

# Agent Based Model of Salmon Migration in the Snake River

Colin McCarthy

May 2017

## Abstract

Agent based Modeling in biological systems is a relatively new approach that has been used increasingly in examining environments that are difficult to describe through traditional statistical models. Studies in ecology have been particularly eager to employ this technique as the emphasis on the agents allow for studying conservation efforts with measurable effects on the population as a whole. In this paper, we showcase an agent based model of the process of salmon returning to their freshwater spawning ground as they pass through the Lower Snake River. Based on a model originally developed by Amy Steimke, we have updated the model to encompass more recent fish run data and provide a statistical analysis of the model that was not in Steimke's original project. In the process of working with the model, we produced a distribution that suggested that the dam objects were causing the salmon population to spread out, a dimension of difficulties to salmon reproduction that is not captured purely in the threat to life that dams pose. After quantifying this observed spread, we developed an addition to the model for opportunity-based reproduction to capture the effects of such a spread on the reproductive success of the salmon population. We find that the submodel results suggest a clear impact of dams upon the reproductive success. This significant finding leads us to conclude with potential implications and expansions to the work presented here.

## 1 Introduction to Agent Based Modeling

Salmon behavior and response to environmental change has been an area of focus in the Pacific Northwest, due to the life cycles exposing the salmon to a variety of environmental dangers. The adjunct addition of dams to the waterways has generated scientific interest in the effects of damming and subsequent restoration approaches have had on the salmon's reproductive success. Initially, the use of Agent Based Modeling provides a novel approach to the quantification of such environmental disruptions, but observations made in the model suggest a problem designed for the ABM approach due to agent interactions. Due to the spreading out of the population observed in the model, a submodel is presented to explore additional impacts on reproductive success of the salmon. Basing the model on a successful model created by Amy Steimke in the programming language netLogo, we extend the model to encapture the full effects of the population spread due to dams and the full scale life cycle. Furthermore we will check the validity of the model based on previous year counts provided by the dams that we are modeling. Ultimately, a successful model is presented to more accurately represent salmon in the Snake River, and provide a basis for further research into the effectiveness of restoration efforts.

### 1.1 ABM strengths and weaknesses

Agent based Modeling (or as it is referred to in ecology, individual-based modeling) is a computational technique that examines the interactions between an individual and its environment, creating the "macroscopic regularities" that emerge as patterns of agent interactions [4, P. 5]. Agent Based models consist of two self-contained subsets, the environment modeled, which is updated by a set of rules that set the current state, and the agents, which make decisions based on a set of behavioral trends and responses to the environment [4, P. 5]. An agent represents an individual organism in the model amongst a larger population. Typically when modeling ecology based systems, the unit of interest is the life form in question(our agent). Because agents have a defined set of properties before they are placed in their environment, heterogeneity among individuals are supported in the model [2, P. 83] allowing us to represent a trait with normal distribution spread across a system.

## 1.2 A history of ABM

Agent Based models can be traced back to von Neumann’s “cellular automation” in the 1940’s that outlined interaction between single units. In this early model von Neumann posed the question of whether it was possible to create a system that could self-replicate, a challenge that was answered in the 1970’s by the mathematician John Conway. In Conway’s “Game of Life”, the model ruled that cells would reproduce with enough space (representing access to nutrition) to do so in the environment or would die off if there was not enough space around the cell, and similarly would die off if there were too few neighbors representing under population [9, P. 130]. Depending on the initialization of the environment, a variety of patterns of growth emerged giving insight into patterns of growth found in nature. In this early example, each cell was treated as its own agent, determining its behavior from the environment. This work acts clearly as an early example of the techniques behind ABM. This style of modeling was picked up by Schelling in 1978 in a model that analyzed housing segregation by “scattering black and white chips and then moving them around on what amounted to a large chess board” [10, P. 6]. Each chip would either stay at its current position, or move to a random empty spot on the board based on what percentage of the neighbors were of similar color. Depending on the sensitivity of the rules applied to the chips, the board became upwards of sixty percent segregated. Schelling’s work brought attention to the power of agent-centric modeling techniques, and spread the practice to many other fields of study. Nowadays the techniques of ABM have been brought to model ecological systems [7], political parties [10], environmental justice movements [4], geology [2], and other forms of social collaboration [9].

## 1.3 Rigor in modeling

An issue arises when we bring any model to bear on a system regarding the validity of the technique. As mathematicians, we often hang our hat on the rigor that a proof brings to any problem. A well-executed proof acts not only to convince the reader of the outcome, but also acts at every step to make statements that are objectively “correct”. However, when it comes to applying models there is no guarantee that the same level of rigor is applied. With the use of ABM modeling, we face the same issue as all modeling techniques. This trade off between the rigor of a proof and the realistic picture of an ABM model can be lessened by “bringing a new standard of rigor, for computational as opposed to formal analytical work, into the equation” [10, P. ix]. This type of rigor is created by the precision required by the model to accurately represent its model system. “[A]gent based models have to be complete, consistent, and unambiguous if they are to be capable of being executed on a computer” [4, P. 6]. This form of rigor is closely tied to the Curry-Howard correspondence (the direct relationship between computational programs and proofs, stating both are built from the same logic and thus programs are a form of proof) to insure our eventual code has a degree of traditional rigor. Thus, so long as our model is constructed with care we can assert a degree of certainty over our result. With this in mind we can now examine the salmon and how we bring ABM to bear.

## 1.4 Systems that apply to modeling

As discussed above, systems that lend themselves to having individual actors are especially apt to Agent Based models. There are a few basic assumptions innate to working with an ABM. The environment must be able to be defined by a set of state equations which confines the agent. Furthermore, agents are assumed to act on “adaptive learning”, in contrast to “forward-looking strategic analysis”, and agents are assumed to have “bounded rather than perfect rationality” in the way they make decisions regarding the environment [10, P. 5]. As such, we make sure that agents act only on the information available to them at the time rather than predicatively. This is why ecological systems and organisms can make especially good systems for study, and why this was one of the earliest areas of study for ABM.

In the research presented here we model the ecological system of the Snake River with salmon set as agents. Of particular interest, we focus on the agents representing the salmon population in the Snake River. This is an area of particular interest due to the recent damming and destruction of such dams in salmon habitats, and investigating the effects of such environmental manipulation on the salmon themselves. While there has been research done on salmonid growth in natural waterways [15], and research on the effects on resource conflict due to obstruction of waterways [1], there has not been an in-depth look at the salmon population with regard to this waterway obstruction using Agent Based Modeling techniques. Advancing from a baseline model constructed by Amy Steimke ??, we account

for more detailed interactions between the salmon and a more complex environment. We also verify Steimke’s model and the new model with the 2012 data used in the previous model and with the most recent 2015 data.

## 2 The Chinook Salmon

The salmonids that are found in the Snake River are the Chinook, Jack Chinook, Coho, Sockeye, Steelhead, Wild Steelhead, Lamprey, and Shad varieties that are commonly found in the Pacific Northwest. We will assume a level of uniformity in aspects of their interaction with the environment unless otherwise noted, with the Chinook being our main focus due to its prevalence in the environment.

### 2.1 A brief summary of the life cycle of salmon

The life cycle presented here is taken largely from the National Parks Service [22]. Salmon start their life cycle as small eggs laid in freshwater environments. Each female lays between 1,500 to 10,000 eggs, and only twenty percent of laid eggs develop [17]. The eggs lay in the river bed as the embryos develop. In the spring the eggs hatch, and the small fish called Alevins survive in the gravel eating the remains of their embryo. As they leave the gravel the fish grow to be Fry which stay in the freshwater stream developing for a time determined by their species. Sockeye salmon spend 1-2 years after migrating to a lake, Coho salmon spend around a year, and Chinook less than five months. After this time the salmon are considered to be adult as they migrate towards the ocean. The process of swimming downstream can be blocked by unnatural obstruction, or natural predation. Once the fish have reached salt water, they begin a process of adapting to the environment called smolting. However, only two percent of hatched eggs make it through the maturation process to become adult and make this journey [17]. Depending on the species, the salmon spend a variable period of time in the oceans before returning to their spawning ground. Sockeye spend two years in ocean water, Coho spend up to seven years with an average of four years, and Chinook can spend up to eight years. The salmon travel back to the freshwater inlets from which they originated, and upon reaching the inlet stop feeding and survive until reproduction purely on the accumulated energy stored in their bodies. The journey takes on average 15 days, although the role of artificial barriers greatly slows down this process. Of particular focus in our model is the variety and survival effects of said barriers that stand between the salmon and their spawning ground. Once they reach their spawning ground, the females build nests for their offspring, while the males fight for dominance. After mating the adult individuals die and the cycle repeats into the next generation.

### 2.2 Dangers to the Chinook

Salmon are a rather odd species behaviorally. The great distances traversed in an effort to reach their spawning grounds places them at great risk for changes in the environment negatively impacting their survival rate. The Salmon travelling upstream have a limited amount of energy due to the aforementioned lack of feeding after returning to fresh water. Time spent stuck on obstacles ticks away precious energy for the salmon’s survival, whether it be from dams blocking, or the flow rate of the river from dam release. Higher level of flow can lead to a shorter travel time for salmon, but too high and the river can prevent returning to the spawning ground. Water velocity can also impact oxygen content, as well as temperature effects directly on the habitability of the environment. Dams also present a danger directly other than obstruction. When travelling to salt water, the turbines can easily pull in fry and kill them, or create pools of water that are more easily predated. The return home includes learning to use salmon ladders or salmon cannons, a skill that can be increased by successive successful dam crossings. It is worth noting that there is no significant difference between genders in ability to travel upstream, so gender was ignored in agents. Thus, in our model we will choose to focus on the times traversing up the river only leaving the rest to be done in further research.

### 2.3 The Snake River

The Snake River is a branch of the Columbia River that runs from Wyoming to Washington where it enters the Columbia and runs to the Pacific Ocean. Of particular interest is the stretch of the Snake river that runs through Washington state to three major tributaries, the Clearwater, Tuccannon, and Grande Ronde River [14]. Each has been documented as a major site for salmon spawning grounds,



Figure 1: The Dams Included in the Model

and thus makes the stretch of the Snake leading up to these tributaries a major thoroughfare of salmon traversal as they return to their tributaries. However, starting in 1909 with the construction of the Swan Falls Dam, a series of large hydroelectric projects were set up to dam the Snake river resulting in a total of fifteen dams being constructed along the length of the Snake [8]. These dams pose hazards for wildlife to traverse the river as they have historically done, with the salmon of the Snake being particularly impacted due to their life cycle. We will in our model focus on the four dams that block the Lower Snake as they directly block access to the three tributaries mentioned above. All four dams were constructed between the years 1962 to 1981, and range from 9.7 miles from the Columbia River to 107.5 miles up the Snake. In order of distance, the dams of interest to our model are the Ice Harbor, Lower Monumental, Little Goose, and Lower Granite Dam. Before their construction Washington state already had in place a law requiring all obstruction of waterways to include a fish passageway, and as such all four dams offer fish ladders for Salmon to traverse up to bypass the dam. However, the physical barrier and threat of turbines still impede the salmon returning to their spawning grounds, and thus their ability to reproduce. We assume that the obstacles that block the salmon can be learned to be more readily navigated, rather than being a constant chance of passing the barriers. In that regard, it is worth mentioning that there are four dams along the stretch of the Columbia before the salmon reach the Ice Harbor dam, as such salmon in the model are expected to have some form of dam learning before starting the model (as we only have counts from the number of fish that have successfully passed the Ice Harbor Dam). With this in mind we move into a discussion of our model.

### 3 Moving to a model

For the construction of our model we will be using the modeling software NetLogo, a software package designed for the purpose of creating multi-agent models. To better understand the final model we first explore what NetLogo has to offer, and then turn to an exploration of Steimke’s model. This will allow us to introduce our new model before moving to verify both models and discuss the data presented in the system. As with any model a move towards higher specificity will always make more assumptions of the system, so there will be a conscious effort to point out where each of these assumptions occur. While this section will be innately more technical in nature, the decisions in the model should naturally follow from the system we have detailed above.

#### 3.1 NetLogo

NetLogo is a software package designed and distributed by Northwestern University, maintained by the Center for Connected Learning and Computer Based Modeling (CCL), and originally created in 1999 by Uri Wilensky who was at the time working for Tufts University [12]. The software presents a user interface for modification directly, as well as access coding with specific NetLogo syntax. Ultimately the

code is translated into Java byte code when run in the model through a built in compiler [12]. In NetLogo two distinct units are presented in the code, mirroring its close ties to ABM fundamentals. “Turtles” are the collection of agents with each agent being a single turtle, where each agent can be referenced by an indexing number identifying the agent in the population. Individual agents can be singled out, or the collection can be universally updated by some general action. For example, the code

```
ask turtle 11 [set color green]
```

Sets the individual agent referenced as turtle 11 to override its current color setting and instead display “green” by use of the primitive color command (primitives being variable names that NetLogo already understands). If instead we looked to change all agents, we simply leave out the indexing number and apply it to all turtle objects. For models with more than one type of turtle object, variable names for types of turtles can be referenced in the reference to the turtle object. The other distinct unit is termed “patches”, where each patch is an element of the environment. While it may be clunky to consider each element of the environment an individual part of the model instead of a uniform set, this allows for each location in the environment to have a different state that can be tracked and referenced. Patches are referenced by an  $x$  and  $y$  coordinate, so to alter a specific patch we would simply type

```
ask patch 37 0 [set pcolor black]
```

Which would ask the patch at (37,0) to change its color to be black. NetLogo uses the convention that the element (0,0) is the far right patch that is horizontally in the middle. The final element of our model is what NetLogo deems a “tick”, or one execution of the code loops. A tick is one run through of the actions described in the “go” component of the code that tells how the model to run. Because of this, all NetLogo models are in some way dependent on a factor of time, as a tick allows the progression of the model. Other dependencies can be created from the tick that can then drive the model, but fundamentally all models are run on a tick based system for updating the turtles and patches in the model. With this somewhat simple outline of NetLogo we can now move to outlining Steimke’s model.

### 3.2 The Foundational Model

The model that we employ throughout our research is a derivation of the model written by Amy Steimke. We will cover the variables implemented the model and the logic behind the design of the model. To start, let us examine what we wish to characterize as a variable throughout the model.

#### Variables tracked throughout the model

We start our model with the declaration of all variables that will be available on a global scope in the model. For the simple model we only use the calculated average velocity of the river as a global agent, as we need to update both the turtles and the patches based on this information. We also set out to describe the variables that are specific to the patches (the river) and for the turtles (the salmon). The information required for the patches would be any elements that interact with the salmon, so in the baseline model we only note the  $x$  coordinates of our dam obstacles. The salmon are naturally more complex. For each turtle, we have to keep track of the Boolean variable dictating if the agent is moving, the speed of such movement, the baseline probability of dying on any update tick of the model, the Boolean representing if the salmon has approached a dam object, the number of ticks spent at a dam, and the current probability of the salmon bypassing the dam object. This is a rather heady list, so do not get caught up in the particulars. We will outline where each variable becomes important to our model.

#### Initialization

The meat of the model begins when the setup line in our model is called, initializing the model before we run it. Using the patch variables dam1 through dam4 mentioned above, we create a line of four patches that are represented in the model as black lines through the environment. We then move to set the average velocity of the salmon, using an equation calculated by Salinger and Anderson [16]. We start by assigning the flow-rate of the river to flow, and converting the river temperature into degrees Celsius. Then, as Salinger and Anderson outline, we have two separate linear regression lines that differ depending on the temperature of the river. If the temperature is below 16.3 degrees, then the first

equation is used to calculate average velocity. Otherwise, the second equation is used to set the value of the velocity variable. This aspect of the model relies on research done previously, but remains an area of the model that can be refined by further research. We present the equation from Salinger and Anderson below in netLogo code.

```
let flow (FlowRate-m3/s/1000)
  let rivertemp ((RiverTemp-farenheit -32)/1.8)
    ifelse rivertemp <= 16.3
[set velocity ((27.3+(2*rivertemp)-(1.5*flow))/24)]
[set velocity ((100.7-(2.5*rivertemp)-(1.5*flow))/24)]
```

The code above is derived from the following equations, wherein  $v$  is the velocity,  $t$  is temperature of the river, and  $q$  is the river flow rate.

$$v = 27.3 + 2t - 1.5q \text{ for } t \leq 16.3 \quad (1)$$

$$v = 100.7 - 2.5t - 1.5q \text{ for } t > 16.3 \quad (2)$$

Variables such as initial probability of passing dams or dying while moving are set by the user defined input to the model, and both the Booleans at-dam? and start-movement are set to false. Furthermore, from our initial velocity calculations, we distribute the velocity among the salmon with a normal distribution, with the standard deviation again able to be input by the user. We then correct for any negative velocities that may occur by simply considering the absolute value to be the velocity of the agent. This can be seen in the code below.

```
set movement-speed random-normal velocity speed-stdev
  if movement-speed <= 0 [set movement-speed movement-speed * -1]
```

Finally, we include a small section detailing the redrawing of patches if we do not include dams in our model, simply by repainting the patches to the river blue. We will get into how the salmon are affected by these dam objects in our next section.

## Movement

We start our movement section with a list of commands that are to be run, just as we did with setup. For each tick of the model, we will check if we can start migration of the salmon, move any available turtles, check if any die randomly, make sure that there is at least one salmon object that has yet to reach the end of our model, and finally run any dam bypassing behavior. The start-migration variable is set to true if a randomly rolled float is less than a user defined probability, and set to false when the agent reaches the end of our model so that it doesn't exit the model. The variable at-dam? is similarly set, simply by using the primitive NetLogo command in-cone. This command creates a limited degree of vision in front of the agent that can report about the patches in front of it. In this case, we care if any of the patches in front of the salmon are black to represent the dam, in which case we can update our variable to be true.

```
ifelse any? patches in-cone movement-speed 1 with [pcolor=2]
```

We also represent the process of bypassing dams and the learned behavior that can emerge from such a process. Bypassing dams is a fairly simple process in this baseline model. A percentage chance of passing dams is given to each agent, and if they roll a random number below that percentage chance, they are allowed to move forward past the dam. However, after passing the first dam every dam encountered subsequently each agent is asked to recalculate its chance of passing the next dam in the system. This is done using the following equations. The first equation details the requirements for dam learning to occur, namely if the time spent at a dam is less than the current number of dams previously crossed divided by two times the passage rate. This can be seen in the equation below.

$$T_{\text{dam}} \leq \frac{N_{\text{dam}}}{2p_i} \quad (3)$$

Once this condition is met, we update the percentage chance of bypassing the next dam encountered. This is done by multiplying the previous passage rate  $P$  by  $1 + R$  where  $R$  is a random float in the

interval  $(0, 1)$ . In this way, learning is always increased, but by a variable amount.

$$P_x = P(R + 1) \tag{4}$$

In our code, we will split this dam learning into two cases, dependent on which dams the salmon have already crossed.

```
[if at-dam? = true and xcor > dam2 and xcor < dam3 and time-at-dam <= 3
  [set probability-of-passage probability-of-passage*(random-float 1+1)]
if at-dam? = true and xcor > dam3 and xcor < dam4 and time-at-dam <= 5
  [set probability-of-passage probability-of-passage*(random-float 1+1)]
```

Finally, we quickly detail the process of fish dying at each tick. If the agent rolls a random float below their percentage chance to die, then we simply call the primitive “die” to remove the agent from the model. We also account for the fact that salmon are much more likely to die when trying to cross a dam by increasing the mortality rate when in front of a dam object. With all of the methods detailed above, we now have a working baseline model for salmon traversing the dams of the Snake River.

### Interface

One of the features that NetLogo allows for each model is an automatically generated user interface that can easily be set up with a variety of buttons, sliders, and entry fields for user input, as well as graphs and tables for noting the data outputs of the model. In our baseline model we keep track of the mean time spent at a dam, the number of fish who have reached the end of our model, the number of fish that are alive, and graph the number of fish crossing each dam as a component of time. Thus when running our model, salmon agents distributed on the far side of the model advance across the environment, pausing at dams until they can successfully reach the end of the model, or die off. From this final report we can then pull the data collected from our reporters, and compare this to the records that record salmon numbers in an attempt to judge the validity of the model. While Steimke reported in her original work the outputs of her run versus the 2012 data, we not only report her findings but additionally update them with 2016 data before we modify the model to be a more detailed look at the migration process.

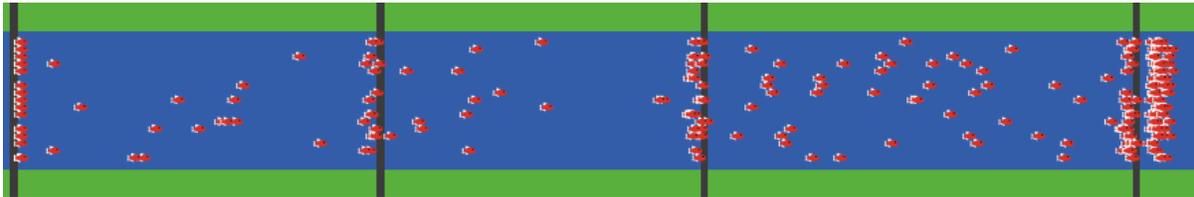


Figure 2: The graphical interface of the baseline model

## 4 Testing the Model

The parameters used throughout the model are presented below, along with references to the source of the data used. For quick reference, the data is summarized in Tables 4 and 5 in Appendix A for the settings used throughout this paper.

**Conditions of the Original Model:** In Steimke’s original model, she ran the system using a Spring 2012 sample of data of Chinook Salmon. Between the dates of May 10<sup>th</sup> to June 10<sup>th</sup>, 65,247 Chinook were recorded crossing the Ice Harbor Dam. A variety of factors were calculated previously as values for the models, with the probability of starting migration set to 1.5 percent, the initial mortality constant set to .04 percent, the probability to make it through a dam at 5 percent, a standard speed deviation to be 0.2, and the dam death multiplier making the mortality rate 3 times higher. Furthermore, data collected by Steimke found the average flow rate to be 3500 cubic meters per second, and the average river temperature to be 54.95 degrees Celsius. Running the model initialized as above, Steimke reported 57,137 passing the fourth dam compared to the reported 57,528, which we confirmed by running the model multiple more times (data of runs can be found in Table 1). While we were never able to

create the same level of accuracy presented in Steimke’s model using the specified parameters in the paper (specifically 63,500 as initial salmon population, and a Dam death multiplier of 5), the above parameters created the accuracy reported by Steimke. As such, we will use them moving forward for other year reports.

**2012 Spring Data:** Using data collected by the Fish Passage Center (FPC) [5], we were able to find the total number of Adult Chinook that cross through the Ice Harbor and Lower Granite Dam. Furthermore, using the definition of a spring run as all Salmon recorded between the dates of the first of May to the first of June, we can query the FPC database for specific counts in the spring run. Unsurprisingly, we find that the Salmon count is very similar to Steimke’s model, with 61,927 passing through Ice Harbor dam and 52,728 reaching Lower Granite Dam. For the spring of 2012, FPC reports an average temperature of 53.3 degrees Fahrenheit. Discharge data for the 2012 year was downloaded from the USGS site 13334300 recording station. Since the USGS reports flow in cubic feet, a simple calculation was used to turn the discharge rate to cubic meters, resulting in a discharge rate of 1,924.5 cubic meters per second. Running this data yields an interesting trend of overestimating the expected return by about three percent. This mirrors the error of evaluating the average throughout the year, which is an unexpected result. Additionally the difference between Steimke’s run conditions and the conditions found above point to a suggestion that the model presented is tuned to the 2012 specific data, with around a three percent error for other systems. Interestingly, this error always occurs in the form of an underestimation of the population due to the weaker flow rate decreasing the average velocity of the salmon agents.

**2016 Spring Data:** Again using the definition of spring as all Salmon recorded between May 1st and June 1st, we find that the FPC database records a rather low count of 57,672 salmon passing the Ice Harbor Dam, and 50,760 reaching the Lower Granite Dam. FPC further reports that the average temperature in this time period is 55.79, with the GIS recording a flow rate of 1,688.7 cubic meters per second of water discharge. Interestingly the spring 2016 data is without the error introduced in the previous terms.

**2012 Average Data:** Taking a full year as our frame allows one to include the three runs that occur each year, and look instead at differences between years. A further level of specificity could be brought to the model by directly comparing the runs of similar times of years. With this in mind, we seed the model with 124,685 Chinook that was reported by the FPC at the Ice Harbor Dam, and 114,217 Chinook recorded at the Lower Granite Dam. Additionally, we can calculate the average temperature of the 2012 year for the Snake river to be 59.27 degrees Fahrenheit. The average flow throughout the year for the 2012 river was 975.6 meters cubed per second. Note that is much lower than the data used in Steimke’s original model as river flow tends to be much higher during the time of salmon migration. This is again a simplification introduced into the model to make yearly comparisons easier, but will be done with a sacrifice in accuracy. We again report the data collected in Table 1. Note that in averaging the year, we introduce around 3 percent error into the system, due to the issues of arrival dates.

**2016 Average Data:** Similar to the 2012 year, we shift from a run specific view to a full year summary view. FPC reported that in 2016 a total of 118,177 adult Chinook salmon passed the Ice Harbor Dam, and 108,839 reached the Lower Granite Dam. Water temperature was recorded to be 61.33 degrees Fahrenheit, with a flow rate calculated to be an average of 800.02 cubic meters per second. Again we introduce error by looking over the course of a year, instead of during peak times of migration, as to avoid having to define what periods of time are considered "run times". Again we see the same approximate error about 3.5 percent difference between expected and recorded numbers. This suggests that as we move forward from the baseline model we can account for this term in adjusting some of our fixed parameters such as Dam death multiplier and likely account for nearly all the error being observed in the models beyond the original.

Table 1: Results from Model runs

Year	Seed	Recorded	Model	Percent Error
Original 2012	65247	57528	57518	0.0
			57574	0.0
Spring 2012	61927	52728	54565	-3.4
			54668	-3.6
Spring 2016	57672	50760	51127	-0.7
			50802	0.0
Full Year 2012	124685	114217	110421	3.3
			110684	3.1
Full Year 2016	118177	108839	104802	3.7
			105085	3.4

The data presented in Table 1 is a brief and informal display of some of the data collected from our model. Table 6 in Appendix B has average results from 11 test runs per condition. We now turn to statistical analysis of our results to see if any of the results can be confidently labelled as significantly different.

## 5 Statistical Analysis of Model

To verify the degree of accuracy that we can ascribe to our model, we turn to the statistical analysis tool of two sided  $t$ -test. The goal of such statistical analysis is to determine if we can confidently say that our findings are significantly close enough to the recorded data as to be accurate. Using  $t$ -test techniques, we are able to answer the question of significance not only for the model’s accuracy in one year, but also in the average of the last 4 years. We will first introduce the  $t$ -test, and then work through our calculations and results for the above model.

### 5.1 Introduction to $t$ -tests

The  $t$ -tests are a statistical technique to determine if two numbers are statistically significant from each other. This is done by comparing the difference in means while accounting for the variability in the data. The equation for calculating a  $t$  statistic is shown in equation (5). This is the two sided  $t$ -test as we are testing that the model mean could be greater, or lesser, than the recorded mean. Other forms of the  $t$ -test exist, one of which we employ to compare years. In this case, we have two groups, the first which we denote with the subscript  $X_T$  and the second we denote with  $X_C$ . The  $\bar{X}$  term refers to the arithmetic mean of each group which is calculated by summing the values of the group and dividing them by the total number of elements summed. Similarly, the  $n$  refers to the total number of members of that group, while the var is a variable for the variance of the group. Variance is calculated by summing the squares of the difference between each member of the group and the mean, as seen in equation (6). Note that  $m$  is the number of elements in the group in question, and  $\bar{X}$  is the mean of said group. Together, the denominator is commonly referred to as the “standard error of the difference”. To interpret the  $t$  statistic, we need to determine the numbers of variables being accounted for; represented by degrees of freedom (in our case two less than the sum of the size of the two groups), and a preset  $\alpha$  level. The  $\alpha$  level is the degree of accuracy we want to assign to our model as it is the percentage chance that our model’s results occurred by chance.

$$t = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\text{var}_T / n_T + \text{var}_C / n_C}} \quad (5)$$

$$\text{var} = \sum_{j=1}^m X_j - \bar{X} \quad (6)$$

#### Paired $t$ -test

We will also use a paired  $t$ -test to compare the multiple years of our model together. A paired  $t$ -test is calculated very similarly to the standard  $t$ -test introduced above, but allows for data which is related as pairs to be entered. This is useful when we compare results that are not all from the same conditions,

such as the conditions that vary between years. A paired  $t$ -test is calculated with equation (7). Because we are dealing with data in pairs, we consider the difference of each pair, and then find the mean of these differences. This value is then entered as  $\bar{d}$  in the equation, and unlike the two tailed  $t$ -test, the mean is calculated after the pairs are evaluated. Then, all that is left is to calculate the standard deviation of the differences ( $s$ ) and use the same interpretation of  $n$  as in the previous  $t$ -test,

$$t = \frac{\bar{d}}{\sqrt{s^2/n}}. \quad (7)$$

We will use the  $t$  statistic in a similar way to the one generated with the two sided  $t$  test as a measure of significance by calculating a corresponding  $p$  value.

### Application of the $t$ -test

In evaluating our model, we calculate  $t$ -tests for two different categories of results. We first examine the seasonal specific accuracy, comparing the recorded value of fish passing the Lower Granite Dam and the group of collected outputs of our model. We note that in this calculation we are assuming that the group of the recorded values is composed of two entries that are equal, and thus a variance of 0. The second category looks to verify the model between years, comparing the group of recorded yearly counts to the group of calculated yearly counts by way of a paired  $t$ -test. In both cases we use the table found in “*Elementary Statistics*” [3] to determine if our model numbers are significantly different from the actual recorded number of salmon. By using the  $p$ -value of 0.05 as the cutoff for significance. If the value is less, we note that there is less than a five percent chance that the difference is due to chance, and if it is greater, then the difference could be from chance. Ideally, to verify the model we hope to find that the calculated are not statistically different from the recorded, but we will use the  $t$  statistics corresponding  $p$  value to determine what we should not rely on as an accurate output from the model without making allowances for the result.

## 5.2 Results from our Analysis

Table 2: Results from Model runs

Year	Recorded	Modeled	$t$ statistic	$p$ value	standard error	significant
Original 2012	57528	57508	0.385	0.7078	21.19	No
Spring 2012	52728	54636	47.8313	0.0001	24.87	Yes
Spring 2016	50760	50926	2.2541	0.0456	30.31	Yes
Full Year 2012	114214	110431	63.0103	0.0001	24.69	Yes
Full Year 2016	108839	104794	46.5	0.0001	35.79	Yes
Comparing Years	N/A	N/A	0.9817	0.3818	N/A	No

### Comparing between model years

Finally, we can calculate a paired  $t$ -test value score to examine the overall accuracy of the model. A paired  $t$ -test accounts for each value of group one to be related to a value in group two, creating an entry pair. Using the average value for each of the categories previously presented, we calculate a  $t$ -value of 0.9817, which gives us a  $p$ -value of 0.3818. This calculation was done with 10 degrees of freedom, just as the two tailed  $t$ -statistic was. Overall, the model while not predicting an individual year with the accuracy we would wish, is generating values that are not significantly different from the recorded values when observed outside of a specific year dependency. This non-significant difference between years is what allows us to move forward and work with this data to make assumptions about the predictive value for trends which is used for further research.

## 6 Expanding the model

Since we have examined and established the validity of the model in reproducing the results found in the Snake River, we now explore the effects of obstructions such as dams on the reproductive success of the salmon, instead of just the survival success. Studies done at spawning grounds for salmon in Canada [6] found that sixty eight percent of female salmon create redds (areas of eggs) in two days of

reaching the spawning grounds. Outside this window the chances of the female being available or alive for mating purposes sharply decreases, effectively removing the female from any area of study interest. We examine if obstructions artificially spread the population so that this reproductive success is on average decreased, suggesting a secondary effect of the dams that is not accounted for when only examining the passage rates through the dams. This suggests the creation of a submodel, which accounts for a normally distributed reproductive chance with standard deviation of 1 day to find a mate. If we consider this a search period, we can allow agents in the model to have a search “tolerance” that is normally distributed for the time in which they will seek out a mate. Outside of that period they will not be allowed to mate, and be considered dead in the model. We examine if this is not an issue without dam obstructions due the fact that runs arrive continuously throughout the run time, but as our model introduces a striation in the population we find a decrease in reproductive success. With this goal in mind we turn to justifying and writing the expansion of our model.

## 6.1 Implications of the Distribution

An initial indication that there is a significant spreading effect in the salmon population are the distributions generated by the netlogo model, as seen in Figure 3. The distribution is best categorized as a Poisson distribution, due to the right skew and decrease in height. However, at each consecutive dam crossing, we are introducing a dependence on the previous time crossing and thus are no longer quite Poisson distributions. Still, the Poisson estimation of our distribution will allow for us to calculate an estimation of the variance  $\lambda$ , which will be inversely related to the spread of the population at each dam. If this  $\lambda$  value decreases at each dam, we have shown that the population is spreading out over time.

### The Poisson Distribution

The Poisson distributions represent the probability of events occurring in a fixed time interval that are independent of previous values. This can easily be translated to the model system, as we are dealing with independent occurrences over a time spread (in this case the time interval of the model running). Poisson distributions must also meet a series of criteria, and we will outline those which are relevant to our model.

1. The Poisson is a discrete distribution, so the data must also be discrete. This is something that is innate to our model as it is based around independent time ticks.
2. Events occur independently. As we mentioned above, this is the main reason why the distributions at each dam are not Poisson, but the distribution of start times are.
3. There is a constant rate at which each event occurs. Again, with the model of dam learning we are creating an inconsistent rate at each stage, and thus only the initial start time has Poisson distributions.

It follows that the distributions in Figure 3 are not in fact Poisson distributions, but closely related. Unfortunately there is no distribution that is more closely related than the Poisson, so we move forward using the Poisson to calculate our parameters, but acknowledging that we have introduced error in using the Poisson as an estimate.

### Parameters of the Poisson

The Poisson distribution has a mean and median parameter, but is largely characterized by the average number of events per time interval which is commonly referred to as  $\lambda$ . The median ( $v$ ) is calculated as

$$\lambda - \ln 2 \leq v < \lambda + \frac{1}{2} \quad (8)$$

and the mean is equivalent to the parameter  $\lambda$ . The variance of the distribution is also equivalent to the mean for all Poisson distributions, so we will refer to both as simply  $\lambda$ . Variance describes the differences between the height (or y values) throughout the distribution. The larger the variance of a Poisson distribution, the more compact the data is. As the data spreads out, the difference between the data in a set neighborhood decreases. Thus the variance  $\lambda$  is inversely related to how spread out our data is in the distribution.

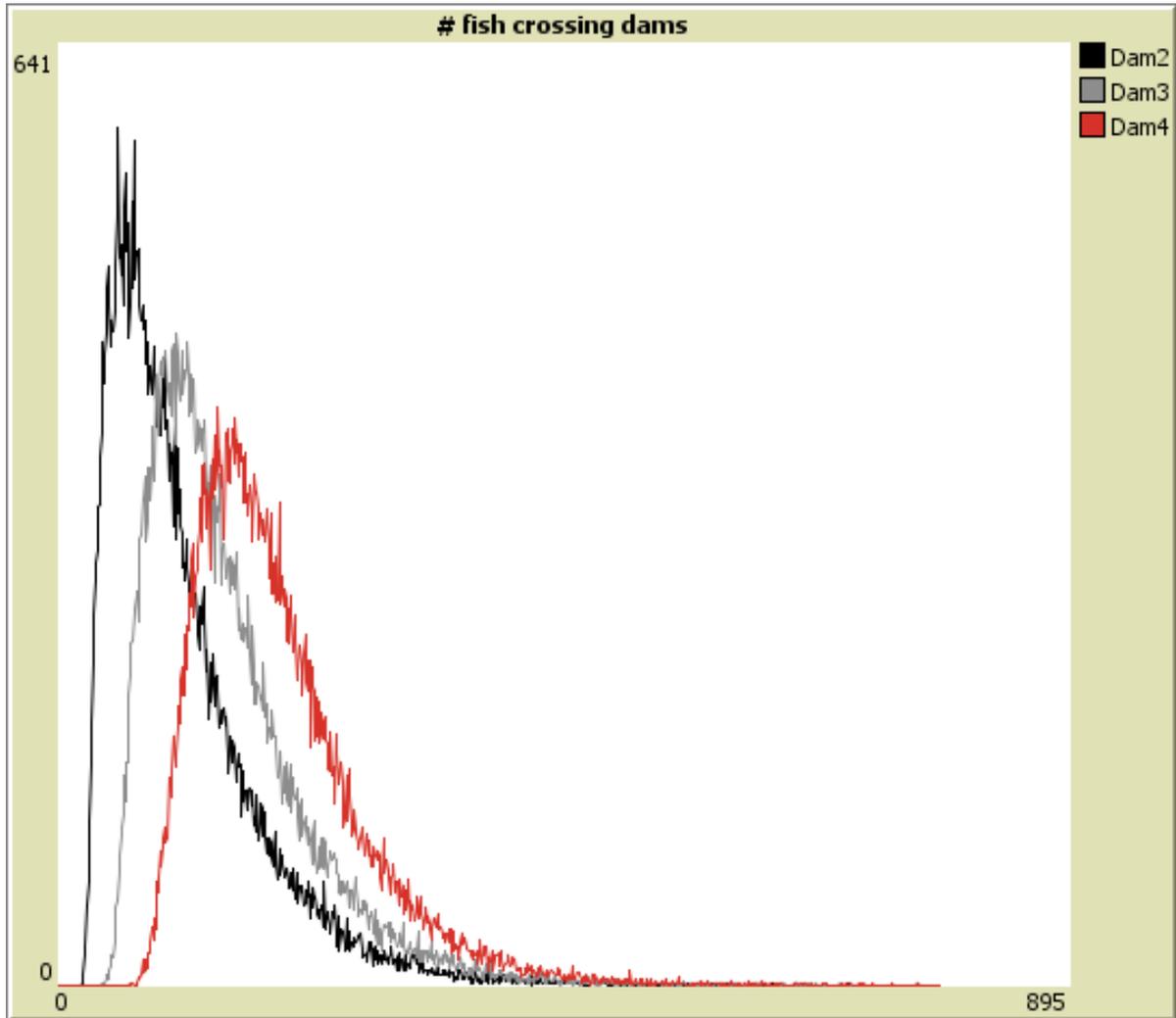


Figure 3: The number of fish crossing each dam at each time tick

### Compound Poisson Distributions

If you have multiple Poisson distributions that you are incorporating at the same time, then a Compound Poisson models the outcome of such an event [18]. The Compound Poisson requires all elements that are being previously accounted for to have Poisson distributions, and then considers the summation of each of these distributions in order to find the final state of the distribution. Additionally, the number of terms being calculated for at a moment is also considered to have a Poisson distribution, as the inputs of the previous models are added as they are calculated. Compound Poisson distributions have a similar mean, variance, and median that is calculated from the previous Poisson distributions, and capture the full system. So while we look at each distribution individually in our calculations, one could instead sum together the three distributions to compare the parameters between conditions instead of within a condition.

## 6.2 Quantifying the Spread of the Model

To quantify the effects of successive dams in our model, we examine the trends in the average variance of each condition at successive dams. To calculate the average variance, each condition was run ten times with variance being calculated for that run, and then the variances at each dam was summed to calculate the mean variance at each dam in the condition. Those results are recorded below in Table 3. Note that for each condition, we see the variance decreasing at successive dams which informs us that the spread is increasing at each successive dam. This pattern of increasing spread at each dam suggests

supports the claims made above numerically.

Table 3: Examining the average variance

Year	$\bar{\lambda}$ at Dam 2	$\bar{\lambda}$ at Dam 3	$\bar{\lambda}$ at Dam 4	Trend at successive dams?
Original 2012	71.28539	69.10276	66.82547	Decrease in variance
Spring 2012	66.17703	64.12771	61.97376	Decrease in variance
Spring 2016	63.73179	61.78017	59.76291	Decrease in variance
Full Year 2012	128.60466	124.81153	120.80781	Decrease in variance
Full Year 2016	131.90042	128.05352	123.99245	Decrease in variance

Paired  $t$ -tests were run on each year, comparing the variances of dam 2 to dam 3, dam 2 to dam 4, and dam 2 to dam 4. In ALL cases, in every year, the  $p$ -values were found to be significantly different ( $p < .001$ ) from each other.

### Calculating Variance

To calculate the variance of the three distributions generated in a single run of the model, we turned to MATLAB. First, after each model run, the distribution was exported as a .csv file and trimmed to only contain the relevant data. The data was ordered Dam2-Dam4 column wise each in export to keep uniformity. This data file was then loaded into MATLAB, and the Poisson parameter estimator function was called on the data, outputting an estimate for the  $\lambda$  parameter. This data was then saved into a master document, to be used to calculate the means for each  $\lambda$  value in each condition. The MATLAB code used is presented below, with # referring to the file name in question.

```
load #.csv
X = X#;
m=poissfit(X(:,1))
n=poissfit(X(:,2))
o=poissfit(X(:,3))
```

We find is the average variance is decreasing at successive dams, and thus the spread is increasing as the fish encounter more dams. In fact, quantifying the correlation between dams and variance by use of the data provided in Table 3, we find a correlation coefficient of  $-.07$  for all conditions,  $-0.4909$  for just seasonal runs, and  $-0.8924$  for only yearly runs (Table 8). As variance is inversely related to spread, we note that there then exists a positive correlation of varied strength between adding a successive dam and increasing the spread of the population. This supports that there is a spreading effects of the dams and allows us to transition to the reproductive submodel that is described below.

## 7 A reproductive opportunity model

To move forward in creating a separate submodel, the first new element we must introduce is a difference in sex of the salmon. We allow the population to be half male and half female as there does not appear to be a significant difference in the population travelling upstream, denoting this by the colors red and purple respectively. Once each salmon has passed the fourth dam, we then invoke our submodel. To start, each agent is moved from their position past the fourth dam into a single patch (point in the environment) to be considered the spawning ground. This single spawning ground can easily be redefined to be multiple spawning grounds as desired, however we stick to only one to present the simplest form of the model. Each agent also has a decaying “mating energy” that is a value normally distributed throughout the population, with a mean of 24 (representing 24 hours) and a standard deviation of 24 (representing the plus or minus one day). At each call of the model, while the agent is in the mating ground, this value will be decreased until they run out of the “mating energy” and subsequently die. During the time frame for which the agent has a positive value, it pairs up with any nearby agents that are of the opposite color. If it finds any, the agent sets its reference variable of partner to the reference number of the agent it has found. Similarly, the partner number of said agent is paired with the agent in question. The male agents are then removed from the model, and the female agents are moved past the mating zone to represent successful mating offspring. The submodel also changes the stop consideration of the model, running until all agents have died in the mating zone. The output of this model then

represents all successful matings, which can be used to calculate the total offspring that will grow into adulthood. A snippet of the code written for the submodel is presented in appendix C.

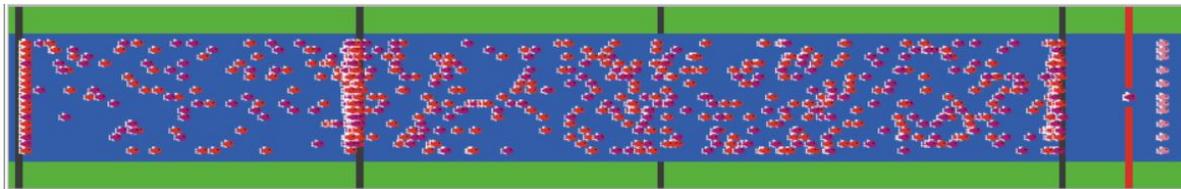


Figure 4: The model with the submodel enabled

## 7.1 Data from the submodel

We collected data from the submodel for each of the year conditions we have been using above, with the addition of comparing data with a variance of dams destroyed. Data with no dams destroyed, one dam destroyed, two dams destroyed, and all dams destroyed was collected. This data can be seen in Table 9 in Appendix B. Note that this data was collected with dams destroyed was done from the lowest dam up to the highest dam, although no significant difference was observed when dams were destroyed in reverse order. The first trend that emerges is the large number of successful matings compared to the number of salmon we expect to reach the fourth dam. While generally 110,000 salmon reach the fourth dam, only around 50,000 successfully mate. This effect is not as notable in the yearly runs, suggesting again that the weakness of yearly runs can be found in an increased spread of the population. Thus, the yearly runs by themselves are not the best basis to note trends, but with the support of the seasonal runs to support them we can still draw conclusions. The main trend observed between conditions appears in each model year. While it is more pronounced in the “year” settings due to the number of agents being run through the model, both conditions show that there is a dramatic decrease between all dams enabled and one dam destroyed. After the first dam is destroyed, the effects of additional dams decreases through the rest of the conditions. However, despite this decrease in the slope between conditions, the overall trend is clear. Running  $t$ -tests comparing the data between conditions for each year finds that for each dam destruction state the results are significantly different. As we destroy each dam we see an increase in the number of successful matings that are too large to account for simply by chance.

## 7.2 Implications for Conservation Decisions

While it is always hard to make absolute statements, it becomes clear that dams are having a dramatic impact on the reproductive success of salmon which is not captured simply by the pass rates for dams. Removing the dams, while not capturing the full effect on the ecosystem, increases the successful number of matings in the salmon population. Often, arguments presented by those supporting the role of dams focus on the low mortality of salmon population at dams, but this does not capture the full picture. Instead, when considering the advantages and disadvantages of dams in major salmon traversed waterways like the Snake River, the role of such a population spread should not be overlooked. With further observation in the field to support this research, the new element in considering the impact of dams could see a shift in the current conservation efforts for rivers with large numbers of salmon travelling to their spawning grounds. Note, that since we simplified to one mating ground, rather than the hundreds that are represented throughout tributaries of the Snake river, we are dealing with an improbable “best case” scenario. In fact, it is almost certain that the number of successful matings is much lower than what we represent. So while our data presents a case that provokes a need for change in our ecosystems, this data may be an understatement of the true effect that dams can have on the population. We now extend our model to be able to predict future years, which can demonstrate the fact that this decrease in population (due to less successful matings) is amplified as each year has population continue to be decreased. Note that the number of returning salmon we calculate is much lower than our seed values due to the fact that recovery efforts have focused on fisheries artificially returning the salmon populations to dam-less populations so that the impact of the dams is decreased.

## 8 Seeding future years

To calculate future seed values (the number of salmon entering the model), we take the number of fish that have successfully reached the end of our model (separate recorded numbers for time of year, as well as what submodels were used) and use them as a basis for the number of successful pairings. In the cases where the submodel for opportunity based reproduction is not called, the number of successful pairings is assumed to be half of the number of salmon that finish the model. This number is then multiplied by the number of offspring each reproduction will lead to entering the next year’s model, on average. This calculation is done by the equation below:

$$A = M \times 5000 \times 0.20 \times .015 \times 0.95 \times 0.40 \times 0.1 \quad (9)$$

Where  $A$  is the number of adult salmon who travel back upstream and  $M$  is number of successful matings. For each successful matings, the female lays on average 5000 eggs. Of those eggs, only 20 percent of the eggs develop into hatchlings [17]. Of the number of successful alevins (hatched eggs), only one to two percent of them will survive to be adults [11]. Of the juveniles, 95 percent of them successfully travel downstream to the ocean to become full adults citeNOAA. Finally around 40 percent are able to return as adults to return to their mating grounds and only 10 percent re-enter the Snake River [19], allowing us to generate the equation above to create the seeding equation for future years.

We can run the seeding calculation for the each year, as shown in Table 10. Taking the 2012 Spring data and running it through our seeding equation, we get 14,614 returning with all dams on, 15,394 with one dam off, 15,582 with two dams off, and 15,652 with all dams off. This is compared to the value calculated by FPC in 2013 between May 10<sup>th</sup> and June 10<sup>th</sup> of 20,001 salmon. Note that the closest estimation is provided by all dams destroyed, but this is because we do not account for the salmon introduced by fisheries, and that these numbers are significantly less than our seed year, as in 2013 the run was not between the above dates.

Extending this beyond what we present here, this calculation can be done with dams on and off, dam learning on and off, and opportunity based reproduction on and off. We leave those not presented here as future work. This number can be done on previous years, and compared to the recorded number of actual seeding fish, or done predicatively in future years to show the amplified effect of the decreased reproductive successes.

## 9 Summaries and Conclusions

Agent Based Modeling can have multiple uses in exploring ecological systems. As we stated in this paper, the model provided a novel approach to a previously documented question of the affect of dams on how salmon travel upstream. Working with Steimke’s base model, we applied the model to recent year salmon counts and differentiating between seasonal runs and full year abstractions of total salmon. Applying statistical tools to the model, we found  $t$ -test results for each individual year indicated a middling degree of inaccuracy in the model. This was amplified in the concept of yearly data due to the generalizations assumed when accounting for periods of seasonal runs, and times where nearly no salmon traverse the Snake River. However, despite these findings when comparing the accuracy of the model between years, we found that each condition was statically different from each other. This lead us to the conclusions that while using the model to predict individual yearly data requires further tuning to match the recorded numbers, the model could be used by altering a few of the assumptions without loss of the patterns described above. This is due to the fact that the output of the model was consistent in the error generated, and thus the error is simply a product of the assumptions we made when defining the variables (temperature, velocity, passage chance, etc.) without impacting the trends that lead to the conclusions.

In this process of analyzing the model, we highlighted a distribution that appeared at first to be Poisson, but under further investigation differed slightly from the standard distribution. With an approximation of the distribution, we were able to quantify the spread and variance of the salmon as they traversed the dams of the model, and noted a strong pattern of the spread of the population increasing at each consecutive dam. Research on the salmon life cycle revealed studies on the availability of female salmon once they have reached their spawning grounds suggested that the number of salmon completing the model was not representative of the number of salmon that were available to reproduce. A submodel was created to capture this interaction, assuming only one mating zone rather than the complicated

system of tributaries that act as the spawning grounds of the salmon. This submodel was utilized with our model conditions, as well as with varied number of dams still intact along the river. Ultimately, the data collected via this submodel pointed us to the final conclusion of this paper.

The model indicated that dams are acting to decrease the number of successful offspring that can grow to be adults, and this factor is not captured in a traditional view of dams as simply obstacles. By varying the dam construction along the Snake, we found that the removal of the first dam (in our case the Lower Granite Dam) had the most dramatic effect, while successive dams had a diminishing benefit. This trend, we suggest is due to the compound effect of dams on the spread. Finishing off, we present a simple equation for estimating the number of adult salmon that will swim up the Snake to return to their mating grounds so that future year data can be estimated, and the effect of dams increasing spread can be seen to be amplified through generations. With additional work on tuning the model, expanding on the fact that the first (sequentially) dam removed has a more dramatic effect, and supporting field work to account for the spread, the concepts presented within the paper could act to inform future conservation efforts and add a new element to the calculation of the impacts of dams on salmon populations.

## Appendix A: Model Settings

### Parameters of the Model

Table 4: Constants of the model

Migration	Starting Probability 0.015	Speed Standard Deviation 0.20
Dam Traversal	Dam Death Multiplier 3.00	Passing Dam Probability 0.05
Reproduction	Successful Reproduction Probability 1.00	Mortality Constant 0.0004

Table 5: Variables for each year

Year	Seed	Temperature	Flow Rates
Original 2012	65247	54.95	3500
Spring 2012	61927	53.30	1924.5
Spring 2016	57672	55.79	1688.7
Full Year 2012	124685	59.27	975.6
Full Year 2016	118177	61.33	800.02

## Appendix B: Data Tables

### Model Data

Table 6: Average run results

Year	Seed	Recorded	Average Model Output
Original 2012	65247	57528	57508
Spring 2012	61927	52728	54636
Spring 2016	57672	50760	50926
Full Year 2012	124685	114217	110431
Full Year 2016	118177	108839	104794

## Variance Data

Table 7: Average Variance

Year	$\bar{\lambda}$ at Dam 2	$\bar{\lambda}$ at Dam 3	$\bar{\lambda}$ at Dam 4
Original 2012	71.2854	69.1028	66.8255
Spring 2012	66.1770	64.1277	61.9738
Spring 2016	63.7318	61.7802	59.7629
Full Year 2012	128.6047	124.8115	120.8078
Full Year 2016	131.9004	128.0535	123.9925

## Correlating Variance

Table 8: Correlation Coefficient of Variance

Correlation Coefficient	All Conditions	Only Spring Runs	Only Yearly Runs
R	-.07	-0.4909	-0.8924.

## Mating Model Data

Table 9: Average Number of Successful Matings

Year	All Dams Enabled	One Dam Disabled	Two Dams Disabled	All Dams Disabled
Original 2012	26989	28463	28707	28881
Spring 2012	25639	27007	27336	27460
Spring 2016	23981	25258	25488	25643
Full Year 2012	47493	54724	55176	55593
Full Year 2016	49314	51976	52437	52721

## Future Seeding

Table 10: Future Seeding with Mating Model Data

Year	All Dams Enabled	One Dam Disabled	Two Dams Disabled	All Dams Disabled
Original 2012	15384	16224	16363	16462
Spring 2012	14614	15394	15582	15652
Spring 2016	13669	14397	14528	14617
Full Year 2012	27071	31192	31450	31688
Full Year 2016	28109	29626	29889	30051

## Appendix C: Code Snippets

The full code for the model can be found [HERE](#).

### Mating Submodel

```
1 to mate
2 if Opportunity-Mating?
3 [ask turtles
4   [if (xcor >= dam4 and xcor < dam4 + 9)
5     [setxy matezone 0]
6     if xcor = matezone
7       [set matelife matelife - 1
8         if partner = 0
9           [set partner one-of other turtles with
10            [partner = 0 and color != [color] of myself and xcor = matezone]
11              ifelse partner = nobody [set partner 0] [ask partner
12                [ set partner myself ] ]
13            ]
14            if matelife != nobody
15              [if matelife <= 0
16                [die]
17              ]
18            if partner != nobody
19              [if partner != 0
20                [if color = magenta
21                  [setxy matezone + 5 8 - random(17)
22                    set color pink]
23                ]
24              ]
25            ]
26          ]
27        ]
28 end
```

## References

- [1] Akhbari, M., & Grigg, N. S. (2013). A Framework for an Agent-Based Model to Manage Water Resources Conflicts. *Water Resources Management*, 27(11), 4039–4052. doi:10.1007/s11269-013-0394-0
- [2] Albrecht, J. (2007). *Key concepts and techniques in GIS*. Los Angeles i.e. Thousand Oaks, CA: SAGE Publications.
- [3] Bluman, A. G. (2012). *Elementary statistics: a step by step approach*. New York: McGraw-Hill Science/Engineering/Math.
- [4] Campbell, H. E., Kim, Y., & Eckerd, A. (2015). *Rethinking environmental justice in sustainable cities: insights from agent-based modeling*. New York, NY: Routledge.
- [5] Form and query by Lynnae Sutton. *Adult Daily Dam Counts - Data Query*. Adult Daily Dam Counts - Data Query. [http://www.fpc.org/web/apps/adultsalmon/Q\\_adultcounts\\_dataquery.php](http://www.fpc.org/web/apps/adultsalmon/Q_adultcounts_dataquery.php). Accessed 1 March 2017
- [6] Grant, J., Hruska, K. A., Hinch, S. G., Patterson, D. A., & Healey, M. C. (2011). Egg retention in relation to arrival timing and reproductive longevity in female sockeye salmon (*Oncorhynchus nerka*). *Canadian Journal of Fisheries and Aquatic Sciences*, 68(2), 250–259. doi:10.1139/f10-153

- [7] Grimm, V., & Railsback, S. F. (2005). *Individual-based Modeling and Ecology*. Princeton: Princeton University Press.
- [8] Harrison, J. Fish passage at dams. *Fish passage at dams*. North west council. <https://www.nwcouncil.org/history/FishPassage>. Accessed 21 February 2017
- [9] Humphreys, P. (2004). *Extending ourselves: computational science, empiricism, and scientific method*. New York: Oxford University Press.
- [10] Laver, M., & Sergenti, E. (2012). *Party competition: an agent-based model*. Princeton: Princeton University Press.
- [11] Many Factors Impact Salmon. *Clean Hydro*. <https://nwriverpartners.org/factors-that-impact-salmon>. Accessed 25 April 2017
- [12] NetLogo User Manual (version 5.3.1). *NetLogo User Manual (version 5.3.1)*. <https://ccl.northwestern.edu/netlogo/faq.html>. Accessed 21 February 2017
- [13] N. O. A. A. (2002). *Anadromous fish agreements and habitat conservation plans for the Wells, Rocky Reach, and Rock Island hydroelectric projects: final environmental impact statement*. Portland, OR: U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.
- [14] The River. *Snake River Waterkeeper*. <http://www.snakeriverwaterkeeper.org/the-river/>. Accessed 21 February 2017
- [15] Phang, S. C., Stillman, R. A., Cucherousset, J., Britton, J. R., Roberts, D., Beaumont, W. R. C., & Gozlan, R. E. (2016). FishMORPH - An agent-based model to predict salmonid growth and distribution responses under natural and low flows. *Scientific Reports*, 6, 29414. doi:10.1038/srep29414
- [16] Salinger, D. H., & Anderson, J. J. (2006). Effects of Water Temperature and Flow on Adult Salmon Migration Swim Speed and Delay. *Transactions of the American Fisheries Society*, 135(1), 188–199. doi:10.1577/t04-181.1
- [17] Salmon Life Cycle. *Salmon Life Cycle*. <http://www.5counties.org/salmoncycle.htm>. Accessed 31 January 2017
- [18] Siegrist, K. Compound Poisson Processes. *Compound Poisson Processes*. Department of Mathematics, University of Utah. <http://www.math.uah.edu/stat/poisson/Compound.html>. Accessed 11 April 2017
- [19] Sockeye Salmon Frequently Asked Questions. *National Parks Service*. U.S. Department of the Interior. <https://www.nps.gov/katm/learn/photosmultimedia/sockeye-salmon-frequently-asked-questions.htm9>. Accessed 25 April 2017
- [20] Steimke, A. Salmon migration in the Snake River. *Salmon migration in the Snake River, by Amy Steimke (model ID 4220)* NetLogo Modeling Commons. [http://modelingcommons.org/browse/one\\_model/4220model\\_tabs\\_browse\\_info](http://modelingcommons.org/browse/one_model/4220model_tabs_browse_info). Accessed 9 May 2017
- [21] Trochim, W. M. K. (2006, October 20). *The T-Test*. The T-Test. <https://www.socialresearchmethods.net/kb/stat.t.php>. Accessed 8 March 2017
- [22] United States. *National Park Service*. The Salmon Life Cycle. National Parks Service. U.S. Department of the Interior. [https://www.nps.gov/olym/learn/nature/the-salmon-life-cycle.htm#CP\\_JUMP\\_224851](https://www.nps.gov/olym/learn/nature/the-salmon-life-cycle.htm#CP_JUMP_224851). Accessed 31 January 2017