CALIFORNIA CENTRAL VALLEY CHINOOK SALMON:
A COMPARISON OF STATISTICAL FORECASTS FOR 2007 & 2008

Kyra L. Mills and William J. Sydeman

Farallon Institute for Advanced Ecosystem Research
P.O. Box 750756
Petaluma, California 94975
kmills@faralloninstitute.org, wsydeman@faralloninstitute.org

July 2008
EXECUTIVE SUMMARY

1) We forecast the Central Valley Index (CVI) for California’s Chinook salmon for 2007 and 2008 using the Pacific Fisheries Management Council’s Jack Model, and a model we developed using seabirds as environmental indicators (Bird Model).

2) The Bird Model predicted reasonably well the low CVI observed in 2007 (a modest underestimate of 14%), whereas the Jack Model overestimated the CVI in 2007 by 116%.

3) The Bird Model predicts considerably more fish (~374,000) in 2008 than the Jack Model, (~157,000).

4) In comparing observed versus predicted numbers of fish over different base periods, the Bird Model outperformed (was closer to 1:1) the Jack Model. The Bird Model appears to be more robust through environmental regime shifts (1976-1977 and 1989-1990) in the California Current.

5) The Jack and Bird Models demonstrate that Chinook Salmon returns are governed, to a large extent, by ocean conditions and not simply by the production of young salmon in streams or their survival before entering the sea.

6) Moreover, the Jack and Bird models also demonstrate that returns are determined by ocean conditions in the year before and in the year of maturation, suggesting key new temporal focal points for management-related research.

7) Ecosystem-based management of California’s Chinook Salmon fisheries could be enhanced by better knowledge and understanding of how temporal environmental and food web variability affect ocean survival of the fish.
INTRODUCTION

Adult Chinook Salmon returns (ocean catch plus escapement) to the Sacramento River watershed in 2007 were the second lowest on record. This prompted severe management actions, as well as a great deal of speculation into possible causes. In response to the unpredicted and precipitous decline, the Pacific Fisheries Management Council (PFMC) closed the fishery for the 2008 season, the most drastic management measure in the history of West Coast salmon fisheries. The socio-economic impacts of this measure will be extreme for fishers and coastal communities.

The unexpected nature of the decline remains a major conundrum for management. To address this issue, with program development support from California Sea Grant, we present a comparison of 2 methods for forecasting salmon returns to the Sacramento River watershed. The first is the statistical regression model currently used by the PFMC to forecast Chinook Salmon. This model is based on returns of precocious males at age-two (known as “jacks”) to forecast the fall run and catch (known as the “Central Valley Index”) one year later. The second involves a model we developed testing the hypothesis that seabirds can be used as indicators of the oceanic ecological conditions experienced by salmon when they are at sea. Details of this model may be found in Roth et al. (2007), and another application using the same approach using seabirds as food web indicators has been made for sockeye salmon in the Bering Sea (Sydeman et al. 2008). In particular, we focus on Cassin’s Auklets, a small planktivorous seabird that breeds on the Farallon Islands, off the coast of central California. We surmise that the predictive co-variation demonstrated by these seabird models function because the salmon and seabirds eat the same prey, and are thus dependent on the same food web. Therefore, assuming that variability in seabird breeding success reflects variability in the food web and food availability in the environment, variation in seabird breeding success may provide a proxy to salmon survival, a different but related demographic trait, in the ocean. Assuming that salmon survival at sea determined population size, and ultimately catch and escapement, the seabird-based models could be predictive. The relationship between Sacramento watershed Chinook Salmon survival at sea and food web variation is not known, but in Oregon compelling relationships between Coho Salmon and food web indicators have been established (Peterson and Schwing 2003, Peterson, unpublished data). Such direct measures of the food web for California fish are currently unavailable, so we have established the bird index as an indirect approximation.

Sacramento River fall Chinook Salmon spawning escapement has exhibited extreme variability over the past 10 years, with the highest number on record in 2002 (775,499) and the second lowest number on record in 2007 (87,966). The 2007 escapement estimate was 33% below the
predicted value of 265,000, and failed to meet the conservation escapement goal range of 122,000 to 180,000 fish for the first time in 15 years (PFMS 2008, 1). Sacramento River Fall Chinook comprise the largest stock of California’s Central Valley Chinook, and are used as an overall indicator of the health of the population.

Chinook Salmon that are caught in ocean fisheries in Oregon and California primarily originate in the Sacramento-San Joaquin River system. There are 4 different runs: late-fall, winter, spring, and fall. In 2007, the Central Valley Index (CVI), a measure of ocean harvest and escapement, was the lowest on record: 232,000 (PFMC 2008, 2). The forecast for 2007, using the “Jack Model” was 499,900, which was 2.16 times greater (116%) than observed (PFMC 2008, 2). Moreover, in 2007, 5,939 “jacks” returned to Central Valley rivers, also the lowest on record, which does not bode well for the CVI in 2008, using the PFMC method of prediction.

METHODS

The Salmon Technical Team (STT) of the PFMC has used and evaluated several predictors of the CVI since 1985, but since 1991 has used a linear regression of the previous year’s (year \(x - 1\)) Central Valley age-two precocial male returns (“jacks”) to predict the following year’s catch and escapement. From 1985-2007, the forecast has ranged from 0.49 to 2.16 times the actual value. Due to recent changes in the rate at which precocial males are returning, the STT is currently considering a different prediction method (PFMC 2008).

We used data extracted from the Preseason Reports 2000-2008 (www.pcouncil.org) to formulate the forecast for 2007 and 2008 based on the traditional “jack” and novel “bird” approaches. This included numbers of age-two fish, pre-season CVI forecasts (for the Jack Model), and post-season CVI estimates.

To compare the Jack Model with a Bird Model, we used linear regression to model the CVI based on the breeding success (number of offspring produced per pair) for the seabird, the Cassin’s Auklet. We followed the same approach as the forecasts based on jacks by modeling the CVI based on auklet breeding success in year \(x-1\) (i.e., time-lagged by one year), noted in tables and figures as “Bird \(x - 1\)”. In addition, we developed a second Bird Model by modeling the CVI as a function of breeding success in the year of return (year \(x\)) and one year earlier (year \(x-1\)). This is denoted as: “Bird \(x, x - 1\)”.

We developed models for two time periods: 1990 – year \(x\) and 1971 – year \(x\), with year \(x\) being the year previous to the forecast year that we were calculating. The shorter time period (1990 – year \(x\)) was chosen because it corresponds with the current approach used by the PFMC to
forecast salmon abundance. In addition, we chose the longer time period (1971 – year x) to assess how robust these models were over a greater range of environmental conditions, including regime shifts of 1976 - 1977 and 1989 - 1990 (McGowan et al. 2003).

Overall, we had 4 different Bird Models: 2 time series (1990 – year x and 1971 – year x), and 2 “bird series” (Bird x - 1, and Bird x, x - 1). For comparative purposes, we kept the Jack Model constant, using year x - 1 with the 2 time series (1990 – year x, and 1971 – year x).

**Forecasts for 1995-2006**

As done by PFMC, we sequentially calculated a CVI forecast for each year using the Bird Model for previous years, and starting at 1995. For the Jack Model, the 1990 – year x time series was what the PFMC/STT uses, but we calculated forecasts using the 1971 – year x time series, as these were not available.

**Forecasts for 2007 and 2008**

Since the CVI in 2007 was overestimated, we were interested in how well the Bird Model would have done to predict returns in this year as well as in 2008. To accomplish this, we used the two time series (1990-2006 and 1971-2006) for our 2007 model comparison forecasts. In addition, we used 1971-2007 and 1990-2007 to predict 2008 returns, and compared them to what the PFMC/STT has predicted based on the Jack Model.

To evaluate the “power” of the forecasts, the STT uses a simple ratio of the forecast to the post-season population estimate. Using this ratio, a perfect prediction would be 1.0, indicating that the pre-season forecast exactly matched the post-season population estimate. A number greater than 1.0 indicates model overestimation, and a number less than 1.0 represents a model underestimate. To summarize over- and under-estimates, we tabulated the number of times the forecast was greater or less than 10% of the observed. This cutoff is arbitrary, but over- and under-estimates of less than 10% probably would not result in any serious management issues for the fisheries or conservation. In addition, we used an ensemble model-averaging approach to combine the different model forecasts and compare the Jack and Bird Models over the two time series. We conducted all analyses using Stata 6.0 (Stata Corporation 1999). Note that all P-values reported in the text are approximations as we did not adjust for autocorrelation in these analyses.
RESULTS

Forecasts for 1995-2006

The coefficient of determination, $r^2$, for the longer time series, 1971- year $x$, was much lower than that for the shorter time series for both the Jack and the Bird Models (Table 1, Fig. 1 and 2). The ratio of pre-season forecast to post-season population estimates yielded underestimates (values < 1.0) for all models, except the Jack Model (1990 – year $x$), which represented a 12% overestimation overall. The model that came closest to 1.0 (using the ensemble model averaging approach for 1995-2006) was the Bird Model (1990 – year $x$), Bird $x$, x-1 (0.97). The Jack Model overestimated the CVI >10% three times, and underestimated 7 times, whereas the Bird Model overestimated (>10%) twice and underestimated 6 times (Table 1, Fig. 3). It is interesting to note that the longer time series models (1971 – year $x$), for both Jack and Bird, had a smaller number of overestimates greater than 10%, than the shorter time series models. On average, the two Bird Models, Bird x-1, and Bird $x$, x-1 underestimated the CVI by 12% and 13%, respectively. However, the Bird Model (Bird x-1) only overestimated the CVI (>10%) once, while the Bird $x$, x-1 overestimated (>10%) 3 times.

Figures 1 and 2 reflect the relationship between the CVI forecasts for the Jack and Bird Models (Bird x-1) and the CVI post-season estimate for the 1990 – year $x$ and 1971 – year $x$ time series. In both cases, the Bird Models yielded higher $r^2$ on average (0.45), than the Jack Models (0.34).

The linear trends for the two types of Bird Models (year $x$ -1 and year $x$, x-1) are illustrated in Fig. 4 and 5, corresponding with the two time series. It is evident that the shorter time series (1990 - year $x$) has a tighter fit to the post-season CVI estimate than the longer time series (1971 - year $x$). It is interesting to note that the peaks in CVI prediction for the longer time series are more “smoothed out” than for the shorter time series. Furthermore, 1995 stands out as a year where the Bird Models (for both time series) grossly underestimated the CVI: by 756,000 salmon for the shorter time series and by almost 600,000 for the longer time series (Fig. 4 and 5). The salmon adult abundance for 1995 represents the highest on record, which was not predicted based on auklet chick production for 1994 and 1995, which was above average. In contrast, both 2001 and 2002 were years of high chick productivity (well above the long-term average), which accurately predicted a high salmon abundance in 2002 (the second highest abundance since 1995), with an average ratio between the pre-season forecast and the post-season estimate of 0.96. However, in 2003, the Bird Models predicted another year of high abundance, whereas the CVI dropped by 31% (Fig. 3 and 4).
**Forecasts for 2007 and 2008**

The Jack Model predicted 499,900 fish in 2007, which was 2.16 times the actual of 232,000 (Table 2). In contrast, the Bird Model (Bird x-1, 1990-2006), forecasted a CVI of 258,100 fish, a slight overestimate (1.11) of the actual value. Using the longer time period (1971-2006), the Bird Model (Bird x -1) forecasted 271,100 fish, 1.17 times the actual, and the Jack Model forecasted 531,600, 2.29 times the actual (Fig. 3). Using the second Bird Model x, x-1 did not improve the 2007 forecast for either the shorter time period (1990-2006) or the longer time period (1971-2006; Table 2). The average ratio between the pre-season forecast and the post-season actual value for the Jack Model was 2.3 (based on 2 models) and for the Bird Model was 1.15 (based on 4 models).

Table 3 summarizes the 2008 CVI forecasts, using the 4 different Bird Models and the 4 Jack Models. The “official” PFMC forecast for 2008, based on the Jack Model (1990-2007) is 157,100. However, this forecast excludes 2005 because it is believed to be an outlier and presents “excessive leverage on the resulting predictor” (Appendix D, PFMC 2008, 2). The 2008 CVI forecast using the Jack Model and including the 2005 data point is 347,500, and using the “new” jack method the forecast is 363,500 adult salmon (Table 3). In 2007, the auklet’s breeding success improved, but remained 50% below the long-term average (Warzybok and Bradley 2007, Goerick et al. 2008). Using the 1990-2007 time series, the Bird Models predict a 2008 return of 505,300 fish. Using an ensemble model-averaging approach, the Bird Models predict a return of 374,200 fish for 2008, whereas the Jack Model predicts 58% fewer salmon, using the “official” prediction.

**CONCLUSIONS**

Overall, for the forecasts made for 1995-2007, the number of overestimates (>10%) of the Jack and the Bird Models was similar (Jack: 3; Bird: 2.75, average), and the number of times the forecast was underestimated (<10%) was also similar (Jack: 7.5; Bird, 6; Table 1). The four years that really stand out as large overestimates for the Jack Model are 1998, 2005, 2006, and 2007 (Fig. 3). Thus, it appears that when ocean conditions are below average, and when Cassin’s Auklet productivity is low, the Jack Model tends to overestimate the CVI. For example, in 2005, the Jack Model projected a CVI return of 1,678,000 fish, and the post-season estimate was only 850,000 fish, representing less than half of the number that had been predicted. In 2007, the pre-season forecast was over twice that of the post-season estimate.

For the Bird Models, the only year that really stands out as an overestimate is 2003, when the forecast for the 1990 – year x Bird Models was 1.5 times the post-season estimate (Fig. 3). The
Bird Models had larger underestimates of the CVI, whereas the Jack Model had larger overestimates. The years that stand out as underestimates of the CVI for the Bird Model, for both the 1990 – year x and 1971 – year x time series, are 2005 and 2006 (Fig. 4 and 5), the only two years on record when auklet productivity was zero. This implies that even though auklets were unable to raise chicks those two years because of poor ocean conditions, salmon returns were not as bad as predicted from the models alone. It is interesting to note that the 1995 predictions were underestimated for all models (Figs. 3, 4, 5). Even though auklets produced an above average number of chicks, 1995 was a year of high salmon returns and the highest on record.

The Bird Models outperformed the Jack Models for both time series; the Bird Models had an average $r^2$ 31% higher than the Jack Model for the 1990-year x time series, and 33% higher than the Jack Model for the 1971-year x time series. The Bird Models perform better than the Jack Models through the longer time period. However, for both the Bird and Jack Models, the ones that performed the best for predicting CVI were the ones using the shorter time series (1990 - year x; Table 1). These models had a higher average $r^2$ (52% greater) than the 1971 – year x models. Between 1971 and 2007 there have been two environmental regime shifts of the California Current (1976-1977 and 1989-1990), which may explain why the shorter time series (1990 - year x) models are better at predicting the CVI.

In comparing observed versus predicted CVI abundance over the various time periods, the Jack Model overestimated the CVI for the shorter time series for the 1990-2006 predictions, whereas both models underestimated for the longer time series for the 1990-2006 predictions. For the 2007 CVI predictions, the Jack Models and the Bird Models overestimated when using both time series, although all 4 Bird Models had a smaller overestimation than the Jack Models.

The PFMC is currently considering a change in its forecast method because of certain limitations of the current method. The new forecasting method will restrict the number of “jacks” used in the forecast to those found in the Sacramento River only. Using this new jack method (described in Appendix D, PFMC 2008, 3), the PFMC predicts a return of 363,500 salmon in 2008 (Allen Grover/CDFG, pers. comm.). However, the “official” prediction for 2008 is 157,100, a conservative prediction and the lowest Central Valley Index forecast on record (PFMC 2008, 2). It is therefore expected that the 2008 adult escapement for the Sacramento River fall Chinook will be below the lower end of the escapement goal of 122,000 to 180,000 adults (PFMC 2008, 2).

The Jack and Bird Models demonstrate that Chinook Salmon returns are governed, to a large extent, by ocean conditions and not simply by the production of young salmon in streams or their survival before entering the sea. Chinook Salmon spend most of their life at sea, and yet
there is surprisingly little information regarding this prolonged stage of their life. Auklet chick production is relatively easy to monitor and is under the influence of ocean conditions. Seabirds and salmon (while in their ocean phase) encounter similar variability in the marine environment and therefore may exhibit similar trends in reproduction (seabirds) and survival (salmon). As has already been shown elsewhere (i.e., Roth et al. 2007, Sydeman et al. 2008), there is significant co-variation between certain seabird parameters, such as timing of breeding and breeding success, and returning salmon. Auklets are sensitive indicators of zooplankton, primarily krill availability, and other oceanographic conditions (Abraham and Sydeman 2004, Jahncke et al. 2008, Wells et al. in press). In years of poor ocean productivity, auklets have poor breeding success. In 2005 and 2006, the auklets experienced complete breeding failure, unprecedented in 3 decades of study (Sydeman et al. 2006, Peterson et al. 2006), indicating exceptionally poor food availability. We hypothesized that krill were either less abundant or distributed inappropriately for the auklets to forage successfully, leading to reproductive failure.

The Jack and Bird Models demonstrate that salmon returns are determined by ocean conditions in the year before and in the year of maturation. When year x (the year of maturation) is included in both models, the fit of the models is improved. Because so little information exists regarding the ocean phase of the salmon, the information gleaned from this study suggests key new temporal focal points for management-related research. Ecosystem-based management of California’s Chinook Salmon fisheries could be enhanced by better knowledge and understanding of how temporal environmental and food web variability affect ocean survival of these fish. By including these Bird Models in Chinook Salmon management, the gap is lessened between ocean ecosystem and its effects on adult salmon numbers.

ACKNOWLEDGEMENTS

Funding for this project was provided by California Sea Grant Program Development Project No. R/FISH-146PD under NOAA Grant NA040AR4170038 issued to University of California. We thank David Manning (SCWA) and Brian Wells (NOAA/NMFS) for comments on previous versions of this report.
LITERATURE CITED


Table 1. Central Valley Index abundance forecasts (1995-2007) using the two time series models (1990 – year x and 1971 – year x) and the 2 bird series, Bird x-1 and Bird x, x-1. The $r^2$ is the average of the individual $r^2$ for the 1995-2007 forecasts. The pre-season/post-season estimate ratio is the average of the individual ratios for the years 1995-2007, and the number of overestimates and number of underestimates represents the number of times the pre-season/post-season ratio was greater than or less than 10%.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave. $r^2$</td>
<td>0.46</td>
<td>0.22</td>
<td>0.63</td>
<td>0.26</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>Ave. pre/post-season</td>
<td>1.12</td>
<td>0.92</td>
<td>0.95</td>
<td>0.84</td>
<td>0.97</td>
<td>0.81</td>
</tr>
<tr>
<td>No. Overestimates</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>No. Underestimates</td>
<td>6</td>
<td>9</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2. Central Valley Index abundance forecast (in thousands) for 2007 based on the Jack Models and the Bird Models using two time-series, 1990-2006 and 1971-2006 and the 2 bird series, Bird x-1 and Bird x, x-1. Pre/Post-season is the ratio between the forecast and the post-season CVI estimate.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CVI Forecast</td>
<td>499.9</td>
<td>531.6</td>
<td>258.1</td>
<td>271.1</td>
<td>263.5</td>
<td>273.2</td>
</tr>
<tr>
<td>Pre/Post-season</td>
<td>2.16</td>
<td>2.29</td>
<td>1.11</td>
<td>1.17</td>
<td>1.14</td>
<td>1.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CVI Forecast</td>
<td>347.5</td>
<td>157.1</td>
<td>456.3</td>
<td>363.5</td>
<td>505.3</td>
<td>465.2</td>
<td>262.7</td>
</tr>
<tr>
<td>r²</td>
<td>0.48</td>
<td>0.71</td>
<td>0.23</td>
<td>0.45</td>
<td>0.69</td>
<td>0.39</td>
<td>0.69</td>
</tr>
<tr>
<td>p-value</td>
<td>0.001</td>
<td>&lt; 0.001</td>
<td>0.003</td>
<td>0.002</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.005</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

**Fig. 1** – Relationship between CVI forecast and post-season estimates (1995-2007) for Jack and Bird Models, using 1990 – year x models time-lagged by one year (year x-1).

**Fig. 2** - Relationship between CVI forecast and post-season estimates (1995-2007) for Jack and Bird Model, using 1971 – year x models time-lagged by one year (year x-1).

**Fig. 3** – CVI forecasts for 1995-2007 for all models, represented as the deviation from 1 for the ratio forecast/post-season estimate. Positive values represent overestimates, and negative values represent underestimates.

**Fig. 4**. Comparison of Bird Model, Bird x-1 and Bird x, x-1 for the 1990 – year x time series.

**Fig. 5**. Comparison of Bird Model, Bird x-1 and Bird x, x-1 for the 1971 – year x time series.
Fig. 1
Fig. 2

Bird: $r^2 = 0.26$
$p = 0.004$

Jack: $r^2 = 0.22$
$p = 0.2$

- Diamond: Jack Model (1971 - year x)
- Square: Bird Model (1971 - year x)
Fig. 3
Fig. 4
Fig. 5