Modeling juvenile salmon migration using a simple Markov chain

E. Ashley Steel   Peter Guttorp

NRCSE

Technical Report Series

Modeling juvenile salmon migration using a simple Markov chain

E. Ashley Steel
and
Peter Guttorp

National Research Center for Statistics and the Environment
University of Washington, Seattle, Washington 98195, U.S.A

SUMMARY

We describe movement patterns of hatchery-raised, juvenile, spring chinook salmon, *Oncorhynchus tshawytscha*, using a two-state Markov chain model. The existence of two states, moving and holding, is suggested by anecdotal information from a large radio-tagging study; yet, data describing such observations of small-scale fish behavior are not adequate to estimate transition probabilities directly. Instead, we estimate the transition probability matrix from travel times within each of 11 river segments using a method of moments approach. Bootstrapped confidence intervals are presented. The precision of parameter estimates is improved by incorporating information on river velocity at each observation station. Results suggest differences in fish behavior between river segments.

1. Introduction

1.1 Background

Recent evidence indicates that substantial mortality of chinook salmon, *Oncorhynchus tshawytscha*, from the Snake River system, located in Washington, Oregon, and Idaho, occurs in the free-flowing segments of the river above Lower Granite Dam (LGR); yet, there is little information about fish behavior in this area. Previous research has focused on the managed sections of the river below LGR because, historically, poor survival has been attributed to difficulties in passage through hydroelectric facilities (Raymond 1979; Raymond 1988). The current estimates for survival of hatchery-produced, yearling, chinook salmon to LGR, the first dam encountered during seaward migration, have been as low as 15 - 35% (Lower Granite Migration Study Steering Committee 1993). The data suggest that improved understanding of behavior during this migration period might enable improved management strategies. Our aim is to develop a stochastic model which provides insights into small-scale fish behavior within the constraints of well-described, larger-scale models of migration processes.

We are interested in developing a simple methodology that can account for the relationship between individual fish behavior and environmental conditions so that the technique might be applied by statisticians and non-statisticians alike. Fisheries radio-tagging technology is advancing rapidly as tags get smaller and biologists become more adept at the required surgical and electronic skills. With these advances, there has been a proliferation of radio-tag data; however, methods for analyzing such information are not readily available in the fisheries literature.

The current data set provides some of the first detailed information on fish behavior and in-stream conditions in the free-flowing segments of the river above LGR reservoir. Research on environmental controls of migratory behavior in other species or in other regions provides evidence that factors such as water velocity, water temperature and available light may regulate fish movements (Jonsson 1991; Berggren and Filardo 1993; Independent Scientific Group 1996). The eventual goal of this research is to develop a model which can be used to test for the importance of environmental covariates in regulating behavior. In this paper, we develop a basic model to describe observed patterns of fish movement. We then incorporate mean river velocity into the model to approximate actual river conditions and we examine it's effect on the precision of parameter estimates.

1.2 Data

The fisheries data for this analysis are from a large radio-tagging study carried out by the National Marine Fisheries Service. Combination radio transmitter/passive integrated transponder (PIT) tags were surgically implanted into 129 yearling chinook salmon at Lookingglass Hatchery in March 1997. The fish were allowed to recover in the hatchery for approximately two weeks after which time they were released into Lookingglass Creek (LGC). Sample size was reduced by mortality both at the hatchery and during migration. During their migration from LGC to the LGR Reservoir, the fish swam past 12 fixed-site receiving stations. Due to signal strength, antenna orientation, tag failures, and other difficulties with the electronic equipment, each fish was detected at only some

of the stations.   Results of this analysis should not be extrapolated to all hatchery fish because comparisons between fish with combination radio transmitter / PIT tags and fish with only PIT tags suggest that the radio-tag may affect migration behavior. For further details on the radio-tagging experiments, see Hockersmith *et al.* (1998).

In this paper, we describe a technique to model travel time between stations. Travel time was calculated between each pair of contiguous stations (not including the initial release point) for each fish observed at both endpoints. There are between 7 and 31 observations for each of the 11 segments.

River velocity data was collected at 8 of the 12 observation stations during the period of out-migration. Velocity (m/sec) is defined by the maximum observable surface velocity, which was estimated from the travel time of floating objects over a fixed distance and, where possible, from a boat using a Global Positioning System (GPS). Velocity generally increases as ones moves downstream; however, the final sites on the Snake River are just upstream of LGR Reservoir and velocity decreases significantly in this area.

2. Markov Chain Model

2.1 Model Summary

To model travel times between stations, a two-state Markov chain model (Guttorp 1995) was selected to meet two criteria. First, the selected model should converge to the inverse Gaussian distribution in the limit. Previous research on migrations of large cohorts of

fish between dams has shown that the distribution of arrival times at dams follows the inverse Gaussian distribution extremely well (Zabel *et al.* 1998). Second, the model should describe migration patterns observed in the field. During the study, mobile tracking was used to pinpoint fish locations between the fixed-site monitoring stations. Contrary to expectations, fish were often observed to stay in the same location for several days at a time before re-initiating downstream movement. A two-state Markov process is a simple model that meets these two criteria.

We use the two-state Markov model to describe fish behavior between observation opportunities. Our model assumes that fish make movement "decisions" 10 times every hour with the probability of movement dependent on the action taken during the previous time interval. For the basic model, a movement action is defined as a downstream travel distance of 1 km. Velocity is incorporated into the model by allowing fish to move a distance that is dependent on the relative mean velocity in each river segment. We observe a fish at a particular monitoring station after it has moved L kilometers downstream with L defined by the length of the river segment. By defining the model in this way, we can use arrival distributions to estimate biologically meaningful parameters within the transition matrix.

2.2 Notation

The following notation will be necessary for the calculations in the next section. Let

$\text{p}_{i,j}$ = probability of movement decision i during the current time interval given movement decision j at the previous time interval,

w = wait time to move one unit of distance,

t_k = travel time through river segment k,

v_k = mean velocity in river segment k,

L_k = length (km) of river segment k,

n_k = number of observations in segment k,

i, j = 0, 1 (0 = stay, 1 = move), and

k = 1, 2, 3, … , 11.

2.3 Estimation of the Transition Matrix Using Method of Moments

The first step in applying the method of moments is to calculate the expected value and the variance of t_k using the Markov model. First, we calculate the expectation and variance of w.   We assume that each fish is initially in the move state, a reasonable assumption given that the fish must be moving to enter each study segment.

$$E(w) \;=\; \sum_{w=0}^{\infty} w p(w)$$

$$=\; 1p_{11} + 2(1-p_{11})(1-p_{00}) + 3(1-p_{11})(1-p_{00})p_{00} + 4(1-p_{11})(1-p_{00})p_{00}^{\,2} \cdots$$

$$=\; p_{11} + (1 - p_{11})(1 - p_{00})\sum_{x=0}^{\infty}(2+x)p_{00}^{\,x}$$

$$=\; p_{11} + (1 - p_{11})\left(2 + \frac{p_{00}}{1 - p_{00}}\right).$$

And,

$$\mathrm{Var}(w) \;=\; \sum_{w=0}^{\infty} w^2 p(w) - E(w)^2 .$$

$$\sum_{w=0}^{\infty} w^2 p(w) = \; 1^2 p_{11} \; + \; 2^2 (1-p_{11})(1-p_{00}) \; + \; 3^2 (1-p_{11})(1-p_{00})p_{00} \; + \; 4^2(1-p_{11})(1-p_{00})p_{00}^2 \cdots$$

$$= \; p_{11} \; + \; (1-p_{11})(1-p_{00}) \sum_{x=2}^{\infty} x^2 p_{00}^{\,x-2}$$

$$= \; p_{11} \; + \; (1-p_{11})(1-p_{00}) \left[\frac{2}{(1-p_{00})^3} \; + \; \frac{1}{p_{00}} \; + \; \frac{1}{p_{00}(1-p_{00})^2} \right].$$

Therefore,

$$\mathrm{Var}(w) \;=\; p_{11} \; + \; (1-p_{11})(1-p_{00}) \left[\frac{2}{(1-p_{00})^3} \; + \; \frac{1}{p_{00}} \; + \; \frac{1}{p_{00}(1-p_{00})^2} \right]$$

$$- \left[p_{11} \; + \; (1-p_{11}) \left(2 \; + \; \frac{p_{00}}{1-p_{00}} \right) \right]^2 .$$

We are interested in the moments of t_k, the total travel time through segment k. Total

travel time can be calculated as the sum of the individual wait times, w. These w are

independent given that the initial state for each interval must be 1. Therefore,

$$E(t_k) = L_k\!\left(\frac{v_k}{\bar{v}}\right) E(w) \quad \text{and} \quad \mathrm{Var}(t_k) = L_k\!\left(\frac{v_k}{\bar{v}}\right) \mathrm{Var}(w), \quad \text{where } L_k\!\left(\frac{v_k}{\bar{v}}\right) \text{ represents the}$$

number of required movements in segment k at relative velocity $\dfrac{v_k}{\bar{v}}$. The parameter v_k is

equal to one for all k river segments in the basic model.

The mean and variance of the inverse Gaussian distribution can be calculated from the

data as $\bar{t}_k$ and $\dfrac{\left(\bar{t}_k\right)^3}{\sum\limits_{n_k}\left(\dfrac{1}{t_k}-\dfrac{1}{\bar{t}_k}\right)}$ respectively (Folks and Chhikara 1978). The method of

moments estimator of the transition probability matrix is calculated by setting the

expected value and variance equal to their observed values and solving for $\hat{p}_{00}$ and $\hat{p}_{11}$.

The estimates $\hat{p}_{01}$ and $\hat{p}_{10}$ can be calculated simply as $1-\hat{p}_{00}$ and $1-\hat{p}_{11}$. Although

these estimators are not necessarily efficient or unbiased, they are reasonable and can be

obtained with a minimum of mathematical difficulty (Larsen and Marx 1986).

3. Results

3.1 Estimated Transition Matrices by River Segment

Tables 1 and 2 display the estimates and bootstrapped 95% confidence intervals (1000

iterations) for both p_{00} and p_{11} by river segment for the basic model (Table 1) and for the

model incorporating relative velocity (Table 2). Confidence intervals were constrained to

(0,1). A lack of significant digits in the estimate identifies occasions where the simulated

estimates were outside this range. Table 1 also includes the number of observations in

each segment and the length of that segment (km).

Velocity is incorporated into the model by multiplying the reach length, L_k, by the relative

mean velocity, $\dfrac{v_k}{\bar{v}}$. The effect of incorporating velocity in the model is to reduce or

increase the number of movements required to complete a given river segment. We

assume that, where the river is faster, a moving fish travels further with each "decision"

to move and vice versa. Table 2 includes the relative velocity in each segment. It

contains only nine of the original segments because velocity data was not available for

two of the segments.

Table 1 - Estimates and Confidence Intervals for p_{00} and p_{11} from the basic model.
Number of observations and segment length are included. GRR = Grande Ronde River,
SR = Snake River.

p_{00}			River Data			p_{11}		
95% CI	estimate	95% CI	Segment #	n_k	L_k	95% CI	estimate	95% CI
0.9735	0.9882	0.9947	GRR (2)	31	41	0.9826	0.9859	0.9918
0.3147	0.9607	0.9881	GRR (3)	11	19	0	0.8902	0.9390
0	0.9899	0.9957	GRR (4)	7	11	0.5807	0.9622	0.9689
0.7458	0.9458	0.9716	GRR (5)	16	25	0.8903	0.9465	0.9580
0	0.8132	0.9244	GRR (6)	25	15	0.4251	0.8572	0.9059
0.1859	0.8662	0.9296	GRR (7)	25	23	0.6856	0.9173	0.9350
0.9308	0.9728	0.9842	GRR + SR (8)	21	6	0.8455	0.8858	0.9062
0.9182	0.9789	0.9909	SR (9)	23	13	0.8581	0.9216	0.9449
0.9969	0.9993	0.9997	SR (10)	20	25	0.9754	0.9876	0.9918
0.9626	0.9996	0.9999	SR (11)	20	1	0.3727	0.9323	0.9647
0.9956	0.9999	1.0000	SR (12)	20	7	0.9332	0.9894	0.9946

Table 2 - Estimates and Confidence Intervals for p_{00} and p_{11} incorporating river velocity.
GRR = Grande Ronde River, SR = Snake River. Relative velocity is calculated as $\frac{v_k}{\bar{v}}$.

p_{00}			River Data		p_{11}		
95% CI	estimate	95% CI	Segment #	Relative Velocity	95% CI	estimate	95% CI

0.9804	0.9893	0.9947	GRR (2)	1.10	0.9864	0.9894	0.9972
0.3638	0.9598	0.9879	GRR (3)	0.90	0	0.8706	0.9292
0.9734	0.9908	1	GRR (4)	1.36	0.9514	0.9771	0.9994
0.7816	0.9499	0.9736	GRR (5)	1.08	0.9232	0.9580	0.9672
0.9450	0.9584	1	GRR (7)	1.44	0.9784	0.9949	0.9999
0.9259	0.9725	0.9836	GRR + SR (8)	0.92	0.8258	0.8723	0.8944
0.9964	0.9992	0.9997	SR (10)	0.80	0.9674	0.9844	0.9901
0.9715	0.9996	0.9999	SR (11)	0.69	0.2404	0.9037	0.9493
0.9970	0.9999	1.0000	SR (12)	0.57	0.8985	0.9813	0.9909

Estimates of both p_{00} and p_{11} are above 0.80 for all segments in both models although the 95% confidence interval is often large. In the basic model, it is difficult to distinguish any pattern by comparing the estimates of p_{00} or p_{11} between segments of the river or between the two rivers and the confluence.

Patterns are somewhat more apparent in the velocity model. For both p_{00} and p_{11}, there is a difference between the estimates for the Snake River (discounting the estimate for segment 11 which is only 1 km long) and for the confluence (segment 8). Estimates for p_{00} are all larger in the Snake River than in the Grande Ronde River although this difference does not appear to be significant.

3.2 Comparing Precision Between the Two Models

The effect of velocity on the width of the 95% confidence intervals around p_{00} and p_{11} is described in Figure 1.

Figure 1: Width of 95% confidence intervals for p_{00} and p_{11} for the basic model (0) and for the model incorporating velocity (V).
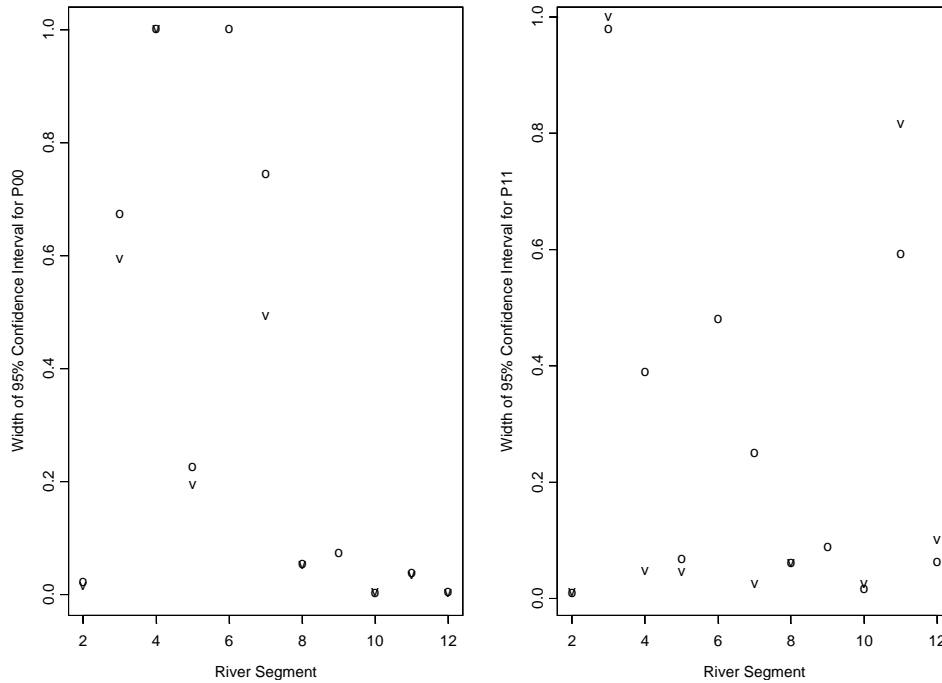
In all cases, the width of the confidence interval around p_{00} was reduced or unchanged when velocity was incorporated into the model. The confidence intervals around p_{11} for segments 2, 9, 10, and 11 increased with the addition of relative velocity. These four segments have a relative velocity < 1. The only other segment with a relative velocity < 1 is segment 8, the confluence. The precision of the estimate of p_{11} for segment 8 was unchanged with the addition of velocity.

4. Discussion

The two-state Markov model is a simple model to describe the process of migration in juvenile salmonids. Parameters within the transition probability matrix yield information about behavior that would be difficult to observe directly. Using the method of moments

approach described here, estimates of the transition probability matrix can be calculated

from arrival time distributions, frequently observed in both radio-tagging studies and in

the large PIT tag studies carried out by the National Marine Fisheries Service.

These results suggest that both p_{00}, the probability of a fish staying at one movement

opportunity given that it stayed at the previous movement opportunity, and p_{11}, the

probability of a fish moving given that it moved at the previous movement opportunity,

are fairly large. As a result, we would expect the actual behavior to include fairly long

runs of moving and staying. Further, there are rarely large differences between $\hat{p}_{00}$ and

$\hat{p}_{11}$ indicating that, on average, time is divided fairly evenly between staying and moving.

In effect, the basic model describes observed patterns (fish behavior plus river movement)

while the velocity model describes fish behavior. Table 2 suggests that fish behavior

might be different in different segments of the river. The larger values of $\hat{p}_{00}$ in the

Snake River indicate that there may be longer runs of staying in these segments, even

after adjusting for mean velocity. Mobile-tracking of radio-tagged fish during the study

period also identified parts of the Snake River as areas of delay (Hockersmith *et al*.

1998). Further, this analysis indicates that fish behavior in the region of the confluence

may be different than behavior further downstream, perhaps having shorter runs of

staying and holding.

The use of mean velocity to adjust the distance traveled in a given movement increased

the precision of most estimates of $\hat{p}_{00}$ and $\hat{p}_{11}$. However, further refinements of this

approach might better accommodate changes in river condition by using velocity during

the exact time interval in which a fish passes through a particular river segment rather

than mean velocity. This approach will also be a useful framework for incorporating other

environmental information, such as temperature and available light, into a model that

describes small-scale fish movement patterns during migration.

REFERENCES

Berggren, T. J. and M. J. Filardo (1993). An analysis of variables influencing the migration of juvenile salmonids in the Columbia River Basin. *North American Journal of Fisheries Management* **13**: 48-63.

Folks, J. L. and R. S. Chhikara (1978). The inverse Gaussian distribution and its statistical application - a review. *Journal of the Royal Statistical Society, B* **40**(3): 263-189.

Guttorp, P. 1995. *Stochastic Modeling of Scientific Data*. Chapman and Hall: Bury St. Edmunds, Great Britain.

Hockersmith, E., W. D. Muir, S. G. Smith, and B.P. Sanford. (1998). *Determining Areas of Loss and Delay for Juvenile Hatchery Salmonids above Lower Granite Dam Reservoir, 1997*. U.S. DOE, Bonneville Power Administration. In Press. Contract # DE-AI79-93BP10891.

Independent Scientific Group, The. (1996). *Return to the River: Restoration of Salmonid Fishes in the Columbia River Ecosystem*. Northwest Power Planning Council. September 10.

Jonsson, N. (1991). Influence of water flow, water temperature and light on fish migration in rivers. *Nordic Journal of Freshwater Resources* **66**: 20-35.

Larsen, R. J. and M. L. Marx (1986). *An Introduction to Mathematical Statistics and its Applications*. Prentice Hall: Englewood Cliffs, NJ.

Lower Granite Migration Study Steering Committee (1993). *Research Plan to Determine Timing, Location, Magnitude, and Cause of Mortality for Wild and Hatchery Spring/Summer Chinook Salmon Smolts Above Lower Granite Dam*. U.S. DOE, Bonneville Power Administration. October. Contract # DE-AI79-92BP61176.

Raymond, H. L. (1979). Effects of dams and impoundments on migrations of juvenile chinook salmon and steelhead from the Snake River, 1966 to 1975. *Transactions of the American Fisheries Society* **108**(6): 505-529.

Raymond, H. L. (1988). Effects of hydroelectric development and fisheries enhancement on spring and summer chinook salmon and steelhead in the Columbia River Basin. *North American Journal of Fisheries Management* **8**(1): 1-24.

Zabel, R. W., J. J. Anderson, and P. Shaw (1998). A multiple reach model to describe the migratory behavior of Snake River yearling chinook salmon (Oncorhynchus tshawytscha). *Canadian Journal of Fisheries and Aquatic Sciences*(In Press).