Descriptive Analyses

Francois Rerolle

August 7, 2018

Report of some descriptive statistics on the final cleaned dataset (A3 + environmental covariates).

1 Description of dataset

1.1 What, where and when

In 2017, we collected all A3 forms available in 4 districts of the southern province of Champasak in Laos: Moonlapamok, Pathoomphone, Sanasomboon and Sukhuma. A3 forms record all suspected malaria cases passively detected that got tested for malaria by RDT and/or microscopy.

A total of 64501 A3 forms were collected. Individuals reported coming from 505 villages (4.2 % missing) in 7 districts (0 % missing). Most individuals lived in districts where A3 form was collected but some lived in other districts:

```
##
##    Champasak        Khong  Moonlapamok     Paksxong Pathoomphone
##          958          578        15741            2        24316
##   Phongthong  Sanasomboon      Sukhuma         <NA>
##            3        14158         8747            0
```

Figure 1 shows a time distribution of A3 forms ranging from $2013-09-25$ to $2016-12-29$, pretty constant over a 3 year period from October 2013 to October 2016. Figure 2 shows that this true across the 4 districts of data collection.

1.2 Description of variables

There are two types of variables available in this dataset:

- Variables that were collected in the A3 form: date, district, village, age, gender, occupation, RDT result, microscopy result and treatment taken.
- Environmental covariates that were extracted from raster layers via GPS coordinates of villages (matched via offical list of villages in Champasak): altitude, temperature, precipitation, population, forest percent in 1km and 10km for the data collection and previous 5 years.

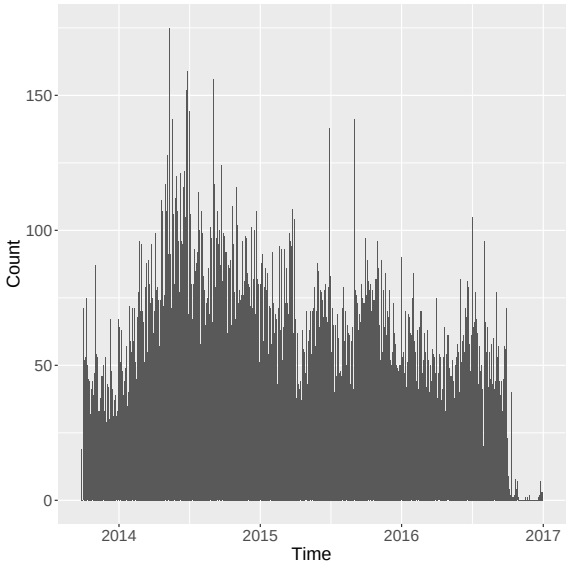

Figure 1: Histogram of A3 data collection times.

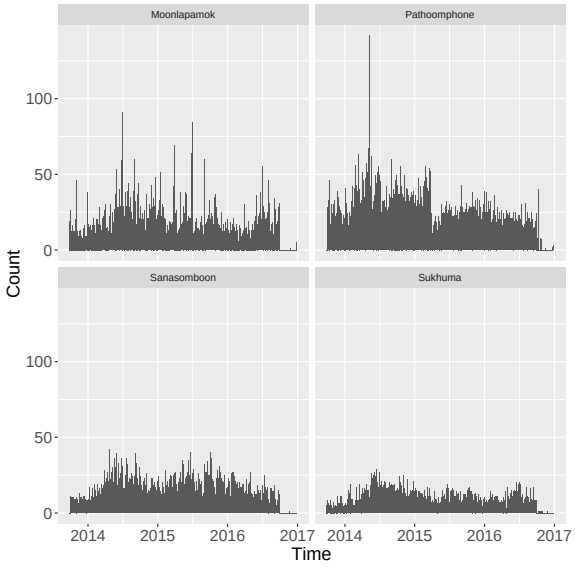

Figure 2: Histogram of A3 data collection times per district.

2 Descriptive statistics

2.1 Basic description

2.1.1 A3 variables

Variable	Description	Missing
Date	Date of the A3 interview, over 3 years from October 2013 to 2016.	0.3 %
District	Reported district where living. 8 different districts of Champasak Province. 97.6 % come from the 4 districts where A3 forms where collected: MP, PT, SB and SK.	0 %
Village	Reported village where living. 506 different villages of Champasak Province, 497 in MP, PT, SB or SK.	4.2 %
Age	Reported age. Ranges from 0 to 98, with a median of 26 years.	0.7 %
Male	Reported gender. 71 % of males.	2.2 %
Occupation	Reported occupation. 6 different occupations. 67.7 % are farmers.	8.7 %
RDT	RDT results. 16.9 % Pf, 19.4 % Pv and 1.8 % mix.	29.4 %
Microscopy	Microscopy results. 8 % Pf, 10.7 % Pv and 1.7 % mix.	75.7 %
Treatment	Treatment provided to A3 individuals. 4 different treatment combinations. 78 % received no treatment and 21.3 % received Coartem alone.	0.1 %
GPS	GPS coordinates of villages matched with official list of villages in Champasak with GPS. Environmental covariates were extracted for all villages with GPS coordinates.	11.1 %

Table 1: Basic description of A3 variables in dataset.

2.1.2 Environmental variables

Variable	Description	Missing
Altitude	Altitude, extracted from SRTM with resolution 1km. Ranges from 89 to 301, with a mean of 128 and a median of 114 meters.	11.4 %
Temperature	Mean annual temperature, extracted from worldclim with resolution 1km. Ranges from 25.6 to 27, with a mean of 27 and a a median of 26.8 °C. We also have a measure of seasonality with the standard deviation.	11.1 %
Precipitation	Total annual precipitation, extracted from worldclim with resolution 1km. Ranges from 1831 to 2550, with a mean of 2193 and a a median of 2193 millimeters. We also have a measure of seasonality with the coefficient of variation.	11.1 %
Population	Population, extracted from worldpop with resolution 100m, aggregated at 1km. Ranges from 9 to 402, with a mean of 61 and a median of 52 habitants in km^2. Available for 2010 and 2015.	11.1 %
Forest	Percent forest within 10 km (*resp.* 1 km), computed from landcover (High-Biomass Vegetation) layers produced via classification algorithm on remote landsat data. Ranges from 0 to 83.1 (*resp.* 0 to 97.8), with a mean of 27.3 % (*resp.* 5.9) and a median of 25.1 % (*resp.* 0.4).	11.5 %
Forest change	Absolute percent forest change within 10 km (*resp.* 1 km), in previous 2 years. Ranges from −11.4 to 8.4 (*resp.* −32.2 to 32.7), with a mean of −2.7 % (*resp.* −1.8) and a median of −3 % (*resp.* −0.2). Available for previous 5 years.	11.5 %

Table 2: Basic description of environmental variables in dataset.

2.2 Descriptive summary

2.2.1 A3 variables

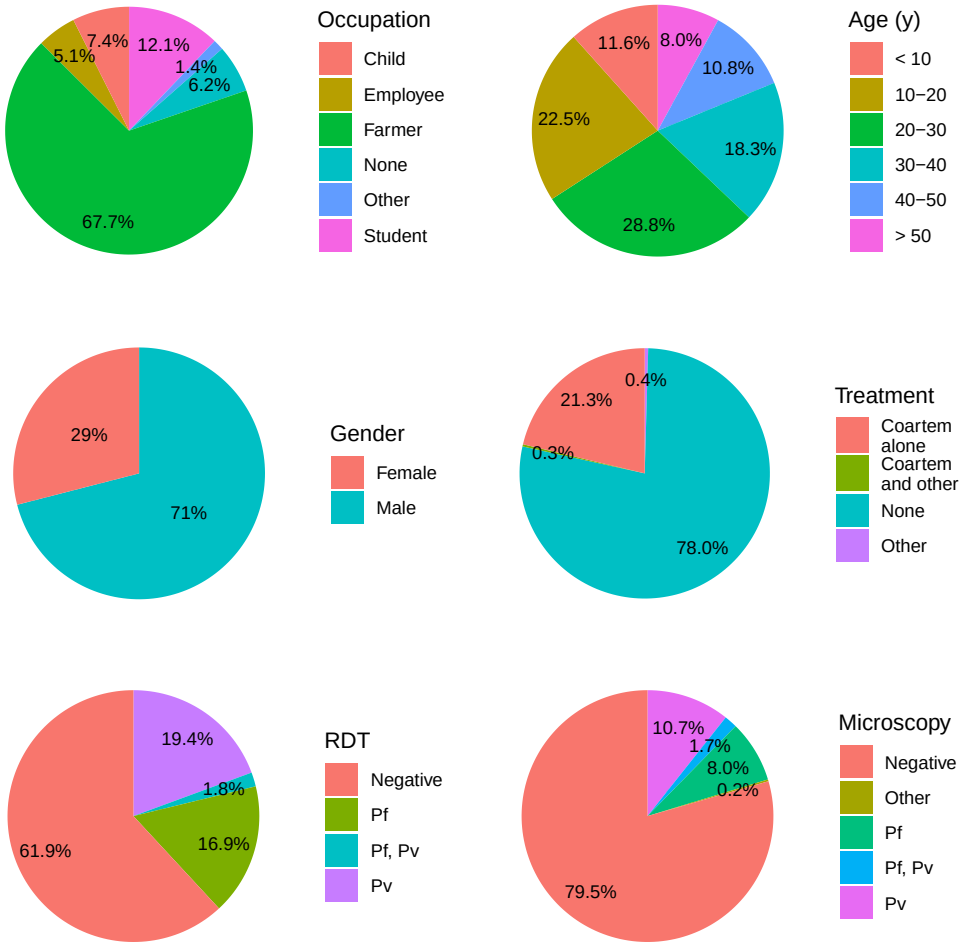

Figure 3: Pie charts of A3 variables.

2.2.2 Environmental variables

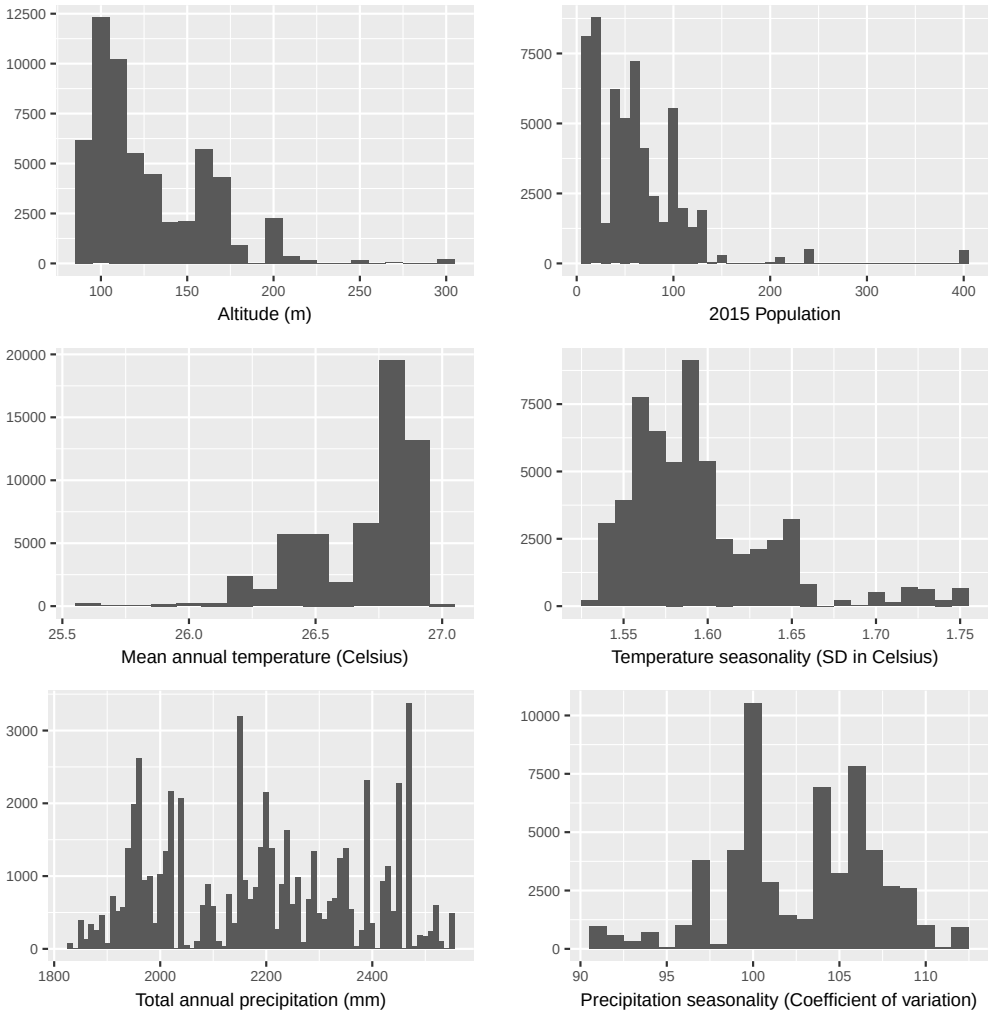

Figure 4: Histogram of environmental covariates at A3 villages.

2.2.3 Forest variables

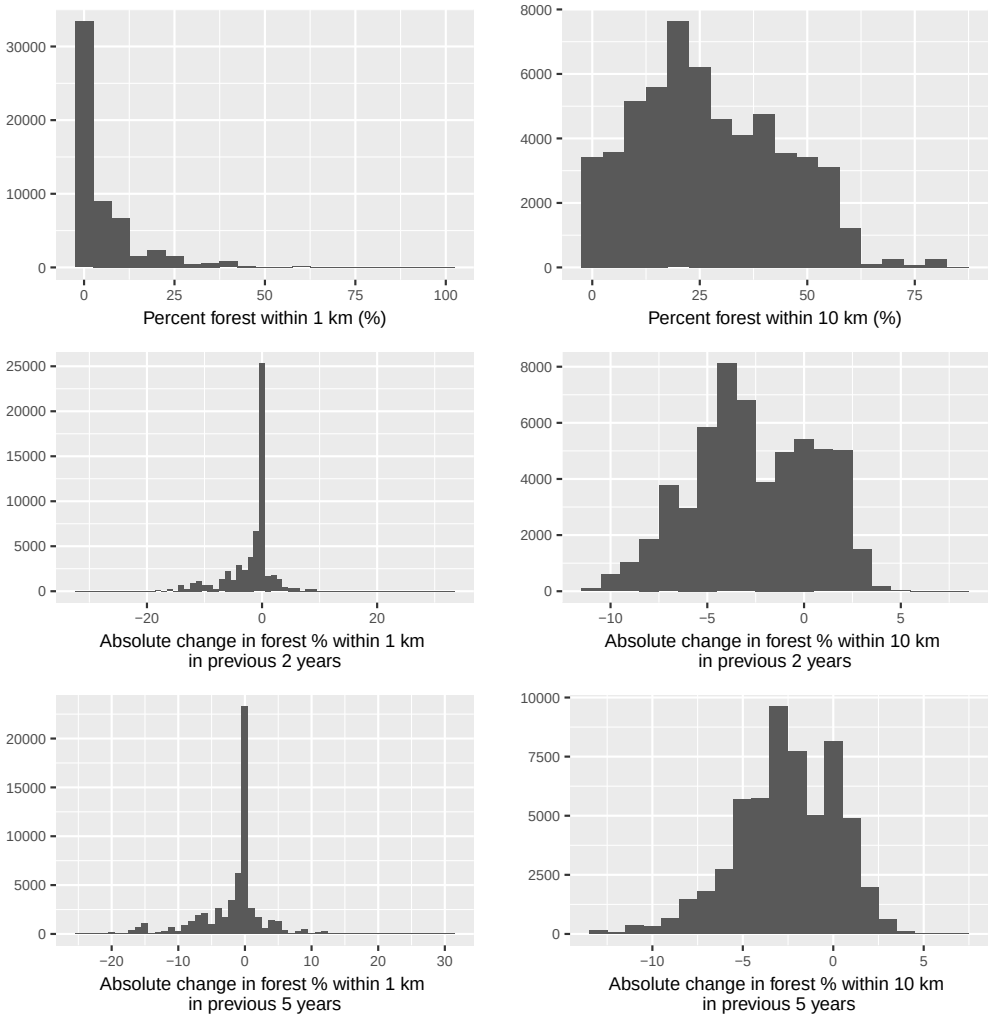

Figure 5: Histogram of forest covariates at A3 villages.

2.3 Variables' correlations

Figure 6: Correlogram of variables in dataset.

2.4 Unadjusted associations between forest and malaria

2.4.1 Plasmodium Falciparum

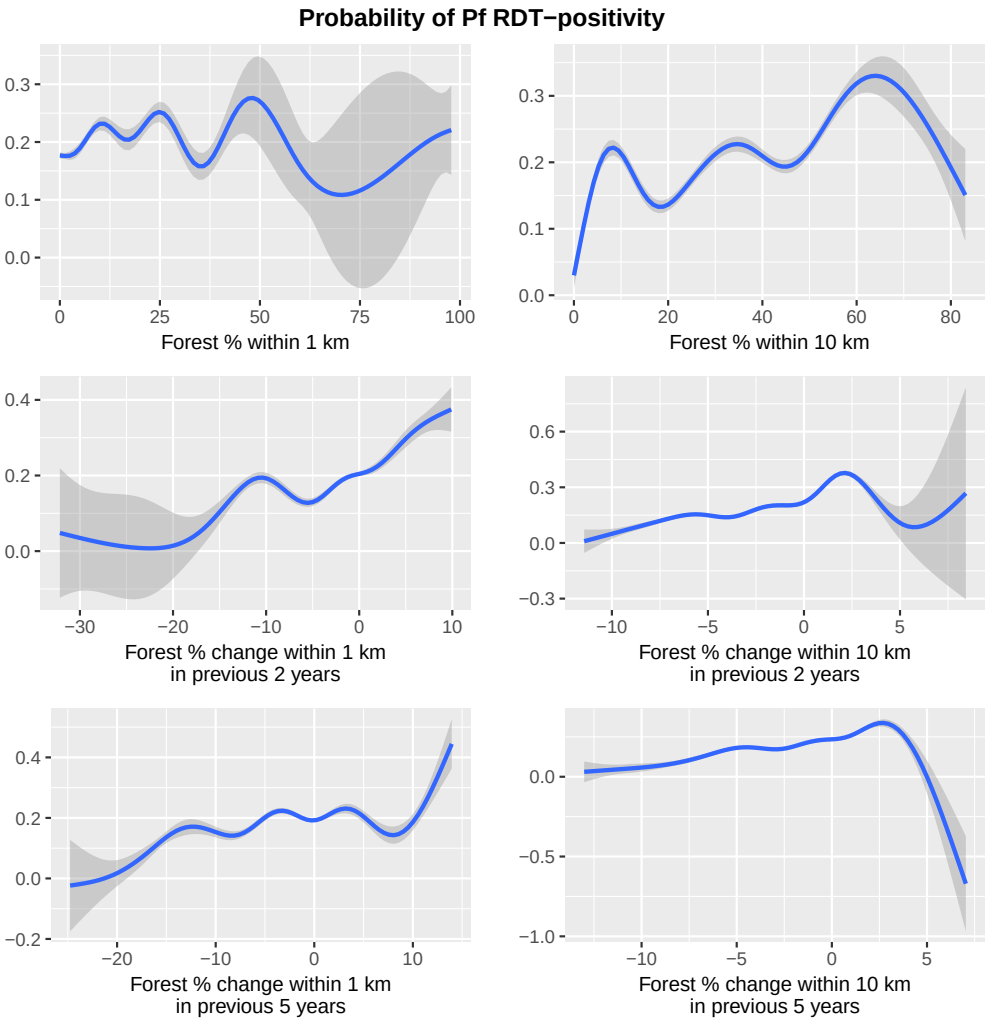

Figure 7: GAM smooth plots of association between forest and Pf RDT positivity.

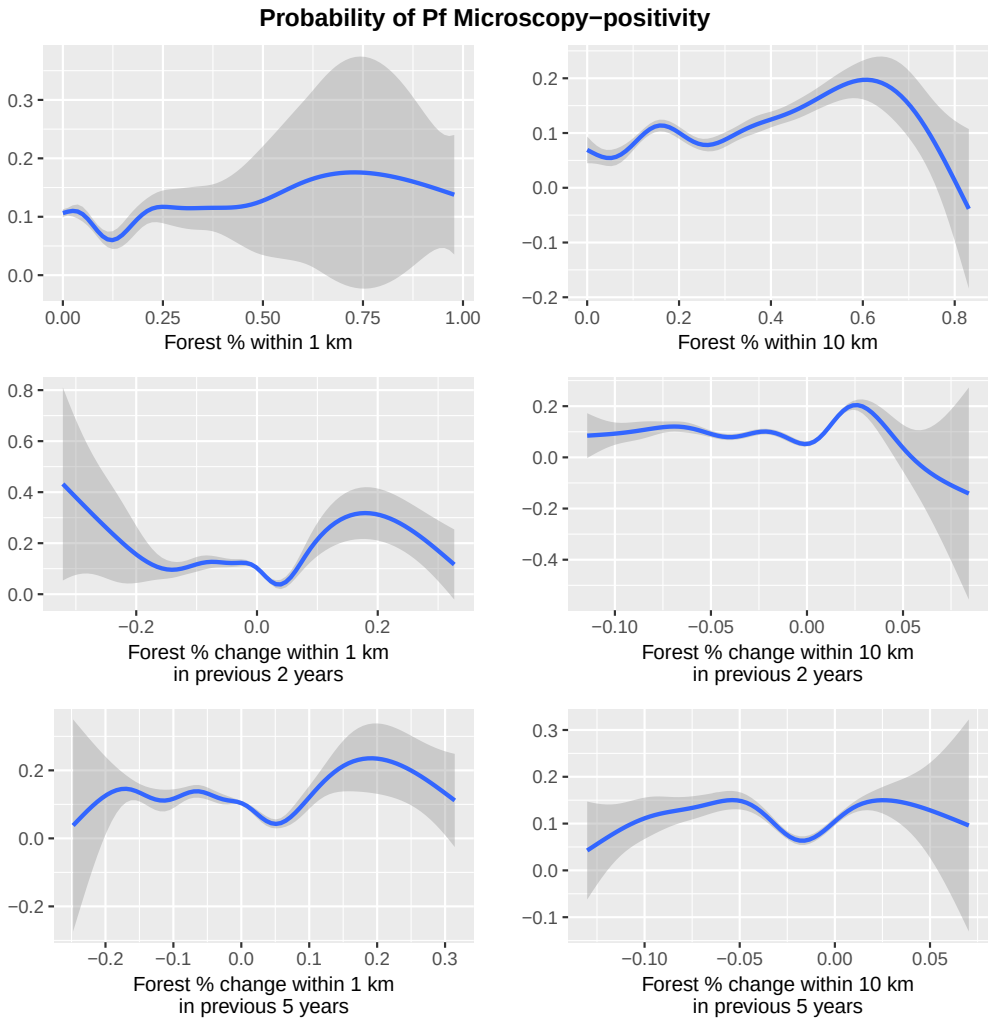

Figure 8: GAM smooth plots of association between forest and Pf Microscopy positivity.

2.4.2 Plasmodium Vivax

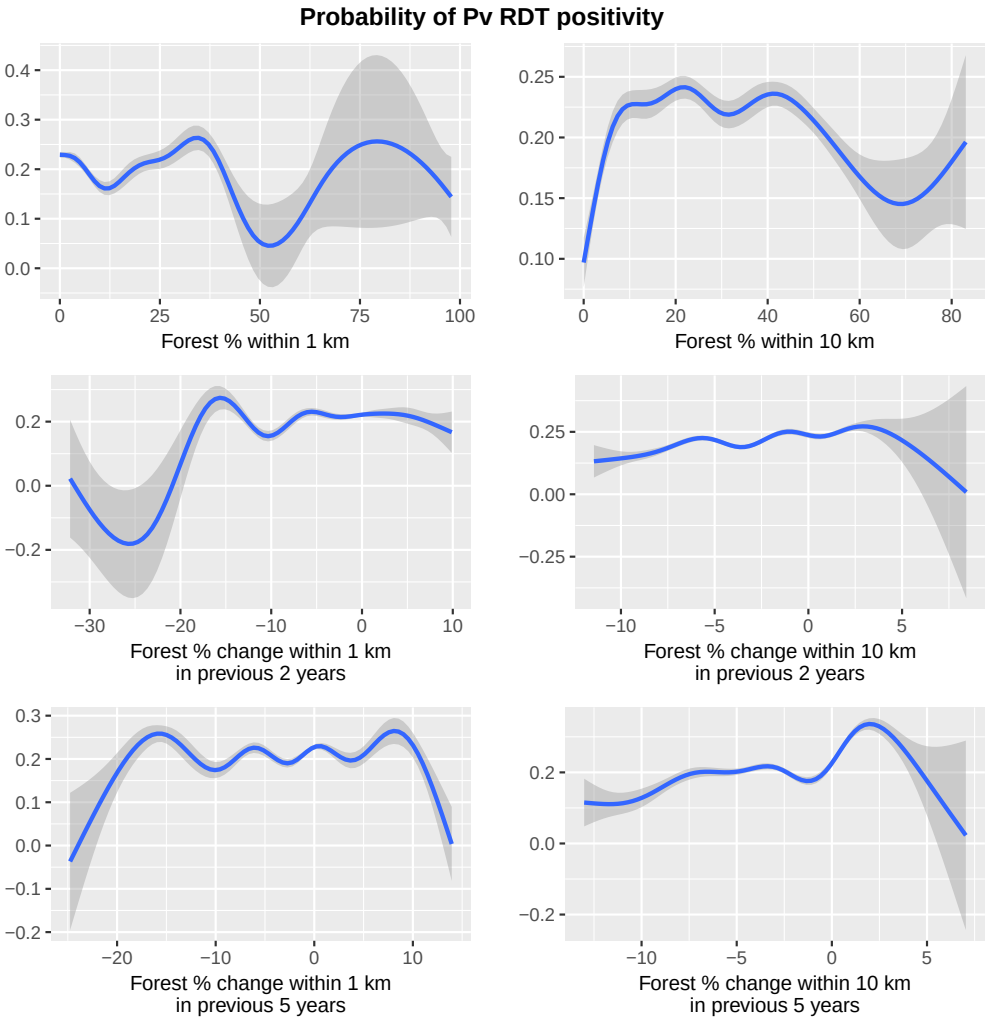

Figure 9: GAM smooth plots of association between forest and Pv RDT-positivity.

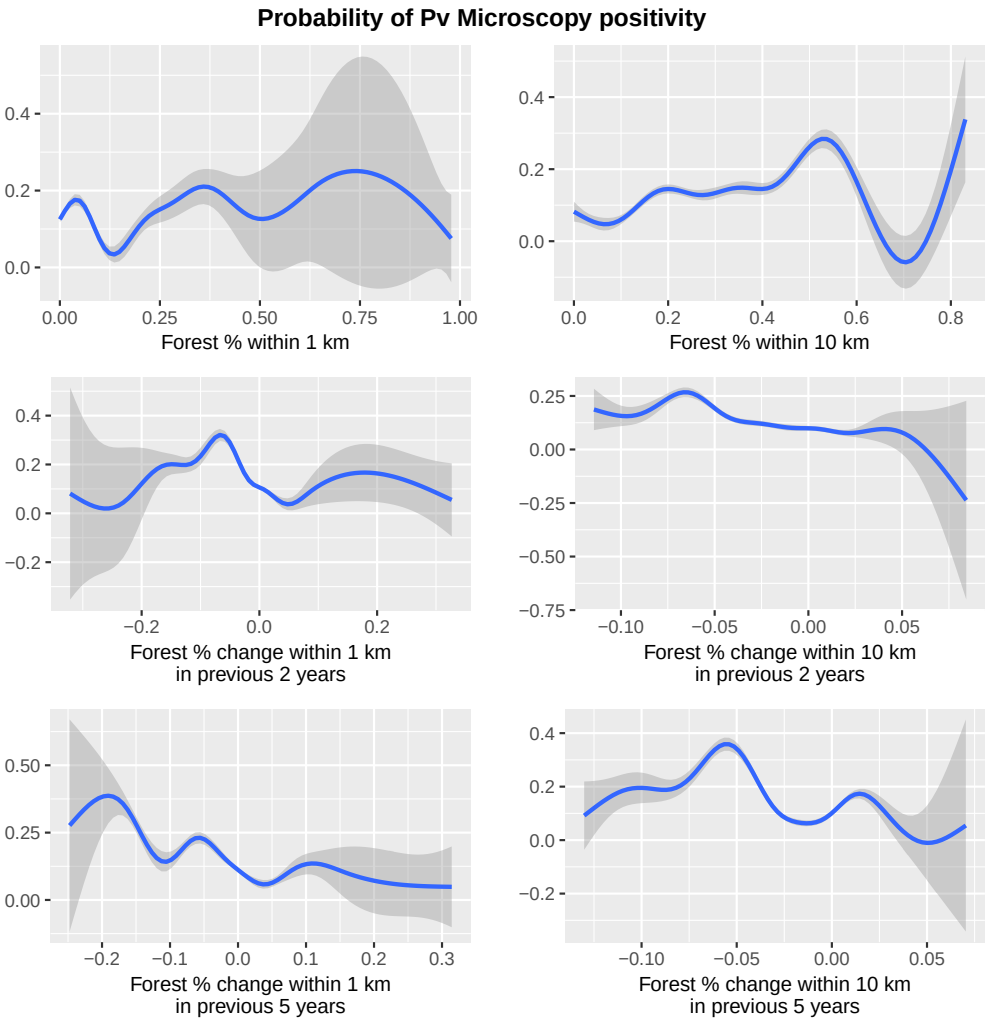

Figure 10: GAM smooth plots of association between forest and Pv Microscopy positivity.