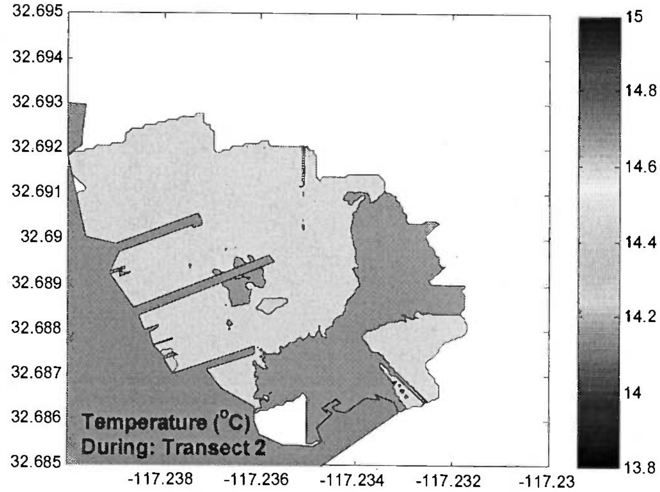
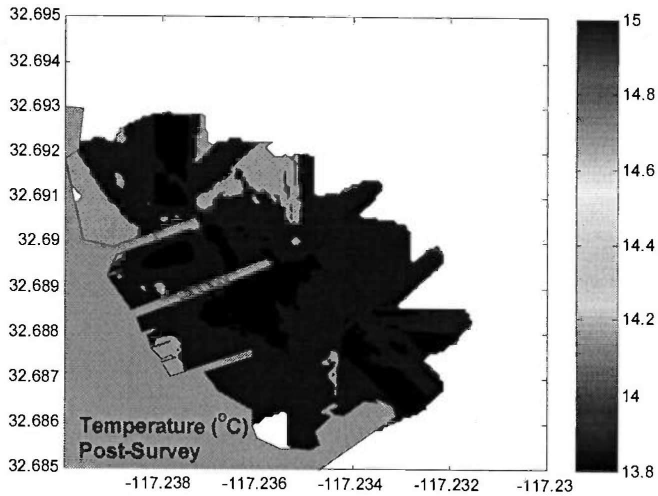
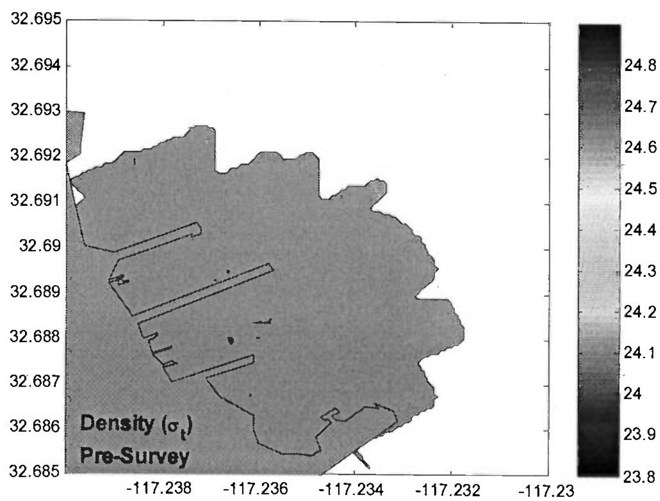


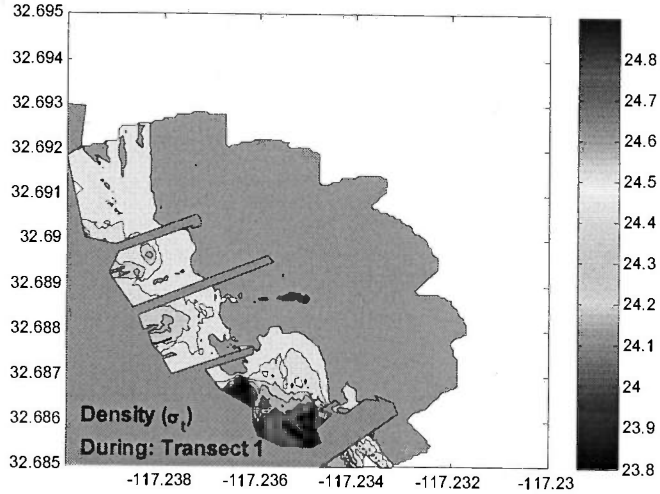
32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
15
14.8
14.6
14.4
14.2
14
13.8
Temperature (°C)
During: Transect 2



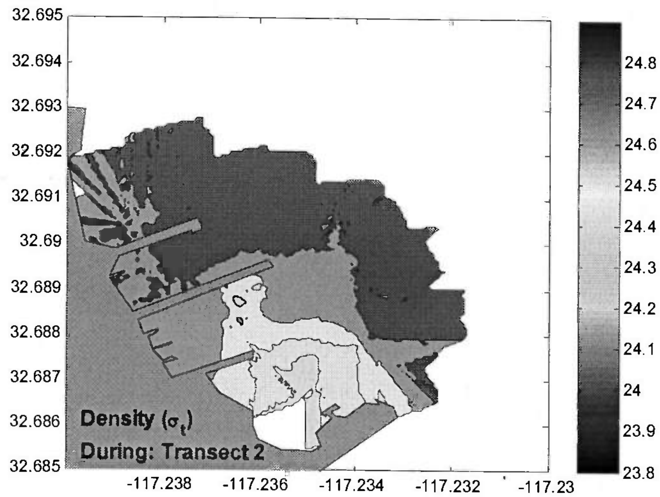
32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
15
14.8
14.6
14.4
14.2
14
13.8
Temperature (°C)
Post-Survey



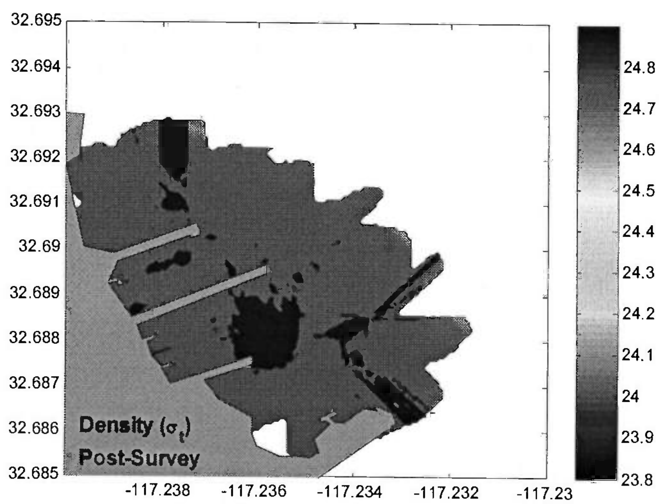
32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
24.8
24.7
24.6
24.5
24.4
24.3
24.2
24.1
24
23.9
23.8
Density (σ_t)
Pre-Survey



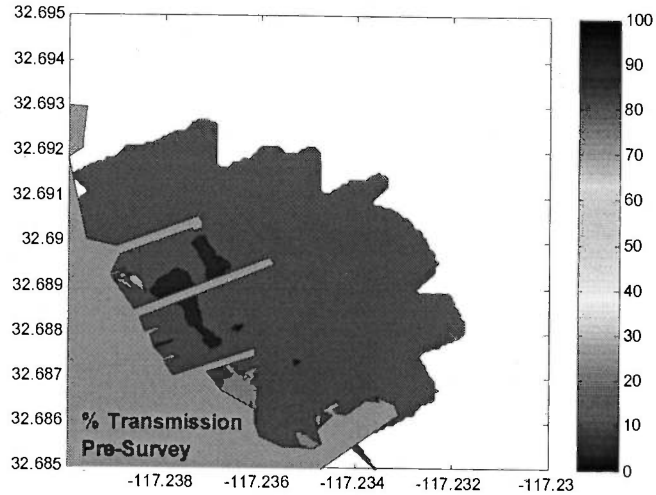
32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
24.8
24.7
24.6
24.5
24.4
24.3
24.2
24.1
24
23.9
23.8
Density (σ_t)
During: Transect 1

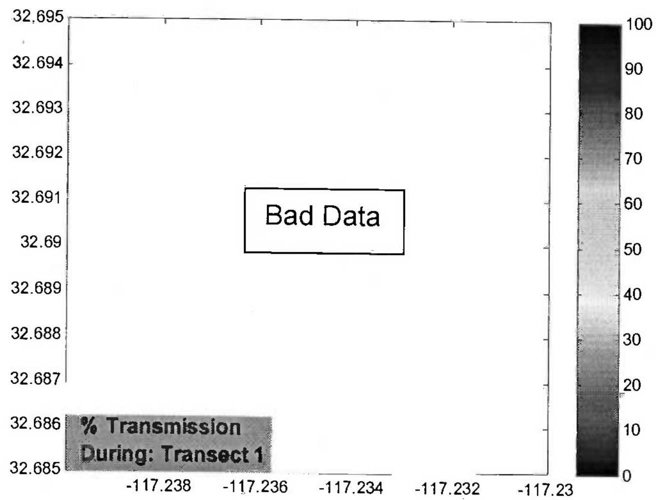


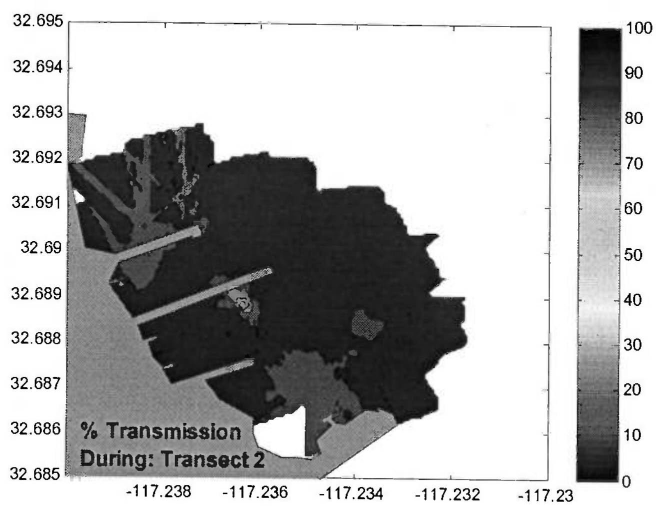
32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
24.8
24.7
24.6
24.5
24.4
24.3
24.2
24.1
24
23.9
23.8
Density (σ_t)
During: Transect 2

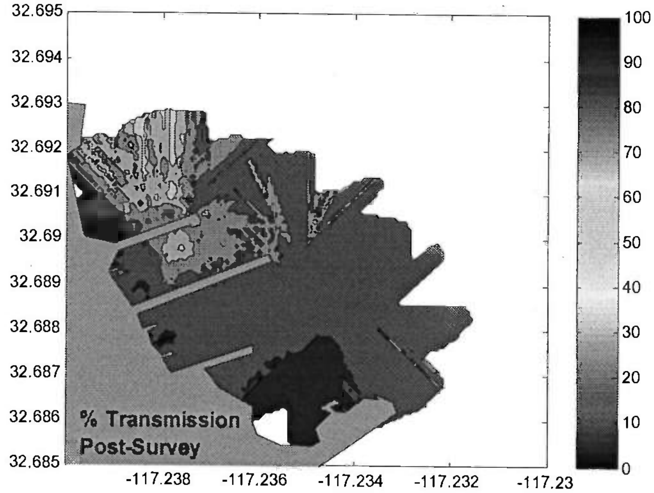


32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
24.8
24.7
24.6
24.5
24.4
24.3
24.2
24.1
24
23.9
23.8
Density (σ_t)
Post-Survey


32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
100
90
80
70
60
50
40
30
20
10
0
% Transmission
Pre-Survey


32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
100
90
80
70
60
50
40
30
20
10
0
Bad Data
% Transmission
During: Transect 1


32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
100
90
80
70
60
50
40
30
20
10
0
% Transmission
During: Transect 2

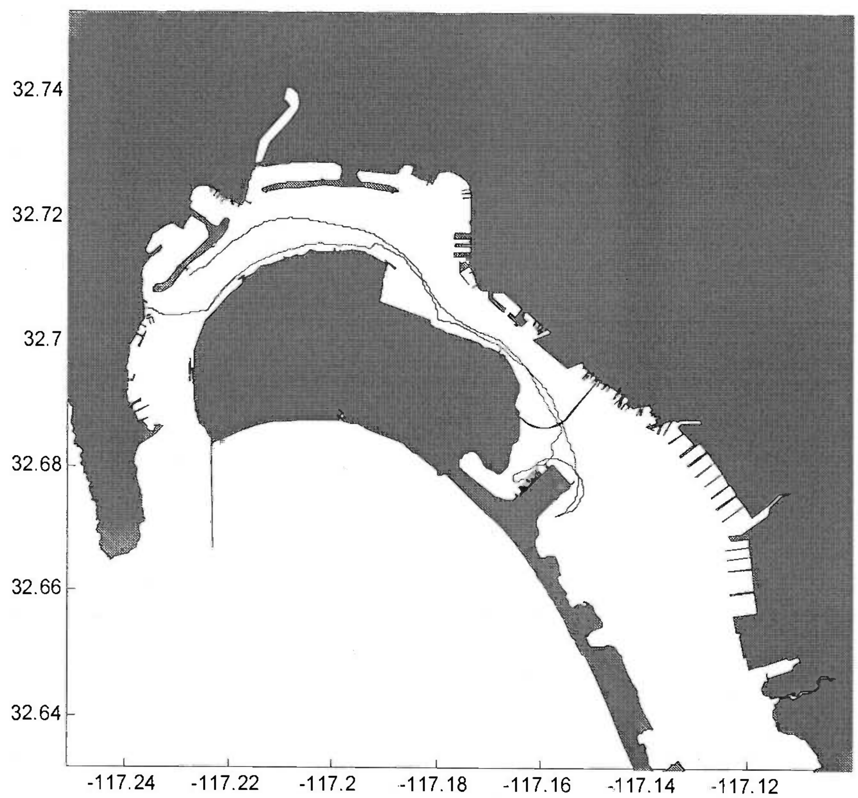

32.695
32.694
32.693
32.692
32.691
32.69
32.689
32.688
32.687
32.686
32.685
-117.238
-117.236
-117.234
-117.232
-117.23
100
90
80
70
60
50
40
30
20
10
0
% Transmission
Post-Survey

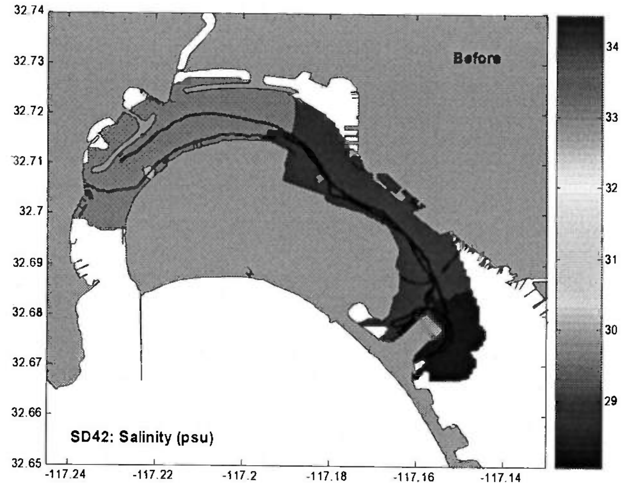
Appendix G3

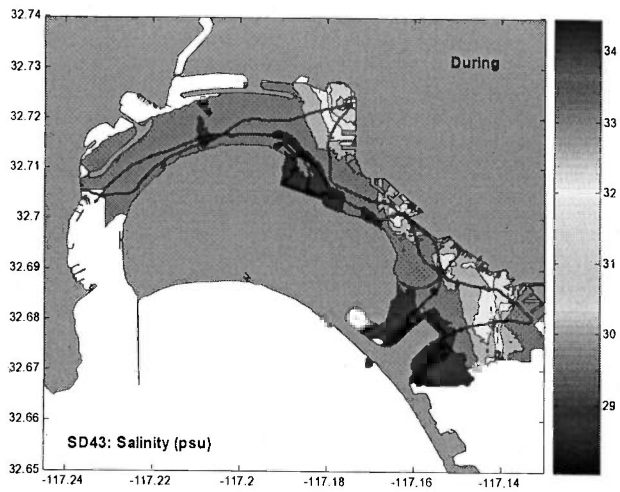
NAB/NI

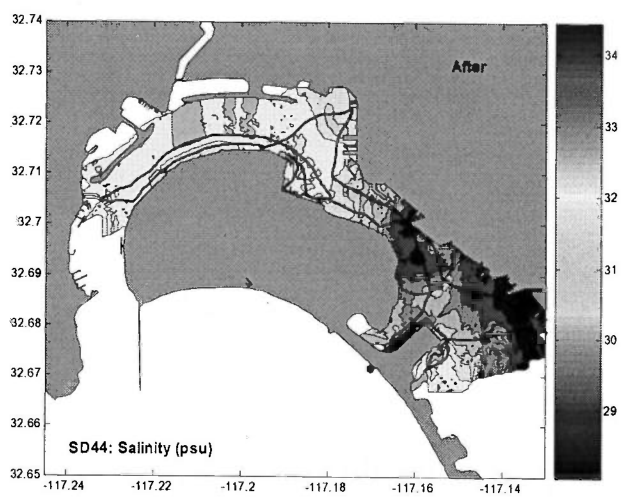
SDB4- 10/17/2004	NAB/NI
SDB6- 2/10/2005	NAB/NI, NAB, NI
SDB7- 4/27/2005	NAB/NI, NAB, NI

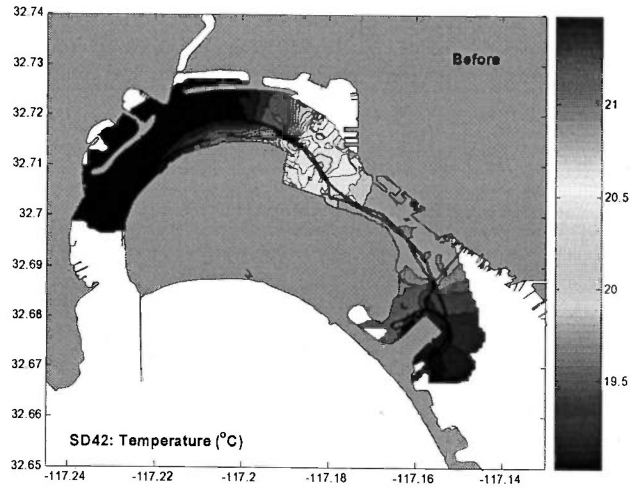
SDB4- 10/17/2004

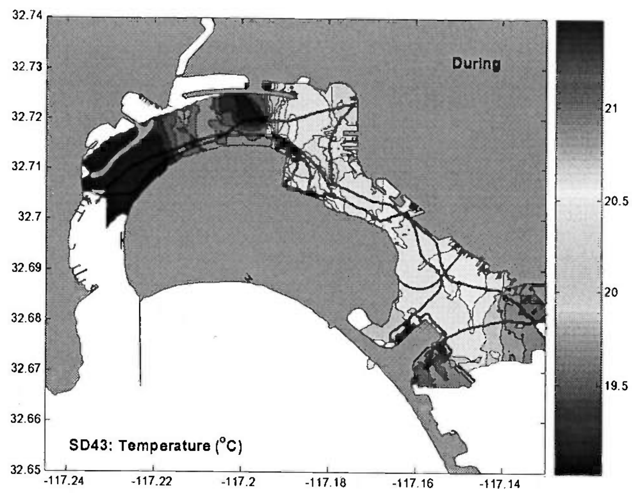


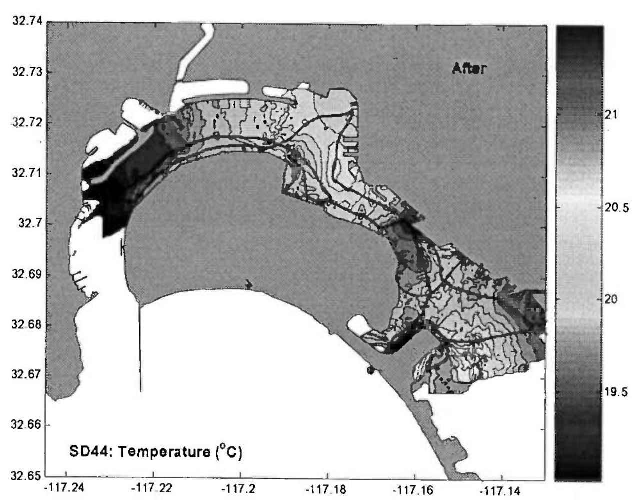

Before
SD42: Salinity (psu)

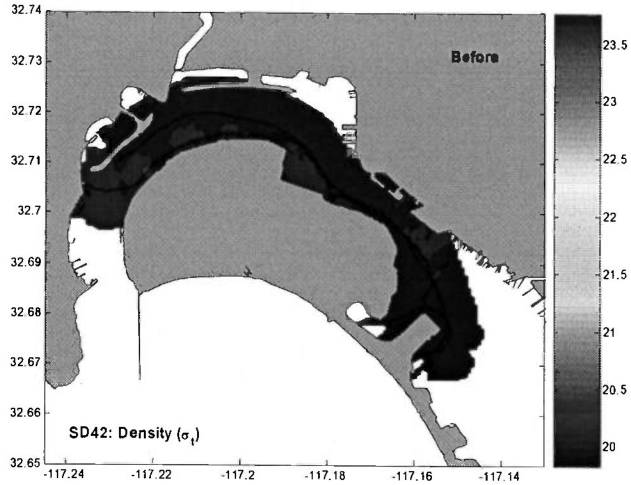

During
SD43: Salinity (psu)

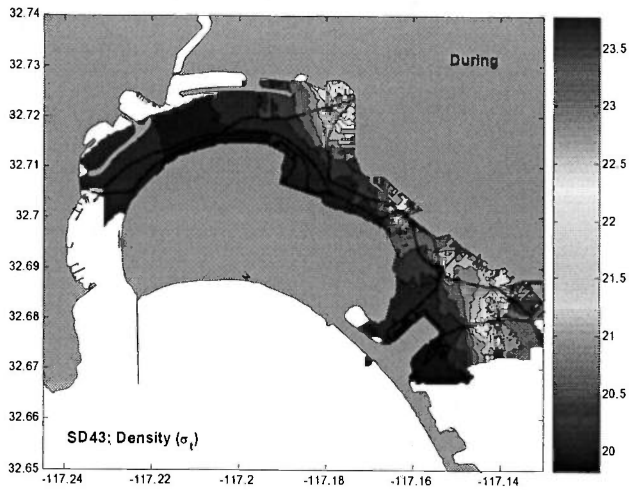

After
SD44: Salinity (psu)

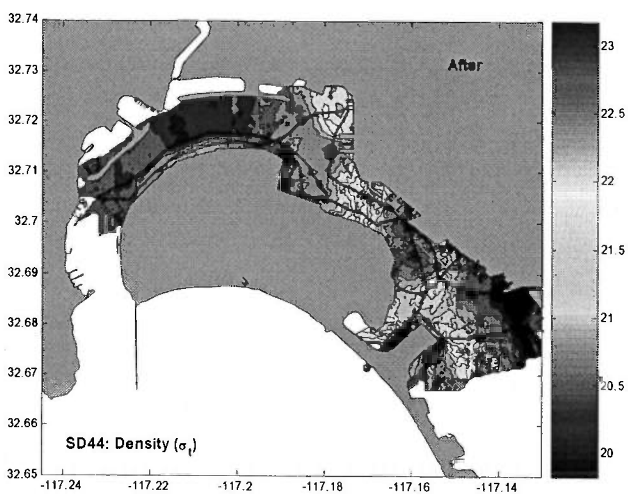

Before
SD42: Temperature (°C)
32.74
32.73
32.72
32.71
32.7
32.69
32.68
32.67
32.66
32.65
-117.24
-117.22
-117.2
-117.18
-117.16
-117.14
21
20.5
20
19.5

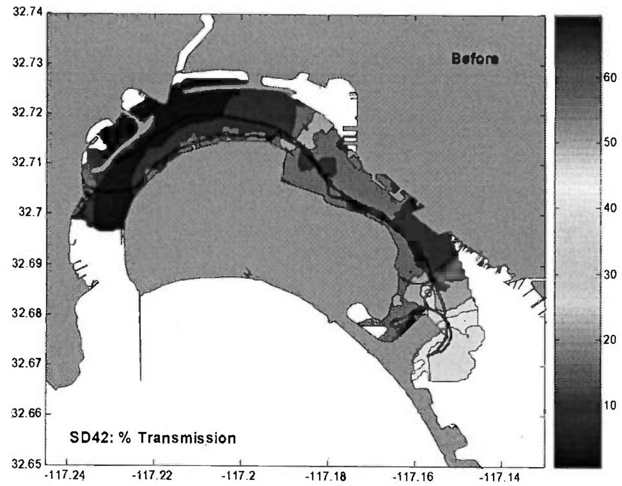

During
SD43: Temperature (°C)
32.74
32.73
32.72
32.71
32.7
32.69
32.68
32.67
32.66
32.65
-117.24
-117.22
-117.2
-117.18
-117.16
-117.14
21
20.5
20
19.5

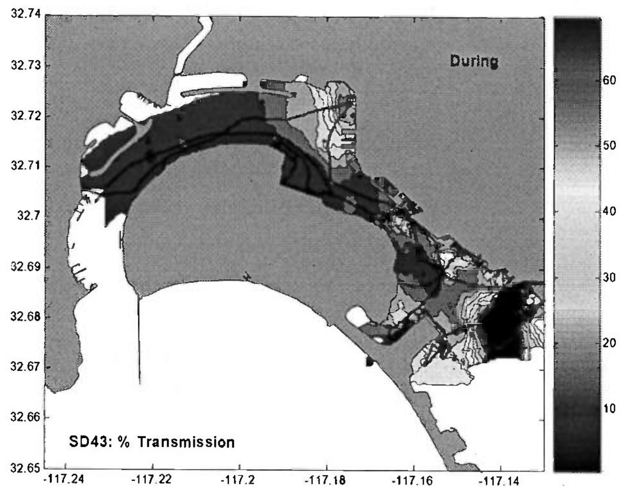

After
SD44: Temperature (°C)
32.74
32.73
32.72
32.71
32.7
32.69
32.68
32.67
32.66
32.65
-117.24
-117.22
-117.2
-117.18
-117.16
-117.14
21
20.5
20
19.5

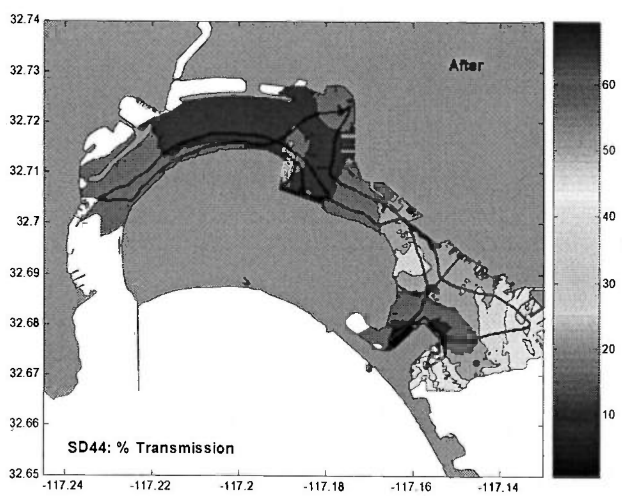

Before
SD42: Density (σ_t)


During
SD43: Density (σ_t)

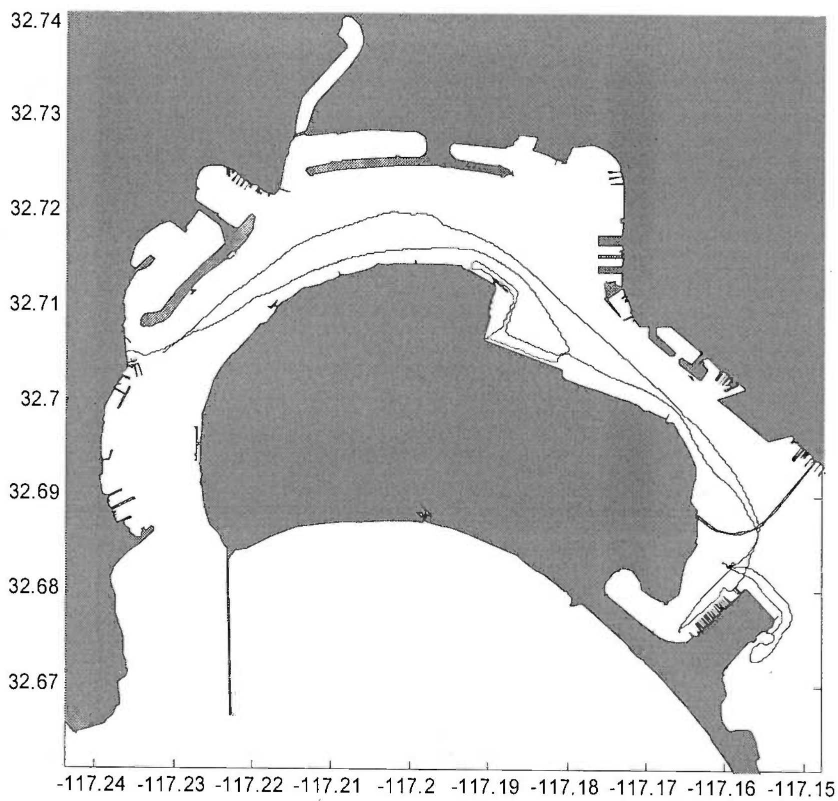

After
SD44: Density (σ_t)


Before
SD42: % Transmission
32.74
32.73
32.72
32.71
32.7
32.69
32.68
32.67
32.66
32.65
-117.24
-117.22
-117.2
-117.18
-117.16
-117.14
60
50
40
30
20
10

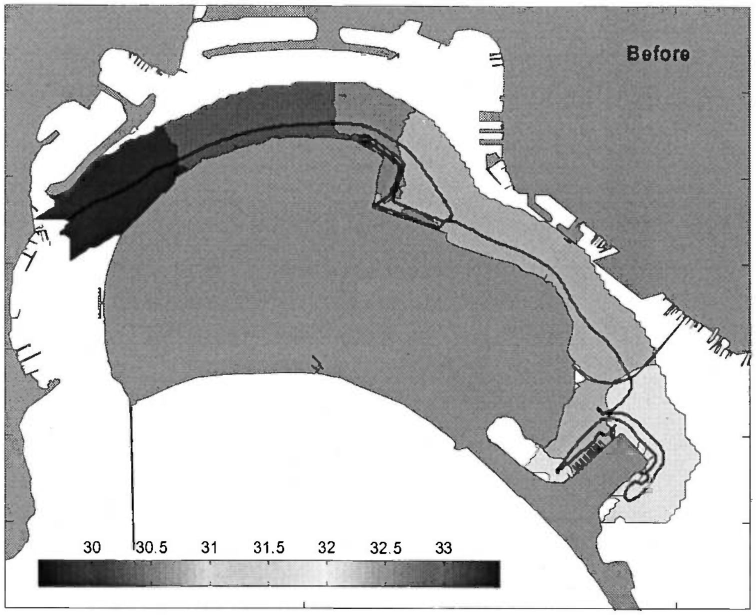

During
SD43: % Transmission
32.74
32.73
32.72
32.71
32.7
32.69
32.68
32.67
32.66
32.65
-117.24
-117.22
-117.2
-117.18
-117.16
-117.14
60
50
40
30
20
10


After
SD44: % Transmission
32.74
32.73
32.72
32.71
32.7
32.69
32.68
32.67
32.66
32.65
-117.24
-117.22
-117.2
-117.18
-117.16
-117.14
60
50
40
30
20
10

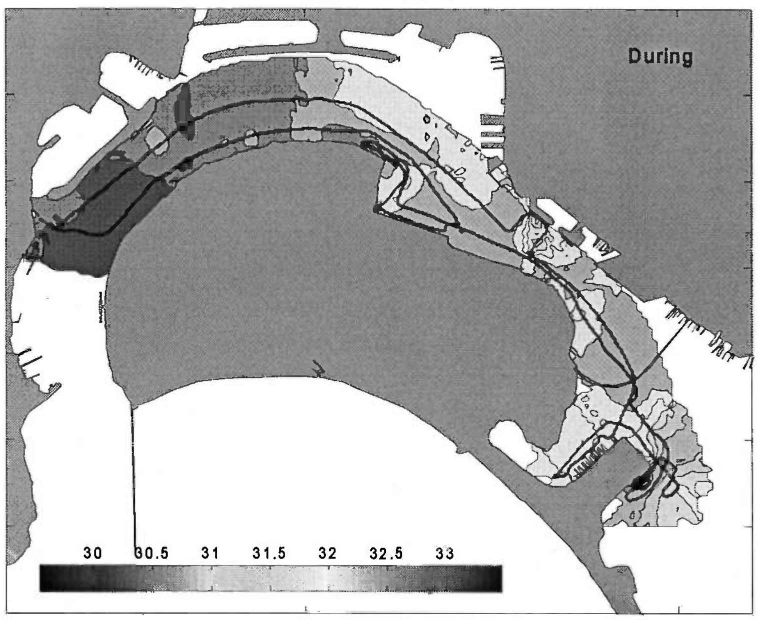
SDB6- 2/10/2005



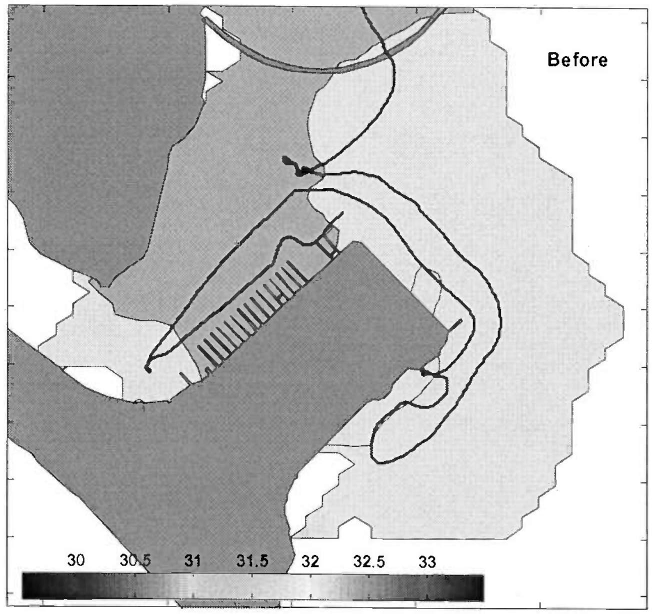
SD47: Salinity (psu)



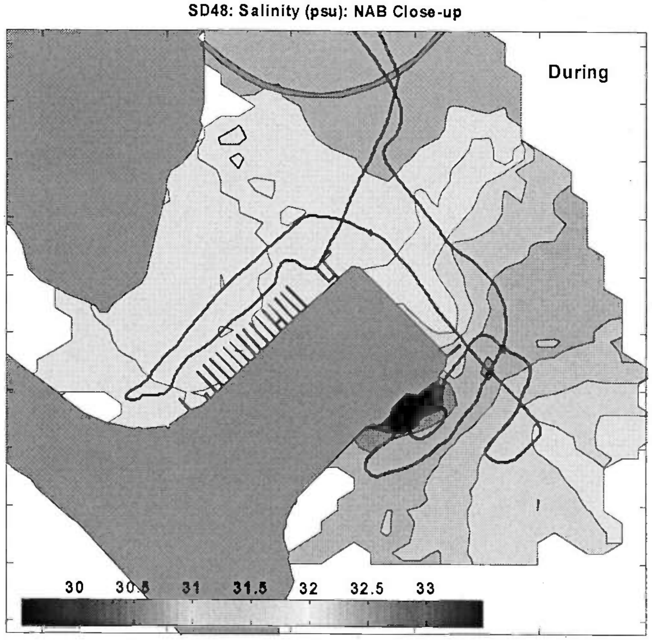
SD48: Salinity (psu)



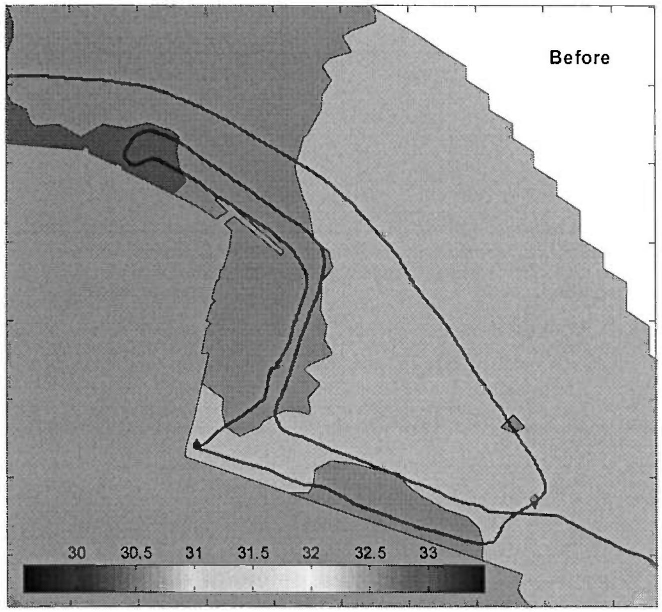
SD47: Salinity (psu): NAB Close-up



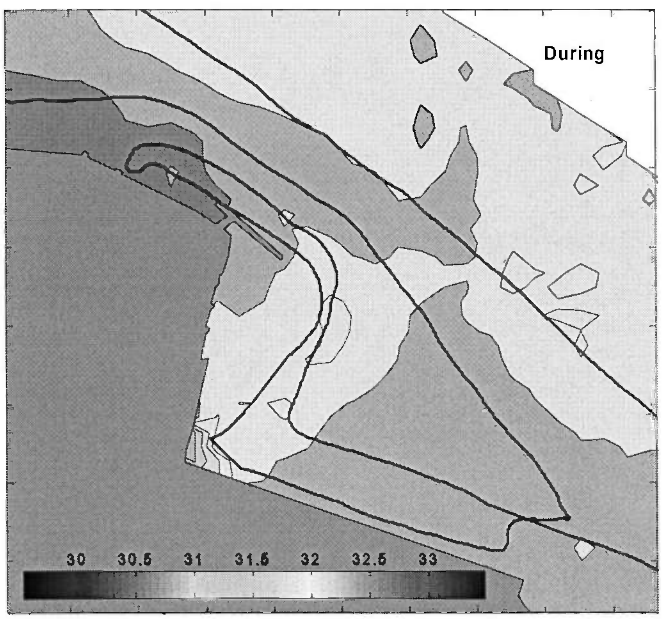
SD48: Salinity (psu): NAB Close-up



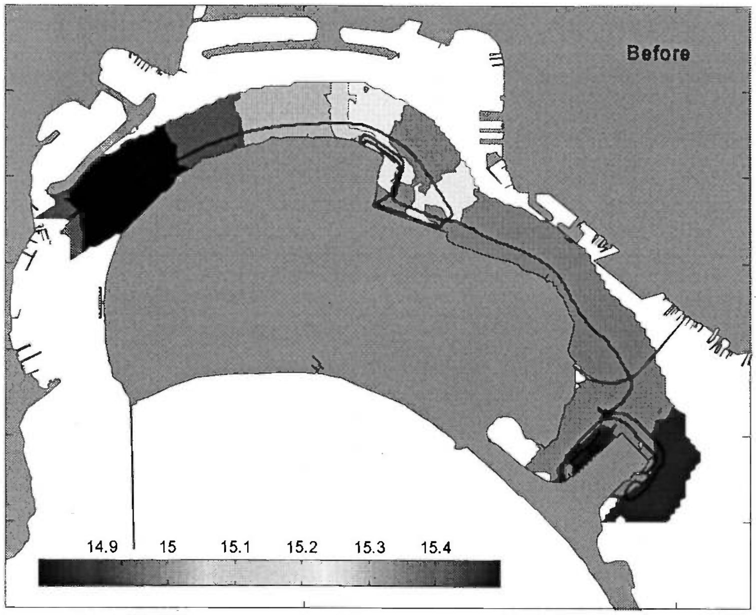
SD47: Salinity (psu): NASNI Close-up



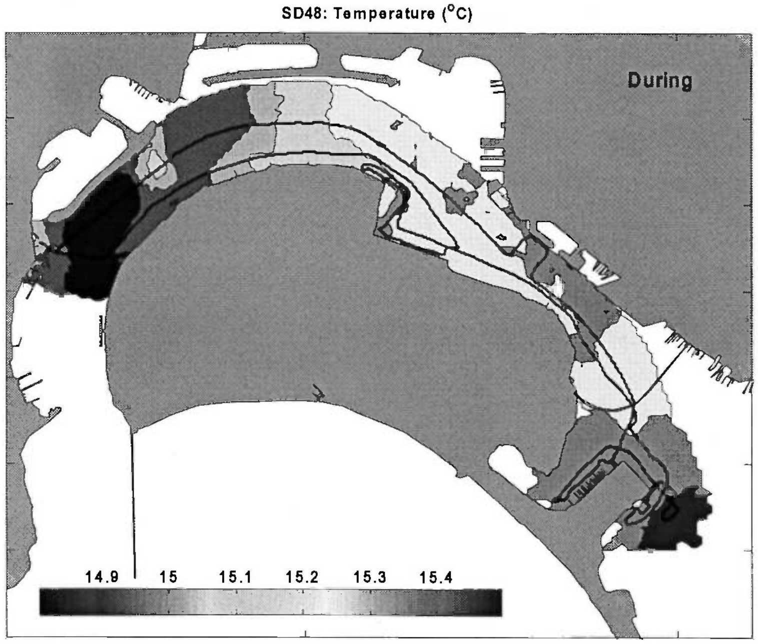
SD48: Salinity (psu): NASNI Close-up



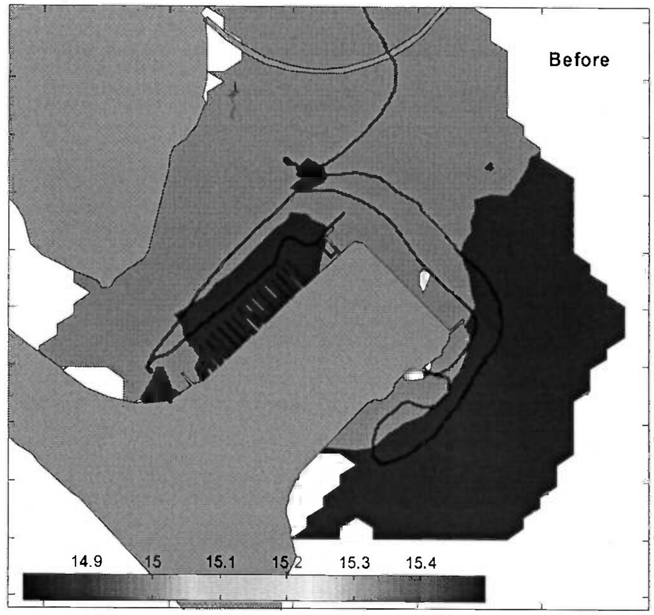
SD47: Temperature (°C)



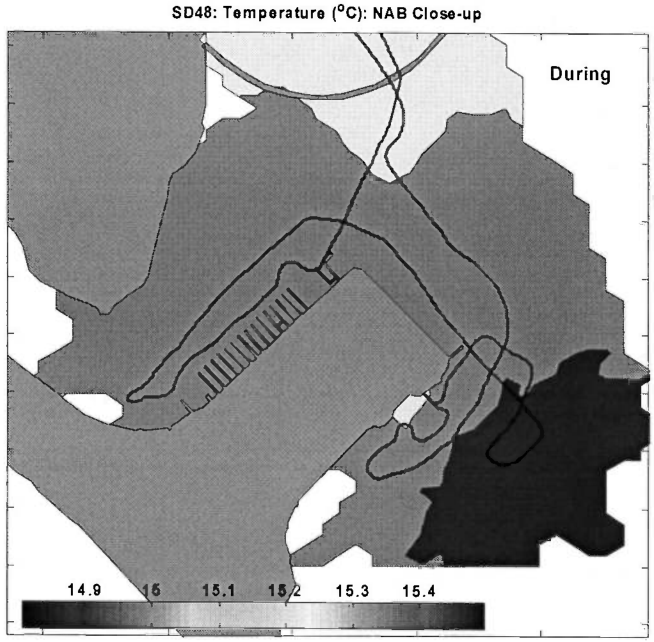
SD48: Temperature (°C)



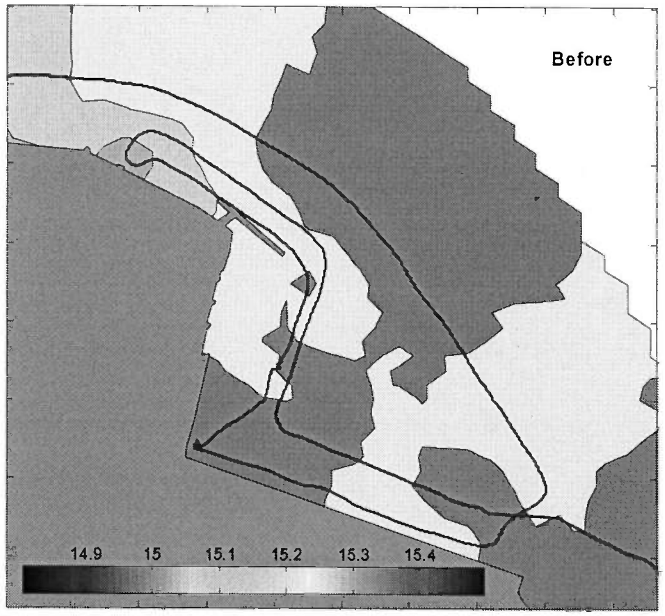
SD47: Temperature (°C): NAB Close-up



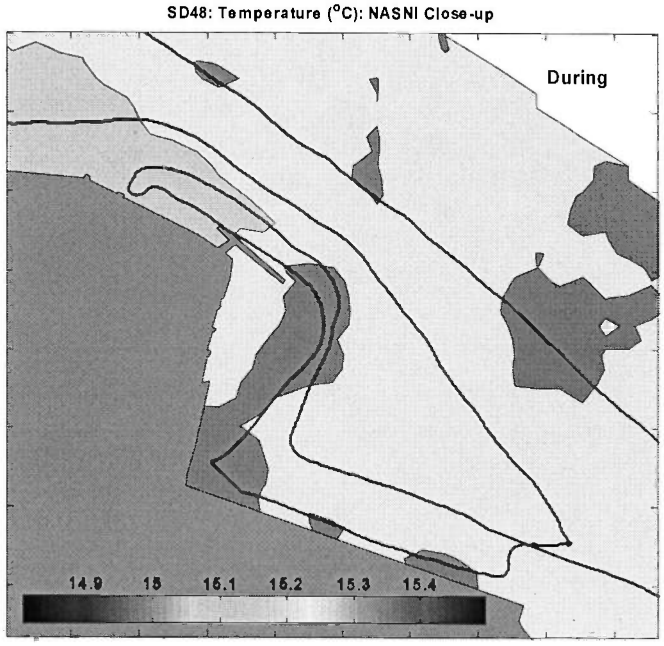
SD48: Temperature (°C): NAB Close-up



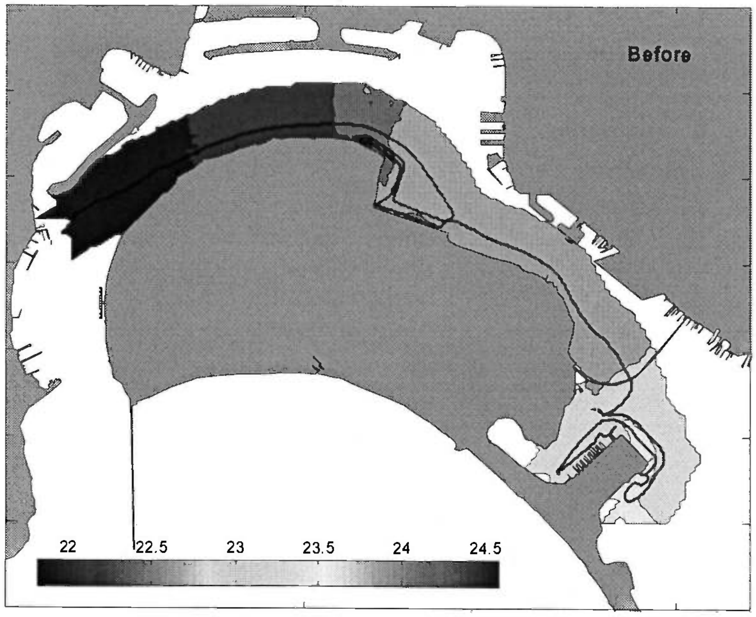
SD47: Temperature (°C): NASNI Close-up



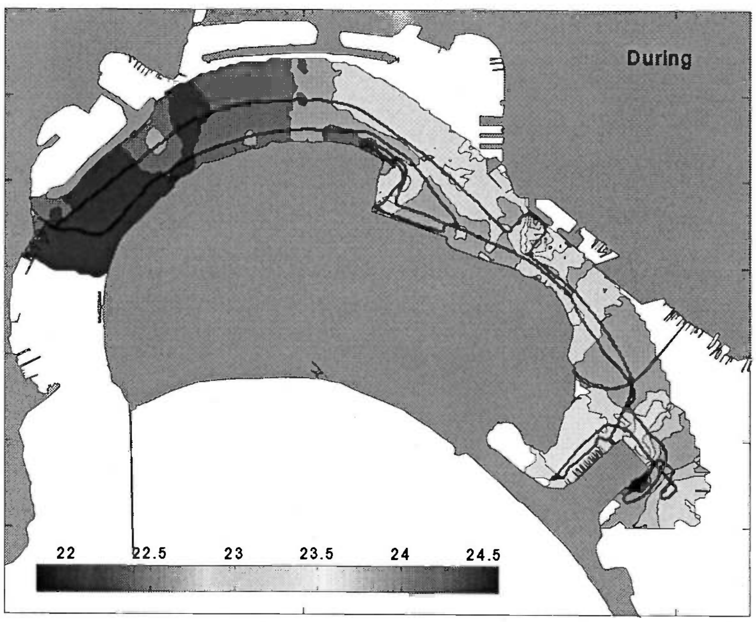
SD48: Temperature (°C): NASNI Close-up



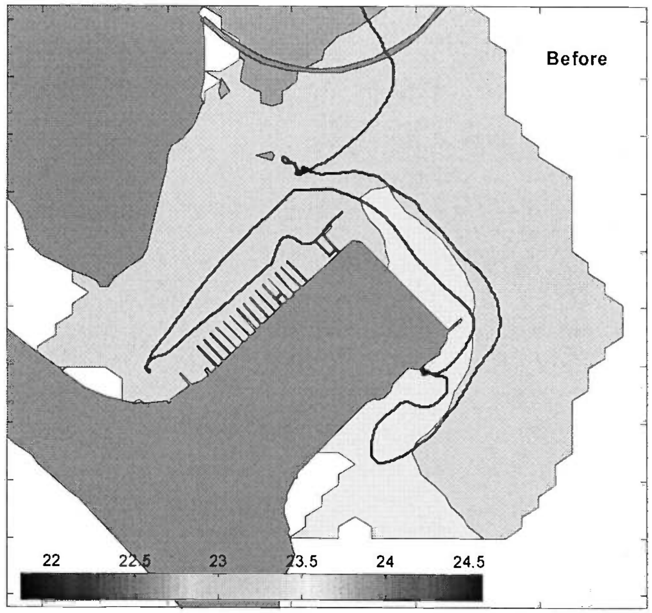
SD47: Density (σ_t)



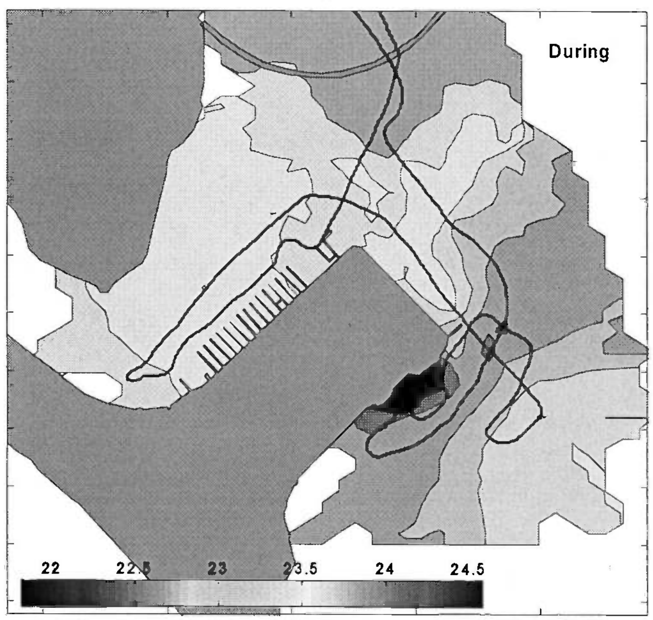
SD48: Density (σ_t)



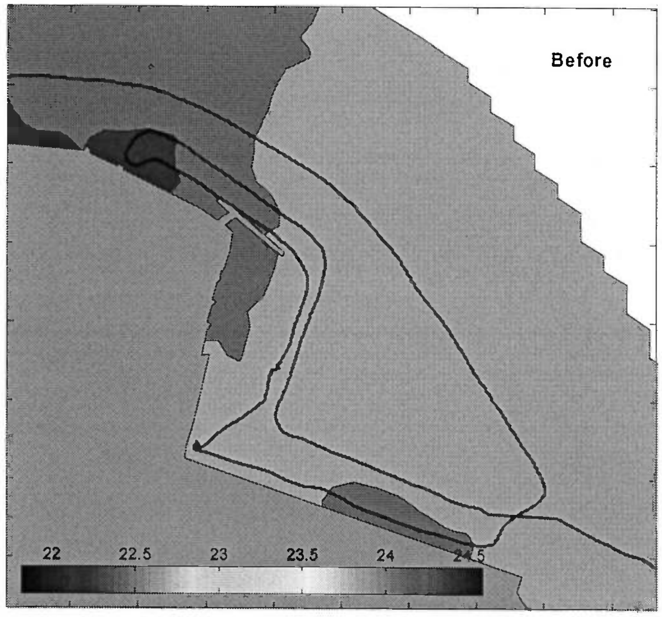
SD47: Density (σ_t): NAB Close-up



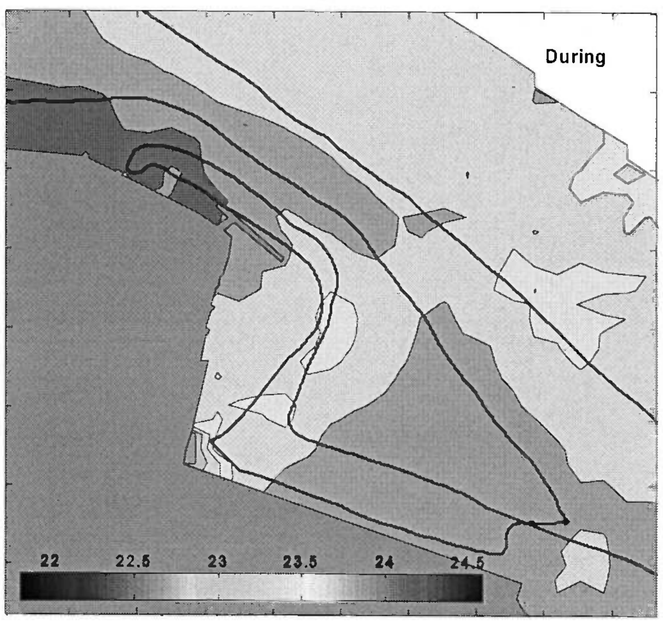
SD48: Density (σ_t): NAB Close-up



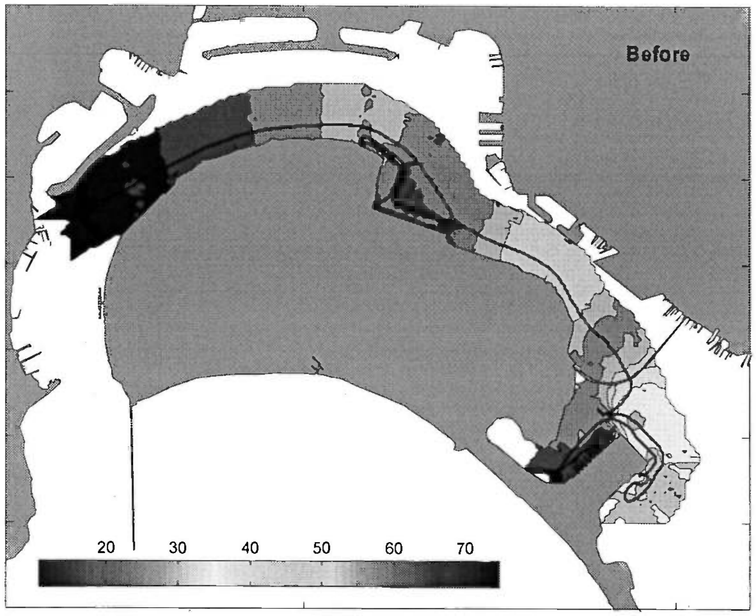
SD47: Density (σ_t): NASNI Close-up



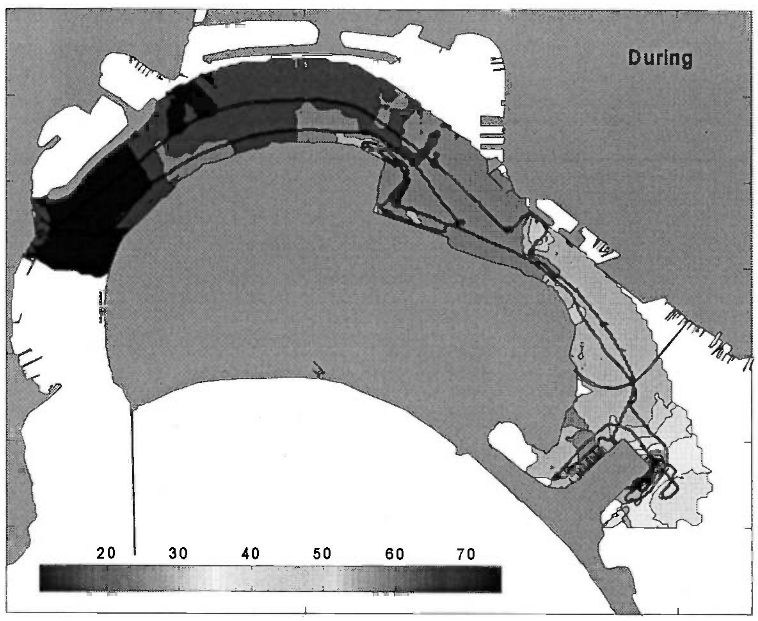
SD48: Density (σ_t): NASNI Close-up



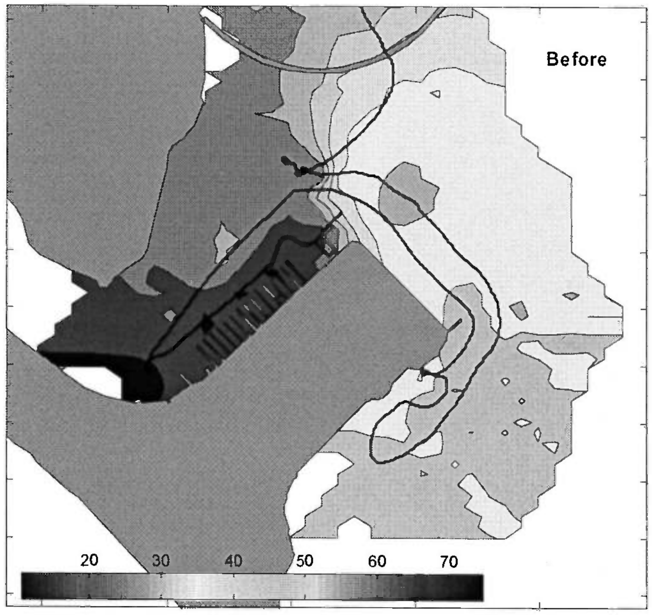
SD47: % Transmission



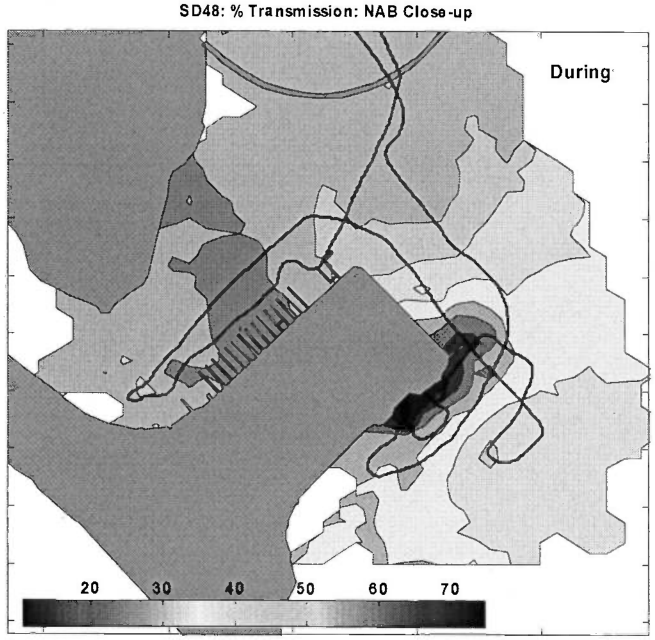
SD48: % Transmission



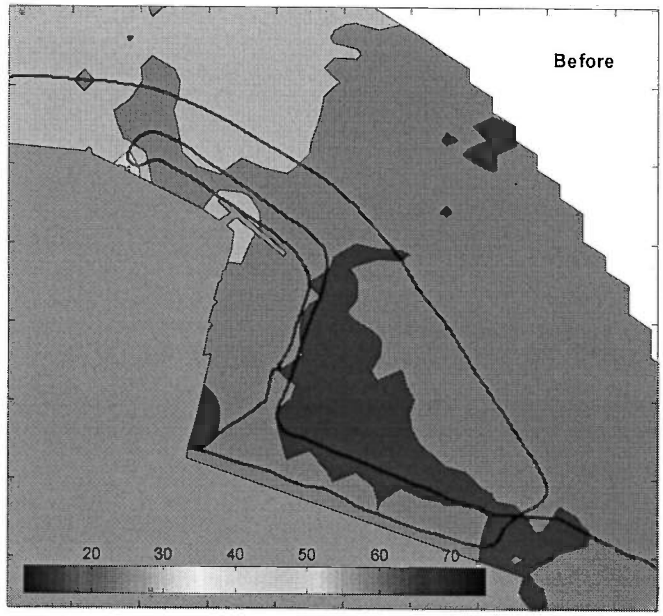
SD47: % Transmission: NAB Close-up



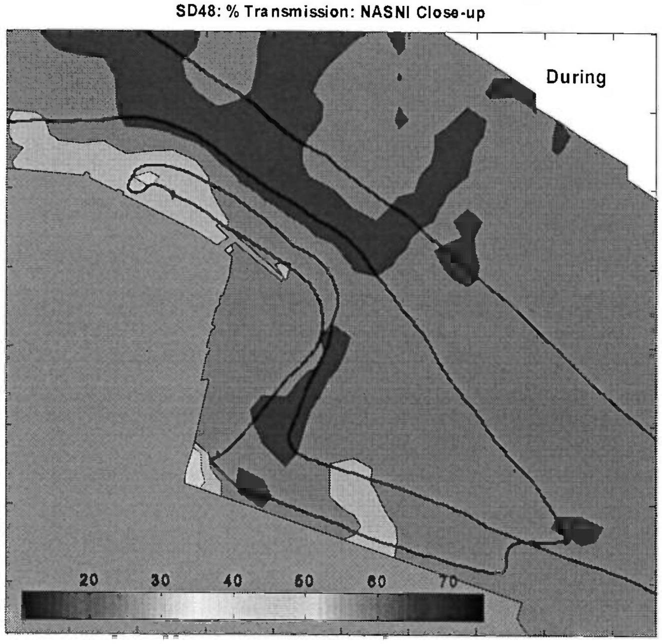
SD48: % Transmission: NAB Close-up



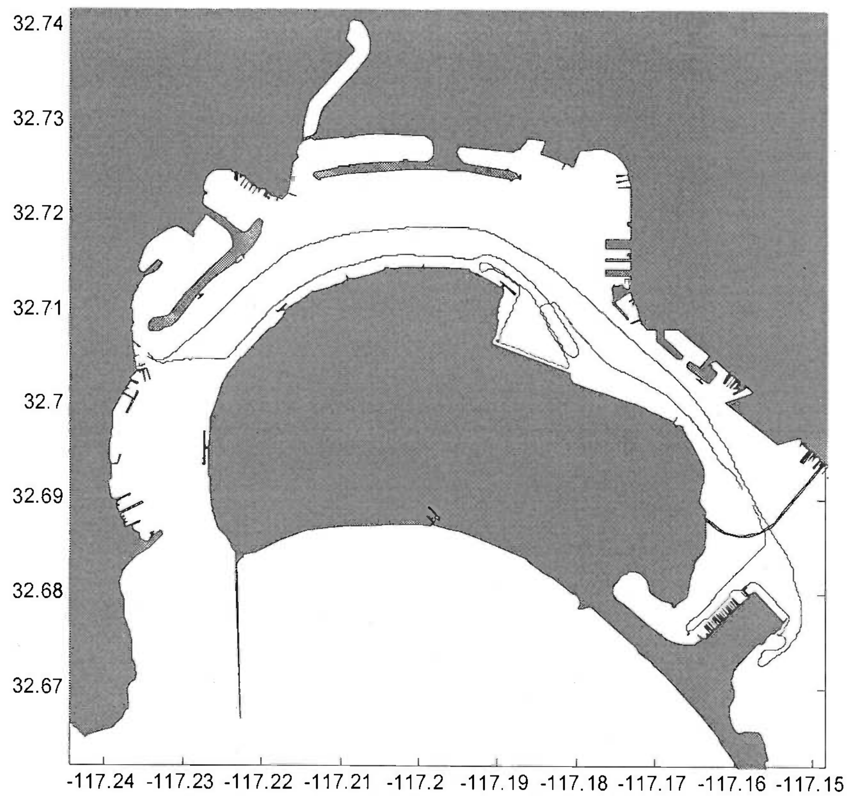
SD47: % Transmission: NASNI Close-up



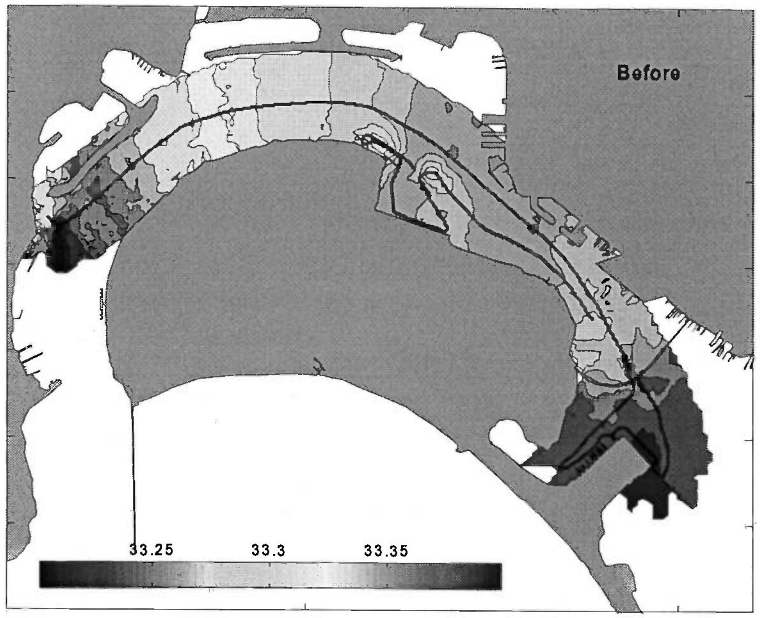
SD48: % Transmission: NASNI Close-up



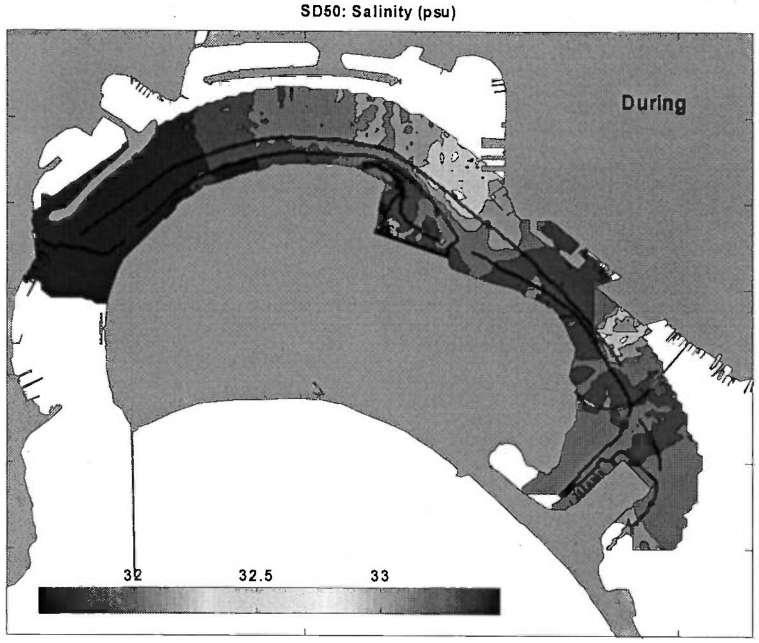
SDB7- 4/27/2005



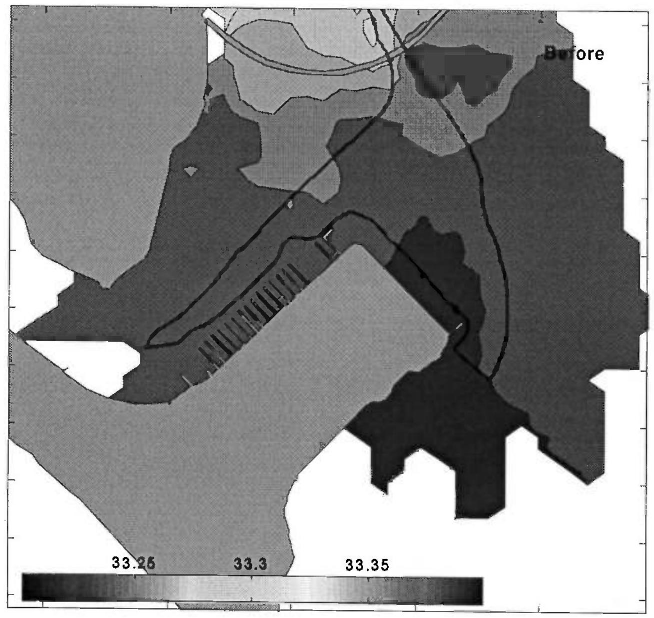
SD49: Salinity (psu)



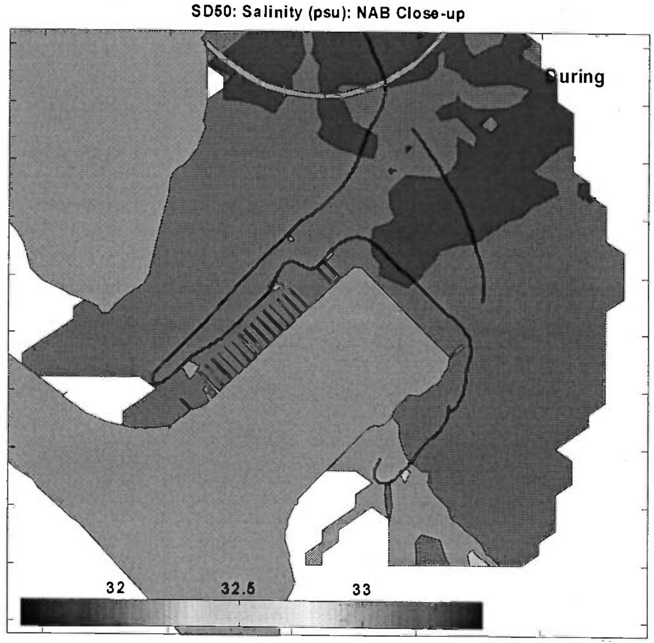
SD50: Salinity (psu)



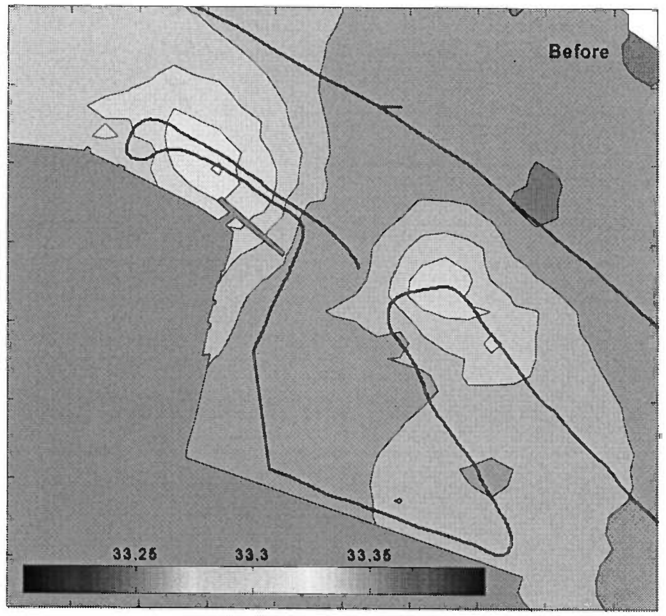
SD49: Salinity (psu): NAB Close-up



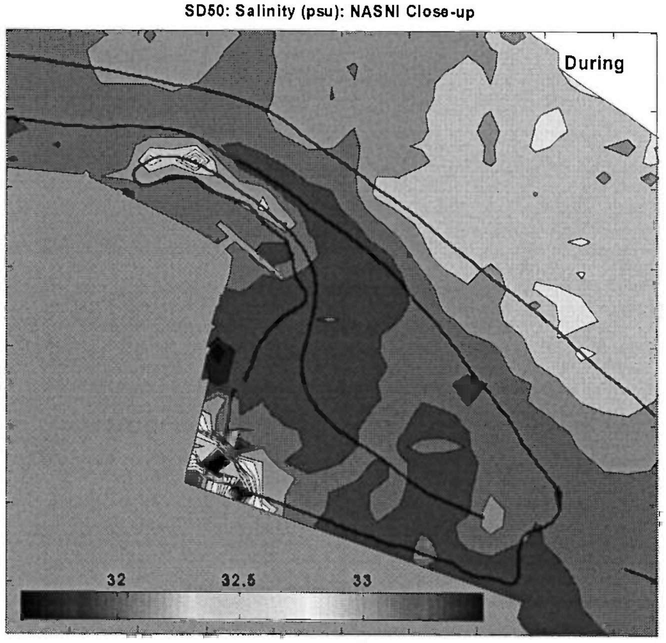
SD50: Salinity (psu): NAB Close-up



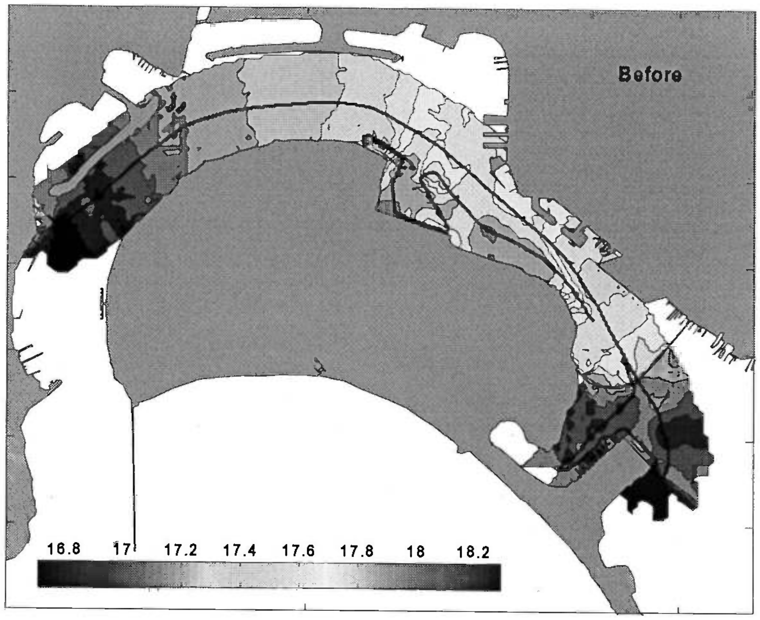
SD49: Salinity (psu): NASNI Close-up



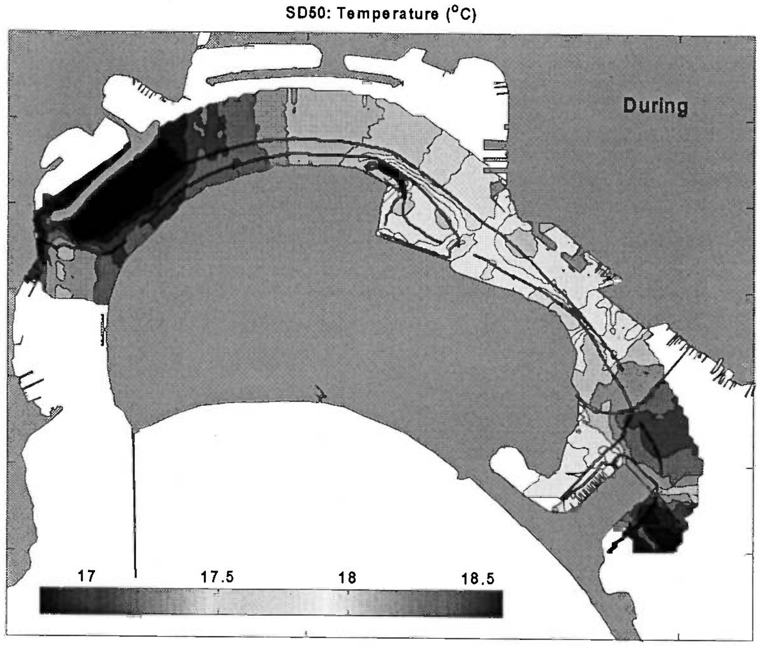
SD50: Salinity (psu): NASNI Close-up



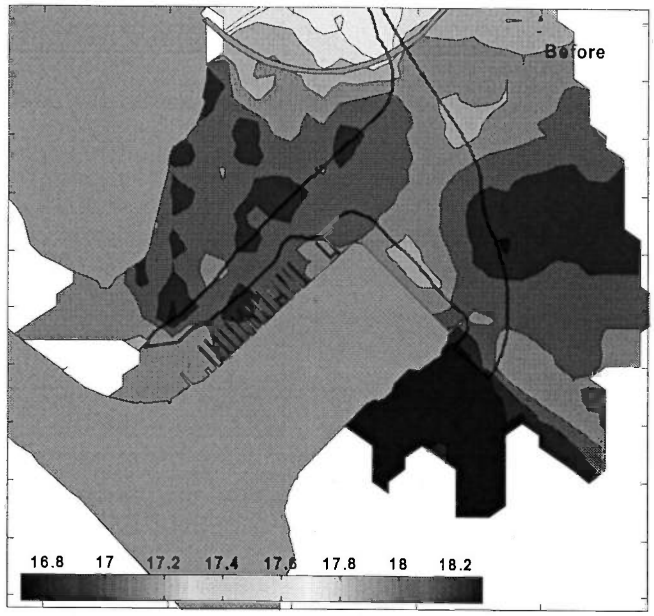
SD49: Temperature (°C)



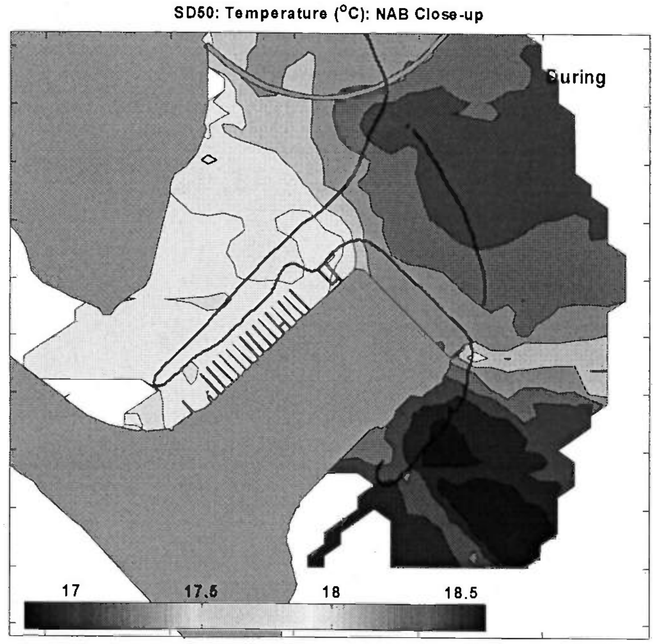
SD50: Temperature (°C)



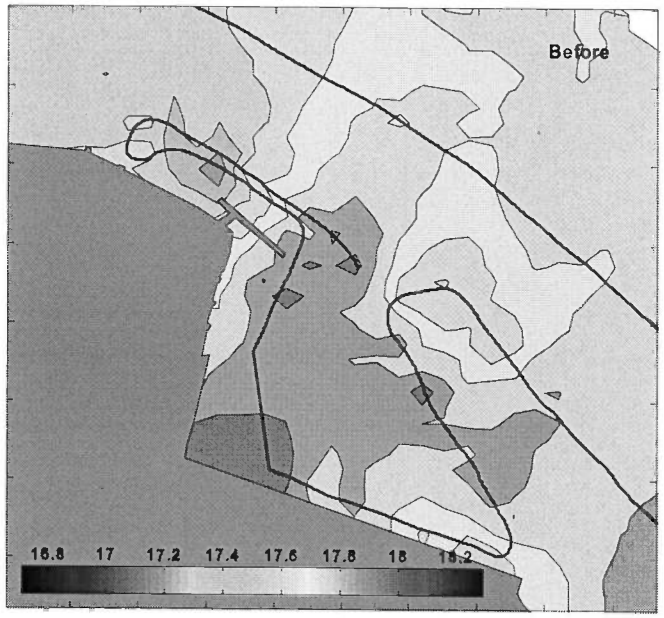
SD49: Temperature (°C): NAB Close-up



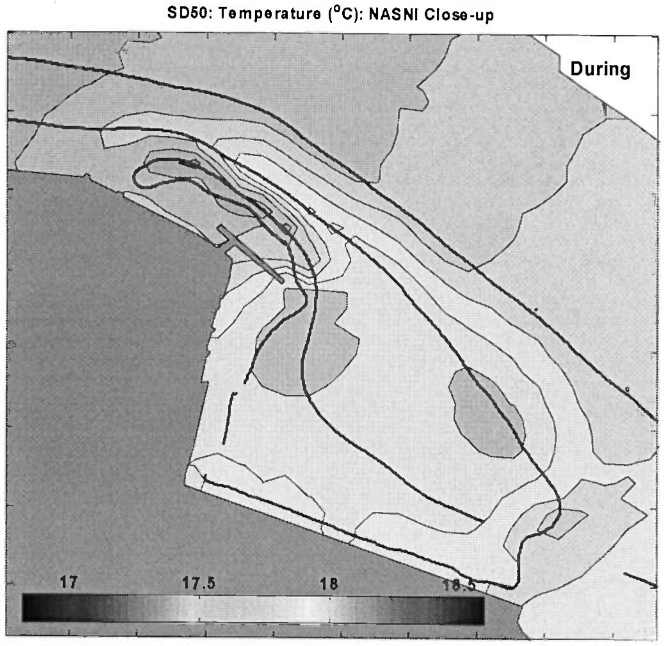
SD50: Temperature (°C): NAB Close-up



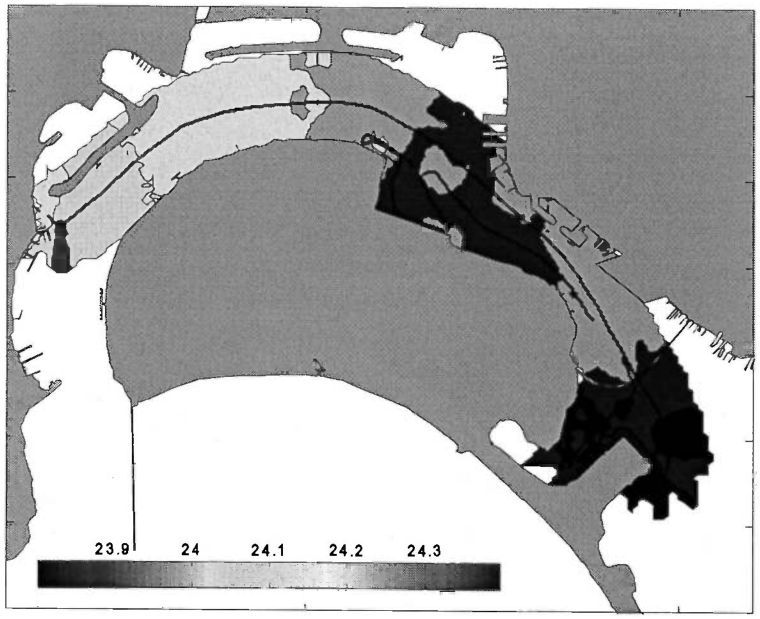
SD49: Temperature (°C): NASNI Close-up



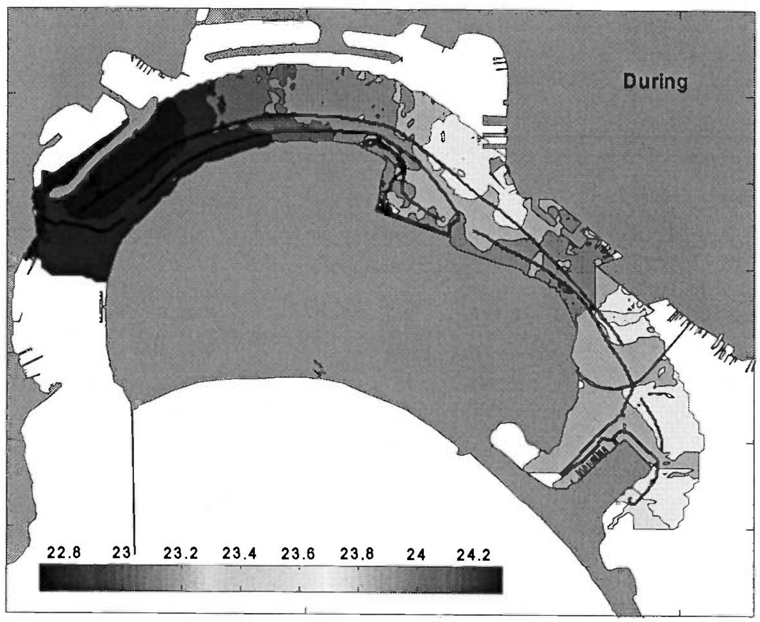
SD50: Temperature (°C): NASNI Close-up



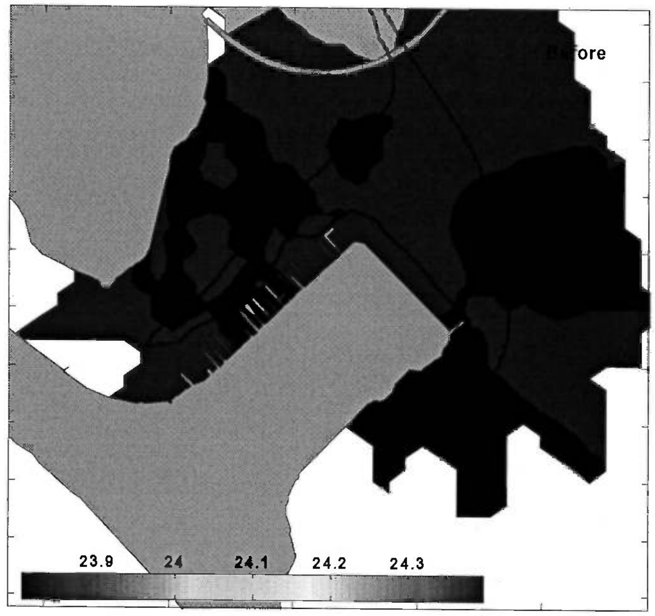
SD49: Density (σ_t)



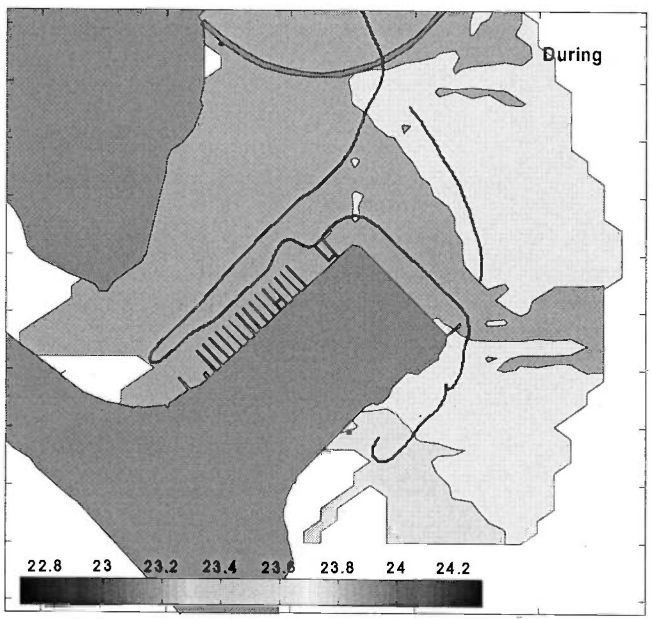
SD50: Density (σ_t)



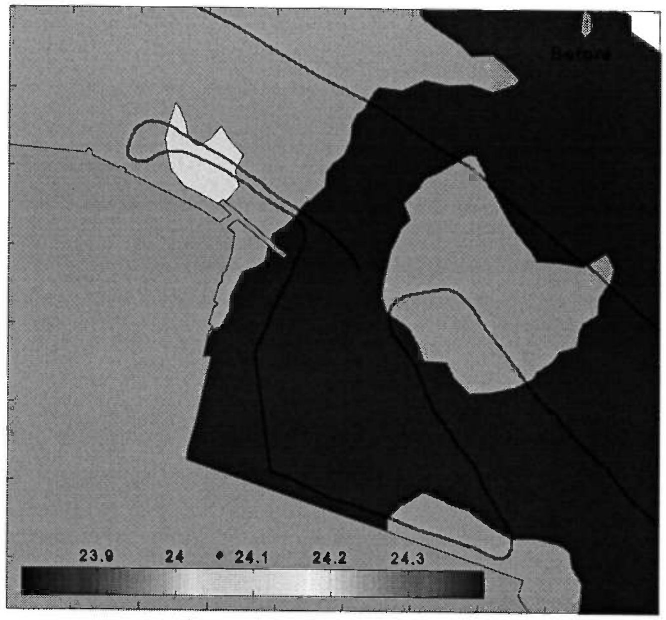
SD49: Density (σ_t): NAB Close-up



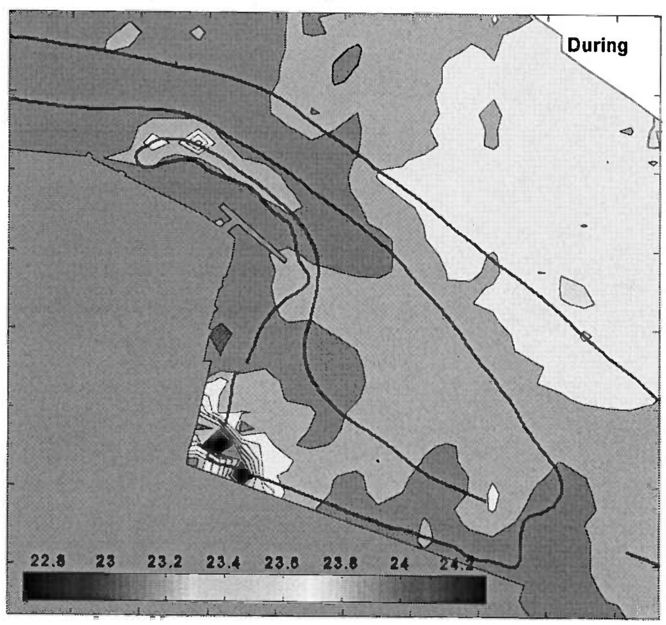
SD50: Density (σ_t): NAB Close-up



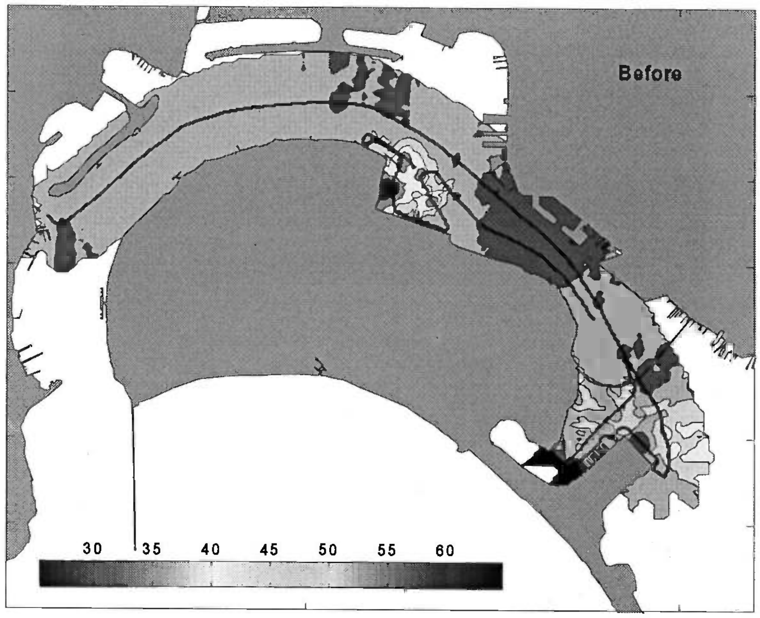
SD49: Density (σ_t): NASNI Close-up



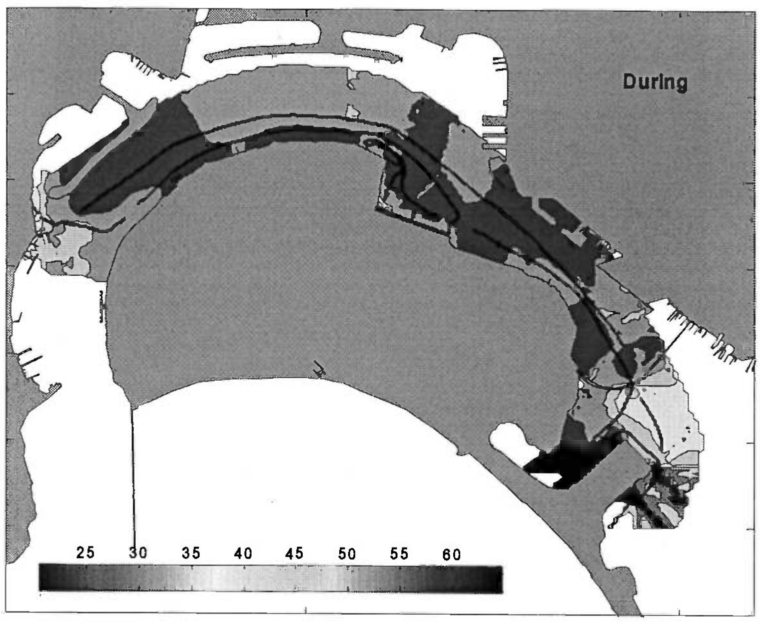
SD50: Density (σ_t): NASNI Close-up



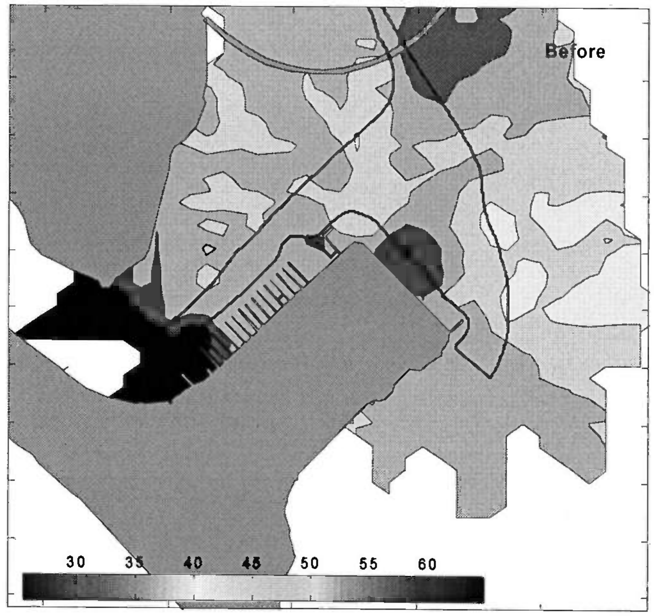
SD49: % Transmission



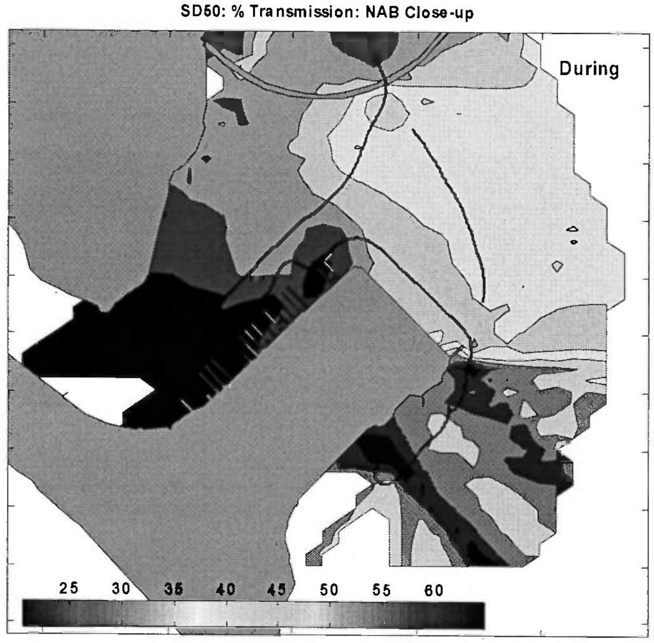
SD50: % Transmission



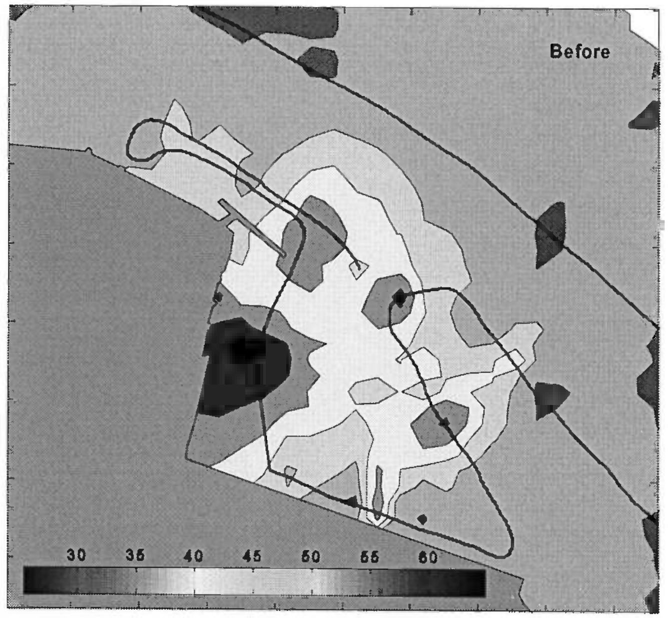
SD49: % Transmission: NAB Close-up



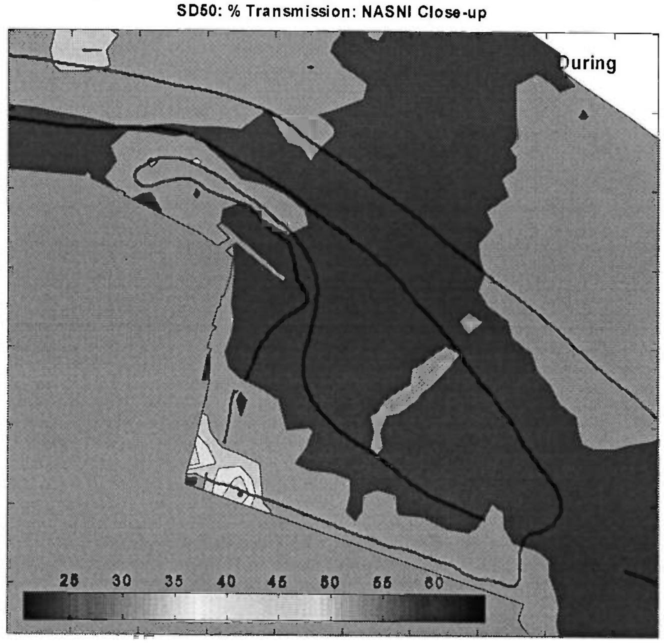
SD50: % Transmission: NAB Close-up



SD49: % Transmission: NASNI Close-up



SD50: % Transmission: NASNI Close-up



APPENDIX H

Floating Bioassay Laboratory Study

Paper presented in Proceedings of Marine Technology Society, Oceans 2005 Conference

Katz, C.N. and G. Rosen, 2005. *Evaluating storm water impacts- monitoring the receiving environment using a floating bioassay laboratory system,* In: Proceedings of the Marine Technology Society, Oceans 2005 Conference, September 18-23, 2005, Washington, D.C. 8 pp.

Evaluating Storm Water Impacts- Monitoring the Receiving Environment Using a Floating Bioassay Laboratory System

C. N. Katz and G. Rosen
Environmental Sciences and Applied Systems Branch, Code 2375
Space and Naval Warfare Systems Center San Diego
San Diego, CA 92152

Abstract - The U.S. Navy is conducting an evaluation of impacts from facility storm water discharges to San Diego Bay. The investigation was prompted by the implementation of local regulations that require a 90% survival rate of fish or mysid shrimp in acute toxicity tests using undiluted storm water. An underlying conceptual approach was to monitor toxicity directly in receiving waters as well as in the undiluted discharge to evaluate impacts. Data collected to date have shown a full range in toxic response in outfall discharge samples. No toxic effects, however, have been observed in bay waters collected immediately outside the outfalls. These results, along with plume mapping, have suggested that the relatively small magnitude and ephemeral nature of these discharges were sufficient to explain the removal of toxicity of the storm discharge once it reaches the bay.

One of the outstanding issues presented by standard toxicity testing is the relevance of 48- or 96-h exposure times to test organisms when actual storm exposures likely occur over much shorter times. To investigate this issue, we conducted toxicity tests with a boat-mounted flow-through bioassay system, which was positioned immediately outside an outfall before, during, and after a storm event. The bioassays included survival of the mysid (*Americamysis bahia*) and topsmelt (*Atherinops affinis*) as well as embryo-larval development of the mussel (*Mytilus galloprovincialis*). Surface bay water was continuously pumped to the test organism containers for the full 48- (mussel) or 96-h (mysid and topsmelt) exposure requirement. Bay water was analyzed continuously for salinity, temperature, dissolved oxygen, pH, light transmission, oil fluorescence, copper, and zinc. Additionally, dilution series toxicity and chemistry were conducted on first-flush and composite samples taken from the outfall prior to discharge. The floating-bioassay system results were consistent with previous monitoring that indicated toxicity of first-flush discharges but no toxicity in the receiving environment. Continuous monitoring showed that storm water was completely mixed out within minutes of discharging to the bay even though the observation point was only 15 feet away from the outfall. The reduction of toxicity in the receiving environment was a result of the very limited time exposure that occurs with this type of discharge.

INTRODUCTION

Industrial storm water discharges from Navy Facilities were investigated in 2003 through 2005 for their impact to San Diego Bay. The investigation was prompted by implementation of a new National Pollutant Discharge Elimination System (NPDES) storm water permit that required storm water collected at the end-of-pipe meet a toxicity requirement of 90% survival using a standard laboratory acute bioassay. The ostensible goal of this requirement is to ensure that bay waters are protected from these discharges. While the 90% threshold should be protective of the receiving environment, the Navy believed that the requirement was overly stringent. The Navy asked the local regulatory agency for permission to conduct an evaluation of storm water toxicity and propose a scientifically-based toxicity requirement.

The Environmental Sciences and Applied Systems Branch at the Space and Naval Warfare Systems Center San Diego (SSC-SD) executed a study to investigate the nature of industrial storm water toxicity and its impacts from four Navy facilities bordering San Diego Bay. The approach taken was to evaluate storm water collected at or near its point of discharge to the bay as well as in bay waters collected immediately outside these discharge points. Additionally, bay waters were monitored before, during, and after storm events using plume tracking techniques to evaluate both their spatial and temporal extent. This approach was designed to evaluate if the measurements made at onshore monitoring locations were predictive of actual receiving water impacts. A summary of the overall results of this investigation is provided here as background.

BACKGROUND

The toxicity investigation was conducted during the October through April wet seasons from 2002 through 2005. During that time period, 11 storms were sampled with rainfall totals ranging from 0.1" up to a record 3.5". Antecedent dry periods ranged from five days up to a record dry period of six months. A total of 13 different industrial storm water drainage areas were sampled at four bases including four piers. The samples represented drainage from 0.5 to 75 acres that included industrial facilities. A total of 41 storm water samples were collected from the end-of-pipe, including 26 first-flush samples (as required in the permit) and 15 full-storm composite samples. A total of 63 bay samples were collected outside these outfalls before, during, and after the storm events. These samples were evaluated using three standard EPA-approved laboratory bioassays: the 96-h survival of *Atherinops affinis* (topsmelt) larvae and *Americamysis bahia* (mysid) juveniles, and 48-h normal embryo-larval development of *Mytilus galloprovincialis* (mussel). The two survival tests were called out in the NPDES permit, whereas the embryo development test was added to provide a highly sensitive test for bay samples. The samples were also analyzed for a suite of chemicals that included total and dissolved metals, dissolved organic carbon (DOC), total suspended solids (TSS), polynuclear aromatic hydrocarbons (PAH), polychlorinated biphenyls (PCB), and chlorinated pesticides.

Results of the storm water bioassays with topsmelt and mysid varied the full range from 0% to 100% survival and averaged about 75%. The tests failed the 90% toxicity requirement about 60% of the time with no significant differences between species. In contrast, the toxicity measured in bay waters immediately outside the outfalls were not toxic and had a very narrow range of results (90 to 100%), averaging ~97% survival for the two species. The mussel embryo-larval development test showed comparable results with a high degree of variability in the storm water samples, ranging from 0 to 97% and averaged 15% normal development. Bay samples averaged 90% normal development, with the exception of two samples collected during a storm event collected after a record six month antecedent dry period. A Toxicity Identification Evaluation (TIE) established copper and zinc as the primary causative agents of the observed toxicity in the storm water samples.

The observed reduction in toxicity was attributed to rapid dilution in the receiving environment, as measured by reduced chemical concentrations and observed from the plume mapping data. As shown in Fig. 1, the maximum amount of fresh water observed (minimum salinity/pre-storm salinity) during storm surveys was about 5%, representing a 20-fold dilution. The storm water signatures were also ephemeral, returning to pre-storm conditions within 12 to 24 hours. Thus, bay waters were able to assimilate the industrial storm water discharges from these facilities without resulting in a toxic impact, thus meeting the Clean Water Act narrative goals of "no toxics in toxic amounts" [1].

STUDY GOALS

The rapid reduction of toxicity of storm water after introduction into San Diego Bay waters was investigated further, using a floating bioassay laboratory system. The goal was to monitor the receiving environment throughout a storm event to evaluate impacts under actual exposure conditions immediately outside the point of discharge. The study was designed to provide a detailed understanding of the interaction of storm water with bay waters to help explain the apparent absence of receiving water toxicity.

METHODS

To perform this task, a flow-through bioassay system was placed aboard the research vessel (RV) ECOS, which also housed the Navy's Marine Environmental Survey Capability (MESC). The MESC is a real-time data acquisition system that was used to continuously monitor surface seawater conditions and to supply water to the bioassay system [2-4]. These techniques provided actual exposure conditions for the test organisms and continuous monitoring of the receiving water conditions.

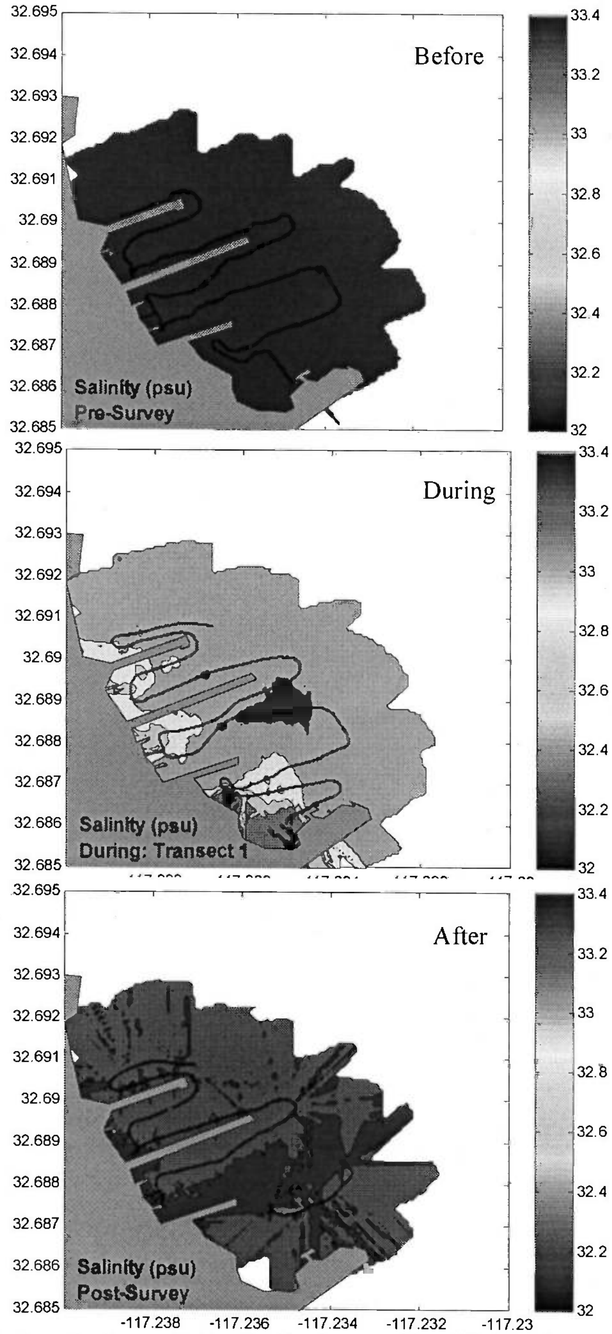


Fig. 1. Surface salinity distribution outside of Submarine Base San Diego before, during, and 28 h after a 25 Feb 2003 storm event.

MONITORING SITE

The site chosen for monitoring was Outfall 14 (OF14) at Naval Base San Diego (NBSD), which enters the bay between Piers 6 and 7 (Fig 2.). The onshore monitoring location was located in a large parking lot about 200 m upstream from the discharge point through the quay wall. The outfall drains ~53 acres, virtually all of which is impervious surface. The onshore location was estimated to effectively sample 92% of the drainage area. Industrial facilities in this drainage area include vehicle maintenance and divers storage facilities. The outfall is tidally influenced with bay water reaching the monitoring location at a tide height of ~1m. The pipe diameter at the monitoring location was 91 cm. This outfall had been monitored on previous occasions and had shown toxicity in storm water samples, particularly to mussel embryos.

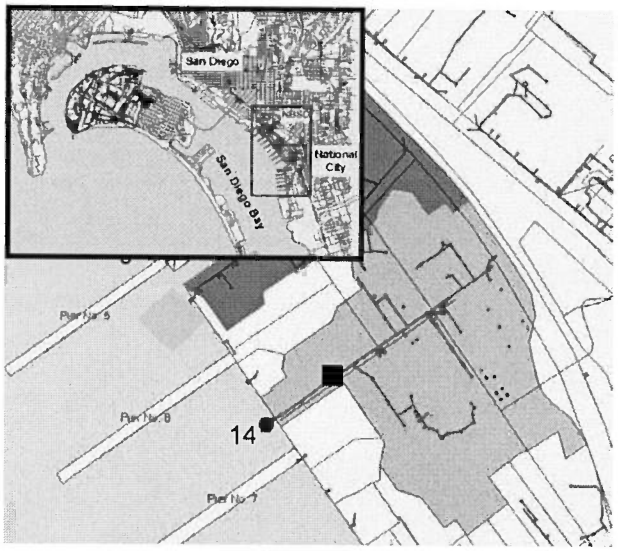


Fig. 2. NBSD Outfall 14, its conveyance system in red and drainage area in pink. The onshore monitoring location is identified by the blue square. The offshore monitoring point was located is identified by the red circle.

STORM EVENT

Monitoring was performed over a four-day period from 26 to 30 October 2004. The storm event, which began @ 0330 on the 27th and ended @ 1145 on the 28th, produced 3.4" of rainfall. The bulk of the rainfall came during two periods; 2.1" during the first six hours of the event, and 0.7" between 1045 and 1140 on the 28th. The remaining rainfall fell during three half-hour periods, each producing about 0.2". The rainfall total was a record for the month of October and came after a five day dry period.

ONSHORE MONITORING

Onshore monitoring was conducted using an automated American Sigma 850 autosampler to collect both first-flush and composite storm water samples, and to measure rainfall and storm water flow. First-flush samples were collected during the first hour of flow, whereas composite samples were collected throughout the first 2.1" of rainfall. The samples collected onshore were analyzed for toxicity and the suite of chemicals identified earlier.

BAY MONITORING

The RV ECOS with MESC system was tied up on the quay wall just outside OF14 (Fig 3.) so that its sensors and water intake system were directly in line with the outfall pipe discharge, about 5 m away from the quay wall. The MESC sensors and water intake were placed at about 1 m depth, though the full water column to about a depth of 7m was periodically evaluated. Surface salinity, temperature, sample depth, light transmission, pH, and oil fluorescence data were collected every four seconds. Two trace metal analyzers, using anodic stripping voltammetry techniques [5], were used to measure dissolved copper and zinc about every 15 minutes. The MESC's trace metal clean Teflon® seawater pumping system was used to supply surface seawater to the bioassay flow- through system at a rate of about 10 L/min, as well as to collect discrete samples for chemical analysis before, during (4 samples), and after (3 samples) the storm event.

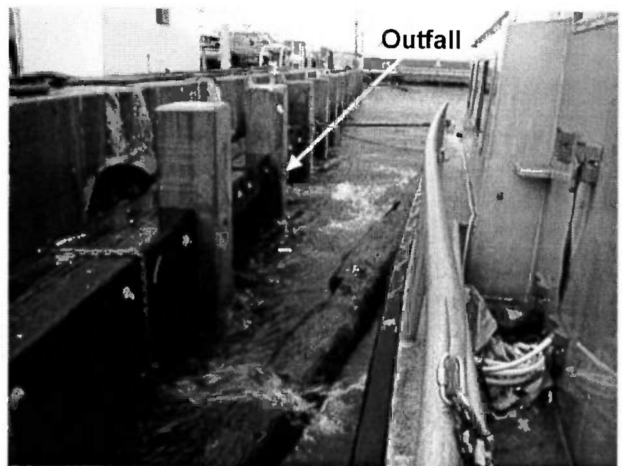


Fig. 3. RV ECOS tied up along quay wall outside OF14. The sensors and pump intake were directly in line with the outfall. Note sheet runoff over quay wall.

FLOATING BIOASSAY LABORATORY SETUP

Water Bath System. A fiberglass water bath measuring 106 cm long X 61 cm wide X 20 cm high was used to house the flow-through exposure chambers (Fig. 4). Water was pumped through a PVC grid fitted with adjustable valves to regulate water flow to individual chambers. Inside the water bath, an acrylic stand with a series of 7.5 cm diameter cutouts held the chambers in place throughout the exposure period. Seawater overflow from the exposure chambers filled the water bath to approximately 5 cm in height to help insulate against temperature shift.

Exposure Chambers. Test organisms were held in clean, seawater-leached 400 mL polyethylene containers (Fig. 4). Matching lids with cutouts were used to prevent organism ejection during boat movement, yet allow access for water flow and feeding. Both control (static) and flow-through chambers contained 250 mL of seawater at all times. Overflow ports on flow-through chambers measured approximately 2 cm and were covered with 300 µm PeCap mesh. The flow rate resulted in an average of 15 turnovers per hour. Control chambers were filled with clean, filtered, natural seawater from the research pier at Scripps Institution of Oceanography. One renewal of the control water was performed for 96-h exposures, while 48-h exposures were not renewed. Topsmelt and mysids swam freely in the chambers, while mussel embryos were contained in 5 cm diameter polycarbonate drums with 20 µm Nitex® mesh on each side, as described in [6].

TOXICITY TESTING

Toxicity testing generally followed standard methodology for assessing acute whole effluent toxicity with topsmelt larvae and mysid juveniles [7] and chronic toxicity with mussel embryos [8]. Mysid and topsmelt exposures were 96-h in duration with a survival endpoint. Mussel exposures were 48-h, with an endpoint based on the proportion of normally developed, D-shaped larvae as examined by microscope. Onshore testing included exposures with effluent representing the first-flush and composite storm water. Effluent salinity was increased to 34 ‰ with bioassay grade synthetic sea salt (Crystal Sea Marinemix) for topsmelt and mysid exposures. Salinity for the mussel exposures was adjusted with hypersaline brine made from clean natural seawater, resulting in a maximum effluent concentration of 61.4%. Onshore tests consisted of 3 replicates of 10 mysids, 4 replicates of 5 fish, or 5 replicates of 150 mussel embryos, for each treatment. At least four dilutions of effluent (0.5 dilution factor) were prepared for each species. An insufficient volume of the composite sample prevented dilutions below 100% for mysid exposures, and any exposure to topsmelt. Negative controls included clean natural seawater and synthetic salt or brine adjusted to 34 ‰ with deionized water. Copper added to natural seawater was used as a positive control, to assess the relative sensitivity of the test organisms.

Offshore testing included two treatments, one under flow-through conditions and the other a "floating" control to assess any impacts associated with being in the field. Six replicates of 10 mysids, 8 replicates of 5 topsmelt, and 6 replicates of 150 mussel embryos were used for each treatment.

All test organisms were purchased from outside vendors and acclimated for ~24 h in the laboratory prior to use. Organisms used in the floating bioassay were acclimated to expected testing temperatures in the exposure chambers over approximately 1 hr and carefully transported to the water bath system aboard the RV ECOS. All topsmelt and mysids were fed twice daily with freshly hatched *Artemia* nauplii. MESC sensors were used to monitor temperature, pH, and salinity for all flow through chambers, and a HOBO® data logger was used to monitor temperature in both static controls and the water bath. Dissolved oxygen was also monitored hourly in all chambers using a YSI oxygen meter.

Fig. 4. Flow-through bioassay setup aboard RV ECOS. Water was continuously dripped into each of the treatment beakers containing topsmelt, mysids, and mussel embryos.

CHEMISTRY

All discrete samples were analyzed for total and dissolved copper and zinc, TSS, and DOC. First-flush and composite storm water samples were also analyzed for total and dissolved aluminum, iron, chromium, manganese, nickel, arsenic, selenium, silver, cadmium, tin, lead, and mercury, 41 PAH analytes, 31 PCB congeners, and chlorinated pesticides including DDT, its metabolites and chlordane.

Chemical analyses were performed in-house and by Battelle's Ocean Sciences and Marine Sciences laboratories, in Duxbury, MA and Sequim, WA, respectively. All analyses were performed using standard NS&T low-detection methods with appropriate QA/QC controls including method blanks, blank-spikes, matrix spikes, duplicates, and standard reference materials. Storm water samples were analyzed for metals using EPA methods 1638m and 1640. Bay water samples were analyzed for metals using trace metal analysis techniques described in [9]. DOC was analyzed using EPA method 415.1. TSS analysis was performed using standard protocols developed at the University of New Hampshire [10]. Water samples analyzed for organic chemicals were extracted using EPA SW846 3510C. Extracts were analyzed for PAH using EPA method EPA SW846 Method 8270C, for PCB congeners using EPA Method 1668A and for chlorinated pesticides using EPA SW846 Methods 8081A and 8082.

RESULTS

ONSHORE

Roughly 13,000 m^3 of water was discharged through OF14 during this storm event. An additional, but unmeasured amount also discharged as sheet runoff (Fig. 3). Maximum observed flow was roughly 0.5 m^3/s.

Undiluted first-flush (FF) storm water was significantly toxic ($p < 0.05$) to mysids and to mussel larvae, but did not negatively impact topsmelt survival (Table 1). Composite (Comp) samples showed a reduced toxic effect with minimal toxicity to mysids and no toxicity to mussel larvae (topsmelt were not tested). Laboratory control survival was ~100% for both mysids and topsmelt, and normal development was 89% for mussel larvae. The positive control, a reference toxicant test with copper, indicated normal sensitivity of all species (within 2 standard deviations of the laboratory control chart mean), with LC50 values of 287, 98, and 7.6 µg/L, for mysids, topsmelt, and mussels, respectively. All water quality data were within acceptable limits.

Table 1. Summary of toxicity data. Data represent percent survival (mysid, topsmelt) or percent normal larval development (mussel). Lowest observable effect concentration (LOEC) and the concentration causing 50% mortality (LC50) or effect (EC50) are included. Dashed lines indicate no data.

Exposure Type	Sample Type	Parameter	Mysid	Topsmelt	Mussel
Onshore	FF	Neg. Control	98.3	100.0	92.6
		100% effluent	63.3	90.0	1.2
		LOEC	100.0	>100.0	50.0
		LC50 or EC50	>100.0	>100.0	49.1
	Comp	Neg. Control	98.3	-	92.6
		100% effluent	80.0	-	86.4
		LOEC	-	-	50.0
		LC50 or EC50	-	-	>61.4
Offshore	Receiving	Floating Control	93.3	70.0[a]/100.0[b]	92.2
		Flow-through	98.3	62.5[a]/89.3[b]	80.5

[a] Actual percent survival

[b] Percent survival relative to floating control

Chemistry results are shown in Table 2. All metals (except aluminum) were lower in the composite samples by about half the amounts measured in the first-flush sample. TSS, however, was higher in the composite sample as were some individual organic, though in the case of the organics the increase may have resulted from being at or near the detection limit. Aluminum, iron, silver, lead, mercury, and tin were nearly all in the particulate phase, with the remaining metals ranging between 30% and 70% dissolved phase. Copper, individual DDT isomers, total PCB (TPCB), and some of the higher molecular weight PAH concentrations were elevated above their respective water quality standards (WQS) in both first-flush and composite samples [11]. Zinc was above its WQS in the first-flush sample but below it in the composite sample. The typical elevation above a WQS was between a factor of 2 and 8 in the first-flush sample and about half that amount in the composite sample.

Table 2. Summary of chemistry data. Metals data are for dissolved fraction only. Organic summations Priority Pollutant PAH (PP PAH), Total PCB congeners (TPCB), Total DDT isomers (TDDT), and Total Chlordane isomers (TCHLOR) were calculated using ½ method detection limit for analytes measured at or below the detection limit. Dashed lines indicate no data.

		ONSHORE		OFFSHORE		
Analyte	Units	FF	COMP	BEFORE	DURING *	AFTER+
Ag	µg/L	0.00601	0.00378	-	-	-
Al	µg/L	14.7	17.7	-	-	-
As	µg/L	2.04	1.72	-	-	-
Cd	µg/L	0.492	0.244	-	-	-
Cr	µg/L	2.22	9.99	-	-	-
Cu	µg/L	18.9	9.89	3.9	4.7	3.7
Fe	µg/L	26.4	25.0	-	-	-
Hg	µg/L	0.00597	0.00330	-	-	-
Mn	µg/L	29.2	13.2	-	-	-
Ni	µg/L	3.67	1.66	-	-	-
Pb	µg/L	0.493	0.441	-	-	-
Se	µg/L	0.848	0.356	-	-	-
Sn	µg/L	0.25**	0.25**	-	-	-
Zn	µg/L	175	68.4	7.8	9	8.7
TSS	mg/L	61.2	78.7	1.4	3.8	2.2
DOC	mg/L	11.7	6.0	0.91	1.2	0.90
PP PAH	ng/L	596	387	-	-	-
TPCB	ng/L	71	30	-	-	-
TDDT	ng/L	7.5	3.6	-	-	-
TCHLOR	ng/L	2.4	1.8	-	-	-

* Average of 4 samples

+Average of 3 samples

** Value=1/2 Method detection limit

OFFSHORE

Mysid survival and mussel normal larval development were high in the floating controls, each exceeding 90% (Table 1). Mysids in the flow-through treatment experienced nearly no mortality. Mussel development in the flow-through treatment was slightly lower than the floating control, but the difference was not statistically significant. Topsmelt survival was reduced in both floating controls and flow-through treatments (Table 1). However, 86% of the topsmelt mortalities occurred in the first 24 hours. Though water quality of the offshore treatments was within a range tolerated by all species, a spike in water temperature (maximum of 26.3 °C) was measured at the beginning of the field exposure for both the floating control and flow-through treatments. This spike occurred when a sun block had not yet been put into place. Because of the low control values, the topsmelt data are also reported as percent survival relative to the controls.

The chemistry of discrete bay water samples collected before, during, and after the storm event are shown in Table 2. For simplicity, data in the table for the four "during" and three "after" samples were averaged. "During" samples were collected during the first 2.1" of rainfall when it was actively raining and there was visual storm discharge to the bay. "After" samples were collected ~ 8, 16, and 40 h after rainfall and storm flow had ceased.

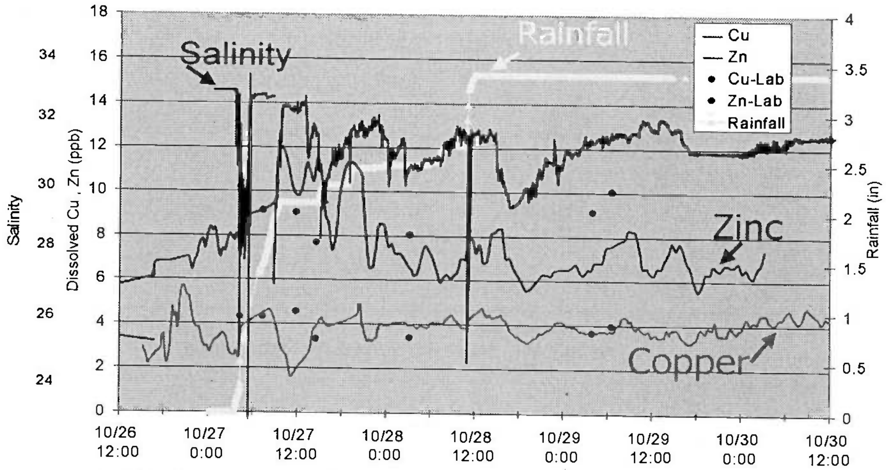


Fig. 5 MESC full-storm monitoring data for salinity, dissolved copper and zinc. Dissolved copper and zinc in discrete samples and cumulative rainfall data are also

Concentrations of dissolved copper and zinc measured in the discrete bay samples during the storm were about 20% higher than those measured in either the pre-storm or post-storm samples, and overall varied about 11% relative standard deviation (rsd). These metals were consistently 60 and 90%, respectively in the dissolved phase. Copper was always above its WQS of 3.1 µg/L, probably due to chronic hull paint leachate. Zinc was well below its WQS of 81µg/L. DOC and TSS levels measured in bay samples were more variable (42 and 59% rsd, respectively) and increased by a factor of ~3 during the storm, but decreased to pre-storm levels in the "after" samples.

REAL-TIME MONITORING

Salinity, temperature, pH, light transmission, and oil fluorescence data measured by MESC were highly variable during storm flow conditions. In particular, salinity varied from a pre-storm value of 33.5 psu to near zero during the most intense rainfall periods. However, these low salinity conditions were maintained for very short periods of time; on the order of minutes or tens of minutes (Fig. 5). Over the exposure time period, salinity averaged 32.4 psu and thus this freshwater signal of 3.5% translated into an average dilution factor of ~30. Some of the observed variations could also be attributed to tidal fluctuations, which were particularly noticeable after the storm was over (Fig. 5). Continuous copper and zinc monitoring, representing between 165 and 265 analyses, showed a slightly lower variability with a maximum change of about a factor of two. The continuous monitoring with the MESC trace metal analyzers produced comparable (but not exact) results to the analysis made in the discrete samples.

DISCUSSION

The storm event monitored was an exceptionally high rainfall event, falling in the 98th percentile for rainfall totals in the region [12]. Therefore, this storm is representative of the upper range of volume discharge to the bay from this drainage area. Though outfall chemistry data are historically quite variable, the measured levels during this storm were uniformly lower than those previously observed at this site, a result likely due to the short 5-d antecedent dry period. Even with lower event mean concentrations than observed on other occasions (by approximately a factor of five), this discharge event still represents an upper bound for contaminant mass load to the bay from this site.

Only copper and zinc were measured in the storm water samples at levels likely to cause the observed acute toxicity. Two separate TIE studies conducted at San Diego Navy facilities (as part of the Navy's overall toxicity investigation) identified both copper and zinc as the primary contributors to observed toxicity. In this study, it also appeared that both copper and zinc concentrations were predictive of toxicity. A strong negative correlation was observed between mysid survival and copper and zinc concentrations (r^2 = 0.977 and 0.966, respectively). For mysids, zinc very likely contributed to the observed toxicity in the effluent, as measured concentrations were high enough to cause lethality to this species (96-h mysid zinc LC50= 303 µg/L; [13]) while copper concentrations were not high enough (96-h mysid copper LC50= 153 µg/L; [13]). Strong relationships between mussel larval development and copper and zinc concentrations were also observed (r^2 = 0.931 and 0.882, respectively). In this case, copper and zinc both likely played a role in the observed effects based on the sensitivity of this species (48-h copper EC50= 6.43 µg/L [9]; 48-h zinc EC50=178 µg/L [14]). The absence of observed effects for any sample with topsmelt is consistent with the relatively low sensitivity of this species to the measured metal concentrations (96-h copper LC50=238 µg/L [15]; 96-h zinc LC50 = 627 µg/L [16]).

Dilution series data for storm water effluent samples resulted in EC50 and LOEC values of ~50% for mussel embryos. These values were higher than the 23% average value observed in samples collected from San Diego Navy facilities. These toxicity values should translate into a dilution factor of between 2 and 5 needed to reduce toxicity to these organisms. Because pre-storm bay water was the diluent in these tests, the dilution factors need no adjustment for bay background conditions.

Even though the magnitude of the contaminant load to the bay during this event was relatively high, the copper and zinc levels measured in the bay were insufficient to cause acute toxicity. The minimum dilution factor determined by dividing the first-flush sample concentration by the difference between the maximum bay water value during the storm and the pre-storm concentration (MESC trace metal analyzer data) was 5 and 22 for copper and zinc, respectively. The average dilution factor, determined comparably by dividing the composite sample concentration by the difference in the average values measured during and before the storm, was 15 for copper and 24 for zinc. The similar calculation using the discrete sample data, yielded an average dilution factor of 12 for copper and 57 for zinc. These dilution factors bracket the average value of 30 calculated from the salinity measurements.

The range in observed dilution factors was more than sufficient to explain the observed reduction in toxicity of bay waters. The rapid mixing that occurred immediately outside the point of discharge led to a significant reduction in both chemical concentrations and limited the exposure duration to minutes rather than the 48- or 96-h exposures used in standard bioassays. The use of standard methods, therefore, overestimates the impact of episodic ephemeral discharges like storm water. These findings support results measured to date at all San Diego Navy facilities, which show that toxicity measured at the end-of-pipe does not reflect actual toxic impacts in the receiving environment.

The use of the floating bioassay laboratory system with the unique MESC continuous monitoring capability provided a useful means to directly evaluate receiving water impacts. Though reports of marine larval fish and invertebrates as field-based bio-monitoring tools are limited [17-20], these results suggest that mysids, topsmelt larvae, and mussel embryos have good potential for use in exposures outside of the laboratory. These test organisms performed well under highly fluctuating seawater conditions. In particular, they did not appear to be impacted from the drop in salinity to near zero detected in the early stages of the storm. The relatively low (70%) floating control survival observed for topsmelt suggests this species may have a heightened sensitivity to fluctuating temperature, which spiked high at the start of the exposure period. This temperature fluctuation, however, could be better controlled in future efforts.

CONCLUSIONS

The data provided by the floating bioassay system confirms that bay exposures are very limited as a result of rapid mixing and dilution immediately outside the point of discharge. Observed exposure times were on the order of minutes rather than the 48- or 96-h exposures used in standard bioassays. Thus, using standard laboratory bioassays of storm water discharges made at the end-of-pipe potentially overestimate the acute toxic impact of storm water discharges to receiving waters by overestimating exposure times. The unique data afforded by continuous monitoring with a floating bioassay laboratory provided a detailed understanding of the interaction of storm water with bay waters in explaining the lack of receiving water toxicity.

Acknowledgments

The authors thank the Navy Pollution Abatement Ashore Prevention Technology and Demonstration/Validation Program, a part of the Naval Facilities Engineering Command (Andy Del Collo, program manager) and Commander Navy Region Southwest (Brian Gordon and Rob Chichester, environmental program managers) for their financial support. We also thank E. Arias, A. Blake, M. Brand, B. Davidson, P. Earley , R. Fransham, R. Gauthier, J. Guerrero, C. Kurtz, and C. Zacharias for their crucial help in the field, in the laboratory, and with data processing.

References

[1] 33USC 1251, United States Code, Chapter 33, Section 1251.

[2] S.H. Lieberman, C. Clavell, and D.B. Chadwick. "Techniques for real-time environmental mapping and monitoring," in: Proceedings, 16th meeting, US-Japan Marine Facilities Panel: United States-Japan Cooperative Program in Natural Resources on Marine Facilities, 1989, pp. 495-499.

[3] D.B. Chadwick and M.H. Salazar. "Integrated measurement technologies for monitoring the marine environment". In: Proceedings of Oceans 91 Meeting, 1-3 October, 1991, Oceanic Engineering Society of IEEE, 91CH3063-5, pp. 343-350, 1991.

[4] C.N. Katz and D.B. Chadwick. "Synoptic mapping and modeling of a sewage outfall plume". In: Proceedings of the Marine Technology Society, 1993, pp. 207-215.

[5] A. Zirino, S.H. Lieberman, and C. Clavell. "Measurement of Cu and Zn in San Diego Bay by automated anodic stripping voltammetry". *American Chemical Society*, vol. 12, no. 1, pp. 73-78, 1978.

[6] B.M. Phillips, B.S. Anderson, J.W. Hunt, P.A. Nicely, R.A. Kosaka, R.S. Tjeerdama, V. De Vlaming, and N. Richard. "In situ water and sediment toxicity in an agricultural watershed". *Environ. Toxiocol. Chem.*, vol 23, pp. 435-442, 2004.

[7] U.S. Environmental Protection Agency. "Methods for measuring the acute toxicity of effluents and receiving waters to freshwater and marine organisms". Fifth Edition. Office of Water. EPA-821-R-02-012. Washington, D.C., USA, 2002.

[8] U.S. Environmental Protection Agency (1995). "Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms". EPA-600-R-95-136. Washington, DC.

[9] G. Rosen, I. Rivera-Duarte, L. Kear-Padilla, and D.B. Chadwick. "Use of laboratory toxicity tests with bivalve and echinoderm embryos to evaluate the bioavailability of copper in San Diego Bay, California, USA," *Environ. Toxiocol. Chem.*, vol. 24, pp.415-422, 2005.

[10] University of New Hampshire, Jackson Estuarine Laboratory. "Standard operating procedure: water sample filtration and analysis for total suspended solids, chlorophyll and phaeopigments". SOP 1.06. 1992.

[11] U.S. Environmental Protection Agency. "Water quality standards; establishment of numeric criteria for priority toxic pollutants for the State of California", 40 CFR Part 131, 1997.

[12] Historical climate for Lindbergh Field International Airport, 1948 to 1990 from: http://www.wrh.noaa.gov/sgx/climate/san-san.htm.

[13] Cripe, G.M. (1994). Comparative acute toxicities of several pesticides and metals to *Mysidopsis bahia* and postlarval *Penaeus duorarum*. Environmental Toxicology and Chemistry: 13, 1867-1872.

[14] B.M. Phillips, B.S. Anderson, J.W. Hunt, P.A. Nicely, S.E. Palmer, F.H. Palmer. "Toxicity of metal mixtures to Purple sea urchins (*Strongylocentrotus purpuratus*) and Bay mussels (*Mytilus galloprovincialis*)." *Society of Environmental Toxicology and Chemistry*, 21st Annual Meeting, Nashville, TN. 2000.

[15] B.S. Anderson, D.P. Middaugh, J.W. Hunt, and S.L. Turpen. "Copper toxicity to sperm and larvae of topsmelt Atherinops affinis, with notes on induced spawning." *Mar. Environ. Res.*, vol. 31, pp. 17-35, 1991.

[16] G. Rosen, unpublished data.

[17] R. Beiras, E. Vazquez, J. Bellas, J.I. Lorenzo, G. Fernandez, J.C. Mach, and L. Casas.. "Sea-urchin embryo bioassay for in situ evaluation of the biological quality of coastal seawater". *Estuar. Coast. Shelf S.*, vol. 52, pp. 29-32, 2001.

[18] B.S. Anderson, B.M. Phillips, J.W. Hunt. "Marine Bioassay Project (Ninth Report): Investigations of receiving water toxicity in coastal waters". State Water Resources Control Board, California Environmental Protection Agency, December 1998, 42 pp.

[19] J.R. Clark, P.W. Borthwick, L.R. Goodman, J.M. Patrick, Jr., E.M. Lores, J.C. Moore. "Comparison of laboratory toxicity test results with responses of estuarine animals exposed to fenthion in the field". *Environ. Toxiocol. Chem.*, vol. 6, pp. 151-160, 1987.

[20] R. Comeleo. "Monitoring water column toxicity by measuring the survival, growth and reproduction of field-deployed caged mysids". Interim Report. SAIC, Environmental Research Laboratory, Narragansett, RI. 1991.

Appendix I

REVIEWER COMMENTS ON DRAFT STUDY REPORT and PROPOSED ALTERNATIVES

Technical Review Team Members:

Dr. Debra Denton, United States Environmental Protection Agency Region IX (no comments on proposed alternatives)
Mr. Shaun Halvax, Southwest Marine Shipyard (no comments provided)
Ms. Ruth Kolb, City of San Diego (no comments provided)
Ms. Eileen Maher, Port of San Diego (comments on report and proposed alternatives)
Mr. Ken Schiff, Southern California Coastal Water Research Program (no comments on proposed alternatives)
Mr. Scott Sobiech, US Fish and Wildlife Service (no comments provided)

Additional Outside Reviewers:

Dr. Allen Burton, Professor and Director of Institute of Environmental Quality, Wright State University
Dr. Robert Spies, Applied Marine Sciences, Livermore, CA.

Appendix Organization:

Comments and response to comments on the draft study report and proposed alternatives are organized alphabetically by reviewer name:

NOTE: The comments made here refer to a draft report that has been modified in producing the final report. References to specific page, table, or figure numbers may not match the final report.

Dr. Allen Burton, Professor and Director of Institute of Environmental Quality, Wright State University

BURTON STUDY COMMENTS:	RESPONSE TO COMMENTS:
a. General Comments: This four year study is the most extensive and advanced on-site stormwater runoff study that I am aware of. The study uses a state-of-the-art approach, combining multiple lines-of-evidence (LoE) (chemistry, lab and field toxicity, plume mapping, and effluent and off-shore monitoring). This study consisted of 136 discrete samples and 350 toxicity tests conducted during a wide range of meterological conditions and seasonal extremes. The LoE were combined to form weight-of-evidence (WoE) based conclusions on the degree of toxicity of stormwater and its effects on the receiving waters, the causes of the toxicity, and its sources. It is well know that stormwater runoff varies widely in quality and potential impacts even on a site specific basis, due to the myriad of interacting and fluctuating factors that affect it (e.g., frequency, duration, magnitude of precipitation events, source variability, seasonal factors affecting physical, chemical and biological characteristics). Nevertheless, this study effectively characterized the bounds of that variability to such a degree that the statistical confidence of the key parameters and study factors is known. This degree of precision and confidence in the data allows for conclusions to be drawn from the study that have a low degree of uncertainty. The principal conclusions and the uncertainties that I find apparent are as follows:	No response necessary.
i. Relatively undiluted stormwater samples from the study area vary in toxicity; and are more likely to be toxic when measured by standard laboratory exposures (48 to 96 hr constant exposure to first flush waters).	No response necessary.
ii. Corresponding samples that are composites (collected throughout the event) are likely to be non-toxic in similar laboratory exposures.	No response necessary.
iii. Bay waters are likely to be non-toxic during a storm event.	No response necessary.
iv. The statistical power (ability to detect differences) of the laboratory toxicity assays was very good, indicating low replicate variability.	No response necessary.
v. More realistic, *in situ* exposures during a storm event showed no toxicity.	No response necessary.

BURTON STUDY COMMENTS:	RESPONSE TO COMMENTS:
vi. Plume mapping showed freshwater runoff stayed nearshore and on the surface of the bay prior to mixing. This suggests that benthic organisms are not affected by the runoff events, even near the outfalls. Note, the most sensitive toxicity test species is a benthic organism.	No response necessary.
vii. Copper and zinc appear to be the toxicants of concern, based on the chemical data and toxicity identification evaluation (TIE) studies. However, the potential for toxicity from new-age pesticides (e.g., diazinon and pyrethroids), and from photo-induced toxicity from polycyclic aromatic hydrocarbons (PAHs) were not assessed. These compounds are common and known to be a source of toxicity in runoff. The low level discharge of PCBs is worrisome, given their propensity for biomagnifications and that there should not be on-going sources. The information from the TIE studies may contribute to source control, but given the complexity of the site and the prevalence of these chemicals (Cu, Zn, PAHs) on impervious surfaces of industrial and urban areas, it will be a challenge to reduce their occurrence.	The focus of this study was to evaluate the potential for toxicity of receiving waters from storm water discharges. The study used a list of CoCs that were thought to be the most likely causative agents generated at these types of sites. One purpose of performing toxicity tests is to evaluate the overall potential impact of all constituents whether or not they were measured in the sample. A lack of toxic response in bay waters would suggest that if these compounds were present in sufficient amounts to cause toxicity that this would have been observed.
viii. The laboratory based exposures using first flush samples do not provide useful information to meet the goals of the Clean Water Act. Resident organisms in the Bay do not remain in the plume for 48 to 96 hrs and the first flush water does not remain in the plume for 48 to 96 hrs. Composite samples would provide a more realistic relationship to receiving water conditions. However, *in situ* exposures of toxicity test organisms provide the greatest degree of realism (thus least degree of uncertainty).	No response necessary.

BURTON STUDY COMMENTS (cont.):	BURTON STUDY COMMENTS:
ix. The continued discharge of low levels of chemicals during runoff events may result in elevated sediment concentrations in the outfall area. These sediments should be assessed for their impact on benthic organisms.	The focus of this study was to evaluate the potential for toxicity of receiving waters from storm water discharges. The best way to evaluate long-term impacts to sediments is within current TMDL programs that are used to evaluate the magnitude and extent of impaired sediments using a weight of evidence approach and identify sources of the impairment. These programs are currently underway at several locations in San Diego Bay including at Navy facilities. Additional programs that can provide a better evaluation of long-term impacts to sediments include the Bay Protection and Toxic Cleanup Program, the Bight'98 program organized by the Southern California Coastal Water Research Program, and the Port of San Diego/Regional Water Quality Control Board Baywide Monitoring program.
x. The USEPA is adopting the Biotic Ligand Model for improving the water quality criteria for Cu. Given the role of DOC at the site, the BLM approach may provide useful information.	We agree that the BLM may provide useful information once it is adopted.
xi. Metal toxicity determinations are confounded by the influence of salinity and rapidly changing salinity during a runoff event. This has been the subject of previous studies and should be investigated.	We agree that metal toxicity may be confounded. Though salinity may vary in the receiving environment, toxicity test protocols require fixed salinity conditions.

BURTON STUDY SPECIFIC COMMENTS:	RESPONSE TO COMMENTS:
b. Specific Comments	
i. The Executive Summary and Discussion/Conclusion sections should be modified, incorporating and better emphasizing some of the issues raised above. These issues include the reliable WoE approach, the lack of reality of the laboratory toxicity test exposures, and the findings of the field exposure test. At present, the field exposure test is virtually not mentioned in any section except Methods. The likely rapidly fluctuating concentration of Cu and Zn in the plume should be discussed in relation to WQS exceedances. Finally, the spatial and temporal limitation of resident organism exposures due to the shallow surface water, nearshore stormwater plume should be emphasized. The following references provide a wealth of information and cited peer reviewed studies that support the points I am making. You may want to include some of these citations and content for additional justification of permit conditions: Burton, G.A., Jr., and R. Pitt. 2001. *Stormwater Effects Handbook: A Tool Box for Watershed Managers, Scientists and Engineers*. CRC/Lewis Publishers, Boca Raton, FL, 924 pp., available online from USEPA at: (http://www.epa.gov/ednnrmrl/publish/book/handbook/index.htm); Burton, G.A., Jr., R. Pitt, and S. Clark. 2000. *The role of traditional and novel toxicity test methods in assessing stormwater and sediment contamination.* CRC Critical Reviews in Environmental Science & Technology 30: 413-447 (pdf attached); and, Burton GA Jr., Greenberg MS, Rowland CD, Irvine CA, Lavoie DR, Brooker JA, Eggert LM, Raymer DFN, McWilliam RA. 2005. *In situ exposures using caged organisms: a multi-compartment approach to detect aquatic toxicity and bioaccumulation.* Environ. Pollut.134:133-144 (pdf attached).	The comments provided above have led us to focus the report findings on the major goal of evaluating the efficacy of using WET testing and its use in "assessing and protecting against impacts upon water quality and designated uses caused by the aggregate toxic effects of the discharge of pollutants" (EPA, 1991). The comments above suggest that our conclusions were a bit broader than the data support and that the conclusions need to be focused on toxicity testing and what it shows. These comments will therefore serve to improve the report. We have reviewed the cited literature (some prior to the reviewed draft) and intend to re-evaluate where their citation can be used to support discussions and conclusions in the report.
ii. Study Goals: While a goal was to "evaluate toxic impacts on the receiving environment", this was done in a limited manner – using only 3 standard test species. A survey of resident benthic organisms along a gradient of exposures would be a useful LoE for future studies.	The study focused on the WET test requirement in the permit which is used in "assessing and protecting against impacts upon water quality and designated uses caused by the aggregate toxic effects of the discharge of pollutants" (EPA, 1991). The study fully evaluated the impacts that can be determined by WET testing. A survey of benthic organisms is best done within a full sediment investigation such as those being done at Navy facilities under the TMDL program to evaluate magnitude and extent of impairment using a weight-of-evidence approach.

BURTON STUDY SPECIFIC COMMENTS:	**RESPONSE TO COMMENTS:**
iii. The field bioassay study by Katz and Rosen (2005) should be presented in greater detail in the Results and Discussion sections.	Additional text was added.
iv. Monitoring sites: Greater than 90% impervious surface area is amazingly high. Consider adding background information from Burton and Pitt (2001) concerning typical contaminants and loadings from similar sites.	Additional text was added where applicable.
v. Spell check needed throughout – occasional misspells.	A spell check was completed after all report modifications to text.
vi. Table 5. Was organism age not a QA/QC objective?	Organism age is a QA/QC objective. Table 5 was altered to include the
vii. Table 6. Units missing.	Units were added.
viii. Recheck all tables and figures to ensure they are stand alone. Some are missing explanations of acronyms/abbreviations.	Additional text added where applicable.
ix. Some levels of PAHs observed in first flush samples have been associated with photo-induced toxicity. This is not addressed in this study, nor could it be measured in typical laboratory exposures with fluorescent lighting.	The issue of photo-activated PAH was not addressed in this study. A comparison of toxic thresholds on an analyte by analyte basis showed that there were a six instances when either fluoranthene or pyrene concentrations exceeded an acute effect threshold (including photoactivation) found in the literature. None of the bay PAHs was above a chronic toxic effect level (also including photoactivation). Additional data found in Scannell et. al., 2005 was evaluated for the report and appropriate text added.
x. Diazinon has been identified as a problem chemical in southern California, but apparently was not evaluated. It is toxic at the ppt level and has been found in rainfall.	Diazinon was not identified as a likely contaminant of concern at the start of the study as the Navy has not allowed its use for several years. The lack of toxicity in the receiving waters suggests that even if diazinon was present in the storm water, it did not lead to toxic effects in the receiving water.
xi. The summation of the pesticides found may be a problem, and some produce synergistic effects in combination (see several papers by Mike Lydy). The pyrethroids are also a problem in California. Note recent studies by Don Weston.	The lack of toxicity in the receiving waters suggests that the amounts present in storm water were insufficient to cause a toxic effect in the receiving water. The value of toxicity testing it that it allows an evaluation of toxic components even if they are not measured independently.
xii. P. 103, paragraph 2 (remove future tense).	Text adjusted.

BURTON STUDY SPECIFIC COMMENTS:	RESPONSE TO COMMENTS:
xiii. The correlations of Cu and Zn with toxicity were weak. Despite their sporadic short term exceedance of WQS, the salinity, DOC and fluctuations in concentration make the causality issue uncertain. Nevertheless, they are often identified in stormwaters as compounds of concern.	No response necessary.

BURTON ALTERNATIVES COMMENTS:	RESPONSE TO COMMENTS:
a. The “Alternative” document was not reviewed until the above “Storm Study” report review was completed. It is good to see the authors of the “Alternative” document have identified the key issues and problems associated with the existing permit language.	No comment necessary.
b. The existing permit toxicity limits are in no way based on scientific evidence that shows a relationship to protecting receiving waters. Such limits also ignore current USEPA draft guidance and a wealth of scientific evidence provided by a range of stakeholders that documents the limitations of WET testing.	No comment necessary.
c. Realistic MSD levels must be incorporated into any permit that utilizes toxicity testing. Given the huge law suit challenging previous WET guidance, the years of effort put into developing the draft guidance, and the extensive documentation which is publicly available, it is amazing that the proposed permit language was used.	No comment necessary.
d. The Ohio EPA has allowed stormwater permittees to focus on receiving water impacts, rather than end-of-pipe limits, given the complex exposure issues that preclude use of conventional WET approaches. This approach is very reasonable and should be considered here.	No comment necessary.
e. The goals of the Clean Water Act (CWA), and issues of anti-degradation and anti-backsliding, are not at risk, when receiving waters are protected. Receiving water protection cannot be extrapolated from laboratory toxicity testing under erroneous exposure scenarios or from chemical data that do not consider fluctuation concentration, complexation and bioavailability. The study site has very complicated exposure and complexation issues, due to tidal mixing, freshwater-saltwater density differences, and salinity-DOC complexation phenomena.	No comment necessary.

BURTON ALTERNATIVES COMMENTS (cont.):	RESPONSE TO COMMENTS:
f. The suggested permit language used in Alternative 1 and 2 is a huge improvement over the current language, since a more realistic dilution is introduced and laboratory exposures times are reduced. The use of realistic MSD levels is appropriate and should dictate acceptability/exceedance criteria. However, the proposed testing procedures are still overly conservative, for the following reasons: i. The plume mapping data shows resident organisms are exposed for up to tens of minutes, IF they were to remain in the plume (a conservative assumption). An exposure of 48 hrs introduces a safety factor greatly exceeding two-fold. A top-to-bottom depth integrated sample would better characterize receiving waters adjacent to the outfall. ii. The resident organisms are marine, and, therefore, will not remain in the freshwater plume during a runoff event. Their exposure period will likely be on the order of a few minutes, maximum. Benthic organisms will have no exposure to unmixed effluent. This reality must be considered if the NPDES permit (and CWA goal) is trying to protect the receiving waters and their biota.	We agree with the comments regarding exposure magnitude and duration. Using conservative exposure conditions will provide an additional level of protection in the permit. We agree that benthic organisms will not be exposed to unmixed storm water.
g. The use of *in situ* exposures would provide realistic assessments of toxicity.	We agree, but *in situ* methods are still in a developmental stage.
h. The use of benthic colonization or transplant studies below the outfalls would provide a realistic assessment of resident organism effects. i. Sampling of benthic macroinvertebrates on a gradient from the outfalls would also provide an assessment of resident organism effects.	The scope of the effort was to evaluate potential toxic effects to receiving waters. Investigations into benthic impairments are best left a part of ongoing TMDL and Baywide monitoring programs where the data are evaluated using multiple lines of evidence including an evaluation of sources.
j. These approaches could be incorporated into permit language for assessment on a yearly basis. These types of tests would better assess whether or not the goals of the CWA are being met, than artificial exposures in the laboratory with single species.	No comment necessary.

Dr. Debra Denton, USEPA Region 9

DENTON STUDY COMMENTS:	RESPONSE TO COMMENTS:
I complement the Navy on undertaking this study to determine the spatial and temporal potential for toxicity at the various Navy sites. For those of us, myself included who have implemented a storm water program, understand and appreciate the complexity of addressing sampling and testing logistics involved with storm water testing. Overall, the Navy has done an extensive job of collecting and analyzing storm water for toxicity assessment including TIEs over multiple storm events and years.	No response necessary.
I submitted comments on the study's design please refer to my letter dated December 8, 2003. The report does not address the following suggestions/recommendations as described in my letter. I will restate my specific recommendations from the December letter and highlight issues and comments.	Responses below:
1. "The Navy should prepare a Quality Assurance Project Plan (QAPP). All methods will utilize standard USEPA procedures and follow the project's approved Standard Operating Procedures (SOPs) and QAPP (see USEPA 2001). The QAPP will be consistent with the SWRCB's Surface Water Ambient Monitoring Program (SWAMP) program requirements. All data will be subjected to a 100% audit by the project QA Officer. Any deviations in SOPs for sample analysis or reporting will be recorded and corrective actions will be implemented according to the QAPP." The section on materials and methods should have clearly discussed and cited the QAPP and how it was followed. This must be added in the final report, an appendix would be appropriate.	This investigation started under a Navy research program to evaluate storm water methods and impacts at Navy facilities without a requirement to develop a QAPP. The project was expanded to evaluate toxicity after the SDRWQCB requested the study. Initial discussions with Regional board staff did not identify a QAPP as a requirement for the study. The recommendation that a QAPP be produced occurred well after initiation of the project, which is contrary to guidance that requires a QAPP be produced and accepted prior to that start of a study. However, the investigation followed the principals and elements that are included in formal QAPP documents. Where possible, we have added text within the appropriate report sections to improve the level of detail that support the elements that are included in a formal QAPP.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
2. "The test methods and test species to be tested must be from the following methods: - 1st edition west coast marine short-term test methods (USEPA 1995) - 4th edition freshwater short-term test methods (USEPA 2002a) - 5th edition freshwater and marine acute test methods (USEPA 2002b). - For both acute (include invertebrate and vertebrates) and chronic test methods (include invertebrate, vertebrate, and plants) multiple test species must be evaluated for both study objectives." Most importantly, I specifically delineated that USEPA toxicity test procedures were to be followed. Therefore, why did the Navy choose to follow the ASTM 1994 bivalve development protocol instead of the USEPA 1995 bivalve development protocol? There are several in discrepancies as to which WET test method was conducted for this species. For example, the report cites ASTM 1994, Appendix F cites ASTM 1999, and Appendix H cites USEPA 1995. It is unclear as to what method and whether the method was consistent for all these analyses? I suggest that in the final report that any differences between these methods be described. The report needs to specifically identify the mysid tested in this section (reader should not have to go to Appendix B or Executive Summary). Why were only three reps employed instead of the required four reps (a required minimum as specified in the manual)?	The permit requires testing survival of either a vertebrate or an invertebrate. The mussel development endpoint was added because it is among the most sensitive short-term tests available. The level of effort required to perform the toxicity tests for each storm event with multiple species, locations, and sample types was immense. Adding another test species would have surpassed the capabilities of most toxicity labs. EPA methods were used for the topsmelt and mysid testing specified in the permit. ASTM 1999 was used for the mussel tests and should have been uniformly cited in the documentation. ASTM methods are nearly identical to USEPA 1995 though they differ in their test acceptability criteria. ASTM protocols were referenced because results were based on normal shell development and not survival. This was considered appropriate because of the sensitivity of the sublethal endpoint, and the speed with which results could be evaluated, particularly because of large testing requirements. References to other than ASTM 1999 were corrected in the report and appendices. Text describing the specific test species used was added to the text. The test methods require two replicates for effluent testing and four replicates for receiving water testing. Three replicates were used for all tests to streamline laboratory work and to provide consistency between effluent and bay water testing. Though this was a departure for the number of replicates required for receiving water testing, test method variability was good. All receiving water test results were above 90% survival and a fourth replicate would not have changed this outcome.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
3. "The proposed technical approach must provide the basis for how the data will be evaluated for data analysis and evaluation steps. For example, how will the data be evaluated for whether the quality assurance/quality control (QA/QC), and test acceptability criteria (TAC) requirements were achieved? The QAPP should include data analysis and evaluation procedures. For each test endpoint and method, the results for various endpoints must be calculated according to EPA flowcharts in the WET manuals. The results must be examined on a test-by-test basis. Test results should be reviewed to ensure that (1) data were properly reported and (2) proper estimates were generated according to EPA flowcharts (see Report Preparation and Test Review section of the manuals). Note, all these steps should be specified in a document in advance of the data analysis to ensure that all data meets the appropriate and required QA/QC, TAC and statistical assumptions before the data is included in the overall data evaluation." The discussion on toxicity data quality assurance/quality control starting on page 28, needs to be refined and more detail provided (e.g., the report has a better description of methods and materials provided in the TIE test method discussion). For example, the sentence "Exceedances in several data quality objective did not automatically invalidate a test" so how was this documented in the QAPP? Also, according to the manuals, the test acceptability criterion (TAC) (90% or greater survival and the controls for acute test methods) is mandatory. Please specify whether all tests achieved the TAC? Where is the discussion of the QAPP data analysis and evaluation procedures (include as an appendix)?	As mentioned previously, a formal QAPP was not generated. Table 5 did show the QA objectives and the test acceptability requirements. As described above, text was added to the methods and results sections that better describes and quantifies the evaluation of performance relative to QA/QC objectives and TAC.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
4. For the evaluating the acute toxicity objective, the study must determine the following statistical analysis for each test method, (1) standard t-test, (2) no-observed adverse effect concentration (NOAEC), and (3) the Lethal Concentration (LC50) accordingly to USEPA 2002. Each of these endpoints should be considered for this study objective. In Appendix B it appears that the analysis of the standard t-test was not conducted. Apparently, Appendix C does provide the statistical t-test analysis. However, a major oversight of Appendix C is that the control response data is not included. I suggest that these statistical endpoints be in one table for ease of comparison for the reader.	The endpoints cited in the comment were all conducted/ determined and reported in the appendices. Pre-storm bay water results act as the controls for the outfall samples. Because of the large amount of data and to minimize redundancy, a choice was made to show the pre-storm water bay results (which acted as controls) with the other bay water data. Appendix B was designed to be a summary only. It does have a column showing the endpoint value calculated relative to control, providing the reader with the ability to determine the control result. Text was added to the report and the appendices to aid the reader in the organization of the tables and to identify applicable control result.
5. The stormwater sample should be collected with a concurrent flow measurement and ability to relate the sample collection timing to the relationship on the storm hydrograph or a hyetograph as may be appropriate. The report does not include the concurrent flow measurement volume and relationship of the storm hydrograph. As the report states on page 2, "the composite samples were collected to provide sample data that was representative of the entire storm discharge as that could be used in mass loadings calculations in future TMDLs." In order to do mass loading calculations flow measurements must be conducted at sample time.	The permit requires that first-flush samples be collected during the first hour of flow, with no measure of flow. Flow measurements were made as part of this study to allow the generation of flow-weighted composite samples (See Section: Onshore Sampling-Composite). Mass loading calculations are not pertinent to evaluating the toxicity threshold and were not included in the report. These data can be used in evaluating TMDLs associated with the watershed at a later date.
I have reviewed the December 2005 draft final report to evaluate Navy facility storm water toxicity and have the following comments. I will not be providing comments on the modeling and chemistry analysis is this is not my direct expertise. I do have extensive comments on the interpretation of PMSD, toxicity tests, and TIE analysis. Considering the amount of information in the report and appendices and my workload, I am sure that I would find additional thoughts and comments on this report.	No response necessary.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
Analytical analysis discussion: • I question why a very limited list of contaminants of concern such as metals, PAH, PCB, and chlorinated pesticides including DDT were only initially targeted. Considering the wide range of land uses within these four stations, such as fuel storage, painting, sandblasting, vehicle repair and maintenance, why were not additional compounds tested (e.g., oil and grease, surfactants, pesticides such as organophosphates and/or pyrethroids etc?). What pesticides are used at these Navy stations? I suggest inserting the water quality criteria for both diazinon and chloryprifos to Table 40. Diazinon CCC = 0.8185 ug/L and CMC = 0.81850 ug/L (USEPA 2005). Chlorpyrifos CCC = 0.0056 ug/L and CMC = 0.011 ug/L (USEPA 1986).	The list of CoCs cited in the comment was originally identified for Naval Station based on their historical presence in sediments within the central pier area of Naval Station (Fairey et al., 1996; Chadwick et al., 1998) as well as from historical NPDES storm water data collected from 1994 through 1999. The sediments in this area are listed by the SDRWQCB for a TMDL investigation. The list of CoCs was vetted with the Technical Review Team (TRT) and matched up with the activities cited in the comment. Though samples were not specifically analyzed for the pesticides noted in the comment, they would have been identified during the TIE analyses if they were present at levels sufficient to cause toxicity. The lack of toxicity in receiving waters suggests that even if these pesticides were present in the storm water samples and not analyzed for TIEs, they did not lead to toxic effects in the receiving water.
• Table 22 needs to include the NPDES performance goals for all the metals in addition to copper and zinc for comparison purposes.	The permit only has performance goals for copper and zinc.
• In the Executive Summary it highlights that that the "toxicity of undiluted first flush stormwater was highly variable, spanning the full range of impact, 0 to 100%, and average 72% for both topsmelt and mysid test species". Coincidentally on page 48, the report states that "the data show considerable variability of the individual metals spanning a range of approximately 25% to 180% for both dissolved and total metal. Variability was typically about the same or lower in the composite samples than in first-flush samples." The Executive Summary text needs to be clear that the variability of toxicity response is related to the fluctuating chemicals and concentrations (in particular metals) as to be expected. This point illustrates that is paramount to collect that first-flush stormwater sample to capture the magnitude of toxicity.	Text was added to clarify that variable toxicity was related to variable concentrations of contaminants. First flush storm water discharge samples only capture the magnitude of toxicity of storm water for the moment in time when the sample is taken. First-flush samples do not capture the toxicity of the discharge, nor do they capture the magnitude of toxicity that occurs in the receiving environment as a result of the discharge.
PMSD discussion: • The tables showing PMSD (Table 12, 20, 27, 34) should show the number of tests included in the data set. For a direct comparison to USEPA 2000, I strongly suggest that the PMSD values for the 10th and 90th percentiles be provided.	Text was added to show the number of tests and 10th and 90th percentiles of the data.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
• The sentence "the lower and upper PMSD bounds are recommended as test acceptability criteria" is inaccurate. The PMSD bounds are not applied as test acceptability criteria (TAC), however, are to be reviewed according to Section 12 (Report Preparation and Test Review) of USEPA 2002a,b.	The text was modified.
• The definition of PMSD should be stated as "the minimum significant difference (MSD) is the smallest difference between the control and another test treatment that can be determined as statistically significant in a given test, and the PMSD is the MSD represented as a percentage of the control response."	The text was modified.
• In looking at the figures 55 - 57, I question how the EPA values were plotted? I would assume to plot the EPA values this report's author would need the individual data points (not available to my knowledge), from the EPA 2000 document.	Tables (B-8a through B-8c) in the EPA document (EPA, 2000) identify the 5, 10, 15, 20, 25, 50, 75, 80, 85, 90, and 95th percentile PMSDs. These are the data plotted in the figures. Individual test data were not available.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
• The sentence "The 90% requirement in the permit language exceeds the EPA's lower boundary for PMSD and has no statistical power to identify actual toxic effects". This sentence is wholly inaccurate and confounding two separate and independent issues into one sentence. "Power can be characterized only by repeated testing. Power is an attribute not of a single test, but of the sequence of many tests conducted under similar conditions and with the same test design (USEPA 2000). Power is the probability of correctly detecting a true toxic effect (i.e., declaring an effluent toxic when in fact it is toxic). The sensitivity of the toxicity test will depend in part on the number of replicates per experimental units per treatment, the alpha and beta (provided beta is used to determine the effect size desired), and the variability (e.g., MSE). The power to detect differences increases (e.g., MSD decreases) as the variability decreases and the effect size increases. The MSD provides an indication of within test their ability, and smaller values of MSD are associated with increased power to detect a toxic effect. EPA recommends upper and lower PMSD bounds for each test method in order to minimize within test variability and increase statistical power (Denton et al., 2003). The expression of the permit language of toxicity shall not produce less than 90 percent survival 50% of the time has nothing to do with either test power or PMSD bounds, as this is a regulatory decision as to the interpretation of the Clean Water Act's interpretation of "no toxics in toxic amounts."	The text was attempting to describe that the 90% requirement in the permit has no statistical basis for identifying a toxicity test result as toxic. Depending on control response the arbitrary 90% cutoff falls at or below the lower bound 10th percentile PMSD indicating that only 10% of labs could declare a result toxic. Additional text and tables were added to the report to clarify these points. The basis of the Board's decision to include a study of storm water toxicity was to develop a science-based toxicity threshold for industrial storm water discharges. The choice of toxicity metric must therefore be scientifically defensible and not based on an arbitrary decision or interpretation. Because the Navy's proposed standards are based on scientific measurements, data, and statistical compilation, PMSD was included in developing the Navy's proposed standards. It should be noted that the Water Quality Control Plan for the San Diego Basin states: "The survival of aquatic life in surface water subjected to a waste discharge or other controllable water quality factors, shall not be less than that for the same water body in areas unaffected by the waste discharge...". It is the Navy's position that the lack of toxicity measured in receiving waters meets the narrative requirement of "no toxics in toxic amounts".

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
• The sentence "even using the PMSD upper bound, the EPA found only 50% of the labs were able to detect a 25% difference from control" should be restated. See USEPA 2000 page 5-7, "EPA found that about half of the labs in the days that were able routinely to detect a 25% difference between control and treatment." I suggest that inserting this sentence solely from the EPA document without context is inappropriate. Labs not achieving routinely the PMSD (i.e. sufficient within-test sensitivity) must reduce the PMSD by decreasing within-test variability, increase control mean and/or increase number of replicates within a test. The point is that labs should be achieving PMSD lower than the 90th percentile in order to not exceed this upper bound.	The text was altered to match the statement "EPA found that about half of the labs in the data set were able routinely to detect a 25% difference between control and treatment."
• I do agree that the PMSD values from this study are reasonable and it was good to examine with the additional test results using the Natilus data set for comparison purposes.	No response necessary.
TIE discussion: • TIE Appendix E cites the use of the ASTM 1993 mussel development test and appendix F cites the use of ASTM 1999 of which neither is the method cited and used in the routine toxicity tests cited in the main report, ASTM 1994. So what test method was conducted, and if different versions (what are the differences)?	As described above, the correct citation for the larval bivalve tests should have been ASTM (1999) and was changed in the text.
• The acronyms mean to be consistent with the body of the main text (e.g., NAVSTA, SUBASE, NASNI).	The acronyms used by the TIE laboratory did not match those used in the final report. These reports were delivered as *.pdf files and we are not able to modify them to match. Instead, an introductory paragraph was added to the Appendices E and F describing the relationships between acronym usage. NAV=NAVSTA, SUBASE=SUB and NI=NASNI.
• On page 3, did the researchers follow the Marine Phase I TIE procedures? If so reference is not included (USEPA 1996). This document details the procedures for the marine test species.	Marine Phase I TIE procedures following methods outlined in EPA/600/R-96/054 were employed throughout both study years. This citation was inadvertently left out of the reports and was inserted as appropriate.
• As stated on page 48, why were only 11 of the 16 outfall samples analyzed for copper and zinc? According to EPA Phase II of the TIE procedures, EDTA also chelates the divalent metals cadmium, magnesium, lead, and nickel. Therefore, complete metal scans should have been conducted for each of these outfall samples.	A metals screen was conducted for all TIE samples. The statement on p. 48 refers to the fact that some outfall samples (not the TIEs) were analyzed only for copper and zinc.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
• A major limitation of the TIE analysis (Appendix E) is that only two manipulations were conducted with the samples, the addition of EDTA and C18 column extraction. It appears that the researchers went in thinking only metals were the toxicants and not wishing to ascertain the total toxicity of the samples? For example, the pH adjustment tests would have provided more information on the nature of pH dependent toxicants. In addition conducting the thiosulfate addition would elicit further which metals are complexed with only one or both additives (EDTA and thiosulfate).	The TIE laboratory was requested to identify the causative agents in storm water samples to three test endpoints. There was no direction given to look for a particular category of toxicants. The approach taken was consistent with the USEPA Marine Phase I TIE manual that states that the number of treatments is only a recommendation and may require modification depending on each application. The Phase I treatments selected were sufficient for a Phase 1 Characterization since they eliminated all toxicity associated with the samples. The choice of Phase 1 treatments reflected prior knowledge of toxicants likely to be present as well as limitations on sample volume. Application of pH-adjustments would have been of limited benefit because of difficulties in controlling pH in highly buffered seawater, as well as intrinsic toxic effects of pH on bivalve larval development. Sodium thiosulfate may have been of some benefit in terms of distinguishing between different metals, but analytical data were available for the metals, and it was more cost-effective to proceed directly to spiking studies to identify specific metals responsible for toxicity.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
• The study did not conduct any Phase III analyses. Unless the researchers consider figures 5 - 6 and the metals additivity study to demonstrate the relationship of TUs to copper and zinc? However, the reported r2 values are weak. In addition, the figure for both copper and zinc, the figures intercept is ~ 1.8 TUs meaning that there is still toxicity unaccounted for in the sample. On page 31, it states for Subbase of 23 c+e that EDTA clearly removed toxicity, however sufficient copper was not present to account for all the toxicity present and there was barely enough to zinc account for all of the toxicity.	Phase III TIE analyses were conducted and included: 1) copper and zinc toxicity studies; 2) studies with mixtures of copper and zinc; 3) comparison of sample metal concentrations with available literature values; 4) statistical comparisons of predicted and actual TUs present in the samples; and 5) comparisons of species sensitivity. The TIE reported that r2 values for copper and zinc alone were weak, indicating that neither metal by itself could account for toxicity observed across all of the samples. However, the relationship for a combination of copper and zinc was strong with an r2 of 0.8 and a p-value of 0.02. The text in the report stated that "…both metals contributed to toxicity…". The accompanying Figure 11 also shows this conclusion. Though Figure 5c shows a relationship with an intercept of 1.8 TUs, all of the data points show that there is more than enough copper and zinc present to account for toxicity in the sample. It is inappropriate to extrapolate results to the intercept because all of the samples exhibited toxicity (above 3 TU) and had sufficient residual copper and zinc to explain toxicity. This is further complicated by the fact that some portion of the metals is not bioavailable.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
• In addition, there is mysid toxicity which was not addressed. I suggest a copper and zinc spiking study to ascertain whether these two toxicants are the cause of toxicity to the mysids? This is important because mysid toxicity was demonstrated with several the samples (see page 28) however only one TIE was conducted?	Mysid toxicity was clearly addressed in this study. However, sample SUBASE OF23c+e was the only sample collected in February 2004 that exhibited a sufficiently strong reduction in survival (approximately 43%) to justify performing a TIE. Toxicity was completely removed by the EDTA treatment, indicating divalent cationic metals were the cause of toxicity in this sample. Other samples collected in February exhibited marginal reductions in survival, the greatest of which was only 18% less than the control, which would be problematic in terms of detecting differences among treatments. Conversely, three samples collected in March 2005 were subjected to TIEs with mysids. While metal spiking studies were not performed with mysids in conjunction with the February samples, toxicity tests on copper and zinc were conducted in support of TIEs performed with mysids on the March 2005 samples, and the results and conclusions presented in the associated report.
• I question why the initial toxicity tests were conducted on February 19, 2004 and the TIE manipulations were not conducted until February 27? As stated in the report, toxicity dissipated during this time, therefore some level was lost in the samples. Given that the mussel development was 0% development for the February 18 sample (see page B-5) therefore, it is reasonable and doable to have initiated a TIE as soon as February 21-22.	The reason was explained on Page 3 of the TIE report. The TIEs were not initiated until February 27 because the lab was awaiting results of the 7-day exposures with mysids and silversides. This test design was intended to ensure that a toxic outcome could be evaluated even if acute toxicity was not identified. When toxicity was identified to the mussel, a decision was made to focus TIEs on the clearly toxic result. The design was modified in the second round of TIEs to evaluate acute exposures only.
• On page 28, was the fraction 95% methanol fraction analyzed with GC/MS?	Yes. GC/MS was performed on the 90, 95,and 100% methanol fractions. The results of this analysis are included in the addendum report prepared on January 12, 2005.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
• Appendix F, Table 4 should note that cadmium and lead are above or near the EPA criterion continuous concentration (CCC) (EPA 2002c). As indicated on page 45 the relationship between topsmelt and mysid survival is reasonably strong with a $r2$ value 0.76 and both copper and zinc for topsmelt $r2 = 0.23$ and 0.44 respectively were relatively weak. The researchers speculate that surfactants may be the primary cause to topsmelt. However, additional Phase II and III TIE procedures are necessary to confirm this hypothesis. Therefore, the Navy must conduct these additional Phase II and III TIEs with additional toxic samples to topsmelt and mysid test species (see Table 10).	Lead and cadmium were present in some of the samples, but were not related to the observed responses. No relationships were calculated between mysid and topsmelt survival; the $r2$ value (0.76) in the second sentence above refers to the relationship between zinc and mysid survival and, compared with the substantially lower $r2$ values obtained between metals and topsmelt survival, suggests that metals were not responsible for adverse effects observed with topsmelt. Surfactant toxicity was implicated for both mysids and topsmelt in one of the samples for which EDTA did not remove toxicity, but C18 extraction and aeration did. Toxicity of the foam fraction collected after aeration was also demonstrated with mysids, and toxicity was strongly correlated with MBAS ($r2 = 0.9$), a measure of anionic surfactants. While this evidence supporting surfactant toxicity is relatively strong, toxicity in the sample dissipated before the investigation could be fully completed. Though the TIE laboratory recommended that additional steps be taken to confirm the findings, there was insufficient sample and toxicity to perform further work.
• Check math errors in Tables 8 and 9.	The numbers were checked and corrected.
Discussion section of report: • The discussion on page 105 text does not match the referenced figure 52. According to the figure, first flush failed 70% of the time for the 90% survival and composites failed ~ 42% of the time. Also, first flush failed ~28% of the time for the 70% survival and composites failed ~ 4% of the time.	Both the figure and the text are correct. A failure of a threshold is the point that falls to the left of the "threshold" line on the graph because the thresholds are for greater than or equal to criteria. For first-flush, the cumulative percentage is 58%. Tables were inserted to clarify these data.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
• Throughout this discussion, it incorrectly refers to the amount of dilution necessary to remove toxicity. The mechanism to remove toxicity is to control and/or treat the toxicant(s) not by dilution.	WET testing of effluents specifically requires that dilution series be run to evaluate the dilution level at which there is no longer a toxic effect. WET guidance states "impact from toxics would only be suspected where effluent concentrations after dilution are at or above toxicity effect concentrations" (EPA1991). The report was trying to describe the observation that toxicity was rarely observed in the receiving environment regardless of the magnitude of toxicity measured in first-flush storm water samples.
• I don't recall reviewing the bay sample study plan? Where is a map showing the relationship of the samples to the various outfalls and timing of the samples to the storm events. Some of these they samples were collected up to 10 to 20 days after a storm event? How is this representative of toxicity after a storm event? I suggest that figure 54 be separated into three different figures showing before, during and after storm events.	Receiving water sampling was one of the critical measurement components of the study. The sampling plan provided to the TRT in September 2003 included this component as part of the plan, described receiving water results from sampling events, and provided a map detailing sample locations during the next phase of the investigation. A brief given to the TRT in September 2004 described the efforts in receiving water sampling and described their results to date. The report has maps showing all receiving water sampling locations (Figures 2, 6, 7, and 11) and text to go along with them. Pre-storm water samples were collected before rainfall began, "During" samples were collected while it was raining or storm water was still flowing out drains, and "after" rainfall samples were collected between 12 and 30 hours after rainfall or storm flow stopped. No samples were collected 10 to 20 days after a storm event. Additional figures were added showing before, during, and after storm events.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
Overarching recommendations: It is apparent that the Navy did an extensive job of assessing the temporal and spatial toxicity for their outfalls (n = 136 of the total 350 toxicity tests only 40% on outfalls vs 60% on bay waters). However, clearly toxicity is demonstrated that the various outfalls as taken from Appendix B of the report (see Table 1). I would consider the test results with less than 10% survival or development to be highly toxic samples. Taken from the conclusion of the report, "The 90% survival requirement in the NPDES permit failed in 58% of first flush samples." Please verify the number 58% (see comment above regarding page 105). Clearly this statement demonstrates that the Navy must be addressing toxicity in first flush samples discharges to the Bay. Partial TIEs were conducted and identified copper and zinc as outfall toxicants. The Navy needs to further verify that these are the sole toxicants in these outfalls as this was not clearly demonstrated. The Navy needs to conduct a Toxicity Reduction Evaluation (TRE) to identify the sources and remove those metal sources from these outfalls. In addition, the Navy must conduct additional Phase II and III TIEs (and possible TRE) with additional toxic samples to determine if surfactants and additional toxicant(s) are causing toxicity to topsmelt and mysid test species.	We appreciate the reviewer's acknowledgment of the effort put forward. The Navy agrees that organisms exposed to 100% storm water for 48 or 96 hours may show a toxic response. However, the study clearly shows that receiving water organisms are not exposed to 100% storm water for even a fraction of that time. This fact was demonstrated by the 200+ observations that there was (except in two cases) no toxicity found in receiving waters. This issue is also partially affected by the use of a 90% survival cutoff that declares a sample as toxic when it is not, using normal toxicity reporting methods. The principal CoCs identified by the TIEs were copper and zinc. The Navy uses a continuous and iterative process of BMP implementation, storm water analysis, CoCs source identification, and BMP enhancements to improve storm water discharge quality.
The report needs to be revised based on the peer review comments. A response to comment document which addresses the peer review comments needs to be prepared along with the revised final report. In addition, I suggest a technical editor review the document for consistency purposes. For example, many tables do not include the number of tests there were used to conduct the analyses (Table 12, 20, 27, 34, 41). All tables need clear headings and proper notation of the information within the table (e.g., what is the meaning of it - within Tables in Appendix B?). The paper titled "Evaluating Stormwater Impacts - Monitoring the Receiving Environment Using a Floating Bioassay Laboratory System" has this been submitted to a peer-reviewed journal? If so how will the peer review comments on this report be reconciled prior to submitting for journal publication?	This peer review, along with others, was added as an appendix. The report was edited for technical and editorial content. The paper cited was published in conference proceedings as indicated in the text. It is currently being modified for submission to a peer review journal. Because of deadlines for generating the report and delivering it to the SDRWQCB, peer reviews for this paper will not be available.

DENTON STUDY COMMENTS (cont):	RESPONSE TO COMMENTS:
The data as prepared and documented in this report does not provide a rationale for an alternative approach to determine a valid survival rate for acute exposure to discharges of stormwater from industrial areas at Naval Base Point Loma Complex.	The data produced in this study provide a basis for an alternative toxicity requirement that is both scientifically defensible and protective of San Diego bay waters. The comment suggests that if the report were prepared and documented differently that it may support an alternative approach. The draft report was modified with additional text, tables, and graphics to respond to comments generated by this and other reviewers to better document and support a rationale for an alternative requirement. The final report, not the draft report should be evaluated when making a determination.

DENTON ALTERNATIVES COMMENTS:	RESPONSE TO COMMENTS:
Lastly, I will not be providing comments on "Navy's proposed alternative toxicity requirements for industrial storm water discharges report" as it is inappropriate based on nature of the document and my regulatory position. I would assume that the other peer reviewers should be excusing themselves because of the regulatory implications. This particular document needs to be reviewed and discussed with the San Diego Regional Water Quality Control Board. As you are aware, the draft final report did not include appendices A – H for our review. However, after a phone call, I did receive electronic versions of these appendices. Therefore, I question whether the other reviewers received and reviewed these appendices?	The document along with appendices will be provided to the San Diego Regional Water Quality Control Board. We regret the oversight that some reviewers did not receive appendices. Reviewers that did not receive them were contacted and offered electronic copies.

DENTON TABLE 1 IDENTIFIED IN COMMENTS

Table 1: Summary table of toxicity test results at the Navy outfalls.

Outfall	Test species	Tests with 0 - <10 % survival or developm ent[a]	Tests with < 70% survival or developm ent[b]	Tests with < 90% survival or developm ent[c]
NAV	**Topsmelt**	**2/13; 16%**	**4/13; 31%**	**6/13; 46%**
	Mysid	**2/13; 16%**	**5/13; 38%**	**6/13; 46%**
	mussel	**4/13; 31%**	**9/12; 70%**	**9/13; 70%**
SUB	**Topsmelt**	**0/10; 0%**	**0/13; 0%**	**4/10; 40%**
	Mysid	**0/10; 0%**	**1/10: 10%**	**7/10; 70%**
	Mussel	**9/9; 100%**	**9/9; 100%**	**9/9; 100%**
NAB	**Topsmelt**	**1/8; 13%**	**2/8; 25%**	**3/8; 38%**
	Mysid	**1/4; 25%**	**1/4; 25%**	**2/4; 50%**
	Mussel	**4/4; 25%**	**4/4; 100%**	**4/4: 100%**
NI	**Topsmelt**	**0/7; 0%**	**0/7; 0%**	**2/7; 29%**
	Mysid	**0/4; 0%**	**2/4; 50%**	**2/4; 50%**
	mussel	**3/4; 75%**	**3/4; 75%**	**3/4; 75%**

a = number of tests that had < 10% survival or larval development/total number of samples (i.e., considered ighly toxic samples)
) = number of tests that had <70% survival or larval development in the 100%/total number of samples.
: = number of tests that had <90% survival or larval development in the 100%/total number of samples.

Vote: no TIE test results were included in these numbers.

DENTON REFERENCES IDENTIFIED IN COMMENTS

Denton DL, Fox JF, Fulk FA. 2003. Enhancing toxicity test performance by using a statistical criterion. Environ Toxicol Chem 22(10)2323-2328.

U.S. Environmental Protection Agency. 1986. Aquatic Life Ambient Water Quality Criteria for Chlorpyrifos. EPA-440/5-86-005.

U.S. Environmental Protection Agency. 1995. Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to West Coast Marine Organisms. Eds: Chapman GA, Denton DL, Lazorchak JM. EPA/600/R-95-136. Office of Research and Development. Cincinnati, OH.

U.S. Environmental Protection Agency. 1996. Marine Toxicity Identification Evaluation (TIE): Phase I Guidance Document. Edis: Burgess R, Ho K, Morrison G,.Chapman GA and Denton DL. EPA/600/R-95/054. Environmental Effects Research Laboratory, Narragansett, RI.

U.S. Environmental Protection Agency. 2000. Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications under the National Pollutant Discharge Elimination System Program. Eds: Denton DL, Fox J, Fulk FA, Greenwald K, Narvaez M, Norberg-King TJ, Phillips L. EPA/833/R-00-003. Office of Water. Washington, DC.

U.S. Environmental Protection Agency. 2001. Requirements for Quality Assurance Project Plans. EPA/QA/R-5, EPA/204/B-01/003.

U.S. Environmental Protection Agency. 2002a. Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms. 4th Edition. EPA/821-R-02-013. Office of Water.

U.S. Environmental Protection Agency. 2002b. Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms. 5th Edition. EPA/821-R-02-012. Office of Water.

U.S. Environmental Protection Agency. 2002c. National Recommended Water Quality Criteria: 2002. US EPA Office of Water and Science and Technology. EPA-822-R-02-047.

U.S. Environmental Protection Agency. 2005. Aquatic Life Ambient Water Quality Criteria for Diazinon. EPA-822-R-05-006.

Ms. Eileen Maher, Port of San Diego

MAHER STUDY COMMENTS:	RESPONSE TO COMMENTS:
1. Early in the report, the abbreviations for each respective naval base were identified (NAV, SUB, NAB, NI) and used. The latter sections of the report (Results, Discussion and Conclusions) do not follow this naming structure. It is recommended that the entire report be consistent when naming the respective areas.	We have checked that acronym use for each base to ensure that is consistent throughout the document.
2. It does not appear that surfactants were measured in the general chemistry constituents, yet they have been identified through the TIEs (P61, 76, 90) as potentially being a cause, or partial cause of toxicity. Is there any way to find out the levels of surfactants at these sites? Furthermore, if the chemistry data for surfactants does not exist, how can the TIE results be validated?	Early data collection did not suggest surfactants as CoCs and they came to light as contributory toxic agents the last year of the project (January 2005). The lack of toxic response in bay waters after storm events reflects that these compounds are below levels that cause a toxic response in the bay. The benefit of performing toxicity tests is not all contaminants need to be measured to identify a potential problem. Nonylphenol was identified as a partial cause of toxicity in one sample in the last year of the project. It was identified as a probable cause of toxicity in the TIE based on previous observations of toxic levels by the toxicity laboratory. However, saltwater aquatic life criteria just became available (EPA, February 2006). The acute criterion is 7.0 ug/L. The estimated concentration of nonylphenol in the TIE first-flush sample was 0.18 ug/L. These latest EPA criteria will be included in the report text.
3. The conclusions section needs to be greatly expanded. It is recommended that the conclusions restate the goals and provide brief findings on how each goal was met or addressed and, more importantly, what final outcome was.	We will take these comments under advisement.
4. The recommendations for the proposed two alternative toxicity requirements seem fair and reasonable.	No comment necessary.

MAHER STUDY COMMENTS (cont.):	RESPONSE TO COMMENTS:
5. Exec Summ, p2, 2nd to last paragraph: The plume discussion does not identify that there are any differences in mixing or dispersion as the sites progress further back into the bay. The general understanding is that residence time of water (and possibly pollutants) in the bay and respective dilution/flushing takes longer as you get further into the bay, however this study does not indicate this. It seems likely that some slight differences in mixing times should be observed in the plume measurements because of the longer residence time. Should there be some mention of why this is/is not found during this study? If so, the text in the Exec Summary, as well as the Discussion, and Conclusion sections should address this.	Localized mixing of storm water plumes with the receiving water is driven by local conditions and not residence time of bay waters.
6. Methods, p9, 2nd paragraph: It is mentioned that several drains discharge into Chollas and Paleta Creeks before going into the bay. However, it does not appear that any of those drains were sampled. It is recommended that if drains leading to these creeks are not sampled as part of this project, that it be clearly stated. Otherwise it leads the reader to infer that they may be a part of the study area, especially given the TMDL concerns at those creeks.	The text was adjusted to remove this confusion.
7. Storm Design Criteria, p22: How does this study's design storm criteria differ from what triggers an acceptable storm for the Permit? If it differs, some wording should be added to identify the differences. Also, it may be appropriate to justify why using a storm design criteria that is different than the Permit, still provides data that can be used to develop adequate alternative standards.	The permits only specify that samples be collected during scheduled facility operating hours during the first hour of discharge when preceded by at least three working days without storm water discharge. As such, our design storm only differed in that we planned for sample collection 24hrs/7days per week. The minimum rainfall we imposed was used to ensure we could obtain sufficient sample. Permit language on the requirement was added to the text.

MAHER STUDY COMMENTS (cont.):	RESPONSE TO COMMENTS:
8. In the metals chemistry tables (Results section), the qualifier states that the "grayed out cells are values at the MDL". Earlier in the section (p38), it states that non-detect results are reported as half the MDL value. Additionally, in Methods (p35, tbl 6) metals MDL values are identified. However, the gray shaded values in Tables 14, 15, 22, 29, 36 do not appear to follow the MDL reporting criteria identified on pg 35 or 38. Please clarify and modify if needed.	Table data reported as non-detect are shown as MDL values. The text was corrected to reflect this. A sentence was added to each appropriate section on organics that summations were made using ½ the MDL when there was a non-detect values.
9. Metals results (p 48, 66, 80, 94): The results for each of the sites show metals, namely copper and zinc, being compared to Permit performance goals (63.6 and 117ug/L). Are there other constituents identified in the Permit as having performance goals? It is recommended that a table be added to the methods sections identifying the benchmarks, chronic wq standards, water quality objectives, etc that each constituents will be compared to later in the report.	No other constituents have a performance goal in the permit. A section and table were added to the methods describing these chemical data benchmarks and comparisons.
10. TIE, p60: The text in this section mentions the 3 outfalls at Naval Station. I believe the site should be the Sub Base. Please review the section to see if 1) the text should be moved to appropriate part of the Naval Station's results, or 2) the site was misnamed. This comment provides additional support for using a consistent naming scheme (see Gen Comment 1 above).	The site was misnamed and corrected in the text as Submarine Base.
11. Discussion, Toxicity Eval, p105: It is not clear if mussel larval development tests are required and/or used for the NPDES Permit.	Mussel larval development tests are not required under the permit. The text was altered to reduce confusion.
12. Conclusions, final paragraph, p122: While, I agree with the general assumption that the stormwater runoff is not causing receiving water toxicity, the text as written is very strong and appears too controversial. It is recommended that the language either be expanded or toned down to get the point across without offending those parties of whom you are hoping to gain buy-in for alternative standards.	We understand the nature of this comment and will take this under advisement in finalizing the text in the final report.

MAHER ALTERNATIVES COMMENTS:	RESPONSE TO COMMENTS:
4. The recommendations for the proposed two alternative toxicity requirements seem fair and reasonable.	No comment necessary.

SCHIFF COMMENTS:	RESPONSE TO COMMENTS:
1. Quality of Work First off, this document contains a lot of work; 11 storms, 136 samples, 350 tox tests. The authors are to be commended for just the shear volume of effort applied to this study. I can only judge the quality of the work based on the summaries provided in the body of the text. The raw data were contained in the appendices that were not included with the document and, therefore, I cannot judge the quality at the raw data level. However, based on the textual summaries the data appear sound and of good quality. Negative and positive controls generally performed well. MSD calculations indicated that replicate variability, on average, was not extreme. In my reading of the document, no glaring discrepancies were discussed that should cause me to dismiss the work based on poor quality. However, greater quantification of deviations (i.e. pg 38, 1st Para) may be appropriate. Similarly, portions of the testing were conducted at another facility (Nautilus Environmental) and an intercalibration, or at least a comparison of reference toxicant responses, would seem appropriate. Similarly, chemical analysis deviations are described and qualified, but not well quantified. For example, an indication of holding times exceedence was described and flagged in the database, but no mention of the holding time exceedence magnitude (i.e. 1 day or 100 days?). Chemistry problems such as those described in the report are common in studies of this magnitude and are to be expected, but descriptions of their magnitude are important to ensure that they are not at a level of concern.	Text was added to better quantify the QA/QC items identified. There was no formal intercalibration between laboratories. A comparison of reference toxicant data was be made and included in the text. The MSD data plots and discussion described in the discussion section do provide one type of comparison of the data generated from the two labs.

SCHIFF COMMENTS (cont.):	RESPONSE TO COMMENTS:
Ability to Answer Study Questions 1. Evaluate the magnitude of industrial storm water toxicity from Navy facilities The study appeared capable of answering this question at the sampled sites. There is a constant struggle between a wet weather sampling design that favors more sites for fewer storms or fewer sites for more storms. While no specific facts are given to support their claim, the authors state that the sites selected are representative of all Navy facilities. With this caveat, the range of storm sizes and intensities fits well within the scale of meteorological extremes for this region. My suggestion to follow up on this thought, since the range of conditions is assumed to have been characteristically sampled, would be to conduct some power analysis to determine an optimum sample size for determining the mean, median or extremes in water quality or toxicity. This will also provide context for estimating the confidence bounds derived from this study when describing Navy water quality.	The reviewer clarified the comment on "power analysis" by phone. The suggestion was that Navy consider the numbers and type of data needed when moving forward rather than a calculation necessarily needed in the report.

SCHIFF COMMENTS (cont):	REPSONSE TO COMMENTS:
2. Evaluate causes of toxicity This question was only partially answered. The causes of toxicity were evaluated in three fashions: 1) toxicity identification evaluation manipulations (TIEs), 2) correlations between toxic responses and chemical concentrations; and 3) comparison of estimated toxicity and measured toxicity for presumed contaminants. This method of evaluation is well founded in the literature and entirely appropriate. The specific details of the implementation are difficult to evaluate however. For example, all of the TIE data are in an appendix, which I did not have, while only summaries are provided in the report. Based on the summaries, certain treatments consistently removed toxicity. The EDTA treatment was particularly effective. Therefore, the assumption regarding trace metals is appropriate and, based upon the chemistry data at these sites, trace metals are likely toxicants. Some discrepancies do occur. For example, NAV suffered from a consistent loss of toxicity between initial tests and TIE baseline testing, indicating that some other potential toxicants may have existed. The confounding issue of copper and zinc removal by the C18 column has been experienced in our lab. In later experiments, the laboratory uses C18 column elutions to confirm the presence/lack of toxicity, which is an appropriate step. In general, my recommendation to the authors is to expand the TIE results text a bit to provide more quantitative evaluations so the reader does not have to take the conclusions on faith (i.e. pg 43). One factor that prohibits me from completely buying into the fact that metals are the only or primary toxicants is the poor predictive relationships between toxic effect in both the mussel and mysid and the predicted toxic units based on copper + zinc. To overcome this obstacle, the authors may want to try predicting toxicity from chemistry using non-linear models. Alternatively, the use of other predictive tools such as the Biotic Ligand Model may work well. At a minimum, the authors should investigate and report on the quantitative relationships between dissolved trace metal concentrations and potential binding agents (i.e. hardness, DOC, etc.).	The appendices with the full report from the contractor will also be included in the final report (we regret the omission of these reports during the review process). We agree the dose-response curves were quite variable and are not definitive for a causal relationship of toxicity to metals. Non-linear models were attempted with similar outcome. Toxicity and chemistry correlations with DOC and other parameters were evaluated with no better outcome. Text was added that describes some of the evaluations performed.

SCHIFF COMMENTS (cont.):	RESPONSE TO COMMENTS:
Additional data analysis The strong suit of this study is the shear volume of data generated and it is in my opinion that the authors have not analyzed the data set completely. Either within this document, or as part of another document, the authors should start exploring a number of analyses well suited to this data set. Here are just a few examples: 1) effect of storm characteristics (i.e. storm size, intensity, timing within season, cumulative rainfall, etc.) on water quality or toxicity; 2) relationship of water quality to catchment activities (clearly some sites are worse than other (i.e. NAB); 3) comparison to other industrial activities (RWQCB 4 has a compilation of tox and WQ from all industrial SIC codes in their region) to see if Navy bases are better than or worse than other industrial types.	We appreciate the reviewer's acknowledgment of the effort put forward. The study design does not support a statistical evaluation of individual storm characteristics. However, the study clearly shows that receiving water quality and toxicity were affected after an exceptionally long dry period. The Navy is investigating ways to mitigate this worst-case potential for effects. The Navy has an active program that is continuously implementing and improving BMPs for industrial drainage areas within their SWPPP (iterative approach). Results of the study along with results of standard storm water monitoring have identified locations that are prioritized for additional efforts. While a comparison of the magnitude of toxicity and chemistry measured in other industrial discharges is interesting, mass loading and the potential for impacts to bay waters is dependent on relative discharge volumes. These evaluations are part of ongoing Bay TMDL investigations and should be made on a watershed approach.

SCHIFF COMMENTS (cont.):	RESPONSE TO COMMENTS:
Interpretation The specific conclusion that the authors pose is that although runoff from Navy facilities may be toxic, it does not impact San Diego Bay receiving waters. From a technical perspective, this is only partially true. The data on receiving water testing is especially compelling to support this statement. I was particularly impressed by the flow through testing used OF 14. The main problem with any toxicity test is trying to simulate actual exposure. In this case, the organisms were exposed to actual Bay waters during/following wet weather for the duration of the test, which is as close to real exposure as one could expect. Storm composites represent the next closest approximation to exposure because composites at least integrate the variations in concentration over the course of an entire storm (which can be extreme). Finally, grab samples represent the least realistic exposure because it is only a single moment in time and does not take into account within storm variability or receiving water dilution. The lack of toxicity between wet weather discharges and receiving water toxicity at a subset of these locations has been observed previously in San Diego Bay, only this time associated with Chollas Creek (Schiff et al 2003).	No comment necessary.
The conclusion is not supported because not all impacts are exerted in the water column at the time of discharge. For example, there are several areas in San Diego Bay near industrial facilities (not just the Navy) that suffer from contaminated sediments (i.e. sediment chemistry, sediment toxicity, and benthic community impairments). This study also showed that, at times, large concentrations of several constituents including trace metals, PAHs, and pesticides/PCBs are discharged through Navy outfalls. These outfalls, in combination with other potential sources, may be contributing to the contaminated sediments that exert their effects at longer time and/or spatial scales.	The focus of the effort was to evaluate the efficacy of the toxicity requirement applied at the end-of-pipe in evaluating the potential for receiving water impacts. Based on this study, the current requirement does not do well at predicting toxic impacts in the receiving water and therefore will do no better at predicting the potential for impact to sediments. Large concentrations do not correspond to large mass loads. The potential for sediment impacts should be evaluated through programs designed for that purpose such as the TMDL program, baywide monitoring program…etc. We will modify the text to focus the goals and results on the efficacy of the WET test requirement rather than the broader question of any impact to the bay.