Python FishRand User Manual

January 2019

Contents

Glossary

Attraction Area The polygon in which a higher density of a certain species of fish resides.

Attraction Factor A weight that is multiplied by the area of a region. An attraction factor of 100 in region 1, is equivalent to saying the density of fish in region 1 is 100 times more than in the surrounding areas.

Domain The Domain is the polygon that bounds the entire area in which fish movement is simulated.

Fish Population A Fish population is used to mean one of the number of populations per species in the simulation. Each population in a species follows a separate path through the pond according to probabilities generated in the spatial modeling step.

Latin Hypercube A method of sampling distributions, in which the distribution is partitioned into different sections and each section is sampled a equal number of times.

Sample Site A Sample Site is the x,y coordinate and associated regional and chemical proprieties at that point.

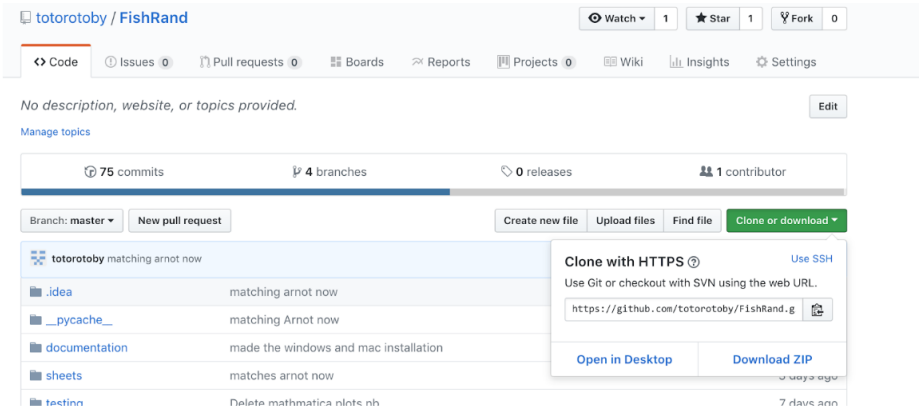
Uncertain Parameter A Parameter that is statistical because a user is unsure of what the exact measurements of the parameter are. Uncertain Parameters are sampled in the outer Monte-Carlo loop.

Variable Parameter A Parameter that is statistical because it inherently varies from sample to sample. Uncertain parameters are sampled with the inner Monte-Carlo loop.

1 Installation

1.1 Getting FishRand

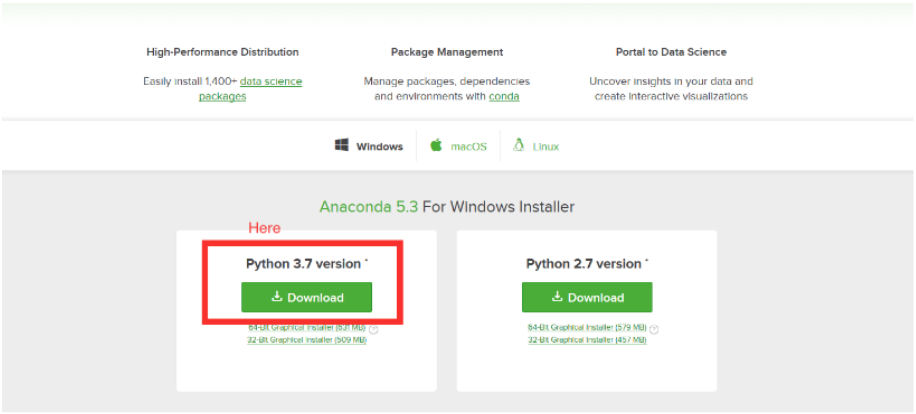
1. Go to `https://github.com/totorotoby/FishRand`
2. Download the directory as a ZIP, from the "Clone or download" button.



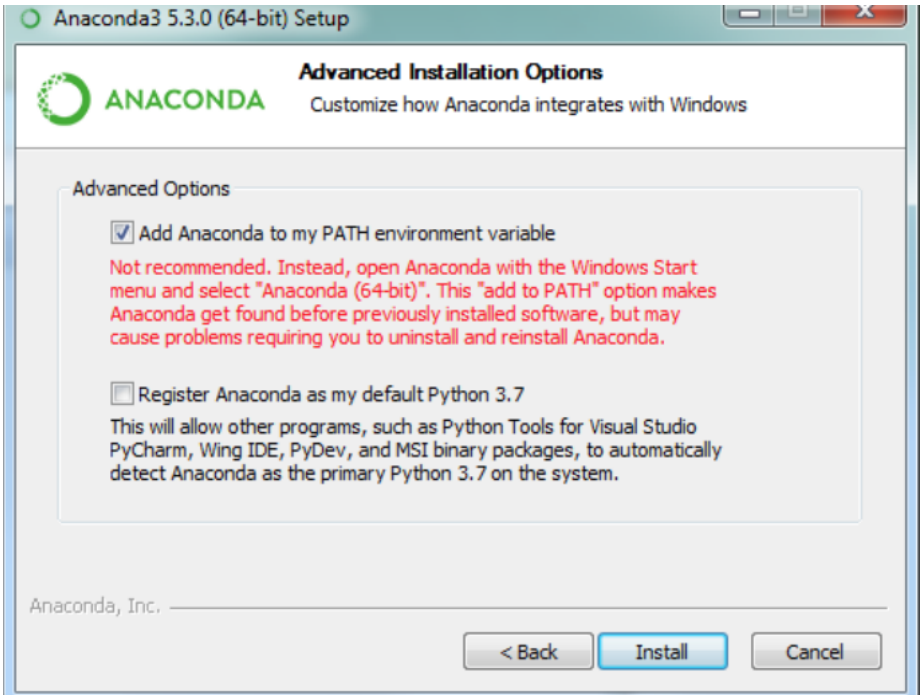
3. Unzip the folder in whatever location you downloaded it to.

1.2 Windows installation

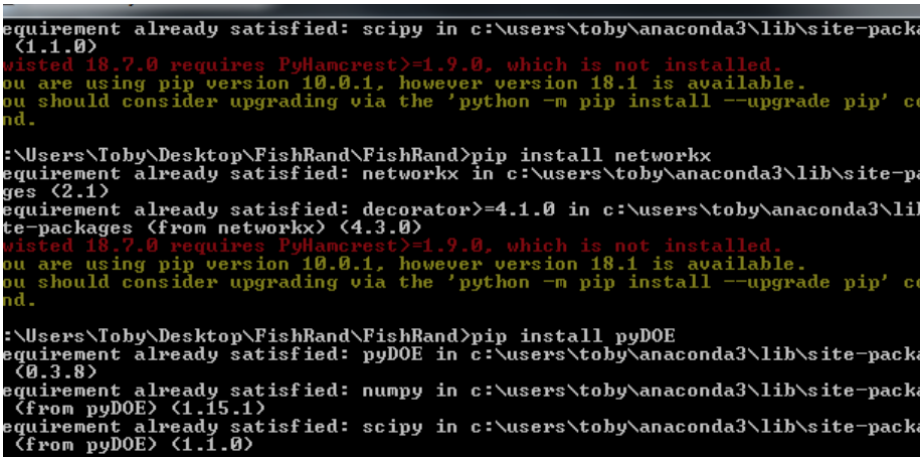
1. Go to: `www.anaconda.com/download/`, and download the python 3.7 version.



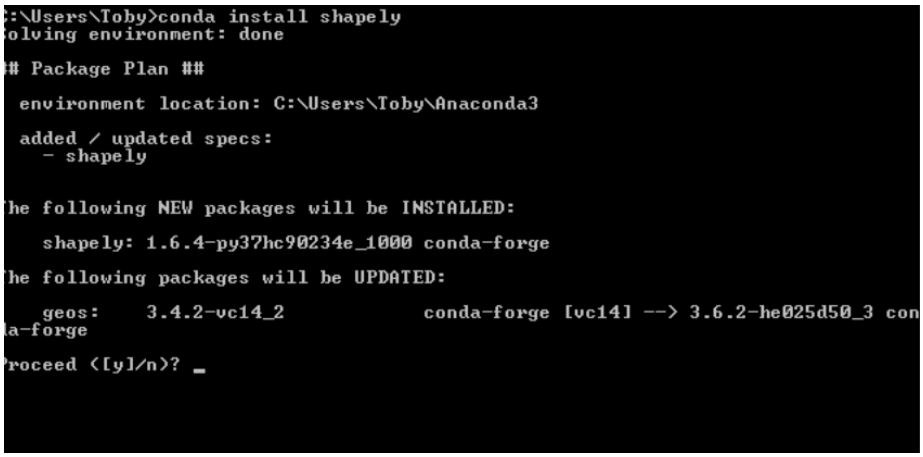
2. Install Anaconda. In the installation prompt check "Add Anaconda to my PATH".



3. Locate "windows_installer.bat" in your FishRand folder, and double click that. A command prompt window should appear and run multiple commands. (There may be some warning messages, ignore these.)



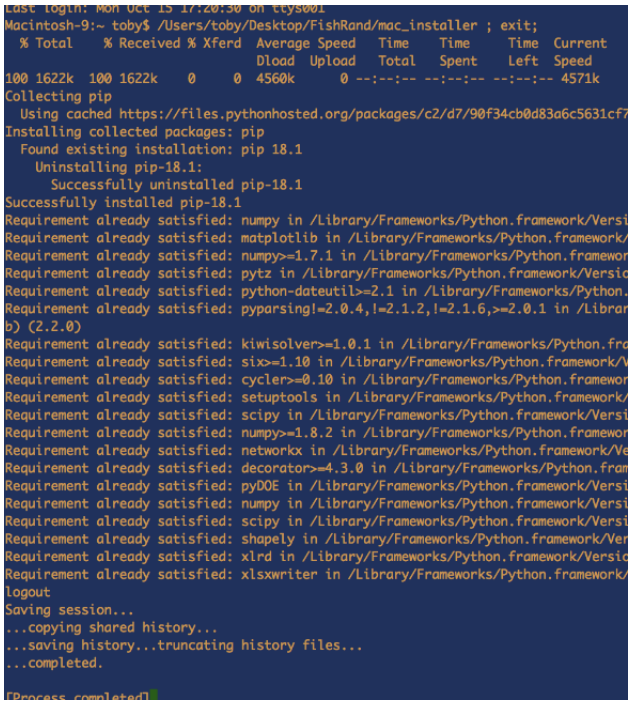
4. Open another command prompt window, and type "conda install shapely" and hit return. After a while there will be some text asking for a "y/n". Type "y" and hit return



1.3 Mac Installation

If you are sure you have Python 3 (Check by typing "python3 -V" into terminal) on your computer you can skip steps 1 and 2.

1. Go to `https://www.python.org/downloads/release/python-370/Download` the corresponding installer for mac and Run the installer.
2. In the FishRand Folder, double click on "mac_installer. A terminal window should appear and execute a bunch of commands.



3. From now on, every time you want to run FishRand you can double click on the file in the FishRand directory named FishRand_mac_linux (with no extension).

2 Model Structure

2.1 Bio-accumulation Model

Similar to previous versions of FishRand, the Python version has a Bio accumulation Model nested inside both Monte Carlo iterations, and spatial realizations for each fish population. The Python version uses the Bio-accumulation Model from Arnot, 2004 with a ability to change some constants (i.e. Octal Proportionality Constants, and Dietary Absorption efficiencies).

2.2 Monte Carlo Simulation

In FishRand, bio-accumulation parameters can be described as statistical distributions that are then randomly sampled for Monte Carlo Simulations.

2.2.1 Sampling

Each distribution is of type Uncertain or Variable, and is sampled in the following way:

Let $X = D_t(p_1, \ldots, p_e)$ be the Distribution of type t (uncertain or variable) with parameters $p_1, \ldots, p_e$. Then random realizations are drawn from D_t by partitioning D_t by b quantiles and randomly drawing the same number of samples from each partition.

If the Distribution is classified as "Uncertain," FishRand creates a vector:
$V_X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, where x_i is the ith sample from X, with the following property:

If b is the number of partitions of X indexed by k, then from each partition $\frac{n}{b}$ samples are drawn from X_k where X_k is the kth partition of X. Each sample is then placed at index i of vector V_X with uniform probability.

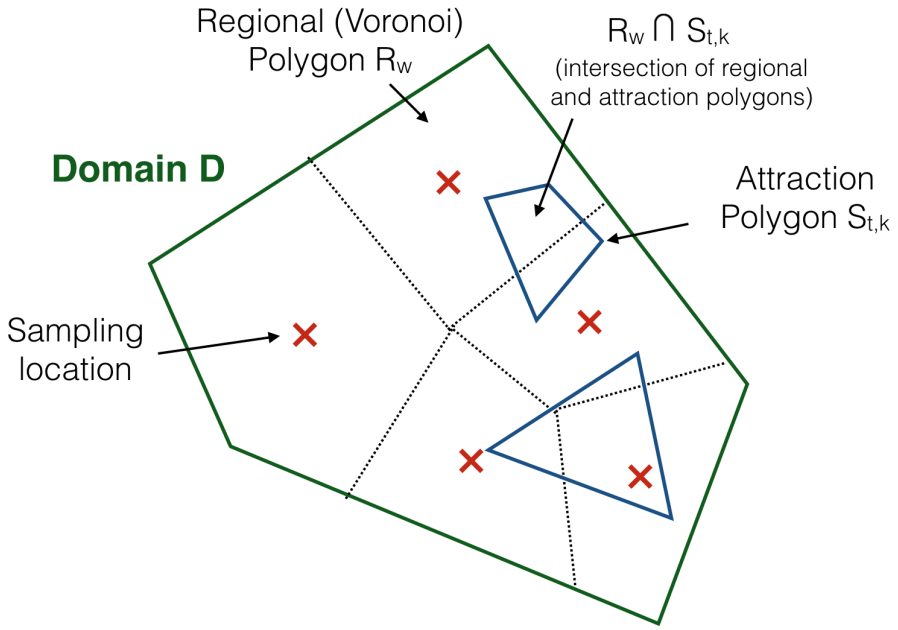
If the Distribution is Variable FishRand creates a matrix:

$$M_Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix}$$

Where variable distributions are sampled nm times instead of just n times. M_X has the same property as V_X except that each sample is placed at indices i, j with uniform probability.

2.2.2 Spatial Structure

Fishrand constructs Regional Polygons R_w around each measurement location to map the areas where the parameter values associated with measurements are applied. Using the coordinates specified for sample site input, FishRand generates Voronoi Polygons R_w around each sample site that are also bounded by the Domain Polygon input in the Domain input section. The area of each Regional Polygon R_w is then calculated (to be used to determine the fish populations location in step 3.) For more info on the Voronoi polygon function used see: `https://docs.scipy.org/doc/scipy-0.18.1/reference/generated/scipy.spatial.Voronoi.html`



The polygons for different attraction areas are then overlaid on regional polygons. The polygons formed by the intersections of these two polygons are determined, and all parameters within these intersection polygons are uniform.

2.2.3 Fish Movement

At each time step, the location of each population of fish is generated by drawing randomly from a discrete set of probabilities that represent each polygon within the domain. If there were no areas with stronger or weaker attraction (no difference in attraction factors) then these probabilities would simply be proportional to the area of each polygon. Attraction factors are used to weight the probabilities based on area to account for how fish are attracted to certain areas. All invertebrates, zooplankton, and phytoplankton do not change location, and hence are not included in the Monte Carlo generation of locations.

This set of probabilities, that describe the odds for different locations, does not change with time. Hence these probabilities are calculated before the transient simulation and then used each time step to move fish populations between polygons. Fishrand uses a three step approach: first randomly assigning whether a population is in the domain, second randomly assigning a population to an attraction polygon, lastly randomly assigning the population to a regional polygon that is located (at least in part) within an attraction polygon. It also allows for the possibility that a fish population is outside any attraction polygon.

1. Using the abundances from the Migratory_data sheet, FishRand determines if a fish population is in the domain at that certain time step. The

probability a fish population is within the domain is the abundance of that this at that time step.

2. The probabilities associated with each Attraction factor polygon is determined. Let D be the entire domain polygon. For Fish type t, there are attraction polygons $S_{t1}, S_{t2}, S_{t3}...$ indexed by k, with corresponding attraction factors $F_{t1}, F_{t2}, F_{t3}....$ The conditional probability give the fish population is within the domain of being in any given attraction factor is:

$$P(S_{tk}) = \frac{F_{tk} \cdot A(S_{tk})}{\sum_k F_{tk} \cdot A(S_{tk}) + (A(D) - \sum_k A(S_{tk}))}$$

where the probability that a fish is in an area in the domain outside of all the attraction factor polygons is:

$$P(D - \bigcup_k S_{tk}) = \frac{(A(D) - \sum_k A(S_{tk}))}{\sum_k F_{tk} \cdot A(S_{tk}) + (A(D) - \sum_k A(S_{tk}))}$$

3. Given that a population is in a particular attraction polygon, a second set of probabilities describes the likelihood that the population is in a particular regional polygon. These probabilities are found by overlaying the attraction factor polygons on the regional polygons $R_1, R_2, ...,$ to generate $S_{tk} \cap R_1, S_{tk} \cap R_2, ...$ The probability of a fish population being in a region R_w is then determined from the areas of all these polygons::

$$P(R_w) = \frac{A(S_{tk} \cap R_w)}{A(S_{tk})}$$

And, if the population is not in an attraction factor polygon:

$$P(R_w) = \frac{A((D - \bigcup_k S_{tk}) \cap R_w)}{A(D - \bigcup_k S_{tk})}$$

From these probabilities a region is determined, and then region specific properties are set.

At each time step, each Fish population takes on a new location by drawing from the above probabilities. Therefore each population has a different time series location history, and there are $T \cdot P \cdot m \cdot n$ number of Monte Carlo simulations of bio-accumulation per fish type, where T is the number of time steps, P is the number of populations, m is number of "uncertainty" samples and n is number of additional "variable" samples.

2.2.4 Operations

To find the concentrations in each organism FishRand needs to preform the algebraic operations in Bio-accumulation model. The result of generic algebraic operation $\odot$ $(+, -, /, *, \text{etc...})$ between a uncertain, and variable parameter matrix R is:

$$R_{ij} = x_i \odot y_{ij}$$

Between two variable parameters and two uncertain parameters normally defined matrix addition is used.

2.3 Fish Concentration Fitting

Fishrand produces a sample distribution of chemical concentrations for each fish type at each time step. After the Monte Carlo simulations have reached a specified time, Fishrand will fit a parametric distribution to the sample distribution.

For a given fish type, all the samples in all the populations are grouped to form a set of $P \cdot m \cdot n$ concentration samples. These samples are assembled to make a sample CDF:

$$F(R_{ij}) = \frac{\#\text{ elements} < R_{ij}}{P \cdot n \cdot m}$$

and then use non-linear least squares to fit find parameters for, Normal, Log-Normal, Uniform, and Gamma fits. Giving 4 distributions G indexed by v. To compare the 4 different fits FishRand uses the Kolmogorov–Smirnov test. The KS test finds, for each fit, the maximum difference between $F(R_{ij})$ and $G_v(R_{ij})$ and then minimizes that value over distribution types (v). Symbolically:

$$\underset{v}{\operatorname{argmin}}(\max_{R_{ij}} |F(R_{ij}) - G_v(R_{ij})|)$$

This fit is then consider optimal by FishRand, and is labeled in the GUI. The main advantage of using the KS statistic to find a good fit is that it is non-parametric, so that no prior assumptions about the underlying type of distribution are made.

3 Excel Input

Python FishRand accepts input from specifically formatted Excel spreadsheets. The basic formatted sheet (Which can be duplicated and passed to FishRand), can be found in the FishRand subfolder: sheets/input/Blank_Input.xlsx The excel spreadsheet consists of 8 different tabs, described below.

3.1 Sample and Time input

In the Time and Sample Input tab, statistical sampling options can be set, and the time step parameters are set. FishRand can be run in 3 different Sampling modes:

3.1.1 Deterministic mode.

In this mode, FishRand does not do Monte Carlo Simulations. Instead, it simply uses one point value for each input parameter. If you plan to run FishRand with no statistical input, set both Total number of Uncertainty samples and Total number of inner loop samples to 1. This insures that no repeated sampling of random variables occurs.

3.1.2 Monte-Carlo Mode but without distinguishing between variable and uncertain parameters.

To avoid distinguishing between variable and uncertain parameters, model all parameters as variable (To see how to input a variable parameter look at section 3.5.2) . Also, set Total number of Uncertainty samples to 1. The number of number inner loop samples, (inner loops number = total sample / uncertainty loop samples) then becomes the total number of samples taken from every distribution that is inputted. For good results make sure that inner loops samples is 500 or above.

3.1.3 Monte-Carlo Mode with Variable and uncertain statistical input.

The Monte Carlo Simulation has two loops: an outer loop for generating samples of uncertain parameters and inner loop for generating samples of variable parameters. Consequentially, the number of samples drawn for "Uncertain parameters" is equal to the number of iterations of the outer loop, and the number of samples generated for "Variable parameters" is larger, the product of the number of outer and inner loop iterations. For good results make sure that Total number of Uncertainty samples is set to at least 500 so that Uncertain samples are not under-sampled. Total number of Inner loop samples can be set to 1 or more. See section 3.5.2 to see how to input distributions.

3.2 Latin Hypercube bins

The number of Latin Hypercube bins can also be set. The number of bins corresponds the number of equally spaced partitions of each input distribution. A equal number of samples are taken from each partition, ensuring that each distribution is well sampled. The default is 10 bins. With a larger number of variable and uncertain samples fewer bins are required to give accurate output, as the chance the samples covering a larger portion of the distribution increase. See section 4.2.1 for a more in depth description of Latin Hypercube sampling.

3.3 Time Input

The beginning, and end times for the simulation are entered under the sampling inputs. Beginning and end times should be entered as "MM,DD,YYYY". The time steps are input as either "Days", "Weeks", "Months", "Quarters".

3.4 Steady State

Set the Steady State option to "YES" for a steady state solution of the model that uses Arnot, 2004 equation 2. Otherwise, for a transient model run, set Steady State to "NO" which will use Arnot, 2004 equation 1, and has the added capacity for spatially explicit modeling. Of course, you an also run the model out to approach steady state with enough time steps. If you have multiple sample sites, you must run the model temporally, because Arnot equation 2 does not take into account any spatial parameters. But, again, you can run the time steps out to find a steady state solution for multiple regions.

3.5 Parameter input formatting

In the next four input tabs, both non-statistical and statistical parameters are accepted.

3.5.1 Adding a deterministic parameter

To input a deterministic parameter locate the "Entry" column in which you would like to input. In the corresponding cell enter the number you would like for that parameter.

3.5.2 Adding a statistical parameter

Locate the "Entry" Column. In the corresponding cell, first input the parameter as either variable or uncertain by adding a "V" or a "U" respectively. Then after a comma and a space, input the distribution type you would like to use. (See Table 1 for distribution types).
In the distribution table: α, and β define the beta function with pdf: $f(x) = \frac{\gamma(\alpha-1)\gamma(x-1)^{\beta-1}}{\gamma(\alpha)\gamma(\beta)}$, and the Log-Normal parameters μ, and σ, are not the corresponding normal μ, and σ, but are the μ, and σ of the actual Log-Normal distribution.
Lastly, add another space, and inside parenthesis add the comma-space separated parameters corresponding to that distribution (See distribution table for parameterizations).

Examples:
For a normal distribution: V, Normal (5, 3)
For a Triangle distribution: U, Triangle (3, 5, 4)

Note: In some situations, if not enough statistical parameters are input, some output concentrations will be deterministic and some will be probabilistic. For example if only Large-mouth Bass weight is variable (and hence Bass output concentrations are probabilistic), Pumpkinseed concentrations will still be deterministic since Bass weight has no effect on Pumpkinseed concentration. In cases like these FishRand is currently not programmed to handle these situations. To avoid this problem make sure that input concentrations are probabilistic in every region.

Name	Parameters
Normal	μ, σ, min (optional), max (optional)
Uniform	beginning, length
Triangle	beginning, ending, peak
Log-Normal	μ, σ
Beta	α, β
Weibull	λ, k

Table 1: Distribution Table

3.6 Adding Multiple Objects

In the multiple objects (i.e. Fish, Sample Sites, Chemicals) can be input. A new Object can be created by copying and pasting a existing object directly below itself. For instance if we wanted to have two Fish in our model, the Fish section would look like:

Fish	Entry
Name	pumpkinseed
Weight (kg ww)	0.047
Lipid Content (kg/kg)	0.037
Non-Lipid Content (deafult = .2) (kg/kg)	
Fraction Pore Water Ventilated (deafult = 0)	
Dietary absorption efficency of lipid Organic Matter (deafult = .72)	
Dietary absorption efficency of nonlipid Organic Matter (deafult = .72)	
Dietary absorption efficency of water (deafult = .25)	
Feeding Rate if known, otherwise leave blank and estimated) (kg/d)	
Growth Rate (if known, otherwise leave blank and estimated) (d^-1)	
Filter Feeder Flag (0 = is, and 1 = is not)	
Fish	Entry
Name	bluegill
Weight (kg ww)	0.098
Lipid Content (kg/kg)	0.0202
Non-Lipid Content (deafult = .2) (kg/kg)	
Fraction Pore Water Ventilated (deafult = 0)	
Dietary absorption efficency of lipid Organic Matter (deafult = .72)	
Dietary absorption efficency of nonlipid Organic Matter (deafult = .72)	
Dietary absorption efficency of water (deafult = .25)	
Feeding Rate if known, otherwise leave blank and estimated) (kg/d)	
Growth Rate (if known, otherwise leave blank and estimated) (d^-1)	
Filter Feeder Flag (0 = is, and 1 = is not)	

3.7 Chemical Concentrations

In the Chemical Concentrations tab, the concentration of each chemical at each time step in each region is inputted. Each column from "B" onward represents a single time step. In the "A" column below the Time steps label (at A3) , add the name of the first region listed in the Regional Input tab, then add the name of the first chemical in the Chemicals Input tab in the cell below. Following the chemical name add the 4 cells labelling the different types of concentrations. Then add the next chemical in the Chemical Input tab, and repeat. In the columns to the left of "A" add the corresponding concentrations for each time step. Note: the region label in yellow is only input a single time, after the region label is input only green chemical labels are input until the next region is ready to be input below! The input all together will look something like this:

A	B	C	D	E	F
Timesteps →	0	1	2	3	
Pegan Cove					
PCB 101					
Concentration in Sediment (ng/g ww)					
Total Concentration in Water (ug/L)					
Dissolved Concentration in Water (ug/L)	5.00E-06	1.50E-05	3.00E-04	5.00E-06	5.00E-0
Concentration in Pore Water (ug/L)	2.65E-05	1.65E-05	3.00E-04	2.65E-05	2.65E-0
PCB 52					
Concentration in Sediment (ng/g ww)					
Total Concentration in Water (ug/L)					
Dissolved Concentration in Water (ug/L)	5.00E-06	1.50E-05	3.00E-04	5.00E-06	5.00E-0
Concentration in Pore Water (ug/L)	2.65E-05	1.65E-05	3.00E-04	2.65E-05	2.65E-0
Pegan Cove Dummy					
PCB 101					
Concentration in Sediment (ng/g ww)					
Total Concentration in Water (ug/L)					
Dissolved Concentration in Water (ug/L)	5.00E-06	1.50E-05	3.00E-04	5.00E-06	5.00E-0
Concentration in Pore Water (ug/L)	2.65E-05	1.65E-05	3.00E-04	2.65E-05	2.65E-0
PCB 52					
Concentration in Sediment (ng/g ww)					
Total Concentration in Water (ug/L)					
Dissolved Concentration in Water (ug/L)	5.00E-06	1.50E-05	3.00E-04	5.00E-06	5.00E-0
Concentration in Pore Water (ug/L)	2.65E-05	1.65E-05	3.00E-04	2.65E-05	2.65E-0

3.8 Organism Diets

In the Organism Diet tab, each Fish and Invertebrates diet must be input. The first row of a Diet Entry specifies the organism we are created the diet for. Leave the Fraction column blank for this row. In subsequent rows a fraction of the diet can entered as a decimal, or should be entered as 0 if this organism does not eat the other organism. The fractions much sum to 1 for all Fish and Invertebrates. The ordering of Diet Entries must move up the food web, otherwise FishRand will crash.

3.9 Spatial Modeling

The Last two tabs of input are for spatial modeling. If only a single Sample site is input, or the model is in steady state, these tabs can be ignored. In the Migratory Input tab at each time step, the fraction of Fish Populations present in the whole domain can be input as a decimal between 0 and 1. In the Sample Sites tab the regional areas, and fish attractions are input. In the first row the boundary of the entire Domain is input as vertices of a polygon in clockwise order, starting at any point. Below the Domain input, a coordinate must be entered for each sample site to create a Thiessen polygon map. Coordinates and vertices are entered as x, y locations. Make sure your coordinates are within the Domain polygon, and the format of the entry is x "commma" "space" y. Lastly, Attraction Areas, and number of fish populations are input. To input attraction factors set the definition property to either "Polygon" or "Fraction", then input its name, associated fish, attraction factor and coordinates, or fractions below. If the definition property is set to fraction, the fractions inputted are simply the fraction of each sample site in which the attraction area is located. More coordinates or fractions can be input by adding them to the corresponding further right columns. Make sure your attraction areas coordinates are input within the boundary.

Example:

Attraction Facto	Assosiated Fish	Factor	Coord 1 or Fraction of Sample 1 Area	Coord 2 or Fraction of Sample 2 Area	Coord 3 or Fraction of Sample 3 Area
dummy yellow p	yellow perch	1	98, 50	3, 339	8, 395
dummy bluegill	bluegill	1	98, 50	3, 339	8, 395
dummy largemo	largemouth bass	1	98, 50	3, 339	8, 395

Each fish must have at least one corresponding area attraction area. This is a downside, but If you wish to not input a attraction area for a fish make 'fake' coordinates or weights, and set the attraction Factor to 1.

3.10 Important Notes on Solving

- There are two different ways of inputting Dissolved Concentration of a Chemical in Water (g/L):
  1. You can directly input it by entry into its cell in the Chemical Concentrations tab.
  2. You can input:
     - Total Concentration in Water in Chemical Concentrations
     - Disequilibrium factor of Dissolved Organic Carbon in the Corresponding cell of Chemical Input
     - Disequilibrium factor of Particulate Organic Carbon in the Corresponding cell of Chemical Input
     - POC–octanol proportionality constant in corresponding cell of Regional Input
     - DOC–octanol proportionality constant in corresponding cell of Regional Input
- If any organisms diet is partly made up of sediment, then both the Fraction of Organic Carbon Content in Sediment, and Concentration in Sediment (ng/g) must be input.
- Some parameters in the excel spreadsheet are labelled with default values that they will take on if nothing is entered.

4 FishRand App

The App is where you can run the excel input file, and view information about the model.

4.1 Opening the App

The App is opened by either double clicking "FishRand_windows" or "FishRand_mac_linux".

4.2 Loading Input

In the loading "Input" section of the App, use "Choose File" button to find the excel input file.

The "Timesteps to Save and Display" section can be left blank if a steady state model is being ran. Otherwise the timesteps are integers representing the number of timesteps since the start, and can be input. Multiple timesteps to save and display can be typed with a comma and a space between them. For instance: '4, 10, 30'. Then click run to run the model. **** more about loading screen, and bugs****

4.3 Visual Output

After FishRand finishes running, different output from the model can be displayed.

A map of the regions can be displayed and saved, by clicking "Show Region". A diagram of the food web can be displayed by clicking "Show Food Web".

In the View Distributions section, of the Timesteps specified in the Input section, concentrations of in Fish can be displayed. In the three pull down menus select the Time, Fish, and Chemical you would like to see the concentration distribution of.

Since FishRand uses a KS-statistic to attempt to fit the best distribution to the calculated concentration samples, you can choose what types of distributions you would like to see on the right side of the View Distributions panel. After you have selected these options, click "Show Distributions" to see the different distribution fits.

If you would instead like to see the concentration of a certain fish over time, you can click the "Show Time Graph", as long has you have a specific fish and chemical selected on the pull down menus to the left.

Note that if you run a model with Steady State set you yes, the there is no visual output the only output is in excel.

4.4 Excel Output

The results of the model can be saved to a excel spreadsheet. If you have specified statistical in the "Type of Distribution to save" pull down menu you can select the type of fit that the excel spreadsheet will report. I recommend that you should choose KS Best which saves the best fit distribution for each concentration. Next click the Save button, which you bring you to a menu where you can specify the location and name of the output file. The information in the output file will vary depending on whether statistical input is specified, and if the model was temporal or not.

5 Tests

The different test sheets can be found in the "sheets/input/tests" Each test involves running different inputs and comparing the results. The test can be run to make sure that your version of FishRand is working, or to demonstrate that this version of the FishRand model is correct in some aspect.

5.1 Test for Convergence to Equilibrium with Arnot Equations

In the "Arnot_Match_DO_NOT_EDIT.xlsx" all the input parameters of the bioaccumaltion model are set to match the parameters in "AQUAWEBv1.2_DO_NOT_EDIT.xlsx", the spread sheet provided by Arnot in Arnot, 2004. If the Arnot Match spread sheet is run out 20 weeks or longer one can observe in the output spreadsheet that the concentrations of organisms in this spreadsheet are approaching the steady state concentrations in the aquaweb spreadsheet.

5.2 Test for Same Results Between Total Water Concentration and Dissolved Water Concentration, and Sediment Concentration and Pore Water Concentration

In the "Arnot_match_Sediment_Total_Water.xlsx" spreadsheet chemical concentrations are input as concentrations in sediment and total concentrations in overlying water. In the "/Arnot_match _Pore_water_Dissolved_Water.xlsx" spreadsheet chemical concentrations are input as concentrations in pore water, and dissolved concentrations in overlying water. If both spreadsheets are run and output sheets generated, one can observe that the output concentrations are the same with a small rounding error due to not adjusting the input decimals to exactly match the outputs. For more information on the two types of input see 3.10.