

Spatial and Temporal Distribution of Horseshoe Crab (*Limulus polyphemus*) Spawning in Delaware Bay: Implications for Monitoring

DAVID R. SMITH^{1,*}, PENELOPE S. POOLER¹, BENJIE L. SWAN², STEWART F. MICHELS³,
WILLIAM R. HALL⁴, PETER J. HIMCHAR⁵, AND MICHAEL J. MILLARD⁶

¹ U.S. Geological Survey, Leetown Science Center, 1700 Leetown Road,
Kearneysville, West Virginia 25430

² Limuli Laboratories, 7 Bay Avenue, Dias Creek, Cape May Court House, New Jersey 08210

³ Delaware Division of Fish and Wildlife, P. O. Box 330, Little Creek, Delaware 19961

⁴ College of Marine Studies, 700 Pilottown Road, Lewes, Delaware 19958-1298

⁵ New Jersey Division of Fish, Game, and Wildlife, P.O. Box 418, Port Republic, New Jersey 08241

⁶ U.S. Fish and Wildlife Service, Northeast Fisheries Center, 308 Washington Avenue, Lamar,
Pennsylvania 16848

ABSTRACT: Concern for the status of horseshoe crab (*Limulus polyphemus*) has increased as harvest for conch and eel bait has increased and spawning habitat has decreased. In early 1999 a workshop was held at the behest of the Atlantic States Marine Fisheries Commission to design a statistically valid survey of horseshoe crab spawning in Delaware Bay. The survey that resulted was a redesign of a volunteer-based spawning survey that began in 1990, and its network of volunteers was relied on to implement the three-stage sampling design in 1999. During May and June of 1999, 163 participants surveyed during the highest of the daily high tides on 16 beaches (8 on each side of Delaware Bay). During the first half of the spawning season, spawning was associated with lunar phases, but moderated by wave height. Disproportionately more spawning occurred within 3 d of the first new and full moons, and spawning activity (measured by an index of female density) was correlated inversely to the percent of beaches with waves ≥ 0.3 m. Spawning was heaviest on the Delaware shore around the full moon in May in spite of low waves in New Jersey during the new and full moons in May. Number of beaches sampled was the most important factor in determining the precision of the spawning index and power to detect a decline. Explicit consideration of statistical power has been absent from the current debate on horseshoe crab status and harvest. Those who argue against harvest restrictions because of a lack of statistically significant declines take on a burden to show that the surveys they cite have high statistical power. We show the Delaware Bay spawning survey will achieve high statistical power with sufficient sampling intensity and duration. We recommend that future Delaware Bay spawning surveys sample on 3 d around each new and full moon in May and June and increase the number of beaches to ensure high statistical power to detect trends in baywide spawning activity.

Introduction

Increases in harvest of Atlantic horseshoe crabs (*Limulus polyphemus*) for bait and loss of their spawning habitat have raised questions about their current status (Botton and Ropes 1987; Berkson and Shuster 1999; Widener and Barlow 1999). Much concern has focused on shorebird, horseshoe crab interactions in Delaware Bay. Horseshoe crab spawning on Delaware Bay beaches coincides with shorebird migration through the region. Shorebirds, such as redknot (*Calidris canutus*), ruddy turnstone (*Arenaria interpres*), sanderling (*Calidris alba*), and semipalmated sandpiper (*Calidris pusilla*), stop over in the Delaware Bay and feed

heavily on horseshoe crab eggs, as well as infaunal benthic invertebrates, to fuel nesting and reproduction in the Arctic (Castro and Myers 1993; Botton et al. 1994; Burger et al. 1997; Tshipoura and Burger 1999). In addition to being harvested for bait, horseshoe crabs are caught and bled to produce *Limulus* amoebocyte lysate, which is used to detect pathogenic endotoxins in medical products (Berkson and Shuster 1999). Bled horseshoe crabs are returned to their point of capture, and Rudloe (1983) reported that survival of bled horseshoe crabs was 90% of the survival of unbled horseshoe crabs.

In spite of the emergence of multiple threats on the long-term stability of horseshoe crabs and the potential that their decline will have a ripple effect through the ecosystem, few resources had been directed at collecting data for stock assessment. In

* Corresponding author: tele: 304/724-4467; fax: 304/724-4465; e-mail: david_r_smith@usgs.gov.

response, the Atlantic States Marine Fisheries Commission (ASMFC) developed a Fisheries Management Plan with a mandate that certain states “formulate standardized and statistically robust methodologies . . . for spawning surveys” (ASMFC 1998a, p. 23). As part of the mandate, a workshop was held in January 1999 to discuss the design of a logistically feasible and scientifically valid spawning survey. Horseshoe crab researchers, natural resource managers, and biological statisticians attended the workshop. Primary goals of the workshop were to develop an index of spawning activity and design a statistically valid survey to track that index over time.

Since 1990, there has been a volunteer-based survey with the expressed goal of indicating status and trends of Delaware Bay horseshoe crabs (Finn et al. 1991). The volunteer-based survey has been criticized because frequent methodological changes have put into question its reliability and accuracy (ASMFC 1998c). Workshop participants recognized that, given available funds, the existing volunteer workforce would continue to be needed to implement a baywide spawning survey. At the same time, the volunteer-based survey would be strengthened by a statistical redesign. Outstanding issues regarding survey design remained, such as when and how often to survey beaches especially in relation to new and full moons, which are the times that horseshoe crabs spawning is believed to be heaviest (Rudloe 1980; Shuster and Botton 1985; Maio 1998). Other survey design issues, such as sample size calculations that rely on large-scale spatial and temporal variation, needed to be more fully addressed than was possible with available data.

In this paper we have three broad objectives. We describe a survey design and present results from the 1999 spawning survey in Delaware Bay. In the presentation of results, we describe the patterns of spawning spatially and temporally within the Delaware Bay. Second, we use the results from the 1999 survey to evaluate the survey design and make recommendations for future horseshoe crab spawning surveys in Delaware Bay. Finally, we discuss statistical power of the spawning survey in the context of risks that management decisions present to other species and fisheries that depend on the horseshoe crab.

Methods

We implemented a three-stage survey design in which beaches were selected at the first stage, dates within a beach were selected at the second stage, and quadrats within a beach and date were selected at the third stage. The target population was identified as that portion of the horseshoe crab

population that spawns in Delaware Bay in a given year. The survey objective was to estimate an index of spawning activity useful for measuring trends in spawning activity over time. We derived formulae for estimating the index based on this survey design and present them in the Appendix.

The first stage of sampling involved selection of spawning habitat within the bay and required defining a baywide sampling frame, which was a list of all spawning habitat that could be sampled in Delaware Bay. Because of physical constraints not all spawning habitat in Delaware Bay was accessible to a volunteer-based survey. As an alternative sampling frame, we included only beaches that could be accessed by public road. The restricted sampling frame, which cannot be used to estimate abundance or population size, is appropriate for trend detection under the assumption that trends in spawning activity on accessible beaches are parallel to trends for the target population. We stratified the beach-level sampling frame by state (Delaware and New Jersey) because in past surveys spawning activity was typically higher on one side of Delaware Bay in a given year (Swan unpublished data) and managers were interested in state-specific assessments. In the 1999 survey, 16 beaches (8 on each side of Delaware Bay) were selected by stratified random sampling (Table 1). We limited sampling to a beach section ≤ 1 km in length because 1 km could be surveyed within 2 h, the time in which most spawning activity is likely to be completed (Maio 1998).

The second stage of sampling involved the time dimension and required defining a temporal sampling frame. Peak spawning activity occurs during May and June in Delaware Bay (Swan unpublished data). The temporal sampling frame could comprise all dates in May and June or it could be restricted to a subset of those dates (similar to the beach-level sampling frame). The issue of which dates to include in the temporal sampling frame was unresolved. Sampling in the 1999 survey was temporally intensive to examine the temporal distribution of spawning and evaluate alternative temporal sampling frames. In the 1999 survey, 16 beaches were scheduled to be sampled 3 d around the new or full moon (2 d prior, the day of, and 2 d after) and 1 d at the neap tide. To augment the temporal sampling even further, 6 of the 16 beaches (3 beaches on each side of the Delaware Bay) were sampled every second or third day (Table 1 and Fig. 1). We made a heuristic judgment to sample one of the diurnal high tides so that limited volunteers' time could be distributed over more beaches and dates. In past surveys, more spawning activity tended to occur during the higher of the daily high tides (Swan unpublished data in Dela-

TABLE 1. Beaches in Delaware (DE) and New Jersey (NJ) where spawning horseshoe crabs were sampled during 1999 in Delaware Bay.

State	Beach	Length of Beach Section (km)	Sampling Schedule
DE	Woodland	0.4	3 d at full/new moon plus 1 neap d
	Kitts Hummock	1.0	every second or third d
	North Bowers	1.0	3 d at full/new moon plus 1 neap d
	Big Stone	1.0	every second or third d
	Slaughter	1.0	3 d at full/new moon plus 1 neap d
	Fowlers	1.0	every second or third d
	Prime Hook	1.0	3 d at full/new moon plus 1 neap d
NJ	Broadkill	1.0	3 d at full/new moon plus 1 neap d
	Sea Breeze	0.2	3 d at full/new moon plus 1 neap d
	Gandys	0.4	3 d at full/new moon plus 1 neap d
	Fortescue	1.0	every second or third d
	Reeds	1.0	every second or third d
	Kimbles	0.8	3 d at full/new moon plus 1 neap d
	Highs	0.5	every second or third d
	South Cape Shore Lab	0.8	3 d at full/new moon plus 1 neap d
	North Cape May	1.0	3 d at full/new moon plus 1 neap d

ware Bay; Rudloe 1980 in Gulf of Mexico; Maio 1998 in mid-Atlantic coastal bays). We sampled on the higher of the daily high tides, which was invariably the evening or nighttime high tide.

The third (and ultimate) stage of sampling involved subsampling a beach to count horseshoe crabs. A 1-m² quadrat was chosen as the sampling unit to increase survey efficiency based on evidence that horseshoe crabs tend to cluster spatially along a beach. Horseshoe crabs were clustered on 82% of the sampling occasions during the 1996

and 1997 surveys (Swan unpublished data); this based on the equality of variance to mean in a Poisson series (Elliott 1977, p. 40). In the case of spatial clustering, Elliott (1977, p. 128) recommends, and cites other studies that support, the use of a relatively small quadrat.

We placed quadrats along a beach by systematic sampling with 2 random starts (Thompson 1992). Systematic sampling is an attractive alternative to simple random sampling because it is convenient to apply under field conditions. Systematic sam-

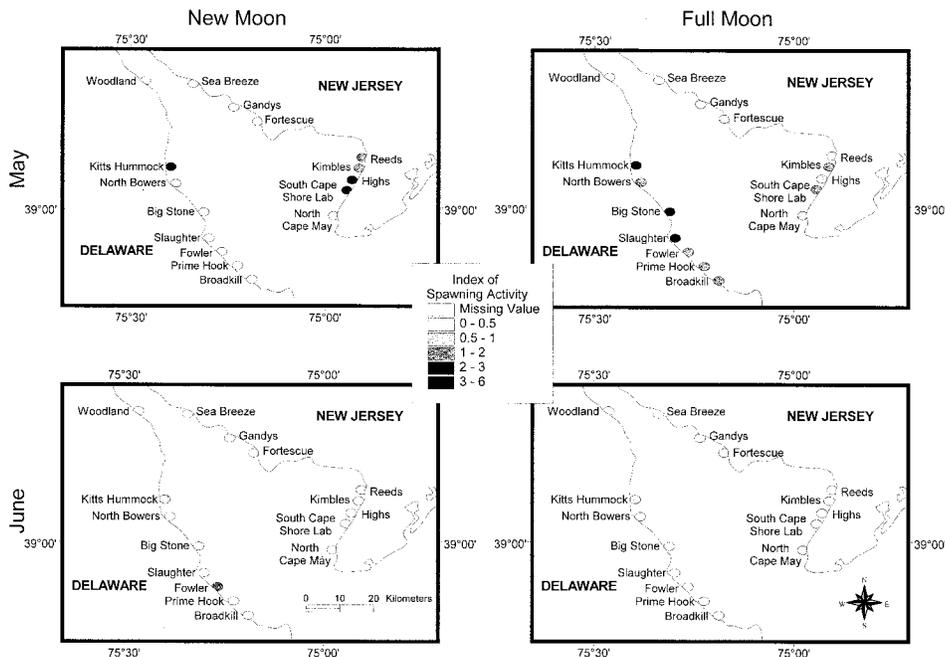


Fig. 1. Spatial distribution of spawning activity in Delaware Bay during the new and full moons of May and June 1999 at 16 beaches selected for the 1999 horseshoe crab spawning survey.

pling performs well for spatially clustered populations (Christman 2000). Multiple random starts allow estimation of variance and decrease the chance that the interval between quadrats coincides with a spatial pattern in the distribution of spawning horseshoe crabs. Hedayat and Singh (1991) recommend a systematic sample with 2 random starts because more random starts would compromise the advantage of easy application. Because we expected to sample 100 quadrats per beach, the interval between quadrats depended on the length of the beach.

Counting horseshoe crabs began when the tide began to recede from the high tide mark. Quadrats were placed systematically along the line formed by the spawning horseshoe crabs that were highest on the beach. Horseshoe crabs that were at least halfway in the quadrat were counted, and sex was recorded. Weather conditions, such as wave height, were also recorded. Protocol and data sheets were distributed to volunteers at training workshops and made available on internet (<http://aegis.er.usgs.gov/groups/stats/Limulus/protocol.html>).

We believe that most of the spawning females can be encountered ≤ 1 m of the high tide line. Brockmann (1990) observed 94% of nesting females ≤ 1 m of the maximum high tide line. Maio (1998) found highest density at shallow depths (≤ 33 cm). We attempted to count horseshoe crabs that were ≤ 2 m (rather than ≤ 1 m) from the high tide line by locating 2 adjacent quadrats extending into the bay at each systematically located point along the beach. In our trials of this approach we encountered two problems. The quadrat furthest in the bay was frequently in water too deep to be observed effectively, and the additional quadrats were time consuming so that surveying a beach exceeded the allotted time (2 h). We discontinued this practice.

The index of spawning activity was the average number of spawning females per quadrat (or per 1 m^2) at the high tide line on the highest of the daily high tides at a given beach. We formulated the index of spawning activity in terms of spawning females because the female segment of the population best represents the fluctuations in reproductive potential and egg availability to shorebirds. A female-based spawning survey was also recommended by a panel assembled by ASMFC to review the horseshoe crab stock assessment (ASMFC 1998b). We averaged across nights to estimate the index for a beach-specific index at a given period and averaged across beaches to estimate the index for the bay. Averages were weighted appropriately to account for survey design attributes, such as stratification and unequal beach length (Appendix).

There were 163 participants in the 1999 Delaware Bay horseshoe crab spawning survey. The participants, who were mostly volunteers, surveyed beaches for a total of 190 h, not counting the travel time or the time waiting for the tide to be at the right height. Typically, a crew of 3 to 4 took 53 min to sample a beach.

We used correlation to describe linear bivariate associations and generalized linear modeling methods, implemented in SAS and S-PLUS, to relate measures of reproductive activity to multiple explanatory variables. Where indicated by examination of residuals, we transformed response variables to diminish influence of outliers or satisfy distributional assumptions.

We used data from the 6 beaches that were most frequently sampled to compare the mean and variance of spawning activity that resulted from several alternative temporal sampling frames. For each alternative, we computed mean and among-beach variance for dates that were sampled within the frame. Our objective was to find the temporal sampling frame that minimized variance in relation to the mean. The alternative temporal sampling frames were defined as all days in May and June, 7 d centered on new or full moon plus 1 d at neap tide, 7 d centered on new or full moon, 7 d beginning on new or full moon, 5 d beginning on new or full moon, and 1 d on new or full moon.

Sample size and power to detect trends are important for effective implementation of the survey. We used the 1999 survey results to assess variation in the index of spawning activity and to calculate sample size and statistical power. The three sources of variation in the estimate of the index are between beach, between dates within each beach, and between systematic sample (i.e., quadrats) within each beach and date (Appendix). We simulated how changing the sample size for beaches, dates, or quadrats affects coefficients of variation (CV). Based on our comparisons of the alternative temporal sampling frames, we used only one temporal sampling frame in our simulations; that was the 7-d periods centered on the full and new moons in May or June. For simplicity, the beach-level sampling frame was set to 25 accessible beaches per state; this was slightly larger than the sampling frame in current use (21 accessible beaches in New Jersey and 19 in Delaware). We calculated CV for estimates of baywide and state-specific spawning. CV is important because it measures reliability of the index and is related inversely to the probability of detecting a decline in the index. We used program TRENDS (Gerrodette 1993) to calculate power to detect trends in the index and assess the effectiveness of the survey to detect declines in spawning activity. Type I error rate (α) was set to 0.10

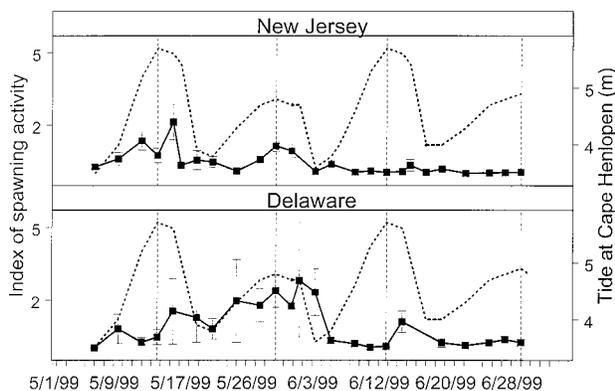


Fig. 2. Index of spawning activity during May and June for beaches on the New Jersey and Delaware sides of the Delaware Bay. The index of spawning activity is based on the density of spawning female horseshoe crabs. Tide height (m) at Cape Henlopen is the dashed line. Vertical reference lines (dashed) indicate new and full moon dates; the first vertical line from the left is a new moon.

and 0.20. We calculated power for exponential declines in spawning activity of 25%, 33%, and 50% over 5 and 10 yr. Power was based on a two-sided t -test, which is a conservative choice because power would be higher if based on a one-sided t -test. An examination of the 1996 and 1997 survey results (Swan unpublished data) indicated that beach-level CV was proportional to $1/\sqrt{\text{spawning activity}}$, so we assumed that relationship in the power calculations. We calculated cost of a spawning survey in terms of person-nights or the number of participants required if each person surveys only one night assuming a crew of three.

Results

SPATIAL AND TEMPORAL DISTRIBUTION OF HORSESHOE CRAB SPAWNING

During the first half of the spawning season, spawning was associated with the lunar phases (Fig. 2). Disproportionately more spawning occurred around the first new and full moons ($t_{79} = 4.27$, $p < 0.001$); 50% of spawning activity occurred within 3 d of the first new and full moons, but only 23% would be expected if spawning was uniformly distributed through May and June. Spawning was also affected by wave height (Fig. 3). When spawning began on the New Jersey shore (early to mid-May), a high percentage of beaches on the Delaware shore were experiencing wave heights in excess of 0.3 m (Fig. 3). Then as wave action subsided on the Delaware shore, spawning increased until it peaked again in early June shortly after the May full moon (Fig. 3).

Tide height alone was a poor predictor of spawning activity. During the first two lunar cycles (prior to June 7) when most of the spawning occurred,

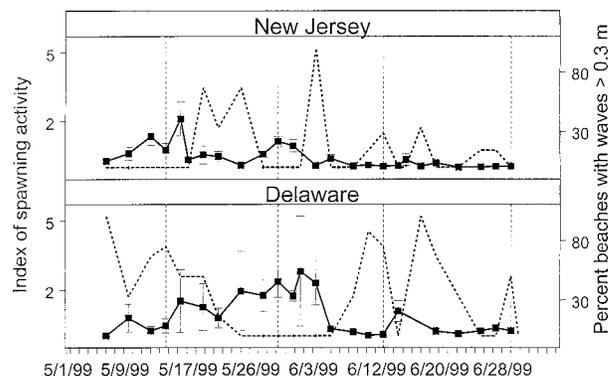


Fig. 3. Index of spawning activity during May and June for beaches on the New Jersey and Delaware sides of the Delaware Bay. The index of spawning activity is based on the density of spawning female horseshoe crabs. Percent of the beaches with waves exceeding 0.3 m is the dashed line. Vertical reference lines (dashed) indicate new and full moon dates; the first vertical line from the left is a new moon.

tide height accounted only for 12% of the variation of females and 6% of males. Continuing to look at results from the first two lunar cycles, correlation of spawning activity with tide height, although low, was higher for females ($r = 0.325$, $p = 0.106$) than for males ($r = 0.196$, $p = 0.338$). Spawning activity correlated more strongly with high waves (% of beaches with waves > 0.3 m), and the correlation was similar for females ($r = -0.531$, $p = 0.005$) and males ($r = -0.576$, $p = 0.002$). Effective modeling of spawning activity required a combination of temporal and geographic strata, weather, and tide height (Table 2).

At the new moon in May, spawning was concentrated in the vicinity of Reeds, Kimbles, Highs, and South Cape Shore Lab beaches (Fig. 1). From the end of May to early June spawning in Delaware was widespread. The exception was Kitts Hummock where spawning was relatively intense from early May and into early June, perhaps because it was locally sheltered from prevailing winds by a breakwater.

Sex ratio was affected by the magnitude of spawning activity, and that effect differed between the states ($F_{1,190} = 11.44$, $p = 0.003$; sex ratio tested on log scale). At beaches with spawning activity > 0.35 females m^{-2} , median sex ratio was 3.5 males to females (90% of those ratios ranged from 1.9 to 5.9) regardless of state ($F_{1,85} = 0.13$, $p = 0.721$; sex ratio tested on log scale). However, at lower spawning activity (≤ 0.35 females m^{-2}) sex ratio differed between the states ($F_{1,105} = 19.58$, $p < 0.001$; sex ratio tested on log scale); sex ratios were higher and more variable in New Jersey, where the median was 2.8 and 90% ranged from 1.0 to 9.4. Sex ratios were lower and less variable in Delaware at low

TABLE 2. Summary statistics for models relating spawning to state, time, tide height, and wave height. The r^2 was 0.63 for females and 0.67 for males. Spawning activity was transformed to the log scale for the regression. Time was categorized into dates before and after June 7, which was halfway between the full moon in May and the new moon in June. Wave height was the percent of beaches with waves ≥ 0.3 m.

Variable	Females			Males		
	Parameter Estimate	SE	p	Parameter Estimate	SE	p
Intercept	4.619	1.195	0.0004	-2.913	1.262	0.0261
State	1.183	0.329	0.0009	1.031	0.348	0.0050
Time	2.105	0.328	0.0001	2.183	0.346	0.0001
Tide height	0.432	0.239	0.0787	0.377	0.253	0.1438
Wave height	-2.128	0.505	0.0001	-3.279	0.533	0.0001

spawning activity, where the median was 1.5 and 90% ranged from 0.8 to 3.3.

Not all beaches were sampled as scheduled. Two of the eight Delaware beaches (Slaughter and North Bowers) were not sampled on any of the 3 d around the new moon in June, and a New Jersey beach (Sea Breeze) was not sampled in June because of problems with access. This missing information is unlikely to weaken our conclusions because spawning appeared to taper off considerably throughout the Delaware Bay during June. Missing sampling dates as scheduled is a concern, especially in a volunteer survey, and this is addressed in the discussion section.

EVALUATION AND RECOMMENDATION OF SURVEY DESIGN FOR MONITORING AN INDEX OF SPAWNING ACTIVITY

Among-beach variance to mean ratio was lowest and magnitude of the spawning activity was intermediate when the temporal (second-stage) sam-

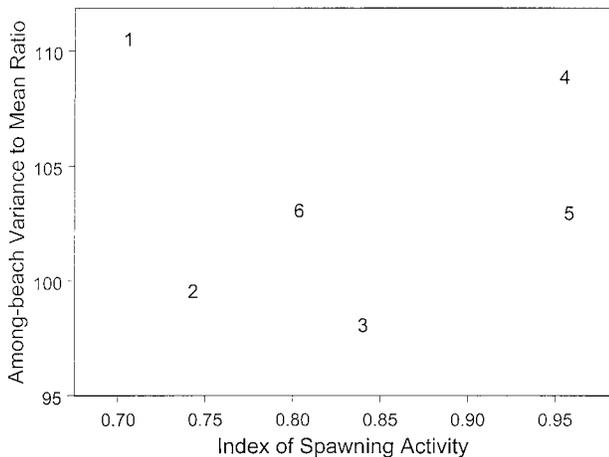


Fig. 4. Ratio of among-beach variance to mean spawning activity for six alternative temporal sampling frames. Numbers indicate alternatives: 1) all days in May and June, 2) 7 d centered on new or full moon plus 1 d at neap tide, 3) 7 d centered on new or full moon, 4) 7 d beginning on new or full moon, 5) 5 d beginning on new or full moon, and 6) d of new or full moon.

pling frame was restricted to the 7 d centered on new or full moons (Fig. 4; Case 3). The magnitude of spawning activity was greatest for a temporal sampling frame of 5 or 7 d beginning on and after the new or full moon (Cases 4 and 5). The magnitude was least when dates around neap tide were included (Cases 1 and 2). Sampling only on the new or full moon (Case 6) decreased the magnitude of the spawning index compared to Cases 3, 4, and 5.

The variance of the spawning activity index, $\hat{v}ar(\bar{y})$, can be partitioned into three components for sampling beaches, dates within beaches, and quadrats within dates and beaches (Appendix). Using data from the six most frequently sampled beaches and the Case 3 temporal sampling frame, the beach-level variance component was 116 times the date-level variance component and over 10,000 times the quadrat-level variance component. Removing stratification of dates by lunar period caused the date-level variance component to increase by a factor of 3.7. These results highlight the value of temporal stratification and suggest that precision of the index will be determined largely by number of beaches.

The effect of the number of beaches, nights, and quadrats on CV was the same whether spawning activity was estimated by state or baywide. We present only comparisons from the baywide index. Number of beaches was most important in determining CV ($F_{1,1256} = 32034.54$, $p < 0.0001$; Fig. 5); 96% of the reduction in CV (as measured by sum of squares) was due to number of beaches. Number of nights was statistically significant in determining CV ($F_{1,1256} = 102.79$, $p < 0.0001$). Less than 1% of the reduction in CV was due to number of nights. Number of quadrats did not significantly affect CV of the baywide index ($F_{1,1256} = 0.04$, $p = 0.85$).

Precision of the state-specific index was lower than for a baywide index. For sample sizes of 5 to 15 beaches per state, on average CV for a state-specific index was 41% higher than for the baywide

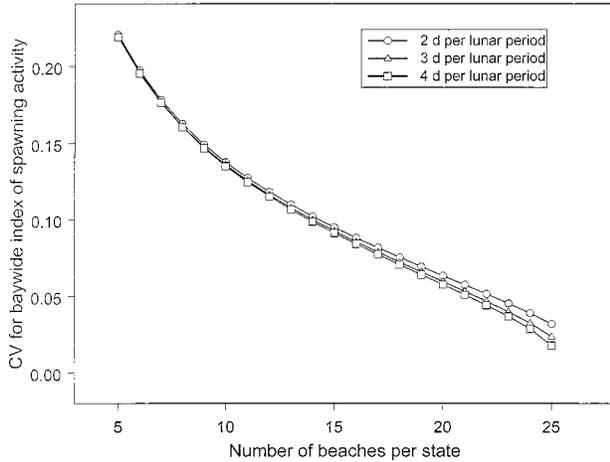


Fig. 5. Coefficient of variation as a function of the number of beaches sampled per state and number of dates sampled per lunar period.

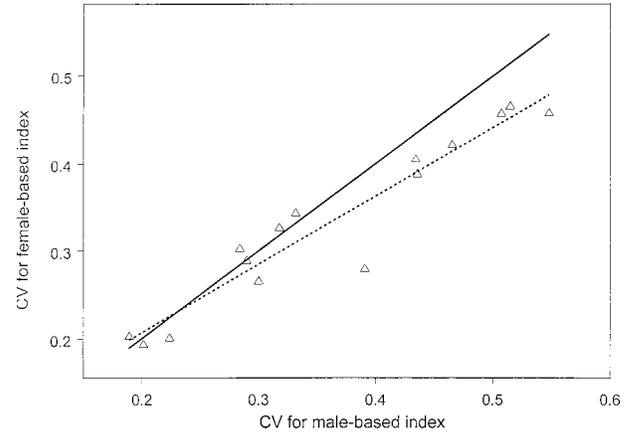


Fig. 6. Coefficient of variation for an index of spawning activity based on females and males for beaches surveyed during May and June 1999 in Delaware Bay. Solid line shows the one-to-one isopleth where CV for males and females based indices would be equal.

index. The degree of clustering and the CV for spawning males and females indicated that the distribution of males was more variable than of females (Fig. 6), so a female-based index of spawning activity tended to be more precise.

Probability of detecting a decline in spawning activity (statistical power) is determined by sample size, sampling variance, decline in spawning activity (magnitude and form), and Type I error rate. Because we found that sampling variance was chiefly determined by number of beaches, we focused on the effect of the number of beaches on statistical power and set the number of sampling occasions to three nights per lunar period and the number of quadrats to 100 per beach (Table 3).

Power to detect declines increased with number

of sampled beaches, duration of monitoring, Type I error rate (α), and magnitude of the decline (Table 3). For 5 yr of monitoring, power did not exceed 0.80 unless the decline was at least 50% when $\alpha = 0.10$. Increasing α to 0.20 resulted in adequate power for smaller declines. For example, power exceeded 0.80 ($\alpha = 0.20$) for a decline $\geq 33\%$ if 15 beaches or more were sampled per state annually for 5 yr. A 25% decline over 5 yr was not detectable (power < 0.80) for ≤ 15 beaches sampled per state but a 25% decline over 10 yr was detectable (power ≥ 0.80) if 15 beaches were sampled per state and $\alpha = 0.10$, or if 13 beaches were sampled per state and $\alpha = 0.20$.

Effort for the 1999 survey as scheduled was 894

TABLE 3. Sample size (beaches per state) for a survey of spawning horseshoe crabs and the resulting cost, coefficient of variance (CV), and power to detect exponential decline in spawning activity. In all cases, 3 nights were sampled in each of the 4 lunar periods in May and June; 100 quadrats were sampled per beach per night. Calculations were based on 25 accessible beaches per state, 7-d lunar periods, and 1-km long beaches. Unit of measurement for cost was person-nights or the number of participants required if each participant surveys only 1 night. Power was based on a two-tailed *t*-test and the assumption that CV was a nonlinear function of spawning activity. Power ≥ 0.80 are shown in italics.

No. of Beaches Sampled per State	Cost	CV for Baywide Index	5 yr of Monitoring						10 yr of Monitoring					
			$\alpha = 0.10$			$\alpha = 0.20$			$\alpha = 0.10$			$\alpha = 0.20$		
			25% Decline	33% Decline	50% Decline	25% Decline	33% Decline	50% Decline	25% Decline	33% Decline	50% Decline	25% Decline	33% Decline	50% Decline
5	360	0.22	0.17	0.22	0.39	0.30	0.38	0.59	0.27	0.41	0.73	0.43	0.57	0.85
6	432	0.20	0.18	0.24	0.44	0.32	0.41	0.64	0.31	0.46	0.79	0.46	0.63	0.90
7	504	0.18	0.20	0.27	0.50	0.34	0.45	0.70	0.35	0.53	0.86	0.51	0.69	0.94
8	576	0.16	0.22	0.32	0.57	0.38	0.50	0.77	0.41	0.61	0.92	0.57	0.76	0.97
9	648	0.15	0.23	0.34	0.62	0.40	0.53	0.81	0.44	0.66	0.95	0.61	0.80	0.98
10	720	0.14	0.25	0.37	0.67	0.42	0.57	0.85	0.49	0.71	0.97	0.65	0.84	0.99
11	792	0.13	0.28	0.41	0.72	0.45	0.61	0.88	0.53	0.76	0.98	0.70	0.88	1.00
12	864	0.12	0.30	0.45	0.77	0.49	0.66	0.92	0.59	0.82	0.99	0.75	0.91	1.00
13	936	0.11	0.34	0.51	0.83	0.53	0.71	0.95	0.65	0.87	1.00	0.80	0.95	1.00
14	1,008	0.10	0.38	0.57	0.88	0.58	0.77	0.97	0.72	0.92	1.00	0.85	0.99	1.00
15	1,080	0.09	0.43	0.64	0.93	0.64	0.83	0.99	0.80	0.96	1.00	0.90	0.99	1.00

person-nights. That cost is similar to the cost of sampling 12 or 13 beaches per state under the proposed design, which calls for sampling fewer dates than in 1999 (Table 3). Effort allocated to sampling dates in 1999 should be shifted to sampling more beaches in future surveys to increase power to detect trends.

Discussion

In 1999, spawning was varied geographically and associated with lunar phase. Spatial and temporal distribution of spawning was moderated by wave height. During the new moon in May, spawning was greater on New Jersey beaches at a time when wave height might have prevented spawning on Delaware beaches. As waves subsided and the full moon in May approached, spawning increased on the Delaware side of the bay. As predicted by Shuster and Botton (1985) spawning in Delaware Bay did not follow lunar phase as sharply as Rudloe (1980) observed for beaches in Apalachee Bay, Florida. Results contradict the assertion of Shuster and Botton (1985) that spawning activity is greatest along the Cape May shore of New Jersey unless prevailing northwesterly winds drive spawning to wind protected Delaware Bay shore. In 1999, spawning was heaviest on the Delaware shore around the full moon in May in spite of low waves in New Jersey during the new and full moons in May. Spawning tapered off in June throughout the bay.

We recommend that the spawning index be based on females for three reasons. First, female abundance determines reproductive potential, which is an important parameter in population assessment. Egg production available for shorebirds is also a function of female abundance (as well as abiotic factors that cause scoring and deposition of beach sediment). Second, the distribution of males was more variable than distribution of females. Third, counting only spawning females would require less time to survey a beach because the majority of the total count was males (ASMFC 1998b; Maio 1998). A female-based index would be biologically informative, more sensitive to trend detection, and less costly than an index based on males or total counts.

Concentration of spawning around the new or full moons suggests that future surveys need only sample during those times. We recommend sampling during the 7-d periods centered on the new and full moons of May and June because that temporal sampling frame minimized among-beach variance in spawning activity relative to the mean. The validity of this restricted temporal sampling frame depends on the assumption that trends in spawning activity during lunar periods are parallel to

trends in spawning activity during periods of neap tides. The pattern of decline in spawning activity reported by Widener and Barlow (1999) supports this assumption. Widener and Barlow (1999) surveyed spawning horseshoe crabs at a Cape Cod beach three times over 15 yr and found that spawning activity during the peaks (lunar period) and valleys (neap tidal periods) dropped uniformly.

Volunteer training, coordination, and dedication can greatly influence the success and credibility of the survey. For example, Slaughter Beach was only sampled around the full moons and several beaches were sampled only 1 out of the scheduled 3 d around a new or full moon. The effect of missing data on the index of spawning activity depends on the reason for not sampling as scheduled. If the beaches were omitted at random, then the index will not be biased. If the decision not to sample was influenced by a perception of the magnitude of spawning then the index will be biased. This issue is critical in a volunteer-based survey because volunteers, unless instructed otherwise, may decide subjectively whether to sample or not.

It is important to note that the value of the spawning survey to conservation of horseshoe crab and associated species is in trend detection. The index of spawning activity cannot be used to estimate population abundance or population size of horseshoe crabs in Delaware Bay. Consider that we counted horseshoe crabs only within 1 m of the high tide line, on the beaches near land-based access points, and on the highest of the daily high tides. For practical reasons we restricted the sampling frames. Certain habitat, such as inaccessible beaches and tidal creeks, were left out of the beach-level sampling frame. It is not known what proportion of the horseshoe crab population spawns in a given year, so we did not attempt to estimate population size or absolute magnitude of spawning activity. The underlying assumption was that the trends in spawning activity for the Delaware Bay could be represented by the trends in spawning activity on accessible beaches during the weeks centered on new or full moons. (This assumption can be tested by a separate study to compare trends in spawning activity among spawning habitats.) Use of the proposed index to compute a population size by simply expanding the number by the length of beaches in Delaware Bay would be highly biased and misleading.

The most effective way to increase the probability of detecting a trend in spawning activity was to increase the number of beaches sampled (Fig. 6). Sampling more quadrats on a beach did not significantly increase the precision of a baywide index. Sampling more dates per lunar period increased precision, although the increase was minor

in comparison to the effect of sampling more beaches. We recommend sampling multiple dates (3 d per lunar period) for statistical and pragmatic reasons. Stratification on lunar period (Appendix) is important for reducing sampling variance; at least 2 d must be sampled per lunar period to estimate sampling variance. We believe that 3 d are preferable to 2 because 3 d allows dates to be selected systematically in a balanced pattern (before, during, and after the new or full moon). Rather than sampling more than 3 d per lunar period, effort should be allocated to sampling more beaches. Power analysis, which highlights the importance of sampling more beaches, is an approximation useful for planning. The survey design that we present calls for an independent random selection of accessible beaches in each year. Power to detect change could be improved somewhat by a design that allows a combination of fixed and random selection of beaches (Fuller 1999).

Programs to monitor horseshoe crabs in estuaries other than Delaware Bay will share the need for a formal survey design to select beaches and dates to assure proper statistical inference (i.e., the first and second stages of sampling). The need to subsample beaches using quadrats (i.e., the third stage of sampling) will be applicable only where spawning activity is high enough to prevent a complete count. Our conclusion that number of beaches and temporal stratification are important determinants of power to detect trends should apply generally to large scale monitoring of horseshoe crab spawning activity.

Trend detection involves balancing risks. The risk of concluding a decrease in spawning activity (or increase) when in fact it has not changed is the Type I error rate (α). We call this risk the fishers' risk because falsely concluding a decrease in spawning activity could lead unnecessarily to harvest reduction. In contrast, there is the species' risk, which is the risk of failing to detect a true and biologically meaningful decrease in spawning activity. In statistical notation, species' risk is Type II error rate (β), and statistical power is its complement ($1 - \beta$). We call this the species' risk because failing to detect a biologically important change in spawning activity could lead to a stock collapse or declines in species that depend on horseshoe crabs. The fishers' risk, which is measured by α , is set prior to data analysis and is known. The species' risk, which is only known if statistical power is calculated, depends on sample size, sampling variance (which is a function of survey design and underlying population variance), true trend in spawning activity, and the fishers' risk (α). All else being equal, the lower the fishers' risk the higher the species' risk. To balance the two risks, survey de-

sign must be improved to reduce sampling variance or sample size needs to be increased.

Under the precautionary principle (Buhl-Mortensen and Welin 1998), it would be risk prone to assume that species' risk is low unless it has been shown to be low by calculation of statistical power, even when survey data shows no statistically significant trend. Peterman and M'Gonigle (1992) outline three outcomes when statistical power is incorporated into monitoring programs and regulatory decision making. The first is a case where survey data shows a statistically significant decline of a magnitude deemed to be biologically important. As a result, harvest restriction would be recommended. Second is the case where survey data fails to show a statistically significant decline, and statistical power has been evaluated and found to be high for a decline deemed to be biologically important. In this case, harvest restriction would not be recommended. Third is the case where survey data fails to show a statistically significant decline, and statistical power has been evaluated and found to be low for a biologically important decline. In such a case, species' risk is high because the possibility of an important decline cannot be ruled out. The precautionary principle stipulates that, in the face of this uncertainty, harvest reduction should be recommended as a risk adverse strategy.

Many sampling programs that have encountered horseshoe crabs have failed to show significant coastwide declines (ASMFC 1998c). Although statistical power has not been considered explicitly in assessments of these sampling programs, past efforts to monitor horseshoe crabs have suffered from methodological inconsistencies and gear inefficiencies (ASMFC 1998c). In the face of this uncertainty and consistent with the precautionary principle, ASMFC (1998a) has adopted a risk adverse management strategy and has called for harvest reduction. Others may point to the lack of statistically significant declines and argue against harvest reduction. However, in doing so they take on a burden to show that the surveys they cite have high statistical power.

We proposed a survey and assessed its statistical power to detect declines in baywide spawning activity. The amount of acceptable risk and magnitude of decline deemed biologically important are judgments that must be debated by stakeholders in the management of horseshoe crabs. We argue for a balance of risks (i.e., fishers' risk = species' risk), suggest biologically important declines should be determined through population modeling, and present results in Table 3 that can be used to guide sample size decisions.

ACKNOWLEDGMENTS

We are grateful to the many volunteers who lost sleep to count horseshoe crabs. Dr. Carl N. Shuster, Jr. generously shared his insights into the life history of *L. polyphemus*, the fascinating animal to which he has devoted much of his professional life. We thank Marion Pohlman, Sherry Hartley, Glen Garvey, Bruce Luebke, Annie Larson, Nate Thomas, Tedor Whitman, and Tom O'Connell for helping to coordinate volunteers. Sherry Hartley was also instrumental in coordinating data entry. Joe Margraf and Kathy Jo Maio provided critical direction to the horseshoe crab survey workshop. Thanks go to Carl Shuster and Mark Botton for comments on an earlier draft, which helped us improve the paper.

LITERATURE CITED

- ATLANTIC STATES MARINE FISHERIES COMMISSION. 1998a. Interstate fishery management plan for horseshoe crab. Fishery Management Report No. 32, Atlantic States Marine Fisheries Commission, Washington, D.C.
- ATLANTIC STATES MARINE FISHERIES COMMISSION. 1998b. Terms of reference and advisory report for the horseshoe crab stock assessment peer review. Stock Assessment Report No. 98-01, Atlantic States Marine Fisheries Commission, Washington, D.C.
- ATLANTIC STATES MARINE FISHERIES COMMISSION. 1998c. Horseshoe crab stock assessment report for peer review. Stock Assessment Report No. 98-01 (Supplement), Atlantic States Marine Fisheries Commission, Washington, D.C.
- BERKSON, J. AND C. N. SHUSTER, JR. 1999. The horseshoe crab: The battle for a true multiple-use resource. *Fisheries* 24:6-12.
- BOTTON, M. L., R. E. LOVELAND, AND T. R. JACOBSEN. 1994. Site selection by migratory shorebirds in Delaware Bay, and its relationship to beach characteristics and abundance of horseshoe crab (*Limulus polyphemus*) eggs. *The Auk* 111:605-616.
- BOTTON, M. L. AND J. W. ROPES. 1987. The horseshoe crab, *Limulus polyphemus*, fishery and resource in the United States. *Marine Fisheries Review* 49:57-61.
- BROCKMANN, H. J. 1990. Mating behavior of horseshoe crabs, *Limulus polyphemus*. *Behaviour* 114:206-220.
- BUHL-MORTENSEN, L. AND S. WELIN. 1998. The ethics of doing policy relevant science: The precautionary principle and the significance of non-significant results. *Science and Engineering Ethics* 4:401-412.
- BURGER, J., L. NILES, AND K. E. CLARK. 1997. Importance of beach, mudflat and marsh habitats to migrant shorebirds on Delaware Bay. *Biological Conservation* 79:283-292.
- CASTRO, G. AND J. P. MYERS. 1993. Shorebird predation on eggs of horseshoe crabs during spring stopover on Delaware Bay. *The Auk* 110:927-930.
- CHRISTMAN, M. C. 2000. A review of quadrat-based sampling of rare, geographically clustered populations. *Journal of Agricultural, Biological, and Environmental Statistics* 5:168-201.
- ELLIOTT, J. M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Freshwater Biological Association, Scientific Publication No. 25. Ambleside, Cumbria, UK.
- FINN, J. J., C. N. SHUSTER, JR., AND B. L. SWAN. 1991. *Limulus* Spawning Activity on Delaware Bay Shores 1990. Finn-Tech Inc., Cape May, New Jersey.
- FULLER, W. A. 1999. Environmental surveys over time. *Journal of Agricultural, Biological, and Environmental Statistics* 4:331-345.
- GERRODETTE, T. 1993. TRENDS: Software for a power analysis of linear regression. *Wildlife Society Bulletin* 21:515-516.
- HEDAYAT, A. S. AND B. K. SINHA. 1991. Design and Inference in Finite Population Sampling. John Wiley & Sons, Inc., New York.
- MAIO, K. J. 1998. Quantitative description of temporal and spatial patterns in spawning activity of the horseshoe crab, *Limulus polyphemus*. Master's thesis. University of Maryland Eastern Shore, Princess Anne, Maryland.
- PETERMAN, R. M. AND M. M'GONIGLE. 1992. Statistical power analysis and the precautionary principle. *Marine Pollution Bulletin* 24:231-234.
- RUDLOE, A. 1980. The breeding behavior and patterns of movement of horseshoe crab, *Limulus polyphemus*, in vicinity of breeding beaches in Apalachee, Florida. *Estuaries* 3:177-183.
- RUDLOE, A. 1983. The effect of heavy bleeding on mortality of the horseshoe crab, *Limulus polyphemus*, in the natural environment. *Journal of Invertebrate Pathology* 42:167-176.
- SHUSTER, JR., C. N. AND M. L. BOTTON. 1985. A contribution to the population biology of horseshoe crabs, *Limulus polyphemus* (L.), in Delaware Bay. *Estuaries* 4:363-372.
- THOMPSON, S. K. 1992. Sampling. John Wiley & Sons, Inc., New York.
- TSIPOURA, N. AND J. BURGER. 1999. Shorebird diet during spring migration stopover on Delaware Bay. *The Condor* 101:635-644.
- WIDENER, J. W. AND R. B. BARLOW. 1999. Decline of a horseshoe crab population on Cape Cod. *Biological Bulletin* 197:300-302.

Received for consideration, August 16, 2000

Accepted for publication, June 13, 2001

Appendix

The survey design that we recommend can be described as a three-stage sampling design. At the first stage, beaches are selected; at the second stage, dates within beaches are selected; and at the third stage, quadrats within dates and beaches are selected. (By beach we mean a section of beach the length of which can be sampled at the third stage of sampling.) Selection at the first and second stages is stratified. Beaches are stratified by state (i.e., New Jersey or Delaware) with equal allocation and random sampling within strata. Dates are stratified by lunar period (e.g., 7-d period centered on the new or full moons in May and June) with equal allocation and systematic sampling of nights within lunar period (e.g., selecting 2 d prior to, the day of, and 2 d after the new or full moon). Quadrats on a beach are selected by systematic sampling with 2 random starts. Given the design of the survey we derive formulae for an estimate of the spawning index and for its variance.

Notation:

- H = number of beach-level strata
 N_h = number of accessible beaches in the h^{th} strata
 P = number of lunar periods in the spawning season
 T = number of dates in a lunar period
 L_{hi} = length of i^{th} beach in h^{th} strata
 S_{hi} = number of possible systematic samples on i^{th} beach in h^{th} strata
 n_h = number of beaches sampled in h^{th} strata
 t = number of dates sampled during a lunar period
 s = number of random starts for a systematic sample
 m_{hijkl} = number of quadrats in the l^{th} systematic sample on the k^{th} date in the j^{th} lunar period on i^{th} beach in h^{th} strata
 y_{hijklp} = count of spawning female horseshoe crab in the p^{th} quadrat in the l^{th} systematic sample on k^{th} date in j^{th} period on i^{th} beach in h^{th} strata

The index of spawning activity for the k^{th} date of the j^{th} period on the i^{th} beach in the h^{th} strata is

$$\bar{y}_{hijk} = \frac{S_{hi} \sum_{l=1}^s y_{hijkl}}{L_{hi} s},$$

where $y_{hijkl} = \sum_{p=1}^{m_{hijkl}} y_{hijklp}$ is the count of spawning female horseshoe crabs in each systematic sample. To compute the index for the i^{th} beach in the h^{th} strata, we have

$$\bar{y}_{hi} = \frac{\sum_{j=1}^P \bar{y}_{hij}}{P},$$

where $\bar{y}_{hij} = (t)^{-1} \sum_{k=1}^t \bar{y}_{hijk}$. Because beaches vary in length, we use the ratio estimator (Thompson 1992, p. 60) to estimate the index for the h^{th} strata

$$\bar{y}_h = r_h \bar{L}_h,$$

where $r_h = \sum_{i=1}^{n_h} \bar{y}_{hi} / \sum_{i=1}^{n_h} L_{hi}$ and $\bar{L}_h = \sum_{i=1}^{n_h} L_{hi} / N_h$. Finally, because of stratification of beaches, the baywide index is

$$\bar{y} = \frac{\sum_{h=1}^H L_h \bar{y}_h}{\sum_{h=1}^H L_h},$$

where $L_h = \sum_{i=1}^{N_h} L_{hi}$. Variance of the baywide index can be estimated by

$$\widehat{\text{var}}(\bar{y}) = \left(\sum_{h=1}^H L_h \right)^{-2} \sum_{h=1}^H L_h^2 \widehat{\text{var}}(\bar{y}_h),$$

where the estimate of strata-specific variance is based on three stage sampling (Thompson 1992, p. 137)

$$\begin{aligned} \widehat{\text{var}}(\bar{y}_h) &= \left(\frac{N_h - n_h}{N_h} \right) \frac{\hat{\sigma}_h^2}{n_h} \\ &+ \frac{1}{N_h n_h P^2} \sum_{i=1}^{n_h} \sum_{j=1}^P \\ &\times \left[\left(\frac{T-t}{T} \right) \frac{\hat{\sigma}_{hij}^2}{t} + \frac{1}{Tt} \sum_{k=1}^t \left(\frac{S_{hi} - s}{S_{hi}} \right) \frac{\hat{\sigma}_{hijk}^2}{s} \right]. \end{aligned}$$

There are three population variance estimates in $\widehat{\text{var}}(\bar{y}_h)$: variance among beaches ($\hat{\sigma}_h^2$), variance among dates within lunar period and beach ($\hat{\sigma}_{hij}^2$) and variance among systematic samples within date, lunar period, and beach ($\hat{\sigma}_{hijk}^2$). To estimate these population variances we have

$$\hat{\sigma}_h^2 = \frac{\sum_{i=1}^{n_h} (\bar{y}_{hi} - r_h L_{hi})^2}{n_h - 1},$$

$$\hat{\sigma}_{hij}^2 = \frac{\sum_{k=1}^t (\bar{y}_{hijk} - \bar{y}_{hij})^2}{t - 1}, \quad \text{and}$$

$$\hat{\sigma}_{hijk}^2 = \frac{\sum_{l=1}^s (\bar{y}_{hijkl} - \bar{y}_{hijk})^2}{s - 1}.$$