Validation Report

22-07-2022
Eradication version: ___
emodpy-malaria version: ___

Table of contents

1. Introduction 2
  1.1 Background 2
2. Results summary 3
  2.1 Performance compared to model version from calibration 3
3. Visual comparison of reference data and matched simulations 4
  3.1 Incidence by age 4
  3.2 Prevalence by age 5
  3.3 Infectiousness to vectors 6
  3.4 Duration of infection - all ages 8
  3.5 Duration of infection - by age 9
  3.6 Asexual parasite density by age 10
  3.7 Gametocyte density by age 15
4. Additional comparisons of reference data and matched simulations ... 20
  4.1 Incidence by age 20
  4.2 Prevalence by age 22
5. Comparisons from prior EMOD publications 24
  5.1 Incidence and prevalence by age 24
  5.2 Infectiousness to vectors 25
  5.3 Parasite densities 26

1. Introduction

1.1 Background

The goal of this report is to help users quickly identify whether updated versions of the malaria model are still well-calibrated to capture a range of relevant real-world malaria observations.

The figures and tables compare simulation output generated with a particular version of the Eradication.exe and of emodpy-malaria with 1) the simulation results generated by earlier versions of Eradication.exe and emodpy-malaria (the versions used to calibrate the model) and 2) reference datasets from real-world observations.

This report was generated by running the malaria model validation workflow available at https://github.com/InstituteforDiseaseModeling/malaria-model_validation. Additional information on the reference datasets and on the simulation assumptions are available from the repo in "Notes on reference datasets and simulation assumptions.docx," and instructions on how to re-run the validation comparisons are in the README file.

2. Results summary

2.1 Performance compared to model version from calibration

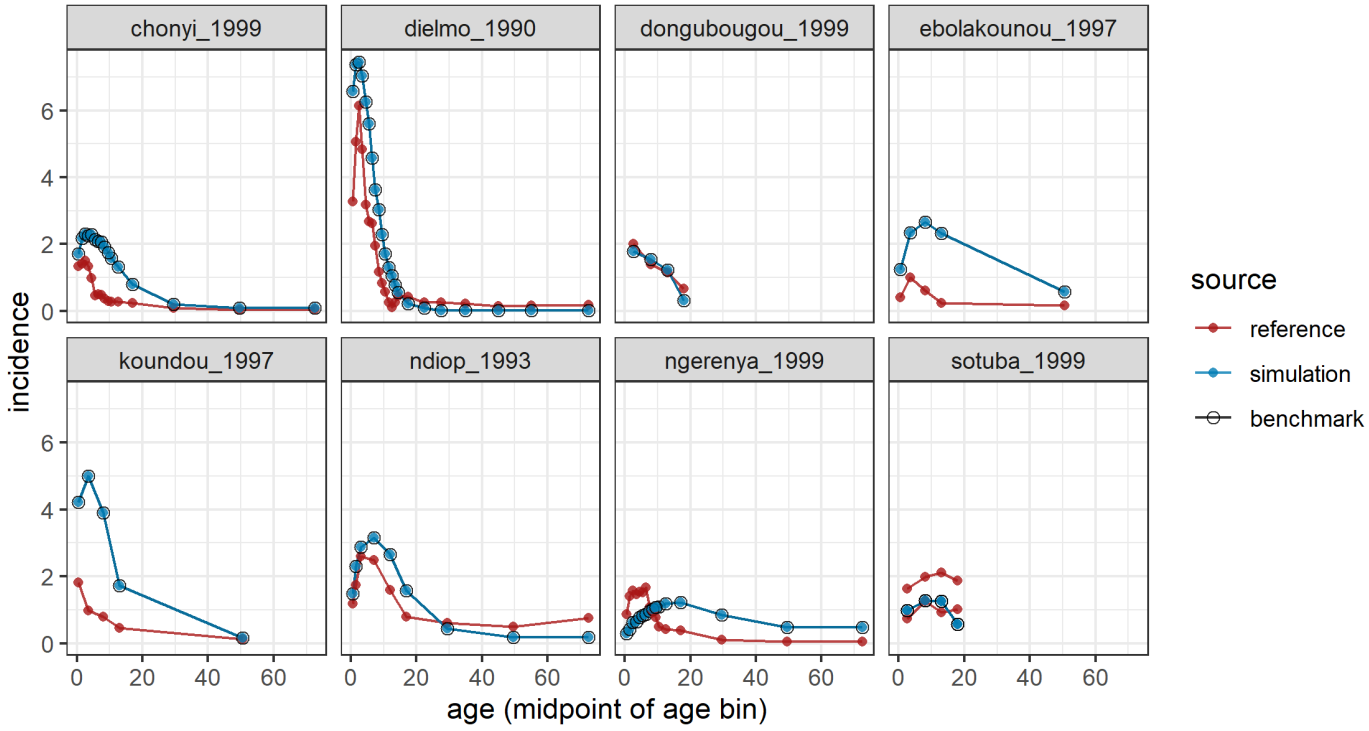
The table below shows, for each validation relationship (rows), the mean absolute difference between all reference and matched simulation datapoints for both the new (first column) and benchmark (second column) simulations. The final three columns of the table show the number of sites where the new simulations matched the reference dataset better, similarly, or worse compared to the benchmark simulations.

validation_relationship	abs_diff_new	abs_diff_bench	num_sites_better	num_sites_similar	num_sites_worse
age_incidence	0.94	0.94	0	8	0
age_prevalence	0.27	0.27	0	17	0
asexual_par_dens	0.19	0.19	0	15	0
gamet_par_dens	0.24	0.24	0	15	0
infectiousness	0.14	0.14	0	6	0

3. Visual comparison of reference data and matched simulations

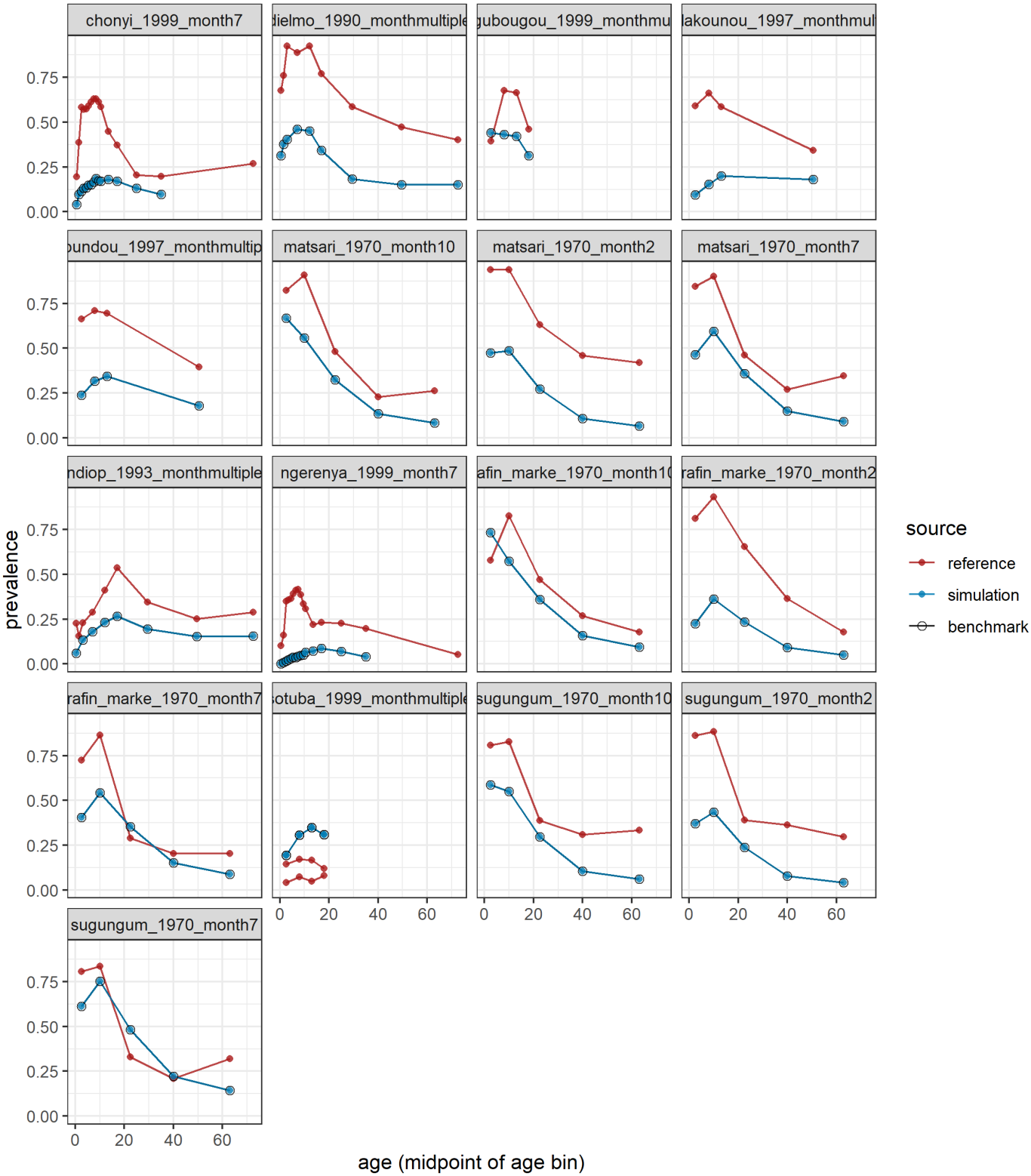
3.1 Incidence by age

The plots below compare the age-incidence relationships from reference datasets and matched simulations.



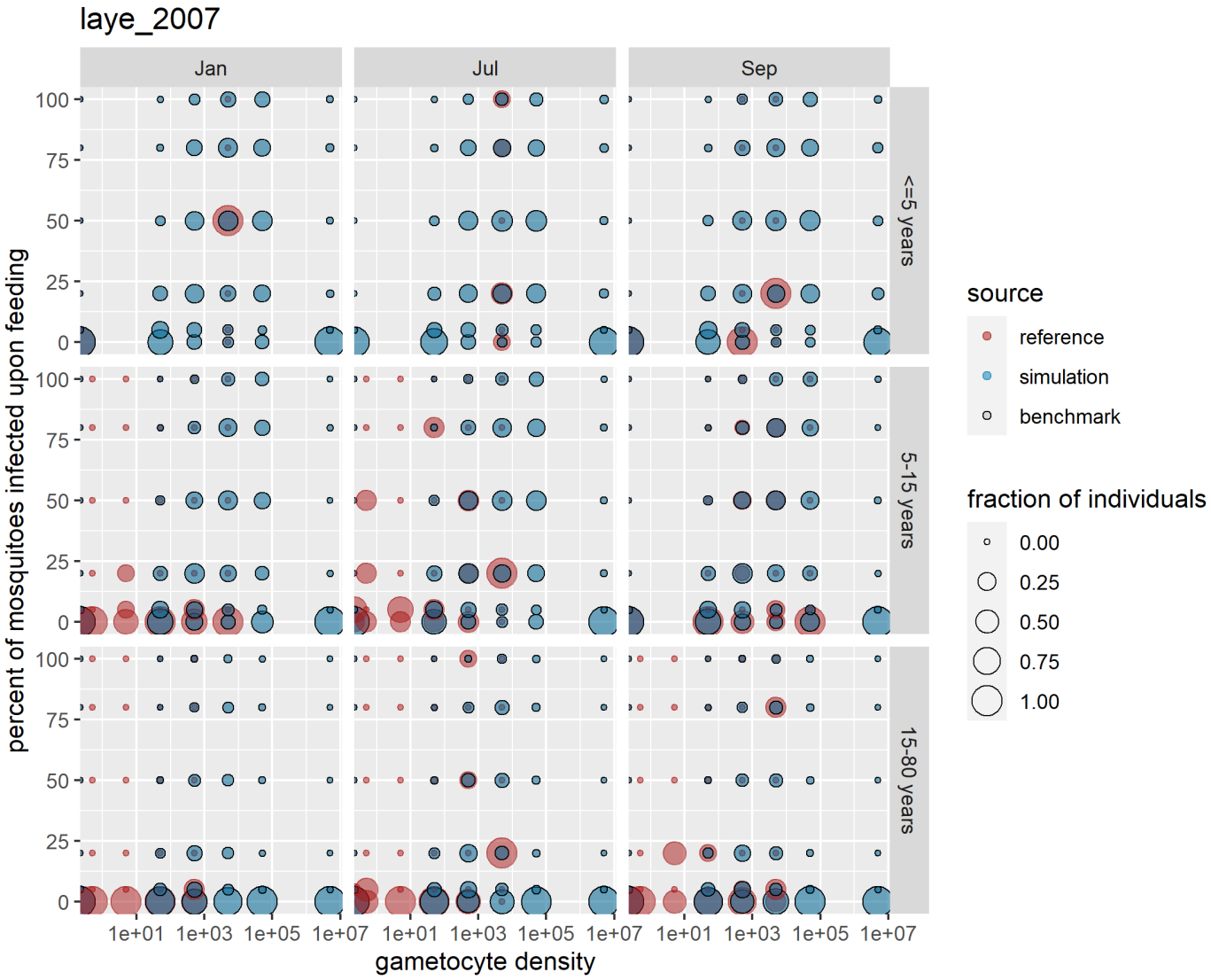
3.2 Prevalence by age

The plots below compare the age-prevalence relationships from reference datasets and matched simulations.



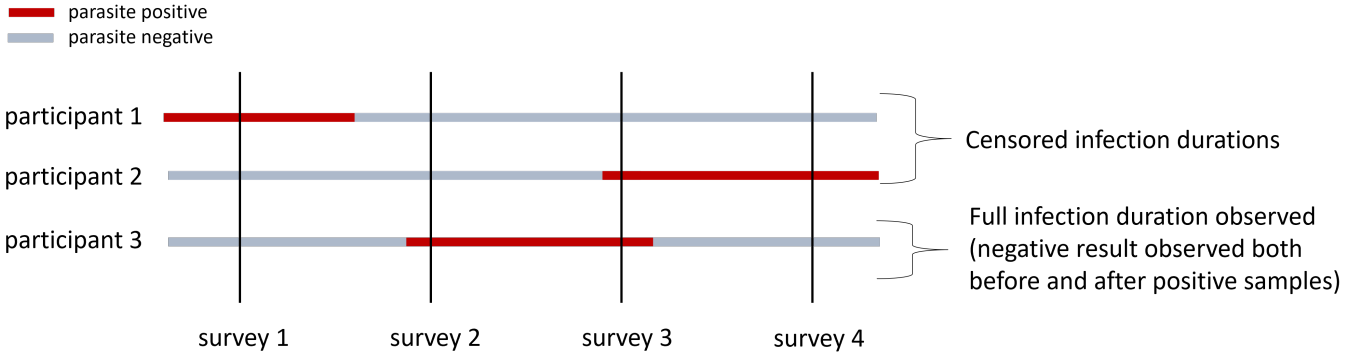
3.3 Infectiousness to vectors

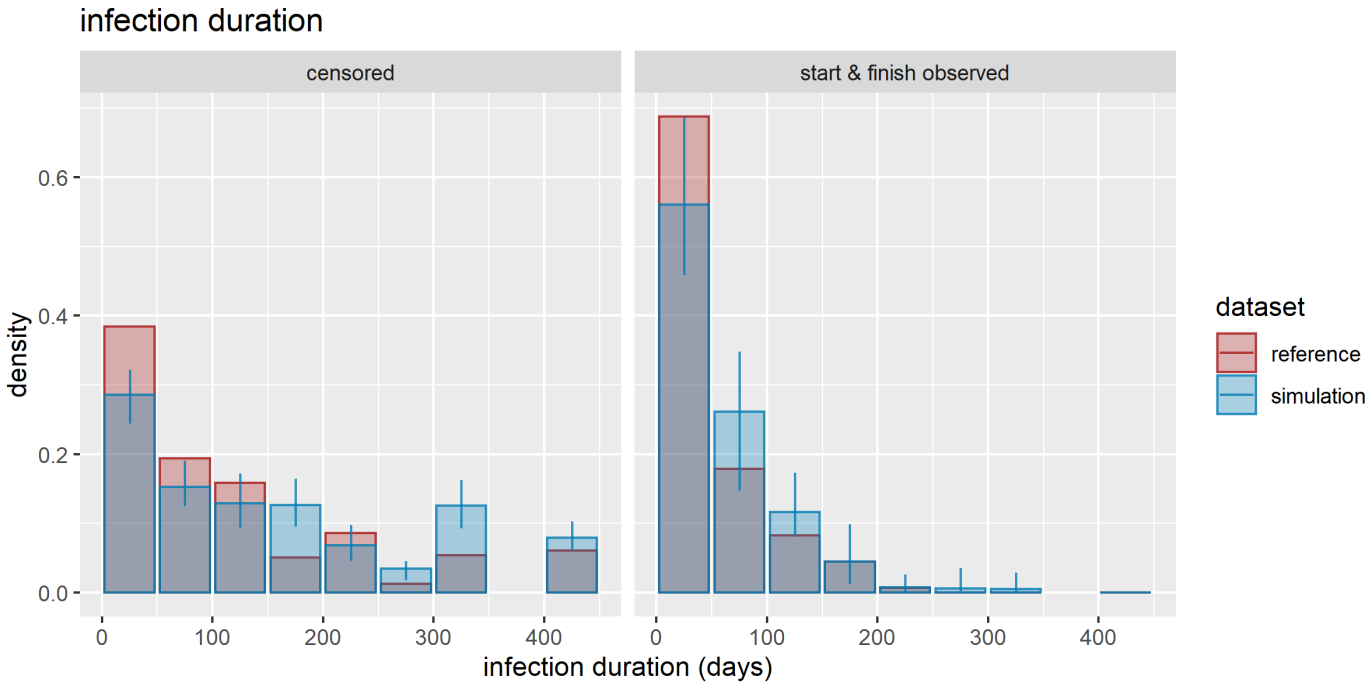
Each of the below plot panels corresponds to a site. Within a plot panel, each row corresponds to an age group and each column corresponds to the month when sampling occurred.
The x-axis shows the gametocyte density in an infection. The y-axis shows how infectious an individual is to mosquitoes. The dot size shows how often a person of a given age and gametocyte density falls into each of the infectiousness bins (each column's dot sizes sum to one).
In the reference datasets, the sample size is sometimes quite small.


laye_2007
Jan
Jul
Sep
<=5 years
5-15 years
15-80 years
percent of mosquitoes infected upon feeding
gametocyte density
1e+01
1e+03
1e+05
1e+07
source
reference
simulation
benchmark
fraction of individuals
0.00
0.25
0.50
0.75
1.00

3.4 Duration of infection - all ages

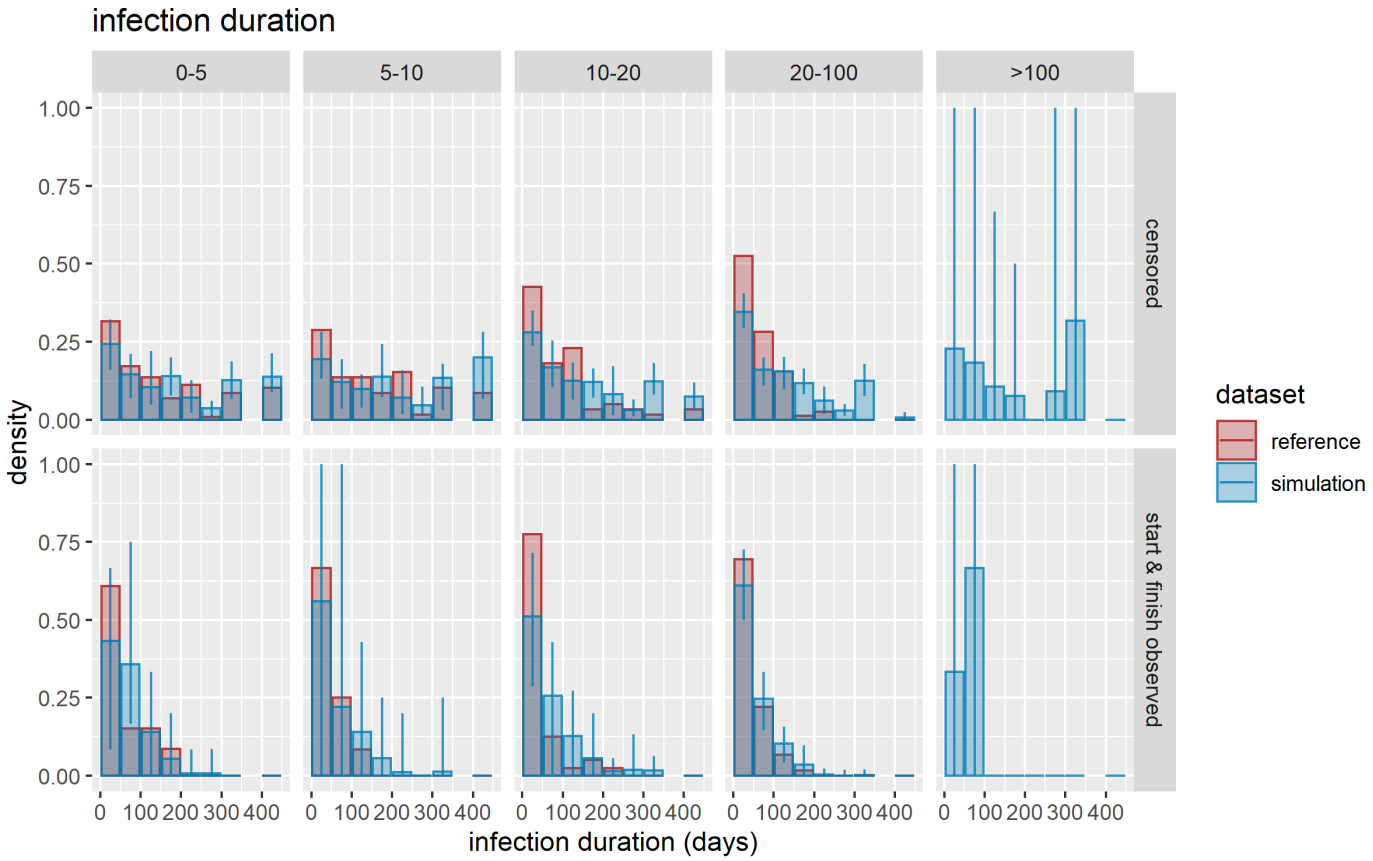
The plots below compare the duration over which individuals had positive tests in the reference dataset and matched simulations. The sampling design from the reference data was matched in the simulations.Observed infections are divided into two groups. "Censored" infections refer to infections where the individual was positive at the first or final survey of the study (so the infection may have extended beyond the period observed). "Start & finish observed" infections refer to infections were the individual was observed to have a negative test at the start and end of the infection. The two types of infection duration records are illustrated in the figure below.





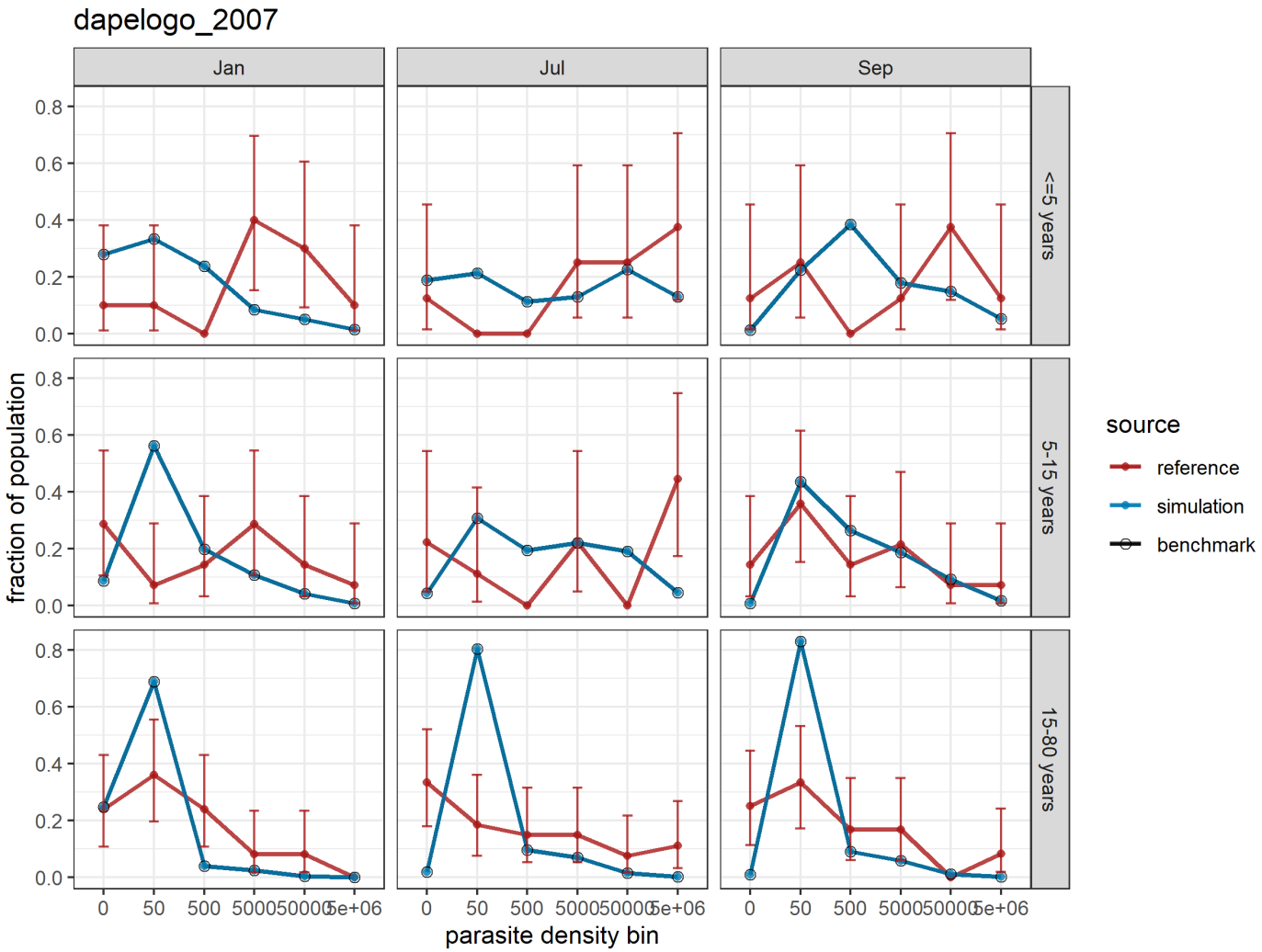
3.5 Duration of infection - by age

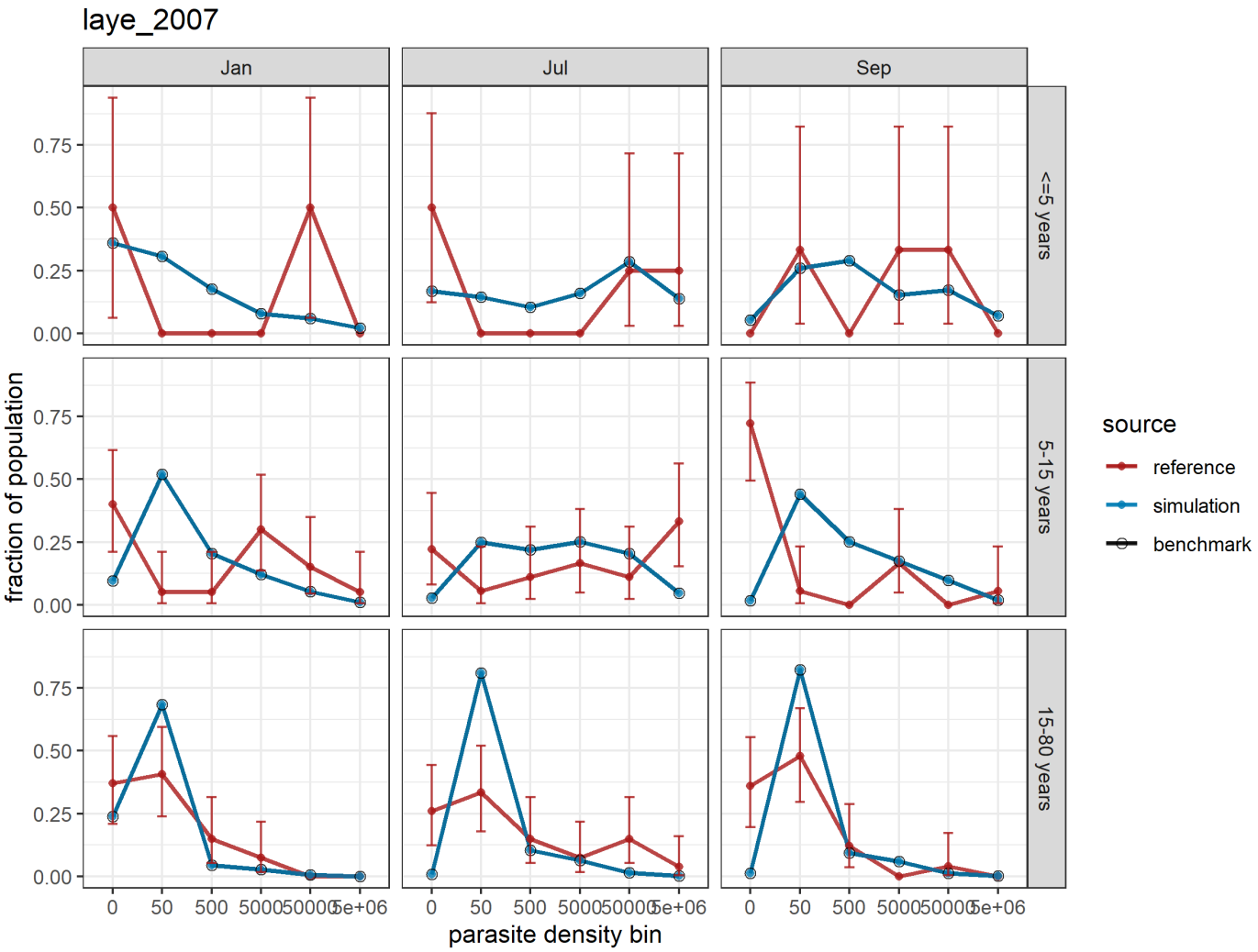
The plots below compare the duration over which individuals had positive tests in the reference dataset and matched simulations. The sampling design from the reference data was matched in the simulations.Observed infections are divided into two groups. "Censored" infections refer to infections where the individual was positive at the first or final survey of the study (so the infection may have extended beyond the period observed). "Start & finish observed" infections refer to infections were the individual was observed to have a negative test at the start and end of the infection. The two types of infection duration records are illustrated in the figure below.
In the plot panel below, columns correspond to the age group (in years) and rows correspond to whether or not the start and end of the infection was observed.

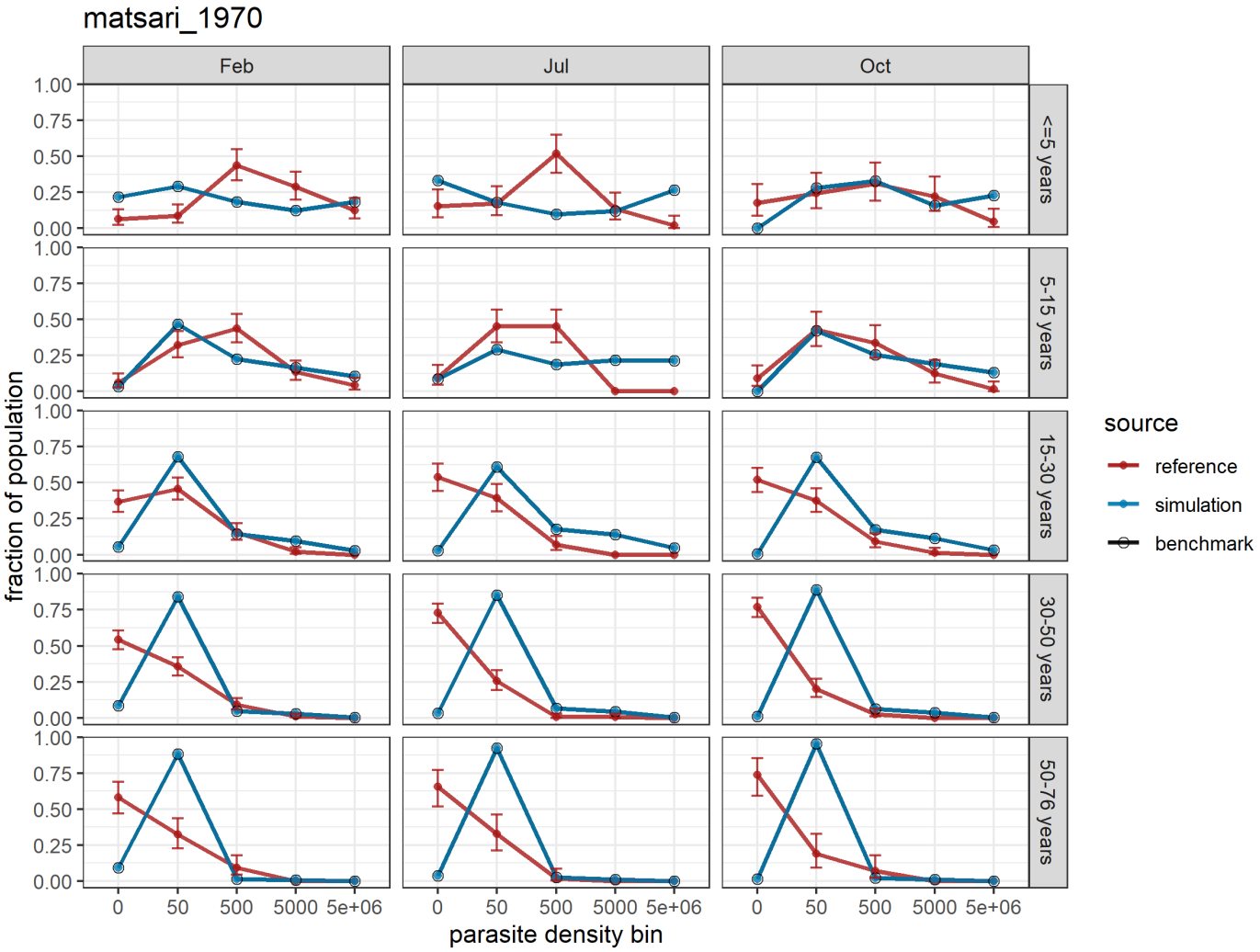


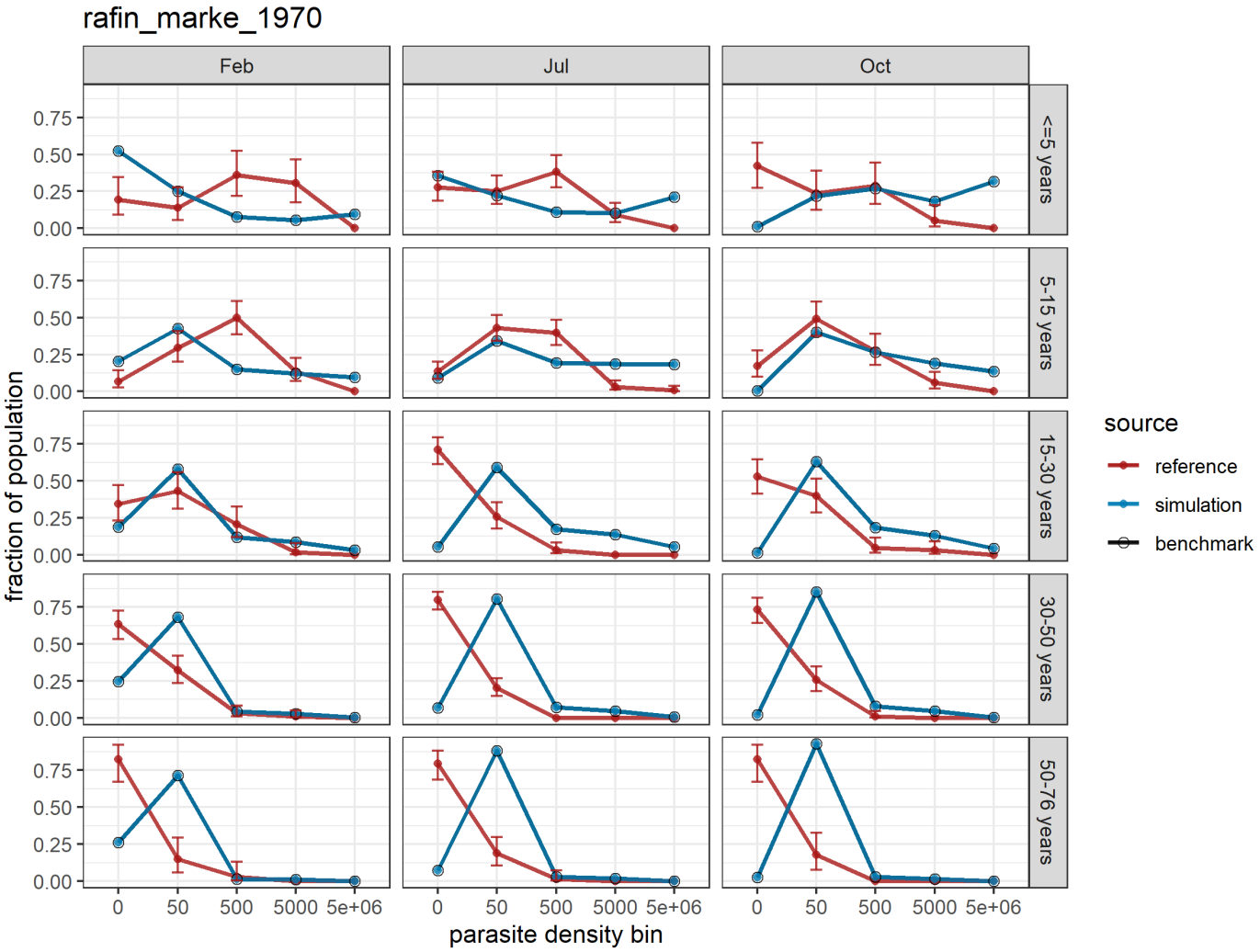
3.6 Asexual parasite density by age

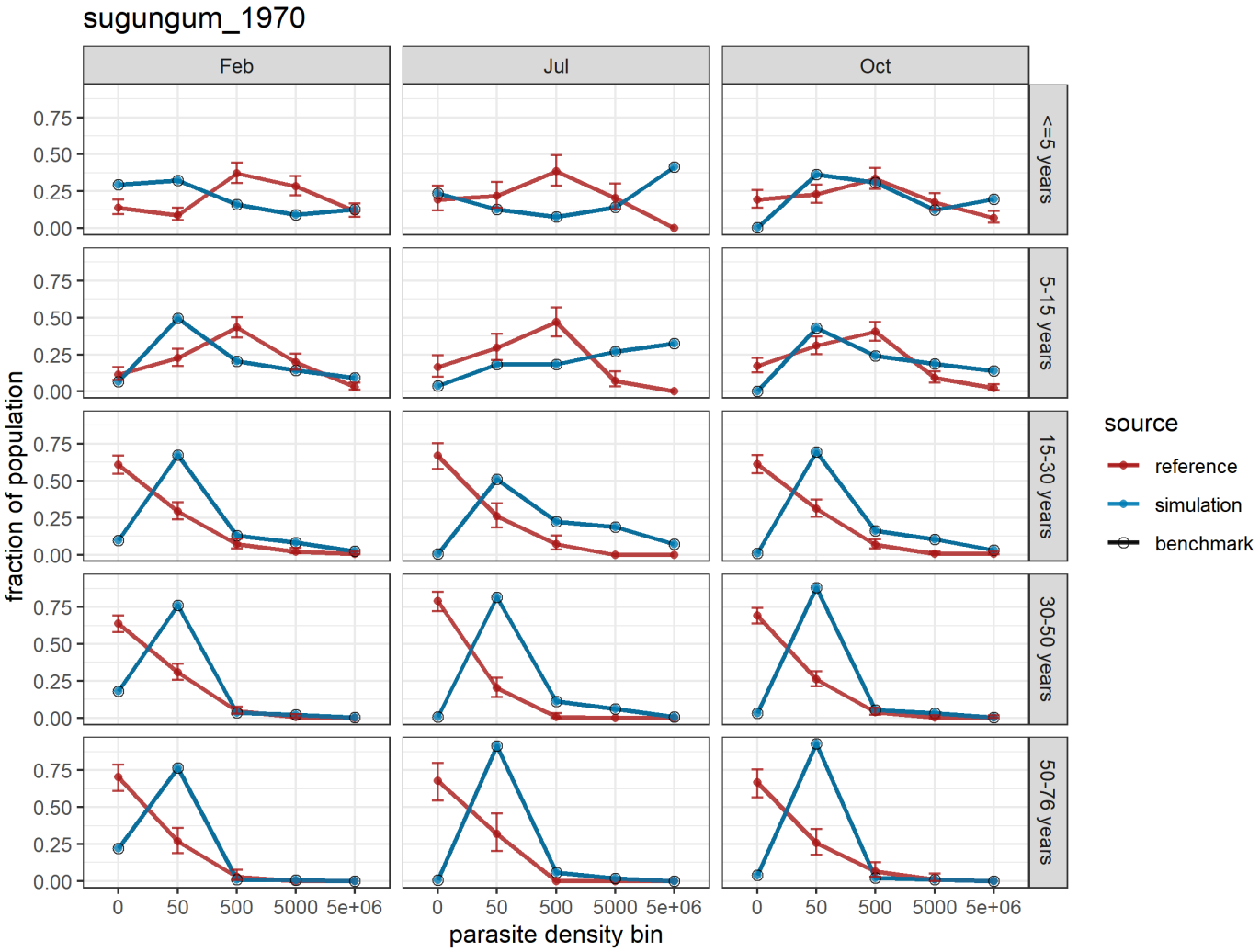
The plots below compare the distribution of parasite densities across ages and seasons from reference datasets and matched simulations. Each plot panel corresponds to a site. Note that some of the reference datasets have small sample sizes, especially in the youngest age groups.




laye_2007
Jan
Jul
Sep
<=5 years
5-15 years
15-80 years
fraction of population
0.00
0.25
0.50
0.75
0
50
500
5000
50000
5e+06
parasite density bin
source
reference
simulation
benchmark

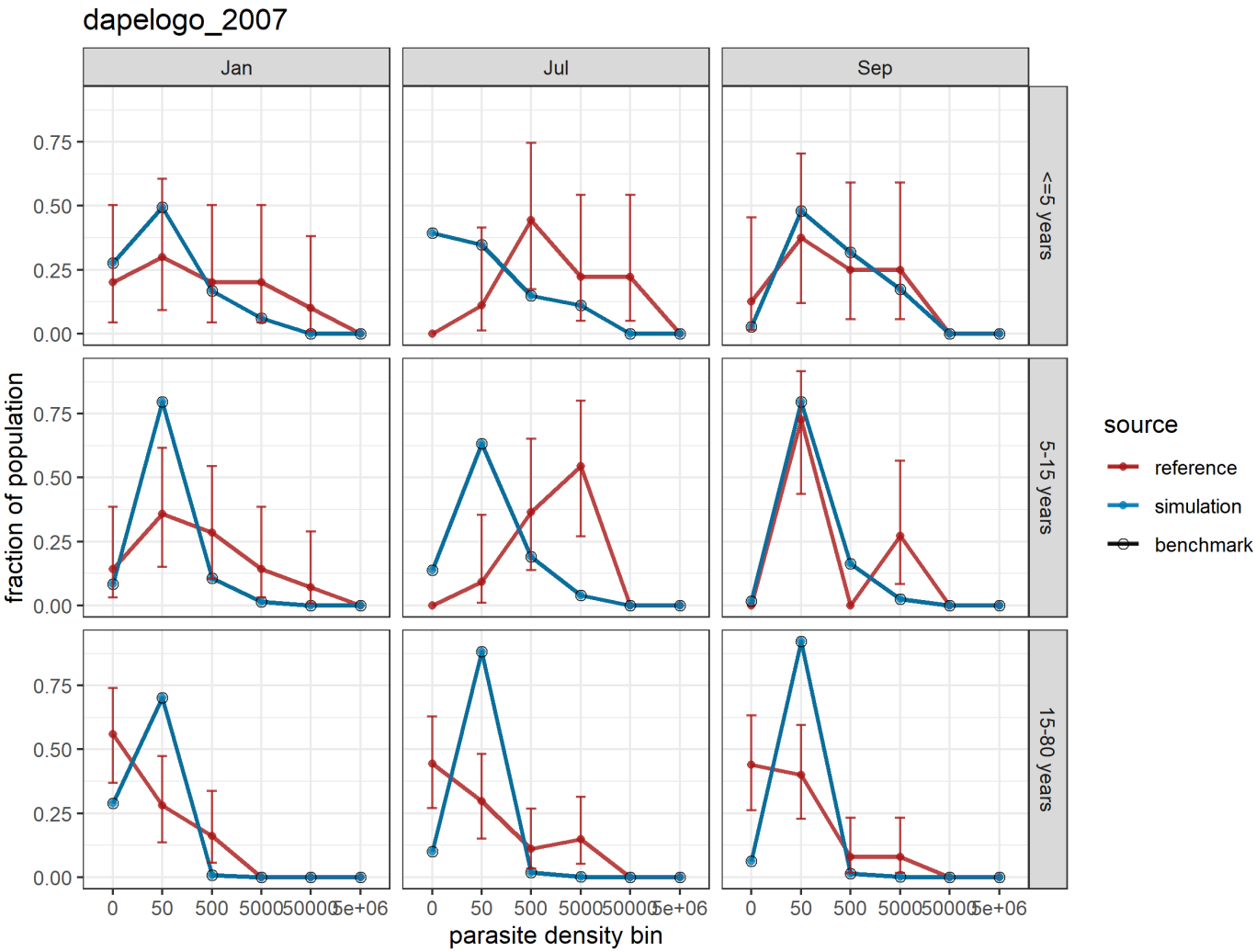

matsari_1970
Feb
Jul
Oct
<=5 years
5-15 years
15-30 years
30-50 years
50-76 years
fraction of population
parasite density bin
0
50
500
5000
5e+06
0.00
0.25
0.50
0.75
1.00
source
reference
simulation
benchmark

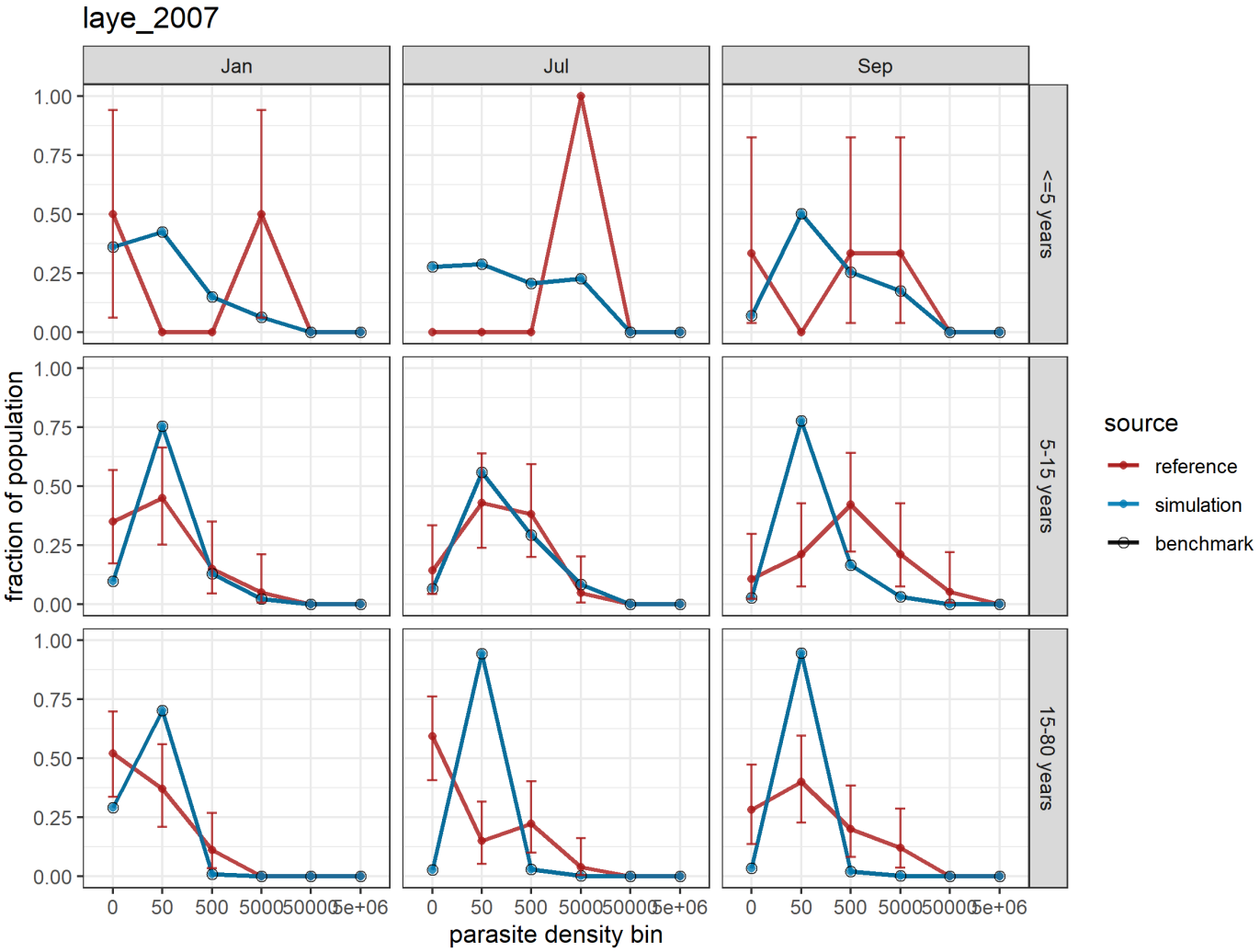

rafin_marke_1970
Feb
Jul
Oct
<=5 years
5-15 years
15-30 years
30-50 years
50-76 years
fraction of population
0.00
0.25
0.50
0.75
0
50
500
5000
5e+06
parasite density bin
source
reference
simulation
benchmark

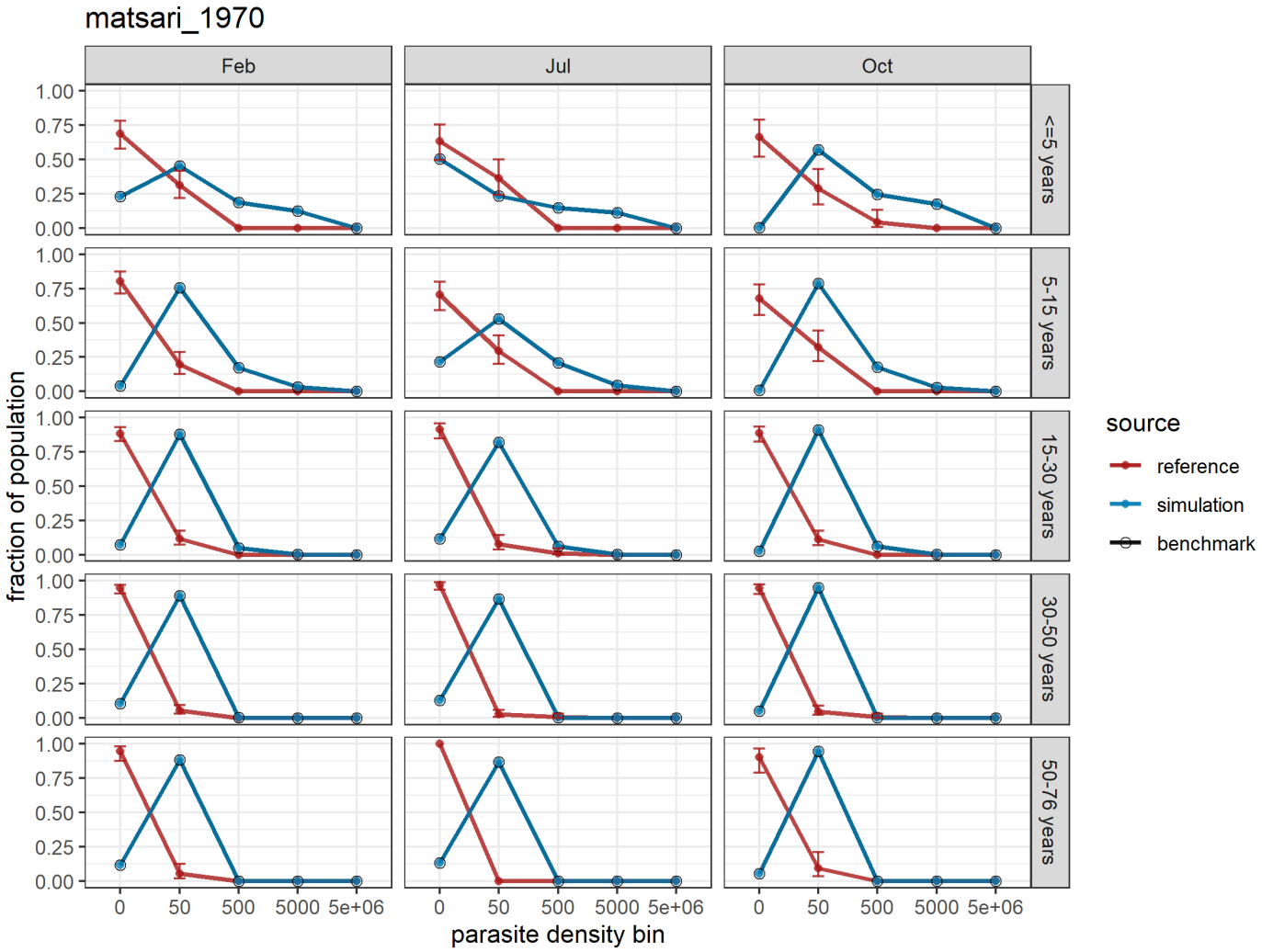

sugungum_1970
Feb
Jul
Oct
<=5 years
5-15 years
15-30 years
30-50 years
50-76 years
fraction of population
0.00
0.25
0.50
0.75
0
50
500
5000
5e+06
parasite density bin
source
reference
simulation
benchmark

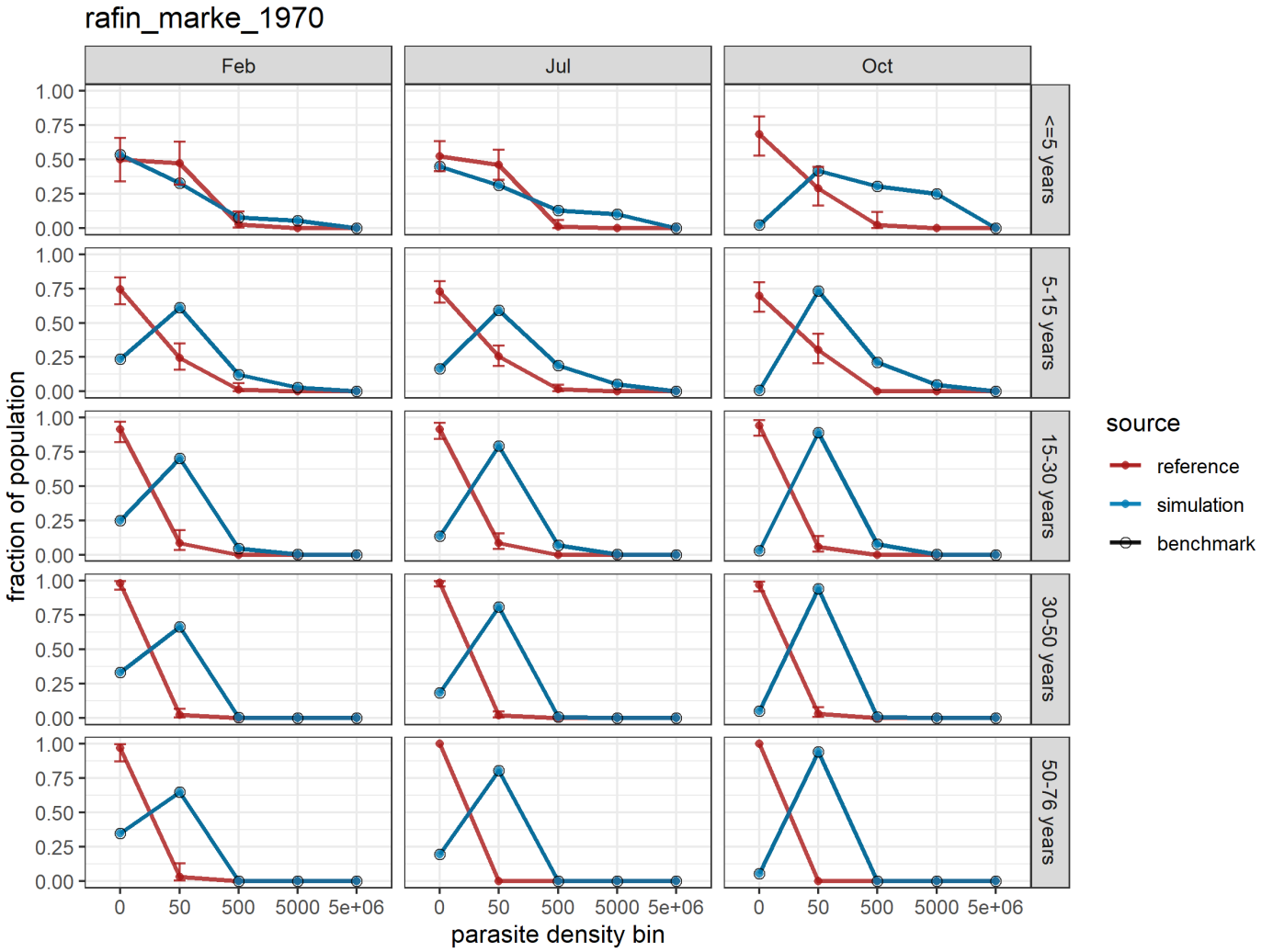
3.7 Gametocyte density by age

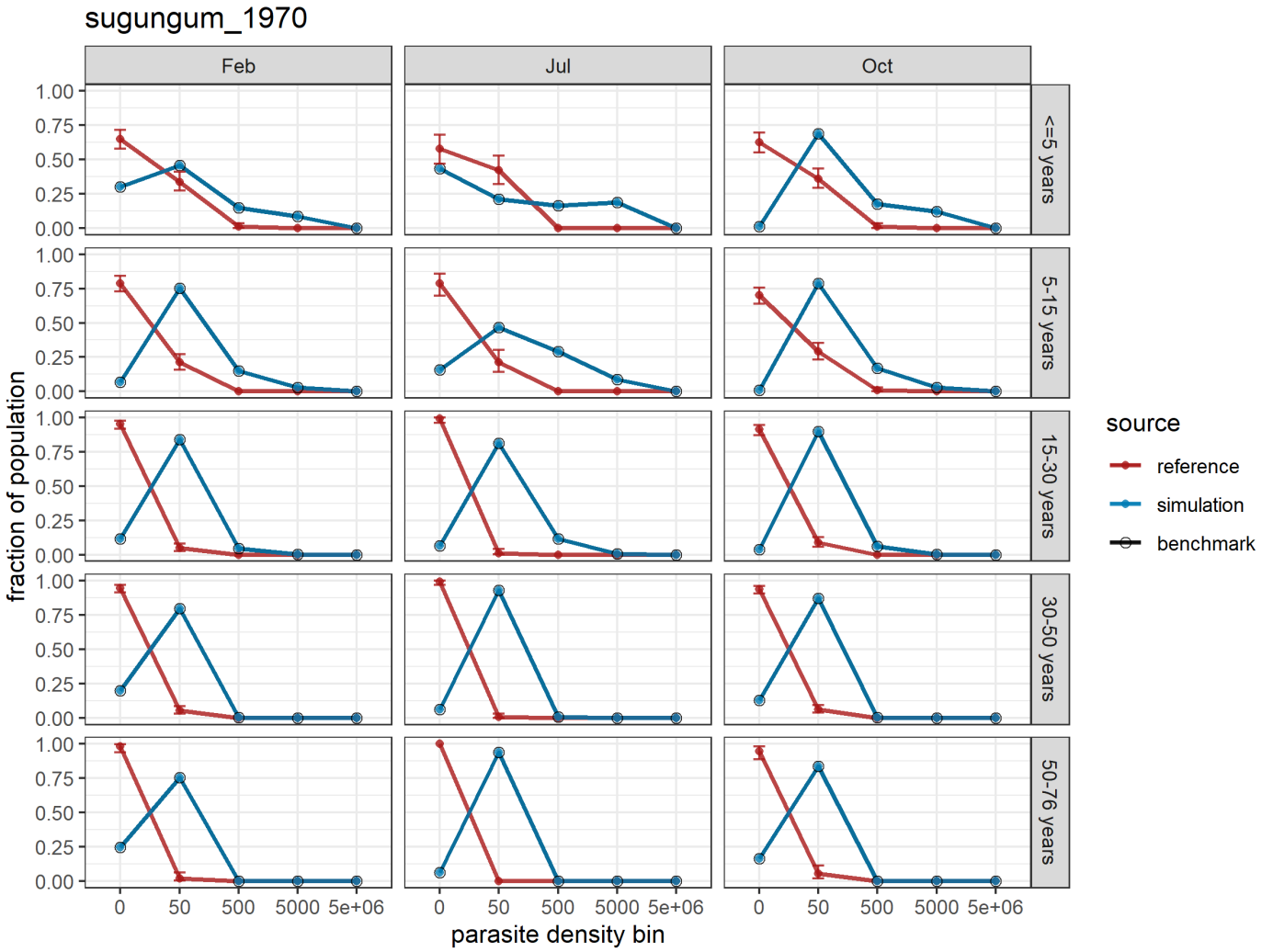
The plots below compare the distribution of gametocyte densities across ages and seasons from reference datasets and matched simulations. Each plot panel corresponds to a site. Note that some of the reference datasets have small sample sizes, especially in the youngest age groups.




laye_2007
Jan
Jul
Sep
<=5 years
5-15 years
15-80 years
fraction of population
parasite density bin
source
reference
simulation
benchmark


matsari_1970
Feb
Jul
Oct
<=5 years
5-15 years
15-30 years
30-50 years
50-76 years
fraction of population
1.00
0.75
0.50
0.25
0.00
0
50
500
5000
5e+06
parasite density bin
source
reference
simulation
benchmark

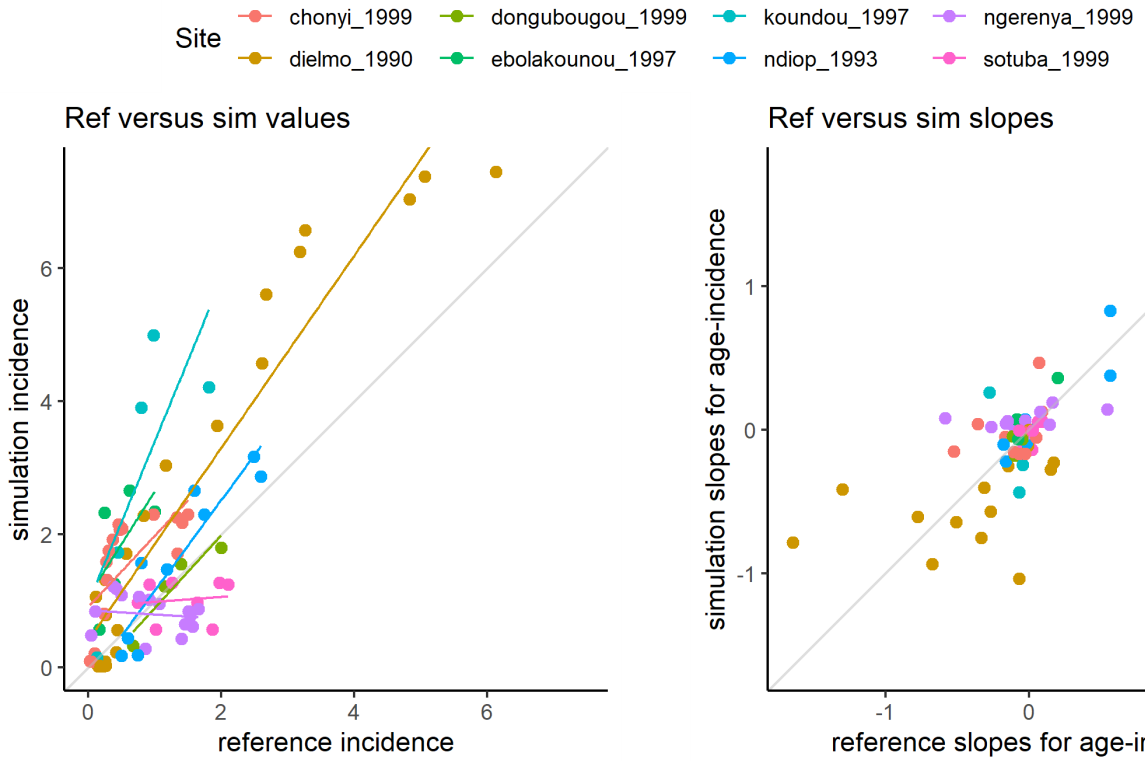

rafin_marke_1970
Feb
Jul
Oct
<=5 years
5-15 years
15-30 years
30-50 years
50-76 years
fraction of population
1.00
0.75
0.50
0.25
0.00
0
50
500
5000
5e+06
parasite density bin
source
reference
simulation
benchmark


sugungum_1970
Feb
Jul
Oct
<=5 years
5-15 years
15-30 years
30-50 years
50-76 years
fraction of population
1.00
0.75
0.50
0.25
0.00
0
50
500
5000
5e+06
parasite density bin
source
reference
simulation
benchmark

4. Additional comparisons of reference data and matched simulations

4.1 Incidence by age

Below, the plots show the correlation between the simulation and reference incidence in each age group (left plot) and the correlation between the simulation and reference slopes when moving between the incidence in one age group and the incidence in the next oldest age group (right plot).

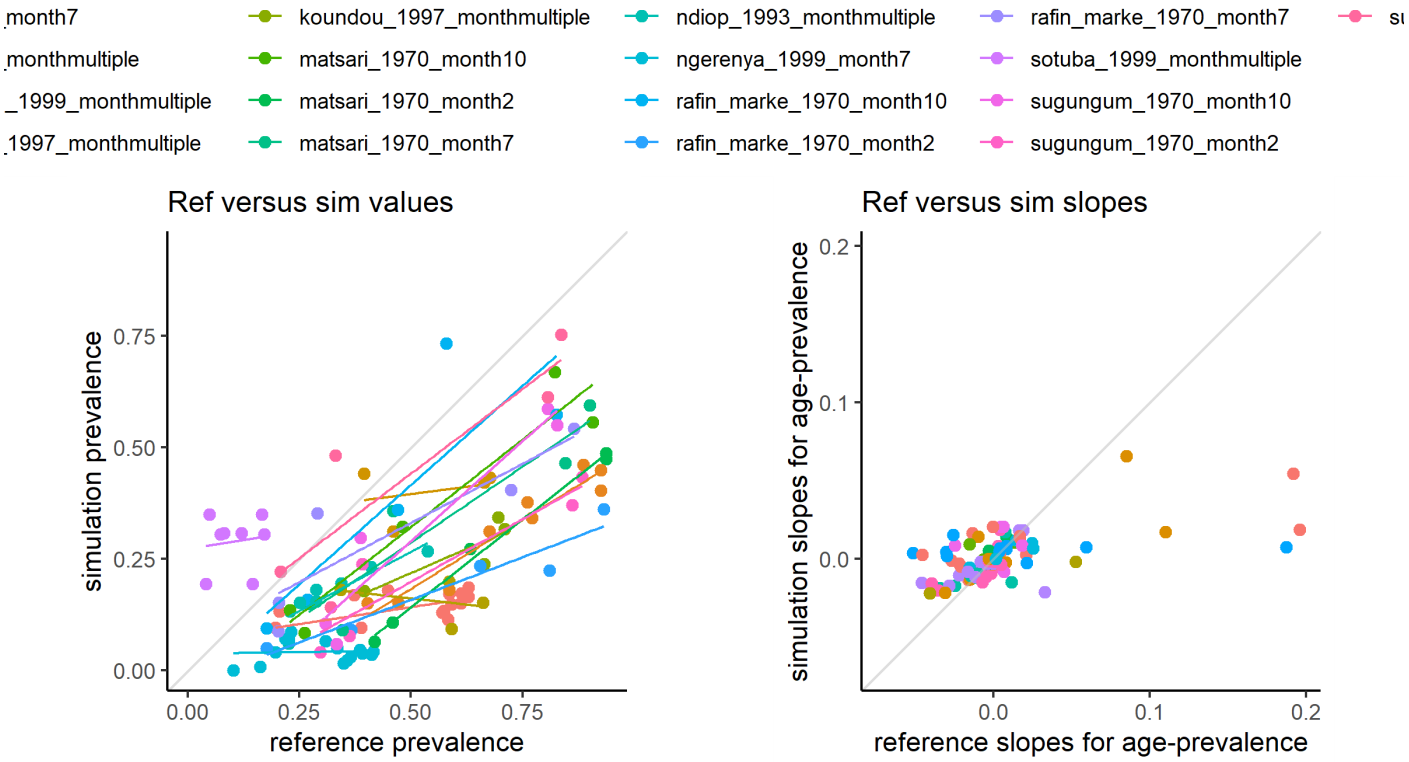


Site	mean_rel_diff	mean_abs_diff	corr_slope	corr_r_squared	mean_rel_slope_diff	mean_abs_slpe_diff
all_sites	1.67	0.94	nan	nan	nan	nan
chonyi_1999	2.34	0.95	1.07	0.48	8.28	0.13
dielmo_1990	1.42	1.23	1.45	0.91	2.95	0.34
dongubougou_1999	0.2	0.19	1.09	0.87	0.61	0.06
ebolakounou_1997	3.51	1.34	1.58	0.36	5.51	0.09
koundou_1997	2.45	2.16	2.43	0.6	3.96	0.28
ndiop_1993	0.47	0.52	1.36	0.85	1.57	0.1
ngerenya_1999	2.19	0.62	-0.06	0.02	1.54	0.15

sotuba_1999	0.37	0.57	0.07	0.02	2.02	0.06

4.2 Prevalence by age

Below, the plots show the correlation between the simulation and reference prevalence in each age group (left plot) and the correlation between the simulation and reference slopes when moving between the prevalence in one age group and the prevalence in the next oldest age group (right plot).



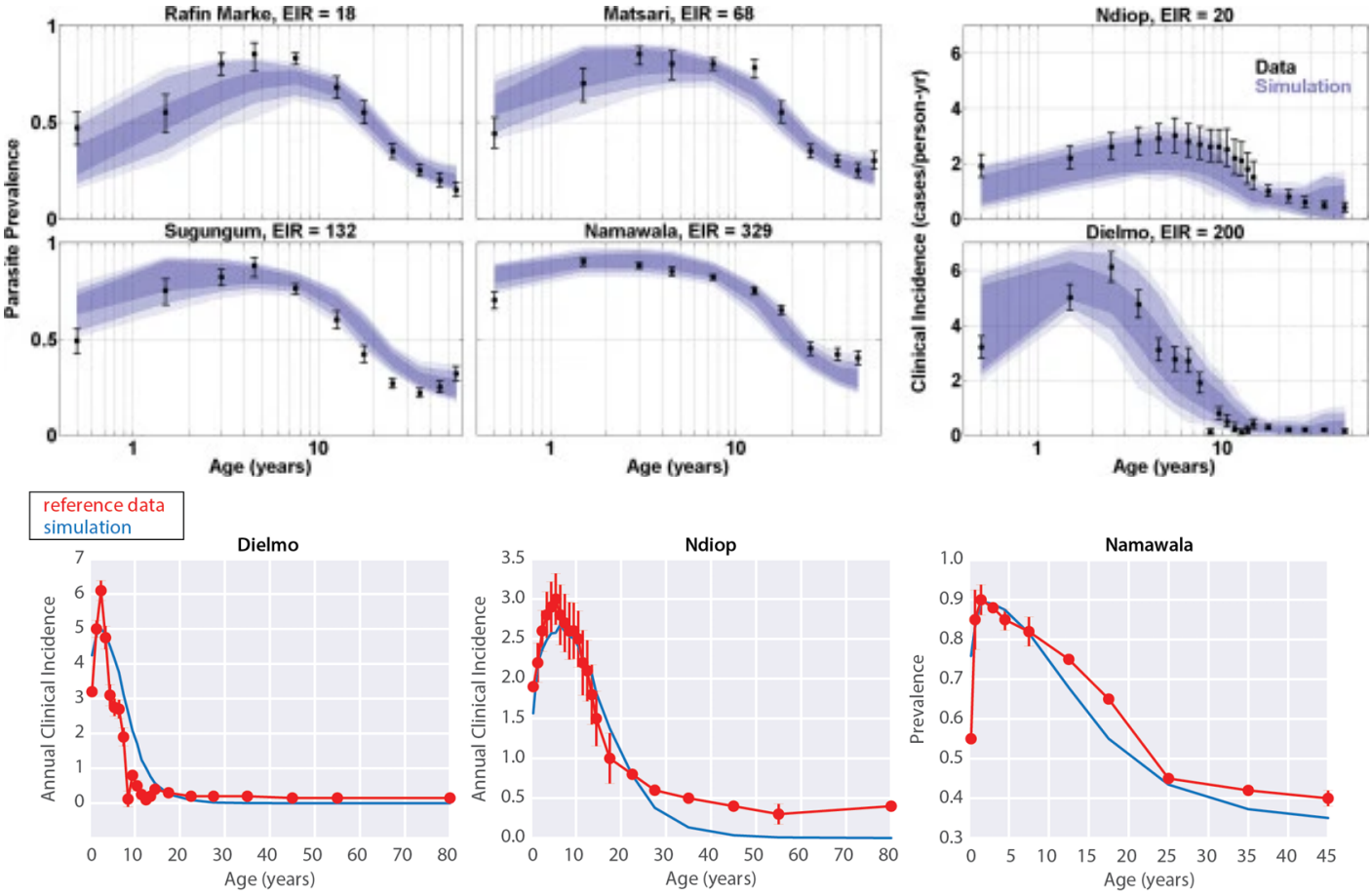
Site	mean_rel_diff	mean_abs_diff	corr_slope	corr_r_squared	mean_rel_slope_diff	mean_abs_slpe_diff
all_sites	0.71	0.27	nan	nan	nan	nan
chonyi_1999_month7	0.68	0.34	0.15	0.41	8.34	0.04
dielmo_1990_monthmultiple	0.57	0.4	0.62	0.93	0.88	0.02
dongubougou_1999_monthmultiple	0.29	0.17	0.13	0.09	0.59	0.02
ebolakounou_1997_monthmultiple	0.69	0.39	-0.11	0.11	0.9	0.01
koundou_1997_monthmultiple	0.56	0.35	0.43	0.74	1.3	0.01
matsari_1970_month10	0.4	0.19	0.78	0.93	1.38	0.01
matsari_1970_month2	0.63	0.4	0.78	0.99	inf	0.0
matsari_1970_month7	0.44	0.23	0.68	0.87	0.92	0.01
ndiop_1993_month						

multiple	0.47	0.15	0.53	0.8	0.6	0.01
ngerenya_1999_month7	0.85	0.26	0.01	0.0	0.92	0.03
rafin_marke_1970_month10	0.34	0.14	0.89	0.72	0.59	0.02
rafin_marke_1970_month2	0.69	0.4	0.38	0.92	0.49	0.01
rafin_marke_1970_month7	0.37	0.17	0.53	0.8	60.8	0.01
sotuba_1999_monthmultiple	2.48	0.18	0.16	0.02	1.62	0.01
sugungum_1970_month10	0.46	0.21	0.89	0.92	1.89	0.01
sugungum_1970_month2	0.62	0.33	0.56	0.87	1.86	0.01
sugungum_1970_month7	0.29	0.12	0.76	0.77	1.76	0.01

5. Comparisons from prior EMOD publications

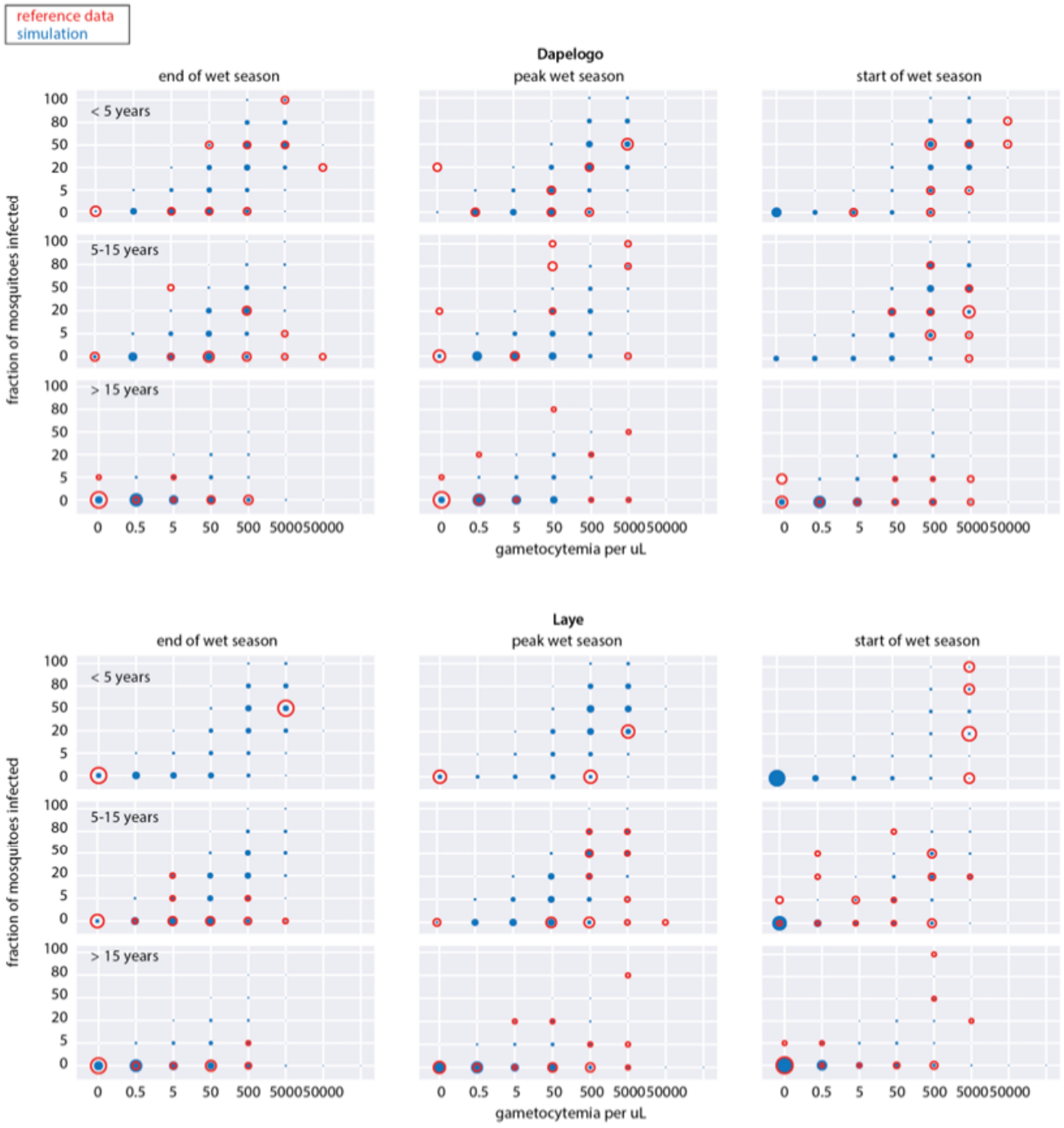
5.1 Incidence and prevalence by age

The top plots come from McCarthy et al. 2015 and the bottom plots come from Selvaraj et al. 2018.



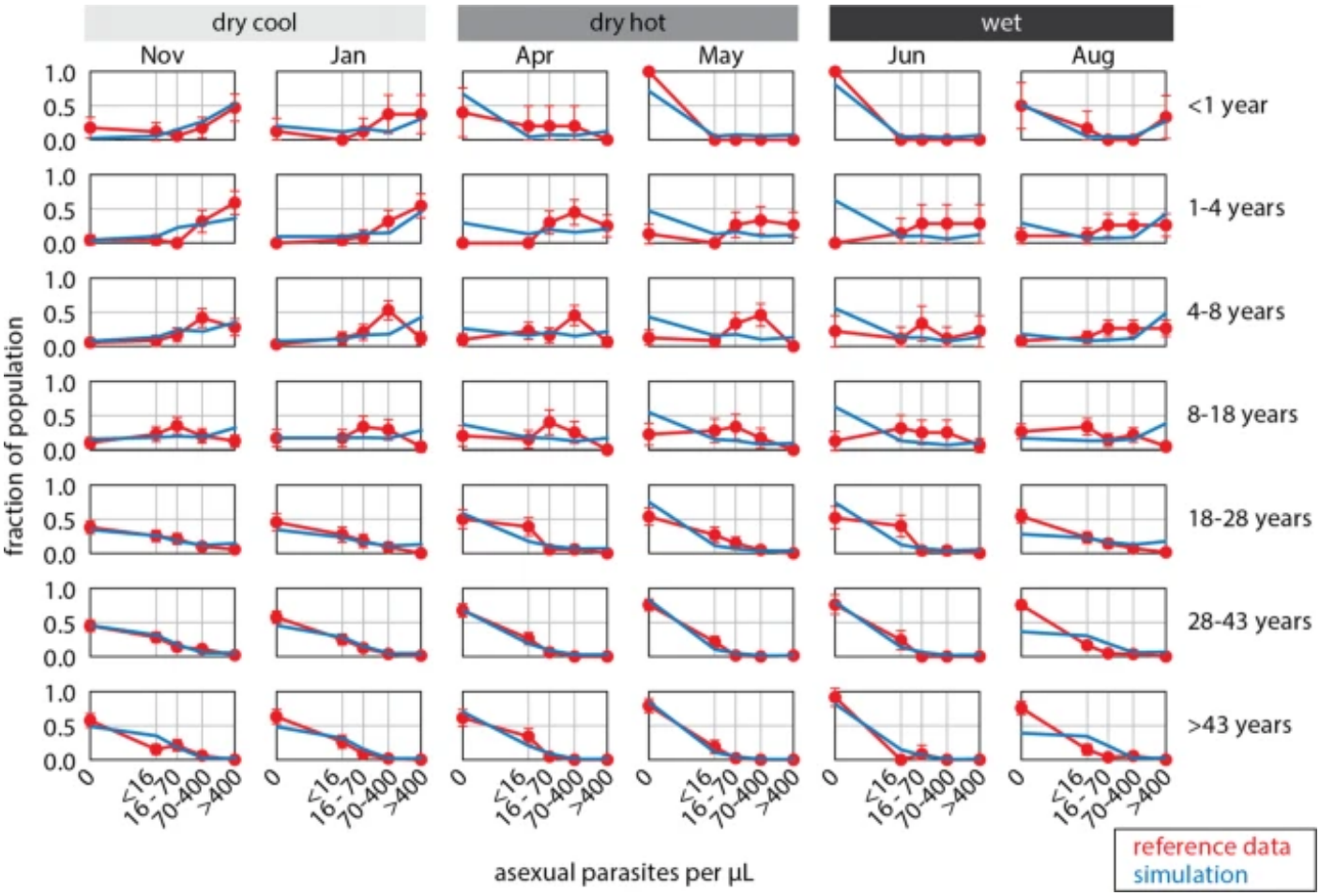
5.2 Infectiousness to vectors

The following plot comes from Selvaraj et al. 2018

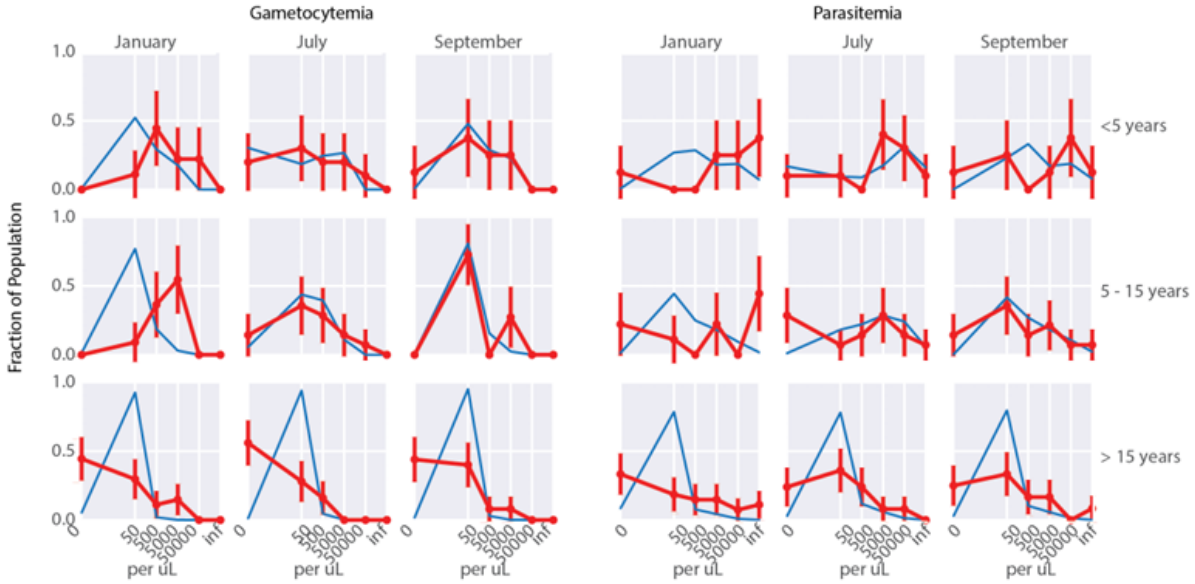


5.3 Parasite densities

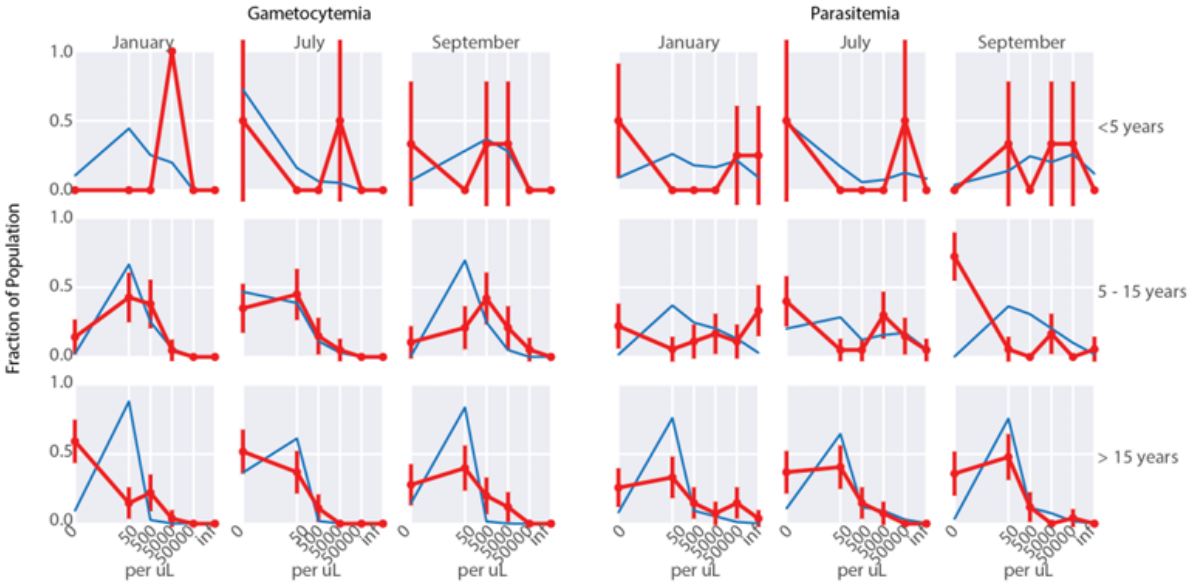
The following plots come from Selvaraj et al. 2018

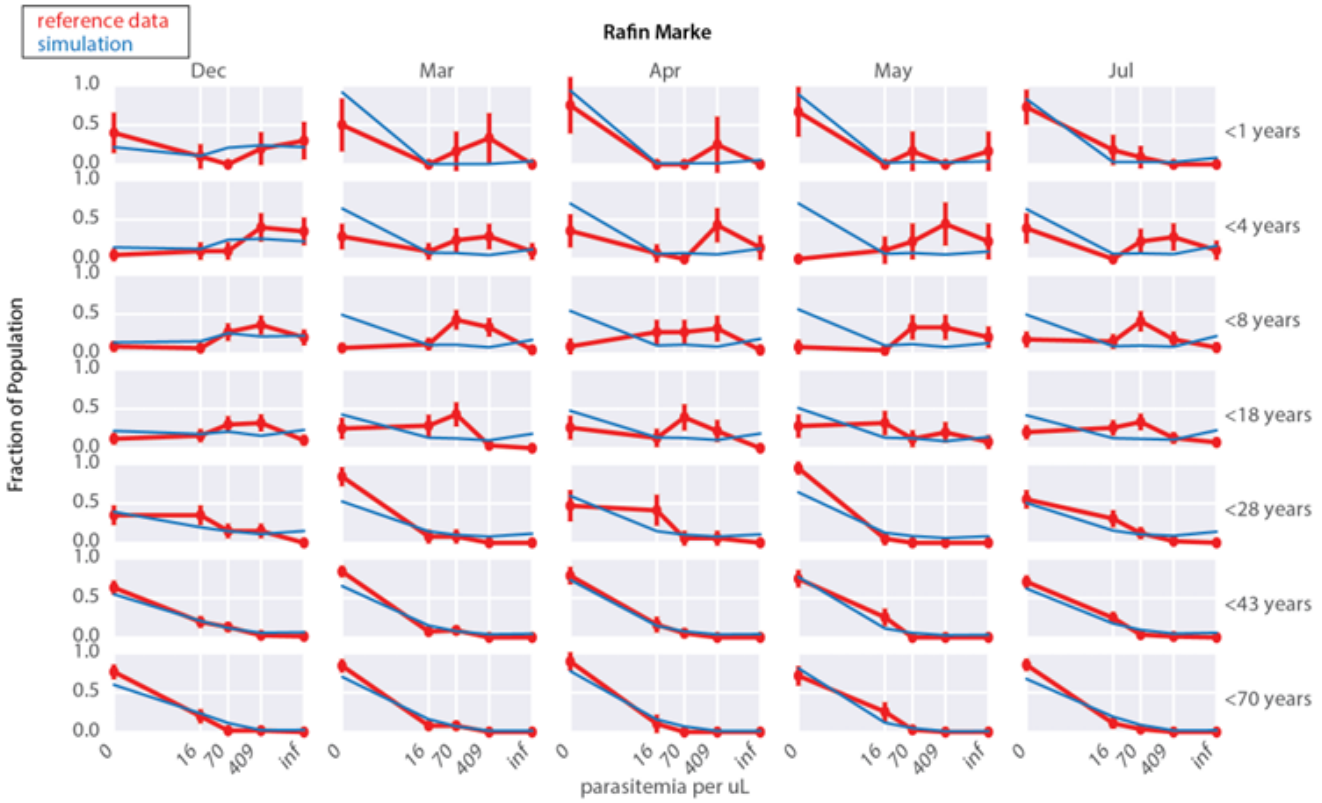


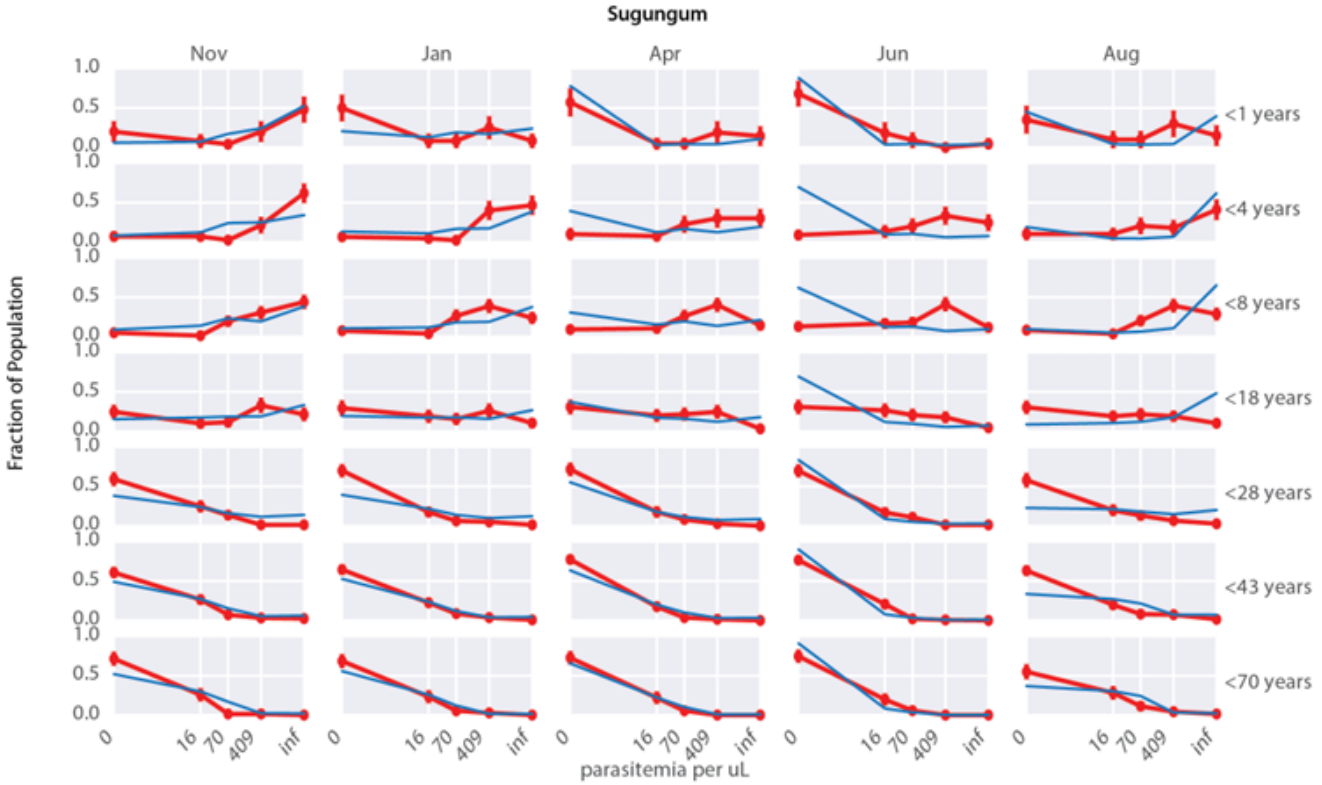
Dapelogo



Laye




reference data
simulation
Rafin Marke
Dec
Mar
Apr
May
Jul
<1 years
<4 years
<8 years
<18 years
<28 years
<43 years
<70 years
Fraction of Population
1.0
0.5
0.0
0
16
70
409
inf
parasitemia per uL


Sugungum
Nov
Jan
Apr
Jun
Aug
<1 years
<4 years
<8 years
<18 years
<28 years
<43 years
<70 years
Fraction of Population
1.0
0.5
0.0
0
16
70
409
inf
parasitemia per uL

This page left unintentionally blank