A Note on Confidence Limits in Raimondi 1 April 09 RWQCB Report

Prepared by Scott A. Jenkins, Ph.D.

April 8, 2009

Introduction: On page 4 of the 1 April 09 report entitled “Review of Impingement study and mitigation assessment – Carlsbad Seawater Desalinization Project”, Dr. Pete Raimondi uses a monotonic function on page 3 (referred to herein as a *flow proportional impingement function*) relating fish impingement with CDP consumption of lagoon water, in order to map a second relationship on page 5 giving mitigation acres required as a function of CDP consumption of lagoon water. He concludes, “If 100 percent of water is needed (304 mgd) then the acreage need to mitigate impingement ranges from ~11-21 (if all the acres are intertidal mudflats or subtidal) to 28-54 (if the wetland is a mixture of habitats – 40% of which are intertidal wetlands or subtidal).” This conclusion appears to be quantitatively flawed because there are several problems with the specification of the flow proportional impingement function and its associated confidence limits on page 3 that ultimately inflate his mitigation numbers. The flow proportional impingement function on page 3 appears below and will be referred to as Exhibit 1.

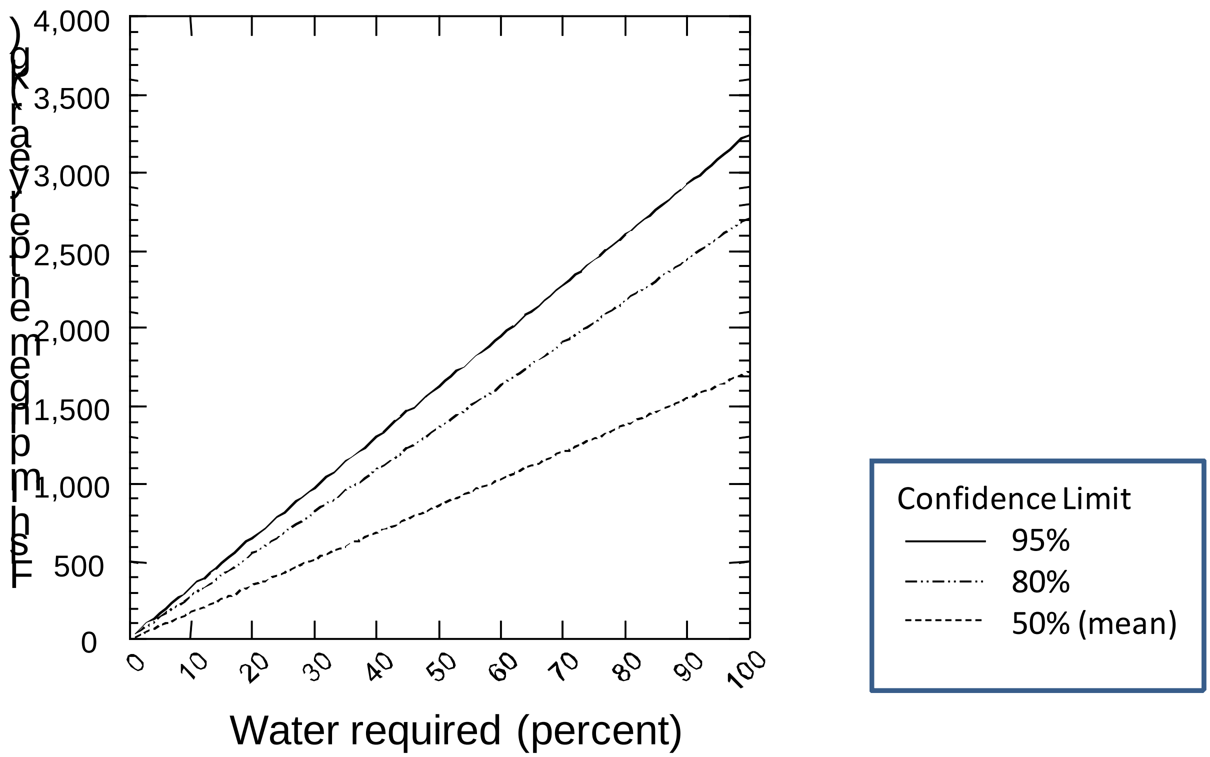

EXHIBIT 1

Problem 1 Change of Variables:

Exhibit 1 does not represent the impingement data in terms of the real units of the sample record. The impingement data were recorded as a certain number of kilograms of fish impinged on a particular sample day, at which time EPS intake opertions pumped a certain number of milions of gallons per day (mgd). In Exhibit 1, the x-axis expresses the intake water consumption as a percentage of 304 mgd (the amount the CDP will consume on a daily basis). It is striking that the x-axis terminates at 100%, even though all the impingement samples were collected at intake flows exceeding 304 mgd; average intake flows were 658 mgd during the impingement sampling, corresponding to a water usage of 216%. The y-axis in Exhibit 1 is an annualized impingement loss, representing a cumulative loss derived from a discrete number of daily samples.

To derive the “50% confidence limit” curve in Exhibit 1, the two outlier samples were thrown out and the total biomass from the remaining samples of the study was multiplied by 365 and divided by the number of remaining samples (50 samples). To illustrate the arithmetic, total biomass impinged (less outliers) = 233.5 kg and yields a daily impingement loss of 233.5/50 samples = 4.67 kg/sample and rounded to 4.7 kg/day, giving an annualized mass of 1,715 kg. This annualized impingement loss was incorrectly assigned a water usage of 100% and connected by a dotted straight line to the origin to get a form of the flow proportional impingement function that is called the “50% (mean) confidence limit” in Exhibit 1 and the blue line in Figure 1. The problem with this derivation is that the annualized impingement loss of 1,715 kg did not occur at 304 mgd (100%); rather, it occurred at EPS flow rates considerably higher, on average the equivalent of 216% (657 mgd). Thus, the 50% confidence limit curve should have been shifted to the right as shown in blue in Figure 1, greatly diminishing the slope of that line.

The mistake is repeated in formulating the “80% confidence limit” curve in Exhibit 1. It appears this curve was derived from the total biomass impinged by all the samples, totaling 372.5 kg, yielding an annualized impingement loss of 372.5 x 365/52, rounded to 2,700 kg. The “80% confidence limit” curve in Exhibit 1 was then formulated by connecting the origin with a straight dashed line to 2,700 kg at the incorrect water usage assignment of 100%. Instead, the annualized total biomass should have been ascribed to the actual flow rate that occurred during the Impingement study that averaged 657 mgd as denoted by the dashed vertical line in Figure 1 at 216% water usage. The 95% confidence limit curve in Exhibit 1 was ascribed to the mean plus 95% of the mean, or a daily impingement of 4.67 kg/d + 0.95(4.67 kg/d) = 9.106 kg /d rounded to an annualized impingement loss of 3,250 kg. But again this loss was ascribed to only 304 mgd, when in fact flow rates were never that low throughout the impingement study and averaged 657 or 216 % water usage. Consequently the 95% confidence limit curve should have drawn through the 216% water utilization point as shown by the dashed sloping black line in Figure 1. Because the remainder of the Raimondi analysis to quantify mitigation acres is linear, his mitigation estimates are too high by at least a factor of 2.16.

Figure 2 illustrates how an annualized flow proportional impingement function can be derived directly from the impingement data without introducing distortion from incorrect assignments of the intake water consumption. Each impingement sample is

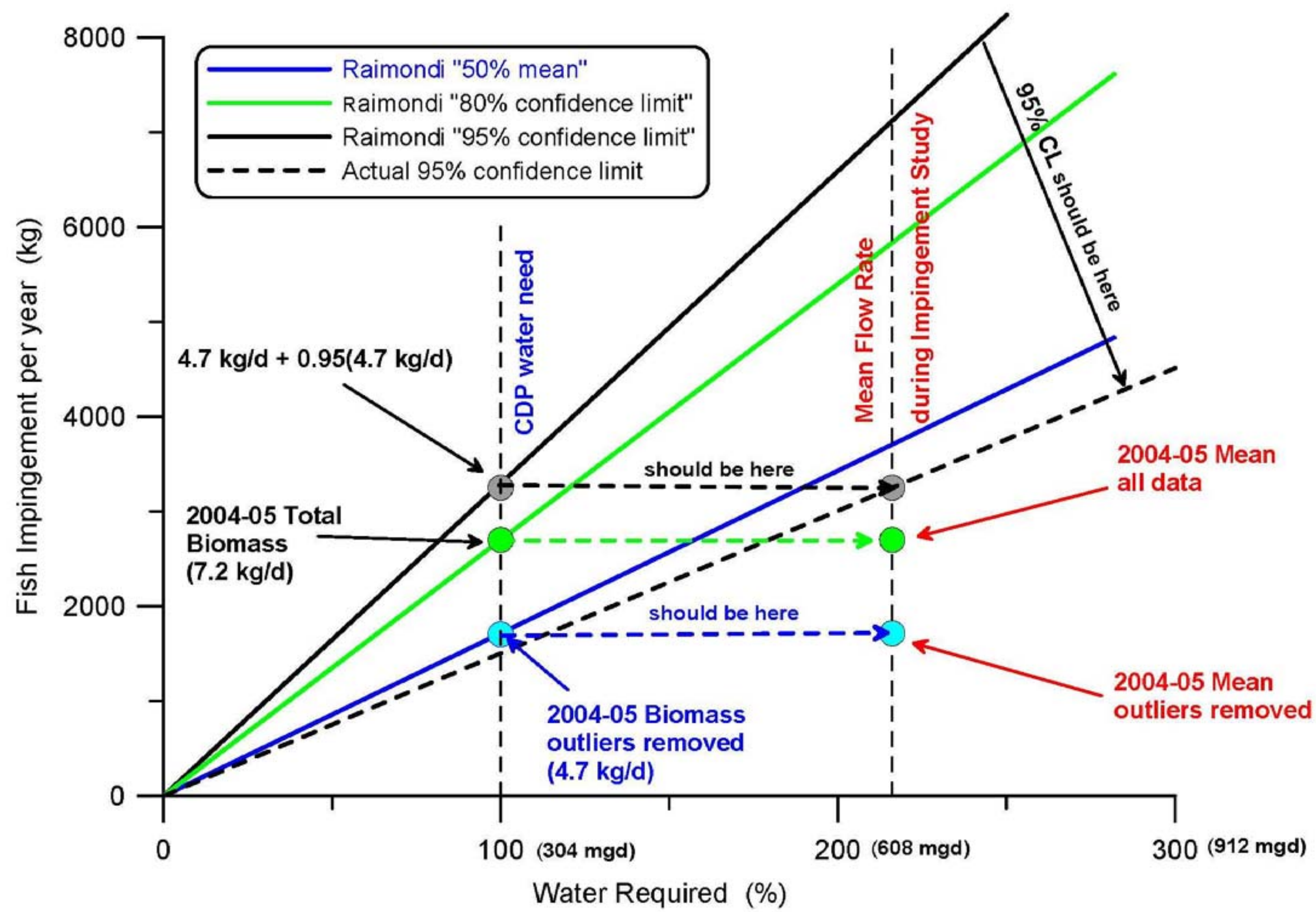

Figure 1: Establishment of Mitigation levels and "confidence limits" relative to actual versus assumed flow rates. Water required is the percentage basis of actual water usage during impingement sampling divided by water required for operation of Carlsbad Desalination Project (304 mgd). Raimondi "confidence limit" curves from p. 4 of 1 April 2009 report to RWQCB.

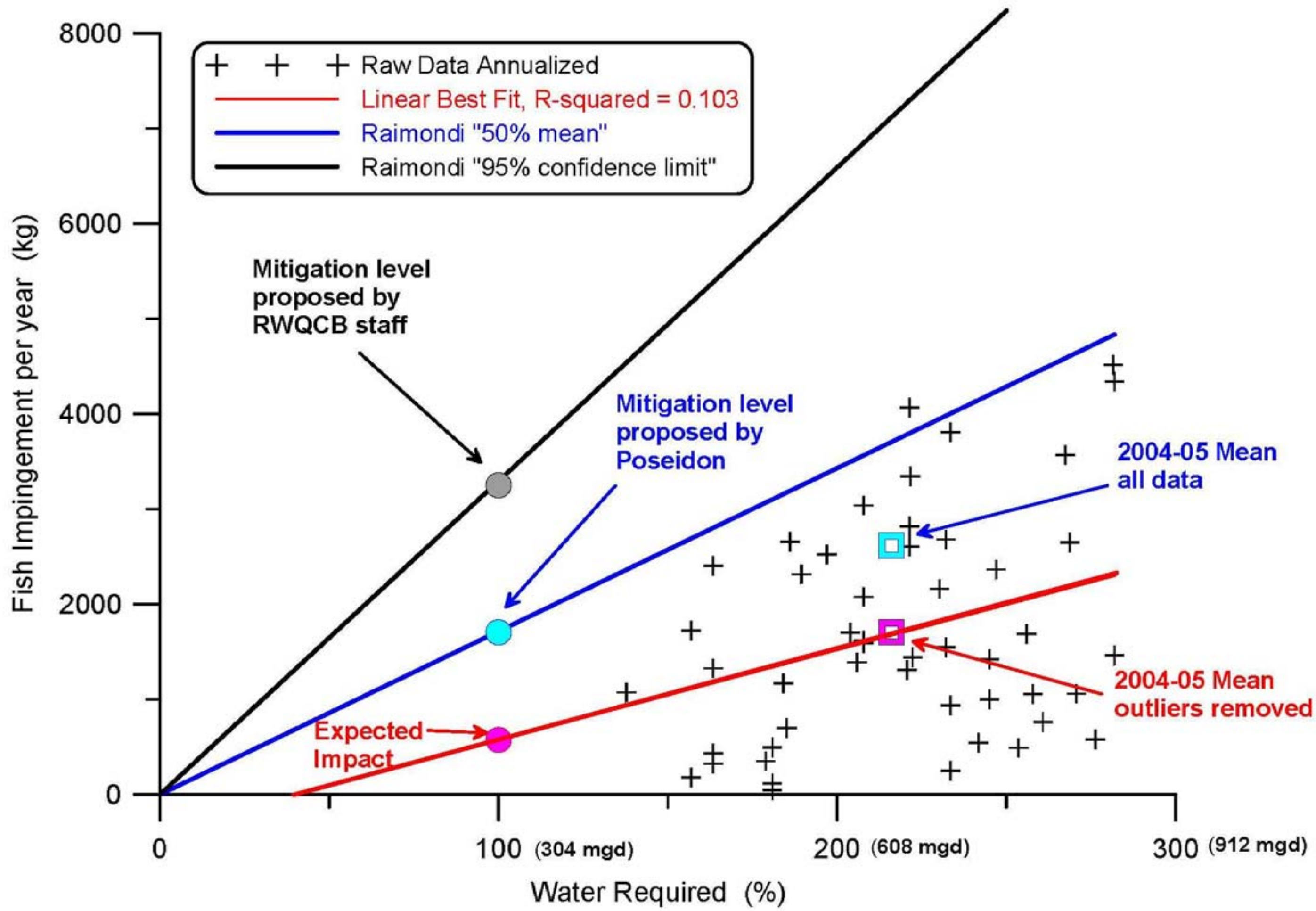

Figure 2: Mitigation levels and "confidence limits" relative to annualized impingement data with outliers removed (50 samples). Water required is the percentage basis of actual water usage during impingement sampling divided by water required for operation of Carlsbad Desalination Project (304 mgd). Raimondi "confidence limit" curves from p. 4 of 1 April 2009 report to RWQCB.

individually annualized by multiplying the impingement mass for that day by 365 days. Omitting the outliers, the impingement data points thus annualized appear as black crosses in Figure 1, all of which correspond to water usage in excess of 100%. A linear best fit to these data by least squares produces the alternative flow proportional impingement function shown in red. In addition, the ensemble average of total annualized biomass impinged (including the outliers) is plotted as a cyan colored square at the mean flow rate of 657 mgd (216% water usage). The 50% and 95% confidence limit curves from Exhibit 1 are shown in blue, and black, respectively. Note that the 50% and 95% confidence limit curves from Exhibit 1 substantially over-predict the mean impingement losses derived from all the data, and only three of the individually annualized data points (including the two outliers) exceed the 50% confidence limit curve from Exhibit 1.

Problem 2 Usage of Confidence Limits:

A confidence interval represents a closed interval where a certain percentage of the sample population is likely to lie. In the case of a confidence limit, the lower bound of the interval is set at zero and the upper bound is set at the prediction, in this case the *flow proportional impingement function*. Raimondi asserts "in typical inferential statistics, confidence limits of 95% are generally used." While that may be true in card games and other games of chance having discrete outcomes, predictive skill in natural systems having a continuum of outcomes is seldom that good or even possible. This is especially true in hydrodynamic transport modeling and the problem at hand clearly represents a coupled biological/hydrodynamic system.

When data is extremely noisy and poorly ordered with respect to the parametric variables one tries to employ, the use of confidence intervals and limits approaches insensibility. For example, the best fit regression line to all the impingement data in Figure 3 has a very low coefficient of determination, R-squared = 0.018, indicating a marginal correlation between the EPS flow rate and impingement. (Confidence limits approach infinity for random processes).

The most likely explanation for the poorly ordered data in this study is that the impingement sample set contains two distinctly different populations, the two outlier samples shown as red crosses in Figure 3, and the remainder of the samples shown as black crosses. When a sample set shows this kind of bifurcation, it is not a logical error (false converse) as Raimondi asserts, but rather logical recourse to suspect other variables (besides EPS flow rate) are influencing the result. After all, natural systems are not in general single valued functions of a single variable such as Exhibit 1 and Figure 3 seek to portray.

In our 3 April 09 submissions to the RWQCB staff, time histories of rainfall, daily high and low water levels and tidal range were presented; and only the rainfall showed any correlation with the high impingement samples (where 7 of the highest 10 impingement sample days were either coincident with or preceded by rainfall). In the most recent correspondence from staff, inference was made that dredging was a causal agent of the two outliers. However dredging is performed round the clock during the dredge cycle by a suction dredge which bypasses its discharge past the EPS intakes and onto the beach via a pipeline. Therefore the spoils (and fish killed by the cutter head) are not available to impingement

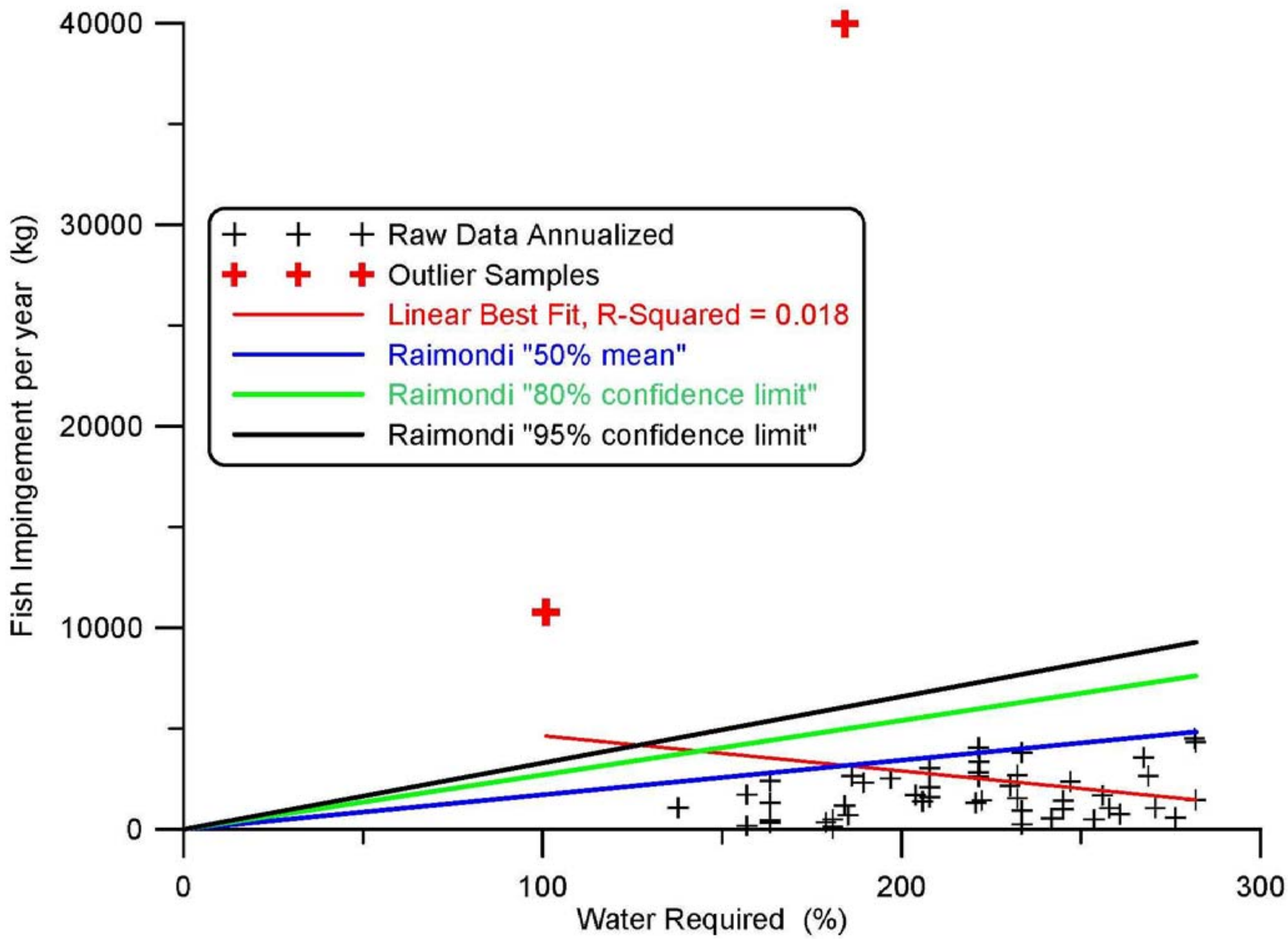

Figure 3: Impingement data (black & red crosses, 52 samples) expressed in terms of change of variables per p. 4 of Raimondi 1 April 2009 report to RWQCB. Outlier samples shown in red, "confidence limits" (blue, green and black). Water required is the percentage basis of actual water usage during impingement sampling divided water required for operation of Carlsbad Desalination Project (304 mgd).

by the EPS intake, and many other nominal or low impingement counts occurred throughout the 2005 dredging cycle.

Conclusion: By a change of variables and incorrect assignments thereof, a distorted annualized form of the *flow proportional impingement function* was developed on p.4 of the Raimondi 1 April 09 report. Slopes of this function are steeper than what would be derived from the impingement data directly, and consequently, the calculations of mitigation requirements that are based on that function are inflated. The impingement data are not sufficiently well ordered to make sensible use of confidence limits. However, the use of these limits in trying to quantify mitigation does not appear necessary since the project now features a post-project impingement monitoring component that will “true up” any significant error in assessing the persistent impingement levels.