Control Charting

STATISTICAL ASPECTS OF TURBIDITY MONITORING

ENVIRONMETRICS AUSTRALIA

April 2007

Limitations Statement

The sole purpose of this document is to identify options for the statistical analysis of turbidity data as provided to Environmetrics Australia Pty. Ltd. by the Port of Melbourne Corporation (PoMC). The analyses and techniques presented herein are intended to be indicative only. The passage of time, manifestation of latent conditions or impact of future events may require further exploration, subsequent data analysis, and re-evaluation of the findings, observations, conclusions, and recommendations expressed in this document. Accordingly, Environmetrics Australia Pty. Ltd. accepts no liability or responsibility whatsoever for or in respect of any use of or reliance upon this document, its recommendations or any other information contained herein by any party.

STATISTICAL ASPECTS OF TURBIDITY MONITORING : CONTROL CHARTING

Prepared for Port of Melbourne Corporation

By

Environmetrics Australia Pty. Ltd.

April 2007

Table of Contents

List of Figures

1. INTRODUCTION

This report should be read in conjunction with the companion report *"Statistical Aspects of Turbidity Monitoring: Setting Environmental Limits"* (Environmetrics Australia 2007a). Having established a 'limit' for turbidity, it is both necessary and desirable to establish a monitoring program which can assess performance against this limit by tracking fluctuations on appropriate spatial and temporal scales. In developing a rigorous and statistically defensible monitoring program it is imperative that *all* users of the monitoring data have a shared understanding of:

- The purpose of monitoring;
- The magnitude of 'natural' or background variation in the parameter(s) being measured;
- The statistical concepts of Type I and Type II errors;
- The consequences of exceeding an agreed limit.

These points are expanded on in the following section.

1.1 Design criteria

A monitoring program must be linked to clearly articulated objectives: at a minimum it will address the following questions: *why monitor?; where to sample?; and how frequently to sample?*. Elevated turbidity levels are expected during the capital dredging phase of the Channel Deepening Project (CDP) and this situation has the capacity to impact various environmental assets. A significant concern is the impact of elevated turbidity on seagrass health and survival. The companion report referred to above provided guidelines on the establishment of turbidity thresholds which will satisfy certain minimum light requirements deemed necessary to sustain healthy seagrass meadows in the most vulnerable areas of Port Phillip Bay. In this regard there are two linked turbidity monitoring objectives for the CDP: (i) to assess current turbidity levels in the vicinity of the environmental assets to be protected and to ensure that previously agreed criteria are being met; and (ii) to provide operators with an 'early-warning' capability so that corrective action may be taken *before* the monitoring program flags the existence of a critical situation with respect to turbidity.

The monitoring program described in this report has been developed and 'tuned' to meet requirements (i) and (ii) above. It should not be assumed that data collected under this program can simply be analysed in alternative ways to answer questions it was never intended to address.

The spatial and temporal components of the proposed monitoring program have been examined. It should be kept in mind that environmental monitoring poses certain challenges that are either non-existent or otherwise controllable in settings such as industrial manufacturing or laboratory-based research. Control over environmental factors or background conditions is rarely achievable. The net effect of this lack of control is increased 'noise' in environmental data which can manifest itself in a number of seemingly aberrant ways. The most common is increased variability and the occurrence of extremes – very high or very low readings that are transient in nature. An overarching requirement of any turbidity monitoring program is that the 'signal-to-noise' ratio be sufficiently high so that false-triggering is minimised while the likelihood of responding to real and imminent situations of potential concern is sufficiently high. These are competing objectives that are a function of the level of monitoring. The situation is depicted in Figure 1.

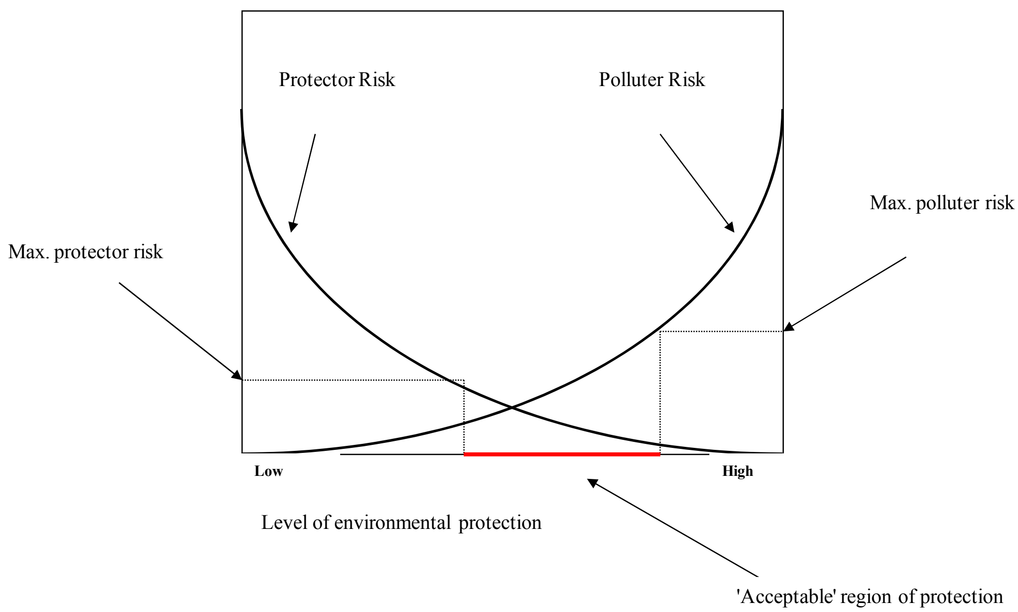

Figure 1. Identification and relationship between competing aspects of risk as a function of the level of monitoring.

In Figure 1 we identify two types of risk. The 'polluter risk' is the risk that a trivial or unimportant (environmental) effect triggers punitive action while the 'protector risk' is the risk that an environmentally important impact goes undetected. These two risks are otherwise referred to as 'Type I' and 'Type II' errors respectively. The competing nature of these risks as a function of the intensity of monitoring effort is readily apparent. The revised water quality guidelines (ANZECC/ARMCANZ 2000) advocate a risk-based approach to water quality monitoring. This represents a significant shift from the previous 'line in the sand' approaches to water quality assessment and regulation and gives explicit recognition to the risks identified in Figure 1 as well as the highly variable nature of water quality parameters. The monitoring program identified in this report attempts to balance both types of error or risk and provides an approach which is entirely consistent with procedures advocated in the water quality guidelines (ANZECC/ARMCANZ 2000).

Another important consideration in the context of environmental monitoring is *spatial variation*. A critical consideration is the determination of the *range of influence* for the parameter of interest. This is essentially the minimum separation between pairs of observations in order for them to be treated as being statistically independent. The notion of statistical independence is an important consideration and its absence can seriously jeopardise the validity of many standard statistical procedures. The monitoring program identified in this report has been developed with due regard to critical aspects of spatial dependency.

Finally, all users and stakeholders of the monitoring program should be agreed on (i) how the data generated from the monitoring program is to be analysed (and by whom); (ii) how the results are to be interpreted; and (iii) what action follows the 'tripping' of an environmental trigger. As has already pointed out, contemporary environmental monitoring/regulation has moved away from 'command and control' approaches which were underpinned by prosaic comparisons of monitoring results with simple numerical limits. These approaches gave little or no recognition to the high levels of natural variation which periodically resulted in excursions beyond a notional limit that were independent of any anthropogenic influence. Risk-based monitoring programs aim to be relatively insensitive to such events and focus on teasing out underlying signals and assessing their significance.

2. PRE-PROCESSING 'RAW' NTU DATA

In-situ turbidity monitoring is to be undertaken by means of autonomous sensors attached to fixed moorings at pre-selected sites around the Bay. The frequency of raw data acquisition is 1Hz. This sampling frequency is much greater than assessment and response time frames. That is, we do not need to be making an assessment of turbidity levels on a second-by-second basis. Also, as discussed in the previous section, the 'raw' NTU data is expected to exhibit fairly high levels of natural (or background) variation. Furthermore other effects such as instrument drift and transient 'spikes' that are unrelated to turbidity (for example, seaweed floating past the sensor) need to be identified and accounted for. Thus, some level of 'pre-processing' of the raw NTU data is warranted. These issues and suggested data processing strategies are discussed below. By way of example, we consider raw NTU data acquired at the Cameron's Bight[1] site between 4-Jan-2005 and 26-Jul-2006[2] (Figure 2). Although the notional sampling frequency was 1Hz, data was only recorded at this rate for 3 x one-minute intervals four times per hour. These were at {0,1,2}, {15,16,17}, {30,31,32}, and {45,46,47} minutes past the hour. On closer inspection, there were missing samples at irregular times as shown in the frequency distribution in Table 1. Overall, a total of 2,189,051 NTU observations were made at this site over the eighteen month period.

Table 1. Frequency distribution for Cameron's Bight NTU data between 4-Jan-2006 and 26-Jul-2007.

minute	Count
0	218900
1	262680
2	48157
15	228649
16	274380
17	50298
30	228550
31	274260
32	50280
45	228550
46	274260
47	50087
N=	**2189051**

[1] Throughout this report, references to Cameron's Bight relate specifically to PoMC site 2601.
[2] These data were obtained from benthic loggers. The proposed monitoring program will use instruments suspended in the water column. This does not affect the statistical explanations and advice given here.

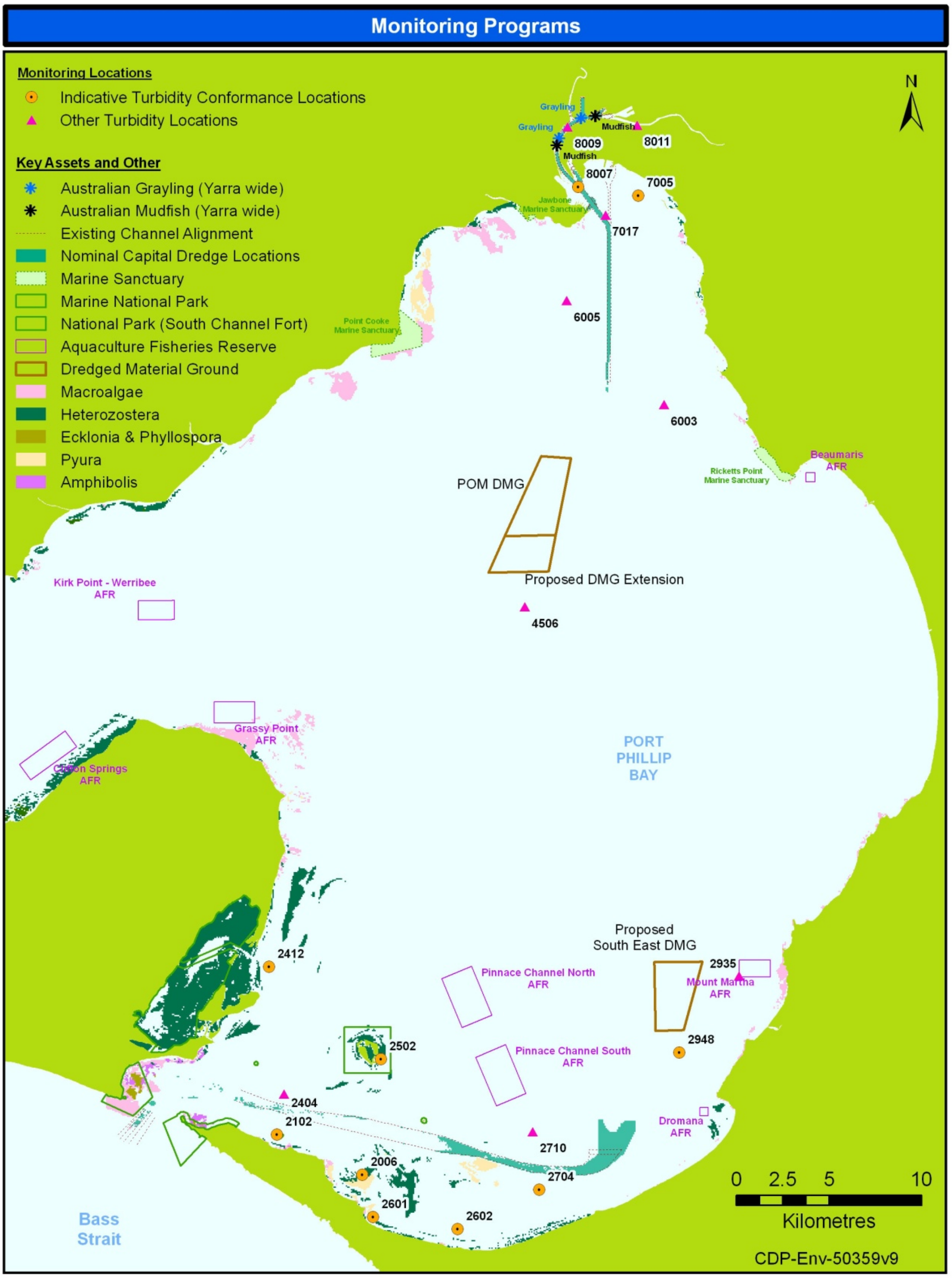

Figure 2. Port Phillip Bay showing elements of proposed dredging program and monitoring sites. (Source: Port of Melbourne Corporation).

To assist in the identification of transient spikes, the raw *NTU* data are differenced. Large (absolute) values of these first differences warrant further investigation as these are indicative of 'aberrant' behaviour that may be indicative of instrument problems, or other situations as described above that are not indicative of any turbidity trend. For the present data, the largest one-second difference in NTU was 87.

Our analysis commences by recoding all observations within a three minute block to the median time. This results in *NTU* data identified as being sampled at 1, 16, 31, or 46 minutes past the hour. A plot of the resulting data is shown in Figure 3. The upward 'drift' in *NTU* readings is clearly evident and this is due to bio-fouling of the instrument sensor. Although the effect is relatively small (increasing by about 0.02 *NTU* per day), it is easily corrected by removing the linear trend component of the *minimum NTU* reading as a function of time. This corrected *NTU* time-series plot is shown in Figure 4.

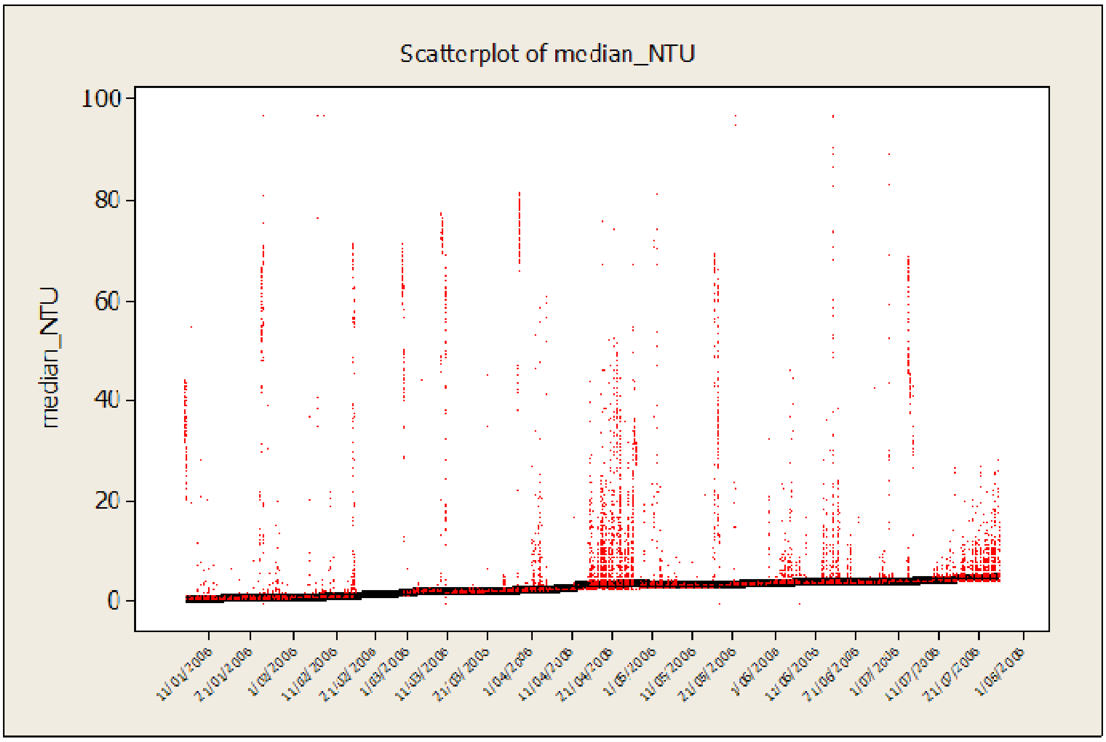

Figure 3. Time-series plot of raw NTU data. The gradual increase over time is indicative of sensor fouling.

The distribution of *NTU* readings is invariably *right-skewed* as shown in Figure 5. This will be an important consideration if the data are to be used for statistical inference using techniques that conventionally assume normality in the distribution of results.

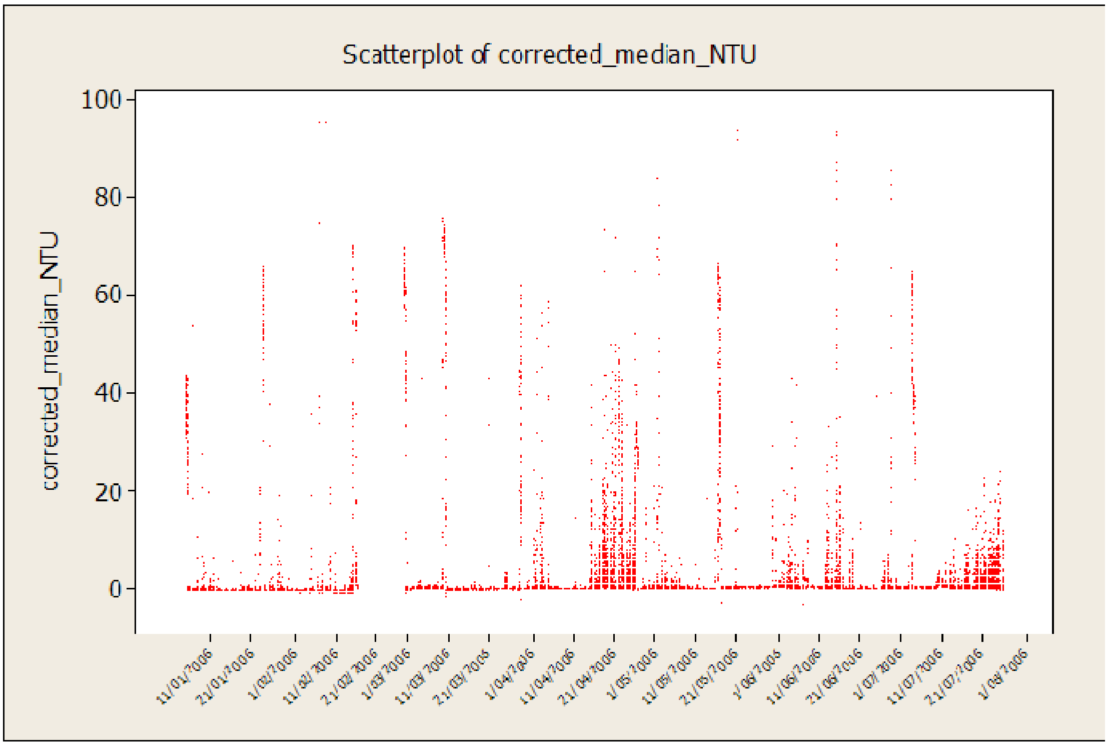

Figure 4. NTU time-series plot corrected for instrument drift.

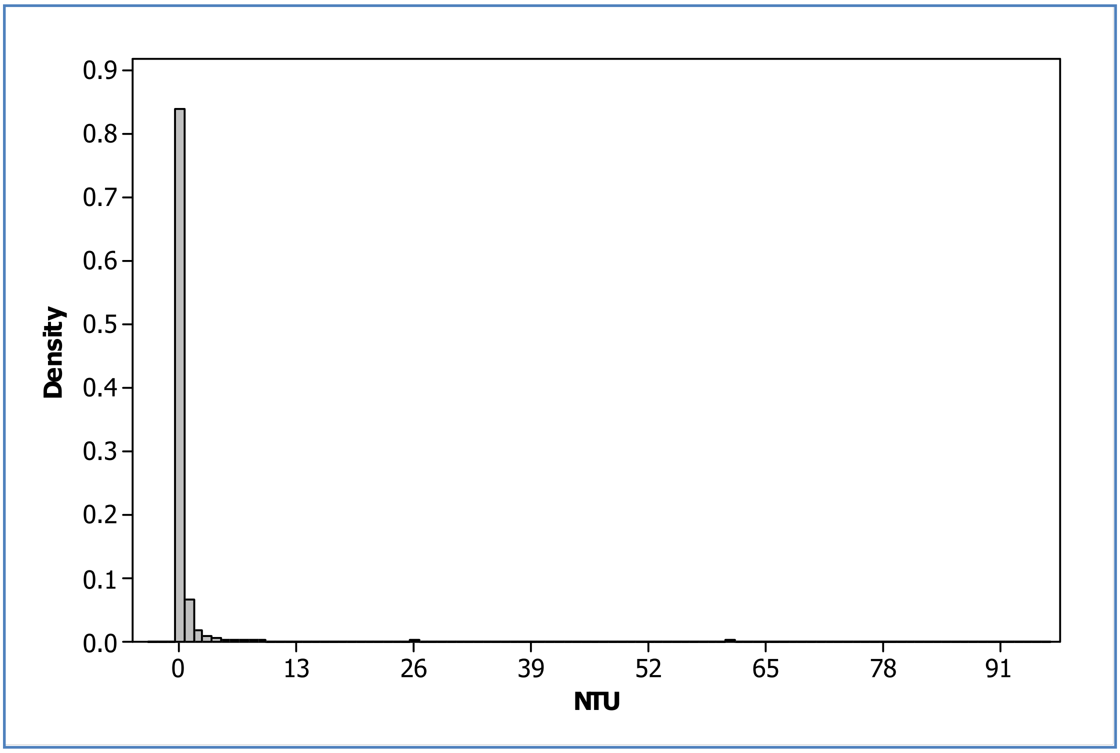

Figure 5. Empirical histogram for NTU (medians of nominal 60 x 1-second readings)

As alluded to above, the identification of an appropriate interval of time for aggregating *NTU* data is an important consideration. Assessments based on raw data acquired at high sampling frequencies may be both unwarranted and possibly distorted by 'aberrant' observations. On the other hand, detail will be lost if the averaging period is too long. Figure 6 shows boxplots for the median of Cameron's Bight *NTU* data as a function of three aggregation periods and a statistical summary for each distribution given in Table 2.

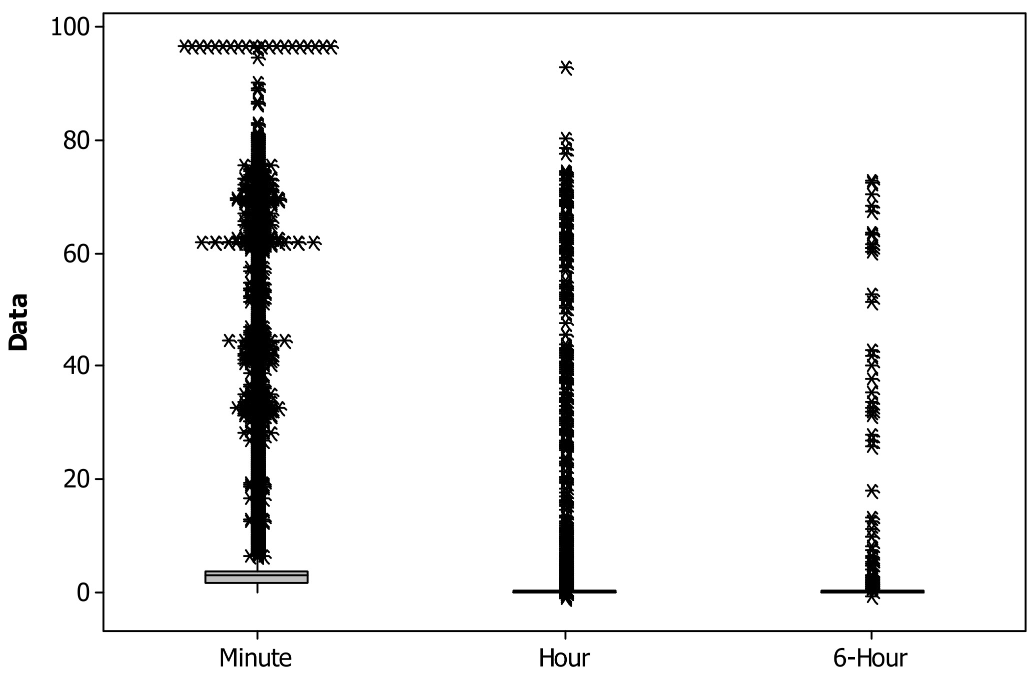

Figure 6. Boxplots for 1Hz NTU data – comparison of medians obtained over 1 minute, 1 hour and 6-hour intervals.

Table 2. Comparison of sample statistics for raw and aggregated NTU data.

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Raw NTU	2147248	4.8463	0.00705	10.3339	-1.9704	1.7925	2.8323	3.7100	96.6280
Minute	54236	4.7039	0.0432	10.0608	-0.1904	1.7850	2.8296	3.6888	96.6280
Hourly	4577	4.584	0.146	9.869	-0.190	1.771	2.813	3.676	96.628
6-hourly	767	4.504	0.353	9.766	0.521	1.782	2.809	3.687	74.812

Both Figure 5 and table 2 illustrate the highly variable nature of *NTU* data which is aggregated on sub-hourly time-frames. Indeed, the *negative* minimum values recorded for these cases are clearly erroneous. From an operational perspective, the aggregation interval needs to be sufficiently long so that the distorting effects of aberrant observations and transient spikes are minimised while still providing a truthful description of overall turbidity conditions. In addition, the aggregation time frame needs to be commensurate with response time-scales, ie. the time lag between intervention and the triggering of an undesirable trend or situation. This trade-off in responsiveness and representiveness is a feature of the competing risks identified in Figure 1. It is clearly impractical to make minute-by-minute assessments of turbidity conditions and to attempt to respond on this time-scale. These and other considerations have led to the identification of a six-hourly aggregation period for processing and reporting *NTU* data. This provides a good representation of actual conditions and implies that there will be a six hour lag between the first warning of an *impending* shift to undesirable turbidity levels and an operational response. It is important to stress that this trigger is an 'early warning' and *not* a declaration of an 'out-of-control' situation. As will be seen in the following sections, the trigger level can be set so that dredging operators are informed of deteriorating conditions which are most likely dredging related and not the result of natural spikes and/or background variability.

3. CONTROL-CHARTING

Control-charts are not new and date back to the early 1920s when their use in industrial and manufacturing contexts was first advocated by Walter Shewhart. The motivation for the development of a 'Shewhart chart' was the partitioning of process variation into 'assignable' and 'chance' causes. Shewhart stressed that bringing a production process into a state of statistical control, where there is only chance-cause variation, and keeping it in control, is necessary to predict future output and to manage a process economically[3]. At its simplest level, a control chart is nothing more than a visual representation of an evolving phenomenon. The chart is generally augmented by a number of visual cues such as:

- A centre line, drawn at the mean of the 'process';
- An upper warning limit[4] placed at k_1 standard deviations above the centre line;
- An upper control-limit placed at k_2 standard deviations above the centre line;
- A lower warning limit placed at k_1 standard deviations below the centre line;

[3] http://en.wikipedia.org/wiki/Walter_A._Shewhart
[4] 'warning' and 'control' limits are standard control-charting terminology – not to be confused with the CDP-specific terminology 'environmental limit' (see footnotes 5 and 7 on page 15).

- A lower control-limit placed at k_2 standard deviations below the centre line

While the choice of constants k_1 and k_2 are arbitrary, the default values are 2 and 3 respectively. In the present context, k_1 and k_2 are to be chosen to provide *operational control* – that is an early-warning capability that could, for example result in a higher level of surveillance (Response level 1) or management action (Response level 2). The process of establishing of response levels or triggers is a statistical one that attempts to balance the degree of forewarning and level of false-triggering. Environmental limits can also be indicated on the chart, although these have been determined by *ecological* rather than statistical factors. An example of the use of control charts for turbidity monitoring is provided in the following sections: one for Hobson's Bay[5] in the north and one for site 2006 in the south. Before moving on to these examples, we pause to briefly consider a particular type of control chart known as the *EWMA* (Exponentially Weighted Moving Average) chart that will be used for turbidity monitoring.

4. THE EWMA CONTROL CHART

A particular advantage of the EWMA chart over other forms of control charting is its ability to be 'tuned' to the circumstances at hand. Because each point plotted on the EWMA chart is a weighted composite of the current observation and past history, it can be made to be more or less responsive to current conditions. The *EWMA* chart was developed by Roberts (1959) as an alternative to the popular CUSUM chart (Page 1954). In addition to the ability to tune the *EWMA*'s responsiveness, it has a number of other desirable features such as:

- It makes use of *all* the available data;
- It can be used to detect both large and small shifts in a process;
- It can be used to track process variability as well as changes in a mean response;
- It is relatively insensitive to non-normal data.

Given data $X_1,\ldots,X_n$, values of the *EWMA* statistic are computed using equation 1.

[5] Throughout this report, references to Hobson's Bay relate specifically to PoMC site 7005.

$$Z_i = \lambda X_i + (1-\lambda) Z_{i-1} \tag{1}$$

with the choice of parameter value λ chosen to weight the relative contributions of the present data and past observations ($0 < \lambda < 1$).

4.1 Illustrative example: Hobson's Bay

Half-hourly background turbidity data was collected between 9/03/2005 13:46 and 21/07/2006 15:01 (a total of 40,053 observations). Figure 6 shows that there is little loss of the essential NTU characteristics by aggregating these data over 6-hourly periods. Furthermore, the log-transformed NTU data (right-hand panels of Figure 7) exhibit a far greater degree of normality than the original NTU data.

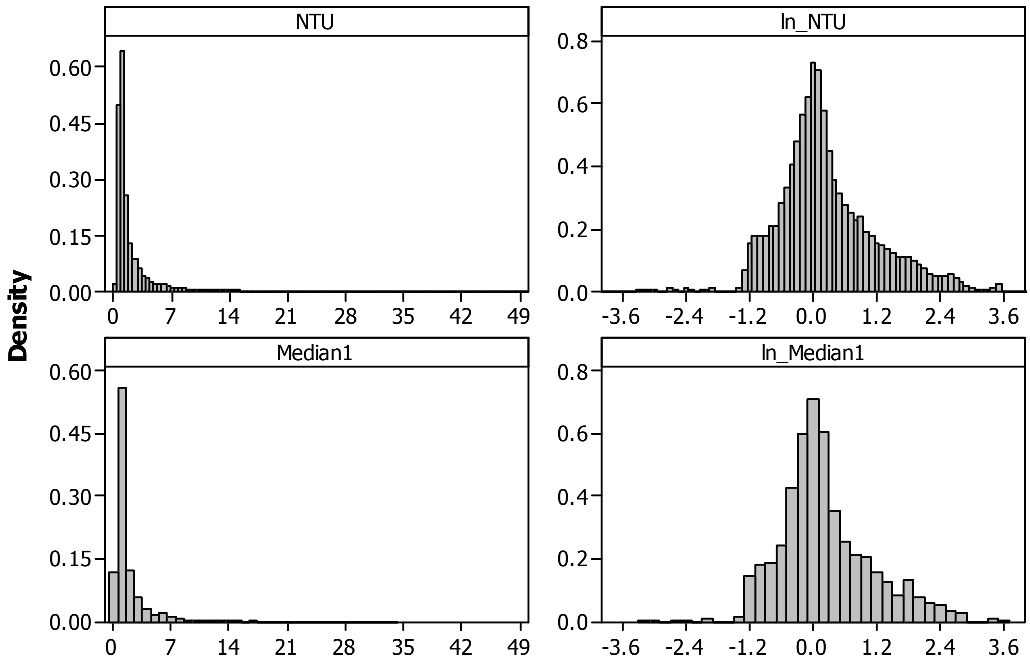

Figure 7. Histogram of Hobson Bay NTU data: 15-minute data (top panel); 6-hourly medians (bottom panel).

Modelled output on incremental suspended sediment concentrations during dredging in Hobson's Bay was provided by PoMC. This data set comprised a total of 14,787 half-hourly predicted TSS concentrations between 1/1/2008 and 11/3/2009. These were converted to incremental NTU values by dividing the TSS concentration (mg/L) by 1.4462 (Environmetrics Australia 2007a). The background and modelled

NTU data was matched to the nearest 30 minute period for each day there was both a modelled *TSS* and a background turbidity reading (without regard to 'year'). A statistical summary of total *NTU* is shown in Figure 8. It is evident from Figure 8 that the predicted additional *NTU* in Hobson's Bay will be less than 4 units during dredging. Histograms of both the background and total *NTU* data is shown in Figure 9 while Figure 10 shows a comparison between the *cdfs* for both data sets. The predicted impact of dredging on background turbidity is seen to be minor.

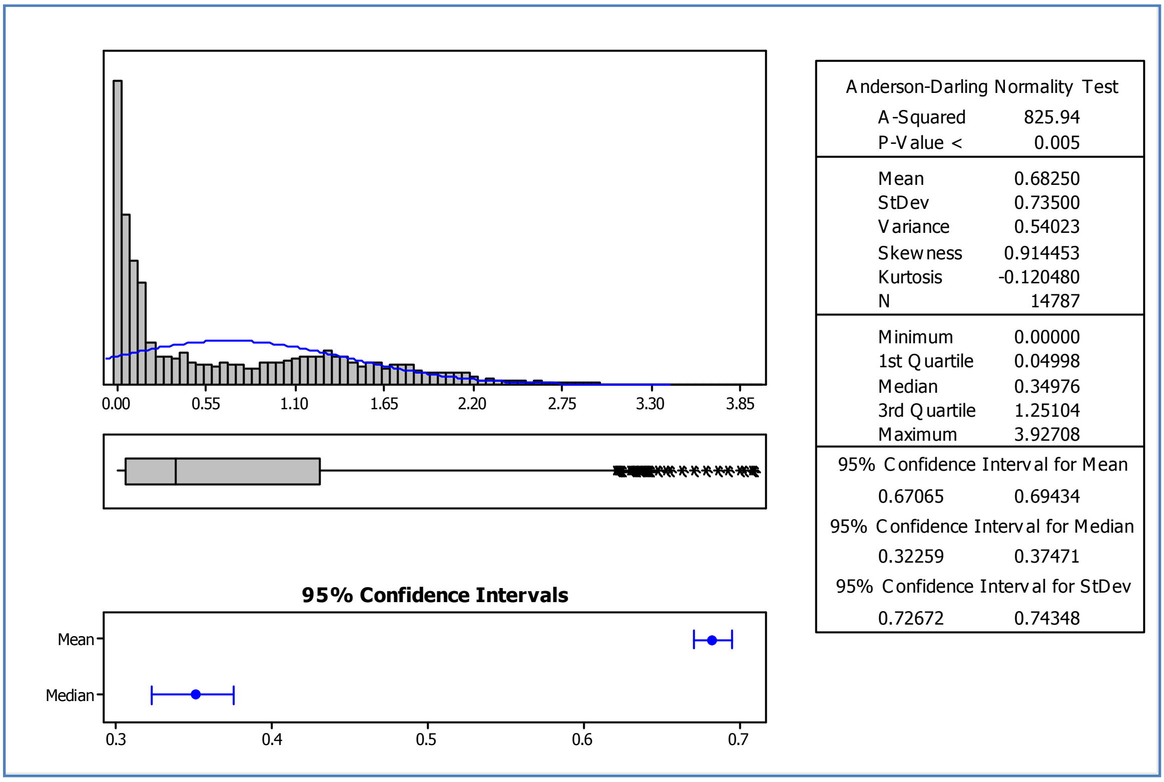

Figure 8. Predicted incremental NTU

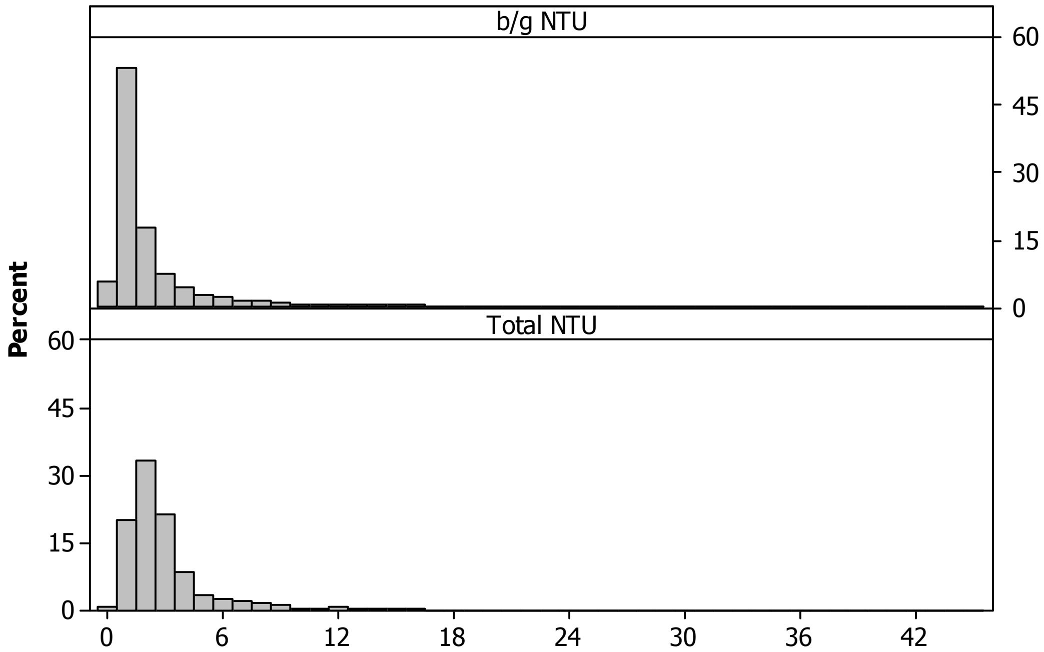

Figure 9. Histograms of background turbidity (top) and background + predicted dredging (bottom).

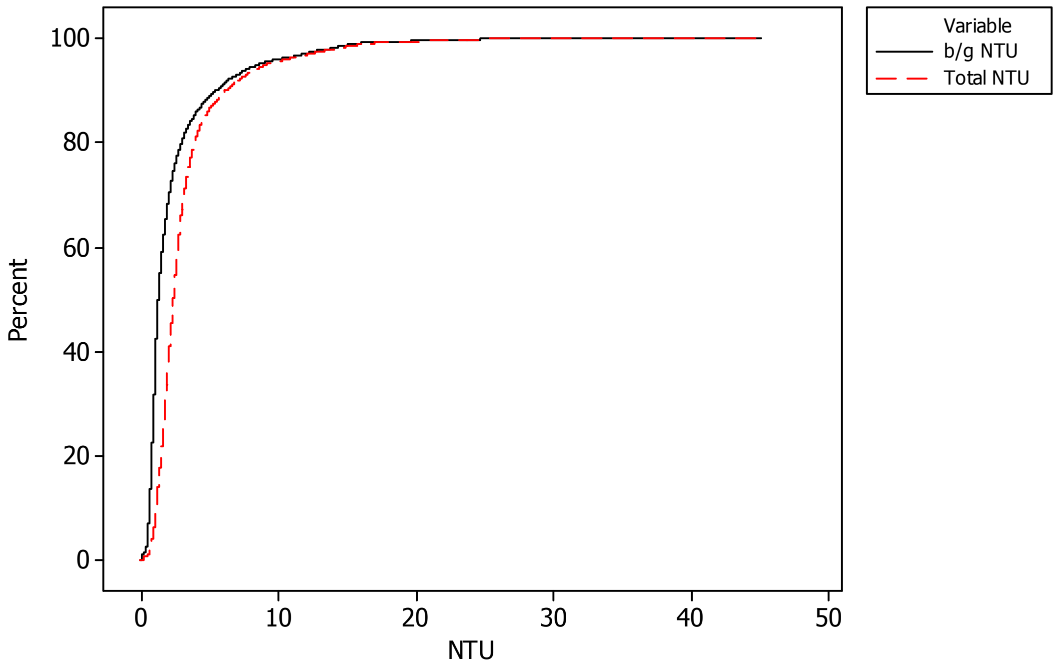

Figure 10. Emprical cdfs for background turbidity (black solid line) and total turbidity (red broken line).

The environmental limit for Hobson's Bay has been set by PoMC at 70 *NTU*[6]. Based on an analysis of existing records, it is exceedingly unlikely that a value as high as this will ever be observed at this site. It is estimated that after sampling continuously every 15 minutes for 5 months there would be only one such reading. Given that the *EWMA* is a smoothing procedure, it is even less likely that the *EWMA* statistic would ever reach this level[7]. To provide an early warning capability, two 'response levels' or triggers are proposed[8]. The first trigger level is set equal to the 99.9th. percentile of the *EWMA* and the second trigger level is set equal to the 99.99th. *EWMA* percentile. Figure 11 shows the 6-hourly *EWMA* chart for Hobson Bay (total) *NTU* during dredging.

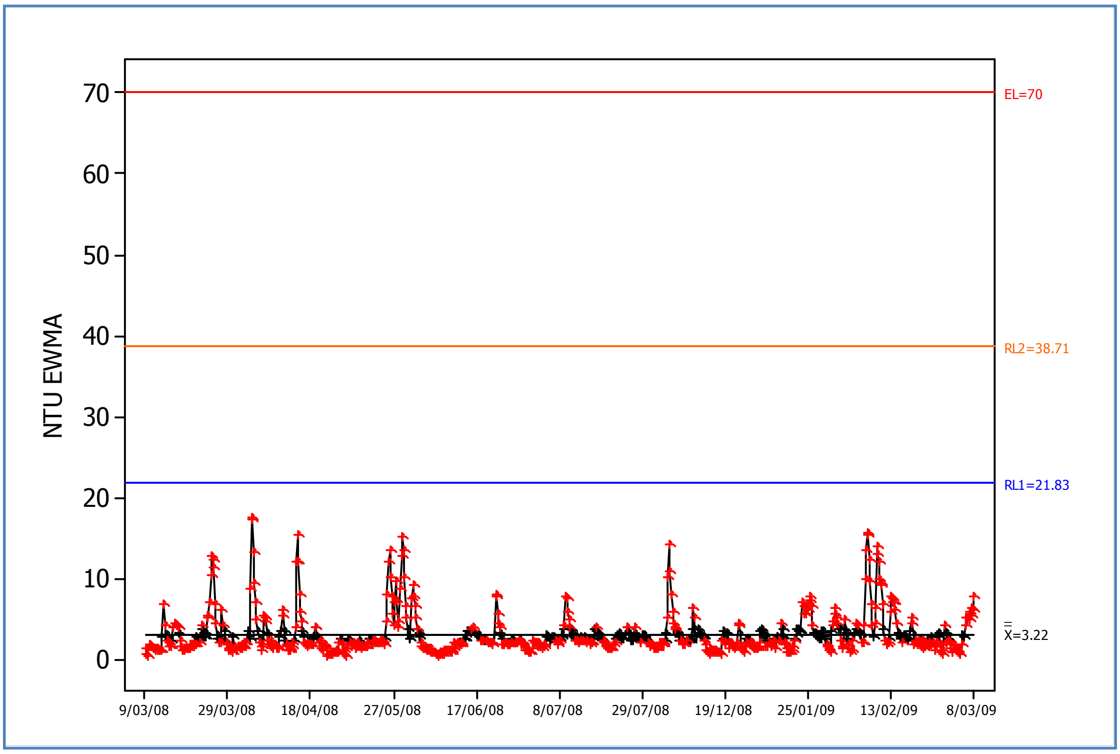

Figure 11. Illustrative example of EWMA control chart for Hobson's Bay turbidity data. $\lambda = 0.6$

[6] PoMC define an environmental limit as the numerical performance standard within which the project must remain.
[7] It turns out that for this example, the probability that the *EWMA* exceeds 41.34 is the same as the probability that an individual, 15-minute *NTU* reading exceeds 70.
[8] PoMC define a response level as a trigger that provides an early warning to enable management action to be taken in order to remain within the environmental limit.

4.2 Illustrative example: Site 2006

Seagrasses in the vicinity of site 2006 in the south-east corner of the Bay have been identified as vulnerable to elevated turbidity levels during dredging. The Port of Melbourne Corporation has established a 25 *NTU* environmental limit for this site. Our companion report "*Statistical Aspects of Turbidity Monitoring: Setting Environmental Limits"* (Environmetrics Australia 2007a) found that, on average, the minimum light criterion for seagrass will be met without the need for any operational intervention to limit turbidity. However, conditions on time-scales of one to two weeks are expected to be sufficiently poor that this light requirement may *not* be met over these shorter periods. In this section we illustrate the use of control charting as an operational management tool and demonstrate how it can be 'tuned' to adjust the level of responsiveness to current conditions. For the purpose of illustration, we have assumed that the dredging operator has decided to monitor conditions using a 15*NTU* reference level so as to provide an 'early warning' that the 25*NTU* environmental limit may be exceeded.

Figure 12 shows (on a logarithmic scale) the histogram for total NTU (background plus predicted dredging) at site 2006. This distribution has been successfully modelled using a three-component mixture of normal distributions with parameters given in table 3.

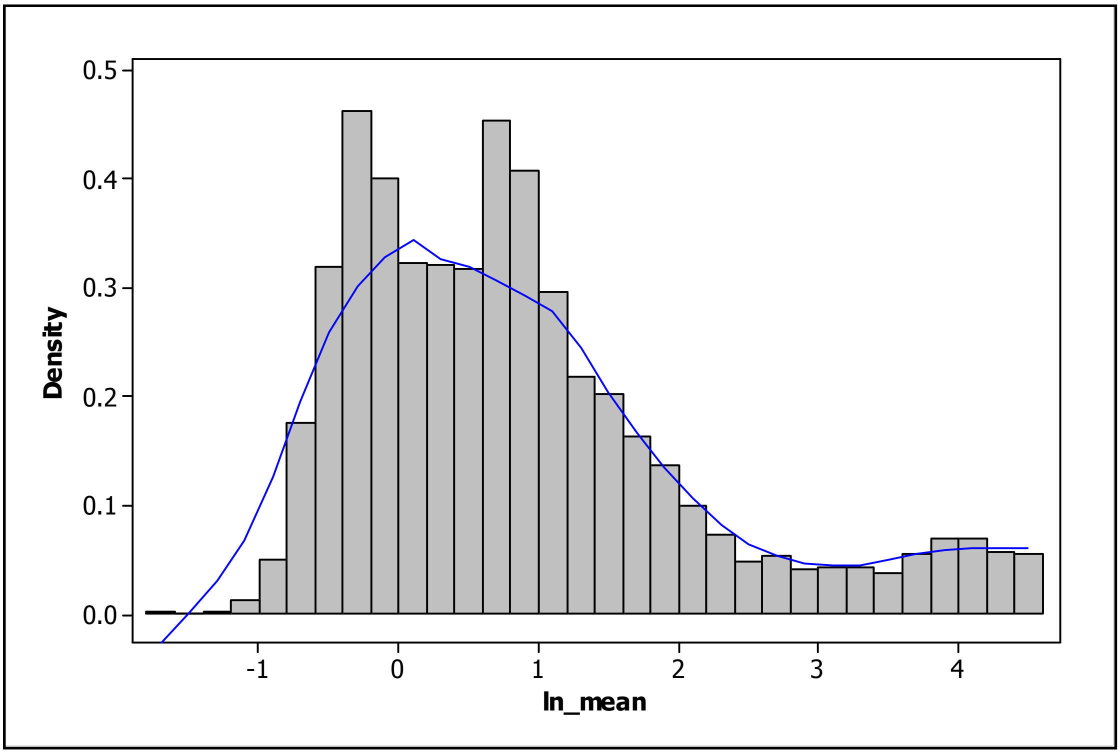

Figure 12. Empirical distribution for *ln(*total NTU) (modelled + background). Blue line is smooth version.

Table 3. Parameters for 3-component mixture. P-values are mixing proportions.

Parameter	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	p_1	p_2	p_3
Value	0.15461	1.4468	3.8463	0.57814	0.70815	0.49417	0.61229	0.30101	0.08670

The resulting fit using the parameters in table 3 is shown in Figure 13. To assess the adequacy of this fit, the empirical *cdf* and theoretical *cdf* were compared (Figure 14). It is clear from Figure 14 that the fitted distribution provides an exceedingly good fit across the entire range of *ln*(*NTU*) values.

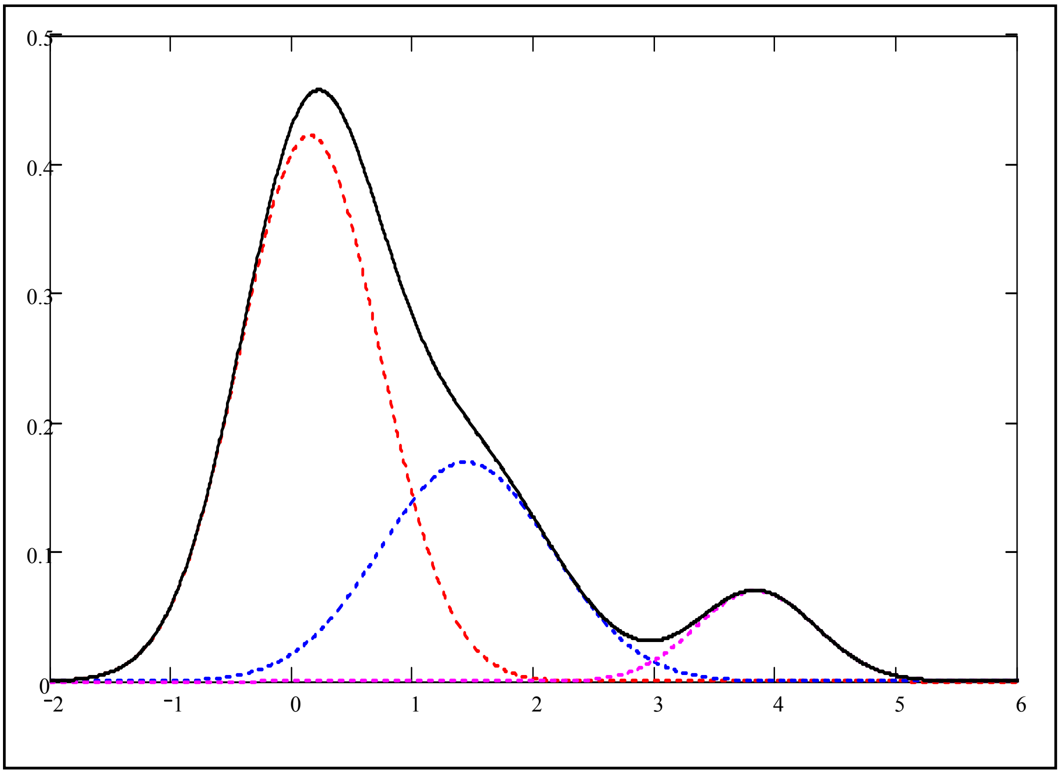

Figure 13. 3-component normal distributions (broken curves) and overall fit (solid curve) for *ln*(*NTU*). Parameter values given in Table 1.

Figure 15 shows the time series plot for predicted *total NTU* during dredging. Although it can be established that, in the absence of any intervention, the 15 *NTU* 'trigger' is expected to be exceeded approximately 10% of the time overall (Figure 13) the minimum light requirement at 3m will still be met at least 50% of the time – having due regard for the variability in background and incremental turbidity as well as the inherent uncertainty in the relationships between *TSS* and *NTU* and *TSS* and Kd (extinction coefficient).

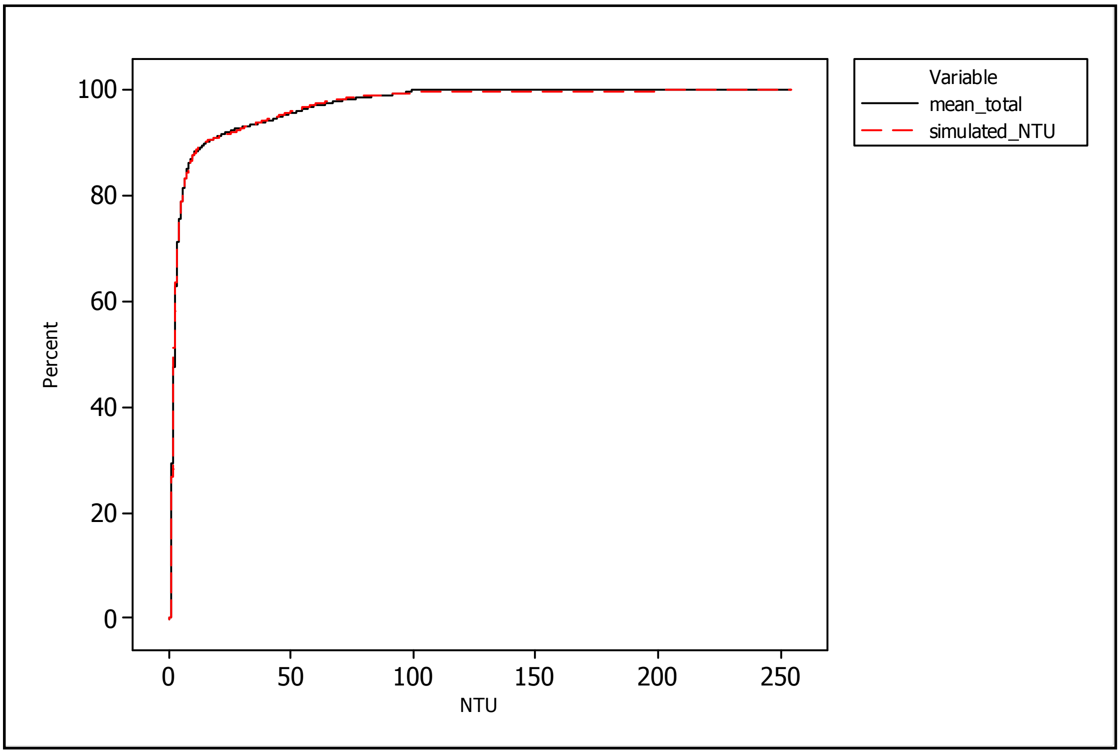

Figure 14. Comparison of empirical *cdf* for *ln*(NTU) (solid black line) and 3-component mixture of normal distributions (broken red line).

Six-hourly *EWMA* control charts have been produced for the *total NTU* using $\lambda = 0.6$ (Figure 16) and $\lambda = 0.2$ (Figure 17). In both cases it is evident that there will be excursions beyond the notional 15 *NTU* 'trigger'. The considerations underpinning the choice of an appropriate choice for λ are discussed elsewhere in this report. However, as can be ascertained from equation 1 and can be seen from Figures 16 and 17, smaller values of λ give less weight to the current data and increase the relative contribution of the past sequence of values whereas the reverse is true for larger values of λ. In practice, the ultimate choice of λ will be a compromise between the level of responsiveness to current conditions versus the level of false or 'inappropriate' triggering.

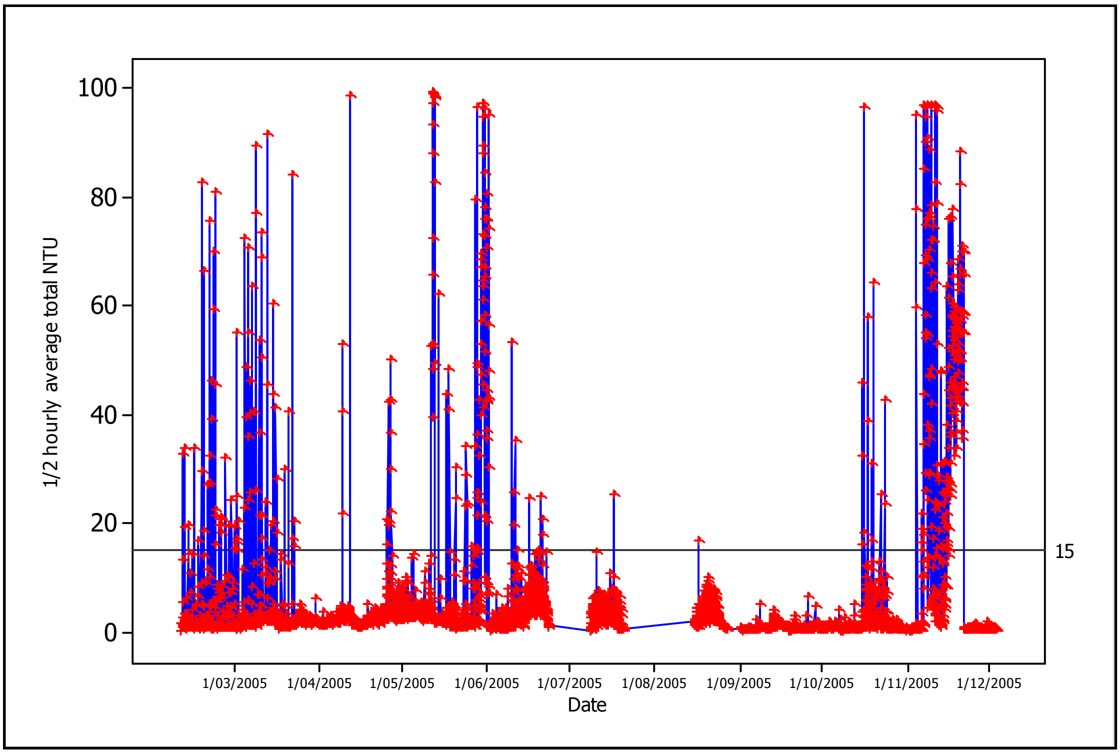

Figure 15. Time-series plot of total *NTU* for the period 9-Feb-2005 to 3-Dec-2005.

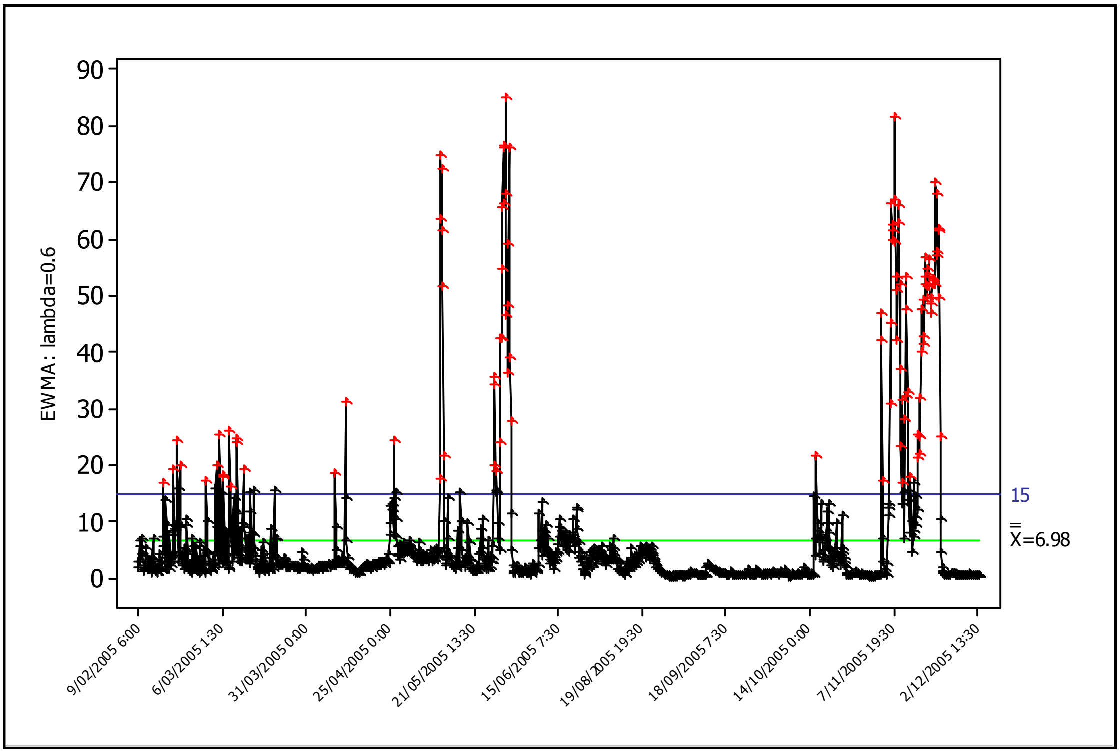

Figure 16. 6-hourly EWMA chart for total *NTU* with lambda=0.6.

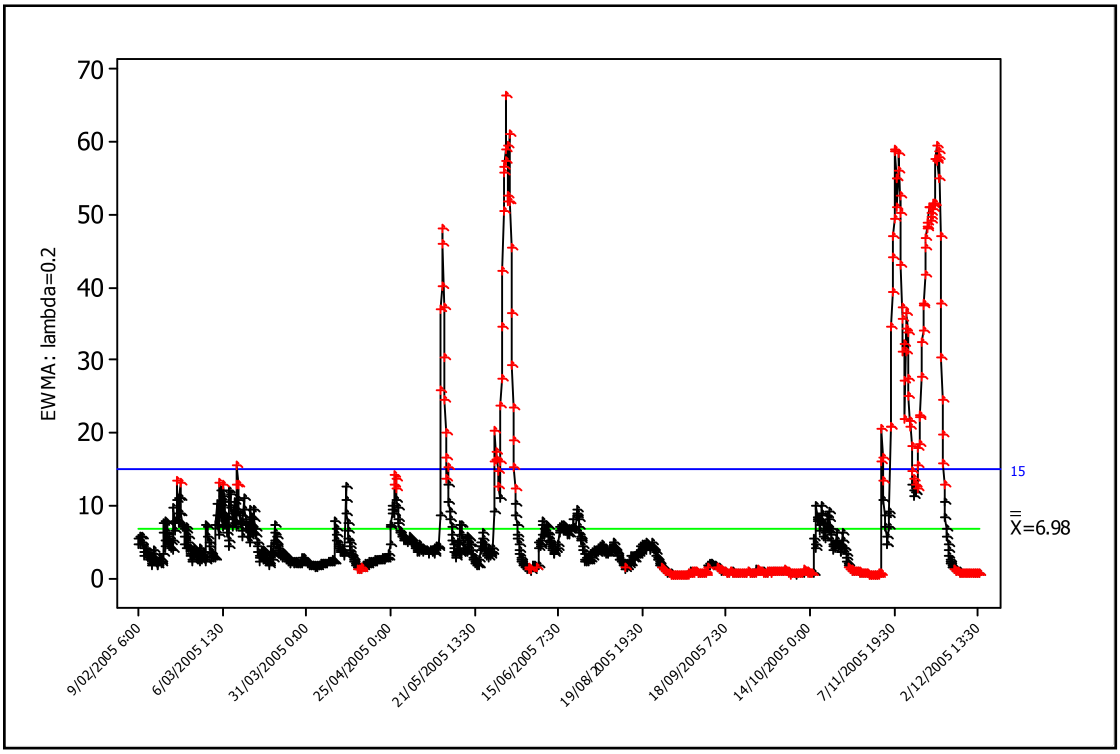

Figure 17. 6-hourly EWMA chart for total *NTU* with lambda=0.2.

5. ADVANCE WARNING SYSTEM[9]

One of the drawbacks of any monitoring program that has a 'compliance' assessment component is that the proclamation of in or out of compliance can made only after the very last observation for the relevant reporting period is available. Devices such as time-series plots and control charts are certainly useful and provide visual cues as to emerging trends, however they do not admit any quantitative assessment of the likelihood that the final compliance target will be met at any interim point in the monitoring period. Although further research and development is required, we illustrate here a potential 'early warning' capability for turbidity monitoring[10]. As stated in the previous section, we are particularly interested in maintaining an *NTU* below 15 units. An analysis of the combined background and predicted *NTU* for site 2006 suggests that this will occur 90.13% of the time. Clearly any increase in turbidity above what is predicted will result in this proportion being something less than 90.13%. Figure 18 illustrates the situation. At any point in time,

[9] The procedures and methods of this section should be regarded as 'experimental' as of the date of this report and subject to further evaluation and possible refinement/modification.

the *current* estimate of the proportion of time that *NTU* is less than 15 is shown by the green line in Figure 18. It is seen that this estimate fluctuates around the notional target of 0.9013 and by the end of the monitoring period it has converged to what we expect under predicted/modelled conditions (ie. proportion of times *NTU* <15 is 90.13%). The red line in the figure provides, at any point in time, the likelihood that the assumed target (=0.9013 in this case) will be achieved by the *end* of the monitoring period. Thus, the red line provides a ‘heads-up’ without having to wait until the very last *NTU* has been recorded. It will be noticed that the red line terminates before the end of the series. This is because some erratic oscillations are possible towards the end of the monitoring period. An arbitrary ‘trigger’ level of 0.1 has been set as a warning/action limit on the red curve (this is shown as a purple horizontal line in the figures). The red line in Figure 18 is always above the trigger value. Figure 19 is essentially the same except the series does *not* converge to the assumed target of 0.9013 by the end of the monitoring period (it’s final value is somewhat less at 0.8962 and corresponds to an overall 10% increase in predicted *NTU*). In the early stages of monitoring the red curve suggests that there is a reasonable likelihood that the 0.9013 target will be attained by the end of the monitoring. However, there is a period during which the red curve approaches the trigger and it is at this point that a decision would need to be made about possible interventions. As it turns out, the red line recovers to very high values in the latter stages of monitoring. This ambiguous case arises because the final result of 0.8962 is not too far removed from the target of 0.9013. A more distinct situation is depicted in Figure 20 where the final proportion is 0.8528 corresponding to a 100% increase in predicted *NTU*. We see that the prospect that a 0.9013 final result being achieved is flagged as almost impossible at an early stage (about 1/10th. of the way through the monitoring period).

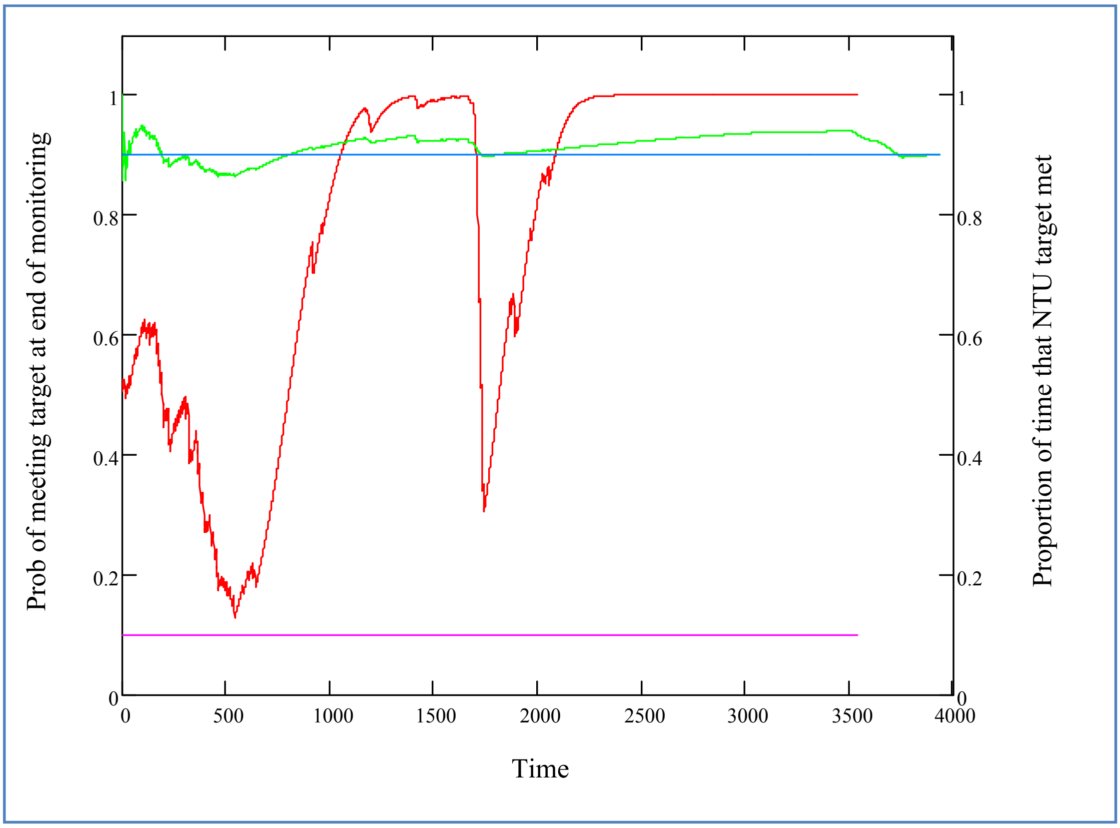

Figure 18. Proportion of time NTU<15 (green curve) and probability that target proportion of 0.8231 (blue horizontal line) will be met at the end of the monitoring period (red curve) for "in-control" situation at site 2006.

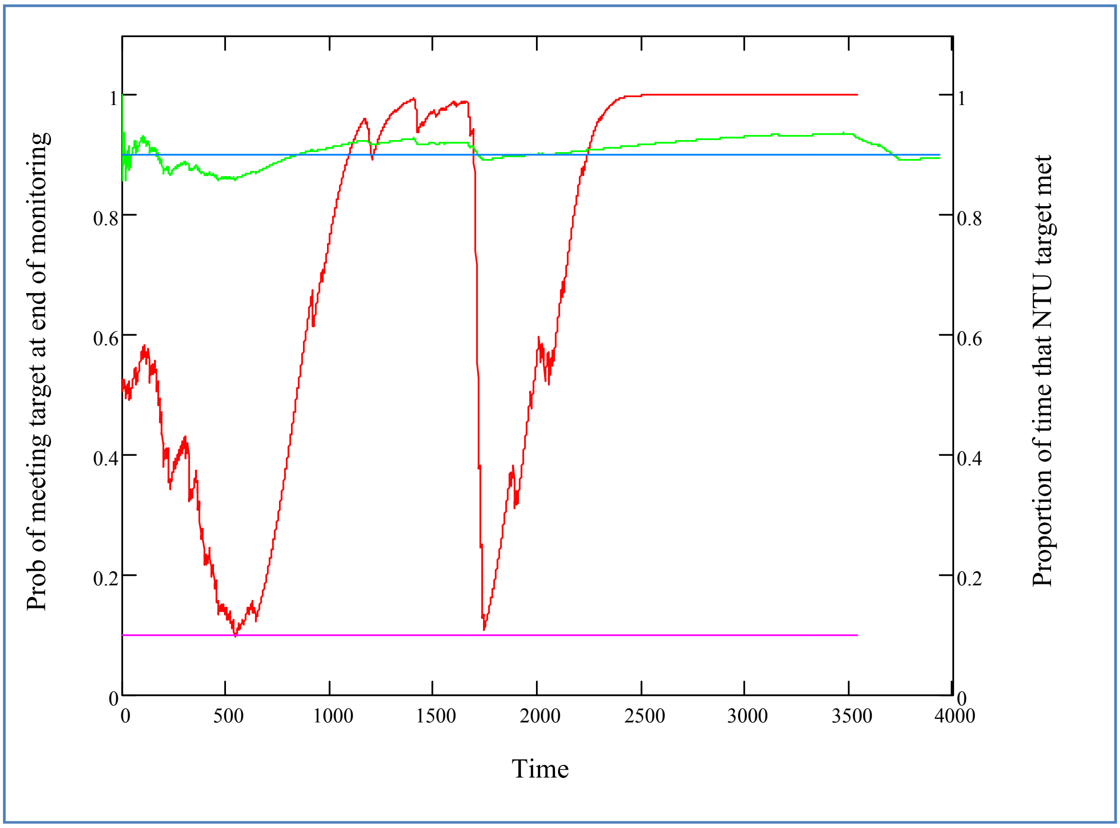

Figure 19. As for Figure 17 except final proportion (0.8962) slightly below target of 0.9013.

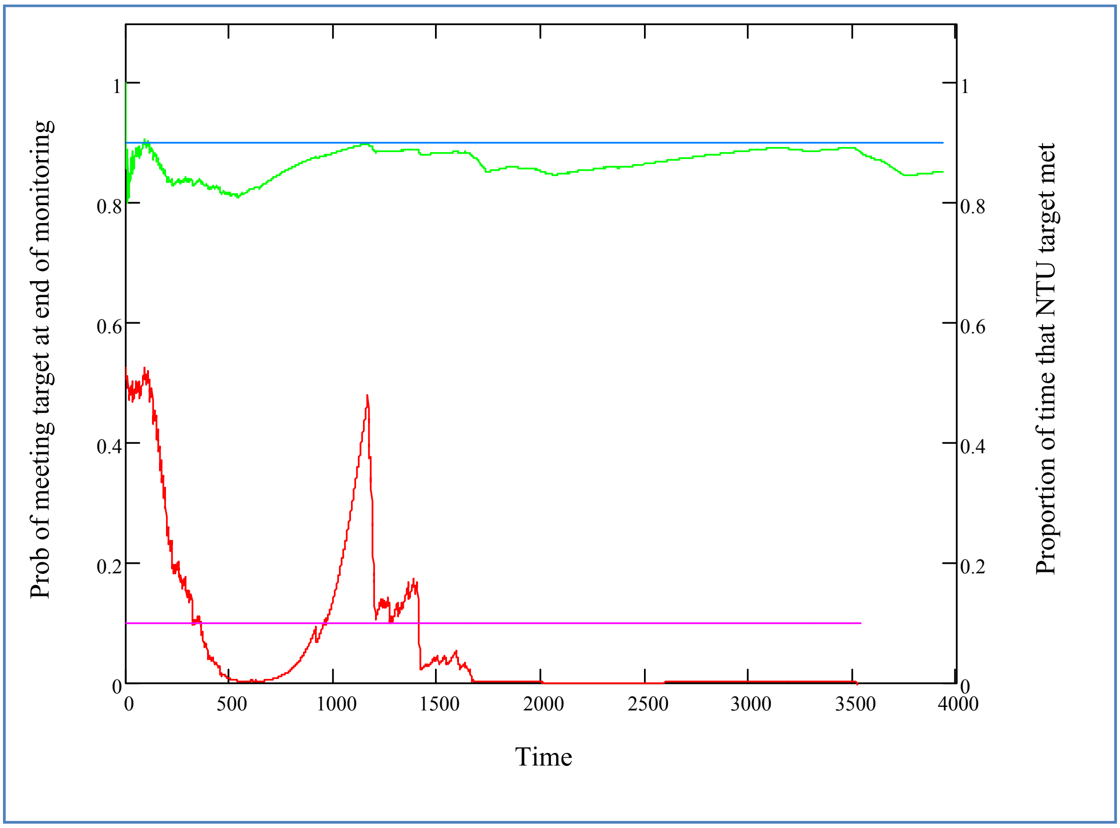

Figure 20. As per Figure 17 except final proportion (0.8528) significantly below target of 0.9013.

6. SAMPLE SIZE AND SPATIAL-TEMPORAL CONSIDERATIONS

In this section we investigate spatial properties of turbidity monitoring and allied issues of replication, sample size, and statistical power. While these are important considerations, an over-emphasis on the quantification of power for example, is often-times unwarranted in environmental monitoring contexts and can result in inefficient (or inappropriate) sampling designs if more fundamental (and important) aspects of spatial dependency are not adequately addressed. The monitoring program proposed in this report is underpinned by the following considerations:

- It should be fit-for-purpose;
- Its performance should be immune to moderate perturbations to assumed conditions;
- It yields timely information on both operational and environmental conditions and performance.

To be clear: the objective of the CDP monitoring program is to provide informed and timely environmental assessments; it is not about testing research hypotheses.

6.1 Spatial characteristics

A complete assessment and characterisation of the spatial correlation structure of turbidity monitoring results is beyond the scope of this report. Nevertheless we have undertaken a number of preliminary analyses at sites in the Southern Channel to provide an indication of the spatial characteristics of turbidity patterns. Approximately 150 (average) daily *NTU* values were obtained from six sites identified as: Capel Sound; Hovell Pile; Mid Ground Shelf ; South Channel Pile; Spoil Ground; and Shoal. A simple but effective visual tool for identifying dependencies among multivariate data sets is the so-called Andrews plot developed by D.F. Andrews (1972). The basic idea is to use a Fourier transform of a *p*-dimensional vector of observations into *p* separate series as follows:

$$f_{\underline{X}}(t) = \frac{X_1}{\sqrt{2}} + X_2 \sin t + X_3 \cos t + X_4 \sin 2t + X_5 \cos 2t + \dots \quad (1)$$

A summary of the features and use of this plot can be found in Fienberg (1979). In short, a high degree of coherence in the resulting series is indicative of dependency between the p components. The Andrews plot for the six sites in the Southern Channel is shown in Figure 21. Figure 21 suggests that overall the six channel sites behave similarly with respect to turbidity. This observation is further supported by the high correlations between pairs of sites (Table 3).

Table 3. Pearson correlation coefficients for daily NTU among pairs of sites in the Southern Channel (first ntry) and p-value (second entry).

	Capel Sound	Hovell Pile	Mid Grnd Shelf	South Chan Pile	Spoil Grn Shoal
Hovell Pile	0.897 0.000				
Mid Grnd Shelf	0.705 0.000	0.783 0.000			
South Chan Pile	0.692 0.000	0.737 0.000	0.539 0.000		
Spoil Grn Shoal	0.856 0.000	0.845 0.000	0.615 0.000	0.945 0.000	
South Hovell	0.860 0.000	0.860 0.000	0.509 0.000	0.796 0.000	0.902 0.000

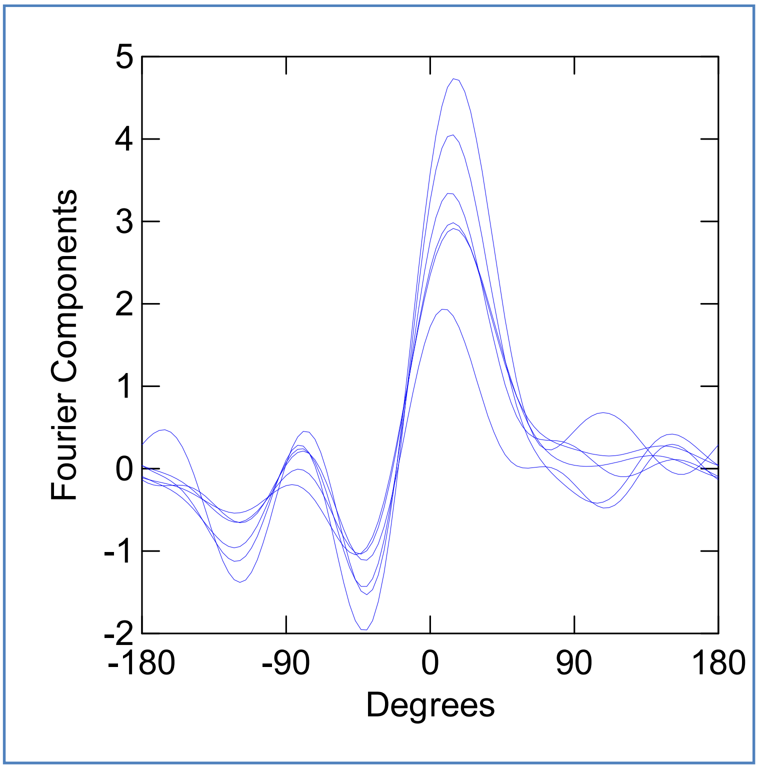

Figure 21. Andrews Fourier plot of daily NTU for six sites in the Southern Channel.

Another (statistical) tool for quantifying spatial dependency is the (semi) variogram. More precisely, the variogram is a measure of spatial dissimilarity as a function of separation. The variogram can be computed for each day for which data is

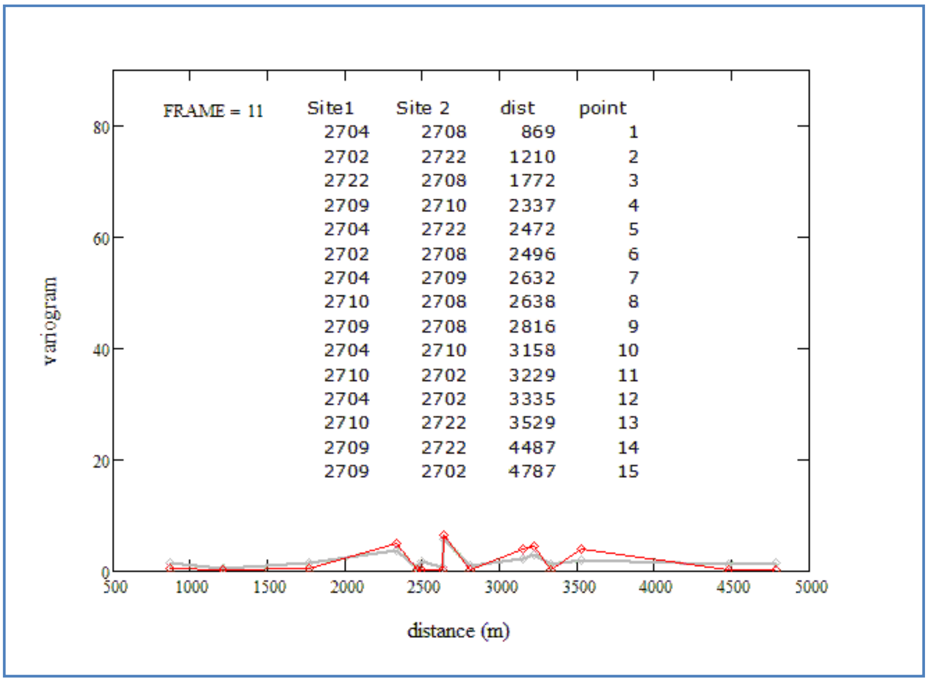

Figure 22. Sample variogram for NTU in Southern Channel on day 11 (red line) and for entire 150-day series (grey line).

available. Two such examples are provided in Figures 22 and 23.

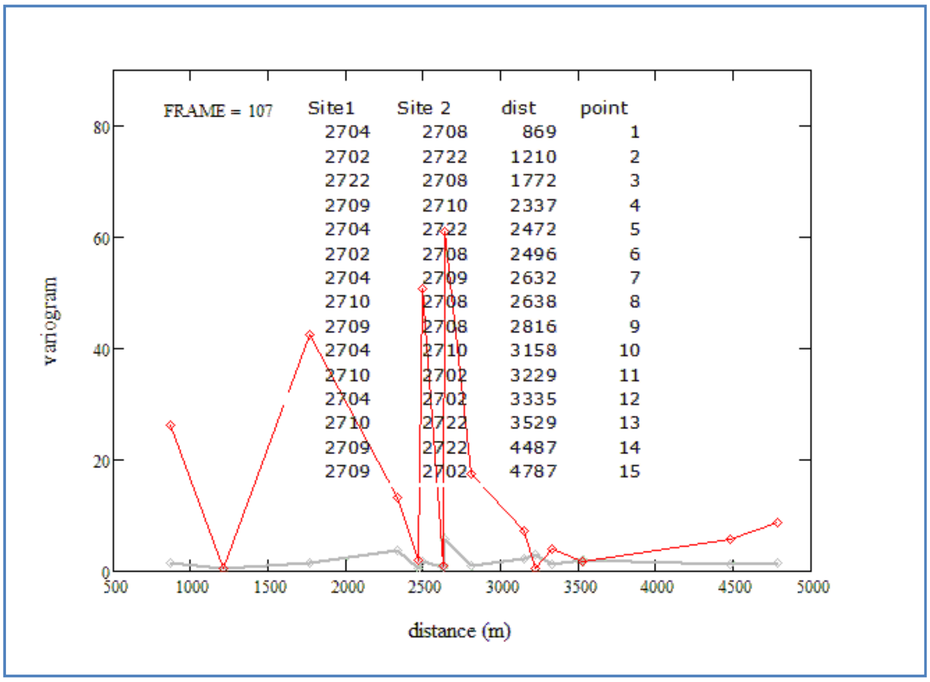

Figure 23. Sample variogram for NTU in Southern Channel on day 107 (red line) and for entire 150-day series (grey line).

Figure 22 shows that on this particular day, there is virtually no dissimilarity in *NTU* across the range of sites while Figure 23 shows an instance where may be some (relatively small) difference in turbidity between sites that are separated by no more than about 3km. Taken together, the results presented in Table 3 and Figures 21 to 23 suggest that there is a high level of spatial contunuity in turbidity readings on scales extending to 5km. While daily fluctuations are possible, turbidy on these scales is relatively uniform as depicted in Figure 24.

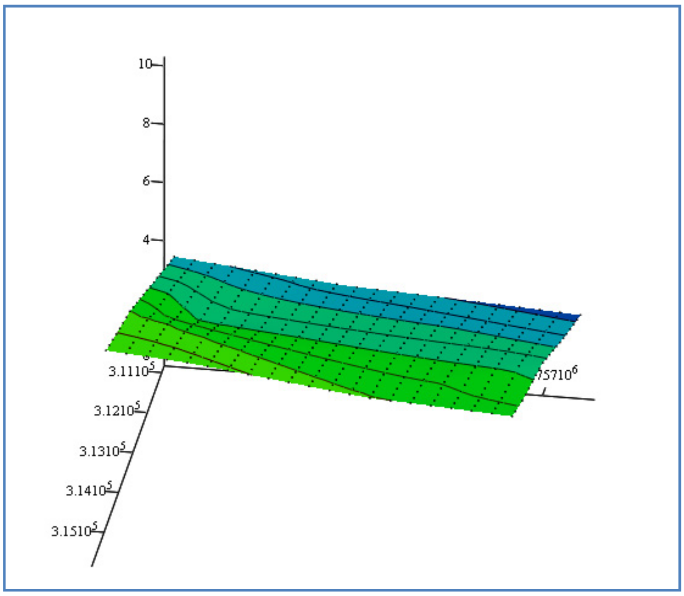

Figure 24. Indicative spatial pattern of turbidity in Southern Channel.

The implication of these reults is that, from a statistical perspective [11]a spatially dense (on sub-5km scales) turbidity monitoring network is not warranted.

6.2 Temporal characteristics

Turbidity data from in-situ loggers will typically be recorded once every 15 minutes and as such a high degree of autocorrelation is expected between successive *NTU* readings. A plot of the sample *ACF* (autocorrelation function) for Cameron's Bight turbidity data is shown in Figure 25. We note that the ACF shows an exponential decay out to a lag of about 24-28. In the parlance of time-series analysis, this phenomenon is characteristic of an *ARIMA(1,1)* process[12]. Thus, from a practical perspective, *NTU* readings need to be separated by at least 6 hours in order to be considered (statistically) independent. This time-frame is consistent with the 6-hourly period used for control-charts of section 3.

[11] There may be environmental, operational, or other reasons why more intense spatial coverage for turbidity monitoring is required.
[12] Autoregressive integrated moving average model with first-order autoregressive and moving average components.

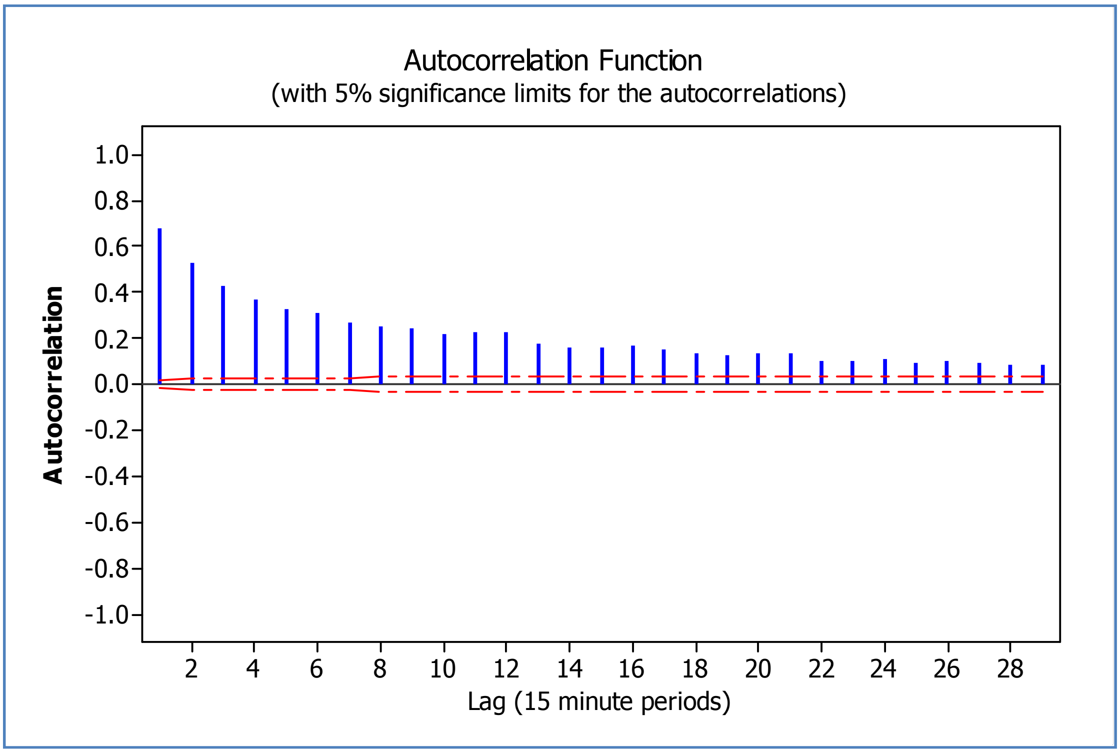

Figure 25. Autocorrelation function for 15 minute NTU data at Cameron's Bight

6.3 Power and sample-size

Our position on power and sample-size calculations for environmental monitoring studies was stated at the beginning of this section. Given the pervasiveness of these analyses and the regulatory focus on this aspect of the montoring design, we have undertaken some indicative power and sample-size calculations for Hovell Pile turbidity data. While formulae for power and sample-size calculations can be found in most statistics textbooks, what is often overlooked or ignored is that these are predicated on the assumption that the data at hand are normally distributed. Violations of the normality assumption produce results that are meaningless. An examination of the histogram of Figure 4 shows that this is likely to be the case if unmodified *NTU* data are used for such calculations. Figure 6 shows that a greater degree of normality is restored if we work with *log*-transformed *NTU* data. The results of a slightly more sophisticated transformation (the Johnson transformation) are shown in Figure 26. The 'optimal' transformation suggested by this procedure is given by equation 2 where Y denotes a 'raw' *NTU* value and X' denotes the transformed value.

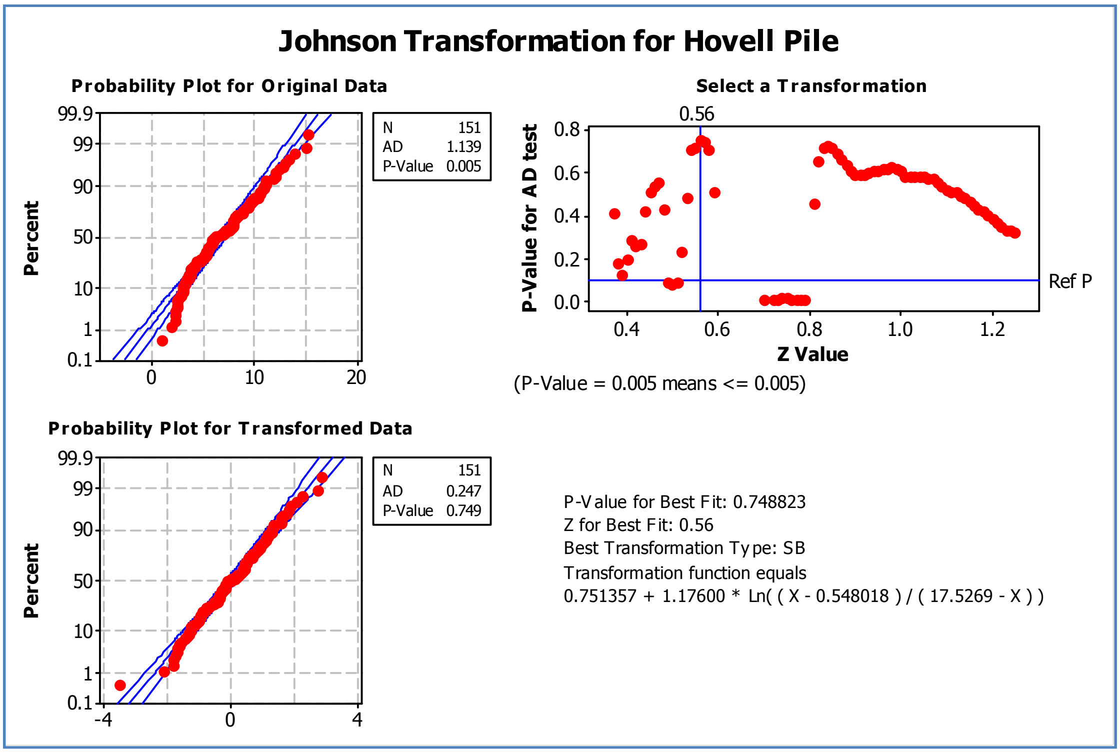

Figure 26. Identification of an appropriate transformation of daily NTU data for Hovell Pile. Normal probability plots for the untransformed data (top left panel) and transformed data (bottom left panel) are shown together with information on the best-fitting Johnson transformation (right panel).

$$Y = 0.751357 + 1.176\ln\left(\frac{X - 0.548018}{17.5269 - X}\right) \quad (2)$$

A plot of the relationship between Y and X is shown in Figure 27.

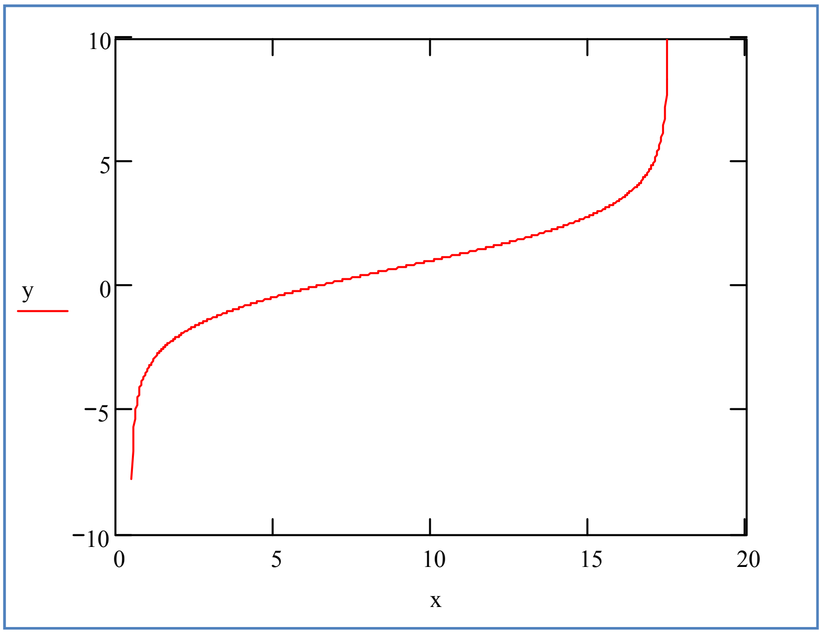

Figure 27. Plot of transformation function (equation 2) for Hovell Pile daily NTU.

We note that equation 2 is a 1-to-one transformation and thus power calculations made with reference to the (normally distributed) Y variable can be translated back to equivalent statements about X. The details of this process are not given here, instead we provide an illustrative example of the results (Figure 28).

Figure 28 shows for example, that in order to detect an approximate 50% increase in the mean *NTU* at Hovell Pile with 80% power using a one-sided t-test at the $\alpha = 0.05$ level, about 10 (independent) samples would be required. Given that a minimum time-lag of about 6 hours is required in order to treat successive observations as 'independent' this means that this assessment would require almost three days of monitoring. This analysis should be regarded as indicative only – other factors such as the dynamic nature of sediment plumes and the constantly changing nature of suspended sediment concentrations means that rather simplistic analyses such as these are of limited value. Other difficulties in the use of power and sample-size analyses in environmental settings are discussed in Fox (2001) and Fox et al. (2007).

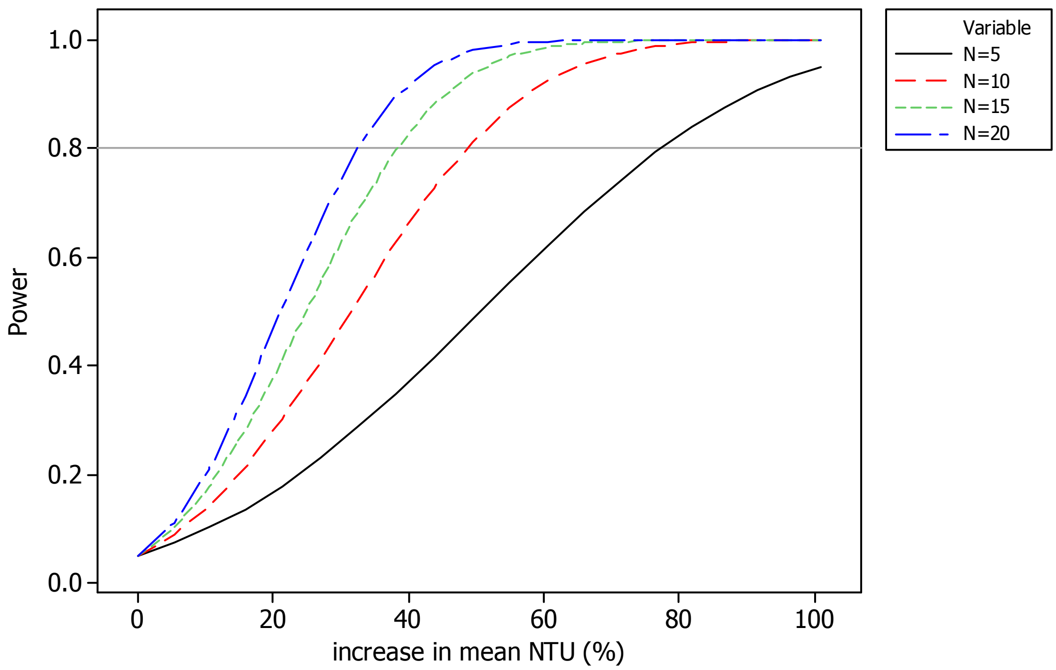

Figure 28. Power curves for various sample sizes for monitoring NTU at Hovell Pile. One-sided alternative hypothesis and 0.05 level of significance assumed.

7. DISCUSSION-RECOMENDATIONS

In this report we have examined a number of fundamental statistical issues associated with the development of a robust, credible, and (statistically) defensible monitoring design. It is recognised however, that there are other *non-statistical* factors and considerations that need to be evaluated and assessed in order to arrive at a final design that achives its stated purpose in a cost-effective and reliable way. The purpose of this report is to outline, by way of example, how important considerations of spatial-temporal variability, power, sample size, sampling frequency, and control-chart parameters influence the monitoring design. It should however, be recognised that further work will ineviatably be required to adpat and modify the suggested approach to provide greater spatial coverage of the Bay. We also point out that we are unaware of previous uses of a 'real-time' control-charting approach to environmental monitoring and further adjustment and 'fine-tuning' will be necessary before the system is fully operational. In the following sections we discuss some of the important issues that have been raised in this report and which may require further evaluation prior to implementation.

7.1 Choice of monitoring tool

There are numerous options for gathering, presenting, and interpreting monitoring data. We have stressed the need for clarity of purpose and have noted a fundamental distinction in the needs and requirements of data collection and analysis techniques in support of testing a research hypothesis with those required to obtain quick and reliable summaries of current environmental conditions and emerging trends. The use of CDP monitoring data to underpin *operational* decisions that aim to ensure compliance with environmental criteria means that the monitoring tools and procedures need to be flexible, adaptive, and rapid. Visual aids are ideally suited to this purpose and control-charts provide additional enhancements that enable them to be used in a *pro-active* fashion. In addition, control-charts are advocated in the National Water Quality Monitoring Strategy (ANZECC/ARMCANZ 2000) for monitoring water quality. This report has advocated the use of the Exponetially Weighted Moving Average (*EWMA*) control chart as the cornerstone of the turbidity monitoring program. This is not to say that other, equally useful devices (such as *CUSUM* and *Xbar* charts) are not applicable. The choice of the *EWMA* chart has been influenced by a number of important features. These include:

- The ability to track / monitor the evolving series over time and to gain a quick, visual appreciation of emerging problems, trends and performance against environmental criterion;
- The inclusion of action limits or ‘response-levels’ that provide triggers for intervention or a management response *before* an environmentally unacceptable situation has occurred;
- The inclusion of both current and previous monitoring data;
- The ability to ‘tune’ the performance characteristics to weight the relative contributions of current and (immediate) past conditions;
- Ease of use and interpretation.

7.2 Choice of control-chart parameters

It is important to note that the use of control-charts for environmental monitoring is relatively new and so experience with them in this context is somewhat limited. Also, as mentioned in section 3, control-charts were originally developed to monitor manufacturing processes where data on say, tolerances of manufactured items was more likely to be normally distributed than environmental data. Thus, the 'default' control limits of $\pm 2\sigma$ for 'warning' limits and $\pm 3\sigma$ for 'action' limits are not necessarily the most appropriate in an environmental setting. For example, the adoption of $\pm 2\sigma$ warning limits implies there is a 5% chance that a warning limit will triggered even when the 'process' is in control. When applied to a 6-hourly control chart over a six-month period this translates to 36 false warnings. Given the cost and inconvenience associated with action based on false alarms, this default alarm level is possibly inappropriate. We advocate an approach that integrates professional judgement (based on an interpretation of the monitored data) supported by numerical triggers that have low rates of false triggering. As a guide, we have suggested the use of the 99.9th. percentile (of the predicted total *NTU* data) as a first-level response (RL1) and the 99.99th. percentile as a second-level response (RL2). The expected number of false-alarms under the conditions just described is about 1 at RL1. What constitutes a 'response' is a matter for the CDP stakeholders and proponents. As a guide, and consistent with ANZECC/ARMCANZ (2000), the first-level response would typically be an investigation or follow-up into the nature and causes for the trigger. If these investigations reveal that the triggering of RL1 was not a false-alarm (as could be caused by instrument malfunction or conditions unrelated to dredging) a prudent response would place increased emphasis and attention on individual (half-hourly) *NTU* readings. An appropriate response to the tripping of RL2 would be the suspension of dredging operations. However, the final limits and associated actions adopted may need to accommodate other considerations such as the level of environmental protection afforded under various scenarios of unanticipated turbidity elevations.

Another important consideration in the design of the *EWMA* control chart is the choice of the parameter λ. While there are no hard and fast rules, it is suggested that the final choice be influenced by the nature of the environmental asset being protected and the required level of chart responsiveness to *current* conditions. Thus, for example, greater responsiveness to current conditions might be appropriate for seagrasses than fish (given the latter have an ability to avoid unfavourable conditions) in which case $\lambda = 0.6$ would be preferred over $\lambda = 0.2$. Such considerations need to undertaken on a site-by-site basis in which case it is entirely possible that a common λ would not be appropriate.

7.3 Spatial replication

A limited investigation into the spatial variation in turbidity data suggests that, overall, *NTU* data are likely to exhibit a high degree of concordance on spatial scales of less than five kilometres[13]. This suggests that increasing the number of sites within a five kilometre radius around an asset of significance (ostensibly on the basis of improving statistical power) is of limited value. A preferred strategy is to apply the resources that would otherwise be devoted to such small-scale replication to deployments on larger spatial scales. Given this preference (and for reasons given in the next section) we have avoided recommending a fixed number of replicate monitoring locations at each site.

7.4 Number of sites (power and sample size issues)

A small, but nevertheless indicative, power and sample size analysis associated with the *NTU* monitoring data has been undertaken. A number of issues were identified in section 6.3. In particular, the highly non-normal nature of *NTU* data means that, in their 'raw' form, they are not amenable to conventional power and sample size calculations. Furthermore, given the dynamic nature of sediment plumes, it is unlikely that other attendant assumptions of independence (in time and space) and constant variance would be met. Thus, in the absence of any remedial action, the results from any such analysis are likely to be erroneous at best, and possibly fatally flawed at worst rendering any reliance upon them unsound. Notwithstanding this concern, we have argued that the present monitoring program is focussed on making rapid and reliable assessments of environmental condition rather than testing an hypothesis about the true mean turbidity[14] (which is implicit in power and sample size calculations).

[13] This is a general statement and does not rule out the occurrence of more localised and transient effects.
[14] At a *particular* time and place.

7.5 Data integrity issues

Based on the analysis of the background monitoring data supplied to Environmentrics Australia, it is apparent that the filed instruments have a number of limitations that result in both censored and erroneous data (eg. negative *NTU* readings). The former is an artefact of the instrument settings and/or limitations which means that *NTUs* above 96 cannot be measured. While this presented no particular problems for the analysis of the background data (since, with the exception of some transient 'spikes', background levels were well below this threshold) it is a potentially significant issue for turbidity monitoring during dredging given that the measured *NTU* will, on occasions exceed 96. In such instances, it would be most inappropriate to use instruments which cannot reliably obtain reliable data over the entire range of expected *NTU* values. Simply truncating the data and reporting a limit of detection would lead to a potentially serious under-reporting of the true turbidity levels. Similar bias will be introduced if adequate (statistical) QA/QC procedures are not implemented which will detect and screen for anomalous and aberrant data as they are recorded. Negative *NTU* readings should be replaced by zero values so as to minimise the effects of negative bias on computed statistics. Guidelines are available for the treatment of outliers and other aberrant observations in the Water Quality Guidelines (ANZECC/ARMCANZ 2000).

Another data integrity issue concerns gaps in the data record. This is most likely to arise from instrument failure. There are no hard and fast rules as to how to treat gaps in a time-series record although there are essentially two choices:

- Reporting as 'missing' or 'NA';
- Data imputation – inferring the missing values by various interpolatory techniques (linear interpolation, cubic splines etc.) or modelling approaches (eg. time-series models).

The second option is preferred and as a guideline, is the recommended approach when no more than approximately 20% of data is missing[15]. The 20% guideline should be adopted for the 6-hourly *EWMA* chart. Thus, assuming 'raw' data is acquired every 15 minutes, then interpolation should not be attempted if more than 4 readings are missing.

[15] This applies to data missing at random and not systematic losses (eg. all data missing for a particular site or time).

8. REFERENCES

Anderson, D.F. (1972) Plots of high dimensional data. *Biometrics*, **28**, 125-136.

ANZECC/ARMCANZ (2000) Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Volume 1, Paper No. 4 October 2000. Australian and New Zealand Environment and Conservation Council (ANZECC) and Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ).

Baker, E.T. and .Lavelle, J.W. (1984) The effect of particle size on the light attenuation coefficient of natural suspensions. *Journal of Geophysical Research,* **89(C12),** 8197-8204.

Environmetrics Australia (2007a) An Examination of Regression Options for TSS and NTU. Report to Port of Melbourne Corporation, January 2007.

Environmetrics Australia (2007b) A probabilistic analysis of the benthic light climate in the Southern Channel during dredging. Rev 01/07, Report to Port of Melbourne Corporation, January 2007.

Feinberg, S.E. (1979) Graphical Methods in Statistics, *The American Statistician*, **33(4)**, 165-178.

Fox, D.R. (2001) Environmental Power Analysis – A New Perspective, *Environmetrics*, **12(5)**, 437-449.

Fox, D.R., Ben-Haim, Y., Hayes, K.R., McCarthy, M.A., Wintle, B., Dunstan, P. (2007) An Info-Gap approach to power and sample size calculations. *Environmetrics*, **18(2)**, 189-203.

Page, E.S. (1954). Continuous Inspection Schemes. *Biometrika*, **41(1),** 100-115.

Roberts, S.W. (1959). Control Chart Tests Based on Geometric Moving Averages. *Technometrics* **42(1),** 97-101.