APPENDIX A: METHODS FOR AERIAL SURVEYS OF SEALS

The following is an almost-verbatim copy of the METHODS section of the draft final report on the 1999 aerial surveys of ringed seals conducted for BP by LGL. Methods for the planned 2001 surveys will be very similar to the 1999 methods described below. The report from which this section is taken is

> Moulton, V.D. and R.E. Elliott, with T.L. McDonald, G.W. Miller, W.J. Richardson and M.T. Williams. 2000. Fixed-wing aerial surveys of seals near BP's Northstar and Liberty sites, 1999. p. 3-1 to 3-72 *In:* W.J. Richardson and M.T. Williams (eds.), Monitoring of ringed seals during construction of ice roads for BP's Northstar oil development, Alaskan Beaufort Sea, 1999. Rep. from LGL Ltd., King City, Ont., and LGL Alaska Res. Assoc. Inc., Anchorage, AK, for BP Explor. (Alaska) Int., Anchorage, AK, and Nat. Mar. Fish. Serv., Anchorage, AK, and Silver Spring, MD. DRAFT, March 2000.

References to Figures and Tables included in the 1999 report but not in this extract have been deleted. The few notable changes anticipated in 2001 relative to 1999 are identified by *[bracketed editorial inserts]*. References cited in this section are listed at the end of this Appendix.

Survey Design

In 1999, as in 1997-98, two "grids" of aerial survey transects were flown between longitudes 147°06'W and 149°04.5'W, an east-west extent of about 75 km or 40 n.mi. (46 mi). Each grid consisted of 40 north-south transects spaced 1.85 km (1.15 mi, 1 n.mi.) apart. Each transect extended from the Beaufort Sea shoreline to roughly 37 km (23 mi) offshore or to the edge of the land-fast ice if it was encountered and recognizable <37 km offshore *[see Fig. 1 under Task 1]*. One of the grids that we surveyed includes some of the same transects flown by ADF&G during their wider-ranging ringed seal surveys. *[ADF&G's last year of surveys was 1999.]* The second or alternate grid was offset from the first by 0.9 km (0.6 mi) to the east. In this report, we define a *survey replicate* as a complete survey of the 80 unique transects. In 1999, as in 1998, two complete survey replicates were completed. Survey replicate 1 was flown on 4-8 June 1999 and replicate 2 was flown on 8-13 June 1999. In total, 5486 linear kilometers (3409 mi) of surveys were flown over land-fast ice by LGL during the 10-day survey period. The BP/LGL surveys were flown later in the season in 1999 than in 1997-98. However, snowmelt and breakup were late in 1999, and surface conditions were generally consistent with those during the 1997-98 surveys.

A 40-transect grid usually required two days to complete and an 80-transect survey replicate took four days to complete. Ideally, the 20 odd-numbered transect lines from one grid were flown on one day, and the 20 even-numbered lines from that grid were flown on the next day. The odd and even numbered lines from the alternate grid were then (ideally) flown on the third and fourth days. Each day's flight was designed to sample 20 of the 80 distinct transects within the study area, rather than sampling the eastern portions one day and the western portions the next. Thus, the entire study area was to be sampled four times during each replicate survey, and eight times during each year. Fog or low cloud prevented the timely completion of transects during many survey days, and it was often not possible to cover 20 full transects on one day. Sometimes it was necessary to cover parts of two 20-transect groups on the same day, and to complete the coverage of those 20-transect groups on another day. No transects were surveyed on 11 June 1999 because of fog.

The northern ends of repeated transects varied somewhat from day to day. Northbound transects were usually terminated when we had flown at least 37 km (23 mi) or when it was apparent that we had reached

the northern edge of the fast ice. In 1999, the fast ice edge (Fig. 1) was often easy to recognize because of large open leads in the ice.

The southern ends of transects were usually defined by the coastline. However, we sometimes avoided flying over narrow nearshore bands of deteriorated ice. Near the Endicott production facilities, we started or ended some transects 1-2 km (0.6-1.2 mi) north of Howe Island to avoid flying close to bird colonies located there.

Survey Procedures

The 1999 surveys were flown in a DHC-6 Twin Otter operated by Corporate Air of Billings, Montana. This twin-engine high-wing aircraft was equipped with turbo-prop engines, a GPS navigation system, and large bubble windows installed in the emergency exits. Two pilots were present during the entire survey. *[In 2000, a TurboCommander operated by Commander Northwest of Wenatchee, WA, and Anchorage, AK, and specially modified for survey work was used. It is anticipated that the TurboCommander will be used again in 2001.]*

The survey procedures generally followed those of Frost and Lowry (1988). We used strip transect methodology, which has been standard for previous aerial surveys of ringed seals in Alaska. Surveys were usually conducted at an altitude of 91 m (300 ft) above sea level (ASL) and a ground speed of 222 km/h (120 knots). Transect numbers 14-16 and 54-56 near Reindeer Island were flown at altitude 152 m (500 ft) on 6-9 June 1999 to accommodate researchers studying ringed seal haulout behavior in that area. There was concern about radio-tagged seals returning to the water during aircraft overflights at 91 m ASL.

We surveyed transect strips 411 m (1350 ft) in width on each side of the aircraft. These strips extended from 135 m (443 ft) to 546 m (1791 ft) from the centerline. Strip boundaries were marked on the aircraft's windows with tape at the appropriate inclinometer angles, which were 9.5° and 34° below the horizontal for surveys at 91 m altitude and 15.6° and 41.5° for surveys at 152 m altitude. Sightings of seals inside 135 m or beyond 546 m were recorded as off-transect sightings. For consistency with previous ringed seal surveys, we have *not* attempted to adjust the strip boundaries or calculated densities to take account of the "earth curvature" corrections described by Lerczak and Hobbs (1998).

The two primary observers occupied the right seat behind the co-pilot and the left seat behind the pilot. A third observer operated a computerized data logger and recorded polar bear sightings and tracks. This third observer did not record seal sightings and was positioned behind the right observer. The two primary observers sat at large bubble windows that allowed greater downward visibility than standard windows. The surveys were usually flown between 8:00 and 16:00 h true local time (10:00-18:00 Alaska Daylight Time) when numbers of seals hauled out on the ice were expected to be highest. When sightability was severely impaired for more than approximately 5 of the 1-min time periods along a transect, that transect was generally re-surveyed later, and the data from the initial incomplete survey were discarded.

Data Recording Procedures

A GeoLink data logger automatically recorded time and aircraft position (latitude and longitude) at 2-s intervals throughout the flights. The GeoLink system consisted of a portable computer, Garmin GPS unit, and GeoLink data logging software (Version 6.0). At keystrokes initiated by the computer operator, the time and position of the aircraft were automatically logged at the start and end of each transect. Polar bear sightings and tracks were also logged via GeoLink.

The two primary observers recorded the time, visibility (n.mi.), ice cover (%), ice deformation (%), meltwater (%), sunglare (none, moderate, or severe), and overall sightability conditions (ranging from "excellent" to "impossible") onto audio tape at the end of each 1-minute (~3.7 km or 2.3 mi) time period. An electronic timer signaled the observers at 1-min intervals. Ice deformation was estimated by the observers on each side of the aircraft. At the end of each 1-min interval, the observers estimated the percent of the on-transect ice surface surveyed during the preceding minute that was deformed rather than smooth ice. The ice deformation estimates were categorized by intervals of 10%. Cracks and leads in the ice were also noted by the observers at the specific times when seen, allowing their locations to be extracted subsequently from the GeoLink files.

Environmental parameters were recorded by the computer operator (with the assistance of the pilots) at the start of each transect. These variables included cloud cover (in tenths), ceiling height (ft), visibility (n.mi.), wind speed (knots), wind direction (° T), and air temperature (°C).

For each seal sighting, the observer dictated onto audio tape the species, number, habitat (hole or crack), and behavior (look, move, dive, or none) of the seal(s), and noted whether the sighting was on or off transect.

When polar bears were sighted, the observer recorded size/age/sex class when this was determinable, behavior, and direction of movement.

The observers also recorded the time of any sightings of industrial sites or activity, including ice roads or artificial islands.

Analysis Procedures

The location of each seal sighting was determined by matching the time of the sighting with the position recorded for that time in the GeoLink GPS logs. Time periods with severely impaired sightability conditions were excluded from all analyses. Each sighting was also linked to the environmental variables recorded for the corresponding one minute (3.7 km) time period. The fast-ice edge was subjectively located by mapping open leads; areas with leads were classified as pack ice and were excluded from analyses.

Hourly (or more frequent) temperature and wind speed data for Deadhorse airport at Prudhoe Bay were obtained from the National Climatic Data Center (Asheville, NC) for the entire study period. Each one-minute time period was assigned a wind speed and air temperature value by interpolating from the values obtained from the nearest preceding and following airport weather records. Airport data and data collected from the plane were highly correlated in all survey years (Pearson's $r > 0.8$ for air temperature and wind speed). The airport data, with the exception of cloud cover, were used in analyses because they provided finer temporal coverage. Cloud cover data collected from the survey aircraft were used in analyses because it differed from the airport data. From the airport data, an index of wind chill called heat loss was calculated by using the following formula (Siple and Passel 1945):

$$H = (12.1452 + 11.6222\,(v^{1/2}) - 1.16222\,(v))\,(33\text{-}t)$$

where H is the heat loss in Watts/m^2, v is the wind speed in m/s, and t is the temperature in °C. Wind chill in °C was also calculated based on the following formula (Siple and Passel 1945):

$$T_{wc} = 33 + (t\text{-}33)\,(0.474 + 0.454\,(v^{1/2}) - 0.0454\,(v))$$

Weather conditions experienced by seals on the ice undoubtedly varied from weather data collected at the Deadhorse airport. However, we feel the airport data provide a good approximation to "on-ice" weather.

The percent ice deformation data collected at one-minute intervals during all surveys were, for corresponding locations, averaged across days and plotted at the midpoint of the one-minute time period. The averaging procedure involved comparing the GPS coordinates for the midpoints of replicated time periods. If the midpoints were within 800 m of each other, the ice deformation data were averaged. If they were more than 800 m (2625 ft) apart, they were treated as independent values. These data were contoured at 5% intervals using Vertical Mapper for MapInfo (Version 5.0.1). The contoured data were used as a GIS layer showing ice deformation. MapInfo was used to compute the portions of the surveyed area that occurred within the various ice deformation categories. Seal sightings were overlaid on the ice deformation layer, and the numbers of on-transect seal sightings/km^2 and individuals/km^2 were determined for each ice deformation category using MapInfo supplemented by specially written MapBASIC computer code.

In a similar manner, water depth contours were developed based on all available depth soundings. Sounding data, obtained on CD-ROMs from NOAA, included Hydrographic Survey Data, Vol. 1, vers. 3.1, and Marine Geophysical Data/Bathymetry, Magnetics, Gravity, vers. 3.2. The 3-m, 5-m, and additional contours by 5-m intervals out to 45 m were derived using Vertical Mapper for MapInfo. These depth contours were used as a GIS layer. MapInfo was used to calculate the surveyed areas within each contour interval. Seal sightings were overlaid onto the depth GIS layer, and densities for both on-transect sightings and individual seals were calculated.

Five kilometer "bins" of distance as measured from the ice edge shoreward were also plotted and used as a GIS layer. The on-transect surveyed area in each bin was calculated. In the same manner as described above, seal sightings were overlaid onto this layer, and seal sightings/km^2 and individuals/km^2 were calculated for each 5-km interval.

A seal density contour map was created using Vertical Mapper for MapInfo by contouring the ringed seal density (seals/km^2) calculated for each time period segment midpoint. More specifically, the density contours were created from the irregularly spaced midpoints of time period segments by using the inverse distance method (Vertical Mapper) with the following parameters: 1 zone, minimum 1 point, maximum 25 points, 200 m cell size, 18,660 m search and display radius, exponent 1.

Date, time-of-day, and weather effects were analyzed using the 1-minute time periods as the common unit of observation. For example, to compare ringed seal densities with respect to time-of-day, all 1-min time periods surveyed at a particular hour were combined in one bin. The number of on-transect seals was divided by the on-transect area surveyed to calculate the density for that hour.

To investigate potential changes in size of ringed seal groups during the survey period, group size (number of individual seals/number of seal sightings) was calculated for every sighting in 1999 and averaged by date. Group size was further divided by sightings at cracks and those at holes. The percent of the total individual seals (and also sightings) observed hauled out along cracks (vs. holes) in the ice was also calculated for each survey date. These procedures were repeated for data from the 1997-98 surveys to permit interannual comparisons.

For the 1999 data, we examined seal sightings in relation to distance from the Northstar ice road. Ten 1-km "bins" of distance from the edge of the ice road (including the ice dump area north of Seal Island) were plotted and used as a GIS layer. (No seals were observed directly on the ice road.) The on-transect area per bin was calculated and the number of seal sightings and individuals were overlaid onto this layer, which permitted density calculations. The results from replicates 1 and 2 were combined because of the relatively small areas and numbers of sightings involved in this localized analysis. Water depths <3 m were excluded from these calculations.

As part of a multiyear analysis (see "Poisson Regression" section, later) to examine the potential influence of industry, a similar approach was taken for examining seal sightings in relation to distance from Tern Island in 1997 and 1998 and two areas of vibroseis operations in 1998. Data were organized in ten 1-km bins around these industrial areas even during non-industrial years to permit comparison of seal densities between industrial and non-industrial years (while controlling other variables).

Statistical Tests

Univariate Tests

We used the chi-square (χ^2) goodness-of-fit test to assess the significance of observed differences in ringed seal densities with respect to physical (e.g., % deformation), weather (e.g., air temperature), and temporal (e.g., time of day) variables. Simultaneous Bonferonni-corrected 95% confidence intervals were calculated for the observed proportions by strata. An expected proportion (based on available survey area) falling outside the confidence interval for the observed proportion for that stratum was considered significantly different (Manly et al. 1993). All tests were done based on numbers of seal sightings (singletons or groups) rather than numbers of individual seals. The different seals within a closely spaced group are not statistically independent. The expected numbers of seal sightings in the various strata (if seal density were unrelated to the variable in question) were assumed proportional to the surveyed amounts of land-fast ice within those strata. Although the statistical tests were always conducted on the basis of seal sightings (total number of singletons or groups seen), we discuss the results in terms of observed seal densities (individuals/km^2).

Two complete survey replicates were conducted in 1999. All 80 transects were surveyed twice. For comparisons of seal densities with respect to physical factors that did not change during the course of the study, such as water depth, we considered the two survey replicates to be non-independent. At any given location along each of those transects, these variables would be the same during each survey, and some of the same seals may have been seen repeatedly. To avoid pseudoreplication problems associated with the lack of independence of these "repeated measures", we examined each survey replicate (group of 80 unique transects) separately whenever possible. It should be noted that the location of survey lines varied somewhat from replicate to replicate. For the analysis of the relationships of observed seal densities to weather and temporal variables, we pooled the data across replicates. We assumed that numbers of seal sightings at a given location would vary as a result of variation in the temporal and weather factors between replicate surveys. This would make each replicate partially independent with respect to these variables. However, there is still concern about interdependence of results given the presumably fixed number of seals in the area and the close spacing of adjacent transects.

The non-parametric Page's L test (Page 1963) was used to test for progressive seasonal trends in group size (at cracks, holes, and overall) and in percent of total ringed seal sightings that were at cracks in the land-fast ice. We hypothesized that group size and percent of seals observed at cracks would increase during the survey period. These analyses were performed separately for each of the three survey years.

Poisson Regression

Poisson regression models (McCullagh and Nelder 1989; Cameron and Trivedi 1998) were used to assess the relationship between seal counts in small segments of the survey transects and several variables known or expected to influence seal abundance and haul-out behavior. The ultimate objective is to quantify any influence of oil industry activities on the number of seals hauled out, after allowing for natural factors that also influence the number of seals seen. (Additional data from years with more intensive construction

operations will be needed before this analysis can be completed.) The remainder of this subsection includes a technical description of the Poisson regression procedures.

Prior exploratory analysis had revealed that the ringed seal count data exhibited a Poisson distribution. The Poisson distribution is a positive discrete distribution in which only positive integers are acceptable values. Tests based on this distribution were more appropriate for the ringed seal count data than tests assuming a normal distribution, where non-integers and negative values would also be assumed to be permissible. Separate Poisson regression models were fitted to the 1999 data alone and to combined data from the years 1997 to 1999.

The unit of observation in these analyses was normally the segment of a survey transect covered within a 1-min period of observation. This was approximately a 3 km^2 area, i.e. a segment about 3.7 km long (the distance traveled in 1 min) × 0.822 km wide (411 m on each side of the aircraft). However, if environmental conditions remained relatively constant during consecutive 1-min time periods, the data were pooled over these time periods; seal counts and segment areas were summed. This was done to reduce concern about possible lack of independence and to reduce the potential for autocorrelation. To account for the fact that larger survey areas would likely contain more ringed seals than smaller areas, the logarithm of the survey area was fitted as an offset variable in all regression models.

The values for environmental parameters, including percent ice deformation, percent melt water, and visibility, were averaged for combined right and left observer data. Treating data from left and right observers separately would have resulted in pseudoreplication, as environmental conditions were highly correlated between left and right sides of the plane. Also, seal sightings by left and right observers were not always independent of each other, e.g., when ringed seals were counted at an ice crack extending across both observers' fields of view beneath the aircraft. Although the same seals were not counted by both observers (their fields of view were separated by a 270-m-wide strip underneath the plane that was considered off-transect), the seals hauled out at the crack were not entirely independent of each other.

The names of the covariates and factors used in the analyses are listed in Table A-1. All covariates were required in order for a transect segment to be included in the Poisson regression analyses. All variables in the Poisson regression, except survey replicate number and year, were continuous. The survey replicate and year were considered discrete and treated as factors in the models. Quadratic terms were included for the covariates time of day and date to investigate possible non-linear trends. The response variable was the number of observed ringed seal sightings (singletons or groups) in a transect segment, with log (segment area) as an offset variable. The number of seals was not used as the response variable because different seals within a closely-spaced haul-out group should not be treated as statistically independent. Aside from the influences of the various factors and covariates that were analyzed explicitly, the probability of detecting a seal from the survey aircraft was assumed to be constant throughout the study period, and over the length and width of each transect. Data collected during conditions of poor sightability, in water depths < 3 m, and over pack ice were excluded from analyses.

Backward model selection (Rawlings et al. 1998) was employed to derive the final model. Backward model selection is an objective variable selection technique that sequentially eliminates the least significant variables from a candidate model until all variables remaining in the model are significant at the $\alpha = 0.05$ level. Significance of terms in the model was assessed by approximate *F*-tests, which account for overdispersion of the raw data using a quasi-likelihood approach. Overdispersion occurs when the variance of the response variable exceeds its mean. If this occurs and no adjustment is done, test statistics and standard errors will be erroneously inflated (Cameron and Trivedi 1998). Calculations were done with S-Plus Version 2000 (Venables and Ripley 1999).

TABLE A-1. List of variables included in the Poisson regression models of seal sightings in relation to environmental parameters; for 1999 and 1997-99 combined. The response variable in both models was the number of seal sightings.

Data Set		Factors	Covariates
A.	1999	Survey replicate	Water Depth (m) Distance from Ice Edge (km) Ice Deformation (%) Melt Water (%) Time of Day (hr-ADST) Time of Day^2 (hr-ADST) Date Date^2 Air Temperature (°C) Wind Speed (m/s) Heat Loss (W/m^2) Cloud Cover (%) Visibility (n.mi.) Distance from Northstar Ice Roads (1 km bins) [a]
B.	1997-99	Year Survey replicate nested within year	Water Depth (m) Distance from Ice Edge (km) Ice Deformation (%) Melt Water (%) Time of Day (hr-ADST) Time of Day^2 (hr-ADST) Date Date^2 Air Temperature (°C) Wind Speed (m/s) Heat Loss (W/m^2) Cloud Cover (%) Visibility (n.mi.) Distance from Industry (1 km bins) [b]

[a] The distance of an observation from the Northstar ice road in ten 1-km bins (see Fig. 3.2). Observations >10 km from the ice road were coded as "11".

[b] This composite variable accounts for distance to industrial areas in all three survey years. It is composed of "Distance from Northstar Ice Roads" for observations in 1999, the minimum of the distance values for distance from "Tern Island", "Eastern Vibroseis Area", and "Prudhoe Bay Vibroseis Area" for observations in 1998, and "Tern Island" for observations in 1997. See Fig. 3.2 for location of these areas. Distances were recorded as ten 1-km bins from the edge of the industry zone. Observations >10 km from industry were coded as "11" and those within the industrial zone were coded as "0".

The initial model for the backward selection process using 1999 data alone included all variables listed in Table A-1A plus the interaction of survey replicate (1 or 2) with each of the other variables. The initial model for 1997-99 data contained all variables listed in Table A-1B plus the interaction of year and each of the other variables. In the 1997-99 model, survey replicate was nested within year and within all year interactions; however, in the model fitted to these data, the nested effects were not significant and were dropped from further consideration. Main effects were not considered for elimination from the model if they were involved in a statistically-significant interaction.

Due to the potential for temporal correlation among seal counts collected on successive 1-min transect segments, the deviance residuals of all final models were checked for correlation (Pearson's r) within each transect line (T.L. McDonald, WEST Inc., pers. comm.). Average correlation between residuals separated by less than 5 min of flight time (about 18.5 km) was calculated for each model. Moran's I statistic (Moran 1950) was computed on residuals of the final model. Assuming n observations existed in a given 5-min interval of flight, Moran's I computed correlation and standard error among the $n(n-1)/2$ pairs of residuals. Five min intervals were chosen for testing because at a minimum, each 5-min interval contained $5(4)/2 = 10$ pairs of points. Intervals less than 5 min might not contain a sufficient number of pairs for testing using Moran's I. An estimate of overall temporal correlation was computed by averaging correlation estimates across all transects present in the analysis. When this overall correlation estimate was non-significant or negative, temporal correlation was deemed to have an insignificant influence on model estimation.

Model fit was examined by computing the minimum, lower quartile (25th percentile), median, upper quartile (75th percentile), and maximum deviance residual for each model. The absolute value of the lower and upper quartile was compared to 2.0 and, if greater, model fit was further examined for systematic factors producing the large number of high residuals. In addition, deviance residuals were plotted against key environmental variables and examined for trends. These tests revealed that there were no residual quartiles greater than 2.0 in absolute value. No trends were observed when deviance residuals were plotted against key variables. This examination of residuals was deemed to validate all final models.

Results from these analyses include the following: estimates and standard errors of the coefficients, approximate F-values, P-values, overdispersion estimates, and Pearson's r values for temporal correlation. The Tables also show the expected percent increase or decrease in the number of seal sightings (with 95% confidence intervals) for a 1-unit change in the value of each covariate. Degrees of freedom (sample size – number of terms in the model) are also reported for each model.

Literature Cited in Appendix A

Cameron, A.C. and P.K. Trivedi. 1998. Regression analysis of count data. Cambridge Univ. Press, New York, NY. 412 p.

Frost, K.J. and L.F. Lowry. 1988. Effects of industrial activities on ringed seals in Alaska, as indicated by aerial surveys. p. 15-25 *In*: W.M. Sackinger et al. (eds.), Port and Ocean Engineering Under Arctic Conditions. Volume II. Symposium on Noise and Marine Mammals. Geophysical Inst., Univ. Alaska, Fairbanks, Fairbanks, AK. 111 p.

Lerczak, J.A. and R.C. Hobbs. 1998. Calculating sighting distances from angular readings during shipboard, aerial, and shore-based marine mammal surveys. **Mar. Mamm. Sci.** 14(3):590-599.

Manly, B.F.J., L.L. McDonald and D.L. Thomas. 1993. Resource selection by animals. Chapman & Hall, New York, NY. 177 p.

McCullagh, P. and J.A. Nelder. 1989. Generalized linear models, 2nd ed. Chapman & Hall, London, U.K. 511 p.

Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. **Biometrika** 37:17-23.

Page, E.B. 1963. Ordered hypotheses for multiple treatments: a significant test for linear ranks. **J. Am. Stat. Assoc.** 58:216-230.

Rawlings, J.O., S.G. Pantula and D.A. Dickey. 1998. Applied regression analysis: a research tool, 2nd ed. Springer-Verlag, New York, NY. 657 p.

Siple, P.A. and C.F. Passel. 1945. Measurements of dry atmospheric cooling in subfreezing temperatures. **Proc. Am. Phil. Soc.** 89:177-199.

Venables, W.N. and B.D. Ripley. 1999. Modern applied statistics with S-Plus, 3rd ed. Springer-Verlag, New York, NY. 501 p.

APPENDIX B: STATISTICAL POWER OF AERIAL SURVEYS OF BOWHEAD WHALES [6]

This report briefly describes methods used in a power analysis designed to determine the ability of intensive aerial surveys to detect a reduction in number of whales near the Northstar Island (NSI) if there was such a reduction. This Appendix presents methods used to conduct the power analysis, and a summary of the results.

It is assumed that the aerial surveys to be conducted near Northstar during one or more post-construction years would include daily (weather permitting) site-specific aerial surveys of the type conducted in 1996-98 around seismic operations, but in this case centered on Northstar Island. In addition, it is assumed that the broad-scale MMS aerial surveys would continue, and that the MMS data from the area around NSI would also be used in the analysis.

Methods

To evaluate the statistical power of aerial surveys, two areas surrounding NSI were defined. One area was called the *impact* zone and consisted of ocean habitat within a certain distance of NSI. The impact zone was hypothetical for this power analysis exercise and varied from 1 mile to 20 miles in radius. The second area was called the *control* zone and consisted of ocean habitat beyond the impact zone to the boundary of the anticipated aerial survey coverage that might be achievable on a daily basis in 2000-2002. NSI was defined to have an impact on whales when the time trajectories of sighting rates in the impacted and control areas were significantly non-parallel.

To clarify the type of NSI impact that the analyses described in this report detect, consider the situation depicted in Figure B-1 where NSI is assumed to reduce the number of whale sightings within 10 miles. In Figure B-1, average sighting rate estimated from historical data on the control area (outside 10 miles) was 3.5 whales per 1000 km of survey effort prior to NSI. Average sighting rate estimated from historical data on the potentially impacted area was 1.6 whales per 1000 km based on survey effort prior to NSI. Assuming 15 whales per 1000 survey km will be seen in the control area each year after construction, and that NSI produces no effect on whales, the analysis expects average sighting rate in the potentially impacted area to be 13.07 whales per 1000 survey km. This expected number of whale sightings, 13.07, was computed assuming perfect "parallelness" of the two average sighting trajectories displayed in Figure B-1. Expected "parallelness" in the absence of an NSI effect has been marked in Figure B-1 with a triangle. The solid lines displayed in Figure B-1 assume a 50% reduction in sighting rate on the impacted area over what would have been expected under perfect "parallelness". In this case, the 50% reduction equates to a sighting rate of 6.53 whales per 1000 km in the impacted area after construction, i.e. 0.5(13.07).

Hypothetical whale sighting rates in the control and impact areas before and after construction of NSI were compared using a two-sample t-test procedure. The t-test procedure tested for differences among yearly sighting rate differences before and after construction. For example, suppose three years of aerial survey data were available prior to construction and that the impact - control differences were -1.2, 0.5, and 0.2 sightings per 1000 kilometers of survey. Suppose further that a single year of survey data after construc-

---

[6] By **Trent L. McDonald,** Western EcoSystems Technology Inc., 2003 Central Ave., Cheyenne, WY 82001.

tion resulted in a impact - control difference of 1.5. The t-test procedure would test for a difference between the true underlying differences using observed (yearly) differences as a basis for replication. In this example, the estimated mean difference prior to construction was -0.16 while the estimated mean difference after construction was 1.5. All historical data from 1978 through 1998 aerial surveys (both MMS and site-specific) were used to establish reasonable hypothetical sighting rates, historical variation in sighting rate across years, and historical differences across impacted and control areas.

For each set of simulation conditions, 500 random sets of data were generated randomly assuming the number of whale sightings followed a Poisson distribution with mean proportional to survey effort. Effort in past years was fixed at the levels that actually occurred. Effort for post-construction monitoring was assumed to be the average of 3.5 times effort in 1996, 2.5 times effort in 1997, and 1.7 times effort in 1998. (These factors represent the amount of additional effort expected to be achievable with a site-specific aerial survey centered on Northstar relative to the amount of effort obtained near Northstar in 1996-98 when the site-specific surveys were usually not centered specifically on Northstar.) The assumed values for post-construction survey effort appear in Table B-1. For each randomly generated data set, lack of "parallelness" was measured by the two-sample t-test procedure. Expressing the situation in terms of means, rather than differences, the t-test procedure tested the hypothesis,

$$H_o: \mu_{i,b} - \mu_{c,b} - \mu_{i,a} + \mu_{c,a} = 0$$

where $\mu_{i,b}$ was the mean sighting rate on the impacted area before NSI, $\mu_{c,b}$ was the mean sighting rate on the control area prior to NSI, $\mu_{i,a}$ was the mean sighting rate on the impacted area after NSI, and $\mu_{c,a}$ was the mean sighting rate on the control area after NSI. By definition, a significant NSI impact was detected when H_o was rejected in favor of the alternative $\mu_{i,b} - \mu_{c,b} - \mu_{i,a} + \mu_{c,a} \neq 0$. H_o was rejected when the observed t-statistic exceeded the $\alpha = 0.05$ and $\alpha = 0.20$ quantile from a Student's t distribution. Two α levels of significance were used to assess power under "liberal" and "conservative" evidence requirements. The $\alpha = 0.05$ level of significance requires strong evidence in favor of the alternative hypothesis in order to reject the null hypothesis of no effect. The $\alpha = 0.20$ significance level requires less evidence than the $\alpha = 0.05$ level in order to reject the null hypothesis. If one-tailed tests were used to detect impacts, significance levels would then be $\alpha = 0.025$ and $\alpha = 0.10$. Degrees of freedom for the t-test were (number of years of surveys prior to construction) + (number of years after construction) - 2. The proportion of NSI impacts detected out of 500 randomly generated data sets was computed and reported as power.

Power to detect a 0%, 50%, 75%, and 99% reduction in whale sightings on impacted areas was computed for impact radii of 1, 2, 4, 10, and 20 miles. For comparative purposes, sighting rates on the control area were assumed to be 3, 5, 10, 15, and 20 whales per 1000 km of survey effort. (In fact, the average sighting rate in the control area in prior years has been less than 5 regardless of the assumed size of the impact zone — see dashed horizontal lines in Figures B-1 to B-5.) Both one year of post-construction monitoring (2000) and three years of post-construction monitoring (2000 – 2002) were evaluated at the $\alpha = 0.05$ ($\alpha = 0.025$ one-tailed) significance level. The three years post-construction situation was also evaluated using $\alpha = 0.20$ ($\alpha = 0.10$ one-tailed). This latter case assumes the longest post-construction survey effort among all assessed situations, and requires the least amount of evidence before an impact is detected.

Results

Figures B-1 through B-5 graph historical and hypothetical data for impact radii of 10, 1, 2, 4, and 20 statute miles, respectively. Three hypothetical years of post-construction monitoring are shown in the Fig-

ures. Situations involving one year of post-construction monitoring were identical to the three-year post construction situations except that the hypothetical information from 2001 and 2002 was deleted.

Results of the power analysis assuming one year of post-construction monitoring appear in Table B-2. Results assuming three years of post-monitoring data and $\alpha = 0.05$ appear in Table B-3. Results assuming three years of post-monitoring data and the less conservative $\alpha = 0.20$ criterion appear in Table B-4. The half-width of an approximate 95% confidence interval on the values in Tables B-2 through B-4 is 0.033.

In general, power to detect an effect was low given typical whale sighting rates of 3 or 5 sightings per 1000 km of survey effort in the control area. With those typical sighting rates, power to detect an effect was low regardless of the size of the impact area (1-20 mile radius), even in the case of near-total avoidance of the impact area. With typical whale sighting rates, power would be low even with three years of post-construction surveys and an $\alpha = 0.20$ criterion (Table B-4).

For assumed sighting rates of 10 or more sightings per 1000 km of survey effort in the control area, the power of the aerial surveys to detect an effect was higher. In general, power was positively related to the sighting rate, to the size of the impact area, and to the number of years of post-construction surveys (Tables B-2 to B-4). On average, power of three years post-construction monitoring to detect an effect was approximately 84% higher than the power of a single year of post-construction monitoring. With three years of post-construction surveys, reasonable power (0.80 or higher) might be achievable if the impact radius were 10 miles or more, if a high proportion of the bowheads avoided the impact zone, and if the sighting rate (in the control area) were 10 or more sightings per 1000 km of surveys (Tables B-3 to B-4). However, this sighting rate and impact radius are both higher than can reasonably be expected. With realistic sighting rates of 3 – 5 sightings per 1000 km, power would be low even if the impact radius were as much as 10-20 miles and surveys continued for 3 post-construction years.

TABLE B-1. Assumed kilometers of survey effort in the control and impact areas each year post-construction.

Impact Area Size	Assumed Effort in Control Area (km)	Assumed Effort in Impact Area (km)
1 km	28,568	83.4
2 km	28,468	183.7
4 km	28,132	520
10 km	26,067	2585
20 km	19,645	9007

TABLE B-2. Power of **1 year** post-construction aerial surveys to detect varying reductions in whale sightings within various radii impact areas using α = **0.05** and assuming different sighting rates in the control area. For one-tailed tests, significance level requirements would be α = 0.025. Power assessed under simulation using *t*-test for differences of differences. The bound on the simulation error was ±0.033.

Impact Area Size (mi)	Effect Size	Sighting rate in control area (per 1000 km)				
		3	5	10	15	20
1	0	0.006	0.006	0.038	0.058	0.142
	0.5	0.004	0	0.012	0.014	0.154
	0.75	0.01	0	0	0.006	0.206
	0.99	0.008	0	0	0	0.294
2	0	0.004	0.006	0.014	0.038	0.062
	0.5	0.002	0.002	0.012	0.052	0.108
	0.75	0	0	0.004	0.072	0.154
	0.99	0	0	0.008	0.158	0.254
4	0	0	0.002	0.008	0.014	0.036
	0.5	0	0	0.014	0.086	0.16
	0.75	0	0	0.02	0.172	0.33
	0.99	0	0	0.04	0.332	0.472
10	0	0	0	0	0	0.002
	0.5	0	0	0	0.028	0.202
	0.75	0	0	0.01	0.268	0.68
	0.99	0	0	0.076	0.648	0.886
20	0	0	0	0	0	0
	0.5	0	0	0	0	0.034
	0.75	0	0	0.002	0.076	0.682
	0.99	0	0	0.016	0.678	0.996

TABLE B-3. Power of **3 years** post-construction aerial surveys to detect varying reductions in whale sightings within various radii impact areas using and $\alpha = 0.05$ and assuming different sighting rates in the control area. For one-tailed tests, significance level requirements would be $\alpha = 0.025$. Power assessed under simulation using *t*-test for differences of differences. The bound on the simulation error was ±0.033.

Impact Area		Sighting rate in control area (per 1000 km)				
Size (mi)	Effect Size	3	5	10	15	20
1	0	0.004	0.004	0.02	0.05	0.068
	0.5	0.004	0	0.002	0.092	0.154
	0.75	0.002	0	0.008	0.21	0.288
	0.99	0.008	0	0.006	0.366	0.434
2	0	0	0.004	0.036	0.036	0.052
	0.5	0	0.002	0.06	0.088	0.172
	0.75	0	0	0.098	0.242	0.35
	0.99	0	0	0.188	0.332	0.432
4	0	0	0	0.006	0.018	0.034
	0.5	0	0	0.066	0.222	0.372
	0.75	0	0	0.188	0.42	0.468
	0.99	0	0.006	0.374	0.468	0.47
10	0	0	0	0	0	0
	0.5	0	0	0.072	0.514	0.858
	0.75	0	0	0.372	0.876	0.906
	0.99	0	0	0.808	0.912	0.918
20	0	0	0	0	0	0
	0.5	0	0	0.01	0.418	0.97
	0.75	0	0	0.404	0.998	1
	0.99	0	0.008	0.964	1	1

TABLE B-4. Power of **3 years** post-construction aerial surveys to detect varying reductions in whale sightings within various radii impact areas using α = **0.20** and assuming different sighting rates in the control area. For one-tailed tests, significance level requirements would be α = 0.10. Power assessed under simulation using *t*-test for differences of differences. The bound on the simulation error was ±0.033.

Impact Area		Sighting rate in control area (per 1000 km)				
Size (mi)	Effect Size	3	5	10	15	20
1	0	0	0	0.102	0.062	0.116
	0.5	0	0	0.182	0.244	0.348
	0.75	0	0	0.29	0.398	0.428
	0.99	0	0	0.45	0.47	0.548
2	0	0	0.004	0.032	0.04	0.086
	0.5	0	0.012	0.142	0.294	0.366
	0.75	0	0.016	0.256	0.462	0.502
	0.99	0	0.026	0.4	0.542	0.56
4	0	0	0.002	0.044	0.066	0.058
	0.5	0	0.012	0.246	0.442	0.516
	0.75	0	0.022	0.408	0.574	0.718
	0.99	0	0.048	0.49	0.712	0.754
10	0	0	0	0	0.01	0.032
	0.5	0	0.004	0.658	0.912	0.892
	0.75	0	0.058	0.894	0.926	0.92
	0.99	0	0.238	0.916	0.916	0.892
20	0	0	0	0	0	0
	0.5	0	0.002	0.656	0.998	1
	0.75	0	0.12	1	1	1
	0.99	0.002	0.692	1	1	1

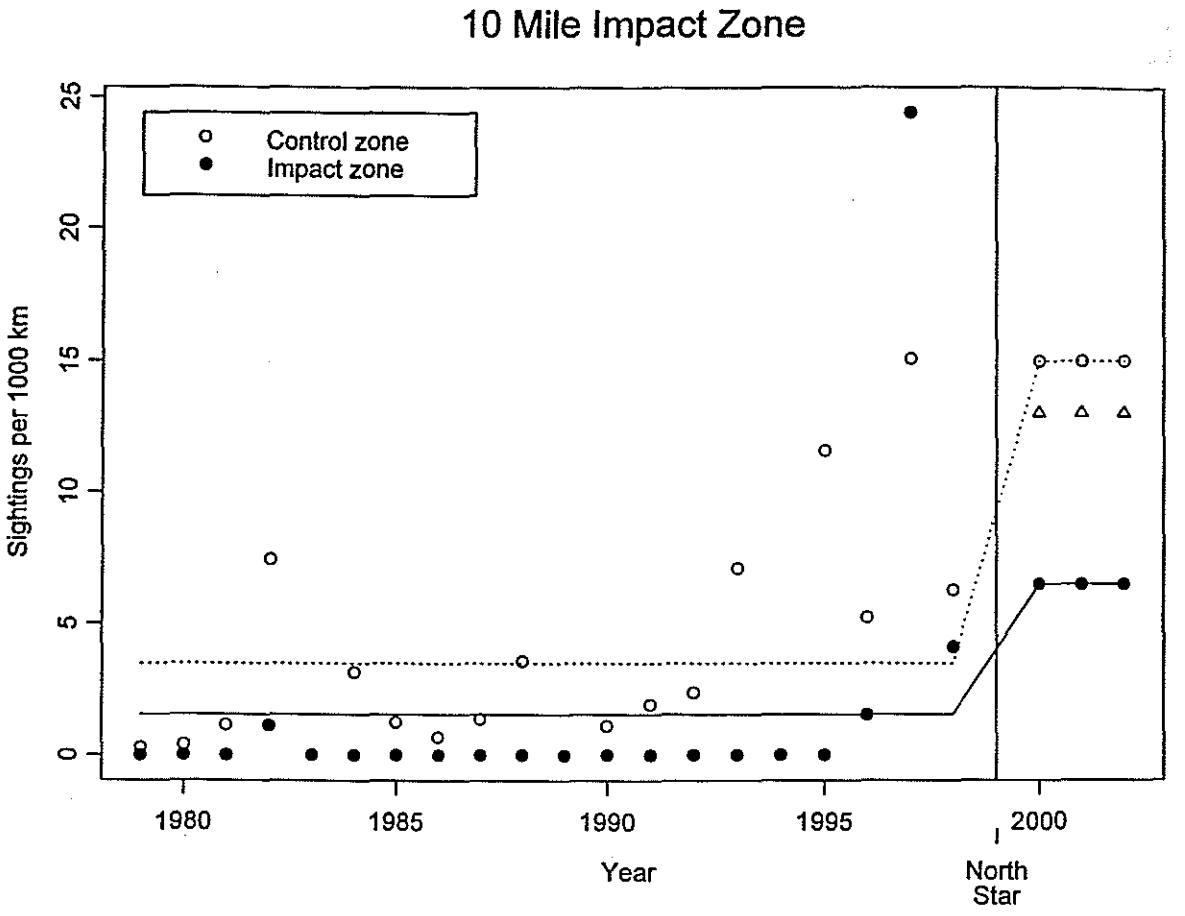

FIGURE B-1. Historical and hypothetical data used to assess power assuming a 10 mile impact zone, 15 sightings per 1000 km in the control area after construction, and assuming a 50% reduction in sightings over what would be expected under the hypothesis of no North Star effect. Filled circles represent sighting rates (per 1000 km) in the potentially impacted area. Hollow circles represent sighting rates on the control area. Expected sighting rates in the impact area under the hypothesis of no effect are shown as triangles.

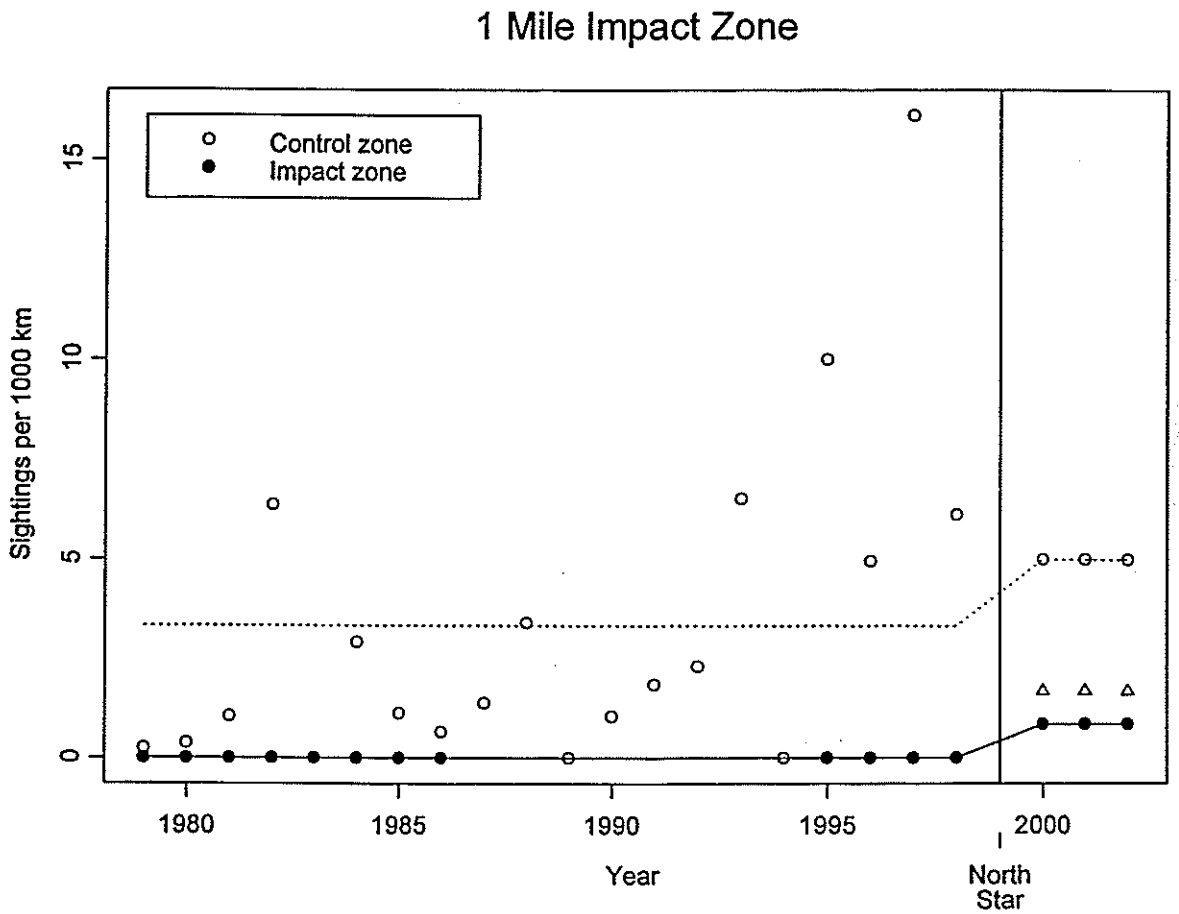

FIGURE B-2. Historical and hypothetical data used to assess power assuming a 1 mile impact zone, five sightings per 1000 km in the control area after construction, and assuming a 50% reduction in sightings over what would be expected under the hypothesis of no North Star effect. Filled circles represent sighting rates (per 1000 km) in the potentially impacted area. Hollow circles represent sighting rates on the control area. Expected sighting rates in the impact area under the hypothesis of no effect are shown as triangles.

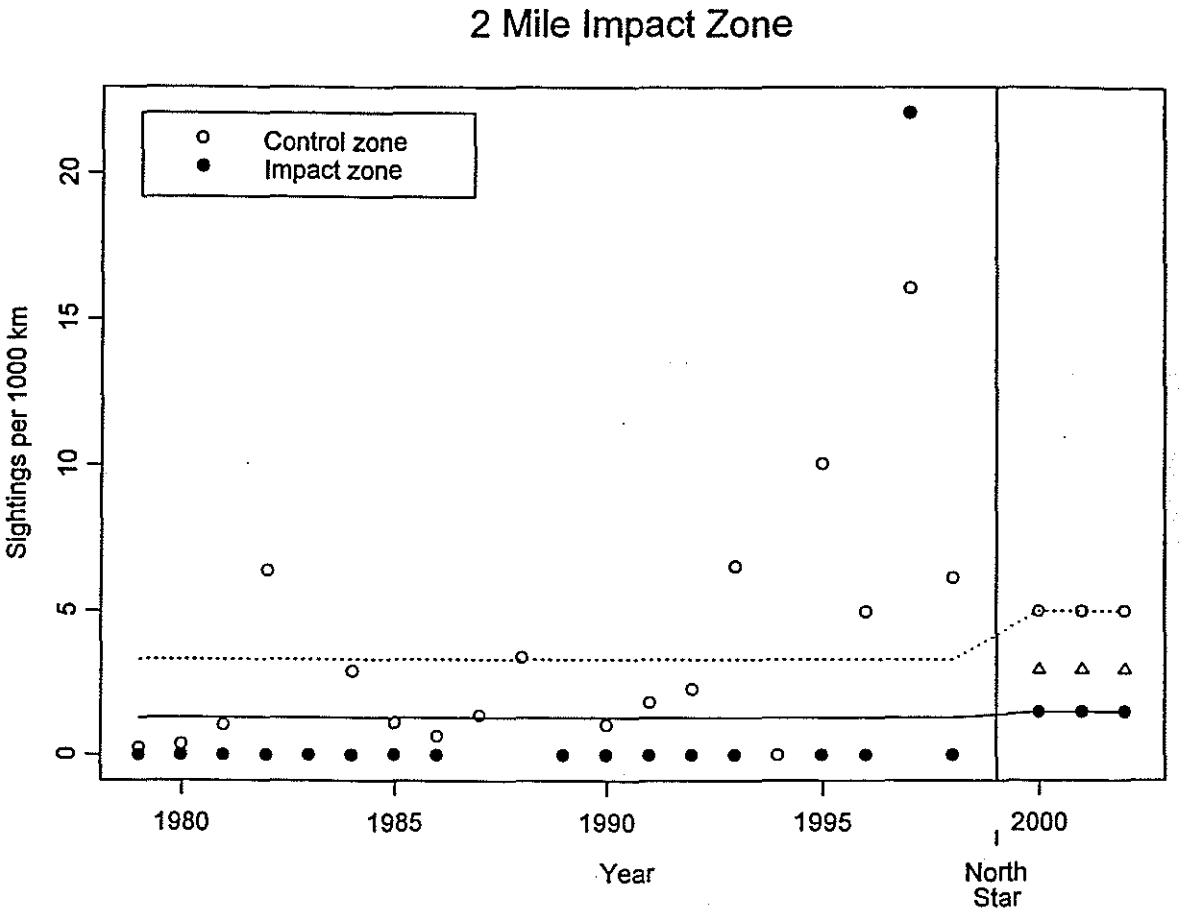

FIGURE B-3: Historical and hypothetical data used to assess power assuming a 2 mile impact zone, five sightings per 1000 km in the control area after construction, and assuming a 50% reduction in sightings over what would be expected under the hypothesis of no North Star effect. Filled circles represent sighting rates (per 1000 km) on the potentially impacted area. Hollow circles represent sighting rates on the control area. Expected sighting rates in the impact area under the hypothesis of no effect are shown as triangles.

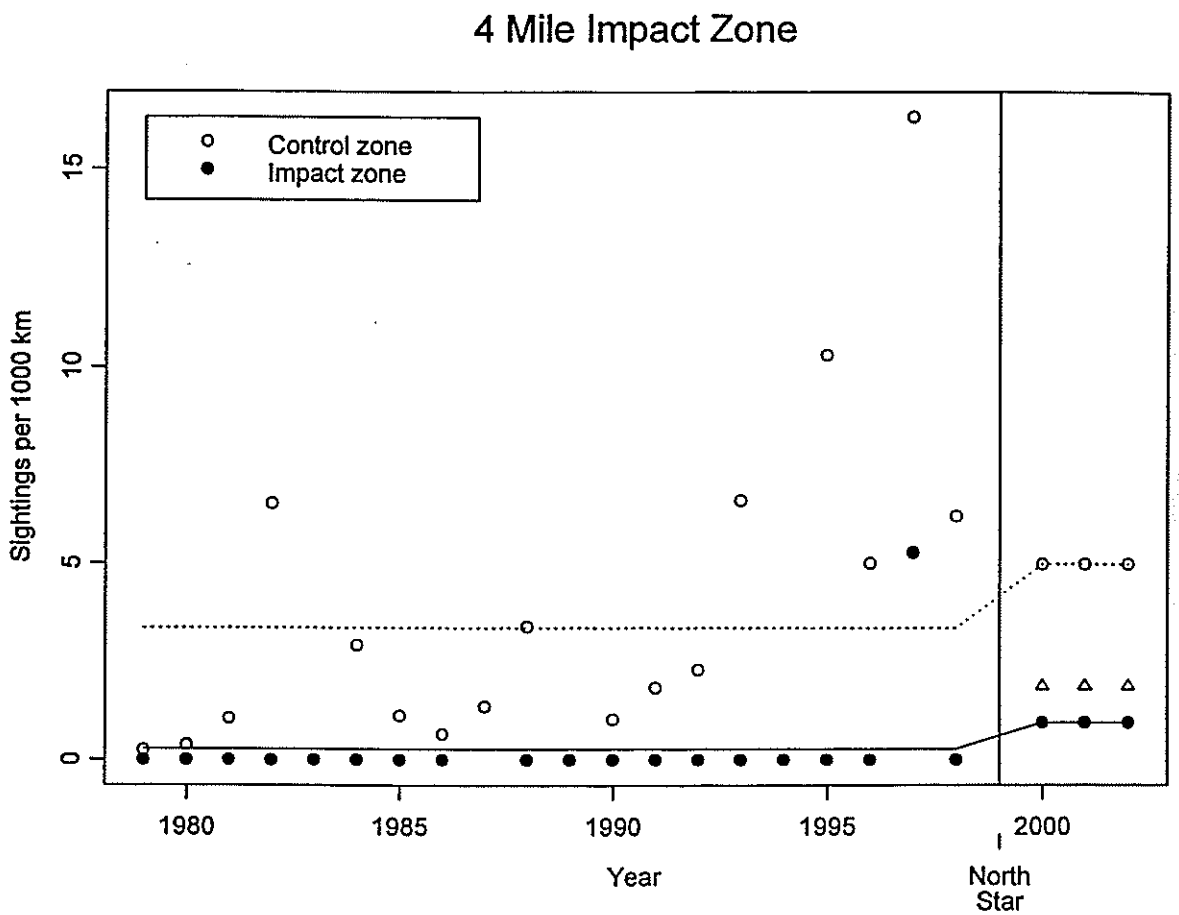

FIGURE B-4: Historical and hypothetical data used to assess power assuming a 4 mile impact zone, five sightings per 1000 km in the control area after construction, and assuming a 50% reduction in sightings over what would be expected under the hypothesis of no North Star effect. Filled circles represent sighting rates (per 1000 km) on the potentially impacted area. Hollow circles represent sighting rates on the control area. Expected sighting rates in the impact area under the hypothesis of no effect are shown as triangles.

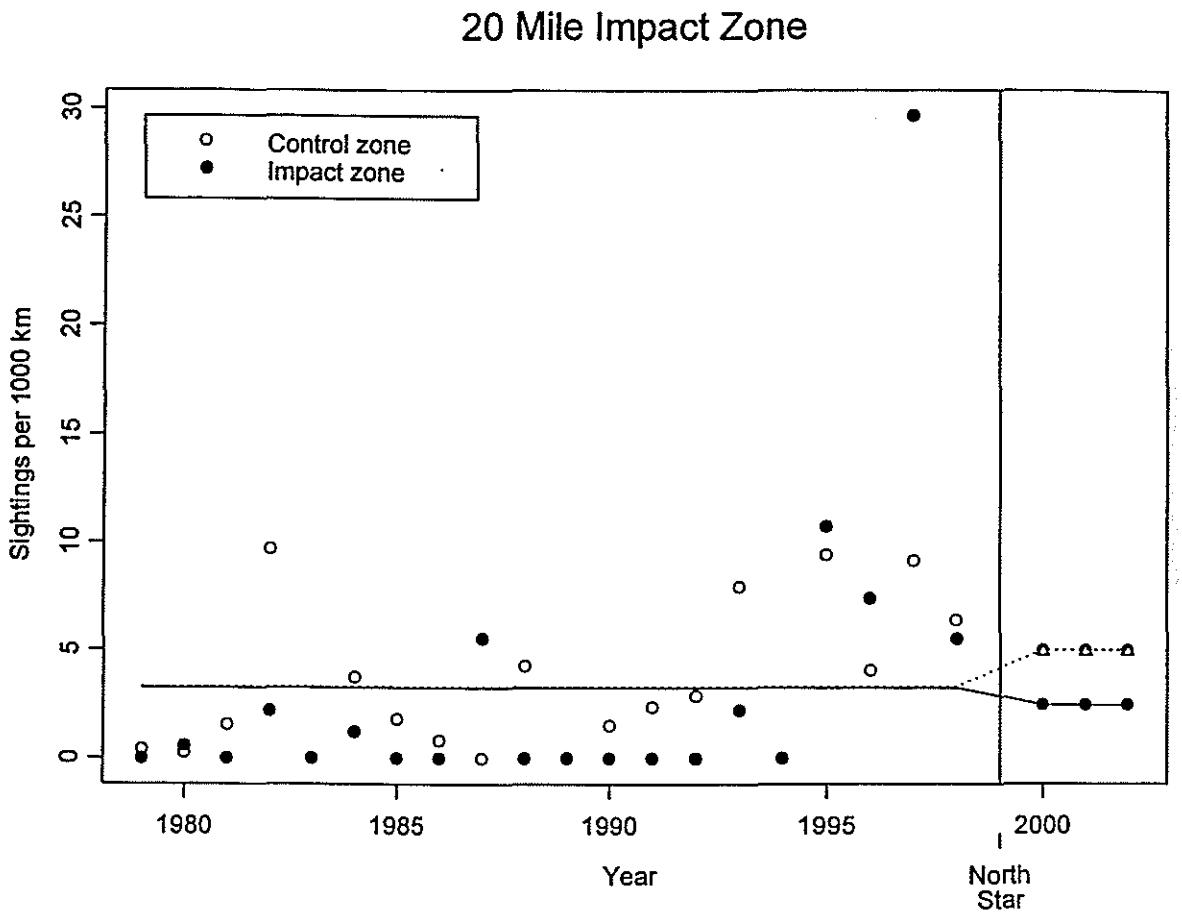

FIGURE B-5. Historical and hypothetical data used to assess power assuming a 20 mile impact zone, five sightings per 1000 km in the control area after construction, and assuming a 50% reduction in sightings over what would be expected under the hypothesis of no North Star effect. Filled circles represent sighting rates (per 1000km) on the potentially impacted area. Hollow circles represent sighting rates on the control area. Expected sighting rates in the impact area under the hypothesis of no effect are shown as triangles.

APPENDIX C: STATISTICAL POWER OF ACOUSTIC SURVEYS FOR CALLING BOWHEAD WHALES [7]

In this report, we describe methods used in a power analysis designed to determine the ability of an acoustic localization technique to detect a reduction in the number of whale calls originating near Northstar Island (NSI) at times when NSI produces substantial underwater noise. We present Methods and a summary of Results, but no discussion is given (see Task 7 section of Monitoring Plan for discussion). The Methods section is organized as follows: First, we discuss a simulation of the expected data, including anticipated distributions of both NSI noise and distance to whales, and a model for the effect of noise on distance. Next, we describe quantile regression as a method of estimating the noise effect, and provide some background on this methodology. Finally, we discuss a power analysis concerning the ability of the proposed methodology to detect and quantify effects of relevant magnitude, including the factors affecting power and our approach to estimating power.

Methods

Data Simulation

Distributions of Noise and Distance to Whales.—In all that follows, we consider noise expressed on a decibel scale, recognizing that this is a logarithmic scale. It is assumed that NSI noise will be measured at a fixed hydrophone located about 500 m offshore of NSI. We assume that the frequency distribution of this noise can be approximated by a gamma distribution. We generated random observations from a gamma distribution with shape parameter 5.6 and scale parameter 2.5, truncated the upper tail of the distribution at 35.5 (less than 0.4% of the distribution was truncated), and translated all remaining observations by adding 86.8. The result was a right-skewed distribution (Fig. C-1) with both mean and median of 100 dB, and range of 88 – 122 dB. The percentiles of our simulated noise distribution closely approximated those from empirical data collected by a fixed hydrophone located 450 m from Sandpiper Island during drilling and other operations in autumn 1985 (data provided by Charles Greene, Greeneridge Sciences Inc.).

Under conditions of no noise from NSI, we assumed that the distribution of distance from NSI to detectable calling whales would be roughly uniform between 17 and 33 km, centered at 25 km, with roughly normal tails below 17 and above 33 km. That is, the probability distribution of whale locations ought to increase as a normal density function between 0 and 17 km, be relatively constant between 17 and 33 km, and then decrease as a normal density function beyond 33 km. This fall-off beyond 33 km offshore is expected because the acoustic monitoring system is designed to cover only the southern portion of the bowhead migration corridor. For calls originating farther offshore, the system will detect and localize a diminishing proportion of the calls at increasing distances beyond 33 km.

We generated such a distribution by simulating observations from a mixture of normal and uniform distributions. Seventy percent of the observations in our mixture distribution came from a normal distribution with mean 25 and standard deviation 6.67, while 30% of the observations came from a uniform

---

[7] By **Chris Nations** and **Trent L. McDonald,** Western EcoSystems Technology Inc., 2003 Central Ave., Cheyenne, WY 82001.

distribution with minimum 17 and maximum 33. We truncated the lower tail of the mixture at 0 km, since a very small percentage (less than 0.01%) of the distribution is negative, and bowheads are not expected to occur inshore of NSI. Figure C-2 depicts 1000 simulated observations of distance to whale calls under the assumption that NSI noise level has no effect on distance from NSI.

Modeling Whale Response to Noise.—Whales might respond to noise from NSI in variety of ways. A simple model of noise avoidance posits that whales will increase their distance from NSI in direct proportion to the noise. At any given noise level, we would expect the effect to be maximum at the noise source and diminish at greater distances from the due to decay in sound level with increasing distance. Irrespective of both the form of the sound decay and the effect of noise on very distant whales, the noise effect should be most apparent on nearer whales. We would be surprised if whales far from a noise source moved more in response to the noise than whales close to the noise.

We simulated the effect of noise by "moving" the whales closest to NSI offshore by a distance proportional to the noise level, so that whales exposed to the minimum noise (88 dB) had no response and whales experiencing the maximum noise (~122 dB) moved the greatest distance offshore. We defined "close" based on the 0.05 quantile line, the line that separates the lower 5% from the upper 95% of the distribution of distance conditioned on noise. If there were truly no noise effect, the quantile line would be perfectly horizontal. For any particular noise effect, say, movement of 0.15 km offshore per 1 dB increase in noise, whales were assumed to move offshore such that 5% of the distribution remained below the line with slope 0.15 km/dB. That is, the "true" 0.05 quantile line would have slope 0.15. For a simulated data set (such as that shown in Figures C-2 and C-3) all observations below the theoretical line with slope 0.15 were moved a distance sufficient to guarantee that 5% of the true distribution remained below the line. This distance, d, was calculated as: d = slope × (Noise Level − 88). As an example, consider a whale close to NSI (below the 0.15 km/dB line). At a noise level of 120 dB, that whale would move $0.15 \times (120 - 88) = 4.8$ km. Another close whale at the median noise level (100 dB) would move 1.8 km. Figure C-3 depicts 1000 simulated observations under the assumptions described in this example.

It might be argued that noise should have some effect on more distant whales in addition to the effect on those closest to NSI. However, our method of assessing the noise effect, quantile regression (see below), is insensitive to changes in whale location for distant whales. That is, because we expect the greatest effect among the nearest 5% of whales, we focus on detecting a shift in location among these nearest 5% of whales. In estimating the 0.05 quantile line, it is immaterial whether the more distant whales move 0, 1, or 10 km. In other words, our method is robust to changes in the shape of the distribution of distances larger than the quantile of interest.

The important consequence of this robustness is that, as long as 5% of observations remain below (south of) the estimated line and 95% remain above it, the line is completely unaffected. For this reason, we determined that it was unnecessary to simulate decreasing detectability of whale calls at large distances from the hydrophone array. Our simulated distribution of large distances may be unrealistic; however, quantile regression for lower quantiles is not affected by details in the upper tail of the distribution.

Estimation

Quantile Regression: Assessing the Relationship between Noise and Distance.—If there is any effect of noise on the distribution of calling whales, we assume that the lower portion of the distribution of distance, i.e., the nearest distances to NSI, ought to shift upward (offshore) as noise increases, as in the example in Figure C-3. We let the 5th percentile represent this lower part of the distribution. Thus, positive slope of the line through the 5th percentile indicates a noise effect, while zero or negative slope is taken as

evidence of no noise effect. Quantile regression (Koenker and Bassett 1978) provides the methodology to assess slope through any percentile of a conditional distribution (here, the distribution of distance offshore is conditional on noise). As with the quantiles of a univariate distribution, quantile regression estimates are less sensitive to assumptions of normality and constant variance than is standard regression through the mean.

Background on Quantile Regression.—Estimation procedures for quantile regression and standard linear regression are similar in some respects. Linear regression estimates, $\hat{\beta}$, may be obtained from minimizing the sum of squared residuals, i.e.,

$$\hat{\beta} = \min_{\beta}\left\{\sum_{i=1}^{n}(y_i - \mathbf{X}_i\beta)^2\right\}$$

where y_i is the observed value of the response variable, $\mathbf{X}_i$ is the vector of explanatory variables, and β is the unknown parameter vector. Likewise, quantile regression estimates may be obtained by minimizing a sum of weighted residuals. If $r_i = (y_i - \mathbf{X}_i\beta)$ are residuals and τ represents the desired quantile with $0 < \tau < 1$, then the regression estimates for the τ quantile are

$$\hat{\beta}_\tau = \min_{\beta}\left\{\sum_{i=1}^{n} r_i[\tau - I(r_i < 0)]\right\}.$$

Here, I(A) represents the indicator function constructed such that all observations below the fitted line receive a constant negative weight of $1-\tau$ and all observations above the line receive a constant positive weight of τ. For example, if $\tau = 0.05$ (the 5th percentile), then observations below the line receive a weight of –0.95 while observations above the line receive a weight of +0.05. Most, if not all, of the literature on estimation for quantile regression focuses on performing this minimization through linear programming (e.g., Buchinsky 1998; Portnoy and Koenker 1997); however, we obtained estimates using widely available nonlinear optimization routines implemented in Matlab (MathWorks 1999). Identical estimates of $\hat{\beta}_\tau$ are obtained by both procedures. The fitted line in Figure C-3(b) represents the regression through the 5th percentile. Significance of $\hat{\beta}_\tau$ was assessed using a randomization test (see below) that did not require the distribution of $\hat{\beta}_\tau$ to be known.

Power Analysis

Factors Affecting Power.—Power to detect an effect will depend on the size of the hypothesized effect, in this case, the slope of the underlying relationship between the 5th-percentile distance offshore and noise. We considered a range of possible effect sizes, simulating data with true slopes of 0, 0.0075, 0.015, 0.03, 0.06, 0.15, and 0.3 km/dB. These slopes for the 5th-percentile distance offshore correspond to a minimum of no noise effect and a maximum increase in 5th-percentile distance of 10.2 km for a noise increase of 34 dB from 88 dB to 122 dB re 1 μPa, as measured a few hundred meters from NSI.

Similarly, we considered sample sizes of 100, 300, 1000, and 3000 localized, independent calls. The largest sample size is a conservative estimate of the total number of calls expected to be localized in a one-month period (Charles Greene, personal communication). However, frequent whale calls may be autocorrelated in time and location. We plan to test for autocorrelation and, if necessary, subsample the available data to obtain independent observations. For example, it may be necessary to take a subsample of every 10th observation to achieve independence; in that case, 300 observations would remain if 3000 calls were

localized in one season. We feel that 100 observations is likely a lower limit on the number of useable observations, given that fewer than 3000 observations may be available and that subsampling may still be necessary. It is also possible that more than 3000 observations will be available. Still, given the estimated power for a sample size of 3000 (see Results), we would expect larger samples to result in very high power to detect even rather small effects.

For testing, we selected $\alpha = 0.10$, a value that requires moderate evidence against the null hypothesis of no effect in order for it to be rejected. Estimated power would have been lower had we used $\alpha = 0.05$.

Estimating Power.—If there were no effect of NSI noise on distance to the closest vocalizing whales, the slope of the quantile regression through the 5th percentile, denoted $\hat{\beta}_{0.05}$, ought to be zero. Conversely, if there is a noise effect, $\hat{\beta}_{0.05}$ ought to be positive. Thus, the null and alternative hypotheses of interest were

H_0: $\beta_{0.05} = 0$ and H_a: $\beta_{0.05} > 0$.

To estimate power, we conducted simulations in two stages, involving primary and secondary data sets. At each of the 24 combinations of sample size and slope, 500 primary data sets were generated. Each primary data set fulfilled the assumptions of the alternative hypothesis, that is, each contained a positive relationship between noise and distance; however, one set of primary data sets was generated based on the assumption of no relationship between noise and distance. For each primary data set, $\hat{\beta}_{0.05}$ was calculated. To calculate a significance level under the null hypothesis of $\hat{\beta}_{0.05} = 0$, a randomization test (Manly 1998) was performed on each primary data set. The randomization test randomly permuted distance values and reassigned these to noise levels. Each random permutation of distance values was called a secondary data set. One hundred secondary data sets were produced for each primary data set. Each secondary data set fulfilled the assumption of the null hypothesis, that is, of no relationship between noise and distance. The proportion of slopes computed on secondary data sets that were greater than the slope calculated from the primary data set provided an estimate of significance.

If the *P*-value computed via randomization was less than α, then the null hypothesis was rejected in favor of the alternative hypothesis. Since the 500 primary data sets were generated under the assumption of the alternative, the number of times the null hypothesis was rejected (out of 500) constituted an estimate of power. That is, power was calculated as the proportion of *P*-values less than α.

Results

As shown in Table C-1 and Figure C-4, power increased with both sample size and effect size (i.e., magnitude of slope). As expected, the probability that the testing procedure led to the conclusion that $\beta_{0.05} > 0$ when in fact $\beta_{0.05} = 0$ (i.e., Type I error) was roughly equal to $\alpha = 0.10$, irrespective of sample size. The power was 0.99 with an expected sample size of 3000 and a slope of 0.15 km/dB, corresponding to a displacement of the 5th-percentile by 4.8 km when received level a few hundred meters from Northstar is 120 dB. For this slope, the power was still high (0.92) for sample size 1000, and 0.65 for sample size 300. With a larger slope of 0.3 km/dB (i.e., a displacement of 9.6 km at 120 dB), the probability of detecting a noise effect was ≥0.95 for sample sizes of 300 or greater.

Other factors that might affect the estimated power to a lesser extent than sample and effect size include (1) distribution of noise (e.g., normal rather than gamma); (2) heteroscedasticity in the conditional

distribution of distance (e.g., increased variation at high noise levels); and (3) nonlinear relationship between noise and distance. We have not examined these other factors.

References

Buchinsky, M. 1998. Recent advances in quantile regression: a practical guideline for empirical research. **J. Human Resour.** 33:88-.

Koenker, R. and G. Bassett, Jr. 1978. Regression quantiles. **Econometrica** 46:33-50.

Manly, B.J.F. 1998. Randomization, Bootstrap and Monte Carlo Methods in Biology (2nd ed.). Chapman and Hall, New York.

MathWorks. 1999. Matlab Function Reference, Volume 1, Version 5.3. The MathWorks, Inc., Natick, MA.

Portnoy, S. and R. Koenker. 1997. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute error estimators. **Statistical Sci.** 12:279-300.

TABLE C-1. Power of acoustic surveys to detect increases in distances to whale vocalizations. The 5th percentile of distance was assumed to be a linear function of noise as measured a few hundred meters offshore of Northstar Island (expressed in dB re 1 μPa, 20-500 Hz band). Power was estimated at 7 different effect sizes (slopes relating distance and noise) and 4 different sample sizes. Significance level was $\alpha = 0.10$ in all cases.

		Sample Size			
Slope (km/dB)	Displacement (km)[1]	100	300	1000	3000
0	0	0.1040	0.1127	0.1134	0.1127
0.0075	0.24	0.1068	0.1258	0.1182	0.1408
0.015	0.48	0.1168	0.1165	0.1014	0.1469
0.03	0.96	0.1095	0.1174	0.1956	0.3448
0.06	1.92	0.0949	0.1956	0.3567	0.6840
0.15	4.80	0.3783	0.6538	0.9175	0.9940
0.30	9.60	0.7699	0.9479	1.0000	1.0000

[1] Projected offshore displacement of 5th-percentile calling whale when received noise level a few hundred meters from Northstar Island is 120 dB re 1 μPa in the 20-500 Hz band.

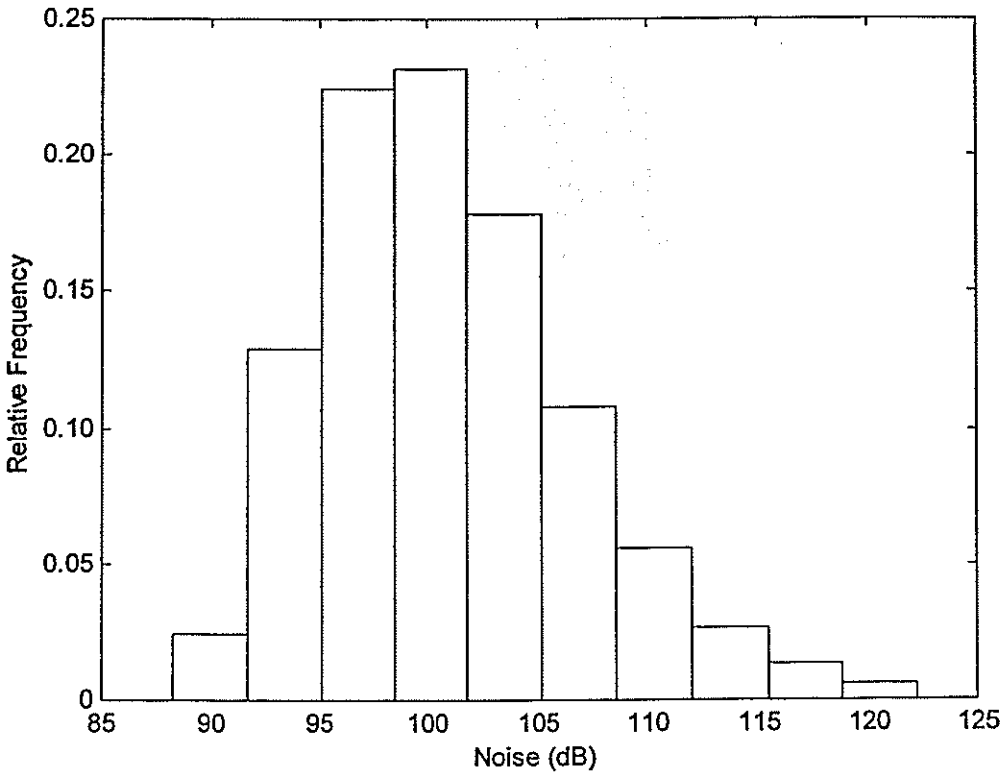

FIGURE C-1. Empirical distribution of noise (in dB re 1 μPa, 20-500 Hz band), based on simulation of 10,000 simulated observations from a gamma (5.6, 2.5) distribution, upper tail truncated at 35.5, and transformed by adding 86.8 to all values.

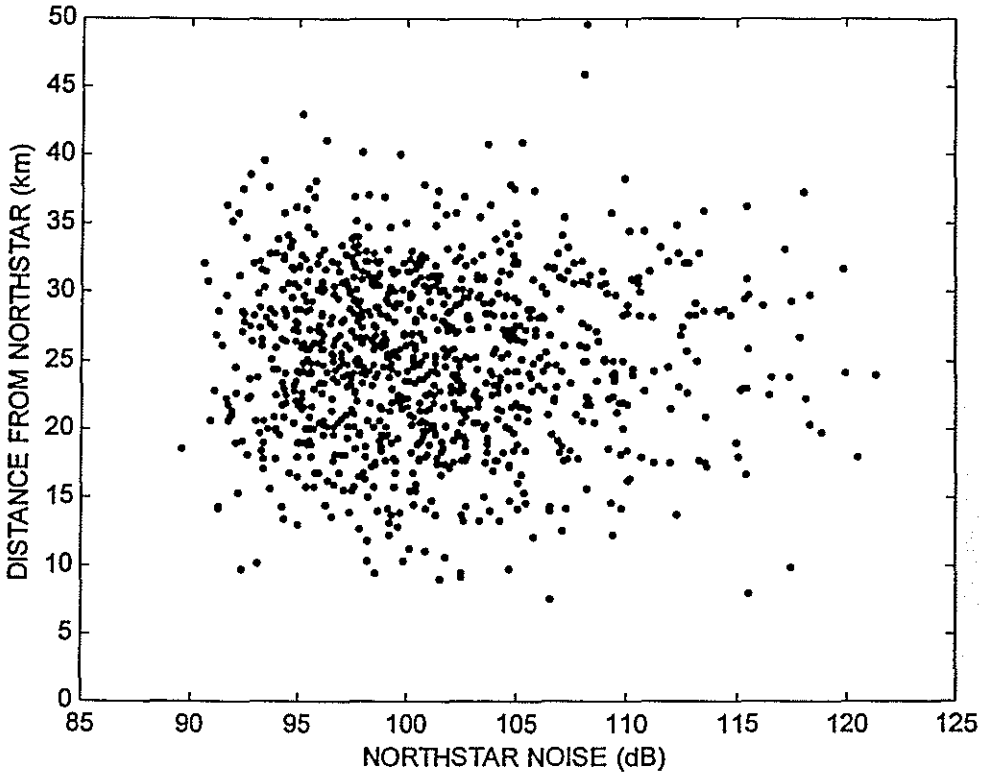

FIGURE C-2. Simulation of 1000 localizations of calling whales under the null hypothesis of no relationship between Northstar noise and distance from Northstar to vocalizing whales. Noise was distributed as a transformed gamma with a mean of approximately 100 dB re 1 μPa in the 20-500 Hz band (as shown in Fig. C-1). Distance was distributed as a mixture of normal and uniform distributions (see text for details), with a mean of 25 km.

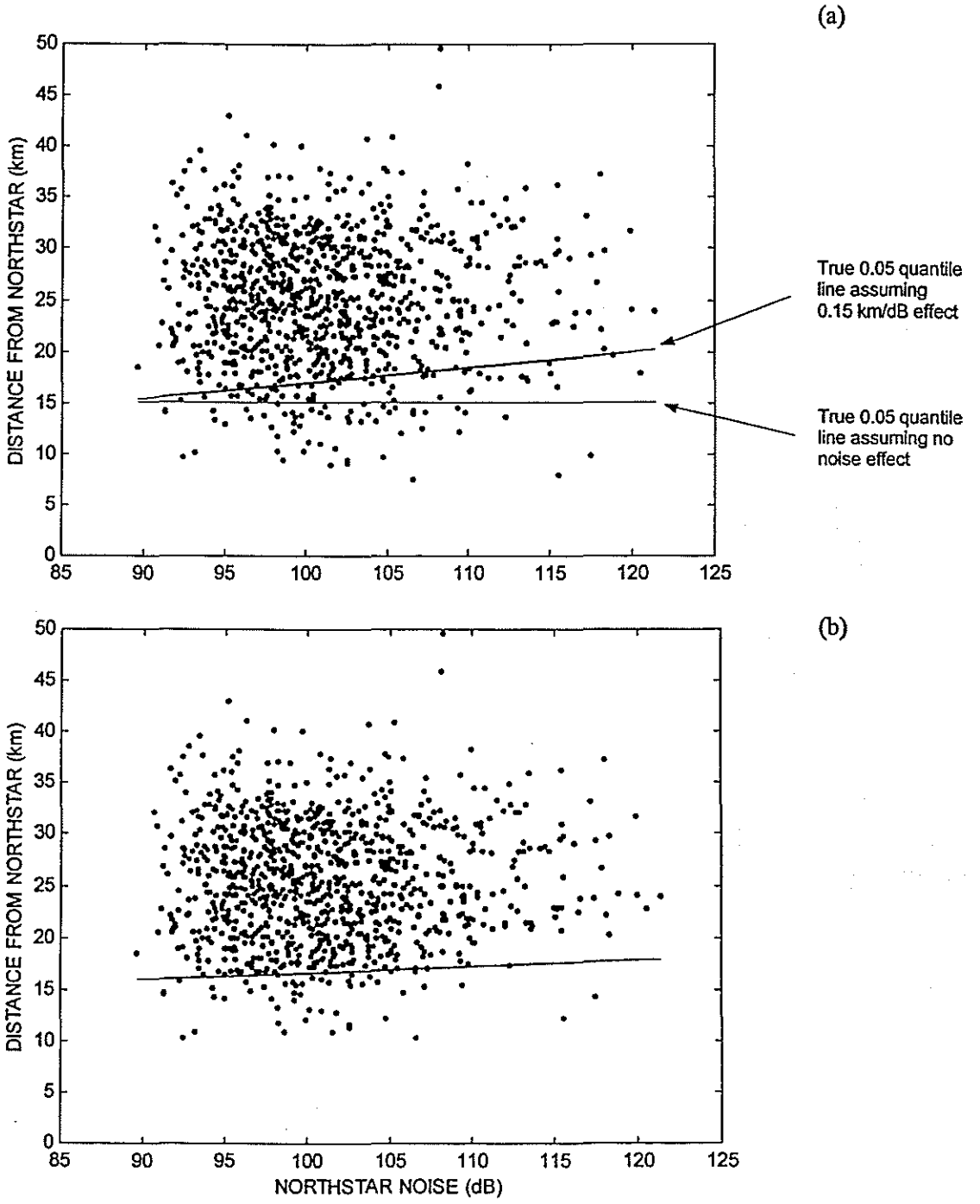

FIGURE C-3. Simulation and subsequent estimation of noise effect. (a) Same data as shown in Figure C-2. Horizontal line represents true 0.05 quantile line for no noise effect. Upper line with positive slope represents a hypothetical effect size of 0.15 km/dB (i.e., 4.8 km farther offshore for 120 dB than for 88 dB). (b) Same data, but with observations below the upper line moved offshore based on the 0.15 dB/km assumption (see text). The line in (b) is the estimated 0.05 quantile line for the altered distribution.

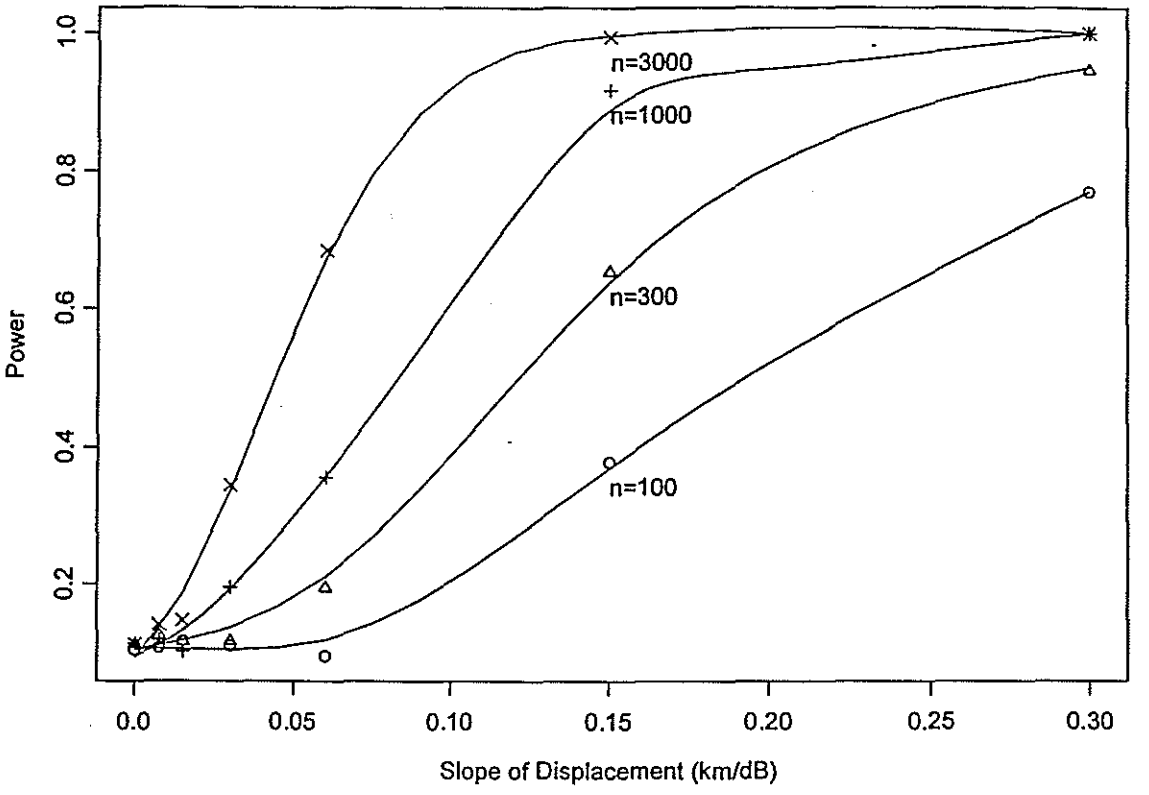

FIGURE C-4. Estimated power as a function of both sample size and magnitude of the slope relating 5th percentile distance offshore and noise. Symbols represent data contained in Table 1. Lines represent a smoothed estimate of the power curve for each sample size.