Using Fuzzy Sets to Deal with Uncertainty in Single-species, Single-season Occupancy Models.

Malcolm L. McCallum

School of Agriculture and Applied Sciences, Langston University, Langston, Oklahoma 73050

e-mail: malcolm.mccallum.tamut@gmail.com

Abstract: Occupancy modeling is a valuable tool for managing wildlife populations. Current occupancy models provide estimates of occurrence based on a point estimate for the species detectability and presence-absence. However, detectability can vary based on many variables ranging from weather to personnel. Therefore, I propose the use of fuzzy sets rather than point estimates for detectability and binomial presence-absence data during calculations of occupancy. Fuzzy occupancy estimates are easier to determine, more robust, and generally more informative than traditional point-based occupancy models. Consequently, managers will have better information available for comparing occupancy values among sites. Fuzzy sets are especially useful when parameters of the study violate key data standards.

Key words: Conservation biology, Fuzzy logic, Fuzzy math, inventory, monitoring, Occupancy modeling, Wildlife management.

INTRODUCTION

Occupancy modeling is a valuable modeling technique in wildlife management (MacKenzie 2005). It uses arbitrarily defined spatial units or discrete naturally occurring sample units to estimate with imperfect detectability what fraction of a given area that a species occupies (MacKenzie et al. 2002). One important factor in constructing an occupancy model is determining the detectability of the species in question (Mazerolle et al. 2007). Accurately estimating this statistic is central to producing a robust occupancy model with portrays closely the actual state of occupancy in the study area (MacKenzie et al. 2003; MacKenzie 2005). The presence of a species is confirmed by detecting it on the study site. However, confirming that a species is actually absent from a site when it was not found is in most cases nearly impossible (MacKenzie et al. 2006). To deal with this problem, an array of mathematical methods is available.

If the detection probability is 1 or known to be without error, an unlikely scenario, our confidence that detection equals presence will be high. However, the more difficult an organism is to detect, the less confident we can be that sites without detection are sites without presence. In surveys for wildlife, especially endangered wildlife, it is often more important to ensure absence than presence. If the organism is absent, no special action may be needed. However, if there is any chance an exceedingly rare or difficult to detect species (e.g., Ivory Billed Woodpecker) is present, it is critical that actions be taken to avoid extinction.

There are many occupancy models and methods of calculating detectability (MacKenzie et al. 2006). Although occupancy modeling, especially with the program PRESENCE or MARK, is not difficult or hard to understand, data sets must fit a series of standards to be used in the mathematical computations. First, problems arise if you are dealing with a small number of study

McCallum. Fuzzy sets and uncertainty in occupancy models. BIORXIV/2018/353623

1 sites, typically 10-20 is a minimum depending on the p-value. Further, if the parameter estimates
2 are close to 1 or 0, such as when there is a very low detection probability, psi estimates and
3 standard error will be absurdly high. Further, all current occupancy models and their
4 detectability estimates use and produce point estimates. A point estimator alone is often
5 insufficient to provide a complete inference because of the importance uncertainty plays in the
6 results (Casella and Berger 2001). The use of fuzzy sets and interval estimates in occupancy and
7 detectability calculations allows their use when data sets are not suited and when detectability is
8 difficult to assess.

9 Both interval analysis and fuzzy math are applicable to all kinds of uncertainty, without the
10 need for subjective interpretations used in Monte Carlo approaches, and its foundation is a
11 consistent axiomatic system that is different from that used in probability theory (Ferson et al.
12 1999). Fuzzy arithmetic is a conservative, nonsubjective generalization of interval analysis. Both
13 deal well with uncertainty and require fewer data than alternative methods (Silvert 1997, 2000;
14 Ferson et al. 1999). Whereas interval analysis uses a simple range of values, all of which are
15 equally ranked in regard to their possibility of being true; fuzzy arithmetic rates data (x-axis)
16 based on degrees of possibility (P) called membership values (y-axis) where $y = 0 =$ lowest
17 possibility ($P_{\min}$) and $y = 1 =$ highest possibility ($P_{\max}$). All other x-values have decreasing
18 membership values (i.e., possibilities, as y approaches zero). Fuzzy arithmetic is particularly
19 useful where high levels of uncertainty such as ambiguity, nonspecificity, discord, and fuzziness
20 exist (Klir and Yuan 1994). Fuzzy arithmetic was used to compare current declines with
21 background extinction (McCallum 2007; 2015), estimate the impacts of climate change on
22 wildlife (McCallum et al. 2009), and for risk assessment (Ferson et al. 1999).

23 The real advantage of fuzzy and interval-based studies is that it provides a much clearer
24 understanding of the situation even if relatively little information is known. For example, when
25 extinction estimates were compared with the fossil record, the actual values for extinction
26 estimates that were calculated were very wide; however, they clearly demonstrated that current
27 losses were not normal and compared more favorably with that previously observed during mass
28 extinctions (McCallum 2007; McCallum 2015; Barnosky 2006; Ceballos et al. 2015). Likewise,

29 There are numerous mathematical approaches to occupancy modelling, all of which involve
30 point estimates at the current time (see review in MacKenzie et al. 2006). Each has weaknesses
31 and strengths when used with point estimates. However, with the use of fuzzy sets and intervals,
32 they become more robust estimators.

33 Herein, I demonstrate using RAMAS RiscCalc (Applied Biomathematics, Setauket, NY) how
34 fuzzy sets can improve detectability estimates used in occupancy modeling.

35 36 A FUZZY APPROACH TO DETECTABILITY

37
38 There are many complex occupancy models for wildlife managers from which to choose
39 (MacKenzie et al. 2006). The most basic model:

$$40 \Psi = s_D / (s)(p^*) \text{ [eq. 1]}$$

41
42 Where:

43 s = the number of sites surveyed

44 s_D = the number of sites where the organism was detected

McCallum. Fuzzy sets and uncertainty in occupancy models. BIORXIV/2018/353623

$p^* = \text{the probability of detecting the organism in 't' surveys} = 1 - (1 - p)^t$

If I surveyed 16 sampling sites in a preserve and detected the organism at least once in eight of those locations over three survey periods, with an estimated $p = 0.8$, then

$$p^* = 1 - (1 - p)^t = 1 - (1 - 0.8)^3 = 0.992 \text{ [eq. 2]}$$

then the point estimate for occupancy at the preserve is:

$$\Psi = 8 / (16 * 0.992) = 0.504 \text{ [eq. 3]}$$

However, the true detection probability for an organism is unknown and we can safely assume that for many organisms the instantaneous detection probability on any given survey varies with many variables such as environmental factors, physiography, and habitat complexity (MacKenzie and Royle 2005). Still, we can certainly assume, barring a case of extreme luck, that the detection probability for the organism is not any less than 0.5 in the above example because that is the actual rate that we detected the organism in this study ($8/16 = 0.5$). If it was less detectable than 0.5, we may not expect to detect it 50% of the time. However, the surveyors may be careless or poorly skilled, and the organism might actually be perfectly detectable ($p = 1$) with good personnel under typical conditions. So, the detection probability could range from 0.5 – 1 with a best estimated $p = 0.8$. Therefore, we can set up a fuzzy set for p and calculate p^*

$$p^* = 1 - (1 - [0.5, 0.8, 1.0])^3 = [0.875, 0.992, 1.0] \text{ [eq. 4]}$$

The graphical representation of p^* (Fig. 1) demonstrates that detectability above 0.992 rapidly falls to zero, suggesting that values that are most representative of what is possible are < 0.992 . At least, if also greater than 0.5.

The detectability estimate (p^*) above is used in the following computation to calculate a fuzzy estimate for occupancy (Fig. 2):

$$\Psi = 8 / (16 * [0.875, 0.992, 1.0]) = [0.5, 0.504, 0.571] \text{ [eq. 5]}$$

This shows that occupancy may range from 0.5 – 0.571. Here, $\Psi = 0.504$ remains the best occupancy estimate, but other rates are certainly possible. In fact, the estimates with the highest degree of possibility tend to be higher than the best estimate. This suggests that occupancy may be higher the best estimate, but it is probably not much lower than it. Notice that the base of the fuzzy estimate is equal to the interval estimate. Furthermore, there is an assurance value for ranges spanning the polygon. We do not expect the occupancy to be at or below 0.5 or at or above 0.571 because the membership values (y-axis) are zero. As occupancy approaches the best estimate ($y = 1$), our confidence that occupancy may fall in that range increases to 50% ($y = 0.5$), 70% ($y = 0.7$) and so on. If one was not confident in assigning a middle value for a fuzzy set, this calculation could be performed as a simple high and low to give an interval estimate instead of a fuzzy estimate. That interval estimate for occupancy would be range 0.5 – 0.571, but interval analysis will not provide the membership values on the y-axis to assess relative confidence of different values within the interval.

The above demonstration uses enough sampling sites for traditional occupancy modeling. However, if detectability is really low with a small number of sampling sites, this can be a

Figure 1. Graphical representation of p (left) and p^* (right).

1 problem with point-based occupancy models. If four sampling sites in a preserve are surveyed
2 and the organism detected at least once in two of those locations over three survey periods, with
3 an estimated $p = 0.05$, then

$$4 \quad p^* = 1 - (1 - p)^t = 1 - (1 - 0.05)^3 = 0.142 \text{ [eq. 6]}$$

5 then the point estimate for occupancy at the preserve is:

$$6 \quad \Psi = 2/(4 \cdot 0.142) = 3.51 \text{ [eq. 7]}$$

7 However, if sets are used, p^* is (Fig. 3)

Figure 2. Fuzzy estimate of occupancy in a model survey site.

$$p^* = 1 - (1 - p)^t = 1 - (1 - [0, 0.05, 1])^3 = [0, 0.143, 1.0] \text{ [eq. 8]}$$

The graphical representation (Fig. 3) shows that detectability for this species won't be much less than the best estimate; however, detectability may be substantially higher. The percent confidence is much higher for much broader intervals in this display than we saw in the earlier example because uncertainty is higher due to low numbers of sampling sites and the low detection. If an investigator does not have an informed reason for using bounds larger than zero of less than one, then it is standard practice to use a full range base on the polygon of 0 – 1. However, using zero as the lower threshold of the fuzzy or interval estimate is problematic. When p^* is substituted into the occupancy formula to obtain the estimate for this preserve, we have a zero in the denominator which results in an undefined relationship.

1

Figure 3. Graphical depiction of p^* where detection and number of sampling sites is very low.

2
$$\Psi = 2/(4*[0, 0.143, 1.0]) = \text{undefined. [eq. 9]}$$

3 To avoid this and obtain usable information, one can substitute 0.000001 for zero.

4
$$\Psi = 2/(4*[0.000001, 0.143, 1.0]) = [0.5, 3.51, 500,000] \text{ [eq. 10]}$$

5 This demonstrates how such a low p^* can create results that are problems for point based
6 approaches. Essentially, there is a very high, narrow peak at 3.51, with occupancy values that
7

Figure 4. Full fuzzy occupancy estimate (left) and the portion of the fuzzy occupancy estimate above $y = 0.1$ (right).

1 slowly tail off in membership from around 10% to zero where it reaches 500,000 (Fig. 4). There
2 is roughly a 50% assurance that occupancy is 2 – 10.

3 This is the value of using fuzzy sets over point estimates. We know that anytime we calculate
4 detection or occupancy that the number we generate is accompanied by a certain degree of
5 confidence, error, or uncertainty. When we apply probability theory to occupancy, we are
6 largely concerned with the accuracy of a single number, or what is the most probable occupancy.
7 However, in situations where we are assessing risk of extirpation, extinction or impairment, it
8 really is more vital to ensure absence than presence. Fuzzy and interval analyses

9 do this. Anything outside of the bounds of the interval is generally not possible. There is
10 much more information revealed by a fuzzy distribution than in a point estimate with error bars.
11 As such, we can be more confident in our absence data, ensuring that actions taken do not
12 encumber populations that were present but went undetected.

13 14 **REPLACING BINOMIAL PRESENCE-ABSENCE WITH FUZZY SETS**

15
16 In the previous section, we used the binomial yes-no results from surveys in the field for one
17 location with 16 sampling sites and another with only four sampling sites. Organisms that are
18 difficult to detect frequently go unreported in surveys. Fuzzy sets can be used to deal with this
19 problem. If our study site has 16 sampling locations sufficiently near to known populations we
20 can assess the confidence in our detection or non-detection very easily. To do this, we use a

McCallum. Fuzzy sets and uncertainty in occupancy models. BIORXIV/2018/353623

Table 1. Fuzzy representation of detection estimates at sixteen hypothetical sampling sites for a single species within a single study site.

Sampling Site	Fuzzy Detection Estimate	Hypothetical Rationale
1	[1]	Detected
2	[1]	Detected
3	[0, 0.5, 1]	A surveyor is sure he saw the animal rapidly dive into the creek.
4	[1]	Detected
5	[0, 0.8, 1]	The animals are present in the pond just outside the sampling site. Even though we did not detect them in this site, the habitat is perfect and they are present so nearby that it is difficult to believe they do not enter this area.
6	[1]	Detected
7	[0, 0.01, 1]	The habitat is very marginal at best, and we saw no signs of the animal
8	[1]	Detected
9	[1]	Detected
10	[0, 0.9, 1]	The animal was in a tree, five feet outside our search radius.
11	[0, 0.2, 1]	We detected the animal the first time, but there are no signs of any here now, and frankly, the habitat is just not right for this species.
12	[1]	Detected
13	[1]	Detected
14	[0, 0.05, 1]	Did not find them and this just isn't a good place for them.
15	[0, 0.05, 1]	See #14
16	[0, 0.05, 1]	See #14

1 sliding scale where $x = 0$ = did not detect, and $x = 1$ = did detect. If we detected the organism,
2 then we use the fuzzy set $x = [1]$. However, if we did not detect the organism, we use a fuzzy set
3 bound by zero and one $[0, ?, 1]$ to represent the extremes that are possible, then insert an estimate
4 on a scale of 0 to 1 representing your confidence the non-detection is accurate. How one
5 assesses confidence may depend on the species and situation in which they are operating. For
6 example, if I am in a parking lot and I survey it for turtles and find none, I am undoubtedly pretty
7 confident due to the simplicity of the habitat, the type of habitat, its location relative to
8 occupied habitat, and my ability to survey the habitat, that the non-detection is accurate. In this
9 situation, I am very confident, so I set up a fuzzy number $[0, 0.05, 1]$ because I am at least 95%
10 confident that my result is accurate. However, if I were in a very complex habitat, under time
11 constraints with a crew that was very inexperienced, I may not be very confident that non-
12 detection is accurate. In such a situation, I account for this by estimating my confidence with a
13 fuzzy number $[0, 0.5, 1]$ meaning that I am only 50% confident that the non-detection is
14 accurate, and that the species does not occur here. At first, this might seem very subjective, but
15 converting cognitive estimates into a quantitative scheme in this way can be very accurate and

McCallum. Fuzzy sets and uncertainty in occupancy models. BIORXIV/2018/353623

informative within the study. In this study site, we had 16 sampling sites. There are various reasons you might choose to use a fuzzy estimate instead a 1 or a 0 in the calculations (Table 1.). However, to remain objective, one should develop a rubric that delineates values based on clear-cut easy to interpret standards. If we were using binomial data, our results for this inventory would 9/16 sampling sites where the animal was detected at least once during the three visits. However, in looking at the notes, we can see that some of these detections and non-detections have problems. Do we record the animal in the tree? Some investigators would be tempted to do so, but this violates the experimental design. However, ignoring its presence in the calculations is counter to the purpose of the study when we can see clearly that the animal is present and undoubtedly uses sampling site. There is even a detection that we have in our notes but makes no sense. Either the animal was well outside its normal habitat, it was misidentified, or there was an error in data recording. With the fuzzy set, we can incorporate some sense into this model to account for our certainty in my results. The fuzzy results for this inventory would give us eight sites where detection was thought accurate. Then, we have one site where the detection is baffling, and another seven sites where non-detection varies in confidence due to experience or difficulty in surveying.

Using the p^* from [eq. 4] to calculate occupancy and then using the fuzzy set estimates for detection in the numerator

Figure 5. Fuzzy occupancy estimate for study site.

McCallum. Fuzzy sets and uncertainty in occupancy models. BIORXIV/2018/353623

$$\Psi = [4, 6.56, 14.5, 20]/(16*[0.875, 0.992, 1.0]) = [0.25, 0.51, 1.14, 1.43] \text{ [eq. 11]}$$

There is much more information available in the fuzzy occupancy estimate [eq. 11, Fig 5] than in a point estimate [eq. 3]. The fuzzy estimate provides that total occupancy could fall from 0.25 to 1.43, but the most credible interval estimate is 0.51 – 1.14. Any values where $1.43 < x < 0.25$ are not possible. It also shows that there are more possible values > 1.14 than possible values less than 0.51. This tells us much about the status of the species that we have surveyed compared to a simple point estimate for occupancy, even if standard deviation, standard error, or confidence intervals are determined. In fact, the fuzzy estimate [0.25, 0.51, 1.14, 1.43] suggests occupancy is actually much lower than previously predicted occupancy ([0.5, 0.504, 0.571]) using a fuzzy detectability estimate and binomial presence-absence data in [eq. 5] (Fig. 2).

In addition to a more informative estimate for occupancy, fuzzy occupancy calculations are much easier to accomplish and more flexible with number of sampling sites and low detectability. In fact, the robustness of fuzzy estimates to deal with uncertainty can also encompass all kinds of variability among sampling sites. They are also much easier to interpret and communicate to others.

There are many different ways to calculate detectability and occupancy, in this study, I used a very simplistic model. However, incorporation of fuzzy sets or interval sets into more complex models is as simple as shown. However, besides fuzzifying the variables as I have done, one can also make coefficients in the functions fuzzy. The approach is the same, but this is probably not needed except under very unique and unusual circumstances.

LITERATURE CITED

- Casella, G., and R. L. Berger. 2001. Estimation: Point and Interval. Pp. 4744-4749 *In* International Encyclopedia of Social & Behavioral Sciences. Elsevier Science Ltd.
- Ferson, S., W. Root, and R. Kuhn. 1999. RAMAS Risk Calc: Risk assessment with uncertain numbers. Applied Biomathematics, Setauket, NY.
- Klir, G. J., and B. Yuan. 1994. Fuzzy sets and fuzzy logic: Theory and Applications. Prentice Hall, Upper Saddle River, NJ.
- MacKenzie, D. I. 2005. What are the issues with presence-absence data for wildlife managers? *Journal of Wildlife Management* 69(3):849-860.
- MacKenzie, D. I. and J. A. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology* 42:1105-1114.
- MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* 84(8):2200-2207.
- 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(8):2248-2255.
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, J. E. Hines. 2006. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Academic Press, Burlington, MA. 324 pp.

McCallum. Fuzzy sets and uncertainty in occupancy models. BIORXIV/2018/353623

- 1 Mackenzie, D. I., J. D. Nichols, N. Sutton, K. Kawanishi, and L. L. Bailey. 2005. Improving
- 2 inferences in population studies of rare species that are detected imperfectly. *Ecology*
- 3 86(5):1101-1113.
- 4 Mazerolle, M. J., L. A. Bailey, W. L. Kendall, J. A. Royle, S. J. Converse, and J. D. Nichols.
- 5 2007. Making great leaps forward: Accounting for detectability in herpetological field studies.
- 6 *Journal of Herpetology* 41(4):672-689.
- 7 McCallum, M.L. 2015. Vertebrate biodiversity losses point to a sixth mass extinction.
- 8 *Biodiversity and Conservation* 24:2497 – 2519.
- 9 McCallum, M.L. 2010. Future climate change spells catastrophe for Blanchard’s Cricket Frog,
- 10 *Acris blanchardi* (Amphibia: Anura: Hylidae). *Acta Herpetologica* 5:119 – 130.
- 11 McCallum, M. L. 2007. Amphibian decline or extinction? Current declines dwarf background
- 12 extinction rate. *Journal of Herpetology* 41(3):483-491.
- 13 McCallum, M. L., J.L. McCallum, S.E. Trauth. 2009. Predicted climate change may spark box
- 14 turtle declines. *Amphibia-Reptilia* 30:259-264.
- 15 Silvert, W. 1997. Ecological impact classification with fuzzy sets. *Ecological Modelling* 96:1-10.
- 16 Silvert, W. 2000. Fuzzy indices of environmental conditions. *Ecological Modelling* 130:111-119.
- 17
- 18
- 19