Experimenting with the past to improve environmental monitoring programs

Easton R. White[1]* and Christie A. Bahlai[2]

[1]Department of Biology, University of Vermont, VT 05405, USA,
[2]Department of Biological Sciences, Kent State University, OH 44242, USA

Abstract Long-term monitoring programs are a fundamental part of both understanding system dynamics and making management decisions. Yet, monitoring programs are often created without considering statistical power, site selection, or the full costs and benefits of monitoring. Further, data from monitoring programs with different goals and protocols are now being combined for comparative analyses. Key considerations can be incorporated into the optimal design of a management program with simulations and experiments. Here, we advocate for the expanded use of a third approach: non-random sampling of previously collected data. This approach conducts experiments with available data to understand the consequences of different monitoring approaches. We first illustrate this approach in the context of monitoring programs to assess species trends. We then apply the approach to a pair of additional, more general case studies.

Keywords: statistical power, population trends, data-poor fisheries, species monitoring

Long-term environmental monitoring

Long-term species monitoring programs are an essential piece of modern ecological research and conservation science (Hughes et al. 2017; Bahlai et al., in review). Numerous studies have demonstrated that long-term monitoring can have disproportionately large contributions in terms of advancing scientific understanding and policy (Giron-Nava et al. 2017). Species monitoring programs, like the USA-based Long Term Ecological Research (LTER) Network, as well as compilations of time series, like the Living Planet Index, show the scope of long-term datasets now available (Magurran et al. 2010, Foundation 2016). Furthermore, with the advent of infrastructure that connects and stores data collected by a wide variety of professional and amateur naturalists, monitoring should continue to become more feasible and cost-effective. Large-scale citizen science programs, like iNaturalist (`https://www.inaturalist.org/`) and eBird (`https://ebird.org/home`), allow for increased data collection as well as data use and resuse (Sullivan et al. 2009, Joppa 2017). Similarly, numerous new technologies, including eDNA and drones, will bring down the cost of monitoring through automation and increasing the sheer taxonomic, temporal, and spatial resolution of observations(Bohmann et al. 2014, Hodgson et al. 2018). All of these efforts will lead to increases in the number of species monitored as well as the quantity and quality of the data collected, to previously unimaginable levels.

*Easton.White@uvm.edu

Panel 1: Key considerations when designing monitoring programs

A long-term monitoring program should be evaluated and designed with several key considerations in mind (Legg and Nagy 2006, Lindenmayer and Likens 2010, White 2019):

- Precision and accuracy of monitoring approach
- The number of sites to monitor
- Correlations between sites (e.g. species biology that will affect the movement of individuals)
- Spatial scale at which to monitor
- The frequency and length of monitoring
- The cost (both fiscally and in terms of person-hours) that should be devoted to monitoring
- System dynamics (e.g. autocorrelation, cycles, trends) that will affect variability in observations
- Changes in the system dynamics over time (e.g. management action that affects species trend)
- The complexity and specifics of the question of interest, e.g. it will take longer to detect rare events than to measure year-to-year variability
- Changes in sampling design over time or variability between observers/sensors

Species monitoring program design

Despite the recognized importance of long-term monitoring programs, key questions remain. Are long-term monitoring programs designed in a way to address key questions of interest (Legg and Nagy 2006, Nichols and Williams 2006, Field et al. 2007, McDonald-Madden et al. 2010, Lindenmayer and Likens 2010, Lindenmayer et al. 2012)? For instance, suppose that a monitoring site was chosen originally to monitor long-term changes in a bird population near a university. A long-term study could certainly reveal the population trend for that specific location. However, the site may have been chosen specifically because it was at high abundance at the beginning of the study—this causes a site-selection bias (Fournier et al. 2019). Populations naturally vary, both in time and in space, so the very act of initially selecting a site to monitor with particular population attributes can potentially confound the very patterns they seek to monitor. Suppose the birds in this population undergo a cyclic dynamic related to resource exploitation, or rotate between different patches for nesting from year to year. Thus, when we ask new questions of long-term monitoring data, we have to think carefully about how the monitoring program was originally designed and whether or not we have adequate statistical power (Lindenmayer and Likens 2010, White 2019; and Panel 1). Furthermore, an understanding of the spatial and temporal scale at which the data is taken—and how that relates to the life history and ecology of the species—is essential. These considerations, amongst others (Panel 1), are especially relevant when data from different sources are combined for comparisons—which is increasingly performed (Maguran et al. 2010, Keith et al. 2015, Giron-Nava et al. 2017, White 2019, Saunders et al. 2019). Lastly, the tradeoff between information gained from monitoring and the cost of monitoring has to be considered (Bennett et al. 2018).

To address these issues, there are three classes of tools available to design and evaluate monitoring programs. First, the most commonly used approach are simulation models

(Gerrodette 1987, Rhodes and Jonzen 2011). Using prior knowledge about the system under question, simulation models can be constructed to incorporate key factors that affect species dynamics. With an appropriate model, simulations can then be run for a variety of scenarios, including changing the number of samples taken per year, altering the number of sites sampled, and sampling for different lengths of time (Rhodes and Jonzen 2011, Barry et al. 2017, White 2019, Christie et al. 2019). Although powerful, this approach is limited to systems in which many aspects of the biology are already known to some extent.

Second, experiments can also be used to test the effect of different sampling protocols. As in the case of simulation models, experiments with different levels of monitoring, or different monitoring approaches, can be used. A related approach would simply be to compare different sampling regimes across systems to evaluate which are the most successful. Indeed, integrated population modelling was developed as an analytical approach to identify and address data discrepancies between data taken by differing methodologies or at differing times in a species's life history (Saunders et al. 2019). This method has been applied with great success to advance understanding of the trajectories of populations of well-monitored taxa such as waterfowl (Arnold et al. 2018). However, the key disadvantage of this approach is, like simulation models, integrative modelling approaches are reliant on the availability of large amounts of data, documenting multiple facets of a species' biology. Of course, these types of experiments providing multi-faceted data are often infeasible or impossible for many systems.

Here, we advocate for the expanded use of a third approach: **non-random sampling of previously collected monitoring data** (White 2019). This concept leverages existing information by starting with long-term monitoring data already collected for a system. The data can then be subsampled, or divided, in various ways depending on the question of interest (Fig. 1a). Then a metric (for example, a mean or a slope) could be calculated for each subsample. Each subsample metric would then be compared to the metric for complete data (all the data combined). The complete data acts as a "true value" for comparison. This is analogous to simulation studies where the true parameters are known (Bolker 2008). We learn about the elements of a good monitoring program by examining which subsamples of the data are most influential and the number of subsamples needed to have a high probability of detecting the true value of the metric.

This approach is best described with a simple example (Fig. 1b). White (2019) studied how many years of monitoring were required to detect population trends. For each time series, White (2019) examined all possible subsamples of different lengths of time. He then calculated the population trend for each subsample. The fraction of subsamples of a particular length, that had the same overall trend as the complete time series (i.e. the "true trend"), is the statistical power. Thus, the minimum time series required was the time series length that met a high enough threshold of statistical power. (White 2019) was able to use this approach on 822 time series, allowing for comparison across species and systems. Using a similar approach, (Wauchope et al. 2019) examined both the minimum time and frequency of sampling required to be confident in determining species trends. Using resampling of the breeding bird survey, they found that sampling for a short period, or infrequently, was adequate to determine the species trend direction, i.e. positive or negative. However, more frequent and longer monitoring was required to estimate the percent changes over time.

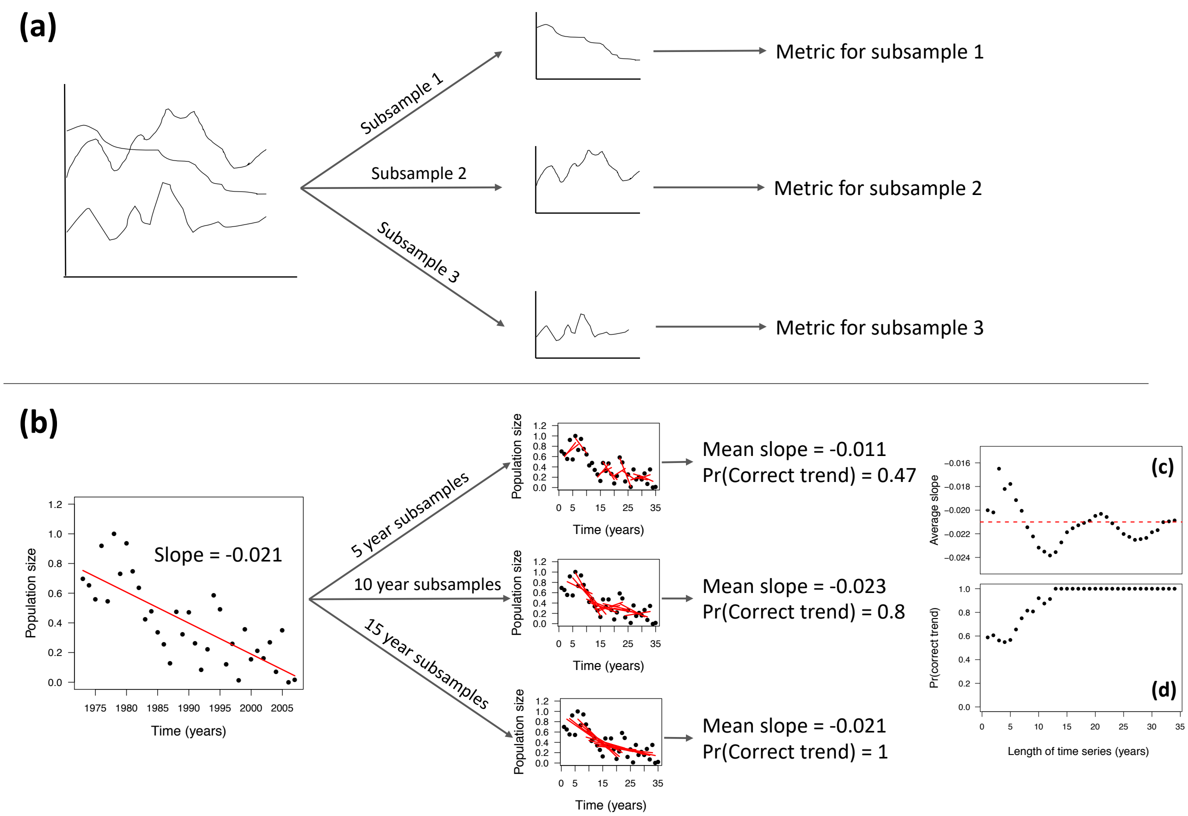

Figure 1: (a) The general process of non-random sampling of past data from left to right (i.e. sequentially, starting with data from farthest into the past) includes: dividing data into subsamples, calculating metrics on those subsamples, and comparing the subsample metrics to the combined (i.e. "true metric") dataset, (b) same process as panel (a), but for specific example of examining the minimum number of years required to detect long-term population trends (White 2019). The pair of figures on the bottom right show how the (c) average slope and the (d) probability of correcting identifying a trend change with the number of years monitored.

Non-random sampling in other contexts

Non-random sampling of past data can be applied to a variety of contexts beyond estimating long-term population trends. For example, consider the importance of studying data-poor fisheries (Dowling et al. 2015 and Table 1 of Chrysafi and Kuparinen 2015). Using mostly simulation models, a lot of work has been done to develop tools when observational data is limited. To study data-poor fisheries using non-random sampling, one should instead study data-rich fisheries. The goal would be to artificially degrade the data-rich examples until the point that the fishery would be considered data-poor (Fig. 2). We can then see how various methods of data-poor fisheries perform given that we have the full data set to act as a true comparison. The idea is the same as in simulation studies where we know the parameter values of simulated data exactly. As an example, we took data on darkblotched rockfish (*Sebastes crameri*) from the U.S. West Coast Groundfish Bottom Trawl Survey data (Keller et al. 2017, Stock et al. 2019). We then played two "experiments" with the data. First, we suppose we only had access to shallow or deep data because of technology limitations. We show that regression estimates of parameters differs differs based on which depths were included 2b). We also examine the effect of degrading the data to only a fraction of the totals records we have available. We see that model estimates for the effect of being within a rockfish conservation area are not accurate until a large fraction of the original data is included (Fig. 2c).

A practical example of the application of non-random sampling applied to management decision-making can be found in Cusser et al. (2019). Agricultural management recommendations are often based on conclusions from short to medium-term field trials (ca. 1-5 years), and it is common to observe contradictory findings between trials. In this study, the authors applied Bahlai's (2019) non-random sequential sampling algorithm to long-term data examining the effects of tillage practice on productivity and return-on investment. They found that, because of high natural variability in the system, 10 years of data was required to observe the "true" pattern of difference between treatments, and that more than a third of the sampled sequences shorter than 10 years led to outright misleading results (i.e. statistically significant trends which showed the opposite relationship between treatments). Whereas it is unlikely that practitioners making management decisions can consistently rely on a decade of data to guide them, the results of the non-random sequential sampling of long-term data provide guidance on reconciling apparently differing trends between trials.

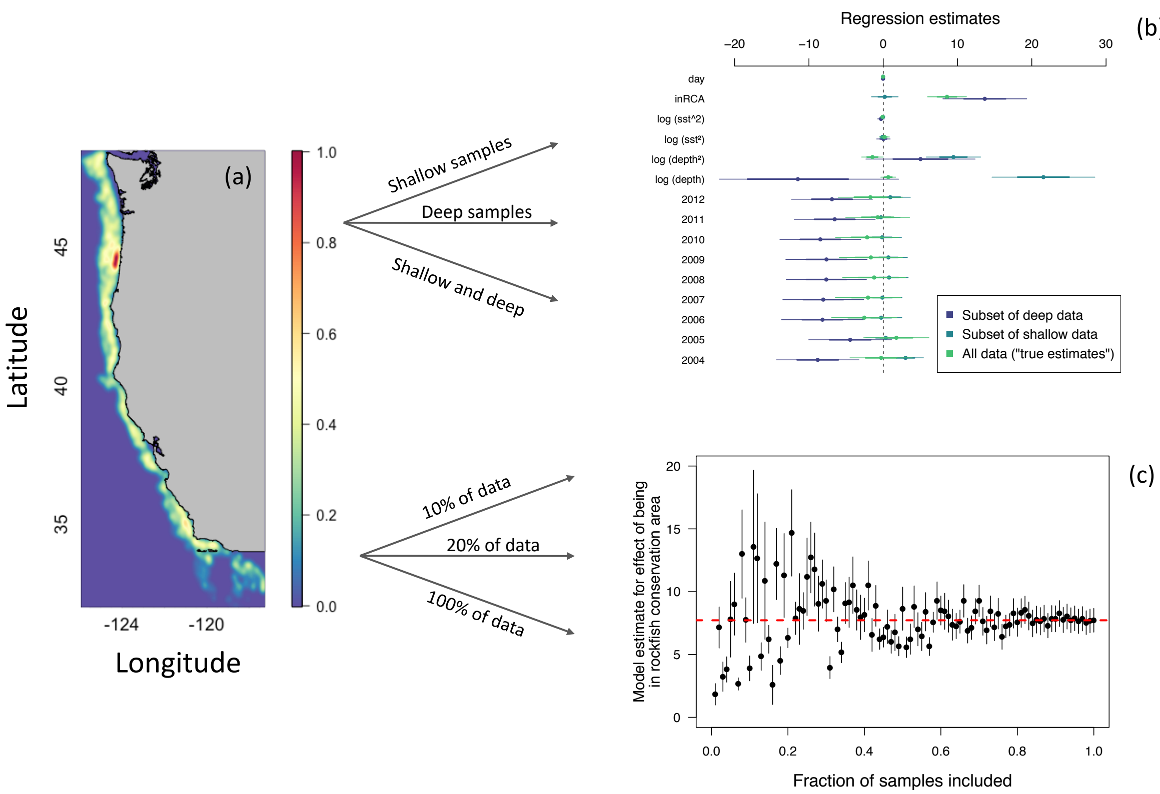

Figure 2: Subsampling process for darkblotched rockfish (*Sebastes crameri*) catch in kg in U.S. West Coast Groundfish Bottom Trawl Survey data from 2003-2012. (a) Bivariate kernel density estimate showing smoothed density of fishing effort (7,161 haul locations) (Keller et al. 2017, Stock et al. 2019) (b) Parameter estimations for linear regression of three subsamples of data: deep trawls, shallow trawls, and all combined data (Note: the number of records was kept consistent for the three groups). (c) Estimate for the effect of being in a rockfish conservation area (inRCA) on catch for different amounts of data included. The horizontal, dashed line is the “true” estimate which is the estimate when 100% of the data is included.

Panel 2: Example questions that could be addressed using non-random sampling

Question	Non-random sampling approach
How many test wells are needed to understand subsurface water flow?	We would start with an example system where a large number of test wells produced accurate dynamics. Then, we artificially degrade this data using less test wells. Last, we would examine when the predicted dynamics change as a result of less test wells.
What is the effect of not being able to identify microorganisms to the species level?	We first select data from a well-resolved tree that does identify organisms to the species level. Then, we artificially degrade the data in a way where we pretend a tree is only resolved to the genus or family level. We could then study the effect of not identifying organisms to the species level.
What is the effect of scuba diving depth limitations on estimates of biological diversity?	We would use high-quality diver survey data that was collected along a gradient of depths. We would then artificially degrade the data by removing deeper dives. We could then compare the diversity metrics when all the data is included versus only shallow dives.

Conclusions

Data from long-term monitoring programs are used in assessing trends in environmental observations, understanding system dynamics, and making management decisions. It is critical that these monitoring programs be designed in order to address our questions of interest. This is particularly relevant when new questions are asked of monitoring data or data from disparate monitoring programs are combined. We show that non-random sampling of past monitoring programs can be used to understand sampling requirements and the consequences of bias (Figs. 1,2). This approach can be applied to a variety of systems and questions (Panel 2). Combined with simulations and experimental approaches, we argue that non-random sampling of past data should be used more widely to study questions related to sampling design. More work in this area will allow scientists and managers to better evaluate past efforts and to design new monitoring programs using evidence-based approaches.

References

Arnold, T. W., Clark, R. G., Koons, D. N., and Schaub, M. Integrated population models facilitate ecological understanding and improved management decisions. *Journal of Wildlife Management*, 82(2):266–274, 2018. ISSN 19372817. doi: 10.1002/jwmg.21404.

Bahlai, C. The bad breakup Algorithm. *Zenodo*, page Version v1.0, 2019. doi: 10.5281/zenodo.2561051.

Barry, J., Maxwell, D., Jennings, S., Walker, D., and Murray, J. Emon: an R-package to support the design of marine ecological and environmental studies, surveys and monitoring programmes. *Methods in Ecology and Evolution*, 8(10):1342–1346, 2017. ISSN 2041210X. doi: 10.1111/2041-210X.12748.

Bennett, J. R., Maxwell, S. L., Martin, A. E., Chadès, I., Fahrig, L., and Gilbert, B. When to monitor and when to act: Value of information theory for multiple management units and limited budgets. *Journal of Applied Ecology*, 55(5):2102–2113, 2018. ISSN 13652664. doi: 10.1111/1365-2664.13132.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., and de Bruyn, M. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, 29(6):358–367, 2014. ISSN 01695347. doi: 10.1016/j.tree.2014.04.003.

Bolker, B. M. *Ecological Models and Data in R.* Princeton University Press, Princeton, New Jersey, 1 edition, 2008.

Christie, A. P., Amano, T., Martin, P. A., Shackelford, G. E., Simmons, B. I., and Sutherland, W. J. Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*, (May):1–13, 2019. ISSN 0021-8901. doi: 10.1111/1365-2664.13499.

Chrysafi, A. and Kuparinen, A. Assessing abundance of populations with limited data: Lessons learned from data-poor fisheries stock assessment. *Environmental Reviews*, 24(1): 25–38, 2015. ISSN 11818700. doi: 10.1139/er-2015-0044.

Cusser, S., Bahlai, C., Swinton, S. M., Robertson, G. P., and Haddad, N. M. Long-term research needed to avoid spurious and misleading trends in sustainability attributes of no-till. *bioRxiv*, pages 1–31, 2019. doi: 10.1101/788240.

Dowling, N. A., Dichmont, C. M., Haddon, M., Smith, D. C., Smith, A. D., and Sainsbury, K. Empirical harvest strategies for data-poor fisheries: A review of the literature. *Fisheries Research*, 171:141–153, 2015. ISSN 01657836. doi: 10.1016/j.fishres.2014.11.005.

Field, S. A., Connor, P. J. O., Tyre, A. J., and Possingham, H. P. Making monitoring meaningful. *Austral Ecology*, 32:485–491, 2007. doi: 10.1111/j.1442-9993.2007.01715.x.

Foundation, W. W. Living Planet Report 2016 - Risk and resilience in a new era. Technical report, 2016.

Fournier, A. M., White, E. R., and Heard, S. B. Site-selection bias and apparent population declines in long-term studies. *Conservation Biology*, 00(0):1–10, 2019. ISSN 15231739. doi: 10.1111/cobi.13371.

Gerrodette, T. A power analysis for detecting trends. *Ecology*, 68(5):1364–1372, 1987. ISSN 00129658. doi: 10.2307/1939220.

Giron-Nava, A., James, C. C., Johnson, A. F., Dannecker, D., Kolody, B., Lee, A., Nagarkar, M., Pao, G. M., Ye, H., Johns, D. G., and Sugihara, G. Quantitative argument for long-term ecological monitoring. *Marine Ecology Progress Series*, 572:269–274, 2017.

Hodgson, J. C., Mott, R., Baylis, S. M., Pham, T. T., Wotherspoon, S., Kilpatrick, A. D., Raja Segaran, R., Reid, I., Terauds, A., and Koh, L. P. Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution*, 9(5):1160–1167, 2018. ISSN 2041210X. doi: 10.1111/2041-210X.12974.

Hughes, B. B., Beas-luna, R., Barner, A. K., Brewitt, K., Brumbaugh, D. R., Cerny-Chipman, E. B., Close, S. L., Coblentz, K. E., Nesnera, K. L. D., Drobnitch, S. T., Figurski, J. D., Focht, B., Friedman, M., Freiwald, J., Heady, K. K., Heady, W. N., Hettinger, A., Johnson, A., Karr, K. A., Mahoney, B., Moritsch, M. M., Osterback, A.-M. K., Reimer, J., Robinson, J., Rohrer, T., Rose, J. M., Sabal, M., Segui, L. M., Shen, C., Sullivan, J., Zuercher, R., Raimondi, P. T., Menge, B. A., Grorud-Colvert, K., Novak, M., and Carr, M. H. Long-term studies contribute disproportionately to ecology and policy. *BioScience*, 67(3):271–281, 2017. doi: 10.1093/biosci/biw185.

Joppa, L. N. The case for technology investments in the environment. *Nature*, 552(7685): 325–328, 2017. ISSN 0028-0836. doi: 10.1038/d41586-017-08675-7.

Keith, D., Akçakaya, H. R., Butchart, S. H., Collen, B., Dulvy, N. K., Holmes, E. E., Hutchings, J. A., Keinath, D., Schwartz, M. K., Shelton, A. O., and Waples, R. S. Temporal correlations in population trends: Conservation implications from time-series analysis of diverse animal taxa. *Biological Conservation*, 192:247–257, 2015. ISSN 00063207. doi: 10.1016/j.biocon.2015.09.021.

Keller, A. A., Wallace, J. R., and Methot, R. D. The Northwest Fisheries Science Center's West Coast Groundfish Bottom Trawl Survey: History, Design, and Description. Technical Report January, 2017.

Legg, C. J. and Nagy, L. Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management*, 78:194–199, 2006. doi: 10.1016/j.jenvman.2005.04.016.

Lindenmayer, D. B. and Likens, G. E. The science and application of ecological monitoring. *Biological Conservation*, 143(6):1317–1328, 2010. ISSN 00063207. doi: 10.1016/j.biocon.2010.02.013.

Lindenmayer, D. B., Gibbons, P., Bourke, M., Burgman, M., Dickman, C. R., Ferrier, S., Fitzsimons, J., Freudenberger, D., Garnett, S. T., Groves, C., Hobbs, R. J., Kingsford,

R. T., Krebs, C., Legge, S., Lowe, A. J., Mclean, R., Montambault, J., Possingham, H., Radford, J., Robinson, D., Smallbone, L., Thomas, D., Varcoe, T., Vardon, M., Wardle, G., Woinarski, J., and Zerger, A. Improving biodiversity monitoring. *Austral Ecology*, 37 (3):285–294, 2012. ISSN 14429985. doi: 10.1111/j.1442-9993.2011.02314.x.

Maguran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M., Elston, D. A., Scott, E. M., Smith, R. I., Somerfield, P., and Watt, A. D. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in Ecology and Evolution*, 25(10):574–582, 2010. ISSN 01695347. doi: 10.1016/j.tree.2010.06.016.

Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M., Elston, D. A., Scott, E. M., Smith, R. I., Somerfield, P. J., and Watt, A. D. Long-term datasets in biodiversity research and monitoring : assessing change in ecological communities through time. *Trends in Ecology and Evolution*, 25:574–582, 2010. doi: 10.1016/j.tree.2010.06.016.

McDonald-Madden, E., Baxter, P. W. J., Fuller, R. A., Martin, T. G., Game, E. T., Montambault, J., and Possingham, H. P. Monitoring does not always count. *Trends in Ecology and Evolution*, 25(10):547–550, 2010. ISSN 01695347. doi: 10.1016/j.tree.2010.07.002.

Nichols, J. D. and Williams, B. K. Monitoring for conservation. *Trends in Ecology and Evolution*, 21(12):668–673, 2006. doi: 10.1016/j.tree.2006.08.007.

Rhodes, J. R. and Jonzen, N. Monitoring temporal trends in spatially structured populations: how should sampling effort be allocated between space and time? *Ecography*, 34(6):1040–1048, 2011. ISSN 09067590. doi: 10.1111/j.1600-0587.2011.06370.x.

Saunders, S. P., Farr, M. T., Wright, A. D., Bahlai, C. A., Ribeiro, J. W., Rossman, S., Sussman, A. L., Arnold, T. W., and Zipkin, E. F. Disentangling data discrepancies with integrated population models. *Ecology*, 100(6):1–14, 2019. ISSN 00129658. doi: 10.1002/ecy.2714.

Stock, B. C., Ward, E. J., Eguchi, T., Jannot, J. E., Thorson, J. T., Feist, B. E., and Semmens, B. X. Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model framework. *Canadian Journal of Fisheries and Aquatic Sciences*, pages 1–66, 2019. doi: https://doi.org/10.1139/cjfas-2018-0281.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., and Kelling, S. eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10):2282–2292, 2009. ISSN 00063207. doi: 10.1016/j.biocon.2009.05.006.

Wauchope, H. S., Johnston, A., Amano, T., and Sutherland, W. J. When can we trust population trends? A method for quantifying the effects of sampling interval and duration. *Methods in Ecology and Evolution*, page 498170, 2019. doi: https://doi.org/10.1111/2041-210X.13302.

White, E. R. Minimum time required to detect population trends: the need for long-term monitoring programs. *BioScience*, 69(1):40–46, 2019.