Inseason Forecasts of Sockeye Salmon Returns to the Bristol Bay Districts of Alaska

Saang-Yoon Hyun

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Washington

2002

Program Authorized to Offer Degree: Interdisciplinary Program in Quantitative Ecology and Resource Management

University of Washington

Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Saang-Yoon Hyun

and have found that it is complete and satisfactory in all respects, and that any and all revisions required by the final examining committee have been made.

Co-Chairs of Supervisory Committee:

____

Ray W. Hilborn

____

James J. Anderson

Reading Committee:

____

Ray W. Hilborn

____

James J. Anderson

____

Katherine W. Myers

Date: ______________________________

University of Washington

Abstract

Inseason Forecasts of Sockeye Salmon Returns to the Bristol Bay Districts of Alaska

by Saang-Yoon Hyun

Co-Chairpersons of the Supervisory Committee:
Prof. Ray Hilborn & Prof. James J. Anderson
Quantitative Ecology and Resource Management

The Bristol Bay sockeye salmon fishery has been the most valuable salmon fishery in North America, and provides season employment for several thousand workers. The fishery consists of five reasonably discrete fishing districts corresponding to watersheds where the salmon are returning to spawn. The long term objective of management is to achieve Maximum Sustained Yield from the fishery, and this is implemented on an annual basis by regulating the time allowed for fishing to allow a predetermined number of fish to pass the fishery and make it to their natal streams and lakes to spawn.

The expected total return of fish to each district is an important part of the fishery management and is equally important to the fishing fleet and the fish processors. I developed a statistical model for inseason run size prediction that uses data from (1) a test fishery at Pt. Moller, (2) the age composition of the catch at Pt. Moller, (3) the total return to date by district and (4) the age composition of the return to each district. Optimization and Bayesian methods are used to obtain both point estimates and distributions of estimates. I found that the temporal pattern in catches at Pt. Moller explained 59% of the variation in run timing in the fishing districts. Using the preseason forecast as a prior significantly improved the performance of the estimation during the initial stage of the season. This method provides a consistent way to incorporate diverse forms of data in a single unified statistical framework that should provide a significant improvement in inseason run forecasting. The method was tested using data from 1999 and found to perform well. In terms of the absolute values of relative errors of forecasts of the returns to the main districts (Kvichak-Naknek, Egegik, and Nushagak) made on

June 24, June 29, and July 4, the mean values were 86.7%, 72.4%, and 59.9% when preseason forecasts were not incorporated, whereas they were 27.6%, 25.4%, and 20.9% when preseason forecasts were incorporated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It has been an adventure to work in the Quantitative Ecology and Resource Management (QERM) program at the University of Washington (UW). I thank God and the people who have helped me reach the finish line. I feel honored to have excellent advisors: Ray Hilborn and James Anderson. Hilborn provided the original ideas for this dissertation. Hilborn's guidance was critical to the development and completion of my dissertation. Anderson was my M.S. thesis advisor as well. Anderson was flexible, helpful, and supportive. I also thank my other committee members: Kate Myers and Steve West. Myers' editing improved my writing. West was my graduate school representative.

Financial support for this project was mainly from two sources: Research Assistantship with the Highseas project, and Teaching Assistantships through the QERM program and the Center for Quantitative Science (CQS). I express my deep thanks to Highseas project scientists (Kate Myers, Trey Walker, and Nancy Davis) and the QERM and CQS faculty for their support. I especially appreciate the thoughtful supervision of Loveday Conquest, co-director of QERM. Ashley Steel of the Northwest Fisheries Science Center (NWFSC) provided one-quarter Research Assistantship.

One of the hardest parts in my thesis was programming in optimization software, ADMB. Billy Ernst and James Ianelli helped me with the programming. Without their help, I could not have completed this thesis work. I highly acknowledge Ernst's versatile characteristics; he is a carpenter, a sailor, and good company as well as an intelligent scientist. UW Alaska Salmon Program (ASP) and the Alaska Department of Fish and Game (ADFG) provided data. My special thanks go to Warner Lew, a former UW ASP staff.

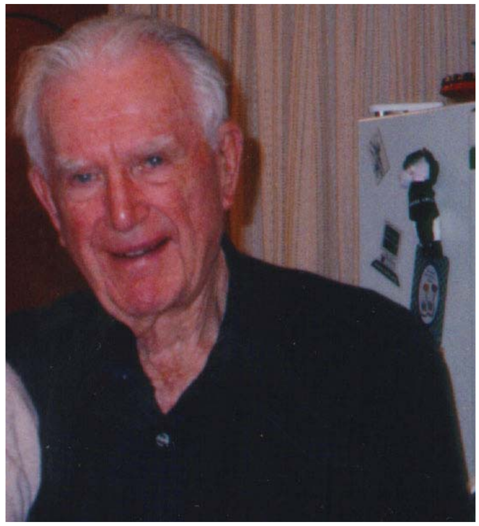
I was fortunate to meet many good people while I was in the graduate school. The International Rotary Foundation, which provided the scholarship for my first year M.S. work, assigned George Harry Jr as my Rotary counselor. His helpfulness and interest not only in my academic progress but also in my personal life were truly supportive. I sincerely and deeply respect him for his academic intelligence, social activity, and rationality. Others to whom I am also indebted are James Norris, Eric Anderson, Beth Babcock, Vincent Gallucci, Steve Mathews, Owen Hamel, Joel Reynolds, Chris Foote, Kevin Brinck, Ray Timm, Kris Ryding, and friends in QERM and Fisheries at the UW. The advertisement poster that Carolina Minte-Vera and Billy Ernst made for my final defense was surprising, beautiful, and entertaining. I will keep the poster forever.

One of the valuable lessons that I have learned during my graduate studies is cooperation: *to help and to be helped*. I have been helped much by many people. I would be glad to help those who need my knowledge.

It took several years for me to work for my M.S. in Fisheries and my Ph.D. in QERM at the UW. While I have worked for my graduate studies in the United States, my family in Korea gained new members. My sister gave birth to two children, letting me become an uncle for them, and my brother married recently. I wished to see my nephew and niece often, and I was sorry not to attend my brother's wedding. I thank my family for having been healthy and cheerful. It was not easy to have lived alone in a foreign country, but the time I spent in Seattle has been one of my best times in my life.

DEDICATION

To my special friend, George Harry Jr



PROFILE OF GEORGE HARRY JR

Academic record

B.S. Oregon State University, 1940, fisheries.

M.S. University of Michigan, 1941, zoology.

Ph.D. University of Washington, 1956, fisheries.

Work record

1941-1946. Sea duty as naval officer.

1947-1954. Commercial fisheries biologist. Oregon Fish Commission, Astoria.

1954-1958. Director, Fisheries research and management of all Oregon Fish Commission projects, including river fisheries and hatcheries. Portland.

1958-1967. Director, National Marine Fisheries Auke Bay Laboratory, Juneau, Alaska.

1967-1970. Director, NMFS research laboratory, Ann Arbor, Michigan.

1970-1980. Director, National Marine Mammalogy laboratory, Seattle.

Retired-1980.

CHAPTER I. INTRODUCTION

1.1. RESEARCH MOTIVATION AND OBJECTIVES

There are some common goals in managing anadromous fish like Pacific salmon (genus, *Oncorhynchus*) in the northwest. The goals are to hit an escapement goal of homing fish, to conserve inherited characteristics of a stock, and to maximize the harvest of surplus fish. Surplus fish mean fish left after subtracting an escapement from a run. The first and second goals concern biological conservation of the management stock in terms of not only abundance but also genetic characteristics. The third goal concerns economic benefit. To achieve these goals, managers need to know in advance the run size and timing of homing fish. Generally there are two kinds of forecasts for anadromous fish management: a preseason forecast and an inseason forecast. A preseason forecast is made before homing fish start to arrive at a local management area. The preseason forecast information is mainly used for fish buyers and processors such as canneries. However the preseason forecast information is usually not accurate enough to be used for management. The managers rely on an inseason forecast to achieve the three goals. Once fish reach a local management area, managers start to monitor the run and collect data through a test fishery. On the basis of these data, an inseason forecast is made. As the data are accumulated, the inseason forecast is updated periodically to improve the estimates of run size and timing. This inseason forecast helps managers regulate fisheries that target the homing fish. The regulations include opening or closing a fishery in a certain area during a certain time.

Sockeye salmon (*O. nerka*) of Bristol Bay, Alaska are also managed with the same goals as those described above. Bristol Bay is located in the southeastern Bering Sea and is surrounded by five estuaries (Figures 1.1 and 1.2). There are mainly eight stocks that compose the Bristol Bay sockeye salmon run. A problem in managing the Bristol Bay sockeye salmon is derived from difference in escapement goal and run timing between stocks. To conserve run timing profiles of stocks, optimal number of spawners should be allowed to reach their spawning grounds over the entire season period. In other words, fishing activity should be properly distributed over the season. Because of these stock-specific goals, it is not a good idea to apply the same fishing effort to all stocks during the season. The current fisheries that target the Bristol Bay sockeye salmon are not allowed to occur in the ocean beyond the five estuaries. But the estuary-specific fishery does not guarantee a stock-specific exploitation because some stocks have to pass a common estuary. A drift gillnet, which is the legal fishing gear in Bristol Bay, is size-selective but cannot be stock-selective unless the fish body size significantly differs between stocks. Another factor that makes the stock-specific management difficult is the short time window of the salmon run. Because about 80% of sockeye salmon usually migrate through Bristol Bay within two weeks, the Alaska Department of Fish and Game (ADFG) must make quick decisions about fishery regulations.

Even though inseason forecasts were first made in 1968 (Eggers and Fried 1984, Helton 1991, Rogers 1994), there is no systematic algorithm for estimating stock-specific runs during the season. The main objective of my research is to estimate stock-specific run sizes on a daily basis during the season. The data are mainly catch and age composition from three fishery sources: an offshore test fishery, estuary fisheries, and escapement fisheries. As the season progresses, the data are updated, and estimates of stock-specific runs are improved. Because of the short duration of the salmon run, 'daily' estimates of stock-specific runs are desired. This information will help ADFG managers to decide inseason regulations to achieve the three management goals. Specific objectives of this research are to develop a computer algorithm for inseason forecasts of returns, and to implement the algorithm into software to be used by managers.

1.2. LITERATURE REVIEW

1.2.1. General features of sockeye salmon

Pacific sockeye salmon are also called red salmon (Alaska), blueback salmon (Columbia River), nerka and krasnaya ryba (Russia), benizake and benimasu (Japan) (Burgner 1991). Sockeye salmon are anadromous like other fish of the genus *Oncorhynchus*, but some sockeye salmon populations called 'kokanee' spend their entire life only in fresh water. Another important feature of sockeye salmon is the semelparous life history; the fish spawn only once during their entire life and then die.

The primary spawning grounds of sockeye salmon range from 47^0N to 63^0N (Burgner 1991). The grounds in North America extend from tributaries of the Columbia River to the Kuskokwim River in western Alaska, and those of Asia are distributed on the Kamchatka Peninsula, Russia.

Mature sockeye salmon return to their natal stream mainly during June through September, and spawn and fertilize eggs before they die. The sockeye eggs in stream gravel develop during September through January, and the sockeye alevin emerge from the natal gravel during January through April (Burgner 1991, Pearcy 1992). The main characteristic of the alevin stage is the presence of ‘yolk.’ Once the fish yolk disappears, the fish stage is called ‘fry.’ The sockeye fry migrate to lakes during May through June. Sockeye salmon require a ‘lake’ rearing environment for the juveniles. This requirement is a distinction of sockeye salmon, which is different from some fish of the genus, *Oncorhynchus*. Chinook (*O. tshawytscha*) and coho salmon (*O. kisutch*) utilize ‘stream’ rearing environments as juveniles. The residence time of the sockeye juvenile fish in lake ranges from one year to three years. When the juvenile fish are ready for ocean life phase by undergoing a series of physiological, behavioral and biochemical changes (Hoar 1976), they are called ‘smolts.’ The sockeye smolts migrate to the ocean during June through July. Their ocean residence time, during which they are maturing, ranges from one year to three years. Ocean growth of the immature and mature sockeye salmon continues while they reside in the ocean (French et al. 1976, Burgner 1991). The ocean distribution of sockeye salmon occurs mainly in the north Pacific Ocean, covering a latitudinal range from 40^0N through 65^0N and a longitudinal range from 150^0E through 125^0W.

1.2.2. Bristol Bay sockeye salmon

Bristol Bay

Bristol Bay of Alaska is located at the southeastern area of Bering Sea and the co-ordinates of the bay center are 58^0N 159^0W (Figures 1.1 and 1.2). The bay is surrounded mainly by five estuaries: Togiak, Nushagak, Kvichak-Naknek, Egegik, and Ugashik (Figure 1.2). As a term of the management unit, these estuaries are often called 'districts.' Districts of Nushagak and Kvichak-Naknek are connected to a few rivers. The flow of Nushagak district is contributed mainly by the Igushik River, the Wood River, and the Nushagak River, and that of Kvichak-Naknek district is contributed mainly by the Kvichak River, the Branch River, and the Naknek River (Figure 1.2). Thus, Bristol Bay has mainly nine river systems: Togiak, Igushik, Wood, Nushagak, Kvichak, Branch, Naknek, Egegik, and Ugashik. Iliamna Lake, the largest of the sockeye-producing lakes in the world (2,622 km^2) is connected to the Kvichak River (Burgner 1991).

Life history of the Bristol Bay sockeye salmon

The Bristol Bay sockeye salmon begin their ocean life phase when they enter the bay. Sockeye smolts enter the bay mainly during late May through June. Despite that the five estuaries of the bay are radially located, the juveniles from those estuaries choose the coastal waters along the southeast side of inner and outer Bristol Bay during their seaward migration (Straty 1974). At the juvenile stage in Bristol Bay, the main food item of the juveniles is zooplankton. By September, substantial numbers of Bristol Bay sockeye juveniles are still only 460-560 km from their estuaries of origin. They tend to remain within about 100 km of the shore during their feeding and migration movement. During the first fall and winter in the marine environment, their migration direction is variable: to the middle of the Bering Sea and to the southward through the Aleutian passes. The main water source of the Bering Sea is the Alaskan Stream, an extension of the Alaskan Current of the Alaskan Gyre (Verkhunov 1995). Stomachs of the juveniles sampled in the Bering Sea included other food items beside zooplankton: larval capelin

(*Mallotus villosus*), sand lance (*Ammodytes hexapterus*), and herring (*Clupea harengus pallasi*) (Foerster 1968, French et al 1976).

The ocean distribution of the Bristol Bay sockeye salmon during their maturing stage ranges from about 46^{o}N in the central North Pacific Ocean to about 64^{o}N in the Bering Sea and from about 175^{o}E to about 145^{o}W of the Gulf of Alaska. The limiting factor of the distribution seems to be water temperature (Burgner 1991, Pearcy 1992).

The return time of the Bristol Bay mature sockeye salmon is almost simultaneous except the Ugashik and Togiak stocks that consistently return a few days later (Figure 1.3). The run duration is very short; it happens mainly during one month from about the middle of June through the middle of July. About 80% of the returns occur only within a two-week period despite their diverse distribution at sea (Burgner 1980). By summer, the Bristol Bay adult sockeye salmon are in much lower abundance in the high-seas as their inshore returns progress. Four age groups account for about 98% of all Bristol Bay returns: 1.2, 2.2, 1.3 and 2.3[1] (Fried et al. 1988), even though the proportions of age groups can vary each year.

In stomachs of the homing adults caught in the basin area of the central Bering Sea, food items were found to be more varied and included squid, fish larvae, amphipods, and euphausiids, whereas in the shelf area to the east, the items were almost exclusively euphausiids, with a small proportion of fish larvae, including walleye Pollock (Nishiyama 1974, 1984). The average stomach content volume was greater in the sockeye salmon sampled from the shelf area, which appeared to coincide with the general trend of zooplankton biomass distribution between the two areas (the central Bering Sea and the shelf area). The caloric value per unit weight of food consumed was also greater in the shelf area. This coincidence in the stomach content volume and the zooplankton biomass supports the idea that salmon are opportunistic feeders (Pearcy 1992).

---

[1] This salmon age is expressed as its European way (Koo 1962). A fish of age 'x'.'y' spent 'x' winter(s) in freshwater after fry stage and 'y' winter(s) in the ocean.

Run characteristics

One of the most remarkable characteristics of the Bristol Bay sockeye salmon run is high annual variability in abundance. Dealing with the returns of year 1958 through 2001, the coefficient of variation (CV)[2] was 0.60 (Figure 1.4). As an extreme example, the return size of 1995 (60,488,000 fish) was almost 27 times as many as that of 1973 (2,245,000 fish).

Another characteristic of the returns is a cyclic pattern (Figure 1.4). This cyclic pattern is mostly due to returns of sockeye salmon to the Kvichak River (Figure 1.5). The Kvichak stock returns have cycled with four or five year periodicity. Because the data of only the Kvichak stock returns are not available, I show the returns to the Kvichak-Naknek district in Figure 1.5. The Kvichak stock generally returns as the highest abundance among the Bristol Bay river stocks except its off-peak years.

Some literature suggests a possible mechanism for the cyclic pattern in the annual return size (Mathews 1967, Eggers and Rogers 1987). Mathews (1967) postulated an interaction between spawning populations of successive years. The postulate is that a large spawning population might change some controlling environmental factor such as food organisms or the intra-gravel environment, and this change might be detrimental to the production of sockeye salmon in the ensuing two years. With the assumption, Mathews (1967) modified the deterministic Ricker model between spawners and recruits into a stochastic version by incorporating an error term in calculating recruits. Besides, the modified model has spawners of the past two years as well as the current year in calculating the resultant recruits. By simulation with this modified version, Mathews (1967) succeeded in producing the cyclic pattern in returns. However, the real data used in the simulated model were limited to those of just seven years, and the values for parameters in the model were arbitrarily chosen.

In addition to the suppression of production following large escapement, the sockeye salmon fishery also seems to have been responsible for the cyclic pattern. The early fishery (before adoption of formal escapement goals) was limited by processing

---

[2] CV = standard deviation / mean

capacity and only loosely regulated for a fifty percent exploitation rate (Eggers and Rogers 1987). During the peak cycle year runs, fish mortality by this fishery tended to be much lower than average. With this reasoning, Eggers and Rogers (1987) called the mechanism 'depensatory fishing.' A 'depensatory' mechanism is defined as a relationship where mortality of a population decreases as the population abundance increases.

1.2.3. Forecasts of the Bristol Bay sockeye salmon returns

Bristol Bay sockeye salmon compose over 50% of the sockeye salmon harvested in North America (Fredin 1980, Rogers 1986). Because of this high productivity, Bristol Bay sockeye salmon are an important economic source in the northwest. From year 1958 through 2001, the annual average catch of the Bristol Bay sockeye salmon was 16.4 million (Figure 1.4). In managing this valuable population, forecasts of returns are a critical part. Recalling the high variability in annual returns (Figure 1.4) and the short duration in return time, accurate forecasts are strongly desired by managers and the fishing industry.

The first forecasts were made by UW FRI (University of Washington Fisheries Research Institute) in about 1950 (Rogers 1998). About 1962, the Alaska Department of Fish and Game (ADFG) started to participate in forecasting the annual runs from inshore observations (escapements, smolts, and age composition) and in 1984, salmon processors asked UW FRI to make forecasts from these data to provide a second opinion.

Preseason forecasts

At present, preseason forecasts of sockeye salmon returns to Bristol Bay are made by both ADFG and UW FRI. Salmon buyers and processors such as canneries use preseason forecasts to determine staff and equipment needed for production of fresh, frozen, and canned products and to plan deployment of tenders and processing vessels (Fried and Yuen 1987). For the industry, a forecast is most useful when available well in advance of the run (at least six months before the run). ADFG also uses preseason forecasts to set a quota for a commercial fishery at False Pass (Hilborn, *Personal communication*); 8% of the forecast run are allocated for the False Pass fishery (Figure

1.1). Run predictions are made for each major age group (usually four ages: 1.2, 2.2, 1.3, 2.3) and summed to obtain a forecast for a river system. Then the river system forecasts are summed to predict the run to a fishing district, and the predicted catch is obtained by subtracting the recent five year average of escapements from the district run.

From 1987 to 1996, the ADFG forecast of the Bristol Bay sockeye salmon run differed from the actual run by an average of 27% (range: 9-56%), and the UW FRI forecast differed by an average of 22% (range: 5-43%) (Rogers 1998). However, in case of forecasts of the 1997 and 1998 runs, the forecasts by both agencies differed from the actual runs by about 100%. The actual runs of 1997 and 1998 turned out to be far smaller than the forecasts. For example, the UW FRI preseason forecasts of the 1997 and 1998 runs were 35.1 millions and 33.8 millions but the actual runs of those years were 18.9 millions and 18.3 millions, respectively. This serious discrepancy between forecasts and actual runs left unreliable the traditional forecast methods that ADFG and UW FRI have used.

I briefly describe the methods of preseason forecasts by ADFG and UW FRI. ADFG uses mainly two ways to forecast individual river system stocks by major age group. The first method is to use spawner-recruit data and its forecasts are calculated through a linear form of the Ricker model (Brannian et al. 1982).

$$\ln(R_{a,s,y} / E_{s,y}) = \ln(a) - b \cdot E_{s,y}$$

where $R_{a,s,y}$: the number of age a fish returning to river system s from spawning during brood year y; $E_{s,y}$: the number of spawners in river system s during brood year y; a and b are parameters.

The second method is to use sibling and smolt data and its forecasts are estimated through a linear form suggest by Peterman (1981, 1982a, b).

$$\ln(R_{a,s,y}) = a + b \cdot \ln(S_{a-1,s,y})$$

where $S_{a\text{-}1,s,y}$: the number of age *a-1* smolts produced by brood year y and migrating seaward from river system s. Forecasts using smolt data are possible only from river systems that have smolt enumerating programs. Smolt enumerating programs were started in Kvichak River system in 1971, Wood River system in 1975, Naknek and Egegik River systems in 1982, and Ugashik in 1983, respectively.

In UW FRI, preseason forecasts have been made traditionally by Rogers since 1985, but the recent forecast of the 2000 run was also made by Hilborn. The traditional methods by Rogers depend on relationships between numbers of fish in a run and estimates of the numbers of fish at earlier times in their life (e.g. the approaching run, immature fish at sea, seaward migrant smolt, fry in lakes, or the number of spawners) (Rogers 1994). By regression models with these variables, Rogers predicts fish return by river system and by age.

Hilborn et al (1999) use mainly four data sources to predict returns: (1) jack returns, (2) sibling returns, (3) spawners, and (4) the past year returns. The return of jacks usually provides a good prediction of the next year's return of 2-ocean fish. And there often exists a strong relationship between the return of 2-ocean fish and subsequent return of 3-ocean fish from the same cohorts. These relationships with jacks return and sibling returns offer a basis in predicting returns by regression models. However, these regression analyses of Rogers and Hilborn do not incorporate an unexpected change in salmon ecosystem. To avoid the serious failure of the forecasts of the 1997 and 1998 runs, Hilborn checks the historical pattern in recruits per spawner and the total return by brood year. He suggests alternative run forecasts by simply averaging the recruits per spawner and the recent past runs over different time horizons.

Inseason forecast

Inseason forecast of sockeye salmon return to Bristol Bay is useful mainly to three entities: (1) ADFG managers, (2) the commercial processing industry, and (3) fishermen. Managers need an idea of the run size to determine when to allow commercial fishing. The industry processors use the inseason forecast to decide how many tenders to employ and how many floating processors to send to the bay. And based on the inseason forecast, fishermen decide whether it is worth gearing up for the fishing season (Hilborn et al. 1999).

The inseason forecast of the return is initially based on catch per unit effort (CPUE) of a test fishery that occurs offshore from Port Moller, Alaska during the salmon return season (Figure 1.1). And the return size estimate is updated also by commercial catch reporting and spawning escapement monitoring every day of the season. The inseason forecast project by the Port Moller test fishery had been operated by ADFG from 1968 to 1985, but it has been taken by UW FRI since 1987 (Eggers and Fried 1984, Helton 1991, Rogers 1994). These inseason forecasts have provided more accurate predictions than preseason forecasts because the relative abundance of the run of a year is estimated just six - eight days before fish arrival in the bay. The inseason forecast of the Port Moller test fishery provides the fishing industry and management agency, or ADFG with fish run timing as well as fish run size.

The Port Moller test fishery gear is a drift gillnet. Its stretched-mesh size is five and 1/8 inches (13.02 cm), and it is 200 fathoms long (366 m) and 60 meshes deep (7.81 m). CPUE at each fishery station is calculated by dividing the catch number by the product of the drift gillnet length times fishing time. When the unit of fishing time is minutes, UW FRI uses the following CPUE formula.

$$\text{CPUE} = 6{,}000 \times \frac{\text{catch}}{\left[200 \text{ fathoms} \times \text{fishing time (minutes)}\right]} \tag{1.1}$$

where 6,000 is a scale factor. Beside the catch data, the Port Moller fishery project collects information about water turbidity by Secchi disc, water temperature, air temperature, cloud cover, wave height, wind speed, wind direction, and tide (Rogers et al. 1999). UW FRI operates Port Moller test fishery from early June to about July 10. The test boat attempts to fish each day at several stations located along a transect line between Port Moller and Cape Newenham (Figure 1.1, Rogers 1999). From onshore to offshore along the transect line, the stations are named 2, 4, 6, 8, 10, and 12. Station 2 is located 33 miles out from Port Moller and the distance between sequential stations of these even numbers is 10 miles (Table 1.1). The daily fishery operation consists of a set of these stations. Traditionally only four stations, 2, 4, 6 and 8 had been considered until stations 10 and 12 were added from year 1999. If many fish are caught from station 8, the crew fishes at station 10 and even at station 12 to detect the offshore distribution of fish passage. The fish spatial distribution over the inshore through offshore (i.e. station 2 through station 8) has not been constant every year. With the Port Moller data set of 1985 through 1989, Helton (1991) found that CPUE at stations 2 and 4 were higher under north, northwest, and west winds. However, my analysis with the data set of 1985 through 1999 produced a different result from that of Helton (1991). The winds of northwest, north, northeast, and east led to more offshore distribution of the fish while those of southeast, south, southwest and west resulted in more onshore distribution (Figure 1.6).

Rogers of UW FRI found that the CPUE from station 8 has been significantly correlated with the actual run size. Rogers weighted the CPUE of station 8 twice those of the other stations 2, 4, and 6. We call the sum of the weighted CPUEs of a day Rogers' index of the day.

$$\text{Rogers' index of day } t = \frac{4}{5}\cdot\left(\mathrm{CPUE}_{2,t} + \mathrm{CPUE}_{4,t} + \mathrm{CPUE}_{6,t} + 2\cdot\mathrm{CPUE}_{8,t}\right) \qquad (1.2)$$

where $\mathrm{CPUE}_{s,t}$ denotes CPUE of the test fishery deployed at station s and day t. The inseason forecast with Port Moller fishery data is made by the ordinary regression model, where its response variable is the historical actual run size and its explanatory variable is the cumulative Rogers index up to the latest fishery date. With the runs of year 1985 through 2001 (that of 1986 is missing) to Bristol Bay and the cumulative Rogers' indices up to July 9 of the corresponding years, the Rogers' regression model was $\hat{Y} = 14.202 + 0.011\cdot X \quad (R^2 = 0.46, p = 0.004)$ (Figure 1.7) where $\hat{Y}$ is the predicted run, and X is the cumulative Rogers' index. In Figure 1.7, the three points of 1997, 1998 and 2001 look outliers. Excluding those points, the regression model improved: $R^2 = 0.86$, $p = 0.000$. The failure of the Rogers' inseason forecasts of the 1997, 1998 and 2001 runs also provoked re-examination of the traditional methodology of the forecast.

Other studies of inseason forecast

Other studies of the inseason forecasts of the Bristol Bay sockeye salmon runs, Mundy (1979) and Fried and Hilborn (1988), differ from the literature described above in methodology. Mundy (1979) defined 'fish migratory timing' as a frequency distribution of time. In other words, fish migratory timing referred to fish abundance per unit time in a fixed geographic reference frame. He showed by literature review that fish migratory timing was unique by fish stock, and used the concept of migratory timing to estimate fish return size. He considered fish arrival time a random variable, and normalized fish migratory timing (i.e. a frequency distribution of time). He called the normalized frequency of the fish arrival date 'the time density.' The return size of sockeye salmon to Bristol Bay was estimated with the time density developed with historical data from an offshore test fishery. When x fish were observed up to day d from a test fishery, the total run size in the season could be estimated by dividing x by the cumulative time density at d.

Fried and Hilborn (1988) used Bayesian law to make an inseason forecast of the Bristol Bay sockeye salmon return. They combined the probability densities of four data sources: (1) data used for a preseason forecast, (2) cumulative commercial CPUE of Unimak fishery, (3) cumulative CPUE of Port Moller test fishery, and (4) cumulative commercial catch and spawning escapement data. The combined probability density was used as the resultant joint density given the return size, which was the parameter of their interest. As prior probability of the return size, they chose a set of 67 alternative hypotheses corresponding to total run sizes ranging from 0 to 66 million (using increments of 1 million sockeye salmon). They fitted a Gamma density to the historical runs of year 1956 through 1987 and used the Gamma density to calculate the prior probability of the respective run in the 67 hypotheses. Because the Gamma density is continuous, they needed to scale the prior probabilities by letting the sum of the prior probabilities become one. Finally they calculated posterior probability of the return size by Bayesian law. This calculation was repeated every day when the inseason data are updated. As results of a hind-casting procedure, where only data prior to the year of interest were used to calculate predictive equations, the Bayesian composite forecast was always more accurate than the least accurate one of the forecasts with individual data sources and was sometimes more accurate than the most accurate one of the forecasts with individual data sources.

1.2.4. Inseason forecasts of salmon runs to other areas

Salmon runs to the Skeena River, B.C., Canada

Walters and Buckingham (1975) developed a control system for inseason salmon management with sockeye salmon and pink salmon (*O. gorbuscha*) of the Skeena River, B.C., Canada. The main idea of their control system was to correct control variables or management actions in the system as data were updated. The management actions were determined on weekly basis. The objective of the management was to achieve target escapements of the two salmon species and to allow a fishery on the surplus fish. They needed to estimate the run size of the respective salmon for the objective. Because of high uncertainty in the preseason forecast, they combined the preseason estimate and the inseason estimate by weighting these two estimates as daily data were updated. That is,

$$R = W_t \cdot R_p + (1 - W_t) \cdot R_i$$

where R: run estimated, W_t: weight based on data to time t $(0 \le W_t \le 1)$, R_p: preseason estimate of run, R_i: inseason estimate of run. When the preseason forecast and the inseason forecast were assumed to be independent of each other, the variance of run was as follows.

$$Var(R) = W_t^{\,2} \cdot Var(R_p) + (1 - W_t)^2 \cdot Var(R_i)$$

The value of W_t was determined as W_t that minimized $Var(R)$ in the above equation. Thus, the W_t was able to be expressed as a function of $Var(R_p)$ and $Var(R_i)$. The W_t was near 1 early in the season, and decreased as $Var(R_i)$ decreased (i.e. as time went by). Because of this role of the weight W_t, the run estimated was affected more by preseason forecast early in the season and more by inseason forecast later in the season.

In making an inseason forecast of the salmon run, Walters and Buckingham (1975) used the observed run to date in the season and the historical daily run proportions of the run. The inseason estimate, R_i was calculated by simply dividing the observed run to date by historical cumulative proportion to the date. The observed run to date in the season was the sum of catch and escapement to the date.

$$R_i = \frac{observed\ run\ to\ date}{cumulative\ proportion\ to\ date}$$

Regarding the calculation of $Var(R_i)$, they directly used the formula of Bigelow of International Institute for Applied Systems Analysis (IIASA) without giving the reasoning (Walters and Buckingham 1975, p. 112). No description was available except that the variance calculation of the formula was approximated. I guess that the approximation may have been from the Taylor series approximation. The variance formula was

$$Var(R_i) \approx \frac{R_t^{\ 2} \cdot Var(P_t)}{P_t^4} \cdot [1 + 2 \cdot \frac{Var(P_t)}{P_t^2}]$$

where R_t: observed run to time t, and P_t: mean cumulative proportion returned at time t.

This information of the estimated run was used to calculate a target exploitation rate[3] each week.

$$target\ rate = \frac{(total\ desired\ catch) - (catch\ to\ date)}{(total\ remaining\ run)}$$

However, where there was difference between sockeye salmon and pink salmon in run timing, applying a common target rate to the two salmon runs would have been problematic.   Walters and Buckingham (1975) let different target rates be applied separately to sockeye salmon and pink salmon over different time zone.

The key control variable in the system of Walters and Buckingham (1975) was the number of open days for the fishery each week. The following equation of a catch curve was used to calculate the number of open days.

$$U = (1 - \exp[-c \cdot (E \cdot d)])$$

[3] exploitation rate = catch / run

where *U*: exploitation rate, *c*: catchability coefficient, *E*: fishing effort per day open, and *d*: days open. When exploitation rate *U* (= catch/run), catchability coefficient *c*, and effort per day open *E* are known, days open *d* can be calculated from the above equation. The calculation of exploitation rate *U* was described in the above paragraph. Weekly fishing effort per day open *E* were empirically calculated on the basis of relations with CPUE of the previous week in the season of this year and with CPUE of the week in the season of last year. Catchability coefficient *c* was calculated from the above relationship of exploitation rate *U* and days open *d* with the data of year 1971 through 1973. Finally days open *d* could be calculated. This procedure from the estimation of run to the determination of fishing days was repeated every week in the season.

Pink salmon runs to southeastern Alaska

Sex ratio information was used to make an inseason forecast of the pink salmon run to southeastern Alaska (McKinstry 1993, Zheng and Mathisen 1998). A remarkable pattern in pink salmon runs was temporal change in sex ratio; male pink salmon were preponderant during the first half of the run and female pink salmon during the second half.

Because the possible number of sexes was two (male or female) and thus sex could be considered a binomial variable, McKinstry (1993) used a logistic regression suited for binomial data. In the logistic regression model, he estimated run timing of pink salmon, specifically the mean timing day (MTD). The proportion of male fish being observed at time *t* was formulated as a logistic function.

$$p(t) = \frac{\exp(\alpha + \beta \cdot t)}{1 + \exp(\alpha + \beta \cdot t)}$$

where α and β are parameters. The inflection of this logistic curve was considered a change in preponderance from males to females. MTD was defined as the time that corresponds to the inflection point of the curve. That is, the value of *t*, that make the above *p(t)* be 0.5, was defined as MTD. The *t* value could be expressed as α and β: $t = -\alpha/\beta$. When taking the logit function in the above equation, we get a linear form:

$$\text{logit}(p(t)) = \log\left(\frac{p(t)}{1-p(t)}\right) = \alpha + \beta \cdot t$$

These parameters α and β were able to be estimated by fitting the linear form of the logistic function to observed *p(t)*. Thus, the predicted MTD was given as '$-\hat{\alpha}/\hat{\beta}$'. The mean value of the historical MTD values was adjusted by the predicted MTD in the season (the mean curve of historical run proportions at a given time was shifted by the adjusted time). On the basis of the shifted run proportion curve, total run size was predicted:

$$\textit{Predicted total run} = \left({\textit{observed run to date } t} \big/ {\textit{run proportion at date } t} \right)$$

Adjusting run timing turned out to improve the inseason forecasts. However a big improvement generally occurred only following the middle of the run.

The setting of Zheng and Mathisen (1998)'s study was the same as that of McKinstry (1993). They developed a sex ratio index with data of cumulative catch by all gears or cumulative CPUE of the seine fishery, and estimated run of pink salmon in the season by three models: a linear model, a non-linear model, and a combined model. In these models, the response variable was the run size and the candidate predictive variables were the sex ratio index, cumulative catch, cumulative CPUE, and cumulative catch. The sex ratio index was derived from deviations of weekly male proportions to the corresponding mean values and the deviation of the sex ratio curve to its mean curve for a given year. Incorporating sex ratios into inseason forecast models correctly adjusted the run timing and thus improved overall forecasts. The forecast errors of Zheng and Mathisen (1998) were much smaller than those reported by McKinstry (1993). The difference was due to different forecast models, methods used to incorporate sex ratio data into forecast models, stock definitions, and periods when the forecasts were conducted.

Chum salmon run to Hood Canal in Puget Sound, Washington

Springborn et al (1998) used a time density model for an inseason forecast of the chum salmon run to Hood Canal in Puget Sound, Washington. The concept of a time density had been first used by Mundy (1979) to estimate fish abundance. The concept was described in the sub-section, 'Other studies of inseason forecast' under section 1.2.3. Springborn et al. (1998) extended the Mundy (1979)'s idea to estimate run size and entry timing for the northern Hood Canal chum salmon fishery. Two kinds of fishing gear were used: drift gillnets and purse seines. Their inseason forecast model consisted of two parts. The first part was to build the time density with daily CPUE data from the drift gillnet fishery and to estimate run size and entry timing in the season. The second part was to correct the run size estimate of the first part by catch data from the purse-seine fishery in the season. Because of a large disparity in the daily harvest rate between the gillnet fishery and the purse-seine fishery, Springborn et al. (1998) did not use catch data from the purse-seine fishery for the time density model. Including daily CPUE from the purse-seine fishery into building the time density would lead to serious inflation of the run size estimate. The model deployed in the second part was a linear regression model where a peak 1-day purse-seine catch and the time density run estimate were used as independent variables to provide a 'corrected' run size estimate. The peak 1-day purse-seine catch had been found significantly correlated with the actual run size.

Because I already described how the time density was used to estimate the run size in the season, I describe here how to estimate run timing from the time density. Letting 'fish arrival time' be a discrete random variable (say Y), the mass function of Y would be a normalized frequency of the arrival time. Thus, the mass function[4] of fish arrival time in year j would be

$$f_{Y_j}(y_{ij}) = \frac{n_{ij}}{N_j}$$

where n_{ij}: the number of fish that pass a reference region at time i in year j, N_j: the total number of fish over the entire time range in year j. The expectation value of the arrival time would be

[4] Even though Mundy (1979) and Springborn et al (1998) call the normalized frequency distribution of fish arrival time 'time density,' I refrain from using the term, 'density' because the random variable Y is discrete not continuous.

$$E(Y_j) = \sum_{i=1}^{l} y_{ij} \cdot f_{Y_j}(y_{ij})$$

where the time interval (day) ranges from 1 to *l*. The overall mean over year 1 through year *m* on the basis of historical data would be

$$\Gamma = \frac{\sum_{j=1}^{m} E(Y_j)}{m}$$

When year *j* was the year of inseason forecast, the parameter of interest (say, β) was difference in run timing between the past and the year *j*.

$$\beta_j = \Gamma - E(Y_j) \tag{1.3}$$

With the historical data, the parameters β_j's of each year *j* could be estimated. However, it is the parameter of the current year not the past year in which we are interested in the forecast.

To estimate the parameter β_j of the current year, Springborn et al. (1998) took the following steps. With the assumption that n_{ij}, the number of fish passing a region at time *i* in year *j* was proportional to $CPUE_{ij}$ of the fishery at the region on time *i* in year *j*, n_{ij} in the above time mass was replaced by $CPUE_{ij}$. That is, $CPUE_{ij} = c \cdot n_{ij}$ where *c* is a proportional constant. And then Springborn et al. (1998) related the cumulative mass function of *Y* to a cumulative distribution function that has properties similar to the cdf of a normal density. That is,

$$F_{Y_j}(t) = \sum_{i=1}^{t} \frac{CPUE_{ij}}{c \cdot N_j} = \frac{1}{\{1 + \exp[-(a + b \cdot t)]\}} \tag{1.4}$$

The right side of this above equation was from Mathisen and Berg (1968). When adding run timing parameter β_j to *t* in the above equation,

$$\sum_{i=1}^{t} CPUE_{ij} = \frac{c \cdot N_j}{\{1 + \exp[-(a + b \cdot (t + \beta_j))]\}} \tag{1.5}$$

This above model was called the 'time-shifted' distribution. For estimation of a, b, and c, the historical data of β_j, $CPUE_{ij}$ and N_j were used. And then β_j of the season was estimated from the observed daily $CPUE_{ij}$ in the season, the estimated values of a, b, and c, and the estimated N_j.

Outmigration of smolts from the Snake River in the Columbia River basin

In the Columbia River Basin, there are concerns about smolt mortality by dams during the outmigration because smolts must pass several dams before reaching the ocean. Since 1988, wild salmon have been PIT-tagged through monitoring and research programs conducted by the Columbia River fisheries agencies and Tribes (Townsend et al. 1996, 1997). With the data of PIT-tagged recoveries and the outmigration time (day), Townsend et al. (1996, 1997) predicted the proportion of a particular population that arrived at an index site on a given date. The forecast of the proportion can be used to adjust daily spill amounts of a dam during the migration season. Regulating the timing and volume of water released from storage reservoirs has become a central mitigation strategy for improving downstream migration conditions for juvenile salmonids in the Snake River. Townsend et al. (1996, 1997) introduced three methods to predict the proportion $\hat{p}$ of the outmigration run at a given day and site, and combined three values from the three algorithms to give the final estimate $\hat{p}$.

In the first method, historical outmigration runs over time were used as an important reference. For each year, the percentage of the cumulative outmigration run by date was calculated. The proportions of the historical cumulative runs were plotted against date. The cumulative run was divided into 100 equal portions and the slopes over each corresponding interval were calculated. The cumulative runs were smoothed to filter out statistical randomness. The slopes of the historical curves at each percentage were to be compared to that of the current year of prediction. The total squared error for each predicted percentage of outmigration run was calculated according to the following.

$$LSE(\hat{p}) = \sum_{i=1}^{n} \sum_{j=0}^{100} (S_{oj} - S_{ij\hat{p}})^2 \cdot W_{ij}$$

where S_{oj}: the slope at the jth percentile (j = 0, 1, 2, …,100) for the current year of prediction, $S_{ij\hat{p}}$: the slope at the jth percentile for the $\hat{p}$ percent the historical year i (i = 1, 2, …, n), and W_{ij}: weight for the jth percentile for ith historical year. LSE denotes Least Squares Error. The goal of this algorithm is to find $\hat{p}$ that minimizes the LSE. The weight W_{ij} is

$$W_{ij} = \frac{D_{oj} + D_{ij}}{R_o + R_i}$$

where D_{oj}: estimated number of days between the (j-1) and the jth percentile for the season, D_{ij}: number of days between the (j-1) and jth percentile for the ith historical year, R_o: range in days of the current observed outmigration, and R_i: range in days of the ith historical year outmigration. The effect of the W_{ij} is to give more weight to the errors generated in the tails of the distribution, where the slopes tend to be flat and the number of days between sequent percentile points is high. The sum of weights is one.

The second algorithm used the PIT-tag data. The proportion $\hat{p}$ for a given day and site was calculated by the following relationship.

$$\hat{p} = \frac{x_d}{\overline{p} \times N}$$

where x_d: total observed smolts to day d, $\overline{p}$: the mean 'recapture' proportion of the previous years, and N: total number of smolts tagged in the season. The denominator represents the expected fish to be recovered.

The third algorithm used the average number of days since the outmigration started, weighted by the number of fish observed per day. The average number was called the mean-fish-run-age (MFRA). Thus, MFRA was formulated as follows.

$$MFRA = \frac{\sum_{d=1}^{n} [fish_d \times (n+1-d)]}{\sum_{d=1}^{n} fish_d}$$

where $fish_d$ = number of fish observed on day d; n = total number of days until the cumulative proportion p of the total smolt outmigration has been observed. The present year's MFRA is matched to the respective historical year's MFRA. The historical observed p corresponding to the matching MFRA is the predicted $\hat{p}$ from that year.

The $\hat{p}$ values from the above three algorithms were given the respective weight in calculating the final value, $\hat{p}$.

1.2.5. Current management of the Bristol Bay sockeye salmon

Commercial and subsistence fisheries mainly target sockeye salmon. Commercial sockeye salmon harvests in Bristol Bay began in 1893 (Minard and Meacham 1987, ADFG 1998). The current commercial fishing is usually limited to five major fishing districts (Figure 1.2). In case of the Wood River and the Naknek River, commercial 'in-river' fisheries sometimes occur. Two kinds of legal fishing gears are allowed in the commercial harvests: (1) 150 fathom (274.5 m) drift gillnets fished from 32 foot (11.1 m) gillnet boats and (2) 50 fathom (91.5 m) set gillnets attached to the beach. A subsistence fishery is operated by Alaska residents. Subsistence salmon fishing is significant in numbers of fish utilized as well as in its cultural importance to watershed residents (Minard and Meacham 1987). The subsistence harvest has a legal priority over commercial and sport harvests. A sport harvest of the Bristol Bay sockeye salmon is not significant.

At present, the agency managing the Bristol Bay sockeye salmon is ADFG. The primary management strategy of ADFG is expressed as three goals. The first goal is to meet the required number of spawners in each of eight major river systems: Togiak, Igushik, Wood, Nushagak, Kvichak, Branch, Naknek, Egegik, and Ugashik. Sockeye salmon return to the Branch river system is not significant. The optimal escapements are set by ADFG. The escapement goals are shown in Table 1.2 (ADFG 1998, Lew *personal communication*). The second goal is to conserve the profile of the escapement return time. To achieve this second goal, ADFG needs to distribute catches and escapements over the entire run and to preserve genetic diversity. The third goal is to maximize harvests of the surplus fish after subtracting the optimal escapements from the returns. The third goal concerns the economic aspect.

To achieve these three goals, ADFG needs to assess the sockeye salmon run timing and strength during the fish return season. The inseason assessment is based on observations of various sources: the Port Moller test fishery, the commercial district fishery, spawning escapement monitoring, aerial surveys over spawning grounds, the district test fishery, and the in-river test fishery (ADFG 1998). The district test fishery is deployed at irregular times while the in-river test fishery is operated at every tide change (every flood tide and every ebb tide). The within-district test fishery is held in every district except the Togiak district. The in-river test fishery is operated in only four rivers: the Kvichak River, the Egegik River, the Ugashik River, and the Igushik River (Minard and Meacham 1987).

The ADFG manages each of the river specific stocks as an individual entity. Commercial fishing openings and closures are predicated on attainment of escapement goals and are implemented by flexible rather than fixed fishing schedules (Minard and Meacham 1987). Authority to open and close fishing districts by emergency order has been given to biologists located near the fishing grounds, allowing rapid management response time. Thus each of the five districts is managed independently to conform to the individual stock characteristics of run timing and strength.

1.3. THESIS STRUCTURE

The technical part of this dissertation comprises three chapters: Chapters 2, 3, and 4. In Chapter 2, I use the Port Moller test fishery data to detect run timing of Bristol Bay sockeye salmon during the season. In Chapter 3, I develop an algorithm for forecasting district-specific run sizes, and show point estimates of runs. In Chapter 4, I use Bayes' law to show probability distributions for runs. All forecast results in this thesis are based on a hind-casting procedure, where only data prior to a forecast time are used to calculate forecasts of run timing and run sizes.

In Chapter 5, I discuss a management application of this thesis work. In Appendices, I show the forecast program written in Automatic Differentiation Model Builder (ADMB), and describe how to run the ADMB program for one who wants to use it.

1.4. ACRONYMS

ADFG: Alaska Department of Fish and Game
ADMB: Automatic Differentiation Model Builder
CPU: Central processing unit
CPUE: Catch per unit effort
CV: Coefficient of variation
K-S test: Kolmogorov-Smirnov goodness of fit test
MCMC: the Markov Chain Monte Carlo method
MB: megabyte(s)
MLE: Maximum likelihood estimate (estimator)
MSE: Error mean square or residual mean square
RAM: Random access memory
RTI: run timing index
UW ASP: University of Washington Alaska Salmon Program
UW FRI: University of Washington Fisheries Research Institute

Table 1.1. Location of Port Moller test fishery stations (Helton 1991, Rogers et al. 1999). These stations are shown as small dots in Figure 1.1.

Station	Miles from Port Moller	Latitude	Longitude
2	33	56°25.48 N	160°44.88 W
4	43	56°35.15 N	160°50.71 W
6	53	56°45.07 N	160°56.96 W
8	63	56°54.43 N	161°01.96 W
10	73	57°03.86 N	161°07.83 W

Table 1.2. The goals of sockeye salmon escapement to eight Bristol Bay river systems (ADFG 1998, Lew *personal communication*.)

District	River system	Escapement goal
Kvichak-Naknek	Kvichak	4-6 millions for off-peak years 6-8 millions for peak and pre-peak years
	Naknek	0.8-1.4 millions
Egegik	Egegik	0.8-1.4 millions
Ugashik	Ugashik	0.5-1.2 millions
Nushagak	Igushik	150,000-250,000
	Wood	0.7-1.2 millions
	Nushagak	340,000-760,000
Togiak	Togiak	100,000-200,000

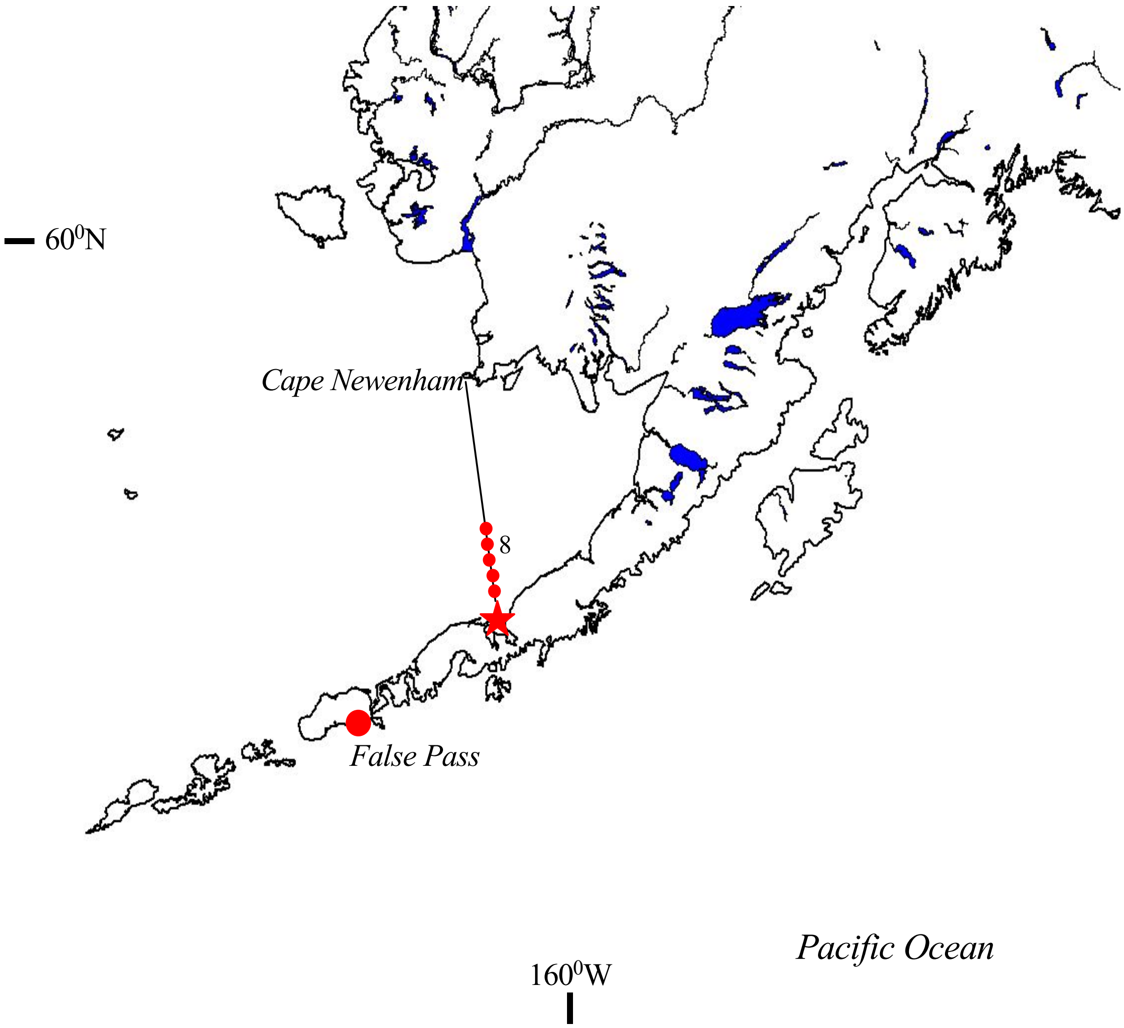


Figure 1.1. Bristol Bay, Alaska. The star mark indicates the location of Port Moller. The Port Moller test fishery occurs along a transect line between Port Moller and Cape Newenham. The small dots on the transect line represent the stations of the test fishery. From onshore to offshore, the stations are named 2, 4, 6, 8, and 10.

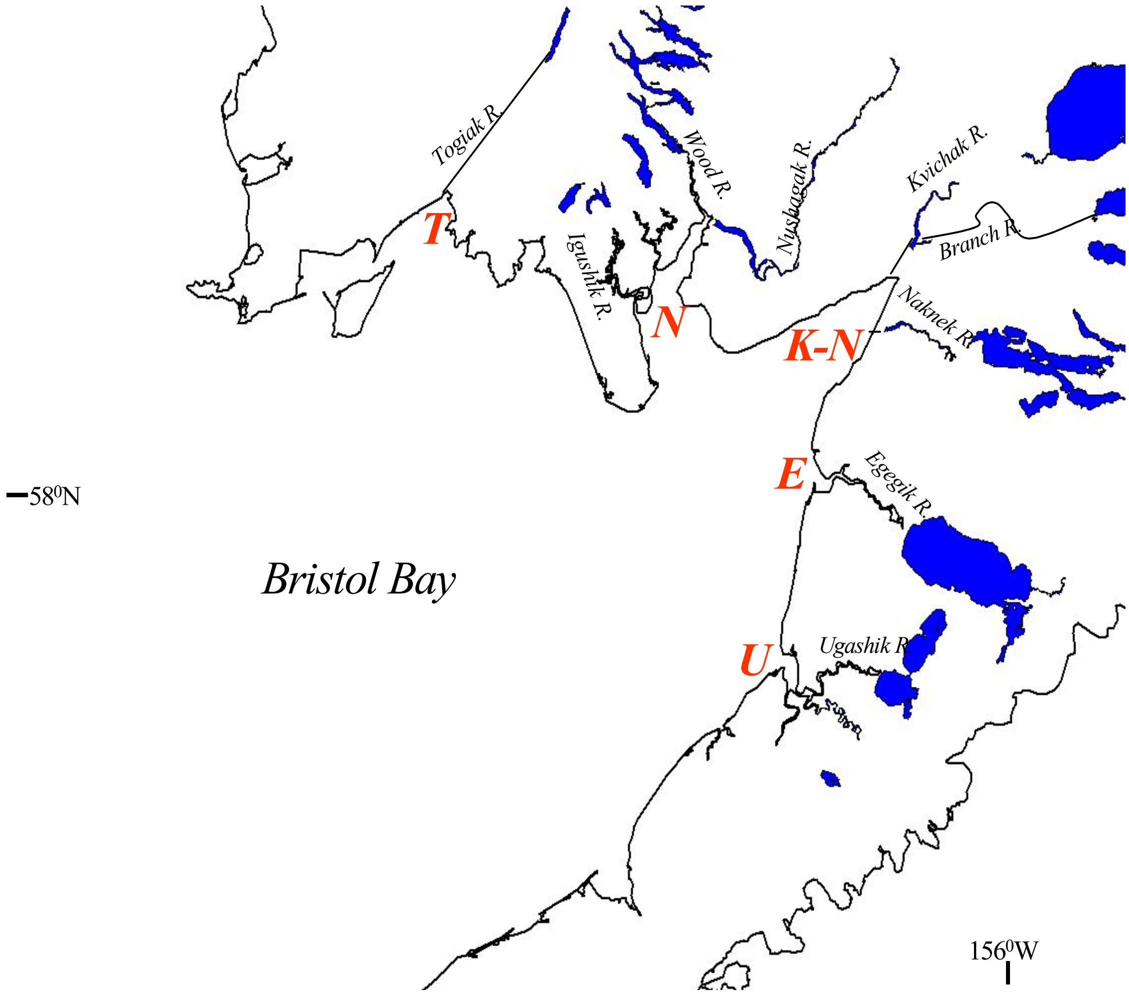


Figure 1.2. Five estuaries and nine river systems in Bristol Bay, Alaska. Names of the respective five estuaries are abbreviated with their first letters (*T*: Togiak, *N*: Nushagak, *K-N*: Kvichak-Naknek, *E*: Egegik, and *U*: Ugashik).

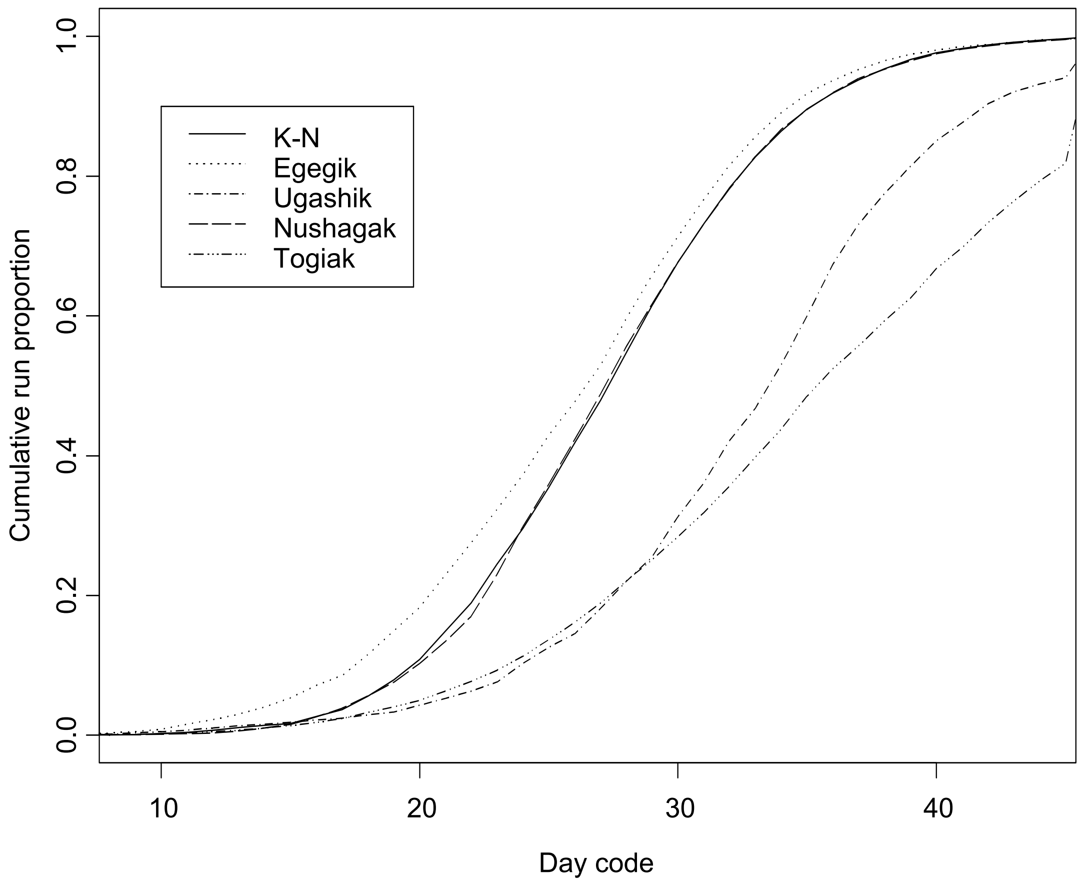


Figure 1.3. The cumulative run proportions of five district stocks. The respective five lines are the mean values of the cumulative run proportions by day of year 1955 through 2001 except for the Togiak district. The data of years 1955 through 1957 for the Togiak district are not available. The cumulative run proportion can be used as a run timing index. I code calendar dates, starting on June 10: day code 1 = June 10, day code 2 = June 11, and so on.

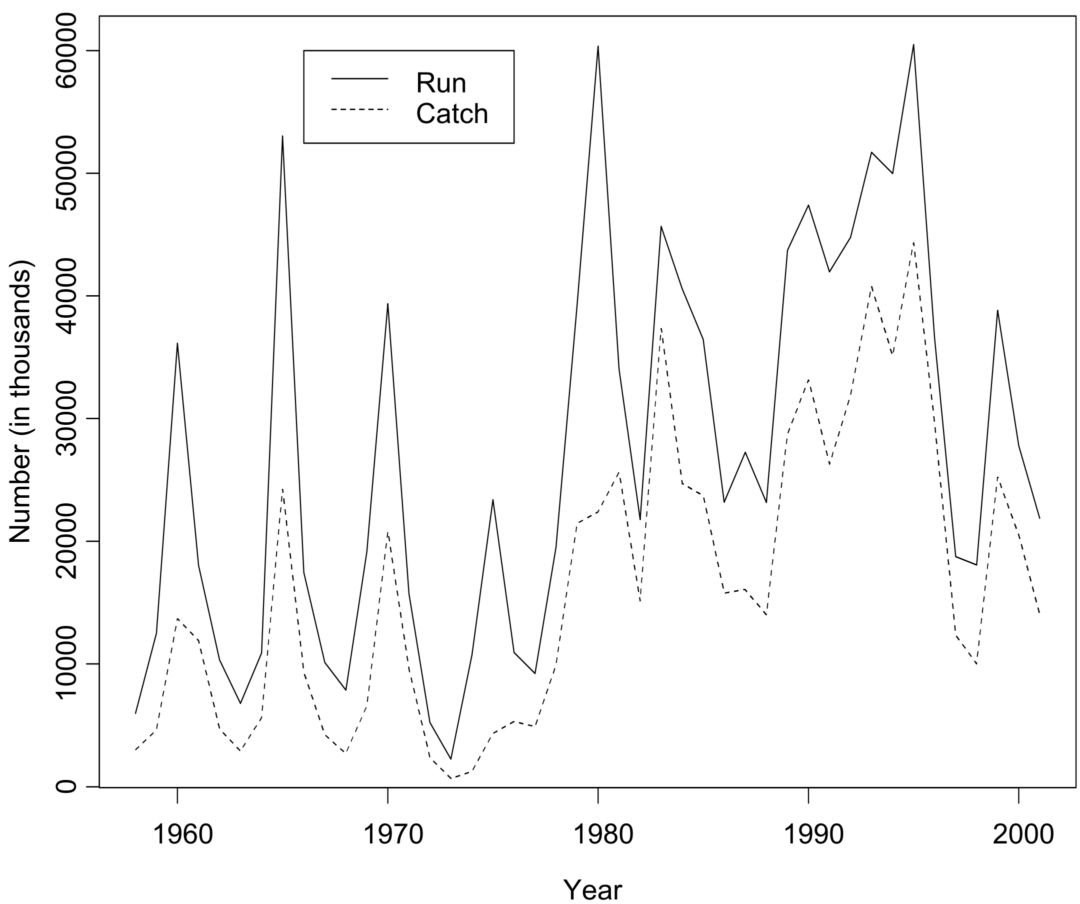


Figure 1.4. Annual returns of the Bristol Bay sockeye salmon from 1958 to 2001. Run size (or return size) of a year is the sum of catch and escapement at the year.

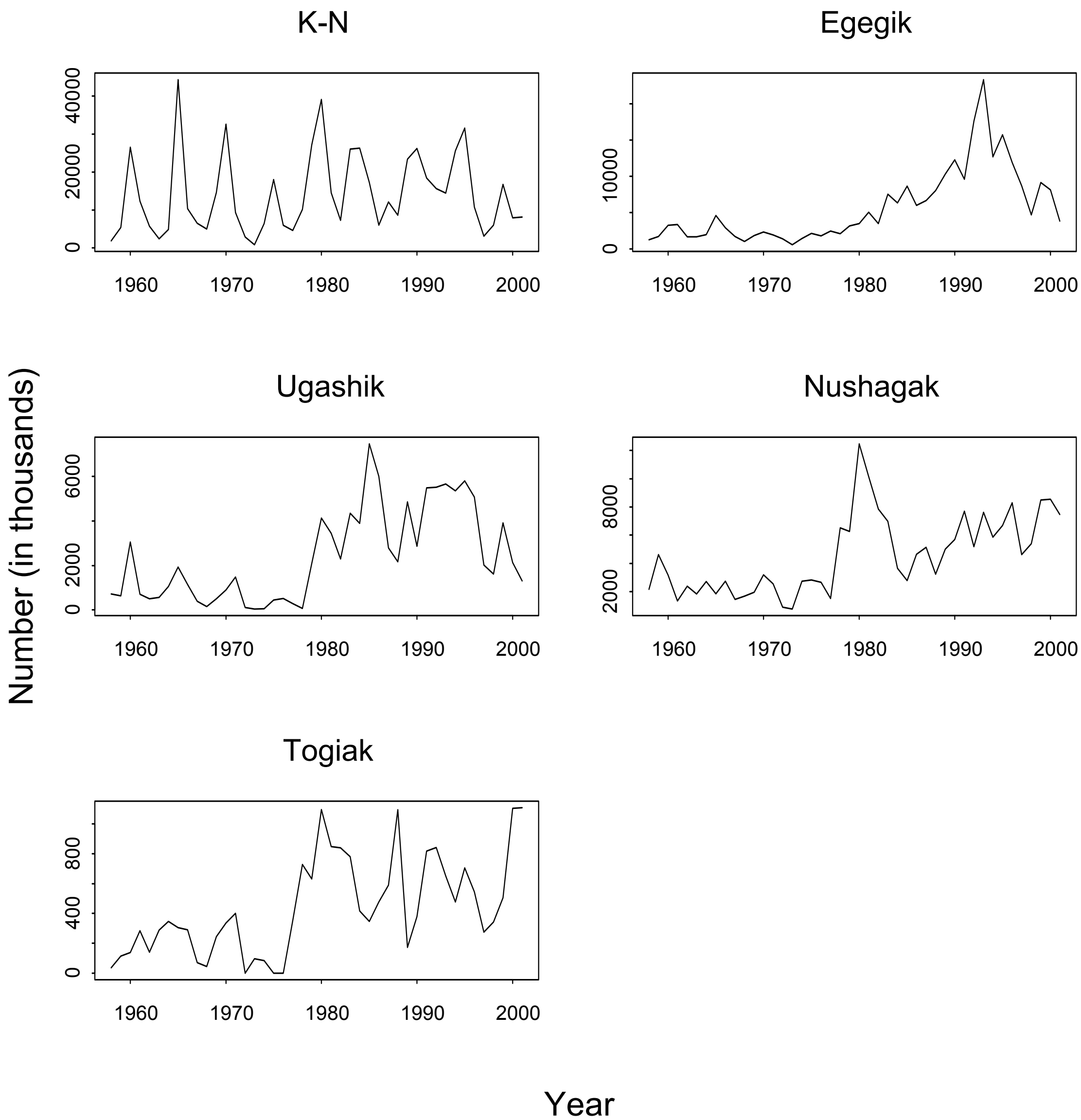


Figure 1.5. Annual returns to five districts from 1958 to 2001. A cyclic pattern in annual returns is most remarkable in the Kvichak-Naknek district. This cyclic pattern is mostly due to runs of sockeye salmon to the Kvichak River.

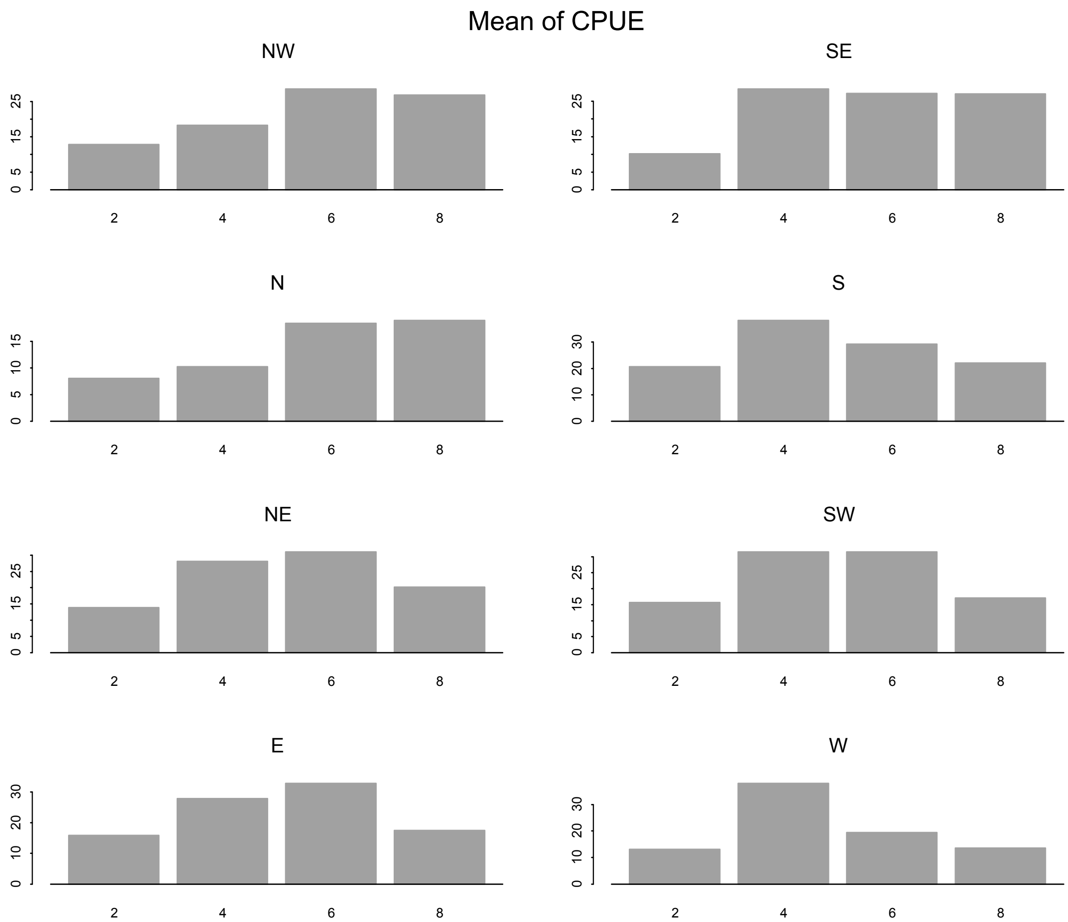


Figure 1.6. The mean values of station- and wind direction- CPUE of the Port Moller fishery deployed during year 1985 through 1999. The winds of northwest, north, northeast, and east led to more offshore distribution of the fish while those of southeast, south, southwest, and west resulted in more onshore distribution.

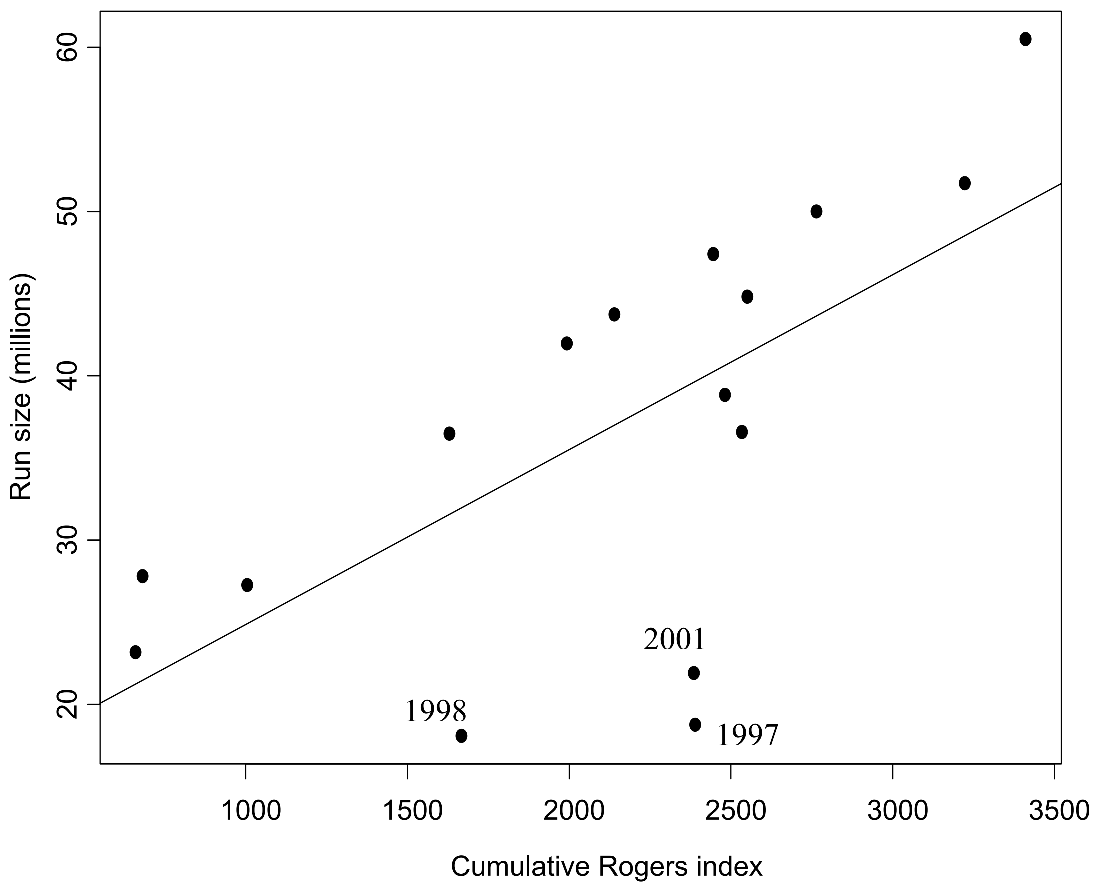


Figure 1.7. The relationship between 1985-2001 runs to Bristol Bay and the cumulative Rogers' indices up to July 9 of the corresponding years. The line represents fitted values of the regression model: $\hat{Y} = 14.202 + 0.011 \cdot X \quad (R^2 = 0.46, \quad p = 0.004)$. Removing the three data points of years 1997, 1998, and 2001, the regression model improves: $R^2 = 0.86, \quad p = 0.000$.

CHAPTER II. INSEASON FORECAST OF RUN TIMING

INTRODUCTION

Variability in fish run timing is one of the main factors that make difficult an accurate inseason forecast of fish run size. Sockeye salmon adults return to Bristol Bay mainly during one month from about the middle of June through the middle of July (see the sub-section of ‘Life history of the Bristol Bay sockeye salmon’ under section 1.2.2). Figure 2.1 displays the historical run proportions against day. The run proportions against day can be a run timing indicator (Figure 1.3). I find large yearly variability in the run timing (Figure 2.1). Not only in Figure 2.1 but also in this entire thesis, I code calendar dates, starting on June 10: day code 1 = June 10, day code 2 = June 11, and so on. In sub-sections, ‘Pink salmon runs to southeastern Alaska,’ and ‘Chum salmon run to Hood Canal in Puget Sound, Washington’ under section 1.2.4, I described ideas about a forecast of fish run timing found in literature.

At present, there is no accepted method for forecasting run timing of Bristol Bay sockeye salmon. As an *ad hoc* index of run timing of the fish, Hilborn (*personal communication*) uses the ratio of the sum of Rogers’ CPUE of June 21 through June 30 (day code 21) to the sum of the indices up to June 20. Rogers’ CPUE is the weighted CPUE calculated with the catch data of the Port Moller test fishery (Equation 1.2). Though Hilborn’s run timing index provides some information about run timing, we have to wait until June 30 during the season to calculate the index.

The objective of this chapter is to develop an acceptable algorithm for forecasting run timing of Bristol Bay sockeye salmon on a daily basis during the run season. I use the Port Moller test fishery data.

METHODS

2.1. PORT MOLLER TEST FISHERY DATA

I used Rogers' CPUE to detect fish run timing. In the historical catch data set of the Port Moller fishery, Rogers' CPUE of some days were missing because the fishery could not be deployed under unexpected circumstances such as bad weather, or damage to the fishing gear or boat. I replaced the missing CPUE with the mean value of those of the days before and after. Figure 2.2 shows Rogers' CPUE against day by year. The Port Moller fishery was not deployed in 1986, so Figure 2.2 does not show data for 1986.

2.2. RUN TIMING INDEX

I standardized the cumulative daily CPUE by setting the final sum equal one (100%). Then, I fit the following logistic curve to the standardized cumulative CPUE.

$$y = \frac{1}{1 + \exp(a + b \cdot x)} \tag{2.1}$$

where x is time (day), y is the cumulative Rogers' CPUE, and a and b are parameters. I defined fish run timing index (RTI) as time (day) that corresponded to the inflection point of the fitted logistic curve. That is, the RTI unit is 'day,' but not necessarily discrete. Figure 2.3 shows an example, where I fit the logistic curve to the standardized cumulative CPUE of year 1999.

The analytical derivation of RTI was simple. Differentiating y of Equation 2.1 twice with respect to x, and then solving '$d^2 y / dx^2 = 0$' for x led to the following.

$$x = -\frac{a}{b} \equiv RTI \tag{2.2}$$

The RTI value is determined by two parameters: a and b (Equation 2.2). I call this index 'Hyun's index' to ease comparison with other alternative indices.

I used the Delta method (Seber 1982) to derive the variance of RTI. The idea of the Delta method is to expand a function of interest by the Taylor series and then to consider the first significant terms. Thus, a variance formula derived by the Delta method is an approximation formula. By the Delta method, I had the following formula for calculation of the RTI variance.

$$\begin{aligned} Var(RTI) &= Var\left(-\frac{a}{b}\right) \\ &\approx \left[\frac{E(a)}{E(b)}\right]^2 \cdot \left[\frac{Var(a)}{E(a)^2} + \frac{Var(b)}{E(b)^2} - \frac{2 \cdot Cov(a,b)}{E(a) \cdot E(b)}\right] \end{aligned} \tag{2.3}$$

I used statistical software, Splus in estimating the two parameters (*a* and *b*) in the non-linear logistic curve (Equation 2.1) and their covariance. The 'nls' function in Splus enabled us to fit a non-linear model to data, and the output provided the estimates of parameters in the model and their variance-covariance matrix.

2.3. VALIDATION OF PORT MOLLER RTI

Because the Port Moller catch data were from a 'test' fishery, the data size was not large enough to produce statistically reliable results. Thus, I compared the Port Moller RTI with run timing of Bristol Bay sockeye salmon, which was inferred from run data (= catch + escapement) of the inshore Bristol Bay. I applied the same idea to the inshore run data. That is, I standardized the cumulative inshore-run, and fit the logistic curve of Equation 2.1 to the cumulative run size. Finally, I defined the inshore RTI as time (day) that corresponded to the inflection point of the fitted curve.

Table 2.1 shows the Port Moller RTI estimates of years 1985 through 2001. Table 2.2 displays the district-specific RTI estimates of the same period. And Table 2.3 has those of the lumped five district fish and the lumped four district fish (excluding the Togiak district fish). Standardizing the cumulative data (Rogers' CPUE for the Port Moller RTI; run size for the inshore RTI) at the final days (July 9 for the Port Moller RTI; the end of the return season for the inshore RTI), I calculated these RTI estimates.

Run timing of year 1994 was latest while that of year 2001 was earliest on the basis of the inshore RTI estimates of years 1985 through 2001 (Table 2.3). I found a high correlation between the yearly Port Moller RTI estimates and those of the inshore RTI except for the Togiak and Ugashik fish (Figure 2.4). The correlation coefficients between the Port Moller RTI estimates and each of those of districts Kvichak-Naknek, Egegik, Ugashik, Nushagak, and Togiak were 0.73, 0.75, 0.53, 0.78, and 0.36, respectively (Figure 2.4). The correlation coefficient between the Port Moller RTI estimates and those of the lumped five districts was 0.75 (Figure 2.4). When comparing the Port Moller RTI estimates with those of the lumped four districts excluding the Togiak district, the correlation coefficient increased a little to 0.77. The high correlations indicate that the Port Moller RTI can detect run timing of the Bristol Bay sockeye salmon.

2.4. RUN TIMING FORECAST

To forecast fish run timing is to compare the run timing estimate of the season with those of the past years. We are interested in *how early* or *how late* the run timing is compared to those of the past years. That is, a forecast of run timing is a relative index on the basis of a comparison between the present and the past. However, it is more desirable to detect run timing before the final day of the Port Moller fishery season because the earlier we forecast fish run timing, the more helpful the forecast information is.

To capture run timing as the meaning of a relative index before the final day of the Port Moller test fishery,

(1) I standardize both the cumulative Rogers' CPUE *at any arbitrary day* during the season and those *at the same day* of the past years,

(2) I calculate the Port Moller RTI of the respective years from the corresponding fitted logistic curves,

(3) finally I compare the Port Moller RTI of the season with those of the past years (Equation 2.4).

Figure 2.5 illustrates an example where fish run timing of year 2001 is evaluated at June 24. I intentionally extended the *x*-axis of Figure 2.5 beyond day code 15 (June 24) to emphasize that the analysis can be done at any day (not necessarily at final day); I ignored the historical data after the day.

I defined the relative index of run timing of the season as the difference between the Port Moller RTI estimate of the season and the average of those of the years prior to the season.

$$\begin{aligned}&\text{Relative index of run timing}\\&=(\text{Port Moller RTI estimate of the season})\\&\quad-(\text{the average of those of the years prior to the season})\end{aligned} \tag{2.4}$$

We can calculate the relative index of run timing at any day during the season. The positive sign of the relative index indicates that run timing of the season is later than the average of the past years, while the negative sign means that that of the season is earlier. In case of the above example (Figure 2.5), the relative index of the 2001 run timing evaluated at June 24 was '-1.1.' That is, the 2001 run timing detected at June 24 was earlier by about one day than the average of the past years. As the season data are accumulated over time, the run timing detection should improve (see Table 2.5).

RESULTS

2.5. PORT MOLLER RTI

Table 2.4 presents the Port Moller RTI estimates of year 1985 through 2001 and their standard deviation estimates evaluated at the following four days: day codes 10 (June 19), 15 (June 24), 20 (June 29), and 25 (July 4). Table 2.1 has those evaluated at day code 30 (July 9). Figure 2.6 compares the yearly Port Moller RTI estimates evaluated at the respective day with those of inshore RTI of four districts (excluding Togiak district). The five lines in Figure 2.6 (B) are the yearly Port Moller RTI estimates evaluated at day codes 10, 15, 20, 25, and 30 in order from the bottom dotted line to the top square box line. The numerical values on the five lines of Figure 2.6 (B) are the correlation coefficients between the respective line and the inshore RTI. The correlation coefficients were high (0.75, 0.76, 0.77, and 0.77) except for the Port Moller RTI estimates of day code 10. It was an encouraging result that the Port Moller RTI estimates of even day code 15 were highly correlated with the inshore RTI estimates (r = 0.75 in Figure 2.6).

2.6. RELATIVE INDEX OF RUN TIMING

Table 2.5 displays the relative index of fish run timing of years 1999, 2000, and 2001 evaluated at day codes, 15, 20, 25, and 30, respectively. The 1999 run timing was slightly later than the average of the past years while those of 2000 and 2001 were earlier. For example, the 1999 run timing detected at July 4 was later by about one day (0.6 in Table 2.5) than the average of the past years, and those of 2000 and 2001 were earlier by about three days and about two days, respectively (-3.4 and -2.4 in Table 2.5). As expected, run timing detection improved as the season data were accumulated over time; absolute values of the relative indices of 2000 and 2001 increased over evaluation time (from 1.4 to 3.8 for year 2000, and from 1.1 to 3.3 for year 2001 in Table 2.5). The results of Table 2.5 are used in Chapters 3 and 4.

DISCUSSION

Port Moller RTI of years 1985 through 2001 were well correlated with those of the Kvichak-Naknek, Egegik, and Nushagak fish while they were poorly correlated with those of the Togiak fish (Figure 2.4). This is not a surprising result because run timing of the Togiak fish is significantly different from those of the other district fish (Figure 1.3).

Port Moller RTI seems to detect run timing, but does not explain a considerable portion of run timing variability. The correlation coefficient between the yearly Port Moller RTI evaluated *at the final day* of the test fishery season and those of the lumped four district fish (excluding the Togiak fish) was 0.77 (Figure 2.6). That is, Port Moller

RTI accounted for *at most* 59% of run timing variability: $0.59 = 0.77^2$ (the determination coefficient = the correlation coefficient^2). If the yearly Port Moller RTI were evaluated before the final day, the proportion of run timing variability explained by Port Moller RTI would be less than 59%.

The correlation coefficient between two sequences indicates how well the fluctuation of elements in a sequence corresponds to that of elements in the other sequence, but the value does not represent the fluctuation *magnitude*. Note that the yearly RTI of the four district fish (excluding the Togiak fish) in Figure 2.6 (A) fluctuate much more remarkably than the yearly Port Moller RTI in Figure 2.6 (B). Because the run timing forecast is a relative index (Equation 2.4), better detection of changes in fluctuation magnitude would improve the forecast.

As an alternative index of run timing, I could use the slope of the line tangent to the fitted logistic curve (Equation 2.1) at $x = 0$. A change in the initial slope of the fitted curve is correlated with that in RTI in Equation 2.2, and thus the choice of the initial slope would not change the current results. However, the slope unit is not time (day), so I prefer the current RTI to the initial slope.

The idea of the relative index of run timing (Equation 2.4) is the same as that of the run time parameter, β in Equation 1.3 (the sub-section of 'Chum salmon run to Hood Canal in Puget Sound, Washington' under section 1.2.4). Applying the idea to the Bristol Bay sockeye salmon run was not successful. To estimate the run time parameter β, I needed to estimate not only three parameters (a, b, and c in Equation 1.5) but also total run size (N in Equation 1.5). The estimation of c, which is the proportion of fish caught by a test fishery on a day out of fish abundance available on the day, requires the prior knowledge of the day-specific proportion of total fish run size over the run season. In case of the Bristol Bay sockeye salmon, the proportion of total run size, which passes the Port Moller on a day, is very variable by year. Besides, the variance of the inseason estimate of total run size is usually large. The large variability in the day-specific proportion and the run size estimate prevented me from applying the idea of Springborn et al. (1998).

Figure 2.7 compares Hilborn's indices (see Introduction of this chapter) for years 1985 through 2001 with the yearly RTI of the lumped four district fish (excluding the Togiak fish). The correlation coefficient between them was 0.68 (Figure 2.7), which was smaller than that between the yearly Port Moller RTI evaluated at day code 20 (June 29) and those of the four district fish (0.76 in Figure 2.6). On the basis of the comparison of the correlation values (0.68 vs. 0.76), Port Moller RTI seems to be better than Hilborn's index. Another merit about Port Moller RTI is that we can estimate it on a daily basis (at any day). The calculation of Hilborn's index requires the Port Moller fishery data up to day code 21 (June 30).

Table 2.1. The Port Moller RTI estimates of years 1985 through 2001, and their standard deviations. The Port Moller fishery was not deployed in 1986. The RTI estimate was estimated with the cumulative Rogers' CPUE standardized at the final day of the fishery.

Year	RTI	S.D.
1985	17.7	0.2
1987	16.6	0.0
1988	17.8	0.1
1989	17.7	0.2
1990	19.0	0.1
1991	17.9	0.2
1992	17.4	0.2
1993	16.8	0.1
1994	20.1	0.1
1995	18.0	0.2
1996	18.0	0.2
1997	18.8	0.2
1998	20.3	0.2
1999	19.5	0.2
2000	14.5	0.2
2001	14.7	0.1

Table 2.2. The district-specific RTI estimates of years 1985 through 2001, and their standard deviations. The RTI estimate was estimated with the cumulative run size standardized at the final day of the return season. 'KN' denotes Kvichak-Naknek.

Year	KN		Egegik		Ugashik		Nushagak		Togiak	
	RTI	S.D.	RTI	S.D.	RTI	S.D.	RTI	S.D.	RTI	S.D.
1985	26.5	0.2	24.1	0.1	29.4	0.1	26.8	0.2	34.9	0.3
1986	29.0	0.1	28.2	0.1	30.9	0.1	30.6	0.1	35.6	0.2
1987	30.2	0.1	25.2	0.1	32.8	0.2	26.8	0.2	36.7	0.2
1988	27.4	0.2	23.7	0.1	33.9	0.1	27.4	0.2	32.8	0.1
1989	24.8	0.1	25.4	0.1	32.2	0.1	25.2	0.1	34.3	0.3
1990	27.8	0.1	27.9	0.1	32.9	0.1	28.3	0.1	35.5	0.4
1991	26.6	0.2	26.9	0.1	32.5	0.1	27.0	0.1	38.6	0.2
1992	28.0	0.1	26.5	0.1	36.2	0.1	28.5	0.1	36.8	0.1
1993	23.3	0.1	21.7	0.1	29.4	0.1	23.3	0.0	33.1	0.2
1994	29.6	0.1	28.1	0.1	34.2	0.1	30.1	0.1	41.1	0.3
1995	27.4	0.1	25.4	0.1	33.4	0.3	25.9	0.1	40.1	0.2
1996	25.1	0.1	22.7	0.1	27.6	0.2	25.3	0.1	37.8	0.3
1997	28.4	0.1	23.9	0.1	30.5	0.1	27.2	0.1	34.4	0.4
1998	30.6	0.1	25.5	0.1	36.4	0.4	27.7	0.1	36.5	0.3
1999	27.7	0.2	25.2	0.2	32.8	0.1	27.6	0.2	40.6	0.1
2000	23.7	0.2	20.3	0.1	27.9	0.3	23.9	0.2	36.0	0.1
2001	21.6	0.1	19.7	0.1	30.8	0.2	22.7	0.1	35.7	0.1

Table 2.3. The inshore RTI estimates of years 1985 through 2001, and their standard deviations. They were estimated with the cumulative run size being standardized at the final day of the return season. 'All five' denotes the lumped five districts, and 'Four districts' means the lumped four districts where the Togiak fish were excluded.

Year	All five		Four districts	
	RTI	S.D.	RTI	S.D.
1985	27.9	0.3	26.8	0.2
1986	30.5	0.2	29.7	0.1
1987	30.1	0.3	28.8	0.3
1988	29.1	0.3	28.3	0.3
1989	28.2	0.3	27.0	0.3
1990	30.1	0.2	29.2	0.2
1991	30.0	0.3	28.4	0.2
1992	31.1	0.4	29.9	0.4
1993	26.0	0.3	24.6	0.3
1994	32.2	0.3	30.6	0.2
1995	30.1	0.4	27.8	0.3
1996	27.1	0.4	25.2	0.2
1997	28.7	0.3	27.7	0.2
1998	31.0	0.4	30.0	0.4
1999	30.4	0.4	28.5	0.2
2000	26.1	0.4	23.9	0.2
2001	25.7	0.4	23.6	0.3

Table 2.4. The Port Moller RTI estimates of years 1985 through 2001 and their standard deviations, evaluated at day codes 10, 15, 20, and 25, respectively.

Year	Day code 10		Day code 15		Day code 20		Day code 25	
	RTI	S.D.	RTI	S.D.	RTI	S.D.	RTI	S.D.
1985	7.2	0.1	10.0	0.2	14.3	0.3	16.6	0.2
1987	7.2	0.3	12.2	0.2	15.5	0.1	16.3	0.1
1988	5.9	0.1	10.7	0.3	15.0	0.2	16.6	0.2
1989	7.2	0.2	10.4	0.2	13.4	0.2	16.3	0.2
1990	7.6	0.2	11.6	0.2	16.0	0.2	18.7	0.1
1991	8.2	0.2	11.8	0.1	13.6	0.1	15.7	0.2
1992	7.7	0.3	10.9	0.2	13.1	0.1	15.8	0.2
1993	6.7	0.1	10.0	0.2	12.7	0.2	15.7	0.2
1994	7.7	0.2	12.2	0.2	15.0	0.1	18.2	0.1
1995	7.1	0.2	9.6	0.1	13.1	0.2	15.8	0.2
1996	6.8	0.1	9.8	0.1	12.8	0.2	16.2	0.2
1997	7.0	0.2	10.6	0.2	13.9	0.2	17.0	0.2
1998	7.3	0.1	10.7	0.1	14.3	0.2	17.6	0.2
1999	7.1	0.4	11.9	0.2	14.4	0.1	17.3	0.1
2000	5.6	0.1	9.5	0.4	10.8	0.3	13.3	0.2
2001	7.3	0.2	9.7	0.1	12.2	0.1	14.0	0.1

Table 2.5. Relative index of fish run timing of years 1999, 2000, and 2001 evaluated at day codes, 15, 20, 25, and 30, respectively. The minus sign (-) indicates that run timing of the season is earlier than the average of those of the past years. These results are to be used in Chapters 3 and 4.

Year	June 24	June 29	July 4	July 9
1999	1.1	0.4	0.6	1.3
2000	-1.4	-3.3	-3.4	-3.8
2001	-1.1	-1.6	-2.4	-3.3

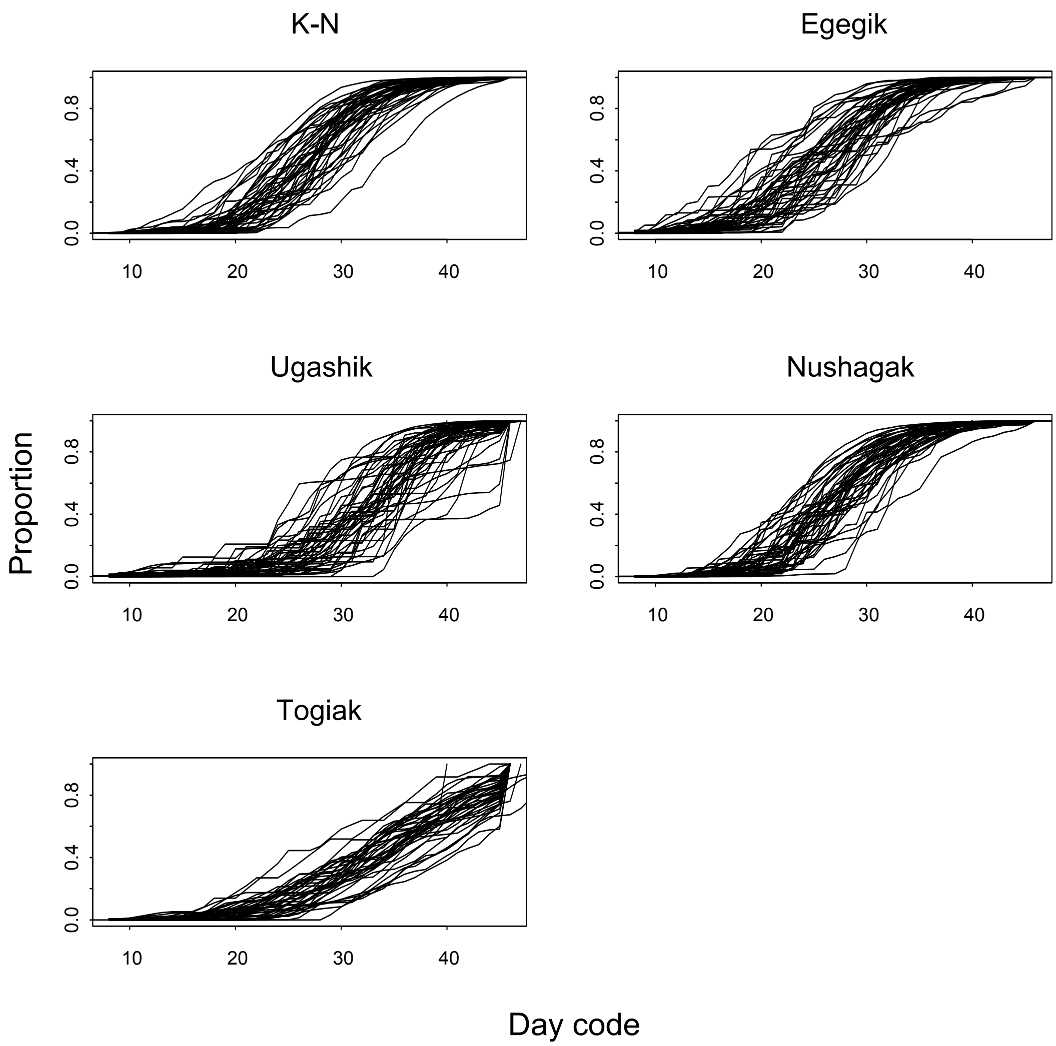


Figure 2.1. Historical run proportions of five stocks against day. The data of years 1955 through 2001 are used except for the Togiak stock. The data of years 1955 through 1957 for the Togiak stock are not available.

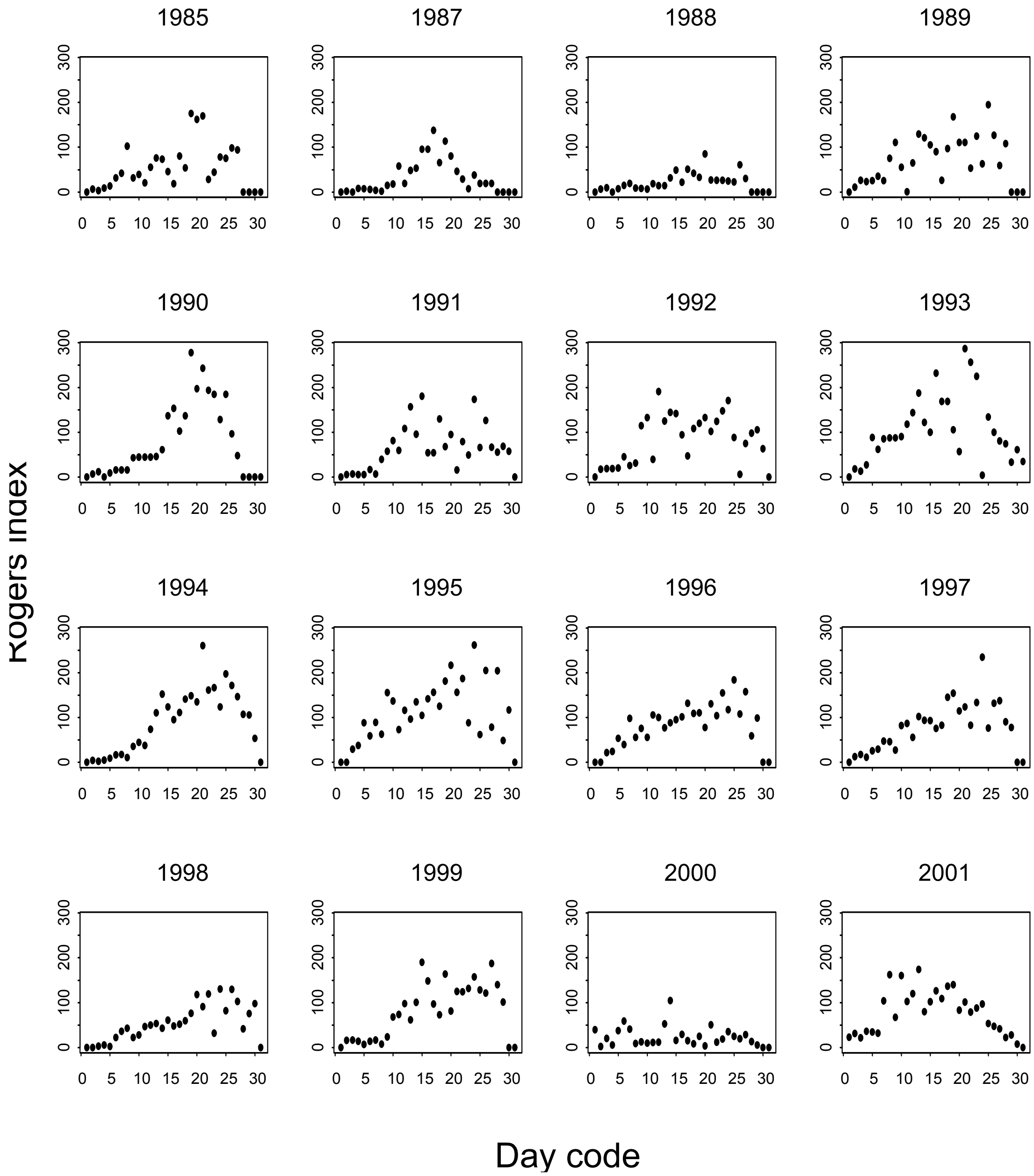


Figure 2.2. Port Moller Rogers' CPUE index against day by year. The Rogers' index unit is '$(6,000 \times \text{catch})/[200 \text{ fathoms} \times \text{fishing time (min)}]$,' where 6,000 is a scale factor. I code calendar dates, starting on June 10: day code 1 = June 10, day code 2 = June 11, and so on.

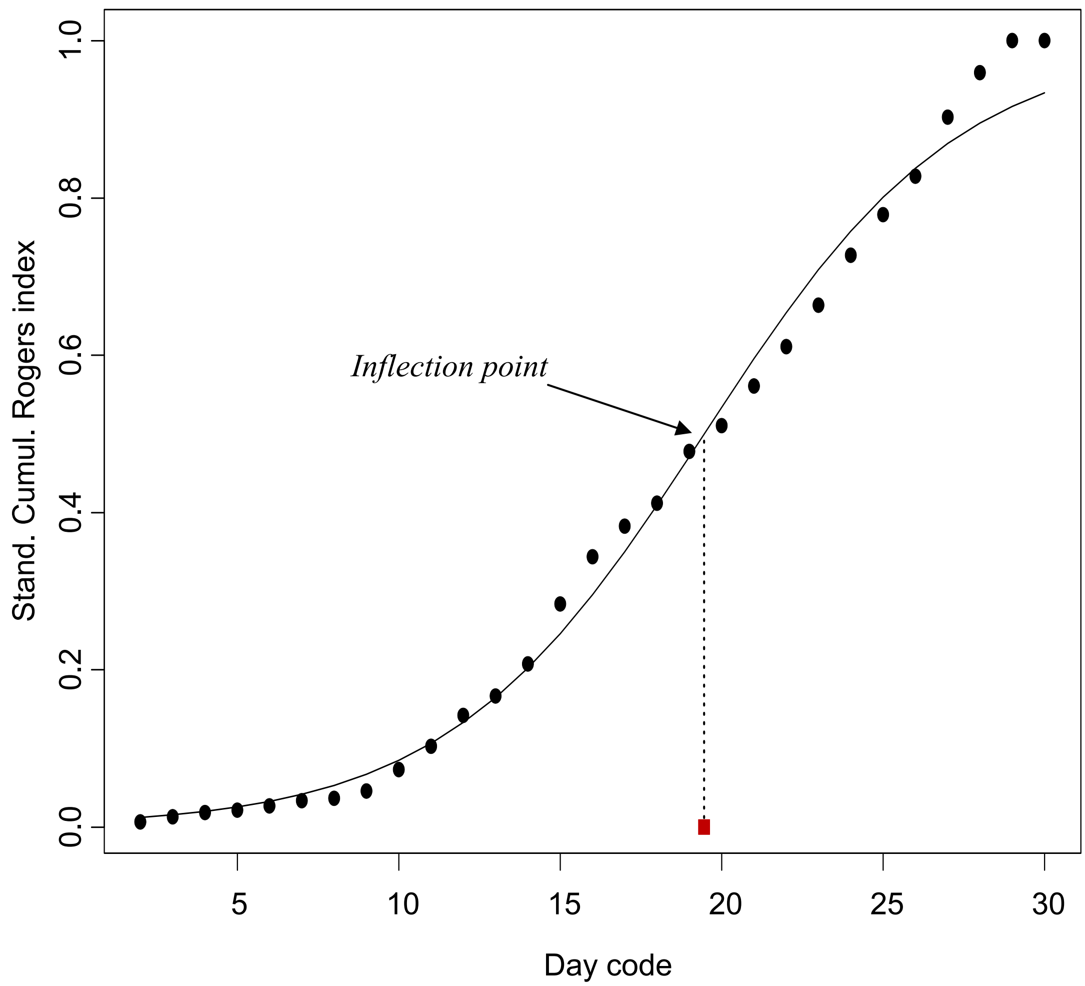


Figure 2.3. The 1999 Port Moller RTI estimate. Dots represent the data of the 1999 cumulative Rogers' indices that are standardized at day code 30 (July 9). The solid line is the logistic curve fitted to the dots. I define RTI as day that corresponds to the inflection point of the fitted logistic curve. The inflection point of the curve is located at the coordinates of (19.45, 0.5), and the RTI estimate is 19.45 (square point).

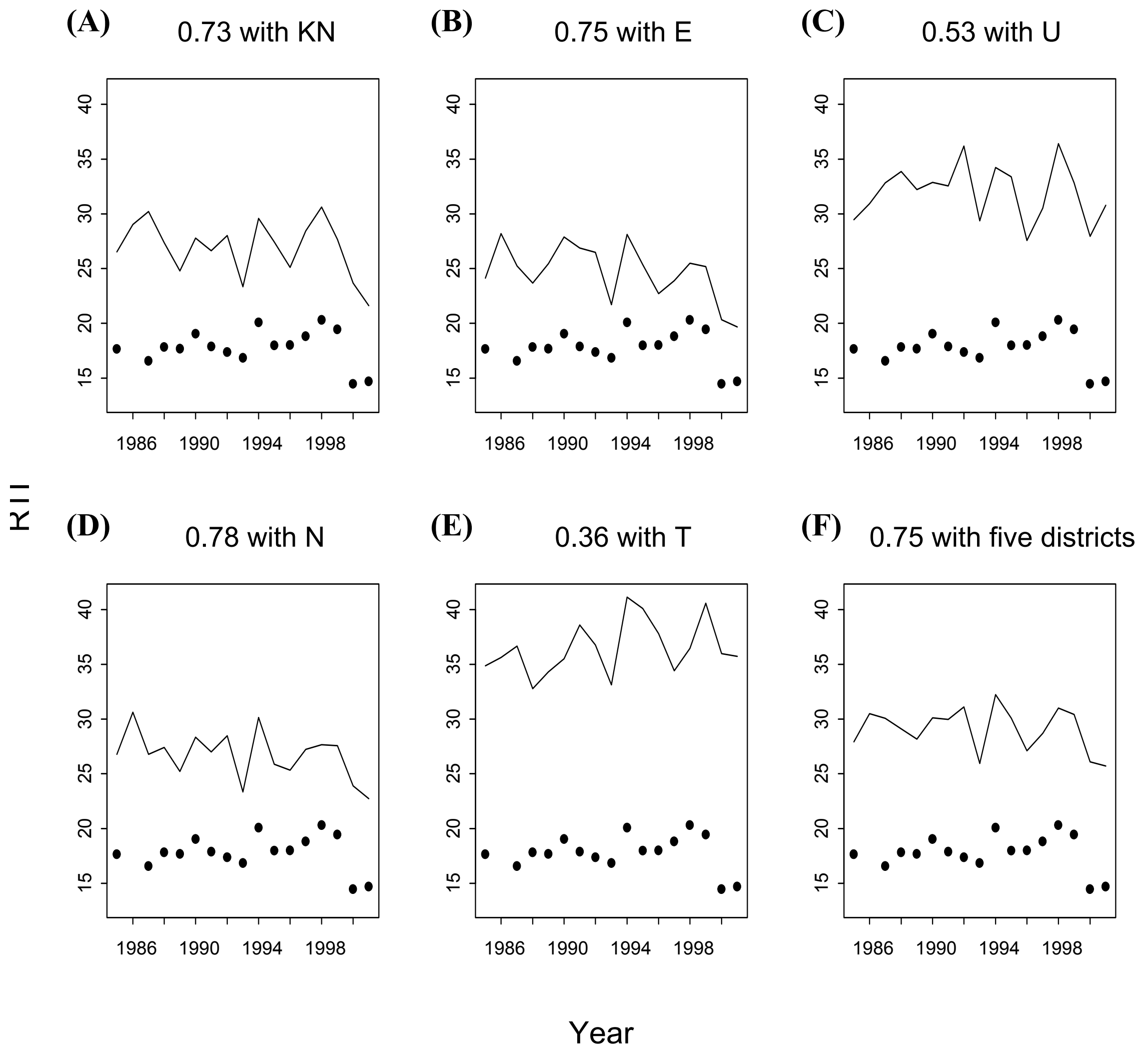


Figure 2.4. Comparison of estimates of the Port Moller RTI of years 1985 through 2001 and those of the inshore RTI of the same period. Dots represent the Port Moller RTI estimates. Lines in the boxes of (A) through (F) indicate the RTI estimates of 'Kvichak-Naknek,' 'Egegik,' 'Ugashik,' 'Nushagak,' 'Togiak,' and 'the lumped five districts,' respectively. The numerical value above the respective box is the correlation coefficient between the dot line and the solid line within the box.

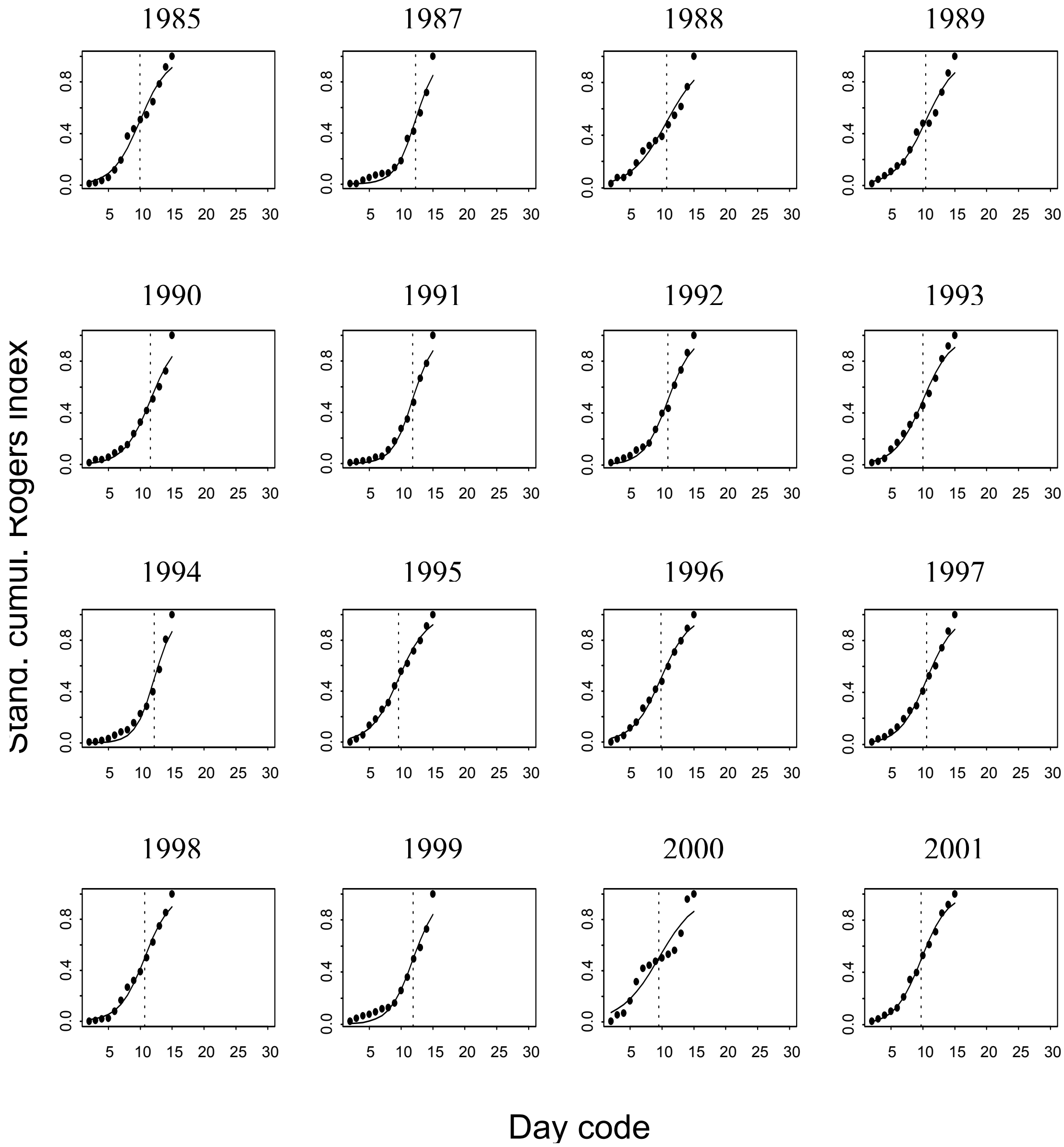


Figure 2.5. An example of forecasting run timing in year 2001 on June 24 (day code 15). When I want to detect run timing of the 2001 season at June 24, I standardize the cumulative Rogers' indices only up to the same day in the past years (before 2001), ignoring the data beyond the day. Intentionally I extend the *x*-axis beyond the day code 15 to show the idea. And then I compare the RTI estimate of the season (2001) with those of the past years. The vertical lines intersect the inflection points of the fitted logistic curves.

(A)

(B)

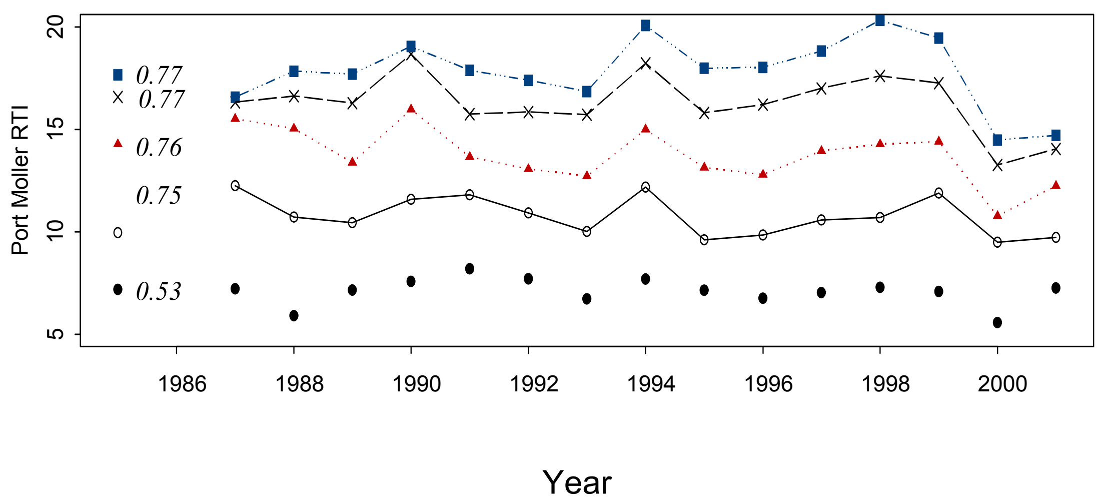


Figure 2.6. (A): Yearly inshore RTI estimates of four stocks (excluding the Togiak stock). (B): Yearly Port Moller RTI estimates evaluated at five day codes 10 (dot line), 15 (circle line), 20 (triangle line), 25 (cross mark line), and 30 (square line). The numerical values on the five lines in the (B) box are the correlation coefficients between the respective line and the inshore RTI in the (A) box.

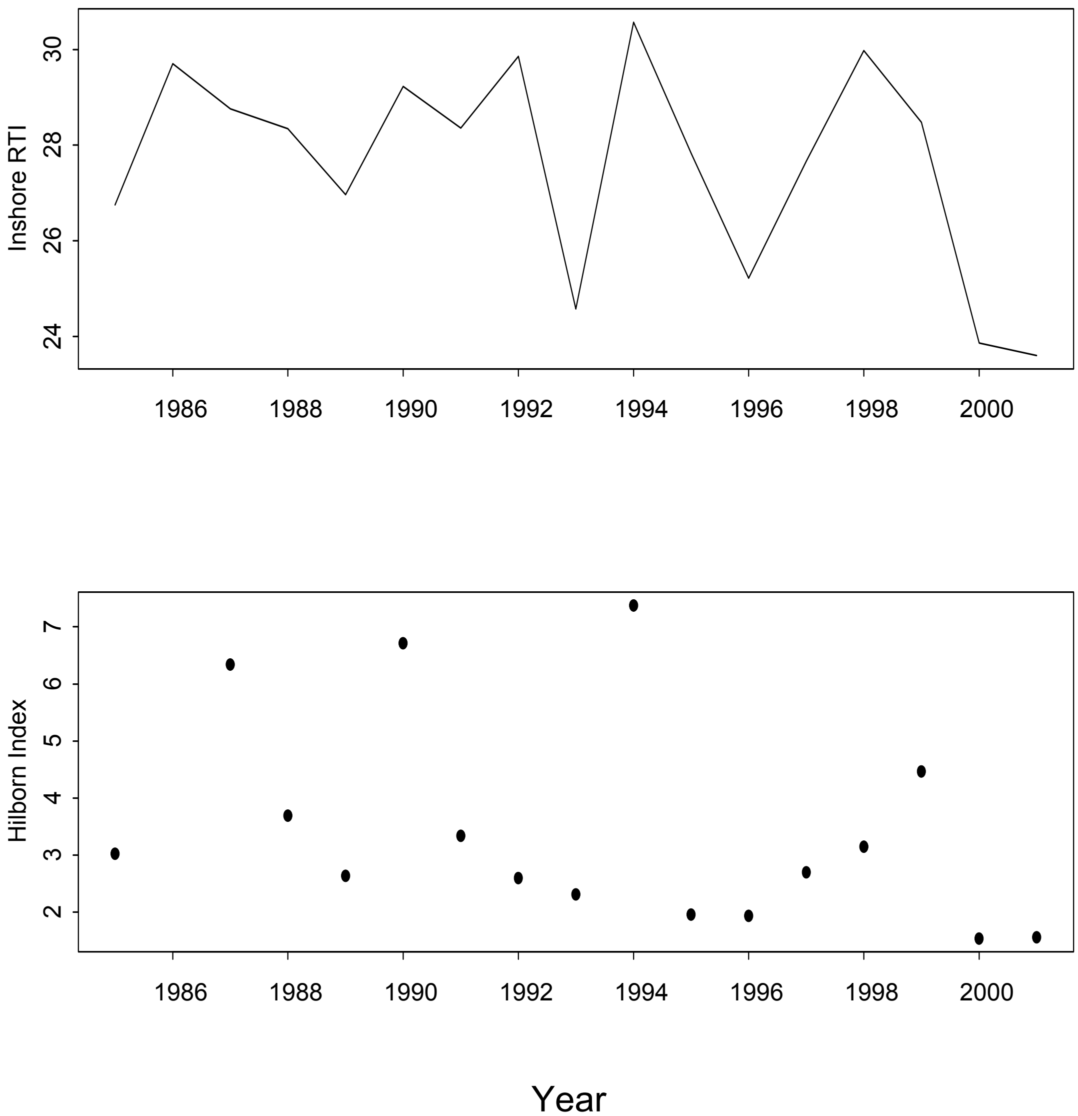


Figure 2.7. (A): Yearly inshore RTI estimates of four stocks (excluding the Togiak stock). (B): Yearly Hilborn RTI estimates. The correlation coefficient between the two lines is 0.68.

CHAPTER III. INSEASON FORECASTS OF RETURNS BY OPTIMIZATION

INTRODUCTION

The objective of this chapter is to estimate stock-specific run sizes on a daily basis. The term, 'stock' in this thesis means district-specific fish. I use all available data sources, with which I develop the objective functions of run sizes. The data sets include the following categories: (1) the catch of the Port Moller test fishery, (2) the age-specific proportions in the Port Moller fishery catch, (3) the catch of commercial and subsistence fisheries, (4) escapements, and (5) the age-specific proportions in stock-specific run size (= catch + escapement). In case of data categories (1), (3), and (4), I use not only the inseason data but also the historical data. We determine fish age by reading fish scales. ADFG collects scales of fish randomly chosen out of the Port Moller catch, the district-specific catch and the escapement fish, and then reports the age-specific proportions on a daily basis.

The main method of this thesis is an optimization technique. As optimization software, I use Automatic Differentiation Model Builder (ADMB) (Anonymous 1994, 2000). ADMB has the following merits: (1) to estimate many parameters or many predictive variables in a non-linear model, (2) to provide not only point estimates but also their variances, (3) to be less sensitive to initial guess values of estimates than other optimization software, and (4) to calculate Bayes' posterior distributions of estimates.

METHODS

3.1. DEFINITION OF TERMS

3.1.1. Variables, parameters, and objective functions

In defining variables and parameters over time, I follow the definitions by Gelman et al. (1995). Figure 3.1 shows the relationship between variables and parameters, the estimation of parameters, and the prediction of unobserved data. Parameters are involved with a function between explanatory variables and response variables. The estimation of parameters is based only on observed data. Once parameters are estimated, we are often interested in predicting 'unobserved response variables' from the function with the parameter estimates[1] and explanatory variables. The unobserved response variables are called the 'predictive variables,' and thus the density or distribution of the predictive variables is called the predictive density or distribution. Gelman et al. (1995) add tilde mark (~) to unobserved variables to distinguish them from observed variables. In this thesis, I estimate 20 predictive variables that are run sizes of five districts and four ages (Figure 3.2).

In section 3.4. 'Objective functions,' I develop the predictive densities of run sizes and the likelihood functions of run sizes. With the densities and likelihood functions of run sizes, we are interested in finding modes (run sizes) that maximize those functions. In this thesis, the objective function of run sizes means both the predictive density and the likelihood function. In most optimization software including ADMB, the objective function is used as its negative logarithm for ease of calculation. In this case, our interest is to find values that minimize the negative objective functions.

[1] The term 'estimate' is different from the term 'estimator.' An estimator is a function of a sample, and an estimate is the realized value of an estimator obtained when a sample is actually taken (Casella and Berger 1990).

3.2. NOTATIONS

The following list shows general notations used in this thesis. Notations valid only in a local subsection may not be found here.

Notation	Description
s	Stock (district)-specific fish. Five stocks were considered. Stock code 1 = Kvichak-Naknek; stock code 2 = Egegik; stock code 3 = Ugashik; stock code 4 = Nushagak; and stock code 5 = Togiak.
a	Age. Four age groups were considered. Age code 1 = age 1.2; age code 2 = age 1.3; age code 3 = age 2.2; and age code 4 = age 2.3.
$r_{s,a}$	Run size of stock s and age a.
$r_{s,\bullet}$	Stock-specific run size ignoring age: $r_{s,\bullet} = \sum_{a=1}^{4} r_{s,a}$
$r_{\bullet,a}$	Age-specific run size ignoring stock: $r_{\bullet,a} = \sum_{s=1}^{5} r_{s,a}$
R	Total run size. The sum of district- and age- specific run sizes: $R = \sum_{s=1}^{5}\sum_{a=1}^{4} r_{s,a}$
t	Time (day). Calendar dates were coded, starting on June 10: June 10 = 1, June 11 =2, and so on.
D_t	Rogers' index at day t. Rogers' index is the weighted CPUE from the Port Moller fishery (see Equation 1.1).
I_t	The cumulative Rogers' index up to day t.
~	Tilde mark (~) refers to an unknown variable. For example, $\tilde{R}$ represents unknown (predictive) total run size while R is known (observed) total run size.
^	Circumflex mark (^) refers to the estimate of an unknown value such as a parameter or a predictive variable. For example, $\hat{\beta}$ is the estimate of β.
$U_{a,t}$	The cumulative number of age a fish caught by the Port Moller fishery up to day t.

$U_{\bullet,t}$	The cumulative catch of the Port Moller fishery up to day *t*, ignoring age.
G_a	The selectivity of the Port Moller gillnet fishery for age *a* fish.
k_t	The proportion of run size that pass a location of interest at day *t*; day-specific proportion of run size.
$j_{s,a,t}$	The observed cumulative run of stock *s* and age *a* up to day *t*. I caution readers not to be confused with the above run sizes ('*r*' and '*R*'). The '*r*' or '*R*' indicates final run size, which is the cumulative run up to the end of the season.
$j_{s,\bullet,t}$	The observed cumulative run of stock *s* up to day *t*.
$h_{s,t,i}$	The cumulative proportion of run size of stock *s* up to day *t* in past year *i*.

3.3. MAIN IDEA

The general idea of the methodology is as follows:

Step 1. I develop the objective functions (the predictive densities or the likelihood functions) of run sizes.

Step 2. I take the negative logarithms of the respective objective functions, and treat the sum of the negative logarithm functions as the joint objective function.

Step 3. With ADMB, I look for run sizes, which minimize the joint objective function.

Step 4. I do the estimation of Step 3 on a daily basis during the season. As the inseason data are updated, the estimation is supposed to improve.

3.4. OBJECTIVE FUNCTIONS

3.4.1. Predictive density of total run size

Total run size means the sum of district- and age- specific runs in the 20 cells in Figure 3.2.

$$R = \sum_{s=1}^{5}\sum_{a=1}^{4} r_{s,a} \tag{3.1}$$

The ordinary regression model between total run size and the cumulative Rogers' index is significant (Rogers and Steen 2000). I set up the ordinary regression model of total run size against the cumulative Rogers' index up to June 20 through July 6, respectively with the data of years 1985 through 2001 excluding the outlier years (1990, 1994, 1997, and 2001) (Figure 3.3). The determination coefficient (R^2) of the regression model ranged from 0.65 to 0.82 (Figure 3.3). The 1986 Port Moller test fishery data were not available because the test fishery was not deployed in that year. I used the following ordinary regression model to develop the predictive density of total run size.

$$\begin{aligned} &R = \beta_{0,t} + \beta_{1,t} \cdot I_t + \varepsilon_t \\ &R \sim N(\beta_{0,t} + \beta_{1,t} \cdot I_t,\ \sigma_t^{2}) \end{aligned} \tag{3.2}$$

where R is the total run sizes of the past years, I_t is the cumulative Rogers' indices of the corresponding historical years at day t, '$\beta_{0,t}$, $\beta_{1,t}$, and σ_t^2' are parameters, and ε_t is the error term. The expected value of R is '$\beta_{0,t} + \beta_{1,t} \cdot I_t$', and its variance is σ_t^2. The estimates of these three parameters vary by day t, and thus the parameters have subscript t in Equation 3.2.

When we have new catch data during the season and predict total run size of the year at day t, I use the empirical relationship of Equation 3.2. Thus, the predictive density of unknown variable $\tilde{R}$ is normal.

$$f(\tilde{R}) = \frac{1}{\sqrt{2\pi \cdot \hat{\sigma}_t^{2}}} \exp\left[-\frac{\left(\tilde{R} - (\hat{\beta}_{0,t} + \hat{\beta}_{1,t} \cdot I_t)\right)^2}{2 \cdot \hat{\sigma}_t^{2}}\right] \tag{3.3}$$

Taking the negative logarithm of the equation, and ignoring the constant terms with respect to $\tilde{R}$, we get the following:

$$-\ln f(\vec{\tilde{r}}) \propto \frac{\left[\left(\sum_s \sum_a \tilde{r}_{s,a}\right) - \hat{\beta}_{0,t} - \hat{\beta}_{1,t} \cdot I_t\right]^2}{2 \cdot \hat{\sigma}_t^{2}} \tag{3.4}$$

I replaced $\tilde{R}$ by '$\sum_{s=1}^{5}\sum_{a=1}^{4}\tilde{R}_{s,a}$' in Equation 3.4.

'$2\cdot\hat{\sigma}_t^{\,2}$' in the numerator of Equation 3.4 cannot be ignored though it is constant with respect to $\tilde{R}$. Only in the single Equation 3.4, the term is constant with respect to $\tilde{R}$ whereas it is not constant in the joint objective function. The term cannot be factored out from the joint objective function.

The following two illustrations may help readers understand the previous sentences.

(1) The following is a function of x, comparable to Equation 3.3.

$$f_1(x)=\exp\left[-\frac{g(x)}{2\cdot\sigma^2}\right]$$

When taking the negative logarithm to the function $f_1(x)$, we have the following:

$$-\ln f_1(x)=\frac{g(x)}{2\cdot\sigma^2}$$

'$2\sigma^2$' is constant with respect to x in this case.

(2) The following is the function $f_1(x)$ times another function of x.

$$f_1(x)f_2(x)=\exp\left[-\frac{g(x)}{2\cdot\sigma^2}\right]\cdot f_2(x)$$

When taking the negative logarithm to the above function, we have the following:

$$-\ln\left[f_1(x)f_2(x)\right]=\frac{g(x)}{2\cdot\sigma^2}-\ln f_2(x)$$

In this case, '$2\sigma^2$' cannot be factored out, and is not constant with respect to x.

Equation 3.4 is the first component of the joint objective function, where I_t is the observed value (data), and '$\hat{\beta}_{0,t}$, $\hat{\beta}_{1,t}$, and $\hat{\sigma}_t^{\,2}$' are the estimated values. I show the parameter estimation in sub-section, 3.5.1. 'Parameters in the predictive density of total run size.'

3.4.2. Likelihood function of age-specific run sizes

The ADFG used fish scales sampled from the Port Moller fishery catch to determine fish age. I found that the age composition of the Port Moller fishery catch generally matched that of returns to Bristol Bay. Figure 3.4 compares the proportions of four age groups (ages 1.2, 1.3, 2.2, and 2.3) of the Port Moller catch with those of returns to Bristol Bay. Regarding the Port Moller data in Figure 3.4, I used the fish only caught before July 5, because sockeye salmon caught after July 5 are likely to return to an area other than Bristol Bay (Hilborn, *personal communication*). In Figure 3.4, age-specific proportions of the Port Moller catch are similar to those of returns to Bristol Bay except for years 1997 and 1998.

I modeled the joint probability distribution of age-specific catches with the multinomial probability mass function. When $U_{a,t}$ denotes the age-specific cumulative catch of the Port Moller fishery up to day t, the joint probability of the age-specific cumulative catches is as follows.

$$f(\vec{U}_t) = \frac{U_{\bullet,t}!}{\prod_{a=1}^{4}\left(U_{a,t}!\right)} \prod_{a=1}^{4}\left[P_{a,t}^{\;U_{a,t}} \right] \tag{3.5}$$

where $P_{a,t}$ is the proportion of age a fish out of the cumulative Port Moller catch up to day t. Equation 3.5 is also the likelihood function of $P_{a,t}$.

$$L(\vec{P}_t) \propto \prod_{a=1}^{4} P_{a,t}^{\;U_{a,t}} \tag{3.6}$$

As the negative log-likelihood,

$$-l(\vec{P}_t) \propto -\sum_{a=1}^{4} [U_{a,t} \cdot \ln(P_{a,t})] \tag{3.7}$$

The maximum likelihood estimate (MLE) of the proportion in the multinomial distribution is as follows.

$$\hat{P}_{a,t} = \frac{U_{a,t}}{U_{\bullet,t}} \tag{3.8}$$

I re-parameterized the age-specific proportion $P_{a,t}$ to have the function with respect to run sizes.

$$
\begin{aligned}
\hat{P}_{a,t} &= \frac{U_{a,t}}{U_{\bullet,t}} = \frac{U_{a,t}}{\sum_{a=1}^{4} U_{a,t}} \\
&= \frac{\sum_{d=1}^{t} r_{\bullet,a}\cdot k_d \cdot V \cdot G_a}{\sum_{a=1}^{4}\sum_{d=1}^{t} r_{\bullet,a}\cdot k_d \cdot V \cdot G_a} = \frac{r_{\bullet,a}\cdot G_a}{\sum_{a=1}^{4} r_{\bullet,a}\cdot G_a}
\end{aligned}
\tag{3.9}
$$

where k_d is the proportion of run size that passes the Port Moller fishery area at day d (day-specific proportion of run size), V is fish vulnerability to the Port Moller fishery, and G_a is the fishery selectivity for age a fish. I assumed that V and G_a were constant regardless of day within a year. Fish vulnerability, V may vary by year because ocean environmental variables are not constant by year. However, the uncertainty in the vulnerability V does not cause a problem, because the vulnerability is canceled out in the numerator and the denominator (Equation 3.9). The gillnet gear selectivity for age 2.3 fish is assumed to be full (i.e. $G_4 = 1$), so there are three parameters: G_1, G_2, and G_3.

Replacing $P_{a,t,}$ in Equation 3.7 by that of Equation 3.9, I have the following likelihood function of run sizes:

$$
-l(\vec{\hat{r}}_{\bullet}) \propto -\sum_{a=1}^{4}\left[U_{a,t}\cdot \ln\left(\frac{\hat{r}_{\bullet,a}\cdot \hat{G}_a}{\sum_{a}\hat{r}_{\bullet,a}\cdot \hat{G}_a}\right)\right]
\tag{3.10}
$$

By the invariance property of MLE (Zehna 1966), the estimates of $\vec{r}_{\bullet}$ in the new likelihood function also become MLE. Now, the original likelihood of age-specific proportions (Equation 3.7) becomes the likelihood of age-specific runs (Equation 3.10). Equation 3.10 is the second component of the joint objective function, where $U_{a,t}$ is the observed value (data) and $\hat{G}_a$ is the estimated value. I describe the estimation of G_1, G_2, and G_3 in sub-section '3.5.2. Parameters in the likelihood function of age-specific runs.'

3.4.3. Predictive density of stock-specific run size

Stock-specific run size means the sum of catch and escapement that belong to its district (Figure 1.2). In Figure 3.2, stock-specific run size is the row sum. That is,

$$r_{s,\bullet} = \sum_{a=1}^{4} r_{s,a} \tag{3.11}$$

To forecast stock-specific run size at an arbitrary day (say t) during season, I used the cumulative run up to day t during the season, and the historical cumulative proportions of the run at day t. If we observe the cumulative run data up to a day during the season, and know the cumulative proportion of the run at the day, we can estimate final run size by dividing the cumulative run by the proportion. I used historical data to calculate the cumulative proportion of run size by day. The run data of year 1955 through the present were available except for the Togiak stock (those of three years 1955, 1956 and 1957 were not available for the Togiak stock). Figure 2.1 shows the historical cumulative proportions of run sizes by day for the five stocks. Large variability in the proportion by day was found in all five stocks (Figure 2.1). When estimating final run size with the observed cumulative run of stock s up to day t and the historical cumulative run proportions at the day t, I used all the proportions rather than the mean value of the proportions to carry the variability in the proportion. That is,

$$\begin{aligned} \vec{\hat{r}}_{s,\bullet} &= \frac{\text{observed cumulative run of stock } s \text{ up to day } t \text{ during the season}}{\text{historical cumulative proportions of the stock } s \text{ run at day } t} \\ &= \frac{j_{s,\bullet,t}}{\vec{h}_{s,t}} = \frac{j_{s,\bullet,t}}{\{h_{s,t,i} \mid i = \text{a past year } (1,\ldots,n)\}} \end{aligned} \tag{3.12}$$

The numerator is a scalar value, and the denominator is a vector. Thus, the resultant dimension is also a vector: $\vec{\hat{r}}_{s,\bullet} = \{\hat{r}_{s,\bullet,i} \mid i = \text{an individual index in the sample}\}$. $\hat{r}_{s,\bullet,i}$ is an ith element out of the estimates of final run size of stock s.

I could consider the histogram of the run size estimates ($\hat{r}_{s,\bullet,i}$) as an estimated distribution for stock-specific run size. For instance, Figure 3.5 shows distribution of the 1999 Egegik run size estimated at the specified day (June 24, June 30, July 6, July 12, July 18, and July 24). In Figure 3.5, the distribution predicted at June 30 extends beyond the *x*-axis limit, but I don't show the part for the same scale of plots in the left column. In Figure 3.5, the dotted vertical line represents the actual run size of Egegik stock in year 1999. The variance of the predictive run distribution was small during the initial and final stage of the return season, while it was large during the middle of the season (Figure 3.5).

I explored various parametric densities to develop the distribution of $\hat{r}_{s,\bullet}$: normal, gamma, lognormal, inverse Gaussian (Wald), and location gamma. The 'location gamma' density is named by me, and the term is not found in a statistics book.

(1) Normal

If $r_{s,\bullet} \sim N(\mu_{s,t}, \sigma_{s,t}^{2})$,

$$f(r_{s,\bullet}) = \frac{1}{\sqrt{2\pi \cdot \sigma_{s,t}^{2}}} \exp\left[-\frac{(r_{s,\bullet} - \mu_{s,t})^2}{2 \cdot \sigma_{s,t}^{2}}\right] \tag{3.13}$$

Though there is no limitation for the domain of a normal random variable, the domain of $r_{s,\bullet}$ is positive in this case. That is, $r_{s,\bullet} > 0$, $\mu_{s,t} > 0$, and $\sigma_{s,t}^{2} > 0$.

(2) Gamma

If $r_{s,\bullet} \sim gamma(\alpha_{s,t}, \beta_{s,t})$,

$$f(r_{s,\bullet}) = \frac{1}{\Gamma(\alpha_{s,t}) \cdot \beta_{s,t}^{\alpha_{s,t}}} r_{s,\bullet}^{(\alpha_{s,t}-1)} \cdot \exp\left(-\frac{r_{s,\bullet}}{\beta_{s,t}}\right) \tag{3.14}$$

where $r_{s,\bullet} > 0$, $\alpha_{s,t} > 0$, and $\beta_{s,t} > 0$. Some statistics textbooks show a different gamma density from the above density by using different parameterization: i.e., $\beta^{*} \equiv (1/\beta)$.

(3) Lognormal

If $r_{s,\bullet} \sim lognormal(\mu_{s,t}, \sigma_{s,t}^{2})$,

$$f(r_{s,\bullet}) = \frac{1}{\sqrt{2\pi \cdot \sigma_{s,t}{}^{2}}} \cdot \frac{1}{r_{s,\bullet}} \cdot \exp\left[-\frac{(\ln r_{s,\bullet} - \mu_{s,t})^2}{2 \cdot \sigma_{s,t}{}^{2}} \right] \tag{3.15}$$

where $r_{s,\bullet} > 0$, $-\infty < \mu_{s,t} < \infty$, and $\sigma_{s,t}{}^{2} > 0$.

(4) Inverse Gaussian (Wald)

If $r_{s,\bullet} \sim \textit{inverse Gaussian}(\mu_{s,t}, \lambda_{s,t})$,

$$f(r_{s,\bullet}) = \sqrt{\frac{\lambda_{s,t}}{2\pi \cdot r_{s,\bullet}{}^{3}}} \cdot \exp\left[\frac{-\lambda_{s,t} \cdot (r_{s,\bullet} - \mu_{s,t})^2}{2 \cdot \mu_{s,t}{}^{2} \cdot r_{s,\bullet}} \right] \tag{3.16}$$

where $r_{s,\bullet} > 0$, $\mu_{s,t} > 0$, and $\lambda_{s,t} > 0$. $\mu_{s,t}$ is a measure of location and $\lambda_{s,t}$ is a reciprocal measure of dispersion. Equation 3.16 is a standard form of the inverse Gaussian distribution (Johnson and Kotz 1970, p. 138).

(5) Location gamma (*ad hoc term*)

When the parameter α in the ordinary gamma density (Equation 3.14) is large, we can never have an asymmetric gamma distribution. As the value of α increases, the shape of a gamma density becomes a symmetric shape. If a random variable, say *T*, is exponential (i.e. $T \sim \varepsilon(\beta)$), then the random variable is also a gamma variable where the parameter α of the gamma distribution is 1 (i.e. $T \sim gamma(1, \beta)$). When T_i are independent, $\sum_{i=1}^{\alpha} T_i \sim gamma(\alpha, \beta)$. Thus when the value of α is large, the sum of T_i should approach a normal (symmetric) distribution by the central limit theorem.

To overcome this problem where an ordinary gamma density with a large value of α cannot be asymmetric, first I shifted the distribution of $\hat{r}_{s,\bullet}$ in Equation 3.12 by subtracting the minimum value of $\hat{r}_{s,\bullet}$ from all values of $\hat{r}_{s,\bullet}$. The shift changes only the location of the distribution but not its shape. I fit an ordinary gamma density to the 'shifted' distribution of $\hat{r}_{s,\bullet}$; I calculated two parameters of the ordinary gamma density with the shifted data. The parameter value α estimated in the shifted distribution was small enough to produce an asymmetric shape. After fitting an ordinary gamma density to the shifted distribution, I moved both the gamma density and the shifted distribution of $\vec{\hat{r}}_{s,\bullet}$ back to its original site. Thus, the modified gamma density has one more parameter besides the original two parameters, α and β. The additional parameter was the minimum value of the original values, $\vec{\hat{r}}_{s,\bullet}$. I call the modified gamma density the location gamma density. Figure 3.6 illustrates the procedure that I have described so far. The histogram in Figure 3.6 is the 1999 Egegik run distribution estimated at day code 30 (July 9) by Equation 3.12. In Figure 3.6, even without doing a goodness of fit test, we can see that the location gamma density fits a very skewed distribution much better than the ordinary gamma density.

I formalize the location gamma density as follows. If $r_{s,\bullet}$ ~ *location gamma*$(\alpha_{s,t}, \beta_{s,t}, \gamma_{s,t})$, where

$$\gamma_{s,t} = \min\left[\vec{\hat{r}}_{s,\bullet}\right] = \min\left[\frac{j_{s,\bullet,t}}{\vec{h}_{s,t}}\right] \tag{3.17}$$

I shift the frequency distribution of $\vec{\hat{r}}_{s,\bullet}$ in Equation 3.12 by subtracting $\gamma_{s,t}$ from each of $\vec{\hat{r}}_{s,\bullet}$. That is,

$$\left(\vec{\hat{r}}_{s,\bullet} - \gamma_{s,t}\right) = \left(\frac{j_{s,\bullet,t}}{\vec{h}_{s,t}} - \min\left[\frac{j_{s,\bullet,t}}{\vec{h}_{s,t}}\right]\right) \tag{3.18}$$

The parameters $\alpha_{s,t}$ and $\beta_{s,t}$ are estimated with the shifted vector, '$\vec{\hat{r}}_{s,\bullet} - \gamma_{s,t}$'. Thus, a location Gamma density with α, β, and γ is an ordinary gamma density shifted by γ. Letting f^* be an ordinary gamma density, I express a location gamma density f as follows.

$$\begin{aligned} f(r_{s,\bullet}) &= f^*(r_{s,\bullet} - \gamma_{s,t}) \\ &= \frac{1}{\Gamma(\alpha_{s,t}) \cdot \beta_{s,t}^{\alpha_{s,t}}} (r_{s,\bullet} - \gamma_{s,t})^{(\alpha_{s,t}-1)} \cdot \exp\left[-\frac{(r_{s,\bullet} - \gamma_{s,t})}{\beta_{s,t}}\right], \quad \text{if } r_{s,\bullet} \geq \gamma_{s,t} \\ &= 0, \quad \text{otherwise} \end{aligned} \tag{3.19}$$

Evaluation of the five parametric densities

I used Kolmogorov-Smirnov goodness of fit test to evaluate the above five densities. As an example, I fit the five densities to the 1999 Egegik run distribution predicted at day codes 15 (June 24), 20 (June 29), 25 (July 4), 30 (July 9), 35 (July 14), and 40 (July 19), respectively. Table 3.1 shows the results of the K-S test for those densities fitted to the 1999 Egegik run distribution; the larger p-value is, the better the fit is. The lognormal density shape always turned out to be almost identical to that of the inverse Gaussian density, so I did not distinguish them differently. Figure 3.7 shows an example where the five densities are fitted to the 1999 Egegik run distribution estimated at day code 30 (July 9). For the distribution estimated at the initial stage of the season (day codes 15 and 20), the best fit was the lognormal density (or the inverse Gaussian density) (p-values 0.581, and 0.245 in Table 3.1) while, for the distribution estimated at other days (day codes 25, 30, and 35), the best fit was the location gamma density (p-values 0.570, 0.655, and 0.400 in Table 3.1). For the distribution estimated at day code 40, every density fit poorly (p-values 0.009, 0.001, 0.000, and 0.001 in Table 3.1). The poor fit was due to a very narrow distribution around the mode of the run estimates. The run distribution estimated near the season end was extremely narrowed (e.g., in Figure 3.5, the run distributions estimated at July 18 and July 24). However, the poor fit is not a problem, because the run size near the season end becomes so obvious that we don't need to forecast it.

Generally the location gamma density and the lognormal density (or the inverse Gaussian density) fit the predicted distribution of stock-specific run size well. Figure 3.8 represents the average value of p-values of five tests in Table 3.1, except for the last test for the distribution predicted at day code 40. The location gamma density and the lognormal (or the inverse Gaussian) density were much better than the gamma density and the normal density in terms of goodness of fit (Figure 3.8). However, I had a problem in ADMB programming when I used the location gamma density for the predictive density of stock-specific run size. The location gamma density was defined over two separate domains (Equation 3.19). The separation prevented ADMB from differentiating the location gamma density with respect to run size over the *smooth* continuous domain.

This situation compelled me to use the lognormal density or the inverse Gaussian density for the predictive density of stock-specific run size. However, the lognormal density is more common than the inverse Gaussian, and the former density is implemented in most statistical software. Thus, I chose the lognormal density. Taking the negative logarithm of the lognormal density (Equation 3.15) and ignoring constants with respect to run size, we have the following.

$$-\ln f(\tilde{N}_{s,\bullet}) \propto \left[\ln \tilde{N}_{s,\bullet} + \frac{(\ln \tilde{N}_{s,\bullet} - \hat{\mu}_{s,t})^2}{2 \cdot \hat{\sigma}_{s,t}^{\ 2}} \right] \tag{3.20}$$

where $\tilde{N}_{s,\bullet} > 0$, $-\infty < \hat{\mu}_{s,t} < \infty$, and $\hat{\sigma}_{s,t}^{\ 2} > 0$. Equation 3.20 is another component of the joint objective function. Because there were five districts, I had to consider the respective five lognormal objective functions. $\hat{\mu}_{s,t}$ and $\hat{\sigma}_{s,t}^{\ 2}$ are estimates of $\mu_{s,t}$ and $\sigma_{s,t}^{\ 2}$, and I describe the estimation in sub-section, '3.5.3. Parameters in the predictive density of stock-specific run size.'

3.4.4. Likelihood function of stock- and age- specific run sizes

Age composition from stock-specific run data was also available. I applied the multinomial mass function to the joint probability distribution of stock- and age- specific proportions. The principle is the same as that of the likelihood function of age-specific proportions in the Port Moller fishery catch (Equation 3.5). That is, the joint probability distribution of the cumulative stock-specific and age-specific runs is:

$$f(\vec{j}_{s,t}) = \frac{j_{s,\bullet,t}\,!}{\prod_{a=1}^{4} \left(j_{s,a,t}\,! \right)} \prod_{a=1}^{4} \left[P_{s,a,t}^{\ \ j_{s,a,t}} \right] \tag{3.21}$$

where $j_{s,a,t}$ is the cumulative run of age *a* fish to district *s* up to day *t*, and $P_{s,a,t}$ is the proportion of age *a* fish out of the cumulative run to district *s* up to day *t*. Equation 3.21 also is the likelihood function of $P_{s,a,t}$.

$$L(\vec{P}_{s,t}) \propto \prod_{a=1}^{4}\left[P_{s,a,t}{}^{j_{s,a,t}} \right] \tag{3.22}$$

Considering the MLE of $P_{s,a,t}$ and re-parameterizing it with run sizes of interest, we have the following.

$$\begin{aligned} \hat{P}_{s,a,t} &= \frac{j_{s,a,t}}{j_{s,\bullet,t}} = \frac{j_{s,a,t}}{\sum_{a} j_{s,a,t}} \\ &= \frac{\sum_{d=1}^{t} r_{s,a} \cdot k_d}{\sum_{a}\sum_{d=1}^{t} r_{s,a} \cdot k_d} = \frac{r_{s,a}}{\sum_{a} r_{s,a}} \end{aligned} \tag{3.23}$$

where k_d is the proportion of run size that enters district s at day d (day-specific proportion of district-specific run size). Finally, I replaced the proportion in the likelihood function (Equation 3.22) by the relation in Equation 3.23, and took the negative logarithm of the function:

$$-l(\vec{\tilde{r}}_s) \propto -\sum_{a=1}^{4}\left[j_{s,a,t} \cdot \ln\left(\frac{\tilde{r}_{s,a}}{\sum_{a} \tilde{r}_{s,a}} \right) \right] \tag{3.24}$$

In Equation 3.24, $j_{s,a,t}$ is observed data, and $\tilde{r}_{s,a}$ is the predictive variable in the objective function. In this case, I did not consider the fishery selectivity for age-specific fish because run data were not only from the gillnet fishery but also from research beach seines. Equation 3.24 is the last component of the joint objective function. I had to consider the respective five multinomial objective functions because there were five districts.

Table 3.2 contains the summary of the objective functions I have described so far. The joint objective function is the sum of the negative logarithms of the respective twelve objective functions: Equation 3.4, Equation 3.10, five of Equation 3.20, and five of Equation 3.24.

3.5. PARAMETERS

There were 16 parameters in the objective functions: three in the predictive density of total run size (Equation 3.4), three in the likelihood function of age-specific runs (Equation 3.10), and 10 in the predictive densities of the respective five stock-specific runs (five Equation 3.20 for each stock).

3.5.1. Parameters in the predictive density of total run size

The predictive density of total run size was normal (Equation 3.4). It had three parameters: σ_t^2, $\beta_{0,t}$, and $\beta_{1,t}$. The likelihood function of these three parameters is:

$$L(\beta_{0,t},\beta_{1,t},\sigma_t^2) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{\sigma_t^2}} \exp\left[-\frac{\left(\left(\sum_s \sum_a r_{s,a,i}\right) - \beta_{0,t} - \beta_{1,t} \cdot I_{t,i}\right)^2}{2 \cdot \sigma_t^2} \right] \tag{3.25}$$

'$\left(\sum_s \sum_a r_{s,a,i}\right)$' is total run size of past year i, and $I_{t,i}$ is the cumulative Rogers' index at day t of the corresponding year i. As the negative log likelihood,

$$-l(\beta_{0,t},\beta_{1,t},\sigma_t^2) \propto \left[\frac{n}{2} \cdot \ln \sigma_t^2 + \frac{1}{2 \cdot \sigma_t^2} \sum_{i=1}^{n} \left(\left(\sum_s \sum_a r_{s,a,i}\right) - \beta_{0,t} - \beta_{1,t} \cdot I_{t,i}\right)^2 \right] \tag{3.26}$$

Though the formulae for ML estimators of the parameters were not necessary due to the benefit of ADMB, I derived them to check the parameter units. After differentiating the above negative log likelihood with respect to the three parameters, and setting them equal to zero,

$$\frac{\partial(-l)}{\partial \beta_{0,t}} = 0; \quad \frac{\partial(-l)}{\partial \beta_{1,t}} = 0; \quad \frac{\partial(-l)}{\partial \sigma_t^2} = 0$$

I solved these equations for the parameters. The solutions are as follows:

$$\hat{\beta}_{1,t} = \frac{\sum_i \left[\left(I_{t,i} - \frac{\sum_i I_{t,i}}{n} \right) \left(\sum_s \sum_a r_{s,a,i} - \frac{\sum_i \sum_s \sum_a r_{s,a,i}}{n} \right) \right]}{\sum_i \left(I_{t,i} - \frac{\sum_i I_{t,i}}{n} \right)^2} \tag{3.27}$$

$$\hat{\beta}_{0,t} = \frac{\sum_i \sum_s \sum_a r_{s,a,i}}{n} - \hat{\beta}_{1,t} \cdot \frac{\sum_i I_{t,i}}{n} \tag{3.28}$$

$$\hat{\sigma}_t^{\,2} = \frac{\sum_i \left[\left(\sum_s \sum_a r_{s,a,i} \right) - \hat{\beta}_{0,t} - \hat{\beta}_{1,t} \cdot I_{t,i} \right]^2}{n} \tag{3.29}$$

Thus, the units of $\beta_{1,t}$, $\beta_{0,t}$ and σ_t^2 are:

$$\left[\beta_{1,t} \right] = \frac{fish\ run\ size}{Rogers'\ index} \tag{3.30}$$

$$\left[\beta_{1,t} \right] = fish\ run\ size \tag{3.31}$$

$$\left[\sigma_t^{\,2} \right] = (fish\ run\ size)^2 \tag{3.32}$$

where ‘$[\ ]$’ denotes the unit notation.

Note that, in the above likelihood function (Equation 3.26), there is no tilde mark (~) for $r_{s,a,i}$, because they are observed values, not predictive values.

3.5.2. Parameters in the likelihood function of age-specific runs

I had three parameters of G_1, G_2, and G_3 in the likelihood function of age-specific runs (Equation 3.10). Subscripts 1, 2, and 3 are age code. These parameters are the Port Moller fishery selectivity for age-specific fish. In case of the selectivity for age 2.3 fish, the full selectivity is assumed: i.e. $G_4 = 1$ where subscript 4 is the age 2.3 code. The following is the likelihood function of the parameters:

$$L(\vec{G}) \propto \prod_{i=1}^{n} \prod_{a=1}^{4} \left[\frac{r_{\bullet,a,i} \cdot G_a}{\sum_a r_{\bullet,a,i} \cdot G_a} \right]^{U_{a,i}} \tag{3.33}$$

$r_{\bullet,a,i}$ is age-specific run size of past year i, and $U_{a,i}$ is age-specific catch from the Port Moller fishery in the corresponding year i. I assumed that the selectivity is constant by time (year as well as day), so G did not have a time subscript. As the negative log likelihood, Equation 3.33 becomes the following.

$$-l(\vec{G}) \propto -\sum_{i=1}^{n}\sum_{a=1}^{4}\left[U_{a,i}\cdot\ln\left(\frac{r_{\bullet,a,i}\cdot G_a}{\sum_a r_{\bullet,a,i}\cdot G_a}\right)\right] \tag{3.34}$$

Differentiating the above negative log likelihood with respect to G_a, and setting it equal zero, I have an implicit equation for G_a. However, the derivation for checking the parameter units is not needed because the selectivity is a fraction whose range is from 0 to 1.

$r_{\bullet,a,i}$ in the above likelihood function (Equation 3.34) does not have a tilde mark(~) because they are observed values.

3.5.3. Parameters in the predictive density of stock-specific run size

The predictive density of stock-specific run size was lognormal (Equation 3.20). It had two parameters in the predictive density for each stock (${\sigma_{s,t}}^2$ and $\mu_{s,t}$), so there were five pairs (i.e., 10 parameters) for five stocks. The likelihood function of those two parameters is:

$$L(\mu_{s,t},{\sigma_{s,t}}^2) \propto \prod_{i=1}^{n}\frac{1}{\sqrt{{\sigma_{s,t}}^2}}\exp\left[-\frac{\left(\ln\hat{r}_{s,\bullet,i}-\mu_{s,t}\right)^2}{2\cdot{\sigma_{s,t}}^2}\right] \tag{3.35}$$

where $\hat{r}_{s,\bullet,i}$ is an *i*th element out of the estimates of final run size of stock *s* (Equation 3.12). As the negative log likelihood, Equation 3.35 becomes the following:

$$-l(\mu_{s,t},{\sigma_{s,t}}^2) \propto \left[\frac{n}{2}\ln{\sigma_{s,t}}^2+\frac{1}{2\cdot{\sigma_{s,t}}^2}\sum_{i=1}^{n}(\ln\hat{r}_{s,\bullet,i}-\mu_{s,t})^2\right] \tag{3.36}$$

I differentiated the above negative log likelihood with respect to the two parameters, set them equal to zero,

$$\frac{\partial(-l)}{\partial\mu_{s,t}}=0;\quad \frac{\partial(-l)}{\partial{\sigma_{s,t}}^2}=0$$

and solved these equations for the parameters. I had the following solutions:

$$\hat{\mu}_{s,t} = \frac{\sum_i \ln(\hat{r}_{s,\bullet,i})}{n} \tag{3.37}$$

$$\hat{\sigma}_{s,t}{}^2 = \frac{\sum_i \left(\ln \hat{r}_{s,\bullet,i} - \hat{\mu}_{s,t}\right)^2}{n} \tag{3.38}$$

Thus, the units of $\mu_{s,t}$, and $\sigma_{s,t}{}^2$ are:

$$\left[\mu_{s,t}\right] = \ln(\textit{fish run size}) \tag{3.39}$$

$$\left[\sigma_{s,t}{}^2\right] = \left[\ln(\textit{fish run size})\right]^2 \tag{3.40}$$

The estimation of $\mu_{s,t}$, and $\sigma_{s,t}{}^2$ are involved not only with historical data ($h_{s,t,i}$) but also with inseason data ($j_{s,\bullet,t}$) (recall Equation 3.12), while the estimation of the other parameters ($\beta_{1,t}$, $\beta_{0,t}$, σ_t^2, G_1, G_2, and G_3) requires only historical data.

3.6. INCORPORATION OF RUN TIMING FORECAST

In Chapter 2, I forecasted fish run timing with Port Moller fishery data. According to run timing forecast, I use historical data that belong to a different day from a forecast day. Figure 3.9 illustrates the idea. Normally, when forecasting run sizes at day t during the season, I estimate parameters with historical data that correspond to day t in past years, and then predict run sizes with the estimates of parameters and inseason data (Figure 3.9 (A)). If I detect run timing earlier or later by q days, I use historical data that correspond to day '$t \pm q$' in past years (Figure 3.9 (B)).

There are three historical data sources: (1) total run size of year i ($\sum_s \sum_a r_{s,a,i}$), (2) the cumulative Rogers' CPUE up to day t in year i ($I_{t,i}$), and (3) the cumulative proportion of run size of stock s at day t in year i ($h_{s,t,i}$). The values of the first two sources ($\sum_s \sum_a r_{s,a,i}$ and $I_{t,i}$) determine the estimates of $\beta_{0,t}$, $\beta_{1,t}$, and σ_t^2 in the normal predictive density of total run size (Equations 3.27, 3.28, and 3.29). The values of the third historical data source ($h_{s,t,i}$) and the observed cumulative run of stock s up to day t

during the season ($j_{s,\bullet,t}$) determine the estimates of the five pairs of $\mu_{s,t}$ and $\sigma_{s,t}^{2}$ in the respective lognormal predictive densities of stock-specific run sizes (Equations 3.37 and 3.38, where $\hat{r}_{s,\bullet,i} = j_{s,\bullet,t} / h_{s,t,i}$). G_1, G_2, and G_3 in the likelihood function of age-specific run sizes are not affected by the run timing incorporation because the three parameters are constant over time.

For example, if I make forecasts of the 2000 run sizes at July 4, I use the historical data of $I_{t,i}$ and $h_{s,t,i}$ that correspond to July 7 not to July 4: i.e. t = July 7, and i = years prior to forecast year 2000. The 2000 Port Moller RTI evaluated at July 4 was earlier by about three days than the average of those in the past years (-3.4 in Table 2.5).

RESULTS

By the hind-casting procedure, I made forecasts of returns of years 1999, 2000, and 2001 at the following days per season (year): day codes 15 (June 24), 20 (June 29), 25 (July 4) and 30 (July 9).

3.7. PARAMETER ESTIMATES

Tables 3.4, 3.5 and 3.6 show the MLEs and standard deviations for parameter estimates, calculated with the likelihood functions in section '3.5. Parameters.' The units of parameter estimates are summarized in Table 3.4. The MLEs are used when forecasting returns of years 1999 (Table 3.4), 2000 (Table 3.5), and 2001 (Table 3.6). The respective table has mainly two columns: 'With' and 'Without.' The values under the 'With' column were estimated with the run timing incorporation, while those under the 'Without' column were estimated without the incorporation. The units of the estimates are shown in Table 3.5. Regarding the Port Moller fishery selectivity for age specific fish (G_a), I show the estimates only once per table because the values are constant by day within a season. In case of forecasting the 1999 returns at June 29, the parameter estimates and the run forecasts under the 'With' column were the same as those under the 'Without' column (Table 3.4), because the Port Moller RTI detected at June 29, 1999 was almost the same as the average of those in the past years (0.4 in Table 2.5).

3.8. FORECASTS OF RETURNS

Tables 3.7 through 3.12 show forecasts of stock- and age- specific returns. The marginal values under the 'sum' column indicate the stock-specific run forecasts, in which ADFG managers are most interested. Tables 3.13, 3.14 and 3.15 compare the stock-specific run forecasts with the actual run sizes, where the difference between the forecast and the actual run size is expressed as relative error (%):

$$\text{Relative error (\%)} = \left(\frac{\text{forecast - actual run}}{\text{actual run}}\right)\times 100 \tag{3.41}$$

The minus (-) sign in an error value indicates an under-forecast.

Generally forecasts (run estimates) approached their actual run sizes as time progressed (Tables 3.13, 3.14, and 3.15; Figures 3.10, 3.11, and 3.12). Absolute values of errors in forecasts made at day code 15 (June 24) ranged from about 5% to about 640%, and those in forecasts made at day code 30 (July 9) ranged from about 1% to about 60%. Forecasts of returns to Togiak district had larger errors than those to the other districts.

3.8.1. Incorporation of run timing forecast

Forecasts of returns with the run timing incorporation were generally less biased than those without the incorporation, except for forecasts of the 1999 runs and of the Ugashik and Togiak returns. Tables 3.13, 3.14, and 3.15 have forecasts of stock-specific returns for years 1999, 2000, and 2001, which were calculated *with* the run timing incorporation and *without* the incorporation, respectively. Figures 3.10, 3.11, and 3.12 show the respective summary of Tables 3.13, 3.14, and 3.15. In those three figures, the horizontal dotted line represents the actual run size, and cross mark (×) points are forecasts of returns with the run timing incorporation while square marks are those without the incorporation. In case of the 1999 run forecasts (Figure 3.10), differences between cross marks and square marks are not significant because the Port Moller RTI of 1999 was almost the same as the average of those of the past years (1.1, 0.4, 0.6, and 1.3 in Table 2.5). However, in case of the 2000 and 2001 run forecasts (Figures 3.11 and 3.12), cross marks are closer to the actual run size than square marks, except for the Ugashik and Togiak stocks.

3.8.2. Incorporation of the Port Moller fishery selectivity for age-specific fish

The incorporation of the Port Moller fishery selectivity for age-specific fish did not improve forecasts of returns. Tables 3.16, 3.17, and 3.18 have the 1999, 2000, and 2001 run forecasts, which were calculated *with* the selectivity incorporation and *without* the incorporation, respectively. When I estimated forecasts of returns ignoring the selectivity, I let the selectivity parameters one: i.e. $G_1 = G_2 = G_3 = 1$. The forecast error values (%) under the 'With' column were not significantly different from those under the 'Without' column (Tables 3.16, 3.17, and 3.18). In both cases, I incorporated the run timing forecast accordingly.

Also the selectivity incorporation did not affect age composition (proportion) in forecasts of returns. In Figures 3.13 (forecasts of the 1999 returns made at July 4), 3.14 (forecasts of the 2000 returns made at July 4), and 3.15 (forecasts of the 2001 returns made at July 4), age-specific proportions in the forecasts made with the selectivity incorporation (solid lines) are compared with those made without the selectivity (dashed lines). In case of the 'Port Moller' boxes in those figures, the lines represent age composition in forecasts of age-specific run sizes ($r_{\bullet,a}$). Dots indicate age composition in observed values (under the 'Port Moller' label, dots are those in the cumulative Port Moller catch up to July 4; under the district name, dots are those in the observed cumulative run to the corresponding district up to July 4). In Figures 3.13, 3.14, and 3.15, solid lines and dashed lines are very close to each other, and they match dots well. Tables 3.19, 3.20, and 3.21 show the proportion values used to draw Figures 3.13, 3.14, and 3.15.

DISCUSSION

3.9. ASSUMPTION OF THE JOINT OBJECTIVE FUNCTION

It would be vulnerable to criticism to treat the joint objective function as the product of the respective objective functions (in the logarithm, as the sum of the respective objective functions). The treatment is based on the assumption where data sets in the joint objective function are independent of one another. The following is the joint probability function of the respective data sets where district- and age- specific run sizes are involved as predictive variables or parameters.

$$\begin{aligned}&\Pr\left(\text{data}_1, \text{data}_2, ..., \text{data}_k\right)\\ &= \prod_{i=1}^{k} \Pr\left(\text{data}_i\right) \qquad (\because \textit{independence})\end{aligned} \tag{3.42}$$

Each ‘ $\Pr(\text{data}_i)$ ’ is the respective objective function before the transformation of the negative logarithm: Equations 3.3, 3.5, 3.15, and 3.21, respectively (also see Equation 4.2 for a different expression).

Rigorously speaking, the independence assumption in Equation 3.42 is not correct. For example, the age composition data of the Port Moller fishery catch are not independent of those of observed run sizes to the five districts. However, the catch of the Port Moller test fishery is usually small, and the catch abundance is not correlated with the observed run sizes especially during the initial stage of the season. Also the run size and age composition of each district are independent of those of the other. Thus, the violation of independence is not serious.

3.10. SAMPLE SIZE IN THE MULTINOMIAL PROBABILITY FUNCTION

Another obstacle to forecasts of returns was data sizes in the multinomial likelihood functions: $U_{\bullet,t} = \sum_a U_{a,t}$ in Equation 3.5, and $r_{s,\bullet,t} = \sum_a r_{s,a,t}$ in Equation 3.21. It is an inherent problem that occurs when a probability distribution of age composition (age-specific groups in number or proportion) from fisheries data is assumed as a multinomial distribution. Usually the sampling designs deployed to collect data, along with the selection protocols utilized in the field, generate estimates of age composition that necessarily depart, to some degree, from a strictly theoretical multinomial distribution (Crone and Sampson 1998). The expanse and dynamics of fisheries prevent us from sampling in a strictly random manner.

Thus, we are advised not to use a real catch size but to scale down the size for total sample size in a multinomial probability (Crone and Sampson 1998). If I used the real data size (the real catch for $U_{\bullet,t}$ in Equation 3.5, and the real run for $j_{s,\bullet,t}$ in Equation 3.21), I would give an over-weight to the multinomial objective function. As a result, the multinomial objective function would dominate the joint objective function. A question would be raised in response to the advice: 'how much should we scale down the sample size?' or 'what is the optimum sample size that most accurately describes the actual variability associated with the sample estimates of age composition?' To determine the optimum sample size, Crone and Sampson (1998) used weighted nonlinear regression analysis with the actual variance measures (e.g. CV) associated with the sample estimates (proportions) of age composition. However, I could not apply the idea because of absence of the data. By a process of trial and error, I set the sizes. Fortunately, estimates of age-specific proportions in Port Moller catch and stock-specific run sizes turned out to be extremely close to the observed values (Figures 3.13, 3.14, and 3.15). Also, forecasts of returns approached the actual returns as forecast time progressed during the season (Figures 3.10, 3.11, and 3.12); forecasts made at day code 30 (July 9) were close to the actual returns.

3.11. SELECTIVITY OF THE PORT MOLLER GILLNET FISHERY

Because the incorporation of the Port Moller fishery selectivity (G_a) for age-specific fish did not reduce bias in forecasts of runs, the parameters did not draw attention. However, the estimates of G_a may be useful for other research. There appears to be no formal report regarding the selectivity of the Port Moller gillnet gear. I succeeded in finding MLE of G_a, assuming the full selectivity for age 2.3 fish ($G_4 = 1$):

G_1= 0.557, G_2 = 0.837, and G_3 = 0.630 in Table 3.6, where subscripts 1, 2 and 3 denote ages 1.2, 1.3, and 2.2. Because of the hind-casting procedure, the estimates associated with the run forecasts of years 1999 (Table 3.4), 2000 (Table 3.5), and 2001 (Table 3.6) are a little different. The difference is negligible: G_1 = 0.553, G_2 = 0.862, G_3 = 0.616 for the 1999 forecasts (Table 3.4); G_1 = 0.560, G_2 = 0.851, G_3 = 0.631 for the 2000 forecasts (Table 3.5). Their likelihood profiles are to be shown in Chapter 4 (Figures 4.1, 4.2, and 4.3).

The mean and standard deviation of length of age-specific sockeye salmon caught by the Port Moller fishery during the 1999 season was as follows: 505.5 mm and 24.43 mm for age 1.2 fish (sample size: 1,738); 557.5 mm and 30.49 mm for age 1.3 fish (sample size: 835); 516.7 mm and 23.28 mm for age 2.2 fish (sample size: 1,021); 563.6 mm and 30.43 mm for age 2.3 fish (sample size: 348). Figure 3.16 displays the relation between fish length, fish age, and the selectivity of the gillnet fishery, though the statistical measures (R^2 = 0.94; p-value = 0.03) are not meaningful because there are only four observations. Ocean age-3 fish (1.3 and 2.3) are remarkably larger than ocean age-2 fish (1.2 and 2.2) (Figure 3.16).

3.12. FORECASTS OF RETURNS

Estimates of district-specific run sizes approached the actual run sizes as time progressed during the season (Figures 3.10, 3.11, and 3.12). The approach is not surprising because data are accumulated and updated over time.

An important finding was that the run timing incorporation improved forecasts of run sizes, except for forecasts of the 1999 runs and of the Ugashik and Togiak runs. Despite that Port Moller RTI is a little biased from true run timing (Figure 2.6), forecasts of returns made with the Port Moller RTI incorporation were less biased than those without the incorporation. If we incorporated *true* run timing, we could further reduce bias in forecasts of runs.

The 1999 Port Moller RTI evaluated at day codes 15 (June 24), 20 (June 29), 25 (July 4), and 30 (July 9) were not different from the average of the past years (1.1, 0.4, 0.6, and 1.3 in Table 2.5). That's why the run timing incorporation did not make significant differences in the 1999 run forecasts (Figure 3.10).

The Ugashik and Togiak fish return significantly later by a few days than the other stocks (Figure 1.3). Yearly Port Moller RTI were poorly correlated with yearly RTI of both Ugashik and Togiak stocks (Figure 2.4: $r = 0.53$ with Ugashik, and $r = 0.36$ with Togiak). Thus, the run timing adjustment on the basis of the Port Moller RTI (Table 2.5) did not improve forecasts of the Ugashik and Togiak runs.

Table 3.1. Results of the K-S test of the five densities fitted to the 1999 Egegik run distribution estimated at day codes 15 (June 24), 20 (June 29), 25 (July 4), 30 (July 9), 35 (July 14), and 40 (July 19), respectively. In each cell, the upper value and the lower value inside parentheses are p-value and K-S test statistic.

Density	Day 15 p-value (K-S)	Day 20 p-value (K-S)	Day 25 p-value (K-S)	Day 30 p-value (K-S)	Day 35 p-value (K-S)	Day 40 p-value (K-S)
Location gamma	0.029 (0.222)	0.028 (0.216)	0.570 (0.115)	0.655 (0.107)	0.400 (0.131)	0.009 (0.244)
Lognormal or Inv. Gaussian	0.581 (0.118)	0.245 (0.151)	0.462 (0.125)	0.293 (0.144)	0.031 (0.213)	0.001 (0.284)
Gamma	0.022 (0.230)	0.014 (0.233)	0.244 (0.151)	0.187 (0.160)	0.023 (0.221)	0.000 (0.391)
Normal	0.000 (0.351)	0.000 (0.356)	0.056 (0.203)	0.076 (0.189)	0.012 (0.236)	0.001 (0.293)

Table 3.2. Summary of the objective functions. After taking the negative logarithm of the respective objective function, the sum of the negative logarithm functions is the joint objective function.

Data source	Variable of objective function	Objective function nature
Port Moller test fishery	Total run size	Normal predictive density function
	Age-specific run sizes	Multinomial likelihood function
Historical daily proportions of district run; district fisheries and escapements	District-specific run size	Lognormal predictive density function
District fisheries and escapements	District- and age- specific run sizes	Multinomial likelihood function

Table 3.3. Units of parameter estimates in Tables 3.4 through 3.6 and Figures 4.1 through 4.3. Subscript *t* represents a date.

Parameter	Unit	Description
$\beta_{0,t}$ $\beta_{1,t}$ σ_t^2	millions millions/Rogers index millions^2	These parameters are from the normal predictive density of the total run (Equation 3.4).
$\mu_{1,t}$ $\sigma_{1,t}^2$	[ln(thousands)] [ln(thousands)]^2	These parameters are from the log-normal predictive density of district-specific run (Equation 3.20). Subscript 1 denotes the Kvichak-Naknek stock.
$\mu_{2,t}$ $\sigma_{2,t}^2$	[ln(thousands)] [ln(thousands)]^2	These parameters are from the log-normal predictive density of district-specific run (Equation 3.20). Subscript 2 denotes the Egegik stock.
$\mu_{3,t}$ $\sigma_{3,t}^2$	[ln(thousands)] [ln(thousands)]^2	These parameters are from the log-normal predictive density of district-specific run (Equation 3.20). Subscript 3 denotes the Ugashik stock.
$\mu_{4,t}$ $\sigma_{4,t}^2$	[ln(thousands)] [ln(thousands)]^2	These parameters are from the log-normal predictive density of district-specific run (Equation 3.20). Subscript 4 denotes the Nushagak stock.
$\mu_{5,t}$ $\sigma_{5,t}^2$	[ln(thousands)] [ln(thousands)]^2	These parameters are from the log-normal predictive density of district-specific run (Equation 3.20). Subscript 5 denotes the Togiak stock.
G_1, G_2, G_3	Fraction whose range is from 0 to 1	Port Moller fishery selectivity for age-specific fish. These parameters are from the likelihood function of age-specific proportions (Equation 3.10). Subscripts 1, 2, and 3 denote age 1.2, age 1.3, and age 2.2, respectively.

Table 3.4. Point estimates (MLE) and their standard deviations for the parameters used to forecast the 1999 returns at June 24, June 29, July 4, and July 9, respectively. Subscript t corresponds to the respective forecast date. S.D. denotes standard deviation. The values under the 'With' column are associated with the run timing incorporation while those under the 'Without' column are not. Because the age-specific gillnet selectivity (G_a) is constant by day within the season, I show the values only once. In case of June 29, the estimates under the 'Without' column are the same as those under the 'With' column, because Port Moller RTI detected at June 29, 1999 was not significantly different from the overall RTI in the past years (Table 2.5).

		With		Without	
Parameter	**Date**	**MLE**	**S.D.**	**MLE**	**S.D.**
$\beta_{0,t}$	June 24	24.138	6.289	23.281	6.447
$\beta_{1,t}$		0.026	0.010	0.024	0.009
σ_t^2		108.330	0.000	106.510	0.000
$\mu_{1,t}$		8.902	0.224	8.566	0.193
$\mu_{2,t}$		7.008	0.171	6.781	0.174
$\mu_{3,t}$		8.754	0.134	8.709	0.148
$\mu_{4,t}$		7.335	0.221	6.943	0.218
$\mu_{5,t}$		4.518	0.144	4.198	0.132
$\sigma_{1,t}^2$		1.998	0.203	1.566	0.342
$\sigma_{2,t}^2$		1.167	0.261	1.242	0.274
$\sigma_{3,t}^2$		0.651	0.153	0.832	0.191
$\sigma_{4,t}^2$		2.000	0.007	2.000	0.012
$\sigma_{5,t}^2$		0.579	0.155	0.504	0.132
G_1		0.553	0.540		
G_2		0.862	0.679		
G_3		0.616	0.502		
$\beta_{0,t}$	June 29	14.797	7.377		
$\beta_{1,t}$		0.021	0.006		
σ_t^2		85.816	0.000		
$\mu_{1,t}$		7.787	0.179		
$\mu_{2,t}$		8.399	0.150		
$\mu_{3,t}$		7.901	0.145		
$\mu_{4,t}$		6.922	0.193		
$\mu_{5,t}$		5.378	0.195		
$\sigma_{1,t}^2$		1.405	0.300		
$\sigma_{2,t}^2$		0.996	0.212		
$\sigma_{3,t}^2$		0.887	0.194		
$\sigma_{4,t}^2$		1.640	0.350		
$\sigma_{5,t}^2$		1.181	0.300		

Table 3.4. (continued)

Parameter	Date	With MLE	With S.D.	Without MLE	Without S.D.
$\beta_{0,t}$		15.947	7.253	14.885	7.250
$\beta_{1,t}$		0.015	0.004	0.014	0.004
σ_t^2		88.133	0.000	84.592	0.000
$\mu_{1,t}$		10.095	0.091	9.874	0.082
$\mu_{2,t}$		9.613	0.073	9.463	0.071
$\mu_{3,t}$		6.978	0.150	6.733	0.140
$\mu_{4,t}$	July	9.451	0.106	9.233	0.101
$\mu_{5,t}$	4	4.799	0.126	4.571	0.108
$\sigma_{1,t}^2$		0.361	0.077	0.297	0.063
$\sigma_{2,t}^2$		0.234	0.050	0.221	0.047
$\sigma_{3,t}^2$		0.944	0.206	0.822	0.179
$\sigma_{4,t}^2$		0.495	0.106	0.446	0.095
$\sigma_{5,t}^2$		0.568	0.134	0.420	0.099
$\beta_{0,t}$		17.335	7.625	17.431	7.533
$\beta_{1,t}$		0.011	0.004	0.011	0.004
σ_t^2		97.238	0.000	96.550	0.000
$\mu_{1,t}$		9.866	0.051	9.752	0.042
$\mu_{2,t}$		9.353	0.040	9.258	0.032
$\mu_{3,t}$		8.382	0.122	8.168	0.111
$\mu_{4,t}$	July	9.161	0.049	9.048	0.039
$\mu_{5,t}$	9	5.582	0.085	5.430	0.071
$\sigma_{1,t}^2$		0.115	0.025	0.078	0.017
$\sigma_{2,t}^2$		0.069	0.015	0.045	0.010
$\sigma_{3,t}^2$		0.621	0.135	0.532	0.115
$\sigma_{4,t}^2$		0.108	0.023	0.065	0.014
$\sigma_{5,t}^2$		0.276	0.063	0.191	0.044

Table 3.5. Point estimates (MLE) and their standard deviations for the parameters used to forecast the 2000 runs at June 24, June 29, July 4, and July 9, respectively. Subscript t corresponds to the respective forecast date. S.D. denotes standard deviation. The values under the 'With' column are associated with the run timing incorporation while those under the 'Without' column are not. Because the age-specific gillnet selectivity (G_a) is constant by day within the season, I show the values only once.

		With		Without	
Parameter	**Date**	**MLE**	**S.D.**	**MLE**	**S.D.**
$\beta_{0,t}$	June 24	21.688	6.000	23.225	6.206
$\beta_{1,t}$		0.023	0.008	0.024	0.009
σ_t^2		90.060	0.000	99.056	0.000
$\mu_{1,t}$		10.953	0.175	11.469	0.190
$\mu_{2,t}$		9.926	0.168	10.188	0.179
$\mu_{3,t}$		8.711	0.137	8.835	0.145
$\mu_{4,t}$		7.382	0.197	8.091	0.216
$\mu_{5,t}$		6.387	0.153	6.663	0.144
$\sigma_{1,t}^2$		1.341	0.286	1.559	0.336
$\sigma_{2,t}^2$		1.277	0.269	1.339	0.292
$\sigma_{3,t}^2$		0.789	0.172	0.815	0.185
$\sigma_{4,t}^2$		1.716	0.366	2.000	0.008
$\sigma_{5,t}^2$		0.704	0.182	0.618	0.160
G_1		0.560	0.516		
G_2		0.851	0.657		
G_3		0.631	0.501		
$\beta_{0,t}$	June 29	14.722	6.690	14.802	7.135
$\beta_{1,t}$		0.016	0.004	0.021	0.006
σ_t^2		74.865	0.000	80.296	0.000
$\mu_{1,t}$		9.485	0.101	10.605	0.180
$\mu_{2,t}$		9.767	0.083	10.592	0.148
$\mu_{3,t}$		8.881	0.146	9.529	0.142
$\mu_{4,t}$		9.535	0.121	10.714	0.194
$\mu_{5,t}$		6.523	0.150	7.158	0.191
$\sigma_{1,t}^2$		0.461	0.097	1.455	0.307
$\sigma_{2,t}^2$		0.309	0.065	0.986	0.208
$\sigma_{3,t}^2$		0.911	0.196	0.870	0.188
$\sigma_{4,t}^2$		0.656	0.138	1.702	0.359
$\sigma_{5,t}^2$		0.810	0.191	1.166	0.291

Table 3.5. (continued)

Parameter	Date	With MLE	With S.D.	Without MLE	Without S.D.
$\beta_{0,t}$		17.022	7.308	14.943	7.026
$\beta_{1,t}$		0.011	0.004	0.014	0.004
σ_t^2		89.146	0.000	79.477	0.000
$\mu_{1,t}$		9.155	0.054	9.669	0.080
$\mu_{2,t}$		9.329	0.046	9.724	0.070
$\mu_{3,t}$		8.676	0.122	9.344	0.141
$\mu_{4,t}$	July 4	9.240	0.070	9.762	0.099
$\mu_{5,t}$		6.358	0.071	6.918	0.114
$\sigma_{1,t}^2$		0.133	0.028	0.291	0.061
$\sigma_{2,t}^2$		0.096	0.020	0.218	0.046
$\sigma_{3,t}^2$		0.641	0.138	0.857	0.185
$\sigma_{4,t}^2$		0.221	0.047	0.437	0.092
$\sigma_{5,t}^2$		0.192	0.044	0.480	0.112
$\beta_{0,t}$		17.673	7.291	17.728	7.314
$\beta_{1,t}$		0.010	0.003	0.010	0.003
σ_t^2		91.092	0.000	91.521	0.000
$\mu_{1,t}$		8.779	0.019	9.055	0.041
$\mu_{2,t}$		8.989	0.014	9.236	0.032
$\mu_{3,t}$		7.951	0.083	8.561	0.109
$\mu_{4,t}$	July 9	8.984	0.015	9.263	0.038
$\mu_{5,t}$		6.428	0.050	6.896	0.072
$\sigma_{1,t}^2$		0.016	0.003	0.076	0.016
$\sigma_{2,t}^2$		0.009	0.002	0.045	0.009
$\sigma_{3,t}^2$		0.309	0.065	0.520	0.111
$\sigma_{4,t}^2$		0.010	0.002	0.064	0.013
$\sigma_{5,t}^2$		0.098	0.022	0.202	0.046

Table 3.6. Point estimates (MLE) and their standard deviations for the parameters used to forecast the 2001 runs at June 24, June 29, July 4, and July 9, respectively. Subscript *t* corresponds to the respective forecast date. S.D. denotes standard deviation. The values under the 'With' column are associated with the run timing incorporation while those under the 'Without' column are not. Because the age-specific gillnet selectivity (G_a) is constant by day within the season, I show the values only once.

		With		Without	
Parameter	**Date**	**MLE**	**S.D.**	**MLE**	**S.D.**
$\beta_{0,t}$		21.033	5.523	22.411	5.758
$\beta_{1,t}$		0.024	0.007	0.025	0.008
σ_t^2		84.774	0.000	93.713	0.000
$\mu_{1,t}$		11.531	0.179	12.045	0.194
$\mu_{2,t}$		10.253	0.169	10.525	0.176
$\mu_{3,t}$		8.914	0.136	9.033	0.144
$\mu_{4,t}$	June	10.660	0.195	11.363	0.213
$\mu_{5,t}$	24	5.108	0.150	5.375	0.139
$\sigma_{1,t}^2$		1.445	0.305	1.662	0.354
$\sigma_{2,t}^2$		1.314	0.274	1.339	0.289
$\sigma_{3,t}^2$		0.795	0.172	0.829	0.185
$\sigma_{4,t}^2$		1.712	0.361	2.000	0.009
$\sigma_{5,t}^2$		0.693	0.176	0.602	0.153
G_1		0.557	0.498		
G_2		0.837	0.625		
G_3		0.630	0.492		
$\beta_{0,t}$		15.796	5.940	15.826	6.182
$\beta_{1,t}$		0.017	0.004	0.020	0.005
σ_t^2		72.216	0.000	75.490	0.000
$\mu_{1,t}$		10.096	0.146	10.811	0.179
$\mu_{2,t}$		9.210	0.127	9.746	0.149
$\mu_{3,t}$		7.924	0.138	8.370	0.145
$\mu_{4,t}$	June	9.969	0.144	10.707	0.194
$\mu_{5,t}$	29	6.111	0.159	6.570	0.185
$\sigma_{1,t}^2$		0.981	0.205	1.480	0.309
$\sigma_{2,t}^2$		0.740	0.154	1.018	0.212
$\sigma_{3,t}^2$		0.837	0.178	0.928	0.198
$\sigma_{4,t}^2$		0.947	0.198	1.723	0.359
$\sigma_{5,t}^2$		0.880	0.210	1.131	0.278

Table 3.6. (continued)

Parameter	Date	With MLE	With S.D.	Without MLE	Without S.D.
$\beta_{0,t}$		17.832	6.410	16.352	6.028
$\beta_{1,t}$		0.011	0.003	0.013	0.004
σ_t^2		84.958	0.000	75.141	0.000
$\mu_{1,t}$		9.443	0.059	9.799	0.080
$\mu_{2,t}$		8.641	0.056	8.888	0.070
$\mu_{3,t}$		6.948	0.126	7.404	0.143
$\mu_{4,t}$	July	9.359	0.087	9.711	0.098
$\mu_{5,t}$	4	6.468	0.081	6.843	0.111
$\sigma_{1,t}^2$		0.161	0.034	0.295	0.062
$\sigma_{2,t}^2$		0.146	0.030	0.225	0.047
$\sigma_{3,t}^2$		0.702	0.150	0.900	0.192
$\sigma_{4,t}^2$		0.352	0.073	0.438	0.091
$\sigma_{5,t}^2$		0.258	0.058	0.468	0.107
$\beta_{0,t}$		18.661	6.225	18.703	6.243
$\beta_{1,t}$		0.010	0.003	0.010	0.003
σ_t^2		85.519	0.000	85.904	0.000
$\mu_{1,t}$		9.147	0.023	9.375	0.040
$\mu_{2,t}$		8.365	0.017	8.567	0.031
$\mu_{3,t}$		7.015	0.074	7.518	0.108
$\mu_{4,t}$	July	9.034	0.017	9.263	0.037
$\mu_{5,t}$	9	6.590	0.054	6.959	0.070
$\sigma_{1,t}^2$		0.024	0.005	0.074	0.016
$\sigma_{2,t}^2$		0.013	0.003	0.045	0.009
$\sigma_{3,t}^2$		0.245	0.052	0.526	0.111
$\sigma_{4,t}^2$		0.014	0.003	0.063	0.013
$\sigma_{5,t}^2$		0.115	0.026	0.197	0.044

Table 3.7. The 1999 run forecasts (thousands) with the run timing information incorporated.

Date to which cumulative data are available	District	Age 1.2	Age 1.3	Age 2.2	Age 2.3	Sum
June 24 (15)	K-N	217	215	218	209	858
	Egegik	73	18	190	66	348
	Ugashik	16,107	3,773	8,254	997	29,132
	Nushagak	19	173	1	10	203
	Togiak	13	13	13	13	51
June 29 (20)	K-N	16,304	3,899	6,378	860	27,442
	Egegik	771	333	1,656	494	3,255
	Ugashik	659	205	295	92	1,251
	Nushagak	15	208	1	8	231
	Togiak	17	17	17	17	67
July 4 (25)	K-N	13,038	2,672	6,280	1,334	23,324
	Egegik	4,576	1,147	5,133	1,000	11,855
	Ugashik	224	68	97	31	421
	Nushagak	3,103	3,302	326	229	6,959
	Togiak	24	53	1	1	80
July 9 (30)	K-N	10,512	2,032	5,253	1,263	19,060
	Egegik	3,904	1,043	4,973	1,020	10,940
	Ugashik	1,687	302	478	126	2,593
	Nushagak	4,328	3,463	590	235	8,616
	Togiak	71	124	4	2	201

Table 3.8. The 1999 run forecasts (thousands) with the run timing information *not* incorporated.

Date to which cumulative data are available	District	Age 1.2	Age 1.3	Age 2.2	Age 2.3	Sum
June 24 (15)	K-N	239	236	240	229	944
	Egegik	54	13	139	49	255
	Ugashik	16,056	3,786	8,252	995	29,089
	Nushagak	14	126	1	7	148
	Togiak	10	10	10	10	40
June 29 (20)	K-N	16,304	3,899	6,378	860	27,442
	Egegik	771	333	1,656	494	3,255
	Ugashik	659	205	295	92	1,251
	Nushagak	15	208	1	8	231
	Togiak	17	17	17	17	67
July 4 (25)	K-N	11,876	2,440	5,713	1,214	21,242
	Egegik	4,180	1,048	4,684	912	10,824
	Ugashik	199	61	86	28	374
	Nushagak	2,804	2,989	294	207	6,294
	Togiak	22	48	1	1	72
July 9 (30)	K-N	9,788	1,885	4,887	1,176	17,736
	Egegik	3,667	976	4,666	956	10,265
	Ugashik	1,514	271	429	113	2,328
	Nushagak	4,085	3,258	557	222	8,122
	Togiak	67	116	3	2	188

Table 3.9. The 2000 run forecasts (thousands) with the run timing information incorporated.

Date to which cumulative data are available	District	Age 1.2	Age 1.3	Age 2.2	Age 2.3	Sum
June 24 (15)	K-N	2,564	11,741	433	276	15,014
	Egegik	1,470	3,368	2,386	3,344	10,569
	Ugashik	611	1,820	216	184	2,831
	Nushagak	53	231	2	2	288
	Togiak	8	290	4	1	302
June 29 (20)	K-N	1,060	5,672	250	331	7,312
	Egegik	825	3,546	2,140	2,750	9,261
	Ugashik	580	1,717	202	165	2,663
	Nushagak	2,425	4,460	25	33	6,944
	Togiak	8	297	4	1	310
July 4 (25)	K-N	1,151	5,439	891	432	7,914
	Egegik	910	3,442	1,892	2,801	9,045
	Ugashik	452	2,168	190	120	2,931
	Nushagak	3,173	5,179	41	41	8,434
	Togiak	12	458	6	1	477
July 9 (30)	K-N	910	4,378	565	542	6,395
	Egegik	792	2,928	1,807	2,361	7,886
	Ugashik	287	1,632	121	78	2,117
	Nushagak	3,189	4,654	58	38	7,939
	Togiak	14	540	7	1	563

Table 3.10. The 2000 run forecasts (thousands) with the run timing information *not* incorporated.

Date to which cumulative data are available	District	Age 1.2	Age 1.3	Age 2.2	Age 2.3	Sum
June 24 (15)	K-N	2,765	12,630	465	296	16,156
	Egegik	1,611	3,683	2,610	3,656	11,560
	Ugashik	659	1,965	233	198	3,056
	Nushagak	81	351	3	3	437
	Togiak	11	409	5	1	426
June 29 (20)	K-N	955	5,160	233	316	6,664
	Egegik	741	3,231	2,015	2,670	8,657
	Ugashik	1,113	3,278	395	313	5,100
	Nushagak	2,346	4,349	25	34	6,754
	Togiak	10	388	5	1	404
July 4 (25)	K-N	1,326	6,189	1,027	506	9,048
	Egegik	944	3,535	1,993	2,999	9,472
	Ugashik	568	2,705	238	152	3,662
	Nushagak	3,433	5,543	45	45	9,066
	Togiak	16	588	8	1	612
July 9 (30)	K-N	1,086	5,216	687	662	7,651
	Egegik	919	3,397	2,149	2,830	9,295
	Ugashik	392	2,222	165	106	2,885
	Nushagak	3,945	5,734	73	48	9,800
	Togiak	20	770	10	1	801

Table 3.11. The 2001 run forecasts (thousands) with the run timing information incorporated.

Date to which cumulative data are available	District	Age 1.2	Age 1.3	Age 2.2	Age 2.3	Sum
June 24 (15)	K-N	720	27,569	1,361	2,312	31,962
	Egegik	32	5,067	835	2,507	8,440
	Ugashik	55	2,920	246	285	3,506
	Nushagak	10	6,104	1	72	6,186
	Togiak	3	90	1	1	95
June 29 (20)	K-N	814	24,052	1,301	1,615	27,782
	Egegik	59	3,578	676	2,228	6,540
	Ugashik	65	1,051	68	88	1,271
	Nushagak	78	8,302	6	85	8,471
	Togiak	4	196	1	2	203
July 4 (25)	K-N	539	12,552	511	690	14,292
	Egegik	59	2,877	904	2,096	5,936
	Ugashik	28	453	29	38	548
	Nushagak	365	10,783	7	81	11,237
	Togiak	7	487	1	7	502
July 9 (30)	K-N	290	8,590	269	433	9,583
	Egegik	42	1,961	739	1,583	4,325
	Ugashik	139	687	33	69	928
	Nushagak	399	7,957	11	68	8,436
	Togiak	10	636	1	10	656

Table 3.12. The 2001 run forecasts (thousands) with the run timing information *not* incorporated.

Date to which cumulative data are available	District	Age 1.2	Age 1.3	Age 2.2	Age 2.3	Sum
June 24 (15)	K-N	737	27,440	1,279	2,250	31,706
	Egegik	40	5,958	1,022	2,911	9,931
	Ugashik	59	3,135	266	305	3,765
	Nushagak	13	7,732	1	90	7,836
	Togiak	4	123	1	1	129
June 29 (20)	K-N	955	29,151	1,473	1,826	33,405
	Egegik	71	4,346	817	2,660	7,894
	Ugashik	93	1,498	97	125	1,813
	Nushagak	63	7,276	4	74	7,417
	Togiak	5	241	1	3	250
July 4 (25)	K-N	822	17,348	760	952	19,882
	Egegik	69	3,482	1,049	2,534	7,134
	Ugashik	36	583	37	49	705
	Nushagak	332	11,836	6	89	12,263
	Togiak	9	572	1	9	590
July 9 (30)	K-N	370	10,964	358	562	12,254
	Egegik	52	2,421	910	1,969	5,352
	Ugashik	184	910	44	91	1,229
	Nushagak	499	10,003	14	87	10,603
	Togiak	13	850	1	13	878

Table 3.13. Comparison of the effect of incorporating the run timing forecast on the 1999 run forecasts and that of ignoring the run timing forecast. 'With' and 'Without' denote 'with incorporation of the run timing forecast' and 'without it.' Units of the forecast and error values are 'numbers in thousands' and '%.' The minus (-) sign indicates an under-forecast.

Date to which cumulative data are available	District	With		Without	
		Forecast	Error	Forecast	Error
June 24 (15)	K-N	858	-94.9	944	-94.4
	Egegik	348	-96.2	255	-97.2
	Ugashik	29,132	643.3	29,089	642.3
	Nushagak	203	-97.6	148	-98.3
	Togiak	51	-89.8	40	-92.0
June 29 (20)	K-N	27,442	63.7	27,442	63.7
	Egegik	3,255	-64.4	3,255	-64.4
	Ugashik	1,251	-68.1	1,251	-68.1
	Nushagak	231	-97.3	231	-97.3
	Togiak	67	-86.7	67	-86.7
July 4 (25)	K-N	23,324	39.1	21,242	26.7
	Egegik	11,855	29.6	10,824	18.3
	Ugashik	421	-89.3	374	-90.5
	Nushagak	6,959	-18.0	6,294	-25.8
	Togiak	80	-84.2	72	-85.6
July 9 (30)	K-N	19,060	13.7	17,736	5.8
	Egegik	10,940	19.6	10,265	12.2
	Ugashik	2,593	-33.8	2,328	-40.6
	Nushagak	8,616	1.6	8,122	-4.3
	Togiak	201	-60.1	188	-62.6

Table 3.14. Comparison of the effect of incorporating the run timing forecast on the 2000 run forecasts and that of ignoring the run timing forecast. 'With' and 'Without' denote 'with incorporation of the run timing forecast' and 'without it.' Units of the forecast and error values are 'numbers in thousands' and '%.' The minus (-) sign indicates an under-forecast.

Date to which cumulative data are available	District	With		Without	
		Forecast	Error	Forecast	Error
June 24 (15)	K-N	15,014	90.6	16,156	105.1
	Egegik	10,569	29.9	11,560	42.1
	Ugashik	2,831	33.0	3,056	43.6
	Nushagak	288	-96.6	437	-94.9
	Togiak	302	-72.7	426	-61.5
June 29 (20)	K-N	7,312	-7.2	6,664	-15.4
	Egegik	9,261	13.8	8,657	6.4
	Ugashik	2,663	25.2	5,100	139.6
	Nushagak	6,944	-18.7	6,754	-21.0
	Togiak	310	-72.0	404	-63.4
July 4 (25)	K-N	7,914	0.5	9,048	14.9
	Egegik	9,045	11.2	9,472	16.4
	Ugashik	2,931	37.7	3,662	72.1
	Nushagak	8,434	-1.3	9,066	6.1
	Togiak	477	-56.9	612	-44.6
July 9 (30)	K-N	6,395	-18.8	7,651	-2.9
	Egegik	7,886	-3.1	9,295	14.2
	Ugashik	2,117	-0.5	2,885	35.6
	Nushagak	7,939	-7.1	9,800	14.7
	Togiak	563	-49.1	801	-27.5

Table 3.15. Comparison of the effect of incorporating the run timing forecast on the 2001 run forecasts and that of ignoring the run timing forecast. 'With' and 'Without' denote 'with incorporation of the run timing forecast' and 'without it.' Units of the forecast and error values are 'numbers in thousands' and '%.' The minus (-) sign indicates an under-forecast.

Date to which cumulative data are available	District	With Forecast	With Error	Without Forecast	Without Error
June 24 (15)	K-N	31,962	291.2	31,706	288.1
	Egegik	8,440	120.3	9,931	159.2
	Ugashik	3,506	167.8	3,765	187.6
	Nushagak	6,186	-17.2	7,836	4.9
	Togiak	95	-91.4	129	-88.4
June 29 (20)	K-N	27,782	240.0	33,405	308.9
	Egegik	6,540	70.7	7,894	106.1
	Ugashik	1,271	-2.9	1,813	38.5
	Nushagak	8,471	13.4	7,417	-0.8
	Togiak	203	-81.7	250	-77.5
July 4 (25)	K-N	14,292	74.9	19,882	143.4
	Egegik	5,936	54.9	7,134	86.2
	Ugashik	548	-58.2	705	-46.2
	Nushagak	11,237	50.4	12,263	64.1
	Togiak	502	-54.7	590	-46.7
July 9 (30)	K-N	9,583	17.3	12,254	50.0
	Egegik	4,325	12.9	5,352	39.7
	Ugashik	928	-29.1	1,229	-6.2
	Nushagak	8,436	12.9	10,603	41.9
	Togiak	656	-40.8	878	-20.8

Table 3.16. Evaluation of the incorporation of the Port Moller gear selectivity for age-specific fish in forecasting the 1999 returns. The forecast (thousands) and error (%) values under the 'Without' column represent those calculated with the selectivity *ignored*. The values under the 'With' column are the same as those under the 'With' column in Table 3.13. In both cases, I incorporated the run time forecast information accordingly. The minus (-) sign indicates an under-forecast.

Date to which cumulative data are available	District	With		Without	
		Forecast	Error	Forecast	Error
June 24 (15)	K-N	858	-94.9	1,177	-93.0
	Egegik	348	-96.2	362	-96.0
	Ugashik	29,132	643.3	27,770	608.6
	Nushagak	203	-97.6	219	-97.4
	Togiak	51	-89.8	51	-89.8
June 29 (20)	K-N	27,442	63.7	647	-96.1
	Egegik	3,255	-64.4	2,299	-74.9
	Ugashik	1,251	-68.1	27,092	591.3
	Nushagak	231	-97.3	235	-97.2
	Togiak	67	-86.7	67	-86.7
July 4 (25)	K-N	23,324	39.1	20,216	20.6
	Egegik	11,855	29.6	12,228	33.6
	Ugashik	421	-89.3	422	-89.2
	Nushagak	6,959	-18.0	9,075	7.0
	Togiak	80	-84.2	80	-84.1
July 9 (30)	K-N	19,060	13.7	18,241	8.8
	Egegik	10,940	19.6	10,982	20.0
	Ugashik	2,593	-33.8	2,418	-38.3
	Nushagak	8,616	1.6	9,162	8.0
	Togiak	201	-60.1	203	-59.8

Table 3.17. Evaluation of the incorporation of the Port Moller gear selectivity for age-specific fish in forecasting the 2000 returns. The forecast (thousands) and error (%) values under the 'Without' column represent those calculated with the selectivity *ignored*. The values under the 'With' column are the same as those under the 'With' column in Table 3.14. In both cases, I incorporated the run time forecast information accordingly. The minus (-) sign indicates an under-forecast.

Date to which cumulative data are available	District	With		Without	
		Forecast	Error	Forecast	Error
June 24 (15)	K-N	15,014	90.6	15,169	92.6
	Egegik	10,569	29.9	10,930	34.3
	Ugashik	2,831	33.0	2,569	20.7
	Nushagak	288	-96.6	285	-96.7
	Togiak	302	-72.7	305	-72.4
June 29 (20)	K-N	7,312	-7.2	8,303	5.4
	Egegik	9,261	13.8	9,755	19.9
	Ugashik	2,663	25.2	2,507	17.8
	Nushagak	6,944	-18.7	5,488	-35.8
	Togiak	310	-72.0	317	-71.3
July 4 (25)	K-N	7,914	0.5	8,045	2.1
	Egegik	9,045	11.2	9,425	15.8
	Ugashik	2,931	37.7	3,029	42.3
	Nushagak	8,434	-1.3	7,500	-12.2
	Togiak	477	-56.9	482	-56.4
July 9 (30)	K-N	6,395	-18.8	6,417	-18.6
	Egegik	7,886	-3.1	7,929	-2.6
	Ugashik	2,117	-0.5	2,152	1.1
	Nushagak	7,939	-7.1	7,866	-8.0
	Togiak	563	-49.1	566	-48.8

Table 3.18. Evaluation of the incorporation of the Port Moller gear selectivity for age-specific fish in forecasting the 2001 returns. The forecast (thousands) and error (%) values under the 'Without' column represent those calculated with the selectivity *ignored*. The values under the 'With' column are the same as those under the 'With' column in Table 3.15. In both cases, I incorporated the run time forecast information accordingly. The minus (-) sign indicates an under-forecast.

Date to which cumulative data are available	District	With		Without	
		Forecast	Error	Forecast	Error
June 24 (15)	K-N	31,962	291.2	28,561	249.6
	Egegik	8,440	120.3	10,940	185.6
	Ugashik	3,506	167.8	3,561	172.0
	Nushagak	6,186	-17.2	6,712	-10.2
	Togiak	95	-91.4	95	-91.4
June 29 (20)	K-N	27,782	240.0	23,932	192.9
	Egegik	6,540	70.7	8,466	121.0
	Ugashik	1,271	-2.9	1,290	-1.5
	Nushagak	8,471	13.4	9,732	30.2
	Togiak	203	-81.7	203	-81.7
July 4 (25)	K-N	14,292	74.9	13,730	68.1
	Egegik	5,936	54.9	6,237	62.8
	Ugashik	548	-58.2	548	-58.1
	Nushagak	11,237	50.4	11,177	49.6
	Togiak	502	-54.7	503	-54.6
July 9 (30)	K-N	9,583	17.3	9,562	17.0
	Egegik	4,325	12.9	4,343	13.4
	Ugashik	928	-29.1	914	-30.2
	Nushagak	8,436	12.9	8,416	12.6
	Togiak	656	-40.8	656	-40.8

Table 3.19. Age-specific proportions in data observed up to day code 25 (July 4) of year 1999, and those in the 1999 run forecasts made at the day. The first category is the observed data available up to the day, the second category is the run forecasts made with the Port Moller selectivity for age-specific fish considered, and the third one is the run forecasts with the selectivity ignored. These values are used to draw Figure 3.13.

Category	Area	Age 1.2	Age 1.3	Age 2.2	Age 2.3
Observed data	Port Moller	0.45	0.21	0.26	0.08
	K-N	0.48	0.14	0.29	0.09
	Egegik	0.35	0.11	0.44	0.11
	Ugashik	0.53	0.16	0.23	0.07
	Nushagak	0.42	0.50	0.05	0.04
	Togiak	0.31	0.68	0.01	0.01
The run forecasts made with the selectivity being considered	Total	0.42	0.23	0.26	0.09
	K-N	0.56	0.12	0.27	0.06
	Egegik	0.39	0.10	0.43	0.08
	Ugashik	0.53	0.16	0.23	0.07
	Nushagak	0.45	0.47	0.05	0.03
	Togiak	0.31	0.67	0.01	0.01
The run forecasts made with the selectivity being ignored	Total	0.44	0.21	0.27	0.08
	K-N	0.50	0.14	0.27	0.09
	Egegik	0.36	0.11	0.43	0.11
	Ugashik	0.53	0.16	0.23	0.07
	Nushagak	0.42	0.50	0.05	0.04
	Togiak	0.31	0.67	0.01	0.01

Table 3.20. Age-specific proportions in data observed up to day code 25 (July 4) of year 2000, and those in the 2000 run forecasts made at the day. The first category is the observed data available up to the day, the second category is the run forecasts made with the Port Moller selectivity for age-specific fish considered, and the third one is the run forecasts with the selectivity ignored. These values are used to draw Figure 3.14.

Category	Area	Age 1.2	Age 1.3	Age 2.2	Age 2.3
Observed	Port Moller	0.15	0.63	0.08	0.14
	K-N	0.14	0.68	0.12	0.06
	Egegik	0.09	0.36	0.22	0.33
	Ugashik	0.15	0.74	0.07	0.04
	Nushagak	0.37	0.62	0.01	0.01
	Togiak	0.03	0.96	0.01	0.00
Predicted with the age selectivity	Total	0.14	0.63	0.08	0.15
	K-N	0.15	0.69	0.11	0.06
	Egegik	0.10	0.38	0.21	0.31
	Ugashik	0.15	0.74	0.07	0.04
	Nushagak	0.38	0.61	0.01	0.01
	Togiak	0.03	0.96	0.01	0.00
Predicted without the age selectivity	Total	0.16	0.61	0.09	0.14
	K-N	0.12	0.72	0.10	0.06
	Egegik	0.08	0.40	0.18	0.35
	Ugashik	0.14	0.75	0.06	0.04
	Nushagak	0.33	0.66	0.00	0.01
	Togiak	0.03	0.96	0.01	0.00

Table 3.21. Age-specific proportions in data observed up to day code 25 (July 4) of year 2001, and those in the 2001 run forecasts made at the day. The first category is the observed data available up to the day, the second category is the run forecasts made with the Port Moller selectivity for age-specific fish considered, and the third one is the run forecasts with the selectivity ignored. These values are used to draw Figure 3.15.

Category	Area	Age 1.2	Age 1.3	Age 2.2	Age 2.3
Observed	Port Moller	0.03	0.82	0.05	0.11
	K-N	0.01	0.93	0.01	0.05
	Egegik	0.01	0.51	0.11	0.37
	Ugashik	0.05	0.83	0.05	0.07
	Nushagak	0.02	0.98	0.00	0.01
	Togiak	0.02	0.97	0.00	0.02
Predicted with the age selectivity	Total	0.02	0.84	0.03	0.11
	K-N	0.04	0.88	0.04	0.05
	Egegik	0.01	0.49	0.15	0.35
	Ugashik	0.05	0.83	0.05	0.07
	Nushagak	0.03	0.96	0.00	0.01
	Togiak	0.02	0.97	0.00	0.02
Predicted without the age selectivity	Total	0.02	0.84	0.04	0.10
	K-N	0.03	0.89	0.02	0.06
	Egegik	0.01	0.48	0.14	0.37
	Ugashik	0.05	0.83	0.05	0.07
	Nushagak	0.03	0.96	0.00	0.01
	Togiak	0.02	0.97	0.00	0.02

(A) Variables and parameters over time

Time index	*1*	*2*	......	*i-1*	*i*
Explanatory variables (data)	X_1	X_2	......	X_{i-1}	X_i
Parameters	θ_1	θ_2	......	θ_{i-1}	
Response variables (data)	Y_1	Y_2	......	Y_{i-1}	$\tilde{Y}_i$

Unobserved data: Predictive variables
(note tilde mark)

(B) Estimation of updated parameters, and prediction of unobserved data

$$\hat{\theta}_{i-1} = f(\text{observed data}) = f(X_1,\ X_2,\ ...,\ X_{i-1},\ Y_1,\ Y_2,\ ...,\ Y_{i-1})$$

$$\begin{aligned}\tilde{Y}_i &= f(\text{explanatory variables at } i, \text{ estimates of updated parameters})\\ &= f(X_i,\ \hat{\theta}_{i-1})\end{aligned}$$

Figure 3.1. (A) Variables and parameters over time, and (B) Estimation of updated parameters, and prediction of unobserved data. Estimation of parameters is based only on observed data, and must not be affected by estimates of predictive variables. After parameters are estimated, predictive variables at time *i* can be calculated with estimates of updated parameters and explanatory variables at time *i*.

Age

	$r_{1,1}$	$r_{1,2}$	$r_{1,3}$	$r_{1,4}$	$r_{1,\bullet}$
	$r_{2,1}$	$r_{2,2}$	$r_{2,3}$	$r_{2,4}$	$r_{2,\bullet}$
Stock	$r_{3,1}$	$r_{3,2}$	$r_{3,3}$	$r_{3,4}$	$r_{3,\bullet}$
	$r_{4,1}$	$r_{4,2}$	$r_{4,3}$	$r_{4,4}$	$r_{4,\bullet}$
	$r_{5,1}$	$r_{5,2}$	$r_{5,3}$	$r_{5,4}$	$r_{5,\bullet}$
	$r_{\bullet,1}$	$r_{\bullet,2}$	$r_{\bullet,3}$	$r_{\bullet,4}$	R

Figure 3.2. Contingency table of the predictive run sizes. '$r_{s,a}$'of each cell denotes the *final* run size of stock s and age a. The final run size means the cumulative run size up to the end of the season. '$r_{s,\bullet}$' and '$r_{\bullet,a}$' represent the respective marginal sums. 'R' is the total sum.

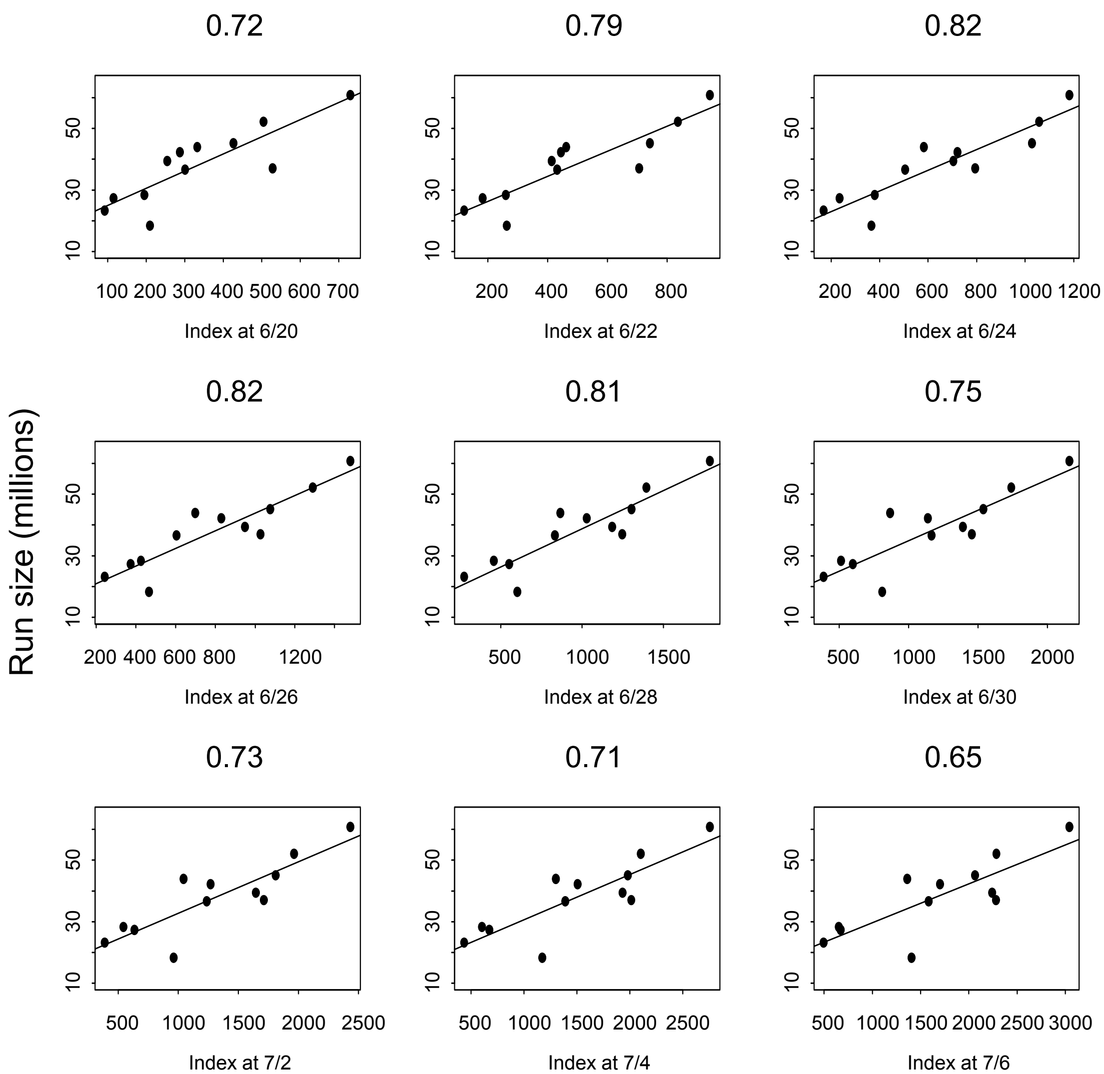


Figure 3.3. Rogers' regression model by day. The determination coefficient (R^2) ranges from 0.65 to 0.82 when outlier data (years 1990, 1994, 1997, and 2001) are excluded.

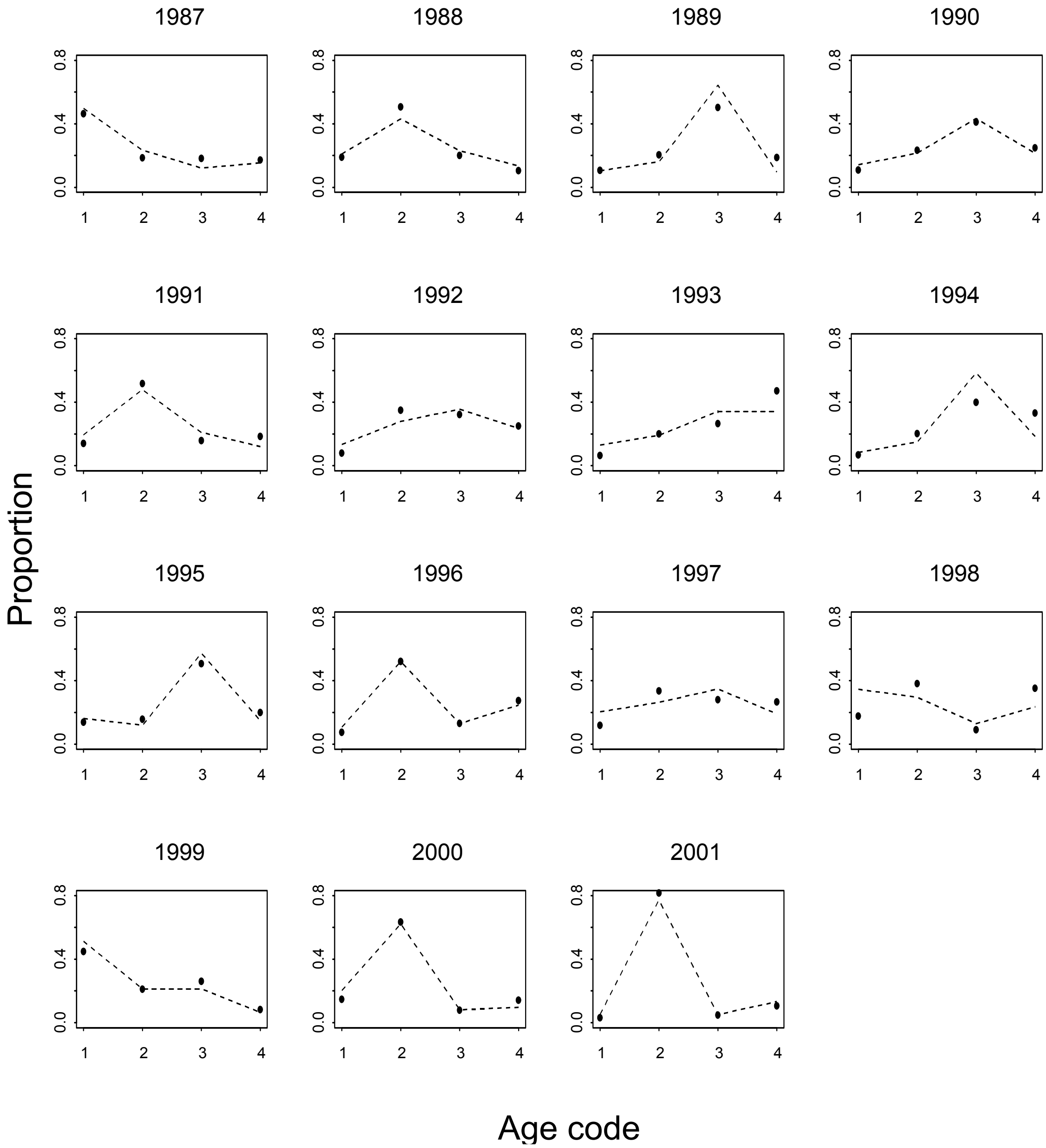


Figure 3.4. Comparison of age composition (in percent) in the Port Moller fishery catch (dots) and that in run size to Bristol Bay (dashed line). Age codes 1 through 4 denote ages 1.2, 1.3, 2.2, and 2.3, respectively.

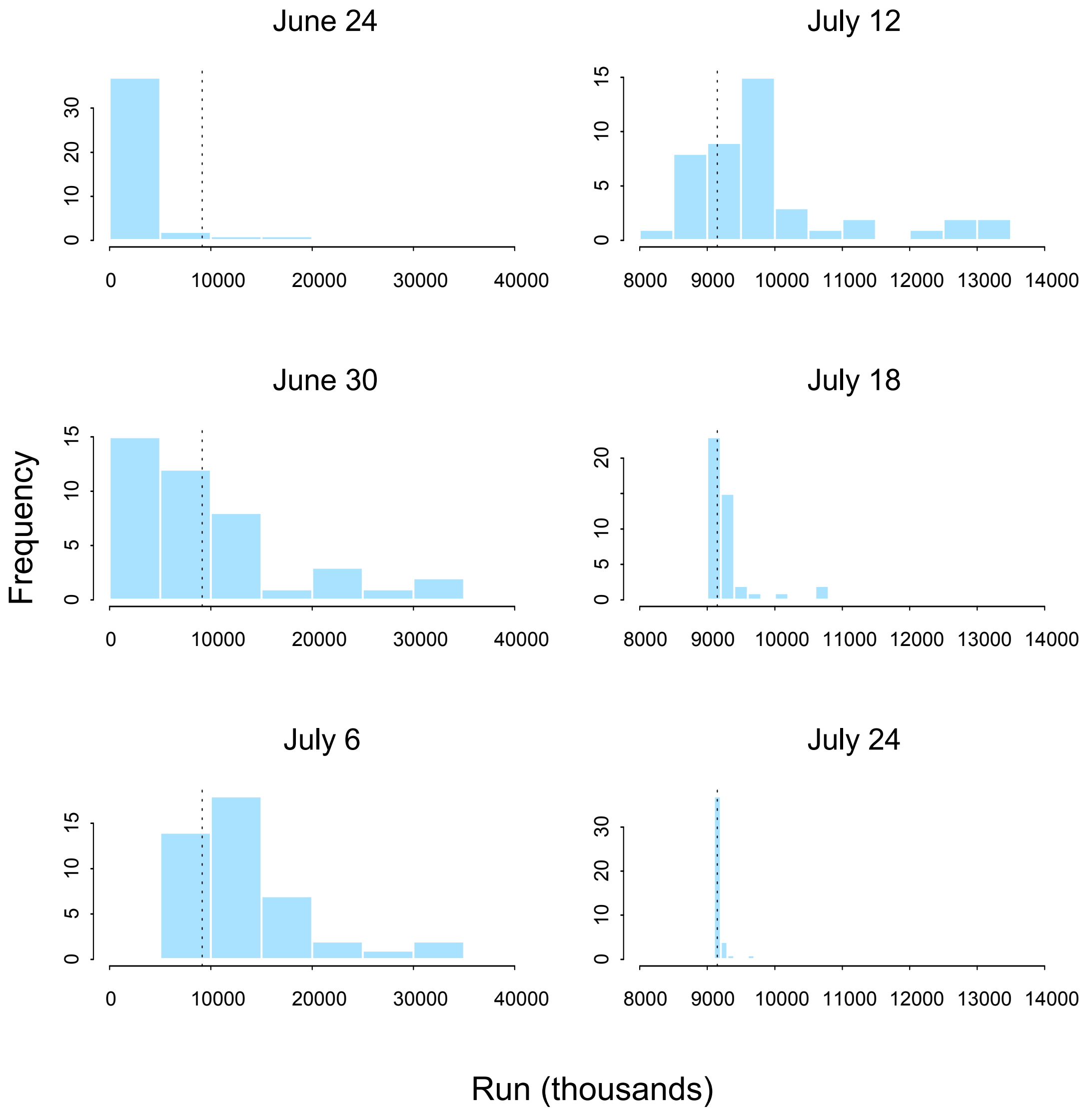


Figure 3.5. The distribution of the 1999 Egegik run size estimated at the specified day. The distribution predicted at June 30 includes values above the *x*-axis limit, but I don't show them for the same scale of plots in the left column. The dotted vertical line is the actual run size.

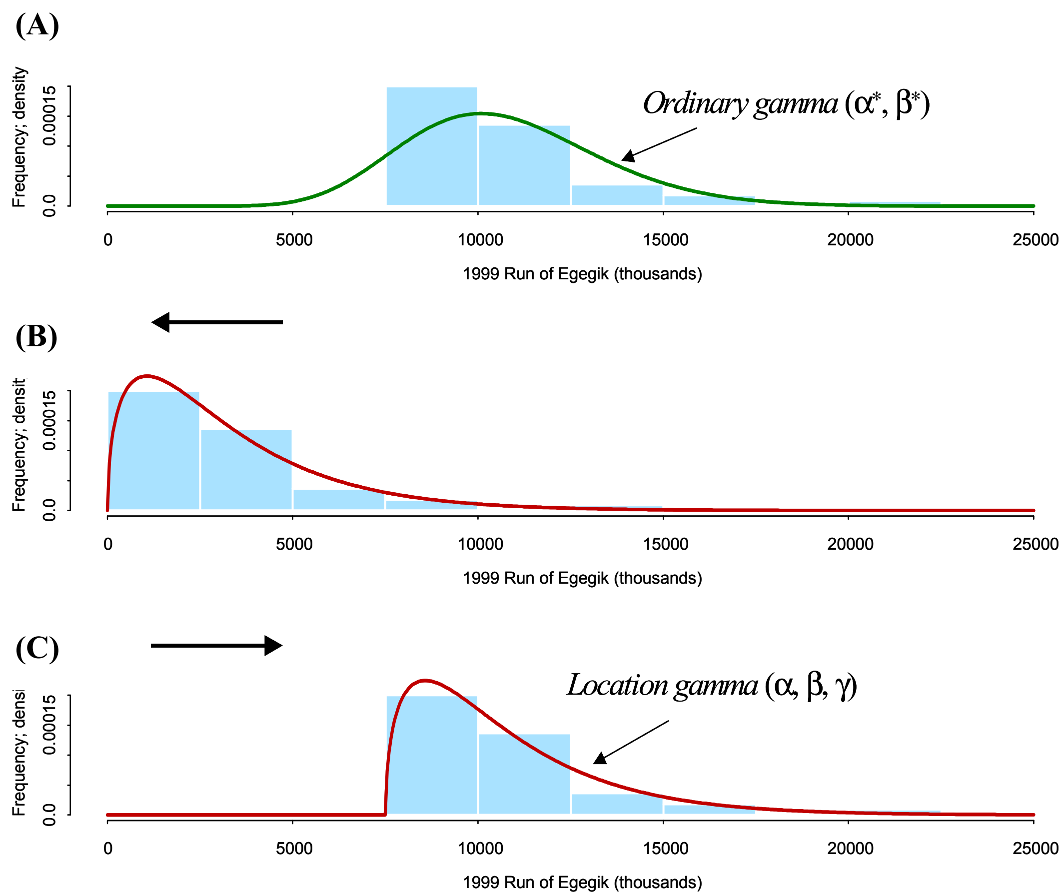


Figure 3.6. An example of developing a location gamma density from an ordinary gamma density. The histogram is the 1999 Egegik run distribution predicted at day code 30 (July 9) by Equation 3.12, and the *y*-axis scale is adjusted as the probability density scale. In A, an ordinary gamma density is fitted to the histogram. The ordinary gamma shape looks symmetric while the histogram shape is not. I added the asterisk (*) mark to the two parameters, α^* and β^* to indicate that they are different from α and β shown in C. After shifting the histogram, not changing the shape, I fit another ordinary gamma density to the shifted histogram (B). And then I shift the new ordinary gamma and histogram back to the original location of the histogram (C). The location gamma has an additional parameter, γ that indicates the minimum value of the histogram.

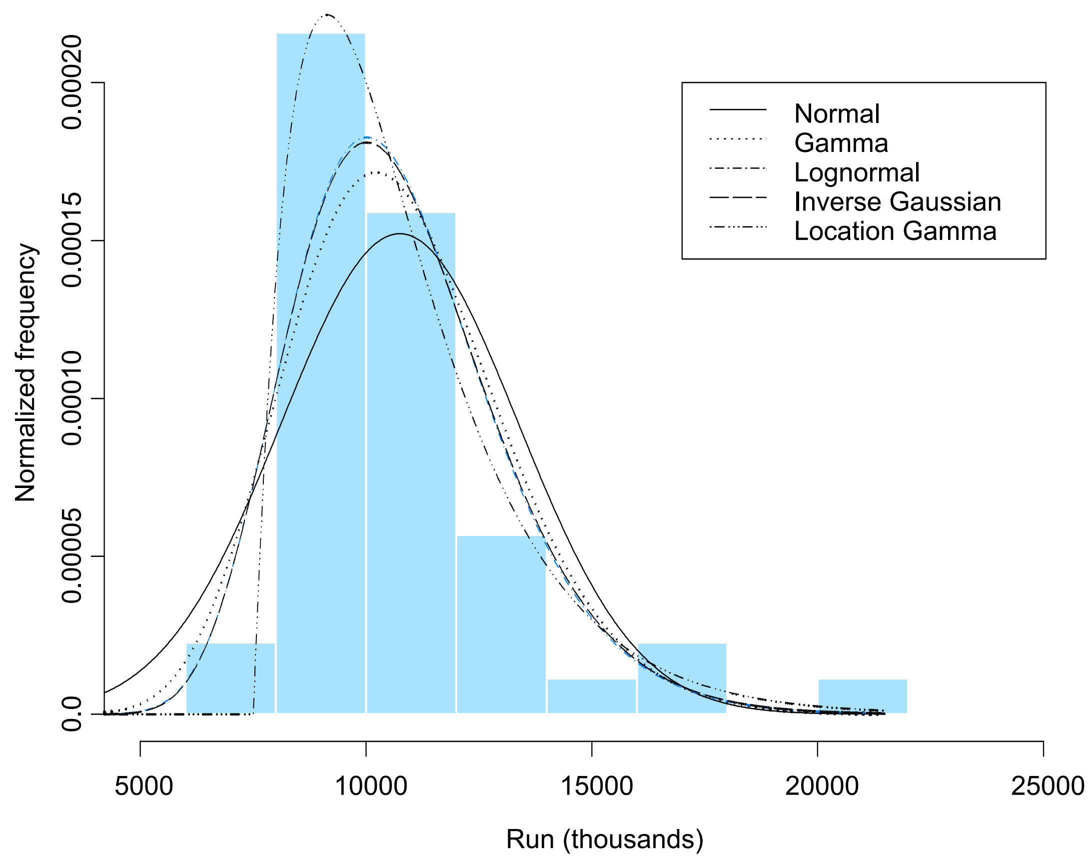


Figure 3.7. An example of five densities fitted to the 1999 Egegik run distribution predicted at day code 30 (July 9). Parameters in the respective six densities are estimated by maximum likelihood method.

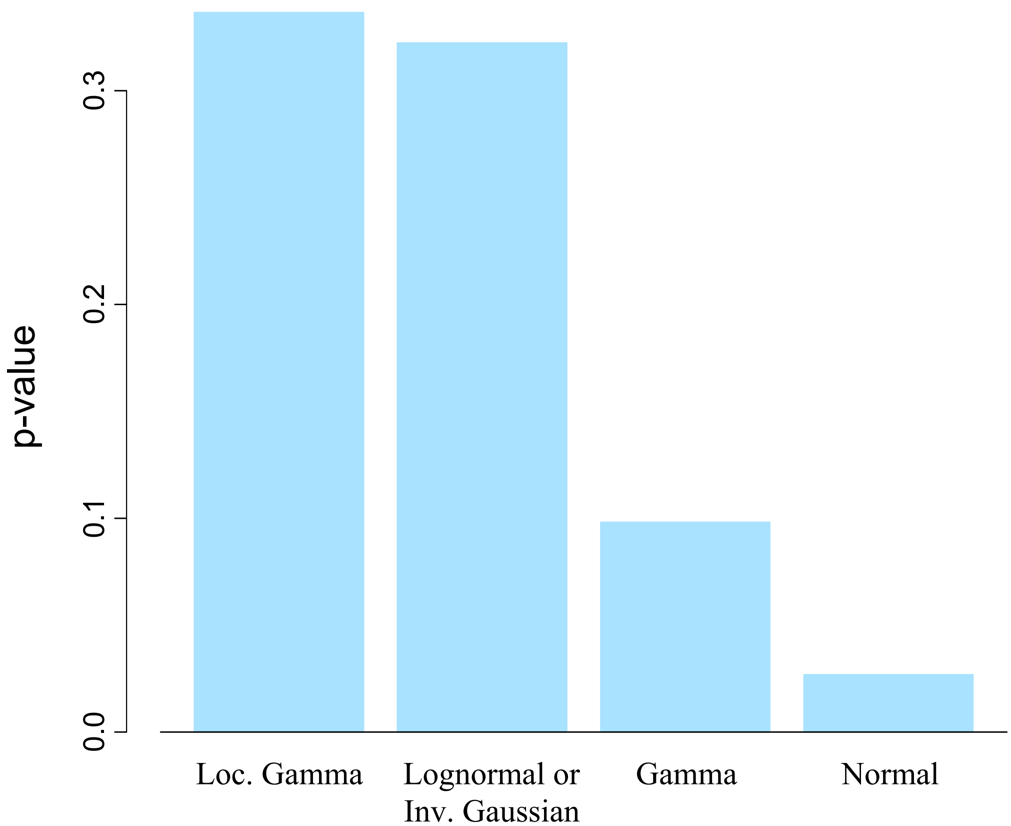


Figure 3.8. The results of K-S goodness fit test of five parametric densities for the 1999 Egegik run size. The p-value is the average of those of five tests in Table 3.1, except for the test for the distribution predicted at day code 40 (July 19).

$D_{t,\vec{y}}$: historical data that belong to day t in past years $\vec{y}$

$\vec{\theta}_t$: parameters in the objective functions that are to be estimated at day t

$\vec{r}$: predictive variables (run sizes) in the objective functions

(A) Normally,

$$D_{t,\vec{y}} \rightarrow \vec{\theta}_t \rightarrow \vec{r}$$

(B) When detecting run timing earlier or later by q days,

$$D_{t\pm q,\vec{y}} \rightarrow \vec{\theta}_t \rightarrow \vec{r}$$

Figure 3.9. An illustration of how I incorporate a run timing forecast into the estimation of run sizes. (A) In forecasting run sizes at day t during the season, I use historical data, which correspond to day t in past years, to estimate parameters in the objective functions. (B) When detecting run timing earlier or later by q days, I use historical data that correspond to day '$t \pm q$' in past years.

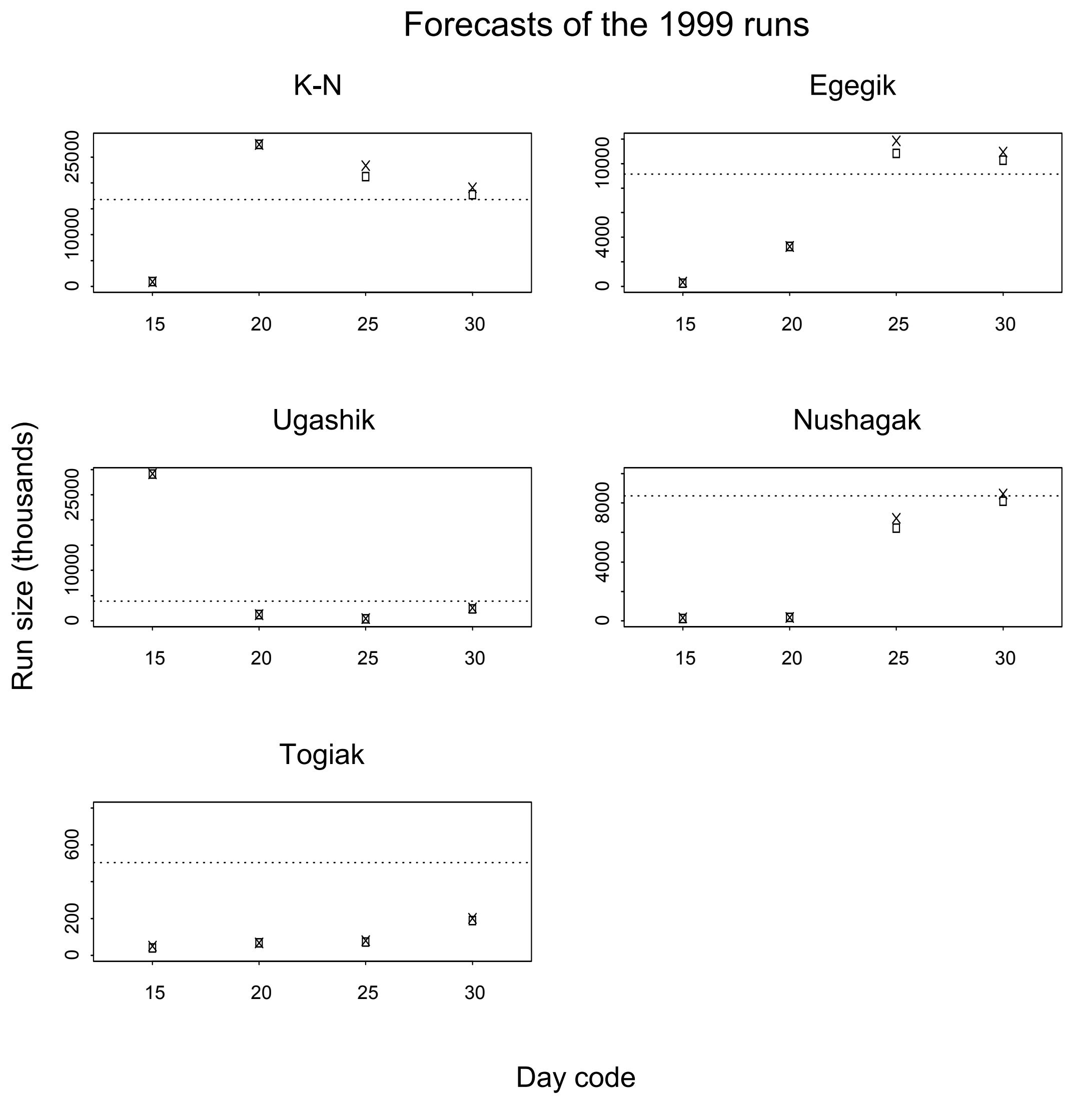


Figure 3.10. Summary of Table 3.13, where the 1999 run forecasts are compared with the actual returns. The horizontal dotted line represents the actual run size. The cross mark (×) points are run forecasts made with the run timing information incorporated, while the square marks are those made with the run timing information *not* incorporated.

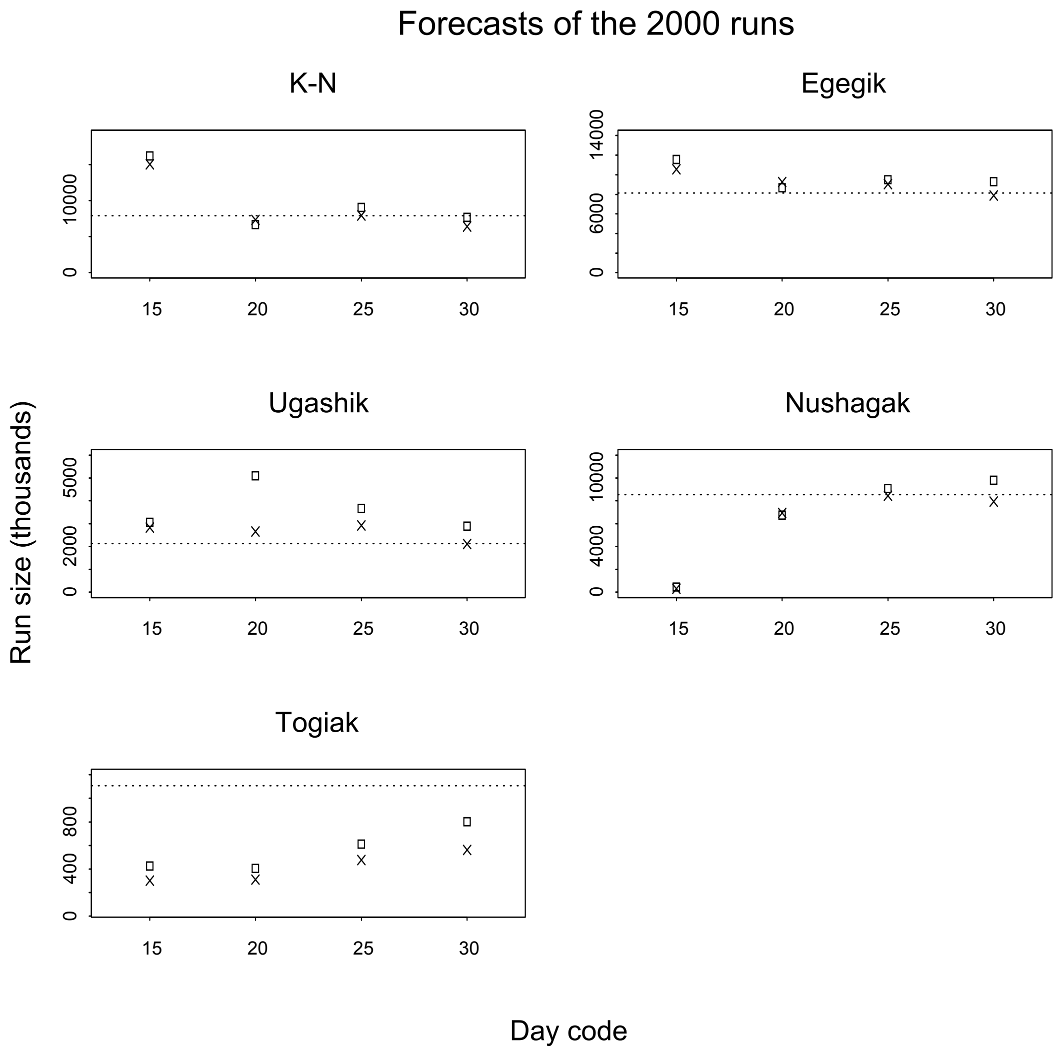


Figure 3.11. Summary of Table 3.14, where the 2000 run forecasts are compared with the actual returns. The horizontal dotted line represents the actual run size. The cross mark (×) points are run forecasts made with the run timing information incorporated, while the square marks are those made with the run timing information *not* incorporated.

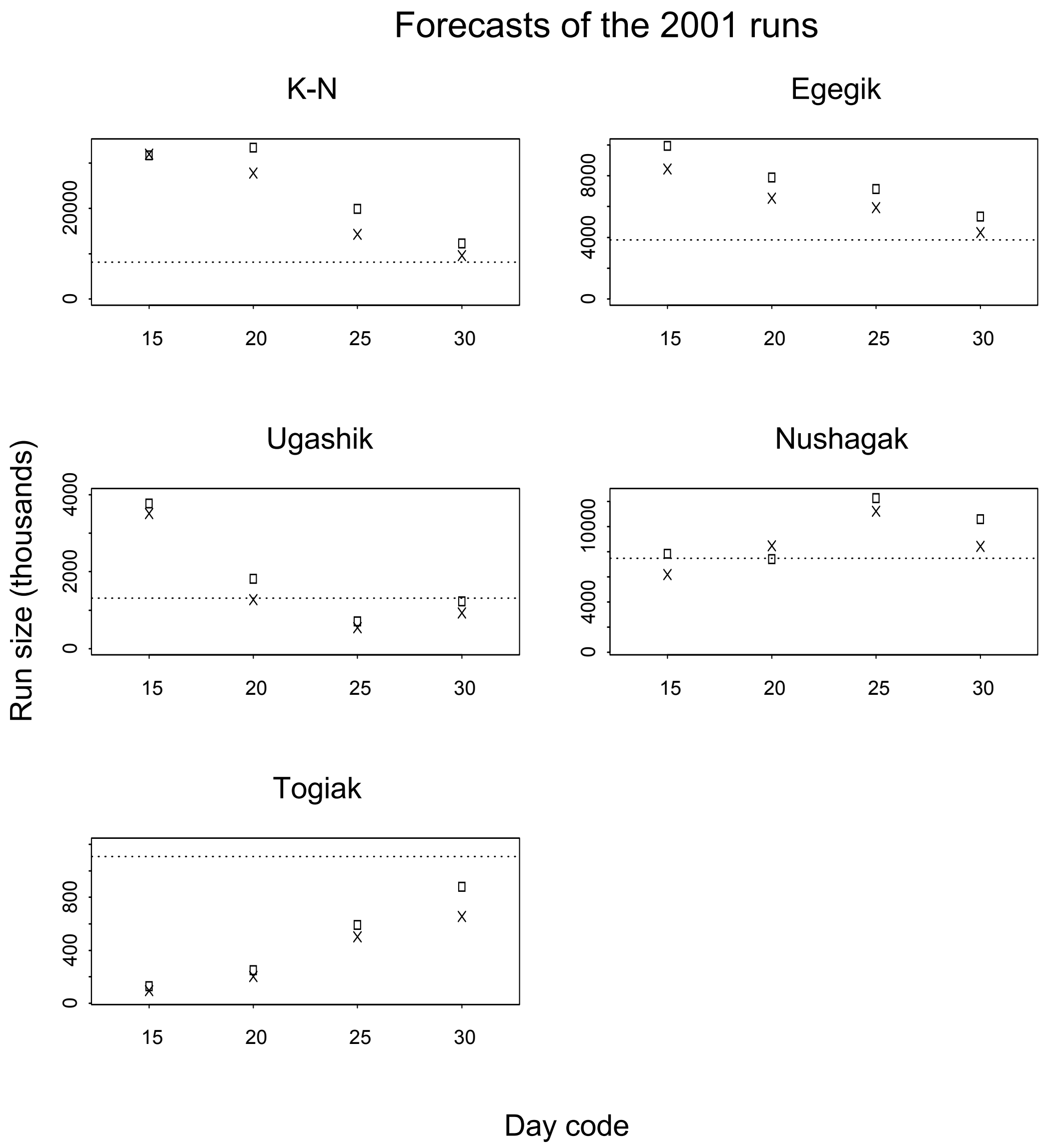


Figure 3.12. Summary of Table 3.15, where the 2001 run forecasts are compared with the actual returns. The horizontal dotted line represents the actual run size. The cross mark (×) points are run forecasts made with the run timing information incorporated, while the square marks are those made with the run timing information *not* incorporated.

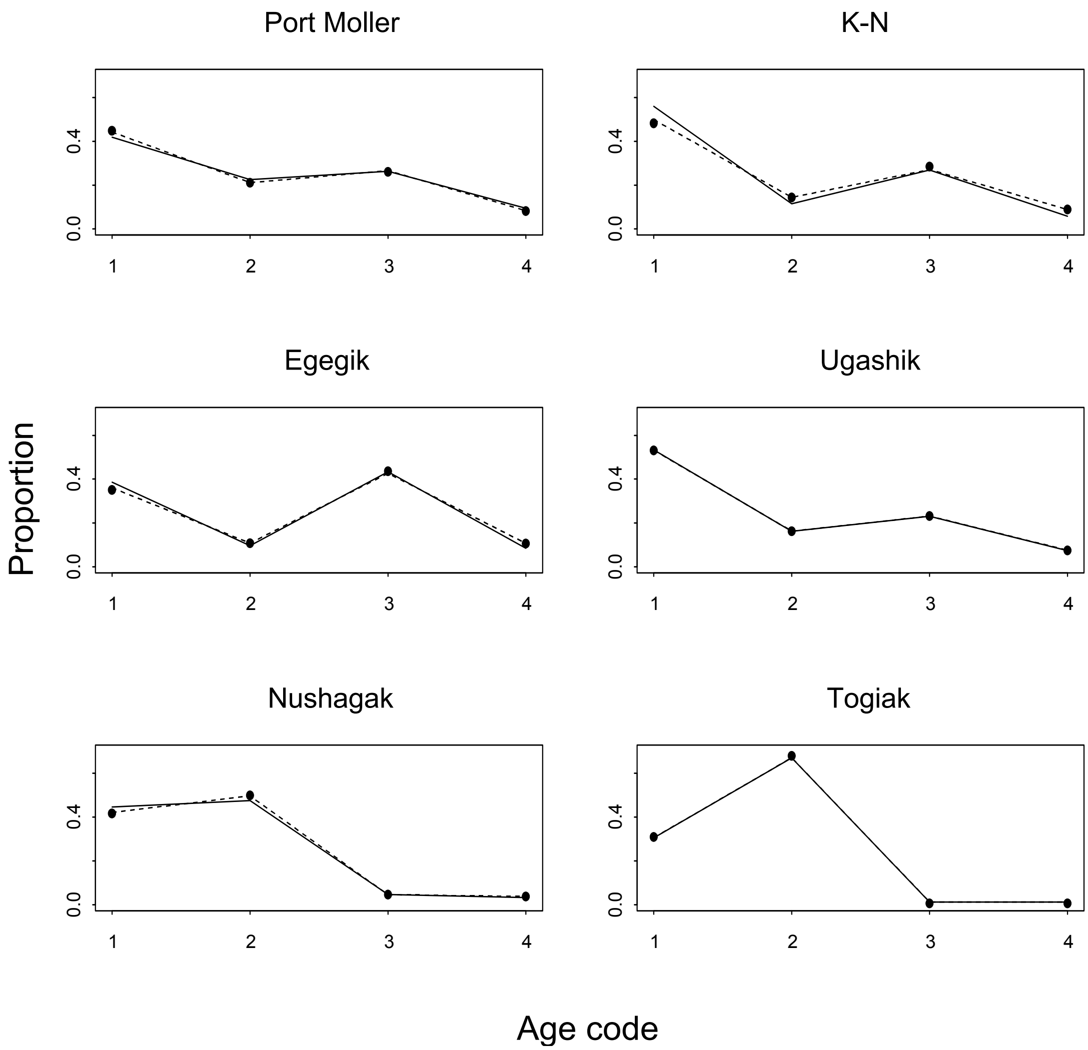


Figure 3.13. Comparison of the effect of considering the Port Moller gear selectivity for age-specific fish on the 1999 run forecasts made at day code 25 (July 4) and that of ignoring the selectivity. Dots indicate age composition (in proportion) observed up to the day, solid lines represent that in run forecasts made with the selectivity considered, and, and dashed lines are that in run forecasts made with the selectivity ignored. In case of the 'Port Moller' box, dots are age composition observed in the Port Moller fishery catch up to the day.

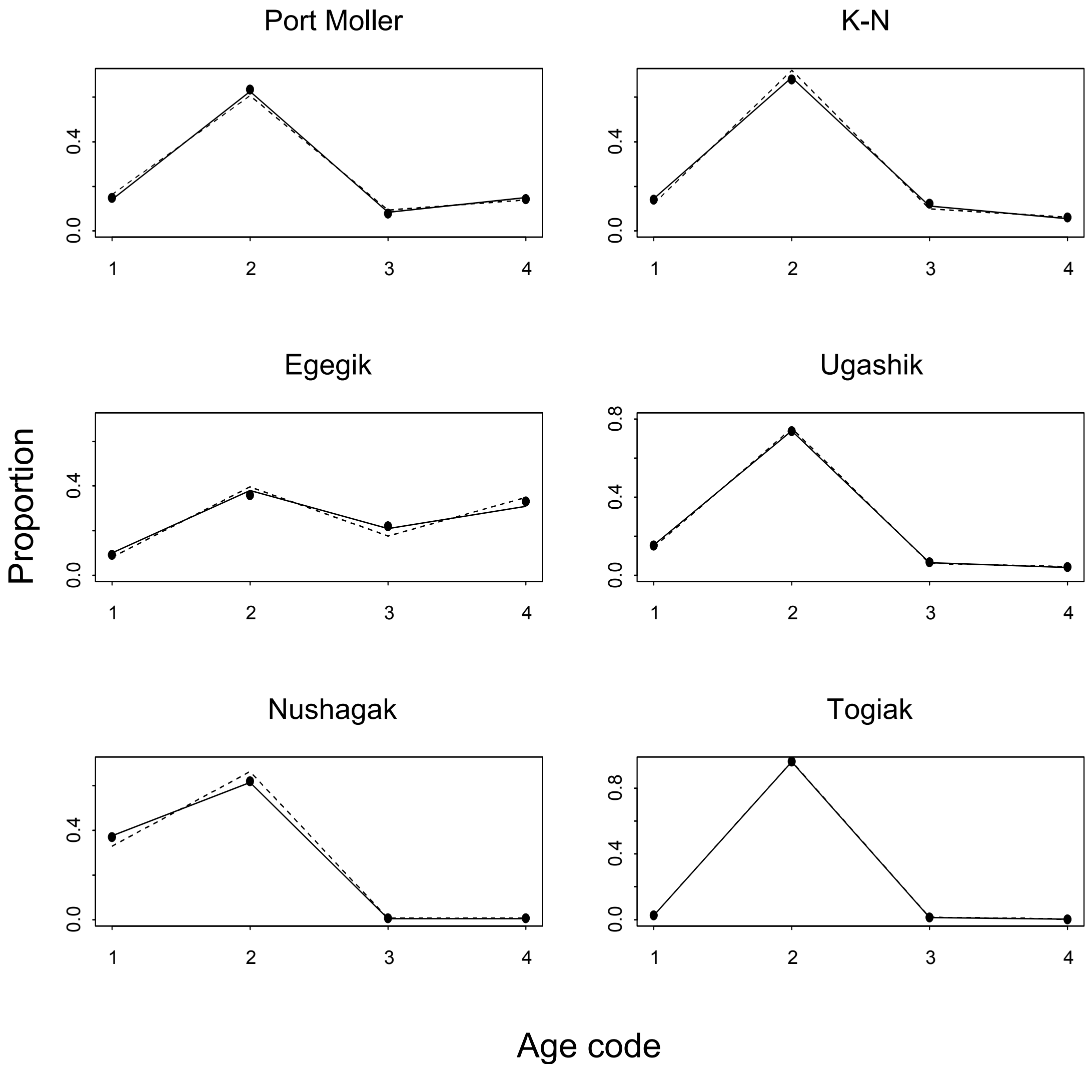


Figure 3.14. Comparison of the effect of considering the Port Moller gear selectivity for age-specific fish on the 2000 run forecasts made at day code 25 (July 4) and that of ignoring the selectivity. Dots indicate age composition (in proportion) observed up to the day, solid lines represent that in run forecasts made with the selectivity considered, and, and dashed lines are that in run forecasts made with the selectivity ignored. In case of the ‘Port Moller’ box, dots are age composition observed in the Port Moller fishery catch up to the day.

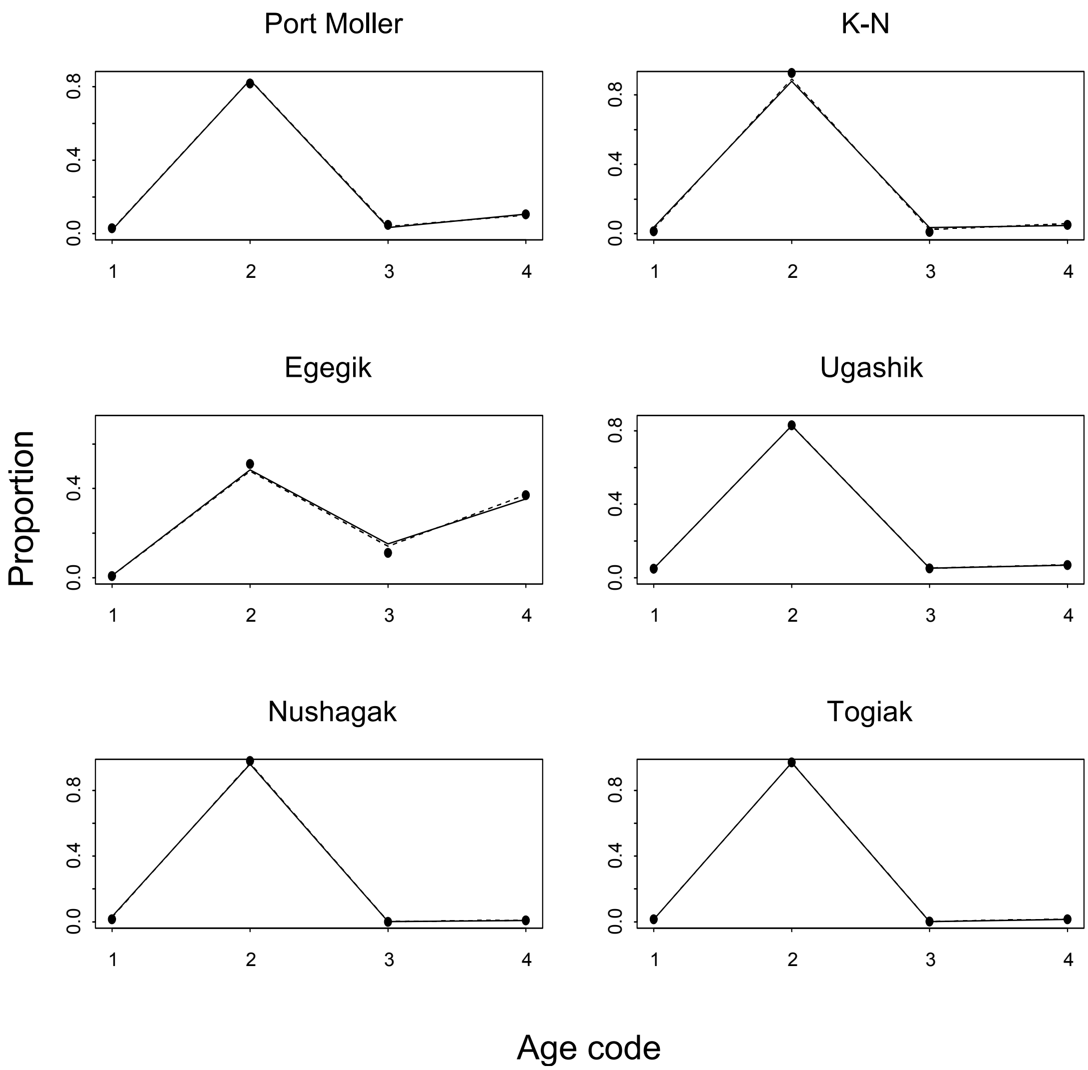


Figure 3.15. Comparison of the effect of considering the Port Moller gear selectivity for age-specific fish on the 2001 run forecasts made at day code 25 (July 4) and that of ignoring the selectivity. Dots indicate age composition (in proportion) observed up to the day, solid lines represent that in run forecasts made with the selectivity considered, and, and dashed lines are that in run forecasts made with the selectivity ignored. In case of the 'Port Moller' box, dots are age composition observed in the Port Moller fishery catch up to the day.

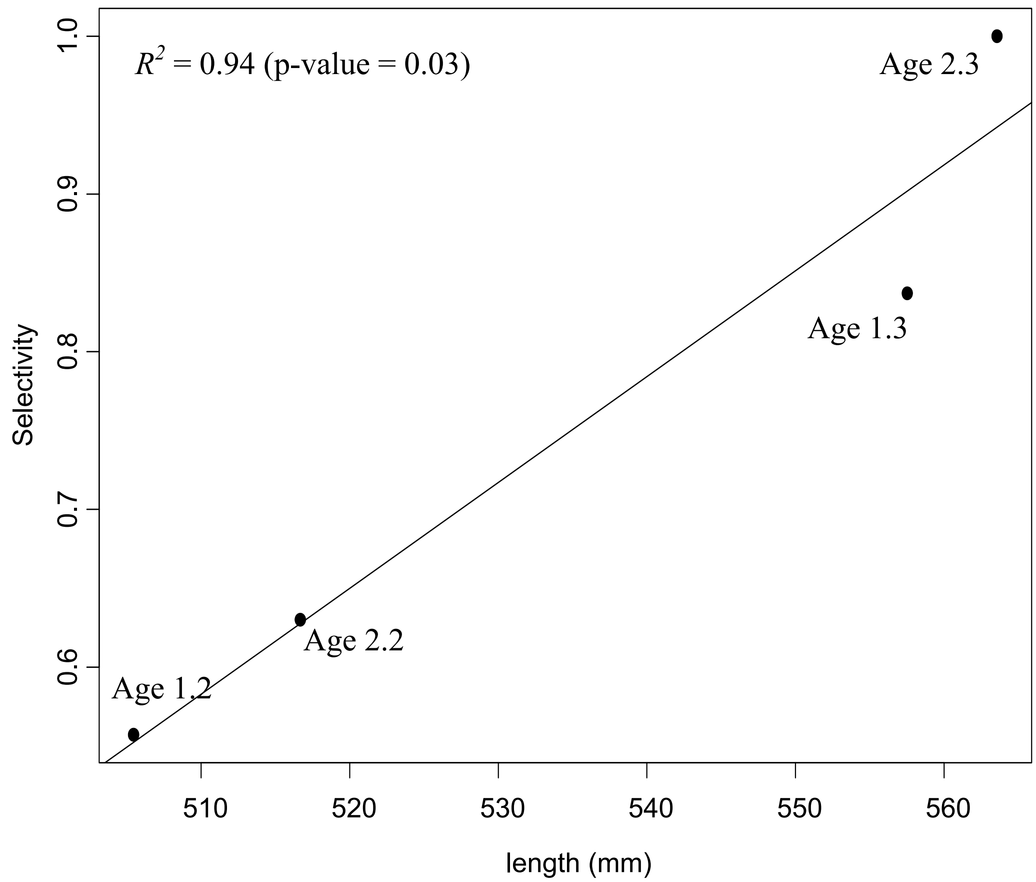


Figure 3.16. Relation between fish length, fish age, and the selectivity of the Port Moller gillnet fishery. The unit of selectivity is a fraction.

CHAPTER IV. UNCERTAINTY IN ESTIMATES OF RETURNS

INTRODUCTION

This chapter is an extension of Chapter 3, where I did the point estimates of returns. The objective of this chapter is to show uncertainty in estimates of returns. I use Bayes' law to build the distributions of forecasts. Fried and Hilborn (1988) used Bayes' law for an inseason forecast of total run size. In subsection '*Other studies of inseason forecast*' under section 1.2.3 of Chapter I, the paper is reviewed. Fried and Hilborn (1988) did not estimate stock-specific returns but only total run size.

METHODS

4.1. PARAMETER DISTRIBUTIONS

There were 16 parameters (Chapter 3). Regarding the parameter uncertainty, we are interested in the parameter distributions in addition to the point estimates. Generally when we estimate a predictive variable associated with distributions of parameters, we have to draw random values from the parameter distributions, and then use the values to build a distribution of the predictive variable (Gelman et al. 1995). If we apply the idea to the forecast algorithm of this thesis, we must:

(1) build distributions of parameters;
(2) draw a set of random values from the parameter distributions and pass the random values to the optimization stage (the next step);
(3) per the set of the random values from the parameter distributions in step 2, find optimized values of the predictive variables (returns) in the joint objective function;
(4) after saving the optimized values, repeat the above steps until the frequency distributions of the run estimates become smooth.

However, this above procedure could not be handled in ADMB (see the 'TPL file structure' section in Appendix I for the reasoning). As an alternative method, I treated the respective likelihood functions of the parameters (Equations 3.26, 3.34, and 3.36) as the objective functions. That is, I let the 16 parameters as well as run sizes become *not-fixed quantities* (see TPL file structure of Appendix I). And then, I estimated both the predictive variables (run sizes) and the parameters, simultaneously. Though this idea can be easily implemented into ADMB, it is not correct because the parameter estimates are affected by estimates of the predictive variables. The parameter estimates should be independent of the predictive variables (Figure 3.8). Because of the incorrectness, I compared the alternative method with the correct method where only data were used for the estimation of parameters. Figures 4.1, 4.2, and 4.3 show the likelihood profiles of the parameter estimates used to forecast the 1999, 2000, 2001 returns at July 4, respectively. In these figures, each solid line indicates the distribution of the respective parameter estimated by the alternative method, while each dashed line represents that estimated by the correct method. When I use the parameter estimates of the correct method for forecasts of returns, I pass the MLEs (i.e., fixed values) of the parameters to the PROCEDURE_SECTION, where the predictive returns are estimated. Regarding the *x*-axis labels and units in Figures 4.1, 4.2, and 4.3, refer to Table 4.5 and Table 3.3. The difference between those two methods in mode and variance of the respective parameter estimate was not significant except for the estimates of 'ga1' (G_1), 'ga2' (G_2), and 'ga3' (G_3) in Figure 4.1, and those of 'beta0' ($\beta_{0,t}$), 'beta1' ($\beta_{1,t}$) , and 'sigma2' (σ_t^2) in Figure 4.2. In sub-section, '4.2.4. Alternative method revisited,' I discuss whether the differences in parameter estimates between the two methods lead to a significant difference in the predictive returns.

4.2. ESTIMATION OF RETURNS

4.2.1. Bayesian framework

Both UW Alaska Salmon Program (ASP) and ADFG make preseason forecasts of stock- and age- specific returns. I used the preseason run forecasts for the prior information of returns in a Bayesian context. The following equation expresses the Bayes' law, ignoring a denominator constant term.

$$\Pr(\vec{r} \mid \text{data}) \propto \Pr(\text{data} \mid \vec{r}) \cdot \Pr(\vec{r}) \tag{4.1}$$

$\Pr(\vec{r})$ denotes the joint prior distribution of stock- and age- returns, and $\Pr(\vec{r} \mid \text{data})$ is the joint posterior distribution of returns.

$\Pr(\text{data} \mid \vec{r})$ represents the probability distribution of data, and it is replaced by the joint objective function of returns. To those who fully understand Chapter 3, the following (Equation 4.2) description may be redundant.

$$\begin{aligned}
&\Pr(\text{data} \mid \vec{r}) = \Pr\left(\text{data}_1, \text{data}_2, \ldots, \text{data}_{12} \mid \vec{r}\right) \\
&= \prod_{i=1}^{12} \Pr\left(\text{data}_\text{i} \mid \vec{r}\right) \\
&= \Pr(\tilde{R}) \cdot \Pr(\vec{U}_t \mid \vec{r}) \cdot \left(\prod_{s=1}^{5} \Pr(\tilde{r}_{s,\cdot}) \right) \cdot \left(\prod_{s=1}^{5} \Pr(\vec{j}_{s,t} \mid \vec{r}) \right)
\end{aligned} \tag{4.2}$$

- $\Pr(\tilde{R})$: Predictive normal distribution of unobserved data, total run size ($R = \sum_s \sum_a r_{s,a}$). This term corresponds to Equation 3.3.

- $\Pr(\vec{U}_t \mid \vec{r})$: Joint multinomial distribution of observed data, age-specific cumulative catches up to day *t*. This term is Equation 3.5, where age-specific returns are parameters.

- $\Pr(\tilde{r}_{s,\cdot})$: Predictive lognormal distribution of unobserved data, stock-specific run size. This term corresponds to Equation 3.15. Five lognormal distributions are considered for the five stocks (note product sign over stock *s*).

- $\Pr(\vec{j}_{s,t} \mid \vec{r})$: Joint multinomial distribution of observed data, stock- and age-specific cumulative runs up to day t. This term is Equation 3.21, where stock- and age- returns are parameters. Five multinomial distributions are considered for the five stocks (note product sign over stock s).

4.2.2. Prior distribution of returns

As the joint prior distribution of stock- and age- specific returns, I used two kinds: a uniform distribution and a normal distribution. When I deployed the uniform prior distribution of returns, the ' $\Pr(\vec{r})$ ' term in Equation 4.1 was just a constant.

In applying the normal distribution of returns to the joint prior distribution, I used preseason forecasts of returns made by UW ASP. That is, for the mean value of stock- and age- specific run, I used preseason forecast (point estimate) of the run size. However, both UW ASP and ADFG do not provide the variances of preseason forecasts. Thus, I needed to infer the variance from error mean square (MSE) of an ordinary regression model where I used the historical preseason forecast and the actual run size for the exploratory variable and the response variable, respectively. For example, when I inferred the variance of preseason forecast of Egegik- and age 1.2- run of year 2000, I built an ordinary regression model with the historical preseason forecasts of the runs prior to 2000 and the actual run sizes of the corresponding years. I took MSE of the regression model for the variance of preseason forecast of Egegik- and age 1.2- run of year 2000. Table 4.1 displays data that were used for the regression model. Tables 4.2, 4.3, and 4.4 show the variance estimates of preseason forecasts of the 1999, 2000, and 2001 returns, respectively.

The MSE values for preseason forecasts of Togiak returns of ocean age-2 were too small (almost zero) or could not be calculated, because the actual run sizes were negligible; Togiak- and age 1.2- run was reported as 0.1 million fish, and Togiak- and age 2.2- run as 0 million fish every year (Table 4.1). For the missing MSE values, I used the average of the MSE values for preseason forecasts of Togiak returns of ocean age-3 (Tables 4.2, 4.3, and 4.4).

The respective variance estimates of preseason forecasts of stock- and age-returns were independent by stock and age, except for Togiak returns of ocean age-2. I assume independence between preseason forecasts of 20 run sizes (Figure 3.1), so the joint prior density of returns is the product of the respective normal densities of returns. That is,

$$\Pr(\vec{r}) = \prod_{s}\prod_{a}\Pr(r_{s,a}) \tag{4.3}$$

$\Pr(r_{s,a})$ is the normal density of stock *s*- and age *a*- run, whose mean and variance are preseason forecast of the run, and MSE of the regression model of the historical actual run sizes against the corresponding preseason forecasts.

$$\Pr(r_{s,a}) \propto \frac{1}{\sqrt{MSE_{s,a}}} \cdot \exp\left(-\frac{\left[r_{s,a} - E(r_{s,a}) \right]^2}{2 \cdot MSE_{s,a}} \right) \tag{4.4}$$

where $E(r_{s,a})$ is preseason forecast of stock- and age- specific run.

4.2.3. Calculation of the joint posterior distribution of returns

The Markov Chain Monte Carlo (MCMC) calculation is implemented in ADMB. MCMC is a well-known method of calculating marginal posterior distributions. When we calculate Bayes' law where a multivariate density is involved, we have to integrate the multivariate distribution over the dimensions. It is almost impossible to analytically integrate a high-dimensional distribution over the dimensions. The MCMC is a method of numerically integrating a high-dimensional distribution and sampling from a posterior distribution to build marginal posterior distributions. ADMB MCMC method uses the Metropolis-Hastings algorithm. Studies associated with the MCMC method cite mainly the following literature: Gelman et al. (1995), and Gamerman (1997).

When I calculated the marginal posterior distributions of the predictive returns, I did one million MCMC runs, and sampled the results at intervals of 30 because of autocorrelation. Because of the sequential correlation of the Markov chain, we are advised to use the run results at intervals of some simulation runs. The procedure is called ‘thinning’ (Raftery and Lewis 1996; Patterson 1999).

4.2.4. Alternative method revisited

Because of differences in the parameter estimates between the alternative method and the correct method (Figures 4.1, 4.2 and 4.3), I checked how different the estimates of the predictive returns made by the alternative method were from those made by the classical method where point estimates (MLE) of parameters were used.

Figure 4.4 shows the posterior distributions of the 1999, 2000, and 2001 returns estimated at July 4 of the respective year by the alternative method (solid line) and the classical method (dashed line). In both cases, I incorporated run timing forecast (Table 2.5), and used the uniform densities for the prior densities of returns. In Figure 4.4, the modes of the posterior distributions made by the classical method are a little closer to the actual returns (vertical dotted line) than those made by the alternative method. But the differences were not significant (Figure 4.4), so I proceeded with the alternative method.

RESULTS

By the hind-casting procedure, I made forecasts of the 1999, 2000, and 2001 returns at three days of the respective year: day codes 15 (June 24), 20 (June 29), and 25 (July 4). Though I also made them at day code 30 (July 9) in Chapter 3, I do not in Chapter 4 because the day (July 9) is after the half point of the return season, and forecasts of returns made after the day are not interesting. I present the marginal posterior distributions of stock-specific returns, because forecasts of stock-specific returns are of the most interest to ADFG.

4.3. POSTERIOR DISTRIBUTIONS

It took about 17 minutes to do one million MCMC simulation runs in ADMB with a personal computer whose CPU speed and RAM size were 750 MHz and 192 MB, respectively.

4.3.1. Marginal posterior distributions of stock-specific returns

Figures 4.5, 4.6, and 4.7 show the marginal posterior distributions of stock-specific returns of 1999, 2000, and 2001 made at three days (day codes 15, 20, and 25). Regardless of the prior densities of returns, the modes of the posterior distributions approach the actual returns (vertical dotted line), and the variances of the posterior distributions become narrow as forecast time progresses during the respective season (year).

4.3.2. Uniform prior vs. normal prior

In Figures 4.5, 4.6, and 4.7, the posterior distributions of returns from the uniform priors (solid lines) are much wider than those from the normal priors. Tables 4.6, 4.7, and 4.8 present forecast errors in terms of relative error (%) between the modes of the posterior distributions of returns and the actual returns. The minus (-) sign indicates an under-forecast error. Generally the normal priors of returns led to smaller errors in forecasts than the uniform priors, except for the 1999 run forecasts (Tables 4.6, 4.7, and 4.8). In case of the 2001 forecasts (Table 4.8), the errors from the normal priors were all smaller than those from the uniform priors, except for the Togiak run forecasts made at June 29 and July 4.

DISCUSSION

4.4. PRESEASON FORECASTS

Preseason forecasts of returns are usually not accurate enough to be used for management. Absolute values of relative errors in preseason forecasts of stock-specific returns of 1999, 2000, and 2001 ranged from 4.4% to 130.3% (Figure 4.8). Despite the uncertainty, using the preseason forecast information for prior densities of returns increased the accuracy of the posterior distributions of returns (Tables 4.6, 4.7, and 4.8; Figures 4.5, 4.6, and 4.7).

However, as the season progresses, we should decrease our reliance on posterior distributions of returns from the normal prior densities of preseason forecasts, and increase our reliance on those from the uniform prior densities. As inseason data are accumulated, it is better not to incorporate the uncertain preseason forecast information. Figure 4.9 illustrates an example where forecasts (posterior distributions) of the 2000 returns were made at June 24 (an initial time of the season) and at July 9 (a middle time of the season). In Figure 4.9, dashed lines depict the posterior distributions from the normal prior, and solid lines depict the posterior distributions from the uniform prior. In the posterior distributions of June 24 (left column of Figure 4.9), the modes of the distributions from the normal prior are closer to the actual returns (vertical dotted line) than those from the uniform prior, but the distributions from the normal prior do not cover the actual returns securely, especially for the Ugashik and Nushagak returns. In the posterior distributions of July 9 (right column of Figure 4.9), the distributions from the uniform prior are obviously better than those from the normal; (1) the distributions from the uniform prior cover the actual returns more securely, and (2) their accuracy (in modes and variances) improves on that of the distributions of June 24 (left column).

4.5. PORT MOLLER FISHERY DATA

It costs about US $100,000 to deploy the Port Moller test fishery per season. There is no literature that evaluates the test fishery's value. The traditional inseason forecast method (Rogers' regression model) with the Port Moller catch data has been questionable. The determination coefficients (R^2) of Rogers' regression model ranged from 0.65 to 0.82, where data of outlier years (1990, 1994, 1997, and 2001) are excluded (Figure 3.2). When data of outlier years were included, the coefficient was only about

0.46 (Figure 1.7). In response to this uncertainty, a question may be raised: say, 'Is such a uncertain forecast worth the monetary value?'

I examine a contribution of the Port Moller catch data to forecasts of returns. Figure 4.10 shows the effect of absence of the Port Moller data on forecasts of returns, displaying forecasts (posterior distributions) of the 2001 returns made at June 24 (an initial time of the season) and July 14 (a final time of the season). In Figure 4.10, dashed lines represent the posterior distributions calculated without the objective functions of Port Moller data (the first and second components in Table 3.2), and solid lines are those calculated with them as well as the other objective functions. For both cases, the uniform prior densities of returns are used, and run timing information is incorporated accordingly. In posterior distributions of July 14 (right column of Figure 4.10), the two lines are almost identical; i.e. the absence of the Port Moller data does not make a difference in forecasts made at a final time of the season. However, in posterior distributions of June 24 (left column of Figure 4.10), the distributions made without the Port Moller data (dashed lines) are extremely inaccurate. This indicates that the Port Moller data are necessary to forecasts made at an initial time of the season.

Table 4.1. Data that are used to calculate the variances of preseason forecasts by stock and age. The unit of the actual and forecast run size is number in thousands. KN denotes Kvichak-Naknek.

Year	District	Age	Actual	Forecast	Year	District	Age	Actual	Forecast
1992	KN	1.2	2,930	3,300	1994	Ugashik	2.2	2,479	1,700
1992	KN	1.3	3,940	2,500	1994	Ugashik	2.3	2,252	400
1992	KN	2.2	5,236	6,400	1995	Ugashik	1.2	2,034	600
1992	KN	2.3	4,157	1,600	1995	Ugashik	1.3	709	1,700
1993	KN	1.2	2,727	3,500	1995	Ugashik	2.2	2,302	1,300
1993	KN	1.3	3,414	1,900	1995	Ugashik	2.3	955	1,400
1993	KN	2.2	5,180	5,100	1996	Ugashik	1.2	191	900
1993	KN	2.3	4,005	2,600	1996	Ugashik	1.3	3,167	3,700
1994	KN	1.2	2,208	5,400	1996	Ugashik	2.2	597	1,200
1994	KN	1.3	2,780	2,300	1996	Ugashik	2.3	1,218	2,000
1994	KN	2.2	20,158	13,400	1997	Ugashik	1.2	265	700
1994	KN	2.3	1,144	2,100	1997	Ugashik	1.3	597	700
1995	KN	1.2	3,434	2,200	1997	Ugashik	2.2	1,013	1,000
1995	KN	1.3	2,552	3,500	1997	Ugashik	2.3	326	500
1995	KN	2.2	22,780	36,000	1998	Ugashik	1.2	333	800
1995	KN	2.3	3,898	4,700	1998	Ugashik	1.3	352	600
1996	KN	1.2	795	2,400	1998	Ugashik	2.2	241	1,100
1996	KN	1.3	6,661	3,700	1998	Ugashik	2.3	827	700
1996	KN	2.2	1,114	2,400	1999	Ugashik	1.2	2,816	600
1996	KN	2.3	2,715	4,700	1999	Ugashik	1.3	328	1,000
1997	KN	1.2	1,272	6,200	1999	Ugashik	2.2	692	1,000
1997	KN	1.3	851	1,900	1999	Ugashik	2.3	198	100
1997	KN	2.2	882	1,700	2000	Ugashik	1.2	402	400
1997	KN	2.3	513	2,100	2000	Ugashik	1.3	0	4,500
1998	KN	1.2	2,476	6,200	2000	Ugashik	2.2	0	400
1998	KN	1.3	2,441	2,000	2000	Ugashik	2.3	0	400
1998	KN	2.2	1,180	4,600	1992	Nushagak	1.2	2,016	1,100
1998	KN	2.3	564	1,700	1992	Nushagak	1.3	1,878	2,500
1999	KN	1.2	10,269	5,800	1993	Nushagak	1.2	2,925	1,300
1999	KN	1.3	2,035	2,500	1993	Nushagak	1.3	3,907	3,800
1999	KN	2.2	4,252	8,400	1993	Nushagak	2.3	131	100
1999	KN	2.3	1,237	1,000	1994	Nushagak	1.2	1,299	1,900
2000	KN	1.2	402	3,000	1994	Nushagak	1.3	3,744	2,600
2000	KN	1.3	2,340	8,800	1994	Nushagak	2.2	73	200
2000	KN	2.2	723	3,600	1995	Nushagak	1.2	3,123	1,300
2000	KN	2.3	338	1,300	1995	Nushagak	1.3	2,890	3,300
1992	Egegik	1.2	413	700	1995	Nushagak	2.2	487	100
1992	Egegik	1.3	4,561	2,000	1995	Nushagak	2.3	96	100
1992	Egegik	2.2	8,863	5,300	1996	Nushagak	1.2	2,670	1,800
1992	Egegik	2.3	4,515	2,300	1996	Nushagak	1.3	4,790	4,800
1993	Egegik	1.2	513	400	1996	Nushagak	2.2	60	200
1993	Egegik	1.3	1,278	1,100	1996	Nushagak	2.3	322	200
1993	Egegik	2.2	11,061	11,000	1997	Nushagak	1.2	1,910	1,700
1993	Egegik	2.3	11,239	5,700	1997	Nushagak	1.3	2,472	3,600
1994	Egegik	1.2	403	400	1997	Nushagak	2.2	107	200

Table 4.1. (continued)

Year	District	Age	Actual	Forecast	Year	District	Age	Actual	Forecast
1994	Egegik	1.3	456	1,200	1997	Nushagak	2.3	110	100
1994	Egegik	2.2	6,063	2,900	1998	Nushagak	1.2	3,066	2,900
1994	Egegik	2.3	5,650	11,700	1998	Nushagak	1.3	2,280	2,900
1995	Egegik	1.2	1,397	800	1998	Nushagak	2.2	150	300
1995	Egegik	1.3	867	1,100	1998	Nushagak	2.3	86	0
1995	Egegik	2.2	9,598	5,500	2000	Nushagak	1.2	804	1,500
1995	Egegik	2.3	3,979	4,700	2000	Nushagak	1.3	0	3,800
1996	Egegik	1.2	335	800	2000	Nushagak	2.2	0	300
1996	Egegik	1.3	3,939	1,700	2000	Nushagak	2.3	0	300
1996	Egegik	2.2	3,113	6,000	1992	Togiak	1.2	111	100
1996	Egegik	2.3	4,721	7,200	1992	Togiak	1.3	575	400
1997	Egegik	1.2	497	1,300	1993	Togiak	1.2	132	100
1997	Egegik	1.3	1,117	3,600	1993	Togiak	1.3	403	400
1997	Egegik	2.2	4,963	6,700	1994	Togiak	1.2	101	100
1997	Egegik	2.3	2,607	2,300	1994	Togiak	1.3	328	400
1998	Egegik	1.2	368	600	1994	Togiak	2.3	53	100
1998	Egegik	1.3	573	1,000	1995	Togiak	1.2	189	100
1998	Egegik	2.2	880	3,700	1995	Togiak	1.3	460	400
1998	Egegik	2.3	3,099	3,300	1996	Togiak	1.2	50	100
1999	Egegik	1.2	3,173	600	1996	Togiak	1.3	429	600
1999	Egegik	1.3	985	1,100	1996	Togiak	2.3	37	100
1999	Egegik	2.2	4,246	4,700	1997	Togiak	1.2	64	100
1999	Egegik	2.3	993	1,300	1997	Togiak	1.3	124	300
2000	Egegik	1.2	0	1,100	1997	Togiak	2.3	29	100
2000	Egegik	1.3	0	5,000	1998	Togiak	1.2	43	100
2000	Egegik	2.2	0	4,400	1998	Togiak	1.3	229	300
2000	Egegik	2.3	0	3,100	1998	Togiak	2.2	6	0
1992	Ugashik	1.2	463	800	1998	Togiak	2.3	30	0
1992	Ugashik	1.3	1,626	1,600	1999	Togiak	1.2	341	100
1992	Ugashik	2.2	1,875	1,100	1999	Togiak	1.3	166	200
1992	Ugashik	2.3	1,750	500	1999	Togiak	2.2	31	0
1993	Ugashik	1.2	694	1,500	1999	Togiak	2.3	15	0
1993	Ugashik	1.3	692	1,600	2000	Togiak	1.2	0	100
1993	Ugashik	2.2	2,144	1,500	2000	Togiak	1.3	0	900
1993	Ugashik	2.3	2,310	900	2000	Togiak	2.2	0	0
1994	Ugashik	1.2	345	600	2000	Togiak	2.3	0	100
1994	Ugashik	1.3	391	900					

Table 4.2. The variances of the 1999 preseason forecasts by stock and age. The MSE values for preseason forecasts of Togiak returns of ocean age-2 could not be calculated, so the values were replaced by the average of the MSE values for preseason forecasts of Togiak returns of ocean age-3. That is, 5,652 is the mean value of 8,266 and 3,037 (see the bottom row). The unit is (number in thousands)^2.

	Age1.2	Age1.3	Age2.2	Age2.3
K-N	3,401,351	402,272	40,354,778	1,845,376
Egegik	111,155	986,878	5,580,280	11,947,424
Ugashik	112,754	178,560	38,138	402,164
Nushagak	426,042	375,807	3,666	1,845
Togiak	5,652	8,266	5,652	3,037

Table 4.3. The variances of the 2000 preseason forecasts by stock and age. The MSE values for preseason forecasts of Togiak returns of ocean age-2 could not be calculated, so the values were replaced by the average of the MSE values for preseason forecasts of Togiak returns of ocean age-3. That is, 5,457 is the mean value of 8,763 and 2,150 (see the bottom row). The unit is (number in thousands)^2.

	Age1.2	Age1.3	Age2.2	Age2.3
K-N	3,154,284	346,226	34,743,877	1,806,658
Egegik	96,060	854,278	4,664,666	10,621,201
Ugashik	94,326	153,293	33,322	435,728
Nushagak	426,042	375,807	3,666	1,845
Togiak	5,457	8,763	5,457	2,150

Table 4.4. The variances of the 2001 preseason forecasts by stock and age. The MSE values for preseason forecasts of Togiak returns of ocean age-2 could not be calculated, so the values were replaced by the average of the MSE values for preseason forecasts of Togiak returns of ocean age-3. That is, 24,482 is the mean value of 45,772 and 3,193 (see the bottom row). The unit is (number in thousands)^2.

	Age1.2	Age1.3	Age2.2	Age2.3
K-N	2,805,521	5,462,567	30,058,333	1,573,821
Egegik	98,919	2,188,391	4,122,386	9,148,752
Ugashik	106,882	1,895,483	56,287	382,698
Nushagak	355,097	609,315	3,238	13,327
Togiak	24,482	45,772	24,482	3,193

Table 4.5. Labels in Figures 4.1, 4.2, and 4.3. Table 3.3 shows a detailed description about the parameters and their units.

Label	**Correct notation**
beta0, beta1, sigma2	$\beta_{0,t}$, $\beta_{1,t}$, σ_t^2
logmu_KN, logsigma2_KN	$\mu_{1,t}$, $\sigma_{1,t}^2$
logmu_E, logsigma2_E	$\mu_{2,t}$, $\sigma_{2,t}^2$
logmu_U, logsigma2_U	$\mu_{3,t}$, $\sigma_{3,t}^2$
logmu_N, logsigma2_N	$\mu_{4,t}$, $\sigma_{4,t}^2$
logmu_T, logsigma2_T	$\mu_{5,t}$, $\sigma_{5,t}^2$
ga1, ga2, ga3	G_1, G_2, G_3

Table 4.6. Forecast errors (%) in posterior distributions of the 1999 returns calculated at the given day with the two prior densities of runs being used: uniform and normal. The error values are a relative difference between the modes of the posterior distributions in Figure 4.5 and the corresponding actual returns. The minus (-) sign indicates an under-forecast.

District	June 24		June 29		July 4	
	Uniform	Normal	Uniform	Normal	Uniform	Normal
K-N	-53.5	-5.7	4.5	-21.1	13.6	-1.7
Egegik	-38.5	-53.1	22.0	-32.0	60.6	-31.2
Ugashik	521.3	-25.5	123.8	-26.2	16.2	-27.1
Nushagak	-7.1	-25.5	-54.5	-29.9	51.3	-19.2
Togiak	22.0	-32.9	355.5	-18.2	-70.4	-31.4

Table 4.7. Forecast errors (%) in posterior distributions of the 2000 returns calculated at the given day with the two prior densities of runs being used: uniform and normal. The error values are a relative difference between the modes of the posterior distributions in Figure 4.6 and the corresponding actual returns. The minus (-) sign indicates an under-forecast.

District	June 24		June 29		July 4	
	Uniform	Normal	Uniform	Normal	Uniform	Normal
K-N	110.5	26.8	70.6	24.0	29.1	33.9
Egegik	54.0	25.1	67.8	21.0	29.1	20.8
Ugashik	339.7	132.5	302.5	132.7	353.7	122.5
Nushagak	-25.4	-27.7	41.7	-22.6	41.3	-21.5
Togiak	-2.9	-21.9	35.1	-20.5	-36.8	-21.7

Table 4.8. Forecast errors (%) in posterior distributions of the 2001 returns calculated at the given day with the two prior densities of runs being used: uniform and normal. The error values are a relative difference between the modes of the posterior distributions in Figure 4.7 and the corresponding actual returns. The minus (-) sign indicates an under-forecast.

District	June 24		June 29		July 4	
	Uniform	Normal	Uniform	Normal	Uniform	Normal
K-N	162.1	-3.4	104.2	0.2	84.7	5.7
Egegik	251.4	81.1	154.0	76.3	75.9	52.2
Ugashik	755.1	176.7	453.1	143.1	391.4	98.3
Nushagak	77.7	0.2	132.3	1.7	153.5	1.8
Togiak	-67.8	-65.8	14.0	-57.9	-24.2	-49.8

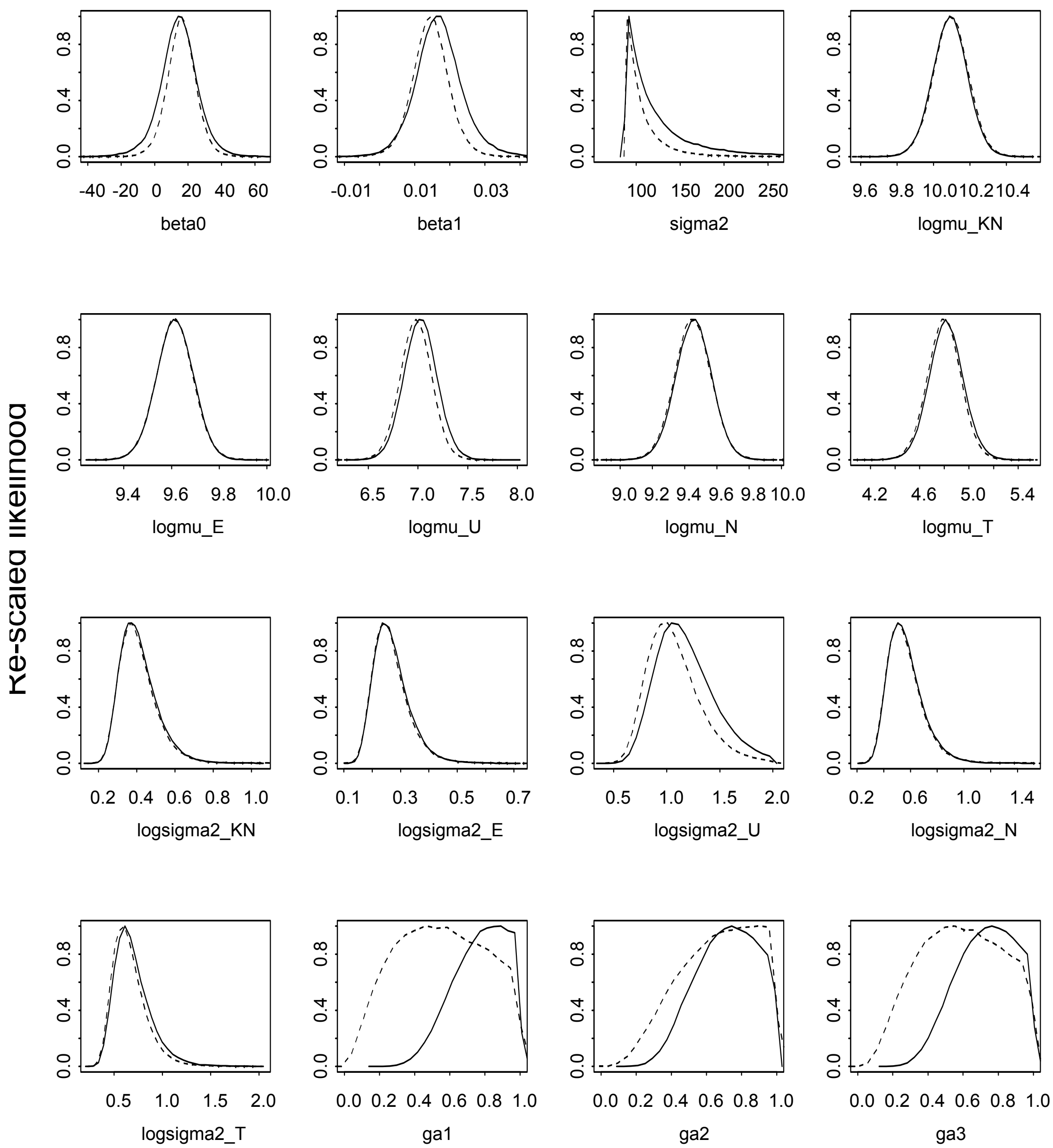


Figure 4.1. Comparison of the alternative method (solid line) and the correct method (dashed line) in the likelihood profiles of the parameter estimates that are used for making the 1999 run forecasts at July 4. Regarding the *x*-axis labels and units, refer to Table 4.5 and Table 3.3.

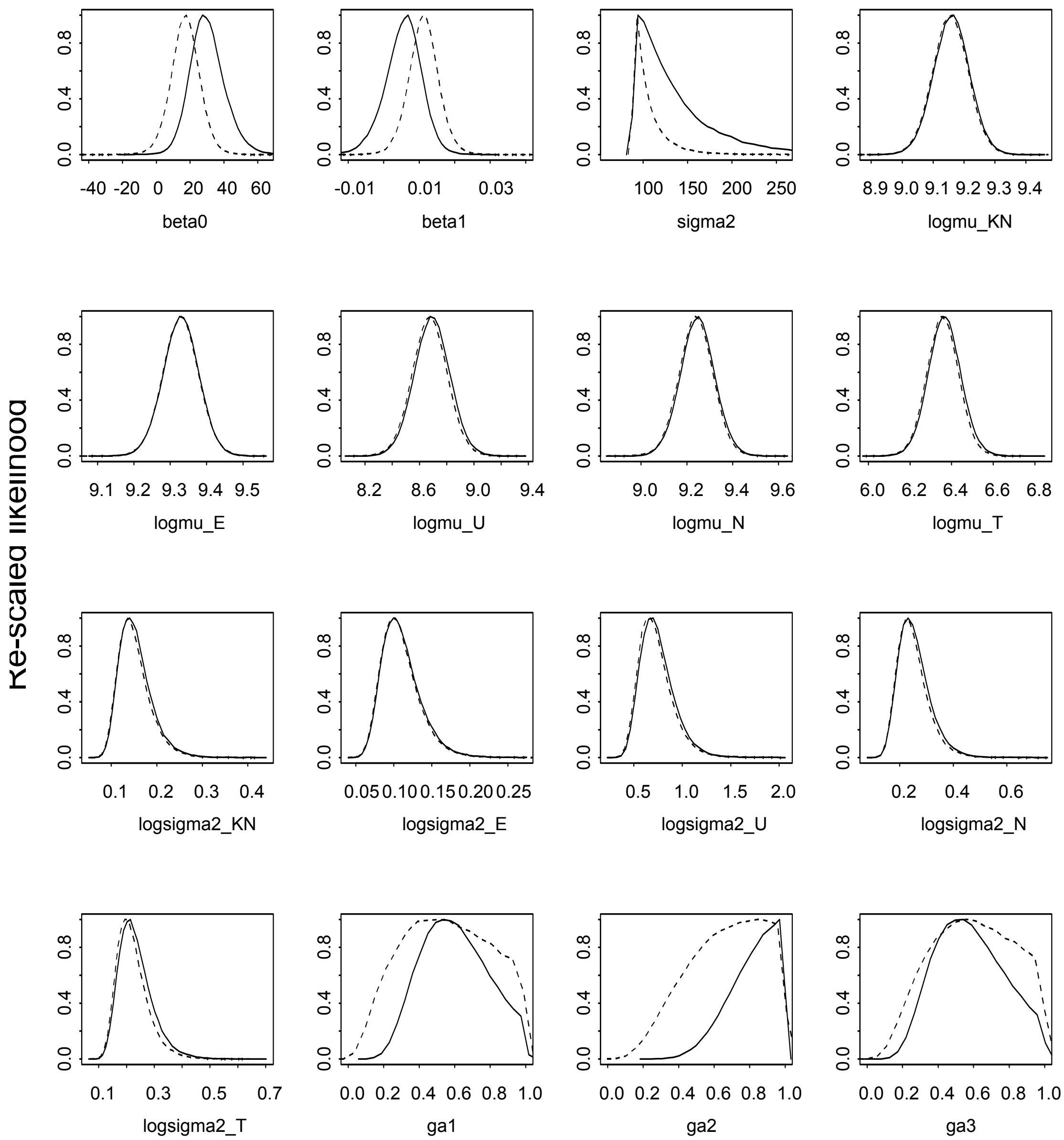


Figure 4.2. Comparison of the alternative method (solid line) and the correct method (dashed line) in the likelihood profiles of the parameter estimates that are used for making the 2000 run forecasts at July 4. Regarding the *x*-axis labels and units, refer to Table 4.5 and Table 3.3.

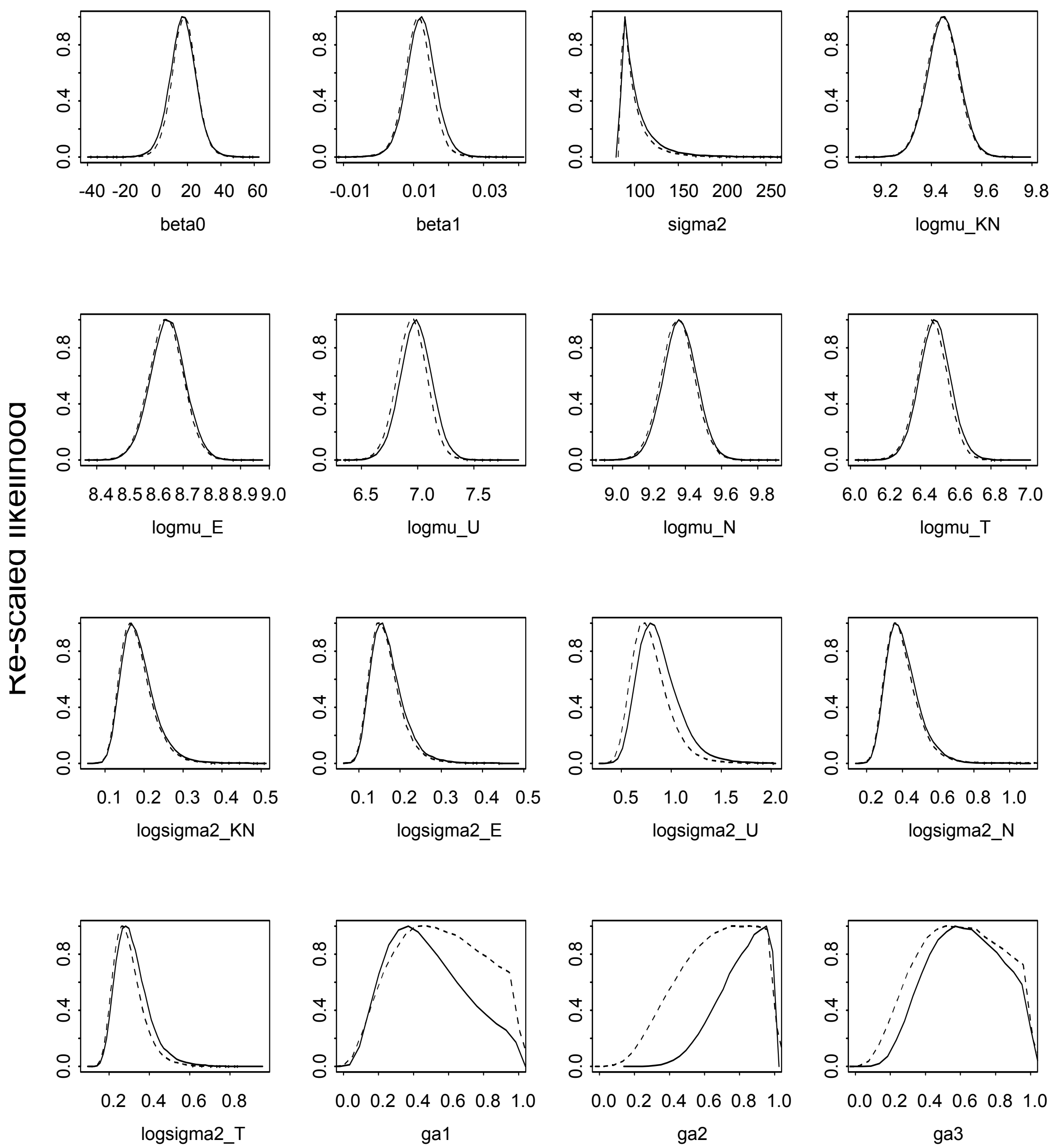


Figure 4.3. Comparison of the alternative method (solid line) and the correct method (dashed line) in the likelihood profiles of the parameter estimates that are used for making the 2001 run forecasts at July 4. Regarding the *x*-axis labels and units, refer to Table 4.5 and Table 3.3.

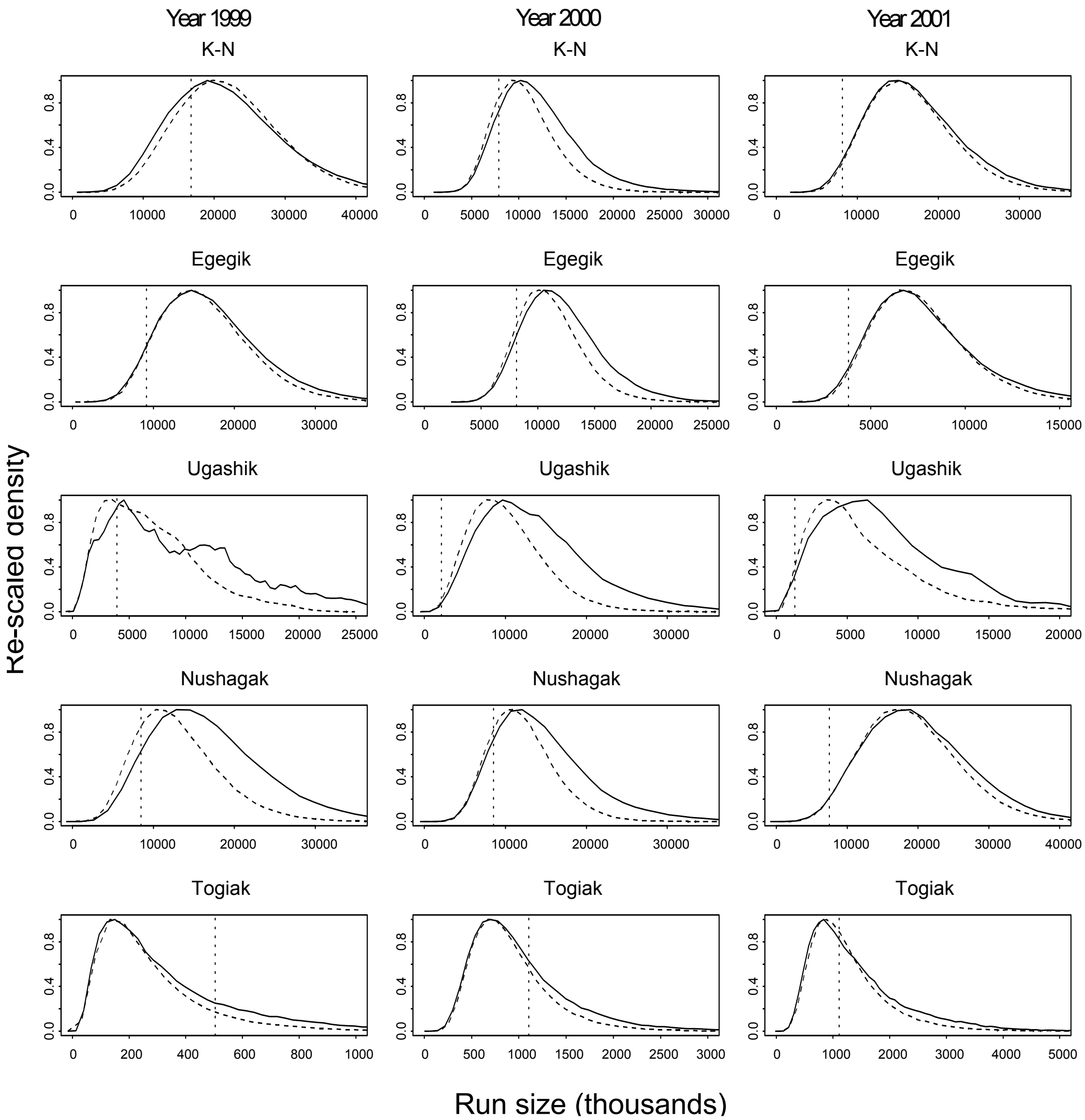


Figure 4.4. Comparison of the alternative method (solid line) and the classical method (dashed line) in posterior distributions of the 1999, 2000 and 2001 returns estimated at July 4 of the respective year. The vertical dot line refers to the actual run size. In both cases, I incorporated run timing forecast (Table 2.5) accordingly, and used the uniform densities for the prior densities of returns.

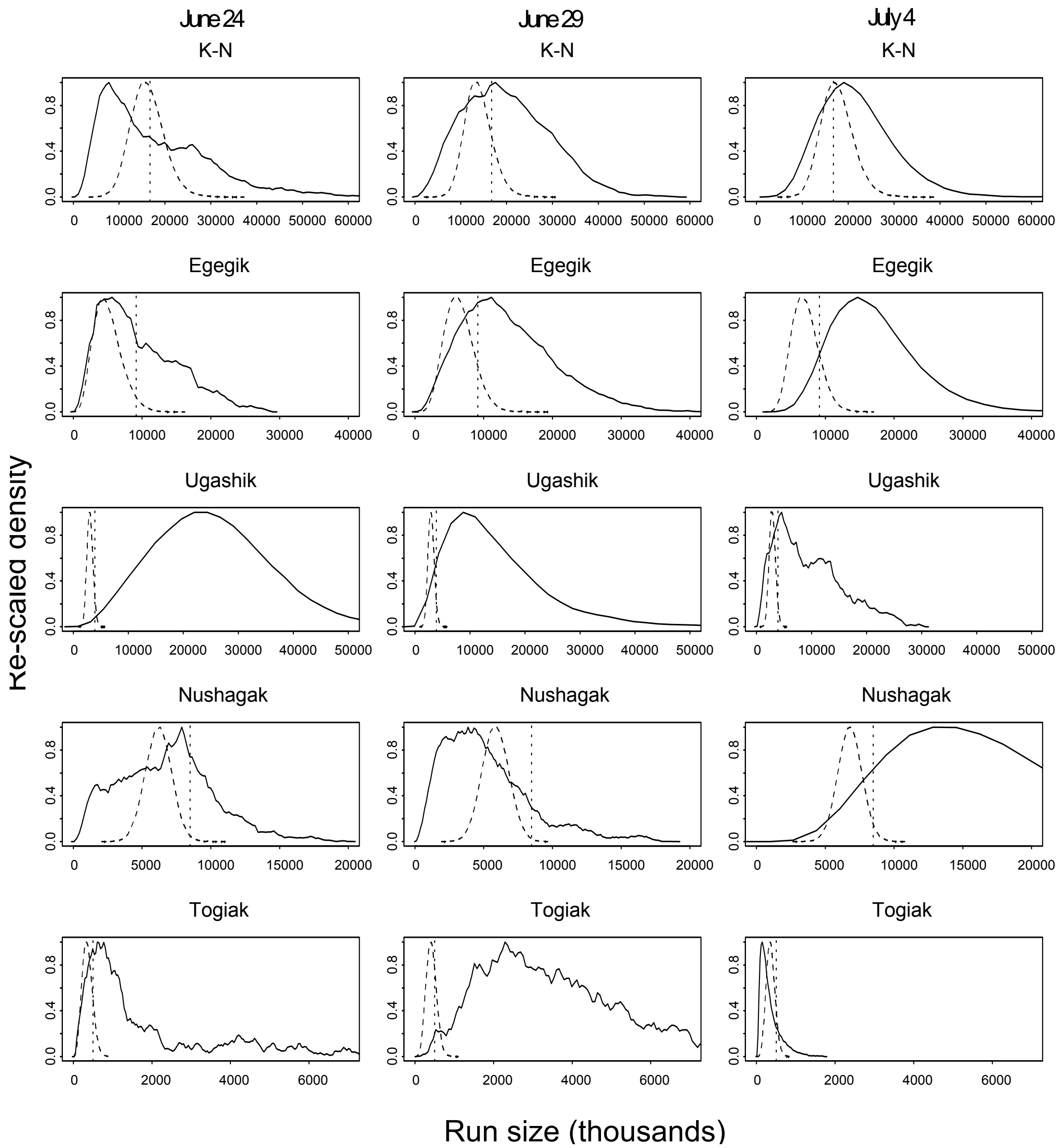


Figure 4.5. Marginal posterior distributions of stock-specific returns of 1999 made at three days: June 24, June 29, and July 4 (each column). Solid lines are the posterior distributions with the uniform prior densities used, and dashed lines are those with the normal prior densities used. Run timing forecast was incorporated accordingly (Table 2.5).

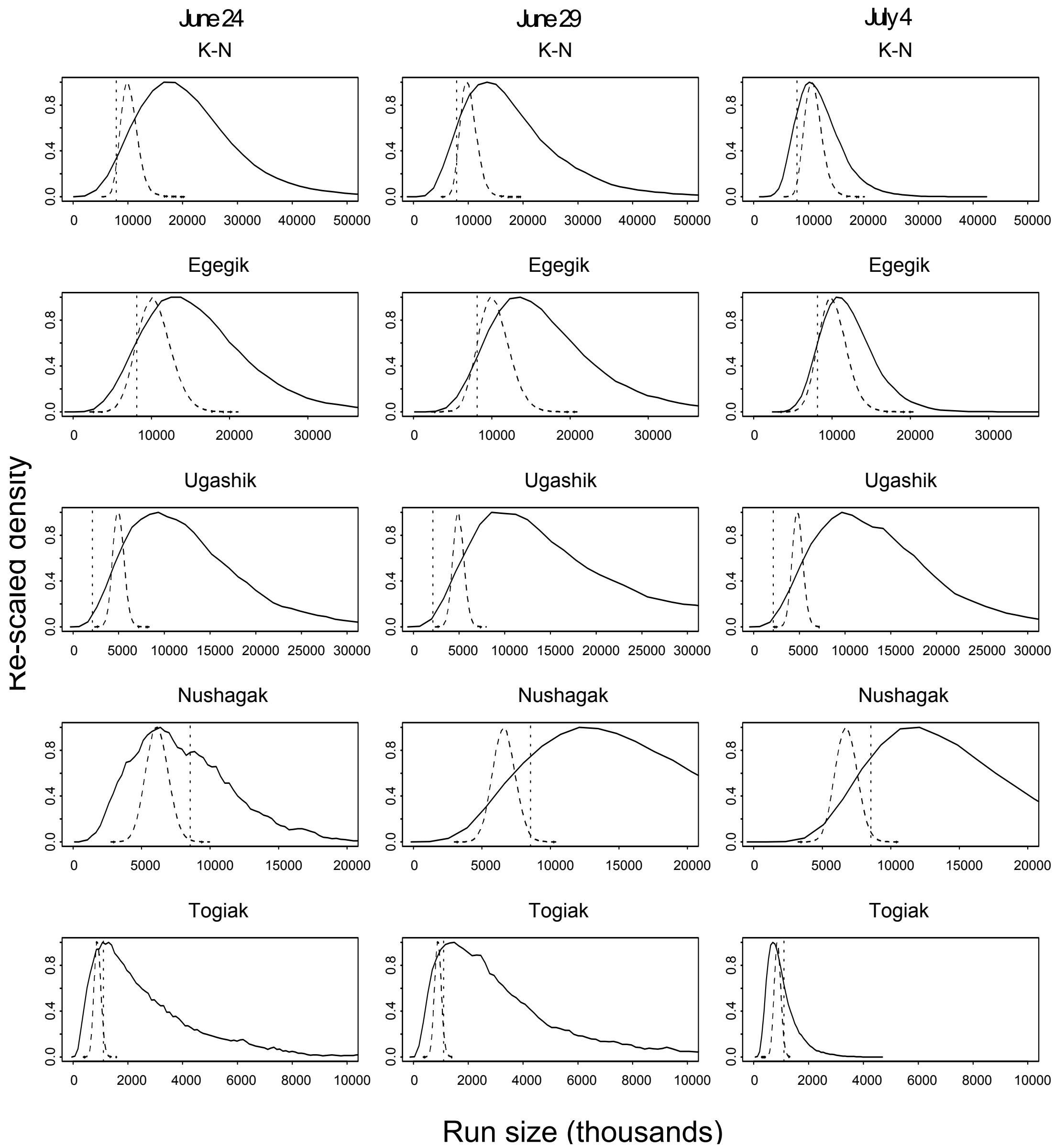


Figure 4.6. Marginal posterior distributions of stock-specific returns of 2000 made at three days: June 24, June 29, and July 4 (each column). Solid lines are the posterior distributions with the uniform prior densities used, and dashed lines are those with the normal prior densities used. Run timing forecast was incorporated accordingly (Table 2.5).

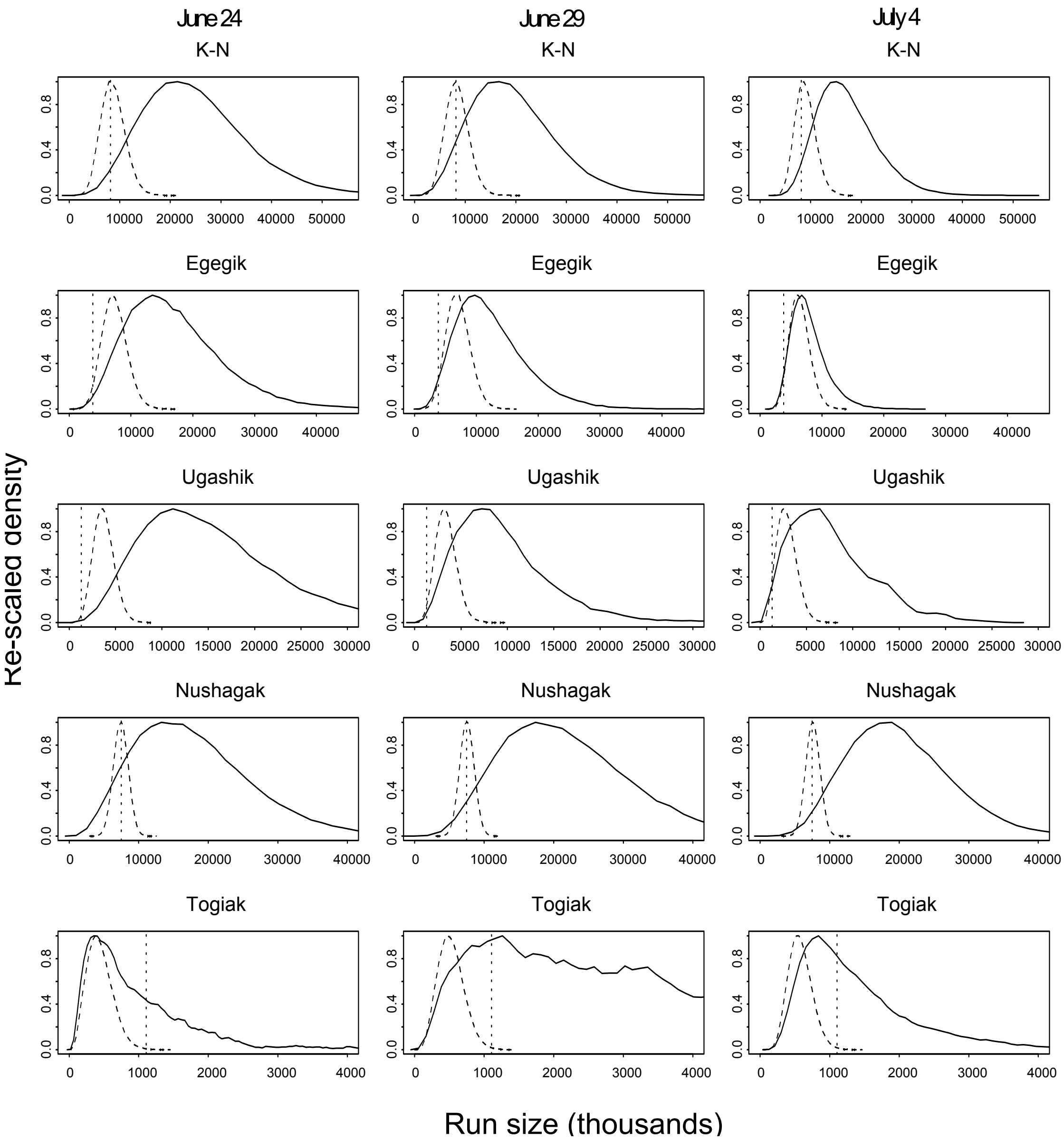


Figure 4.7. Marginal posterior distributions of stock-specific returns of 2001 made at three days: June 24, June 29, and July 4 (each column). Solid lines are the posterior distributions with the uniform prior densities used, and dashed lines are those with the normal prior densities used. Run timing forecast was incorporated accordingly (Table 2.5).

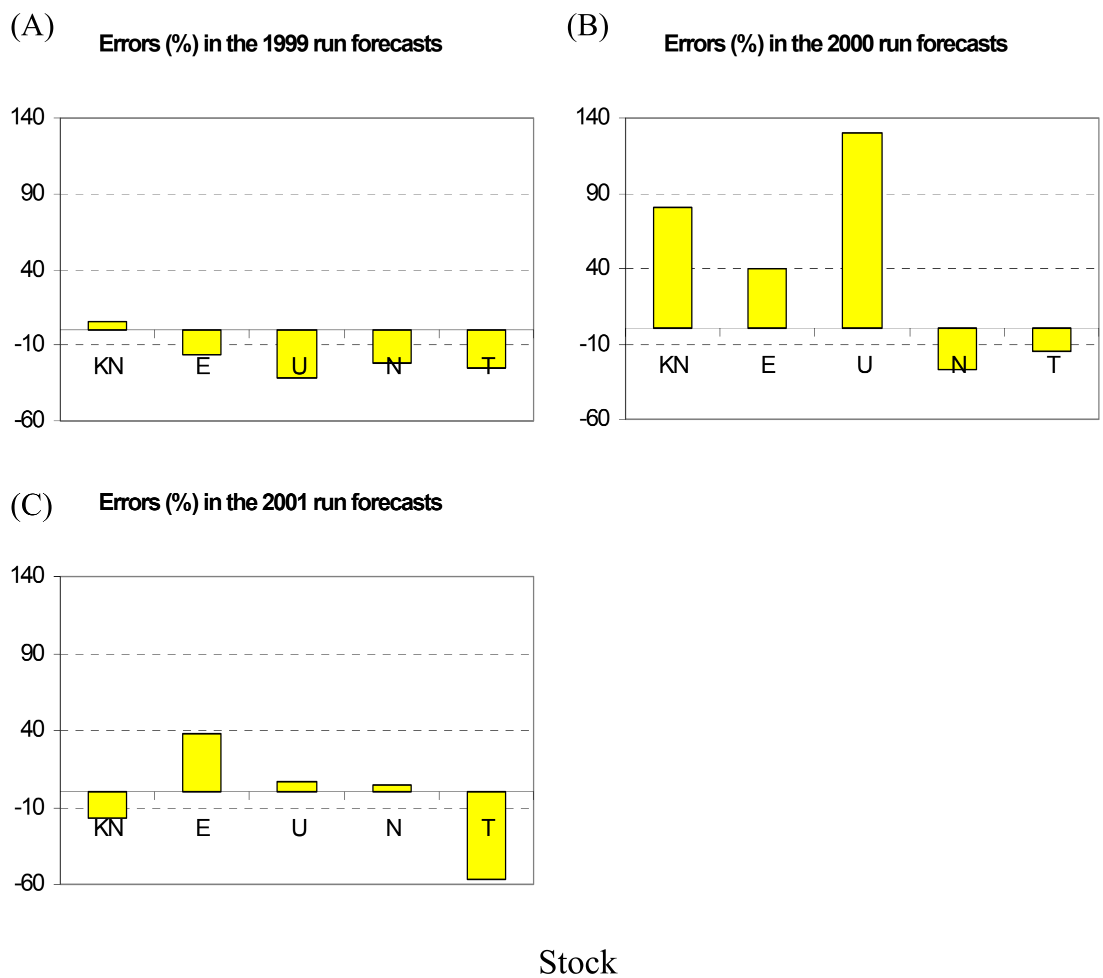


Figure 4.8. Relative errors (%) in preseason forecasts of stock-specific returns of 1999, 2000, and 2001. The negative errors indicate under-forecasts. KN: Kvichak-Naknek; E: Egegik; U: Ugashik; N: Nushagak; T: Togiak.

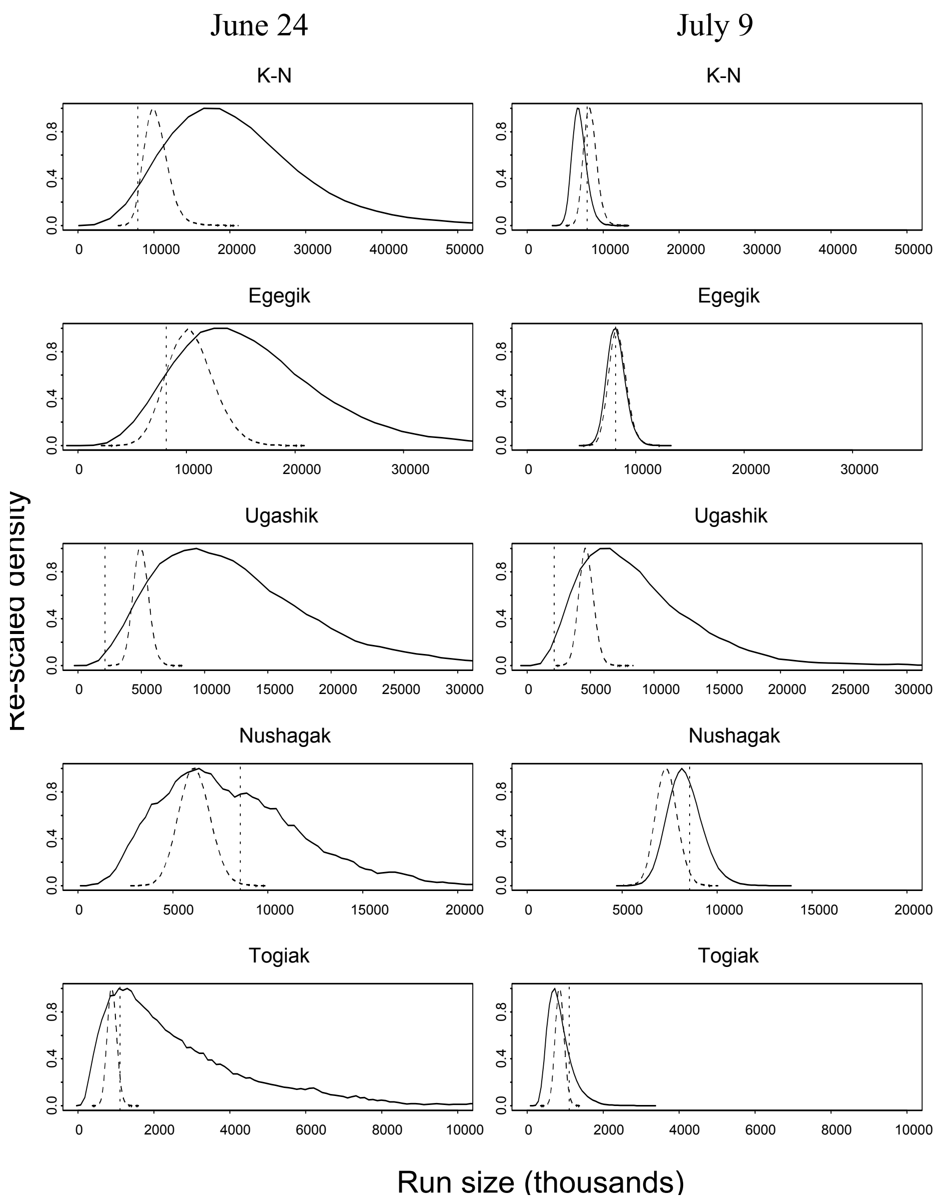


Figure 4.9. An example illustrating that the normal prior densities of returns decrease the forecast accuracy as time progresses during the season. Forecasts (posterior distributions) of the 2000 returns are made at June 24 (an initial time of the season; left column) and July 9 (a middle time of the season; right column). Dashed lines are the posterior distributions with the normal prior used, and solid lines are those with the uniform prior used.

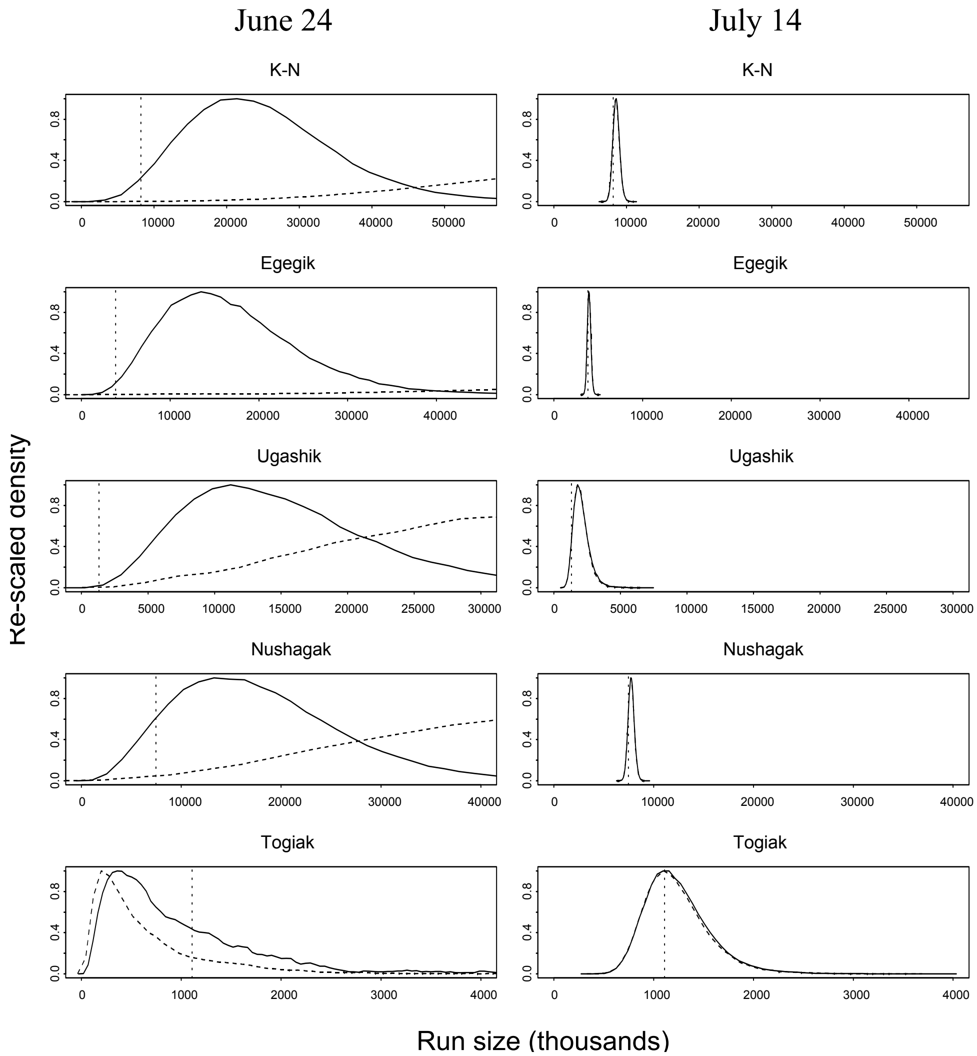


Figure 4.10. An example showing importance of the Port Moller fishery data. The 2000 run forecasts are made at June 24 (an initial time of the season; left column) and July 14 (a final time of the season; right column) without the Port Moller fishery data (dashed line) and with the data (solid line). The absence of the Port Moller data leads to extremely poor forecasts of returns during the initial stage of the season, but it does not make any difference during the final stage.

CHAPTER V. CONCLUSIONS

An accurate forecast of fish run timing will improve a forecast of salmon run size because it indicates the percentage of final run size that pass an area of interest on a certain day. Forecasts of returns, made with the incorporation of run timing detected from the Port Moller test fishery data, were less biased than those made without the run timing incorporation (Tables 3.13, 3.14, and 3.15; Figures 3.8, 3.9, and 3.10). However, the run timing detection by the Port Moller data is quite uncertain. For example, yearly Port Moller RTI evaluated even at the last day of the test fishery (day code 30) does not account for about 41% of variation in yearly run timing of four district fish (except the Togiak stock); in Figure 2.6, the determination coefficient (R^2) between Port Moller RTI of day code 30 and the inshore RTI is about 59% (= 0.77^2).

Also yearly Port Moller RTI does not capture well the fluctuation *magnitude* in yearly run timing of four district fish (Figure 2.6). Because I subtract the average of Port Moller RTI estimates of years prior to a season of interest from Port Moller RTI of the season to detect *how* early or *how* late fish run timing of the season is different from those of the past years (Equation 2.4), the fluctuation magnitude is an important statistic.

Because of the uncertainty in Port Moller RTI as a run timing estimate, I suggest that managers should also use other indicators in judging fish run timing. For example, seawater temperature may be a run timing index of the Bristol Bay sockeye salmon. A few studies reported that seawater temperature was negatively correlated with the run timing of the Bristol Bay sockeye salmon (Burgner 1980, Nishiyama 1984): the warmer the ocean is, the earlier the fish return. Variability in the fish ocean distribution in response to ocean temperature may explain the negative correlation. The distribution is farther north and closer to coastal waters near their natal streams during warm years and thus the fish can arrive at their home streams earlier than during cold years when the ocean distribution is farther south to the open waters of the North Pacific ocean (Rogers 1984). However, seawater temperature also is an uncertain indicator of fish run timing. Burgner (1991) reported that inter-annual differences in early spring ocean surface temperatures accounted for about 50% of the deviation in run timing of the Bristol Bay sockeye salmon.

Given the uncertainty in a run timing forecast, managers may opt to compare several hypotheses regarding fish run timing based on as much information as they can collect such as Port Moller RTI, environmental data, anecdotal stories, and experiences. Managers can simulate those hypotheses by simply changing day code in the forecast ADMB program (section 3.6. Incorporation of run timing forecast; Appendix). Figure 5.1 shows an example of the idea where forecasts (posterior distributions) of 2001 returns are estimated at day code 20 (June 29) under three hypotheses regarding fish run timing: (1) fish run timing in the 2001 season is earlier by five days than the average of those in the past year seasons (dashed line), (2) it is not different (dashed and three-dotted line), and (3) it is later by five days (solid line). Posterior distributions in the left column of Figure 5.1 are from the uniform prior densities, and those in the right column are from the normal prior densities of preseason forecasts.

In estimating returns, I used all data available (inseason and historical data) (Table 3.2). In addition, I could add information of preseason forecasts of returns into the estimation by the Bayesian method. During the initial stage of the season, posterior distributions of returns from the normal prior of preseason forecasts were generally better than those from the uniform prior. However, the contribution of the preseason forecasts will depend on how accurate they are. As inseason data are accumulated in time during the season, we should decrease reliance on the preseason forecast information

Regarding forecasts of the Bristol Bay sockeye salmon returns, many studies have been done. One of the recent studies is the research of Adkison and Peterman (2000). Adkison and Peterman (2000) examined errors of various forecast models with possible permutations of the following predictors: spawner-recruit relationships, air temperature, sea surface temperature, Pacific Decadal Oscillation, North Pacific Index, sibling returns to date, and last year's deviation from the expected return. Adkison and Peterman (2000) found that the accuracy of any model was not better that the historical accuracy of the ADFG and UW ASP forecasts. The study reminds me that it is very hard to make an accurate forecast in a large ecosystem.

The value of this thesis is the development of a forecast algorithm for the Bristol Bay sockeye salmon returns rather than a remarkable improvement of the forecast accuracy. Managers may find the forecast algorithm useful for a management tool.

Recently Flynn and Hilborn (*In preparation*) developed a new model whose forecast of sockeye salmon run size to Bristol Bay is much less biased than the traditional forecast model. As future work, I suggest that the model should be incorporated into the current forecast algorithm of this thesis.

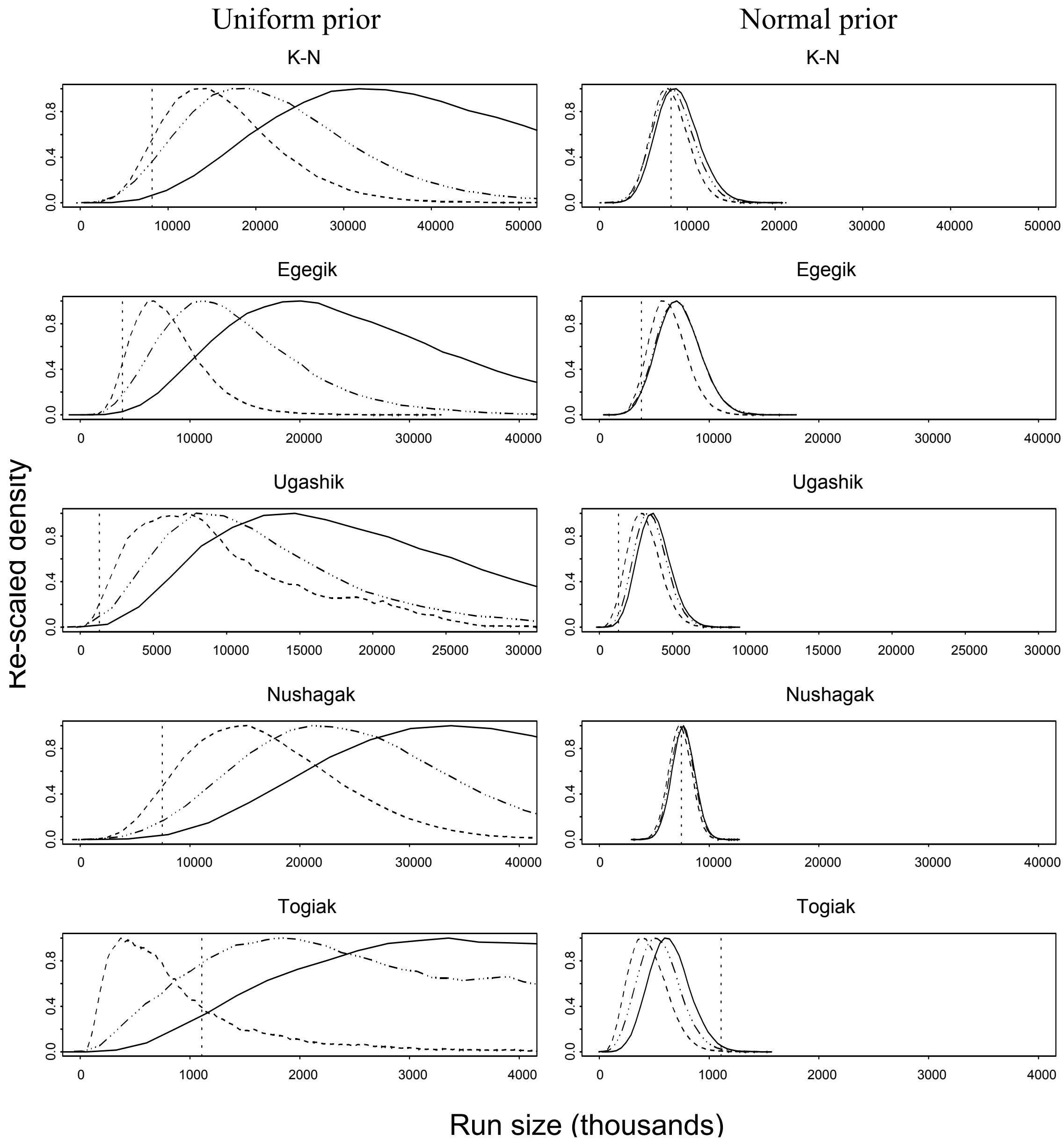


Figure 5.1. An example exploring several hypotheses regarding run timing. Posterior distributions of the 2001 returns are estimated at day code 20 (June 29) under three hypotheses: (1) fish run timing in the 2001 season is later by five days than those of the past year seasons (solid line), (2) it is not different from those of the past year seasons (dashed and dotted line), and (3) it is earlier by five days than those of the past year seasons (dashed line). Posterior distributions in the left column are from the uniform prior, and those in the right column are from the normal prior (preseason forecasts).

REFERENCES

Adkison, M.D., and R.M. Peterman. 2000. Predictability of Bristol Bay, Alaska, sockeye salmon returns one to four years in the future. N. Am. J. Fish. Manage. 20: 69-80.

Alaska Department of Fish and Game (ADFG). 1998. Division of Commercial Fisheries, annual management report. Bristol Bay area. Regional Information Report, No. 2A99-18. 155 pp.

Anonymous. 1994. *Manual*. Autodif: A C++ array language extension with automatic differentiation for use in nonlinear modeling and statistics. Otter Research Ltd., Sidney.

Anonymous. 2000. *Manual*. An introduction to AD Model Builder version 4: For use in nonlinear modeling and statistics. Otter Research Ltd., Sidney.

Brannian, L.K., O.A. Mathisen, and D.A. McCaughran. 1982. Variance estimates of sockeye salmon predictions with reference to the Egegik River system of Bristol Bay, Alaska. Final Report for the period January 1, 1982-June 30, 1982 to Alaska Department of Fish and Game. Contract No. 82-0769. University of Washington, Fisheries Research Institute, Seattle. 105 pp.

Burgner, R.L. 1980. Some feature of ocean migrations and timing of Pacific salmon, p. 153-164. *In* W.J. McNeil and D.C. Himsworth (eds.). Salmonid Ecosystems of the North Pacific. Oregon State University Press, Corvallis.

Burgner, R.L. 1991. Life history of sockeye salmon (*Oncorhynchus nerka*), p. 1-117. *In* C. Groot and L. Margolis (eds.). Pacific salmon life histories. University of British Columbia Press, Vancouver, BC, Canada.

Casella, G., and R.L. Berger. 1990. Statistical inference. Duxbury Press, Belmont, California. 650 pp.

Chen, X.H., and P.A. Shelton. 1996. A nonparametric forecast model of inshore Atlantic cod (*Gadus morhua*) landings based on biomass, cumulative landings, and water temperature. Can. J. Fish. Aquat. Sci. 53: 558-562.

Crone, P.R., and D.B. Sampson. 1998. Evaluation of assumed error structure in stock assessment models that use sample estimates of age composition, p. 355-370. *In* Funk, F, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang (eds.). Fishery stock assessment models. Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks.

Eggers, D., and S. Fried. 1984. 1982 Bristol Bay test fishing projects. Alaska Department of Fish and Game (ADFG) Technical Data Report 117: 81 pp.

Eggers, D.M., and D.E. Rogers. 1987. The cycle of runs of sockeye salmon (*Oncorhynchus nerka*) to the Kvichak River, Bristol Bay, Alaska: cyclic dominance or depensatory fishing?, p. 343-366. *In* H.D. Smith, L. Margolis, and C.C. Wood (eds.). Sockeye salmon (*Oncorhynchus nerka*) population biology and future management. Can. Spec. Publ. Fish. Aquat. Sci. 96.

Flynn, L., and R. Hilborn. *In preparation*. Test fishery indices for sockeye salmon as affected by age composition and environmental variables.

Foerster, R.E. 1968. The sockeye salmon, *Oncorhynchus nerka*. Bull. Fish. Res. Board Can. 162.

Fredin, R.A. 1980. Trends in North Pacific salmon fisheries, p. 59-119. *In* W.J. McNeil and D.C. Himsworth (eds.). Salmonid Ecosystems of the North Pacific. Oregon State University Press, Corvalis.

Fredin, R.A., R.L. Major, R.G. Bakkala, and G. Tanonaka. 1977. Pacific salmon and the high seas salmon fisheries of Japan. Natl. Ocean. Atmos. Admin., Natl. Mar. Fish. Serv., Northwest and Alaska Fish. Center., Seattle, Proc. Rep., 327 p.

French, R., H. Bilton, M. Osako, and A. Hartt. 1976. Distribution and origin of sockeye salmon (*Oncorhynchus nerka*) in offshore waters of the North Pacific Ocean. Int. North Pac. Fish. Comm. Bull. 34.

Fried, S.M., and R. Hilborn. 1988. Inseason forecasting of Bristol Bay, Alaska, sockeye salmon (*Oncorhynchus nerka*) abundance using Bayesian probability theory. Can. J. Fish. Aquat. Sci. 45: 850-855.

Fried, S.M., and H.J. Yuen. 1987. Forecasting sockeye salmon (*Oncorhynchus nerka*) returns to Bristol Bay, Alaska: A review and critique of methods, p.273-279. *In* H.D. Smith, L. Margolis, and C.C. Wood (eds.). Sockeye salmon (*Oncorhynchus nerka*) population biology and future management. Can. Spec. Publ. Fish. Aquat. Sci. 96.

Fried, S.M., B.A. Cross, and H.J. Yuen. 1988. A synopsis and critique of forecasts of sockeye salmon returning to Bristol Bay, Alaska in 1988. Alaska Department of Fish and Game, Division of Commercial Fisheries. 32 pp.

Gamerman, D. 1997. Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference. Texts in statistical science. Chapman & Hall, London. 245 pp.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 1995. Bayesian data analysis. Texts in statistical science. Chapman & Hall, London. 526 pp.

Helton, D.R. 1991. An analysis of the Port Moller offshore test fishing forecast of sockeye and chum salmon runs to Bristol Bay, Alaska. M.S. thesis, University of Washington, Seattle. 122 pp.

Hilborn, R., D. Rogers, R. Steen, and W. Lew. 1999. Forecast of the 2000 run of sockeye salmon to Bristol Bay. Univ. Washington. School of Fisheries, FRI-UW 9910. Seattle.

Hilborn, R. *Personal communication*. Email address: rayh@u.washington.edu

Hoar, W.S. 1976. Smolt transformation: evolution, behavior, and physiology. J. Fish. Res. Board Can. 33: 1234-1252.

Johnson, N.L., and S. Kotz. 1970. Continuous univariate distributions 1. Houghton Mifflin Company, Boston. 300 pp.

Koo, T.S.Y. 1962. Age designation in salmon. Contribution No. 106., p. 37-48. *In* T.S.Y. Koo (ed.). Studies of Alaska red salmon. University of Washington Press. Seattle, Washington. College of Fisheries, University of Washington. 1: 449 pp.

Lew, W.R. *Personal communication*. Email address: wlew@fish.washington.edu

Mathews, S. 1967. The economic consequences of forecasting sockeye salmon runs to Bristol Bay, Alaska: A computer simulation study of the potential benefits to a salmon canning industry from accurate forecasts of the runs. Ph.D. dissertation, Univ. of Washington, Seattle. 238 pp.

Mathisen, O.A., and M. Berg. 1968. Growth rates of the char in the Vardnes River, Troms, northern Norway. Institute of Freshwater Research, Drottningholm, Report 48.

McKinstry, C.A. 1993. Forecasting migratory timing and abundance of pink salmon (*Oncorhynchus gorbuscha*) runs using sex-ratio information. MS thesis, Univ. of Washington, Seattle. 74 pp.

Minard, R.E., and C.P. Meacham. 1987. Sockeye salmon (*Oncorhynchus nerka*) management in Bristol Bay, Alaska, p. 336-342. *In* H.D. Smith, L. Margolis, and C.C. Wood (eds.). Sockeye salmon (*Oncorhynchus nerka*) population biology and future management. Can. Spec. Publ. Fish. Aquat. Sci. 96.

Mundy, P.R. 1979. A quantitative measure of migratory timing illustrated by application to the management of commercial salmon fisheries. Ph.D. dissertation, Univ. of Washington, Seattle, Wash. 86 pp.

Nishiyama, T. 1974. Energy requirements of Bristol Bay sockeye salmon in the central Bering Sea and Bristol Bay, p. 321-343. *In* D.W. Hood and E.J. Kelly (eds.). Oceanography of the Bering Sea. Univ. Alaska, Inst. Mar. Sci., Occ. Publ. 2.

Nishiyama, T. 1984. Two important factors in oceanic stages of salmon: food and temperature, p. 227-258. *In* Proceedings of the Pacific salmon biology conference (USSR, USA, Canada, Japan), Yuzhno-Sakhalinsk, USSR, 1978. Tikhookeanskii Nauchno-Issledovatel'skii Institut Rybnogo Khoziaistva i Okeanografii, Vladivostok, USSR.

Patterson, K.R. 1999. Evaluating uncertainty in harvest control law catches using Bayesian Markov Chain Monte Carlo virtual population analysis with adaptive rejection sampling and including structural uncertainty. Can. J. Fish. Aquat. Sci. 56: 208-221.

Pearcy, W.G. 1992. Ocean ecology of North Pacific salmonids. Washington Sea Grant Program, University of Washington Press, Seattle and London.

Peterman, R.M. 1981. Form of random variation in salmon smolt-to-adult relations and its influence on production estimates. Can. J. Fish. Aquat. Sci. 38: 1113-1119.

Peterman, R.M. 1982a. Nonlinear relation between smolts and adults in Babine Lake sockeye salmon (*Oncorhynchus nerka*) and implications for other salmon populations. Can. J. Fish. Aquat. Sci. 39: 904-913.

Peterman, R.M. 1982b. Model of salmon age structure and its use in preseason forecasting and studies of marine survival. Can. J. Fish. Aquat. Sci. 39: 1444-1452.

Raftery, A.E., and S.M. Lewis. 1996. Implementing MCMC, p. 115-127. *In* W.R. Gilks, S. Richardson, and D.J. Spielgehalter (eds.). Markov Chain Monte Carlo in practice. Chapman and Hall, London.

Rogers, D.E. 1984. Trends in abundance of northeastern Pacific stocks of salmon, p. 100-127. *In* W.G. Pearcy (ed.). The Influence of Ocean Conditions on the Production of Salmonids in the North Pacific. Oregon State University Sea Grant Program, Corvallis.

Rogers, D.E. 1986. Pacific salmon, p. 461-476. *In* D.W. Hood and S.T. Zimmerman (eds.) The Gulf of Alaska: physical environment and biological resources.

U.S. National Ocean Service, Ocean Assessments Division, Alaska Office; U.S. Minerals Management Service, Alaska OCS Region, Washington, D.C.

Rogers, D.E. 1987. The regulation of age at maturity in Wood River sockeye salmon (*Oncorhynchus nerka*), p. 78-89. *In* H.D. Smith, L. Margolis, and C.C. Wood (eds.). Sockeye salmon (*Oncorhynchus nerka*) population biology and future management. Can. Spec. Publ. Fish. Aquat. Sci. 96.

Rogers, D.E. 1994. Forecasts of the 1995 sockeye salmon runs to Bristol Bay. FRI-UW 9417. Univ. Washington, School of Fisheries, Seattle.

Rogers, D.E. 1998. Forecasts of the 1999 sockeye salmon runs to Bristol Bay using traditional FRI methodology. FRI-UW-9818. Univ. Washington, School of Fisheries, Seattle.

Rogers, D.E. 1999. A Bristol Bay almanac for 1999. FRI-UW 9906. Univ. Washington, School of Fisheries, Seattle.

Rogers, D.E., and R.P. Steen. 2000. A Bristol Bay almanac for 2000. SOF-UW-2005. Univ. Washington, School of Fisheries, Seattle. 59 pp.

Rogers, D.E., B.J. Rogers, T.P. Quinn, R. Steen, R. Hilborn, G.T. Ruggerone, T. Rogers, C. Boatright, and B. Ernst. 1999. Operations manual for Fisheries Research Institute Field Campus in Alaska. UW, FRI, Alaska Salmon Program. 74 pp.

Seber, G.A.F. 1982. The estimation of animal abundance and related parameters. Macmillan Publishing Co., Inc. New York. 654 pp.

Springborn, R.R., N.D. Lampsakis, and V.F. Gallucci. 1998. A time density model to estimate run size and entry timing in a salmon fishery. N. Am. J. Fish. Manage. 18: 391-405.

Straty, R.R. 1974. Ecology and behavior of juvenile sockeye salmon (*Oncorhynchus nerka*) in Bristol Bay and the eastern Bering Sea, p. 285-320. *In* D.W. Hood and E.J. Kelly (eds.). Oceanography of the Bering Sea with emphasis on renewable resources. Univ. Alaska, Inst. Mar. Sci., Occ. Publ. 2.

Townsend, R.L., P. Westhagen, D. Yasuda, J.R. Skalski, and K. Ryding. 1996. Evaluation of the 1995 predictions of the run-timing of wild migrant yearling chinook in the Snake River Basin using Program Real Time. Center for Quantitative Science, School of Fisheries, University of Washington, Seattle, Washington. Technical Report to Bonneville Power Administration Project No. 91-051. DOE/BP-35885-9. 64 pp.

Townsend, R.L., D. Yasuda, and J.R. Skalski. 1997. Evaluation of the 1996 predictions of the run-timing of wild migrant spring/summer yearling chinook in the Snake River Basin using Program Real Time. School of Fisheries, University of Washington, Seattle. Technical Report to Bonneville Power Administration Project No. 91-051. DOE/BP-91572-1. 30 pp.

Verkhunov, A.V. 1995. Climate change and variability of the Bering Sea current system, p. 73-80. *In* R.J. Beamish (ed.). Climate change and northern fish populations. Can. Spec. Publ. Fish. Aquat. Sci. 121.

Walters, C.J., and S. Buckingham. 1975. A control system for intraseason salmon management, p. 105-137. *In* Proceedings of a workshop on salmon management, February 24-28. CP-75-2. International Institute for Applied Systems Analysis, Laxenburg.

Zehna, P.W. 1966. Invariance of maximum likelihood estimators. Annals of Mathematical Statistics 37: 744.

Zheng, J., and O.A. Mathisen. 1998. Inseason forecasting of southeastern Alaska pink salmon abundance based on sex ratios and commercial catch and effort data. N. Am. J. Fish. Manage. 18: 872-885.

APPENDIX I. INTRODUCTION TO ADMB

In this appendix, I briefly describe ADMB for those who wish to use the forecast ADMB program developed in this thesis. For a detailed description about ADMB, refer to the manual (Anonymous 1994, 2000) available free from the following website: http://otter-rsch.com/admodel.htm

Quick recipe for running ADMB program

For source code, two text files are required: a TPL file and a DAT file. Implement computation code in a TPL file, and put data in a DAT file (see examples of a TPL file and a DAT file in Appendix II).

When a TPL file and a DAT file are ready, we have to compile and link the TPL file before running the program. The ADMB command, 'makeadm' compiles the TPL file and links it to ADMB library. Once the compile and link are successful, an executable file will be generated automatically. The name of the executable file is the same as that of the TPL file except for the extension name. The extension name of a TPL file is 'TPL,' while that of its executable file is 'EXE.' If the executable file is successfully run with the DAT file, several files are generated. Of these, some important files are a PAR file, an STD file, a COR file, and a REP file. The PAR file has the parameter estimates, the STD file has not only the parameter estimates but also the standard deviations, and the COR file has the variance-covariance matrix of the estimates. The REP file contains the output of REPORT_SECTION in the TPL file. As an option, we may want the MCMC or likelihood profile computation. An HST file has the marginal distribution of the respective parameter after the MCMC computation. A PLT file contains the likelihood profile of the respective parameter after the likelihood profile computation.

In the following lists, I summarize the above paragraph, assuming the names of a TPL file and a DAT file are 'general.tpl' and 'yr2001d20.dat,' respectively. I display ADMB key words in bold font.

- **makeadm** general

  (compile and link TPL file, *general*)

- general **-ind** yr2001d20.dat

  (run executable file, *general* with DAT file, *yr2001d20.dat*)

- general **-ind** yr2001d20.dat **-mcmc** 1000000 **-mcsave** 30

  (optional computation: do one million MCMC runs and save the results every 30 MCMC runs)

- general **-ind** yr2001d20.dat **-lprof**

  (optinal computation: compute the likelihood profile)

TPL file structure

A TPL file consists of up to nine sections. Three of these sections are required: DATA_SECTION, PARAMETER_SECTION, and PROCEDURE_SECTION. In DATA_SECTION, data values are set, and they are treated as constants. In PARAMETER_SECTION, quantities of estimation interest are declared, and they are treated as *not-fixed values* (*quantities*). The reason why I use term, 'not-fixed values' is to avoid confusion. The general meaning of parameters is different from that of parameters in ADMB, where only quantities of estimation interest are called parameters[1]. Quantities, which we do not intend to estimate, must be declared outside PARAMETER_SECTION, and be treated as *fixed values*. The estimation of not-fixed quantities is done in PROCEDURE_SECTION, where the objective function of not-fixed quantities is repeatedly differentiated with respect to all the respective not-fixed quantities.

In implementing the forecast algorithm of this thesis into ADMB, stock- and age-specific returns are not-fixed quantities. The 16 parameter estimates associated with the

---

[1] Some of ADMB users call quantities of estimation interest 'free parameters,' but the terms are not found in ADMB manual.

joint objective function are fixed quantities, and should be declared outside PARAMETER_SECTION; they should remain fixed during PROCEDURE_SECTION (during the estimation of returns). The ADMB requirement prevents the use of random values from the parameter distributions (see four-step procedure in section 4.1).

APPENDIX II. FORECAST ADMB CODE

TPL FILE

I show an example of a TPL file here. '//' is followed by comments. The first value in DAT file is to be passed to 'daycode' in the first line under DATA_SECTION. By changing the value, users can explore as many hypotheses regarding run timing as they want (section 3.6. Incorporation of run timing forecast; Chapter 5).

```
DATA_SECTION
  init_int daycode;   //today's daycode

  init_ivector whichLike(1,4); //which likelihood

  //Port Moller regression component
  init_number pmx; // daily cumulative index (cpue) for Port M. for the regression

  init_vector Ud(1,5);  //Cumulative runs of the five districts

  vector ncump(1,5); //length of the elements in the historical cump (e.g. KNcump)

  // Data for Multinomial A (PM)
  init_vector pmUa(1,4);  //the cumulative age-specific catches from PM fishery
  init_int ssize_MA // effective sample size for Mutinomial A
  number pm_offset
  number runs_offset

  // change data to be proportions (and use assumed sample size)
  !! pmUa /= sum(pmUa);
  !! pm_offset =  -1.0* ssize_MA * (pmUa* log(pmUa)); //Added assumed sample size

  // Data for Multinomial B (Samples from fisheries of dif districts)
  init_matrix Uda(1,5,1,4);  //five districts and four ages
  init_vector ssize_MB(1,5);
  matrix ssizemat_MB(1,5,1,4);

 LOCAL_CALCS
   for (i=1;i<=5;i++)   {
     ssizemat_MB(i)=ssize_MB(i);
     // change data to be proportions (and use assumed sample size)
     Uda(i) /= sum( Uda(i) );
     runs_offset +=  -1.0*ssize_MB(1)*(Uda(i)* log(Uda(i)));
   }
 END_CALCS

  init_int prior_type

  init_matrix mean_prior(1,5,1,4)
```

```
 init_matrix cv_prior(1,5,1,4)

 //Data for the Port Moller index calculation
 init_int betadim; //the number of betas in PM regression
 init_int rpmindD; //rows of PM index data
 init_int cpmindD; //columns of PM index data
 init_int pmn; //rows of data matrix Xmat

 init_ivector drows(1,5);  //rows of the five-district data sets
 init_int dcols;  //columns of the five-district data sets

 init_int gagen; //rows of the age ratio data
 number gagen2;

 matrix agefmat1(1,gagen,1,4); //PM age freq
 matrix agefmat2(1,gagen,1,4); //district age freq
 init_matrix agefreqpmmat(1,gagen,1,5); //age frequency of Port Moller catch
 init_matrix agefreqdistmat(1,gagen,1,5); //age freq. of district run

LOCAL_CALCS
   for(int i=1; i<=gagen; i++)   {
     for(int j=1; j<=4; j++)  {
        agefmat1(i,j)=agefreqpmmat(i,j+1);
        agefmat2(i,j)=agefreqdistmat(i,j+1);
     }
   }
END_CALCS

 matrix xmat(1,pmn,1,betadim);
 matrix ymat(1,pmn,1,1); //column vector
 init_matrix pmmat(1,rpmindD,1,cpmindD); //Port Moller data

 int r;
 !!r=1;
LOCAL_CALCS
   for(i=1; i<=rpmindD; i++)
     {
     if(pmmat(i,2)==daycode) //2nd column is daycode
       {
       xmat(r,1)=pmmat(i,4);
       xmat(r,2)=pmmat(i,5);
       ymat(r,1)=pmmat(i,6);
       r=r+1;
       }
     }
END_CALCS

 !! int dr1=drows(1);
 init_matrix KN_m(1,dr1,1,dcols)  //matrix form of KN data

 !! int dr2=drows(2);
 init_matrix E_m(1,dr2,1,dcols)  //matrix form of Egegik data

 !! int dr3=drows(3);
```

```
 init_matrix U_m(1,dr3,1,dcols)  //matrix form of Ugashik data

 !! int dr4=drows(4);
 init_matrix N_m(1,dr4,1,dcols)  //matrix form of Nushagak data

 !! int dr5=drows(5);
 init_matrix T_m(1,dr5,1,dcols)  //matrix form of Togiak data

 int counter;
 int i;
 int ii;
 int j;
 int k;

 vector oldKNcump(1,50);
 vector oldEcump(1,50);
 vector oldUcump(1,50);
 vector oldNcump(1,50);
 vector oldTcump(1,50);

// start  the DATA SET 1
LOCAL_CALCS
  ii=0;
  for(i=1;i<=drows(1);i++)
    {
     if(KN_m(i,5)==daycode) //5th column is daycode
       {
       ii= ii+1;
             if(KN_m(i,6)==0)
         oldKNcump(ii)=0.00001;
       else
         oldKNcump(ii)=KN_m(i,6); //6th column is proportion
       }
    }
END_CALCS

 vector KNcump(1,ii)

LOCAL_CALCS
  for(k=1; k<=ii; k++)
    KNcump(k)=oldKNcump(k);
END_CALCS
//end  the DATA SET 1

// start  the DATA SET 2
LOCAL_CALCS
  ii=0;
  for(i=1;i<=drows(2);i++)
    {
     if (E_m(i,5)==daycode) //5th column is daycode
       {
       ii= ii+1;
       if(E_m(i,6)==0)
         oldEcump(ii)=0.00001;
```

```
        else
          oldEcump(ii)=E_m(i,6); //6th column is proportion
        }
    }
END_CALCS

 vector Ecump(1,ii)

LOCAL_CALCS
  for(k=1; k<=ii; k++)
     Ecump(k)=oldEcump(k);
END_CALCS
//end  the DATA SET 2

//start  the DATA SET 3
LOCAL_CALCS
  ii=0;
  for(i=1;i<=drows(3);i++)
    {
      if (U_m(i,5)==(daycode)) //5th column is daycode
        {
         ii= ii+1;
         if(U_m(i,6)==0)
            oldUcump(ii)=0.00001;
         else
            oldUcump(ii)=U_m(i,6); //6th column is proportion
        }
    }
END_CALCS

 vector Ucump(1,ii)

LOCAL_CALCS
  for(k=1; k<=ii; k++)
      Ucump(k)=oldUcump(k);
END_CALCS
//end  the DATA SET 3

//start  the DATA SET 4
LOCAL_CALCS
  ii=0;
  for(i=1;i<=drows(4);i++)
    {
      if (N_m(i,5)==daycode) //5th column is daycode
        {
         ii= ii+1;
         if(N_m(i,6)==0)
             oldNcump(ii)=0.00001;
         else
             oldNcump(ii)=N_m(i,6); //6th column is proportion
        }
    }
END_CALCS
```

```
 vector Ncump(1,ii)

LOCAL_CALCS
  for(k=1; k<=ii; k++)
    Ncump(k)=oldNcump(k);
END_CALCS
//end  the DATA SET 4

// start  the DATA SET 5
LOCAL_CALCS
  ii=0;
  for(i=1;i<=drows(5);i++)
     {
      if (T_m(i,5)==(daycode)) //5th column is daycode
        {
        ii= ii+1;

        if(T_m(i,6)==0)
           oldTcump(ii)=0.00001;
        else
           oldTcump(ii)=T_m(i,6); //6th column is proportion
        }
     }
END_CALCS

 vector Tcump(1,ii)

LOCAL_CALCS
  for(k=1; k<=ii; k++)
     Tcump(k)=oldTcump(k);
END_CALCS
//end  the DATA SET 5

PARAMETER_SECTION
 init_bounded_vector pops_tmp(1,20,1.,100000.,2);

 //betas in PM regression
 init_number b0pm(1); //phase 1
 init_number b1pm(1); //phase 1

 number sig2pm;  //sigma squared in PM regression
               //sig2pm can be expressed as a function of data, b0pm and b1pm

 init_vector lognmu_tmp(1,5,1); //mu in lognormal;//phase 1
 init_bounded_vector lognsig2_tmp(1,5,0.,2.,1); //sigma2 in lognormal; //phase 1

 init_bounded_vector gage_tmp(1,3,0.,1.,1);

 sdreport_number totr; //total r
 sdreport_vector pops_dist(1,5);
 sdreport_matrix pops(1,5,1,4); //five districts and four ages

 sdreport_vector betaPM(1,betadim);
 sdreport_number varpm;  //equal to sig2pm
```

```
  sdreport_vector lognmu(1,5);
  sdreport_vector lognsig2(1,5);

  sdreport_vector gage(1,3);

  number fbetasvar;   //negative log likelihood of betas and var in PM regression
  number flognparam1; //negative log likelihood of log normal parameters
  number flognparam2; //negative log likelihood of log normal parameters
  number flognparam3; //negative log likelihood of log normal parameters
  number flognparam4; //negative log likelihood of log normal parameters
  number flognparam5; //negative log likelihood of log normal parameters
  number fgearages;   //negative log likelihoodd of age selecivity by PM gillnet fishery

  number fpm; //Port Moller predictive density
  number fpmage; //multinomial with PM age data
  number fage; //multinomial with runs age data

  number flognormal; //likelihood for the lognormal/gamma

  vector pm_multinom(1,4); //declare//PM multinomial elements
  vector pmPredprop(1,4); //vector of PM age-specific proportions

  matrix multinom(1,5,1,4); //declare//Run multinomial elements
  matrix Predprop(1,5,1,4); //declare
  matrix Normal_value(1,5,1,4);

  matrix smat(1,betadim,1,betadim);

  objective_function_value f; //negative logarithm

INITIALIZATION_SECTION
  pops_tmp 1000.
  b0pm 25.7
  b1pm 0.02
  lognmu_tmp 7.0
  lognsig2_tmp 0.5
  gage_tmp 0.5

PRELIMINARY_CALCS_SECTION
  ncump(1)=KNcump.indexmax();
  ncump(2)=Ecump.indexmax();
  ncump(3)=Ucump.indexmax();
  ncump(4)=Ncump.indexmax();
  ncump(5)=Tcump.indexmax();
  gagen2=gagen;  //number of data rows for estimating MLE of age-specific gear selectivity

PROCEDURE_SECTION
  f=0.0;
  pops.initialize(); //assign zero values
  betaPM.initialize();
  lognmu.initialize();
  lognsig2.initialize();
  varpm=0.0;
```

```
betaPM(1)=b0pm;
betaPM(2)=b1pm;
varpm=sig2pm;

fbetasvar=negloglike_betasvarF(betaPM);
f=fbetasvar;

for(i=1;i<=5;i++)
  for(ii=1;ii<=4;ii++)
    pops(i,ii)=pops_tmp(ii+4*(i-1));

totr=sum(pops);

for(i=1;i<=5;++i)
  pops_dist(i)=sum(pops(i));

if(whichLike(3)==1)
{
 fpm=(1.0/(2.0*varpm*square(1000)))*square(totr-(betaPM(1)+betaPM(2)*pmx)*1000);
 f+=fpm;
}

gage=gage_tmp;

fgearages=negloglike_gearage(gagen2, gage);
f+=fgearages;

if(whichLike(1)==1)
{
 pm_multinomial(); //multinomial A //returns pm_multinom
 fpmage=sum(pm_multinom) - pm_offset;
 f+=fpmage;
}

if(whichLike(2)==1)
{
 multinomial();  //multinomial B
 fage=sum(multinom) - runs_offset;
 f+=fage;
}

lognmu=lognmu_tmp;
lognsig2=lognsig2_tmp;

flognparam1=negloglike_logmusig2F(lognmu(1), lognsig2(1), ncump(1), Ud(1)/KNcump);
flognparam2=negloglike_logmusig2F(lognmu(2), lognsig2(2), ncump(2), Ud(2)/Ecump);
flognparam3=negloglike_logmusig2F(lognmu(3), lognsig2(3), ncump(3), Ud(3)/Ucump);
flognparam4=negloglike_logmusig2F(lognmu(4), lognsig2(4), ncump(4), Ud(4)/Ncump);
flognparam5=negloglike_logmusig2F(lognmu(5), lognsig2(5), ncump(5), Ud(5)/Tcump);

f+=flognparam1;
f+=flognparam2;
f+=flognparam3;
```

```
  f+=flognparam4;
  f+=flognparam5;

  if(whichLike(4)==1)
  {
    flognormal=0.0;
    for (i=1;i<=5;++i)
        {
         flognormal += -1.0*log(1.0/sum(pops(i)) )+(1/(2*lognsig2(i))) * square( log(sum(pops(i)))-lognmu(i)
);
        }
    f+=flognormal;
  }

  if(prior_type==1)
     {
      Normal_prior(); //each elements by age and district
      f+=sum(Normal_value);
     }

FUNCTION dvariable negloglike_betasvarF(dvar_vector betas)
   dvariable negl;
   dvar_matrix bmat(1,betadim,1,1);//column vector
   bmat(1,1)=betas(1);
   bmat(2,1)=betas(2);
   varpm=sum(square(ymat-xmat*bmat))/pmn;
   negl=(pmn/2.0)*log(varpm)+(1/(2.0*varpm))*sum( rowsum(square(ymat-xmat*bmat)) );
   return negl;

FUNCTION dvariable negloglike_logmusig2F(dvariable mu, dvariable sig2, double& n, dvector& yv)
   dvar_vector negl(1,n); //each negative log likelihood
   dvariable negL;  //sum of the respecive log likelihoods
   for(int i=1; i<=n; i++)
     negl(i)=(1.0/2.0)*log(sig2)+(1.0/(2.0*sig2))*square(log(yv(i))-mu);
   negL=sum(negl);
   return negL;

FUNCTION dvariable negloglike_gearage(double& n, dvar_vector G)
   dvar_vector negl(1,n);  //each negative log likelihood
   dvariable negL;  //sumof the respective neg. log likelihood
   dvar_vector gearagev(1,4); //vector of (g1,g2,g3,1)
   gearagev(1)=G(1);
   gearagev(2)=G(2);
   gearagev(3)=G(3);
   gearagev(4)=1; //assuming full selectivity for age 2.3
   dvar_vector subnegl(1,4);

   for(int i=1; i<=n; i++)  {
     for(int j=1; j<=4; j++)   {
       subnegl(j)=agefmat1(i,j)*log(agefmat2(i,j)*gearagev(j)/(agefmat2(i)*gearagev));
            //the denominator is vector*vector
     }
     negl(i)=-1.0*sum(subnegl);
   }
```

```
  negL=sum(negl);
  return negL;

FUNCTION Normal_prior
  Normal_value=  elem_div( square(pops - mean_prior),(2*square(elem_prod(cv_prior,mean_prior))));

FUNCTION pm_multinomial //returns vector form
 dvar_vector gearagev(1,4); //vector of (g1,g2,g3,1)
 gearagev(1)=gage(1);
 gearagev(2)=gage(2);
 gearagev(3)=gage(3);
 gearagev(4)=1;  //assuming full selectivity for age 2.3

 for(i=1;i<=4;i++)
    pmPredprop(i)=gearagev(i)*(pops(1,i)+pops(2,i)+pops(3,i)+pops(4,i)+pops(5,i));
 pmPredprop /= sum(pmPredprop);
 pm_multinom =  -1.0*ssize_MA*elem_prod(pmUa, log(pmPredprop));

FUNCTION multinomial //returns matrix form
 for(i=1;i<=5;i++)
   Predprop(i)=pops(i)/sum(pops(i));
 multinom = -1.0*elem_prod(ssizemat_MB, elem_prod(Uda,log(Predprop) ) );

REPORT_SECTION
 report<<"Age frequencies of Port Moller & Districts"<<endl;
 report<<"observed"<<endl;
 report<<pmUa<<endl;  //observed PM age freq
 report<<Uda<<endl;  //observed district age freq
 report<<"predicted"<<endl;
 report<<pmPredprop<<endl;
 report<<Predprop<<endl;
 report<<"PM regression from PM CPUE index"<<endl;
 report<<"mean of regression model"<<endl;
 report<<" "<<(betaPM(1)+betaPM(2)*pmx)*1000<<endl;
 report<<"predicted total run"<<endl;
 report<<" "<<totr<<endl;
 report<<"mean values of district-run distribution"<<endl;
 report<<" "<<mean(Ud(1)/KNcump)<<" "<<mean(Ud(2)/Ecump)<<" "<<mean(Ud(3)/Ucump);
 report<<" "<<mean(Ud(4)/Ncump)<<" "<<mean(Ud(5)/Tcump)<<endl;
 report<<"predicted district-runs"<<endl;
 report<<" ";
 for(i=1;i<=5;i++)
   report<<sum(pops(i))<<" ";
 report<<endl;
 report<<"beta0, beta1 and var in PM regression model"<<endl;
 report<<" "<<betaPM(1)<<" "<<betaPM(2)<<" "<<varpm<<endl;
 report<<"lognormal_mu in district-run model"<<endl;
 report<<lognmu<<endl;
 report<<"lognormal_sigt in district-run model"<<endl;
 report<<lognsig2<<endl;
 report<<"Age-specific selectivity of PM gillnet"<<endl;
 report<<gage<<endl;
 report<<"negative log of PM multinomial: fpmage"<<endl;
 report<<" "<<fpmage<<" "<<whichLike(1)<<endl;
```

```
report<<"negative log of Inshore multinomial: fage"<<endl;
report<<" "<<fage<<" "<<whichLike(2)<<endl;
report<<"negative log of PM regression: fpm"<<endl;
report<<" "<<fpm<<" "<<whichLike(3)<<endl;
report<<"negative log of Lognormal: flognormal"<<endl;
report<<" "<<flognormal<<" "<<whichLike(4)<<endl;
report<<"Total objective: f"<<endl;
report<<" "<<f<<endl;
report<<"Prior (Normal: 1 or Uniform: 0): "<<prior_type<<endl;
report<<"Adjusted day code: "<<daycode<<endl;
report<<"duga and dtog: "<<duga<<" "<<dtog<<endl;
```

DATA FILE

I display an example of a DAT file here. '#' is followed by comments. Because the historical data are so large, I omit the middle parts of the data from the example.

```
#Which day?
#adjusted day code(change)  #TODAY CODE 20
#20  #add 2 because of being earlier by 1.6 than the past yrs (I sense it from PM data)
22

#Which likelihood (PM_Multinomial/ inshore_Multinomial/PM_Regression/ inshore_lognormal)
1 1 1 1

#pmx: Mean of cumulative Roger's weighted CPUE from PM fishery (change)
#Unit of this value: 6000*catch/(fishing gear length (fm) * mean fishing time (minutes))
1821.905

#For the lognormal(integrated across ages) Cummulative runsize to a particular day
#Ud: unit of these values: (1000's) (change)
3206 2079 124 2439 29

#Data for Multinomial A (PM)
#pmUa: cumulative age-specific catches by PM fishery (change)
74      2960    164     385

# sample size of multinomial A
100

#Data for Multinomial B (Samples from fisheries of dif districts)
# Uda: Cumulative runs by district and age (change)
# Real sample size
30      3347    46      221
29      2156    333     1337
25      419     26      35
14      1954    1       20
5       253     0.1     3   #crash when 5 253 0 3
#
# sample size of multinomial B
20 20 20 20 20
```

```
# PRIORS from preseason forecasts
# Switch for prior type (1: normal/ 0: uniform)
1

# Prior Means (thousands)
2500 1600 1500 1200
600 2200 1200 1300
400 200 300 500
2500 4800 200 300
100 300 40 40

# Prior CV
0.6700   1.4608   3.6550   1.0454
0.5242   0.6724   1.6920   2.3267
0.8173   6.8838   0.7908   1.2373
0.2384   0.1626   0.2845   0.3848
2.2128   0.7131   3.9117   1.4126

#betadim: dimension of betas
2
#rpmindD: Rows of Rogers PM index data
660
#cpmindD: Columns of Rogers PM index data
6
#pmn: rows of data matrix, Xmat #(1998-1984-1)
15

#drows: Rows in the five district-specific data sets
1714     1750     1664     1687     1385
#dcols: The respective five district-specific data sets have 6 columns
6

#gagen: rows of age frequency data
14

#Age ratioes of Port Moller catch
#pmyr   a1.2    a1.3    a2.2    a2.3
1987    0.462   0.185   0.182   0.171
1988    0.190   0.505   0.200   0.105
1989    0.106   0.203   0.503   0.187
1990    0.108   0.234   0.410   0.249
1991    0.140   0.517   0.158   0.185
1992    0.079   0.349   0.322   0.250
1993    0.063   0.200   0.266   0.471
1994    0.068   0.202   0.398   0.332
1995    0.138   0.157   0.506   0.199
1996    0.075   0.520   0.131   0.275
1997    0.119   0.335   0.279   0.267
1998    0.176   0.381   0.090   0.352
1999    0.447   0.211   0.260   0.082
2000    0.147   0.633   0.078   0.142

#Age ratioes of district run
```

#distyr	a1.2	a1.3	a2.2	a2.3
1987	0.496	0.232	0.119	0.153
1988	0.208	0.430	0.229	0.133
1989	0.103	0.161	0.641	0.095
1990	0.141	0.215	0.432	0.213
1991	0.194	0.479	0.209	0.119
1992	0.133	0.278	0.356	0.234
1993	0.128	0.189	0.341	0.342
1994	0.084	0.147	0.585	0.184
1995	0.161	0.118	0.572	0.148
1996	0.107	0.520	0.128	0.245
1997	0.203	0.262	0.346	0.189
1998	0.344	0.294	0.129	0.233
1999	0.513	0.212	0.211	0.064
2000	0.202	0.622	0.080	0.095

#pmmat:Rogers Port Moller index #660 rows
#yr daycd dayindex one cumudayindex actrun

1985	2	7.04	1	7.04	36.5
1985	3	3.44	1	10.48	36.5
1985	4	8.88	1	19.36	36.5
1985	5	13.92	1	33.28	36.5
1985	6	31.84	1	65.12	36.5
1985	7	42.24	1	107.36	36.5
1985	8	102.32	1	209.68	36.5
1985	9	31.84	1	241.52	36.5

..............(omitting) ....

2000	25	25.23	1	604.12	27.8
2000	26	19.11	1	623.23	27.8
2000	27	28.47	1	651.70	27.8
2000	28	13.02	1	664.72	27.8
2000	29	5.70	1	670.42	27.8
2000	30	0.00	1	670.42	27.8
2000	31	0.00	1	670.42	27.8
2000	32	0.00	1	670.42	27.8
2000	33	0.00	1	670.42	27.8
2000	34	0.00	1	670.42	27.8
2000	35	0.00	1	670.42	27.8
2000	36	0.00	1	670.42	27.8
2000	37	0.00	1	670.42	27.8
2000	38	0.00	1	670.42	27.8
2000	39	0.00	1	670.42	27.8
2000	40	0.00	1	670.42	27.8
2000	41	0.00	1	670.42	27.8
2000	42	0.00	1	670.42	27.8
2000	43	0.00	1	670.42	27.8
2000	44	0.00	1	670.42	27.8
2000	45	0.00	1	670.42	27.8

#distcd	mo	day	yr	daycd	cumrunpro
1	6	22	1955	13	0.0003
1	6	23	1955	14	0.0003

1	6	24	1955	15	0.0003
1	6	25	1955	16	0.0084
1	6	26	1955	17	0.009
1	6	27	1955	18	0.0685
1	6	28	1955	19	0.0692
1	6	29	1955	20	0.0718

.... (omitting)...............

5	8	7	2000	59	0.9946
5	8	8	2000	60	0.996
5	8	9	2000	61	0.9977
5	8	10	2000	62	0.9985
5	8	11	2000	63	0.9985
5	8	12	2000	64	0.9985
5	8	13	2000	65	0.9985
5	8	14	2000	66	0.9989
5	8	15	2000	67	1

VITA

Saang-Yoon Hyun

2002

Education

Ph.D. (Fall quarter 1997 - Spring quarter 2002) in Quantitative Ecology and Resource Management (QERM) at the University of Washington (Seattle, WA).

M.S. (Fall quarter 1993 - Fall quarter 1996) in Fisheries at the University of Washington (Seattle, WA).

B.F. (March 1986 - February 1993; on-leave from 1988 to 1990 for military service) in Aquaculture and fisheries management at the Cheju National University (Cheju, Korea).

High school diploma (February 1986) Ohyun High School (Cheju, Korea).

Research fields of interest

Quantitative fisheries management, biostatistics.

TA work experience

Fall 2000 - Spring 2002: Teaching assistant for QERM and Quantitative Science classes at the UW.

RA work experience

Summer 2001: Research assistant for salmon habitat project (PI: Dr. Ashley Steel) at the UW.

1997 - Spring 2000: Research assistant for Highseas project (PI: Dr. Katherine Myers) at the UW.

1994 - 1996: Research assistant for Columbia River Salmon Passage Model (CRiSP) project (PI: Dr. James Anderson) at the UW.

Other work experience

1988 - 1990: Military service (mandatory in Korea).

Publications

Hyun, S. 2002. Inseason forecasts of sockeye salmon returns to the Bristol Bay districts of Alaska. Ph.D. dissertation, University of Washington, Seattle, Wash.

Norris, J.G., **S. Hyun**, and J.J. Anderson. 2000. Ocean distribution of Columbia River Upriver Bright Fall Chinook salmon stock. N. Pac. Anadr. Fish Comm. Bull. No. 2: 221-232.

Myers, K.W, R.V. Walker, N.D. Davis, K.Y. Aydin, **S. Hyun**, R.W. Hilborn, and R.L. Burgner. 2000. Migrations and abundance of salmonids in the North Pacific, 2000. Annual report to ABL, NMFS, High Seas Salmon Research Project. Fish Res. Inst. Univ. Wash., Seattle (SAFS-UW-0009). 93 p.

Myers, K.W, R.V. Walker, N.D. Davis, K.Y. Aydin, **S. Hyun**, R.W. Hilborn, and R.L. Burgner. 1999. Migrations, abundance, and origins of salmonids in offshore waters of the North Pacific, 1999. Annual report to ABL, NMFS, High Seas Salmon Research Project. Fish Res. Inst. Univ. Wash., Seattle (FRI-UW-9909). 125 p.

Hyun, S., K.W. Myers, and J.G. Sneva. 1998. The time of annulus formation in chinook salmon caught in Washington coastal waters. (NPAFC Doc. 352.) FRI-UW-9808. Fisheries Research Institute, University of Washington, Seattle, WA.

Myers, K.W, R.V. Walker, N.D. Davis, K.Y. Aydin, **S. Hyun**, R.W. Hilborn, and R.L. Burgner. 1998. Migrations, abundance, and origins of salmonids in offshore waters of the North Pacific, 1998. Annual report to ABL, NMFS, High Seas Salmon Research Project. Fish Res. Inst. Univ. Wash., Seattle (FRI-UW-9810). 72 p.

Hyun, S. 1996. Ocean distributions of the Columbia River Hanford Reach and Snake River fall chinook salmon (*Oncorhynchus tshawytscha*) stocks in the effect of interannual ocean conditions on their survival. M.S. thesis, University of Washington, Seattle, Wash.

Scholarships and awards

Awarded 1993 -1994 Rotary Foundation Ambassadorial Scholarship for an M.S. in Fisheries at the UW.

Awarded the prize of Dean at College of Oceanic Sciences in Cheju National University (CNU) for an honor graduate (1993).

Hobbies

Traveling, hiking, biking, swimming, skiing.