

A Review of R.B. Nerf's Critique of Don Flora

Carl Shipley Ph.D., July 2010

Summary

Don Flora, a retired Ph.D. living on Bainbridge Island, has written an analysis of data from two large Battelle Institute studies of Kitsap shoreline, pointing out that these studies failed to document clear effects of proposed environmental stressors such as bulkheads on measures of nearshore environmental health.¹ Flora's analysis has caused some controversy because it challenges basic assumptions used in the development of shoreline management policies. In a paper distributed to the media, R.B. Nerf criticized Flora's work on mathematical grounds, arguing that Flora's use of linear statistics was not adequate for analysis of complex data which might have had nonlinear characteristics.² The purpose of the following paper is to examine Nerf's analysis to determine if it constitutes a reasonable criticism of Flora's work.

Nerf apparently believes that because the Battelle studies made frequent use of composite scores made by summing or averaging individual measures, negative effects of proposed stressors such as bulkheads would be obscured or completely eliminated in the Battelle data. To support his argument, Nerf presents a hypothetical situation in which strong trends in individual measurements are obscured in composite scores formed by combining these measurements. To a casual reader, Nerf's analysis suggests that direct measurements in the Battelle studies may have shown strong harmful effects of features such as bulkheads and that, due to a mathematical complexity Flora ignored, these effects disappeared in the data Flora analyzed.

However, this argument is not valid because, where the data are available in the Battelle studies, direct measurements of the relationship between proposed stressors and indicators of environmental health do not reveal any clear negative effects of the stressors. In addition, examination of Nerf's theoretical model reveals that it constructs hypothetical composite scores in a specific way that we would not expect to see in real world data. Combining the data in a more logical way eliminates Nerf's effect. Nerf's general assertion that linear measures are somehow inappropriate for analysis of the type of data reported in the Battelle studies ignores the fact that linear measures are routinely used to good effect in describing this kind of data. In fact, as will be shown below, linear correlation is even quite powerful when used to describe Nerf's own model of nonlinear data.

NOTES

¹ Don Flora, Ph.D., "Evidence of Near-Zero Habitat Harm From Nearshore Development"

² R.B. Nerf, "Critique of D.F. Flora's 'Evidence...'"

The Battelle Studies

Before examining Flora's and Nerf's arguments it is necessary to briefly review the Battelle studies Flora examined. These studies divided the entire perimeter of Bainbridge Island (the first Battelle study) and the eastern shoreline of Kitsap County (the second Battelle study) into sectors in which investigators measured the presence of several possible environmental stressors. These stressors, generally called "controlling factors" in the Battelle model, were things associated with human development such as the presence of bulkheads or docks. Measures tabulating the presence of individual stressors were reported and also summed or averaged to form a composite stressor score for each study area. In the first Battelle study, but not the second, extensive data were also provided for several measures of environmental health, such as the existence of eelgrass or fish spawning areas; these data were also combined to form a composite environmental health score for each sector.

In the second Battelle study, the authors essentially stopped reporting useful direct measurements on environmental health. Extensive data were presented for features thought to be stressors, but measures of environmental health were presented for only 14 of 518 study areas. A weak association between stressors and health was reported for this sample, but Flora pointed out that the very small sample size and the fact that the 14 areas were not chosen randomly makes the scientific value of this result questionable.

Flora's Analysis

Based on examination of both direct measurements where they were available and composite scores formed by combining these measurements, Flora argued that the Battelle studies failed to find clear relationships between proposed stressors and environmental health. Flora's point should not be in dispute because the Bainbridge Battelle study itself, which is the only one that presented extensive data on both stressors and health, clearly says that examination of the data only "suggest a slight correlation between impaired controlling factors and reduced ecological function" (Bainbridge Island Nearshore Assessment, p. 98).

Even this "slight suggestion" of an association is questionable. The Bainbridge Battelle study presents maps with detailed information on the location of both proposed stressors, such as bulkheads, and measures of environmental functioning, such as eelgrass beds and fish spawning areas. Based on direct measurements of these maps, Flora found that about half of Bainbridge beaches with eelgrass also have a bulkhead and about half do not. This is the definition of lack of a clear relationship. Flora also provides statistical analysis showing that bulkheads had no clear negative effect on the existence of nearshore spawning areas for sand lance or surf smelt. These statistics are useful to a scientist, but simple percentages are probably more meaningful for the average reader. Flora found that on Bainbridge, 72 percent of surf smelt spawning is in front of bulkheads. He also found that 49 percent of sand lance spawning beach is in front of bulkheads.³

³ Percentages for sand lance and surf smelt spawning are from Don Flora, personal communication.

Nerf's Critique

Nerf's critique focuses exclusively on Flora's analysis of composite scores, which are the most common kind of statistic in the Battelle reports. Nerf's criticism is based on the mathematical fact that clear relationships in individual measurements can sometimes be obscured when these data are combined to form composite scores. To demonstrate this, Nerf first constructs three hypothetical functions, each of which presents a strong, clear relationship between a hypothetical environmental stressor caused by some form of human activity and a hypothetical measure of environmental health.

Figure 1, below, presents Nerf's hypothetical functions (the data are identical to those in Nerf's Figure 3 found on page 5 of his paper). The three stressors in Nerf's hypothetical model are "Herbicide," washed into the water after use by a property owner, "Turbidity," representing clouding of water by storm runoff from a drain of some sort, and beach "Steepening," caused by the presence of a bulkhead. In each case, the level of Nerf's hypothetical stressor increases as we go to the right on the X axis while the fraction of Nerf's hypothetical measure of environmental health, "lee grass," an imaginary surrogate for eelgrass, increases from 0 to 1 as we go up the Y axis. Thus, higher levels of stressor are clearly associated with lower levels of lee grass, a very strong negative correlation.

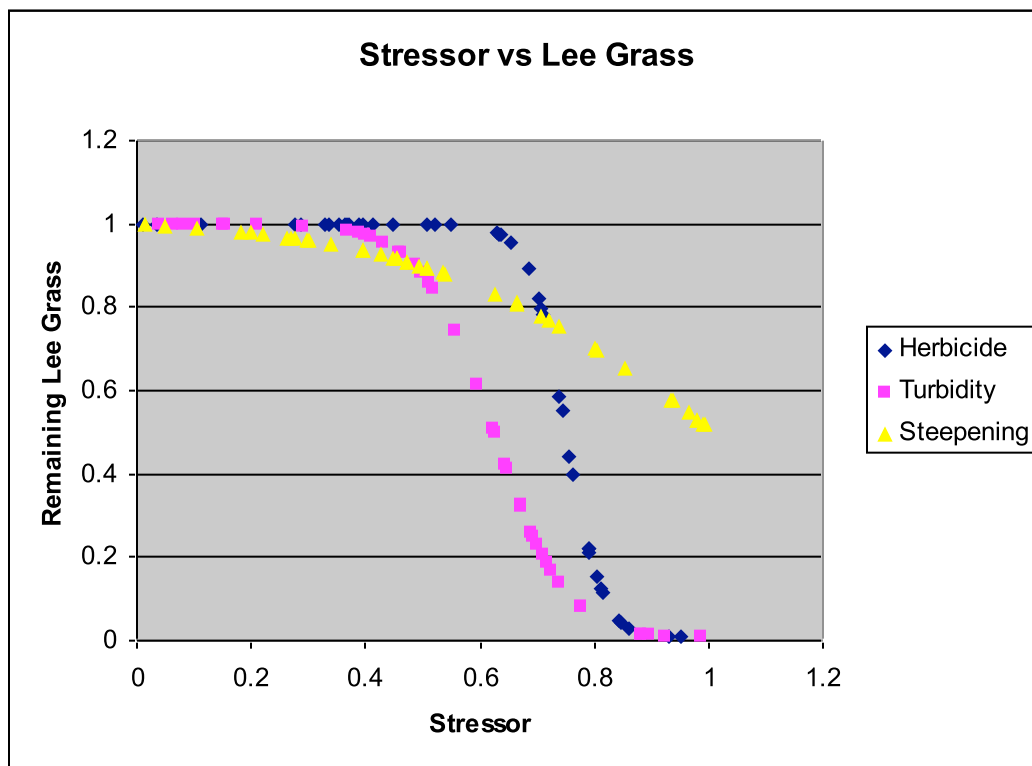


Figure 1 – Nerf's three hypothetical functions

After presenting these functions, Nerf develops a hypothetical example in which composite measures are used to evaluate the effect of stressors on environmental health. In this example, measures of each of Nerf's three stressors are taken at 41 sites and these values are added to form a composite stressor score. The expected level of lee grass for each site is obtained by multiplying together the predicted lee grass fractions for each stressor in the triplet. Figure 2 presents a replica of Nerf's plot of these composite stressors and predicted lee grass fractions (data for the graph are presented in Table 1 of the Appendix below). As can be seen, the level of correlation between stressors and lee grass is greatly reduced. Nerf asserts that this example shows that when nonlinear data which have a clear trend are combined to form composite measures, the original trend may disappear. He suggests that this kind of effect was present in the Battelle composite measures Flora analyzed. In other words, Nerf argues that there may have been strong effects of stressors on environmental health in the Battelle research, but these effects were obliterated by the use of composite scores. Nerf says Flora ignored this issue and, thus, his work is invalid.

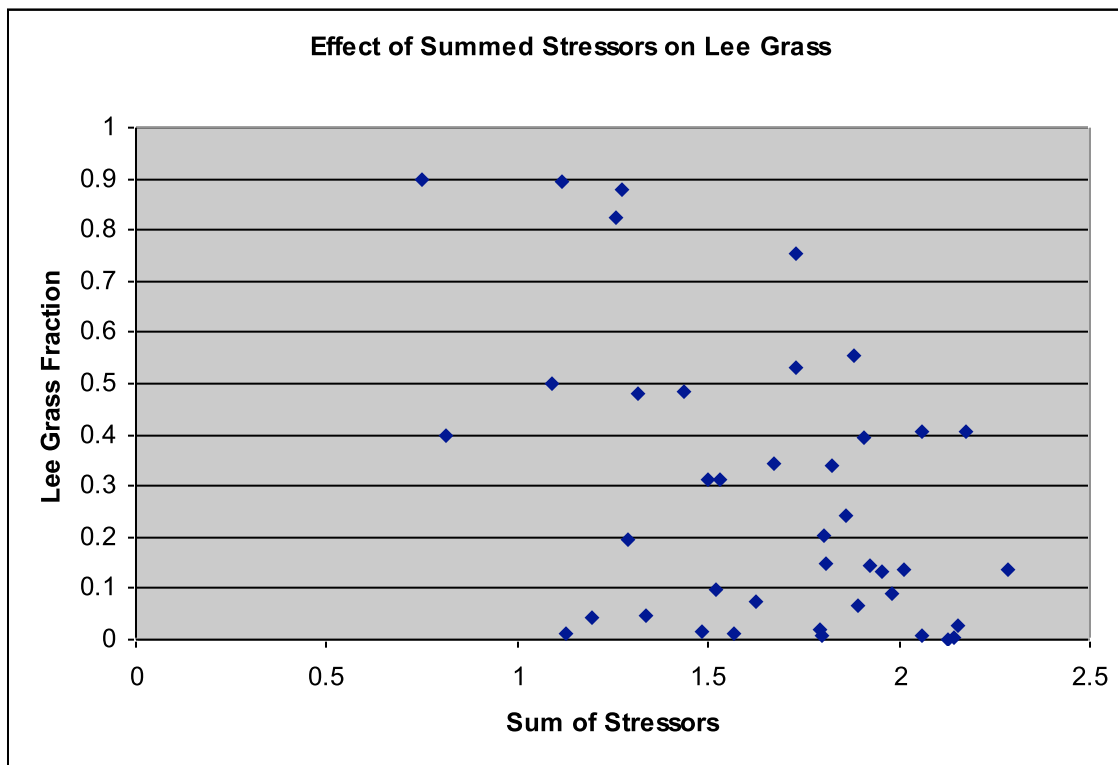


Figure 2 – Nerf's composite stressor scores plotted against the fraction of lee grass remaining

This is not a reasonable criticism of Flora's work. In the first place, the force of Nerf's criticism depends on the idea that direct measures, as opposed to composite scores formed by combining these measures, show strong negative effects of stressors on health. This is not the case in the Battelle studies. Nerf ignores the fact that Flora examined direct measures where they were available. Data from the Battelle study of Bainbridge Island, which is the only study that reported the detailed information one would need to construct real world functions of the kind

hypothesized by Nerf, as illustrated in Figure 1, do not contain clear correlations between stressors and health. In Nerf's hypothetical world, beach steepening caused by the presence of bulkheads, has a dramatic negative effect on lee grass, Nerf's surrogate for eelgrass. However, in the real world, when Flora examined the location of bulkheads and eelgrass on maps of Bainbridge Island, he found that bulkheads had no effect on the likelihood that a beach would have eelgrass.

Secondly, even given the imaginary world reflected in Figure 1, Nerf's suggestion that trends seen in direct measurements will necessarily be obscured in composite scores is untrue. In fact, whether trends of the kind seen in Figure 1 are obscured in composite data depends solely on how the composites are formed. Nerf formed his composite scores in a way that tends to obscure the effects seen in the individual functions. In order to understand this, it is instructive to examine some of Nerf's three-value data sets. In many of these triplets, a high observation for a stressor theoretically caused by human activity is paired with a low observation for other stressors theoretically caused by the same thing. For example, the first set of three stressors (see Table 1 in the Appendix) contains a high value for herbicide (.763) and two low values for turbidity (.037) and steepening (.013). The second set contains a low score for herbicide (.113) and high values for turbidity (.647) and steepening (.737). The third set contains a high value for turbidity (.987) and low values for herbicide (.035) and steepening (.105). These kinds of pairings are not present in all data sets, but they exist in enough of them to cause the relationship between stressor and health that is so obvious in Figure 1 to be greatly reduced in Figure 2.

To appreciate the mathematical effect of organizing the data sets such that they contain both high and low stressor values, remember that the predicted level of lee grass for each triplet is obtained by multiplying each of the three lee grass fractions together. In Nerf's imaginary world, very high values for herbicide or turbidity essentially wipe out lee grass – they introduce a near zero term into the multiplication used to compute the predicted fraction of lee grass at each site. Thus, a high value for herbicide or turbidity tends to remove any influence of other stressors in the triplet. It is almost like multiplying any effect of the other members of the triplet by zero. This reduces the level of correlation between stressors and health.

Put another way, what Nerf has done in forming his data sets is to move from an imaginary world in which there is a great deal of lee grass – more than half the predicted fractions for lee grass in Figure 1 are greater than 0.8 – to one in which there is relatively little lee grass – less than one tenth of the predicted lee grass fractions in Figure 3 are greater than 0.8. Mathematically, truncating the distribution of lee grass in this way tends to remove correlation between stressors and health. To help understand this, suppose that we added a few more stressors to the model. Suppose we formed composite scores using 6 imaginary deadly lee grass stressors and that each composite had a very high value for at least one of these killers - and thus a very low number associated with this stressor, something near zero, for the predicted amount of lee grass remaining. The composite predicted level of lee grass would be formed by multiplying together 6 numbers, one of which would always be near zero. We would then have a hypothetical world with essentially no lee grass. In this case, there would be almost no

correlation between composite stressor values and predicted lee grass because, regardless of the values of 5 of the 6 terms in each composite score, the level of lee grass would be the same.

The fact that herbicide and turbidity are such deadly lee grass killers in Nerf's model is contrary to what we expect in the real world. The Battelle studies do not provide direct measures of herbicide and turbidity but the Bainbridge Island study provides information on the location of both bulkheads and eelgrass. We would expect turbidity from storm drains and herbicide applied by humans to be present in highly developed areas, areas that would also have bulkheads. In fact, we might expect herbicide to come from lawns behind bulkheads and turbidity to be caused by runoff from drains in or near bulkheads. We would, then, expect that if herbicide and turbidity were powerful killers of eelgrass, there would be little eelgrass in front of bulkheads. However, we know that about half the eelgrass on Bainbridge Island is in front of a bulkhead.

Nerf attributes the decline in correlation he sees in composite data partially to nonlinearity and partially to what he describes as "collapsing what is really four dimensional data into a two dimensional plot," explanations that are unlikely to be meaningful to the reader and ones that suggest the decline is mathematically inevitable. However, the decline is not a mathematical inevitability; it is simply a result of the fact that Nerf has sprinkled high values for imaginarily powerful stressors into many of his data triplets. Forming data sets in this way does not conform to what we would expect in the real world. Triplets with both high and low stressor values are illogical because the three-value sets are supposed to represent measures of things related to human activity taken at a single site. We would expect a site with a high level of human development to have high levels of stressors related to human development - and a site with no human development to have uniformly low levels of stressors associated with development. This is the logic inherent in the Battelle studies. However, this logic has been destroyed in Nerf's example.

If Nerf's composite data is constructed in an artificial way designed to eliminate correlation we would expect that if we construct his three-value sets in a more logical fashion the decrease in correlation would disappear. This is, in fact, the case. Figure 3 presents a graph of composite scores created by combining Nerf's hypothetical data in a way such that high values of one stressor are observed with high values of other stressors - presumably these are observations taken at sites with high levels of human development - and low values are found with other low values - presumably observations taken at sites with little or no human activity (data for this graph are presented in Table 2 of the Appendix; they were obtained by simply sorting the values in Table 1).

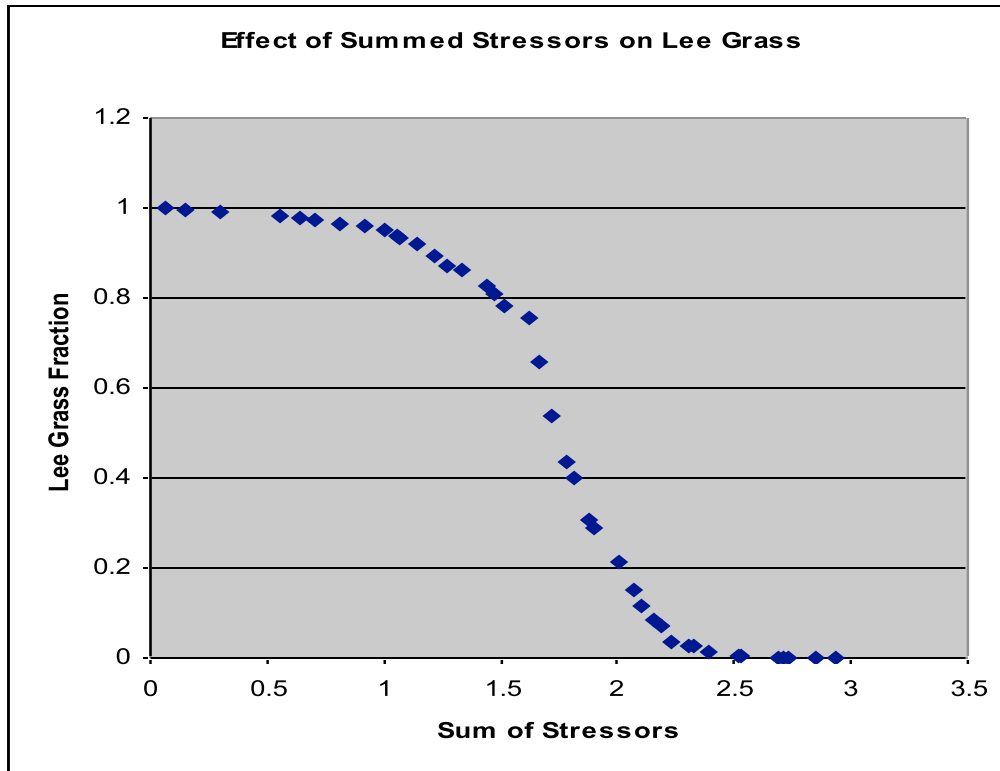


Figure 3 – Composite scores formed from correlated observations.

It's important to realize that ordering composite scores in this way doesn't change the original stressor and lee grass pairs. If we re-plot Figure 1 from the individual stressor and lee grass pairs in Table 2, we will get Figure 1 again, because we will be plotting the same pairs. All we have changed is that the values in the composite scores are now combined in a more logical way. If we form the data triplets in this fashion, we see the same relationship in the plot of the composite data as we see in the original functions. The correlation coefficient between the stressor composite and the predicted presence of "lee grass" in Figure 3 is -0.93. Even though the stressor function is somewhat nonlinear in form, the correlation coefficient, a linear measure, can be considered to account for the overwhelming majority of variation in the data.

I am not arguing that Nerf's example is mathematically incorrect, simply that it is misleading and not to the point. Given the real world of the kind of data generated by the Battelle studies, we would not expect composite measures to completely obscure or eliminate strong effects present in direct observations. The Battelle studies make very extensive use of composite measures, indeed, the Battelle study of east Kitsap shoreline focuses almost entirely on this kind of measure. If Nerf's argument is that Flora should not have been analyzing data formed by combining individual scores, then Nerf should be criticizing the Battelle authors themselves because they make very extensive use of such data. The logic of the Battelle research is that human influences at a given site will tend to be correlated. This logic is completely valid and is, in fact, widely present in ecological research. In this situation, using composite measures is not only mathematically sound, it is a powerful way to examine complex issues that researchers want

to study. Use of composite scores allows scientists to ask this very important question – do small effects of environmental stressors sum to cause large effects on environmental health?

Nerf's general argument rests on the suggestion that linear measures are not useful in understanding data that may have some nonlinear aspects. This is clearly untrue. In fact, linear correlation is a very common statistic used to determine if an association exists in real world sets of paired data, even if these data may have some nonlinear trends.⁴ The value of using linear correlation to understand even nonlinear data can be seen in Nerf's own model. The correlation coefficients for Nerf's three stressor functions, shown in Figure 1, range from -0.8 to -0.95, all of which are significantly different from zero at the .0001 level. Even though Nerf's three functions are designed to be nonlinear, linear correlation is still quite powerful in describing them because each function contains a very strong linear component.

In considering the value of linear correlation when it is applied to data that may have some nonlinear characteristics, it is also interesting to note that Nerf's composite scores (shown in Figure 2) still contain a statistically significant linear correlation reflecting the effect of stressors on environmental health. Nerf states that the coefficient of determination for his combined scores is .21, implying a correlation with an absolute value of about 0.45 (the coefficient of determination is formed by squaring a correlation coefficient). Nerf says this value does not rise to the level Flora required for statistical significance, ignoring the fact that Flora was discussing data with a different sample size. A correlation with an absolute value of .45 is significantly different from zero at the .01 level for Nerf's sample of 41 points.⁵ This result reflects the fact that the correlations in Nerf's original three functions are so strong their effects can still be detected, albeit at a greatly reduced level, even in data formed in a way designed to obscure these effects.

One further point relating to Nerf's discussion of Flora deserves attention. Nerf asserts that the Battelle studies were not interested in identifying relationships between human stressors and environmental functions. It is a somewhat frightening thought to believe scientists argue they can simply assume proposed stressors have harmful effects even when these effects are not clearly present in

⁴ Since the correlation coefficient is an important statistic for this discussion, it is worthwhile to provide a little detail about the measure. Correlations vary between +1.0 and -1.0. An example of a data set for which a correlation coefficient might be applied could be height and weight for a set of individuals. In this case, we would expect to see a strong positive correlation – perhaps something on the order of 0.5 – because taller people, on average, tend to be heavier. The correlation would not be perfect (i.e., 1.0), because there are some relatively shorter but heavier people and some relatively taller but thinner people. However, the correlation would be expected to be positive and, in a sample of reasonable size, statistically significant. A negative correlation would be seen in a case where a high value of one variable might be associated with a low value of another variable - thus, the three theoretical functions Nerf presents, shown in Figure 1 of this paper, have very large negative correlations. A small correlation is found in situations where there is no clear relationship between two variables. For example, this would be the case if half the time when the first variable is large the second is small and half the time when the first is large the second is also large.

⁵ Statistical tables or calculators are widely available on the web. One with a good user interface can be found at: <http://faculty.vassar.edu/lowry/tabs.html>

their own data. However, while it is true that some language in each of the Battelle studies indicates that the authors assume stressors have negative effects on nearshore environmental health, the Battelle authors are also clearly interested in trying to examine these effects – an interest revealed by the design of the studies, by explicit language in the studies, and by much of the treatment of the data in the studies. For example, section 1.1 of the Bainbridge study (page 1) contains this very explicit statement of what the authors call one of the “needs of the study” – to examine the question, “What effects do typical shoreline modifications have on nearshore habitat (especially salmonid habitat)?” Also, both Battelle studies explicitly discuss correlations, or even suggestions of correlations, between stressors and environmental health in cases where the authors feel an association may be present in the data.

Conclusion

I think it is worth repeating that a casual reader of Nerf’s paper may assume that, like the theoretical functions Nerf presents, direct measures in the Battelle studies show strong, clear negative relationships between stressors associated with human activity and measures of nearshore environmental health, relationships that are somehow obscured in composite data. This is not true. Flora’s point is that the Battelle studies found surprisingly little relationship between proposed environmental stressors and measured environmental health – in both direct measurements and in composite data formed from these measurements. This fact clearly suggests that current models of how human activity affects the nearshore environment are far from complete. That is an important observation, one that is not well publicized, and one that is worth emphasizing as we try to develop effective shoreline management policies.

Appendix

Table 1 – Data Nerf used to illustrate the reduction in association seen in composite measures. This table is a copy of Nerf’s Table 1 in his Appendix B.

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	0.922	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

Table 2
Data for Figure 3

Herbicide	Turbidity	Steepening	L(h)	L(t)	L(s)	X axis H+T+S	Y axis L(h, s, t)
0.953	0.987	0.992	0.010	0.010	0.518	2.932	0.000
0.932	0.922	0.991	0.010	0.010	0.518	2.845	0.000
0.861	0.894	0.980	0.027	0.013	0.530	2.735	0.000
0.846	0.883	0.979	0.045	0.016	0.530	2.708	0.000
0.842	0.883	0.965	0.049	0.016	0.545	2.690	0.000
0.814	0.777	0.936	0.113	0.080	0.574	2.527	0.005
0.811	0.777	0.933	0.124	0.081	0.577	2.521	0.006
0.803	0.738	0.853	0.156	0.141	0.655	2.394	0.014
0.791	0.725	0.806	0.211	0.168	0.697	2.322	0.025
0.790	0.718	0.802	0.219	0.185	0.701	2.310	0.028
0.789	0.709	0.737	0.221	0.206	0.755	2.235	0.034
0.763	0.701	0.719	0.398	0.229	0.769	2.183	0.070
0.757	0.694	0.708	0.444	0.248	0.776	2.159	0.085
0.744	0.690	0.664	0.551	0.260	0.808	2.098	0.116
0.739	0.672	0.663	0.588	0.321	0.809	2.074	0.153
0.710	0.670	0.626	0.783	0.328	0.832	2.006	0.214
0.708	0.647	0.540	0.795	0.413	0.879	1.895	0.289
0.702	0.644	0.535	0.823	0.423	0.881	1.881	0.307
0.684	0.625	0.509	0.891	0.501	0.893	1.818	0.399
0.654	0.623	0.507	0.956	0.509	0.894	1.784	0.435
0.635	0.596	0.492	0.976	0.613	0.900	1.723	0.538
0.634	0.558	0.472	0.976	0.744	0.908	1.664	0.659
0.629	0.519	0.471	0.980	0.846	0.909	1.619	0.754
0.549	0.512	0.456	0.998	0.858	0.915	1.517	0.784
0.522	0.498	0.455	0.999	0.883	0.915	1.475	0.807
0.507	0.487	0.449	1.000	0.901	0.917	1.443	0.826
0.447	0.462	0.428	1.000	0.932	0.925	1.337	0.862
0.412	0.460	0.396	1.000	0.933	0.935	1.268	0.872
0.394	0.432	0.395	1.000	0.957	0.935	1.221	0.895
0.390	0.410	0.338	1.000	0.969	0.951	1.138	0.922
0.371	0.398	0.301	1.000	0.974	0.960	1.070	0.935
0.369	0.387	0.298	1.000	0.978	0.960	1.054	0.939
0.363	0.368	0.275	1.000	0.984	0.965	1.006	0.950
0.355	0.292	0.270	1.000	0.995	0.966	0.917	0.961
0.337	0.211	0.263	1.000	0.999	0.967	0.811	0.966
0.330	0.153	0.221	1.000	1.000	0.975	0.704	0.975
0.287	0.150	0.201	1.000	1.000	0.978	0.638	0.978
0.276	0.101	0.181	1.000	1.000	0.981	0.558	0.981
0.113	0.081	0.105	1.000	1.000	0.991	0.299	0.991
0.035	0.062	0.049	1.000	1.000	0.996	0.146	0.996
0.010	0.037	0.013	1.000	1.000	0.999	0.060	0.999