Computing the Importance of Physical and Hydraulic Variables to Fish Habitat Selection Through Multiple Linear Regression

Izzy Brown
August 24, 2014
EcoInformatics Summer Institute Engineered Log Jam Study

Abstract:

The purpose of this study is to quantify the importance of turbulent kinetic energy, strain, water velocity, fish depth, and fish distance to wood to the decision making process of overwintering coho salmon during habitat selection. Although many studies have evaluated this habitat selection of salmon, few have included hydraulic variables in the equation, while recent research indicates that these parameters could affect salmon behavior. More so, greater understanding of the influence of these parameters can aid in future river and habitat restoration efficiency. Successful coho salmon habitat restoration projects can increase the abundance, biomass, and biodiversity of such salmon. This study was conducted using high resolution (0.1 m) mapping of the flow field through an acoustic Doppler velocimeter array and underwater videogrammetry of fish locations at the Oregon Hatchery Research Center in Alsea, Oregon. Such a study is of even greater importance because studies have shown that overwinter survival success is one of the greatest indicators of annual coho salmon populations [1]. Although inconsistencies were discovered within the multiple linear regression model used, such inconsistencies highlight fish length dependent habitat selections as well as coho salmon affinities for low velocity levels, low distances to wood, deeper depths, and low turbulent kinetic energy values.

Introduction:

Ecologists and hydrologists have been interested in effective riverine and fish habitat restoration methods for many years. As human population and development has increased, the effect on river systems and habitat stability within such systems has multiplied for the sake of economic and recreational human gain. In addition to increased channelisation, wood removal from rivers has been shown to constrain salmonid populations and biomass. [2] As riverine restoration studies evolve, much focus has been given to habitat preferences of fish in relation to wood debris. Many studies have been conducted in order to determine the effect of depth and water velocity on fish habitat selection, in assumption that these conditions explain fish preference for wood debris. However, some have noted fish remaining beneath cover in shallower and faster conditions than those shown to be preferable to fish.[3] Although turbulence has been shown to be ecologically meaningful and abundant in and near covered areas, it also

---

[1] Howard, C. (2006), Mill Creek Fisheries Monitoring Program: Ten Year Report. Final Report to Save-the-Redwoods League. *Stillwater Science*s, *Save-the-Redwoods League*. Received at http://www.savetheredwoods.org/media/pdf_howard.pdf.

[2] Hafs, A.W.(2014).Quantifying the role of woody debris in providing bioenergetically favorable habitat for juvenile salmon. *Ecolofy Modelling, 285*, 30-38. doi: 10.1016/j.ecolmodel.2014.04.015

[3] Ayllón, D., Almodóvar, A., Nicola, G. G. and Elvira, B. (2009), Interactive effects of cover and hydraulics on brown trout habitat selection patterns. River Res. Applic., 25: 1051–1065. doi: 10.1002/rra.1215

3

has been excluded from many related studies[4]. Thus, the purpose of this study is to quantify the importance of physical and hydraulic variables to fish habitat selection through the means of multiple linear regression.The physical and hydraulic variables in question are turbulent kinetic energy (TKE), strain, water velocity, fish depth, and fish distance to wood. TKE is defined as the change in velocity over time, while strain is defined as the change in velocity over space. Though we know TKE and strain are inevitably correlated, as they are both dependent on velocity, they are included in this model as we hypothesize temporal and spatial velocity fluctuations are also important to habitat selection.

Other studies defined habitats according to what was available in a given river system, such as by substrate or vegetations types[5]. However as our study site was self designed, for the purpose of controlled physical and hydraulic variation, as well as relatability to other study sites, such definitions were trivial to our goal. Instead habitats within this study were defined by regions of given spatial and hydraulic attribute ranges, as done in common Instream Flow Incremental Methodology (IFIM)[6]. Such a methodology allows for more generalized interpretations surrounding questions of salmon juveniles and habitat selection, while veering away from study site specific conclusions. However while other studies using IFIM habitat definitions have taken temporally dependent variables into account by computing and comparing fish habitat preference between different time series[7], the data at hand does not have enough fish observations per time series to compute such a comparison with significant results. We therefore move forward with a IFIM habitat definition structure, while defining all available locations as those observed within a fish trajectory. Such a definition allows future studies to quantify and compare, for example, the importance of distance to fish and TKE values throughout a given fish's trajectory. We thus quantify the the amount a given fish prefers a range of variable values, and the "used" conditions, by the length of time one remains in area with those attribute ranges. In conjunction with this quantification is the assumption that fish remain in areas for longer periods of time when those areas meet their habitat requirements.

Methods:

The study at hand derives from an investigation into the discrimination, by overwintering juvenile coho salmon, of microhabitats, and whether or not TKE and strain play roles in this decision making process along with distance to wood, depth, and velocity. It is hypothesized that TKE will be a strong factor in coho salmon habitat selection. The investigation took place at the Oregon Hatchery Research Center in Asea, Oregon, and involved the overlaying of mapped flow fields at a 0.1 m scale onto underwater videogrammetry of fish locations(Tullos and Walter, in review).

Distance to wood was computed for each fish using Geographic Information Systems(GIS) in ArcMap and fish coordinate data. In order to do so a linear model was fit to

---

[4] Lacey, R.W. Jay, et. al (2012). The IPOS Framework: Link Fish Swimming Performance in Altered Flows from Laboratory Experiments to Rivers. Received from wileyonlinelibrary.com. DOI: 10.1002/rra.1584

[5] Kemp, J.L., et.al. (10/4/2000). *The habitat-scale ecohydraulics of rivers.* Recieved from http://www.sciencedirect.com/science/article/pii/S0925857400000732 DOI: 10.1016/S0925-8574(00)00073-2

[6] Stewart, G. , et. al. (2005). *Two- Dimensional Modelling of Habitat Suitability as a Function of Discharge on two Colorado Rivers.* Received at (www.interscience.wiley.com). DOI: 10.1002/rra.868

[7] Stewart, G. , et. al. (2005). *Two- Dimensional Modelling of Habitat Suitability as a Function of Discharge on two Colorado Rivers.* Received at (www.interscience.wiley.com). DOI: 10.1002/rra.868

each line which outlined a log in the study site. After conversion of each log polygon into polylines, random points were created along these lines, each with their own coordinate data. The "Near" tool in GIS allowed for the computation of the distance of each fish to the nearest random point on a log polyline. Any fish located beneath a log was determined to have a "distance to wood" of zero, as the function of this variable was to represent the distance between each fish and cover.

A dataset was compiled containing each fish's location, time of observation, and the values of TKE, strain, water velocity, fish depth, and fish distance to wood at those locations in order to quantify and analyze the importance of those variables to a fish's habitat selection. In order to determine the time spans within which a fish remained in a habitat (an area of little to no change in attribute ranges), the normalized derivative of each variable was computed and plotted throughout a fish's trajectory and tested with varying threshold values. Normalization was computed by dividing the difference between variable values for consecutive fish observations by the earlier of the two observed values. Territory time was defined as time spans within which 3 out of 5 these normalized derivative trajectories were below a threshold line at the same time .

In order to maintain objectivity, each variable was tested with the same threshold line. Thus a threshold line of r = 0.4 highlights the time spans within which 3 out of 5 variables changed less than 40% throughout consecutive fish observations. For each threshold value, each matrix of variable trajectories returned a new matrix of the same observations excluding those which spiked above the threshold line. The amount of effect a threshold line had on a variable could therefore be deduced by seeing how the size of each variable trajectory matrix changed. As the size of each of these matrices symbolized the maximum number of fish observations beneath the threshold for that variable, this relationship between the maximum number of fish observations beneath the threshold and the threshold value is plotted in Figure 1. As the threshold value nears 0.5, the slope of the line representing the total number of territory times produced from each threshold begins to decrease. Initially the slope is high, including clusters of territory times beneath the threshold, yet a decrease in slope signifies the inclusion of one or two territory times, rather than clusters. We choose a threshold of 0.5 in order to optimize the inclusion of territory time clusters. The yellow line in Figure 1 displays the leveling out of size of the depth trajectory matrix at a threshold of r = 0.2. Since the maximum rate of change for depth is 0.27 m/s, we thus denoted 0.1 m/s as the threshold for depth and 0.5 as the threshold for all other variables. 1

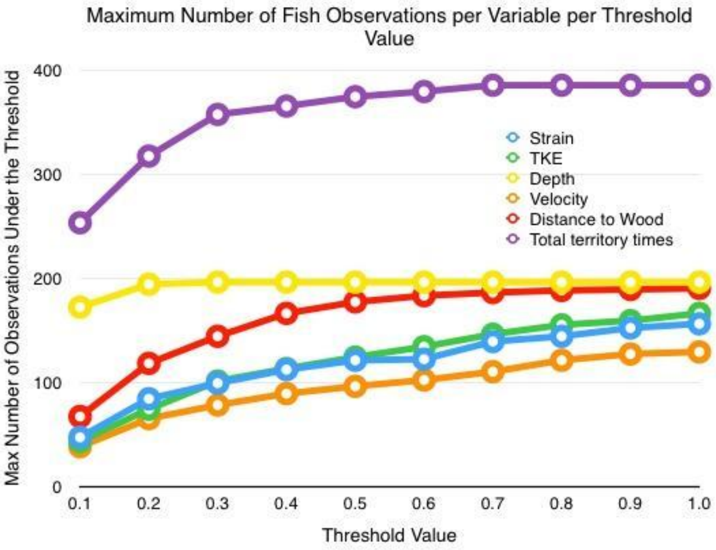
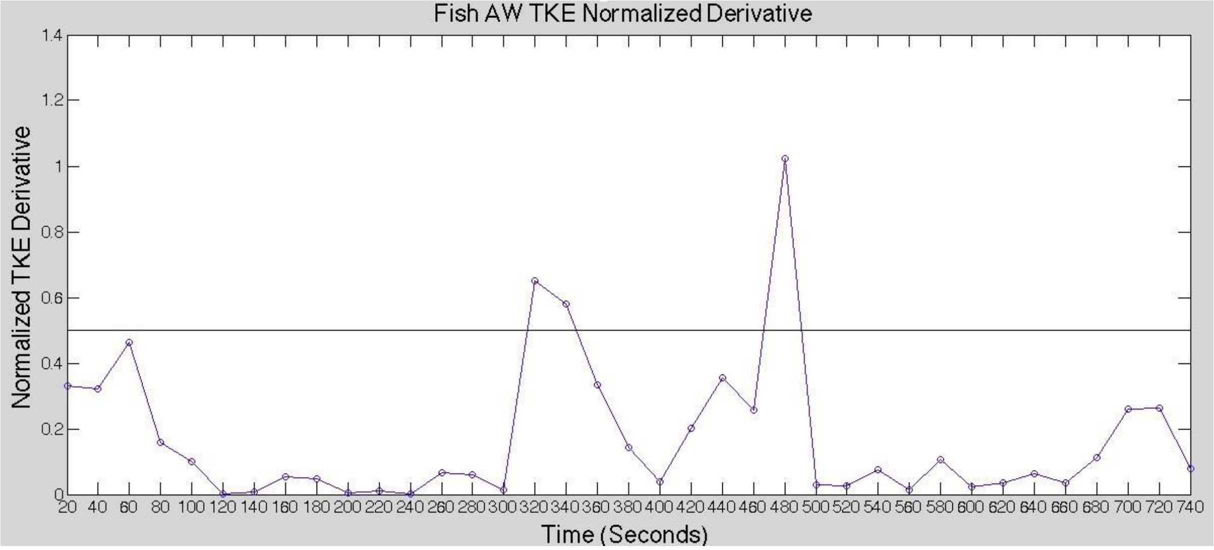


Figure 1. Comparison of the maximum number of fish observation per variable trajectory matrix for each threshold value. Additionally comparing the total number of territory times(purple) produced from each threshold to the corresponding threshold value.

The threshold value determines the section of each fish's trajectory with which our study focuses. Each normalized derivative point under the threshold in the example plot, Figure 2, corresponds to a timestamp at which a fish observation was made. By working backwards to receive the initial and final timestamps, one can deduce the territory time length and relate it to the values of the variables within that time and place. For example, given that at least 2 other variables are below the threshold during this time, Fish AW in Figure 2 has a territory time between the 100 and 300 second timestamps.



2

Multiple Linear Regression:

After defining and quantifying fish territory preference through territory time spans, the importance of each variable to a fish's habitat selections was quantified through multiple linear regression. Before computing the multiple linear regression, outliers were identified by consulting box plots comparing each fish's total variable range and removed when their variable value ranges did not overlap with the others (Appendix A). The quantification of the importance of each physical and hydraulic variable to fish habitat selection is encompassed by the coefficients ßn for n=1:5 within the multiple linear regression equation:

$$\log(\mathit{Territory\ time}) = \sum_{i=1}^{n} B_n X_n \tag{2}$$

$$X_1 = \mathit{Normalized\ average\ of\ TKE\ in\ territory}$$

$$X_2 = \mathit{Normalized\ average\ of\ Strain\ in\ territory}$$

$$X_3 = \mathit{Normalized\ average\ of\ Water\ Velocity\ in\ territory}$$

$$X_4 = \mathit{Normalized\ average\ of\ Depth\ in\ territory}$$

$$X_5 = \mathit{Normalized\ average\ of\ Distance\ to\ Wood\ in\ territory}$$

Since data collection methods hindered the ability to observe many long fish trajectories, and thus long territory times, the territory times are not normally distributed. The multiple linear regression was thus computed with the log of the territory times instead to account for this issue. Additionally, the independent variables were the normalized averages of the variable values in these habitats in order to accurately compare each variable's effect on territory time. Averaging the values in these time spans allowed for the grouping of autocorrelated variable values between consecutive observation times.

The results of this multiple linear regression are shown in Table 1. The coefficient estimate represents the difference in the predicted territory time given a one unit increase in the given variable. The standard error of the coefficients measures how precisely the model estimates the coefficients unknown value. The t-Stat is the ratio between the the coefficient estimate and its standard error. It confirms the variable belongs in the model by displaying the standard deviations of the coefficient from zero. Typically a t-Stat greater than 2 or less than -2 is significant with 95% confidence. P-values determine the significance of results by testing the validity of the null hypothesis. Typically a $p\text{-value} < 0.05$ indicates statistically significant results and a $p\text{-value} > 0.05$ does not. The intercept is the predicted length of a territory time given that the averaged values of the variables in a given place and time are zero. The F-statistic value was obtained through ANOVA analysis. The F-statistic value is a test statistic used to decide whether the sample means are within sampling variability of each other. It is the ratio of the variances of group means to the mean of within group variances (Model Mean Square ÷ Error Mean Square). The null hypothesis is rejected in the F-statistic value is large.There cannot be statistical significance if the F-statistic value is less than or equal to 1. The ordinary R squared value indicates the predictive ability of the model and range from zero to 1. Thus a lower R squared value displays an inability by the model to accurately predict territory time length. Finally, a collinearity test was computed to find the Variance Inflation Factor. Higher collinearities are displayed by higher variance inflation factors, however should only become a cause for concern when greater than five.

Results

Scatter Plots:

Before computing the multiple linear regression model, we plotted the average of each variable value within a territory according to the the log of the respective territory time in order to get a better sense of the relationship. The resulting plots are shown below in Figures 2 through 6 and are color coded according to fish length.

3

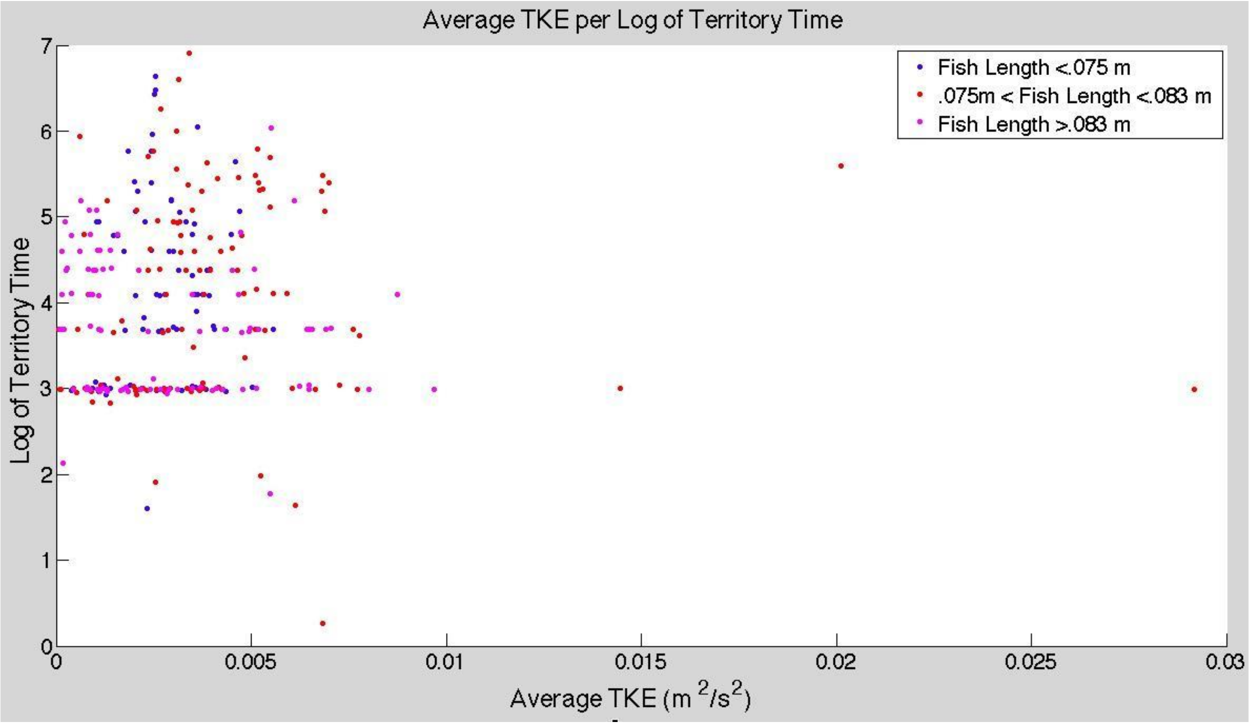


Figure 2. Comparing the average value of TKE in each territory to the log of the length of the respective territory time. Larger fish with shorter territory times are in lower TKE areas than smaller fish with longer territory times. The majority of all territories have average TKE values below 0.1 m²/s².

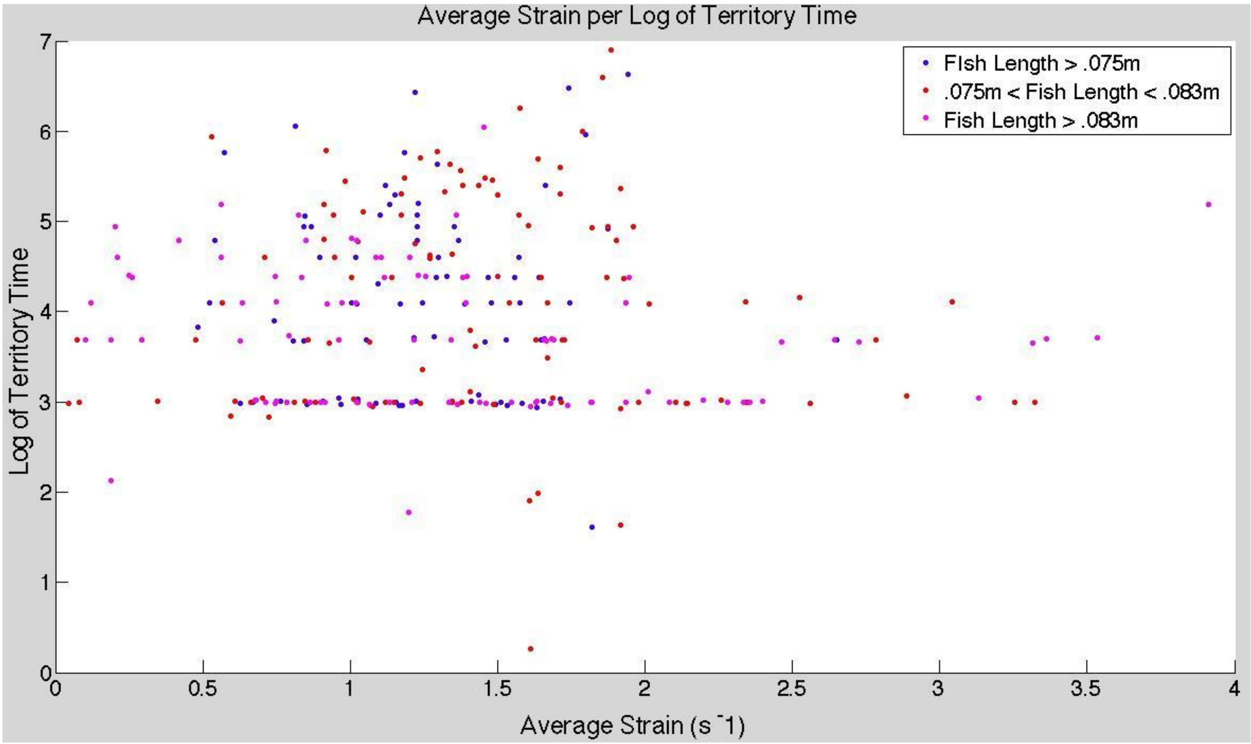


Figure 3. Comparing the average value of strain in each territory to the log of the length of the respective territory time. Scatter points seem to be relatively scattered, though longer territory times seem to cluster with average strain values below 2 s^-1. Clustered according to fish length does not seem to appear.

5

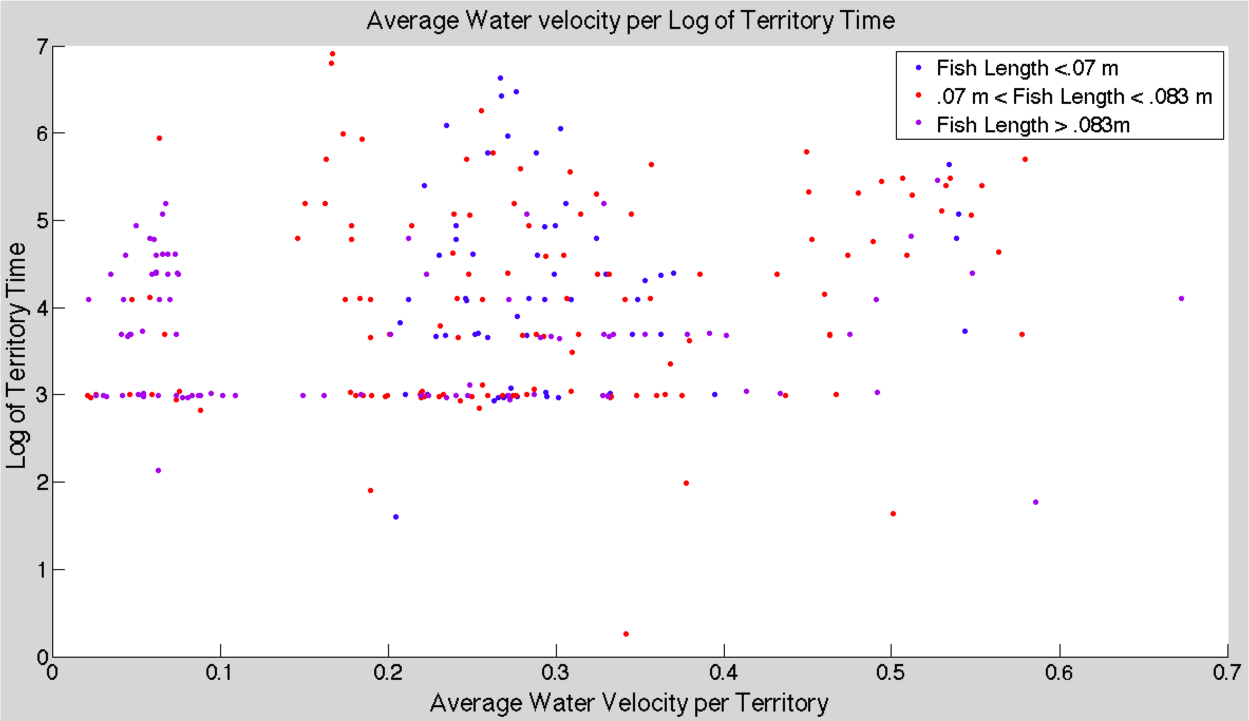


Figure 4. Comparing the average value of water velocity in each territory to the log of the length of the respective territory time. Larger fish with shorter territory times cluster at the velocity ranges 0.0 - 0.1 m/s while small and medium sized fish cluster 0.2 - 0.4 m/s and less so between 0.5 - 0.6 m/s

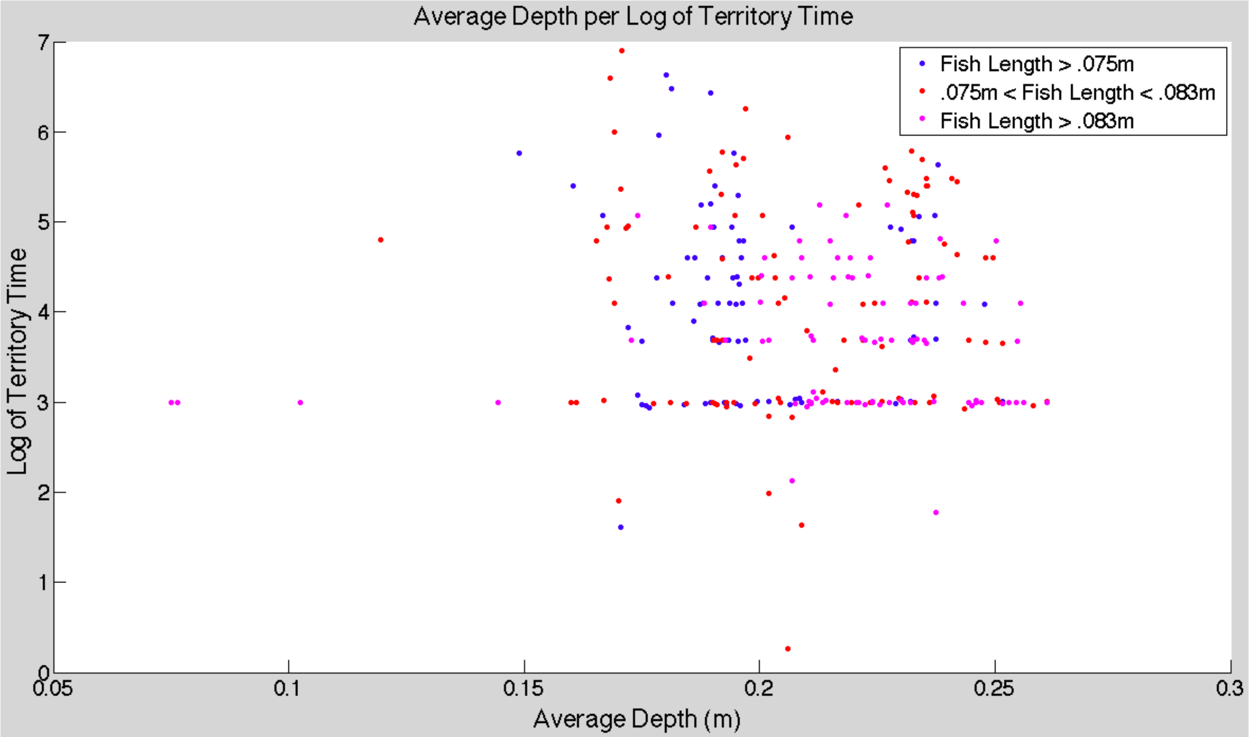


Figure 5. Comparing the average value of fish depth in each territory to the log of the length of the respective territory time. Larger fish with shorter territory times cluster at depths of 0.2 - 0.25 m while small to medium sized fish cluster at depths of 0.15 - 0.2 m with longer territory times.

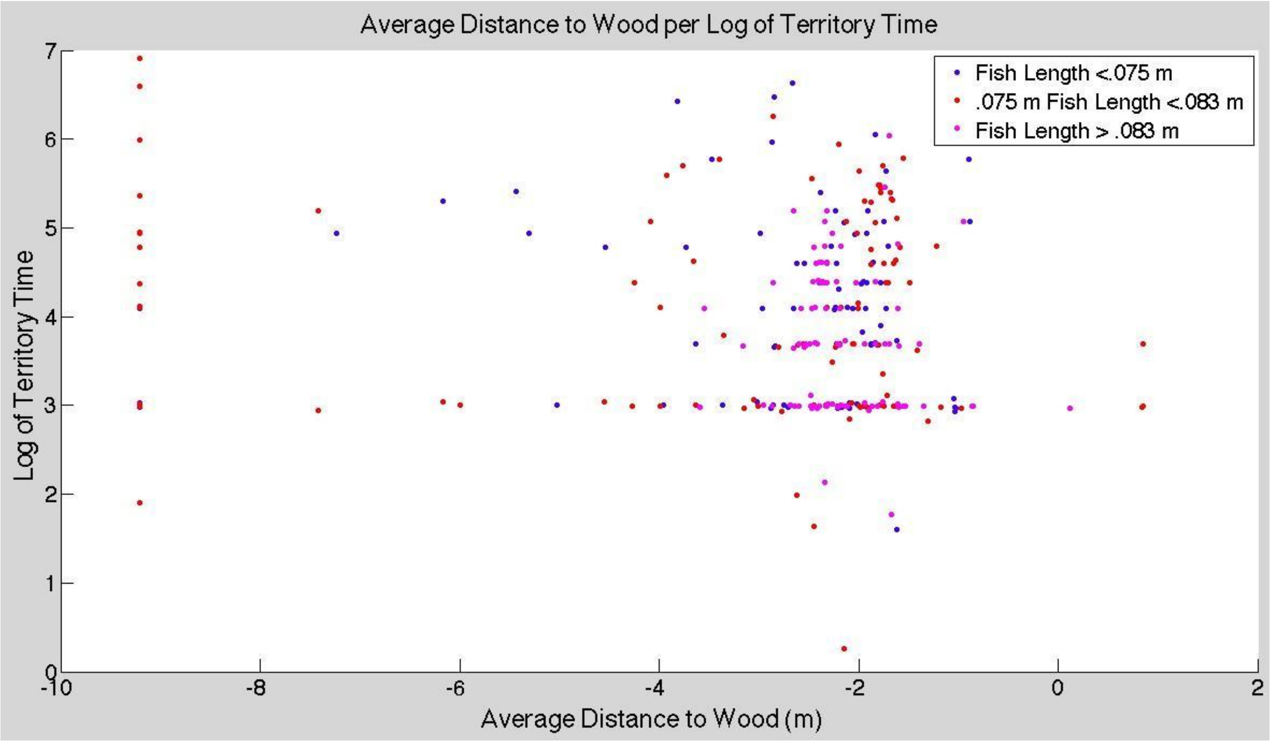


Figure 6. Comparing the log of the average value of fish distance to wood in each territory to the log of the length of the respective territory time. The majority of territories exist within a meter of wood. Smaller fish longer territory times exist closer to wood than do larger fish with shorter territory times.

Below, Table 1 shows the results of the multiple linear regression which indicate that velocity and distance to wood are the two significant factors a juvenile coho salmon habitat selection. Additionally, the R squared value of 0.12 shows that this model is not fit to accurately predict territory time length. According to the coefficients, increased velocity and increased distance to wood significantly increase territory time length.

	Coefficient Estimate	**Standard error of the coefficients**	**tStat = Estimate/SE**	**p-Value**	**F Statistic**	**Variance Inflation Factor**	**Ordinary R Squared**
Intercept	4.88	0.50	9.76	0.00			0.12
TKE Coefficient	-27.14	24.05	-1.13	0.26	1.27	4.74	
Strain Coefficient	-0.24	0.12	-1.91	0.06	3.67	1.54	
Velocity Coefficient	3.60	1.71	2.10	0.04	4.43	4.25	
Depth Coefficient	-3.05	2.27	-1.35	0.18	1.81	1.02	
Distance to Wood Coefficient	-0.59	0.26	-2.30	0.02	5.27	1.06	

Table 1. Multiple linear regression results, including the coefficient estimate, standard error of coefficients, the t-Stat, and the p-Value of each variable.

Additionally, we sought to check the distribution of the residuals. Residuals are deduced by the difference between the observed territory times and those predicted by the multiple linear regression model. The purpose of this is test the normality of the distribution of the residuals. A normal distribution of residuals indicates accuracy in the model. The histogram of the residuals is shown in Figure 6 and the residuals and relatively normally distributed.

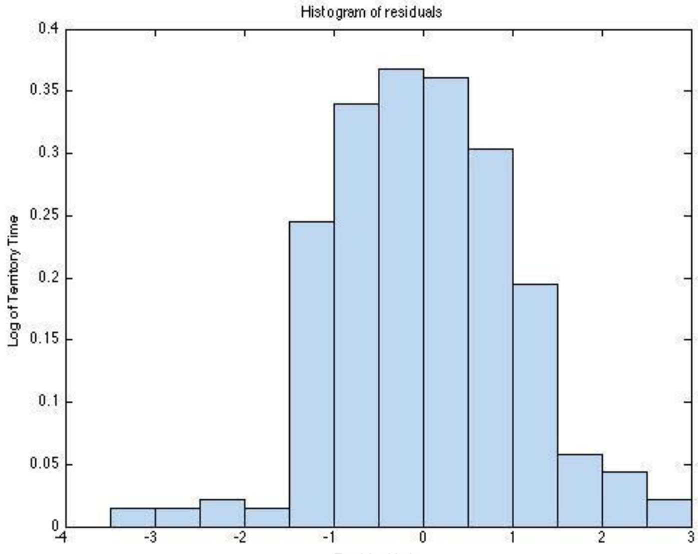


Figure 6. A histogram of residuals from the multiple linear regression comparing the log of territory time and TKE, strain, water velocity, fish depth, and fish distance to wood.

To further analyze effect of significant interactions between variables on territory time, we plotted the depth and distance to wood values both for total fish observations and for territory times in Figure 12. The purpose of such a plot was to see, out of the total locations and variable values available to fish, which were used most heavily through territory use. Territories are shown to cluster overall in deeper waters closer to wood.

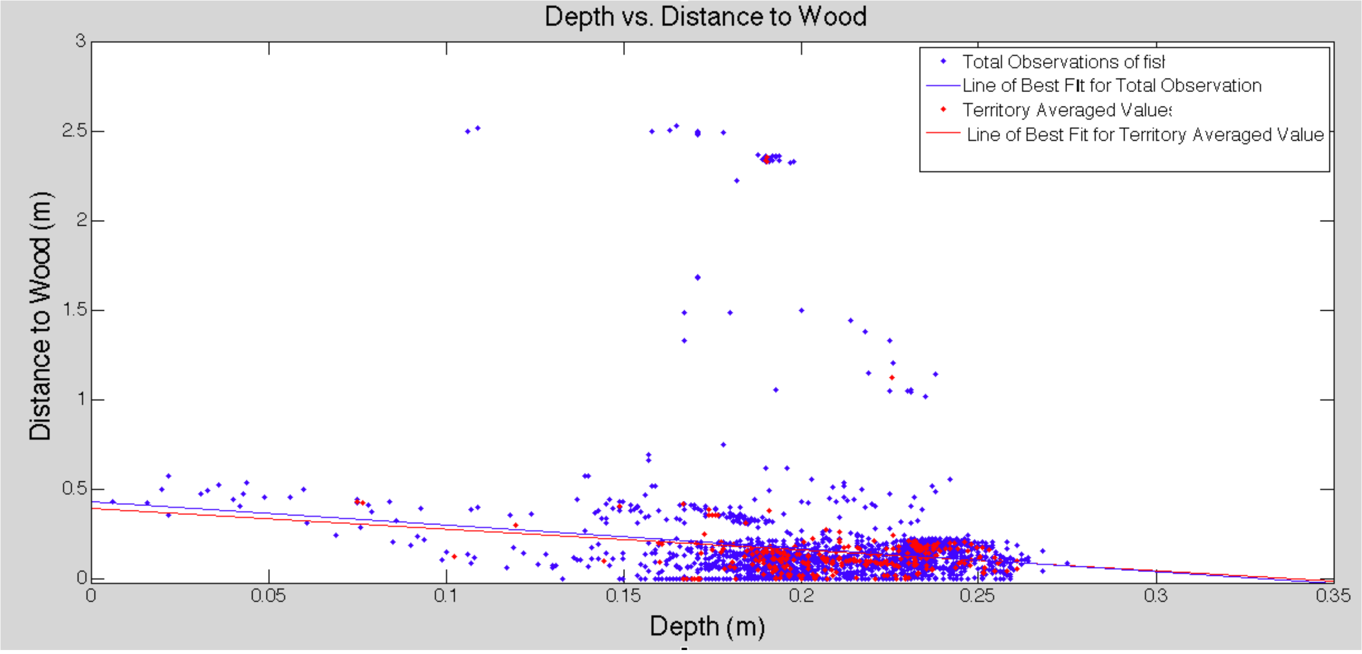


Figure 7. Shows the interactions between depth and distance to wood and compares the variable values used through total fish observations, and those used throughout territory use.

Discussion

In consulting the scatter plots of each physical and hydraulic variable across the log of territory time, strain seems to show no relationship to lengthening territory times. Little research could be found on the effects of strain on coho salmon behavior, and much more should be done.

However, much research has been conducted on the water velocity preferences of fish and these results have predominantly disagreed with the findings of the multiple linear regression model whose results are in Table 1. If we solely review the p-values of this table, the results are a bit unsurprising. Water velocity and distance to wood are the two most significant factors in determining territory time. Many studies have noted this trend as well as the preference by coho salmon to low velocity areas that typically occur in wood covered refuges for fish. This trend is usually even stronger during the winter, the season during which this study occurred and in which uncovered regions have extremely high water velocity levels. However, the velocity coefficient expressed by this model indicates that a one unit increase in water velocity lengthens the territory time of a given fish by 3.6 seconds. In addition to this, a one unit increase in distance away from wood decreases territory time by .6 seconds. Thus by the model at hand, coho salmon like to be close to wood yet in high velocity areas, instead of taking refuge from them.

While collinearities between velocity and TKE are no doubt existent within the observations, VIFs of less than 5 typically do not skew coefficients by that much. The answer may lie instead in fish length comparisons. Figure 4 displays a comparison of averaged water velocities per territory with the log of the lengths of those territory times. Color coded clusters appear throughout this plot with colors dependent on fish length. Larger fish, with lengths greater than .083 meters, cluster in the 0 - 0.1 m/s velocity range and exhibit logs of territory times less than 5.5. Meanwhile, small to medium fish with lengths below .083 meters cluster at higher water velocities(0.2 - 0.4 m/s) with a maximum log of territory time of 7. Such findings are not out of the ordinary, since smaller coho salmon have been found to expend less energy in such velocities than larger salmon[8].

In reference to Figure 6, a similar comparison of small to large fish emerges. While smaller and medium sized fish cluster beneath wood, with a log of distance to wood equal to -8, larger fish to not have to act in such a way. Such behavior reflects the protection wood cover has been shown to have for fish from their predators. While this protection is pertinent for fish of smaller sizes, it is less so for larger ones. Although the difference between fish lengths may be between hundredths of a meter, within the realm of the study these larger fish were still the largest in the channel and freer from risk of predation. This leg up has allowed them to receive some benefits of wood cover, most notably a low velocity, and thus energy expenditure area, while remaining at a presumably small distance from high velocity areas where foraging success is typically the most likely [9] .

We may infer that the although the model does not have a high probability of accurately predicting territory times, as apparent by the low ordinary R squared value, such a model occurred because of behavioral differences dependent on fish length. Thus, in comparing small to large territory times to deduce the importance of each variable to habitat selection, we ultimately compare large and small fish territory variable values with preference for small to medium sized fish territories. It is thus crucial in analyzing these multiple linear regression results to bring attention to the range of these scatter plots. While the smaller fish with the longest territory times have higher velocities than the larger fish with shorter territory times, their average water velocity was 0.2 m/s. Other studies have defined low water velocities for overwintering coho salmon to be 0.15 m/s. In addition to this, almost all fish have selected habitats within a meter of wood (as shown by a log of distance to wood less than zero). Thus we may still infer these fish are selecting habitats with low water velocities and distances to wood. Consideration of scope and range is important in interpreting these results and in future evolutions of the model.

Likewise, the fish length dependent relationship may have skewed the coefficient results for depth within Table 1. The related literature supports the hypothesis that coho salmon habitat selection is highly correlated with deeper depths[10]. Thus, a positive depth coefficient estimate was expected yet unreceived. Such a coefficient may be a result of the relative territory length difference between large fish, clustered at depths 0.2 - 0.25 meters, and smaller fish, cluster at depths of 0.15 - 0.2 meters. Again, the average depth selection of coho salmon is about 0.2 meters, which is shallower than the displayed preference of the larger fish in this study [11]. These scatter plot clusters are in agreement with studies which note a similar behavior between larger and smaller trout, in addition to observations of competition between such fish size groups in

---

[8] Press, G.R, et. al, (2011). Juvenile Salmon Response to the Placement of Engineered Log Jams (ELJS) in the Elwha River, Washington State, USA. *River Research and Applications* (Impact Factor: 2.43). DOI: 10.1002/rra.1481

[9] Press, G.R, et. al, (2011)

[10] Howard, C. (2006), Mill Creek Fisheries Monitoring Program: Ten Year Report. Final Report to Save-the-Redwoods League. *Stillwater Science*s*, Save-the-Redwoods League*. Received at http://www.savetheredwoods.org/media/pdf_howard.pdf.

[11] Mills, T.J, et.al. (2004). Matrix of Life History and Habitat Requirements for Feather River Fish Species - Coho Salmon. *State of California Department of Water Resources*v

which larger individuals "won" these preferred habitats[12]. If we infer greater validity in the scatter plots, and that preferred conditions are actually displayed in shorter territory times, as those are the time spans associated with the larger fish who win them, then we may be able to infer preferences for velocity and TKE.

Although TKE had a relatively higher VIF, its collinearity does not skew the data, and thus can still signify a quantified role in the coho salmon decision making process. While high turbulent kinetic energy can cause turbulent shear stress on fish bodies, it can also propel fish forward and reduce energy expenditure. Such reasons to prefer and avoid high TKE levels could cause statistical insignificance when considering territory times. [13] However, the scatter plot in Figure 2 displays the same fish length dependent territoriality preferences as the scatter plots mentioned here. Almost all territories exist with average values of TKE below 0.01 m^2/s^2 while larger fish with shorter territory times cluster at TKE values below 0.005 m^2/s^2 . If assume larger fish are winning these lower TKE values through competition with the smaller ones, we may infer that lower TKE values are preferable to coho salmon habitat selection.

Finally, the insignificance of the depth coefficient contrasts the majority of findings which name depth, as well as velocity and distance to wood, as a significant factors in determining coho salmon habitat selection. Figure 7 indicates habitat selections by these fish should typically in be in deeper water closer to wood, in comparison to the total fish observations made in the study. The current p-value of the depth coefficient may be result of threshold sensitivity by the depth variable values. In reference to Figure 1, we see that depth is unaffected by threshold values greater than 0.2. For this reason, we gave depth a threshold of 0.1, however such a threshold only excludes about 30 observations from the depth trajectory matrix, indicating that depth is much more sensitive to threshold values that other variables in question. In the interest of time, the threshold value remained at 0.1 for depth, however future studies should further analyze this relationship and its effect on the depth coefficient estimate significance.

Conclusion

As humans move into an age of undoing many of the harmful manipulations they have caused in river ecosystems, it is our responsibility to act as environmentally aware as possible. Riverine habitats with the appropriately preferred characteristics have been shown to increase fish biomass, biodiversity, and abundance[14]. Such studies of the influence of these preferred characteristics can better parse out where river restoration projects should focus. The goal of this project was to quantify the importance of five physical and hydraulic characteristics in coho salmon habitat selection. It is possible that the quantifications presented in Table 1 of this paper present valid contradictions to the norm of the related literature and bring into question the basis of other studies. However before such claims may be met, further work must be done into the effect of fish length and competition on our habitat definition. Possibly the greatest question of all presented through this work is why larger are remaining in habitats for shorter amounts of time than smaller fish. Additionally, future studies should further analyze the effects of varying

---

[12] Kielland, O.N. (). Size-dependent habitat use of juvenile brown trout (Salmo trutta L.) in an artificial river. Received at http://www.diva-portal.org/smash/get/diva2:640170/FULLTEXT01.pdf

[13] Lacey, R.W.J., et. al.(2012). The IPOS Framework: Finking Fish Swimming Performance in Altered Flows from Laboratory Experiments to Rivers. *River Restoration Application* (28: 429-443). doi:10.1002/rra.1584.

[14] Railsback, S.F. (2009). InSTREAM: The Individual-Based Stream Trout Research and Environmental Assessment Model. *U.S. Department of Agriculture, Forest Service, General Technical Report PSW-GTR-218*.

threshold on each independent variable in question. For the sake of time, little threshold analysis could be done prior to the completion of this paper, however it is quite probable that depth thresholds are impacting the p-value of the depth coefficient, and possible that similar processes could be affecting other variables in the study.