Preprint template

M.D. Catchen [1,2]

[1] McGill University [2] Québec Centre for Biodiversity Sciences

Correspondance to:

Last revision: *July 17, 2021*

Abstract: This template is used by the Poisot lab at Université de Montréal to write manuscripts using github. It uses github actions as a way to generate a website that can be annotated using hypothes.is, a PDF document for copy-editing and submission to journals, and a PDF document for submission to preprint servers. At every push on the master branch, the whole series of documents will be updated automatically.

Introduction

Being able to predict how Earth's ecosystems will change in the future—ecological forecasting—is increasingly an imperative to mitigate climate and land-use change (**cite?**). The fundamental problem of forecasting is to predict how some system will change in the future given a time-series of of observations of this system in the past (**Strydom2021PreNet?**). When stated at this level of abstraction, it is clear that there is widespread potential application of the forecasting problem. Consequently many tools have been developed for time-series analysis across a variety of fields. It would serve ecologists well to have sense how well these models perform in ecological systems, in order to decide on the proper tools for forecasting a particular system at a particular scale.

In the machine learning literature, one form of model that has proven successful in predicting temporal/sequential data are Recurrent Neural Networks (RNNs). There are various forms of RNNs: classic RNN, LTSM, GRU, RNN-Turing Machine, … etc.

In this paper we use a LTER dataset of freshwater fish from (**cite?**) which describes the occurrence of 14 fish species across 10 Wisconsin lakes over the span of 20 years. To simulate the experience of one attempting to forecast the dynamics of this system, we fit multiple models to the data, one year at a time By doing this, we demonstrate that an RNN forecasts more effectively than (competing *boring* models)…

Methods

Data

We use data from (**datacite?**), an LTER project. The data describes the occurrence of 14 species (species list), across 10 lakes (lakes list). The data was collected across 20 years. Sometimes there are more than one data point per year. We consider any occurence within a year as a 1 for that year.

RNN setup

A recurrent neural network (RNN) is a type of neural network which takes a vector of inputs $x = [x_1, x_2, x_{\ldots,x_n}]$, and each particular input x_i is weighted by a hidden value from the previous input in the sequence, x_{i-1} .

Specifics to this network training. Final network setup: GRU, Dense, Sigma, etc… Done in Flux v12 in Julia v1.6.

Results

[Figure 1 about here.]

Discussion

References

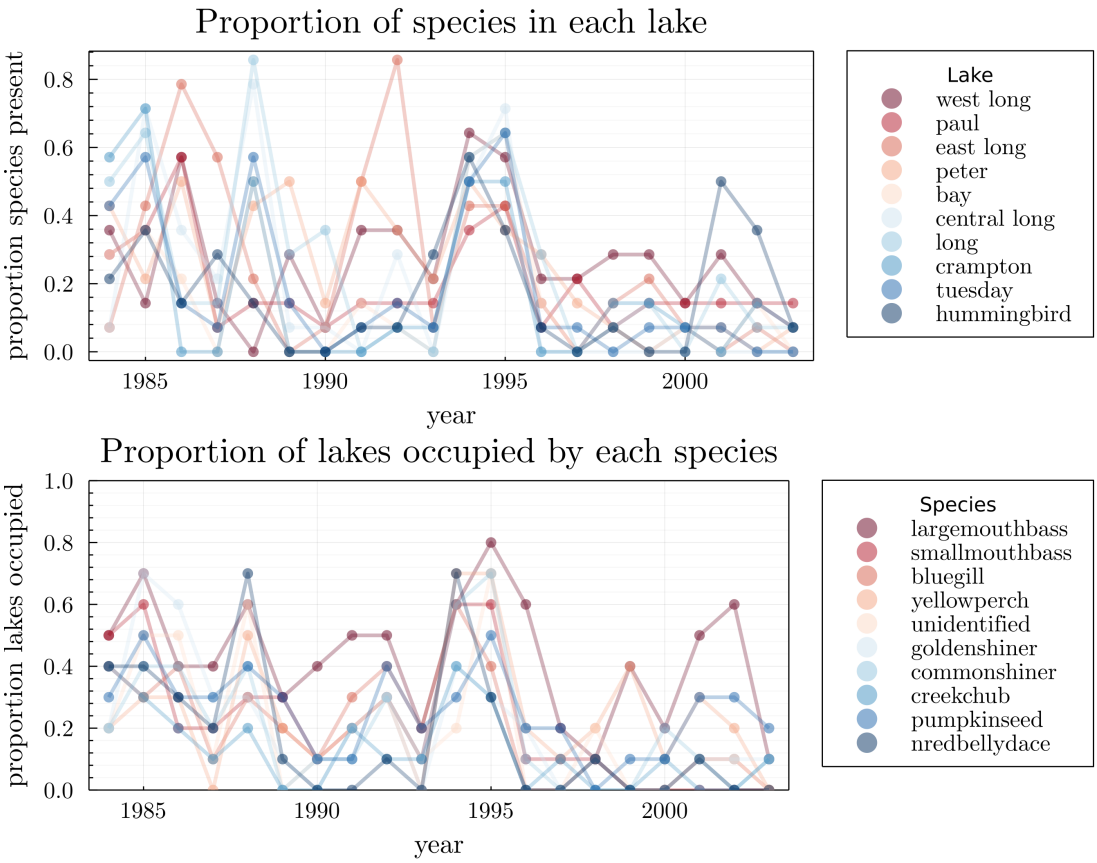

Figure 1: TODO: occupancy data. 1994? whats up with that?