

# PhD plan and Research Proposal

Nathan Khadaroo - 8520000

## Introduction / General theme:

This document serves as an outline for the PhD. It is generated using an R Markdown script which can be found here, alongside all previous versions of the document.

The general theme of the PhD is to explore the use of statistical modeling to address bias in citizen science generated data in the social sciences.

I will first provide a non-exhaustive review of the literature, I will then outline some proposals for projects/papers that would constitute the substantive of the PhD thesis.

## Literature review:

In this section I provide a non-exhaustive review of the literature on bias in citizen science data sets.

Bird, et al (2014) provide a detailed overview of statistical solutions to issues...

as well as a list of available R packages for their implementation.

## Project proposal ideas:

**Systematic review:** The first project I plan to undertake is a systematic review of citizen science projects which use statistical techniques to address bias in the data collection process.

There is no existing systematic review of the ways statistically modeling has been applied to citizen science data sets across fields. This would therefore constitute a meaningful contribution to the literature.

There are existing reviews such as Bird, et al (2014). However they are field specific (in this case biology), and are non systematic. This exposes the findings to a higher risk of bias, and prevents the authors from making quantitative statements about the relative prevalence of various approaches.

Key aspects of studies which would have to be recorded would be the type of bias which is being addressed  
( ) This could potentially also be my masters dissertation.

**Evaluating solutions:** This project, which would necessarily take place after and build upon the systematic review, would involve evaluating various solutions to bias in

## Developing a unified framework:

## Training needs:

In this section I outline skills which I would need to acquire to accomplish this project.

- A good understanding of a framework (Bayesian?) with which to evaluate and compare various approaches.
- Systematic review training, the university does run some of these (the most recent unfortunately clashed with Sheffield) I think based on Prisma and Cochrane approach. I will also look into NIRO.
- Occupancy modeling *appears* to be the most prominent approach in species monitoring, it could be fruitful to audit a biology module on this, though I can't see any. Additionally MacKenzie et al (2017) looks like a useful overview.
- Small area estimation is quite common with crowd-sourced data (could occupancy be considered a special case of this??). I understand there is a NCRM course on the r sae package but I have been told it isn't fantastic, David has kindly sent me the slides.

## **Elective module choice:**

It is fairly difficult to get detailed information on most courses.

Some potential takes could be:

- Longitudinal data analysis with Alex Cernat.
- Possibly Complex Survey Designs and Analysis though I'm not sure of the software used, and the MARD module by the same lecture was underwhelming and seemed outdated (the focus was mainly on the now quite stale debate between qualitative and quantitative approaches to social research, no mention of currently active debates such as those surrounding causal inference or reproducibility).

## **Ethics:**

I am committed to making all the output from the PhD freely available and as reproducible as possible. This involves exclusively using scripted open-sourced software for analysis, making data used open whenever possible. I also aim to pre-register any analysis I will be undertaking (including the systematic review) on my Open Science Framework page.

## **Bibliography:**

Bird, T.J., Bates, A.E., Lefcheck, J.S., Hill, N.A., Thomson, R.J., Edgar, G.J., Stuart-Smith, R.D., Wotherpoon, S., Krkosek, M., Stuart-Smith, J.F. and Pecl, G.T. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173, 144-154.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., & Hines, J. E. (2017). *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Elsevier.

Wijewardhana, U. A., Meyer, D., & Jayawardana, M. (2020). Statistical models for the persistence of threatened birds using citizen science data: A systematic review. *Global Ecology and Conservation*, 21, e00821.

Appendix:

PhD plan (so far!)

