# Critique of D.F. Flora's "Evidence..."

***R.B. Nerf, Jr.***

The statistical analysis and conclusions in the paper "*Evidence of Near-Zero Habitat Harm from Nearshore Development*," by D.F.Flora, are invalid. This article illustrates why. I demonstrate that Mr. Flora's approach can not reliably detect the presence or absence of a causal link between stressors and ecological function. From comments about the paper on *kitsapsun.com* and *pugetsoundblogs.com* I understand that Flora intended his paper to be understandable to the non-scientist, so I will try to wrap the necessary technical details of my argument in a narrative that gives some context to the science.

When I first encountered Flora's paper, I thought it was intended for a scientific audience because he used footnotes and technical terms such as *coefficient of determination* and *regression*, and he asserted numerical thresholds that separated significant from insignificant for the statistics he had calculated. But as a scientific work, it fails to give enough detail on what he actually did. Another scientist can not form a judgment on the merit of his arguments. Following his references, I looked at the shoreline assessments that created the data which he had analyzed, and found some large data sets and long reports: about 250 pages each for Bainbridge Island[1] and Eastern Kitsap[2] shorelines.

Not wanting to diligently plow through 500 pages of detail, I tried to just trace the original observations ("stressors" such as bulkheads, moorings, floats, and "ecological function indicators" such as eelgrass and geoduck beds), as they wended their way through intermediate calculational stages to the data set that Flora had used for his analysis. The use of a conceptual model was interesting: stressors such as a bulkhead don't kill eelgrass directly; they do so by changing a "controlling factor" such as sediment supply that directly affects the organism (but are too expensive to measure directly). Similarly, I noted that the ecological function indicators were limited to those that were most economical to acquire.

As I looked at the details of the intermediate processing of the observations, I became concerned because any causal link between stressor and ecological function was being obscured by changing observations into "scores", binning, averaging, etc. Puzzled, I went back and read the introductory material to the assessment and found that it was never intended to supply evidence of causality; causality is explicitly assumed in the Bainbridge report:

> "*This assessment is based on the general assumption that alteration of shorelines often results in a change in nearshore ecological functions*." page 2

And the reason behind the scoring/binning/averaging was:

> "...the primary objectives of the Bainbridge Island nearshore habitat assessment effort were to
> - Delineate management areas (MAs) and appropriate subareas
> - Characterize the ecological features and conditions within those MAs
> - Provide a baseline assessment of nearshore ecological functions using repeatable methods
> - Consolidate this information into a single, GIS-based database that can be used by planners and resource managers.

So Flora had done his analysis on a data set that was not collected with the causality question in mind, and, furthermore, the data had been processed in a way that might obscure any connections that were there.

I'm not a statistician, but a physicist. However, I've intensively used regression analysis on at least three major projects. One of these was a baptism of fire because of high scientific interest from my colleagues combined with substantial financial consequences to my employer. A thorough series of “peer reviews” taught me some of the myriad pitfalls that await the regression practitioner. (I've collected a few contemporary examples in Appendix A - Pitfalls.) The fact that Flora gave no indication that he was aware of these potential problems piqued my skepticism of his analysis and conclusions.

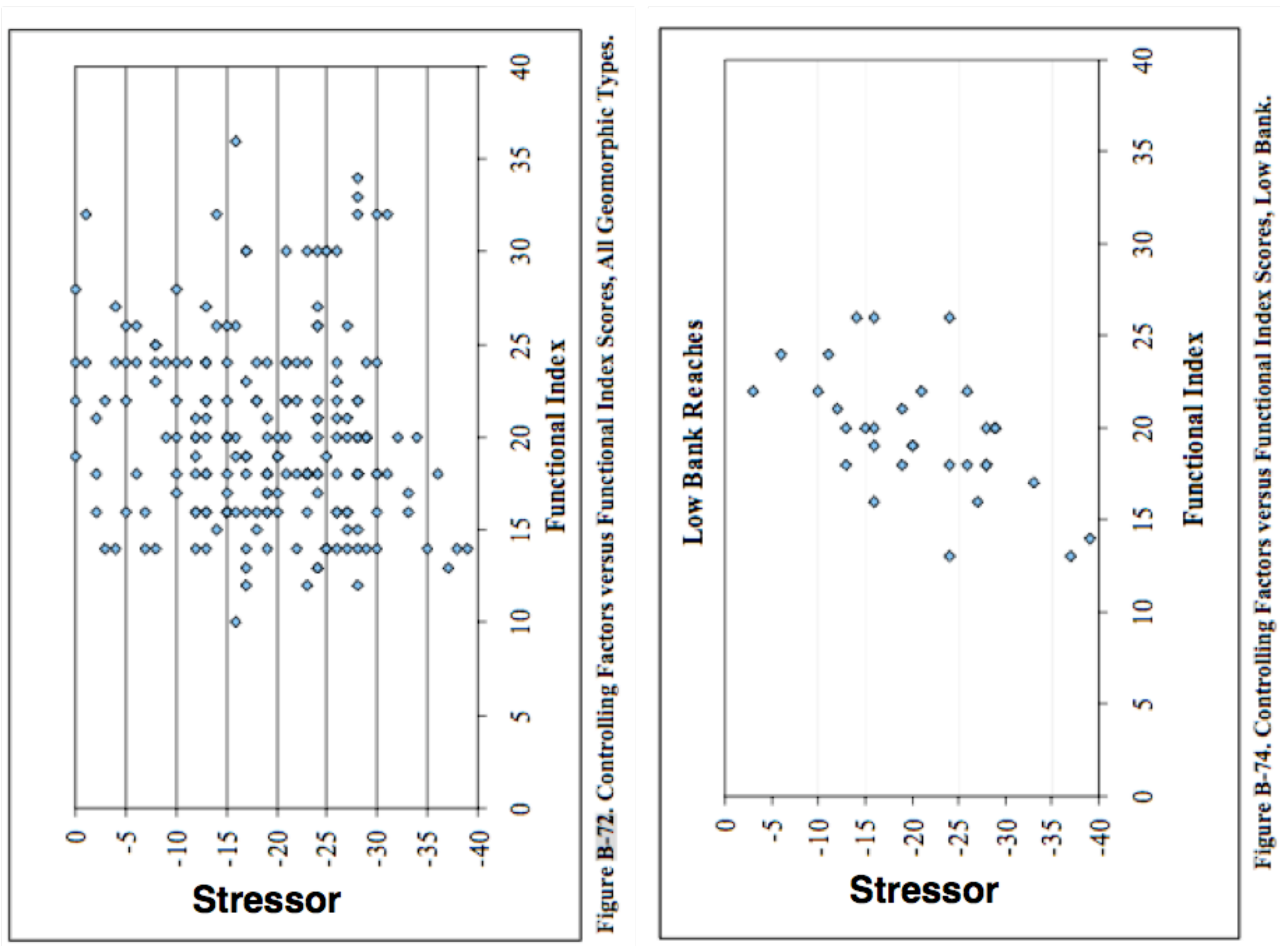


*Figure 1: Bainbridge data analyzed by Flora*

Figure 1 shows two of the Bainbridge data sets commented on by Flora; both are rotated 90 degrees CCW so they will have the same orientation as the other Figures below. Each plotted point represents a summary of all the observations collected at a particular place on the shoreline. The point's horizontal location indicates how heavily stressed that location is (worse to the right), and the vertical indicates how the ecology is functioning (worse to the bottom). The left plot is of the complete data, the right shows the low-bank subset. My uncalibrated “Mark I eyeball” sees little structure at left, but could be convinced that there is a 45 degree down-sloping trend in the low-bank data at right.

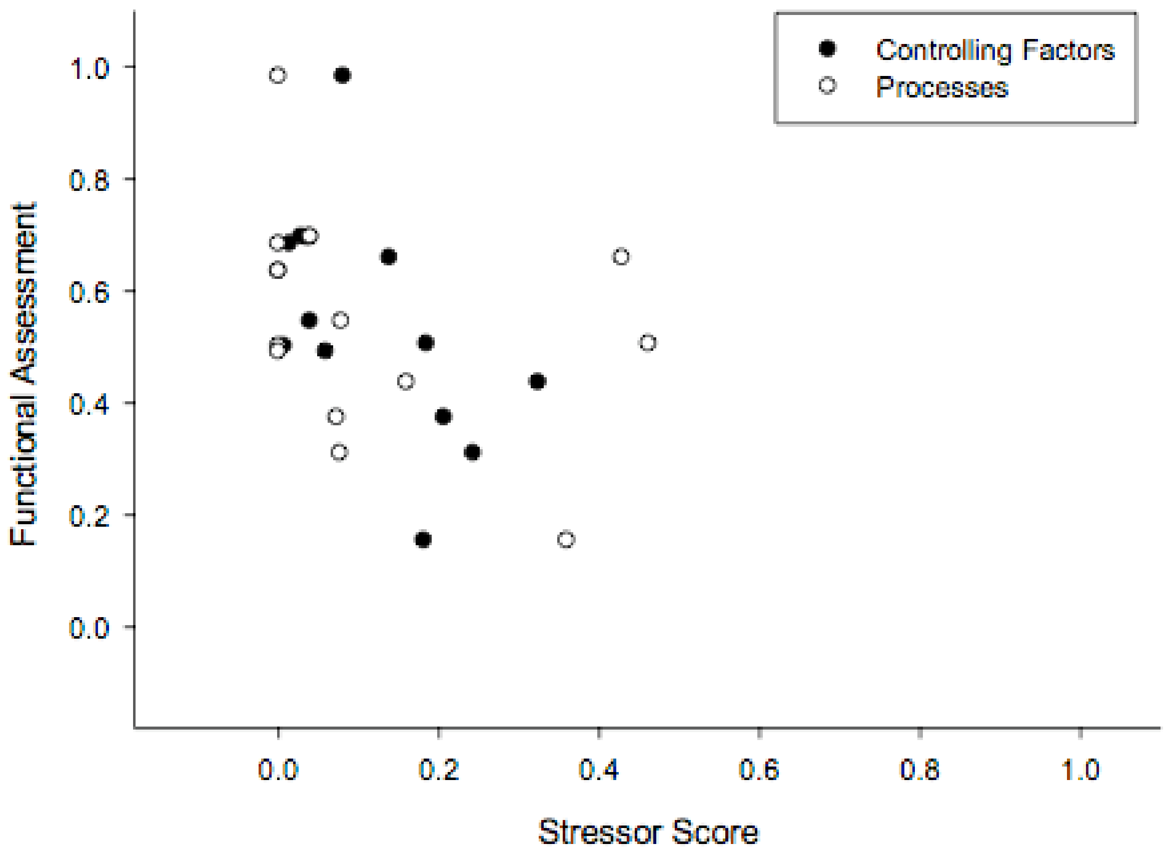


*Figure 2: East Kitsap data analyzed by Flora*

Figure 2 Shows data from East Kitsap assessment that was analyzed by Flora; here the stressors are categorized by whether they represent local conditions or physical processes affecting them. If my “Mark I eyeball” could ignore the two highest stressor points, it would draw a trend curve, concave to the right, through the remaining points.

Regression, also called curve fitting, is a mathematical technique for relating a set of observations of a system to a mathematical model of that system. (It is a way to “calibrate the eyeball.”) It was invented around 1800 to relate astronomical observations of a comet's or a planet's location on the sky on a particular date (the system) to the object's orbital elements (the model). The analysis works by using the dates and the model as input to calculate the corresponding locations of the object on the sky. The model is adjusted so that all the computed locations “best” match the observed locations.

Linear regression is the most commonly-used regression technique because it is computationally inexpensive; the updating of the model occurs in a single step. Indeed, its functions are commonly available in spreadsheet applications, such as *Excel* or *Open Office*. In Flora's analysis, the model is the straight line that "best" passes through the cloud of data points in the three plots above. Once the best line is derived, the *coefficient of determination* can be calculated to show how tightly the points cluster around the line. Flora does not state what the best line looks like for any of the plots. He gives values for the coefficient of determination for some. Not surprisingly, it is near zero for the first plot. He does not give a numeric value for the low-bank data, simply characterizing it as "extremely low". For Figure 2 he gives a value of 12%, far from the 66% threshold he asserts is needed for "significance". Note, however, that if the true trend is a curve, rather than a straight line, the coefficient's value will be unreliable.

If I presume that Flora's statistical tests are correct, and that the apparent patterns in the above plots are meaningless, then I would like to know whether an actual causal relationship would be visible when the observations are "cooked-down" into a similar plot. A common ploy is to test an analysis technique by feeding it artificial data that contains the "signal" one is looking for. If the signal is detected, then noise can be added to get an idea of how robust the analysis is likely to be when faced with real-world data.

Out of ~~laziness~~ economy of effort, I would like to test with data that passes the Einstein test: "Make it as simple as possible, but no simpler." My experience has been that the average Mark I eyeball is very capable of seeing underlying patterns in 2-dimensional and 3-dimensional data, so we'll try 4-dimensional data (3 stressors and 1 ecological function). It is important, as well, that the artificial data mimic the processes that are in operation in the real world. At first, no random noise should be added; sometimes even simple processing schemes can add a surprising amount of chaos. What follows is the construction of a simple-as-possible data set that mimics the real world

Let us model a shoreline in hypothetical Pastik County, where imaginary lee grass is being adversely affected by herbicide washed off of golf courses, shaded from light by turbid water from storm runoff, and restricted in area by beach steepening caused by bulkheads. Our lee grass is a biostatistician's dream organism; it responds to each of the stressors in exactly the same way in every location, independent of any other facet of its environment. For each of the stressors it responds similarly. At first it responds little to increasing stress until its natural defenses are overcome. Then it declines more or less linearly until it is at very low abundance. The decline is well-modeled by shifted and scaled versions of the so-called *logistic function*, a curve shaped like a backwards "S".

Pastik County's budget allows data to be taken at only 41 locations, which are chosen randomly in order to sample the entire range of stressor levels that occur in the county. Because lee grass is such a model organism, the choice of location is the only source of randomness in our model. So, at each of these 41 locations the staff have recorded a triplet of stressor levels denoted (h,t,s) (i.e. a measured value for herbicide, a value for turbidity, and a value for slope), and the level of lee grass abundance, L, denoted L(h,t,s) based on the supposition that L is related to values for h, t, and s. The observations and the model are tabulated in Appendix B - Data.

Figure 3 shows the "causal" view of our data set. The plot shows the individual contributions from herbicide L(h), turbidity L(t), and steepness L(s). The vertical axis is the fraction of lee grass remaining as the intensity of each of the stressors increases to its maximum value in the county (1.0 on the horizontal axis). The backwards "S" shape of each curve is well-sampled by our 41 random locations. From the shape of the curves we can see that about 50% of the lee grass will remain when herbicide reaches ~78%, or when turbidity reaches ~62% or when beach steepening reaches ~100% of its range.

To calculate the abundance from the individual contributions, we can apply the effects of each stressor sequentially. Since each stressor acts independently, we wind up with the product L(h)L(t)L(s). Thus, for a shoreline with 78% herbicide <u>and</u> 62% turbidity <u>and</u> 100% steepening, only 1/8th of the lee grass will remain.

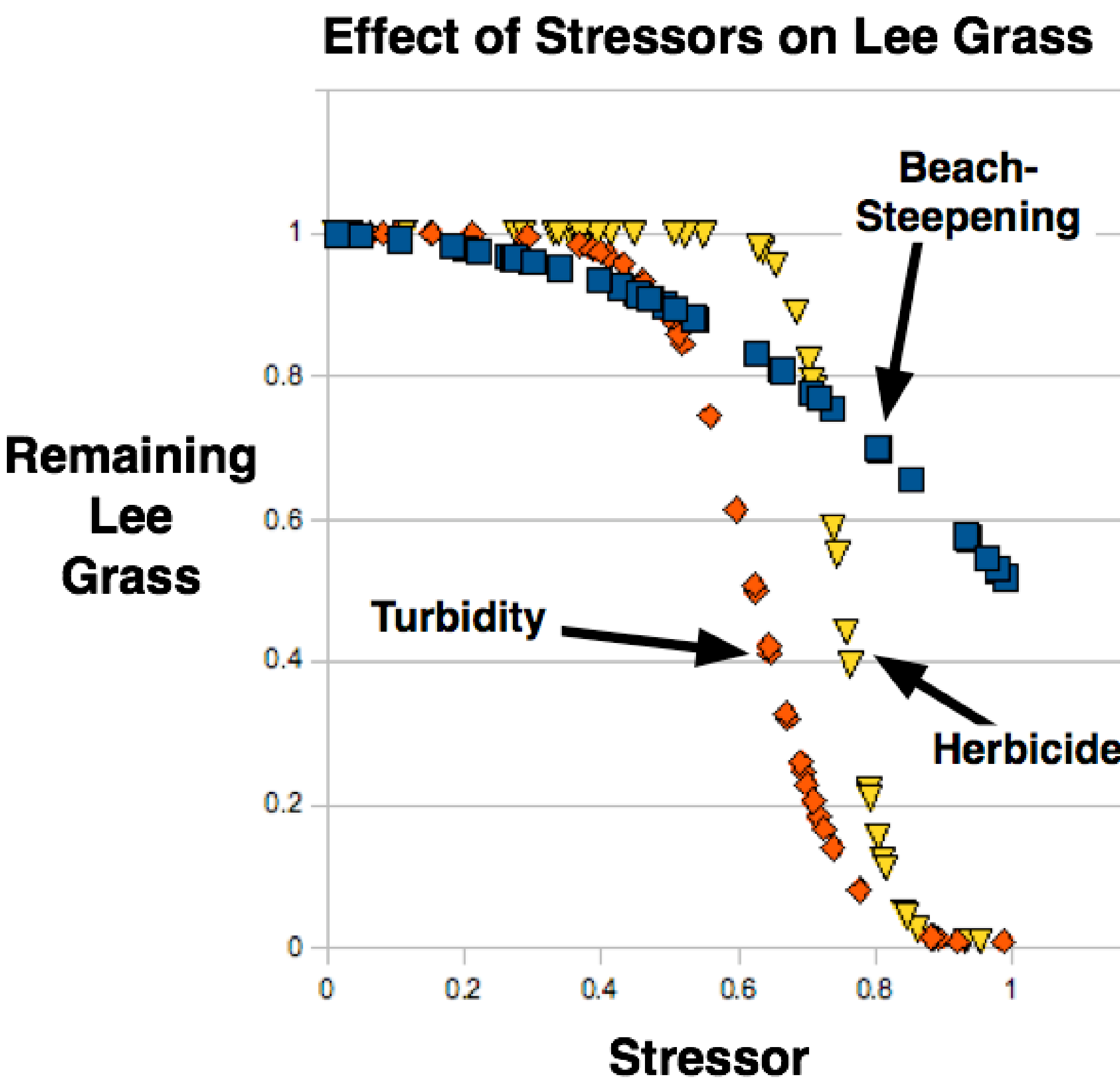


*Figure 3: Effect of three stressors on lee grass*

Figure 4 shows how the artificial data set looks when plotted following Figures 1&2; the horizontal axis is the sum of the stressors (h+t+s) and the vertical axis is the lee grass abundance L(h,t,s). (Note that the plot symbols mark subsets, not stressors in this plot.) It is hard to believe that this chaotic-appearing data is causally derived from the clearly-defined noise-free curves in Figure 3. Part of the "noise" is due to the non-linearity introduced by the multiplication of the individual stressor components, and part is due to compressing what is really 4-dimensional data into a 2-dimensional plot.

The sidebars in Figure 4 show the coefficient of determination computed for the full 41 location data set, and subsets of 21 and 11 locations. It ranges from 5% to 21% , far below the threshold given by Flora, even though derived from a noise-free causal data set. This illustrates how applying an inappropriate statistical test can lead one astray.

## Conclusion

Applying linear regression analysis to the summary data of the Bainbridge and East Kitsap nearshore assessments cannot reliably detect the presence or absence of a causal link between stressors and ecological function.

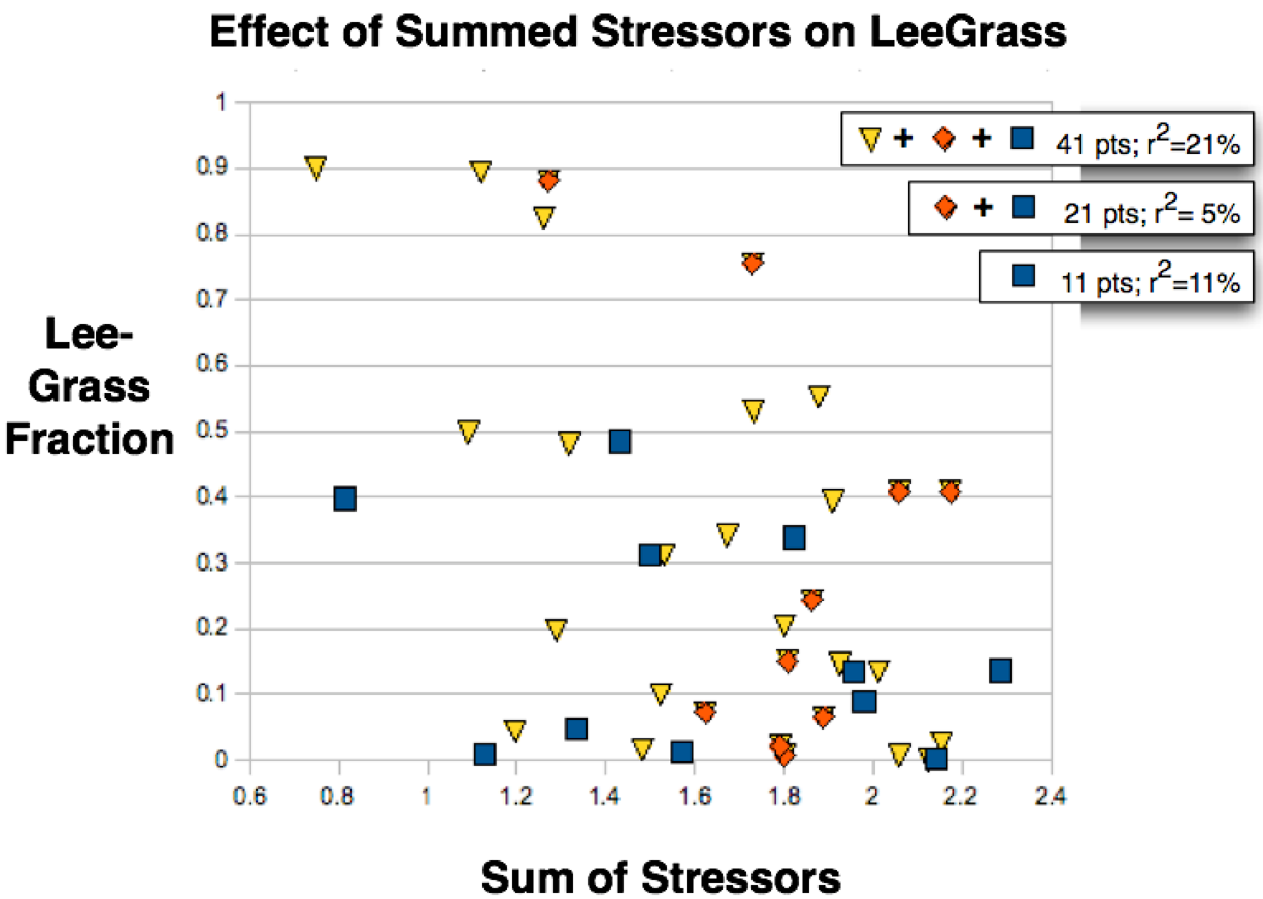


*Figure 4: Pastik County displayed similarly to Kitsap County data*

## *Appendix A - Pitfalls*

From *Regression analysis* in Wikipedia, the free encyclopedia

> "However when carrying out inference using regression models, especially involving small effects or questions of causality based on observational data, regression methods must be used cautiously as they can easily give misleading results."

From *Regression Diagnostics* in Wikipedia

> "Commonly used checks of goodness of fit include the R-squared, analyses of the pattern of residuals and hypothesis testing."
>
> "Interpretations of these diagnostic tests rest heavily on the model assumptions. Although examination of the residuals can be used to invalidate a model, the results of a t-test or F-test are sometimes more difficult to interpret if the model's assumptions are violated."

From *Coefficient of Determination* in WikiDoc

> "Notes on interpreting $R^2$
>
> $R^2$ does NOT tell whether:
>
> - the independent variables are a true cause of the changes in the dependent variable
> - omitted-variable bias exists
> - the correct regression was used
> - the most appropriate set of independent variables has been chosen
> - there is co-linearity present in the data
> - the model might be improved by using transformed versions of the existing set of independent variables"

## *Appendix B - Data*

| Observations | | | | Theory | | |
|---|---|---|---|---|---|---|
| Herbicide | Turbidity | Steepening | L(h,t,s) | L(h) | L(t) | L(s) |
| 0.763 | 0.037 | 0.013 | 0.397 | 0.398 | 1.000 | 0.999 |
| 0.113 | 0.647 | 0.737 | 0.312 | 1.000 | 0.413 | 0.755 |
| 0.035 | 0.987 | 0.105 | 0.010 | 1.000 | 0.010 | 0.991 |
| 0.412 | 0.894 | 0.263 | 0.013 | 1.000 | 0.013 | 0.967 |
| 0.803 | 0.883 | 0.456 | 0.002 | 0.156 | 0.016 | 0.915 |
| 0.811 | 0.460 | 0.708 | 0.090 | 0.124 | 0.933 | 0.776 |
| 0.744 | 0.150 | 0.540 | 0.484 | 0.551 | 1.000 | 0.879 |
| 0.654 | 0.694 | 0.936 | 0.136 | 0.956 | 0.248 | 0.574 |
| 0.842 | 0.292 | 0.201 | 0.048 | 0.049 | 0.995 | 0.978 |
| 0.702 | 0.718 | 0.535 | 0.134 | 0.823 | 0.185 | 0.881 |
| 0.739 | 0.153 | 0.933 | 0.339 | 0.588 | 1.000 | 0.577 |
| 0.635 | 0.725 | 0.449 | 0.150 | 0.976 | 0.168 | 0.917 |
| 0.684 | 0.498 | 0.992 | 0.407 | 0.891 | 0.883 | 0.518 |
| 0.790 | 0.672 | 0.428 | 0.065 | 0.219 | 0.321 | 0.925 |
| 0.710 | 0.368 | 0.980 | 0.408 | 0.783 | 0.984 | 0.530 |
| 0.932 | 0.062 | 0.806 | 0.007 | 0.010 | 1.000 | 0.697 |
| 0.861 | 0.211 | 0.719 | 0.021 | 0.027 | 0.999 | 0.769 |
| 0.355 | 0.777 | 0.492 | 0.073 | 1.000 | 0.081 | 0.900 |
| 0.634 | 0.432 | 0.663 | 0.755 | 0.976 | 0.957 | 0.809 |
| 0.757 | 0.596 | 0.509 | 0.243 | 0.444 | 0.613 | 0.893 |
| 0.390 | 0.410 | 0.472 | 0.880 | 1.000 | 0.969 | 0.908 |
| 0.330 | 0.690 | 0.991 | 0.135 | 1.000 | 0.260 | 0.518 |
| 0.371 | 0.558 | 0.979 | 0.395 | 1.000 | 0.744 | 0.530 |
| 0.846 | 0.081 | 0.270 | 0.043 | 0.045 | 1.000 | 0.966 |
| 0.549 | 0.883 | 0.049 | 0.016 | 0.998 | 0.016 | 0.996 |
| 0.789 | 0.738 | 0.626 | 0.026 | 0.221 | 0.141 | 0.832 |
| 0.276 | 0.387 | 0.455 | 0.895 | 1.000 | 0.978 | 0.915 |
| 0.629 | 0.701 | 0.471 | 0.204 | 0.980 | 0.229 | 0.909 |
| 0.394 | 0.625 | 0.298 | 0.481 | 1.000 | 0.501 | 0.960 |
| 0.363 | 0.644 | 0.664 | 0.342 | 1.000 | 0.423 | 0.808 |
| 0.369 | 0.398 | 0.965 | 0.531 | 1.000 | 0.974 | 0.545 |
| 0.337 | 0.920 | 0.802 | 0.007 | 1.000 | 0.010 | 0.701 |
| 0.522 | 0.670 | 0.338 | 0.312 | 0.999 | 0.328 | 0.951 |
| 0.708 | 0.709 | 0.507 | 0.146 | 0.795 | 0.206 | 0.894 |
| 0.010 | 0.462 | 0.275 | 0.899 | 1.000 | 0.932 | 0.965 |
| 0.791 | 0.101 | 0.395 | 0.197 | 0.211 | 1.000 | 0.935 |
| 0.814 | 0.487 | 0.221 | 0.099 | 0.113 | 0.901 | 0.975 |
| 0.287 | 0.623 | 0.181 | 0.499 | 1.000 | 0.509 | 0.981 |
| 0.507 | 0.519 | 0.853 | 0.554 | 1.000 | 0.846 | 0.655 |
| 0.953 | 0.777 | 0.396 | 0.001 | 0.010 | 0.080 | 0.935 |
| 0.447 | 0.512 | 0.301 | 0.824 | 1.000 | 0.858 | 0.960 |

## Revision History

1. 11/20/09 – First public release.
2. 12/03/09 – Minor edits for clarity suggested by reviewers.

## *References*

1 - Williams, G.D., Evans NR, and RM Thom 2004. *Bainbridge Island Nearshore Habitat Assessment Management Strategy Prioritization, and Monitoring Recommendations*. PNWD-3391. Prepared for the City of Bainbridge Island, Bainbridge Island, Washington, by Battelle Marine Sciences Laboratory, Sequim, Washington.

2 - Borde, A.B., C. Judd, N.K. Sather, and R.M. Thom 2009. *East Kitsap County Nearshore Habitat Assessment and Restoration Prioritization Framework*. PNWD-4053. Prepared for the Kitsap County, Washington, by Battelle Marine Sciences Laboratory, Sequim, Washington.