Measuring the Biotic Integrity of Stream Ecosystems with Restoration

Robin M. Henderson and James R. Pratt

Abstract.

Millions of dollars are spent annually to restore biodiversity and ecosystem services in streams and rivers degraded by land use change and other human activities. To determine if stream restoration has improved the biotic integrity of the Middle Fork John Day River (MFJDR) over the past 5 years, we tested alternative methods for assessing the status of stream ecosystems. Promising approaches include developing predictive models based on the estimation of observed to expected ratios (O/E) using numerical clustering and classification techniques. Macroinvertebrate assemblage data from the Oregon Department of Environmental Quality (ODEQ) was used to compare 105 reference sites and 442 test sites in three Oregon ecoregions. We used discriminant function analysis (DFA) and random forest (RF) models to predict the probability that a test site is a member of each reference site group by applying predictor variables developed for a variety of scales. The best performing model, based on model performance metrics, was chosen to determine if the 20 MFJDR sites have improved biotic integrity scores following restoration. Results indicate that there is not yet evidence of improvement in biotic integrity scores for restored MFJDR sites. Ongoing analysis is needed to characterize interannual variability and the sources of biological variability in the river ecosystem.
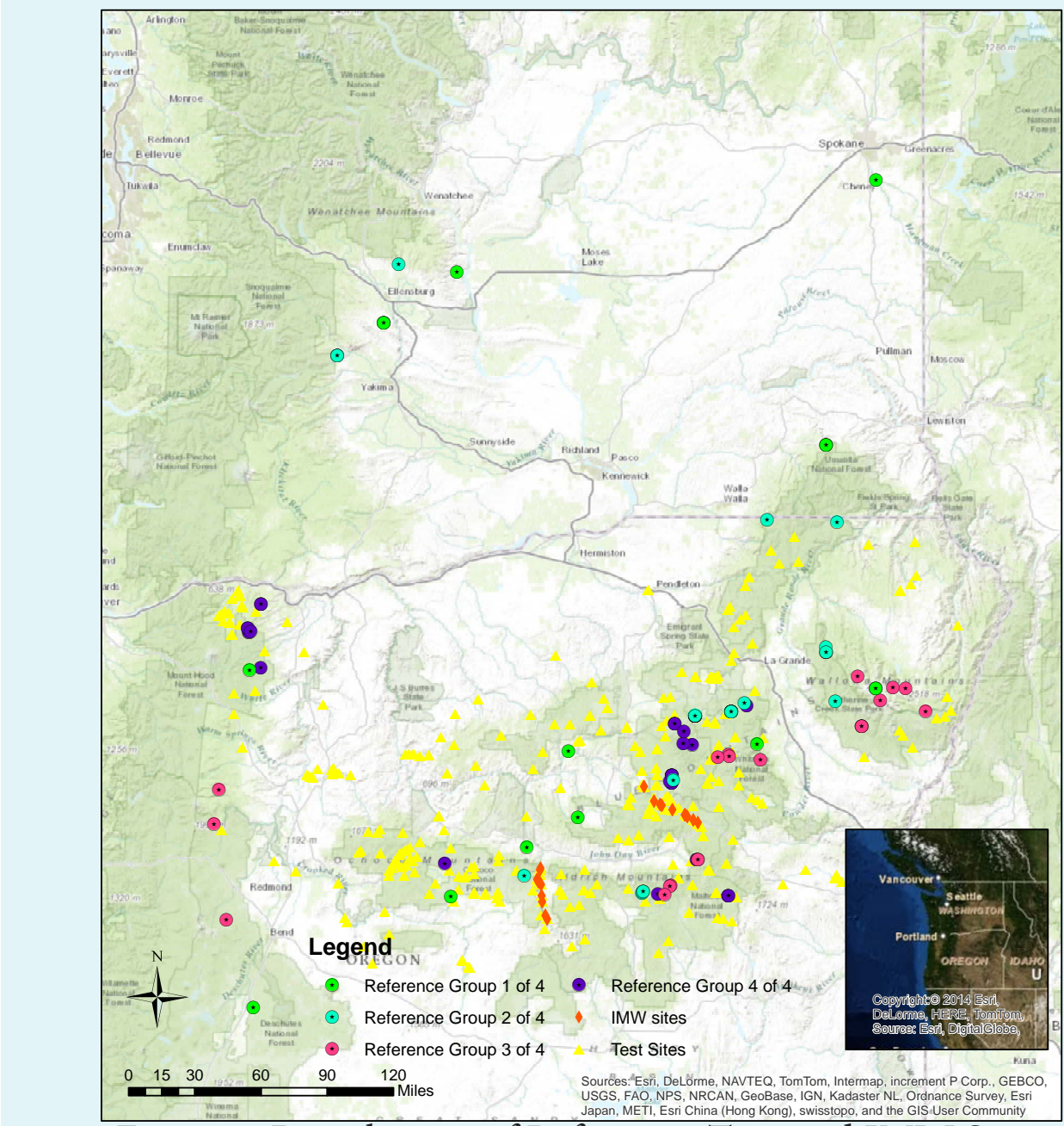
Figure 1. Distribution of Reference, Test, and IMW Sites.

Figure 2. Dendrogram of Reference Groups. This figure shows the 2 different reference groups classifications used for this research.

Figure 3. Monitoring and Restoration Activities in IMW. A smolt trap (left) and preparations for channel realignment (right).

Introduction.

Recently, stream restoration has become a primary conservation strategy, therefore, impacts from environmental stressors need to be detected, monitored and assessed. Biotic indices can be used to assess aquatic ecosystem conditions, to set protection and restoration goals, identify stresses to the stream, and evaluate the effectiveness of management actions. One index is the O/E ratio, generated as the output of a predictive model, which estimates the macroinvertebrate assemblage expected at a stream site if it were in a minimally disturbed reference condition. Traditional O/E indices, based on the River InVertebrate Prediction and Classification System (RIVPACS) framework, utilize DFA to predict the probability that a test site is a member of each reference site group, including the RIVPACS-type model developed by the ODEQ entitled the PREDictive Assessment Tool for Oregon (PREDATOR) model. Newer statistical methods such as RF may help practitioners develop predictive models that better support the needs of bioassessment programs. Thus, the goals of this research were to:

1. To evaluate how different classification methods based on the same set of candidate predictor variables affected the performance of O/E indices in detecting biological alteration associated with landscape and waterway modification.
2. To determine if stream restoration of the MFJDR has improved biotic integrity over the past 5 years.
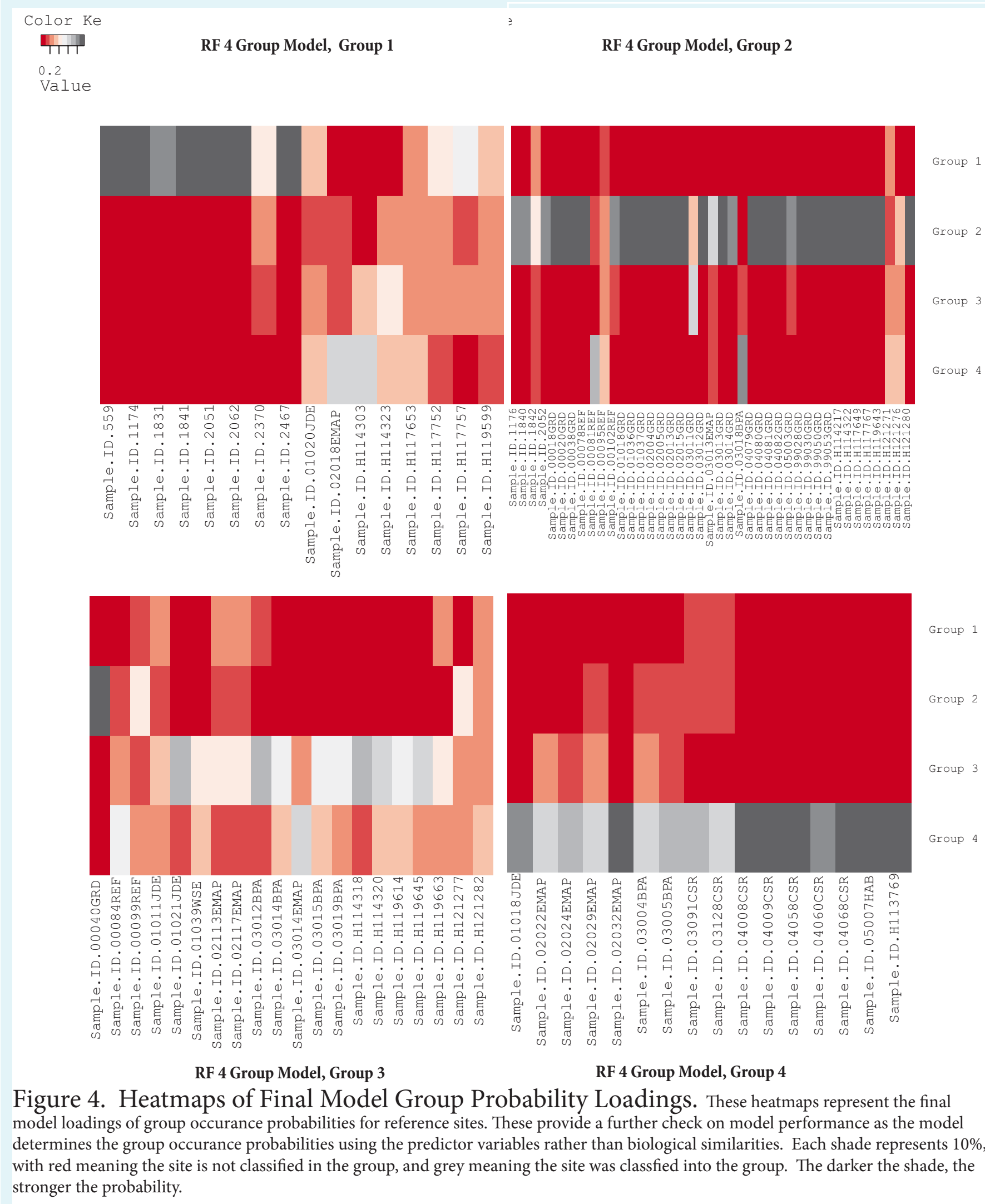

Figure 4. Heatmaps of Final Model Group Probability Loadings. These heatmaps represent the final model loadings of group occurance probabilities for reference sites. These provide a further check on model performance as the model determines the group occurance probabilities using the predictor variables rather than biological similarities. Each shade represents 10% with red meaning the site is not classified in the group, and green meaning the site was classified into the group. The darker the shade, the stronger the probability.
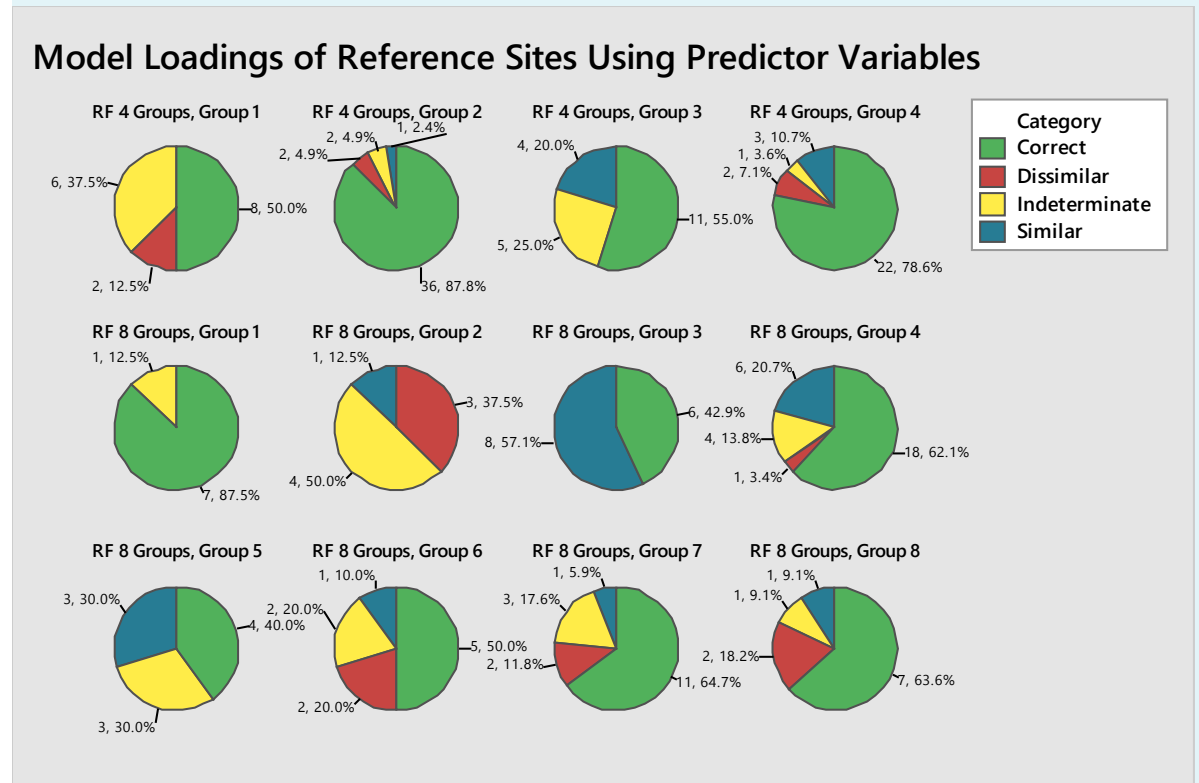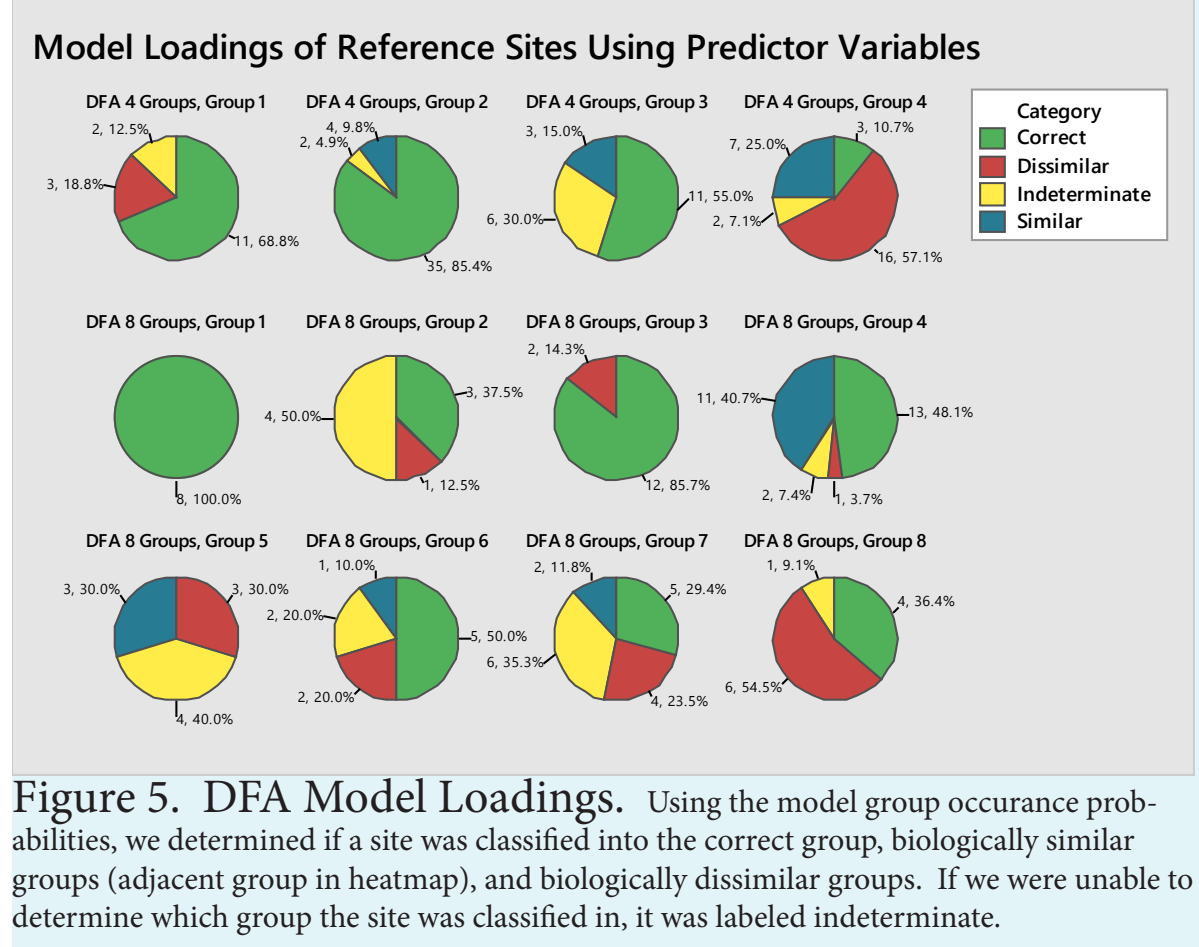

Figure 5. DFA Model Loadings. Using the model group occurance probabilities, we determined if a site was classified into the correct group, biologically similar groups (adjacent group in heatmap), and biologically dissimilar groups. If we were unable to determine which group the site was classified in, it was labeled indeterminate.

Figure 6. RF Model Loadings. Using the model group occurance probabilities, we determined if a site was classified into the correct group, biologically similar groups (adjacent group in heatmap), and biologically dissimilar groups. If we were unable to determine which group the site was classified in, it was labeled indeterminate.

Methods.

The methodology for deriving RIVPACS-type models was followed in this research. We used datasets previously aggregated by ODEQ as well as another dataset from the MFJDR Intensively Monitored Watershed (IMW) from Central and Eastern Oregon and Washington for model development (Figure 1). To evaluate model performance, we compared performance metrics (accuracy, precision, bias, sensitivity to stressors, and responsiveness) of the 2 classification techniques and the PREDATOR model. The final model was selected from the 2 classification techniques based upon these measures, which was then used to determine if the biotic integrity of the MFJDR has improved with restoration. Performance metrics were defined as follows.

1. Accuracy: Accuracy was examined with 10 fold cross-validation (c.v.) and a regression plot of O versus E for reference sites. Accurate models have a scatterplot that resembles a 1:1 line.
2. Precision: Precision was measured as the observed variability of O/E values among reference sites, represented by the standard deviation (SD) of O/E values and by comparing the amount of variation in O that is predicted by E, represented by the r2 value from regression of O to E for reference sites. Precise RIVPACS-type models produce SDs of approximately 0.15 and r2 values between 0.5-0.75.
3. Bias: The mean O/E score, SD, and distributions were compared for model calibration sites and validation sites using boxplots and the Student's t-test.
4. Sensitivity to stressors: We measured sensitivity as the percent of test sites that fell below the 10th percentile of reference site O/E scores. We used McNemar's test to determine if sensitivities varied significantly from one another since the percentage of sites declared as most disturbed by different models cannot be extrapolated to other models.
5. Responsiveness: We measured responsiveness as the Student's t value estimated from the comparison of reference and test site O/E scores.

Table 1. Predictor Variable Statistics.

This table shows the mean value of the top 10 predictor variables, as determined by the final model. 1. For categorical variables, the median is provided rather than the mean.

Table 2. Predictor Variable Importance.

Figure 7. Oxbow Conservation Area Project within the MFJDR IMW. The first photo (left) was taken before large woody debris placement in 2011. The second and third photos show the project in 2011 and 2013, after restoration.

Figure 8. A portion of the channel is blocked off during the course of a restoration project in the MFJDR IMW.

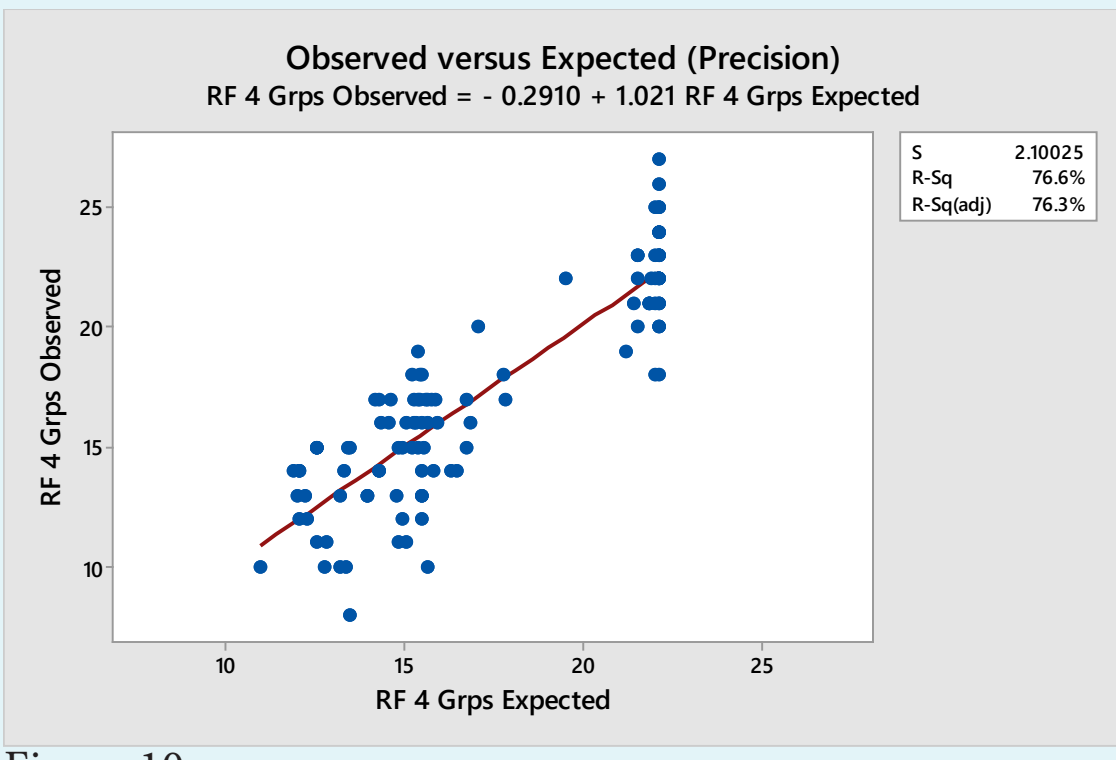Table 3. Benchmarks of Biological Condition.

Figure 9. Panoramic View of MFJDR.

Figure 10. RF 4 Groups Model Regression Plot. Observed was regressed on Expected to determine model precision.

Results.

1. Accuracy: Results produced from the 10-fold c.v. of the DFA and RF models show that the RF models have a greater accuracy rate compared to the DFA models and the PREDATOR model using both 2-sample t-tests. These results were significant at the 0.05 level, with p<0.01 and a sample size of 100. Comparisons across different group sizes of the 10-fold c.v. accuracy rates show that the models with the 4 group sizes classification have a statistically higher accuracy rate compared to the 8 group models, also significant at the 0.05 level, with p<0.01 and a sample size of 100. Furthermore, all models generally produced a 1:1 scatterplot; however, the RF models produced less scatter and intercepts closer to one (Figure 10).
2. Precision: Generally speaking, all SDs produced were very similar between the RF and DFA models. Both RF models produced comparable SDs to the PREDATOR model, but one DFA model exceeded the SD of the PREDATOR model. The r2 values of the RF models, which represents the percent of O explained by E, exceeded the values produced from the DFA models and the PREDATOR model (Table 4).
3. Bias: Mean O/E values for both RF and DFA models were very close to unity, which showed that both models produced unbiased estimates of biological condition.
4. Sensitivity to Stressors: Based on the results of the Mcnemar's test, the RF models were more sensitive to stressors than DFA models and the 8 group models were more sensitive to stressors than the 4 group models; however, both the 4 group and 8 group DFA and RF models were significantly more sensitive to stressor's compared to the PREDATOR model (Table 4, Figure 10).
5. Responsiveness: The 8 group models produced lower O/E scores compared to the 4 group models. All models were comparable in terms of the Student's t statistic to use the PREDATOR model (Table 4).

Ultimately, the RF 4 Group model was selected as the final model. This model was used to evaluate the biotic integrity of the MFJDR, where there have been recent major restoration events such as channel reconfiguration, buffer zone establishment, and large woody debris placement over the last 5 years. The t-test results indicate that there is not enough evidence (p<0.05) to show that there have been improvements in the biotic integrity scores from 2010 to 2011, similar to the PREDATOR model; however, the DFA and RF model boxplots indicate that the O/E scores from 2012 were lower than 2011 in both treatment and control sites for the IMW.

Discussion.

1. The RF classification technique produces a better performing model and biotic index compared to DFA models as well as the current PREDATOR model, as measured by the 5 performance metrics used in this research.
2. Currently, we do not possess enough macroinvertebrate data to determine how management actions within the IMW have affected the MFJDR with regards to the biotic integrity. General trends observed from the distributions and boxplots of the MFJDR IMW O/E scores indicate a decline in biological integrity has occurred from 2011 and 2012. Review of regional climate data indicate that a strong la nina event occurred in 2010 and a moderate el nino event in 2009, which may have impacted the biotic integrity. Furthermore, snowpack trends in the Columbia River basin from 2009-2012 varied from month to month and year to year, but these trends were not analyzed statistically. Ongoing analysis is needed to determine the regional climate drivers which may impact the biological integrity of regional streams. Additionally, further research will attempt to trace if restoration events have affected the biotic integrity of the MFJDR and to characterize interannual variability in the MFJDR.
3. Assessing candidate predictor variables helps identify the dominant natural factors that control macroinvertebrate assemblages at reference sites; thus, the results of this study may help environmental and land use managers understand the effects of human land use and make more effective land use decisions to address watershed impairment.

Table 4. Model Statistics.

Metric	DFA 4 Grps	RF 4 Grps	DFA 8 Grps	RF 8 Grps	PREDATOR
Calibration Sites					
10-fold Crossvalidation	0.75	0.82	0.53	0.62	0.59
Predictive Model Mean O/E	1.01	1	1	1.01	1.01
Predictive Model SD O/E	0.16	0.13	0.15	0.15	0.15
Null Model Mean O/E	1	1	1	1	1
Null Model SD O/E	0.18	0.18	0.18	0.18	0.18
Outliers	0	0	0	0	NA
R²	0.52	0.77	0.59	0.65	33.3
Validation Sites					
Predictive Model Mean O/E	1.01	1	1	1.01	NA
Predictive Model SD O/E	0.16	0.13	0.15	0.15	NA
Test Sites					
Predictive Model Mean O/E	0.88	0.9	0.81	0.86	0.93
Predictive Model SD O/E	0.18	0.18	0.2	0.19	0.11
Outliers	4	0	16	0	NA
Sensitivity	31.2	33.8	49.8	40.1	24.6
Student's t	7.45	6.98	10.86	8.6	6.8

This table provides the statistics of each model developed in this research.

Figure 11. Model Biological Classifications. Model sensitivity was calculated using the percent of sites classified as most disturbed, sites with O/E scores in the 10th percentile of reference site O/E scores.

Figure 12. IMW O/E Scores. This shows the distributions of the IMW O/E scores by year, separated by the treatment and control streams.

Figure 13. Oxbow Conservation Area. A view of the Oxbow Conservation Area, a management unit within the MFJDR IMW.

Figure 14. Variance of IMW O/E Scores. These graphs show how each O/E score for each site has varied from 2010-2012.

References.

1. Clarke, R., Wright, J. & Furse, M., 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. Ecological Modelling, Volume 160, p. 219–233.
2. Hawkins, C., Norris, R., Hogue, J. & Feminella, J., 2000. Development and Evaluation of Predictive Models for Measuring the Biological Integrity of Streams. Ecological Applications, Volume 10.
3. Hubler, S., 2008. PREDATOR: Development and use of RIVPACS-type macroinvertebrate to assess the biotic condition of wadeable Oregon streams, s.l.: Oregon Department of Environmental Quality; Laboratory Division; Watershed Assessment Section.
4. Van Sickle, J., Huff, D. D. & Hawkins, C. P., 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. Freshwater Biology, Volume 51, p. 359–372.
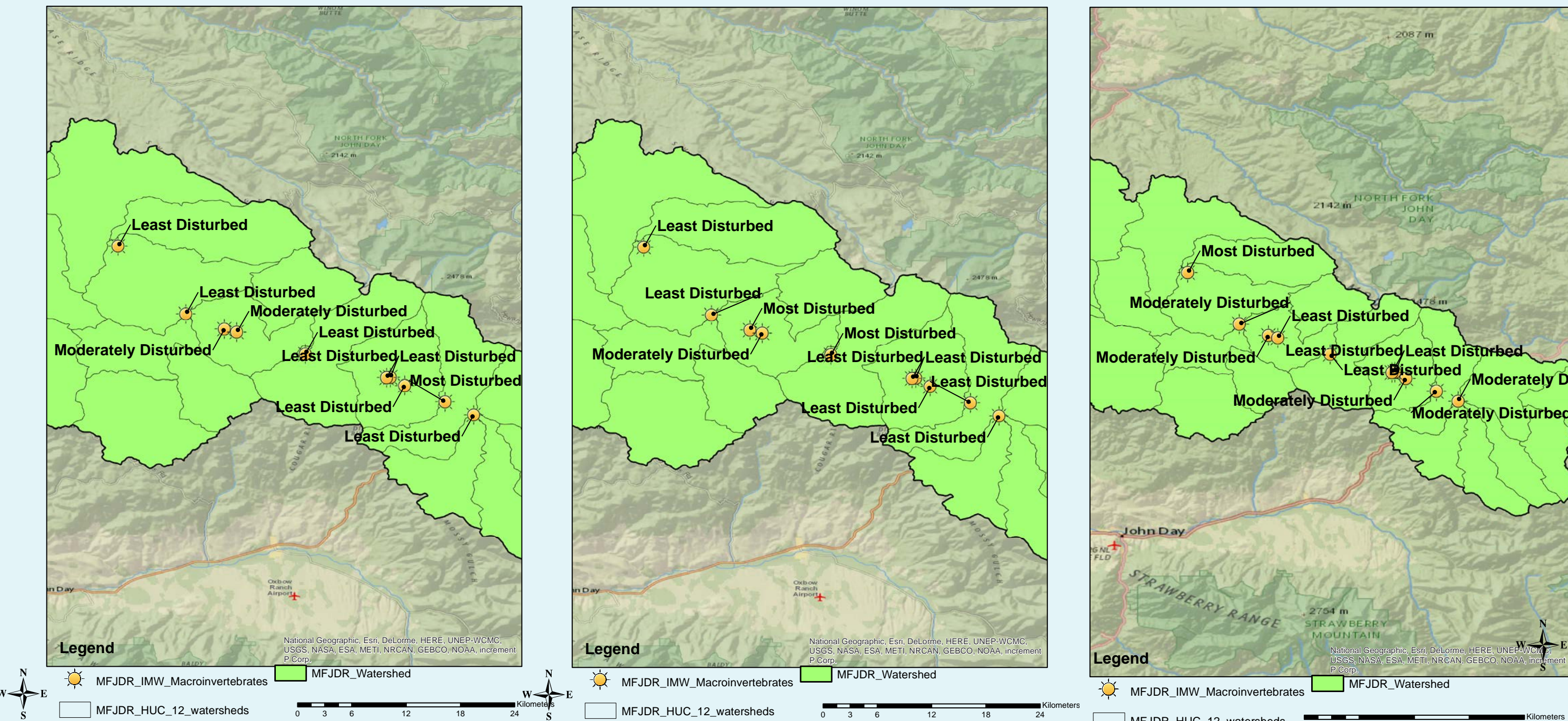
Figure X. Biotic Integrity of MFJDR, 2010-2012. These figures show a time series representing the biological class of the MFJDR IMW sites.

WASHINGTON STATE UNIVERSITY TRI-CITIES