Trends in
Ecology & Evolution

CellPress
OPEN ACCESS

Opinion

Evaluating Impact Using Time-Series Data

Hannah S. Wauchope [1,2,*] Tatsuya Amano [3,4] Jonas Geldmann,[1,5] Alison Johnston [1,6] Benno I. Simmons [1,2,7] William J. Sutherland [1] and Julia P.G. Jones [8]

Humanity's impact on the environment is increasing, as are strategies to conserve biodiversity, but a lack of understanding about how interventions affect ecological and conservation outcomes hampers decision-making. Time series are often used to assess impacts, but ecologists tend to compare average values from before to after an impact; overlooking the potential for the intervention to elicit a change in trend. Without methods that allow for a range of responses, erroneous conclusions can be drawn, especially for large, multi-time-series datasets, which are increasingly available. Drawing on literature in other disciplines and pioneering work in ecology, we present a standardised framework to robustly assesses how interventions, like natural disasters or conservation policies, affect ecological time series.

Highlights

Ecologists have called for more robust studies on the impact of conservation interventions, or environmental shocks, on outcomes of interest, such as populations, habitat loss, or pressures.

Time-series data are increasingly available and can, if appropriately analysed, allow such causal inferences.

However, there are important pitfalls that make large-scale analyses involving multiple time series problematic.

There has been progress in a range of fields, but the literature is fragmented and not all is easily accessible to ecologists.

A framework is presented, with clear and consistent terminology, to support ecologists to conduct effective impact evaluation with time-series data. This will allow them to contribute to better-informed environmental management decisions.

Impact Evaluation in Ecology

Ecologists often seek to understand the impact of a conservation intervention (e.g., a protected area or reintroduced species), or an environmental shock (e.g., an oil spill or hurricane) on one or more response variables (e.g., population counts or habitat loss) [1–9]. In recent years, there has been a surge of literature calling for more rigorous impact evaluation in ecology and conservation [10–14]. While the terms **impact evaluation** (see Glossary) and **intervention** are often used to describe the impact of a deliberate intervention, such as a policy change [10], here they are used to consider the general problem of **causal inference**.

To determine the impact of an intervention, one must understand what would have happened if the intervention had not occurred [15]. Ideally this is achieved through an experimental setup, where units are randomly allocated to treatment and control groups. However, while experimental manipulation of whole ecosystems, or random application of conservation interventions at scale do exist [16,17], such experiments are seldom feasible [10,18,19], or indeed possible in the case of events such as natural disasters. Instead, researchers commonly try to estimate the '**counterfactual**' using **quasi-experimental** impact evaluation methods [14,20]. A commonly used approach is to examine outcomes before and after the intervention, [**before after (BA)** analysis], or to identify a separate control group that shares as many characteristics as possible with the intervention group, except for the intervention, [**control intervention (CI)** analysis]. These approaches can be combined to compare before to after, between control and intervention groups, [i.e., **before after control intervention (BACI)** analysis].

Impact Evaluation with Time Series

Time-series data are a common and powerful [21] way to conduct impact evaluation in ecology. The methods used by ecologists to conduct impact evaluation with time-series data or **cross-sectional time series,** have remained largely unchanged since seminal papers published in the 1980s and 1990s [21,22]. The standard framework tends to only consider the **average change** between control and intervention; in a BACI time-series context, this analysis has been termed **BACI Paired Series** (**BACIPS**) [21]. These methods assume that a change in an average response variable, can capture how the time series responds to the

[1]Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, CB2 3QZ, UK
[2]Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn, TR10 9FE, UK
[3]School of Biological Sciences, University of Queensland, Brisbane, Australia
[4]Centre for Biodiversity and Conservation Science, University of Queensland, Brisbane, Australia
[5]Center for Macroecology, Evolution and Climate, Globe Institute, University of Copenhagen, Copenhagen, Denmark
[6]Lab of Ornithology, Cornell University, Ithaca, New York, USA
[7]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK
[8]School of Natural Sciences, Bangor University, Bangor, LL57 2UW, UK

*Correspondence:
hannah.wauchope@gmail.com
(H.S. Wauchope).

intervention, by assuming that the data fluctuate around pre- and post- intervention averages. In reality, many time series show trends through time independently of the intervention. They could therefore respond to an intervention not only with an **immediate change**, but with a change in trend, which is not always captured by comparing average differences.

Other fields, such as medicine [23,24], public health [25,26], and education [27] have long recognised this, and methods to account for changes in trend were developed as early as the 1970s in the psychological sciences [28] (see S1 in the supplemental information online for a brief overview of terminology in other fields). In ecological impact evaluation, trends have only very recently been included. Thiault et al. [29,30] introduced the '**Progressive Change BACIPS**' approach in 2017, as an extension of BACIPS that considers trends in the 'after' period, but still assumes no change through time in the 'before' period. Chevalier et al. [9] addressed this by considering trends and averages both before and after, and also introduced 'CI-measures', which further quantify the nature of impact in a BACI framework [31]. This pioneering work is built upon, by arguing that average change can be misleading when trends are present in the dataset, and that only **trend change** should be considered in these cases.

This is best understood with an example. Imagine monitoring a population of African elephants (*Loxodonta africana*) before and after a hunting ban, which actually reverses a downward population trajectory (Figure 1, first column). If the average count of the years before the ban is compared to the average count of the years after, one may conclude that the ban has had a negative effect (Figure 1A, average change). However, if the trend of the counts is considered, it becomes clear that the population was declining pre-ban but has begun to increase post-ban; suggesting that the ban has actually had a positive impact on the population (Figure 1C, trend change). Thus, change in trend is needed to accurately assess the impact of an intervention on a time series with trends. Including trend does not negate that there can also be an abrupt shift following the intervention; this immediate change can also be analysed (Figure 1E). In this example, if conservationists had used average change, they would have erroneously concluded that the hunting ban had a negative impact on elephant populations, and the beneficial intervention may have been stopped.

Since most ecological time-series impact evaluation studies have focused on only one or a handful of time series [32–37], using only average change has not been a big problem, as each time series could be visually checked for indications of trends. However, risks increase when using large numbers of time series (e.g., counts of multiple bird species at multiple sites [38]), where checking individual series becomes impractical. Large datasets of time series are increasingly available from long-term surveys, remote-sensing, and national monitoring schemes [39–43]. A clearer framework for analysis is needed to avoid inadvertent mistakes in large datasets, and to improve clarity of analysis in smaller ones (see S2 in the supplemental information online, which details all potential mistakes that the presented framework averts).

Drawing from other disciplines and pioneering work in ecology [9], we present a framework for conducting impact evaluation with ecological time-series data. Our framework is directed towards those working with time series that show trends through time; if there is no expectation of trend change, the models given here for 'average change' can be used, though more sophisticated **fixed effects panel regression** methods also exist [44–46] (however, there are limitations to this approach when interventions are staggered through time [47]). Finally, a frequentist statistical framework is generally used through this paper for simplicity, but the concepts are readily transferrable to Bayesian or information theoretic approaches.

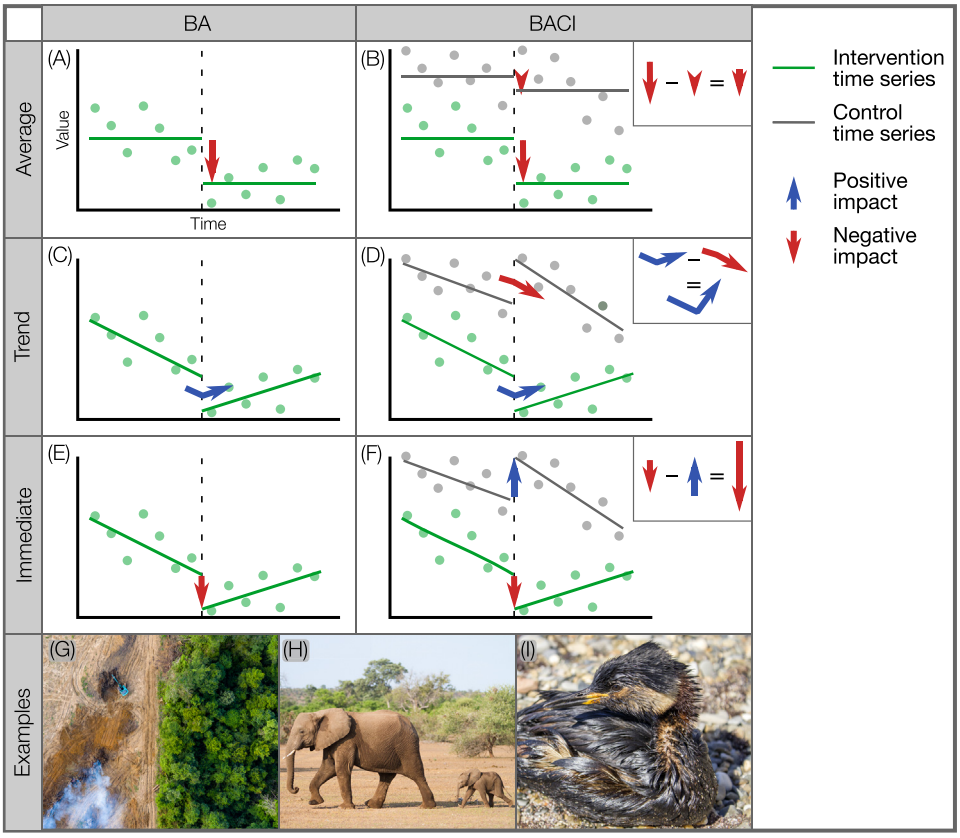

Figure 1. Average, Trend, and Immediate Change When Assessing the Impact of an Intervention (broken vertical line) Using BA or BACI Data. Blue arrows indicate positive change and red indicate negative change. Impact can be defined by change in average (A, B), change in trend (C, D) and/or an immediate change (E, F). BACI comparisons show the **BACI Contrast**, (i.e., the difference in the change in before to after, between control (grey) and intervention (green) time series. In this example, average and immediate change indicate a negative impact, but trend change indicates a positive impact. Many impact evaluation questions could be considered in this framework including investigating the impact of carbon payments on tropical deforestation (G; Richard Whitcombe/shutterstock.com), a hunting ban on elephant populations (H; Villiers Steyn/shutterstock.com), or oil spills on populations of waterbirds (I; Mike Shooter/shutterstock.com).

Choosing a Control

BA, CI, and BACI methods are all techniques to infer the counterfactual. BA analysis assumes that, were it not for the intervention, the trajectory of the time series would not have changed, while CI assumes that the difference observed between the control and intervention time series is a result of the intervention, and that no other unobserved differences exist between the control and intervention sites. BACI addresses these assumptions, by combining BA and CI components. Further discussions of the relative merits of the three approaches are given in other papers [20,48], but BACI is the best option if data allows [48]. Note that if using a BA comparison method, results should be examined for the possibility that regression to the mean is occurring (see S7 in the supplemental information online). As CI analysis is unable to effectively consider changes in trend [20], the focus here is on BA and BACI analysis. For simplicity, 'control' is referred to as a control in either space or time, and 'temporal-control' or 'spatial-control' are used to distinguish between the 'before period' or 'control' time series, respectively.

For BACI comparisons, statistical matching can be used to identify a control time series that is as equivalent as possible to the intervention time series, based on a set of matching variables

Glossary

Before after (BA): a method of analysis that estimates the counterfactual by comparing values from before to after the intervention.
Before After Control Intervention (BACI): a method of analysis that estimates the counterfactual by comparing the change from before to after between control and intervention groups.
BACI Contrast: a commonly used term for BACI average change. Given by subtracting the before–after difference of the control group from the before–after difference of the intervention group. That is, BACI Contrast = $(\bar{A}_I - \bar{B}_I) - (\bar{A}_C - \bar{B}_C)$.
BACI Paired Series (BACIPS): methods that discuss BACI time-series analysis with multiple paired groups, typically considering average change.
Causal inference: the statistical process of concluding that an observed association is due to causation not correlation.
Control intervention (CI): a method of analysis that estimates the counterfactual by comparing values between control and intervention groups.
Controlled or Comparative Interrupted Time Series Analysis: The term used in some disciplines to refer to models such as ours that calculate BACI trend and immediate change.
Counterfactual: what would have occurred in the absence of an intervention.
Cross-sectional time series/panel data: time-series data of many entities, each followed through time. For instance, annual counts from many identified prides of lions, or monthly deforestation for many regions in a country.
Difference in Differences: A term typically used in econometrics to refer to the BACI Contrast (i.e., average change in a BACI framework) but is now sometimes also used to refer to a trend change BACI analysis.
Fixed effects panel regression: an analysis method for **panel data** where there is *not* an expectation of any trend in the data. Time is included in the model as a fixed factor to control for temporal shocks.
Impact evaluation: determining how an intervention has causally affected outcomes (examples of outcomes in

[49,50]. Crucially, the matched control and intervention time series should show similar trends in the 'before' time period [51]; this is often referred to as the parallel trends assumption. For example, if trends in populations of elephants were being compared before and after a hunting ban was introduced in one population, the parallel trends assumption is met if the elephants at both sites were declining at approximately the same rate over the 'before' years (Figure 1D). Many matching methods are available to facilitate the matching process [49,50,52,53], though it always requires careful consideration of the assumptions involved [50]. [31] provides 'CI- measures' to estimate the similarity of control and intervention groups before and after an intervention, which can be useful for further interrogating the nature of impact, especially when matches are imperfect.

Choosing a Change Metric

Time series can respond to an intervention either by an abrupt change when the intervention is introduced, a gradual change over time, or both. The change can be measured in three ways: average, trend, or immediate (Figure 1; note these terms were introduced to ecology by [9]). Box 1

Box 1. Simplified Formulae for Calculating Average, Trend, and Immediate Change

Average Change

To compare the change in average with BA data, each value of the time series (where value could be population count, percentage forest cover, etc.) is predicted by a binary Before-After variable (*BA*), which is 0 pre-intervention and 1 after.

$$Value \sim BA \qquad \text{[I] (BA average)}$$

If the coefficient of *BA* is significantly positive, the average value of the time series is higher post-intervention.

In a BACI analysis, this average change is then compared between intervention and spatial-control time series (Figure 1b), this is often termed the **BACI contrast** [31,60]. A Control-Intervention variable (*CI*) is included, which is 0 for the spatial-control time series and 1 for the intervention time series.

$$Value \sim BA + CI + (BA \times CI) \qquad \text{[II] (BACI average)}$$

The interaction between *BA* and *CI* describes how the intervention affects the change from before to after. A positive coefficient indicates that the average difference from before to after is more positive in the intervention time series.

Trend and Immediate Change

To estimate trend and immediate change, we must include time in the model. In order to compare immediate change between the last time step before intervention, to the first time step after intervention, the simplest way to construct the model is for time to be centred around 0, with 0 being the first time step after intervention {similar to [9], though see [54,61] and https://nawmp.org/sites/default/files/2018-01/1986%20OriginalNAWMP.pdf for an alternative method}.

For a BA study, the value of the time series is predicted by *Time* and the *BA* coefficient.

$$Value \sim Time + BA + (Time \times BA) \qquad \text{[III] (BA trend/immediate)}$$

The *BA* coefficient gives the change in values from before to after at Year 0 – this is the immediate change. The interaction between *BA* and *Time* gives the trend change from before to after. A positive coefficient indicates the trend after the intervention is more positive than before the intervention (note that it could still be negative, just less negative than before).

As before, for a BACI study an intervention (*CI*) coefficient is introduced.

$$Value \sim Time + BA + CI + (BA \times CI) + (BA \times Time) + (CI \times Time) + (BA \times CI \times Time) \qquad \text{[IV] (BACI trend/immediate)}$$

The interaction between *BA* and *CI* describes the BACI immediate change and *BA, CI,* and *Time* the BACI trend change, (i.e., the difference in before-after immediate, or trend change, respectively, between control and intervention time series). A positive coefficient for these interactions indicates that the intervention time series has had a more positive change in immediate/trend than the control time series.

ecology include population counts, body mass, or rate of habitat loss).

Interrupted Time Series Analysis: The term used in some disciplines to refer to models such as ours which consider BA trend and immediate change.

Intervention: an event that disturbs a system. The event could be intentional, accidental, or natural, for instance the designation of a protected area, an oil spill, or a wildfire.

Average/Trend/Immediate Change: compares change in average, trend, or immediate from before to after in BA analysis, and the difference in change from before to after between control and intervention time series in BACI analysis.

Progressive Change BACIPS: a modified form of BACIPS that considers multiple linear and nonlinear responses in the after period. These methods assume a steady state, or no trend, in the before period.

Quasi-experimental: a range of approaches used to estimate the causal impact of an intervention without randomisation.

Time-series data: a series of measurements at intervals through time. For example, annual counts of lions in a pride, monthly measures of deforestation in a region, and annual hunting rate of a bird species.

gives simplified models to calculate these metrics, and Box 2 gives a real-world example. Full models can be found in S3 in the supplemental information online. An example of interpreting coefficients can be found in S4 in the supplemental information online.

Which metric to use will depend on the response expected, and the type of data used. If there is no expectation that the time series shows trends through time, then comparing average change between intervention and control is suitable. If the time series shows trends through time and no immediate change is expected after the intervention, then trend change should be used (note that without any immediate change, average change will still give a correct direction of response under

Box 2. Case Study: Interpreting the Impact of a Protected Area on Populations of the Common Merganser

We use trends in a population of common merganser, or goosander (*Mergus merganser*), from a protected and unprotected site as a worked example to demonstrate how to interpret model coefficients from BACI time series analysis, and the importance of using the right metric of change.

The common merganser is a sea duck that is distributed across Europe and North America. It is currently classified as 'Least Concern' but has previously been a target species for conservation in the United States [61]. We analysed Christmas Bird Count data to determine how a wintering population of common merganser was impacted by a protected area's designation in 1997, when compared to a similar but unprotected population, identified through statistical matching (Figure I, see [38] for detailed methodology).

We ran Model 4 (see Equation IV in Box 1), using a generalised linear model with a negative binomial distribution [38], and conducted a robustness check by comparing models constructed with and without the *BA* term (see 'Is there any change at all?'). The model with *BA* had a better fit to the data, so analysis proceeded (Figure I). To evaluate the impact of the protected area, we were most interested in three coefficients: (i) the *BA:CI* term, which indicates immediate change; (ii) the *BA:CI:Year* term; which indicates trend change; and (iii) the *CI:Year* term, which tells us if the parallel trends assumption is satisfied (see S4 in the supplemental information online for details of how to interpret all coefficients).

The *CI:Year* term was not significant ($P = 0.09$), meaning the parallel trends assumption was satisfied. The immediate change term (*BA:CI*) was also insignificant ($P = 0.97$), however, the trend change term (*BA:CI:Year)* was significantly positive (Estimate = 0.31 ± 0.12, $P = 0.01$), indicating that the trend change from before to after was more positive in the protected site.

If we had applied an average change model to this data (i.e., Model 2, see Equation II in Box 1), we would have detected no significant impact of the protected area ($P = 0.38$), indicating the importance of adopting the trend and immediate change model when working with time series that show trends.

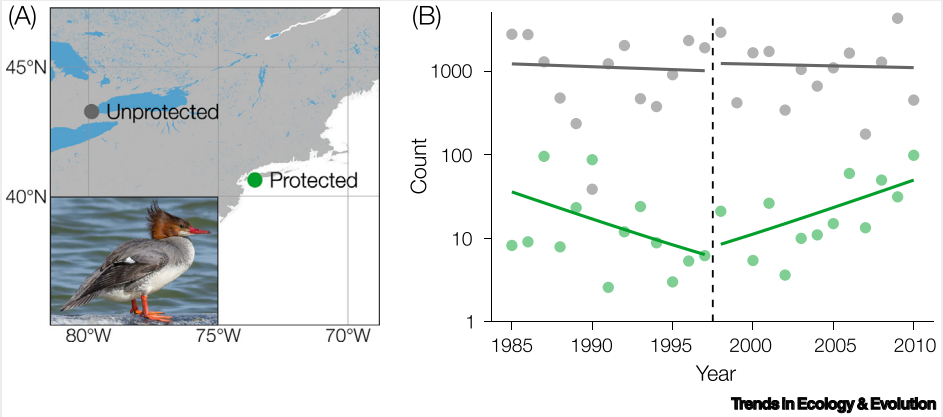

Figure I. Common Merganser (*Mergus merganser*) Case Study Site Locations and Count Data (A) Map Showing Protected (light green) and unprotected (grey) sites where merganser (*Mergus merganser*) have been monitored in the North East USA (Photo: Frank Schulenburg/Wikimedia Commons/CC BY-SA 3.0) (B) Merganser counts, with lines showing modelled trend; vertical broken line shows the year of protected area designation.

BACI, because of the parallel trends assumption, but may not give the correct magnitude; see S5 in the supplemental information online). If the time series shows trends through time and an immediate change is expected following the intervention, then both trend and immediate change must be compared, as average change will be misleading (as demonstrated in Figure 1). Sometimes, it may not be known what response is expected, in which case it may be best to assess all three types of change, and carefully examine them in the light of knowledge of the system, to ascertain which is most appropriate. CI-measures that quantify how much of the change is occurring in control versus intervention time series, can also be useful [31].

For example, rates of deforestation are often not considered to show consistent trends through time, and so to estimate the impact of a new policy aiming to reduce deforestation (Figure 1G), average deforestation rates before and after the policy could be compared (Figure 1A,B). Conversely, a population of elephants that has been declining due to hunting is likely to respond to a hunting ban by a gradual increase in individuals (Figure 1H). If analysing only BA data, trend change is needed, but either average change or trend change would be appropriate with BACI data (Figure 1C,D; as long as the parallel trends assumption is satisfied). An oil spill affecting a population of waterbirds, would cause immediate mortality for many individuals (Figure 1I), so comparing immediate change would best reflect the impact of the spill (Figure 1E,F). However, a waterbird population could respond to the restoration of a wetland both with an immediate change, as more adults are able to migrate to the area, as well as a gradual increase in individuals, through improved breeding success. In this case, both trend and immediate change should be compared; Box 2 gives an example of this.

In large datasets, it can sometimes be difficult to summarise the signal of impact across many time series, especially if they do not all respond in the same way. Box 3 discusses the complexities of such cases and provides some suggested solutions.

Box 3. Making Conclusions from Large Datasets

When working with datasets containing many time series (e.g., annual counts of many species across many sites), further decisions must be made to effectively analyse the data and deduce the impact of an intervention.

Where possible, the best strategy is to analyse all data in one model that includes random or fixed factors to account for correlations within groups (e.g., species and sites, see [9] for an example). However, this strategy relies on some key assumptions. First, that there are no all-zeroes in the dataset, as these could be interpreted incorrectly by the model (Figure 2A), and second, that each time series will respond to the intervention in a similar way, as random intercept models assume parallel slopes. Random slope models are a solution but they can be very time consuming to run with a large dataset.

An alternative solution, if these assumptions are not met, is to run individual models of each time series separately and summarise outcomes. For example, if there are not cases of all-zeroes in the dataset, the immediate and trend coefficients from individual models can be compared on a scatter plot, as demonstrated in Figure I. In this case, we can see that most points cluster in the positive immediate change and positive trend change quadrant. Although there are some cases of opposing outcomes, it seems that overall, there has been a positive impact from this intervention.

If there are some time series with all-zeroes in the dataset, the situation becomes more complex as these time series can only have a categorical outcome (e.g., counts in the before period, all-zeroes in the after period would be a negative impact), while the time series that do not have all zeroes will have model coefficients. One solution is to categorise the data, based on careful consideration of the different outcomes expected, and how they could be reflected. For example, a study assessing protected area effectiveness, might define a positive impact as any positive change, whether immediate or trend change from a model, or where a population has immigrated to the site. Each population time series could then be classified as a 'positive', 'neutral', or 'negative' outcome (see S6 in the supplemental information online), and the number of time series in each category examined to assess impact.

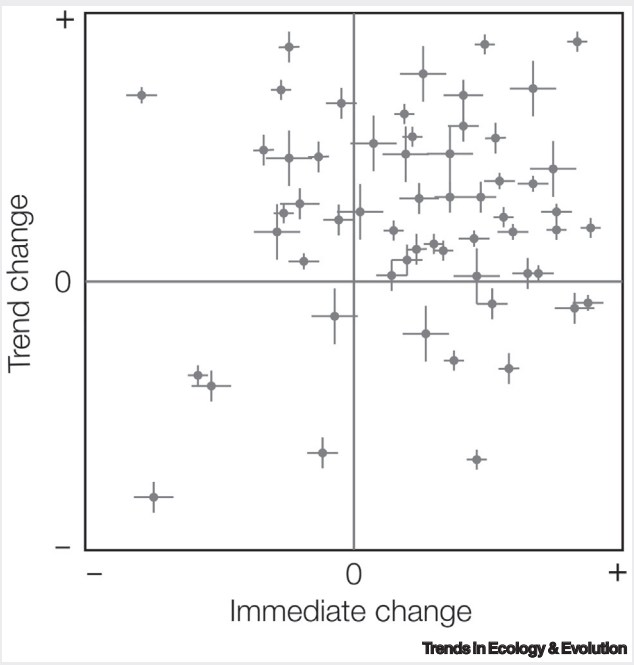

Figure I. Hypothetical Pot of Model Coefficient Output from Comparisons of Immediate and Trend Change Across Many Time Series (each point).

Considerations

There are certain attributes of observational data that can further complicate time-series impact evaluation, especially when time series have trends. If not properly considered, these can lead to additional inadvertent mistakes (see S2 in the supplemental information online). Here, are three considerations and how to take them into account.

All-Zeroes

Time series in ecology can have 'all zeroes' (or, more often, almost all zeroes, but we use the term for any such case), in the years before or after an intervention. For example, if a restoration project establishes new habitat, a population might immigrate to the area, and a survey of the site would return only zero counts in the years before the restoration, but non-zero counts after. Similarly, a fire might destroy all suitable habitat of a population, leaving it locally extinct in the after period. Analysing such cases is difficult, as trend change models will interpret an 'all-zero' population as stable (no trend) at zero. Figure 2A shows a case where a population has gone locally extinct post-intervention, but where the trend change has gone from declining to stable, and so shows a positive impact. For analyses with a small number of time series this is easy enough to manage manually, or by changing to an average/immediate change framework. However, for large datasets that are being analysed in a trend framework, all-zero cases need to be analysed separately. See Box 3 for suggestions on how to do this.

Time Lags and Breakpoints

In some cases, there may be the expectation of a time lag between when an intervention occurs, and when it impacts a population. For example, there may be a lag period of a number of years between a new conservation policy being introduced, and when management starts. In these cases, the 'lag period' (between the two vertical lines in Figure 2B) can be excluded from analysis by removing the

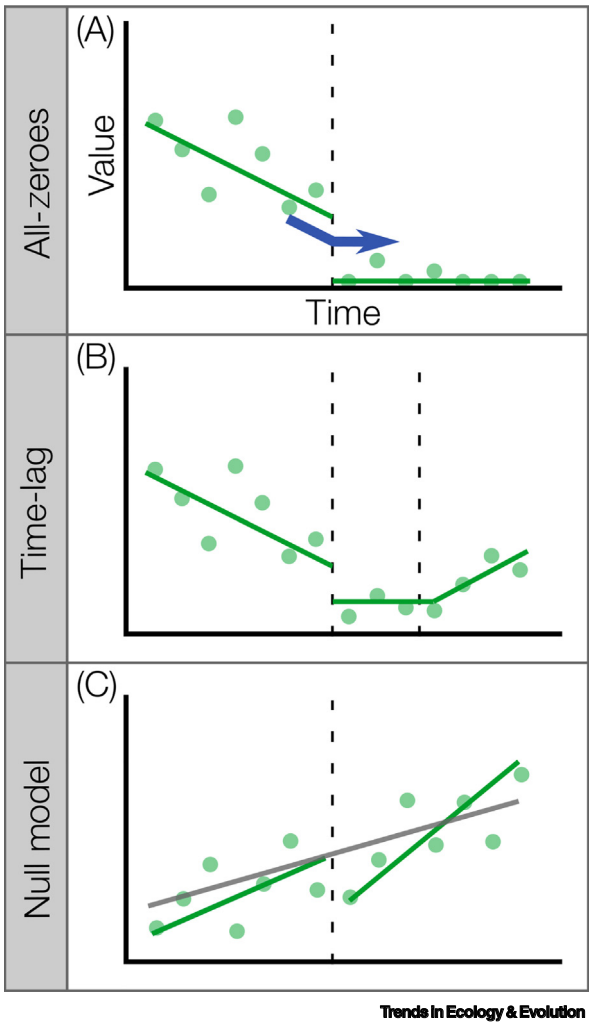

Figure 2. Three Considerations when Conducting BA or BACI Time Series Impact Evaluation (intervention shown by vertical broken line). Only BA shown in this figure for simplicity. (A) are there cases of all-zeroes? (B) Is there likely to be a lag time between the intervention (left broken line) and when the population responds (right broken line)? (C) Does the impact evaluation model perform better than a null model which does not include the intervention (grey line)?

lag years, or by being modelled separately [54,55]. Alternatively, the time of the intervention can be shifted to when it would be expected to start taking effect. Often, this point is not known, in which case various methods are available to statistically identify the most likely breakpoint between before and after [56–59], though this can still be difficult to ascertain at times and there is more work to be done in this area. Once the breakpoint has been identified either through knowledge of the system, or through breakpoint estimation, analysis can proceed, adjusting the 'before' and 'after' years of the datasets to centre around the estimated breakpoint.

Is There Any Change at All?

Finally, it is prudent to ascertain whether there is, in fact, any impact at all from the intervention, as in some cases it may be plausible that the outcome variable has not responded (Figure 2C). The models discussed here explicitly fit a break in the time series from before to after, but the model may fit better without this. To check, we recommend running two models, one with the full model of interest (average/immediate/trend and BA/BACI; Box 1), and the other the same model but with the BA term removed. The performance of these two models can be compared (e.g., through hypothesis testing or information-theoretic approaches), and if the model without

the BA term fits better, then the most parsimonious model does not include any impact from the intervention.

Concluding Remarks

Effective conservation decisions require a robust understanding of how interventions and environmental shocks affect biodiversity. Time series data in ecology offer rich opportunities for causal inference, but care is needed to avoid drawing erroneous conclusions, especially with large datasets. Devising simple and generalisable ways to include non-linear responses would greatly increase the power of impact evaluations (see Outstanding Questions), particularly as the length of available time series increase. Some methods are explored by [29], but there is further work to be done. By using the framework presented, ecologists and conservationists can avoid misinterpreting the effectiveness of conservation measures, and the impact of environmental disasters, providing the best opportunities for effective and efficient conservation decisions.

Outstanding Questions

Is there a generalisable way to conduct impact evaluation when responses are not linear? By necessity, the models we describe, and those that are used by most disciplines for impact evaluation with trends, assume linearity, or that linearity can be achieved through transformation. Though some methods are described to deal with nonlinear models, there is as yet no framework for making robust conclusions from nonlinear data.

Are there better ways to incorporate all-zeroes into a trend change modelling framework? This article suggests analysing cases of all-zeroes separately, but a better solution would be a model that can distinguish between a stable trend, and a case of all-zeroes, indicating extinction or immigration.

How can analyses that identify the breakpoint be improved when there is no obvious change?

How can the terminology used in impact evaluation be aligned between fields, to improve sharing of methods? Different fields have long traditions of particular terminologies.

Acknowledgements

We thank Andrea Manica and Rob Freckleton for comments that greatly improved the manuscript. H.S.W. was supported by the Cambridge Trust Poynton Scholarship, Cambridge Department of Zoology J.S. Gardiner Studentship, and Cambridge Philosophical Society; T.A. was supported by the Australian Research Council Future Fellowship (FT180100354), and the University of Queensland strategic funding; J.G. was supported by European Union's Horizon 2020 Marie Skłodowska-Curie program (No. 706784), and VILLUM FONDEN (VKR023371); B.I.S. was supported by a Royal Commission for the Exhibition of 1851 Research Fellowship; W.J.S. is funded by Arcadia and J.P.G.J. was supported by the Leverhulme Trust: RPG-2014-056. CBC. Data are provided by National Audubon Society and through the generous efforts of Bird Studies Canada and countless volunteers across the western hemisphere.

Supplemental Information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tree.2020.11.001

References

1. Bruner, A.G. *et al.* (2001) Effectiveness of parks in protecting tropical biodiversity. *Science* 291, 125–128
2. Cuenca, P. *et al.* (2016) How much deforestation do protected areas avoid in tropical Andean landscapes? *Environ. Sci. Policy* 56, 56–66
3. Geldmann, J. *et al.* (2013) Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. *Biol. Conserv.* 161, 230–238
4. Ripple, W.J. and Beschta, R.L. (2012) Trophic cascades in Yellowstone: the first 15 years after wolf reintroduction. *Biol. Conserv.* 145, 205–213
5. McCleery, R. *et al.* (2014) Better off in the wild? Evaluating a captive breeding and release program for the recovery of an endangered rodent. *Biol. Conserv.* 169, 198–205
6. Moreno, R. *et al.* (2013) Ten years after the Prestige Oil Spill: seabird trophic ecology as indicator of long-term effects on the coastal marine ecosystem. *PLoS One* 8, e77360
7. Krauss, J. *et al.* (2010) Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels. *Ecol. Lett.* 13, 597–605
8. Woodcock, B.A. *et al.* (2016) Impacts of neonicotinoid use on long-term population changes in wild bees in England. *Nat. Commun.* 7, 12459
9. Chevalier, M. *et al.* (2019) Changes in forest bird abundance, community structure and composition following a hurricane in Sweden. *Ecography* 42, 1862–1873
10. Baylis, K. *et al.* (2016) Mainstreaming impact evaluation in nature conservation. *Conserv. Lett.* 9, 58–64
11. Lesbarrères, D. and Fahrig, L. (2012) Measures to reduce population fragmentation by roads: what has worked and how do we know? *Trends Ecol. Evol.* 27, 374–380
12. Woodhouse, E. *et al.* (2015) Guiding principles for evaluating the impacts of conservation interventions on human well-being. *Philos. Trans. R. Soc. B Biol. Sci.* 370, 20150103
13. Jung, M. *et al.* (2019) Impacts of past abrupt land change on local biodiversity globally. *Nat. Commun.* 10, 5474
14. Butsic, V. *et al.* (2017) Quasi-experimental methods enable stronger inferences from observational data in ecology. *Basic Appl. Ecol.* 19, 1–10
15. Ferraro, P.J. (2009) Counterfactual thinking and impact evaluation in environmental policy. *New Dir. Eval.* 2009, 75–84
16. Fayle, T.M. *et al.* (2015) Whole-ecosystem experimental manipulations of tropical forests. *Trends Ecol. Evol.* 30, 334–346
17. Wiik, E. *et al.* (2019) Experimental evaluation of the impact of a payment for environmental services program on deforestation. *Conserv. Sci. Pract.* e8, 1
18. Pynegar, E.L. *et al.* (2018) The effectiveness of Payments for Ecosystem Services at delivering improvements in water quality: lessons for experiments at the landscape scale. *PeerJ* 6, e5753
19. Pynegar, E.L. *et al.* (2019) What role should randomized control trials play in providing the evidence base for conservation? *Oryx* Published online October 24, 2019. https://doi.org/10.1017/S0030605319000188
20. Larsen, A.E. *et al.* (2019) Causal analysis in control–impact ecological studies with observational data. *Methods Ecol. Evol.* 10, 924–934
21. Stewart-Oaten, A. *et al.* (1986) Environmental impact assessment: 'pseudoreplication' in time? *Ecology* 67, 929–940
22. Underwood, A.J. (1994) On beyond BACI: sampling designs that might reliably detect environmental disturbances. *Ecol. Appl.* 4, 3–15
23. Karapinar-Çarkıt, F. *et al.* (2019) The effect of a pharmaceutical transitional care program on rehospitalisations in internal medicine patients: an interrupted-time-series study. *BMC Health Serv. Res.* 19, 717
24. Kontopantelis, E. *et al.* (2015) Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *BMJ* 350, h2750

25. Laliotis, I. *et al.* (2016) Total and cause-specific mortality before and after the onset of the Greek economic crisis: an interrupted time-series analysis. *Lancet Public Heal.* 1, e56–e65
26. Lopez Bernal, J. *et al.* (2018) The use of controls in interrupted time series studies of public health interventions. *Int. J. Epidemiol.* 47, 2082–2093
27. Markowitz, A.J. (2018) Changes in school engagement as a function of no child left behind: A comparative interrupted time series analysis. *Am. Educ. Res. J.* 55, 721–760
28. Gottman, J.M. and Glass, G.V. (1978) Analysis of interrupted time-series experiments. In *Single Subject Research: Strategies for Evaluating Change* (Kratochwill, T.R., ed.), pp. 197–236, Academic Press
29. Thiault, L. *et al.* (2017) Progressive-change BACIPS: a flexible approach for environmental impact assessment. *Methods Ecol. Evol.* 8, 288–296
30. Thiault, L. *et al.* (2019) Ecological evaluation of a marine protected area network: a progressive-change BACIPS approach. *Ecosphere* 10, e02576
31. Chevalier, M. *et al.* (2019) New measures for evaluation of environmental perturbations using before-after-control-impact analyses. *Ecol. Appl.* 29, e01838
32. Fisher, R. *et al.* (2019) epower: an r package for power analysis of before-after-control-impact (BACI) designs. *Methods Ecol. Evol.* 10, 1843–1853
33. Sansom, A. *et al.* (2016) Negative impact of wind energy development on a breeding shorebird assessed with a BACI study design. *Ibis* 158, 541–555
34. Vieira, J.V. *et al.* (2016) Assessment the short-term effects of wrack removal on supralittoral arthropods using the M-BACI design on Atlantic sandy beaches of Brazil and Spain. *Mar. Environ. Res.* 119, 222–237
35. Jones, G.M. *et al.* (2016) Megafires: an emerging threat to old-forest species. *Front. Ecol. Environ.* 14, 300–306
36. Nelson, D. *et al.* (2017) Experimental whole-stream warming alters community size structure. *Glob. Chang. Biol.* 23, 2618–2628
37. Terraube, J. *et al.* (2020) Assessing the effectiveness of a national protected area network for carnivore conservation. *Nat. Commun.* 11, 2957
38. Wauchope, H. *et al.* (2019) Quantifying the impact of protected areas on near-global waterbird population trends, a pre-analysis plan. *PeerJ Prepr.* 7, e27741v
39. Bálint, M. *et al.* (2018) Environmental DNA time series in ecology. *Trends Ecol. Evol.* 33, 945–957
40. Pasquarella, V.J. *et al.* (2016) From imagery to ecology: leveraging time series of all available Landsat observations to map and monitor ecosystem state and dynamics. *Remote Sens. Ecol. Conserv.* 2, 152–170
41. Horns, J.J. *et al.* (2018) Using opportunistic citizen science data to estimate avian population trends. *Biol. Conserv.* 221, 151–159
42. Dennis, E.B. *et al.* (2017) Using citizen science butterfly counts to predict species population trends. *Conserv. Biol.* 31, 1350–1361
43. Hughes, B.B. *et al.* (2017) Long-term studies contribute disproportionately to ecology and policy. *Bioscience* 67, 271–281
44. Brüderl, J. and Ludwig, V. (2015) Fixed-effects panel regression. In *The Sage Handbook of Regression Analysis and Causal Inference* (1st edn) (Best, H. and Wolf, C., eds), pp. 327–357, Sage, Los Angeles
45. 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, e0141380
46. Ferraro, P.J. and Miranda, J.J. (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
47. Callaway, B. and Sant'Anna, P.H.C. (2020) Difference-in-Differences with Multiple Time Periods. *arXiv* Published online August 18, 2020. https://arXiv.org1803.09015
48. Christie, A.P. *et al.* (2019) Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *J. Appl. Ecol.* 56, 2742–2754
49. Stuart, E.A. (2010) Matching methods for causal inference: a review and a look forward. *Stat. Sci.* 25, 1–21
50. Schleicher, J. *et al.* (2020) Statistical matching for conservation science. *Conserv. Biol.* 34, 538–549
51. Linden, A. (2018) A matching framework to improve causal inference in interrupted time-series analysis. *J. Eval. Clin. Pract.* 24, 408–415
52. Ho, D. *et al.* (2011) MatchIt: nonparametric preprocessing for parametric causal inference. *J. Stat. Softw.* 42, 1–28
53. King, G. and Nielsen, R. (2019) Why propensity scores should not be used for matching. *Polit. Anal.* 27, 435–454
54. Wagner, A.K. *et al.* (2002) Segmented regression analysis of interrupted time series studies in medication use research. *J. Clin. Pharm. Ther.* 27, 299–309
55. Torres, A. *et al.* (2011) Assessing the effects of a highway on a threatened species using Before–During–After and Before–During–After-Control–Impact designs. *Biol. Conserv.* 144, 2223–2232
56. Zeileis, A. *et al.* (2002) Strucchange: an R package for testing for structural change in linear regression models. *J. Stat. Softw.* 7, 1–38
57. Zeileis, A. *et al.* (2003) Testing and dating of structural changes in practice. *Comput. Stat. Data Anal.* 44, 109–123
58. Muggeo, V.M.R. (2017) Interval estimation for the breakpoint in segmented regression: a smoothed score-based approach. *Aust. N. Z. J. Stat.* 59, 311–322
59. Linden, A. and Yarnold, P.R. (2016) Using machine learning to identify structural breaks in single-group interrupted time series designs. *J. Eval. Clin. Pract.* 22, 855–859
60. McDonald, T.L. *et al.* (2000) Analysis of count data from before-after control-impact studies. *J. Agric. Biol. Environ. Stat.* 5, 262
61. Linden, A. (2015) Conducting interrupted time-series analysis for single- and multiple-group comparisons. *Stata J.* 15, 480–500