Measuring Malaria in Complex Transmission Systems

IND E 519 Term Paper

Alec Georgoff

June 6, 2018

UNIVERSITY *of*
WASHINGTON

Contents

1 Background

1.1 The Ross-Macdonald Model

Basic malaria transmission has been modeled using the Ross-Macdonald (RM) model since the early 20^{th} century; it is the simplest basic model of transmission and is often used as the "theoretical scaffolding" of a malaria transmission model. The RM model depends on many ecological parameters, such as mosquito blood feeding rate and probability of infection upon receiving a bite from an infectious mosquito. The combination of these parameters for a specific location is summarized in the *basic reproductive number*, or R_0. This value gives a convenient threshold for transmission: if $R_0 < 1$, transmission will eventually reduce to zero over time; if $R_0 > 1$, transmission will be sustained and malaria will be endemic in the given area. This R_0 threshold is useful for modeling the effect of malaria interventions; any intervention that is modeled aims to reduce the R_0 value below 1 so that malaria transmission will die off. The R_0 value also provides a useful metric for comparing interventions against one another - if one intervention has a lower R_0 value, it will cause malaria transmission to die off quicker.

1.2 Forest Malaria

A unique malaria transmission landscape is present in the case of forest malaria, which is a problem in countries such as those that make up the Greater Mekong Subregion: Cambodia, Laos, Vietnam, Thailand, Myanmar, and China. In these areas, people live in villages that surround a forest and frequently travel into the forest for various activities, where they contract malaria and proceed to carry it back to the village. This landscape is especially unique because it involves multiple R_0 values, one for each village and one for the forest; therefore, interventions that aim to eliminate malaria transmission must account for their effect on all of the R_0 values in the system. This paper will analyze how basic reproductive numbers affect malaria transmission in the simple system of one forest and one village, corresponding to two R_0 values: $R_{0,F}$ and $R_{0,V}$. This unique transmission environment presents challenges to successful interventions, which need to take into account these specific features. For instance, deploying insecticide-treated bednets (ITNs) in the village may reduce $R_{0,V}$ to below 1;

however, if $R_{0,F}$ remains above 1, forestgoers can still bring back enough malaria to the villages to sustain transmission.

2 Objective

2.1 Mathematical Relationship Between $R_{0,V}$, $R_{0,F}$, and Village Malaria Prevalence

Starting with basic differential equations that govern the model, this paper will use dimensional analysis and equilibrium assumptions to simplify the mathematics and obtain a relationship between $R_{0,V}$, $R_{0,F}$, and malaria prevalence in the village.

Once a mathematical relationship between $R_{0,V}$, $R_{0,F}$, and village prevalence is derived, visualization tools will be utilized to show the relationship between the three values. In particular, a "boundary" will be sought, which is the curve composed of combinations of R_0 values that mark the transition between malaria elimination and malaria endemicity.

2.2 Policy Advice Using Real Data

The relationships presented in this paper will be useful in devising policy advice in malaria-endemic areas; however, for the results to be tractable, the inputs to the model must be based off of real data for the area being targeted for intervention. This data input strategy will be discussed more in the conclusion of the paper.

3 Methodology

3.1 Differential Equations

The model is based on four ordinary differential equations (ODEs) that describe the interactions between humans and mosquitoes in both the forest and the village. These

equations are based off of the Ross-Macdonald model. These ODEs are as follows:

- Humans in the forest:

$$\frac{dX_F}{dt} = \left(b(1-p)a_V \frac{Z_V}{H_F(1-p)+H_V} + bpa_F \frac{Z_F}{H_F}\right)(H_F - X_F) - rX_F \tag{1}$$

- Humans in the village:

$$\frac{dX_V}{dt} = \left(ba_V \frac{Z_V}{H_F(1-p)+H_V}\right)(H_V - X_V) - rX_V \tag{2}$$

- Mosquitoes in the forest:

$$\frac{dY_F}{dt} = a_F c \frac{X_F}{H_F}(V_F - Y_F) - g_F Y_F \tag{3}$$

- Mosquitoes in the village:

$$\frac{dY_V}{dt} = a_V c \left(\frac{X_F(1-p)+X_V}{H_F(1-p)+H_V}\right)(V_V - Y_V) - g_V Y_V \tag{4}$$

Note the following expressions for Z_F and Z_V:

$$Z_F = e^{-g_F n} Y_F$$

$$Z_V = e^{-g_V n} Y_V$$

3.2 Parameters

The following are the parameters that are used in the model:

Symbol	*Parameter*
b	Proportion of bites by infectious mosquitoes that cause an infection
r	Rate that humans recover from an infection
c	Proportion of mosquitoes infected after biting infectious human

n	Time for sporogonic cycle
a_F	Human blood feeding rate in forest
a_V	Human blood feeding rate in village
g_F	Per capita death rate of mosquitoes in forest
g_V	Per capita death rate of mosquitoes in village
H_F	Human population density in forest
H_V	Human population density in village
X_F	Number of infected forest goers
X_V	Number of infected villagers
p	Proportion of time forest goers spend in the forest
$1-p$	Proportion of time forest goers spend in the village
V_F	Vector population in forest
V_V	Vector population in village
Y_F	Number of *infected* vectors in forest
Y_V	Number of *infected* vectors in village
Z_F	Number of *infectious* vectors in forest
Z_V	Number of *infectious* vectors in village

Many of these parameters are determined by the specific landscape under study; most of them can be summarized by the R value:

$$R = \frac{Va^2bce^{-gn}}{Hgr} \tag{5}$$

3.3 Equilibrium Simplification

In order to simplify the equations, two assumptions are made: **equilibrium infected mosquito population** and **equilibrium infected human population**. In this context, *equilibrium* populations are those populations achieved at extremely large time scales; they approximate the population as $t \to \infty$. In an epidemiological sense, *equilibrium* populations are of interest because they are the long-term results of any interventions that take place.

3.3.1 Equilibrium Infected Mosquito Population

At equilibrium, the rate of change of the number of infected mosquitoes is zero:

$$\frac{dY_F}{dt} = 0 = a_F c \frac{X_F}{H_F}(V_F - Y_F) - g_F Y_F$$

$$Y_F = \frac{a_F c V_F X_F}{H_F g_F + a_F c X_F} \tag{6}$$

$$\frac{dY_V}{dt} = 0 = a_V c \left(\frac{X_F(1-p) + X_V}{H_F(1-p) + H_V}\right)(V_V - Y_V) - g_V Y_V$$

$$Y_V = \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p) + H_V) + a_V c((1-p)X_F + X_V)} \tag{7}$$

3.3.2 Equilibrium Infected Human Population

The equilibrium mosquito populations can be inserted into the infected human equations:

$$\begin{aligned}\frac{dX_F}{dt} = \Bigg(&\frac{b(1-p)a_V e^{-g_V n}}{H_F(1-p) + H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p) + H_V) + a_V c((1-p)X_F + X_V)} \\ &+ \frac{bpa_F e^{-g_F n}}{H_F} \cdot \frac{a_F c V_F X_F}{H_F g_F + a_F c X_F}\Bigg)(H_F - X_F) - rX_F\end{aligned} \tag{8}$$

$$\frac{dX_V}{dt} = \left(\frac{ba_V e^{-g_V n}}{H_F(1-p) + H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p) + H_V) + a_V c((1-p)X_F + X_V)}\right)(H_V - X_V) - rX_V \tag{9}$$

Before simplifying these equations, it is useful to introduce a few new parameters:

$$\chi_F = \frac{X_F}{H_F} \qquad \chi_V = \frac{(1-p)X_F + X_V}{(1-p)H_F + H_V}$$

$$R = \frac{Va^2bce^{-gn}}{Hgr} \qquad S = \frac{a}{g}$$

Forest Population At equilibrium, the rate of change of the number of infected humans is zero:

$$\frac{dX_F}{dt} = 0 = \left(\frac{b(1-p)a_V e^{-g_V n}}{H_F(1-p) + H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p) + H_V) + a_V c((1-p)X_F + X_V)} + \frac{bpa_F e^{-g_F n}}{H_F} \cdot \frac{a_F c V_F X_F}{H_F g_F + a_F c X_F} \right)(H_F - X_F) - rX_F \tag{10}$$

By rearranging and substituting parameters, the following relationship is derived (see appendix):

$$0 = \left(R_{0,V}(1-p)\frac{\chi_V}{1 + S_V c \chi_V} + R_{0,F} p \frac{\chi_F}{1 + S_F c \chi_F} \right)(H_F - X_F) - X_F \tag{11}$$

Equation (11) provides a relationship between the R_0 values and the number of infected humans (X_F and X_V).

Village Population At equilibrium, the rate of change of infected humans is zero:

$$\frac{dX_V}{dt} = 0 = \left(\frac{ba_V e^{-g_V n}}{H_F(1-p) + H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p) + H_V) + a_V c((1-p)X_F + X_V)} \right)(H_V - X_V) - rX_V \tag{12}$$

By rearranging and substituting parameters, the following relationship is derived (see appendix):

$$0 = \left(R_{0,V}(1-p)\frac{\chi_V}{1 + S_V c \chi_V} \right)(H_V - X_V) - X_V \tag{13}$$

Equation (13) provides a relationship between $R_{0,V}$ and the number of infected humans in the village (X_F).

3.4 Solving for Roots

Equations (11) and (13) provide a set of constraints for the relationship between $R_{0,F}$, $R_{0,V}$, X_F, and X_V. Because these equations equal zero at equilibrium, finding the roots of the equations at different values of $R_{0,F}$ and $R_{0,V}$ reveals the values of X_F and X_V at those R values.

3.4.1 R Code - Set Up Equations as Function

In order to solve for the roots of the equation, the equations need to be set up as a function in R:

```
###################################
#
# Set up the equations as a function
#
###################################

model <- function(X, R_0_v, R_0_f, H_v, H_f, S_v, S_f, c_val, p_val) {
  # X[1] = X_f
  # X[2] = X_v

  # convert to prevalence:
  chi_f = X[1] / H_f
  chi_v = ((1 - p_val) * X[1] + X[2]) / ((1 - p_val) * H_f + H_v)

  equation_village <- (R_0_v * (1 - p_val) * (chi_v / (1 + S_v * c_val * chi_v)))
                                                * (H_v - X[2]) - X[2]

  equation_forest <- (R_0_v * (1 - p_val) * chi_v / (1 + S_v * c_val * chi_v)
                                                + R_0_f * p_val * chi_f /
                        (1 + S_f * c_val * chi_f)) * (H_f - X[1]) - X[1]

  return(c(equation_village, equation_forest))
}
```

3.4.2 R Code - Solve for Roots

A custom R function is written to solve for the roots of the equations:

```
###################################
#
# Solve for roots
#
###################################

find_roots <- function(R_0_v, R_0_f,
                       H_v. = H_v, H_f. = H_f,
                       S_v. = S_v, S_f. = S_f,
                       c_val = c, p_val = p,
                       chi_v_start. = chi_v_start, chi_f_start. = chi_f_start) {

  # convert start point from prevalence to # of humans:
  X_v_start <- chi_v_start. * H_v
  X_f_start <- chi_f_start. * H_f

  # use multiroot solver to find roots:
  ss <- multiroot(f = model, start = c(X_v_start, X_f_start),
                  R_0_v = R_0_v, R_0_f = R_0_f,
                  H_v = H_v., H_f = H_f.,
                  S_v = S_v., S_f = S_f.,
                  c_val = c_val, p_val = p_val)

  # convert results to prevalence:
  chi_v_SS <- ss$root[1] / H_v
  chi_f_SS <- ss$root[2] / H_f

  return(c(chi_v_SS, chi_f_SS))
}
```

3.4.3 R Code - Cycle Through R Values

A set of nested *for* loops is constructed to cycle through all possible combinations of R values and determine the roots of the equations at those values:

```
###################################
#
# Cycle through R values
#
###################################

# set R values to cycle through:
R_0_v_values <- seq(0, 10, 0.01)
R_0_f_values <- seq(0, 10, 0.01)

# create data table to store results:
results <- data.table(R_0_v = rep(0, times = length(R_0_f_values) * length(R_0_v_values)),
                      R_0_f = 0, chi_v = 0, chi_f = 0)

i <- 1

for (v in R_0_v_values) {
  for (f in R_0_f_values) {
    # record current R values:
    results[i, R_0_v := v]
    results[i, R_0_f := f]
    # solve for roots at those R values:
    results[i, chi_v := find_roots(v, f)[1]]
    results[i, chi_f := find_roots(v, f)[2]]

    # print progress:
    cat("R_0_v =", v, ", R_0_f =", f, " \r", file = "", sep = " ")
    flush.console()

    i <- i + 1
  }
}
```

4 Results

By looping through a large set of combinations of $R_{0,V}$ and $R_{0,F}$, a landscape of prevalence values was created. Village prevalence was chosen as the focus, since the goal of any interventions is to eliminate malaria in the village. Two visualization tools were used: a heatmap and a 3D surface.

4.1 Heatmap

Figure 1 shows the distribution of village malaria prevalence as a function of $R_{0,V}$ and $R_{0,F}$.

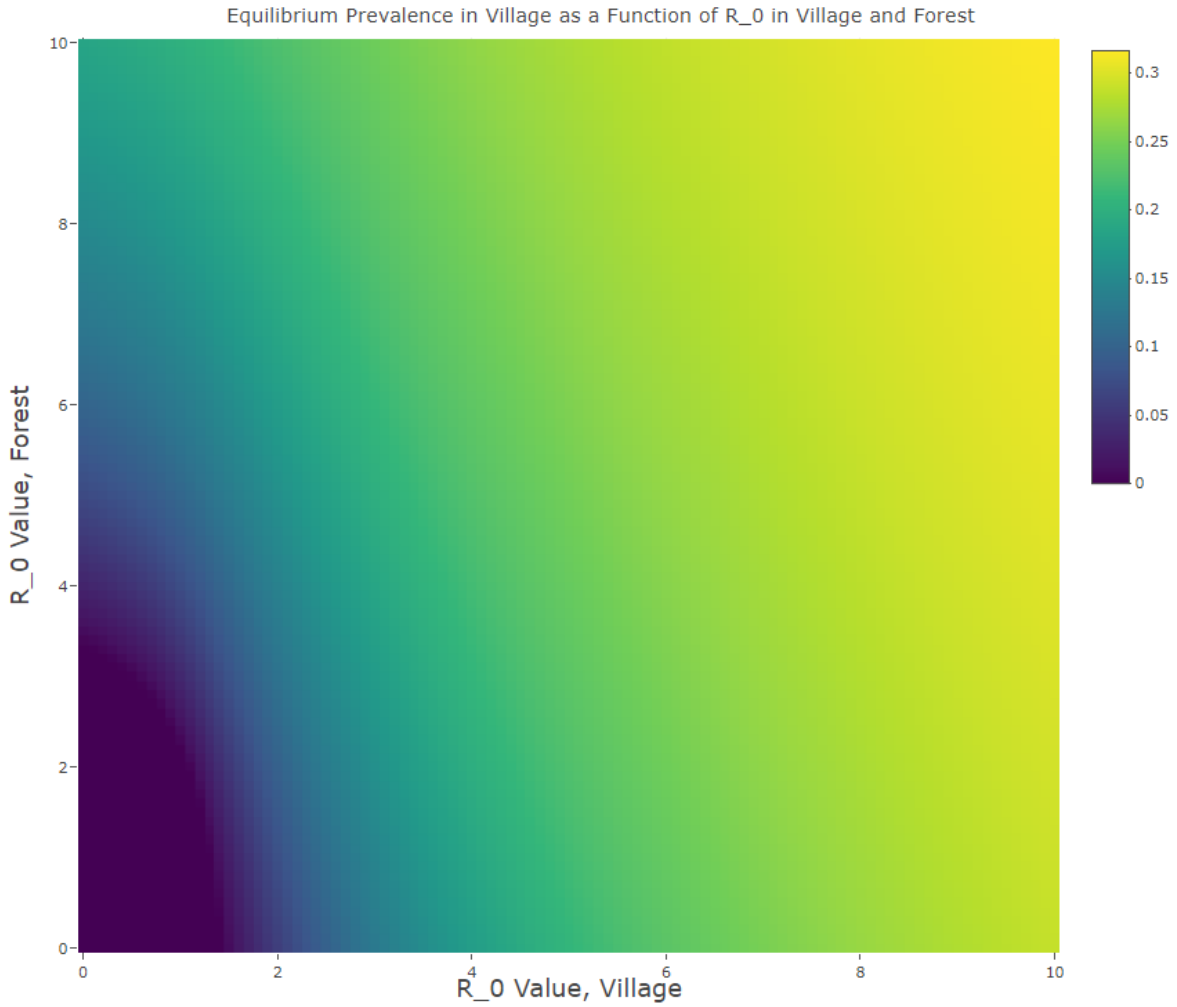

Figure 1. Heatmap of village malaria prevalence as a function of $R_{0,V}$ and $R_{0,F}$

As expected, the prevalence in the village increases monotonically with increasing R_0 values in either location. Interestingly, the region on the heatmap of zero malaria prevalence is larger than expected. Logic dictates that, in cases where either R_0 value is less than one, malaria will not be persistent. This is seen clearly in the results; however, the heatmap also shows regions where malaria is eliminated despite the fact that R_0 is greater than one in one (or both)

locations. For example, it is possible to have $R_{0,V} = 0.4$ and $R_{0,F} = 2.7$ and achieve zero malaria prevalence in the village.

4.2 3D Surface

Figure 2 shows a 3D representation of the village malaria prevalence as a function of $R_{0,V}$ and $R_{0,F}$. Note that this figure includes a smaller number of data points than the heatmap; this was done to improve the visualization of the surface

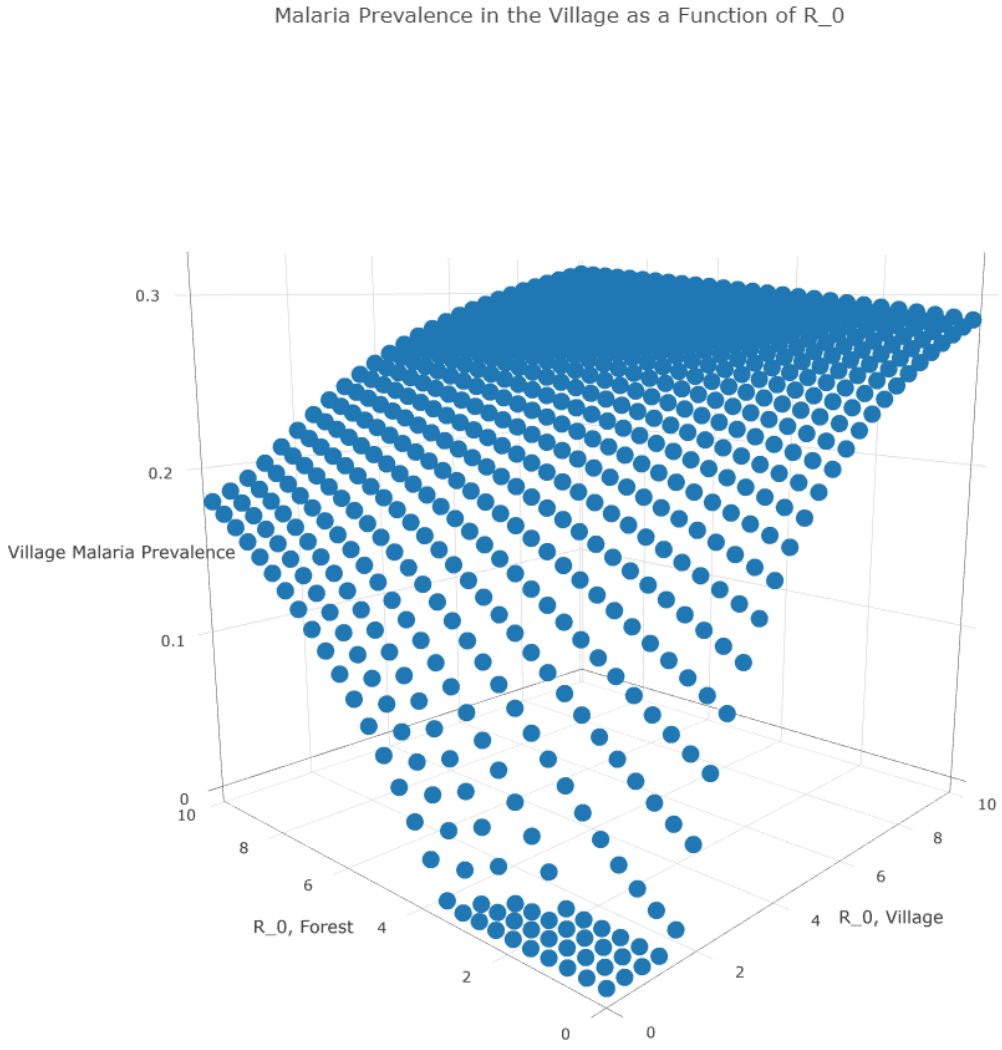

Figure 2. 3D surface of village malaria prevalence as a function of $R_{0,V}$ and $R_{0,F}$

The 3D surface reiterates what was shown in the heatmap; the village malaria prevalence rapidly crashes down in a "wave" as $R_{0,V}$ and $R_{0,F}$ decrease.

4.3 Interactive HTML Results

Both the heatmap and the 3D surface are available as interactive HTML pages at the following addresses:

Heatmap: `https://georgoff.github.io/IND_E_519/heatmap.html`

3D Surface: `https://georgoff.github.io/IND_E_519/3D_surface.html`

These HTML pages are fully interactive and allow the user to explore the results more in-depth; specific prevalence values can be displayed by hovering the mouse on the heatmap or surface. In addition, the 3D surface can be rotated in space.

5 Conclusion

5.1 Applicability of Results

This analysis is useful in describing how the values of $R_{0,V}$ and $R_{0,F}$ are related to the transmission of malaria, specifically in the village. When public health professionals are planning interventions, they can use the tools developed in this project to determine what R_0 values they should aim to achieve. These results show that both $R_{0,V}$ and $R_{0,F}$ must be addressed for any intervention to be tractable.

5.2 Future Directions

There are two main goals for improving the results of this research: further investigation into the region of the R_0 landscape where there is no malaria transmission, and the utilization of real-world input data for the model. In addition, expanding the model to multiple villages

and/or forests could yield interesting new results.

5.2.1 Investigate Region of Zero Malaria Transmission

As seen in the heatmap (Figure 1), there is a region within the R_0 landscape where malaria prevalence drops to zero. The boundary of this region is not yet well-defined. By setting a threshold value for malaria prevalence, it would be possible to create a binary value of malaria transmission. The boundary between these two regions would be of great interest; specifically, it would be useful to know how the boundary curve relates to the various input parameters.

5.2.2 Use Real Input Data

For these tools to be useful, the results must be based off of actual data; specifically, the parameters used in the model should be reasonable approximations of real-world values. For instance, in the Greater Mekong Subregion, it would be worthwhile to collect field data about the human and mosquito populations within the village(s) and forest(s) along with information about how much time people spend in the forest (the p value). These values could then be used when designing interventions for those locations.

5.2.3 Expand to Multiple Villages and/or Forests

To keep the math reasonable, this analysis used the simple case of one village and one forest. In the real world, this scenario is rather unlikely. It is more likely that there are multiple villages that surround one or more large forests; also, movement of humans between villages is certainly a possibility. Accounting for these factors in the model would require much more complex mathematics, but could be worthwhile for designing region-specific interventions.

6 Appendix

6.1 Mathematics Derivations

Determination of Equilibrium Mosquito Populations

Forest Mosquito Population:

$$\frac{dY_F}{dt} = 0 = a_F c \frac{X_F}{H_F}(V_F - Y_F) - g_F Y_F$$

$$a_F c \frac{X_F}{H_F} V_F = \left(g_F + a_F c \frac{X_F}{H_F} \right) Y_F$$

$$Y_F = \frac{a_F c V_F X_F}{H_F g_F + a_F c X_F}$$

Village Mosquito Population:

$$\frac{dY_V}{dt} = 0 = a_V c \left(\frac{X_F(1-p) + X_V}{H_F(1-p) + H_V} \right) (V_V - Y_V) - g_V Y_V$$

$$\left[\left(g_V + a_V c \frac{X_F(1-p) + X_V}{H_F(1-p) + H_V} \right) Y_V = a_V c \frac{X_F(1-p) + X_V}{H_F(1-p) + H_V} V_V \right] \frac{1}{a_V c} \frac{H_F(1-p) + H_V}{X_F(1-p) + X_V}$$

$$\left(\frac{(H_F(1-p) + H_V) g_V}{a_V c (X_F(1-p) + X_V)} + 1 \right) Y_V = V_V$$

$$Y_V = \frac{a_V c V_V ((1-p)X_F + X_V)}{g_V (H_F(1-p) + H_V) + a_V c ((1-p)X_F + X_V)}$$

Rearranging Equations to Involve R_0

Forest Population:

Since we're looking at equilibrium, we can set $\frac{dX_F}{dt}$ to 0:

$$0 = \left(\frac{b(1-p)a_V e^{-g_V n}}{H_F(1-p)+H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p)+H_V) + a_V c((1-p)X_F + X_V)} \right.$$
$$\left. + \frac{bpa_F e^{-g_F n}}{H_F} \cdot \frac{a_F c V_F X_F}{H_F g_F + a_F c X_F} \right)(H_F - X_F) - rX_F$$

Divide through by r:

$$0 = \left(\left(\frac{b(1-p)a_V e^{-g_V n}}{H_F(1-p)+H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p)+H_V) + a_V c((1-p)X_F + X_V)} \right) \frac{1}{r} \right.$$
$$\left. + \frac{bpa_F e^{-g_F n}}{rH_F} \cdot \frac{a_F c V_F X_F}{H_F g_F + a_F c X_F} \right)(H_F - X_F) - X_F$$

Now we need to introduce a variable for prevalence:

$$\chi_F = \frac{X_F}{H_F}$$

$$\chi_V = \frac{(1-p)X_F + X_V}{(1-p)H_F + H_V}$$

We also need to define additional parameters:

$$R = \frac{V a^2 b c e^{-gn}}{Hgr}$$

$$S = \frac{a}{g}$$

Let's focus on the first large term:

$$\left(\frac{b(1-p)a_V e^{-g_V n}}{H_F(1-p)+H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p)+H_V) + a_V c((1-p)X_F + X_V)} \right) \frac{1}{r} \tag{14}$$

$$= \frac{1}{r} \left(\frac{V_V {a_V}^2 b c e^{-g_v n}}{H_F(1-p)+H_V} \cdot \frac{g_V}{g_V}(1-p) \frac{\chi_V}{g_V + a_V c \chi_V} \right)$$

$$= R_{0,V}(1-p) \frac{g_V \chi_V}{g_V + a_V c \chi_V}$$

$$= R_{0,V}(1-p)\frac{\chi_V}{1+S_V c\chi_V}$$

And now the second large term:

$$\frac{bpa_F e^{-g_F n}}{rH_F} \cdot \frac{a_F c V_F X_F}{H_F g_F + a_F c X_F}$$

$$= \frac{V_F {a_F}^2 bce^{-g_F n}}{rH_F} \cdot \frac{pX_F}{H_F(g_F + \frac{a_F c}{H_F} X_F)}$$

$$= R_{0,F} p \frac{\frac{X_F}{H_F}}{1 + \frac{a_F c}{H_F g_F} X_F}$$

$$= R_{0,F} p \frac{\chi_F}{1 + S_F c \chi_F}$$

Putting the terms back together:

$$0 = \left(R_{0,V}(1-p)\frac{\chi_V}{1+S_V c\chi_V} + R_{0,F} p \frac{\chi_F}{1+S_F c\chi_F} \right)(H_F - X_F) - X_F$$

Village Population:

Since we're looking at equilibrium, we can set $\frac{dX_V}{dt}$ to 0:

$$0 = \left(\frac{b(1-p)a_V e^{-g_V n}}{H_F(1-p)+H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p)+H_V) + a_V c((1-p)X_F + X_V)} \right)(H_V - X_V) - rX_V$$

Divide through by r:

$$0 = \left(\left(\frac{b(1-p)a_V e^{-g_V n}}{H_F(1-p)+H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p)+H_V) + a_V c((1-p)X_F + X_V)} \right) \right) \frac{1}{r}(H_V - X_V) - X_V$$

Now we need to introduce a variable for prevalence:

$$\chi_F = \frac{X_F}{H_F}$$

$$\chi_V = \frac{(1-p)X_F + X_V}{(1-p)H_F + H_V}$$

We also need to define additional parameters:

$$R = \frac{Va^2bce^{-gn}}{Hgr}$$

$$S = \frac{a}{g}$$

Let's focus on the large term in parentheses:

$$\left(\frac{b(1-p)a_V e^{-g_V n}}{H_F(1-p)+H_V} \cdot \frac{a_V c V_V((1-p)X_F + X_V)}{g_V(H_F(1-p)+H_V) + a_V c((1-p)X_F + X_V)}\right)\frac{1}{r} \tag{15}$$

$$= \frac{1}{r}\left(\frac{V_V {a_V}^2 bce^{-g_v n}}{H_F(1-p)+H_V} \cdot \frac{g_V}{g_V}(1-p)\frac{\chi_V}{g_V + a_V c\chi_V}\right)$$

$$= R_{0,V}(1-p)\frac{g_V\chi_V}{g_V + a_V c\chi_V}$$

$$= R_{0,V}(1-p)\frac{\chi_V}{1 + S_V c\chi_V}$$

Putting the terms back together:

$$0 = \left(R_{0,V}(1-p)\frac{\chi_V}{1+S_V c\chi_V}\right)(H_V - X_V) - X_V$$

6.2 R Code

The R code used in this analysis can be downloaded at the following address:

`https://georgoff.github.io/IND_E_519/multiroot_solve.R`