This file describes the six datasets, six R scripts, and three NetLogo agent-based model (ABM) code used to run the analysis for “Nonrandom foraging and resource distributions affect the relationships between host density, contact rates, and parasite transmission”. Netlogo was run using version 6.1.1 and R version 4.0.0 was used to analyze all the data.

**Authors**: Zachary Gajewski1, Philip McEmurray2,3, Jeremy Wojdak4, Cari McGregor4, Lily Zeller1, Hannah Cooper1, Lisa K. Belden2, Skylar Hopkins1

1. Department of Applied Ecology, North Carolina State University, Raleigh, NC
2. Department of Biological Sciences, Virginia Tech, Blacksburg, VA
3. Department of Anthropology, Washington University in St. Louis, St. Louis, MO
4. Department of Biology, Radford University, Radford, VA

**Corresponding Author:** Zachary Gajewski, [gajewskizach@gmail.com](mailto:gajewskizach@gmail.com)

**Experimental Study Design**

To understand how resource distributions influenced host contacts were performed an 8x7 factorial microcosm experiment that varied snail density (2,4,6,8,10,12,14, or 16 snails) and periphyton resource distributions (ranging from uniformly distributed to completely clustered). Periphyton was grown on tiles and arranging these tiles along with empty tiles allowed for the creation of different resource distributions in the microcosm. In each of the microcosm experiments, snails were uniquely colored and visually observed for 45 minutes. During that 45 minutes, all snail contacts, durations of those contacts, number of patches visited, and durations of those visits were recorded. Each treatment was replicated twice and all treatments were videoed to be able to check observations. An agent-based model (ABM), described below, was built based on the microcosm experiments and used to examine how host foraging behavior and resource distribution influence host contact rates, and thus, disease dynamics. Using Bayesian methods, we fit Holling type II functions to the ABM and experimental contact data. All data files, R scripts used for analysis, and the NetLogo files for the ABM used in this study are described below.

**NetLogo v6.1.1 Files**

**SnailModel.nlogo**

This is an agent-based model based on microcosm experiments conducted with an aquatic snail that links individual-level behavior with population-level processes, focusing on how foraging behaviors and resource distribution affect contact rates between individuals in a population.

The agent-based model consists of patches arranged in a 12 by 12 grid, and each patch has a state variable representing the biomass of the algae on the patch. The model allows for seven different algae distributions, ranging from “0” or completely uniform to increasingly clustered to completely clustered at “6”. Agents, representing snails, move around during the agent-based model simulations and are allowed to forage completely randomly, forage optimally, or somewhere in between those two settings. Foraging behavior is controlled by stochasticity, where 0 indicates optimal foraging and 100 indicates random foraging. Optimal foraging is defined by agents staying on a resource patch until it drops below a certain biomass and then they would randomly search for a new patch.

Before starting the simulation, parameters specific to patches and agents can be adjusted. The number of agents included in the simulation can be adjusted by using the number-of-snails slider. The snail-feeding-rate is the amount of algae a snail consumes per time step and can be adjusted by typing a new number in the text box. GUD represents giving up density or the amount of resources that are left on a tile when a snail will move off the tile when foraging optimally and can be adjusted with the slider. Lastly, the agents’ foraging behavior can be set using the stochasticity slider. A stochasticity of 0 indicates that the agents will forage completely optimally, while a stochasticity of 100 means agents will forage completely randomly.

The patch parameters that can be adjusted include initial-algae-coverage and algae-distribution. The algae-distribution box allows the user to enter the aggregation index of 0,1,2,3,4,5, or 6. Resources begin uniformly spaced at an aggregation index of 0 and begin to cluster as aggregation index increases, until they are completely clustered at an aggregation index of 6. The initial-algae-coverage determines the total amount of algae at the beginning of the experiment, and can be adjusted with the slider.

Lastly, simple disease dynamics in the simulation can be adjusted with the disease parameters. The number of starting infected agents can be adjusted with the starting-infections slider. Additionally, how infectious the pathogen is can be increased by increasing the infectiousness slider. Lastly, the recover-time, or time that it takes for a snail to recover from the infection can be adjusted with the slider.

The model outputs the average number of contacts that agents made, the duration of those contacts, the average number of patches visited, the average patch time, the average proportion of resources left, and the average edge time.

When all parameters are set, the simulation is started by using the SETUP button to set up the agents and patches and the GO button starts the simulation.

**SnailModel\_PatchRegeneration.nlogo**

This is an agent-based model that is the same as SnailModel.nlogo, described above, except resource patches were able to regenerate to full after hosts moved off them. This model was created to ensure that the loss of resources was not influencing host contacts. These model results are shown in the supplemental material of the paper.

**SnailModel\_NoPatchLimits.nlogo**

This is an agent-based model that is the same as SnailModel.nlogo, described above, except that this model removes the limits placed on the number of snails allowed on a patch. Therefore, this model allows any number of snails to be on a patch. This model was created to examine how the number of hosts allowed on a patch influenced contact rates. This model results are shown in the supplemental material of the paper.

**R v4.0.0 Script Files**

**FormatNetlogoCode.R**

This file uses

* Output from Netlogo with the first seven rows removed. Netlogo output can be created using any of the .nlogo files described above.
* nlogo\_cleaner.R, this R script file contains the function used in this manuscript to transform the Netlogo output from horizontal format to vertical format.

This R script prepares simulation outputs from NetLogo for analysis in R. Output from NetLogo with the first seven rows removed is converted from a horizontal format to a vertical format. For larger files, the terminal (macOS) or command prompt (window) can be used to remove the first seven rows. This is done by loading in the NetLogo cleaning function, found in nlogo\_cleaner.R. This code outputs OutputABMSimFinal.csv, used in ModelFits.R.

**nlogo\_cleaner.R**

This file uses

* Output from Netlogo with the first seven rows removed. Netlogo output can be created using any of the .nlogo files described above.

This function prepares simulation outputs from NetLogo for analysis in R. The function takes the data and the number of different types of rows and outputs, given by the user in the function, and transforms the data from horizontal to vertical. Additionally, this function takes the setting rows and adds them into the settings to the vertical format.

**ModelFits.R**

Uses the following data files:

* SnailExpData.csv, this file contains the experimental count data used to fit the Holling Type II function.
* OutputABMSimFinal.csv, this file contains all of the agent-based model simulation data and settings. This file is output from the FormatNetlogoCode.R file.
* ExpSnailDurationContactData.csv, this file contains the contact duration data from the microcosm experiments and is used to find the average contact duration used in the experimental Holling Type II model fits.

In this script file, we use Bayesian methods to fit Hollings type II functions to the experimental and ABM contact data across host density. One Hollings type II function is fit to the experimental data, while 6 models are fit to the ABM data, one per stochasticity level. This script also uses the ABM file to calculate R0 values. Lastly, this R script creates main figures 2 and 3, and supplemental figure 7.

**Exp\_ABM\_PatchComparison.R**

Uses the following data files:

* OutputABMSimFinal.csv, this file contains all of the agent-based model simulation data and settings.
* ExpPatchData.csv, this file contains all of the experimental patch information data, including number of patches visited and patch time.

In this script file, we compare experimental and ABM data to ensure that the ABM is producing realistic results. We compare the patches visited and patch duration time between the ABM and experiment. Additionally, this file makes figure 4.

**CleaningGUD\_Data.R**

Uses the following data files:

* GUD\_Data.csv, this file contains experimental data on the amount of algae left on a tile when a snail left a tile.
* TankInfo\_Data.csv, this file contains general information about the microcosm experiment setups.

In this script, we read in data collected from ImageJ and microcosm experiment videos, which have the area of algae left on a tile when a snail leaves. This script uses that data and calculates giving up density and feeding rates, both are used in the agent-based model. Additionally, this script is used to make supplemental figure 1.

**ResourceRegrowthComparison.R**

Uses the following data files:

* Output data from the SnailModel\_PatchRegeneration.nlogo file. These data are not included as a data file due to the large size, but you can create it by running the SnailModel\_PatchRegeneration.nlogo and cleaning the output with FormatNetlogoCode.R.
* ABM model fits created in the ModelFit.R file

In the Netlogo model used in the main text of the manuscript, resource did not regrow after being consumed. In the supplement and this script file, we examine whether the loss of resources within the timeframe of the ABM (50 ticks) impacted host contacts, comparing simulations with and without resource replenishment. This script fits cleans output from the FormatNetlogoCode.R, from both theSnailModel.nlogo file and SnailModel\_PatchRegeneration.nlogo file, by examining how many resource patches drop below the giving up density and fit the contact rate model for each output. This script file was used to create supplemental figures 4 and 5.

**Data Files**

**OutputABMSimFinal.csv**

File Description:

This data file is cleaned output from the agent-based model. We ran an 8x7x6 factorial experiment, each with 150 replicates (N = 50400 simulations). There were eight agent densities (8, 16, 24, 32, 40, 48, 56, and 64), seven resource distributions, and six levels of movement stochasticity ranging from optimal foraging to random foraging (0, 8, 18, 30, 50, and 100). This data is originally given in wide format and is rearranged into long format in this data file. This file contains columns for all the settings used in the ABM and observed variables in the simulation. Each row is a timestep in one of the simulations.

Column Description:

*start\_infection:* This column shows the number of agents that started the simulation as infected. All of our simulations started with 1 infected agent, but this number could be changed by others who use the code for different purposes. This column is a setting and will remain constant.

*Feeding:* This shows the feeding rate of the agents and is set at the start of the simulation. This feeding rate is constant across treatments and simulations and based on our microcosm data, but this constant could be changed by others who use the code for different purposes.

*algae\_conc:* This column shows the starting concentration of algae on an algae patch. This is set at the start of the simulation. This is constant across treatments and simulations, but this constant could be changed by others who use the code for different purposes.

*Snails:* This column shows the agent density in the simulation. This column was 8,16,24,32,40,48,56, or 64. These were four times the amount of densities used in the microcosm experiments because the ABM used four times the number of patches.

*stoch:* This column indicates the stochasticity of the agents’ behavior in the simulation. A lower stochasticity represents more random foraging (0 is completely random), while a higher stochasticity represents more optimal foraging. In our simulation, we compared 0, 8, 18, 30, 50 or 100.

*Infect:* This column represents the constant infectiousness probability, where the higher the value, the more likely uninfected agents are to become infected when making contact with an infected agent. This was constant across treatments and simulations, but this constant could be changed by others who use the code for different purposes.

*Algae:* This column shows the aggregation index. The aggregation index is a numeric variable describing the aggregation of resources. An aggregation index of 0 indicates that the resources are uniformly spaced. As aggregation index increases, the clustering of resources also increases, and for distributions 4, 5, 6, distance to the nearest resource patch also increases with clustering.

*recover:* The recover time is set at the start of the simulation and is the amount of time it takes infected agents to become healthy again. This was constant across treatments and simulations, but this constant could be changed by others who use the code for different purposes.

*GUD:* This column shows the set giving up density, or the amount of resources left on a patch when an agent leaves, when the agent is optimally foraging. This value is set at the start of the simulation and was based on experimental data. This was constant across treatments and simulations, but this constant could be changed by others who use the code for different purposes.

*resources:* This column shows the average amount of resources left on resource patches at the end of the simulation. These data were used to compare resource depletion rates in the ABM and the experiment.

*edge:* This column shows the average amount of time or ticks that agents spent on edge patches.

*patch\_time:* This column shows the average amount of time or ticks that agents spent on a patch. This variable is also called “patch duration” in the paper.

*Ss:* The number of susceptible agents in the simulation at Time 0. This variable was set as the starting number of agents minus 1 individual (the one infected individual).

*Is:* This column shows the number of infected agents at Time 0. All of our simulations started with 1 infected agent, but this number could be changed by others who use the code for different purposes

*Rs:* This column shows the number of recovered agents at Time 0. All of our simulations started with 0 recovered agents, but this number could be changed by others who use the code for different purposes.

*contacts:* This column gives the average number of contacts in the simulation at a given tick. A contact in the simulation is defined as two agents being on the same patch at the same time.

*total\_contacts:* This column shows the total number of contacts made in the simulation across all agents and all ticks.

*patches:* This column shows the average number of patches that agents visited.

*unique\_contacts:* This column represents the number of unique contacts made in the simulation. If two agents make contact twice this column will only show 1, because the contact is only unique the first time.

*contact\_dur:* This column shows the average contact duration across all the contacts in the simulation.

*ticks:* This column shows the number of ticks or time steps in the simulation. For our purposes, ticks represented minutes. In our simulations, ticks always went from 0 (start of the simulation) to 50.

**ResourceDistributionExperiment\_ContactData.csv**

File Description:

This data file contains all the contact durations made between snails in the microcosm experiments. This dataset contains the contact start time, end time, and total duration for each contact in each microcosm replicate. Each row represents an individual contact between two snails.

Column Descriptions:

*Date:*  Date that we ran the replicate. There were two trial weeks and treatment groups were stratified across days. Format is Month/Day/Year.

*Trial:* The trial or week during which we ran the microcosm replicate. This column is either 1 or 2, for the first trial or second. Note that all treatment groups were run in each trial.

*Tank:* The unique tank number for the replicate during which the contact took place.

*Density:* Host density or the number of snails in that tank/trial.

*AggregationIndex*: A numeric variable describing the aggregation of resources. An aggregation index of 0 indicates that the resources are uniformly spaced. As aggregation index increases, the clustering of resources also increases, and for distributions 4, 5, 6, distance to the nearest resource patch also increases with clustering.

*Snail1*: The unique identifier for one snail involved in the contact. The letters are based on the color(s) of nail polish used on the snail.

*Snail2:* The unique identifier for the other snail involved in the contact. The letters are based on the color(s) of nail polish used on the snail.

*StartTime:* The time that the contact first started between the two snails (*Snail1* and *Snail2*). The format is given as HH:MM:SS (e.g., 00:45:00 is the 45 minutes and zero seconds, the end of the trial).

*EndTime:* The time that the contact ends between the two snails (*Snail1* and *Snail2*). The format is given as HH:MM:SS (e.g., 00:45:00 is the 45 minutes and zero seconds, the end of the trial).

*Duration:* The total contact duration between the two snails (*Snail1* and *Snail2*), given in seconds.

**SnailExpData.csv**

File Description:

This data file contains information on each snail in the microcosm experiments, including the number of contacts each snail made. Additionally, information about the microcosm set up, such as *Snaildensity* and *AggregationIndex* are given. Each row represents an individual snail.

Column Description:

*Trial:* The trial or week during which we ran the microcosm replicate. This column is either 1 or 2, for the first trial or second. Note that all treatment groups were run in each trial.

*Tank:* Unique tank or replicate number

*ColorID*: A unique identifier for each snail in a given tank/microcosm. The letters are based on the color(s) of nail polish used on the snail.

*SnailDensity:* The snail density for the tank/microcosm. Snail density was either 2,4,6,8,10,12,14, or 16. If any snails were found dead or extremely lethargic, we removed it and adjusted the density for that tank/microcosm accordingly.

*AggregationIndex:* A numeric variable describing the aggregation of resources. An aggregation index of 0 indicates that the resources are uniformly spaced. As aggregation index increases, the clustering of resources also increases, and for distributions 4, 5, 6, distance to the nearest resource patch also increases with clustering.

*Length:* The length of the snail, in millimeters. These data were sometimes useful for identifying which snail was which in videos, where the color dots were sometimes difficult to distinguish.

*Count:* The total number of contacts made by the snail with any other snails, including repeat contacts with the same individuals.

**ExpPatchData.csv**

File Description:

This data file contains the movement information of every snail in the microcosm experiment, as determined from observations with videos. Each snail was watched for 45 minutes and we recorded which tile they were on, the duration of time spent on the tile, and whether the tile was on the edge or contained periphyton resources. Additionally, we noted when snails were floating and thus not foraging or moving on tiles. Floating observations were excluded from analyses. Each row represents one snail’s time spent on a single tile.

Column Description:

*date:* Date that we ran the replicate. There were two trial weeks and treatment groups were stratified across days. Given as Month/Day/Year.

*replicate:* The trial or week during which we ran the microcosm replicate. This column is either 1 or 2, for the first trial or second. Note that all treatment groups were run in each trial. In other datasets, this column is called “Trial”.

*tank:* Unique tank number within the trial

*snail.density:* The snail density for the tank/microcosm. Snail density was either 2,4,6,8,10,12,14, or 16. If any snails were found dead or extremely lethargic, we removed it and adjusted the density for that tank/microcosm accordingly.

*snail:* A unique identifier for each snail in a given tank/microcosm. The letters are based on the color(s) of nail polish used on the snail.

*tile.x:* The x coordinates for the tile that the snail (given in *snail* column) was on.

*tile.y:* The y coordinates for the tile that the snail (given in *snail* column) was on.

*edge.y.n.:* A yes, y, if the tile that the snail was on was on the edge of the microcosm and no, n, if it was not.

*algae.y.n.:* A yes, y, if the tile that the snail was on had periphyton (“algae”) on it and no, n, if the tile was an empty or resource-free tile.

*start.time:* The time when the snail first moved on to the current tile. Times of 00:00:00 indicate that the snail was on the tile at the start of the observation window.

*end.time:* The time when the snail moved off of the tile it was on to a new tile. Times of 0:45:00 indicate that the snail was on the tile at the end of the observation window, rather than an observation of the snail moving off the tile.

*duration:* The duration that the snail was on the listed tile (in *tile.x* and *tile.y* columns), given in seconds. This was calculated as the end.time minus the start.time.

*aggregation:* A numeric variable describing the aggregation of resources. An aggregation index of 0 indicates that the resources are uniformly spaced. As aggregation index increases, the clustering of resources also increases, and for distributions 4, 5, 6, distance to the nearest resource patch also increases with clustering.

*floating:* A column indicating whether the snail was floating over the indicated tile, marked as a 1, or was on the tile, marked as a 0. We excluded floating observations from analyses.

**GUD\_Data.csv**

File Description:

This data file contains information about how much periphyton (“algae”) remained on tiles at the times when snails left the resource tiles, as observed from screenshots from the video recordings made during the microcosm experiments. Screenshots were taken each time a snail left a tile, and screenshots were cropped and loaded into ImageJ to calculate the total area of the white tile and the total green area (covered by periphyton) remaining on the tile when the snail left. Each row represents a movement off a tile by a snail. Note that empty tiles or tiles that are not resource tiles contain 0’s and NA’s. Additionally, snails moving off empty times stopped being included in this data set because there was not green area to find in imageJ.

Column Description:

*tank:* Unique tank/microcosm number within a trial.

*snail:* A unique identifier for each snail in a given tank/microcosm. The letters are based on the color(s) of nail polish used on the snail.

*NeverMoved:* This column is used to distinguish observations for snails that were moving off tiles (0) from observations of snails that were on a tile at the end of the observation period (1), at time 0:45:00. When calculating average GUD, we excluded observations taken at the end of the observation period, where NeverMoved was 1.

*MoveTime*: The time that the snail moved off the tile. Given as Hours:Minutes:Seconds or H:MM:SS.

*TotalArea:* The total area of the tile that the snail was leaving. The area was found using ImageJ and a cropped image from a video of the microcosm. Due to slight differences in camera perspectives, the tile areas appear somewhat different, and using a proportion with the TotalArea as the denominator controls for differences in perspective.

*GreenArea..in.2.*: The total green area or area covered by periphyton left on the white tile when the snail moved off the tile. The green area was found using a cropped image taken from the video of the microcosm experiment and ImageJ.

*snail.density:* The snail density for the tank/microcosm. Snail density was either 2,4,6,8,10,12,14, or 16.

*aggregation.index:* A numeric variable describing the aggregation of resources. An aggregation index of 0 indicates that the resources are uniformly spaced. As aggregation index increases, the clustering of resources also increases, and for distributions 4, 5, 6, distance to the nearest resource patch also increases with clustering.

**TankInfo\_Data.csv**

File Description:

This data file contains information about the snail density, resource distribution, and snails in each of the microcosm experiments. Each row is a different tank/microcosm.

Column Description:

*replicate:* The trial or week during which we ran the microcosm replicate. This column is either 1 or 2, for the first trial of or second. Note that all treatment groups were run in each trial. In other datasets, this column is called “Trial”.

*tank:* Unique tank number within a trial

*snail density:* The snail density for the tank/trial. Snail density was either 2,4,6,8,10,12,14, or 16. If any snails were found dead or extremely lethargic, we removed it and adjusted the density for that tank/microcosm accordingly.

*aggregation:* A numeric variable describing the aggregation of resources. An aggregation index of 0 indicates that the resources are uniformly spaced. As aggregation index increases, the clustering of resources also increases, and for distributions 4, 5, 6, distance to the nearest resource patch also increases with clustering.

*snail color:* A unique identifier for each snail in a given tank/microcosm. The letters are based on the color(s) of nail polish used on the snail.