

-----Readme file -----

## **Habitat Risk Software**

Cornell Wildlife Health Lab,  
Animal Health Diagnostic Center,  
College of Veterinary Medicine,  
Cornell University  
Ithaca, New York, USA

### **Contacts**

#### *Software programmers:*

Brenda Hanley, bjh262@cornell.edu  
Corey Mitchell, coreyirene@gmail.com  
W. David Walter, wdw12@psu.edu

#### *Research collaborators:*

James Kelly, james.kelly@tn.gov  
Rachel Abbott, rca74@cornell.edu  
Nick Hollingshead, nah88@cornell.edu  
Lauren Miller, lauren.miller@dec.ny.gov

#### *Research supervisor:*

Krysten Schuler, ks833@cornell.edu

### **Overview of the Habitat Risk Software**

The **Habitat Risk Software** leverages a previously published Bayesian hierarchical model framework (Clayton and Kaldor 1987; Banderjee et al. 2004; Gelman et al. 2004; Evans et al. 2016) with opportunistic (hunter-harvest) wildlife surveillance disease testing data and publicly available geospatial (raster) data to estimate the geographical risk that a hunter will harvest a white-tailed deer (*Odocoileus virginianus*) that tests positive for Chronic Wasting Disease (CWD) in a small “study area” portion of the state.

The four-part R scripts of the **Habitat Risk Software** (1) prepare the surveillance (testing) and geospatial (raster) data for model inclusion, (2) parameterize and estimate coefficient values for 25 predetermined candidate model structures, (3) select the model structure with the lowest Deviance Information Criterion (DIC) given the data, (4) gather diagnostic plots to verify modeling assumptions have been met, and then (5) display the results of the best model in geographical and tabular context via an interactive web user interface (UI). UI capabilities include interactive and downloadable maps of estimated risk (and associated error), an overlay

map of disease data with the spatial covariates, and detailed statistical information about the best model and the model selection process. The first three of the four scripts must be run in sequence, as the data outputs of one become the data inputs of the next, while the fourth script is a “command center script” that may be used to run the others in sequence via one click.

The 1<sup>st</sup> R script (“1\_(state name)\_Data\_Prep.R”) of the **Habitat Risk Software** prepares the input data for model specification, which includes the standardization and stacking of disease testing data over spatial covariates. The command codes further create the sampling grid, load the spatial covariate data, assign the mean values of each covariate to cells in the sampling grid, overlay the disease testing data onto the grid, and save the stack to the working directory. Data inputs necessary to run the 1<sup>st</sup> script of the **Habitat Risk Software** include the preprocessed spatial layers (Mitchell et al. 2021) and surveillance disease testing data (more information below). Outputs from the 1<sup>st</sup> script of the **Habitat Risk Software** include a shape file of the sampling grid, the average spatial covariate values for each grid cell, and a full data set (where spatial covariate and grid cell designations are appended to each record of the disease testing data).

The 2<sup>nd</sup> R script (“2\_(state name)\_Pipeline.R”) of the **Habitat Risk Software** uses the outputs of the 1<sup>st</sup> script of the **Habitat Risk Software** to fit the pre-programmed model structures to data, estimate coefficient values, then pinpoint the best statistical model to estimate the risk that any given grid cell will harbor a CWD+ deer. The 2<sup>nd</sup> script of the **Habitat Risk Software** contains 25 pre-programmed models (determined *a priori*) that include additive combinations of age (fawn, yearling, adult), sex (male or female), digital elevations model (DEM; in meters above sea level), slope (in percent), forest (in percent canopy cover), streams (distance to the nearest water body in meters), clay (in percent clay in soil), local clustering (“CAR”, i.e., the variation ascribed to the proximity of CWD+ animals to a grid cell of interest), and regional clustering (“HET”, i.e., the variation ascribed to CWD+ at the regional scale) covariates. The 25 model structures are in Table 1. Each model is fit in WinBUGS (Spiegelhalter et al 2002; 2003) using the R2WinBUGS package in R (Sturtz et al. 2003). The 2<sup>nd</sup> **Habitat Risk Software** script then uses the Deviance Information Criterion (DIC) to identify which of the 25 candidate model structures exhibits the best fit given the data (Burnham and Anderson 2002). The 2<sup>nd</sup> script of the **Habitat Risk Software** then uses the best model to model response (CWD+ risk probabilities) and their upper and lower credible intervals for each cell in the sampling grid. Finally, the 2<sup>nd</sup> script of the **Habitat Risk Software** generates and saves the diagnostics plots for the best model. Data inputs necessary to run the 2<sup>nd</sup> script of the **Habitat Risk Software** include the files generated from the 1<sup>st</sup> script of the **Habitat Risk Software**. Software requirements necessary to run the 2<sup>nd</sup> of the **Habitat Risk Software** script include WinBUGS (Spiegelhalter et al 2002; 2003; more information below). Outputs from the 2<sup>nd</sup> script of the **Habitat Risk Software** include the best model, its parameter estimates, and the condensed spatial shape file of the sampling grid.

The 3<sup>rd</sup> R script (“3\_(state name)\_HabitatRisk.R”) of the **Habitat Risk Software** incorporates the outputs from the 2<sup>nd</sup> script of the **Habitat Risk Software** to produce the graphics in the UI. Graphics displayed in the UI include (1) an interactive map representing the computed results

of the best model for each cell in the sampling grid, (2) an interactive map that overlays spatial and testing data, (3) information regarding the model selection process including best model, and (4) a downloadable version of the risk estimates.

The 4<sup>th</sup> R script (“4\_(state name)\_Command\_Center.R”) of the **Habitat Risk Software** may be used to run the first 3 scripts of the **Habitat Risk Software** in the appropriate computational sequence via one click.

The user interface of the **Habitat Risk Software** may be executed for any area with disease testing records that contain exact latitude/longitude coordinates. The Bayesian hierarchical model and the user interface may be executed for any area containing disease testing records with exact latitude/longitude coordinates and at least one CWD+ animal (Evans et al. 2014). This packet does not furnish the real testing data necessary to run the **Habitat Risk Software**.

### Inputs Needed to Run the Habitat Risk Software

Data inputs necessary to run the 1<sup>st</sup> script of the **Habitat Risk Software** include (1) the preprocessed spatial layers (Mitchell et al. 2021) and (2) the surveillance disease testing data. Instructions for the procurement of spatial (raster) data and the specific processing requirements appear in Mitchell et al. 2021 and in the section **Preprocessing of Geospatial Data** (below). Table 2 depicts the data structure required for the disease surveillance testing data and the section **Template for the Disease Testing Data** (below) displays the necessary variables names and category notation.

Data inputs necessary to run the 2<sup>nd</sup> script of the **Habitat Risk Software** include (1) the spatial (raster) layers processed and autosaved during the execution of the 1<sup>st</sup> script of the **Habitat Risk Software** and (2) the data files processed and autosaved during the execution of the 1<sup>st</sup> script of the **Habitat Risk Software**. The software requirements necessary to run the 2<sup>nd</sup> script of the **Habitat Risk Software** include WinBUGS (Spiegelhalter et al 2002; 2003), found at <https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>.

Data inputs necessary to run the 3<sup>rd</sup> script of the **Habitat Risk Software** include (1) the spatial (raster) layers processed and autosaved during the execution of the 2<sup>nd</sup> script of the **Habitat Risk Software** and (2) the data files processed and autosaved during the execution of the 2<sup>nd</sup> script of the **Habitat Risk Software**.

Data inputs necessary to run the 4<sup>th</sup> script of the **Habitat Risk Software** include (1) the preprocessed spatial layers (Mitchell et al. 2021) and (2) the surveillance disease testing data. Instructions for the procurement of spatial (raster) data and the specific processing requirements appear in Mitchell et al. 2021 and in the section **Preprocessing of Geospatial Data** (below). Table 2 depicts the data structure required for the disease surveillance testing data and the section **Template for the Disease Testing Data** (below) displays the necessary variables names and category notation. The software requirements necessary to run the 4<sup>th</sup>

script of the **Habitat Risk Software** include WinBUGS (Spiegelhalter et al 2002; 2003), found at <https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>.

## Habitat Risk Software Tutorial

### *Preparing and Running the Command Code (the 4<sup>th</sup> R Script)*

**Step 1:** Open R and verify that you are running the appropriate version of the R Software. The appropriate version is 4.0.2 (2020-06-22) -- "Taking Off Again"

Copyright (C) 2020 The R Foundation for Statistical Computing  
Platform: x86\_64-w64-mingw32/x64 [64-bit].

**Step 2:** Install the appropriate versions of the dependencies (R packages) in R. Packages include:

- (1) "devtools" version 2.3.2,
- (2) "sp" version 1.4-4,
- (3) "rgdal" version 1.5-23,
- (4) "raster" version 3.4-5,
- (4) "spdep" version 1.1-5.
- (5) "dplyr" version 1.0.4,
- (6) "rgeos" version 0.5-5.
- (7) "coda" version 0.19-4,
- (8) "R2WinBUGS" version 2.1-21,
- (9) "car" version 3.0.10,
- (10) "shinydashboard" version 0.7.1,
- (11) "shinydashboardPlus" version 0.7.5,
- (12) "shiny" version 1.6.0,
- (13) "leaflet" version 2.0.3,
- (14) "leaflet.extras" version 1.0.0,
- (15) "RColorBrewer" version 1.1-2,
- (16) "rgl" version 0.103.5,
- (17) "shinyBS" version 0.61,
- (18) "rmarkdown" version 2.6,
- (19) "MASS" version 7.3-53,
- (20) "png" version 0.1-7,
- (21) "sf" version 0.9-7,
- (22) "leafpop" version 0.0.6,
- (23) "mapview" version 2.9.0,
- (24) "rmapshaper" version 0.4.4,
- (25) "htmlwidgets" version 1.5.3.

To install a particular version of an R package, run the code:

```
library(devtools)
```

```
devtools::install_version("package name", version = "#.#.#").
```

**Step 3:** Create a folder on your computer that will function as your working directory.

**a:** Create a folder then name it in accordance with your project. This folder will be your working directory.

**b:** Set the path to the working directory in R by clicking the R Console, clicking “File”, clicking “Change dir...”, then navigating through the file paths to your working directory. Click “OK”.

**c:** Verify that R has the correct path to your working directory by clicking the R console, typing “`dir()`”, and hitting enter. The file names of the contents of your working directory will print in the R console.

**Step 4:** Load **WinBUGS 14** into your working directory.

**a:** Go to <https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/> and download the **WinBUGS 14** software onto your machine.

**b:** Unzip (if necessary) the contents of the WinBUGS 14 software and save the packet *as is* in your working directory.

**c:** Double check that the folder name is “winbugs14\_full\_patched” and the subfolder name is “WinBUGS14”.

**d:** Double check this folder is in your working directory by clicking the R console, typing “`dir()`”, and hitting enter. The file names of the contents of your working directory (including this packet) will print in the R console.

**Step 5:** Prepare the **Habitat Risk Software** on your machine.

**a:** Download, unzip, and save **Habitat Risk Software** contents in your working directory.

**b:** To double check all contents are in your working directory, click the R console, type “`dir()`”, and hit enter. The file names of the contents of your working directory will print in the console.

**Step 6:** Pre-process the spatial (raster) data using the instructions in Mitchell et al. (2021).

**a:** Go to and download the contents from Mitchell et al. (2021).

**b:** Open the Readme of the **Geospatial Software** and follow the directions.

**c:** Save the prepared spatial files (raster images) into your working directory.

**Step 7:** Open the “4\_(state)\_Command\_Center.R” script in R.

**Step 8:** Hit “Run all”.

From this moment in **Step 8**, your script will automatically call the appropriate libraries, then begin running the 1<sup>st</sup> script of the **Habitat Risk Software**. Several lines of code will be executed without any further interaction from you. Your script will automatically load the shape files, initiate projections, draw a map of your state then populate the sampling data, reproject the map under the desired projection, add the outline around the state boundaries, then appear to halt computations. The halt will occur when the line

```
e <- drawExtent(show=TRUE, col="red")#click twice in plot window  
around positive cluster to subsample an area of interest
```

appears on your console. This is the moment when you must draw your desired sampling grid. Your sampling grid will constitute your study area and should represent a small portion of the state. To draw your sampling grid, you will left click your mouse on the map at right two times: (1) your 1<sup>st</sup> click on the map in the upper left hand spatial area of interest will define the northwest corner of your sampling grid and (2) your 2<sup>nd</sup> click on the map in the lower right hand spatial area of interest will define the southeast corner of your sampling grid. After you draw your sampling grid (i.e., after you complete both clicks on the map), the computer will automatically define the remaining boundaries of your sampling grid, populate the grid cell boundaries, name grid cells, crop the rasters to your sampling grid, compute the mean value of each spatial covariate in each cell, append grid numbers and spatial covariate values to each testing record, then save sampling grid, cropped rasters, and appended data to your working directory.

The script will then automatically open the 2<sup>nd</sup> script of the **Habitat Risk Software**, where it will load additional R packages, uptake the computational products of the 1<sup>st</sup> script of the **Habitat Risk Software**, load and prepare the shape file of the sampling grid, then reformat the data for inclusion in the WinBUGS computing language. From this moment, the script assigns the study area into one of two possible mutually exclusive categories depending on CWD status:

- 1) If the wildlife surveillance disease testing data contains location of each deer in exact latitude/longitude coordinates *and* at least one record (deer) is CWD+, then the computer automatically selects and executes the command lines in *PATH A*.
- 2) If the wildlife surveillance disease testing data contains location of each deer in exact latitude/longitude coordinates *and* all records (deer) are CWD-, then the computer automatically selects and executes the command lines in *PATH B*.

*PATH A*. Your script will initialize, parameterize, and activate WinBUGS to fit the 1<sup>st</sup> of the 25 candidate model structures. A window depicting the UI of WinBUGS will pop-up on your desktop and you will see the automation occurring on that screen. Once WinBUGS has fully updated the 1<sup>st</sup> model, R will automatically close WinBUGS, summarize the parameter estimates, find the DIC, then save all the WinBUGS outputs to your working directory. Your script will then automatically reactivate WinBUGS to fit the 2<sup>nd</sup> of the 25 candidate model structures, then follow the same pattern of fitting and saving. Your script will repeat this pattern until all 25 models have been run in sequence. The script will then select the model with the lowest relative DIC, identify and save the median (50%), upper (97.5%), and lower (2.5%) credible intervals for the CAR and HET values for each grid cell (if necessary), identify and save the median (50%), upper (97.5%), and lower (2.5%) credible intervals for the risk of adult bucks, adult does, yearling bucks, yearling does, fawn bucks, and fawn does for each grid cell, compress and save the cropped shapefiles for rapid load/display in the user interface in the 3<sup>rd</sup> **Habitat Risk Software** script, then generate and save the diagnostic plots for the best model. All files are autosaved to your working directory.

*PATH B*. Your data does not have CWD+ animals, so you cannot run the Bayesian hierarchical model. However, you can still visualize how your testing data overlays spatial covariates in your study area of interest. Accordingly, your script will skip over WinBUGS model fitting and model selection portion of the 2<sup>nd</sup> **Habitat Risk Software**, create placeholder (empty) data files, then compress and save the cropped shapefiles for rapid load/display in the user interface in the 3<sup>rd</sup> **Habitat Risk Software** script. All files are autosaved to your working directory.

The script will then automatically open the 3<sup>rd</sup> script of the **Habitat Risk Software**, where it will load additional R packages, uptake the computational products of *PATH A* or *PATH B* of the 2<sup>nd</sup> script of the **Habitat Risk Software**, load the shape file of the sampling grid, load the testing data, load the compressed raster images, then display the spatial data, testing data, and modeling results via the UI. If a study area satisfied the modeling assumptions for *PATH A*, all UI tabs will be auto-populated with data or results. If a study area satisfied the modeling assumptions for *PATH B*, only the UI tab that overlays the disease testing data onto the condensed spatial rasters will auto-populate.

**Step 9:** Once done interacting with the UI, close the tab depicting the UI, stop the R code, and close the R program.

### **Important Computational Considerations of the 1<sup>st</sup> Script of the Habitat Risk Software**

Modeling results will hinge on the selection of your sampling grid in the context of the testing data. If your study area has CWD+ animals (i.e., *PATH A*), the sampling grid that you select should contain (1) disease-negative (black) points, (2) little open space (i.e., most cells should contain data), and (3) at least one disease-positive (red) point. If your jurisdiction does not have CWD+ animals (i.e., *PATH B*), the sampling grid that you select should only contain disease-negative (black) points.

Larger spatial extents of your sampling grid, greater numbers of data points internal to the grid, or both, will lengthen the runtime of the models in WinBUGS.

If you mis-key or do not otherwise like the selection of your sampling grid, you may stop the script, close R, then restart the **Habitat Risk Software** at **Step 7**.

The **Habitat Risk Software** is pre-programmed to auto-populate your sampling grid to a cell size of 6 square kilometers, which is the estimated home range size of white-tailed deer (J. Kelly, expert opinion). If this home range size is not appropriate to your study area, you may alter the cell size by changing the command

```
Home_range_size=6
```

to your appropriate home range size. Home range size should be in square kilometers.

Subsequent processing of spatial and disease data will automatically ensue for your sampling grid through raster cropping and stacking and during these processing commands, the script will automatically display in and remove from the plotting window some graphical representations of the sampling grid and raster layers. This is normal and you should allow the computer the time necessary to complete these computations.

### **Important Computational Considerations of the 2<sup>nd</sup> Script of the Habitat Risk Software**

The 2<sup>nd</sup> script takes up to 14 days of run time, depending on the spatial extent of the sampling grid and the density of the disease testing data. During this runtime, let the script do what it is intended to do. If you are not going to monitor this computer while it runs this script, ensure (1) it has a consistent power source, (2) the programs will not be halted by automatic updates [perhaps power down the internet connection], (3) the screen is off. You may also run this script in the background while you continue other work. In either case, make sure your computer has a constant power source and your screen is turned off when you are not directly using it (you don't want the lengthy runtime to burn out the lights on your screen). If the script breaks at any point, you will need to start over at that point by either (1) starting over at Step 7 or (2) opening and running the 2<sup>nd</sup> script by hand (more on that, below).

Because WinBUGS is functionally embedded in the computational pathways of the command code, errors that occur in any of the 25 WinBUGS deployments will halt the computational pipeline and inhibit the **Habitat Risk Software** from functioning as intended. Errors that occur in WinBUGS may arise from inappropriate parameterization of priors, the number of iterations, the length of the burn-in, the thinning parameters, errors in your testing data, errors in your spatial covariates, missing of spatial covariates in your sampling grid, storage or RAM issues, etc. You must troubleshoot these errors using the WinBUGS documentation [<https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf>, Accessed June 2021].

The **Habitat Risk Software** is preprogrammed to pick the model that has the lowest DIC. However, you may wish to override that selection to pick a more parsimonious model with similar DIC. You may manually override the pre-programmed model selection and enter the model of interest by removing the

```
BEST=as.numeric(which.min(Vector))
```

command and replacing it with

```
Best=X,
```

where X is the numeric index of the desired model (see Table 1).

All computations are automatically saved in the working directory. Should you wish to look closely at the output for any model, you can do so by visiting the files for that model in your working directory.

After model fitting and selection, the covariate estimation and subsequent data crunching procedures will automatically run in both the console and plotting windows. This is normal and you should allow the computer the time necessary to complete these computations.

### **Important Computational Considerations of the 3<sup>rd</sup> Script of the Habitat Risk Software**

It is imperative to run the first two scripts of the *Habitat Risk Software* before you open the 3<sup>rd</sup> script for the first time. However, if you do not intend to add more data or modify the sampling grid, you may run the 3<sup>rd</sup> script repeatedly without rerunning its predecessors.

### **Preprocessing of the Geospatial Data Needed to Run the Habitat Risk Software**

Pre-processed geospatial data must be taken into the 1<sup>st</sup> script of the *Habitat Risk Software* in a precise format and structure. The software used to pre-process geospatial data archived in public geospatial repositories is available in Mitchell et al. (2021).

The processed geospatial data include “(POSTAL CODE)\_(EPSG CODE).gpkg”, “(POSTAL CODE)\_(EPSG code)\_buff\_50km.gpkg”, “Soils\_gNATSGO\_30m\_(POSTAL CODE)\_(EPSG CODE).tif”, “Streams\_NHDstrahler4plus\_distance\_(POSTAL CODE)\_(EPSG CODE).tif”, “NLCD2016\_Open\_(POSTAL CODE)\_(EPSG CODE).tif”, “NLCD2016\_Forested\_(POSTAL CODE)\_(EPSG CODE).tif”, “NLCD2016\_Developed\_(POSTAL CODE)\_(EPSG CODE).tif”, “DEM\_slope\_30m\_(POSTAL CODE)\_(EPSG CODE).tif”, “DEM\_elevation\_30m\_(POSTAL CODE)\_(EPSG CODE).tif”.

The naming format for each layer includes the two-digit state postal code (POSTAL CODE), for example NY, OH, VA, and EPSG code of the selected State Plane Coordinate System for that state (EPSG CODE), for example 32116. For New York, the first geospatial layer read into the script is named “NY\_32116\_buff\_50km.gpkg”.

### **Template for the Disease Testing Data Needed to Run the Habitat Risk Software**

The disease testing data must be taken into either the command code or the 1<sup>st</sup> script of the *Habitat Risk Software* in a precise format and structure.

The disease testing data include “(state name)FineScaleDataCleanedAll.csv”.

The naming format for the disease testing data are *Sex*, *Age.Group*, *Sample.Dat*, *longitude*, *latitude*, *Result*, *County*, and *Scale*, in that order. Each row of data represents an individual deer that was harvested and tested for the disease. *Sex* is limited to categories “male” and “female”. *Age.Group* is limited to categories “adult”, “yearling”, and “fawn”. *Sample.Dat* represents the year that the deer was harvested from the wild and must be of the format of YYYY. *longitude* and *latitude* are the longitude and latitude of the deer at the exact location of harvest. *Result* is the disease test result, limited to the disease categories “Not Detected” and “Positive”. *County* represents the county in which the deer was harvested, and *Scale* is equal to 1 for all deer. The programming language R is sensitive to capital letters, spaces, and characters, so you must convert your own data into the precise headings and categories found in the templates.

## Technical Details of the Habitat Risk Software

The command center script was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86\_64-w64-mingw32/x64 (64-bit) and requires all the packages listed for the other three scripts.

1<sup>st</sup> (of 4) R script (“Data\_Prep”) was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86\_64-w64-mingw32/x64 (64-bit), and requires seven packages: (1) “devtools” version 2.3.2, (2) “sp” version 1.4-4, (3) “rgdal” version 1.5-23, (4) “raster” version 3.4-5, (5) “spdep” version 1.1-5, (6) “dplyr” version 1.0.4, and (7) “rgeos” version 0.5-5.

2<sup>nd</sup> (of 4) R script (“Pipeline”) was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86\_64-w64-mingw32/x64 (64-bit), and requires eight packages: (1) “devtools” version 2.3.2, (2) “coda” version 0.19-4, (3) “R2WinBUGS” version 2.1-21, (4) “sp” version 1.4-4, (5) “rgdal” version 1.5-23, (6) “spdep” version 1.1-5, (7) “car” version 3.0.10, and (8) “raster” version 3.4-5. The 2<sup>nd</sup> script further requires that you download WinBUGS from: <https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>.

3<sup>rd</sup> (of 4) R script (“HabitatRisk”) was written under R version 4.0.2 (2020-06-22) -- "Taking Off Again" Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86\_64-w64-mingw32/x64 (64-bit), and RStudio version 1.3.1093 Copyright © 2009-2020 “Apricot Nasturtium” (aee44535, 2020-09-17) for Windows, and requires the packages: (1) “shinydashboard” version 0.7.1, (2) “shinydashboardPlus” version 0.7.5, (3) “shiny” version 1.6.0, (4) “leaflet” version 2.0.3, (5) “dplyr” version 1.0.4, (6) “leaflet.extras” version 1.0.0, (7) “RColorBrewer” version 1.1-2, (8) “rgl” version 0.103.5, (9) “shinyBS” version 0.61, (10) “rmarkdown” version 2.6, (11) “MASS” version 7.3-53, (12) “png” version 0.1-7, (13) “rgdal” version 1.5-23, (14) “sp” version 1.4-4, (15) “sf” version 0.9-7, (16) “leafpop” version 0.0.6, (17) “mapview” version 2.9.0, (18) “raster” version 3.4-5, (19) “rmapshaper” version 0.4.4, (20), “devtools” version 2.3.2, and (21) “htmlwidgets” version 1.5.3.

4<sup>th</sup> (of 4) R script (“HabitatRisk”) was written under R version 4.0.2 (2020-06-22) -- “Taking Off Again” Copyright (C) 2020 The R Foundation for Statistical Computing Platform: x86\_64-w64-mingw32/x64 (64-bit), and RStudio version 1.3.1093 Copyright © 2009-2020 “Apricot Nasturtium” (aee44535, 2020-09-17) for Windows, and requires the packages: (1) “devtools” version 2.3.2, (2) “sp” version 1.4-4, (3) “rgdal” version 1.5-23, (4) “raster” version 3.4-5, (5) “spdep” version 1.1-5, (6) “dplyr” version 1.0.4, (7) “rgeos” version 0.5-5, (8) “coda” version 0.19-4, (9) “R2WinBUGS” version 2.1-21, (10) “car” version 3.0.10, (11) “shinydashboard” version 0.7.1, (12) “shinydashboardPlus” version 0.7.5, (13) “shiny” version 1.6.0, (14) “leaflet” version 2.0.3, (15) “leaflet.extras” version 1.0.0, (16) “RColorBrewer” version 1.1-2, (17) “rgl” version 0.103.5, (18) “shinyBS” version 0.61, (19) “rmarkdown” version 2.6, (20) “MASS” version 7.3-53, (21) “png” version 0.1-7, (22) “sf” version 0.9-7, (23) “leafpop” version 0.0.6, (24) “mapview” version 2.9.0, (25) “rmapshaper” version 0.4.4, and (26) “htmlwidgets” version 1.5.3. The 4<sup>th</sup> script further requires that you download WinBUGS from: <https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>.

## Citations

- Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2004. Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, New York.
- Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer-Verlag. 1-488 p.
- Clayton, D., and J. Kaldor. 1987. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43:671-681.
- Evans, T. S., Kirchgessner, M. S., Eyer, B., Ryan, C. W., and Walter, W. D. 2016. Habitat Influences Distribution of Chronic Wasting Disease in White-tailed Deer. *Journal of Wildlife Management*, 80(2):284-291
- Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. Bayesian data analysis. New York: Chapman and Hall/CRC. 1-696 p.
- Mitchell, C., Walter, W. D., Hollingshead, N., & Schuler, K. 2021. Processing of Geospatial Data for the Habitat Risk Software [Software]. Cornell University Library eCommons Repository. <https://doi.org/10.7298/2tt1-yy48>
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64:583-639.
- Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. 2003. WinBUGS Version 1.4 user manual. MRC Biostatistics Unit, Cambridge.
- Sturtz, S., U. Ligges, and A. Gelman. 2005. R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software* 12:1-16.
- Walter, W. D., D. P. Walsh, M. L. Farnsworth, D. L. Winkelman, and M. W. Miller. 2011b. Soil clay content underlies prion infection odds. *Nature Communications* 2:1-6.

## Financial Support

Financial support comes from (1) Michigan Disease Initiative – *Optimizing CWD Surveillance: Regional Synthesis of Demographic, Spatial, and Transmission-Risk Factors* (2019); (2) Tennessee Wildlife Resources Agency - *Modeling Risk of Infection for Individually Harvested Deer & Estimating Prevalence When Sampling is Limited* (2020); (3) Michigan Disease Initiative - *SOP4CWD Dashboard: A Web Application for Disease Visualization and Data-Driven Decisions* (2020); (4) Multistate Conservation Grant Program - *Surveillance Optimization Project for Chronic Wasting Disease: Streamlining a Web Application for Disease Visualization and Data-Driven Decisions* (2021).

## Acknowledgements

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We thank J. Fleegle, A. Korman, J. Peaslee, C. Them, and four anonymous professionals for their helpful suggestions to improve the functionality and appearance of the UI.

We further acknowledge the R Shiny internet forums, including <https://stackoverflow.com/questions/48432061/turn-states-on-a-map-into-clickable-objects-in-shiny>, <https://stackoverflow.com/questions/45953741/select-and-deselect-polylines-in-shiny-leaflet>, <https://rstudio.github.io/leaflet/shiny.html>, <https://stackoverflow.com/questions/44259716/how-to-save-a-leaflet-map-in-shiny>, and <https://github.com/andrewsali/shinycssloaders> for providing examples that we used to program various capabilities.

## License

The ***Habitat Risk Software*** is shared under a MIT License.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN

AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

**Suggested Citation for Re-Use of the *Habitat Risk Software*:**

Hanley, B., Mitchell, C., Walter, W. D., Kelly, J., Abbott, R., Hollingshead, N., Miller, L., & Schuler, K. 2021. Habitat Risk Software [Software]. Cornell University Library eCommons Repository. <https://doi.org/10.7298/rcz8-nw50>

Table 1. The set of candidate models determined *a priori* and used by the **Habitat Risk Software** to investigate the effect of demographic and environmental covariates on odds of CWD infection. The intercept corresponds to the baseline class (in the case of CWD), *sex* is coded for male=1 and female=0, *age* is coded for fawn=0, yearling=1, and adult=2. Environmental covariates include *dem*, the digital elevation model, *slope*, the localized gradient of the area, *for*, a compound category representing forest cover (which is also highly correlated with the percentage of open and developed land covers), *streams*, the distance to the nearest water source, and *clay*, the percentage of clay in the soil, *CAR*, the spatial effects include local clustering, and *HET*, the region-wide heterogeneity.

Model Index	Model Structure
1	<i>Intercept + sex + age + dem + slope + for + streams + clay + CAR + HET</i>
2	<i>Intercept + sex + age + dem + slope + ---- + streams + clay + CAR + HET</i>
3	<i>Intercept + sex + age + dem + ----- + for + streams + clay + CAR + HET</i>
4	<i>Intercept + sex + age + ----- + slope + for + streams + clay + CAR + HET</i>
5	<i>Intercept + sex + age + dem + ----- + ---- + streams + clay + CAR + HET</i>
6	<i>Intercept + sex + age + ----- + slope + ---- + streams + clay + CAR + HET</i>
7	<i>Intercept + sex + age + ----- + ----- + for + streams + clay + CAR + HET</i>
8	<i>Intercept + sex + age + ----- + ----- + ---- + streams + clay + CAR + HET</i>
9	<i>Intercept + sex + age + ----- + ----- + ---- + ----- + clay + CAR + HET</i>
10	<i>Intercept + sex + age + ----- + ----- + ---- + streams + ----- + CAR + HET</i>
11	<i>Intercept + sex + age + ----- + ----- + ---- + streams + clay + ----- + HET</i>
12	<i>Intercept + sex + age + ----- + ----- + ---- + streams + clay + CAR + -----</i>
13	<i>Intercept + sex + age + ----- + ----- + ---- + ----- + ----- + CAR + HET</i>
14	<i>Intercept + sex + age + ----- + ----- + ---- + ----- + clay + ----- + HET</i>
15	<i>Intercept + sex + age + ----- + ----- + ---- + ----- + clay + CAR + -----</i>
16	<i>Intercept + sex + age + ----- + ----- + ---- + streams + ----- + ----- + HET</i>
17	<i>Intercept + sex + age + ----- + ----- + ---- + streams + ----- + CAR + -----</i>
18	<i>Intercept + sex + age + ----- + ----- + ---- + streams + clay + ----- + -----</i>
19	<i>Intercept + sex + age + ----- + ----- + ---- + ----- + ----- + ----- + HET</i>
20	<i>Intercept + sex + age + ----- + ----- + ---- + ----- + ----- + CAR + ----</i>
21	<i>Intercept + sex + age + ----- + ----- + ---- + ----- + clay + ----- + ----</i>
22	<i>Intercept + sex + age + ----- + ----- + ---- + streams + ----- + ----- + ---</i>
23	<i>Intercept + sex + age + ----- + ----- + ---- + ----- + ----- + ----- + ----</i>
24	<i>Intercept + ----- + ---- + ----- + ----- + ---- + ----- + ----- + CAR + HET</i>
25	<i>Intercept + ----- + ---- + ----- + ----- + ---- + ----- + ----- + ----- + ----</i>

Table 2. An example of the wildlife disease testing data structure containing *fictitious* data. This structure is required for the surveillance disease testing. *Fictitious data is used in this table for illustrative purposes only.* *Sex* is limited to categories “male” and “female”. *Age.Group* is limited to categories “adult”, “yearling”, and “fawn”. *Sample.Dat* represents the year that the deer was harvested from the wild and must be of the format of YYYY. *Longitude* and *latitude* are the longitude and latitude of the deer at the exact location of harvest. *Result* is the disease test result, limited to the disease categories “Not Detected” and “Positive”. *County* represents the county in which the deer was harvested, and *Scale* is equal to 1 for all deer.

Sex	Age.Group	Sample.Dat	longitude	latitude	Result	County	Scale
male	adult	2018	-86.123	32.000	Not Detected	BARTHOLOMEW COUNTY	1
female	yearling	2024	-85.321	35.123	Positive	ALLEGHENY COUNTY	1
male	fawn	2017	-83.000	34.321	Not Detected	BENSON COUNTY	1