American Fisheries Society Symposium 6:181–189, 1989


# Model-Based Sampling Methods for Effort and Catch Estimation

JOHN M. HOENIG[1]

*Rosenstiel School of Marine and Atmospheric Science, University of Miami*
*4600 Rickenbacker Causeway, Miami, Florida 33149, USA*

FRANK B. MARTIN

*University of Minnesota, Department of Applied Statistics, St. Paul, Minnesota 55108, USA*

C. MARK HEYWOOD

*Minnesota Department of Natural Resources*
*Post Office Box 6247, Rochester, Minnesota 55903, USA*

*Abstract.*—Model-based sampling procedures are described in which a regression equation is developed to model trends in observed counts of fishing boats. Total fishing effort is estimated by integrating the area under the regression curve. Variance of the estimated fishing effort is computed by the delta (Taylor's-series) method and use of the covariance matrix of the regression coefficients. Model-based sampling has three possible advantages over classical sampling: it can provide a detailed description of variation in boat counts as a function of explanatory variables; it can allow more convenient sampling schedules; and it can provide greater precision if there are strong trends in the counts with respect to an explanatory variable such as time. The same approach can be used to estimate other quantities that vary systematically with some auxiliary variable, such as catch that varies over time or space.

Designing a fishery sampling program is difficult when the variable of interest varies systematically over time or space (Barnard et al. 1985). For example, if the sampling unit consists of a (day) × (location) combination, a randomly selected sampling plan might, by chance, call for several widely separated areas to be sampled simultaneously or in near succession. If followed literally, this sampling plan would require several people to conduct the sampling.

Rather than focusing on the difficulties presented for classical sampling techniques by temporal and spatial trends in data, we describe here a model-based approach to sampling whereby the trends in the data are modeled by regression techniques. This approach has three advantages: (1) it provides a detailed description of the fishery in relation to available explanatory variables; (2) it allows convenient sampling schedules to be established; and (3) it can provide efficient estimates by using information from auxiliary variables. A disadvantage of the method is that the design and analysis of the survey require more thought. Consequently, these aspects of the survey become less routine.

[1]Present address: Department of Fisheries and Oceans, Post Office Box 5667, St John's, Newfoundland A1C 5X1, Canada.

We first describe a general approach to model-based sampling and illustrate it with a simple hypothetical example. For more background on this subject, the reader is referred to Cassel et al. (1977). Next, we apply the approach to an effort survey based on aerial counts of boats in a sport fishery. This survey was originally designed with two-stage sampling within stratified random sampling. Consequently, the model-based approach can be compared directly to results from classical sampling. Finally, we discuss the implementation of the model-based approach to other sampling problems.

## Model-Based Approach to Sampling

Suppose the daily landings of a particular species in a port follow a distinct seasonal trend. Then the observations of daily landed catch, $C_i$, can be modeled by a polynomial regression,

$$C_i = b_0 + b_1 t_i + b_2 t_i^2 + \ldots + b_k t_i^k + e_i, \qquad (1)$$

or by a regression of some other suitable form, in which $t_i$ is the time when the $i$th observation is made, the $b$ coefficients are estimated by regression, and $e_i$ is the residual for the $i$th observation.

The total landings over any portion of the season, such as the interval $(t_1,t_2)$, can be estimated by integrating the regression equation

$$\hat{C}(t_1,t_2) = \int_{t_1-1/2}^{t_2+1/2} (b_0 + b_1 t + b_2 t^2 + \ldots + b_k t^k)dt. \quad (2)$$

The fractions in the limits are a consequence of approximating a discrete variable (day) with a continuous one.

The approximate bias and variance of $\hat{C}(t_1,t_2)$ can be found by the "delta" or Taylor's-series method (Kendall and Stuart 1977; Seber 1982): for a function $f$ of the random variable $\vec{X} = \{X_1, X_2, \ldots, X_k\}$,

$$\text{bias } [f(\vec{X})] \doteq \sum_{i=1}^{k} {}^1\!/_2 V(X_i) \frac{\partial^2 f}{\partial X_i^2} + \sum\sum_{i<j} \text{Cov } (X_i,X_j) \frac{\partial^2 f}{\partial X_i\, \partial X_j}; \quad (3)$$

$$V[f(\vec{X})] \doteq \sum_{i=1}^{k} V(X_i) \left(\frac{\partial f}{\partial X_i}\right)^2 + 2 \sum\sum_{i<j} \text{Cov } (X_i,X_j) \frac{\partial f}{\partial X_i} \frac{\partial f}{\partial X_j}. \quad (4)$$

Estimates of bias and variance of $f(\vec{X})$ are obtained by substituting sample estimates for the population variances and covariances in equations (3) and (4). Cramér (1946) established the asymptotic validity of the delta method for functions of sample moments such as regression coefficients.

It should be noted that, whereas the variable of integration in our example was time ($t$), the statistical variables in the bias and variance formulae are the regression coefficients, $b_i$, and the time limits are treated as known constants. For example, if the catches in Figure 1 are modeled by a quadratic equation, the estimated total catch would be given by

$$\hat{C}(t_1,t_2) = \int_{t_1}^{t_2} (b_0 + b_1 t + b_2 t^2)dt$$

$$= (b_0 t + b_1 t^2/2 + b_2 t^3/3)\Big|_{t_1}^{t_2}$$

$$= b_0(t_2 - t_1) + b_1(t_2^2 - t_1^2)/2$$

$$+ b_2 (t_2^3 - t_1^3)/3. \quad (5)$$

For notational simplicity, the fractional limits used in equation (2) are dropped until the aerial survey example is considered. The variance of

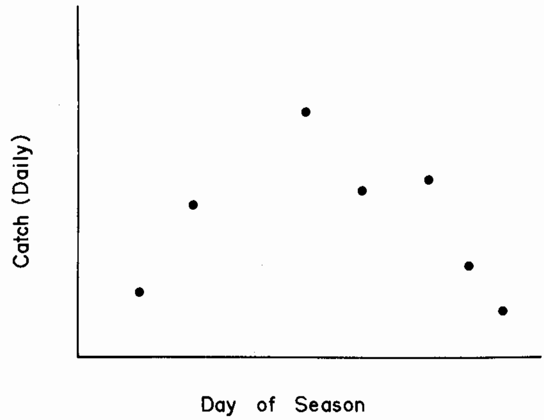


FIGURE 1.—Hypothetical example in which daily landings follow a dome shaped trend over time.

$\hat{C}(t_1,t_2)$ can be found from equation (4) as

$$\begin{aligned} \hat{V}[\hat{C}(t_1,t_2)] \doteq\ & (t_2 - t_1)^2\, \hat{V}(b_0) + (t_2^2 - t_1^2)^2 \hat{V}(b_1)/4 \\ & + (t_2^3 - t_1^3)^2 \hat{V}(b_2)/9 \\ & + (t_2 - t_1)(t_2^2 - t_1^2)\, \widehat{\text{Cov}}(b_0,b_1)/2 \\ & + (t_2 - t_1)(t_2^3 - t_1^3)\widehat{\text{Cov}}(b_0,b_2)/3 \\ & + (t_2^2 - t_1^2)(t_2^3 - t_1^3)\, \widehat{\text{Cov}}(b_1,b_2)/6. \end{aligned}$$

The delta method can also be used to approximate covariances. For two functions, $f$ and $g$, of the random variable $\vec{X} = \{X_1, X_2, \ldots, X_k\}$,

$$\text{Cov}[f(\vec{X}),g(\vec{X})] \doteq \sum_{i}^{k} \sum_{j}^{k} \text{Cov}(X_i,X_j) \frac{\partial f}{\partial X_i} \frac{\partial g}{\partial X_j}. \quad (6)$$

Again, to estimate the covariance of $f(\vec{X})$ and $g(\vec{X})$, the sample estimates of $\text{Cov}(X_i,X_j)$ are substituted for the population values in equation (6). In our example from Figure 1, if we were interested in the covariance of estimates made for two portions of the season, $t_1$ to $t_2$ and $t_3$ to $t_4$, we would have, from equation (6),

$$\begin{aligned} & \widehat{\text{Cov}}[\hat{C}(t_1,t_2),\hat{C}(t_3,t_4)] \\ & = \hat{V}(b_0)(t_2 - t_1)(t_4 - t_3) \\ & \quad + \hat{V}(b_1)(t_2^2 - t_1^2)(t_4^2 - t_3^2)/4 \\ & \quad + \hat{V}(b_2)(t_2^3 - t_1^3)(t_4^3 - t_3^3)/9 \\ & \quad + \widehat{\text{Cov}}(b_0,b_1)[(t_2 - t_1)(t_4^2 - t_3^2)/2 \\ & \qquad + (t_2^2 - t_1^2)(t_4 - t_3)/2] \end{aligned}$$

$$+ \widehat{\text{Cov}}(b_0,b_2)[(t_2 - t_1)(t_4^3 - t_3^3)/3 + (t_2^3 - t_1^3)(t_4 - t_3)/3]$$
$$+ \widehat{\text{Cov}}(b_1,b_2)[(t_2^2 - t_1^2)(t_4^3 - t_3^3)/6 + (t_2^3 - t_1^3)(t_4^2 - t_3^2)/6].$$

The computations are lengthy but not difficult.

### Comparison of Methods in an Aerial Survey of Effort

We now apply the model-based approach to estimation of fishing effort from a series of instantaneous boat counts made during aerial flights in 1984 over Lake Vermillion, a large (20,000-hectare) lake in Minnesota. Although the example deals with a sport fishery, the methodology is easily adapted to commercial fisheries. The structure and complexity of the problem make the example informative for those considering using a model-based approach.

The fishing season was divided into eight strata according to day type (weekday versus weekend day) and time of day (four periods corresponding to quarter fishing days). The sampling unit was a quarter day. Within a stratum, sampling units were selected by simple random sampling. As a consequence of sampling strata independently, it was possible to have a sampling schedule that required some days to be sampled more than once.

Each sampling unit (quarter day) was subsampled, in theory, by random selection of an "instant" in which to make a count of boats fishing. This count represents an unbiased estimate of the average number of boats fishing during the quarter day (based on a sample of one instant). Consequently, the product of the count times the duration of the quarter day provides an unbiased estimate of the effort, in boat-hours, during the quarter day (Robson 1961).

In practice, the aerial flights required about an hour to cover the whole lake. However, each location on the lake was only observed for an "instant." Flights followed one of three possible routes, selected randomly, and flight starting times were also varied to approximate random observations over time and space.

In setting up this sampling plan, days were randomly selected from a 19-week period. In addition, 11 extra flights were scheduled after the regular season was finished to check on fishing effort during the autumn. The latter data were not collected according to the above sampling scheme, so they cannot be combined with the regular season data if estimates are to be calculated according to stratified random sampling. They can be analyzed as part of the model-based procedure. However, to keep results for the two methods comparable, only data from the first 19 weeks are analyzed. In addition, only two observations were made during period 1 on weekdays, and no attempt is made to model this stratum.

#### *Classical Results*

The boat counts were multiplied by the length of a quarter day at the corresponding time of the season to convert them to fishing effort (Figure 2). The efforts were averaged within each stratum to obtain the mean effort and then multiplied by the number of days in the season to obtain the seasonal total (Table 1).

Within a stratum, an unbiased estimate of the variance of an estimated mean obtained by two-stage sampling is (Cochran 1977)

$$\hat{V}(\bar{\bar{y}}) = \frac{1 - f_1}{n} s_1^2 + \frac{f_1(1 - f_2)}{nm} s_2^2;$$

$\bar{\bar{y}}$ = estimated population mean;
$f_1$ = sampling fraction at the first stage (proportion of days sampled);
$f_2$ = sampling fraction at the second stage (proportion of "instants" sampled; zero in our example);
$n$ = number of first-stage units (days) sampled;
$m$ = number of second-stage units (instants) sampled (=1);
$s_1^2$ = variance at first stage, estimated by $\sum_i(\bar{y}_i - \bar{\bar{y}})^2/(n - 1)$, $\bar{y}_i$ being the sample mean from the $i$th primary sampling unit (day);
$s_2^2$ = variance at second stage estimated by $\sum_i \sum_j (y_{ij} - \bar{y}_i)^2/[n(m - 1)]$, $y_{ij}$ being the $j$th observation from the $i$th primary sampling unit (day).

Because only a single observation was taken at the second stage (i.e., one observation per quarter day sampled), the variance cannot be estimated in this manner. However, when $f_1$ is negligible, a close approximation is

$$\hat{V}(\bar{\bar{y}}) \doteq s_1^2/n; \quad (7)$$

that is, the variance of the mean effort is approximately the sample variance of the mean computed from all the estimated efforts within the stratum. If $f_1$ is not negligible, then equation (7) overes-

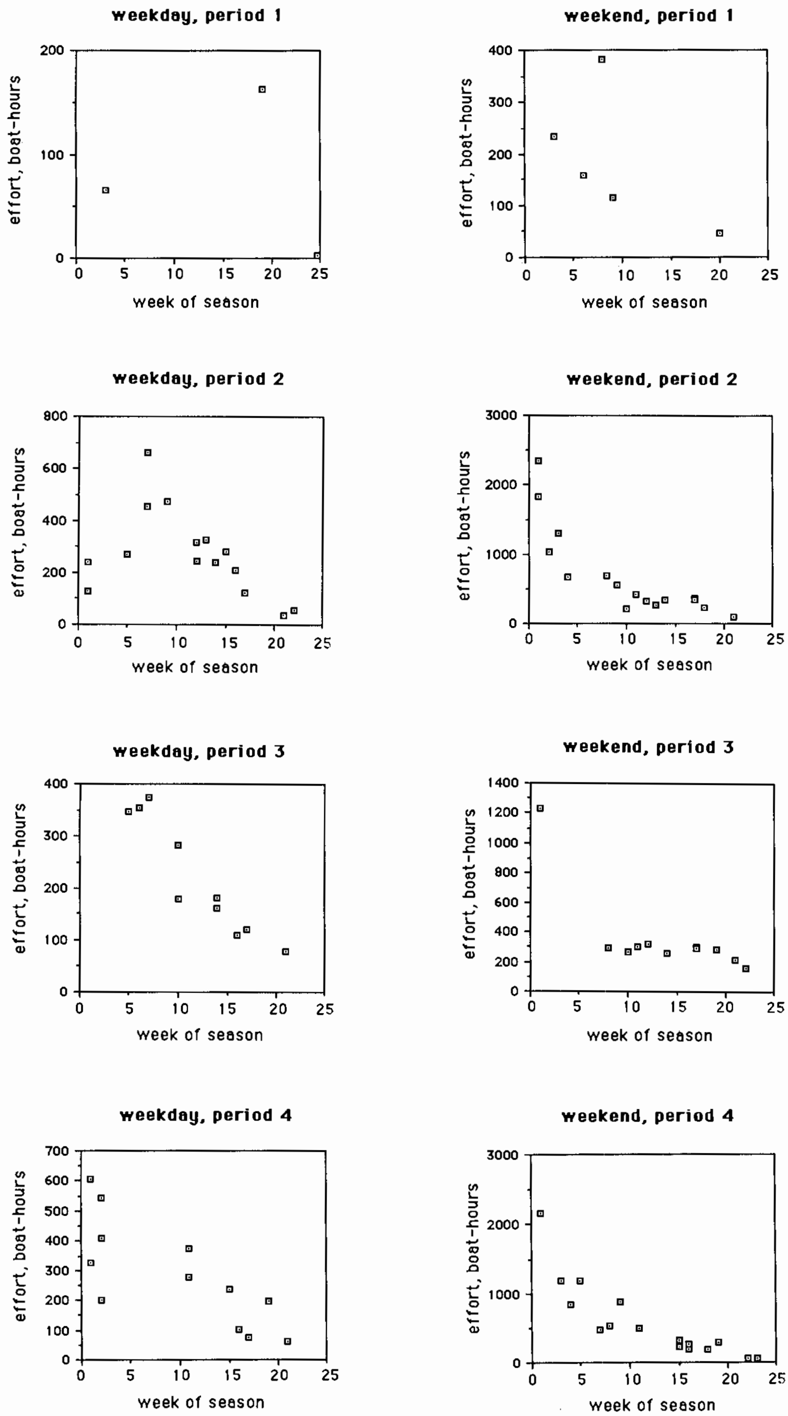


FIGURE 2.—Estimated fishing efforts on Lake Vermillion (in boat-hours per quarter fishing day) in eight time strata.

TABLE 1.—Aerial survey estimates of fishing effort on Lake Vermillion, Minnesota, based on a two-stage sampling design within stratified random sampling.

| Stratum | Day type[a] | Period[b] | Mean daily effort (boat-hours) | Days in season | Seasonal total (boat-hours) | Number of counts | 2 × CV (%)[c] |
|---|---|---|---|---|---|---|---|
| 1 | Weekday | 1 | 114 | 95 | 10,792 | 2 | 87 |
| 2 | Weekday | 2 | 305 | 95 | 29,022 | 13 | 27 |
| 3 | Weekday | 3 | 234 | 95 | 22,230 | 9 | 30 |
| 4 | Weekday | 4 | 304 | 95 | 28,928 | 11 | 34 |
| 5 | Weekend | 1 | 222 | 38 | 8,447 | 4 | 53 |
| 6 | Weekend | 2 | 730 | 38 | 27,732 | 15 | 45 |
| 7 | Weekend | 3 | 396 | 38 | 15,037 | 9 | 53 |
| 8 | Weekend | 4 | 662 | 38 | 25,171 | 14 | 45 |
| Total | | | | | 167,359 | 77 | 15 |

[a] Weekdays: Monday–Friday exclusive of holidays. Weekend days: Saturdays, Sundays, holidays.
[b] Quarter of the fishing day.
[c] Coefficient of variation: CV = 100(standard error)/estimate; 2 × CV is half the width of the 95% confidence interval about the estimated seasonal total, as a percentage of the estimated total.

timates by the amount $f_1\ S_1^2/n$, $S_1^2$ being the true value of the first-stage variance (Cochran 1977).

The variance of the total within a stratum (Table 1) was computed as

$$\hat{V}\text{ (total within stratum)} \\ = (\text{number days in stratum})^2 \\ \times (\text{variance of mean from equation 7}).$$

The variance of the total over all strata is simply the sum of the variances because strata are, by definition, independent.

### *Model-Based Results*

Fourteen explanatory variables were examined by regression analysis for their ability to account for variability in the boat counts during the first 19 weeks of the season. Seven of these were

- wk = week of the season;
- wksq = square of the week of the season;
- wkcub = cube of the week of the season;
- $D$ = indicator variable for day type ($D$ = 0 for weekday, 1 for weekend day);
- $P1$ = indicator variable for first period of the day ($P1$ = 1 for the first period; 0, otherwise);
- $P2$, $P3$ = indicator variables for second and third periods of the day.

The other variables were first-order interaction terms between week and day type, week and periods of the day, and day type and periods of the day. These variables can be combined in over 8,000 possible models, and it was not feasible to compute regressions for all possible subsets. Visual inspection of the data (Figure 2) suggested that time, day type, and period of the day might be important. Therefore, we computed all possible regressions after forcing the model to contain wk, wksq, wkcub, $D$, and the three indicators for period of the day ($P1$, $P2$, and $P3$). We also computed all possible subsets of the seven variables listed above; after removing nonsignificant variables from the list of seven, we then forced the remainder to be in all possible subsets with the rest of the available variables.

Under this procedure, the best model for effort was

$$\text{effort}_i = b_0 + b_1\,\text{wk}_i + b_2\,\text{wksq}_i + b_3D_i + b_4\,P1_i + b_5D \times \text{wk}_i + b_6\,P1 \times \text{wk}_i + e_i; \quad (8)$$

the $e_i$ are residual errors. The coefficient estimates are

$$b_0 = 620;$$
$$b_1 = -84;$$
$$b_2 = 3.8;$$
$$b_3 = 1{,}037;$$
$$b_4 = -1{,}311;$$
$$b_5 = -67;$$
$$b_6 = 104.$$

The $R^2$ for this model is 0.75. The covariance matrix for the coefficients is given in Table 2.

The estimates of the seasonal totals were obtained by integrating the regression equation over time for each combination of the indicator variables. Thus, for example, for weekend days during periods 2 through 4 combined,

$$\widehat{\text{effort}}\ (\text{weekend}, P2\text{–}P4)$$
$$= 2 \times 3 \int_{0.5}^{19.5} [b_0 + b_3 + (b_1 + b_5)t + b_2t^2]\, dt.$$

The factor of 2 arises because there are two weekend days per week; the factor of 3 is to

TABLE 2.—Estimated covariance matrix for the model-based sampling regression coefficients developed for boat counts on Lake Vermillion. The regression model was developed to explain variation in estimated fishing effort (per quarter fishing day) as a function of the time within the season (week and week$^2$), the day type, the time of the day (whether or not first quarter of the day), and two interaction terms.

| Coefficient | Constant (const) | Week (wk) | Day type (*D*) | First quarter of day (*P*1) | (*D* × wk) | (*P*1 × wk) | Week$^2$ (wksq) |
|---|---|---|---|---|---|---|---|
| const | 9,011 | | | | | | |
| wk | −1,401 | 373 | | | | | |
| *D* | −5,930 | 304.7 | 12,480 | | | | |
| *P*1 | −1,615 | 574.6 | −6,247 | 125,800 | | | |
| *D* × wk | 431.1 | −27.96 | −940.1 | 473.7 | 93.07 | | |
| *P*1 × wk | 382 | −135.9 | 513.8 | −16,940 | −50.93 | 2,601 | |
| wksq | 49.5 | −17.63 | 10.26 | −31.36 | −1.193 | 7.408 | 0.962 |

estimate the effort in three time periods of the day simultaneously. Variances were computed according to equation (4). The results are summarized in Table 3.

The model described by equation (8) has six coefficients, including two interaction terms. We therefore tried to find a simpler model that would be easier to explain to managers and other user groups. We fitted a model using just the a priori information that main effects for day type, period of the day, week, and square of the week would likely be important. (It is commonly believed by fisheries managers in Minnesota that fishing effort declines rapidly over time at the beginning of the season and then tapers off gradually as the season progresses.) This model had an $R^2$ value of 0.57. Since the probability values for the coefficients for period 2 and period 3 were each equal to or greater than 0.50, we simplified the model further by dropping the main effects for these periods. The resulting $R^2$ value was 0.56 (Table 4).

TABLE 3.—Lake Vermillion aerial survey results based on a model-based regression procedure. A regression model was developed to explain estimates of fishing effort (per quarter fishing day) in terms of available explanatory variables. Seasonal totals were estimated by integrating the regression model over time.

| Stratum | Day type[a] | Period[b] | Seasonal total effort (boat-hours) | Variance | 2 × CV (%)[c] |
|---|---|---|---|---|---|
| 1 | Weekday | 1 | Unestimable | | |
| 2 | Weekday | 2 | 26,358 | 14,254,275 | 28 |
| 3 | Weekday | 3 | 26,358 | 14,254,275 | 28 |
| 4 | Weekday | 4 | 26,358 | 14,254,275 | 28 |
| 5 | Weekend | 1 | 14,172 | 65,557,284 | 114 |
| 6 | Weekend | 2 | 24,504 | 2,004,244 | 11 |
| 7 | Weekend | 3 | 24,504 | 2,004,244 | 11 |
| 8 | Weekend | 4 | 24,504 | 2,004,244 | 11 |
| Total | | | 166,758 | $2.08 \times 10^8$ | 17 |

[a] Weekdays: Monday–Friday exclusive of holidays. Weekend days: Saturdays, Sundays, holidays.

[b] Quarter of the fishing day.

[c] Coefficient of variation: CV = 100(standard error)/estimate; 2 × CV is half the width of the 95% confidence interval about the estimated seasonal total, as a percentage of the estimated total.

## Discussion

The two methods of analysis gave similar estimates of effort for strata 2–8 combined (156,567 boat-hours for stratified random sampling [Table 1] versus 166,758 for model-based sampling [Table 3]). The estimated confidence limits were also extremely close (±15% of the estimate for stratified random sampling and ±17% for model-based sampling). It is disappointing that the model-based procedure did not produce narrower confidence bands. However, sampling dates were chosen randomly (so that stratified random sampling could be used) rather than optimally, and only a portion of the data (the first 19 weeks) were analyzed so that the temporal trend was less evident. It appears that the trend in the data must be quite strong for the model-based approach to be efficient. This also was noted by Hansen et al. (1983). A strong temporal trend, suggesting that model-based sampling might be appropriate, would be expected for seasonal fisheries such as those exploiting migratory stocks.

With careful planning, the performance of the model-based approach can be improved. Certain intuitive considerations are likely to be useful. Thus, near-orthogonal designs will minimize covariance terms. Sampling the extremes of the model space (i.e., the beginning and end of the season, the edges of the geographical region of the interest) should aid in model discrimination. An appropriate formal design criterion would be to minimize the integrated variance of prediction ($I_\lambda$-optimal design). Cook and Nachtsheim (1982) provided an extremely flexible method (and algorithm) for finding optimal designs. Their method

TABLE 4. Lake Vermillion aerial survey results based on the simplified regression model: $\text{effort}_i = b_0 + b_1\text{wk}_i + b_2\text{wksq}_i + b_3D_i + b_4P1_i + e_i$; $R^2 = 0.56$. The $e_i$ are residual errors.

**Parameter estimates**

| | | | |
|---|---|---|---|
| Constant (const) | $b_0 = 914$ | Day type ($D$) | $b_3 = 361$ |
| Week (wk) | $b_1 = -100$ | Period ($P1$) | $b_4 = -534$ |
| Week$^2$ (wksq) | $b_2 = 2.77$ | | |

**Covariance matrix**

| | const | wk | wksq | $D$ | $P1$ |
|---|---|---|---|---|---|
| const | 11,570 | | | | |
| wk | −2,655 | 5,028 | | | |
| wksq | −2,080 | 37.34 | 599.2 | | |
| $D$ | 90.08 | −3.058 | −29.64 | 1.565 | |
| $P1$ | 331.7 | −2,471 | −443.8 | 31.29 | 25,660 |

**Stratum estimates**

| Stratum | Day type[a] | Period[b] | Effort (boat-hours) | Variance | 2 × CV (%)[c] |
|---|---|---|---|---|---|
| 1 | Weekday | 1 | Unestimable | | |
| 2–4[d] | Weekday | 2–4 | 78,383 | $2.19 \times 10^8$ | 38 |
| 5 | Weekend | 1 | 3,855 | $2.20 \times 10^6$ | 77 |
| 6–8[d] | Weekend | 2–4 | 72,460 | $5.71 \times 10^8$ | 66 |
| Total overall | | | 154,698 | $7.89 \times 10^8$ | 36 |

[a] Weekdays: Monday–Friday exclusive of holidays. Weekend days: Saturdays, Sundays, holidays.
[b] Quarter of the fishing day.
[c] Coefficient of variation: CV = 100(standard error)/estimates; 2 × CV is half the width of the 95% confidence interval about the estimated seasonal total, as a percentage of the estimated total.
[d] Three strata combined.

provides for uncertainty in the functional form of the regression model by averaging the results over a series of possible polynomial models ($\bar{L}$-optimal design). The user can specify the weighting of the different models. Thus, one can specify, for example, that the functional form is most probably a quadratic but that it might be a cubic or, less likely, a linear model. If, on the other hand, one has no idea of the degree of the polynomial, other than that it is probably less than or equal to a quartic (say), one could compute the optimal design with uniform weights.

It might be argued that the computed variances for the model-based approach are likely to be too optimistic because they are based on an assumption that the regression model is appropriate, whereas, in reality, uncertainty is associated with both the parameter estimates and the model formulation. To a certain extent, the computed variances will reflect the inappropriateness of the model through lack of fit. However, Monte Carlo simulation studies would be useful to evaluate the importance of model specification errors. As more information is gained about the system being modeled (by the accumulation of samples over several years), there should be less uncertainty in the specification of a suitable model.

The model-based approach can provide a convenient way to avoid logistical problems. It provides a measure of robustness against missed sampling opportunities and can also be used to avoid having to sample widely separated locations simultaneously. However, if the variable of interest (e.g., effort) varies over both time and space, care must be taken to avoid confounding the effects of the explanatory variables by, for example, always sampling one location early in the season and another region later.

Another reason for choosing model-based sampling is to obtain a detailed description of a fishery. In the Lake Vermillion example (equation 8), the model is easily interpreted as follows.

(1) Effort at the beginning of the season on weekdays during periods 2, 3, and 4 averages 620 boat-hours per period ($b_0$).

(2) For weekend days, 1,037 boat-hours ($b_3$) are added per period for periods 2, 3, and 4.

(3) For period 1 on weekends, 1,311 boat-hours ($b_4$) are subtracted from the weekend average ($b_0 + b_3$); effort during period 1 on weekdays is not estimable from the regression.

(4) For weekday periods 2–4, effort drops off with time according to the quadratic equation: decline = −84 wk + 3.8 wksq (based on $b_1$ and $b_2$).

(5) For weekend days, periods 2–4, effort drops off according to: decline = (−84 − 67)wk + 3.8 wksq.

(6) For period 1 on weekend days , effort drops off according to: decline = (−84 −67 + 104) wk + 3.8 wksq.

The other regression model employed (Table 4) describes the essential features of the fishery adequately and is even simpler to explain to managers and user groups.

The model-based approach can be used as an exploratory tool. For example, the effect of time of the season was modeled as a polynomial. The effect of time within a day also could be explored. One approach would be to plot the residuals from the fitted model against the time of day to look for a trend that could be modeled (i.e., use an added variable plot: Weisberg 1980). If time of day and time within the season are both significant, one could integrate over both variables.

In our example, the order of the polynomial function of time was decided on the basis of a general $F$-test (Weisberg 1980). If the observations within a week are regarded as nearly equivalent to multiple observations at a single time, a test for lack of fit can be used to decide on the order of the polynomial. This procedure was used by Chapman (1965) for a structurally similar problem involving tag returns over time from a single-release tagging experiment. Chapman pointed out that, if a notable trend occurs within weeks, the estimate of error sum of squares will be inflated. To test for linearity, Chapman suggested fitting regression lines to the data within each week and pooling the residual sums of squares. This procedure should be appropriate in general but may be rather tedious for complex models.

Model-based sampling can be used in a variety of situations involving temporal and spatial trends, including estimation of catch, effort, and catch rate. If catch rate and effort are each modeled by functions of time, say $f(t)$ and $g(t)$, respectively, catch would be estimated by $\int f(t)g(t)dt$. Estimates of (relative) stock abundance can be obtained by integration of counts along a transect, such as in a hydroacoustic survey.

To date, model-based sampling has received scant attention in the fisheries literature. A few simple examples exist in the sportfishing literature on creel surveys (Lambou 1961; Bulkley 1966; Powell and Bowden 1981). Mundy (1979) used a model-building approach to study migratory timing of salmonid fishes, and Reilly et al. (1983) fitted a model to counts of migrating whales to estimate the size of the population. Pope and Woolner (1985) fitted regression models to mackerel egg counts and suggested the use of numerical integration for computing estimates of total production. In ICES (1986)[2], the total egg production by the sole *Solea solea* was estimated by trapezoidal integration to compute the area under a temporal production curve. It seems likely that increased use will be made of model-based sampling in the future.

**Acknowledgments**

We are grateful for the helpful comments received from Michael Sissenwine, Tim Smith, Stephen Reilly, and John Pope. The International Council for the Exploration of the Sea graciously permitted us to cite the report of a Working Group. Nancy Anderson provided secretarial support. Partial support for this study was provided through the Cooperative Institute for Marine and Atmospheric Studies by National Oceanic and Atmospheric Administration cooperative agreement NA85-WCH-06134.

## References

Barnard, J., W. Meyers, J. Pearce, F. Rumsey, M. Sissenwine, and W. Smith. 1985. Surveys for monitoring changes in renewable resources: forests and marine fisheries. American Statistician 39:363–373.

Bulkley, R. V. 1966. Catch of the 1965 tidewater cutthroat sport fishery and notes on the life history of the coastal cutthroat trout in the Siuslaw River, Oregon. Oregon State University, Fisheries Research Report 4, Corvallis.

Cassel, C.-M., C.-E. Sarndal, and J. Wretman. 1977. Foundations of inference in survey sampling. Wiley, New York.

Chapman, D. G. 1965. The estimation of mortality and recruitment from a single-tagging experiment. Biometrics 21:529–542.

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

Cook, R. D., and C. J. Nachtsheim. 1982. Model robust, linear-optimal designs. Technometrics 24:49–54.

Cramér, H. 1946. Mathematical methods of statistics. Princeton University Press, Princeton, New Jersey.

Hansen, M. H., W. G. Madow, and B. J. Tepping. 1983. An evaluation of model-dependent and probability-sampling inferences in sample surveys. Journal of the American Statistical Association 78:776–793.

ICES (International Council for the Exploration of the Sea). 1986. Report of the ad hoc working group on the 1984 and 1985 sole (*Solea solea* L.) egg surveys. ICES, C.M. 1986/G:95, Copenhagen.

[2]Cited with permission of the International Council for the Exploration of the Sea.

Kendall, M., and A. Stuart. 1977. The advanced theory of statistics, volume 1, 4th edition. Macmillan, New York.

Lambou, V. 1961. Determination of fishing pressure from fishermen or party counts with a discussion of sampling problems. Proceedings of the Annual Conference Southeastern Association of Game and Fish Commissioners 15:380–401.

Mundy, P. R. 1979. A quantitative measure of migratory timing illustrated by application to the management of commercial salmon fisheries. Doctoral dissertation. University of Washington, Seattle.

Pope, J. G., and L. Woolner. 1985. Improving the sampling design of western mackerel egg surveys. International Council for the Exploration of the Sea, C.M. 1985/D:11, Copenhagen.

Powell, T., and D. Bowden. 1981. Fisherman survey and harvest analysis. Colorado Division of Wildlife, Federal Aid in Fish and Wildlife Restoration, Job Completion Report F-55R, Fort Collins.

Reilly, S. B., D. W. Rice, and A. A. Wolman. 1983. Population assessment of the gray whale, *Eschrichtius robustus*, from the California shore censuses, 1967–80. U.S. National Marine Fisheries Service Fishery Bulletin 81:267–281.

Robson, D. 1961. On the statistical theory of roving creel census of fishermen. Biometrics 17:415–437.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters, 2nd edition. Macmillan, New York.

Weisberg, S. 1980. Applied linear regression. Wiley, New York.