AN ABSTRACT OF THE THESIS OF

Maarika Teose for the degree of Honors Baccalaureate of Science in Mathematics presented on May 30, 2008. Title: The Effect of Temperature on the Survival of Chinook Eggs and Fry.

Abstract Approved: _____

Edward Waymire

_____

Vrushali Bokil

ABSTRACT: Temperature data from above and below the Cougar Dam collected by the U.S. Geological Survey prior to the construction of the temperature control structure was analyzed to determine how the differing temperature regimes affect the growth and survival of threatened spring-run Chinook salmon. An ARIMA time-series model was used to approximate each temperature record and a MATLAB program was written to simulate the growth of an individual Chinook salmon from egg to fry. Given a spawning distribution, the model was used to "grow" several hundred fish in each temperature regime for a given period of time, and the distributions of fish weights were recorded in histograms. The distributions of fish weights above and below the dam were significantly different ($p < 0.001$) and the temperature regime above the dam produced larger fish than the temperature regime below the dam ($p < 0.001$). We conclude that the temperature regime below the dam (prior to the construction of the temperature control structure) is less conducive to the survival of Chinook than the temperature regime above.

Key Words: Chinook salmon growth, Cougar Dam, simulation model, temperature.

E-mail: maarika.teose@gmail.com

The Effect of Temperature on the Survival of Chinook Eggs and Fry

by

Maarika Teose

A PROJECT

submitted to

Oregon State University

University Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Mathematics (Honors Scholar)

Presented May 30, 2008

Commencement June 2008

<u>Honors Baccalaureate of Science in Mathematics</u> project of <u>Maarika Teose</u> presented on <u>May 30, 2008.</u>

APPROVED:

_____

Co-Mentor, representing Mathematics

_____

Co-Mentor, representing Mathematics

_____

Thesis Advisor, representing Fisheries and Wildlife

_____

Dean, University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, University Honors College. My signature below authorizes release of my project to any reader upon request.

_____

Maarika Teose, Author

Acknowledgements

I would like to extend my heartfelt gratitude to all three of my thesis advisors: Dr. Edward Waymire, Dr. Vrushali Bokil, and Dr. Jason Dunham. Rarely does an undergraduate get the opportunity to work so closely with such wonderful faculty for such a long time.

I am grateful to Uran Chu for his generous assistance with the statistical analysis in this project, and to Jorge Ramirez for helping me get started on this project nearly a year ago.

I would like to thank my parents Erik and Liina Teose for making all of this possible.

Finally, I want to thank all the lovely people who kept me happy and sane during the final whirlwind weeks.

Contents

List of Figures

List of Tables

Preface

In the Pacific Northwest of the United States, several runs of wild salmon have been listed under the Endangered Species Act (ESA) as threatened or endangered during the past decade. Dams constitute major obstacles for migrating fish and adversely affect their survival rates. In extreme cases, dams not only affect the structure of a river, but also its water chemistry and temperature regime.

One such dam is the Cougar Dam on the South Fork of the McKenzie River in Oregon. Since its construction by the Army Corps of Engineers in 1963 there has been no fish passage above the dam, and consequently the wild anadromous fish populations above the dam have become extinct. In 1999, spring-run Chinook salmon (*Oncorhynchus tshawytscha*) below the Cougar Dam were ESA-listed, and it was believed that the dam was directly responsible for this. The reservoir outlet was at the base of the reservoir, meaning that the water being released downstream was unnaturally cold during the summer and unnaturally warm during the winter. By "unnaturally cold" and "unnaturally warm," we mean that the water temperatures below the dam differed from the temperature of the water flowing into the reservoir, which we call the "natural" temperature regime of the river in the area. In an effort to revive populations of wild Chinook salmon, the Army Corps of Engineers installed a \$52 million water temperature control structure in the reservoir which became operational in 2004. This structure makes it possible to draw water from different levels in the water column, enabling the engineers to control the water temperature below the dam and allowing them to implement a water temperature regime which mimics the natural temperature regime. It is hoped that the more natural temperature regime will increase the rate of survival of Chinook in the South Fork of the McKenzie.

The Effect of Temperature on the Survival of Chinook Eggs and Fry

1 Introduction

The early life histories of Chinook salmon are heavily temperature-dependent. Spring-run Chinook in the South Fork of the McKenzie River begin their upstream migration when river temperatures have warmed suitably and river flow is high enough to permit access to upper reaches of streams. McKenzie River salmon spawn in the fall, approximately between the 15th of August and the 15th of October, with a spawning peak in September [16]. The fertilized eggs then incubate in redds - nests that the females dig into the gravel river bed - until late February or early March [16]. Generally there is high mortality in the egg stage as a result of predation, and also from competition among females for the best nesting sites.

Once a Chinook egg hatches, the larval fish (known as an alevin) lives in the gravel for a few weeks, subsisting entirely on its yolk-sac. When the yolk-sac is absorbed, the young fish "swims-up" from the gravel and begins feeding. These fish are known as fry and they grow during the spring and summer months, after which - depending on genetic predisposition - they either immediately migrate downstream to the open ocean or remain in freshwater over the winter and migrate to the ocean the following year.

The effect of water temperature on the growth of Pacific salmon eggs and juveniles has been studied extensively ([13], [3], [4], [14], [1], [18], [20], [10]). Decades of laboratory and hatchery research have yielded substantial information about the conditions for optimum fish growth. However, this information is largely based on experiments in which temperatures are held constant over all or part of the fish's lifetime. Relationships between temperature and growth are then extrapolated from this data. It is unknown if these equations have been tested in the field in order to verify their accuracy in predicting fish weight in a variable temperature regime.

1.1 The "Egg-Fry Conflict"

In general, the eggs of salmon incubate better in colder water, whereas fry fare better in warmer water. For instance, normal embryonic growth for Chinook salmon lies between $4.5°$ and $14.2°C$ (average $9.35°C$), and optimum fry growth occurs between $15.6°$ and $19°C$ (average $17.3°C$) [20]. Additionally, it has been observed that eggs incubated in colder waters have lower mortality rates than fish that have incubated at warmer temperatures, even though they often hatch slower and smaller than their warm-water incubated counterparts [20]. This temperature trade-off has been termed the "egg-fry conflict" [17].

Thus it is the author's intention to construct a model of the egg-fry system and to use this model to analyze the ways in which differing water temperature regimes may affect the survival and fitness of a single generation of salmon. In the process, it is hoped that the "egg-fry conflict" will come into clearer focus.

1.2 Summer Model

In the summer of 2007 the author had the opportunity to begin her work on this problem at the H.J. Andrews Experimental Forest in Blue River, Oregon as a part of the EcoInformatics Summer Institute (EISI). Over the course of ten weeks a rough initial model was built, which which we will call the "summer model" [19].

One measure of a generation's fitness is its biomass. In general, it is assumed that a large number of big fish will produce more viable offspring than a small number of small fish. Biomass can be estimated by multiplying the mean weight of the population by the number of individuals in the population. The summer model focuses heavily on finding an expression for the expected value of the weight of the fry at some time t_m after spawning.

The summer model makes several assumptions about the egg-fry system. Firstly, it assumes that all spawning in the reach occurs on the same day every year, the 1st of September. Additionally, water temperature is assumed to be constant during September and October of each year, since one of the inputs of the summer model is the (constant) incubation temperature for an egg. Furthermore, water temperature is also assumed to be

constant during in-stream fry growth, which, for the salmon in this system, occurs over the course of a full calendar year. These constant temperatures are obtained by averaging mean monthly temperatures.

In the summer model, time of hatching is assumed to be a random variable denoted by T_h that depends on incubation temperature[1], T, and has a probability density function $f_{Th}(t, T)$. Equations in the literature ([7], [13], [1]) give the time (in days) to fifty percent hatching - in other words, the median hatching time - as a function of temperature, denoted $D_2(T)$. The form of these functions is $D_2(T) = \gamma(T + \beta)^c$, where T is temperature in degrees Celsius, and γ, β and c are constants. In the summer model, the distribution of hatching times is assumed to be triangular, peaking at $D_2(T)$.

Additional assumptions in the summer model are that fry and alevin have the same growth rates, i.e. once an egg hatches it is essentially a fry. Furthermore, growth is assumed to be density independent.

Under these assumptions, the author formulated an expression for the expected value and variance of fish weight at some time t_m after spawning. Using USGS water temperature data from two gauges - one situated above the reservoir (14159200) and the other below the Cougar Dam (14159500) - the model was implemented for each temperature regime and expected values of the fish weights were calculated and compared. What the author had hoped for was that the model would predict a higher biomass for the "natural" tempera-ture regime, and a smaller biomass of fish for the "unnatural" temperature regime below the dam; the spring-run Chinook salmon in the McKenzie River have, after all, evolved to grow optimally in the historically natural temperature regime of the river. A higher biomass from the "natural" temperature regime would have lent support to the idea that the water temperature control structure installed at the Cougar Dam will help to increase the probability of survival of Chinook salmon.

Unfortunately, the summer model did not give the desired results. According to the summer model, the "unnatural" water temperature regime below the dam will produce a

---

[1]Please note that although temperature is denoted by T in the summer model, we shall be using the symbol θ further on.

higher biomass of fish than the "natural" temperature regime above the reservoir! This is certainly not what has been observed in reality, and so it is obvious that the summer model's assumptions are too unrealistic, since the model failed to reflect even the qualitative aspects of the system.

2　Objectives

One of the primary reasons for the failure of the summer model may be that temperature is defined too coarsely (constant incubation temperature, constant fry growth temperature). Additionally, the underlying physical processes - spawning, incubation, hatching, and temperature-dependent growth - are not all represented.

The overall objective of this thesis is to improve upon the summer model and obtain a model which more closely represents reality, at least qualitatively. Since we believe temperature to be the major culprit in the downfall of the summer model, our first objective is to better understand the temperature data we are working with. This will mean performing a time series analysis of temperature data gathered from the USGS water temperature gauges located above and below the dam, and fitting an auto-regressive moving average (ARMA) model to each set of data.

Additionally, we will use MATLAB to build an individual-based simulation model that is capable of using a record of daily water temperatures to essentially "grow" a fish. The output will be the weight of a fish at some time t_m after hatching. This model will incorporate several things that the summer model did not (e.g. spawning, alevin life stage, daily temperature inputs).

The final step will be to run the MATLAB script several times, "growing" several hundred fish in each temperature regime. We will look for differences in the resulting histograms of fish weights.

We will also briefly analyze the supposed "egg-fry conflict" and attempt to characterize it mathematically.

3   Methods

3.1   Empirical versus Mechanistic Models

Empirical (predictive) models are based on statistics and are good for predicting values of response variables within the range of data obtained, but not very good for predicting values outside of the data range, or into future, because these models only look at the ways in which response variables change with respect to (supposed) explanatory variables, but make no attempt to explain the underlying reasons for the observed relation. For example, a linear regression is a type of empirical model for predicting the value of a response variable given the value of the independent variable.

Mechanistic (explanatory) models, on the other hand, are based entirely on the underlying physical processes of the system, relating response and explanatory variables with equations that come from the observed interactions between them. A classic example of a mechanistic model is Newton's Law of Cooling. These models are usually much more difficult to construct. The pay-off, however, is that the modelers can feel confident about using their model to predict values of response variables outside of the data range and into the future [5].

Empirical models predicting egg survival, incubation time and juvenile growth based on water temperature do exist ([7], [8], [4], [1]), however, to date there has been relatively little effort to relate these equations to one another in a meaningful way. More recently, mechanistic models relating temperature and salmon growth have been appearing, for example [11], [5]. Predictably, they are extremely complicated and highly parameterized, requiring numerical methods for their implementation. Both types of models are useful, and we use both types in our simulation.

3.2 Introduction to Autoregressive Integrated Moving Average (ARIMA) Models

We would like to describe the behavior of the temperature regimes above and below the Cougar Dam. We do this by performing a time-series analysis of the temperature records and producing an autoregressive integrated moving average (ARIMA) model for each data set.

The idea behind univariate time-series models is that given some series of observations $\Theta = \{\theta_t \mid t = 0, 1, \ldots\}$ the tth element in the series can be described as a function of previous elements in the time series. For example, we can say that today's water temperature depends (to some extent) on yesterday's water temperature, and perhaps also on the temperature two days ago, or three, and so on. This is known as an autoregressive series, and is said to be of order p if each element of the series at most depends on the p previous elements in a sense made more precise below.

At each time step, a small amount of stochasticity is included in the time series by adding a white-noise disturbance term in the form of a random number drawn from a normal distribution with mean zero and variance σ^2. If, say, the tth element of our time series depends not only on the white-noise disturbance term at time t, but also on the disturbance term at time $t - 1$, and $t - 2$, all the way back to time $t - q$, then we say that our time series is a moving average process of order q.

In the event that our time series exhibits certain undesirable behaviors, such as nonstationarity (we will discuss this phenomenon below), then it is often possible to transform the original series into a more well-behaved series by removing, or "differencing out" trends. For example, if our time series had a first-order (linear) trend and we wished to remove this trend, we let

$$\theta_t^1 \equiv \Delta\theta_t = \theta_t - \theta_{t-1} \tag{1}$$

and work with the transformed series $\Theta^1 = \{\theta_t^1 \mid t = 1, 2 \ldots\}$. If the trend is quadratic, then we let $\theta_t^2 \equiv \Delta\theta_t^1$, and so on. The order of the trend in a time series is denoted d, and

such a process is called an integrated series. Thus the shorthand for an p-order autoregressive, d-order integrated, and q-order moving average time series is ARIMA(p,d,q). If we assume that our time series is not integrated, or if we are working with a transformed series in which the trend has already been differenced out, then we may ignore the "integrated" component of the ARIMA model and our model reduces to an ARMA(p,q) model. The functional form of the ARMA(p,q) model is

$$\theta_t = m + \alpha_1 \theta_{t-1} + \alpha_2 \theta_{t-2} + \cdots + \alpha_p \theta_{t-p} + (\epsilon_t - \beta_1 \epsilon_{t-1} - \cdots - \beta_q \epsilon_{t-q}) \qquad (2)$$

where m is the mean of the series, α_i and β_i are constants, and ϵ_i, $i = t, t-1, \ldots, t-q$ is a white noise process that is a sequence of independently and identically distributed normal variables with mean zero and variance σ^2 (denoted iid N(0,σ)).

It is important to note that the ARMA(p,q) model given in Equation 2 can only be applied to our time series if we assume that our process satisfies stationarity. In a stationary time series the mean and variance of the sequence are independent of time. For example, if we imagine our time series being plotted as a movie running forward in time, then stationarity implies that our time series will hover around a single mean rather than slowly increase or decrease over time; additionally, the variance around this single mean will remain constant for all time, i.e. the path of the time series will not tend to get further and further (or closer and closer) to the mean. This is called first moment stationarity. More generally, if we define the autocovariance coefficients of our series to be

$$\rho(k) = \mathbb{E}(\theta_t \theta_{t-k}), \quad k = 0, 1, 2, \ldots$$

then second moment stationarity means that $\rho(k)$ will remain the same regardless of the time t. As an example, if we have a time series of length 10, then

$$\rho(5) = \mathbb{E}(\theta_{10} \theta_5) = \mathbb{E}(\theta_6 \theta_1).$$

We note that $\rho(0) = \mathbb{E}(\theta_t^2)$ gives us the variance, and stationarity assumes that this remains constant for all t.

The concepts underlying the construction of an ARMA(p,q) model can be understood by restricting ourselves to the ARMA(1,0) model, better known as the AR(1) model, which contains only two parameters. Although the algebra involved is much simpler than for the general ARMA(p,q) case, the ideas can easily be extended to cover the entire class of models (for a more thorough introduction to modeling ARMA(p,q) processes see [12]).

We can express the white noise terms in Equation 2 as $\epsilon_i = \sigma z_i$, $i = t, t-1, \ldots, t-q$, and thus the AR(1) model with zero mean has the form

$$\theta_t = \alpha \theta_{t-1} + \sigma z_t \tag{3}$$

where z_t is iid N(0,1). The two parameters we must estimate are α and σ. We do this by squaring both sides of Equation 3 and taking expectations, which yields

$$\rho(0) = \alpha^2 \mathbb{E}(\theta_{t-1}^2) + \mathbb{E}(\sigma^2 z_t^2) + 2\alpha \mathbb{E}(\theta_{t-1} \sigma z_t) \tag{4}$$

The AR(1) model requires that θ_{t-1} and z_t are independent of each other, and thus the last term in Equation 4 can be written as

$$2\alpha \mathbb{E}(\theta_{t-1})\mathbb{E}(\sigma z_t) = 2\alpha\sigma \mathbb{E}(\theta_{t-1})\mathbb{E}(z_t).$$

We recall from above that since z_t is iid N(0,1), then $\mathbb{E}(z_t) = 0$ and thus the last term in Equation 4 vanishes, leaving

$$\rho(0) = \alpha^2 \rho(0) + \sigma^2 \tag{5}$$

$$= \frac{\sigma^2}{1 - \alpha^2} \tag{6}$$

We also calculate the lag-1 covariance:

$$\rho(1) \quad = \quad \mathbb{E}(\theta_t \theta_{t-1}) \tag{7}$$

$$= \quad \mathbb{E}\left[(\alpha\theta_{t-1} + \sigma z_t)\theta_{t-1}\right] \tag{8}$$

$$= \quad \alpha\mathbb{E}(\theta_{t-1}^2) + 0 \tag{9}$$

$$\rho(1) \quad = \quad \alpha\rho(0) \tag{10}$$

Thus we find

$$\alpha \quad = \quad \frac{\rho(1)}{\rho(0)} \tag{11}$$

$$\sigma^2 \quad = \quad \rho(0)\left[1 - \left(\frac{\rho(1)}{\rho(0)}\right)^2\right] \tag{12}$$

The values of the parameters $\rho(0), \rho(1)$ are estimated from the data using the Yule-Walker Estimates [12],

$$\rho(0) \approx \widehat{\rho(0)} \quad = \quad \frac{1}{n}\sum_{j=1}^{n}\theta_j^2 \tag{13}$$

$$\rho(1) \approx \widehat{\rho(1)} \quad = \quad \frac{1}{n}\sum_{j=2}^{n}\theta_j\theta_{j-1} \tag{14}$$

and from these it is possible to calculate approximate values for α and σ.

3.3   Construction of the ARMA Model

The computation of the ARMA specifications and parameter values is best done by statistical software; we use SAS 9.1 to compute our time-series model. The three steps in ARMA modeling are as follows [12]:

1. Check for stationarity. If the series is nonstationary, induce stationarity by transforming the series or differencing out trends.

2. Using the autocorrelation properties as guides, choose a few ARMA models for esti-

mation and testing, arriving at a preferred model with white noise residuals.

3. Use the preferred model to produce forecasts.

3.3.1 Below the Dam

We wish to fit an ARMA model to temperatures below the Cougar Dam (gauge 14159500). We begin by making sure we have the correct data to work with. In early April of 2002 the Cougar Reservoir was drawn down in order to begin construction on the temperature control structure. Since we are interested in comparing the "unnatural" temperature regime produced by the dam to the "natural" regime above the dam, we do not want to contaminate our data with temperature records during the construction or operation of the temperature control structure. Thus we obtain data from USGS for the daily mean temperatures, averaged from 1978 to 2001.

We first check the data for evidence of non-stationarity. In practice, this is done by using various statistical tests (e.g. augmented Dickey-Fuller and Phillips-Perron) to check for a unit root in the autoregressive part of the model [12]. This property of having a unit root ("unit" meaning "equal to 1") is undesirable because it induces time-dependence. Consider again the AR(1) model in Equation 3. If we disregard the σz_t we can treat Equation 3 as a recurrence relation of the form

$$\theta_t = \alpha\theta_{t-1} \tag{15}$$

To find the solution to this recurrence – that is, if we wished to remove dependency on previous elements in the series [9] – we use the ansatz $\theta_t = r^t$ and obtain

$$r^t = \alpha r^{t-1} \tag{16}$$

By dividing through by r^{t-1} we find that $r = \alpha$. This is the characteristic equation of our recurrence relation, and we observe that it is linear. In our case, a unit root in the characteristic equation implies that $\alpha = 1$. When we substitute this back into Equation 3,

we obtain

$$\theta_t = \theta_{t-1} + \sigma z_t.$$

Expanding the term θ_{t-1} as a telescoping series, we can rewrite the above as

$$\theta_t = \theta_0 + \sigma \sum_{j=1}^{t} z_j \tag{17}$$

Recall that each z_j is independent and is drawn from N(0,1), and thus if we square both sides of Equation 17 and take expectations, we find that

$$Var(\theta_t) = \sigma^2 t + Var(\theta_0) \tag{18}$$

We observe that the variance is not independent of time, and thus a series with a unit root does not satisfy stationarity.

Our initial exploration indicates that we do not have sufficient evidence to reject the null hypothesis that we have a unit root (all test statistics are greater than 0.05), so we take the first difference of our data in an attempt to transform it into a stationary time series. By "differencing" we mean the process described above in Equation 1. Using the same unit root tests as above, we find that the first-differenced data exhibits stationarity.

We now perform the ARMA procedure in SAS to estimate possible ARMA(p,q) models for our first-differenced data using a few approximation techniques. We decide that the ARMA(1,2) model is sufficient because the ARMA procedure statistics indicate that it is a good fit. We note that this is in fact an ARIMA(1,1,2) model, since we are fitting an ARMA(1,2) model to the first-differenced data. SAS gives us the values of the autoregressive and moving average coefficients for this model, and thus for the temperature data below the dam we have

$$w_t = \mu(1 + \alpha_1) - \alpha_1 w_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} \tag{19}$$

where

- w_i is the difference between the temperature on day i and day $i - 1$.

- $\mu = -0.02277$ is the mean of the differenced series.

- $\alpha_1 = -0.98824$ is the first autoregressive coefficient.

- ϵ_i is the error term in the ith time step.

- $\beta_1 = -1.14192$ is the first moving average coefficient.

- $\beta_2 = 0.30612$ is the second moving average coefficient.

The set of values w_t given by Equation 19 are the modeled first-differences of the temperature regime. If we wish to compare our model to the observed data, we must now do the following. Let us define $\bar{\Theta}$ to be the overall (grand) mean of the temperature data, and let

$$\nu_t \equiv \theta_t - \bar{\Theta}.$$

Then the value w_t in Equation 19 can be represented by $w_t = \nu_t - \nu_{t-1}, \quad t = 1, 2, \ldots$ We now observe that we can write the term ν_t as a telescoping series, that is,

$$\nu_t = \sum_{j=1}^{t} (\nu_j - \nu_{j-1}) + \nu_0 \tag{20}$$

Using this yields

$$
\begin{aligned}
\theta_t &= \nu_t + \bar{\Theta} & (21)\\
&= \sum_{j=1}^{t} (\nu_j - \nu_{j-1}) + \nu_0 + \bar{\Theta} & (22)\\
&= \sum_{j=1}^{t} w_j + \nu_0 + \bar{\Theta} & (23)\\
&= \sum_{j=1}^{t} w_j + \theta_0 & (24)
\end{aligned}
$$

Based on our data in the series $\Theta_t = \{\theta_s : s \leq t\}$, we now compute the value $\hat{w}_t =$

$\mathbb{E}(w_t|\Theta_{t-1})$. The value $\hat{w}_t$ is the predicted or average value of w_t based on the data in Θ. Using this data, we compute the expected value of Equation 19. Recalling that $\mathbb{E}(\epsilon_t) = 0$ we have the deterministic equation for the predicted mean values as

$$\hat{w}_t = \mu(1 + \alpha_1) - \alpha_1 w_{t-1} + \beta_1 \hat{\epsilon_{t-1}} + \beta_2 \hat{\epsilon_{t-2}} \tag{25}$$

for $t = 1, 2, \ldots$. We now perform an in-sample fit of our ARMA model via the following algorithm. We first let $w_0 = \nu_0 = \theta_0 - \bar{\Theta}$ and proceed thus:

$$\hat{w}_1 \quad = \quad \mu(1 + \alpha_1) - \alpha_1 w_0 \tag{26}$$

$$\hat{\epsilon}_1 \quad = \quad w_1 - \hat{w}_1 \tag{27}$$

$$\hat{w}_2 \quad = \quad \mu(1 + \alpha_1) - \alpha_1 w_1 + \beta_1 \hat{\epsilon}_1 \tag{28}$$

$$\hat{\epsilon}_2 \quad = \quad w_2 - \hat{w}_2 \tag{29}$$

$$\hat{w}_3 \quad = \quad \mu(1 + \alpha_1) - \alpha_1 w_2 + \beta_1 \hat{\epsilon}_2 + \beta_2 \hat{\epsilon}_1 \tag{30}$$

For all succeeding estimates we can write

$$\hat{w}_t = \mu(1 + \alpha_1) - \alpha_1 w_{t-1} + \beta_1 \hat{\epsilon_{t-1}} + \beta_2 \hat{\epsilon_{t-2}} \tag{31}$$

Finally we plot the values $\hat{\theta}_t = \sum_{j=1}^{t} \hat{w}_j + \theta_0$ for each t. The result is shown in Figure 1. True temperature data is plotted as the dotted line, and the in-sample ARIMA(1,1,2) fit is shown as the solid line. We see that the fit to the true data is quite good.

3.3.2 Above the Dam

To fit an ARMA model to temperature data from USGS gauge 14159200, which lies above the Cougar Reservoir, we follow a nearly identical procedure as that described in Section 3.3.1. We find that the series is nonstationary and that differencing once transforms the data into a stationary series. Once again we estimate possible ARMA(p,q) models for our first-differenced data and discover that the model with the best fit to the differenced-data
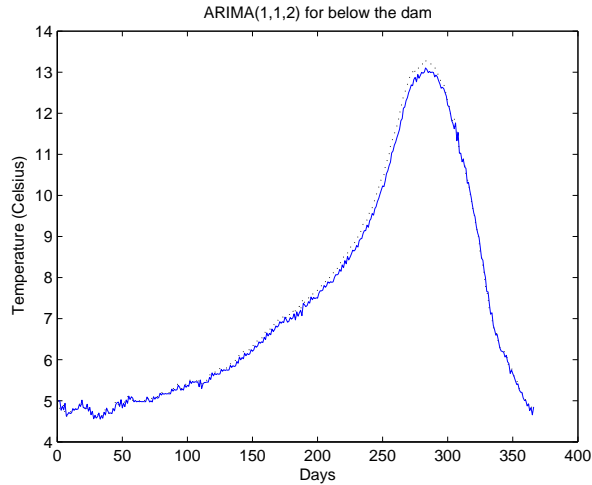
Figure 1: ARIMA(1,1,2) model for temperatures below the dam.

is ARMA(9,7), which gives us an ARIMA(9,1,7) model. We plot the observed temperature data along with our ARIMA(9,1,7) model approximation in Figure 2. The dotted line represents the true temperature data, and the solid line represents the ARIMA(9,1,7) in-sample fit.

Note on the ARMA Procedure

Our goal with constructing the ARMA models was to generate hypothetical temperature records that behave similarly to the true temperature time series, at least in regards to the underlying correlations. Using these simulated temperature regimes, we would be able to use our code (see below; also see Appendix A) to "grow" our fish in slightly different temperature regimes each year, strengthening our conclusions about the distributions of fish weights above and below the dam. Unfortunately, the ARMA procedure is not an appropriate method of generating hypothetical temperature records, that is, forecasting. This is because a crucial element of a yearly temperature record is that the temperature on December 31 of one year must be similar to the temperature on January 1 of the following year; in essence, we require periodic boundary conditions for our deterministic model. Although an in-sample fit of the data approximates the true data well, Figure 3 shows the difficulty in
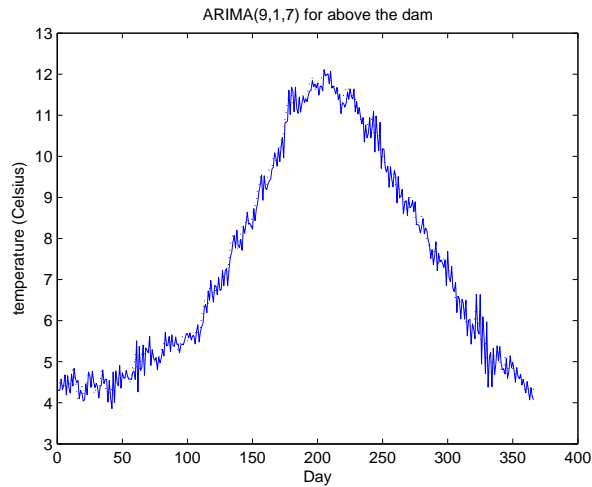
Figure 2: ARIMA(9,1,7) model for temperatures above the dam.

using the ARMA model to forecast temperatures beyond a few days (true temperature data shown as dotted line, simulated temperature regimes shown as solid lines). For comparison, see Figure 4, in which we observe the natural year-to-year variation in daily temperatures below the dam. Although the ARMA model is able to capture the average behavior of our temperature time series, it is not able to describe the fluctuations. Since we are not able to generate realistic hypothetical temperature regimes using our ARMA models, for the remainder of this project we use the raw daily temperature data obtained from USGS.

3.4 Modeling Growth

Simulation models intending to "grow" a fish have been built before, for example see [11], [5]. The author's intention, however, is slightly different than these previous models. There are several variables that contribute to in-stream salmon growth, such as food availability and population density, but for the purposes of this model these will be disregarded. Additionally, habitat variables that affect the early life histories of salmon, such as river flow and stream-bed composition, will be ignored as well. We are purely interested in the effect of temperature on the growth of Chinook eggs and fry.

Furthermore, this study is comparative, focusing on the temperature regimes above and
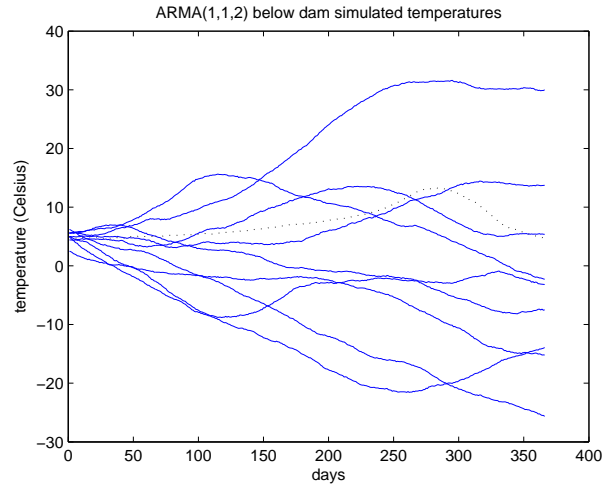
Figure 3: Simulated out-of-sample temperature regimes below the dam using ARIMA(1,1,2) forecasts.

below the Cougar Dam on the South Fork of the McKenzie River. As noted above in Section 2, the only intention is to see how the distributions of fish weights differ depending on which temperature regime is used, and not to see how closely these distributions reflect the true distributions of fish weights observed above and below the dam. This would be impossible, since the wild anadromous fish populations above the dam became extinct after the construction of the dam. Thus, our simulation model will be very specific to this system, and not necessarily generalizable to other rivers or salmon populations.

3.4.1 Growth Equations

The first step, coded in MATLAB as the function

```
[Th,Ta,wt] = fishalevin(Ts,Tm,site),
```

"grows" an individual fish, modeling its development from spawning all the way through to the fry stage (here we include the alevin life stage). The inputs for this function are the time of spawning (Ts), the time at which we wish to know the final fish weight (Tm), and the site number of the USGS gauge whose temperature data we will be using (one of 14159200 or 14159500). The outputs will be the time of hatching (Th), time of alevin swim-up (Ta),
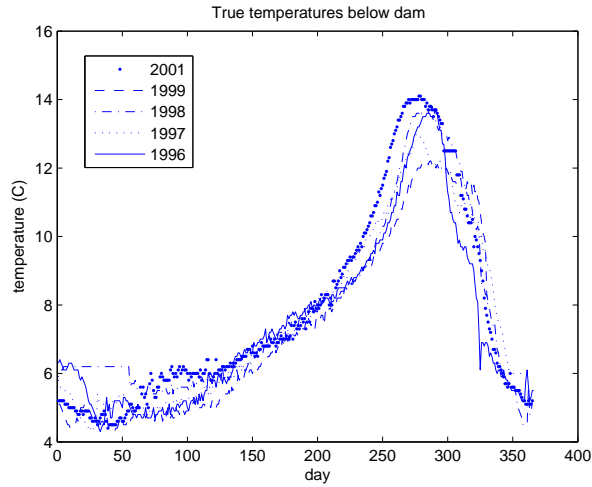
Figure 4: Several years of daily temperature data from below the dam.

and weight at time Tm, denoted wt (see Appendix A for this and other MATLAB codes). The growth of the fish is broken up into three distinct phases: incubation, alevin, and fry. The equations used to calculate growth depend on the life stage of the fish.

As noted above, length of incubation is temperature-dependent. We calculate length of incubation according to the log-inverse Bělehrádek equation given in [1].

$$\ln P = \ln k + b \ln(\theta - c) \tag{32}$$

This equation gives the percentage development P per day of the egg for a given temperature θ. In fact, P is defined as $P = 100/t$, where t is incubation time to 50% hatch; that is, t is the average incubation time. By "development" we mean progress towards hatching. The total percentage development of the fish egg at day T is given by summing the daily percentages from T_{init} to T. Equation 32 is one of several possible empirical models presented in [1] to calculate incubation time. Each of these models is fit to pooled data to obtain the parameter estimates, and Equation 32 is the expression with the best fit. Coding this method of calculating incubation time is done by keeping track of the accumulated development percentage and stopping the loop when this value exceeds 100.

While incubating, we assume that the embryo and yolk sac grow according to the system of differential equations (3) given in [6]. These differential equations are mechanistic since they are based on the anabolism and catabolism of the fish, which is temperature-dependent. As in [6], we assume that both embryo and yolk sac are spherical and are connected by the embryo's vitelline system. If we let p be the radius of the yolk sac in millimeters at time t, and r be the radius of the embryo in millimeters at time t, we have

$$\frac{dr}{dt} = \frac{acA}{r^2} - br \tag{33}$$

$$\frac{dp}{dt} = -\frac{aA}{p^2} \tag{34}$$

In the above,

- $a = \alpha_1 \exp(\alpha_2 \theta)$, where α_1, α_2 are constants and θ is temperature. From [6] we have approximations to the parameter values α_1, α_2. These are given by

  - $\alpha_1 = 0.01024$

  - $\alpha_2 = 0.1315$

- $b = \beta_1 \exp(\beta_2 \theta)$, where β_1, β_2 are constants. From [6] we have

  - $\beta_1 = 0.000464$

  - $\beta_2 = 0.1906$

- c is the water absorption coefficient. It is assumed that before hatching $c = 1$, and after hatching $c = 2.433$.

- $A = A_t = \min(r_t^2, p_t^2)$.

The computational forms of Equations 33 and 34 are derived using the Forward Euler numerical method [2], which is a finite difference scheme that approximates the first

derivatives in Equations 33 and 34 by:

$$\frac{dr}{dt} \approx \frac{r_{t+1} - r_t}{\Delta t} \tag{35}$$

$$\frac{dp}{dt} \approx \frac{p_{t+1} - p_t}{\Delta t} \tag{36}$$

Thus we have from [5] that

$$r_{t+1} = r_t + \left(ac\frac{A_t}{r_t^2} - br_t \right) \cdot \Delta t \tag{37}$$

$$p_{t+1} = p_t - \frac{aA_t}{p_t^2} \cdot \Delta t \tag{38}$$

The weight of the embryo together with the yolk sac is calculated as density $\times$ volume, and assuming that embryo and yolk sac have the same density ρ, the weight (in milligrams) at time t is given by

$$w_t = \rho \cdot \frac{4\pi}{3}(r_t^3 + p_t^3) \tag{39}$$

In implementing Equations 37, 38 and 39, we take parameter values from [6]:

- Time step is $\Delta t = 0.25$ days

- Initial radius of embryo is $r_0 = 0$ mm

- Initial radius of yolk sac is $p_0 = 4.46$ mm

- Density of embryo and yolk sac is $\rho = 1.05$

- Additionally, the parameters a and b are temperature-dependent; that is, we actually have $a = a(\theta)$ and $b = b(\theta)$. More precisely, since our temperature data is measured daily (rather than 4 times per day, our time step Δt), we observe that for t being a multiple of four, $a(\theta_t) = a(\theta_{t+1}) = a(\theta_{t+2}) = a(\theta_{t+3})$ and similarly for $b(\theta_t)$.

Thus for the incubation stage, Equations 37 and 38 take the form

$$r_{t+1} \quad = \quad r_t + (a(\theta_t) - b(\theta_t)r_t) \cdot \Delta t \tag{40}$$

$$p_{t+1} \quad = \quad p_t - a(\theta_t)\frac{r_t^2}{p_t^2} \cdot \Delta t \tag{41}$$

Alevin growth is also calculated using Equations 37, 38 and 39; however, this life stage is divided into two parts for computation, and growth during each part is modeled slightly differently. As long as the radius r of the embryo (now referred to as the larva) is smaller than the radius p of the yolk sac, the equations governing growth of the alevin are

$$r_{t+1} \quad = \quad r_t + (2.433\, a(\theta_t) - b(\theta_t)r_t) \cdot \Delta t \tag{42}$$

$$p_{t+1} \quad = \quad p_t - a(\theta_t)\frac{r_t^2}{p_t^2} \cdot \Delta t \tag{43}$$

As soon as the larva grows larger than its yolk sac, the equations of growth are given by

$$r_{t+1} \quad = \quad r_t + \left(2.433\, a(\theta_t)\frac{p_t^2}{r_t^2} - b(\theta_t)r_t \right) \cdot \Delta t \tag{44}$$

$$p_{t+1} \quad = \quad p_t - a(\theta_t) \cdot \Delta t \tag{45}$$

Alevin remain in the stream-bed gravel and subsist entirely on their yolk sacs until they reach what is referred to as "maximum alevin wet weight", or MAWW. Nicholas Beer (1999) states that "maximum alevin wet weight is a key point in a chinook salmon's development because it is physiologically significant, easy to identify, and seems to coincide chronologically with emergence from the gravel" [5]. Once the alevin emerges, or "swims-up," it begins external feeding, and it is at this point that we decide the fish is physiologically a fry, and thus grows according to the fry growth equation given in equation (2) of [8]. We assume that the fry feed to satiation and that growth is not density-dependent. We also assume no predation, nor any mortality whatsoever. Given these assumptions, the weight

(in grams) of the fish at time t is denoted W_t and is given by

$$W_t = \left[W_0^b + bc \frac{(\theta_t - \theta_{LIM})t}{100 \left(\theta_M - \theta_{LIM} \right)} \right]^{1/b} \tag{46}$$

where

- W_0 is the initial weight of the fish.

- b and c are parameters.

- θ_t is the temperature at time t.

- θ_M is the optimal growing temperature for the species of fish.

- θ_{LIM} is a value that changes depending on the water temperature. If we let θ_U be the upper temperature limit for growth, and θ_L be the lower temperature limit for growth, then $\theta_{LIM} = \theta_L$ when $\theta \leq \theta_M$, and $\theta_{LIM} = \theta_U$ when $\theta > \theta_M$.

Equation 46 is an empirical model for fry growth, which contrasts with our earlier mechanistic model for egg and alevin growth. Although bioenergetic (mechanistic) models for fry growth do exist – the University of Wisconsin Center for Limnology and the Wisconsin Sea Grant Institute have created software entitled Fish Bioenergetics 3.0 which computes fish growth in terms of bioenergetics – we will be using Equation 46 because it is generally accepted to be a good model for fish growth [7], and it is relatively simple to implement. The computational form of Equation 46 we use is actually given by the Forward Euler approximation to the derivative

$$\frac{d(W^b)}{dt} \approx \frac{W_{t+1}^b - W_t^b}{dt} \tag{47}$$

and since our time step dt is taken to be one day, we have

$$W_{t+1} = \left[W_t^b + bc \frac{(\theta_t - \theta_{LIM})}{100 \left(\theta_M - \theta_{LIM} \right)} \right]^{1/b}. \tag{48}$$

The parameter values we use are for Atlantic salmon (*Salmo salar*) because we are unable to find parameter values for Chinook salmon. Although Pacific and Atlantic salmon are different, we assume that they are close enough relatives that we can use Atlantic salmon parameter values as estimates of Chinook salmon parameter values [7]. From [8] we have

- $b = 0.31$

- $c = 3.53$

- $\theta_M = 15.94$

- $\theta_L = 5.99$

- $\theta_U = 22.51$

In implementing the above growth equations to "grow" our fish, we take the final conditions of one life stage to be the initial conditions for the next life stage.

3.4.2   Weight Distribution

The second step of the simulation model is the MATLAB file

```
[M,S,Q,K]= FishDistribution(T,site).
```

This function returns the mean, standard deviation, skewness, and kurtosis of fish weights at day T, which is the Julian day of year 1 (the spawning year is taken to be year 0). The distribution of spawning times is assumed normal with standard deviation $\sigma = 15.3333$ days and mean 274 (Julian day of the year corresponding to September 30) [16]. The site input is one of 14159200 or 14159500. We grow 200 fish in the specified temperature regime and plot a histogram of fish weights at day T.

4   Results

We begin by first looking at our model for fish growth. Figure 5 is a plot of the weight of a fish from egg to fry (weight is measured in milligrams). Egg and alevin stages are plotted
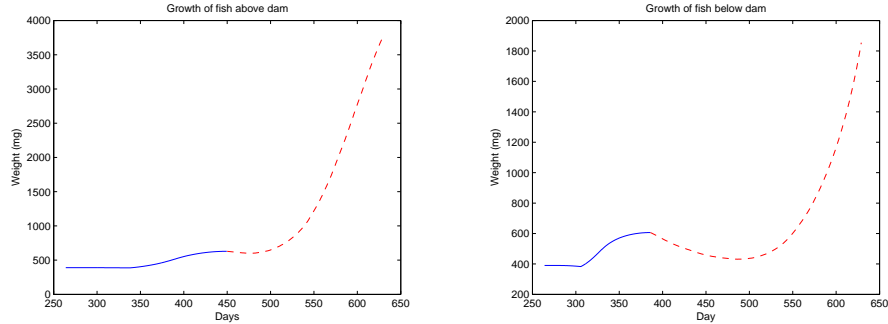
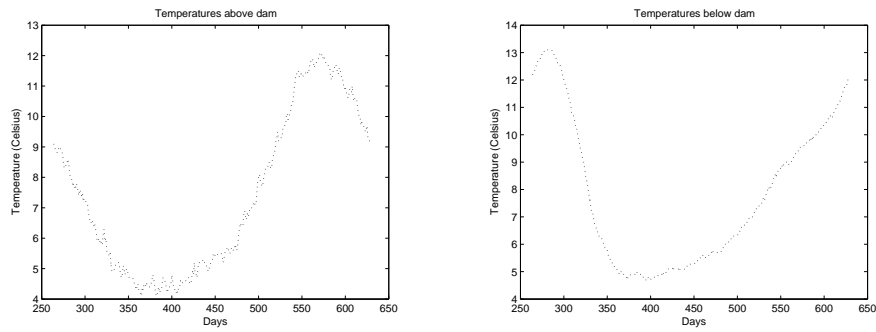Figure 5: Growth of fish above and below the dam.

Figure 6: Temperatures experienced during growth of fish above and below the dam.

as solid lines, and the fry stage is plotted as a dashed line. In each plot the spawning day is taken to be September 21 and the fish grow for one year. We also plot the temperatures experienced during the growth of each of these fish in Figure 6.

Interestingly, we observe a possible indication of the egg-fry conflict in Figure 5. We notice in both plots of fish weight that as soon as the fish hatches (becomes a fry), growth slows and becomes negative for a short period before increasing. Whether or not such weight loss may be physically possible, it supports experimental observations [13] that while eggs and alevin seem to tolerate the cold winter temperatures well (as evidenced by positive growth),the fry do not tolerate it well, as evidenced by a decreased growth rate.

We now turn to the distribution of fish weight. We chose 30 time points between April 20th and the December 16th which will be our values for T, the inputs for the code described in Section 3.4.2. For each value of T we "grow" 200 fish in each temperature regime. For
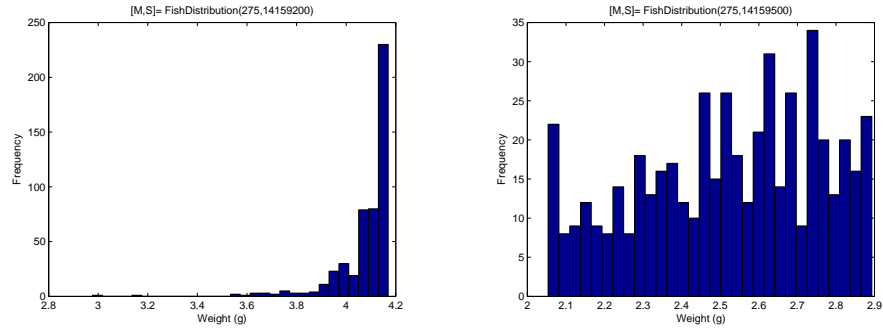
Figure 7: Histograms of fish weights above and below the dam on October 1, corresponding to the beginning of smolt outmigration [16].

Figure 8: Histograms of fish weights above and below the dam on October 15, corresponding to the beginning of the peak of smolt outmigration [16].

the generated data, including the mean and standard deviation of fish weights above and below the dam, see Appendix B. We show, as an example, four of pairs of histograms (see Figures 7, 8, 9, 10). Each pair of histograms shows the distribution of fish weights at time T depending on whether the 200 fish were "grown" above the dam or below the dam.

By observation it is clear that the distributions of fish weights above and below the dam are different. We especially note that the skewness of the above-dam data seems to be much different than the skewness of the below-dam data. To verify this we will use the two-sample t-tests to compare the first through fourth moments of each distribution; that is, the mean, standard deviation, skewness, and kurtosis (a measure of the "peakedness" of a distribution [15]) of the below-dam and above-dam weight distributions. We note that
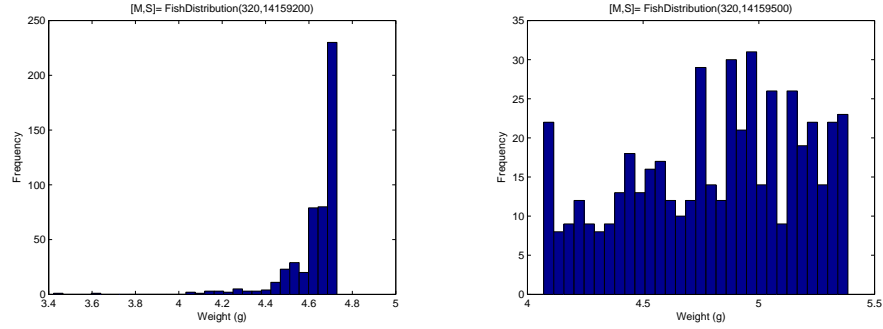
Figure 9: Histograms of fish weights above and below the dam on November 15, corresponding to the middle of the peak of smolt outmigration [16].
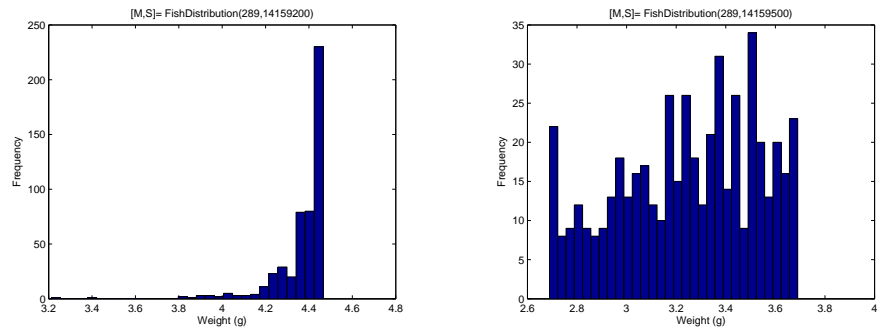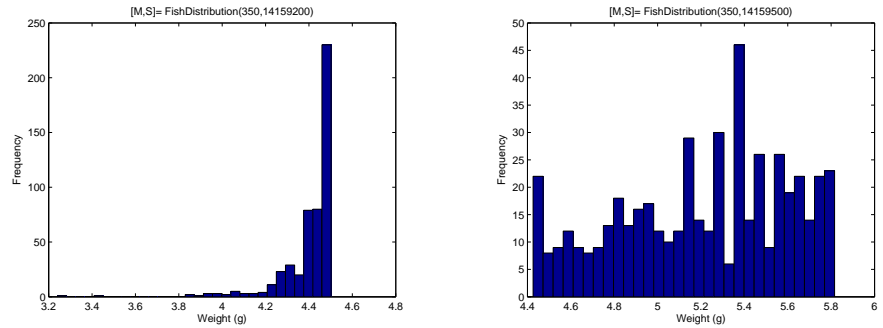

Figure 10: Histograms of fish weights above and below the dam on December 15, corresponding to the end of the peak of smolt outmigration [16].

in-sample and between-group independence is preserved since the model uses deterministic individual growth.

In order to conclude that the distributions of fish weights above and below the dam are different, only one of our four two-sample t-tests must be deemed statistically significant. We do, however, change the p-value that determines significance to $p < 0.05/4$, or approximately 0.01. This is called a Bonferroni correction [21]. We find that although the two-sample t-test performed on the means returns an insignificant p-value ($p = 0.16$), the three remaining two-sample t-tests gives us strong evidence to reject the null hypothesis that the above and below dam weight distributions are the same (for standard deviation: $p = 3.85e - 4$; for skewness: $p = 9.25e - 89$; for kurtosis: $p = 1.83e - 76$). Thus we conclude that our

distributions above and below the dam are very different.

Furthermore, we observe that fish grown above the dam tend to be larger than fish grown below the dam for the majority of the growing period. We use a Wilcoxon signed-rank test to see if the vector $M = Ma - Mb$ has a mean significantly different from zero, where Ma is a vector of mean weights above the dam recorded at regular time intervals, and Mb is a vector of mean weights below the dam recorded at regular time intervals. We find that indeed, the vector M does have a mean significantly different from zero (p = 2.6744e-004). The average of the vector Ma is 2.633 grams, and the average of vector Mb is 2.0305 grams.

5   Conclusions

Since the vector $M = Ma - Mb$ (where Ma is a vector of mean weights above the dam and Mb is a vector of mean weights below the dam) has a mean significantly different from zero, and also since the mean of Ma is greater than Mb, we can conclude that the temperatures above the dam tend to produce fish that are larger for most of the growing period than the temperature regime below the dam. If we assume that larger fish have a higher chance of survival, then our model shows that the temperature regime below the dam is less conducive to the survival of Chinook fry than the temperature regime above.

We note, however, that this assumption of "larger fish" being "fitter fish," i.e. more capable of surviving and reproducing, may not be true for all life stages of the salmon; in fact we mentioned in Section 1.1 that smaller fry, if they incubated as eggs at lower temperatures, tend to have a lower mortality rate than larger fry that incubated as eggs in warmer water [20].

We were not able to explicitly find a way to characterize the "egg-fry conflict," other than to observe that temperatures that promote positive growth in egg and alevin life stages may not promote positive growth in the fry life stage. One possible improvement to the current model would be to describe fry growth using a bioenergetic model. If this bioenergetic model included terms that modeled the temperature-dependent metabolism of the fish, then this term could be compared to the anabolism and catabolism terms in the embryo-yolk sac system described in [6]. This comparison might help to reveal the "egg-fry conflict."

Another improvement to the current model would be to implement a method of generating hypothetical temperature records that is more appropriate than ARMA modeling. For instance, we might instead construct a Markov chain with serial dependence, otherwise known as a Markov Chain of order m. This method requires decades' worth of daily temperature data, which gives us many temperature readings for each day of the calendar year. We determine conditional probabilities for each day that can be colloquially described as

"given the probability that yesterday's temperature was above (or below) the mean temperature for that day of the year, what is the probability that today's temperature is above (or below) the mean?" We do not have to restrict ourselves to looking back only one day, however, we can look back as far as m days, hence the Markov chain of order m. Once we have these conditional probabilities for each day of the year, we choose whether or not "today's" temperature will be above (or below) the mean. When we have decided (by selecting from a bag of white and red beads, or using a computer program, or any other suitable method), we randomly select a value for today's temperature drawn from the distribution of temperatures that have (historically) occurred on this day. If the temperature we select agrees with our decision for today (that the temperature today will be above – or below – the mean), then we choose this to be today's temperature. If it does not agree with our decision, we "throw out" our value and draw again until we have a temperature reading that agrees with our decision.

This is a highly involved process and our simulated temperature regimes will be rather noisy. However, we can be guaranteed that our simulated temperature on any given day will be within the historically accepted temperature range for that day. We will not get the types of impossible temperature regimes that the ARMA(p,q) models produce.

The results of this model are much more believable than the results of the summer model. This is because the underlying physical processes of the system are represented and the temperature inputs are more realistic. But, as mentioned before, this model applies to a very specific system and thus may not be generalizable to other temperature regimes, other geographic locations, or other salmon species. This model is merely to be used as a tool to compare different temperature regimes.

References

[1] D.F. Alderdice and F.P.J. Velsen. Relation between temperature and incubation time for eggs of chinook salmon (Oncorhynchus tshawytscha). *Journal of the Fisheries Research Board of Canada*, 35:69–75, 1978.

[2] K. Atkinson and W. Han. *Elementary Numerical Analysis*. John Wiley and Sons, Inc., Hoboken, NJ, 3rd edition, 2004.

[3] T.D. Beacham and C.B. Murray. Variation in developmental biology of sockeye salmon (Oncorhynchus nerka) and chinook salmon (O. tshawytscha) in British Columbia. *Canadian Journal of Zoology*, 67:2081–2089, 1989.

[4] T.D. Beacham and C.B. Murray. Temperature, egg size, and development of embryos and alevins of five species of Pacific salmon: a comparative analysis. *Transactions of the American Fisheries Society*, 119:927–945, 1990.

[5] W. Nicholas Beer. Comparison of Mechanistic and Empirical Methods for Modeling Embryo and Alevin Development in Chinook Salmon. *North American Journal of Aquaculture*, 61:126–134, 1999.

[6] W. Nicholas Beer and James J. Anderson. Modelling the Growth of Salmonid Embryos. *Journal of Theoretical Biology*, 189:297–306, 1997.

[7] D.T. Crisp. *Trout and Salmon: Ecology, Conservation and Rehabilitation*. Blackwell Science, Oxford, England, 2000.

[8] J.M. Elliott and M.A. Hurley. A functional model for maximum growth of Atlantic salmon parr, salmo salar, from two populations in northwest England. *Functional Ecology*, 11:592–603, 1997.

[9] Mary Flahive and Kenneth Bogart. *Discrete Mathematics Through Guided Discovery: Classnotes for Math 355/399*. OSU, 2007.

[10] U.H. Humpesch. Inter- and intra-specific variation in hatching success and embryonic development of five species of salmonids and Thymallus thymallus. *Archiv fur Hydrobiologie*, 104:129–144, 1985.

[11] Henriette I. Jager, Hal E. Cardwell, Michael J. Sale, Mark S. Bevelhimer, Charles C. Coutant, and Webb Van Winkle. Modelling the linkages between flow management and salmon recruitment in rivers. *Ecological Modelling*, 103:171–191, 1997.

[12] Jack Johnston and John DiNardo. *Econometric Methods*. The McGraw-Hill Companies, Inc., New York, NY, 4th edition, 1997.

[13] D.A. McCullough. A review and synthesis of effects of alterations to the water temperature regime on freshwater life stages of salmonids, with special reference to chinook salmon, 1999. Prepared for the U.S. Environmental Protection Agency.

[14] C.B. Murray and J.D. McPhail. Effect of incubation temperature on the development of five species of Pacific salmon (Oncorhynchus) embryos and alevins. *Canadian Journal of Zoology*, 66:266–273, 1988.

[15] NIST/SEMATECH. *Engineering Statistics e-Handbook of Statistical Methods*. U.S. Commerce Department Technology Administration. http://www.itl.nist.gov/div898/handbook/, accessed 5/23/2008.

[16] USACE (US Army Corps of Engineers). Biological assessment of the effects of the Willamette River Basin flood control project on species listed under the Endangered Species Act. Technical report, R2 Resource Consultants, Inc and USACE, Portland District Office, Portland, OR, April 2000. Prepared for the National Marine Fisheries Service and the U.S. Fish and Wildlife Service.

[17] Tom Quinn. *The behavior and ecology of Pacific salmon and trout*. University of Washington Press, Seattle, WA, 2005.

[18] W.E. Ricker. Growth Rates and Models. In W.S. Hoar, D.J. Randall, and J.R. Brett, editors, *Fish Physiology: Bioenergetics and Growth*, volume 8, chapter 11, pages 677–743. Academic Press, Inc., New York, NY, 1979.

[19] M. Teose. The effect of temperature on the survival of chinook salmon eggs and fry: a probabilistic model. Final paper written for the 2007 EcoInformatics Summer Institute, Oregon State University, August 2007.

[20] A.H. Weatherley and H.S. Gill. Growth. In C. Groot, L. Margolis, and W.C. Clarke, editors, *Physiological Ecology of Pacific Salmon*, chapter 3, pages 101–158. UBC Press, Vancouver, B.C., 1995.

[21] E.W. Weisstein. Bonferroni Correction. Online, 2008. From MathWorld, A Wolfram Web Resource. http://mathworld.wolfram.com/BonferroniCorrection.html.

Appendices

A   MATLAB Code

A.1   fishalevin.m

```
function [Th,Ta,wt]=fishalevin(Ts,Tf,site)
%
% function [Th,Ta,wt]=fishalevin(Ts,Tf,site)
% The function fishalevin.m calculates the weight of a single fish Tf days
% after spawning on day Ts, having experienced the thermal regime recorded
% at the specified site.
%
%     INPUTS
%         Ts = time of spawning [integer day of the year]
%         Tf = number of days after spawning at which you wish to know the
%         weight of the fish [days since spawning]
%         site = one of '14159500'(below dam) or '14159200' (above dam)
%
%     OUTPUTS
%         Th = time of hatching [integer day of year]
%         Ta = time of alevin swim-up [integer day of year]
%         wt = fish weight at time Tm [grams]
%
% Sample call sequence:
%
% [Th,Ta,wt]=fishalevin(264,365,14159500)
%
% Length of incubation is calculated using the log-inverse Belehradek
% equation from Alderdice and Velsen (1978). Embryo growth during egg and
% alevin stages is calculated using Eqns (3) in Beer & Anderson (1997). Fry
% growth calculated using eqn 2 from Elliott & Hurley (1997).


%% Step 0: Decide which temperature regime we're using.

if site == 14159500
    theta_n = [5 ... 5];

thetaN = zeros(1,2*length(theta_n));
thetaN(1:length(theta_n)) = theta_n;
thetaN(length(theta_n)+1:length(thetaN)) = theta_n;


elseif site == 14159200
        theta_n = [4.3 ... 4.4]';

thetaN = zeros(1,2*length(theta_n));
thetaN(1:length(theta_n)) = theta_n;
thetaN(length(theta_n)+1:length(thetaN)) = theta_n;

else
```

```
    disp('ERROR: Invalid Site Selection.')
    Th = NaN; Ta = NaN; wt = NaN;
    return
end


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Set-up
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

P = 0;                % percent embryonic development
p = [];
p(1) = 4.46;          % radius of yolk sac, in millimeters
r = [];
r(1) = 0;             % radius of embryo, in millimeters
W = [];
rho = 1.05;           % density of yolk and embryo (assumed to be equal)
W(1) = rho*(4/3)*pi*(p(1)^3+r(1)^3);
dt = 0.25;

kbel = 0.08646;    % these are paramters in log-inv Belehradek: pooled data
bbel = 1.23473;    %
cbel = -2.26721;   %

a = [];
b = [];
a1 = 0.01024;    % these are parameters in eqns (3) of Beer & And (1997)
a2 = 0.1315;     %
b1 = 0.0006464;  %
b2 = 0.1906;     %
c = 2.433;       %

for i=1:(1/dt):2*length(thetaN)
    time = (i-1)*dt + Ts;
    a(i:i+3) = a1*exp(a2*thetaN(time)); % parameter a changes with temp
    b(i:i+3) = b1*exp(b2*thetaN(time)); % parameber b changes with temp
end

Tm = Tf + Ts;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Incubation
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

day = 0;        % days since fertilization

for j = 2:500
    P = P + kbel*(thetaN(j-2+Ts)-cbel)^bbel; % calculate the hatching day
    day = day + 1;

    if P > 100
        break
    end
end
```

```
jmax = day/dt;

for j=2:jmax
    r(j) = r(j-1) + (a(j-1)-b(j-1)*r(j-1))*dt;        % radius of embryo
    emb(j) = rho*(4/3)*pi*(r(j))^3;                   % weight of embryo
    p(j) = p(j-1) - ((a(j-1)*r(j-1)^2)/p(j-1)^2)*dt;  % radius of yolk
    yolk(j) = rho*(4/3)*pi*(p(j))^3;                  % weight of yolk
    W(j) = emb(j) + yolk(j);                          % weight of fish
end

Th = Ts + day;  % time of hatching = time of spawning+length of incubation

%% Possibility of Tm < Th
if Tm < Th
    k = (Tm-Ts)/dt;
    Th = NaN; Ta = NaN;
    disp('Pre-hatching weight')
    wt = W(k)*0.001; % convert milligrams to grams
    return
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Alevin
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


jinit = jmax + 1;
jmax2 = jmax*3;

day2 = 0;    % day2 is the number of days since hatching


% Part one: Yolk sac is larger than embryo

for j=jinit:jmax2
    p(j) = p(j-1) - (a(j)*r(j-1)^2/p(j-1)^2)*dt;
    r(j) = r(j-1) + (a(j)*c - b(j)*r(j-1))*dt;
    W(j) = rho*(4/3)*pi*(p(j)^3+r(j)^3);

    if r(j)^2 > p(j)^2
        break
    end

    day2 = day2 + dt;
end

Q = day + day2;

jinit2 = Q/dt + 1;

jmax3 = (Tm-Ts)/dt;


% Part two: Embryo is larger than yolk sac
```

```
for j=jinit2:jmax3
    p(j) = p(j-1) - a(j)*dt;
    r(j) = r(j-1) + ((a(j)*c*(p(j-1)^2/r(j-1)^2) - b(j)*r(j-1)))*dt;
    W(j) = rho*(4/3)*pi*(p(j-1)^3+r(j-1)^3);

    if W(j) < W(j-1)
        break
    end

    day2 = day2 + dt;
end

% Q = day + day2;
maxtime = (length(W)-1)*dt;

t = (0:dt:maxtime) + Ts;


plot(t,W)
hold on

[C,I] = max(W);     % Find the maximum alevin weight C (in mg) and also...
                    % the time I in days*dt (after fertilization) this occurs

F = ceil(I*dt);

Ta = Ts + F;        % Ta is the day of swim-up

%% Possibility of Tm < Ta
if Tm < Ta
    k = (Tm-Ts)/dt
    Ta = NaN;
    disp('Pre-emergence weight')
    wt = W(k)*0.001;  % convert milligrams to grams
    return
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Fry
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

B = 0.31;           % Parameter values for fry growth eqn from Elliott & Hurley (1997)
c = 3.53;           % We use an empirical growth equation rather than a bioenergetic...
theta_M = 15.94;    % growth model for simplicity
theta_L = 5.99;     %
theta_U = 22.51;    %

w = [];
w(1:Ta-1) = 0;
w(Ta) = 0.001*max(W);   % set initial fry weight to be weight at swim-up (grams)

t2 = Ta:Tm;
```

```
for j=Ta+1:Tm
    if thetaN(j-1) <= theta_M
        theta_LIM = theta_L;
    else theta_LIM = theta_U;
    end
    w(j) = (w(j-1)^B + (B*c/100)*(thetaN(j-1)-theta_LIM)/(theta_M - theta_LIM))^(1/B);
end

wt = w(Tm);     % weight in grams

plot(t2,1000*w(Ta:Tm),'r--')
hold off
```

A.2   FishDistribution.m

```
function [M,S,Q,K]= FishDistribution(T,site)

% function [M,S,Q,K]= FishDistribution(T,site)
% Returns the mean, standard deviation, skewness, and kurtosis of fish weights at day
% T, which is the Julian day of the year *after* spawning occurs. Spawning times are
% randomly chosen from a discrete normal distribution with mean 274
% (Julian day of the year corresponding to September 30) and st.dev. of
% 15.3333 days. Site is one of 14159200 or 14159500. Sample call sequence:
% [M,S,Q,K]= FishDistribution(200,14159500)

%% Initialize vectors

Ts = [];
Th = [];
wt = [];
randn('state',100);
% Assume mean spawning day is September 30rd, standard deviation of
% spawning is 15.3333 days.
R = 15.3333*randn(1,500) + 274;

%% Define spawning distribution and grow 500 fish according to fishalevin.m algorithm.

for k = 1:200
    Ts(k) = round(R(k));
    Tf = (366 - Ts(k)) + T;
    [Th(k),Ta(k),wt(k)] = fishalevin(Ts(k),Tf,site);
end

%% Find first through fourth central moments of wt

M = mean(wt);
S = std(wt);
Q = skewness(wt);
K = kurtosis(wt);

%% plot results

% figure;
```

```
% hist(wt,30)
```

A.3 FishHistogramData.m

```
Ma = [];
Sa = [];
Qa = [];
Ka = [];

Mb = [];
Sb = [];
Qb = [];
Kb = [];

s = 110;
dt = 8;
f = 350;

D = s:dt:f;
d = [];
d(1) = s;

disp(sprintf('day \t Ma \t\t Mb \t\t Sa \t\t Sb \t\t Qa \t\t Qb \t\t Ka \t\t Kb'))

for i=1:length(D)
    [Ma(i),Sa(i),Qa(i),Ka(i)] = FishDistribution(D(i),14159200);
    [Mb(i),Sb(i),Qb(i),Kb(i)] = FishDistribution(D(i),14159500);
    d(i+1) = s + dt*i;
    disp(sprintf('%d \t%0.5f \t%0.5f \t%0.5f \t%0.5f \t%0.5f \t%0.5f \t%0.5f ...
    \t%0.5f',d(i),Ma(i),Mb(i),Sa(i),Sb(i),Qa(i),Qb(i),Ka(i),Kb(i)))
end
```

B   Simulation Output

day	Ma	Mb	Sa	Sb	Qa	Qb	Ka	Kb
110	0.60179	0.50224	0.03696	0.07537	-3.92336	-0.24968	23.75996	2.0689
118	0.60898	0.49784	0.0369	0.07526	-4.05057	-0.23201	24.81245	2.07826
126	0.62489	0.49827	0.03761	0.07537	-4.00053	-0.22898	24.58527	2.0802
134	0.65003	0.50326	0.03918	0.07587	-3.83128	-0.23079	23.25107	2.07966
142	0.69147	0.51275	0.0413	0.07687	-3.76921	-0.23175	22.61187	2.0802
150	0.74473	0.52804	0.04357	0.07846	-3.7824	-0.23326	22.74818	2.08105
158	0.8173	0.54924	0.04658	0.08064	-3.79853	-0.23525	22.91531	2.08219
166	0.90616	0.57851	0.05015	0.08362	-3.81595	-0.23785	23.09609	2.08367
174	1.02077	0.61768	0.05462	0.08753	-3.8354	-0.24106	23.29868	2.08553
182	1.17884	0.66932	0.06055	0.09257	-3.85805	-0.24492	23.53514	2.08779
190	1.35873	0.73241	0.06702	0.09856	-3.87947	-0.24912	23.75959	2.09029
198	1.5688	0.80314	0.07425	0.1051	-3.90026	-0.25331	23.97813	2.09281
206	1.80666	0.8869	0.08211	0.11262	-3.91984	-0.25768	24.18453	2.09548
214	2.07065	0.98455	0.09048	0.12112	-3.938	-0.26215	24.37645	2.09823
222	2.33839	1.09542	0.09865	0.13045	-3.95358	-0.26656	24.54154	2.101
230	2.63325	1.22274	0.10733	0.14083	-3.96827	-0.27095	24.69745	2.10379
238	2.9163	1.37047	0.11539	0.15246	-3.98048	-0.27535	24.8274	2.10663
246	3.20856	1.54106	0.12347	0.16541	-3.99158	-0.27972	24.94561	2.10947
254	3.47917	1.74443	0.13076	0.1803	-4.00074	-0.28416	25.04336	2.11241
262	3.72192	1.99344	0.13715	0.19782	-4.0082	-0.28876	25.12309	2.11549
270	3.93763	2.29541	0.14273	0.21818	-4.01432	-0.29342	25.18852	2.11865
278	4.13908	2.65806	0.14786	0.24158	-4.01966	-0.29805	25.2456	2.12183
286	4.31243	3.07459	0.15221	0.26727	-4.02399	-0.30244	25.29193	2.12489
294	4.44291	3.51891	0.15545	0.29352	-4.0271	-0.30634	25.32525	2.12763
302	4.55383	3.96337	0.15818	0.31877	-4.02965	-0.30964	25.3526	2.12998
310	4.61347	4.37213	0.15964	0.34124	-4.03099	-0.31227	25.36695	2.13186
318	4.6255	4.72163	0.15994	0.35993	-4.03125	-0.31428	25.36982	2.13331
326	4.62122	4.98716	0.15983	0.37385	-4.03116	-0.31568	25.3688	2.13432
334	4.55807	5.13784	0.15829	0.38165	-4.02974	-0.31644	25.35363	2.13487
342	4.48553	5.19045	0.1565	0.38436	-4.02809	-0.31669	25.33586	2.13506
350	4.39899	5.19465	0.15436	0.38457	-4.02606	-0.31671	25.31417	2.13507

Table 1: FishHistogramData.m Output

Ma is mean weight above the dam. Mb is mean weight below the dam. Sa is standard deviation of weights above the dam. Sb is standard deviation of weights below the dam. Qa is skewness of weights above the dam. Qb is skewness of weights below the dam. Ka is kurtosis of weight data above the dam. Kb is kurtosis of weight data below the dam.