A Predator Susceptibility Model of Juvenile Salmon Survival and a Voronoi Tessellation-based Approach for Generating Hypothetical Forest Landscapes

Gregor Passolt

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2012

Committee:

James J. Anderson

Sándor F. Tóth

Program Authorized to Offer Degree:
Quantitative Ecology & Resource Management

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thanks to my advisers who have guided me through this with understanding, my friends in the lab and in QERM who have provided insight, distraction and commiseration as required, and my family and friends for constant encouragement and support.

Chapter 1

PREDATOR SUSCEPTIBILITY MODEL OF JUVENILE SALMON SURVIVAL

Abstract

In relating juvenile salmonid size to adult survival, past observational studies have shown mixed results, while modeling has been limited to regression methods lacking in ecological mechanism. This paper develops a simple 4-parameter model relating the survival of juvenile Chinook salmon (*Oncorhynchus tshawytscha*) to their size based on susceptibility to predation. The model is parameterized to test the hypothesis that there is a critical size for smolts beyond which they experience significantly lower mortality due to a reduction in gape-limited predatory pressure. The model is validated on spring-run Chinook salmon tagged in the Columbia River Basin, tracking either in-river survival during out-migration from hatcheries to Lower Granite Dam or adult returns for fish tagged at Lower Granite Dam and at the mouth of the Grande Ronde River returning to Bonneville Dam. The predator susceptibility model outperforms a generalized linear model for adult return data and for in-river survival when mortality is high, while providing interpretable parameters and a mechanistic structure. These results show strong support for the critical-size critical-period hypothesis of Beamish & Mahnken (2001) while giving a plausible functional form for the response. There is also evidence that larger fish realize greater survival increases from barging than smaller fish. This provides a framework through which interannual variations in size-dependent survival can be separated from overall survival, which could help guide management targets in both hatchery and dam management as well as wild habitat restoration.

1.1 Introduction

It is generally accepted that the survival of juvenile fish depends on their size, and that bigger is better, however many studies contradict each other about the importance of this

relationship and the form it takes, while others find equivocal or negative results (Parker, 1971; Sogard, 1997). Previous work on the length-survival relationship in Pacific salmon (*Oncorhynchus* spp.) focused on the presence or absence of a length effect on recruitment, but only recently have attempts been made to give it a mathematical form other than a linear model (Satterthwaite et al., 2009; Hunsicker et al., 2011; Ohata et al., 2011). Koenings et al. (1993) used non-parametric Loess smoothing to find common patterns in survival with respect to length, but no general theory was developed, only the recognition of at least two "non-linear, functionally distinct zones." Even the presence of a length effect can be difficult to detect. Though usually present in data (Saloniemi et al., 2004; Jokikokko et al., 2006; Henderson & Cass, 1991; Holtby et al., 1990), some studies observed no effect of size on survival (Rechisky, 2010) or even a negative effect (Ewing & Ewing, 2002). Many of the equivocal results had limiting factors in the data, for example acoustic tags have a larger size requirement than passive integrated transponder (PIT) tags, thus studies dependent on acoustic tag data necessarily truncate the smaller fish, which can affect the results by preventing comparisons with smaller individuals. Rechisky (2010) examined Dworshak Hatchery Chinook salmon smolts tagged with acoustic tags, and binned the data into 5 mm length classes starting at 130 mm. In an AIC-based regression analysis of her 2008 data, fork length was a significant factor in the best-rated model for adult returns to Lower Granite Dam—until the largest length class (150–154 mm) was removed, at which point a model without fork length provided the best fit, though the new model included travel time as a parameter, which was correlated with fork length. The minimum recommended size for PIT-tagging is 65 mm (MPM, 1999), which enables consideration of wild fish and smaller hatchery fish that are unavailable for acoustic tagging. In the case of Ewing & Ewing (2002), the authors found a negative correlation between length and survival, but suggested that unusually warm ocean conditions coupled with high testosterone levels in the larger juveniles to created the anomalous effect. In fact, Duffy & Beauchamp (2008) conducted an analysis of the same stocks as the Ewing paper for more recent years and found a strong length effect, attributed primarily to predation in the early marine period.

Migrating from one ecosystem to another, juvenile Chinook salmon face different sets of predators. In-river, there is heavy predation due to smallmouth bass (*Micropterus dolomite*),

northern pikeminnow (*Ptychocheilus oregonensis*), channel catfish (*Ictalurus punctatus*), and walleyes (*Stizostedion vitreum*) (Vigg et al., 1991; Tabor et al., 1993; Ward et al., 1995; Shively et al., 1996a; Petersen & Ward, 1999; Fritts & Pearsons, 2004). After entering the ocean, potential predators include cutthroat trout (*Oncorhynchus clarkii clarki*) and adult salmon; avian predators are present throughout (Parker, 1971; Holtby et al., 1990; Duffy & Beauchamp, 2008). The critical-size, critical-period hypothesis of Beamish & Mahnken (2001) (see also Farley et al., 2007) does not address freshwater predation and mortality, but makes a strong case that marine survival is determined in two stages, both of which depend on size and growth. In the first stage, which starts upon ocean entry (usually early to mid-summer for spring Chinook salmon), size-selective predation is the primary cause of mortality. It is well established that gape-limitation results in linear relationship between predator size and maximum prey size, with slope and intercept dependent on the species involved (Poe et al., 1991), and that these relationships in large part determine size-selectivity in predation. The second stage begins at the end of the first marine summer, where it is theorized that survival depends on accumulating enough energy reserves to survive the first marine winter, which is signified by maintaining a high enough growth rate to achieve a critical size. The critical size varies depending on ocean condition and resource availability, but those falling short are most susceptible to predation. Farley et al. (2007) observed that resource availability should be taken into account in managed stocks as in a year of low ocean productivity producing more juveniles could decrease survival overall if competition is too intense. Ocean conditions can affect survival in both size-dependent and size-independent ways.

Past studies on a length or size effect on survival have focused at one of two scales. Many are at a stock or location scale, using mean lengths and survival rates for different sites or for the same site in different years. Requiring less data, these studies were especially popular historically. The question they sought to answer was if increasing the overall size within a population results in higher survival. Many of them used only a handful of data points and fit simple linear models (e.g. Beckman et al., 1999; Henderson & Cass, 1991), but as Henderson & Cass pointed out, means tend not to vary widely which makes effects difficult to detect. Reisenbichler et al. (1982) plotted 18 years of data from three Sacramento River Chinook salmon hatcheries and found a strong, non-linear length effect, and Koenings et al.

(1993) found both a similar length effect in a study of sockeye salmon from 12 sites in Alaska over a long time interval.

Studies conducted within a cohort look at how the length distribution of survivors relates to the overall length distribution. This can be done prospectively by measuring and PIT-tagging (Zabel & Williams, 2002) or retrospectively, using a method such as scale radius analysis, where the size at ocean entry is inferred by examining a scale taken from the adult fish, and the distribution of survivors' sizes is then compared to a previously sampled distribution of all out-migrating fish (Claiborne et al., 2011). These studies generally found increased survival in larger fish (see also Moss et al., 2005; Cross et al., 2008), but as Saloniemi et al. (2004) began to do for Atlantic salmon, I would like to know how the size effects vary with year. Zabel & Achord (2004) used both within- and across-cohort approaches on Chinook salmon and found that length was an excellent predictor of survival within a cohort, but that mean length was a poor predictor of cohort survival. Their analysis was based solely on Generalized Linear Models (GLM), where a single parameter was responsible for the entire length effect. A similar analysis, also based on GLMs was conducted by Jokikokko et al. (2006) for Atlantic salmon, where length and type (hatchery or wild) were observed to have significant effects on survival. Good et al. (2001), also working with Atlantic salmon, found that the strength of size-selective effects depended on ocean conditions.

Both Reisenbichler et al. (1982) and Koenings et al. (1993) noted diminishing returns in the benefits of size which can help explain some of the inconsistency of size effects, but this also raises the issue of functional form: is a GLM an appropriate model to fit the observed survival patterns? Diminishing returns would suggest a sigmoidal function with more flexibility than a GLM, which always approaches 1 asymptotically. Furthermore, a functional response with additional parameters allows for analysis off *how* the length-survival relationship changes with other factors. In this paper, I develop a mechanistic model that fits the observed patterns in survival better than a GLM while still remaining tractable and parsimonious. With parameters that correspond to different response characteristics, the model aims to decompose the size effect into independent functional components.

1.2 Theory

Dividing prey mortality into size-dependent and size-independent causes, a differential equation can express the mortality experienced by salmon of a certain size l. Beginning at a certain point on the juvenile salmon migration,

$$\frac{dN_l}{dt} = -c_1 N_l - c_2 N_l P_l, \tag{1.1}$$

where N_l is the number of prey of size l, P_l is the proportion of size-limited predators able to consume prey of length l, and c_1 and c_2 are constants controlling the rates of size-independent and size-dependent mortality, respectively. Solving Equation 1.1 with the initial condition that at $t = 0$, $N_l = N_l(0)$ yields

$$N_l(t) = N_l(0) \cdot \exp(-c_1 t) \cdot \exp(-c_2 P_l t). \tag{1.2}$$

Survival rates, $N_l(t)/N_l(0)$, are usually considered rather than amounts. The available data is not continuous with time. Rather, it records survival from one location to another. Evaluating at a certain time $t = T$ (and suitably adjusting the constants as $c_3 = \exp(-c_1 T)$ and $c_4 = -c_2 T$) survival rate as a function of size becomes

$$S(l) = \frac{N_l(T)}{N_l(0)} = c_3 \exp(-c_4 P_l). \tag{1.3}$$

This assumes that travel speed—and therefore time—is independent of length, though in actuality larger fish tend to migrate more quickly which also serves to reduce the time spent in high-risk areas and can therefore be wrapped into the constants. In this form, the survival curve has an upper asymptote at c_3 (which is related to c_1, the size-independent mortality coefficient) and a lower asymptote at $c_3 \exp(-c_4)$, however other interesting characteristics such as the location of the inflection point are intractable. A first-order Taylor approximation about the point $P_l = p_0$ yields a functional form with more readily interpreted parameters.

$$S(l) \approx c_3 \exp(-c_4 p_0) - c_3 c_4 p_0 \exp(-c_4 p_0) \cdot (P_l - p_0) \tag{1.4}$$

$$= c_3(1 + c_4 p_0) \exp(-c_4 p_0) - c_3 c_4 p_0 \exp(-c_4 p_0) \cdot P_l. \tag{1.5}$$

Making the following substitutions in the constants,

$$k_1 = c_3 c_4 p_0 \exp(-c_4 p_0) \tag{1.6}$$

$$k_0 = c_3 \exp(-c_4 p_0)\big[1 + c_4 p_0 - c_4\big], \tag{1.7}$$

Equation 1.5 simplifies to

$$S(l) \approx k_0 + k_1(1 - P_l). \tag{1.8}$$

In this form the lower and upper asymptotes are k_0 and $k_0 + k_1$ and the curve connecting them will share properties, such as symmetry and the inflection point, with P_l, the proportion of predators capable of consuming a smolt of size l. An intuitive explanation for this form of the model follows.

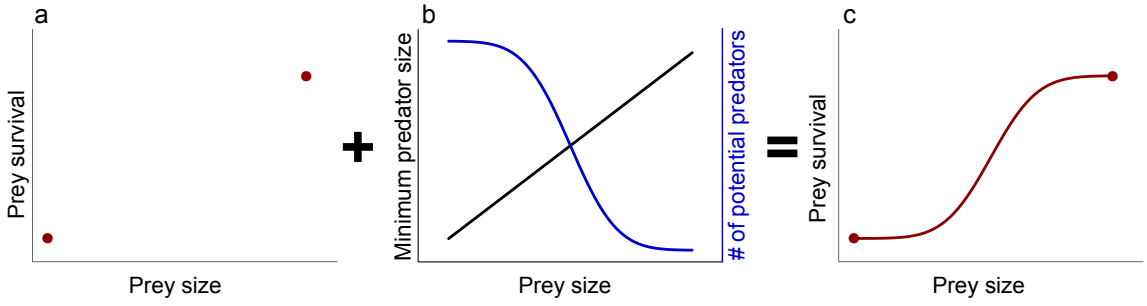

Figure 1.1: Panel (a) shows an intuitive starting assumption, with the extreme sizes taking on extreme survival rates. Panel (b) shows how the number of predators decreases with prey size, assuming a normally distributed predator population and a linear relationship between prey size and minimum predator size. Panel (c) combines the information in the other panels as in Equation 1.8 to produce a sigmoidal survival curve.

Assume that the smallest smolt, equally available to all predators, has a probability of survival k_0. Similarly, assume that a very large smolt, essentially free from size-selective predation, has a probability of survival that is greater by k_1. This gives survival probabilities k_0 and $k_0 + k_1$ at two points on opposite ends of the length spectrum. The remaining question is how to connect them. Within this modeling framework, the only difference between the limiting points of the length curve is the number of potential predators, so this is the basis for the path connecting the dots. There is some distribution of predators by the maximum

size of the prey they can consume, and the shape of this distribution will determine the shape of the survival curve. This assumes that predation is only limited by the maximum prey size, i.e. a predator capable of hunting a certain-sized prey is equally able and likely to consume any smaller prey. (This assumption is addressed further in the discussion.) Under this assumption, the cumulative distribution function (CDF) of the predator-density in terms of maximum prey size determines the risk of predation, thus the parameters that must be estimated for model fitting are k_0 and k_1 for the position of the asymptotes and whatever parameters are necessary to describe the predator-density distribution. Figure 1.1 builds this model graphically assuming a normal distribution of predators, while Figure 1.2 shows a hypothetical system where two overlapping normally distributed predator populations with different prey-size restrictions produce a survival curve with two distinct ramps.

For a certain species, Poe et al. (1991) showed that, due to gape limitation, there is a strong linear relationship between the length of the largest possible prey and the length of a predator. Using the initial assumption that predator sizes are normally distributed by length, this means that the number of predators just big enough to successfully hunt prey of a given length is normally distributed by prey length through a linear transform of the predator length distribution. While this holds easily for a single predator population, extending it as an assumption of the net distribution for all size-limited predators is reasonable and requires estimating just 2 additional parameters. The mean of this distribution, l_c, equal to the length of the largest prey available to the mean predator, is a critical or threshold length because it marks the inflection point of the cumulative distribution function. This results in survival probability for a fish of length l of

$$S(l) = k_0 + k_1 \Phi\left(\frac{l - l_c}{\sigma}\right) \tag{1.9}$$

where k_0 is the low survival asymptote, k_1 is the difference between the low and the high asymptotes, l_c is the threshold length (the maximum prey length of the mean predator), and σ_c is the standard deviation of the maximum prey length distribution which affects the steepness of the ramp between the asymptotes, and Φ is the standard Gaussian CDF.

Due to changing ocean and river conditions, I hypothesize that k_0 and k_1 both are heavily susceptible to a year effect, whereas l_c and σ_c should be relatively constant as long

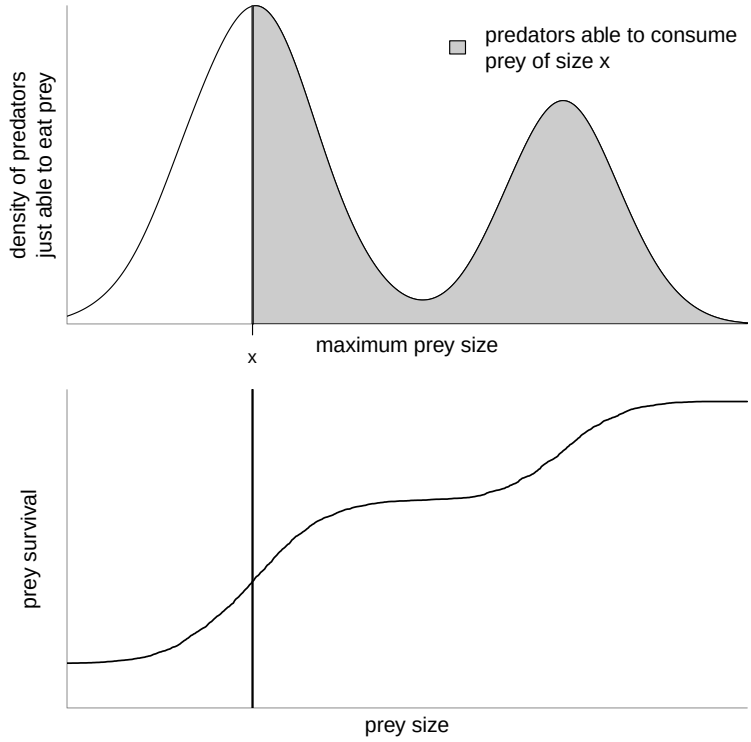

Figure 1.2: The modeling principles can be easily generalized to different distributions of pressure due to predation. In the example above, a bi-modal predator distribution (more accurately a bi-modal distribution in the maximum size of prey that predators are capable of hunting) results in a survival curve with two "steps". The predator susceptibility model can be as flexible as the maximum prey size distribution.

as predator populations and size distributions are relatively constant. All parameters should depend heavily on distance traveled in-river and prey characteristics and behavior, thus are location- and stock-specific.

1.2.1 Using a GLM

A standard GLM with a logit transform is used as a comparison. At its simplest, this follows the equation

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot l \tag{1.10}$$

where p is the probability of survival, l is length and the β's are the model parameters. Fixed effects for changes by year are added to both the intercept and length terms, and the best model is chosen by comparing Akaike Information Criterion (AIC) scores on data with normalized predictors. The GLM is valid over the entire real line of possible length values, with asymptotes at 0 and 1 survival probability as lengths approach positive and negative infinity. The best fitting GLM provides a base case from which to assess the performance of the predator susceptibility model. The GLM has the advantage of having fewer parameters, however its asymptotes are fixed at 0 and 1, and the parameters are difficult to interpret ecologically, for example β_1 is the change in the log odds ratio due to unit change in length. The predator susceptibility model, on the other hand, introduces 2 additional parameters, but has flexible asymptotes and all four parameters are easily interpreted. Taken by itself, the only assumption the GLM tests is whether or not there is a length effect. If the GLM results in a non-significant or negative coefficient for length, it indicates that there is no evidence of a length effect, and thus the assumptions of the predator susceptibility model do not hold.

When both GLM and predator susceptibility models are fit to the same data they are compared using the AIC, which assigns a score to how well a model fits a particular data set including a penalty for extra parameters. Using standard rules of thumb, if the difference in AIC scores is less than 2 the models fit equally well, from 2 to 7 indicates a moderately significant improvement in fit by the lower scoring model, and difference greater than 10 indicates highly significant improvement (Burnham & Anderson, 1998). The GLM by itself can test for the presence of a length effect, then comparison with the predator susceptibility model will show if the length effect takes the sigmoidal shape I hypothesize with asymptotes between 0 and 1. The shape of the predator susceptibility model can approximate a logistic GLM under certain parameter values (large σ_c, $k_0 \approx 0$, $k_1 \approx 1$), so when the GLM outperforms the predator susceptibility model it is due mostly to the difference in the number of parameters. However, when the AIC of the predator susceptibility model is more than 10 lower than that of the GLM, the predator-susceptibility model's flexibility has significantly improved the fit, and these fits are classified as "very significant".

The predator susceptibility model is fit to individual binary survival data using the `mle2`

function from the `bbmle` package in R (Bolker & R Development Core Team, 2011). Though they will not be individually presented, model fits are validated by likelihood profiling, and these are used to generate confidence intervals on the parameters. An example likelihood profile plot is show in Figure 1.3. Graphs of the model fits presented with binned data never indicate that binned data were used to compute parameter estimates, the bins are merely a convenience to visually assess the model fits.

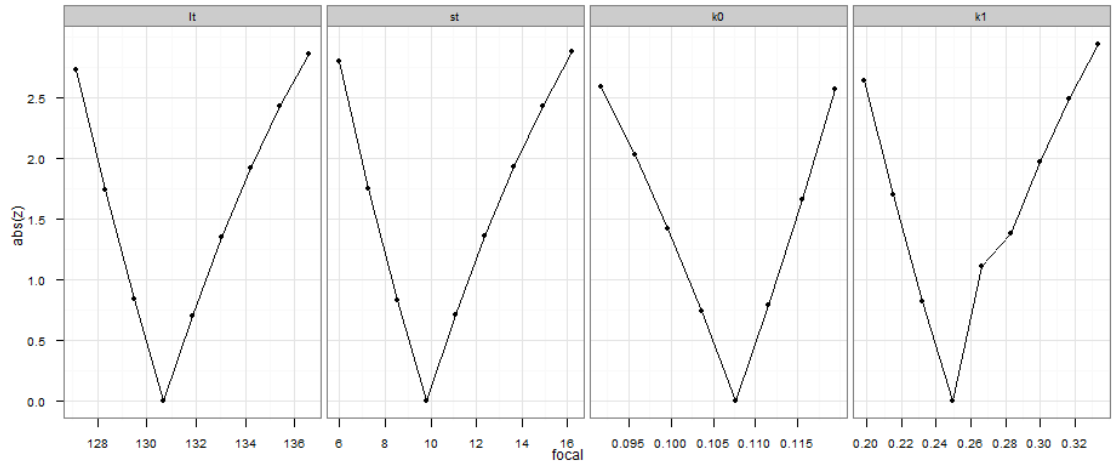

Figure 1.3: A sample likelihood profile plot. This corresponds to the fit to the Sawtooth Hatchery data. The steep "V" shape indicates definite minima in the negative log-likelihood surface, showing that the model has converged on a stable solution. The z-scores on the vertical axis are used to estimate confidence intervals for each of the four model parameters in the horizontal axes.

1.3 Data

Three distinct data sets are used to test the predator-susceptibility model. All three data sets follow out-migrating spring Chinook salmon from the Columbia River basin fitted with Passive Integrated Transponder (PIT) tags and recorded in the Columbia Basin PIT Tag Information System (PTAGIS). Nonexistent and extremely high length values, temporal outliers, as well as length values less than 65 mm were omitted. (It is not recommended that fish smaller than 65 mm be PIT-tagged (MPM, 1999).) As travel time plays a strong

role in mortality rates of Chinook salmon and is strongly correlated with fish size (Smith et al., 2002; Anderson et al., 2005) in addition to distance, it must be considered in the analysis as it could confound the effect of length on survival. The most practical way of dealing with distance effects is to select fish based on PIT-tagging location. For in-river survival, I selected fish PIT-tagged pre-release at the Sawtooth Fish Hatchery on the Salmon River, with survival defined by the existence of any positive detections at Lower Granite Dam or farther downstream. Adult survival, defined as any positive detection event as an adult—usually at Bonneville Dam—was tracked for fish PIT-tagged at the mouth of the Grande Ronde River and at Lower Granite Dam (treated separately). These sites are shown on the map in Figure 1.4.

1.3.1 In-River Survival: Sawtooth and Rapid River Fish Hatcheries

The least complicated data set is of spring Chinook salmon PIT-tagged and released at two hatcheries upriver from Lower Granite Dam with survival tracked to the dam. These two data sources are fish reared at the Sawtooth Fish Hatchery on the Salmon River, 1,442 km from the mouth of the Columbia (747 km above Lower Granite Dam), and Rapid River Hatchery, 978 km from the mouth (283 km above Lower Granite Dam). For these smolts survival was measured to Lower Granite Dam. Individual PIT-tagged fish were marked as survivors if their tag was detected at Lower Granite or any site downstream, making these conservative survival estimates depending on detection probabilities. Zabel et al. (2005) showed that smaller fish have greater detection probabilities, though this depends on both year and site. This could lead to underestimating the survival rates of large subjects. The Sawtooth data consists of 15,765 fish released in 10 years, although the vast majority of the taggings were in 1989 and 1991, summarized in Table 1.1. The lengths and survival rates of the Sawtooth population changed significantly from the early 90's to the late 90's and early 2000's. The 1989 and 1991 fish exhibited low survival rates compared to the later years. The length distributions changed as well. With different data characteristics on either side of a 6-year gap in the Sawtooth data, and the majority of observations coming before the break, the early and later Sawtooth data is treated separately.

The Rapid River data set, summarized in Table 1.2, is more uniform than the Sawtooth data. There are no gaps in the years covered, and the lengths and survivals fluctuate normally. The Rapid River survival rates are higher than the Sawtooth rates as the fish have less than half as far to travel to Lower Granite Dam. A potential issue with fish tagged in the hatchery is that the smolts are not necessarily tagged and measured near the release date. The records include both the tagging and release dates, allowing us to precisely calculate the latency time between tagging and release. This is of concern because fish in the hatchery system are under no predation threat and, if they are growing rapidly, fish tagged and measured 30 days before release (high latency) would not be comparable with those tagged very close to release (low latency). These particular hatcheries were chosen because a preliminary examination of the latency distributions for all hatcheries with over 10,000 PIT-tagged fish suggested that the Sawtooth and Rapid River data had good natural breaks in the latencies, allowing for easy grouping. As it turned out, the fish experience such cold river conditions in their final months in the hatchery—often the Sawtooth raceways are covered in ice—that their growth during this period is negligible (Philip Coonts, Sawtooth Hatchery, personal communication).

1.3.2 Adult Returns: Grande Ronde River Mouth

Wild and hatchery stocks are known to have different survival characteristics, and their length distributions are different enough that they are treated separately in this analysis. Preliminary analysis was conducted using 45,395 PIT-tagged spring Chinook salmon captured and released within at the mouth of the Grande Ronde River, about 800 km from the ocean, accessed from the PTAGIS database on 9 November 2010 and updated on 8 June 2011. The division examined in this data set is wild versus hatchery with breakdowns by year. The response variable for this data set is Smolt-to-Adult Return (SAR), which for this paper is defined as an adult detection at Bonneville Dam or above.

Year	Tagged	Survival to Lower Granite (%)	Mean length (mm)
1989	7039	15.71	117.4
1991	7081	6.62	113.6
1998	287	50.87	138.7
2001	500	51.20	134.9
2002	248	39.92	131.8
2004	203	54.68	112.4
2005	100	20.00	118.1
2006	100	46.00	117.8
2007	207	46.38	112.0

Table 1.1: Summary of the fish tagged at the Sawtooth Fish Hatchery. No fish were tagged at Sawtooth between 1991 and 1998, and after that relatively few fish were tagged in any given year. Three years in the middle, 1998, 2001, and 2002 have exceptionally high lengths, while the survival rates were much lower in 1989 and 1991 compared to the later years.

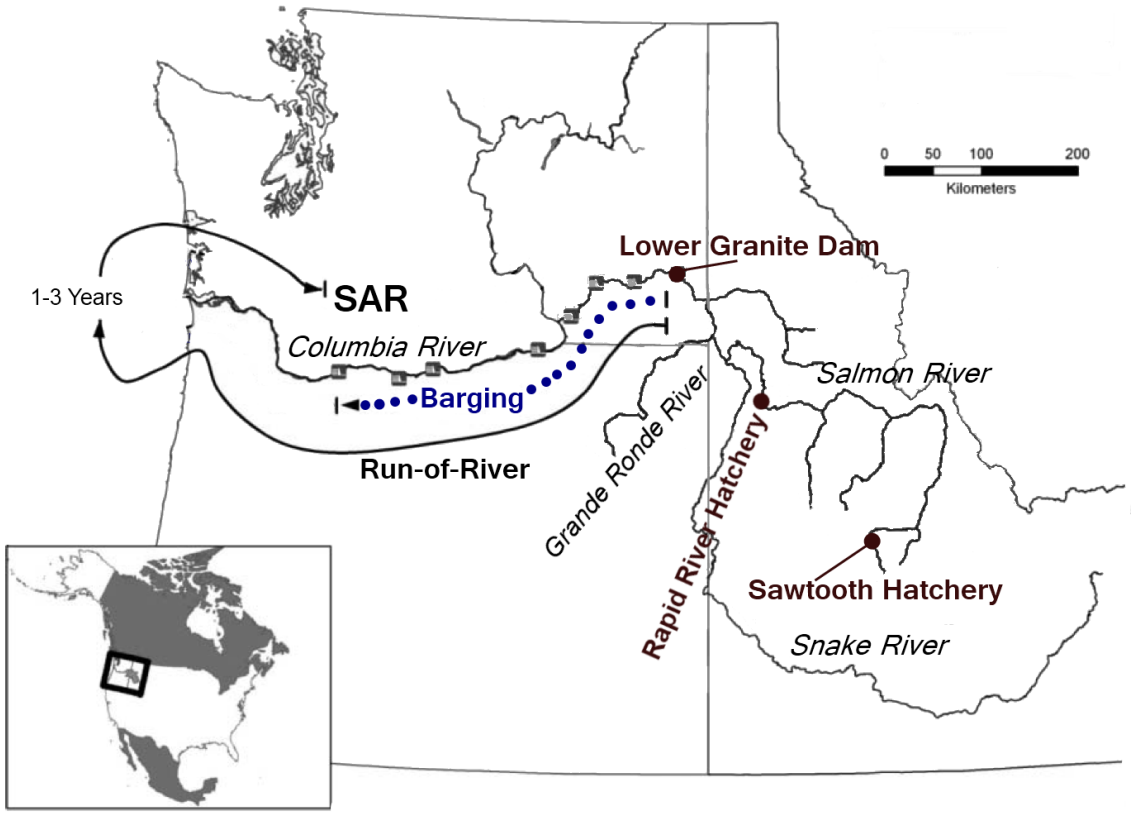

Figure 1.4: Data sets consist of fish PIT-tagged at the mouth of the Grande Ronde River (where it enters the Snake River), at Lower Granite Dam on the Snake River, and at Sawtooth Hatchery on the Salmon River. Survival of the Sawtooth fish is in-river pre-hydrosystem, from the hatchery to Lower Granite Dam, whereas smolt-to-adult returns (SAR) from the tagging site to Bonneville Dam are measured in the Lower Granite and Grande Ronde data. Some of the fish passing Lower Granite Dam were loaded onto barges and taken downriver to be released below Bonneville Dam (BON, on the map), while the rest made the same trip as run-of-river fish. (Map adapted from Haeseker et al. (2012).)

Year	Tagged	Survival to Lower Granite (%)	Mean length (mm)
1993	3285	54.82	114.3
1994	2910	35.12	119.2
1995	1961	61.45	120.8
1996	19072	43.73	124.4
1998	48339	59.62	117.5
1999	45409	61.65	120.8
2000	47577	54.69	119.6
2001	54915	67.11	119.0
2002	182913	60.26	122.7
2003	135717	47.43	120.1
2005	5277	70.21	123.9
2006	49871	61.63	122.0
2007	55584	55.33	115.2
2008	70711	65.61	122.8

Table 1.2: Summary of the fish tagged at Rapid River Hatchery. The variations in length and survival are small compared to the variations in the Sawtooth data set. The survival rates are higher than Sawtooth due to Rapid River's proximity to Lower Granite Dam.

	Wild					Hatchery				
Year	Tagged	Returned	SAR (%)	Mean length (mm)	Median DOY	Tagged	Returned	SAR (%)	Mean length (mm)	Median DOY
1998	1567	4	0.14	113.5	93	846	3	0.35	130.3	99
1999	3237	38	2.11	109.2	110	1771	7	0.40	127.7	83
2000	589	9	1.81	118.6	130	257	3	1.17	141.4	130
2001	2043	2	0.30	120.7	110	1375	0	0.00	134.3	99
2002	2733	29	1.42	113.1	109	1397	10	0.72	135.7	107
2003	4062	14	0.30	109.7	96	1396	6	0.43	150.5	83
2004	4501	19	0.45	112.0	105	1401	5	0.36	126.6	83
2005	3375	10	0.46	113.4	121	1401	1	0.07	129.0	90
2006	5029	49	1.02	111.4	115	1400	12	0.86	124.4	94
2007	3964	51	1.44	115.0	102	1394	14	1.00	129.3	104
2008	3831	105	3.09	114.4	116	1407	30	2.13	129.7	108
2009	4856	67	1.36	111.8	119	1396	20	1.43	123.3	93

Table 1.3: Summary table for the fish tagged at the mouth of the Grande Ronde River. There is significant yearly variation in the length distributions of hatchery fish (indicated by the variable means). The median day-of-year (DOY) of tagging is also given. The absolute number of returns in any given year is often too low to accurately estimate a length distribution of returners.

1.3.3 Lower Granite Dam Adult Returns

The largest and most complex data set consists of Chinook salmon smolts PIT-tagged as they passed Lower Granite Dam, 695 km from the estuary, for all years with substantial data available and mostly completed returns by the time of writing: 1998-2009 (out-migration year). Many of the smolts PIT-tagged at Lower Granite Dam are also selected for transport: they are loaded onto barges and shipped downriver. After a 36-hour trip, they are offloaded below Bonneville Dam. Run-of-river fish remaining in the hydrosystem for those 234 km take 2-4 weeks to cover the distance. PIT-tagging and barging practices differ from year to year at Lower Granite, with smolts of different types being tagged (or not) and assigned to barge transport (or not) with different treatment plans. These data are summarized in Table 1.4 with breakdowns by year, type, and transport. As in the Grande Ronde River data, the response variable is SAR to Bonneville Dam. Table 1.4 shows that the SAR rates fluctuate by an order of magnitude with year, however Figure 1.5 shows that the length distributions for the wild- and hatchery-type are consistent, and the differences between the types are clear.

To increase the sample size, along with those smolts positively identified as spring Chinook salmon, smolts with the "15H" run-type code were included, the classification given when the tagger is unable to determine if they are spring or summer Chinook salmon. Temporally, the data were truncated after a tagging date of June 30, which is early enough that the vast majority of the remaining fish can be expected to be spring-run (Beamesderfer et al., 1996). This increased the sample size from about 16,000 (positively identified spring Chinook salmon) to 1.2 million, substantial enough to justify the inclusion of some other runs as well.

	Wild				Hatchery			
Year	Barge	Barge SAR (%)	River	River SAR (%)	Barge	Barge SAR (%)	River	River SAR (%)
1998	5482	0.60	8683	0.61	38770	0.66	61523	0.38
1999	8119	2.35	11829	1.39	43169	2.16	61478	1.49
2000	0		58496	1.77	0		0	
2001	17506	1.01	0		0		0	
2002	4899	1.47	33935	1.03	0		0	
2003	7101	0.42	43051	0.17	0		0	
2004	11194	0.54	0		0		0	
2005	12668	0.31	0		0		36094	0.09
2006	19505	0.91	10095	0.77	33935	0.81	148258	0.57
2007	16752	1.10	14513	0.68	23935	1.02	78111	0.24
2008	19022	3.25	9412	1.77	42452	2.36	98288	1.22
2009	13315	0.98	13815	0.43	0		91253	0.41

Table 1.4: Fish PIT-tagged at Lower Granite Dam. The PIT-tagging practices differ from year to year; the fluctuations in barge rates and the relative numbers of a wild and hatchery fish tagged are due to changing PIT-tagging policies. The SAR rates also vary widely by year from 0.09% to 3.25%. Wild fish tend to exhibit higher overall survival. Only for barged fish in 1998 and run-of-river fish in 1999 do hatchery reared fish have SAR rates exceeding their wild counterparts.

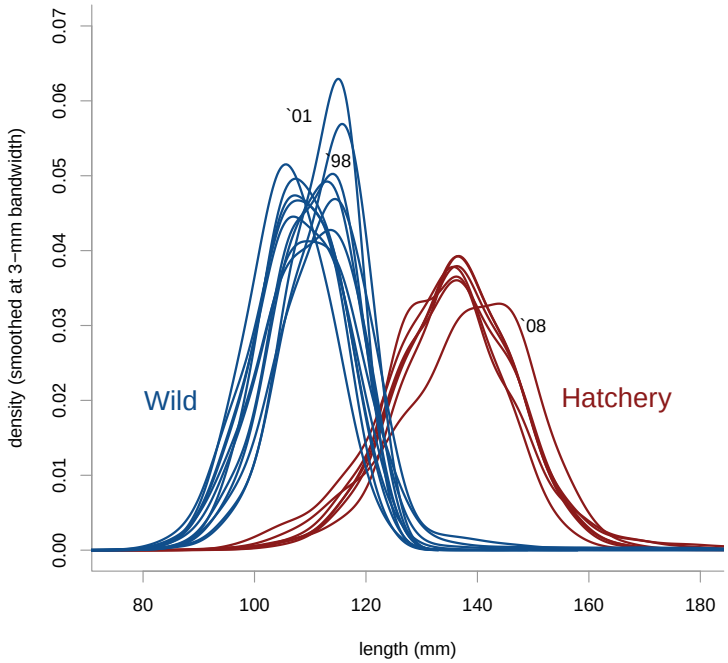

Figure 1.5: The differing distributions of fork length of wild and hatchery spring Chinook salmon stocks tagged at Lower Granite Dam by year. A 3 mm bandwidth smoother was applied to eliminate extraneous noise. The distributions are remarkably consistent from year to year. Due to their different survival characteristics and length distributions, they are always treated separately in modeling.

1.4 Results

1.4.1 Juvenile In-River Survival: Hatchery Release to Lower Granite

The predator susceptibility model fit to the two early years of the Sawtooth Hatchery data (1989 and 1991), which measures survival from the hatchery to Lower Granite Dam, displays clear sigmoidal ramps. Figure 1.6 shows the Predator Susceptibility Model fit to these data with year effects on k_0 and k_1 and shared values for l_c and σ_c. The joint critical length is $l_c = 127.2$ mm (95% confidence interval 119.7 to 135.2 mm) with $\sigma_c = 18.7$ mm (95% CI 11.6 to 20.4 mm). A complete table of parameter estimates with confidence intervals is in Appendix A. Survival increases from 11.2% to 25.5% in 1989 and from 5.1% to 11.1% in 1991. The σ_c term can be used as the half-width of the interval around the critical length in which the increase in survival is highest, indicating that fish in the Sawtooth data begin seeing gains in survival due to size at about 108.5 mm. The mean fish lengths (and standard deviations) for these years are 117.4 mm (13.5 mm) and 113.6 mm (9.6 mm), respectively. Thus the majority of smolts are large enough to see some reduced mortality due to their size, but 89.33% of the fish fall below the critical length.

Sawtooth Hatchery is 747 km from Lower Granite Dam, while the Rapid River Hatchery is only 283 km away, a difference in distance that is reflected in mortality rates. Survival to Lower Granite was 57.8% overall for Rapid River Hatchery, but only 11.15% for the early Sawtooth years. The Rapid River fish do not exhibit any critical size effect in their survival to Lower Granite: the PSM fails to converge while a GLM with fixed effects by year gives a significant positive coefficient for length overall, but the year effects of length combine with it in any given year resulting in slopes that are often not significantly different from 0. Figure 1.7 show the flat length-mortality relationships for the Rapid River fish by year. As mentioned in Section 1.3.1, the later years of Sawtooth data have higher lengths and survivals than 1989 and 1991 and resemble the Rapid River data in the length-survival relationship and also are not fit well by the Predator Susceptibility Model.

1.4.2 Adult SAR: Grande Ronde River

The Grand Ronde River data are a smaller and noisier data set. In most years, there were between 3,000 and 5,000 smolts PIT-tagged, but there are only 5 instances where more than 20 adults returned for a given out-migration-year:rear-type combination, eliminating the possibility of many individual year fits. These data reveal consistently positive correlations between survival and length, with the critical size appearing in the aggregated data. Results are summarized in Appendix A.2.

In some of the data, especially the Grande Ronde wild data but also in some Lower Granite subsets, there is a drop in survival for the very largest smolts. This is similar to a pattern observed by Koenings et al. (1993). In their work, observing the shapes of Loess-smoothed survival curves plotted against mean length from many different stocks they noted strongly diminishing returns in survival as length increased beyond a threshold, and eventually a decline. I hypothesized that this decline could be due to timing, i.e. that the biggest of fish might arrive later in the season, and temporal factors overwhelm the length factors in determining survival. This is not the case, however, as temporal distribution of smolt size is nearly identical for all but the smallest smolts—which were especially common toward the end of the season—this does not seem to be the case in the data set (see Figure 1.8).

1.4.3 Adult SAR: Passing Lower Granite

The largest and most complex data set addressed in this paper is of all the smolts that were PIT-tagged at Lower Granite Dam from 1998–2009. For each of the four type:transport subsets of the data the PSM fit best with temporally constant l_c and σ_c and fixed effects by year for k_0 and k_1. The parameter values with confidence intervals are shown in Table 1.5. The optimization calls in the maximum likelihood estimation did not fully converge when estimates for k_0 and k_1 were not significantly different from 0 (their lower bound), but examining the likelihood profiles shows that this did not adversely affect the estimates of the other parameters; only the confidence intervals for that particular term were affected. Figure 1.9 shows the PSM fit to 2008 with data for comparison, and Figure 1.10 shows the

model fits for each year.

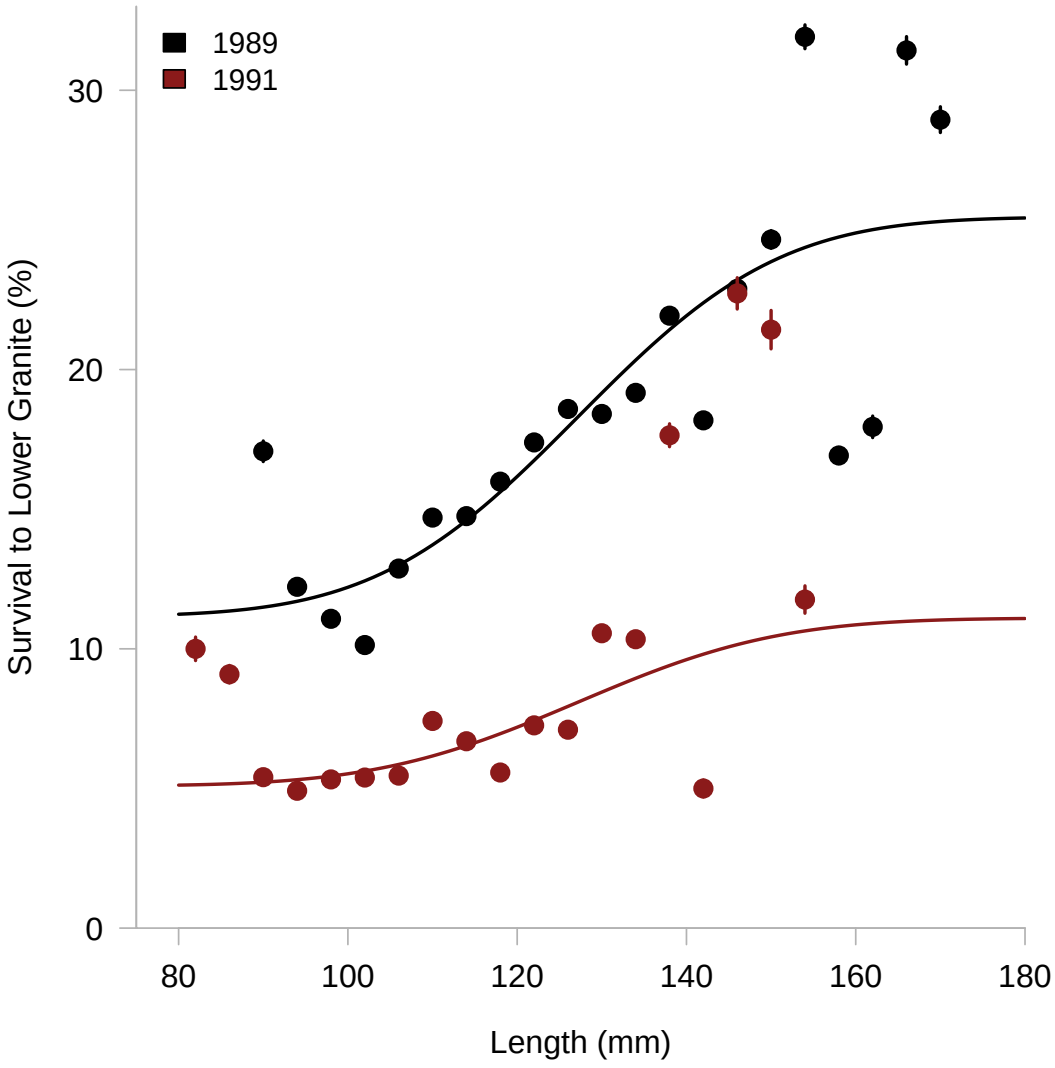

Figure 1.6: The two early Sawtooth years exhibit sigmoidal shape, with a ramp between low and high horizontal asymptotes. Using AIC, the Predator Susceptibility Model fits the data significantly better than a GLM with fixed year effects (ΔAIC $= -8.37$). Data are spring Chinook salmon smolts PIT-tagged at the Sawtooth Hatchery in 1989 and 1991, with survival measured to Lower Granite Dam plotted against length at tagging. Each point on the upper graph represents a bin, with the number of PIT-tagged fish in each bin indicated in the histogram below. The model was fit to individual, not binned data. Binomial 95% confidence intervals are shown as vertical bars.

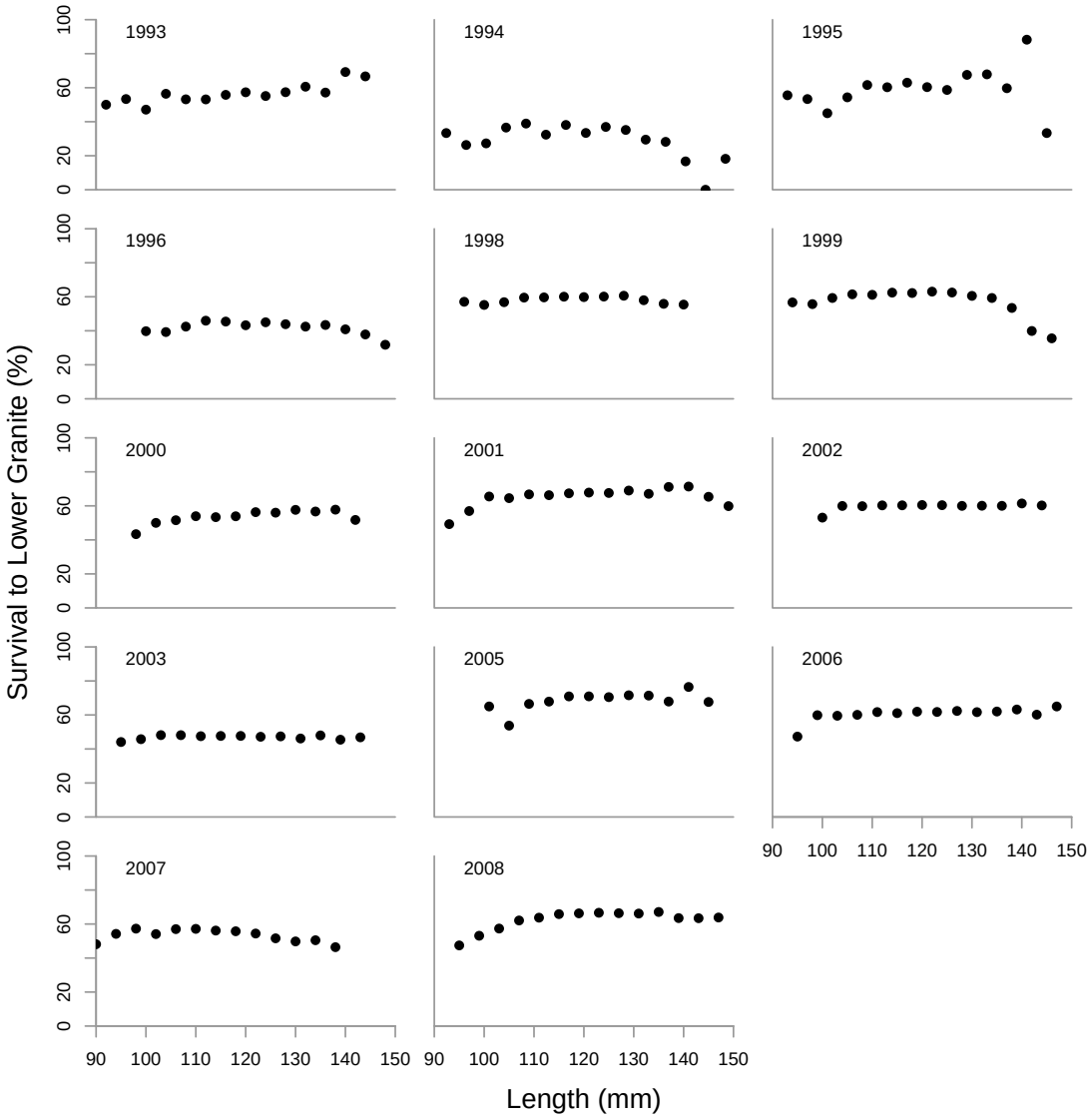

Figure 1.7: Survival to Lower Granite Dam versus length for the Rapid River data set. While a GLM fits a significant positive coefficient for length, there is no evidence of a critical length effect in these data.

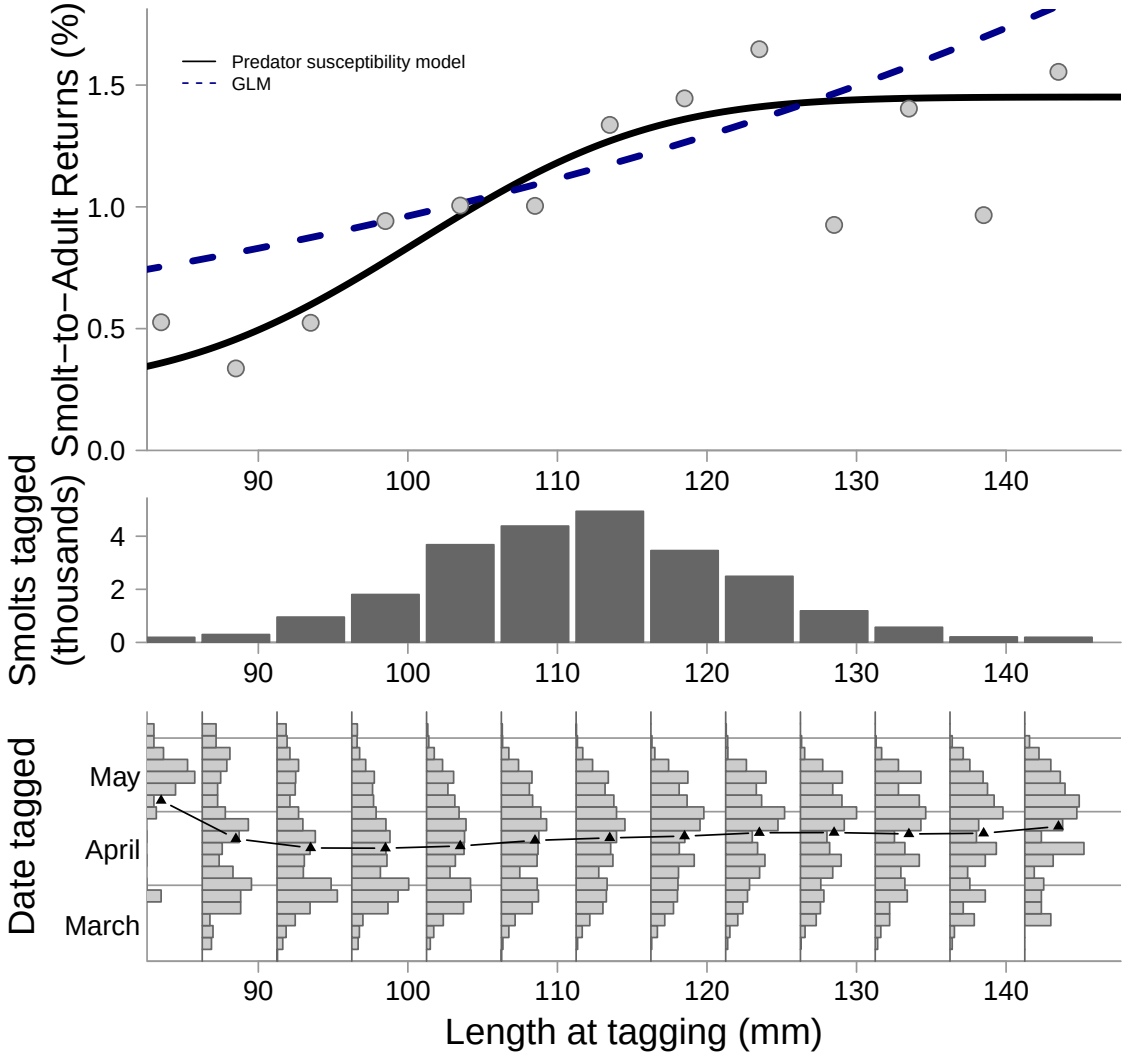

Figure 1.8: Wild fish PIT-tagged at the mouth of the Grande Ronde River (1998–2009), with the tagging date (the date they entered the Snake River) plotted as a histogram for each of the length bins. The mean tagging date for each length bin is marked with a triangle. The timing distribution changes minimally among the higher length bins, and timing parameters added to the GLM are not significant.

			l_c			σ_c			
Type	Transport	Mean length (mm)	$\hat{l_c}$	Lower 95% CI	Upper 95% CI	$\hat{\sigma_c}$	Lower 95% CI	Upper 95% CI	ΔAIC
Wild	Barge	110.5	107.6	106.3	109.0	5.6	4.2	7.6	−30
Wild	Run-of-river	109.1	103.1	101.0	105.2	8.2	5.9	11.5	−5
Hatchery	Barge	136.2	145.5	141.0	150.2	15.7	12.6	19.6	−37
Hatchery	Run-of-river	135.6	141.6	138.0	145.3	21.4	18.3	24.9	15

Table 1.5: Summary of PSM parameter estimates for the Lower Granite data. The mean lengths of the wild fish are slightly larger than the mean critical length estimates, falling outside the 95% confidence interval. The mean lengths of the hatchery fish fall below the l_c confidence interval. The ΔAIC column indicates compares the PSM with the best fitting GLM, with negative values indicating that the PSM is a better fit.

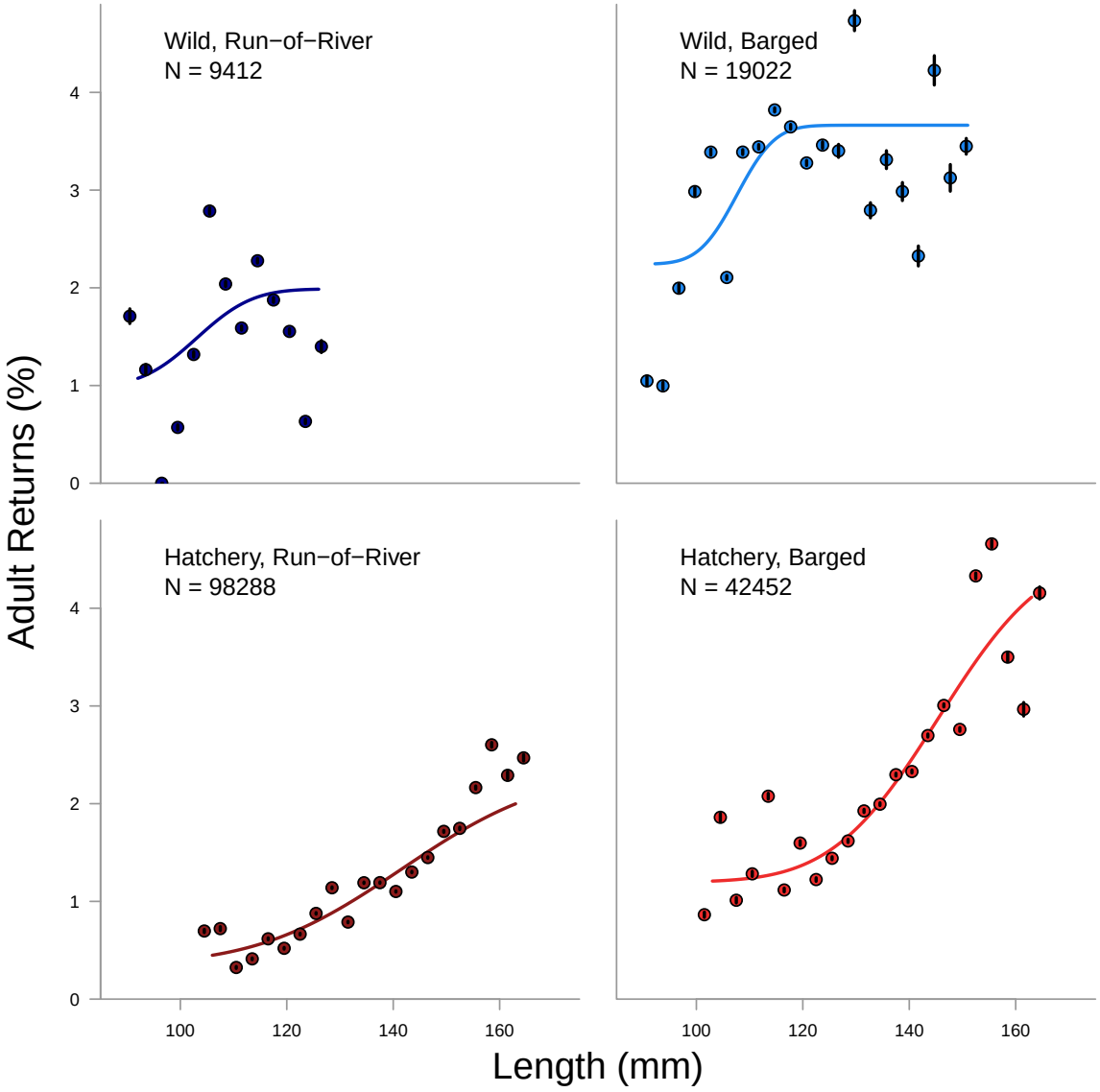

Figure 1.9: The Predator Susceptibility Model fits to the 2008 Lower Granite data. Fits for all years can be found in Figure 1.10. Here the model predictions are shown with data points (binned by length in 3 mm intervals). With smaller sample sizes, more noise is present for the wild fish, especially the wild run-of-river subset.

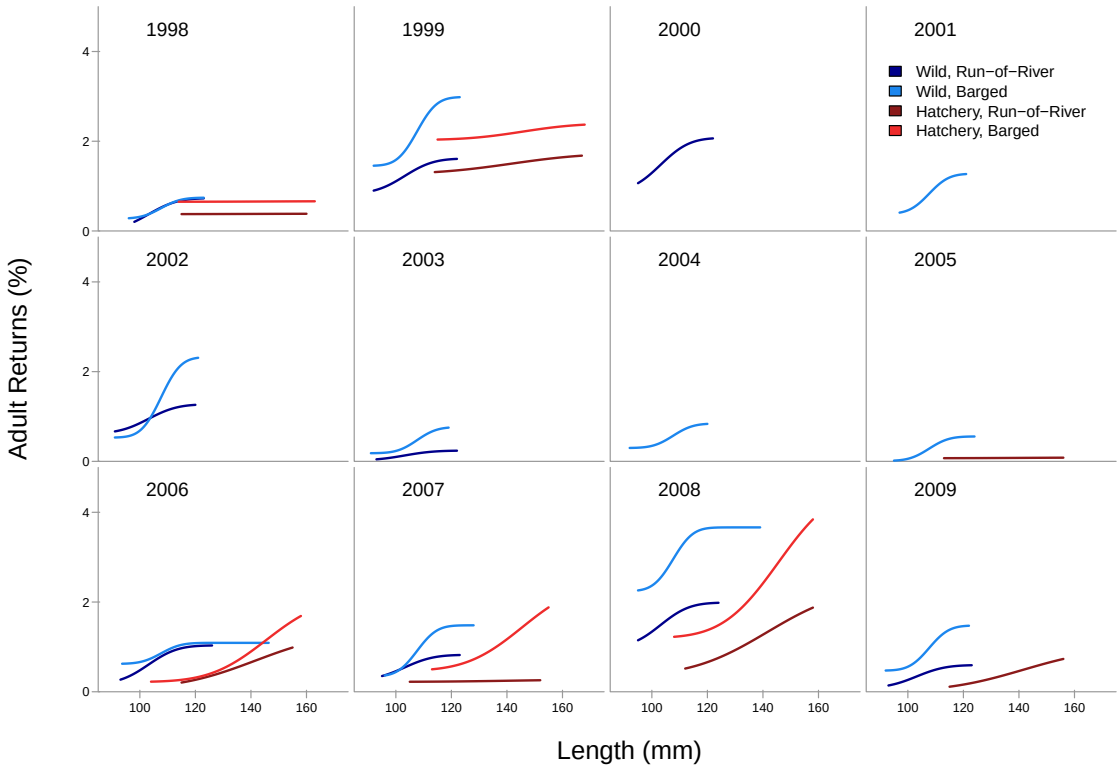

Figure 1.10: SAR curves fit by the Predator Susceptibility Model for fish PIT-tagged at Lower Granite Dam, by year, with colors indicating brood type and transport treatment. Horizontally the curves extend from the 2.5^{th} to 97.5^{th} data percentiles. The model is fit to each type:tranport interaction group with the asymptotic parameters k_0 and k_1 free to vary by year, but l_c and σ_c parameters fit constant within each group. Fit summaries are in Table 1.5

1.5 Discussion

The in-river survival data using fish from Sawtooth Fish Hatchery exhibits a clear sigmoidal pattern in survival as it relates to fork length in the early years. Of the data sets examined, this one has the fewest possible confounding factors: all the fish involved are from the same stock and the same hatchery, and no significant mortality occurred before they were released. Tracking survival to Lower Granite Dam avoids confounding factors of hydrosystem passage, the complexities of transport systems and the variability of ocean conditions. Mortality is due almost entirely to a single mechanism, predation, and there is strong evidence of a threshold effect in a juvenile's salmon susceptibility to predation. In the later Sawtooth years both rearing techniques and river conditions changed to significantly increase the sizes and survival rates of the Sawtooth fish. I hypothesize that in the later years of the Sawtooth data and in the Rapid River data mortality is too low for length effects to be observable.

The "critical-size, critical-period" hypothesis of Beamish & Mahnken (2001) identifies two stages where threshold effects on size and growth impact salmon marine mortality; the PSM shows that there can be a significant threshold effect prior to ocean entry. Benton et al. (2006) stated that "threshold effects abound in nonlinear systems and biological systems are typically nonlinear," thus it should come as no surprise that such effects are present in any life stage with significant mortality. These results could be combined with the critical-size critical-period hypothesis either by adding a stage for freshwater mortality prior to ocean entry or by redefining the beginning of the first stage to include freshwater predation in addition to marine predation. Regardless, the Sawtooth data demonstrate that critical size effects are very present during the freshwater phase of out-migration.

Adult return data is noisier than in-river survival and, while the signal of a threshold effect comes through, the response aggregates all-cause mortality across life stages. Restricting analysis to specific PIT-tagging locations eliminates some confounding factors, but the fish PIT-tagged at Lower Granite Dam or the mouth of the Grande Ronde River come from different stocks, they spawned at different locations, and the populations have undoubtedly experienced some mortality prior to tagging—mortality which the Sawtooth data suggests would have preferentially afflicted the smaller fish, thus the length distribution has been

altered. Despite these limitations, when enough data is present a sigmoidal SAR:length relationship is evident.

This research was inspired in large part as an alternative to basic linear models applied to survival data binned by length. Any statistical comparison between the model developed here, which follows the form of a 4-parameter sigmoidal curve, and these linear models is problematic: the linear models are not valid models when the data are not binned, as they will predict a survival probability less than 0 for fish that do occasionally survive. However, if the data are binned when a model is fit, then the result depends on *how* the data are binned. For survival data, it is preferable to fit models in a framework assuming a binomial error structure, such as the maximum likelihood approach used here. Many papers (e.g. Koenings et al., 1993) seek to compare stocks, but this analysis shows that how the lengths are distributed around the mean can impact survival estimates. The length distribution of the smolts has long tails. These data have been artificially truncated at 65 mm on the low end due to the size constraints of PIT-tags, but are otherwise intact and cover the length distribution of both hatchery and wild fish shown in Figure 1.5.

Zabel & Achord (2004) noted that average length does not help predict the survival of a population, but relative length is a valuable predictor of survival within a population. In their analysis of in-river survival, they did find year and site effects, and found length a better predictor of juvenile survival than a more complicated condition index, but they were using just a GLM. With the predator-susceptibility model, the *way* in which the response changes by year and site can be inferred from the parameter values. Especially dealing with in-river survival, where the variability of ocean conditions is absent, the asymptotes may be affected by a strong year effect, however I expect the threshold length to be relatively stable between years because the size distribution of the predator populations are probably stable, although it may vary by site. (Of course, if the predator length distributions are not stable, that would manifest in self in unstable critical length values.) I suggest this as a promising direction for future work using a model such as this one capable of dissecting a length effect into components and finding relationships between these length parameters and other factors affecting salmon survival.

Using the approach of Beckman et al. (1999), where SAR rates are regressed against mean

lengths by year, there is no statistically significant length effect. Using just the hatchery fish PIT-tagged at Lower Granite and omitting the years 2000–2004 when few fish were tagged, a linear model fit to SAR vs. mean length has $R^2 = 0.42$ and a marginally insignificant length coefficient with a p-value of 0.067. This is remarkably similar to the Beckman *et al.* results, which reported a correlation of 0.57 between SAR and length with a p-value of 0.06. They were mainly looking for a measure of smolt quality which would allow comparisons between hatcheries at different locations, however looking at the length distribution *within* a single location and year to reveal a strong SAR-length relationship.

There are two cases in which critical lengths are consistently absent: survival greater than 40% for in-river data (e.g. Rapid River Hatchery and the later years of Sawtooth Hatchery) and SAR less than 0.4% in the adult return data. In these extremes it is no surprise that a critical length effect is absent. When the mortality is low as in the Rapid River data set, it's possible that the critical length effect is emerging, but not strong enough to be evident as the number of fish at the tails of the length distribution is small, though their mortalities are essential for accurately observing a critical length. For adult returns, some years have extremely low survival due to various environmental factors, many of which are length-independent. When several length-independent effects coincide to create a year with exceptionally low survival, any length-dependent mortality is masked by massive overall mortality. Figure 1.12 shows an example of an extremely low survival year in the Lower Granite data. In these data, run-of-river hatchery fish are the most prone to experience extremely low survival, which encourages barging.

1.5.1 Effect of Barging on SAR

Even with the aid of barging, these results imply that length is a crucial factor in determining survival of Chinook salmon. However, the modeled response to barging suggests that barging is most effective on large smolts. The effects of barging on the model parameters are consistent by brood type: for both wild and hatchery stocks barging increases l_c by about 4 mm and increases σ_c by 2.6 mm for wild fish and 5.7 mm for hatchery fish. More interesting than the small changes in l_c and σ_c are the changes in k_0 and k_1. Barging increases SAR

for all lengths, thus both k_0 and k_1, but generally k_1 increases more than k_0 with barging, indicating that bigger fish realize greater benefits from barging. Looking back at the wild fish in Figure 1.10, 1998 and 2006 are exceptions, but in every other year with data (1999, 2002, 2003, 2007–2009) barging helps large wild fish to a greater extent than small wild fish. Though there are fewer years of data in which to compare hatchery raised run-of-river and barged fish, they may also experience the same effect.

There is a risk that timing has a confounding effect. Throughout the season the individual out-migrating juveniles grow rapidly and the environmental conditions in the river and estuary are changing. Growth does not appear to be a consistent confounding factor in the Lower Granite data. Pearson correlation coefficients between length and the day of year passing Lower Granite Dam are generally weak and sometimes negative, although there are 3 instances (2002 wild barged and run-of-river, 2008 hatchery barged) where the correlation exceeds 0.3. In comparing run-of-river to barged fish there is even less of an issue; the correlation of correlations is 0.888 for wild and 0.697 for hatchery stocks, which shows that the length:date correlations follow the same trends for the different transportation treatments. The correlation coefficients are available in Tables A.8 and A.9 in the appendix.

Unfortunately, I do not have survival data from Lower Granite Dam to Bonneville Dam for the run-of-river fish, so the differential delayed mortality coefficient, D (defined as the survival ratio of barged fish to run-of-river fish post-Bonneville Dam), cannot be computed (Anderson et al., 2011). Survival rates during barging are close to 100%, but Haeseker et al. (2012) show that in-river survival depends strongly on the average percent spill rates, but is generally between 40% and 70% for Chinook salmon. Applying these survival rates uniformly across the length distribution I can make a very rough estimate that D would be about 1 for the hatchery fish and larger wild fish, but that D would tend to be less than 1 for wild fish shorter than about 100 mm. The predator susceptibility model provides a foundation on which to build future work examining the size-dependency of D.

1.5.2 Translating into Overall Survival

The overall management goal is to maximize the number of smolts that return as adults, which requires looking beyond parameter estimates. The mean lengths of the Lower Granite data set are a little below the threshold lengths for hatchery fish and a little above the critical lengths for wild fish, but within σ_c of the critical length in both cases—not stuck in the asymptotes. For the Sawtooth data set (hatchery fish), the mean length is again falls about σ_c below l_c. This indicates that hatchery fish are just reaching the sizes where survival is most sensitive to length. Figure 1.13 shows hypothetical overall SAR rates for the 2008 wild barged smolts tagged at Lower Granite if their length distribution was shifted. Extreme predictions are probably very unreliable—and 20 mm increases in length are probably unattainable—but the model is believable in the neighborhood of the actual state. For these wild fish, the status quo is above the inflection point, but substantial gains still could be attained through size increases, while a similar graph for hatchery fish would have the status quo below the inflection point with even more potential for increased survival.

1.5.3 Conclusion

The fundamental result of this study is that there is a strong nonlinear connection between length and survival. Care should be taken when drawing conclusions about the presence or absence of a length effect; examining only means, small samples, samples not covering the length distribution, or comparing fish measured at different locations can obscure the length-survival relationship. The data are very noisy, and a critical size effect is not detectable in every subset of the data, but it is observed often enough that it cannot be discounted. Its identifiability is sensitive to mortality levels, sample size, and length range.

Also important is the result that a critical length effect can influence in-river survival of out-migrating spring Chinook salmon, in addition to the ocean-entry and first-winter critical lengths identified in other studies. In the early Sawtooth Hatchery data there is a clear sigmoidal relationship between length and survival. However, if the data were truncated the pattern would be lost. In the Rapid River data mortality is low, and no strong length-dependency is evident. At the other extreme, some of the Lower Granite subsets have

mortality so high that not enough survivors remain to make inferences, even with tens of thousands of fish tagged. However, a sigmoidal size effect is evident in the Lower Granite data as well. That mean lengths of hatchery fish are below critical lengths suggests that, to maximize survival, hatchery managers should continue to focus on maximizing size. With the majority of hatchery-reared smolts close to the low survival asymptote there is potential for relatively large gains in survival.

Further analyses treating in-river and marine survival separately, but using a model such as the predator susceptibility model that can isolate how survival patterns change under variable conditions, would bring a greater understanding of how the length effect changes based on stock, distance traveled in-river, river temperature, etc., which could then lead to specific management targets for the length of hatchery fish based on current conditions, habitat restoration for wild fish, and even length-selective barging.

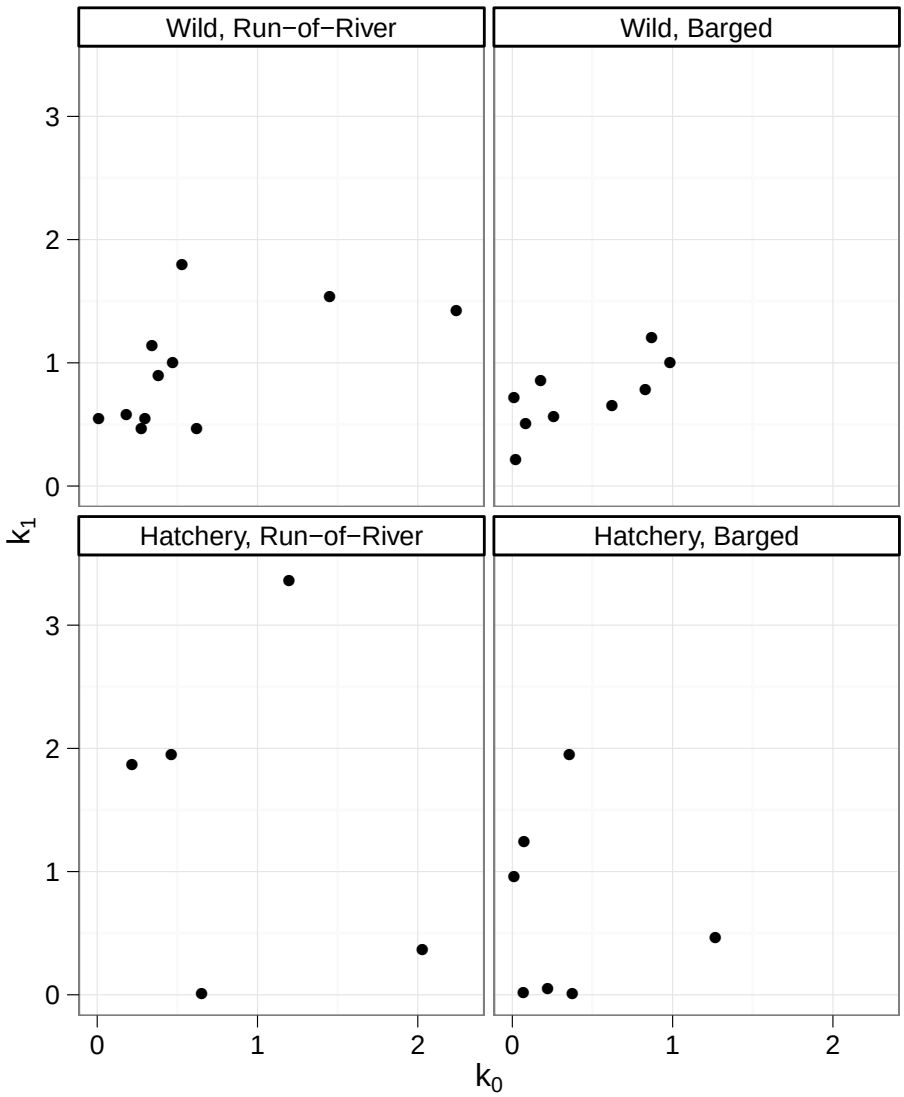

Figure 1.11: Yearly values of k_0 and k_1 fit to the Lower Granite data set. The range of k_1 (vertical scale) is wider for hatchery fish (bottom row) than for wild fish (top row).

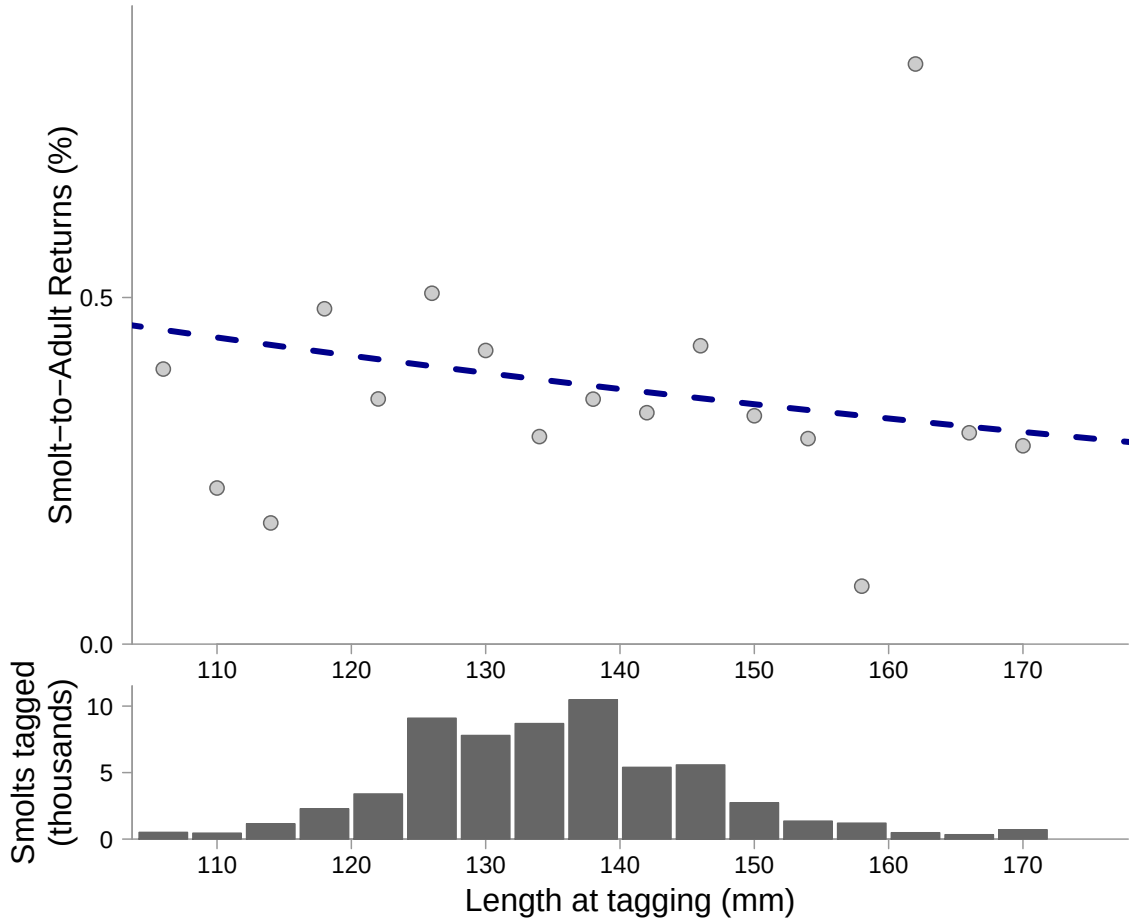

Figure 1.12: A length effect is absent for the run-of-river hatchery smolts PIT-tagged at Lower Granite in 1998. Of 38,802 fish tagged, there were 233 adult returns (0.38%), but a GLM (dotted line) fits a non-significant negative trend in survival by length. In this particular year mortality was exceptionally high; non-size-selective effects dominated the survival patterns. The histogram bars correspond to the points on the upper plot.

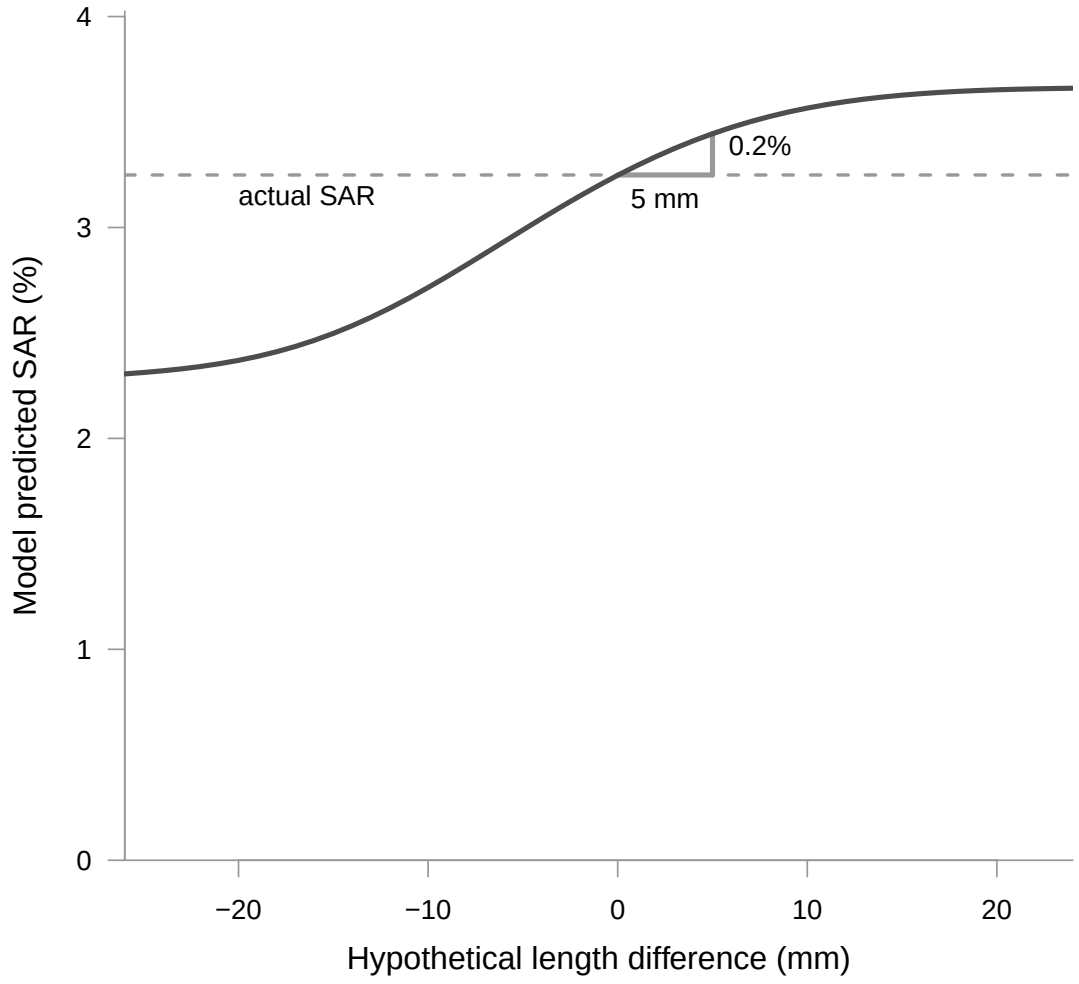

Figure 1.13: Using the Lower Granite fit of the PSM and the 2008 length data, this plot shows the predicted overall SAR for the wild barged smolts PIT-tagged at Lower Granite Dam in 2008 given hypothetical shifts in the length distribution. A 5 mm increase from the current length would, under the model, increase the overall survival from 3.25% to 3.45%.

Chapter 2

A VORONOI TESSELLATION-BASED APPROACH FOR GENERATING HYPOTHETICAL FOREST LANDSCAPES

Abstract

Optimization models used for forest management are highly complex and the demand for data to test them far exceeds the supply of real data sets, thus hypothetical test forest landscapes are often used. However, current hypothetical landscape generators are limited in their ability to match the characteristics of real forests and offer little or no control over important landscape metrics such as average vertex degree. This paper identifies four characteristic landscape statistics that could affect the efficiency of spatially explicit optimization algorithms and can be compared between simulated and real forest landscapes. Using these as guidelines, we describe a method for generating hypothetical landscapes capable of generating landscapes of specified characterization. A generated landscape is based on a Voronoi tessellation, created from points chosen by a mixture of random point processes, which is then edited to include gaps and non-convex polygons, overcoming the shortcomings of Voronoi tessellations in this application. Through a series of multiple regressions the algorithm determines appropriate control parameters so that the output landscape will have high probability of matching targeted characteristic statistics within a tolerance. The method produces landscapes with a wide range statistics, covering the characteristics of real forests and extending into extreme cases unlikely to be encountered in real data, while providing greater flexibility and control over the generated landscapes than previous methods for generating hypothetical forests.

2.1 Introduction

Forest managers often consult with spatially explicit optimization models to decide on a management strategy. Historically, timber production was the top priority, but Tóth &

McDill (2009) show that many objectives for forest use, including recreational value, carbon sequestration, wildlife habitat management and watershed protection can be considered jointly with timber production under budgetary and regulatory constraints to develop management schemes that balance all of these considerations and use the forest resources as efficiently as possible. Forests are typically divided into harvest management units to give forest managers small, well-defined regions on which different management actions can be carried out (Petroski, 2006); such actions could include diverse treatments such as thinning, cutting, doing nothing, or developing recreational facilities. Finding optimal management schemes requires the use of optimization models which develop management schedules for each management unit extending for a fixed management horizon, often close to a century.

Researchers classify models depending on the spatial constraints of the problems they are used to solve. The most common spatial constraint is one of maximum area for a given treatment in a time period, often used to address harvest scheduling in the presence of regulations limiting the maximum contiguous clear-cut area. These problems are divided into two main types. Unit Restriction Models (URM) are for problems where no two adjacent management units can be harvested in the same period because the total area of any two adjacent management units would exceed the maximum clear-cut. Area Restriction Models (ARM) form a generalization of URMs but include the general case when management units are small enough that it is possible for adjacent management units to be harvested in the same period. Thus, in an ARM the area of each management unit must be taken into account in determining whether a specific prescription is feasible (Murray, 1999). ARMs are more difficult to formulate efficiently and to solve. Other types of spatial restrictions are used for other types of problems. For example, the minimum patch size problem is applied to preserve wildlife habitat that must contain contiguous patches of a certain area (Rebain & McDill, 2003), and the connectivity problem (Önal & Briers, 2006; Conrad & Smith, 2012) sites selected for a wildlife reserve must also be connected by a network of protected land.

All of these models rely on complex systems of linear constraints to ensure that the spatial restrictions are met. The models are tested by running them on existing forests, sometimes with hypothetical data about the age structure of the forest, but appropriate data sources are often proprietary and are too few to provide adequate analysis. Ideally,

the models would be tested on data sets covering a wide variety of potential topologies and geometries to see under what conditions they perform well and, more importantly, to identify the circumstances that negatively affect the computational time for optimal solutions, which could be of great use for model selection and development. To make up for the lack of real data sets, researchers frequently test models on hypothetical forests (e.g. McDill & Braze, 2000; McDill et al., 2002; Rebain & McDill, 2003; Constantino et al., 2008; McNaughton & Ryan, 2008). Unfortunately, the tools used to produce these hypothetical forests do not provide explicit control over most of the output characteristics, thus they are limiting factors in the analysis of the optimization models.

For the purposes of these forest planning models, a forest landscape is defined as a set of polygons, representing the management units in a forest. In a review of existing generators, Li et al. (2010) divided the existing algorithms by method. The two most promising candidates were based on random graphs and Voronoi diagrams, but in their current states none of the existing methods meet the goals outlined above. Specifically, the ability to vary the output characteristics in a controlled manner is absent. An overview of the relative merits of these methods and one other method follows.

Optimization models "see" a landscape as a table of adjacencies and a list of areas and other pertinent information corresponding to each management unit, thus it is natural to think of a landscape as a graph where management units correspond to nodes and adjacencies are indicated by edges, with a vector of areas mapped in correspondence with the nodes. Associating other characteristics with nodes (such as age class or distance to a road) is a trivial addition. Under this system, every possible landscape corresponds to a planar graph with an area mapping. The generation of random graphs has been a topic of significant mathematical research. Erdős & Rényi (1960) analyzed the properties of random graphs created by generating specifying a degree (number of nodes) and then selecting a certain number of edges uniformly from the set of all possible edges. Unfortunately for this application, their results show that it is unlikely for purely random graphs that have more than half as many edges as nodes to be planar. This low upper bound on the number of adjacencies for planarity would be a challenge for following a random graph based approach for generating hypothetical forests.

Constantino et al. (2008) generated hypothetical forests in a deterministic way by using as building blocks two specific graphs, a 4-node cycle denoted as an "F-instance" and a more complicated 10-node graph called a "G-instance". Grids of F- and G-instances are connected to each other, with each node assigned unit area. This method gives good control over the number of management units, but is in no way random. Topologically there is little to differentiate hypothetical landscapes of equal sizes constructed by this method. Many implementations of the ARM and URM problems improve efficiency or solution quality by supplementing the model formulation with information on specific subgraphs, like cycles or cliques of certain sizes (e.g. Goycoolea et al., 2005). In a case like this, it is theoretically possible that a model implementation could work unusually efficiently or inefficiently on a certain local topology, so when hypothetical test forests are composed entirely of repeating subgraphs, evaluation of the implementation could be biased. To control for such possibilities in testing a model some degree of randomness should be present in generated data.

The MakeLand program of McDill & Braze (2000) operates by first generating a set of random nodes and then connecting each node to a given number of its nearest neighbors. These initial nodes are placed randomly with an inhibition parameter, a minimum separation distance between points. To create planarity, the algorithm deletes intersecting arcs until no lines cross, always deleting the arc that has the most intersections per unit length. The faces of the graph are taken as the forest management units, and arcs continue to be deleted until the number of faces matches the user input number of management units, always deleting the longest arc of the smallest face. This approach offers control over the resulting adjacencies through two parameters: the inhibition parameter governing point placement and the connection parameter giving the number of arcs initially drawn per point. The authors noted that a high inhibition parameter "tends to create maps with a more even distribution of polygon sizes." Thus the algorithm allows for some control over the area distribution, but the authors did not comment on the degree distribution of their results.

Another approach to generating planar maps uses Voronoi tessellations or diagrams. Voronoi diagrams (also called Thiessen polygons) are defined based on a set of points, where a polygon for each point is created enclosing all the area that is closer to it than to any other point. Voronoi diagrams have wide-ranging applications, such as behavioral ecology,

image compression, and cell biology (Du et al., 1999), and algorithms for computing Voronoi diagrams are readily available. Barrett (1997) applied Voronoi tessellations to a similar forest management unit-related problem: automatically generating management units from a map of a real, undivided forest. Wyszomirski & Weiner (2009) used Voronoi diagrams to look at crowding and competition in plants and found that the area distribution of the polygons was dependent on the level of clustering in the initial points. As Li et al. (2010) noted, landscapes generated from Voronoi diagrams have polygons that are always convex. Additionally, the corners of the Voronoi polygons are formed by the intersection of 3 boundaries, whereas corners of degree 4 (as would be formed by two intersection lines) are common in landscapes that have been divided into management units by humans. However, I will show that these limitations can be overcome in the initial point placement and in post-processing the resulting hypothetical landscape.

This paper describes the development of an algorithm, Rlandscape, that improves on the existing methods of hypothetical landscape generation. We identify four landscape characteristics important for the computational complexity of harvest scheduling models and present a landscape generating procedure that uses Voronoi tessellations and allows the analyst to have tight control over these parameters. By drawing the initial points from a mixture of four random distributions and editing the resulting diagram, all of the known weaknesses of the Voronoi diagram method are eliminated. The range of the characteristics of output landscapes is shown to more than cover the range of characteristics observed in real landscapes and those generated by MakeLand, and the ability to target characteristic ranges is shown to work efficiently enough to quickly produce a large number of landscapes with specified characteristics.

2.2 Methods

We start with a description of metrics we will use to characterize a landscape, then detail the point processes used to create Voronoi tessellations and post-processing to correct for the aforementioned weaknesses. Finally we describe the modeling performed to relate input parameters and output characteristics through the stochastic landscape creation process.

2.2.1 Choosing metrics

The number of polygons, area distribution, and degree distribution are the primary landscape characteristics that have been shown to affect the computational complexity of harvest scheduling models (Tóth et al., 2012; McDill & Braze, 2000; Constantino et al., 2008). The number of polygons in a landscape, n, requires no additional explanation. Presented below are the metrics used to describe the area and degree distributions. Taken together, all of these metrics define what we are calling the *characteristics* of a given landscape.

Area

For a hypothetical landscape, total area (or equivalently, average area of a polygon given the number of polygons) can be set *post-hoc* without affecting the geometry of the landscape because the distance units used are arbitrary. However, the distribution of area among the polygons is important. The difference between the ARM and URM models introduced above is the size of management units relative to the maximum allowable contiguous harvest area (Murray, 1999), and the solution methods of the much more difficult ARM problems usually involve groupings of small polygons (e.g. McNaughton & Ryan, 2008; Önal & Briers, 2006; Goycoolea et al., 2005). We need to consider a measure of the spread of the distribution of areas. Standard deviation is a poor choice, because in a hypothetical landscape its units are as arbitrary as the units of area. The mean polygon area, and thus the standard deviation as well, will vary significantly with both the number of polygons and the area of the total landscape, which would prevent comparisons between landscapes, thus a relative measure is required. The coefficient of variation (CV, defined as standard deviation divided by mean times 100%) is an appropriate measure of the spread of the area distribution. The CV is a standardized standard deviation, and will allow for comparisons between forest landscapes, both real and hypothetical, at any scale.

Degree

Spatially explicit optimization models consider adjacencies as well as areas, thus the degree distribution also should be part of the landscape characteristics. Recent work has shown

that the distribution of the degree of each polygon, i.e. how many polygons it is adjacent to, can be significant in evaluating the potential for a strengthening procedure, which modifies the constraints used and often outperforms other methods (Tóth et al., 2012). This is similar to work done on even broader generalizations of these optimization problems which uses degree extensively in bounding complexity and simplifying problem formulation (e.g. Berman & Fürer, 1994). Unlike area, the degree distribution is scale-invariant, thus the standard measures of a distribution, mean and standard deviation, μ_d and σ_d, are appropriate. In spatially constrained harvest scheduling algorithms, both *weak* and *strong* notions of adjacency are used. Two polygons are *weakly adjacent* if they share a finite number of points, and are *strongly adjacent* when they share an infinite number of points, i.e. they have a common border (Goycoolea et al., 2005). Strong adjacency is the type considered here. By viewing adjacency in this manner, any landscape is equivalent to planar graph, where each node corresponds to a polygon, and edges between nodes correspond to a shared border. Bollobás (1998) gives a proof that a planar graph with $n > 3$ vertices, the maximum number of edges is $e = 3n - 6$. Since each edge connects two vertices, the maximum average degree $\max(\mu_d) = 2(3n-6)/n = 6(n-2)/n$, which approaches 6 in the limit as n approaches infinity. The lower bound on μ_d is dependent on how connected the landscape is. Technically, μ_d would be 0 if each polygon was completely isolated, but it approaches 2 for connected graphs. Our method does not enforce connectivity (as many real forests contain disconnected pieces), but extremely disconnected landscapes are not complex spatially, so in general landscapes with $\mu_d < 2.5$ are not of interest. Individual polygons, of course, may have degree equal to any non-negative integer.

These four landscape statistics, n, CV, μ_d and σ_d compose the landscape characteristics; they are readily calculated from typical forest data: adjacency and area tables. As a reference point, we calculated statistics for real forests with adjacency tables or area tables posted as public data on the University of New Brunswick's Integrated Forest Management Lab website (www.unb.ca/fredericton/forestry/research/ifmlab/index.html). Some of the landscapes' degree distributions were omitted because they used soft adjacency, making μ_d and σ_d incomparable.

2.2.2 Generating a landscape

This section describes how a landscape is generated. First, points are chosen, then a Voronoi diagram is created from the points, and finally the diagram is edited. Voronoi diagrams are well established, the original work here is on the point processes used to pick the initial points and on the edits performed on the diagram; these two steps are detailed below.

Point processes

Voronoi diagrams depend entirely on the points used to create them, thus the selection of initial points is fundamental to the creation of a hypothetical landscape. A *random point process* is an algorithm for creating a set of random points. Many such algorithms exist with various properties (see Baddeley & Turner, 2005). We settled on 4 point processes to serve as the basis for landscape creation. Others were tried, but found to be redundant or too unpredictable. The baseline point process is the random uniform process, where x and y coordinates are determined by independent random draws to uniform distributions covering the feasible x and y ranges. This offers no control over point placement other than the bounds. The only control parameter for this method is the number of points to place.

An inhibition process, where points are placed uniformly randomly but never within a certain (*inhibition distance*) from another point, produces a landscape with regularly sized polygons. Higher inhibition distance results in more evenly spaced points which create a more regular landscape, lowering CV. Setting the inhibition distance to 0 makes this method equivalent to the random uniform process, which allows for clustering without actively creating clusters. This algorithm is commonly known as the *simple sequential inhibition* (SSI) point process (Baddeley & Turner, 2005). (McDill & Braze (2000) used a similar point process in MakeLand.)

To achieve the opposite extreme, landscapes with high area CV, we introduced a clustering process. Initial trials used a Thomas clustering point process, where "parent" locations are chosen uniformly randomly, and then a random number of "child" points are created around each parent, with their displacement from the parent points determined by a bivariate Gaussian distribution. This introduced an unwanted level of variability in n, as the number of

both parent and child points were determined by random draws from a Poisson distribution. Instead, we adapted the method by making the number of points per cluster an explicit parameter and calculating an appropriate number of clusters based on the total number of points to place, but preserving the Gaussian displacement from parent points.

When comparing the polygons created from these three point processes with the management unit divisions in real forests, the lack of right angles was striking. The *lattice grid* method was designed specifically to allow for right-angled intersections of polygon borders. In a Voronoi diagram based on points with coordinates drawn from continuous distributions, all border corners will be adjacent to 3 polygons, and there will be no weakly adjacent polygons. The only way for the corner of a border to be adjacent to 4 (or more) polygons is if it equidistant from the four closest chosen points, an event which has probability 0 in any point process with continuous probability density. However, in actual landscapes with borders chosen by humans, right-angle intersections and weakly-adjacent polygons are not uncommon. Restricting point coordinates to a regular discrete scale, orthogonal border intersections and weak adjacencies become possible. The lattice grid method overlays a grid over the feasible area, and the samples points from the lattice points of that grid. The numbers of horizontal and vertical grid lines are control parameters. In addition to right angles, this method tends to produce horizontal and vertical borders, reminiscent of how a forester might have divided a forest section into management units in the absence of natural breaks. This point process works in conjunction with other processes to mimic a landscape that has some natural breaks in management units as well as some imposed artificial boundaries.

Landscapes can be created using any combination of the point processes; Figure 2.1 shows examples of Voronoi tessellations resulting from points used by each of the point processes individually. However, using mixtures of these point processes allows for greater flexibility in the characteristics of the output landscape.

Editing the Voronoi tessellation

Landscapes produced by Voronoi tessellations have certain properties not described directly by the characteristic statistics that could be different real landscapes. Right angles and weak adjacency are imposed by using the lattice point process, but additional differences remain: the Voronoi landscapes are completely connected with no gaps or holes, and the polygons are strictly convex. We resolve these two issues by editing the polygons of the Voronoi tessellation.

Polygon deletion creates holes in the landscape and opens the possibility of a disconnected forest, which occurs in real forests when a river or road cuts through the region. The final patchiness of the landscape is given by a control parameter, p_H, equal to the proportion of the polygons in the final landscape that are classified as holes. (Polygons that are deleted to become holes in the final landscape are omitted when calculating other landscape statistics.) Polygons are selected for deletion randomly with no preference given for size or adjacencies. When a polygon is deleted all of its neighbors' degrees decrease by 1, effectively lowering the degree mean. The effects on σ_d and CV are negligible.

We address the convexity issue in a similar way: by deleting borders between two polygons and merging them into one. As in polygon deletion, a control parameter p_M is defined as the proportion of polygons in the final landscape that are the product of a merge event. Borders are selected for deletion uniformly among all possibilities, which then merges the polygons they had been separating. Merging is more complicated that simply deleting polygons, because it's possible that more edges than those selected must be deleted. For example, if polygon A is merged with polygon B, and B with C, the adjacencies must be checked so that if there is a border between A and C it is also be deleted. Both deleting and merging reduces the number of polygons, thus the number of points chosen initially is inflated from the target n so that the expected value of n for the final landscape matches the target. However, the nature of the merging prevents precise knowledge of the output n as we cannot tell *a priori* (continuing the above example) whether the border between A and C will also be selected for deletion or if it must be deleted afterwards in the clean-up phase. Thus merging introduces some randomness in the final n.

2.2.3 Meeting constraints

The previous section described how to create landscapes; this section describes how inputs are chosen to generated landscapes with characteristics that fall within a specified range. The control parameters relating to point processes are the number of points to be placed by each method, n_{unif}, n_{lat}, n_{ssi}, and n_{clust}, the number of horizontal and vertical gridlines for the lattice method, the inhibition distance for the SSI method, and the number of points per cluster and spread term for the cluster method. Additionally, to create a landscape two more control parameters are introduced: p_H and p_M. These control parameters are the inputs to the landscape creating algorithm. But, the algorithm is highly stochastic, and we want to be able to create landscapes with specific characteristics.

Some of the relationships between control parameters and landscape characteristics can be readily intuited. For example, when a polygon is deleted its neighbors' degrees all decrease by 1, so μ_d will also decrease. This is illustrated in Figure 2.2. Which point processes are used has a strong effect on the area CV. Pairs plots, such as the one in Figure 2.3 show the relationships between several variables simultaneously and hint at relationships that can then be explored in models.

We considered three options for determining appropriate control parameters for landscape characteristics. Two involve modeling explicit relationships between the two sets of variables, using linear models or generalized additive models (GAMs). The third way is to estimate the distributions of control parameters for specific landscape characteristics through a Markov Chain Monte Carlo procedure without making any assumptions or inferences about the relationships between the two. Of the two modeling approaches, linear models are simpler than GAMs, which would only be necessary if the linear models proved inadequate. The distribution-based approach has one major drawback: to provide good estimates it would rely on repeatedly estimating multivariate distributions, which would depend on a large amount of data, resulting in a computationally expensive algorithm. In the modeling approach, the parameter estimation is done once up front, subsequent evaluations of the models at certain points is as easy as evaluating an expression. With these considerations, we decided to start with the linear model approach.

Picking parameters

The first control parameters to estimate are the number of points to place with each process, which for generality will be expressed as proportions of the total number of points placed, p_{unif}, p_{lat}, p_{clust}, and p_{SSI}. Of the characteristics, CV is most closely related to the type of point process as clustering increases CV and inhibition decreases CV. To avoid confounding, defaults were set for the inhibition distance and the lattice control parameters

Based on 500 simulations, the random uniform point process produces landscapes of mean CV 56.1 with standard deviation 5.8 when no other point processes are used and moderate merge and hole proportions of 0.1 are applied. We use this mean as a spline point in the regressions on whether cluster points are added to increase the CV or lattice and SSI points added to lower the CV. The additional control parameters for these point processes, that is the numbers of horizontal and vertical segments for the lattice grid, the inhibition distance for the SSI and the cluster radius are related to the distance metric as well, and thus dependent on the number of polygons and the dimensions of the entire forest. Of even greater concern is when the bounds depend on distance specific concerns. For example, if the inhibition distance is set too high, it will be impossible to place enough points, and number of possible points placed by the lattice method is clearly limited by the total number of lattice points. To avoid confounding, in these two cases we selected default values for the point process-related control parameters, excepting the number of points placed by each method. The inhibition distance is set at $0.639\sqrt{a/n_{SSI}}$, where a is the horizontal to vertical aspect ratio of the entire landscape and n_{SSI} is the number of points to be placed by the SSI method. In 100 simulations of the point process, this was the (approximate) largest distance that resulted in all the points being placed. For the lattice method, the number of horizontal and vertical grid lines, v and h respectively, are set to defaults given by $h = \lceil \sqrt{n_{lat}/a} \rceil$ and $v = \lceil \sqrt{n_{lat} \cdot a} \rceil$, the minimum values such that $hv \geq n_{lat}$ with the number of horizontal and vertical segments corresponding to the aspect ratio of the landscape. With these defaults, rather than regressing on the control parameters with potential bounding problems, the regressions can focus on the mixing parameters p_{lat} and p_{SSI}, the proportions of points placed by each method, with the rest of the points being placed by the uniform method. In

the clustering process, regressions on initially tended to increase the number of points per cluster to absurdly high levels when the CV was higher than average (often on the order of 50 points per cluster). To prevent this, the default number of points per cluster is set to 5, with the spread along with p_{clust} to be estimated based on characteristics.

Simulations were then run (50,000 for each non-uniform point process) generating landscapes with randomly varying proportions of points placed by each process and p_H and p_M. These data formed the bases for sequential regressions were to find each necessary model parameter based on all of the specified landscape characteristics. We selected models using step-wise model selection based on Bayes Information Criterion (BIC) beginning with full models. Further evaluating included model parameters based on their significance using t-tests, we found an appropriate model for each parameter. Box-Cox transformations were applied to the response when applicable. An example model (for p_{lat}) is presented below, see Appendix B for all of the models. The proportion point process parameters are the first estimated parameters, so they depend only on the characteristics, however once they are estimated they are added in as predictors for estimating the remaining parameters.

$$\begin{aligned} p_{lat}^{2.6} = {} & \beta_0 + \beta_1 \mathrm{CV} + \beta_2 \mu_d + \beta_3 \sigma_d + \beta_4 a + \beta_5 \mathrm{CV}\mu_d + \beta_6 \mathrm{CV}\sigma_d + \\ & \beta_7 \mu_d \sigma_d + \beta_8 \mathrm{CV} a + \beta_9 \mu_d a + \beta_{10} \sigma_d a + \beta_{11} \mu_d \sigma_d \mathrm{CV} + \\ & \beta_{12} \mathrm{CV}\mu_d a + \beta_1 3 \mathrm{CV}\sigma_d a + \beta_{14} \mu_d \sigma_d a + \beta_{15} \mathrm{CV}\mu_d \sigma_d a. \end{aligned} \tag{2.1}$$

The β's from the above model are all significant at the 0.001 level, and the 2.6 power transformation of p_{lat} was determined by a Box-Cox test. This particular regression is unique in that the aspect ratio is significant; it is insignificant for all of the other parameters. This is understandable because the geometry of the lattice grid depends on the aspect ratio, while all the other point processes ignore it. The next sequential regression, for p_H, adds p_{lat} to the pool of predictors. We apply a similar series of multiple regressions to determine appropriate parameters for the high CV case.

2.2.4 Mass-producing landscapes

Mass landscape creation with specific targets is handled by two algorithms illustrated in Figure 2.4. Both are implemented in R (R Development Core Team, 2011) and follow the

naming convention of an "r" prefix to denote random generating function. Rlandscape generates a single landscape from control parameters following the steps in Section 2.2.2. The second algorithm, rland, takes as inputs the number of landscapes to be produced and the acceptable bounds for their characteristics. It then picks specific targets for each landscape to be generated and uses the regressions from Section 2.2.3 to select control parameters for these targets. It then generates landscapes using Rlandscape and tests to see if their characteristics are within bounds, and repeats this process until enough landscapes have been generated.

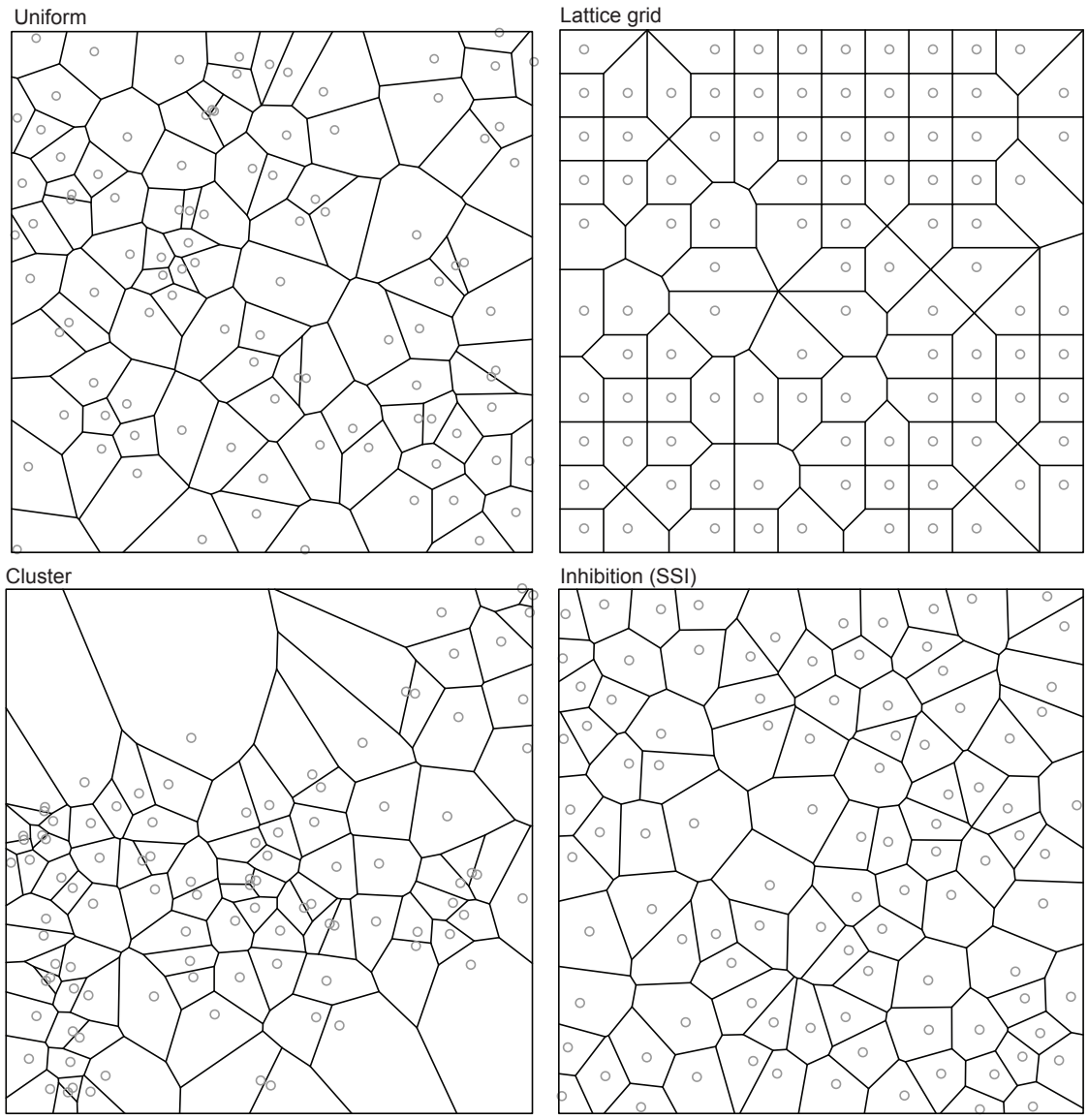

Figure 2.1: Sample Voronoi tessellations of 100 polygons created using each of the four point processes. The locations of the initial points are marked with gray circles. Visually, the difference in area distributions is readily apparent: the lattice and inhibition methods produce regularly sized cells, the uniform method produces more variation, and the clustering method very high area variation. Also evident are the right angles and intersections of more than three borders produced exclusively by the lattice grid method.

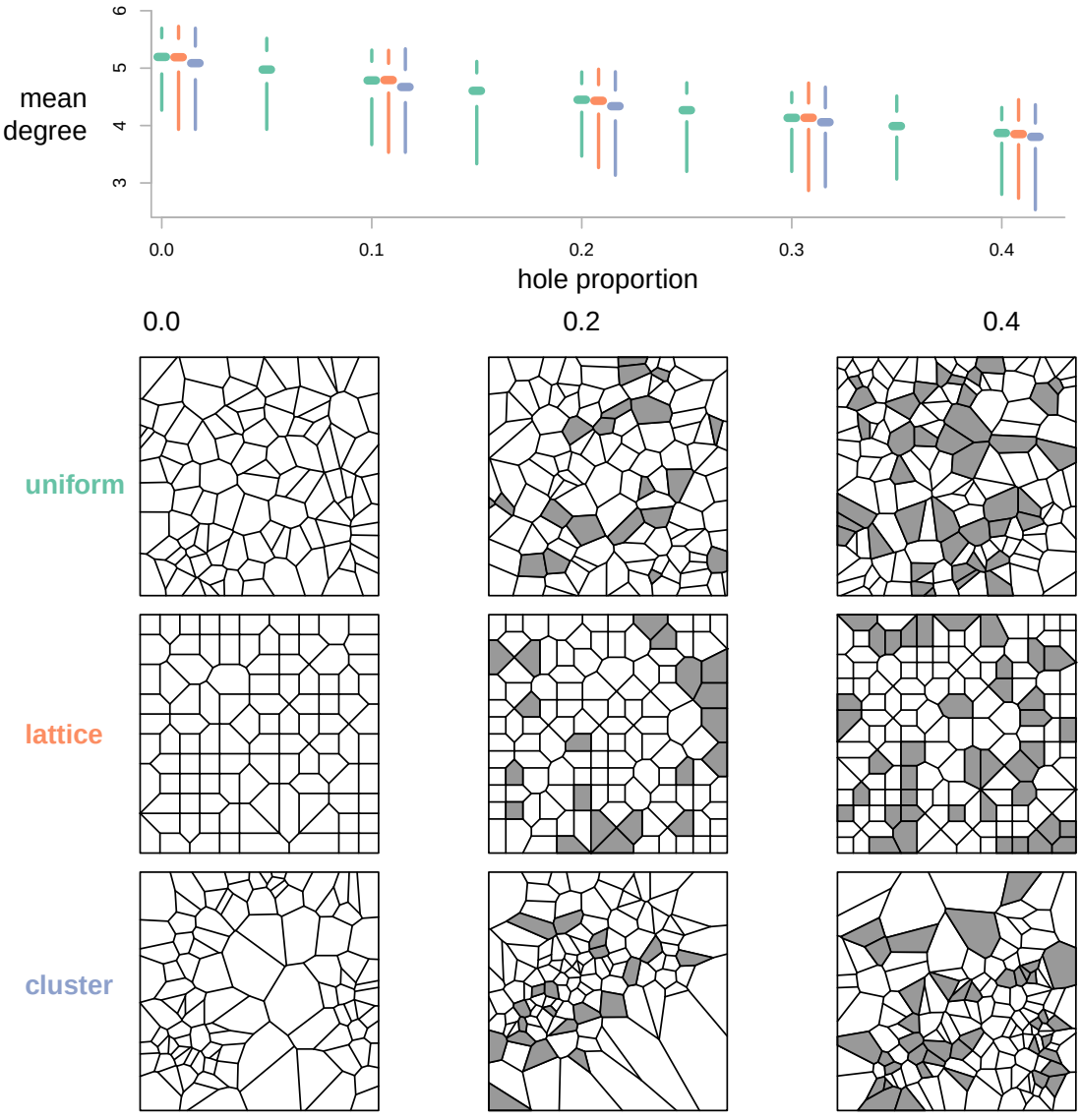

Figure 2.2: This figure shows how hole proportion can lower the degree mean of landscapes. Degree means (μ_d) from approximately 15,000 simulations are plotted in the top graph using Tufte-style boxplots (the mean is indicated by the horizontal bar, vertical lines extend from quartile to max or min), corresponding by color to sample landscapes by each of three of the point processes.

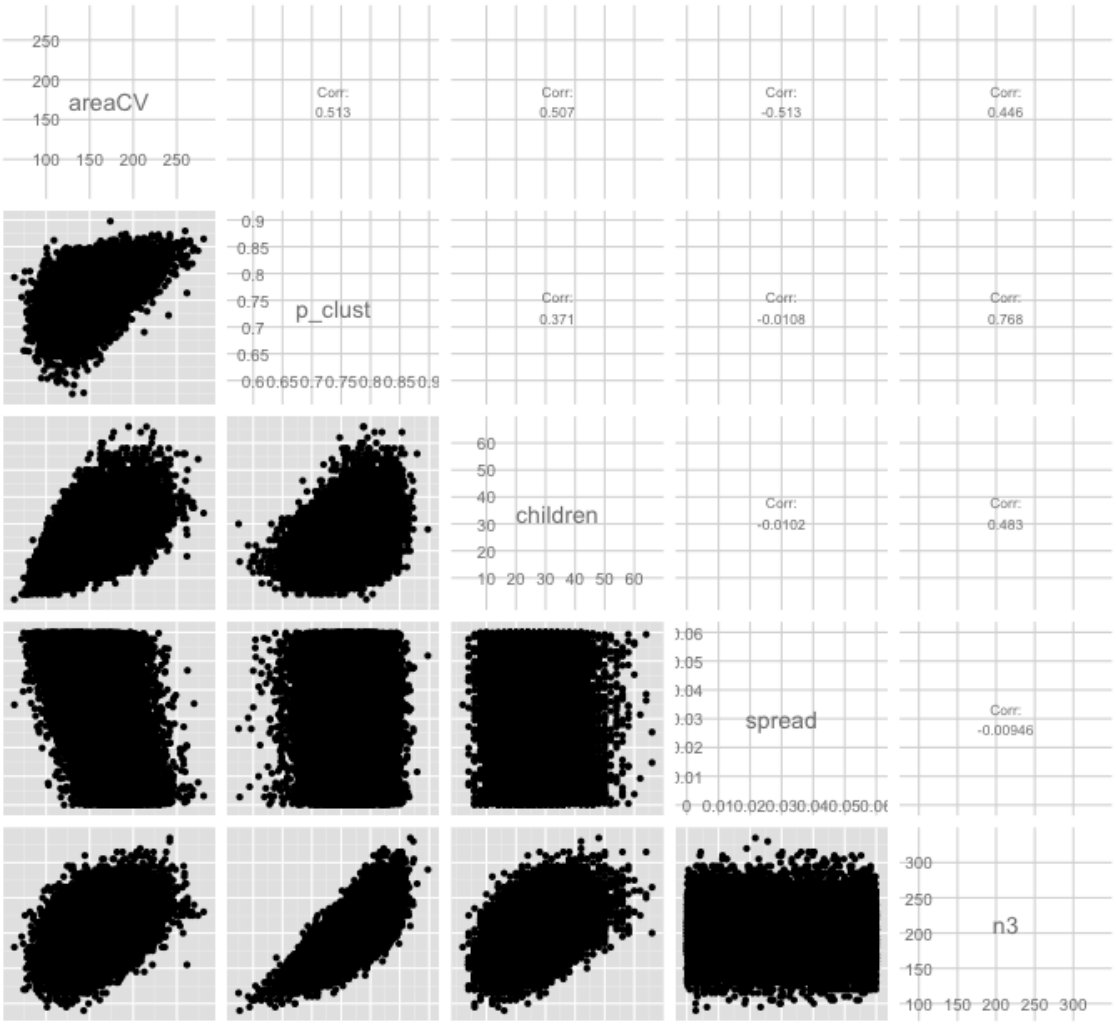

Figure 2.3: An example of a pairs plot, showing the relationship between pairs of variables. This one illustrates that the area CV statistic (examining the left column of plots) is strongly related to the cluster parameters. Correlation coefficients are displayed in the upper right half of the figure.

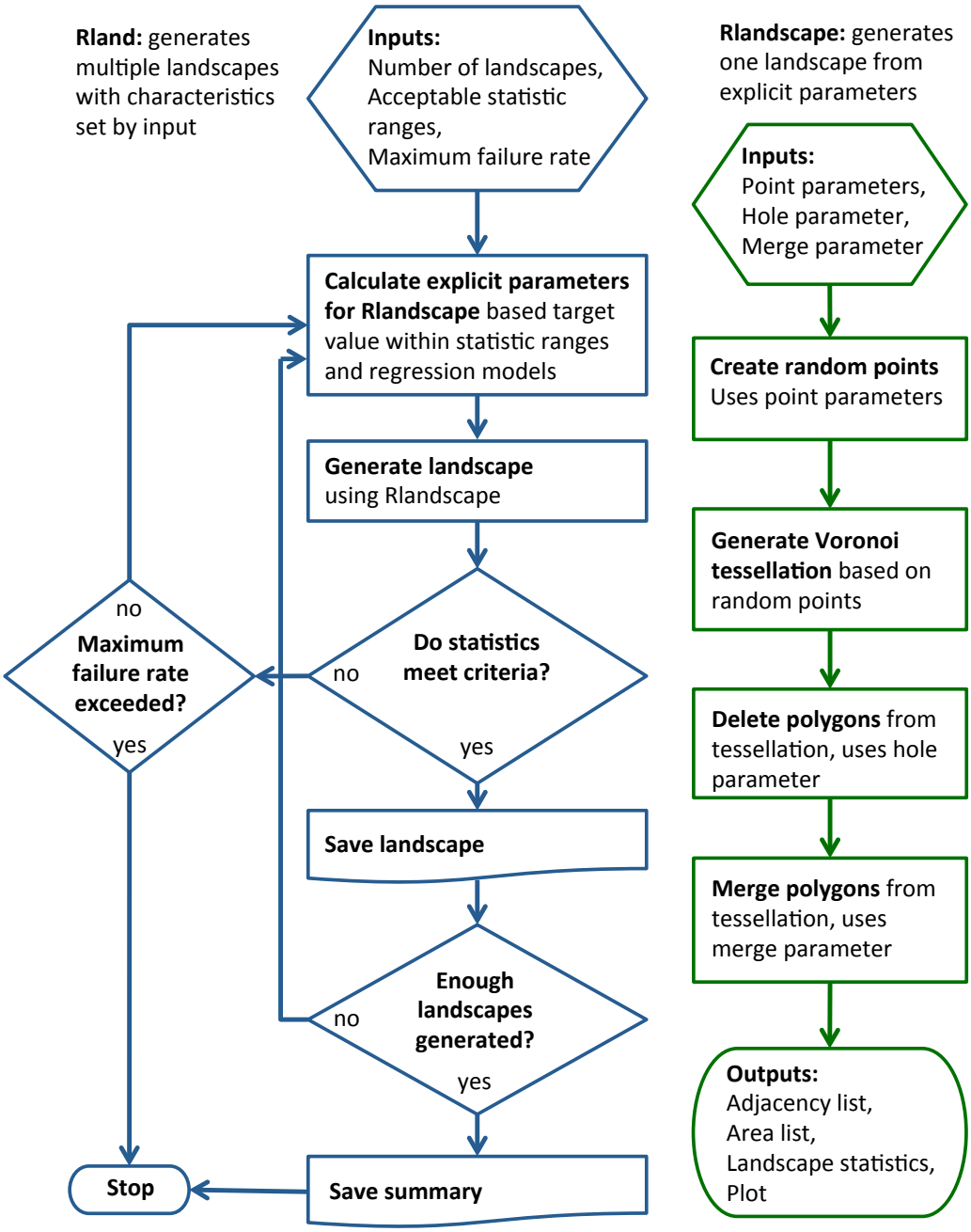

Figure 2.4: This flowchart shows the approach taken to produce hypothetical forest landscapes with statistics that meet input criteria. The right column (in green) shows the flow of the Rlandscape program, which generates a single landscape by editing a Voronoi tessellation created from points chosen using a mixture of random point processes. The blue flow chart shows how the wrapper program, Rland, estimates parameters for Rlandscape based on input landscape statistics, tests the output for compliance, and saves a summary of the results when the run is complete (or aborts if a maximum allowable failure rate is exceeded).

2.3 Results and Discussion

2.3.1 Range

To test if Rlandscape is capable of producing hypothetical landscapes with similar characteristics to actual forests, we calculated characteristic statistics for landscapes with publicly posted adjacencies and areas on the University of New Brunswick's Integrated Forest Management Lab's website (ifmlab.for.unb.ca/fmos/datasets/). We then simulated 500 landscapes using rlandscape with randomly varying parameters. The results are shown in Figure 2.5, along with the characteristics of 20 landscapes produced by MakeLand. The figure makes clear that Rlandscape exceeds the ranges seen in these real forests, whereas MakeLand is constrained in both area CV and σ_d.

2.3.2 Timing

Figure 2.6 shows the relationship between run time and the number of polygons in a landscape. Run time increases approximately linearly with n (at least for $n < 700$)with a large part of the variation explained by p_H and p_M. This is because they effectively increase the number of polygons that must be generated; e.g. a 100-polygon landscape with no holes or merges is 100 polygons pre- and post-edit, but if $p_H = 0.5$ than 200 polygons must be generated so that 100 remain after 50% are deleted in the edit step. As implemented, we have successfully created landscapes of up to 10,000 polygons (see Figure 2.7). The algorithm failed at 100,000 due to a "evaluation nested too deeply: infinite recursion" error. Capable of producing a 1,000-polygon landscape in under 10 seconds on 2.4 GHz Intel Core 2 Duo Macintosh with 6 GB of RAM, Rlandscape works efficiently enough to be a practical landscape generator.

2.3.3 Accuracy

There is significant stochasticity built into Rlandscape; Figure 2.8 shows the variation in characteristics of landscapes generated using the exact same control parameters, however the regression methods are effective in selecting control parameters to produce acceptable landscapes. To illustrate this, Figure 2.9 shows the number of trials and time needed to produce 500 landscapes under a variety of specifications. For common characteristic values

overlapping with observations of real forests, the efficiencies tend to be above 50%. At the extreme values of the ranges there is higher variability, which results in lower success rates. This is particularly apparent in the case of high CV, however the steep increases is time needed occur beyond the range of CV exhibited in real landscapes. High values of μ_d require many tries, but as p_H is very low in these cases the run time remains relatively short. Applying multiple constraints compounds the efficiency reductions, typically resulting in efficiencies of between 10% and 20%. The parameter picking algorithm is currently calibrated for producing intermediate values efficiently, however it could be modified to evaluate the input constraints and change methods if they are above certain thresholds.

2.3.4 Conclusions

This project accomplished two goals: to create a method for producing varied, realistic forest landscapes, and to target the landscapes so that landscapes of specific characteristics can be produced. The success of the first goal is demonstrated in Figure 2.5, which shows that Rlandscape output well covers the ranges of characteristics observed in real forests. The benchmarking shown in Figure 2.9 show that Rlandscape is viable for generating data sets of any feasible set of characteristics.

It's our hope that Rlandscape will prove to be a useful tool with a variety of applications. The addition of other point processes and modifications or the tracking of additional statistics could expand Rlandscape's already flexible framework. It would also be possible to modify the Voronoi diagram creation, for example Okabe et al. (2008) formulated several generalized Voronoi diagrams which employ inward/outward distances and additively/multiplicatively weighted shortest path distances. Additional data can be easily simulated with or without spatial correlation on top of an Rlandscape output to mimic other landscape features.

A number of modifications could be made to these algorithms to enhance their utility. Incorporating support for soft adjacency constraints, for example, would be a minor modification that would increase the set of models that the output landscapes could test. Other modifications could alter the polygon deletion and merging to better mimic real forests or to improve fine control of landscape characteristics. We view the current, purely

random implementation of these processes as an advantage, spatially correlating deletion probabilities could simulate discontinuities present in actual forests, like rivers and roads. Alternatively, the deletions and merges could target polygons that are outliers in degree or area to reduce variability (or the opposite to increase variability) which would improve efficiency and precision in producing landscapes of targeted characteristics at the expense of randomness. Rland does not do well with extreme values of parameters. For example, some real forests are divided into equally sized stands such that area CV= 0. The only way for Rlandscape to mimic this behavior is to exclusively use the lattice point process, which then drastically reduces the flexibility of the characteristic space. As a way around this, areas could be assigned *post-hoc*, using only the adjacencies from the generated landscape. No visual representation of the landscape would be available, but it would be feasible. The current regression methods for choosing control parameters are sufficient, but using GAMs might yield more robust fits resulting in higher efficiency in producing landscapes.

Harvest scheduling models require data on more than the layout of the stands. At a minimum, information about the state of the forest in each stand (e.g. age class) is necessary, which Rlandscape does not provide. Currently, we recommend that researchers assign age classes and any other information as they see fit. Similarly, if a specific area distribution is required for a set of trials (for example all stands having equal area) we encourage researchers to use Rlandscape to generate the adjacencies and ignore the provided area list replacing it with one that meets their own qualifications. The graphic representation of the landscape will then be inaccurate, but the output landscape could be deformed into a topologically equivalent landscape with the new area distribution, so the adjacency/area combination remains feasible.

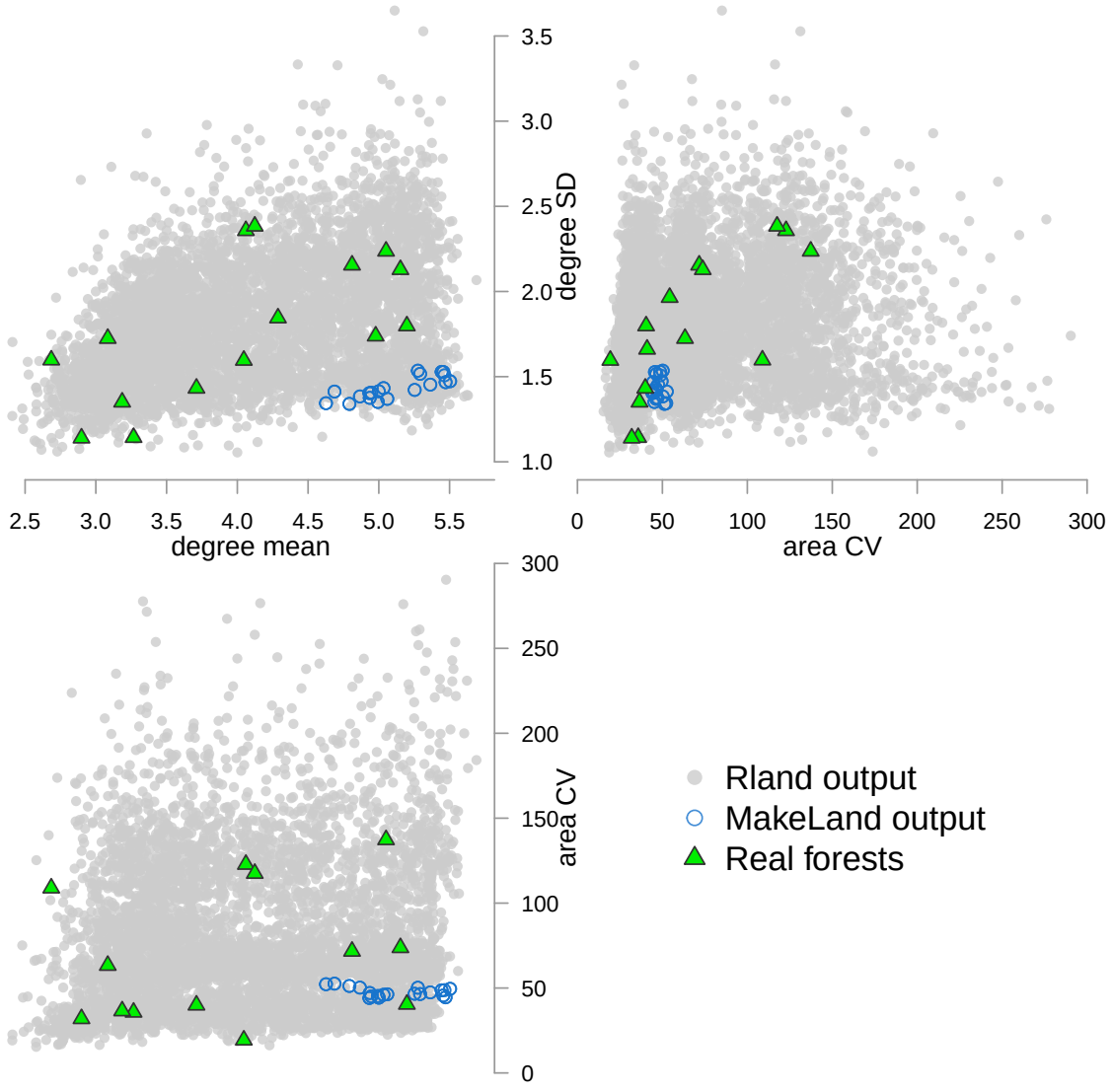

Figure 2.5: Landscape statistics from 17 real forests (green triangles), 20 landscapes generated by MakeLand (blue circles), and 500 landscapes generated by Rlandscape with targets varying over the range of the real forest statistics. Rlandscape more than covers the characteristic space of the real forests. MakeLand's output has somewhat varied μ_d, correlated with the small variation in σ_d, however in this output the CV is static. The real forests are those with publicly posted adjacencies and areas on the University of New Brunswick's Integrated Forest Management Lab (ifmlab.for.unb.ca/fmos/datasets/).

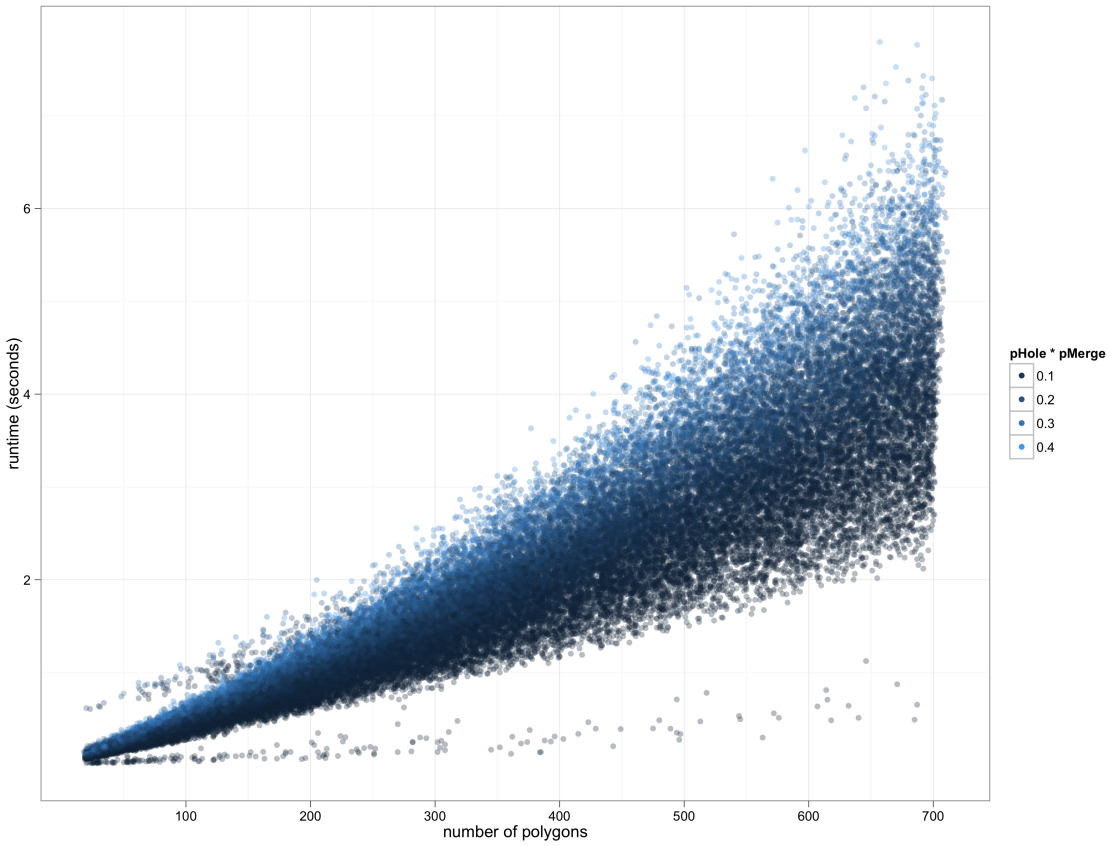

Figure 2.6: Run time in seconds to produce a single landscape is plotted against the total number of polygons in that landscape, with color showing the product of the hole proportion and merge proportion terms for about 45,000 simulations. Both the number of polygons and deleting and merging parameters are positively correlated with the run time.

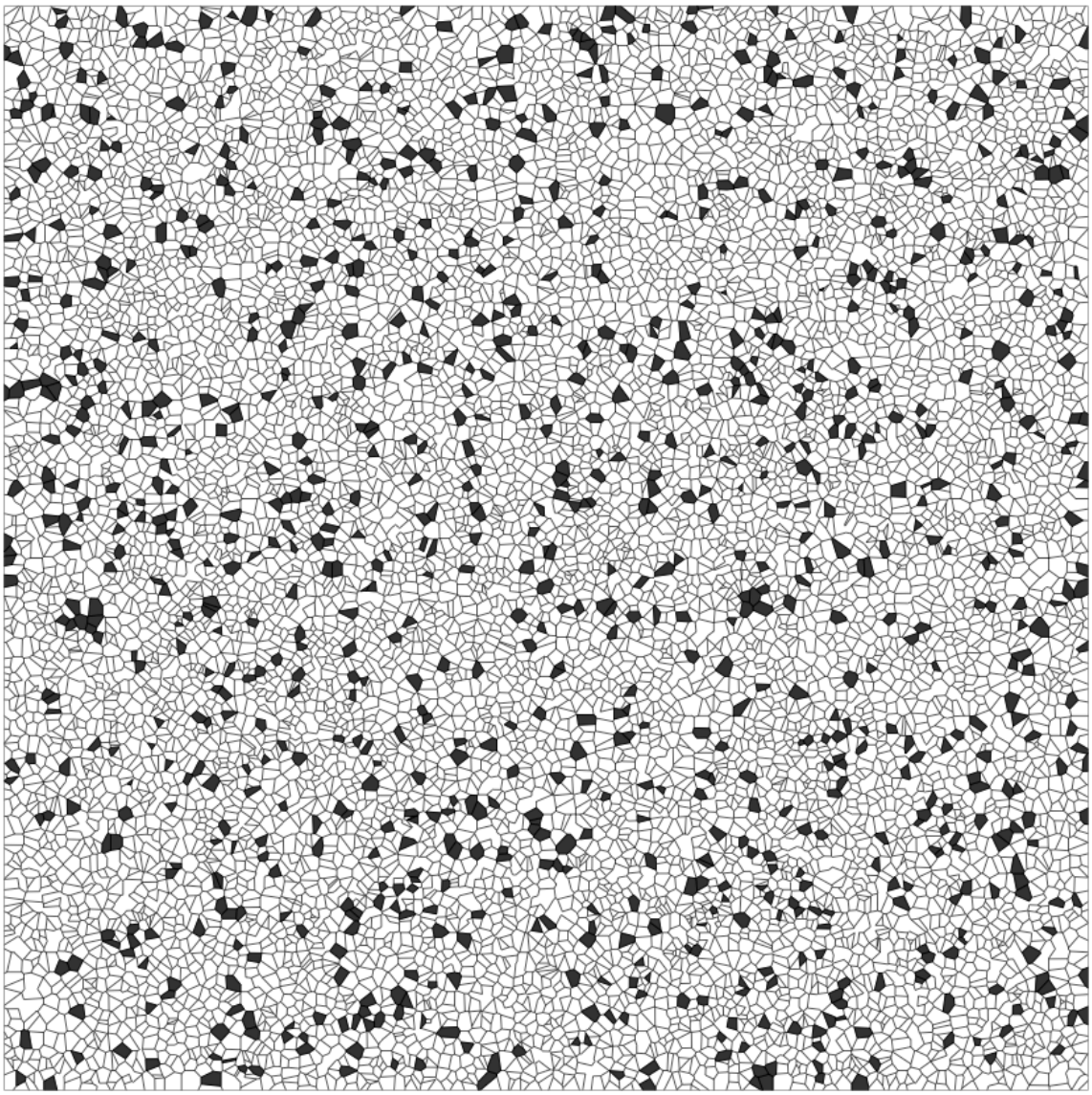

Figure 2.7: A landscape with 10,000 polygons, 25% placed by each point process. For this landscape, $\mu_d = 5.4$, $\sigma_d = 1.6$, and CV = 42.9. This landscape took 199.485 seconds to generate.

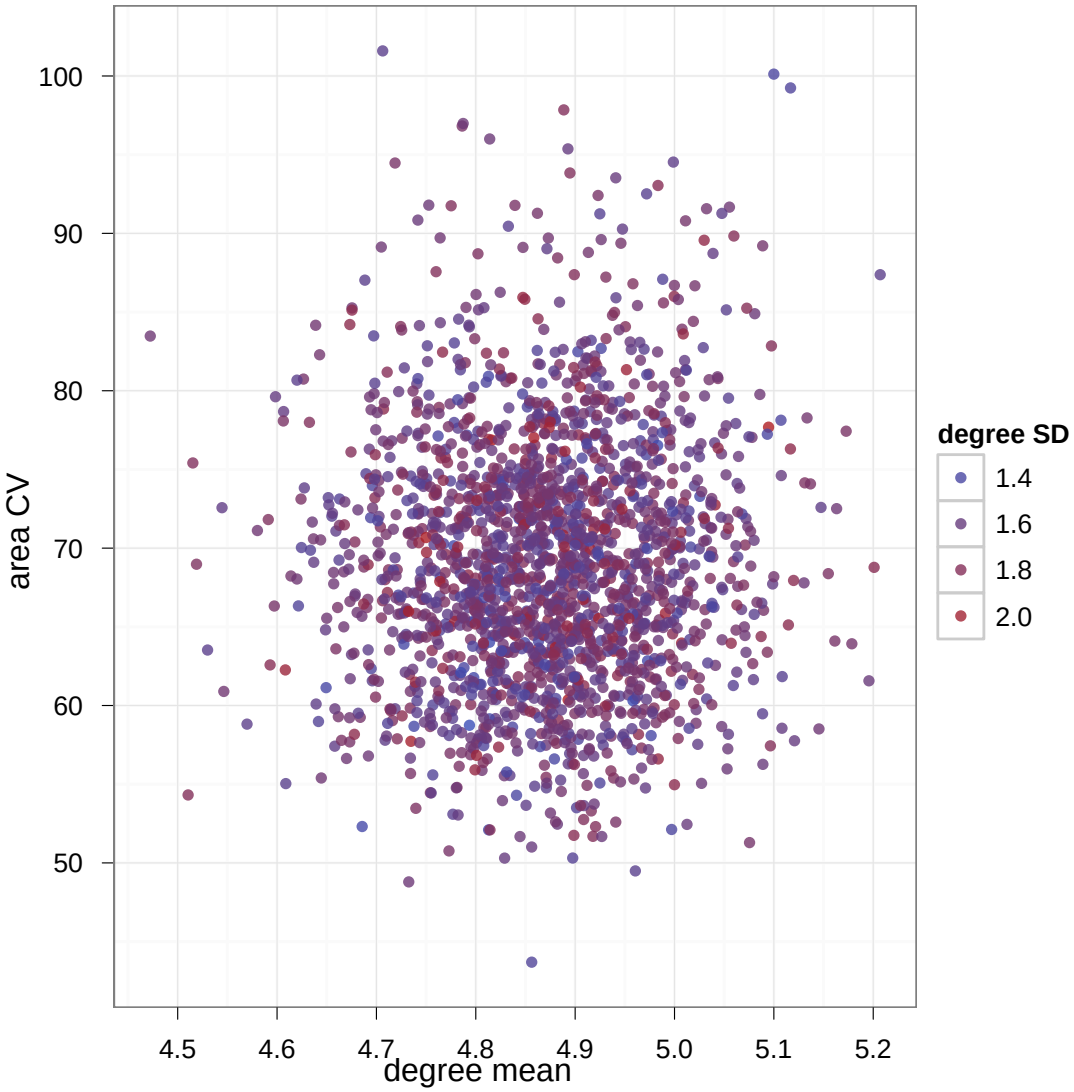

Figure 2.8: The resulting statistics of 2000 landscapes of 100 polygons each created by giving Rlandscape the same input parameters. Degree mean, on the horizontal axis, varies from about 4.5 to 5.2, though the total feasible range is from about 3 to 6. Area CV on the vertical axis and degree standard deviation indicated by color also vary significantly, though all three are roughly normally distributed. Not shown is the output n: of the 2000 trials 1958 have $n = 100$, and the range is from 97 to 101.

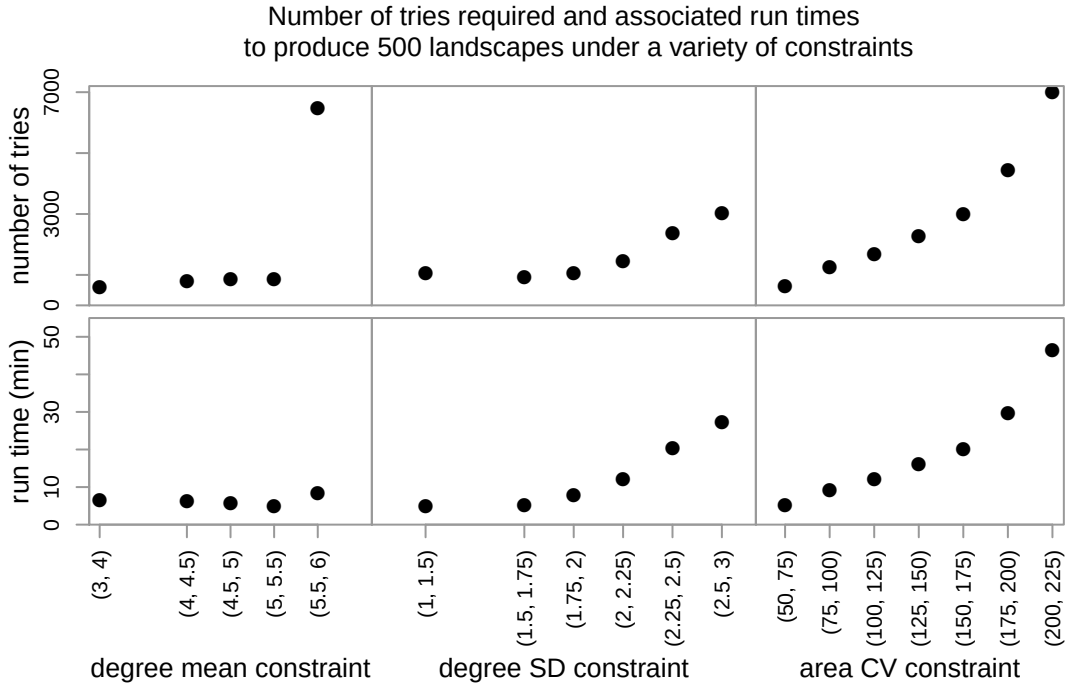

Figure 2.9: This shows the results of testing the time and number of tries necessary for Rland to produce 500 landscapes under varying constraints. The number of polygons in these landscapes was held between 95 and 105, and other constraints were set as indicated on the horizontal axis. Only one landscape statistic was imposed at a time; when μ_d was constrained, σ_d and CV were allowed to vary freely.

BIBLIOGRAPHY

(1999). *PIT Tag Marking Procedures Manual*. Columbia Basin Fish and Wildlife Authority PIT Tag Steering Committee, 2nd ed.

Anderson, J. J., Gildea, M. C. & Williams, D. W. (2006). Linking environmental and population hetergeneity to salmon survival through the stochastic evolution of vitality. Unpublished.

Anderson, J. J., Gosselin, J. L. & Mui (2010). Differential post-hydrosystem mortality of transported and in-river passing spring Chinook from the Snake River. Unpublished.

Anderson, J. J., Gurarie, E. & Zabel, R. W. (2005). Mean free-path length theory of predator-prey interactions: Application to juvenile salmon migration. *Ecological Modelling* 186 196–211.

Anderson, J. J., Ham, K. D. & Gosselin, J. L. (2011). Snake River Basin differential delayed mortality synthesis. Tech. rep., U.S. Army Corps of Engineers.

Anderson, J. T. (1988). A review of size dependent survival during pre-recruit stages of fishes in relation to recruitment. *Journal of Northwest Atlantic Fisheries Science* 8 55–56.

Arendt, J. D. (1997). Adaptive intrinsic growth rates: An integration across taxa. *The Quarterly Review of Biology* 72 149–177.

Baddeley, A. & Turner, R. (2005). Spatstat: an *R* package for analyzing spatial point patterns. *Journal of Statistical Software* 12 1–42.

Barrett, T. M. (1997). Voronoi tesselation methods to delineate harvest units for spatial forest planning. *Canadian Journal of Forest Resources* 27 903–930.

Beamesderfer, R., Ward, D. L. & Nigro, A. A. (1996). Evaluation of the biological

basis for a predator control program on northern squawfish (*Ptychocheilus oregonensis*) in the Columbia and Snake rivers. *Can. J. Fish. Aquat. Sci.* 53 2898–2908.

Beamesderfer, R. C. & Rieman, B. E. (1991). Abundance and distribution of northern squawfish, walleyes, and smallmouth bass in John Day Reservoir, Columbia River. *Transactions of the American Fisheries Society* 120 439–447.

Beamish, R. J. & Mahnken, C. (2001). A critical size and period hypothesis to explain natural regulation of salmon abundance and the linkage to climate and climate change. *Progress in Oceanography* 49 423–437.

Beamish, R. J., Thomson, B. L. & McFarlane, G. A. (1992). Spiny dogfish predation on Chinook and coho salmon and the potential effects on hatchery-produced salmon. *Transactions of the American Fisheries Society* 121 444–455.

Beckman, B. R., Dickhoff, W. W., Zaugg, W. S., Sharpe, C., Hirtzel, S., Schrock, R., Larse, D. A., Ewing, R. D., Palmisano, A., Schreck, C. B. & Mahnken, C. V. W. (1999). Growth, smoltification, and smolt-to-adult return of spring Chinook salmon from hatcheries on the Deschutes River, Oregon. *Transactions of the American Fisheries Society* 128 1125–1150.

Beckman, B. R., Larsen, D. A., Lee-Pawlak, B. & Dickhoff, W. W. (1998). Relation of fish size and growth rate to migration of spring Chinook salmon smolts. *North American Journal of Fisheries Management* 18 537–546.

Benton, T. G., Plaistow, S. J. & Coulson, T. N. (2006). Complex population dynamics and complex causation: Devils, details and demography. *Proceedings: Biological Sciences* 273 1173–1181.

Berman, P. & Fürer, M. (1994). Approximating maximum independent set in bounded degree graphs. In *Proceedings of the fifth annual ACM-SIAM symposium on Discrete algorithms*, SODA '94. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 365–371.

Bolker, B. & R Development Core Team (2011). *bbmle: Tools for general maximum likelihood estimation*. R package version 1.0.3.

Bollobás, B. (1998). *Modern Graph Theory*. Springer Science+Business Media, Inc.

Burnham, K. P. & Anderson, D. R. (1998). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer.

Claiborne, A. M., Fisher, J. P., Hayes, S. A. & Emmett, R. L. (2011). Size at release, size-selective mortality, and age of maturity of Willamette River hatchery yearling Chinook salmon. *Transactions of the American Fisheries Society* 140 1135–1144.

Conrad, J. M. & Smith, M. D. (2012). Nonspatial and spatial models in bioeconomics. *Natural Resource Modeling* 25 52–92. URL `http://dx.doi.org/10.1111/j.1939-7445.2011.00102.x`.

Constantino, M., Martins, I. & Borges, J. G. (2008). A new mixed-integer programming model for harvest scheduling subject to maximum area restrictions. *Operations Research* 56 542–551.

Cowan, J. H., Jr., Rose, K. A. & DeVries, D. R. (2000). Is density-dependent growth in young-of-the-year fishes a question of critical weight? *Reviews in Fish Biology and Fisheries* 10 61–89.

Craig, J. K., Burke, B. J., Crowder, L. B. & Rice, J. A. (2006). Prey growth and size-dependent predation in juvenile estuarine fishes: experimental and model analyses. *Ecology* 87 2366–2377.

Cross, A. D., Beauchamp, D. A., Moss, J. H. & Myers, K. W. (2009). Interannual variability in early marine growth, size-selective mortality, and marine survival for Prince William Sound pink salmon. *Marin and Coastal Fisheries: Dynamics, Management, and Ecosystem Science* 1 57–70.

Cross, A. D., Beauchamp, D. A., Myers, K. W. & Moss, J. H. (2008). Early marine

growth of pink salmon in Prince William Sound and the Coastal Gulf of Alaska during years of low and high survival. *Journal of the American Fisheries Society* 137 927–939.

Du, Q., Faber, V. & Gunzburger, M. (1999). Centroidal Voronoi tesselations: Applications and algorithms. *SIAM Review* 41 637–676.

Duffy, E. J. & Beauchamp, D. A. (2008). Seasonal patterns of predation on juvenile Pacific salmon by anadromous cutthroat trout in Puget Sound. *Transactions of the American Fisheries Society* 137 165–181.

Erdős, P. & Rényi, A. (1960). On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*. 17–61.

Ewing, R. & Ewing, G. (2002). Bimodal length distributions of cultured Chinook salmon and the relationship of length modes to adult survival. *Aquaculture* 209 139 – 155.

Farley, E. V., Moss, J. H. & Beamish, R. J. (2007). A review of the critical size, critical period hypothesis for juvenile Pacific salmon. *North Pacific Anadromous Fish Commission* 4 311–317.

Fresh, K. L. & Schroder, S. L. (2003). Predation by northern pikeminnow on hatchery and wild coho salmon smolts in the Chehalis River, Washington. *North American Journal of Fisheries Management* 23 1257–1264.

Fritts, A. L. & Pearsons, T. N. (2004). Smallmouth bass predation on hatchery and wild salmonids in the Yakima River, Washington. *Transactions of the American Fisheries Society* 133 880–895.

Fritts, A. L. & Pearsons, T. N. (2006). Effects of predation by nonnative smallmouth bass on native salmonid prey: the role of predator and prey size. *Transactions of the American Fisheries Society* 135 853–860.

Good, S. P., Dodson, J. J., Meekan, M. G. & Ryan, D. A. J. (2001). Annual variation in size-selective mortality of Atlantic salmon (*Salmo salar*) fry. *Can. J. Fish. Aquat. Sci.* 58 1187–1195.

Goycoolea, M., Murray, A. T., Barahona, F., Epstein, R. & Weintraub, A. (2005). Harvest scheduling subject to maximum area restrictions: Exploring exact approaches. *Operations Research* 53 490–500.

Gray, R. H. & Dauble, D. D. (2001). Some life history characteristics of cyprinids in the Hanford Reach, mid-Columbia River. *Northwest Science* 75 122–136.

Haeseker, S. L., McCann, J. A., Tuomikoski, J. & Chockley, B. (2012). Assessing freshwater and marine environmental influences on life-stage-specific survival rates of Snake River spring–summer Chinook salmon and steelhead. *Transactions of the American Fisheries Society* 141 121–138. URL `http://dx.doi.org/10.1080/00028487.2011.652009`.

Hastings, A., Arzberger, P., Bolker, B., Collins, S., Ives, A. R., Johnson, N. A. & Palmer, M. A. (2005). Quantitative bioscience for the 21st century. *BioScience* 55 511–517.

Henderson, M. A. & Cass, A. J. (1991). Effect of smolt size on smolt-to-adult survival for Chilko Lake sockeye salmon (*Oncorhynchus nerka*). *Can. J. Fish. Aquat. Sci.* 48 988–994.

Hinrichsen, R. A. (1994). *Optimization Models for Understanding Migration Patterns of Juvenile Chinook Salmon*. Ph.D. thesis, University of Washington.

Holtby, L. B., Andersen, B. C. & Kadowaki, R. K. (1990). Importance of smolt size and early ocean growth to interannual variability in marine survival of coho salmon (*Oncorhynchus kisutch*). *Can. J. Fish. Aquat. Sci.* 47 2181–2194.

Hunsicker, M. E., Ciannelli, L., Bailey, K. M., Buckel, J. A., White, J. W., Link, J. S., Essington, T. E., Gaichas, S., Anderson, T. W., Brodeur, R. D., Chan, K.-S., Chen, K., Englund, G., Frank, K. T., Freitas, V., Hixon, M. A., Hurst, T., Johnson, D. W., Kitchell, J. F., Reese, D., Rose, G. A., Sjodin, H., Sydeman, W. J., van der Veer, H. W., Vollset, K. & Zador, S. (2011). Functional responses and scaling in predator-prey interactions of marine fishes: contemporary issues and emerging concepts. *Ecology Letters* 14 1288–1299.

Ibbotson, A. T., Beaumont, W. R. C. & Pinder, A. C. (2011). A size-dependent migration strategy in Atlantic salmon smolts: Small smolts favour nocturnal migration. *Environ Biol Fish* 92 151–157.

Johnson, D. W. (2009). *Selection on Larval and Adult Body Size in a Marine Fish: Potential Evolutionary Responses and Effects on Population Dynamics*. Ph.D. thesis, Oregon State University.

Jokikokko, E., Kallio-Nyberg, I., Saloniemi, I. & Jutila, E. (2006). The survival of semi-wild, wild and hatchery-reared Atlantic salmon smolts of the Simojoki River in the Baltic Sea. *Journal of Fish Biology* 68 430–442.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *The Journal of the American Statistical Association* 90 773–795.

Koenings, J. P., Geiger, H. J. & Hasbrouck, J. J. (1993). Smolt-to-adult survival patterns of sockeye salmon (*Oncorhynchus nerka*): Effects of smolt length and geographic latitude when entering the sea. *Can. J. Fish. Aquat. Sci.* 50 600–611.

Ledgerwood, R. D., Naman, S. W., Ryan, B. A., Sandford, B. P., Banks, C. Z. & Ferguson, J. W. (2004). Detection of PIT-tagged juvenile salmonids in the Columbia River Estuary using a pair-trawl, 2000 and 2001. Tech. rep., Fish Ecology Division, Northwest Fisheries Science Center, NOAA.

Ledoux, H. & Gold, C. M. (2008). Modelling three-dimensional geoscientific fields with the Voronoi diagram and its dal. *International Journal of Geographical Information Science* 22 547–574.

Li, R., Bettinger, P. & Weiskittel, A. (2010). Comparisons of three different methods used to generate forest landscapes for spatial harvest scheduling problems with adjacency restrictions. *Mathematical and Computational Forestry & Natural-Resource Sciences* 2 53–60.

Martel, G. (1996). Growth rate and influence of predation risk on territoriality in juvenile coho salmon (*Oncorhynchus kisutch*). *Can. J. Fish. Aquat. Sci.* 53 660–669.

Mathews, S. B. & Buckley, R. (1976). Marine mortality of Puget Sound coho salmon (*Oncorhynchus kisutch*). *J. Fish. Res. Bd. Can.* 33 1677–1684.

Mathews, S. B. & Ishida, Y. (1989). Survival, ocean growth, and ocean distribution of differentially timed releases of hatchery coho salmon (*Oncorhynchus kisutch*). *Can. J. Fish. Aquat. Sci.* 46 1216–1226.

McDill, M. E. & Braze, J. (2000). Comparing adjacency constraint formulations for randomly generated forest planning problems with four age-class distributions. *Forestry Science* 46 423–436.

McDill, M. E., Rebain, S. A. & Braze, J. (2002). Harvest scheduling with area-based adjacency constraints. *Forest Science* 48 631–642.

McNaughton, A. J. & Ryan, D. (2008). Adjacency branches used to optimize forest harvesting subject to area restrictions on clearfell. *Forest Science* 54 442–454.

Moss, J. H., Beachamp, D. A., Cross, A. D., Myers, K. W., Jr., E. V. F., Murphy, J. M. & Helle, J. H. (2005). Evidence for size-selective mortality after the first summer of ocean growth by pink salmon. *Transactions of the American Fisheries Society* 134 1313–1322.

Mu, L. & Radke, J. (2009). A weighted difference barrier method in landscape genetics. *Journal of Geographical Systems* 11 141–154.

Muir, W. D., Marsh, D. M., Sandford, B. P., Smith, S. G. & Williams, J. G. (2006). Post-hydropower system delayed mortality of transported Snake River stream-type Chinook salmon: unraveling the mystery. *Transactions of the American Fisheries Society* 135 1523–1534.

Murray, A. T. (1999). Spatial restrictions in harvest scheduling. *Forest Science* 45 45–52.

Nelson, T., Boots, B., Wulder, M. & Feick, R. (2004). Predicting forest age classes from high spatial resolution remotely sensed imagery using Voronoi polygon aggregation. *GeoInformatica* 8 143–155.

Nemhauser, G. & Sigismondi, G. (1992). A strong cutting plane/branch-and-bound algorithm for node packing. *Journal of the Operational Research Society* 43 443–457.

Ohata, R., Masuda, R. & Yamashita, Y. (2011). Ontogeny of antipredator performance in hatchery-reared japanese anchovy *Engraulis japonicus* larvae exposed to visual or tactile predators in relation to turbidity. *Journal of Fish Biology* 79 2007–2018.

Okabe, A., Satoh, T., Furata, T., Suzuki, A. & Okano, K. (2008). Generalized network Voronoi diagrams: Concepts, computational methods, and applications. *International Journal of Geographical Information Science* 22 965–994.

Önal, H. & Briers, R. A. (2006). Optimal selection of a connected reserve network. *Operations Research* 54 379–388. `http://or.journal.informs.org/content/54/2/379.full.pdf+html`.

Parker, R. M., Zimmerman, M. P. & Ward, D. L. (1995). Variability in biological characteristics of northern squawfish in the lower Columbia and Snake rivers. *Transactions of the American Fisheries Society* 124 335–346.

Parker, R. R. (1971). Size selective predation among juvenile salmonid fishes in a British Columbia inlet. *Journal of the Fisheries Research Board of Canada* 28 1503–1510.

Peacor, S. D., Bence, J. R. & Pfister, C. A. (2007). The effect of size-dependent growth and environmental factors on animal size variability. *Theoretical Population Biology* 71 80–94.

Perkins, T. A. & Jager, H. I. (2011). Falling behind: Delayed growth explains life-history variation in Snake River fall Chinook salmon. *Transactions of the American Fisheries Society* 140 959–972.

Petersen, J. H. & Ward, D. L. (1999). Development and corroboration of a bioenergetics model for northern pikeminnow feeding on juvenile salmonids in the Columbia River. *Transactions of the American Fisheries Society* 128 784–801.

Petroski, J. R. (2006). *Treatment Unit Delineations for Spatial Planning on Pennsylvania State Forest Land*. Master of science, Pennsylvania State University.

Poe, T. P., Hansel, H. C., Vigg, S., Palmer, D. E. & Prendergast, L. A. (1991). Feeding of predaceous fishes on out-migrating juvenile salmonids in John Day Reservoir, Columbia River. *Transactions of the American Fisheries Society* 120 405–420.

Quinn, T. P. (2005). *The Behavior and Ecology of Pacific Salmon and Trout*. Bethesda, MD: American Fisheries Society.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rebain, S. A. & McDill, M. E. (2003). A mixed-integer formulation of the minimum patch size problem. *Forest Science* 49 608–618.

Rechisky, E. L. (2010). *Migrationandsurvivalofjuvenilespring Chinookand sockeyesalmondeterminedbyalargescaletelemetryarray*. Ph.D. thesis, University of British Columbia (Vancouver).

Reed, T. E., Martinek, G. & Quinn, T. P. (2010). Lake-specific variation in growth, migration timing and survival of juvenile sockeye salmon *Oncorhynchus nerka*: seperating environmental from genetic influences. *Journal of Fish Biology* 77 692–705.

Reisenbichler, R., McIntyre, J. & Hallock, R. (1982). Relation between size of Chinook salmon, *Onchorhynchus tshawytscha*, released at hatcheries and returns to hatcheries and ocean fisheries. *California Fish and Game* 68 57–59.

Saloniemi, I., Jokikokko, E., Kallio-Nyberg, I., Jutila, E. & Pasanen, P. (2004). Survival of reared and wild Atlantic salmon smolts: size matters moire in bad years. *ICES Journal of Marine Science* 61 782–787.

Satterthwaite, W. H., Beakes, M. P., Collins, E. M., Swank, D. R., Merz, J. E., Titus, R. G., Sogard, S. M. & Mangel, M. (2009). Steelhead life history on

California's central coast: Insights from a state-dependent model. *Transactions of the American Fisheries Society* 138 532–548.

Shively, R. S., Poe, T. P. & Sauter, S. T. (1996a). Feeding response by northern squawfish to a hatchery release of juvenile salmonids in the Clearwater River, Idaho. *Transactions of the American Fisheries Society* 125 230–236.

Shively, R. S., Poe, T. P., Sheer, M. B. & Peters, R. (1996b). Criteria for reducing predation by northern squawfish near juvenile salmonid bypass outfalls at Columbia River dams. *Regulated Rivers-research & Management* 12 493–500.

Smith, S. G., Muir, W. D. & Willams, J. G. (2002). Factors associated with travel time and survival of migrant yearling Chinook salmon and steelhead in the lower Snake River. *North American Journal of Fisheries Management* 22 385–405.

Sogard, S. (1997). Size-selective mortality in the juvenile stage of teleost fishes: A review. *Bulletin of Marine Science* 60 1129–1157.

Stamps, J. A. (2007). Growth-mortality tradeoffs and 'personality traits' in animals. *Ecology Letters* 10 355–363.

Sutton, T. M. & Ney, J. J. (2001). Size-dependent mechanisms influencing first-year growth and winter survival of stocked striped bass in a Virginia mainstream reservoir. *Transactions of the American Fisheries Society* 130 1–17.

Tabor, R. A., Shively, R. S. & Poe, T. P. (1993). Predation on juvenile salmonids by smallmouth bass and northern squawfish in the Columbia River near Richland, Washington. *North American Journal of Fisheries Management* 13 831–838.

Tóth, S. F. & McDill, M. E. (2008). Promoting large, compact mature forest patches in harvest scheduling models. *Environmental Modeling and Assessment* 13 1–15. URL `http://dx.doi.org/10.1007/s10666-006-9080-4`.

Tóth, S. F. & McDill, M. E. (2009). Finding efficient harvest schedules under three conflicting objectives. *Forest Science* 55 117–131.

Tóth, S. F., McDill, M. E., Könnyű, N. & George, S. (2012). A strengthening procedure for the path formulation of the area-based adjacency problem in harvest scheduling models. *Mathematical and Computational Forestry & Natural-Resource Sciences* 4 1–23.

Tóth, S. F., McDill, M. E. & Rebain, S. A. (2006). Finding the efficient frontier of a bi-criteria, spatially explicit, harvest scheduling problem. *Forest Science* 52 93–107.

Vigg, S., P, P. T., Prendergast, L. A. & Hansel, H. C. (1991). Rates of consumption of juvenile salmonids and alternative prey fish by northern squawfish walleyes smallmouth bass and channel catfish in John Day Reservoir, Columbia River. *Transactions of the American Fisheries Society* 120 421–438.

Ward, D. L., Petersen, J. H. & Loch, J. J. (1995). Index of predation on juvenile salmonids by northern squawfish in the lower and middle Columbia River and in the lower Snake River. *Transactions of the American Fisheries Society* 124 321–334.

West, C. J. & Larkin, P. A. (1987). Evidence for size-selective mortality of juvenile sockeye salmon (*Oncorhynchus nerka*) in Babine Lake, British Columbia. *Can. J. Fish. Aquat. Sci.* 44 712–721.

Wyszomirski, T. & Weiner, J. (2009). Variation in local density results in a positive correlation between plant neighbor sizes. *The American Naturalist* 173 705–708.

Zabel, R. W. & Achord, S. (2004). Relating size of juveniles to survival within and among populations of Chinook salmon. *Ecology* 85 795–806.

Zabel, R. W., Wagner, T., Congleton, J. L., Smith, S. G. & Williams, J. G. (2005). Survival and selection of migrating salmon from capture-recapture models with individual traits. *Ecological Applications* 15 1427–1439.

Zabel, R. W. & Williams, J. G. (2002). Selective mortality in Chinook salmon: what is the role of human disturbance? *Ecological Applications* 12 173–183.

Zens, M. S. & Peart, D. R. (2003). Dealing with death data: individual hazards, mortality and bias. *Trends in Ecology and Evolution* 18 366–373.

Zimmerman, M. P. (1999). Food habits of smallmouth bass, walleyes, and northern pikeminnow in the lower Columbia River basin during outmigration of juvenile anadromous salmonids. *Transactions of the American Fisheries Society* 128 1036–1054.

Appendix A

SUPPLEMENTARY MATERIAL FOR CHAPTER 1

This appendix presents the actual parameter estimates with confidence intervals for the predator susceptibility model fits, as well as other tables that supplement Chapter 1. Values for the asymptote-controlling parameters k_0 and k_1 are given in units of percent, l_c and σ_c are in units of mm. When a variable is a fixed effect corresponding to a certain year the last two digits of the year are notated as a left-hand-side superscript, e.g. ${}^{99}k_0$ is the value for k_0 in 1999.

A.1 Hatchery Survival to Lower Granite Dam

Sawtooth early years			
Parameter	Fitted value	Low (95%)	High (95%)
l_c	127.2	119.7	135.2
σ	18.7	11.6	30.4
${}^{89}k_0$	11.15	9.17	13.56
${}^{91}k_0$	5.08	4.27	6.05
${}^{89}k_1$	14.31	10.19	20.09
${}^{91}k_1$	6.02	3.58	10.11

Table A.1: Parameter values for the Predator Susceptibility Model fit with fixed effects to the early Sawtooth data and 95% confidence intervals. Preceding superscripts indicate the applicable year for fixed effects terms.

A.2 Adult Returns: Grande Ronde River

	Wild			Hatchery		
	Estimate	Low	High	Estimate	Low	High
l_c	96.96	92.78	101.32	155.09	106.26	226.37
σ_c	15.05	9.91	22.84	39.48	18.14	85.93
k_0	0.00	0.00	NA	0.00	0.00	NA
k_1	1.46	1.26	1.70	2.54	0.80	8.05

Table A.2: Parameter estimates with upper and lower 95% confidence bounds for wild and hatchery fish PIT-tagged at the mouth of the Grande Ronde River. Confidence intervals for k_0 and k_1 invalid (marked NA) if the parameter estimate is not significantly different from 0. The k_* parameters are given as percents.

A.3 Adult Returns: Lower Granite Dam

AIC	Wild		Hatchery	
Model Comparison	Barge	River	Barge	River
PSM	0	0	0	15
GLM(1)	30	7	37	0
GLM(2)	47	5	94	87
Omitting 1998 and 1999				
PSM	0	0	0	33
GLM(1)	28	5	26	5
GLM(2)	47	7	24	0

Table A.3: Relative AIC scores of the PSM with fixed effects asymptotes to competing GLM fits. GLM(1) has year effects for both intercept and slope, while GLM(2) has year effects for the intercept only. As 1998 and 1999 exhibit exceptionally small length effects, especially for hatchery fish, the models were fit with and without those years, though it did not qualitatively change the results regarding the PSM. Lower AIC scores are better, with a difference of 2 indicating significant improvement and a difference of 10 indicating highly significant improvement. The PSM improves on the GLM approach for all but the run-of-river hatchery fish, though the gains are most significant for barged fish.

Lower Granite, Wild, Barged			
Parameter	Estimate	Low	High
l_c	107.64	106.33	108.96
σ_c	5.62	4.17	7.58
${}^{98}k_0$	0.28	0.13	0.58
${}^{99}k_0$	1.45	1.17	1.79
${}^{01}k_0$	0.38	0.23	0.62
${}^{02}k_0$	0.53	0.32	0.87
${}^{03}k_0$	0.18	0.11	0.30
${}^{04}k_0$	0.30	0.21	0.43
${}^{05}k_0$	0.01	0.00	0.07
${}^{06}k_0$	0.62	0.50	0.77
${}^{07}k_0$	0.34	0.21	0.55
${}^{08}k_0$	2.24	1.95	2.58
${}^{09}k_0$	0.47	0.35	0.63
${}^{98}k_1$	0.47	0.25	0.89
${}^{99}k_1$	1.54	1.12	2.11
${}^{01}k_1$	0.89	0.66	1.21
${}^{02}k_1$	1.79	1.34	2.41
${}^{03}k_1$	0.58	0.39	0.88
${}^{04}k_1$	0.54	0.35	0.84
${}^{05}k_1$	0.55	0.45	0.67
${}^{06}k_1$	0.47	0.30	0.72
${}^{07}k_1$	1.14	0.91	1.42
${}^{08}k_1$	1.42	1.06	1.91
${}^{09}k_1$	1.01	0.77	1.31

Table A.4: Parameter estimates with upper and lower 95% confidence bounds for wild barged fish PIT-tagged at Lower Granite Dam. The k_* parameters are given as percents.

Lower Granite, Wild, Run-of-river			
Parameter	Estimate	Low	High
l_c	103.11	101.04	105.22
σ_c	8.22	5.88	11.47
${}^{98}k_0$	0.01	0.00	NA
${}^{99}k_0$	0.83	0.57	1.22
${}^{00}k_0$	0.87	0.68	1.12
${}^{02}k_0$	0.62	0.49	0.79
${}^{03}k_0$	0.02	0.00	0.14
${}^{06}k_0$	0.18	0.06	0.50
${}^{09}k_0$	0.08	0.02	0.33
${}^{98}k_1$	0.72	0.65	0.80
${}^{99}k_1$	0.78	0.46	1.35
${}^{00}k_1$	1.21	0.95	1.53
${}^{02}k_1$	0.65	0.46	0.92
${}^{03}k_1$	0.22	0.16	0.30
${}^{06}k_1$	0.86	0.62	1.19
${}^{07}k_1$	0.56	0.33	0.96
${}^{08}k_1$	1.00	0.53	1.91
${}^{09}k_1$	0.51	0.36	0.73

Table A.5: Parameter estimates with upper and lower 95% confidence bounds for wild run-of-river fish PIT-tagged at Lower Granite Dam. Confidence intervals for k_0 and k_1 invalid (marked NA) if the parameter estimate is not significantly different from 0. The k_* parameters are given as percents.

Lower Granite, Hatchery, Barged			
Parameter	Estimate	Low	High
l_c	145.52	141.03	150.15
σ_c	15.73	12.64	19.57
${}^{98}k_0$	0.65	0.62	0.69
${}^{99}k_0$	2.03	1.89	2.17
${}^{06}k_0$	0.22	0.14	0.33
${}^{07}k_0$	0.46	0.36	0.60
${}^{08}k_0$	1.20	1.05	1.36
${}^{98}k_1$	0.01	0.00	NA
${}^{99}k_1$	0.37	0.15	0.89
${}^{06}k_1$	1.87	1.47	2.38
${}^{07}k_1$	1.95	1.46	2.61
${}^{08}k_1$	3.37	2.66	4.25

Table A.6: Parameter estimates with upper and lower 95% confidence bounds for hatchery barged fish PIT-tagged at Lower Granite Dam. Confidence intervals for k_0 and k_1 invalid (marked NA) if the parameter estimate is not significantly different from 0. The k_* parameters are given as percents.

Lower Granite, Hatchery, Run-of-river			
Parameter	Estimate	Low	High
l_c	141.60	137.96	145.34
σ_c	21.35	18.28	24.93
${}^{98}k_0$	0.37	0.35	0.40
${}^{99}k_0$	1.27	0.00	NA
${}^{05}k_0$	0.07	0.06	0.08
${}^{06}k_0$	0.07	0.03	0.19
${}^{07}k_0$	0.22	0.21	0.23
${}^{08}k_0$	0.36	0.28	0.46
${}^{09}k_0$	0.01	0.00	NA
${}^{98}k_1$	0.01	0.00	0.29
${}^{99}k_1$	0.47	0.00	NA
${}^{05}k_1$	0.02	0.00	NA
${}^{06}k_1$	1.24	1.02	1.51
${}^{07}k_1$	0.05	0.00	NA
${}^{08}k_1$	1.95	1.65	2.31
${}^{09}k_1$	0.96	0.84	1.10

Table A.7: Parameter estimates with upper and lower 95% confidence bounds for hatchery run-of-river fish PIT-tagged at Lower Granite Dam. Confidence intervals for k_0 and k_1 invalid (marked NA) if the parameter estimate is not significantly different from 0. The k_* parameters are given as percents.

Year	Barged	Run-of-River	Difference
1998	0.052	0.004	0.048
1999	0.159	0.129	0.030
2002	0.371	0.315	0.056
2003	0.042	0.160	-0.118
2006	-0.115	-0.038	-0.077
2007	0.131	0.021	0.110
2008	-0.173	-0.259	0.087
2009	0.050	0.055	-0.005

Table A.8: Pearson correlation coefficients between length and day-of-year when passing Lower Granite (coincident with tagging and barging, when applicable) for wild spring Chinook salmon. The final column has the difference in correlations between the barged and run-of-river treatment groups. The correlation of the correlations is 0.888.

Year	Barged	Run-of-River	Difference
1998	0.265	0.227	0.038
1999	0.048	0.044	0.004
2006	-0.167	-0.035	-0.133
2007	0.000	-0.368	0.368
2008	0.393	0.256	0.137

Table A.9: Pearson correlation coefficients between length and day-of-year passing Lower Granite (coincident with tagging and barging, when applicable) for hatchery spring Chinook salmon. The final column has the difference in correlations between the barged and run-of-river treatment groups. The correlation of the correlations is 0.697.

Appendix B

RLANDSCAPE USER'S GUIDE AND MODEL FITS

B.1 User's Guide

Rlandscape was built using R version 2.15.0, and an up-to-date R installation is recommended for running it. R is an open-source and free statistical language, available for UNIX, Windows, and MacOS. The most recent R version can be found at www.r-project.org. It will also be necessary to install the Spatstat package (for example, by entering `install.packages("spatstat")` in the R console). To use the Rlandscape functions, they must first be loaded by using the source command. It is a good practice to change the working directory to the directory where they Rlandscape files are located, either through `File\Change dir` menu or by using the `setwd()` command, e.g. `setwd("c:/directory")`. Once R knows where to look for the files, entering `source("rlandscape.R")` will load all the appropriate functions. The working directory is also the location where output files will be saved by default, but this can be modified to subdirectories of the working directory or any other folder by giving a full file path as an output argument.

B.1.1 Rland

The primary function users will interact with is `rland`. Its arguments are detailed below, but default values are set so that simply entering `rland()` will produce 10 landscapes. They will be saved in the working directory as `landscape_%id_%datatype.extension`, where `%id` is a unique number for each landscape, `%datatype` is either `adj` or `areas` for the adjacency or area data. A summary file called `landscape_summary.csv` will be saved that includes the statistics for each of the generated landscapes.

Common usage requires R's native concatenation command, `c()`, to create 2-element vectors specifying minimum and maximum values. These can be passed to arguments `n`, `degMean`, `degSD`, `areaCV`, and `hAsp` to set bounds on the number of tiles, degree mean,

degree standard deviation, area coefficient of variance, and horizontal:vertical aspect ratio, respectively. Arguments that are left out default to unrestrictive values, thus the call `rland(n = c(240, 260), degMean = c(4.5, 5), degSD = c(1, 2))` will produce landscapes with between 240 and 260 tiles that have mean degree between 4.5 and 5 and degree standard deviation between 1 and 2. The absolute bounds of these statistics is mentioned above. Practically, degree means between 3 and 5.5 are easily attainable, especially if the user can allow a tolerance range of about 0.5. Degree standard deviation has a narrow-seeming range, tending to be from about 1.5 to 2.1. The recommended range for area coefficient of variation is from about 50 to 200. Both σ and CVa have no real upper bounds, but producing extremely high values is accomplished by having a very high `pMerge` which usually results in one or two tiles dominating the landscape, or extremely tight clusters of tiles that are minuscule when compared to the average. We encourage users to calculate these statistics for real data sets to get a sense of how they vary in actual applications. The other arguments taken by `rland` are listed below.

`reps`, a positive integer (default 10), which sets how many landscapes are to be produced.

`varied`, a boolean (default true), indicating whether the targets should be spread across the allowable ranges (the default) or set consistently in the middles.

`effort`, a numeric between 0 and 1 (default 0.95), which is a maximum acceptable failure rate evaluated every 500 iterations.

`filename`, a character string (defaults to `"landscape"`) giving the initial part of the file name under which to save results, which can optionally include a file path.

`savePlot`, a boolean (default false) indicating whether or not a plot of each landscape should be saved as a `.png` file.

`saveLand`, a boolean (default false) indicating whether or not to save each landscape object as a `.Rdata` file.

`saveFails`, a boolean (default false) indicating whether or not to save summaries about the landscapes that are unsuccessful, probably useful for development purposes only. Additional parameters will be passed to the plotting function if `savePlot` is set to true.

B.1.2 Rlandscape

The function that generates an individual landscape is `rlandscape`. It only generates one landscape at a time (which it returns), so any batch projects will need to call it either from `rland` or in a loop. The arguments it takes are

`n`, a vector of up to 4 non-negative integers (default `c(100, 0, 0, 0)`) of how many tiles in the landscape should be based on points from the uniform, lattice, cluster, and inhibition methods, respectively. The length of the vector can be shorter than 4, in which case it will be filled out with zeroes.

`hAspect`, a numeric (default 1) which gives the horizontal to vertical aspect ratio.

`pMerge`, a numeric in $[0, 1)$ (default 0.1) giving the proportion of edges in the landscape that are deleted.

`pHole`, a numeric in $[0, 1)$ (default 0.1) giving the proportion of tiles in the landscape that are holes.

`plot`, a boolean (default FALSE) indicating whether or not a plot of the resulting landscape is displayed immediately. (It will not be saved automatically in either case.)

`control`, a list of control parameters passed to the point processes. The list can contain up to three numeric vectors named "lattice", "cluster", and "ssi". The lattice element should be a vector of two integers, giving the number or horizontal and vertical segments in the lattice grid. The cluster element should be a vector of an integer and a numeric, giving the number of points per cluster and the spread around each

cluster. The SSI element takes only on numeric: the inhibition distance. An example input would be `list(lattice = c(10, 12), cluster = c(15, 0.001), ssi = .001)`. Any arguments that are left out will be defined by the defaults of the point processes.

`savePath`, a character vector (defaults to `NULL`) which, if defined, will save the landscape object, its adjacency table, its area table, and its summary statistics similar to the saving in `rland`.

Additional arguments will be passed to the plotting routine if `plot` is set to `true`. As mentioned earlier, both `pHole` and `pMerge` are accounted for when the actual points are selected so that the output number of tiles matches the input `n`. The `rlandscape` function returns a landscape object which may be assigned to a variable, e.g. by entering `myLand <- rlandscape(n = c(40, 60), pHole = 0.3))` at the console. Its attributes may then be accessed within R by using the the `$` operator, for example `myLand$stats` will print the statistics of `myLand` and `myLand$adjacencies` will print the adjacency list. The complete list of attributes of an landscape object can be viewed by the command `attributes(myLand)`. A landscape object can also be drawn with the plot command `plot(myLand)`.

Plotting landscapes

The plotting routine defined for landscape objects takes *many* parameters for flexibility. The defaults should produce nice plots, but the following arguments can be given to `plot` either directly or via a call to `rland` or `rlandscape`:

`number`, a boolean (default false) that indicates whether the unique ID number of a tile should be plotted on it.

`points`, a boolean (default is the opposite of `number`) indicating whether or not the point corresponding to each tile should be drawn.

`axes`, a boolean (default false) indicating whether axes should be drawn.

`lineCol`, a color (default is `"black"`) with which to draw the borders between tiles.

`holeCol`, a color (default is `"gray20"`) with which to fill in the deleted tiles.

There are 13 other graphical parameters giving control over colors, margins, and text sizes used. Their details will be omitted, but they are well documented in the code (the code can be viewed by entering `plot.landscape()`).

Output

Any output specified will be appropriately named and saved as described in Section B.1.1. The adjacency files are saved in a tab-delimited format, where each line consists of 2 tile indices that are adjacent to each other. They are sorted in order of index number, with the first column dominating. Pairs are included so that if tile 1 is adjacent to tile 5, both 1 5 and 5 1 will be included as rows. The areas are stored in a comma-separated values (`.csv`) format with columns indicating the index number, the total area of the tile, and the x and y position of the original point that created the tile. Index numbers will not be strictly sequential. The index values are set when the initial random points are placed, and gaps are created when tiles are deleted or merged. If tile 17 is a hole, or was merged with tile 12, then the index 17 will not occur in the area or adjacency tables. The summary file is also in a CSV format, with columns recording the landscape name (with ID number), the total number of tiles, degree mean, degree standard deviation, area coefficient of variation, the number of points assigned by each point process, the merge and hole proportions, the horizontal aspect ratio, and the control parameters passed to `rlandscape`. If the landscape object is saved, it is written as a `.Rdata` file using the function `dput`; it can be loaded back into R using `dget`. If plots are saved in `rland`, they will be saved as `.png` files with appropriate aspect ratios.

B.2 Model Fits

Box-Cox transformations were applied when appropriate. This is indicated in the text preceding the table, and given as an exponent to which the response was raised. Predictors with a b exponent indicate that a previously defined Box-Cox exponent is applied. hAsp is

the horizontal to vertical aspect ratio, s is the spread of points around a cluster center in the cluster method, and all other parameters are as defined in Chapter 2.

B.2.1 Low CV

Low CV, Low degre mean: $\mu_d < 4.5$

Parameter	Coefficient
Intercept	0.0018916802
CV	0.0242357612
μ_d	0.3448884426
σ_d	0.7975754620
hAsp	-0.1285168319
μ_dCV	-0.0122143735
CV σ_d	-0.0272013806
$\mu_d\sigma_d$	-0.2134670817
CV hAsp	0.0047607477
μ_d hAsp	0.0306755782
σ_d hAsp	0.0890567504
$\mu_d\sigma_d$ hAsp	0.0069329260
μ_d CV hAsp	-0.0009202953
σ_d CV hAsp	-0.0032959800
$\mu_d\sigma_d$ hAsp	-0.0192665653

Table B.1: Linear model for estimating p_{lat}. Box-Cox transformation: 2.6.

Parameter	Coefficient
Intercept	-0.111726129
CV	0.020055917
μ_d	0.102445411
σ_d	1.240468628
p_{lat}^b	0.971029828
μ_d CV	-0.005162486
CV σ_d	-0.014424910
$\mu_d\sigma_d$	-0.262783129
CV p_{lat}^b	-0.009592290
$\mu_d p_{lat}^b$	-0.376047080
$\sigma_d p_{lat}^b$	-0.522937725
CV $\mu_d\sigma_d$	0.003245974
CV $\mu_d p_{lat}^b$	0.004181883
CV $\sigma_d p_{lat}^b$	0.008188504
$\mu_d\sigma_d p_{lat}^b$	0.148101201
CV $\mu_d\sigma_d p_{lat}^b$	-0.002921096

Table B.2: Linear model for estimating p_H. Box-Cox transformation: 0.71.

Parameter	Coefficient
Intercept	-0.851223375
CV	0.030557526
μ_d	-0.154347713
σ_d	0.605047039
p^b_{lat}	-2.163712546
p^b_H	-0.323784893
μ_d CV	-0.001861974
CV σ_d	-0.009958089
$\mu_d\sigma_d$	0.073295652
CV p^b_{lat}	-0.026261358
$\mu_d p^b_{lat}$	0.882713600
$\sigma_d p^b_{lat}$	1.019102973
CV p^b_H	-0.006171021
$\mu_d p^b_H$	0.032948069
$\sigma_d p^b_H$	0.292099442
$p^b_{lat} p^b_H$	1.679892644
CV $\mu_d p^b_{lat}$	-0.003114309
CV $\sigma_d p^b_{lat}$	0.019323158
$\sigma_d\mu_d p^b_{lat}$	-0.378709781
$\sigma_d p^b_{lat} p^b_H$	-0.884864371

Table B.3: Linear model for estimating p_M. Box-Cox transformation: 0.75.

Low CV, high degree mean: $\mu_d \geq 4.5$

Parameter	Coefficient
Intercept	1.861864714
μ_d	-0.006812436
σ_d	-0.029767572
CV	-0.031320774

Table B.4: Linear model for estimating p_{SSI}. Box-Cox transformation: 2.5.

Parameter	Coefficient
Intercept	1.463375069
p^b_{SSI}	-0.098484600
mu_d	-0.150494846
σ_d	0.342471110
CV	-0.011659322
$p^b_{SSI}\mu_d$	-0.018858640
$p^b_{SSI}\sigma_d$	0.088031132
$\mu_d\sigma_d$	-0.109025814
p^b_{SSI} CV	-0.002353975
σ_d CV	0.004181020

Table B.5: Linear model for estimating p_H. Box-Cox transformation: 0.7.

Parameter	Coefficient
Intercept	-3.229551001
p^b_{SSI}	0.858816217
μ_d	0.401164457
σ_d	1.141137939
CV	0.048860426
p^b_H	0.926730792
$p^b_{SSI}\mu_d$	-0.117577910
$p^b_{SSI}\sigma_d$	-0.076030390
$\mu_d\sigma_d$	-0.074082431
μ_d CV	-0.007196297
σ_d CV	-0.005547122
$p^b_{SSI}p^b_H$	-0.298113908
CV p^b_H	-0.017367741

Table B.6: Linear model for estimating p_M. Box-Cox transformation: 0.75.

B.2.2 High CV

Parameter	Coefficient
Intercept	.45857
CV	.00055712

Table B.7: Linear model for estimating p_{clust}.

Parameter	Coefficient
Intercept	1.006e+00
CV	-1.161e-02
μ_d	-2.409e-02
σ_d	2.048e-02
n hAsp	1.148e-04
p_{clust}	-7.454e-01
CV n hAsp	-1.353e-06
CV p_{clust}	1.339e-02
$\mu_d n$ hAsp	5.324e-06
n hAsp p_{clust}	5.715e-05
CV σ_d	-1.054e-03
CV μ_d	2.870e-04
$\mu_d \sigma_d$	7.616e-03
σ_d n hAsp	-4.681e-06

Table B.8: Linear model for estimating s, the spread of the clusters. Box-Cox transformation: 0.46.

Parameter	Coefficient
Intercept	1.233e+00
CV	-1.282e-03
μ_d	-1.812e-01
σ_d	5.212e-01
p_{clust}	-1.869e-01
s	6.585e-01
$\mu_d \sigma_d$	-8.997e-02
$\mu_d s$	-1.367e-01
CV s	-3.949e-03
CV σ_d	-9.294e-04
$\sigma_d s$	-1.346e-01
CV p_{clust}	2.477e-03
$\sigma_d p_{clust}$	5.556e-02

Table B.9: Linear model for estimating p_H. Box-Cox transformation: 2/3.

Parameter	Coefficient
Intercept	-6.026e+00
CV	2.244e-03
μ_d	9.834e-01
σ_d	3.157e+00
n	-6.879e-04
s	-1.372e+00
p_{clust}	-4.003e-02
p_H	2.283e+00
$\mu_d\sigma_d$	-4.755e-01
CV s	2.694e-02
$\sigma_d p_H$	-9.757e-01
CV σ_d	-2.095e-03
$\mu_d p_H$	6.487e-02
$\sigma_d n$	3.241e-04

Table B.10: Linear model for estimating p_M. Box-Cox transformation: 0.8.