A VALIDATION TECHNIQUE FOR ASSESSING PREDICTIVE ABILITIES OF RESOURCE SELECTION FUNCTIONS

SHAY HOWLIN[1], Western EcoSystems Technology, Inc. 2003 Central Avenue, Cheyenne, WY 82001, USA
WALLACE P. ERICKSON, Western EcoSystems Technology, Inc. 2003 Central Avenue, Cheyenne, WY 82001, USA
RYAN M. NIELSON, Western EcoSystems Technology, Inc. 2003 Central Avenue, Cheyenne, WY 82001, USA

Abstract: Two variants of a proposed technique for the validation of resource selection functions (RSF) were compared. Each validation technique involves validation of a model with a set of used points held out from the original data or obtained through additional sampling. A simulation was designed to examine the ability of the validation methods to judge predictive abilities of specified models to data generated under a known RSF. The two validation techniques use covariates and the estimated RSF to predict the expected selection for the validation data set, and then the expected selection is compared to the observed selection. One of the validation methods involves taking a census of use in the entire study area while the other method involves the predicted relative probability of selection for the new used points. A normal theory linear regression model is used to evaluate the relationship between observed and expected selection from the estimated model. The slope of the regression relationship between observed and expected provides an assessment of the predictive abilities of the RSF model. This method is computationally easy to use and provides distinction between RSF models with poor, acceptable, and good predictive abilities.

Key words: Resource selection functions, model validation, prediction, habitat selection.

Resource selection functions are generally developed for one study area using data from a set of used points and a set of available points (Manly et al. 2002). The RSF is often used to quantify the characteristics of the physical environment influencing the selection process or to assess the importance of a characteristic of interest in the selection process (Burnham and Anderson 1998). Perhaps the largest benefit of modeling resource selection for a given population is the ability to predict future resource selection or make predictions in a new geographic area that is ecologically similar to the area used to estimate the RSF. To use RSF predictions in the development of management plans or to plan further research (i.e. survey strata), users should determine if the model predicts selection accurately.

Goodness-of-fit tests are commonly employed in statistical modeling during and at the end of a model-building process. Regression diagnostics such as model residuals, influence statistics, and collinearity can be used to evaluate the performance of the RSF model on the model-building data (Rawlings et al. 1998). Model validation describes the general assessment of the model through the use of data that were not involved in the model-building process. Since model assessment with the data used to create the model generally concludes the model fits (Hosmer and Lemeshow 1989), we recommend running model validation with data set aside during the model building process and not a more traditional k-fold validation. Model assessment ensures the model is robust and applicable to data other than the model data and is necessary when the model is intended for predictive use (Hosmer and Lemeshow 1989).

Validation of a RSF model is inherently difficult because of the nature of the data. The RSF model predicts a continuous value, the relative probability of selection, and a set of data for the validation process can be thought of as discrete, i.e. whether a point was used. Even when the predicted values are scaled from 0 to 1, there is no statistical support for defining a cutoff value above which points are expected to be used, and below which points are expected to be unused.

Here we describe two model validation methods based on a comparison of selection in a new set of used points to the predicted relative probability of resource selection by the model under validation. Validation data can be held out from the original data or obtained through additional sampling. The first validation method, the census method, assumes that if the model has good predictive abilities the high use areas observed in the validation dataset will be in locations with higher predicted relative probabilities of selection (Johnson et al. 2000). The second validation method, the sample method, assumes that if the model has good predictive abilities the validation used points should have higher values of the predicted relative probability of selection compared to a random sample of points from the study area.

[1]Email: showlin@west-inc.com

VALIDATION METHODS

The two techniques proposed here for validating an RSF model involve a new set of *m* used locations with associated covariates. Each validation method compares the predicted resource selection probability from the model to one of two forms of observed selection. Grid cells in the study area (census method) or the available sample (sample method) are grouped into 20 bins based on percentiles of the predicted relative probabilities of resource selection from the model under validation. The observed selection is then compared to the predicted selection in each bin using simple linear regression. The slope of the regression model, with the predicted selection as the predictor of observed selection, is a measure of the predictive ability of the model (Table 1). When the slope is not significantly different from zero there is no correlation between the predictions and actual resource use, and the predictive abilities of the model are unacceptable. When the slope is positive and significantly different from zero there is significant positive correlation between predictions and actual use, and the predictive abilities of the model are acceptable. Perfect predictive abilities would come from a model that provides a 1 to 1 linear relationship with observed resource use. The best predictive model is identified when the slope of the relationship between expected and observed selection is not significantly different from 1 and the regression line runs through the origin (i.e. β_o=0). When the slope is significantly greater than 1, there is significant positive correlation between predictions and actual use, and the predictive abilities of the model are acceptable.

The census validation method requires validation data in the form of counts of use in each grid cell of the study area (Johnson et al. 2000). This method is most useful when many animals are radio collared and the location for each animal can be easily obtained anywhere in the study area. The observed selection for a grid cell is the relative frequency of use obtained by dividing the count of use in the grid cell by the total count of use in the study area. The observed selection for a bin is the sum of these values for each grid cell in the bin. The expected selection for a bin is the sum of the predicted relative probabilities of selection, scaled to sum to 1, for each grid cell in the bin (Table 2). See Appendix A for formulas to calculate observed and expected selection.

The sample validation method requires validation data in the form of a sample of used locations in a study area. This method requires a less intensive survey effort, compared to the census method, to obtain the model validation dataset. The expected selection for a location is the relative probability of selection predicted from the model under validation. The observed selection for a bin is the count of the number of use points in the bin. The expected selection for a bin is calculated as the sum of the expected relative probabilities of selection, scaled to sum to *m*, for each available point in the bin (Table 3). See Appendix A for formulas to calculate observed and expected selection.

SIMULATION METHODS

We designed a simulation to compare the effectiveness of the two validation techniques. A population of 10,000 grid cells was created with a random value for the covariate, X_1, distributed normally with mean equal to 200 and variance equal to 27. The probability of selection was assigned to each cell using an exponential function: $\exp(0.008*X_1)$. The RSF was then scaled to a resource selection probability function (RSPF) so that each value was between 0 and 1. Each grid cell was either used or unused during a given time period as determined by comparing the probability of selection to a random number from a uniform distribution. When the probability of selection was larger than the random number, the point was considered used at that time. When the probability of selection was smaller than the random number, the point was considered not used at that time.

From this population, a random sample of used cells and a random sample of all cells were drawn to simulate the used and available data (Manly et al. 2002) for model building. In addition, model validation data was simulated by comparing the same probability of selection of the grid cells in the population to a new uniform random number, simulating use during another time period. A sample of size 2000, 1000, or 500 was then drawn from the population for the sample validation method. For the census validation method, fifteen time periods were simulated and the sum of use over all times was taken as the count of use. All cells in the population were used in the census validation.

For each validation technique, the predicted relative probability of selection was calculated using the true resource selection probability function:

$$\text{RSPF} = \exp(\beta_o + 0.008*X_1),$$

and six intentionally mis-specified models: $\text{RSF} = \exp(\beta_o + 0.002*X_1)$, $\text{RSF} = \exp(\beta_o + 0.004*X_1)$, $\text{RSF} = \exp(\beta_o + 0.006*X_1)$, $\text{RSF} = \exp(\beta_o + 0.010*X_1)$, $\text{RSF} = \exp(\beta_o + 0.012*X_1)$, and $\text{RSF} = \exp(\beta_o + 0.014*X_1)$. Where β_o was chosen to ensure the function was a resource selection probability function and probabilities were between 0 and 1.

The census validation method was used to obtain expected and observed selection for each bin, and the sample validation method was used to obtain the expected and observed count of use points for each bin. A simple linear regression model was fit to evaluate the relationship between the observed and expected selection (census method) or the observed and expected counts (sample method). The estimated slope and variance were saved for each of the 500 iterations in the simulation. We assume the impact of assuming an incorrect form of the RSF would provide additional information on the robustness of the technique.

Table 1. The predictive ability of the model under validation as determined by the slope and 95% confidence interval of the validation regression.

Slope ($\hat{\beta}$)	95% Confidence interval on $\hat{\beta}$	Predictive abilities of model
+ or -	CI includes 0	Unacceptable
$0<\hat{\beta}<1$	CI excludes 0 and 1	Acceptable
$\hat{\beta}>0$	CI excludes 0 and includes 1	Good
$\hat{\beta}>1$	CI excludes 0 and 1	Acceptable

Table 2. Expected and observed selection of all cells in each bin, for one iteration of the census validation method. Slope of the linear regression model predicting observed selection with expected selection was 1.09.

Bin	Number of cells in bin	Sum of expected selection	Sum of observed selection
1	500	0.032	0.029
2	500	0.036	0.034
3	500	0.038	0.035
4	500	0.040	0.041
5	500	0.042	0.047
6	500	0.043	0.040
7	500	0.044	0.040
8	500	0.046	0.053
9	500	0.047	0.051
10	500	0.048	0.042
11	500	0.050	0.048
12	500	0.051	0.049
13	500	0.053	0.059
14	500	0.054	0.051
15	500	0.056	0.057
16	500	0.057	0.048
17	500	0.060	0.068
18	500	0.062	0.059
19	500	0.067	0.069
20	500	0.076	0.082
Total	10000	1	1

Table 3. Expected and observed counts of used cells in each bin, for one iteration of the sample validation method. Slopes of the liner regression model comparing observed with expected counts were 1.55 for the 0.006 model, 1.17 for the 0.008 model and 0.94 for the 0.010 model.

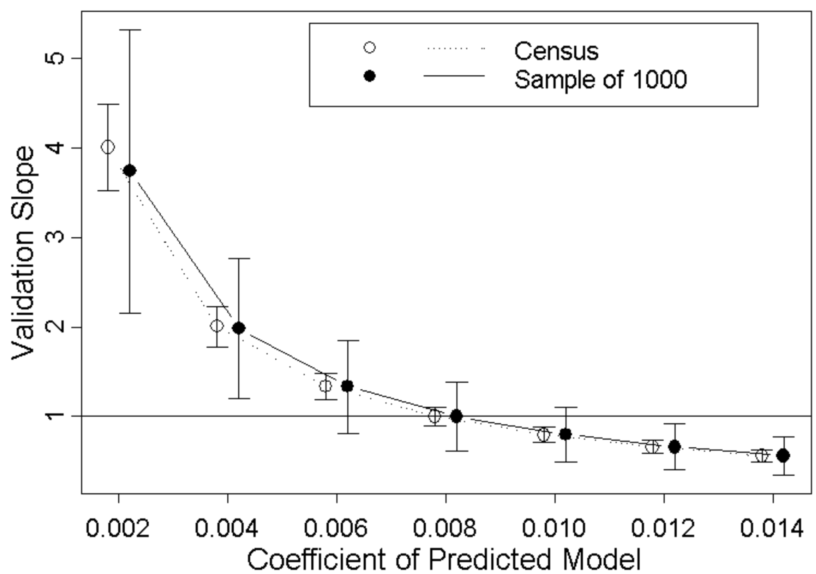
Bin	Number of available cells in bin	Expected count with 0.006 model	Expected count with 0.008 model	Expected count with 0.010 model	Observed count
1	51	34	31	27	32
2	51	38	34	32	37
3	51	40	37	34	34
4	51	41	38	36	38
5	51	42	40	38	49
6	51	43	41	40	40
7	51	44	43	41	35
8	51	45	44	43	49
9	51	46	45	44	35
10	51	47	46	46	49
11	51	48	47	47	38
12	51	49	49	49	41
13	51	50	50	50	53
14	51	51	52	52	54
15	51	52	53	54	44
16	51	53	55	56	55
17	51	55	57	59	61
18	51	56	59	62	57
19	51	59	63	67	75
20	51	65	71	78	80

SIMULATION RESULTS

The slope and 95% upper and lower confidence limits were averaged over the 500 iterations of the simulation for each predicted RSF model (Figure 1). When the predictions of the resource selection probabilities were made with the same model as that of the underlying model of the data, the mean slope for the census validation method was 1.00 (95% CI: 0.89, 1.11). The mean slope for the sample validation method with a sample of 2000 use points was 1.00 (95% CI: 0.67, 1.33), the mean slope for a sample of 1000 use points was 1.00 (95% CI: 0.61, 1.39) and the mean slope for a sample of 500 use points was 1.00 (95% CI: 0.52, 1.48).

When the predicted resource selection probabilities were from models with coefficients smaller than that of the underlying model of the data, the mean slope for both validation methods was larger than 1 (Figure 2). The 95% confidence interval for these estimates did not include 1 for the census method, indicating the census validation method could discern a difference in the model fit when the true model was proportional to $\exp(0.008*X_1)$ and the predictive model was proportional to $\exp(0.006*X_1)$. The mean slope for the sample validation method was comparable to the census method for all model specifications, though the 95% confidence intervals were larger and in some cases included 1.

Figure 1. Slope (and 95% confidence interval) of validation regression predicting observed selection with predictive RSF model with different coefficients. A slope of 1 indicates a one-to-one relationship between the observed and expected values. The underlying data were simulated with a model coefficient of 0.008.



When the predicted resource selection probabilities (expected counts) were from models with coefficients larger than that of the underlying model of the data, the mean slope for both validation methods was smaller than 1. Again, the 95% confidence interval for these estimates did not include 1 for the census method and did include 1 for the sample method with exp(0.006*X1) and exp(0.010*X1) model specifications (Figure 3).

DISCUSSION

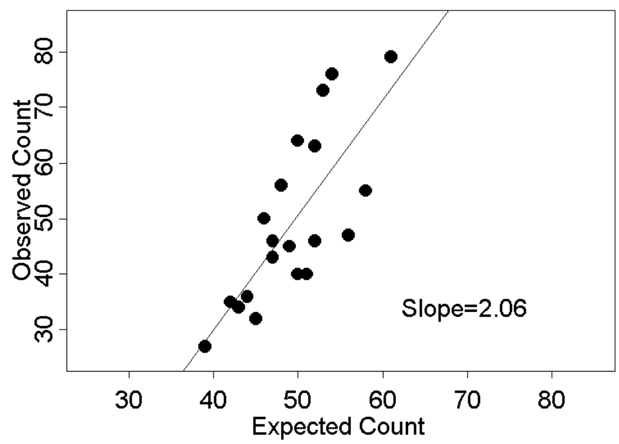
The census validation method is very good at discerning when the correct model for the underlying data structure has been used for predictions. When models with the wrong coefficients are used to predict selection, the census method declares the model predictions acceptable, but not good according to our definition. But, perhaps this method is too sensitive and the models with small bias are biologically meaningful and the predictions are adequate for most applications.

Conversely, for the sample validation method, some of the mis-specified models (coefficients of 0.006 and 0.010) were found to have good predictive abilities. However, we should also consider the possibility that this method is too lenient. Are these model predictions truly useful? Real data will be required to discover whether the model validation method is assessing biologically meaningful results.

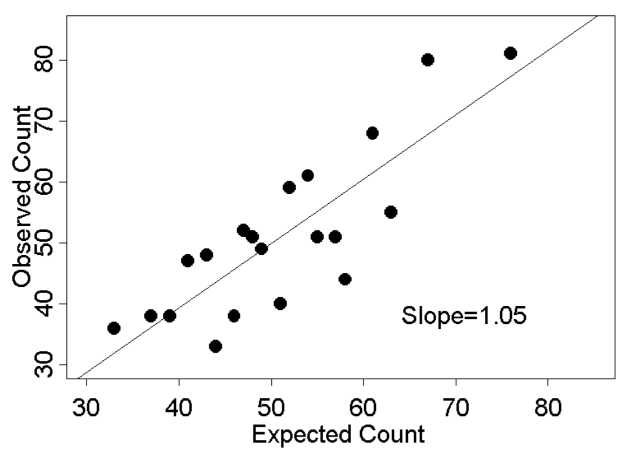
When a validation method has determined that the RSF cannot adequately predict resource selection in the new data, the model may still provide useful information. RSF models with poor predictive abilities for a new year of data or a new region should still be considered useful for explaining the selection process that occurred in the time and space of the data used to develop the model. Interpretation of this model should provide insight into the physical environment influencing the selection process and lead to the identification of differences in selection processes in other times or areas.

Figure 2. Examples of slopes of the linear regression lines for predictive RSF models a) scaled $\exp(0.006*X_1)$, b) scaled $\exp(0.008*X_1)$, and c) scaled $\exp(0.010*X_1)$, and the sample validation method.

a.



b.



c.

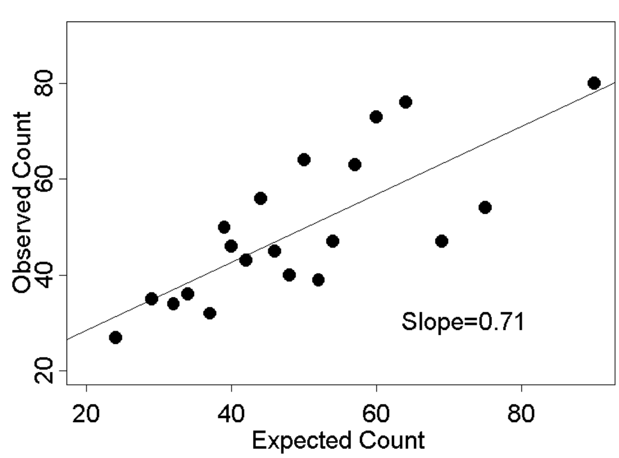
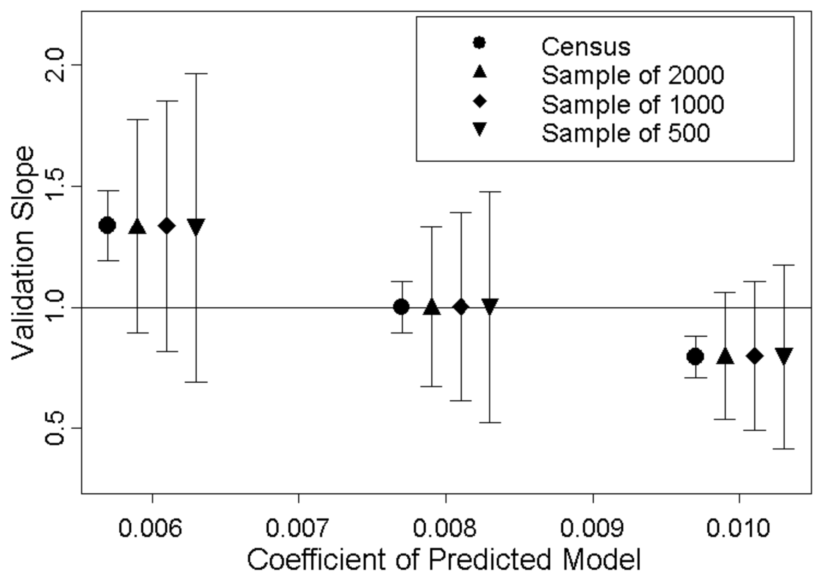


Figure 3. Slope of validation regression predicting observed selection with predictive RSF model with different coefficients. A slope of 1 indicates a one-to-one relationship (i.e. slope = 1) between the observed and expected values. The underlying data were simulated with a model coefficient of 0.008.



We have described the use of the validation technique with validation data held out from the original data or obtained through additional sampling. Boyce et al. (2002) recommend a k-fold validation design, where the models are fit and validated k times. The major difference between the validation technique proposed by Boyce et al. (2002) and our proposed technique is that we use a different definition of predicted selection and provide a clear distinction between RSFs with poor, acceptable, and good predictive abilities. If a k-fold validation design is used, the slope from each validation dataset can be averaged and the standard error calculated across the k datasets, or the proportion of the k datasets with resulting good, acceptable, or unacceptable predictive abilities can be reported.

Questions arise when determining the set of validation points to withhold. Since a randomly selected sample is thought to be representative of the model-building data, validation with this set of data may be optimistic (Manly, pers. comm.). Alternatively, withholding blocks of data in contiguous geographic locations for validation may result in a model unrepresentative of the study area. Rawlings et al. (1998) contended that validation data need to represent the areas for which predictions are desired. For data collected across several years, withholding blocks of data in one time frame may result in a model unrepresentative of the entire time period. We hypothesize it might be best to withhold many small blocks of data within the spatial extent of the study area or within the time frame of the study.

The sample validation technique we propose is also applicable to many discrete choice resource selection models if there is more than one choice in each choice set. Choice sets can be defined based on choices (used points) made within a specified space (e.g. home ranges), or time (e.g. year). In any case, a selection of choices and corresponding choice sets can be withheld from the fit of the model under validation (withheld from the original dataset or obtained through additional sampling). Each choice set can function as the available data in the validation, and the corresponding set of choices can function as the use data. A separate validation regression could be estimated for each choice set to validate one model. The slope from each validation dataset can be averaged and the standard error calculated across the datasets, or the proportion of the datasets with resulting good, acceptable, or unacceptable predictive abilities can be reported. Alternatively, the relationship between the predicted and observed selection in each validation bin of each choice set can be assessed in one validation regression.

EXAMPLE CALCULATIONS FOR SAMPLE VALIDATION TECHNIQUE

This section provides an example of the steps required to conduct the sample validation technique, as described in Appendix B. A set of 1000 observations compose the example validation available dataset (Table 4), while 50 observations compose the example validation dataset (Table 5). The model to be validated is: RSF = $\exp(\beta_o + 0.0155*X)$. Using this function, the predicted relative probability of selection was estimated for the example validation available and used datasets.

Predicted selection is calculated using the example validation available dataset. The predicted relative probability of selection for this dataset is scaled so that the sum of the predictions over all the observations equals the number of used points set aside for validation (50). To scale the dataset, the sum of the predicted relative probability of selection for all the observations in the example validation available dataset is calculated (23936.90). Each observation is divided by this sum and multiplied by 50 to get the scaled predicted relative probability of selection.

Twenty percentiles of the scaled predicted relative probabilities of selection are calculated with the example validation available dataset (Table 6). Each observation in the available dataset is placed into a bin based on these percentiles (e.g. bin 1 contains observations in the available dataset with the scaled predicted relative probability of selection less than or equal to the 5^{th} percentile, bin 2 contains observations in the available dataset with the scaled predicted relative probability of selection less than or equal to the 10^{th} percentile and greater than the 5^{th} percentile, etc.). Each observation will be put into a bin resulting in an equal number of observations in each bin (1000/20bins = 50 observations per bin). The sum of the scaled predicted relative probability of selection is calculated across the observations in each bin. This is the expected selection for each bin. The sum of the expected selection across all 20 bins should equal the number of used points in the validation data (50).

Observed selection is calculated using the example validation used dataset. The predicted relative probability of selection for this data set is scaled identically to the available dataset (divided by 23936.90 and multiplied by 50). Each observation in the used dataset is placed into one of the 20 bins created above for the available points (Table 6). The number of used points in each bin is counted to obtain the observed selection for each bin. The sum of the observed selection across the 20 bins should equal the number of used points in the validation dataset.

The validation is performed by regressing the observed count on the expected count with normal linear regression using the bins as the sample units (Table 7). The slope of this regression is taken as an assessment of the predictive abilities of the model. For this example, the slope of the validation regression is 0.99 with the 95% confidence interval from 0.38 to 1.59. According to table 1, the model in this example has good predictive abilities.

LITERATURE CITED

BOYCE, M. S., P. R. VERNIER, S. E. NIELSON, F. K. A. SCHMIEGELOW. 2002. Evaluating resource selection functions. Ecological Modelling 157: 281-300.

BURNHAM, K. P. AND D. R. ANDERSON. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York.

HOSMER, D. W., AND S. LEMESHOW. 1989. Applied logistic regression. John Wiley and Sons, New York.

JOHNSON, B. K., J. W. KERN, M. J. WISDOM, S. L. FINDHOLT, AND J. G. KIE. 2000. Resource selection and spatial separation of mule deer and elk during spring. Journal of Wildlife Management 64:685-697.

MANLY, B. F. J, L. L. MCDONALD, D. L. THOMAS, T. L. MCDONALD, W. P. ERICKSON. 2002. Resource selection by animals: statistical design and analysis for field studies, 2nd edition. Kluwer Academic Publishers.

RAWLINGS, J. O., S. G. PANTULA, AND D. A. DICKEY. 1998. Applied regression analysis: a research tool, 2nd edition.Springer-Verlag,NewYork.

Table 4. Twenty-four observations of the example validation available dataset.

Observation number	X	Predicted relative probability of selection	Scaled predicted relative probability of selection	Bin
1	169.03	13.73	0.0287	3
2	199.60	22.06	0.0461	11
3	191.37	19.42	0.0406	9
4	200.64	22.42	0.0468	11
5	188.87	18.68	0.0390	8
6	224.47	32.44	0.0678	17
7	196.70	21.09	0.0441	10
8	185.48	17.72	0.0370	7
9	202.88	23.21	0.0485	12
10	178.97	16.02	0.0335	5
11	195.48	20.70	0.0432	10
12	206.42	24.52	0.0512	13
13	149.70	10.18	0.0213	1
14	187.75	18.36	0.0383	7
15	193.91	20.20	0.0422	9
16	241.82	42.45	0.0887	19
17	237.92	39.95	0.0835	19
18	179.91	16.26	0.0340	6
19	231.29	36.05	0.0753	18
20	201.22	22.62	0.0473	11
.				
.				
.				
997	190.27	19.09	0.0399	8
998	202.30	23.00	0.0481	11
999	189.09	18.74	0.0392	8
1000	206.25	24.46	0.0511	13
TOTAL		23936.91	50	

Table 5. Predicted relative probability of selection and bins for the 50 used points in the example validation used dataset.

Observation number	X	Predicted relative probability of selection	Scaled predicted relative probability of selection	Bin	Observation number	X	Predicted relative probability of selection	Scaled predicted relative probability of selection	Bin
1	178.03	15.79	0.0330	5	26	188.24	18.50	0.0386	8
2	208.60	25.36	0.0530	13	27	234.90	38.13	0.0797	18
3	200.37	22.33	0.0466	11	28	218.24	29.45	0.0615	15
4	209.64	25.77	0.0538	14	29	189.18	18.77	0.0392	8
5	197.87	21.48	0.0449	10	30	198.32	21.63	0.0452	10
6	233.47	37.29	0.0779	18	31	225.04	32.72	0.0684	17
7	205.70	24.25	0.0507	12	32	220.51	30.51	0.0637	16
8	194.48	20.38	0.0426	9	33	179.48	16.15	0.0337	5
9	211.88	26.68	0.0557	14	34	224.25	32.32	0.0675	17
10	187.97	18.42	0.0385	8	35	193.26	20.00	0.0418	9
11	204.48	23.79	0.0497	12	36	216.59	28.71	0.0600	15
12	215.42	28.19	0.0589	15	37	259.54	55.86	0.1167	20
13	158.70	11.70	0.0245	2	38	224.18	32.29	0.0675	17
14	196.75	21.11	0.0441	10	39	231.31	36.06	0.0753	18
15	202.91	23.22	0.0485	12	40	216.62	28.72	0.0600	15
16	250.82	48.80	0.1019	20	41	204.51	23.81	0.0497	12
17	246.92	45.93	0.0960	20	42	230.88	35.83	0.0748	18
18	188.91	18.69	0.0390	8	43	238.66	40.42	0.0844	19
19	240.29	41.45	0.0866	19	44	219.44	30.00	0.0627	16
20	210.22	26.01	0.0543	14	45	196.70	21.09	0.0441	10
21	252.86	50.36	0.1052	20	46	165.72	13.05	0.0273	3
22	180.09	16.30	0.0341	6	47	230.61	35.67	0.0745	18
23	202.68	23.14	0.0483	11	48	220.95	30.71	0.0642	16
24	188.12	18.46	0.0386	8	49	251.74	49.50	0.1034	20
25	234.51	37.90	0.0792	18	50	196.40	20.99	0.0439	10

Table 6. Percentiles of the scaled predicted relative probability of selection for the example validation available dataset. Observations in bin 1 had scaled predicted relative probability of selection less than or equal to the 5th percentile, observations in bin 2 had scaled predicted relative probability of selection greater than the 5th percentile and less than or equal to the 10th percentile, etc.

Percentile	Value	Percentile	Value
5th	0.0224	55th	0.0484
10th	0.0267	60th	0.0509
15th	0.0293	65th	0.0534
20th	0.0313	70th	0.0568
25th	0.0339	75th	0.0617
30th	0.0359	80th	0.0669
35th	0.0384	85th	0.0717
40th	0.0406	90th	0.0797
45th	0.0430	95th	0.0933
50th	0.0460	100th	0.1484

Table 7. Expected and observed counts for the sample validation used dataset.

Bin	Number of available points	Sum of unscaled predicted relative probability of selection	Expected Count (Sum of unscaled predicted relative probability of selection)	Observed count
1	50	448.26	0.94	0
2	50	588.86	1.23	1
3	50	671.89	1.40	1
4	50	724.66	1.51	0
5	50	779.81	1.63	2
6	50	833.36	1.74	1
7	50	890.32	1.86	0
8	50	944.10	1.97	5
9	50	1002.50	2.09	2
10	50	1061.96	2.22	5
11	50	1127.72	2.36	2
12	50	1186.30	2.48	4
13	50	1247.69	2.61	1
14	50	1317.32	2.75	3
15	50	1416.59	2.96	4
16	50	1537.91	3.21	3
17	50	1660.42	3.47	3
18	50	1801.93	3.76	6
19	50	2074.32	4.33	2
20	50	2620.98	5.47	5
TOTAL	1000	23936.90	50	50

APPENDIX A. Formulas for the calculation of expected and observed selection for the census and sample validation techniques.

Expected selection for the census validation method is the predicted RSF for each grid cell, summed by bin:

$$P_i = \sum_{j}^{N_i} \hat{y}_{ij}$$

where i is a bin indicator from 1 to b bins, j is a grid cell indicator from 1 to N grid cells in the study area, y-hat is the predicted relative probability of selection with the model under validation scaled to sum to 1.

Observed selection for the census validation method is the frequency of use points in each grid cell, summed by bin:

$$O_i = \sum_{j}^{N_i} \frac{r_{ij}}{r}$$

where r is the count of use in the study area in the validation dataset.

Expected selection for the sample validation method is the predicted RSF for each grid cell in the available sample used to generate the model, summed by bin:

$$P_i = \sum_{k}^{n_i} \hat{y}_{ik}$$

where k is a grid cell indicator from 1 to n grid cells in the available sample, y-hat is the predicted relative probability of selection with the model under validation scaled to the number of use points in the validation set, m.

Observed selection for the sample validation method is the count of the number of used points in each bin:

$$O_i = m_i$$

where m is the number of used points in the study area in the validation dataset.

APPENDIX B. Steps for conducting the sample validation technique.

1. Identify the model under validation.
2. Determine the appropriate available and used data for the validation dataset. The available dataset may be the same dataset used to create the model, but the used dataset for validation should be a new set of used points held out from the original data or obtained through additional sampling.
3. Use the model under validation to predict the relative probability of selection for both the available and used validation datasets.

Calculate the predicted selection
4. Scale the available data
 Sum the predicted relative probabilities of selection for the available data (Sum)
 Divide each value by Sum and multiply by the number of used points (n)
5. Calculate the 5^{th} to 100^{th} percentiles (by 5) of the scaled relative probability of selection for the available data.
6. Put each observation in the available dataset into bins based on the percentiles. Bin 1 will contain observations in the available dataset with the scaled predicted relative probability of selection less than or equal to the 5^{th} percentile. Bin 2 will contain observations in the available dataset with the scaled predicted relative probability of selection less than or equal to the 10^{th} percentile and greater than the 5^{th} percentile. Etc.
7. Sum the scaled predicted relative probabilities of selection of each observation in each bin. This is the expected selection for each bin. The sum of the expected selections across all bins will equal the number of used points in the validation data.

Calculate the observed selection

8. Put each observation in the used validation dataset into the same bins calculated with the above percentiles (based on available data). Bin 1 will contain observations in the used dataset with the scaled predicted relative probability of selection less than the 5^{th} percentile. Bin 2 will contain observations in the available dataset with the scaled predicted relative probability of selection less than the 10^{th} percentile and greater than the 5^{th} percentile. Etc.
9. Sum the number of used points in each bin. This is the observed selection for each bin. The sum of the observed selections should equal the number of used points in the validation data.

Validation
10. Regress the observed count on the predicted count using a normal linear regression model.
11. Use the slope and confidence interval of the slope of the validation regression and table 1 to evaluate the predictive abilities of the model.