Eco-Informatics Summer Institute 2009

Wood in Streams

Hoan-Vu Do
Jonathan Gillick
Gregory Reeb

Wood in Streams

Since the 1980's, much effort has been devoted to studying the dynamics of large wood in streams, especially in the Pacific Northwest where high winter flows and frequent landslides make wood more dynamic than in other parts of the country. We find two foci in the literature on large woody debris of relevance to our study. These are: prediction of wood volumes per stream length throughout a stream network, and accelerated stream habitat restoration of a reach through the placement of large woody debris.

At the network scale of prediction, volume of wood per stream length is seen as a useful measure of stream habitat functioning. Therefore, since the 1980's there have been repeated calls to better predict the effects of various land management practices and large natural disturbances on the volume of wood in a network. Studies have examined effects of the presence of roads near streams, clear-cutting practices, and the tendencies of wood to redistribute in a stream network as a result of large winter floods and landslides. This work has resulted in many positive changes in forestry policy with regards to riparian zone protection and the maintenance of healthy volumes of large wood in streams. Additionally, the study of the effects of large natural disturbances on wood dynamics in streams has greatly increased our understanding of system functioning and has led to many removals of stream road crossings to allow for necessary wood mobility (Czarnomski et al. 2008).

Separate, although not completely, from the focus on network-scale wood volume prediction is the literature on large woody debris installation for reach-scale restoration projects. Millions of dollars are spent each year on watershed restoration and stream habitat improvement in the U.S. Pacific Northwest in an effort to increase fish populations (Roni et al. 2002), so there is great encouragement to discover the best ways to implement the placement of large wood. In addition to the economic investment in restoration projects, there is a liability associated with large wood breaking free and possibly damaging property downstream. Also, it is easy to see that installed wood that leaves a restoration reach can neither function in that site nor help accumulate new wood. These economic as well as ecological pressures amount to the necessity to impose relative stasis upon installed wood.

As we hinted, network scale considerations are not entirely separate from those on the reach scale. In order to identify a stream reach as in need of habitat restoration, one may use a network scale. Considering the whole network allows for the identification and understanding of sections of a stream with persistent low wood volume. Once a stream reach is identified as having unacceptably low volumes of wood, we must ask why. The cause is often the remnant of anachronistic forestry policy that allowed for the disturbance of riparian old-growth forests. Once the direct cause has been addressed—by, say, replanting clear-cut riparian forests or thinning young riparian forests to encourage the growth of large conifers—planners can choose to accelerate wood accumulation in the reach through the artificial placement of large wood in the stream.

Having used the network scale to identify and understand low wood volumes in sections of a stream, we now shift to a reach scale and see the network as its context. Within the reach scale our main question aims at learning under what conditions wood moves the least and accumulates the most additional wood. Transitioning from the network scale to the reach scale involves changing our environmental factors and considering wood in a new way (Fig. 1).

Network Scale Prediction		Reach Scale Restoration
Floods/landslides	→	Annual peak flow
Stream order	→	Channel width
Wood volume per stream length	→	Each log: Length, diameter, volume, root wad, zonation, structure type

Figure 1. Comparisons between entities in network scale wood volume prediction and those in reach scale restoration of woody debris.

In a reach scale we now greatly expand on wood attributes, considering for each log its length, diameter, volume, root wad presence, zonation, and type of structure in which it was installed (for some analyses). We also reduce environmental context for each log to the annual peak flow and the average channel width for the entire restoration reach. We aimed to test whether this was a sufficient system to allow for prediction of wood movement in a specific reach and to learn which of these factors played the largest role in determining wood movement. We also looked for correlation between the number of key pieces (defined as logs greater than or equal to the average channel width) at a position in a reach and the number of accumulated logs.

For our analyses we used data from the Quartz Creek restoration site—Quartz Creek being a major tributary of the McKenzie River in Oregon. In 1988, multiple logs were installed into the Quartz Creek restoration reach in various structural configurations. All logs in the reach have been monitored yearly since 1988.

Separately, we investigated whether Mark Meleason's Streamwood model could accurately predict wood movement in Quartz Creek.

Methods. We are interested in predicting the probability of movement of a log based on the data we have from 20 years of wood mapping at Quartz Creek. From attributes of individual logs—length, diameter, volume, position in the stream, presence or absence of a root wad—and the peak annual flow for Quartz Creek, we constructed three models with two objectives:

1: model probability of movement of a log

2: determine significance of each factor in affecting probability of movement

The Quartz Creek data set contains locations of approximately 4000 logs tracked between 1988 and 2007. Since some logs entered the stream after 1988 and some left before 2007, and since flows changed from year to year, we decided to consider logs on a yearly basis for our models. For our purposes, then, the data set contains about 33,000 log/year pairs. Each model gives the probability of movement for a log in a year, and to compute probability over a longer time scale, we can use a geometric distribution. So if the probability of movement in a year given by the model is p, the probability of moving in year n, $P(n) = p(1-p)^{n-1}$. Probability of moving within the first n years is given by $P(1) + P(2) + \ldots + P(n)$. Note that the response variable in each model is a binary "Yes" or "No" for whether a log moves. We consider a change in location of greater than ten meters as movement, and change of less than ten meters as no movement. We chose ten meters because our analyses are within the context of restoration—i.e restoring habitat functioning—and both extensive reports submitted about the Quartz Creek restoration (one in 1993 and another in 1998) stated that the majority of logs and structures that moved ten meters or less still performed their initially intended functions.

Our first model is a linear equation created by linear regression on the data set. We used the method of least squares to get the following equation, which is explained in table 1:

(Eq. 1) $p = B1 + (B2 * L) + (B3 * RW) + (B4 * V) + (B5 * D) + (B6 * Z) + (B7 * PF)$

Symbol	L	RW	V	D	Z	PF
Meaning	Length	Root Wad	Volume	Diameter	Zonation	Peak Flow
Units	meters	Binary	m^3	meters	No Units	m^3/s

Table 1: Inputs and abbreviations

B1	B2	B3	B4	B5	B6	B7
0.2365	6.459e-3	-2.863e-3	7.775e-3	-0.1518	-3.106e-4	5.338e-04

Table 2: Coefficients in Linear Model

In the Quartz Creek data set, zonation is denoted by the percent of a log in zones 1 through 4 of the stream. Each log has values z_1(percent of log in active channel), z_2(percent in bankfull channel), z_3(percent above bankfull channel), and z_4(percent outside bankfull channel). We calculated our zonation value Z by Z = $(z_1 + 3z_2 + 5z_3 + 7z_4)$, yielding values for Z between 100 and 700.

For our second model, we determined the logistic equation for probability of movement shown in equation (2) by logistic regression on the data set using the method of least squares. Logistic equations are of the form p = 1 / (1+exp(-x)), where x is a function of the inputs to the model. Table 3 shows the coefficients in the logistic equation.

(Eq. 2) p = 1 / (1 + exp(- (C1 + (C2 * L) + (C3 * RW) + (C4 * V) + (C5 * D) + (C6 * Z) + (C7 * PF))))

C1	C2	C3	C4	C5	C6	C7
-0.806	-0.0786	0.0618	-0.0438	-1.775	-0.0029	0.00409

Table 3: Coefficients in Logistic Model

Our third model is a Bayesian Belief Network (BBN). BBN's consist of conditional probability tables with input variables affecting the probability of response variables. In our model, movement is the single response variable, and all other inputs affect movement. Figure 2 shows the setup of our model in Netica, a graphical interface to run BBN's.

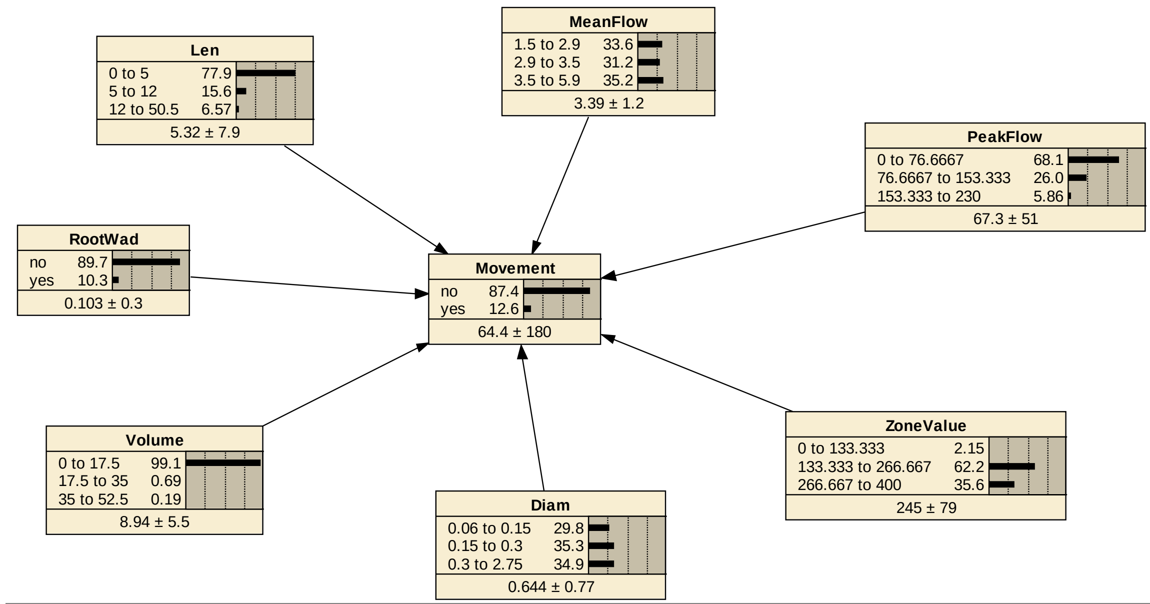

Figure 2: Bayesian Belief Network

Results. To test the effectiveness of our regression models, we randomly selected 25,000 log/year pairs as training data, ran the regressions, and then tested the models on the other 7859 data points. To test the models, we made a prediction based on each model of whether each log in the test case would move and then compared the predictions to the true results. The models output probabilities, so to make predictions on movement, we selected a threshold k, where $0 < k < 1$ and predicted movement if the model gave probability greater than k.

Only 13 percent of the logs in the data actually moved, and we did not find any identifiable subset of logs for which the probability was very high, so we expected predicting movement to be difficult. For the logs in the test set, the highest probability outputted by the linear model was 0.32, and the highest probability given by the logistic equation was 0.42. A typical threshold might be set at $k = 0.5$, but given our data, we had to choose a lower value. To find an optimal threshold, we tested our models on 250 values for k between 0 and 0.25. We found that for the linear model, k = 0.113 gave the best results, and for the logistic model, k = 0.122 worked best. These values are quite close to the fraction of logs that actually moved, so we are effectively saying that if the probability of movement is higher than the mean, we predict movement.

Since probabilities of movement are so low, it turned out that to get the best accuracy (the highest proportion of correct predictions), we should just predict no movement for every log. With this model, we predict correctly on 87 percent of the logs. However, we predict correctly on none of the logs that actually move. No value of k for either model achieved a better accuracy than 87 percent.

Although we cannot beat the accuracy of predicting zero movement for each log, there are other measures typically used to measure effectiveness of models. We looked at two measures, **Precision** and **Recall**. Recall is the fraction of logs that moved that we predicted to move, and precision is the fraction of logs that we predicted to moved that actually did move. Recall says something about how often we recognize that a log will move, and precision denotes the ratio between false positives and true positives.

Precision = True Positives / (True Positives + False Positives)

Recall = True Positives / (True Positives + False Negatives)

We decided to calculate precision and recall for each model and use a measure that combines the two, called an **F-measure** or **F-score**. The F-score that we used weights the two factors equally by taking their harmonic mean. Precision, recall, and F-scores all give values between 0 and 1. An F-score of one represents perfect prediction.

F = 2 * (precision * recall) / (precision + recall)

We determined the optimal k values by the highest F-score. Results are summarized in table 4.

Model	**Precision**	**Recall**	**F-Score**	**Accuracy**
Predicting 0	undefined	0	Undefined	0.87
Random Guessing	0.13	0.47	0.20	0.50
Linear	0.16	0.83	0.27	0.41
Logistic	0.17	0.74	0.28	0.48

Table 4: Test Results

Note that we did not perform calculate precision, recall, and F-score for the BBN because of time constraints. We do not expect to have gotten much better results than the linear or logistic regression models, though, because the BBN predicts probability of movement by the fraction of logs that moved that fit into certain ranges for each parameter, and these probabilities were similar to ones from the other models.

Our second objective in creating these models, besides predicting movement, was to determine which factors were more influential in whether or not a log moved. Each regression model outputted t-values for each parameter, which measure the statistical significance of the correlation between mean values of two parameters. A high t-value for a parameter, then, indicates that it is strongly connected to whether a log moved. For the BBN model, Netica provides a “Sensitivity Analysis” function, which measures how much the variance of the response variable changes when values are removed from an input variable. A high sensitivity value indicates a correlation between movement and an input variable. Tables 5, 6, and 7 show the t-values and sensitivity values for the models. Table 8 shows the relative importance of parameters according to each model.

Parameter	**Diameter**	**PeakFlow**	**Length**	**Zonation**	**Volume**	**RootWad**
t-Value	-13.092	12.238	-9.936	-9.243	7.037	-0.400
P(>|t|)	< 2e-16	< 2e-16	< 2e-16	< 2e-16	2.02e-12	0.689

Table 5: t-Values for Linear Model

Parameter	**Diameter**	**PeakFlow**	**Zonation**	**Length**	**Volume**	**RootWad**
t-Value	-11.971	11.913	-9.052	-8.707	3.365	-0.607
P(>|t|)	< 2e-16	< 2e-16	< 2e-16	< 2e-16	7.66e-4	0.5436

Table 6: t-Values for Logistic Model

Parameter	**PeakFlow**	**Volume**	**Zonation**	**Diameter**	**Length**	**RootWad**
Variance Reduction	1288	322.2	62.09	50.62	6.201	0.2586
Variance of Beliefs	5.76e-3	1.44e-3	2.78e-4	2.27e-4	2.78e-5	1.2e-6

Table 7: Sensitivity Values for BBN model

Linear	**Logistic**	**BBN**
Diameter	Diameter	Peak Flow
Peak Flow	Peak Flow	Volume
Length	Zonation	Zonation
Zonation	Length	Diameter
Volume	Volume	Length
Root Wad	Root Wad	Root Wad

Table 8: Relative Significance of Inputs (in Order of Importance)

Conclusions. Our logistic model for probability of movement seems to be the best model of the ones we have examined. The linear model does not necessarily give probabilities between 0 and 1, and so might not make sense if applied to new data sets, and it also has a slightly lower F-score than the logistic model. Still, the logistic model is not a very good predictor for whether a log will move or not, and the results suggest that we need to take other factors into account to make accurate predictions. The equation is useful though to understand generally what types of logs will move, and what how a particular change in one or more of the parameters will affect the probability of movement.

Future Work. Our objective in creating these models was to predict probability of movement within a one-year time frame, which has proven to be a difficult and perhaps impossible task given our data. A more reasonable, and perhaps more useful goal, however, is to predict log movement over longer time frames. The inherent problem with trying to predict movement within one year is that no matter what model we use, probabilities are going to be very low—i.e. 13 percent on average—so if we ever predict a log to move, we will do so with little confidence. If we look at a longer time frame, though, we can be more certain. A log that our logistic model gives a 13 percent chance of moving in one year would have a 65 percent chance of moving in 10 years (without correcting for change in flow). Our models could easily be adjusted to predict movement over any number of years, and we expect much higher accuracy in predicting movement over longer time frames. The expected number of years before a log moves, or survival rate, might be a useful number.

Streamwood. OSU Streamwood is a computer modeling program that predicts wood accumulation for stream orders less than five in forested streams in the Pacific Northwest (Meleason et al., 2001). OSU StreamWood is an individual-based, stochastic model composed of two submodels: a forest model and a wood model. The forest model simulates riparian forest growth under various management regimes and outputs logs, which are then inputted to the wood model. The wood model simulates recruitment of trees to the channel and subjects all logs in the stream to in-channel process such as breakage, movement, and decomposition.

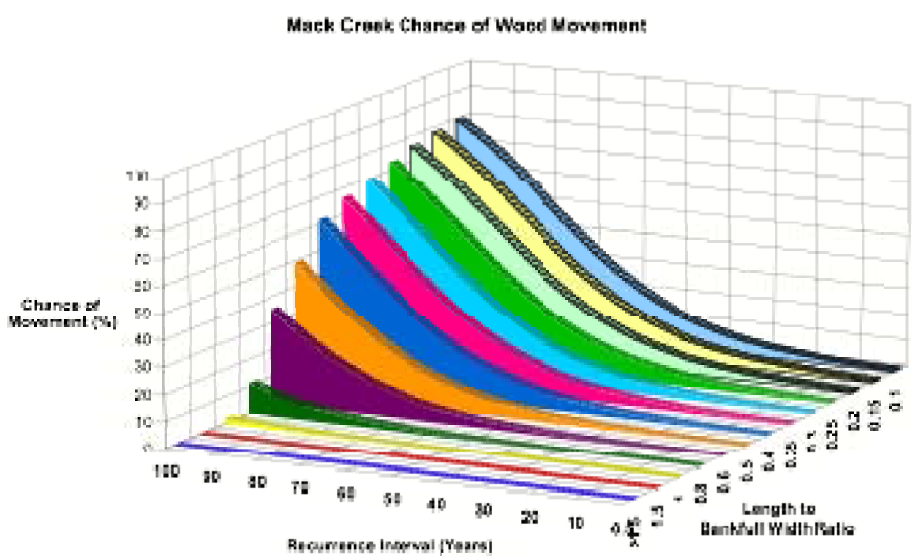

Figure 3. Theoretical model for chance of wood movement in streams. Chance of movement is a function of flow (recurrence interval in years), piece length to bankfull width ratio, proportion of the piece outside the channel, and number of key pieces in the reach.

The probability of a log moving according to Meleason et al. (2001) is a function of flow (recurrence interval in years), the piece length to stream width ratio, number of key pieces, and the proportion of a piece outside the stream channel. Meleason's theoretical model for chance of wood movement in streams showed that as the length to width ratio decreases and recurrence interval increases, the chance of wood movement increases. Figure 3 shows the relationship between two of the four variables that affect the chance of wood movement in the StreamWood model.

Comparing this theoretical model with actual data from Quartz Creek, we found that the StreamWood model does not do well at predicting the chance of movement for small length to bankfull width ratios. Figure 4 shows the relationship between length to bankfull width ratio and recurrence interval for Quartz Creek based on the probability of movement generated with Netica software. We can see from Figure 4 that for bankfull width ratio from 0.15 to 0.75, the chance of movement is relatively high even for small recurrence intervals. Meleason's model showed that the predicted chance of movement is very low or non-existent for bankfull width ratio from 0.1 to 0.8 with low recurrence interval. One possible reason why the prediction might be different is that Meleason's theoretical model only has four parameters, which certainly do not capture all the variables that affect wood movement. In the natural environment, there exist other important parameters that are not easily modeled, such as the interaction of logs with one another, the substrate that a log rests on (rocks and boulders), and local stream characteristics (such as calm section versus rapid section of a stream). Also, the Streamwood model considers each log individually, although in actuality, neighboring logs tend to move as a set. One other interesting aspect from the Quartz Creek graph is a large dip in the chance of movement to almost zero at the third recurrence interval. One possible explanation for this dip is that during a high flow, wood may move. When the wood stops moving—i.e. when it is trapped—it can be seen as stable relative to the high flow it is experiencing. Then, in the next year, if there is a high winter flow, but not as high as the previous year, wood pieces that had moved the year before to stable positions will tend not to move the next year, barring an even higher flow than the year before. Potentially, we think, this can explain the dip in figure 4, but the relationship between log movement and flows in successive years needs to be investigated.

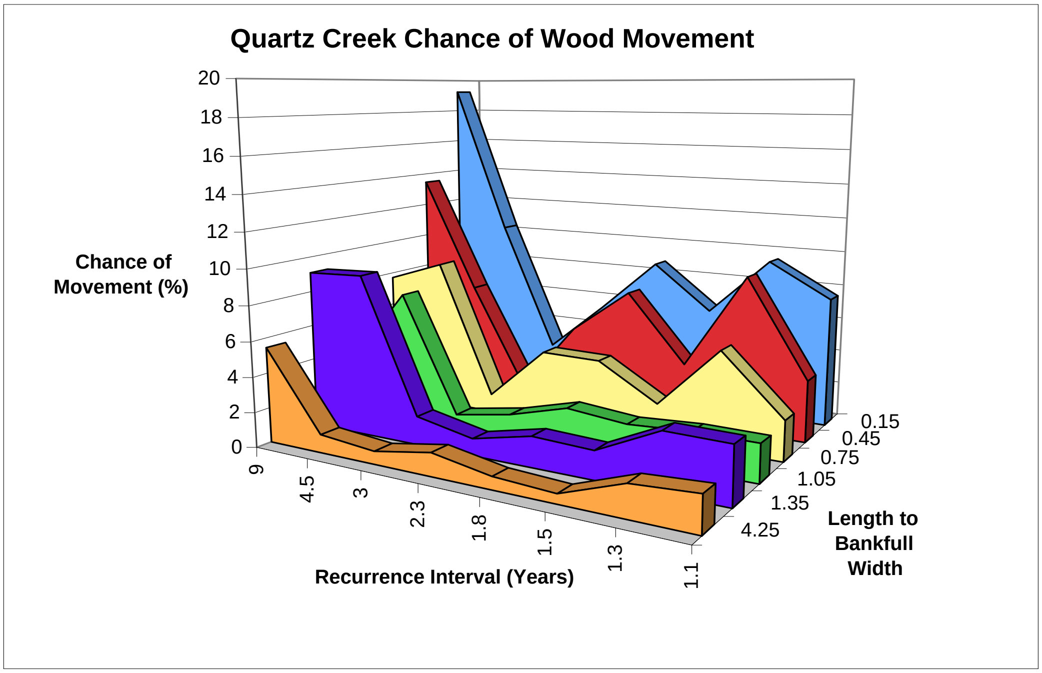

Figure 4. Wood Movement at Quartz Creek

Figure 5 shows the total wood volume for Quartz Creek from 1988 to 2007 over the 1 kilometer survey area. The 335 m position at the site had a higher wood volume than any other position in the reach, and a consistently high volume until 2004. Then, in 2005, the total wood volume at that position began to decrease, eventually down to just over half the total volume of 2004. The graph suggests that there must have been a high flow that took place that moved some of the wood out of the system. This information supports the dip seen in Figure 4, because the 335m position has the highest volume and it lost almost half of its original volume in a single year due to a high flow that moved a lot of tagged logs out of the reach. Thus, with more than half of the original wood volume moved out of the system, there are fewer tagged logs for which to record movement. When generating probability of movement for different categories of length to bankfull width ratio for a particular recurrence interval, we see a big dip in the graph for probability of wood movement. We believe this is because large amounts of wood moved out of the system before it was recorded.

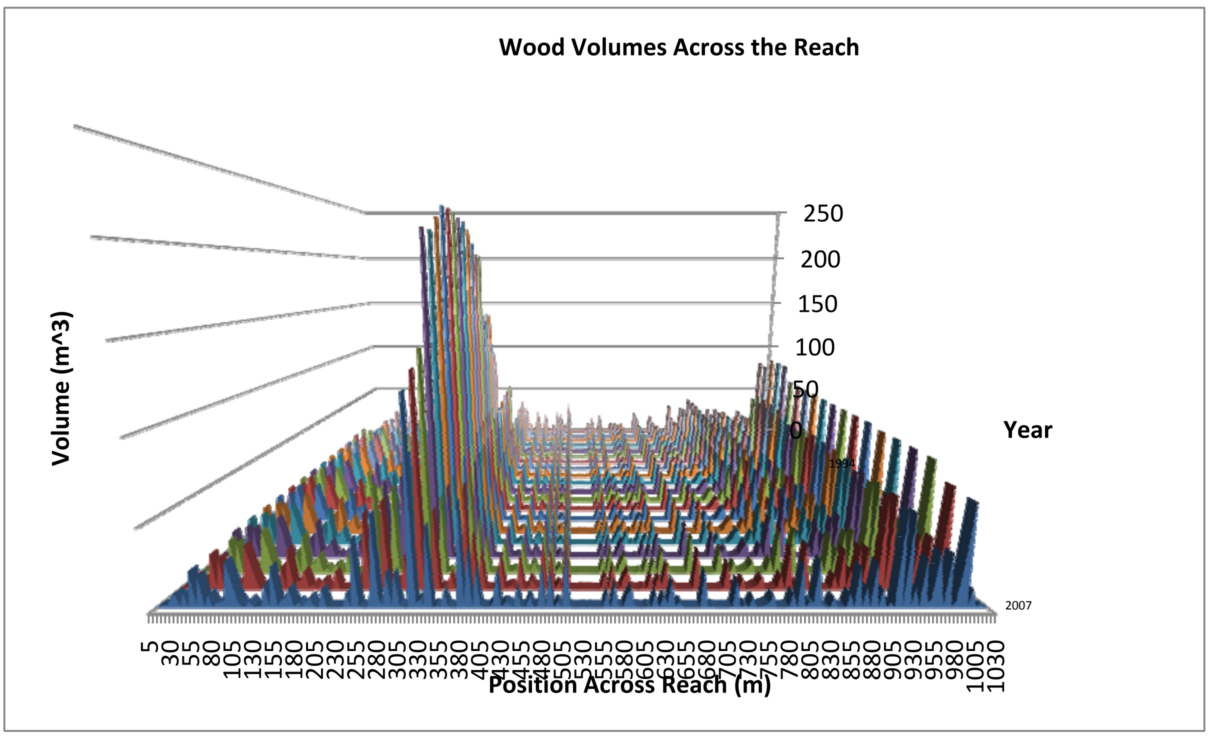

Figure 5. Total wood volume across Quartz Creek Restoration site from 1988 to 2007.

References

Czarnomski, N.M., D. Dreher, K Snyder, J. Jones, F. Swanson. 2008. Dynamics of wood in stream networks of the western Cascades Range, Oregon. NRC Research Press.

Roni, P., T. Beechie, R. Bilby, F. Leonetti, M. Pollock, G. Pess. 2002. A review of stream restoration techniques and a hierarchical strategy for prioritizing restoration in Pacific Northwest watersheds. North American Journal of Fisheries Management 22:1-20.