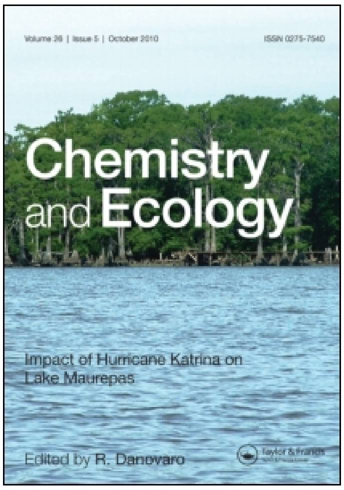

Chemistry and Ecology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713455114

The Effects of Ocean Disposal of Sewage Sludge On the Relative Abundance of Benthic Megafauna

Robert A. Pikanowski[a]
[a] Sandy Hook Laboratory, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northeast Fisheries Center, Highlands, New Jersey

To cite this Article Pikanowski, Robert A.(1992) 'The Effects of Ocean Disposal of Sewage Sludge On the Relative Abundance of Benthic Megafauna', Chemistry and Ecology, 6: 1, 199 — 212
To link to this Article: DOI: 10.1080/02757549208035272
URL: http://dx.doi.org/10.1080/02757549208035272

Chemistry and Ecology, 1992, Vol. 6, pp 199–212
Reprints available directly from the publisher
Photocopying by license only from:
Gordon & Breach Science Publishers S.A.

THE EFFECTS OF OCEAN DISPOSAL OF SEWAGE SLUDGE ON THE RELATIVE ABUNDANCE OF BENTHIC MEGAFAUNA

ROBERT A. PIKANOWSKI

U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northeast Fisheries Center, Sandy Hook Laboratory, Highlands, New Jersey 07732

(Received 5 August, 1991; in final form 8 October, 1991)

Disposal of sewage sludge in the New York Bight Apex (12-Mile Dump Site) ceased at the end of December 1987. Previous efforts to quantify the effects of sludge were hindered by the inability to obtain true replication. The cessation of dumping afforded the opportunity to apply the technique of replication in time, also known as a Before/After, Control/Impact design. Conditionally, this method allows one to separate treatment effects from the natural differences that confound many environmental impact studies. The Environmental Processes Division of the Northeast Fisheries Center, National Marine Fisheries Service sampled the benthic environment of the New York Bight Apex from June 1986 through September 1989 using a sample design based on the technique of replication in time.

Three dominant species (rock crab, *Cancer irroratus;* little skate; *Raja erinacea;* and winter flounder, *Pseudopleuronectes americanus*) and total demersal finfish, collected by otter trawl, showed no statistically significant response to the cessation of disposal. American lobster (*Homarus americanus*) increased in local abundance, but this result was possibly confounded by a change in fishing effort.

KEY WORDS: Sewage sludge, marine disposal, New York Bight, benthic ecology, demersal finfish

INTRODUCTION

It is difficult to quantify environmental impacts precisely using inferential statistics (Hurlbert, 1984) since it is usually not feasible to conduct experiments in the real environment. Replication in time (RIT) is suggested as a way to overcome this dilemma (Stewart-Oaten *et al.*, 1986). The Environmental Processes Division, Northeast Fisheries Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration (NOAA) used the RIT paradigm to investigate effects of the termination of sewage sludege disposal on measures of benthic ecology. The results for a subset of this study, catch per tow of dermersal finfish and benthic mega-invertebrates, are presented.

The 12-Mile Dump Site

Municipalities surrounding the New York Bight (NYB) have disposed of sewage sludge at the 12-Mile Dump Site (12-MDS, Figure 1) since 1924. The maximum annual disposal rate was 8.3 million metric tons released in 1983 (Santoro, 1987). Details of the hydrodynamic transport of the sludge are still under investigation (Manning, 1991; Davis and Walker, 1991). Empirical evidence obtained from sewage markers such as coprostanol, tomato seeds, or bacterial spores, suggests that sludge material accumulated in low potential energy areas primarily to the west and

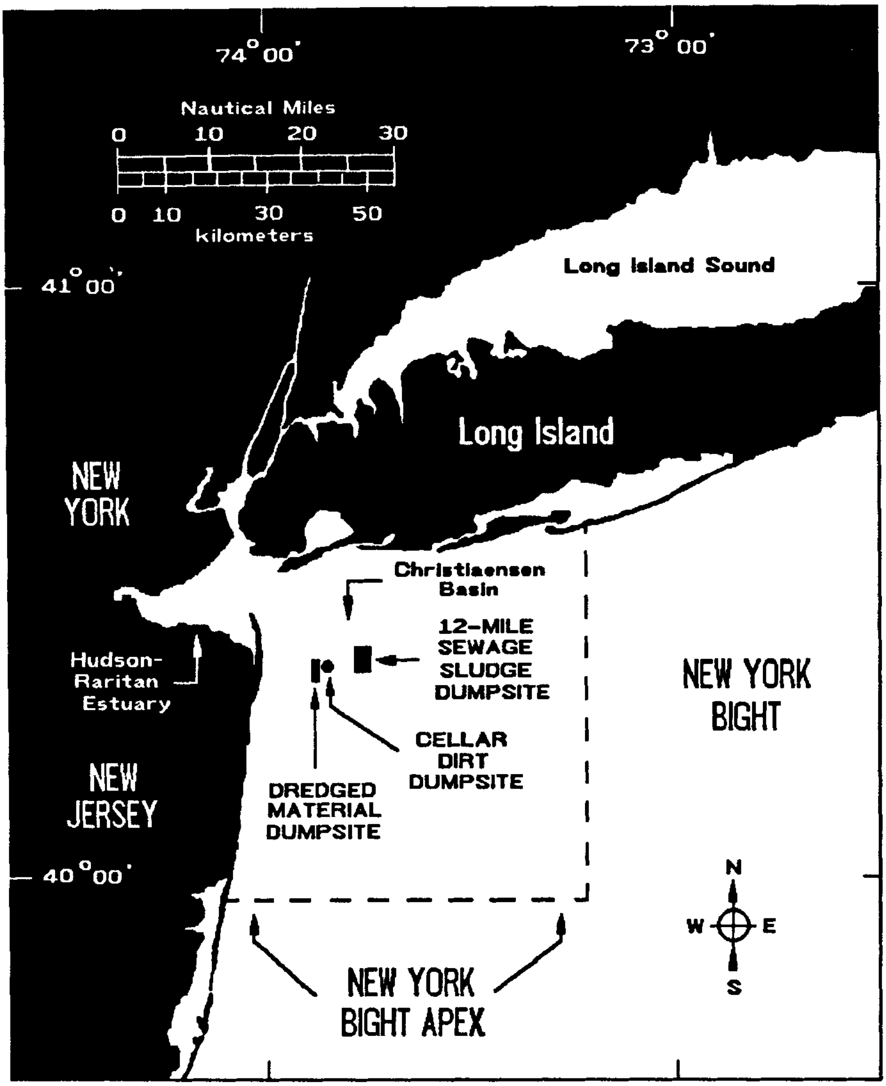

Figure 1 Location of the 12-Mile Dump Site and the New York Bight.

to the south of the dump site in the Christiaensen Basin (Environmental Processes Division, 1988). The sludge provided a large input of labile carbon, an energy source for secondary benthic production which, under appropriate weather driven conditions, could exacerbate local hypoxia (Boesch, 1982). The sludge contained measurable quantities of microbial pathogens, heavy metals, and toxic organic compounds (Stanford and Young, 1988). Elevated levels of pathogens led to closure of surf clam (*Spisula solidissima*) beds, contributing in large part to the justification of terminating disposal at the 12-MDS. Closure of the 12-MDS occurred at the end of December 1987. It is intuitive that sludge disposal had some effect, at least locally, on benthic ecology. This study was designed to quantify these changes.

The Statistical Dilemma

Monitoring and survey programmes conducted during the sludge disposal era generated considerable descriptive information (Gross, 1976; Mayer, 1982, Reid, *et al.*, 1987). Unfortunately, this information could not easily answer a basic question: "What is the ecological effect of ocean disposal of sewage sludge in the NYB?" The framework of these programmes simply did not provide a proper statistical model for hypothesis testing.

The basic question can be restated as the null hypothesis $-_1H_0$: *Oceanic disposal of sewage sludge has no effect on benthic ecology.* To test this hypothesis one would need replicate experimental units (potential worldwide dump sites). The treatment (disposal) could then be randomly allocated to half of them, with the other half left untreated. Clearly this experiment is impractical.

A more realistic hypothesis could be $-_2H_0$: *Disposal of sewage sludge has no effect on the benthic ecology of the New York Bight.* This experiment cannot be performed because there is only one NYB; there are no experimental units over which the treatment can be replicated.

Consider then $-_3H_0$: *Disposal has no effect on those sites in the NYB which accumulate sewage sludge.* This hypothesis appears testable using data from the historic studies. The ecology of sites which accumulate sludge can indeed be statistically compared to nearby sites that do not evidence accumulation. Interpretation of the results is then problematic. Sites that accumulate sludge are depositional and may be expected to have a different basic ecology than the nondepositional sites not accumulating sludge. In addition, depositional sites are likely to have differential exposures to nutrients, pathogens, and contaminants coming from other sources in the NYB. The effects of sludge, therefore, cannot be separated from other differences between the two groups because the treatment was not randomly allocated between hydrographic groups.

It follows to pose the hypothesis $-_4H_0$: *Disposal has no effect on depositional sites in the NYB that accumulate sludge compared to nearby depositional sites that do not accumulate sludge.* The dispersion of the sludge is contiguous; sites that accumulate sludge are geographically segregated from nearby control sites. Ecological gradients, now due to locational differences, can be expected to confound interpretation of results; the treatment is not randomly allocated in a spatial sense.

With the termination of disposal, one could pose $-_5H_0$: *The ecology of sites that accumulate sludge does not change after the cessation of disposal.* Samples taken before and after termination can be compared statistically. However, in a dynamic system such as the NYB, ecological measurements (such as abundance or distribution of species) are likely to change over time. Ecological gradients, due to such natural variation, will confound interpretation of results; the treatment is not randomly allocated in a temporal sense. However, cessation of disposal does hold a key to proposing a hypothesis in which the chance occurrence of confounding effects is minimized.

Replication in Time

Consider the narrower hypothesis $-_6H_0$: *For effects that are reversible, the difference in ecology between an affected site and unaffected site does not change after the cessation of disposal.* The qualifying clause is necessary because the experiment was, in essence, run backwards. Theoretically, sludge disposal may have irreversibly changed a component of the system (at least within the time-frame of the study). For such a component, the experiment will lead to the false conclusion that sludge has no effect when the opposite is indeed the case. Because the population pool of NYB species has a much greater range than the small area affected by sludge, irreversibility is unlikely. To test ${}_6H_0$, an affected site and a nearby unaffected (control) site are selected. The expectation is that they are similarly influenced by large scale driving forces and random events, such as climate or fluctuations in abundance. That is, on average, the sites are both exposed to and react to such events equally. Both sites are simultaneously sampled on multiple occasions, before and after the onset of the treatment (which in this case is the cessation of disposal). On each sampling occasion, the difference between sites is calculated. This procedure is known as a Before/After, Control/Impact (BACI) or RIT design. For an event unrelated to the treatment to be confounding, it must be a perturbation which: 1) affects one site only, 2) is large in magnitude, 3) is long-lasting, and 4) occurs mostly in the before or after cesation period. Events not fitting these criteria contribute to variance and should not bias period means. If no confounding event occurs, statistically significant change in the difference between the sites, between the before and after periods, is due to the cessation of disposal or chance at the stated significance level of the test. The Environmental Processes Division designed a study making use of this method to test ${}_6H_0$ in the NYB (Environmental Processes Division, 1988).

METHODS

The 12-Mile Dump Site Recovery Study

The Environmental Processes Division selected three sites (circumscribed by ellipses in Figure 2) to be used in the RIT model. Although the historical studies cannot be used to quantify sludge effects, they can provide an assessment of exposure. Based upon information generated by the historical studies, station NY6 was selected as the contaminated site and station NY11 was selected as the uncontaminated (or control) site. Station R2 was a nutrient enriched site that may have been intermediate in exposure to sludge (Environmental Processes Division, 1988). Here, R2 is designated an alternative control. This increases the likelihood of having a control site which responds similarly to events unrelated to the treatment.

Eight subsamples were collected at each site in odd numbered months, plus August, making a total of seven replicate sampling opportunities per calendar year. Fish and mega-invertebrates were collected with an otter trawl having an 11.0 m footrope and a 9.8 m headrope. The body and cod end of the trawl were constructed of 76 mm and 51 mm stretch mesh nylon, respectively. The trawl was towed at an average of 4 km h^{-1}. Tows of fifteen minutes yielded an area swept by the gear of approximately one hectare. Sampling began in July 1986 and the last samples were collected in September 1989. Thus, there were 11 replicate sampling opportunities before the cessation of disposal and 13 after. The experimental design, the statistical model, the variable list, and sampling methodologies are detailed in a formal study plan (Environmental Processes Division, 1988).

Statistical Considerations

Replicates in time (henceforth, simply replicates) were generated by first averaging the values of the eight subsamples collected at a site each month. Thus, relative megafaunal abundance was estimated by the mean capture weight per tow. The

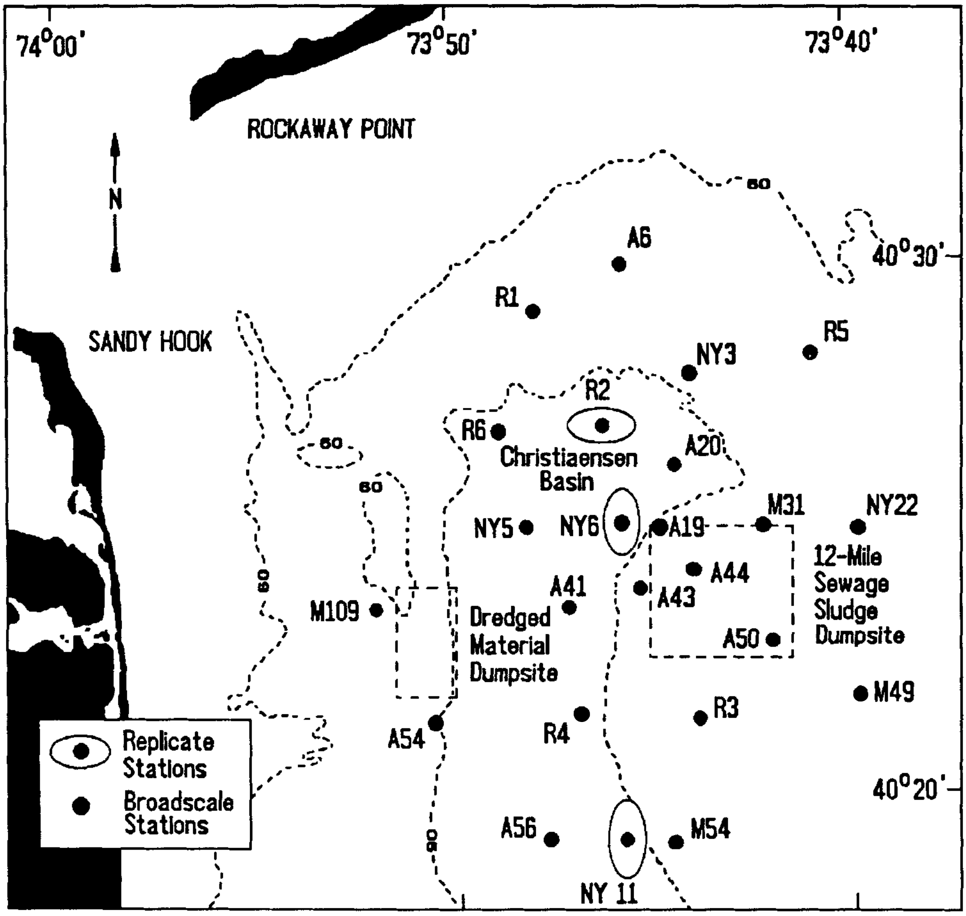

Figure 2 Sites sampled in the 12-Mile Dump Site Recovery Study. The contaminated (NY6) and control (NY11, R2) sites used in the replication in time analysis are circumscribed by ellipses. The other stations were part of a broad scale study.

average value at a control site (NY11 or R2) was subtracted from the average value at the contaminated site (NY6), yielding a replicate. Because the sample size (24 replicates) is too small to test reliably for normality, both the Student's t and Mann-Whitney U tests (Zar, 1984) were used to detect statistical differences between the before and after periods. In general, the two tests will differ only if the results are borderline or if the data are severely non-normal. In either case a more detailed treatment of the analysis would be called for, such as a transformation of the data. The tests are two-sided (with $\alpha = 0.05$); the objective is to determine change and there was no prior reason to predict a direction.

The treatment was not truly replicated over independent experimental units because the same sites were repetitively sampled. It is fully expected that the value of a variable at any sample site will predict, to some degree, the value at that site on the next sampling opportunity (serial correlation). The expectation is that the differences between sites will be uncorrelated; that is, effects not due to the treatment either are equal at both sites and cancel when the difference is calculated, or are unequal but do not persist. For example, a storm or fishing effort may lead to a change in abundance at one site but not the other. If this perturbation does not persist, the expectation is that it and other similar perturbations will merely contribute an unbiased variability. If the perturbation does persist, it may be detected as serial correlation. Serial correlation of the replicates, within periods, was tested nonparametrically with a runs test for randomness (whether successive replicates are above or below the median) and parametrically with the C test (Zar, 1984). These tests are one-sided ($\alpha = 0.05$); only positive serial correlation (high/low values followed by high/low values) is confounding. When serial correlation is found within the before period, the experiment may be confounded; some random or long term process may be causing any difference between sites to change independently of the treatment.

Serial correlation between replicates may be due to additivity that is periodic (Stewart-Oaten, *et al.*, 1986). Periodic additivity can occur when the difference between sites is correlated with the magnitude of the site values, yielding a time plot with a sinusoidal shape. For example, times of high abundance may yield larger differences between sites than times of low abundance. This correlation may be unavoidable for some variables that evidence seasonality. Seasonal periodic additivity will increase variance and decrease the power of the test, but will not bias the test if the before and after periods span similar calendar intervals. If the correlation is not due to periodic additivity, adjustments using cofactors are necessary to avoid qualifying the results (Stewart-Oaten, *et al.*, 1986). Inspection of a graphical display of the data will usually be sufficient to determine the type of serial correlation found.

RESULTS AND DISCUSSION

Tests for serial correlation were not significant for any of the replicate biological data. Thus, there are no *detectable* confounding events requiring qualification of the results. Results for megafauna are summarized in Table 1. NY6-NY11 and NY6-R2 represent replicates in time using sites NY11 and R2, respectively, as controls. Also included are before and after results for NY6 alone. These results can be used to test the inappropriate ${}_5H_0$ and are shown for comparison with RIT.

Temperature

Temperature is not a response variable; here, the application of RIT is heuristic. Temperature controls the rates of biological processes and is a major factor in determining the distribution of marine species. If the temperature difference between the contaminated and control sites changed between the before and after periods, results from the experiment could be misleading. That is, the temperature differential could be a confounding factor. The summer of 1988 was anomalously cold (Figure 3; Wilk, *et al.* in review). Average temperatures at NY6 before cessation were 2.5°C colder than after cessation. Although this comparison is not proper (the before and after periods span different seasons), it illustrates that a test of ${}_5H_0$ could be confounded by this anomaly. However, the anomaly occurred at all three sites and thus did not appear to bias the replicates (Figure 4).

However, treatment site NY6 was in fact 0.36°C warmer than control site NY11

Table 1 Results of the 12-Mile Dump Site Study. Values represent the mean catch per trawl tow, in kilograms, before and after abatement of disposal of sewage sludge.[a]

		MEAN (Std. Error)				
TEST GROUP	*SITE*	*BEFORE*	*AFTER*	*CONTRAST*	*t*	*u*
Little Skate	NY6	4.25(1.13)	4.55(0.90)	−0.31(1.42)	−0.21	80.0
	NY6–NY11	−1.78(0.89)	−0.23(0.77)	−1.55(1.17)	−1.32	87.0
	NY6–R2	−4.66(1.34)	−2.00(1.71)	−2.65(2.23)	−1.19	99.0
Winter Flounder	NY6	1.31(0.27)	2.11(0.29)	−0.80(0.40)	−2.00	104.0
	NY6–NY11	0.66(0.26)	1.27(0.26)	−0.62(0.37)	−1.67	99.0
	NY6–R2	−0.53(0.43)	0.15(0.45)	−0.68(0.62)	−1.09	89.0
Demersal Finfish	NY6	9.07(2.40)	10.90(1.46)	−1.83(2.72)	−0.67	90.0
	NY6–NY11	0.60(1.79)	1.54(1.56)	−0.94(2.36)	−0.40	81.0
	NY6–R2	−4.73(1.80)	−2.42(1.54)	−2.30(2.36)	−0.98	91.0
Rock Crab	NY6	4.85(1.85)	2.12(0.63)	2.72(1.83)	1.49	93.5
	NY6–NY11	3.31(1.74)	0.45(0.44)	2.86(1.67)	1.72	97.0
	NY6–R2	−0.15(1.98)	−2.92(1.04)	2.77(2.14)	1.29	78.0
American Lobster	NY6	0.38(0.10)	1.01(0.27)	−0.63(0.31)	−2.07	101.5
	NY6–NY11	−0.05(0.19)	0.33(0.12)	−0.38(0.22)	−1.76	98.5
	NY6–R2	0.14(0.13)	0.69(0.20)	−0.56(0.25)	−2.20	105.5

[a] Values under the column labelled contrast are differences (before minus after) being tested. Student's t-test has 22 degrees of freedom. Negative values of t imply that biomass at NY6 has increased after abatement. N_1 is 11 and N_2 is 13 for the Mann-Whitney U test. Critical values are $t \geq 2.074$ and $U \geq 106$. NY6 is the contaminated site. NY11 and R2 are alternative control sites. NY6–NY11 and NY6–R2 represent test results using the technique of replication in time.

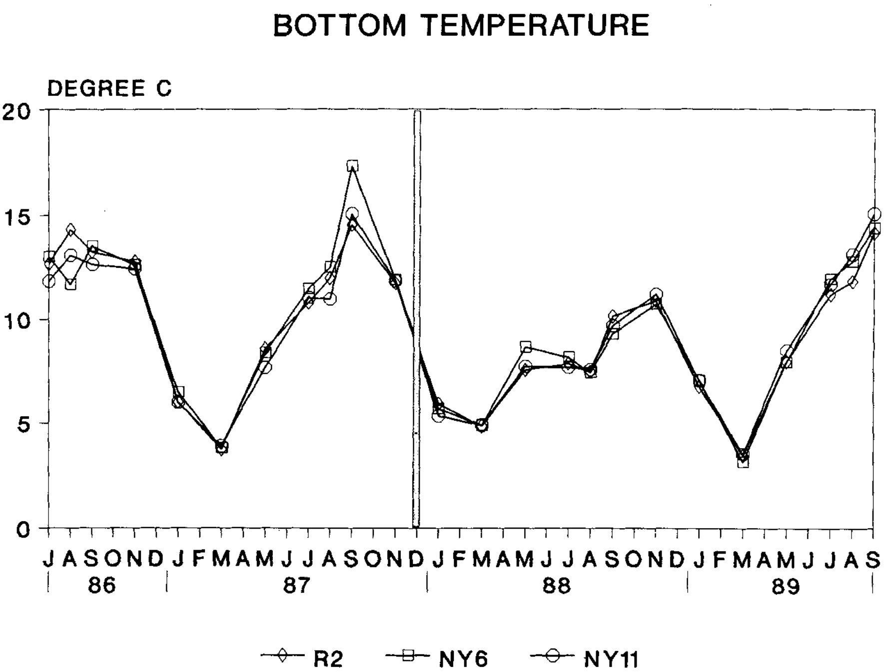

Figure 3 Averaged bottom temperature at the study sites. The vertical bar denotes the cessation of sewage sludge disposal.

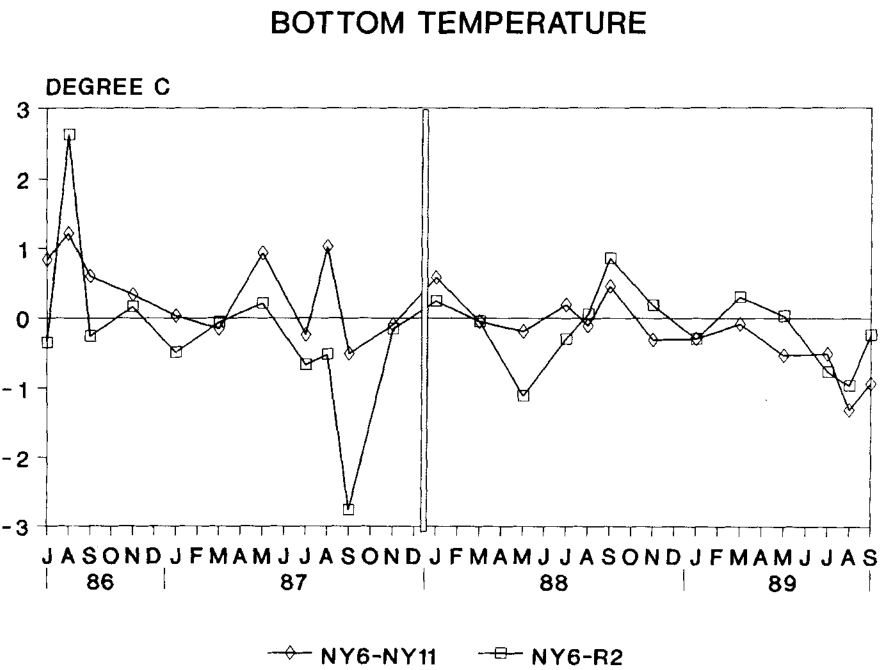

Figure 4 Temperature differentials between the contaminated (NY6) and control (NY11, R2) sites. The vertical bar denotes the cessation of sewage sludge disposal.

before cessation and 0.24°C cooler after cessation. The parametic test on the difference between period means (0.60°C, the amount NY6 cooled with respect to NY11) was significant (t_{37} = 2.65, $P < 0.02$). The nonparametric test was also significant ($U_{11,13}$ = 106, P = 0.05). There was no serial correlation between replicates in the before period ($C < 0.0$). There was significant parametric serial correlation of the replicates (C = 0.60, $P < 0.01$) in the after period, although the nonparametric test yielded P = 0.10. It seems that this correlation can be explained by the downward trend in the NY6–NY11 temperature difference from August 1988 through to the end of the survey (Figure 4). The value for March 1989 was slightly above the median value (May 1988), otherwise the nonparametric statistic would also have been highly significant. A similar result for a response variable would lead one to conclude that the sites are still equilibrating or some outside factor is operating, such as this temperature trend. Usually, serial correlation in a response variable in the after period, as opposed to the before period, requires no special consideration because one is looking for a change. However, temperature is a driving force; therefore, results for response variables sensitive to temperature should be interpreted carefully.

The relative cooling of NY6 with respect to NY11 was statistically significant. An investigator must judge whether the change of 0.6°C is ecologically significant, thus possibly confounding a particular result. Station NY6 warmed by 0.05°C with respect to R2. This small difference was not significant and no serial correlation of the replicates was evident. If an investigator is studying a variable that is highly dependent on temperature, R2 would make a better control than NY11, other things being equal. Here, this difference is judged to be nonconfounding because neither site strays outside the temperature range of any of the dominant species (Bigelow and Schroeder, 1953).

Catch composition

Six species accounted for ~ 86% of the demersal finfish biomass captured by trawl in the study (Wilk, *et al.* in review). Individual percentages were: little skate (*Raja erinacea*), 54%; winter flounder (*Pseudopleuronectes americanus*), 14%; ocean pout (*Macrozoarces americanus*), 9%; fourspot flounder (*Paralichthys oblongus*), 4%; red hake (*Urophycis chuss*), 3%; and windowpane (*Scophthalmus aquosas*), 2%. Two species accounted for 97% of the crustacean biomass: rock crab (*Cancer irroratus*), 84%, and American lobster (*Homarus americanus*), 13%. Other species (for example: spiny dogfish (*Squalus acanthias*); butterfish (*Peprilus triacanthus*); longfin squid (*Loligo pealei*); or horseshoe crab (*Limulus polyphemus*)) sometimes comprised a large part of the catch but are excluded from consideration because they are nondemersal or are known to be transient.

Demersal finfish

Little skate at contaminated site NY6 increased from 4.25 to 4.55 kilograms per tow after cessation (Table 1). The NY6–NY11 replicates showed an increase of 1.55 kg/tow and the NY6–R2 replicates increased 2.65 kg/tow (Table 1). Neither of these increases were significant (α = 0.05, two-sided) because the standard errors of the differences between means were 1.17 and 2.23, respectively (Table 1). Although the general abundance of little skate seemed to decline over the course of the study (as evidenced by calculated decreases at NY11 and R2 of 1.3 and 2.4 kg/tow respectively), this species seemed to find conditions equally favourable at NY6 after cessation. The implication is that an unperturbed NY6 may carry a higher density of skate than the two control sites. However, the study results do not confirm this; that is, it cannot be said that sludge disposal affects the local abundance of little skate. A test of ${}_5H_0$ leads to a strong conclusion that sludge has little effect ($t_{22} = -0.21$, Table 1).

Winter flounder increased from 1.31 to 2.11 kg/tow at NY6 (Table 1). The mean difference for NY6–NY11 increased by 0.62 kg/tow and for NY6–R2 increased by 0.68 (Table 1). Here, general abundance appeared to remain constant while abundance at NY6 seemed to increase. Once again the study failed to confirm this as the results were not significant ($P > 0.1$ and $P > 0.2$, respectively). In contrast, a test of ${}_5H_0$ yields a borderline result ($t_{22} = -2.00$, $P \sim 0.07$, $U_{11,13} = 104$, $P \sim 0.07$).

Upon examination of the data, the study could not be used to test for an effect on the abundance of fourspot flounder, ocean pout, or red hake. Red hake and ocean pout were present in January through May only. These months were sampled only once in the before period so there are insufficient data to draw a conclusion. This situation is illustrated for fourspot flounder, which were most abundant in May (Figure 5).

Demersal finfish as a group increased from 9.07 to 10.90 kg/tow at NY6 (Table 1). NY6-NY11 increased by 0.94 kg/tow and NY6-R2 increased by 2.30 (Table 1). Neither of these results are significant. The reason for this is the large standard error (2.36 kg/tow for both replicate contrasts). One could assume that this variability may be a function of seasonality. A plot of the data (Figure 6) does not evidence any pattern. Indeed, a test for serial correlation of the replicates is highly nonsignificant ($C < 0.0$).

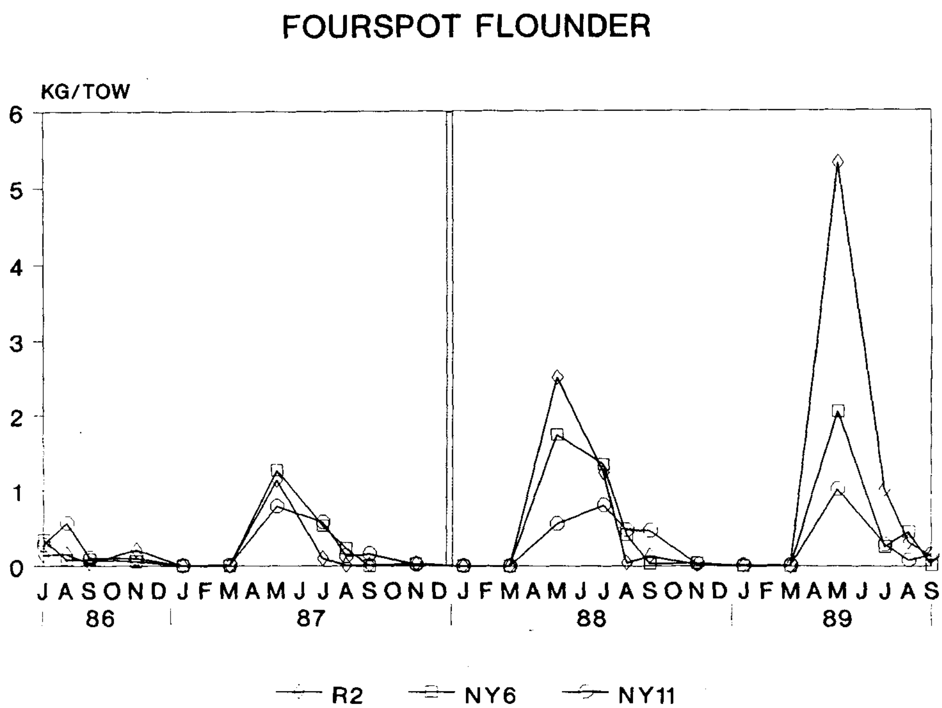

Figure 5 Abundance of fourspot flounder (*Paralichthys oblongus*) at the study sites. The vertical bar denotes the cessation of sewage sludge disposal.

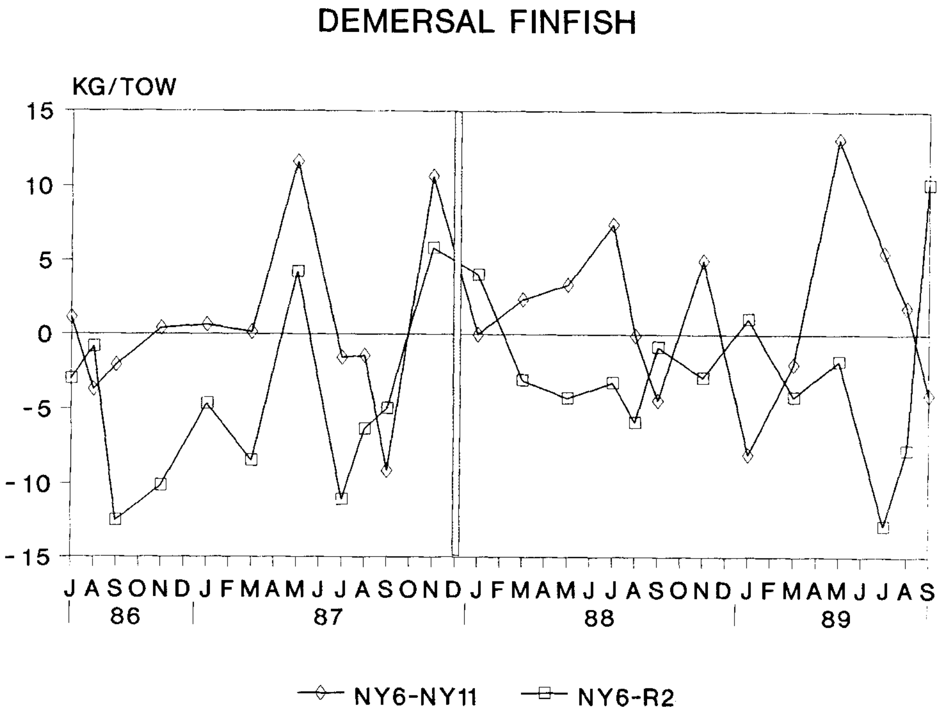

Figure 6 The replicate data (differences between the monthly means of the contaminated site, NY6, and the control sites, NY11 and R2) for demersal finfish biomass. The vertical bar denotes the cessation of sewage sludge disposal.

Crustaceans

Rock crab dominated the crustacean biomass; thus, the results for total crustacean biomass were similar to the results for rock crab and are not reported. Rock crab decreased from 4.85 kg/tow to 2.12 at NY6 (Table 1). NY6–NY11 decreased by 2.86 kg/tow and NY6–R2 decreased by 2.77 (Table 1). Although these decreases represent a halving of biomass at NY6, they are not statistically significant. Part of the explanation for the large decrease (and the accompanying large standard error) rests in a data point that is a statistical outlier as tested by the method given by Sokal and Rohlf (1981). In the first month (July 1986) of the survey, 20.6 kg/tow of rock crab were caught. This is double the next highest value. When the tests are run with this point removed, NY6–NY11 decreased by 1.31 kg/tow and NY6–R2 decreased by 0.8. The standard error also decreased, but the results were still nonsignificant. If the recalculated results had been significant, the question would have had to be posed as to whether July 1986 could indeed be removed from the analysis. During this month, NY6 experienced a bloom of the tube building worm *Asabellides oculata*. This "spaghetti worm" was in sufficient abundance to foul the trawl net (personal observation). By the next sampling opportunity (August 1986), it had disappeared. Assuming that crabs were drawn to this abundance of prey, the question then becomes whether the bloom was a random event unrelated to sludge, in which case the data could be treated as an outlier; or, whether the bloom was a result of disposal, thus making the data valid. At this time, the question remains unanswered.

American lobster increased from 0.38 kg/tow to 1.01 at NY6 ($P \sim 0.05$), NY6–NY11 increased by 0.38 kg/tow ($P \sim 0.09$), and NY6–R2 increased by 0.56 ($P \sim 0.04$) (Table 1). One could conclude (under either ${}_5H_0$ or ${}_6H_0$) that lobster prefer an unperturbed NY6. This conclusion could be wrong. Because sludge fouls lobster gear, the NY6 site was not fished commercially until after cessation. The number of pot strings increased from essentially zero to dozens at the NY6 site (personal observation). This is the type of event that may cause the replication in time technique to yield erroneous results. It is long lasting, affects one site only, occurs in one period only, and may be at the magnitude of the treatment effect. If potting reduced the local lobster population, the mean abundance at NY6 in the after period was biased downward. In this case, the sludge effect was greater than the study indicated. Conversely, if the introduction of bait and structure on an otherwise featureless bottom attracted lobster (thereby making them more accessible to the trawl), the mean abundance in the after period was biased upward. Here, the sludge effect would be less than the study indicated. If an estimate of the effect of potting was available, it could be used to correct the data; then, the results could be accepted without qualification.

CONCLUSIONS

Confidence intervals around the estimate of the effect of sludge (difference between the before and after means) can be generated. These can be used to specify a range of ecological costs (in relative abundance) of sludge disposal. In general however, ${}_6H_0$ was not rejected. Except for the possibly confounded results for American lobster, the study was unable to detect a significant response, measured in biomass per trawl tow, of benthic megafauna to the cessation of sewage sludge disposal at the 12-MDS. This situation can be interpreted in two ways. First, the lack of a response is real. This implies that benthic megafauna are insensitive to the effects of sludge and are thus poor "indicator" species. Because these species are robust and mobile, it could be theorized that their direct response to the presence of sludge is weak. They perhaps respond only indirectly through trophic interactions. A strong response may have been possible if sludge had locally exacerbated a global hypoxia. However, hypoxia did not occur during the course of the study. The minimum mean value of dissolved oxygen concentration (2.8 mg/l) was recorded in September 1989 at R2 (Arlen *et al.*, in preparation).

Second, the experiment, as executed, may not have possessed sufficient power to detect an actual difference. The question then becomes – "What difference did the experiment have the power to detect?". Using the sample variance as an estimate of the population variance, the minimum detectable difference δ can be calculated as:

$$\delta \geq SE\cdot(t_{\alpha(2),22} + t_{\beta(1),22}) \qquad \text{(Zar, 1984)},$$

$$\delta \geq 3.395\cdot SE,$$

where SE is the standard error of the difference between the means (listed in Table 1) and the t's are the tabulated values for 22 degrees of freedom. Here, the significance level α is set at 0.05 ($t_\alpha = 2.074$, two-sided) and the power $1\text{-}\beta$ is set at 0.9 ($t_\beta = 1.321$, one-sided).

For demersal finfish, SE is equal to 2.36 with either NY11 or R2 as controls, so that $\delta = 8.01$. Holding other estimates constant with NY11/R2 as the control, the actual after period biomass at NY6 needed to be less than or equal to 1.95/0.58 or greater than or equal to 17.97/16.60 to have had a 90% chance of being found significantly different. Suppose a manager decided that a change of 5.4 kg/tow was ecologically significant. Using the method given by Zar (1984), the power of the test would be approximately 0.5, i.e. the experiment would detect an ecologically significant difference only half the time. For differences less than the SE (as both the sampled differences in fact are), the power of the test approaches the significance level of the test. That is, the chance of detecting a difference is equal to the chance of falsely rejecting the null hypothesis! Because any result is then utterly ambiguous, such an experiment surely can be said to have failed. On the other hand, if the δ declared necessary to be ecologically significant is greater than 8.01, this experiment would have a detection power greater than 90%. The experiment would then be a success and the result could be confidently accepted.

When designing experiments with a single variable (or multiple variables and the ability to optimally allocate samples), reliable variance estimates, and prespecified δ's, the number of samples necessary to assure success (within the probabilities selected for α and β) can be calculated (for example: Green, 1989 or Saila, *et al.*, 1976). A decision can then be made as to whether the resources are available to perform the experiment. This could not be done for the 12-MDS study. Reliable variance estimates were not available. Logistics demanded that the number of samples be the same for all variables measured. Thus, sampling could not be optimized for individual variables with differing sample size requirements. Management requirements for δ were unspecified. In essence, the 12-MDS study was a pilot study (albeit at full scale). In such situations, managers must evaluate the success of an experiment by specifying what they feel is ecologically significant and determining the power of the experiment as above.

Acknowledgements

I would like to thank members of the Environmental Analysis Investigation for their work in collection, data processing and assistance in preparation of this manuscript. Included in this group are Judith Rugg, Donald McMillan, Beth Phelan, Linda Stehlik, Anthony Pacheco and Stuart Wilk. I have much gratitude for the operators of the NOAA research vessel KYMA – Capt. Fred Farwell and Sherman Kingsley. Without their consummate professionalism this study would have been a much more difficult undertaking.

References

Arlen, L., Draxler, A.F.J., and Bruno, R. (in preparation) Hydrographic observations in the bottom water of the New York Bight at the "12 Mile" Dump Site: 1983–1990.

Bigelow, H.B. and Schroeder, W.C. (1953) Fishes of the Gulf of Maine. U.S. Fish and Wildlife Service, *Fishery Bulletin* **74,** 577 pp.

Boesch, D.F. (1982) Ecosystem consequences of alterations of benthic community structure and function in the New York Bight region. pp. 543–568 In: Mayer, G.E. (Editor). *Ecological stress and the New York Bight: Science and management.* Estuarine Research Federation, Columbia, South Carolina.

Davis, W. and Walker, H. (1991) Sediment erosion studies. pp. 10–12. In: *Response of the habitat and biota of the Inner New York Bight to abatement of sewage sludge dumping.* Third annual progress report. U.S. Department of Commerce, NOAA, NMFS, NEC, Technical Memorandum, NMFS F/NEC-82, Washington, D.C., 57 pp.

Environmental Processes Division. (1988) A plan for study: Response of the habitat and biota of the Inner New York Bight to abatement of sewage sludge dumping. U.S. Department of Commerce, NOAA, NMFS, NEC, Technical Memorandum, NMFS F/NEC-55, Washington, D.C., 34 pp.

Green, R.H. (1989) Power analysis and practical strategies for environmental monitoring. *Environmental Research* **50,** 195–205.

Gross, M.G. (Editor) (1976) *Middle Atlantic Continental Shelf and the New York Bight,* Special Symposia, Vol. 2, American Society of Limnology and Oceanography, Lawrence, Kansas, 441 pp.

Hurlbert, S.H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54,** 187–211.

Manning, J. (1991) Circulation: Current meter observations. pp. 4–7 In: Studholme, A.L., Ingham, M.C., and Pacheco, A. (Editors). *Response of the habitat and biota of the Inner New York Bight to abatement of sewage sludge dumping:* Third annual progress report. U.S. Department of Commerce, NOAA, NMFS, NEC, Technical Memorandum, NMFS F/NEC-82, Washington, D.C.

Mayer, G.F. (Editor) (1982) *Ecological Stress and the New York Bight: Science and Management.* Estuarine Research Federation, Columbia, South Carolina, 715 pp.

Reid, R.N., Ingham, M.C., and Pearce, J.B. (1987). NOAA's northeast monitoring program (NEMP): A report on the progress of the first five years (1979–1984) and a plan for the future. U.S. Department of Commerce, NOAA, NMFS, NEC, Technical Memorandum, NMFS F//NEC-44, Washington, D.C., 138 pp.

Saila, S.B., Pikanowski, R.A., and Vaughan, D.S. (1976). Optimum allocation strategies for sampling benthos in the New York Bight. *Estuarine and Coastal Marine Science* **4,** 119–128.

Santoro, E. (1987) Status report: phaseout of ocean dumping of sewage sludge in the New York Bight apex. *Marine Pollution Bulletin* **18,** 278–280.

Sokal, R.R., and Rohlf, F.J. (1981) *Biometry,* Second Edition. W.H. Freeman and Company, New York, New York, 859 pp.

Stanford, H.M., and Young, D.R. (1988) Pollution loadings to the New York Bight apex. In: Oceans 88 Conference Record. Marine Technological Society, Washington DC. pp. 745–751.

Stewart-Oaten, A., Murdoch, W.W., and Parker, K.R. (1986) Environmental impact assessment: "Pseudoreplication in time?" *Ecology* **67,** 929–940.

Studholme, A.L., Ingham, M.C., and Pacheco, A. (Editors) (1991) Response of the habitat and biota of the Inner New York Bight to abatement of sewage sludge dumping: Third annual progress report. U.S. Department of Commerce, NOAA, NMFS, NEC, Technical Memorandum, NMFS F/NEC-82, Washington, D.C., 57 pp.

Wilk, S.J., Pikanowski, R.A., Pacheco, A.L., McMillan, D.G., Phelan, B.A., and Stehlik, L.L. (In review) Fish and megainvertebrates collected in the New York Bight apex during the 12-Mile Dump Site Recovery Study, July 1986–September 1989.

Zar, J.H. (1984) *Biostatistical Analysis,* Second Edition. Prentice-Hall, Englewood Cliffs, New Jersey, 718 pp.