

**Susitna-Watana Hydroelectric Project
(FERC No. 14241)**

**Fish and Aquatics Instream Flow Study
Study Plan Section 8.5**

**Discussion of Habitat Suitability Criteria Model
Validation**

Prepared for

Alaska Energy Authority



SUSITNA-WATANA HYDRO

Clean, reliable energy for the next 100 years.

Prepared by

R2 Resource Consultants

October 2016

TABLE OF CONTENTS

1.	Introduction.....	1
2.	Background and Methods	2
2.1.	Evaluation of Model Assumptions.....	3
2.1.1.	Independence of Predictor Variables	4
2.1.2.	Normality of Random Effects.....	4
2.1.3.	Compound Symmetry of Errors.....	4
2.1.4.	Correct Linear (Polynomial) Relationships	5
2.1.5.	Predictors Measured Without Error	5
2.1.6.	Selected Methods for HSC Validation.....	5
2.2.	Apparent Validity.....	5
2.2.1.	Selected Methods for HSC Validation.....	6
2.3.	Internal Validity	7
2.3.1.	Selected Methods for HSC Validation.....	8
2.4.	External Validity.....	8
3.	Results	8
3.1.	Evaluation of Model Assumptions.....	9
3.1.1.	Independence of Predictor Variables	9
3.1.2.	Normality of Random Effects.....	9
3.1.3.	Correct Polynomial Relationships	9
3.2.	Apparent Validity.....	10
3.3.	Internal Validity	10
4.	Conclusions.....	10
5.	Literature Cited	11
6.	Tables	15
7.	Figures.....	16

LIST OF TABLES

Table 3.2-1. Summary of apparent and internal model validity results for draft final Chum Salmon spawning HSC model. 15

LIST OF FIGURES

Figure 3.1-1. Histogram showing the distribution of random sampling event effects. Binned values are the difference of the intercept estimated for each site minus the estimated overall intercept for the model. 16

Figure 3.1-2. Scatter plot of model deviance residuals plotted against each predictor. Continuous predictors have a constant line at the mean (indicating no trend in residuals) and a local regression smoothed line fit to utilization data only. 17

Figure 3.2-1. Plot of observed versus predicted probabilities of occurrence for 10 equal-sized bins on range of (0,1) (left), and on nine alternative bit cutoff series (right). 18

LIST OF ACRONYMS AND SCIENTIFIC LABELS

ABBREVIATION	DEFINITION
AEA	Alaska Energy Authority
AIC	Akaike's Information Criteria
FERC	Federal Energy Regulatory Commission
GVI	Generalized Variance Inflation Factor
HSC	Habitat Suitability Criteria
HSI	Habitat Suitability Index
IFS	Fish and Aquatics Instream Flow Study 8.5
ISR	Initial Study Report
Project	Susitna-Watana Hydroelectric Project, FERC No. 14241
ROC	Receiver Operator Characteristic
RSP	Revised Study Plan
SIR	2014-2015 Study Implementation Report
TM	Technical Memorandum
TWG	Technical Workgroup
VIF	Variance Inflation Factor

1. INTRODUCTION

On December 14, 2012, Alaska Energy Authority (AEA) filed with the Federal Energy Regulatory Commission (FERC) its Revised Study Plan (RSP), which included 58 individual study plans (AEA 2012). The FERC-approved RSP Section 8.5 (Fish and Aquatics Instream Flow Study [IFS]) (AEA 2012) as modified by FERC's April 1, 2013 Study Plan Determination (FERC 2013), the *Open Water HEC-RAS Flow Routing Model Technical Memorandum (TM)* (R2 et al. 2013), the *Selection of Focus Areas and Study Sites in the Middle and Lower Susitna River for Instream Flow and Joint Resource Studies – 2013 and 2014 TM* (R2 2013a), and the *Adjustments to Middle River Focus Areas TM* (R2 2013b) is collectively referred to herein as the "Study Plan Section 8.5." Study Plan Section 8.5 focused on establishing an understanding of important biological communities and associated habitats, and of the hydrologic, physical, and chemical processes in the Susitna River that directly influence those communities and habitats. Study Plan Section 8.5 also described the study methods that would be used to evaluate Susitna-Watana Hydroelectric Project, FERC No. 14241 (Project) effects, including the selection of study sites, collection of field data, data analysis, and modeling.

The goal of the IFS and its component study efforts was to provide quantitative indices of existing aquatic habitats that enable a determination of the effects of alternative Project operational scenarios. As part of this effort and represented in Objective 4 of the IFS Study Plan (RSP Section 8.5.4.5) (AEA 2012), AEA is developing site-specific Habitat Suitability Criteria (HSC) and Habitat Suitability Indices (HSI) for various species and life stages of fish for biologically relevant time periods. The development of HSC/HSI criteria has been a central topic for discussion during a number of Technical Workgroup (TWG) and Technical Team meetings in 2013 and 2014, with methods more fully described in RSP Section 8.5.2.6. Detailed information describing the data collection and analytical methods being used for HSC development were provided in the Study 8.5 Initial Study Report (ISR) Part A, Section 4.5 (R2 2014a); Study 8.5 ISR Part C, Appendix M (*Habitat Suitability Curve Development*) (R2 2014b); Study 8.5 2014-2015 Study Implementation Report (SIR), Section 4.5 (R2 2015a); and most recently and comprehensively in Study 8.5 SIR, Appendix D (*Habitat Suitability Curve Development*) (R2 2015b).

To date, and after two years (2013 and 2014) of HSC data collection, over 7,000 site-specific microhabitat use and availability measurements have been collected during 267 unique sampling events (Study 8.5 SIR, Appendix D, Tables 5.2-1 and 5.2-2 [R2 2015b]). This has resulted in the development of 12 statistically robust multivariate HSC models that represent select life stages of nine high priority fish species. Specifically, models were developed for Chinook Salmon (*Oncorhynchus tshawytscha*) fry and juvenile, Chum Salmon (*O. keta*) spawning, Coho Salmon (*O. kisutch*) fry and juvenile, Sockeye Salmon (*O. nerka*) spawning, Arctic Grayling (*Thymallus arcticus*) fry and juvenile, whitefish fry and juvenile, and Longnose Sucker (*Catostomus catostomus*) juvenile and adult. These models have been delivered to FERC in the IFS Study 8.5 Initial Study Report (ISR) Part D (R2 2015a).

Although preliminary statistical model validation was applied during the modeling process, formal discussion of validation of these draft models has not been provided to FERC. This Technical Memorandum (TM) (a supplement to the Study 8.5 2014-2015 Study Implementation

Report [November 9, 2015] and filed with FERC as Attachment 5 to *Response of the Alaska Energy Authority to Comments on the Initial Study Report*) provides an overview of model validation methods and proposes those most relevant for the statistical HSC models developed for the Susitna River.

For the purposes of this discussion, validation is defined based on three important questions regarding a statistical model:

1. Is the model appropriate for the data and the question being asked?
2. How well does the model fit the data?
3. How well does the model function in the application for which it is intended?

The following sections present background information and discuss the appropriateness of various validation methods for the specific case of the Susitna River HSC models. Selected relevant methods are then applied to an example draft HSC model, Chum Salmon (*Oncorhynchus keta*) spawning, and discussed relative to the validation process. Confidence or prediction intervals are not presented in this TM, as they are not a form of statistical validation. However, confidence intervals for one HSC model are presented in *Decision Support System Uncertainty Technical Memorandum* (a supplement to the Study 8.5 2014-2015 Study Implementation Report [November 9, 2015] and filed with FERC as Attachment 6 to *Response of the Alaska Energy Authority to Comments on the Initial Study Report*).

2. BACKGROUND AND METHODS

The generalized polynomial mixed effects HSC models presented in the Study 8.5 SIR, Appendix D (R2 2015b) provide a relative ranking of habitat use for each species and life stage. This ranking is related to the probability of fish using a particular habitat, but because the model is based on presence/available habitat data rather than presence/absence data, it is only proportional to that probability (Johnson et al. 2006). The ranking is *relative* because the mixed effects model essentially averages the habitat relationships across different sampling events and sites in the river with potential different overall levels of fish prevalence. Some species of fish, for example, may be more commonly found in the Lower River Segment of the Susitna River or in the vicinity of particular tributaries.

This HSC/HSI process is not well-suited to determining large-scale locations of fish use, or what areas of a large river system have or will have the highest prevalence of fish. Sampling for such a purpose would require a hierarchical sampling scheme with sufficient replication at multiple spatial scales to estimate all variables impacting fish habitat selection. Instream flow models are focused on a subset of reach-scale variables that have a predictable relationship with flow. Although multiple 50/100 meter reaches have been sampled, they are sampled as clusters of spatial polygons with varying potential as fish habitat. The overall level of use of the reaches varies widely for each species and life stage. There are three options for addressing this differential use. First, the differences among reaches not explainable by the reach-scale variables could simply be ignored. This would lead to large residual variance and uneven precision in habitat predictions across the river. More problematic is the combination of two levels of

sampling effort assuming independence among all sampled habitat units. The second option would be to estimate a different fixed intercept for each sampling event. This option is untenable because the model would then only be applicable to the sampled sites and could not be extended to the rest of the river. The third option, and that selected for the Susitna River HSC models, is to add sampling event into the model as a random effect. The resulting mixed effects model removes these larger-scale influences by assuming the overall fish prevalence for each sampling event is a random variable with an approximately normal distribution. This model formulation conveniently accounts for lack of independence among samples within a site with the assumption of compound symmetry. Random effects models are generally considered to provide more robust ecological inferences (Gillies et al. 2006).

Attempts to validate HSC in the literature are limited and tend to focus on additional sampling to determine whether more fish are found in areas predicted to have higher suitability by HSC models (Guay et al. 2000). Although this method of validation may be useful for smaller study areas and studies with presence/absence data (rather than presence/availability data), it is not applicable here, because there is no consideration of the spatial and temporal variability in prevalence (species abundance), nor in the difference between absence and availability data (Phillips and Elith 2013). For example, sites with highly suitable in-stream habitat within low-prevalence areas would not necessarily be expected to contain higher numbers of fish than sites in high-prevalence areas with lower suitability.

More appropriate validation methods for the Susitna River HSC models fall into the following four categories that are discussed in detail below: 1) evaluation of model assumptions; 2) apparent validity; 3) internal validity; and 4) external validity. All statistical analyses were performed in R statistical software (version 3.1.1; R Core Team 2014).

2.1. Evaluation of Model Assumptions

The first method for assessing the validity of a statistical model is to verify the model assumptions. This is an evaluation of whether or not the selection of the best-fit model was a valid process which should yield reasonable predictions for the population of interest. In addition to the basic assumption that the outcome variable is binary (0/1), the assumptions of the multivariate mixed effects logistic regressions used for HSC modeling are:

1. predictor variables are independent;
2. random site effects are approximately normally distributed;
3. errors follow compound symmetry structure of dependence;
4. relationship between the selected independent and dependent variables is linear; and
5. independent variables are measured without error.

Note that logistic regression does not make any assumptions of normality or homogeneity of variance.

2.1.1. Independence of Predictor Variables

Correlation or lack of independence among predictors in a multiple regression is labeled multicollinearity, and it impacts the precision of individual regression coefficients and their interpretation. It does not impact the strength or predictive capabilities of the model if the prediction space is in the range of data used to fit the model, and if the correlations among variables also remain the same (Neter et al. 1990). However, the relative importance of each predictor cannot be interpreted based on the magnitude of the regression coefficients, which are subject to change depending on which variables are included in the model. Because HSC models are mainly concerned with prediction (e.g., which habitats are most suitable to spawning) as opposed to interpretation of most important habitat covariates, collinearity is not an important issue for the HSC analysis. Nonetheless, variance inflation factors (VIFs) can be used as a formal check of collinearity. The square-root of the VIF is an estimate of the multiplicative inflation of the confidence interval around the coefficient estimates. When some predictors are categorical with more than two levels (i.e., more than one degree of freedom), the generalized variance inflation factors (GVIFs) are used with a similar interpretation (Fox and Monette 1992). If there are p degrees of freedom in a term, then $GVIF^{1/2p}$ is a one-dimensional expression of the decrease in the precision of estimation due to collinearity. Polynomials of the same variable are inherently correlated, but do not need to be interpreted separately. The VIF is estimated for the model without random effects because it is unclear how to estimate the VIF for mixed effects models. Some authors recommend that $VIF > 10$ indicates a problem with collinearity (Ahmadi-Nedushan et al. 2006; Neter et al. 1990). The dependencies among the utilized predictors were not found to be an issue impacting any of the HSC models.

2.1.2. Normality of Random Effects

The assumption of normal random effects is routine in mixed effects models, and studies indicate that most aspects of inference are highly robust to violations of this assumption (McCulloch and Neuhaus 2011). The exception to this is a possible bias in the estimation of the intercept or overall average of, in our case, fish prevalence in the river. According to a review of the published literature by McCulloch and Neuhaus (2011), multiple studies on non-normality of the random effects distribution have confirmed little impact on within-cluster covariates, which include the habitat covariates in the HSC models. Rather, the impact is to the estimates of the random effects (i.e., the mean prevalence at each site). These random effect values are not used in the HSC model, so the impact from non-normal random effects should be minimal. Although the random effects were highly skewed in the Susitna River HSC models, the skewness was reduced by excluding sites where the modeled species and life stage were not found.

2.1.3. Compound Symmetry of Errors

A mixed effects regression was used for the HSC analysis largely because the errors were not expected to be independent if the data from all sampling events were analyzed together. The mixed effects model corrects for this lack of dependence due to sampling events with an assumption of compound symmetry, which essentially means that all samples within a sampling event are equally correlated with each other. This assumption would be violated if there were spatial autocorrelation that diminishes with distance among utilization locations within a sampling event. This is one reason that multiple fish utilizing the same microhabitats were

counted as a single observation in the analysis. This assumption is difficult to test with binary data, and such a test would require large sample sizes within sites. For example, residuals could be plotted against x-y location (separately by site) to look for a non-random distribution.

2.1.4. Correct Linear (Polynomial) Relationships

The HSC models were fit with polynomial regressions, allowing restricted non-linear relationships between predictors and preference. Using Akaike's Information Criteria (AIC), the best-fit model was selected from the choice of linear, quadratic, or third-order polynomial models. An assessment of how closely the data match with the assumed relationship is typically a visual assessment, which is somewhat complicated by having 0/1 categorical data. However, the residuals can be plotted with a smoothed fit to visually compare the model to the assumed relationship.

2.1.5. Predictors Measured Without Error

Unless specialized methods are applied, all standard regression methods assume that the predictor variables are measured without error. This assumption is not met, and cannot realistically be met for field measurements. The result of violation of this assumption, quite common in ecological applications, is an unknown bias in the coefficient estimates (Neter et al. 1990).

2.1.6. Selected Methods for HSC Validation

Based on the above discussion, three indicators are proposed to assess the validity of the model assumptions for the mixed effects logistic regressions on presence/availability data:

1. variance inflation factors: <10 indicates that multicollinearity is not problematic;
2. distribution of random effects: provided as information only – model is robust to violations; and
3. residuals plotted against the predictors with smoothed fit: a general assessment of the usefulness of polynomial regression for fitting the relationships – large outliers would also be evident from these plots.

These indicators were used in the example Chum Salmon spawning Susitna River HSC model in Section 3.1.

2.2. Apparent Validity

Apparent validity or “goodness-of-fit” is an evaluation of model performance on the existing data. There are five standard statistics typically used for apparent validity in logistic regression:

1. sensitivity – the proportion of true positives correctly identified by the model;
2. specificity – the proportion of true negatives correctly identified by the model;
3. percent correctly classified by the model;

4. Cohen's kappa statistic – percent correctly classified adjusted for chance predictions; and
5. area under the receiver operator characteristic (ROC) curve – a measure of model accuracy across all potential “thresholds”, where the threshold is the probability value used to discriminate between a 0/1 classification.

The first four metrics depend upon a selected threshold probability, interpreted as a prediction of species presence (e.g., <50 percent model prediction is interpreted as “absence”; >50 percent = presence). The proper threshold to use is unclear. For example, if the maximum probability predicted by the model is less than 0.5 (i.e., a relatively rare species), selecting a threshold of 0.50 would mean only absences are predicted by the model. A standard process for threshold selection is to choose the threshold that maximizes Cohen's kappa statistic (Boyce et al. 2002). Methods 3 through 5 are based on a combination of sensitivity (true positives) and specificity (true negatives).

One difficulty in evaluating the validity of the Susitna River HSC models using these statistics is that the existing data are utilization and availability data, rather than utilization and unused habitat data. In other words, the “0” data represent total habitat, not necessarily “absence”. It is certainly possible (albeit unlikely) that one exact point in the river is recorded as both a utilization and availability point. Thus, it is not sensible to judge the performance of the model in terms of predicting “0” responses (i.e., where fish will not be), since these locations are not expected to be necessarily unsuitable habitats. Of the apparent validity metrics listed above, only the sensitivity metric exclusively applies to the utilization points, and is therefore a potentially sensible measure of model performance on used/available datasets. But this measure remains dependent on the selected threshold value.

Boyce et al. (2002) recommend a revised sensitivity index (hereafter referred to as the Boyce Index) for presence/availability modeling that does not rely upon a single threshold probability, but looks at the percent of observations that were utilized in binned prediction values. An increasing relationship between the bin mid-points and the percent utilized, as measured by a significant non-parametric correlation, would indicate a valid model. Hirzel et al. (2006) further modified the Boyce Index using a moving-window sensitivity assessment that further does not depend on the choice of bin endpoints.

No matter what measures are used, apparent validity is limited to the data used to fit the model, and therefore may not be reflective of the actual performance of the model in terms of prediction of new data (Logo et al. 2007; Manel et al. 2001). However, these statistics are relatively simple to estimate, and they are in standard use as measures of model validity (Beakes et al. 2014).

2.2.1. Selected Methods for HSC Validation

For apparent validity, three metrics were computed and presented in this report: 1) model sensitivity with 0.50 and the maximum Cohen's kappa threshold as two objective thresholds; 2) the Boyce Index (Boyce et al. 2002); and 3) Hirzel's modified Boyce Index (Hirzel et al. 2006). The sensitivity and Boyce Index were estimated using the *PresenceAbsence* library (Freeman and Moisen 2008) in *R* (version 3.1.1; R Core Team 2014). The Hirzel approach was further modified, as their application seems to provide an arbitrarily increasing sample size for the non-parametric correlation test by including points estimated from all bin cutoff points in a single

correlation test. Their use of a moving window was followed, but rather than reporting a single non-parametric correlation test, results from each window were provided separately. The following set of bin cutoffs was used for the evaluation:

(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1) [Original bins]

(0.01, 0.11, 0.21, 0.31, 0.41, 0.51, 0.61, 0.71, 0.81, 0.91) [Note: for these and following bins, the first and last bin (e.g., [0, 0.1], and [.91, 1]) are not included because they are not the same size as the other bins.]

(0.02, 0.12, 0.22, 0.32, 0.42, 0.52, 0.62, 0.72, 0.82, 0.92)

(0.03, 0.13, 0.23, 0.33, 0.43, 0.53, 0.63, 0.73, 0.83, 0.93)

(0.04, 0.14, 0.24, 0.34, 0.44, 0.54, 0.64, 0.74, 0.84, 0.94)

(0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95)

(0.06, 0.16, 0.26, 0.36, 0.46, 0.56, 0.66, 0.76, 0.86, 0.96)

(0.07, 0.17, 0.27, 0.37, 0.47, 0.57, 0.67, 0.77, 0.87, 0.97)

(0.08, 0.18, 0.28, 0.38, 0.48, 0.58, 0.68, 0.78, 0.88, 0.98)

(0.09, 0.19, 0.29, 0.39, 0.49, 0.59, 0.69, 0.79, 0.89, 0.99)

2.3. Internal Validity

Internal validity (also known as cross-validation) is an evaluation of the reproducibility of the model. For example, would a similar model arise if different sample data were selected from the same underlying population? Methods for internal validity include splitting the dataset into separate calibration and validation subsets (Buisson et al. 2008), *k*-fold cross-validation (Kornis and Vander Zanden 2010; Boyce et al. 2002), and bootstrapping (Steyerberg et al. 2001).

Splitting the dataset for calibration and validation is a process that may be reasonable when ample data are available and can be randomly assigned to these two groups. For the HSC analysis, such an assignment would be problematic due to limited sample sizes and the grouped nature of the data (i.e., random site effects). The *k*-fold cross-validation method (Fielding and Bell 1997) is an expansion of the data splitting method, whereby a random portion of the dataset (of sample size *k*) is left out of the model fitting process, then used as a test case for the model fit on the reduced dataset. This process is repeated until all of the data have been left out in turn. The extreme version of this method is jackknifing: leaving out one sample at a time (*k*=1). The *k*-fold cross validation method is often recommended and applied (Olden et al. 2002; Mull and Wilzbach 2007; McRae et al. 2012), but Fang (2011) points out that the results are asymptotically equivalent to the AIC criteria, which have already been used to fit all HSC models. Also, for the mixed effects model, because the samples are grouped into sites by the random effect, it makes more sense to leave out one sampling event at a time, rather than single samples.

The bootstrap procedure is strongly recommended as the best internal validation technique by Steyerberg et al. (2001), because the *k*-fold cross-validation measure is inefficient and can be biased. For bootstrapping validation, a sample from the dataset is selected *with* replacement, and

the model is refit (with the same predictors) and tested for goodness-of-fit on both the original dataset and the bootstrapped dataset. The goodness-of-fit to the bootstrap sample is a measure of apparent validity, whereas the goodness-of-fit to the original sample is a measure of internal validity. The average difference between these two measures is an estimate of “optimism” or optimistic bias in the apparent validity in the original model. This bias is subtracted from the measures used for apparent validity (Section 2.2) for an overall measure of internal model validity.

2.3.1. Selected Methods for HSC Validation

For the mixed model presence/availability approach used for the Susitna River HSC models, the bootstrapping method, applied as described above, is proposed as most appropriate. There are two levels of sampling in the mixed effects model, the sampling event, and the locations within each sampling event, so there are two levels of sampling in the bootstrap. First, the sampling events used in the model are sampled with replacement. Then, within each selected sampling event, the utilization sites are also selected with replacement. The availability locations are a measure of available habitat, and sampling these locations with replacement would lead to skewed, nonsensical habitat assessments within sites. Instead, the original availability locations are retained for each sampling event bootstrap.

For each bootstrap sample, the sensitivity at 0.50 and at the maximum kappa threshold were estimated, as well as the Boyce statistic with ten bins (apparent validity for bootstrap sample). Then, these same statistics are calculated for the bootstrap sample model predictions for the original data (internal validity for the bootstrap sample). The average difference (over 1,000 bootstrap samples) between the apparent and internal validity statistics for the bootstrap model is a measure of optimism for the apparent validity statistics estimated on the original model. This optimism statistic is then subtracted from the apparent validity (described in Section 2.2) for an estimate of internal validity.

2.4. External Validity

External validity is an evaluation of whether the model is transportable (i.e., whether it would apply to data collected outside of the current sampling strategy and area). An evaluation of external validity would require a separate data collection process, and it is not clear whether a measure of external validity would be useful for the HSC/HSI process, which is meant only to apply to the Susitna River. As a result, external validity was not measured as part of the analysis.

3. RESULTS

The draft Susitna River HSC multivariate model for Chum Salmon spawning (R2 2015b) is:

$$\log\left(\frac{\widehat{p}}{1-\widehat{p}}\right) = C_k + 0.999depth - 0.155depth^2 + 0.408vel - 1.23vel^2 \\ - 0.225temp + 0.247(vel * temp),$$

where:

p is the approximate probability of Chum Salmon spawning,

k indexes four intercept values for substrate/upwelling combinations:

$C_1 = 0.811$ (all gravel substrate)

$C_2 = 0.382$ (gravel dominant mixed substrate)

$C_3 = -0.131$ (gravel subdominant mixed substrate)

$C_4 = -0.999$ (no gravel, but cobble dominant),

(Note: other substrates are assumed NOT suitable for spawning, HSC=0)

The methods for model validation selected and described above were applied to the Chum Salmon spawning HSC model as an example of how model validation can be reported.

3.1. Evaluation of Model Assumptions

3.1.1. Independence of Predictor Variables

As reported in the Study 8.5 SIR, Appendix D (R2 2015b), the highest adjusted VIF for the full fixed-effects model (including all potential predictor variables) for Chum Salmon spawning was 1.40, indicating that confidence intervals around predicted coefficients may be 18 percent (the square-root of 1.40 = 1.18) wider than they would be with uncorrelated predictors. This VIF (1.4) was well below the threshold of 10 typically used to indicate a concern for multicollinearity.

3.1.2. Normality of Random Effects

The distribution of random effects is displayed in Figure 3.1-1. The distribution is symmetric, but is leptokurtic (i.e., thinner peak than normal distribution). Thus, there is some minor deviation from normal distribution.

3.1.3. Correct Polynomial Relationships

Three residual plots for continuous covariates are displayed in Figure 3.1-2 to visually evaluate if there are large departures from the functional relationships selected by the model for depth, velocity, and stream temperature. Because the data are presence-availability data, the residuals from the availability points are of no concern (i.e., all of the [observed – fitted] residuals in the lower cloud of points below zero are from availability points). The blue smoothed lines fit to the utilization points only do not indicate that the linear and quadratic fits were systematically biased. The boxplot of residuals versus substrate type in the lower right corner of Figure 3.1-2 is for utilization points only, and illustrate that utilization points in cobble substrates (with no gravel) were the most poorly fit (i.e., greatest departures from 1). That means that the chum that were spawning in cobble substrates were spawning in areas which were predicted to be low suitability by the model, i.e., the model is under predicting suitability of cobble substrates.

In addition to systematic departures for functional relationships, the plots in Figure 3.1-2 can also be used to identify individual sample outliers; the plots have a dotted line at deviance residual =

2 to highlight points that are fit the most poorly (these are utilized points with very low predicted probabilities). The outlying sample points are all located in cobble substrate with zero velocity and depths less than 1 foot.

3.2. Apparent Validity

The threshold value that resulted in the highest value for Cohen's kappa statistic from the Chum Salmon spawning model was a predicted probability of 0.54. Of the observed utilization data points, 78 percent had predicted probabilities of use > 0.50 , and 69 percent had predicted probabilities > 0.54 (Table 3.2-1). Thus, the sensitivity of the model is in the range of 69-78 percent, depending upon the selected threshold.

The observed versus predicted probabilities are displayed for ten bins in the left plot of Figure 3.2-1. There is an obvious increasing relationship, showing good predictive capability (Spearman's correlation coefficient = 0.98, $p = 0.00005$; Table 3.2-1). The slight exception to this high agreement is in the highest bin. Thus, locations with model predictions of 80-90 percent probability of utilization have substantially lower observed probability of utilization.

Using nine alternative moving windows for the bin cutoffs along with the original 10 bins, there are ten potential comparisons between observed and predicted probabilities, as displayed by 10 point types in the right plot in Figure 3.2-1. The Spearman correlation coefficients for these 10 series range from 0.88 to 1.0, with significance p -values ranging from 0.000006 to 0.003, indicating good to very good predictive capability.

3.3. Internal Validity

Apparent and internal validity estimates for the bootstrap models (mean of estimates from 1,000 models) are displayed in Table 3.2-1. The average optimism for sensitivity was 1.5 percent when using a probability threshold based on the maximum Kappa result, and 1.2 percent for the 50 percent threshold. These optimism estimates can be used to adjust the apparent validity estimated derived in Section 3.2 to a more realistic estimate of overall model sensitivity. The Boyce Index results showed that the bootstrap models were not consistently better at predicting the observed probability of occurrence in the bootstrap sites used for calibrating the model than they were for the original data set. Thus, there was no evident "optimism" in the Boyce estimate of apparent validity (0.98).

4. CONCLUSIONS

For the draft Susitna River Chum Salmon spawning HSC model, most modeling assumptions have been shown to be reasonable. There is some evidence that the model fits better for gravel substrates, and does not fit as well for spawning locations in cobble substrates. There are fewer data points in this substrate group (6 percent of utilization observations), so the reduced accuracy may just be a function of the group not being equally represented in the model fitting. However, it does indicate that the Chum Salmon spawning HSC could be improved by more investigation into the selection of habitat within cobble substrates.

The Susitna River Chum Salmon spawning HSC model has good apparent validity for sensitivity (predicting true utilization locations; Table 3.2-1) ranging from 69-78 percent depending on what objective threshold is used as a cutoff for predicting redd presence. According to the validity measure recommended by Boyce et al. (2002) and Hirzel et al. (2006), the model provides very good agreement between observed and predicted values, with the correlation estimate of 0.98 (ranging from 0.88 to 1 depending on bin cut-off values). As displayed in the graph in Figure 3.2-1, the model may be over-predicting species occurrences in the most suitable modeled habitats, which suggests caution in over-interpreting the “highly suitable” category. There is also some over-prediction in the lower range near 20 percent probability, suggesting that the suitability of habitat does not rise as quickly as the model would predict.

Another consideration when assessing model validity is whether reasonable alternatives exist if the model validation indicates less than ideal fit. For example, what alternative models would provide more valid results, or would the issues be solved with further data collection. Alternative models would, in turn, have to be compared using the same validation approaches described here. There is no question that all models are improved by more data, but it is often difficult to define a standard or objective target for the sufficiency of sampling.

A more complete assessment of validity measures for all HSC models will be provided in the Updated Study Report.

5. LITERATURE CITED

- Alaska Energy Authority (AEA). 2012. Revised Study Plan. Susitna-Watana Hydroelectric Project, FERC Project No. 14241 Submittal: December 14, 2012. <http://www.susitna-watanahydro.org/study-plan>.
- Ahmadi-Nedushan, B., A. St-Hilaire, M. Berube, E. Robichaud, N. Thiemonge, and B. Bobee. 2006. A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. *River Research and Applications* 22(5):503-523.
- Beakes, M.P, J.W. Moore, N. Retford, R. Brown, J.E. Merz, and S.M. Sogard. 2014. Evaluating statistical approaches to quantifying juvenile Chinook salmon habitat in a regulated California river. *River Research and Applications* 30(2):180-191. <http://onlinelibrary.wiley.com/doi/10.1002/rra.2632/abstract>.
- Boyce, M.S., P.R. Vernier, S.E. Nielsen, F.K.A. Schmiegelow. 2002. Evaluating resource selection functions. *Ecological Modelling* 157:281-300
- Buisson, L., L. Blanc, and G. Grenouillet. 2008. Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. *Ecology of Freshwater Fish* 17:244-257.
- Fang, Y. 2011. Asymptotic equivalence between cross-validations and akaike information criteria in mixed-effects models. *Journal of Data Science* 9:15-21.

- Federal Energy Regulatory Commission (FERC). 2013. Study Plan Determination on 14 remaining studies for the Susitna-Watana Hydroelectric Project. Susitna-Watana Hydroelectric Project, FERC No. P-14241. April 1, 2013. http://elibrary.FERC.gov/idmws/file_list.asp?accession_num=20130401-3022.
- Fielding, A.H. and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 1:38-49.
- Fox, J. and G. Monette. 1992. Generalized collinearity diagnostics. *Journal of the American Statistical Association* 87:178–183.
- Freeman, E.A. and Moisen, G. 2008. PresenceAbsence: an R Package for presence-absence model analysis. *Journal of Statistical Software* 23(11):1-31.
- Freeman, E.A., and G.G. Moisen. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling* 217(1):48-58.
- Gillies, C.S., M. Hebblewhite, S.E. Nielsen, M.A. Krawchuk, C.L. Aldridge, J.L. Frair, D.J. Saher, C.E. Stevens, and C.L. Jerde. 2006. Application of random effects to the study of resource selection by animals. *Journal of Animal Ecology* 75:887-898.
- Guay, J.C., D. Boisclair, D. Rioux, M. Leclerc, M. Lapointe, and P. Legendre. 2000. Development and validation of numerical habitat models for juveniles of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* 57:2065-2075.
- Guisan, A. and N.E. Zimmerman. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.
- Hirzel, A.H., B. Le Lay, V. Helfer, C. Randin, and A. Guisan. 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199:142-152.
- Johnson, C.J., S. Nielsen, E. Merrill, T. McDonald, and M. Boyce. 2006. Resource selection functions based on use–availability data: theoretical motivation and evaluation methods. *Journal of Wildlife Management* 70(2):347-357.
- Kornis, M.S. and M.J. Vander Zanden. 2010. Forecasting the distribution of the invasive round goby (*Neogobius melanostomus*) in Wisconsin tributaries to Lake Michigan. *Canadian Journal of Fisheries and Aquatic Sciences* 67(3):553-562.
- Logo, J.M., A. Jimenez-Valverde, and R. Real. 2007. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*. 17(2):145-151.
- Manel, S., H.C. Williams, and S.J. Ormerod. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38:921-931.
- McCulloch, C.E. and J.M. Neuhaus. 2011. Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science* 26:388-402.

- McRae, C.J., K.D. Warren, and J.M. Shrimpton. 2012. Spawning site selection in interior Fraser River coho salmon *Oncorhynchus kisutch*: an imperiled population of anadromous salmon from a snow-dominated watershed. *Endangered Species Research* 16:249-260.
- Mull, K.E. and M.A. Wilzbach. 2007. Selection of spawning sites by coho salmon in a northern California Stream. *North American Journal of Fisheries Management* 27:1343-1354.
- Neter, J., W. Wasserman, and M.H. Kutner. 1990. *Applied Linear Statistical Models*. Third Edition. Irwin. Homewood, Illinois.
- Olden, J.D., D.A. Jackson, and P.R. Peres-Neto. 2002. Predictive models of fish species distributions: A note on proper validation and chance predictions. *Transactions of the American Fisheries Society* 131:329-336.
- Phillips, S.J. and J. Elith. 2013. On estimating probability of presence from use-availability or presence-background data. *Ecology* 94(6):1409-1419.
- R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>.
- R2 Resource Consultants, Inc. (R2). 2013a. Selection of Focus Areas and Study Sites in the Middle and Lower Susitna River for Instream Flow and Joint Resource Studies – 2013 and 2014. Susitna-Watana Hydroelectric Project, FERC No. P-14241 Submittal: March 1, 2013, Attachment C, Joint Resource Study Technical Memorandum. Prepared for Alaska Energy Authority, Anchorage, Alaska. <http://www.susitna-watanahydro.org/wp-content/uploads/2013/09/TechMemoSelectionOfFocusAreas.pdf>.
- R2 Resource Consultants, Inc. (R2). 2013b. Adjustments to Middle River Focus Areas. Susitna-Watana Hydroelectric Project, FERC No. P-14241 Submittal: May 31, 2013, Joint Resource Study Technical Memorandum. Prepared for Alaska Energy Authority, Anchorage, Alaska. http://www.susitna-watanahydro.org/wp-content/uploads/2013/05/SuWa_R2_TM-FocusAreasAdjustments.pdf.
- R2 Resource Consultants, Inc. (R2), Geo-Watersheds Scientific (GWS), Brailey Hydrologic, and Geovera. 2013. Open-water HEC-RAS Flow Routing Model. Susitna-Watana Hydroelectric Project, FERC No. P-14241 Submittal: January 31, 2013, Attachment A, Technical Memorandum. Prepared for Alaska Energy Authority, Anchorage, Alaska. <http://www.susitna-watanahydro.org/wp-content/uploads/2013/02/Att-A-Flow-Routing-Model.pdf>.
- R2 Resource Consultants, Inc. (R2). 2015a. Supplemental Information to June 2014 Initial Study Report. Susitna-Watana Hydroelectric Project, FERC No. P-14241 Submittal: November 6, 2015, Initial Study Report, Study 8.5, Part D. Prepared for Alaska Energy Authority, Anchorage, Alaska. http://www.susitna-watanahydro.org/wp-content/uploads/2015/11/08.5_IFS_ISR_PartD.pdf.

R2 Resource Consultants, Inc. (R2). 2015b. Habitat Suitability Criteria Development. Susitna-Watana Hydroelectric Project, FERC No. P-14241 Submittal: November 6, 2015, 2014-2015 Study Implementation Report, Study 8.5, Appendix D. Prepared for Alaska Energy Authority, Anchorage, Alaska. http://www.susitna-watanahydro.org/wp-content/uploads/2015/11/08.5_IFS_SIR_App_D_HSC.pdf.

Steyerberg, E.W., F.E. Harrell Jr., G.J.J.M. Borsboom, M.J.C. Eijkemans, Y. Vergouwe, and J.D.F. Habbema. 2001. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology* 54:774-781.

6. TABLES

Table 3.2-1. Summary of apparent and internal model validity results for draft final Chum Salmon spawning HSC model.

	Apparent Validity	Bootstrap Model Apparent Validity	Bootstrap Model Internal Validity	Average Difference Apparent-Internal for Bootstrap	Estimated Internal Validity for Final Model
Sensitivity at Max Kappa Threshold	0.692	0.8	0.785	0.015	0.678
Sensitivity at 0.50 Threshold	0.778	0.758	0.746	0.012	0.766
Boyce Index	0.98	0.932	0.952	n/a ¹	0.98

Notes:

¹Correlation of observed and expected bin proportions was not consistently higher with bootstrap applied to the calibration data (median difference was zero).

7. FIGURES

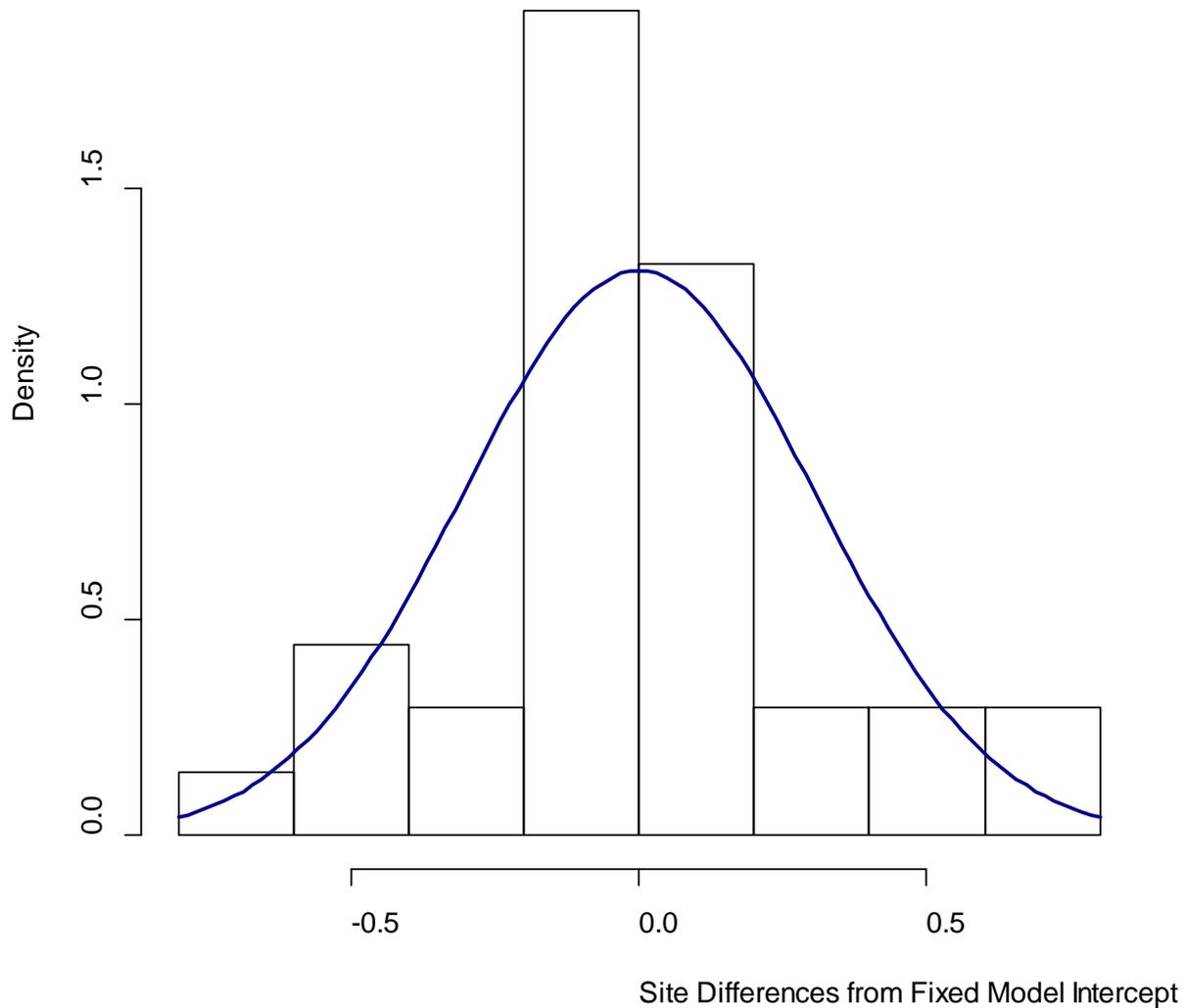


Figure 3.1-1. Histogram showing the distribution of random sampling event effects. Binned values are the difference of the intercept estimated for each site minus the estimated overall intercept for the model.

Note: A normal distribution with zero mean and observed variance is overlaid for reference.

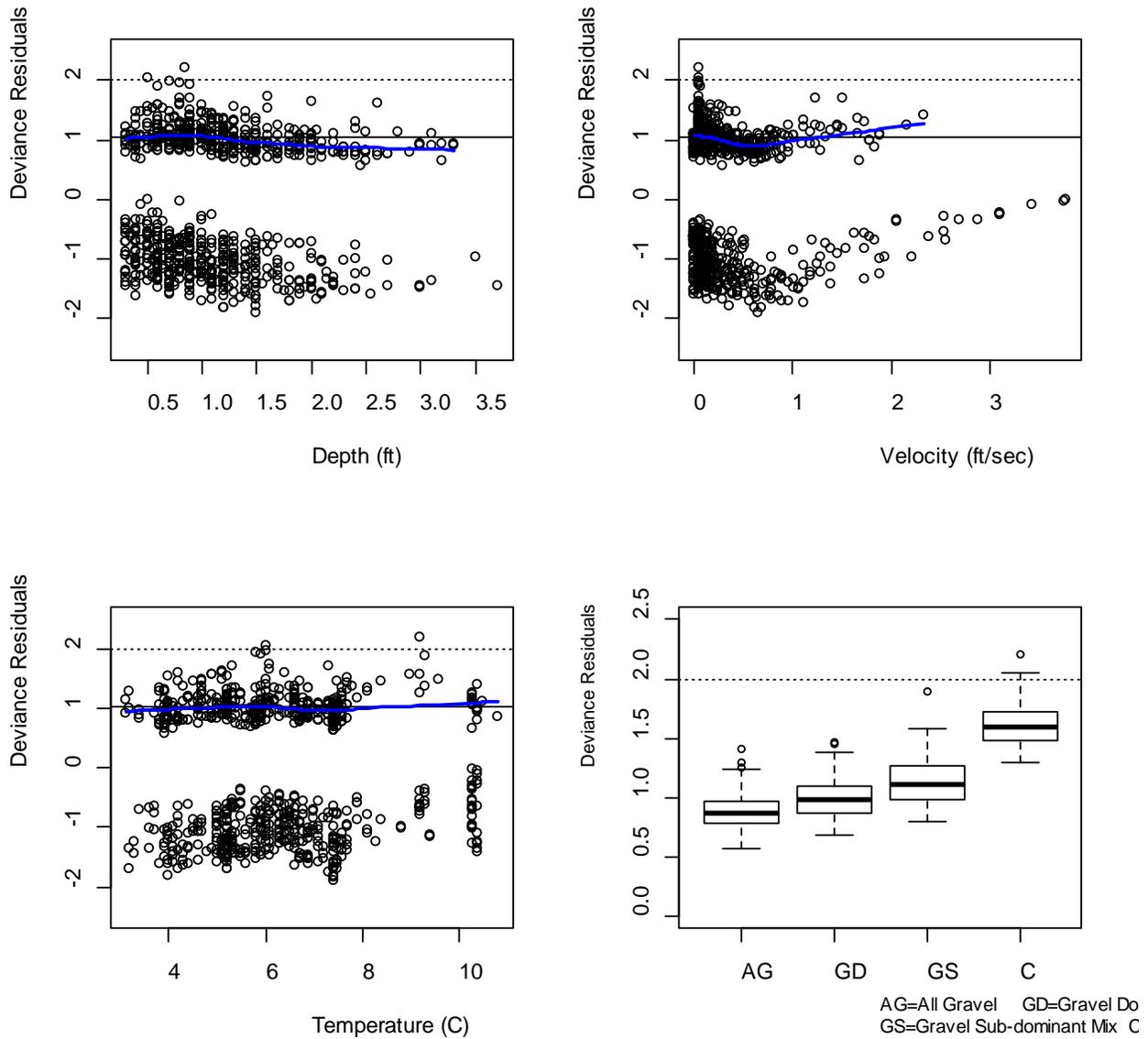


Figure 3.1-2. Scatter plot of model deviance residuals plotted against each predictor. Continuous predictors have a constant line at the mean (indicating no trend in residuals) and a local regression smoothed line fit to utilization data only.

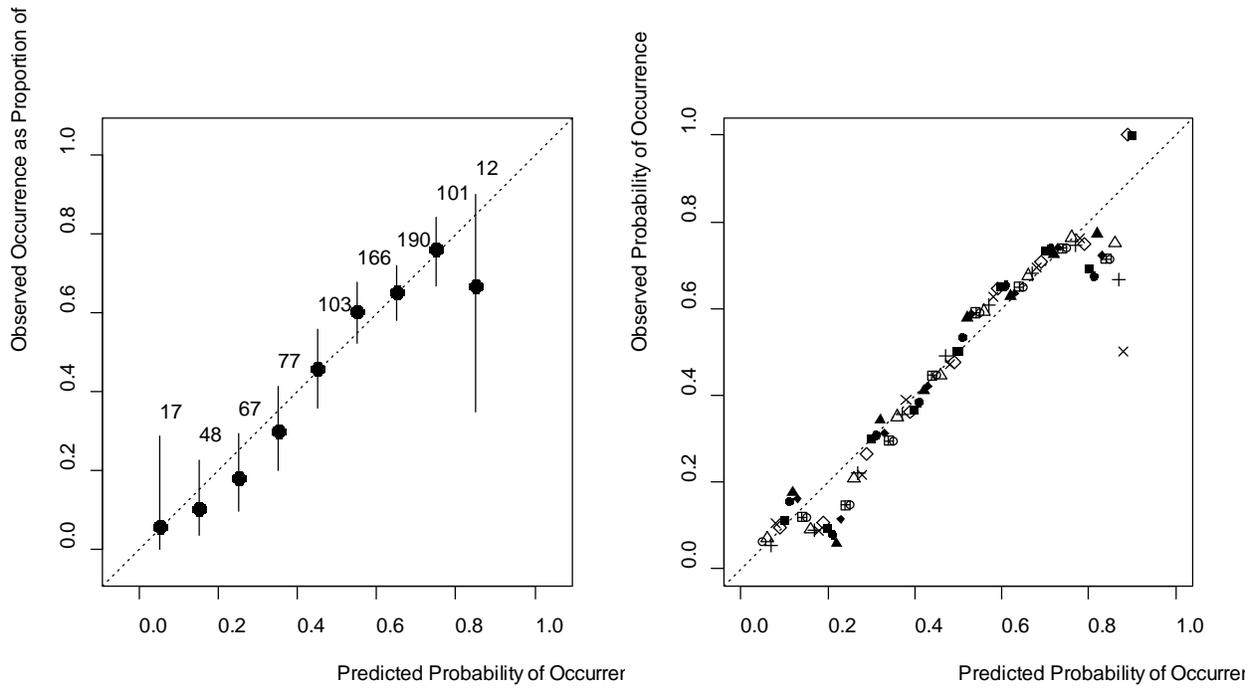


Figure 3.2-1. Plot of observed versus predicted probabilities of occurrence for 10 equal-sized bins on range of (0,1) (left), and on nine alternative bit cutoff series (right).