ABSTRACT

ALLEN, SHANAE D. An Integer Programming Approach to Selecting Individuals for Transfer in Pedigreed Populations. (Under the direction of Dr. Yahya Fathi and Dr. Kevin Gross).

Extinctions of various species are becoming more prevalent and as a result, zoos now play an active role in the conservation of endangered species. It is widely recognized that responsible population management practices are necessary to ensure the long term survival of species residing in zoos. A central concern is to preserve genetic diversity of captive populations in order to avoid detrimental effects on reproductive fitness and to maintain the adaptive potential of the population.

When given the task of selecting individuals to transfer to new or existing populations, zoo managers must take into consideration the genetic effects on all involved populations. This is a challenge because the addition and removal of individuals change the genetic composition of the population. The proposed integer program identifies a group of individuals to transfer that maximizes genetic diversity within all populations. This model is based on a measure of genetic diversity, proportional gene diversity, which is deduced from pedigree analysis.

First, an intuitive, quadratic integer program is presented which considers transferring a given number of individuals from one source population to a transfer site that does not contain existing individuals. This model is reformulated as a linear IP, which in turn lends itself to further simplification. Two extensions of this model allow for (1) the specification of demographic constraints and (2) the consideration of preexisting individuals at the transfer site.

Data are obtained from the California condor studbook and performances of the presented models are compared. Both the linear model and the simplified linear model achieve optimality for a small demand. For larger demand, the latter comes close to optimality in a reasonable amount of time. The performance of the linear IP models is compared to that of an existing program used to balance genetic diversity between two populations, MetaMK. MetaMK employs an iterative, user-based method. A user of this program selects individuals one by one for transfer, as opposed to considering groups of individuals. This method does not guarantee optimality. For the considered demand values, the IP model identifies more genetically diverse groups of individuals than those found by using MetaMK, although one selection procedure in MetaMK comes close to optimality.

An Integer Programming Approach to Selecting Individuals for Transfer in Pedigreed Populations

by
Shanae D. Allen

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Operations Research

Raleigh, North Carolina

2009

APPROVED BY:

_____	_____
Dr. Yahya Fathi	Dr. Kevin Gross
Chair of Advisory Committee	Co-chair of Advisory Committee

_____
Dr. Ted Emigh

BIOGRAPHY

This lucky girl was born in State College, PA where her mother and father, Christina Wyrwas and Jeffrey Allen, were attending Penn State University. Shanae has never been too far removed from school since. She spent the majority of her childhood in Altoona, PA, where she was surrounded by many loving relatives and friends.

Shanae then ventured to the big city of Pittsburgh, PA to attend La Roche College followed by the University of Pittsburgh. During that time, she studied abroad in Rome where she fell in love with Italy and traveling in general. She left Pittsburgh with a Bachelor of Science in mathematics and was admitted to the Master of Operations Research degree program at NCSU. After taking a breadth of courses in biomathematics and OR, she found her niche applying optimization techniques to animal conservation. After graduation, she hopes to apply her knowledge in this context.

ACKNOWLEDGMENTS

I would like to express my thanks to my professors and committee members, especially Dr. Yahya Fathi, for his patience, helpful suggestions, and encouraging words. Throughout this process I have learned the value of balancing passion and reason. Furthermore, I'm grateful to the Department of Operations Research for offering me an assistantship, as this would not have been possible otherwise; also for providing me with a quiet office space to study and for loaning me a department laptop and computer software. Finally, I must thank Ms. Walls, our department secretary, for being on top of deadlines, paperwork, etc.

Colleen Lynch, of the Lincoln Park Zoo, and Michael Mace, of the San Diego Zoo, both contributed to crucial components of my thesis. Michael Mace offered data on the California condor population; while Colleen Lynch answered numerous questions on population management practices, gave valuable insight on the application of my model, and pointed me in the direction of MetaMK.

Most significantly, I'd like to thank my mother whose sometimes daily words of encouragement helped me through and overall perspective gave me calm. It is because of her guidance and support that I have reached this milestone in my life. Also, the love regularly shown by my grandparents and the rest of my family has been invaluable in pursuing my degree.

To my friends in the biomath department: if it wasn't for "biomath date night", rock climbing, salsa classes, camping, etc., I would have graduated semesters ago. In all honesty, you have made my time in grad school much more enjoyable, and I hope we never "break up".

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The combined effect of human population growth, diminished natural habitats, and climatic changes have led to decreases in species diversity and vitality. The International Union for Conservation of Nature (IUCN) predicts that one in four terrestrial vertebrates (mammals, birds, amphibians, and reptiles) is currently threatened with extinction [7]. For most critically endangered species, the only hope of survival lies in the protection of zoos, creating the need for responsible population management strategies and conservation goals.

The preservation of genetic diversity is recognized as one of the top priorities in global conservation by the IUCN [31]. Genetic diversity refers to the present differences in the genomes of all individuals in a given population [15]. The amount of genetic diversity within a closed, captive population is measured relative to the genetic diversity of the wild population from which the captive population originated, hereafter referred to as the base population [22], [29], [12]. Due to the dynamics of small populations, retention of the base population's genetic diversity within the captive population decreases through time (disregarding mutations) [23]. A common measurement of genetic diversity is expected heterozygosity, or the expected proportion of heterozygotes for single loci [15]. This measurement is used herein and the calculation of it will be addressed in the following chapter.

The motivation for thorough genetic management of captive populations is twofold. First, genetic diversity for traits responsive to selection is what enables adaptive, evolutionary change. For many captive species with diminished or nonexistent wild populations, one goal is to once again establish wild, viable, free-ranging populations. This can occur through reintroductions, the creation of new wild populations in part of the species'

former range, or translocations, the addition of individuals to a pre-established population [17]. A genetically diverse population would fare better after release, as the long term survival of any population depends heavily on its ability to adapt to changing environments.

Reintroductions are a means to restore biodiversity and ecological function to altered habitats. This is a fairly recent development in conservation practices: the number of projects involving reintroductions of vertebrates was 124 during 1900-1992 and increased to 424 by 2005 [33]. Some of these include the black-footed ferret (*Mustela nigripes)*, red wolf (*Canis rufus)*, California condor (*Gymnogyps californianus)*, Przewalski's horse (*Equus caballus przewalskii),* and the Arabian oryx (*Oryx leucoryx*) [34], [29].

Another motivation for stringent genetic management is that a lack of genetic diversity for some traits also has a more immediate effect on a population. Reduced genetic diversity can be a result of inbreeding, which is commonly correlated to a reduction in reproductive fitness, or fecundity. A reduction in fecundity within an inbred population is referred to as inbreeding depression. This embodies reproductive rate, survival, mating ability, etc. [15]. Several studies involving both captive and wild populations have shown compelling evidence that inbreeding negatively affects fitness and survivorship. Ralls, Ballou, and Templeton [30] considered 40 captive populations belonging to 38 mammalian species and found that mortality was 33% higher in offspring from full sibling and parent-offspring mating. Crnokrak and Roff [8] performed a meta-analysis on papers containing 157 data sets on measures of fitness traits in wild inbred populations of mammals, birds, snakes, fish, snails, and plants. Significant levels of inbreeding depression were found in 54% of wild, inbred populations.

A practice that effectively slows the rate of decrease of genetic diversity is managing the removal, transfer, and breeding of individuals by consulting a complete pedigree of the population [23]. This level of management is optimal but requires complete control over establishing successful matings and/or the ability to assess parentage of all offspring in the subpopulations of interest. Additionally all births and deaths must be known and recorded.

For the aforementioned reasons, intensive and careful planning is required to ensure that all populations under the control of zoo managers remain as genetically unaltered as possible [15]. This is a challenge because the addition and removal of individuals change the genetic composition of the population [11]. Frequently, when selecting individuals to remove from one population and add to another, the ideal genetic management for the respective

populations is negatively correlated. The focus of this thesis is to manage the transfer of individuals between subpopulations such that genetic trade-offs are balanced.

Other aspects of management include consideration of the demographic needs of subpopulations and the logistic factors of moving individuals among subpopulations; such as, budget, transportation, and space constraints. Therefore, efficient planning is required to ensure that all transfers between subpopulations best benefit the population as a whole while satisfying these constraints.

Herein, an optimization model is developed to select individuals from captive populations to create a new captive or wild population such that genetic diversity is maximized among both subpopulations. This basic model is extended in two ways; (1) chosen individuals must satisfy several demographic constraints, and (2) selected individuals are added to a pre-existing subpopulation. The California condor pedigree is applied as a case study, as this is one population for which the wild component is included in the pedigree. The performance of the model is compared to results obtained from an existing program, MetaMK, which is currently used to balance genetic diversity between populations [1].

The application and motivation of the proposed model changes from program to program, as species' population structures vary and as different goals are assessed for populations in different phases of conservation. The need for such decision support models is evident as zoo managers must transfer individuals between subpopulations for a multitude of reasons – space constraints in zoo facilities, to prevent highly inbred individuals from breeding, to create new captive populations, and to supply wild populations.

In addition to the previously stated reasons for preserving genetic diversity, two current issues in population management are addressed. First, the proposed method for selecting individuals could potentially increase the success rate for reintroduced populations by supplying the wild population with more genetically diverse individuals. The success rate is currently very low (13%), although the definition of success is debatable [14]. Second, even in captive populations that allow for control over mate selection, unplanned births are problematic. Cronin, Mitchell, Lonsdorf, and Thompson [9] evaluated births and the results of transfers in 35 captive species and found that approximately 50% of births across species resulted from non-recommended pairings. In this setting, ensuring maximal genetic diversity in captive subpopulations would make these births less detrimental to the population as a whole. Also, an optimization model is desirable because optimal solutions can be

identified and additional constraints such as maximum allowable transportation costs are easily incorporated.

Chapter 2

Background

2.1 Overview

This section introduces the basic principles behind measuring genetic diversity through pedigree analysis. A major assumption of this method is that the genetic diversity under consideration is for traits unaffected by selection; changes in genetic diversity are random and do not convey an advantage or disadvantage to the species. Accordingly, all genotypes for traits of this type have equal survivorship. As previously mentioned, the genetic diversity mostly of conservation interest is for that of adaptive traits and traits related to fecundity. These traits are heavily influenced by selection and also by many genic and environmental factors. The measurement of these types of traits may vary continuously across the population, and are termed quantitative traits [27].

This is an obvious dichotomy between theory and practice. Kimura [19] introduced the neutral theory, which hypothesizes that most genetic variation present in populations is neutral or nearly neutral; so that, as asserted by Lacy, Ballou, Princée, Starfield, and Thompson [23], "models of change in neutral variation serve as an adequate baseline for overall genetic change in a population" (p. 58). Estimates of neutral variation are also expected to be an indication of quantitative variation in captive populations because effects of selection are minimized in captive populations due to a lack of environmental stressors [23]. Still, the relationship between factors affecting neutral and quantitative variation remains unclear [32]. Montgomery et al. [25] used replicate inbred populations of *Drosophila melanogaster* and provided empirical evidence that using pedigree analysis to manage ge-

netic diversity best preserved microsatellite gene diversities. Another interesting assertion of the neutral theory is that the great majority of evolutionary changes at the molecular level are not caused by selection but by random drift of neutral genes [19]. A discussion of random drift follows in a subsequent section.

The following theory is applied to genetic diversity of one autosomal (not a sex chromosome) locus for which there is no selection, mutation, or linkage occurs among alleles, and alleles follow Mendelian segregation. Because alleles present at different loci are independent, the results for the amount of genetic diversity at one locus can be applied to all other loci that obey these same principles.

2.2 Measures of Genetic Diversity for an Ideal Population

In a diploid species, each individual possesses two alleles per locus on each pair of chromosomes; one allele is inherited from each parent. Therefore an individual is either homozygous at a particular locus, or has two of the same alleles, or is heterozygous, has two different alleles. Genetic diversity of single loci can be measured as observed and expected heterozygosity, and as the number of distinct alleles present [15].

Observed heterozygosity (H_o), or the proportion of observed heterozygotes in a sample, is determined by genotyping randomly sampled individuals at a particular locus. Frequencies for each allele and genotype can then be calculated. For some populations, the Hardy-Weinberg Principle provides a method for extrapolating these results to predict the heterozygosity of the entire population. This law is a mathematical model of genotype frequencies in an ideal population for a locus that does not undergo genetic drift, mutation, selection, or gene migration. An ideal population in this case is a closed population of constant, large size with random mating occurring among individuals. In such a population, allelic frequencies will remain the same from generation to generation; therefore predicted or "expected" and observed heterozygosities are equivalent [15].

Expected Heterozygosity (H_e) is the probability that two genes at a given locus, selected at random from the relevant population(s), will be different [22], [26]. If the locus is in Hardy Weinberg equilibrium the distribution of frequencies for all genotypes can be mathematically expressed as a binomial (for two alleles) or multinomial (for multiple alleles) distribution. In this context, there are two independent trials and random variables are

defined as X_l = the number of times allele A_l is observed. Accordingly, if there exists alleles $A_1, ..., A_L$,with frequencies $q_1, ..., q_L$ respectively, the frequency of A_lA_l genotypes is q_l^2 while the frequency of A_lA_m genotypes is $2q_lq_m$ $(l, m = 1, ..., L$ and $l \neq m)$. Therefore the total expected frequency of heterozygotes is $\sum_{l \neq m} q_l q_m$ or equivalently, $1 - \sum_{l=1}^{L} q_l^2$ [26].

As previously stated Hardy Weinberg equilibrium can only be applied to large, randomly mating populations when there are no perturbing forces on allele and genotype frequencies; although deviations from Hardy Weinberg equilibrium due to one or several of these factors can be mathematically quantified. A locus in a population that is not considered ideal may still be in Hardy-Weinberg equilibrium if the population is relatively large and mating is near random. To test if a locus in such a population is in Hardy Weinberg equilibrium, a chi-squared test is performed on the observed genotype frequencies in the sample. If differences are not significant, then the expected heterozygosity of the population is $1 - \sum_{l=1}^{L} \widehat{q}_l^2$, where allele frequencies, $\widehat{q}_l$, are gathered from the sample [15].

Hardy Weinberg equilibrium is not directly applicable to many captive populations due to small population sizes and the violation of random mating. Genetic drift and inbreeding are two catalysts of genetic diversity loss; both have a minimal effect on the loss of genetic diversity in large, randomly mating populations but are the most significant causes of genetic diversity loss in small or fragmented randomly mating populations. In this respect, the expected heterozygosity of a small, nonrandom mating population will be significantly less than that predicted by the Hardy Weinberg law.

Genetic drift is the process of random change in allele frequencies. Each offspring receives two alleles, one allele randomly sampled from each parent. As a result, allele frequencies vary throughout generations. In randomly mating populations, the expected frequency of each allele averaged over many populations stays constant through time, but the variance of allele frequencies increases. This is due to the mechanisms of sampling.

The population structure of a typical captive population is such that the overall population is divided into several to hundreds of subpopulations; where each subpopulation exists in a zoo or the wild. This population structure leads to the violation of random breeding as the gene flow between zoo facilities and supplied to wild populations is left up to managers. This is rarely done randomly.

If the captive population is wisely managed, individuals are transferred and bred to produce an avoidance of close inbreeding; while mismanagement or a lack thereof, leads to

increased inbreeding. Mismanagement occurs when managers unknowingly transfer related individuals among facilities. A lack of management is typified by groups of individuals kept in isolation among zoos and left to breed freely causing inbreeding to occur at an increased rate. Only recently has attention been given to preserving genetic diversity in captive populations, as many were not managed until the 1980's [15]. The management strategy suggested in this thesis and by others ([3], [22]), in addition to maximizing genetic diversity in future generations, also attempts to rectify past losses in heterozygosity.

Another factor that exacerbates losses in genetic diversity in captive populations of endangered species is that most are founded by a small number of individuals, called founders. According to Frankham, Ballou, and Briscoe [15], a minimum of 20-30 founders are needed to capture the genetic diversity in the base population. Currently there are many species that stem from a far less number of founders than this minimum; such as the Arabian Oryx, black-footed ferret, European bison, California condor, Guam rail, and the Przewalski's horse. The California condor population will be further considered.

2.3 Measure of Genetic Diversity for a Finite, Nonrandom Mating Population

To review, a measure of genetic diversity for a neutral locus is expected heterozygosity. For an ideal population, this is $H_e = 1 - \sum_{l=1}^{L} q_l^2$. For a finite, nonrandom mating population (therefore one in which genotypes do not follow Hardy Weinberg equilibrium frequencies) expected heterozygosity is termed *gene diversity (GD)* by Nei [26]. All other assumptions are assumed to be true. A commonly used maximum likelihood estimator of gene diversity is $\widehat{d}_t = 1 - \sum_{l=1}^{L} \widehat{q}_{l,t}^2$ [18]. The interest herein is to quantify gene diversity for N considered pedigreed individuals alive at time t. The expectation of the estimator $\widehat{d}_t$, sampling two alleles out of the N individuals alive at time t is

$$GD_t = E(\widehat{d}_t) = 1 - \sum_{l=1}^{L} E(\widehat{q}_{l,t}^2). \tag{2.1}$$

The notation used by Lacy (1995), GD_t, is used instead of $E(\widehat{d}_t)$. To calculate $\sum_l E(\widehat{q}_{l,t}^2)$, the work of Cockerham [6] will be closely followed.

First, random variables, y_{iln}, are introduced. Let $y_{iln} = 1$ if individual i possesses allele A_l in the n^{th} position and $y_{iln} = 0$ otherwise. For diploid individuals, $n = 1, 2$. In

this way, the order of alleles, for a particular locus, is differentiated within individuals. The population under consideration consists of the N living individuals recorded in the pedigree so that $i = 1, ..., N$. Due to the assumptions and theory previously stated, there are no perturbing forces acting on expected allele frequencies as averaged over many groups of individuals. Therefore, the probability that individual i inherits allele A_l in position n remains constant through generations. Then, $E(y_{iln}) = q_{l,0}$, where $q_{l,0}$ is the frequency of allele A_l in the base generation ($t = 0$). Also,$E(y_{iln}^2) = q_{l,0}$ because $y_{iln}^2 = y_{iln}$. Then from the definition of variance, $Var(y_{iln}) = q_{l,0}(1 - q_{l,0})$.

From this basis, the following covariances will be considered; the covariance between frequencies of allele A_l in positions $n = 1$ and $n = 2$ within a particular individual and the covariance between frequencies of allele A_l from a particular pair of individuals. From these developments, the expected proportion of homozygotes for allele A_l will be calculated for the group of N individuals under consideration.

At the individual level, $y_{il1} * y_{il2} = 1$, only if $y_{il1}$and y_{il2} are both 1; therefore $E(y_{il1}y_{il2}) = prob$(individual i is homozygous for allele $A_l) * 1$. The probability is calculated by considering the two possibilities an individual is homozygous; either the individual inherited the same allele from two different ancestors, the two alleles are *identical by state* (IBS), or the individual received two of the same alleles from the same ancestor by way of inbreeding, the alleles are *identical by descent* (IBD) The probability individual i contains alleles *identical by descent* is denoted f_i, and is called the *inbreeding coefficient* [12]. f_i is calculated from a pedigree and this will be illustrated in a later section. So that,

$$\begin{aligned} prob(y_{il1} = 1, y_{il2} = 1) &= prob(individual\ i\ is\ IBS\ for\ allele\ A_l) \\ &\quad + prob(individual\ i\ is\ IBD\ for\ allele\ A_l) \\ &= q_{l,0} * q_{l,0}(1 - f_i) + q_{l,0} * f_i. \end{aligned}$$

It then follows,

$$Cov(y_{il1}y_{il2}) = E(y_{il1}y_{il2}) - Ey_{il1}Ey_{il2} = f_i * q_{l,0}(1 - q_{l,0}).$$

This relationship demonstrates the correlation between the frequency of allele A_l at position 1 and frequency of allele A_l at position 2 is f_i. Next, consider the relationship between frequencies of allele A_l, in the first allele position in two individuals, i and j. The

quantity, $y_{il1} * y_{jl1} = 1$, only if both individuals contain allele A_l at position 1. Similar to the previous argument, this can happen if the allele A_l shared by the two individuals is IBD or IBS. Just as f_i is the probability an individual is IBD for a random allele, k_{ij} is the probability that two randomly chosen alleles, one from individual i and one from individual j, are IBD. Thus, $k_{ij} = k_{ji}$. This probability, called the *kinship coefficient* between individuals i and j, comes from pedigree analysis and will be calculated for all pairs of individuals [12]. The kinship coefficient is the inbreeding coefficient of the pair's offspring. Therefore,

$$\begin{aligned} prob(y_{il1} = 1, y_{jl1} = 1) &= prob(individuals\ i\ \&\ j\ share\ allele\ A_l\ IBS\ at\ position\ 1) \\ &\quad + prob(individuals\ i\ \&\ j\ share\ allele\ A_l\ IBD\ at\ position\ 1) \\ &= q_{l,0} * q_{l,0}(1 - k_{ij}) + q_{l,0} * k_{ij}. \end{aligned}$$

Due to the independent inheritance of allele A_l for each position within an individual and the definition of kinship coefficient, this probability is the same for any two chosen positions so that,

$$\begin{aligned} prob(y_{il1} = 1, y_{jl1} = 1) &= prob(y_{il1} = 1, y_{jl2} = 1) = prob(y_{il2} = 1, y_{jl1} = 1) \\ &= prob(y_{il2} = 1, y_{jl2} = 1) \end{aligned} \tag{2.2}$$

Consequently,

$$\begin{aligned} Cov(y_{il1} y_{jl1}) &= Cov(y_{il1} y_{jl2}) = Cov(y_{il2} y_{jl1}) = Cov(y_{il2} y_{jl2}) \\ &= E(y_{il1} y_{jl1}) - E y_{il1} E y_{jl1} = k_{ij} * q_{l,0}(1 - q_{l,0}). \end{aligned}$$

As above, the kinship between two individuals, k_{ij}, is the correlation between the frequency of a random allele chosen from one individual with the frequency of a randomly selected allele from the other individual [6].

To find the expected proportion of homozygotes for allele A_l in the group of N living individuals at time t, consider first,

$$\widehat{q}_{l,t} = \frac{\sum_{i=1}^{N} \sum_{n=1}^{2} y_{iln}}{2N} \tag{2.3}$$

Then,

$$E(\widehat{q}_{l,t}^2) = E\left(\left(\frac{\sum_{i=1}^{N}\sum_{n=1}^{2} y_{iln}}{2N}\right)^2\right) \tag{2.4}$$

Written equivalently as,

$$\frac{1}{4N^2}E\left(\sum_{i=1}^{N}\sum_{n=1}^{2} y_{iln} * \sum_{i=1}^{N}\sum_{n=1}^{2} y_{iln}\right)$$

$$\begin{aligned} &= \frac{1}{4N^2}\left(\sum_{i=1}^{N}\sum_{n=1}^{2} E(y_{iln}y_{iln}) + 2\sum_{i=1}^{N} E(y_{il1}y_{il2})\right) \\ &\quad + \frac{1}{2N^2}\left(\sum_{i=j+1}^{N}\sum_{j=1}^{N-1}\sum_{m=1}^{2}\sum_{n=1}^{2} E(y_{iln}y_{jlm})\right) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{4N^2}\left(2Nq_{l,0} + 2\left(\sum_{i=1}^{N} Cov(y_{il1}y_{il2}) + \sum_{i=1}^{N} E(y_{il1})E(y_{il2})\right)\right) \\ &\quad + \frac{1}{2N^2}\sum_{i=j+1}^{N}\sum_{j=1}^{N-1}\sum_{m=1}^{2}\sum_{n=1}^{2} Cov(y_{iln}y_{jlm}) \\ &\quad + \frac{1}{2N^2}\sum_{i=j+1}^{N}\sum_{j=1}^{N-1}\sum_{m=1}^{2}\sum_{n=1}^{2} E(y_{iln})E(y_{jlm}) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{2N^2}\left(Nq_{l,0} + (q_{l,0} - q_{l,0}^2) * \sum_{i=1}^{N} f_i + Nq_{l,0}^2\right) \\ &\quad + \frac{1}{N^2}\left((q_{l,0} - q_{l,0}^2) * 2\sum_{i=j+1}^{N}\sum_{j=1}^{N-1} k_{ij} + N(N-1)q_{l,0}^2\right) \end{aligned}$$

$$= \frac{(q_{l,0} - q_{l,0}^2)\left(\frac{1}{2}\sum_{i=1}^{N}(1+f_i) + 2\sum_{i=j+1}^{N}\sum_{j=1}^{N-1} k_{ij}\right)}{N^2} + q_{l,0}^2$$

The quantity below, $\overline{MK}_t$, is termed the *average mean kinship* of the group of N individuals alive at time t (this quantity is designated as the group coancestry coefficient, Θ_l, by Cockerham [6]) and will be further considered in this analysis.

$$\overline{MK}_t = \frac{\frac{1}{2}\sum_{i=1}^{N}(1+f_i) + 2\sum_{i=j+1}^{N}\sum_{j=1}^{N-1} k_{ij}}{N^2} \tag{2.5}$$

Turning the attention back to finding GD_t, let L be the total number of alleles present in the base population. Then by (2.1),

$$GD_t = 1 - \sum_{l=1}^{L} E(\widehat{q}_{l,t}^2) \tag{2.6}$$

$$GD_t = \left(1 - \sum_{l=1}^{L} q_{l,0}^2\right) * (1 - \overline{MK}_t) \tag{2.7}$$

Further simplification is achieved by substituting $GD_0 = 1 - \sum_{l=1}^{L} q_{l,0}^2$. Giving the following,

$$\frac{GD_t}{GD_0} = 1 - \overline{MK}_t \tag{2.8}$$

GD_t/GD_0 is referred to as *proportional gene diversity* and represents the measure of genetic variation within the population relative to the base population [22].

Another quantity worth noting is the variance of the frequency of some allele A_l at time t.

$$var(\widehat{q}_{l,t}) = \overline{MK}_t * q_{l,0}(1 - q_{l,0}), \tag{2.9}$$

Equivalently stated as

$$var(\widehat{q}_{l,t}) = \left(\frac{1 + \overline{f}}{2N} + \frac{N-1}{N}\overline{k}\right) q_{l,0}(1 - q_{l,0}), \tag{2.10}$$

where $\overline{f}$ is the average inbreeding coefficient among individuals and $\overline{k}$ is the average kinship coefficient among all pairs. As Cockerham [6] so succinctly states "the variance among gene frequencies for neutral genes of groups of individuals depends on the inbreeding and relatedness of individuals in the group as well as the number of individuals" (p. 73). So that average mean kinship addresses cumulative loss of genetic diversity throughout generations lost as a result of inbreeding, as well as genetic drift.

Cockerham [6] asserts that the expected frequency of heterozygotes should reflect the variation of genes within individuals, whereas proportional gene diversity reflects variation within and between individuals. Accordingly, Cockerham [6] defines expected frequency of heterozygotes to be

$$2\sigma_w^2 = 2q_{l,0}(1 - q_{l,0}) * (1 - \overline{f}), \tag{2.11}$$

where σ_w^2 is the variance between genes within individuals. An important distinction is that this definition pertains to a population containing only two alleles.

Proportional gene diversity is a more appropriate metric to use in population management when considering the following. Average mean kinship is the probability two randomly sampled alleles from the population are identical by descent [22]. In a similar sense, as Cockerham [6] and Ballou and Foose [2] allude to, average mean kinship is the expected mean inbreeding coefficient of N randomly sampled offspring generated from the random mating of the N considered individuals. In this fashion, selecting individuals for breeding based on this metric allows one to control expected heterozygosity in future generations. Therefore, minimizing average mean kinship among breeders maximizes gene diversity in the next generation.

Mathematically, the relationship between average mean kinship of individuals at time t and the expected mean inbreeding coefficient of N offspring considered at time $t+1$ is,

$$\overline{MK}_t = E(\overline{f}_{t+1}). \tag{2.12}$$

The above definition asserts that the population at time t is randomly mating. Thus, the application of this definition is limited to populations that are hermaphroditic, meaning individuals can self-fertilize and there is no separation of sexes. Cockerham [6] expounded on this definition to account for non-hermaphroditic populations with equal sex ratios. In this case, $\overline{MK}_t = E(\overline{f}_{t+2})$. Also, Cockerham [6] presented a formulation to find $var(\widehat{q}_{l,t})$ in non-hermaphroditic populations with unequal sex ratio. Herein, it is assumed the population of interest has equal sex ratio.

Another benefit to using this metric as opposed to one such as the calculation of gene diversity due solely to genetic drift ($\frac{GD_t}{GD_0} = (1 - \frac{1}{2N})^t$) is that the calculation of average mean kinship does not require non-overlapping generations (as is typical in many animal populations) [22]. Although, the translation of average mean kinship to be the expected mean inbreeding coefficient of offspring only holds if all individuals have equal probability of mating and producing surviving offspring. Ballou and Lacy [3] provide an individual metric similar to mean kinship, called kinship value (defined in Literature Review), that

accounts for varying reproduction probabilities.

2.4 Pedigree Analysis

The parameters, inbreeding and kinship coefficients, needed to measure proportional gene diversity are deduced from pedigree analysis in the following manner. A recursive method is used to construct the symmetrical kinship matrix which includes both pairwise and self kinships. This method accounts for overlapping generations so that the group of N individuals alive during time t may contain individuals from multiple generations. The procedure and principle behind this method is presented in a similar manner to that of Chang et al. [5].

Consider individuals X and Y with alleles x_m, x_p and y_m, y_p, denoting the maternally and paternally inherited allele in each individual. Additionally, let x and y be random alleles drawn from X and Y respectively (i.e. $x = x_m$ or $x = x_p$). Furthermore, let individuals M_x and P_x be the parents of X, and M_y and P_y be the parents of Y. Then,

$$\begin{aligned} K_{XY} &= prob(\text{a random allele drawn from } X \text{ is IBD to a random allele drawn from } Y) \\ &= prob(x \text{ is IBD to } y_m | y = y_m) * prob(y = y_m) \\ &\quad + prob(x \text{ is IBD to } y_p | y = y_p) * prob(y = y_p). \end{aligned}$$

Note: K_{XY} is the kinship coefficient between x and y as k_{ij} denoted the kinship coefficient between i and j in the previous section.

If X is not a direct descendent of Y, the previous statement can be equivalently expressed as,

$$\begin{aligned} K_{XY} &= \frac{1}{2} prob(x \text{ is IBD to a random allele drawn from } M_y) \\ &\quad + \frac{1}{2} prob(x \text{ is IBD to a random allele drawn from } P_y). \end{aligned}$$

Thus,

$$K_{XY} = K_{YX} = \frac{1}{2}(K_{XM_y} + K_{XP_y}). \tag{2.13}$$

To calculate self kinships, recall the inbreeding coefficient, f_X , of individual X is the kinship coefficient of the individuals parents, $K_{M_xP_x}$. Self kinship, K_{XX}, is the

probability two alleles taken at random from the same individual, with replacement, are IBD. So that,

$$\begin{aligned} K_{XX} &= prob(x \text{ is IBD to } x_m | x = x_m) * prob(x = x_m) \\ &\quad + prob(x \text{ is IBD to } x_p | x = x_p) * prob(x = x_p). \end{aligned}$$

Equivalently,

$$K_{XX} = \frac{1}{2}(1 + f_X) = \frac{1}{2}(1 + K_{M_x P_x}). \tag{2.14}$$

In this fashion,

$$\overline{MK} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} K_{ij}}{N^2}. \tag{2.15}$$

It is important to clearly define the base population as the inbreeding coefficients, f_i, and kinship coefficients, k_{ij}, are measured relative to this population. Setting the base population as the point of reference means that pedigreed relationships between individuals in the base population and to individuals in past generations are ignored. Therefore, the individuals in the base population are assumed to be unrelated and noninbred, or in other words, have inbreeding and kinship coefficients of zero [12].

If the founders are thought to be related (i.e. all came from the same clutch or confirmation by molecular analysis), individuals called "analytical founders" may be added to the pedigree [29]. The addition of these individuals and their hypothetical relationships to the real founders create a deeper pedigree that accurately reflects the presumed kinships and self-kinships of the founders. In this scenario, the base population would be defined as the population from which the parents of the founders originated. This is demonstrated by Ralls and Ballou [29].

The recursive algorithm in (2.13) can be easily used to construct the kinship matrix if all offspring follow parents. The base population should be listed first in the kinship matrix, so that the kinship matrix contains an identity submatrix between these individuals. But, the value of average mean kinship does not change regardless of whether "analytical founders" precede the true founders in the kinship matrix or if only true founders are considered in the kinship matrix, as long as the kinships between the true founders remain the same. It is beneficial though for other types of genetic analysis to include unrelated individuals from the base population in the pedigree. For one, the pedigree then lends itself

for use in gene drop analysis [22]. This type of analysis would be used to estimate the allelic diversity in the present population.

The inclusion of relationships between founders will increase kinships in following generations as opposed to using unrelated founders. It is usually assumed, due to a lack of data, that founders are unrelated and noninbred. Consequently, proportional gene diversity may be an overestimate.

Another detail worth noting is that it is standard in many population management programs (GENES, PM2000) for founders to be excluded in the calculation of $\overline{MK}$ [22]. This allows the metric, proportional gene diversity, to be a reflection of gene diversity propagated in captivity versus in the wild. The inclusion of wild individuals would lead to a skewed value as the calculation of the metric is sensitive to outliers. This is ignored in the presented problem instance as there were only a small number of wild individuals in the pedigree. Also the goal of the example is to demonstrate the selection of individuals rather than produce an accurate assessment of gene diversity within the population.

Chapter 3

Literature Review

3.1 Mean Kinship in Population Management

Ballou and Lacy [3] were the first to use the concept of mean kinship and average mean kinship to identify genetically valuable individuals. Therein, mean kinship for individual i, denoted mk_i, is defined as

$$mk_i = \frac{\sum_{j=1}^{N} k_{ij}}{N}, \tag{3.1}$$

where N is the total number of living individuals and k_{ij} is the kinship between individuals i and j as previously described.

This metric provides a way to rank an individual based on its relatedness to all living individuals and its inbreeding coefficient. An individual's mean kinship illustrates its genetic importance among other individuals, whereas average mean kinship describes the genetic diversity of an entire group.

The authors also suggested an iterative approach, using average mean kinship, to manage breeding in five simulated pedigrees. They found this method to outperform several other breeding strategies in both retention of heterozygosity and allelic diversity. This approach involved first ranking individuals by mean kinship and selecting the male and female with the lowest such value to produce one offspring. Then mean kinship was recalculated for each individual with respect to the added individual. This procedure is continued until the number of desired offspring is reached. This assumes complete control over breeding and also the survival of all offspring.

Haig, Ballou, and Derrickson [16] compared several selection schemes – one closely resembled minimizing average mean kinship – for the purpose of creating a genetically diverse group of individuals. Some selection criteria were based on reproductive fitness, maximizing an empirical measure of variation using molecular data, equalizing founder contribution, maximizing allelic diversity, and maximizing founder genome equivalents (FGE). The number of founder genome equivalents, was originally defined by Lacy [21], and as it is used by Haig et al. [16] is related to average mean kinship by $FGE \approx \frac{1}{2\overline{MK}}$ [22]. Under the genetic criteria, individuals were chosen based on simulated genotypes. To evaluate the success of each selection option, genotypes of hypothetical offspring of the selected individuals were simulated and compared. The heterozygosity and allelic diversity of simulated offspring of the individuals selected by maximizing FGE compared favorably against other selection criteria. This study, however, does not consider the effects of the removal of individuals on the source population.

3.2 The Consideration of Genetic Trade-offs

Few publications consider the effects of transferring individuals on both the source and destination populations. Two such studies by Earnhardt [11] and Ralls and Ballou [29] selected individuals based on mean kinship. Earnhardt [11] compared several strategies for selecting individuals for reintroduction using a metric directly related to average mean kinship to measure genetic trade-offs between captive and wild populations of several species. The genetic-based selection schemes involved ranking individuals according to mean kinship and selecting those for reintroduction that were either genetically detrimental or genetically valuable to the source population. For all species, neither method simultaneously benefited both populations. In other words, the average mean kinship of one population was less than that of the original population before the removal of individuals. Also, both methods resulted in very low genetic diversity within one of the created populations. This is expected, as an improved selection strategy would simultaneously consider kinships between all selected individuals and the effect of removing those individuals from the source population.

Ralls and Ballou [29] illustrated a method that allows for the quantification of genetic importance of each individual within both the source and destination populations

and shows how this amount is dependent upon previously chosen individuals. This is accomplished by using MetaMK. This program measures the effect of transferring a specific individual on the proportional gene diversity in both the source and destination populations [1]. Using this program, individuals are chosen iteratively based on given criteria (i.e. minimize genetic loss within the source population; maximize genetic gain in the destination or source population; and a combination thereof, etc.). The downside of using such an iterative method is that, even by sorting the individuals according to certain criteria, an optimal solution is not apparent. There may be many individuals that satisfy given criteria. Therefore, trial and error must be used to distinguish between choices but it still does not guarantee finding an optimal solution. MetaMK is explained in further detail and the program's performance is compared to that of the proposed model in the Results and Computation section.

Ralls and Ballou [29] utilized MetaMK to manage transfers between all California condor subpopulations, including several preexisting wild populations. Different selection strategies were employed to best accomplish various goals. Several of these selection schemes are further discussed in the Results and Computation section of this thesis. The result of the placement of all individuals on the overall genetic diversity among all populations was not quantified but Ralls and Ballou [29] state that the total genetic diversity within all populations was increased.

3.3 The Use of Optimization Techniques in Population Management

When optimization techniques are employed, the focus changes from considering genetic importance of all individuals to considering the genetic importance of all sets of individuals. By using optimization models, the genetic composition of all possible groups of d individuals are evaluated, where d is the given number of individuals to be transferred, and an optimal solution, if one exists, can be identified.

Kostreva, Ogryczak, and Tonkyn [20], were among the first to apply an optimization model to identifying individuals for transfer. The goal of this model was to adjust founder contribution within the living individuals, as opposed to using mean kinship. The contribution of a particular founder to the current generation is the total of all kinships be-

tween the founder and each living individual. Founder contribution varies among founders as some lines are bred more than others leading to the over- and under- representation of founder lines.

The model proposed by Kostreva, Ogryczak, and Tonkyn [20] is an integer program that selects a group of individuals for reintroduction that more evenly represents all founders. Specifically, the model identifies the group of individuals that contains the greatest representation of the founder with the lowest founder contribution. This method increases the representation of underrepresented founders but does not adjust for over-represented founders. An attempt to equalize founder contribution, however, is no longer recommended in population management. From the concept of "target founder contribution" presented by Ballou and Foose [2], managing according to minimizing average mean kinship best adjusts founder representation. Average mean kinship is minimized when each founder's contribution is proportional to the number of their unique alleles surviving in the population [2]. This assures the greatest heterozygosity.

Minimizing weighted average mean kinship by way of an integer program model is employed by Fernández, Toro, and Caballero [13] for the purpose of finding the optimal number of progeny individuals should contribute. The concept of their model is based on the iterative approach presented by Ballou and Lacy [3]. This is done by minimizing the sum of all weighted kinships such that kinship between two individuals is multiplied by the product of the number of gametes to be generated by each individual. The conception of one individual requires two gametes, or reproductive cells (i.e. sperm, egg). The decision variables are v_j, the number of gametes to be generated by individual j. The objective function first considered by Fernández, Toro, and Caballero [13] is

$$\min \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} v_i v_{j.} \tag{3.2}$$

Due to the mathematical complexity of this model, a heuristic (simulated annealing) is used to find a close approximation to the global minimum.

Chapter 4

Model Formulations

When faced with the task of creating new wild or captive populations, zoo managers recognize the necessity of managing genetic diversity within populations, but the sizable number of choices creates the need for a decision making tool. Without such a tool there is a total of $\binom{N}{d}$ number of selection schemes, when selecting d individuals out of N individuals. This number quickly increases with population size and demand.

Herein, genetic diversity of a group of individuals is measured by that group's average mean kinship, due to the relationships previously stated. The objective of the following formulations is to select individuals for transfer such that average mean kinship within the source and destination populations is minimized. First, an intuitive formulation is presented for the simplest case where a group of individuals are transferred to a site in which no other individuals exist. This model is mathematically complex in that it involves a nonlinear objective function as well as a nonlinear constraint. This is especially problematic given a large number of integer variables; which is typical for the presented problem. An alternative but equivalent model follows; this model is linear thereby reducing computational complexity.

In all formulations, the solution must satisfy two core constraints; the number of selected individuals must satisfy a given demand and the resulting average mean kinship of the source population must be less than or equal to that of the transfer site. The latter constraint was chosen on both a biological and a computational basis. As the demand increases, it is possible that several different groups of transferred individuals result in the same total average mean kinship. By adding this constraint, the source population's gene

diversity is less affected by the transfer. Computationally, this constraint decreases run time by decreasing the number of equivalent solutions.

Two extensions of the model follow. The first is formulated to allow for a specified demographic composition of the transferred individuals; while the other includes an altered objective function and constraints to account for the existence of other individuals in the transfer site.

4.1 Assumptions

The assumptions for the proposed model can be summarized as follows:

- The parentage for all individuals is known and individuals originating from wild populations are unrelated and noninbred.
- All individuals in the source population can be transferred.
- The kinships between and within all individuals contained in the kinship matrix are considered in the calculation of average mean kinship of relevant populations.
- The total number of individuals in the source population that are to be selected for transfer is fixed.

4.2 Parameters

Most of the data needed to solve the stated objective for all proposed models are based on a complete pedigree comprised of N living individuals. This includes C individuals currently residing in the source site and T individuals present in the transfer site. Each individual is assigned a unique studbook number at birth; S shall denote the set of studbook numbers of all N living individuals, where $N = C + T$ and $S = [s_1, s_2, ..., s_N]$. An element of S refers to a particular individual, s_i. The order of the elements of S will change depending upon the model formulation.

The kinship matrix, K, (also termed *additive genetic variance-covariance matrix*) is an $N \times N$ symmetric matrix that contains kinship values for each living pair of individuals, including self kinships. The rows and columns of K are ordered according to the set S. For instance for S=[1, 103, 11, ..., 99], k_{12} is the kinship coefficient of individuals #1 and #103.

All of the following formulations require that a fixed number of individuals, d, be transferred from the source population to the transfer site.

4.3 Notation for decision variables

The decision as to which individuals shall be selected for transfer can be represented as a 1-by-C solution vector, $\mathbf{x}$, containing C binary variables in the form x_i. An element, x_i, equals one if the individual, s_i, is chosen for transfer from the source site to the transfer site; x_i is equal to zero otherwise.

In this fashion, the kinship between two individuals should be counted toward average mean kinship of the source site, $\overline{MK}_{SOURCE}$, only if both are not chosen for transfer; and the kinship between these two individuals should be counted toward average mean kinship of the transfer site, $\overline{MK}_{TRANSFER}$, if both are selected for transfer. The kinship between two individuals will not be counted if one individual is selected for transfer and one is not. In this situation, it is assumed the two individuals reside in different subpopulations and can not breed; therefore the pair will not contribute to the genetic diversity of future generations.

4.4 Nonlinear Model

This model accurately reflects the goal of minimizing average mean kinship within the source and transfer sites when no other individuals exist in the destination site prior to transfer; therefore $T = 0$ and equivalently $N = C$.

$$\min_x f(x) \tag{4.1}$$

where

$$f(x) = \overline{MK}_{SOURCE} + \overline{MK}_{TRANSFER} \tag{4.2}$$

According to definition (2.15),

$$f(x) = \frac{1}{(N-d)^2} \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij}(1-x_i)(1-x_j) + \frac{1}{d^2} \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} x_i x_j \tag{4.3}$$

subject to

$$\frac{1}{(N-d)^2}\sum_{i=1}^{N}\sum_{j=1}^{N}k_{ij}(1-x_i)(1-x_j) \leq \frac{1}{d^2}\sum_{i=1}^{N}\sum_{j=1}^{N}k_{ij}x_ix_j$$

$$\sum_{i=1}^{N}x_i = d$$

and

$$x_i \in \{0,1\} \qquad \text{for } i = 1,...,N. \tag{4.4}$$

4.5 Linear Model

Expansion of the objective function, f, of the original model leads to

$$\begin{aligned} f(x) \quad = \quad & \frac{[(N-d)^2+d^2]\sum_{i=1}^{N}\sum_{j=1}^{N}k_{ij}x_ix_j}{d^2\times(N-d)^2} - \\ & \frac{\left(\sum_{i=1}^{N}\sum_{j=1}^{N}k_{ij}x_i + \sum_{i=1}^{N}\sum_{j=1}^{N}k_{ij}x_j\right) - \sum_{i=1}^{N}\sum_{j=1}^{N}k_{ij}}{(N-d)^2} \end{aligned} \tag{4.5}$$

The terms $\sum_{i=1}^{N}\sum_{j=1}^{N}k_{ij}x_i$ and $\sum_{i=1}^{N}\sum_{j=1}^{N}k_{ij}x_j$ are equivalent due to the symmetry of the K matrix. Also a new variable z_{ij} is introduced such that,

$$z_{ij} = x_ix_j \tag{4.6}$$

Therefore, $z_{ij} \in \{0,1\}$ such that $z_{ij} = 1$ only when $x_i = 1$ and $x_j = 1$, otherwise $z_{ij} = 0$. To avoid introducing a nonlinear constraint, the following set of $3 \times N^2$ linear constraints may be substituted,

$$z_{ij} \leq x_i \qquad \forall \;\; i = 1,...N,\;\; j = 1,...,N$$

$$z_{ij} \leq x_j \qquad \forall \;\; i = 1,...N,\;\; j = 1,...,N$$

$$z_{ij} \geq x_i \; + x_j - 1 \qquad \forall \;\; i = 1,...N,\;\; j = 1,...,N.$$

Additionally, the substitution results in $N(N-1)$ new variables. This leads to the formulation of the linear model,

$$\min_x f'(x) \tag{4.7}$$

where

$$f^{'}(x) = [(N-d)^2 + d^2] \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} z_{ij} - 2d^2 \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} x_i \tag{4.8}$$

subject to

$$\frac{N^2 - 2Nd}{d^2} \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} z_{ij} \geq \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} - 2 \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} x_i \tag{4.9}$$

$$\sum_{i=1}^{N} x_i = d \tag{4.10}$$

$$z_{ij} \leq x_i \qquad \forall \;\; i = 1, ...N, \;\; j = 1, ..., N \tag{4.11}$$

$$z_{ij} \leq x_j \qquad \forall \;\; i = 1, ...N, \;\; j = 1, ..., N \tag{4.12}$$

$$z_{ij} \geq x_i \;\; + x_j - 1 \quad \forall \; i = 1, ...N, \;\; j = 1, ..., N \tag{4.13}$$

and

$$x_i \in \{0, 1\}, \qquad \text{for } i = 1, ..., N. \tag{4.14}$$

$$z_{ij} \in \{0, 1\}, \qquad \text{for } i = 1, ..., N, \;\; j = 1, ..., N \tag{4.15}$$

4.6 Model Extensions

4.6.1 Demography

Demography, especially in reintroduced groups, is extremely important in maximizing reproductive potential. For the first extension, demographic data on the current population are utilized so that sex ratio and age distribution within the transferred group may be specified in the model's constraints. After reordering the set of individuals, S, the model follows closely to the previous model (4.7). The set S shall be reordered such that males precede females; or $S = [s_1, ..., s_m, s_{m+1}, ..., s_{m+f}]$, where m is the total number of males and f is the total number of females. Furthermore, each sex group will be ordered

by age. This should be straightforward as individuals are given a studbook number based on their acquisition date; so that studbook numbers follow chronological order. Problems do arise, however, if studbooks of the same captive species are kept by multiple studbook keepers with different numbering systems, or if the living population contains individuals that were born outside of the population.

This formulation then easily lends itself to the addition of constraints that specify the number of males and females of certain ages to select. The objective function and most other constraints remain the same as those previously presented in (4.7). One difference is if the number of males and females are specified, the demand constraint (4.10) is obsolete. Again, this is assuming there are no other individuals at the transfer site.

4.6.2 Occupation of Transfer Site

The final formulation presented herein considers when individuals are present at the transfer site. In this scenario, $T \neq 0$, and kinships among those present in the destination population and between individuals in the source population and those in the destination population must be accounted for. First, the set S must be reordered such that $S = [s_1, ..., s_C, s_{C+1}, ..., s_{T+C}]$. In this manner, individuals in the source population (total=C) precede those present in the transfer site (total=T). From here, the K matrix can be divided as illustrated:

$$K = \left[\frac{C \times C \mid C \times T}{T \times C \mid T \times T}\right]$$

The formulation for the average mean kinship within the source population after transfers remains the same as described in (4.3). This is not the case with the average mean kinship of the transfer site since the addition of new individuals must be taken into consideration. The total kinships within and between all individuals in the transfer site after the addition of selected individuals can be partitioned into the following summations: the total kinship (including self kinships) among those present in the destination population prior to transfers, the total kinship between transferred individuals and those in the destination population, and the total kinships (including self kinships) among selected individuals. Average mean kinship can then be straightforwardly computed by dividing this amount by $(T + d)^2$. The objective function can be expressed intuitively as,

$$\min\ g(x) = \frac{1}{(C-d)^2}\sum_{i=1}^{C}\sum_{j=1}^{C}k_{ij}(1-x_i)(1-x_j) + \frac{1}{(T+d)^2}\left(\sum_{i=C+1}^{T+C}\sum_{j=C+1}^{T+C}k_{ij} + 2\sum_{i=C+1}^{T+C}\sum_{j=1}^{C}k_{ij}x_j + \sum_{i=1}^{C}\sum_{j=1}^{C}k_{ij}x_ix_j\right) \tag{4.16}$$

The definition of decision variables remains the same. This expression can then be handled in a similar manner to that of the original nonlinear model regarding simplification and the use of the substitution presented in (4.6). The constraint (4.9) is no longer appropriate, but a related constraint would be one that requires the new average mean kinship of the source site to be less than or equal to the average mean kinship prior to transfers.

Chapter 5

Problem Instance

Data pertaining to the California condor pedigree are applied to the nonlinear and linear models presented in (4.3) and (4.7), respectively. Neither of these formulations considers demographic data or the existence of other individuals at the transfer site. A description of the data precedes the results. Solutions are obtained by running the aforementioned models in Ilog OPL Development Studio (version 10) on a 2.13 GHz Intel Core2 computer with 2046 MB of memory. Lastly, a comparison between the performance of the linear integer programming (IP) model and MetaMk is presented.

5.1 Data Description

Michael Mace, Studbook Keeper of the California condor, provided parentage, births, and deaths of all California condors as of September 10, 2007 [24]. The first individuals included in the studbook are the 14 founders; the first of which was taken into captivity in February 1967 and was still living as of September 10, 2007. In total, there were 173 living individuals as of this date. The program, PM2000, was used to construct the kinship matrix [28]. The file required by PM2000 (.ped) was one that contained information on individuals accurate as of June 9, 2005. Because of the demonstrative purpose of this case study, these data were used in lieu of more recent data. Therefore, the results presented hereafter are relevant to the population as of June 9, 2005, which consisted of 150 living individuals. This population size is comparable to that of the more recent population and also to the size of many other captive endangered species populations [11]. A sample of the

data is contained in Appendix I.

These data contain individuals that reside in three wild populations and ten captive populations. In this problem instance, all of these individuals are considered to belong to the source population and are therefore equally likely to be transferred. This is unrealistic for those in wild populations but there were only a few of these wild individuals (five) included in the analysis. Also, the sex ratio of the considered population was approximately equal (73 females and 77 males) so that the calculation of gene diversity as previously presented is appropriate. Finally, all 14 founders were included in the analysis.

5.2 Results and Computation

To calculate kinships, PM2000 uses the recursive algorithm presented in (2.13) and (2.14) and assumes all individuals with wild parents are noninbred and unrelated to each other. The kinship matrix was exported in an Excel spreadsheet to be readily called by Ilog OPL. The original model presented in (4.1), the linear model presented in (4.7), and a simplified linear model (described below), were encoded in Ilog OPL using data and code presented in Appendices I and II, respectively. This program uses the powerful optimization solver, Cplex, to generate solutions to linear, integer, and quadratic programming problems.

First, for comparison purposes, all three models were executed with the same data and a demand of five transferred individuals ($d = 5$). The original (nonlinear) model was run for over an hour without producing an optimal solution. Upon termination, Ilog OPL found eight feasible solutions with a best integer solution value of .15198 and 7648 remaining branch and bound nodes. The linear model, (4.7), was then run. It should be noted that the objective function producing the true value of the sum of average mean kinships was alternatively used. This being,

$$\begin{aligned} \min f(x) \quad = \quad & \frac{1}{d^2 \times (N-d)^2} \left[[(N-d)^2 + d^2] \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} z_{ij} \right] \\ & + \frac{1}{(N-d)^2} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} - 2 \sum_{i=1}^{N} \sum_{j=1}^{N} k_{ij} x_i \right]. \end{aligned} \tag{5.1}$$

Compared to the objective function of the linear model presented in (4.8), the function (5.1) includes additional constants for the purpose of clearly illustrating the differences in

Table 5.1: Performance of the original model (Model 1), linear model (Model 2), and the simplified linear model (Alt. Mod. 2) in Ilog OPL for d=5.

	Model 1	**Model 2**	**Alt. Mod. 2**
Branch and Bound Nodes:	8685	301	0
Best Feasible Solution Value:	.15195	.15151	.15151
Elapse Time:	3740.89s	197.53s	28.10s

performance between the IP model and MetaMK. Using the same function, f, for both models assures equal objective function values for a given vector, $\mathbf{x}$. The linear model was run to completion in Ilog OPL and optimal value of .15151 was found in about three minutes.

For the linear model, an observation of the nature of term coefficients within the objective function and constraints leads to an alternative formulation. If the constraints that ensure $z_{ij} = 0$ when $x_i = x_j = 0$, (4.11) and (4.12), are removed, the remaining constraints and objective function are not altered; therefore, the $\mathbf{x}$ solution vector remains the same, although the Z matrix will change. To clarify, the coefficients of z_{ij} terms in both the objective function (4.8) and constraint (4.9), k_{ij}, is greater than or equal to zero, while the coefficient of the x_i term is less than or equal to zero. As a result, if $k_{ij} > 0$, $z_{ij} = 1$ only to satisfy the constraint that $z_{ij} = 1$ when $x_i = x_j = 1$, (4.13). Alternatively when $k_{ij} = 0$, z_{ij} can equal 0 or 1, but in this instance, values of z_{ij} are inconsequential. Hence by removing constraints (4.11) and (4.12), the relationship $z_{ij} = x_i x_j$ no longer holds indefinitely but the objective function value and $\mathbf{x}$ solution vector remain unchanged. Relatively simple adjustments to the objective function will allow for the initial relationship, $z_{ij} = x_i x_j$, to be upheld when the aforementioned constraints are removed. Herein, the simplified formulation of the linear model includes the true value objective function in (5.1) and constraints (4.9), (4.13) - (4.15).

The removal of constraints (4.11) and (4.12) did indeed decrease the runtime. Running this simplified formulation when $d = 5$ produced the same optimal value and $\mathbf{x}$ vector as the linear model but in less elapse time ($\approx$ 30 sec.). In this instance, the model did not require the exploration of any branch and bound nodes because the solution to the linear relaxation was integer. Results are summarized in Table 5.1.

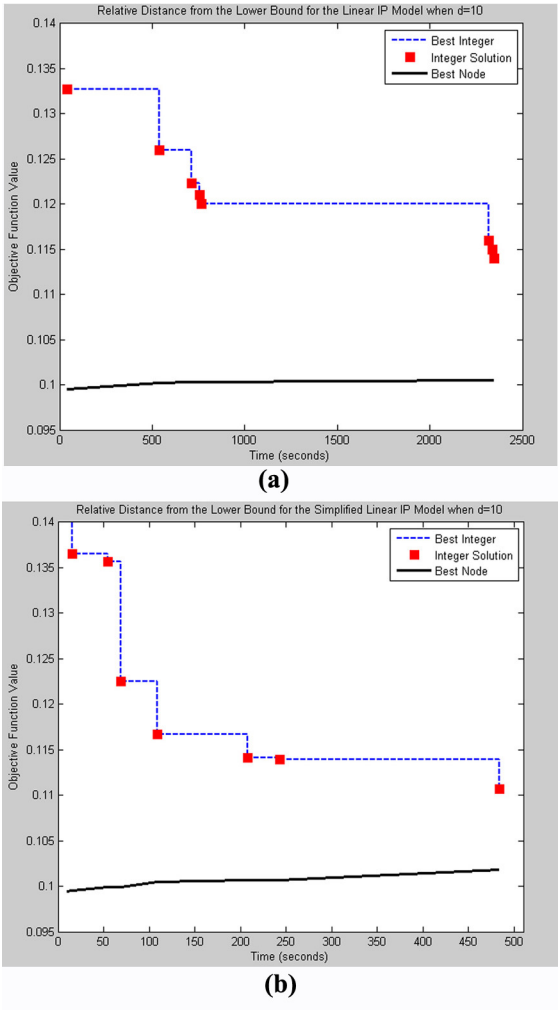

Figure 5.1: Relative distance from the lower bound with respect to time for (a) the original linear IP model and (b) the simplified linear IP model when d = 10.

The optimal solution of the two linear models is to select and transfer the group of individuals with studbook numbers 27, 31, 32, 36, and 45. This gives $\overline{MK}_{SOURCE}$ and $\overline{MK}_{TRANSFER}$ of 0.05151 and 0.10, respectively. The average mean kinship of the transfer population, 0.10, is the absolute minimum value for a group of five individuals. In other words, all pairs of individuals have a kinship value of zero, and a self kinship

Table 5.2: Performance of the linear model (Model 2) and the simplified linear model (Alt. Mod. 2) for varying demand. Note: $\overline{MK}_S = \overline{MK}_{SOURCE}$, $\overline{MK}_T = \overline{MK}_{TRANSFER}$

	$d = 6$	$d = 7$	$d = 8$	$d = 9$	$d = 10$
$\overline{MK}_S$	.0518	.0522	.0527	.0530	.0527
$\overline{MK}_T$	.0833	.0714	.0625	.0586	.0575
$\overline{MK}_S + \overline{MK}_T$	.1351	.1236	.1152	.1116	.1102
Elapse Time (Model 2)	121.07s	1874.79s	-	-	-
Elapse Time (Alt. Mod. 2)	186.01s	313.49s	447.59s	2674.00s	>2h 30s

value of 0.5. Compared to the average mean kinship of the source population prior to transferring individuals, 0.05157, this solution gives an improvement. To put these values into perspective, an average mean kinship value of 0.063 is equivalent to the population being related on average at level of first cousins [29]. Recall that this measure is sensitive to population size, so that a small group of unrelated individuals may have a relatively high average mean kinship. In small populations, most individuals will be related after only a few generations of random mating.

Next, both linear formulations were used to assess the performance of the optimization model with an increase in demand. Results are presented in Table 5.2. Note that for $d = 10$, the program was terminated after 2.5 hours; the corresponding solution reflects the best integer solution found in this time limit. For most instances, the simplified linear formulation was much more efficient in achieving or coming close to optimality. Figure 5.1 illustrates this graphically by showing how quickly each linear IP formulation converges to the linear relaxation optimal solution when $d = 10$. For this reason, the simplified linear model was used for larger values of d in lieu of the linear model initially proposed.

One observation in Table 5.2 is that, as d increases, the best integer solution lends decreasing values of the total average mean kinship ($\overline{MK}_{SOURCE} + \overline{MK}_{TRANSFER}$). This could suggest there is an optimal demand, not considered here, for which gene diversity is maximized among both populations. If it is feasible to transfer a large group of individuals, analysis such as this would prove beneficial as there is greater potential to maximize gene diversity within populations if d is variable.

5.3 Comparison to MetaMK

As previously stated, MetaMK measures the effect of transferring a specific individual on the proportional gene diversity in both the source and destination populations. For each individual, the change in proportional gene diversity relative to the source and destination populations are denoted as ΔGD_A and ΔGD_B in MetaMK, respectively. These values are calculated as the change in proportional gene diversity that would occur if the individual is transferred to the other population. Specifically, the change in proportional gene diversity of the source population, defined more descriptively as ΔGD_{SOURCE}, resulting from the transfer of individual i, is $\Delta GD_{SOURCE} = GD_{NEW} - GD_{PRIOR} = (1 - \overline{MK}_{NEW}) - (1 - \overline{MK}_{PRIOR})$. GD_{PRIOR} and GD_{NEW} refer to the proportional gene diversity in the source population before and after individual i is transferred; similarly with $\overline{MK}_{PRIOR}$ and $\overline{MK}_{NEW}$.

A comparison between the performance of MetaMK when using several different selection procedures to that of the proposed model follows. The data are the same as that used in the previous section to compare the IP models. First, a description of how individuals are selected under each selection strategy, as well as the goal of the proposed strategy is presented. Then, the results from applying each strategy to the given data and the comparison to the proposed integer program model are summarized. The first two selection procedures are similar to those used by Ralls and Ballou [29] and are described in the context of population management of the California condor. The last is a general, alternative procedure.

Selection Procedure 1 (SP 1 in Table 5.3): New Captive Population. This selection procedure was considered by Ralls and Ballou [29] for the purpose of selecting breeding pairs to establish a new captive population. Herein, this procedure is altered by selecting individuals rather than pairs. Ralls and Ballou [29] use MetaMK to accomplish this by first choosing the individual whose removal least affects the gene diversity present in the source population and iteratively selecting subsequent individuals that lend the greatest increase in the gene diversity of the new zoo population. Alternatively, if the first individual is chosen that maximizes genetic gain in the source population, individuals of rare lineages are selected thereafter as a result. The performed method avoids transferring as many

genetically valuable individuals. Therefore, there is less of a reduction of gene diversity within the source population but yet not as much gain in gene diversity within the new zoo population.

Selection 2 (SP 2 in Table 5.3): Transferring Individuals Between Two Wild Populations. Ralls and Ballou [29] used the following procedure to place chicks originating in captivity among two wild populations. The placement of chicks was decided by arbitrarily assigning all chicks to one wild location, then using MetaMK to relocate chicks whose removal maximized the average change in gene diversity. Ralls and Ballou [29] first assigned all chicks to the CA population and transferred chicks to the AZ population whose removal maximized the average change in gene diversity in CA and AZ. Some chicks transferred to CA negatively affected the genetic diversity of that population. It was not made clear, however, if the average mean kinship of this population was increased due to these transfers.

Selection 3 (SP 3 in Table 5.3): General Method. The goal for this selection strategy is similar to that of Selection Procedure 1, in that it avoids decreasing the gene diversity of the source population. The method is to first sort the individuals in the source population by $\Delta GD_{TRANSFER}$, the change in proportional gene diversity of the destination population given the individual was transferred. Then select the individuals with the highest $\Delta GD_{TRANSFER}$ and ΔGD_{SOURCE} such that $\Delta GD_{SOURCE} \geq 0$. If two or more individuals have the same $\Delta GD_{TRANSFER}$ and ΔGD_{SOURCE}, the individual with the highest mean kinship, mk, in captivity is selected. This secondary criterion is also used in the first two selection strategies but this was not explicitly stated as a criterion used by Ralls and Ballou [29].

The results in Table 5.3 show that the IP model outperformed the iterative method, with Selection Procedure 2 producing the smallest difference of .00021. Observe also that SP 2 increased the average mean kinship of the source population by 0.00015, while SP 3 decreased the mean kinship of the source population but significantly increased that in the destination population.

If additional individuals are required to transfer, the solution set obtained by using MetaMK would contain the previously chosen individuals. In general, the time using

Table 5.3: A summary of the results of each selection method with a comparison to the linear model (IP Model 2). Note: $\overline{MK}_S = \overline{MK}_{SOURCE}$, $\overline{MK}_T = \overline{MK}_{TRANSFER}$

	IP Model 2	SP 1	SP 2	SP 3
Selected Individuals	27,31,32,36,45	389,33,1,5,31	27,32,33,44,45	44,32,45,76,36
$\overline{MK}_S$	.0515	.0524	.0517	.0512
$\overline{MK}_T$	.1000	.1025	.1000	.1150
$\overline{MK}_S$+$\overline{MK}_T$	.1515	.1549	.1517	.1662

Table 5.4: Comparison of solutions obtained by the simplified linear model (IP Alt. Mod. 2) and MetaMK (SP 2) for d=10. Note: $\overline{MK}_S = \overline{MK}_{SOURCE}$, $\overline{MK}_T = \overline{MK}_{TRANSFER}$

	IP Alt. Mod. 2	MetaMK SP 2
Selected Individuals	1,120,20,27,31,32,36,44,5,7	27,32,33,44,45,31,36,43,5,1
$\overline{MK}_S$	.0527	.0523
$\overline{MK}_T$	.0575	.0600
$\overline{MK}_S$+$\overline{MK}_T$	.1102	.1123

MetaMK increases linearly with the number of individuals chosen, although this would vary with the user and selection criteria. For the considered data set, the time to select one individual was about 5 seconds; this does not include time spent to initialize the program. Table 5.4 compares solutions obtained through the simplified linear model (Alt. Mod. 2) and MetaMK for $d = 10$.

The solution to the simplified IP model considered in Table 5.4 pertains to the best feasible solution found at the time of termination (2.5 hrs) as illustrated in Table 5.2. Although optimality was not proven, the IP model again found a better solution than MetaMK. In this instance, the difference between solutions is more significant (0.0021). The disadvantage to using the IP model in this case, is that computation time was significantly greater. It should be noted, however, that the IP model found a feasible solution that lends an improved objective function value (.11067) compared to that found in MetaMk, in a little over seven minutes.

In summary, there are several advantages to using the IP model as opposed to MetaMK. First, without comparison to the optimization model, the performance of the it-

erative method is unknown as this method does not prove optimality or give a lower bound for all possible solutions like the IP model. For this problem instance, as confirmed by the solution to the IP model, SP 2 gives a near optimal solution when d is small. Second, the time to reach a feasible solution may be much less using MetaMK, but this is not necessarily true if several individuals equally satisfy the same criteria, such as $\Delta GD_{TRANSFER}$ and ΔGD_{SOURCE}. In this case, trial and error is involved in finding a group to transfer as subsequently selected individuals depend upon those previously chosen. Another consideration is that for these types of management decisions, it is presumed that computation time is not of major concern.

Another aspect of MetaMK is that it gives users freedom in selecting individuals, which may lead to beneficial or detrimental management decisions. The results suggest that SP 2 best achieves the goal of minimizing average mean kinship among two populations. There may be motivation to use alternative selection procedures, such as those previously described. The goals of the selection procedure must be thoroughly understood by the user, as each produces varying solutions. It is much more beneficial to quantify selection criteria and use the IP model, as constraints can be readily added and the relative distance to the lower bound can be assessed. Lastly, this iterative method does not protect against user error. It is quite easy to overlook individuals that best fit given criteria, and as previously mentioned, this will affect which individuals are subsequently selected. Selecting individuals becomes a more difficult task as number of criteria increases. User error is much less a factor with the IP model because unless there is an error in coding, the same model can be run repeatedly with multiple data sets.

5.4 Model Improvements

It should be noted that to more accurately reflect the goal of managing gene diversity within populations several modifications should be made. First, as previously mentioned, individuals past reproductive age should not be included in the calculation of proportional gene diversity as they no longer contribute to the genetic diversity of future populations. Also constraints should be added to avoid the selection of certain individuals, such as founders, from being relocated (i.e. to a wild population) if this is undesirable. Furthermore it is suggested by Ralls and Ballou [29] that pairs of individuals with differences

in mean kinship more than 1.5% should be avoided. Therefore, any chosen male and female should satisfy this constraint.

Chapter 6

Conclusion

The objective of this thesis was to provide a straightforward optimization model for the purpose of maximizing proportional gene diversity within the source and destination populations, and also to give a comprehensive overview of the statistical theory and methods in pedigree analysis involved in calculating this measure of genetic diversity. Few publications present the management practice of minimizing average mean kinship in this type of framework (but see: [4], [22]). This background knowledge is essential in correctly applying and interpreting the average mean kinship of a population in the context of population management. As previously mentioned, for this metric, as computed herein, to be an accurate assessment of gene diversity of future generations many assumptions must be met. Primarily, mating is random between individuals for which average mean kinship is computed. In other words, matings are uncontrolled within populations and all individuals are equally likely to breed and produce offspring. If breeding is controlled to some extent so that there is an avoidance of inbreeding in the population, the average mean kinship of the population will give an estimate of the minimum gene diversity of the offspring.

As shown, the optimization approach to the considered problem instance outperforms the existing iterative method; although as supported by the solution to the IP model, one selection procedure comes close to optimality when few individuals are selected. There are several key advantages to using an optimization model in this context, as opposed to the iterative method of MetaMK. First, for smaller problem instances optimality can be achieved within a reasonable amount of time, and at all times the nearness to the lower bound can be assessed for every feasible solution. As presented, a simplification to

the presented linear IP model quickly converges to the neighborhood of optimality for the considered demand values. Furthermore, the optimization model can easily accommodate additional constraints and may be altered to satisfy problem extensions. Lastly, the use of the optimization model reduces the risk of user error which is substantial in MetaMK.

The relocation of animals, especially for the purpose of reintroduction, is a complex process and the best management practice for one species may not be desirable or feasible for another. The presented models only consider one particular population management goal and ignore many factors, such as logistic and biological constraints. To identify the best strategy, managers must define the program goal based on the biology of the species of interest and the phase of conservation for which the species is in. The following are two examples of when the present goal could be considered inappropriate. When a species is in the first stage of reintroduction, due to the uncertainty in survivorship at the reintroduction site, genetically less valuable animals are transferred until survivorship is more predictable. Additionally, the presented model would be less applicable to large captive populations, which are less likely to suffer consequences from genetic losses resulting from transfers.

In a broader scope, modeling problems in population management in an optimization framework is a way to consider many factors in a quantitative way, instead of relying on professional judgment that may overlook optimal solutions. Finding optimal management strategies is especially important in conservation programs to best ensure the long term survival of populations. It is hoped that more optimization models such as this, are incorporated into decision making processes of species population management.

Chapter 7

Avenues of Future Research

There are several direct extensions of the proposed model which would be applicable to a wider array of management tasks. The first would allow for individuals to be placed among more than one transfer sites. The current model with several adjustments could accommodate this, but the resulting increase in the number of decision variables could lead to substantial computation time for large data sets. Next, the incorporation of measures of allelic diversity would ensure a proper distribution of founder alleles within subpopulations; average mean kinship does not explicitly account for allele loss, only heterozygosity. Through the use of gene drop analysis, the amount of a particular founder's genome surviving in the population can be estimated by performing and averaging many Monte Carlo pedigree simulations. Additionally, gene drop analysis can provide an individual measure, termed *genomic uniqueness*, which gives the frequency of simulations for which that individual contained alleles unique to the population. Many population management programs that base breeding recommendations on the ranking of individuals by mean kinship, also consider each individual's genomic uniqueness [3]. This increases the probability that rare alleles are passed on to following generations. The extent to which minimizing average mean kinship retains allelic diversity may vary depending on the species' pedigree. Ralls and Ballou [29] performed gene drop analysis to assess the distribution of founder alleles resulting from the placement of individuals based on minimizing average mean kinship. They found that founder alleles were sufficiently represented in all subpopulations.

Another area of needed improvement is finding a measure that gives a more realistic estimate of genetic diversity within a population. The current measure ignores the diversity

of linked genes and genetic diversity which is a result of mutations. Wray [36] adapted the previously presented methods on constructing the kinship matrix to account for the random effects of mutations and the inheritance of these genes through the pedigree. Furthermore, Toro et al. [35] incorporated information on molecular genetic markers to find more accurate estimates of kinship coefficients via Monte Carlo Markov Chains.

Finally, several improvements to the integer program and branch and bound algorithm could result in a more time efficient model. First, generating added constraints (cuts) which tighten bounds can reduce run time. Furthermore, cuts or further simplifications which take advantage of the sparsity of the kinship matrix, K, may prove beneficial. Lastly, one may also produce a more efficient branch and bound algorithm specifically for this problem that branches on variables more intelligently.

Bibliography

[1] Ballou, J. D. (1999). MetaMK: Metapopulation Mean Kinship Analysis Software (Version 2.0.0a) [Software]. Available from http://consgen.mq.edu.au/Links%20and%20Downloads.htm.

[2] Ballou, J. D., & Foose, T. J.(1996). *Demographic and Genetic Management of Captive Populations.* In D. G. Kleiman, M. E. Allen, K. V. Thompson, S. Lumpkin (Eds.), *Wild Mammals in Captivity* (pp. 263-283). Chicago and London: The University of Chicago Press.

[3] Ballou, J. D. & Lacy, R. C. (1995). *Identifying Genetically Important Individuals for Management of Genetic Variation in Pedigreed Populations.* In J. D. Ballou, M. Gilpin, & T. J. Foose (Eds.), *Population Management for Survival and Recovery* (pp. 76-111). New York: Columbia University Press.

[4] Caballero, A., & Toro, M. A. (2000). Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetics Research, 75*(3) 331-343.

[5] Chang, H. L., Fernando, R. L., & Grossman, M. (1991). On the principle underlying the tabular method to compute coancestry. *Theoretical and Applied Genetics, 81,* 233-238.

[6] Cockerham, C. C. (1969). Variance of gene frequencies. *Evolution, 23,* 72-84.

[7] Collen, B., Ram, M., Dewhurst, N., Clausnitzer, V., Kalkman, V., Cumberlidge, N., & Baillie, J. E. M. (2008). *Broadening the Coverage of Biodiversity Assessments.* Retrieved August 21, 2008, from http://cmsdata.iucn.org/downloads/ broadening_the_coverage_of_biodiversity_assessments.pdf.

[8] Crnokrak, P., & Roff, D. A. (1999). Inbreeding depression in the wild. *Heredity, 83*(3), 260-270.

[9] Cronin, K. A., Mitchell, M. A., Lonsdorf, E. V., & Thompson, S. D. (2006). One year later: evaluation of PMC-recommended births and transfers. *Zoo Biology, 25*(4), 267-277.

[10] Crow, J. F., & Kimura, M. (1970). *An Introduction to Population Genetics Theory.* Minneapolis, MN: Burgess Publishing Company.

[11] Earnhardt, J. M. (1999). Reintroduction programmes: genetic trade-offs for populations. *Animal Conservation, 2,* 279-286.

[12] Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Harlow, England: Prentice Hall.

[13] Fernández, J., Toro, M. A., & Caballero, A. (2001). Practical implementation of optimal management strategies in conservation programmes: a mate selection method. *Animal Biodiversity and Conservation, 24*(2), 17-24.

[14] Fischer, J. & Lindenmayer, D. B. (2000). An assessment of the published results of animal relocations. *Biological Conservation 96*(1), 1-11.

[15] Frankham, R., Ballou, J.D., & Briscoe, D.A. (2002). *Introduction to Conservation Genetics*. Cambridge, UK: Cambridge University Press.

[16] Haig, S. M., Ballou, J. D., & Derrickson, S. R. (1990). Management options for preserving genetic diversity: reintroduction of Guam rails to the wild. *Conservation Biology, 4*(3), 290-299.

[17] IUCN (1998). *Guidelines for Re-introductions.* Prepared by the IUCN/SSC Reintroduction Specialist Group. IUCN, Gland, Switzerland and Cambridge, UK.

[18] Johnson, A. M. (2004). *Estimation and sampling properties of gene diversity, heterozygosity and FST*. Retrieved from NCSU Electronic Theses and Dissertations (etd-12232004-163646).

[19] Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

[20] Kostreva, M. M., Ogryczak, W., & Tonkyn, D. W. (1999). Relocation problems arising in conservation biology. *Computers and Mathematics with Applications, 37,* 135-150.

[21] Lacy, R. C. (1989). Analysis of founder representation in pedigrees: Founder equivalents and founder genome equivalents. *Zoo Biology, 8*(2), 111-123.

[22] Lacy, R. C. (1995). Clarification of genetic terms and their use in the management of captive populations. *Zoo Biology, 14,* 565-578.

[23] Lacy, R. C., Ballou, J. D., Princée, F., Starfield, A., & Thompson, E. A. (1995). Pedigree Analysis for Population Management. In J. D. Ballou, M. Gilpin, & T. J. Foose (Eds.), *Population Management for Survival and Recovery* (pp. 57-75). New York: Columbia University Press.

[24] Mace, M. (2008). California Condor studbook, 2008. San Diego Zoo, San Diego, CA.

[25] Montgomery, M. E., Ballou, J. D., Nurthen, R. K., England, P. R., Briscoe, D. A., & Frankham, R. (1997). Minimizing kinship in captive breeding programs. *Zoo Biology 16*(5), 377-389.

[26] Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America, 70*(12), 3321-3323.

[27] Pierce, B. A. (2003). *Genetics: A Conceptual Approach*. New York: W. H. Freeman and Company.

[28] Pollak, J. P., Ballou, J. D., & Lacy R. (2007). PM2000: Analysis and Management of Pedigreed Populations Software (Version 1.213) [Software]. Available from http://www.vortex9.org/pm2000.html.

[29] Ralls, K., & Ballou, J. D. (2004). Genetic status and management of California condors. *The Condor, 106*(2), 215-227.

[30] Ralls, K., Ballou, J. D., Templeton, A. (1988). Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation Biology, 2*(2), 185-194.

[31] Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology, 17*(1), 230-237.

[32] Rodríguez-Clark, K. M. (1999). Genetic Theory and Evidence Supporting Current Practices in Captive Breeding for Conservation. In L. F. Landweber & A. P. Dobson (Eds.). *Genetics and the Extinction of Species* (pp. 47-73). Princeton, NJ: Princeton University Press.

[33] Seddon, P. J., Armstrong, D. P., & Maloney, R. F. (2007). Developing the science of reintroduction biology. *Conservation Biology, 21*(2), 303-313.

[34] Stanley Price, M. R. & Soorae, P. S. (2003). Reintroductions: whence and whither? *International Zoo Yearbook 38*(1), 61-75.

[35] Toro, M., Silió, L., Rodrigañez, J., Rodriguez, C., & Fernández, J.(1999). Optimal use of genetic markers in conservation programmes. *Genetics Selection Evolution, 31*, 255-261.

[36] Wray, N. R. (1990). Accounting for mutation effects in the additive genetic variance-covariance matrix and its inverse. *Biometrics 46*(1), 177-186.

Appendices

APPENDIX I: Sample of Data

California Condor Studbook (Partial)

CALIFORNIA CONDOR Studbook (Gymnogyps californianus) Page 3

Stud #	Sex	Hatch Date	Sire	Dam	Location	Date	Local ID	Event	Death Date	Rearing	Tag/Band	Name
22	F	~ 1980	6	10	CALIFORNI	~ 1980	______	Hatch		Parent		
						~30 Jun 1980		Death	~30 Jun 1980			
23	M	~ 1981	4	8	CALIFORNI	5 Dec 1982	NONE	Capture		Parent		PAXA
					LOSANGELE	5 Dec 1982	001130	Transfer				
					SD-WAP	29 Aug 1984	027003	Transfer				
					PORTLAND	20 Nov 2003	030292	Transfer				
24	?	~ 1981	3	12	CALIFORNI	~ 1981	______	Hatch		Parent		WGI
						~ 1982		Death	~ 1982			
25	M	~ 1982	3	12	CALIFORNI	13 Aug 1982	NONE	Capture		Parent		XOLXOL
					SD-WAP	13 Aug 1982	012173	Transfer				
					LOSANGELE	28 Mar 1983	001131	Transfer				
					SD-WAP	29 Aug 1984	012173	Transfer				
26	F	~ Apr 1982	6	10	CALIFORNI	~ Apr 1982	______	Hatch		Parent		BOS
						~ Nov 1983		Death	~ Nov 1983			
27	M	~ Apr 1983	4	8	CALIFORNI	4 Aug 1983	CUYAMA	Capture		Parent		CUYAMA
					LOSANGELE	4 Aug 1983	001359	Transfer				
					BRDS PREY	18 Nov 2003	______	Transfer				
28	M	30 Mar 1983	2	11	SANDIEGOZ	30 Mar 1983	012174	Hatch		Hand		SISQUOC
					SD-WAP	31 Mar 1983	012174	Transfer				
29	F	25 May 1983	2	11	SANDIEGOZ	25 May 1983	012176	Hatch		Hand		SESPE
					SD-WAP	25 May 1983	012176	Transfer				
					LOSANGELE	28 Aug 1984	001853	Transfer				
					SD-WAP	5 Sep 1992	012176	Transfer				
30	F	5 Apr 1983	3	12	SANDIEGOZ	5 Apr 1983	012175	Hatch		Hand		TECUYA
					SD-WAP	6 Apr 1983	012175	Transfer				
					BRDS PREY	23 Sep 1993	TECUYA	Transfer				
31	F	2 Jul 1983	3	12	CALIFORNI	8 Nov 1983	CACHUM	Capture		Parent		CACHUMA
					LOSANGELE	8 Nov 1983	001475	Transfer				
					SD-WAP	2 Feb 2000	800008	Transfer				
32	F	27 May 1983	6	10	SANDIEGOZ	27 May 1983	012177	Hatch		Hand		ALMIYI
					SD-WAP	28 May 1983	012177	Transfer				
33	M	16 May 1984	7	9	CALIFORNI	15 Sep 1984	SEQUOI	Capture		Hand		SEQUOIA
					LOSANGELE	15 Sep 1984	001988	Transfer				

Kinship Matrix (Partial)

	1	104	11	111	120	121	125	135	137	138	139	140	141	145	147	154	155	157	159
1	0.5	0	0	0	0	0	0	0	0	0	0	0.25	0.25	0	0	0	0	0	0.25
104	0	0.5	0.125	0.0313	0	0	0.0313	0.0625	0.0625	0	0	0.0625	0.0625	0.0625	0.0625	0.0313	0.0625	0	0.0625
11	0	0.125	0.5	0.0625	0	0	0.0625	0.125	0.125	0	0	0	0	0	0	0.0625	0.25	0	0
111	0	0.0313	0.0625	0.5	0.0625	0.0625	0.25	0.0313	0.0938	0.0625	0.0625	0	0	0.0313	0	0.25	0.0938	0.0625	0
120	0	0	0	0.0625	0.5	0.25	0.0625	0	0	0.25	0.25	0	0	0	0	0.0625	0.0625	0.25	0
121	0	0	0	0.0625	0.25	0.5	0.0625	0	0	0.25	0.25	0	0	0	0	0.0625	0.0625	0.25	0
125	0	0.0313	0.0625	0.25	0.0625	0.0625	0.5	0.0313	0.0938	0.0625	0.0625	0	0	0.0313	0	0.25	0.0938	0.0625	0
135	0	0.0625	0.125	0.0313	0	0	0.0313	0.5	0.0625	0	0	0	0	0	0	0.0313	0.0625	0	0
137	0	0.0625	0.125	0.0938	0	0	0.0938	0.0625	0.5	0	0	0	0	0	0	0.0938	0.125	0	0
138	0	0	0	0.0625	0.25	0.25	0.0625	0	0	0.5	0.25	0	0	0	0	0.0625	0.0625	0.25	0
139	0	0	0	0.0625	0.25	0.25	0.0625	0	0	0.25	0.5	0	0	0	0	0.0625	0.0625	0.25	0
140	0.25	0.0625	0	0	0	0	0	0	0	0	0	0.5	0.25	0.0625	0.0625	0	0	0	0.25
141	0.25	0.0625	0	0	0	0	0	0	0	0	0	0.25	0.5	0.0625	0.0625	0	0	0	0.25
145	0	0.0625	0	0.0313	0	0	0.0313	0	0	0	0	0.0625	0.0625	0.5	0.0625	0.0313	0	0	0.0625
147	0	0.0625	0	0	0	0	0	0	0	0	0	0.0625	0.0625	0.0625	0.5	0	0	0	0.0625
154	0	0.0313	0.0625	0.25	0.0625	0.0625	0.25	0.0313	0.0938	0.0625	0.0625	0	0	0.0313	0	0.5	0.0938	0.0625	0
155	0	0.0625	0.25	0.0938	0.0625	0.0625	0.0938	0.0625	0.125	0.0625	0.0625	0	0	0	0	0.0938	0.5	0.0625	0
157	0	0	0	0.0625	0.25	0.25	0.0625	0	0	0.25	0.25	0	0	0	0	0.0625	0.0625	0.5	0
159	0.25	0.0625	0	0	0	0	0	0	0	0	0	0.25	0.25	0.0625	0.0625	0	0	0	0.5

APPENDIX II: Ilog Cplex Code

Nonlinear Model

```
/*********************************************
 * OPL 5.5 Model
 * Author: sdallen
 * Creation Date: 10/16/2008 at 4:14 PM
 *********************************************/

{int} N = ...; //studbook number of individuals listed in order of
               //kinship matrix
int T = ...;  //Total number of living individuals
int d = ...;  //Wild demand
float kin_val[N][N] = ...;  //kinship matrix

dvar boolean x[N];  //1 if x[N] is chosen for transfer;
                    //0 otherwise

minimize
    (sum (j in N, n in N) kin_val[j,n]*x[j]*x[n])/(d*d) +
(sum (j in N, n in N) kin_val[j,n]*(1-x[j])*(1-x[n]))/((T-d)*(T-d));

subject to{
    ctdemandismet:
    sum (j in N) x[j]==d;

    ctmkislessincaptive:
    (sum (j in N, n in N) kin_val[j,n]*x[j]*x[n])/(d*d) >=
(sum (j in N, n in N) kin_val[j,n]*(1-x[j])*(1-x[n]))/((T-d)*(T-d));
}
```

Linear Model with True Objective Function Value

```
/*********************************************
 * OPL 5.5 Model
 * Author: ShanaeDomenica Allen
 * Creation Date: 10/20/2008 at 1:13 PM
 *********************************************/

{int} N = ...; //studbook number of individuals listed in order of
               //kinship matrix
int T = ...;  //Total number of living individuals
int d = ...;  //Transfer site demand
float kin_val[N][N] = ...;  //kinship matrix

dvar boolean x2[N];  //1 if x2[N] is chosen for transfer;
                     // 0 otherwise
dvar boolean z2[N][N]; // z2[i][j]=x2[i]x2[j]

minimize
    (((T-d)*(T-d) +d*d)* sum (j in N, n in N) kin_val[j,n]*z2[j,n]
      -(2*d*d)*sum (j in N, n in N) kin_val[j,n]*x2[n]
      +d*d*(sum (j in N, n in N) kin_val[j,n]))/((T-d)*(T-d)*d*d);

subject to{
    ctdemandismet:
    sum (j in N) x2[j]==d;

    ctmkislessincaptive:
    ((T*T - 2*T*d)/(d*d))*sum (j in N, n in N) kin_val[j,n]*z2[j,n]>=
     sum (j in N, n in N) kin_val[j,n]
     - 2*sum (j in N, n in N) kin_val[j,n]*x2[n];

    forall( i in N, j in N)
    ctz_iszero_when_xi_iszero:

    z2[i][j]<=x2[i];

    forall( i in N, j in N)
    ctz_iszero_when_xj_iszero:

    z2[i][j]<=x2[j];
```

```
    forall( i in N, j in N)
    ctz_isone_when_xixj_isone:

    z2[i][j]>= x2[i]+x2[j] -1;
}
```