

Welcome to the Seminar  
**Resource Selection Functions and  
Patch Occupancy Models:  
Similarities and Differences**

**Lyman McDonald  
Senior Biometrician  
WEST, Inc.  
Cheyenne, Wyoming  
and  
Laramie, Wyoming**

**lmcdonald@west-inc.com  
<http://www.west-inc.com>**



**To hear the seminar, dial**

**(605) 772-3434, access code 388-726-214**

1. The seminar is being recorded in Windows Media format.
2. If Successful, the wmv file will be available on our web site early next week for download.
3. If you are on an expensive long distance telephone line, I suggest that you hang up and watch the slides. Download the wmv file next week and listen to the seminar at your leisure.



4. If the slides download too slow, listen to the seminar, then download and play the wmv file next week.

**5. To hear, dial (605) 772-3434, with access code 388-726-214.**

## Ground Rules

1. Panelists. Please set telephone on mute, or limit background noise.
2. Panelists. Hold questions and comments for the discussion period.
3. All. We will try to answer questions and hold discussion during the second hour, however it may not work very well. E-mail questions and comments to me later.
4. Minimize the box in your upper left corner using the tab on the left hand side of the box.



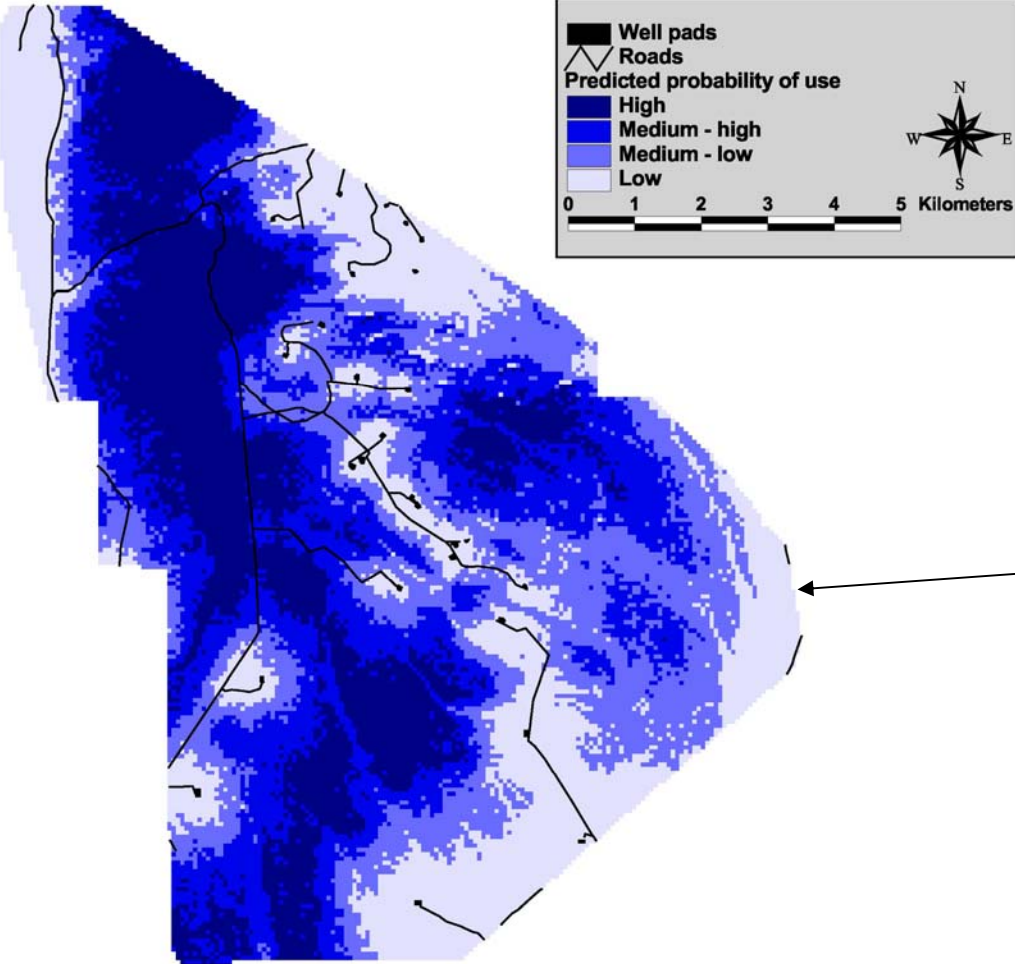
**To hear the seminar, dial**

**(605) 772-3434, access code 388-726-214**

A scenic view of a mountain valley with a river and forested hills. The background shows a wide valley with a river winding through it, surrounded by lush green forests and distant mountains under a clear sky.

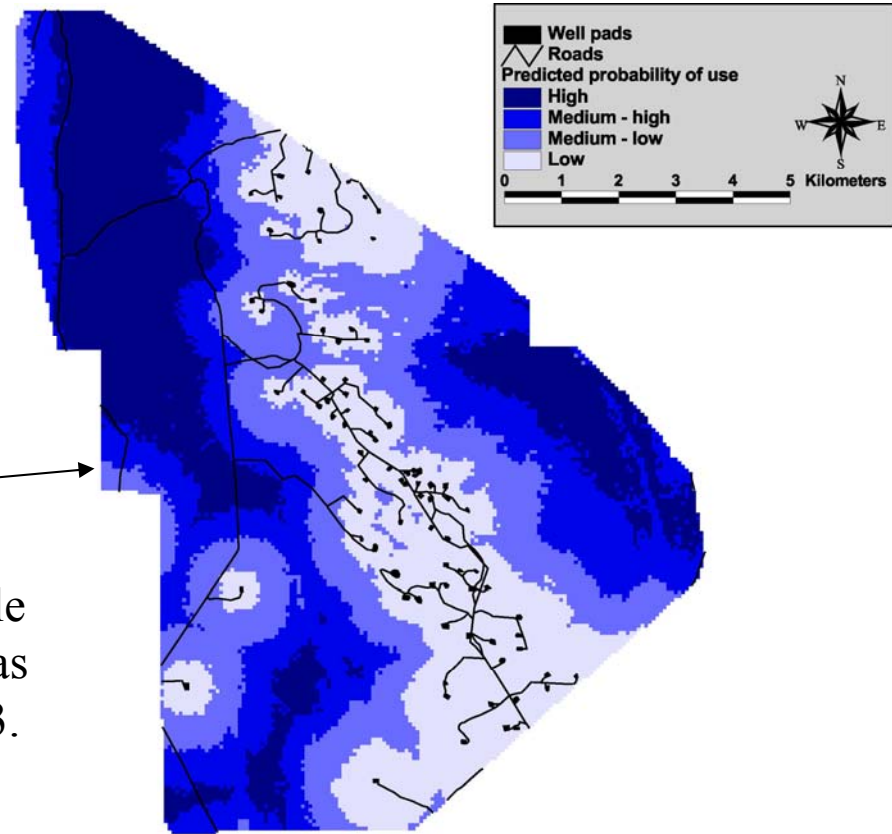
## • **Resource selection functions and patch occupancy models**

- Powerful methods of identifying areas within a landscape that are highly used by a population of plants or animals.
  - It is generally assumed that if individuals select certain habitat units or food resources disproportionately to their availability or ‘patches’ with certain characteristics, it improves their fitness, reproduction, or survival.
  - Justify management actions on natural resources.
  - Monitor distributions of populations.



Estimated relative probability of ‘use and detection of use’ by deer in first year of study of effect of drilling for natural gas.

Third year of study with increased drilling activity.



Sawyer et al. 2006. Winter habitat selection of mule deer before and during development of a natural gas field. *Journal of Wildlife Management* 70: 396-403.

# References

- Manly, B.F.J., L.L. McDonald, D.L. Thomas, T.L. McDonald, and W.P. Erickson. 1996, 2002. *Resource selection by animals: Statistical design and analysis for field studies, Second Edition*. Kluwer Academic Publishers, Dordrecht.
- Journal of Wildlife Management, No. 2, 2006. Papers by
  - Dana Thomas and Eric Taylor
  - Rich Alldredge and James Griswold
  - Chirs Johnson, Scott Nielsen, Eve Merrill, Trent McDonald, and Mark Boyce.
  - Steve Buskirk and Josh Millspaugh
  - Darryl MacKenzie
  - Trent McDonald, Bryan Manly, Ryan Nielson, and Lowell Diller
  - Josh Millspaugh and seven co-authors
  - Hall Sawyer, Ryan Nielson, Fred Lindzey, and Lyman McDonald
- MacKenzie, D.I., J.D. Nichols, J. A. Royle, and K.H. Pollock. 2006. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Academic Press, Burlington, MA.

# Resource Selection (Probability) Functions

- Models for the relative probability (or probability) that a unit in the study area or a food item is used and detected to be used by the sampling protocol.

- $\text{Pr}(\text{used and detected to be used}) = \text{Pr}(\text{used}) * \text{Pr}(\text{detected to be used} \mid \text{used}).$

**Patch Occupancy Models.** Attempt to estimate both terms in the equation, specifically the probability that a unit is used.

- $\text{Pr}(\text{used})$

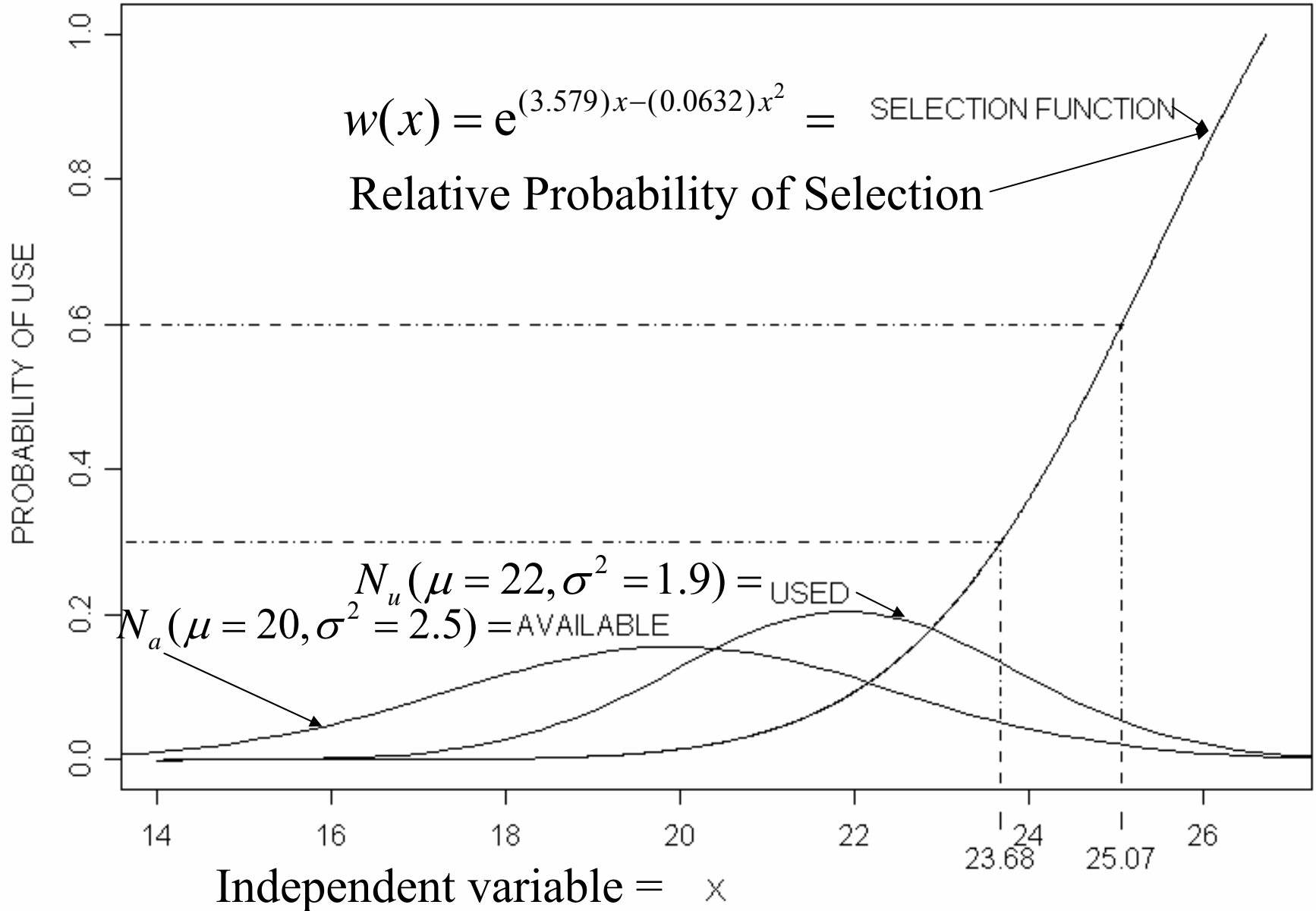
# Resource Selection Functions

## Pr(used and detected to be used)

Typical Applications.

- Relocations of radio tagged animals define points (units) that are ‘used’, i.e. ‘used and detected to be used.’ Model covariates include: slope, aspect, elevation, habitat type, etc.
- Samples of a prey species available before and after predatory fish are introduced to a lake. Model covariates include: size of prey and color.
- Nest boxes are classified as used or unused.

**Hypothetical Example:** The selection function,  $e^{ax-bx^2}$ , will change one normal distribution into another.



SELECTION FUNCTIONS—*McDonald, Manly, and Raley*

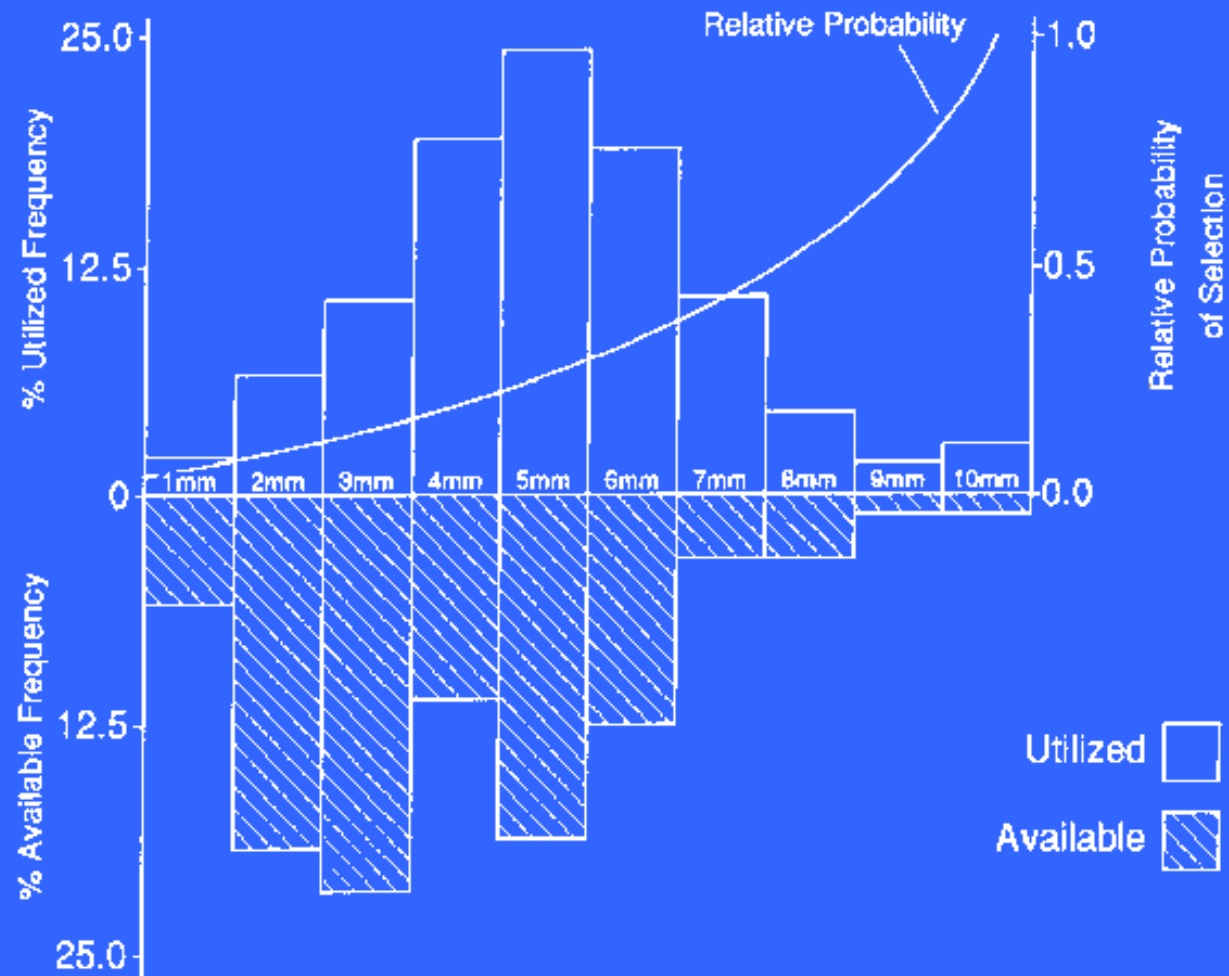


FIGURE 2. Relative probability of selection of insects by Tree Swallows plotted as a function of prey length. The graph is superimposed on the distributions of insect lengths in the samples of available and used prey.

My first estimated resource selection function.

$$f_u(x) = \frac{w(x) f_a(x)}{E_{f_a}[w(x)]}$$

## Weighted distribution theory:

- $x$  is a vector of covariates measured on ‘units.’
- $w(x)$  is a non-negative weight or selection function.
- $f_u(x)$  is the distribution of  $x$  for used units.
- $f_a(x)$  is the distribution of  $x$  for units in the study area (available units).
- If the constant,  $E_{f_a}[w(x)]$ , can be evaluated then we obtain a probability selection function  $w^*(x)$ , where

$$f_u(x) = w^*(x) f_a(x)$$

$$f_u(x) = \frac{w(x) f_a(x)}{E_{f_a}[w(x)]}$$

Relationship to Information Theory:

Kullback-Liebler directed distance

from  $f_u(x)$  to  $f_a(x)$  is

$$E_{f_a}[-\log_e(w(x))] = E_{f_a}[\text{entropy}]$$

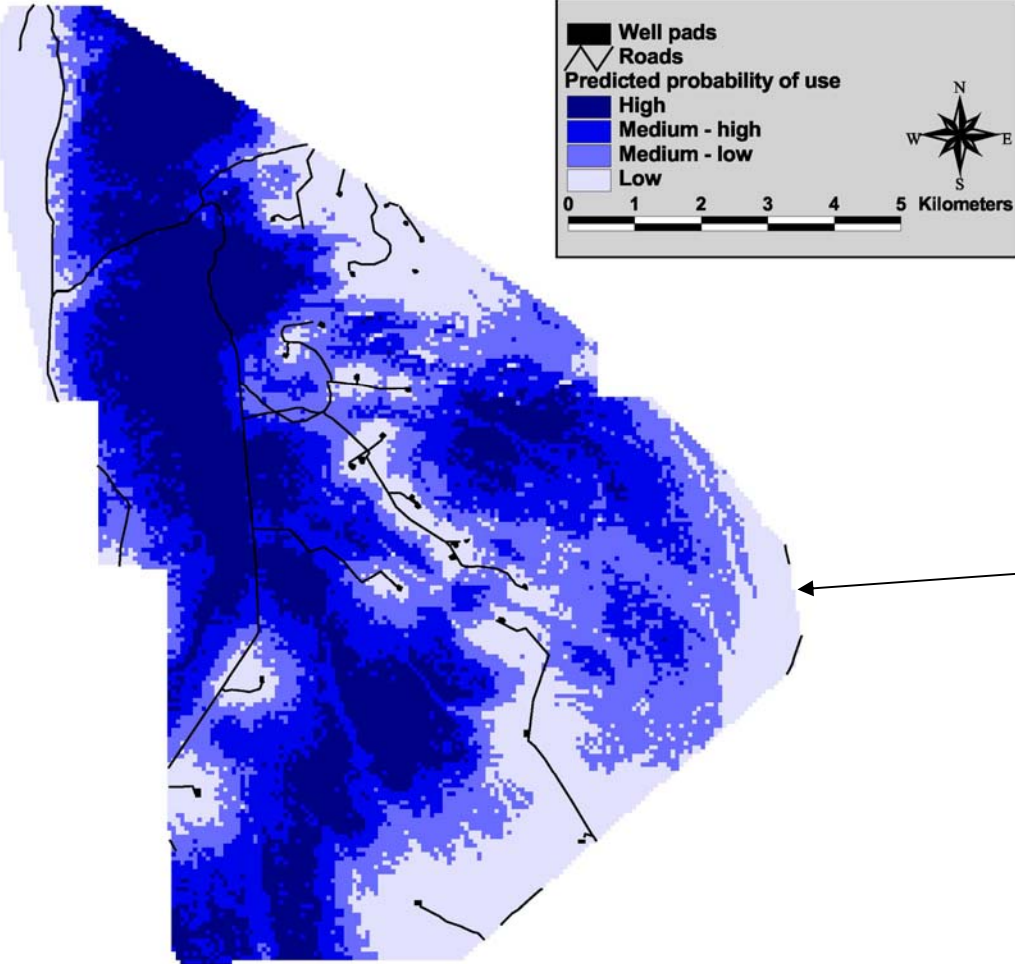
$$= E_{f_a}[-\log(\text{selection function})].$$

$$f_u(x) = \frac{w(x)f_a(x)}{E_{f_a}[w(x)]}$$

- Given estimates of two of the three functions, we can estimate the third.
- $w(x)$  is the fitness function in study of natural selection.
- Horvitz-Thompson estimates.
  - Given biased observed data [e.g., unequally sampled data,  $f_u(x)$ ]
  - selection function  $w(x)$  [e.g., unequal sampling probabilities]
  - we can estimate parameters of the unbiased data,  $f_a(x)$  [e.g, Horvitz-Thompson estimators].
- Line transect sampling:  $x$  is the perpendicular distance to detected objects.
  - $w(x)$  (*i.e.*,  $g(x)$ ) is the detection function,  $f_a(x)$  is a uniform distribution given random placement of transects, and  $f_u(x)$  is the distribution of observed perpendicular distances.

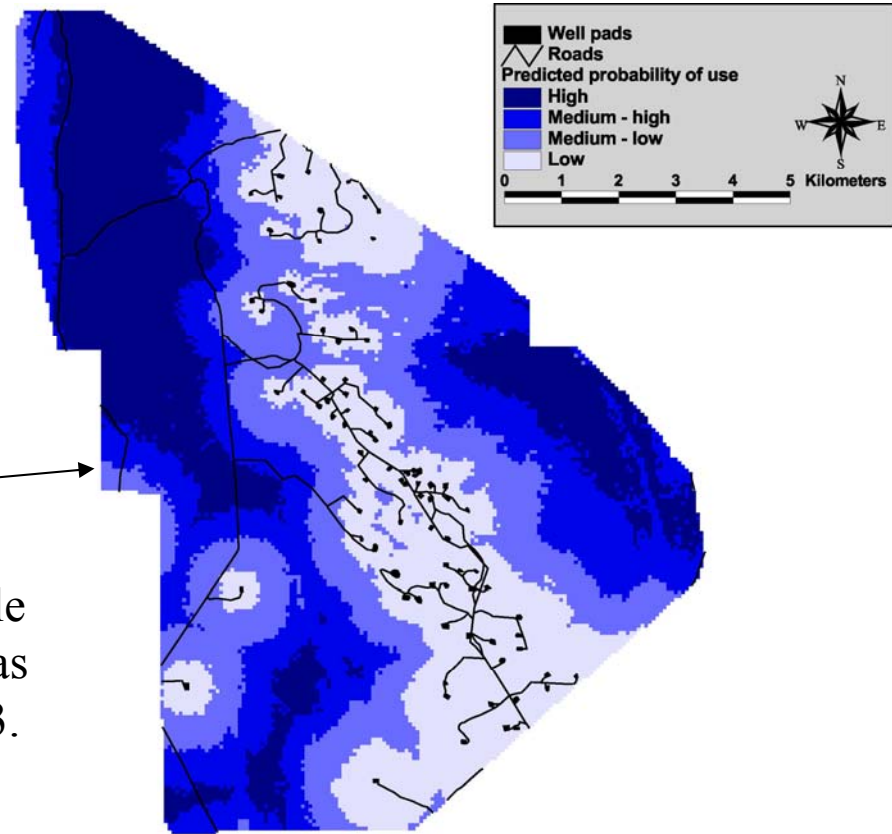
$$f_u(x) = \frac{w(x)f_a(x)}{E[w(x)]}$$

- The most common application in resource selection studies.
  - Sample of units (patches, points) in a study area with pdf  $f_a(x)$  i.e., the units ‘available’ (I have grown to hate that word!!)
  - Sample of units (patches, points) ‘used **and detected to be used**’ by the animals with, pdf  $f_u(x)$
  - Estimate  $w(x)$ , the Resource Selection Function (RSF), an estimate of relative probability of selection as a function of  $x$ .
  - Usually, sampling fractions are not known and  $w(x)$  cannot be scaled to a probability selection function.
  - $\text{Pr}(\text{use}) \times \text{Pr}(\text{detected} \mid \text{use})$  cannot be unscrambled without additional information.



Estimated relative probability of ‘use and detection of use’ by deer in first year of study of effect of drilling for natural gas.

Third year of study with increased drilling activity.



Sawyer et al. 2006. Winter habitat selection of mule deer before and during development of a natural gas field. *Journal of Wildlife Management* 70: 396-403.

# Patch Occupancy Models (MacKenzie et al. 2006)

- **The original study design.**
  - One sample of patches (units, points, etc.) from a study area ‘available’ to the animals.
  - Repeated independent visits to the units over time.
  - Record ‘detection’ (1) or ‘non detection’ (0).
  - Data are a matrix of 1’s and 0’s (rows correspond to units, columns correspond to times)
- **Assumptions**
  - Independent visits.
  - Closure (i.e., if a unit is used, then it is used on all survey times & if unused, it is unused on all survey times).

# Patch Occupancy Models (MacKenzie et al. 2006)

- For example, likelihoods for units with data

0101

0000

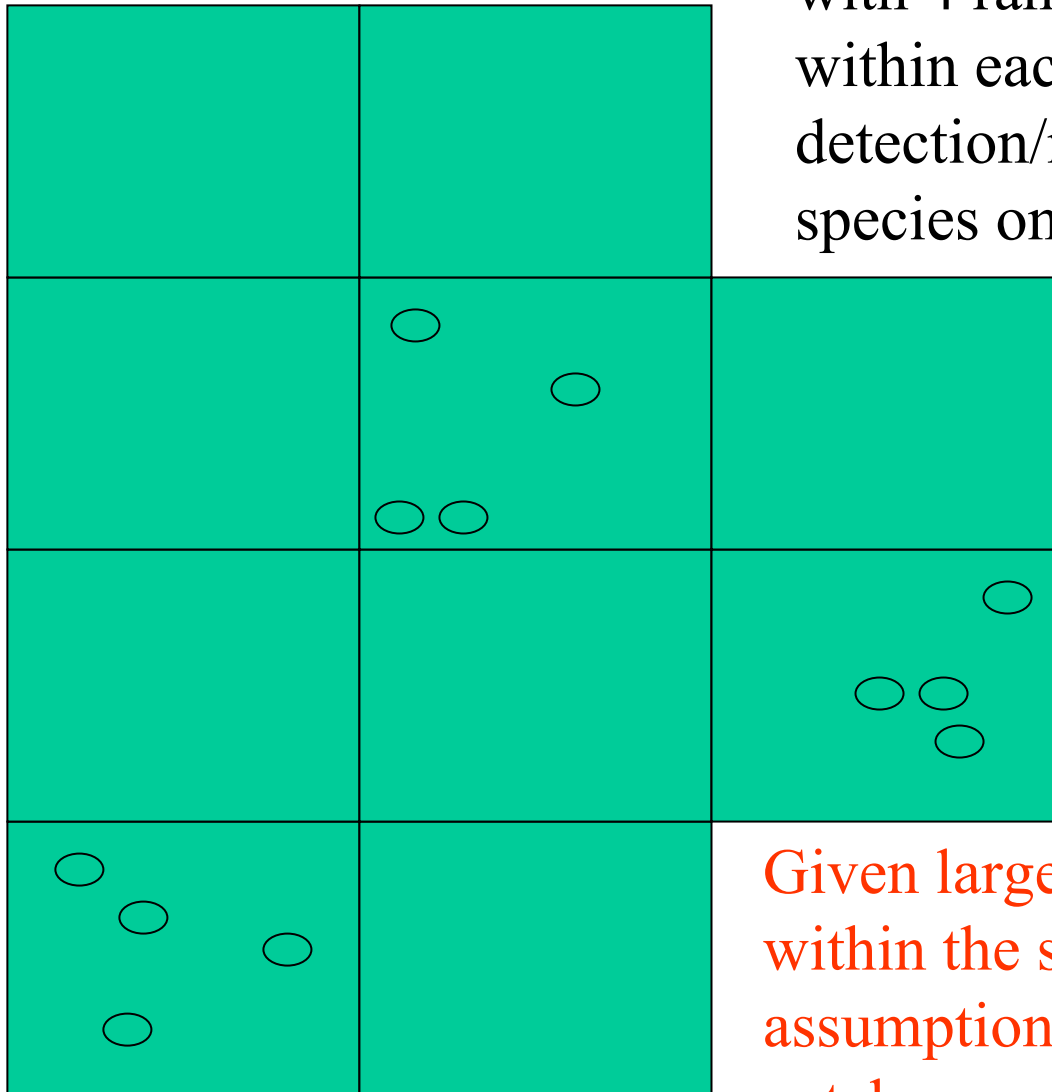
1000

- $w^*(x_i)(1-p)p(1-p)p$
  - $w^*(x_i)(1-p)^4$
  - $w^*(x_i)p(1-p)^3$
- $p = \text{Pr}(\text{detection} \mid \text{used})$ , but could be modeled.
- $w^*(x_i)$  might be modeled by a logistic function of  $x_i$ .
- Combined likelihood function can be maximized for estimates of  $p$  and  $w^*(x_i)$ .
- Theory and estimation methods are similar to those for analysis of capture-recapture studies.

# Alternatives for Repeated Independent Surveys

- Conduct multiple ‘independent’ surveys during a single visit to a sample of sites.
  - Independent surveyors.
- Within large sites, conduct surveys at multiple smaller subplots.
  - **Closure assumption is easily violated!**
  - If there is one animal in the site, then at most one subplot can be occupied.

Study area with 10 units.



Random sample of 3 units, with 4 random sub-units within each. Record detection/non-detection of a species on single visit to each

sub-unit.

Data matrix.

1 0 0 1

0 0 0 0

0 1 0 0

Given large mobile animals within the study area, are the assumptions for estimation of patch occupancy models satisfied?

# Relax Closure Assumption

- MacKenzie et al. (2006, pages 105, 213)
- If animals are moving ‘at random’ then the same theory can be applied to estimate probability a unit is used.
- Animals occupy an area larger than a unit with non-zero probability of being present at the time of the surveys.
  - Theory is not correct unless units are sampled with replacement.
  - Assumption of closure (independence) is violated, e.g., if there is one animal in the study area then at most one sample unit can be occupied.
- Methods require a large number of animals moving at random. Estimates are easily biased.

# Definition of 'Available' Units.

- Ideally, 'Patch Occupancy Models' and 'Resource Selection Probability Functions' are identical.
- The study area defines the units under study, that is, the units 'available.'
- If the study area is changed, estimated coefficients in the estimated resource selection function and patch occupancy model will change.
- **Both methods depend equally on the units defined to be 'available' in the study area!**

# Problem

- Units are random points in the study area.
- Units are baited to attract animals, e.g. scent stations to attract coyotes.
- Units are checked multiple times and presence or absence of evidence of ‘use’ is recorded, e.g., tracks of coyote(s) in smooth sand.
- History of visits at the points might be:
  - 100010
  - 000000
  - 001011
  - Etc.
- Is it OK to estimate a Patch Occupancy Model using standard methods in MacKenzie et al. (2006)?

# Problem

- Units are random points in the study area.
- Units are baited to attract animals, e.g. scent stations to attract coyotes.
- Units are visited multiple times and presence or absence of evidence of ‘use’ is recorded, e.g., tracks of coyote(s) in smooth sand.
- History of visits at the points might be:
  - 100010 The point is visited.
  - 000000 The point is not visited.
  - 001011 The point is visited.
  - Etc.
- Is it OK to estimate a Resource Selection Probability Function using logistic regression of visited and non-visited points on predictor variables?

# Research Problem

- Periodic re-location of radio tagged animals.
- It is common to have gaps of missing data. Probability of recording the location of an animal may depend on the habitat (predictor variables).
  - ×
- RSFs from samples of used and available units predict
  - $\Pr(\text{use}) \times \Pr(\text{detection} \mid \text{use})$
- Research problem. How can the patch occupancy methods be used to separate these two components?
- One idea. Large number of tagged animals. Record presence-absence of tagged animals in each unit periodically, i.e., we are sampling with replacement. It may work!
  - Problem. Maintain independence of sample data for MLE.

# Application to detection of a condition, i.e., disease, in a population.

- One sample of animals from the population in study area, i.e., sample ‘available’ animals.
- One sample of animals with the disease, i.e., they have the disease and are detected by the study protocol.
- Use RSF to model
  - $\Pr(\text{disease and detection of disease})$   
 $= \Pr(\text{disease}) * \Pr(\text{detection} | \text{disease})$
- Medical biostatisticians may not be aware of RSFs.
- $\Pr(\text{detection} | \text{disease})$  is not 100% and depends on the predictor variables measured on the subject?

# Application to detection of a condition, i.e., disease, in a population.

- Fit the Resource Selection Function  $\exp(\beta'x)$  using standard software.
  - Run the two samples through any software for logistic regression. The MLEs for the model  $\exp(\beta'x)$  are the same as produced by the software for  $\exp(\beta'x)/(1 + \exp(\beta'x))$ .
  - $\exp(\beta'x)$  is an estimate of the relative probability of the disease as a function of the covariates  $x$ .
  - Estimates  $\Pr(\text{disease}) * \Pr(\text{detection} | \text{disease})$ .
- Trent McDonald in Johnson et al. (2006) JWM, No. 2, finally proved that the likelihood is maximized by this ‘short cut,’ originally suggested by Manly, McDonald, et al. (1996, 2002)

# Use Patch Occupancy Models to Break the Confounding

- Random sample of animals in the population.
- 1<sup>st</sup> Assumption: Independence.
  - Obtain evaluations of each animal by ‘independent’ methods (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> independent opinions).
- 2<sup>nd</sup> Assumption: Closure.
  - If subject has the disease, they have it during all examinations.
- Patch occupancy methods apply.
- Problem. With rare diseases, sample size may have to be very large.
- Medical biostatisticians may not be aware of methods.

You may download a copy of the PowerPoint Slides or, if I am successful, a wmv copy of the entire presentation including audio, from our web site next week: [www.west-inc.com](http://www.west-inc.com), or from our ftp site

<ftp://ftp.west-inc.com>

Username: webinar

Password: west

E-mail questions or comments to [lmcdonald@west-inc.com](mailto:lmcdonald@west-inc.com)

The End.

06/03/2006