

Project Proposal

Anne Marie Saunders

Imperial College London

MSc Computational Methods in Ecology and Evolution

April 3rd, 2020

Keywords: vector abundance, time series analysis, interpolation, model selection, vector borne disease, mosquitos

1 Introduction

1. After what temporal lag are mosquito abundance dynamics dependent on past temperature and precipitation?
2. What effect does interpolation between available abundance measurements have on the best-fit temporal lag of temperature and precipitation?
3. - subject to change - How well does the best-fit model predict mosquito abundances based on time series of temperature and precipitation?

2 Proposed Methods

2.1 Temporal Lags

Trap count data will be obtained from the VectDyn database and aggregated to each location by week. Temperature and precipitation data will be obtained from the National Oceanic and Atmospheric Administration Climate Prediction Center and mapped to each trap count location.

I will assess three GLM models of mosquito abundance- as a function of temperature, precipitation, and both temperature and precipitation- using Akaike's Information Criterion (AIC) to determine which meteorological variables most commonly best describe the patterns of mosquito abundance for the datasets. I will then utilize AIC to determine the best-fit temporal lag of each meteorological variable by assessing ten separate univariate models of both temperature and precipitation lagged between zero and four weeks. The best-fit lag for each variable will be incorporated into a bivariate GLM describing mosquito abundance as a function of time-lagged temperature and precipitation.

2.2 Interpolation

I will then use three methods of interpolation- forward filling, backward filling, and mean filling- to interpolate count data to daily intervals. Based on these interpolated abundances, I will find the best-fit temporal lag of both temperature and precipitation for intervals between zero and twenty-eight days for each of the three interpolation techniques. I will use AIC to assess the best-fit temporal lag for temperature and precipitation for each interpolation method. These will be combined to produce three bivariate GLM models whose AICs, in combination with the AIC of the non-interpolated model, can be compared.

2.3 Cross-Validation

I would like to do a 3rd question if there is time but I'm still figuring out exactly what I would want this to be. Right now I would plan to do a leave-one-out cross validation procedure where I train the best-fit model on a collection of time series in each location and test the resultant model on a sample time series from a year that was not used to train the model. I can then assess the performance of the prediction model against the observed data. I could also compare the performance of interpolated versus non-interpolated best-fit models in predicting population dynamics.

3 Anticipated Outputs and Outcomes

- best-fit time lags for meteorological variables from both interpolated and non-interpolated data sets
- Perhaps an understanding of which datasets work well with interpolation versus datasets with better fit from non-interpolated data
- An understanding of how interpolation affects the quality of predictions made from the best-fit model

4 Project Feasibility and Timeline

5 Budget

Category	Item	Cost
High Performace Computing	Computing time to fit interpolated and non-interpolated models to each dataset	??
Lab Equipment	16 GB RAM to help me better handle very large dataset	£120
	Laptop screws and screwdriver to help install RAM and because I have a couple of screws loose	£10