

# Development and evaluation of sediment quality guidelines for Florida coastal waters

DONALD D. MACDONALD<sup>\*1</sup>, R. SCOTT CARR<sup>2</sup>, FRED D. CALDER<sup>3</sup>,  
EDWARD R. LONG<sup>4</sup> and CHRISTOPHER G. INGERSOLL<sup>5</sup>

<sup>1</sup>*MacDonald Environmental Sciences, Ltd., 2376 Yellow Point Road, R.R. no. 3, Ladysmith, British Columbia V0R 2E0, Canada*

<sup>2</sup>*National Biological Survey, Texas A. & M. University – Corpus Christi, Campus Box 315, 6300 Ocean Drive, Corpus Christi, TX 78412, USA*

<sup>3</sup>*Florida Department of Environmental Protection, 3900 Commonwealth Drive, Tallahassee, FL 32399-3000, USA*

<sup>4</sup>*Coastal Monitoring and Bioeffects Assessment Division, National Oceanic and Atmospheric Administration, Bin C15700, 7600 Sand Point Way NE, Seattle, Washington 98115, USA*

<sup>5</sup>*National Biological Survey, Midwest Science Center, 4200 New Haven Road, Columbia, MO 65201, USA*

Received 12 December 1994; accepted 5 May 1995

The weight-of-evidence approach to the development of sediment quality guidelines (SQGs) was modified to support the derivation of biological effects-based SQGs for Florida coastal waters. Numerical SQGs were derived for 34 substances, including nine trace metals, 13 individual polycyclic aromatic hydrocarbons (PAHs), three groups of PAHs, total polychlorinated biphenyls (PCBs), seven pesticides and one phthalate ester. For each substance, a threshold effects level (TEL) and a probable effects level (PEL) was calculated. These two values defined three ranges of chemical concentrations, including those that were (1) rarely, (2) occasionally or (3) frequently associated with adverse effects. The SQGs were then evaluated to determine their degree of agreement with other guidelines (an indicator of comparability) and the percent incidence of adverse effects within each concentration range (an indicator of reliability). The guidelines also were used to classify (using a dichotomous system: toxic, with one or more exceedances of the PELs or non-toxic, with no exceedances of the TELs) sediment samples collected from various locations in Florida and the Gulf of Mexico. The accuracy of these predictions was then evaluated using the results of the biological tests that were performed on the same sediment samples. The resultant SQGs were demonstrated to provide practical, reliable and predictive tools for assessing sediment quality in Florida and elsewhere in the southeastern portion of the United States.

**Keywords:** sediment quality guidelines; contaminants; biological effects; marine; estuarine.

## Introduction

Sediment chemistry data indicate that Florida coastal sediments in several areas are contaminated (Long and Morgan 1990, Delfino *et al.* 1991, FDEP 1994). For example, sediments in Tampa Bay, Pensacola Bay and Biscayne Bay are contaminated with trace metals, polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs) and certain pesticides. Additionally, Choctawhatchee Bay and St Andrews Bay sediments are contaminated with metals, PAHs and pesticides and elevated levels of PCBs have been

\*To whom correspondence should be addressed.

detected in St Andrews Bay, Apalachicola Bay, Charlotte Harbor, Naples Bay, Rookery Bay and the St Johns River (Long and Morgan 1990).

While sediment chemistry data are essential for evaluating sediment quality, they do not provide a comprehensive basis for classifying or managing contaminated sediments (Long *et al.* 1995). Interpretive tools are also required to relate sediment chemistry data to the potential for adverse biological effects. Various toxicity and bioaccumulation tests can be performed to evaluate the biological significance of sediment-associated contaminants (Burton 1992). In addition, numerical sediment quality guidelines (SQGs) can be used to help interpret the significance of chemical concentrations in sediments, identify contaminants of concern and prioritize areas for regulation or restoration (Long and Morgan 1990, Di Toro *et al.* 1991, Persaud *et al.* 1992, MacDonald *et al.* 1992).

A variety of approaches have been developed to support the derivation of numerical SQGs in the United States and Canada (see Chapman 1989, Adams *et al.* 1992, MacDonald *et al.* 1992, USEPA 1992, MacDonald, 1994 for reviews). The procedures used for deriving the SQGs and the philosophy behind their development are approach specific. Consequently, each approach has certain advantages and limitations that affect the applicability of the resultant guidelines. Presently there is little agreement as to which approach provides the most reliable guidelines, therefore, each of the major approaches was evaluated to determine which would best address Florida's unique requirements for SQGs (MacDonald 1994). The results of this evaluation indicated that the weight-of-evidence approach, developed to provide informal guidelines for evaluating data collected under the National Status and Trends Program of the National Oceanic and Atmospheric Administration (NOAA; Long and Morgan 1990, Long 1992), would provide scientifically defensible and broadly applicable guidelines for assessing sediment quality. The evaluation by MacDonald (1994) also indicated that some modifications to the original approach could potentially increase the applicability of the procedure for deriving SQGs for Florida's coastal waters.

The objectives of this study were to (1) describe the modifications made to the weight-of-evidence approach to support the development of SQGs for Florida coastal waters, (2) present the SQGs that were derived and (3) evaluate the applicability of these SQGs for Florida coastal waters. The comparability of the SQGs was assessed by comparing them with those derived using other data or other methods. The reliability of the guidelines was evaluated by determining the percent incidence of adverse effects within each of three ranges of contaminant concentrations. The predictability of the SQGs was assessed using several independent data sets from Florida and the Gulf of Mexico, which contain both sediment chemistry and biological effects data.

## Methods

The weight-of-evidence approach to the development of numerical sediment quality guidelines has been described in detail elsewhere (Long and Morgan 1990, Long 1992, Long and MacDonald 1992, Long *et al.* 1995), so only an overview of the approach and the modifications that were adopted in Florida will be presented here. The derivation of numerical guidelines using the weight-of-evidence approach consists of three main steps. First, all of the available information which described associations between contaminant concentrations and adverse biological effects in sediments was collected and evaluated. These data were collated in a biological effects database for sediments (BEDS;

MacDonald *et al.* 1994). Next, two SQGs, a threshold effects level (TEL) and a probable effects level (PEL), were derived for each chemical substance. These guidelines were intended to define three ranges of contaminant concentrations that were rarely, occasionally and frequently associated with adverse biological effects. Finally, the SQGs were evaluated to determine their applicability to Florida coastal waters.

#### *Development of a biological effects database for sediments*

The first step in the derivation of SQGs involved the development of a biological effects database for sediments (BEDS) to compile matching chemical and biological data from numerous studies conducted throughout North America. Over 350 reports were reviewed and critically evaluated for this purpose. These reports provided information from equilibrium-partitioning models, laboratory spiked-sediment toxicity tests and field studies used to investigate the toxicity and/or benthic community composition of marine and estuarine sediments. Each record in the database included the citation, type of test and/or biological effects observed or predicted approach that was used, study area, test duration (if applicable and reported), species tested or the benthic community considered, TOC concentrations (if reported) and the concentration(s) of each chemical (expressed on a dry weight basis). More detailed descriptions of the BEDS are provided in Long (1992), Long and MacDonald (1992), CCME (1994), MacDonald (1994), Long *et al.* (1995).

#### *Derivation of sediment quality guidelines*

The BEDS was used as the sole source of information about the potential effects of sediment-associated contaminants. Data for each of the chemicals of concern in Florida coastal waters were retrieved from the database, incorporated into data tables and sorted in ascending order of the chemical's concentration. Each ascending data table summarized all available information that related the concentrations of the chemical to effects on aquatic organisms (MacDonald *et al.* 1994). The information in these tables was then sorted into two data sets, including (1) an 'effects data set', which included all of the 'effects' data entries (i.e. those for which an adverse biological effect was observed in association with at least a 2-fold elevation in the chemical concentration above reference conditions; Long *et al.* 1995) and (2) a 'no effects data set', which included all of the 'no effects' data entries (i.e. those for which either no adverse biological effect was observed or an effect was observed but the chemical concentration was not elevated relative to reference conditions; less than a 2-fold elevation). Both a TEL and a PEL were derived for each substance that had at least 20 data entries in both the 'effects' and 'no effects' category.

In this study, the original derivation procedures were modified to develop a TEL and a PEL for each analyte. Originally, the 10th (ER-L) and 50th (ER-M) percentile values in the effects data set were used to establish sediment quality guidelines (Long and Morgan 1990, Long 1992, Long and MacDonald 1992, Long *et al.* 1995). This method was similar to the procedure used by Klapow and Lewis (1979) to establish marine water quality standards in California. These authors reasoned that the use of percentiles of aquatic toxicity data effectively minimized the influence of single (potentially outlier) data points on the resultant assessment values (e.g. Barrick *et al.* 1988). The original procedure did not utilize the information in the no effects data set, however. Nonetheless, data on the concentrations of contaminants that are not associated with

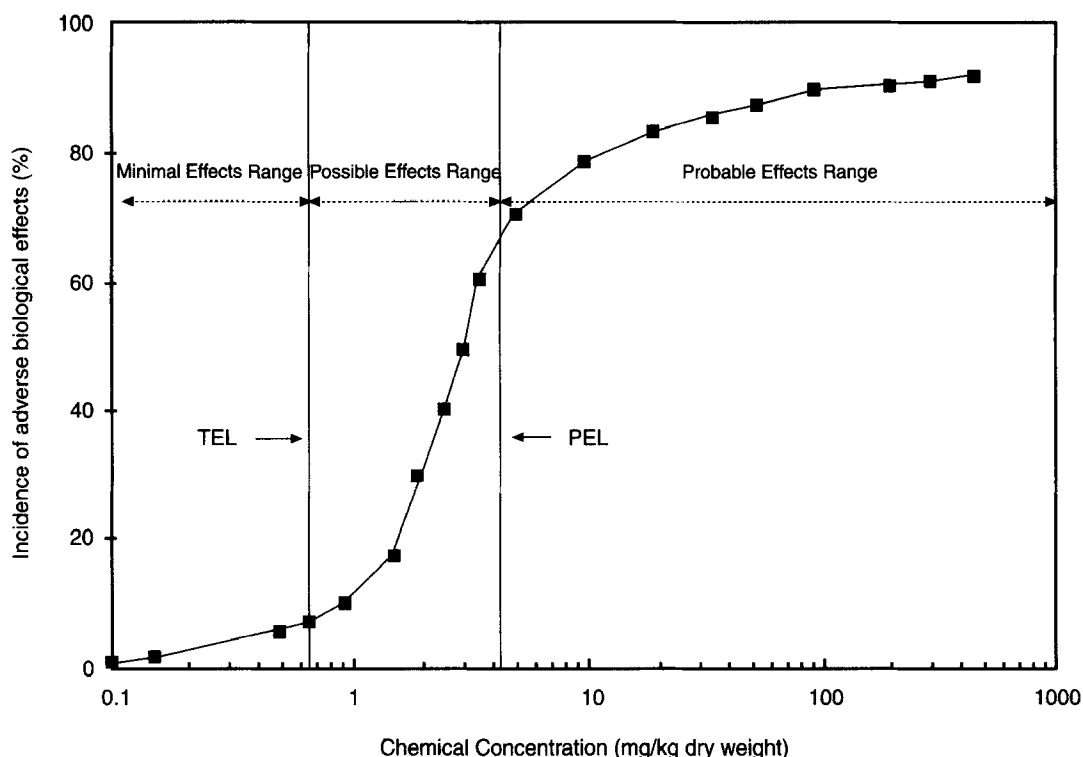
adverse effects may provide additional information for defining the relationships between contaminant exposure and biological effects and was, therefore, used in this investigation.

In the present study, two SQGs were derived for each analyte using the information in both the effects and the no effects data sets, with the distributions of these data sets determined using percentiles (Byrkit 1975). For each analyte, a TEL was derived by calculating the geometric mean of the 15th percentile of the effects data set and the 50th percentile of the no effects data set. Similarly, a PEL was developed for each chemical by determining the geometric mean of the 50th percentile of the effects data set and the 85th percentile of the no effects data set. These arithmetic procedures were established by testing a variety of options using data for cadmium, copper, fluoranthene and phenanthrene and, subsequently, evaluating the resultant guidelines relative to their narrative objective. In this respect, the TEL was intended to estimate the concentration of a chemical below which adverse effects only rarely occurred (i.e. a minimal effects range; Fig. 1). Similarly, the PEL was intended to provide an estimate of the concentration above which adverse effects frequently occurred (i.e. probable effects range). Therefore, the TEL and PEL were intended to define three concentration ranges for a chemical, including those that were rarely, occasionally and frequently associated with adverse effects (Fig. 1). The extent to which the tested options satisfied these narrative objectives was determined by calculating the percent incidence of adverse effects below the TEL and above the PEL (Long *et al.* 1995).

The arithmetic procedures used to derive the SQGs in Florida were similar to those that have been used in other applications. For example, the Canadian Council of Ministers of the Environment (CCME, 1992) calculated the geometric mean of the lowest-observed-effect concentration (LOEC) and the no-observed-effect concentration (NOEC) to establish effects-based soil quality criteria. Similarly, Rand and Petrocelli (1985) calculated maximum acceptable toxicant concentrations (MATCs) from LOECs and NOECs from aquatic toxicity tests. The Canadian Council of Resource and Environment Ministers (CCREM, 1987) established a protocol that utilizes an analogous method for deriving water quality guidelines for the protection of agricultural water uses. In each of these cases, the geometric mean, rather than the arithmetic mean, was calculated due to uncertainty regarding the distributions of the data sets (i.e., the data were not expected to be normally distributed; Sokal and Rohlf 1981). The arithmetic procedures used in this study recently were adopted for deriving national SQGs in Canada (CCME 1994).

#### *Evaluation of sediment quality guidelines*

The SQGs developed in this study were evaluated using three procedures. The comparability of the SQGs was evaluated by comparing the TELs and PELs with similar assessment tools that have been derived using different approaches or procedures. The reliability of the SQGs was evaluated by calculating the percent incidence of adverse effects within ranges of contaminant concentrations defined by the TELs and PELs, using the information contained in the BEDS (MacDonald 1994, Long *et al.* 1995). Finally, the predictability of the SQGs was assessed using independent sediment chemistry and biological effects data sets from areas throughout the southeastern portion of the United States (MacDonald 1994).



**Fig. 1.** Conceptual representation of the ranges of contaminant concentrations defined by SQGs and the potential for observing adverse effects within these ranges.

**Comparability.** The SQGs were compared to a variety of sediment quality criteria, guidelines, objectives and standards that were developed for use in other applications (Tables 1 and 2). Guidelines developed using different approaches or different procedures were considered to be comparable if they agreed within a factor of three of the SQGs developed in this study (Lorenzato *et al.* 1991).

We interpret the threshold effects levels (TELs) as the concentrations below which adverse biological effects rarely occur. Hence, the TELs 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 derived in this study, including (1) the NOAA effects range-low values (ER-Ls; Long *et al.* 1995), (2) the screening levels (PSDDA-SL) developed for the Puget Sound Dredged Disposal Analysis Program (USACOE 1988), (3) the USEPA chronic sediment quality criteria (SQC-chronic; Lyman *et al.* 1987, Pavlou 1987, Hansen *et al.* 1993a,b,c,d,e) and (4) the sediment quality objectives (SQOs) developed for Burrard Inlet (Swain and Nijman 1991).

We interpret the probable effects level (PELs) as the concentrations above which adverse biological effects frequently occur. Hence, the PELs are considered to provide a lower level of protection for aquatic organisms. The four sets of guidelines that were compared to the PELs included (1) the NOAA effects range-median values (ER-Ms; Long *et al.* 1995), (2) the lowest Puget Sound apparent effects threshold values (LAETs; Barrick *et al.* 1988), (3) the USEPA acute sediment quality criteria (SQC-

**Table 1.** A comparison of the TELs to other sediment quality guidelines for coastal and marine waters

Substance	TEL <sup>a</sup>	ER-L <sup>b</sup>	PSDDA-SL <sup>c</sup>	SQC-Chronic <sup>d</sup>	SQO <sup>e</sup>	Number of SQGs <sup>f</sup> comparable to TEL
Metals (SQGs in mg kg <sup>-1</sup> )						
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 <sup>g</sup>	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
Polychlorinated biphenyls (PCBs; SQGs in µg kg <sup>-1</sup> )						
Total PCBs	21.6	22.7*	130	NG	30*	2
Pesticides (SQGs in µg kg <sup>-1</sup> )						
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 (SQGs in µg kg <sup>-1</sup> )						
Bis(2-ethylhexyl)phthalate	182	NG	3100	NG	NG	0

Polycyclic aromatic hydrocarbons (PAHs; SQGs in  $\mu\text{g kg}^{-1}$ )

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-Methylnaphthalene	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	18 000	160*	1
Chrysene	108	384	670	1200	140*	1
Dibenz(a,h)anthracene	6.22	63.4	120	12 000	60	0
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

<sup>a</sup>TEL, threshold effect level (this study).<sup>b</sup>ER-L, effects range low (Long and Morgan 1990; Long *et al.* 1995).<sup>c</sup>PSDDA-SL, screening level used in the Puget Sound Dredged Disposal Analysis Program (USACOE 1988).<sup>d</sup>SQC-chronic, chronic sediment quality criterion (assuming 1% TOC; EqPA; Lyman *et al.* 1987, Pavlou 1987, Hansen *et al.* 1993a,b,c,d,e).<sup>e</sup>SQO, sediment quality objective (Swain and Nijman 1991).<sup>f</sup>SQG, sediment quality guideline.<sup>g</sup>NG, no guideline available.

\*Indicates that the SQGs are within a factor of three of the TEL.

**Table 2.** A comparison of the PELs to other sediment quality guidelines for coastal and marine waters

Substance	PEL <sup>a</sup>	ER-M <sup>b</sup>	LAET <sup>c</sup>	SQC <sup>d</sup> -acute	SLC <sup>e</sup>	Number of SQGs <sup>f</sup> comparable to PEL
Metals (SQGs in mg kg <sup>-1</sup> )						
Arsenic	41.6	70*	57*	16	NG <sup>g</sup>	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
Polychlorinated biphenyls (PCBs; SQGs in µg kg <sup>-1</sup> )						
Total PCBs	189	180*	130*	NG	36.6	2
Pesticides (SQGs in µg kg <sup>-1</sup> )						
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 (SQGs in µg kg <sup>-1</sup> )						
Bis(2-ethylhexyl)phthalate	2647	NG	1900*	NG	NG	1



Polycyclic aromatic hydrocarbons (PAHs; SQGs in  $\mu\text{g kg}^{-1}$ )

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	10 500	414*	1
2-Methylnaphthalene	201	670	670	NG	NG	0
Phenanthrene	544	1500*	1500*	14 000	368	2
Total low molecular weight PAHs	1442	3160*	5200	NG	NG	1
Benz(a)anthracene	693	1600*	1300*	55 000	261	2
Benzo(a)pyrene	763	1600*	1600*	450 000	397	2
Chrysene	846	2800	1400*	115 000	384	1
Dibenz(a,h)anthracene	135	260*	230*	NG	NG	2
Fluoranthene	1494	5100	1700*	9000	644*	2
Pyrene	1398	2600*	2600*	49 500	665	2
Total high molecular weight PAHs	6676	9600*	12 000*	NG	NG	2
Total PAHs	16 770	44 792*	NG	NG	NG	1

<sup>a</sup>PEL, probable effect level (this study).<sup>b</sup>ER-M, effects range median (Long *et al.* 1995, Long and Morgan 1990).<sup>c</sup>LAET, lowest apparent effects threshold (PTI, 1988).<sup>d</sup>SQC-acute, acute sediment quality criterion (assuming 1% TOC; EqPA; Lyman *et al.* 1987, Pavlou 1987).<sup>e</sup>SLC, national screening level concentration (Neff *et al.* 1987).<sup>f</sup>SQG, sediment quality guideline.<sup>g</sup>NG, no guideline available.

\*Indicates that the SQGs are within a factor of three of the PEL.

acute; Lyman *et al.* 1987, Pavlou 1987) and (4) the national screening level concentrations (SLCs; Neff *et al.* 1987).

**Reliability.** The reliability of the SQGs developed in this study was evaluated using the information contained in the ascending data tables. To facilitate this evaluation, a scoring system was devised to integrate information on three distinct attributes of the SQGs, including (1) the incidence of adverse biological effects within the minimal effects range, (2) the incidence of adverse biological effects within the probable effects range and (3) the degree of concordance between the concentrations of sediment-associated contaminants and the incidence of adverse biological effects. Good concordance between these two variables is indicated by marked increases in the incidence of effects over the three ranges of contaminant concentrations (Long *et al.* 1995).

First, a TEL score (TS) was determined for each analyte to quantify incidence of adverse biological effects within the minimal effects range. Specifically, the number of effects data entries and the total number of data entries that were contained within the minimal effects range were determined for each substance. Subsequently, the percent incidence of adverse effects was calculated for each substance by dividing the number of effects data entries by the total number of data entries within the minimal effects range and multiplying this value by 100. A TS of 2, 1 or 0 was assigned if the incidence of adverse biological effects within the minimal effects range was <10%, 10–25%, or >25%, respectively (MacDonald 1994).

Next, a PEL score (PS) was determined for each substance for which SQGs were derived. Consistent with the procedures that were used to determine the TS, the percent incidence of adverse biological effects within the probable effects range was calculated for each substance. A PS of 2, 1 or 0 was subsequently assigned if the incidence of adverse biological effects within the probable effects range was >65%, 50–65% or <50%, respectively (MacDonald 1994).

A concordance score (CS) was determined to assess the degree of agreement between contaminant concentrations and the incidence of adverse biological effects. The CS was calculated by first determining 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. Long *et al.* (1995) indicated that there should be a consistent and marked increase in the incidence of effects within the three concentration ranges. Therefore, the presence of at least a 2-fold increase in the incidence of effects between adjacent ranges of concentrations was used as an indicator of concordance. A CS of 2 was assigned if the percent incidence of adverse biological effects was a factor of 2 or more higher in the probable effects range than in the possible effects range and in the possible effects range compared to the minimal effects range. A CS of 1 was assigned if the factor of 2 difference in the incidence of effects was apparent between only two ranges. A CS of 0 was assigned if there was no apparent concordance between chemical concentrations and the incidence of adverse effects.

Finally, the overall reliability of the guidelines for each substance was evaluated by calculating a total reliability score (TRS; MacDonald 1994). The TRS was determined by calculating the sum of TS, PS and CS for each substance. The guidelines for a substance were considered to have a high degree of reliability if they had a TRS of 6 (i.e. the maximum score). A moderate degree of reliability was assigned when

intermediate scores were obtained for one or two of the parameters or when a low score was obtained for one parameter but high scores were assigned for the other two (i.e. with a TRS of 4 or 5). SQGs with a TRS that was  $<4$  were considered to have a low degree of reliability.

*Predictability.* The predictability of the TELs and PELs was evaluated separately using independent data from field surveys conducted at sites in Tampa Bay, Pensacola Bay, coastal Gulf of Mexico and offshore Gulf of Mexico. First, the concentration of each substance in each sediment sample was compared to the SQGs. 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 TELs. Sediment samples with concentrations of one or more contaminants above the TEL but below the PEL (i.e. within the possible effects range) were neither predicted to be toxic nor non-toxic and were not used to evaluate the predictability of the SQGs.

The accuracy of the predictions was then evaluated by comparing them with the results of the biological investigations. The predictability of the PEL was calculated as the ratio of the number of samples that were correctly predicted to be toxic and the number of samples that were originally predicted to be toxic (expressed as a percentage). Similarly, the predictability of the TEL was calculated as the ratio of the number of samples that were correctly predicted to be non-toxic and the number of samples that were originally predicted to be non-toxic (expressed as a percentage). In this assessment, toxic samples were defined as those in which one or more of the measured bioassay end-points were significantly different from control or reference samples.

## Results

Numerical SQGs were derived for a total of 34 substances, including nine trace metals, total PCBs, 13 individual PAHs, three groups of PAHs, seven pesticides and bis(2-ethylhexyl)phthalate. The TELs are listed in Table 1 and the PELs are listed in Table 2. Fewer than 20 effects data records were available for a number of additional substances (e.g. tributyltin, pentachlorophenol, dioxins, furans and a suite of pesticides) that are also contaminants of concern in Florida coastal sediments (MacDonald 1994). Therefore, SQGs could not be derived for these substances.

### *Comparability*

The evaluation of the comparability of the SQGs was impaired by the lack of guidelines for certain substances. For example, guidelines from three or more approaches were available for only 19 of the substances for which TELs have been developed and for only 18 of the substances for which PELs had 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 SQGs compare favourably to guidelines that were derived for other applications.

The TELs for 17 of the 34 substances were within a factor of three for two or more other guidelines (Table 1). The best agreement was observed for metals and the poorest

agreement was observed for high molecular weight (HMW) PAHs. The TELs were usually lower than values developed using other guidelines, indicating that the TELs could be more protective. The PELs for 14 of the 34 substances were within a factor of three for two or more of the other guidelines listed in Table 2. Once again, greatest agreement among the various guidelines was observed for metals. Relatively poor agreement was observed among the guidelines for pesticides and low molecular weight (LMW) PAHs. As was the case for the TELs, the PELs were generally lower than values developed using guidelines based on other procedures.

### *Reliability*

Using information in the ascending data tables, the reliability of the TELs for 30 substances was found to be relatively high ( $TS = 2$ ), as indicated by the low incidence of effects ( $<10\%$ ) within the minimal effects range (Tables 3 and 4). This group included the TELs for nine metals, 14 individual PAHs or groups of PAHs, six pesticides and bis(2-ethylhexyl)phthalate. Moderate reliability ( $TS = 1$ ) was indicated for the TELs for fluorene, dibenz(a,h)anthracene and total PCBs. Low reliability ( $TS = 0$ ) was indicated for only one substance (total DDT).

The reliability of the PELs generally was lower than that of the TELs. The PELs for 16 substances had a relatively high degree of reliability ( $PS = 2$ ), as indicated by a high incidence of adverse biological effects ( $>65\%$ ) within the probable effects ranges (Tables 3 and 4). Of the highly reliable PELs, 14 were for individual PAHs or groups of PAHs. 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 five of the nine metals; a low degree of reliability ( $PS = 0$ ) was indicated for arsenic, mercury and nickel. The PELs for pesticides and total PCBs had 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 SQGs (Tables 3 and 4). The incidence of adverse effects consistently and markedly increased with increasing concentrations for all trace metals except mercury, nickel and silver. Two-fold increases in the incidence of effects between the minimal effects range and possible effects range and the possible effects range and probable effects range were also observed for ten of the 16 individual PAHs and groups of PAHs (Table 4). The concordance scores for three pesticide guidelines were high ( $CS = 2$  for dieldrin, p,p'-DDD and p,p'-DDE), while those for four other pesticides and total PCBs were lower ( $CS = 0$  or 1).

Overall, total reliability scores of 4 or more were calculated for the majority of the guidelines, indicating high or moderate reliability (Table 4). A high degree of reliability ( $TRS = 6$ ) was indicated for one trace metal (cadmium), ten individual PAHs or groups of PAHs and bis(2-ethylhexyl)phthalate. The SQGs for 16 other substances were 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 reliability of the guidelines for mercury, nickel, total PCBs, chlordane, lindane and total DDT was lower ( $TRS < 4$ ).

### *Predictability*

The predictability of the SQGs was evaluated using four independent data sets from the southeastern portion of the United States. In Tampa Bay, Florida, matching sediment

chemistry and biological effects data were available for 61 sediment samples (Long *et al.* 1994). The analytes in these samples included metals, PAHs, PCBs and a suite of pesticides. Based on a comparison of the sediment chemistry data with the PELs, 46 of these samples were predicted to be toxic (Table 5). A suite of toxicity tests was also conducted on these samples, including a 10 day amphipod (*Ampelisca abdita*) survival test (using bulk sediments) and a 1 hr sea urchin (*Arabacia punctulata*) fertilization test (using undiluted pore water). Based on the results of these toxicity tests, 40 samples were observed to be toxic (i.e. toxicity was observed in one or more tests). Therefore, the predictability of the PELs was 87% (40 out of 46 samples). Only two of the 61 sites were predicted to be non-toxic; the results of the biological tests conducted on these two samples revealed that neither sample was toxic. Thirteen samples were not classified, as the concentrations of all analytes were below the PEL but the concentrations of one or more substances exceeded the TEL (i.e. within the possible effects range).

In Pensacola Bay, data for 20 samples were available for assessing the predictability of the guidelines (E.R. Long, unpublished data). The concentrations of metals, PAHs and a suite of pesticides were measured in each of these samples. Sediment toxicity was assessed using a 10 day amphipod (*A. abdita*) survival test, a sea urchin (*A. punctulata*) fertilization (1 h) and embryo development test (48 h) and a microbial bioluminescence test (Microtox). Based on comparisons of the metals and organics data with the PELs, 12 samples were predicted to be toxic (Table 6). 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 non-toxic; both of these samples were observed to be toxic. It should be noted that these samples contained elevated levels of one or more substances that could not be identified using a range of analytical techniques and which could have caused or contributed to the toxicity.

As part of the Environmental Monitoring and Assessment Program (EMAP), administered by the US Environmental Protection Agency, matching sediment chemistry and biological effects data were collected from eight areas in the Gulf of Mexico region in 1991 (USEPA, unpublished data). The areas sampled in this survey included Galveston Bay (TX), Matagorda Bay (TX), Mississippi River (LA), Mississippi Sound (LA), Mobile Bay AL), Pensacola Bay (FL), Florida Panhandle (FL) and West Central Florida (FL). Sediment chemistry data were collected on metals, PAHs, PCBs, a suite of pesticides and several additional substances. Sediment toxicity was assessed using acute toxicity (lethality) tests with the amphipod, *A. abdita* (10 days) and the mysid, *Mysidopsis bahia* (4 days). As no statistical evaluation of the toxicity test results were reported, a 20% difference between the survival of test organisms in Gulf of Mexico sediments versus control sediments was assumed to indicate toxicity in this evaluation (USEPA/USACOE 1991, Schimmel *et al.* 1994). Of the 47 samples collected in this survey, three were predicted to be toxic and 16 were predicted to be non-toxic (Table 7). The results of the two toxicity tests indicated that none of the samples that were predicted to be toxic were observed to be toxic (predictability = 0%). In contrast, 15 of the 16 samples that were predicted to be non-toxic were in fact non-toxic (predictability = 94%).

As part of the Minerals Management Services (MMS) Gulf of Mexico Offshore Operators Monitoring Experiment (GOOMEX), sediment chemistry and toxicity data were collected in the vicinity of petroleum exploration and production platforms in the Gulf of Mexico (R.S. Carr, D.C. Chapman, B.J. Prestley, J.M. Biedenbach, L. Robertson

**Table 3.** Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments

Substance	% effects in the minimal effects range ( $\leq$ TEL <sup>a</sup> )	% effects of the possible effects range ( $>$ TEL to $<$ PEL <sup>a</sup> )	% effects in the probable effects range ( $\geq$ PEL)
<b>Metals</b>			
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
<b>Polychlorinated biphenyls (PCBs)</b>			
Total PCBs	15.7	36.9	54.9
<b>Polycyclic aromatic hydrocarbons (PAHs)</b>			
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
Total low molecular weight 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
Dibenz(a,h)anthracene	15.8	11.6	65.1
Fluoranthene	9.5	20.2	79.7
Pyrene	7.4	19.3	83.0
Total high molecular weight PAHs	9.5	15.0	65.5
Total PAHs	7.3	19.3	76.7
Pesticides			
Chlordane	9.0	12.1	17.0
Dieldrin	3.5	13.2	50.0
Lindane (gamma-BHC)	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)phthalate	8.5	21.2	66.7

In the ascending data tables for each chemical, entries were identified as effects data entries if an adverse biological effect was reported and concordance was apparent between the observed biological response and the measured chemical concentration.

% effects = (number of effects data entries in the range/total number of data entries in the range) × 100.

<sup>a</sup>TEL, threshold effect level; PEL, probable effect level.

**Table 4.** An evaluation of the overall reliability of the sediment quality assessment guidelines for priority substances in Florida coastal waters

Substance	TEL <sup>a</sup> score (TS)	PEL <sup>a</sup> score (PS)	Concordance score (CS)	Total reliability score (TRS)	Overall <sup>b</sup> reliability
<b>Metals</b>					
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
<b>Polychlorinated biphenyls (PCBs)</b>					
Total PCBs	1	1	1	3	L
<b>Pesticides</b>					
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
<b>Phthalates</b>					
Bis(2-ethylhexyl)phthalate	2	2	2	6	H
<b>Polycyclic aromatic hydrocarbons (PAHs)</b>					
Acenaphthene	2	1	1	4	M
Acenaphthylene	2	1	1	4	M
Antracene	2	2	2	6	H
Fluorene	1	2	1	4	M
Naphthalene	2	2	2	6	H
2-Methylnaphthalene	2	2	2	6	H
Phenanthrene	2	2	2	6	H
Total low molecular weight PAHs	2	2	2	6	H
	2	2	1	5	M
Benz(a)anthracene					
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 PAHs	2	2	1	5	M
Total PAHs	2	2	2	6	H

<sup>a</sup>TEL, threshold effect level; PEL, probable effect level.<sup>b</sup>H, high (TRS = 6); M, moderate (TRS = 4–5); L, low (TRS < 4).



**Table 5.** Predictability of the TELs<sup>a</sup> and PELs<sup>a</sup> in Tampa Bay sediments (Long *et al.* 1994)

Category	Number of samples per category	Observed number of toxic samples: amphipod test (%)	Observed number of toxic samples: sea urchin test (%)	Predictability (using one to two tests; % correct)
Actual toxicity	61	8 of 61 (13)	50 of 61 (82)	–
Not toxic (<TELs)	2	0 (0)	0 (0)	100 (2 of 2)
Possibly toxic (>TELs; <PELs)	13	1 (8)	10 (77)	–
Toxic (one PEL exceeded)	22	0 (0)	16 (73)	73 (16 of 22)
Toxic (two to five PELs exceeded)	3	2 (67)	3 (100)	100 (3 of 3)
Toxic (six to nine PELs exceeded)	5	1 (20)	5 (100)	100 (5 of 5)
Toxic (>10 PELs exceeded)	16	4 (25)	16 (100)	100 (16 of 16)
Toxic (one or more PELs exceeded)	46	7 (15)	40 (87)	87 (40 of 46)
Toxic (two or more PELs exceeded)	24	7 (29)	24 (100)	100 (24 of 24)

<sup>a</sup>TEL, threshold effect level; PEL, probable effect level.**Table 6.** Predictability of the TELs<sup>a</sup> and PELs<sup>a</sup> in Pensacola Bay sediments (E.R. Long, unpublished data)

Category	Number of samples per category	Number of toxic samples: amphipod test (%)	Number of toxic samples: sea urchin test (%)	Number of toxic samples: Microtox test (%)	Predictability (using one to three tests; % correct)
Actual toxicity	–	0 of 20 (0)	13 of 20 (65)	15 of 20 (75)	–
Not toxic (<TELs)	2	0 (0)	1 (50)	2 (100)	0 (0 of 2)
Possibly toxic (>TELs; <PELs)	6	0 (0)	3 (50)	4 (67)	–
Toxic (one PEL exceeded)	2	0 (0)	2 (100)	2 (100)	100 (2 of 2)
Toxic (two to five PELs exceeded)	5	0 (0)	3 (60)	4 (80)	80 (4 of 5)
Toxic (six to nine PELs exceeded)	2	0 (0)	2 (100)	2 (100)	100 (2 of 2)
Toxic (>10 PELs exceeded)	3	0 (0)	2 (67)	3 (100)	100 (3 of 3)
Toxic (one or more PELs exceeded)	12	0 (0)	9 (75)	11 (92)	92 (11 of 12)
Toxic (two or more PELs exceeded)	10	0 (0)	7 (70)	9 (90)	90 (9 of 10)

<sup>a</sup>TEL, threshold effect level; PEL, probable effect level.

**Table 7.** Predictability of the TELs<sup>a</sup> and PELs<sup>a</sup> in coastal Gulf of Mexico sediments (USEPA, unpublished data)

Category	Number of samples per category	Number of toxic samples: amphipod test (%)	Number of toxic samples: mysid test (%)	Predictability (using one to two tests; % correct)
Actual toxicity	47	7 of 47 (15)	0 of 47 (0)	–
Not toxic (<TELs)	16	1 (6)	0 (0)	94 (15 of 16)
Possibly toxic (>TELs; <PELs)	28	6 (27)	0 (0)	–
Toxic (one PEL exceeded)	3	0 (0)	0 (0)	0 (0 of 3)
Toxic (two to five PELs exceeded)	0	–	–	–
Toxic (six to nine PELs exceeded)	0	–	–	–
Toxic (>10 PELs exceeded)	0	–	–	–
Toxic (one or more PELs exceeded)	3	0 (0)	0 (0)	0 (0 of 3)
Toxic (two or more PELs exceeded)	0	–	–	–

<sup>a</sup>TEL, threshold effect level; PEL, probable effect level.

and S.P. Boothe submitted). Five platforms were sampled during the first cruise and three platforms were sampled during the second cruise. Twenty-five stations were sampled per platform per cruise, with the stations arranged in a radial array around each platform. The sediment chemistry data consisted of a suite of trace metals and petroleum hydrocarbons; however, no chlorinated organic compounds were measured. Both sea urchin (*A. punctulata*) fertilization and embryological development tests were conducted with sediment pore water from the 125 samples from the first cruise, but only the embryological development test was used for the samples collected during the second cruise, as the latter test proved to be more sensitive for assessing the effects of the contaminants present at these platforms.

In the GOOMEX study, 16 of the 200 samples tested were toxic (Kennicutt 1993). Based on comparisons of the sediment chemistry data to the guidelines, 27 of the 200 samples were predicted to be toxic and 50 were predicted to be non-toxic (Table 8). Thirty-seven percent (ten out of 27) of the samples predicted to be toxic (i.e. one or more PELs exceeded) were observed to be toxic. The predictability increased to 58% (seven out of 12) when two or more PELs were exceeded in sediment samples. All of the non-toxic samples with contaminant concentrations that exceeded one or more PELs were predicted to be toxic based on the concentrations of zinc or lead. Ninety-six percent of the 50 samples predicted to be non-toxic were observed to be non-toxic using these very sensitive toxicity tests (Carr and Chapman 1992).

Considering all of the data collected in the Gulf of Mexico region, the results of biological investigations indicated that 59 of the 88 samples that were predicted to be toxic actually were toxic (Table 9). Hence, an overall predictability of 67% was calculated for the PELs. By comparison, 66 of the 70 samples that were predicted to be non-toxic actually were non-toxic to all of the organisms tested. The predictability of the TELs was, therefore, calculated to be 94%. When data from coastal areas of the Gulf of Mexico only were considered, the predictability of the TELs and PELs were 85 and 84%, respectively. It should be noted that slightly more than one-half of the samples collected in these surveys had concentrations of one or more contaminants that fell within the possible effects range; therefore, it was not possible to predict whether or not these samples would be toxic. On average, approximately one-fifth of these samples were observed to be toxic, based on the results of a battery of tests.

**Table 8.** Predictability of the TELs<sup>a</sup> and PELs<sup>b</sup> in offshore Gulf of Mexico sediments using the sea urchin embryological development test (Carr *et al.* 1995)

Category	Number of samples per category	Number of toxic samples (%)	Predictability (%)
Actual toxicity	200	16 of 200 (11)	—
Not toxic (<TELs)	50	2 (4)	96 (48 of 50)
Possibly toxic (>TELs; <PELs)	123	11 (9)	—
Toxic (one PEL exceeded)	15	3 (20)	20 (3 of 15)
Toxic (two to five PELs exceeded)	12	7 (58)	58 (7 of 12)
Toxic (six to nine PELs exceeded)	0	—	—
Toxic (>10 PELs exceeded)	0	—	—
Toxic (one or more PELs exceeded)	27	10 (37)	37 (10 of 27)
Toxic (two or more PELs exceeded)	12	7 (58)	58 (7 of 12)

<sup>a</sup>TEL, threshold effect level; PEL, probable effect level.

**Table 9.** Evaluation of the predictability of the TELs<sup>a</sup> and PELs<sup>a</sup> in southeastern United States sediments

Location	Number of samples correctly predicted as not toxic (%)	Number of samples correctly predicted as toxic (%)
Coastal Gulf of Mexico		
Tampa Bay	2 of 2 (100)	40 of 46 (87)
Pansacola Bay	0 of 2 (0)	11 of 12 (92)
Coastal Estuaries	15 of 16 (94)	0 of 3 (0)
Total coastal Gulf of Mexico	17 of 20 (85)	51 of 61 (84)
Offshore Gulf of Mexico	48 of 50 (96)	10 of 27 (37)
Gulf of Mexico (total coastal plus offshore)	65 of 70 (93)	61 of 88 (69)

<sup>a</sup>TEL, threshold effect level; PEL, probable effect level.

## Discussion

The modified weight-of-evidence approach developed here is characterized by a number of attributes that make it attractive for deriving SQGs for Florida coastal waters. Unlike many other approaches to the development of SQGs, the weight-of-evidence approach does not attempt to establish absolute sediment quality assessment values. Instead, the approach delineates ranges of contaminant concentrations that are probably, possibly and not likely to be associated with adverse biological effects. This approach recognizes the uncertainty associated with the prediction of biological effects under a variety of field conditions and relies upon the evidence assembled from numerous independent studies.

One of the more important attributes of the weight-of-evidence approach is its overall practicality. Guidelines for 34 potentially toxic substances were calculated relatively quickly and inexpensively using available data. In addition, by considering matching sediment chemistry and biological effects data from studies conducted in the field, the influence of mixtures of chemicals in sediments is incorporated in the resultant SQGs. This feature increases the degree of environmental realism and, thus, applicability of the guidelines. Furthermore, the information in the BEDS is highly relevant to the guidelines derivation process because it applies to a wide range of biological organisms and end points, incorporates a large number of direct measurements on organisms that are normally associated with bedded sediments and includes many data from various studies conducted in the southeastern United States (including Florida). These attributes are likely to give the SQGs derived using the modified weight-of-evidence approach broad applicability in the southeast, increasing the probability that the guidelines would be appropriate for implementation in Florida.

In addition to the other advantages of the approach, the arithmetic procedures used in this study for calculating SQGs considered the information in both the effects and no effects data sets. Hence, the resultant guidelines were more likely to satisfy narrative objectives. And, in contrast to the apparent effects threshold approach (Barrick *et al.* 1988), the weight-of-evidence procedure does not rely heavily on individual data points. Therefore, outliers do not excessively influence values in the overall guidelines derivation process.

Despite the benefits associated with this approach, a number of limitations were also evident which could restrict application of these guidelines. First, the weight-of-

evidence approach does not fully support the quantitative evaluation of cause and effect relationships between contaminant concentrations and biological responses. Although information from spiked-sediment toxicity tests and equilibrium partitioning models is included in the BEDS, the weight-of-evidence approach is still largely based on associations between contaminant concentrations and biological responses. Various factors other than concentrations of the contaminant under consideration could have influenced the actual response observed in any given investigation, including (but not limited to) the additive and synergistic effects of co-occurring contaminants. While the assembly of extensive information from numerous estuarine and marine sites across North America into a single database reduces this limitation, the level of uncertainty associated with the resultant SQGs is still not quantified. We are currently investigating the use of toxic units models to reduce uncertainty (i.e. R.C. Swartz, D.W. Schults, R.J. Ozretich, J.O. Lamberson, F.A. Cole, T.H. DeWitt, M.S. Redmond and S.P. Ferraro submitted).

Application of the recommended approach may also be restricted by other limitations on the available information. Presently, few data exist on the chronic responses of marine and estuarine organisms to contaminated sediments. Furthermore, few data are available for some potentially important contaminants of sediments in Florida, including tributyltin, pentachlorophenol, dioxins and furans and various pesticides. These data gaps impair our ability to evaluate the overall applicability of the approach to Florida.

Sediment-associated contaminants can accumulate in the tissues of aquatic organisms and thus have the potential to adversely affect human and non-human consumers of aquatic biota. However, the guidelines developed in this study do not address either the potential for bioaccumulation or the associated adverse effects of bioaccumulation on higher trophic levels.

One shortcoming of the weight-of-evidence approach is associated with the limitations on the data that describe the potential bioavailability of sediment-associated contaminants (Di Toro *et al.* 1990). Large differences in toxicity of sediment-associated contaminants have been reported for relatively small ranges in concentration for total organic carbon (TOC) and/or acid volatile sulphide (AVS) (Swartz *et al.* 1990, Di Toro *et al.* 1991, Adams *et al.* 1992). However, data on sediment grain size, levels of TOC and concentrations of AVS were not provided in most of the reports reviewed in this study. Thus, it was not possible to express the guidelines in terms of the factors that may influence the bioavailability of these contaminants. While reliance on ranges of concentrations instead of absolute values and consideration of the no effects data set reduces this limitation, sediment quality guidelines are less defensible if they do not account for the factors that control bioavailability (e.g. Di Toro *et al.* 1991). For this reason, the SQGs derived in this study were evaluated to determine their comparability, reliability and predictability. Sediment quality guidelines derived using other approaches and being considered for use in national or regional programs should also be thoroughly evaluated before being implemented.

The results of this evaluation indicate that the SQGs can be used with a high or moderate degree of confidence to assess sediment quality. The SQGs for approximately half of the substances were comparable to other guidelines that have been developed using different approaches or different procedures. Additionally, the SQGs for 28 substances had a moderate or high degree of reliability, as indicated by the data contained in the BEDS. Furthermore, the SQGs for 34 substances, when used

collectively, provided predictive tools for correctly classifying marine and estuarine sediments with respect to adverse biological effects. Overall, the predictability of the dry weight-normalized PELs and TELs was >67 and >94%, respectively, indicating that the potential limitations identified previously do not seriously compromise the applicability of the SQGs. Interestingly, the predictability of the PELs was higher (>80%) for sediment samples that had complex mixtures of contaminants than it was for sediment samples that were contaminated by one or two substances only (predictability = 37%). This increased ability to predict toxicity in sediments with complex mixtures is likely because the database, upon which the guidelines are based, primarily contains information from sites with sediments that contained many chemical substances. Therefore, the guidelines are particularly relevant for assessing sediment quality in areas with multiple contaminant inputs.

A critical consideration in sediment risk assessment is the potential for incorrectly classifying sediments. Based on the results of this assessment, it is apparent that there is an approximately 33% probability of incorrectly classifying non-toxic samples as toxic using the PELs (i.e. false positives). In contrast, there was a relatively low probability of incorrectly classifying toxic samples as non-toxic using the TELs (i.e. false negatives). Therefore, the guidelines are considered to be conservative tools for assessing contaminated sediments (i.e. they err on the side of environmental protection; see C.G. Ingersoll, P.S. Haverland, E.L. Brunson, T.J. Canfield, F.J. Dwyer, C.E. Henke, N.E. Kemble, D.R. Mount and R.C. Fox (submitted) for a more comprehensive evaluation of type I and type II errors using SQGs). Together, the evaluations of comparability, reliability and predictability indicate that the SQGs are likely to be appropriate for use in a variety of applications and sites in Florida and, perhaps, elsewhere.

The GOOMEX study, conducted in the vicinity of offshore oil and gas exploration and production platforms in the Gulf of Mexico, provided a unique environment for testing the predictability of the SQGs. The primary contaminants of concern were metals (Cd, Cu, Hg, Pb and Zn) and the sampling design provided a gradient in contaminant concentrations along transects extending from the point source discharge. This situation was unique compared with most urban coastal environments where complex mixtures of organic and inorganic contaminants are usually present due to point and non-point sources of pollutants. Compared with the two studies in coastal areas which used the sea urchin pore water tests, the GOOMEX study had a much lower concordance between predicted and observed toxicity. This may be due to the fact that most of the data used to develop the SQGs was obtained from studies conducted near urbanized coastal areas, where complex mixtures of contaminants are more prevalent. The additive and synergistic effects of these co-occurring chemicals are of necessity incorporated into the SQGs. In this unique study in the Gulf of Mexico, where the numbers and classes of chemicals present were limited compared to urbanized coastal areas, the PELs may be overprotective due to the limited additive or synergistic effects of co-occurring contaminants. Another explanation might be that the particular forms of the metals were insoluble or ligands were present which minimized their bioavailability. The TELs, however, provided accurate predictions of non-toxicity in the GOOMEX study.

The PELs accurately predicted sediment toxicity in two out of three studies used to assess the predictability of the SQGs in coastal areas of the Gulf of Mexico. No

concordance was observed between predicted and observed toxicity in the EMAP study; this disparity might be explained by differences in the types of toxicity tests that were used among studies. The EMAP study utilized whole sediment toxicity tests with adult crustaceans, whereas the GOOMEX study used sea urchin pore water toxicity tests. The sea urchin fertilization and embryo development tests with pore water provide estimates of chronic toxicity and uses, as an end-point, responses of a sensitive life stage of a sensitive species. In contrast, the whole sediment assessments used in the EMAP study involved estimates of acute toxicity to adult crustaceans. The PELs appear to be more predictive of chronic, sublethal effects than the more obvious acute, lethal effects. Likewise, the TELs appear to be protective of chronic, sublethal effects, as demonstrated by the high degree of concordance between the TEL predictions and the observed toxicity with the sea urchin tests. These results emphasize the fact that the organisms used in various toxicity tests differ with respect to their ability to estimate effects of sediment-associated contaminants. Therefore, comprehensive sediment quality assessments should employ a battery of biological tests, at least one of which should be sensitive enough to detect chronic, sublethal effects in sensitive species.

The results of this investigation indicate that the SQGs generally provide reliable and predictive tools for assessing coastal sediment quality in Florida and elsewhere in the southeast. However, the SQGs should not be used as stand-alone sediment quality criteria. The applicability of these guidelines in other coastal areas of the United States has not been fully evaluated. For this reason, the predictability of the guidelines derived using the original (Long and Morgan 1990) and modified (this study) approaches will be compared using independent data sets from throughout the United States and Canada. The results of this comparison, which will be published in a subsequent manuscript, will provide additional guidance in the use and applicability of the respective SQGs.

### **Acknowledgements**

Encouragement, suggestions, and advice were provided by Herb Windom (Skidaway Institute of Oceanography), Steve Schropp (Taylor Engineering Inc.), Gail Sloane and Tom Seal (Florida Department of Environmental Protection), Pam Haverland (National Biological Survey), Sherri Smith (Environment Canada) and Jay Field (National Oceanic and Atmospheric Administration). Technical support was also provided by M.L. Haines, B. Charlish, K. Brydges, B. Moore and M. Popadyne (MacDonald Environmental Sciences Ltd). Initial drafts of this manuscript were reviewed by Sherri Smith, Pam Haverland and Jim Dwyer. Helpful and constructive comments were also provided by two anonymous reviewers.

### **Disclaimer**

The methods and guidelines presented in this report do not necessarily represent the policy of the National Biological Survey, the Florida Department of Environmental Protection or the National Oceanic and Atmospheric Administration.

## References

- Adams, W.J., Kimerle, R.A. and Barnett, J.W., Jr (1992) Sediment quality and aquatic life assessment. *Environ. Sci. Technol.* **26**, 1863–76.
- Barrick, R., Becker, S., Pastorok, R., Brown, L. and Beller, H. (1988) *Sediment Quality Values Refinement: 1988 Update and Evaluation of Puget Sound AET*. Prepared by PTI Environmental Services for the US Environmental Protection Agency, Bellevue, WA.
- Burton, G.A., Jr (1992) *Sediment Toxicity Assessment*, Chelsea, MI: Lewis Publishers.
- Byrkit, D.R. (1975) *Elements of Statistics*, 2nd edn, New York: D. Van Nostrand Company.
- Carr, R.S. and Chapman, D.C. (1992) Comparison of solid-phase and pore-water approaches for assessing the quality of marine and estuarine sediments. *Chem. Ecol.* **7**, 19–30.
- Carr (1995)
- CCME (Canadian Council of Ministers of the Environment) (1992) *A Protocol for the Derivation of Ecological Effects-based and Human Health-based Soil Quality Criteria for Contaminated Sites*, Ottawa: CCME Subcommittee on Environmental Quality Criteria for Contaminated Sites, Environment Canada.
- CCME (Canadian Council of Ministers of the Environment) (1994) *Protocol for the Derivation of Canadian Sediment Quality Guidelines for the Protection of Aquatic Life*, Ottawa: CCME Task Group on Water Quality Guidelines.
- CCREM (Canadian Council of Resource and Environment Ministers) (1987) *Canadian Water Quality Guidelines. Appendix XV: Protocols for Deriving Water Quality Guidelines for the Protection of Agricultural Water Uses (October 1993)*, Ottawa: Task Force on Water Quality Guidelines.
- Chapman, P.M. (1989) Current approaches to developing sediment quality criteria. *Environ. Toxicol. Chem.* **8**, 589–99.
- Delfino, J.J., Coates, J.A., Davis, W.M., Garcia, K.L., Jacobs, M.W., Marincic, K.J. and Signorella, L.L. (1991) *Toxic Pollutants in Discharges, Ambient Waters, and Bottom Sediments*, Volume I and II, submitted to Florida Department of Environmental Regulation, Tallahassee, FL. University of Florida, Gainesville, FL.
- Di Toro, D.M., Mahony, J.D., Hansen, D.J., Scott, K.J., Hicks, M.B., Mayr, S.M. and Redmond, M.S. (1990) Toxicity of cadmium in sediments: the role of acid volatile sulfide. *Environ. Toxicol. Chem.* **9**, 1487–502.
- Di Toro, D.M., Zarba, C.S., Hansen, D.J., Berry, W.J., Swartz, R.C., Cowan, C.E., Pavlou, S.P., Allen, H.E., Thomas, N.A. and Paquin, P.R. (1991) Technical basis for establishing sediment quality criteria for non-ionic organic chemicals using equilibrium partitioning. *Environ. Toxicol. Chem.* **10**, 1541–83.
- FDEP (Florida Department of Environmental Protection) (1994) *Florida Coastal Sediment Contaminants Atlas*, Tallahassee, FL: Office of the Secretary. 112 pp.
- Hansen, D.J., Berry, W.J., Di Toro, D.M., Paquin, P., Davanzo, L., Stancil, F.E., Jr and Kollig, H.P. (1993a) *Proposed Sediment Quality Criteria for the Protection of Benthic Organisms: Endrin*, Washington, DC: US Environmental Protection Agency, Office of Water and Office of Research and Development, Office of Science and Technology, Health and Ecological Criteria Division.
- Hansen, D.J., Berry, W.J., Di Toro, D.M., Paquin, P., Davanzo, L., Stancil, F.E., Kollig, H.P. and Hoke, R.A. (1993b) *Proposed Sediment Quality Criteria for the Protection of Benthic Organisms: Dieldrin*, Washington, DC: US Environmental Protection Agency, Office of Water and Office of Research and Development, Office of Science and Technology, Health and Ecological Criteria Division.
- Hansen, D.J., Berry, W.J., Di Toro, D.M., Paquin, P., Davanzo, L., Stancil, F.E., Jr and Kollig, H.P. (1993c) *Proposed Sediment Quality Criteria for the Protection of Benthic Organisms: Phenanthrene*, Washington, DC: US Environmental Protection Agency, Office of Water and Office of Research and Development, Office of Science and Technology, Health and Ecological



## Criteria Division.

- Hansen, D.J., Berry, W.J., Di Toro, D.M., Paquin, P., Stancil, F.E., Jr and Kollig, H.P. (1993d) *Proposed Sediment Quality Criteria for the Protection of Benthic Organisms: Acenaphthene*, Washington, DC: US Environmental Protection Agency, Office of Water and Office of Research and Development, Office of Science and Technology, Health and Ecological Criteria Division.
- Hansen, D.J., Berry, W.J., Di Toro, D.M., Paquin, P., Davanzo, L., Stancil, F.E., Jr and Kollig, H.P. (1993e) *Proposed Sediment Quality Criteria for the Protection of Benthic Organisms: Fluoranthene*, Washington, DC: US Environmental Protection Agency, Office of Water and Office of Research and Development, Office of Science and Technology, Health and Ecological Criteria Division.
- Kennicutt, M.C., II (ed.) (1993) *Gulf of Mexico Offshore Operations Monitoring Experiment. Phase I: Sublethal Responses to Contaminant Exposure, Interim Report: Year 1*, New Orleans, LA: Gulf of Mexico OCS Regional Office, Minerals Management Service, United States Department of the Interior.
- Klapow, L.A. and Lewis, R.H. (1979) Analysis of toxicity data for California marine water quality standards. *J. Water Pollut. Control Fed.* **51**, 2051–70.
- Long, E.R. (1992) Ranges of chemical concentrations in sediments associated with adverse biological effects. *Mar. Pollut. Bull.* **24**, 38–45.
- Long, E.R. and MacDonald, D.D. (1992) National Status and Trends Program Approach. In *Sediment classification methods compendium*, pp. 14–1–18. Washington, DC: Office of Water, United States Environmental Protection Agency.
- Long, E.R. and Morgan, L.G. (1990) *The Potential for Biological Effects of Sediment-sorbed Contaminants Tested in the National Status and Trends Program*, Seattle, WA: National Oceanic and Atmospheric Administration.
- Long, E.R., Wolfe, D.A., Carr, R.S., Scott, K.J., Thursby, G.B., Windom, H.L., Lee, R., Calder, F.D., Sloane, G.M. and Seal, T. (1994) *Magnitude and Extent of Sediment Toxicity in Tampa Bay, Florida*, Seattle, WA: National Status and Trends Program, National Oceanic and Atmospheric Administration.
- Long, E.R., MacDonald, D.D., Smith, S.L. and Calder, F.D. (1995) Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ. Manage.* **19**, 81–97.
- Lorenzato, S.G., Gunther, A.J. and O'Connor, J.M. (1991) *Summary of a Workshop Concerning Sediment Quality Assessment and Development of Sediment Quality Objectives*, Sacramento, CA: California State Water Resources Control Board.
- Lyman, W.J., Glazer, A.E., Ong, J.H. and Coons, S.F. (1987) *An Overview of Sediment Quality in the United States: Final Report*. Prepared for the Environmental Protection Agency, Washington, DC. US Environmental Protection Agency, Office of Water and Office Water Regulations and Standards.
- MacDonald, D.D. (1994) *Approach to the Assessment of Sediment Quality in Florida Coastal Waters. Volume 1 – Development and Evaluation of the Sediment Quality Assessment Guidelines*. Report prepared for Florida Department of Environmental Protection, Tallahassee, FL.
- MacDonald, D.D., Smith, S.L., Wong, M.P. and Mudroch, P. (1992) *The Development of Canadian Marine Environmental Quality Guidelines*, Ottawa: Ecosystem Sciences and Evaluation Directorate, Conservation and Protection, Environment Canada.
- MacDonald, D.D., Charlish, B.L., Haines, M.L. and Brydges, K. (1994) *Approach to the Assessment of Sediment Quality in Florida Coastal Waters. Volume 3 – Supporting Documentation: Biological Effects Database for Sediments*. Report prepared for Florida Department of Environmental Protection, Tallahassee, FL.
- Neff, J.M., Word, J.Q. and Gulbransen, T.C. (1987) *Recalculation of Screening Level Concentrations for Nonpolar Organic Contaminants in Marine Sediments. Final Report*, Prepared for Environmental Protection Agency Region V, Washington, DC.

- Pavlou, S.P. (1987) The use of the equilibrium partitioning approach in determining safe levels of contaminants in marine sediments. In K.L. Dickson, A.W. Maki and W.A. Brungs (eds) *Fate and effects of sediment-bound chemicals in aquatic systems. Proceedings of the Sixth Pellston Workshop*, pp. 388–95. Toronto, Ontario: Pergamon Press.
- Persaud, D., Jaagumagi, R. and Hayton, A. (1992) *Guidelines for the Protection and Management of Aquatic Sediment Quality in Ontario*, Toronto, Ontario: Water Resources Branch, Ontario Ministry of the Environment.
- PTI (1988)
- Rand, G.M. and Petrocelli, S.R. (1985) *Fundamentals of Aquatic Toxicology*, Washington, DC: Hemisphere Publishing Corporation.
- Schimmel, S.C., Melzian, B.D., Campbell, B.E., Strobel, C.J., Benyi, S.J. Rosen, J.S. and Buffum, H.W. (1994) *Statistical Summary. EMAP – Estuaries Virginian Province – 1991*, Washington, DC: Environmental Monitoring and Assessment Program, Office of Research and Development, US Environmental Protection Agency.
- Sokal, R.R. and Rohlf, F.J. (1981) *Biometry*, San Francisco, CA: W.H. Freeman and Company.
- Swain, L.G. and Nijman, R.A. (1991) An approach to the development of sediment quality objectives for Burrard Inlet. In P. Chapman, F. Bishay, E. Power, K. Hall, L. Harding, D. McLeay, M. Nassichuk and W. Knapp (eds) *Proceedings of the Seventeenth Annual Aquatic Toxicity Workshop*, Vol. 2, pp. 1026–37, 5–7 November 1990, Vancouver, British Columbia: Canadian Technical Report of Fisheries and Aquatic Sciences 1774.
- Swartz, R.C., Schults, D.W., DeWitt, T.H., Ditsworth, G.R. and Lamberson, J.O. (1990) Toxicity of fluoranthene in sediment to marine amphipods: a test of the equilibrium partitioning approach to sediment quality criteria. *Environ. Toxicol. Chem.* **9**, 1071–80.
- USACOE (United States Army Corps of Engineers) (1988) *Evaluation Procedures Technical Appendix – Phase I (Central Puget Sound). Puget Sound Dredged Disposal Analysis Report*, Seattle, WA: Washington State Department of Natural Resources.
- USEPA (United States Environmental Protection Agency) (1992) *Sediment Classification Methods Compendium*, Washington, DC: Office of Water.
- USEPA/USACOE (United States Environmental Protection Agency/United States Army Corps of Engineers) (1991) *Ecological Evaluation of Proposed Discharge of Dredged Material in Ocean Waters*, Washington, DC: Office of Marine and Estuarine Protection.