

# **Monitoring the Stephen's Kangaroo Rat:**

## **An Analysis Of Monitoring Methods And Recommendations For Future Monitoring.**

James E. Diffendorfer<sup>1</sup> and Douglas H. Deutschman<sup>2</sup>.

Department of Biology  
San Diego State University  
San Diego, CA 92104

<sup>1</sup> [jdifffen@sunstroke.sdsu.edu](mailto:jdifffen@sunstroke.sdsu.edu), 619-594-0311

<sup>2</sup> [doug@sciences.sdsu.edu](mailto:doug@sciences.sdsu.edu), 619-594-5391

## TABLE OF CONTENTS

LIST OF FIGURES.....	4
LIST OF TABLES.....	5
EXECUTIVE SUMMARY .....	6
SYNOPSIS AND CRITIQUE OF AVAILABLE DATA .....	8
ANALYSIS OF TRAPPING DATA .....	10
1. Point Estimation of SKR Density .....	10
A. Analysis of Burrow Counting Methods.....	10
B. Power Analysis of Burrow Counting Methods.....	12
C. Relationship between Burrow Counts and Density Estimates.....	13
D Conclusions Regarding Burrow Counts and Density. ....	17
2. Power Analysis of Trapping Methods .....	19
3. Point estimation using Program MARK.....	23
A. Performance of MARK. ....	24
B. Number of Unique Individuals Captured and Capture Probability. ....	27
C. Estimation in MARK and Sampling Effort. ....	30
D. Numbers of Captures as an Estimate of Density. ....	33
E. Conclusions from Analyses of Program MARK. ....	34
4. Monitoring Changes in SKR Density Through Time and Across Space: .....	34
A. Spatial and Temporal Variability of Burrow Counts.....	35
B. Spatial and Temporal Variability of Burrow Counts and Density.....	36
C. Spatial and Temporal Variability of Density .....	38
D. Temporal Autocorrelation in Density Data and Sampling Frequency. ....	40
CONCLUSIONS AND RECOMMENDATIONS .....	41
1. CONCLUSIONS AND RECOMMENDATIONS .....	41
A. Point Estimation of Abundance.....	41
B. Monitoring Across Space and Through Time.....	41
C. Effort and timing of trapping for trend detection. ....	42
D. The use of MARK and estimating vital rates. ....	43

2. RECOMMENDATIONS FOR ADDITIONAL MONITORING .....	43
A. Synchronized Trapping Across All Reserves. ....	44
B. Estimating Demographic Processes. ....	45
C. Habitat and environment relationships. ....	46
D. Population Viability or Simulation Modeling of SKR. ....	47
E. Prioritizing Management Actions. ....	47
LITERATURE CITED .....	49
APPENDIX 1 .....	50
APPENDIX 2 .....	51

## LIST OF FIGURES

<b>Figure 1:</b> Relationship between Burrow Counts and Burrow Lines based on 37 sites in 1989 and 1990. ....	11
<b>Figure 2:</b> Accuracy of transect lines as a function of effort. ....	13
<b>Figure 3:</b> Relationship between burrow count (X, burrows per hectare) and density (Y, unique individuals per hectare). ....	15
<b>Figure 4:</b> Prediction of density (Individuals per ha) from burrow counts (Burrows per ha) at 94 samples spanning three studies between 1989-2000. ....	16
<b>Figure 5:</b> Year-to-Year variability in the Burrow Count – Density Relationship. ....	18
<b>Figure 6:</b> Relationship between numbers of individuals captured with different levels of effort (x-axis) and full sampling (Y-axis) for data collected by Kelt at two grids with relatively high abundance. ....	21
<b>Figure 7:</b> Relationship between numbers of individuals captured with different levels of effort (x-axis) and full sampling (Y-axis) for data collected by O'Farrell at two grids with relatively high abundance. ....	22
<b>Figure 8:</b> Accuracy as a function of sampling effort from representative trapping grids used by Kelt and O'Farrell. ....	23
<b>Figure 9:</b> Estimates of population size from MARK ("N_HAT", y-axis) against the number of individuals captured during a trapping session for Kelt's Plot NO1. ....	26
<b>Figure 10:</b> A histogram and fitted normal curve for capture probabilities taken across data collected by Kelt and O'Farrell. ....	28
<b>Figure 11:</b> Capture probability by trapping session for Kelt's grid, NO1. ....	28
<b>Figure 12:</b> The number of individuals captured in a trapping session versus the captures per individual per day (CAP_PER_IND) during that trapping session. ....	29
<b>Figure 13:</b> The relationship between estimates of probability of capture from MARK and the number of individuals captured during a session for a combined series of data. ....	30
<b>Figure 14:</b> Precision of Survival Estimates from Program MARK as a function of grid size. ....	32
<b>Figure 15:</b> Number of captures in a trapping session plotted against the number of unique individuals captured. ....	33
<b>Figure 16:</b> Variation in burrow counts through time at two sites. ....	37
<b>Figure 17:</b> Time series of the number of unique individuals per trapping session on 4 plots trapped from 1996-2000 by D. Kelt. ....	38
<b>Figure 18:</b> Partial autocorrelation plot. ....	40

## LIST OF TABLES

<b>Table 1.</b> <i>General descriptions of the datasets received.</i> .....	9
<b>Table 2.</b> <i>Results of the unconstrained regression of burrow cells (Y) against burrow lines (X) from 37 sites at Eastside Reservoir in 1989 and 1990.</i> .....	12
<b>Table 3.</b> <i>Summary of unconstrained regression between burrow count (X, burrows per hectare) and density (Y, unique individuals per hectare.</i> .....	15
<b>Table 4.</b> <i>Results from separate regressions of density (Y; animals per ha) against burrow lines (X; burrows per ha) from the three major studies.</i> .....	17
<b>Table 5.</b> <i>Trap effort by full (shaded) and reduced efforts of sampling.</i> .....	19
<b>Table 6.</b> <i>Number of individuals captured <math>M_{(t+1)}</math>, estimate of population size from MARK as well as the associated error terms for plot NO1 of Kelt.</i> .....	25
<b>Table 7.</b> <i>Sums of Squares decomposition for March AFB (4 grids, 3 years).</i> .....	38
<b>Table 8.</b> <i>Sums of Squares decomposition for Eastside Reservoir (8 grids, 2 years).</i> .....	38
<b>Table 9.</b> <i>Sums of Squares decomposition for Shipley (4 grids, 4 years).</i> .....	39

## **EXECUTIVE SUMMARY**

We performed a series of analyses to explore and summarize previous monitoring efforts of Stephens Kangaroo Rats (SKR) and make recommendations for future monitoring and management. We reviewed a large amount of data collected from 1989 – 2001 provided to us by USFWS. These studies have generated large amounts of relatively high quality data. The utility of the studies are reduced by the heavy reliance on burrow counts as a surrogate for SKR abundance, inadequate sampling across space and through time, and lack of standardization of protocols. We used the strongest and most complete datasets to address topics related to the estimation of population parameters at the scale of individual sites or trapping grids, and to the estimation of population trends across space and through time as it relates to understanding overall population trends within a reserve. We then made a series of recommendations regarding 1) a monitoring protocol for collecting data synchronously across all reserves for use in refining and updating a reserve-wide monitoring program and, 2) collecting additional information about biological processes/mechanisms responsible for SKR population dynamics.

Despite nearly 12 years of monitoring, there are still large gaps in our knowledge of the population dynamics of SKR. Given this lack of information, it is impossible to develop a cost effective monitoring program that will produce defensible estimates of population sizes across all reserves, or even within most individual reserves. Our strongest recommendation is to revitalize the monitoring program by adopting uniform standards for data collection, analysis, and reporting across the reserve network. The SKR reserve system must collect 3-5 years of data synchronously across all reserves to describe the pattern of population change across space and through time both within and across reserves. Once in place, these data will drive decision-making on a longer-term, reserve-wide monitoring program that will produce the highest quality data for the most cost-efficient effort.

## INTRODUCTION

This document describes exploratory analyses of SKR trapping data from a number of different studies. In our initial grant we described 5 tasks, which we address in this document:

*"We propose to review and analyze all SKR trapping and monitoring related data collected since 1990 and develop a model of the relationship between the monitoring protocol and the statistical power, or ability of the resultant data to detect population changes and long term trends."      Original proposal, March 2001*

- 1) Data summary/critique of effort and precision of estimated demographic variables.
- 2) Comparison of different sampling protocols (live trapping without marking individuals, live trapping with recaptures of uniquely marked individuals, burrow transects, and radial burrow counts).
- 3) Analysis of spatial and temporal trends in the data and how this relates back to questions of trend identification.
- 4) A power analysis that takes into account #1-#3, and critiques the current sampling methodology.
- 5) A description of a process through which additional data collected in future sampling can be used to refine and update sampling methods.

This document loosely follows the above tasks. We do not organize this report along the 5 tasks, but instead break the document into 3 main sections, this introduction, a data analysis section, and final conclusions. The data analysis section leads the reader through our analyses, results, and conclusions in a narrative fashion. We chose not to describe each set of analyses from the standard "introduction, methods, results, discussion" format typical in scientific papers. Instead, we introduce a topic, discuss our reasoning and methodologies then results and end with summary conclusions. The final section of the document summarizes the key findings of the work and outlines a strategy for future monitoring.

Our analyses include estimating SKR population parameters at three spatial scales. First, we address issues related to estimating parameters at specific trapping locations, or "points" in space. This will typically come from data collected across a single trapping grid or transect. Second, data from multiple points can be combined to estimate population parameters across a larger geographic area, or "site". Sites are typically a reserve or part of a reserve. Finally, we can combine data across points and sites to estimate parameters about the entire SKR population. We call this the "region."

## **SYNOPSIS AND CRITIQUE OF AVAILABLE DATA**

Monitoring data has been collected using different techniques since at least 1989. The first step of our analysis was to understand the spatial and temporal extent of sampling, the methods used, and the level of synchrony in both time and methodology across different sampling efforts (summarized in Table 1). We received 120 Excel files containing more than 250 worksheets of data. Individual worksheets contained information ranging from a single trapping effort with no SKR captures to 6000 rows of detailed capture information. Extensive data on burrows counts and trapping was available from several of these studies.

Data were collected using a number of protocols and at varying intensities. Researchers performed both burrow counts and live trapping, but with a variety of methods. Burrow counts were performed using either cells or 4-8 assessment lines, in which burrows were counted. Trapping was performed using grid sizes of 4x10, 7x7, 8x8, 9x9 for 3, 4, 5 and even 6 consecutive nights. Many PI's estimated abundance from burrows by using the formula "burrows\*0.243" reported in O'Farrell (1992), or by counting the number of captures, unique individuals or by adjusting their estimates by area trapped. The number of trapping sessions per year ranged from 1 to 12, while the number grids per location (i.e. reserve, or larger area of land) ranged from 3 to 47. Finally the duration of a study ranged from a single year to 5 years.

As a result of the incredible array of methodology, effort, timing and location, no unified, single meta-analysis can be performed on these data. Instead, we combined studies whenever possible to address specific questions, then used these results to draw conclusions regarding general protocols for SKR monitoring.



**Table 1.** General descriptions of the datasets received. *PI* refers to the principle investigator conducting research. Types *b*, *t*, and *v* refer to burrow counts, traps, and vegetation data, respectively.

Site	Years	Researcher	Plots / Sampling	Notes on Data	Type
Lake Mathews	1998-2000	Baxter	20 plots / burrow counts and trapping	<b>2 files:</b> 8 worksheets per file overview, sites and history, burrow counts, captures etc	<b>t, b</b>
Lake Mathews	1989-1994	O'Farrell	40 plots / trapping and burrow counts. General trapping in 89, translocation in 91, burrow counts 91-94.	<b>50 files:</b> 16 general, 24 translocation and 7 trap comparison files, 1 burrow counts	<b>t, b</b>
Lake Mathews (vegetation)	1991-1993	O'Farrell	40 plots / vegetation	<b>4 files:</b> 16 worksheets of vegetation data	<b>v</b>
March AFB	1996-2001	Sullivan	47 plots / burrow counts (4 plots with trapping)	<b>16 files:</b> 1 worksheet per file, burrow counts, trapping	<b>t, b</b>
Shipley / Skinner	1997-2000	Kelt	17 plots / trapping (4 plots account for 80% of captures)	<b>1 file:</b> trapping histories	<b>t</b>
Shipley / Skinner	1993-1996	O'Farrell	20 sites / trapping and burrows	<b>21 files:</b> 50+ sheets, variety of vegetation, treatment, burrow and trapping data.	<b>t, b</b>
Shipley / Skinner (vegetation)	1991-1996	O'Farrell	20 plots / vegetation	<b>1 file:</b> 23 worksheets, vegetation data, data analysis, etc.	<b>v</b>
Motte	1998, 2000	Unknown	3 plots / trapping (565 captures in 1998, only 26% were SKR)	<b>2 files:</b> Not all data collected in both years.	<b>t, b</b>
Lake Perris	1997-2000	O'Farrell	4 plots (+ 3 blocks experiment?) / trapping and burrow counts	<b>1 file:</b> 8 worksheets, burrow counts summary (4 plots) and burrow counts. <b>16 files:</b> blocks and treatments.	<b>t, b</b>
Eastside	1989-1990	O'Farrell	26 plots (89) and 10 plots (90) / trapping and burrow counts	<b>4 files:</b> 40 worksheets, 36 with trapping histories 2 with burrow lines and 2 with burrow cells	<b>t, b</b>
Estelle Mtn	1985	O'Farrell	6 plots / trapping an burrow counts	<b>1 file:</b> 6 sheets, 41 animals in total	<b>t, b</b>
Lake Henshaw	1983-1985	O'Farrell	4 plots + inventory plots / trapping	<b>5 files:</b> 17 sheets, capture histories (~5 animals per plot)	<b>t</b>
Riverside Nat'l Cemetery	1993	O'Farrell	2 sites – trap out and mitigation / trapping	<b>2 files:</b> 4 sheets, 26 animals in trap out and 4 in mitigation.	<b>t</b>
Skinner	1992	O'Farrell	10 sites – trap out and translocation/ trapping	<b>10 files:</b> 20 sheets, hundreds of animals in trap out, (~5 animals per trapping session in translocation)	<b>t</b>

## **ANALYSIS OF TRAPPING DATA**

This section of the report has four main parts. The first 3 sections all relate to the small-scale, or point estimation of SKR population parameters. First, we investigate burrow count methods and compare burrow counts to density estimates. We directly analyze the relationship set forth by O'Farrell (1992). Second, we include a power analysis on different trapping protocols and their ability to estimate density accurately. Third, we present a series of analyses comparing estimates of density from Program MARK (White and Burnham 1999) to "crude" estimates derived directly from trapping data. The fourth (and last) section of the analyses deals with the second spatial scale, "sites", and we present information regarding the extent and scale of changes in SKR densities across points within a reserve. We use the broad-scale analysis to make recommendations regarding sampling strategies for collecting data across the reserve system.

### **1. Point Estimation of SKR Density**

The success of the monitoring program depends critically on the ability to monitor changes in population density through time and across space. This, in turn, hinges on the ability to accurately estimate the density of SKR at each point being monitored. Two techniques have been used to estimate density, live trapping and burrow counts. To date, each of these techniques has been implemented in many different ways. A trade off exists between the two methods. Trapping is time consuming and expensive because traps need to be visited frequently and individual animals must be marked before being released. In contrast, burrow counts are fast and efficient. As a result, they can be conducted at many sites for very little cost relative to trapping. However, the utility of burrow counts depends on the strength and stability of the relationship between burrow counts and SKR density.

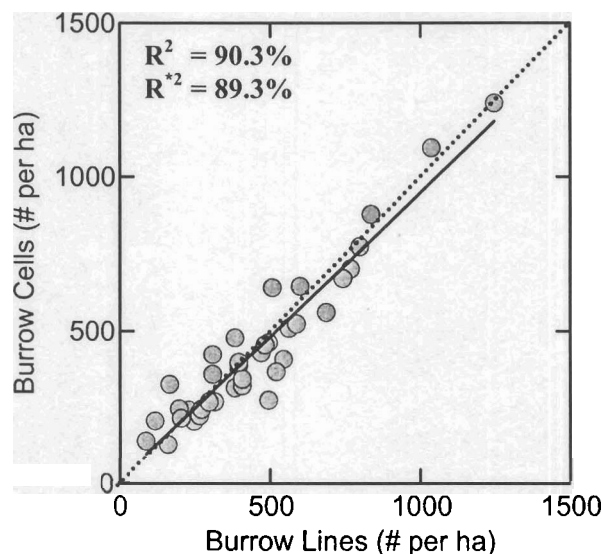
#### A. Analysis of Burrow Counting Methods

Two methodologies for burrow counts have been used extensively following the pioneering work of O'Farrell. One method (Lines) involves counting burrows in several belt transect. The implementation of transect lines varies but commonly 4 transects each 135m long

by 3m wide were used. The second method involves counting burrows in large circles (Cells) laid out in a grid across the landscape. Cells were typically 7.5m radius circles.

Burrow cells (7.5 m radius around each trap) and Burrow lines (3m swath on each of 4 transects) were counted for 35 sites at Eastside Reservoir in 1989 and 1990. Burrow cells included much larger areas searched (.707 ha versus .162 ha) than lines and we presume required much more effort. For comparison, we normalized both types of burrow counts to the density of burrows in a hectare (Burrows  $\text{Ha}^{-1}$ ). After normalizing, we expected these two methods to give identical results. If so, a regression of burrow density for Lines versus Cells should not deviate from a 1:1 relationship (i.e. intercept = 0, slope = 1; see Appendix 1 for details).

We start by regressing the density of burrows from large cells (Y) against the coarser measure using transect lines (X) (Figure 1). The unconstrained regression line has an intercept of 15.7 ( $\pm 27.0$ ) and a slope of 0.934 ( $\pm .052$ ). The slope is not significantly different from 1.0 ( $t = -1.27$ ,  $p = .212$ ) and the intercept is not significantly different from 0.0 ( $t = 0.58$ ,  $p = .565$ ). Thus, burrow lines are an unbiased estimator of burrow density. The unconstrained regression has an  $R^2$  of 90.3%. Since the unconstrained regression line is not significantly different from the 1:1 line, the amount of variation explained by the 1:1 line ( $R^{*2}$  – see appendix 1) is 89.3%, nearly identical to the standard measure,  $R^2$ .



**Figure 1:** Relationship between Burrow Counts and Burrow Lines based on 37 sites in 1989 and 1990.  $R^2$  represents the variation explained by best-fit line (solid line, intercept and slope unconstrained).  $R^{*2}$  represents the amount of variation explained by a 1:1 relationship (dashed line, intercept = 0, slope = 1).

**Table 2.** *Results of the unconstrained regression of burrow cells (Y) against burrow lines (X) from 37 sites at Eastside Reservoir in 1989 and 1990.*

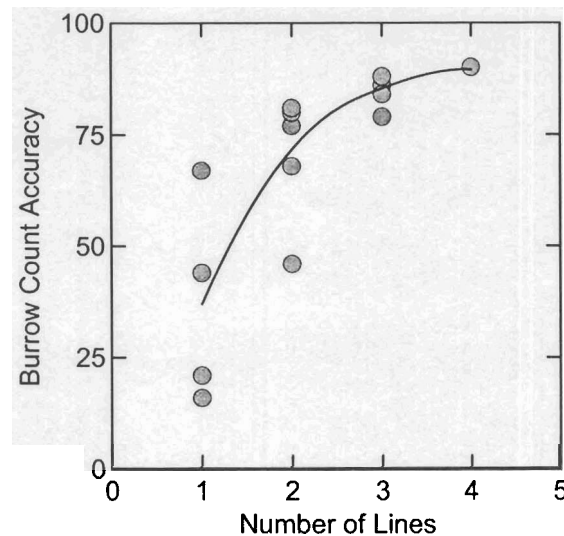
Parameter	Estimate	Std Err	t	p-value
Intercept	15.7	27.0	0.581	0.565
Slope	.934	.052	-1.27	0.212

This example provides compelling evidence for the utility of this statistical analysis. The strong performance of the constrained regression ( $R^2=89\%$ ) shows that the coarser sampling associated with burrow lines (.162 ha) is nearly equivalent to the much more extensive burrow cells (.707 ha). In other words, burrow lines, a sampling method that covers less than 25% of the area of the burrow cells, explains 89% of the variation of the more exhaustive counting technique.

#### B. Power Analysis of Burrow Counting Methods

The minimum sampling effort needed to produce reliable burrow estimates can be further refined. In O'Farrell's data from Lake Matthews, the burrow cells covered an area approximately 4 times greater than the transect lines (.707 ha versus .162 ha). The number of transect lines needed to accurately count burrows was estimated by calculating the  $R^2$  for all subsets of the 4 transects.

Because the study used 4 transect lines, we can form 14 subsets, 4 single transects, 6 transect pairs, and 4 transect triplets. The accuracy of the burrow counts (compared to the much larger cell counts) can be plotted as a function of the number of transects. We estimated accuracy using the coefficient of regression for the constrained regression ( $R^2$ ) as described above. Accuracy of burrow counts approached 90% with 3 or 4 transects (Figure 2). This suggests rapid burrow count assessments using 3 transects at a minimum will likely be adequate. As importantly, it demonstrates that estimates of burrow density from only 1 or 2 transect lines are unreliable.



**Figure 2:** Accuracy of transect lines as a function of effort. Accuracy of estimates was based on  $R^2$ . Sampling as little as 3 transect lines (.12 ha) was adequate to estimate burrow density.

### C. Relationship between Burrow Counts and Density Estimates

The utility of burrow counts as a surrogate for SKR density depends on the strength and stability of the relationship. If the relationship is weak, it will be difficult to use burrow counts in place of the more expensive trapping grids. If the relationship between burrow counts and density is strong, but varies from place-to-place or year-to-year, the relationship will need to be re-calibrated frequently. Burrow counts will provide an inexpensive surrogate for SKR density only if the relationship is strong and stable from place-to-place and year-to-year.

The relationship developed in O'Farrell (1992) is based on data from 30 sites sampled between 1989 and 1991 throughout Riverside County ( $n = 44$ ). Some of the calculations presented in O'Farrell's 1992 paper contain errors and it is important to review them before continuing. O'Farrell calculated the density of SKR from the number of unique animals captured divided by area sampled (A). The area sampled was based on his earlier work (O'Farrell et al 1977). For a 4 by 10 grid with 15 meter spacing, O'Farrell calculates that the area sampled was 1.03 ha. Using the same assumptions about the movement of SKR (7.5 meters), we calculated that the area sampled was only 0.895 ha (See Appendix 2 for details). As a result of this change,

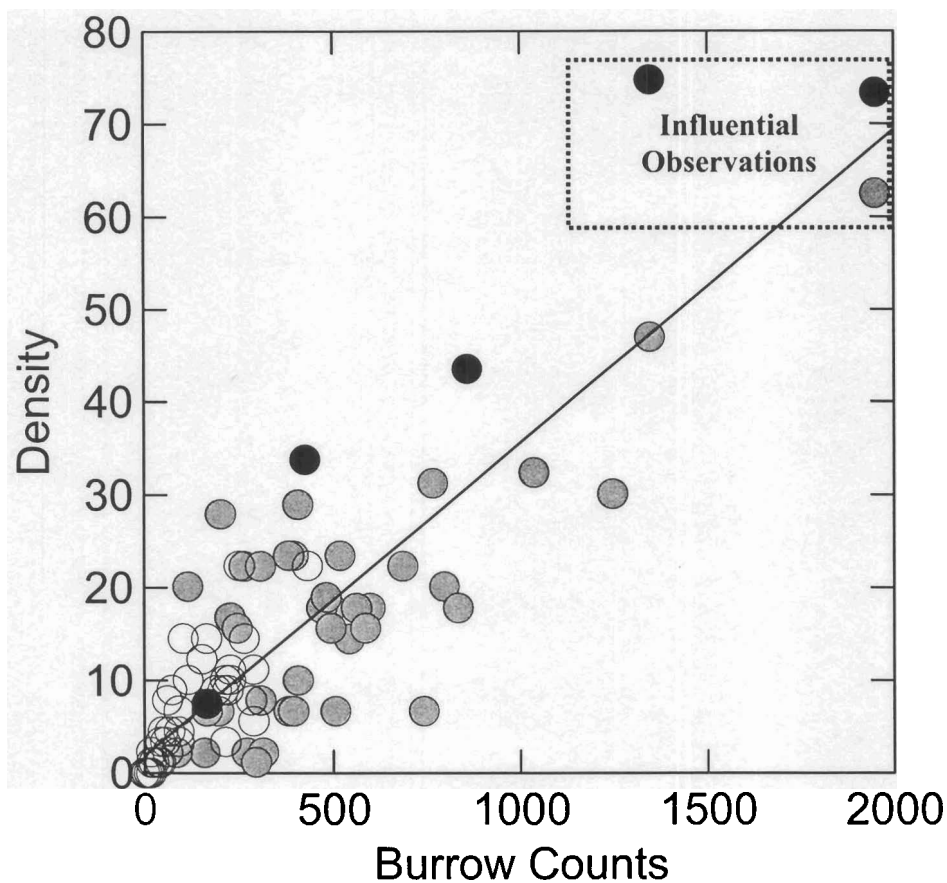
density of trapped animals are actually 15% higher than used in O'Farrell's calculations. This correction leads to a 15% increase in the slope reported by O'Farrell (.279 instead on the reported .243).

A second concern about the O'Farrell paper is the reporting of the strength of the relationship from regressions constrained to go through the origin. This inflates the apparent strength of the relationship. Although O'Farrell mentions this issue in his paper, he does not offer a more meaningful statistic. In the paper, the  $R^2$  is reported as 95.4%. A more reasonable  $R^2$  (Sums-of-Squares explained by the line divided by the Total Sums-of-Squares from the standard regression) is 79.8%.

The third and most important concern stems from the selection of only 5 points of the 42 to build the final model. O'Farrell justified the choice based on his preference for mesh traps. However, the reduction in sample size from 42 to 5 is dramatic and potentially dangerous. The density of SKR burrows and animals was unusually high at several of those 5 sites. Burrow density averaged 949 per hectare ( $\pm 717$ ) at those 5 sites compared with 457 ( $\pm 255$ ) at the other 37 sites. It appears that these sites may have been unusual in ways that were not related to the use of Mesh traps.

We re-analyzed the relationship between using data from 94 sites across 12 years of sampling. In this work, we develop the relationship between SKR densities and burrow counts from 94 pairs of observations, 44 from the original work by O'Farrell, 12 from March AFB, and 38 from Lake Mathews (See Figure 3). These grids include some that were manipulated by experimental treatments. These grids were included in order to cover the broadest range of habitat conditions.

This larger analysis confirms the positive relationship between burrow counts and SKR density. The relationship estimated from the larger dataset is more variable than previously reported. The strength of the relationship (measured as  $R^2$ ) between burrow counts and density across these 94 grids is approximately 77% (Figure 3). The regression is strongly influenced by four influential observations (Sites 26 and 31 with both Sherman and Mesh traps in 1991). Removal of these points results in a marked decrease in the strength of the relationship ( $R^2 = 57\%$ ).



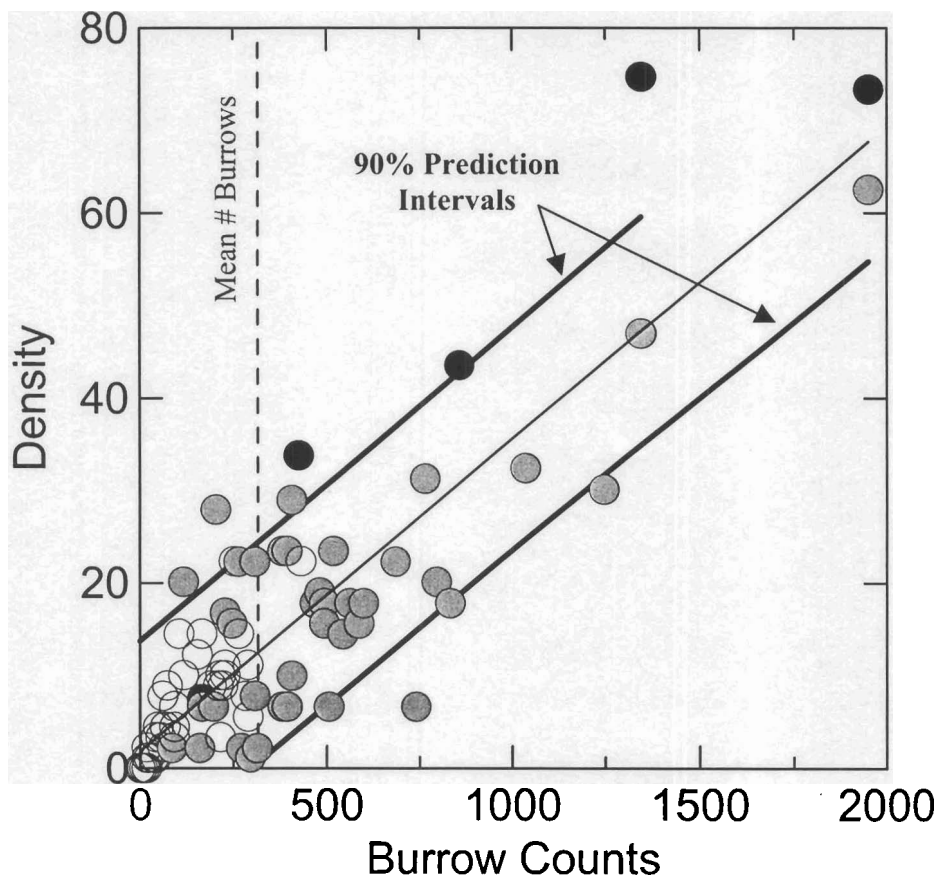
**Figure 3.** Relationship between burrow count ( $X$ , burrows per hectare) and density ( $Y$ , unique individuals per hectare). Three subsets are identified by color; data from mesh traps in O'Farrell 1992 (black), Sherman traps reported in O'Farrell 1992 (gray), and data collected after 1992 (white). The best-fit regression line is plotted for all 94 points.

**Table 3.** Summary of unconstrained regression between burrow count ( $X$ , burrows per hectare) and density ( $Y$ , unique individuals per hectare).

$$Y = 1.83 + .034 * X$$

Parameter	Estimate	Std Err	t	p-value
Intercept	1.83	0.96	1.90	0.060
Slope	.034	.002	17.50	<0.001

On average, there is a clear positive relationship between burrow counts and density. The relationship, however, is not adequate to make accurate predictions at a single grid. When the best-fit relationship is used to estimate density at each grid, the standard errors around the estimate are quite large (Figure 4). At the average burrow count of 317 burrows per hectare, the relationship predicts a density of 12.5 ( $\pm 7.2$ ) animals per hectare. Due to the variability in the relationship, the 90% prediction interval ranges from 0.6 to 23.4 animals per hectare. This prediction interval is unacceptably wide.



**Figure 4.** Prediction of Density (Individuals per ha) from Burrow Counts (Burrows per ha) at 94 samples spanning three studies between 1989-2000. 90% prediction interval is graphed for the density of SKR (thick lines). The dashed line represents the mean number of burrows counted (317 burrows per hectare). At this density, the predicted number of SKR is 12.5 animals per hectare with a standard error of  $\pm 7.2$ .



There is also some evidence that the relationship varies among the three major datasets. The slope of the regression line differs significantly among the three datasets ( $F_{2,83} = 4.22$ ,  $p = .018$ ). The relationship is also influenced by the unusually dense burrow counts encountered by O'Farrell. Excluding the four influential observations from the O'Farrell study (the four points in Figure 4 with the highest burrow counts), the slope of the relationship is estimated at 0.021 ( $\pm .005$ ,  $N = 40$ ). The slope estimated from the Baxter study is .045 ( $\pm .004$ ,  $N = 38$ ) an increase of more than 100% (Table 4)

**Table 4.** Results from separate regressions of density (Y; animals per ha) against burrow lines (X; burrows per ha) from the three major studies. The O'Farrell study is analyzed two ways to demonstrate the influence of four unusual observations.

<b>Study</b>	<b>N</b>	<b>Slope (SE)</b>	<b>R<sup>2</sup></b>
O'Farrell (Eastside; Trimmed)	40	.021 (.005)	29%
O'Farrell (Eastside; All Sites)	44	.033 (.003)	68%
Sullivan (March)	12	.038 (.016)	36%
Baxter (Lake Mathews)	38	.045 (.004)	79%

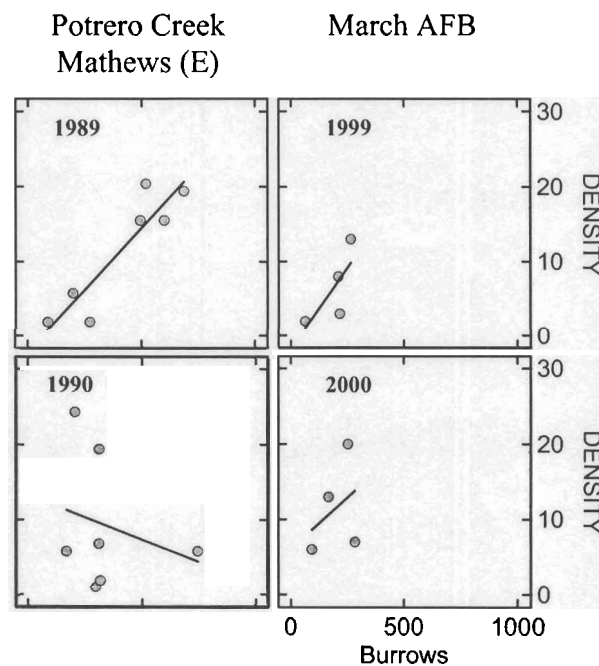
Striking variability was also observed at the level of individual grids. In his seminal paper, O'Farrell sampled several sites in consecutive years using the same burrow counting and trapping protocols. This includes 4 grids at Potrero Creek and 3 grids at Matthews East in 1989 and 1990. In addition, 4 grids at March AFB were sampled annually from 1998 to 2000. These sites offer the most compelling evidence of site-to-site and year-to-year variability in the relationship between burrow counts and density (Figure 5).

#### D Conclusions Regarding Burrow Counts and Density.

Although a clear relationship between burrow counts and density exists, it is much more variable than previously reported. Three lines of evidence suggest there are significant statistical problems with using burrow counts as an estimator of density at a either a specific grid, or a specific site (reserve), during a specific sampling session. First, the standard errors for predicted density using the regression from 80 grids were large, indicating a high degree of imprecision when converting from burrow counts to density. Second, the relationship between burrow

counts and density varied across sites and through time. Third, at a specific site, the relationship between burrow counts and density can change from one year to the next. As a result, it is not obvious how useful burrow counts are for predicting density at a single site during a specific sampling session.

We recommend against the use of burrow counts for estimating density at either a specific grid or a specific site at a specific moment in time. However, there was a strong overall relationship between burrow counts and density when combining data across many sites and many years. Thus, useful information can be extracted from burrow counts, but perhaps at larger spatial and temporal scales than specific grids/reserves (“sites”) at just one trapping session. We make recommendations to further investigate the use of burrow counts for larger scale assessment in the conclusions section.



**Figure 5:** Year-to-Year variability in the Burrow Count – Density Relationship. In each panel, the best-fit regression is plotted (dark line). Notice the relationship is strong in 1989 at Potrero Creek and Mathews E but reversed in 1990. In contrast, the relationship is fairly stable in 1999 and 2000 at March AFB.

## 2. Power Analysis of Trapping Methods

Despite the cost, live trapping of animals may represent the best method to monitor SKR density. Over the past 14 years, trapping has been conducted in many different ways. Grids have ranged from 36 traps (6x6) to 81 traps (9x9) and to rectangular 40 trap (4 x 10) grids. Trapping has been conducted for 3 to 6 consecutive nights. Trapping frequency has ranged from monthly to a single session each year. If trapping is to be possible across large areas, identifying the minimum level of effort necessary to accurately track changes in SKR density is imperative. One method to address this issue is to look at the accuracy of point estimates at a site based on the size of the grid and the number of nights that the grid was run.

Two studies allowed for a preliminary examination of this question. O'Farrell's work at Lake Mathews included 6 months of trapping on a 9x9 grid for 3 nights. Kelt's work used a 7x7 grid for 3 nights. In both cases, we can resample the data with reduced effort and compare the resulting estimate of N to that from the more exhaustive sampling effort. We simulated reduced effort by reducing the size of the grid and/or reducing the number of consecutive nights of trapping. Grid size was reduced by taking those captures occurring in traps located in the outside "rows and columns" out of the raw data.

*Table 5. Trap effort by full (shaded) and reduced efforts of sampling.*

Cumulative Effort (trap nights)	O'Farrell		Kelt	
	Grid Size	Days	Grid Size	Days
243	9 x 9	3	-	-
162	9 x 9	2	-	-
147	7 x 7	3	7 x 7	3
98	7 x 7	2	7 x 7	2
75	5 x 5	3	5 x 5	3
50	5 x 5	2	5 x 5	2

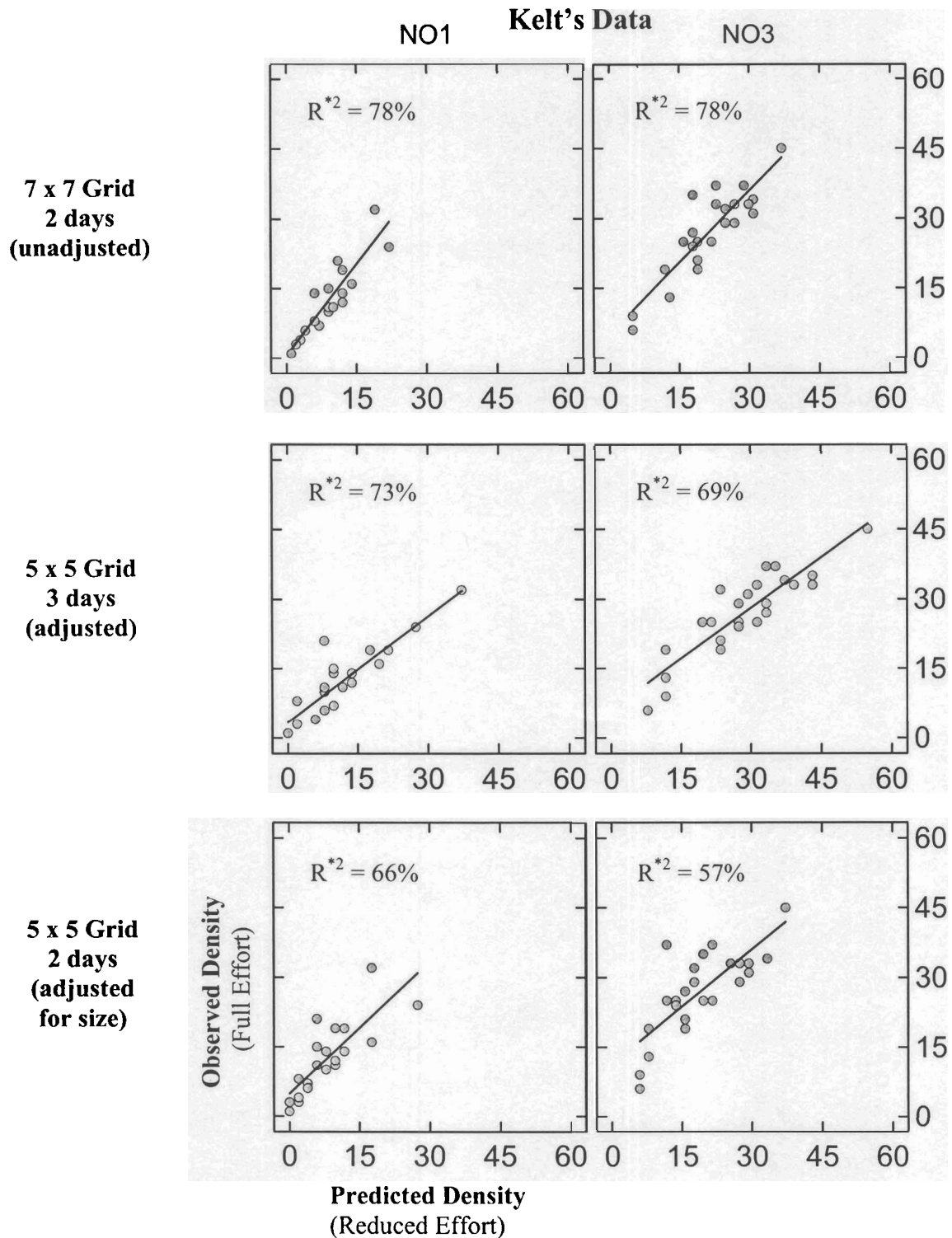
Statistical analysis of the results is based on testing for deviations from the 1:1 regression line and estimating the accuracy of the relationship ( $R^2$ ). Density efforts using smaller grids are adjusted for the area sampled using the formula:

$$N_{adj} = N_{reduced} * \frac{\#Traps_{Full}}{\#Traps_{reduced}}$$

For example, if we observed 10 individuals on a 5x5 grid (25 traps), we would expect that there would be  $10 * (81/25) = 32.4$  individuals on the 9x9 grid. This adjustment simply accounts for the area sampled. No adjustment was made for the number of nights sampled because the relationship between animals caught and the number of nights sampled is not linear.

Reduced trapping effort often led to lower estimates of abundance (numbers of unique individuals/ha) even after differences in area were corrected (Figures 10 and 11). As a result, the reduced trapping effort estimates were conservative. The analyses indicated reducing grid size had a greater impact on accuracy than did a single day reduction in trapping effort.

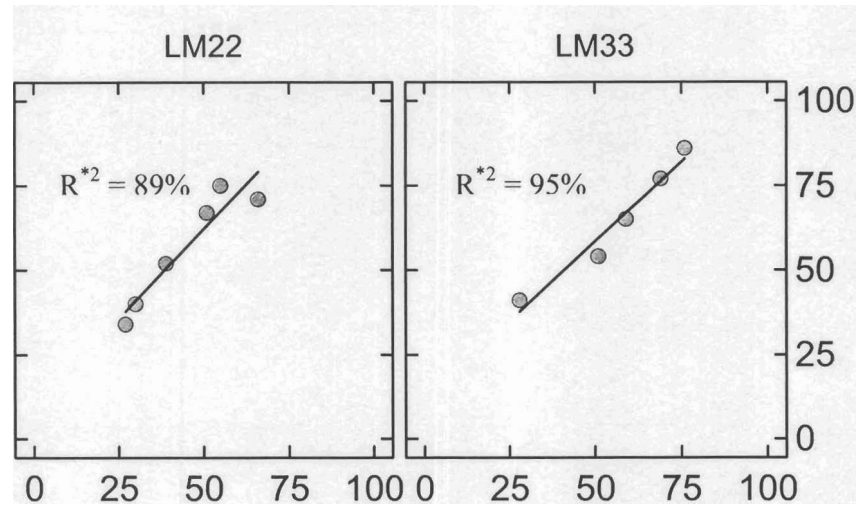
The resampling experiment suggests density estimates converge on the correct values with modest effort (Figure 12). Grid sizes of 7x7 performed nearly as well as the 9x9 grid ( $R^2$  averages 88.5%). In addition, two nights of sampling with 7x7 grids also performed well ( $R^2$  averages 83.3%). These results should not be used to justify a sampling effort of 7x7 grids and 2 nights. Although it appears a trapping protocol with reduced effort relative to what has been done yields similar results, two issues must be considered when recommending a sampling effort. First, we chose grids from O'Farrell and Kelt with reasonably high numbers of individuals captured. Thus, these plots represent relatively high population densities. As such, under average or low abundances, 7 x 7 grids may be too small. Furthermore, these relationships may not be similar when fewer animals are captured. Second, as we mention below, a larger grid size may be necessary to capture an adequate number of individuals for proper functioning of program MARK, or for accurate estimation of specific demographic parameters. Thus, sampling effort should be determined by the need for adequate sample sizes. Given the performance of trapping grids we observed, we recommend an 8 x 8 or 9 x 9 grid with 2 days of effort as a minimum, realizing that 2 days of effort precludes the use of certain capture-recapture models. Our data indicate in most cases, trapping an additional night will not result in a large proportion of new, unmarked individuals, in part because SKR have relatively high capture probabilities.



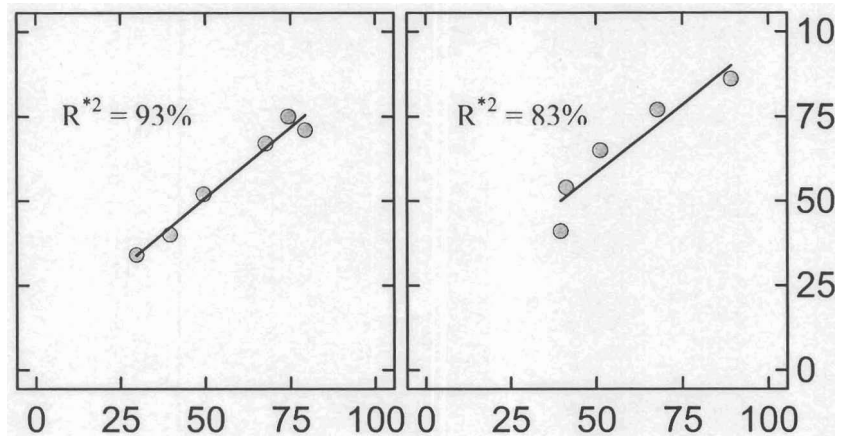
**Figure 6.** Relationship between numbers of individuals captured with different levels of effort (x-axis) and full sampling (Y-axis) for data collected by Kelt at two grids with relatively high abundance. We adjusted for area differences when we reduced grid size by estimating density (individuals/ha).

### O'Farrell Data

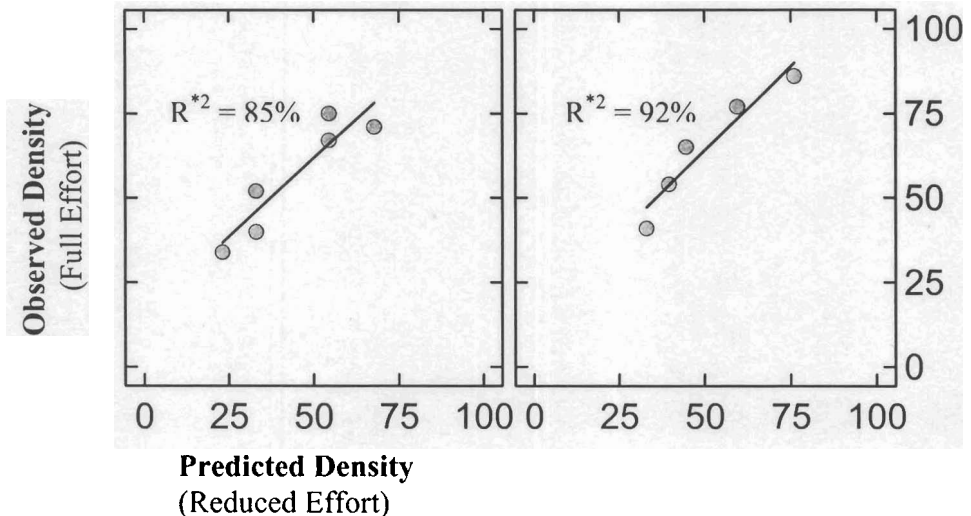
9 x 9 Grid  
2 days  
(unadjusted)



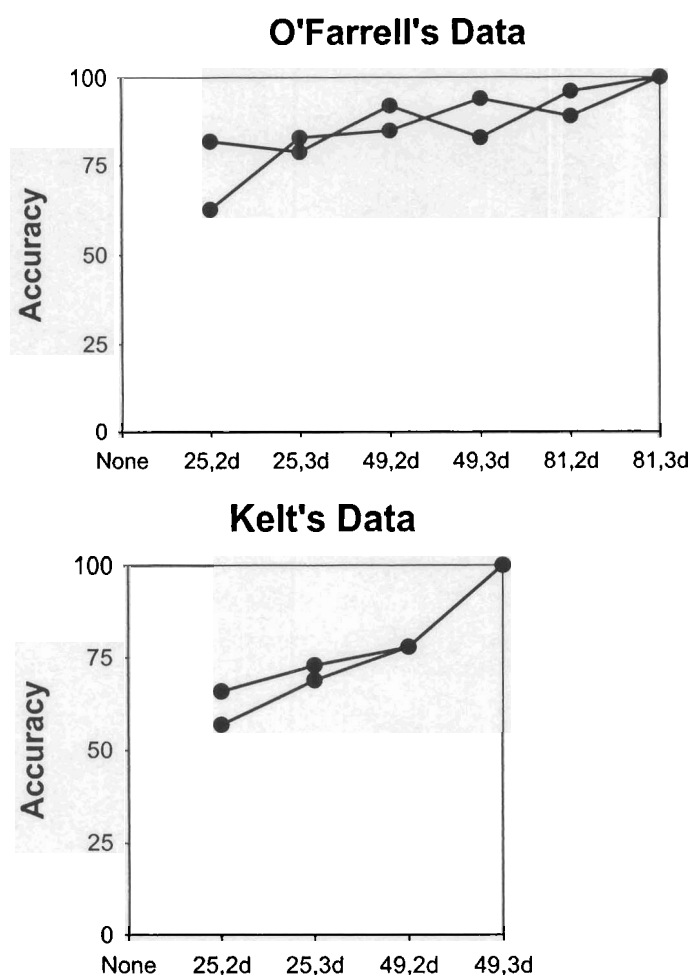
7 x 7 Grid  
3 days  
(adjusted)



7 x 7 Grid  
2 days  
(adjusted for size)



**Figure 7.** Relationship between numbers of individuals captured with different levels of effort (x-axis) and full sampling (Y-axis) for data collected by O'Farrell at two grids with relatively high abundance. We adjusted for area differences when we reduced grid size by estimating density (individuals/ha).



**Figure 8.** Accuracy as a function of sampling effort from representative trapping grids used by Kelt and O'Farrell. Accuracy is estimated as the  $R^2$  for a 1:1 regression between the density estimate from the full effort and that from a reduced level of effort. The gray line and arrow are estimates and interpolation of accuracy vs. effort.

### 3. Point estimation using Program MARK

Successful SKR management will require detailed demographic information about some populations. For example, a Population Viability Analysis was performed on the species during the original design of the reserve system, which relied heavily on estimates of survival and reproduction. It is likely an updated PVA, explicitly incorporating relationships between habitat

type and demographic rates, will greatly enhance the ability to successfully manage SKR populations. The census techniques we've analyzed thus far may be useful for detecting trends in population sizes through time and space, but how well they function for estimating demographic processes such as survival, or allow more accurate estimation of density, required different analyses.

Program MARK represents the "state-of-the art" statistical methods for estimating demographic rates from capture-recapture data. Recently, a paper by McKelvey and Pearson (2001) investigated the utility of using complex estimation procedures with capture-recapture data when sample sizes were small (<50 animals). They showed these complex methods worked poorly when sample sizes were small. Given that most grids produce small numbers of individuals captured, we evaluated, on a preliminary level, the utility of using MARK on the existing SKR data. We analyzed data from 4 grids with relatively high numbers of individuals captured. Thus, these results should be viewed as "best-case scenarios" for a given grid size as most trapping sessions will yield fewer captures. Our approach was to run the data through MARK and check for evidence indicating poor program or model performance such as 1) the inability of MARK to converge on a model or problems with convergence, and 2) demographic estimates with very large or very small standard errors. Both conditions result from the failure of Program Mark to produce reliable estimates.

#### A. Performance of MARK.

In general, sample sizes were too small to allow reliable estimation using MARK. In most cases, complex models that included sex or time dependent changes in survival simply would not converge and even the simplest models occasionally resulted in convergence problems. Given these issues, we investigated how similar estimates from the poorly performing models in MARK were to "crude" estimates of population size and capture probabilities derived directly from the data. We chose a subset of plots with different captures/individual/day and ran them in program MARK to estimate capture probabilities and population size. We then compared the results from MARK with the crude estimates. In these cases, we used a closed capture "null" model in MARK where capture and recapture rates did not vary in any manner. Though simple, this model allowed numerical convergence across all the datasets we ran. In all

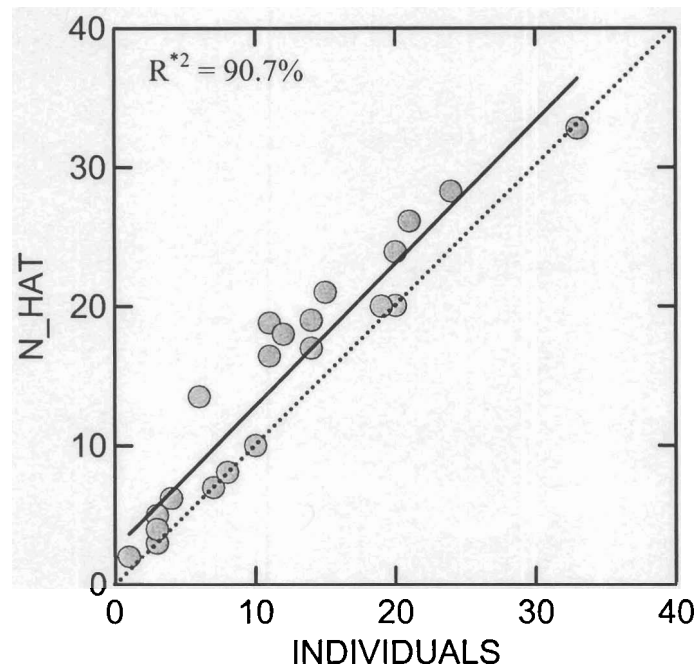


cases, the estimates of population size from MARK were extremely similar to  $M_{(t+1)}$  (Table 6, Figure 9). In many cases, the standard errors were near zero, indicating a lack of fit. In other cases, the prediction intervals were as wide (imprecise) as the confidence intervals from the estimate of abundance based on the burrow counts (Figure 4). Thus, with these small sample sizes, some estimates of  $N$  from MARK were as unreliable as burrow counts or more so.

**Table 6.** Number of individuals captured  $M_{(t+1)}$ , estimate of population size from MARK as well as the associated error terms for plot NO1 of Kelt. Shaded areas show some examples of poor estimates with very small or very large standard errors. Model form was  $\{p=c, p(.,T), s(.,T), N(T), g'(.), g''(.)\}$ .

Session	Numbers captured	Estimate of N from Mark	Standard Error	Lower bound	Upper bound	Prediction Interval Width
1	20	23.907	1.425	23.102	31.072	7.970
2	24	28.267	2.105	26.483	36.631	10.148
3	14	16.988	2.350	15.318	27.425	12.107
4	10	10.000	0.000	10.000	10.000	0.000
5	11	16.419	1.172	16.023	23.551	7.528
6	7	7.000	0.000	7.000	7.000	0.000
7	3	5.000	0.000	5.000	5.000	0.000
8	4	6.222	2.310	5.109	18.695	13.586
9	1	2.000	0.000	2.000	2.000	0.000
10	3	3.000	0.000	3.000	3.000	0.000
11	3	4.000	0.000	4.000	4.000	0.000
12	8	8.100	1.052	8.001	15.053	7.052
13	6	13.505	4.756	9.826	33.562	23.736
14	21	26.115	4.456	21.702	41.974	20.273
15	11	18.803	2.723	16.567	29.857	13.290
16	15	21.001	2.786	18.640	32.072	13.432
17	20	20.044	0.906	20.000	25.458	5.457
18	12	18.000	0.000	18.000	18.000	0.000
19	14	19.000	0.001	19.000	19.001	0.001
20	19	20.000	0.000	20.000	20.000	0.000
21	33	32.797	1.357	32.081	39.844	7.763

As depicted in Figure 9, the number of individuals captured in a trapping session was highly correlated with the estimate of N produced by MARK. The regressions were extremely strong as  $M_{(t+1)}$  was tightly related to the estimates of N from Mark. We found similar results using the other plots trapped by O'Farrell [Plot LM22,  $R^2 = 0.98$ ; Plot LM33,  $R^2 = 0.89$ ].



**Figure 9.** Estimates of population size from MARK ("N\_HAT", y-axis) against the number of individuals captured during a trapping session for Kelt's Plot NO1. Other datasets produced similar results.

Across all the datasets we ran, the estimates of N and the number of unique individuals captured were similar, despite differences in capture probabilities. Two explanations exist for this pattern. First, all individuals in a population were captured so estimates of N must be equal to the number of individuals captured. We do not feel this is an adequate explanation. For example, O'Farrell's grid PL24-89, capture probabilities were 0.25. Thus the probability of not capturing an animal during a 3 night trapping session was  $(1 - 0.25)^3 = 0.42$ . As a result approximately 42% of the population should not be captured so estimates of population size from MARK should be substantially higher than the number of individuals captured. However the estimates of N from Mark were nearly identical to  $M_{(t+1)}$  on PL24-89.

A more reasonable interpretation is that even on these grids of relatively high densities, the number of individuals sampled was so small at each trapping session MARK could not adequately estimate a model. Within most 3-5 night sampling sessions, between 0 and 35 individuals were captured (e.g. Table 6). This represents extremely small sample sizes of only 0 to 12 animals captured per night. Thus, during many sampling sessions, MARK simply could not estimate  $N$  accurately.

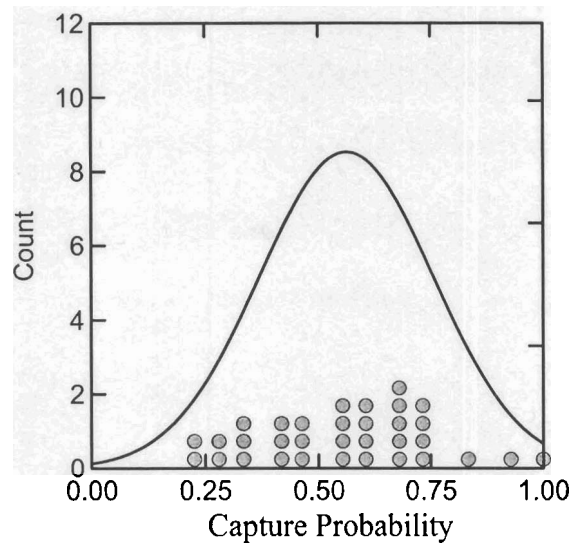
In situations like this, McKelvey and Pearson (2001) indicate the number of unique individuals captured ( $M_{(t+1)}$ ) may be a reasonable index of population size and might be a useful tool for monitoring and management. A key feature with  $M_{(t+1)}$  is that it can be highly sensitive to changes in capture probability. For example, suppose the population declines from 100 to 50 individuals. During the first trapping session, capture probabilities are 0.25 so  $100 \times 0.25 = 25$  individuals are captured. However, during the second trapping session, capture probabilities double to 0.50. In this case,  $50 \times 0.5 = 25$  individuals are also captured and it appears the population has remained stable. Thus, indices, such as  $M_{(t+1)}$  must be used with caution and checked to see how sensitive they are to changes in capture probability.

#### B. Number of Unique Individuals Captured and Capture Probability.

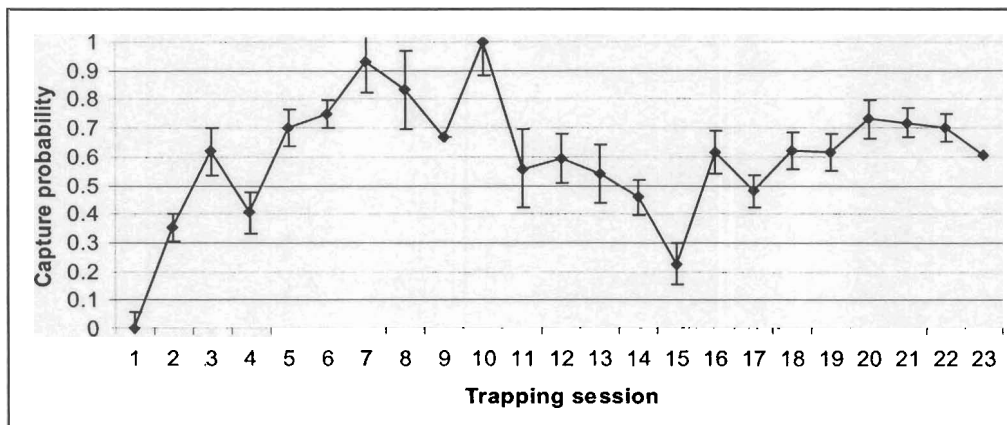
We investigated the potential for using  $M_{(t+1)}$  as an index of population size in two ways. First, we determined if capture probabilities varied. Second, we then investigated the relationship between capture probabilities and  $M_{(t+1)}$ . If capture probabilities vary consistently with density,  $M_{(t+1)}$  will be systematically biased.

Capture probabilities varied from site to site and through time at a specific site. We combined estimates of capture probability produced by MARK for Kelt's data and O'Farrell's Lake Matthews data to better visualize the variability in capture probabilities across different trapping sessions and sites (Figure 10). The average capture probability was 60% but it varied widely ranging from 0.22 to 1.0. In addition, we also found large amounts of variation in the probability of capture within sites through time (Figure 11).

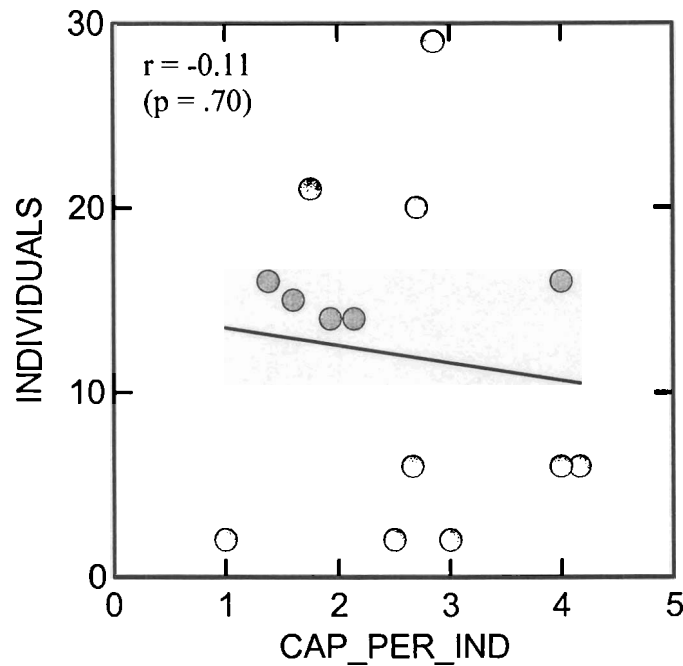
Two lines of evidence suggest that  $M_{(t+1)}$  was relatively insensitive to changes in capture probabilities. First, we found no relationship between crude estimates of capture probabilities (Number of captures per individual) and population size for the data in (Figure 12). This indicates that for these plots, differences in capture probabilities were not the cause of the observed differences in abundances between sites.



**Figure 10.** A histogram and fitted normal curve for capture probabilities taken across data collected by Kelt and O'Farrell.

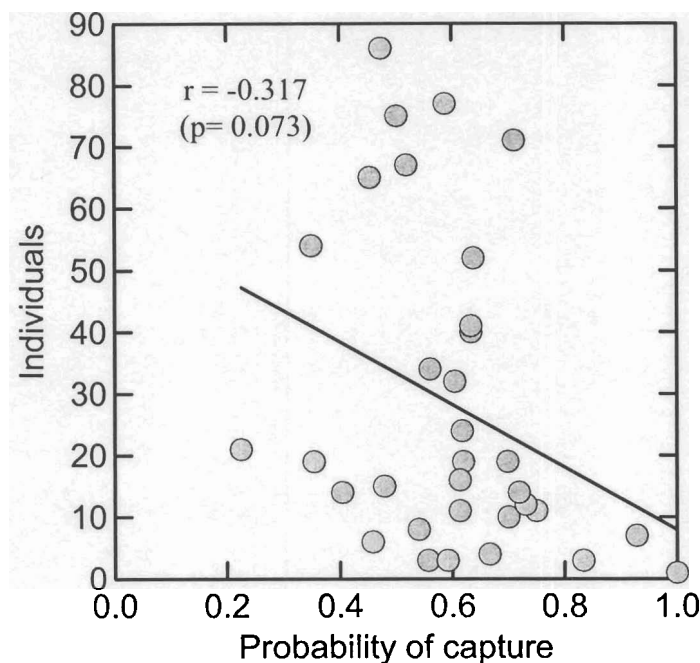


**Figure 11.** Capture probability by trapping session for Kelt's grid, NO1. Error bars represent standard errors estimated from MARK. The pattern held for all grids.



**Figure 12.** The number of individuals captured in a trapping session versus the captures per individual per day (CAP\_PER\_IND) during that trapping session.

Second, to confirm the lack of relationship between capture probabilities and numbers of animals caught, we combined the capture probabilities estimated from MARK and the number of individuals captured for Kelt's plots NO1, LM22, and LM33. In this case, there was a negative and slightly statistically significant relationship between capture probability and the number of individuals captured (Figure 13). In a few cases, capture probabilities were high when the number of individuals was low, driving the seemingly negative relationship. These were situations where only a few individuals were captured, but repeatedly within a session so that MARK estimated a high, but likely very inaccurate, capture probability. Another striking pattern in the data were large amounts of variation in abundances for capture probabilities between 0.40 and 0.65. Again, this indicates variation in abundance is not driven variation in capture probability.



**Figure 13.** *The relationship between estimates of probability of capture from MARK and the number of individuals captured during a session for a combined series of data*

#### C. Estimation in MARK and Sampling Effort.

We investigated the relationship between sampling effort and estimates of survival rates from MARK. Our analyses of population size estimation indicated MARK performed poorly with typical number of captures. However, in our experience, MARK occasionally produces reasonable estimates of survival despite producing poor estimates of N.

We took data from 4 plots and sub-sampled these to simulate reduced trapping efforts. We used data from Plots NO1 and NO3 (D. Kelt's data) and LM22 and LM33 (O'Farrell's data) and compared the estimates of survival between the full and the simulated reduced trapping effort. We used the same methods for simulating reduced trapping effort as described above. We assumed a 7x7 (O'Farrell only) or 5 x 5 trapping grid instead of the full grids by eliminating all captures in the data from the outside "ring" of traps in the original data. To simulate less trapping per session, we simply made all trapping sessions 2 days long. We ran the following

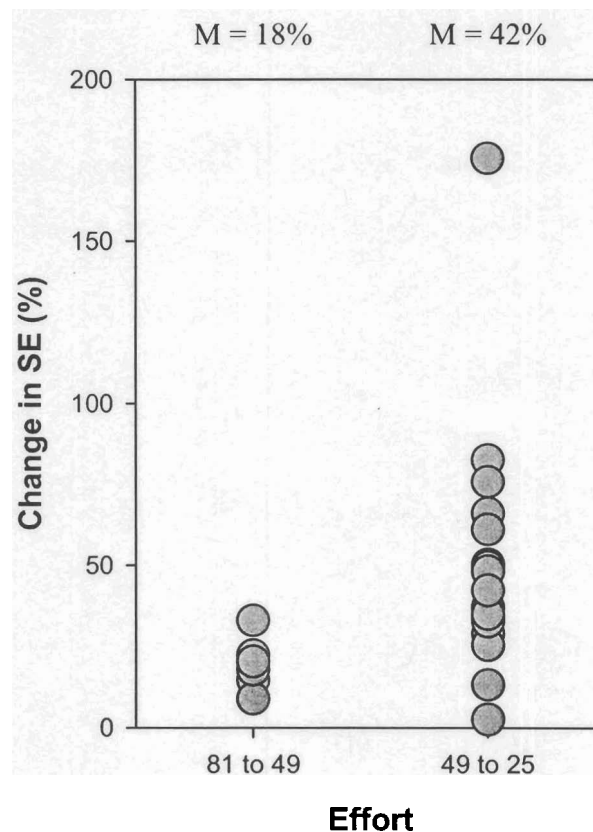
combinations of data through an identical model in MARK and compared both the estimates of survival and the standard errors around these estimates:

- 1) Full data: All traps and all trapping days
- 2) Reduced grid: 7 x 7 or 5 x 5 grid, with full trapping sessions
- 3) Reduced trapping session: All traps, but only 2 days per session
- 4) Reduced grid and trapping session: both reduced grid and 2 days per session.

Because the number of individuals captured was very small during some trapping sessions, complex models in MARK resulted in convergence problems and were not feasible. As such, we reduced model complexity when we ran MARK. We used a standard model that maintained some level of complexity, but was simple enough to converge in all datasets. In all cases, we used the robust design model where we set capture rates ( $p$ ) equal to recapture rates ( $c$ ). Furthermore, we allowed capture rates to vary from trapping session to trapping session, but not between mornings within a trapping session. Rates of survival ( $s$ ) and population size ( $N$ ) varied between trapping sessions. Finally estimates of temporary emigration ( $g'$ ) and immigration ( $g''$ ) off the grid were constant. In MARK nomenclature, the model form was  $\{p=c, p(.,T), N(T), s(T), g'(.), g''(.)\}$ . In one case, Kelt's data, plot NO3, this model would not converge so we used a simplified model  $\{p=c, p(.,.), N(T), s(T), g'(.), g''(.)\}$  where capture probability was fixed across all sessions and all mornings. Finally, to further reduce model complexity, we collapsed morning and evening trap checks into one trap check. Thus, if an animal was captured in both the morning and evening check, it was counted as being captured on that day only.

In general, the utility of MARK for estimating survival was limited. In some cases, MARK produced good estimates of survival with relatively small standard errors, while at other times standard errors were broad. Furthermore in some cases, such as plot NO3, numerical convergence in MARK was suspect for even the simplest model and many estimates of survival had extremely small standard errors, which we interpret as evidence of poor model behavior. Ignoring these drawbacks, we evaluated the performance of smaller grids. Estimates from 7x7 grids performed somewhat worse than those from larger 9x9 grids (median standard error increased only 18%, Figure 14). On the other hand, the 5x5 grids were much more variable (median standard error increased 42%).

The evidence from two nights of trapping is more difficult to interpret. For more than half of the grids, the standard errors were actually smaller than those estimated from three nights of trapping. This is problematic, since the third night should add information leading to improved (smaller error) estimates. The observed results suggest that, 1) given the small sample sizes error estimates from program MARK may not be stable after only two nights of trapping, or 2) our use of a single simplified model for all runs in MARK resulted in models that fit the data less well with 3 vs. 2 days of trapping. Given these patterns, we did not include these results here.

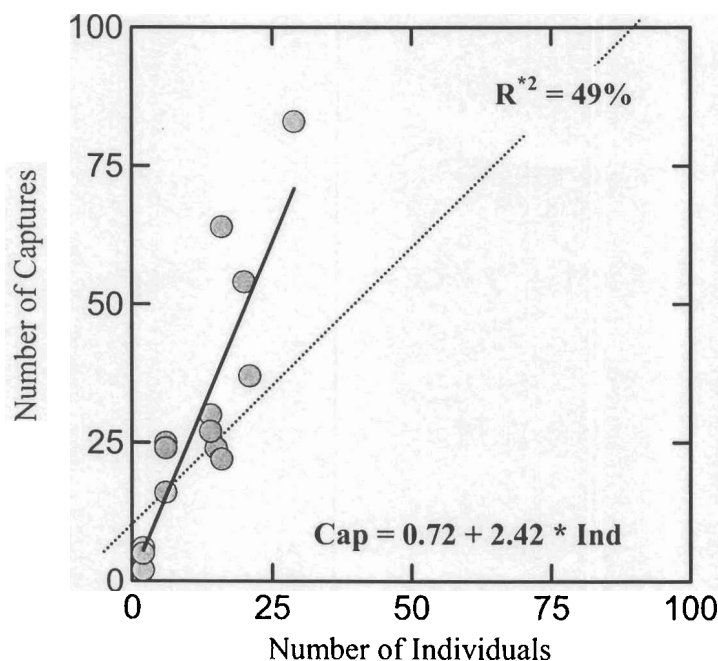


**Figure 14.** Precision of Survival Estimates from Program MARK as a function of grid size. Each point represents the percentage increase in the standard error around the estimate of survival. Data are compiled from the O'Farrell 9x9 grids and Kelt 7x7 grids for three consecutive nights.



#### D. Numbers of Captures as an Estimate of Density.

In the past, some researchers have used the number of captures in a trapping session as an index of abundance. Given the strong relationship between number of unique individuals and population estimates from MARK, we checked the adequacy of using numbers of captures as an estimate of density. We compared the number of captures to the number of individuals captured in a trapping session for a subset of Kelt's data (Figure 15). Overall, the accuracy of numbers of captures was low, and we feel that this estimator of abundance is not useful. Many animals were captured more than once. The number of recaptures was not consistent so the relationship between the number of captures and the number of unique individuals was variable. In general, all trapping of SKR should uniquely mark individuals to allow the direct enumeration.



**Figure 15.** Number of captures in a trapping session plotted against the number of unique individuals captured. Data are for Kelt NO1.

### E. Conclusions from Analyses of Program MARK.

We conclude the current sampling methods result in sample sizes too small to allow the straightforward use of program MARK. McKelvey and Pearson (2001) recently reviewed 4 years of literature on small mammal population studies, and found 98% of the samples collected in these studies were similar to the SKR datasets we investigated -- too small to allow robust selection of the appropriate capture-recapture model. In these cases, they concluded the use of an index, the number of unique individuals captured (" $M_{(t+1)}$ "), is a more reliable estimate of population size. These conclusions hold for the SKR data we analyzed. Though future studies may conclude specific models in MARK may produce reliable estimates, typical population sizes observed in the SKR monitoring program resulted in 1) poor estimates of specific population parameters characterized by extremely small or large standard errors, 2) instances where numerical convergence was not met or suspect in model runs, and 3) the inability to select between different model types when estimating population parameters.

If managers want to use MARK for estimating population parameters, then trapping effort must be increased to capture approximately 100 individuals on a grid. Based on the numbers of individuals captured for the different grid sizes used on SKR to date, the density estimates reported in the reports we reviewed, and estimates of homerange size from the draft recovery plan, we estimate a grid size of approximately 200m x 200m would be needed. With 10 m trap spacing, this would be a large grid, ( $21 \times 21 = 441$  traps). With 15 m trap spacing a  $14 \times 14$  or  $15 \times 15$  trap grid may work. Either way, the level of effort is high, and when populations are low, this level of effort may still not result in adequate sample sizes for sophisticated capture-recapture models.

### **4. Monitoring Changes in SKR Density Through Time and Across Space:**

Monitoring changes in density requires more than an accurate estimate of density at a single point. Monitoring requires the design of a network of sites sampled on a regular basis. Further, the analysis of monitoring data requires the analysis of spatially and temporally structured data. There is a small but growing literature on monitoring ecological populations through time (e.g. see papers by Fuller, Olsen, Overton, Urquhart and Vos in Literature Cited).

Much of this literature is focused on the optimal allocation of sampling effort across space and through time. The optimal design of a monitoring program depends on matching sampling effort to the system being studied and the questions of interest.

An important part of designing an efficient monitoring program is the description of the variability of the population. Populations that are more variable will require greater sampling effort to achieve adequate precision. As importantly, the spatial and temporal scale of this variability needs to be described. Populations that vary at fine scales (i.e. over small distances from site to site, or rapidly through time) will be harder to quantify than populations that change at coarser scales or more smoothly. The first step in our evaluation of a monitoring program was to describe the nature of population fluctuations for SKR. These analyses were used to describe both the amount of variability in SKR populations and the scale at which populations vary. We note that this analysis was limited by a lack of large amounts of density data collected simultaneously across many locations.

#### A. Spatial and Temporal Variability of Burrow Counts

The O'Farrell study at Lake Mathews and Shipley from 1991 to 1994 provided a comprehensive survey. In this study, 60 sites were re-visited for a minimum of four years using burrow counts. We would prefer direct estimates of density, but we felt that the large scale of this study relative to any of the other datasets we had justified this preliminary analysis. The analysis is presented in two ways.

First, the density of burrows was visualized for each of the four consecutive years from 1991 to 1994. The response surface was generated using a distance weighted least squares smoother on square-root transformed data. The visualization provides striking evidence of grid-to-grid variability in abundance (Figure 16). At Lake Mathews, grids in the north and east increased in burrows observed in 1992 compared to 1991. This increase was not observed in the western portion of this site. In 1993 and 1994 the density declined in the east and increased in the west. As a result, the locations with the highest densities observed changed from year to year. Taking each location as a sequence of four observations (1991-1995), we scored the year in which burrow counts peaked at each location. The peak number of burrows was observed at 20% of the grids in 1992, 45% in 1993, and 32% in 1994. High levels of variability were also observed at Shipley. At this site, the peak number of burrows was observed at 25% of the grids

in 1992, 20% in 1993, and 45% in 1994. This demonstrates that the number of SKR burrows changes rapidly through time and across fairly short distances.

Second, the variation in density of burrows was quantified using an orthogonal sums of squares decomposition (Figure 16). This is equivalent to an ANOVA, but formal significance tests are not possible since grids were not replicated within a year. The decomposition shows that relatively little of the change in density can be attributable to a consistent trend through time. Instead, a large amount of variation exists among grids. As importantly, the large interaction term confirms that the spatial pattern of burrow numbers change through time.

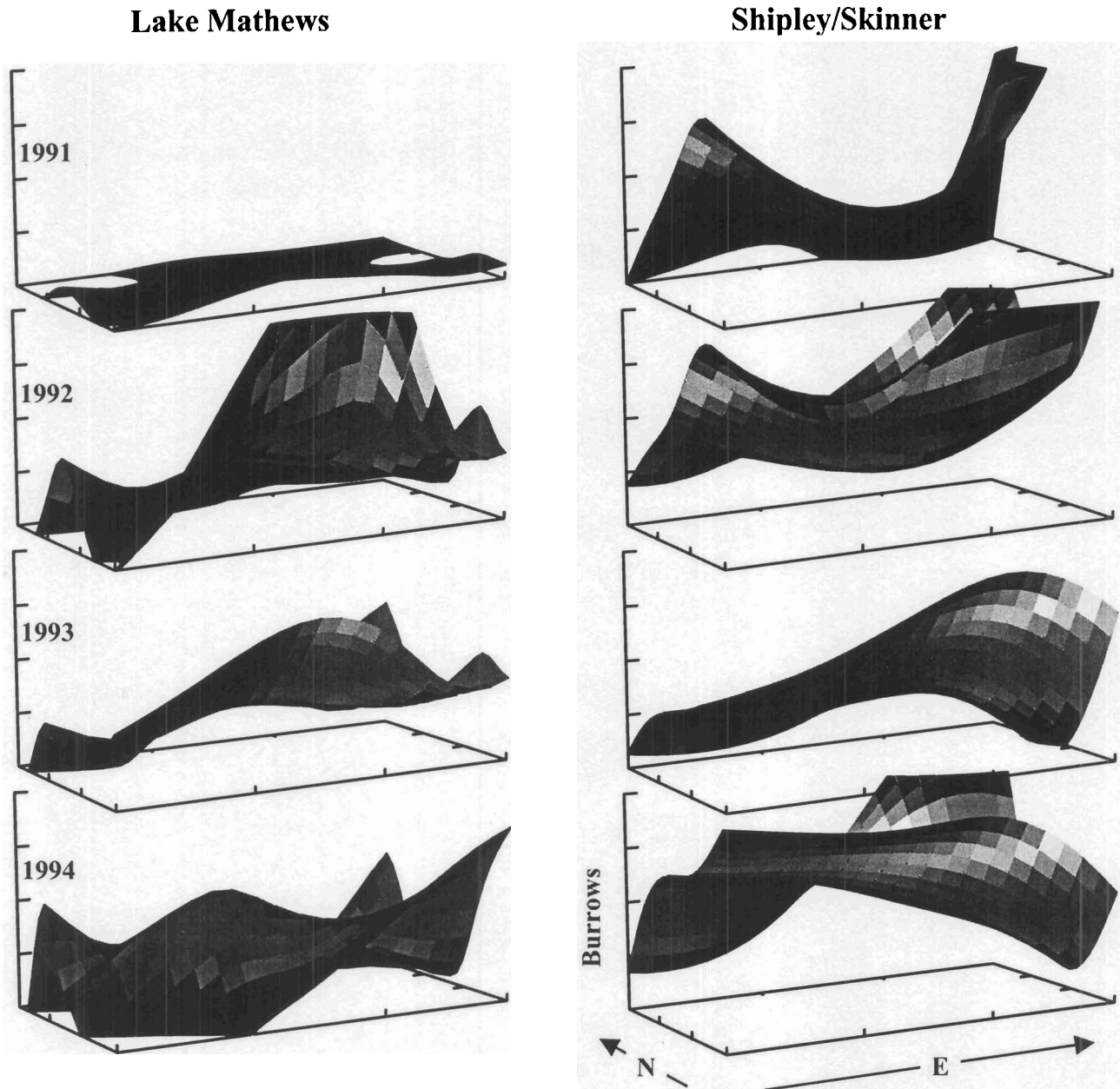
#### B. Spatial and Temporal Variability of Burrow Counts and Density

Over the past 10 years, there have been several sites monitored with both burrow counts and trapping. Some of these have multiple grids monitored over several consecutive years. These provided an opportunity to compare the variability of burrow counts to the variability in density estimates. None of these comparisons is as large and comprehensive as the burrow counts presented above, but they do provide a more direct measure of variability in SKR populations.

At March AFB, four grids were monitored with burrow counts and trapping from 1998 to 2000. Analysis of the burrow counts showed 53% of the variation in density attributable to temporal trend (Table 7). This is higher than the 20-40% observed at Lake Mathews and Shipley. However, a large fraction of the variability was associated with grid-to-grid variability. More importantly, the decomposition for trapped individuals was nearly identical to the decomposition for burrows.

At Eastside reservoir, eight grids were monitored with burrow counts and trapping in 1989 and 1990. Analysis of the burrow counts showed that the majority of the variability was associated with grid-to-grid variability (Table 8). Again, the decomposition for trapped individuals was nearly identical to the decomposition for burrows.

These analyses demonstrate that the amount of variation attributable to simple changes through time observed across all grids (i.e. trend) is highly variable. The proportion of variation attributable to trend ranges from <1% to 50%. Second, the sums-of-squares decomposition observed for individuals captured correspond well to the burrow counts. There is tremendous variation in numbers of individuals among grids and years. Unfortunately, this variation appears erratic at both broad and fine spatial scales.



Source	Sum-of-Sq	df	Mean-Sq
Year	841 (38%)	3	280
Grid	874 (39%)	31	28.2
Year*Grid	507 (23%)	93	5.5

Source	Sum-of-Sq	df	Mean-Sq
Year	123 (19%)	3	40.4
Grid	267 (40%)	13	20.6
Year*Grid	274 (41%)	39	7.0

**Figure 16.** Variation in burrow counts through time at two sites. Each response surface is estimated from at least 14 grids and fit using distance weighted least squares. Hot colors (red and orange) represent many burrows while cool colors (green and blue) represents few burrows. Sums of Squares decomposition partitions the variation into trend (Year), spatial variation (Grid) and their interaction. Data are compiled from the O'Farrell burrow counts.

**Table 7. Sums of Squares Decomposition for March AFB (4 grids, 3 years)**

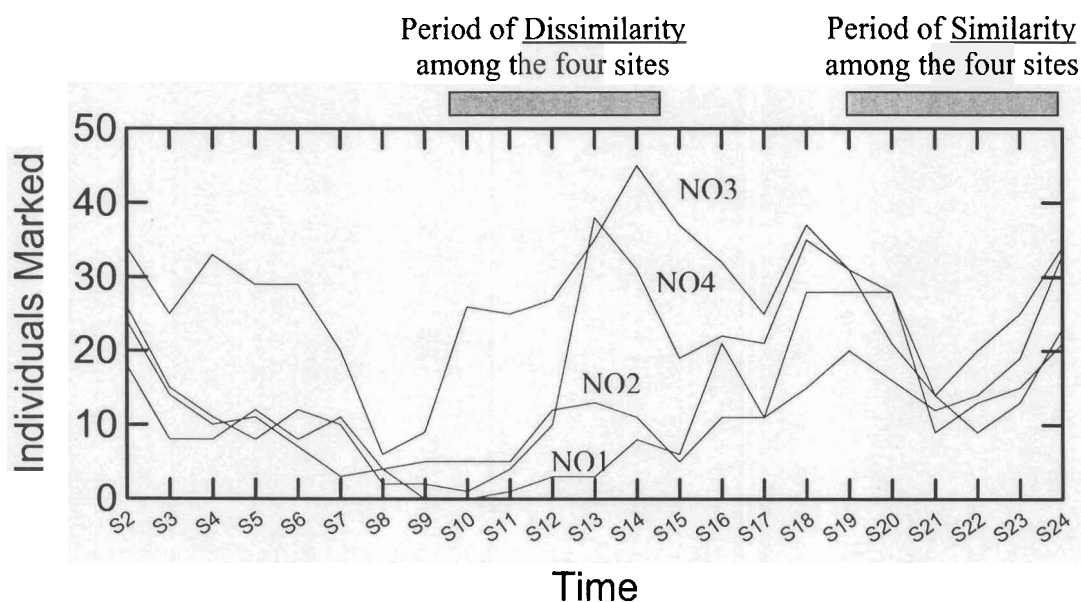
Source	df	Sums of Squares Burrows	Sums of Squares Individuals
Year	2	1401.5 (53%)	186.00 (49%)
Grid	3	944.9 (36%)	144.33 (38%)
Year*Grid	6	299.8 (11%)	48.67 (13%)

**Table 8. Sums of Squares Decomposition for Eastside Reservoir (8 grids, 2 years)**

Source	df	Sums of Squares Burrows	Sums of Squares Individuals
Year	1	400 (<1%)	16 (2%)
Grid	7	146,702 (78%)	624 (65%)
Year*Grid	7	41011 (22%)	320 (33%)

### C. Spatial and Temporal Variability of Density

We also analyzed changes in density through time and space in Kelt's detailed trapping data from Shipley. This study had the best record of change through time with 24 sessions across 4 years. Since many of the 14 trapping grids had very few animals it was difficult to include all grids in an analysis. Here we focus on four grids that had sufficient numbers to allow an analysis using the sums of squares decomposition (Figure 17).



**Figure 17. Time series of the number of unique individuals per trapping session on 4 plots trapped from 1996-2000 by D. Kelt.**

Patterns in the variance decomposition were intriguing. Overall, the relative amount of variation explained by trend through time vs, differences across sites vs, the time by site interaction, was comparable to those found in the previous analyses (Figure 17, Table 9 - top). However, this property of the entire time series consisted of periods when the density across the grids changed in concert (Figure 17 - period of similarity, Table 9 - middle) and other periods when density changes were uncorrelated (Figure 17 - period of dissimilarity, Table 9 - bottom). During the period of dissimilarity, the majority of the variance was explained by differences in abundance across grids. During the period of similarity, the opposite occurred and large amounts of variation (66%) were explained by variation through time.

Taken as whole, the patterns observed in Kelt's data, match well with the patterns we observed in the other datasets we analyzed. For example, March AFB and Eastside Reservoir differed greatly in the amount of variation attributable to time (Table 7 and 8). It is possible March was sampled during a period of similarity while Eastside reservoir was sampled during a period of dissimilarity. Given the rapid changes in abundance across space and through time (Figures 16 and 17), it is possible the patterns in the variance decomposition would have reversed had sampling continued.

**Table 9.** *Sums of Squares Decomposition for Shipley (4 grids, 4 years)*

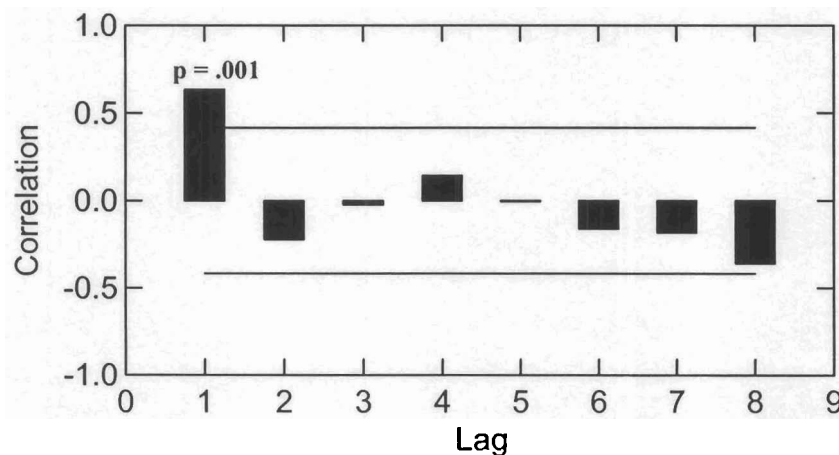
Source	df	Sums of Squares Individuals
Entire Record		
Year	22	4628 (43%)
Grid	3	3515 (32%)
Year*Grid	66	2676 (25%)
Period Of Similarity		
Year	3	576.5 (66%)
Grid	3	210.5 (24%)
Year*Grid	69	87 .0 (10%)
Period Of Dissimilarity		
Year	3	629 (20%)
Grid	3	1965 (64%)
Year*Grid	9	484 (16%)

#### D. Temporal Autocorrelation in Density Data and Sampling Frequency.

Because Kelt's data was sampled so frequently and for such a long period of time, we were able to perform an autocorrelation analysis to determine if samples taken relatively near each other in time were correlated and help determine an adequate sampling frequency. If sampling occurs frequently enough that the numbers of individuals at time  $t$ , are correlated with the numbers of individuals at time  $t+1$ , then the data collected is measuring, to some degree, the same population and is redundant.

The numbers of individuals captured in a trapping session were correlated with numbers of individuals captured the preceding trapping session. This pattern can be seen by looking at the autocorrelation plot for grid NO4 (Figure 18). Correlations for the four grids analyzed ranged from 0.56 to 0.68 ( $p < .006$  for all grids). This suggests that the frequency of sampling used by Kelt could be reduced without significant loss of information. On average, Kelt sampled approximately once every 2 months, so a sampling frequency of 3 or 4 times a year may be adequate for tracking changes in density through time. We emphasize this conclusion is preliminary and has only been done on four grids from a single site. This issue should be analyzed more fully before robust guidelines can be developed.

#### Partial Autocorrelation Plot



**Figure 18.** Partial autocorrelation plot. This depicts the direct correlation between pairs of points separated by time lags of 1 to 9 trapping sessions. Notice that only a lag of 1 (i.e. adjacent trapping sessions) was significantly correlated.



## CONCLUSIONS AND RECOMMENDATIONS

Effective management of SKR requires knowledge in two main areas: Understanding patterns of population change through time and across space (i.e. what/where) and understanding the biological processes driving these changes (i.e., how/why?). The key to creating this level of understanding is a monitoring program that obtains high quality information relevant to understanding both pattern and process in SKR populations. Our analyses lead us to a number of conclusions and recommendations to improve current SKR monitoring efforts. We describe these below, synthesizing our findings into a recommended framework for monitoring SKR over the next 3-5 years.

### 1. CONCLUSIONS AND RECOMMENDATIONS

#### A. Point Estimation of Abundance.

For estimates of population size at a single point, we recommend live trapping, marking and counting unique individuals. Although this is time-consuming and expensive, it is the only direct method to estimate abundance. Neither burrow counts nor total number of captures are correlated strongly enough with density to be used as surrogates. Burrow counts cannot be easily related back to population size because the relationship between burrow counts and  $N$  varies tremendously with space and time (Figure 4). Similarly, the total number of captures is unreliable as shown in our analyses (Figure 15) and, by a host of other researchers (McKelvey and Pearson 2001). *We recommend that all SKR monitoring protocols should require trapping and the unique marking of all individuals captured. The number of unique individuals captured or the Minimum Number Known Alive should be used as indices of population size.*

WHY MARK  
UNIQUELY  
FOR  $M_{t+1}$ ?

#### B. Monitoring Across Space and Through Time.

SKR populations are highly variable across space and through time making large-scale trend detection extremely difficult. Changes in SKR populations across a site were often uncoupled (Figure 16). In addition, populations showed pronounced fluctuations both seasonally and annually (Figure 17). These complex changes in density make detecting trends in population sizes across larger areas (i.e. regionally or within reserves) extremely difficult. Monitoring protocols such as rotating panel designs where sites are revisited on a rotating basis will likely be

ineffective for SKR. Instead, a monitoring protocol including simultaneous sampling of many sites across large areas is necessary. ***We recommend synchronous trapping of SKR across a large number of grids at each site or reserve.***

The correlation between burrow counts and SKR abundance is strong enough to warrant further research. The similarities in the variance partitioning analyses (Figure 16, Tables 7-9) suggest that burrow counts fluctuate in a manner that is similar to abundance. Unlike trapping, burrow counts can be done quickly, so many more sites can be visited, and replication through space is critical to detect reserve-wide changes in SKR abundance. It is possible that burrow counts will be a cost-effective adjunct to trapping or can be used in estimating general trends in SKR abundance across larger regions. ***We recommend continuing to study the relationship between burrow counts and abundance and to assess whether burrow counts can be used as an adjunct to trapping.***

#### C. Effort and timing of trapping for trend detection.

Our power analyses showed the expected loss in accuracy with reduced trapping effort (Figures 6-8). Precision was fairly high when grids were reduced from 9x9 to 7x7 but not when they were further reduced to 5x5. Capture probabilities were typically around 60% (Figure 10) so trapping for two consecutive nights was usually adequate. These analyses must be interpreted with caution because we were only able to analyze sites with abundant captures. In addition, our analyses assumed the more extensive trapping efforts represented the true abundance at a location. Given these caveats, the most cost-effective protocol would include sampling for 2 days on a 7 x 7 trap grid. ***We recommend a minimum grid size of 7x7 for 2 nights.***

It is likely that an extra night of trapping is more expensive than using slightly larger grids. As a result it may be more cost effective to increase sampling effort by using larger grids, than by trapping additional days at a given site. On the other hand, shorter trapping periods limit the ability to use certain capture-recapture models and estimating probabilities of capture. Increases in grid size (to 9x9) and the addition of a third night may improve estimation. ***We recommend that sampling not exceed 9x9 grids for 3 nights unless the cost is justified by a specific research or management objective.***

Abundance estimates on 4 grids trapped by Kelt at 2-month intervals were positively correlated (Figure 18). This suggests that sampling 6 times a year was not necessary to detect changes in population size. It is likely that sampling 2 to 4 times a year will be adequate for

trend detection. We note however, that for some studies, such as estimating survival, or reproductive activity, sampling less than 4 times a year will be inadequate. To better plan the timing of trapping, SKR biologists should be consulted to determine the optimal months or seasons for trapping given available resources. ***We recommend that sampling be conducted between 2 and 4 times per year and that sessions should be at least 2 months apart.***

**D. The use of MARK and estimating vital rates.**

The small sample sizes associated with most trap sessions make it difficult for any closed capture, or robust design model to estimate population parameters (Table 6). Using alternative capture recapture methods such as Cormack-Jolly-Seber approaches, model averaging in program MARK, or incorporating burrow counts as a covariate, may all increase the ability to estimate population parameters. However, the utility of these techniques when sample sizes are small are still uncertain. Estimates from capture-recapture models are highly correlated with the number of individuals captured because of the relatively high capture probabilities exhibited by SKR. ***We recommend against the routine use of Mark to estimate population parameters unless sampling effort is increased to attain adequate sample sizes of ~100 individuals per grid.***

A key strength of MARK is the ability to fit different types of models to a capture-recapture data set to test specific hypotheses about a population. In a number of cases, even relatively simple models would not converge in MARK, making it impossible to compare different models. Thus, the sampling efforts employed to date will limit the use of MARK for testing hypotheses (i.e. animals of different ages, sexes, or body masses have different survival rates, or survival rate depends on habitat type). ***Though not a top priority, we recommend further analyses of the current data as well as simulation studies using program MARK to investigate: 1) the ability of MARK, to distinguish between different models of biological process in SKR populations and 2) to define the optimal sampling effort needed to allow accurate estimation of  $N$  and hypothesis testing.***

## **2. RECOMMENDATIONS FOR ADDITIONAL MONITORING**

Here we describe a monitoring program, starting with what we consider the minimal monitoring effort to employ over the next 3-5 years, and then continuing with 3 additional steps that add additional, critical information about SKR. The primary goal is to begin producing

detailed data on the *pattern* of SKR population dynamics, allowing us to better gauge the necessary number of trapping grids and the spacing of these across reserves in future monitoring. The current data on SKR, though nearly 12 years long, was not planned well enough to sufficiently address this question, thus we must “monitor to understand how to monitor” for the next 3-5 years. The remaining 3 recommendations are designed to go beyond the pattern gathering methods described initially and develop information that informs managers about the underlying *processes* driving SKR population dynamics.

#### A. Synchronized Trapping Across All Reserves.

**Recommendation:** We strongly recommend a coordinated trapping effort, implemented immediately, across all reserves. In all cases, trapping should take place at least 2 but no more than 4 times per year, simultaneously at all locations (a 1-2 month “window” of trapping should be adequate), using 7x7 trapping grids or larger, trapped for 2 or 3 days. Within each reserve, if funds are available, 10-20 sites should be sampled. At each site, SKR should be individually marked and a burrow count assessment done using at least 3 assessment lines. This effort should last for at least 3 years.

If 10-20 sites is too expensive, a subset of at least 6 sites should be trapped while simultaneously surveyed for burrow counts. Burrows should also be counted at an additional suite of sites (~20). The simultaneous sampling of individuals and burrows allows for the validation of the relationship between burrow counts and density at that point and time, increasing the ability to understand patterns in SKR abundance.

*Grid shape:* Based on an earlier draft, we were asked to comment specifically on the use of 4 x 10 trapping grids. We do not recommend their use for a number of reasons. First, they are too small to capture an adequate number of individuals, even if indices are employed. The recommended 7 x 7 (49 trap) grids are a bare minimum, and reducing the number of traps to 40 is unwarranted. Second, square grids have much lower perimeter to area ratios and will cover the entire homerange of more animals than will a narrower rectangular grid. Thus, a 4 x 10 grid will capture proportionately more animals whose homeranges just overlap the grid. This could 1) increase variability in indices based on rectangles relative to grids, 2) make estimation of movement difficult because fewer animals are fully encompassed in the trapping area and 3) result in fewer recaptures of individuals, perhaps reducing the ability to estimate survival or use capture-recapture algorithms. Although 4 x 10 grids have been used in conjunction with 4-

transect burrow counts, there is no reason that the burrow count protocol cannot be modified to include 7 shorter transects. As a result, we strongly favor square grids.

**Rationale:** The data collected to date suggest the pattern of change in SKR abundances is extremely complex, despite the lack of coordination compared across reserves or through time. This conclusion is based on four studies. The three studies that used trapping grids were limited to a few grids and were of short duration. In contrast, O'Farrell's burrow count data is much more extensive. Although the results from the four studies were similar, the data is too sparse to allow a more definite conclusion. Thus, to better estimate patterns of variation in population size, we need 3-5 years of data, sampled simultaneously across a minimum of 10-20 locations, within each reserve. This effort will allow: 1) An accurate estimation of how SKR populations vary across space, through time, and the complexity of the pattern. This information is critical to assessing the number of trapping sites, and their distribution across reserves necessary for a high quality, long-term monitoring program. 2) A spatial autocorrelation analysis, similar to the temporal autocorrelation analysis we performed on Kelt's data, to ascertaining the level of spacing necessary to have independent samples across a reserve. 3) A detailed analysis of the relationship between burrow counts and abundance, particularly focused on the use of burrow counts to estimate trends in population size at the reserve, or across-reserve scale.

With these three types of information in place, we can then determine the most cost effective method for monitoring SKR that produces information at sites, across sites within reserves, and across the entire SKR reserve system.

#### B. Estimating Demographic Processes.

**Recommendation:** We recommend 2-3, large 15 x 15 trapping grids placed in each reserve. These grids should be trapped with 4-5 day trapping sessions, 6-12 times a year. If possible, the grids should be stratified across habitat types, or disturbance gradients considered critical for SKR.

**Rationale:** The larger grids, and more intense sampling would allow a detailed investigation of demographic processes in SKR populations. This effort would allow: 1) A detailed investigation of the relationship between true population size and the number of individuals captured, a necessary component if less labor intensive methods will be used in the

final monitoring program. 2) High quality estimates of survival, stratified by sex, season of first capture, weight, age, or body condition. 3) Information on the timing of breeding, reproduction, and recruitment into the population. 4) Information on short distance movements and homerange size. 5) Finally, if the grids are stratified across habitat types, changes in demographic processes across these habitat types could be quantified.

### C. Habitat and environment relationships.

**Recommendation:** We recommend collecting data and conducting Before, After, Control, Impact (BACI) experiments on particular environmental factors hypothesized to influence the abundance of SKR. Initially, vegetation and rainfall data should be collected at each site SKR are trapped. The method for vegetation sampling should allow the gathering of variables thought to most impact SKR (e.g. percent cover, percent bare ground, diversity/cover of grasses, forb to grass ratios etc). In addition, sampling sites could be stratified across environmental gradients such as the time since last fire (or any successional gradient), soil type, anthropogenic disturbance, or other variables considered by SKR biologists. BACI experiments should be conducted by manipulating habitats while simultaneously performing detailed capture-recapture studies, or integrated into planned management activities, such as controlled burns. Finally, remotely sensed data that can estimate aspects of the vegetation community or habitat disturbance might be useful in predicting SKR abundance.

**Rationale:** Our preliminary analysis illustrated a very high level of variability in SKR populations that was evenly distributed through time and space. Managers must begin understanding the factors influencing these patterns of variation to effectively manage SKR populations. Collecting environmental data at each site would allow biologists to see relationships between patterns of SKR abundance and particular environmental variables, some of which might be manipulated to allow increased SKR populations or more stable populations. The BACI experiments would 1) experimentally confirm the impact of an environmental variable on SKR abundances, 2) help managers gauge the magnitude of a particular management practice on SKR, and 3) elucidate the biological mechanisms causing the changes in abundance to a particular variable.

O'Farrell conducted a BACI experiment from approximately 1991-1994 at Lake Matthews using 40 trapping grids, some of which were burned or disked. Unfortunately, O'Farrell estimated abundance using burrow counts only. Given the high levels of variation in the relationship between burrow counts and true abundance across sites and through time, the results of this well designed experiment are limited in our opinion.

#### D. Population Viability or Simulation Modeling of SKR.

**Recommendation.** Though not specifically a monitoring protocol, a detailed model of SKR population dynamics should be developed.

**Rationale:** The ultimate goal of population modeling is to accurately predict the future state of a population. Once a verified model is in place, managers can compare the outcome of proposed/different management practices on the population in question. This ability to predict population viability under different management scenarios is what makes population modeling so effective and beneficial. The information and understanding of SKR population ecology generated from the first 3 recommendations should be synthesized in the modeling effort.

Modeling populations is an iterative, time-consuming process that could/should drive the allocation of resources in future monitoring efforts, and why we include it here. For example, what if 2 years into the modeling effort, it becomes apparent that extremely accurate estimates of survival across burned versus unburned sites are needed? If so, then some of the ongoing monitoring effort should be stratified across these types of habitat for a number of years. In addition, once a functional model is in place, monitoring in new locations might be necessary to validate model predictions. Thus, if a long-term goal of SKR management is to predict the future state of SKR under various management scenarios, a modeling effort should begin, and the data needs of the model incorporated into a monitoring program.

#### E. Prioritizing Management Actions.

Perhaps the hardest task SKR managers face is prioritizing the use of limited research dollars. We realize we have created a list of expensive recommendations, which if implemented simultaneously, or even in pairs, would surely "break the bank." Our recommendations boil down to a trade off between meeting immediate, applied objectives of estimating population size, versus a longer-term, thorough understanding of SKR population ecology. We feel a reserve-

wide, organized, synchronized sampling protocol would meet the immediate, applied objectives, and lay the ground work for the longer-term understanding of SKR biology. We would place synchronized trapping across all reserves as the highest priority, while simultaneously admitting our lack of knowledge regarding SKR biology and historical work done to date on the species. We hope reserve managers take our results and recommendations not as a “directive for action” from two academic ecologists but as additional information to synthesize into their current expertise when making management decisions for the species.



- in the United States. *Environmental Monitoring and Assessment* 54: 1-45.
- Overton, W. S., and S. V. Stehman. 1996. Desirable design characteristics for long-term monitoring of ecological variables. *Environmental and Ecological Statistics* 3: 349-361
- Urquhart, N. S., S. G. Paulsen, and D. P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. *Ecological Applications* 8: 246-257.
- Urquhart, S. N., and T. M. Kincaid. 1999. Designs for detecting trend from repeated surveys of ecological resources. *Journal of Agricultural, Biological, and Environmental Statistics* 4: 404-414.
- Vos, P., E. Meelis, and W. J. Ter Keurs. 2000. A framework for the design of ecological monitoring programs as a tool for environmental and nature management. *Environmental Monitoring and Assessment* 61: 317-344.
- White, G. C., Burnham, K. P. (1999) Program MARK: Survival estimation from populations of marked animals. *Bird Study* 46 (suppl.): 120-139

## **APPENDIX 1**

### **Use of $R^2$ to compare two Estimation Techniques**

Throughout the report, we compare estimates derived from different sampling or estimation methods. Describing the concordance of two sampling or estimation techniques requires two components. First, we regress the method that is presumed to be more accurate or of greater interest (Y: e.g. the number of unique individuals trapped per hectare) against the second, presumed less accurate method (X: e.g. density of burrows). This choice of X and Y is non trivial since linear regression assumes that X is known without error and that all prediction error is associated with the measurement of Y (Zar 1999). Second, we describe the accuracy of the prediction based on the coefficient of determination ( $R^2$ ) used in linear regression.

In ordinary linear regression the coefficient of determination is defined as:

$$R^2 = 1 - \frac{RSS}{CSS}$$

where RSS is the residual sums of squares unexplained by the regression line and CSS is the corrected sums of squares of the response variable (Neter et al 1996, Zar 1999). The coefficient of determination is a measure of the amount of variation in Y explained by the linear relationship between X and Y. Values range from 0% to 100% with higher values indicating more accurate prediction. Accuracy of 100% is only achieved when the independent variable (X) is a perfect predictor of the response variable (Y).

In many of the regressions presented in this work, we wish to evaluate the utility of a coarser measure to accurately estimate a more refined measure (e.g. coarser burrow lines instead of more costly and extensive burrow cells). In these applications, the regression model of interest is the 1:1 regression line (i.e. regression line with intercept = 0 and slope = 1). This is the appropriate model to test whether the coarse measure is equivalent to the more refined measure. This requires a different measure of concordance than the  $R^2$  from ordinary (unconstrained) linear regression. For these relationships, the calculation of the coefficient of determination is more difficult (Neter et al 1996) and the values reported from common statistical packages are often flawed.

In constrained regressions we use the same approach to calculating the accuracy of the relationship. We define  $R^{*2}$  as

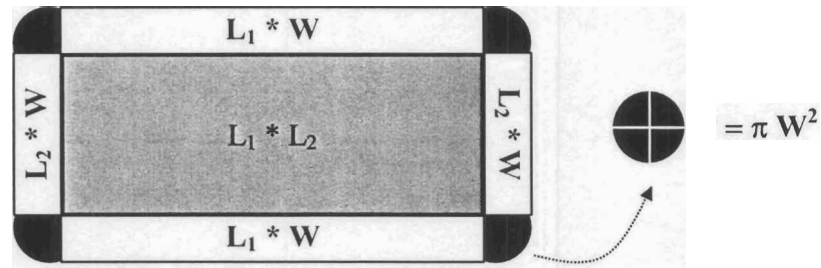
$$R^{*2} = 1 - \frac{RSS^*}{CSS}$$

where  $RSS^*$  is the residual sums of squares around the 1:1 regression line and CSS is defined as above. This provides an appropriate metric for this application of regression (Neter et al. 1996, Ryan 1997, Zar 1999). Since the 1:1 regression has fewer free parameters than the traditional linear regression model, it will have a larger  $RSS^*$  value and thus a smaller  $R^{*2}$ . However, when the constrained regression is a good approximation, the difference will be small.

trapped). The correct formula for rectangular grids is:

$$\text{Area} = L_1 L_2 + 2L_1 W + 2L_2 W + \pi W^2 \quad (\text{A2.2})$$

where  $L_1$ ,  $L_2$  and  $W$  are defined as above. The correct formula can easily be verified with a diagram (Figure A1). Equation (A2.2) accounts the size of the grid ( $L_1 * L_2$ ) as well as a border extending from the perimeter of the trapping grid.



**Figure A1:** Calculation of the area sampled for a rectangular grid. It is assumed that animals can move a distance of  $W$  from outside the grid and be caught. Thus the quantities  $L_1 W$ ,  $L_2 W$  and  $\pi W^2$  are used to account for this border zone.