Measuring Malaria in Complex Transmission Systems

A Time-at-Risk-Based Approach

Alec Georgoff

UNIVERSITY *of*
WASHINGTON

Contents

1 Background

1.1 Ross-Macdonald Equations

The Ross-Macdonald equations describe malaria transmission through both human and mosquito populations. They take the following form:

$$\frac{dX}{dt} = abe^{-gn}\frac{Y}{H}(H - X) - rX \tag{1}$$

$$\frac{dY}{dt} = ac\frac{X}{H}(V - Y) - gY \tag{2}$$

Equation 1 describes the dynamics of infected humans, while Equation 2 describes the dynamics of infected mosquitoes. The variables of interest are X and Y, which represent the number of infected humans and mosquitoes, respectively. The rest of the parameters are intrinsic to the system under study:

Parameter	*Meaning*
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	Human blood feeding rate
g	Per capita death rate of mosquitoes
V	Vector population
H	Human population

Note that, in Equation 1, an assumption is made about the equilibrium number of infectious mosquitoes Z:

$$Z = e^{-gn}Y$$

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 transmission levels, 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 transmission levels in the system.

1.3 Time-at-Risk

When considering multiple villages and/or forests, a key parameter to consider is travel between these areas. The differential equations we use require that each mobile population has its own set of equations describing its movement. For this analysis, we assume that each human lives in a village and spends some proportion of their time in forests and/or other villages. These proportions are represented by $\mathbf{\Psi}$, the Time-at-Risk (TaR) matrix. The TaR matrix takes the following form:

$$\mathbf{\Psi} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix}$$

where n is the total number of locations (villages and forests) and m is the total number of mobile populations. p_{ij} is the proportion of time that a person from location i spends in location j. Note that the proportions in each row must sum to one:

$$\sum_{j=1}^{n} p_{ij} = 1 \quad \text{for } i \in [1, m]$$

Because each village can be comprised of multiple mobile populations, it is possible (likely,

even) for a village to have multiple rows in the TaR matrix describing where each of its mobile populations spends its time.

2 Objective

The goal of this analysis is to determine how reproductive numbers affect malaria transmission in complex systems with multiple interconnected locations. This unique transmission environment presents challenges to successful interventions, which must take into account these specific features. For instance, deploying insecticide-treated bednets (ITNs) in all the villages may reduce their transmission levels below the endemic level; however, if the transmission in the forest remains high enough, forest-goers may still bring enough malaria back to the villages to sustain transmission there.

3 Simplification of Ross-Macdonald Equations

3.1 Equilibrium Assumptions

By assuming equilibrium populations of mosquitoes and humans, it is possible to reduce the system to one equation describing the number of infected humans.

Start by assuming equilibrium mosquito population:

$$\frac{dY}{dt} = 0 = ac\frac{X}{H}(V - Y) - gY$$

Then solve for Y:

$$Y = \frac{acVX}{Hg + acX}$$

Next, plug the equilibrium Y value into the equation for human population at equilibrium:

$$\frac{dX}{dt} = 0 = abe^{-gn} \cdot \frac{1}{H} \cdot \frac{acVX}{Hg + acX}(H - X) - rX \tag{3}$$

3.2 Simplified Parameters

Since most of the parameters in the Ross-Macdonald equations do not vary with time, it can be useful to combine them into one parameter. This reduces the dimensions of the equations, and can be accomplished by rearranging Equation 3.

First, group together constant parameters, divide everything by r, and multiply the first term by $\frac{1}{g}/\frac{1}{g}$:

$$0 = \frac{Va^2bce^{-gn}}{Hgr} \cdot \frac{X}{H + \frac{a}{g}cX}(H - X) - X \tag{4}$$

3.2.1 Stability Index

The first simplified parameter is the *stability index*, which is defined as

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

The stability index can be interpreted as the number of human bites per mosquito over its lifetime.

3.2.2 Reproductive Rate

The next parameter is the reproductive rate:

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

The reproductive rate makes sense intuitively. In the malaria transmission cycle, mosquito biting occurs twice, hence the a^2 factor. Transmission is helped by large numbers of mosquitoes per human host (large V/H) and by large b and c. Transmission is hindered by high mosquito death rates (large g) and by fast disease recovery (large r). Finally, quick sporogonic cycles (low n) produce mosquitoes more quickly, supporting transmission.

Introducing the stability index and the reproductive rate simplifies Equation 4:

$$0 = R\frac{X}{H + ScX}(H - X) - X \tag{5}$$

3.2.3 R as a Threshold

The R value in a system describes the average number of new cases that each case of malaria will cause. For example, in a system with an R value of 3, each case of malaria would cause, on average, 3 more cases to occur.

The use of R introduces a convenient threshold for malaria endemicity:

$$R\begin{cases} < 1 & \text{transmission dies out over time} \\ > 1 & \text{sustained endemic transmission} \end{cases}$$

This threshold can be used when planning malaria interventions; a successful intervention will reduce the R value below 1 and eliminate malaria. R can also be used when comparing interventions against one another. An intervention with a lower R value, assuming both values are below 1, will lead to elimination faster than interventions with higher R values.

4 Matrix Math

4.1 Converting Ross-Macdonald Variables to Vectors

Recall Equations 1 and 2:

$$\frac{dX}{dt} = abe^{-gn}\frac{Y}{H}(H - X) - rX$$

$$\frac{dY}{dt} = ac\frac{X}{H}(V - Y) - gY$$

In a system with multiple locations, each location must have both of these equations describing the infection of humans and mosquitoes. In order to consolidate notation, it is convenient to

describe the system using vectors: each variable is represented by a vector containing the value of that variable for each location or population in the system. We are assuming no mosquito movement between locations, only human movement; therefore, the variables for humans (H and X) must be treated differently than the variables for mosquitoes (V and Y).

4.1.1 Mosquito Variables

Because we assume that mosquitoes stay within one location, there only needs to be one value for V and Y in each location. For n locations, the vectors take the following form:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix}$$

4.1.2 Human Variables

Human populations move between multiple locations, and there needs to be a variable describing each unique mobile population. For m unique mobile populations, the vectors take the following form:

$$\mathbf{H} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_m \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$$

4.2 Matrix Notation

The use of matrices invokes the need for specific notation:

Matrix Product:

If $\mathbf{A}$ is an $n \times m$ matrix and $\mathbf{B}$ is an $m \times p$ matrix,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mp} \end{bmatrix}$$

the matrix product $\mathbf{C} = \mathbf{AB}$ is defined to be the $n \times p$ matrix

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{np} \end{bmatrix}$$

such that

$$c_{ij} = a_{i1}b_{1j} + \cdots + a_{im}b_{mj} = \sum_{k=1}^{m} a_{ik}b_{kj}$$

Hadamard Product (Elementwise Multiplication):

For two matrices, $\mathbf{A}$, $\mathbf{B}$, of the same dimension, $m \times n$, the Hadamard product, $\mathbf{A} \circ \mathbf{B}$, is a matrix, of the same dimension as the operands, with elements given by

$$(\mathbf{A} \circ \mathbf{B})_{i,j} = (\mathbf{A})_{i,j} (\mathbf{B})_{i,j}$$

Scalar Multiplication:

The scalar multiplication of a matrix $\mathbf{A}$ with a scalar λ gives another matrix $\lambda\mathbf{A}$ of the same size as $\mathbf{A}$. The entries of $\lambda\mathbf{A}$ are defined by

$$\lambda (\mathbf{A})_{ij} = (\lambda\mathbf{A})_{ij}$$

Transposition:

Transposing a matrix reflects it over its main diagonal, such that

$$\left[\mathbf{A^T}\right]_{ij} = [\mathbf{A}]_{ji}$$

If $\mathbf{A}$ is an $m \times n$ matrix, then $\mathbf{A^T}$ is an $n \times m$ matrix.

5 Ross-Macdonald Equations in Matrix Form

5.1 Scaling by Time-at-Risk

The Time-at-Risk (TaR) matrix $\mathbf{\Psi}$ from Section 1.3 can be used to scale the human populations (both infected and total populations). This is done by matrix multiplying the transpose of the TaR matrix by the respective population matrix:

$$\mathbf{X}_\Psi = \mathbf{\Psi^T X} \qquad\qquad \mathbf{H}_\Psi = \mathbf{\Psi^T H}$$

The act of scaling by $\mathbf{\Psi}$ accounts for the fact that different populations of humans spend different amounts of time in different places.

5.2 Mosquito Populations

Start with the Ross-Macdonald equation for mosquito populations (Equation 2), modified to use matrices:

$$\frac{d\mathbf{Y}}{dt} = ac\frac{\mathbf{X}_\Psi}{\mathbf{H}_\Psi} \circ (\mathbf{V} - \mathbf{Y}) - g\mathbf{Y}$$

At equilibrium, $\frac{d\mathbf{Y}}{dt} = 0$

$$0 = ac\frac{\mathbf{X}_\Psi}{\mathbf{H}_\Psi} \circ \mathbf{V} - \left(ac\frac{\mathbf{X}_\Psi}{\mathbf{H}_\Psi} + g\right) \circ \mathbf{Y}$$

Solve for $\mathbf{Y}$:

$$\mathbf{Y} = \frac{ac\frac{\mathbf{X}_\Psi}{\mathbf{H}_\Psi} \circ \mathbf{V}}{ac\frac{\mathbf{X}_\Psi}{\mathbf{H}_\Psi} + g} = \frac{ac\mathbf{X}_\Psi \circ \mathbf{V}}{ac\mathbf{X}_\Psi + \mathbf{H}_\Psi g} \tag{6}$$

5.3 Human Populations

Start with the Ross-Macdonald equation for human populations (Equation 1), modified to use matrices:

$$\frac{d\mathbf{X}}{dt} = abe^{-gn}\left(\mathbf{\Psi}\frac{\mathbf{Y}}{\mathbf{H}_\Psi}\right) \circ (\mathbf{H} - \mathbf{X}) - r\mathbf{X}$$

At equilibrium, $\frac{d\mathbf{X}}{dt} = 0$. By inserting Equation 6, we obtain the following:

$$0 = abe^{-gn}\left(\mathbf{\Psi}\frac{1}{\mathbf{H}_\Psi} \circ \frac{ac\mathbf{X}_\Psi \circ \mathbf{V}}{ac\mathbf{X}_\Psi + \mathbf{H}_\Psi g}\right) \circ (\mathbf{H} - \mathbf{X}) - r\mathbf{X}$$

Divide through by r and multiply the first term by g/g:

$$0 = \left(\mathbf{\Psi}\left(\frac{\mathbf{V}a^2bce^{-gn}}{\mathbf{H}_\Psi gr} \circ \frac{g\mathbf{X}_\Psi}{ac\mathbf{X}_\Psi + \mathbf{H}_\Psi g}\right)\right) \circ (\mathbf{H} - \mathbf{X}) - \mathbf{X}$$

We can now substitute in a matrix of R values:

$$\mathbf{R}^* = \frac{\mathbf{V}a^2bce^{-gn}}{\mathbf{H}_\Psi gr}$$

$$0 = \left(\mathbf{\Psi}\left(\mathbf{R}^* \circ \frac{g\mathbf{X}_\Psi}{ac\mathbf{X}_\Psi + \mathbf{H}_\Psi g}\right)\right) \circ (\mathbf{H} - \mathbf{X}) - \mathbf{X} \tag{7}$$

Note that Equation 7 is similar to Equation 5, except adapted for time-at-risk-scaling and matrices of variables.

We can also now introduce a term for prevalence:

$$\mathbf{\Theta}^* = \frac{\mathbf{X}_\Psi}{\mathbf{H}_\Psi}$$

$$0 = \left(\mathbf{\Psi}\left(\mathbf{R}^* \circ \frac{\mathbf{\Theta}^*}{cS\mathbf{\Theta}^* + 1}\right)\right) \circ (\mathbf{H} - \mathbf{X}) - \mathbf{X} \tag{8}$$

Note the $*$ affixed to the $\mathbf{R}$ and $\mathbf{\Theta}$ terms. This is used because our definitions of reproductive number and prevalence are different than directly scaling prevalence by time-at-risk:

$$\mathbf{R}^* \neq \mathbf{R}_\Psi$$

$$\mathbf{R}^* = \frac{\mathbf{V}a^2bce^{-gn}}{\mathbf{H}_\Psi gr} \qquad\qquad \mathbf{R}_\Psi = \mathbf{\Psi^T R}$$

$$\mathbf{\Theta}^* \neq \mathbf{\Theta}_\Psi$$

$$\mathbf{\Theta}^* = \frac{\mathbf{\Psi^T X}}{\mathbf{\Psi^T H}} \qquad\qquad \mathbf{\Theta}_\Psi = \mathbf{\Psi^T \Theta}$$

This discrepancy is further explained in Section 6.

Equation 8 is the governing equation for any system of villages and forests at equilibrium. By plugging in values in the $\mathbf{R}^*$ matrix, we can solve for prevalence at equilibrium; this establishes a link between R value and malaria prevalence in complex transmission systems.

6 Simple Example: One Village, One Forest

To show the advantage of using matrices for systems such as these, it is useful to look at a simple example. Given one village and one forest, what are the equations that describe transmission of malaria?

6.1 Establish Parameters & Variables

The following vectors represent the total human population $\mathbf{H}$, the population of infected humans $\mathbf{X}$, and the reproductive numbers $\mathbf{R}^*$ in each location:

$$\mathbf{H} = \begin{bmatrix} H_{V_{\text{only}}} \\ H_{V \leftrightarrow F} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} X_{V_{\text{only}}} \\ X_{V \leftrightarrow F} \end{bmatrix} \qquad \mathbf{R}^* = \begin{bmatrix} R_V^* \\ R_F^* \end{bmatrix}$$

Because there are two locations and two mobile populations in this model, the TaR matrix is 2×2. Permanent villagers (V_{only}) spend 100% of their time in the village and 0% of their time in the forest. Forest-going villagers ($V \leftrightarrow F$) spend $(p)(100\%)$ of their time in the forest and

$(1-p)(100\%)$ of their time in the village:

$$\mathbf{\Psi} = \begin{array}{c} \\ V_{only} \\ V \leftrightarrow F \end{array} \begin{array}{c} \begin{array}{cc} \text{V} & \text{F} \end{array} \\ \begin{pmatrix} 1 & 0 \\ 1-p & p \end{pmatrix} \end{array}$$

We can now use the TaR matrix to scale the human populations and number of infected humans:

$$\mathbf{H}_\Psi = \mathbf{\Psi^T H} = \begin{bmatrix} 1 & 1-p \\ 0 & p \end{bmatrix} \begin{bmatrix} H_{V_{\text{only}}} \\ H_{V \leftrightarrow F} \end{bmatrix} = \begin{bmatrix} H_{V_{\text{only}}} + (1-p) H_{V \leftrightarrow F} \\ p H_{V \leftrightarrow F} \end{bmatrix}$$

$$\mathbf{X}_\Psi = \mathbf{\Psi^T X} = \begin{bmatrix} 1 & 1-p \\ 0 & p \end{bmatrix} \begin{bmatrix} X_{V_{\text{only}}} \\ X_{V \leftrightarrow F} \end{bmatrix} = \begin{bmatrix} X_{V_{\text{only}}} + (1-p) X_{V \leftrightarrow F} \\ p X_{V \leftrightarrow F} \end{bmatrix}$$

Using these scaled values, we can calculate prevalence in the village and forest:

$$\mathbf{\Theta}^* = \frac{\mathbf{X}_\Psi}{\mathbf{H}_\Psi} = \begin{bmatrix} \Theta^*_{V_{\text{only}}} \\ \Theta^*_{V \leftrightarrow F} \end{bmatrix} = \begin{bmatrix} \dfrac{X_{V_{\text{only}}} + (1-p) X_{V \leftrightarrow F}}{H_{V_{\text{only}}} + (1-p) H_{V \leftrightarrow F}} \\ \dfrac{X_{V \leftrightarrow F}}{H_{V \leftrightarrow F}} \end{bmatrix}$$

We can use this example to show why $\mathbf{\Theta}^*$ is the correct way to scale prevalence:

$$\mathbf{\Theta}^* = \frac{\mathbf{\Psi^T X}}{\mathbf{\Psi^T H}} = \frac{\mathbf{X}_\Psi}{\mathbf{H}_\Psi} = \begin{bmatrix} \Theta^*_{V_{\text{only}}} \\ \Theta^*_{V \leftrightarrow F} \end{bmatrix} = \begin{bmatrix} \dfrac{X_{V_{\text{only}}} + (1-p) X_{V \leftrightarrow F}}{H_{V_{\text{only}}} + (1-p) H_{V \leftrightarrow F}} \\ \dfrac{X_{V \leftrightarrow F}}{H_{V \leftrightarrow F}} \end{bmatrix}$$

$$\mathbf{\Theta}_\Psi = \mathbf{\Psi^T \Theta} = \begin{bmatrix} 1 & 1-p \\ 0 & p \end{bmatrix} \begin{bmatrix} \dfrac{X_{V_{\text{only}}}}{H_{V_{\text{only}}}} \\ \dfrac{X_{V \leftrightarrow F}}{H_{V \leftrightarrow F}} \end{bmatrix} = \begin{bmatrix} \dfrac{X_{V_{\text{only}}}}{H_{V_{\text{only}}}} + (1-p) \dfrac{X_{V \leftrightarrow F}}{H_{V \leftrightarrow F}} \\ p \dfrac{X_{V \leftrightarrow F}}{H_{V \leftrightarrow F}} \end{bmatrix}$$

Directly scaling $\mathbf{\Theta}$ with $\mathbf{\Psi}$ results in an incorrect addition of fractions. For example, if $\frac{X_{V_{\text{only}}}}{H_{V_{\text{only}}}} = 0.6$ and $\frac{X_{V \leftrightarrow F}}{H_{V \leftrightarrow F}} = 0.9$, the resulting village prevalence would be 1.23 for $p = 0.3$.

6.2 Simplify Equations

Our governing equation:

$$0 = \left(\boldsymbol{\Psi}\left(\mathbf{R}^* \circ \frac{\boldsymbol{\Theta}^*}{cS\boldsymbol{\Theta}^* + 1}\right)\right) \circ (\mathbf{H} - \mathbf{X}) - \mathbf{X}$$

First, expand this term:

$$\frac{\boldsymbol{\Theta}^*}{cS\boldsymbol{\Theta}^* + 1} = \frac{\begin{bmatrix}\Theta^*_{V_{\text{only}}} \\ \Theta^*_{V\leftrightarrow F}\end{bmatrix}}{\begin{bmatrix}cS\Theta^*_{V_{\text{only}}} + 1 \\ cS\Theta^*_{V\leftrightarrow F} + 1\end{bmatrix}} = \begin{bmatrix}\dfrac{\Theta^*_{V_{\text{only}}}}{cS\Theta^*_{V_{\text{only}}} + 1} \\ \dfrac{\Theta^*_{V\leftrightarrow F}}{cS\Theta^*_{V\leftrightarrow F} + 1}\end{bmatrix}$$

Now include that in this bigger term:

$$\begin{aligned}\left(\boldsymbol{\Psi}\left(\mathbf{R}^* \circ \frac{\boldsymbol{\Theta}^*}{cS\boldsymbol{\Theta}^* + 1}\right)\right) &= \boldsymbol{\Psi}\begin{bmatrix}R^*_V \\ R^*_F\end{bmatrix} \circ \begin{bmatrix}\dfrac{\Theta^*_{V_{\text{only}}}}{cS\Theta^*_{V_{\text{only}}} + 1} \\ \dfrac{\Theta^*_{V\leftrightarrow F}}{cS\Theta^*_{V\leftrightarrow F} + 1}\end{bmatrix} \\ &= \begin{bmatrix}1 & 0 \\ 1-p & p\end{bmatrix}\begin{bmatrix}R^*_V \cdot \dfrac{\Theta^*_V}{cS\Theta^*_V + 1} \\ R^*_F \cdot \dfrac{\Theta^*_F}{cS\Theta^*_F + 1}\end{bmatrix} \\ &= \begin{bmatrix}R^*_V \cdot \dfrac{\Theta^*_V}{cS\Theta^*_V + 1} \\ (1-p)\cdot R^*_V \cdot \dfrac{\Theta^*_V}{cS\Theta^*_V + 1} + p \cdot R^*_F \cdot \dfrac{\Theta^*_F}{cS\Theta^*_F + 1}\end{bmatrix}\end{aligned}$$

Tie it all together:

$$0 = \left(\boldsymbol{\Psi}\left(\mathbf{R}^* \circ \frac{\boldsymbol{\Theta}^*}{cS\boldsymbol{\Theta}^* + 1}\right)\right) \circ (\mathbf{H} - \mathbf{X}) - \mathbf{X}$$

$$0 = \begin{bmatrix}R^*_V \cdot \dfrac{\Theta^*_V}{cS\Theta^*_V + 1} \\ (1-p)\cdot R^*_V \cdot \dfrac{\Theta^*_V}{cS\Theta^*_V + 1} + p \cdot R^*_F \cdot \dfrac{\Theta^*_F}{cS\Theta^*_F + 1}\end{bmatrix} \circ \begin{bmatrix}H_{V_{\text{only}}} - X_{V_{\text{only}}} \\ H_{V\leftrightarrow F} - X_{V\leftrightarrow F}\end{bmatrix} - \begin{bmatrix}X_{V_{\text{only}}} \\ X_{V\leftrightarrow F}\end{bmatrix}$$

$$0 = \begin{bmatrix} R_V^* \cdot \dfrac{\Theta_V^*}{cS\Theta_V^* + 1} \cdot \left(H_{V_{\text{only}}} - X_{V_{\text{only}}}\right) - X_{V_{\text{only}}} \\ \left((1-p) \cdot R_V^* \cdot \dfrac{\Theta_V^*}{cS\Theta_V^* + 1} + p \cdot R_F^* \cdot \dfrac{\Theta_F^*}{cS\Theta_F^* + 1}\right) \cdot (H_{V \leftrightarrow F} - X_{V \leftrightarrow F}) - X_{V \leftrightarrow F} \end{bmatrix}$$

This example shows that, for the simplest possible system (one village and one forest), the equations that need to be solved become quite cumbersome. This is why the implementation of matrices is advantageous; it ties all of the complexity into one equation.

6.3 Solve Over Many R Values

Now that the equations are written out, we can use a root solver in R to solve for the equilibrium prevalence in the village at different combinations of R values (see Figure 1).

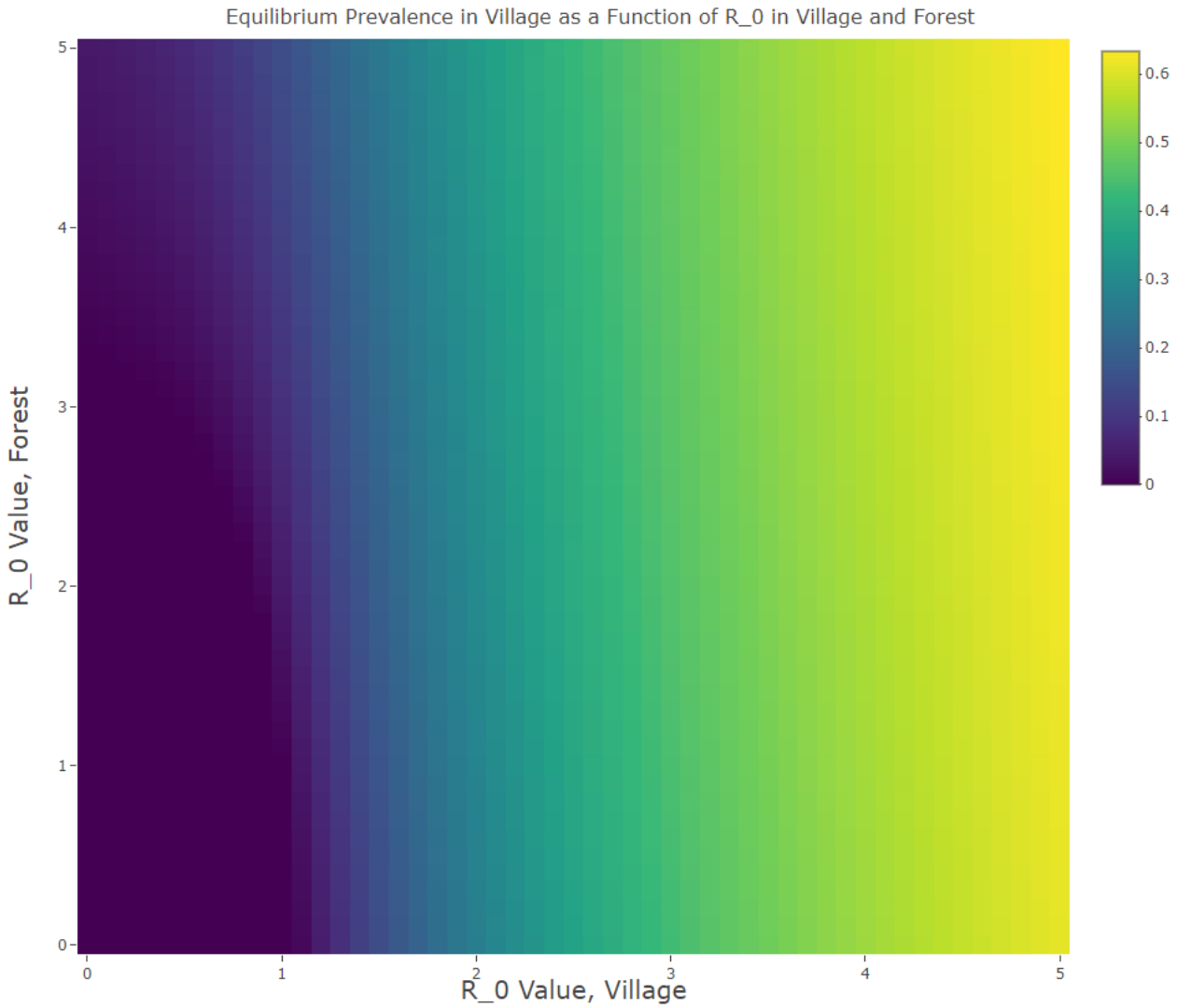

Figure 1. Heatmap of equilibrium village malaria prevalence at given R values ($p = 0.3$)

As expected, when R is less than one in both the village and the forest, there is zero prevalence of malaria in the village. As the R value increases above one in either location, the prevalence of malaria in the village steadily increases; however, the rate of increase is much higher when the R value increases in the village. This is reasonable since increased village transmission has a more direct impact on village prevalence compared to forest transmission. Increasing the R value in the forest has less of an impact on village prevalence because people only spend a fraction of their time in the forest.

7 Expand to More Villages and Forests

As seen in Section 6, vectors and matrices save us from dealing with an absurd amount of complex equations. To show how these vectors and matrices are filled out, let us consider a more complex example with four villages and two forests.

7.1 Human Population Vector

The vector of human populations takes the following form:

$$\mathbf{H} = \begin{bmatrix} 2000 \\ 5000 \\ 2000 \\ 5000 \\ 2000 \\ 5000 \\ 300 \\ 1000 \end{bmatrix} \begin{matrix} \text{V1} \leftrightarrow \text{F} \\ \text{V1 only} \\ \text{V2} \leftrightarrow \text{F} \\ \text{V2 only} \\ \text{V3} \leftrightarrow \text{F} \\ \text{V3 only} \\ \text{V4} \leftrightarrow \text{F} \\ \text{V4 only} \end{matrix}$$

Each of the four villages has two human populations: humans that travel to forests and humans that stay within the village.

7.2 TaR Matrix

The TaR matrix takes the following form:

$$\boldsymbol{\Psi} = \begin{array}{c} \\ \text{V1} \leftrightarrow \text{F} \\ \text{V1 only} \\ \text{V2} \leftrightarrow \text{F} \\ \text{V2 only} \\ \text{V3} \leftrightarrow \text{F} \\ \text{V3 only} \\ \text{V4} \leftrightarrow \text{F} \\ \text{V4 only} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{V1} & \text{V2} & \text{V3} & \text{V4} & \text{F1} & \text{F2} \end{array} \\ \begin{bmatrix} 0.3 & 0 & 0 & 0 & 0.4 & 0.3 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 & 0 & 0.4 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 & 0.2 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

This example shows the flexibility of the TaR matrix. To determine where a population of humans spends their time, we just have to find the row corresponding to that population. For example, the forest-goers from Village 1 spend 30% of their time in the village, 40% in Forest 1, and 30% in Forest 2.

8 R Code

Input Files

In order to provide a straightforward means of specifying parameters, the code accepts .csv files to create $\boldsymbol{\Psi}$ and $\mathbf{H}$. The files can easily be created or edited in Excel, and take the following forms (values taken from Section 7):

H

id	H
v1_f	2000
v1_v	5000
v2_f	2000
v2_v	5000
v3_f	2000
v3_v	5000
v4_f	300
v4_v	1000

Ψ

id	v1	v2	v3	v4	f1	f2
v1_f	0.3	0	0	0	0.4	0.3
v1_v	1	0	0	0	0	0
v2_f	0	0.6	0	0	0	0.4
v2_v	0	1	0	0	0	0
v3_f	0	0	0.8	0	0	0.2
v3_v	0	0	1	0	0	0
v4_f	0	0	0	0.2	0.8	0
v4_v	0	0	0	1	0	0

Set Parameters

The user must specify where the parameter files are saved and set the constant parameters of the model. Currently the code finds solutions across the same range of R values for every location, but this could change in the future to allow more flexibility. The user can specify if they want a PDF output of the prevalence graphs and set the filepath.

```r
# specify filepaths for parameter .csv files:

params_path <- "/homes/georgoff/georgoff.github.io/forest_malaria/params.csv"
psi_path <- "/homes/georgoff/georgoff.github.io/forest_malaria/psi.csv"

# set R values to cycle over:

R_min <- 0
R_max <- 2
R_step <- 0.5

# PDF output settings:
output_PDF <- TRUE
pdf_filepath <- "/homes/georgoff/georgoff.github.io/forest_malaria/test.pdf"

# this script assumes that the following parameters are the same for every
# location; in reality this may not be accurate. an update may be made that
# allows for custom parameters in every location

a <- 0.88   # human blood feeding rate
b <- 0.55   # proportion of bites by infectious mosquitoes that cause an infection
c <- 0.15   # proportion of mosquitoes infected after biting infectious human
g <- 0.1    # per capita death rate of mosquitoes
r <- 1/200  # rate that humans recover from an infection
n <- 12     # time for sporogonic cycle
S <- a/g    # stability index
```

Establish Vectors & Matrices of Variables

Once the parameters are read in, they need to be organized into vectors ($\mathbf{H}$, $\mathbf{X}$) and matrices ($\mathbf{\Psi}$). $\mathbf{H}$ and $\mathbf{X}$ also need to be scaled to create $\mathbf{H}_\Psi$ and $\mathbf{X}_\Psi$.

```r
# read in village and forest parameters from .csv file:

params <- as.data.table(read.csv(params_path))

n_villages <- nrow(params)

H <- params$H
X <- vector(mode = "numeric", length = length(H))
```

```
Psi <- as.data.table(read.csv(psi_path))
Psi[, id := NULL]
Psi <- as.matrix(Psi)
Psi_dt <- as.data.table(Psi)

locs <- names(Psi_dt)

H_psi <- t(Psi) %*% H
X_psi <- t(Psi) %*% X

# choose starting point for root solver:
theta_start <- vector(mode = "numeric", length = length(H))
theta_start[1:length(theta_start)] <- 0.9

# convert to number of humans:
X_start <- theta_start * H
```

Set Up Equations as a Function

Once all the variables are set up, they can be used to form a vector of equations that need to be solved. The root solver requires a function that returns this vector.

```
model <- function(X, Psi, R, c_val, S_val, H) {

  theta_psi <- (t(Psi) %*% X) / (t(Psi) %*% H)

  equation_vector <- (Psi %*% (R * (theta_psi/(c_val*S_val*theta_psi + 1)))) *
    (H-X) - X

  return(equation_vector)

}
```

Solve For Roots

A custom function is created that accepts a vector of R values and returns the solutions to the equations at those R values.

```
find_roots <- function(R,
                       Psi. = Psi,
                       H. = H,
                       S. = S,
                       c_val = c,
                       X_start. = X_start) {

  # use multiroot solver to find roots:
  ss <- multiroot(f = model, start = X_start.,
                  positive = TRUE, maxiter = 1000,
                  ctol = 1e-20,
                  Psi = Psi.,
                  R = R,
                  c_val = c_val,
                  S_val = S.,
                  H = H.)

  return(ss)
}
```

Set Up Results Table

A table is set up to hold all of the results. It includes every possible combination of R values for all the locations.

```
all_R_values <- seq(R_min, R_max, R_step)

list_of_R_values <- list(NULL)

for (i in 1:length(locs)) {
  list_of_R_values[[i]] <- all_R_values

  names(list_of_R_values)[i] <- locs[i]
}

# fill results table with every possible combination of
# R values:

results <- as.data.table(expand.grid(list_of_R_values))

# put in placeholder for theta values:
```

```
theta_holder <- as.data.table(matrix(data = 0, nrow = nrow(results),
                                     ncol = length(locs)))

for (k in 1:length(locs)) {
  names(theta_holder)[k] <- paste0("theta_", locs[k])
}

results <- cbind(results, theta_holder)
```

Cycle Through R Values

A `for` loop is created to loop through every row of the results table and determine the solutions at the specified R values.

```
for (i in 1:nrow(results)) {
  cat("Working on ", i, " of ", nrow(results), "\n")

  these_R_values <- unlist(results[i, 1:length(locs)], use.names = FALSE)

  X_solutions <- find_roots(these_R_values)$root

  theta_solutions <- (t(Psi) %*% X_solutions) / H_psi

  for (j in (1 + length(locs)):ncol(results)) {
    results[i, j] <- theta_solutions[j - length(locs)]
  }
}
```

Create PDF of Results

A PDF can be created that plots a simple bar graph of prevalence values at all locations.

```
if (output_PDF) {
  pdf(pdf_filepath)

  this_row_locs <- vector(mode = "character", length = length(locs))
```

```
  for (j in 1:nrow(results)) {
    for (location in 1:length(locs)) {
      this_row_locs[location] <- paste0(locs[location],
                                        "\nR = ",
                                        as.character(results[j, ..location]))
    }

    this_row <- data.table(loc = this_row_locs,
                           theta = unlist(results[j, (1+length(locs)):ncol(results)],
                                          use.names = F))

    bar <- ggplot(data = this_row,
                  aes(x = loc, y = theta)) +
      geom_col() +
      geom_text(aes(label = round(theta, 3), y = theta + 0.02)) +
      coord_cartesian(ylim = c(0,0.31))

    print(bar)
  }

  dev.off()
}
```

The PDF contains a graph for each possible combination of R values and displays the prevalence value in each location, as seen in Figure 2.

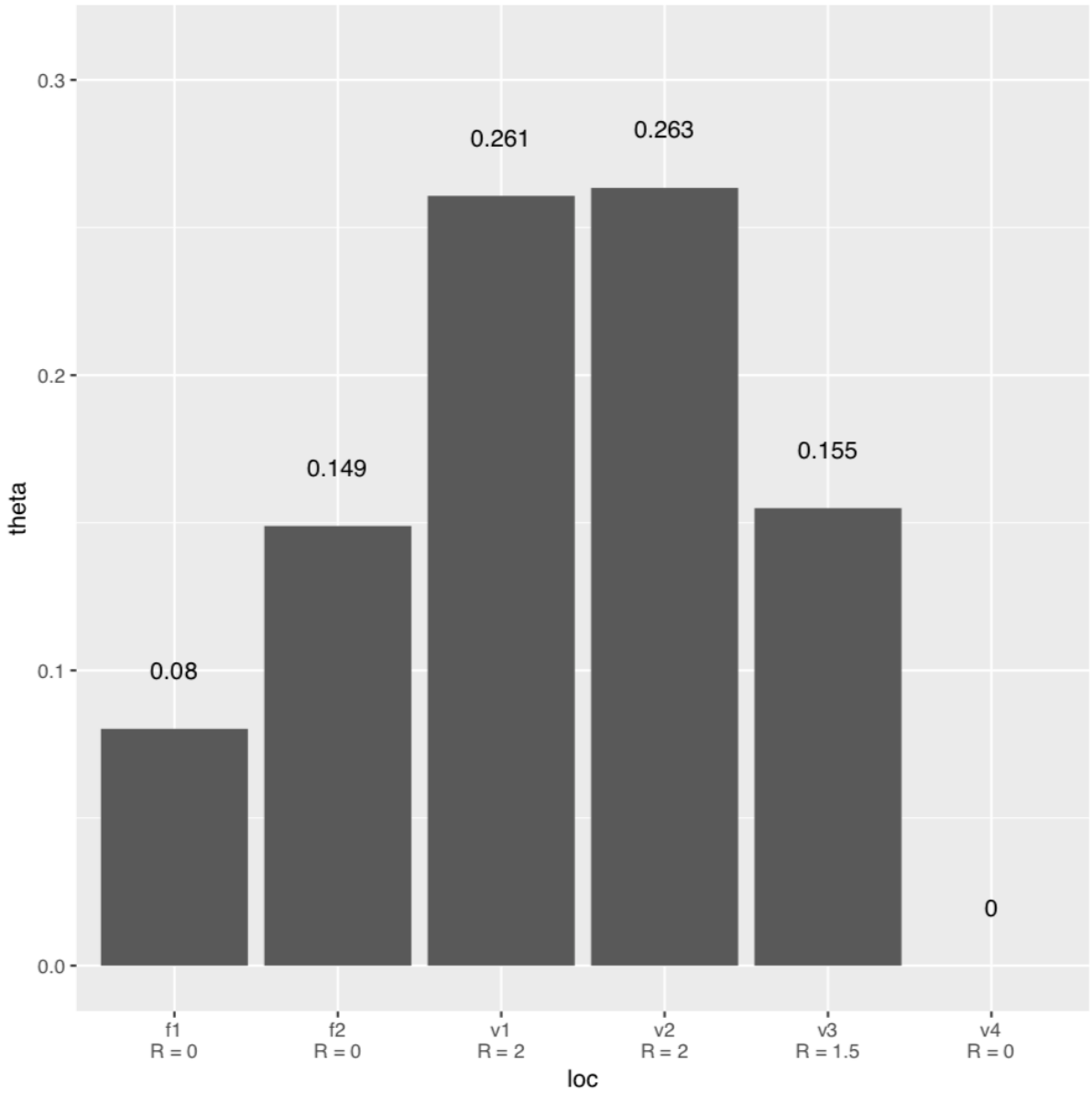

Figure 2. Graph of prevalence values for each location at a given R value