

Shifting baselines in conservation: developing spatial methods and analysis pipelines in R

Jake Curry

December 6, 2018

Supervisor: Dr Natalie Cooper
Natural History Museum
natalie.cooper@nhm.ac.uk

Keywords: GIS, R, Conservation, Biodiversity patterns, Statistical Computing, Reproducible Techniques

Introduction

Global biodiversity is under threat, with the current era being termed the sixth mass extinction [1]. Habitat loss is said to be the leading cause of species extinction[2], with climate change compounding the issue [5]. Whilst we have current species range data, we do not know if this is an accurate baseline for conservation efforts. Species ranges may have shifted greatly before we started cataloging species distributions with the aim of conservation. This is problematic as it means we are possibly using incorrect baselines for determining the effects humans have had on species distributions, which has profound consequences for conservation efforts [3]. For example, policy makers require baselines to be established to determine the effectiveness of conservation strategies, if we are measuring against the wrong baseline we will naturally draw incorrect conclusions as to the effectiveness of a strategy [8][7]. The data for historical species ranges exists as specimen metadata in museum collections, such as at the Natural History Museum, London (NHM). This historic data is also of particular interest as it has been identified as an area of 'key deficiency' by the Royal Society[4]. This project therefore aims to address a key gap in conservation literature, and address practical concerns of how we routinely incorporate the wealth of data contained within natural history museum archives into current conservation science. Researchers at NHM have already carried out preliminary analyses of these data, but would like a more computationally efficient and reproducible protocol. My aim is to improve upon the existing protocol whilst also developing new spatial analysis tools within R to deal with these issues.

Questions Proposed

In this project I will answer the following questions:

How do you build an effective workflow incorporating version control of large datasets?

What is the 'best' method for computing error in locality metadata?

How much error is inherent in natural history museum type specimen metadata?

How much of an impact has the Anthropocene had on species range change? i.e. How much do historic specimen localities differ from current IUCN Red List range maps, within a case study clade?

Methods

Pipelines will be developed using R, as well as potentially Bash, Python and C. The data sets involved will be large (covering many thousands of specimens), R will be used to handle these data sets. The spatial analysis will be carried out in R using both pre-built packages (e.g. sf) and functions that will be developed specifically for this project. Graphics will be produced in R. A flow chart of the proposed methods and rough workflow is shown in Figure 1. Methods for calculating error will include Maximum Likelihood, Bayesian, the point-radius method [6] and any other methods in the literature that appear relevant.

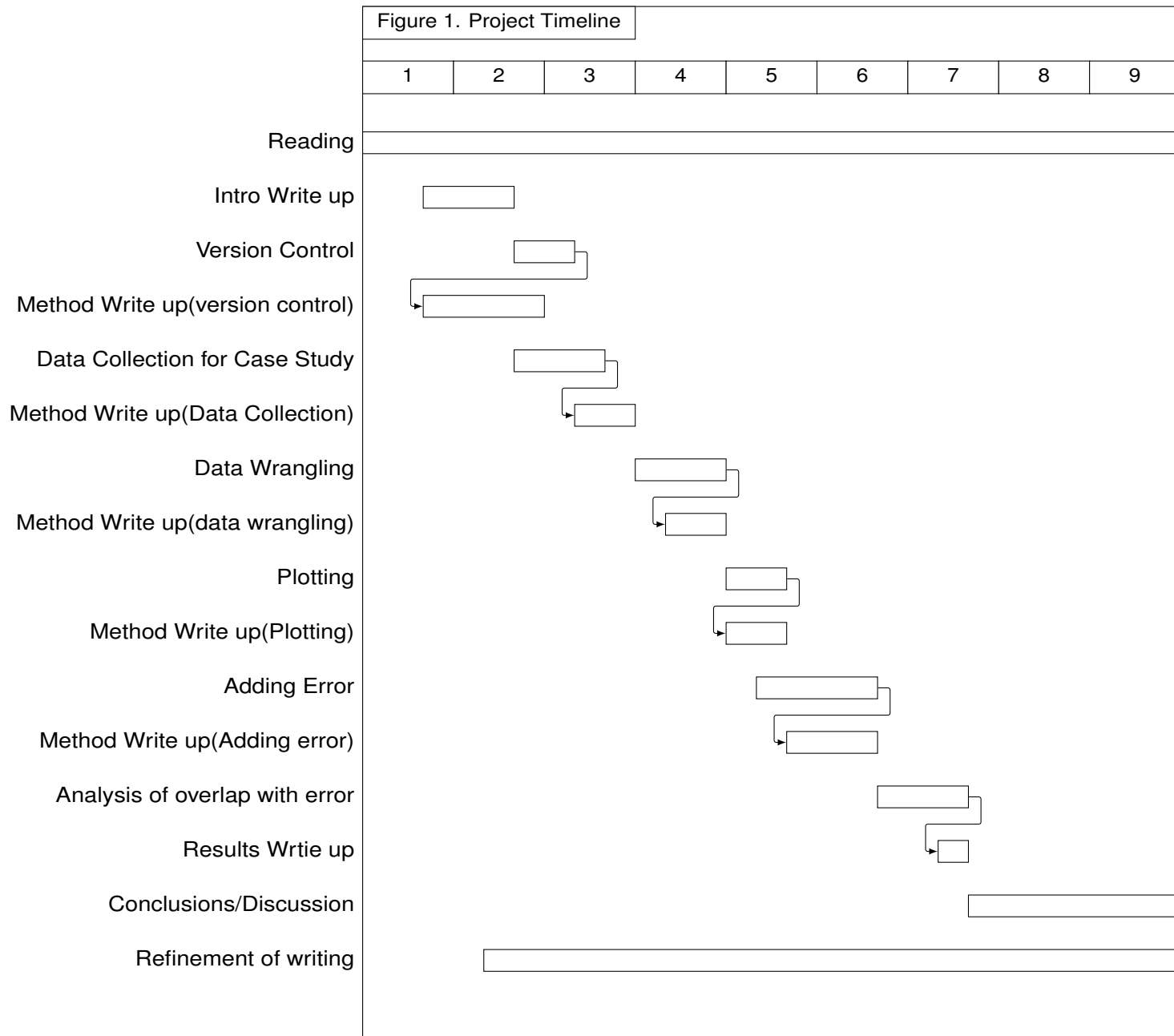
Anticipated outputs and outcomes

1. Development of an analysis pipeline for current and future data collected at the NHM.
2. Version control of the data using datastorr, Git and GitHub.
3. Adding error to already existing point localities/ species range overlap data sets and incorporating this into spatial analyses.
4. Building on the current data sets to include new taxa.
5. A case study establishing how much ranges of a specific vertebrate clade have already shifted since initial type specimen collection. The number of species included will depend on time constraints.

Budget

1. travel - £1000 - Price of return ticket from Sunningdale Station to London Tube zones 1-4 is £15.90, this is a trip I will be taking twice a week minimum. If taking this trip for the duration of the project the cost will be £1208.40.

Timeline



References

- [1] Anthony D. Barnosky, Nicholas Matzke, Susumu Tomiya, Guinevere O. U. Wogan, Brian Swartz, Tiago B. Quental, Charles Marshall, Jenny L. McGuire, Emily L. Lindsey, Kaitlin C. Maguire, Ben Mersey, and Elizabeth A. Ferrer. Has the Earth's sixth mass extinction already arrived? *Nature*, 471(7336):51–57, mar 2011.
- [2] Thomas M. Brooks, Russell A. Mittermeier, Cristina G. Mittermeier, Gustavo A. B. da Fonseca, Anthony B. Rylands, William R. Konstant, Penny Flick, John Pilgrim, Sara Oldfield, Georgina Magin, and Craig Hilton-Taylor. Habitat Loss and Extinction in the Hotspots of Biodiversity. *Conservation Biology*, 16(4):909–923, aug 2002.
- [3] C.A. Froyd and K.J. Willis. Emerging issues in biodiversity & conservation management: The need for a palaeoecological perspective. *Quaternary Science Reviews*, 27(17-18):1723–1732, sep 2008.
- [4] Royal Society Working Party. Measuring biodiversity for conservation. *The Royal Society*, 2003.
- [5] Chris D. Thomas, Alison Cameron, Rhys E. Green, Michel Bakkenes, Linda J. Beaumont, Yvonne C. Collingham, Barend F. N. Erasmus, Marinez Ferreira de Siqueira, Alan Grainger, Lee Hannah, Lesley Hughes, Brian Huntley, Albert S. van Jaarsveld, Guy F. Midgley, Lera Miles, Miguel A. Ortega-Huerta, A. Townsend Peterson, Oliver L. Phillips, and Stephen E. Williams. Extinction risk from climate change. *Nature*, 427(6970):145–148, jan 2004.
- [6] John Wiecezorek, Qinghua Guo, and Robert Hijmans. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18(8):745–767, dec 2004.
- [7] K. Willis and H. Birks. What Is Natural? The Need for a Long-Term Perspective. *Science*, 314(5803):1261–1266, 2006.
- [8] Katherine J Willis, Lindsey Gillson, Terry M Brncic, and Blanca L Figueroa-Rangel. Providing baselines for biodiversity measurement. *Trends in ecology & evolution*, 20(3):107–8, mar 2005.