

November 20, 2013

Final Report for NPS I&M Task Agreement J8W07100041

Kirk Steinhorst, Professor Emeritus, Department of Statistical Science, University of Idaho

This project was initiated on September 15, 2010 and concluded on November 3, 2013. It was, in fact, a continuation of a 5 year TA designed for the same purposes: J8W07060004. The objectives in both cases were described in the abstract of the current TA:

PROJECT ABSTRACT: The National Park Service (NPS) Inventory & Monitoring (I&M) networks within the Pacific West Region will collaborate with the University of Idaho (UI) in a technical assistance project to: 1) design and test robust sampling strategies to achieve sampling objectives of several monitoring protocols; 2) identify, test, and revise statistical approaches to determine status and trends in ecological indicators; 3) respond to peer review comments on sampling design and statistical analyses; 4) prepare narratives, SOPs, and appendices for incorporation into monitoring protocols; and 5) report the results of preliminary tests and analyses.

In short, the task agreement provided statistical support for 7 NPS I&M networks: KLMN, MEDN, MOJN, PACN, SFAN, SIEN, and UCBN. It also provided statistical support for giant sequoia monitoring through an agreement with the USGS.

During the current TA, the PI hired Leigh Ann Starcevich, Steve Hayes, and Matt Nahorniak to help with statistical analysis. Steinhorst as PI provided senior statistical oversight, project management, communication with network managers, and accounting.

We were involved with a large number of projects some of which are summarized below. First I want to give an overview of the statistical principles and techniques employed.

1. Design of long-term monitoring protocols from a statistical viewpoint generally followed the approach and notation given in McDonald (2003). The basic premise is that we define a finite population of sampling units whose spatial extent defines the target of the monitoring protocol. Selecting these units using some random mechanism allows us to generalize to the population as a whole. That generalization is compromised somewhat when nonrandom (e.g. legacy) sampling units are included. The randomly selected sampling units are then measured according to some temporal revisit design. McDonald devised a notation for dealing with the spatial and temporal design. Sets of sampling units are organized into “panels” where all sampling units within a panel are sampled the same year. As an example, he describes a 3 panel design $[(1-2)^1, (2-1)^2]$ where the (1-2) means that the panel is visited the first year and then skipped for two years. The (2-1) notation implies that those two panels are visited two years in a row and then skipped the following year. Sampling units can be assigned to panels using a variety of sampling designs—simple random, systematic random, stratified random, GRTS, etc.

Note that the sampling units will generally be some land unit—a quadrat, a transect, a map unit, a river segment, etc. Informally, we often think of the biological population—all invasive plants or all snowy plover—instead of the statistical population. This doesn't work statistically because we can't randomly sample without having a list of the sampling units first. If we had a list of all of the snowy plover at Point Reyes National Seashore, we wouldn't have to survey them ☺. The sampling units in this case are beach segments. For invasive plants, the sampling units might be trail or road segments. In some cases the sampling units might not be areas of land. For example, if you have an aerial map that shows all osprey nests, then nests could be the sampling unit.

2. For reporting status (as opposed to trend), we most often use “design-based” inference. Design-based inference lets us calculate estimates, standard errors, and confidence intervals knowing only the randomization design used to select the sampling units. We do not need to specify a statistical distribution for the data. Consider the N sampling units making up the (statistical) population of interest, SU_1, SU_2, \dots, SU_N . These sampling units can be arranged in strata or clusters or linearly (along a trail, for example) or in some other configuration. Let $\pi_1, \pi_2, \dots, \pi_N$ be the probabilities of sampling each of the N sampling units. The probabilities do not have to be equal. They just have to be known. If a sampling probability were zero, then its sampling unit would not be part of the population. Given data on a random sample of these units, y_1, y_2, \dots, y_n , then the Horvitz-Thompson theorem (Horvitz and Thompson, 1952) states that unbiased estimates of the total, τ , and mean, μ , are given as

$$\hat{\tau} = \sum_{i=1}^n \frac{y_i}{\pi_i} \quad \text{and} \quad \hat{\mu} = \frac{1}{N} \sum_{i=1}^n \frac{y_i}{\pi_i}.$$

In the case of simple random sampling, $\pi_i = n / N$ so that $\hat{\tau} = N\bar{y}$ and $\hat{\mu} = \bar{y}$. The standard errors of these estimates depend on the joint probability of two sampling units being chosen, π_{ij} . Sometimes these joint probabilities are easy to find, sometimes not. In the latter situations, we use approximations to the standard error formula which avoid needing to know π_{ij} .

In short, if we know how the sampling units were sampled, then we can devise estimators, standard errors, and confidence intervals.

3. For most of the trend analyses, we have been using “model-based” statistical techniques as opposed to “design-based” techniques although sampling weights can be incorporated if desired. The statistical model proposed by Piepho and Ogutu (2002) provides a rich basis for much of this work. The model is specified as,

$$y_{ijk} = \mu + w_j \beta + b_j + a_i + w_j t_i + c_{ij} + e_{ijk}$$

where $i = 1, \dots, \# \text{ sites}$, $j = 1, \dots, \# \text{ years}$, and $k = 1, \dots, \# \text{ visits / year}$. $w_j \beta$ is the linear trend over years (w_j is years since the start of monitoring); b_j is the random (and possibly nonlinear) effect of year; a_i is the random effect of the i^{th} site; $w_j t_i$ is the multiplicative interaction of year and site; c_{ij} is the random interaction of site and year; and e_{ijk} is random error. If there is only one site visit per year, then c_{ij} and e_{ij1} cannot be distinguished. We usually assume that y_{ijk} is measured on a log scale, because many biological variables are skewed to the right—suggesting a log-normal distribution. We are particularly interested in estimating β , its standard error and in testing $H_0: \beta = 0$.

4. When we don't use the model in 3., we use a variety of statistical approaches for detecting trend—zero-inflated count models, nonparametric correlation, or unique models associated with a particular response variable (for example, dynamic occupancy models for terrestrial herps at Santa Monica Mountains National Recreation Area).
5. The question of sample size requires different approaches for design-based inference and model-based inference. For sample surveys (design-based) we usually choose sample size to ensure that confidence intervals will be acceptably narrow so that we can be assured that status is precisely measured. For trend analyses, we usually do a power analysis using analytic or simulation based methods.

In the last three years, we have helped with quite a list of protocols. To give you an idea, I list some of the projects by network.

KLMN: landbirds, lake index of ecological integrity, 5 needle pine, R/statistics workshop

MEDN: terrestrial vegetation, aquatic herps, terrestrial herps, landscape dynamics, invasives

MOJN: integrated upland, groundwater and small springs, riparian vegetation and large springs

PACN: vegetation, invasives, marine fishes and benthos

SFAN: Coho, vegetation, prairie falcon, snowy plover, R/statistics workshop

SIEN: lake chemistry, wetlands, rivers and streams, 5 needle pine; giant sequoia analyses through USGS cooperative agreement

UCBN: aspen, water quality, 5 needle pine

Sampling of SOPs and reports

Trend and Power Analysis for National Park Service Marine Data from 2006 to 2010

Pacific Island Network Natural Resource Technical Report NPS/PACN/NRTR—2013/771

Leigh Ann Harrod Starceвич

Upper Columbia Basin Network Aspen Monitoring Report 2008-2012

City of Rocks National Reserve (CIRO) Natural Resource Technical Report NPS/UCBN/NRTR—2013/795

Eva K. Strand, Ph.D., Stephen C. Bunting, Ph.D, Matthew T. Nahorniak, Leigh Ann Starceвич,

MEDN_InvasivePlantProtocol_Sept26_LAS.pdf **DRAFT**

MOJN_GRTS_AnalysisFiles_20131103.zip – GRTS analysis files for 3 parks—Starceвич

PACN Invasive Status Estimation_20130727.pdf—Starceвич

SFAN_PRFA trend_20131101.pdf – prairie falcon status and trends analysis report. Trends in occupancy, productivity, and phenology and status estimation for total prairie falcon population size—Starceвич

SIEN_LakeChem_PowerAnalysis_20131103.doc – updated power analysis for SIEN lake chemistry based on monitoring data from 2008-2011—Starceвич

WEI_Power_RCode.zip – power analysis commands for SIEN wetlands ecological integrity—Starceвич

UCBN Water Quality Trend Analysis 20131010.pdf—Starceвич

Sequoia Seedling Population Dynamics: Power Analysis and Sampling Design Recommendations—

Matt Nahorniak, Oct 2013

Detecting Mortality Increases over Time in Mature Sequoia Trees: Power Analysis Comparison of

Various Sampling Protocols—Matt Nahorniak, Oct 2013

SAMO Terrestrial Herp Occupancy Status and Trend (2000-2010) / Power to Detect Trend (MEDN)—

Stephen Hayes and Kirk Steinhorst, September 2011

Terrestrial Vegetation Monitoring in the Mediterranean Coast Network

Cabrillo National Monument, Channel Islands National Park, and Santa Monica Mountains National Recreation Area—

John Tiszler, Tarja Sagar, Timothy Handley, Christy A. Brigham, Irina C. Irvine, Marti S. Witter, Dirk

Rodriguez, Keith Lombardo, Tom Philippi, Kathryn McEachern, Leigh Ann Harrod Starceвич,

September 2011

Mojave Desert Inventory and Monitoring Network Streams and Lakes Protocol Natural

Resource Report NPS/MOJN/NRR—2012/001 Christopher C. Caudill, Debra Hughson,

Geoff Moret, Alice Chung-MacCoubrey, Jennifer Burke, Gretchen Baker, Nita Tallent,

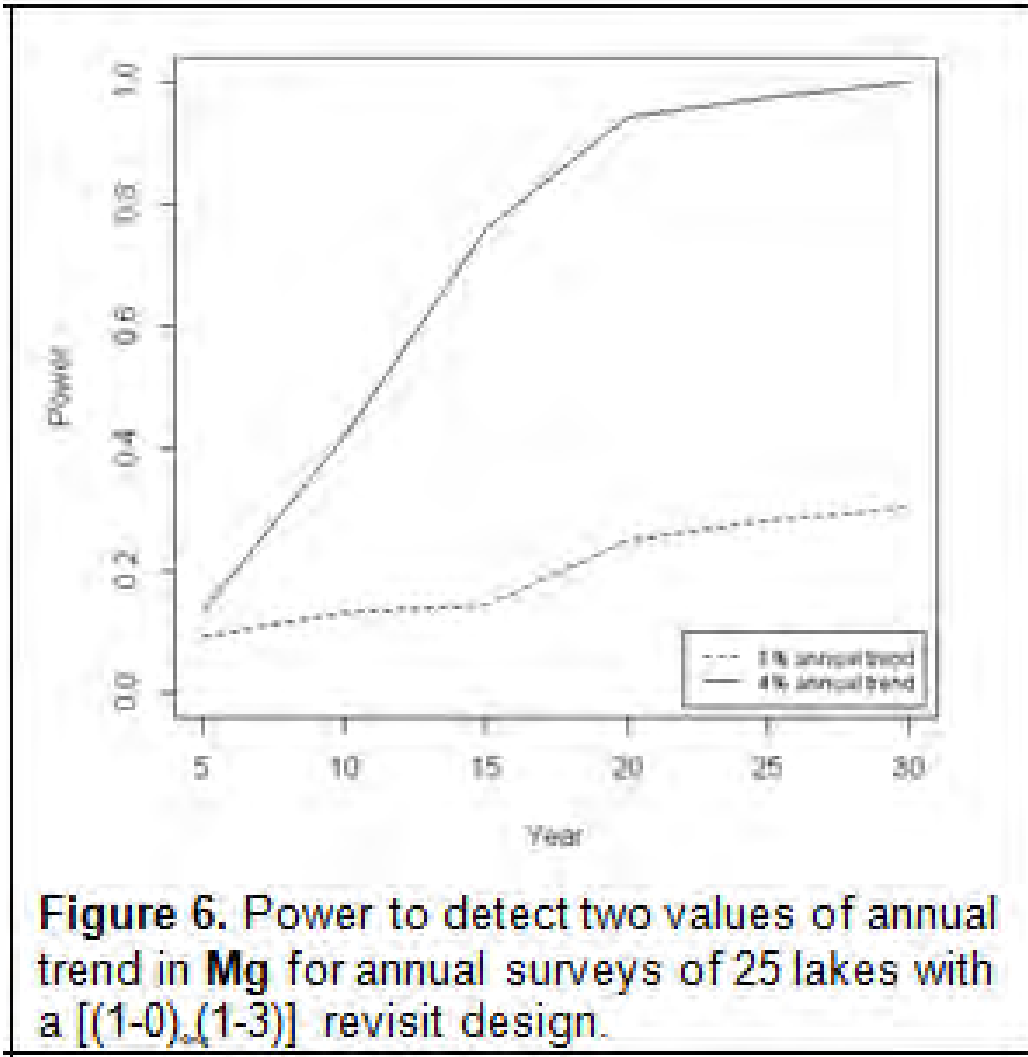
Leigh Ann Starceвич, Kirk Steinhorst

Illustrative Examples

The following examples illustrate the various methods.

A. **Power Analysis for National Park Service Lake Chemistry Data from 2008 to 2011** *Sierra Nevada Network* Leigh Ann Harrod Starcevich November 2013.

The Piepho and Ogutu model with additional covariates was fit to water chemistry data (17 parameters including ANC, Ca, Cl, DIN:TP, P, Mg, Na, NO_3) collected from 24-25 lakes from 2008 to 2011. The models were fit using the REML method in the statistical package R. This provided estimates of means and components of variance which formed the basis of a Monte Carlo estimate of power assuming a $\log(Y)$ response with 1% or 4% annual increases in the mean of Y. The original lakes were selected using the GRTS method with unequal probabilities depending on travel times to each lake. An example power curve for magnesium is



B. **Power analysis for stream discharge in Great Basin streams, Baker Creek and Lehman Creek, in MOJN Streams and Lakes Protocol, version 1—analysis by Leigh Ann Starcevich, February 2012.**

A simplified Piepho and Ogutu model with terms for creek and linear trend in years was fit to pilot data collected since 1992. Based on parameters estimated from this model, Monte Carlo methods were used to estimate power as a function of possible trends and number of years. Power for annual mean daily discharge and annual maximum daily discharge appear below.

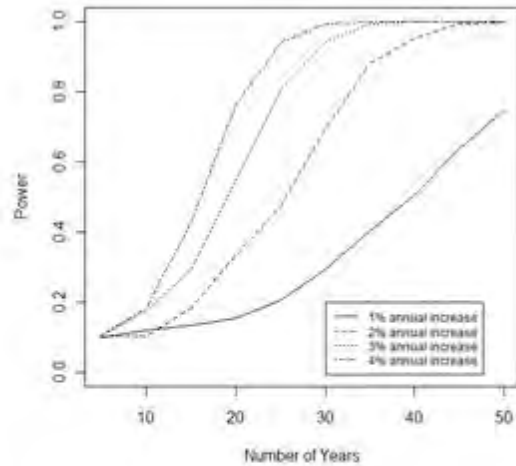


Figure 2.5. Power to detect trend in annual mean daily discharge in Baker Creek or Lehman Creek as a function of the number of survey years.

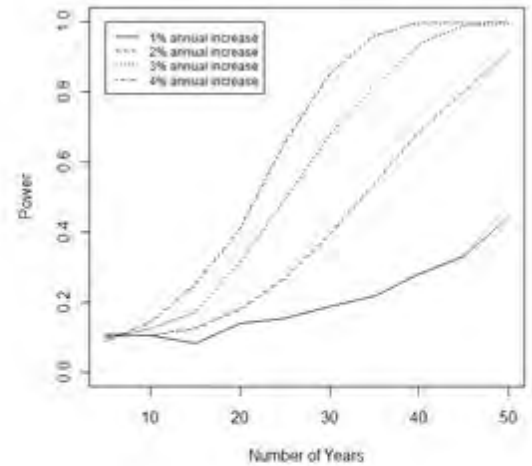


Figure 2.7. Power to detect trend in annual maximum daily discharge in Baker Creek or Lehman Creek as a function of the number of survey years.

C. SAMO Terrestrial Herp Occupancy Status and Trend (2000-2010) / Power to Detect Trend
Mediterranean Network, Hayes and Steinhorst, September 2011

We used dynamic occupancy models (MacKenzie et al. 2006) and maximum likelihood in the R (R Development Core Team 2011) package “unmarked” (Fiske and Chandler 2011) to identify the best fitting model and to estimate trend effects. Occupancy data were available for 10 reptile species collected from 2000-2010 using pitfall array trapping. Habitat type, region, and elevation data were available for site covariate modeling. We used hidden Markov model smoothing, a procedure available in package “unmarked”, to estimate trend (i.e. site-specific occupancy by year) .

For each species, we simulated 500 occupancy data sets (Kery 2010) to estimate power to detect trends. In each iteration, we generated a data set from a dynamic occupancy model with equally declining site colonization and site extinction rates. Declines were chosen so overall occupancy would decrease by 2.5% or 5.0% over 10 years. As an example, the power for detecting changes in western skink numbers is given in the following table:

Table 4. EUSK power simulation results (500 simulations) for Model m1 (SRS allocation) and Model m1.2 (Proportional and Optimal allocation), conducted at $\alpha=0.10$ for 4 survey periods.

% Annual Decline	Years	N	Allocation	Power
2.5 ¹	10	80	SRS	0.428
5 ²	10	80	SRS	0.792
2.5	10	80	Proportional ³	0.534
5	10	80	Proportional	0.842
2.5	10	80	Optimal ⁴	0.490
5	10	80	Optimal	0.790

¹Average absolute decline = $\Psi_{Year,1} - \Psi_{Year,10} = 0.6744 - 0.5358 = 0.1386$

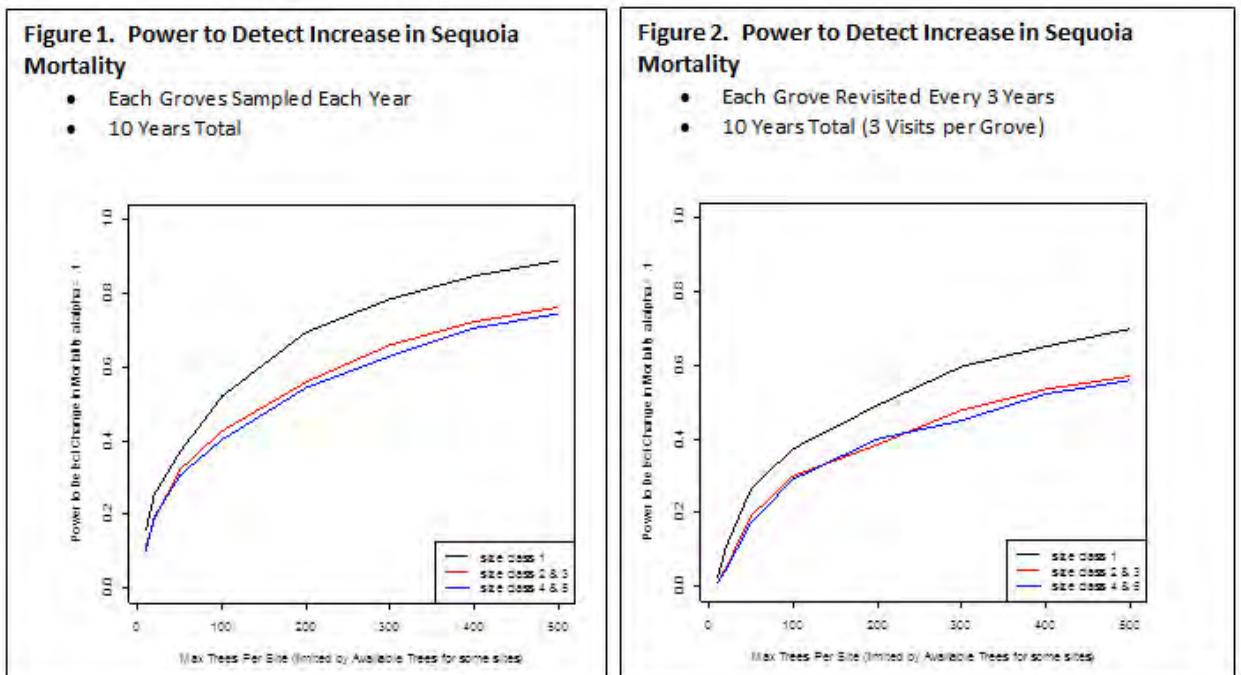
²Average absolute decline = $\Psi_{Year,1} - \Psi_{Year,10} = 0.6744 - 0.4248 = 0.2496$

³Proportional allocation sample sizes: $n_{chacoarral}=39, n_{cscs}=22, n_{grass}=7, n_{woodland}=12$

⁴Optimal allocation sample sizes: $n_{chacoarral}=35, n_{cscs}=19, n_{grass}=3, n_{woodland}=23$

D. Detecting Mortality Increases over Time in Mature Sequoia Trees: Power Analysis Comparison of Various Sampling Protocols, Matt Nahorniak, October 2013

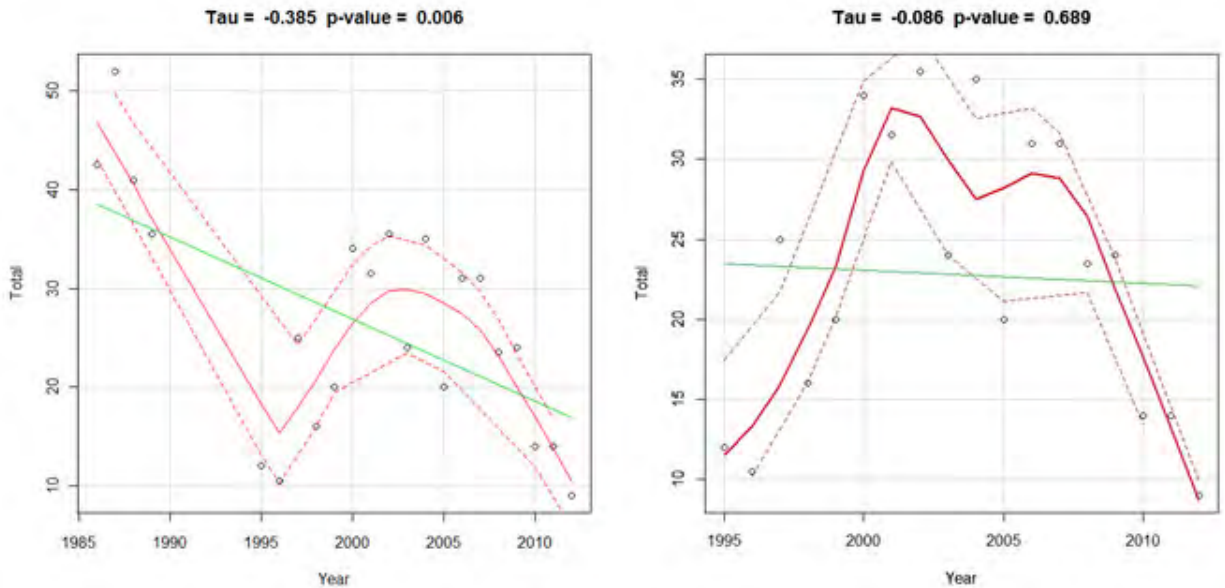
Data on sequoia mortality is not readily available, but a zero-inflated negative binomial model was fit to mortality data for a related species collected from 1982 to 2008. Assuming that sequoia mortality is about a tenth of the modeled mortality, sequoia mortality was simulated for sequoia groves where numbers of trees of various size classes have been recorded. There are 25 groves varying in size from 772 ha to 4 ha where tree numbers by size class have been recorded. Assuming that each grove is sampled using varying revisit designs, then the question remains of how many trees per size class per grove should be sampled in what revisit plan. The power to detect an increase of 8% mortality per year (i.e. a doubling of mortality in 10 years) is given below for two revisit designs.



The three lines on each graph correspond to trees of various size classes. The horizontal axis is the maximum (subject to availability) number of trees to sample per size class per grove.

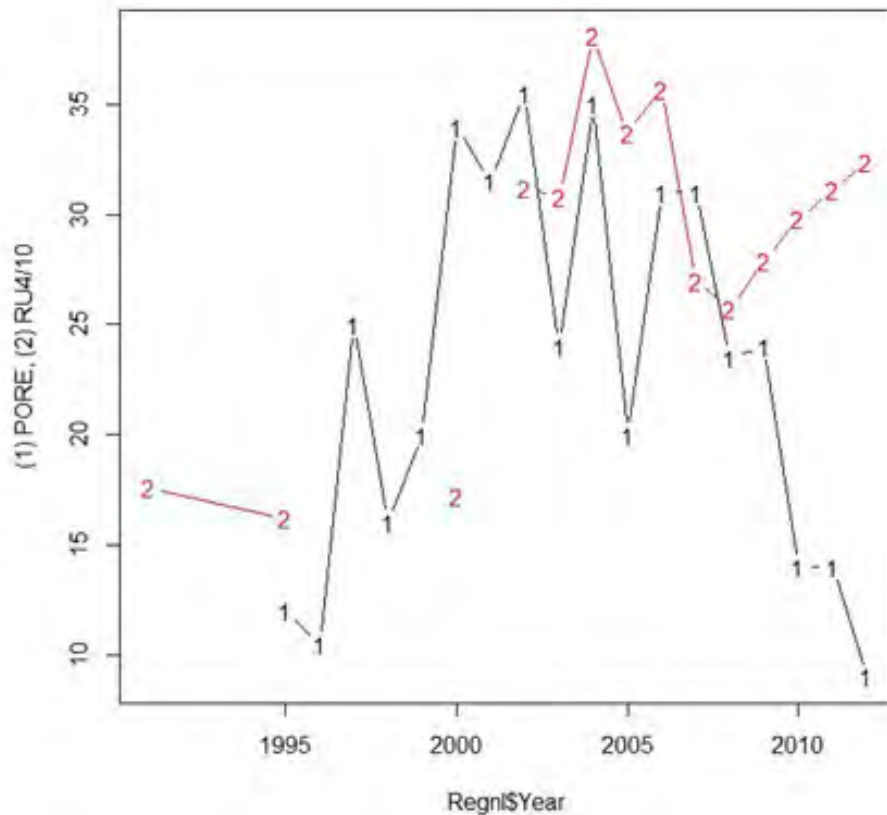
E. Trend Analysis for Snowy Plover at Point Reyes (SFAN) Kirk Steinhorst, August 2013

Data are available for 1986-9 and then from 1995 to 2012. Parameters recorded are winter counts, numbers of nests, nest success, egg numbers, percent hatch, percent fledged, and fledged per egg. The protocol outlines graphical and nonparametric analyses of these variables over years. The data are somewhat like the discharge data described in B. above. You obtain only one number per variable per year. Except for nest counts, there is only one location associated with each time series—Point Reyes National Seashore. With 23 years of data (with a gap after the first 4 years), graphical trends along with Mann-Kendall correlations constitute the appropriate analysis. For example, the winter counts appear to have a decreasing linear trend when considered over all years, but a quadratic trend when only data from 1995 to 2012 are used.



Park biologists are now challenged to explain why plover numbers were higher in the 1980s, were low in the mid-1990s, increased in the early 2000s, and are down again in recent years. One suggestion they made was to compare Point Reyes numbers with regional numbers.

Here is a plot of Point Reyes and regional numbers (from recovery unit 4) graphed on the same scale.



It is interesting to note that plover numbers at Point Reyes track plover numbers in the region up to the last three years. We should ask if the last three counts are cause for concern or just a local aberration.

Discussion

With such a wide range of networks and scientists to work with, we had a rich panoply of statistical issues to deal with. We took each statistical project on its own merits and considered the unique aspects of each. By necessity, we divided up the work to make it more efficient for us to respond to all the requests for service. Leigh Ann became the expert on the Piepho and Ogutu trend model and GRTS sampling. Steve dealt mostly with designs and models for monitoring animal populations. Matt was the “free-agent” who dealt with unique problems. On the other hand, all of us got to see all of the statistical approaches to some degree.

As this TA comes to a close, we should ask if there are lessons to be learned. The answer, of course, is yes. Listed in no particular order:

- How were parameters chosen for detailed analysis when there are multiple parameters to choose from? The answer is that we let the NPS scientist pick the parameters of interest and then built designs and power analyses for all parameters for which there was sufficient data.
- What was the most recurring problem faced? For scientists with modest quantitative training, we had difficulty getting them to give us data in useful forms. We spent a lot of time formatting and rearranging data. That time could have been better spent on statistical analysis. The takeaway message is that we eventually need data in a “row and column” format where rows are observations and columns are variables (or parameters to nonstatisticians). Each row represents a quadrat or a stream reach or a transect or ... Columns are site, date, pH, temperature, etc. or site, date, species 1, species 2, etc.
- While there are a number of scientists with quite a good understanding of the R statistical package, there is a great deal of interest in additional training in the use of R among other scientists. We put on two R workshops—one for SFAN and one for KLMN.
- Large-sample approximations to power can have larger-than-nominal test size and produce inflated power approximations compared to power assessed with a Monte Carlo power simulation when sample sizes are modest.
- Several networks wanted originally to assess the power to detect whether or not an indicator was below/above a threshold value. Determining the threshold value proved to be the more difficult task. Hopefully, continued consistent sampling will provide the basis for determining meaningful threshold values for hypothesis testing.
- The issue of including legacy sites in long-term monitoring designs proved to be a thorny issue. As noted on page 1 above, the use of legacy sites diminishes one’s ability to generalize to a park or network as a whole.
- The conventional wisdom that returning to a set of randomly chosen sites year after year after year is a very efficient way to detect trend is false. In general, it is good to have a panel of sites which are visited every year for a few years to establish a baseline and otherwise one wants to spread the monitoring effort among several panels. Putting all your eggs into a few sites that are sampled intensively over years is not the best plan for determining status either.
- Having the statistician write the data analysis and power SOPs is tempting, but may not be optimal for two reasons: 1) the NPS scientists become more involved and aware when they provide the lead in SOP writing, and 2) there is more time for analysis if the statistician does not spend a great deal of time writing. Collaborative writing with NPS scientists taking the lead seems the best.

References

- Fiske, I. and R. Chandler 2011. unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software*, 43(10), 1-23.
URL <http://www.jstatsoft.org/v43/i10/>.
- Horvitz, D.G. and D.J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663-685.
- Kery, M. 2010. Introduction to WinBUGS for Ecologists: A Bayesian approach to regression, ANOVA, mixed models, and related analyses. Academic Press. Burlington, MA. 302 pp.
- MacKenzie, D., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, J. E. Hines. 2006. Occupancy Estimation and Modeling. Academic Press, New York. 324 pp.
- McDonald, T.L. 2003. Review of environmental monitoring methods: survey designs. *Environmental Monitoring and Assessment*, 85: 277-292.
- Piepho, H-P. and J.O.Ogut. 2002. A simple mixed model for trend analysis in wildlife populations. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(3):350-360.