P.-O. and Mermet, L. (2009). The rise of research on futures in ecology: rebalancing scenarios and prediction. *Ecology Letters* 12,1277-1286

Desmet, P., and Cowling, R. (2004). Using the species-area relationship to set baseline targets for conservation. *Ecology and Society* 9(2)**,** 11.

Dobrowski, S.Z., Thorne, J.H., Greenberg, J.A., Safford, H.D., Mynsberge, A.R., Crimmins, S.M., et al. (2011). Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. *Ecological Monographs* 81(2)**,** 241-257.

Drakare, S., Lennon, J.J., and Hillebrand, H. (2006). The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters* 9(2)**,** 215-227.

Ferraro, P.J., and Hanauer, M.M. (2014). Advances in measuring the environmental and social impacts of environmental programs. *Annual Review of Environment and Resources* 39**,** 495-517.

Ferraro, P.J., Hanauer, M.M., Miteva, D.A., Nelson, J.L., Pattanayak, S.K., Nolte, C., et al. (2015). Estimating the impacts of conservation on ecosystem services and poverty by integrating modeling and evaluation. *Proceedings of the National Academy of Sciences* 112(24)**,** 7420-7425. doi: 10.1073/pnas.1406487112.

Ferraro, P.J., Hanauer, M.M., and Sims, K.R. (2011). Conditions associated with protected area success in conservation and poverty reduction. *Proceedings of the National Academy of Sciences* 108(34)**,** 13913-13918.

Ferraro, P.J., and Pattanayak, S.K. (2006). Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology* 4(4)**,** 482-488. doi: E105 10.1371/journal.pbio.0040105.

Ferraro, P.J., and Pressey, R.L. (2015). Measuring the difference made by conservation initiatives: protected areas and their environmental and social impacts. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370(1681). doi: 10.1098/rstb.2014.0270.

Firn, J., Maggini, R., Chadès, I., Nicol, S., Walters, B., Reeson, A., et al. (2015). Priority threat management of invasive animals to protect biodiversity under climate change. *Global Change Biology* 21(11)**,** 3917-3930.

Fisher, B., Balmford, A., Ferraro, P.J., Glew, L., Mascia, M., Naidoo, R., et al. (2014). Moving Rio forward and avoiding 10 more years with little evidence for effective conservation policy. *Conservation Biology* 28**,** 880-882.

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

Gerst, M.D., Wang, P., and Borsuk, M.E. (2013). Discovering plausible energy and economic futures under global change using multidimensional scenario discovery. *Environmental Modelling & Software* 44**,** 76-86.

Grace, J.B., Scheiner, S.M., and Schoolmaster Jr, D.R. (2015). Structural equation modeling: building and evaluating causal models. Chapter 8 in Fox, G.A., Negrete-Yankelevich, S., and Sosa, V.J. *Ecological Statistics: Contemporary Theory and Application*. Oxford University Press, Oxford, UK, pp 168-199.

Haddaway, N., Woodcock, P., Macura, B., and Collins, A. (2015). Making literature reviews more reliable through application of lessons from systematic reviews. *Conservation Biology* 29(6)**,** 1596-1605.

Hanauer, M.M., and Canavire-Bacarreza, G. (2015). Implications of heterogeneous impacts of protected areas on deforestation and poverty. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370(1681)**,** 20140272.

He, F., and Hubbell, S.P. (2011). Species-area relationships always overestimate extinction rates from habitat loss. *Nature* 473(7347)**,** 368-371.

Höfler, M., Gloster, A.T., and Hoyer, J. (2010). Causal effects in psychotherapy: counterfactuals counteract overgeneralization. *Psychotherapy Research* 20(6)**,** 668-679.

Hurlbert, S.H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological monographs* 54(2)**,** 187-211.

Hyttinen, A., Eberhardt, F., and Järvisalo, M. (Year). "Do-calculus when the true graph is unknown", in: *Proceedings of the 31th Conference on Uncertainty in Artificial Intelligence*.

Iftekhar, M.S., and Pannell, D.J. (2015). “Biases” in adaptive natural resource management. *Conservation Letters* 8(6)**,** 388-396.

Jones, K.W., and Lewis, D.J. (2015). Estimating the counterfactual impact of conservation programs on land cover outcomes: the role of matching and panel regression techniques. *PloS One* 10(10)**,** e0141380.

Joppa, L.N., and Pfaff, A. (2009). High and far: biases in the location of protected areas. *Plos One* 4(12)**,** e8273. doi: 10.1371/journal.pone.0008273.

Kasprzyk, J.R., Nataraj, S., Reed, P.M., and Lempert, R.J. (2013). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software* 42**,** 55-71.

Kline, R.B. (2015). *Principles and Practice of Structural Equation Modeling*. Guilford publications.

Kwakkel, J.H., Auping, W.L., and Pruyt, E. (2013). Dynamic scenario discovery under deep uncertainty: the future of copper. *Technological Forecasting and Social Change* 80(4)**,** 789-800.

Lamb, E.G., Mengersen, K.L., Stewart, K.J., Attanayake, U., and Siciliano, S.D. (2014). Spatially explicit structural equation modeling. *Ecology* 95(9)**,** 2434-2442.

Lempert, R. (2013). Scenarios that illuminate vulnerabilities and robust responses. *Climatic Change* 117(4)**,** 627-646.

Manski, C.F. (2007). Partial identification of counterfactual choice probabilities. *International Economic Review* 48(4)**,** 1393-1410.

Margoluis, R., Stem, C., Swaminathan, V., Brown, M., Johnson, A., Placci, G., et al. (2013). Results chains: a tool for conservation action design, management, and evaluation. *Ecology and Society* 18(3)**,** 22.

Martin, A., Gross-Camp, N., Kebede, B., and McGuire, S. (2014). Measuring effectiveness, efficiency and equity in an experimental Payments for Ecosystem Services trial. *Global Environmental Change* 28**,** 216-226.

Martin, L.J., Blossey, B., and Ellis, E. (2012a). Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment* 10(4)**,** 195-201.

Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M., et al. (2012b). Eliciting expert knowledge in conservation science. *Conservation Biology* 26(1)**,** 29-38. doi: 10.1111/j.1523-1739.2011.01806.x.

Martin, T.G., Murphy, H., Liedloff, A., Thomas, C., Chadès, I., Cook, G., et al. (2015). Buffel grass and climate change: a framework for projecting invasive species distributions when data are scarce. *Biological Invasions* 17(11)**,** 3197-3210.

McConnachie, M.M., Romero, C., Ferraro, P.J., and Wilgen, B.W. (2015). Improving credibility and transparency of conservation impact evaluations through the partial identification approach. *Conservation Biology*.

McKinnon, M., Cheng, S., Garside, R., Masuda, Y., and Miller, D. (2015). Sustainability: map the evidence. *Nature* 528(7581)**,** 185.

Meyfroidt, P. (2015). Approaches and terminology for causal analysis in land systems science. *Journal of Land Use Science***,** 1-27.

Miteva, D.A., Pattanayak, S.K., and Ferraro, P.J. (2012). Evaluation of biodiversity policy instruments: what works and what doesn't? *Oxford Review of Economic Policy* 28(1)**,** 69-92.

Norton, J. (2015). An introduction to sensitivity assessment of simulation models. *Environmental Modelling & Software* 69**,** 166-174.

Nuzzo, R. (2015). How scientists fool themselves-and how they can stop. *Nature* 526(7572)**,** 182-185.

Oliver, T.H., and Roy, D.B. (2015). The pitfalls of ecological forecasting. *Biological Journal of the Linnean Society* 115(3)**,** 767-778.

Pascual, M., Miñana, E.P., and Giacomello, E. (2016). Integrating knowledge on biodiversity and ecosystem services: Mind-mapping and Bayesian Network modelling. *Ecosystem Services* 17**,** 112-122. doi: http://dx.doi.org/10.1016/j.ecoser.2015.12.004.

Pearl, J. (2009). *Causality.* Cambridge university press.

Pfaff, A., and Robalino, J. (2012). Protecting forests, biodiversity, and the climate: predicting policy impact to improve policy choice. *Oxford Review of Economic Policy* 28(1)**,** 164-179.

Pressey, R.L., Visconti, P., and Ferraro, P.J. (2015). Making parks make a difference: poor alignment of policy, planning and management with protected-area impact, and ways forward. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370(1681). doi: 10.1098/rstb.2014.0280.

Rissman, A.R., and Smail, R. (2015). Accounting for results: how conservation organizations report performance information. *Environmental Management* 55(4)**,** 916-929. doi: 10.1007/s00267-014-0435-3.

Rosenzweig, M.L. (1995). *Species Diversity in Space and Time.* Cambridge University Press.

Runge, J., Petoukhov, V., Donges, J.F., Hlinka, J., Jajcay, N., Vejmelka, M., et al. (2015). Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications* 6. doi: 10.1038/ncomms9502.

Rybicki, J., and Hanski, I. (2013). Species–area relationships and extinctions caused by habitat loss and fragmentation. *Ecology Letters* 16(s1)**,** 27-38.

Saltelli, A., and Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software* 25(12)**,** 1508-1517.

Shipley, B. (2002). *Cause and Correlation in Biology: a User's Guide to Path Analysis, Structural Equations and Causal Inference.* Cambridge University Press.

Sills, E.O., Herrera, D., Kirkpatrick, A.J., Brandao, A., Jr., Dickson, R., Hall, S., et al. (2015). Estimating the Impacts of Local Policy Innovation: The Synthetic Control Method Applied to Tropical Deforestation. *PLoS One* 10(7)**,** e0132590. doi: 10.1371/journal.pone.0132590.

Sinclair, S.J., White, M.D., and Newell, G.R. (2010). How useful are species distribution models for managing biodiversity under future climates. *Ecology and Society* 15(8).

Sutherland, W.J., Pullin, A.S., Dolman, P.M., & Knight, T.M. (2004). The need for evidence-based conservation. *Trends in Ecology & Evolution,* 19(6), 305-308.

Stocks, G., Seales, L., Paniagua, F., Maehr, E., and Bruna, E.M. (2008). The geographical and institutional distribution of ecological research in the tropics. *Biotropica* 40(4)**,** 397-404.

Strassburg, B.B.N., Rodrigues, A.S.L., Gusti, M., Balmford, A., Fritz, S., Obersteiner, M., et al. (2012). Impacts of incentives to reduce emissions from deforestation on global species extinctions. *Nature Climate Change* 2(5)**,** 350. doi: 10.1038/nclimate1375.

Visconti, P., Bakkenes, M., Smith, R.J., Joppa, L., and Sykes, R.E. (2015). Socio-economic and ecological impacts of global protected area expansion plans. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370(1681)**,** 20140284.

Whittaker, R.J., and Fernández-Palacios, J.M. (2007). *Island Biogeography: Ecology, Evolution, and Conservation.* Oxford: Oxford University Press.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics* 5(3)**,** 161-215. doi: 10.1214/aoms/1177732676.

BOX 1: What is causal inference?

Causal inference is an analysis of the causal relationship between variables, for example the effect of a treatment on an outcome. It distinguishes causation from association though clarifying and justifying the model assumptions required for its inference (Pearl, 2009). While a range of techniques are used to infer causality, here we refer to the 'counterfactual' or 'potential outcomes' model, i.e. the Neyman-Rubin Causal Model.

Causal inference is typically framed around a causal model: hypotheses regarding how a *treatment* affects an *outcome*, a description of the causal pathway and possible *mechanisms, confounders,* and *moderators* (i.e. the causal model describes the structural assumptions). A *treatment* is the variable that is hypothesized to cause the *outcome* of interest. *Mechanisms* are the path by which the *treatment* causes the *outcome* (in some literatures, an intermediate node along this path is also termed a 'mediator'). *Confounders* (or confounding factors) are rival explanations: variables that are systematically associated with the outcome and the treatment or mechanisms along the casual pathway. *Confounders* may result in an association between a treatment and an outcome that is not direct or causal, or alternatively could mask a direct treatment effect. For example, because the selection bias in the location of protected areas, these areas are likely to experience lower rates of deforestation regardless of whether they were protected or not. Naïve estimates that do not account for this selection bias can severely overestimate protected area effectiveness (Joppa and Pfaff, 2009). It is particularly important to distinguish mechanisms and confounders, as controlling for the influence of a mechanism will essentially remove the impact being sought, while controlling for the influence of a confounding variable is advisable to reduce bias. A *moderator* is an interaction effect, a variable that affects the outcome of the treatment, but not correlated with exposure to the treatment.

A challenge when framing a causal analysis is defining the *counterfactual* outcome: the unobserved outcome for a given unit (e.g., area, species, individual), if the unit's treatment status were different from what is observed. For example in a protected forest we can observe deforestation rates, but we cannot observe (counterfactual) deforestation rates should the same area of forest have instead remained unprotected. The difference between a unit's actual state and its counterfactual state is the causal effect (the *estimand*) that we seek to estimate (also called the treatment effect; Fig. B1). Experimental designs, such as randomized controlled trials, permit causal inference by introducing variation in treatment assignment that is unrelated to potential outcomes. In other words, effective randomization eliminates all rival explanations other than sampling variability, thus giving validity to the assumption that the counterfactual is well represented by the 'control' sample. Where experimental designs are not feasible, quasi-experimental designs can approximate them, by identifying an observable stand-in for the unobservable counterfactual (Fig. B1). Quasi-experimental designs rely on a strong understanding of how treatment was assigned and on statistical techniques to control for confounding factors. These techniques include matching (to control observable confounders)(Ferraro et al., 2011), use of panel data and synthetic controls (to control time-invariant unobservable confounders)(Jones and Lewis, 2015; Sills et al., 2015), instrumental variables, and discontinuity designs (to eliminate unobservable confounders).

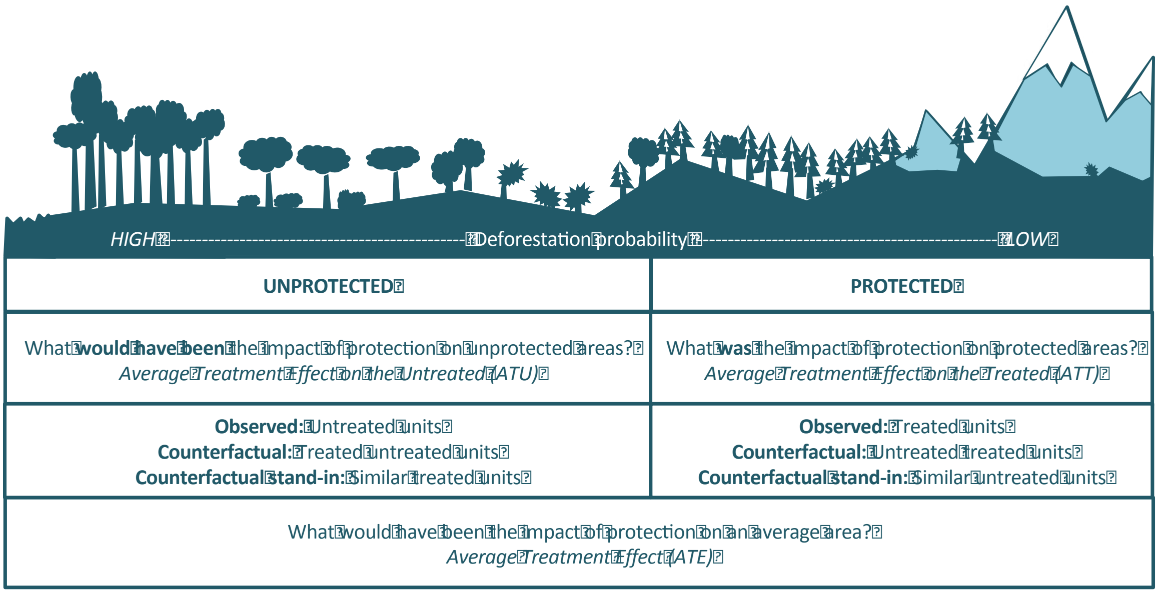

Figure B1: In treatments with a strong selection bias, for example the implementation of protected areas, several different treatment effects may be of interest in impact evaluation. The Average Treatment Effect of the Treated (ATT) is often the sought-after estimand: the expected difference between the observed and counterfactual outcome for the treated population only. As the counterfactual is unobservable, a stand-in is assumed to represent this. The Average Treatment Effect on the Untreated (ATU) may also be policy relevant: the expected effect of a treatment on the untreated population. In rare cases, the expected treatment effect on a randomly chosen unit from the population (treated and untreated) may be relevant: this estimand is called the Average Treatment Effect on the Treated (ATE). This can be calculated proportionally from ATT and ATU.

BOX 2. An illustration of causal models using the Species Area Relationship

We illustrate different types of variables and models using the example of the Species Area Relationship (SAR)(Arrhenius, 1921). The SAR is perhaps the most ubiquitous causal model to explain patterns in species richness, with over 21,000 papers citing it (Web of Science, July 2015). The ‘simple’ SAR model, which posits a positive relationship between habitat area and species richness, can underpin a prospective evaluation of conservation intervention by assuming that some form of land-use change (e.g., establishment of a protected area) is the ‘treatment’ and a change in habitat area is the mechanism through which the treatment affects an ecological ‘outcome’ (Figure B2). The SAR has informed numerous aspects of conservation policy (Drakare et al., 2006) including biodiversity targets (Desmet and Cowling, 2004), land clearing (Brooks et al., 2002), and incentive mechanisms such as payments for ecosystem services and REDD+ (Strassburg et al., 2012). In cases where the SAR is used in prospective evaluations, most studies consider broad types of conservation actions, such as land-use zoning (Brooks et al., 2002; Desmet and Cowling, 2004).

While the simple SAR model is elegant in its simplicity, this oversimplification means that the mechanisms through which projected interventions are proposed to operate are not clear. Further, the model fails to recognize important moderators. Such applications are therefore likely to systematically under or over-estimate impacts (He and Hubbell, 2011; Rybicki and Hanski, 2013). Several variables have been proposed to influence the SAR (Rosenzweig, 1995; Drakare et al., 2006; Whittaker and Fernández-Palacios, 2007). Exploratory models might frame these as ‘covariates’ of the ‘response’ variable (Figure B2). However, these could be more explicitly characterized as ‘mechanisms’ through which the SAR operates (e.g. habitat heterogeneity, population size, immigration, and evolutionary processes including mutation, selection, and drift); ‘confounders’ that may also cause changes in species richness, but for reasons independent of area (e.g. fragment characteristics and edge effects, invasive or predatory species, differences in climate and disturbance regimes and anthropogenic impacts); or ‘moderators’ that lead to variation in the SAR parameters (e.g. taxa, matrix permeability and habitability). These variables can mean similar ‘treatments’, such as the establishment of protected areas, can have substantially different effects in different contexts (Ferraro et al., 2011; Hanauer and Canavire-Bacarreza, 2015). While not all of these variables will be important in any specific context, the basic model implies that there is no variation in mechanisms, moderators, or rival explanations across different proposed conservation interventions or contexts (Figure B2). Analyses which evaluate prospective interventions could be improved by greater consideration of these processes, or by identifying specific on-ground conservation management actions, such as how invasive species might be managed (Firn et al., 2015).

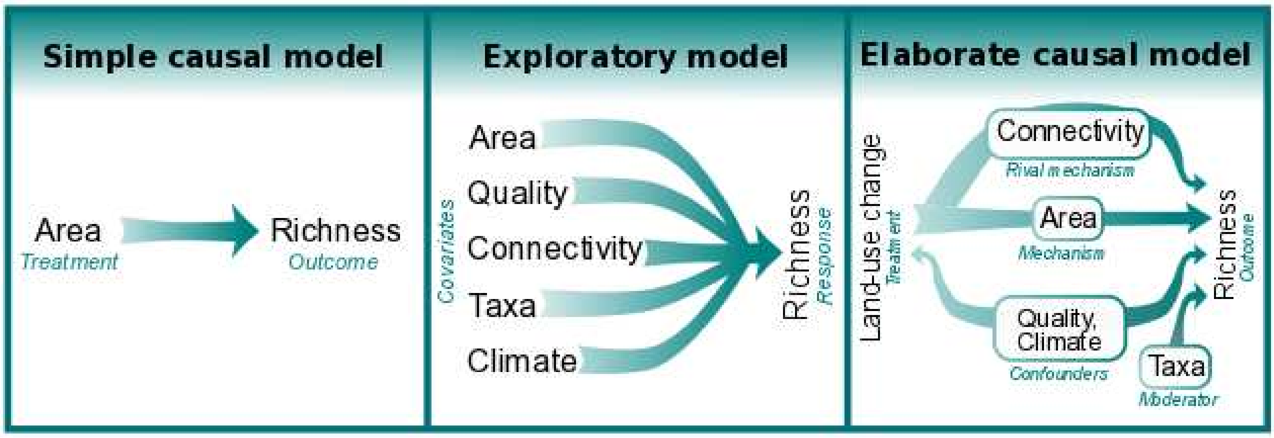

Figure B2. Different modeling frameworks are appropriate at different stages of projection analyses. We illustrate several types using the example of the Species Area Relationship. Simple causal models clarify the causal relationship of interest (i.e. the impact of the treatment on the outcome), but typically need to be elaborated for analysis. Exploratory models may identify useful covariates of the response variable, but are not ideal for attribution of causal impacts. Elaborate causal models make explicit the structure of underlying causal assumptions, and identify the different characteristics of variables and their interactions: key requirements for developing both theories of change and confidence in model projections derived from such analyses. Illustrated here is an elaborated model that may be suitable for 'reductionist' causal inference, whereas an example of more complex 'all-cause' models can be seen in Box 3.

BOX 3: Use of SEM to develop causal models for song sparrow conservation

Structural equation models (SEM, or Path Analysis)(Wright, 1934; Shipley, 2002) offer one approach to developing elaborate causal models (Bollen and Pearl, 2013). Their usefulness in causal inference, particularly for interrogating model structure in complex contexts (Pearl, 2009), has led to their widespread use in health, social sciences, and ecology (Shipley, 2002; Grace et al., 2015; Kline, 2015).

Analysis of song sparrow (*Melospiza melodia*) populations demonstrates the utility of SEMs for conservation. Several subspecies are subject to stochastic variation in climate, brood parasitism and nest depredation, with each of these factors capable of driving local extinction (Arcese & Norris, submitted). Aiming to resolve debate regarding which of these factors were most important for management, Arcese & Norris (submitted) studied an island population over a 40-year period. Resulting SEMs revealed that adult and juvenile survival each exerted about three times more influence on population growth rate, *r*, than reproductive rate, and that juvenile survival determined *r* in most years (Figure B3). Arcese and Norris show that, despite severe winter weather severely limiting populations in the past, climate change has ameliorated these exogenous limits on *r* and increased the influence of density-related limits on *r* via competition for space and food.

If the results from the island population can be transferred to other song-sparrow populations that are currently threatened, the model implies that expanding suitable habitat and re-establishing locally extinct populations by translocating juveniles from extant populations at or near carrying capacity represents a more reliable route to minimizing extinction risk than controlling parasites or predators (Arcese & Norris, submitted). Such model 'transportation' – extrapolation or generalisation of impact estimates from one sample to the population of interest – is already often done informally, often qualitatively, as in the narrative example. However, these narratives are often subject to narrative criticism (i.e. narratives of why such extrapolations might not be appropriate)(e.g. Höfler et al., 2010). More recent work has developed structural causal theory to formally define model transportability, and thereby derive "licensing assumptions" including transport formulae under which model transportability is acceptable (Bareinboim and Pearl, 2013). Such transportability theory may be useful for transportation of impact estimates from experiments or pilots to larger populations (i.e. sample-selection bias), between study systems, to identify useful instrumental or surrogate variables, and in meta-analysis (Bareinboim and Pearl, 2013; 2014). We see many opportunities to engage with this frontier of causal research in the domain of conservation and environmental management.

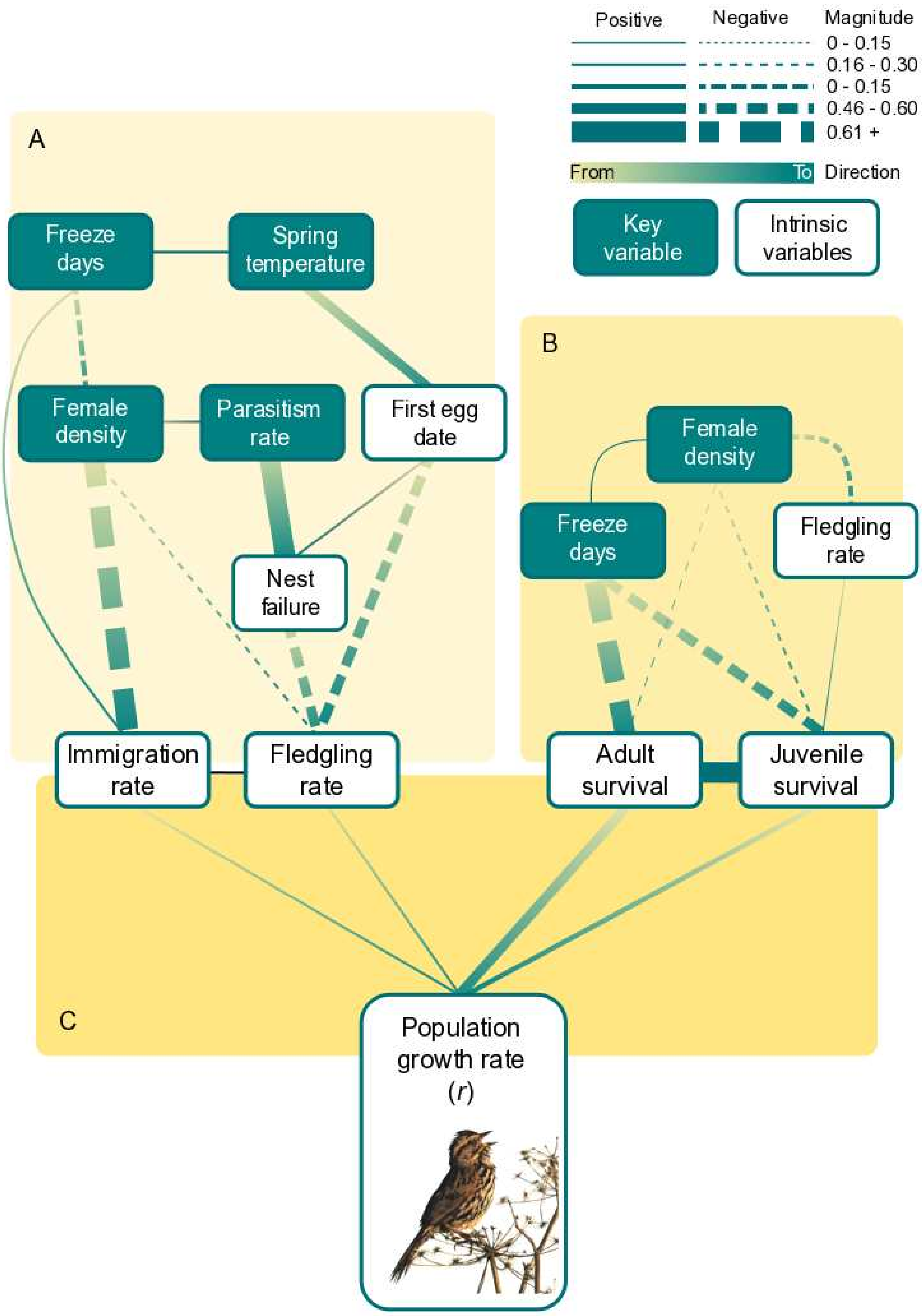

Figure B3: SEM results for song sparrow management (simplified). Lines represent standardized partial regression coefficients (β; directional) or covariances (non-directional), with key variables of management interest highlighted. SEM models were constructed separately to explain A) variation in immigration rate and fledging rate (reproduction), B) variation in adult and juvenile survival, and C) influence of vital rates on population growth rate (*r*). For simplicity, significance levels are not shown, only positive effects on *r* are given, and minor covariances are absent in C. See Arcese and Norris (submitted) for full results.

BOX 4: Partial identification for examining assumptions

In retrospective modelling analyses, partial identification recognizes that the assumptions underlying estimates of the counterfactual, and hence the impact estimates, may have varying levels of credibility (Manski, 2007; McConnachie et al., 2015). It provides an analysis framework that sequentially explores assumptions increasing in strength, in effect a special case of sensitivity analyses. While rarely used in the evaluation of conservation programs to date (McConnachie et al., 2015), this process has a number of potential benefits deriving from the transparent assessment of bias and plausibility of assumptions. These benefits include providing constructive direction when point estimates are potentially biased or contentious, or when information on the potential behavior of participants during future policy or program implementation is limited (McConnachie et al., 2015).

Credibility of assumptions may vary depending on how strong the assumptions are, how well supported the assumptions are by evidence, and how contentious the claims are (for example, due to existing personal biases) (Figure B4). ‘Identification regions’ show the range of values that contain the impact estimate. The identification region with the highest credibility, after the maximum potential bounds, is the ‘no assumptions’ estimate. This is constructed by clipping the minimum and maximum theoretically possible estimates for the counterfactual, with the values of the observed ‘treated’ units. As this makes no claims in regard to the counterfactual, it can engender little controversy aside from measurement error. Slightly stronger assumptions may include a ‘monotone treatment response’ estimate, which constrains the ‘no assumptions’ bounds further, by assuming that treatment impacted the outcome positively (or negatively, should this be a more credible assumption). A ‘monotone treatment selection’ estimate further constrains the bounds, by assuming that the treatment was selectively applied to areas that were in worse (or better) condition than others prior to treatment. These identification regions make some claims on what the counterfactual might be, thus they may not be considered credible if these claims are not supported by evidence.

Point estimates need to make stronger assumptions, requiring them to be backed by more evidence. The most credible point estimate may be identified using conditioning, a causal inference technique that matches samples based on observable covariates (McConnachie et al., 2015). Other impact point estimates include those from Before-After-Control-Intervention (BACI) designs, and the simpler Before-After or Treated-Control comparisons, which may be credible if evidence is shown to suggest the samples were representative, and if the design reasonably accounts for change over time and selection bias (Ferraro and Pressey, 2015; McConnachie et al., 2015).

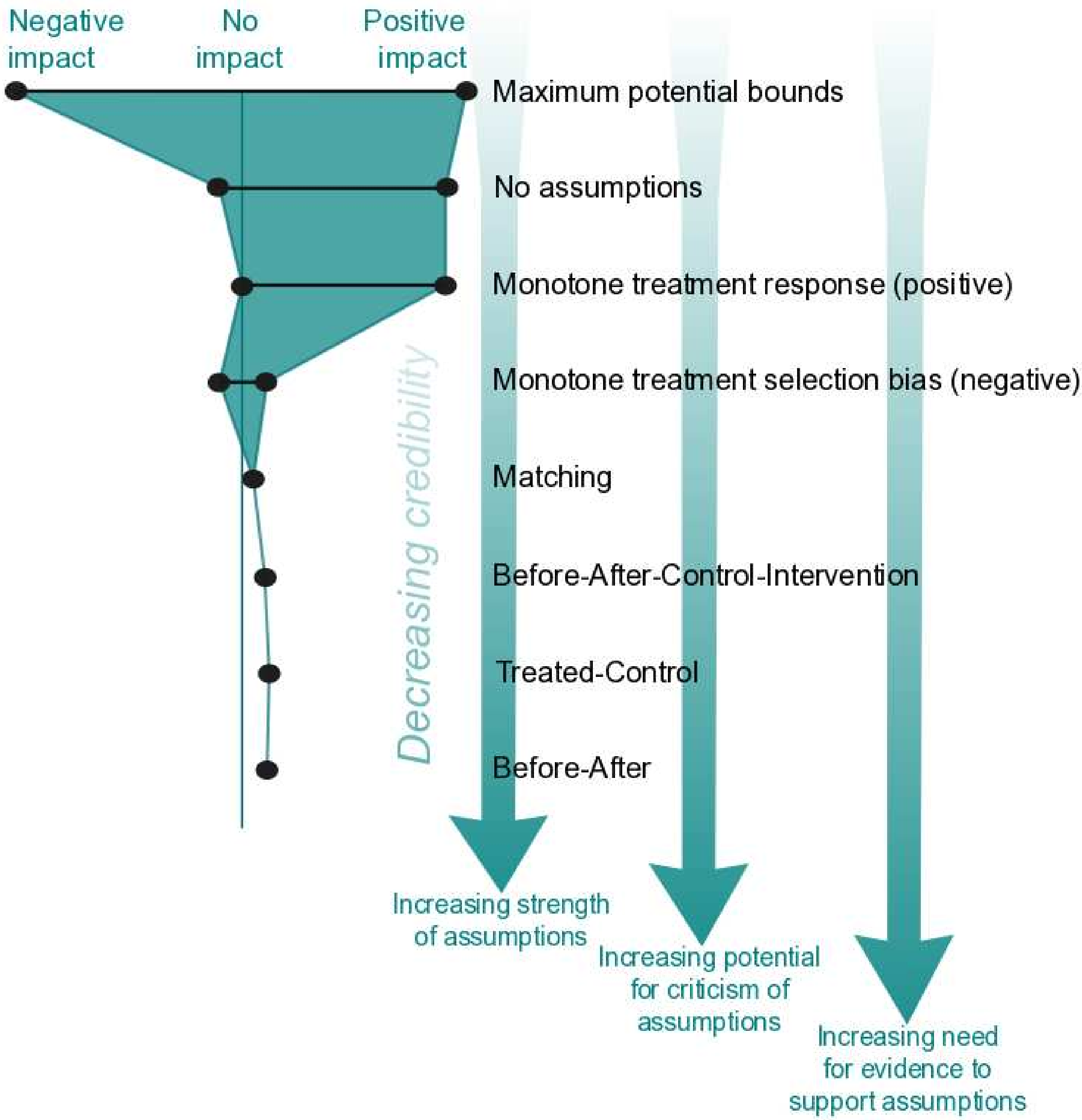

Figure B4: Partial identification is an analysis framework that sequentially explores the implications of assumptions regarding the counterfactual. Assumptions decrease in credibility due to increasing strengths of the claims regarding the counterfactual, which increases the potential for criticism, and the need for evidence to support the claims.