Bat virus underroost shedding model

Benny Borremans

Contents

bennyborremans@bbresearch.org

Last update: 08 Jun 2022.

The table of contents can be clicked to jump straight to specific sections.

Goal = create a model of underroost bat virus shedding.

Sheets placed below roosts collect urine from an estimated number of bats.
Urine samples on the sheet are pooled, and tested for RNA concentration using RT-PCR.
Total sample volume depend on the number and volume of urine droplets on the sheets.
The number and species of bats above the sheet are estimated, but not all bats can always be observed, and bats can move after/before observation.
Samples are stored in one of two buffers (AVL/VTM), or without buffer (NB).
Buffer type affects PCR sensitivity.

The end result (Ct value in a pooled sample) depends on all these factors,
which makes it difficult to estimate shedding prevalence in the population.

The goal of this project is two-fold:
(1) Create a simulation model of the different processes that are believed to be involved, to get better insights and build intuition.
(2) Create a model to estimate shedding prevalence in the population from the available data, capturing as much of the observation process as reasonably possible.

Simulation model

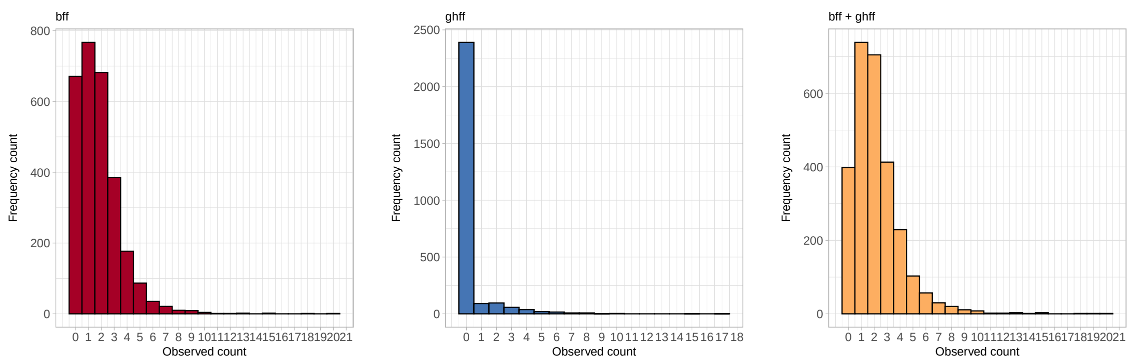
Model number of bats above a sheet

Observed counts and species

There are counts for 2 species, black flying foxes and grey-headed flying foxes.
Counts can be morning, afternoon, and/or overall.
Observations have a level of confidence in the count and/or species.
These confidence levels are available for bff only (do they cover both bff and ghff, or bff only?).

Using only observations with a high level of confidence.
Using only morning observations (as these seem to be the most common).
Removing observations that are not exact (e.g. “5+”).
All sites pooled.

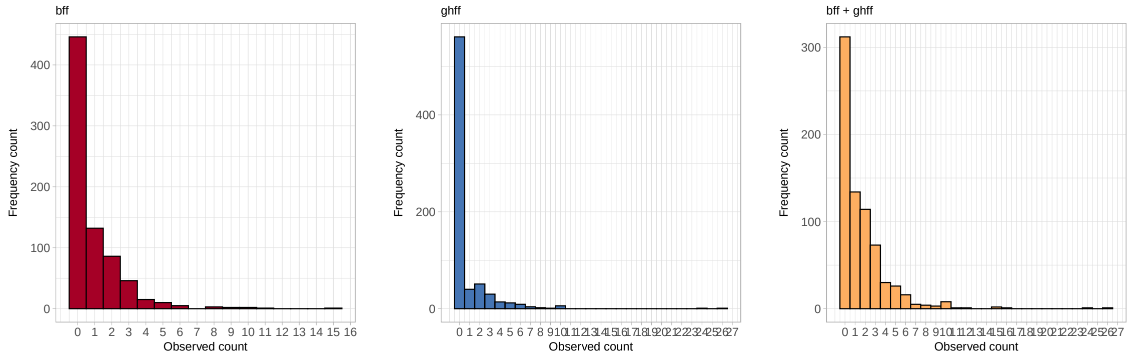
Histograms of counts:



Are higher numbers actually more rare, or just more uncertain?

==> check lower confidence counts.

Histograms of counts:



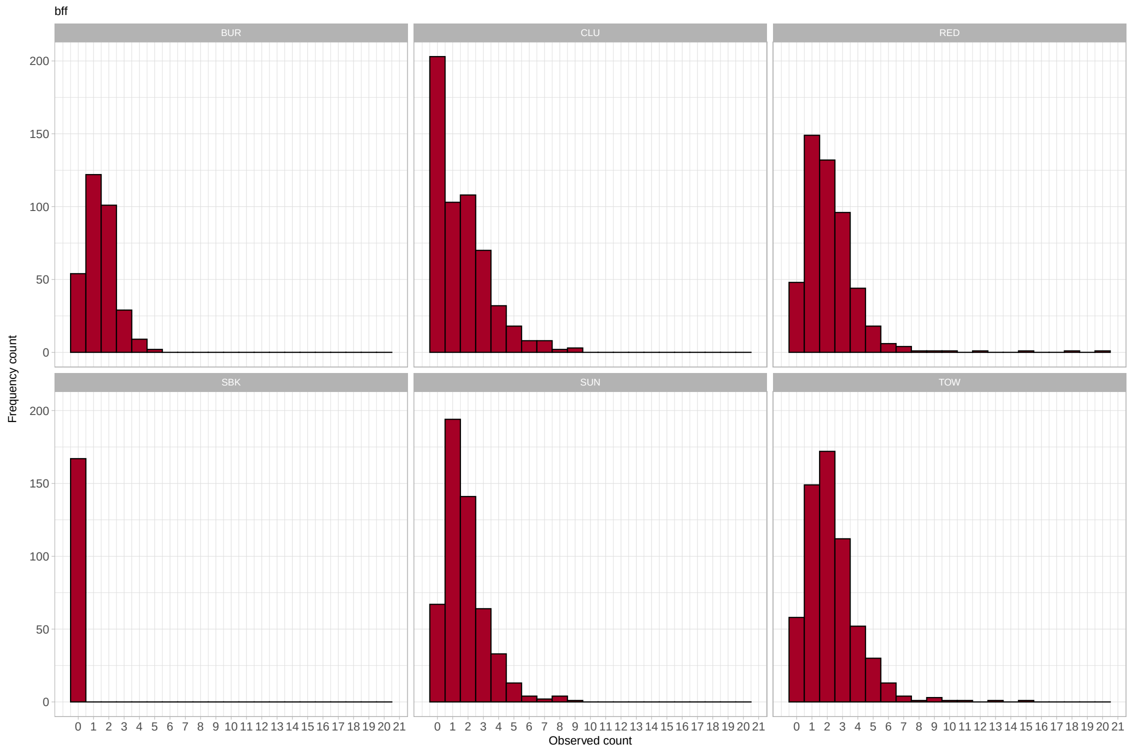
Lower-confidence counts are not higher, except for 2 ghff counts.

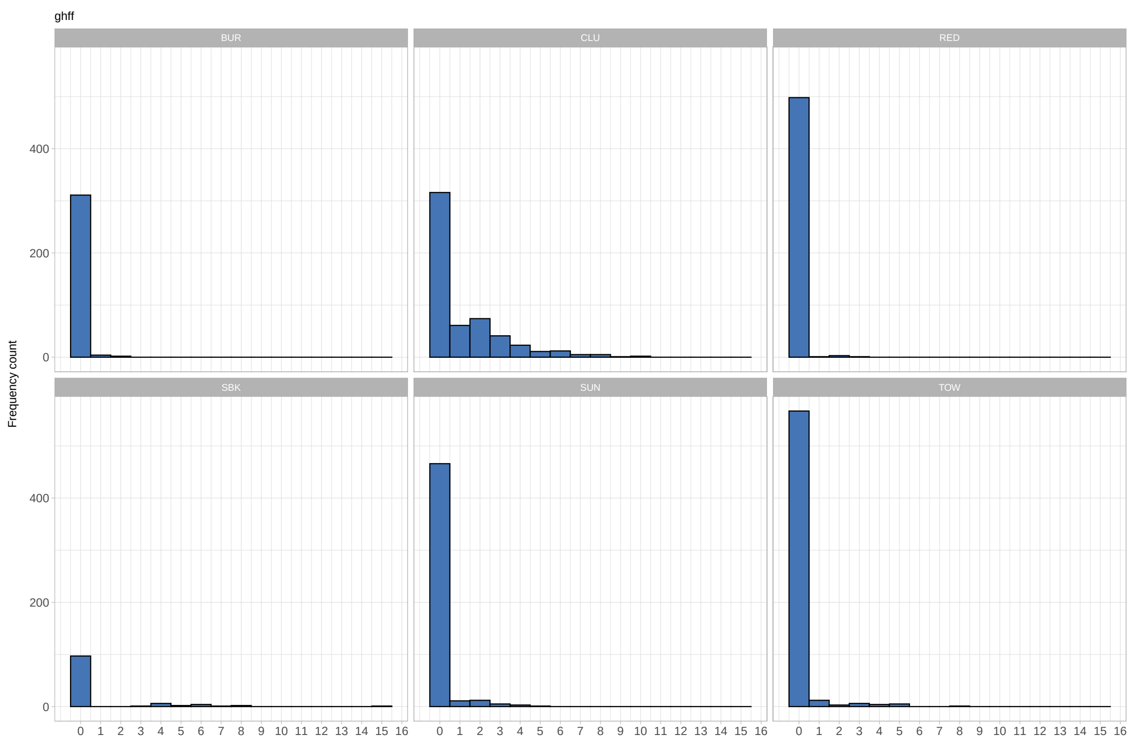
Are higher numbers actually more rare, or just written down as N+?

Not likely, there are only 25 entries with a + sign,
and these are one of: 5+, 10+, 3+, 1+.

Are there differences between sites?

Using only sites with more than 100 observations.

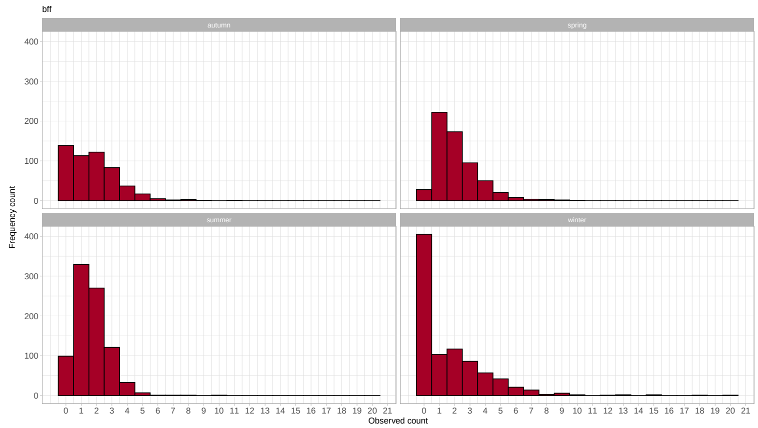




==> all look very similar, except many more 0 counts at CLU.
Any reason for this? Different conditions?
I didn't find anything in the notes, and the "bats" column mostly says "stable".

Any difference between seasons?

Histograms for different seasons.
bff only, most data available, don't need figure overload.



There seems to be an effect of season, that probably should be taken into account.

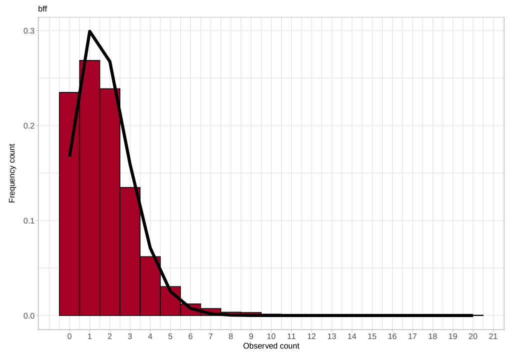
While spring and summer look like Poisson distributions,
autumn and winter seem closer to a mixture of a Bernoulli and a Poisson (probability of seeing any bats + if there are bats, how many).

The distributions look like Poisson, which is convenient.
==> fit Poisson distributions.

All bff data pooled, across seasons and sites:

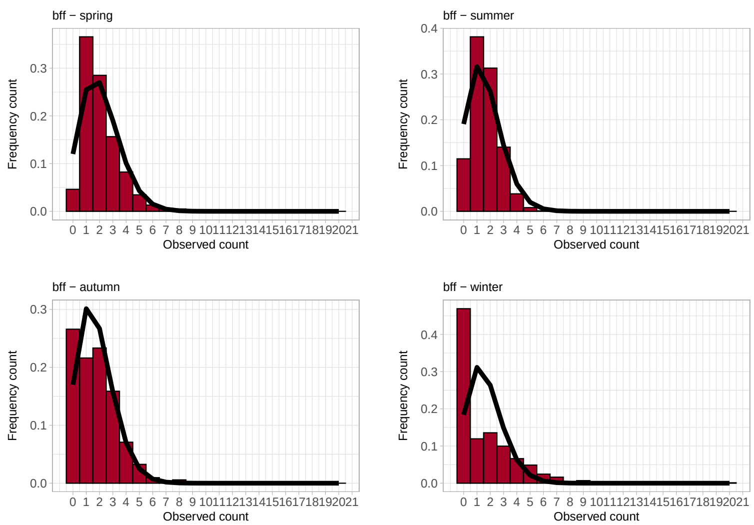
Lambda = 1.7871148.

Fitted distribution:

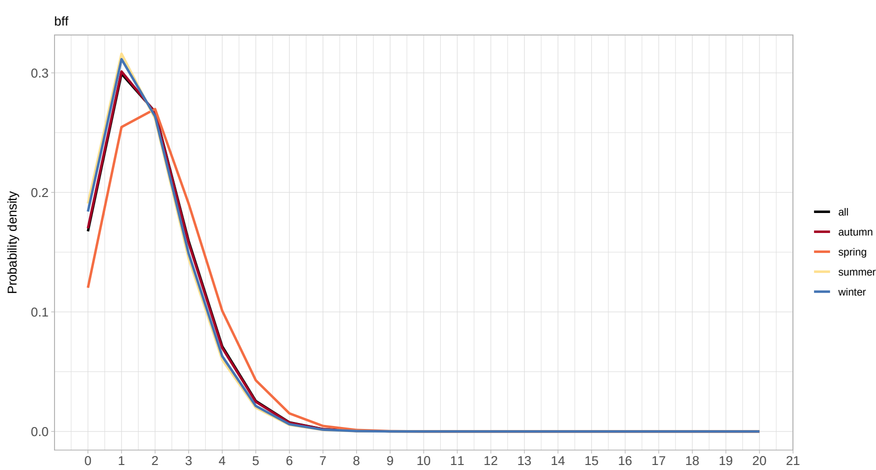


All bff data pooled, per season, across all sites:

Fitted distributions:



All combined:



==> when using only a Poisson distribution, they are very similar across seasons.
While it is clear that in autumn and winter there seems to be a mixture of distributions, for the purposes of the simulation we can probably keep it simple and stick to a Poisson.

Conclusion:
Bat count can be modeled using a Poisson distribution, with:
$N_{bats} \sim Poisson(1.8)$

Correlation counts and sample volume

Choose adequate distribution for generating bat counts

Prevalence estimation model