Python FishRand User Manual

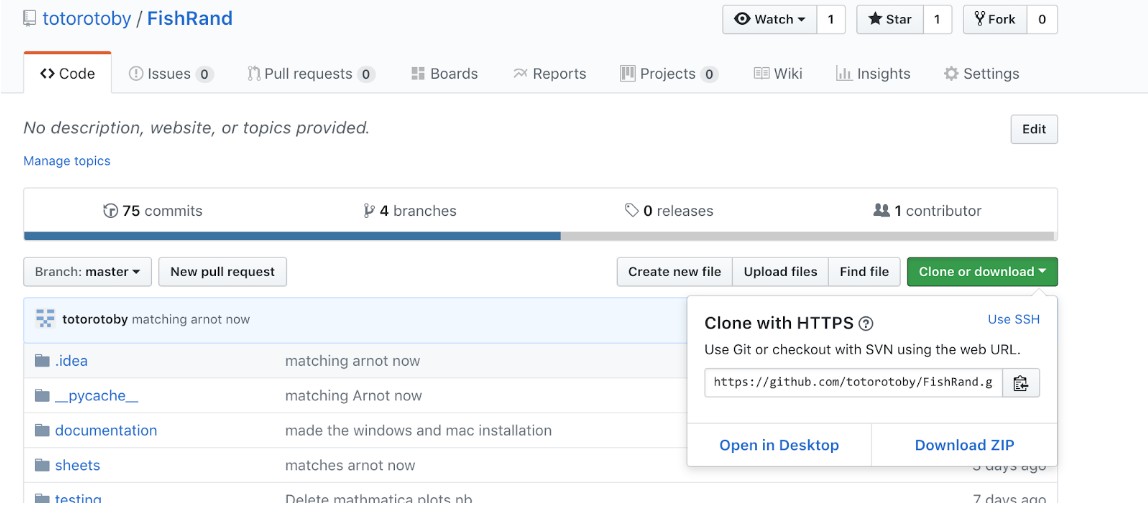
November 2018

**Contents**

1. [Installation 2](#_TOC_250023)
   1. [Getting FishRand 2](#_TOC_250022)
   2. [Windows installation 2](#_TOC_250021)
   3. [Mac Installation 4](#_TOC_250020)
2. [Excel Input 5](#_TOC_250019)
   1. [Sample and Time input 5](#_TOC_250018)
      1. [Deterministic mode. 5](#_TOC_250017)
      2. [Monte-Carlo Mode, but without distinguishing between variable and uncertain parameters. 6](#_TOC_250016)
      3. [Monte-Carlo Mode with Variable and uncertain statistical input. 6](#_TOC_250015)
   2. [Latin Hypercube bins 6](#_TOC_250014)
   3. [Time Input 6](#_TOC_250013)
   4. [Steady State 6](#_TOC_250012)
   5. [Parameter input formatting 6](#_TOC_250011)
      1. [Adding a non-statistical parameter 7](#_TOC_250010)
      2. [Adding a statistical parameter 7](#_TOC_250009)
   6. [Adding Multiple Objects 7](#_TOC_250008)
   7. [Organism Diets 8](#_TOC_250007)
   8. [Spatial Modeling 8](#_TOC_250006)
   9. [Important Notes on Solving 9](#_TOC_250005)
3. [FishRand App 9](#_TOC_250004)
   1. [Opening the App 9](#_TOC_250003)
   2. [Loading Input 9](#_TOC_250002)
   3. [Visual Output 10](#_TOC_250001)
   4. [Excel Output 10](#_TOC_250000)
4. **Model Structure** **10**
   1. Bio-accumulation Model 10
   2. Monte Carlo Simulation 11
      1. Sampling 11
      2. Spatial Sampling 11
      3. Operations 13
   3. Fish Concentration Fitting 13

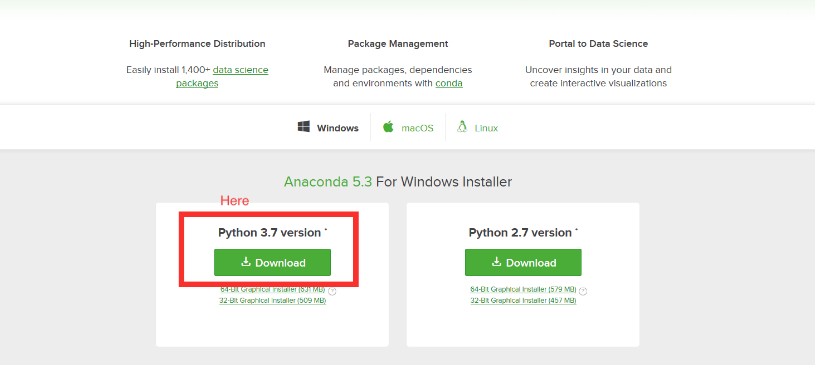
# Installation

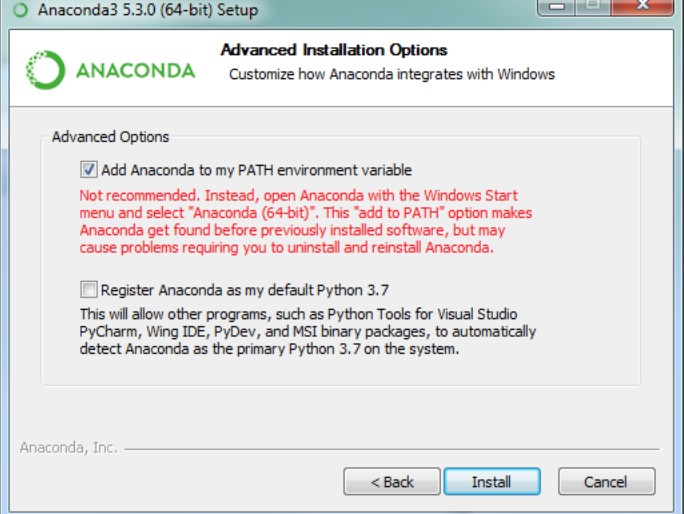
## Getting FishRand

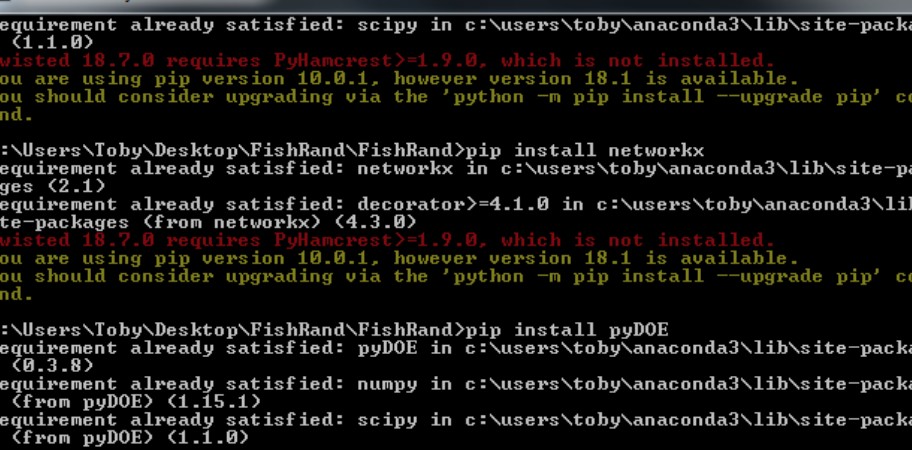
* + 1. Go to https://github.com/totorotoby/FishRand
    2. Download the directory as a ZIP, from the ”Clone or download” button.
    3. Unzip the folder in whatever location you downloaded it to.

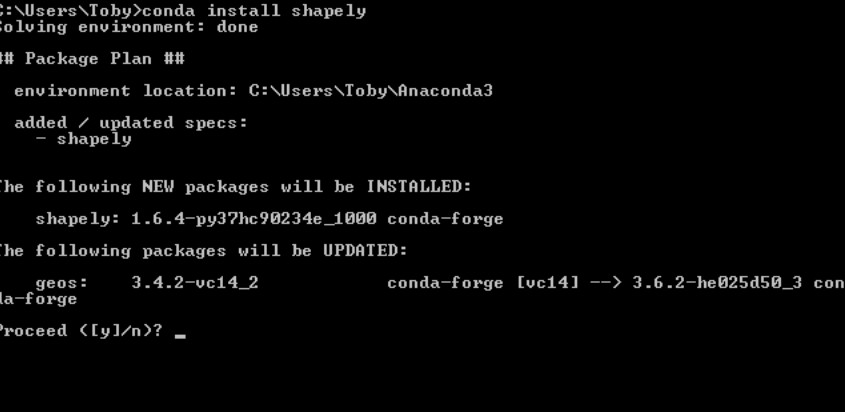
## Windows installation

* + 1. Go to: [www.anaconda.com/download/,](http://www.anaconda.com/download/) and download the python 3.7 ver- sion.



* + 1. Install Anaconda. In the installation prompt check “Add Anaconda to my PATH”.
    2. Locate “windows installer.bat” in your FishRand folder, and double click that. A command prompt window should appear and run multiple com- mands. (There may be some warning messages, ignore these.)

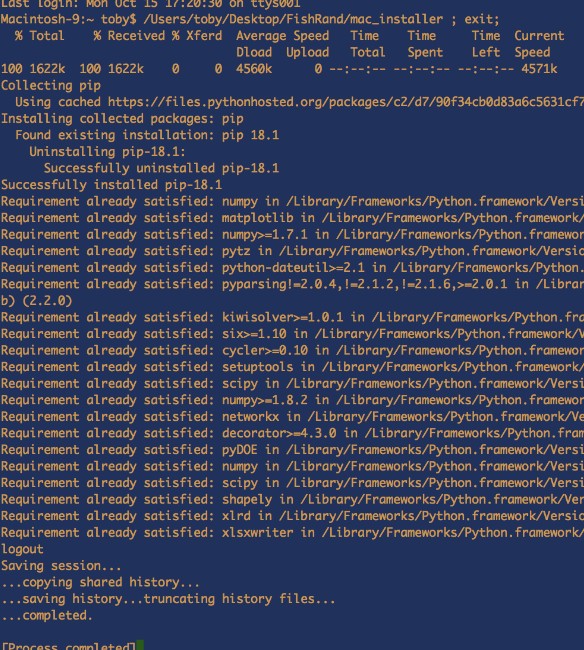


* + 1. Open another command prompt window, and type “conda install shapely” and hit return. After a while there will be some text asking for a “y/n”. Type “y” and hit return

## 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](http://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 win- dow should appear and execute a bunch of commands.



* + 1. 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).

# Excel Input

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

## Sample and Time input

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

### 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.

### Monte-Carlo Mode without distinguishing between vari- able and uncertain parameters.

To avoid distinguishing between variable and uncertain parameters, model all parameters as variable. Set Total number of Uncertainty samples to 1.

### Monte-Carlo Mode with Variable and uncertain statistical in- put.

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.

## Latin Hypercube bins

The number of Latin hypercube bins can also be set. The default is 10. With a larger number of samples fewer bins are required to give accurate output. See section 4.2.1 for a description of Latin Hypercube sampling.

## 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 step are defined in either ”Week”, ”Month”,

## Steady State

Set the Steady State option to ”YES” for a steady state solution of the model. Otherwise, for a transient model run, set Steady State to ”NO”. Of course, you an also run the model out to approach steady state with enough time steps. If you have multiple zones, then no steady state is possible because ….{explain}

## Parameter input formatting

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

Name Parameters

Normal *µ, σ*

Uniform beginning, length Triangle beginning, ending, peak

Log-Normal *µ, σ*1

Beta *α, β*2

Weibull *λ, k*

Table 1: Distribution Table

{move tale down to section 2.5.2}

### Adding a deterministic parameter

To define 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.

### Adding a statistical parameter

Locate the ”Entry” Column. In the corresponding cell, first define the parameter as either variable or uncertain by adding a ”V” or a ”U” respectively. Then after a comma and a space, define the distribution type you would like to use. (See distribution table for distribution types). Entry examples: ”V, Log-Normal” ”U, Weibull”, ”V, Beta”. 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, Triangular (3, 5, 4)

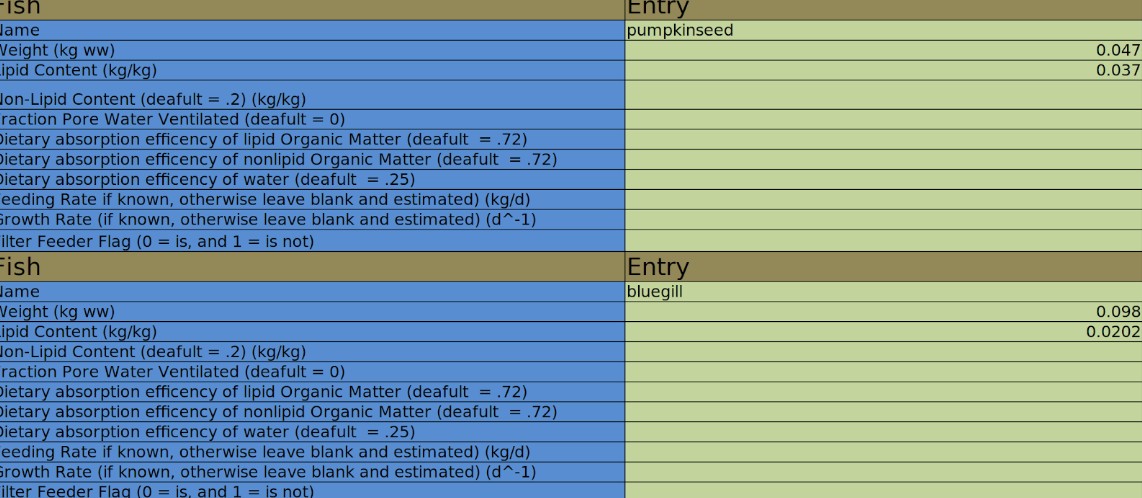
## Adding Multiple Objects

In tabs 1 through 7 multiple objects (i.e. Fish, Sample Sites, Chemicals) can be defined. 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:

1The 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.

2*α*, and *β* define{the math has a an b –fix} the beta function with pdf: *f* (*x*) = *γ*(*α−*1)*γ*(*x−*1)*b−*1

*γ*(*a*)*γ*(*b*)



{the edges of this figure are cut off and also latter figures}

## Organism Diets

In the Org diet tab, each Fish and Invertebrates diet must be defined. 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.

## Spatial Modeling

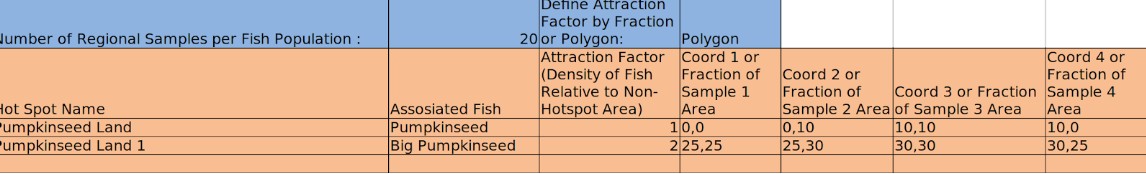
The Last two tabs of input are for spatial modeling. If only a single Sample site is defined, or the model is in steady state, these tabs can be ignored.

In the Migratory data tab at each time step, the fraction of Fish Populations present in the whole domain can be defined as a decimal between 0 and 1.

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 are entered as x,y locations. Make sure your coordinates are within the Domain polygon.

Lastly, Attraction Areas, and number of fish populations are input. To define attraction factors set the Defintion property to either ”Polygon” or ”Fraction”, then define its name, associated fish, attraction factor (The Density of fish relative to the outside area. For instance if a attraction factor is 2, this is equivalent to saying the density of associated fish in the Hotspot is twice that of the outside area) and coordinates, or weights below. More coor- dinates or weights can be defined by adding them to the corresponding further right columns. Make sure your Hotspots are defined within the boundary.

Example:



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

## Important Notes on Solving

There are two different ways of defining Dissolved Concentration of a Chemical in Water (g/L):

*•*

1. You can directly define it by entry into its cell.
2. You can define:
   * Total Concentration in Water
   * Disequilibrium factor of Dissolved Organic Carbon
   * Disequilibrium factor of Particulate Organic Carbon
   * POC–octanol proportionality constant
   * DOC–octanol proportionality constant

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 defined.

*•*

Some parameters in the excel spreadsheet are labelled with default values that they will take on if nothing is entered.

*•*

* The model currently assumes unfiltered water.

# FishRand App

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

## Opening the App

The App is opened by either double clicking “FishRand windows” or “FishRand mac linux”.

## 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\*\*\*\*

## 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.

## 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.

# Model Structure

* 1. **Bio-accumulation Model**

Similar to previous versions of FishRand, the Python version has a Bio accumu- lation 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 Propor- tionality Constants, and Dietary Absorption efficiencies).

* 1. **Monte Carlo Simulation**

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

* + 1. **Sampling**

Each distribution is of type Uncertain or Variable, and is sampled in the follow- ing way:

Let *X* = *Dt*(*p*1*, . . . , pe*) be the Distribution of type *t* (uncertain or variable) with parameters *p*1*, . . . , pe*. Then random realizations are drawn from *Dt* by parti- tioning *Dt* 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:

 

*x*1 *x*2

 

*VX* = ... , where *xi* is the *i*th sample from *X*, with the following property:

*xn*

If *b* is the number of partitions of *X* indexed by *k*, then from each partition *n* samples are drawn from *Xk* where *Xk* is the *k*th partition of *X*. Each sample is then placed at index *i* of vector *VX* with uniform probability.

*b*

If the Distribution is Variable FishRand creates a matrix:

*y*11 *y*12 *. . . y*1*m y*21 *y*22 *. . . y*2*m*

 

*MY* = 

...

...

. . . ...



*yn*1 *yn*2 *. . . ynm*

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

* + 1. **Spatial Structure**

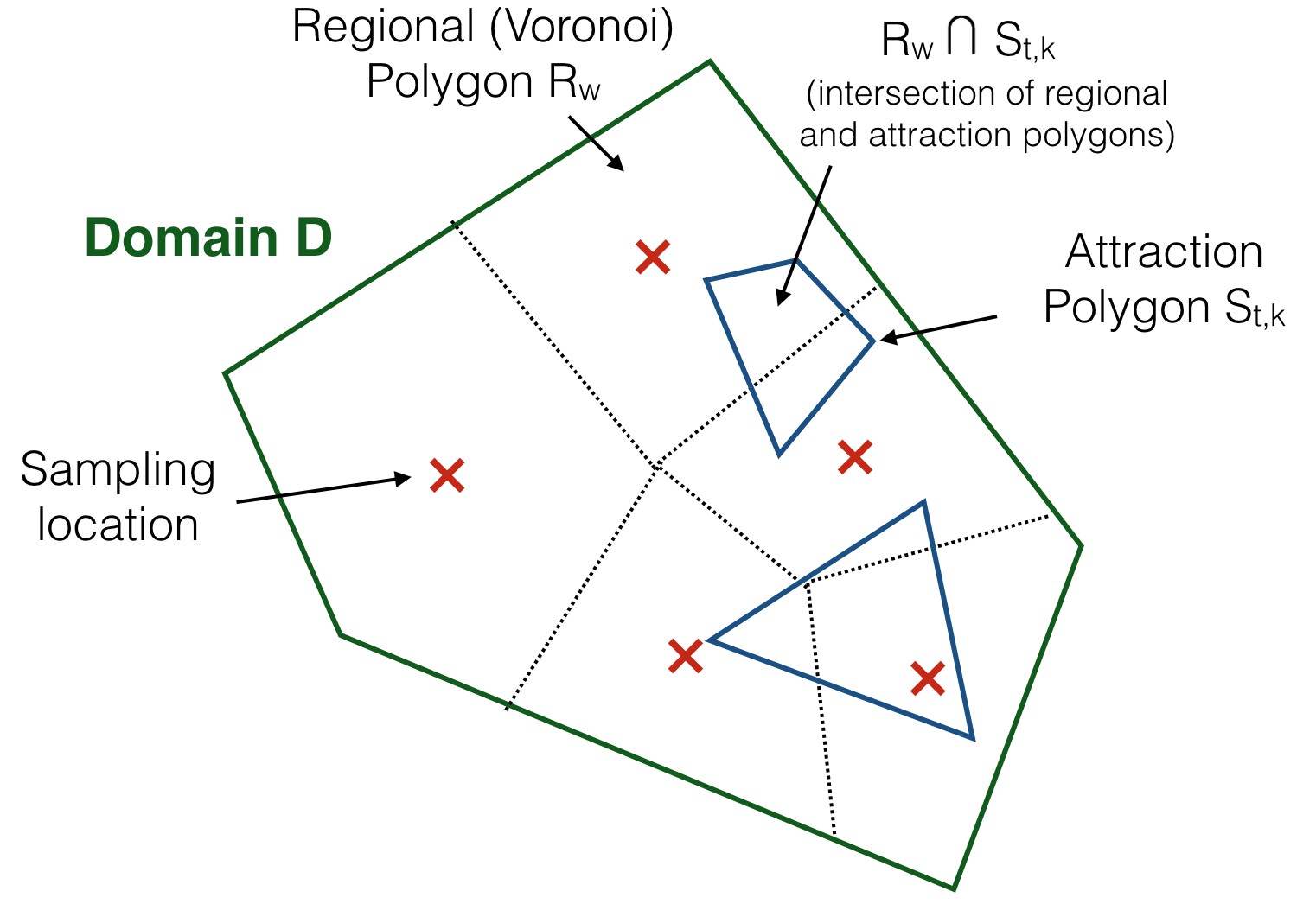
Fishrand constructs Regional Polygons *Rw* to map areas with the parameter values associated with each measurement location. (see figure ?). for ,*Rw* eachthat are alsoDomain area Regional Polygonis(to be )

The polygons for different attraction areas are then overlaid on regional polygons. The polygons formed by the intersections of these two polygons are determined (figure ?). All parameters are uniform within these intersection polygons.

**4.2.3 Fish Movement**

At each time step the location of a population of fish must be determined (All invertebrates, zooplankton, and phytoplankton do not change location so no additional sampling is done). If there are *P* populations of a type of Fish (This is set in the Sample Sites tab), then there are *Pnm* Monte Carlo simulations per Fish type per time step. By specifying a fish populations location all location dependent parameters are also specified to run the Bio-accumulation model (i.e. all input from the Regional Input tab, Temperature, and Chemical Concentrations). A fish populations location an and by extension the parameters listed above are found in a four step process, and is done for *P* populations per fish:

Attraction factor polygon in which a fish population is located is de- termined. Let *AD* be the area of the entire domain. For Fish type *t*, there are attraction polygons *St*1*, St*2*, St*3*...* indexed by *k*, with corresponding attraction factors *Ft*1*, Ft*2*, Ft*3*...*. The conditional probability give the fish population is within the domain of being in any given attraction factor is:



* + - 1. Using the abundances from the Migratory data sheet, FishRand deter- mines 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.

*Ftk · A*(*Stk*)

*P* (*Stk*) =

*k*

*Ftk*

* *A*(*Stk*

) + (*AD*

*− k*

*A*(*Stk*))

* + - 1. Once the fish population is assigned to a attraction polygon or lack there of, the corresponding polygon is overlaid on the regional polygons *R*1*, R*2*, ...*,

to generate *Stk ∩R*1*, Stk ∩R*2*, ...* The probability of a fish population being

in a region *Rw* is then:

*P* (*Rw*

*A*(*Stk ∩ Rw*)

*A*(*Stk*)

) =

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 according to the above algorithm. Therefore each population has a different time series location history, and there are *TP mn* number of Monte Carlo simulations per fish type, where *T* is the number of time steps.

* + 1. **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 (+*, , /, ,* etc...) between a uncertain, and variable parameter ma- trix *R* is:

*O − ∗*

*Rij* = *xi O yij*

Between two variable parameters and two uncertain parameters normally de- fined matrix addition is used.

**4.3 Fish Concentration Fitting**

Once all Monte Carlo simulations have been run up to a specified time step FishRand attempts to fit a distribution to the resulting chemical concentrations in the fish. FishRand the *P mn* simulations from a single time step, to construct a sample cdf where:

*F* (*Rij*

# elements *< Mij*

) =

*nm*

and then use non-linear least squares to fit find parameters for, Normal, Log- Normal, Uniform, and Gamma fits. To compare the different fits FishRand uses the Kolmogorov–Smirnov test. The KS test finds, for each fit, the maxi- mum difference between *F* (*Rij* ) and *G*(*Rij* ) and then minimizes that value over distrubtion types or

argmin(max *|F* (*Rij* ) *− G*(*Rij* )*|*)

dist type *Rij*

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.