

APPLICATION

rSPACE: Spatially based power analysis for conservation and ecology

Martha M. Ellis¹, Jacob S. Ivan², Jody M. Tucker³ and Michael K. Schwartz⁴

¹U.S.D.A. Forest Service, Rocky Mountain Research Station, Montana State University Campus, Bozeman, MT 59715, USA;

²Colorado Parks and Wildlife, Wildlife Research Center, Fort Collins, CO 80526, USA; ³U.S.D.A. Forest Service, Pacific Southwest Region, Sequoia National Forest, Porterville, CA 93257, USA; and ⁴U.S.D.A. Forest Service, Rocky Mountain Research Station, Missoula, MT 59801, USA

Summary

1. Power analysis is an important step in designing effective monitoring programs to detect trends in plant or animal populations. Although project goals often focus on detecting changes in population abundance, logistical constraints may require data collection on population indices, such as detection/non-detection data for occupancy estimation.

2. We describe the open-source R package, rSPACE, for implementing a spatially based power analysis for designing monitoring programs. This method incorporates information on species biology and habitat to parameterize a spatially explicit population simulation. A sampling design can then be implemented to create replicate encounter histories which are subsampled and analysed to estimate the power of the monitoring program to detect changes in population abundance over time, using occupancy as a surrogate.

3. The proposed method and software are demonstrated with an analysis of wolverine monitoring in a U.S. Northern Rocky Mountain landscape.

4. The package will be of use to ecologists interested in evaluating objectives and performance of monitoring programs.

Key-words: detection probability, occupancy estimation, population monitoring, population trends, power analysis, sampling design, spatial simulation

Introduction

Monitoring change in population abundance is often a primary goal for conservation and management programs. The success of monitoring projects requires careful planning, including well-articulated and realistic objectives (Yoccoz, Nichols & Boulenger 2001; Nichols & Williams 2006). Power analyses help ensure that project goals are reasonable and evaluate trade-offs in sampling effort (Field, Tyre & Possingham 2005; Rhodes *et al.* 2006).

Evaluating monitoring program design can be complicated. Traditional power analyses require knowledge of process variation and measurement error that are often unavailable (Gibbs, Droege & Eagle 1998; Morrison 2007). Contemporary analyses are often so complex, computation of power necessitates simulation-based approaches (Robert & Casella 2004). Furthermore, defining an effect size can be complicated when surrogate measures are used in place of abundance. Detection/non-detection data for occupancy estimation are often used in this context (Joseph *et al.* 2006). The relationship between occupancy and abundance depends on density of individuals

and may vary across the landscape with the spatial distribution of resources (Gaston *et al.* 2000).

Standard occupancy-based power analyses are non-spatial and thus assume uniform distribution of individuals across the landscape (e.g. Bailey *et al.* 2007; Guillera-Arroita, Ridout & Morgan 2010; Guillera-Arroita & Lahoz-Monfort 2012), which is unrealistic of natural populations. Additionally, such analyses typically evaluate the ability to detect a trend in occupancy alone without investigating the underlying relationship between occupancy and abundance. To this end, we have designed a framework to facilitate spatially based power analyses for population monitoring. Our approach is unique in that it (1) incorporates available spatial information on habitat and species biology, (2) estimates power to detect trend in abundance using occupancy as a surrogate metric and (3) uses data readily obtainable by most scientists and managers.

We have created the package, rSPACE, for the R statistical environment (R Development Core Team 2014) to implement our approach. The program is designed to provide a flexible shell in which alternative approaches for a replicated population simulation, sampling design and analysis can be customized to match a given scenario. The current implementation focuses on occupancy-based monitoring of territorial

*Correspondence author. E-mail: martha.ellis@gmail.com

carnivores. We anticipate providing additional options and templates for designs and analyses as these are developed.

Below we describe the background and work flow of rSPACE (version 1.1; Ellis *et al.* 2015) and illustrate its use with an example for wolverine in the Northern U.S. Rocky Mountains.

Methods

POPULATION SIMULATION

A raster layer describing habitat suitability for the species and landscape of interest provides the basis for rSPACE simulations. Habitat suitability values range from 0 to 1 and reflect the relative probability of use on the landscape. Habitat suitability indices can range from a categorical habitat/non-habitat assessment to continuous quantitative information from previously studied habitat relationships.

Initially, rSPACE creates a spatially based simulated population from which to sample. Main parameters include initial population size, buffer distances between individual activity centres (often treated as the average distance between home range centres) and the minimum habitat suitability value at which activity centres can occur. Multiple types of individuals (e.g. juveniles, adults, males, females) in the population are allowed; each type of individual is distributed on the landscape independently of other types. For example, for wolverine in the Northern U.S. Rocky Mountains, the locations for female activity centres must be at least 16 km apart, reflecting average female home range sizes, but do not depend on locations of males.

Finally, movement parameters for individuals are defined to represent the expected movement of individuals during the sampling season (i.e. the period of time between the beginning of the first sampling occasion and the end of the last sampling occasion, where occasions are independent periods of time during which detection/non-detection data are collected at each sampling unit). These parameters are used to create a bivariate normal movement distribution which is then modified by the underlying habitat suitability layer. Thus, each individual is assigned a unique, centre-weighted, movement distribution during the sampling season. Long-distance movements during the sampling season can greatly influence population estimation results; therefore, rSPACE allows an option to truncate the tails of the movement distribution.

EFFECT SIZE

The effect size in rSPACE is defined by the magnitude of the trend in population abundance. For example, a monitoring program may be focused on detecting a 20% decline in abundance over a 10-year period. The input value for this trend parameter would be a population growth rate (λ) of 0.977, following a standard exponential growth equation. This effect size is implemented in rSPACE by adding or removing the appropriate number of individuals from the simulation in each year to achieve the specified population growth rate.

IMPLEMENTING A MONITORING SCENARIO

Each replicate in rSPACE starts with a random distribution of individual activity centres according to habitat suitability, population size and territoriality rules (Fig. 1a,b); in each simulation, a movement distribution is developed for every individual during the sampling season (Fig. 1c). The landscape is divided into a rectangular grid, and the probability of at least one individual from the population using a cell

during the sampling season is calculated as $U_j = 1 - \prod_{i=1}^N (1 - U_{ij})$ where U_{ij} is the probability of use for individual i in cell j , computed by integrating over the portion of movement distribution i that overlaps cell j (Fig. 1d). A complete encounter history for each cell, assuming perfect detection, is created using a Bernoulli random draw with probability U_j (1 = used, 0 = unused) for each occasion within the sampling season. Each simulation year, individuals are randomly added or removed to the population to achieve the desired λ . The entire process, starting with the random distribution of activity centres, repeats for each replicate simulation. Thus, in each replicate, a complete encounter history is created for each cell for the maximum number of years and occasions per year, with detection probability (p_{sim}) set to 1.0 (i.e. if a cell is used during an occasion, the species always detected).

Due to the large number of parameter inputs, parameters are entered to all rSPACE functions via a parameter list (see Table 1). The `encounter.history` function provides error checking on the steps used to create each replicate encounter history, and `createReplicates` produces replicated complete encounter histories for the landscape over time.

ANALYSIS

The first analysis step in rSPACE is to subset each complete replicate encounter history file to create 'observed' encounter histories with different sampling effort. Subsetting affects the number of cells sampled, the number of sampling occasions per season, the detection probability for each occasion and possible alternative model formulations. By default, rSPACE applies the following subsets: number of cells ranges from 10% to 90% of the grid (currently implemented as a simple random sample), number of occasions per season ranges from 2 to maximum specified by user, detection probabilities (per occasion) of 0.2 and 0.8 and sampling every year vs. every other year.

In `testReplicates`, users provide the folder location containing the complete encounter history replicates. Each complete encounter history file produces a set of observed encounter histories based on subsetting routines. The `function_name` argument in `testReplicates` supplies an analysis to run on each observed encounter history. The default (`function_name = "wolverine_analysis"`) is a robust design (multiseason) occupancy model as applied in Ellis, Ivan & Schwartz (2014). Details for this function, which provides a template for customized analyses, are included with rSPACE documentation. For each replicate, occupancy estimates and variance-covariance matrices are extracted and then a linear trend is fit with a generalized linear model. If the $(1-\alpha)\%$ confidence interval on the trend parameter is different from zero, then a trend is detected.

`testReplicates` produces a results file, labelled 'sim_results.txt' by default. Each line contains the analysis results for an observed encounter history, with columns for the input encounter history filename, subsetting parameters and results of the analysis function. As an end result, power in rSPACE is calculated as the number of times a trend was detected at a given significance level divided by the total number of replicates.

EXAMPLE: WOLVERINE IN NORTHERN U.S. ROCKIES

To demonstrate the application of rSPACE, we present a simplified analysis of wolverine monitoring in the Northern U.S. Rocky Mountains (Ellis, Ivan & Schwartz 2014). Full code for this example is available in Appendix S1.

The raster habitat layer for wolverine describes areas with and without persistent spring snow (Copeland *et al.* 2010). To reduce computation time, we have reduced the study landscape to the Bitterroot

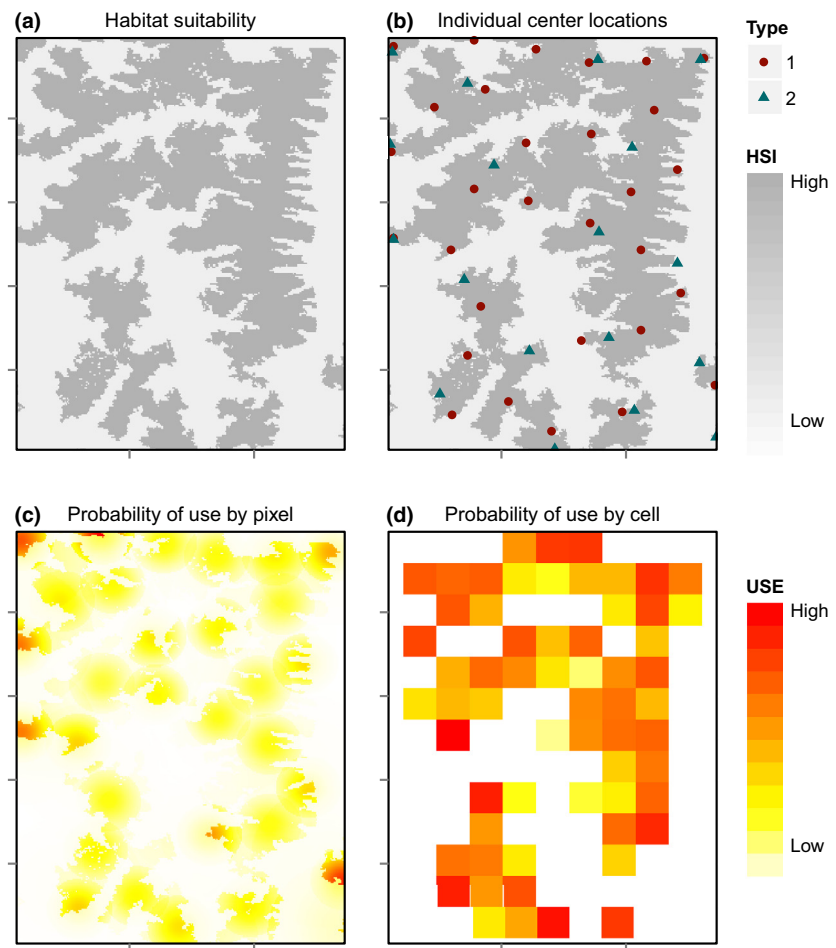


Fig. 1. Steps to building a simulated encounter history: input habitat suitability layer (a), distribute individual activity centres according to habitat suitability and territorial rules (b), build movement distribution for each individual (c), grid landscape and calculate probability of ≥ 1 individual using each sample unit (d). To simulate trend, individuals are added or removed to (b) for each year, then repeats steps (c) and (d). Each replicate simulation starts with a different arrangement of individuals in (b).

Mountain Range along the Montana/Idaho border. These data are included in rSPACE (see Appendix S1).

Parameter estimates for territoriality and movement in this example were derived from the literature and expert opinion (Table 1). We used a 100 km² grid size and tested power over a range of sampling intensities for both number of cells sampled (10–90% of the grid included) and number of occasions (3–6 per year). The `enter.parameters` function opens a dialogue box to assist with parameter list construction:

```
Base_parameters<-enter.parameters()
```

For this example, we considered a 2 × 2 factorial design of population size and growth rate scenarios, with population sizes of $N = 25$ or $N = 50$ individuals and either a 50% decline or increase in population size over a 10-year period (i.e. $\lambda = 0.933$ or $\lambda = 1.041$). For each scenario, we edit the parameter list and use `createReplicates` to produce a set of 100 replicate encounter histories for the landscape. Our first scenario ($N = 25$, $\lambda = 0.933$) would be run as follows:

```
Plist_Scenario1<-Base_parameters
Plist_Scenario1$N<-25
Plist_Scenario1$lambda<-0.933
createReplicates(n_runs=100,map=WolverineHabitat,
Parameters=Plist_Scenario1)
```

`createReplicates` creates a folder to house the replicate encounter histories. Once these files are created, `testReplicates`

can be run by providing the directory for the rSPACE scenario and the parameter list for that scenario:

```
testReplicates(folder="./rSPACE_X",
Parameters=Plist_Scenario1)
```

The default analysis function in `testReplicates` uses a robust design occupancy model, run through the RMark/Program MARK interface (White & Burnham 1999; Laake & Rexstad 2007). For details, see `help(wolverine_analysis)`.

The results for each scenario can be summarized using the functions, `getResults` or `findPower`. In this example, we used a significance level of $\alpha = 0.1$ (type I error rate of 10%). We were interested in how many cells would need to be sampled to obtain 80% power (type II error rate of 20%). The results from our first scenario ($N = 50$, $\lambda = 0.933$; Fig. 2) were obtained using `getResults`, with argument `CI = 0.9`. Under this scenario, with 5 sampling occasions per year, we would need to sample 30 cells to have 80% power at the $\alpha = 0.1$ significance level.

```
getResults(folder="./rSPACE_X",CI=0.9)

findPower(folder="./rSPACE_X",CI=0.9,pwr=0.8)

n_visits  n_grid
3 55.24128
4 35.08446
5 29.30485
6 21.75605
```

Table 1. Input parameter labels and descriptions

	Label	Description	Example values (Scenario 1)
Population simulation	N	Initial population size	50
	lmda	Annual population multiplication rate (λ)	0.933
	n_yrs	Maximum number of years in simulation	10
	MFratio	Proportion of individuals in population (number of types inferred by length)	c(0.6, 0.4)
Movement parameters	buffer	Distance between activity centres (km) for each type of individual.	c(16, 25)
	moveDist	Movement radius (km). In combination with moveDistQ, parameters describing expected movement during the sampling season. Used to determine standard deviation of bivariate normal movement distribution	c(8.5, 12.5)
	moveDistQ	Proportion of movements within moveDist radius (e.g. 0.5 if moveDist represents median movement distance)	c(.9, .7)
	maxDistQ	Truncation for long-distance movements during sampling season. Quantile/proportion of movement distribution to include	c(.95, .95)
Sampling design	grid_size	Area of each grid cell (km ²)	100
	habitat.cutoff	Minimum habitat value where individual activity centres should be located	1
	sample.cutoff	(Optional) Minimum proportion of pixels in cell above habitat.cutoff to include in sampling frame	0.5
	n_visits	Maximum number of sampling occasions during the sampling season each year	6
Subsetting parameters	n_visit_test	(Optional) Number of occasions per year to test	c(3, 4, 5, 6)
	detP_test	(Optional) Per occasion detection probability	c(0.8)
	grid_sample	(Optional) Proportion of grid to sample	seq(0.05, 0.95, by = .1)
	alt_model	(Optional) Alternative model specifications	0

Comparisons among scenarios are valuable for understanding the effect of different parameterizations on the expected power of a monitoring plan. For wolverine, population sizes and growth rates are unknown prior to monitoring, so testing power under a range of possible options was appropriate (Fig. 3).

Discussion

Power analyses assessing the ability to detect trend for a given sampling strategy are critical in designing effective population monitoring programs. Often efforts designed to assess trend were created with inadequate power to ever have been able to detect that trend. Our goal with rSPACE was to provide a simple tool to evaluate the amount of effort required for a survey to estimate statistical trend.

Ideally monitoring programs would measure changes in abundance over time, but in practice, estimating abundance directly is often logistically or financially unfeasible. Consequently, monitoring programs frequently rely on indirect indices such as occupancy to assess trend. The relationship between occupancy and abundance varies depending on the spatial distribution and density of individuals on the landscape. Therefore, the ability to account for this spatial variation is an important but frequently overlooked aspect in conducting power analyses. By incorporating spatial data about the distribution of habitat along with the biological characteristics of the species in question, rSPACE allows for power analyses that accounts for the spatial variation characteristic of wild populations.

In the initial release of rSPACE, we have tried to balance simplicity for which parameter space can be explored vs.

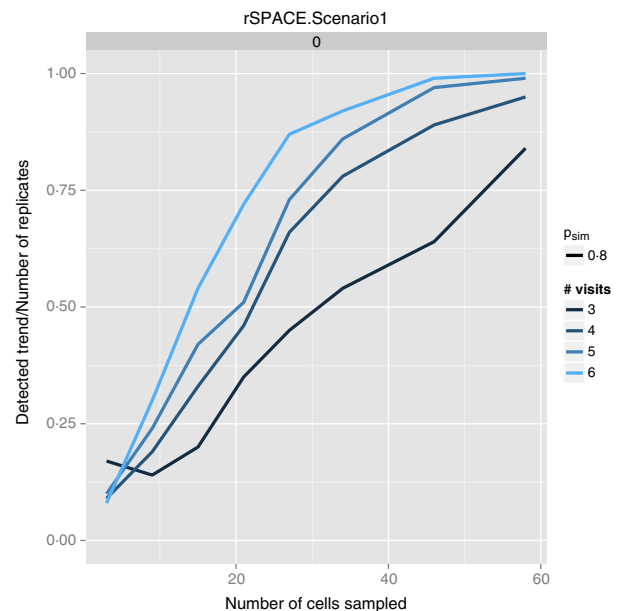


Fig. 2. Example output figure from `getResults`, using Scenario 1 of the wolverine example with $CI = 0.9$. # visits indicates the number of sampling occasions, and p_{sim} indicates per visit detection probability given that a cell was used by ≥ 1 wolverine.

complexity in matching specific scenarios. In addition to options currently available through the rSPACE package on CRAN, we have an active development website (<http://github.com/mmellis/rSPACE>) with example scripts demonstrating alternative uses. We welcome feedback on both current implementations and future suggestions. Through this effort,

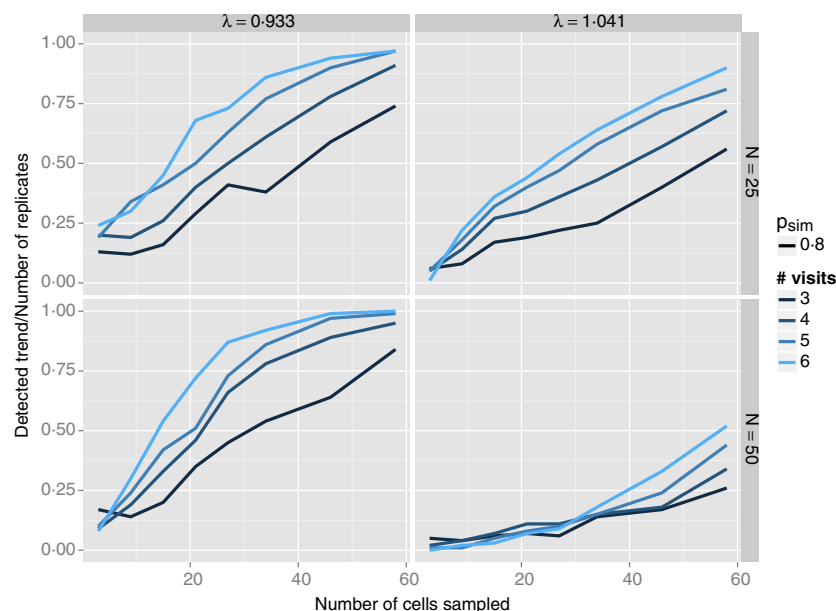


Fig. 3. Comparison of population scenarios for wolverine example. Scenarios include two initial population sizes of $N = 25$ and $N = 50$, and population trends $\lambda = 0.933$ (20% decline) and $\lambda = 1.041$ (20% increase). # visits indicates the number of sampling occasions and p_{sim} indicates per visit detection probability given that a cell is used by ≥ 1 wolverine. Code to produce this figure is included in Appendix S1.

we hope to foster careful monitoring design and encourage multijurisdictional collaboration to enable large-scale monitoring efforts.

Acknowledgements

Initial funding for this project was provided by USFS, the RMRS, a PECASE award to M.K.S., and a Section 6 grant from USFWS Region 6 to Colorado Parks and Wildlife. We thank J. Laake, P. Lukacs and G. White for invaluable technical advice, and we are grateful to M. Alldredge, D. Tripp and two anonymous reviewers whose comments on earlier drafts greatly improved this article.

Data accessibility

All data used in this manuscript are included in the R package: <http://CRAN.R-project.org/package=rSPACE>

References

- Bailey, L.L., Hines, J.E., Nichols, J.D. & MacKenzie, D.I. (2007) Sampling design trade-offs in occupancy studies with imperfect detection: examples and software. *Ecological Applications*, **17**, 281–290.
- Copeland, J., McKelvey, K., Aubry, K., Landa, A., Persson, J., Inman, R. *et al.* (2010) The bioclimatic envelope of the wolverine (*Gulo gulo*): do climatic constraints limit its geographic distribution? *Canadian Journal of Zoology*, **88**, 233–246.
- Ellis, M.M., Ivan, J.S. & Schwartz, M.K. (2014) Spatially explicit power analyses for occupancy-based monitoring of Wolverine in the US Rocky Mountains. *Conservation Biology*, **28**, 52–62.
- Ellis, M., Ivan, J., Tucker, J. & Schwartz, M. (2015) rSPACE: Spatially-Explicit Power Analysis for Conservation and Ecology. <http://CRAN.R-project.org/package=rSPACE>.
- Field, S.A., Tyre, A.J. & Possingham, H.P. (2005) Optimizing allocation of monitoring effort under economic and observational constraints. *Journal of Wildlife Management*, **69**, 473–482.
- Gaston, K.J., Blackburn, T.M., Greenwood, J.J.D., Gregory, R.D., Quinn, R.M. & Lawton, J.H. (2000) Abundance–occupancy relationships. *Journal of Applied Ecology*, **37**, 39–59.
- Gibbs, J.P., Droege, S. & Eagle, P. (1998) Monitoring populations of plants and animals. *BioScience*, **48**, 935–940.

Guillera-Arroita, G. & Lahoz-Monfort, J.J. (2012) Designing studies to detect differences in species occupancy: power analysis under imperfect detection. *Methods in Ecology and Evolution*, **3**, 860–869.

Guillera-Arroita, G., Ridout, M.S. & Morgan, B.J.T. (2010) Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, **1**, 131–139.

Joseph, L.N., Field, S.A., Wilcox, C. & Possingham, H.P. (2006) Presence–absence versus abundance data for monitoring threatened species. *Conservation Biology*, **20**, 1679–1687.

Laake, J. & Rexstad, E. (2007) RMark—an alternative approach to building linear models in MARK. *Program MARK: A Gentle Introduction* (eds E. Cooch & G.C. White). Appendix C, C1–C113. <http://www.phidot.org/software/mark/docs/book/>

Morrison, L.W. (2007) Assessing the reliability of ecological monitoring data: power analysis and alternative approaches. *Natural Areas Journal*, **27**, 83–91.

Nichols, J.D. & Williams, B.K. (2006) Monitoring for conservation. *Trends in Ecology & Evolution*, **21**, 668–673.

R Development Core Team. 2014. R: A language and environment for statistical computing. <http://www.R-project.org/>

Rhodes, J.R., Tyre, A.J., Jonzén, N., McALPINE, C.A. & Possingham, H.P. (2006) Optimizing presence-absence surveys for detecting population trends. *Journal of Wildlife Management*, **70**, 8–18.

Robert, C.P. & Casella, G. (2004) *Monte Carlo Statistical Methods*, 2nd edn. Springer, New York.

White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, S120–S139.

Yoccoz, N.G., Nichols, J.D. & Boulenger, T. (2001) Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, **16**, 446–453.

Received 15 December 2014; accepted 5 March 2015

Handling Editor: Timothée Poisot

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Example R script to demonstrate power analysis for Wolverine.