GENPRES

by

Jim Hines
USGS, Patuxent Wildlife Research Center
Laurel, MD, 20708, USA
email:jhines@usgs.gov
www.mbr-pwrc.usgs.gov

This file last modified: <29 July 2019>

This program simulates presence/absence data to be input to programs Presence or PRESENCE. It can be used to get an idea of how precise the estimates are for given sample effort or design, or the bias of estimates when heterogeneity exists. See the following papers for a description of the methods involved in estimating parameters from presence/absence data:

[1]MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83: 2248-2255

[2]MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson and A. B. Franklin. 2003. Estimating site occupancy, colonization and local extinction probabilities when a species is not detected with certainty. Ecology 84: 2200-2207

[3]Bailey LL, Hines JE, Nichols JD, MacKenzie DI (2007) Sampling Design Trade-offs in Occupancy Studies with Imperfect Detection: Examples and Software. Ecological Applications: Vol. 17, No. 1 pp. 281–290

Definitions:

Single,Multi-season models

PSI : Initial occupancy rate (proportion of sites occupied)
$P(i)$: detection probability for survey i
$P10(i)$: probability of a 'false' detection for survey i
$B(i)$: probability a detection is known for survey i
$EPS(i)$: probability of species extinction from survey i to $i+1$ (= 1-PHI)
$PHI(i)$: probability of species survival from survey i to $i+1$ (= 1-EPS)
$GAMMA(i)$: probability colonization just after survey i

Two-species models

PSI-A : Initial occupancy rate for species A (regardless of occupancy of species B)
PSI-B1 : Initial occupancy rate for species B, given occupancy of species A
PSI-B2 : Initial occupancy rate for species B, given non-occupancy of species A
pA : detection probability for species A, given only A present
pB : detection probability for species B, given only B present
rA : prob. detect only species A, given both species present
rB1 : prob. detect species B, given both species present and species A also detected
rB2 : prob. detect species B, given both species present and species A not detected

Multi-method models

THETA(i) : probability that site i is locally occupied in survey$i$, given site i-1 is not locally occupied
THETA'(i) : probability that site i is locally occupied in survey$i$, given site i-1 is locally occupied

Multi-state, single-season models

PSI1 : Initial occupancy rate (proportion of sites occupied)
PSI2 : Prob. site is in state 2, given occupancy
p1 : Detection prob., given site is in state 1
p2 : Detection prob., given site is in state 2
dlta : Prob. of identifying site as belonging to state 2, given it's in state 2

Multi-state, multi-season models

Psi : Vector of initial occupancy rates (indexed by state)
psi(i-j): Vector of subsequent occupancy rates (indexed by prev. state, subsequent state)
p(i-j) : Vector of detection probs (indexed by detected state, true state)

Multi-state, multi-season models(R,dlta parameterization)

Psi0 : Initial occupancy rate
R0 : Initial prob. of being in state 2
Cpsi0 : Vector of subsequent conditional occupancy rates (Pr(occ|prev.state=0)
Cpsi1 : Vector of subsequent conditional occupancy rates (Pr(occ|prev.state=1)
Cpsi2 : Vector of subsequent conditional occupancy rates (Pr(occ|prev.state=2)
CR0 : Vector of subsequent conditional occ2 rates (Pr(state2|prev.state=0)
CR1 : Vector of subsequent conditional occ2 rates (Pr(state2|prev.state=1)
CR2 : Vector of subsequent conditional occ2 rates (Pr(state2|prev.state=2)
p1 : Vector of detection probs, given site is in state 1
p2 : Vector of detection probs, given site is in state 2
dlta : Prob. of identifying site as belonging to state 2, given it's in state 2

Royle-Nichols models

Lambda : Population size
p : Species detection probability

Survey Design

Occupancy studies fall into two categories: (1) single-season, or (2) multi-season. In the single-season study, sites are surveyed multiple times over a short period of time to estimate the proportion of sites which are occupied and detection probability. In the multiple-season study, sites are surveyed multiple times in two or more seasons, where there is an interval of time between seasons for changes in occupancy to occur. These changes in occupancy are reflected by the values of GAMMA (species colonization rate) and EPS (species extinction rate).

Both survey designs can be implemented by *GENPRES*. Default input values are provided for the multi-season design, and the single-season design can be implemented by changing all values of EPS to 0.0 and all values of GAMMA to 0.0. In fact, the program determines which surveys belong to which season by examining the values of EPS. Anytime the value of EPS is 0.0, the succeeding survey is in the same season as the preceeding survey. When EPS is greater than 0.0, the succeeding survey is the first survey of a new season.

Modeling heterogeneity

This program can be used to examine the effects of heterogeneity on the estimates of occupancy, detection, or change in occupancy (EPS, GAMMA). Although not all individuals are expected to have exactly the same probability of occupying a site, this probability is assumed to be approximately equal for all individuals. If a sub-group of the sampled population have a substantially different probability of occupying the surveyed sites than the rest of the population, then the population is said to be heterogeneous. If the sub-group can be identified when observed, then there is no problem. Each sub-group would be analyzed separately. If there is no way of identifying which group the observation belongs to, then the overall estimate of occupancy will be biased.

This situation can be easily modeled in the program. Simply specify the parameters for one of the sub-groups, then click the 'Add Group' button and enter the values for the other group.

Standard design vs *Panel* design

'Standard' design refers to the situation where each site is surveyed each season. In a 'panel' design, some sites are visited in some seasons, but not in others. The panel design might be used to cover a larger area at a smaller cost, or may be the result when a group of sites become inaccessible.

This situation can be handled by setting the detection probability to zero for a group of sites in a particular season(s). When data are generated, these sites will contain a '.' corresponding to the surveys when they were not visited.

Example:Distribution of sampling effort across 4 bi-weekly survey periods

```
Design 1                      Design 2                      Design 3
_____           _____           _____

Num                           Num                           Num
of                            of                            of
Sites   1   2   3   4         Sites   1   2   3   4         Sites   1   2   3   4
12      xx  --  --  --        6       xx  xx  xx  xx        6       xx  --  xx  --
12      --  xx  --  --        6       xx  --  --  --        6       --  xx  --  xx
12      --  --  xx  --        6       --  xx  --  --        6       xx  --  --  --
12      --  --  --  xx        6       --  --  xx  --        6       --  xx  --  --
                              6       --  --  --  xx        6       --  --  xx  --
                                                            6       --  --  --  xx

Total    s=48 sites                   s=30 sites                    s=36 sites
```

Program installation

Download and unzip the Windows setup program (setup_presence.zip), then execute setup_presence.exe from the zipfile. Data can be generated and analyzed with this program.

Program operation

Once the program starts, a tabbed window appears with default values for each of the parameters (PSI,P(i)). Values can be changed by clicking on the value and typing in a new value (duh!). Changing the number of surveys adds or deletes columns for the survey-specific parameters.

The default scenario indicates that there are 100 sites which are visited a total of 5 times. Initial occupancy is 75% and detection probability is 50% for each survey.

To simulate heterogeneity amoung sites, create multiple groups with different occupancy (psi) or detection (p) probabilities. To test design methods (as in the example above), create groups with detection probabilities equal to zero for surveys which are skipped ('--' in the diagram), and the desired detection probability for surveys which were not skipped ('xx' above).

For example, to simulate 'Design 1' above, change the number of surveys to 8 and numbe of sites to 12. Enter the desired occupancy probability, and detection probabilities for surveys 1 and 2. Enter '0.' for the detection probabilities for surveys 3 through 8 Then, add a 2nd group and change the detection probabilites to 0. for surveys 1 and 2. Change the detection probabilities to the desired value for surveys 3 and 4, and leave the rest at 0. Add groups 3 and 4, changing the detection probabilities to 0. for surveys which are skipped for that group.

When the 'Analyze w/ expected values' button is clicked, data will be generated for this situation and analyzed with program *Presence*. The output from program *Presence* will appear in a new window. If you were to look at the input data file, you would see a sequence of '1's and '0's indicating detection (1) or non-detection (0) for each survey. The number following the detection history is the number of sites which had that exact detection history. (Although in the real world there cannot be fractions of a site, the expected number of sites can be a fraction depending on the input values.)

The parameter estimates from program *Presence* appear at the end of the output file (scroll down to the end using "cntl-end").

Running other models

Nine model-types are available in *GENPRES* which are listed in the definitions section above. Select a different type by clicking the 'Model-type menu and selecting the desired type.

There are several models available for each model type. Once a model-type is selected (for generation of data), a specific model must be chosen to analyze the data. This is done by selecting models under the 'Model' menu. You can generate the data by changing the parameter values in the input table, then choose the model to use to compute the estimates. This could be used to investigate the bias of the parameter estimates when data are generated with different values for each occasion, but analyzed with a model assuming constant values over time.

One of the last models in the 'model' menu, 'user-defined' allows you to analyze the generated data with your own customized model (e.g., model PSI(.),p(T), where p(T) indicates that detection probabilities are forced to a linear trend (logit scale) using the design-matrix). Click 'Help' to see a sample of a model using the design matrix.

The menu choices, 'Define model by name' allow you to specify models similar to the ones in the menu by entering a model name and letting GENPRES create the Presence input file based on whether '(.)' or '(t)' appears after each of the parameters, Psi, Gam, Eps, or p.

Options

- *save all Presence output*
- *specify location of Presence* - tells GENPRES where Presence is installed (in case it's not c:\program files\Presence)
- *specify location of tempfiles* - tells GENPRES where to put temp files (default=MyDocuments)
- *specify alt editor* - shows output in another editor, instead of Notepad
- *simulate by individual* - applies rates to individual sites, following 'binary tree' approach
- *specify # bootstraps for PRESENCE GOF*
- *don't compute variances* (for faster simulations where you don't need SE's)

How to run in '*Batch*' mode

The easiest way to run a series of simulations with different input parameters is to run GENPRES from within a script for another computer language (eg., R, Python,...). In the script, use the function to call an external program to run GENPRES. In R, the call would look something like this:

```
cmd="c:/progra~1/presence/genpres8.exe T=5 N=100 psi=.75 p=.1"
i=shell(cmd)
```

Code to generate input for Presence would then need to be added to the script in order to analyze the generated data.

The R package, RPresence (www.mbr-pwrc.usgs.gov/softwre/bin/RPresence.zip"), makes the process easier.

The input options for GENPRES are:

```
usage genpres8 T=? N=? psi=? p=?,?,?... eps=?,?,?... gam=?,?,?...
       other parameters: psiA psiBA psiBa pA pB rA rBA rBa 2SP                  or
                         theta theta' psi0 psi1 psi2 R p1 p2 dlta
                         lambda psiB etaAA etaBA gamAA gamAB gamBA gamBB epsAA epsAB epsBA epsBB pA pB
                         e d p11 p10 b pi psiBA psiBa gamAB gamAb gamBAA gamBAa gamBaA gamBaa epsAB epsAb
                         epsBAA epsBAa epsBaA epsBaa pA pB 2SP MULTISEASON
  where T=number of sampling occasions,
        N=total number of surveyed sites,
        psi=probability species is present at a site,
        p=detection probability
        eps=extinction rate from sample i to i+1,
        gam=prop of extinct sites which colonize between sample i and i+1,
        e=probability that species enters the superpopulation just before sample i,
        d=probability that species departs the superpopulation just after sample i,
        p10=probability that species is detected, given no occupancy (false positive)
        p11=probability that species is detected, given occupancy
        b=probability that species is certain, given detected
        r=probability that an individual of the species is detected (R/N model
        c=probability that an individual of the species is detected (N/H model
     opts: STOCHASTIC, P0MISS, LIST, QUIET NOBS=n SIMTYPE=2
        STOCHASTIC - data simulated instead of expected value data generated,
        P0MISS - missing value in history instead of zero when P=0,
        LIST - list histories on screen,
        QUIET - don't print much on screen,
        NOBS=n - set history frequency = n, instead of 1,
        SIMTYPE=2 - simulate by individual sites, instead of recursively by group.

        SRVYPERSEASN=n - specify number of secondary periods per primary period.

psi0= prob adults of species are present|not present in i-1
psi1= prob adults of species are present|adults present in i-1
psi2= prob adults of species are present|breeders present in i-1
R   = prob breeders of species are present|adults present
p1  = prob of detecting non-breeding adults
p1  = prob of detecting breeding adults
dlta= prob of detecting young with breeding adults

psiAB= prob of both species present                           N
psiA= prob species A present, regardless of species B      / \psiA
psiB= prob species B present, regardless of species A      x   A
psiBA=prob species B present | species A present          / \ / \psiBA
psiBa=prob species B present | species A not present      0 B A   AB
           psiB=psiBa*(1-psiA)+psiAB
pA=prob detect A | only A present                            AB
pB=prob detect B | only B present                          / \rA
rA=prob detect A | both species present                    0   A
rBA=prob detect B | both present & A detected             / \ / \rBA
rBa=prob detect B | both present & A not detected         0 2 1   3

psi(0)= vector of psi for just before occasion 1
psi(1)= matrix of psi for just before occasion 2
p(1)= matrix of p at  occasion 1
psi(i,r,s)= [in state r @ i] [in state s @ i-1]
p(i,r,s)= [cap in r @ i] [true state=s]
```