Essential Fish Habitat project status report

Reporting date: October 31, 2008

Project number: 2008-4

Title: Physical and temporal aspects of pollock spawning habitat utilization

PIs: Kevin Bailey and Kung-sik Chan

Funding year: 2008

Funding amount: $33,980

Status: ☐ Complete ☒ Incomplete, on schedule ☐ Incomplete, behind schedule

Planned completion date if incomplete: June 2009 for the 2008 contract year. The contract to Dr. Chan for the modeling work was not in place until August 2008. A contract report was filed in September detailing progress. His contract is for one year, but we anticipate he will complete the project no later than June 2009.

Reporting: Have the project results been reported? If yes, where were the results reported? The work only began in August 2008 due to the time lags of funds arriving at AFSC and contracts being let out to University of Iowa. From preliminary results we expect to report the results in a high-ranking journal.

Results: What is the most important result of the study?

This project has 2 components: first to assess the temporal aspects of pollock spawning habitat utilization and second to assess the spatial aspects. Since the project was designed as a 2 year project, we decided to tackle the temporal aspects of pollock spawning habitat utilization in the first year of funding 2008 and in the second year to complete that analysis and then examine the spatial aspects.

We pulled together data on the hatchdate distributions of walleye pollock spawning from 19 years (Table 1) and critical environmental and biological indices (Table 2). The funding for the modeling work arrived at the University of Iowa in August 2008. The important preliminary results are reported here.

We applied a generalized additive model (GAM) to examine variability in the hatchdate distributions of pollock in 19 years of survey data. We tested 7 local environmental factors, 5 climate indices, 2 stock assessment aspects and 1 human intervention as potential covariates in the model. To briefly summarize, our preliminary results indicate that the hatchdate distributions in time are influenced by factors influencing the timing of spawning, including the age composition of the spawning population, temperature prior to spawning, and the 1989 regime shift. The survival of larvae between hatching and sampling is influenced by temperature, transport out of the Shelikof Sea Valley and the 1989 Exxon Valdez oil spill. These results are described in more detail below.

Table 1. The time series of years with hatchdate distributions for larval walleye pollock collected in mid to late May surveys in the Shelikof Strait sea valley, Gulf of Alaska.

1981	1988	1993	1998
1983	1989	1994	1999
1985	1990	1995	2000
1986	1991	1996	2001
1987	1992	1997	

Table 2. List of environmental and biological indices utilized in analyzing hatchdate distribution data. These data were accumulated from a variety of sources.

Shelikof transport index	Shelikof sea valley spring SST
Shelikof Strait winter bottom temperature	GAK station winter bottom temperature
Daily wind mixing	Daily alongshore wind index
GOA river discharge	Shelikof flow index
Shelikof retention index	PDO
NPI	AO
EP-NP	MEI
Shelikof MACE spawning biomass	Shelikof MACE age structure

The following generalized additive model (GAM) structure was applied to study the pollock's abundance distribution over hatchdates and the effects from environmental factors and human activities:

$$l_{y,t,k} = n_y + \sum_a \alpha_{y,a}[\eta_a + s_a(t)] + \sum_b [\eta_b + s_b(t)] \times 1_b + \beta_0 - \beta_1 k^\gamma + \sum_q s_q(W_{q,y,t,k}) + \varepsilon_{\mathrm{y,t,k}} \quad (1)$$

where y denotes the year and t represents the day in the hatching date window; in other words, the number of days since March 1. $l_{y,t,k}$ is the log abundance of the caught larval pollock that were hatched in day t of year y, and survived k or more days. n_y is the log spawning biomass in year y, and is an offset variable in model 1. Additionally, we assume the baseline hatching time distribution is a mixture of the hatchdate distribution of the old pollock spawners and that of the young spawners. Such a mixing hatching time distribution is shown as $\sum_a \alpha_{y,a}[\eta_a + s_a(t)]$, where $\alpha_{y,a}$ is the proportion of group-a in year y; a can be 1 or 2, corresponding to old (older than say 7 years) or young spawner groups respectively; $[\eta_a + s_a(t)]$ gives the non-centered smooth function for baseline hatadate distribution for group a. Furthermore, there may exist regime shifts of the hatchdate distributions in specific periods, and such regime shifts are represented by $\sum_b [\eta_b + s_b(t)] \times 1_b$. 1_b's are the dummy variables to indicate the periods with the regime shift of the hatchdate distributions from the baseline. Moreover, $-\beta_1 k^\gamma$ represents the baseline probability, on a log-scale, for a larva to survival k or more days. $s_q(W_{q,y,t,k})$ is the smooth function of the qth covariate, which is applied to access the environmental effects on the pollock's spawning and survival. Finally, ε is the error process that is independent and identically normally distributed with zero mean and constant variance.

The symbol s, with various subscripts, denotes a smooth function. For model identifiability, these smooth functions are centered so that they are of zero means, and their corresponding η's represent the (possibly non-zero) mean effects. In particular, $\eta_{a=1}$ is the baseline mean (log) hatching abundance of the old spawners, and $\eta_{a=2}$ corresponds to that of the young spawners. The parameter η_b is the difference of the mean log hatching abundance in the period indicated by 1_b relative to the baseline.

Model 1 provides a framework for estimating the pollock's baseline hatchdate distributions and their larvae's baseline survival rate, but also for assessing the environmental and human intervention effects on the pollock's spawning and survival. $n_y + \sum_a \alpha_{y,a}[\eta_a + s_a(t)] + \sum_b [\eta_b + s_b(t)] \times 1_b + \beta_0 - \beta_1 k^\gamma$ equals the expected log abundance of larvae spawned on day t and survived for k or more days, given the spawning biomass and its age distribution and assuming average environmental conditions and no human and natural interventions. Therefore, the difference between the observed log larval abundance and the above expected value gives the residuals from the expected log abundance. Hence, model 1 encapsulates the nonparametric effects of the environmental conditions, the Exxon Valdez intervention and the regime shift, after adjusting for the pollock spawning dynamics and the natural mortality process of the larvae. In particular, our approach improves on the ad hoc approach by Bailey and Macklin (1994) who assessed the association of the environmental effects with the hatchdate distribution based on residuals computed using a fixed egg spawning distribution and a fixed survival distribution. Moreover, the new approach allows us

to study the possibly nonlinear environmental effects, and estimate the baseline hatchdate distributions and survival probabilities from the data.

Our preliminary work indicates the following generalized additive model (GAM) provides the best fit to the pollock's abundance distribution over hatchdates and the effects of environmental factors and human activities (Figure 1).

$$Labundance_{t,d} = \alpha_1 + s_1(d) + [\alpha_2 + s_2(d)] \times 1_{CatchDate.group_{t,d}=2} + [\alpha_3 + s_3(d)] \times 1_{CatchDate.group_{t,d}=3}$$
$$+ [\alpha_4 + s_4(d)] \times 1_{t=1989} + \phi^{t-1990}[\alpha_5 + s_5(d)] \times 1_{t \geq 1990} + \mathrm{s}_6(\mathrm{TransCul}_{\mathrm{t,d}}) + \mathrm{s}_7(\mathrm{SSTBHatch}_{\mathrm{t,d}})$$
$$+ \mathrm{s}_8(\mathrm{SSTSurvAve}_{\mathrm{t,d}}) + \mathrm{s}_9(\mathrm{PDOBHatch}_{\mathrm{t,d}}) + \mathrm{s}_{10}(\mathrm{PDOSurvAve}_{\mathrm{t,d}}) + \varepsilon_{\mathrm{t,d}}$$

$$Labundance_{t,d} = \alpha_1 + s_1(d) + [\alpha_2 + s_2(d)] \times 1_{CatchDate.group_{t,d}=2} + [\alpha_3 + s_3(d)] \times 1_{CatchDate.group_{t,d}=3}$$
$$+ [\alpha_4 + s_4(d)] \times 1_{t=1989} + \phi^{t-1990}[\alpha_5 + s_5(d)] \times 1_{t \geq 1990} + \mathrm{s}_6(\mathrm{TransCul}_{\mathrm{t,d}}) + \mathrm{s}_7(\mathrm{SSTBHatch}_{\mathrm{t,d}})$$
$$+ \mathrm{s}_8(\mathrm{SSTSurvAve}_{\mathrm{t,d}}) + \mathrm{s}_9(\mathrm{PDOBHatch}_{\mathrm{t,d}}) + \mathrm{s}_{10}(\mathrm{PDOSurvAve}_{\mathrm{t,d}}) + \varepsilon_{\mathrm{t,d}}$$

where *TransCul* is the cumulative transport from the hatchdate to the catchdate. *SSTAnomBHatch* and *SSTAnomSurv* are the 30 days' average anomalous sea surface temperature before the hatchdate and the average anomalous sea surface temperature from the hatchdate to the catchdate respectively. Similarly defined are *WindSAnomBHatch* and *WindSAnomSurv*.

Before interpreting the fitted model, it is important to check whether the model provides good fit to the data. We do this by checking the validity of the key assumptions of model 1. The upper left QQ normal plot and the bottom left histogram plot in Figure 2 suggest that the errors are approximately normal. Several points on the lower and upper tails of the QQ normal curve depart from the linear trend, and we use black and red crosses to indicate the corresponding log abundances in Figure 1. These are potential outliers. The plot of residuals against fitted values at the upper right suggests the validity of the constant variance assumption for the errors. The bottom right plot shows that the responses have a positive linear relationship with the fitted values.

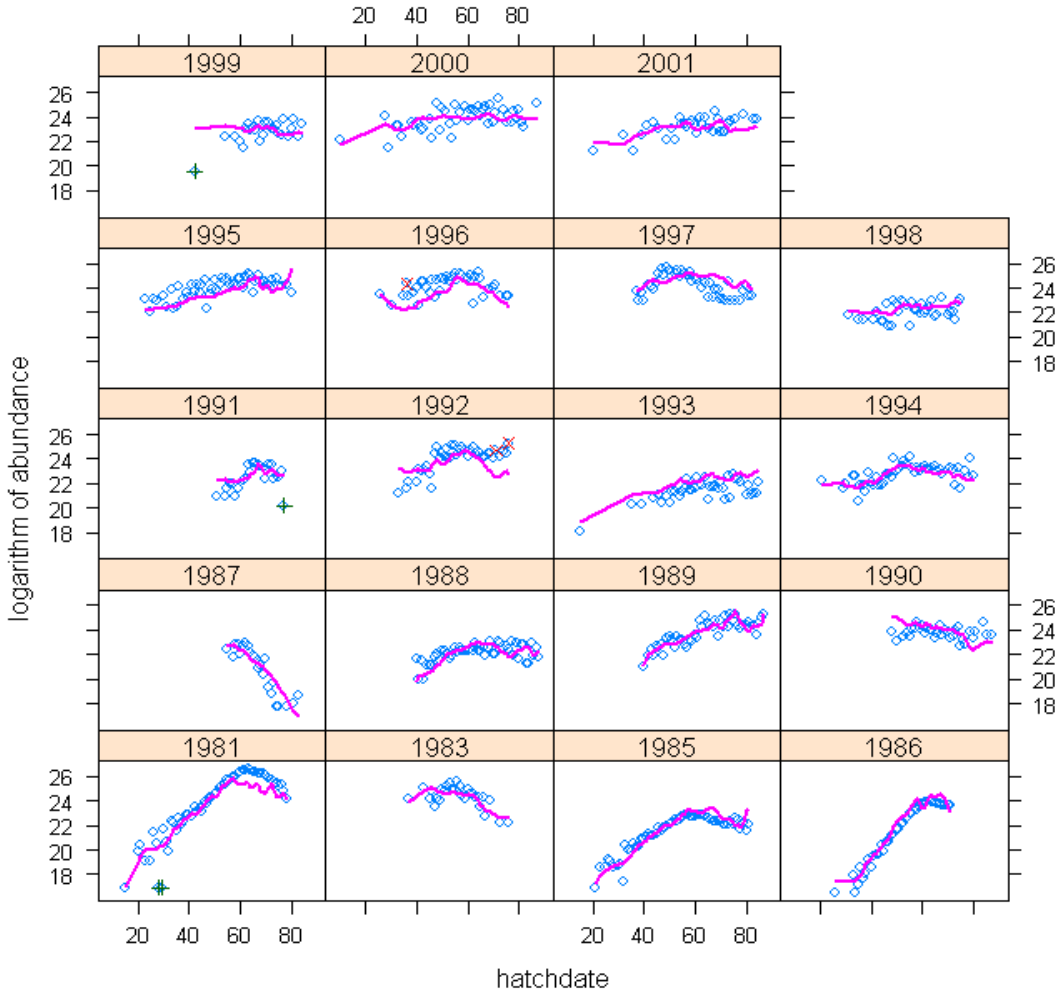

Figure 1: Logarithm of pollock abundances against hatchdates (observed data = circles) and the fitted log abundance curves derived from model 1 (lines). Potential outliers are marked as x's and crosses.

Additionally, the fitted log abundances curves over the hatchdates (solid curves in Figure 1) superimposed on the scatter plots of the observed log abundances against the hatchdates in many years over the period from 1981 to 2001 (open circles in Figure 1). However, we also find certain mismatches of the fitted curves and the observed data in some years. First, we find that model 1 underestimates the pollock's abundance in 1995 and 1996, and overestimates in 1993. Second, the fitted log abundance curves in 1992 and 1997 do not follow the data trends closely. Third, the right tail of the fitted curves in 1981 is obviously lower than the observed log abundance values. These preliminary model defects suggest that there is room for improvement, and we have several strategies in mind.

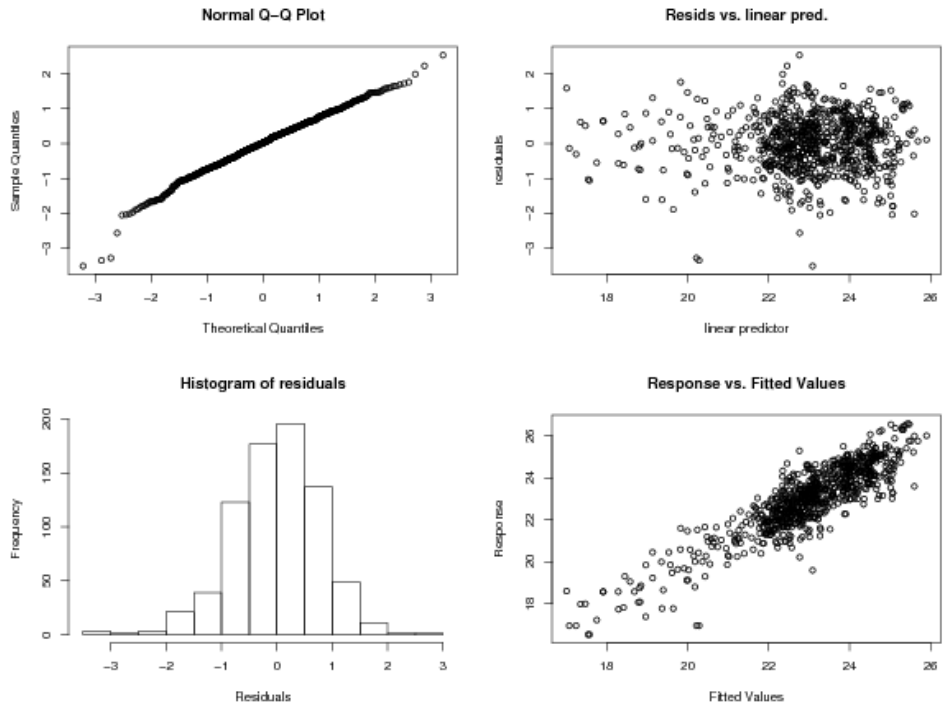

Figure 2: Residual check for the selected model.

Based on the estimates of $\eta_{a=1}$ and $\eta_{a=2}$ in the Table 3, we find that the baseline mean (log) hatching abundance of the old spawners is a little higher than that of the young spawners. We do not list the estimated η_3, since it is not significantly away from zero. This indicates that the mean (log) hatching abundance in 1989 is close to the baseline. The mean log hatching abundance increases by 1.51 (with standard error of 0.13) after 1989. Combining the estimated β_1 (0.12) with the estimated γ (0.9) from grid search, we find that the estimated baseline probability for the pollock's larvae to survive k or more days follows the curve specified as $\exp(-0.12k^{0.9})$. Such a survival curve is very close to the step-survival curve in Bailey and Macklin (1994), which is shown in Figure 3.

Table 3: Parameter estimation in the preliminary model.

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
β_0	12.71	0.36	35.39	< 2e-16 ***
$\eta_{a=1}$	6.66	0.45	14.85	< 2e-16 ***
$\eta_{a=2}$	6.05	0.23	26.05	< 2e-16 ***
η_4	1.51	0.13	11.57	< 2e-16 ***
β_1	0.12	0.017	-7.15	2.19e-12 ***

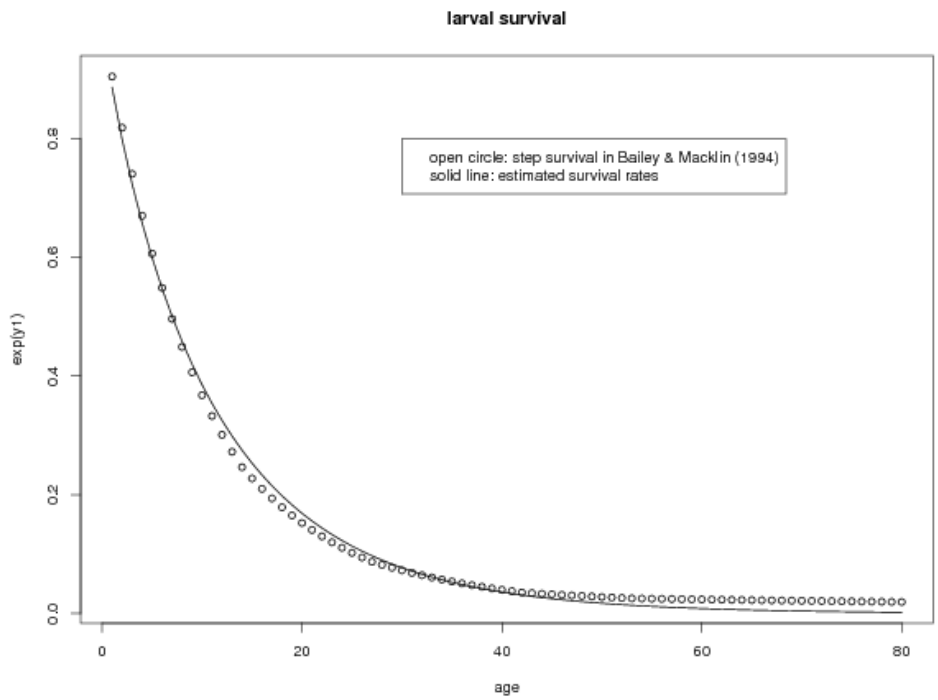

Figure 3: Survival probability of pollock larvae (age in days). Solid line is the fitted model result from our study. Circles are the values from Bailey and Macklin (1994).

The smooth function fits (Figure 4) provide useful information about the abundance distributions and the effects of the environmental and anthropological variables on pollock spawning and survival. In explaining these smooth curves, we focus on the data intensive parts of the functions.

- The smooth functions with label "hatchdate dist (Old Spawner)" and "hatchdate dist (Young Spawner)" show the baseline hatching distributions (on the log-scale) of the two spawning age groups respectively. The baseline hatching abundance of the old spawners reaches its maximal value at day 55 (April 24) in the hatching date window, and this gives the most desirable spawning period for the old spawners about 2 weeks before hatching (April 10). Additionally, the baseline hatching abundance of the young spawners is much flatter than that of the old spawners.
- The smooth function with the heading "hatchdate dist (1989)" indicates that the peak of the hatchdate distribution is a little later than the baseline. Thus the oil spill appears to have truncated the hatchdate distribution through an effect on mortality rates early in the hatching period.
- The smooth curve labeled as "hatchdate dist (after 1989)" has a "U shape", which is opposite to the hump shape of the baseline hatching distribution in "hatchdate dist (Old Spawner)". From this, we deduce that the baseline hatchdate distributions become flatter after 1989, which is consistent with the pattern displayed in Figure 1. This could be a regime shift effect. We are currently trying to determine the nature of this effect.
- The smooth curve with the labeled "transport effect" gives a shape that first increases gradually and then decreases. When the drift out of GOA (Transport) is very strong, it will flush pollock larvae out of the Shelikof Sea Valley and reduce the larval abundance there.
- The smooth curve labeled as "SST effect on hatching" shows that when the sea surface temperature before hatching increases, the larval abundance tends to decrease. This result may be explained as follows. First, the pollock's spawning process may be more protracted in relative colder environment. Second, in a colder environment, the female pollock may spawn later than usual. This brings a shorter period between the hatchdate and the specific catchdate, which means the samples in the survey need to survive a relative shorter period and consequently more larvae can be observed in the specific catchdate. We are currently trying to obtain an acceptable time

series of bottom temperatures. We have used SST here under the assumption that the water column is well-mixed prior to the spawning period. This assumption is well-supported in the literature.

- The smooth function with the title "SST effect on survival" indicates that higher the sea surface temperature results in higher survival rates for larval pollock.
- From the data intensive parts of the curves labeled as "Winds effect on hatching" and "Winds effect on survival", we induce that the strong sea surface winds tend to be related to factors that delay the pollock's spawning and reduce larval survival in the GOA.

Overall, the shapes of most smooth functions for the covariates are consistent with their expected effects on the spawning and the survival of the pollock in GOA. Anomalous relationships may occur at the ends of the distribution which tend to be data-poor.

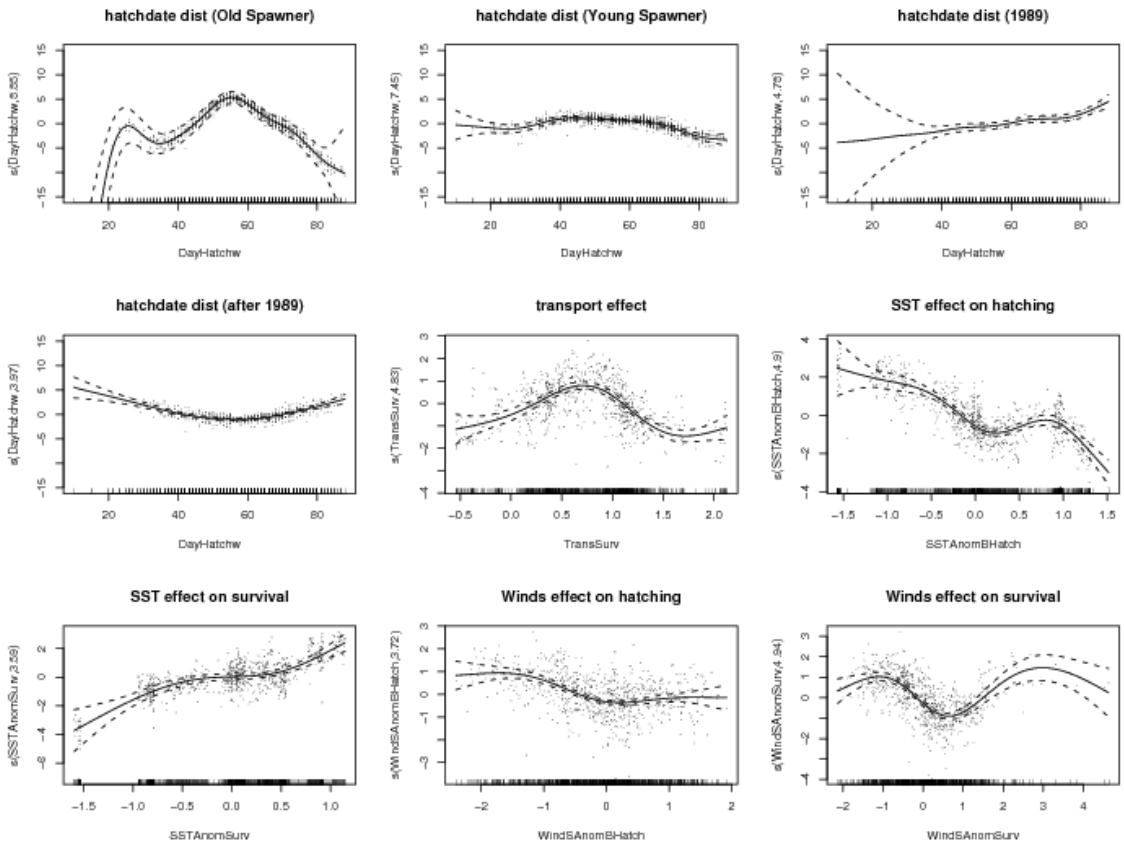

Figure 4: Fitted smooth functions for the factors in the preliminary model.

Summary

Based on our analysis of the preliminary fitted GAM model, we can infer the following conclusions. The distribution of hatchdates over time are influenced by the processes of spawning and survival. From the spawning perspective, the model depicts that the baseline shape of the spawning time distribution is hump shaped with a higher peak in the old spawner group than in the young group. Therefore the older spawners are responsible for a more concentrated period of spawning in Shelikof Strait. The shape of the hatchdates distribution was influenced by regime shift after 1988, and became flatter. Cold temperatures prior to spawning delayed the spawning activity. In the larval survival process, pollock larvae have a higher survival rate in warmer environments and strong winds and transport out of the Shelikof Sea Valley decrease the survival of pollock larvae. **T**he Exxon Valdez spill in 1989 reduced pollock abundances significantly in 1989 and truncated the hatchdate distribution.

We tentatively have identified a novel statistical method to assess recovery of habitat from disturbance. We were able to apply this method to the recovery of pollock spawning from the Exxon Valdez oil spill. The oil spill truncated the earlier part of the hatchdate distribution in 1989, and postponed the peak of the pollock's hatchdate distribution somewhat in 1990. However, the post-oil-spill effects are found to decay with an annual rate ϕ, 0.61, so that the pollock system has a rather weak memory of the oil spill. The parameter ϕ is estimated by minimizing the generalized cross validation score (GCV), with the parameter range being [0, 1]. Since the estimated $\phi = 0.61$ is a bit less than 1, the Exxon Valdez spill leaves a very weak memory in the hatchdate distributions. Quantitatively, the Exxon Valdez spill effects have a half-life of about 1.4 years (Cryer and Chan, 2008, Exhibit 11.2) in the sense that it takes about one year and five months for the oil spill effects to diminish by one half.

We believe that this method may be potentially useful to look at the temporal aspects of habitat recovery time after a disturbance. Future work will be on this aspect of recovery from disturbance, as well as refining the overall model of temporal occupation of the spatial habitat. Finally during the next year of funding we will complete the analysis of the spatial modeling of habitat occupation.