Estimating the Size of Steelhead Runs by Tagging Juveniles and Monitoring Migrants

David A. Boughton*

*National Marine Fisheries Service, Southwest Fisheries Science Center,
110 Shaffer Road, Santa Cruz, California 95060, USA*

Abstract.—The number of individuals in a spawning run of anadromous fish can be estimated by tagging juveniles with passive integrated transponders during the freshwater phase and the subsequent monitoring of adult upstream migrants by use of instream tag readers. The method may enable monitoring of steelhead *Oncorhynchus mykiss* in systems where other methods are intractable. I developed a hierarchical capture–resighting model and applied it to simulated data to develop a relation between the number of marked juveniles and the precision of run size estimates. Precision is primarily controlled by the number of tagged spawners that ultimately return and are detected; only 30–90 tagged spawners are required to obtain relatively precise estimates of run size. For typical marine survival rates, 0.0033–0.0330, this translates to a tagging effort of between 3,400 and 45,000 juveniles/cohort. Estimates are robust to imperfect detection of tagged fish if at least two instream readers independently scan for tags. Reach sampling allows for estimation of run size in large stream systems. Based on my simulations, the number of reaches could be as low as 30–40 under scenarios of high marine survival. Computer code is provided for the estimator using the freely available statistical software R.

Anadromous fish species typically produce annual spawning runs, the size of which is a key metric for tracking the health and status of populations. Spawner abundance is not always monitored in practice, however, because such monitoring is difficult and expensive. A direct estimate of run size requires in situ counting of upstream migrants or spawning adults, usually during times when weather and stream conditions make field work difficult or hazardous. However, without such monitoring, the size of anadromous fish runs must often be conjectured from anecdotal information and indirect reasoning about habitat conditions, and management outcomes in terms of fish cannot be directly linked to management actions, thus inhibiting the learning process.

Runs of winter steelhead *Oncorhynchus mykiss* (anadromous rainbow trout) are especially problematic for direct monitoring because they tend to migrate during the season of highest flows and most turbid

conditions. In areas such as southern California, the run sizes themselves are small and can be concentrated into the few days after high-flow events, when the use of weirs, redd surveys, or direct observation is impractical (Boydstun and McDonald 2005).

An alternative to direct monitoring of steelhead runs is to infer run sizes indirectly from estimates of juvenile abundance. Obviously, estimates of juvenile abundance bear some relationship to adult run size, but they provide a reduced ability to detect trends over time (Shea and Mangel 2001) or to estimate extinction risk; also, because of density dependence, their relationship to adult run size may have an unknown form. Moreover, in steelhead populations, the abundance of juveniles can be a misleading index of adult run sizes because anadromous juveniles are indistinguishable from the nonanadromous rainbow trout, with which they often co-occur. Without the ability to distinguish between the anadromous and nonanadromous forms, one cannot unambiguously monitor the anadromous component via juvenile abundances.

Even so, juvenile abundances have the advantage that they are straightforward to estimate using well-known methods of electrofishing that can be deployed under benign summer low-flow conditions (e.g., Rosenberger and Dunham 2005; Temple and Pearsons 2006). Estimates are made by placing block nets prior to electrofishing, which thus allows for a variety of abundance estimators based on closed-population assumptions. What is needed to disambiguate such data is (1) a method for estimating the probability of transition from juvenile steelhead to anadromous out-migrant and (2) a method for estimating subsequent marine survival to spawning.

Specifically, if N is the number of juvenile steelhead in a given season (prior to out-migration), then the number of resulting spawners can be represented as a binomially distributed random variable:

$$S \sim \text{Binomial}(N, s_1 s_2), \qquad (1)$$

where s_1 is the combined survival and transition rate to out-migration (i.e., rate of successful smolting) and s_2 is subsequent marine survival. Here, I explore the idea that the parameters s_1 and s_2 can be estimated by tagging juveniles with passive integrated transponder
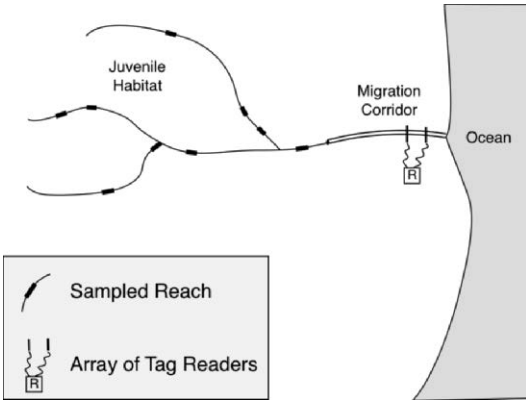
* E-mail: David.Boughton@noaa.gov

FIGURE 1.—Schematic diagram of a two-stage T-JAMM (tagging juveniles and monitoring migrants) design. Reaches are randomly sampled from the stream network. Juveniles are sampled from these reaches, implanted with passive integrated transponder tags, released back to their home reach, and monitored for emigration and return migration using an array of tag readers deployed in the main migration corridor of the stream network.

(PIT) tags and monitoring the adult upstream migrants. This method of tagging juveniles and monitoring migrants (hereafter, T-JAMM design) requires that tagged fish be detectable during their out-migration and return migration, a feat that is now achievable due to recent advances in tagging technology. These advances enable tagged fish to be detected at distances of 20–30 cm from the antennae of tag readers, suitable for a variety of passive monitoring devices placed instream and operated continuously at low cost (Zydlewski et al. 2006). Arrays of antennae allow estimation of per-reader detection probabilities (Zydlewski et al. 2001, 2006; Bond et al. 2007), and the general approach has so far been adapted to streams as wide as 4.5 m (Ibbotson et al. 2004).

My purpose here is to explore key statistical parameters of T-JAMM designs: namely, what scale of tagging effort is required to obtain reasonably precise estimates of run size? Are hundreds, thousands, or tens of thousands of tagged fish necessary to obtain good estimates, and what aspect of the study design or statistical model most limits precision? Can precision be appreciably improved through the use of hierarchical statistical models? To address these questions, I develop a hierarchical Bayesian resighting model for fish that are sampled from reaches that are in turn sampled from stream networks. The supplementary material has a computer code package, bfisher (available at http://swfsc.noaa.gov; select Data Portal → Research Software → bfisher), which can apply the model to real data sets using the freely available

computer language R (R Development Core Team 2006).

Methods

Model Development

Hierarchical models.—In small stream systems, it may be possible to electrofish the entire network of channels, in which case it is simple to develop an estimator for S using equation (1) as a special case of Pollock's (1982) robust design for mark–recapture studies. The more typical situation, however, is a stream network that is too extensive to permit electrofishing of every meter of channel. A useful strategy for such cases is a two-stage sampling scheme (Figure 1). In a two-stage design, reaches are randomly sampled from the stream network (reach sampling) and then fish are sampled from the reaches (fish sampling). The resulting data are used to make statistical assertions about all of the fish in the entire stream network. Typically, the reach sampling is conducted using a computerized geographical information system (GIS) that contains a representation of the stream network digitally segmented into individual reaches on the order of 100–300 m long. A set of reaches is randomly selected and identified in the field using a Global Positioning System unit, and fish are then sampled from each reach using block nets and electrofishers in the conventional way.
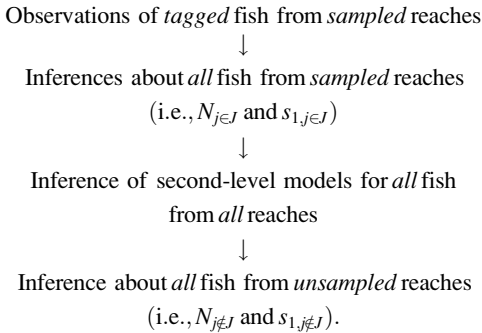
For two-stage sampling, it is necessary to extend equation (1) to multiple reaches:

$$S_j \sim \text{Binomial}(N_j, s_{1,j}s_2), \qquad (2)$$

where j indexes individual stream reaches and total run size is $S = \Sigma \ S_j$ (for simplicity, s_2 is assumed to be the same for all fish from all reaches). The reaches fall into two sets: those that are in the sample (symbolized by $j \in J$, where J represents the sample) and those that are not in the sample (symbolized by $j \notin J$). The statistical problem here is that the sampled fish must be used to estimate the parameters of sampled reaches ($N_{j \in J}$ and $s_{1,j \in J}$), which must then be used to estimate the parameters of unsampled reaches ($N_{j \notin J}$ and $s_{1,j \notin J}$). This entails a second level of statistical modeling that is not present in the conventional robust design but that can readily be developed using concepts such as hierarchical or multilevel generalized linear modeling (Gelman et al. 2004; Gelman and Hill 2007). A hierarchical model takes a Bayesian view of probability—that parameters, not data, have probability distributions—and explicitly models those probability distributions using another statistical model at a second level.

In the problem here, the second level of modeling comprises (1) a model of how juvenile abundance N_j

varies among all reaches in the stream network and (2) a model of how $s_{1,j}$ varies for fish from different reaches in the stream network. The flow of information in this modeling process is

Observations of *tagged* fish from *sampled* reaches
↓
Inferences about *all* fish from *sampled* reaches
$$(\text{i.e.}, N_{j\in J} \text{ and } s_{1,j\in J})$$
↓
Inference of second-level models for *all* fish from *all* reaches
↓
Inference about *all* fish from *unsampled* reaches
$$(\text{i.e.}, N_{j\notin J} \text{ and } s_{1,j\notin J}).$$

The first two steps involve Bayesian inference, which (except for the simplest sorts of problems) cannot be solved analytically. However, the last 10 years have seen enormous improvements in a family of simulation techniques, known collectively as Markov-chain Monte Carlo (MCMC) methods, that together with today's fast computers have allowed complex Bayesian inference to become a routine pursuit by nonspecialists (see Gelman et al. 2004; Congdon 2005). Here, I use open-source MCMC algorithms implemented in two R packages: R2WinBUGS and BRugs.

The last step involves data augmentation (Gelman et al. 2004), which simply means using a fitted statistical model to generate probability distributions of "unobserved" data—in this case, the abundances and fates of untagged fish. The underlying assumption of this step is that sampled and unsampled reaches are exchangeable, a generalization of the "independent, identically distributed" assumption of frequentist statistics that is applicable to the case of random parameters assumed by the Bayesian approach.

The simulation approach of Bayesian MCMC and data augmentation, while numerically intensive, has some attractive features. Most notably, it allows one to dispense with normal approximations for computing such features as means, medians, or confidence limits. Instead, one can obtain an arbitrarily large set of simulated "draws" from the probability distribution of interest (typically 1,000 draws are used), and one can then estimate quantities such as the median or 95% limits directly from these draws without any assumptions as to the parametric form of the underlying probability distribution. This is convenient in the current context because many statistics for mark–recapture models are quite nonnormal, especially when numbers of recaptures are low (Ricker 1975).

Another feature of the approach is that results have a simple interpretation. In contrast to conventional frequentist statistics, which assume some "true" statistical model, the Bayesian approach treats models as useful constructs for making predictions about unobserved data. I believe that this accords better with how most biologists view models, and thus the interpretation of results is more intuitive than for the frequentist approach. For example, the 95% confidence interval for S can be interpreted simply as "There is a 95% probability that the interval contains the actual S," which matches how most such results are interpreted in practice. Under the conventional frequentist paradigm, such an interpretation is incorrect.

The price of the Bayesian approach, which treats parameters as random variables, is that each parameter must be assigned a "prior probability distribution," a statement of belief about its plausible values, prior to making any observations relevant to the parameter. The problem is that if prior information is absent or neutral, for many distributions (such as the multinomial) there is no way to specify a prior completely devoid of information (i.e., assumptions). In practice, this is not a terrible problem, for even though one cannot specify a *completely* noninformative prior, one can specify a *relatively* noninformative prior (e.g., for mortality, a uniform probability between 0 and 1). It is generally found that even with modest amounts of data, the observations quickly outweigh the influence of an uninformative prior (Jaynes 2003). Of course, when only one or two tagged fish are recaptured, an uninformative prior may have a notable influence on the estimate; if zero tagged fish are recaptured, one can still obtain an estimate, but it will be completely dominated by the prior. In other words, one has learned nothing from the data.

I now turn to specifying a series of Bayesian hierarchical models for the T-JAMM design under two-stage sampling.

Tagging phase: estimating juvenile abundances.— Assume that fish are sampled by two electrofisher passes over a reach that is temporarily isolated by block nets. Juveniles captured in the first pass are marked with PIT tags and returned to the reach prior to the second pass. Rosenberger and Dunham (2005) argued that this mark–recapture approach is more robust than traditional depletion-based methods. The data consist of counts of fish captured only in the first pass (denoted F_j), only in the last pass (L_j), or in both passes (B_j; all $j \in J$). In what follows, I make extensive use of the equivalence of the multinomial and Poisson distributions when conditioned on total counts (Gelman et al. 2004; Congdon 2005); thus, familiar formulations of mark–recapture models based on the

multinomial are here portrayed as sets of Poisson-distributed variables following the example of Gelman et al. (2004: 431).

If the probability of capture per pass (p_j) is exchangeable across fish and passes of the electro-fisher, then

$$F_j \sim \text{Poisson}\left[\lambda_j p_j (1 - p_j)\right]$$
$$B_j \sim \text{Poisson}(\lambda_j p_j^2) \qquad j \in J \qquad (3)$$
$$L_j \sim \text{Poisson}\left[\lambda_j p_j (1 - p_j)\right],$$

from which it is possible to estimate the capture probability (p_j) and the expected number of fish in the reach (λ_j). From the exchangeability assumption, it follows that the uncaptured fish (U_j) in reach $j \in J$ are distributed as

$$U_j \sim \text{Poisson}\left[\lambda_j (1 - P_j)^2\right]; \qquad j \in J, \qquad (4)$$

which is one of the quantities of interest. Once the parameters $p_{j \in J}$ and $\lambda_{j \in J}$ have been estimated for the sampled reaches, it is straightforward to obtain a predictive distribution of $U_{j \in J}$ using data augmentation (Gelman et al. 2004).

The other quantity of interest is the number of juveniles in unsampled reaches ($N_{j \notin J}$):

$$N_j \sim \text{Poisson}(\lambda_j); \qquad j \notin J, \qquad (5)$$

which can be estimated from $\lambda_{j \notin J}$ using data augmentation. This requires a second-level model for how the expected count λ_j varies across reaches, so that the $\lambda_{j \notin J}$ of unsampled reaches can be estimated from the $\lambda_{j \in J}$ of sampled reaches. A general-purpose model is the lognormal linear model with covariates (e.g., Wyatt 2002, 2003):

$$\log_e(\lambda_j) = \beta X_j + \varepsilon_j + \varepsilon_k + \log_e(w_j) \qquad (6)$$

for all reaches j. Here, X_j is a vector of regression covariates that can be obtained for both sampled and unsampled reaches, such as attributes obtained from a GIS; and β is a vector of estimated regression weights. The middle terms denote random effects, with ε_j being the estimated variation among reaches, ε_k being the estimated variation among groups of reaches, and k indexing the group. This grouping effect can be useful when different parts of the stream network naturally group into distinct units with potentially different character, such as different tributaries. It thus represents a discrete form of spatial autocorrelation. The last term, w_j, is an offset provided by the user and can be used to account for reaches of different size. For example, the w_j could be set to the lengths of the reaches, in which case the other parameters are standardized per unit length of channel rather than per reach. The offsets must be available for the unsampled reaches as well as the sampled ones.

This general-purpose model is easily simplified by omitting any or all of the following: (1) the random group effect, (2) the covariates, and (3) the offsets. When all are omitted, the model is simply

$$\log_e(\lambda_j) = \beta + \varepsilon_j, \qquad (7)$$

where β is the log-mean abundance of juveniles across sampling units. Simpler models will often be preferable: thus, if one lacks any useful covariates, then omit the terms for covariates; if there is no meaningful way of grouping sample reaches (e.g., a single stream rather than a network), then omit the group effect; and if all sample reaches are the same length, then there is little point in including the offset. The bfisher code allows any combination of these three components to be omitted if so desired. If one is unsure whether to omit a term, both versions of the model can be fit and the best one can be selected via some model selection criterion, such as the deviance information criterion implemented in R2WinBUGS and elsewhere.

The regression weights and random effects can be estimated using standard MCMC techniques (the bfisher code uses the packages R2WinBUGS and BRugs; see Sturtz et al. 2005; R2WinBUGS software and documentation are available from the Comprehensive R Archive Network [CRAN 2009], as is R itself. Unfortunately, the compiled version of BRugs is no longer available from CRAN; I have included a copy with the bfisher code, which can be installed using the "Packages & Data/Package Installer/Local Binary Package" menu item in R. For additional information on BRugs, see also Thomas 2009). From this, one produces a predictive distribution of the $\lambda_{j \notin J}$ for unsampled reaches and thence distributions of $N_{j \notin J}$ using data augmentation.

The capture probability, p_j, is a nuisance parameter, the precision of which may limit the overall success of the model. Indeed, Rosenberger and Dunham (2005) conducted an extensive campaign of field sampling and found that p_j had stronger relationships to site-level covariates than did rainbow trout abundance itself. Such covariates need only be measured at the sites actually sampled and may improve the precision of estimates, particularly for sites with low numbers of fish recaptured. The bfisher code allows covariates for p_j, which is implemented with a complementary log–log link but is otherwise similar to equation (6):

$$\log_e\left[-\log_e(1 - p_j)\right] = \beta X_j + \varepsilon_j + \varepsilon_k + \log_e(w_j) \quad (8)$$

for $j \in J$. The complementary log–log link on the left side of equation (8) is attractive because it gives the right side the interpretation of "log hazard"—the log of capture rate for uncaptured fish. This will often be more intuitive than the interpretation of the familiar logistic link, which is the log-odds that fish are captured versus not captured. The complementary log–log link implies that the hazard rate is a product of the effects of the various model terms (βX_j, ε_j, etc.) and also allows w_j to play the same role as in equation (6), scaling those effects per unit channel length.

Monitoring phase: estimating survival to spawning.—The purpose of the monitoring phase is to estimate $s_{1,j}$, s_2, and a set of nuisance parameters, the

per-capita detection rates achieved by instream tag readers. Detection rates typically fall in the range of 0.55–1.00 (Zydlewski et al. 2006). Estimating detection rate requires two or more readers placed in series along the migration corridor and negligible mortality of migrants between the two readers; this enables the use of closed-population assumptions.

To illustrate the model for a two-reader design, let Ro_{11} be the number of tagged out-migrants detected by both readers, let Ro_{01} be the number missed by the first reader but detected by the second, and so forth with subscripts indicating detection patterns. This gives the model

$$\begin{bmatrix} Vo \\ Ro_{01} \\ Ro_{11} \\ Ro_{10} \end{bmatrix}_j \sim \text{Multinomial}\left\{ M_j, \begin{bmatrix} s_{1,j}(1-d_1)(1-d_2) + (1-s_{1,j}) \\ s_{1,j}(1-d_1)d_2 \\ s_{1,j}d_1 d_2 \\ s_{1,j}d_1(1-d_2) \end{bmatrix} \right\} \tag{9}$$

where M_j is the total number of tagged fish and is calculated as $F_j + B_j$ (note that $M_j = F_j + B_j + L_j$ if fish were tagged on both electrofishing passes. M_j would also need to be discounted by any fish that are collected or killed while electrofishing—see Discussion); and $Vo_j = M_j - Ro_{+,j}$, the number of tagged fish that were never detected migrating out (where $Ro_{+,j}$ is short-hand for $Ro_{01,j} + Ro_{11,j} + Ro_{10,j}$). The new parameters, d_1 and d_2, are detection rates of tagged out-migrants at the first and second readers, respectively. The unobserved fish (Vo_j) are actually a mixture of fish that migrated but were not detected, fish that never migrated (became freshwater residents), and fish that died before migrating. In equation (9), the probability on the top right-hand side contains a term $(1 - s_{1,j})$ for the dead and nonmigratory fish.

Tagged spawners returning 1 or 2 years later also produce counts of detected fish (symbolized by $Ri_{01,j}$, $Ri_{11,j}$, and $Ri_{10,j}$) and a calculated number of unobserved fish ($Vi_j = M_j - Ri_{+,j}$). The corresponding model is similar to equation (9) but with two differences: (1) two new parameters (d_3 and d_4) describe detection rates for incoming adults at each reader and (2) all occurrences of $s_{1,j}$ are replaced by the product $s_{1,j}s_2$.

The $s_{1,j}$ must be estimated for unsampled reaches as well as sampled reaches, but one might expect this parameter to vary quite markedly across the stream network. This variation is addressed using a hierarchi-

cal model similar to equation (8) but with a new set of regression parameters and with p_j being replaced by $s_{1,j}$. As with equations (6) and (8), this allows one to make estimates of a parameter in unsampled reaches without the restrictive assumption that the parameter is identical for all reaches in the system.

Similarly, the detection probabilities (d_1, d_2, $d_3 \ldots$) can be modeled hierarchically as a simple random effects model,

$$\log_e[-\log_e(1 - d_a)] = \beta + \varepsilon_a, \tag{10}$$

in which a subscripts each particular reader–migrant combination. Alternatively, the detection rates can be assumed identical,

$$\log_e[-\log_e(1 - d_a)] = \beta, \tag{11}$$

in which case one can refer simply to $d = d_1 = d_2 = \cdots = d_a$.

Estimating Run Size

Given the model just described, S is a sum of four quantities: (1) the known number of incoming tagged migrants that were detected, (2) an estimate of incoming tagged migrants that were not detected by the readers, (3) an estimate of untagged migrants from sampled reaches, and (4) an estimate of migrants from unsampled reaches. The sum of the above four quantities is

TABLE 1.—Range of parameter values used to simulate data based on steelhead distribution in the Arroyo Seco River system, California. Models are described in Methods.

Parameter	Model[a]		
	Simplest	Imperfect detection	Reach sampling
Detection probability (d)[b]	$d = 1$	$0.50 < d < 0.99$	$d = 1$
Number of juveniles (N)	$500 < N < 40{,}000$	$500 < N < 40{,}000$	$263{,}000 < N < 371{,}000$
Capture probability (p)[c]	$0.1 < p < 0.4$	$0.1 < p < 0.4$	$0.50 < p < 0.78$
Survival to out-migration (s_1)	$0.2 < s_1 < 0.8$	$0.2 < s_1 < 0.8$	$0.2 < s_1 < 0.8$
Marine survival (s_2)	$10^{-3} < s_2 < 10^{-1}$	$10^{-3} < s_2 < 10^{-1}$	$s_2 = 0.033$ or 0.0033
Expected run size $(S = Ns_1s_2)$	$0.1 < Ns_1s_2 < 3{,}200$	$0.1 < Ns_1s_2 < 3{,}200$	$433 < Ns_1s_2 < 6{,}121$

[a] Population parameters (N, s_1, s_2) for the first two models are intended to represent a range of generic situations; parameters for the reach sampling model represent the situation in the Arroyo Seco River system.
[b] Detection probability is assumed to be the same for all tag readers; hence, $d_1 = d_2 = d_3 \ldots = d$.
[c] For the first two models, the range of p is that obtained by Rosenberger and Dunham (2005; their Figure 1). In the reach sampling model, the range of p is that obtained empirically for the Arroyo Seco River system.

$$S = \sum_{j \in J} Ri_{+j}$$
$$+ \sum_{j \in J} \text{Binomial}\left[Vi_j, s_{1,j}s_2(1-d_3)(1-d_4)\right]$$
$$+ \sum_{j \in J} \text{Binomial}(U_j, s_{1,j}s_2)$$
$$+ \sum_{j \notin J} \text{Binomial}(N_j, s_{1,j}s_2),$$
$$(12)$$

which are estimated using MCMC techniques followed by data augmentation.

Characterizing Tagging Effort versus Precision

The cost of the T-JAMM design includes a fixed cost of acquiring and operating the in situ tag readers and the nonfixed costs of tagging the fish, the latter being adjustable depending on the annual sampling effort. Intuitively, one expects that a greater investment in tagging will deliver a more precise estimate, but by how much? My performance metric for this question is relative precision (P), defined for my purposes as the interquartile range of the estimate scaled by the median of the estimate (i.e., $P = [S_{75\%} - S_{25\%}]/S_{50\%}$, analogous to a coefficient of variation). Simulated data were used to explore the relation between the number of fish tagged and the resulting precision of the run size estimate.

The full statistical model was rather numerically intensive, so for the simulations I developed simplified versions to examine specific tradeoffs. The simplest model assumed that all reaches were electrofished (i.e., there were no unsampled reaches) and that there was perfect detection of tagged fish by in situ readers. The next model added imperfect detection, and the next assumed that some reaches were unsampled.

Analyses were programmed in R using BRugs and

R2WinBUGS to iterate the Markov chains. Except for the model with unsampled reaches, data sets were simulated by randomly drawing parameter values from the log-uniform distributions described in Table 1 (which corresponded to prior probabilities of the Bayesian statistical model). All models assumed two-pass electrofishing, with PIT tag implantation occurring after the first pass.

For the model with unsampled reaches, data sets were simulated using a predictive model of steelhead distribution in the Arroyo Seco River system of central California, described by Boughton et al. (2009). This study was based on snorkel counts of juveniles in a random sample of 31 reaches stratified into eight creeks, and abundance was estimated using lognormal models (Wyatt 2002). A predictive distribution was used to generate 200 random abundance values for each of the 1,930 unsampled reaches in the sampling frame. From each of the 200 simulated data sets, I drew a sample of 256 reaches and simulated mark–recapture data. First, an estimate of S was made using the data from the entire sample of 256 reaches. Half of the reaches were then discarded and a new estimate was made, iterating until the sample dropped to eight reaches (a sampling fraction of 0.130–0.004). To reduce computer time, detection of upstream migrants was assumed to be error-free. For half of the 200 data sets, I assumed a high s_2 of 0.033 (Bond 2006), and for the other half I assumed a low s_2 of 0.0033.

Results

Simplest Model

The simplest model is somewhat unrealistic, as it assumes that the entire stream network gets electrofished and that all tagged fish are detected during their downstream and return migrations. It thus represents a best-case scenario for achievable precision. Since Bayesian probability can be interpreted as an optimal

representation of uncertainty (Jaynes 2003; Van Horn 2003), the simplest model implies a fundamental limit on what one can learn about run size using a T-JAMM design.

As expected, in an analysis of 1,000 simulated data sets, the 95% credible intervals of S contained the true value 95% of the time, except for cases where zero tagged fish survived to spawn (in which case the prior probability of s_2 dominated the estimate). These zero-spawner data sets were omitted from further analysis.

The value of P varied widely from 0.05 to 2.00. It did not vary with most population parameters, such as N, s_1, or s_2, though it was somewhat related to the total S (Figure 2). It did not vary with juvenile tagging probability but was weakly related to the number of tagged juveniles and the number of spawners without tags. However, P was closely related to the number of tagged spawners (Figure 2, bottom right corner). The data exhibited a memorable "30 for 30%" rule: to achieve P less than 0.30, it is sufficient to recover 30 tagged spawners. However, due to the log-linear nature of the relationship, to guarantee P less than 0.10 one must recover a much greater number of tagged spawners: at least 200.

Of course, the number of tagged spawners will depend on the number of tagged juveniles, which is generally under one's control to some extent, but it also depends on s_1 and s_2. If one can make an informed guess at s_1 and s_2 beforehand, the approximate number of juveniles to tag is

$$M_{\text{Total}} \approx \frac{4}{s_1 s_2 P^{1.7}}. \qquad (13)$$

Performance Cost of Imperfect Detection

Next, I simulated imperfect detection during the monitoring phase for installations of two, three, or four readers. Zydlewski et al. (2006) described three field deployments in which detection probabilities of tagged fish in freshwater creeks ranged from 0.55 to 1.00 per detector, so I randomly selected values of d between 0.50 and 0.99 and assumed that all readers had the same detection probability.

Figure 3 shows the influence of the number of readers in terms of 5% exceedence curves for P. The exceedence curves were calculated using a normal approximation of the results from simulations. Imperfect detection only degrades P appreciably if just two tag readers are deployed and if fewer than six tagged spawners are detected. The "infinity" entry in Figure 3 depicts results from the simplest model of the previous section.

In these simulations, the model assumed all readers to have the same detection probability for all fish

(equation 11). Another model was used to explore the importance of this assumption, incorporating random effects of the detection probabilities (equation 10). The results (not shown) indicated a slight cost to precision for the case of 10 or more tagged spawners, such that 40 tagged spawners were required to guarantee a P of 0.30. Below 10 tagged spawners, one starts risking a bad situation in which the prior for detection rates at individual antennae sometimes dominates and in which accuracy is unreliable.

Performance Cost of Reach Sampling

Under reach sampling, the "30 for 30%" rule found earlier was changed to a "90 for 30%" rule when s_2 was high (0.033; left side of Figure 4). The number of tagged juveniles necessary to guarantee a given P would be about

$$M_{\text{Total}} \approx \frac{12}{s_1 s_2 P^{1.7}}. \qquad (14)$$

For a P of 0.3, this translates to a tagging effort of about 14,000 juveniles for an s_1 value of 0.2 and about 3,500 juveniles for an s_1 of 0.8, the low and high ends of the smolt transition rates I assumed.

When s_2 was low (0.0033), the "30 for 30%" rule was retained (right side of Figure 4). However, to achieve it, one had to sample more reaches to compensate for lower s_2. The higher number of sampled reaches evidently allowed the parameters of the linear model to be estimated with high precision. Thus, there was no performance cost for reach sampling relative to tagging the whole stream system. Of course, the necessary tagging effort was still much higher than under the high s_2, but it was only around 3.3 times higher rather than proportionate (i.e., 10 times higher). Under the "30 for 30%" rule (equation 13), the tagging effort would have to be around 47,000 juveniles for an s_1 of 0.2 and around 11,700 juveniles for an s_1 of 0.8.

Discussion

Key Assumptions

In general, mark–recapture estimators of the type described in this paper rely on six key assumptions for their accuracy (Ricker 1975: 81–82). The six assumptions are listed here with a discussion of their implications for T-JAMM designs.

Assumption 1: all tags must be detected.—This assumption must be met during the repeat passes of the electrofisher in the tagging phase or else estimates of juvenile abundance can be inflated. Since fish are physically captured during this phase, it should be straightforward to devise handling procedures to ensure that all tags get detected. In the monitoring phase,
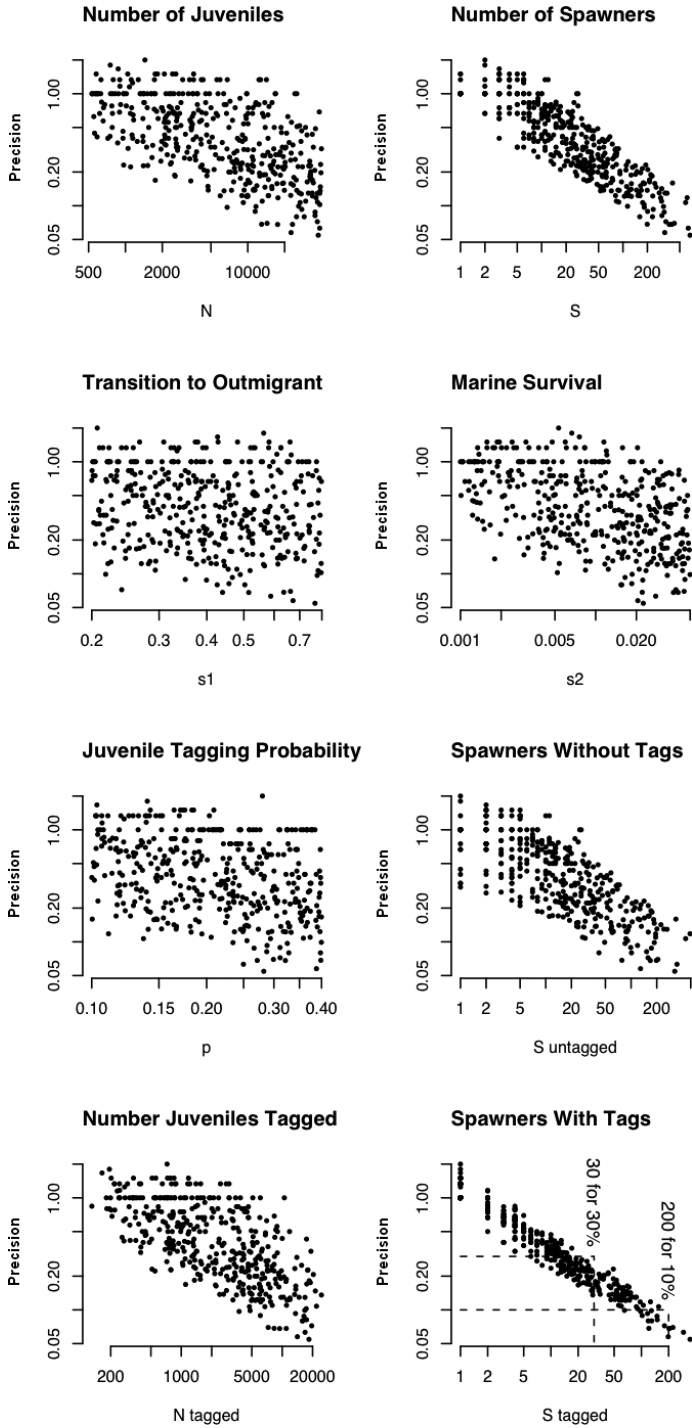
FIGURE 2.—Relative precision (*P*) for steelhead run size estimates using the simplest model applied to simulated data. Horizontal axes depict various key parameters of data sets. Depicted are 421 simulated data sets (out of 1,000) in which at least one tagged steelhead survived to spawn. All axes are on logarithmic scales.
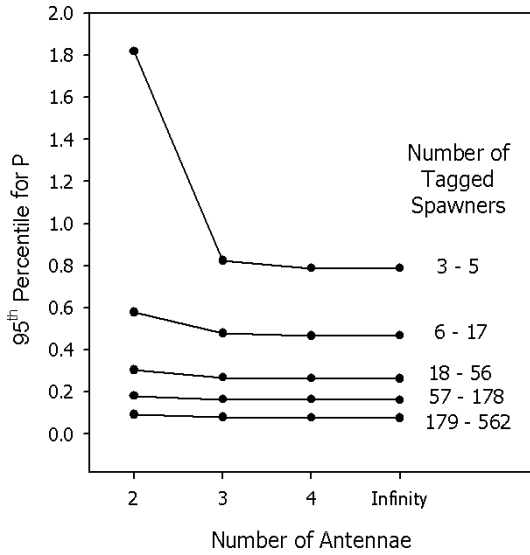
FIGURE 3.—Five-percent exceedence curves for relative precision (*P*) as a function of the number of in situ tag readers (antennae) and the number of tagged steelhead spawners. See Table 1 for parameters of simulated data. Five-percent exceedence is the one-tailed 95% limit for *P* (mean + 1.67·SD), meaning that only 5% of data sets would have *P* exceeding the given curve. The "infinity" column is for a model that assumes perfect detection of tags.

antenna, which could be violated if some tags malfunction or if some aspect of implantation (e.g., location or orientation in the fish's body cavity) affects detection rate. One way to formally test for violations of this assumption would be to test whether partial detections (fish detected at some antennae but not others) occur less frequently than expected by chance. This would require at least three antennae in the detection array, with more antennae providing greater power to detect violations.

Assumption 2: negligible recruitment occurs between capture and detection sessions.—Here, "recruitment" should be understood in a general sense as being recruitment to the spawning run from segments of the population not available for tagging. For example, if some juveniles are too small to tag at the time of the tagging phase but later become smolts, they are effectively "recruited" between the tagging and monitoring phases. Their contribution to the spawning run cannot be estimated without introducing additional assumptions into the models.

A common form of such recruitment arises when steelhead spawning runs consist of fish that have smolted at different ages and have spent different numbers of years in the ocean. For simplicity, the model described here did not explicitly incorporate this heterogeneity, but it can still be used to estimate such spawning runs by using either the covariates or the grouping variable to separate tagged fish into different combinations of age at tagging, year of tagging, and years from tagging to smolt transition. This of course
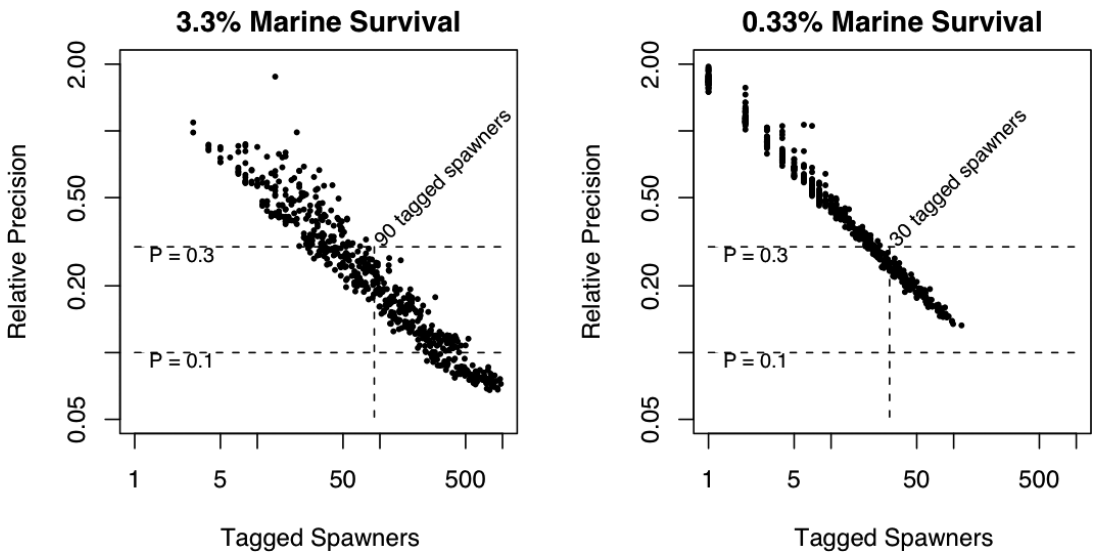
redundancy of tag detectors allows the assumption to be relaxed, as described previously. However, the estimator still makes a weaker assumption: that each tag has an independent risk of detection failure at each

FIGURE 4.—Relative precision (*P*) achieved under reach sampling for two scenarios of marine survival from simulated data based on steelhead of the Arroyo Seco River, central California.

would require an ongoing effort of tagging and monitoring over multiple years to assemble the required data set.

A more technical but elegant solution would be to extend the model structure itself to allow for multiple life history pathways. A straightforward framework for doing so is to situate the T-JAMM models within a broader matrix population model (Clark et al. 2005). In this framework, the projection matrix and population vectors can be viewed as a simple way to organize the explosion of parameters when the full diversity of life history paths is explicitly addressed. In addition, the parameters of the model (transitions, survivals, and abundances of specific life stages) can each be given a hierarchical structure describing their evolution from year to year. This provides a framework for modeling environmental stochasticity and density dependence that is ultimately useful for forecasting future run size using the data augmentation technique.

It is beyond to scope of the paper to assess these more complex models, but it seems likely that the number of tagged spawners will still be the main limit on achievable precision. This is because the number of tagged spawners will always be much less than the number of tagged juveniles, recaptured juveniles, or detected out-migrants, and so the uncertainty in ocean mortality will usually dominate the uncertainty of the run size estimate.

Assumption 3: tagged and untagged fish have the same vulnerability to capture.—Although stream fish often vary in catchability, mark–recapture estimates as used here appear to be relatively robust to violations of this assumption. For example, in an extensive empirical study of rainbow trout, Rosenberger and Dunham (2005) found that unequal catchability of fish led to large biases in depletion-based estimates of abundance but no detectable bias in mark–recapture estimates.

Assumption 4: tagged fish become randomly mixed with untagged fish and/or distribution of capture effort is proportional to fish density.—In the tagging phase, this assumption requires carefully designed procedures for re-introducing tagged fish to the entire sampling reach and for distributing the electrofishing effort over the entire reach in subsequent passes. In the monitoring phase, the chief danger is that portions of the reader array may tend to fail during high-flow events, which may be precisely the time that fish tend to migrate. If so, the distribution of detection effort may be *inversely* proportional to migrant density! This potential problem emphasizes the need to meet the alternative assumption, that tagged and untagged fish are randomly mixed (i.e., migration timing is randomly intermixed). Assuming that migration timing correlates with spatial aspects of geography, the simplest way to achieve this

random intermixing would probably be to randomly sample fish from across the entire watershed via reach sampling. This idea could use empirical validation.

Assumption 5: tagged and untagged fish have the same natural mortality.—Electrofishing and the field implantation of radio frequency identification tags are stressful for fish and can increase the chances of infection or outright mortality. The effect of tag-induced or handling-induced mortality on the estimates of run size depends on when it occurs:

- Mortality during the first capture (e.g., during the act of tagging the fish) effectively removes the fish from the study and should not be included in F_j. It thus causes λ_j to be slightly low (by the number of mortalities). If such mortalities occur frequently enough, they could have an appreciable effect on the parameters of the linear predictor (equation 6), so that juvenile abundance is underestimated in the unsampled reaches as well (by roughly $\Sigma\ X_{j \in J}/J$ juveniles per reach, where $X_{j \in J}$ is the number of mortalities in reach $j \in J$).

- Tag-induced mortality between the first and second captures of the tagging phase invisibly reduces the number of fish available for recapture, fooling the estimator into estimating λ_j too high (which, if common enough, would also affect the linear predictor). The magnitude of this effect per fish mortality will be larger than the negative effect of mortality during the first capture (by a factor of $\sim 1/p_i$); in addition, the number of such mortalities would probably not be unobservable. Thus, it is probably a good idea not to release tagged fish that seem to be in poor condition, and one should instead count them as mortalities during the first capture. If they recover in captivity, after the second electrofishing session they could potentially be released with the other fish and added back into M_j.

- Handling-induced mortality during the second capture does not affect the estimate of juvenile abundance. However, for this case and indeed any mortality during the tagging phase, the number of mortalities should be subtracted from M_j (i.e., the number of tagged fish potentially available for monitoring). Otherwise, s_1 would be underestimated by approximately $\Sigma\ X_j/\Sigma\ M_j$ (where X_j represents all handling-related mortalities at reach j).

- Tag-induced mortality during the monitoring phase would cause underestimation of the s_1 or s_2 of untagged fish and thus underestimation of run size. Unfortunately, this sort of tagging mortality is effectively unobservable under most circumstances. On the other hand, because the bias is toward an underestimate, it is essentially conservative in the

realms of both risk management and harvest management.

Assumption 6: fish do not lose their tags.—The final key assumption is that fish do not shed their tags. From the perspective of the estimation model, tag loss has generally the same effect as tag-induced mortality, with perhaps slightly larger biases being produced if loss occurs during the tagging phase. However, shedding of tags during the tagging phase should be observable—as recaptured juveniles with surgery scars but no tags, shed tags detected in the substrate of the stream, or both. Thus, it should be possible to roughly assess the magnitude of any tag loss that occurs.

Covariates

Covariates can improve the precision of spawner estimates in a number of ways. First, during the tagging phase, covariates of p_j might improve estimates of λ_j, particularly when small numbers of fish were caught (Rosenberger and Dunham 2005). These covariates need only be observed at sampled sites and include traits such as instream wood, emergent vegetation, and water depth. Note that the simulations had the same p_j in all reaches, so they already indicate a best-case scenario for precision and cannot be improved by covariates.

Second, covariates of juvenile abundance may also improve precision. Suitable covariates could improve the "90 for 30%" rule produced by the simulations, but they would never surpass the "30 for 30%" rule since this was a fundamental ceiling on precision of the T-JAMM design itself. Some follow-up simulations suggested that improvements to the "90 for 30%" rule would be modest at best (Figure 5). For example, a covariate having a moderate association with juvenile density (Pearson's product-moment correlation coefficient $r = 0.50$) generally improved the P by just a factor of 2.5% (median for the solid lines in Figure 5), although it occasionally improved P by a factor of 15–20% (Figure 5, solid lines). A covariate with a strong association (Pearson's $r = 0.90$) generally improved P by a factor of 14% (median), with occasional improvements around 20–25% (Figure 5, dashed lines). The number of reaches sampled does not appear to have much effect on the amount of improvement (Figure 5: compare top and bottom panels).

Practical Matters

Precision might also be improved by adjustments of the field technique. Rosenberger and Dunham (2005) showed that mark–recapture accuracy at individual reaches can be improved by making four passes rather than two passes. Temple and Pearsons (2006) con-
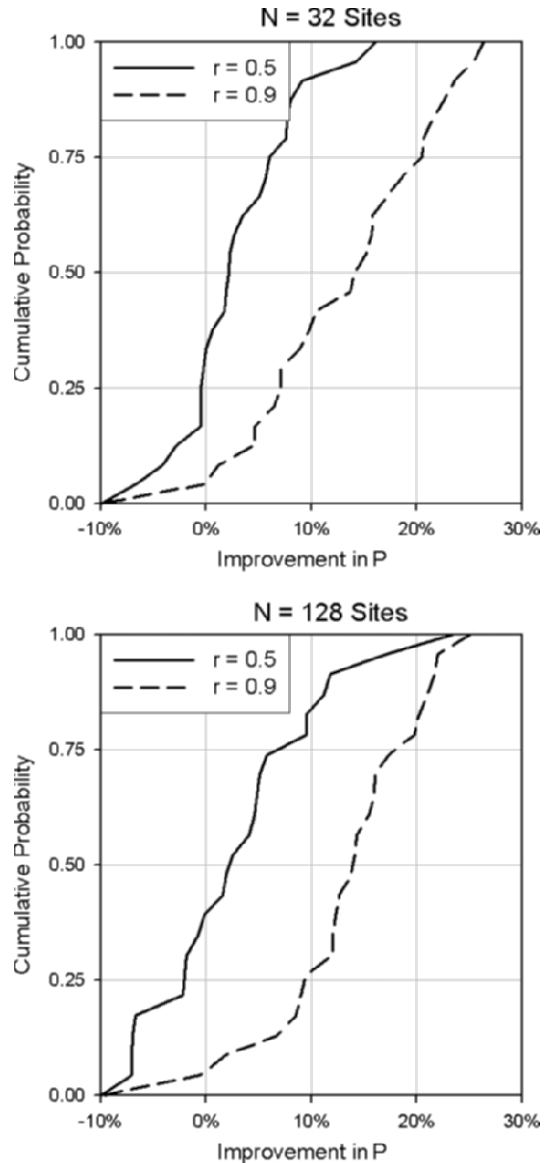

FIGURE 5.—Improvement in relative precision (P) when reach-level covariates are included in the estimation process. Shown are results for 24 simulations of the Arroyo Seco River (California) steelhead data with covariates for juvenile density and with marine survival set at 0.033. The strength of association is in terms of Pearson's product-moment correlation coefficient (r) between the covariate and the log(juvenile density) (i.e., $\log_e[\lambda_j/w_j]$; terms are defined in Methods). Improvement in P is calculated as [(old − new)/old] × 100%. The y-axis shows the probability of achieving an improvement lower than the score on the x-axis; for example, 0.50 indicates the median improvement (half larger, half smaller), and 0.95 indicates that nearly all improvements are smaller. Solid lines depict a covariate with a moderate association with juvenile abundance ($r = 0.50$); dashed lines depict a covariate with a strong association with juvenile abundance ($r = 0.90$).

ducted an experiment showing that electrofishing passes can be separated by as little as 3 h rather than the conventional 24 h without invalidating the key assumption (equal catchability of tagged and untagged fish). In fact, they showed that the conventional 24-h wait entails significant violations of other model assumptions (i.e., nonclosure due to fish escaping under block nets and overnight block-net failure that is probably nonrandom). Another advantage of the 3-h option is that it is less time intensive and allows more reaches and more fish to be sampled overall.

Nevertheless, the implementation of the T-JAMM design clearly requires a spirited tagging effort. The fundamental constraint on precision is the problem of tagging a sufficient number of juveniles to eventually produce around 30–90 tagged spawners/year. As a rough guide to cost, assume about $8,000 per reader array for a half-duplex tagging system, $2.75 per tag for 23-mm tags, and the capability of a three-person field crew to tag around 250 fish/d during the summer field season (T. Williams and D. Rundio, National Marine Fisheries Service, personal communication). Assuming field technicians at $3,000 per month in wages and benefits, this comes out to around $4 per tagged fish plus the fixed cost of the reader array. The scenarios discussed in the results range from a tagging effort of 3,500 fish/year, which translates to an annual cost of $22,000, to a tagging effort of 47,000 fish/year, which translates to an annual cost of $196,000. The middling tagging effort of 14,000 fish/year translates to a $64,000 annual cost. These estimates omit travel and maintenance costs and the fixed costs of maintaining the reader array. The assumption of 250 tagged fish/d may be too optimistic for inexperienced field crews, remote sites that require large travel times, or streams with low densities of juveniles.

If necessary, the tagging effort could be spread across runs in different stream systems even if the runs occur in separate stream basins and are considered demographically distinct populations. For example, if the effort to tag 14,000 juveniles/year is spread across 10 populations, it would translate to about 1,400 tagged juveniles per population per year and 3–9 tagged spawners per population per year. This would cause little loss of precision in the aggregate run size of the 10 populations since the estimator involves the same calculations as for a single population. However, it assumes that the 10 populations experience the same ocean mortality, so that a single parameter, s_2, applies to all 10 runs. This may be a restrictive assumption when s_2 is related to watershed-dependent characteristics of the fish. For example, s_2 can depend on smolt size (Ward et al. 1989), which in turn depends on the feeding opportunities and water temperatures experi-

enced in freshwater habitats (Harvey et al. 2005; Bond 2006; Boughton et al. 2007).

The T-JAMM design clearly has promise for monitoring populations of anadromous fish, especially when other methods are intractable or when the fish co-occur with nonanadromous conspecifics. Only a modest number of tagged spawners are needed for reasonable estimates, but the fundamental constraint of high ocean mortality means that many juveniles must be tagged to achieve this modest number. This appears to be the limiting constraint for practical implementation of T-JAMM designs.

Acknowledgments

References

Bond, M. H. 2006. Importance of estuarine rearing to central California steelhead (*Oncorhynchus mykiss*) growth and marine survival. Master's thesis. University of California, Santa Cruz.

Bond, M. H., C. V. Hanson, R. Baertsch, S. A. Hayes, and R. B. MacFarlane. 2007. A new low-cost instream antenna system for tracking passive integrated transponder (PIT)-tagged fish in small streams. Transactions of the American Fisheries Society 136:562–566.

Boughton, D. A., H. Fish, J. Pope, and G. Holt. 2009. Spatial patterning of habitat for *Oncorhynchus mykiss* in a system of intermittent and perennial streams. Ecology of Freshwater Fish 18:92–105.

Boughton, D. A., M. Gibson, R. Yedor, and E. Kelley. 2007. Stream temperature and the potential growth and survival of juvenile *Oncorhynchus mykiss* in a southern California creek. Freshwater Biology 52:1353–1364.

Boydstun, L. B., and T. McDonald. 2005. Action plan for monitoring California's coastal salmonids. Final report to NOAA Fisheries, Contract Number WASC-3-1295, Santa Cruz, California.

Clark, J. S., G. A. Ferraz, N. Oguge, H. Hays, and J. DiCostanzo. 2005. Hierarchical Bayes for structured, variable populations: from recapture data to life-history prediction. Ecology 86:2232–2244.

Congdon, P. 2005. Bayesian models for categorical data. Wiley, Chichester, UK.

CRAN (Comprehensive R Archive Network). 2009. Comprehensive R Archive Network: frequently used pages. Vienna University of Economics and Business, Department of Statistics and Mathematics, Vienna. Available: cran.r-project.org. (July 2009).

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian Data Analysis, 2nd edition. Chapman and Hall/CRC Press, Boca Raton, Florida.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge, UK.

Harvey, B. C., J. L. White, and R. J. Nakamoto. 2005. Habitat-specific biomass, survival, and growth of rainbow trout (*Oncorhynchus mykiss*) during summer in a small coastal stream. Canadian Journal of Fisheries and Aquatic Sciences 62:650–658.

Ibbotson, A. T., W. R. C. Beaumont, D. Collinson, A. Wilkinson, and A. C. Pinder. 2004. A cross-river antenna array for the detection of miniature passive integrated transponder tags in deep, fast flowing rivers. Journal of Fish Biology 65:1441–1443.

Jaynes, E. T. 2003. Probability theory: the logic of science. Cambridge University Press, Cambridge, UK.

Pollock, K. H. 1982. A capture–recapture design robust to unequal probability of capture. Journal of Wildlife Management 46:752–757.

R Development Core Team. 2006. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Fisheries Research Board of Canada Bulletin 191.

Rosenberger, A. E., and J. B. Dunham. 2005. Validation of abundance estimates from mark-recapture and removal techniques for rainbow trout captured by electrofishing in small streams. North American Journal of Fisheries Management 25:1395–1410.

Shea, K., and M. Mangel. 2001. Detection of population trends in threatened coho salmon (*Oncorhynchus kisutch*). Canadian Journal of Fisheries and Aquatic Sciences 58:375–385.

Sturtz, S., U. Ligges, and A. Gelman. 2005. R2WinBUGS: a package for running WinBUGS from R. Journal of Statistical Software 12(3):1–16.

Temple, G. M., and T. N. Pearsons. 2006. Evaluation of the recovery period in mark-recapture population estimates of rainbow trout in small streams. North American Journal of Fisheries Management 26:941–948.

Thomas, A. 2004. BRugs user manual, version 1.0. Department of Mathematics and Statistics, University of Helsinki, Helsinki. Available: www.math.helsinki.fi/openbugs/data/Docu/BRugs%20Manual.html. (July 2009).

Van Horn, K. S. 2003. Constructing a logic of plausible inference: a guide to Cox's theorem. International Journal of Approximate Reasoning 34:3–24.

Ward, B. R., P. A. Slaney, A. R. Facchin, and R. W. Land. 1989. Size-biased survival in steelhead trout (*Oncorhynchus mykiss*)—back-calculated lengths from adult scales compared to migrating smolts at the Keogh River, British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 46:1853–1858.

Wyatt, R. J. 2002. Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model. Canadian Journal of Fisheries and Aquatic Sciences 59:695–706.

Wyatt, R. J. 2003. Mapping the abundance of riverine fish populations: integrating hierarchical Bayesian models with a geographic information system (GIS). Canadian Journal of Fisheries and Aquatic Sciences 60:997–1006.

Zydlewski, G. B., A. Haro, K. G. Whalen, and S. D. McCormick. 2001. Performance of stationary and portable passive transponder detection systems for monitoring of fish movements. Journal of Fish Biology 58:1471–1475.

Zydlewski, G. B., G. Horton, T. Dubreuil, B. Letcher, S. Casey, and J. Zydlewski. 2006. Remote monitoring of fish in small streams: a unified approach using PIT tags. Fisheries 31(10):492–502.