A comparative study of monitoring programs for coherence in quantifying the dynamics of American lobster fisheries in the state of Maine

Final report

Prepared by

Yong Chen, Jie Cao, Jui-Han Chang and Bai Li
School of Marine Sciences, University of Maine, Orono, Maine 04469

Carl Wilson
Marine Department of Marine Resources
West Boothbay Harbor, Maine

Award number: NA10OAR4170081

Dates of the award period: February 1, 2010 – September 30, 2014

Dates of the reporting period: February 1, 2010 – September 30, 2014

Name(s) of the grantee: University of Maine

Principal investigator: Yong Chen and Carl Wilson

Maine Sea Grant

Research Project Reporting Form

The Maine Sea Grant College Program's reporting period to the National Sea Grant Office and to the University of Maine's Vice-President for Research now corresponds with the fiscal year for Sea Grant. The current reporting period is February 1, 2010 -January 31, 2011, essentially 2010.

Please complete all sections of the form that apply to your Sea Grant or MOSAC/DEP project-related activities in the current reporting period and return the form electronically to Sea Grant at wardwell@maine.edu.

Principal investigator(s): *Yong Chen and Carl Wilson*

Project title: *A comparative study of different monitoring programs for coherence in quantifying the dynamics of American lobster fisheries in the State of Maine.*

Project number: *NA10OAR4170081*

Start date: *Febuary1, 2010*   End date: *September 30, 2014*

Type of report: ☐ Progress
☒ Completion

Table of Contents

Executive summary

Monitoring programs developed by the Mane Department of Marine Resources (DMR) for collecting data on the American lobster (*Homarus americanus*) fisheries in Maine vary greatly in their history, design, data collected, temporal and spatial scales of data collection, costs, utility of the data collected and their impacts on the lobster stock assessment and management. These programs also share similarities and overlaps in data collected and targeted populations. A careful evaluation of current designs can lead to improved monitoring program design and optimization of sampling efforts in the Maine DMR lobster program. This can eventually lead to merge of some sampling programs, resulting in the development cost effective sampling program for the American lobster in the state of Maine.

The overall scientific hypothesis we tested in this project is that various Maine DMR monitoring programs are effective in quantifying the temporal variability of the American lobster stock and fishery along the coast of Maine. The more specific objectives of this study include evaluating the design of the key monitoring programs for their effectiveness in quantifying spatial and temporal variability of the lobster fishery and comparing the relevant monitoring programs for coherence in quantifying the dynamics of lobster fishery. After various discussing with the stakeholders, we were focused on the following three tasks: (1) evaluating the port sampling program and comparing it with the newly developed dealers reporting program in quantifying the lobster landings in the state of Maine; (2) evaluating effectiveness of bottom trawl survey program in capturing temporal trends of lobster stock in the coastal Gulf of Maine; and (3) evaluating the effectiveness of lobster settlement program in quantifying the temporal variability of lobster pre-recruits. These three tasks were selected because catch statistics estimated in the port sampling program, abundance index derived from the bottom trawl survey program, and pre-recruit abundance index derived from the settlement program play critical roles in the Gulf of Maine (GOM) lobster stock assessment.

For task I, a comparison between the lobster port sampling and the dealer database was conducted by randomly subsampling dealer data and comparing port sampling expanded estimates. Initial analysis indicated that the required 10-dealers sampled each month were insufficient to characterize the expanded catch. Combined with the development of mandatory dealer reporting and a subset of harvester reporting (mandated 10% of all license holders), the department decided to suspend the port sampling program after the 2011 season. This change, allowed existing DMR to switch duties of staff to other areas of priority. The DMR has recorded direct savings of approximately $70,000 annually.

For task II, we evaluated the performance of six possible sampling designs to estimate the population abundance index for American lobster using computer simulations. These designs include simple random sampling (SRS), systematic sampling (SYS) and stratified random sampling with four stratification schemes (i.e., based on region, depth, sediment and region × depth). For the stratified random design with region and depth being used for stratification, we evaluated the performances of different strategies for allocating sampling efforts. Simulations were implemented on the “true” populations which were estimated annually from 2002 to 2008 for both spring and fall based on a general additive model developed in a separate study. Relative Estimation Error (REE), Relative Bias (RB) and design effect were used to measure the precision, accuracy and efficiency of mean estimation for different designs. On average, SYS tended to yield the most precise and efficient estimate of mean with specified sample size. However, its estimates tended to be biased and its performance varied with sample sizes and realizations of “true” population, thus changed with lobster distribution. Appropriate stratification, such as using depth to determine strata, significantly improved the precision and efficiency over SRS. Sediment, which is related to lobster distribution, was found to have little contribution to the improvement of the performance over SRS when it is used to determine strata. Also, allocating samples to each stratum based on variance or mean of previous year improved precision and efficiency. This study suggests that current design (i.e., region-depth stratified design) used in the survey had stable performance across years and seasons.

For task III, we evaluated the lobster settler sampling program which follows a fixed-station design to determine if this design can capture the temporal dynamics of settlers, in particular under large changes in spatial distribution of lobster in the last two decades. We compared the fixed-station design with a random sampling design. We developed a generalized additive model (GAM) to quantify the relationship between lobster settler abundance and habitat variables for the mid-coast region of the Gulf of Maine from1989 to 2012. The GAM model was then used to simulate “true” populations using the habitat variables. The two different sampling designs were applied to sample the simulated “true” populations. The fixed-station sampling design tended to under-estimate the “true” abundance, but could capture the temporal trend of settler abundance. A persistence index analysis suggests that the fixed-station design could identify inter-annual change of the Lobster settler abundance. This study suggests that fix-station design is effective in monitoring temporal changes in settler abundance, but could not be used for the estimates of absolute abundance of settlers.

List of figures

Figure 14. The mean of 1000 simulated true population for American Lobster at potential sampling stations from 1989 to 2012.

Figure 15. Temporal trends of means of fixed-station designg and random-station design from 1989 to 2012. The shadows represent 95% confidence interval. The random-station design was repeated 100 times for each given simulated population.

Figure 16. Performance index relative estimation error (%) of two sampling designs from 1989 to 2012.

Figure 17. Performance index relative bias (%) of two sampling designs from 1989 to 2012.

Figure 18. Mean Lobster density (# m^{-2}) from the settlement survey in the mid-coast region of the Gulf of Maine across years (1989 - 2013). The error bars on each time series represent the variability of Lobster density between sampling sites.

Figure 19. Persistence index matrix for given two years (1989 - 2013). The smaller persistence index ϖ value is, the greater persistence the fixed-station sampling obtains and the greater the power of differentiating inter-annual changes in the settler abundance is.

Figure 20. Temporal trend of means of fixed and random designs from 1989 to 2012. The random-station sampling process was only repeated once for each given simulated population. The shadows indicate 95% confidence intervals.

General introduction

Quantitative fisheries assessment plays a central role in fisheries management. The quality of stock assessment, which is large replies on the quality and quantity of fishery-dependent and fishery-independent data available, can determine if fisheries management is successful in achieving its management goals. Thus, monitoring programs that collect the data form the foundation of fisheries stock assessment and management.

Monitoring programs developed for collecting data on the American lobster (*Homarus americanus*) fisheries in Maine vary greatly in their history, design, data collected, temporal and spatial scales of data collection, costs, utility of the data collected and their impacts on the lobster stock assessment and management. These programs also share similarities and overlaps in data collected and targeted populations. A comparative study and careful evaluation of current designs can lead to improved monitoring program design and optimization of sampling efforts in the Maine DMR lobster program. This can eventually lead to merge of some sampling programs, resulting in the development cost effective sampling program for the American lobster in the state of Maine.

The overall scientific hypothesis we tested in this project is that various Maine DMR monitoring programs are effective in quantifying the temporal variability of the American lobster stock and fishery along the coast of Maine. The more specific objectives of this study include evaluating the design of the key monitoring programs for their effectiveness in quantifying spatial and temporal variability of the lobster fishery and comparing the relevant monitoring programs for coherence in quantifying the dynamics of lobster fishery. After various discussing with the stakeholders, we were focused on the following three tasks: (1) evaluating the port sampling program and comparing it with the newly developed dealers reporting program in quantifying the lobster landings in the state of Maine; (2) evaluating effectiveness of bottom trawl survey program in capturing temporal trends of lobster stock in the coastal Gulf of Maine; and (3) evaluating the effectiveness of lobster settlement program in quantifying the temporal variability of lobster pre-recruits. These three tasks were selected because catch statistics estimated in the port sampling program, abundance index derived from the bottom trawl survey program, and pre-recruit abundance index derived from the settlement program play critical roles in the GOM lobster stock assessment (ASMFC 2009).

We used computer simulation studies to evaluate the effectiveness of current design and sampling efforts in quantifying the lobster fishery. Statistical methods and relevant computer programs were developed to analyze the data for each program.

Task I: Evaluation of American Lobster Port Sampling

Initiated in August 1966, the program was designed to survey the Maine lobster fishery using a stratified multistage sampling program. This program allows for unbiased estimates of total catch and effort by strata. Monthly expanded estimates have been generated through this stratified sampling program from 1967 to present. Each month, 10 dealers are randomly selected from a list of potential buying stations that have been verified as buying from a minimum of five fishermen. On each selected sampling day, fishermen selling their catch at the dealer are interviewed for catch and effort information; the catch is counted, and a biological sub-sample of the catch is examined.

Monthly-expanded estimates are a function of pounds surveyed (LB), potential dealers open for the month (PD), potential days fishing (DF) and days sampled (DS): .

$$annual\_\exp anded\_estimate = \sum \frac{(LB)*(PD)*(DF)}{DS}$$

Over the 42-year time series, expansion factors DF and DS have varied without trend while the PD have changed by month and year. Independent of expansion factors, annual pounds surveyed have increased nearly three folds over the time-series.

A comparison between the lobster port sampling and the dealer database was conducted by randomly subsampling dealer data and comparing port sampling expanded estimates. Initial analysis indicated that the required 10-dealers sampled each month were insufficient to characterize the expanded catch (Fig. 1). At that time the DMR decided that the sample increase was beyond the capacity of the department to achieve. Combined with the development of mandatory dealer reporting and a subset of harvester reporting (mandated 10% of all license holders), the department decided to suspend the port sampling program after the 2011 season. This change, allowed existing DMR to switch duties of staff to other areas of priority. The DMR has recorded direct savings of approximately $70,000 annually.

Task II: An Evaluation of an Inshore Bottom Trawl Survey Design for American lobster

The objectives of this task are: (1) to evaluate the performance of current design (i.e., stratified random sampling) in terms of its accuracy, precision and efficiency by comparing with other possible sampling strategies; (2) to compare alternative allocations of sampling efforts for current stratified sampling design used in the survey; (3) to evaluate the robustness of evaluated sampling schemes over time in order to understand the impacts of lobster spatial dynamics resulting from possible environment changes on sampling strategies. A study such as this one is important to understand the overall performance of the current survey design for monitoring lobster and it could also provide knowledge for designing a fishery-independent survey.

II-1. Materials and methods

Maine-New Hampshire inshore trawl survey

The Maine-New Hampshire inshore trawl survey evaluated in this study is a biannual multiple-species fishery-independent survey conducted by the Maine Department of Marine Resources (DMR) each spring and fall since fall of 2000. It follows a stratified random design with four depth strata (9–37 m, 37−64 m, 64−100 m, and >100 m with 12 km offshore limit) and five longitudinal regions based on oceanographic and geological features (Fig. 2). A target of 115 sampling stations was designed for each survey and the number of sample size per stratum was apportioned according to its total area. Groundfish species are the main target species of this survey in its design. However, an estimate of abundance index for American lobster is also a primary sampling objective (Chen *et al*., 2006). The net is a modified version of shrimp net design used in Maine waters and designed to fish for a variety of near-bottom dwelling species without targeting any specific component.

Simulation of a "true" population

The spatial distribution of American lobster is influenced by many factors such as temperature (Aiken and Waddy 1986), salinity (Jury *et al*., 1994), and shelter availability (Wahle and Steneck 1991) and it differs greatly by season, sex, and size class (Chen *et al*., 2006). Chang *et al*., (2010) developed a habitat modeling approach for quantifying season-, size-, and sex-specific lobster distribution in the Gulf of Maine. They used a 2-stage general additive model (GAM), with a stage 1 GAM to estimate the probability of presence of lobsters and stage 2 GAM to estimate the lobster density and multiplied the 2 stage model results to estimate the comprehensive lobster density. The model results suggested that lobster distribution was strongly associated with temperature and depth and different seasonally by sex and size classes, which are consistent with the ecology of the American lobster. In this study, the GAM models with bottom temperature, bottom salinity, latitude, longitude, depth, distance offshore, and two substratum features as the explanatory variables were used to estimate the season-, size-, and sex-specific lobster density distribution from 2002 to 2008. The model predictions were summed over size and sex to produce the spatial distribution of total lobster density (per tow) for spring and fall of each year from 2002 to 2008. We considered these time-series spatial distributions as "true" populations in evaluating alternative sampling designs. These "true" populations changed over time with respect to changes in temperature and salinity variables (see details in Chang *et al*., 2010). The temperature and salinity information for 2002 to 2008 was produced by the Gulf of Maine Ocean Observing System circulation nowcast/forecast system (Xue *et al*., 2005).

Simulation survey designs

The 3698 potential sampling stations generated by overlaying 1 nautical mile (NM) × 1(NM) grids over the survey area were considered as the sampling frame of this study. Areas that could not be towed were excluded (Fig. 2). Three types of sampling designs were considered:

- SRS: *n* stations of the potential 3698 sites were randomly selected and sampled;
- Stratified random sampling: four stratification schemes were defined, including four depths, five regions, seven sediments (i.e., gravel, gravel-sand, sand, clay-silt/sand, sand-clay/silt, clay, and sand/silt/clay), and four depths × five regions, and *n* stations

were allocated proportionally to the size of the strata. The stratified mean $\bar{y}_{\text{str}}$ was estimated by taking the weighted mean over all strata (Lohr 2009):

$$\bar{y}_{\text{str}} = \sum_{h=1}^{H} \frac{N_h}{N} \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h}$$

where n_h is the number of stations sampled in stratum h, $n = \sum_{h=1}^{H} n_h$ is the total number of stations sampled, N_h is the total number of possible stations in stratum h, H is the number of strata, $N = \sum_{h=1}^{H} N_h$ is the total number of possible stations in the survey area, and y_{hi} is the number of lobster density in station i of stratum h.

- SYS: the first station was randomly selected from the total of 3698 grids and the remaining $n - 1$ stations were evenly spaced in the survey area.

Based on the above three designs, a total of six survey designs were evaluated in this study (Table 1).

For the stratified survey design currently used by Maine DMR, Neyman allocation scheme was used to evaluate if such an approach can improve the precision of estimates. Neyman allocation is the special case of optimal allocation when the costs in the strata are approximately equal (Lohr 2009). The sample size in the stratum, n_h, is proportional to $N_h S_h$, where S_h is the variance of stratum h (Lohr 2009). S_h was assumed to be equal to the population variance of the previous year in stratum h which is estimated based on habitat model. In this case, we allocated more sample to highly variable strata and large strata of the previous year. Also we considered the case that n_h is just proportional to S_h which means we just allocate more samples to highly variable strata forecasted by previous year. In most fisheries surveys, mean and variance are related (Smith and Lundy 2006). Therefore, we also investigated the two allocation schemes with mean substituted for variance. Thus we considered four scenarios of sample allocations for the survey design currently used by the Maine DMR:

- Scenario 1: allocating samples based on variances of strata weighted by area
- Scenario 2: allocating samples just based on variances of strata
- Scenario 3: allocating samples based on means of strata weighted by area
- Scenario 4: allocating samples just based on means of strata

Evaluating survey designs

Three indices were used to measure the performance (e.g., accuracy, precision and efficiency) of each sampling scheme. Relative Estimation Error (REE) was used to quantify the accuracy and precision of estimated mean (Chen 1996):

$$REE = \frac{\sqrt{\frac{\sum_{i=1}^{N} (Y_i^{estimated} - Y^{true})^2}{N}}}{Y^{true}} 100\%$$

We also calculated Relative Bias (RB) for the estimated mean as (Paloheimo and Chen 1996):

$$RB = \frac{\frac{\sum_{i=1}^{N} Y_i^{estimated}}{N} - Y^{true}}{Y^{true}} 100\%$$

where $Y_i^{estimated}$ is the estimated mean in the ith simulated survey, Y^{true} is the true mean, N is the number of simulation times. The REE and RB values reflect both bias and variation in the estimation, and a smaller REE or RB value suggests a better performance (Chen 1996). The RB value could also indicate whether the sampling design tends to underestimate or overestimate the population mean.

The variance of sample mean of each sampling strategy was calculated from the distribution of sample mean generated by repeating the sampling process on the "true" population. Such a variance reflects the variability of sample mean. In theory, the sampling designs considered in this study produce unbiased estimates of population mean. However the unbiasedness does not mean that estimate of mean for a particular simulation run would be equal to the true population mean. Rather, the unbiased estimators have variability; sometimes they would be too low or too high. If the estimates of mean are too variable based on certain design, it would be considered of low precision and less efficient. Design effect, $deff(\overline{m})$, was used to quantify the difference of sample-to-sample variability between a specified sampling design and SRS:

$$deff(\overline{m}) = \frac{V(k, \overline{m})}{V(SRS, \overline{m})}$$

where $\overline{m}$ is the sample mean, $V(k, \overline{m})$ is the variance of sample mean under the k^{th} sampling design, $V(SRS, \overline{m})$ is the variance of sample mean under the SRS design.

Simulation procedure

The sampling process was simulated for each design by spring and fall from 2002 to 2008 based on the "true populations". For Design I to Design VI (Table 1), three sample sizes were considered (87, 115, 144) in order to test the impacts of sample size. Simulations could be divided into two steps for each sampling design: (1) draw samples according to a particular design from the "true" population for 1000 times and calculate each performance index; and (2) repeat step 1 for 100 times to capture variability in the simulation and get the distribution of performance indices.

II-2. Results

Simulated populations

The predicted spatial pattern of lobster distribution was stable over time for both spring and fall from 2002 to 2008, therefore, only the distributions of 2006 were shown as an example (Fig. 3). In general, lobster density was predicted to be higher in inshore waters. The hot spots were located in the mid-coast region. Those patterns were similar for both spring and fall.

Survey designs

The values of REE and design effect showed consistently that the performance (i.e., efficiency and precision) of the six survey designs had the following ranking (from best to worst): Design II > Design VI > Design IV > Design V > Design III > Design I. This performance ranking was the same for both spring and fall populations, for different sample sizes (i.e., 87, 115 and 144; Table 2, Fig. 4), and for different years (i.e., from 2002 to 2008; Table 2,

Fig. 5). The same pattern in these two indices was apparent for the spring population (not illustrated).

SYS yielded the most precise and efficient estimates of population mean. However, its performance indices (e.g., REE and design effect) showed large variation with the change of sample size (Table 2, Fig. 4). For example, the annual average REE for fall population decreased from 7.11% to 4.92% when sample size increased from 87 to 115. However it increased from 4.92% to 6.49% when sample size increased from 115 to 144. The annual average design effect showed the same pattern. Thus, increased sample size might lead to decreased performances for SYS. Such variation in the design effects and REE also existed in the spring population. In addition, the design effect of SYS differed for spring and fall population with the same sample size, suggesting SYS is likely to be sensitive to different realization of population spatial distribution (Table 2).

The current region-depth stratified design used by DMR performed slightly better in annual average design effect and REE compared to the depth-stratified design alone, when the same sample size was the used. Stratification by regions only contributes a little to the improvement of the efficiency since it just resulted in less than 10% improvement in design effect. Most of the improved efficiency due to the current depth-region-based 20-strata design came from the depth component of the stratification scheme. Sediment-stratified design had the similar efficiency as region-based design. However, its performance differed by season (Table 2). The REE and design effect obtained by region-stratified design and sediment-stratified design were close (Table 2, Fig. 4). The annual average REE values of stratified designs for the estimation of both spring and fall populations decreased as the sample size became larger (Fig. 4). Such a decrease was gained by increasing sample size from 87 to 115 and was larger than the decrease resulting from increasing sample size from 115 to 144. The improvement of precision by increasing the sample size varied with different designs.

The RB values of all the designs except for Design II were distributed evenly around zero (e.g., annual means were less than 0.1%) for any given sample size and population which indicates that the biases of these designs have no tendency to be either negative or positive (Fig. 6). However, for SYS biases tended to be positive consistently across all the years for both spring and fall populations when sample size was 87 and tended to be negative when sample size increased (Fig. 6). The annual average RB values of SYS had relatively large variation with the change of sample size. The SYS might yield overestimated or underestimated population mean compared to the other sampling designs.

The variations of REE and RB between different years are shown in Figs. 5 and 7. The patterns of REE across the years associated with six sampling designs are almost the same and variations are relative small (Fig. 5). The values of RB across years are stable and show no bias on average for all the sampling designs except SYS (Fig. 7), indicating that different realizations of population distribution in the simulation might not exert a large influence on the performance of those sampling designs.

In conclusion, SYS gave the most precise and efficient estimates of population mean; however, these estimates were biased. Its precision differed by season and its bias varied across years. Stratified design produced unbiased estimates and its precision and efficiency depends on the stratification strategy. All the stratification strategies evaluated had stable performance across years and seasons except sediment-stratified design whose performance varied with season. However, season-specific performance of sediment-stratified strategy was stable across years.

Sample allocations

The results of reallocating samples for Design VI showed that REE of four scenarios reduced by about 2%, suggesting that reallocating samples based on variance or mean of previous year only improved precision slightly. The design effects of the four scenarios decreased by 20% for the years from 2003 to 2008 of both fall and spring populations, suggesting that reallocating samples improved efficiency by about 20%. The RB values were so small (less than 0.1%) that they could be ignored. The performances of the four scenarios are shown in Table 3. Scenario 1 and 2 performed best for both spring and fall population through all the years and the performance indices of those two scenarios are very close.

For the fall population, Scenario 2 performed best for the years of 2003, 2005 and 2008 in which both the REE and design effect were smallest (Table 3). For the year of 2006 the performance of Scenarios 3 was the best. For the year of 2004 and 2007 Scenario 1 had the highest precision and efficiency. For the spring population, best scenarios were not consistently suggested by the values of REE and design effect. However the values of those two indices were very close (Table 3).Scenario 4 did not perform well in any given scenario for both spring and fall populations. Scenario 3 only performed best in the year of 2006 for the fall population. Variance or weighted variance of immediately previous year was a better indicator for allocating samples to each stratum than the mean.

II-3. Discussion

The performance of several sampling designs and different sample sizes in their ability of estimating abundance indices in fishery-independent surveys especially for benthic invertebrate species was examined in several studies (Cabral and Murta 2004; Smith and Lundy 2006; Smith and Tremblay 2003). Although these studies generated insights about performance of various survey designs and sample size, either the number of designs involved or the number of "true" spatial distributions was limited. In this study, the bias relative to the "true" population value and precision and efficiency relative to the variance obtained by SRS, stratified random sampling and SYS were compared, and alternative sampling effort allocation schemes were explored and evaluated using computer simulation based on 14 "true" populations (e.g., spring and fall population each year from 2002 to 2008).

The currently used stratified random sampling design was on average less precise and efficient for estimating population mean than SYS for both spring and fall surveys in all the years investigated. This is consistent with previous studies which suggest that SYS tends to be more accurate than stratified random design (Cochran 2007; Ripley 2004). The desirable properties of SYS generally embodies in providing better support for kriging methods which aim to obtain estimates of spatial distribution (Liu *et al*., 2009). In this study, SYS was demonstrated to out-perform random and stratified random designs for estimating the population mean in terms of precision and efficiency. However, given that RB is non zero for SYS on average, this suggests that the sampling design either over- or under-estimates 'true' population mean. Another striking feature of SYS in this study is that precision and design effect are not always improved with increased sample sizes. For example, increasing the sample size by 25% from 115 to 144 would not reduce sampling errors, rather would actually increase the REE and design effect by 32% and 100%, respectively. REE and variance of sample mean did decrease when the sample size approached the entire population globally (Fig. 8). However, local behaviors of REE and variance were complex. There are some major peaks and a lot of small fluctuations in the curves of REE and variance versus sample size (Fig. 8). Therefore, it is difficult to select a specific sample size to reach a specified performance for a SYS design.

Stratified random design can spread out the samples and often improve the precision and efficiency of survey means compared to SRS (Lohr 2009). However, this study demonstrates that stratification, if determined inappropriately, such as only using regions to determine strata, makes little contribution to the improvement of precision and efficiency. It is critical to select suitable variables to determine strata. Variables that may greatly influence spatial distribution and population structure of target species are considered to be good choices because the strata determined by such variables tend to make homogeneity within a stratum and heterogeneity between stratum. For example, the stratification based on depth in this study improved the efficiency and precision greatly over SRS. Previous studies revealed that lobster distribution and size composition vary with water depth (Chen *et al*., 2006; Wahle and Steneck 1991). However, the stratification based on sediment, which is another variable used in the GAM model for generating the 'true' population, did not improve the performance over SRS as much as depth-stratified design did. Although studies have revealed that high lobster density occurs in substrates with boulders (Cooper and Uzmann 1980) and rocks (Steneck 2006). Due to the limitation of gear type used in this trawl survey such substrates had limited coverage the trawl survey. Also, variable sediment is not as significant as variable depth in the GAM developed by Chang *et al*., (2010). This study suggests that no all variables that may influence spatial distribution of lobster are suitable for survey stratficiation.

Reallocating samples among strata can significantly improve the ability of estimating population mean. A reduction of 20% samples from the current sample size (115) could obtain similar precision and efficiency for estimating population mean by reallocating the sampling efforts based on the variances estimated in the previous year. The four scenarios considered in this study yield improvement in efficiency and precision, indicating that variance and mean might be correlated. However, variance tends to be better than mean as an indicator of allocating samples among the strata. The difference between scenario 1 and scenario 2 is that variance used in scenario 1 is weighted by area. The impact of area weight to variance was related to how well the variance of previous year predicting the next’s. For the years that scenario 1 outweighs scenario 2, the reason is that the weighted variances of previous year are more approaching the true variances than those un-weighted.

The current stratified sampling design was found to be robust to different realizations of lobster population, and its performance was stable between seasons and among years. This suggests that the change in temporal and spatial distributions driven by environmental factors such as bottom temperature and salinity has no effect on the ability of appropriate stratified sampling design to estimate the mean. Smith (1996) simulated two very different populations (with and without spatial structure) to show that the underlying distribution and spatial structure of population have no effect on the performance of stratified sampling design in estimating mean and its standard error. Our study is consistent with his study. Such a result indicates that the relative abundance trend of lobster could be well tracked based on the current design without any standardization.

For a fishery independent survey targeting multiple species as the one evaluated in this study stratified random design is more appropriate. Because different species tend to have different spatial distributions, SYS may perform well on one or some, but not all. Additionally, it's hard to decide a particular sample size for SYS since its performance could dramatically fluctuate with small change of sample size. However appropriate stratified random design is robust to different distributions. Given the variability in fish population distribution over time and space and nature of targeting multispecies in a fishery-independent survey program, stratified random survey design is more desirable for a fishery-independent survey. Defining the sampling frame is a critical issue in a fishery-independent survey. For example, the size of sampling unit can influence the performance of certain sampling designs (Pennington and Volstad 1991). In this study the sampling unit was defined as 1NM × 1NM and some potential sampling units were excluded due to the operability of gear type. Studies may be needed to evaluate the impacts of sampling frame on the inshore bottom trawl survey for the American lobster.

This study suggests that stratified random survey design used in the Maine bottom trawl survey can yield abundance index estimates that can reliably capture spatial and temporal variability of American lobster population along the coast of Maine covered by the survey program. The use of the abundance index in the lobster stock assessment (ASMFC 2009) is thus desirable. Similar approach used in this study can also be used for other fish species to evaluate the reliability of abundance index derived from a fishery-independent survey program in capturing fish stock dynamics in stock assessment.

Task III: Evaluation of Effectiveness of Fixed-station Sampling for Monitoring American Lobster Settlement

The objective of this study is to evaluate if the fixed-station sampling design that has been used in the American lobster settlement survey can capture the temporal dynamics of lobster settlers. Specifically, we simulated temporal "true" populations of the distribution of newly settled lobster in the mid-coast region of the GOM based on a two-stage GAM model; and then applied both fixed and random designs to sample the simulated population. We compared the estimated and "true" population abundance for both the fixed and random survey designs to calculate estimation error. The estimation errors were then compared between the fixed and random survey designs to determine their performance in capturing the temporal variability of the lobster settlers. Additionally, persistence indices were calculated to evaluate the fixed-station sampling's power of detecting temporal trend of the lobster density.

III-1. Methods and materials

Lobster settlement data

The American lobster settlement index program is designed for monitoring annual density of newly settled lobsters (i.e., YOY lobster settlers and juveniles) (Wahle et al. 2010) and it is the target survey program to be evaluated in this study. The program follows a fixed-station design and covers coastal areas from Nova Scotia to Rhode Island. This diver-based suction sampling was conducted in all the sites at the end of the settlement season for each year. Divers collected newly settled lobster from 12 to 20 0.5 m^2 quadrats by using an air-lift suction sampler (Pershing et al. 2012). The size of newly settled lobster varied between 10.5 to 60 mm CL (Chang et al. 2010). For this study, mid-coast region of GOM was chosen as the study area as it has one of longest time series data among the whole coastal areas (Fig. 9). In this region, 10 fixed sampling stations are revisited every year (Fig. 9). The information collected from these 10 sites from 1989 to 2013 was used in the analytical analysis of persistent index (Warren 1994).

Environmental and spatial data

The two-stage generalized additive model (GAM) was developed to quantify the distribution of newly settled lobster using environmental and spatial variables. Bottom water temperature, salinity, latitude, longitude, depth, distance offshore, sediment type, and distance to sediment boundary were identified as environmental variables influencing the distribution of lobster (ASMFC, 2009; Chang et al. 2010), and they were included in the GAM in this study. The bottom water temperature (°C) and salinity for trawl survey stations were measured directly during the survey. The bottom water temperature and salinity data from 1989 to 2012 for potential sampling stations of the mid-coast region in the simulation were obtained through spatially interpolating the data from Finite-Volume, primitive equation Community Ocean Model (FVCOM). We extracted the depth (m) data from U.S. Coastal Relief Model for Northeast Atlantic region. The sediment information was gathered from the map of sediment grain-size distribution for U.S. east coast (Continental Margin Mapping Program). The distance to sediment boundary was calculated using ArcGIS 10.3 (Chang et al. 2010).

Two-stage generalized additive model

A two-stage GAM was developed to predict the lobster abundance distribution along the mid-coast region of the GOM. Following Chang et al. (2010), the environmental variables built in the model include water temperature (*T*, °C), salinity (*S*), settlement type (*Se*), depth (*D*, m), distance offshore (*DO*, degree), distance to the settlement boundary (*DS*, degree), latitude (*La*), and longitude (*Lo*). The first stage GAM estimates the presence of lobsters (p) by using a logic link function with a binomial error distribution. The second stage GAM estimates the log-transformed lobster density (d) by using an identity link function with a Gaussian error distribution (Berry and Welsh 2002). The comprehensive log-transformed lobster density (y) was estimated by multiplying the results generated from the first and second stages of the GAM,

GAM1: $logit\ (p) = s(T) + s(S) + s(De) + s(DO) + s(DS) + s(La) + s(Lo) + Se + \varepsilon$

GAM2: $ln\ (d) = s(T) + s(S) + s(De) + s(DO) + s(DS) + s(La) + s(Lo) + Se + \varepsilon$

$\ln(y) = p \times ln(d)$

where s is spline smoother.

The two-stage GAM was evaluated by degrees of freedom for each environmental or spatial variable. We conducted a preliminary analysis to evaluate the significance of variables in achieving the best fit with fewest independent variables. Eight variables were included initially in the each stage of GAM, then the most significant terms were selected and included as the main effects in the models ($p < 0.05$; Chang et al. 2010).

The performance of the derived model was evaluated using cross-validation approach (Franklin and Miller 2009). We divided the fall trawl survey and environmental data into training and testing data sets prior to modeling for model validation. The partitioning of training and testing data sets are random and based on a ratio of 3:1 since the number of predictors are more than five (Franklin and Miller 2009). We compared the lobster abundance ln(y) predicted based on the model developed using training data with the observed lobster abundance ln(y') of the testing data by using following simple linear-regression model:

$$ln(y') = a + bln(y)$$

We ran the cross-validation 100 times and averaged the estimated performance measures (Fieldling and Bell 1997). The averaged a and b values indicate bias in predicted abundance. An $a = 0$ and $b = 1$ imply that predicted lobster abundance and observed lobster abundance (i.e., testing data) have similar spatial patterns and the model has good predictive performance.

Simulation approach

The spatial distributions of newly settled lobster in the mid-coast region from 1989 to 2012 were predicted using the GAM model developed based on the inshore trawl survey data. These distributions were considered as ‘true’ populations for applying fixed and random sampling schemes. There are 1971 potential sampling stations identified in the mid-coast region of the GOM (Fig. 9). The GAM model yielded the prediction of lobster abundance and associated standard deviation for each potential sampling station for each year from 1989 to 2012. For each year, 1000 realizations of the ‘true’ population were generated based on the variation of the predicted lobster abundance among potential sampling stations.

Both fixed and random sampling designs were applied to the 1000 realizations of the ‘true’ population each year with the sample size of 10. For the fixed sampling scheme, 10 stations out of the 1971 potential sampling stations that are closest to the 10 fixed-stations used in the actual settlement survey were selected. For the random sampling scheme, sampling process was repeated for 100 times for a given realization of ‘true’ population. As a result, the two sampling schemes yielded 1000 sets of estimated mean lobster settler abundance, respectively, for each year. The 1000 sets of the estimated mean lobster abundance for each sampling scheme were compared with the mean of “true” population parameter V^{true} (Yates 1946; Chen 1996; Kimura and Somerton 2006). We estimated relative estimation error (*REE*) and relative bias (*RB*) to quantify the comparison:

$$REE = \frac{\sqrt{\frac{\sum_{i=1}^{N}(V_i^{estimated} - V^{true})^2}{N}}}{V^{true}} \times 100\%$$

where $V_i^{estimated}$ is the estimated mean lobster abundance of 10 sampling stations in the ith sampling; i is from 1 to 1000; V^{true} is the mean lobster abundance of 1971 potential sampling stations for each simulation; and *N* is the sampling times for a given realization of "true" population. The *REE* values reflect the difference between sampling results and true lobster abundance in an area over time and measure both bias and variation in the evaluation. The *RB* measures the estimation bias and is quantified as:

$$RB = \frac{\frac{\sum_{i=1}^{N} V_i^{estimated}}{N} - V^{true}}{V^{true}} \times 100\%$$

A sampling approach with smaller *REE* and *RB* values indicates better performance (Chen 1996). The fixed-station sampling design was thought to have biased estimation compared to random station sampling design, and was expected to have a higher value of *RB.*

Analytical approach

We estimated presence/absence of species and stability in species abundance by measuring persistence (Philips and Johnston 2004; Robinson and Yakimishyn 2013). The index of persistence (ϖ) was estimated from lobster density data of fixed-stations for American lobster settlement survey (1989 - 2013). We did a pairwise comparison of lobster density of all the years and estimated the fixed-station sampling's power of detecting temporal trend of the American lobster density in mid-coast region of the GOM. The value of ϖ can be calculated as (Warren 1994):

$$\varpi = \frac{s_y^2/4}{s_s^2 - s_y^2/4}$$

$$s_s^2 = \frac{\sum_{y=1}^{2}\sum_{y=1}^{n_i}(x_{iy} - \bar{x}_y)^2}{(m_1 + m_2)}$$

$$s_y^2 = \sum_{i=1}^{m}(d_i - \bar{d})^2/(m-1)$$

where x_{iy} is observed lobster density in site i and year y, $\bar{x}_y$ is the mean observation of year y, m_1 *and* m_2 are the number of stations in the 1st and 2nd years, respectively, included in the pairwise comparison, d_i is the difference in density between the two years in site i, $\bar{d}$ is the mean of the density difference, and m is the number of fixed-stations. s_s^2 reflects the difference in lobster density between different sites in the same year, and s_y^2 measures the difference in lobster density of the same site between different years. The estimated ϖ values for each pair of years indicate the degree of persistence. A smaller value of ϖ indicates a greater degree of persistence (Table 4).

III-2. Results

Two-stage GAM selection and performance

The initial two-stage GAM model with the eight variables explained 36.9% and 47.1% variances for the first and second stages, respectively. There were six variables in each stage of the GAM after non-significant variables ($p > 0.05$) were removed (Table 5). Salinity (*S*) and distance to sediment boundary (*DS*) were found not significant in both stages of GAM. The final first stage GAM had a value of 36.8% for deviance and 0.398 for adjusted R^2; and the second stage GAM had a 46.2% for explained deviance and an adjusted R^2 of 0.451 (Table 2).

The response curves were presented in Fig. 3 for significant variables latitude (*La*), longitude (*Lo*), bottom water temperature (*T*), depth (*De*), and distance offshore (*DO*). lobster presence and abundance were found to have linear relationship with temperature. Both the presence probability and density of the American lobster increased with temperature. The other environmental variables showed complex relationships with lobster presence and abundance (Fig. 11). The probability of presence and abundance of American lobster were significantly higher in the gravel sediment and lower in the sand/silt/clay sediment.

Model evaluation

The adjusted R^2 values for the 100 cross-validation runs varied from 0.29 to 0.52. There was a positive relationship between predicted and observed lobster abundances (Fig. 12). The intercept values (mean ± standard error) were 0.94 ± 0.26, and the slope values were (mean ± standard error) 0.84 ± 0.07. The intercept values were significantly larger than 0 ($p < 0.001$), and the slope values were not significantly different with 1 ($p = 1.60$). This indicates that the two-stage GAM might have biased predictions for lobster abundance in the mid-coast region of the GOM and that abundance was increasingly under-estimated with increased abundance. However, it is sufficient for the use in the simulation of “true” populations in this study.

Predicted distribution of newly settled lobsters

The predicted density varied from 7 to 569 during 1989 to 2012 (Fig. 12). The lowest density was found in 1993 with the mean abundance (mean ± SE) 33.08 ± 1.62. The density increased dramatically during 2012 with the mean abundance (mean ± SE) 120.34 ± 1.65. The models predicted stable spatial patterns of the lobsters on sampling stations. The lobster abundance was higher in the inshore region of Kennebec and Damariscotta rivers than in the Sheepscot River (Fig. 13). There were several hot spots that have high lobster abundance in the mouth of the rivers. The lobster abundance decreased with the increased distance offshore. The spatial patterns were similar for all predicted years.

The mean of the “true” population of 1000 realizations for each year was measured (Fig. 14). The mean abundance of the American lobster had generally low abundance during the late 1990s, but increased dramatically from 2008 to 2012 (Fig. 14).

The mean abundance of 1000 simulations for each sampling design (i.e., fixed-station and random design) was estimated (Fig. 15). The mean abundance showed temporal trends similar to the “true” population for both the sampling designs; however the variability of estimates for the fixed-station sampling was much greater than that for the random sampling design. The mean abundance of random station sampling was same with the mean abundance of "true" population, suggesting that random design yielded unbiased estimates. The fixed-station sampling, however, underestimated the "true" simulated population (Fig. 15).

Relative estimation error and relative bias

Based on the *REE* values the random station sampling design had better performance than the fixed-station sampling design (Fig. 16). The mean *REE* of the fixed-station sampling over years ranged from 10.26 ± 3.88% (mean ± SD) to 14.03 ± 4.84%. The mean *REE* for the random station sampling varied from 4.82 ± 0.35% to 6.52 ± 0.47%. Year 2012 had the smallest *REE* values for both sampling designs. Year 1998 had the largest *REE* values for fixed-station sampling, and year 1993 showed the greatest *REE* value for random station sampling. The *REE* temporal changes for both the sampling designs showed positive correlations. The precision of the fixed-station sampling design was apparently worse than random station sampling design over years. The fixed-station sampling design had larger variation of *REE* values compared to random station sampling design.

The random sampling design was unbiased, but the fixed-station sampling design did not have evenly distributed *RB* values around zero (Fig. 17). The annual mean *RB* of the random station sampling ranged from -0.03 ± 0.59% to 0.03 ± 0.64%. The average mean *RB* of fixed-station sampling varied from -14.03 ± 4.84% to -10.26 ± 3.89%. The year 1993 had the smallest bias, and the year 1998 showed the largest bias for the fixed-station sampling. The annual mean *RB* of 1000 simulations for random station sampling design was less than 0.1%. The fixed-station sampling design showed relatively larger variation of *RB* values than random station sampling design (Fig. 17).

The index of persistence for given two years

The mean settlement density was calculated from settlement survey data in the GOM for 25 years (1989 to 2013). There were 10 fixed-stations in this area, and the number of successful sampling sites varied from 8 to 10. There were only 8 sites that had lobster density data during 1989 to 1994. The mean lobster density showed considerable interannual variability (Fig. 18). The average settlement density of the American lobster was over 1 lobster per m^2 during most years.

The index of persistence was estimated for each pair of years by using the newly settled lobster density data from settlement survey. The lower values of index ϖ, the greater persistence occurred between two adjacent years. The persistence between two pairs of years was strong during most years (Fig. 19). The persistence index for year pair 1993 and 2008 was the lowest with value 3.02 and implied the worst degree of persistence. The mean ϖ value (mean ± SD) for the 24 successive year pairs was 0.39 ± 0.21. The corresponding probability that fixed-station sampling will detect the temporal trend of the lobster density in the mid-coast region of the GOM was greater than 81.4% (Table 4). The mean ϖ value for all year pairs was 0.51 ± 0.35. The corresponding probability that fixed-station sampling will detect the temporal trend of the lobster density was around 77.9%.

III-3. Discussion

Fixed and random station sampling designs of monitoring programs for marine benthos were examined in several studies (Warren 1994; Van der Meer 1997). The objectives of monitoring program need to be clearly identified before designing because they may influence the choice of monitoring designs. Comparing temporal change of species abundance in an area or observing spatial patterns of a species between two areas can have a different optimal sampling design (Bijleveld et al. 2012). Sample size can also affect the monitoring results. Smaller sample size reduces estimation precision and even influence the ability to detect temporal changes (Quinn and Keough 2005). The lobster settlement survey is designed to monitor temporal changes of settlers and juveniles, which can then be used to monitor the dynamics of recruitment (Pershing et al. 2012). Hence, a small sample size with enough power to detect temporal trend is an ideal sampling design.

For historical reasons, a fix-station design is used in the settlement survey. In this study, we evaluated the ability of fixed sampling design detecting the temporal changes in the density of the newly settled lobster in the mid-coast region of the GOM. The results from the simulation study indicate that the fixed-station sampling design underestimated the absolute abundance of American lobster settlers (Fig. 16). However, the fixed-station sampling results yielded temporal patterns of the settler abundance similar to the "true" population trend over time. For known spatial pattern of an organism in an area, sampling only in the relative high or low abundance area can result in high relative estimation error for fixed-station sampling design. Those results are consistent with previous research that suggests the fixed-station sampling yielded the biased estimation of the species abundance over years (Van der Meer 1997; Cao et al. 2014). The random station sampling is an unbiased sampling design for monitoring the newly settled lobster abundance, and captured both the "true" population values and the temporal trend of the lobster settler abundance.

The variance of means was used to define the effectiveness of the monitoring program in previous research (Millard and Latternmaier 1986). The relatively high variation of the mean abundance was obtained in the fixed-station sampling process compared to random station sampling (Fig. 15). That is because the variability of random sampling for a given realization of 'true' population was averaged out. Random sampling was repeated 100 times for a given realization and estimated mean was the average of the 100 sets of sample mean. If we conducted fixed-station sampling and random station sampling once in each year for one simulated population, then the random station sampling design would yield larger variance of mean abundance compared with fixed-station sampling design (Fig. 20). This result of large variance for the random station sampling (Fig. 20) is consistent with those of other studies (Van der Meer 1997).

Although, the fixed-station sampling was considered biased, the effectiveness of this sampling design or the power of estimating temporal trend can be evaluated in terms of the probability of detecting temporal change (Millard and Lettenmaier 1986). The power calculation in Van der Meer's study indicates that the fixed-station sampling yielded the higher power of detecting temporal change than random station sampling. The index of persistent results of this study gave the degree of persistence with associated probability that the fixed-station sampling can detect the temporal change of species abundance. The mean ϖ value for the 24 year pairs indicate the fixed-station sampling design has greater than 81.4% of the probability to estimate change of abundance accurately. The information we obtained behind the ϖ value is the changes in spatial distribution between two successive years. A sudden change between two stations for one given year or two successive years can induce a loss in persistence. The reduced persistence may cause a loss in precision of the estimated change in the American lobster abundance (Warren 1994).

The challenge of the study is to identify the mechanisms that affect the accuracy of the prediction for the abundance of the American lobster. Statistical modeling provides an effective method for simulating the population of the American lobster (Cao et al. 2014). However, the predictive ability of the model may be limited by the data collected from different sources. We developed a two-stage GAM for predicting the newly settled lobster abundance in the potential sampling area. The model validation suggests that under-prediction occurs when lobster abundance was relatively low by using the two-stage GAM. The predictive ability may be affected by the period of the data. The bottom trawl survey data were collected from 2000 to 2012, which is short compared with the prediction period year 1989 to 2012 (Jensen et al. 2005). Six variables were included in the two-stage GAM, and bottom water temperature is the most important variable correlated with the temporal prediction of the American lobster abundance from 1989 to 2012 (Chang et al. 2010). The other five variables such as sediment type and depth kept the same for each year. It is difficult to collect the bottom water temperature data for the potential sampling area in the mid-coast region of the GOM. The water temperature data extracted from the FVCOM model highly affected the temporal prediction of the newly settled lobster abundance. The spatial distribution of the American lobster at each year shows reasonable prediction. The American lobsters migrate into deep water towards the coast late fall (Chen et al. 2006). However, the model is not suitable to predict the American lobster abundance in the mid-coast region of the GOM if there are significant changes in the ocean condition (Chang et al. 2010).

The fixed-station sampling design was evaluated by comparison with random station sampling design. The focus here is the estimation of temporal change of the lobster settler in the mid-coast region of the GOM. This study suggests that fixed-station sampling design can detect the substantial change and the temporal trend of the lobster settlers, suggesting that the abundance index from the American lobster settlement index program can capture the temporal variability of the lobster settler density. However, because the fixed-station design tends to yield biased estimates, lobster settlement program cannot be used for estimating absolute settler abundance.

Acknowledgement

This study was supported by a grant from the Maine Sea Grant College program to Chen and Wilson (NA10OAR4170081-11417). We thank Maine Department of Marine Resources for providing the bottom trawl survey data, settlement data and the information for potential sampling stations. We would like to thank Drs. R Wahle, D. Bradley, H. Xue, and R. Russell for their help and discussions in interpreting settlement data and environmental data. Bottom trawl survey data were made available to this study by Maine Department of Marine Resources. In particular we would like to thank Sally Sherman and Keri Stepanek for helping compile the data and provide relevant information.

References

Aiken, D. E., and S. L. Waddy. 1986. Environmental influence on recruitment of the American lobster Homarus americanus: a perspective. *Canadian Journal of Fisheries and Aquatic Sciences* 43, 2258-2270.

ASMFC (Atlantic States Marine Fisheries Commission). 2009. Stock assessment report No. 09-01 (supplement) of the Atlantic States Marine Fisheries Commission: American Lobster stock assessment report for peer review. *Atlantic States Marine Fisheries Commission*, VA, pp. 4.

Berry, S.C., A. H. Welsh. 2002. Generalized additive modeling and zero inflated count data. *Ecological Modeling*. 157: 179-188.

Bijleveld, A.I., J.A.Van Gils, J. Van der Meer, A. Dekinga, C. Kraan, H.W. Van der Veer, and T. Piersma. 2012. Designing a benthic monitoring programme with multiple conflicting objectives. *Methods in Ecology and Evolution*. 3: 526-536. Doi: 10.1111/j.2041-210X.2012.00192.x.

Bowlby, H.D., J.M. Hanson, and J.A. Hutchings. 2008. Stock structure and seasonal distribution patterns of American Lobster, *Homarus americanus*, inferred through movement. *Fisheries Reseaerch*. 90: 279-288.

Cabral, H. N., and A. G. Murta. 2004. Effect of sampling design on abundance estimates of benthic invertebrates in environmental monitoring studies. *Marine Ecology Progress Series* 276, 19-24.

Cao, J., Y. Chen, J.H. Chang, and X. Chen. 2014. An evaluation of an inshore bottom trawl survey design for American Lobster (*Homarus americanus*) using computer simulations. *J. Northwest Atlantic Fisheries Science* 46: 27-39.

Chang, J.H., Y. Chen, D. Holland, J. Grabowski. 2010. Estimating spatial distribution of American Lobster *Homarus americanus* using habitat variables. *Mar Ecol Prog Ser*. 420: 145-156. Doi: 10.3354/meps08849.

Chen, Y. 1996. A Monte Carlo study on impacts of the size of subsample catch on estimation of fish stock parameters. *Fisheries Research*. 26: 207-223.

Chen, Y., S. Sherman, C. Wilson, J. Sowles, M. Kanaiwa. 2006. A comparison of two fishery-independent survey programs used to define the population structure of American Lobster (*Homarus americanus*) in the Gulf of Maine. *Fish Bull*. 104: 247-255.

Cochran, W. G. 2007. ‘Sampling Techniques.’ 3rd Edn. (Wiley India Pvt. Ltd.)

Cooper, R.A., R.A. Clifford , and C.D. Newell. 1975. Seasonal Abundance of the American Lobster, *Homarus americanus*, in the Boothbay region of Maine. Transactions of the American Fisheries Society. 104 (4): 669-674.

Cowan, D.F., A.R. Solow, A.R., and A. Beet. 2001. Patterns in abundance and growth of juvenile Lobster, *Homarus americanus*. *Marine Freshwater Research*. 52: 1095-1102.

Fielding, A.H. and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation*. 24: 38-49

Folmer, O., and M. A. Pennington. 2000. Statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. *Fisheries Research* 49, 165-178

Franklin, J. and J.A. Miller. 2009. Mapping species distributions: spatial inference and prediction. *Cambridge: Cambridge University Press*.

Gavaris, S., and Smith, S. J. 1987. Effect of allocation and stratification strategies on precision of survey abundance estimates for Atlantic Cod (*Gadus morhua*) on the Eastern Scotian Shelf. *Journal of Northwest Atlantic Fishery Science* 7, 137-144.

Green, B.S., C. Gardner, J.D. Hochmuth, and A. Linnane. 2014. Environmental effects on fished Lobsters and crabs. *Reviews in Fish Biology and Fisheries*. 24: 613-638.

Hastie, T. and R. Tibshirani. 1990. Generalized additive models. *Monographs on statistics and applied probability*. Chapman & Hall, London.

Jensen, O.P., R. Seppelt, T.J. Miller, and L.J. Bauer. 2005. Winter distribution of blue crab Callinectes sapidus in Chesapeake Bay: application and cross-validation of a two-stage generalized additive model. *Marine and Ecological Progress Ser*ies 299: 239-255.

Jury, S. H., M. T. Kinnison, W. H. Howell, and W. H. Watson. 1994. The behavior of lobsters in response to reduced salinity. *Journal of Experimental Marine Biology and Ecology* 180, 23-37.

Kimura, D.K. and D.A. Somerton. 2006. Review of statistical aspects of survey sampling for marine fisheries. *Reviews in Fisheries*. 14: 245-283. Doi: 10.1080/10641260600621761.

Kish, L. 1965. ‘Survey sampling.’ (John Wiley & Sons: New York.)

Lawton, P., K. L. Lavalli. 1995. Postlarval, juvenile, adolescent and adult ecology. In: Factor JR (Ed) Biology of the lobster: *Homarus americanus*. Academic Press, San Diego, CA, pp. 47-88.

Liu, Y., Y. Chen, and J. Cheng. 2009. A comparative study of optimization methods and conventional methods for sampling design in fishery-independent surveys. ICES Journal of Marine Science. 66: 1873-1882.

Lohr, S.L. 2009. ‘Sampling: design and analysis.’ 2nd End. (Brooks/Cole: California).

Millard, S.P., and D.P. Lettenmaler. 1986. Optimal design of biological sampling programs using the analysis of variance. *Estuarine, Coastal and Shelf Science*. 22: 637-656.

National Research Council. 2000. Improving the collection, management, and use of marine fisheries data. *National Academy Press*. Washington, D.C., pp. 75-79.

NOAA Annual Commercial Landing Statistics. 2012. http://www.st.nmfs.noaa.gov/commercial-fisheries/commercial-landings/annual-landings/index.

PaloheimO, J. E., and Y. Chen. 1996. Estimating fish mortalities and cohort sizes. *Canadian Journal of Fisheries and Aquatic Sciences* 53, 1572-1579.

Pennington, M., and J. H. Volstad, J. H. 1991. Optimum size of sampling unit for estimating the density of marine populations. *Biometrics* 47, 717-723.

Pershing, A.J.,R.A. Wahle, P.C. Meyers, and P. Lawton. 2012. Large-scale coherence in New England Lobster (*Homarus americanus*), settlement and associations with regional atmospheric condtions. *Fisheries Oceanography*. 21 (5): 348-362.

Philips, B.W. and C.E. Johnston. 2004. Fish assemblage recovery and persistence. *Ecology of Freshwater Fish*. 13: 145-153.

Pooler, P. S., and D. R. Smith, D. R. 2005. Optimal sampling design for estimating spatial distribution and abundance of a freshwater mussel population. *Journal of the North American Benthological Society* 24, 525-537

Quinn, G.P. and M.J. Keough. 2005. Experimental design and data analysis for biologists. 4th edition. *Cambridge University Press*, Cambridge.

Quist, M.C., K.G. Gerow, M.R. Bower, and W.A. Hubert. 2006. Random versus fixed-site sampling when monitoring relative abundance of fishes in headwater streams of the upper

Colorado river basin. *North American Journal of Fisheries Management*. 26: 1011-1019. Doi: 10.1577/M05-153.1.
Ripley, B. D. 2004. 'Spatial statistics.' (Wiley-Interscience.)
Rivoirard, J., J. Simmonds, and K. G. Foote. 2000. 'Geostatistics for estimating fish abundance.' (Blackwell Science: Oxford.)
Robinson, C.L.K. and J. Yakimishyn. 2013. The persistence and stability of fish assemblages within eelgrass meadows (*Zostera marina*) on the Pacific coast of Canada. *Can. J. Fish. Aquat. Sci*. 70: 775-784.
Selgrath, J.C., K.A. Hovel, and R.A. Wahle. 2007. Effects of habitat edges on American Lobster abundance and survival. *Journal of Experimental Marine Biology and Ecology*. 353 (2): 253-264.
Seng, Y.P. 1951. History survey of the development of sampling theories and practice. *Journal of the Royal Statistical Society*. 2: 214-231.
Simmonds, E. J., and R. J. Fryer. 1996. Which are better, random or systematic acoustic surveys? A simulation using North Sea herring as an example. *ICES Journal of Marine Science* 53, 39-50.
Skibo, K. M., C. J. Schwarz, and R. M. Peterman. 2008. Evaluation of sampling designs for red sea urchins Strongylocentrotus franciscanus in British Columbia. *North American Journal of Fisheries Management* 28, 219-230.
Smith, S. J. 1996. Analysis of data from bottom trawl surveys. *NAFO Scientific Council Studies* 28, 25-53.
Smith, S. J., and M. J. Lundy. 2006. Improving the precision of design-based scallop drag surveys using adaptive allocation methods. *Canadian Journal of Fisheries and Aquatic Sciences* 63, 1639-1646.
Smith, S. J., and M. J. Tremblay. 2003. Fishery-independent trap surveys of lobsters (Homarus americanus): design considerations. *Fisheries Research* 62, 65-75.
Steneck, R. S. 2006. Possible demographic consequences of intraspecific shelter competition among American lobsters. *Journal of Crustacean Biology* 26, 628-638.
Steneck, R.S., and R.A. Wahle. 2013. American Lobster dynamics in a brave new ocean. Canadian Journal of Fisheries and Aquatic Sciences, 70: 1612-1624.
Taylor, J. R. 1997. 'An introduction to error analysis: the study of uncertainties in physical measurements.' (University Science Books.)
Thompson, S. K. 1990. Adaptive cluster sampling. Journal of the American Statistical Association 85, 1050-1059.
Van der Meer, J. 1997. Sampling design of monitoring programmes for marine benthos: a comparison between the use of fixed versus randomly selected stations. *Journal of Sea Research*. 37: 167-179.
Wahle, R.A., J.S. Cobb, L.S. Incze, P. Lawton, M. Gibson, R. Glenn, C. Wilson, and J. Tremblay. 2010. The American Lobster settlement index at 20 years: looking back - looking ahead. *Journal of the Marine Biological Association of India*. 52: 180-188.
Wahle, R.A., L.S. Incze, and M.J. Fogarty. 2004. First projections of American Lobster fishery recruitment using a settlement index and variable growth. *Bulletin of Marine Science*. 74: 101-114.
Wahle, R. A., and R. S. Steneck. 1991. Recruitment habitats and nursery grounds of the American lobster Homarus americanus: a demographic bottleneck? Marine Ecology Progress Series 69: 231-243.

Warren, W.G. 1994. The potential of sampling with partial replacement for fisheries survey. *ICES Journal of Marine Science*s 51: 315-324.
Wang, Y. G., Y. Ye, and D. A. Milton. 2009. Efficient designs for sampling and subsampling in fisheries research based on ranked sets. *ICES Journal of Marine Science* 66: 928-934.
Xue, H., L. Incze, D. Xu, N. Wolff, and N. Pettigrew. 2008. Connectivity of Lobster populations in the coastal Gulf of Maine: Part I: Circulation and larval transport potential. 210 (1-2): 193-211.
Xue, H. J., L. Shi., Cousin, S., and N. R. Pettigrew. 2005. The GoMOOS nowcast/forecast system. *Continental Shelf Research* 25: 2122-2146.
Yates, F. 1946. A review of recent statistical developments in sampling and sampling surveys. *Journal of the Royal Statistical Society*. 109: 12-30.

Table 1. List of sampling designs

Design I	Design II	Design III	Design IV	Design V	Design VI
Simple random design	Systematic design	Stratified design with 5 regions strata	Stratified design with 4 depths strata	Stratified design with 7 sediments strata	Stratified design with 20 strata (4 depths × 5 regions)

Table 2. The design effects estimated in simulation for seven years with small, medium and large sample size

		Design / spring						Design / fall					
Sample size	Year	I	II	III	IV	V	VI	I	II	III	IV	V	VI
87	2002	1	0.341	0.964	0.689	0.807	0.604	1	0.244	0.993	0.662	0.915	0.582
	2003	1	0.349	0.964	0.674	0.807	0.600	1	0.244	0.991	0.662	0.914	0.581
	2004	1	0.409	0.983	0.690	0.847	0.614	1	0.260	0.993	0.636	0.907	0.578
	2005	1	0.375	0.973	0.688	0.825	0.612	1	0.261	0.994	0.656	0.930	0.593
	2006	1	0.552	0.940	0.699	0.871	0.579	1	0.265	0.998	0.637	0.930	0.584
	2007	1	0.358	0.973	0.700	0.820	0.620	1	0.249	0.996	0.641	0.923	0.575
	2008	1	0.425	0.975	0.712	0.875	0.611	1	0.249	1.000	0.664	0.938	0.588
115	2002	1	0.120	0.959	0.690	0.810	0.605	1	0.170	0.985	0.646	0.915	0.577
	2003	1	0.118	0.971	0.662	0.823	0.576	1	0.168	0.988	0.642	0.920	0.586
	2004	1	0.129	0.968	0.678	0.836	0.600	1	0.161	0.985	0.624	0.913	0.574
	2005	1	0.125	0.964	0.682	0.817	0.600	1	0.160	1.001	0.633	0.937	0.583
	2006	1	0.198	0.924	0.698	0.865	0.556	1	0.157	0.990	0.625	0.922	0.579
	2007	1	0.123	0.956	0.683	0.816	0.593	1	0.155	0.998	0.628	0.919	0.571
	2008	1	0.155	0.949	0.682	0.853	0.576	1	0.159	0.979	0.633	0.920	0.584
144	2002	1	0.419	0.942	0.697	0.802	0.600	1	0.312	0.983	0.666	0.919	0.589
	2003	1	0.470	0.954	0.685	0.823	0.591	1	0.294	0.971	0.667	0.921	0.597
	2004	1	0.458	0.953	0.691	0.835	0.605	1	0.330	0.975	0.645	0.921	0.587
	2005	1	0.488	0.944	0.697	0.831	0.610	1	0.329	0.978	0.644	0.925	0.598
	2006	1	0.499	0.921	0.714	0.865	0.563	1	0.333	0.976	0.644	0.928	0.592
	2007	1	0.484	0.950	0.710	0.813	0.606	1	0.329	0.987	0.649	0.927	0.587
	2008	1	0.501	0.934	0.707	0.856	0.586	1	0.346	0.966	0.658	0.932	0.589

Table 3. The performance indices for four scenarios of sample allocation based on stratified design with 20 strata

		REE (%)				RB (%)				Design effect			
	Year	1	2	3	4	1	2	3	4	1	2	3	4
FALL	2002	10.067	9.562	10.101	10.107	-0.019	-0.009	0.021	0.024	0.551	0.496	0.552	0.549
	2003	8.099	**7.943**	8.254	8.460	-0.014	-0.002	0.000	0.012	0.388	**0.372**	0.406	0.420
	2004	**8.426**	8.433	8.647	9.044	-0.012	-0.057	-0.022	0.015	**0.393**	0.394	0.416	0.451
	2005	8.849	**8.688**	8.947	9.131	0.057	-0.040	0.023	0.059	0.401	**0.383**	0.407	0.419
	2006	8.822	8.715	**8.663**	8.728	-0.039	0.033	-0.039	0.007	0.401	0.396	**0.388**	0.393
	2007	**7.930**	8.096	8.013	8.431	-0.023	-0.004	0.042	0.025	**0.356**	0.371	0.364	0.402
	2008	8.054	**7.977**	8.549	8.715	-0.034	0.031	-0.008	-0.027	0.360	**0.354**	0.407	0.419
SPRING	2002	10.277	10.359	10.333	10.332	0.002	-0.023	-0.010	-0.010	0.589	0.600	0.597	0.597
	2003	**7.973**	7.993	8.219	8.414	0.030	-0.023	-0.011	-0.019	**0.377**	0.378	0.398	0.421
	2004	8.106	**8.071**	8.510	8.693	-0.003	0.014	0.027	-0.047	**0.386**	0.387	0.431	0.446
	2005	8.584	**8.425**	8.532	8.935	-0.018	0.010	0.019	0.009	0.391	**0.375**	0.390	0.422
	2006	8.237	**8.236**	8.500	8.707	-0.046	0.018	0.065	0.017	**0.333**	0.334	0.359	0.373
	2007	**8.347**	8.367	8.637	8.916	-0.014	0.001	-0.008	0.017	**0.367**	0.369	0.394	0.423
	2008	8.368	**8.317**	8.621	8.792	0.007	0.025	0.015	0.019	0.370	**0.363**	0.389	0.407

The smallest REE and design effect for each scenario are emboldened.

Table 4. The relationship between ϖ values and the probability of fixed-station design being able to estimate inter-annual change. The probability value indicates the power of the fixed-station design that can detect the temporal trend of the Lobster settler density in the mid-coast region of the GOM.

ϖ	Probability (%)
0.1	98.1
0.2	91.7
0.3	85.9
0.4	81.4
0.5	77.9
0.6	75.2
0.7	73.0
0.8	71.1
0.9	69.6

Table 5. Model selection and performance for first-stage general additive model (GAMI; presence/absence) and second-stage GAM (GAMII; abundance). The significance test p-values of initial 8 variables were given for each model. Significant variables were in bold ($p < 0.05$). The data size N and adjusted R^2 were also explained for each model.

Model	***T***	*S*	***De***	***DO***	*DS*	***La***	***Lo***	***Se***	N	R^2 adj
GAMI	0.011	0.738	<0.001	0.002	0.449	<0.001	0.005	0.004	877	0.396
GAMII	<0.001	0.135	<0.001	0.002	0.235	<0.001	0.010	0.015	658	0.456

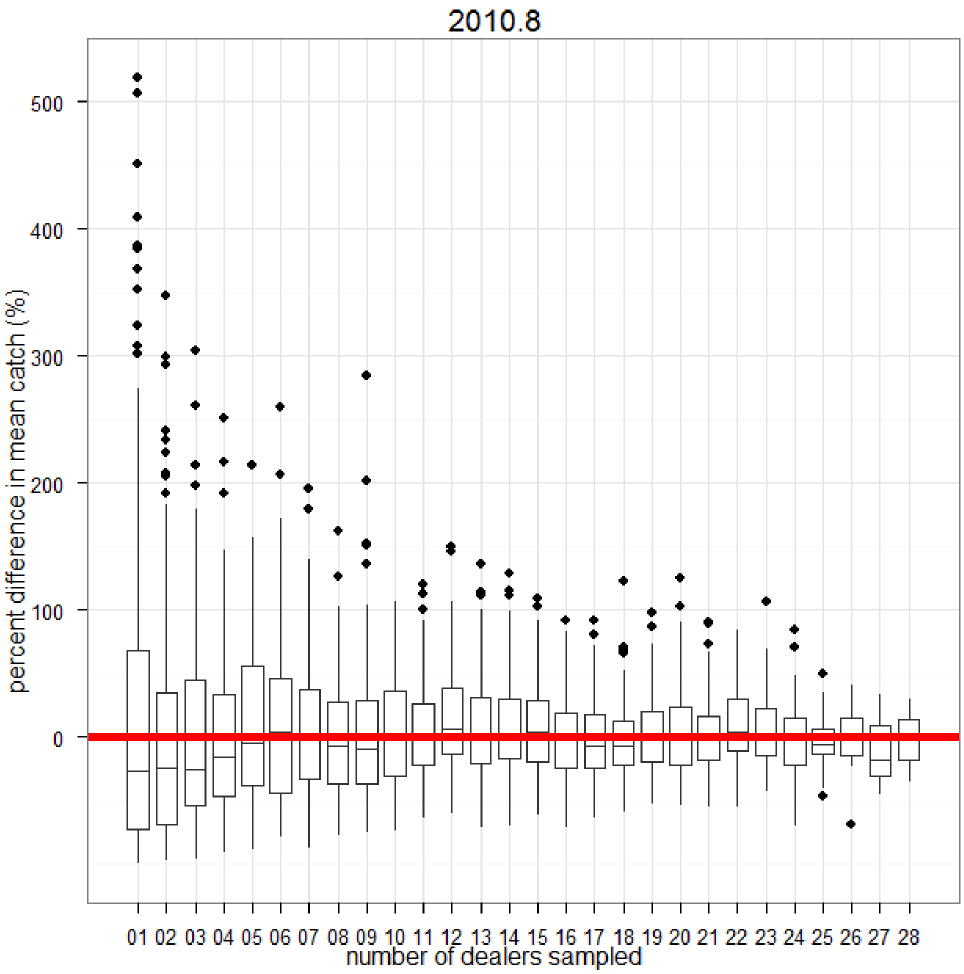

Figure 1. The percent difference in the mean catch with respect to the number of dealers sampled.

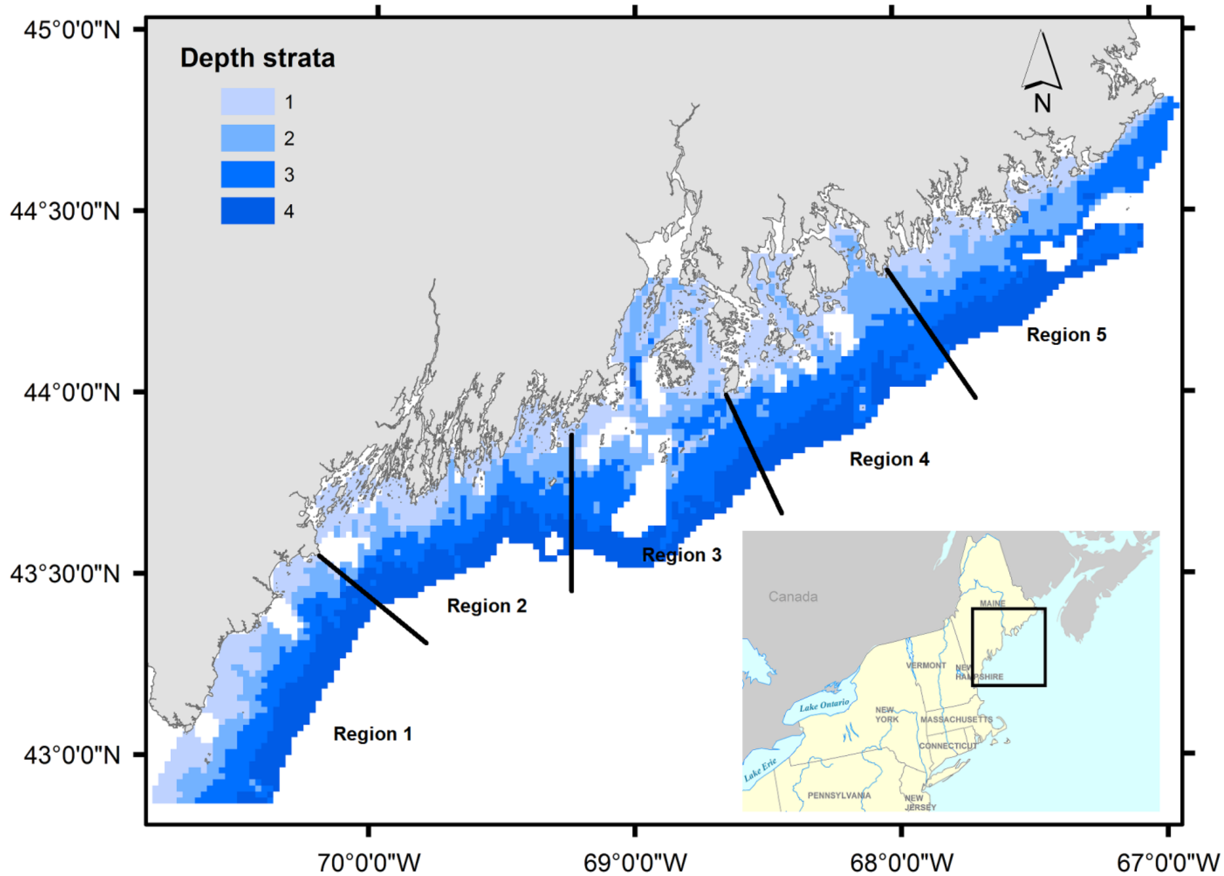

Figure 2. Region and depth strata for the Maine-New Hampshire inshore trawl survey (white areas are the areas that could not be towed)

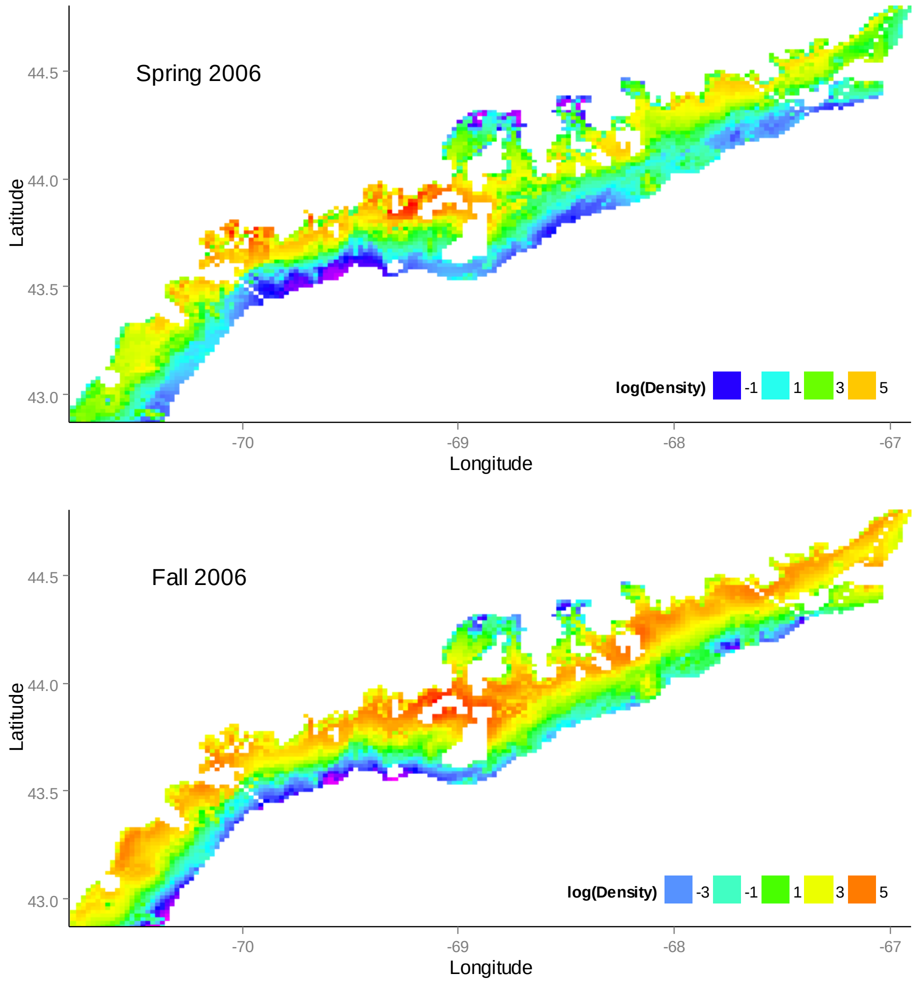

Figure 3. Simulated 'true' population distribution of American lobster in the Gulf of Maine for 2006

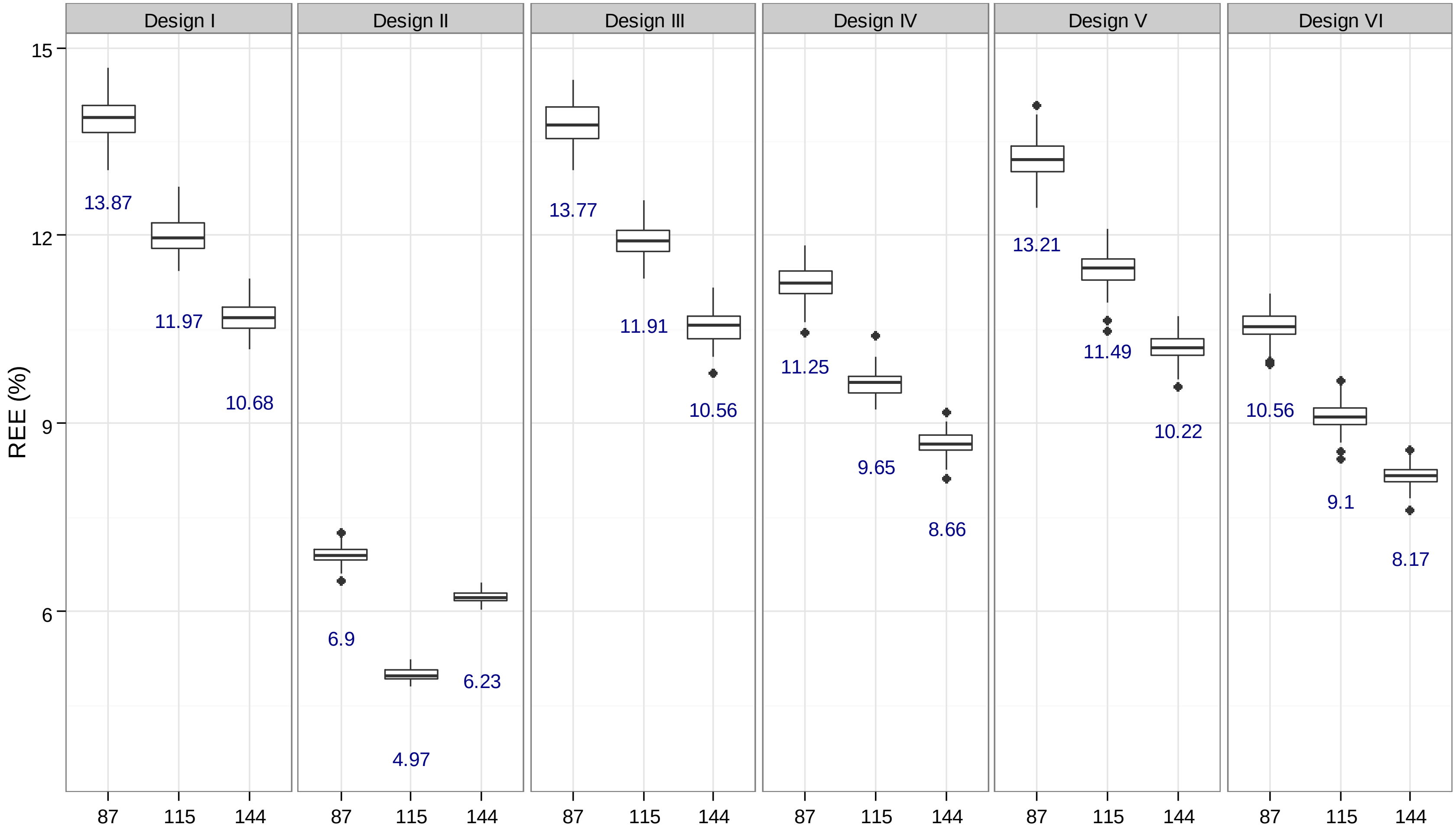

Figure 4. Comparison of index REE yielded by five evaluated sampling designs with small (87), medium (115) and large (144) sample sizes for fall population of 2002(values in the plot are medians)

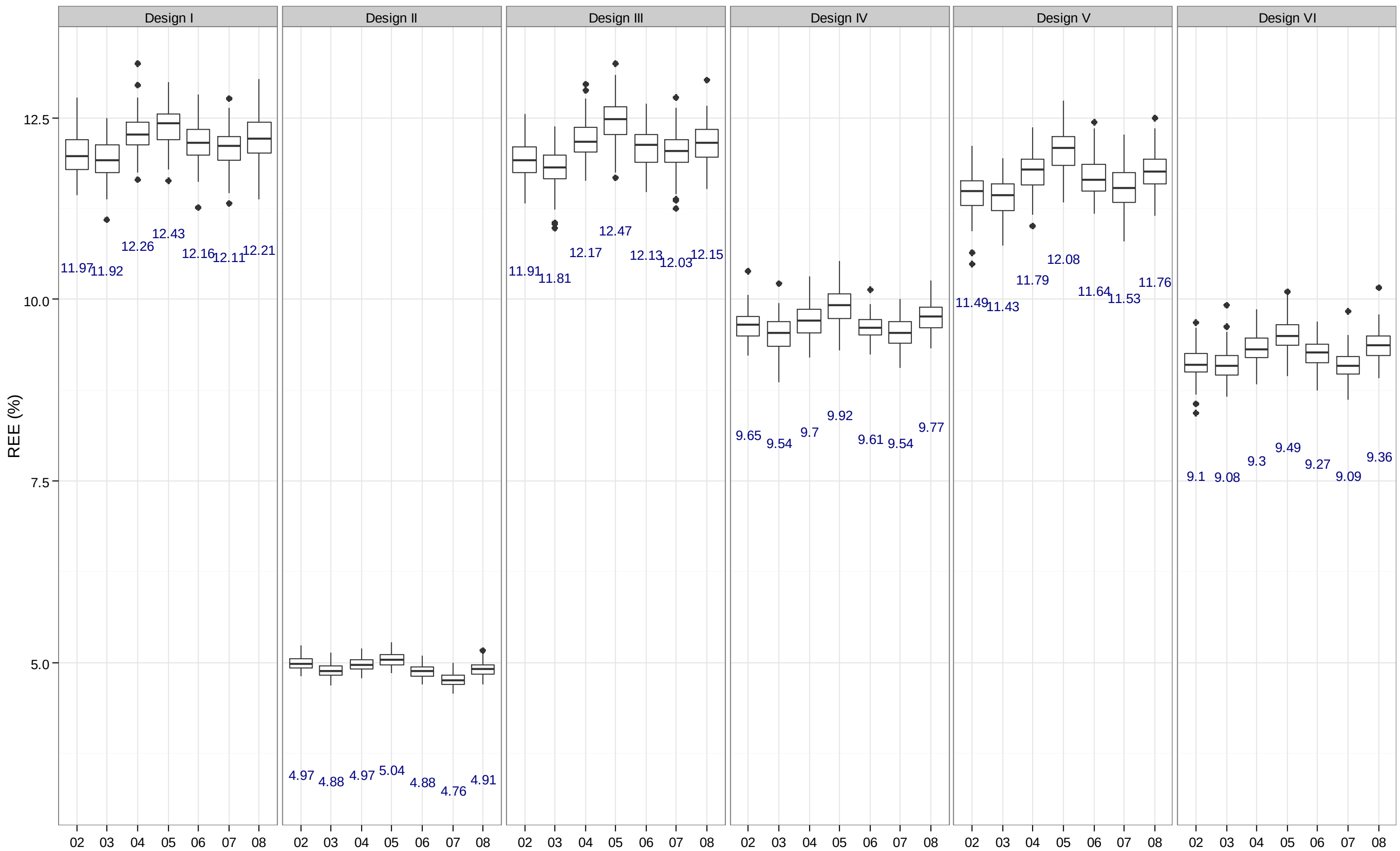

Figure 5. Performance index (i.e., REE) of five evaluated sampling designs with sample size being 115 across years (i.e., 2002-2008) for fall population (values in the plot are medians)

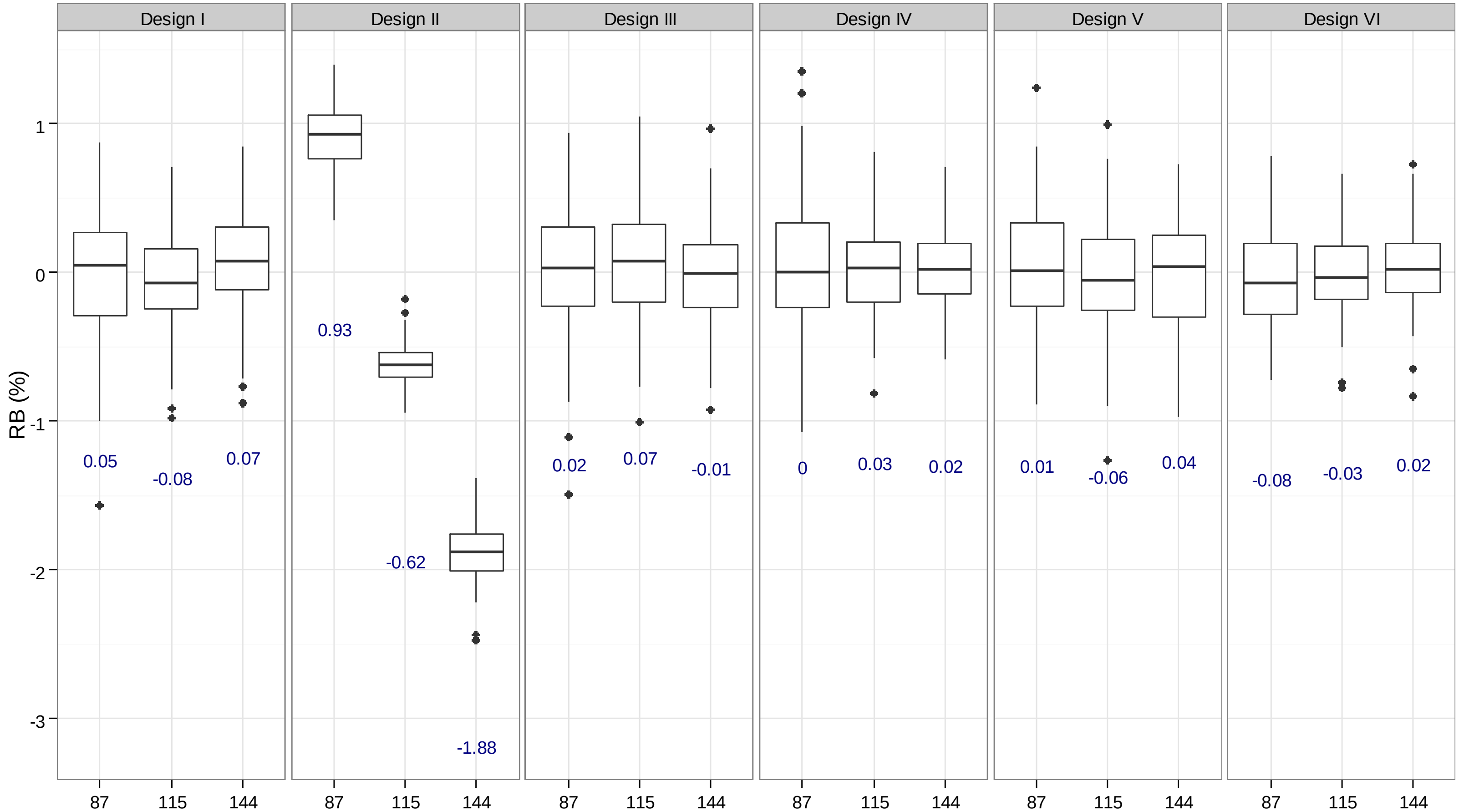

Figure 6. Comparison of index RB yielded by five evaluated sampling designs with small (87), medium (115) and large (144) sample sizes for fall population of 2002 (values in the plot are medians)

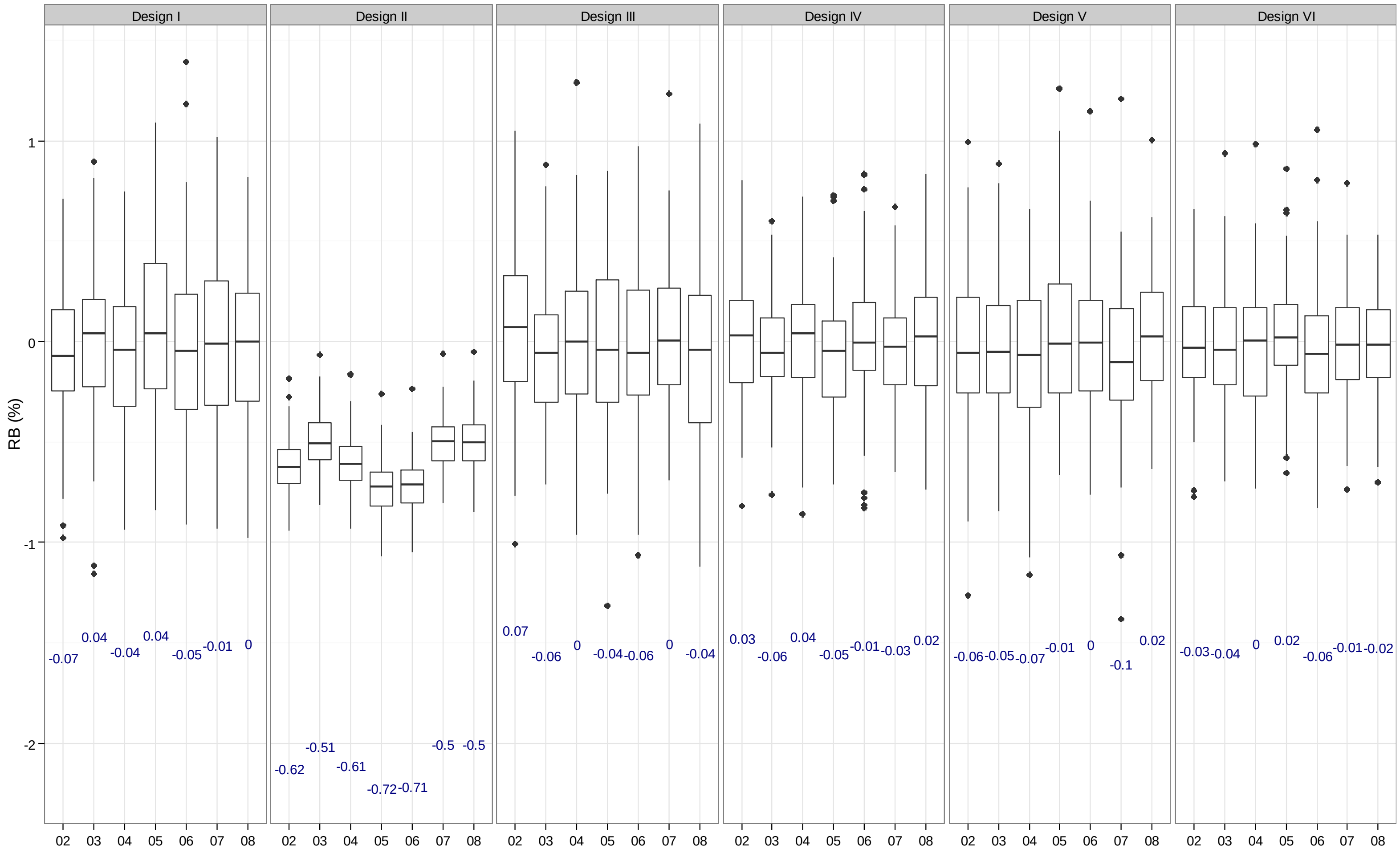

Figure 7. Performance index (i.e., RB) of five evaluated sampling designs with sample size being 115 across years (i.e., 2002-2008) for fall population (values in the plot are medians)

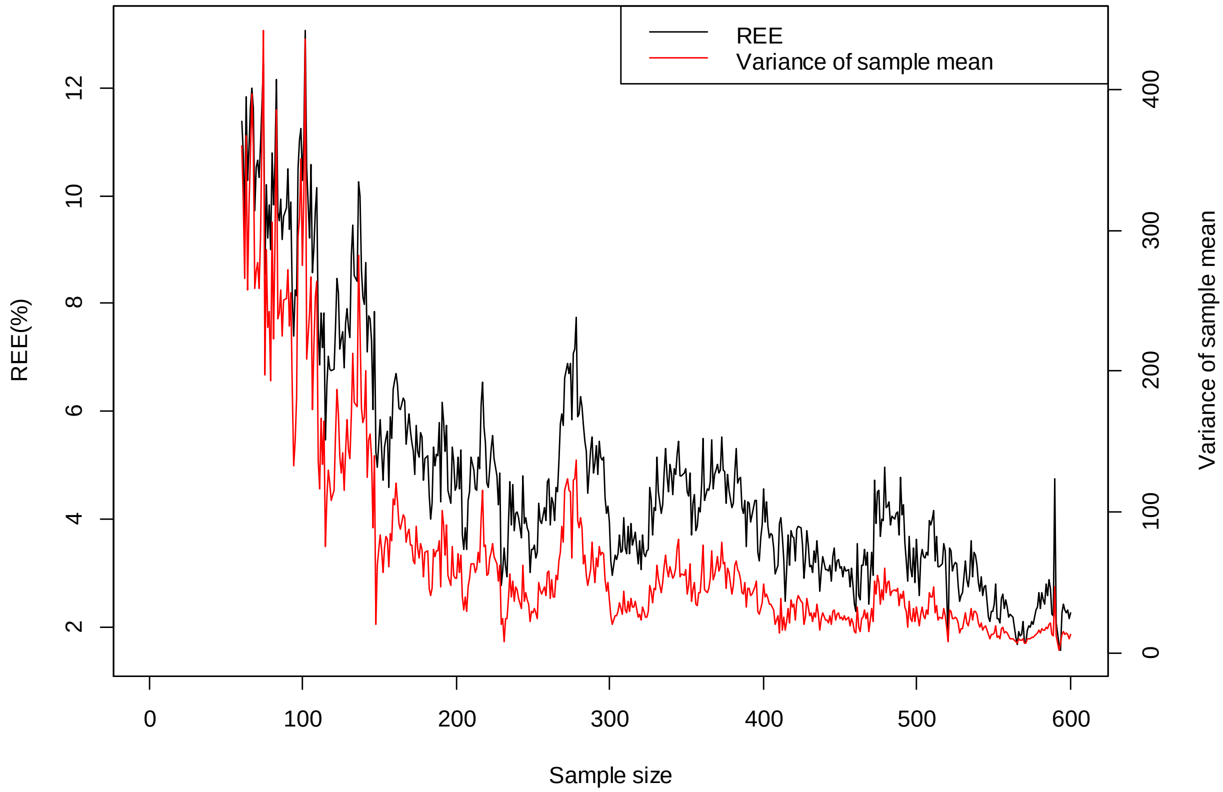

Figure 8. Change of REE and variance of sample mean yielded by systematic design with sample size

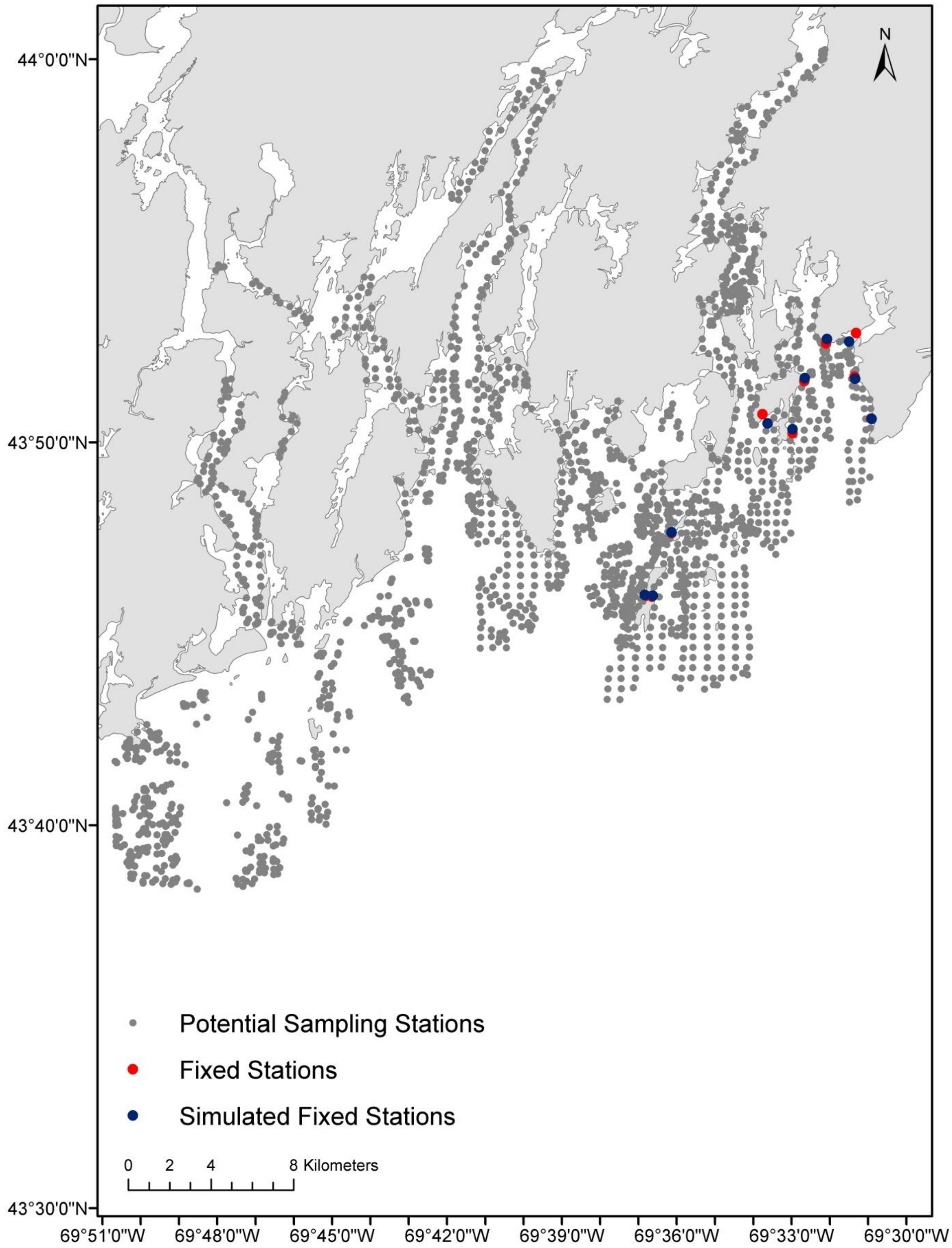

Figure 9. Map of mid-coast region of the Gulf of Maine and potential sampling stations.

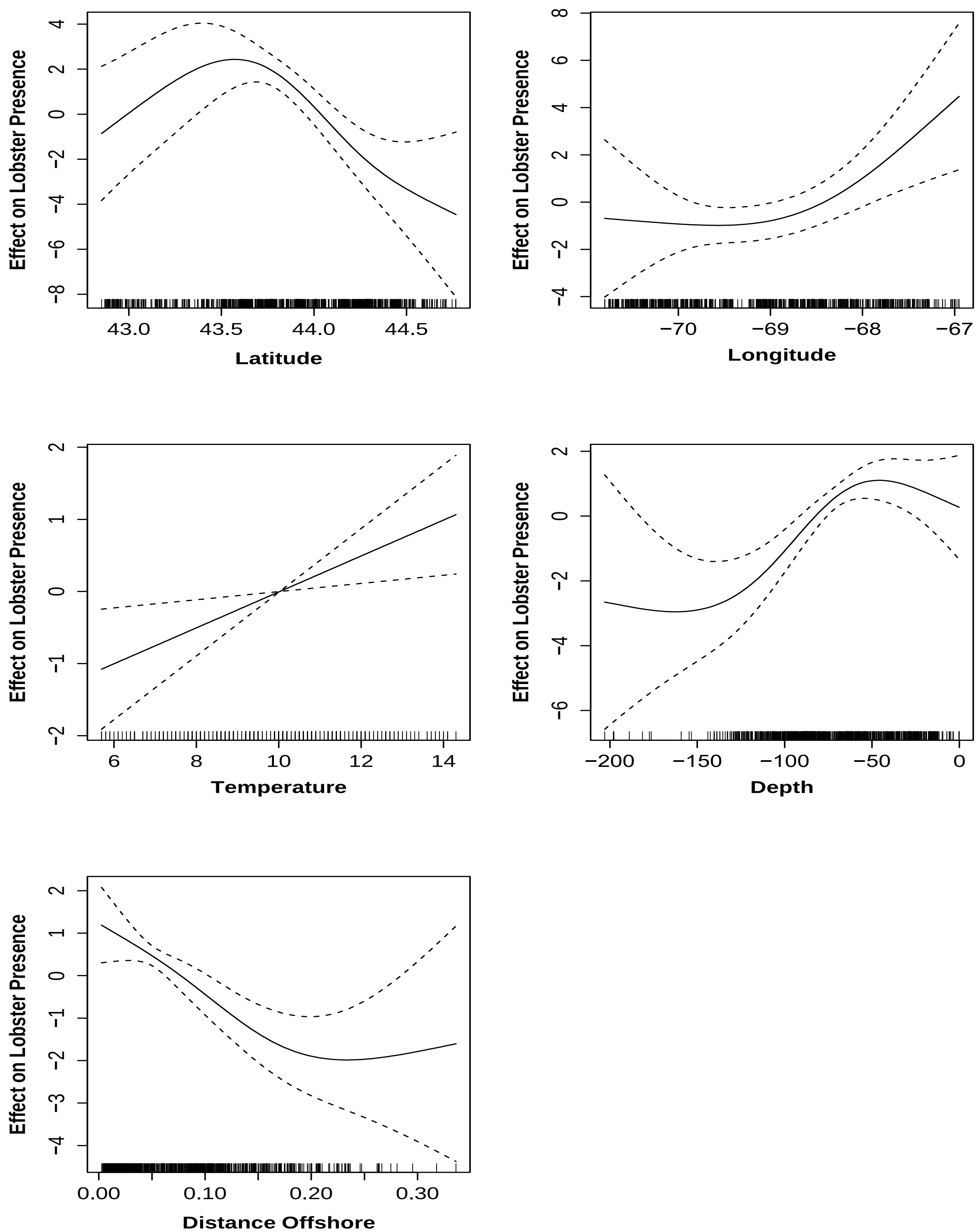

Figure 10. Response curve for significant variables of first stage GAM. The y-axis is the normalized effect of the variables on presence/absence component. The x-axis is the observation values. Dashed lines give 95% confidence intervals.

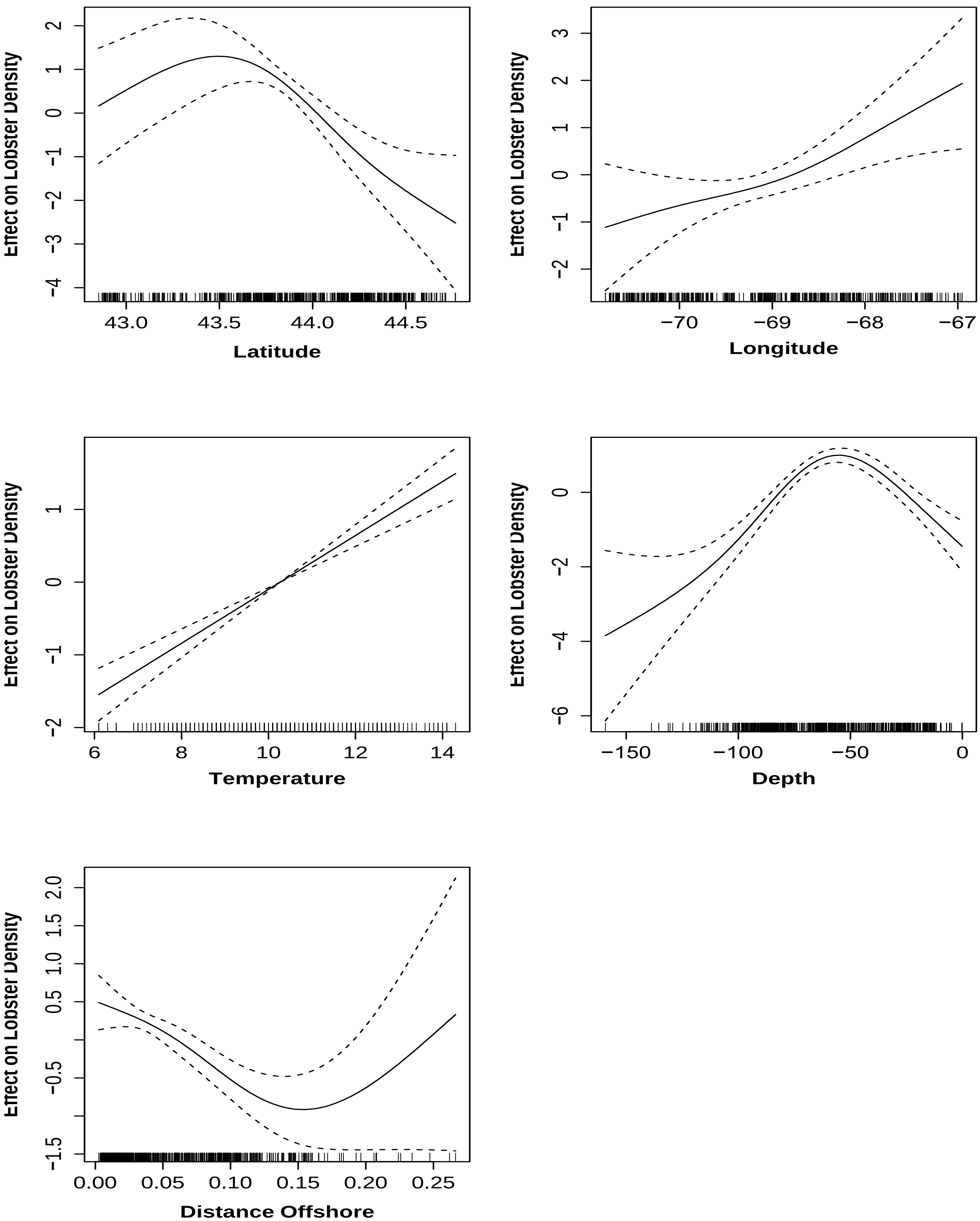

Figure 11. Response curve for significant variables of second stage GAM. The y-axis is the normalized effect of the variables on Lobster abundance component. The x-axis represents the observation values. 95% confidence intervals are shown as dashed lines.

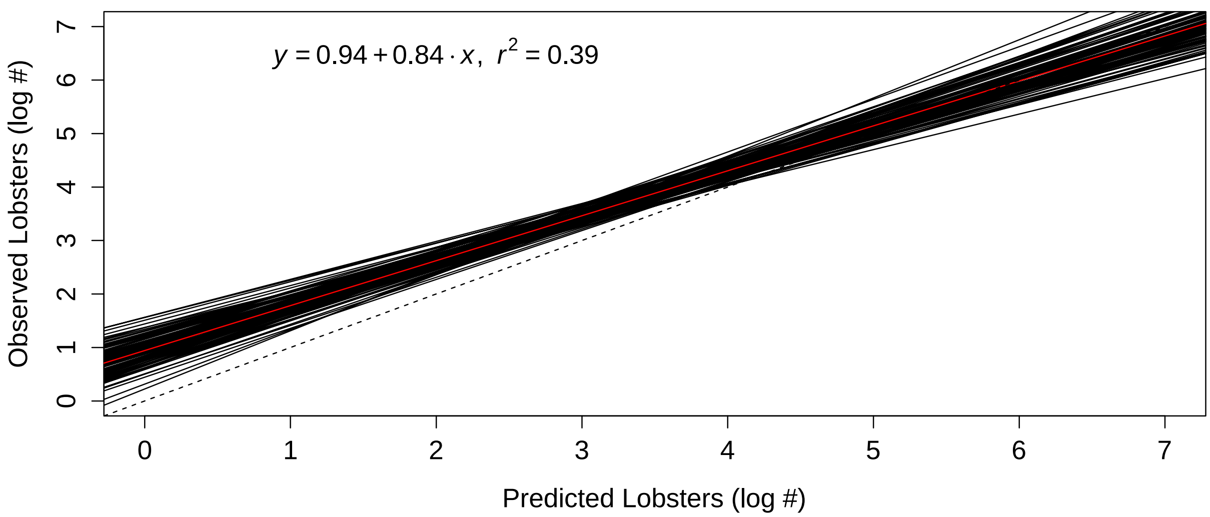

Figure 12. Model cross-validation. The predicted Lobster abundance vs. observed Lobster abundance for the bottom trawl survey data. The black solid lines are 100 linear regression lines fit all the data. The red solid line is the mean of cross validation results. The dashed line is the one-to-one line.

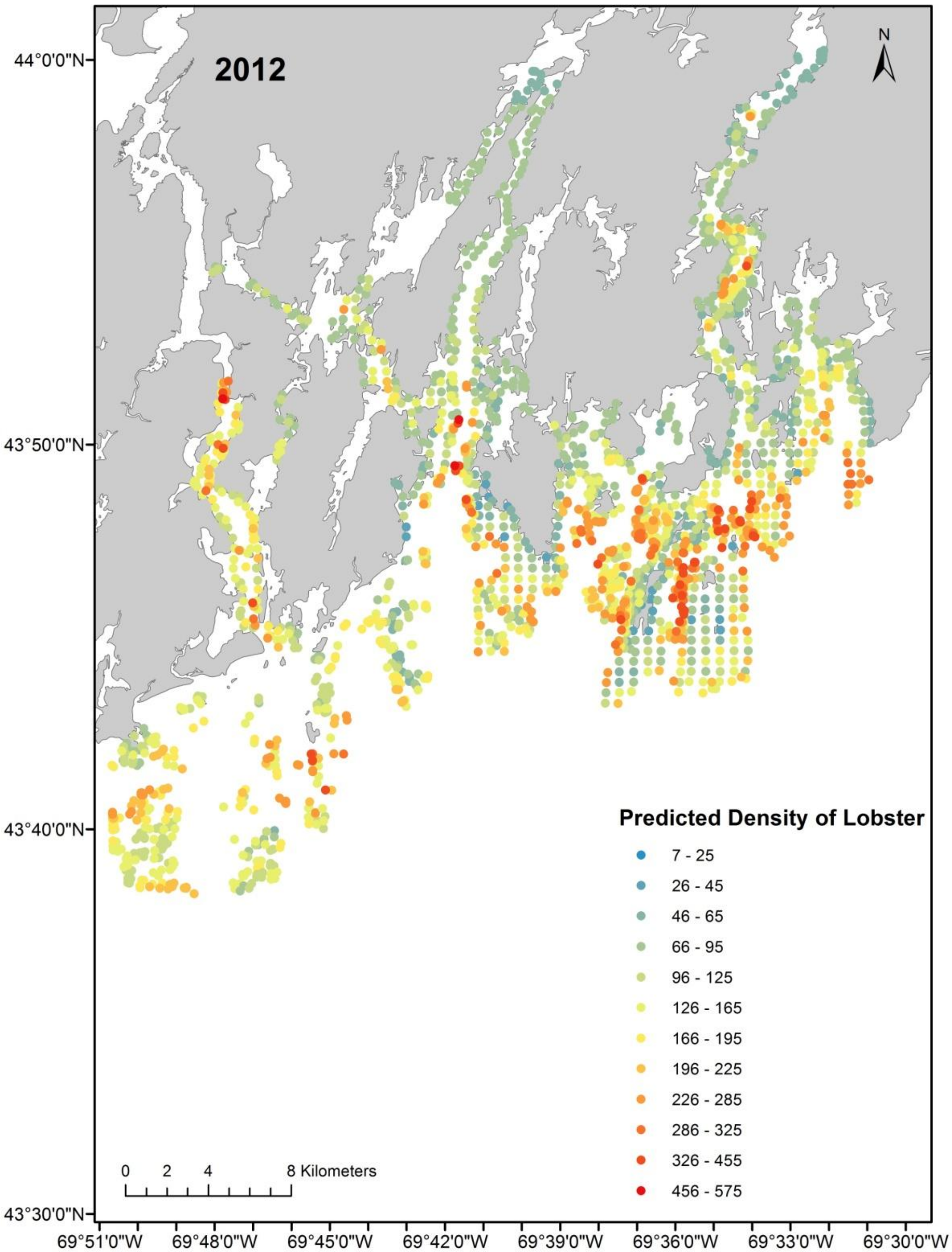

Figure 13. Predicted mean Lobster abundance at potential sampling stations on 2012.

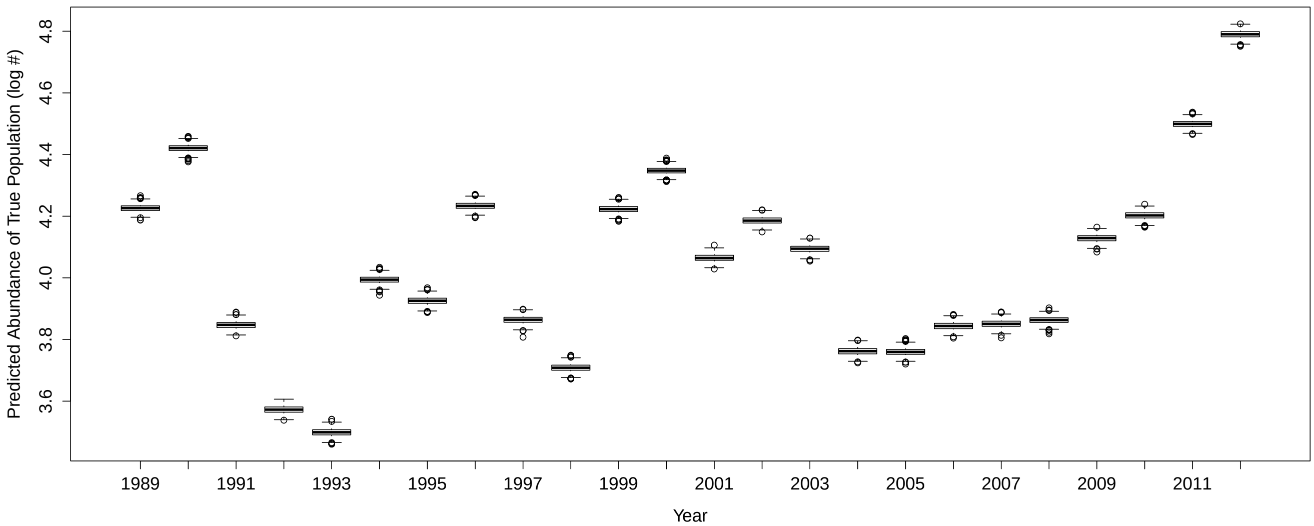

Figure 14. The mean of 1000 simulated true population for American Lobster at potential sampling stations from 1989 to 2012.

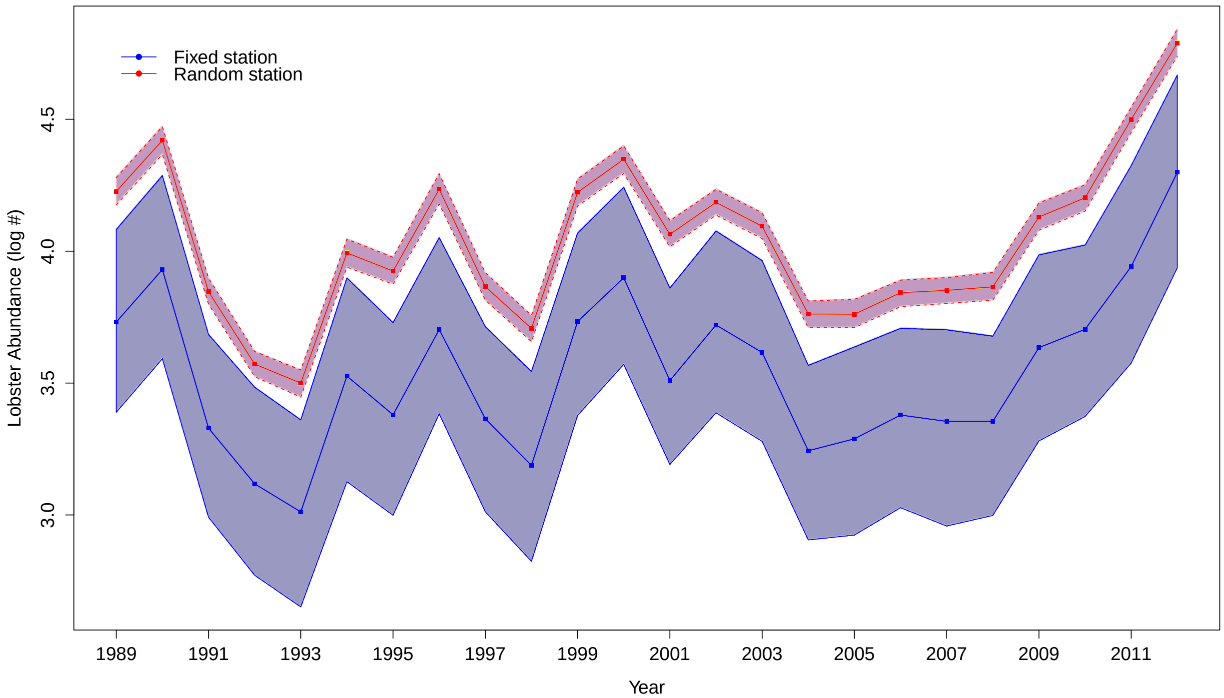

Figure 15. Temporal trends of means of fixed-station designg and random-station design from 1989 to 2012. The shadows represent 95% confidence interval. The random-station design was repeated 100 times for a given simulated population.

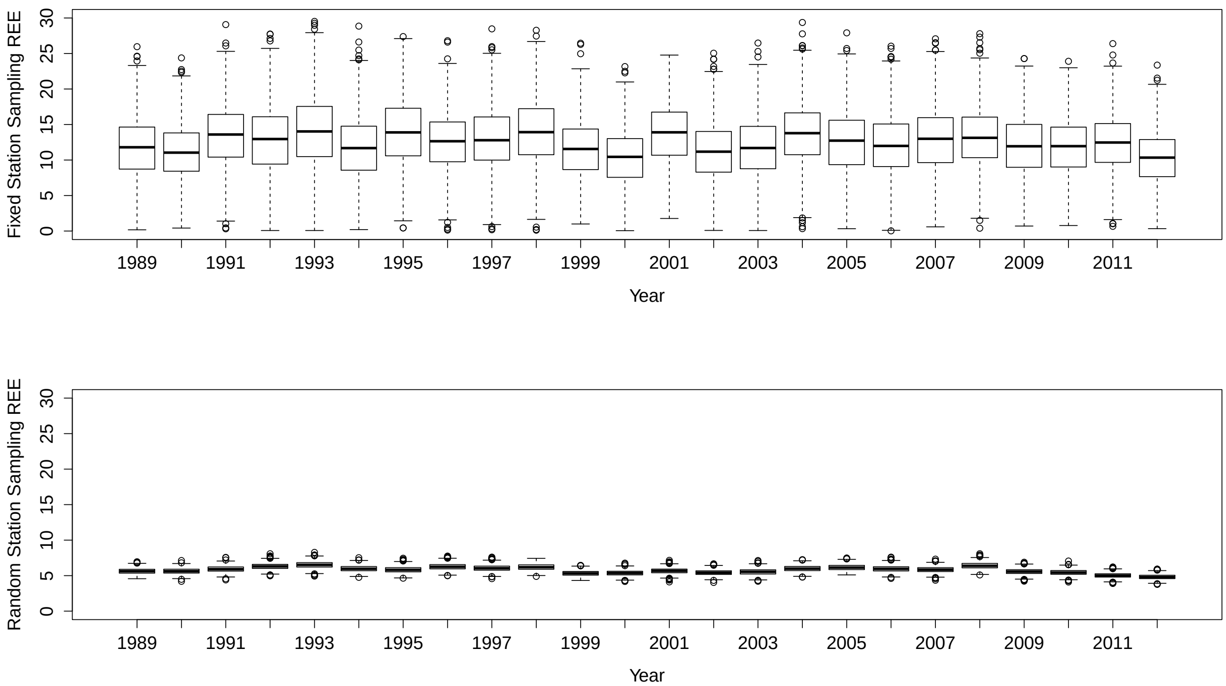

Figure 16. Performance index relative estimation error (%) of two sampling designs from 1989 to 2012.

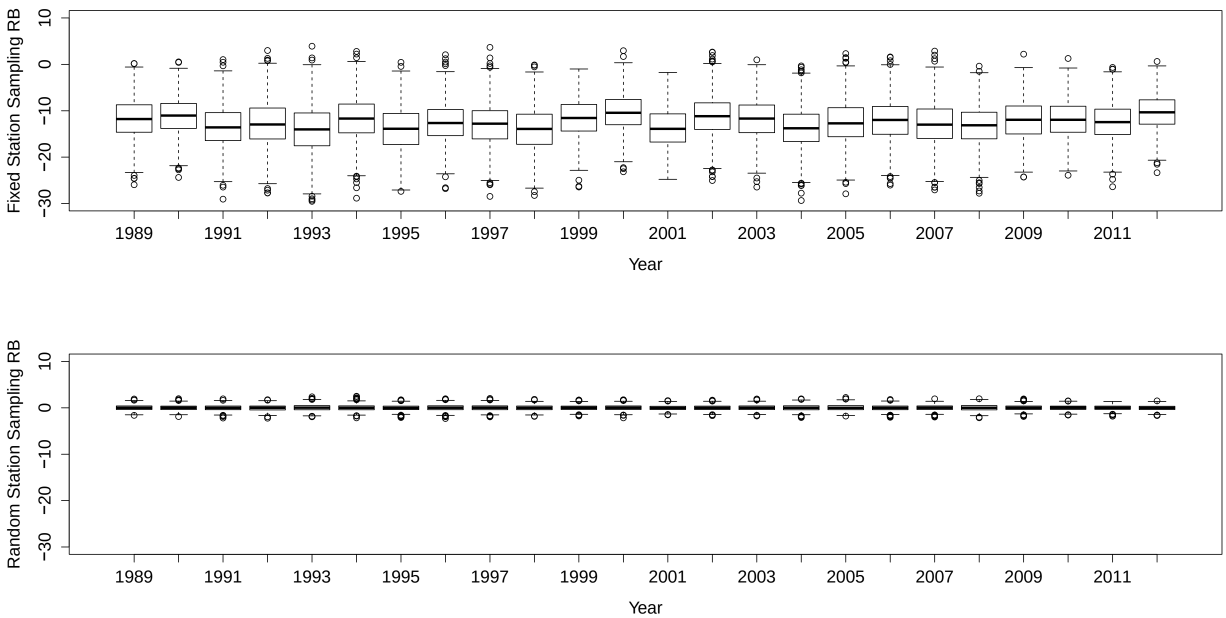

Figure 17. Performance index relative bias (%) of two sampling designs from 1989 to 2012.

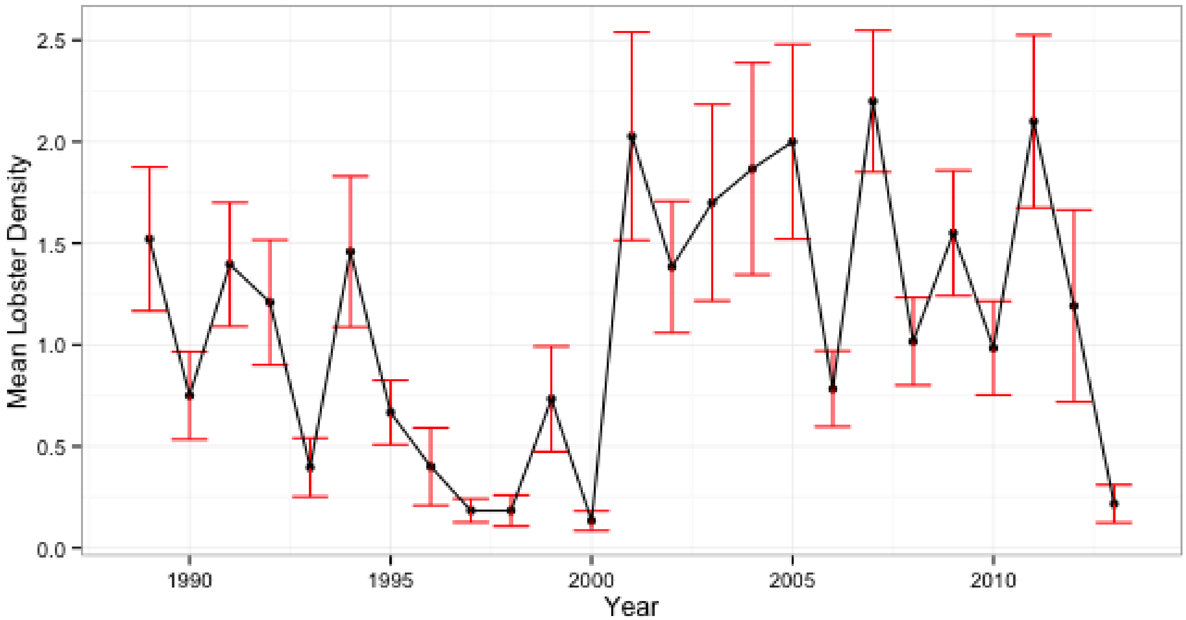

Figure 18. Mean Lobster density (# m^{-2}) from the settlement survey in the mid-coast region of the Gulf of Maine across years (1989 - 2013). The error bars on each time series represent the variability of Lobster density between sampling sites.

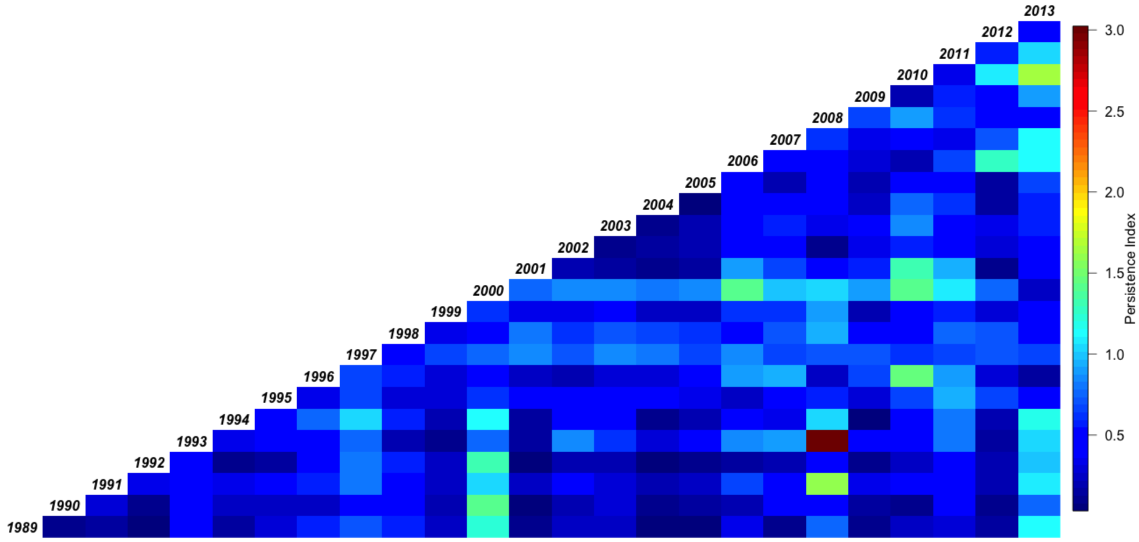

Figure 19. Persistence index matrix for given two years (1989 - 2013). The smaller persistence index ϖ value is, the greater persistence the fixed-station sampling obtains and the greater the power of differentiating inter-annual changes in the settler abundance is.

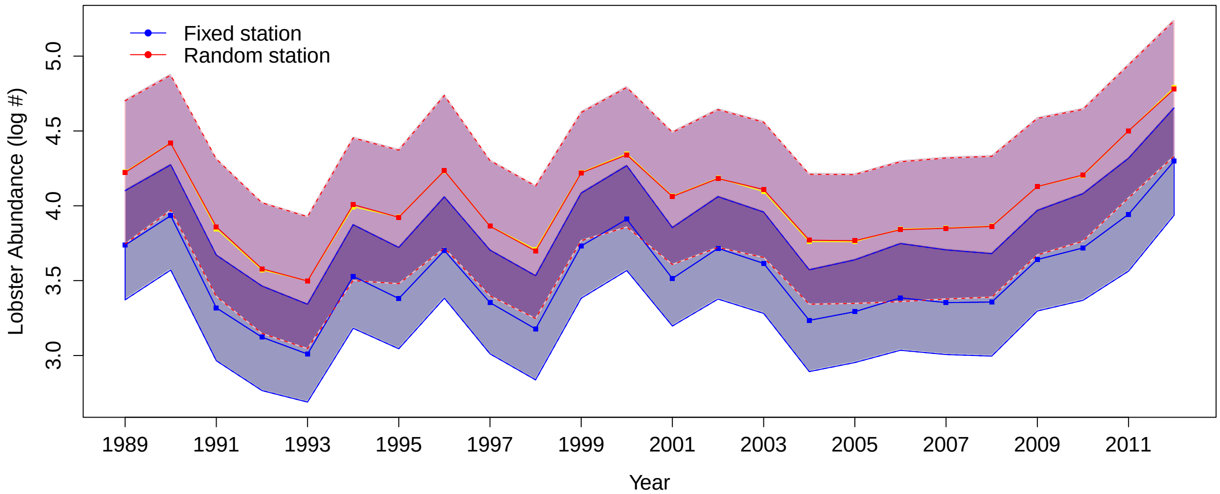

Figure 20. Temporal trend of means of fixed and random designs from 1989 to 2012. The random-station sampling process was not repeated 100 times for a given simulated population. The shadows indicate 95% confidence intervals.