Jim Irvine

Assessing infilling methods for missing data in spawning salmon estimates

by

Ruth Joy

B. Sc. University of Victoria, 1996

A MASTER'S PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Statistics and Actuarial Science

SIMON FRASER UNIVERSITY
July 2002

APPROVAL

Name: Ruth Joy

Degree: Master of Science

Title of thesis: Assessing infilling methods for missing data in spawning salmon estimates

Examining Committee:

Dr. Richard Lockhart
Chair

____

Dr. Richard Routledge
Senior Supervisor

____

Dr. Carl Schwarz

____

Dr. James Richard Irvine
External Examiner
Research Scientist
Fisheries and Oceans Canada

Date Approved: July 5, 2002

Assessing infilling methods for missing data in spawning salmon estimates

Monitoring of populations is a key component of an effective conservation program. Trends in abundance must be monitored to ensure that timely action is taken before conservation risks become too severe. Unfortunately this monitoring is expensive and in many instances, only the abundance of a portion of widespread species can feasibly be estimated in a given year. For spawning estimates of British Columbia's Thompson River coho salmon, 41% of the data are missing between 1976 and 2001. Accurate abundance estimates for this aggregate are particularly important as the abundance of these salmon had declined so severely by the late 1990's that a major, continuing conservation effort was initiated.

This project presents an examination of seven imputation methods for infilling missing data in such records. We assessed the performance of these methods through a simulation study that modeled widely accepted features of the population dynamics of coho from the Thompson River watershed, specifically including the recent decline. The study also incorporated the historical record of missing estimates. Performance was measured through jackknifed sums-of-squares estimates to evaluate bias, chance error and total error of the infilled values. We found that the infilling methods that use a multiplicative analysis-of-variance-style model outperfomed the others, with the preferred version within this class of methods application-dependent.

We also investigated a sockeye salmon population where the missing data pattern was extreme (72% missing). In this extreme case, with little time overlap between

data records for different subsets of spawning areas, no method for imputing missing values worked well. For methods based on modified analyses of variance, this difficulty can be related to the concept of balance in an experimental design. We conclude the project with an exploration of the advantages of sampling schemes that promote this sort of balance. In particular, we demonstrate the drop in bias and chance error by incorporating balance into the sampling design.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A major component of a conservation program is the monitoring of population size. Trends in abundance need to be monitored if timely action is to be taken. Unfortunately, the monitoring is almost always expensive and complicated. This is particularly so for populations such as Pacific salmon (*Oncorhynchus* spp.) that occupy numerous, more or less discrete habitat units. Attempts to census entire species are doomed to failure. At best, accurate estimates can be obtained for relatively few local populations in any given year. In the case of British Columbia's coho salmon (*O. kisutch*), the existing record of spawner-abundance data is irregular.

As a result, it is typically difficult to obtain clear, unambiguous evidence of abundance trends from the irregular records. Here we examine two such datasets: Thompson River coho and North Coast sockeye (*O. nerka*). Both coho and sockeye salmon are protected in areas of Washington, Oregon and California by the US Endangered Species Act. In British Columbia, Thompson River coho salmon are recognised as both genetically unique and severely depressed by a decline in marine survival in the last decade (Bradford and Irvine 2000), such that in May 2002, this population was officially designated as endangered by COSEWIC (COSEWIC 2002).

Clear evidence of abundance trends is key to the implementation of management actions. Without such evidence, declines may go undetected before a crisis ensues.

This project addresses potential improvements to the estimation procedures for assessing trends in spawner abundance. The questions addressed specifically are:

1. How might overall abundance estimates be constructed from partial records of abundance with estimates for individual spawning populations for some creeks missing in some years? This question will be addressed by considering several methods for imputing the missing elements in the data record (Chapter 2).

2. Which of these imputation methods would make this task more reliable, and for what sorts of patterns of missing data can a reliable estimate not be constructed (Chapter 3)?

3. What sorts of sampling schemes for deliberately generating partial records would make this task easier (Chapter 4)?

Chapter 2

Missing Data Methods

In this chapter, seven methods for imputing values for missing data in spawning salmon records are presented. These kinds of data containing information about an entire river system, is typical for coastal North America. Numbers of spawning salmon are estimated in each of the creek tributaries for several years. Typically spawning salmon numbers are not recorded for every year, but instead creeks are irregularly sampled, especially for minor but potentially important subpopulations. Likewise, intensity of sampling varies between years, depending on fluctuating budgets and changing government priorities. The following imputation methods will be discussed with records for a single creek running across a row (rows $1, \ldots, c$), and records for a single year running down a column (columns $1, \ldots, t$), such that the data are in a $c \times t$ matrix.

2.1 Seven Imputation Methods

2.1.1 Zero-Infilled

The zero-infilled method replaces all missing values with zeros.

2.1.2 Nearest-Value

The nearest-value method imputes from the same creek from the nearest year for which there is an entry (i.e., the closest entry in the same row). In the case of a tie, the entry for the closest preceding year is chosen (i.e., the closest left-hand value).

2.1.3 Interpolation/Extrapolation

This method calculates the slope and intercept of a line between the closest left and closest right values for a creek and then infills based on this equation. If there are no left-hand values then the equation is extrapolated based on the closest two right-hand values. Likewise, if there are no right-hand values then the equation is based on the closest two left-hand values.

2.1.4 Averaging of scaled values

The averaging of scaled values is a method that has been used by the Department of Fisheries and Oceans to infill missing values. The method begins by scaling all the observed values by dividing each by the maximum observation for the same creek across all years. The missing values in this rescaled record are then filled in by the average of all the observed scaled values for that year. The missing values are then converted back to the original scale by multiplying by the creek maxima. For example: a single missing value in the j^{th} column (i^{th} row) is replaced with:

$$y_{ij} = y_{i\cdot max} \times \frac{1}{\tilde{c}} \sum_{i}^{\tilde{c}} \frac{y_{ij}}{y_{i\cdot max}}$$

where $\sum_i^{\tilde{c}}$ refers to the sum over all observed creeks for the j^{th} year, and $y_{i\cdot max} = \max_{j=1,\ldots,t}(n_{ij})$.

2.1.5 Transformed Linear Model

This method fits an analysis of variance model to the observed data to estimate creek and year effects. One assumption of this analysis of variance model is that the responses are normally distributed and have constant variance independent of the mean. However, these data are count data and the variance is likely to be a function of the mean. One approach to this problem is to perform a variance stabilizing transformation before fitting the model. The natural log transform (n_{ij} transformed to $\ln n_{ij} + 0.5$) suits this purpose, where the addition of 0.5 is to avoid problems with observed zero counts. The second assumption is that the systematic effects combine additively with no interactions. Only main effects (1, ..., c creeks and 1, ..., t years) are considered, and interactions between creek and year are assumed negligible as they are totally confounded with the error. The log transformation is often also a good choice as it can help to linearize the fit. Missing values are infilled by applying the fitted linear model and back transforming.
The linear model may be represented by:

$$\ln(Y_{ij} + .5) = \mu + c_i + t_j + \epsilon_{ij}, \tag{2.1}$$

where μ is the intercept or grand mean; c_i is the i^{th} creek effect: $i = 1, ..., c$ and $\sum_{i=1}^{c} c_i = 0$; t_j is the j^{th} year effect: $j = 1, ..., t$ and $\sum_{j=1}^{t} t_j = 0$; and ϵ_{ij} is the error term.

However, such transformation-based methods are now considered by most statisticians to have been superceded by generalized linear models. We therefore considered two such models, one based on the Poisson distribution, the other on the gamma distribution.

2.1.6 Model Based on the Poisson Distribution

This method assumes that the data are Poisson-distributed for which the variance is proportional to the mean. We used the following model:

$$Y_{ij} = c_i t_j (1 + \epsilon_{ij}), \tag{2.2}$$

where c_i is the i^{th} creek effect: $i = 1, ..., c$ and to ensure identifiability the product of the c_i's is restricted to 1; t_j is the j^{th} year effect: $j = 1, ..., t$; and ϵ_{ij} is the error term. This gives $E(y_{ij}) = c_i t_j$ and $Var(y_{ij}) = c_i t_j$, with $Y_{ij} \sim Poisson(c_i t_j)$ and hence

$$P(Y_{ij} = y_{ij}) = \frac{e^{-c_i t_j}(c_i t_j)^{y_{ij}}}{y_{ij}!}$$

and the constraint, $\prod_i^c c_i = 1$. Then up to a constant the log-likelihood function is:

$$l(\mathbf{c}, \mathbf{t}|y_{ij}) = \sum_i^{\tilde{c}} \sum_j^{\tilde{t}} \left[y_{ij} \ln(c_i t_j) - c_i t_j \right], \tag{2.3}$$

where $\tilde{c}$ and $\tilde{t}$ indicate that the sum was taken over all observed creeks and all observed years. This log-likelihood can be maximized to produce maximum likelihood estimate for the parameters. To maximize this function (2.3) under the constraint that $\prod_i^c c_i = 1$ or equivalently that $\sum_i^c \ln(c_i) = 0$, we can form the Lagrangian equation:

$$\text{Ł}(\mathbf{c}, \mathbf{t}, \lambda) = l(\mathbf{c}, \mathbf{t}|y_{ij}) + \lambda \sum_i^c \ln(c_i),$$

and set the partial derivatives with respect to c_i and t_j equal to zero. Thus

$$\frac{\partial \text{Ł}}{\partial c_i} = \sum_j^{\tilde{t}} \left(\frac{y_{ij}}{c_i} - t_j \right) + \frac{\lambda}{c_i} = 0$$

$$\text{and} \quad \frac{\partial \text{Ł}}{\partial t_j} = \sum_i^{\tilde{c}} \left(\frac{y_{ij}}{t_j} - c_i \right) = 0$$

Lagrange multipliers indicate the rate at which the maximum value increases as the constraint is relaxed. Here, because the maximum value is independent of the constraint, λ must be zero, and therefore the maximum likelihood estimates satisfy:

$$c_i = \frac{\sum_j^{\tilde{t}} y_{ij}}{\sum_j^{\tilde{t}} t_j}$$

$$t_j = \frac{\sum_i^{\tilde{c}} y_{ij}}{\sum_i^{\tilde{c}} c_i}$$

and can be solved iteratively.

The parameters can also be estimated by the Iterated Reweighted Least Squares (IRLS) procedure. The following procedure is iterated minimized through IRLS: With c_{i0} and t_{j0} fixed from the last iteration, find values of c_i and t_j that minimize:

$$\sum_i^{\tilde{c}} \sum_j^{\tilde{t}} \frac{(y_{ij} - c_i t_j)^2}{c_{i0} t_{j0}} , \qquad \text{subject to } \prod_i^c c_i = 1 \; .$$

The solution converges to the maximum likelihood estimates. We fit the model using the IRLS method from SPLUS. These estimates can also be justified when there is extra-Poisson variation as long as the variance remains proportional to the mean.

As in the previous method, only main effects are considered and interactions between creek and year are assumed to be negligible as they are inseparable from the error component. Missing values are infilled by applying the fitted Poisson model.

2.1.7 Model Using Gamma Distributions

The final imputation method explored here uses a family of gamma distributions in which the variance is proportional to the square of the mean; thus allowing a stronger dependence of the variance on the mean. We used the model as in (2.2) with $E(y_{ij}) = c_i t_j$ and $\mathrm{Var}(y_{ij}) \propto (c_i t_j)^2$. Thus $Y_{ij} \sim gamma(c_i t_j, \nu)$ with

$$P(Y_{ij} = y_{ij}) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu y_{ij}}{c_i t_j} \right)^{\nu} \frac{1}{y_{ij}} e^{-\nu \frac{y_{ij}}{c_i t_j}} ,$$

and the constraint, $\prod_i^c c_i = 1$. Interaction effects between creeks and years are not considered as they are confounded with the error.
The log-likelihood is therefore a fixed constant plus:

$$l(\mathbf{c}, \mathbf{t} | y_{ij}) = \sum_i^{\tilde{c}} \sum_j^{\tilde{t}} \left[-\ln\Gamma(\nu) + \nu \ln \left(\frac{\nu y_{ij}}{c_i t_j} \right) - \frac{\nu y_{ij}}{c_i t_j} \right] \; .$$

As in the Poisson model, to maximize the likelihood function under the constraint, $\prod_i^c c_i{=}1$, we form the Lagrangian equation:

$$Ł(\mathbf{c}, \mathbf{t}, \lambda) = l(\mathbf{c}, \mathbf{t}|y_{ij}) + \lambda \sum_i^c \ln(c_i).$$

By taking the first derivative of the Lagrangian equation with respect to c_i and t_j, and setting these derivatives equal to zero, one can find solutions for the maximum likelihood estimators.

$$\frac{\partial Ł}{\partial c_i} = \sum_j^{\tilde{t}} \left[-\frac{\nu}{c_i} + \frac{\nu y_{ij}}{t_j c_i^2} \right] + \frac{\lambda}{c_i} = 0.$$

As in the case of the Poisson model, the Lagrange multiplier equals zero, and the maximum likelihood equations simplify to:

$$c_i = \frac{1}{\tilde{t}} \sum_j^{\tilde{t}} \frac{y_{ij}}{t_j} \quad \text{for all } i.$$

Similarly

$$\frac{\partial Ł}{\partial t_j} = 0 \qquad \text{gives } t_j = \frac{1}{\tilde{c}} \sum_i^{\tilde{c}} \frac{y_{ij}}{c_i} \text{ for all } j\,.$$

These equations can be solved iteratively for c_i and t_j. In addition, the same estimates can be found through iteratively reweighted least squares. We chose to use the maximum likelihood equations to numerically find the solutions.

The maximum likelihood algorithm starts by making reasonable guesses for the starting values of $c_1, \ldots, c_c$. Each unknown year parameter (t_j) is considered separately and the estimation problem is reduced to an estimation of a mean:

$$t_j = \frac{1}{\tilde{c}} \sum_i^{\tilde{c}} \frac{y_{ij}}{c_i}$$

where $\tilde{c}$ is the number of observed creeks for the j^{th} year. To update the estimate of c_i, the t_j's from above are used to solve for each creek parameter (c_i):

$$c_i = \frac{1}{\tilde{t}} \sum_i^{\tilde{t}} \frac{y_{ij}}{t_j}$$

and the maximum likelihood equations are iterated until convergence. Once the parameters are determined, the model is then used to fill in the gaps in the abundance record.

Chapter 3

Comparison of imputation methods

This chapter discusses the strengths and weaknesses of the seven imputation methods described in the previous chapter. The chapter contains a simulation study that investigates performance in the context of the Thompson River coho population with 40.7% of the observations missing between 1976 and 2001. We then examine performance when the missing value pattern is extreme (72.0% missing between 1950 and 1997) by evaluating the sockeye salmon abundance record for an area on the British Columbia North Coast.

3.1 Simulation Study

3.1.1 Review of Coho Life Cycle Ecology

Coho salmon in southern British Columbia have a 3-year life cycle in which they reproduce only once (Sandercock 1991). From November to January, adults migrate from the ocean to their natal streams, where spawning occurs. After spawning, the adult salmon die. Coho fry emerge the next spring and remain in fresh water usually for one year before migrating to the sea as smolts. This one-year residency in creeks is a potential bottleneck through limited carrying capacity specific to each creek. The majority of these fish remain in the ocean for 18 months before returning to fresh water to begin the three-year cycle again (Sandercock 1991).

Between 1976 and 1990, many spawning coho populations in British Columbia were relatively healthy. In the past decade, however, there has been a considerable decline in numbers of spawning coho salmon in some areas, including the Thompson River where abundance has declined as much as 90% (Bradford and Irvine 2000). This decline is thought to be in large part due to poor marine survival. Specifically, these declines have been correlated with various ocean parameters including upwelling and nearshore temperatures, which cause declines in body size, fecundity and proportion of females (Bradford and Irvine 2000, Nickelson et al. 1994). We incorporated the three-year life cycle and the population trend for this aggregate into the design of the simulation study with spawning abundances based on the Thompson River coho records.

3.1.2 Details of Simulation

The goal of the simulation study was to evaluate the performance of the infilling methods by simulating data to approximate the Thompson River coho salmon records. We generated a data array of 100 matrices each with simulated data for 89 creeks over 26 years. These data were based on a stock-recruitment curve estimated from Black Creek on Vancouver Island, the most reliably observed wild coho salmon population in B.C. (Routledge and Irvine 1999). The stock-recruitment curve was generated through three parameters chosen to specify:

r the ratio of expected number in year j to the number in year $j-3$,

k the carrying capacity in terms of number of spawners needed to fill the freshwater habitat with fry, estimated as the average number of fish in the creek during the first 10 years of records (from 1976 to 1985) when coho salmon populations were considered stable,

rk the maximum number of returning adults that can be produced, and

τ an extra variance parameter.

Specifically if y_{ij} is the number in creek i in year j, then the stock-recruitment curve was modeled by

$$\begin{aligned} y_{ij} &= r y_{i,j-3} \text{ for } y_{i,j-3} \leq k, \\ &= rk \quad\quad \text{for } y_{i,j-3} > k, \end{aligned}$$

where $y_{i,j-3}$ is the number of salmon spawning in the previous generation. Thus the parameter r is the ratio of population size from one parent generation to the next, given that the population size is below the carrying capacity of the stream and that resources are unlimited. This parameter was set at 1.9 which was based on data collected from Black Creek (Routledge and Irvine 1999). It was set at the same level for all 89 creeks in the simulation.

The carrying capacity parameter (rk) is the maximum sustainable number of adult returns for each creek. If a creek's population was below the carrying capacity, it would approach rk in the subsequent generations, increasing by the ratio r=1.9 per generation. If more coho salmon were in the spawning areas than k, then the number of salmon in the following generation could be no more than if k salmon had been on the spawning grounds. This number k is fixed for each creek and specifies the number of spawning salmon required to fully stock the creek with juvenile salmon.

For this simulation study we introduced chance fluctuations about the recruitment curve using the Poisson-inverse-Gaussian distribution. The PIG distribution is widely used as a parametric model for extra-Poisson variability. Unlike its major competitor, the negative binomial, it can have a long right-hand tail without a sharp spike at zero (Dean et al. 1989). It is also easier to manipulate analytically than the Poisson log-normal distribution, which can nonetheless also assume the above form.

Consider the mixed Poisson model where:

$$f(Y = y) = \int_0^\infty e^{-\nu\mu} \frac{[\nu\mu]^y}{y!} g(\nu) d\nu, \quad y = 0, 1, 2, ..., \tag{3.1}$$

Y, the number of spawning salmon, has a conditional Poisson distribution with mean $\nu\mu$. Here ν is a random effect, and $g(\nu)$ is a probability density function such as the inverse-Gaussian density:

$$g(\nu) = \frac{1}{\sqrt{2\pi\tau\nu^3}} e^{\frac{-(\nu-1)^2}{2\tau\nu}} \quad , \nu > 0 \tag{3.2}$$

The distribution of Y (3.1) then has a Poisson-inverse-Gaussian (PIG) distribution with mean and variance functions: μ and $\mu(1+\mu\tau)$, respectively. For a PIG distribution, τ is the variance of the random effect, ν, and dictates the amount of extra Poisson variation. The value of τ was determined for each creek from the variability of the first 10 years of data (σ^2) and the number of spawning salmon three years before $y_{i,j-3}$.

$$\tau = \frac{\sigma^2}{y_{i,j-3}^2}.$$

This extra-Poisson parameter is desirable here because of the considerable environmental variability inherent in the system.

Because of the three-year generation time for coho salmon, values in the time series for each creek were simulated based on the observation three years before ($y_{i,j-3}$) and the extra variance parameter τ. The number of fish at time t_j was generated by multiplying the value three years before $y_{i,j-3}$ by the parameter r. If the value three years before was greater than the carrying capacity, then k was used instead of $y_{i,j-3}$.

$$y_{i,j} \sim \mathrm{PIG}(ry_{i,j-3}, \tau_i) \quad or \quad \sim \mathrm{PIG}(rk_i, \tau_i)$$

An extra 20 years of data prior to the actual start of the record were generated. This was to ensure that the population had reached an equilibrium around rk at the start of the comparative analysis.

Impact of declining marine survival

Bradford and Irvine (2000) suggest that productivity of the Thompson River coho are being negatively affected by changes in ocean conditions. Thus, a progressive loss in marine survival could account for the observed population decline in coho spawning numbers.

We reproduced the decline of spawning coho salmon in the Thompson River system over the past decade (Bradford and Irvine 2000) by using a logistic decline equation. We used the Verhulst-Pearl equation to introduce the decline in the carrying capacity and the intrinsic growth rate of each creek (Renshaw 1991, p.50). These changes were introduced to approximate the decline in survival of Thompson coho salmon over the past decade and to evaluate the performance of the different imputation methods in detecting the trend and in imputing reasonable values for missing data. The decline followed a deterministic logistic curve with a rate of decline:

$$\frac{dY(t)}{dt} = -RY\left(1 - \frac{Y}{rk}\right)$$

where R gives the rate of decline of the spawning population $Y(t)$, rk is maximum number of adult returns and

$$Y(t_0) = \frac{rk}{1+e^{-C}} \approx rk \qquad \text{(for some large positive } C\text{)}.$$

The number of fish at time t is given by integrating with respect to t:.

$$Y(t) = \frac{rk}{1+e^{-C}e^{R(t-t_0)}} \tag{3.3}$$

For a declining curve, the intrinsic growth parameter was arbitrarily set at $R = 0.5$, and the onset of decline was set at $t_0 = 11.5$, the constant was set at $C = 6$ to give from (3.3) the following logistic equation for Y:

$$Y(t) = \frac{rk}{1+e^{-6}e^{0.5(t-11.5)}}$$

During the decline phase, the extra-variance component τ was held constant, as was the ratio r. This was to ensure that the PIG variability declined as the population declined and to ensure that the extra variation was a constant multiple of the expected population size.

The simulation was thus set to compare the seven imputation methods. We used the same missing data pattern as existed in the Thompson River dataset for each of the 100 simulated data matrices. We then infilled the gaps with values from the seven different imputation methods.

3.2 Evaluating Performance of Methods from the Simulation Study

3.2.1 Methods

Plotting Annual Totals

We plotted the yearly totals averaged over all creeks and all 100 simulations for each of the seven imputation methods against the average yearly totals from the known model (Figures 3.1, 3.2). This was done to visually assess how the imputation methods were performing on average. In particular, we were interested to see if there were consistent biases associated with any of the methods.

Jackknifing Dependent Samples

We used a jackknifed sums-of-squares estimate of dependent samples as described below to evaluate three components of error from the infilling methods relative to the known model. In this simulation, we know the actual model from which the data were generated. Thus it is straightforward to compare performance of imputation methods.

We used the following notation in the sums-of-squares equations:

Y_{ijk} is the number of fish in the i^{th} creek, j^{th} year and k^{th} simulation, and is an element in the $89 \times 26 \times 100$ array generated from the known model. Therefore, $\bar{Y}_{.jk} = \frac{1}{89}\sum_{i=1}^{89} Y_{ijk}$ is the average number of fish across all 89 creeks for the j^{th} year and the k^{th} simulation and $\bar{Y}_{.j.} = \frac{1}{100}\sum_{k=1}^{n} \bar{Y}_{.jk}$ tracks the average decline across all years in the known model.

Z_{ijk} is the number of fish in the i^{th} creek, j^{th} year and k^{th} simulation with gaps filled by imputation, and is thus an element in the $89 \times 26 \times 100$ imputed array. Likewise, $\bar{Z}_{.jk} = \frac{1}{89}\sum_{i=1}^{89} Z_{ijk}$ is the average number of fish across all 89 creeks for the j^{th} year and the k^{th} simulation and $\bar{Z}_{.j.} = \frac{1}{100}\sum_{k=1}^{n} \bar{Z}_{.jk}$ tracks the average decline across years in the imputed array.

We used the following equations in calculating three components to the error sums of squares (bias, chance error, and total error) for each imputation method.

$$SS_{bias} = \sum_{j=1}^{26} \frac{(\bar{Z}_{.j.} - \bar{Y}_{.j.})^2}{(\bar{Y}_{.j.})^2}$$

where $\bar{Y}_{.j.}$ and $\bar{Z}_{.j.}$ should be very close and SS_{bias} small if there is little systematic bias. The denominator is a weighting factor to account for the greater variability in years with more abundant fish.

$$SS_{chance} = \frac{1}{100} \times \sum_{j=1}^{26} \sum_{k=1}^{100} \frac{(\bar{Z}_{.jk} - \bar{Z}_{.j.})^2}{(\bar{Y}_{.j.})^2}$$

where SS_{chance} gauges how much chance variation there is about the means.

$$SS_{total} = \frac{1}{100} \times \sum_{j=1}^{26} \sum_{k=1}^{100} \frac{(\bar{Z}_{.jk} - \bar{Y}_{.j.})^2}{(\bar{Y}_{.j.})^2}$$

where SS_{total} is algebraically the sum of the above two equations, and is a gauge of the overall error.

Because the imputed arrays for each of the seven imputation methods are based on the same simulated dataset, the arrays are not independent samples. Therefore we made pairwise comparisons and tested the null hypothesis that these limiting values are the same for any pair of imputation methods, i.e., that their differences is zero.

The sums of squares are calculated from only 100 simulations, therefore we were concerned that the sum-of-squares estimates might be biased estimates for an indefinitely large number of simulations. Therefore we elected to obtain numerical approximations for the sums-of-squares estimates using a jackknife estimation procedure. This jackknife procedure is now described before relating back to the sums-of-squares procedure.

The jackknife procedure consists of taking repeated subsamples of the original sample of n independent observations by omitting a single observation at a time. Thus, each subsample consists of $(n-1)$ observations formed by deleting a different observation from the sample. The jackknife estimate and its standard error are then calculated from these truncated subsamples. For example, suppose θ is the parameter of interest and $\hat{\theta}_{(1)}$,$\hat{\theta}_{(2)}$, ..., $\hat{\theta}_{(n)}$ are estimates of θ based on n subsamples, each of size $(n-1)$ as calculated as below. The jackknife estimate of θ is calculated as the mean of the subsample estimates of θ:

$$\hat{\theta}_{(.)} = \frac{\sum_{k=1}^{n} \hat{\theta}_{(k)}}{n} \tag{3.4}$$

The jackknife estimate of the standard error of $\hat{\theta}_{(.)}$ is

$$SE\,(\,\hat{\theta}_{(.)}\,) = \sqrt{\frac{var\left[\,\hat{\theta}_{(k)}\,'s\,\right]}{n}} \tag{3.5}$$

In our jackknife procedure, we tested the hypothesis that the expected difference in sums of squares was equal to zero: $H_0 = E(\theta) = 0$, where $\theta = SS_{method\,A} - SS_{method\,B}$, and A and B are from methods 1 through 7 as described in Chapter 2. The parameter of interest is then θ and the asymptotic expansion of the raw estimate's expectation is:

$$E(\hat{\theta}_n) = \theta + \frac{a}{n} + \frac{b}{n^2} + ...$$

and of raw estimate based on a sample of (n-1) values is:

$$E(\theta_{n-1,k}) = \theta + \frac{a}{n-1} + \frac{b}{(n-1)^2} + ...$$

Assuming that terms higher than first order are negligible, this gives two equations:

$$\begin{aligned} nE(\hat{\theta}_n) &= n\theta + a\ , \\ \text{and}\ \ (n-1)E(\hat{\theta}_{n-1,k}) &= (n-1)\theta + a\ . \end{aligned}$$

These can be subtracted to obtain $\hat{\theta}_{(k)}$ an unbiased estimate of θ with the bias term of $\frac{1}{n}$ eliminated:

$$\hat{\theta}_{(k)} \;=\; n(\hat{\theta}_n) - (n-1)(\hat{\theta}_{n-1,k}) \tag{3.6}$$

We calculated this for the $n = 100$ jackknifed estimates of θ, $\hat{\theta}_{(k)}$ for $k = 1, \ldots, 100$. The $\hat{\theta}_{(k)}$ are functions of U-statistics and by jackknifing a U-statistic, we get an asymptotically normally distributed random variable with a mean as in (3.4) and standard error as in (3.5) (Arveson 1969).

The differences in sums of squares between methods A (SS_A) and B (SS_B) in bias, chance error and total error following (3.6), are estimated by:

$$\bar{\hat{\theta}}_{(.)} = \frac{1}{100} \times \sum_{k=1}^{100} \left[100 \left(SS_{A;\, non-jack.} - SS_{B;\, non-jack.} \right) - 99 \left(SS_{A;\, jack.[k]} - SS_{B;\, jack.[k]} \right) \right]$$

were SS_{jack} and $SS_{non-jack}$ are the jackknifed and non- jackknifed sums-of-squares estimates respectively.

Also, the relationship between the jackknifed differences in sums of squares is:

$$\hat{\theta}_{(.)\ bias} + \hat{\theta}_{(.)\ chance} = \hat{\theta}_{(.)\ total}.$$

B ↓	A →	ZI	NV	I/E	AS	TL	PM	GM
ZI	$\hat{\theta}_{(\cdot)bias}$ ($st.error$)	0 (0)						
	$\hat{\theta}_{(\cdot)chance}$ ($st.error$)	0 (0)						
NV	$\hat{\theta}_{(\cdot)bias}$ ($st.error$)	**1.45** (0.02)	0 (0)					
	$\hat{\theta}_{(\cdot)chance}$ ($st.error$)	-0.04 (0.02)	0 (0)					
I/E	$\hat{\theta}_{(\cdot)bias}$ ($st.error$)	**1.47** (0.02)	**0.02** (0.006)	0 (0)				
	$\hat{\theta}_{(\cdot)chance}$ ($st.error$)	-0.290 (0.13)	-0.25 (0.13)	0 (0)				
AS	$\hat{\theta}_{(\cdot)bias}$ ($st.error$)	**1.46** (0.02)	0.005 (0.07)	-0.02 (0.007)	0 (0)			
	$\hat{\theta}_{(\cdot)chance}$ ($st.error$)	0.02 (0.05)	0.02 (0.05)	0.27 (0.10)	0 (0)			
TL	$\hat{\theta}_{(\cdot)bias}$ ($st.error$)	**1.42** (0.02)	**-0.03** (0.005)	**-0.06** (0.007)	**0.04** (0.004)	0 (0)		
	$\hat{\theta}_{(\cdot)chance}$ ($st.error$)	-0.06 (0.02)	**0.10** (0.02)	0.35 (0.13)	-0.08 (0.04)	0 (0)		
PM	$\hat{\theta}_{(\cdot)bias}$ ($st.error$)	**1.48** (0.02)	**0.02** (0.004)	0.00 (0.007)	**0.02** (0.005)	**0.06** (0.003)	0 (0)	
	$\hat{\theta}_{(\cdot)chance}$ ($st.error$)	**-0.24** (0.03)	**-0.20** (0.02)	0.05 (0.01)	**-0.22** (0.04)	**-0.30** (0.02)	0 (0)	
GM	$\hat{\theta}_{(\cdot)bias}$ ($st.error$)	**1.47** (0.02)	0.02 (0.005)	-0.009 (0.007)	0.01 (0.005)	**0.05** (0.003)	**-0.01** (0.001)	0 (0)
	$\hat{\theta}_{(\cdot)chance}$ ($st.error$)	0.03 (0.02)	0.07 (0.02)	0.32 (0.13)	0.05 (0.04)	**-0.04** (0.005)	**0.26** (0.02)	0 (0)

Table 3.1: Jackknifed estimates of differences in error sums of squares between methods A and B. Zero-infilled (ZI), Nearest-value (NV), Interp./Extrap. (I/E), Average of scaled values (AS), Transformed linear model (TL), Poisson model (PM), Gamma model (GM). **Bold-face text** represents significant differences at the 5% level with Bonferroni adjusted p-values. Positive values indicate method A has greater error sum of squares, negative values indicate method B has greater.

Figure 3.1: Annual averages of numbers of spawning coho salmon in the Thompson River system of British Columbia; Zero-infilled, Nearest-value, Interpolation/extrapolation, Averaging scaled values. Standard error bars were calculated as the square root of the variance divided by the number of datapoints for that year.

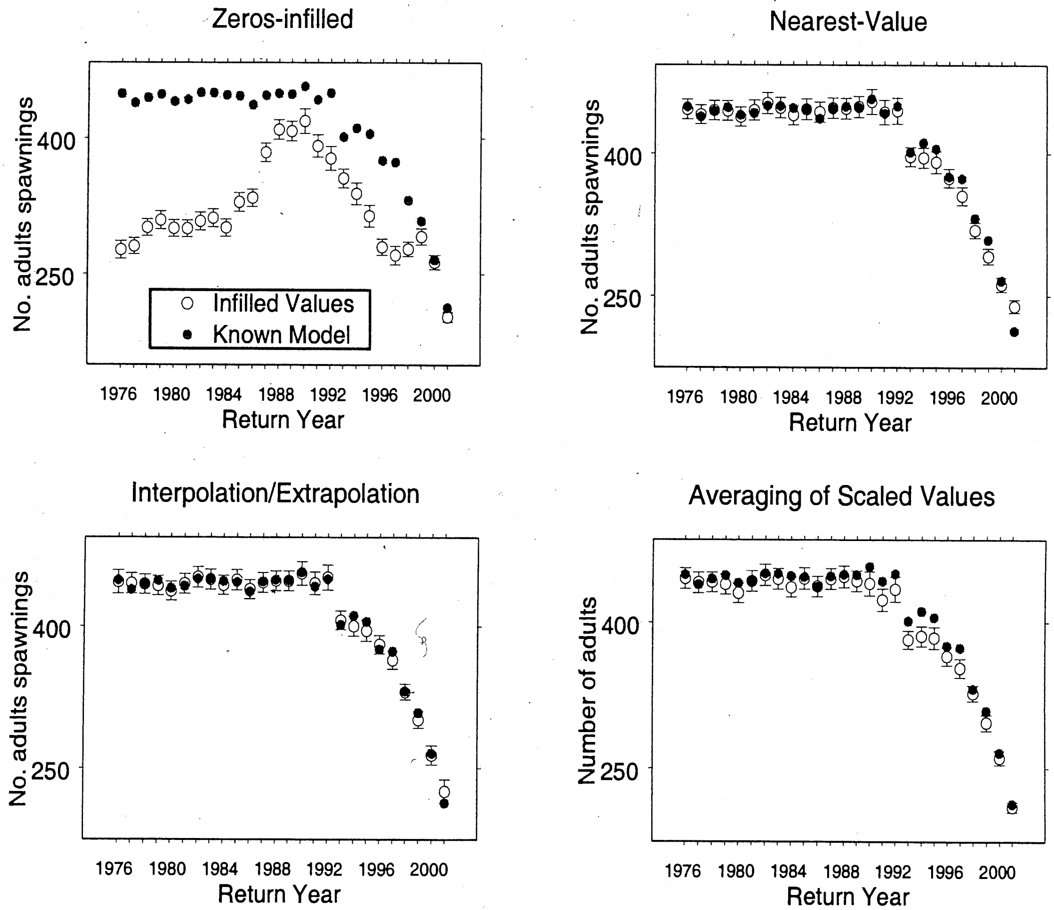


Figure 3.2: Annual averages of numbers of spawning coho salmon in the Thompson River system of British Columbia; ANOVA-based methods. Standard error bars were calculated as the square root of the variance divided by the number of datapoints for that year.

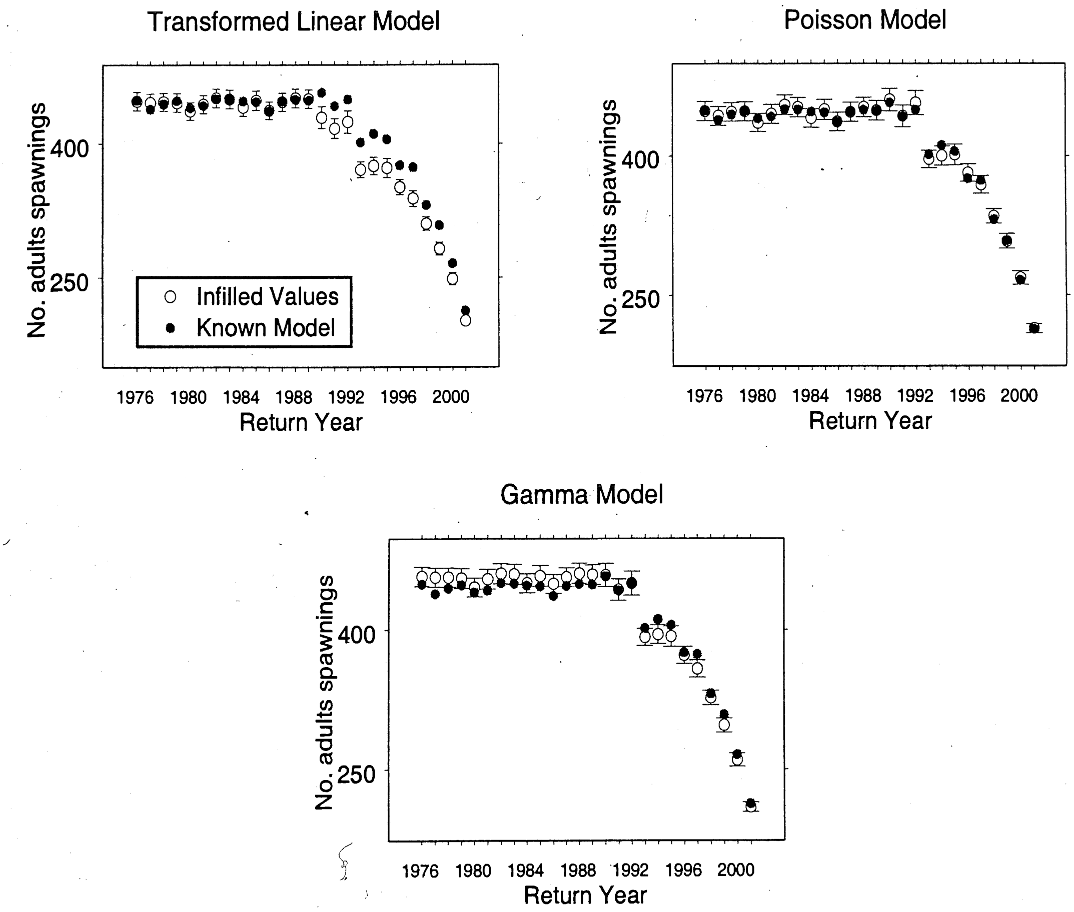


Figure 3.3: Plot of differences between the infilled values and the known values for the seven imputation methods

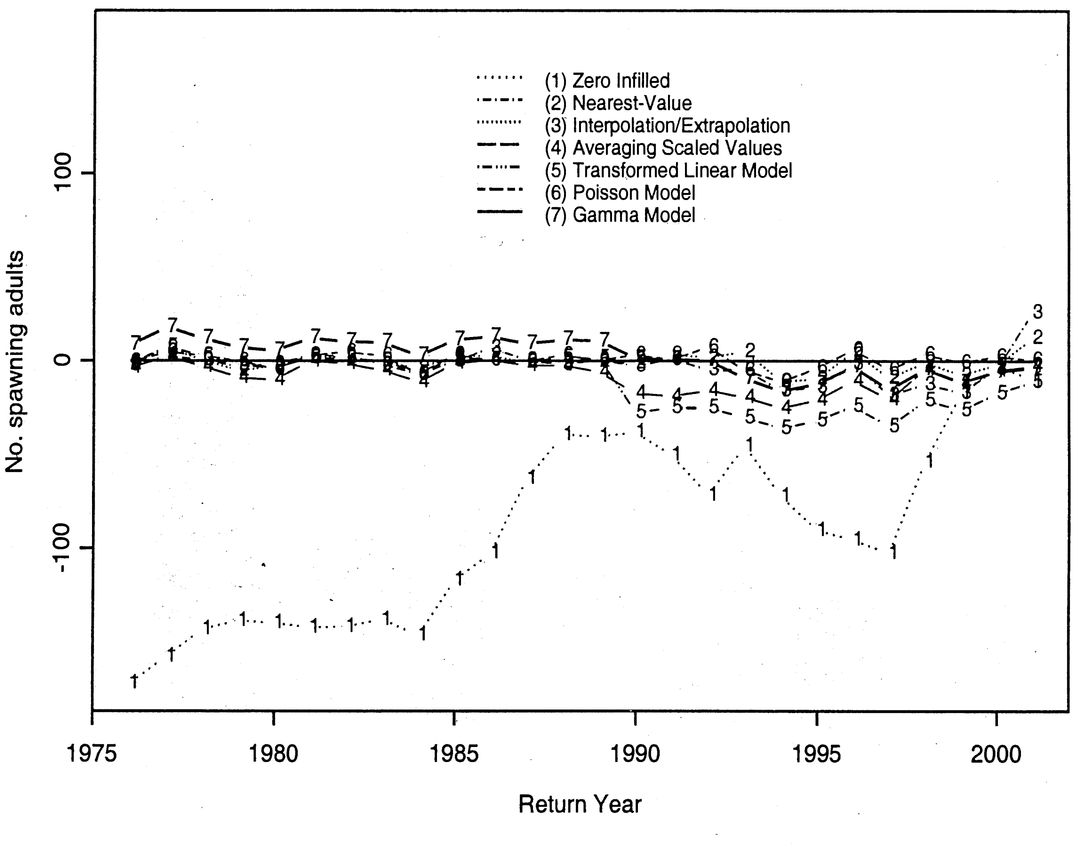
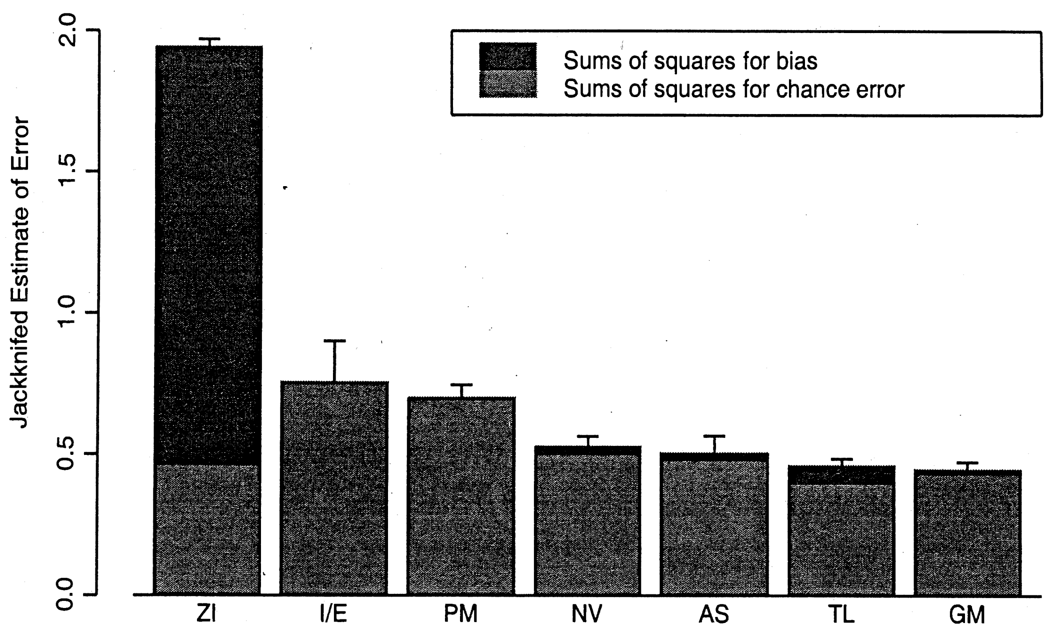


Figure 3.4: Jackknifed Bias and Chance Error Sums of Squares with Standard Error bars calculated from the Overall Error Sums of Squares. Zero-infilled (ZI), Nearest-value (NV), Interp./Extrap. (I/E), Average of scaled values (AS), Transformed linear model (TL), Poisson model (PM), Gamma model (GM)



3.2.2 Results

Of the seven methods, and for the pattern of missing values observed in the Thompson coho spawning records, three methods are not as effective interpolators as the remaining four. We found the zero-infilled, nearest value, and interpolation/extrapolation methods unsuitable, whereas the averaging of scaled values, tranformed linear, Poisson, gamma models seem to perform better. (Table 3.1) gives the jackknifed sums-of-squares estimates of bias and chance error for each pair of the seven imputation methods. Insight from this error analysis and comparative data exploration aided in selecting the better methods. Interpretation of results is presented below.

Zero-infilled

The zero-infilled method is clearly inadequate. The assumption that when data are missing there were no fish, is clearly untrue and introduces an obvious and unacceptable bias. The jackknifed sums-of-squares analysis (Figure 3.4) and the graph of annual totals (Figure 3.1 and Figure 3.3) show clearly the degree of bias introduced by this method.

Nearest value

Although the performance of the nearest-value method as indicated by the jackknifed sums-of-squares analysis is not unreasonable, we have concerns about bias in this method, particularly in the decline phase (Figure 3.1). A major concern is that this method does not allow for trends to be observed within individual creeks when a gap is to be filled. Only jumps between horizontal steps are possible. This may be the cause of the bias evident in the decline phase of Figure 3.1. Furthermore, because infilled values rely only on one other data point, an unrepresentative observation combined with large data gaps will cause this method to work poorly. Therefore there are better tools for monitoring trends in populations.

Interpolation/Extrapolation

Although the interpolation/extrapolation method has a small bias component to the total error, the standard error of the method is very large. Thus the method is unreliable. Furthermore, when the imputation involves extrapolation, impossibly large, or worse, negative infilled values are possible. As a result, we regard this method as unsuitable.

Averaging of Scaled Values

This method performs adequately in this sums-of-squares analysis. However, there are some concerns about this method as it consistently underestimates the population in the decline phase and introduces a non-trivial source of bias (Figure 3.1). In particular, infilled values are sensitive to the maximum observed value in that they can never be greater than this observed maximum.

Gamma Model

Although this method tends to have smaller bias and chance error sums of squares than the Averaging of Scaled Values method, it is not statistically significant at $\alpha = .05$ from this method in any of the three sums of squares components (Table 3.1). However, according to the graphs of annual averages (Figure 3.2 and Figure 3.3), imputed values are consistently higher than the known data during the stable population period, and consistently lower during the decline phase. There are two problems with fitting these data to the gamma distribution. The first is that this kind of datum is discrete and the family of gamma distributions are continuous. A second concern with this model is that it entails assuming that the variance is proportional to the square of the mean. When the mean is small, the model used in the simulations will contain a non-negligible linear term as well. For datasets containing substantial numbers of small entries, especially zeros, as is the case here, this appears to detract from the method's performance. Nonetheless, for reasons we do not yet comprehend, this disadvantage seems to disappear for more controlled patterns of missing entries, as demonstrated in Chapter 4.

Transformed Linear Model

The responses after the transformation were assumed to be normally distributed, have constant variance independent of the mean, and have systematic effects that combine additively. A single transformation is asked to produce three things simultaneously. It should not be surprising then that this method is less effective than the Poisson method described below. The sums-of-squares analysis showed that a large component of the error in this infilling method is attributed to bias (Figure 3.4). The graph of annual totals shows that the bias is substantial in the decline phase indicating that this method is particularly poor at following downward trends.

Poisson Model

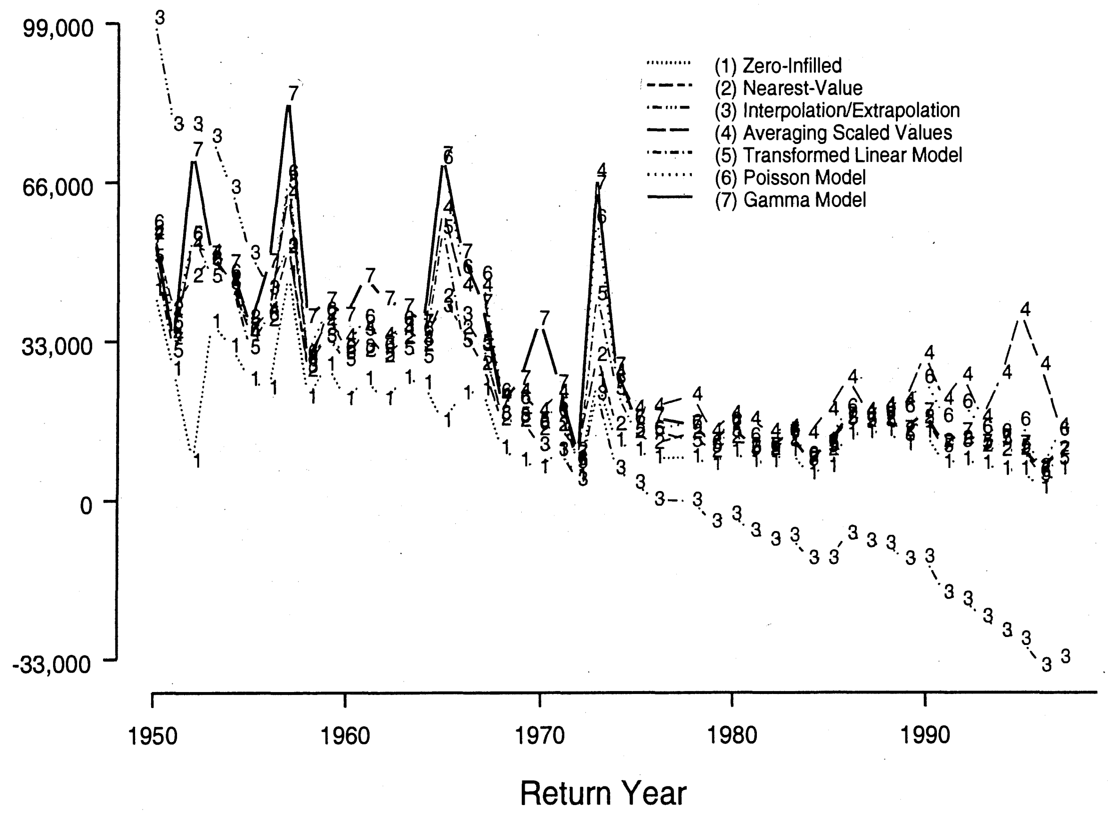
A big advantage over the traditional transformation approach in the Poisson GLM approach is the freedom to specify the variance to mean relation and the error distribution. Therefore, when GLM's are used to fit the data, a single transformation is not trying to do several jobs. The sums-of-squares analysis shows that a small component of the error is attributable to bias. Nonetheless, this method gives a larger chance error component than the transformed linear model method. We would preferentially select a method with a larger amount of chance error over one with a large bias component particularly. The graphs of annual totals (Figure 3.2) suggests this method works the best on average.

3.3 Evaluating methods when the missing data pattern is extreme: North Coast sockeye

These sockeye salmon (*O. nerka*) spawning records were collected between 1950 and 1997 for 68 creeks in Fisheries and Oceans' statistical district 8 on the North Coast of British Columbia. Sampling intensity could be described at best as sporadic with

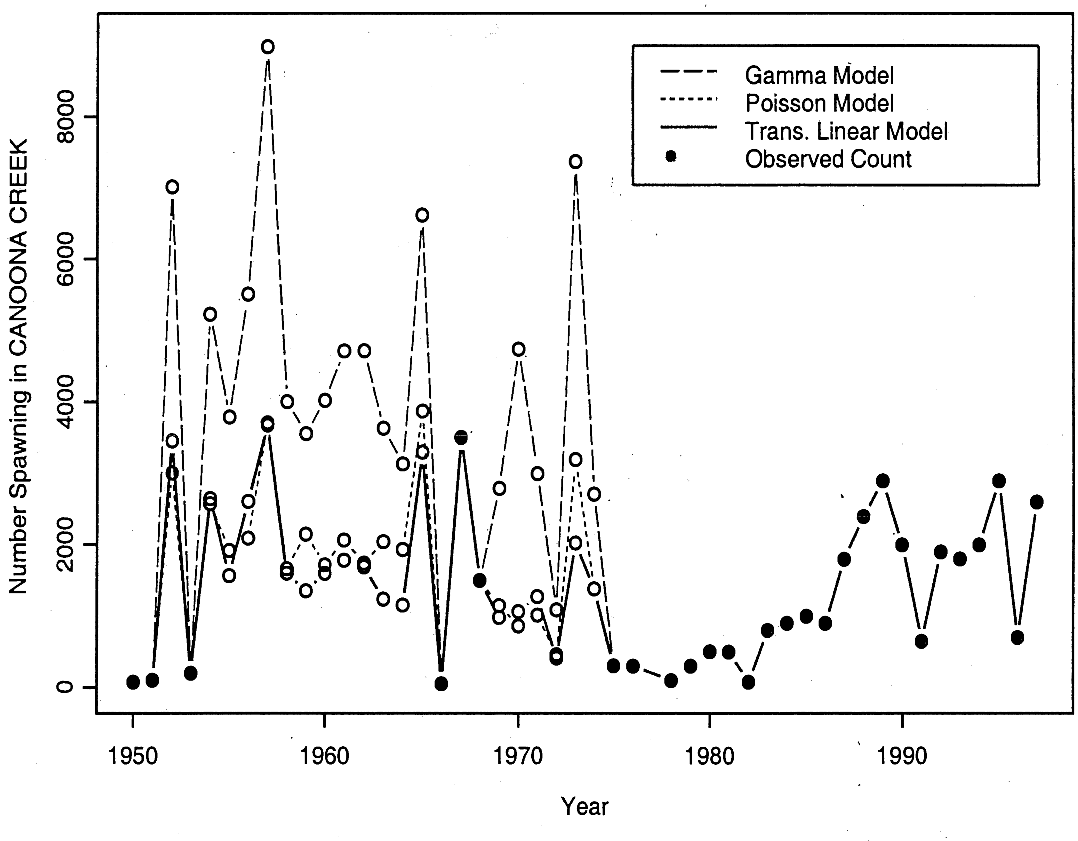
72.0% of the records missing. The scarcity of records is particularly high in the latter half of the series. This large number of missing values and the lack of balance in this pattern caused problems with many imputation methods. In addition, the lack of area-wide consistency in trends across creeks and years causes instabilities with the more sophisticated methods. These instabilities are examined here.

Figure 3.5: Annual Totals of spawning salmon for 68 creeks on the North Coast of British Columbia



We plotted the total numbers of sockeye salmon for the North Coast area as estimated by the seven imputation methods (Figure 3.5). From this plot, it is obvious

Figure 3.6: Annual numbers of spawning sockeye salmon for Canoona Creeks on the North Coast of British Columbia



that three of the methods do not perform well.

Zero-Infilled, Interpolation/Extrapolation, Averaging of Scaled Values Methods

The zero-infilled method substantially underestimates the number of fish spawning, particularly in the later years where the number of observations was lowest.

The weakness of the interpolation/extrapolation method is clearly seen in (Figure 3.5). At either end of the data record where the method extrapolates rather than interpolates, absurdly large and negative annual totals are obtained.

The right-hand section of (Figure 3.5) shows that the Averaging Scaled Values method predicts substantially higher totals than any other method. As most of the data are more complete for the first half of the record, and most of the data have a roughly decreasing trend, the creeks that contain data in the latter years have a big influence on the infilled values. This method fails here because some of the creeks that have more complete data records for the later dates, also have an increasing trend. This causes the average scaled column value to be unduly large, which results in the missing values being inflated without much apparent justification. This is clearly seen in the latter half of the series in (Figure 3.5).

Transformed Linear, Poisson and Gamma Models

The lack of consistent trends across all the creeks creates problems for those methods that model creek and year effects but ignore the interaction effects (transformed linear, Poisson and gamma models). Inconsistent trends across creeks correspond to interactions between years and creeks, and because these interactions cannot be modeled, the imputations are untrustworthy. To demonstrate, we plotted the imputed values of these three methods for Canoona Creek (Figure 3.6). Canoona Creek is a creek unlike most of the other creeks in this area as the sockeye population does not decrease over the time period and missing values are mostly in the first half of the time series. In particular, the gamma model method does a poor job.

The transformed linear model method acts as if the residuals are lognormal. Hence, a large, positive residual will not be unexpected as the lognormal distribution has a long, right-hand tail. The GLM of the Poisson and gamma models minimize weighted sums of squares on an untransformed scale, and hence a large, positive residual will be less consistent with the model. These two methods will react to a large positive residual by increasing the estimated "year" effect, and hence produce the larger peak. The variance in the gamma model is proportional to the square of the mean, and therefore the compensation is less in this model, and the peaks are higher in Cannoona Creek.

Nearest Value

The nearest value method considers only the information from the closest value for that creek. It cannot include information from other creeks for a given year to help with the imputation. In this circumstance however, when there are conflicting trends occuring within the same area, the nearest value method performs moderately well.

3.3.1 Chapter Summary

If the pattern of missing values is very extreme and if the time trends are inconsistent across different spawning creeks, then even the best methods fail. This emphasizes the importance of a good long-term sampling design, an issue to which we now turn.

Chapter 4

Survey Design

Accurate monitoring of salmon populations is a critical step in ensuring the persistence of a sustainable population of a species. This chapter emphasizes the need for a good sampling design for monitoring of populations. In particular, we focus on a pattern of missing values, called a balanced pattern, for which the ANOVA-based methods will work well. We stress the benefits of balanced designs, in which their optimal properties make them an important component in surveys with missing values. We also take a brief look at further refinements that might be appropriate in light of the three-year life history, followed by some summary conclusions.

A Simulation Study Showing the Importance of Balance in Design

Balance in this context refers to designs in which the pattern of missingness is "balanced" across years. Because there is value in sampling larger creeks more frequently than smaller creeks, balance can be modified within creeks such that sampling larger creeks occurs more frequently. This can be approximately achieved through taking weighted independent samples of creeks in each year.

We designed a simple scenario to demonstrate the importance of balance in survey design. Using the same 100 data matrices as were generated for the Thompson River coho simulation in Chapter 3, we removed the same proportion of observations from the record but redistributed the missing pattern such that the pattern was balanced throughout the creeks and years. We removed 38 of the 89 observations for each of the 26 years and for all of the 100 simulated data matrices. Furthermore, we assigned weights to the sampling scheme. Three weight categories were introduced based on creek size: small, medium and large to which the weights 1, 2, and 3 were assigned respectively. The proportion of creeks within each category was roughly equivalent. We then used a random number generator to take a weighted random sample without replacement from the 89 creeks.

To contrast with the balanced sampling scenario, we devised another sampling scheme with the same proportion of missing values per year (38/89) but with an unbalanced pattern. We divided each of the three abundance-based categories of creeks into two, roughly equal subsets (Table 4.1). We also divided the range of years for which the data are used into two roughly equal halves. Within each abundance-based category of creeks, we selected one half of the creeks to have a high portion of the values missing from the first half of the year range, and no missing values from the second half of the year range. Then for the other half of the creeks in the abundance category, we reversed the pattern such that there were no missing values in the first half of the year range and a high portion of missing values in the second half of the year range. The weighting scheme was observed as well as possible.

The changes in jackknifed sums of squares for bias and chance error were substantial for infilled values from the Poisson and gamma models. It is evident that balance affects the degree of chance error more in the Poisson model (almost 3×) than for the gamma model (less than 2×; Table 4.2) and bias is affected more in the gamma model than the Poisson model (Figure 4.1). If the design is balanced, then the amount of bias in the gamma model becomes negligible and it is a better model to use. This is because the precision of the infilled methods is better as the chance error component is almost a third that of the Poisson model. However, if the design is unbalanced then the GLM method has a smaller bias, and is a relatively better model to select.

The sampling pattern for the Thompson River coho dataset had an added dimension of imbalance in the above simulation. In some years, creeks were sampled intensively and in others not, so that there might be 89 creeks sampled in one year

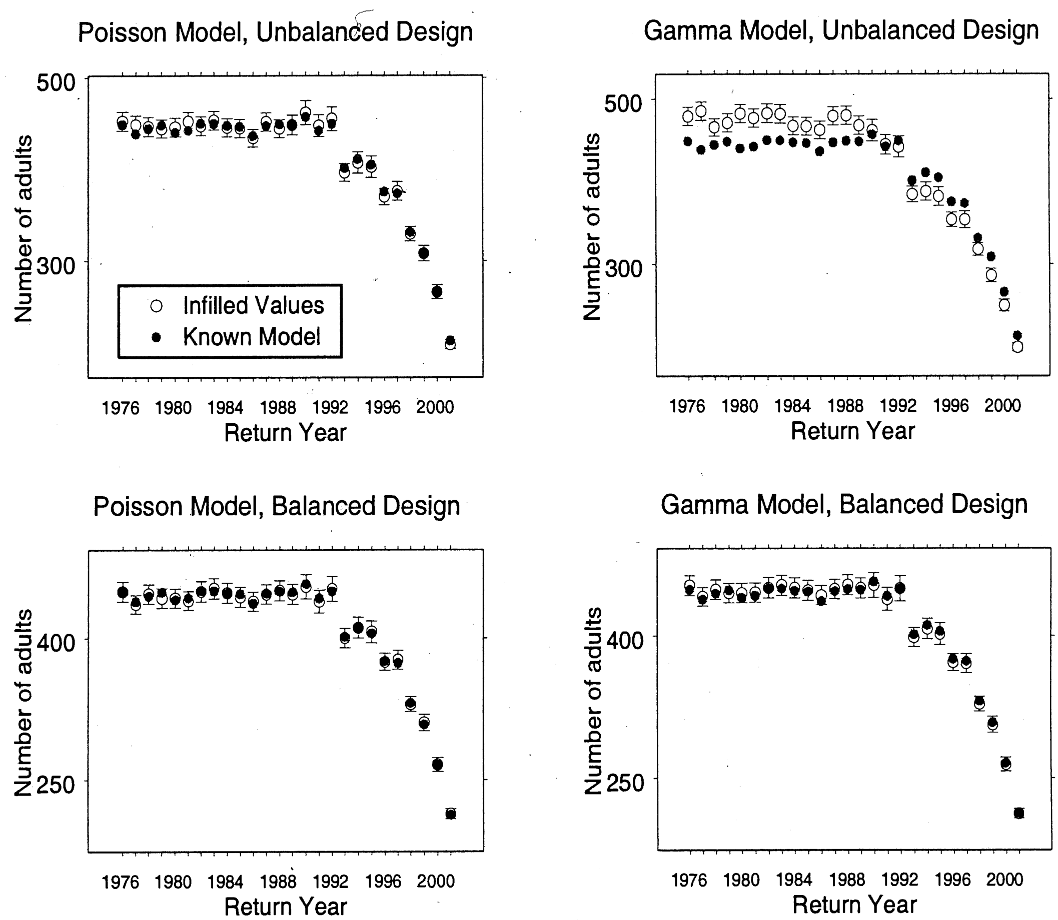
	1^{st} 13 years	2^{nd} 13 years
weight 1	no missing values	94% missing
	94% missing	no missing values
weight 2	no missing values	85% missing
	85% missing	no missing values
weight 3	no missing values	71% missing
	71% missing	no missing values

Table 4.1: Missing pattern used in the simulation of an unbalanced sampling design

		$\hat{\theta}_{bias}$ *(st. error)*	$\hat{\theta}_{chance}$ *(st. error)*
Balanced	**Poisson Model**	0.00136 (0.013)	0.237 (0.022)
	Gamma Model	0.0009 (0.0011)	0.134 (0.018)
Unbalanced	**Poisson Model**	-0.0009 (0.0020)	0.623 (0.048)
	Gamma Model	0.0838 (0.013)	0.218 (0.023)

Table 4.2: Comparison of jackknifed sums-of-squares errors from a balanced and unbalanced sampling design

Figure 4.1: Comparing bias in the infilled values from a GLM using the gamma distribution when the design is balanced and unbalanced.



and 10 the next. Our simulations did not consider this added dimension of imbalance, but it is reasonable to anticipate that this would have produced even more substantial errors in the infilled estimates. This simple demonstration highlights the serious need for better design criteria.

The measure of efficiency is a scale free measure that is used to compare different sampling designs. Efficiency is calculated as the average of the estimated variances for all pairs of treatments or in our case years (John 1971, p. 10). In the simulations discussed here, the balanced and unbalanced designs have the same efficiency. Improvements in the efficiency of sampling design could be made if a balanced incomplete block (BIBD) design could be selected. If E is the efficiency factor of an incomplete block design (IBD) as measured by

$$E = \frac{\lambda t}{mn}$$

where:

λ is the number of times each pair of creeks appear in the same year and must be an integer,

t is the number of years,

n is the number of creeks measured in a given year

m is the number of years measured for a given creek; $m < t$

then for all incomplete block designs E is less than one. But in the class of designs of block size m, and t years (treatments), the most efficient design is the balanced incomplete block design if one exists (Kempthorne 1951). The efficiency of a BIBD (John 1971, p.265) is:

$$E = \frac{\lambda t}{mn} = \frac{t(m-1)}{m(t-1)}$$

where efficiency only depends on t and m. For a given t, efficiency is an increasing fuction of m and therefore the number of measurements per creek (m) should be as large as is reasonable (John 1971, p.265). Values of E may be useful in deciding on the best model.

Connectivity

Related to the notion of balance is the concept of connectivity. The analysis of the North Coast sockeye dataset (Section 3.3) was an example of how lack of connectivity combined with inconsistent trends can cause the best methods to fail. For example, if creeks A & B are censused in year 1, and creeks B & C are censused in year 2, then creeks A & C are connected. The relationship A connected to C is an equivalence relation which forms disjoint equivalence classes for the treatments. A design is said to be connected if there is one equivalence class (i.e., if every pair of treatments is connected). Problems arise when creeks A & B are censused in year 1 and creeks C & D in year 2, then creeks A & C, A & D, B & C, and B & D are disjoint and the design is then disconnected. This means that creek effects are confounded with year effects, such that time trends will be indistinguishable from differences in productivity between sets of creeks. This implies that the analysis of variance methods will breakdown completely without connectivity. Therefore, the ultimate goal of survey design should be to maintain conectivity and balance.

Rotating Panel Surveys

Incorporation of the 3-year life cycle of the Thompson River coho into a rotation design could potentially further improve infilled estimates by reducing within creek variability. The variance can often be reduced by using the same sampling units (creeks) in the two successive salmon generations (Kish 1965, p.462). The variance tends to be least when the overlap is high for elements whose correlations are large (Kish 1965, p.470).

For those Thompson River creeks, in which there were enough observed data to do a time series analysis, a lag-3 serial correlation for 9 of the 11 creeks was demonstrated (coefficients of autoregression: 0.46969, 0.44629, 0.44234, 0.37814, 0.31173, 0.20537, 0.20271, 0.16816, 0.05177, -0.02945, -0.11979. If this correlation had been incorporated into the survey design in the form of some kind of rotating panel design with many of the same creeks appearing in samples separated by three years, then further improvements might be achievable (Kish 1965, p.474). A rotating panel survey that considers the correlation structure between succesive generations of salmon is worth investigating.

Chapter 5

Summary of Conclusions

This thesis presents seven imputation methods for infilling missing data into spawning salmon records and examines their performance in three patterns of missing data. For a reasonable amount of randomly missing data, we found the zero-infilled, nearest value, and interpolation/extrapolation methods to be inferior methods compared to the averaging of scaled values, transformed linear Poisson and gamma models to perform better. In particular, our simulation study shows the Poisson model which assumes the variance proportional to the mean to be the most reliable in this context.

When the missing data pattern is extreme and trends are not consistent across all creeks, problems arise for those methods that model creek and year effects but ignore interactions. In these cases, inconsistent trends across creeks correspond to interactions between years and creeks and because these interactions cannot be modeled, imputations are untrustworthy. In this context of imbalance and inconsistent trends, the nearest value method is the most reliable.

Balance, connectivity and consideration of life history are key components in designing an area-wide survey. In the final chapter, we show that the degree of balance is critical in reducing bias and chance error and the reduction appears not to be uniform across all infilling methods. If balance is incorporated into the design, the gamma model outperforms the Poisson model. In light of this, further investigation into modeling using the Poisson-inverse-Gaussian distribution, which is a mixed distribution that is more flexible in its capacity to model the variance component, may prove to be instructive.

Bibliography

[1] ARVESON, W. (1969). Jackknifing U-statistics. *Annals of Math. Stat.* 40:2076-2100

[2] BRADFORD, M.J., J. IRVINE. (2000). Land use, fishing, climate change, and the decline of Thompson River, British Columbia, coho salmon. *Can. J. Fish. Aquat. Sci.* 57:13-16.

[3] COSEWIC. (2000) Canadian Species at Risk, May 2002. *Committee on the Status of Endangered Wildlife in Canada.*

[4] DEAN, C. and LAWLESS, J.F., and WILLMOT, G.E. (1989). A mixed-Poisson-inverse Gaussian regression model. *Can. J. Statist.* 17:171-181.

[5] JOHN, P.W.M. (1971). Statistical Design and Analysis of Experiments. *The Macmillan Company, New York.*

[6] KEMPTHORNE, O. (1956). The efficiency factor of an Incomplete Block Design. *Ann. Math. Statist.* 27:846-849.

[7] KISH, L. (1965). Survey Sampling. *John Wiley & Sons, Inc., New York.*

[8] NICKELSON, T. E., J. NICHOLAS, H. WEEKS, and K. KOSTOW. (1994). Oregon coho salmon biological status assessment. *Oreg. Dep. Fish Wildl., Corvallis, OR.*

[9] RENSHAW, E. (1991). Modelling biological populations in space and time *Cambridge University Press.*

[10] ROUTLEDGE, R.D., and IRVINE, J.R. (1999). Chance fluctuations and the survival of small salmon stocks. *Can. J. Fish. Aquat. Sci.* 56:1512-1519.

[11] SANDERCOCK, F.K. (1991). Life history of coho salmon (*Oncorhynchus kisutch).* p.357-455. in *C. Groot and L. Margolis,* editors. Pacific salmon life histories. *University of British Columbia Press, Vancouver, B.C.*

Appendix 1: Raw Data

1. Thompson River spawning coho records: 1976 - 2001.

2. North Coast spawning sockeye records: 1950 - 1997.

ek no.	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
[1,]	200	10	338	150	100	200	100	100	100	650	500	150	150	500	350	100	100	250	20	70	75	16	40	159	96	48
[2,]	60	50	150	100	475	75	100	200	300	200	500	1100	500	700	1100	NA	100	0	NA	NA	NA	120	105	345	100	128
[3,]	50	NA	NA	75	0	10	50	25	20	50	NA	50	50	50	NA	NA	30	80	NA	NA	0	0	0	0	2	22
[4,]	750	25	70	0	50	60	10	15	20	50	50	60	60	50	40	50	0	100	NA	100	150	32	38	167	39	NA
[5,]	NA	25	40	10	25	16	15	0	0	50	50	50	50	70	35	50	0	50	30	8	0	0	0	0	6	0
[6,]	NA	50	0	50	50	20	55	100	50	10	100	50	0	50	35	35	0	0	NA	0	20	0	NA	0	0	0
[7,]	25	10	0	100	10	60	30	80	20	30	100	30	100	75	100	50	30	20	25	6	10	0	10	0	0	0
[8,]	15	NA	10	40	125	NA	35	225	50	36	NA	100	75	50	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0
[9,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	NA	0	0	NA	NA	0	0	3
[10,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	5	0	NA	NA	0	0	0	0
[11,]	NA	40	2	30	30	10	20	20	20	100	500	60	100	200	150	25	25	150	12	NA	50	0	5	33	22	1
[12,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	39	6	12
[13,]	NA	325	94	400	300	350	250	200	500	750	600	750	750	800	700	350	0	500	12	NA	200	30	0	41	54	14
[14,]	1500	1100	2694	2000	2500	1500	850	1000	1200	7100	3500	4075	10584	9556	5114	3885	1362	3032	761	1485	800	254	150	2512	1789	1901
[15,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	0	NA	0	NA	NA	NA	NA	0	6	0
[16,]	NA	105	0	150	150	70	65	85	70	100	200	250	40	250	300	50	25	200	20	NA	75	0	NA	20	2	16
[17,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	4	3
[18,]	25	25	0	200	75	42	25	50	50	125	0	0	0	150	120	30	75	120	NA	NA	60	2	20	117	23	20
[19,]	NA	NA	NA	15	45	32	30	30	30	20	30	25	25	40	30	0	0	30	NA	NA	25	1	2	20	9	2
[20,]	NA	NA	NA	NA	NA	NA	NA	40	0	100	25	50	50	90	50	50	30	50	24	6	25	8	15	0	4	3
[21,]	25	10	62	10	25	0	25	100	75	25	25	80	120	150	50	0	0	45	60	32	25	0	50	NA	0	7
[22,]	15	NA	10	NA	25	NA	5	NA	30	20	NA	50	80	10	0	NA	NA	NA	NA	NA	NA	NA	NA	115	41	2
[23,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	NA	0	0	30	34
[24,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	25	15	15	0	0	0	0	NA	NA	0	0	NA	0	0	1
[25,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	0	20	0	60	30	25	15	20	NA	NA	NA	0	0	0	0	1
[26,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	24	0	0
[27,]	750	900	1588	1500	2000	1300	500	800	1000	1550	3800	2205	1960	4407	3100	914	278	1990	409	344	751	106	34	80	57	58
[28,]	25	5	0	0	0	NA	NA	20	25	NA	0	50	50	100	50	50	40	50	10	10	20	16	2	0	6	44
[29,]	NA	1	25	0	40	6	20	NA	NA	50	0	0	50	0	50	50	0	0	NA	NA	0	NA	0	NA	NA	48
[30,]	100	40	100	300	300	350	250	300	200	300	500	600	350	400	250	200	200	250	20	100	25	0	0	612	208	152
[31,]	150	60	594	350	500	550	250	350	250	700	1200	650	500	1200	500	200	300	800	20	300	50	120	200	365	168	664
[32,]	60	18	40	55	140	30	60	15	10	40	75	80	120	80	40	0	0	0	0	0	NA	NA	NA	9	8	96
[33,]	NA	20	40	50	60	20	20	50	10	25	50	50	NA	75	50	50	25	85	NA	2	25	0	2	9	0	1
[34,]	25	1	12	2	3	0	15	5	0	20	30	30	30	40	35	15	0	15	NA	0	10	0	0	3	NA	2
[35,]	NA	NA	NA	4	45	10	10	35	10	30	20	60	20	50	50	50	10	50	6	NA	0	8	0	6	9	0
[36,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	10	NA	NA	NA	0	0	0	NA	0	NA	NA	NA	NA	2
[37,]	150	20	516	300	400	250	100	225	80	150	250	200	450	250	250	200	75	300	NA	180	50	NA	35	93	156	364
[38,]	NA	NA	440	180	200	325	500	550	1000	NA	NA	NA	800	850	2300	800	50	200	100	0	0	0	0	14	445	10
[39,]	NA	NA	NA	5	NA	NA	NA	NA	40	0	0	0	0	0	0	10	20	120	100	126	40	20	20	88	169	41
[40,]	300	360	420	400	400	60	350	450	250	500	425	100	500	600	175	NA	0	100	100	NA	85	NA	0	180	529	242
[41,]	250	10	510	600	600	300	300	450	350	NA	0	NA	500	250	450	50	20	30	1	0	0	0	0	494	200	250
[42,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	NA	NA	NA	NA	NA	NA	12	0	0	0	10	NA	0	0	0
[43,]	NA	NA	15	15	175	40	30	30	90	NA	0	NA	80	NA	60	20	0	0	NA	0	0	0	0	0	0	0
[44,]	400	500	1500	400	400	100	200	200	NA	NA	NA	50	50	150	120	100	NA	10	NA	NA	NA	10	NA	144	20	NA
[45,]	NA	NA	NA	60	60	10	45	50	100	NA	0	65	200	25	70	100	10	20	4	0	0	0	0	0	0	2
[46,]	NA	NA	NA	NA	NA	NA	NA	NA	40	NA	0	10	5	10	0	NA	NA	NA	NA	NA	13	10	4	0	NA	0
[47,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	12
[48,]	350	460	530	700	400	210	550	500	100	950	700	425	1250	700	375	1400	1574	1990	587	1063	280	150	335	962	668	2514
[49,]	60	30	18	110	120	25	60	75	100	250	140	250	100	225	160	NA	0	0	50	0	50	10	NA	99	4	2
[50,]	90	95	380	300	600	40	100	450	280	700	450	1250	580	800	60	200	NA	NA	NA	50	0	NA	0	2614	1745	833
[51,]	15	8	6	100	15	25	110	10	450	NA	NA	200	200	450	150	20	NA	10	NA	NA	3	NA	0	35	109	5
[52,]	NA	NA	NA	NA	NA	NA	NA	NA	20	NA	0	NA	NA	NA	NA	NA	NA	NA	NA	0	0	0	NA	NA	NA	NA
[53,]	NA	NA	NA	NA	30	5	NA	5	NA	1	0	5	0	NA	4	NA	NA	10	NA	0	0	0	NA	9	0	0
[54,]	400	250	600	600	200	180	550	400	325	1200	500	850	0	950	600	600	1408	667	564	708	300	131	377	605	466	1308
[55,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	7	6
[56,]	600	500	600	2300	250	300	700	1200	1000	1150	1100	1600	500	1500	650	600	230	530	100	150	1000	500	550	2299	2640	662
[57,]	1200	1330	2200	1300	1400	700	950	750	200	700	650	1000	1525	600	500	110	416	364	290	259	352	211	135	195	471	460
[58,]	25	NA	10	12	5	NA	NA	NA	NA	NA	0	5	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	15	0	NA
[59,]	8	26	60	20	NA	20	NA	20	45	110	NA	60	60	60	15	NA	NA	0	NA	1	78	20	65	245	109	0
[60,]	20	10	65	80	40	10	35	35	15	20	66	60	0	20	NA	NA	NA	NA	NA	NA	32	0	NA	NA	14	NA
[61,]	1500	1500	400	300	125	100	300	90	125	700	100	500	500	600	680	774	667	740	350	358	150	92	200	267	479	404
[62,]	500	250	350	250	120	90	110	200	250	960	NA	800	400	650	170	50	200	100	50	301	40	15	NA	603	324	106
[63,]	NA	50	8	20	5	10	15	15	5	25	0	25	0	15	22	200	24	70	1	0	0	20	30	90	205	85
[64,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	15	2	0	4	20	0	7	5	10	0	7	0
[65,]	6	NA	10	10	4	4	NA	2	50	25	25	80	0	0	25	26	0	45	12	2	2	0	40	21	65	0
[66,]	NA	NA	NA	NA	NA	30	30	100	150	NA	25	45	0	60	130	150	0	400	15	0	70	10	20	18	77	2
[67,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	25	0	NA	100	31	10	26	42	NA	NA	6	64	48	62
[68,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1477	1600	2000	2500	3000	1500	3000	4000	3500	NA	NA	800	691	1812	2870
[69,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	1500	550	390	2176	1722	231	273	1561	1225	3438	747	573	353	407	131	1089	773
[70,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	NA	NA	NA	0	0	165	0
[71,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	49	4	4	29	3	0
[72,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	47	208	6
[73,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	300	375	100	527	1000	663	600	860	NA	NA	NA	3000	1271	1032	292
[74,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	6	0
[75,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	25	8
[76,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	600	750	170	550	850	595	400	300	900	900	150	220	125	200	548	298	276
[77,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	344	152
[78,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	584	475	200	NA	NA	NA	150	NA	0	0	NA	NA	NA	NA	NA	70	NA
[79,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	790	212	223
[80,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	52	228	480
[81,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	200	NA	250	1500	1565	600	NA	NA	NA	NA	NA	NA	NA	5460	4096	2719
[82,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	80	90	70	100	175	195	150	100	6	150	NA	NA	NA	150	1055	80	60
[83,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	114	7	NA
[84,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	20	14	7
[85,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	120	180	70	60	70	45	55	25	68	60	50	NA	NA	9	NA	NA	0
[86,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	105	100	NA	NA	NA	NA	NA	NA	NA	NA	0	NA
[87,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0	0	68
[88,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	54	0	92
[89,]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	75	NA	NA	100	100	NA	NA	0	NA

ek no.	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
[1.]	15000	6000	NA	3500	3500	3500	1500	750	1500	7500	3500	750	3500	7500	3500	NA	10000	7500	3500	NA	750	200	25	NA	1000	1000	1000	500	500	1000	1000	1000	1500	500	500	500	1200	1000	800	1500	350	800	1200	1800	650	1000	800
[2.]	3500	5800	NA	3500	3500	1500	1500	3500	1500	3500	750	3500	1500	1500	3500	NA	NA	3500	750	5000	1500	3500	400	5000	3000	3000	1000	1000	1000	2000	1000	1500	1500	1000	1000	2000	2500	1300	300	50	350	NA	500	NA	NA	NA	NA
[3.]	3500	3500	NA	1500	3500	750	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1000	NA	NA	225	NA	400	750	75	NA	75	1000	520	200	400	2516	850	1210	620	1000	5010	2500	5000	6030	8795	5800	4500	2000	1000	1500	30	2000
[4.]	7500	4000	NA	3500	3500	750	1500	3500	200	400	750	400	200	3500	1500	7500	NA	3500	1500	500	750	750	750	4000	800	500	100	NA	500	500	700	600	500	100	500	500	500	800	400	700	100	100	1200	325	500	NA	NA
[5.]	800	810	1124	1750	1500	1500	3500	7500	1500	3500	3500	3500	1500	750	3500	3500	3500	750	1500	750	75	NA	25	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	50	25	50	NA	NA	NA	NA	NA	NA	NA	NA	NA
[6.]	750	700	NA	1500	750	750	750	1500	3500	1500	25	750	3500	3500	3500	1500	700	1500	750	NA	NA	1500	1200	800	1000	700	500	100	300	1500	50	600	400	200	300	400	250	200	200	100	250	NA	NA	300	278	52	130
[7.]	1500	800	NA	200	750	3500	750	1500	1500	3500	750	400	25	75	1500	NA	40	25	200	NA	750	400	300	500	800	2500	1500	4500	700	1500	500	150	600	100	30	350	175	385	200	90	190	110	225	240	250	170	400
[8.]	75	100	NA	200	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	50	3500	1500	NA	NA	NA	NA	NA	NA	300	300	100	300	500	500	75	800	900	1000	900	1800	2400	2900	2000	650	1900	1800	2000	2900	700	2600
[9.]	3500	2500	NA	3500	750	3500	1500	1500	400	400	200	400	400	400	200	NA	NA	NA	NA	NA	NA	NA	NA	25	400	NA	NA	30	150	NA	50	200	500	NA	200	NA	650	700	75	400	NA	NA	NA	300	NA	350	25
[10.]	750	400	NA	750	750	750	400	750	400	200	750	400	400	200	75	NA	NA	NA	NA	300	400	400	600	4000	1200	450	1400	500	900	750	500	500	400	400	700	750	750	500	225	NA	NA	NA	NA	NA	NA	NA	NA
[11.]	400	100	NA	3500	1500	1500	1500	3500	1500	750	1500	750	400	750	1500	750	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	100	NA	75	NA	75	250	NA	NA	250	250	250	NA	NA	NA	NA	NA	NA	NA	NA	NA
[12.]	750	750	1500	3500	3500	1500	1500	3500	750	75	1500	750	750	200	750	NA	200	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[13.]	NA	NA	NA	NA	NA	400	750	1500	400	400	750	1500	1500	400	400	1500	1500	400	400	NA	NA	NA	NA	NA	250	200	500	500	200	300	500	300	400	400	600	1100	500	500	NA	20	NA	800	300	NA	NA	NA	370
[14.]	400	200	3500	3500	1500	1500	1500	1500	1500	400	750	400	200	400	750	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	50	500	200	NA	NA	NA	NA	NA	NA	NA	NA	1
[15.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	750	NA	NA	NA	750	400	NA	NA	1500	400	500	400	1500	400	3000	1800	700	400	450	100	750	400	400	500	500	300	500	500	300	1	NA	NA	NA	NA	NA	NA	NA	NA
[16.]	NA	NA	NA	NA	NA	NA	NA	3500	400	NA	NA	7500	NA	NA	NA	NA	5000	NA	NA	NA	NA	400	75	NA	NA	25	NA	NA	NA	NA	NA	50	100	NA	NA	NA	250	10	10	30	NA	NA	25	25	NA	16	10
[17.]	400	300	1500	1500	3500	1500	1500	1500	750	200	400	750	750	1500	400	400	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[18.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	300	400	200	300	6000	1000	500	200	NA	1500	500	20	400	400	400	NA	250	400	250	NA	NA	NA	20	NA	NA	NA	NA	NA
[19.]	NA	NA	NA	NA	NA	NA	NA	3500	750	400	400	400	750	NA	25	NA	NA	400	NA	NA	200	200	200	NA	250	NA	400	300	200	75	NA	50	200	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[20.]	75	100	NA	400	750	400	750	1500	750	200	750	1500	750	200	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[21.]	75	NA	NA	400	200	200	NA	200	NA	25	NA	NA	NA	25	75	NA	NA	NA	25	75	NA	NA	NA	1000	200	NA	200	NA	500	200	200	800	500	200	400	500	500	100	NA	NA	NA	NA	NA	NA	NA	NA	NA
[22.]	75	50	NA	1500	200	200	1500	1500	750	200	200	200	200	75	200	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	200	NA	NA	NA	NA	25	100	NA	NA	NA	50	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[23.]	200	50	NA	750	400	200	400	400	750	200	750	200	400	200	750	NA	NA	75	NA	NA	NA	NA	NA	NA	300	350	NA	NA	NA	NA	NA	NA	NA	NA	NA	50	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[24.]	750	50	NA	400	750	25	400	400	400	400	NA	NA	75	1500	25	750	NA	NA	75	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[25.]	1500	100	NA	400	400	75	750	400	400	25	750	NA	400	75	75	NA	NA	NA	NA	NA	NA	NA	NA	NA	50	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[26.]	400	500	NA	200	NA	400	400	750	NA	400	400	NA	NA	NA	400	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[27.]	NA	NA	NA	NA	NA	NA	NA	NA	200	200	750	NA	750	200	400	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	150	75	NA	NA	250	250	NA	200	200	250	100	NA	NA	NA	5	NA	NA	NA	NA	NA
[28.]	400	100	NA	750	400	400	200	400	400	200	400	400	200	200	400	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[29.]	NA	NA	NA	NA	NA	NA	NA	1500	750	400	75	NA	750	75	75	400	NA	25	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[30.]	400	50	NA	75	200	25	200	200	25	200	75	75	750	200	25	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	75	50	NA	125	80	50	27	25	25	25	NA	NA	NA
[31.]	400	NA	400	NA	NA	NA	NA	NA	NA	1500	NA	NA	NA	NA	NA	NA	NA	NA	200	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	10	NA	NA	NA	NA	NA	NA	NA	NA
[32.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	25	75	25	NA	25	NA	NA	NA	NA	500	1500	400	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	40	20	NA	NA	NA	2	NA	40	15	10	16	15	7	16	1
[33.]	75	50	NA	75	75	75	400	750	200	25	25	200	750	25	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[34.]	NA	NA	NA	NA	NA	NA	NA	6	3	NA	NA	NA	NA	NA	NA	NA	NA	200	NA	NA	NA	400	25	NA	NA	20	NA	100	150	25	50	70	50	25	50	100	50	100	100	100	NA	NA	100	200	10	40	60
[35.]	NA	NA	NA	NA	NA	NA	NA	NA	75	200	200	NA	400	750	200	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	100	NA	NA	NA	NA	NA	NA	NA	NA	4	NA
[36.]	NA	NA	NA	NA	NA	NA	NA	NA	25	400	NA	25	75	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	400	200	NA	NA	NA	75	NA	NA	NA	NA	NA	NA	10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[37.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	750	200	NA	400	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	20	NA	NA	NA	NA
[38.]	NA	50	NA	NA	NA	NA	NA	NA	NA	NA	200	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	50	NA	NA	NA	NA	NA	NA	NA	NA	NA	7	2
[39.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	25	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	4	NA	NA	3	NA	7	3	NA	3	3	4	8	5	9	1	6	NA
[40.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	25	2	4	4	5	3	55	25	35	20	15	10
[41.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	100	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	1	NA	7	7	4	NA	1	NA
[42.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	6	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[43.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	NA	NA	NA	NA	NA	1	NA	4	NA	2	NA	NA	NA	NA	2	2	10	11	NA
[44.]	25	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	10	1	NA	NA	NA	NA	1	2	6	NA	NA	NA
[45.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	4	NA	1	1	NA	NA	NA	NA	NA	1	2	NA
[46.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	25	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA
[47.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	NA	NA	NA	3	NA	NA	1	5	1	2	NA
[48.]	400	75	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2	NA
[49.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	20	NA	NA	NA	25	NA	NA	NA	NA	NA	NA	25	NA	NA	NA	NA	NA	NA	NA	NA
[50.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	25	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[51.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	1	1	NA	NA	1	NA	NA	NA
[52.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	1	NA	NA	NA	2	NA
[53.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	25	NA	NA	NA	NA	NA	NA	NA	NA	NA	7	NA
[54.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3	2	NA
[55.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[56.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
[57.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	10	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA
[58.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	100	NA	NA
[59.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA
[60.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	9	5	NA
[61.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	100	NA	NA	NA	NA	NA	NA	NA
[62.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2	NA	5	NA	NA	NA
[63.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	NA
[64.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	4	NA	NA
[65.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	4	2	NA
[66.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	13	NA
[67.]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	4	NA
	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2	NA