Exxon Valdez Oil Spill
Restoration Project Final Report

Evaluation of Alaska Harbor Seal *(Phoca vitulina)* population surveys: A simulation study

Restoration Project 00509
Final Report

M.D. Adkinson
T.J. Quinn II
Juneau Center
School of Fisheries and Ocean Sciences
11120 Glacier Hwy.
Juneau, Alaska 99801

R.J. Small
Alaska Department of Fish & Game
1255 West 8th Street
Juneau, Alaska 99802

January 2001

The Exxon Valdez Oil Spill Trustee Council conducts all programs and activities free from discrimination, consistent with the Americans with Disabilities Act. This publication is available in alternative formats upon request. Please contact the Restoration Office to make any necessary arrangements. Any person who believes she or he has been discriminated against should write to: EVOS Trustee Council, 645 G Street, Suite 401, Anchorage, Alaska 99501; or O.E.O. US Department of the Interior, Washington, D.C. 20240.

Study History: The objective of this one-year project was to evaluate current monitoring programs for harbor seals in Alaska. Specifically, the evaluation examined program sampling design, accuracy and precision, and application to the management and conservation needs of harbor seals. This evaluation and review used data and information from three primary sources: (1) Results of Trustee Council restoration project 064: Monitoring, Habitat Use, and Trophic Interactions of Harbor Seals in Prince William Sound, Alaska; (2) Annual reports of the ADF&G statewide harbor seal research program, and (3) NMFS reports on annual population abundance surveys of harbor seals in Alaska.

Abstract: We used simulation to investigate robust designs for the Alaska harbor seal population surveys. We employed an operating model approach, creating simulated harbor seal population dynamics and haulout behavior that incorporated all factors thought to potentially affect the performance of aerial surveys. We found that adjusting counts for the effects of covariates such as survey date or stage of tide was both possible and essential. Annual estimates of the average fraction of the population hauled out are also advised.

Key words: harbor seal, *Phoca vitulina*, survey design, aerial count, Gulf of Alaska, simulation.

Project Data: This project is primarily a re-analysis of existing empirical data collected by the ADF&G and NMFS during studies to assess the population trend and abundance of harbor seals in the Gulf of Alaska, including Prince William Sound. These data are primarily counts of harbor seals obtained from both aerial surveys and land-based monitoring studies. Additional data include the geographic location of the seal haulout where the counts were obtained, and records of environmental variables associated with the count data: e.g., date, time, tidal stage. The format of these data is an EXCEL spreadsheet. The primary custodians for these data are: (1) Kathy Frost, 1550 Coyote Trail Fairbanks, AK, 99709-6009; 455-6885; kjfrost@eagle.ptialaska.net; (2) Robert J. Small, ADF&G, 1255 West 8th Street, Juneau, AK, 99802-2256; 465-6167; bob_small@fishgame.state.ak.us; (3) Dave Withrow, NMFS, 7600 Sand Point Way NE, Seattle, WA 98115-0070; 206-526-4019; Dave.Withrow@noaa.gov

Citation: Adkison, M. D., T. J. Quinn II, and R. J. Small. 2001. Evaluation of Alaska Harbor Seal (*Phoca vitulina*) population surveys: A simulation study, *Exxon Valdez* Oil Spill Restoration Project Final Report (Restoration Project 00509), Alaska Department of Fish & Game, Division of Wildlife Conservation, Juneau, Alaska.

RE-DESIGN OF THE ALASKA HARBOR SEAL (*PHOCA VITULINA*) SURVEY: A SIMULATION STUDY

Milo D. Adkison, Terrance J. Quinn II

Juneau Center, School of Fisheries and Ocean Sciences
11120 Glacier Hwy, Juneau AK 99801

Robert J. Small
Alaska Department of Fish & Game
1255 West 8th Street
Juneau, Alaska 99802

ABSTRACT

We used simulation to investigate robust designs for population surveys of Alaska harbor seals. We employed an operating model approach, creating simulated harbor seal population dynamics and haulout behavior that incorporated all factors thought to potentially affect the performance of aerial surveys. We found that adjusting counts for the effects of covariates such as survey date or stage of tide was both possible and essential. Annual estimates of the average fraction of the population hauled out are also advised. A key factor governing the robustness and power of harbor seal population surveys is inter-site variability in trend. This factor is well understood for Prince William Sound and Kodiak trend sites, but better information is needed for other sites in the *Exxon Valdez* Oil Spill area.

Key words: harbor seal, *Phoca vitulina*, survey design, aerial count, Gulf of Alaska, simulation.

Harbor seal population abundance and trends in abundance are primarily assessed using aerial counts obtained during the August molt period. Photographs of all seals at a haulout are taken from an aircraft, and precise counts are then obtained when the image is later scrutinized in the lab. Logistical and weather related factors preclude surveys being conducted under constant environmental conditions, most of which vary asynchronously, thereby resulting in difficulties inferring harbor seal abundance and population trends from aerial counts. For example, not all seals haul out at any one time, and the number hauled out varies by date, time of day, stage of the tide, etc. (Frost et al. 1999, Small et al. 1999).

In order to monitor trends in seal populations, the Alaska Department of Fish & Game (ADF&G) annually counts seals on a consistent set of trend sites within several distinct survey areas (Frost et al. 1999; Small et al. 1999). They have used sophisticated statistical methodologies to standardize these counts to common units by estimating the effects of covariates such as date, stage of tide, time of day, etc. on the number of animals seen (Frost et al. 1999; Small et al. 1999; Jay ver Hoef personal communication). This attempt at standardization may suffer from being based on observational rather than experimental data. In at least one instance, counts had been made progressively earlier in the year across a 5 year period. This could lead to potentially confounding the effect of date on the fraction of the population hauled out with the underlying true trend in the population (Small et al. 1998). ADF&G does not attempt to convert their standardized counts to an estimate of absolute abundance.

The National Marine Fisheries Service (NMFS), mandated by the Marine Mammal Protection Act (MMPA) to estimate absolute abundance, periodically obtains aerial counts of seals over their entire range in Alaska. Both the maximum and mean count for the sites are calculated and summed across sites, and mark-recapture methods are applied to presence/absence data obtained from VHF radio tagged animals to estimate the fraction in the water (Huber 1995; Withrow and Loughlin 1995, 1996b). Based on this fraction, NMFS has applied a single 'correction factor' to their mean count to estimate absolute abundance. Covariate data has been collected during surveys, and methodologies for incorporating this information into population estimates are being investigated.

The correction factor that should be applied to estimate absolute abundance necessarily varies with the state of the covariates that affect the fraction of the population hauled out (Figure 1). In addition, other concerns about the harbor seal surveys have arisen in recent years. In the Prince William Sound (PWS) region, the reported decline has been disputed, based on the premise that seals have moved to locations outside of the survey area, and that the current survey sites are not representative of the entire PWS area. Problematically, the areas that the seals have purportedly moved to contain glacially dominated fiords, where it is suspected that both haulout behavior and visibility of the seals may differ (Hoover 1983, Mathews and Kelly 1996). Recent evidence that seals separated by relatively short distances may exhibit significant genetic differences (Westlake et al. personal communication) suggests that different population trends might occur over small spatial scales. The timing of the molt period may also vary across years (Jemison and Kelly 2001), resulting in temporal variation in haulout attendance and thus the number of seals counted during surveys. A more methodological concern is the disappearance of animals from several haulouts on aerial survey routes; whether these sites should be retained or replaced is not obvious.

In what follows, we examine harbor seal population surveys in that portion of the Gulf of Alaska affected by the *Exxon Valdez* oil spill (EVOS) using computer simulation. The simulation approach is the only way to examine alternative survey designs and evaluate the effects of animal movement, variation in molt period timing, inter-site variability in trends, and other factors thought to affect the surveys and harbor seal populations. We compare the robustness of alternative survey designs to the biological phenomena that might affect our ability to estimate population trends. Based on these analyses, we provide guidance on how to improve existing survey practice.

METHODS

Operating model approach

The operating model approach (also known as operational management procedure) is an emerging technology that is becoming widely used in evaluating the robustness of natural resource management strategies. Pioneered by the Scientific Committee of the

International Whaling Commission (Kirkwood 1997), it has found use in fisheries management, particularly in South Africa (Butterworth and Bergh 1993, Butterworth et al. 1997, Cochran et al. 1998). The goal is to find a management procedure that achieves resource conservation objectives by conducting large-scale simulations over a wide variety of operating conditions and hypotheses.

In our context, the management strategy is the design of the Gulf of Alaska harbor seal population abundance and trend surveys conducted by NMFS and ADF&G, respectively. We use the operating model approach to provide a framework, using a known population, for evaluating the robustness of alternative survey designs. By robustness, we mean designs that produce reasonable estimates (low bias and variability) even when assumptions are not strictly met. The operating model contains all the known and hypothesized factors affecting the population and the survey.

The operating model contains an estimation module that translates the data into estimates of key quantities. Often a simplified estimation model replaces the usual one, in order that large numbers of simulations can be performed.

The inputs to the operational model consist of relevant information about the Gulf of Alaska harbor seal surveys, as well as basic principles of sampling and survey design. We obtained raw data from the survey, gathered literature results and values, and attended meetings with marine mammal scientists to obtain expert opinions. We found that the Bayesian hierarchical model of Ver Hoef and Frost (1999, 2000) provided a logical starting point for development of our model. Knowing that this model would be too complex to use for large-scale simulations, we reanalyzed the raw count data and came up with a general linear model formulation (described below) to approximate results from the Bayesian model.

Reanalysis of Prince William Sound trend route 'A' sites

We performed individual analyses for each of the 25 sites that comprise the ADF&G PWS 'A' trend survey route using a general linear model formulation similar to Ver Hoef and Frost (1999). The dependent variable was ln(count+1). The model contained a categorical variable for year and covariates for date, time of day, tide, and tide height (with quadratic terms for the first three covariates). Because tide height and time of day were the least significant variables, we dropped those variables and reran a reduced general linear model. Results were fairly similar to Ver Hoef and Frost (1999), so we used parameters from the reduced model in subsequent work.

The estimated deviation of the trend at each site from the average trend is shown in Figure 2. There is large inter-site variability in trend, with deviations ranging from –0.28 to 0.22. These deviations are a key ingredient in the operating model.

The estimated effect of day on ln(count+1) is shown for each site and averaged across sites (Figure 3). On this scale the estimated effects range generally from –1 to 1 from 10 days before the mean survey day of August 28 to 20 days after that day (corresponding to the range in the data). The effects also range from minimal (flat) to concave downward (the expected effect) to concave upwards (somewhat counterintuitive). Averaged across sites, the overall effect is concave downward. The peak cannot be seen on the graph but occurs about 15 days before August 28.

The estimated effect of tide on ln(count+1) is shown for each site and averaged across sites (Figure 4). On this scale the estimated effects range generally from –2 to 1 from 3 hr before low tide to 3 hr after low tide (corresponding to the range in the data). The effects range mainly from minimal (flat) to concave downward (the expected effect). Averaged across sites, the overall effect is concave downward with a peak about low tide.

The individual site parameters (linear and quadratic coefficients) for day and tide are used in the operating model. The average across sites in the above graphs can be used to predict the average impact of each factor on the counts.

Spatial structure

The model was multi-year (12 years) to correspond to the longest length of survey trend data available for the Gulf of Alaska Region. The model was multi-area (PWS, Kodiak, AK Peninsula, Cook Inlet, Kenai, Figure 5) to correspond to the 5 regions of interest in the EVOS spill area. All potential sites within the area were classified by substrate, as either ice or non-ice. Sites were identified from the 1996 NMFS survey and the ADF&G Kodiak and PWS trend routes, with some pooling of sites in close proximity, resulting in a total of 244 sites (Figure 5).

The sites were sorted, mainly by decreasing longitude [west to east], and classified into 8 survey sub-areas within the 5 major areas (Table 1).

Table 1 Survey sub-areas

Survey sub-area	Main area	Number of sites
Survey 1	Area 1(AK Peninsula):	53
Survey 2	Area 2 (Kodiak):	37
Survey 3	Area 2 (Kodiak):	35
Survey 4	Area 3 (Cook Inlet):	34
Survey 5	Area 4 (Kenai):	12
Survey 6	Area 5 (PWS A stations):	19
Survey 7	Area 5 (PWS B stations):	27
Survey 8	Area 5 (PWS C+Copper River):	27

The survey sub-areas were chosen to approximate actual practice during the 1996 survey. With the exception of survey 8, these survey routes correspond to entire areas or sub-divisions of areas as defined for population dynamics (see below).

Population dynamics

The model follows the abundance of harbor seals at all sites over the course of the simulation period. Because the survey design was the key factor in this study, we kept the model's population dynamics simple and deterministic. We define abundance as the count of animals that would occur on the date of peak attendance during the molt period at low tide with no unknown covariate present. By doing so, we avoid the problems of modeling the calibration between counts and abundance, such as occurs in the calculation of the minimum abundance estimate.

1. Movement- Using all sites, a matrix Θ of movement proportions is constructed: $\theta_{j\rightarrow i}$ is the probability that an animal at site j will be at site i the next year, so that the "from" region is in columns and the "to" region is in rows. For simplicity it is assumed that movement declines linearly with distance between sites at rate z_m up to a distance d_m, so that $\theta_{j\rightarrow i} = z_m\,(1\text{-}d/\,d_m)$, where d is the distance between sites i and j. Since movement among sites is thought to be small, the default setting is no movement among sites. Movement among sites is incorporated in studies of differing substrate preferences (see below).

2. Time trend- A hierarchical model is constructed for the trend parameter τ_1 at each site, which is defined as the log ratio of abundance in two successive years:

$$\tau_1 = \ln(N_{t+1} / N_t) = \mu + A_i + L_k \text{, where}$$

$$\mu = \text{overall trend}$$

$$A_i = \text{adjustment for area}$$

$$L_k = \text{site - specific adjustment (like an error term).}$$

We assumed that substrate does not affect trends in abundance; rather it affects behavior of the seals through movement or attendance parameters. Note that the value of τ_1 is negative when the population at a site is decreasing. The overall trend (μ) and the adjustments for each of the 5 areas (A_i) are input by the user to the simulation model. We obtained site adjustments (L_k) from reanalysis of the 25 sites in the PWS 'A' trend route, by assuming that inter-site variability in trend from those sites is the same for all sites in the study area. The site adjustments of all 244 locations in the simulation were then sampled with replacement from the 25 deviations in PWS site trend.

3. Initial conditions- We used the 1996 NMFS database (Withrow and Loughlin 1997) to set initial conditions for abundance in the first year. If the NMFS count was < 10 in 1996, we replaced it with 10, so that all sites would have some animals at the start.

*4. Population abundance-*Abundance at the start of year t is written in vector notation as N_t. It is assumed that movement occurs immediately after the beginning of the year, so that abundance after movement is $\tilde{\mathbf{N}}_t = \Theta\,\mathbf{N}_t$. The population trend is accounted for thereafter. For each site i, abundance after movement is written $N_{i,t+1} = \tilde{N}_{i,t}\exp(\tau_{1i})$, where τ_{1i} is the site-specific trend.

Attendance at haul-out sites

*1. Factors-*In the model, we wish to include those factors that may affect the number of seals present at a site (attendance). These include survey date, time of day, time relative to low tide, and tide height (Ver Hoef and Frost, 2000; Small and Pendleton, 2000). Survey date is one of the main survey design features and must be included. For simplicity, we subsume the remaining factors into a single factor, which is known. In addition, we want an additional factor in the model that is unknown to the researchers to see how covariate adjustments can be negatively impacted by an unknown and unmeasured factor. Therefore, we assume that attendance is related to:

a) variable x: day relative to the mean PWS survey day of August 28,
b) variable y, a known covariate: a variable that affects attendance within a day, which we took from the variable "time relative to low tide" in the Ver Hoef and Frost analysis and our reanalysis,
c) variable z, an unknown covariate: a lurking variable not related to low tide timing, and
d) random events.

Day and tide are modeled as quadratic effects, and the unknown covariate is modeled as a linear effect. The expected overall effect E_i of the these three variables for each site i is loglinear:

$$E_i = \exp(\beta_{1i}x_i + \beta_{2i}x_i^2 + \beta_{3i}y_i + \beta_{4i}y_i^2 + \beta_{5i}z_i)$$

[labeling differs from Ver Hoef and Frost]. Differing effects by substrate are included here when they are present in the scenario being examined.

The residual variability and day-to-day autocorrelation were estimated from 6 years of almost daily counts taken at Tugidak Island in the Gulf of Alaska (Jemison et al. 1998). To approximate the deterministic biological processes affecting the magnitude of the counts, counts were modeled as the sum of two bell-shaped curves. The first curve represented the pupping period, and the second the molt period. The shape of these curves was a normal distribution multiplied by a scalar.

$$C_{y,t} = H_{1,y} \exp\left(- \frac{(t - \mu_{1,y})^2}{2\sigma^2} \right) + H_{2,y} \exp\left(- \frac{(t - \mu_{2,y})^2}{2\sigma^2} \right)$$

Here $C_{y,t}$ refers to the count on day t in year y. The subscripts *1* and *2* refer to the pupping and molting periods, respectively. The scalar (H) and mean date (μ) of each curve differed among years and periods, but the width parameter (σ) was shared among curves to achieve parameter stability.

Curves were fit to the Tugidak count data by maximum likelihood assuming lognormal error (Figure 6). The peak date (μ) for the pupping and molting periods had standard deviations of 8.3 and 11.4 days, respectively. The residuals of the log-transformed data were found to have a standard deviation of approximately 0.35 with an autocorrelation of approximately 0.40. For modeling purposes, these were taken as conservatively large estimates of the day-to-day variability in counts unrelated to explanatory covariates. Random variation in counts was thus modeled as lognormal error with a standard deviation sd=0.35, and successive counts within a site and year were assumed to be autocorrelated with c=0.40. The random effect and the annual abundance at each site are then multiplied to obtain the simulated number of animals present on survey day k (i.e., the count on that day):

$C_{i,t,k} = N_{i,t}R(E_{i,t,k}, sd, c)$, where R is the lognormal random generator.

2. Covariates- Variable x, survey day: For each of the 8 surveys in the 5 areas, our base case defines a 10 day survey period starting with a random day 0 to 10 days before the mean survey date (August 28). Flights are conducted randomly on 7 of the 10 days. Variable y, tide: This variable is actually standardized time relative to low tide. It is drawn for each site and survey as a Uniform variable between –1.5 and 1.5. This range is slightly shorter than the range found in actual surveys (–1.9 to 1.8), but helped to simplify the modeling.

Two alternatives are considered for the unknown covariate:

Variable z, unknown covariate 1. This variable is assumed to have a linear positive trend from –1.5 to 1.5 over the 12 years. For each site and survey in a year, uniform random variation between –1.5 to 1.5 is added to the resultant annual value.

Variable z, unknown covariate 2. This variable is assumed to have a random uniformly distributed mean from –1.5 to 1.5 over the 12 years. For each site and survey in a year, uniform random variation between –1.5 to 1.5 is added to the resultant annual value.

Figure 7 shows a realization of these two unknown covariate structures for a single site. In the base case, it is assumed that there is no unknown covariate effect, so what is shown is the uniform tide variable across years and 7 surveys within a year. In study 1, covariate 1 is assumed to be present, which has the distinct linear trend in Figure 7 with uniform variation. In study 2, covariate 2 is present, which is seen by the random trend in pattern over years. The two covariates are not used in the same simulation.

*3. Attendance parameters-*The parameters $\{\beta\}$ are chosen from the reanalysis of PWS Trend A data (described above). For each of the 244 sites, each coefficient is randomly drawn with replacement from the 25 site-specific parameters.

For the unknown covariates 1 and 2, the central tendency of the linear parameter (there is no quadratic coefficient) is chosen to give a mean adjustment in the dependent variable of 0.5 at covariate value –1.5 and of –0.5 at 1.5. This is a similar effect to that for the tide variable, resulting in a coefficient of –0.32. Each site-specific coefficient is drawn from a normal distribution with mean –0.32 and standard deviation 0.16 (presuming that a 50% CV would be a reasonable amount of inter-site variability). This effect of the unknown covariate is shown in Figure 8 as a function of the unknown covariate's values. This is translated into an annual effect in Figure 9 for unknown covariate 1, which has a linear trend over time.

An analysis of multiple years of daily count data (Figure 6) indicates that the peak date of attendance has an inter-annual fluctuation with a standard deviation of about 10 days. While our base model assumes that peak date of attendance is constant, we do construct an alternative scenario with the standard deviation in peak date.

For the purposes of this study, we do not have an additional parameter for the overall lack of attendance of animals not present given optimal conditions (molt peak, low tide). We are more interested in results related to optimal survey design rather than exact estimation of the population. We address survey issues related to this overall lack of attendance based on first principles rather than these simulations.

*4. Counting errors-*We assume that all animals surveyed are counted without error. We believe that the aerial survey probably does have some measurement error in counts due to weather conditions, glare, pelage color, and so on, but that the magnitude of this error is small relative to the effects of attendance factors. Therefore we ignore counting errors as a second-order problem.

Estimation models

Raw count estimates (RCE)-The maximum count for each site and year across replicate surveys is calculated and totaled. No adjustments for covariates are made. The trend is calculated as the slope of an ordinary linear regression of ln(total max count) versus year. In some analyses, the mean count is substituted for the maximum count.

Covariate-corrected estimates (CCE) -We approximate the approach used by ADF&G (a Bayesian hierarchical procedure (Ver Hoef and Frost 1999)) in which counts are adjusted to account for the effects of covariates on the fraction of seals hauled out with the simpler general linear model procedure described above. The rationale for this replacement is that we need a computationally non-intensive method to approximate the technique of adjusting for covariates. Since the goal is to compare relative rather than absolute performance, we feel that the general linear model is acceptable for this purpose. Variables used in the general linear model are year, standardized date, standardized tide, and their quadratic counterparts. The unknown covariate is not used in estimation, except to test whether the general linear model operates properly. The trend is calculated by regression of the estimated annual effects from the general linear model versus year.

Simulations

The algorithm used for conducting simulations is:
1. Generate the population dynamics and hold fixed for each survey design.
2. Simulate the survey to obtain counts.
3. Use the estimation models to get estimates of trend and abundance.
4. Repeat steps 2 and 3 many times (100-1000).
5. Compare with the truth from step 1 to determine bias, accuracy, and power to detect trends.
6. Repeat steps 1 to 5 for various scenarios of interest.

The "base case" simulation contained the following assumptions: the survey was done annually at all sites over 12 years, 7 replicate counts were made over a 10-day period roughly centered about the peak of attendance, seals did not move among haulouts, the average trend was zero, the covariates affecting attendance were known, and there was no difference in behavior due to the haulout substrate. Variations on this base case were explored to test the sensitivity of estimates of trend to characteristics of the survey and the seal populations. Performance was defined as the bias and accuracy (defined as the square root of the mean squared residual, or RRMSE) of estimates of trends and of the effects of covariates.

Modifications of the survey-The first set of simulations looked at obvious modifications of the survey design. We also looked at changes that had been recorded during previous surveys, such as a linear trend in the survey date (Table 2).

Table 2. Survey modifications investigated

Scenario	Values Investigated
Years of data available	4, 8, 12
Replicate counts at each site w/in a year	2, 4, 7, 14
Timing relative to peak attendance	20 days before, at peak, 20 days after, split survey with 4 counts at peak and 3 counts 20 days later
Linear trend in survey date	First survey centered ~ 10 days after peak, survey is two days earlier each year

Lurking covariates-The next set of simulations looked at the effect of a "lurking covariate"; i.e., some unknown factor that affected the fraction of animals available to be counted. As this would have the greatest impact if the value of the variable fluctuated from year to year, two such scenarios were considered. In both, the magnitude of the effect of this lurking covariate was chosen to equal that of the "tide" covariate. In the first scenario, the lurking covariate fluctuated from year to year. In the second, the value of this lurking covariate declined linearly over the 12 years. (See the covariates section above for further details.)

Substrate differences-We investigated the effect of different covariate relationships in different substrates. These simulations were run only for the 46 sites in Prince William Sound, where a particular concern about substrate effects has been expressed. We effected a systematic difference among ice and rock substrate types by removing the effect of "tide" in all ice sites and selecting an identical "average" tide effect at all other sites. As there are relatively few ice sites in Prince William Sound, we ran some simulations in which we assigned an ice designation to a random half of the 46 sites. Because of the small number of ice sites, in all substrate trials we also set all site-specific variations in trend (L_k) to zero to minimize among-site variability not due to substrate type.

We also looked at the effect of movement among haulouts, in particular, directed movement to particular substrates. In one scenario, in each year up to 20% of the population was assumed to move from rock to ice sites, or vice versa. The percentage and direction of this movement was a random uniformly distributed variable that varied from year to year. In the second scenario, every year 2% of the population on rocky haulouts migrated to icy haulouts.

Power to detect individual site trends

Power is the ability to reject a null hypothesis when the null hypothesis is indeed false. In application to the harbor seal survey, power is the probability of detecting a trend in population abundance when one is present. Type II error β is 1 – power, or the probability of not detecting a trend in population abundance when one is present. Conversely, Type I error is the probability of detecting a trend in population abundance when one is not present. Type I error is usually set to a particular level α, typically 0.05.

In conducting a power analysis, it is necessary to have an unbiased procedure; otherwise detection of a trend may be confused with bias in the estimation.

Trends at a single site: For site k, the site's trend over time is the combination of an overall trend, an area-specific trend, and the site-specific trend,

$$\tau_1 = \ln(N_{t+1} / N_t) = \mu + A_i + L_k \text{, where}$$

$$\mu = \text{overall trend}$$

$$A_i = \text{adjustment for area}$$

$$L_k = \text{site - specific adjustment (like an error term).}$$

Either the RCE or CCE procedure produces a confidence interval (τ_{1L}, τ_{1U}) from a linear regression of the GLM estimates, and the confidence statement is

$$P(\tau_{1L} \le \tau \le \tau_{1U}) = 0.95 \quad \text{or} \quad P(\tau_{1L} \le \mu + A_i + L_k \le \tau_{1U}) = 0.95.$$

The random site-effects $\{L_k\}$ complicate the determination of power, because they will likely result in confidence intervals far away from 0 and therefore to rejection of the null hypothesis of no trend in almost all cases. The power approach we use adjusts for the random effects by rewriting the confidence statement as:

$$P(\tau_{1L} - L_k \le \mu + A_i \le \tau_{1U} - L_k) = 0.95.$$

In our power analyses we set μ to the trend magnitude we're interested in detecting, set A_i to zero, and calculate confidence intervals. This is equivalent to setting L_k to zero. However, actually doing so would provide misleading confidence in our ability to detect area-wide trends (see below).

If the confidence interval does not contain 0, then we conclude that a significant trend has been detected. Over 1000 replications, the proportion of significant trends is counted, which represents power. The power values for the 244 sites are averaged as a summary statistic.

Power to detect area-wide trends

Determining the power to detect a trend across an area is more complicated. Intuitively, one would sum up estimates by site using either the CCE or RCE approach and then estimate the trend using these sums. This procedure would effectively weight each site by its abundance. Unfortunately, the lognormal error structure of the counts induces a bias in abundance weighting that we have been unable to correct for.

Alternatively, we average the trend estimates for each site (which are on a log scale), and this preserves the unbiased nature of the results. As before, we set μ to the trend magnitude we're interested in detecting and set A_i to zero. We calculate confidence intervals as before, except we subtract the average L_k across all sites in the area. This subtraction is necessary to avoid confusing site effects with area trends, which might be important for areas with few sites. For areas with many sites, the random effects average out. Type I error is reduced by averaging across sites: there is almost no chance that a trend is detected when no trend is present.

Power vs. the number of sites surveyed

In the next power study, we further investigated the effect of the number of sites on power to detect trends. The previous power studies had different survey days for the 8 survey areas, which tends to introduce autocorrelation into the power statistics among areas. In the new study, a set of covariates was generated for all 244 sites with the same survey dates within a year. The operating model was then run with a set of pseudo-areas having from 2 to 75 sites.

Sensitivity analyses on power

A final power study looked at the effects of other factors. All runs were made with 25 sites in a single hypothetical survey area, and both average site power and average area power were evaluated to encompass the range of power likely to be encountered in practice. True annual trends from 0 to –0.1 were examined. A base run with typical inter-site variability (var(L_k) used in the other power studies was contrasted with no inter-site variability and inter-site variability twice as high (2X). The number of counts per site (ncounts) was reduced from 7 (base) to 5 and to 3 in two runs. The number of years (nyears) was reduced from 12 (base) to 5 to examine the power to detect a trend over 5 years. Finally, the RCE method was evaluated using average counts instead of maximum count for both 12 years and 5 years of data.

RESULTS & DISCUSSION

Modifications of the survey

Accuracy declined dramatically as the number of years of data available decreased, whereas bias was not affected (Figure 10). The number of replicate counts obtained each year did not appear to be as important, as accuracy increased only modestly as the number of replicates increased from 2 to 14 (Figure 11). The effect of the date of the survey was mixed. Trend estimates were relatively insensitive to the particular date surveyed, although performance was slightly improved when surveying at the peak (Figure 12). However, the estimates of the effect of date on the fraction of the population counted were very sensitive to when the survey was conducted. Surveying 20 days before or after the peak greatly reduced the accuracy of estimates of this effect, particularly of the linear component (Figure 13). Splitting the survey so that some counts were taken 20 days apart resulted in a substantial improvement in accuracy.

When the date of the survey evolved over time, the covariate-corrected trend estimates lost little accuracy. However, the raw count-based abundance estimates were severely biased (Figure 14). In contrast, natural fluctuations in the date of peak attendance resulted in little loss of accuracy.

Lurking covariates

Including an unknown covariate that fluctuated from year to year had only a minor effect on the accuracy of trend estimates (Figure 15). With 12 years of data, the trend component of the counts was easily distinguishable from the additional noise due to this covariate. However, we would expect a larger impact if fewer years of data were available. Both trend and abundance estimates were highly biased when the lurking covariate had a linear trend over time.

Substrate effects

Including a systematic difference in the effects of the "tide" covariate between ice and rock sites in Prince William Sound did not appreciably affect the performance of the survey, even if half of the 46 sites were assumed to be ice-type (not shown). Movement among sites did not affect trend estimates any more when movement was substrate-oriented than when it was random with respect to substrate (Figure 16).

However, in these scenarios only the covariate effects differed from site to site. In retrospect, we realize that even eliminating the effect of the major covariate "tide" results in differences between ice and rock sites that are no larger than the differences among rock sites (Figure 4). As these differences in covariate effects are adequately estimated in the base case, we should not expect to see any degradation in the performance of the survey from differences in the effects of the covariates. We did not include differences in average visibility between ice and rock sites, where movement between substrate types obviously might cause large biases.

Power to detect individual site trends

Covariate correction appeared to increase the power to detect trends at individual sites (Fig. 17). Power for the CCE approach was approximately 0.75 for an annual trend of –0.05. With the RCE approach, power was only about 0.50. These results are applicable to survey areas in which most of the harbor seal abundance is found at a single site.

Power to detect area-wide trends

When harbor seal abundance is scattered among several sites (the usual case), then the average trend among sites in a given survey area must be examined. Power increased with the number of sites in the area for both the CCE and RCE methods (Fig. 18). When the true annual trend was small, the power to detect an area-wide trend was often smaller than the power to detect a site-specific trend of similar magnitude, because inter-site variability is the major determinant of area-wide power. The opposite happened when the trend was large, because the averaging over sites reduced estimator variance sufficiently to counteract inter-site variability. For example, using a CCE approach, the power to detect an annual trend at a single site of magnitude –0.05 was approximately 0.75 (Fig. 17), but less power was obtained for detecting an area-wide trend of the same magnitude in survey routes comprised of 27 or fewer sites (Fig. 18). For most sites, the area-wide power was higher than the single-site power when the true trend was –0.6 or larger.

Neither the CCE method nor the RCE method consistently had greater power to detect area-wide trends (Fig. 18).

Power vs. the number of sites surveyed

The power statistics for the CCE method are shown in Figure 19a. As expected, the power generally increases as the number of sites per area increases. Oddly, the power is lower for 50 sites than for 35 sites. We attribute this to an artifact of the particular set of covariate values generated for this study. The main message from this figure is that with 25 or more sites, there is at least a 90% chance of detecting an annual trend of no less than –0.05. For 15 or more sites, there is at least a 90% chance of detecting a trend of no less than –0.075.

The power statistics for the RCE method shown in Figure 19b are similar to those for the CCE estimates. There is at least a 90% chance of detecting a trend of no less than –0.05 with 25 or more sites, and a trend of no less than –0.075 with 15 or more sites.

Sensitivity analyses on power

Not surprisingly the average area power with the CCE method was higher with no inter-site variability and lower with twice as much inter-site variability (Table 3). Interestingly, there was almost no power to detect a trend of –0.1 or less in the latter case. Reducing the number of counts to 5 or 3 decreased average area power only slightly, because inter-site variability was the dominating factor. Reducing the number of years to 5 increased power a slight amount when the true trend was less than –0.05 but decreased power when the true trend was greater than or equal to –0.05. At –0.05, the power decreased from about 90% to about 45%.

Average site power is appropriate when harbor seal abundance is mainly at a single site. Obviously, inter-site variability had no effect on average site power (Table 3). Reducing the number of counts to 5 or 3 decreased average site power, especially when the true trend was near –0.5. Reducing the number of years to 5 decreased power greatly, especially when the true trend was greater than or equal to –0.02. At –0.05, the power decreased from about 80% to about 10%.

With the RCE method, the effects of the factors on the average area power was similar to those for the CCE method (Table 4). Inter-site variability was again the dominating factor. Using average counts instead of maximum count had little effect when there were 12 years of data. Power was slightly higher for average counts when there were 5 years of data. Reducing the number of counts to 5 or 3 decreased average site power, especially when the true trend was above –0.4. Reducing the number of years to 5 decreased power greatly, especially when the true trend was greater than or equal to –0.01. At –0.05, the power decreased from about 50% to under 10%.

Power vs. survey interval

Our simulations assumed that each site was surveyed each year. While this reflects current practice at trend sites, only a fraction of the remaining sites are surveyed in a given year. Different regions within the state of Alaska are surveyed on a rotating schedule. Some power is necessarily lost by having fewer surveys. However, larger changes in population abundance will occur between surveys, somewhat mitigating this effect.

The power to detect a trend for a survey conducted every n^{th} year can be obtained from our results. For instance, to look up the power to detect a trend of –0.02/year in a survey conducted every third year for 13 years (5 surveys in total), we'd look up the power of detecting a trend of –0.06 (-0.02 x n) given 5 annual surveys. From Table 3, this value is 10.2%, smaller than the 18.6% power that would be obtained from annual surveys.

CONCLUSIONS

1. Estimation methods that adjust for the effects of covariates on the fraction of the population hauled out are more robust than methods based on raw counts.
2. The existing survey design seems to provide enough information to simultaneously estimate both the population trend and the effects of covariates at each haulout. This was not obvious to us before this study, as surveys are conducted in such a way as to minimize the range of values of the covariates (flying at the same time of year, stage of the tide, etc.). It was plausible that the covariate effects would be so poorly estimated that the corrections would do more harm than good.
3. To the extent possible, two rules should be followed in surveying. First, the bulk of the counts should be conducted under as standard a set of conditions as possible. These should be timed to coincide with the peak haulout numbers. However, some effort should also be devoted to counting under contrasting conditions, e.g., allocating some counts to later in the season. The effect of a covariate on the fraction hauled out is much better estimated when observations from contrasting conditions exist.
4. If animals move among sites, trend estimation methodology may also need to include corrections to absolute abundance. This is critical if there are substantial differences in the average fraction hauled out among sites, and a substantial fraction of animals move from one site to another between years. If so, trend estimates may be biased.
5. An evolution over time in the fraction hauled out can cause a substantial bias. One obvious candidate for such a phenomenon would be a trend in the availability of food such that the time spent foraging increased or decreased. Annual mark-recapture studies could detect such trends.
6. A key factor governing the robustness and power of harbor seal surveys is inter-site variability in trend. In particular, this affects our ability to detect area-wide

trends. This factor is well understood for Prince William Sound and Kodiak trend sites, but better information is needed for other sites in the EVOS area.

7. Current trend routes might conceivably span several demographically isolated populations with differing trends. If the boundaries of such populations are not known, then managers must respond in a risk-averse fashion to any adjacent group of haulouts exhibiting a decline. If the population boundaries are known, our ability to detect a trend in each population is a simple power calculation. The trend route would contain n sites from the population (possibly a subset of a larger group of haulouts used by the population). The ability to detect a trend of a given magnitude in the population is simply the power to detect a trend from a group of n sites.
8. Although we did not explicitly investigate correction of counts to absolute rather than relative abundance, such methods will probably be affected by covariates. Therefore, methods for correcting counts to absolute abundance that include covariates should be investigated. If not, specific counts will be over- or under-adjusted, depending upon the conditions when they were observed. One such study is currently being conducted at Tugidak Island. A photographic mark-recapture approach is being combined with repeated counts under varying conditions to create covariate-based corrections to absolute abundance.

ACKNOWLEDGMENTS

The authors wish to thank Peter Boveng, Kathy Frost, Lauri Jemison, Brendan Kelly, Beth Matthews, Grey Pendleton, Brian Taras, Dave Withrow, and Kate Wynne for their generosity with data and their insights. Special thanks are due to Jay Ver Hoef for generously sharing his preliminary Bayesian analyses of the ADF&G PWS 'A' survey data. The research described in this paper was supported by the *Exxon Valdez* Oil Spill Trustee Council. However, the findings and conclusions presented by the authors are their own and do not necessarily reflect the views or position of the Trustee Council.

LITERATURE CITED

Butterworth, D.S, and M.O. Bergh. 1993. The development of a management procedure for the South African anchovy resource. In Risk Evaluation and Biological Reference Points for Fisheries Management, Can. Spec. Pub. Fish Aquat. Sci. 120: 83-90.

Butterworth, D.S., K.L. Cochrane, and J.A.A. Oliveira. 1997. Management procedures: a better way to manage fisheries? The South African experience. In Global Trends: Fisheries Management, Amer. Fisheries Soc. Symp. 20: 83-90.

Cochrane, K.L., D.S. Butterworth, J.A.A. Oliveira, and B.A. Roel. 1998. Management procedures in a fishery based on highly variable stocks and with conflicting objectives: experiences in the South African pelagic fishery. Rev. Fish. Biol. Fisheries 8: 177-214.

Frost, K. F., L. F. Lowry, and J. M. Ver Hoef. 1999. Monitoring the trend of harbor seals in Prince William Sound, Alaska, after the Exxon Valdez oil spill. Marine Mammal Science 15:494-506.

Hoover, A. A. 1983. Behavior and ecology of harbor seals (*Phoca vitulina richardsi*) inhabiting glacial ice in Aialik Bay, Alaska. MS Thesis. University of Alaska, Fairbanks.

Huber, H. R. 1995. The abundance of harbor seals (*Phoca vitulina richardsi)* in Washington, 1991-1993. Master of Science thesis, University of Washington, Seattle. 37p.

Jemison, L. A., and B. P. Kelly. in prep. Pupping phenology and demography of harbor seals (*Phoca hispida richardsi*) on Tugidak Island, Alaska. Marine Mammal Science.

Kirkwood, G.P. 1997. The revised management procedure of the International Whaling Commission. In Global Trends: Fisheries Management, Amer. Fisheries Soc. Symp. 20: 91-99.

Mathews, E. A., and B. P. Kelly. 1996. Extreme temporal variation in harbor seal (*Phoca vitulina richardsi)* numbers in Glacier Bay, a glacial fjord in southeast Alaska. Marine Mammal Science 12:483-489.

Small, R. J., G. W. Pendleton, and K. Wynne. 1998. Harbor seal population trends in the Ketchikan, Sitka, and Kodiak Island areas of Alaska. Pages 7-26 *in* R. J. Small, editor. Harbor Seal Investigations in Alaska, Annual Report, NOAA Grant NA57FX0367. Division of Wildlife Conservation, Alaska Department of Fish and Game, Anchorage, Alaska.

Small, R. J., Hastings, K., and L. A. Jemison. 1999. Investigations of harbor seals in Alaska, Annual Report, NOAA Award Number NA87FX0300. Division of Wildlife Conservation, Alaska Department of Fish and Game, Anchorage, Alaska.

Thompson, P., and P. Rothery. 1987. Age and sex differences in the timing of the moult in the common seal, *Phoca vitulina*. Journal of Zoology, London 212:597-603.

Ver Hoef, J.M., and K.J. Frost. 1999. Bayesian hierarchical models for estimating harbor seal changes in Prince William Sound Alaska. Alaska Department of Fish and Game, Fairbanks, Alaska. Manuscript dated June 9, 1999.

Ver Hoef, J.M., and K.J. Frost. 2000. Bayesian hierarchical models for estimating harbor seal changes in Prince William Sound Alaska. Alaska Department of Fish and Game, Fairbanks, Alaska. Manuscript dated May 26, 2000.

Withrow, D. E., and T. R. Loughlin. 1995. Abundance and distribution of harbor seals (*Phoca vitulina richardsi*) along the Aleutian Islands during 1994. Annual Report for 1994 under the NMFS, MMPA Population Assessment Program, Office of Protected Resources, NOAA, NMFS, Silver Spring, MD 20910. May 1995. 16 p.

Withrow, D. E., and T. R. Loughlin. 1996b. Haulout behavior and a correction factor estimate for the proportion of harbor seals missed during molt census surveys near Cordova, Alaska. Pages 89-116 *in* Marine Mammal Protection Act and Endangered Species Act Implementation Program 1995, Office of Protected Resources, NMFS, NOAA, 1335 East-West Highway, Silver Spring, MD 20910. [Available from National Marine Mammal Lab, 7600 Sand Point Way, NE, Seattle, WA 98115].

Withrow, D. E., and T. R. Loughlin. 1997. Abundance and distribution of harbor seals *Phoca vitulina richardsi* along the south side of the Alaska Peninsula, Shumigan Islands, Cook Inlet, Kenai Peninsula and the Kodiak Archipelago in 1996. Pages 119-

155 *in* Marine Mammal Protection Act and Endangered Species Act Implementation Program 1997, Office of Protected Resources, NMFS, NOAA, 1335 East-West Highway, Silver Spring, MD 20910.

Table 3. Effects on average area power and average site power of inter-site variability var(L_k), number of counts (ncounts), and number of years (nyears), using CCE.

		Trend									
		0	-0.01	-0.02	-0.03	-0.04	-0.05	-0.06	-0.07	-0.08	-0.10
Base	Area Avg.	0.0%	0.0%	0.0%	0.0%	0.0%	87.5%	96.7%	100.0%	100.0%	100.0%
var(L_k)=0	Area Avg.	5.4%	85.4%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
var(L_k)=2X	Area Avg.	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
ncounts=5	Area Avg.	0.0%	0.0%	0.0%	0.0%	0.0%	81.6%	93.0%	100.0%	100.0%	100.0%
ncounts=3	Area Avg.	0.0%	0.0%	0.0%	0.0%	0.3%	75.4%	87.5%	100.0%	100.0%	100.0%
nyears=5	Area Avg.	0.0%	0.0%	0.5%	2.8%	7.0%	45.2%	51.1%	83.8%	94.1%	99.6%
Base	Site Avg.	4.3%	7.9%	18.6%	37.4%	59.3%	78.7%	91.3%	97.1%	99.2%	100.0%
var(L_k)=0	Site Avg.	4.3%	7.9%	18.6%	37.4%	59.3%	78.7%	91.3%	97.1%	99.2%	100.0%
var(L_k)=2X	Site Avg.	4.3%	7.9%	18.6%	37.4%	59.3%	78.7%	91.3%	97.1%	99.2%	100.0%
ncounts=5	Site Avg.	5.3%	8.1%	16.6%	31.2%	48.8%	67.2%	81.6%	91.2%	96.4%	99.6%
ncounts=3	Site Avg.	5.4%	6.3%	11.7%	20.2%	32.4%	46.9%	62.5%	74.7%	84.4%	95.3%
nyears=5	Site Avg.	5.2%	5.2%	5.6%	6.1%	7.4%	8.8%	10.2%	12.0%	14.5%	19.7%

Table 4. Effects on average area power and average site power of inter-site variability var(L_k), number of counts (ncounts), number of years (nyears), and using average count instead of maximum count, using RCE.

		Trend									
		0	-0.01	-0.02	-0.03	-0.04	-0.05	-0.06	-0.07	-0.08	-0.10
Base	Area Avg.	0.0%	0.0%	0.0%	0.0%	0.2%	80.8%	97.3%	100.0%	100.0%	100.0%
var(L_k)=0	Area Avg.	7.1%	68.4%	96.2%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
var(L_k)=2X	Area Avg.	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.8%
ncounts=5	Area Avg.	0.0%	0.0%	0.0%	0.0%	0.1%	59.1%	96.1%	100.0%	100.0%	100.0%
ncounts=3	Area Avg.	0.0%	0.0%	0.0%	0.0%	0.1%	98.2%	72.3%	99.8%	100.0%	100.0%
nyears=5	Area Avg.	0.0%	0.1%	0.2%	9.8%	1.8%	42.9%	45.1%	64.2%	89.9%	98.4%
avg.count	Area Avg.	0.0%	0.0%	0.0%	0.0%	0.0%	79.8%	99.7%	100.0%	100.0%	100.0%
avg.count(5 yr)	Area Avg.	0.0%	0.0%	0.0%	9.2%	0.2%	55.7%	60.0%	84.3%	99.8%	100.0%
Base	Site Avg.	4.7%	6.1%	9.3%	19.2%	32.8%	50.9%	66.6%	72.7%	81.8%	97.1%
var(L_k)=0	Site Avg.	4.7%	6.1%	9.3%	19.2%	32.8%	50.9%	66.6%	72.7%	81.8%	97.1%
var(L_k)=2X	Site Avg.	4.7%	6.1%	9.3%	19.2%	32.8%	50.9%	66.6%	72.7%	81.8%	97.1%
ncounts=5	Site Avg.	5.6%	7.7%	11.8%	22.1%	29.0%	42.2%	60.4%	69.0%	81.4%	95.1%
ncounts=3	Site Avg.	6.1%	6.3%	12.3%	18.1%	26.2%	43.0%	48.5%	54.8%	66.0%	90.8%
nyears=5	Site Avg.	5.4%	5.0%	5.5%	5.2%	6.0%	7.5%	8.5%	8.8%	9.5%	14.1%
avg.count	Site Avg.	4.6%	6.4%	12.2%	29.4%	47.3%	67.4%	86.1%	85.9%	95.1%	99.6%
avg.count(5 yr)	Site Avg.	5.6%	5.1%	5.5%	5.6%	7.3%	8.8%	10.0%	11.7%	13.8%	19.3%

LIST OF FIGURES

Figure 1. The fraction of seals hauled out as a function of a hypothetical covariate (e.g. tide). The vertical bars illustrate the differing amount of correction necessary to calculate absolute abundance from counts for different states of the covariate.

Figure 2. Deviations of estimate site trends from the average trend across sites for the 25 sites in the Prince William Sound Trend A survey from reanalysis of the raw data.

Figure 3. Estimated effects of survey day for the 25 sites in the Prince William Sound Trend A survey from reanalysis of the raw data. The average effect across sites is shown with the bold line. Standardized day is (Day – August 28)/100.The range of the observed data is used for the range of the x-axis, which corresponds to 10 days before August 28 to 20 days after August 28.

Figure 4. Estimated effects of tide for the 25 sites in the Prince William Sound Trend A survey from reanalysis of the raw data. The average effect across sites is shown with the bold line. Standardized tide is (Time relative to low tide)/100.The range of the observed data is used for the range of the x-axis, which corresponds to 2.5 hr before low tide to 2.5 hr after low tide.

Figure 5. The 244 haulout locations included in this simulation study (triangles) and the subdivision of the study region into 5 component areas (shaded).

Figure 6. Six years of harbor seal counts from Tugidak Island, Alaska (diamonds). The dashed lines are bell-shaped curves fit to the pupping and molting peaks, and the solid line denotes the summed curves. The residuals of this 'fit' are used to estimate random variation and autocorrelation in count data generated by the simulation model.

Figure 7. A single realization of known and unknown covariates for a single site across years with 7 surveys each year. The Base Case shows the known uniform covariate always present, "Study 1" shows the unknown linear covariate that is only present in Study 1, and "Study 2" shows the unknown random-mean covariate that is only present in Study 2.

Figure 8. The effect of either of the two unknown covariates for 25 sites (to correspond to the PWS Trend A site reanalysis) as a function of the unknown covariate's values. The coefficient for each site is drawn from a normal distribution with mean –0.32 and standard deviation 0.16. The average across sites is also shown.

Figure 9. The average effect of unknown covariate 1, which has a linear trend over years, as a function of year.

Figure 10. Performance of estimates of trend as a function of the number of years of data available. Results from both a covariate-corrected estimator (CCE) and an estimator based on the maximum count (RCE). Top graph shows bias, bottom accuracy. Results from 100 simulations with 244 haulout sites.

Figure 11. Performance of estimates of trend as a function of the number of replicate counts at each site. Results from both a covariate-corrected estimator (CCE) and an estimator based on the maximum count (RCE). Top graph shows bias, bottom accuracy. Results from 100 simulations with 244 haulout sites.

Figure 12. Performance of estimates of trend as a function of the average date of the survey relative to Aug. 28. Results from both a covariate-corrected estimator (CCE) and an estimator based on the maximum count (RCE). Top graph shows bias, bottom accuracy. Results from 100 simulations with 244 haulout sites.

Figure 13. Performance of estimates of the covariate effect as a function of the average date of the survey relative to Aug. 28. Results for both the linear (beta1) and quadratic (beta2) effects. Top graph shows bias, bottom accuracy. Results from 100 simulations with 244 haulout sites.

Figure 14. Performance of estimates of trend relative to the base case for variation in the peak date of attendance (peak fluctuates) and for the case where the survey date averages two days later each year (datetrend). Results from both a covariate-corrected estimator (CCE) and an estimator based on the maximum count (RCE). Top graph shows bias, bottom accuracy. Results from 100 simulations with 244 haulout sites.

Figure 15. Performance of estimates of trend when lurking covariates exist. Shown are the case where the unknown covariate has a linear trend over years (linear cov) and when it varies from year to year (random cov). Results from both a covariate-corrected estimator (CCE) and an estimator based on the maximum count (RCE). Top graph shows bias, bottom accuracy. Results from 100 simulations with 244 haulout sites.

Figure 16. Performance of estimates of trend when animals move between substrate types that differ in their covariate effects. Shown are the case when this movement varies from year to year (ran move) and when there is an annual net movement from rocky to ice substrate (dir move) Results from both a covariate-corrected estimator (CCE) and an estimator based on the maximum count (RCE). Top graph shows bias, bottom accuracy. Results from 100 simulations with 244 haulout sites.

Figure 17. Power to detect site-specific trends of various magnitudes using covariate-corrected estimates (CCE) or raw count estimates (RCE).

Figure 18. Power to detect area-wide trends of various magnitudes, for each of the eight simulated survey routes, using covariate-corrected estimates (CCE) or raw count estimates (RCE).

Figure 19. Power to detect area-wide trends of various magnitudes, for areas containing various numbers of sites, using covariate-corrected estimates (CCE) or raw count estimates (RCE).

Appendix. Substrate type and harbor seal abundance at Prince William Sound and Kodiak population trend route sites were compared to all haulout sites within the EVOS spill area (Table A1, Figures A1-A3). Additionally, substrate type and abundance within the mean of the maximum distances recorded between any 2 haulout sites by adult (~15 km) and subadult (~35 km) seals were also examined.

Table A1.

		Kodiak & PWS			Kodiak			PWS		
	Spill Area	Trend Sites	at 15 km	at 35 km	Trend Sites	at 15 km	at 35 km	Trend Sites	at 15 km	at 35 km
Substrate										
Non-Ice	244	55	10	27	30	1	7	25	9	20
Ice	6	0	0	2	0	0	0	0	0	2
Abundance	23,267	4,742 (20.4)[a]	794 (14.3)[b]	2,828 (37.4)[b]	3,615 (15.5)[a]	81 (2.2)[b]	413 (10.3)[b]	1,127 (4.8)[a]	713 (38.8)[b]	2,415 (68.2)[b]

[a]Percentage of total abundance within EVOS spill area at trend sites.
[b]Percentage of additional abundance within the maximum distance typically traveled by adult seals between two haulouts.

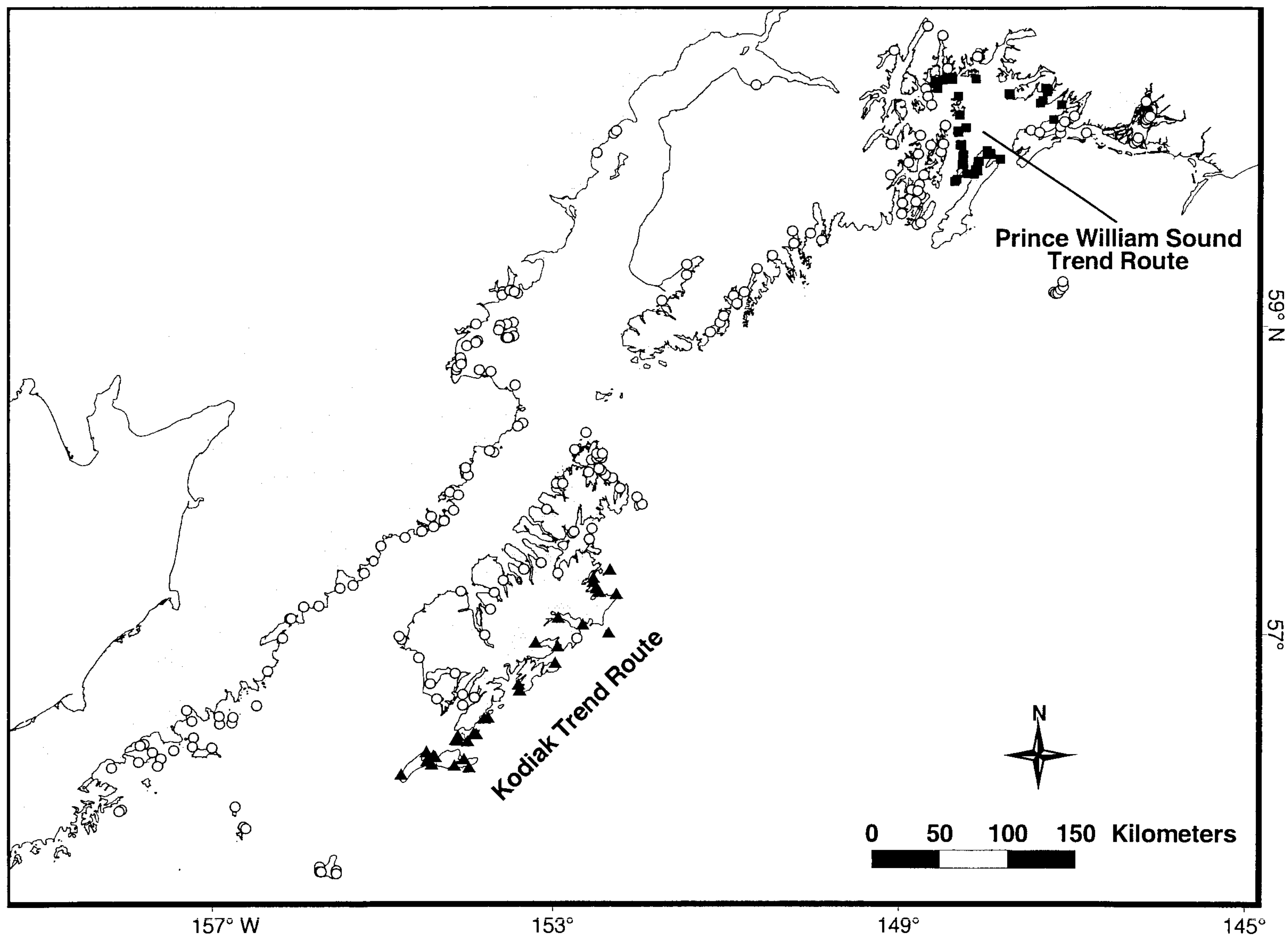


Figure A1. Overview of all haulouts used: Kodiak Trend Route (▲), Prince William Sound Trend Route (■), and non-trend route (○) haulouts.

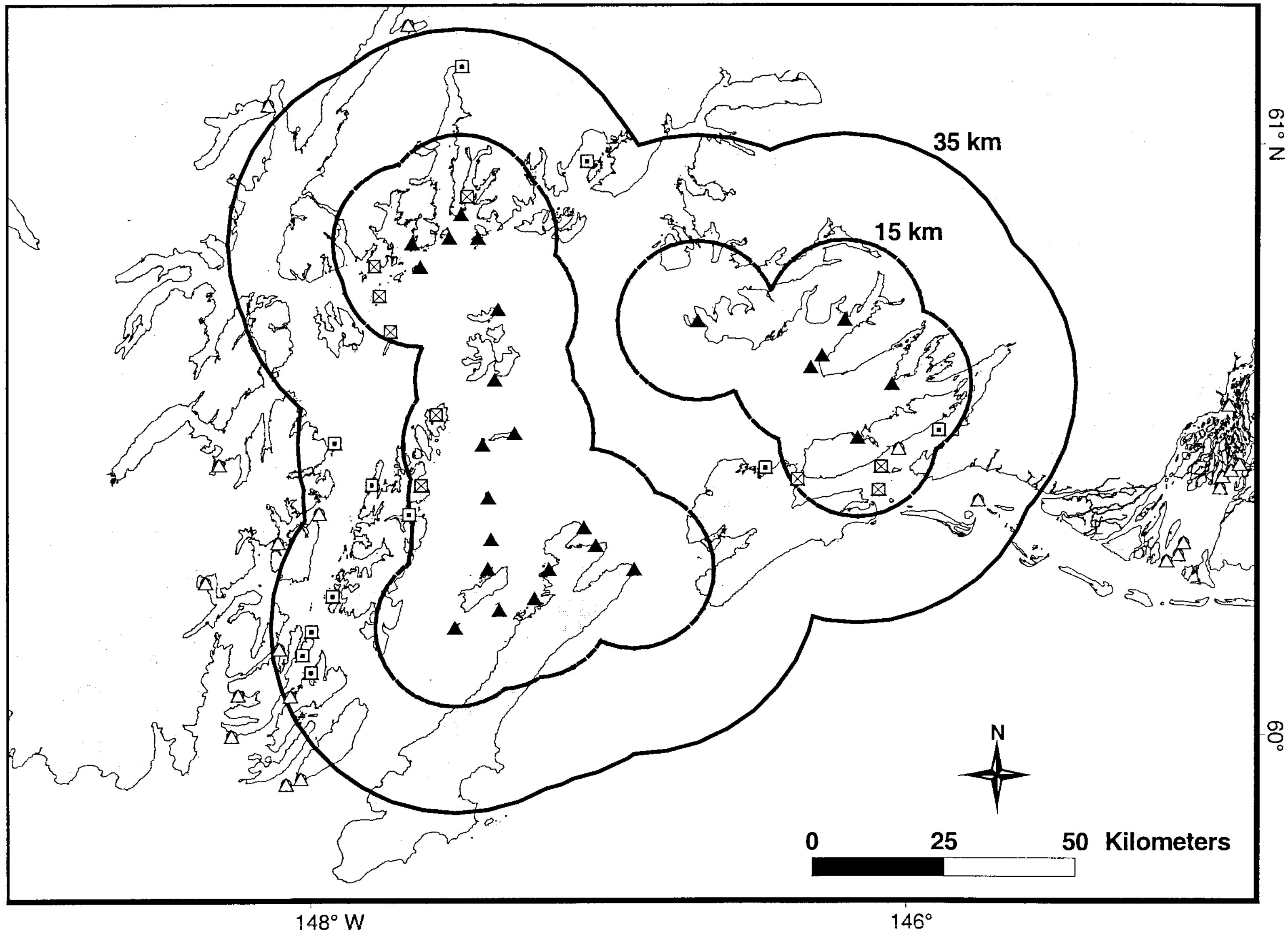


Figure A2. Prince William Sound haulouts, outside the Trend Route Area, but within typical swimming distance of adult (☒) and subadult (▣) harbor seals. Other model haulouts on the Trend Route (▲) and outside the Trend Route (△) are indicated.

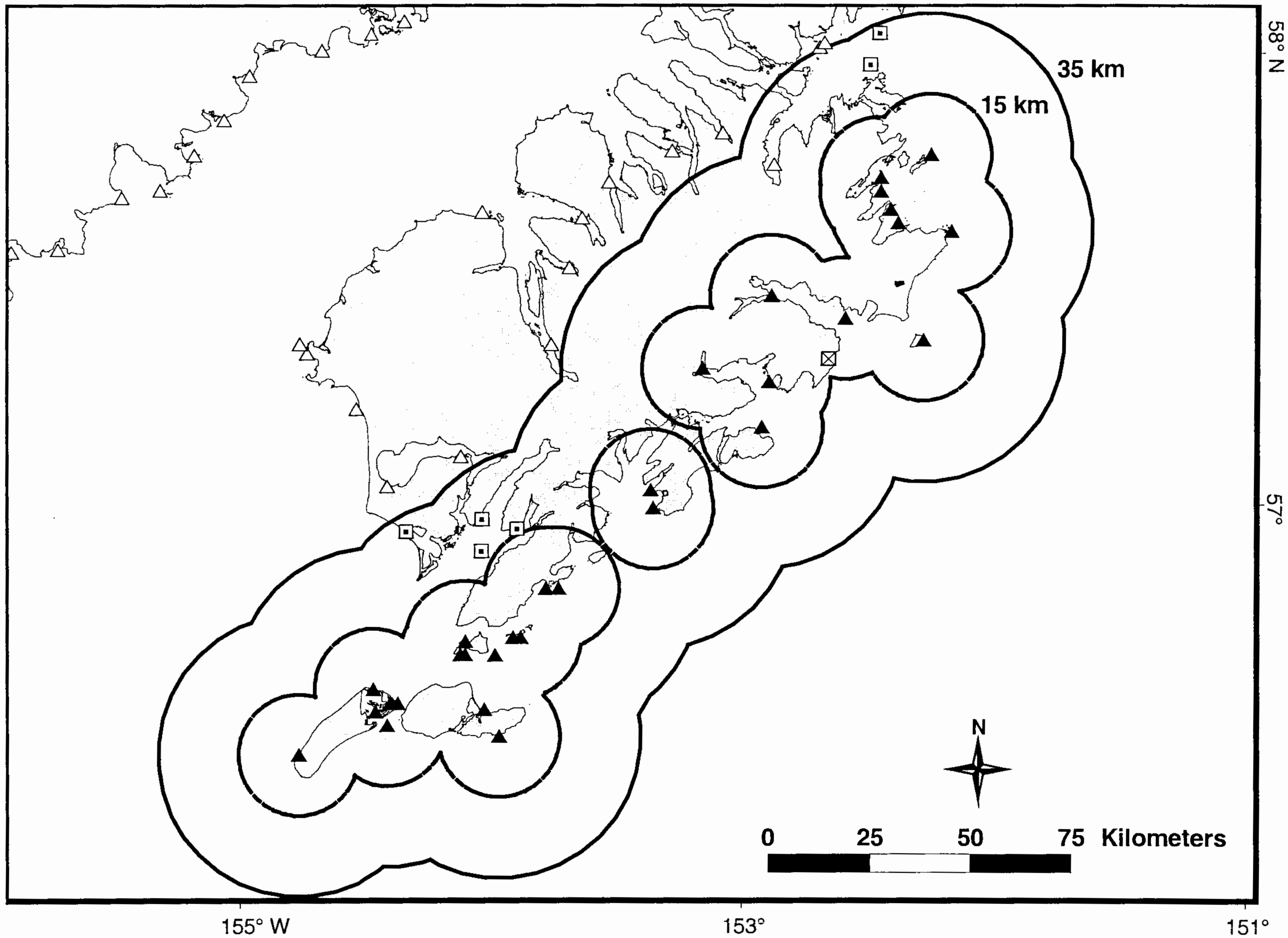


Figure A3. Kodiak haulouts, outside the Trend Route Area, but within typical swimming distance of adult (⊠) and subadult (⊡) harbor seals. Other model haulouts on the Trend Route (▲) and outside the Trend Route (△) are indicated.

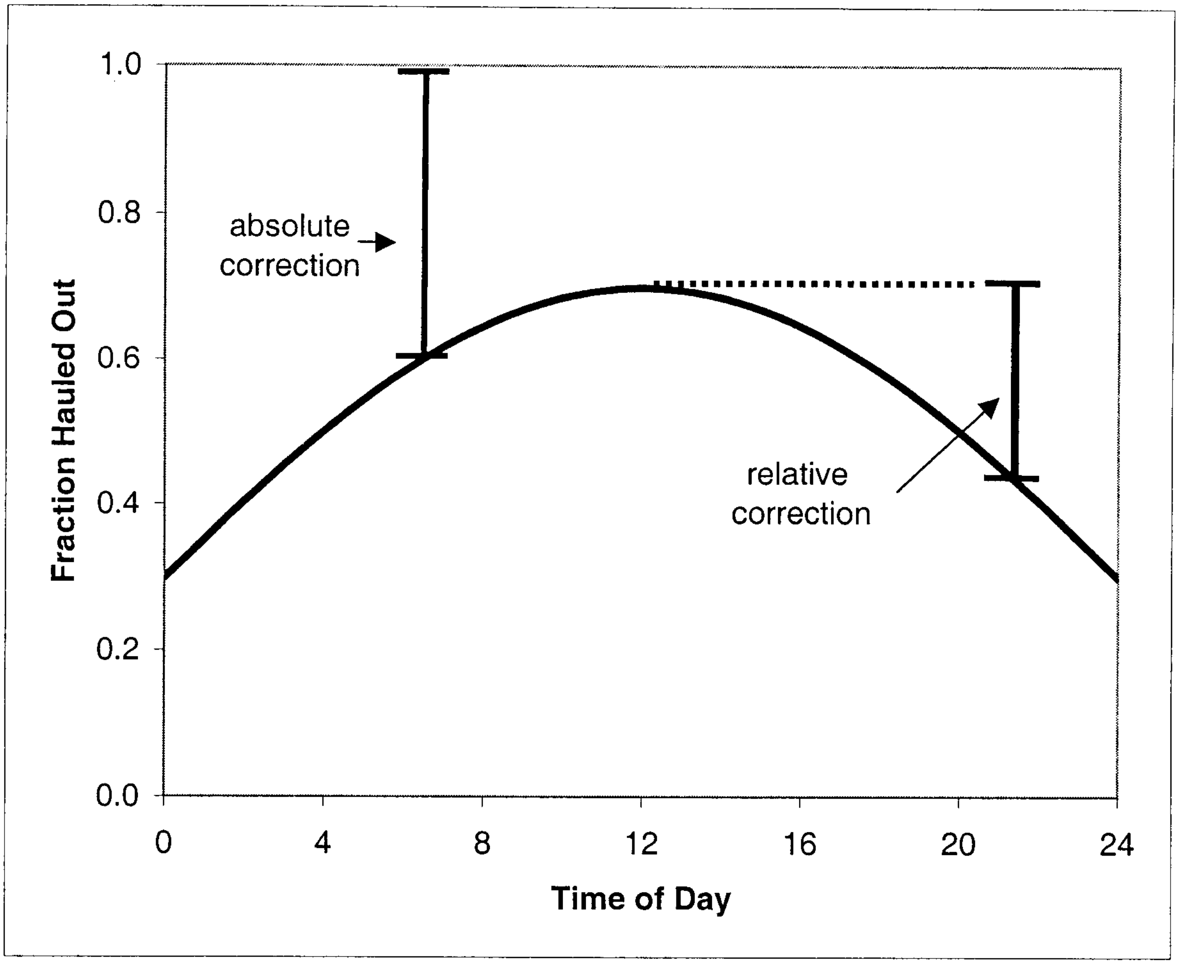

1.0
0.8
0.6
0.4
0.2
0.0
Fraction Hauled Out
absolute
correction
relative
correction
0
4
8
12
16
20
24
Time of Day

Figure 1

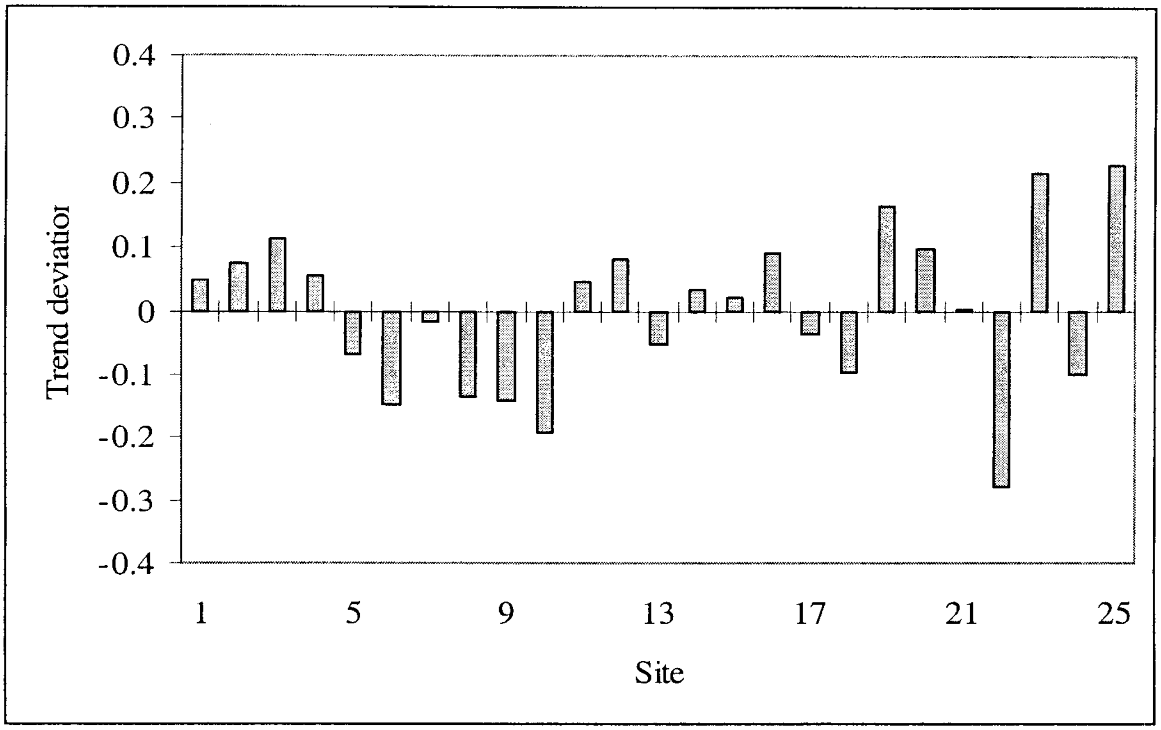


Figure 2

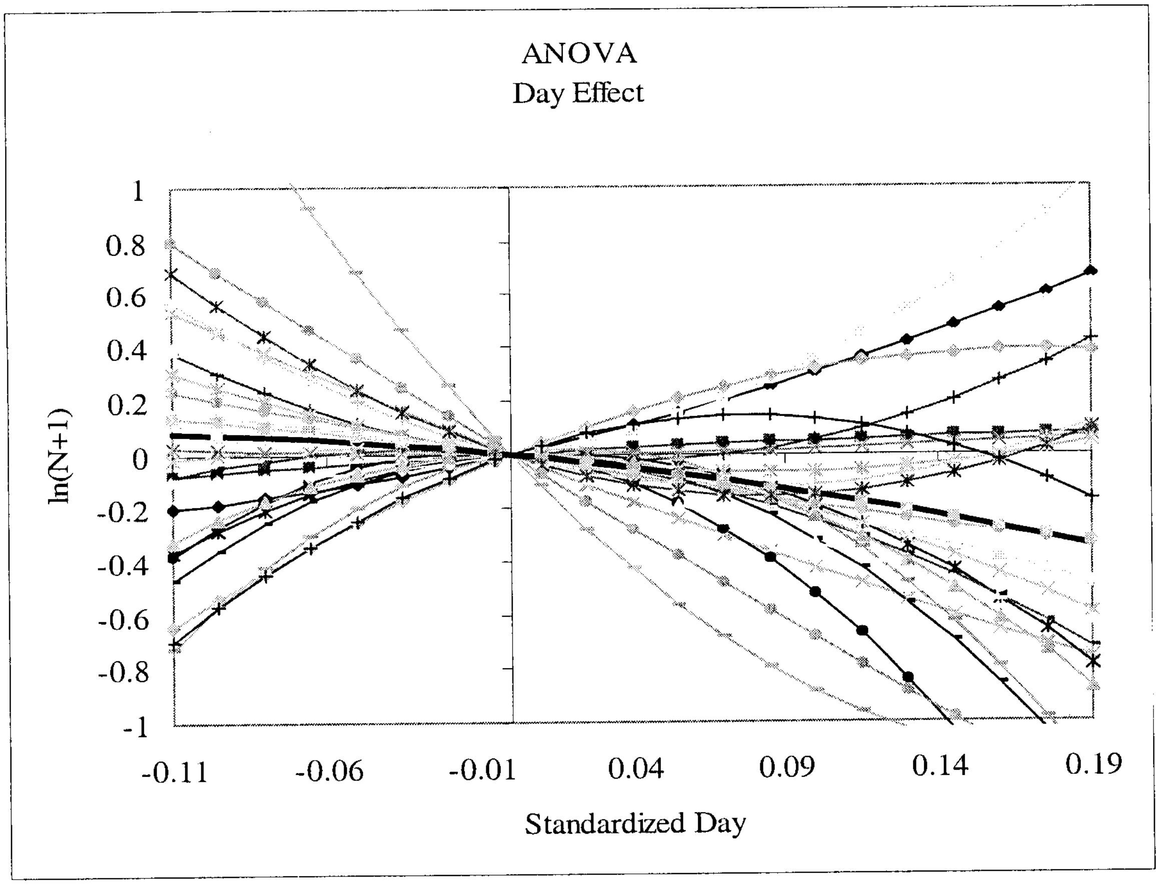
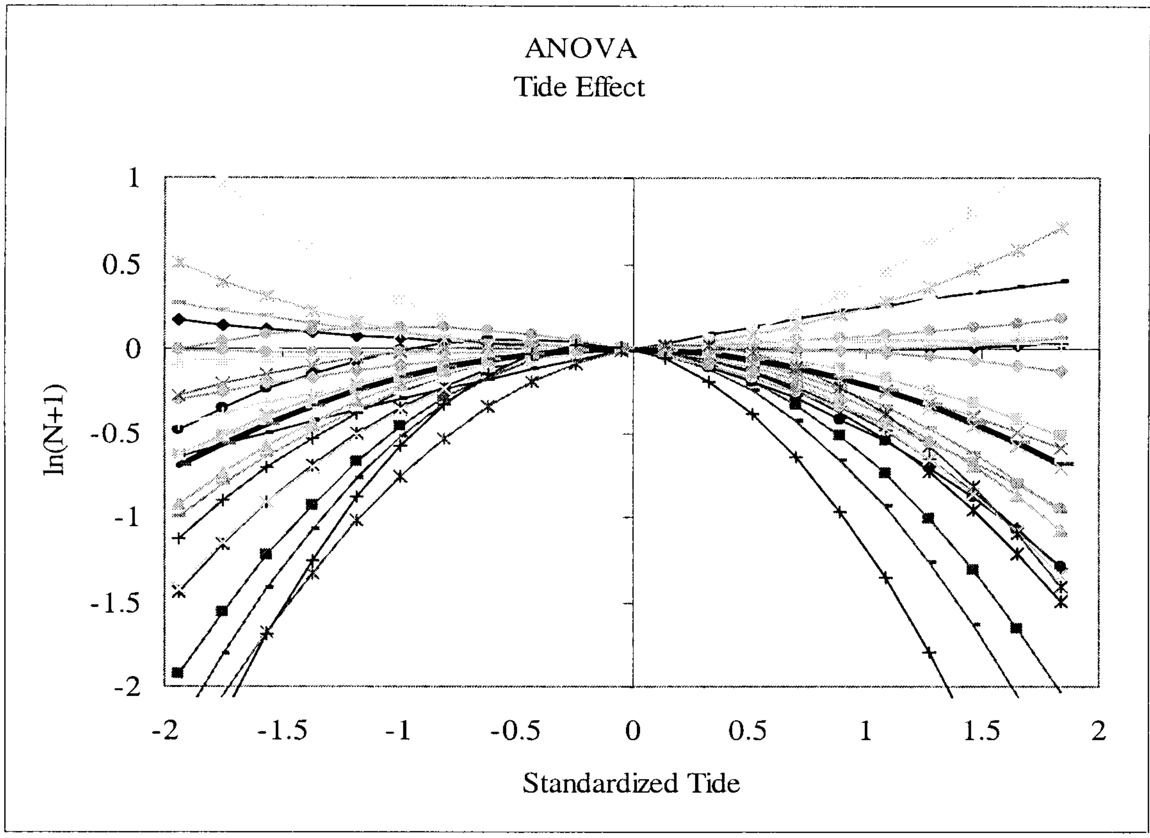

ANOVA
Day Effect
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
ln(N+1)
-0.11
-0.06
-0.01
0.04
0.09
0.14
0.19
Standardized Day

Figure 3



ANOVA
Tide Effect
1
0.5
0
-0.5
-1
-1.5
-2
ln(N+1)
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
Standardized Tide

Figure 4

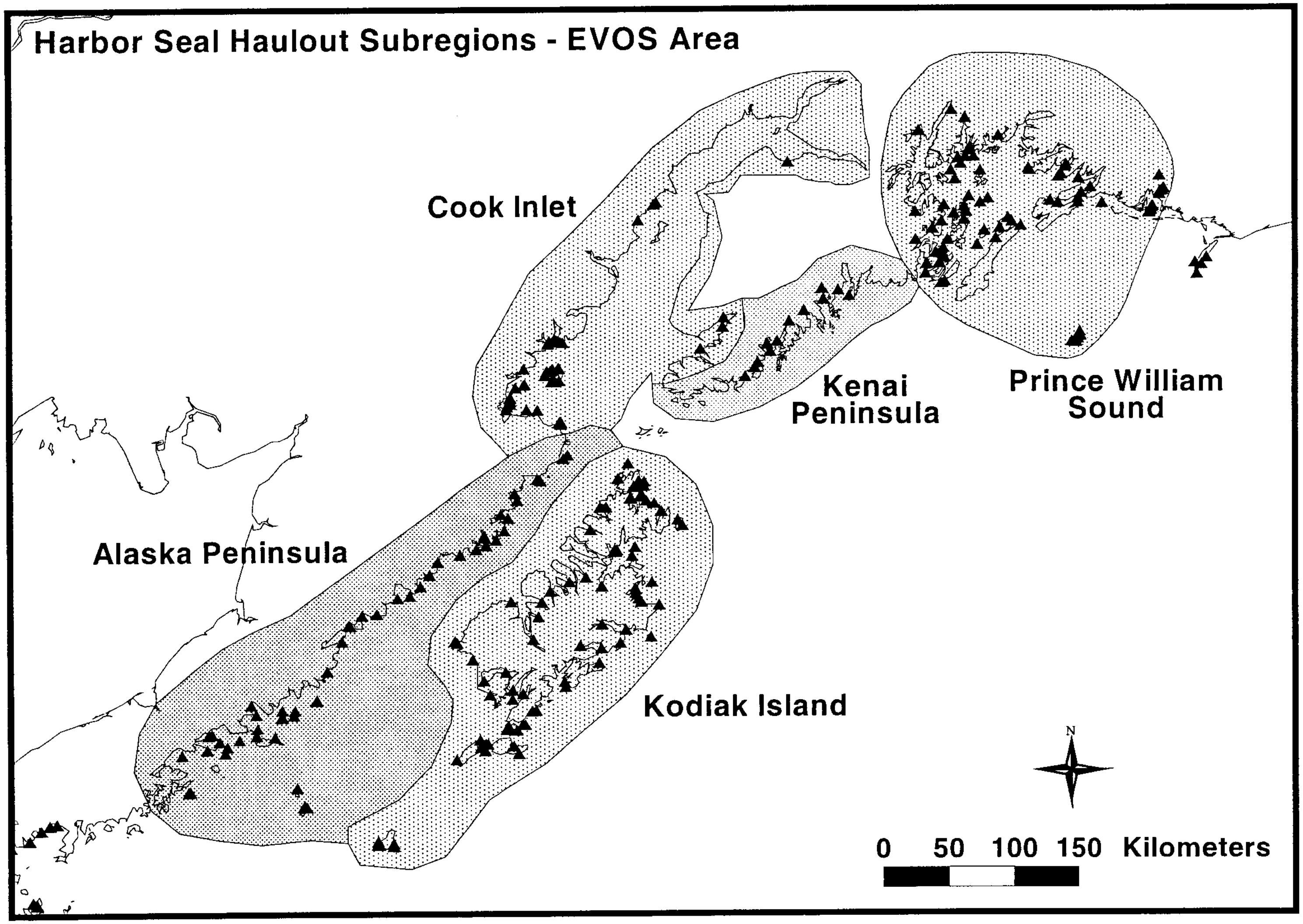

Harbor Seal Haulout Subregions - EVOS Area
Cook Inlet
Kenai Peninsula
Prince William Sound
Alaska Peninsula
Kodiak Island
N
0 50 100 150 Kilometers

Figure 5

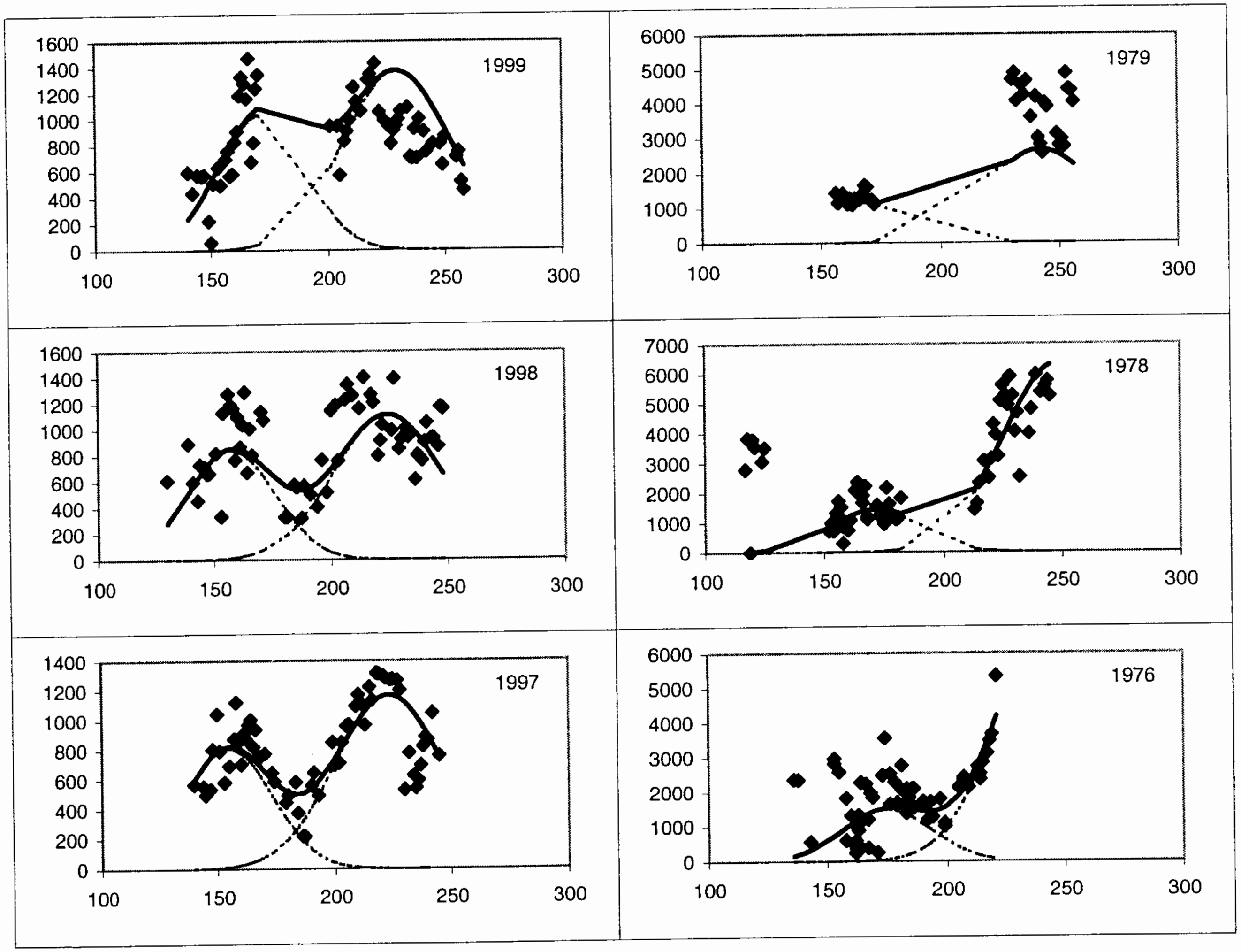
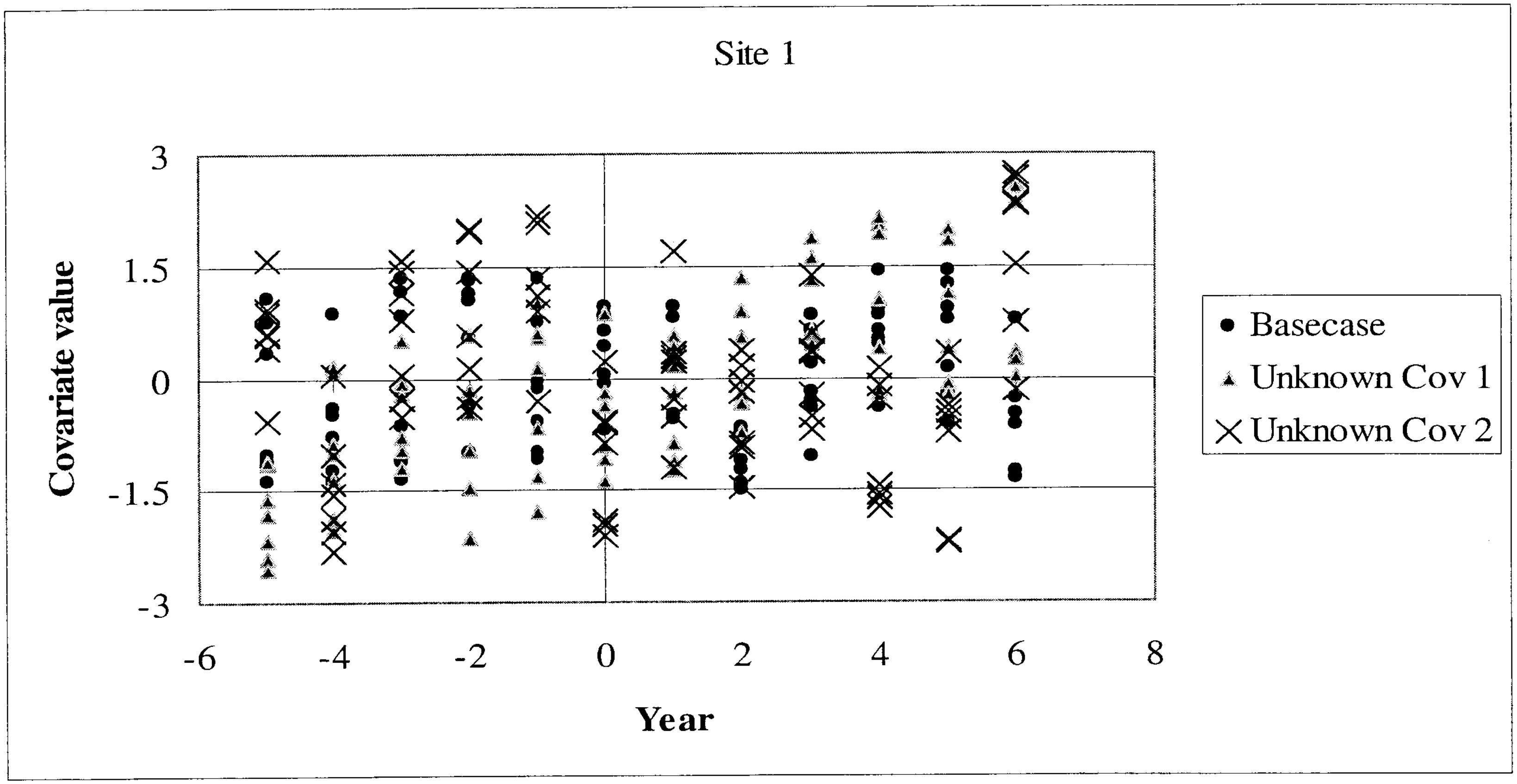
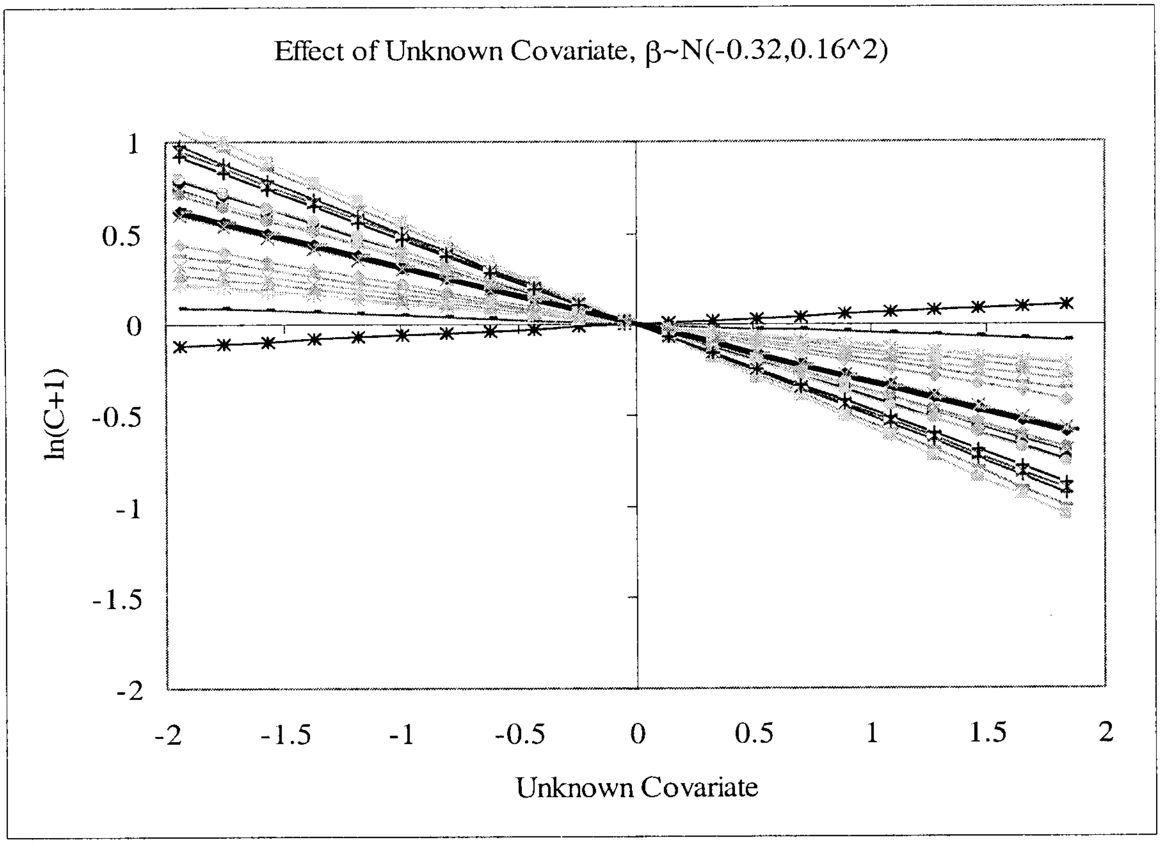


Figure 6



Site 1
Covariate value
3
1.5
0
-1.5
-3
-6
-4
-2
0
2
4
6
8
Year
Basecase
Unknown Cov 1
Unknown Cov 2

Figure 7



Effect of Unknown Covariate, β~N(-0.32,0.16^2)
1
0.5
0
-0.5
-1
-1.5
-2
ln(C+1)
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
Unknown Covariate

Figure 8

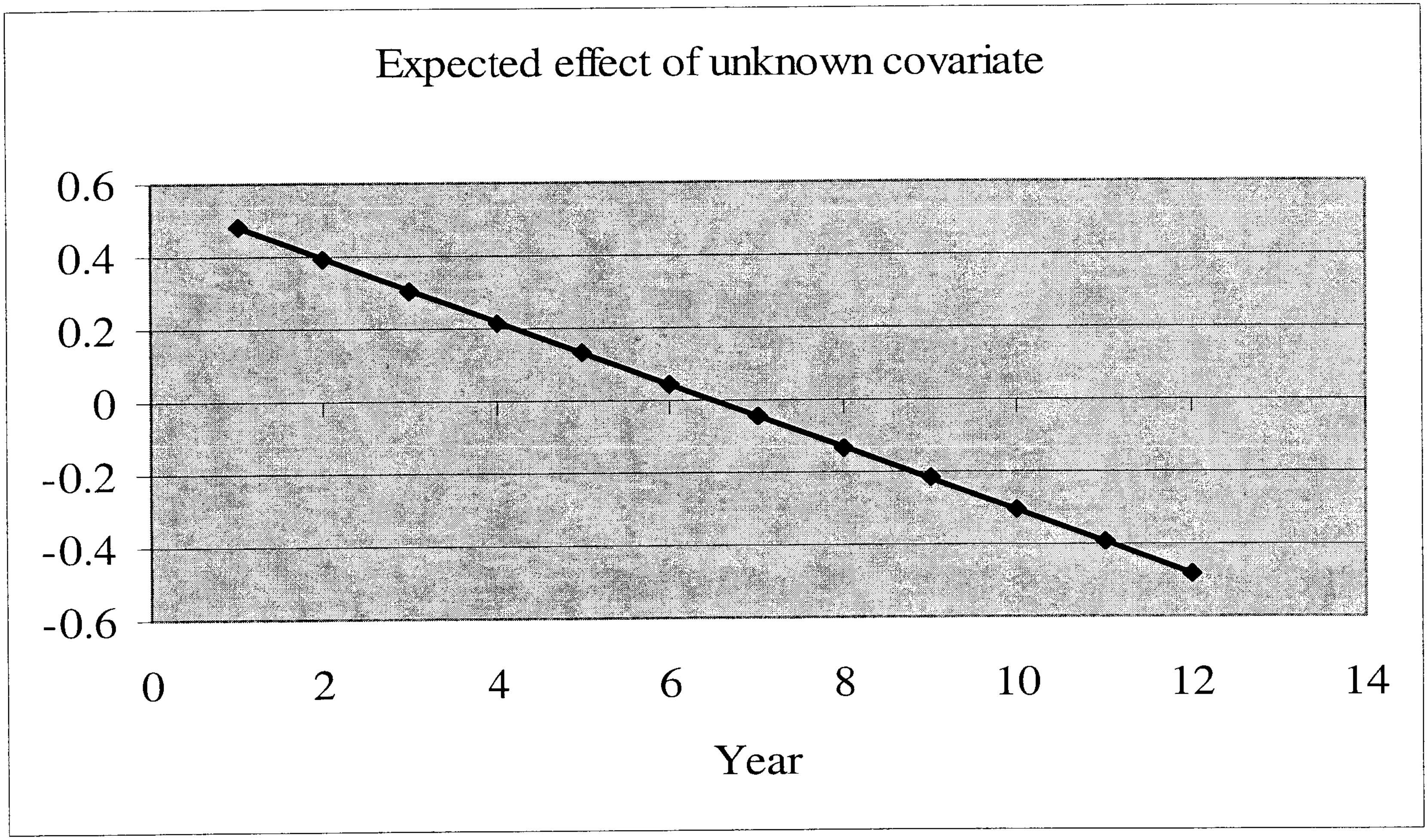

Expected effect of unknown covariate
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
2
4
6
8
10
12
14
Year

Figure 9

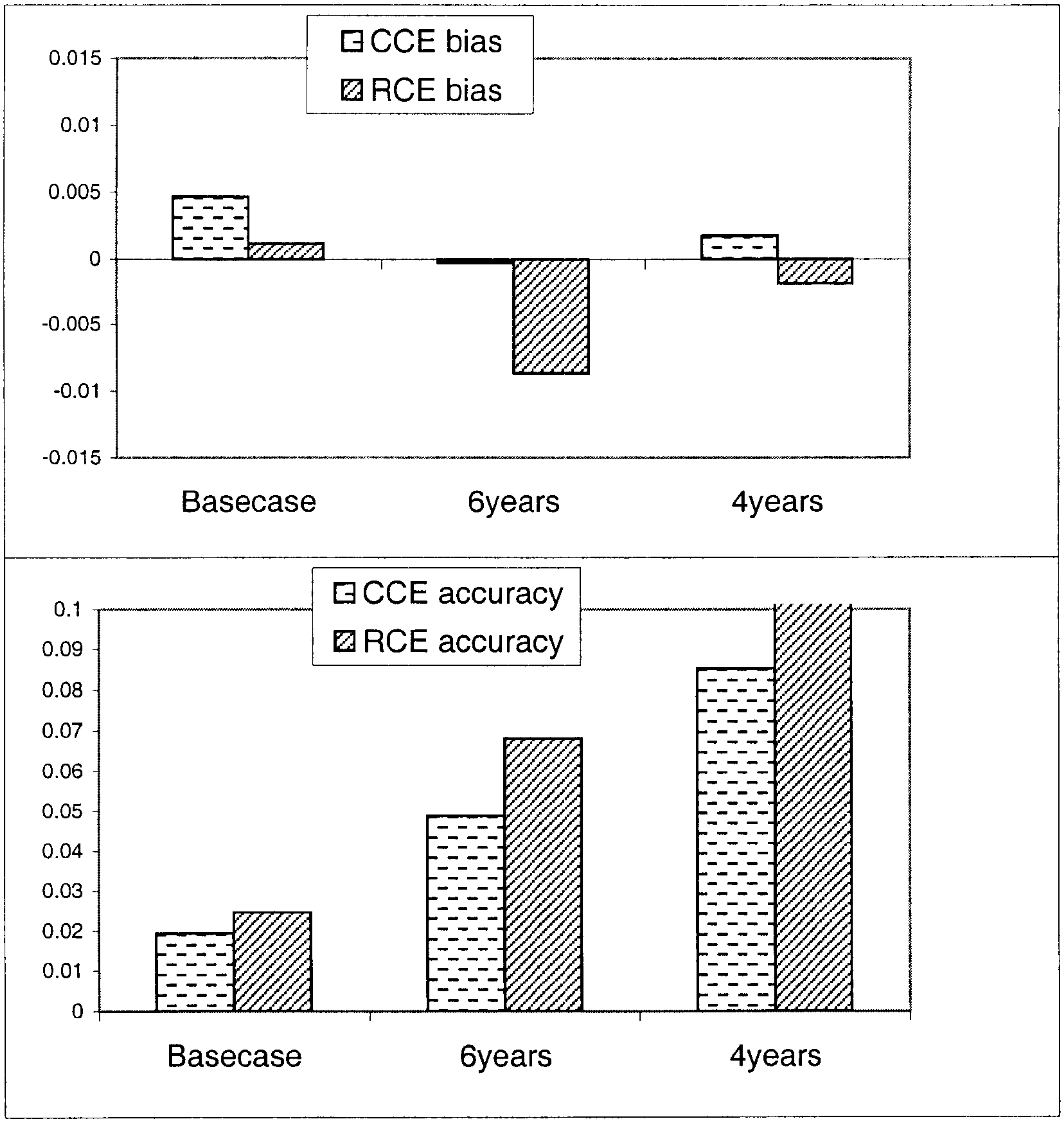

CCE bias
RCE bias
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
Basecase
6years
4years
CCE accuracy
RCE accuracy
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Basecase
6years
4years

Figure 10

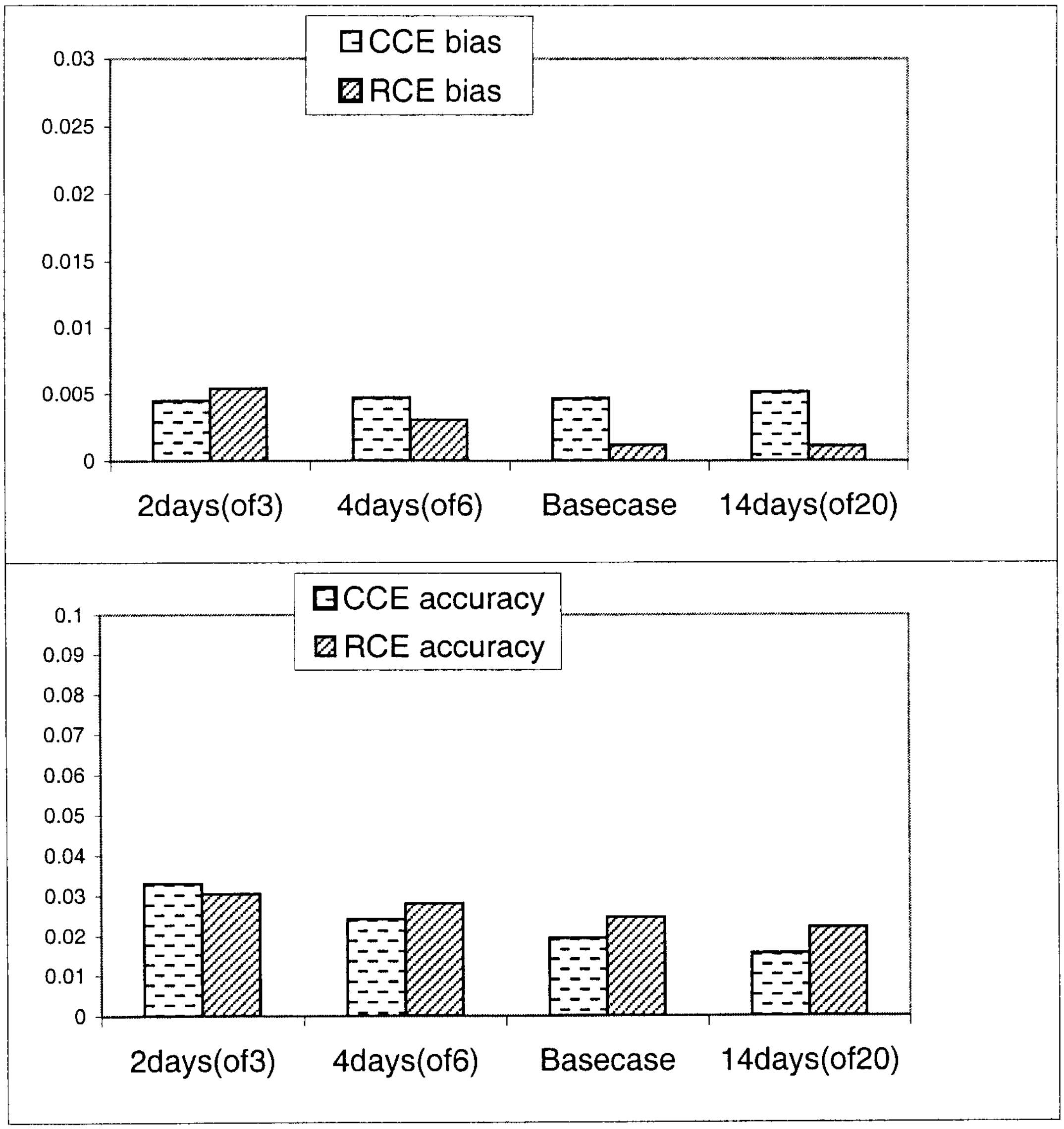

CCE bias
RCE bias
0.03
0.025
0.02
0.015
0.01
0.005
0
2days(of3)
4days(of6)
Basecase
14days(of20)
CCE accuracy
RCE accuracy
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
2days(of3)
4days(of6)
Basecase
14days(of20)

Figure 11

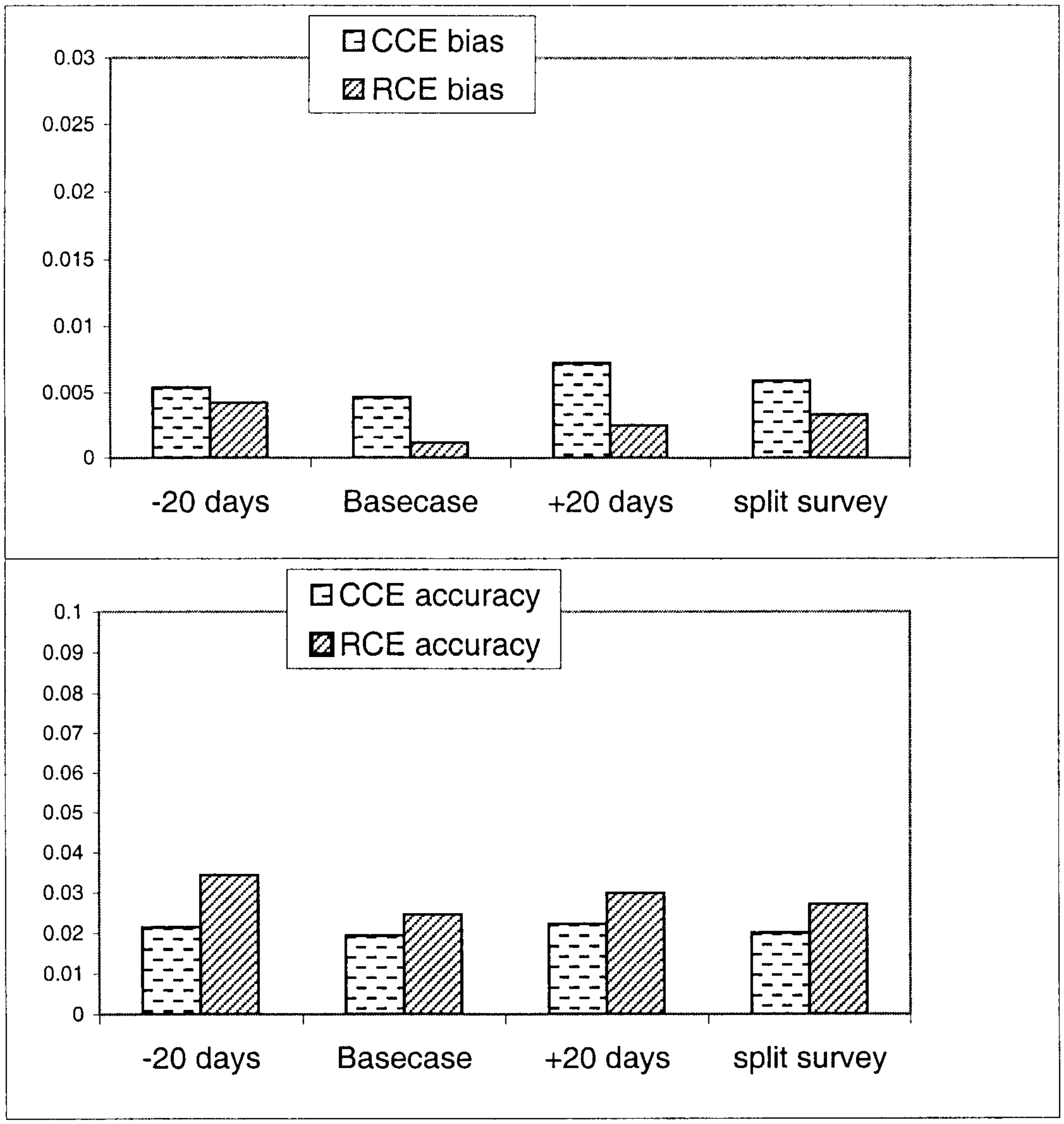
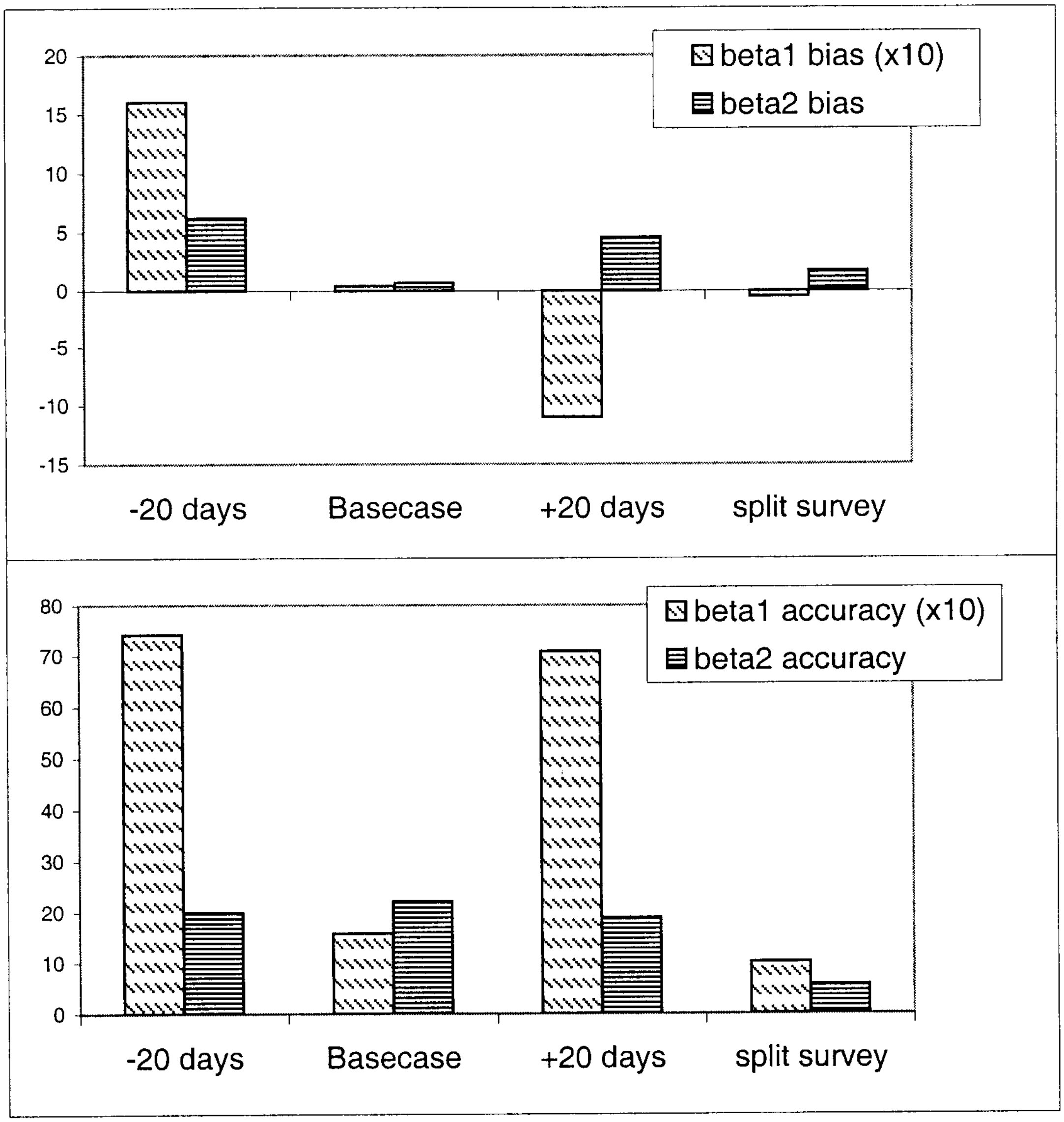

CCE bias
RCE bias
0.03
0.025
0.02
0.015
0.01
0.005
0
-20 days
Basecase
+20 days
split survey
CCE accuracy
RCE accuracy
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-20 days
Basecase
+20 days
split survey

Figure 12



20
15
10
5
0
-5
-10
-15
beta1 bias (x10)
beta2 bias
-20 days
Basecase
+20 days
split survey
80
70
60
50
40
30
20
10
0
beta1 accuracy (x10)
beta2 accuracy
-20 days
Basecase
+20 days
split survey

Figure 13

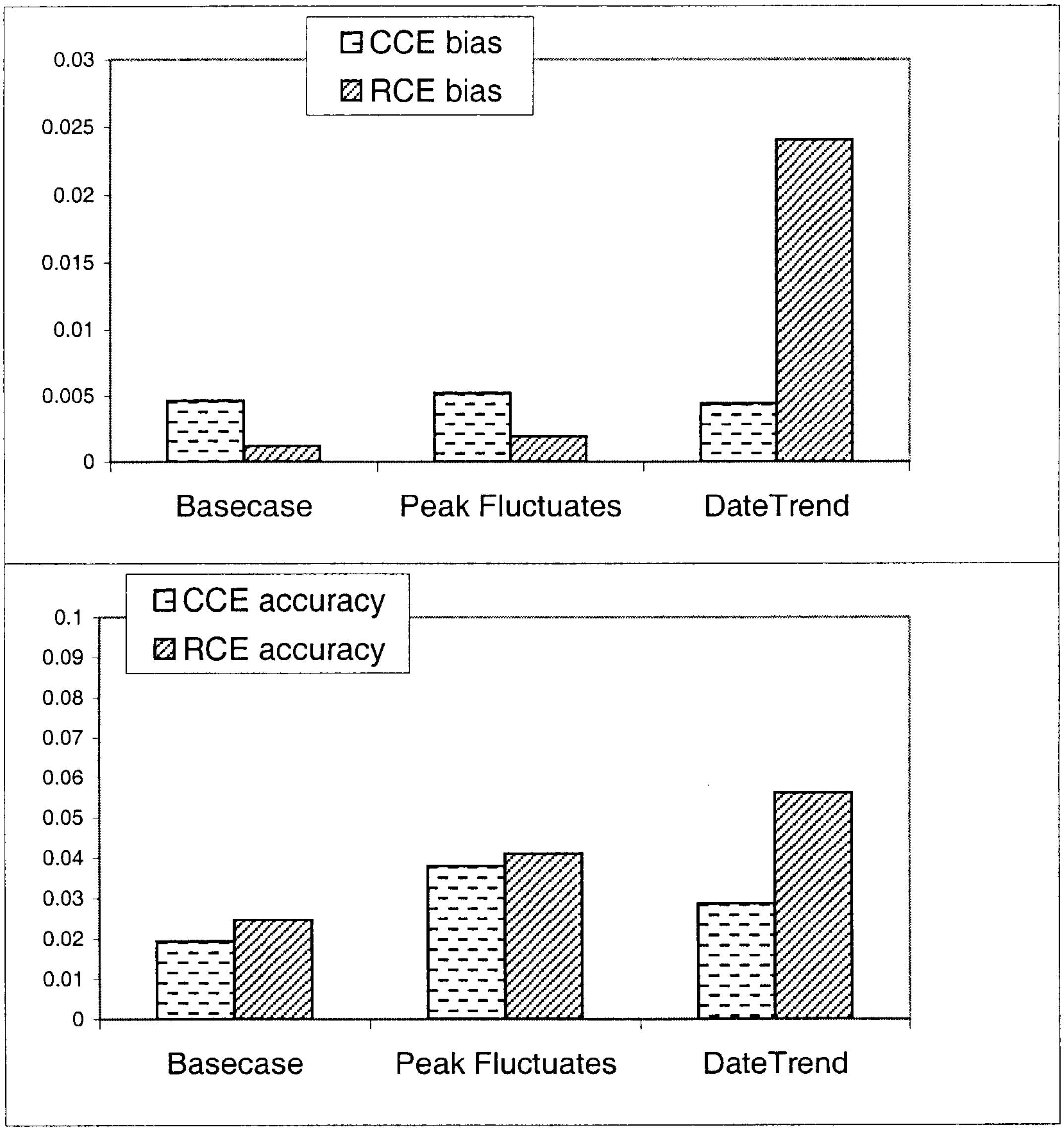

CCE bias
RCE bias
0.03
0.025
0.02
0.015
0.01
0.005
0
Basecase
Peak Fluctuates
DateTrend
CCE accuracy
RCE accuracy
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Basecase
Peak Fluctuates
DateTrend

Figure 14

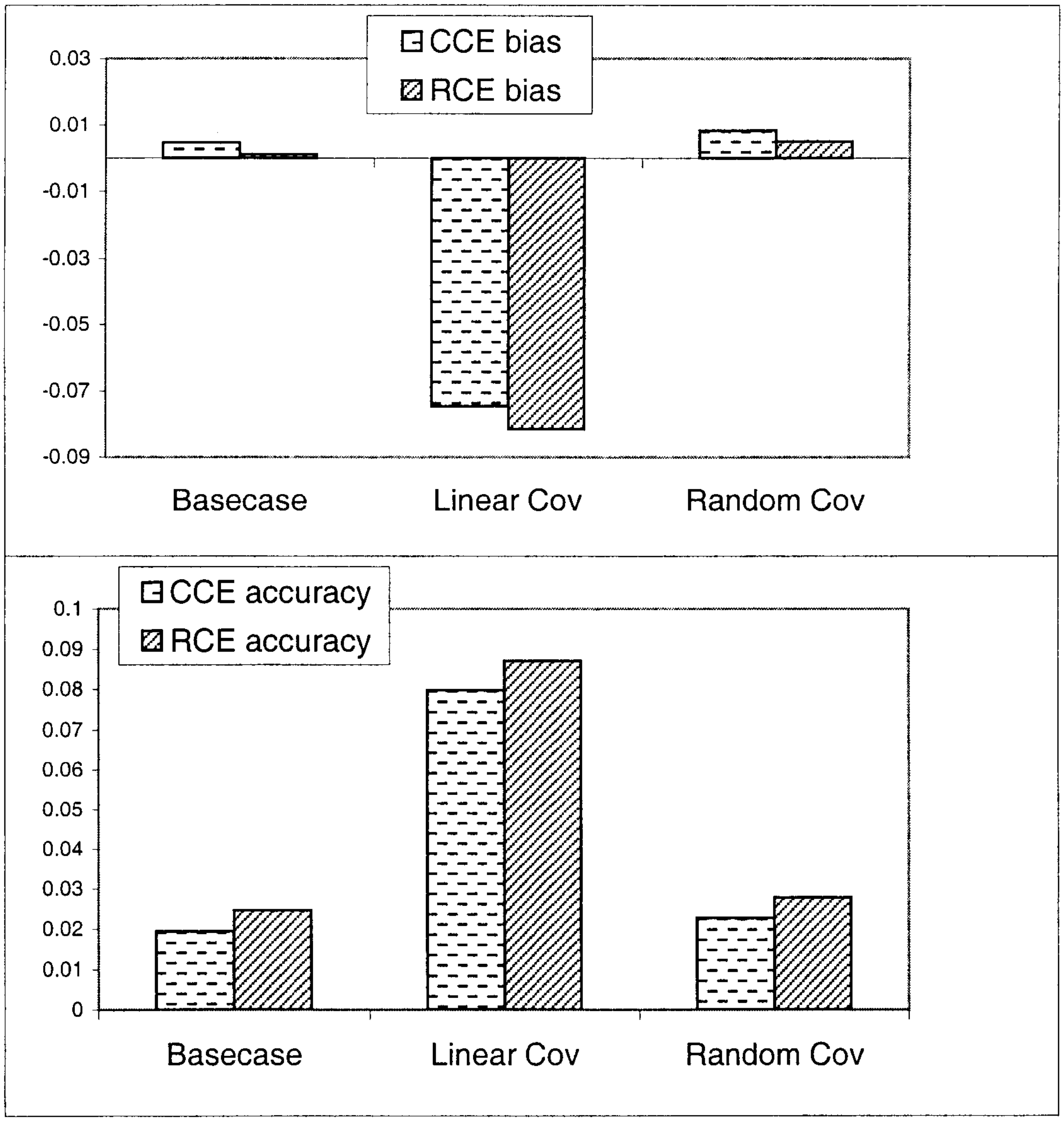

CCE bias
RCE bias
0.03
0.01
-0.01
-0.03
-0.05
-0.07
-0.09
Basecase
Linear Cov
Random Cov
CCE accuracy
RCE accuracy
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Basecase
Linear Cov
Random Cov

Figure 15

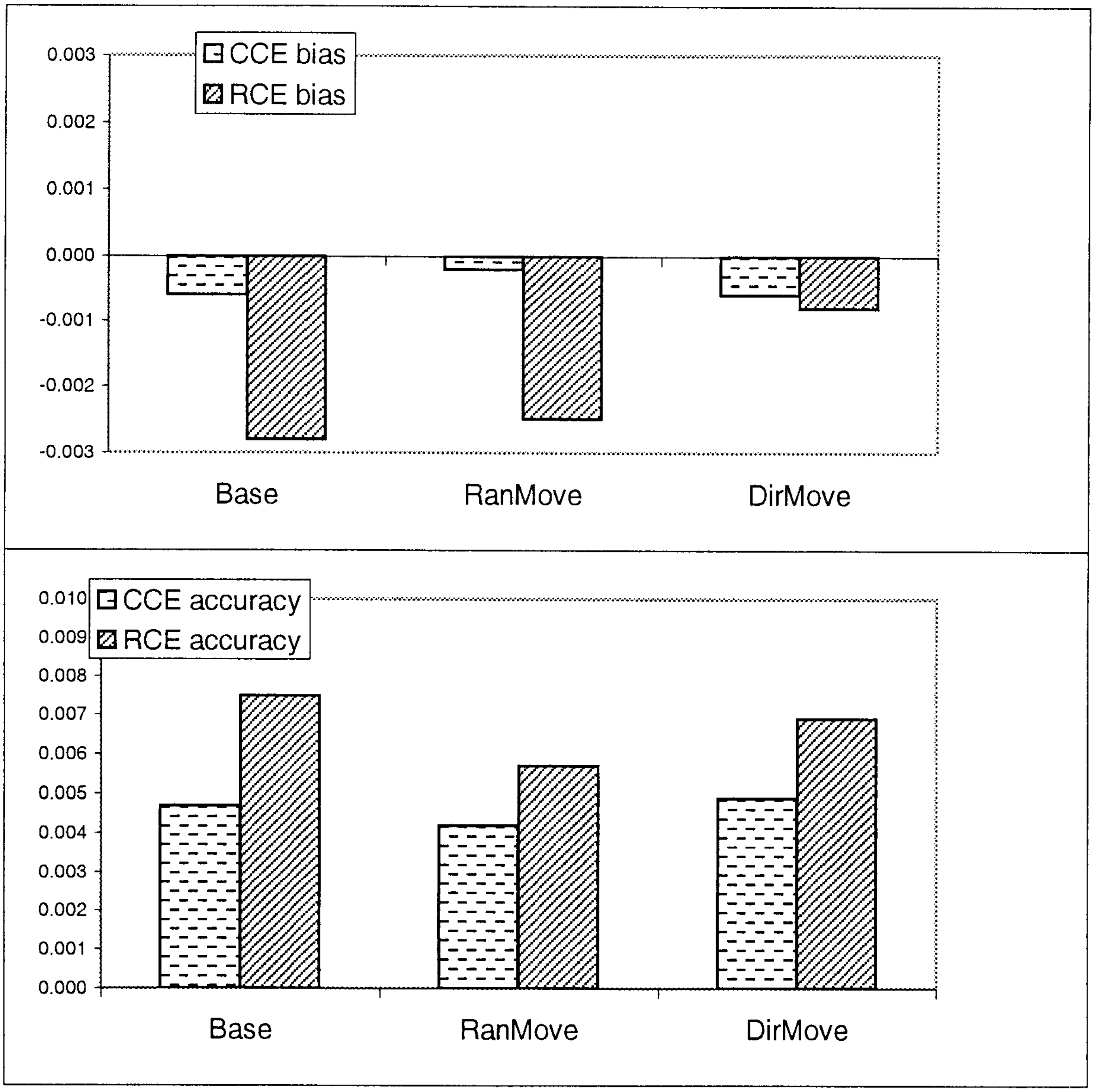
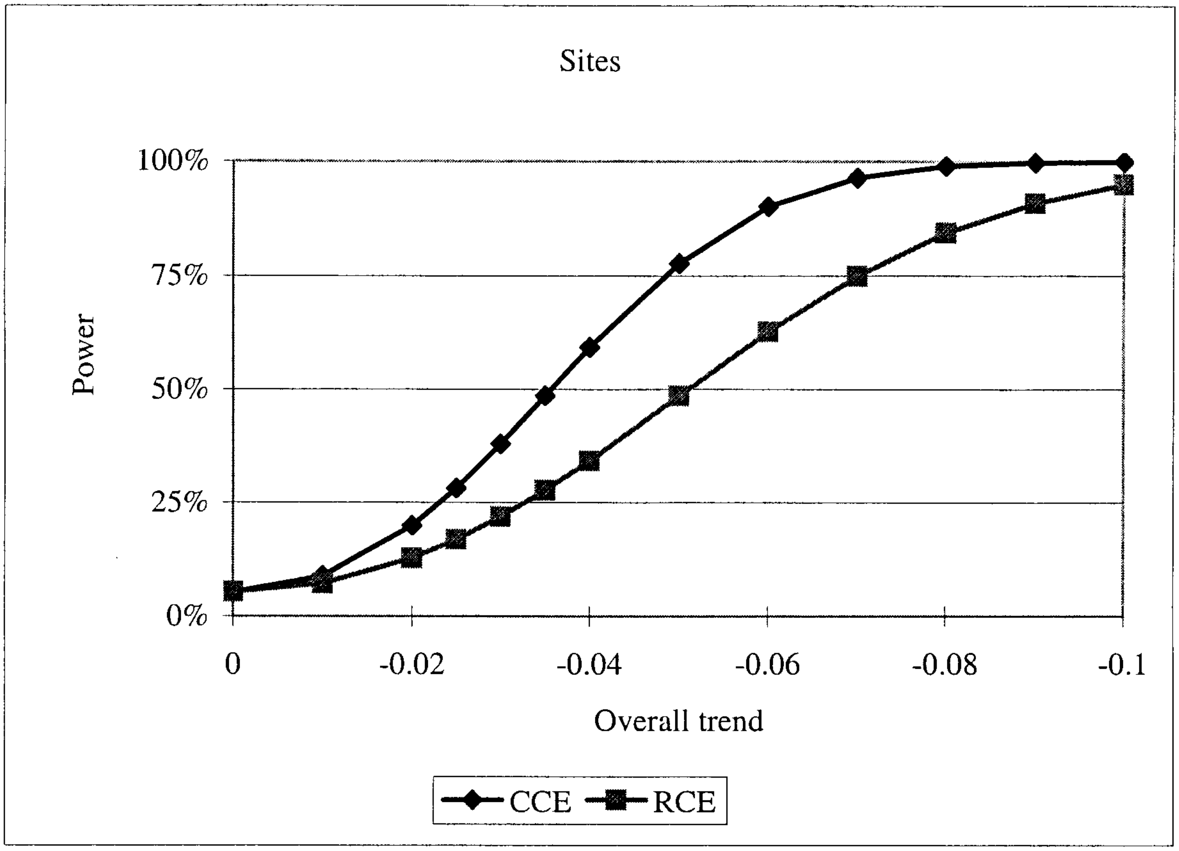

CCE bias
RCE bias
0.003
0.002
0.001
0.000
-0.001
-0.002
-0.003
Base
RanMove
DirMove
CCE accuracy
RCE accuracy
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
Base
RanMove
DirMove

Figure 16


Sites
Power
100%
75%
50%
25%
0%
0
-0.02
-0.04
-0.06
-0.08
-0.1
Overall trend
CCE
RCE

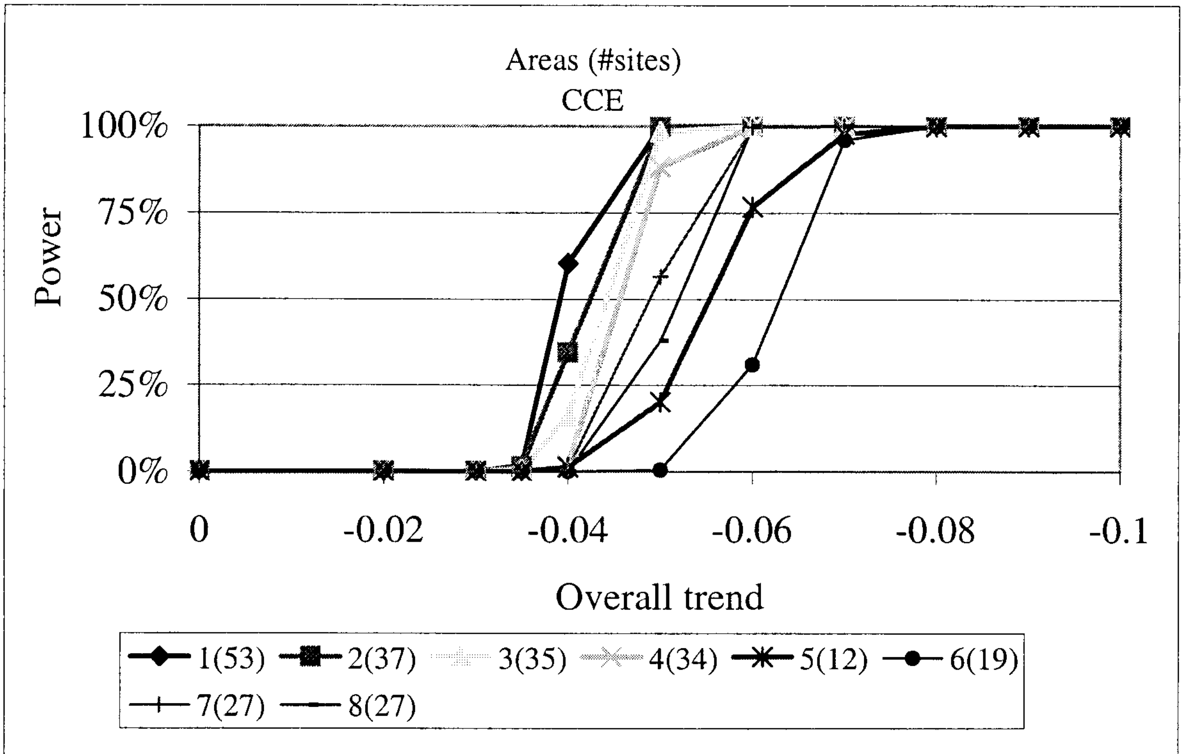
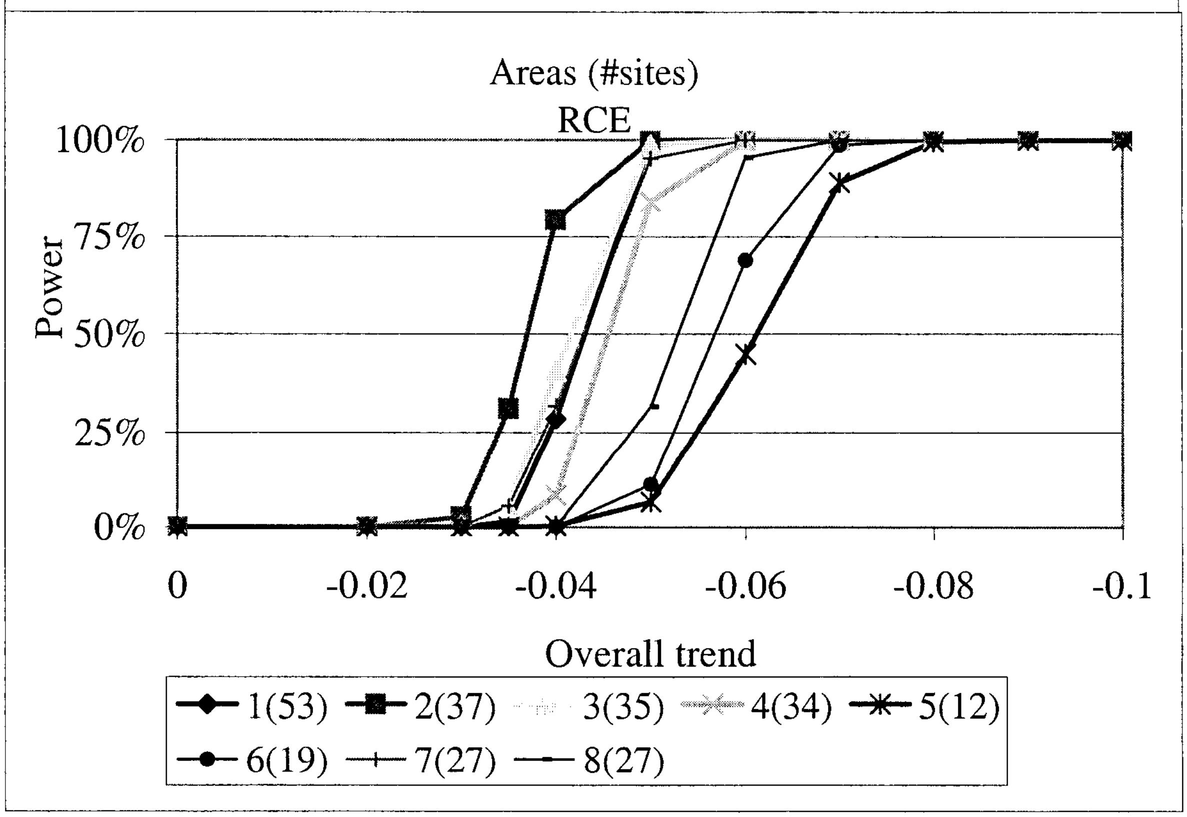

Areas (#sites)
CCE
100%
75%
50%
25%
0%
Power
0
-0.02
-0.04
-0.06
-0.08
-0.1
Overall trend
1(53)
2(37)
3(35)
4(34)
5(12)
6(19)
7(27)
8(27)

Areas (#sites)
RCE
100%
75%
50%
25%
0%
Power
0
-0.02
-0.04
-0.06
-0.08
-0.1
Overall trend
1(53)
2(37)
3(35)
4(34)
5(12)
6(19)
7(27)
8(27)

Figure 18

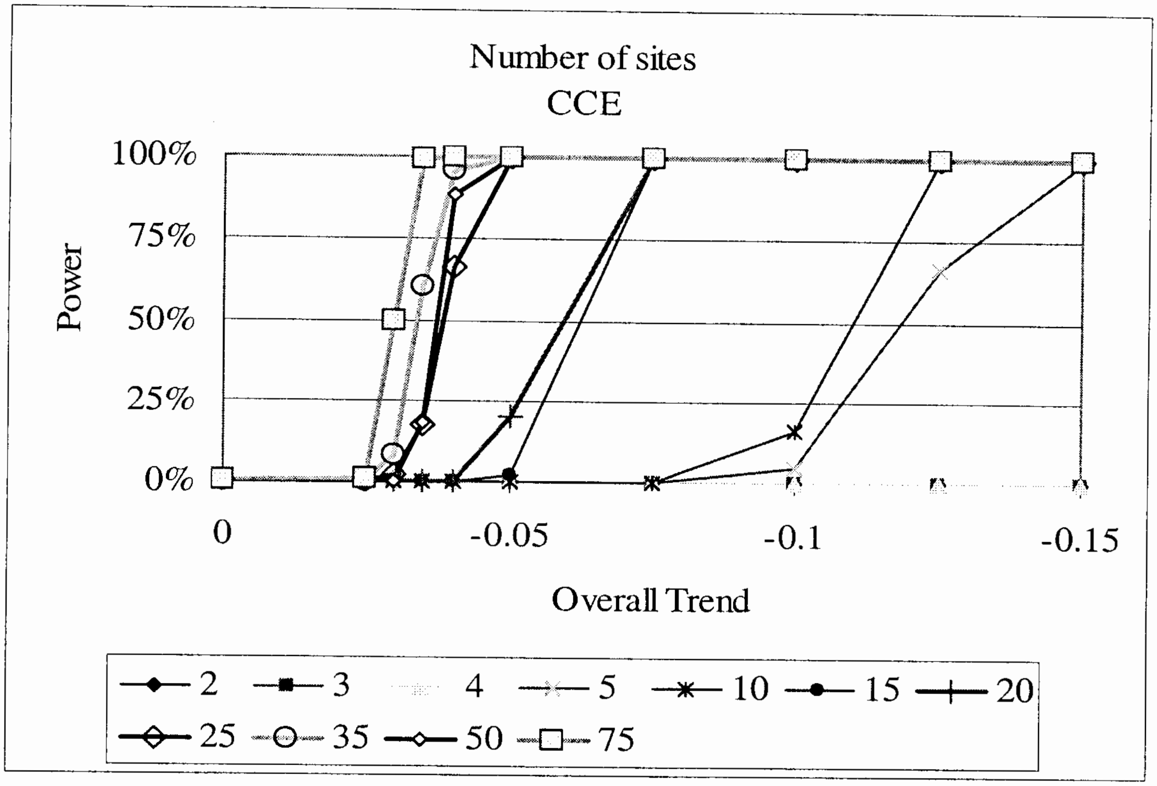

Number of sites
CCE
100%
75%
50%
25%
0%
Power
0
-0.05
-0.1
-0.15
Overall Trend
2
3
4
5
10
15
20
25
35
50
75

Figure 19