North American Journal of Fisheries Management

An Empirical Approach for Estimating the Precision of Hydroacoustic Fish Counts by Systematic Hourly Sampling

Yunbo Xie[a] & Fiona J. Martens[a]

[a] Pacific Salmon Commission, 600-1155 Robson Street, Vancouver, British Columbia V6E 1B5, Canada
Published online: 08 May 2014.

PLEASE SCROLL DOWN FOR ARTICLE

ARTICLE

An Empirical Approach for Estimating the Precision of Hydroacoustic Fish Counts by Systematic Hourly Sampling

Yunbo Xie* and Fiona J. Martens

Pacific Salmon Commission, 600-1155 Robson Street, Vancouver, British Columbia V6E 1B5, Canada

Abstract

Systematic hourly sampling is a logistically favorable method for hydroacoustic estimation of fish passage through constricted passageways or finite sections of a river monitored by fixed-location sonar systems. Similar to simple random and hourly stratified random sampling, systematic hourly sampling produces unbiased estimates of fish passage. Variances of estimates produced by systematic hourly sampling are determined by the sampling fraction, f, variance of the underlying process, S^2, and correlations among all hourly strata of the process. This intrahour correlation dependency makes it difficult to accurately evaluate the precision of the estimate when the complete temporal pattern of the process is unavailable. Fish passage is rarely counted continuously, so the uncertainty of estimated mean by systematically subsampled fish counts is traditionally estimated using variance estimators. Variance estimates by these estimators are likely biased and subject to potentially large errors. We present an alternative approach for estimating the precision of systematic hourly sampling using an empirical relation between precision and sampling fraction established from continuous fish counts acquired by imaging sonar for a wide range of migration scenarios of Pacific salmon *Oncorhynchus* spp. in the lower Fraser River. The empirical relation indicates that a CV (100·SD/mean) of 5.5% can be achieved by the systematic hourly counting of fish passage at a counting effort of 10 min/h, while increasing the counting effort to 20 min/h only leads to a marginally improved CV of 4.1%. The data-based analysis also shows that the precision of systematic sampling can be gained more efficiently and consistently by increasing the sampling rate than by lengthening the sampling time.

Most fixed-location hydroacoustic systems use time-stratified systematic sampling to sample fish passage in rivers and constricted passageways at hydroelectric facilities (Steig et al. 1988; Enzenhofer et al. 1998; Maxwell and Gove 2004; Xie et al. 2005). While estimation methods based on simple random, stratified random, or systematic sampling all produce unbiased estimates of a finite population, systematic sampling is a logistically favorable fish-counting method that facilitates systematic allocations of sampling effort of visual or acoustic fish counters in a field program. For processes exhibiting large intrastrata variability, systematic sampling is more precise than simple random sampling (Cochran 1977). Skalski et al. (1993) evaluated precisions of several sampling methods, including both time-stratified random sampling and systematic sampling, using continuous hydroacoustic fish counts acquired at two

hydroelectric dams on the Columbia River. The study concluded that stratified random sampling is a statistically more favorable sampling method than others. It is intuitive that the precision of an estimate for a finite population, N, such as the number of fish passing through a counting station over a 24-h period, depends on (1) the variability of the underlying process (quantified by process variance, S^2) and (2) the sampling fraction, f, of the process.

For simple random and stratified random sampling, once f is chosen and S^2 is estimated from sampled units, the corresponding precision of estimates is completely determined by the two variables. There is also a straightforward expectation that increasing the sampling fraction improves the precision of estimates produced by either of the two sampling methods. However, it is not as intuitively understood that, in addition to f and

S^2, the precision of estimates produced by systematic sampling also depends on the temporal correlation ρ among all the time strata of the process. Since ρ cannot be accurately estimated from inadequately subsampled data without imposing highly simplified assumptions about the underlying process, such as assuming that the process is a Gaussian random noise, it is not possible to derive an unbiased estimate of precision from the subsampled data. The impact of ρ on the precision of systematic sampling varies greatly with the phase structure of the process and can produce counterintuitive outcomes. The phase structure of a fish count time series over a stratified scale can be quantified by the relative change in fish counts within this temporal scale. Based on the analyses of the continuous fish count data, Skalski et al. (1993) found that increasing sampling effort did not guarantee decreases in sampling variance for systematic sampling, making it difficult to predict sampling precision with alternative levels of effort. Using systematically sampled fish counts from imaging sonar data in a spawning tributary of the Fraser River, British Columbia, for Sockeye Salmon *Oncorhynchus nerka*, Lilja et al. (2008) assessed the relationship between precision and sampling effort for systematic sampling by ignoring temporal correlations in the underlying fish passage time series. The study concluded that the precision of systematic sampling increases monotonically with increasing sampling effort. This conclusion contradicts the findings of Skalski et al. (1993) and Cochran (1977) that showed that the precision of systematic sampling does not generally follow a monotonic relationship with sampling effort.

While it would be ideal to acquire a complete census of fish passage by continuously counting each fish passing a monitoring site, the full-counting method is neither necessary nor practical for estimating migratory fish abundance in rivers and streams. The common practice is to count fish passage for 10–20 min/h either visually or acoustically and to linearly expand the hourly subsampled counts to the full hourly counts (Becker 1962; Maxwell and Gove 2004; Xie et al. 2005; Holmes et al. 2006; Enzenhofer et al. 2010). When acoustically counting migrating fish in a large riverine space, the sound beam is often required to scan the water column by systematically alternating its aiming angle and sounding range within each hour (Xie et al. 2005; Enzenhofer et al. 2010). Such spatially stratified sampling designs make it impossible for the system to obtain full-count time series of fish passage. Since full-count data are rarely available, the uncertainty of the total fish count derived from nonreplicated, systematically subsampled counts is usually estimated by a set of variance estimators recommended for systematic samples (Wolter 1984, 2007). Although these estimators offer a practical approach to address the uncertainty they are biased (Skalski et al. 1993; Reynolds et al. 2007).

In this paper, we present an alternative approach for estimating the precision of systematic sampling using an empirical relation between precision and sampling fraction. This relation is established from a set of continuous fish count data for a wide range of migration scenarios observed at a monitoring site

on the lower Fraser River. The full-count data were generated from acoustic recordings of six daily fish passages by a dual-frequency identification sonar (DIDSON) system (Belcher et al. 2002). We demonstrated the use of this empirical method to estimate the precision of systematically subsampled fish counts with a numerical example provided in the Appendix. We also used the full-count data to reveal the bias of precision estimates by the traditional estimators. Finally, our data-based analyses indicated that increasing the sampling rate was more efficient in gaining the precision of systematic sampling than lengthening the sampling time for the same sampling effort. The focus of this study was on the precision of systematic fish counts, and steps were taken in the data collection and processing stages to ensure that the full-count data were free of systematic bias due to acoustic equipment and data-processing errors.

METHODS

Study site.—The fish-counting site is located at 49°08.175′N and 122°16.466′W, approximately 70 km upstream from the Fraser River's terminus at the Strait of Georgia. This stretch of the river is subjected to tidal influence with a daily change in water level of up to 1.5 m in September. The river is approximately 450 m wide during the spring freshet and has a maximum water depth of nearly 20 m. The water is turbid with a near-zero visibility preventing visual detections of fish targets. The stronger offshore current in the deepest channel near the right bank forces 70% of the fish, primarily Sockeye Salmon, to migrate near the left bank (Xie et al. 2012a).

Fish count data.—The data for this study were acquired by a side-looking DIDSON sonar system in the 2004 and 2010 Sockeye Salmon return seasons. Fish migrating within 30 m of the left bank were insonified by the system. The aim of the sonar was carefully chosen to ensure that it captured the bulk portion of the migration over the sounding range. A total of 6 d of fish-passage data were recorded continuously for 55 min/h by the DIDSON, with a 5-min recording pause between adjacent hours to allow the system to open and close hourly files. The key acquisition parameters for the data are summarized in Table 1. A sonar image clip of 82 salmon-sized fish in the DIDSON's field of view is shown in Figure 1. The recorded image files were read by four experienced readers to produce six time series of fish counts. Each data point in the time series is the net upstream count, which is defined as the upstream minus the downstream counts (Xie et al. 2002) of fish swimming through the field of view in a 5-min time interval. The criteria for counting fish were based on the recognition of the distinctive swimming behavior of fish displayed in the sonar image. To assess the counting error between readers, a total of 24 h of DIDSON image files were chosen from three data sets with hourly passage rates ranging from 600 to 5,800 fish per hour. These files were counted independently by three of the four readers. The counting error was found to be less than 3% in terms of the CV ($100 \cdot$ SD/mean) with the three readers producing statistically similar counts

TABLE 1. Summary of key acquisition parameters for continuous recordings of fish passage data acquired by the DIDSON sonar from the left bank of the study site in 2004 (Aug 22) and 2010 (Aug 17–19 and 21 and Sep 25).

Parameter	Date					
	Aug 22	Aug 17	Aug 18	Aug 19	Aug 21	Sep 25
DIDSON model	Standard	Standard	Standard	Standard	Standard	Long range
Start range (m)	1.5	3.33	3.33	3.33	3.33	12.5
End range (m)	35.8	23.33	23.33	23.33	23.33	32.5
Vertical aim (°)	−4	−8	−8	−8	−8	−8
Bearing (°)	340	340	340	340	340	340
Frequency (MHz)	1.1	1.8	1.8	1.8	1.8	1.2
Frame rate (frames/s)	8	4	4	4	4	5

($P = 0.214$, 0.427, and 0.482, respectively, for the three data sets). These 5-min counts formed the basic units of the time series for the analysis. There were a total of 12 units in each hour. The 12th unit, y_{12}, was interpolated from y_{11}, the 11th unit of the hour, and y_{13}, the first unit of the following hour using a modified version of the simulation model by Reynolds et al. (2007) to account for the possibility of downstream counts being greater than upstream counts as follows:

$$y_{12} = y_{13} + \varepsilon, \qquad (1)$$

FIGURE 1. One frame of a DIDSON sonar image showing 82 upstream-moving fish in the field of view. The fish were estimated from the data to have a mean length of 63.5 cm and an average swimming speed of 0.57 m/s. The data were collected on September 25, 2010, by a side-looking DIDSON sonar system from the left bank.

where $\varepsilon \sim uniform\,(-|y_{13} - y_{11}|, |y_{13} - y_{11}|)$.

Using the observed counts (11 counts per hour), we examined the effect of this simulation model on the first-lag autocorrelation (ac_1) by replacing 1 of the 11 counts in each hour with the simulated count and tested the significance of Δac_1, the difference of ac_1 between the simulated time series and the true time series for each replacement. We performed the test over 5,000 times for each of the six time series used in this study (see Figure 2). The results from these tests indicated that the probability of Δac_1 being significantly greater than zero ($P < 0.05$) was less than 3%. Therefore, using the simulated count by equation (1) to interpolate the 12th unit for each hour is unlikely to increase the autocorrelation of the underlying time series. The time series of the 5-min counts can be readily converted to series of 10-, 15-, 20-, or 30-min counts for the analysis of the relation between precision and sampling effort.

Sampling of the full-count data.—The 5-min counts can be treated as basic units of a finite population $Y = [y_1, y_2, \dots, y_N]^T$ with $N = 288$ over a 24-h period. One can stratify the units by hour and express Y as a $k \times n$ matrix $[y_{ij}]$ where y_{ij} is the ith unit located in the jth hour with $i = 1, 2, \dots, k, j = 1, 2, \dots, n$, and $k \times n = N$. Since Y is stratified into 24 hourly time slots, n (the number of stratifications) is 24. Using the 5-min counts as basic units, k (the number of units per hour) is 12. From the full-count data Y, one can draw n units by one of three sampling methods: (1) simple random sampling, which randomly draws n units regardless of their hourly stratification; (2) hourly random sampling, which randomly draws one unit from each of the n hourly time slots; and (3) systematic hourly sampling, which sequentially draws one of the k units from each of the n hourly time slots.

The total number of samples by the three methods is $\frac{N!}{n!(N-n)!}$, k^n and k, respectively. All three sampling methods produce unbiased estimates of population mean $\bar{Y}$ with sample mean $\bar{y}$, i.e., $E(\bar{y}) = \bar{Y}$.

True variances of estimated mean.—The variances of estimated mean $V(\bar{y})$ of the three sampling methods are given by the following equations (Cochran 1977):
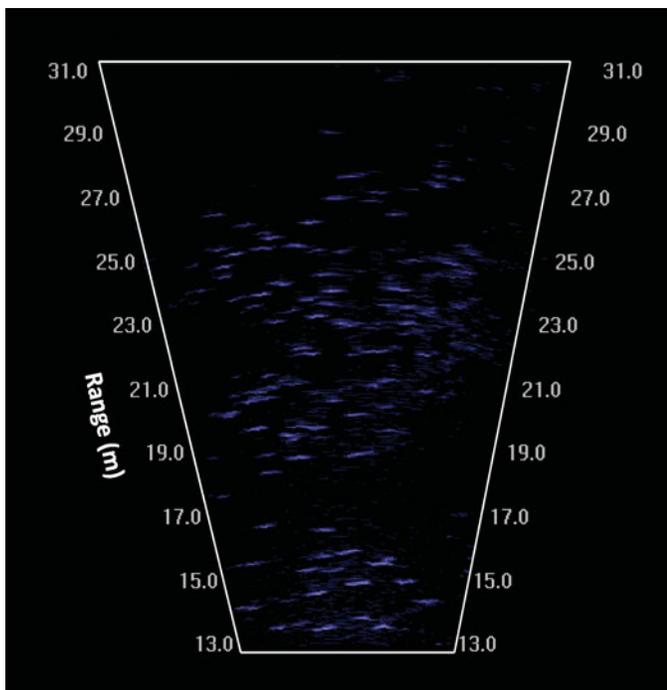
for simple random sampling,

$$V(\bar{y}_{ran}) = \frac{S^2}{n} \cdot (1 - f); \qquad (2)$$

for hourly stratified random sampling,

$$V(\bar{y}_{st}) = \frac{S^2_{wst}}{n} \cdot (1 - f); \qquad (3)$$

for systematic hourly sampling,

$$V(\bar{y}_{sy}) = \frac{S^2_{wst}}{n} \cdot (1 - f) \cdot [1 + (n - 1) \cdot \rho]. \qquad (4)$$

For comparison purposes, we present estimated precisions for the three methods in the Results. In equations (2), (3), and (4), the subscripts *ran*, *st*, and *sy* that accompany $\bar{y}$ represent sample means from simple random, stratified random, and systematic sampling, respectively, and the subscript *wst* is the abbreviation for within strata. Other variables are defined as follows:

$f = \frac{n}{N}$ is sampling fraction;

$S^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{Y})^2}{N-1}$ is the total variance of the population;

$S^2_{wst} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{.j})^2}{n(k-1)}$ is the within-hour variance of y_{ij};

$\rho = \frac{2}{n(n-1)(k-1)}\sum_{i=1}^{k}\sum_{j \neq u}\frac{(y_{ij} - \bar{y}_{.j})(y_{iu} - \bar{y}_{.u})}{S^2_{wst}}$ is the intrahour correlation coefficient between pairs of units in the same sample.

The variable S^2_{wst} is at least an order of magnitude smaller than S^2 for processes exhibiting larger intra- than within-hour variability. Many time series data of daily fish passages, including the data presented in this paper, exhibit large intrahour variability due to tidal, diurnal, and fisheries effects on their migration behavior (Levy and Cadenhead 1995; Reynolds et al. 2007; Xie et al. 2012a). This intrahour pattern makes hourly random sampling more precise than simple random sampling. However, a definitive relationship cannot be drawn about the precisions between hourly random sampling and systematic hourly sampling. According to equation (4), systematic hourly sampling is less precise than hourly random sampling if $\rho > 0$ but it is more precise if $\rho < 0$. For cases where there is little intrahour correlation ($\rho = 0$), the two methods are equally precise. However, the impact of ρ on the precision of systematic sampling is amplified through a multiplying factor of $(n - 1)$. As evidenced from equation (4), the variance of systematic sampling approaches zero when the intrahour ρ (or referenced as intrastrata correlation for more general applications) approaches a negative critical value defined as $\rho_c = -1/(n - 1)$. With a stratification of the full-count data into 24 h, it follows that $(n - 1) = 23$ and $\rho_c = -0.0435$. Therefore, a seemingly small valued ρ can have a large impact on the precision. Furthermore, if the hourly phase structure of a 24-h time series of fish counts results in ρ approaching -0.0435, the systematic hourly sam-

pling can yield extremely precise estimates. Cochran (1977) showed that stratified systematic sampling could yield a considerably higher precision than would random sampling for a time series exhibiting alternating trends that produce a negative coefficient of intrastrata correlation.

Metrics to measure precision.—Sample mean $\bar{y}$ can be converted to the hourly mean by multiplying $\bar{y}$ by k, number of units per hour, to produce an estimated hourly mean of $k\bar{y}$. Using CV as a metric for precision, we can measure the precision of $k\bar{y}$ by

$$CV = \frac{\sqrt{V(k\bar{y})}}{E(k\bar{y})} = \frac{k\sqrt{V(\bar{y})}}{kE(\bar{y})} = \frac{\sqrt{V(\bar{y})}}{\bar{Y}}. \qquad (5)$$

Therefore, the evaluation of the precision of the estimated hourly mean is equivalent to that of the sample mean. As a result, we can use CV to quantify the precision of the three sampling methods from the sample means regardless of the time scale of the units. In addition to CV, other metrics have been chosen to measure the precision of hydroacoustic fish counts such as percent of 95% confidence limits (PCL_{95}) by Lilja et al. (2008) and average percent error (APE) by Chilton and Beamish (1982) and Maxwell and Gove (2007). Assuming $\bar{y} - \bar{Y}$ is normally distributed, it follows that $PCL_{95} = \pm 1.96 \times CV$. Holmes et al. (2006) used both CV and APE to measure the counting error among different readers of DIDSON files and found both metrics led to similar assessments. Coefficient of variation was chosen by Banneheka et al. (1995) to measure the precision of daily salmon escapement estimated hydroacoustically and by Bromaghin and Bruden (1999) to measure the precision of simulated daily escapement models. In this study, we choose CV as a formal metric because of its simple relation to the 95% CIs for normally distributed errors. All the calculations were carried out using statistical software package R, version 2.15.2 (R Foundation for Statistical Computing 2012).

Precision versus sampling fraction for systematic sampling.—With the full-count data, we derived an empirical relation for systematic sampling between precision and sampling fraction via equation (5). We selected hourly fish counts of 5 min ($f = 1/12$), 10 min ($f = 1/6$), 15 min ($f = 1/4$), 20 min ($f = 1/3$), and 30 min ($f = 1/2$) to estimate the CV responses to these sampling fractions. Since the hourly time slot of 60 min is not divisible by 25 min, the 25-min counting units were excluded from this analysis as they may introduce biased estimates of the hourly mean (Skalski et al. 1993). While sample size remains unchanged ($n = 24$) as the hourly sampling time increases from 5 to 30 min, k (the number of systematic samples) decreases from 12 to 2. Using equations (4) and (5), the precision was calculated as a function of f for the six time series of full-count data.

Assessing the bias of variance estimators with the full-count data.—Eight estimators were presented by Wolter (1984) for estimating the variances of systematic sampling from subsamples. These estimators are biased with different magnitudes of error that depend on the temporal pattern of the underlying process. Assessments of the bias through simulation approaches have

been reported by Skalski et al. (1993), Bromaghin and Bruden (1999), and Reynolds et al. (2007). Using simulated data, Reynolds et al. (2007) examined three of these estimators for nonreplicated systematic sampling as follows:

$$\hat{V}_2 = \frac{1-f}{n} \sum_{j=2}^{n} \frac{(s_j - s_{j-1})^2}{2(n-1)}, \tag{6}$$

$$\hat{V}_4 = \frac{1-f}{n} \sum_{j=3}^{n} \frac{(s_j - 2s_{j-1} + s_{j-2})^2}{6(n-1)}, \tag{7}$$

and

$$\hat{V}_5 = \frac{1-f}{n} \sum_{j=5}^{n} \frac{(0.5s_j - s_{j-1} + s_{j-2} - s_{j-3} + 0.5s_{j-4})^2}{3.5(n-4)}, \tag{8}$$

where s_j is the jth unit in a systematic sample that consists of n units. With the full-count time series and the true variance of the sample mean $V_T = V(\bar{y})$, we assessed the bias of variance estimated from equations (6), (7), and (8) for the migration scenarios represented by the six full-count time series. The bias of each variance estimator was measured by a reversed form of relative error RE_B, i.e.:

$$RE_B = \frac{\bar{\hat{V}}_B}{V_T} - 1, \tag{9}$$

where $\bar{\hat{V}}_B$ is the mean of the estimated variances using estimator B over all k systematic samples and a positive RE_B corresponds to overestimation.

Bias of estimated precisions from nonreplicated systematic samples.—In most visual and acoustic fish-counting applications, only a single sample of the k systematic samples is acquired. For instance, if the sample corresponds to the first systematic sample (denoted as S_1) at a sampling fraction of 10 min/h over a 24-h period, the sample will contain 24 fish counts with each count being the number of fish passing a counting site during the first 10 min of each hour. In the absence of replicated samples, the precision of the counts has to be estimated from the single sample via equation (6), (7), or (8). We used the full-count data to assess bias of precision estimates from the single sample for each of the three estimators. The CV estimated from S_1 using estimator B (denoted as $\hat{V}_B\_CV_{S_1}$) was calculated from the following:

$$\hat{V}_B\_CV_{S_1} = \frac{\sqrt{\hat{V}_{S_1}}}{\hat{Y}_{S_1}}, \tag{10}$$

where $\hat{V}_{S_1}$ is the estimated systematic sampling variance from sample S_1 using estimator B and $\hat{Y}_{S_1}$ is the mean of S_1. The

comparison of $\hat{V}_B\_CV_{S_1}$ with the true CV provided a perspective from the real data scenarios on bias when using equation (6), (7), or (8) to estimate the precision from nonreplicated systematic fish counts.

RESULTS

Time Series of Fish-Count Data

A total of six full-count time series of fish passage were obtained from continuous DIDSON sonar recordings of imaging data with a 24-h total fish count ranging from 2,500 to nearly 170,000 (Table 2). The data encompassed a wide range of migration scenarios of low (Figure 2a), medium (Figure 2b, c), and high daily abundance (Figure 2f). The temporal pattern of the data from August 22, 2004, (Figure 2a) resulted from an opening of a food, societal, and ceremonial fishery for the First Nations fishers downstream from the site. This fishery began on August 20 and ended on August 22. The pattern of data from August 19, 2010, (Figure 2d) corresponded to a 24-h opening of a commercial fishery on August 18. The pattern observed on August 21 (Figure 2e) corresponded to the openings of the First Nations food, societal, ceremonial, and economic opportunity fisheries from August 19 to August 22.

Precision of the Three Sampling Methods

Of the three methods, simple random sampling was consistently and significantly less precise than systematic hourly sampling for the six full-count data ($P = 0.003$), whereas the latter was similarly precise to hourly random sampling ($P = 0.340$).
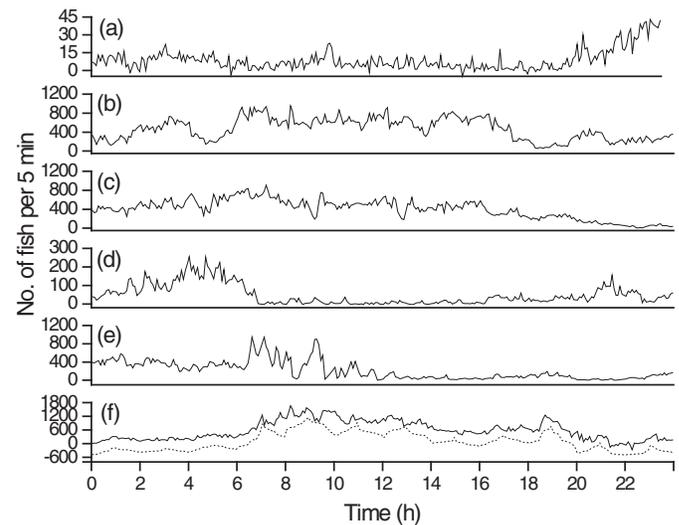
FIGURE 2. Time series plots (solid lines) of daily fish-passage data with each data point representing a 5-min fish count from the DIDSON sonar image data. The data were acquired from the left bank on (**a**) August 22, 2004, and in 2010 on (**b**) August 17, (**c**) August 18, (**d**) August 19, (**e**) August 21, and (**f**) September 25. The dotted line, offset by 400 to avoid overlapping with the solid line, is a reconstructed time series from the September 25 full-count data (see texts for descriptions).

TABLE 2. Key statistics of the six full-count data (Figure 2) and CV of estimated means by the three sampling methods. The two listed *P* values are the results of the *t*-tests on the CV differences with respect to the CV of systematic hourly sampling. Dates were in 2004 (Aug 22) and 2010 (Aug 17–19 and 21 and Sep 25).

	Date						
Statistic	Aug 22	Aug 17	Aug 18	Aug 19	Aug 21	Sep 25	*P*
Mean 5-min fish count	8.7	468.4	400.4	48.1	210.7	584.3	
Within-hour variance S_{wst}^2	20.5	9,409.6	7,081.2	561.3	12,778.0	20,343.9	
Total variance S^2	78.8	51,689.2	40,902.6	3,058.7	40,677.8	175,490.7	
Intrahour correlation ρ	−0.0096	0.0239	0.0140	−0.0242	−0.01831	0.0150	
24-h total fish count	2,517	134,892	115,319	13,857	60,676	168,278	
CV (simple random) (%)	19.5	9.5	9.9	22.4	18.7	14.0	0.003
CV (hourly random) (%)	10.3	4.1	4.1	9.6	10.5	4.8	0.340
CV (hourly systematic) (%)	9.2	5.0	4.7	6.4	8.0	5.5	

Because of the negative intrahour correlations in the time series of data for August 19, 21, and 22, systematic hourly sampling yielded slightly higher precisions than hourly random sampling. Key statistics of the time series and the precision performances by the three sampling methods are listed in Table 2.

Precision versus Sampling Fraction for Systematic Hourly Sampling

The numerical responses showed that the precision of systematic hourly sampling did not improve monotonically as the sampling fraction increased for the migration scenarios on August 17 and 21, whereby increasing the hourly sampling time from 15 to 20 min actually inflated the CV (see the bold italic values in Table 3). This counterintuitive outcome is a result of the less intuitive behavior of the intrahour correlation ρ. The most pronounced effect of ρ on the precision was observed from the dramatic shrinking of the CV to 1.7% at $f = 1/3$ for the data from

August 19. A detailed examination of this outcome revealed that the hourly phase structure of the data had an intrahour coefficient of −0.0418, which was very close to the critical value of −0.0435. The response of CV to f is displayed in Figure 3. The numerical values of estimated relation and observed variability are listed in Table 3. The responses of CV, not estimated directly from the presented data, can be interpolated from the tabulated values at sampling fractions of 1/10 (or 6 min/h) and 1/5 (or 12 min/h).

Highly Precise Estimates of Systematic Hourly Sampling for Time Series with Negative Intrahour Correlation Approaching ρ_c.

We reordered the 12 counts within each hour for the September 25 full-count data such that counts ascended in even hours

TABLE 3. The empirical relation (the mean CV) between CV and sampling fraction for systematic hourly sampling of the full-count data. The values in bold italic text correspond to scenarios in which an increased sampling fraction did not reduce the CV (relative to the CV to the left in the shaded area). Dates were in 2004 (Aug 22) and 2010 (Aug 17–19 and 21 and Sep 25).

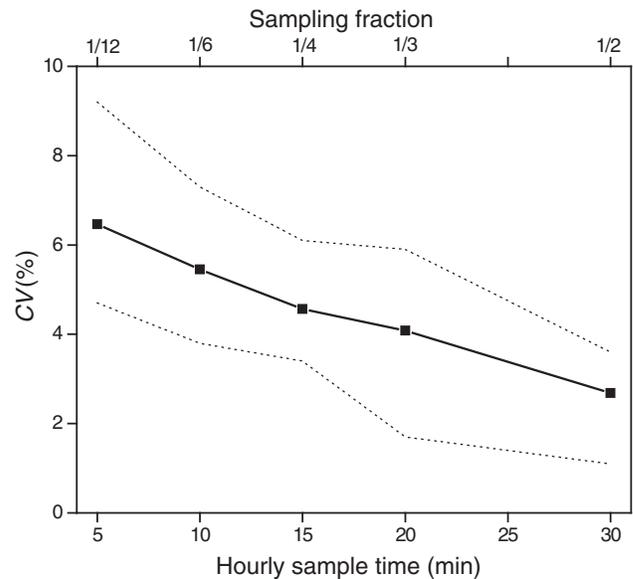|	Sampling fraction (upper row) and sampling time (lower row)				
CV statistic	1/12 5 min	1/6 10 min	1/4 15 min	1/3 20 min	1/2 30 min
CV (%), Aug 22	9.2	7.3	6.1	5.7	3.1
CV (%), Aug 17	5.0	4.6	3.4	*3.7*	1.1
CV (%), Aug 18	4.7	3.8	*3.8*	3.4	*3.6*
CV (%), Aug 19	6.4	5.6	4.2	1.7	*2.5*
CV (%), Aug 21	8.0	6.6	5.3	*5.9*	2.9
CV (%), Sep 25	5.5	4.8	4.6	4.1	2.9
Minimum CV (%)	4.7	3.8	3.4	1.7	1.1
Maximum CV (%)	9.2	7.3	6.1	5.9	3.6
Mean CV (%)	6.5	5.5	4.6	4.1	2.7

FIGURE 3. The empirical relationship between CV and sampling fraction f for systematic hourly sampling (the solid line), which was derived by the mean CVs for the migration scenarios represented by the six full-count data. The dashed lines are the maximum and minimum CVs.

TABLE 4. Bias of the three variance estimators for the systematic samples of the full-count data at $f = 1/6$ or 10 min/h. The listed P-values are the results of the t-tests on RE_B with the alternative hypothesis of $RE_B > 0$. Dates were in 2004 (Aug 22) and 2010 (Aug 17–19 and 21 and Sep 25).

Statistic	Date						P
	Aug 22	Aug 17	Aug 18	Aug 19	Aug 21	Sep 25	
V_T	1.6	1,872.7	914.0	29.5	773.2	3,170.0	
$\bar{\hat{V}}_4$	6.5	28,353.7	20,718.2	607.9	8,904.2	54,987.2	
$\bar{\hat{V}}_2$	3.0	2,982.2	1,062.5	123.8	2,195.9	4,929.6	
$\bar{\hat{V}}_5$	2.2	1,595.1	791.0	71.5	3,026.7	3,458.7	
RE_4	3.00	14.14	21.67	19.63	10.52	16.35	0.002
RE_2	0.87	0.59	0.16	3.20	1.84	0.56	0.024
RE_5	0.38	−0.15	−0.13	1.43	2.91	0.09	0.093

and descended in odd hours (the dotted line in Figure 2f). The alternation of the time series did not affect the precisions of simple random (CV = 14%) and hourly random sampling (CV = 4.8%) but produced an intrahour correlation coefficient of −0.0314, which was much closer to the ρ_c of −0.0435 than the ρ value of 0.015 of the original data (Table 2). As a result of this close proximity, the CV of systematic hourly sampling decreased from 5.5% to 2.6%. This example demonstrates that for the two similar daily fish-passage profiles with identical total and within-hour variances and a cross-correlation coefficient of 0.9388, the precision of systematic hourly sampling can vary greatly due to the difference in the hourly phase structures.

Bias Assessment of the Variance Estimators

The three estimators differed greatly in their bias and $\hat{V}_2$ performed the worst followed by $\hat{V}_4$, whereas $\hat{V}_5$ produced the least biased estimates. Both $\hat{V}_2$ and $\hat{V}_4$ significantly overestimated the variance. These results are, in general, consistent with the assessments of Reynolds et al. (2007) who used simulated full-count data to quantify the bias of the three estimators for the same sampling fraction of 1/6. The evaluated relative errors for each estimator at a sampling fraction of 1/6 (10 min/h) for systematic fish counting are provided in Table 4. Our analysis using the real data showed that $\hat{V}_5$ could either overestimate or

underestimate the variance. To examine how close a simulation model could replicate the real data for the true mean and the true precision by systematic hourly sampling, we used the Reynolds et al. (2007) model to simulate the processes represented by the six full-count data by linearly spreading the difference between adjacent counts from the same systematic sample S_1 to interpolate the unsampled counts. We found that the simulated full-count data statistically replicated the true means ($P = 0.691$) but significantly underestimated the variances of the systematic sampling ($P < 10^{-4}$) resulting in unrealistically small CVs that were on average only 25% of the true CVs.

Bias in Precision Estimation from a Single Systematic Sample

The differences between the true mean and sample mean $\hat{Y}_{S_1}$ were within the 95% CI for the six data sets so $\hat{Y}_{S_1}$ statistically replicated the true mean. Therefore, the deviation of $\hat{V}_B\text{-}CV_{S_1}$ from the true CV was caused by the bias of $\hat{V}_{S_1}$. Of the three estimators, $\hat{V}_4\text{-}CV_{S_1}$ produced the most biased estimates and both $\hat{V}_4\text{-}CV_{S_1}$ and $\hat{V}_2\text{-}CV_{S_1}$ significantly overestimated the CV. While $\hat{V}_5\text{-}CV_{S_1}$ produced the least biased estimates, in comparison with the empirical estimator, the latter outperformed the former with a mean relative error of 6% versus 36%. The estimated CVs for each estimator at $f = 1/6$ are listed in Table 5.

TABLE 5. Comparison of the performance for the estimated CV by a single systematic sample of the full-count data against the empirical CV (see Table 3) at $f = 1/6$ (or 10 min/h). The listed P-values are the results of the t-tests for the difference between $\hat{V}_B\text{-}CV_{S_1}$ and the true CV with the alternative hypothesis of $\hat{V}_B\text{-}CV_{S_1} > $ True CV. Dates were in 2004 (Aug 22) and 2010 (Aug 17–19 and 21 and Sep 25).

CV statistic	Date						P
	Aug 22	Aug 17	Aug 18	Aug 19	Aug 21	Sep 25	
True CV (%)	7.3	4.6	3.8	5.6	6.6	4.8	
$\hat{V}_4\text{-}CV_{S_1}$ (%)	16.0	17.8	18.0	27.3	23.2	19.8	<0.001
$\hat{V}_2\text{-}CV_{S_1}$ (%)	11.9	5.5	3.7	12.9	10.6	5.5	0.028
$\hat{V}_5\text{-}CV_{S_1}$ (%)	13.3	3.4	2.3	11.3	12.9	5.1	0.078
Empirical CV (%)	5.5	5.5	5.5	5.5	5.5	5.5	

DISCUSSION

We present an alternative approach to quantify the precision of fish counts by systematic sampling. This approach, though empirical, is less biased than using the variance estimators or simulated data. The full-count data used in this study provided a rare opportunity to appreciate, with real data scenarios, the bias in estimated precisions for systematic sampling with commonly used variance estimators and simulation approaches. We also demonstrated the less predictable impact of intrahour correlation coefficient ρ on the precision, which has received much less attention in the study of precisions of systematic sampling than has sampling effort. The revelation of the relationship between ρ and the precision from the real data scenarios provided us with a further insight and understanding on the performance of systematic sampling. For fish counting with multiple sonar systems, we have provided a numerical example in the Appendix to demonstrate the application of this method in estimating the uncertainty of total salmon escapement enumerated by an array of sonar counters. Although the empirical relation between the precision and the sampling fraction presented in this paper was established with hydroacoustic fish counts for a specific site on the lower Fraser River, this empirical approach is applicable to assessment of precisions of systematic hourly counting of fish passage in other rivers or for other types of data. We hereby summarize a few key features of systematic hourly sampling from this study.

1. The precision of the estimate does not improve monotonically with sampling fraction due to the less predictable behavior of intrahour correlation ρ.
2. Although hourly random sampling is statistically more robust than systematic hourly sampling, the latter yielded statistically similar precisions to the former in all the data scenarios we analyzed.
3. The hourly sampling time of 10 min (corresponding to a sampling fraction of 1/6) achieved, on average, a practically useful precision with a CV of 5.5%.
4. Doubling the hourly sampling time from 10 to 20 min reduced the CV by 25% as the CV decreased from 5.5% to 4.1% (or a decrease of PCL_{95} from 11% to 8%). This doubling of the sampling fraction did not reduce the CV by one-half as predicted by Lilja et al. (2008) using simulation approaches.
5. Systematic hourly sampling achieved highly precise estimates for time series of fish-passage data with a negative coefficient of intrahour correlation due to hourly alternating trends. In general, if a process is known to statistically alternate its trend over a temporal or spatial scale, then stratifications of the process with this intrastrata trending scale can yield a highly precise estimate by systematic sampling. The relationship between the intrastrata correlation ρ and the precision can serve as a guide in the development of a sampling design in which the objective is to find the number of stratifications, n, and the number of systematic samples, k, to minimize the distance between ρ and ρ_c.

While the focus of this study was to examine the precision of systematic hourly sampling with the full-count data, the data also allowed for the study of the precision of systematic subhourly sampling with sampling rates (frequencies) higher than one sampling unit per hour. Sampling rate is defined for the subsequent discussions as the number of sampling units per hour. Choosing 5-min fish counts as basic sampling units, we can perform systematic sampling of the full-count time series at a sampling rate of 2 sampling units/h by stratifying the 24-h time series data into 48 strata of 0.5-h duration ($n = 48$) to acquire six systematic samples ($k = 6$). With this semihourly sampling of 5-min counts we obtain not only a mean CV of 3.3%, which is lower than 5.5% ($P = 0.004$) obtained with the hourly sampling of 10-min counts, but also consistently lower CVs for all the individual full-count data (Table 6). The favorable precision response to sampling frequency is consistent with the Nyquist–Shannon sampling theorem (Oppenheim and Schafer 2010), which states that subsampled data can completely replicate an underlying process comprised of frequency components not exceeding one-half of the sampling frequency. While fish-passage time series (subject to no anti-aliasing filtering) may contain higher frequency components than can be replicated with the hourly sampled data, by increasing the sampling rate beyond 1 sampling unit/h the data can capture an increased amount of information with subhourly variations, thus better replicating the time series. But increasing the sampling rate causes a decrease of the time interval between adjacent sampling units. We denote this time interval as nonsampling time (NST), which is usually allocated to counting fish passages in other passage areas of the river with the same counter. This is a common fish-counting practice in a large riverine or marine space with a limited number of counters (Maxwell and Gove

TABLE 6. Comparisons of precision gain by the two sampling approaches: lengthening the sampling time versus increasing the sampling rate at three sampling fractions. The shaded values are the sampling parameters and the results by the approach of increasing the sampling rate beyond 1 sampling unit/h. Dates were in 2004 (Aug 22) and 2010 (Aug 17–19 and 21 and Sep 25).

| Parameter | Sampling fraction | | | | | |
	$f = 1/6$		$f = 1/4$		$f = 1/3$	
Sampling rate (units/h)	1	2	1	3	1	4
Sampling time (min)	10	5	15	5	20	5
NST (min)	50	25	45	15	40	10
CV (%), Aug 22	7.3	3.1	6.1	5.0	5.7	0.9
CV (%), Aug 17	4.6	1.8	3.4	2.8	3.7	1.0
CV (%), Aug 18	3.8	2.9	3.8	1.2	3.4	1.9
CV (%), Aug 19	5.6	4.1	4.2	2.0	1.7	2.7
CV (%), Aug 21	6.6	5.2	5.3	4.3	5.9	0.2
CV (%), Sep 25	4.8	2.7	4.6	1.2	4.1	1.6
Minimum CV (%)	3.8	1.8	3.4	1.2	1.7	0.2
Maximum CV (%)	7.3	5.2	6.1	5.0	5.9	2.7
Mean CV (%)	5.5	3.3	4.6	2.8	4.1	1.4

2004; Enzenhofer et al. 2010; Xie et al. 2012b). In summary, NST is a key variable in practical sampling designs at fixed sampling fractions as it must balance two competing demands for (1) an increased capacity of multistrata spatial sampling using a single counter during a prolonged NST, and (2) an improved representation of the true process by data samples with a shortened NST to achieve the smallest resolvable time scale of $2 \times$ NST defined by Nyquist–Shannon sampling theorem.

To compare precision gain by the two sampling approaches (lengthening the sampling time of each count versus increasing the sampling rate beyond 1 sampling unit/ h), we enlist in Table 6 the numerical responses of CV for all the full-count data at $f = 1/6$, $1/4$, and $1/3$, respectively. It is evident that both approaches can meet demand (2), but increasing the sampling rate is more efficient than lengthening the sampling time. On the other hand, the hourly sampling approach provides a greater flexibility to meet demand (1). For applications where a counter is delegated solely to counting fish passage in a fixed area, increasing the sampling rate is a more effective choice than lengthening the sampling time for precision gain under the same sampling fractions. For instance, upon increasing the sampling rate to 4 sampling units/h (quadrupling the hourly rate), we can yield extremely precise estimates for the full-count data in comparison with the corresponding hourly sampling estimates. For applications where a counter is required to sample multiple spatial strata, hourly sampling remains a practically favorable approach for multistrata sampling of fish passage with a greater degree of freedom while still allowing the users to lengthen the sampling time (shortening the NST) for an improved resolution of the sampled data. Ultimately, it is the specified precision and the associated cost for a survey that determine the optimal sampling design.

ACKNOWLEDGMENTS

REFERENCES

Banneheka, S. G., R. D. Routledge, I. C. Guthrie, and J. C. Woodey. 1995. Estimation of in-river fish passage using a combination of transect and stationary hydroacoustic sampling. Canadian Journal of Fisheries and Aquatic Sciences 52:335–343.

Becker, C. D. 1962. Estimating red salmon escapements by sample counts from observation towers. U.S. Fish and Wildlife Service Fishery Bulletin 192(61):355–369.

Belcher, E. O., W. Hanot, and J. Burch. 2002. Dual-frequency identification sonar. Pages 187–192 in Marine Technology Society/Institute of Electrical and Electronics Engineers (MTS/IEEE). Underwater technology 2002. Proceedings of the 2002 international symposium on underwater technology. MTS/IEEE, New York.

Bromaghin, J. F., and D. A. Bruden. 1999. A simulation study of temporal sampling designs for salmon counting towers. Alaska Department of Fish and Game, Regional Information Report 3A99-35, Anchorage.

Chilton, D. E., and R. J. Beamish. 1982. Age-determination methods for fishes studied by the groundfish program at the Pacific Biological Station. Canadian Special Publication of Fisheries and Aquatic Sciences 60:10–12.

Cochran, W. G. 1977. Sampling techniques, 3rd edition. Wiley, New York.

Enzenhofer, H. J., G. M. W. Cronkite, and J. A. Holmes. 2010. Application of DIDSON imaging sonar at Qualark Creek on the Fraser River for enumeration of adult pacific salmon: an operation manual. Canadian Technical Report of Fisheries and Aquatic Sciences 2869.

Enzenhofer, H. J., N. Olsen, and T. J. Mulligan. 1998. Fixed-location riverine hydroacoustics as a method of enumerating migrating adult Pacific salmon: comparison of split-beam acoustics vs. visual counting. Aquatic Living Resources 11:61–74.

Holmes, J. A., G. M. W. Cronkite, H. J. Enzenhofer, and T. J. Mulligan. 2006. Accuracy and precision of fish-count data from a "dual-frequency identification sonar" (DIDSON) imaging system. ICES Journal of Marine Science 63:543–555.

Levy, D. A., and A. D. Cadenhead. 1995. Selective tidal stream transport of adult Sockeye Salmon (Oncorhynchus nerka) in the Fraser River Estuary. Canadian Journal of Fisheries and Aquatic Sciences 52:1–12.

Lilja, J., T. Ridley, G. M. W. Cronkite, H. J. Enzenhofer, and J. A. Holmes. 2008. Optimizing sampling effort within a systematic design for estimating abundant escapement of Sockeye Salmon (Oncorhynchus nerka). Fisheries Research 90:118–127.

Maxwell, S. L., and N. E. Gove. 2004. The feasibility of estimating migrating salmon passage rates in turbid rivers using a dual-frequency identification sonar (DIDSON) 2002. Alaska Department of Fish and Game, Regional Information Report 2A04-05, Anchorage.

Maxwell, S. L., and N. E. Gove. 2007. Assessing a dual-frequency identification sonars' fish-counting accuracy, precision, and turbid river range capability. Journal of the Acoustical Society of America 122:3364–3377.

Oppenheim, A. V., and R. W. Schafer. 2010. Digital signal processing, 3rd edition. Pearson Higher Education, Upper Saddle River, New Jersey.

R Foundation for Statistical Computing. 2012. R version 2.15.2. Available: http://www.r-project.org./ (October 2012).

Reynolds, J. H., C. A. Woody, N. E. Gove, and L. F. Fair, 2007. Efficiently estimating salmon escapement uncertainty using systematically sampled data. Pages 121–129 in C. A. Woody, editor. Sockeye Salmon ecology, evolution, and management. American Fisheries Society, Symposium 54, Bethesda, Maryland.

Skalski, J. R., A. Hoffmann, B. H. Random, and T. W. Steig. 1993. Fix-location hydroacoustic monitoring designs for estimating fish passage using stratified random and systematic sampling. Canadian Journal of Fisheries and Aquatic Sciences 50:1208–1221.

Steig, T. W., G. A. Raemhild, and B. H. Ransom. 1988. Hydroacoustic evaluation of juvenile salmon and steelhead approaching a submersible traveling screen at Rocky Reach Dam in 1988. Chelan County, Public Utility District 1, Wenatchee, Washington.

Wolter, K. M. 1984. An investigation of some estimators of variance for systematic sampling. Journal of the American Statistical Association 79:781–790.

Wolter, K. M. 2007. Introduction to variance estimation, 2nd edition. Springer-Verlag, New York.

Xie, Y., A. P. Gray, F. J. Martens, J. L. Boffey, and J. D. Cave. 2005. Use of dual-frequency identification sonar to verify split-beam estimates of salmon flux and to examine fish behaviour in the Fraser River. Pacific Salmon Commission Technical Report 16.

Xie, Y., F. J. Martens, and J. L. Nelitz. 2012a. Implementation of stationary sub-sampling systems to estimate salmon passage in the lower Fraser River: year 1 of 2011 and 2012 project report to southern boundary restoration and enhancement fund. Pacific Salmon Commission, Vancouver.

Xie, Y., C. G. Michielsens, and F. J. Martens. 2012b. Classification of fish and non-fish acoustic tracks using discriminant function analysis. ICES Journal of Marine Science 69:313–322.

Xie, Y., T. J. Mulligan, G. M. W. Cronkite, and A. P. Gray. 2002. Assessment of potential bias in hydroacoustic estimation of Fraser River Sockeye and Pink salmon at Mission, B.C. Pacific Salmon Commission Technical Report 11.

Appendix: Precision Estimate for Subsampled Fish Counts from Multiple Counters

Assuming the total fish passage A is estimated by the sum of spatially nonoverlapping estimates A_i (where $i = 1, 2, \ldots, M$), we can express A as

$$A = \sum_{i=1}^{M} A_i \tag{A.1}$$

and the variance of A, σ_A^2, as

$$\sigma_A^2 = \sum_{i=1}^{M} \sigma_i^2 + 2 \sum_{i \neq j} \rho_{i,j} \sigma_i \sigma_j, \tag{A.2}$$

where σ_i^2 is the variance of A_i and $\rho_{i,j}$ is the correlation coefficient between A_i and A_j. The CV of A can be expressed as

$$CV_A = \frac{\sigma_A}{A} = \frac{\sqrt{\sum_{i=1}^{M} \sigma_i^2 + 2 \sum_{i \neq j} \rho_{i,j} \sigma_i \sigma_j}}{\sum_{i=1}^{M} A_i}. \tag{A.3}$$

Assuming that all the CVs for A_i ($i = 1, 2, \ldots, M$) are known (denoted as CV_i), equation (A.3) can be rewritten in terms of CV_i as

$$CV_A = \frac{\sqrt{\sum_{i=1}^{M} A_i^2 \cdot CV_i^2 + 2 \sum_{i \neq j} (\rho_{i,j} \cdot A_i \cdot A_j CV_i \cdot CV_j)}}{\sum_{i=1}^{M} A_i}. \tag{A.4}$$

In general, $\rho_{i,j}$ can vary from -1 to $+1$. We do not consider scenarios of negative correlations here but restrict the discussion of equation (A.4) for two limiting cases where $\rho_{i,j} = 0$ or 1 for all i and j from 1 to M. That is, A_i and A_j are either uncorrelated or linearly correlated (in phase).

Case 1: $\rho_{i,j} = 0$ for all i and j from 1 to M.

In this case, equation (A.4) is reduced to

$$CV_A|_{\rho=0} = \frac{\sqrt{\sum_{i=1}^{M} A_i^2 \cdot CV_i^2}}{\sum_{i=1}^{M} A_i} \tag{A.5}$$

Case 2: $\rho_{i,j} = 1$ for all i and j from 1 to M.

Substituting $\rho_{i,j} = 1$ into equation (A.4) and recognizing that $\sum_{i=1}^{M} A_i^2 \cdot CV_i^2 + 2 \sum_{i \neq j} (A_i \cdot A_j \cdot CV_i \cdot CV_j)$ is a polynomial expansion of $(A_1 \cdot CV_1 + A_2 \cdot CV_2 + \ldots + A_M \cdot CV_M)^2$, equation (A.4) is reduced to

$$CV_A|_{\rho=1} = \frac{\sum_{i=1}^{M} A_i \cdot CV_i}{\sum_{i=1}^{M} A_i}. \tag{A.6}$$

It is evident that $CV_A|_{\rho=0} \leq CV_A|_{\rho=1}$ and for all other positive values of $\rho_{i,j}$, CV_A is bounded between equations (A.5) and (A.6). Therefore, we can define a precision interval for A as

$$\frac{\sqrt{\sum_{i=1}^{M} A_i^2 \cdot CV_i^2}}{\sum_{i=1}^{M} A_i} \leq CV_A \leq \frac{\sum_{i=1}^{M} A_i \cdot CV_i}{\sum_{i=1}^{M} A_i} \tag{A.7}$$

The estimator, equation (A.4), provides an accurate measure of precision if accurate information on all $\rho_{i,j}$ is available. If $\rho_{i,j}$ are unknown or their values are uncertain then equation (A.7) is a practical estimator for the precision interval of the total estimate derived from a sum of individual estimates from multiple fish counters or multiple sounding ranges and aims. Precision interval estimator, equation (A.7), can also be applied to the estimation of the precision interval of a seasonal total of fish abundance if treating A_i in equation (A.7) as daily abundance estimates.

A Numerical Example

To demonstrate the practical use of equation (A.7), we present a numerical example for utilizing this equation to estimate the precision of estimated daily total salmon passage for September 12, 2010, when the migration peaked as a large number of late-run Sockeye Salmon passed the counting site. A total of three sonar systems were used to estimate the total passage. The three systems consisted of a split-beam system on the left bank (denoted as L), a split-beam system on a transecting vessel (T), and a DIDSON system on the right bank (R). The two shore-based systems used hourly systematic sampling to count fish at a sampling fraction of 1/10 (6 min/h) for L and 1/6 (10 min/h) for R. The L system sampled the fish passage within 70 m from the left bank. The mobile system T sampled the offshore passage between 70 and 320 m from the left bank. The R system sampled the near-shore passage within 30 m from the right bank. Xie et al. (2012) provided a detailed description of the sampling designs. The L, T, and R estimated a daily total of 491,042, 41,073, and 58,042 salmon, respectively, for a daily total salmon passage of 590,157, which were mainly the late-run Sockeye Salmon heading for the Adams River in the interior of British Columbia. The CVs for L and R are estimated from the empirical relationship (Table 3) as 6.3% and 5.5%, respectively. The CV for T was estimated by Banneheka et al. (1995) as 4%. Substituting the three estimates and their corresponding CVs into equation (A.7), we obtained a CV interval of 5.3% $\leq$

$CV_A \leq 6.1\%$, which adequately quantifies the precision of the estimate of total salmon escapement for this date.

APPENDIX REFERENCES

Banneheka, S. G., R. D. Routledge, I. C. Guthrie, and J. C. Woodey. 1995. Estimation of in-river fish passage using a combination of transect and stationary hydroacoustic sampling. Canadian Journal of Fisheries and Aquatic Sciences 52:335–343.

Xie, Y., F. J. Martens, and J. L. Nelitz. 2012. Implementation of stationary sub-sampling systems to estimate salmon passage in the lower Fraser River: year 1 of 2011 and 2012 project report to southern boundary restoration and enhancement fund. Pacific Salmon Commission, Vancouver.