Capturing Transients

An application of Biostatistics to Astronomy

van Dyk A [1,2] , McBride VA [1,2] , Groot PJ [1,2,3]

[1] South African Astronomical Observatory
[2] Department of Astronomy, University of Cape Town
[3] Department of Astrophysics, Radboud University

Email: anke@saao.ac.za

Introduction

- The aim is to estimate the underlying population size of transients and/or variable stars using the biostatistical CAPTURE-RECAPTURE method.
- *Recurring* transients or variables can be classified as a statistical **closed population.**
- The population remains constant throughout the duration of sampling (observation).

A simple case scenario is the Schnabel Estimator with equal probability of capturing an individual (Eq 4, Laycock 2017):

$$Ns_i = \frac{\sum N_i \times Nc_{(i-1)}}{(\sum Nr_i) + 1}$$

where Ns_i is the estimated quantity at observation i, N_i is the transient count observation i, $Nc_{(i-1)}$ is the cumulative count at observation $i-1$ and Nr_i is the count of individuals re-encountered at observation i.

Simulating a population

- Model a period distribution
- Simulate a population of Be/X-ray binaries, assuming outburst at periastron
- Randomly sample the population at a specified cadence
- Apply a threshold and make note of a detection/non-detection

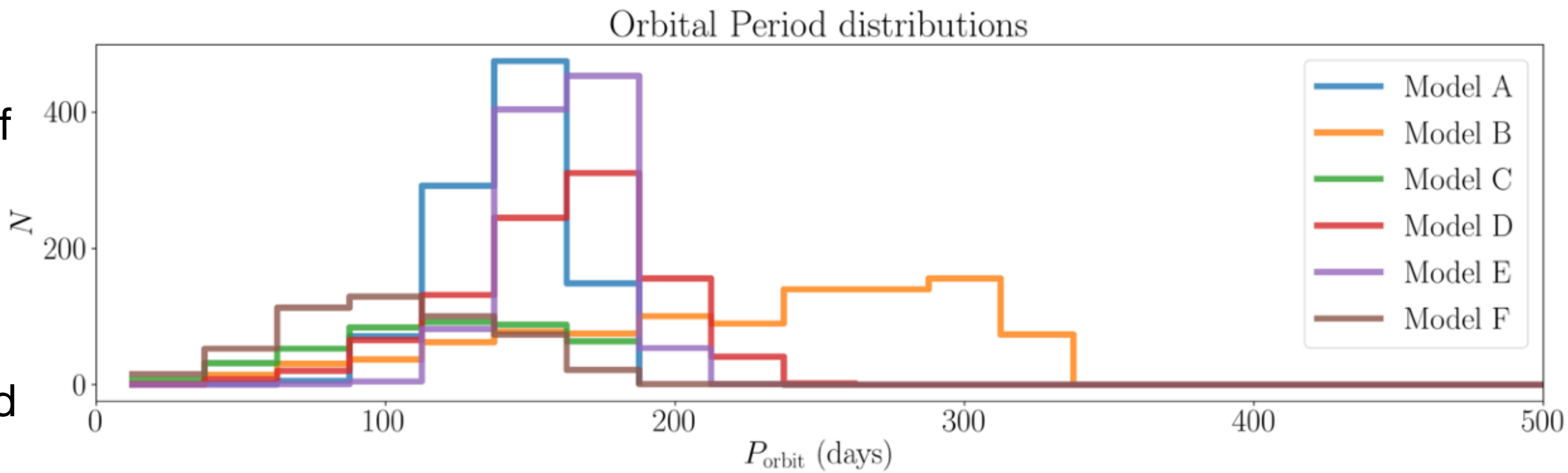


Figure 1: *The different orbital period distribution models represent possible underlying distributions of Be/X-ray binaries. Random samples are drawn from these models to simulate a population of Be/X-ray binaries.*

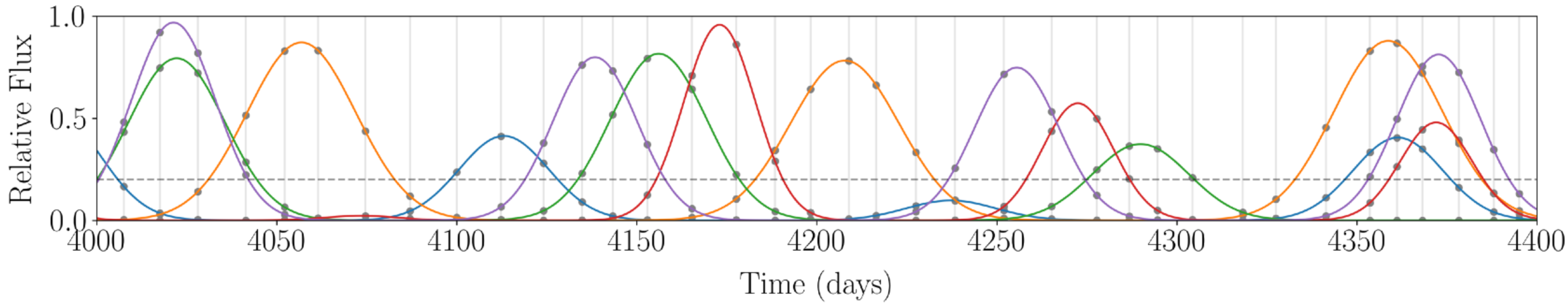


Figure 2: *The figure depicts the simulated lightcurves of small population of Model A Be/X-ray binaries with randomly scaled amplitudes, representing the outburst at periastron. The population in is concurrently but randomly sampled at a 7-14 day cadence that represents a typical survey recurrence observation time. An arbitrary threshold at 0.2 discriminates between a detection and a non-detection.*

Estimating the underlying population

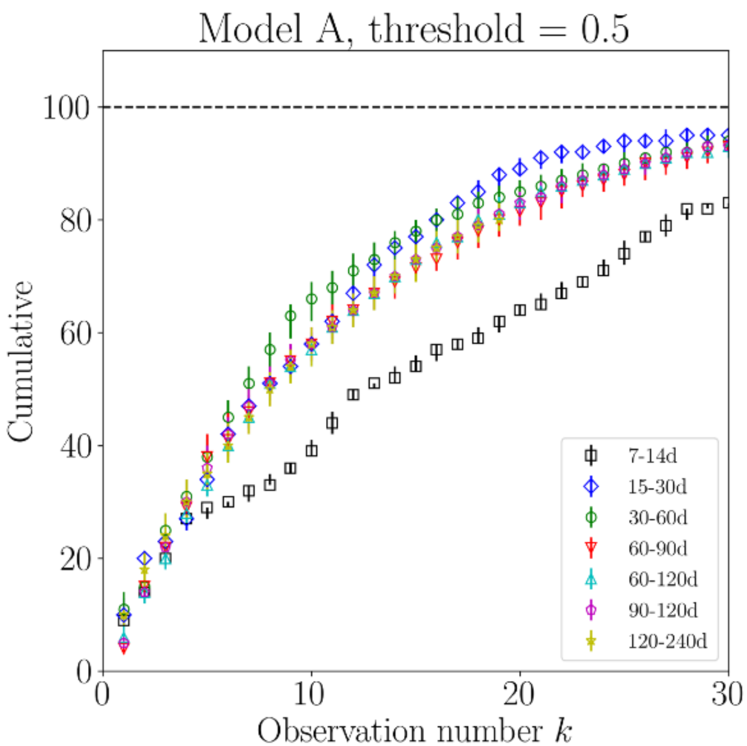
Information is stored in a capture (or encounter) history for each identifier at each observation.
There are various estimator types with different assumptions on capture probability, temporal probability etc.
Logistic regression is employed to estimate the population size given representative sampling at each occasion.

Table 1: *An example capture history*

	ID#1	ID#2	ID#3	ID#4	ID#5	ID#6
t=1	1	1	0	0	1	0
t=2	1	1	1	1	0	0
t=3	0	0	1	1	0	1

Parameters that are built into the simulation and estimation include:

- Amplitude and duration of outburst
- Brightness detection threshold



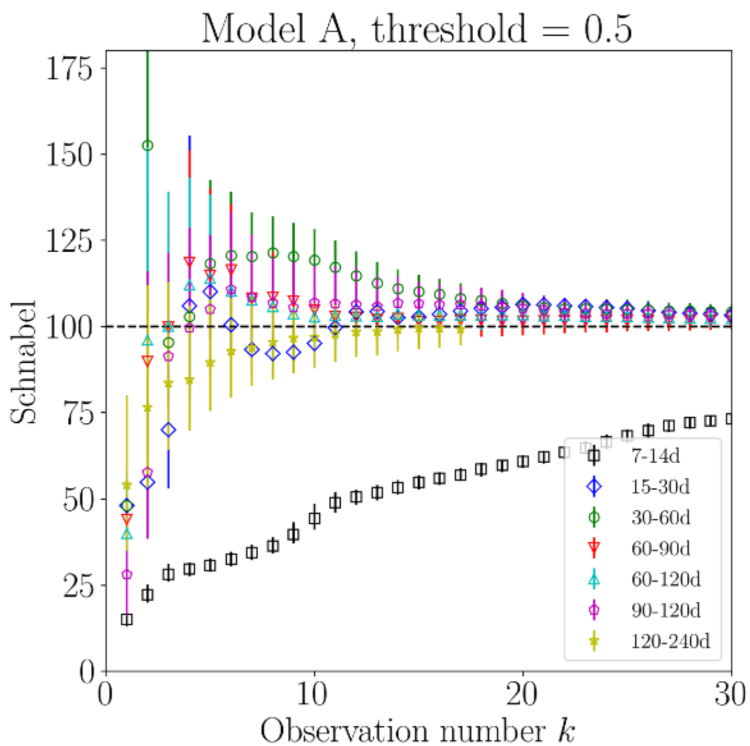


Figure 3: *Both plots show the count/estimator as a function of observation for an underlying population of N=100 at seven different cadences. LEFT: A cumulative count. RIGHT: The Schnabel estimation. For Model A with a median orbital period of ~150 days, we reach the underlying population to within 20% in < 15 observations for all but the 7 to 14 day cadence.*

Next steps

- The variation of estimators needs to be quantified w.r.t. parameters such as magnitude, amplitude and duration of outbursts, and threshold limitations.
- We are currently investigating estimators that allow for heterogeneous probabilities of transient capture and the effect on the accuracy of estimation.
- We are busy developing a methodology and science case for real astronomical data and potential other applications, such as Dwarf Novae Cataclysmic Variables and FRB populations.

References

Laycock, S.T. 2017. From blackbirds to black holes: Investigating capture-recapture methods for time domain astronomy. *New Astronomy.* 54:91-102. Available: http://dx.doi.org/10.1016/j.newast.2017.01.003.

Schnabel, Z.E. 1938. The Estimation of Total Fish Population of a Lake. *The American Mathematical Monthly.* 45(6):348-352. Available: https://www.jstor.org/stable/2304025.