Chapter 7

An Evaluation of the Sediment Quality Assessment Guidelines for Florida Coastal Waters

7.0 Introduction

The numerical sediment quality assessment guidelines (SQAGs) presented in Chapter 6 were derived to help FDEP assess sediment quality in Florida coastal waters. When used with high quality sediment chemistry data and in conjunction with other assessment techniques, these guidelines provide a sound technical basis for assessing the potential for adverse biological effects in contaminated sediments (see Volume 2 for more information on the potential uses of the SQAGs). This chapter provides further information on the reliability, comparability, and predictive ability of these management tools. The following evaluation should be reviewed by potential users to judge the appropriateness of the SQAGs for their specific situations.

A variety of procedures have been used by state and federal agencies to evaluate the applicability of SQAGs. For example, the California State Water Resources Control Board (Lorenzato *et al.* 1991) considered it important to evaluate the precision or comparability of SQAGs. The Board defined precision as the level of agreement among guidelines based on different data and/or different approaches. Long *et al.* (In press) considered that reliability was an essential attribute of SQAGs and evaluated reliability by determining the extent to which guidelines satisfied previously established narrative objectives (i.e., does the PEL define the lower limit of a range of contaminant concentrations that is usually associated with adverse biological effects). In turn, this information may be used to assess the degree of confidence that could be placed in numerical SQAGs for each substance. Finally, SQAGs have been evaluated in terms of their ability to identify impacted and unimpacted sites, using sediment chemistry data alone (PTI 1988). The present evaluation addresses all three of these considerations and includes the following elements:

- An assessment of the comparability of the SQAGs with similar assessment tools derived using different approaches and procedures;

- An assessment of the reliability of the SQAGs using the matching sediment chemistry and biological effects data contained within the expanded NSTP database; and,

- An assessment of the predictive capability of the SQAGs using independent data sets (i.e., not included in the expanded NSTP database) from the southeast and elsewhere in the United States.

The methods used for each of these assessments and their results are discussed in the following sections. Potential users should consider the results of these evaluations before using the SQAGs in the applications outlined in Volume 2.

7.1 Comparability of the Sediment Quality Assessment Guidelines

Several governmental agencies in North America have derived sediment quality criteria, guidelines, objectives, or standards to support the assessment, management, and/or remediation of coastal sediments. These sediment management tools, developed using a wide array of approaches and procedures, often span several orders of magnitude for any given substance (MacDonald *et al.* 1992; Haines *et al.* 1994). Nonetheless, many of these guidelines share common assumptions. For example, certain sediment quality guidelines are assumed to provide a high level of protection for marine organisms, while other guidelines are intended to identify the concentrations of contaminants above which adverse effects are expected. Determination of the level of agreement among similar sediment quality guidelines provides a basis for assessing their comparability (Long *et al.* In press).

In this section, the SQAGs are compared with various sediment quality criteria, guidelines, objectives, and standards developed for other applications. The first step in this process involved collating available North American sediment quality assessment values. The expanded NSTP database was a major source of these assessment values; however, several documents containing compilations of sediment quality criteria, guidelines, objectives, and standards also were used (MacDonald *et al.* 1992; Haines *et al.* 1994). Next, the guideline values that were similar to the threshold effect levels (TELs) were identified. Likewise, guidelines that were similar to the PELs were assembled for each substance. Freshwater guidelines were not considered in this evaluation. Subsequently, the guideline values from each source were tabulated on a substance by substance basis for comparison (Tables 5 and 6).

Threshold effect levels derived from the expanded NSTP database are considered to provide a high level of protection for aquatic organisms. A total of four sets of similar guidelines were identified for comparison with the TELs, including:

- The effects range-low values (ER-Ls) promulgated under the NOAA National Status and Trends Program (Long and Morgan 1990);

Table 5. A comparison of sediment quality assessment guidelines (SQAGs) applicable to coastal and marine waters: Threshold effect level (TEL) equivalents.

Substance	TEL	ER-L	PSDDA - SL	SQC - Chronic	SQO	Number of SQAGs Comparable to TEL
Metals (SQAGs in mg/kg)						
Arsenic	7.24	8.2*	70	8.2*	20	2
Cadmium	0.68	1.2*	0.96*	7.7	1*	3
Chromium	52.3	81*	NG	NG	60*	2
Copper	18.7	34*	81	34*	100	2
Lead	30.2	46.7*	66*	33*	30*	4
Mercury	0.13	0.15*	0.21*	0.01	0.15*	3
Nickel	15.9	20.9*	NG	NG	45*	2
Silver	0.73	1*	1.2*	NG	NG	2
Zinc	124	150*	160*	190*	150*	4
Polycyclic Aromatic Hydrocarbons (PAHs; SQAGs in µg/kg)						
Acenaphthene	6.71	16*	63	2400	50	1
Acenaphthylene	5.87	44	64	NG	60	0
Anthracene	46.9	85.3*	130*	190	10	2
Fluorene	21.2	19*	64	59	50*	2
Naphthalene	34.6	160	210	500	200	0
2-Methylnapthalene	20.2	70	67	NG	NG	0
Phenanthrene	86.7	240*	320	2400	15	1
Total Low Molecular Weight PAHs	312	552*	610*	NG	NG	2
Benz(a)anthracene	74.8	261	450	1600	130*	1
Benzo(a)pyrene	88.8	430	680	18000	160*	1
Chrysene	108	384	670	1200	140*	1
Dibenz(a,h)anthracene	6.22	63.4	120	12000	60	0

Table 5. A comparison of sediment quality assessment guidelines (SQAGs) applicable to coastal and marine waters: Threshold effect level (TEL) equivalents (continued).

Substance	TEL	ER-L	PSDDA - SL	SQC - Chronic	SQO	Number of SQAGs Comparable to TEL
Polycyclic Aromatic Hydrocarbons (PAHs; SQAGs in µg/kg; cont.)						
Fluoranthene	113	600	630	1600	170*	1
Pyrene	153	665	430*	850	260*	2
Total High Molecular Weight PAHs	655	1700*	1800*	NG	NG	2
Total PAHs	1684	4022*	NG	NG	NG	1
Polychlorinated Biphenyls (PCBs; SQAGs in µg/kg)						
Total PCBs	21.6	22.7*	130	NG	30*	2
Pesticides (SQAGs in µg/kg)						
Chlordane	2.26	0.5	NG	0.3	NG	0
Dieldrin	0.72	0.02	NG	200	NG	0
p,p'-DDD	1.22	2*	NG	NG	NG	1
p,p'-DDE	2.07	2.2*	NG	NG	NG	1
p,p'-DDT	1.19	1*	NG	1.6*	NG	2
Total DDT	3.89	1.58*	NG	1.6*	NG	2
Lindane (gamma-BHC)	0.32	NG	NG	3.1	NG	0
Phthalates (SQAGs in µg/kg)						
Bis(2-ethylhexyl)phthalate	182	NG	3100	NG	NG	0

ER-L = Effects range low (Long and Morgan 1990).
PSDDA - SL = Screening level used in the Puget Sound Dredged Disposal Analysis program (USACOE 1988).
SQC - Chronic = Chronic sediment quality criterion (EqPA; Lyman et al. 1987; Pavlou et al. 1987; Hansen et al. 1993a,b,c,d,e).
SQO = Sediment quality objective (Swain and Nijman 1991).
* = Indicates that the SQAGs are within a factor of three of the PEL.

NG = No guideline.

Table 6. A comparison of sediment quality guidelines (SQAGs) applicable to coastal and marine waters: Probable effect level (PEL) equivalents.

Substance	PEL	ER-M	LAET	SQC - Acute	SLC	Number of SQAGs Comparable to PEL
Metals (SQAGs in mg/kg)						
Arsenic	41.6	70*	57*	16	NG	2
Cadmium	4.21	9.6*	5.1*	24	NG	2
Chromium	160	370*	260*	NG	NG	2
Copper	108	270*	390	54	NG	1
Lead	112	218*	450	840	NG	1
Mercury	0.7	0.71*	0.41*	0.15	NG	2
Nickel	42.8	51.6*	> 140	NG	NG	1
Silver	1.77	3.7*	> 0.56	NG	NG	1
Zinc	271	410*	410*	560*	NG	3
Polycyclic Aromatic Hydrocarbons (PAHs; SQAGs in µg/kg)						
Acenaphthene	88.9	500	500	NG	NG	0
Acenaphthylene	128	640	> 560	NG	47.4*	1
Anthracene	245	1100	960	NG	163*	1
Fluorene	144	540	540	NG	101*	1
Naphthalene	391	2100	2100	10500	414*	1
2-Methylnapthalene	201	670	670	NG	NG	0
Phenanthrene	544	1500*	1500*	14000	368	2
Total Low Molecular Weight PAHs	1442	3160*	5200	NG	NG	1
Benz(a)anthracene	693	1600*	1300*	55000	261	2
Benzo(a)pyrene	763	1600*	1600*	450000	397	2
Chrysene	846	2800	1400*	115000	384	1
Dibenz(a,h)anthracene	135	260*	230*	NG	NG	2

Table 6. A comparison of sediment quality guidelines (SQAGs) applicable to coastal and marine waters: Probable effect level (PEL) equivalents (continued).

Substance	PEL	ER-M	LAET	SQC - Acute	SLC	Number of SQAGs Comparable to PEL
Polycyclic Aromatic Hydrocarbons (PAHs; SQAGs in µg/kg; cont.)						
Fluoranthene	1494	5100	1700*	9000	644*	2
Pyrene	1398	2600*	2600*	49500	665	2
Total High Molecular Weight PAHs	6676	9600*	12000*	NG	NG	2
Total PAHs	16770	44792*	NG	NG	NG	1
Polychlorinated Biphenyls (PCBs; SQAGs in µg/kg)						
Total PCBs	189	180*	130*	NG	36.6	2
Pesticides (SQAGs in µg/kg)						
Chlordane	4.79	6*	NG	NG	NG	1
Dieldrin	4.3	8*	NG	NG	NG	1
p,p'-DDD	7.81	20*	16*	NG	NG	2
p,p'-DDE	374	27	9	NG	NG	0
p,p'-DDT	4.77	7*	34	NG	NG	1
Total DDT	51.7	46.1*	NG	210	505	1
Lindane (gamma-BHC)	0.99	NG	NG	NG	NG	0
Phthalates (SQAGs in µg/kg)						
Bis(2-ethylhexyl)phthalate	2647	NG	1900*	NG	NG	1

ER-M = Effects range median (Long and Morgan 1990).
LAET = Lowest apparent effects threshold (PTI 1988).
SQC - Acute = Acute sediment quality criterion (EqPA; Lyman et al. 1987; Pavlou et al. 1987).
SLC = National screening level concentration (Neff et al. 1987).
* = Indicates that the SQGs are within a factor of three of the TEL.
NG = No guideline available.

- The screening levels (PSDDA-SL) developed for use in Washington State under the Puget Sound Dredged Disposal Analysis Program (USACOE 1988);

- The USEPA chronic sediment quality criteria (SQC - Chronic) developed using the equilibrium partitioning approach (assuming 1% TOC; Lyman *et al.* 1987; Pavlou *et al.* 1987; Hansen *et al.* 1993a, b, c, d, and e); and,

- The sediment quality objectives (SQOs) developed for Burrard Inlet, British Columbia (Swain and Nijman 1991). These objectives were derived using a variation of the sediment background approach and are intended to protect the most sensitive use of bed sediments.

Probable effect levels define the lower limit of the range of contaminant concentrations that has a high probability of being associated with adverse biological effects. Hence, the PELs are considered to provide a low level of protection for aquatic organisms. The four sets of guidelines identified for comparison with the PELs included:

- The effects range median (ER-M) values promulgated under the NOAA National Status and Trends Program (Long and Morgan 1990);

- The lowest apparent effects threshold values (LAET) calculated for Puget Sound (PTI 1988);

- The acute sediment quality criteria (SQC - Acute) developed using the equilibrium partitioning approach (assuming 1% TOC; Lyman *et al.* 1987; Pavlou *et al.* 1987); and,

- The national screening level concentrations developed using the screening level concentration (SLC) approach (Neff *et al.* 1987).

While it is difficult to define an acceptable level of comparability for the SQAGs, a panel of experts assembled in California recommended that agreement within a factor of three or less among various guidelines was indicative of good concordance (Lorenzato *et al.* 1991). Using this guidance, the comparability of the SQAGs was evaluated by comparing them with those that have been developed using other approaches and other procedures. Comparable guidelines (i.e., agreeing within a factor of three of the SQAGs) were designated with an asterisk (*) in Tables 5 and 6.

Evaluation of the comparability of the SQAGs was impaired by the lack of guidelines for certain substances. For example, four or more guidelines were available for only 19 and 18 of the substances for which TELs and PELs, respectively, have been developed. An adequate number of guidelines were not available for chromium, nickel, silver, bis(2-ethylhexyl)phthalate, several PAHs, and most of the pesticides. Nonetheless, the results of this evaluation indicate that many of the SQAGs are comparable to the guidelines that have been derived for other applications and, hence, provide a precise basis for evaluating sediment quality conditions.

The recommended TELs for 17 of the 34 substances were comparable to two or more other guidelines, which were intended to provide a high level of protection for aquatic biota (Table 5). The best agreement was observed for metals, while the poorest agreement was observed for high molecular weight PAHs. This evaluation revealed that the TELs were usually lower than the other guidelines with which they were compared.

The recommended PELs for 14 of the 34 substances were comparable to two or more other guidelines (Table 6). Once again, the best agreement among the various guidelines was observed for metals. Relatively poor agreement was observed among the guidelines for pesticides and the low molecular weight PAHs. As was the case for the TELs, the PELs were generally lower than the other guidelines that were derived using different approaches and different procedures.

While agreement between guideline values provides users with additional confidence in these assessment tools, comparability does not guarantee that the guidelines will accurately identify impacted and unimpacted sites. Therefore, the reliability and predictability of the guidelines were also evaluated to determine if they can be used to identify impacted and unimpacted sites in the field.

7.2 Reliability of the Sediment Quality Assessment Guidelines

An evaluation of the reliability of the SQAGs was conducted to provide practitioners with an indication of the level of confidence that should be placed in these management tools. The recommended SQAGs were derived by applying a consistent arithmetic procedure to the information contained in the expanded NSTP database. This procedure was designed to support the derivation of two guideline values for each substance, which could be used together to define three ranges of contaminant concentrations: the minimal effects range; the possible effects range; and, the probable effects range. A high degree of confidence in the guidelines would be indicated if a low, moderate, and high incidence of adverse biological effects were actually observed within these three ranges of concentrations, respectively. In addition, confidence in the SQAGs would be warranted if the incidence of effects increased consistently and markedly with increasing chemical concentrations (Long *et al.* In press).

In this section, the reliability of the SQAGs was evaluated using information in the expanded NSTP database. To conduct this evaluation, a scoring system was devised to integrate the following information:

(i) The incidence of adverse biological effects within the minimal effects range;

(ii) The incidence of adverse biological effects within the probable effects range; and,

(iii) The degree of concordance between the concentrations of sediment-associated contaminants and the incidence of adverse biological effects.

The first step in applying this scoring system was to determine the TEL Score (TS). This parameter quantifies the extent to which the TEL fulfills the objective of defining the upper limit of a range of chemical concentrations within which adverse biological effects occurred only infrequently. The key metric in this evaluation was the incidence of adverse biological effects within the minimal effects range, as calculated using the expanded NSTP database. Specifically, the number of "hits" (*) and total number of records within the minimal effects range were identified on a substance by substance basis. Subsequently, the percent incidence of adverse effects was calculated for each substance by dividing the number of hits by the total number of records in the minimal effects range and multiplying by 100 (Table 7). A TS of two (2), one (1), or zero (0) was assigned if the incidence of adverse biological effects was < 10%, 10 to 25%, or > 25%, respectively, within the minimal effects range (Table 8; Long *et al.* In press).

Next, PEL Scores (PS) were determined for each substance and group of substances for which SQAGs were derived. This second parameter assesses the extent to which the PEL fulfills the objective of defining the lower limit of a range of chemical concentrations within which adverse biological effects usually or always occurred. Consistent with the procedures used to determine the TS, the percent incidence of adverse biological effects within the probable effects range was calculated on a substance by substance basis (i.e., by dividing the number of hits by the total number of records in the probable effects range and multiplying by 100). A PS of two (2), one (1), or zero (0) was assigned if the incidence of adverse biological effects was > 65%, 50 to 65%, or < 50%, respectively, within the probable effects range.

A Concordance Score (CS) was then determined to assess the agreement between contaminant concentrations and the incidence of adverse biological effects. The first step in this process was to calculate the incidence of adverse biological effects within the possible effects range (i.e., between the TEL and PEL). Next, the percent incidence of adverse biological effects within each of the three ranges of contaminant concentrations were compared. As there should be a consistent and marked increase in the incidence of hits

Table 7. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments.

Substance	% 'Hits' in the Minima Effects Range (< =TEL)	% 'Hits' in the Possible Effects Range (>TEL to <PEL)	% 'Hits' in the Probable Effects Range (>= PEL)
Metals			
Arsenic	2.7	12.9	46.8
Cadmium	5.6	20.1	70.8
Chromium	3.5	15.4	52.9
Copper	9.0	21.9	55.9
Lead	5.8	25.8	58.4
Mercury	7.8	23.6	36.7
Nickel	3.3	8.4	9.4
Silver	6.6	9.8	60.5
Zinc	3.8	27.2	64.8
Polycyclic Aromatic Hydrocarbons (PAHs)			
Acenaphthene	7.5	29.1	57.4
Acenaphthylene	7.4	13.9	51.4
Anthracene	8.7	20.5	75.0
Fluorene	11.7	20.5	70.0
2-methylnaphthalene	0.0	23.4	81.5
Naphthalene	2.6	19.3	71.2
Phenanthrene	8.0	22.8	77.8
Sum LMW-PAHs	8.7	19.4	65.6
Benz(a)anthracene	8.7	15.7	78.4
Benzo(a)pyrene	8.5	22.1	70.9
Chrysene	9.2	18.8	72.4
Dibenzo(a,h)anthracene	15.8	11.6	65.1
Fluoranthene	9.5	20.2	79.7
Pyrene	7.4	19.3	83.0
Sum HMW-PAHs	9.5	15.0	65.5
Total PAHs	7.3	19.3	76.7
Polychlorinated Biphenyls (PCBs)			
Total PCBs	15.7	36.9	54.9

Table 7. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments (continued).

Substance	% 'Hits' in the Minima Effects Range (< =TEL)	% 'Hits' in the Possible Effects Range (>TEL to <PEL)	% 'Hits' in the Probable Effects Range (>= PEL)
Pesticides			
Chlordane	9.0	12.1	17.0
Dieldrin	3.5	13.2	50.0
Lindane	2.9	21.1	25.6
p,p'-DDD	3.6	10.9	46.2
p,p'-DDE	5.3	16.5	50.0
p,p'-DDT	7.9	4.8	58.6
Total DDT	47.6	25.6	64.0
Phthalates			
Bis(2-ethylhexyl)phthal	8.5	21.2	66.7

% 'Hits' = Number of effects entries/total number of data entries within each range.
TEL = Threshold effect level.
PEL = Probable effect level.
H = High; M = Moderate; L = Low.
Confidence in the TEL was considered to be H, M, and L when % 'hits' was <10%, 10-25%, and >25%, respectively.
Confidence in the PEL was considered to be H, M, and L when % 'hits' was >65%, 50-65%, and <50%, respectively.

Table 8. An evaluation of the reliability of the sediment quality assessment guidelines for Florida coastal waters.

Substance	TEL Score (TS)	PEL Score (PS)	Concordance Score (CS)	Total Reliability Score (TRS)	Degree of Reliability
Metals					
Arsenic	2	0	2	4	M
Cadmium	2	2	2	6	H
Chromium	2	1	2	5	M
Copper	2	1	2	5	M
Lead	2	1	2	5	M
Mercury	2	0	1	3	L
Nickel	2	0	1	3	L
Silver	2	1	1	4	M
Zinc	2	1	2	5	M
Polycyclic Aromatic Hydrocarbons (PAHs)					
Acenaphthene	2	1	1	4	M
Acenaphthylene	2	1	1	4	M
Anthracene	2	2	2	6	H
Fluorene	1	2	1	4	M
Naphthalene	2	2	2	6	H
2-Methylnapthalene	2	2	2	6	H
Phenanthrene	2	2	2	6	H
Total Low Molecular Weight PAH	2	2	2	6	H
Benz(a)anthracene	2	2	1	5	M
Benzo(a)pyrene	2	2	2	6	H
Chrysene	2	2	2	6	H
Dibenz(a,h)anthracene	1	2	1	4	M
Fluoranthene	2	2	2	6	H
Pyrene	2	2	2	6	H
Total High Molecular Weight PAH	2	2	1	5	M
Total PAHs	2	2	2	6	H

Table 8. An evaluation of the reliability of the sediment quality assessment guidelines for Florida coastal waters (continued).

Substance	TEL Score (TS)	PEL Score (PS)	Concordance Score (CS)	Total Reliability Score (TRS)	Degree of Reliability
Polychlorinated Biphenyls (PCBs)					
Total PCBs	1	1	1	3	L
Pesticides					
Chlordane	2	0	0	2	L
Dieldrin	2	1	2	5	M
p,p'-DDD	2	0	2	4	M
p,p'-DDE	2	1	2	5	M
p,p'-DDT	2	1	1	4	M
Total DDT	0	1	0	1	L
Lindane (gamma-BHC)	2	0	1	3	L
Phthalates					
Bis(2-ethylhexyl)phthalate	2	2	2	6	H

H = High (TRS = 6); M = Moderate (TRS = 4-5); L = Low (TRS <4).
TEL = Threshold effect level; PEL = Probable effect level.

within the three concentration ranges (Long *et al.* In press), the presence of at least a two fold increase in the incidence of hits between adjacent ranges of concentrations was used as an indicator of concordance between increasing chemical concentrations and increased incidence of adverse biological effects. A CS of two (2) was assigned if the percent incidence of adverse biological effects was a factor of two or more higher in the possible effects range compared to the minimal effects range ***and*** a factor of two or more higher in the probable effects range compared to the possible effects range. Lower CSs were assigned if this criterion was not met. Thus, a CS of 1 was assigned if the factor of two difference was apparent between only two of the ranges, while a CS of 0 was assigned if there was no apparent concordance.

Finally, the overall reliability of the SQAGs for each substance and group of substances was evaluated by calculating a Total Reliability Score (TRS). The TRS was determined by calculating the sum of the TEL, PEL, and Concordance Scores. The SQAGs were considered to have a high degree of reliability if they met all of the evaluation criteria identified previously (i.e., TRS = 6). A moderate degree of reliability was considered to exist when intermediate scores were obtained for one or two of the parameters, or a low score was obtained for one parameter but high scores were assigned for the other two parameters resulting in a TRS of four (4) or five (5). A lower degree of reliability was assigned to SQAGs when TRSs of less than four (4) were calculated.

The results of this evaluation indicate that, in general, the recommended SQAGs provide reliable tools for evaluating sediment quality in Florida coastal waters. The TELs for a total of 30 substances were considered to have a high degree of reliability (TS = 2), as indicated by the low incidence of effects within the minimal effects range. Nine trace metals, 14 individual PAHs or groups of PAHs, six pesticides, and bis(2-ethylhexyl)phthalate were included in this group (Table 8). Moderate reliability (TS = 1) was indicated for fluorene, dibenzo(a,h)anthracene, and total PCBs. A low degree of confidence was placed on the TEL for only one substance, total DDT (TS = 0).

In general, the reliability of the PELs was lower than that for the TELs. The PELs for 16 substances were considered to have a high degree of reliability (PS =2), as indicated by a high incidence of adverse biological effects within the probable effects ranges. Of the highly reliable PELs, 14 were for individual PAHs or groups of PAHs (Table 8). The PELs for cadmium and bis(2-ethylhexyl)phthalate were also considered to be highly reliable. A moderate degree of reliability (PS = 1) was indicated for the PELs for most of the trace metals (5 of 9); while a low degree of reliability (PS = 0) was apparent for three metals (arsenic, mercury, and nickel). The PELs for pesticides and total PCBs were considered to have either a moderate or low level of reliability.

A high degree of concordance between contaminant concentrations and the incidence of adverse biological effects was observed for the majority of the recommended SQAGs (Tables 6 and 7). The incidence of adverse effects increased consistently and markedly with increasing concentrations of all trace metals, except mercury, nickel, and silver. Two-fold increases in the incidence of effects between the minimal and possible effects ranges, ***and*** the possible and probable effects ranges were also observed for ten of the 16 individual PAHs and groups of PAHs. The concordance scores for dieldrin, p,p-DDD, and p,p-DDE were also high (CS = 2), while those for four other pesticides and total PCBs were lower (CS = 0 or 1).

As indicated previously, the overall reliability of the SQAGs for each substance and group of substances was determined using all three measures of reliability. The results of this evaluation indicate that the majority of the SQAGs are highly or moderately reliable (TRS ≥ 4) and can be used with confidence to conduct sediment quality assessments (Table 8). A high degree of reliability (TRS = 6) is indicated for the guidelines for one trace metal (cadmium), ten individual PAHs or groups of PAHs, and bis(2-ethylhexyl)phthalate. The SQAGs for 16 other substances were considered to be moderately reliable (TRS = 4 or 5), including those for six trace metals, five individual PAHs, total HMW-PAHs, dieldrin, p,p-DDE, p,p-DDD, and p,p-DDT. The guidelines for mercury, nickel, total PCBs, chlordane, lindane, and total DDT were all considered to have a low level of reliability (TRS ≤ 3). It is important to consider the results of this evaluation when applying the SQAGs.

7.3 Predictability of the Sediment Quality Assessment Guidelines

The SQAGs are intended to provide effective tools for screening sediment quality data to identify priorities for further actions. For this objective to be realized, the SQAGs must support accurate predictions regarding the biological effects of sediment-associated contaminants in the field. Therefore, an initial evaluation is provided in this section to determine their ability to correctly identify biologically impacted and non-impacted sites. As the applicability of the SQAGs to Florida coastal waters and other areas in the southeastern portion of the United States is the focus of this evaluation, the predictability of the guidelines was assessed using the results of field studies from Florida (Tampa Bay and Pensacola Bay) and nearby areas (Gulf of Mexico). However, data from field studies conducted in California (San Pedro Bay) and in New York (Hudson-Raritan Estuary) were also used to evaluate the predictability of the SQAGs in more highly contaminated areas. It should be noted that these independent data sets were not used to derive the SQAGs (i.e., they are independent data sets).

The predictability of the TELs and PELs was determined separately for each of the data sets considered in this evaluation. The first step in the evaluation process was to assemble the sediment chemistry data for each geographic area. Next, the concentrations of each chemical in each sediment sample were compared to the SQAGs. These comparisons formed the basis of the predictions that were made regarding the toxicity of each sediment sample. Sediment samples with concentrations of one or more substances that exceeded their respective PELs were predicted to be toxic. Sediment samples were predicted to be non-toxic if the concentrations of all measured substances were below the respective TELs that had been recommended for those substances. Samples with concentrations of one or more substances greater than the TELs and no exceedances of the PELs were considered to be possibly toxic. These samples were not considered in the evaluation of predictability; nonetheless, the incidence of toxicity was calculated for these samples to provide an indication of the potential for observing toxic effects when contaminant concentrations fall within the possible effects range.

The predictability of the SQAGs was evaluated by comparing the predictions with the results of the biological investigations. The predictability of the PELs and the TELs were calculated as the ratio of the number of stations that were observed to be impacted and the total number of stations that were predicted to be impacted (expressed as a percentage). In this assessment, toxic samples were defined as those in which one or more of the measured bioassay endpoints were significantly different from those in control/reference samples.

During 1991 and 1992, a total of 144 sediment samples were collected to assess the toxicity of Tampa Bay sediments (Long *et al*. 1994). A suite of bioassays was used in this assessment, including a 10-d amphipod (*Ampelisca abdita*) survival test, a 1-hr sea urchin (*Arbacia punctulata*) fertilization test (using porewater), and a Microtox test (using organic extracts). Matching sediment chemistry data, including metals, PAHs, PCBs, and a suite of pesticides, were collected on 61 of these samples. Based on a comparison of the sediment chemistry data with the PELs, a total of 46 samples were predicted to be toxic (Table 9). Of these, 40 samples were observed to be toxic. Therefore, the predictability of the PELs was roughly 87%. Only two sites were predicted to be not toxic, based on comparisons of the sediment chemistry data with the TELs. The results of the biological tests conducted on these samples revealed that neither sample was toxic. Hence, the predictability of the TELs was calculated to be 100% in Tampa Bay.

In 1993, a total of 40 samples were collected to assess sediment quality in Pensacola Bay (Long Unpublished data). Sediment chemistry data were obtained on 20 of these samples, with the analytes including metals, PAHs, and a suite of pesticides. In this study, toxicity was assessed using a 10-d amphipod (*Ampelisca abdita*) survival test, a 48-hr sea urchin (*Arbacia punctulata*) fertilization and embryo development test (using porewater), and Microtox (using organic extracts). Based on comparisons of the chemistry data with the PELs, a total of 12 samples were predicted to be toxic (Table 10). Of these, 11 of the samples were observed to be toxic. Therefore, the predictability of the PELs was 92%. Using the TELs, two samples were predicted to be not toxic; both of these were observed to be toxic. Hence, the predictability of the TELs was calculated to be 0%.

As part of the USEPA Environmental Monitoring and Assessment Program (EMAP 1991), matching sediment chemistry and biological effects data were collected from eight areas in the Gulf of Mexico region in 1991. The areas sampled in this survey included Galveston Bay (TX), Matagora Bay (TX), Mississippi River (LA), Mississippi Sound (LA), Mobile Bay

Table 9. Predicted toxicity vs. observed toxicity in Tampa Bay sediments (1991 and 1992 survey; Long et al. 1994).

Predicted Toxicity	Number of Toxic Samples Amphipod Test (%)	Number of Toxic Samples Sea Urchin Test (%)	Number of Toxic Samples Microtox Test (%)	Predictability (Using All Tests; % Correct)
Overall Toxicity	8 of 61 (13%)	50 of 61 (82%)	7 of 16 (44%)	-
Not Toxic (< TELs)	0 of 2 (0%)	0 of 2 (0%)	0 of 2 (0%)	100% (2 of 2)
Possibly Toxic (> TELs; < PELs)	1 of 13 (8%)	10 of 13 (77%)	1 of 5 (20%)	-
Toxic (1 PEL exceeded)	0 of 22 (0%)	16 of 22 (73%)	1 of 2 (50%)	73% (16 of 22)
Toxic (2-5 PELs exceeded)	2 of 3 (67%)	3 of 3 (100%)	0 of 2 (0%)	100% (3 of 3)
Toxic (6-9 PELs exceeded)	1 of 5 (20%)	5 of 5 (100%)	1 of 1 (100%)	100% (5 of 5)
Toxic (> 10 PELs excceded)	4 of 16 (25%)	16 of 16 (100%)	4 of 4 (100%)	100% (16 of 16)
Toxic (1 or more PELs exceeded)	7 of 46 (15%)	40 of 46 (87%)	6 of 9 (66%)	87% (40 of 46)
Toxic (2 or more PELs exceeded)	7 of 24 (29%)	24 of 24 (100%)	5 of 7 (71%)	100% (24 of 24)

TEL = Threshold effect level; PEL = Probable effect level.

Table 10. Predicted toxicity vs. observed toxicity in Pensacola Bay sediments (Long Unpublished data).

Predicted Toxicity	Number of Toxic Samples Amphipod Test (%)	Number of Toxic Samples Sea Urchin Test (%)	Number of Toxic Samples Microtox Test (%)	Predictability (Using All Tests; % Correct)
Overall Toxicity	0 of 20 (0%)	13 of 20 (65%)	15 of 20 (75%)	-
Not Toxic (< TELs)	0 of 2 (0%)	1 of 2 (50%)	2 of 2 (100%)	0% (0 of 2)
Possibly Toxic (> TELs; < PELs)	0 of 6 (0%)	3 of 6 (50%)	2 of 6 (33%)	-
Toxic (1 PEL exceeded)	0 of 2 (0%)	2 of 2 (100%)	2 of 2 (100%)	100% (2 of 2)
Toxic (2-5 PELs exceeded)	0 of 5 (0%)	3 of 5 (60%)	4 of 5 (80%)	80% (4 of 5)
Toxic (6-9 PELs exceeded)	0 of 2 (0%)	2 of 2 (100%)	2 of 2 (100%)	100% (2 of 2)
Toxic (> 10 PELs excceded)	0 of 3 (0%)	2 of 3 (67%)	3 of 3 (100%)	100% (3 of 3)
Toxic (1 or more PELs exceeded)	0 of 12 (0%)	9 of 12 (75%)	11 of 12 (92%)	92% (11 of 12)
Toxic (2 or more PELs exceeded)	0 of 10 (0%)	7 of 10 (70%)	9 of 10 (90%)	90% (9 of 10)

TEL = Threshold effect level; PEL = Probable effect level.

(AL), Pensacola Bay, Florida Panhandle, and West Central Florida. Sediment chemistry data were collected on metals (including organotins), PAHs, PCBs, a suite of pesticides, and several additional substances. Sediment toxicity was assessed using acute toxicity tests on two crustacean species, including the amphipod, *Ampelisca abdita*, and the mysid shrimp, *Mysidopsis bahia*. As no statistical evaluation of the toxicity test results were reported by the investigators, a 20% difference between the survival of test organisms in Gulf of Mexico sediments and control sediments was considered to be indicative of toxicity. Of the 47 samples collected in this survey, three were predicted to be toxic and 16 were predicted to be not toxic (Table 11). The results of the two bioassays indicated that none of the samples that were predicted to be toxic were observed to be toxic (predictability = 0%). In contrast, 15 of the samples that were predicted to be not toxic were observed to be not toxic (predictability = 94%).

Although additional evaluations of the predictability of the SQAGs are needed, they appear to provide predictive tools for evaluating sediment quality in the southeastern portion of the United States. Considering the field survey results from the Gulf of Mexico region, 51 of the 61 sites that were predicted to be toxic actually were toxic to one or more species (Table 12). Hence, an overall predictability of roughly 84% was calculated for the PELs. The PELs were more predictive of toxicity when two of more of the SQAGs were exceeded in bed sediments (predictability = 97%; Table 12). As might be expected, these data suggest that the likelihood of observing toxic effects increases with the number of chemicals that are present at concentrations that fall within the probable effects range. By comparison, 17 of the 20 samples that were predicted to be not toxic actually were not toxic to the organisms tested. The predictability of the TELs was, therefore, calculated to be 85%. However, nearly half of the samples collected in these surveys had concentrations of one or more contaminants that fell within the possible effects range; it was not possible to predict whether or not these samples would be toxic. It should be noted that roughly 40% of the samples with contaminant concentrations within the possible effects range were observed to be toxic, based on the results of several tests.

While the recommended SQAGs were developed to assess sediment quality in Florida coastal waters, it has become apparent that these guidelines may be used in broader coastal applications throughout the United States. Data from other regions are important because they provide a means of evaluating the predictability of the SQAGs in more contaminated areas and in other sediment types. Therefore, the predictability of the SQAGs was evaluated using matching sediment chemistry and biological effects data from two additional locations, including San Pedro Bay (CA) and the Hudson-Raritan Estuary (NY). In the San Pedro Bay survey (Long Unpublished data), sediment samples were collected from 44 sites, with toxicity evaluated using a 10-d amphipod (*Eohaustorius estuarius*) survival test and a 48-hr abalone (*Haliotis rufescens*) embryo development test (using porewater). Using the PELs, 36 of these samples were predicted to be toxic (Table 13). Thirty-four of the 36 samples were observed to be toxic in one or more assays, resulting in a predictability of 94%. Based on comparisons of the sediment chemistry data to the TELs, none of the remaining samples had

Table 11. Predicted toxicity vs. observed toxicity in Gulf of Mexico sediments (EMAP 1991).

Predicted Toxicity	Number of Toxic Samples Amphipod Test (%)	Number of Toxic Samples Mysid Test (%)	Predictability (Using Both Tests; % Correct)
Overall Toxicity	7 of 47 (15%)	0 of 47 (0%)	-
Not Toxic (< TELs)	1 of 16 (6%)	0 of 16 (0%)	94% (15 of 16)
Possibly Toxic (> TELs; < PELs)	6 of 28 (21%)	0 of 28 (0%)	-
Toxic (1 PEL exceeded)	0 of 3 (0%)	0 of 3 (0%)	0% (0 of 3)
Toxic (2-5 PELs exceeded)	-	-	-
Toxic (6-9 PELs exceeded)	-	-	-
Toxic (> 10 PELs excceded)	-	-	-
Toxic (1 or more PELs exceeded)	0 (0%)	0 (0%)	0% (0 of 3)
Toxic (2 or more PELs exceeded)	-	-	-

TEL = Threshold effect level; PEL = Probable effect level.

Table 12. Predicted toxicity vs. observed toxicity in Gulf of Mexico sediment quality surveys.

Predicted Toxicity	Number of Toxic Samples Amphipod Test (%)	Number of Toxic Samples Mysid Test (%)	Number of Toxic Samples Sea Urchin Test (%)	Number of Toxic Samples Microtox Test (%)	Predictability (Using All Tests; % Correct)
Overall Toxicity	15 of 128 (12%)	0 of 47 (0%)	63 of 81 (78%)	22 of 36 (61%)	-
Not Toxic (< TELs)	1 of 20 (5%)	0 of 16 (0%)	1 of 4 (25%)	2 of 4 (50%)	85% (17 of 20)
Possibly Toxic (> TELs; < PELs)	7 of 47 (15%)	0 of 28 (0%)	13 of 19 (68%)	3 of 11 (27%)	-
Toxic (1 PEL exceeded)	0 of 27 (0%)	0 of 3 (0%)	18 of 24 (75%)	3 of 4 (75%)	67% (18 of 27)
Toxic (2-5 PELs exceeded)	2 of 8 (25%)	-	7 of 8 (88%)	4 of 7 (57%)	88% (7 of 8)
Toxic (6-9 PELs exceeded)	1 of 7 (14%)	-	7 of 7 (100%)	3 of 3 (100%)	100% (7 of 7)
Toxic (> 10 PELs excceded)	4 of 19 (21%)	-	19 of 19 (100%)	7 of 7 (100%)	100% (19 of 19)
Toxic (1 or more PELs exceeded)	7 of 61 (11%)	0 (0%)	51 of 58 (88%)	17 of 21 (81%)	84% (51 of 61)
Toxic (2 or more PELs exceeded)	7 of 34 (21%)	-	33 of 34 (97%)	14 of 17 (82%)	97% (33 of 34)

TEL = Threshold effect level; PEL = Probable effect level.

Table 13. Predicted toxicity vs. observed toxicity in San Pedro Bay sediments (Long Unpublished data).

Predicted Toxicity	Number of Toxic Samples Amphipod Test (%)	Number of Toxic Samples Abalone Test (%)	Predictability (Using Both Tests; % Correct)
Overall Toxicity	22 of 44 (50%)	40 of 44 (91%)	-
Not Toxic (< TELs)	-	-	-
Possibly Toxic (> TELs; < PELs)	3 of 8 (38%)	7 of 8 (88%)	-
Toxic (1 PEL exceeded)	8 of 18 (44%)	16 of 18 (89%)	94% (17 of 18)
Toxic (2-5 PELs exceeded)	6 of 12 (50%)	10 of 12 (83%)	92% (11 of 12)
Toxic (6-9 PELs exceeded)	4 of 6 (67%)	6 of 6 (100%)	100% (6 of 6)
Toxic (> 10 PELs excceded)	-	-	-
Toxic (1 or more PELs exceeded)	19 of 36 (53%)	33 of 36 (92%)	94% (34 of 36)
Toxic (2 or more PELs exceeded)	10 of 18 (56%)	16 of 18 (89%)	94% (17 of 18)

TEL = Threshold effect level; PEL = Probable effect level.

concentrations of contaminants which all fell within the minimal effects range; therefore, it was not possible to evaluate the predictability of the TELs.

In the Hudson-Raritan Estuary, 38 samples were collected in 1992 to evaluate sediment quality conditions (Long Unpublished data). Four biological tests were conducted to assess the toxicity of these samples, including a 10-d amphipod (*Ampelisca abdita*) survival test, 48-hr bivalve (*Mulinea lateralis*) embryo survival and development tests (using elutriates), and a Microtox assay. Of the 34 samples that were predicted to be toxic, 26 samples exhibited toxicity in one or more of the biological tests (Table 14). Hence, the predictability of the PELs was calculated to be 76%. Two samples were predicted to be not toxic and toxicity was not observed in any of the tests conducted with these samples. Therefore, the predictability of the TELs was determined to be 100%. Of the two samples that were considered to be possibly toxic, both samples were observed to be toxic in one or more biological tests.

In addition to providing predictive tools for evaluating sediment quality conditions in Florida, the results of this preliminary evaluation indicate that the recommended SQAGs may be predictive in other areas, as well. Overall, the predictability of the PELs in San Pedro Bay and the Hudson-Raritan Estuary was greater than 85% (60 of 70 samples were correctly classified as toxic). Both of the samples that were predicted to be not toxic were correctly classified. Nine of the 10 of the samples that were classified as possibly toxic were found to be toxic using a battery of tests.

Considering information from all five areas used in this evaluation, the SQAGs appear to provide predictive tools for assessing sediment quality (Table 15). Of the 22 samples that were predicted to be not toxic, 19 were observed to be not toxic in two or more bioassays. The overall predictability of the TELs was, therefore, calculated to be 88%. It is possible that water-borne contaminants (e.g., ammonia, hydrogen sulfide, etc.), unmeasured contaminants in sediments, or other factors were responsible for the observed toxicity in the samples that were incorrectly classified. Of the 131 samples that were predicted to be toxic (i.e., the concentrations of one or more substances exceeded the PELs), 111 were observed to be toxic in one or more biological tests. Therefore, the PELs were determined to have a predictability of roughly 85%. Predictability increased to 89% when the PELs for two or more substances were exceeded in sediment samples. Approximately 74% of the samples collected had contaminant concentrations that fell within either the minimal or probable effects ranges. Therefore, it was possible to classify as toxic or not toxic the majority of the samples collected in these studies.

7.4 Summary

Table 14. Predicted toxicity vs. observed toxicity in Hudson-Raritan Estuary sediments (Long Unpublished data).

Predicted Toxicity	Number of Toxic Samples Amphipod Test (%)	Number of Toxic Samples Any of 4 Tests (%)	Predictability (Using All Tests; % Correct)
Overall Toxicity	21 of 38 (55%)	28 of 38 (74%)	-
Not Toxic (< TELs)	0 of 2 (0%)	0 of 2 (0%)	100% (2 of 2)
Possibly Toxic (> TELs; < PELs)	2 of 2 (100%)	2 of 2 (100%)	-
Toxic (1 PEL exceeded)	0 of 1 (0%)	0 of 1 (0%)	0% (0 of 1)
Toxic (2-5 PELs exceeded)	7 of 13 (54%)	10 of 13 (77%)	77% (10 of 13)
Toxic (6-9 PELs exceeded)	4 of 11 (36%)	7 of 11 (64%)	64% (7 of 11)
Toxic (> 10 PELs excceded)	8 of 9 (89%)	9 of 9 (100%)	100% (9 of 9)
Toxic (1 or more PELs exceeded)	19 of 34 (56%)	26 of 34 (76%)	76% (26 of 34)
Toxic (2 or more PELs exceeded)	19 of 33 (58%)	26 of 33 (79%)	79% (26 of 33)

TEL = Threshold effect level; PEL = Probable effect level.

Table 15. Overall predictability of the recommended sediment quality assessment guidelines (SQAGs; using observed toxicity in any bioassay).

Location	Predictability of the SQAGs (% Correct)			
	Not Toxic (< TELs)	Possibly Toxic (>TELs, < PELs)	Toxic (1 or More PELs Exceeded)	Toxic (2 or More PELs Exceeded)
Tampa Bay (FL)	100% (2 of 2)	(10 of 13)	87% (40 of 46)	100% (24 of 24)
Pensacola Bay (FL)	0% (0 of 2)	(1 of 2)	92% (11 of 12)	90% (9 of 10)
Gulf of Mexico	94% (15 of 16)	(6 of 28)	0% (0 of 3)	-
San Pedro Bay (CA)	-	(7 of 8)	94% (34 of 36)	94% (17 of 18)
Hudson-Raritan Estuary (NY)	100% (2 of 2)	(2 of 2)	76% (26 of 34)	79% (26 of 33)
Overall	**86% (19 of 22)**	**(26 of 53)**	**85% (111 of 131)**	**89% (76 of 85)**

TEL = Threshold effect level; PEL = Probable effect level.

Note: Samples with contaminant concentrations within the possible effects range (i.e., > TEL and < PEL) were not predicted to be toxic or not toxic. Therefore, the predictability of the SQAGs was not evaluated using these samples.

A preliminary evaluation of the recommended sediment quality assessment guidelines was conducted to determine their potential applicability in Florida coastal waters and other areas in the United States. The results of this evaluation indicate that, in general, the SQAGs can be used with a high or moderate degree of confidence to conduct sediment quality assessments. The SQAGs for 18 to 20 substances were comparable to the functionally-similar guidelines that have been developed using different approaches and/or different data. In addition, the SQAGs for 28 substances had a moderate or high degree of reliability, as indicated by the data contained in the expanded NSTP database. Furthermore, the SQAGs for up to 34 substances, when used collectively, were found to provide predictive tools for classifying marine and estuarine sediments in terms of their potential for being associated with adverse biological effects. Overall, the predictability of the PELs and TELs was 85% and 86%, respectively. Therefore, the results of all three evaluations indicated that the SQAGs are likely to be appropriate for use in a variety of applications in Florida and elsewhere.

While the foregoing evaluation provides important information for assessing the SQAGs, additional data are required to fully evaluate the applicability of these assessment tools. Specifically, matching sediment chemistry and biological effects data from additional areas in Florida are needed to determine the predictability of the SQAGs in various types of sediments and in sediments affected by distinct contaminant sources. For example, the predictability of the SQAGs could be different in carbonate-type sediments and/or sediments that are affected principally by industrial or agricultural contaminant sources (i.e., contaminated with substances for which no SQAGs are currently available). Additional data sets, from sediment quality surveys conducted elsewhere in the United States, are also required to evaluate the regional applicability of the SQAGs.