Parity results

Ben Lambert

August 21, 2019

1 Grouping level

See `grouping_holdouts.xlsx` for fits of species-level versus complex-level vs continent-level and for continent-level vs overall-level. For the former, the only interesting case is for the Americas, where the species-level model fits better than the complex level grouping - driven by *A. albitarsis* which has species with strongly varying lifespan estimates. For the continent vs overall fits, Africa was better fit by a model grouping all species together than one allowing continental differences in lifespan; for Asia, the opposite trend was found, with the data being better fit by a model allowing continental differences in lifespan.

2 Effect of temperature on lifespan

In all cases, see Table X (from '`temperature_holdout.xlsx`) for a summary.

2.1 Mean temperature

The overall hold-out log-likelihood was: 20,400.3 for the model without temperature terms; -20,164.61 for the model with linear temperature terms; and -19,131.74 for the model with up to quadratic temperature terms. The p values comparing to the model without temperature terms were: for the linear model, $p = 0.18(n = 724, z = 0.91)$; and for the quadratic model, $p < 0.01(n = 724, z = 3.64)$. The quadratic fit, however, is almost entirely driven by the result for *A. albitarsis s.l.* (see '`detinova_temperature_fits`'), which had significantly higher lifespan at both low or high mean temperatures. When this species is removed from the results, the difference versus the model without temperature terms is no longer significant $(p = 0.38(n = 666, z = 0.30)$.

2.2 Daily (nightly) temperature range

This is the model where we compare min-max for temperature extracted at midnight each day. The overall hold-out log-likelihood was: 20,400.3 for the model without temperature terms; -19,222.3 $(p = 0.01, n = 724, z = 2.33)$ for the model with linear temperature terms; and -18,741.22 $(p < 0.01, n =$

$724, z = 3.60$) for the model with up to quadratic temperature terms. This result was driven by the fit of *A. gambiae s.l.*, where our estimation determined that lifespan increased with daily temperature range.

3 Anopheles gambiae investigations

3.1 Malaria prevalence vs parity

There were only $n = 142$ studies between 2000-2015 (when MAP estimates are available). For these cases, I've plotted 2-10 prevalence versus parity and find no correlation (see `detinova_gambiae_pfpr.pdf`).

3.2 Country-level estimates

I have estimated country-level lifespans for the $n = 541$ data points where GPS location was available. The lifespans are plotted in `detinova_lifespan_gambiae_country.pdf`, which determines that Senegal has, on average, the longest lifespan. K-fold CV, however, indicated that these country-level differences were not significant, since a model grouping at the continent level performed better on out-of-sample prediction. This better performance was almost uniformly observed across the countries in the dataset.

3.3 With or without Reunion

On inspection of the plots of lifespan against temperature range (see `detinova_gambiae_daynight_range_l` it looked as if the results may hinge on the data from Reunion. I, hence, repeated the analysis with/without this country. With this country included (see `fit_complex_daynightmax.rds`), the models with linear and quadratic temperature terms fitted the data better. Without Reunion, the linear model fitted the data best but the difference in fit was weaker (compare `detinova_gambiae_daynight_range_lifespan_cou` with `detinova_gambiae_daynight_range_lifespan_noreunion_country.pdf`) and no longer significant at the 5% level (`fit_complex_daynightmax_noreunion.rds`).