19 COMMENTARY ON STATISTICS FOR TURBIDITY TRIGGERS

PO Box 7117
Beaumaris VIC, 3193
+61-3-9018-7121
www.environmetrics.net.au

An Environmetrics Australia Technical Report

Statistical Considerations Associated with the Establishment of Turbidity Triggers

Candidate Methodologies for Large-Scale Dredging Projects

By
David R. Fox,
PhD.,C.Stat.,P.Stat.,C.Sci.

16 September 2016

Limitations Statement

This report documents statistical issues associated with the establishment of turbidity trigger values for a large-scale capital dredging project. Its findings, recommendations, and conclusions are based on desk-top investigations using *indicative* data sets. As such, no claim is made as to the applicability of the approaches to any specific project. 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.

Executive Summary

This report provides a comprehensive review of statistical issues associated with the development of turbidity trigger values for use during large-scale dredging projects. It has been commissioned by the Lyttelton Port Company's (LPC) as part of their Consent Application process which seeks approval to deepen, widen and extend the LPC's existing navigational channel. The collection of activities required to achieve this outcome is known as the Channel Deepening Project (CDP).

The CDP involves a large-scale dredging program in order to remove a total of approximately 18 million cubic metres of sediment and place it in a 1,250 hectare off-shore disposal site. This activity has the potential to adversely impact the marine ecosystem and controls are therefore required to provide an early warning mechanism of potentially unacceptable water quality.

The use of 'turbidity trigger values' has become *de facto* industry best practice for large-scale dredging projects such as the CDP. Not only is this approach endorsed by the Australian and New Zealand governments, but recent experience (particularly in Australia) with projects of similar scope and objectives has demonstrated the dredging activity can be managed to successful completion without any long-term environmental harm and/or impacts that were not predicted by the environmental impact assessment.

While this experience provides a level of assurance that the use of turbidity trigger values and companion data processing activities will achieve the desired outcome, the science underpinning environmental trigger values is incomplete with issues associated with the treatment of aberrant observations, smoothing techniques, treatment of missing data, and trigger-value computation still requiring further investigation and possible refinement. Furthermore, the recent trend to incorporate the additional dimensions of *frequency* and *duration* (of exceedances) has been guided by flawed research and a lack of 'road-testing' and statistical assurance. This report addresses these issues and provides a comprehensive discussion of the strengths and weaknesses of various approaches as well as offering a number of candidate approaches which we believe are robust, statistically credible, and scientifically defensible.

List of Figures

Table of Contents

1. INTRODUCTION

The purpose of this report is to investigate and recommend appropriate options for: (i) processing 'raw' turbidity data (measured in Nephelometric Turbidity Units or NTUs); and (ii) establish turbidity 'trigger values' or TVs for use during a capital dredging program. The aims of (i) are to identify 'aberrant' readings and 'outliers' and feature/signal extraction while the objective of (ii) is to provide ecologically relevant metrics that provide natural resource managers and port authorities with a tiered, early-warning capability of unacceptable water quality.

For the purpose of this report, an *aberrant* observation is one that is in some way 'unusual' in that it does not appear to belong to or fit the observed pattern of monitored data. We have chosen the word 'aberrant' rather than the more commonly used descriptor 'outlier' to differentiate each case on the basis of 'legitimacy'. An outlier is an aberrant observation with the added property that it is also 'illegitimate' in the context of the statistical characteristics of the variable or process being measured and as such can be considered for *exclusion from the data record*. Outliers are typically the result of transcription errors (eg. wrong placement of the decimal point) in manual data recording and equipment malfunction (eg. a transient voltage spike) in autonomous data recording. An *aberrant* observation on the other hand, is simply 'unusual' but not necessarily erroneous and as such should *not* be automatically flagged for deletion from data record. The distinction and subsequent treatment of outliers and aberrant data is crucial in the context of environmental monitoring and the setting of TVs. Retention of true *outliers* will generally lead to extreme levels of bias and variance inflation in all subsequent statistical processing and modelling of the data. On the other hand, the automatic removal of *aberrant* observations cannot be justified statistically or by a need to 'cleanse' the data or enhance the visual characteristics of some plot of the data. As a simple case in point, statistical theory tells us that there is a 0.27% chance that a normally-distributed random variable will lie beyond 3 standard deviations from the mean value. Thus, if an instrument is continuously sampling at a rate of one reading every minute, then we would expect 3 or 4 such extreme (but nonetheless entirely legitimate) readings every day. So while data values beyond 3 standard deviations of the mean are certainly appear to be unusually high or low, they don't necessarily deserve to be removed from the data record. In any event, we believe the processes to both classify and deal with aberrant data in environmental monitoring programs should be (i) clearly articulated and fully described in

an Environmental Monitoring Program (EMP) document; (ii) be statistically robust; (iii) be scientifically defensible; (iv) be ecologically relevant; and finally (v) be agreed to by a representative cross-section of all stakeholders.

In the context of autonomously acquired turbidity data used to characterise 'background' conditions, the distinction between an *aberrant* reading and *outlier* can have important ramifications. As mentioned above, a turbidity outlier could occur as the result of an equipment malfunction. Other possibilities include fouling of the instrument sensor or the transient obstruction of the sensor due to drift algae or some other material floating in the water column. In such cases it is not uncommon for readings in excess of 2,000 to 3,000 NTU to be recorded – albeit for a short space of time. So, if the ambient turbidity is fluctuating around 20NTU, say, then the retention of even a small number of readings in the thousands of NTUs may seriously bias the statistical summary of 'background' conditions. In the scenarios just described there is no way of knowing what the *correct* reading should have been (unless of course multiple instruments are being used at the same site) and so the data analyst must flag and remove the offending data. This leaves gaps or missing values in the temporal sequence which can create additional problems for the subsequent statistical analysis.

Perhaps more common than turbidity outliers are the occurrences of aberrant turbidity readings. For example, instruments moored in the receiving water body will often 'see' (and hence record) both spatial and temporal 'patchiness' in water clarity. So while the turbidity readings taken when a 'slug' of highly turbid water passes the instrument sensor are entirely legitimate, they are somewhat atypical and as such not representative of the water quality more generally. One way of reducing sampling bias is to increase both the spatial and temporal resolution however this becomes prohibitively expensive.

Atypical conditions can occur on longer time-scales as well. For example, a 12-month background monitoring campaign is desirable since seasonal effects in the turbidity signal are captured. However, the representativeness of the 12-month data record can be significantly compromised by an over-representation of extreme oceanographic and meteorological conditions caused by an abnormally high number of extreme weather events during the monitoring period. Advice on how to best deal with this type of situation does not exist and we have witnessed first-hand the tensions created by a clash of philosophies and approaches. This occurs for example when the Project Proponent

argues for the retention of data collected during the ‘excess’ abnormal weather events while the environmental Regulator insists on the removal of ‘compromised’ data.

Our own view is that if the statistical distribution of turbidity data is significantly (in the statistical sense) altered by an abnormally large number of extreme oceanographic and/or meteorological events, then there is a *prima facie* case for the statistical adjustment of the results in order to better reflect long-term, natural background conditions. It is difficult to be prescriptive about exactly *how* this would be achieved in any given instance, but in general terms the approach would seek to adjust the data through a weighting scheme that de-emphasises the influence of the ‘excess’ storm events.

In the following sections of this report we consider in more detail issues surrounding the pre-processing, analysis, and interpretation of turbidity data before examining candidate procedures and metrics to underpin the triggering process. It is important to appreciate from the outset that this is an imprecise science. Implicit in the use of turbidity to monitor a marine environment for adverse impacts is the strong, but largely untested assumption that the aquatic ecosystem and all that it comprises will be ‘protected’ as long as turbidity is kept below a threshold level. While this assertion has intuitive appeal (for example, we know seagrass need light and turbidity attenuates the photosynthetically active component of light), the *level* or threshold value that achieves the omnibus objective of ecosystem protection more generally is unknown and perhaps unknowable – even if it exists.

Current best practice as articulated in the ANZECC/ARMCANZ (2000a,b) Australian and New Zealand Water Quality Guidelines (and their recent update, Warne et al. 2014) provides an initial starting point but is not prescriptive. While the Guidelines outline a framework for water quality assessments, they acknowledge the need and indeed advocate the use of locally-derived procedures and metrics that are best suited to the specific environment and circumstances under consideration. With this in mind, we believe the Guidelines provide a substantive ‘fall-back’ position when the science and data are insufficient to refine and enhance the recommended monitoring and reporting procedures.

2. DATA PROCESSING

Claims have recently been made in the published literature that the data collection and processing activities associated with large-scale capital dredging projects are inconsistent with the environmental monitoring and reporting requirements required by such projects (Falkenberg and Styan 2014). The somewhat counter-intuitive claim is that too much data is being collected and analysed rather than too little (Table 1). However, this is not a shared view with others questioning the ability of many monitoring programs to achieve their stated objectives (Legg and Nagy 2006 and references therein). Indeed, Peterman (1990) suggested "inadequate monitoring can be both misleading and dangerous not only because of their inability to detect ecologically significant changes, but also because they create the illusion that something useful has been done".

While we certainly do not subscribe to the "more is necessarily better" philosophy, our long experience in *environmetrics* suggests there are very few instances where scientific understanding and environmental protection are not enhanced by access to data which is *spatiotemporally* dense. We do however agree with Gibbs (2013) that short-term, project-driven monitoring has the potential to distort assessments of natural and anthropogenically-derived sediment sources and sinks and that "having a high background turbidity implicitly incentivises dredging operations to produce dredge plumes that push up to the envelope of natural variability".

Table 1.

Year commenced	Programme	# sites	Dredge volume Mm^3	Classification of sites[a]	Collection interval (min)	# records per day (24 h)[b]	Duration of dredging (months)	# Records for programme [b]
2006	Hay Point	4	9	2 impact, 2 reference	10	576	6	103,680
2007	Cape Lambert (Port A)	8	2.5	3 impact, 1 sentinel (uncertain), 4 reference	30	461	10	115,200
2007	Pluto	3	14	3 impact	10	432	3	38,880
2008	Cape Lambert (Port B)	13	16	6 impact, 5 influence, 2 reference	30	624	13	243,360
2009	Gladstone	18	25	6 compli-ance, 12 supple-mentary	15	3072	30	2,764,800
2010	Gorgon	24	7.6	7 moderate impact, 9 influence, 8 reference	30	1296	18	622,080
2012	Wheatstone	22	45	2 high impact, 4 moderate impact, 7 influence, 9 reference	30	1056	54 [c]	1,710,720[c]
2012	Darwin Harbour	19	14.6	8 reactive, 11 non-reactive	15	1824	15–21[c]	820,800–1,149,120[c]
Total		**111**					**149–155**[c]	**6,419,520–6,747,840**[c]

Returning to the issue of data profligacy during dredging campaigns (Falkenberg and Styan 2014) we believe the relatively high sampling frequencies reported for the Gladstone Project in Table 1 are appropriate by virtue of the following:

1. Once monitoring instruments have been deployed, the marginal cost of data collection is relatively insensitive to the sampling frequency;

2. In view of 1 above, a higher sampling frequency provides a level of 'insurance' in terms of an increased ability to identify and correct anomalies in the data record;

3. Data collected at a high sampling frequency can always be resampled to a lower frequency whereas the converse is not possible;

4. A higher sampling frequency allows for better feature and signal extraction than a lower sampling frequency; and

5. The sampling frequency and companion data processing procedures for the Gladstone Project (Table 1) were finely tuned to work in concert with the

statistical smoothing technique (Exponentially Weighted Moving Average) and statistical algorithms underpinning TV computations and to that end, the data collection strategy was fit-for-purpose.

In the remainder of this report it is assumed that the 'raw' data used to characterise 'background' conditions have already undergone preliminary QA/QC checks by the contractor responsible for data collection. We refer to this pre-processing of the raw turbidity data as **functional QA/QC** or *F-qaqc* and the resulting data as *functionally-assured*.

The purpose of the *F-qaqc* step is to check the consistency and integrity of the data obtained from the monitoring instruments and, where appropriate, to take remedial action. These activities include, but are not limited to:

- Flagging and if necessary, removing readings obtained when equipment was known to be faulty, unreliable, or unserviceable;
- Flagging, but **not** removing readings obtained during adverse weather or oceanographic conditions;
- In the case of dual-instrument deployments, aggregating readings in accordance with agreed protocols;
- Implementing agreed protocols in the case of instrument failure for a dual-instrument deployment.

In the following sub-sections, we discuss options for the processing of functionally-assured data *prior* to the computation of trigger values. Collectively, these procedures constitute the **statistical QA/QC** step or S-*qaqc* and the resulting data is referred to as *statistically-assured*. The focus of *S-qaqc* activities is to:

- Identify extreme and unusual data in terms of their *statistical* properties;
- Use statistical data imputation techniques in accordance with agreed protocols to overcome problems created by blocks of missing data;
- Use statistical smoothing techniques in accordance with agreed protocols to attenuate the influence of aberrant observations; and
- In accordance with agreed protocols, apply low-pass statistical filters to characterise the nature of the underlying response-generating process and of the stochastic error component.

While the distinction between *F-qaqc* and *S-qaqc* might seem to be academic, it is nevertheless important in practice as it: (i) clearly delineates responsibilities for different

levels of data processing; (ii) avoids duplication of mathematical and statistical procedures; and (iii) provides a data audit trail from 'raw' to 'final' forms. Figure 1 is indicative of the type of activity undertaken as part of *S-qaqc* but not *F-qaqc* and shows a plot of the joint confidence ellipse for two correlated water quality variables.

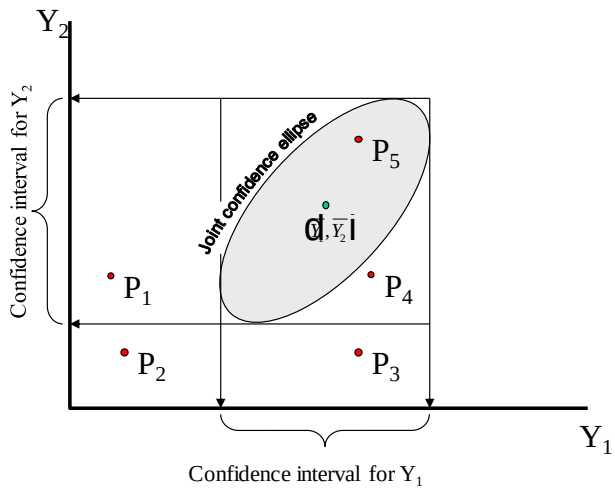

Figure 1. Depiction of joint confidence ellipse and univariate confidence intervals for correlated variables, Y_1 and Y_2. Points P_1, P_2, P_3 , and P_4 are all aberrant since they lie outside the joint confidence ellipse although P_1, P_2, and P_3 lie within one of the univariate confidence intervals and P_4 is not judged be to aberrant at all by reference to the individual intervals.

Conventionally, the separate univariate confidence *intervals* for Y_1 and Y_2 are used to identify 'outliers' in the data set. However, this approach ignores the additional information afforded by the *joint* relationship between the variables and results in the misclassification of P_4 (Table 2).

Table 2.

Point	Joint classification	Univariate classification using:	
		Y1	Y2
P1	aberrant	aberrant	normal
P2	aberrant	aberrant	aberrant
P3	aberrant	normal	aberrant
P4	aberrant	normal	normal
P5	normal	normal	normal

2.1 Smoothing

Statistical smoothing may be viewed as a companion activity to the aberrant data detection issue discussed in the previous section although the emphasis is somewhat different. In general terms, an observed (turbidity) time-series can be simply described as:

$$data = signal + noise \tag{1}$$

The *signal* represents the underlying 'driving force' and integrates all the components of predictable variation in the data. In a perfect world, all of these components would be known and quantified. Of course the reality is different and what we end up with is a representation or *model* of the signal. The adequacy of this model is largely a function of the number of components and how well they are described and brought together. This invariably means that no model is perfect (an observation that lead famous statistician G.E.P. Box to proclaim "all models are wrong but some are useful"). Thus the true *signal* can then be further decomposed as:

$$signal = model + LoF \tag{2}$$

Where *LoF* refers to *lack-of-fit* and represents all the components, processes, factors, etc. that our model did not account for. Bringing (1) and (2) together gives:

$$data = model + LoF + noise \tag{3}$$

The term *'noise'* represents the *stochastic* or purely random variation that can only be described in probabilistic terms. It is the *noise* that gives rise to the scatter in a scatter diagram.

It is important to realise that *statistical modelling* (or modelling more generally) is simply an attempt to estimate the signal. Often, this is achieved by fitting a mathematical function $f(t;\underline{\Theta})$ to the data where *t* is time and $\underline{\Theta}$ is a vector of unknown parameters – the latter being estimated from the data and denoted $\underline{\hat{\Theta}}$. Thus from Equation 3

$$data = f(t;\underline{\hat{\Theta}}) + LoF + noise \tag{4}$$

The difference between *data* and *model* is $\left\{data - f(t;\underline{\hat{\Theta}})\right\}$ and is referred to as the *residual* term. It is clear from Equation (4) that the *residuals* are in fact observations on the quantity $\left\{LoF + noise\right\}$. The *LoF* term is generally ignored in most modelling exercises (since it represents what we don't know!) and hence the *residuals* are simply treated as observations on the *noise* or stochastic error term. Statistical procedures have been devised to help decompose residuals into *LoF* and *noise* components although these will not be discussed here.

Hydrodynamic modelling is a core component of the environmental impact assessment and approval process for any dredging project and is in effect a more elaborate version of equation (4) where instead of a single function $f(\cdot)$ there are many functions and systems of differential equations describing the combined effects of key physical processes such as winds, waves, currents, tides and other relevant 'forcings' and factors. Assessments of the adequacy of a hydrodynamic model are made through an examination of residuals, which as we've seen represent the combined effects of *noise* and *LoF*. In practice, most (hydrodynamic) models ignore the stochastic component and as such these models are

deterministic. Assessments of the variability of outputs from a hydrodynamic model are usually achieved by varying some of the model parameters, re-running the model and examining the changes in the output. While this type of analysis is often undertaken in fulfilment of an 'uncertainty analysis' the reality is that it's not an uncertainty analysis. Changing the parameters in a deterministic model leads to an entirely predictable change in the outputs – there's nothing that's uncertain. Admittedly, the assessment might not be straightforward when the model is a complex computer model, but nonetheless, there's nothing in the output that is a function of a *chance* (and hence unpredictable) process. The type of analysis just described is more accurately referred to as a *sensitivity analysis* where the analyst is interested in seeing how sensitive the output is to minor perturbations of inputs and/or assumptions.

When the functional form $f(\cdot)$ is prescribed by the data analyst, we say that the modelling process is *parametric*. Other strategies have been developed to estimate the signal without being prescriptive about the nature of $f(\cdot)$. In this case, the modelling process is *non-parametric*. Non-parametric procedures are a popular choice since they effectively 'let the data speak for themselves' and free the data analyst from the task of having to identify and parametrise the function $f(\cdot)$. Of course this doesn't work for hydrodynamic modelling since the objective is to develop a model which can *predict* the behaviour of a system under conditions that are yet to occur rather than simply describing what has occurred.

There are numerous statistical smoothing techniques available to smooth turbidity data and estimate the underlying signal. We next describe two such methods.

The Exponentially Weighted Moving Average (EWMA)

The exponentially weighted moving average or EWMA is a statistical smoothing procedure that has its origins in a branch of industrial statistics known as Statistical Process Control (SPC) and is one of a number of devices that are collectively referred to as "control charts". Over recent years, control charting techniques have been successfully used to monitor environmental processes following their recommended use in the ANZECC/ARMCANZ (2000b) National Water Quality Guidelines. Whether the setting is environmental or industrial, the objective is the same: to visualise the evolution over time of a metric of 'performance' and to set limits or 'triggers' on that metric that warn of an actual or impending 'out-of-control' situation.

As the name implies, a moving average (MA) is a locally-weighted average that moves in time in an attempt to track the underlying signal. The simplest MA is based on the concept of stepping an 'window' across the time series and plotting the (arithmetic) mean of the points falling in the window. The *degree of smoothing* is controlled by the width of the window.

Weighted moving averages operate as described above except that whereas in a simple moving average every observation in the window is assigned equal weight, a weighted moving average uses *unequal* weights. An example of a *Gaussian* weighted moving average scheme is depicted in Figure 2.

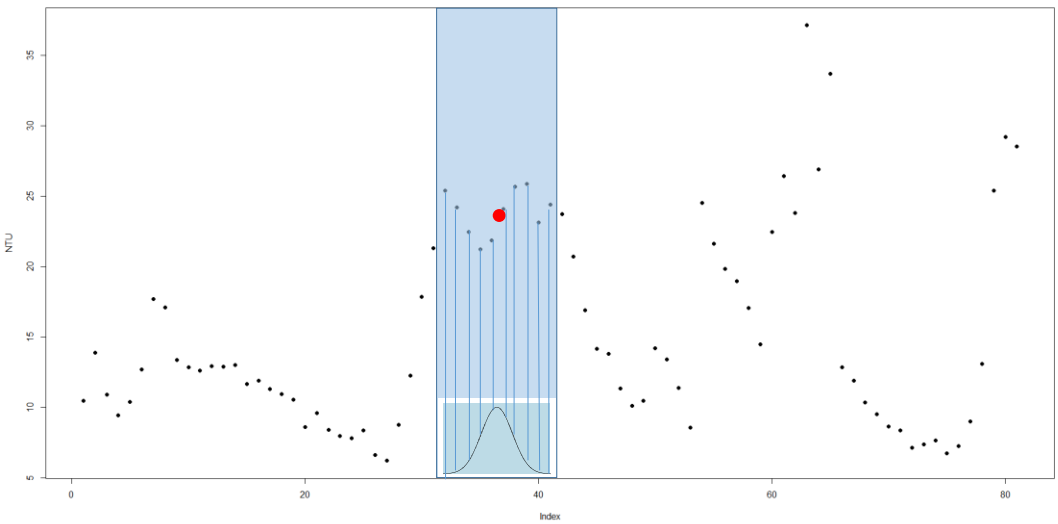

Figure 2. Turbidity time series data (solid circles) with Gaussian-weighted moving average (solid red circle). Weights applied to the individual data values are derived from the Gaussian curve (solid black line).

The EWMA is another variant of a weighted moving average scheme where the weights assigned to the current and all previous observations decrease *exponentially* (Figure 3). This is a consequence of the formula for an EWMA (Equation 5).

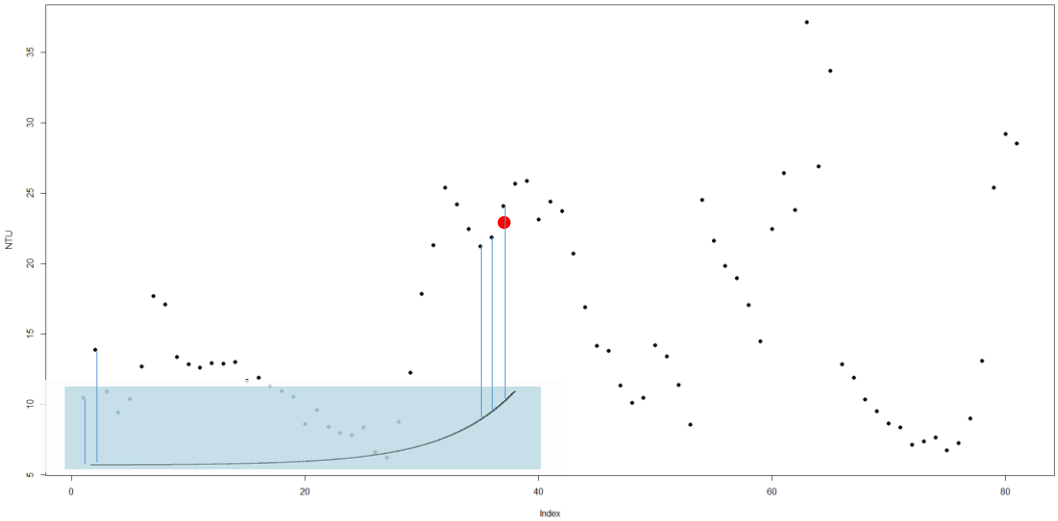

Figure 3. As for Figure 2 excepted averaging is achieved by applying series of weights that decrease exponentially.

$$Z_i = \lambda \bar{X}_i + (1-\lambda) Z_{i-1} \quad ; \quad 0 < \lambda < 1 \tag{5}$$

Equation 5 shows that the current value of the EWMA $\left(Z_i\right)$ is weighted combination of the *current* mean $\left(\bar{X}_i\right)$ and the *previous* value of the EWMA $\left(Z_{i-1}\right)$ where the value of weighting factor λ is an arbitrary constant chosen by the analyst to give greater or lesser importance to the balance between current and historical values. To implement Equation 5, a starting value Z_0 is required and this is usually set equal to the long-term average of the process under consideration.

The EWMA was successfully applied as both a *management* and *compliance* tool during the Port of Melbourne's Channel Deepening Project (https://en.wikipedia.org/wiki/Port_Phillip_Channel_Deepening_Project) and Gladstone Port Corporation's Western Basin Dredging and Disposal Project (http://www.gpcl.com.au/development/western-basin-dredging-and-disposal-project). In both instances, a value of 0.6 was used for the weighting term λ in Equation 5 which meant that the value of the EWMA at any time *t* was comprised of 60% of the mean at time *t* and 40% of the value of the EWMA for the immediately preceding time period. This 60:40 split was found to provide a good balance between smoothing and responsiveness.

For the Gladstone Project, the EWMA was updated every 6 hours. With raw turbidity being logged every 15-minutes this meant that the average $\left(\bar{X}_i\right)$ in Equation 5 was calculated using 24 turbidity readings. While the overall performance of this scheme was

judged to be highly successful and appropriate, operationally, the 6-hour time delay introduced by the EWMA computation was at times problematic in that it unnecessarily delayed response times when worsening water quality was self-evident (L. Andersen *pers. comm.*).

Another, potentially more serious drawback of the EWMA is its inability to be calculated once missing data are encountered. This is due to the *recursive* nature of Equation 5 whereby the current value of the EWMA is dependent on the previous value. Thus if missing data prevent the calculation of $\overline{X}_i$ for the current period, then Z_i cannot be computed as must also be recorded as 'missing'. Even though data may be available for the following periods, the 'missingness' of Z_i means Z_{i+1} cannot be computed which in turn means Z_{i+2} cannot be computed and so on. A robust smoothing technique that does not suffer from this drawback is the Kolmogorov-Zurbenko (KZ) filter.

The Kolomogorov-Zurbenko (KZ) Filter

The Kolmogorov-Zurbenko (KZ) filter belongs to the class of low-pass filters and as such is potentially useful for smoothing turbidity time-series data. In essence the KZ filter is computed by taking *k* time iterations of a moving average (MA) filter of *m* points. It therefore has only 2 parameters – *k* and *m* both of which have clear physical interpretations. In addition to its ease of computation and unlike the EWMA, the KZ filter easily deals with missing data situations and is near optimal (Yang and Zurbenko 2010).

The KZ filter applied to a time series $X(t), t = 0, \pm 1, \pm 2, \ldots,$ is given as:

$$KZ_{m,k}\left[X(t)\right] = \sum_{s=-k(m-1)/2}^{k(m-1)/2} \frac{a_s^{m,k}}{m^k} X(t+s) \tag{6}$$

where $a_s^{m,k}$ are given by the polynomial coefficients of $\left(1 + z + \cdots + z^{m-1}\right)^k$

$$\sum_{s=-k(m-1)/2}^{k(m-1)/2} z^{s+k(m-1)/2} a_s^{m,k} = \left(1 + z + \cdots + z^{m-1}\right)^k \tag{7}$$

Implementation

As mentioned above, the KZ filter computed by taking *k* time iterations of a moving average (MA) filter of *m* points as follows as described in Yang and Zurbenko (2010):

1. First iteration is to apply a MA filter to *m* points:

$$KZ_{m,k=1}\left[X(t)\right]=\frac{1}{m}\sum_{s=-(m-1)/2}^{(m-1)/2}X(t+s) \tag{8}$$

2. Second iteration is to apply a MA operation to the result of the first iteration:

$$KZ_{m,k=2}\left[X(t)\right]=\sum_{s=-(m-1)/2}^{(m-1)/2}\frac{1}{m}KZ_{m,k=1}\left[X(t+s)\right]$$

$$=\sum_{s=-2(m-1)/2}^{2(m-1)/2}\frac{a_s^{m,k=2}}{m^2}X(t+s) \tag{9}$$

3. k^{th} iteration:

$$KZ_{m,k}\left[X(t)\right]=\sum_{s=-k(m-1)/2}^{k(m-1)/2}\frac{a_s^{m,k}}{m^k}X(t+s) \tag{10}$$

The steps identified by Equations 8-10 are amenable to simple spreadsheet coding and as such this smoothing method is well suited to both the *F-qaqc* and *S-qaqc* functions. A more comprehensive set of tools developed around the KZ algorithm is available in the R package `kza` (Close and Zurbenko 2016). We next provide an example of smoothing turbidity time-series data using the `kza` package.

Example

Some typical 15-minute turbidity data (measured in NTU) acquired autonomously from a moored turbidity logger is shown in Figure 4.

Also shown in Figure 4 are traces of the output from the KZ filter for smoothing windows of 1hr, 2hr, 3hr, and 4hr. In each case the *k* parameter in Equation 6 was fixed at 3.

The choice of an appropriate smoothing window (i.e. the *m* value in Equation 6) depends in part on the objective of the smoothing exercise. If we are mostly interested in signal extraction, then the 4hr window is appropriate (for these data). However, if the smoothing is part of the *F-qaqc* process whereby the aim is to simply filter out short runs of high-frequency oscillations in the turbidity signal, then a 1hr moving window would be appropriate (for these data). This choice of *m* also results in minimal loss of fidelity of the original signal (Table 3) resulting in a compact distribution of residuals (= difference between original and smoothed data) (Figure 5). From Table 3 we see that the desired

objective of ‘knocking out’ the extreme, transient peaks has been achieved with the maximum reduced from 326.8 to 288.5 but all other statistics relatively unaffected.

Table 3. Summary statistics for original and smoothed series.

Series	Minimum	First quartile	Median	Mean	Third quartile	Maximum	NA's
Original	0.000	8.350	16.100	22.030	29.700	326.800	377
1hr smooth	1.200	8.809	16.920	22.020	29.870	288.500	267
2hr smooth	1.470	9.276	17.870	22.000	29.860	274.300	225
3hr smooth	1.528	9.550	18.230	21.980	29.980	258.100	201
4hr smooth	1.542	9.677	18.440	22.000	30.190	257.100	181

We next turn our attention to the related activity of *signal extraction*.

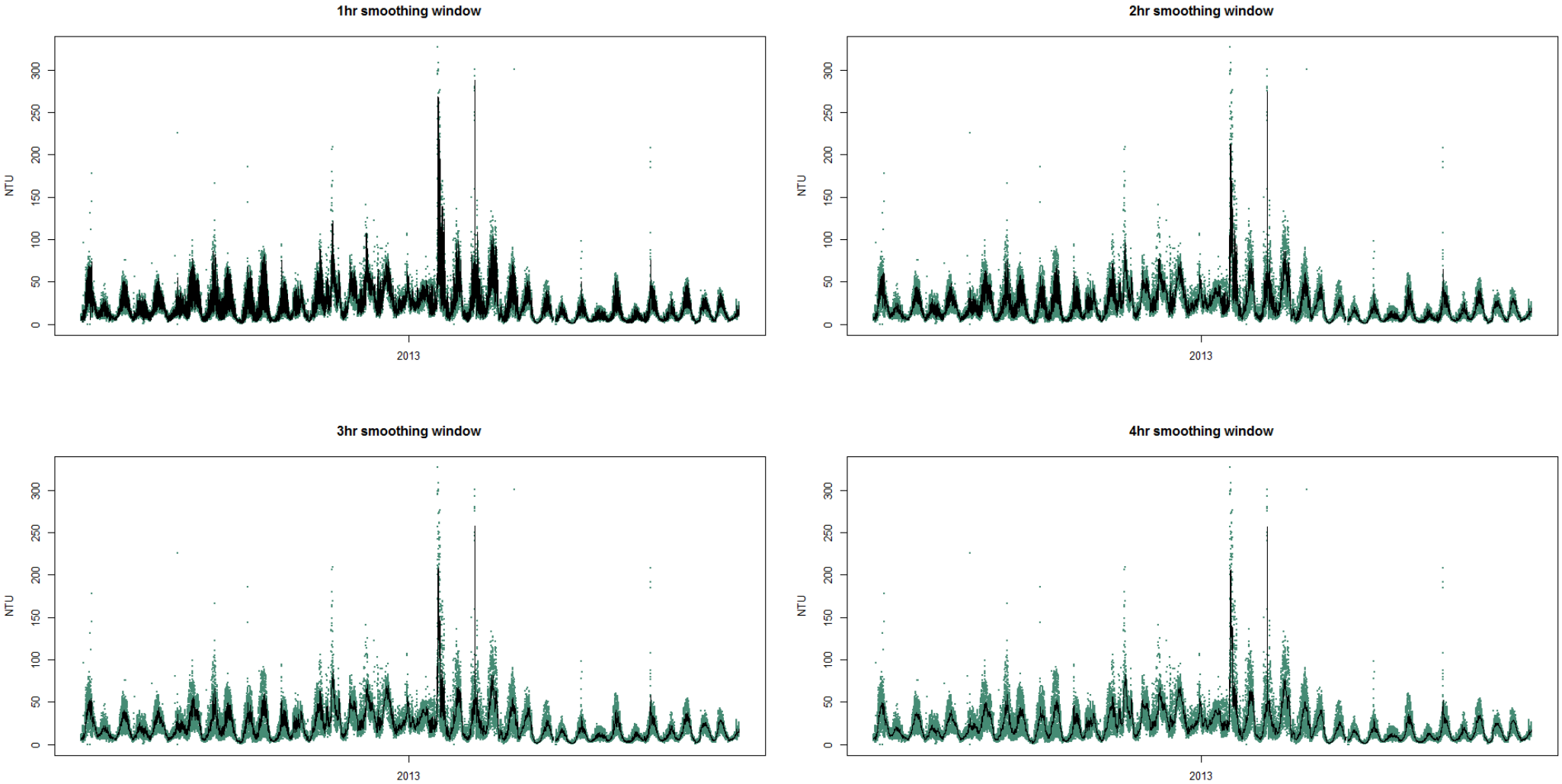

Figure 4. Effect of varying the *m* parameter in the KZ filter. Original data shown as green solid circles with KZ filter output overlaid (black lines).

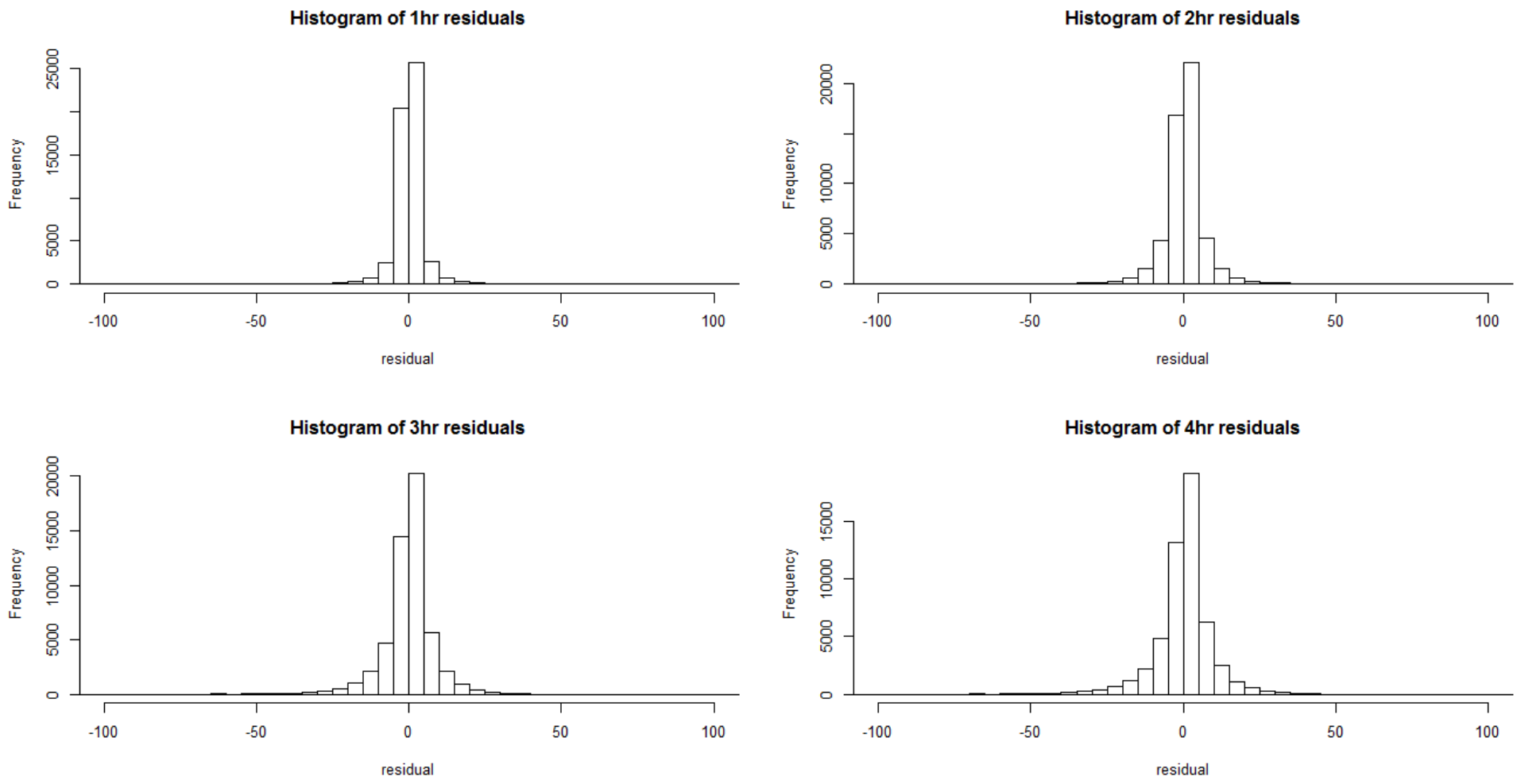

Figure 5. Residual plots after apply the KZ filter shown in previous figure.

2.2 Signal extraction and pattern recognition to infer dredging impacts

The `kza` package used in the previous example has a number of other powerful capabilities. One of these is the KZ Adaptive Filter, or KZA which has the ability to identify abrupt breaks in noisy signals. This is potentially useful in the analysis of turbidity data which is a measure of the *combined* effects of 'natural' or 'background' turbidity plus a dredge-derived contribution.

For large-scale, capital dredging projects, it is advantageous both economically and generally environmentally, to limit the dredging to an intensive, but short campaign. Hence the objective is to minimise operational downtime. However, that objective can be compromised by: (i) regulated stand-downs; (ii) equipment malfunction; (iii) scheduled maintenance; and (iv) adverse weather and/or sea conditions. Thus the binary dredge status (dredging / not dredging) fluctuates somewhat sporadically over the duration of the project. Since dredging contributes to the overall sediment load in the water column, this fluctuating binary pattern can result in break-points in the turbidity signal. However, in waters which are both naturally highly turbid and temporally very variable, these break-points will be masked. A fundamental, but extraordinarily difficult task during any dredging project is to understand and quantify the dredge-related contribution to the observed turbidity at any time or place. For the Port of Gladstone project this was accomplished by applying a 'turbidity prism' concept developed by *Environmetrics Australia* (Figure 6).

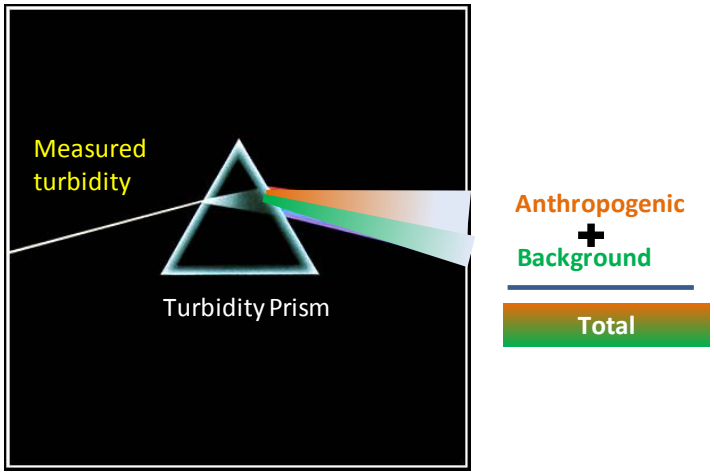

Figure 6. Conceptualisation of the 'turbidity prism'.

For the Gladstone project, a sophisticated *statistical* model was developed which provided *predictions* of the *background* or *natural* turbidity levels that would have been expected under prevailing conditions of wind, currents, tide, rainfall etc. in the absence of any dredging. This predicted background component was subtracted from the measured total turbidity to provide an estimate of the dredge-derived contribution.

The development of the predictive background turbidity model is a complex process that relies on the availability of good quality background turbidity data at multiple sites, on short-time scales (minutes) for at least a whole year. For many projects, this is a luxury that is simply unavailable. Nevertheless, the ability to decompose the measured turbidity data into its constituent components as shown in Figure 6 is highly desirable since it provides a consistent, rigorous, and objective method of quantifying the relative contributions to the measured turbidity. Without this capability, Project Proponents are defenceless against claims of being responsible for all instances of elevated turbidity during a dredging project. Experience in highly variable natural environments shows that exceedingly high levels of natural / background turbidity are not uncommon.

In the absence of a predictive model for background turbidity, the following strategy is proposed: the KZA filter first identifies potential time intervals when a break occurs. These intervals are investigated more intensively by reducing the window size which increases the resolution of the smoothed series. Comparison of the smoothed series against a pre-determined threshold generates a 'box-car' series which can be compared against the known pattern of dredging operations. A high correlation between these two series suggests a strong dredge-related contribution. Conversely, a lack of correlation between these two series suggests dredging activities are not 'driving' the turbidity signal.

The process is depicted in Figures 7 and 8. Figure 7 shows the construction of a synthetic turbidity time-series that is obtained by increasing the background turbidity signal during periods when dredging activities are taking place. This 'amplified' background signal is then fed into the KZA filter to obtain a series of spikes which reflect the estimated positions of break-points (Figure 8). Further processing of this pattern of spikes reduces it to a binary response to be compared against the *actual* pattern of dredging operations (Figure 9).

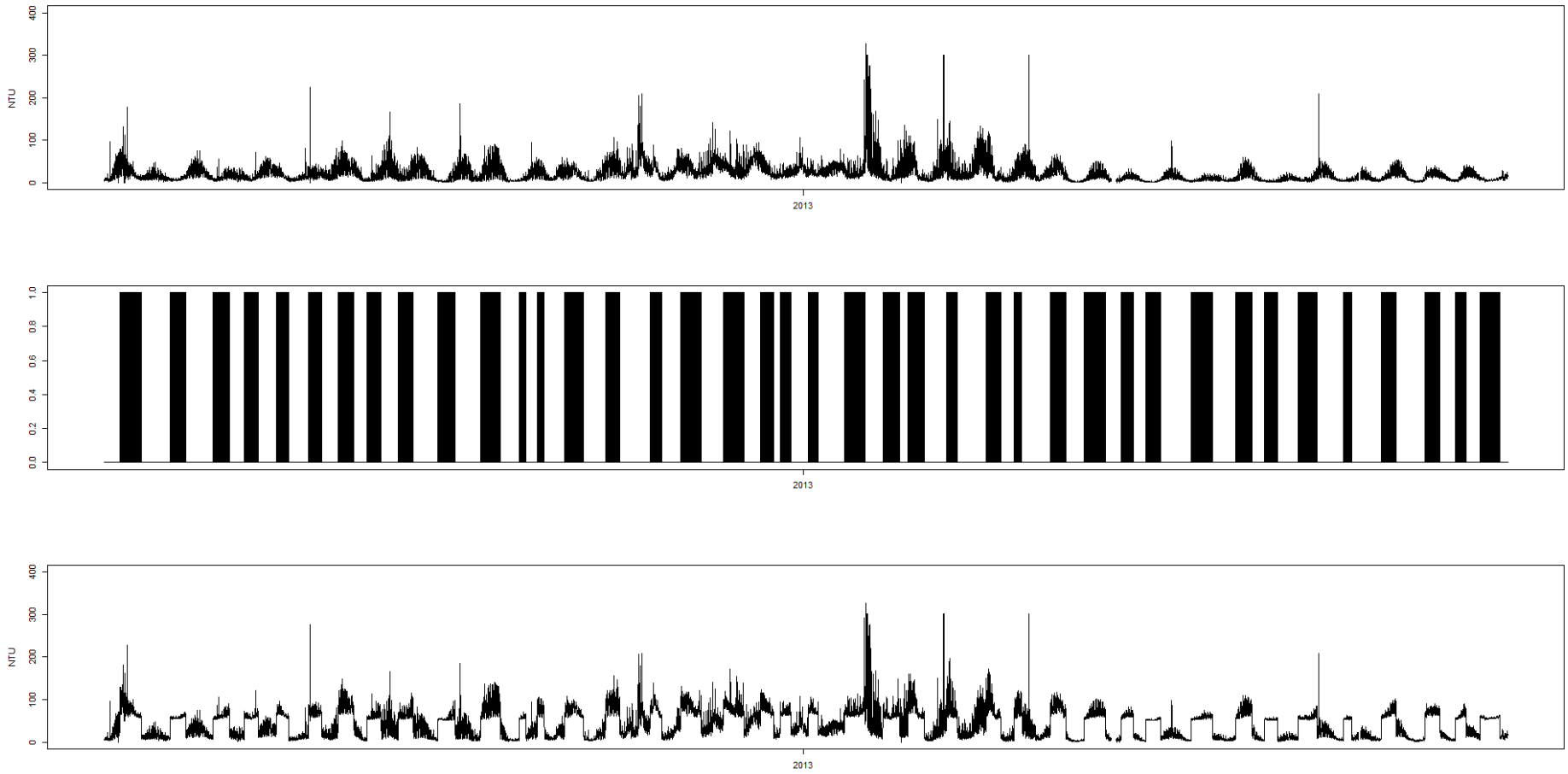

Figure 7. Construction of synthetic time-series (bottom) reflecting dredging activity obtained by intensifying background turbidity (top) during periods of dredge activity (solid black bars, middle).

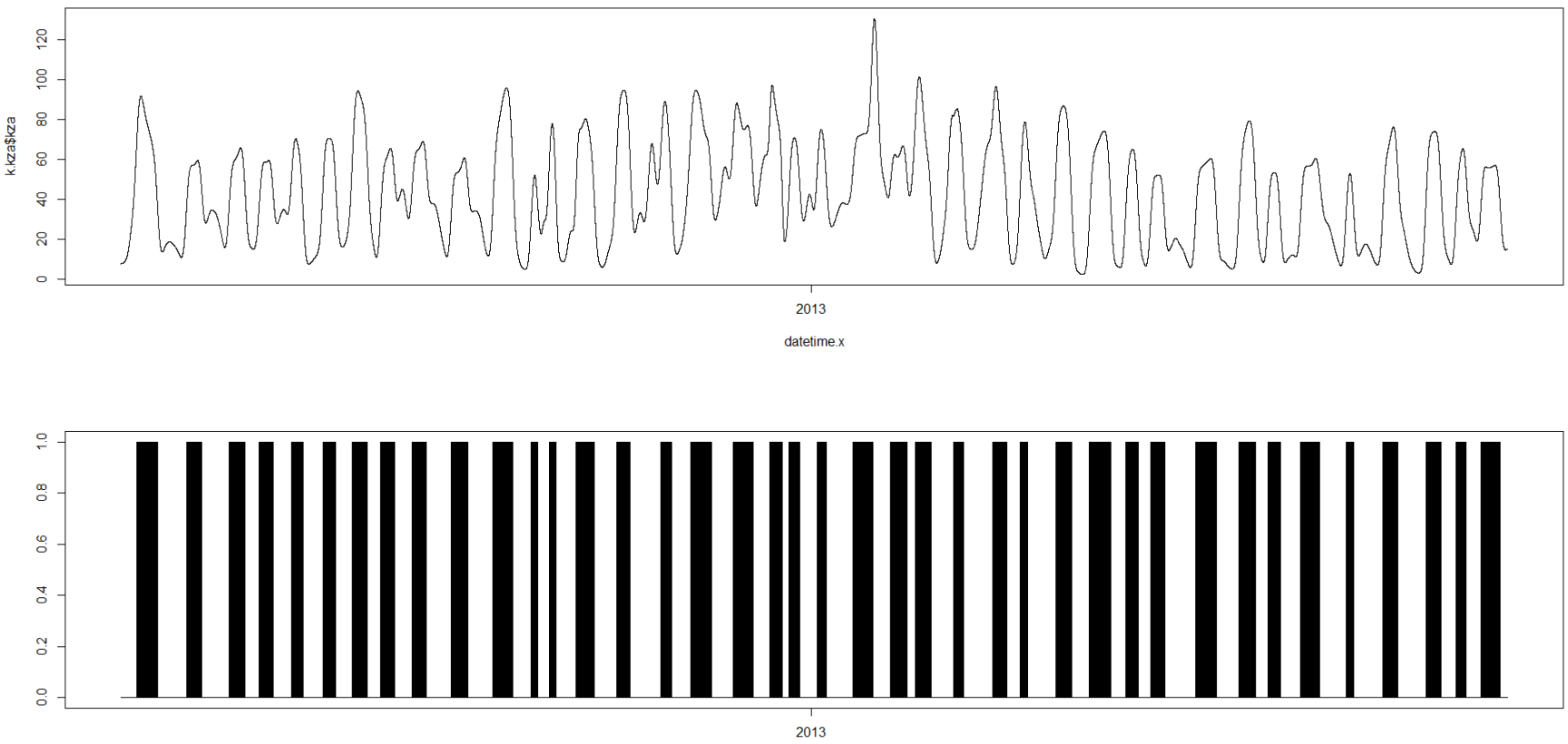

Figure 8. Output (top) after applying adaptive KZ filter to synthetic turbidity series (bottom of Figure 7) with further processing to convert into a binary sequence of break-points (bottom).

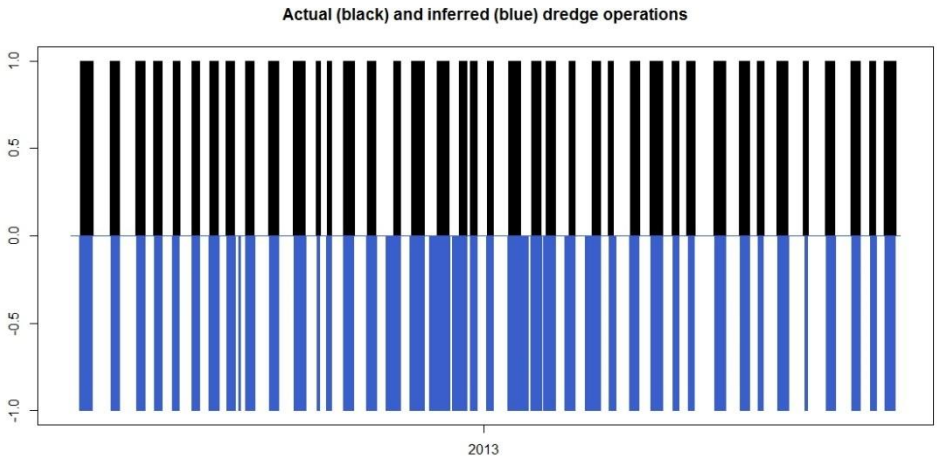

Figure 9. Comparison of estimated pattern of break-points from analysis of turbidity signal (solid blue bars) with the actual pattern of dredging operations (solid black bars).

2.3 Signal Reconstruction

The smoothing techniques discussed in the previous section provide a non-parametric method of extracting a signal from a noisy series and while useful, they do not attempt to *model* or more fully describe the underlying response-generating process. While this type of investigation is generally not required or undertaken during dredging projects, a brief outline is provided here for completeness.

Many time-based phenomena (such as recordings of turbidity and other water quality parameters) exhibit periodic behaviour on varying time-scales. For example, a very strong driver of turbidity is tidal stage which induces a highly regular rise and fall in mean turbidity every 6 hours as well as other 'harmonics' associated with various lunar phases. These frequency constituents can be identified by various mathematical and statistical techniques through a process known as *spectral decomposition*. Common among these is the *discrete Fourier transform* (DFT), *wavelets*, and more recently the Kolmogorov-Zurbenko Fourier transform (KZFT).

The KZFT can be thought of as an iteration of the conventional Fourier transform. It is a linear band-pass filter that belongs to the category of Short-time Fourier Transforms (STFT) (Hlawatsch and Boudreaux-Bartels 1992). The linearity property of the KZFT makes is particularly suited to the analysis of turbidity time-series data as it separates multicomponent signals by reconstructing one component at each time (Yang and Zurbenko 2010).

While the mathematical detail of the KZFT is beyond the technical scope of this document, its implementation is relatively straightforward via the `kzft` function in the R package `kza` (Close and Zurbenko 2016). A demonstration of the kzft function to tease out the underlying signal from noisy time-series data is shown in Figure 10.

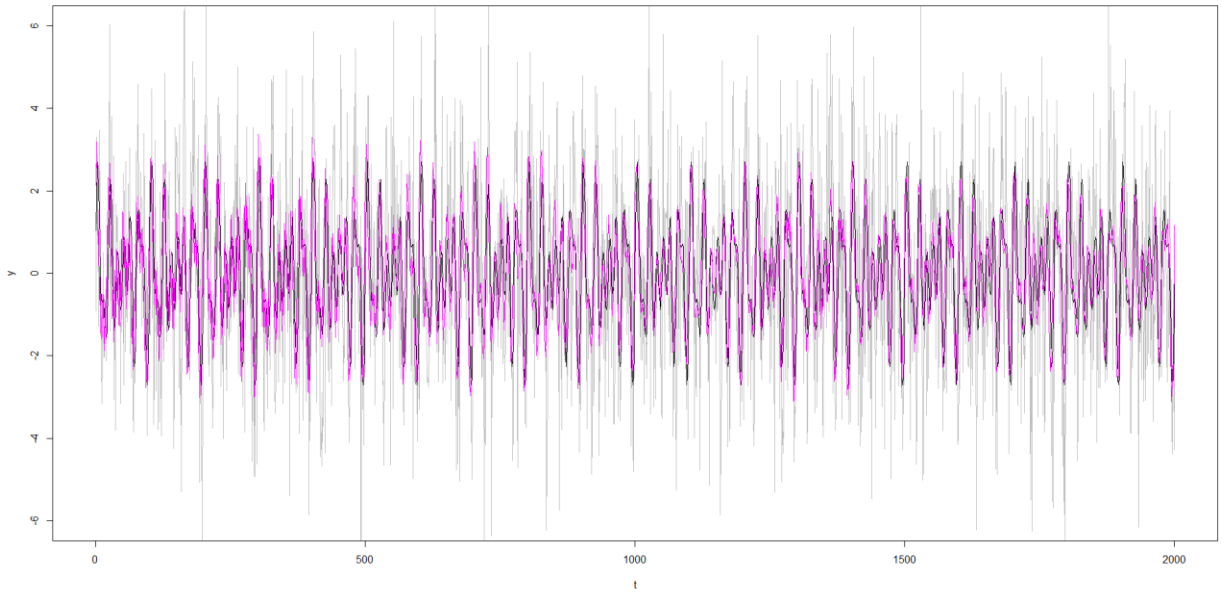

Figure 10. Example of signal reconstruction using kzft. actual signal (black); signal + noise (gray); signal reconstructed from noisy data (magenta).

2.4 Missing values

Missing values are problematic for statistical analyses generally, but pose particular challenges in environmental settings since different methods of treating the 'missingness' will produce different outcomes thus potentially leading to different environmental assessments. Fortunately, this problem is largely avoided when dual turbidity loggers are used. However Environmental Management Plans need to anticipate periods of missing data and articulate data management procedures during such times.

An example of a turbidity data having some relatively large gaps in the temporal sequence is shown in Figure 11. These large gaps cannot be overcome by smoothing techniques such as the KZ filter and, as previously mentioned, they cause the complete breakdown of the EWMA.

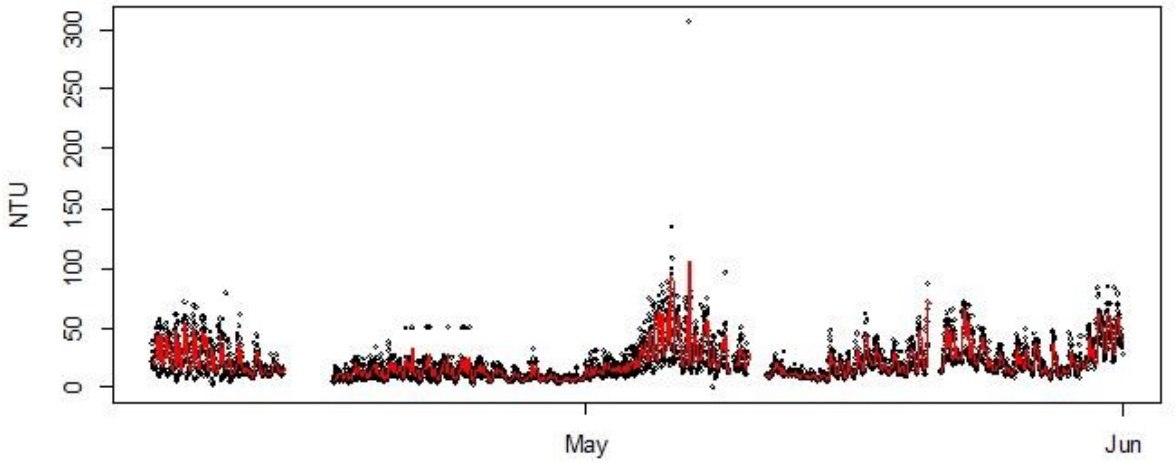

Figure 11. Example of raw turbidity data (open circles) with missing values. Result of KZ smoothing with 1 hr window shown in red.

To date, the only effective and credible means of dealing with the level of 'missingness' indicated in Figure 11 is to *impute* the values of the missing data using advanced statistical modelling techniques. Models developed by *Environmetrics Australia* utilise information on ancillary variables such as wind speed, wind direction, tide, currents, and rainfall together with the autoregressive properties of the sequence prior to and following the missing period to reconstruct the missing temporal sequence. An example of the results of this approach are shown in Figure 12 which shows the in-filling of a 3-day gap in turbidity readings.

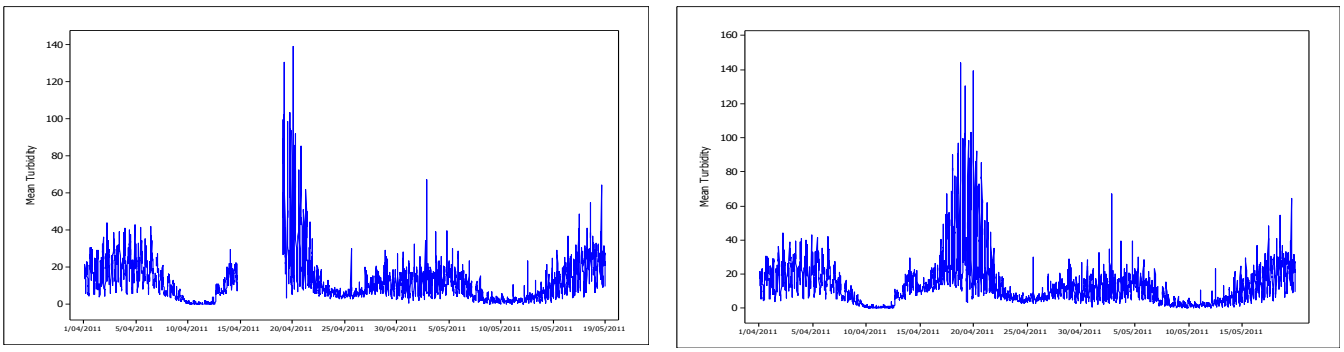

Figure 12. Illustration of temporal 'in-filling' of raw turbidity data. Original series with 3-day gap (left) and after data imputation (right).

Whether or not project proponents commission R&D work to develop local adaptations of predictive background turbidity models and project-specific methods of data imputation is largely a cost-benefit decision for them. Without these tools there is little that can be done to overcome blocks of missing data other than to simply record "NA" in the data

record and revert to a management response driven more by heuristics than science. Whatever these actions and management decisions are, they need to be documented in the EMP.

3. CHARACTERISING BACKGROUND TURBIDITY

In support of its Consent Application to deepen the existing shipping channel the Lyttelton Port Company (LPC) has commissioned Vision Environment to undertake a survey of background turbidity conditions during a 12-month period prior to the commencement of dredging in 2017. The proposed configuration of sites is shown in Figure 13.

To date, relatively little information has been collected on levels of background turbidity inside Lyttelton Harbour and off-shore although the consensus of subjective assessments from those familiar with the local marine environment is that turbidity levels are generally high and are highly variable in both space and time. This anecdotal picture is contradicted by the only significant record of actual measurements taken over a number of years by Environment Canterbury (ECan) (Figure 14).

Tables 4 and 5 show respectively, the frequency of ECan. sampling within Lyttelton Harbour and off-shore between 2002 and 2014. On average, only 1 or 2 samples were taken each month and this level of monitoring falls well short of what is required to develop a comprehensive picture of background turbidity levels. Furthermore, we are advised that the ECan. sampling tended to be undertaken only during relatively calm weather conditions which would explain the inconsistency between the analysis of these data (see later in this section) and the general perception of water quality.

Statistical Analysis of the ECan turbidity data

Notwithstanding previously stated concerns about the representativeness of the ECan. data set, a preliminary analysis may provide some insights into spatial gradients and key features of the background turbidity signal.

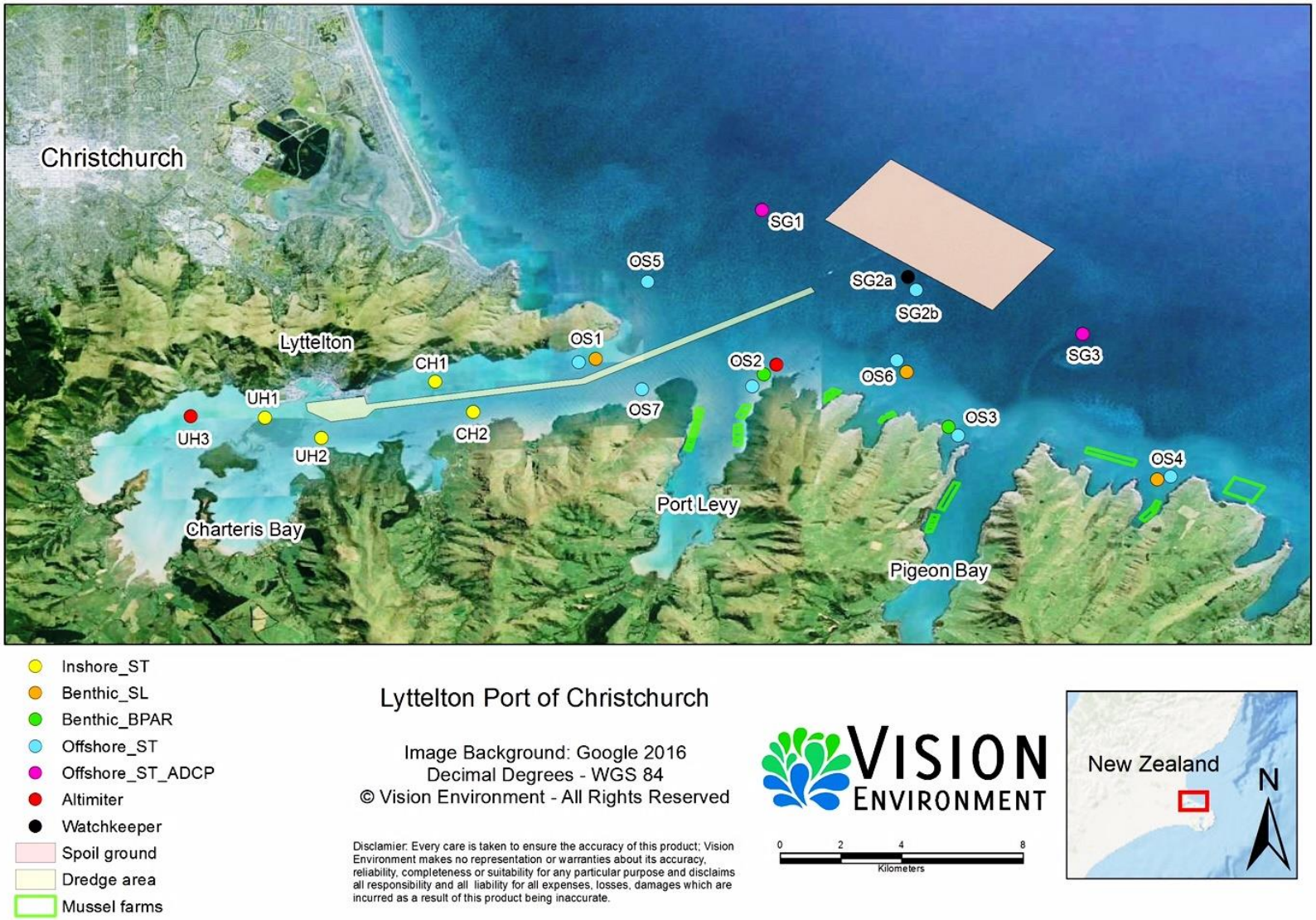

Figure 13. Vision Environment's proposed monitoring locations.

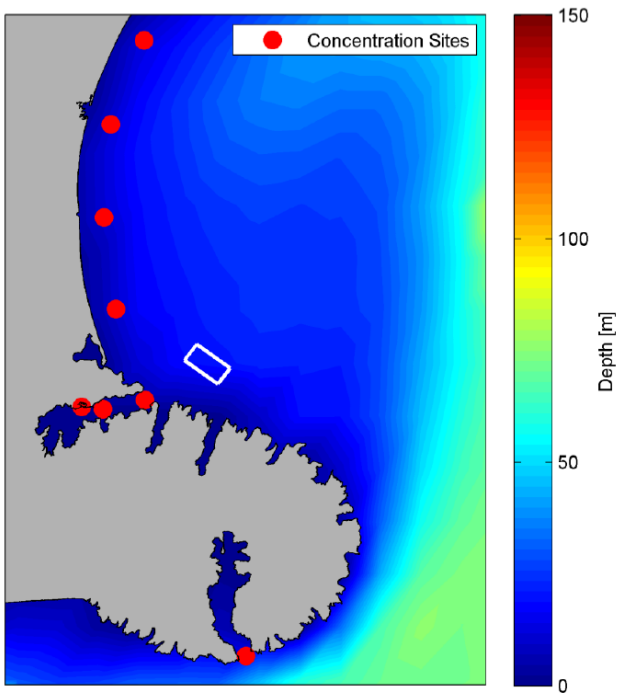

Figure 14. ECan. Surface turbidity monitoring sites

Table 4. Number of turbidity samples taken in Lyttelton Harbour by Environment Canterbury between 2002 and 2013.

	2002	2003	2007	2008	2009	2010	2011	2012	2013	All
Jan	0	2	0	3	0	0	0	1	2	8
Feb	0	2	0	3	0	0	0	1	2	8
Mar	0	2	0	5	2	2	3	4	5	23
Apr	0	0	0	3	0	0	0	1	2	6
May	0	2	0	3	0	0	0	1	2	8
Jun	0	2	0	5	2	2	3	4	7	25
Jul	2	0	3	0	0	0	1	1	0	7
Aug	0	0	3	0	0	0	1	2	0	6
Sep	2	0	5	2	2	6	4	5	3	29
Oct	4	0	3	0	0	3	0	2	0	12
Nov	2	0	3	0	0	0	1	2	0	8
Dec	2	0	5	2	2	3	4	5	3	26
All	12	10	22	26	8	16	17	29	26	166

Table 5. Number of turbidity samples taken at Pegasus Bay & Akaroa Heads by Environment Canterbury between 2002 and 2014.

	2003	2004	2007	2008	2009	2010	2011	2012	2013	2014	All
Jan	0	1	0	0	1	0	0	0	0	1	3
Feb	0	0	0	0	1	0	0	0	0	0	1
Mar	0	1	0	5	6	5	5	5	5	0	32
Apr	0	2	0	0	1	0	0	0	0	0	3
May	0	1	0	0	1	0	0	0	0	0	2
Jun	0	1	0	5	6	5	5	5	5	0	32
Jul	1	0	0	1	0	0	1	0	1	0	4
Aug	0	0	0	1	0	0	0	0	1	0	2
Sep	1	0	5	6	5	10	5	5	6	0	43
Oct	2	0	0	1	0	5	0	0	1	0	9
Nov	1	0	0	1	0	0	0	0	1	0	3
Dec	1	0	5	6	5	5	6	5	6	0	39
All	6	6	10	26	26	30	22	20	26	1	173

Boxplots of both TSS (mg/L) and NTU for the within Harbour (Figure 15) and off-shore data (Figure 16) show similar distributions of overall background turbidity and the absence of any significant spatial variation. Overall TSS concentrations within the Harbour are approximately 11 mg/L (about 4 NTU) and approximately 8 mg/L (about 2 NTU) at the off-shore sites. By any measure, these are exceedingly low.

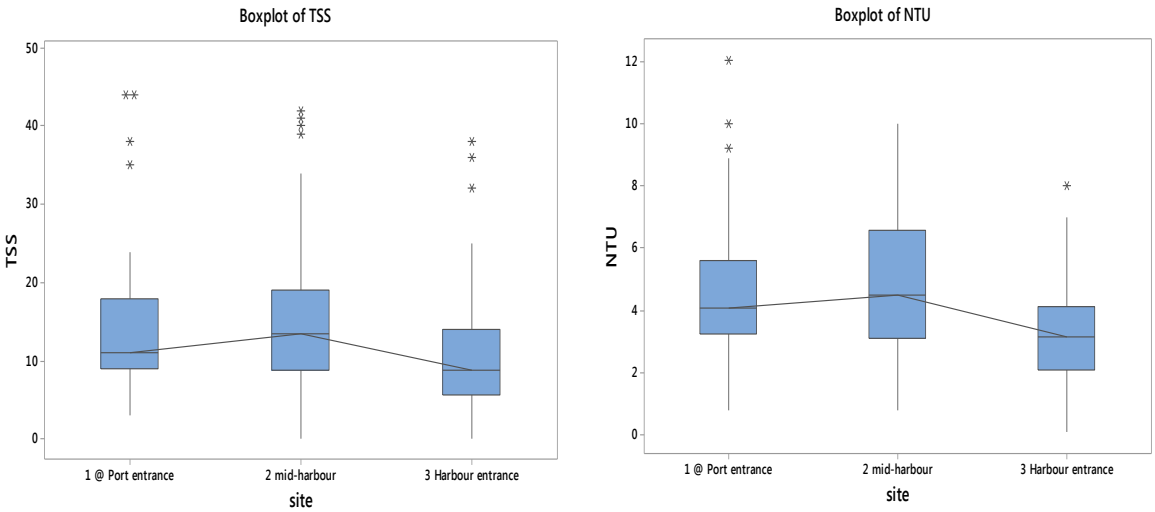

Figure 15. TSS and NTU boxplots for Ecan. monitoring of Lyttelton Harbour. Width of a box represents the Inter-Quartile Range (IQR) while horizontal lines within a box indicate position of median. Asterisks denote aberrant observations.

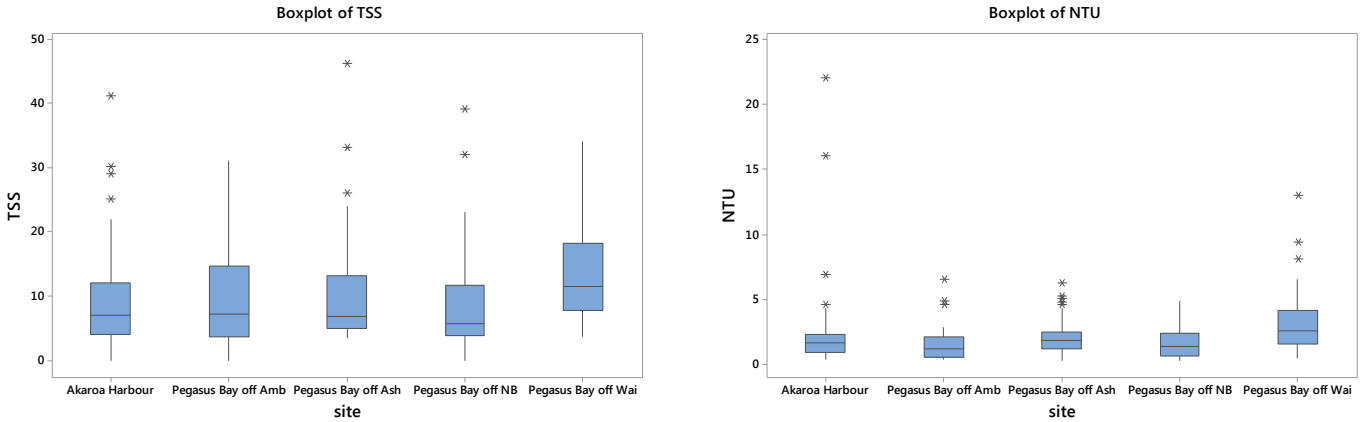

Figure 16. TSS and NTU boxplots for Ecan. off-shore monitoring. Width of a box represents the Inter-Quartile Range (IQR) while horizontal lines within a box indicate position of median. Asterisks denote aberrant observations.

The overall distributions of turbidity mid-harbour and at the Harbour entrance are virtually identical while there is a slight shift (lowering) of the turbidity distribution at the Harbour entrance (Figure 17).

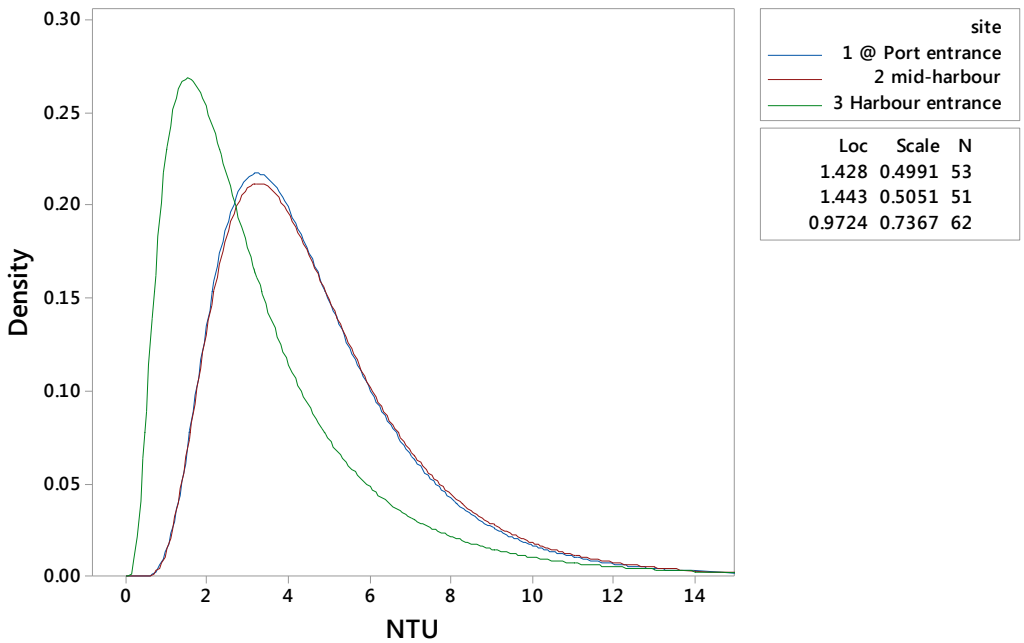

Figure 17. Theoretical log-normal probability models for NTU at ECan's 3 Harbour sites.

Tidal effects

Overall, the ECan. data provides no evidence of any tidal influence on measured turbidity (Table 6) although again, this contradicts local understanding and the results of an ADCP deployment (Figure 18). However, given this data has been pooled across days, months and years, the lack of any tidal signal is not surprising. Again, more intensive sampling is required to investigate this more fully.

Table 6. Mean turbidity (NTU) by Harbour location and tidal stage. (First cell entry is mean; second entry is sample size).

	ebb	flood	All
Port entrance	4.881	4.391	4.677
	31	22	53
mid-harbour	4.952	4.525	4.726
	24	27	51
Harbour entrance	3.428	2.998	3.206
	30	32	62
All	4.388	3.885	4.143
	85	81	166

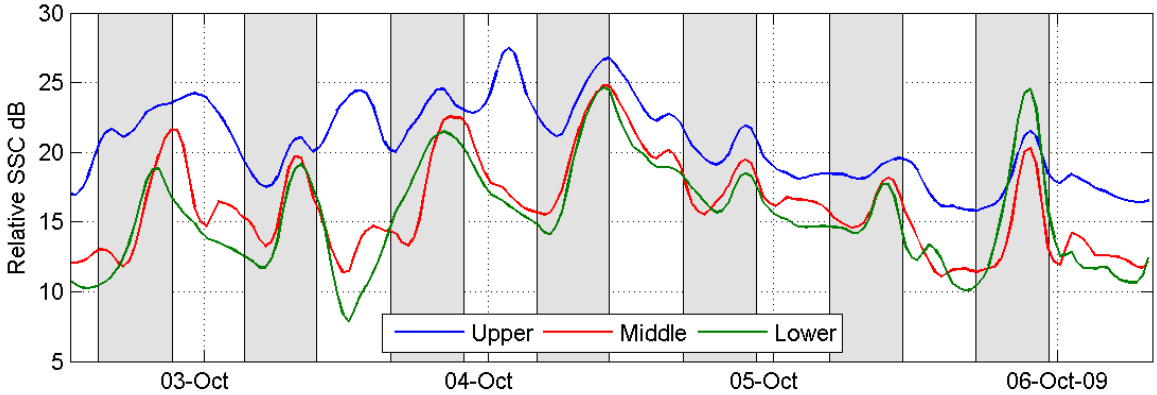

Figure 18. Relative suspended sediment concentration (SSC) from ADCP measurements at Godley Head indicating 10 to 25-fold change in SSC over a tidal cycle. *(Source: Mulgor Consulting Ltd.).*

TSS-NTU relationship

Another important aspect of the background turbidity characterisation is the relationship between the two common measures, NTU and TSS or SSC as mg/L. This relationship is often assumed to be spatially-temporally invariant. However, experience in other projects suggests this is definitely not the case and that pronounced differences in the TSS-NTU

(and light attenuation, K_d) relationship(s) exist. Failure to adequately take account of this spatio-temporal variation can result in gross errors when converting between TSS and NTU. These conversions are required in order to harmonise *measurements* in NTU with output from hydrodynamic models expressed as TSS in mg/L.

As limited as it is, the ECan. data provides evidence of spatial variation in the TSS-NTU relationship as indicated by Figure 19 and the following regression models:

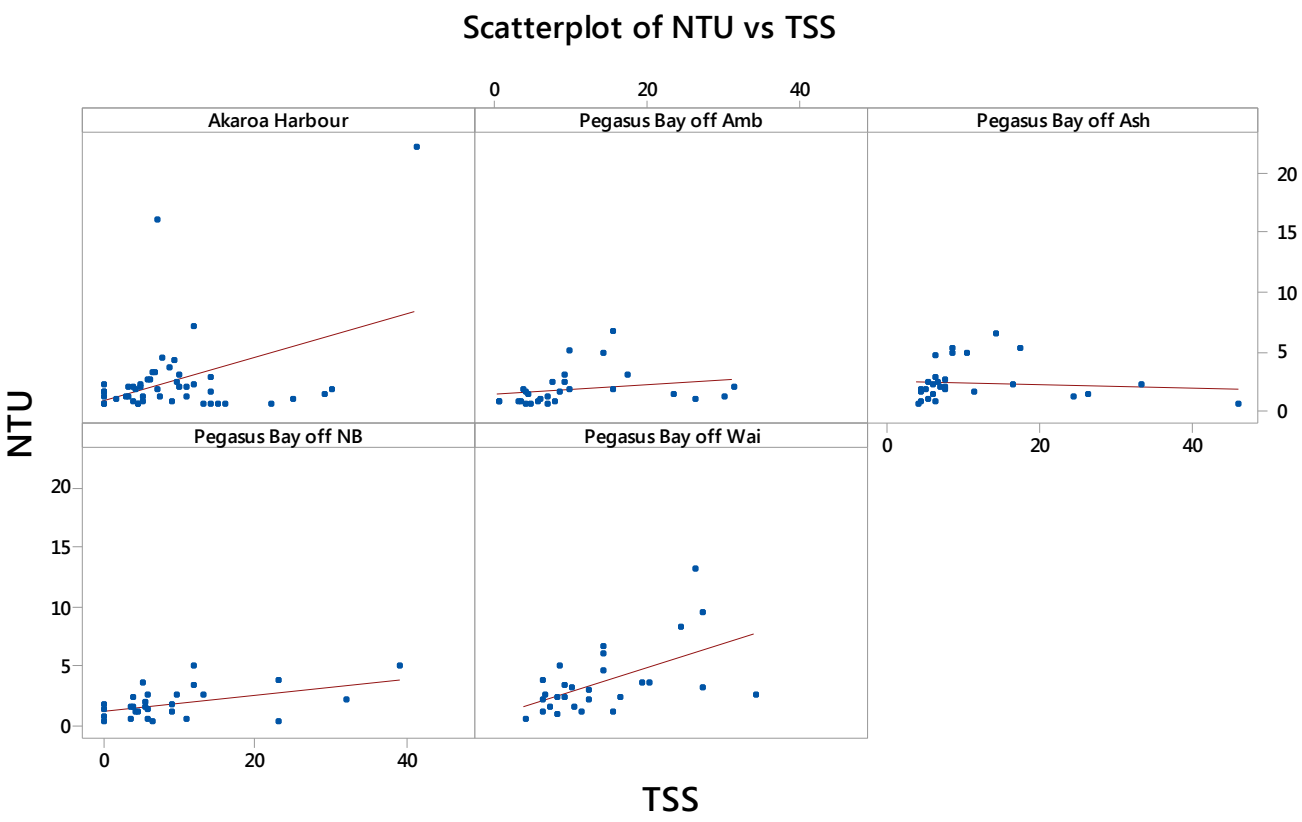

Figure 19. Relationship between NTU and TSS at ECan. off-shore monitoring sites.

```
Akaroa Harbour       NTU = 0.719 + 0.1836 TSS

Pegasus Bay off Amb  NTU = 1.266 + 0.0421 TSS

Pegasus Bay off Ash  NTU = 2.447 - 0.0144 TSS

Pegasus Bay off NB   NTU = 1.126 + 0.0699 TSS

Pegasus Bay off Wai  NTU = 0.804 + 0.1992 TSS
```

Overall, these relationships are relatively weak, but this may be due to confounding by temporal effects and the fact that turbidity is naturally low in this region. The negative coefficient in the relationship at Pegasus Bay (off-shore Ashley River mouth) is spurious and again reflects the lack of any clear signal at this site (Figure 19).

While temporal confounding may be responsible for the observed weak relationships in Figure 19, another possibility is that the turbidity is primarily *organic* in nature. However further data collection and additional laboratory analyses would be required in order to establish whether or not this was the case.

3.1 Distributional Properties

One of the main objectives of the background turbidity characterisation is the establishment of trigger values used to manage water quality during a dredging project. This is an exercise in balancing competing risks. Triggers that are set too low will result in unnecessary and expensive stoppages while triggers that are set too high will not be sufficiently protective of environmental assets. A key element of any risk analysis is *uncertainty* – if there was no uncertainty, there would be no risk. Mathematically, uncertainty is handled through the calculus of *probabilities* and *probability density functions* (pdfs) are central to the statistical treatment of data arising from chance events. Thus, if TVs are to be accompanied by probabilistic assessments of *sensitivity* (true positive rate of triggering) and *specificity* (true negative rate of triggering), then an estimate of the *pdf* of the underlying variable (TSS, NTU, SSC etc) needs to be obtained. Background data collected during the baseline monitoring period will be used for this purpose. This is a fairly straightforward task as water quality parameters are invariably right-skewed and well represented by a *log-normal* probability model (Figure 20).

The steps involved are as follows:

1. Filter the data as described in section 2.1 of this report;

2. Use maximum likelihood estimation techniques (*mle*) to estimate the parameters of either a 2 or 3-parameter log-normal probability distribution for the smoothed data generated at step 1.

3. Examine diagnostic plots such as those shown in Figure 21 together with summary goodness-of-fit statistics to confirm the adequacy of the fitted distribution.

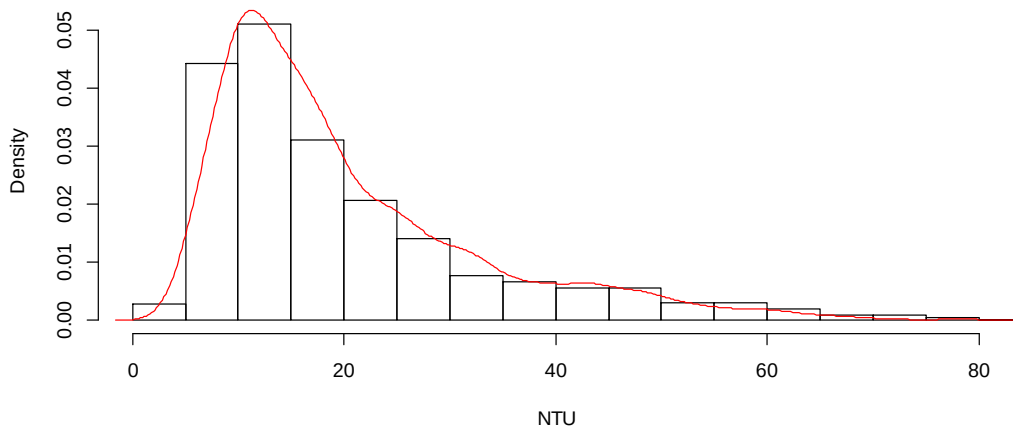

Figure 20. Empirical distribution of raw turbidity data (rectangles) with smooth estimate obtained from KZ-filtered data using 1 hr bandwidth as suggested in this report.

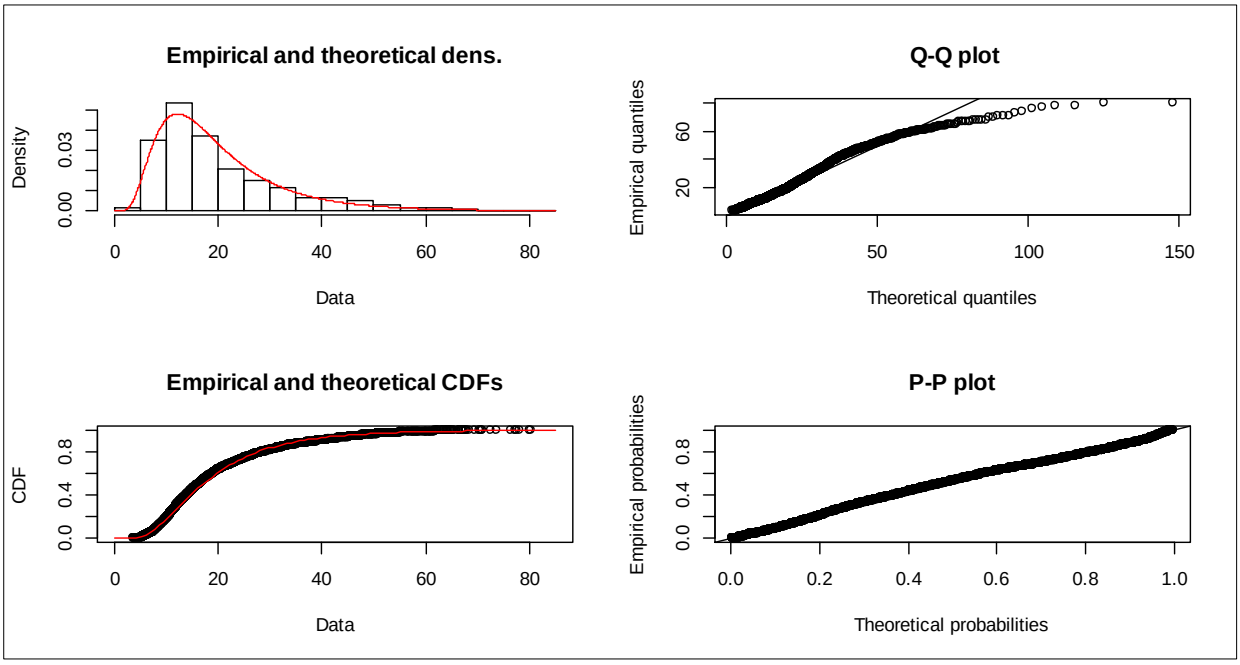

Figure 21. Diagnostic plots for fitting of log-normal distribution to KZ-filtered data using 1 hr bandwidth.

3.2 Influence of Extreme Events

The introduction to this report spent some time clarifying the distinction between *aberrant* and *outlier* as well as highlighting the possibility of bias introduced by the occurrence of an abnormally large number of extreme weather and/or ocean events during the background data collection period.

Researchers in the physical sciences often quantify extreme events by their *return period* which is the inverse of the probability that the event will be exceeded in any one year. Thus a flood having a 0.01 probability of occurring in any one year has a return period of 100 years. However, the assessment of 'extreme' is more complex than just examining the frequency of occurrence. The World Meteorological Organization's Expert Team on Climate Change Detection and Indices has defined a core set of 27 descriptive indices of extremes. The indices describe particular characteristics of extremes, including *frequency*, *amplitude* and *persistence* (Tank et al. 2009). Interestingly, it is these same concepts that have been invoked for the development of trigger values (discussed in Section 4).

The indices are readily computed using the R package `climdex.pcic` while modelling of extreme values can be achieved using the `extRemes` package developed by the Research Applications Laboratory at the National Center for Atmospheric Research, Boulder, Colorado.

At the present time, the use of extreme indices and/or more sophisticated modelling has not been used prior, during, or after dredging projects in order to make assessments of the 'representativeness' or otherwise of the snapshots in time used to describe the natural environment. Nevertheless, we think this is an important topic and one that needs to be given more attention in the environmental impact assessment process for dredging projects.

4. TRIGGER VALUES FOR IMPACT ASSESSMENT

Since its incorporation into the Australian and New Zealand Water Quality Guidelines (ANZECC/ARMCANZ 2000b) some 16 years ago, the concept of a 'trigger value' has gained widespread acceptance among natural resource managers. We commence this section with a brief review of the development and use of TVs in Australia and New Zealand.

4.1 Background

The thinking behind the TV development was that water quality assessments needed to move away from prosaic comparisons of a measured parameter against a reference as this fostered a binary "pass/fail" notion of water quality. Although TVs still rely on a numerical comparison, the subtle differences are (a) generally a tiered system of TVs is used and (b) exceedance of a TV initiates some sort of management response rather than punitive action. On this latter point, the Guidelines were quite clear: "exceedances of the

trigger values are an 'early warning' mechanism to alert managers of a potential problem. **They are not intended to be an instrument to assess 'compliance' and should not be used in this capacity**" (ANZECC/ARMCANZ 200b, page 7.4-4). Furthermore, the ANZECC/ARMCANZ (2000b) document advocated the use of locally-derived trigger values to replace *de facto* standards.

Information on New Zealand's trigger values is available on the Ministry for Environment's website although it would appear this has not been updated for some time (http://www.mfe.govt.nz/publications/fresh-water-environmental-reporting/trigger-values-rivers-may00/trigger-values-rivers).

According to the New Zealand Ministry for Environment:

> *"Running medians of water quality data measured in monitoring programmes may be compared with these trigger values. If the median value of a water quality variable for a particular site exceeds the trigger value, then it is intended to "trigger" a response on the part of water managers, which might be to initiate special sampling or carry out an investigation of reasons for the degraded water quality … The 80 percentile has been (arbitrarily) chosen where high concentrations or values indicate degraded water quality, and the 20 percentile where a low benchmark is appropriate".*

The basis of this procedure was the following suggestion by Fox (1999) which was subsequently adopted as the *de facto* method of setting TVs for physico-chemical stressors (ANZECC/ARMCANZ 200b):

> *"A trigger for further investigation will be deemed to have occurred when the median concentration of n independent samples taken at a test site exceeds the eightieth percentile of the same analyte at a suitably chosen reference site".*

As noted in Box 1, the comparison procedure is adaptive in that both the median being tested and the TV are continually updated (Figure 22).

Box 1. Extract from section 7.4.4 of ANZECC/ARMCANZ (2000b) - *Comparing test data with guideline trigger values*

The computational requirements of the approach are minimal and can be performed without the need for statistical tables, formulae, or computer software. As demonstrated later in this section, the temporal sequence of trigger events is readily captured in a simple plot. It should be understood that the trigger protocol is responsive to shifts in the location (i.e. 'average') of the distribution of values at the test site. While differences in shape of the reference and test distribution may be important in some instances, this is a secondary consideration that is not specifically addressed by this protocol. It is also important to note that the role of the 80th percentile at the reference site is to simply quantify the notion of a *'measurable perturbation'* at the test site. The protocol is not a statistical test of the equivalence of the 50th and 80th percentiles *per se*. The advantages of using a percentile of the reference distribution are 1) it avoids the need to specify an absolute quantity and 2) because the reference site is being monitored over time, the trigger criterion is being constantly updated to reflect temporal trends and the effects of extraneous factors (e.g. climate variability, seasonality). Implementation of the trigger criterion is both flexible and adaptive. For example, the user can identify a level of routine sampling (through the specification of the sample size *n*) that provides an acceptable balance between cost of sampling and analysis and the risk of false triggering. The method also encourages the establishment and maintenance of long-term reference monitoring as an alternative to comparisons with the default guideline values provided in Section 3.3 that do not account for site-specific anomalies.

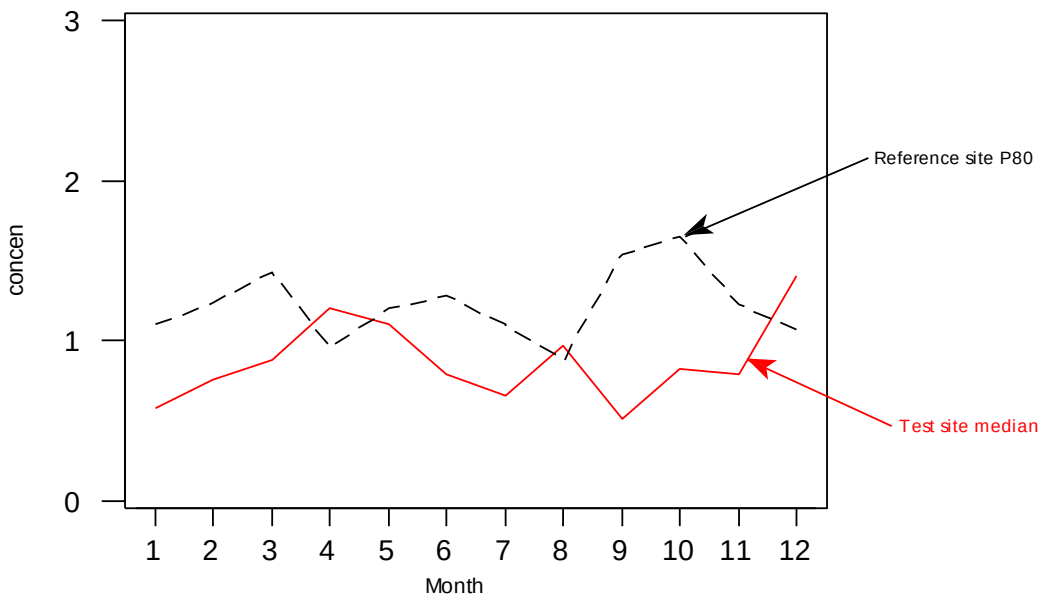

Figure 22. Illustration showing the dynamic nature of the P_{50}-P_{80} comparison procedure for physico-chemical stressors *(Taken from Fox 1999)*.

Fox (1999) examined the statistical performance characteristics of the P_{50}-P_{80} comparison procedure and concluded that, overall it performed reasonably well. For example, with as few as 5 observations at the test site, the procedure had a 5.8% Type I error rate and a power of 50%.

Since its inclusion in the National Water Quality Guidelines (ANZECC/ARMCANZ 200a,b) the P_{50}-P_{80} comparison procedure has been widely used in many environmental monitoring contexts and has underpinned most water quality assessments during capital dredging projects in Australia. However, these dredging projects have mostly adopted a *static* version of the procedure outlined above by treating the 80th. (and other reference) percentiles as *fixed* in time. Furthermore, there has been a tendency to widen the scope of this comparison procedure by including the additional concepts of *frequency* and *duration* of exceedance as suggested by McArthur et al. (2002). As noted in section 3.2, the incorporation of *frequency* and *duration* makes sense and is used in other disciplines. Unfortunately, there are gaping holes in the McArthur et al. (2002) methodology - a point conceded by McArthur himself (*pers. comm. February 2011)*. These issues are addressed in later in section 4.4.

4.2 Purpose

The rationale behind the development and use of Trigger Values based on an analysis of the background data record is straightforward and guided by the following logic:

- Physical-chemical stressors are naturally present;
- The existence and persistence of valued ecosystem components (for example, benthic infauna, macroalgae, molluscs etc.) is evidence of resilience to all levels and exposure patterns of all naturally-occurring stressors – including episodic extremes; and therefore
- Provided anthropogenic inputs of a stressor do not elevate stressor levels beyond some high-order percentile of 'background' or 'natural' levels, then the system will be protected.

While this line of reasoning has intuitive appeal and seems reasonable, there are nevertheless some implicit and largely untested assumptions and unresolved questions. The most vexing are: *what percentile should be chosen to achieve the desired outcome?; what is the interplay between level of exceedance and patterns of exposure?; and how will we know the activity hasn't caused an effect?*

Although there are gaps in our knowledge and understanding, these are slowly being filled. For example, the Port of Melbourne (PoMC) used risk-based turbidity triggers for its channel deepening project (Environmetrics Australia 2007). Unlike conventional light-based thresholds (based on a fixed percentage surface irradiance), these new trigger values were determined *probabilistically* in such a way that as long as turbidity levels remained below the threshold, the *probability* that seagrasses would be protected was some arbitrarily high value. This pioneering work was further advanced during the course of the Gladstone Port Corporation's (GPC) Western Basin Dredging and Disposal Project (http://www.gpcl.com.au/development/western-basin-dredging-and-disposal-project) whereby independent research was undertaken to establish both a *quantum* and *duration* for a minimum light level required to protect sensitive seagrass species. This enabled the development of companion turbidity trigger values that ensured this minimum light level was achieved at all times. Both the PoMC and GPC projects were completed with no adverse environmental outcomes. In particular, the GPC approach (http://environmetrics.net.au/index.php?news&nid=77) has been widely recognised as the 'gold standard' in turbidity and light monitoring for large capital dredging projects. However, its successful implementation is dependent upon all of the following:

(i) a clearly identified sentinel environmental asset to be protected at all costs;

(ii) at least one year's monitoring data of background conditions on finely resolved spatial-temporal scales;

(iii) contemporaneous monitoring of the condition of the environmental asset to be protected and its response to naturally occurring changes in ambient conditions;

(iv) contemporaneous measurements of benthic light regime;

(v) completion of successful research program linking the asset's response to changes in natural conditions.

These requirements do not all exist for LPC's proposed activity and hence the approach is not feasible.

4.3 Limitations

As alluded to in the previous section, the background-based trigger-value approach has embedded within it the assumption that the marine ecosystem will be protected as long as turbidity levels are kept below some statistically-derived threshold of pre-dredging, background conditions. Although there is no hard evidence to support this assumption, there is a *lack of evidence* of any harm during projects which have used TVs to manage turbidity. Furthermore, Ross Sneddon (Marine Ecologist for the project), considers the assumption is appropriate for the CDP and a turbidity trigger based on predicted conditions should be protective of the ecological communities within the wider areas of the project location (pers comm).

Although the science has progressed to the point where the specific light requirements have been quantified for a single species, at a single location it is unlikely that this method will achieve widespread uptake by virtue of the time-consuming and expensive R&D required for its implementation. It also requires identification of an indicator species, which will act as the 'canary in the cage' for the wider ecosystem. For the time being, and in the absence of a more scientifically defensible approach, the use of percentile-based TVs is recommended – albeit with modification as detailed in section 4.4.

As an endnote to this review of turbidity TVs, the issue of *ecological significance* was countenanced when the TV method adopted by ANZECC/ARMCAZ (2000b) was first mooted (Box 2).

Box 2. Trigger-value development: Consideration of ecological significance issue (Fox 1999).

> "The concept of 'ecological significance' is elusive but is implicit in our trigger rule. While we have not sought to provide a definition, it is impossible to escape a quantification of a 'significant perturbation' (my terminology). Our de facto definition of a significant perturbation corresponds to the size of the shift between the 50th. and 80th. percentiles at a reference site. This definition is arbitrary, although seems to have broad acceptance and intuitive appeal among experts. Whether such a shift represents an ecologically significant change is a matter for others to debate, although universal agreement is unlikely to emerge. It has also been suggested that the median – P80 trigger comparison being advocated here is flawed since like statistics are not being compared (eg. the median at the reference site is not being compared with the median at the test site). This is not the case. We are not attempting to make any inference about the similarity or otherwise between the median at the test site and the 80th. percentile at the reference site. The sole function of introducing the 80th. percentile at the reference site is to provide a measure of shift in statistical location of the test distribution that is generally regarded as being representative of a significant perturbation. The advantage of the suggested method is that the trigger threshold is not time invariant and is regionally sensitive".

4.4 Sample and Time-based Percentiles

The case for the retention of percentile-based TVs was made in previous sections of this report although it has been acknowledged that refinements are required to address deficiencies of the method described by McArthur at al. (2002). However, before doing so we outline some important concepts.

Preliminaries

We first consider the duality between a *time-based* percentile and a *sample* percentile using the depiction of a turbidity time-series in Figure 23.

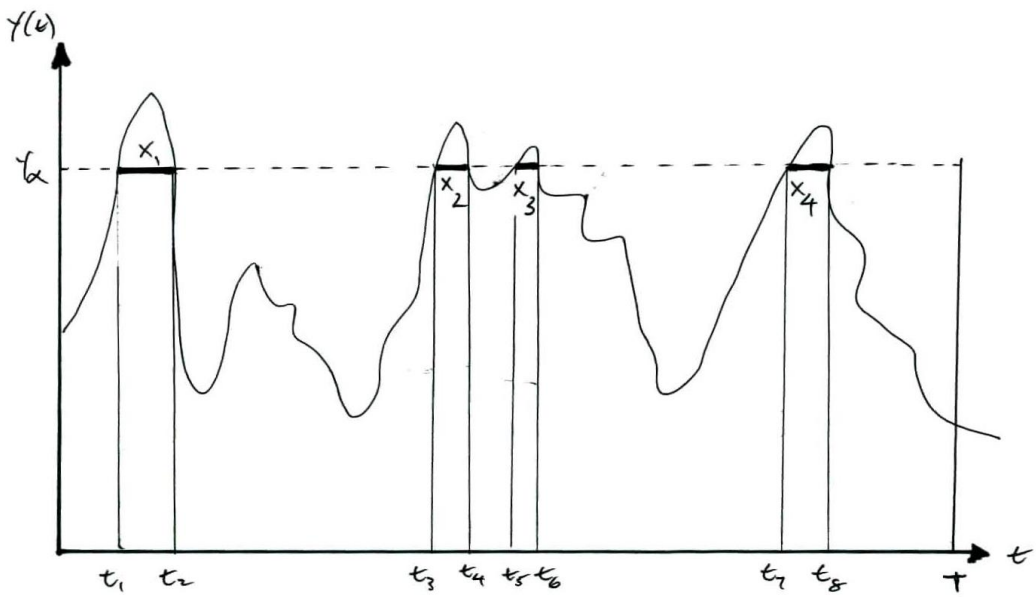

Figure 23. Representation of a turbidity time-series. The α percentile (horizontal broken line) defines the *intensity*; individual *durations* are denoted by X_i; and the number of durations in the interval [0,T] is the *frequency*.

The statistical definition of the α percentile, Y_α is that $\alpha \cdot 100\%$ of all (sample) values are numerically less than or equal to Y_α or conversely $(1-\alpha)\cdot 100\%$ are greater than Y_α. Intuitively, one would expect that the proportion of *time* that a *continuous* function $Y(t)$ exceeded Y_α would also be $(1-\alpha)\cdot 100\%$. To see that this intuition is correct, consider the regular *discrete* sampling of $Y(t)$ as shown in Figure 24. In this case the time interval is divided into 100 equally spaced sampling points. The proprtion of time the *continuous* trace is above a nominal threshold of 20 units is computed to be 71.4%. The number of discrete sample points which are above the 20 unit threshold is 69 and so the sample estimate of this proportion is 69/100 = 0.69.

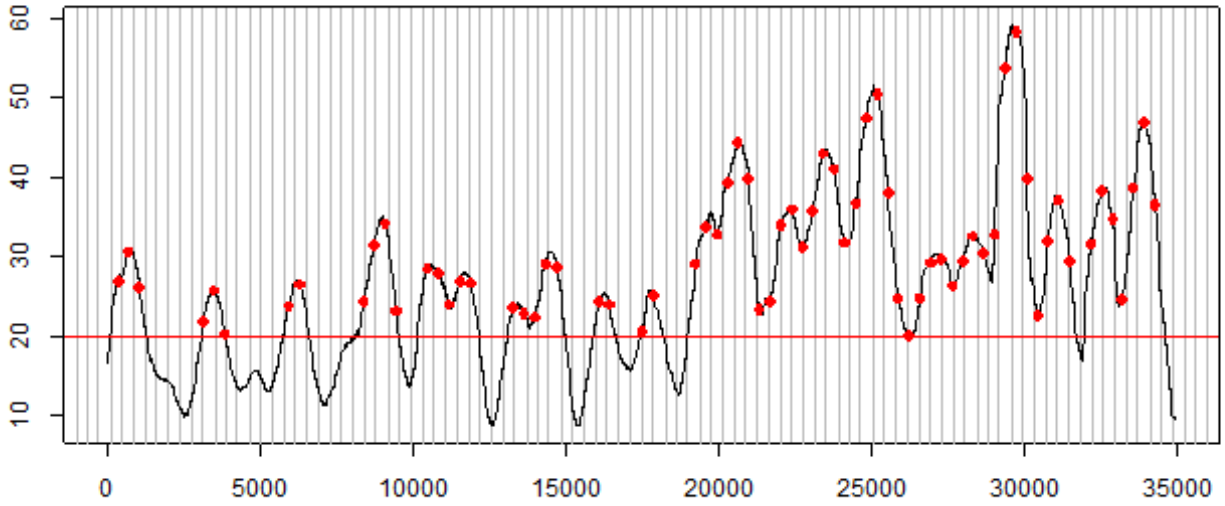

Figure 24. Discrete sampling (solid circles) of continuous signal *Y(t)* (black curve). Nominal threshold of 20 (red line).

The asymptotic equivalence of the *time-based* percentile and the *discrete sample* estimate is readily established as follows.

Consider a continuous signal $Y(t)$ over the interval $[0 \le t \le T]$. Let Y_α be the $\alpha \cdot 100\%$ percentile of $Y(t)$ and suppose the number of exceedences of the form $Y(t) > Y_\alpha$ is *k*. Furthermore, let the *duration* of the ith exceedence be X_i. A time-based defintion of Y_α thus implies:

$$\frac{\sum_{i=1}^{k} X_i}{T} = (1-\alpha) \qquad (7)$$

Next, we can write $X_i = t_{2i} - t_{2i-1}$; $i = 1,\ldots,k$ where the $\{t_{2i-1}\}$ are the times at which the ith exceedence commenced and the $\{t_{2i}\}$ are the times at which the ith exceedence ceased. For regular discrete sampling we divide the time interval $[0 \le t \le T]$ into a total of *N* intervals each of width Δt and hence

$$T = N \cdot \Delta t \qquad (8)$$

Now the *number* of discrete sampling points n_i falling in an interval of length X_i is

$n_i \approx \frac{X_i}{\Delta t}$ and so the overall *proportion* of *samples* (call it *p*) exceeding Y_α is $\frac{\sum_{i=1}^{k} n_i}{N}$.

We therefore have:

$$p = \frac{\sum_{i=1}^{k} n_i}{N} \approx \frac{\sum_{i=1}^{k} \frac{X_i}{\Delta t}}{N}$$

$$= \frac{\sum_{i=1}^{k} X_i}{N\Delta t} = \frac{\sum_{i=1}^{k} X_i}{T} \qquad (9)$$

Comparing equations 7 and 8 we see that $p \approx (1-\alpha)$ thus establishing the equivalence of *discrete* percentile estimates with their time-based counterparts. Note, the approximation in Equation 9 improves as *N* increases with equality attained as $N \to \infty$ $(\Rightarrow \Delta t \to 0)$.

4.5 The IFD Approach

It is now generally accepted that the trigger-value process has three dimensions: the trigger *level* (referred to by McArthur et al. 2002 as the *intensity*); the *frequency* of exceedance; and the *duration* of exceedance. This is similar to the World Meteorological Organization's classification of extremes using the concepts of *frequency*, *amplitude* and *persistence* and the Canadian Council of Ministers of the Environment's (CCME) Water Quality Index (WQI) based on the concepts of *scope* (percentage of indicators not meeting the relevant water quality objective)*, frequency* (percentage of comparisons where the guideline was not met)*,* and *amplitude* (a normalised measure of the extent to which failed comparisons deviated from the guideline) (Canadian Council of Ministers of the Environment 2001). The CCME thus talks of a water quality measurement in a 3-D *exceedance space* (Figure 25).

The previous discussion concerning the asymptotic equivalence of discrete sample percentiles and continuous, time-based percentiles is important as it makes clear the relationship between the three quantities *intensity*, *frequency*, and *duration* as used by McArthur et al. (2002).

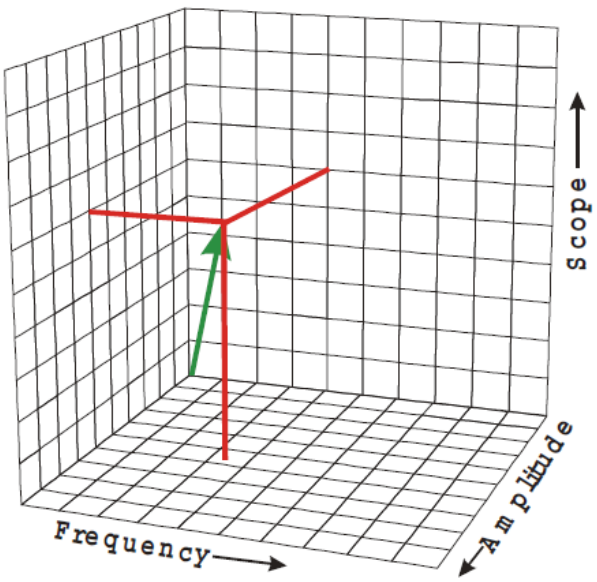

Figure 25. Representation of CCME WQI in three-dimensional exceedance space. *(from Canadian Council of Ministers of the Environment 2001).*

From Equation 7 we have immediately:

$$k \cdot \bar{X} = (1-\alpha) \cdot T \tag{10}$$

and it is thus apparent that the quantities $k, \bar{X},$ and α *cannot* all be varied independently of each other while still maintaining the equality of Equation 10 as was assumed by McArthur et al. (2002). Figure 26 illustrates this point using an indicative turbidity time-series. For the 20% exceedance rate, there were 18 exceedances of relatively small duration; at 50% and 60% there were 23 exceedances and durations tended to be longer; while at 80% there were fewer exceedances but of longer duration. Further analysis revealed some well-defined relationships between exceedance rate and average duration (Figure 27) and exceedance rate and number of exceedances (Figure 28). These results make intuitive sense. For example, when the exceedance rate is zero, *k* *must* also be zero but the duration curve is asymptotic at this point since, from Equation 10 , $\bar{X} = \frac{(1-\alpha) \cdot T}{k}$. At the other end of the scale, an exceedance rate of 100% means $Y(t)$ is always above the threshold and so there is a single exceedance of duration T : a result confirmed by Equation 10 with $\alpha = 0$ and *k=1*. Between these two extremes, *k* can only increase which is clearly evident in Figure 28.

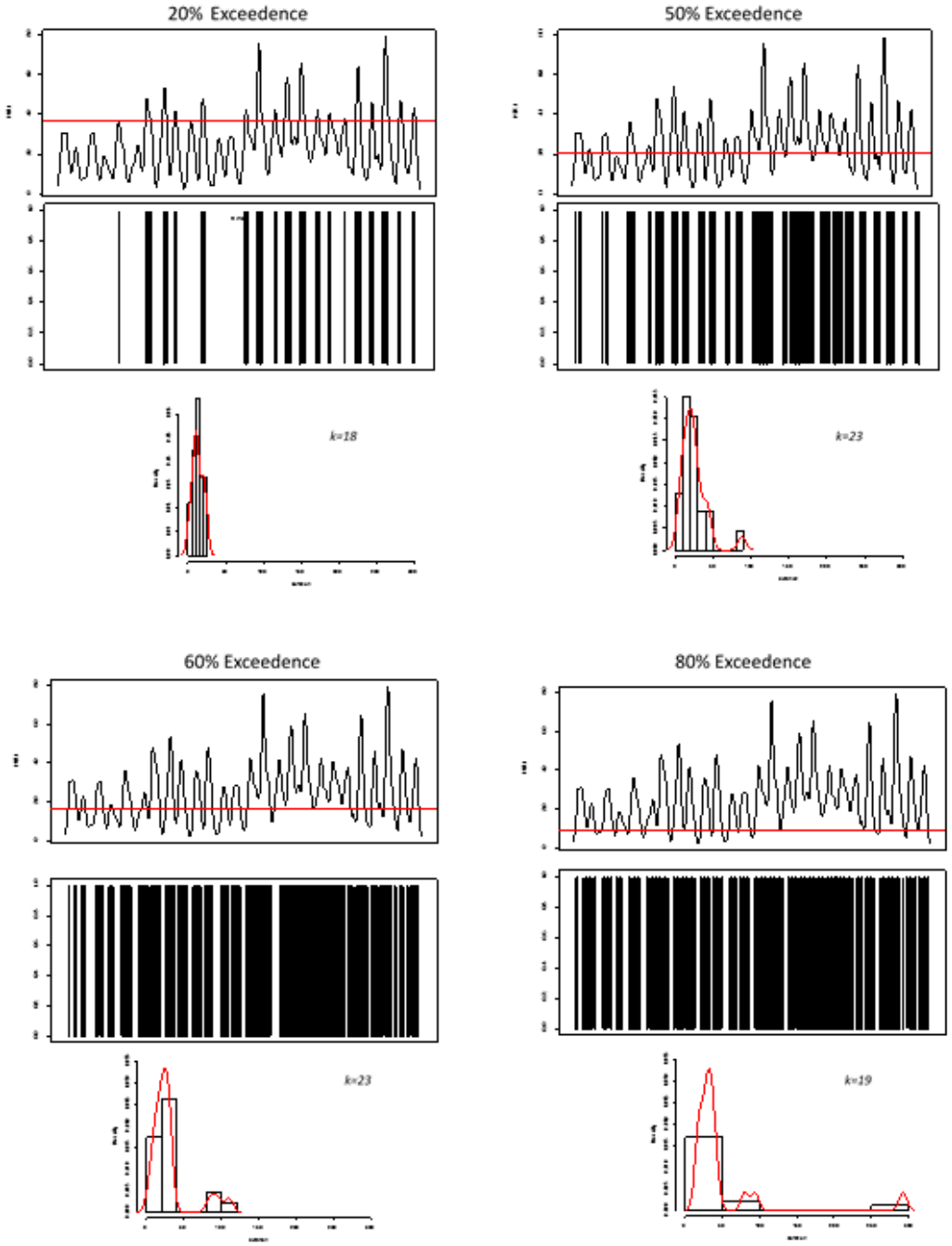

Figure 26. Effects of changing *intensity* (expressed as exceedance rate and depicted by red horizontal line) on *frequency* (*k* – the number of vertical black bars) and *duration* distributions.

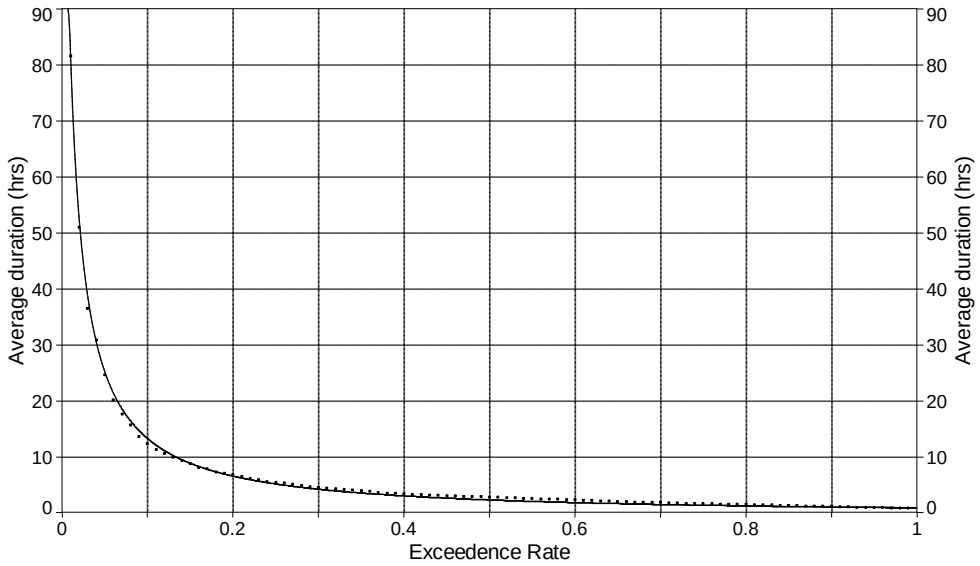

Figure 27. Empirical relationship between exceedance rate and average duration for the time-series used in Figure 26.

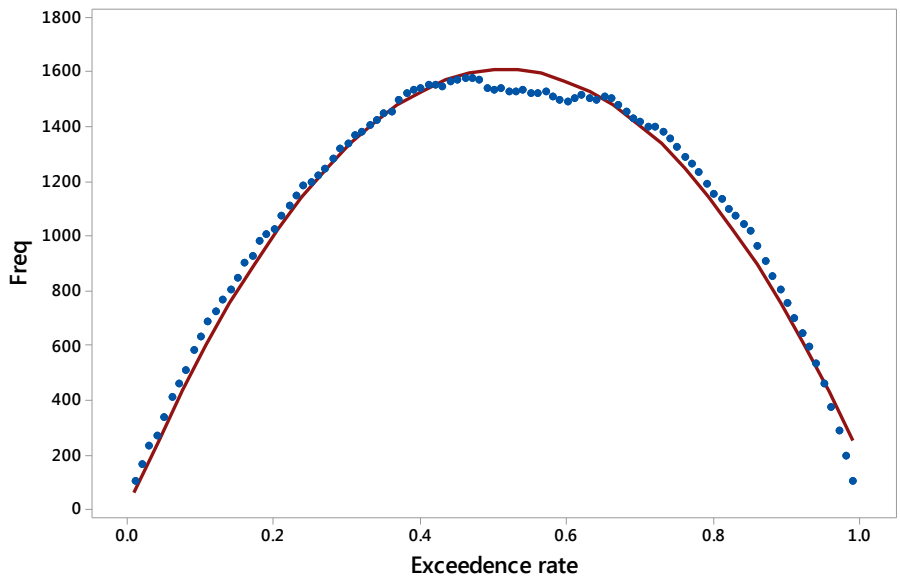

Figure 28. Empirical relationship between exceedance rate and number of exceedances for the time-series used in Figure 26.

An outline of the McArthur et al. method is described next.

The McArthur et al. (2002) method

The McArthur et al. (2002) approach was an attempt to incorporate the extra dimensions of *frequency* and *duration* of exceedance into the use of trigger values. While this makes sense and potentially increases the utility of the trigger-value method, the proposed methodology was flawed in so far as it extended the percentile concept to all three dimensions (intensity, frequency, and duration) *independently of each other*. As was explained in section 4.5, it is impossible to vary all three of these while simultaneously maintaining the equality of Equation 10.

The first step in the IFD method as proposed by McArthur et al. (2002) was to compute the 95th. and 99th. percentiles of the turbidity (SSC) distribution. The 95th. percentile was referred to as the *threshold concentration* and the 99th. percentile the *intensity guideline* (Figure 29).

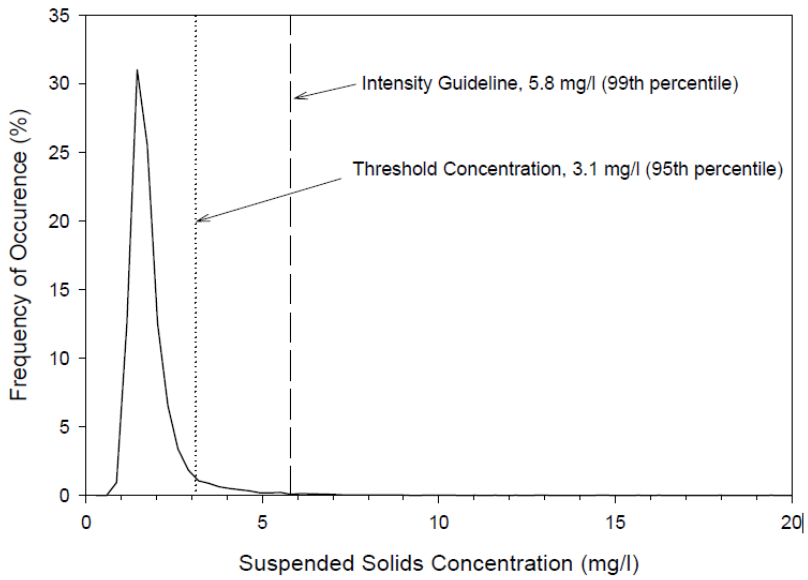

Figure 29. Distribution of suspended sediment concentrations used to derive the intensity guideline and threshold concentration (Source: McArthur et al. 2002).

The data was then re-examined to determine the distribution of all *duration* events during which the threshold value was exceeded. The 95th. percentile of this *duration distribution* was then used as a *duration guideline value* (Figure 30). The procedure thereafter is opaque and not well defined. According to McArthur et al. (2002):

> *"To develop frequency guidelines, all events exceeding the threshold value were grouped into classes by duration. For each duration class, a frequency distribution was developed over the time frame of interest. In*

this case, the time frame selected is a week. The 95% confidence limit was then selected as the total allowable frequency. The principle behind the guidelines requires that natural SSCs plus that due to disposal cannot exceed the natural bounds. Therefore, the mean frequency is subtracted from the total allowable frequency to determine the frequency guideline for dredged material disposal".

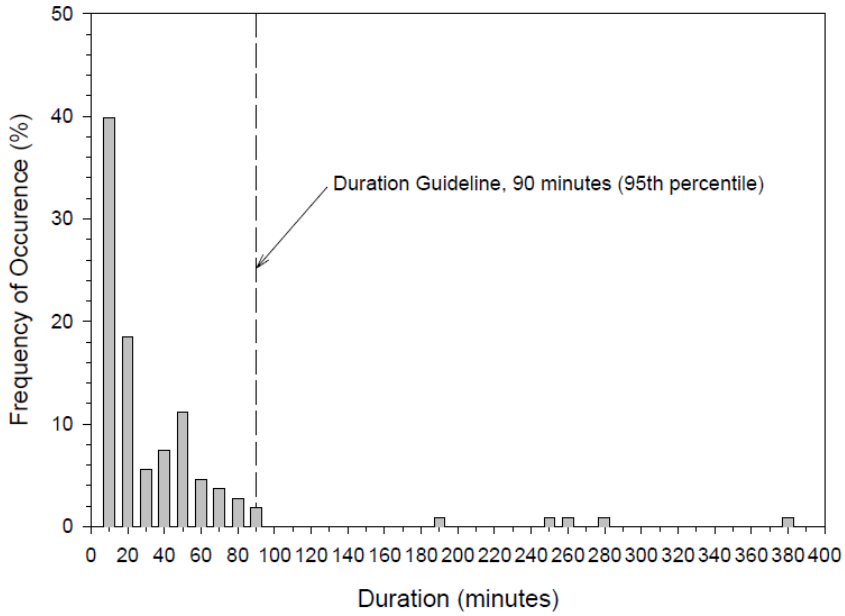

Figure 30. Frequency distribution of exceedance durations used to establish a duration guideline value.

The mechanics of the above method are not well understood and a conversation with the lead author in 2011 resulted in the frank admission that he couldn't recall himself as the work had been done 10 years prior and had never been implemented. Although the lack of application and trialling of the methodology was foreshadowed in the original paper (*"It has yet to be determined exactly how the guidelines will be used to manage the Miami ODMDS"*, McArthur et al. 2002) it was nevertheless picked up and used without modification in a number of major dredging projects in Australia (Darwin Harbour, Cape Lambert, and Dampier) (Appendix 1).

In view of the developer's own misgivings and the lack of 'road-testing', it is our view that neither the McArthur et al. (2002) method as originally proposed or projects that made use of it, should be cited as examples of "best practice" for turbidity monitoring during large-scale dredging projects. Accordingly, we have sought to modify the basic approach to overcome these limitations.

4.6 A Modified IFD Approach

It is worth re-iterating the guiding principle adopted by McArthur et al. (2002): "*the guidelines requires that natural SSCs plus that due to disposal cannot exceed the natural bounds".*

Further analysis of the *background* data presented in Figure 26 illustrates how both duration and frequency vary as a function of exceedance rate. In essence, for a given environment, there is only one combination of average duration and average frequency that correspond to a given exceedance rate – as defined by the points *on the curve* in Figure 31. The MacArthur et al. (2002) approach *independently* manipulates frequency and duration but this results in an infeasible outcome (in the background IFD space) that cannot honour the guiding principle stated above (Figure 31).

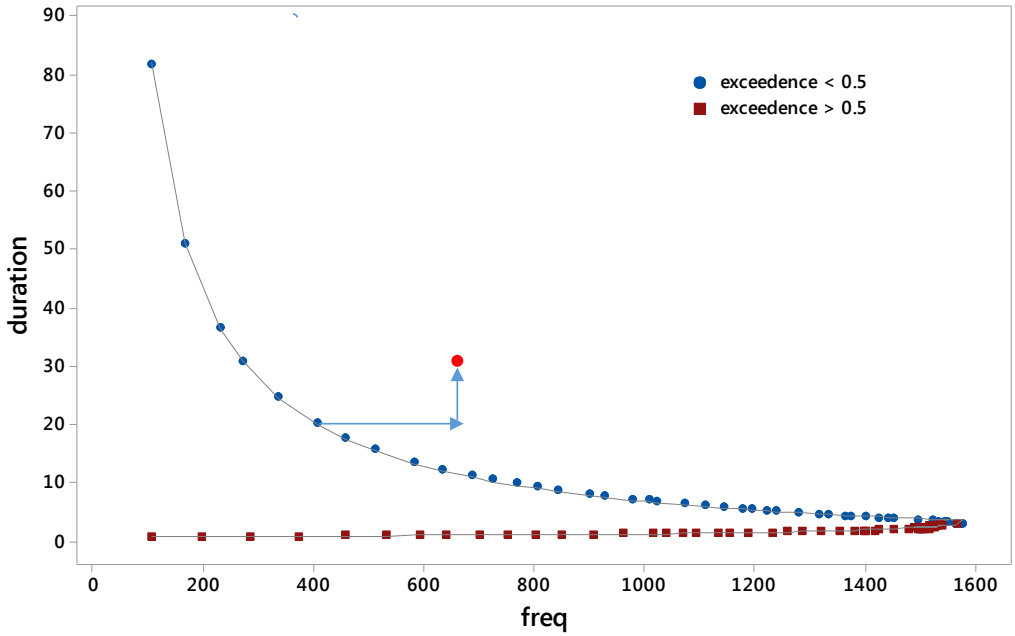

Figure 31. Illustration showing the duration-frequency relationship as a function of exceedance rate for background data. The McArthur et al. (2002) approach takes an arbitrary point on this curve and simultaneously increases its frequency and duration (blue arrows) in an attempt to represent the effects of dredging. However, this results in a combination that is *never* observed under 'natural' conditions and thus represents an outcome that cannot honour the objective *"that natural SSCs plus that due to disposal cannot exceed the natural bounds".*

We believe a more meaningful way of implementing the IFD approach is to incorporate the hydrodynamic model output to determine how the curve in Figure 31 is modified. This also recognises that dredging activities will, as predicted by the modelling, have an effect on turbidity levels near to the activity, but will not at distant locations.

Preliminary investigations using simulated data illustrating this effect are shown in Figure 32.

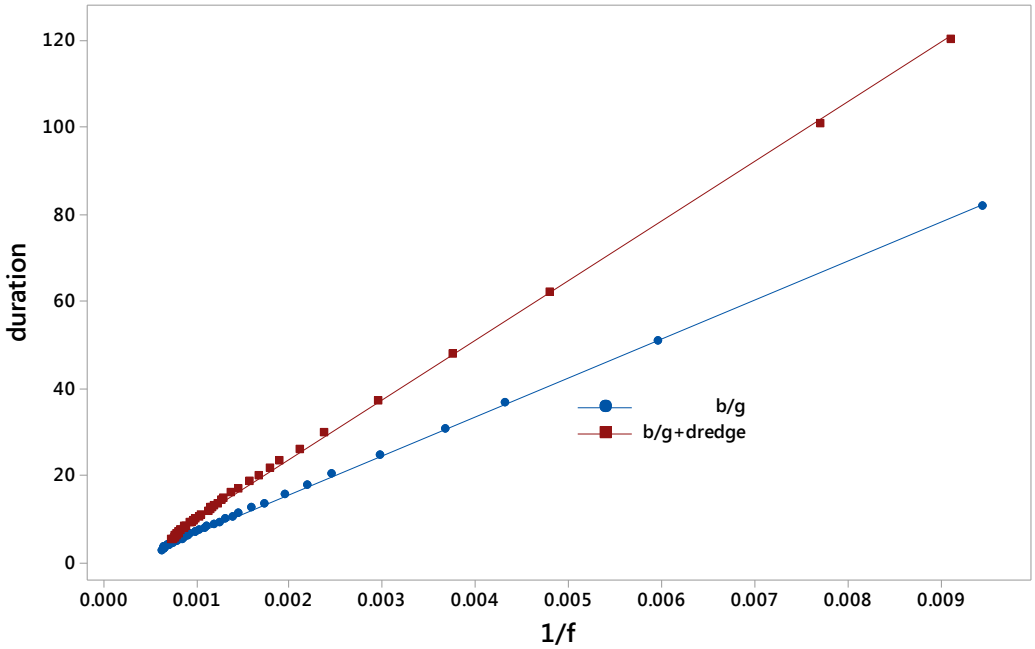

Figure 32. Modification to the IFD approach based on predicted increase in turbidity.

Thus, the establishment of trigger values in the 3-dimensional IFD space would proceed as follows:

1. fix the exceedance rate at some agreed value;
2. identify the position on the red line in Figure 32 corresponding to the exceedance rate set in 1.;
3. the TVs for frequency and duration are then read off at the position identified in 2.

The data and information needed to perform steps 1-3 above are:

- annual time series of background turbidity data (in NTU);
- routine contemporaneous measurements of TSS and NTU at selected monitoring sites;
- modelled *incremental* TSS on appropriate spatial and temporal scales covering the physical extent of the dredge plume over the project duration.

Finally, because of the complex interplay between all three components (intensity, frequency, and duration) it is not possible to predict in advance how these measures will

change after taking account the effects of dredging. Figure 33 illustrates this point which shows frequency and duration curves for both background turbidity and background + dredging as a function of exceedance rate using the same data as above.

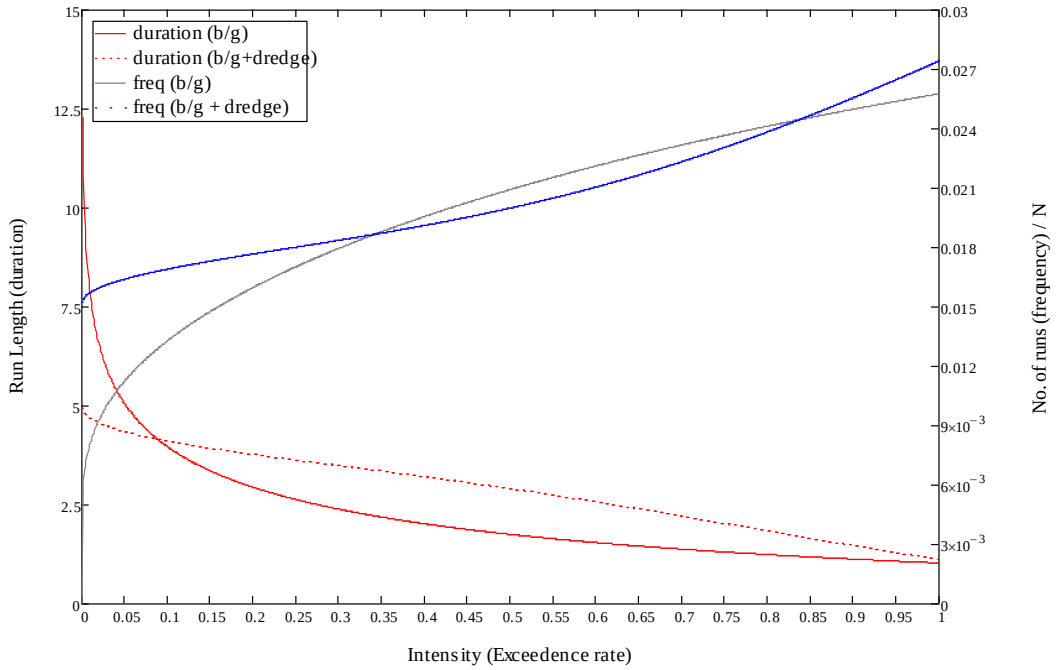

Figure 33. Duration curves (red) and frequency curves (blue) for background and background + dredge-related turbidity as a function of exceedance rate.

Figure 33 shows that there is a multitude of possibilities, not all of which are expected. For example, with a 10% exceedance rate, the average background duration would not change although the number of exceedances (as a fraction of the total data record) would increase from about 0.013 to 0.017 (about 140 additional exceedances of individual turbidity readings/yr). On the other hand, if the exceedance rate is not to exceed 5%, then compared to background, an additional 175 exceedances per year would be permitted but the average duration of those exceedances would reduce from 5 hours to 4 hours.

It is important to note that these results are illustrative only and pertain to one particular set of (synthetic) data. Further work using actual data sets obtained from the environment to be monitored is required in order to ensure the integrity and reliability of the proposed strategy.

APPENDIX 1: SUMMARY OF TRIGGER VALUES USED IN RECENT AUSTRALIAN DREDGING PROJECTS

Summary of water quality trigger values in past and proposed dredging and dredge material placement projects *(Source: SKM 2013).*

Project	Trigger Value	Program Specifications
Port of Hay Point Apron Areas and Departure Path Capital Dredging Project (QLD)	100 NTU	Trigger values set at 100 NTU over a continuous period of six hours at two fringing reefs.
Hay Point Coal Terminal Expansion Phase 3 (QLD)	110 NTU (derived as 80th percentile of baseline turbidity at the compliance site plus 100 NTU)	Trigger value of 110 NTU based on a 6 hour daily median during daylight hours as a daily trigger, the occurrence of 4 daily triggers within any 7 day period constituted an exceedance.
Western Basin Dredging and Disposal Project (QLD)	Level 1 – 80th percentile (NTU) Level 2 – 95th percentile (NTU) Level 3 – 99th percentile (NTU)	Site -specific turbidity trigger values with different levels of reporting ranging from 2 to 46 NTU (dry season) averages over 6 hours. Level 1 and Level 2 triggers internal and external reporting triggers and a Level 3 action trigger when eight continuous values (48 hours) exceed the trigger value.
	PAR less than 6 mol/m^2/d	Light-based trigger values established for seagrass sites based on a two-week moving average. Light-based triggers only apply during the period of active seagrass growth.
Townsville Marine Precinct Project (QLD)	109 NTU	Trigger values used were based on ANZECC/ARMCANZ and QWQG Central coast guidelines. More specific criteria of consecutive minutes above 109 NTU and the number of times incidences that were allowed per week (10 and 20 minute exceedances), as well as number of incidences allowed during the dredging project (30 minute to 12 hour exceedance).
Cairns CityPort Dredging Project (QLD)	35 NTU	A threshold of 35 NTU was set for waters around seagrass receptors, outside of this receptor areas a threshold of ambient plus 100 per cent was adopted as a management trigger based on a period of greater than 6 hours.
Rosslyn Bay State Boat Harbour Maintenance Dredging (QLD)	42.9 mg/L at Bluff Rock (80th percentile) 31.4 mg/L at Wreck Point (80th percentile)	Proposed trigger value for 2012 maintenance dredging is the 80th percentile of baseline measurements.
Port of Dampier Dredging Project (WA)	10 mg/L (offshore) 35 mg/L (inshore)	Frequency of exceedances of 10 mg/L TSS (offshore) and 35 mg/L TSS (inshore) for various durations (hrs) in a period of a month, depending on impact zone.

Project	Trigger Value	Program Specifications
Fremantle Harbour Dredging Program (WA)	PAR	Water quality triggers based on minimum light requirement for seagrasses and corals for boundaries of zones of Loss, Effect and Influence. Tiered management responses based on trigger levels with no reduction in coral or seagrass cover in zone of Influence.
Pluto LNG Dredging Program (WA)	95th percentile (NTU)	Exceedances of 95th percentiles for intensity, duration and frequency of SSC (derived from the baseline data) for different zones.
Cape Lambert – Port B (WA)	Level 1 80th percentile (NTU) Level 2 95th percentile (NTU) Level 3 99th percentile (NTU)	Exceedances of NTU from potentially impacted sites versus reference sites at different percentiles of water quality tested over a frequency of exceeding daily values for 14 days in 21 day duration. Management responses of Level 1, 2, and 3 include: exceedance of Level 1 meant reducing dredging, Level 2 meant relocating the dredge, and Level 3 meant reducing dredging further.
Port Hedland RGPS (WA)	80th percentile (NTU) 95th percentile (NTU)	Exceedances of 80th and 95th percentiles of background turbidity data (expressed as NTU levels), with the percentile chosen depending on tide, season and impact zone.
Gorgon Project (WA)	25 mg/L TSS for 2 in 6 days 10 mg/L TSS for 7 in 21 days 5 mg/L for 20 in 60 days	TSS triggers represent short-term, medium-term and long-term water quality criteria. These criteria were applied to the Zone of Moderate Impact and Zone of Influence. For the sake of ongoing monitoring of *in situ* loggers a TSS/NTU conversion was used.
Wheatstone Project (WA)	NTU to be confirmed, project in approval stage.	Exceedance of NTU at varying levels for different periods of time (expressed as a fraction of days above NTU values depending on management zone); trigger values being derived from data on Gorgon Project coral morality.

5. References

ANZECC and ARMCANZ (2000a). National Water Quality Management Strategy, Paper No. 4 Australian and New Zealand Guidelines for Fresh and Marine Water Quality Volume 1: The Guidelines (Chapters 1–7), Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand.

ANZECC and ARMCANZ (2000b). National Water Quality Management Strategy, Paper No. 7a Australian Guidelines for Water Quality Monitoring and Reporting, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand.

Close, B. and Zurbenko, I. (2016). kza: Kolmogorov-Zurbenko Adaptive Filters. R package version 4.0.0. https://CRAN.R-project.org/package=kza

Canadian Council of Ministers of the Environment (2001) Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index 1.0, Technical Report. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Environmetrics Australia (2007). Statistical Aspects of Turbidity Monitoring: Setting Environmental Limits. Report prepared for Port of Melbourne Corporation, April 2007. (Available at http://environmetrics.net.au/index.php?p=2_1).

Falkenberg, LJ. and Styan, CA. (2014). Too much data is never enough: A review of the mismatch between scales of water quality data collection and reporting from recent marine dredging programmes. *Ecological Indicators*, 45:529–537.

Fox, DR. (1999) An assessment of the proposed trigger method for physical-chemical stressors. Unpublished report, CSIRO, Perth, Western Australia.

Gibbs, MT. (2013). Environmental perverse incentives in coastal monitoring. *Marine Pollution Bulletin*; 73:7–10.

Hlawatsch F, Boudreaux-Bartels GF. (1992). Linear and quadratic time-frequency signal representations. *IEEE Signal Process Magazine*, 9:21–67.

Legg, CJ. And Nagy, L. (2006). Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management*, 78:194–199.

McArthur, C., Ferry, R., Proni, J. (2002) Development of guidelines for dredged material disposal based on abiotic determinants of coral reef community structure. *Proc. Third Specialty Conference on dredging and dredged material disposal*, Coasts, Oceans, Ports and Rivers Institute (COPRI) of ASCE, May 5, 2002, Orlando, FL USA.

MetOcean Solutions (2016). Lyttelton Harbour Dredging Project: Simulations of suspended sediment plumes and associated deposition from offshore disposal. Report P0201-02, March 2016.

Pavanelli, D. and Bigi, A. (2005). Indirect Methods to Estimate Suspended Sediment Concentration: Reliability and Relationship of Turbidity and Settleable Solids. *Biosystems Engineering*, 90:75–83.

Peterman, RM. (1990). Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Science* 47, 2–15.

SKM (2013). Improved dredge material management for the Great Barrier Reef Region, Great Barrier Reef Marine Park Authority, Townsville.

Tank, AMGK., Zwiers, FW., Zhang, W. (2009). Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation. World Meteorological Organization, Geneva, Switzerland.

Warne, M.J., Batley, G.E., Braga, O., Chapman, J.C., Fox, D.R., Hickey, C.W., Stauber, J.L., Van Dam, R. (2014). Revisions to the derivation of the Australian and New Zealand guidelines for toxicants in fresh and marine waters, *Environmental Science and Pollution Research*, 21(1), 51-60.

Yang, W. and Zurbenko, I. (2010). Kolmogorov-Zurbenko Filters. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2:340-351.