CBP Response To the June 19, 2009 Technical Memorandum Submitted by Malcolm Pirnie on behalf of V/MAMWA

July 16, 2009

Comment #1

The CBP position is that relaxation of either the sample size requirement or the standard deviation criterion increases the probability of inaccurately classifying a degraded segment as healthy, or vice versa. This position is based on extensive discussions with benthic IBI experts (Llanso and Dauer, personal communications 2009) as well as peer-reviewed publications (Llanso et al 2009). While the relaxation of these criteria does result in small increases in the number of segment-periods identified as "healthy" or "degraded," examination of the curves in question suggests that some of these segments may be inaccurately classified. Thus, inclusion of these segments may distort the bioreference curve, reducing its ability to distinguish between degrees of hypoxia that allow for healthy benthic communities and those which degrade benthic communities.

However, in the interest of full disclosure, the CBP has responded to the request by Malcolm Pirnie on behalf of V/MAMWA to examine the effects of relaxing the data screening criteria to accept segment-period combinations with sample size ≥ 8 and/or standard deviation ≤ 1.2.

It has been reported by Malcolm Pirnie that relaxation of the *n* requirement from 10 to 8 and/or the maximum standard deviation from 1.0 to 1.2 substantially increases the number of "healthy" segment-periods for analysis. These results appear to have been obtained using a flawed dataset. Per email communications with Clifton Bell of Malcolm Pirnie dated May 26, 2009, the original dataset provided to V/MAMWA and Malcolm Pirnie contained duplicate records. Specifically, for the "fixed station" samples both "total_score" and "grand_score" records were included. "Total_score" records are replicate measurements of the same sampling event; the average of these is reported as the "grand_score." Benthic experts (Llanso) recommend using the "grand_score" in our analyses. When "total_score" records are removed from the benthic dataset, the numbers reported in Table 1 of the Malcolm Pirnie Technical Memorandum (June 19, 2009) are somewhat reduced – see Table 1 below.

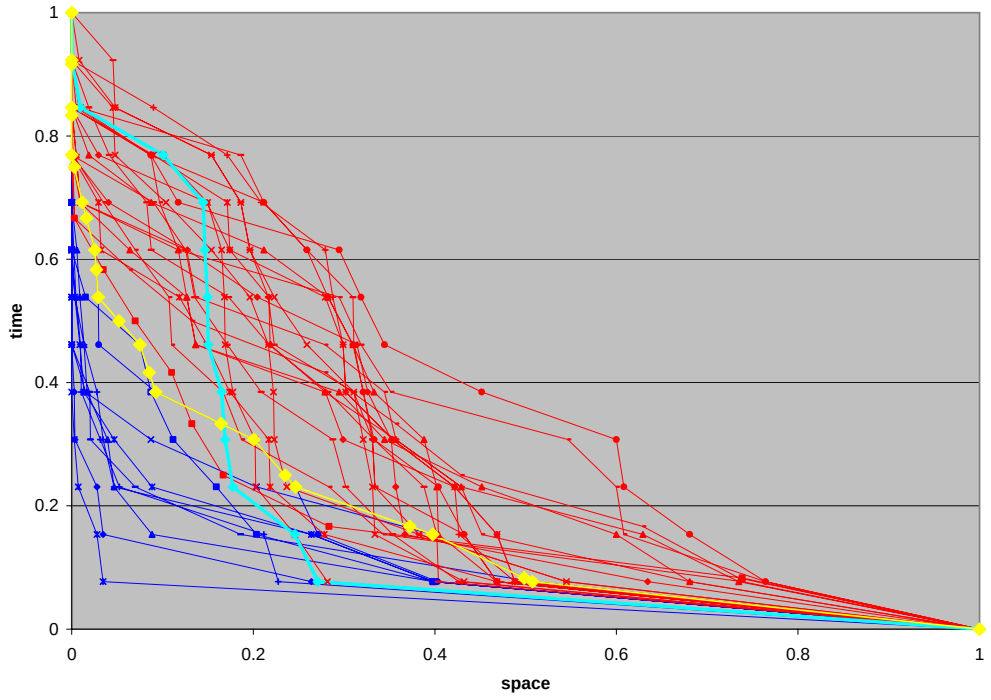
Table 1: Revised From Malcolm Pirnie Tech. Memo. To Reflect Corrected Data

	Scenario A (Default) B-IBI ≥ 3.0 $n \geq 10$ S.D. < 1.0	**Scenario B** B-IBI ≥ 3.0 $n \geq 10$ S.D. < 1.2	**Scenario C** B-IBI ≥ 3.0 $n \geq 8$ S.D. < 1.0	**Scenario D** B-IBI ≥ 3.0 $n \geq 8$ S.D. < 1.2
Total number of "healthy" deep water segment-periods	10	11	13	16

Relaxation of the criteria results in moderate increases (ranging from 1 to 6) in the number of segment-periods classified as "healthy." Due to the increased risk of inaccurate classification, it is important to examine not just the number of additional segment-periods, but also the shape of these curves. If a curve is classified as "healthy" but its location in CFD space is consistent with DO violation CFDs of segment-periods classified as "degraded," then it is reasonable to question whether an inaccurate classification has occurred.

In the case of Scenario B (relaxing the standard deviation criterion from a maximum of 1.0 to a maximum of 1.2), a single curve (CB5MH 1999-2001) is added to the group of "healthy" segment-periods. In Figure 1 below, this curve is visible as a light blue line, while the population of 10 curves identified in Scenario A are presented by dark blue lines. Degraded segment-periods are visible as red lines. The bioreference curve generated from the 100th percentile of "Scenario A" violations at each time step is visible as a yellow line.

Figure 1: Scenario B

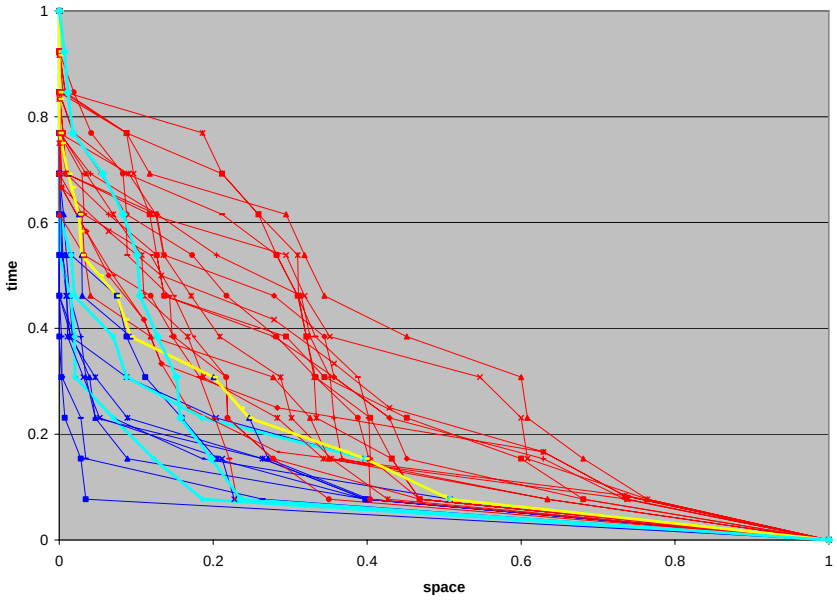


The shape of the CB5MH 1999-2001 curve (light blue line in Figure 1) raises the question of whether increasing the uncertainty of our screening criteria resulted in erroneous classification of this segment-period as healthy. In particular, the location of the top half of this curve in CFD space that is dominated by degraded curves decreases our confidence in the accuracy of its classification. The addition of this curve, particularly in combination with the methodology of taking the 100th percentile of each curve at each point in time, would increase the potential for the resulting

bioreference curve to allow rates of hypoxia that result in degradation of the benthic community.

In the case of **Scenario C**, the standard deviation is kept consistent with our recommended criteria but the sample size criterion is relaxed from 10 to 8. This relaxation of the recommended criteria results in the classification of 3 additional segments as "healthy." The CFD curves for these additional segments are shown as light blue curves in Figure 2 below.

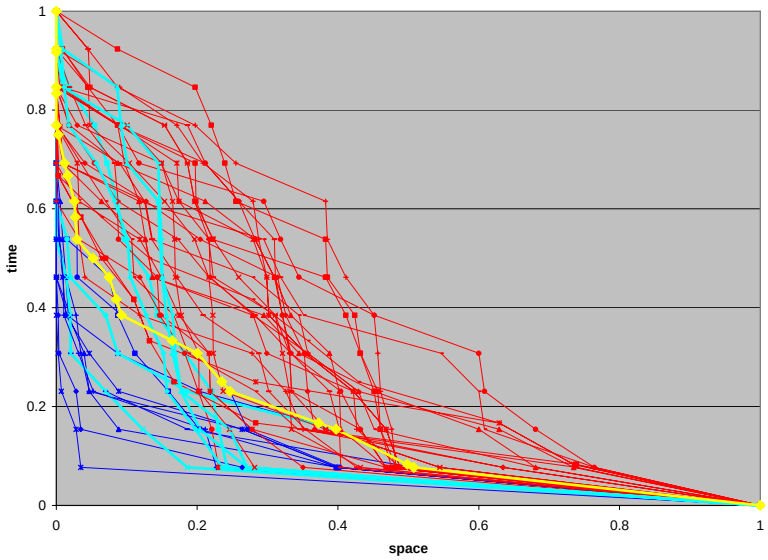
Figure 2: Scenario C



While two of the additional curves (CB6PH 1998-2000 and CB6PH 2000-2002) fall within the cloud of violation rates deemed "acceptable," one curve (CB3MH 1996-1998) once again extends into the cloud of data dominated by CFDs associated with degraded segment-periods (Figure 2). As described earlier, this raises the concern that relaxation of our criteria has resulted in the inaccurate classification of a degraded segment-period as healthy.

The relaxation of both the sample size and the standard deviation criteria (Scenario D) increases the number of segment-periods classified as "healthy" from 10 to 16. However, 4 of these additional CFD curves extend into "degraded" CFD space to a degree that calls into question the accuracy of their classification as healthy (see Figure 3 on next page).

Figure 3: Scenario D



Conclusion

It has been suggested that relaxing our criteria with respect to minimum sample size and maximum standard deviation increases the number of healthy segments that can be used to generate the bioreference curve. Based on our findings described here, our position is that the increased uncertainty of accurate classification resulting from relaxation of the criteria far outweighs the potential benefit of increased sample size. Our methodology results in a total sample size of 24 segment-periods, of which 10 are classified as healthy and 14 are classified as degraded. We find this to be a sufficient sample size for elucidating the boundary between acceptable (i.e. those which allow a healthy benthic community to persist) and unacceptable violations of the Deep Water D.O. criteria.

Comment #2

It is suggested that the reduction in false positives and false negatives resulting from our change in methodology is caused not by the production of a more accurate curve, but by the removal from analysis of many segment-periods that were previously found to be false positives. The comment stated that there are numerous "healthy" segment-periods that are excluded from our analysis.

It is accurate to say that many of the segment-periods that generated false positives have been removed from our analysis. However, it is misleading to conclude that numerous healthy segment-periods were excluded.

The purpose of these method modifications is to increase the accuracy with which we classify segment-periods as "healthy" or "degraded." Therefore, segment-periods that contained insufficient data for accurate classification were excluded. Our analyses illustrate that relaxing the criteria with respect to sample size and standard deviation results in the inclusion of segment-periods for which the accuracy of the classification

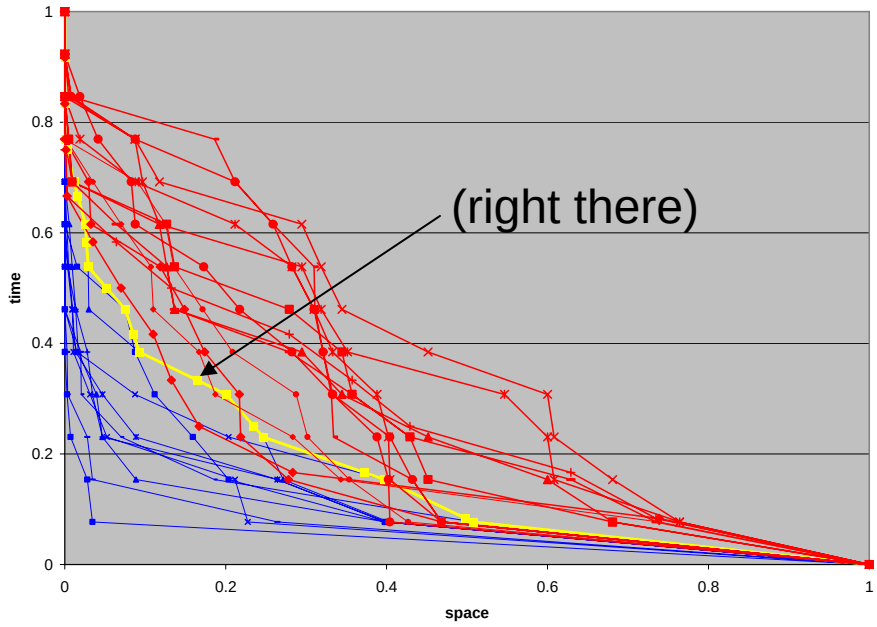
can reasonably be called into question. To include these segments would increase uncertainty with respect to the accuracy of the bioreference curve. For instance, with regard to the example cited in Comment #2: we have illustrated that the shape and location in CFD-space of the curve associated with CB5MH 1999-2001 calls into question the accuracy of its classification as "healthy" under the relaxed screening criteria see Figure 1).

Comment #3

It is proposed that comparing the total area under a CFD assessment curve to the total area under the bioreference curve is a better measure of the degree to which healthy biological communities can tolerate violations of the DO criteria than the existing "point" method. Arguments put forth to support this proposal include: (1) a segment-period may exceed the bioreference curve in one area of CFD space while the overall area of its exceedance is within than that represented by the bioreference curve; (2) there is high variability in the shape of CFD curves and the data do not allow identification of combinations of time and volume that lead to poor B-IBI scores in a segment; (3) the proposed "area" method has lower error rates than the published "point" method, even with the modifications proposed by CBP to the latter method.

With regard to arguments 1 and 2, CBP's position is that with application of the method modifications, we are now accurately classifying benthic communities as "healthy" or "degraded" when we have sufficient data to do so. As a result, the data are able to inform us of a rather specific combination of time and volume that forms the boundary between healthy and degraded benthic communities in the Deep Water designated use (see Figure 4).

Figure 4: D.O. violation curves associated with healthy (blue) and degraded (red) benthic communities in Deep Water. The Deep Water bioreference curve (yellow) is also shown.



With regard to Argument 3, tables 2 and 3 of the submitted comments appear to have been constructed using a flawed dataset, as communicated to Jim Pletl of VAMWA via email on May 14, 2009, and discussed with Clifton Bell of Malcolm Pirnie in email communications occurring between May 20 and May 25, 2009. Using a dataset with all duplicate records removed and all appropriate criteria applied, the error rate for the "Point Method" is zero. In other words, all segment-periods classified as "healthy" using our proposed criteria pass the proposed Deep Water bioreference curve, and all segment-periods classified as "degraded" fail the proposed bioreference curve. Thus Tables 2 and 3 in the Malcolm Pirnie Technical Memorandum (June 19, 2009) can be revised as follows:

Table 2: Revised From Malcolm Pirnie Tech. Memo. To Reflect Corrected Data

Method	**Correct**		**Incorrect**	
	Healthy Segments Passing	**Degraded Segments Failing**	**Healthy Segments Failing**	**Degraded Segments Passing**
Published "Point" Method	100%	100%	0%	0%
Proposed "Area" Method	100%	100%	0%	0%

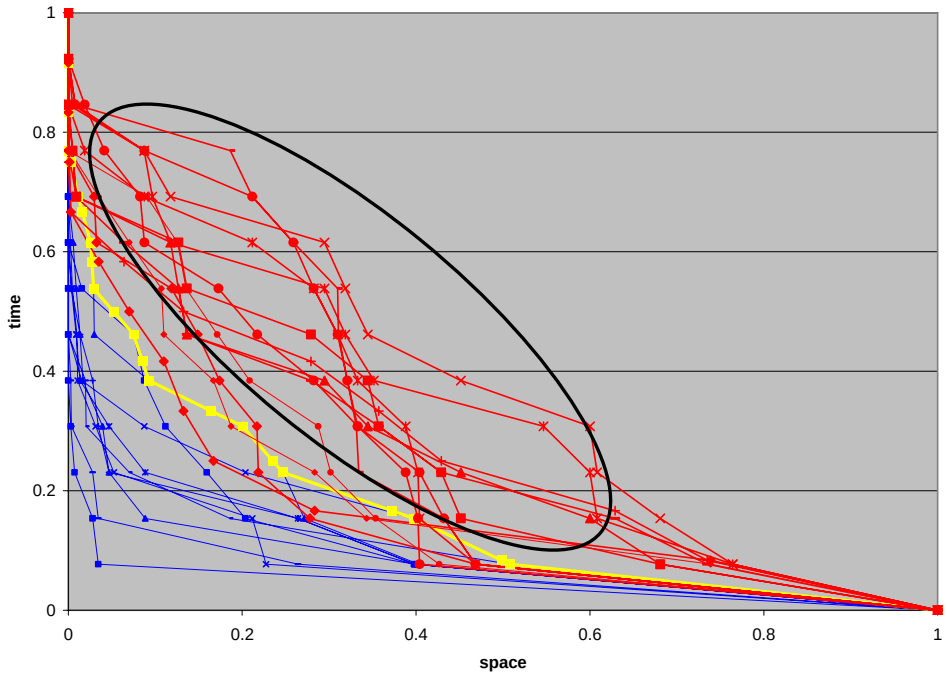
Table 3: Revised From Malcolm Pirnie Tech. Memo. To Reflect Corrected Data

Method	**Correct**		**Incorrect**	
	Healthy Segments Passing	**Degraded Segments Failing**	**Healthy Segments Failing**	**Degraded Segments Passing**
Published "Point" Method	CB6PH_1996_1998 CB7PH_1996_1998 CB6PH_1997_1999 CB7PH_1997_1999 CB7PH_1998_2000 CB6PH_1999_2001 CB7PH_1999_2001 CB7PH_2000_2002 CB6PH_2004_2006 CB7PH_2004_2006	PAXMH_1996_1998 POTMH_1996_1998 PAXMH_1997_1999 POTMH_1997_1999 POTMH_1998_2000 PAXMH_1999_2001 POTMH_1999_2001 PAXMH_2000_2002 RPPMH_2000_2002 PAXMH_2001_2003 PAXMH_2002_2004 PAXMH_2003_2005 PAXMH_2004_2006 RPPMH_2004_2006		
Proposed "Area" Method	CB6PH 1996-1998 CB6PH 1997-1999 CB6PH 1999-2001 CB6PH 2004-2006 CB7PH 1996-1998	POTMH19992001 POTMH19982000 RPPMH20022004 PAXMH19992001 PAXMH20012003		

	CB7PH 1997-1999 CB7PH 1998-2000 CB7PH 1999-2001 CB7PH 2000-2002 CB7PH 2004-2006	PAXMH20042006 POTMH19971999 PAXMH20032005 PAXMH20002002 POTMH19961998 RPPMH20002002 PAXMH20022004 PAXMH19961998 PAXMH19971999		

Thus both methods result in the same error rates when duplicate records are removed and CBP's proposed criteria are applied to the classification of benthic communities. However, in contrast to Argument 2 as described above, it is our position that this dataset does provide us with convincing biological information with regard to the degree and distribution of Deep Water DO criteria violations that can be tolerated by the benthic community. Furthermore, by using the worst violation rate allowed by any healthy community at each point in time, we have allowed for greater violation rates in regions of CFD-space where CFD curves from healthy and degraded communities overlap. As illustrated and described in our response to Comment #1, it is reasonable to postulate – based on the distribution in CFD-space of curves associated with healthy and degraded benthic communities – that violations occurring in the CFD-space circled in black in Figure 5 below lead to degradation of the benthic community.

Figure 5:



Thus we propose that the shape of the bioreference curve is an important factor in identifying acceptable violations of the DO criteria.

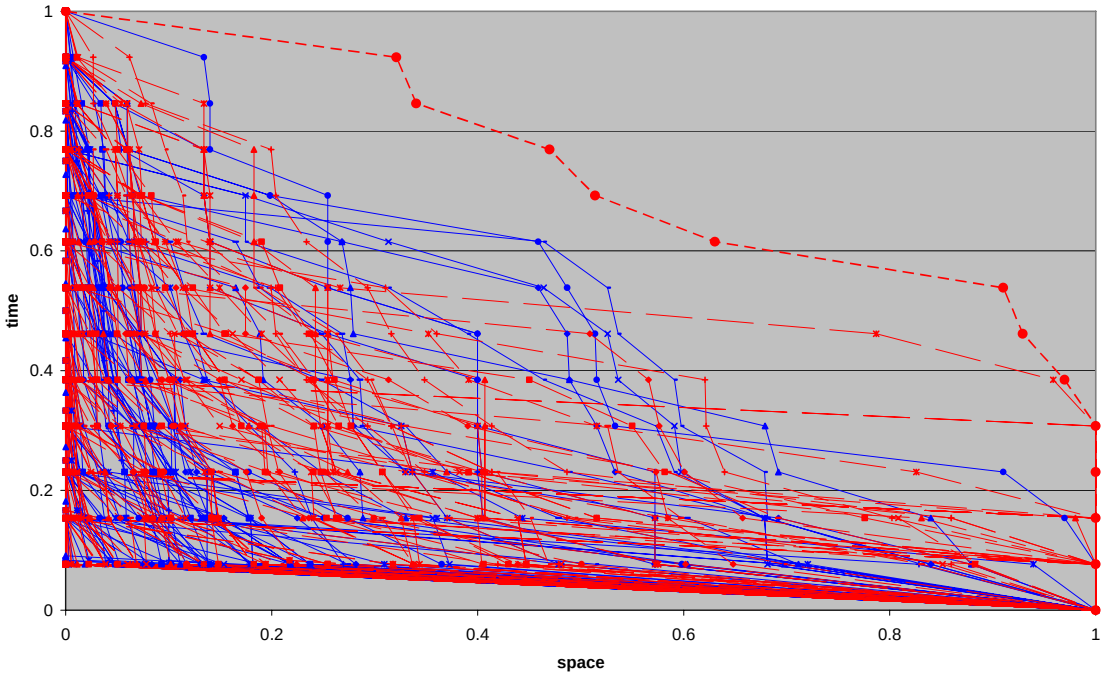
The shape of this biological reference curve also provides further support for the suitability of the hyperbolic 10% reference curve, in that it illustrates the sensitivity of biological communities to chronic violations of DO criteria.

Comment #4:

It is suggested that the benthic community is a more appropriate measure of acceptable violations of Open Water DO criteria than the 10% default curve.

Our analyses show that the benthic community is not sensitive to violations of the Open Water DO criterion of 5.0 mg/L. The complete overlap of violation rates in CFD-space from healthy and degraded benthic communities (see Figure 6) indicates that violations of the 5.0 mg/L DO criterion are not related to benthic community degradation in these areas.

Figure 6: violation rate CFD curves of healthy and degraded benthic communities in the Open Water designated use



This is further illustrated and confirmed by the observation that the best possible error rate that can be obtained is approximately 50-50, or the equivalent of flipping a coin.

Given the irrelevance of DO criteria violation rates to benthic community degradation in the Open Water designated use, it is inappropriate to use benthic community health as a reference for identifying acceptable rates of DO criteria violation. In cases where no biological reference is available for determining acceptable exceedance of a water quality

standard, the precedent exists to apply a 10% allowance. The CBP position is that until a reference dataset exists that is relevant to the biological communities for whose protection the open water DO criterion was designed, we must default to the EPA published guidance and apply the hyperbolic 10% reference curve.