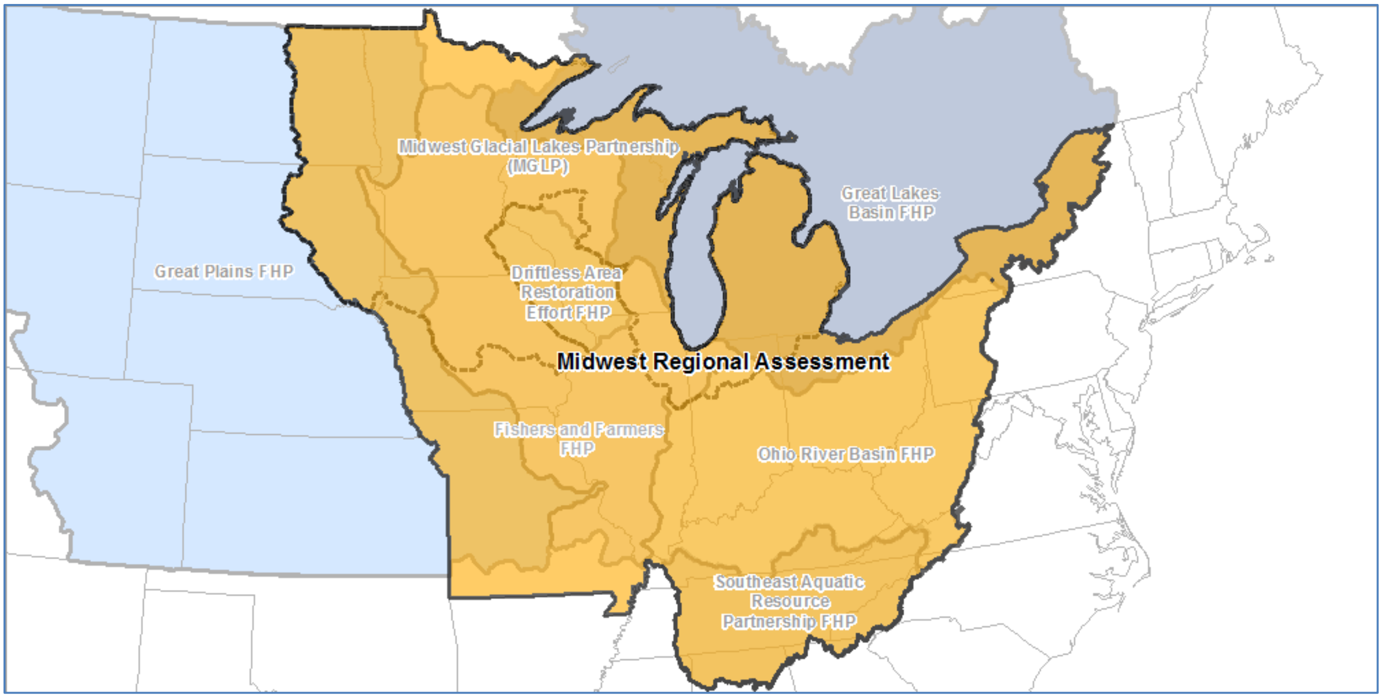
Midwest Fish Habitat Partnership Fish Habitat Modeling Results

Regional Assessment

Model Summaries
9/30/2013

Coldwater Guild: Probability of Presence
Coolwater Guild: Probability of Presence
Warmwater Guild: Probability of Presence



Authors
Jason Clingerman, Aquatic Ecologist/Modeler – Downstream Strategies
Todd Petty, Aquatic Ecologist – West Virginia University
Fritz Boettner, Project Manager/GIS – Downstream Strategies
Sally Letsinger, GIS – Geodata Basics
Jackie Strager, GIS – West Virginia University
Evan Hansen, Chief Scientist – Downstream Strategies

Downstream Strategies
295 High Street, Suite 3
Morgantown, WV 26505
www.downstreamstrategies.com

Report date:
September 3, 2013

ACKNOWLEDGEMENTS

Principal Investigator: Maureen Gallagher, USFWS

Midwest and Great Plains FHP Science Team:

Steve Krentz, USFWS; Mark Brouder USFWS, Jeff Thomas ORSANCO, Emily Grandstaff SARP, Louis Mauldin USFWS

Fish Data: States of Indiana, Illinois, Iowa, Kentucky, Maryland, Michigan, Minnesota, Missouri, New York, North Carolina, North Dakota, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, Wisconsin

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVIATIONS

BRT	boosted regression tree
ASI	anthropogenic stress index
HQI	natural habitat quality index
CV	cross-validation
DS	Downstream Strategies
FHP	Fish Habitat Partnership
GIS	geographic information systems
GLB	Great Lakes Basin
NHD	National Hydrography Dataset
NPDES	National Pollutant Discharge Elimination System
ROC	receiver operating characteristic
USFWS	United States Fish and Wildlife Service

1. INTRODUCTION

1.1 Document outline

This report provides a summary of the key outcomes resulting from models developed by DS for use in assessing aquatic habitats for the Midwest Fish Habitat Partnerships. The appendices provide additional maps, charts, and metadata useful for evaluating the results of the models.

This document is divided into six major sections. This section, Section 1, summarizes the project goals, structure, and methodology. Sections 2, 3, and 4 summarize the model input and results for each of the three response variables. Section 5 details the methods and results from the combined analysis where all three of the responses were considered simultaneously. Section 6 summarizes some of the limitations to this modeling effort, and outlines suggestions for future similar works.

The following are included for each model's results summary.

- Subsection one, *Modeling inputs*, discusses details of the predictor and response variables used in the analyses.
- Subsection two, *Modeling process*, covers the basic details and outcomes of the statistical modeling process using BRTs, including information on model certainty. Variable influence and functional relationships between predictor and response variables are included under corresponding headings as well.
- Subsection three, *Post-modeling*, contains information resulting from the post-modeling process, including information on the top stressors and natural habitat variables and their role in the calculation of the final indices.
- Subsection four, *Mapped results*, contains maps for visualizing conditions at the 1:100k catchment scale and includes maps of expected current probability of presence, stress, and natural quality; it also provides examples of how the two post-modeling indices (i.e., HQI and ASI) can be combined to inform restoration priorities and how those priorities can be visualized in a spatially explicit manner.

1.2 Project background

Fishery and aquatic scientists often assess habitats to understand the distribution, status, stressors, and relative abundance of aquatic resources. Due to the spatial nature of aquatic habitats and the increasing scope of management needs, traditional analytical assessment methods are often limited in their ability to address complex and dynamic aquatic systems. Advancements in the geographic information systems (GIS) field and related technologies have enabled scientists and managers to more effectively collate, archive, display, analyze, and model spatial and temporal data. For example, spatially explicit habitat assessment models allow for a more robust interpretation of many terrestrial and aquatic datasets, including physical and biological monitoring data, habitat diversity, watershed characteristics, and socioeconomic parameters.

Downstream Strategies (DS) was contracted by the United States Fish and Wildlife Service (USFWS) to create a spatially explicit data analysis and modeling system for assessing fish habitat condition for several individual Fish Habitat Partnerships (FHP) across the Midwest and Great Plains based on a range of metrics. These analyses provided data and tools for specific aquatic species for each FHP, and were constructed at the scale of the individual FHP. These results were useful, but lacked a region-wide assessment of overall habitat quality and aquatic stressors. This project built upon the knowledge gained and framework designed during the individual FHP-scale modeling efforts and provided the consistent region-wide aquatic endpoints. Additionally, for this analysis, a new methodology was developed for assigning stress and determining natural quality of aquatic habitats, and this analysis also included an assessment of aquatic habitat vulnerability to climate change.

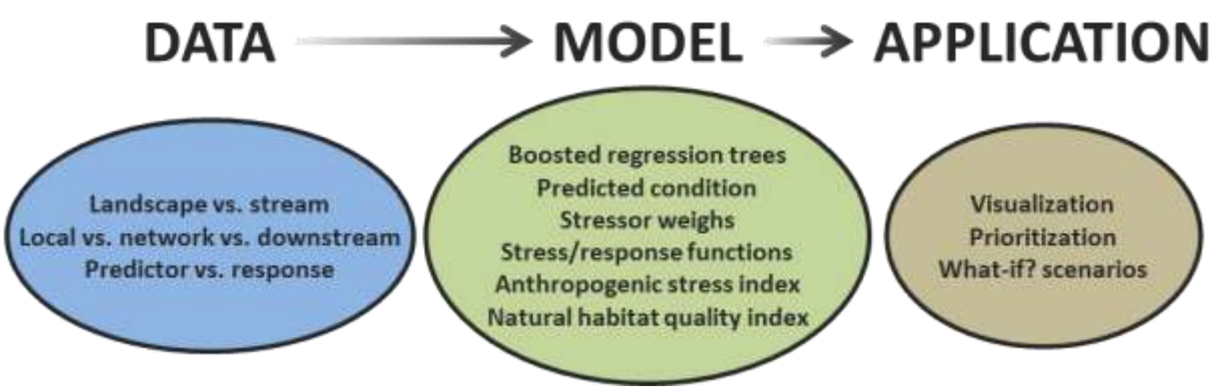
Generally, the models, analyses, and data produced as a result of this project are intended to enable a unique, broad, and spatially explicit understanding of the links between natural habitat conditions, human influences on aquatic habitats, and aquatic health. Specifically, the outcomes can be utilized to conduct fish habitat condition assessments based on a range of stakeholder-specified metrics and modeling endpoints to help determine natural drivers of aquatic conditions, as well as major stressors within the Midwest and Great Plains regions. The ultimate goal is to improve understanding of how local and regional processes influence stream conditions in the region and to provide additional knowledge, data, and tools to help prioritize and drive conservation action throughout the Midwest and Great Plains.

1.3 Overview of the assessment process

1.3.1 *Modeling*

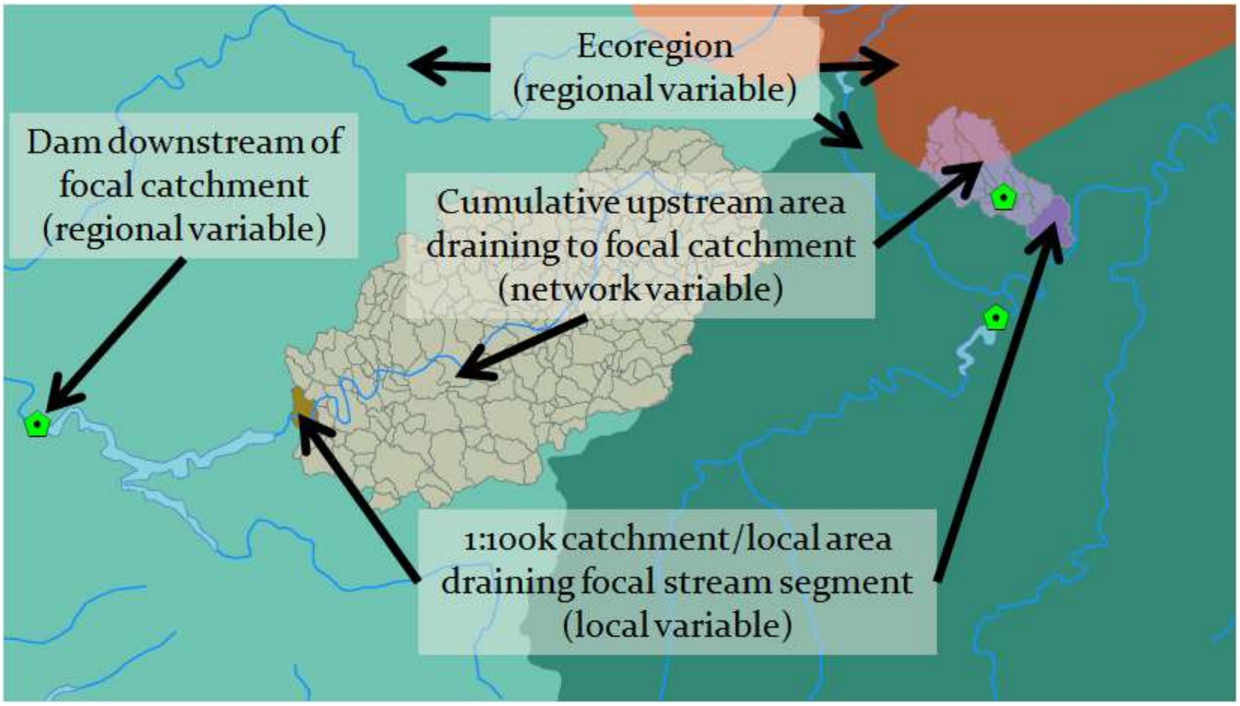
A diagram of the general assessment process is outlined in Figure 1. DS received landscape and aquatic data specified and provided by the individual FHPs to develop models and tools for visualizing expected current and potential future conditions and prioritizing management actions.

Figure 1: Diagram of the habitat assessment process



The data provided by FHPs for use in the modeling process can be broken down into two categories: response variables and predictor variables. The response variables for this project are presence-absence datasets of freshwater stream fish guilds. For this assessment, a fish guild is defined as a group of fish that have similar habitat requirements and are relatively intolerant to habitat degradation. There were three response variables used in this assessment: coldwater, coolwater, and warmwater guilds. Each response variable represented a separate model. The predictor variables are typically measures of land use or land cover derived from GIS, such as percent impervious surface area or road crossing density. Although the response variable is always measured at the same local scale (e.g., individual sample site on a stream), the predictor variables are compiled at multiple scales (Figure 2), including the local scale (e.g., single 1:100k National Hydrography Dataset (NHD) stream catchment), the network scale (e.g., all upstream catchments and the local catchment), or the regional scale (e.g., ecoregion).

Figure 2: Diagram and examples of different scales of data used for predictor variables



For this assessment nearly all of the predictor and response data necessary was already held by DS from prior individual FHP assessments across the Midwest. This data was simply compiled into one large, region-wide dataset. DS, along with the Midwest and Great Plains FHP science team, agreed to limit the predictor variables to those that were strong predictors in other modeling efforts. The final list of potential predictor variables is shown below in Table 1.

Table 1: Predictor variables

Variable description	Source	Variable Type
Network drainage area	NHD+	Natural
Minimum catchment elevation	NHD+	Natural
Slope of catchment flowline	NHD+	Natural
Mean annual precipitation	NHD+	Natural
Mean annual air temperature	NHD+	Natural
Mean annual baseflow index	USGS	Natural
Mean annual recharge rate	USGS	Natural
Bedrock geology	USGS	Natural
Landcover classification	NLCD 2006	Varies
Impervious surface data	NLCD 2006	Anthropogenic
Groundwater use rate	USGS	Anthropogenic
Surface water use rate	USGS	Anthropogenic
Population density	NOAA	Anthropogenic
Road/stream crossing density	TIGER	Anthropogenic
Road density	TIGER	Anthropogenic
Dam density	National Inventory of Dams	Anthropogenic
Active mine density	USGS	Anthropogenic
Toxic release inventory site density	USEPA	Anthropogenic

NPDES permit density	USEPA	Anthropogenic
Superfund site density	USEPA	Anthropogenic

The process then employs a statistical modeling approach, called boosted regression trees (BRT), to relate the instream response variable to the landscape-based predictor variables. BRT models combine decision trees (i.e. classification and regression trees [CART]) and boosting methodologies, which result in better cross-validated models than other methods (Elith et al., 2006[i]), including CART. Decision trees are advantageous because (1) they can incorporate any type of predictor data (binary, numeric, categorical); (2) model outcomes are unaffected by differing scales of predictors; (3) irrelevant predictors are rarely selected; (4) they are insensitive to outliers and non-normalized data; (5) they can accommodate missing predictor data; and (6) they can automatically handle interactions between predictors (Elith et al., 2008). The boosting algorithm used by BRT improves upon the accuracy of a basic CART approach by following the idea that averaging many rough models offers efficiency over finding a single prediction rule that is highly accurate (Elith et al., 2008[ii]).

This process results in a series of quantitative outcomes, including predictions of expected current conditions to all catchments in the FHP (on the scale of the response), measures of the accuracy of those predictions, a quantification of each predictor variable's relative influence on the predictions (i.e., variable importance), and a series of plots illustrating the modeled functional relationship between each predictor and the response (e.g., plot of impervious area vs. presence-absence). The predictions of current conditions are created by extrapolating the BRT model to each catchment within the modeling area. The units of the predicted current condition for this assessment are probability of presence for the fish guild. These current conditions are useful for assessments of suitable habitats and mapping the expected range of species.

Predictive accuracy is quantified using an internal cross-validation (CV) method (Elith et al., 2008). The method consists of randomly splitting the input dataset into ten equally-sized subsets, developing a BRT model on a single subset and testing its performance on the remaining nine, and then repeating that process for the remaining nine subsets. Thus, the accuracy measures, such as the CV receiver operating characteristic (ROC) score (for presence-absence responses) or the CV correlation coefficient, are actually averages of ten separate ROC or correlation measurements. A standard error for the ten estimates is also given. CV measures are designed to estimate how well the model will perform using independent data.

1.3.2 *Post modeling*

Characterizing anthropogenic stress and natural habitat quality of aquatic habitats is a useful and necessary process for helping land and fisheries managers identify place-based conservation and restoration strategies. For each of the three models, a post-modeling process was used to characterize anthropogenic stress and natural habitat quality for all catchments within the study area. Stress and natural habitat quality indices and metrics were identified and calculated based on BRT model outputs, and details of those calculations are below.

Once developed, these indices of stress and habitat quality can be used to generate and visualize restoration and protection priorities by analyzing how stress reduction or habitat improvement can increase the probability of presence. For example, areas of high natural quality and low stress could represent protection priorities, whereas areas of high natural quality and high stress may represent restoration priorities.

Anthropogenic stress

Stress indices are critical for evaluating anthropogenic landscape drivers that structure aquatic responses. Managers can use stress indices and metrics to assess how anthropogenic processes are impacting aquatic responses and can utilize this information to cite restoration projects in order to maximize efficiency. Individual stressors were identified by examining BRT model outputs, both the variable influence table and the functional relationship between predictor variables and response variables. Any predictor variable significantly affected by anthropogenic disturbance was included as a potential stressor. Stressors were not utilized for calculation of stress in the model when the functional relationship between a potential stressor and the response variable was not indicative of a mechanistic relationship (e.g. regional trends were overwhelming mechanistic relationships).

Individual stress metrics were calculated by determining the increase in probability of presence for each catchment when stress for that predictor variable was hypothetically removed. A new predictor variable dataset was produced to calculate each individual stressor metric. The new predictor dataset contained the same values as the original predictor dataset except for a single anthropogenic variable for which a stress metric was calculated. For this variable, the values were all hypothetically set to reflect "no stress." This provided a hypothetical baseline that represented the removal of all stress from that predictor variable. The existing BRT model was then applied to the new hypothetical landscape data to provide an extrapolation of the current model assuming zero stress for that stressor. The difference between the current predicted probability of presence and the probability of presence under this "no stress" situation indicated the change that could be attributable to stress. This process was repeated for each stressor to generate individual metrics of stress on a potential scale of 0-1. Higher stress values indicated a larger change in predicted probability of presence after removing stress, and lower stress values indicated that the catchment was relatively unaffected by removing stress (Table 2).

For each catchment, the individual stress metrics (e.g. agriculture stress, impervious surface stress, etc) were then summed to produce an overall stress metric, the anthropogenic stress index (ASI). The generalized formula for calculating individual stress metrics and ASI is as follows:

individual stress metric = probability of presence$_{no\ stress}$ – probability of presence$_{current}$

anthropogenic stress index (ASI) = individual stress metric 1 + individual stress metric 2 + ….

Table 2: Example of stress calculations

Comid	Current Condition Predictions	Stressor 1 Predictions	Stressor 1 Metric	Stressor 2 Predictions	Stressor 2 Metric	Anthro. Stress Index (ASI)
Catchment ID	**Predictions using current landscape data**	**Predictions when stressor 1 removed**	**(Stressor 1 pred – Current Pred)**	**Predictions when stressor 2 removed**	**(Stressor 2 pred – Current Pred)**	**Stressor 1 Metric + Stressor 2 Metric**
1234567	0.80	0.90	0.10	0.80	0	0.10
1234568	0.25	0.50	0.25	0.35	.10	0.35
1234569	0.5	0.7	0.2	0.55	.05	0.25

Natural habitat quality

Natural habitat quality metrics provide critical baseline information on the optimal potential condition of a catchment. We defined natural quality as the maximum probability of presence under a zero-stress situation; essentially, the highest attainable condition in the catchment. These metrics allow managers to further classify each catchment and target specific land-based conservation or restoration actions.

The natural habitat quality index (HQI) was calculated directly from the BRT output. Metrics for 'natural' predictor variables were calculated using a different approach than the stressor calculations detailed above. Individual natural quality metrics were not seen as useful by the Midwest and Great Plains FHP Science Team since individual habitat variables were not considered practical management targets (e.g., elevation is a relatively fixed value) and therefore were not used in the calculation of HQI. A single hypothetical 'no stress' dataset was created where all stressors were removed. The existing BRT model was then applied to this hypothetical predictor dataset, and the resulting probability of presence indicated the maximum condition attainable by removing all stress. The probability of presence calculated by the BRT model for this hypothetical 'no stress' dataset is the HQI and this value indicates the maximum condition expected in each catchment.

$$\textit{natural habitat quality index (HQI)} = \textit{probability of presence}_{\textit{all stressors removed}}$$

1.3.3 *Future Climate Scenarios*

In addition to the analyses described above, potential future climate scenarios were also analyzed for each model. While current climactic data were used as predictor variables in the individual FHP's models described in the project background, the impact of future climate scenarios was not assessed until this project. The methodology used was similar to the methods for calculating stress and natural quality. A new predictor dataset was created where the values for temperature and precipitation variables were replaced with predicted temperature and precipitation values for the year 2050 under the Intergovernmental Panel on Climate Change's (IPCC) fourth assessment, global climate model (GCM) ensemble, high A2 emission scenario (IPCC 2007). This scenario showed a region-wide mean annual increase in both temperature (Figure 3) and precipitation (Figure 4). Mean annual temperature increased from 9.86°C to 12.67°C while mean annual precipitation average increased from 983.3mm to 1037mm. The maps illustrate that temperature changes are predicted to be most dramatic in the north, while precipitation increases occur more throughout the Mid-Atlantic states, and in the extreme northwest portion of the study area. These analyses assumed that all other landscape features were held at present levels, with only precipitation and temperature changing to assess changes expected from future climate scenarios.

This process was completed for each of the three responses, but was only applied during the combined analysis. This avoided indicating climate-induced changes on catchments that did not contain the fish assemblage in question. For example it would be inappropriate to assess the effects of climate change on the coldwater response in catchments that would likely contain a warmwater assemblage. Instead, we only applied climate vulnerability to catchments where the condition in the catchment matched the modeled response.

Figure 3: Predicted 2050 percent change in mean annual temperature

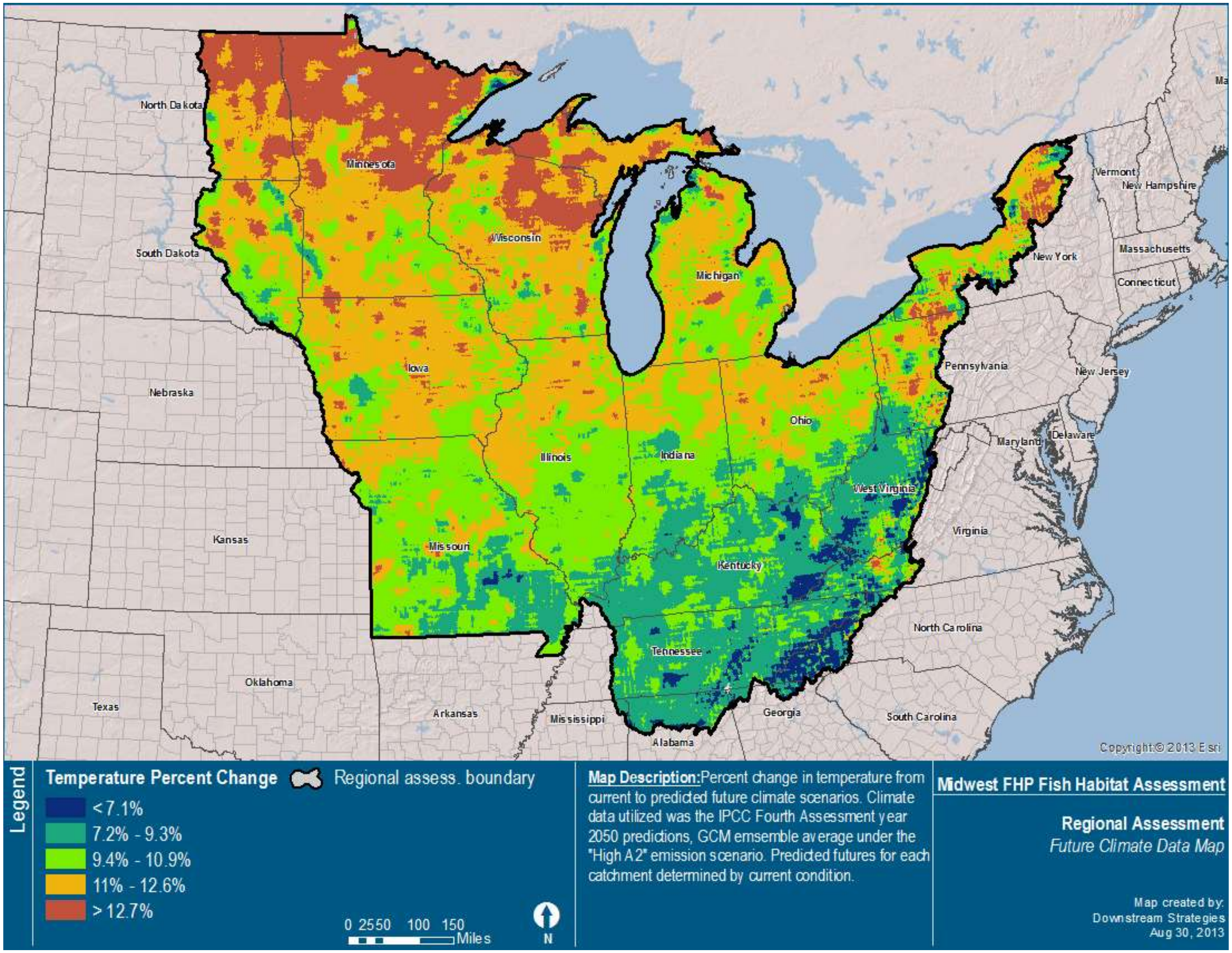
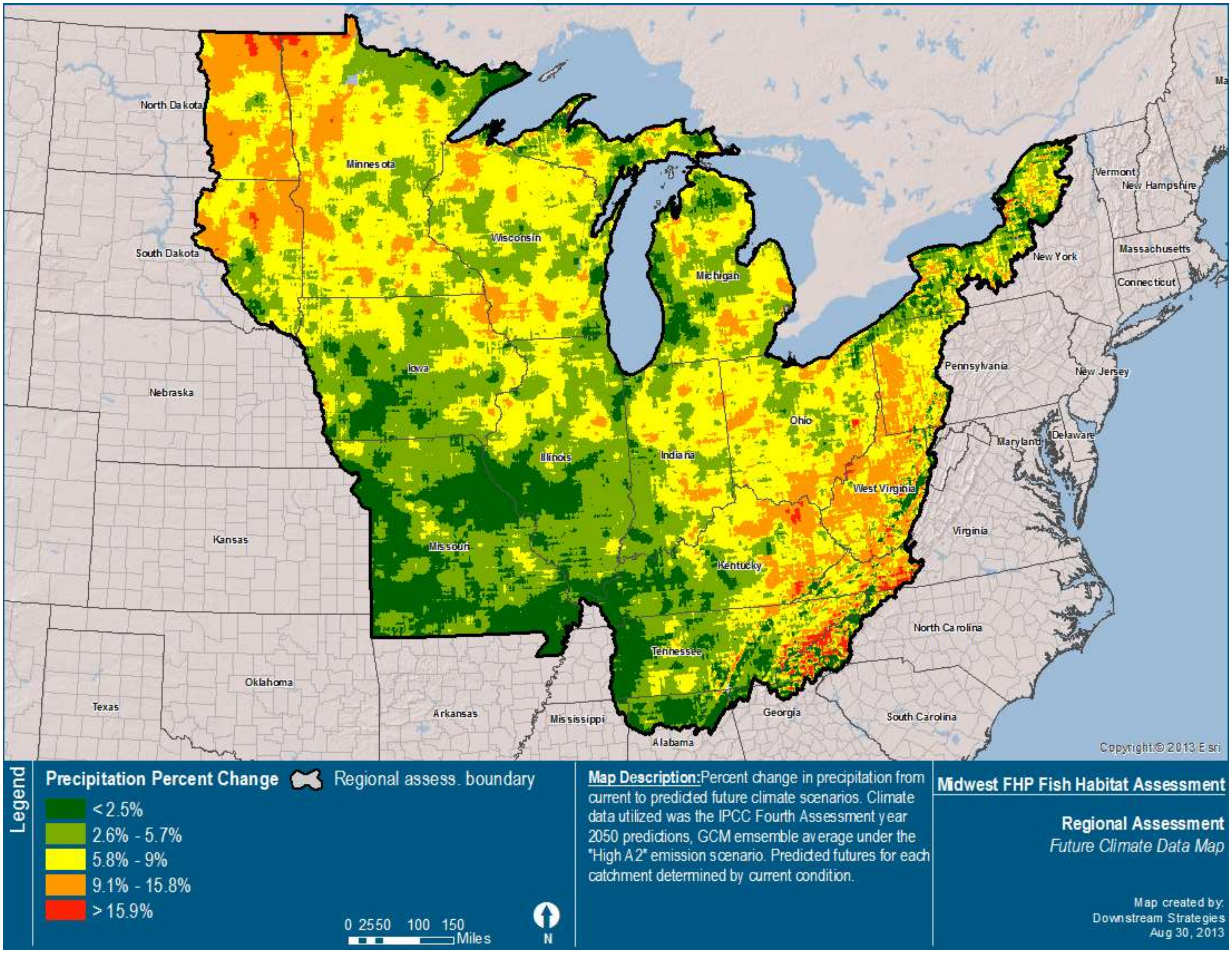


Figure 4: Predicted 2050 percent change in mean annual precipitation.



1.3.4 *Assessment Summary*

These methods provide current predictions of probability of presence, ASI scores, HQI scores, and potential future probability of presence for each of the three models. Metrics and indices were generated at the 1:100k NHD catchment scale and then mapped in GIS.

2. COLDWATER GUILD

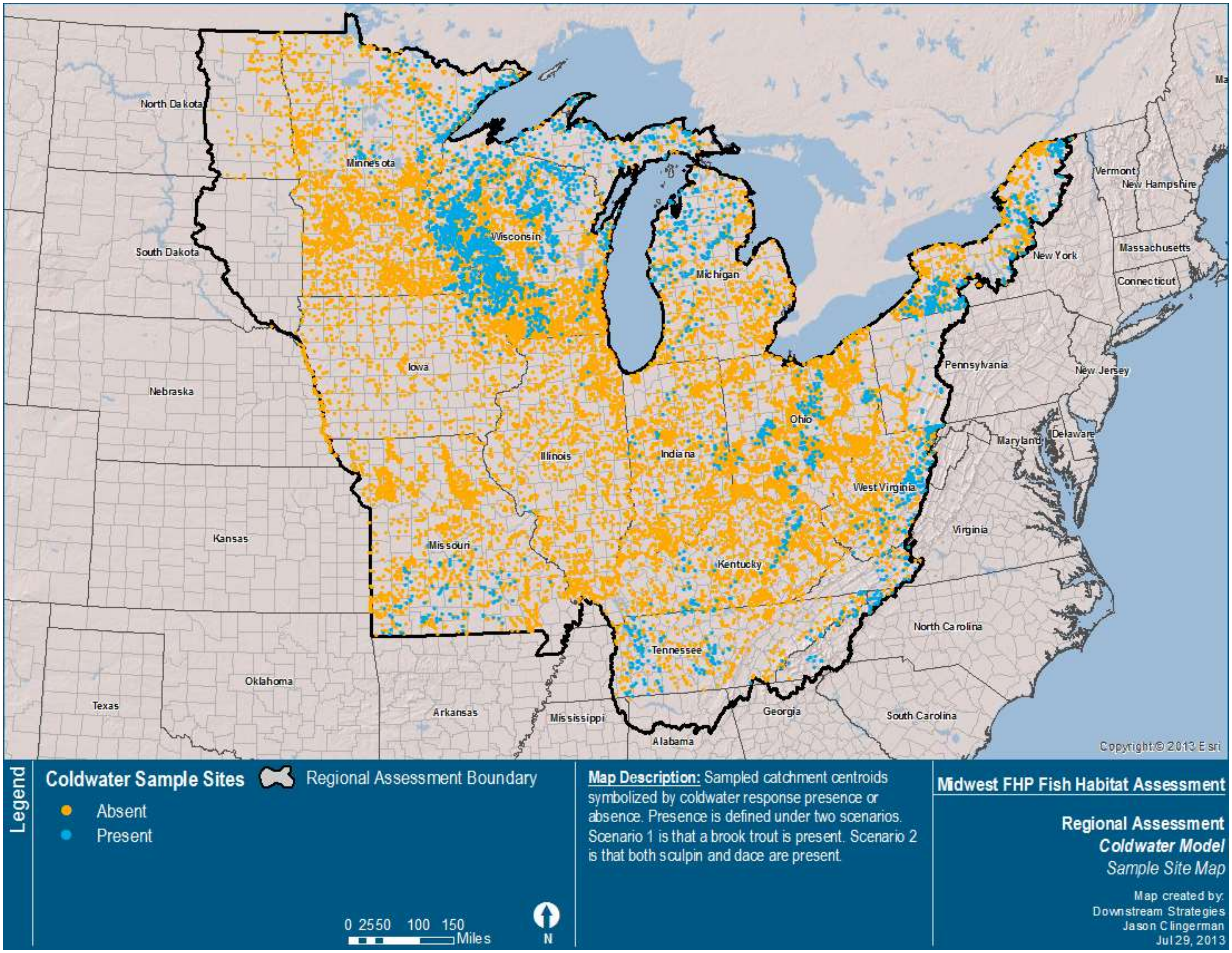
2.1 Modeling inputs

DS coordinated with a team of Fish Habitat Partnership scientists from the Midwest and Great Plains to construct a coldwater guild response variable and model the predicted probability of presence across the Midwest region. For this assessment, a fish guild is defined as a group of fish that have similar habitat requirements and are relatively intolerant to habitat degredation. This response was a presence-absence response, with presences being indicated when A) brook trout (*Salvelinus fontinalis*) were present or B) both a sculpin (*Cottus* sp.) and a coldwater dace (*Phoxinus* sp., *Clinostomus* sp., or *Rhinicthys cataractae*) were present. Absences were assumed where these species/scenarios were not found in community sample data.

Individual Fish Habitat Partnerships (FHPs) provided DS with fish data collected in streams over a time frame spanning from 1995 to 2011. Data collected by the FHPs generally came from state wildlife and fisheries agencies or from other reliable sources such as universities. DS then processed that data to create a presence-absence dataset for this coldwater fish guild which was comprised of 18,908 observations. Figure 5 maps all of the sampling sites that were used to construct the model and outlines the regional assessment boundary. Model outputs were applied to all 1:100k catchments within the regional assessment boundary.

DS cooperated with Midwest and Great Plains FHP Science Team to arrive at a list of landscape-based habitat variables used to predict coldwater guild habitat throughout the region. These variables were also used to characterize habitat quality and anthropogenic stress. Building on the science team's input, DS compiled a list of 67 predictors for evaluation. Preliminary exploratory models were then run to identify variable predictive performance and statistical redundancy. From that list, 57 variables were removed due to statistical redundancy ($r > 0.6$), logical redundancy, or poor predictive performance (relative influence < 1.0 in preliminary model run). This resulted in a final list of 10 predictor variables for the BRT model and assessment. See Appendix A for a full data dictionary.

Figure 5: Coldwater guild modeling area and sampling sites



2.2 Modeling process

2.2.1 *Predictive performance*

The final selected model was comprised of 6,200 trees. The model had a CV correlation statistic of 0.541±0.004 and a CV ROC score of 0.868±0.002.

2.2.2 *Variable influence*

The BRT output includes a list of the predictor variables used in the model ordered and scored by their relative importance. The relative importance values are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman and Meulman, 2003). The relative influence score is scaled so that the sum of the scores for all variables is 100, where higher numbers indicate greater influence. The relative influence table for the coldwater guild model is shown below in Table 3.

Table 3: Relative influence of all variables in the final coldwater guild model

Variable code	Variable description	Relative influence
BFI_MEANC	Network mean baseflow index	23.49
TEMP	Mean annual air temperature	22.15
SLOPE	Slope of catchment flowline	17.02
MINELEVRAW	Minimum catchment elevation	9.99
PRECIP	Mean annual precipitation	9.00
IMPSURF_MC	Network impervious surface cover	6.82
AREASQKMC	Network drainage area	4.51
AG_PC	Network agriculture land cover	4.24
WATER_GWC	Network groundwater use	1.47
WATER_SWC	Network surface water use	1.32

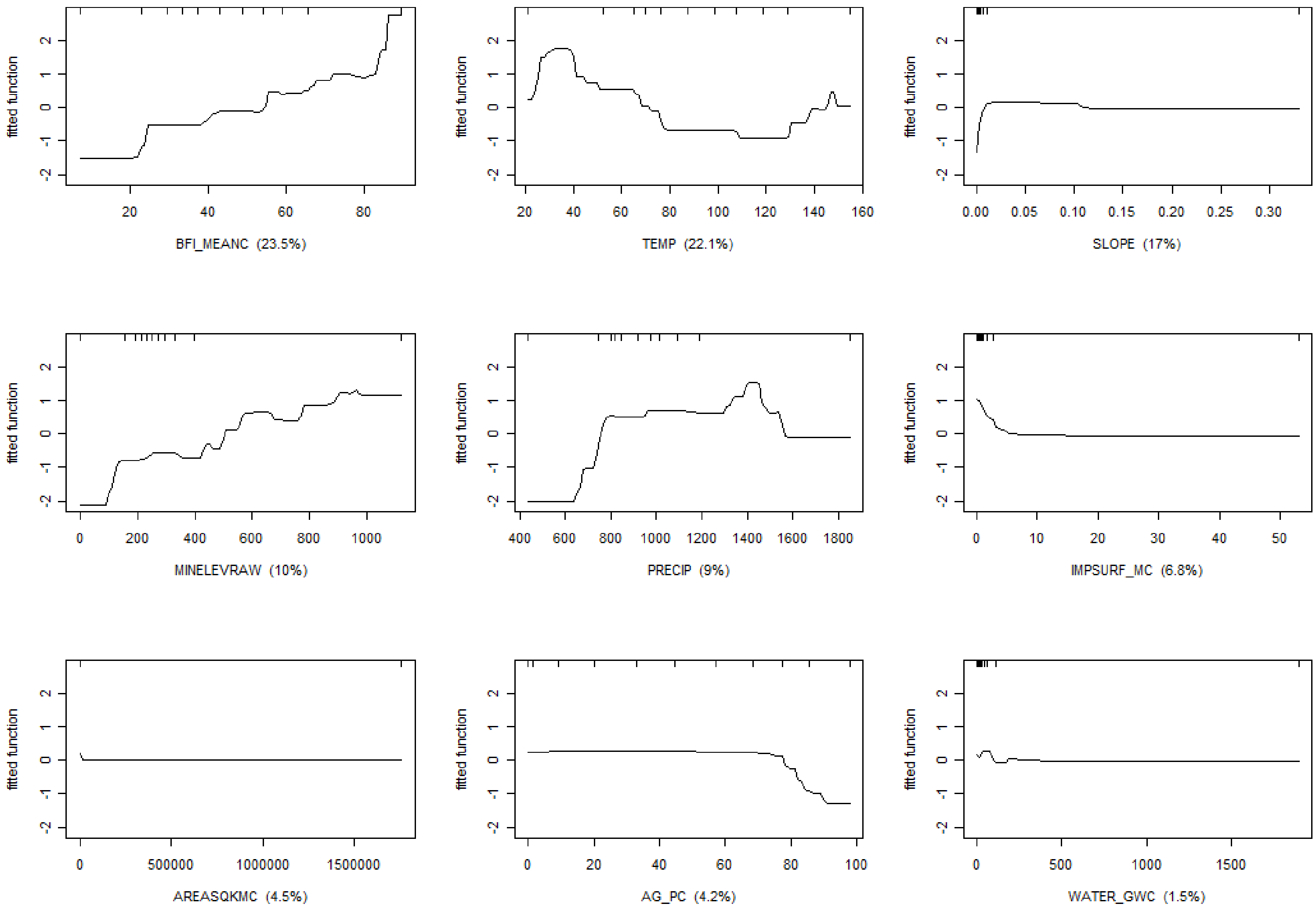
Note: Individual variables are highlighted according to whether they were determined to be anthropogenic (grey shading) or natural (no shading).

2.2.3 *Variable functions*

The BRT output also contains quantitative information on partial dependence functions that can be plotted to visualize the effect of each individual predictor variable on the response after accounting for all other variables in the model. Similar to the interpretation of traditional regression coefficients, the function plots are not always a perfect representation of the relationship for each variable, particularly if interactions are strong or predictors are strongly correlated. However, they do provide a useful and objective basis for interpretation (Friedman, 2001; Friedman and Meulman, 2003).

These plots show the trend of the response variable (y-axis) as the predictor variable (x-axis) changes. The response variable is transformed (usually to the logit scale) so that the magnitude of trends for each predictor variable's function plot can be accurately compared. The dash marks at the top of each function represent the deciles of the data used to build the model. The function plots for the nine most influential variables in the coldwater guild model (Table 3) are illustrated in Figure 6. The plots for all variables are shown in Appendix B.

Figure 6: Functional responses of the dependent variable to individual predictors of coldwater guild



Note: Only the top nine predictors, based on relative influence (shown in parentheses; see Appendix A for descriptions of variable codes), are shown here. See Appendix B for plots of remaining predictor variables.

2.3 Post-modeling

2.3.1 *Stress and natural quality*

The variable importance table and partial dependence functions of the final BRT model were used to assess the potential stressors for the coldwater guild model. Within the model, there were four variables considered anthropogenic in nature (Table 3). After reviewing the functional relationships of these four potential stressors, two of the four stressors were removed from ASI calculations. These variables ('WATER_SWC' and 'WATER_GWC') had function plots that were unintuitive: their relationships to the response likely captured some sort of regional variation in the model rather than a mechanistic relationship with the response. The two remaining stressors, network agriculture land cover (AG_PC) and network impervious surface cover (IMPSURF_MC), were used to calculate ASI for the coldwater guild model. Section 1.3.2 details how ASI and HQI were calculated for each model.

2.3.2 *Potential future climate scenario*

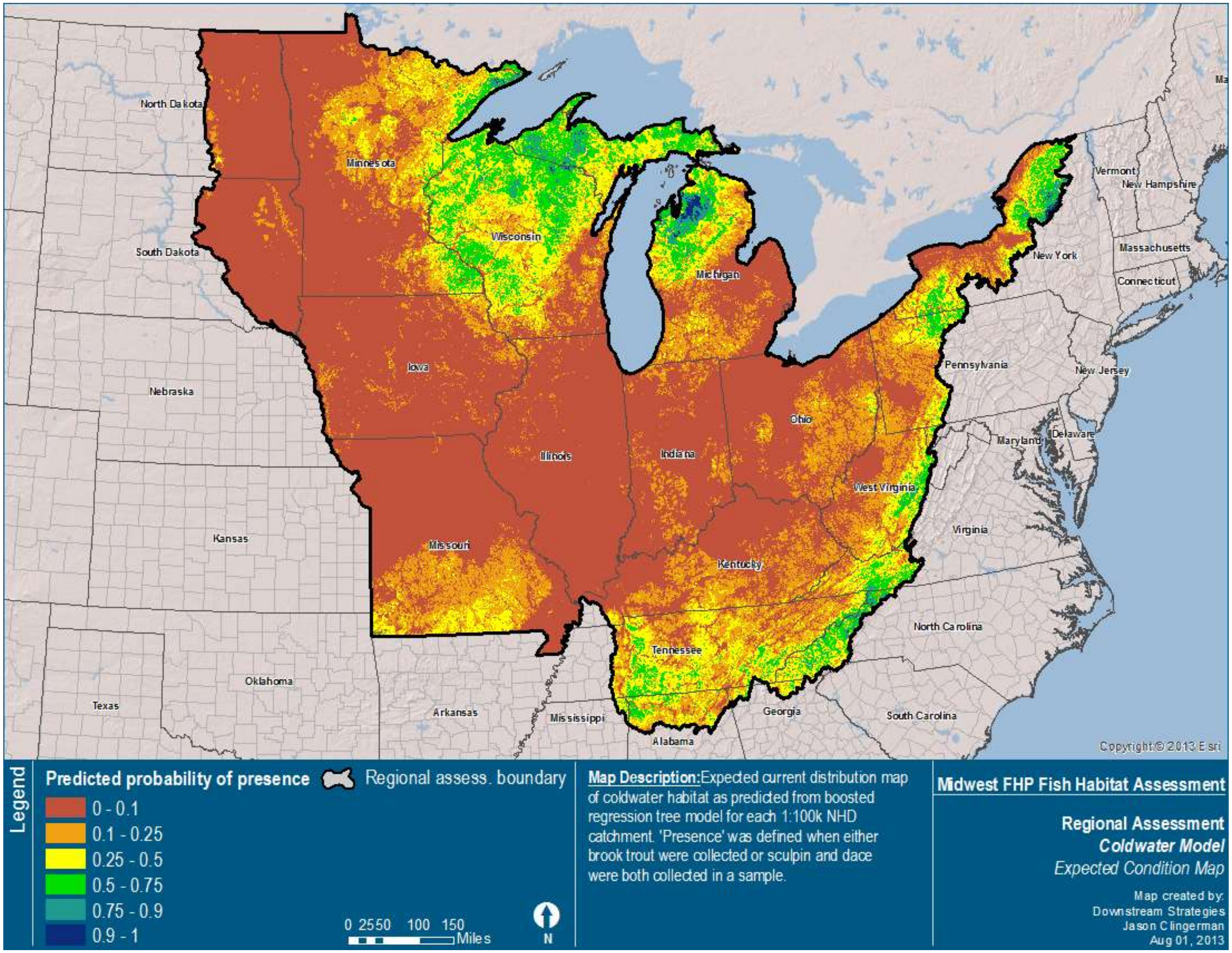
The coldwater BRT model was extrapolated onto a dataset that contained future climate data as described in Section 1.3.3. The potential future predictions were then compared to the current predictions. Percent change in probability was calculated for each catchment to assess climate change vulnerability.

2.4 Mapped results

2.4.1 *Expected current conditions*

Coldwater guild probability of presence was calculated for all 1:100k stream catchments in the study area using the BRT model. The predicted probability values ranged from 0 to 1, where 0 = absent and 1 = 100% probability of presence. The mean predicted probability of presence across the region was 0.143. Of the total 641,615 catchments, less than 1% (5,458 catchments) had a predicted probability of presence greater than 0.75, and about 5.5% (36,286 catchments) had a predicted probability of presence between 0.5 and 0.75. These results are mapped in Figure 7.

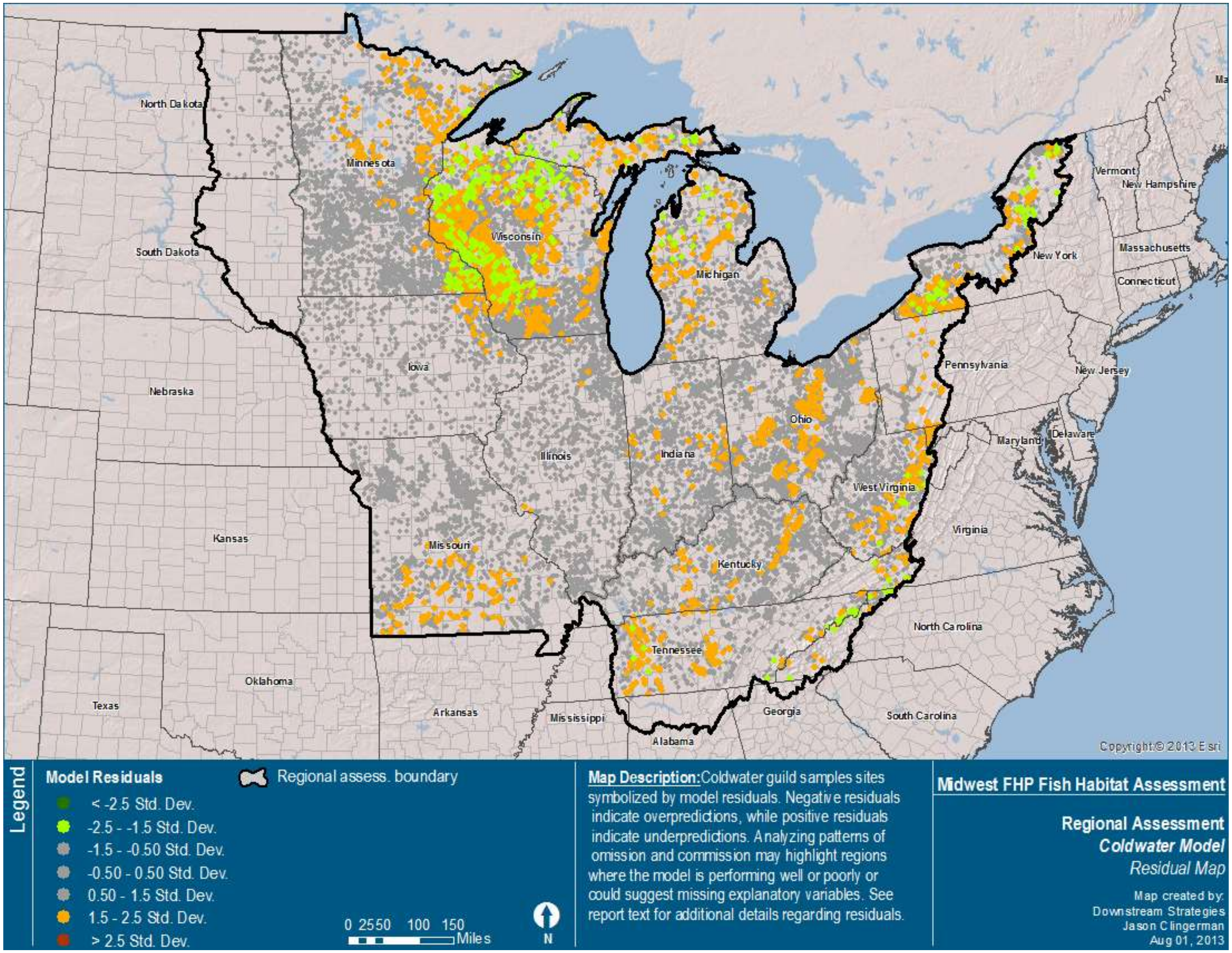
Figure 7: Expected coldwater guild distribution



2.4.2 *Spatial variability in predictive performance*

Analyzing patterns of omission and commission may highlight regions where the model is performing well or poorly or could suggest missing explanatory variables. To assess omission and commission, residuals were calculated by the BRT model. The residuals are a measure of the difference in the measured and modeled values (measured value *minus* modeled value). Negative residuals indicate overpredictions (predicting higher values than are true), while positive residuals indicate underpredictions (predicting lower values than are true). Figure 8 shows the distribution of model residuals per sampling site.

Figure 8: Distribution of coldwater guild model residuals by sampling site



2.4.3 *Indices of stress and natural quality*

Maps of HQI and ASI illustrate the spatial distribution of natural habitat potential (i.e., HQI score) and anthropogenic stress (i.e., ASI score) throughout the Midwest region. HQI and ASI scores are mapped in Figure 9 and Figure 10, respectively. The two variables contributing toward the calculation of ASI are mapped in Figure 11 and Figure 12. HQI, ASI, and their metrics are all scaled on a 0-1 scale (see Sections 1.3.2 and 2.3 for more details on HQI and ASI calculation). For HQI, higher values indicate higher natural quality, while higher values for ASI indicate higher levels of anthropogenic stress.

At first glance, it may seem that regional stress conditions are overly optimistic, but it is necessary to consider that the stress index is showing areas where probability of presence for this response is reduced because of stressors. It is likely that stress on aquatic systems in general is much more widespread than is indicated in any individual model's stress maps. Stress from all three models is considered together in Section 5 of this report. For all stress and natural quality indices, all catchments are shown, even in areas where the probability of presence is low. This is necessary and useful to consider areas outside of the current expect range where stress could have caused a historic population to be extirpated.

Figure 9: Natural habitat quality index for coldwater guild

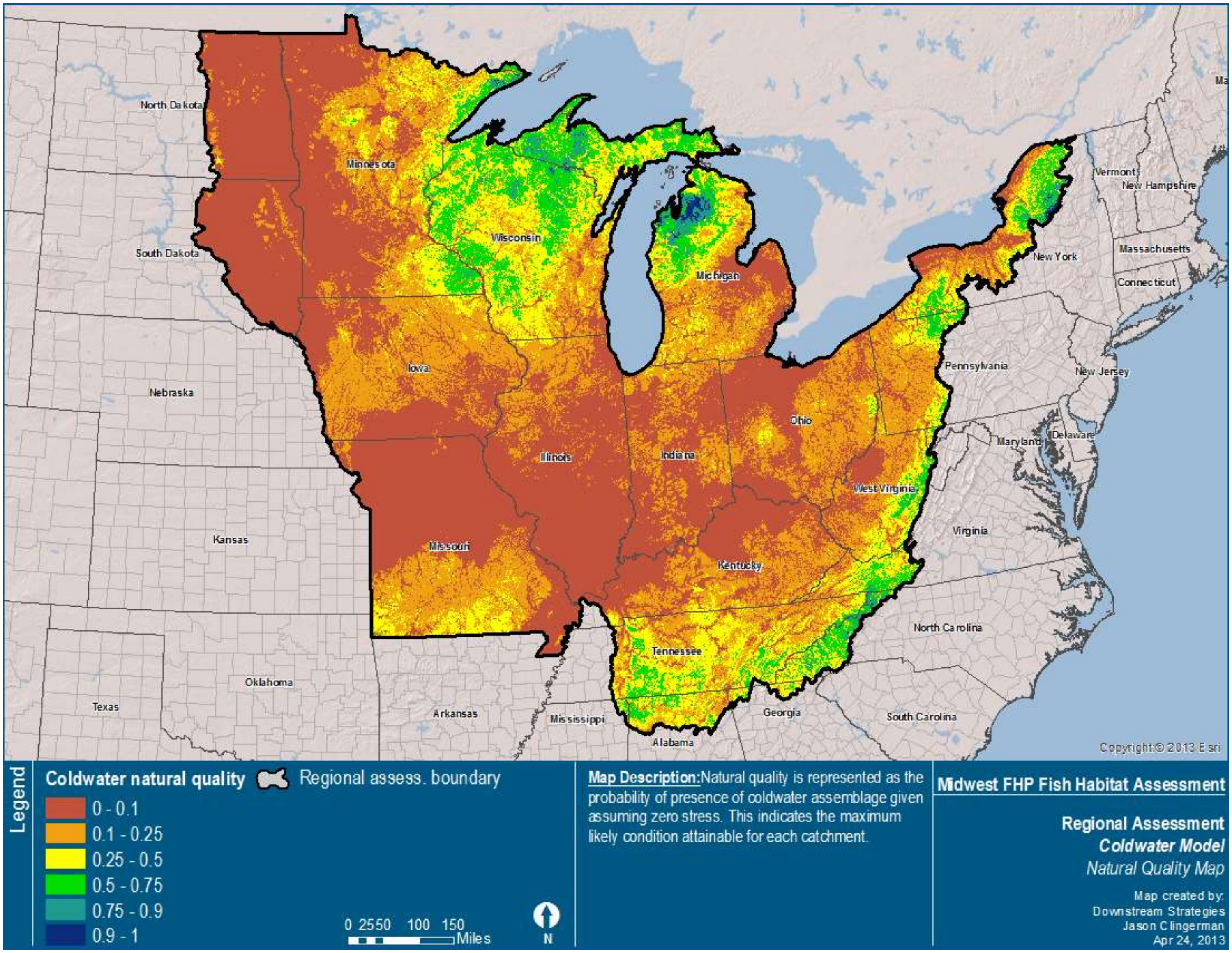


Figure 10: Anthropogenic stress index for coldwater guild

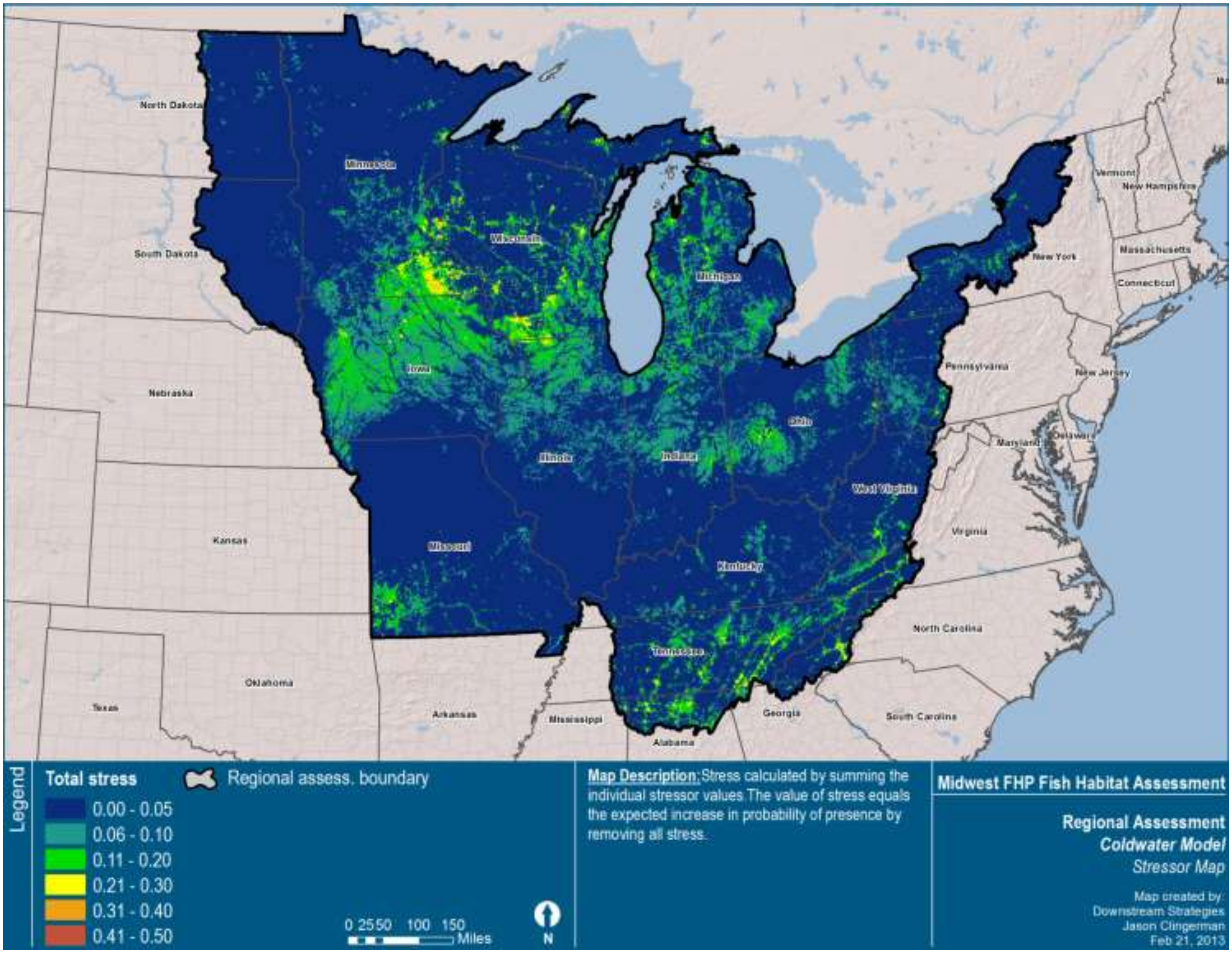


Figure 11: Agriculture stressor metric for coldwater guild

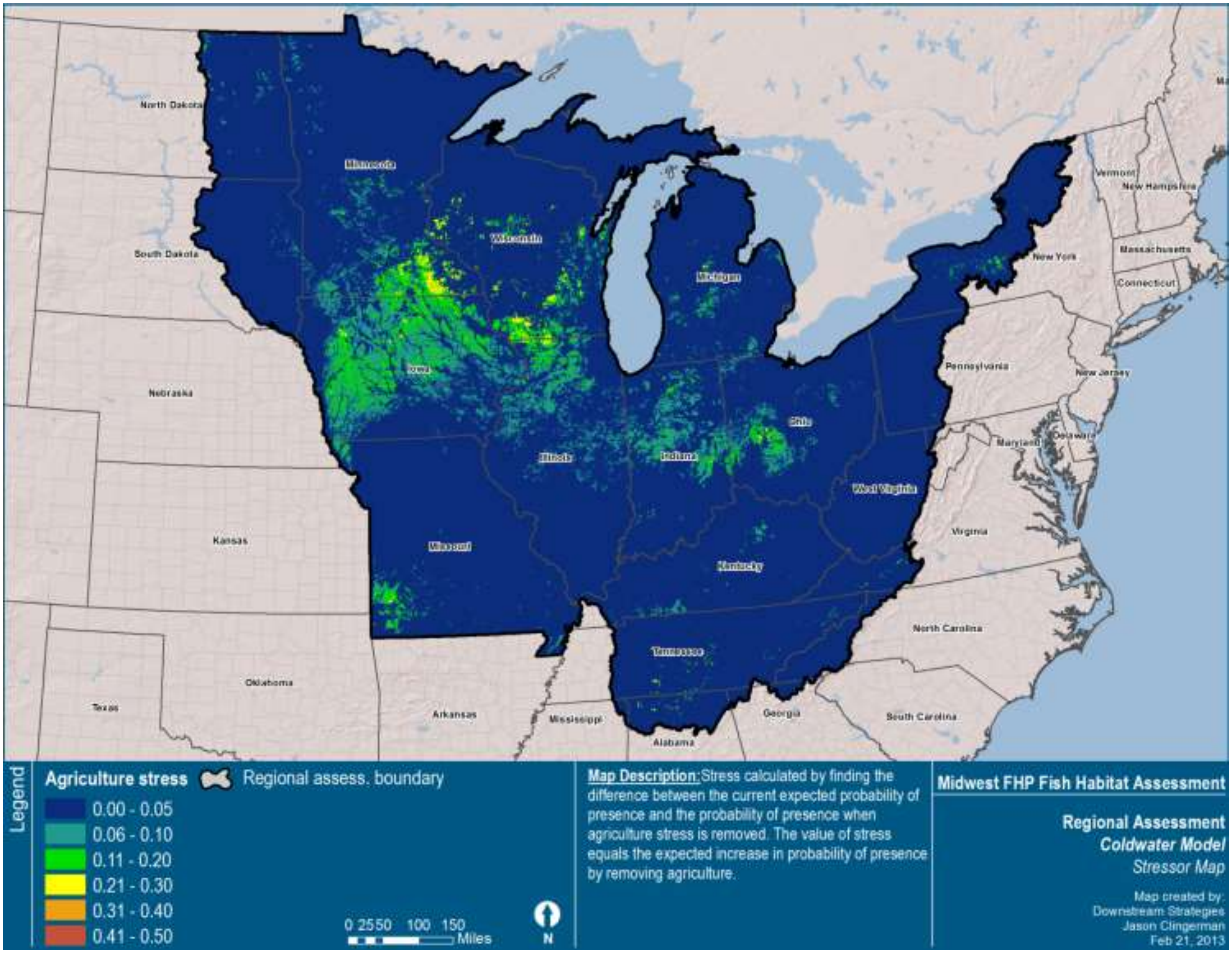
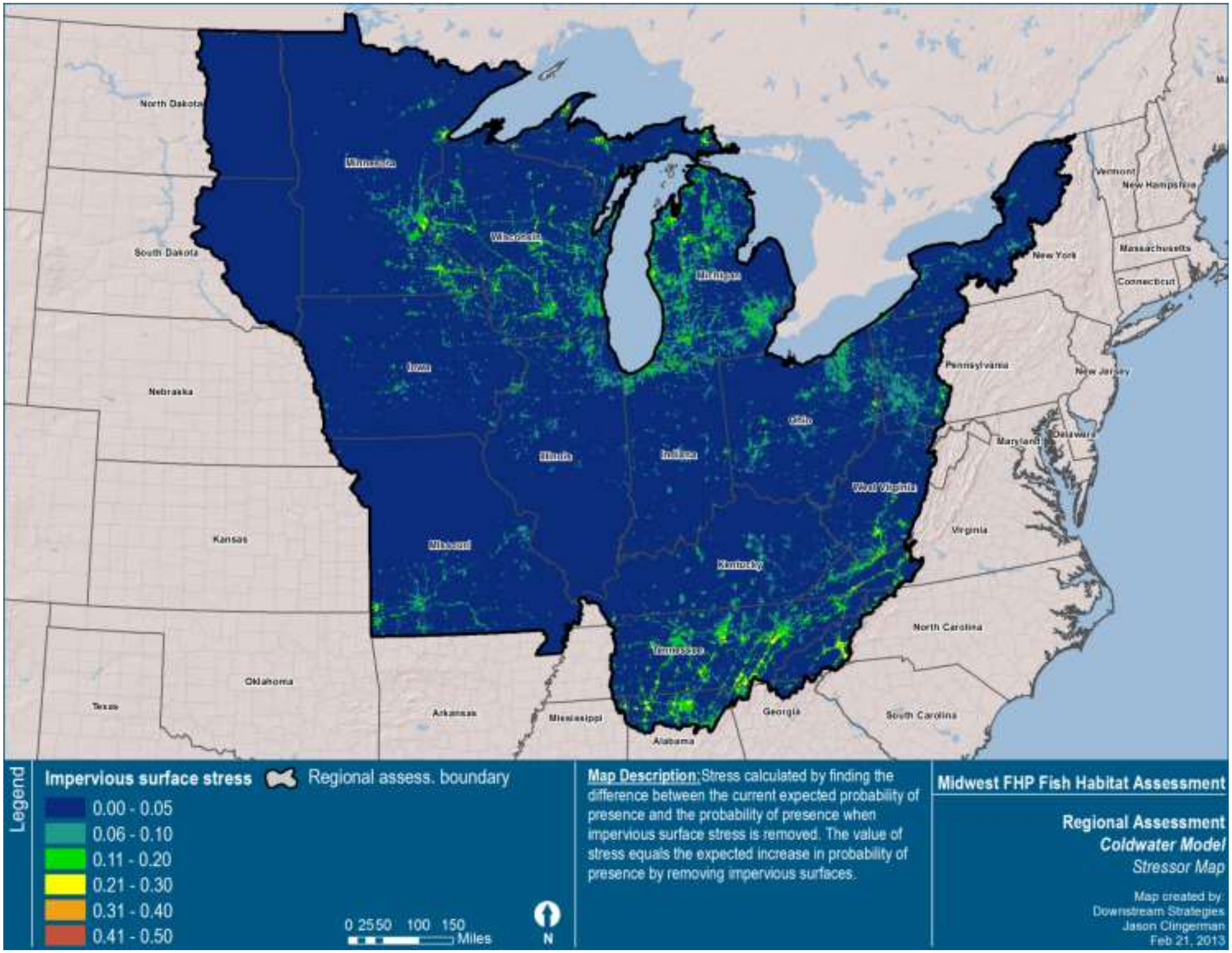


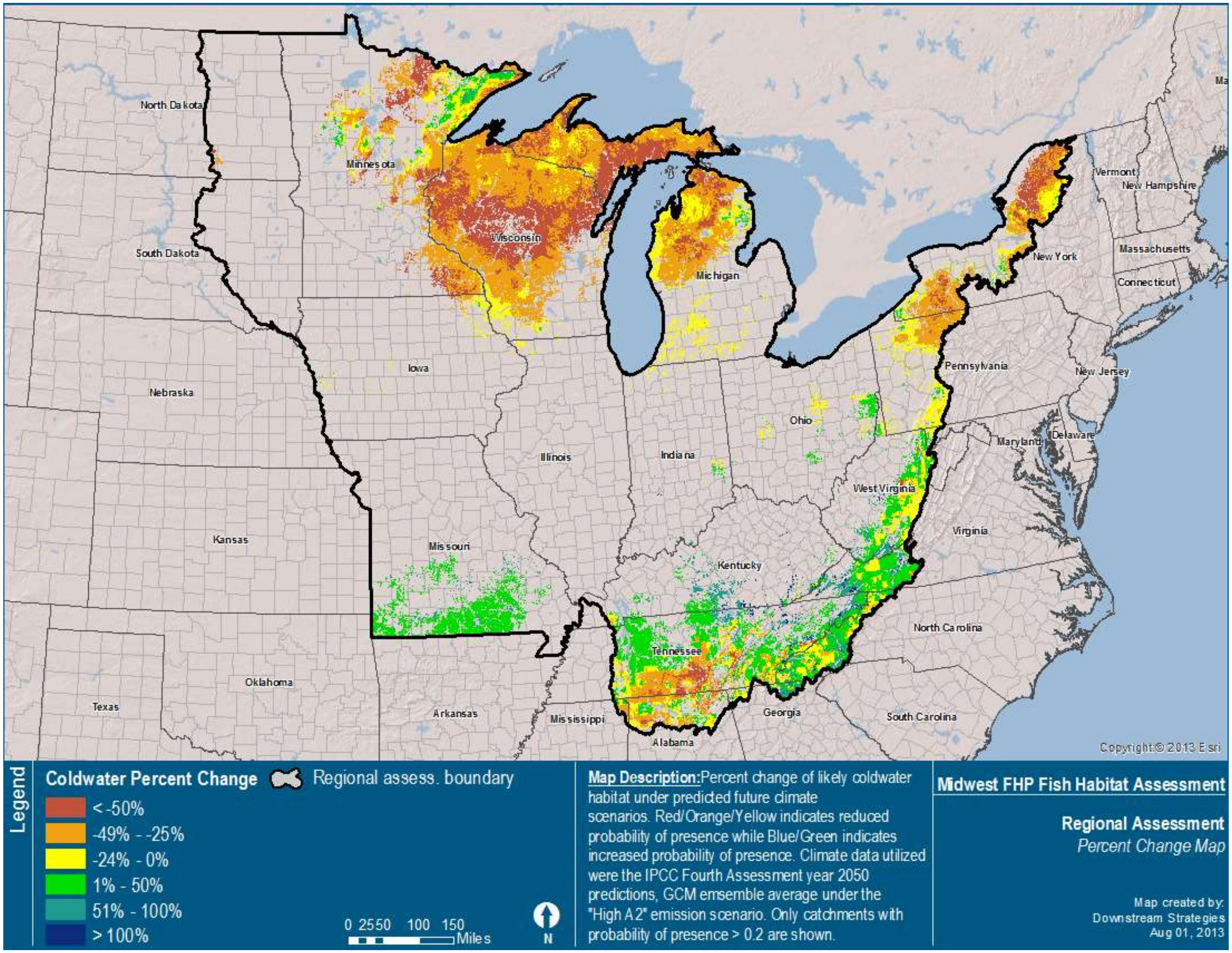
Figure 12: Impervious surface stressor metric for coldwater guild



2.4.4 *Potential future climate scenario*

A map of percent change in coldwater habitat probability of presence based on a 2050 IPCC A2 future climate change scenario (see Section 1.3.3 for further explanation) is shown in Figure 13. The percent change values should be interpreted as a measure of how susceptible each catchment may be to climate change. Positive percent change indicates the probability is expected to increase under the future climate scenario, while a negative percent change indicates a lower probability to be expected under the 2050 climate scenario. To ensure that only habitats that are likely to contain coldwater fish guilds; only catchments where the current probability of presence is greater than 0.20 are shown in this figure. This cutoff level was selected after visualizing the data to ensure that it was effective at removing areas not likely to contain coldwater habitat while still adequately portraying the potential effect of future climate scenarios upon the expected coldwater habitat.

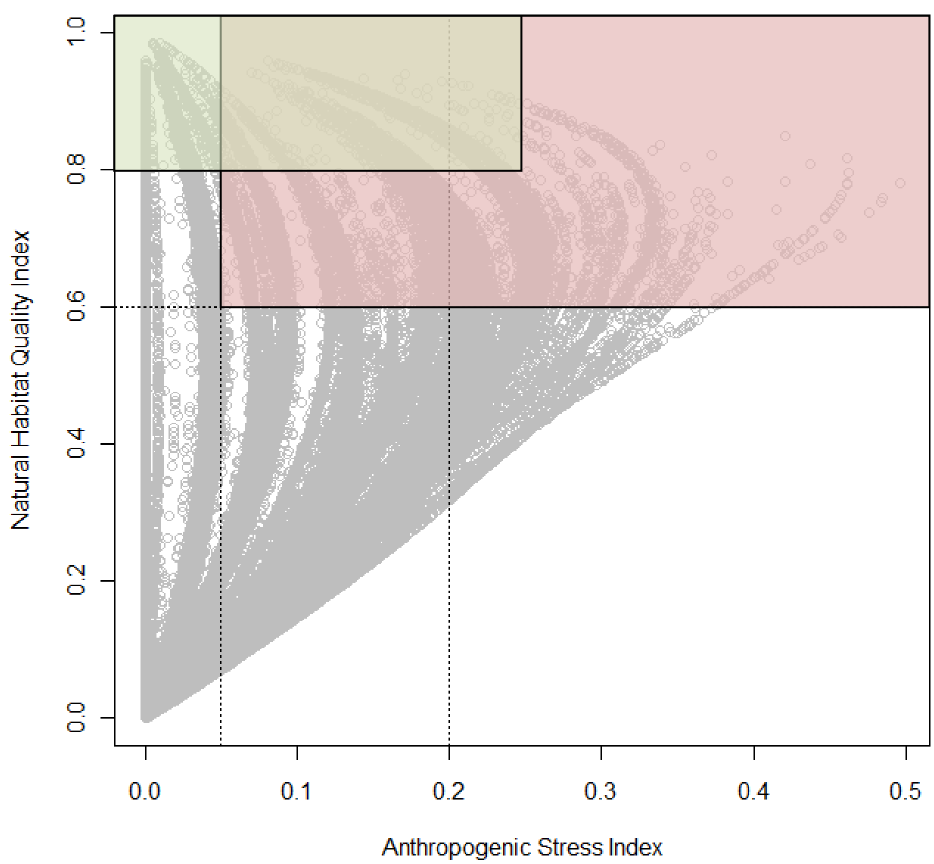
Figure 13: Potential climate change scenario for coldwater habitat



2.4.5 *Restoration and protection priorities*

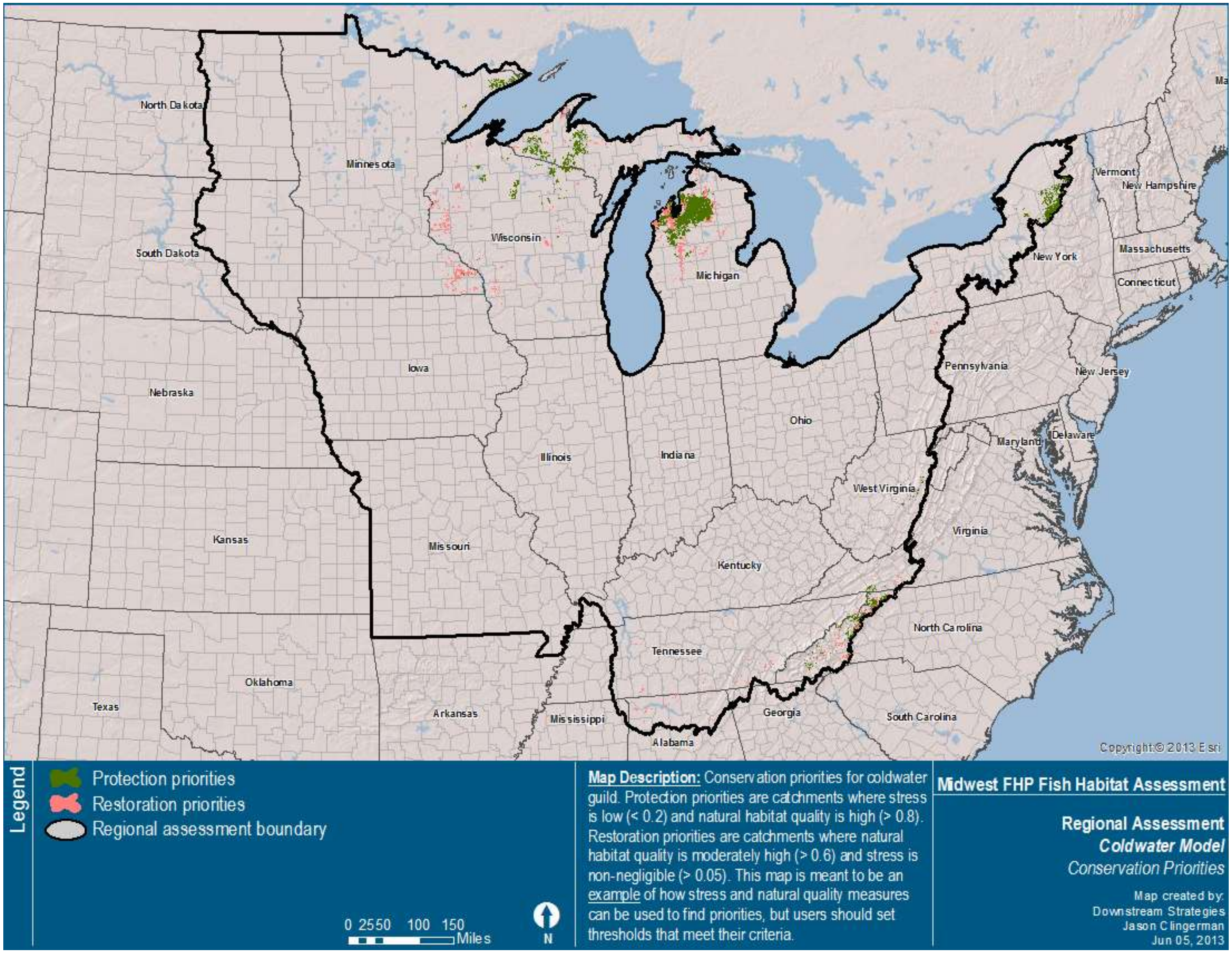
A plot of HQI versus ASI values for all catchments in the study area can be used as a reference to define HQI and ASI thresholds when evaluating restoration and protection priorities (Figure 14). In the example shown (Figure 15), thresholds for protection priorities were defined as catchments with high natural habitat quality and low anthropogenic stress; these thresholds were based on HQI greater than 0.8 and ASI less than 0.2. The thresholds used to identify restoration priorities were defined as catchments with high natural habitat quality and moderate to high anthropogenic stress; these thresholds were based on HQI greater than 0.6 and ASI greater than 0.05. Due to the methodology used to set these thresholds, there is potential for certain catchments to be classified as both restoration and protection priorities (Figure 14), in these cases protection priority overrides restoration when mapped in Figure 15. These thresholds were solely based on the relative scores for natural quality and stress indices. Though this example scenario provides an informed set of criteria for identifying conservation priorities, it is only intended to demonstrate the functionality of querying catchments based on these attributes to identify areas that meet user-defined criteria to guide conservation, protection, and restoration planning.

Figure 14: Coldwater guild HQI versus ASI values for all catchments



Note: The red box indicates catchments defined as restoration priorities under the example scenario. The green box indicates catchments defined as protection priorities under the same scenario.

Figure 15: Example restoration and protection priorities for coldwater guild



3. COOLWATER GUILD

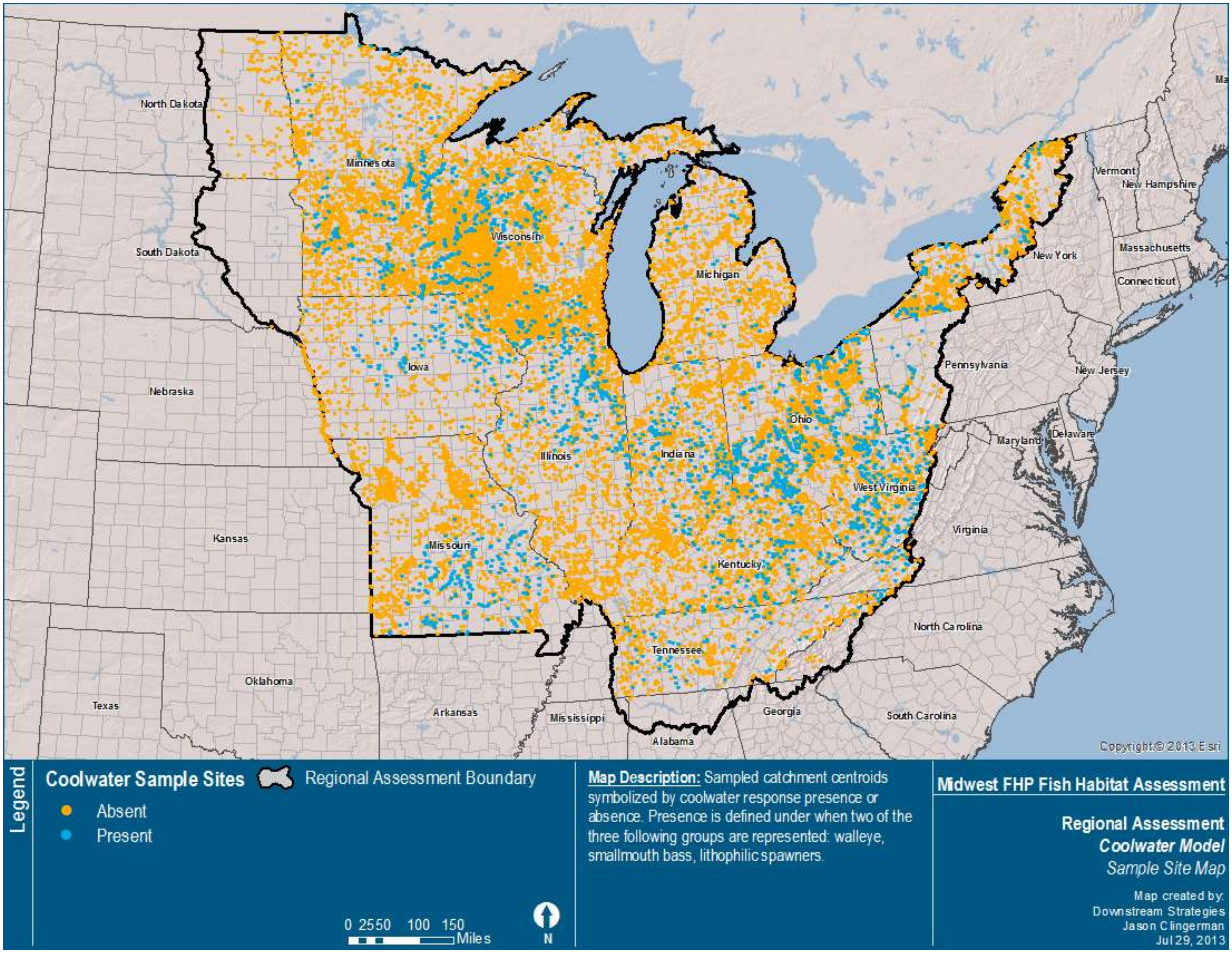
3.1 Modeling inputs

DS coordinated with a team of Fish Habitat Partnership scientists from the Midwest and Great Plains to construct a coolwater guild response variable and model the predicted probability of presence across the Midwest region. For this assessment, a fish guild is defined as a group of fish that have similar habitat requirements and are relatively intolerant to habitat degredation. This response was a presence-absence response, with presences being indicated when a fish from two of the three following groups was present in a sample. Group one contained only smallmouth bass (*Micropterus dolomieu*), group two contained only walleye (*Sander vitreus*) and group 3 contained several lithophilic spawning species. The lithophilic species used in group three were: northern hogsucker (*Hypentelium nigricans*), gravel chub (*Erimystax x-punctatus*), slenderhead darter (*Percina phoxocephala*), rainbow darter (*Etheostoma caeruleum*), gilt darter (*Percina evides*), channel darter (*Percina copelandi*), bigeye chub (*Hybopsis amblops/Notropis amblops*), banded darter (*Etheostoma zonale*), and Iowa darter (*Etheostoma exile*). Absences were assumed where these species/scenarios were not found in community sample data.

Individual Fish Habitat Partnerships (FHPs) provided DS with fish data collected in streams over a time frame spanning from 1995 to 2011. Data collected by the FHPs generally came from state wildlife and fisheries agencies or from other reliable sources such as universities. DS then processed that data to create a presence-absence dataset for this coolwater fish guild which was comprised of 18,908 observations. Figure 16 maps all of the sampling sites that were used to construct the model and outlines the regional assessment boundary. Model outputs were applied to all 1:100k catchments within the regional assessment boundary.

DS cooperated with Midwest and Great Plains FHP Science Team to arrive at a list of landscape-based habitat variables used to predict coolwater guild habitat throughout the region. These variables were also used to characterize habitat quality and anthropogenic stress. Building on the science team's input, DS compiled a list of 67 predictors for evaluation. Preliminary exploratory models were then run to identify variable predictive performance and statistical redundancy. From that list, 53 variables were removed due to statistical redundancy ($r > 0.6$), logical redundancy, or poor predictive performance (relative influence < 1.0 in preliminary model run). This resulted in a final list of 14 predictor variables for the BRT model and assessment. See Appendix A for a full data dictionary and the metadata document for variable processing notes.

Figure 16: Coolwater species modeling area and sampling sites



3.2 Modeling process

3.2.1 *Predictive performance*

The final selected model was comprised of 7,300 trees. The model had a CV correlation statistic of 0.493±0.001 and a CV ROC score of 0.859±0.005.

3.2.2 *Variable influence*

The BRT output includes a list of the predictor variables used in the model ordered and scored by their relative importance. The relative importance values are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman and Meulman, 2003). The relative influence score is scaled so that the sum of the scores for all variables is 100, where higher numbers indicate greater influence. The relative influence table for the coolwater guild model is shown below in Table 4.

Table 4: Relative influence of all variables in the final coolwater species model

Variable code	Variable description	Relative influence
AREASQKMC	Network drainage area	58.02
TEMP	Mean annual air temperature	9.87
MINELEVRAW	Minimum catchment elevation	4.69
SLOPE	Slope of catchment flowline	3.87
AG_PC	Network agriculture land cover	3.35
WET_P	Catchment wetland cover	3.01
BR5PC	Network sand/gravel bedrock geology	2.58
SHR_PC	Network shrub/scrub cover	2.41
BR1PC	Network carbonate bedrock geology	2.39
IMPSURF_MC	Network impervious surface cover	2.09
ROADLEN_den	Catchment road density	2.06
BR4PC	Network metamorphic bedrock geology	2.02
BR3PC	Network mafic bedrock geology	1.97
TRIC_den	Network toxic release inventory density	1.68

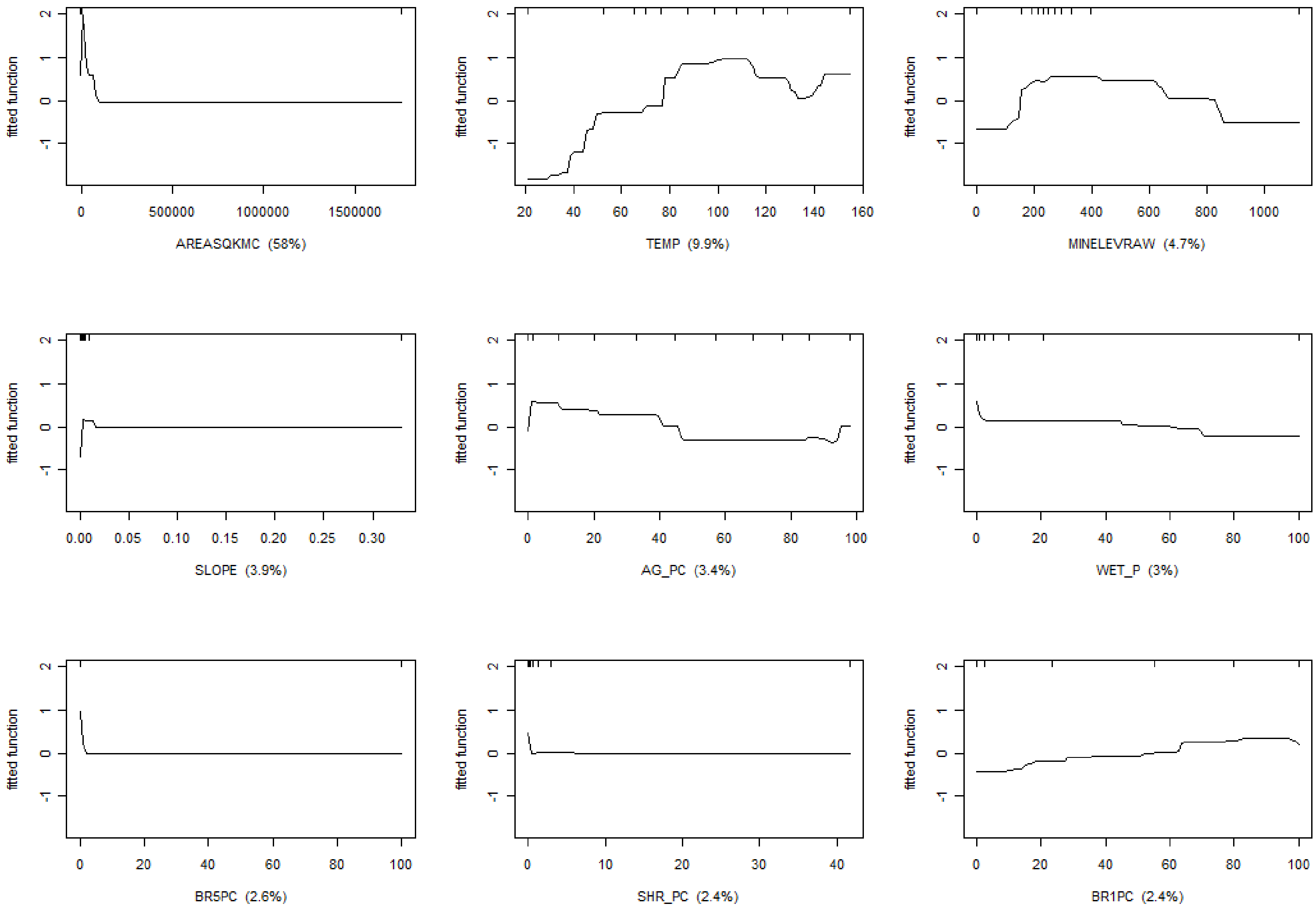
Note: Individual variables are highlighted according to whether they were determined to be anthropogenic (grey shading) or natural (no shading).

3.2.3 *Variable functions*

The BRT output also contains quantitative information on partial dependence functions that can be plotted to visualize the effect of each individual predictor variable on the response after accounting for all other variables in the model. Similar to the interpretation of traditional regression coefficients, the function plots are not always a perfect representation of the relationship for each variable, particularly if interactions are strong or predictors are strongly correlated. However, they do provide a useful and objective basis for interpretation (Friedman, 2001; Friedman and Meulman, 2003).

These plots show the trend of the response variable (y-axis) as the predictor variable (x-axis) changes. The response variable is transformed (usually to the logit scale) so that the magnitude of trends for each predictor variable's function plot can be accurately compared. The dash marks at the top of each function represent the deciles of the data used to build the model. The function plots for the nine most influential variables in the coolwater species model (Table 4) are illustrated in Figure 17. The plots for all variables are shown in Appendix B.

Figure 17: Functional responses of the dependent variable to individual predictors of coolwater species



Note: Only the top nine predictors, based on relative influence (shown in parentheses; see Appendix A for descriptions of variable codes), are shown here. See Appendix B for plots of remaining predictor variables.

3.3 Post-modeling

3.3.1 *Stress and natural quality*

The variable importance table and partial dependence functions of the final BRT model were used to assess the potential stressors for the coolwater guild model. Within the model, there were four variables considered anthropogenic in nature (Table 3). After reviewing the functional relationships of these four potential stressors, two of the four stressors were removed from ASI calculations. These variables ('ROADLEN_den' and 'TRIC_den') had function plots that were unintuitive: their relationships to the response likely captured some sort of regional variation in the model rather than a mechanistic relationship with the response. The two remaining stressors, network agriculture land cover (AG_PC) and network impervious surface cover (IMPSURF_MC), were used to calculate ASI for the coolwater guild model. Section 1.3.2 details how ASI and HQI were calculated for each model.

3.3.2 *Potential future climate scenario*

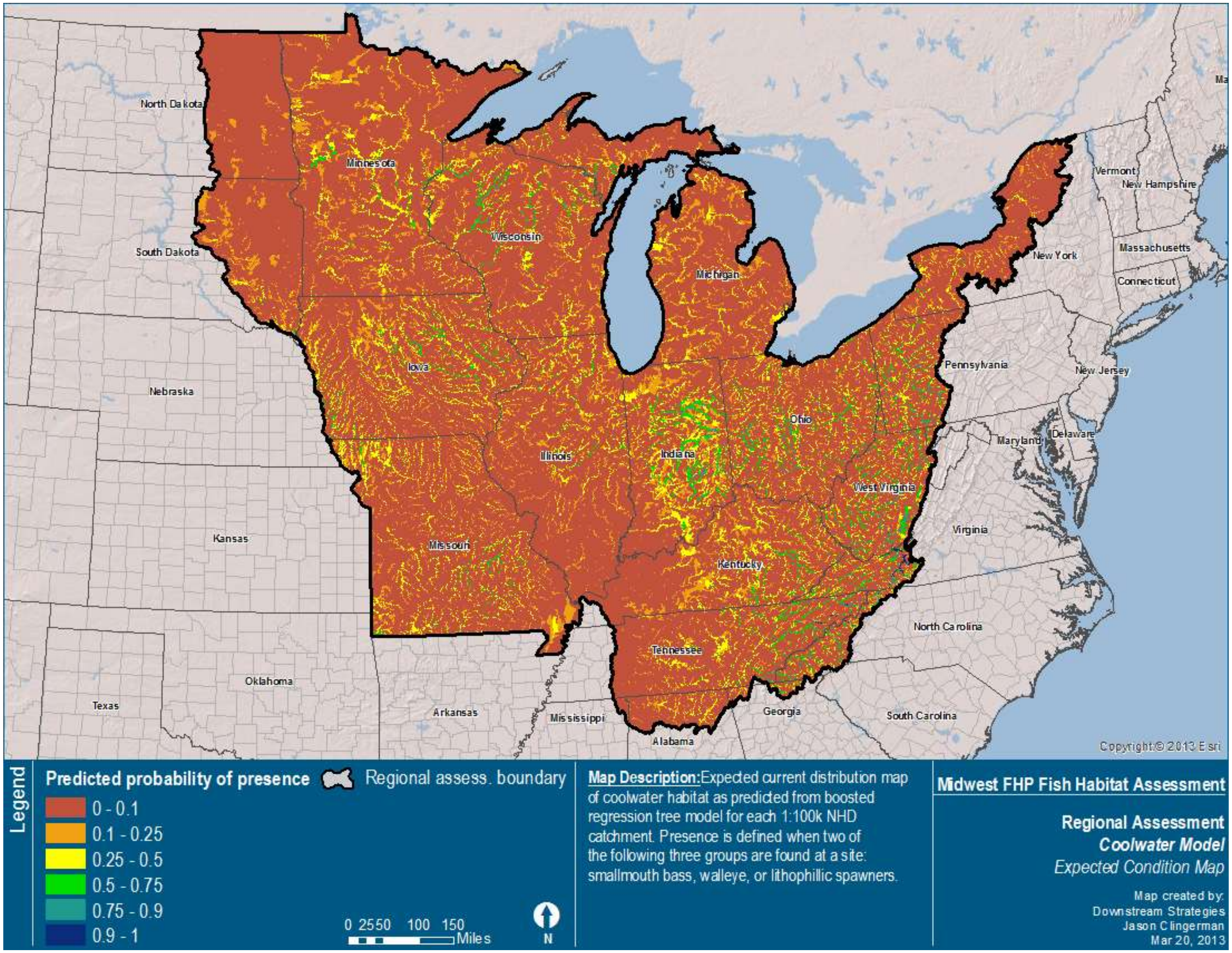
The coolwater BRT model was extrapolated onto a dataset that contained future climate data as described in Section 1.3.3. The potential future predictions were then compared to the current predictions Percent change in probability was calculated for each catchment to assess climate change vulnerability.

3.4 Mapped Results

3.4.1 *Expected current conditions*

Coolwater species probability of presence was calculated for all 1:100k stream catchments in the study area using the BRT model. The predicted probability values ranged from 0 to 1. The mean predicted probability was 0.079. Of the total 641,615 catchments, less than 0.5% (2,637 catchments) had a predicted probability of presence greater than 0.75, and about 2.5% (16,088 catchments) had a predicted probability of presence between 0.5 and 0.75. These results are mapped in Figure 18.

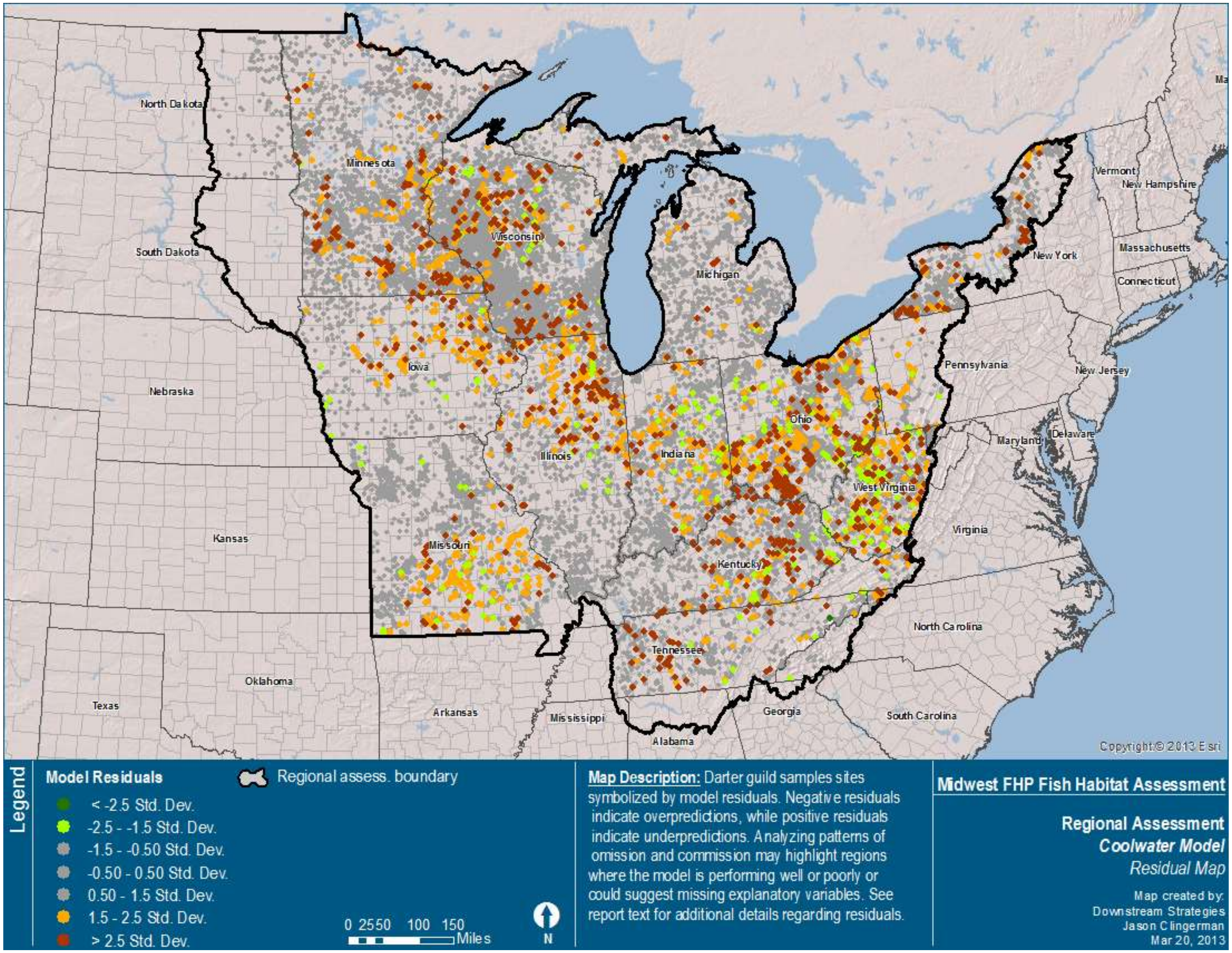
Figure 18: Expected coolwater species distribution



3.4.2 *Spatial variability in predictive performance*

Analyzing patterns of omission and commission may highlight regions where the model is performing well or poorly or could suggest missing explanatory variables. To assess omission and commission, residuals were calculated by the BRT model. The residuals are a measure of the difference in the measured and modeled values (measured value *minus* modeled value). Negative residuals indicate overpredictions (predicting higher values than are true), while positive residuals indicate underpredictions (predicting lower values than are true). Figure 19 shows the distribution of model residuals per sampling site.

Figure 19: Distribution of coolwater species model residuals by sampling site



3.4.3 *Indices of stress and natural quality*

Maps of HQI and ASI illustrate the spatial distribution of natural habitat potential (i.e., HQI score) and anthropogenic stress (i.e., ASI score) throughout the Midwest region. HQI and ASI scores are mapped in Figure 20 and Figure 21, respectively. The two variables contributing toward the calculation of ASI are mapped in Figure 22 and Figure 23. HQI, ASI, and their metrics are all scaled on a 0-1 scale (see Sections 1.3.2 and 3.3 for more details on HQI and ASI calculation). For HQI, higher values indicate higher natural quality, while higher values for ASI indicate higher levels of anthropogenic stress.

At first glance, it may seem that regional stress conditions are overly optimistic, but it is necessary to consider that the stress index is showing areas where probability of presence for this response is reduced because of stressors. It is likely that stress on aquatic systems in general is much more widespread than is indicated in any individual model's stress maps. Stress from all three models is considered together in Section 5 of this report. For all stress and natural quality indices, all catchments are shown, even in areas where the probability of presence is low. This is necessary and useful to consider areas outside of the current expect range where stress could have caused a historic population to be extirpated.

Figure 20: Natural quality index for coolwater species

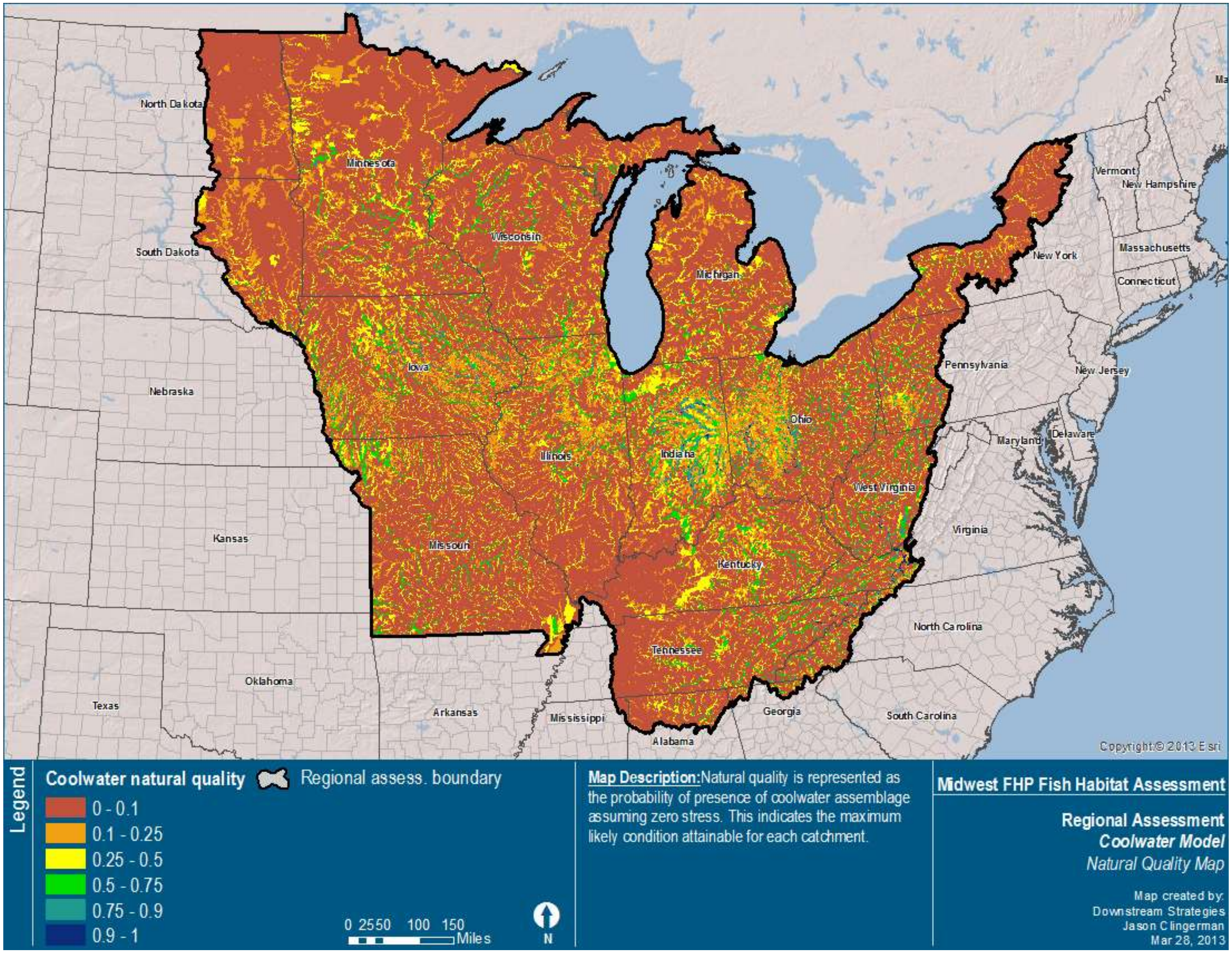


Figure 21: Anthropogenic stress index for coolwater species

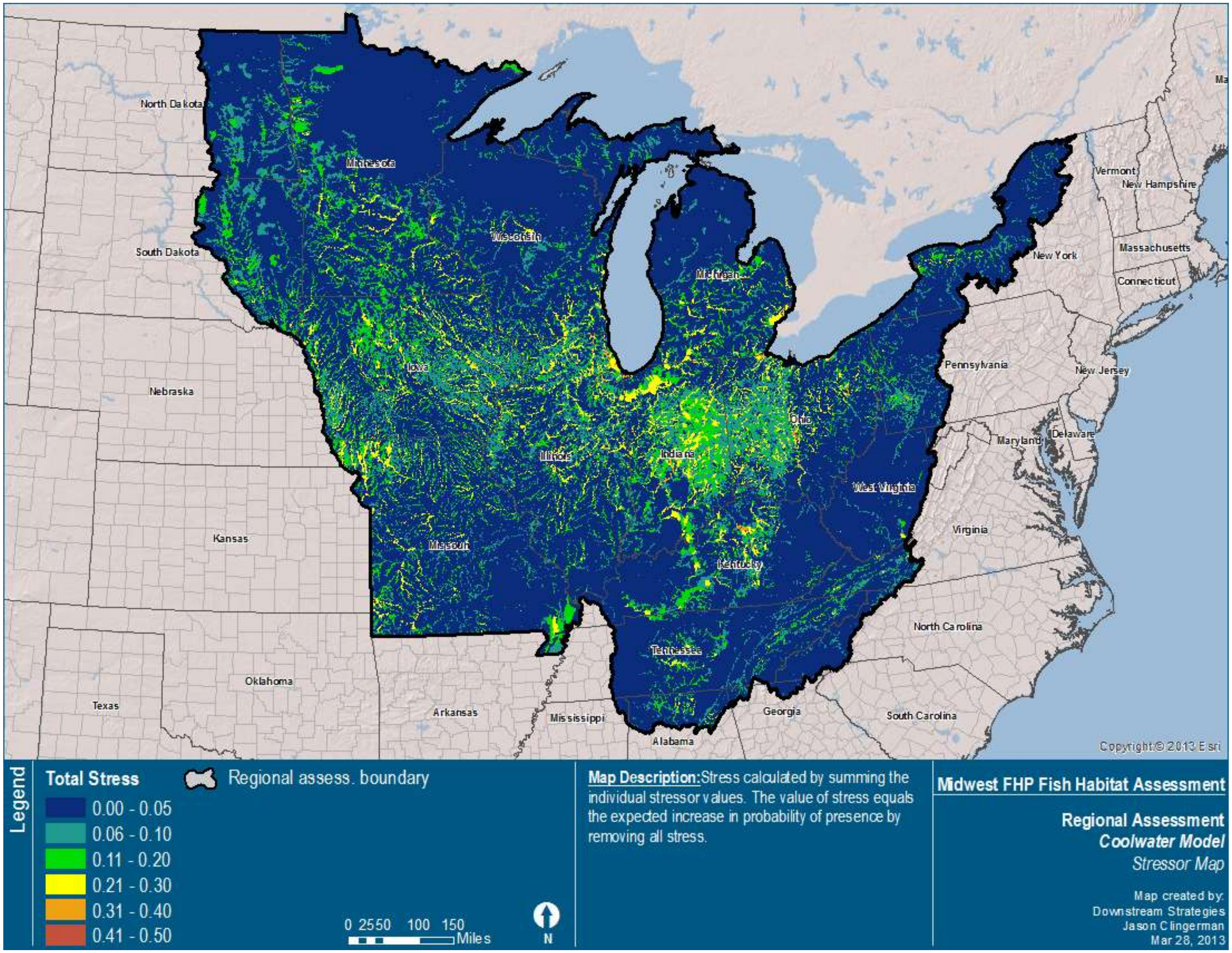


Figure 22: Agriculture stressor metric for coolwater species

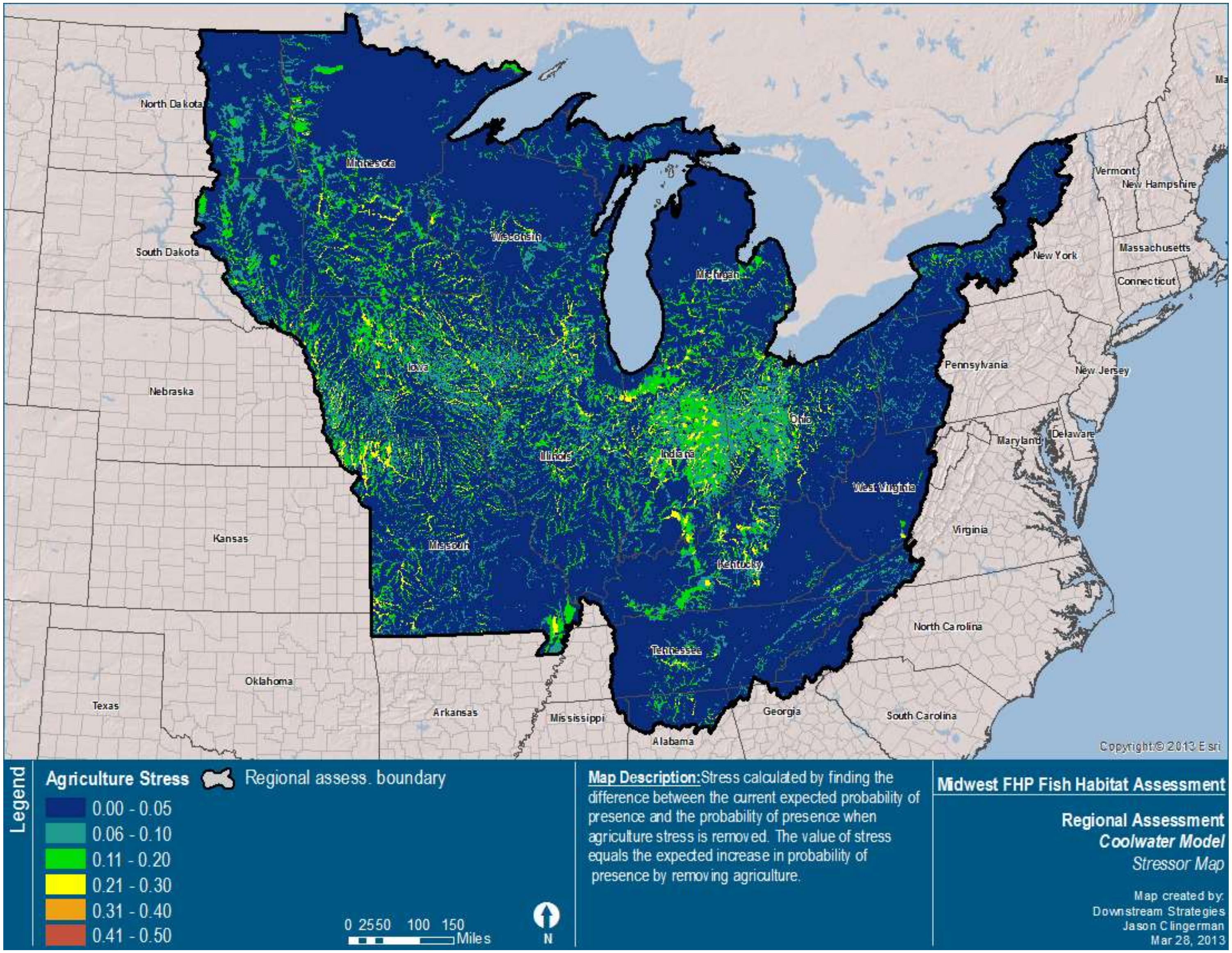
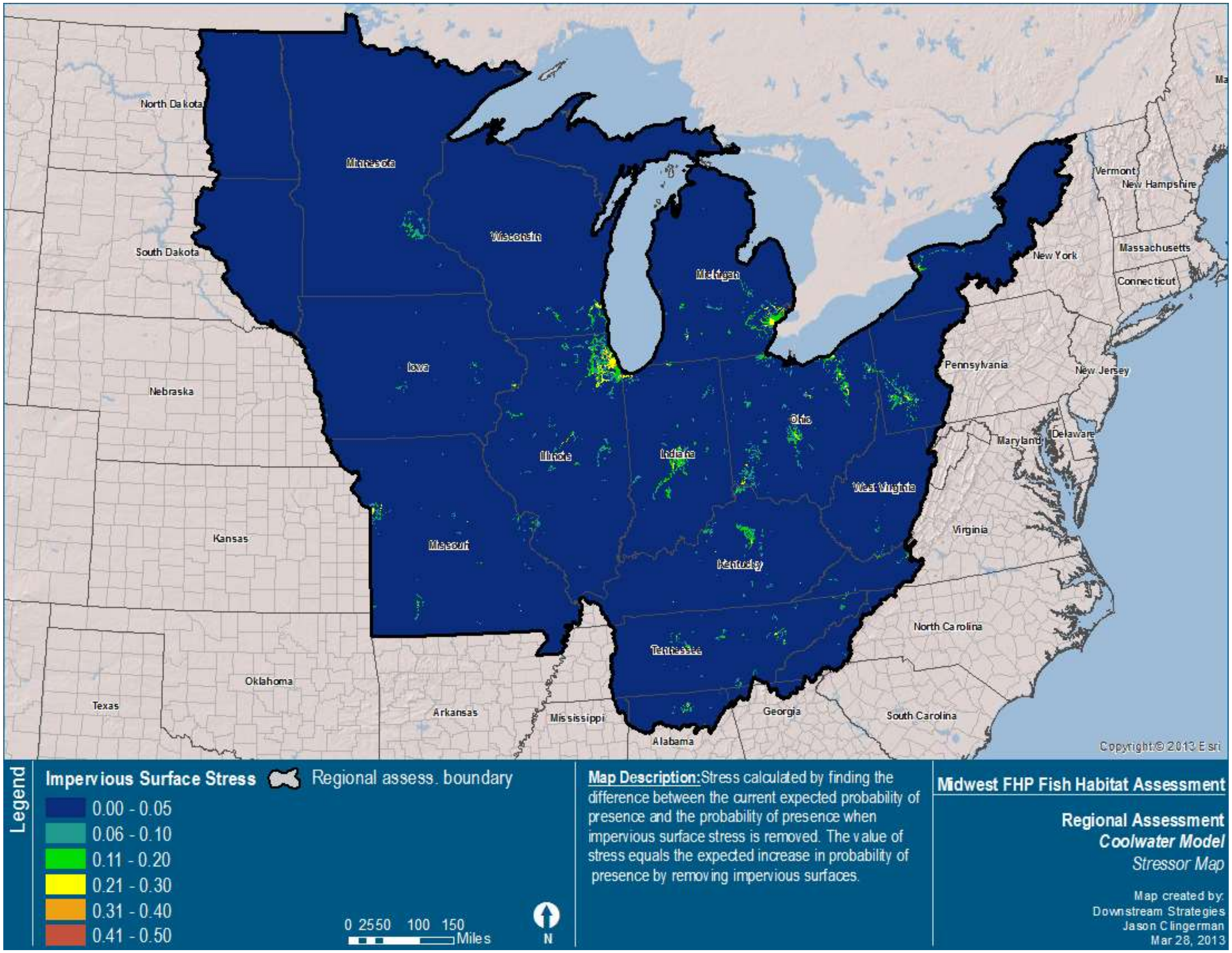


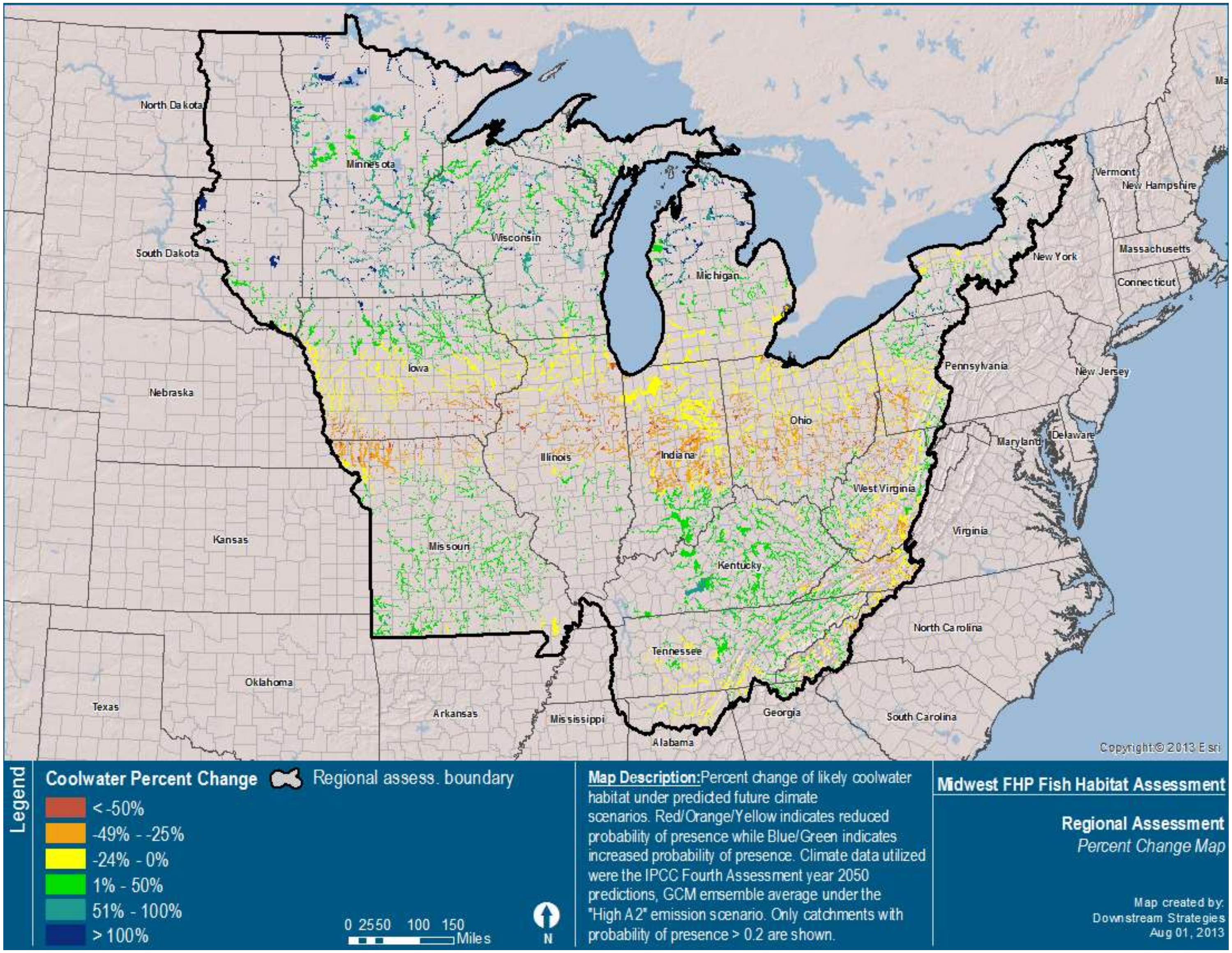
Figure 23: Impervious surface stressor metric for coolwater species



3.4.4 *Potential future climate scenario*

A map of percent change in coolwater habitat probability of presence based on a 2050 IPCC A2 future climate change scenario (see Section 1.3.3 for further explanation) is shown in Figure 24. The percent change values should be interpreted as a measure of how susceptible each catchment may be to climate change. Positive percent change indicates the probability is expected to increase under the future climate scenario, while a negative percent change indicates a lower probability to be expected under the 2050 climate scenario. To ensure that only habitats that are likely to contain coolwater fish guilds; only catchments where the current probability of presence is greater than 0.20 are shown in this figure. This cutoff level was selected after visualizing the data to ensure that it was effective at removing areas not likely to contain coolwater habitat while still adequately portraying the potential effect of future climate scenarios upon the expected coolwater habitat.

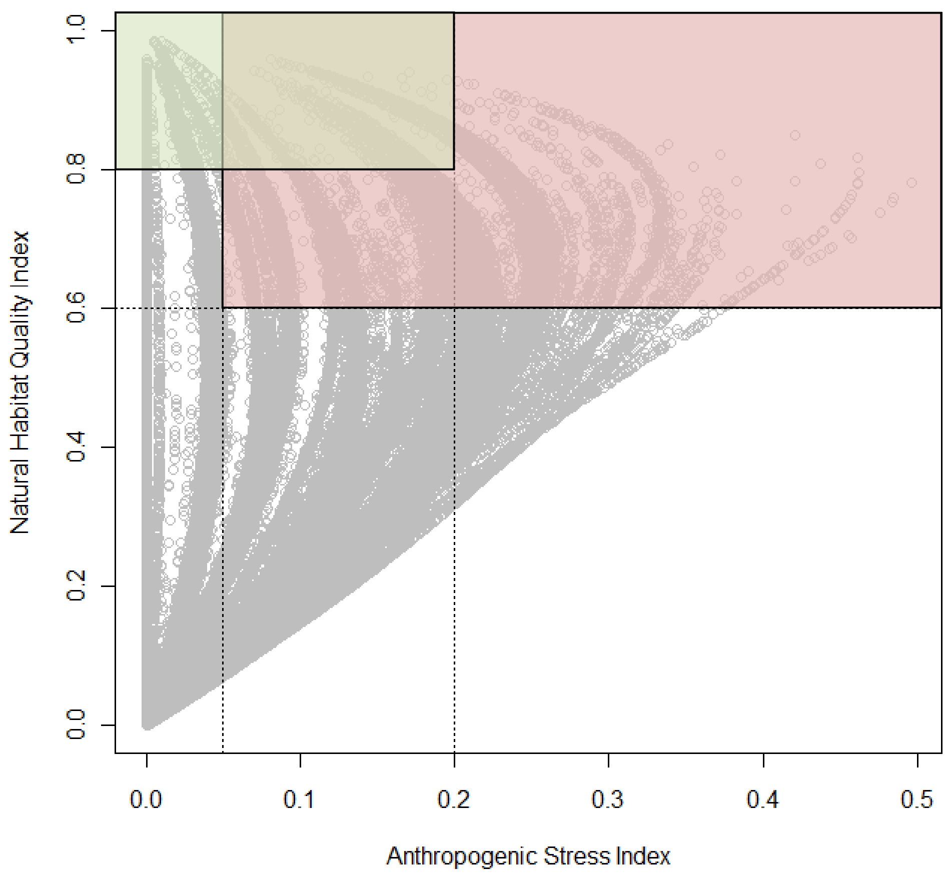
Figure 24: Potential climate change scenario for coolwater habitat



3.4.5 *Restoration and protection priorities*

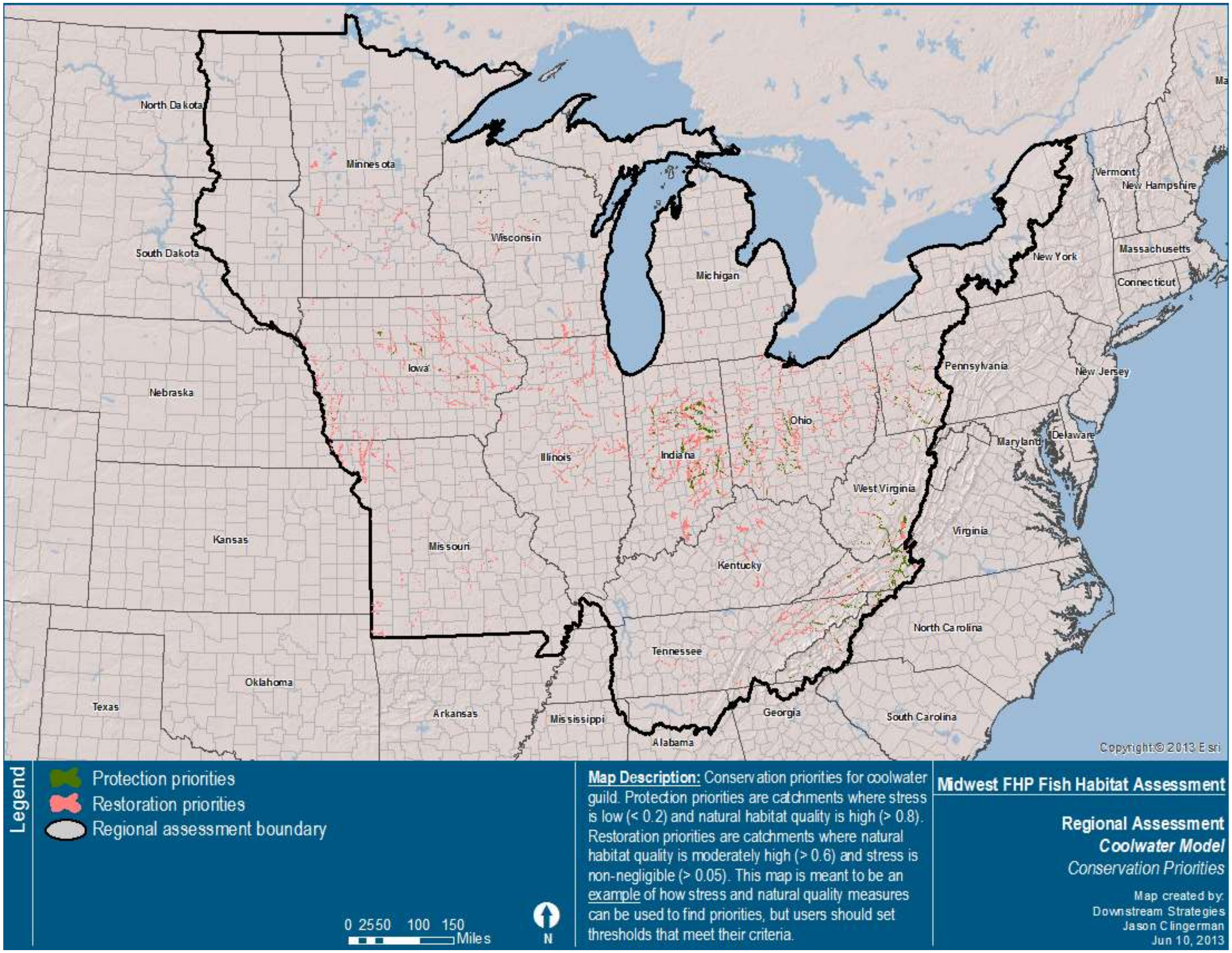
A plot of HQI versus ASI values for all catchments in the study area can be used as a reference to define HQI and ASI thresholds when evaluating restoration and protection priorities (Figure 25). In the example shown (Figure 26), thresholds for protection priorities were defined as catchments with high natural habitat quality and low anthropogenic stress; these thresholds were based on HQI greater than 0.8 and ASI less than 0.2. The thresholds used to identify restoration priorities were defined as catchments with high natural habitat quality and moderate to high anthropogenic stress; these thresholds were based on HQI greater than 0.6 and ASI greater than 0.05. Due to the methodology used to set these thresholds, there is potential for certain catchments to be classified as both restoration and protection priorities (Figure 25), in these cases protection priority overrides restoration when mapped in Figure 26. These thresholds were solely based on the relative scores for natural quality and stress indices. Though this example scenario provides an informed set of criteria for identifying conservation priorities, it is only intended to demonstrate the functionality of querying catchments based on these attributes to identify areas that meet user-defined criteria to guide conservation, protection, and restoration planning.

Figure 25: HQI versus ASI values for all catchments for coolwater species



Note: The red box indicates catchments defined as restoration priorities under the example scenario. The green box indicates catchments defined as protection priorities under the same scenario.

Figure 26: Restoration and protection priorities for coolwater species



4. WARMWATER GUILD

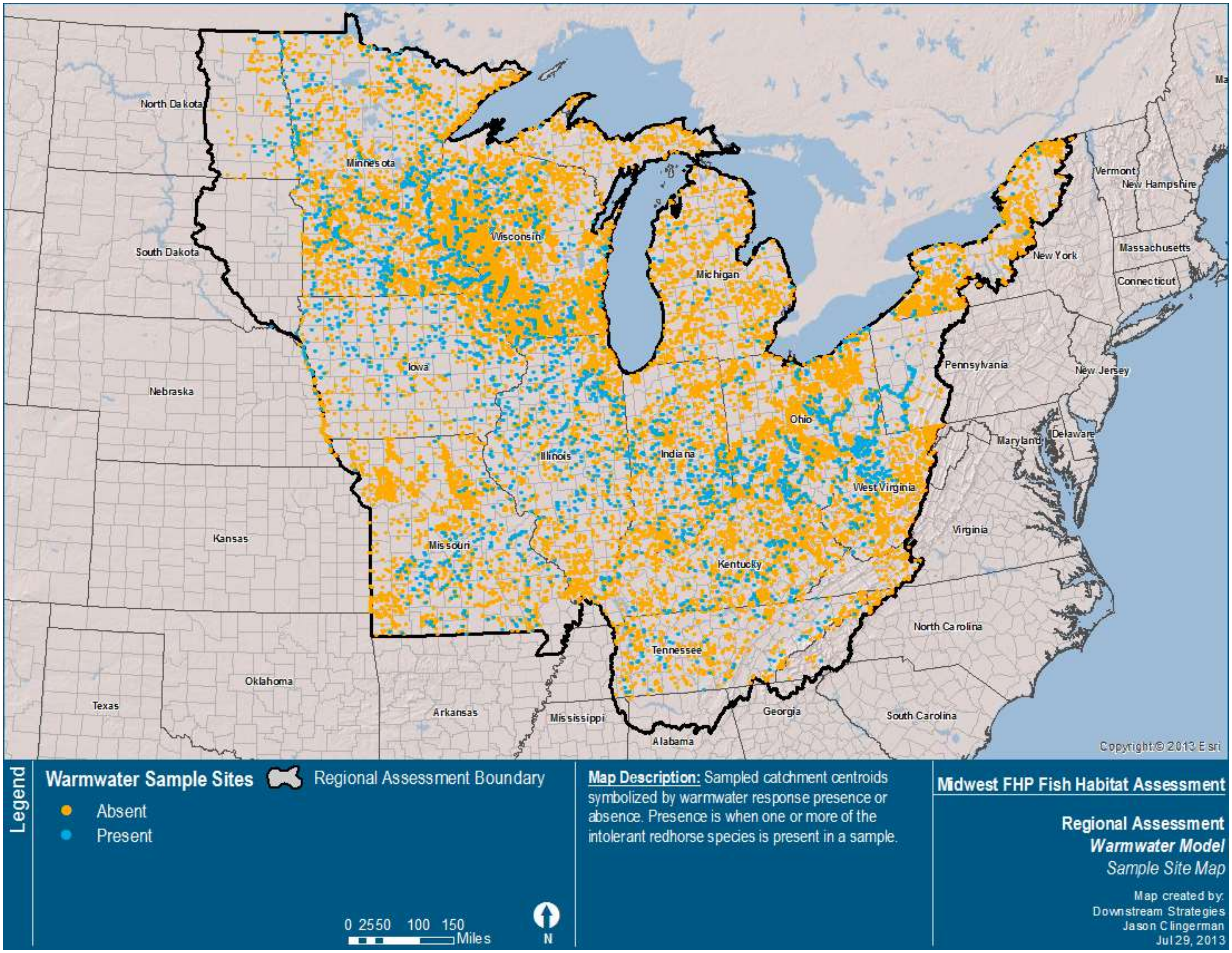
4.1 Modeling inputs

DS coordinated with a team of Fish Habitat Partnership scientists from the Midwest and Great Plains to construct a warmwater guild response variable and model the predicted probability of presence across the Midwest region. For this assessment, a fish guild is defined as a group of fish that have similar habitat requirements and are relatively intolerant to habitat degradation. This response was a presence-absence response, with presences being indicated when one of the following species was encountered in a sample: black redhorse (*Moxostoma duquesnei*), shorthead redhorse (*Moxostoma macrolepidotum*), river redhorse (*Moxostoma carinatum*), greater redhorse (*Moxostoma valenciennesi*), smallmouth redhorse (*Moxostoma breviceps*). Absences were assumed where these species/scenarios were not found in community sample data.

Individual Fish Habitat Partnerships (FHPs) provided DS with fish data collected in streams over a time frame spanning from 1995 to 2011. Data collected by the FHPs generally came from state wildlife and fisheries agencies or from other reliable sources such as universities. DS then processed those data to create a presence-absence dataset for this warmwater fish guild which was comprised of 18,908 observations. Figure 27 maps all of the sampling sites that were used to construct the model and outlines the regional assessment boundary. Model outputs were applied to all 1:100k catchments within the regional assessment boundary.

DS cooperated with Midwest and Great Plains FHP Science Team to arrive at a list of landscape-based habitat variables used to predict warmwater guild habitat throughout the region. These variables were also used to characterize habitat quality and anthropogenic stress. Building on the science team's input, DS compiled a list of 67 predictors for evaluation. Preliminary exploratory models were then run to identify variable predictive performance and statistical redundancy. From that list, 58 variables were removed due to statistical redundancy ($r > 0.6$), logical redundancy, or poor predictive performance (relative influence < 1.0 in preliminary model run). This resulted in a final list of 9 predictor variables for the BRT model and assessment. See Appendix A for a full data.

Figure 27: Warmwater guild modeling area and sampling sites



4.2 Modeling process

4.2.1 *Predictive performance*

The final selected model was comprised of 4,450 trees. The model had a CV correlation statistic of 0.565±0.006 and a CV ROC score of 0.874±0.004.

4.2.2 *Variable influence*

The BRT output includes a list of the predictor variables used in the model ordered and scored by their relative importance. The relative importance values are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman and Meulman, 2003). The relative influence score is scaled so that the sum of the scores for all variables is 100, where higher numbers indicate greater influence. The relative influence table for the warmwater guild model is shown below (Table 5).

Table 5: Relative influence of all variables in the final warmwater guild model

Variable code	Variable description	Relative influence
AREASQKMC	Network drainage area	82.39
RECH_MEANC	Network mean recharge rate	5.60
MINELEVRAW	Minimum catchment elevation	4.11
TEMP	Mean annual air temperature	1.70
SHR_PC	Network shrub/scrub cover	1.66
TRIC_den	Network toxic release inventory density	1.33
IMPSURF_MC	Network impervious surface cover	1.12
AG_PC	Network agriculture land cover	1.07
BR5PC	Network sand/gravel bedrock geology	1.02

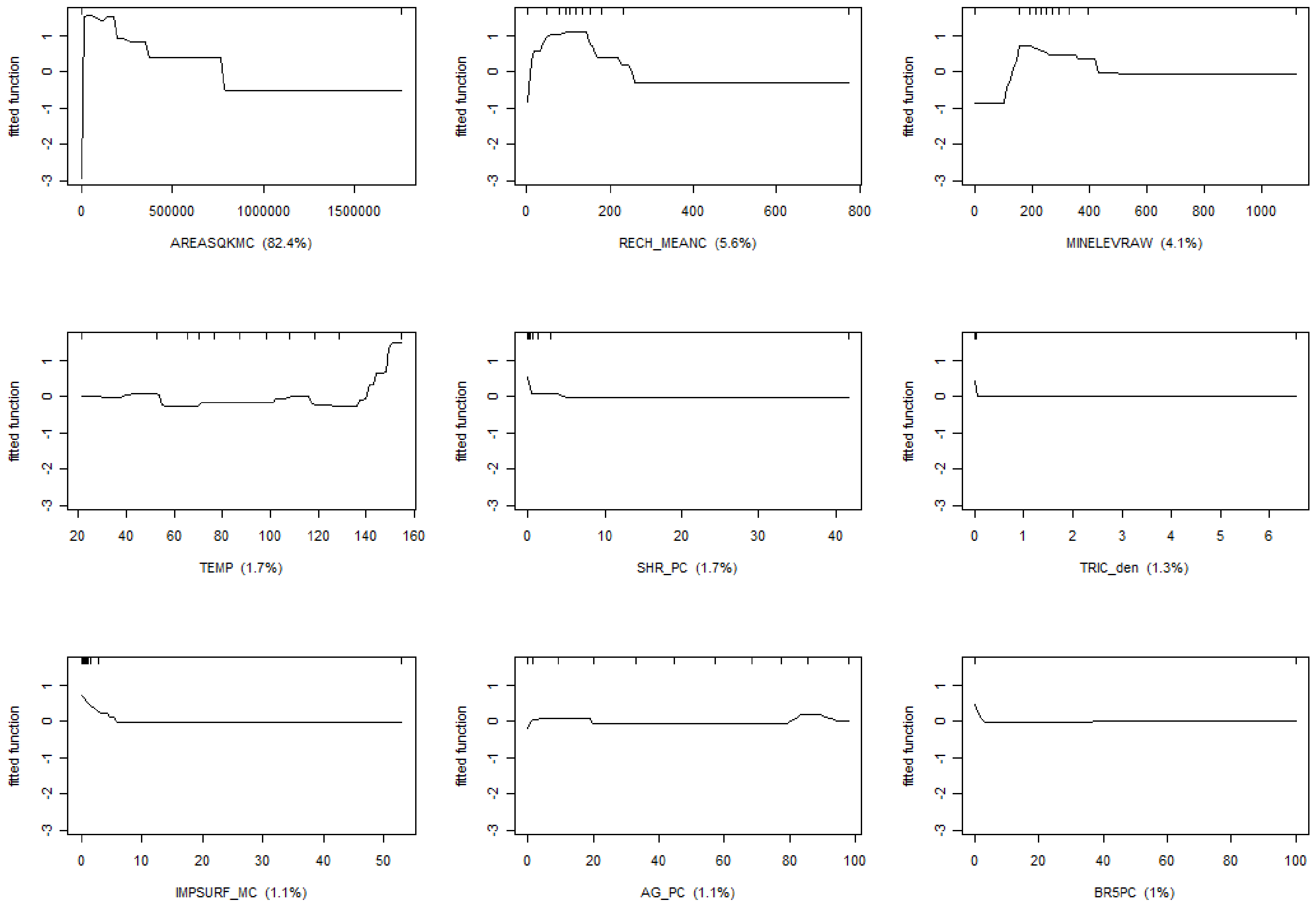
Note: Individual variables are highlighted according to whether they were determined to be anthropogenic (grey shading) or natural (no shading).).

4.2.3 *Variable functions*

The BRT output also contains quantitative information on partial dependence functions that can be plotted to visualize the effect of each individual predictor variable on the response after accounting for all other variables in the model. Similar to the interpretation of traditional regression coefficients, the function plots are not always a perfect representation of the relationship for each variable, particularly if interactions are strong or predictors are strongly correlated. However, they do provide a useful and objective basis for interpretation (Friedman, 2001; Friedman and Meulman, 2003).

These plots show the trend of the response variable (y-axis) as the predictor variable (x-axis) changes. The response variable is transformed (usually to the logit scale) so that the magnitude of trends for each predictor variable's function plot can be accurately compared. The dash marks at the top of each function represent the deciles of the data used to build the model. The function plots for the nine most influential variables in the warmwater species model (Table 5) are illustrated in Figure 28. The plots for all variables are shown in Appendix B.

Figure 28: Functional responses of the dependent variable to individual predictors of warmwater guild



Note: Only the top nine predictors, based on relative influence (shown in parentheses; see Appendix A for descriptions of variable codes), are shown here. See Appendix B for plots of remaining predictor variables.

4.3 Post-modeling

4.3.1 *Stress and natural quality*

The variable importance table and partial dependence functions of the final BRT model were used to assess the potential stressors for the warmwater guild model. Within the model, there were three variables considered anthropogenic in nature (Table 5). After reviewing the functional relationships of these three potential stressors, one of the three stressors were removed from ASI calculations. This variable (AG_PC) had a function plot that was unintuitive: the relationship to the response likely captured some sort of regional variation in the model rather than a mechanistic relationship with the response. The two remaining stressors, network toxic release inventory site density (TRIC_den) and network impervious surface cover (IMPSURF_MC), were used to calculate ASI for the warmwater guild model. Section 1.3.2 details how ASI and HQI were calculated for each model.

4.3.2 *Potential future climate scenario*

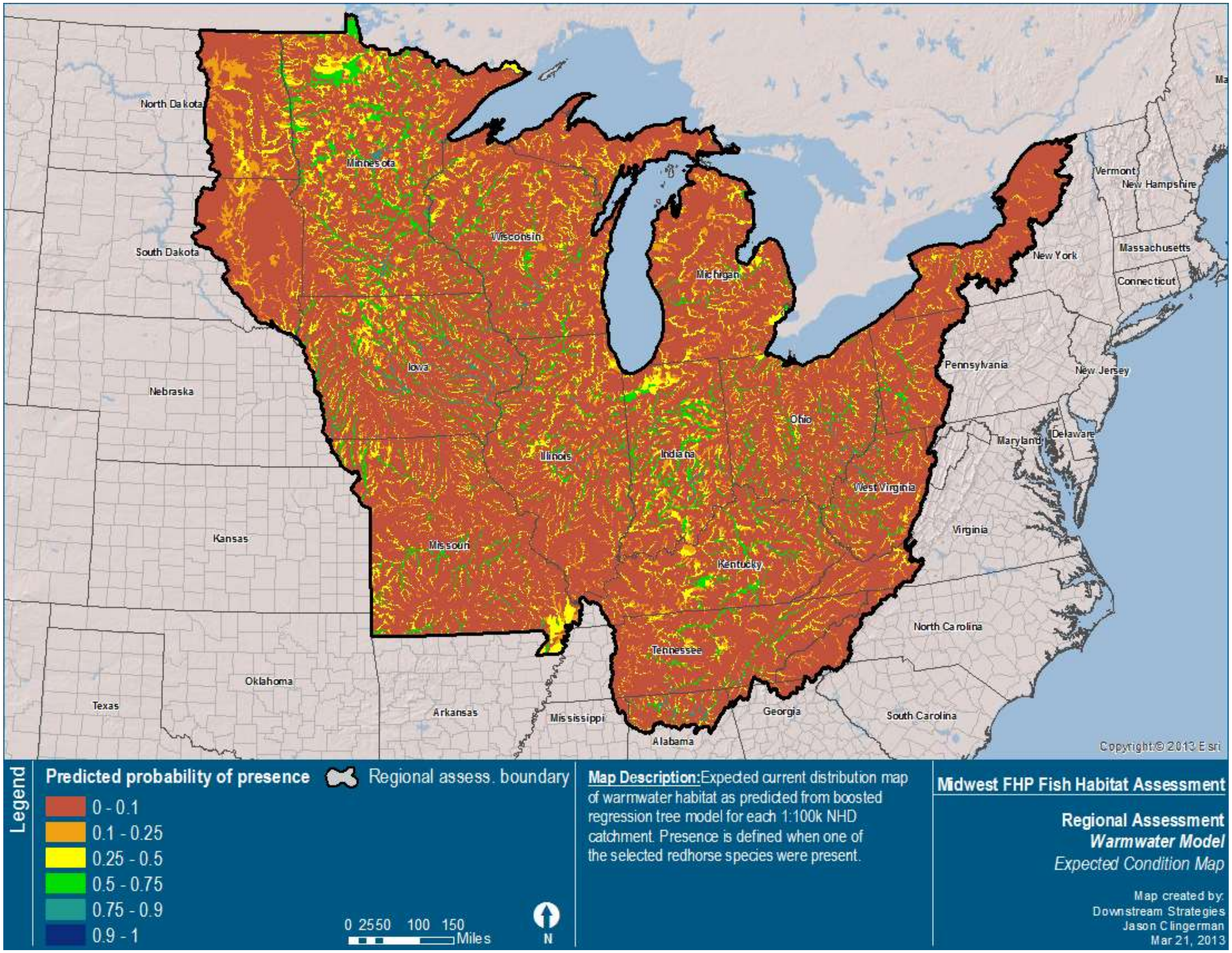
The warmwater BRT model was extrapolated onto a dataset that contained future climate data as described in Section 1.3.3. The potential future predictions were then compared to the current predictions. Percent change in probability was calculated for each catchment to assess climate change vulnerability.

4.4 Mapped Results

4.4.1 *Expected current conditions*

Warmwater species probability of presence was calculated for all 1:100k stream catchments in the study area using the BRT model. The predicted probability values ranged from 0 to 1. The mean predicted probability was 0.088. Of the total 641,615 catchments, about 0.5% (3,274 catchments) had a predicted probability of presence greater than 0.75, and about 4.6% (29,625 catchments) had a predicted probability of presence between 0.5 and 0.75. These results are mapped in Figure 29.

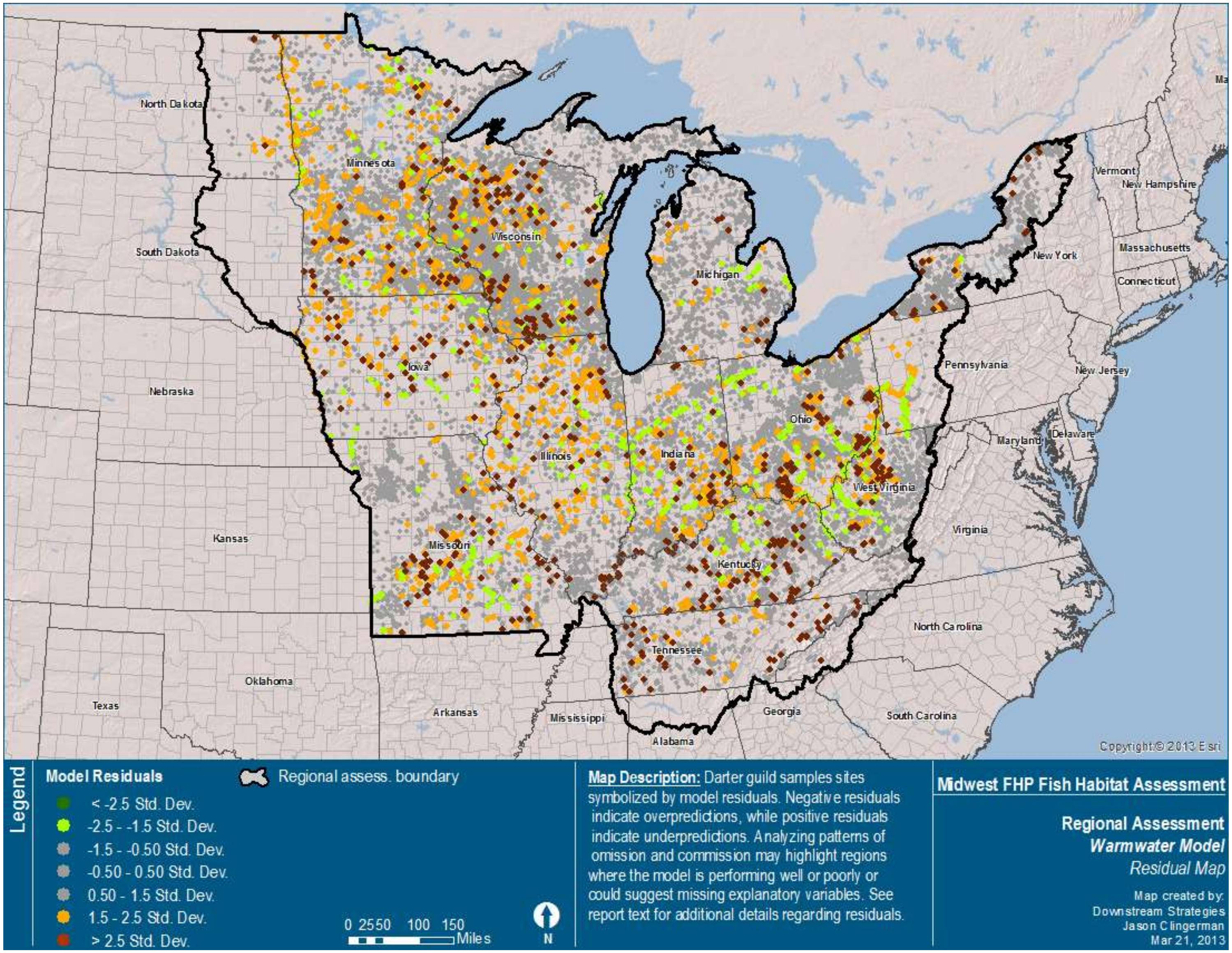
Figure 29: Expected warmwater guild distribution



4.4.2 *Spatial variability in predictive performance*

Analyzing patterns of omission and commission may highlight regions where the model is performing well or poorly or could suggest missing explanatory variables. To assess omission and commission, residuals were calculated by the BRT model. The residuals are a measure of the difference in the measured and modeled values (measured value *minus* modeled value). Negative residuals indicate overpredictions (predicting higher values than are true), while positive residuals indicate underpredictions (predicting lower values than are true). Figure 30 shows the distribution of model residuals per sampling site.

Figure 30: Distribution of warmwater guild model residuals by sampling site



4.4.3 *Indices of stress and natural quality*

Maps of HQI and ASI illustrate the spatial distribution of natural habitat potential (i.e., HQI score) and anthropogenic stress (i.e., ASI score) throughout the Midwest region. HQI and ASI scores are mapped in Figure 31 and Figure 32, respectively. The two variables contributing toward the calculation of ASI are mapped in Figure 33 and Figure 34. HQI, ASI, and their metrics are all scaled on a 0-1 scale (see Sections 1.3.2 and 4.3 for more details on HQI and ASI calculation). For HQI, higher values indicate higher natural quality, while higher values for ASI indicate higher levels of anthropogenic stress.

At first glance, it may seem that regional stress conditions are overly optimistic, but it is necessary to consider that the stress index is showing areas where probability of presence for this response is reduced because of stressors. It is likely that stress on aquatic systems in general is much more widespread than is indicated in any individual model's stress maps. Stress from all three models is considered together in Section 5 of this report. For all stress and natural quality indices, all catchments are shown, even in areas where the probability of presence is low. This is necessary and useful to consider areas outside of the current expect range where stress could have caused a historic population to be extirpated.

Figure 31: Cumulative natural quality index for warmwater guild

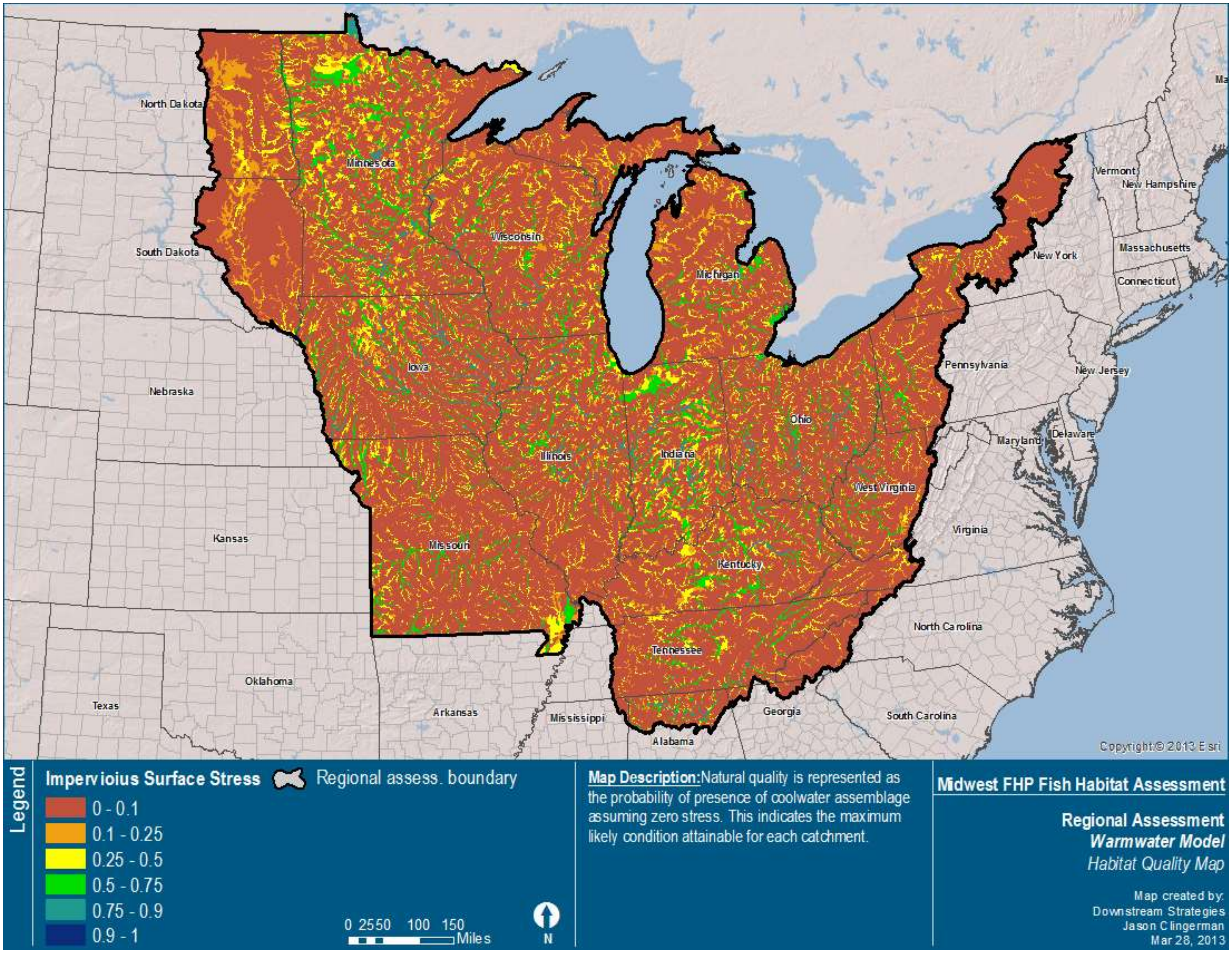


Figure 32: Cumulative anthropogenic stress index for warmwater guild

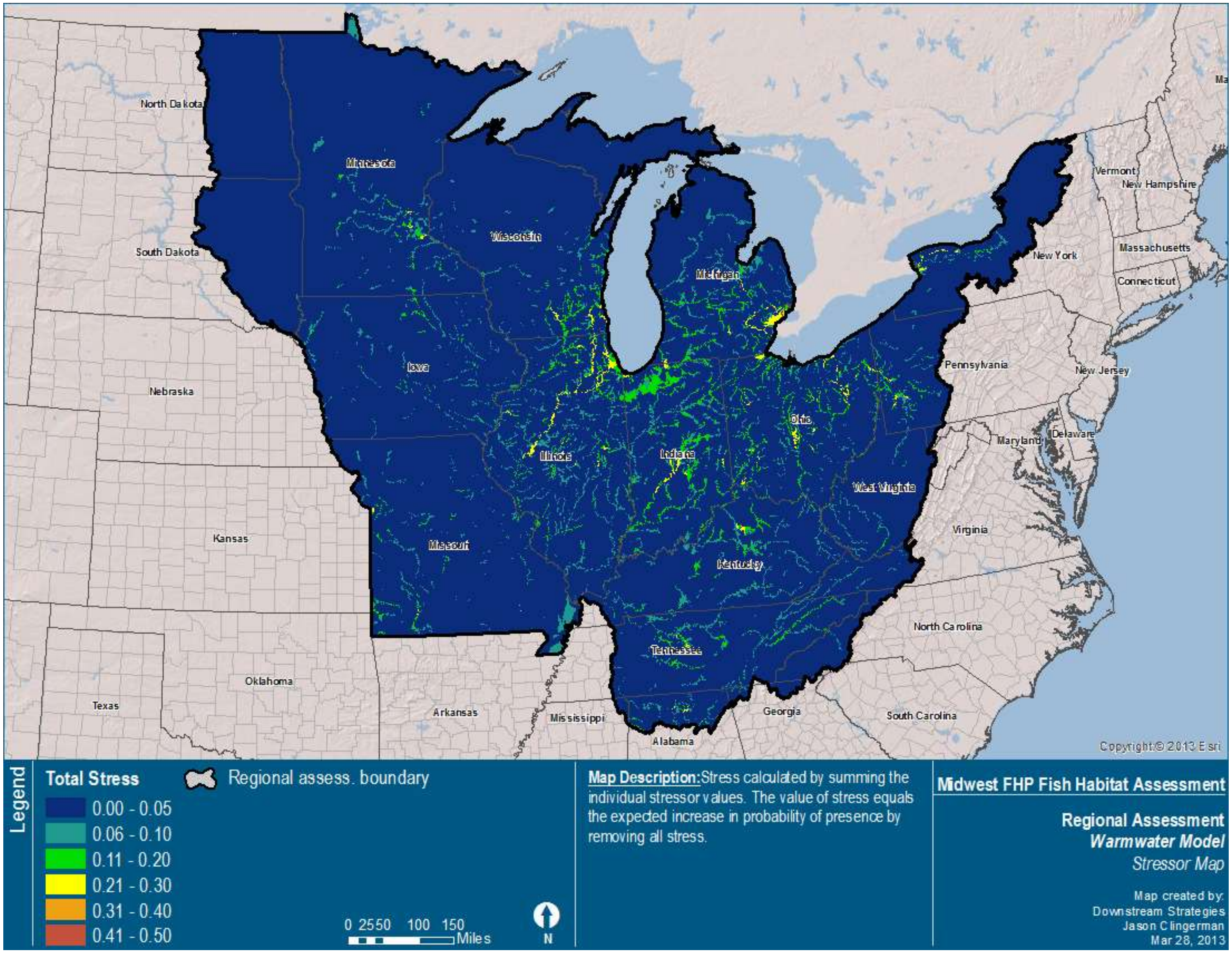


Figure 33: Toxic Release Inventory stressor index metric for warmwater guild

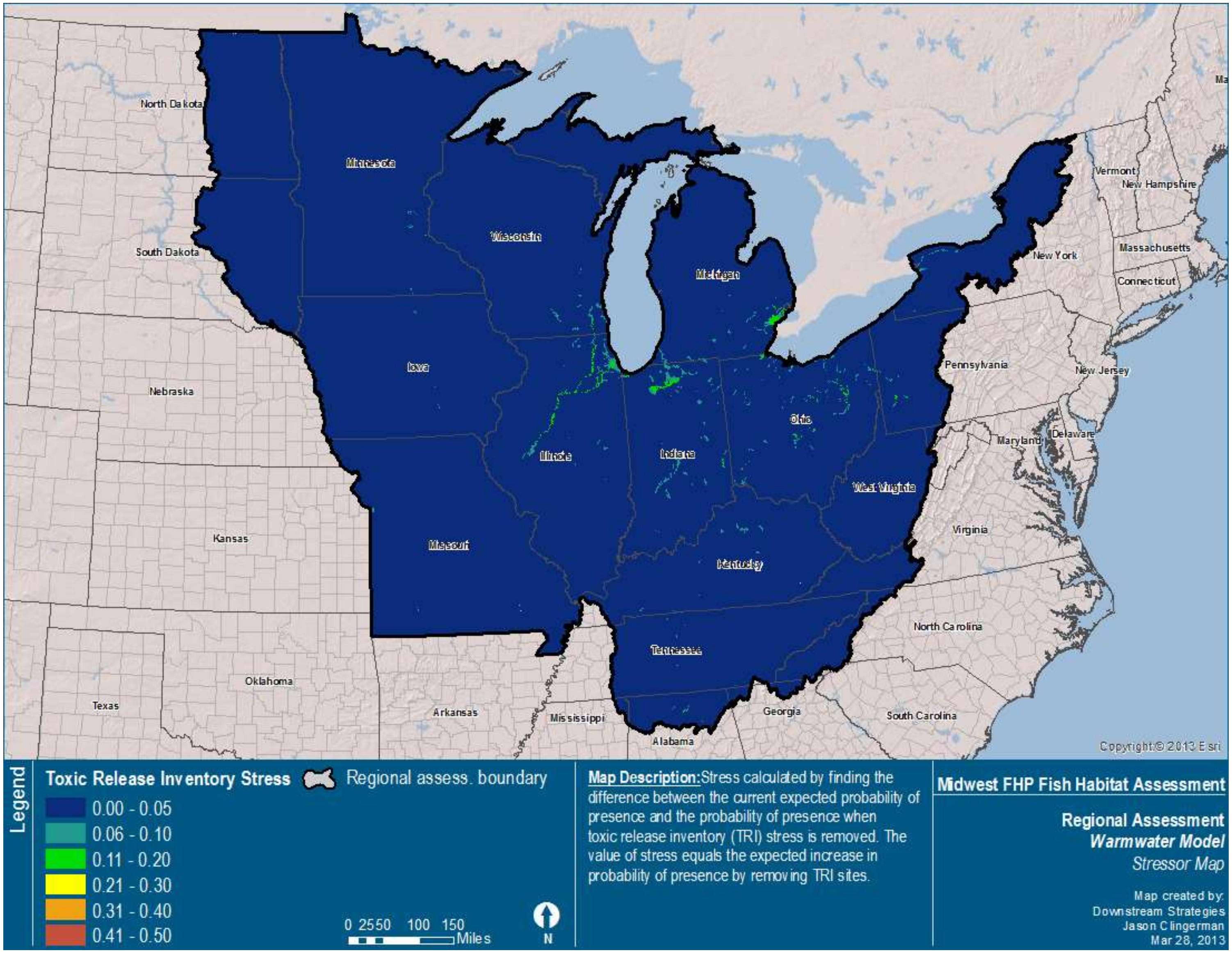
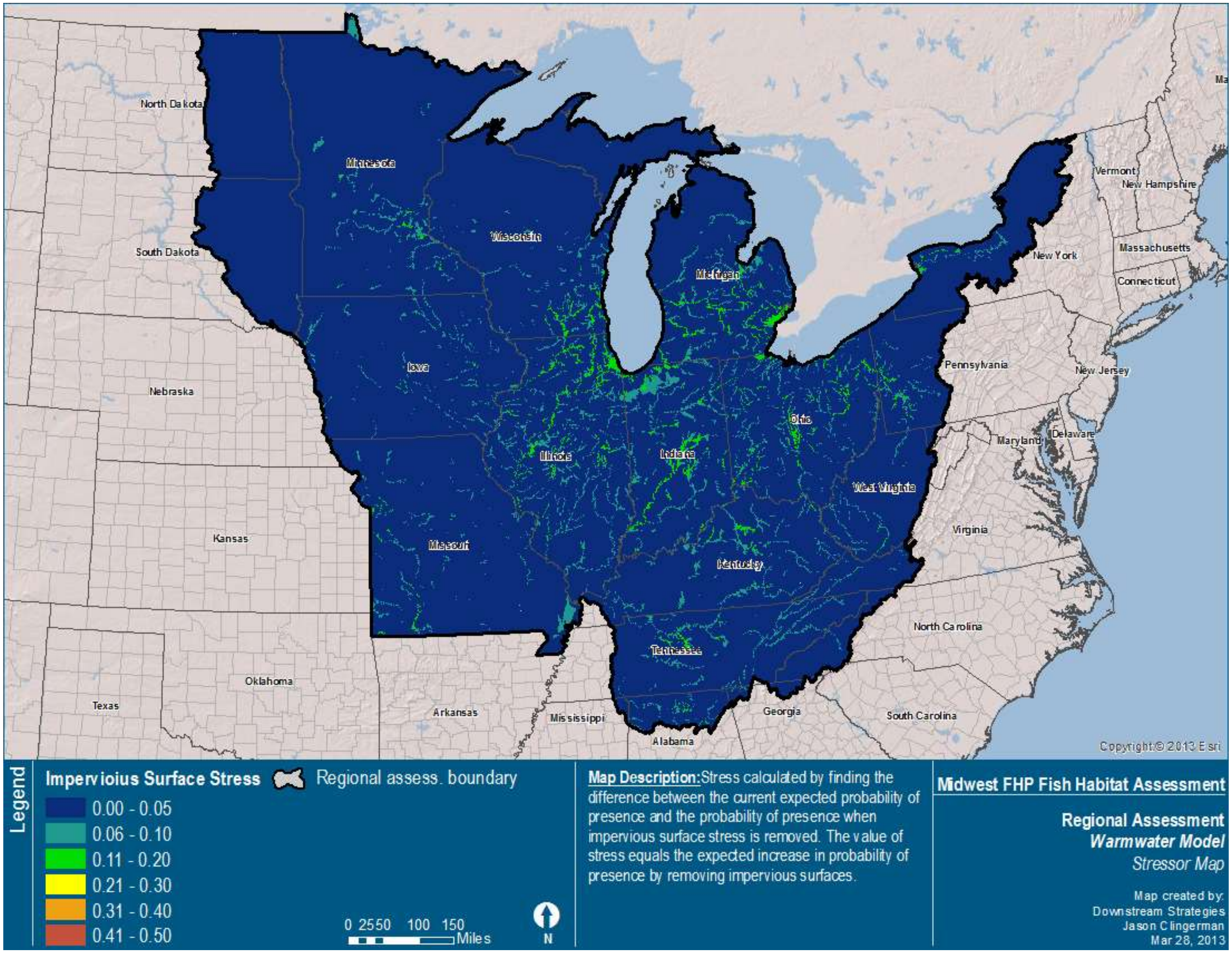


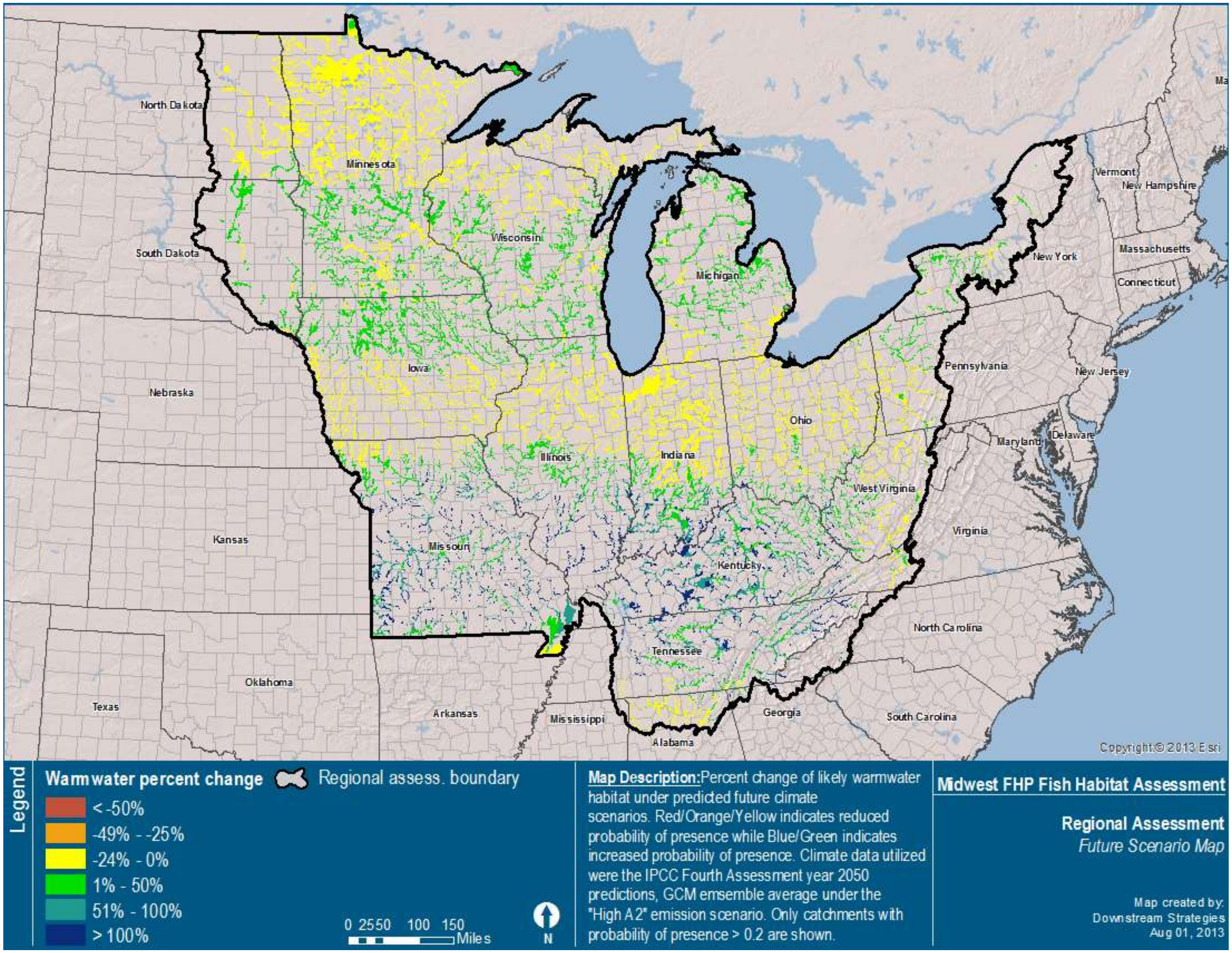
Figure 34: Impervious surface stressor metric for warmwater guild



4.4.4 *Potential future climate scenarios*

A map of percent change in warmwater habitat probability of presence based on a 2050 IPCC A2 future climate change scenario (see Section 1.3.3 for further explanation) is shown in Figure 35. The percent change values should be interpreted as a measure of how susceptible each catchment may be to climate change. Positive percent change indicates the probability is expected to increase under the future climate scenario, while a negative percent change indicates a lower probability to be expected under the 2050 climate scenario. To ensure that only habitats that are likely to contain warmwater fish guilds; only catchments where the current probability of presence is greater than 0.20 are shown in this figure. This cutoff level was selected after visualizing the data to ensure that it was effective at removing areas not likely to contain warmwater habitat while still adequately portraying the potential effect of future climate scenarios upon the expected warmwater habitat.

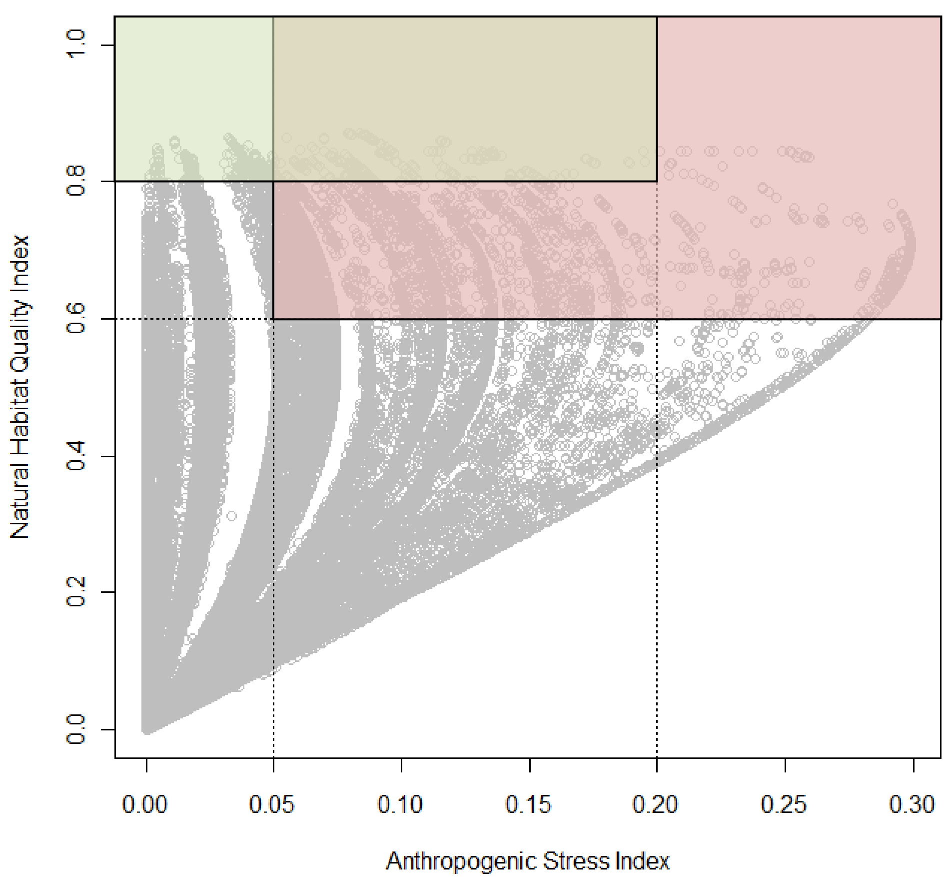
Figure 35: Potential climate change scenario for warmwater habitat



4.4.5 *Restoration and protection priorities*

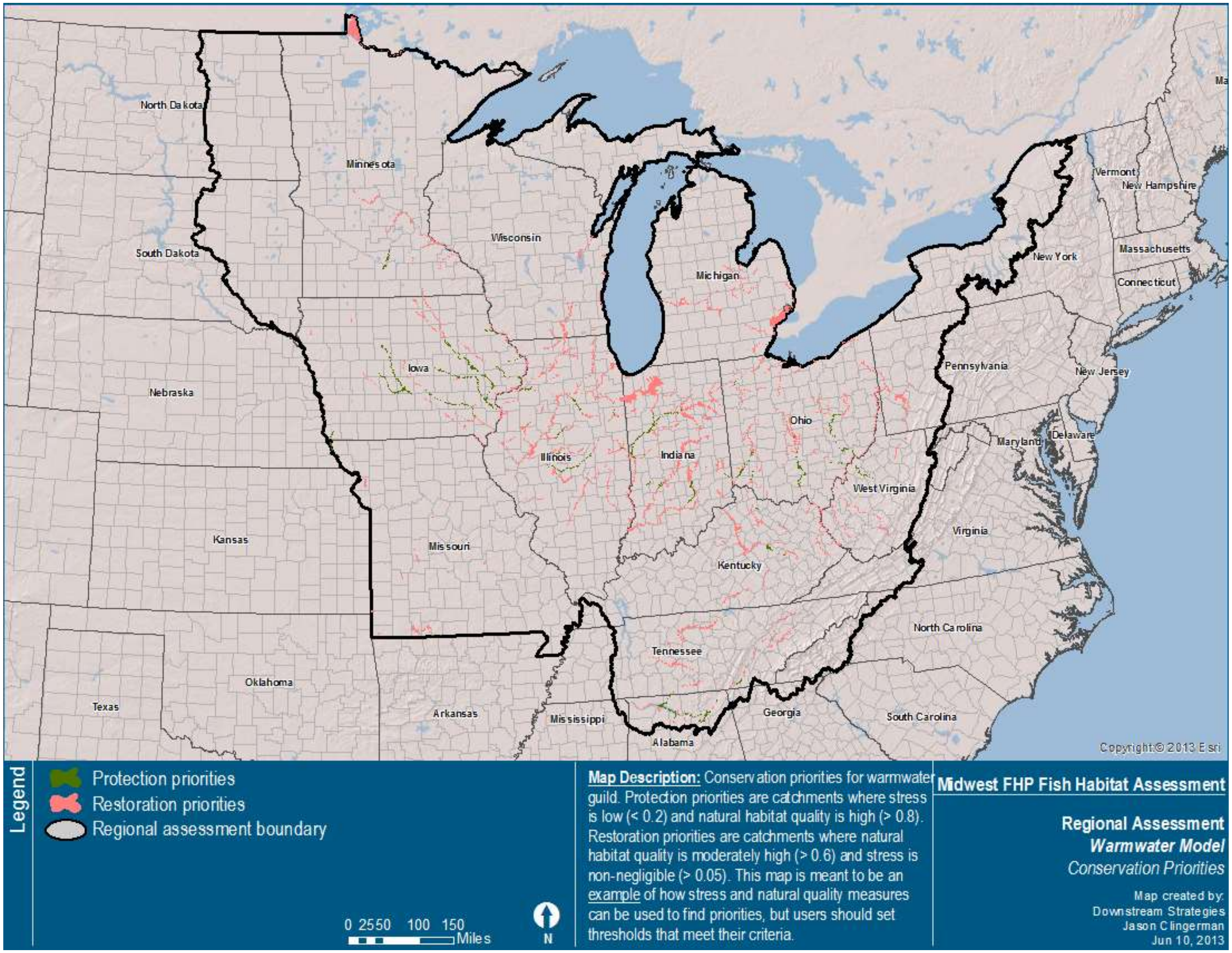
A plot of HQI versus ASI values for all catchments in the study area can be used as a reference to define HQI and ASI thresholds when evaluating restoration and protection priorities (Figure 36). In the example shown (Figure 37), thresholds for protection priorities were defined as catchments with high natural habitat quality and low anthropogenic stress; these thresholds were based on HQI greater than 0.8 and ASI less than 0.2. The thresholds used to identify restoration priorities were defined as catchments with high natural habitat quality and moderate to high anthropogenic stress; these thresholds were based on HQI greater than 0.6 and ASI greater than 0.05. Due to the methodology used to set these thresholds, there is potential for certain catchments to be classified as both restoration and protection priorities (Figure 36), in these cases protection priority overrides restoration when mapped in Figure 37. These thresholds were solely based on the relative scores for natural quality and stress indices. Though this example scenario provides an informed set of criteria for identifying conservation priorities, it is only intended to demonstrate the functionality of querying catchments based on these attributes to identify areas that meet user-defined criteria to guide conservation, protection, and restoration planning.

Figure 36: HQI versus ASI values for all catchments for warmwater species



Note: The red box indicates catchments defined as restoration priorities under the example scenario. The green box indicates catchments defined as protection priorities under the same scenario.

Figure 37: Restoration and protection priorities for warmwater guild



5. COMBINED ANALYSIS

5.1 Introduction

Managers often require aggregated measures of habitat stress, quality, and long-term vulnerability based on multiple input models to identify regional conservation strategies that meet the needs of multiple species and habitat types. Until this point, all modeling performed had been done for a key species or species group, and for a specific region. The necessity for a more comprehensive, region-wide assessment of aquatic health remained.

For broad-scale assessments of overall aquatic health, biologists and managers often utilize Indices of Biotic Integrity (IBI), which is an index calculated from either fish or macroinvertebrate sample data. IBIs work well at summarizing complex biological community data into a single index that relays generalized health and quality of a habitat, but for the Midwest Regional Assessment, there were no existing IBIs that could be applied to the entire study area as they are routinely constructed as much smaller scales; for instance, at the state level.

Since there was not an existing single index we could model for the entire region, we chose to model three complementary fish community responses, and then analyze them in aggregate to produce a single region-wide assessment of overall aquatic health. Region-wide indices of generalized habitat condition (stress, natural quality, and climate vulnerability) were created by combining the modeled outputs for each of the three species guilds described in Sections 2-4. This information will aid managers making broad-scale conservation decisions across the region.

5.2 Methodology

5.2.1 *Species selection*

The model assessments detailed in sections 2-4 created indices of stress and natural quality for each specific response variable. These three responses were chosen to represent distinct, but complimentary guilds of species that inhabit the majority of stream habitats within the Midwest. Species (or species groups) chosen had to meet several criteria. The first criterion was that they needed to be found across the majority of the study area, in order to exclude any results driven by localized presence or absence of species with limited ranges. Another requirement was that species or species groups needed to be sensitive to chemical or physical habitat degradation. These types of species, sometimes called “intolerant”, are more useful when assessing aquatic stress than species that are more tolerant to disturbances. The exact species included in each of the three models are described above in sections 2-4.

5.2.2 *Determining current stream type*

The current stream type was determined by identifying the single model (i.e. cold-, cool-, or warmwater) that produced the highest probability of presence. This classification provides information about the most likely species composition given the three models utilized. This information can allow managers to understand the generalized distribution of the three species groups across the region.

5.2.3 *Determining optimal stream type*

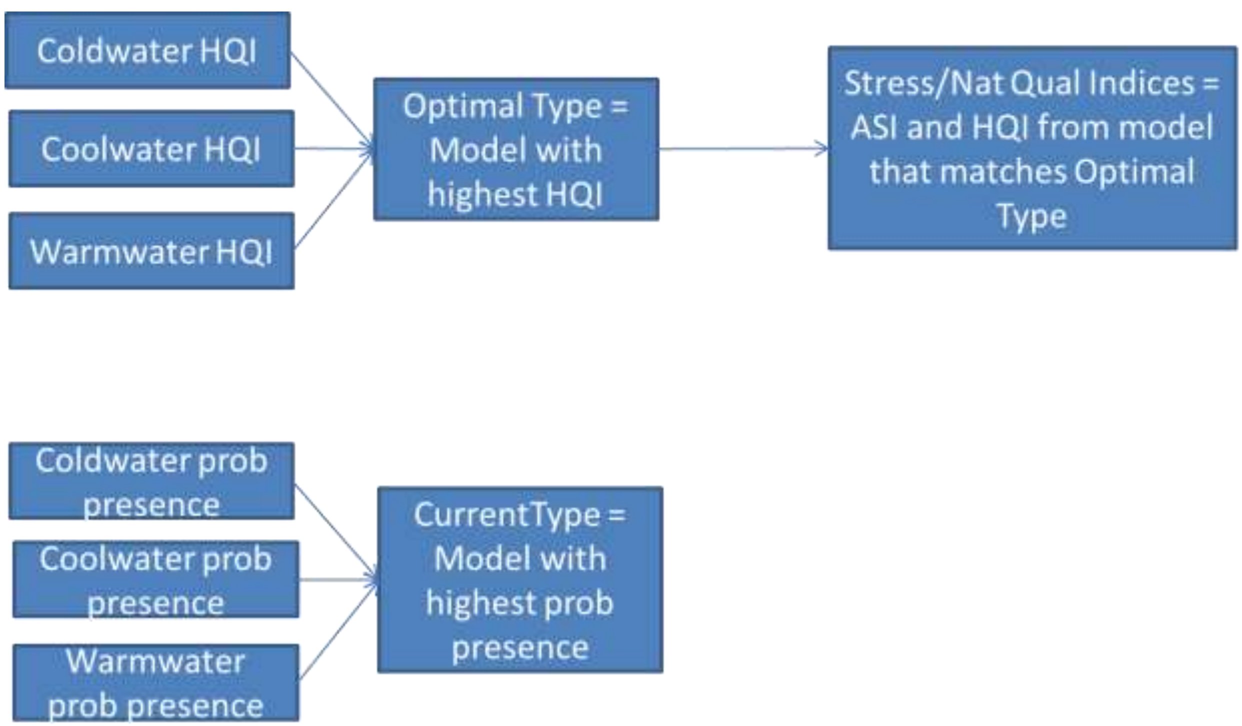
We then determined which of the models produced the highest HQI score for each catchment. By determining the highest HQI score for each catchment, we were able to identify which of the three models (species guilds) that each catchment would most likely contain under optimal, no-stress habitat conditions. Essentially this provided a rule-based method for classifying each catchment as an ideal stream “type” – coldwater, coolwater, or warmwater. Since type was determined by analyzing the HQI score (i.e. the optimal habitat condition under no stress), this type indicates which of the three species guilds would most likely inhabit each catchment under idealized conditions, or the best possible conditions attainable through removal of stress.

It is likely that many catchments would have the same optimal and current type, especially if stress is low. Other catchments might have a current type that differs from the optimal type. This would indicate current stress is driving the current type away from the optimal type of fish community. For instance, a catchment could optimally classify as coldwater, but, due to current stressors (e.g. high agriculture reducing stream shading), could be classified as a current coolwater catchment. By understanding the optimal and current stream types, managers can identify proper areas for specific restoration actions. In application, a manager attempting to restore coldwater fish communities might search for catchments where the optimal type is coldwater, but the current type is coolwater. Before this analysis, only information on current and optimal condition of a single model was available. This analysis combined data from three complementary models to allow for cross-model comparisons.

5.2.4 *Selecting stress and natural quality indices*

Stress and natural quality indices are not always applicable across all catchments when viewed irrespective of the optimal type, or natural condition, of the stream at a given catchment. For example, a warmwater model stressor index should not be applied to small, high elevation headwater catchments. In order to assign each catchment appropriate index values among the three sets of models available (i.e. from the cold, cool, and warmwater models), we used a set criterion to apply one model's results to each catchment based on the optimal stream type. Further, rather than utilizing the mean values, we used a rule-based criterion to avoid indicating stress on a catchment when that model's index did not match the optimal stream type, based on the inherent natural condition of the stream. Figure 38 details this rule-based approach.

Figure 38: Flowchart outlining stress and natural quality calculations for combined analysis



We used the optimal type (described above) to determine what stress and natural quality indices to select for each catchment. This allowed us to produce region-wide indices of stress and natural quality, by applying one ASI score and one HQI score to each catchment based the optimal stream type for that catchment. For example, if a single catchment had the following HQI scores: coldwater = 0.2, coolwater = 0.4, and warmwater = 0.8, the highest score (i.e., warmwater in this example) was selected as the catchment type, and only the indices associated with the warmwater model were applied to the regional analysis for that specific catchment.

The region-wide assessments of stress and natural quality allow managers to understand both current and optimal habitat conditions within each catchment, along with the overall stress and natural quality for the three species guilds modeled in the assessment. These regionally consistent measures of stress and quality can further inform restoration and conservation management approaches at varying scales.

5.2.5 *Combined climate vulnerability*

In an effort to understand regional scale impacts from potential climate change scenarios, we analyzed climate change impacts at the catchment scale. We compared the current conditions within each catchment, which was determined by the current stream type identified for each catchment, to the future climate scenario for that stream type. This resulted in one measure of climate change vulnerability for each catchment, and that value was taken from the model which corresponded with the current stream type.. For each catchment we then utilized the current probability of presence and the predicted probability of presence under future climate scenarios—both of which were derived from the model associated with current type--to calculate a percent change in probability of presence. The percentage of change from current conditions to the 2050 High A2 climate change scenario (see Section 1.3.3 for more information on the specific climate change scenario) was used as the climate vulnerability metric for each catchment. This index illustrates the general vulnerability of potential climate change on aquatic habitats across the region.

Potential transitions in likely habitat type under this climate change scenario were also calculated. This was determined by selecting the current model response with the highest probability of presence (*current* type), and then selecting the future model response with the highest probability of presence under the future climate scenario (most likely *future* type). The transitions between habitat types were then calculated.

5.3 Results and discussion

5.3.1 *Current and optimal stream type*

Using the methodology described above we were able to determine the most likely current stream type for each of the 641,615 catchments (Figure 39). This designation of current stream type is based only upon the three modeled endpoints, and does not consider other assemblages not represented. Results should be interpreted as the most likely of the three modeled responses to be present at each catchment.

We also determined the optimal stream type by assessing the natural quality scores within each catchment (Figure 40). This designation was also based on the three responses modeled, and should be interpreted as the most likely of the three responses to be present within each catchment under no-stress situations.

Mapped results indicate that the most drastic difference between optimal and current conditions occurs throughout the Cornbelt, where agricultural stress is likely contributing to stream warming and causing more warmwater species to be present rather than coolwater fish communities.

5.3.2 *Stress and natural quality indices*

This procedure produced one index for natural quality (Figure 41) and one index for region-wide stress (Figure 42). It is important to remember that while these indices and maps provide a general indication of overall quality and stress on aquatic habitats, they were built based on the three specific assemblages that were identified by fishery professionals as indicative of high quality habitats across the range of this study. Areas of the study area where none of the three assemblages modeled are naturally distributed may not classify well. The most obvious example of this is in the small streams across much of the low-elevation or southern latitude areas. It is plausible that for these areas, none of the assemblages chosen would naturally inhabit these very small, warm streams.

Areas in Wisconsin, Michigan, New York, Pennsylvania, West Virginia, Virginia, North Carolina and Tennessee have the highest natural quality values for the coldwater model (Figure 41). Since the coldwater model focuses on smaller, headwater catchments, the areas with strong coldwater habitats are the easiest to recognize visually. Alternatively, cool and warmwater responses occur in larger streams, and can be seen throughout much of the rest of the Midwest as thin lines of high quality habitat surrounded by catchments with lower natural quality scores.

Stress (Figure 42) is only driven by agriculture, impervious surfaces, and toxic release inventory in this analysis (See sections 2.3.1, 3.3.1, and 4.3.1). In the map, the majority of stress is occurring across the Cornbelt, and is most likely a result of the agriculturally dominated landscape, but the effect of impervious surfaces can be seen in other areas as well, such as Chicago, Minneapolis-St. Paul, and Pittsburgh, as well as along some major highways throughout the study area.

Furthermore, stress is only indicated in areas where one of these stressors decreases the probability of presence of the model indicated for each catchment. For instance, some catchments may have high agriculture, but if it is not contributing to a reduction in the probability of presence for the model selected for that catchment, stress may not be indicated at that location. Therefore, areas identified as lower stress should be viewed as areas where stress was not assigned to any of the three responses chosen for this assessment, but not necessarily as areas definitively free from all human-induced aquatic stress.

5.3.3 *Combined climate vulnerability*

The combined climate vulnerability analysis produced results that indicated overall percent change in probability of presence (Figure 43) and the predicted habitat transitions (Figure 44). The climate change scenarios utilized for this analysis indicated an increase in both mean annual precipitation and mean annual temperature (see section 1.3.3). From the model function plots (Figure 6, Figure 17, Figure 28) we see that generally increased precipitation results in increases in the coldwater predicted presence, while increases in temperature would have varying effects depending on the model and exact temperature ranges of a particular area. As expected, the coldwater response is most likely to occur at the lowest temperatures, coolwater response is most likely at moderate to high temperatures, and warmwater responses are most likely at higher temperatures.

From the map of overall percent change in probability of presence, we can see that much of the northern half of the study area would be expected to experience a decrease in probability of current assemblages under the given 2050 climate scenario. Much of the southern half of the study area shows potential increases in probability of presence under the 2050 climate scenario utilized for this analysis. The decreases in the northern half (and in places along the Appalachian Mountains), are likely due to the increase in temperature, while the increases across the study area are likely due to the predicted increase in precipitation. This map is useful for assessing the general trends in potential climate change vulnerability for the three responses modeled.

Figure 44 shows the catchments expected to transition from one habitat type to another may be more useful for localized assessments. The transition types are mapped for all catchments where probability of presence is greater than 0.2. This cutoff level was selected after visualizing the data to ensure that it was effective at removing areas not likely to contain any of the three habitat types, while still adequately portraying the potential effect of future climate scenarios on the expected cold, cool, and warm-water assemblages modeled. Warming of stream assemblages are shown in red to yellow gradients (red, yellow, orange), while cooling of stream assemblages are shown in shades of blue. There are a few noticeable trends we see from this analysis. One is that the coldwater habitat predicted to transition to coolwater habitat occurs mostly through Wisconsin and Michigan, with some occurances in New York and Pennsylvania. These transitions are likely driven by the increase in annual temperature. In the northern half of the study area, there are also many areas currently predicted as warmwater which are predicted to transition to coolwater habitats. In the southern half of study area, the dominant predicted transitions are cool to warm and cold to warm. Much of the area expected to transition from cold to warm occurs across an area where coldwater habitats are generally rare. This could be a result of the inability of the models to appropriately classify the habitat in these areas, but also indicates that these areas may be the most at risk of losing already uncommon species guilds. The transitions from cool to warm are likely being predicted as a result of the predicted increase in annual temperature.

All results here are based on three region-wide models and should only be used to assess general trends in expected climate vulnerability for these specific responses, and should be combined with localized knowledge of threats and conditions before implementing specific local conservation actions. Ideally, similar analyses at finer spatial scales would provide more locally-relevant results and further inform this regional assessment of climate change vulnerability.

Figure 39: Current stream habitat type

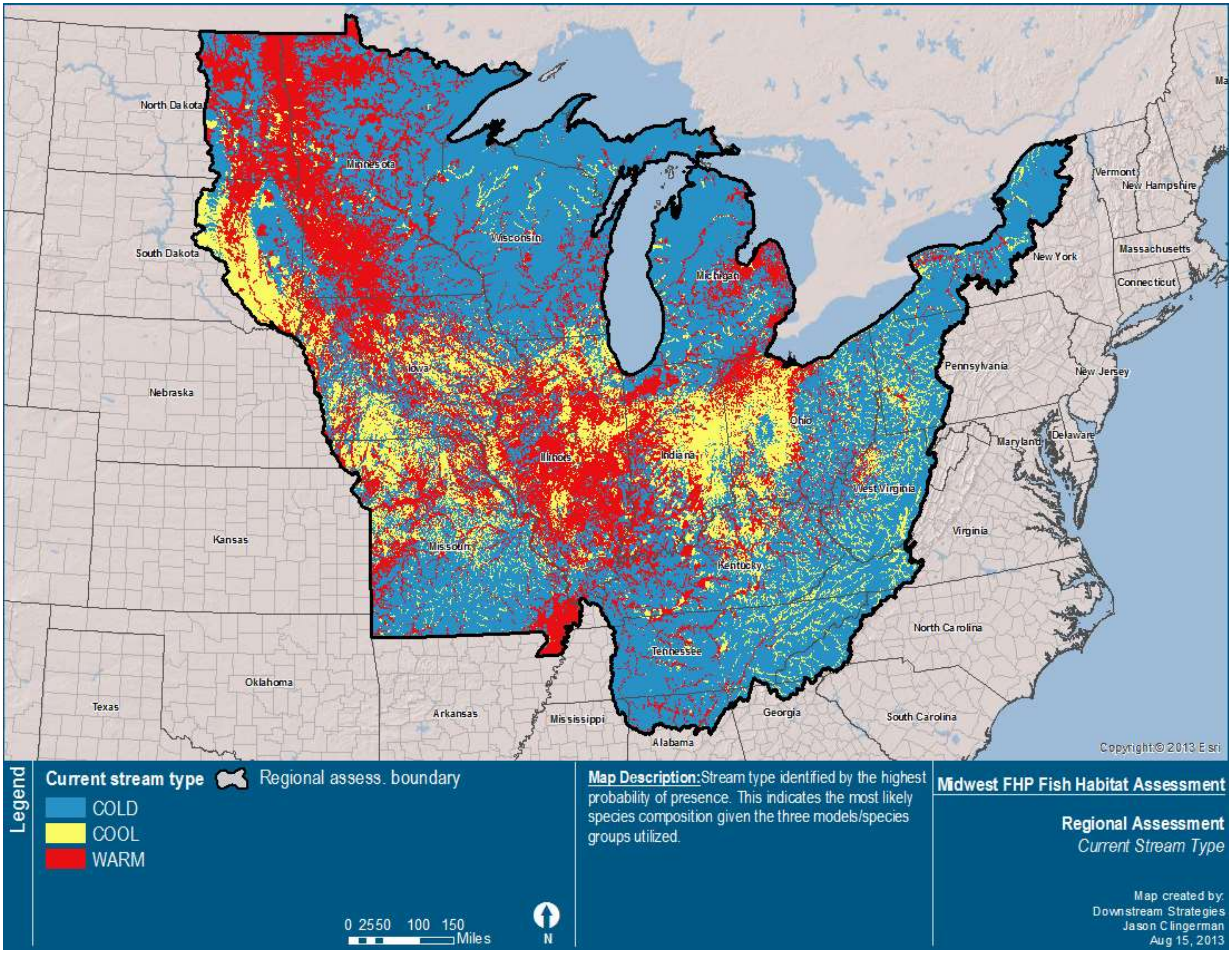


Figure 40: Optimal stream habitat type

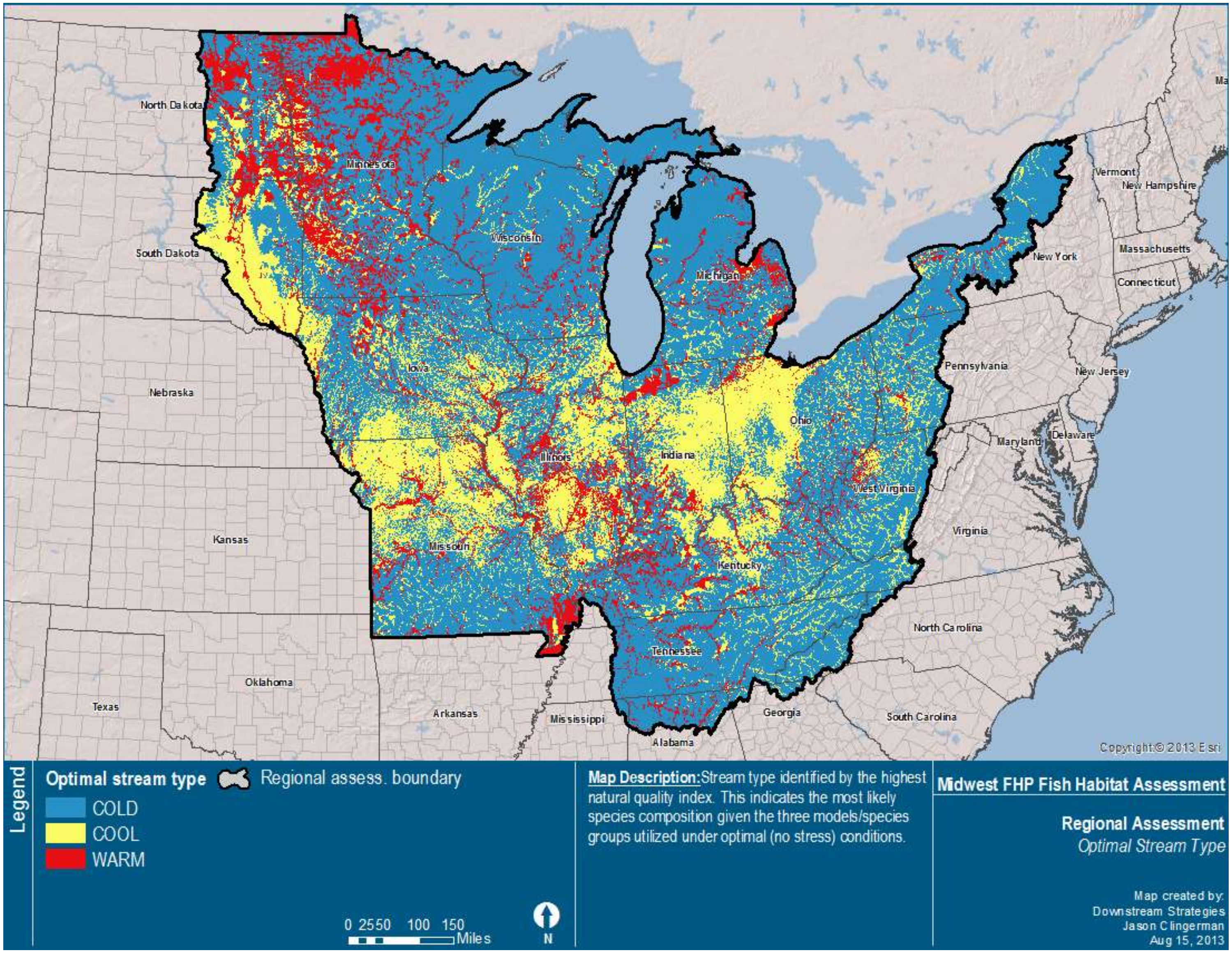


Figure 41: Region-wide aquatic natural quality index

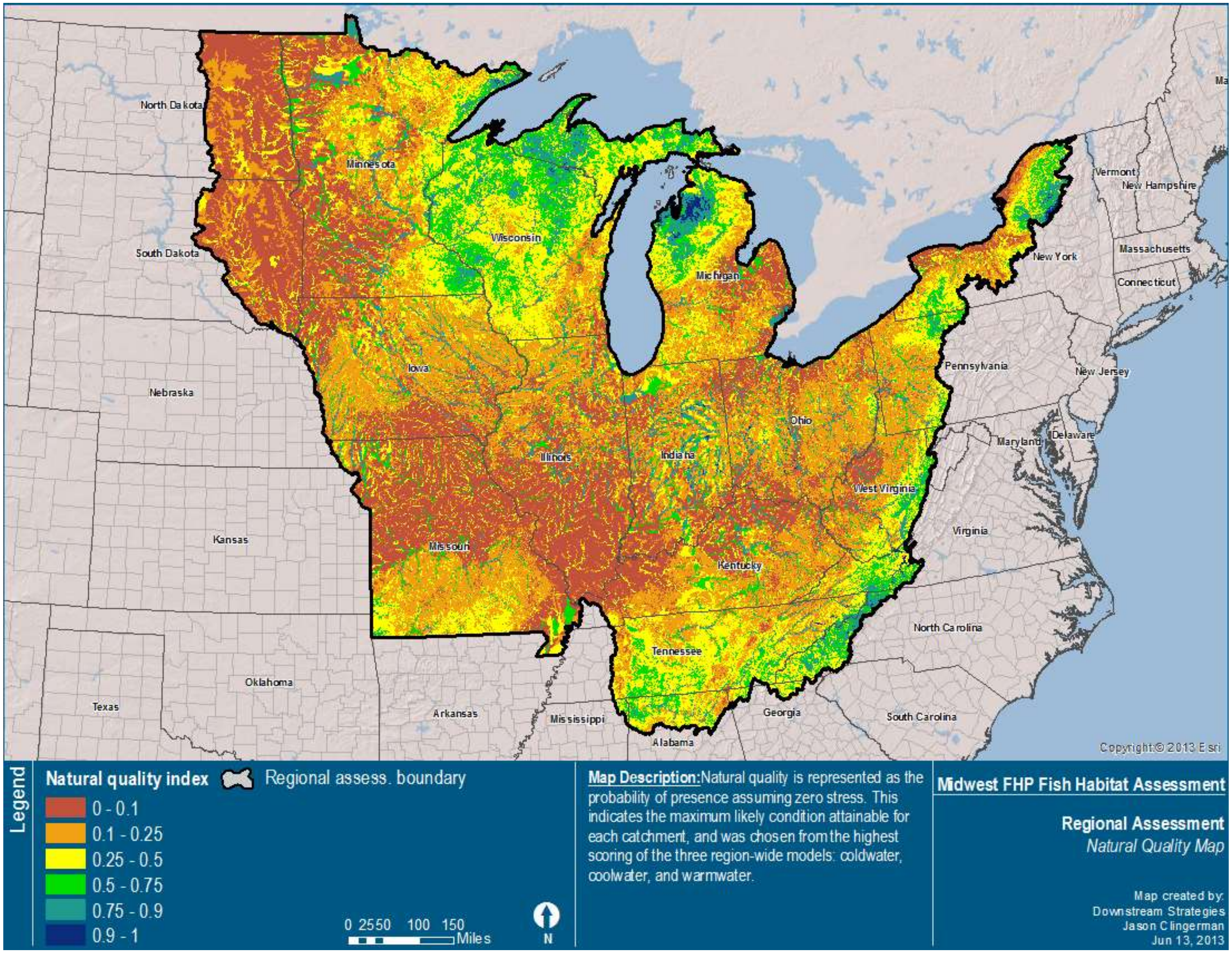


Figure 42: Region-wide aquatic stress index

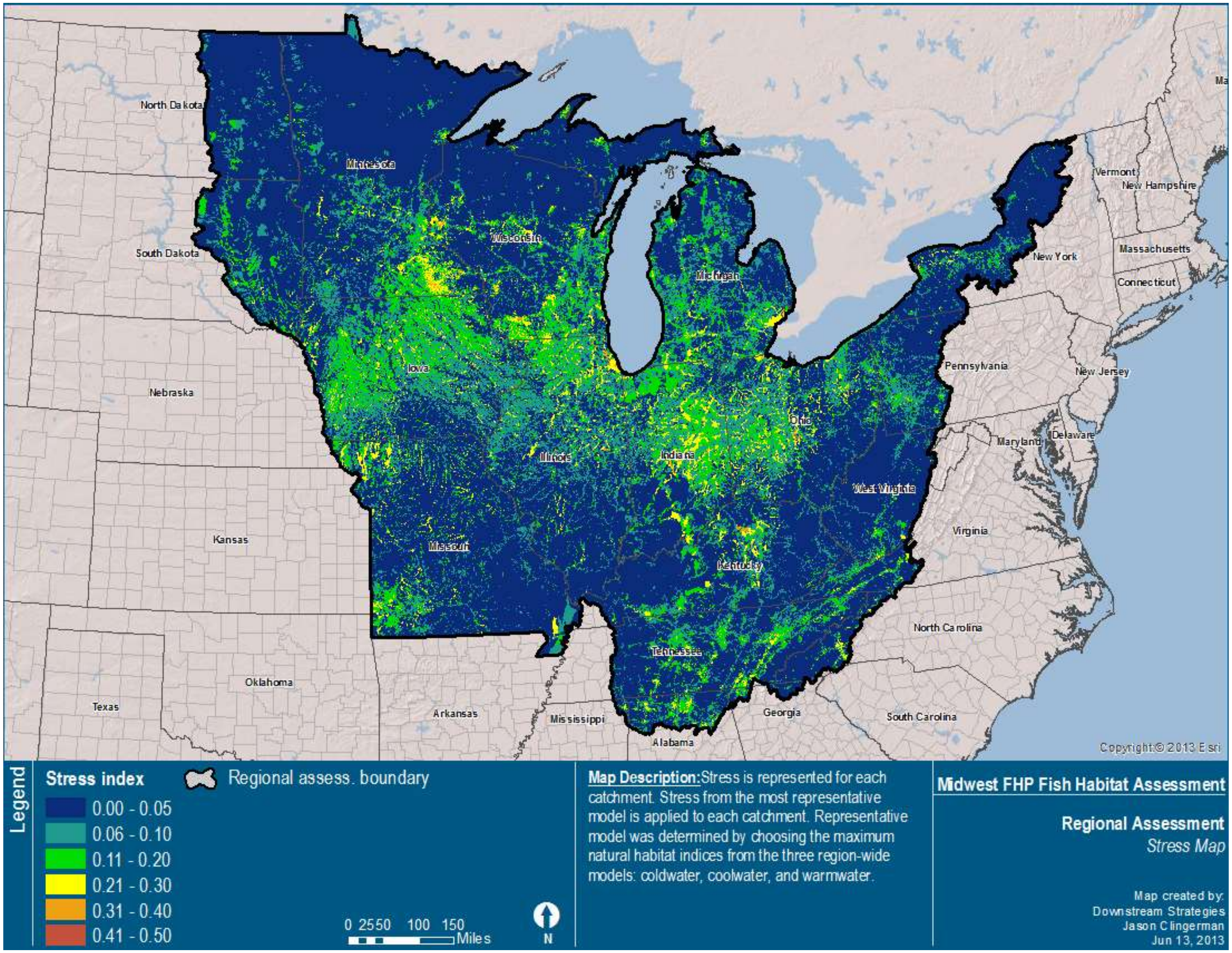


Figure 43: Combined climate change vulnerability

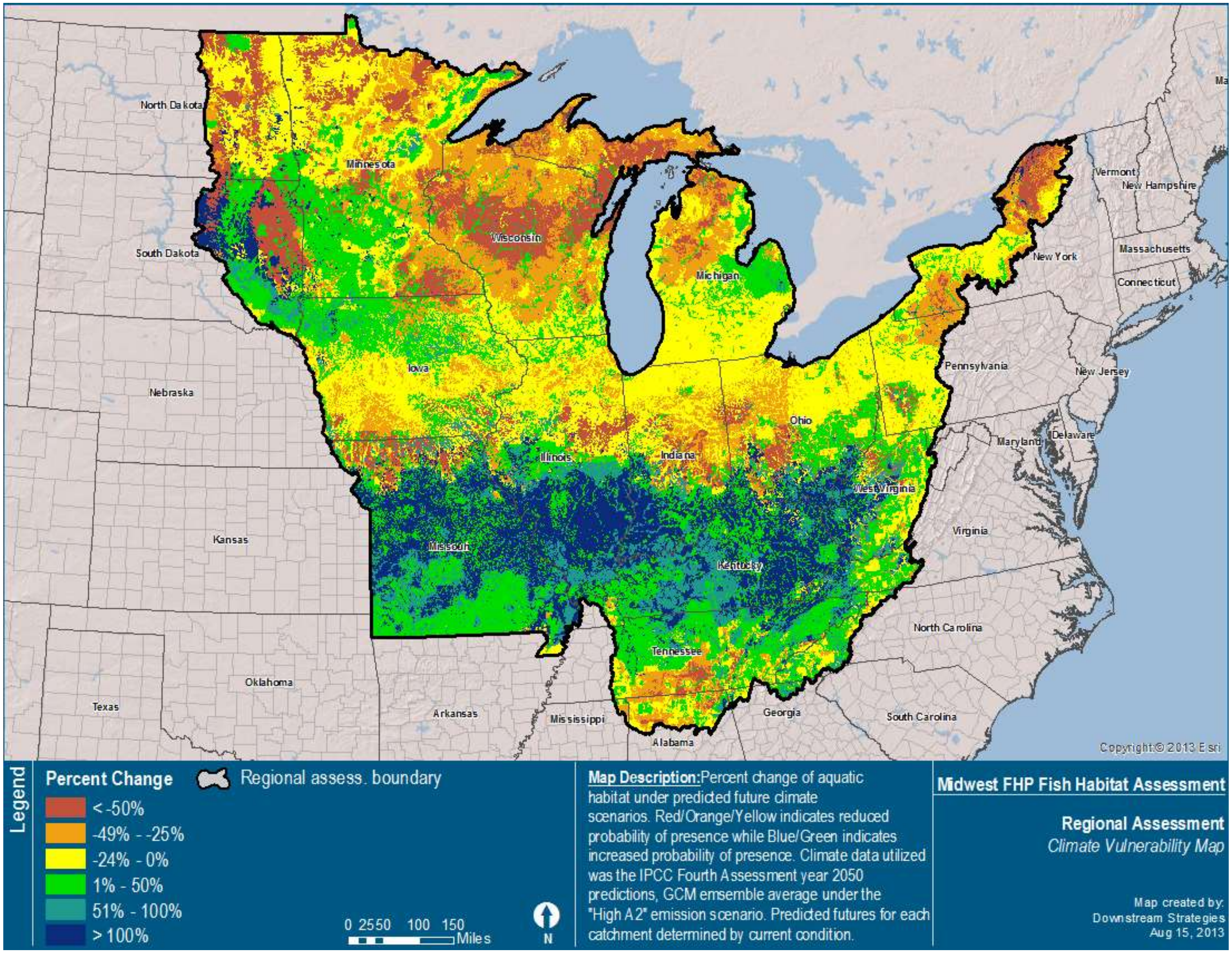
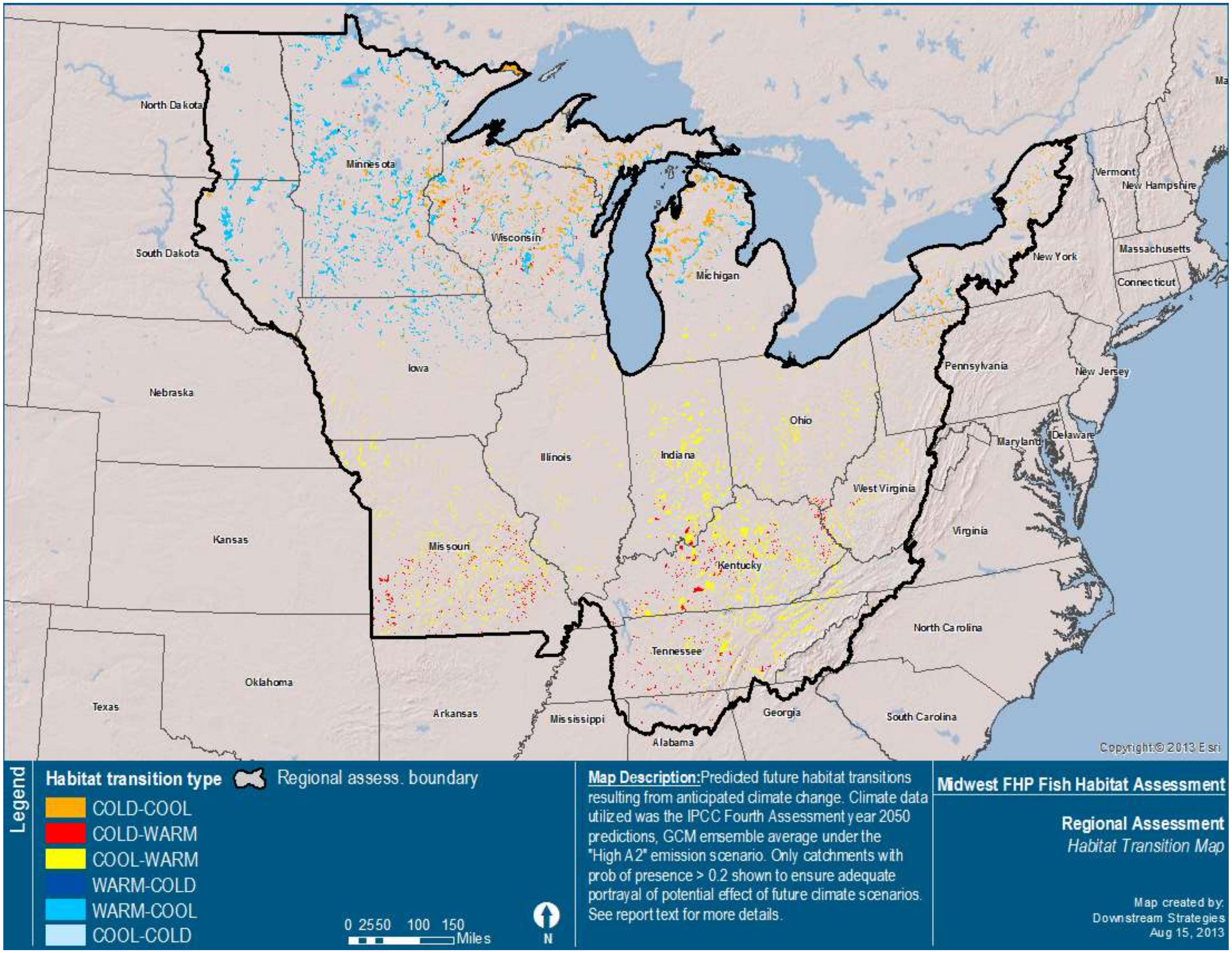


Figure 44: Predicted habitat transitions from potential climate change



6. LIMITATIONS AND SUGGESTIONS FOR FUTURE WORK

In general, while the estimates of probability of presence, index scores, HQI, and ASI generated through this assessment represent a useful and objective means for assessing aquatic habitat and prioritizing habitats for restoration or protection, there are some limitations that are important to consider. Results generated through the modeling process are ultimately limited by the quality of data used to generate them. In the future, the model can be improved by improving the resolution and precision of the data. For example, some county-level data were used as predictor variables although the data likely generalize conditions at the catchment scale. Although these variables—such as network surface water consumption—were limited in spatial resolution, they still had high relative influence in the BRT model and were important to retain for predictive performance. In the future, refinement of these county-level variables or inclusion of higher resolution surrogates could improve both the precision of the BRT model predictions and post-modeling indices.

A second limitation is that the data and maps represent only a snapshot in time. Therefore, the models may not represent conditions before or after the data were collected or created. For example, any habitat lost or gained due to increased impervious surface cover since the 2006 National Land Cover Database (NLCD) was not considered in this assessment. The amount of such changes would likely be minimal, and at the regional scale of study analyzed here would not likely influence the model results significantly. Similarly, a portion of the uncertainty can be attributable to the temporal mismatches between the fish collection data and landscape data, although we used the best available matching data to avoid this as much as possible, though improving the temporal match between those datasets for future work would be beneficial.

These broad-scale models offered valuable insight into which landscape-level stressors and natural conditions were structuring aquatic responses across the region. However, they do not assess conditions at a scale that is appropriate for the determination of more local-level conditions, as the broad patterns overshadowed variables that may structure aquatic responses at finer scales. Recent modeling efforts at the regional and FHP scale have indicated that smaller-scale models are likely necessary to pinpoint localized stressors. From discussions with experienced modelers and fishery professionals, HUC8 watersheds were agreed upon as the most appropriate scale. Please see the DS report "Analysis of scale on boosted regression tree fish habitat models" for a case study of how scale influences the importance of stressor variables.

There were also a few important issues that were beyond the scope of this project. Acid precipitation, biological interactions, and local habitat variation are all important in structuring fish communities. These variables were not directly used as predictor variables, although, when possible, surrogates were used to approximate variation in the model resulting from these processes.

Local habitat measures such as water quality (pH, alkalinity, instream temperature), physical habitat complexity, and substrate size are examples of local measures important to structuring fish communities. These measures could not be directly quantified in this analysis given the scope and scale of the project. However, since each catchment's land cover and geology was included in the analysis, some aspects of water quality were indirectly modeled. Likewise, habitat complexity and substrate size could be partially captured by the combination of stream slope and bedrock and surficial geology. Nonetheless, exclusion of detailed local measures likely accounts for some uncertainty in the model results. Thus, the results from this analysis should be combined with local expert knowledge and additional field data to arrive at the most accurate representation of habitat conditions.

In addition, inclusion of biological interactions in future models could improve the precision of the model and the ability to quantify its influence on the response variables. Specifically, important biological interactions in this system could include the negative interactions resulting from the introduction of non-native or other stocked fishes, such as brown trout or Asian carp.

There are inherent limitations to the climate vulnerability analyses. All estimates of climate vulnerability used downscaled global climate models of predicted temperature and precipitation changes in the 2050 decade based on the IPCC (SOURCE). Global climate models are based on various assumptions of future economic development, energy use, policies, population growth, geophysical responses, and more. As with any predicted values, there is uncertainty and variability associated with the predicted climate data used in this study. The climate change model used in this study is intended to be interpreted as one example of potential future changes based on the best available science. When updated assessments are completed those values should be incorporated to ensure the most up to date information is utilized.

REFERENCES

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. Journal of Animal Ecology 77: 802-13.

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Annals of Statistics 29: 1189-1232.

Friedman JH, Meulman JJ (2003) Multiple additive regression trees with application in epidemiology. Statistics in Medicine 22: 1365-81.

IPCC. 2007. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Core Writing Team, Pachauri, R.K. and Reisinger, A. (Eds.) IPCC, Geneva, Switzerland. pp 104.

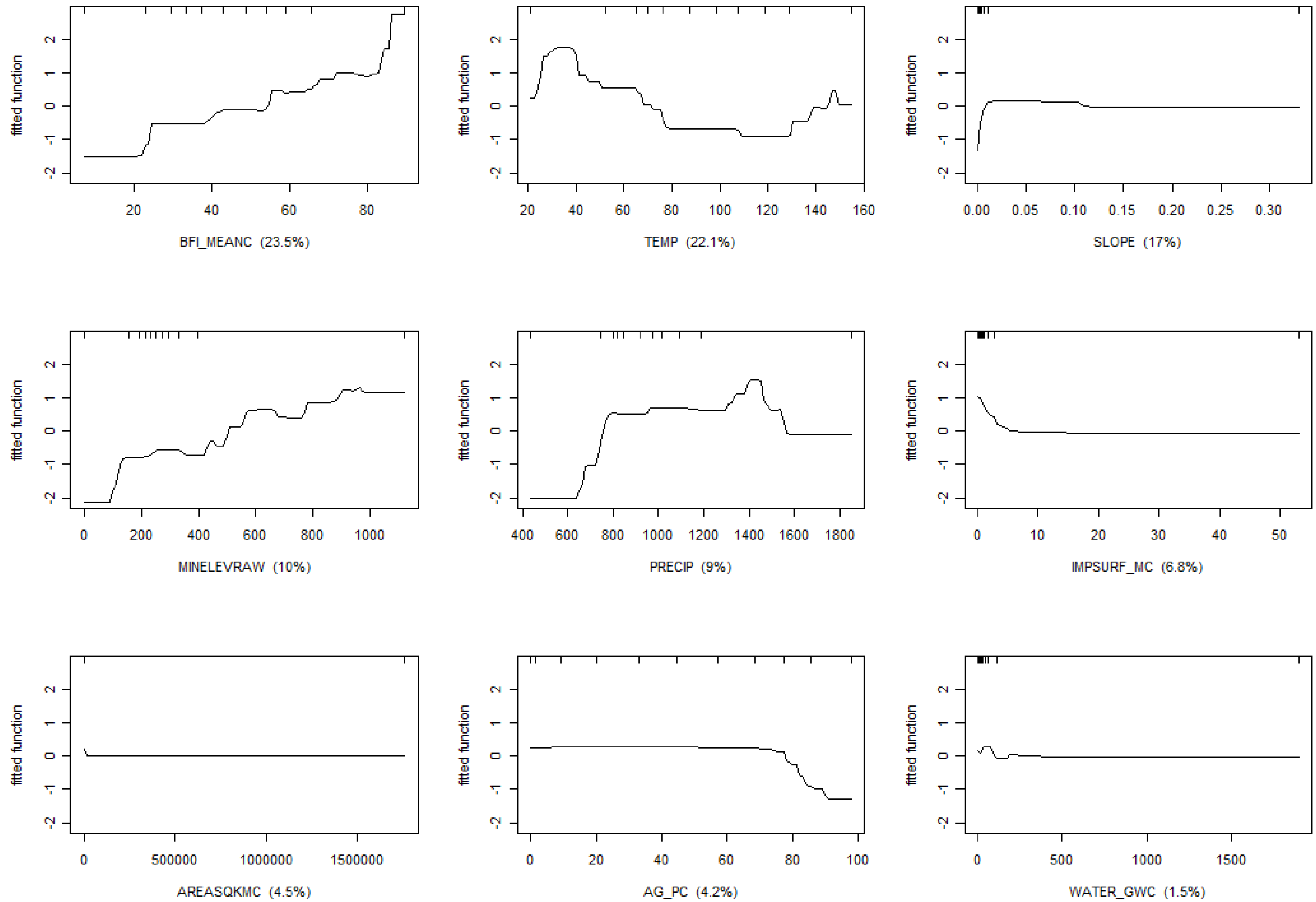
Appendix A: DATA DICTIONARY

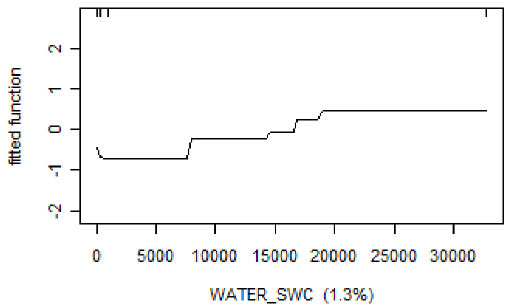
Field	Description	Source
Comid	catchment comid (unique identifier)	NHDPlus
Areasqkm	area of catchment, sq km	NHDPlus
Cumdrainag	Cumulative drainage area in square kilometers	NHDPlus
Minelevraw	Minimum elevation (unsmoothed) in meters	NHDPlus
Slope	Slope of flowline (cm/cm)	NHDPlus
Precip	Mean annual precipitation in mm	NHDPlus
Temp	Mean annual temperature in degrees centigrade * 10	NHDPlus
Impsurf_M	mean percent impervious, catchment	NLCD 2006
Impsurf_MC	mean percent impervious, cumulative	NLCD 2006
BR1P	Catchment bedrock geology, percent Carbonate (LOCAL)	USGS (Reclassified by Letsinger)
BR2P	Catchment bedrock geology, percent Felsic (igneous) (LOCAL)	USGS (Reclassified by Letsinger)
BR3P	Catchment bedrock geology, percent Mafic (igneous) (LOCAL)	USGS (Reclassified by Letsinger)
BR4P	Catchment bedrock geology, percent Metamorphic (LOCAL)	USGS (Reclassified by Letsinger)
BR5P	Catchment bedrock geology, percent Sand and gravel (LOCAL)	USGS (Reclassified by Letsinger)
BR6P	Catchment bedrock geology, percent Sandstone (LOCAL)	USGS (Reclassified by Letsinger)
BR7P	Catchment bedrock geology, percent Shale (LOCAL)	USGS (Reclassified by Letsinger)
BR8P	Catchment bedrock geology, percent Unconsolidated (LOCAL)	USGS (Reclassified by Letsinger)
BR1PC	Network bedrock geology, percent Carbonate (CUMULATIVE)	USGS (Reclassified by Letsinger)
BR2PC	Network bedrock geology, percent Felsic (igneous) (CUMULATIVE)	USGS (Reclassified by Letsinger)
BR3PC	Network bedrock geology, percent Mafic (igneous) (CUMULATIVE)	USGS (Reclassified by Letsinger)
BR4PC	Network bedrock geology, percent Metamorphic (CUMULATIVE)	USGS (Reclassified by Letsinger)
BR5PC	Network bedrock geology, percent Sand and gravel (CUMULATIVE)	USGS (Reclassified by Letsinger)
BR6PC	Network bedrock geology, percent Sandstone (CUMULATIVE)	USGS (Reclassified by Letsinger)
BR7PC	Network bedrock geology, percent Shale (CUMULATIVE)	USGS (Reclassified by Letsinger)
BR8PC	Network bedrock geology, percent Unconsolidated (CUMULATIVE)	USGS (Reclassified by Letsinger)
DEV_P	NLCD 2006, % of developed land cover classes (0 to 100), (NLCD classes 22, 23, 24)	NLCD 2006
AG_P	NLCD 2006, % of agricultural land cover classes (0 to 100), (NLCD classes 81, 82)	NLCD 2006
BAR_P	NLCD 2006, % of barren land cover classes (0 to 100), (NLCD class 31)	NLCD 2006
FOR_P	NLCD 2006, % of forest land cover classes (0 to 100), (NLCD classes 41,42,43)	NLCD 2006

WET_P	NLCD 2006, % of wetland cover classes (0 to 100), (NLCD classes 90,95)	NLCD 2006
GRS_P	NLCD 2006, % of grassland cover classes (0 to 100), (NLCD classes 71)	NLCD 2006
SHR_P	NLCD 2006, % of shrub/scrub cover classes (0 to 100), (NLCD classes 52)	NLCD 2006
DEV_PC	Network NLCD 2006, % of developed land cover classes (0 to 100), (NLCD classes 22, 23, 24)	NLCD 2006
AG_PC	Network NLCD 2006, % of developed land cover classes (0 to 100), (NLCD classes 81, 82)	NLCD 2006
FOR_PC	Network NLCD 2006, % of developed land cover classes (0 to 100), (NLCD classes 41,42,43)	NLCD 2006
WET_PC	Network NLCD 2006, % of developed land cover classes (0 to 100), (NLCD classes 90,95)	NLCD 2006
GRS_PC	Network NLCD 2006, % of grassland cover classes (0 to 100), (NLCD class 71)	NLCD 2006
SHR_PC	Network NLCD 2006, % of shrub/scrub cover classes (0 to 100), (NLCD class 52)	NLCD 2006
Rechg	recharge, total mean (mm/year) (catchment)	USGS
BFI_mean	Mean baseflow index (catchment)	USGS
Rechgc	recharge, total mean (mm/year) (Network)	USGS
BFI_meanc	Mean baseflow index (network)	USGS
Water_gw	LOCAL: USGS National Atlas of the US: Ground Water Use by COUNTY 2000: Millions gallons per day/km2	NFHAP
Water_sw	LOCAL: USGS National Atlas of the US: Surface Water Use by COUNTY 2000: Millions gallons per day/km2	NFHAP
Cattle	LOCAL: Agricultural Census 2002, 1:2M scale, INTEGER: average number of cattle/acre farmland	NFHAP
Popdens	LOCAL: US Population Density 2000, NOAA, scale 1km, #/km2	NFHAP
Roadcr	LOCAL: Census 2000 TIGER Roads, 1:100K scale, road crossings identified by INTERSECT, with points generated, #/km2	NFHAP
Roadlen	LOCAL: Census 2000 TIGER Roads, 1:100K scale, units not given - m/km2	NFHAP
Dams	LOCAL: National Inventory of Dams, 2002-2004, #/km2	NFHAP
Mines	LOCAL: USGS Active Mines and Mineral Processing Plants, 2003, #/km2	NFHAP
Tri	LOCAL: USEPA, 2007: #/km2 Toxics Release Inventory Program sites	NFHAP
Npdes	LOCAL: USEPA, 2007: #/km2 National Pollutant Discharge Elimination System sites	NFHAP
Cerc	LOCAL: USEPA, 2007: #/km2 Compensation and Liability Information System sites	NFHAP
Water_gwc	NETWORK: USGS National Atlas of the US: Ground Water Use by COUNTY 2000: Millions gallons per day/km2	NFHAP
Water_swc	NETWORK: USGS National Atlas of the US: Surface Water Use by COUNTY 2000: Millions gallons per day/km2	NFHAP
Cattlec	NETWORK: Agricultural Census 2002, 1:2M scale, INTEGER: average number of cattle/acre farmland	NFHAP
Popdensc	NETWORK: US Population Density 2000, NOAA, scale 1km, #/km2	NFHAP
Roadcrc	NETWORK: Census 2000 TIGER Roads, 1:100K scale, road crossings identified by INTERSECT, #/km2	NFHAP
Roadlenc	NETWORK: Census 2000 TIGER Roads, 1:100K scale, units not given - m/km2	NFHAP
Damsc	NETWORK: National Inventory of Dams, 2002-2004, #/km2	NFHAP
Minesc	NETWORK: USGS Active Mines and Mineral Processing Plants, 2003, #/km2	NFHAP
Tric	NETWORK: USEPA, 2007: #/km2 Toxics Release Inventory Program sites	NFHAP
Npdesc	NETWORK: USEPA, 2007: #/km2 National Pollutant Discharge Elimination System sites	NFHAP
Cercc	NETWORK: USEPA, 2007: #/km2 Compensation and Liability Information System sites	NFHAP

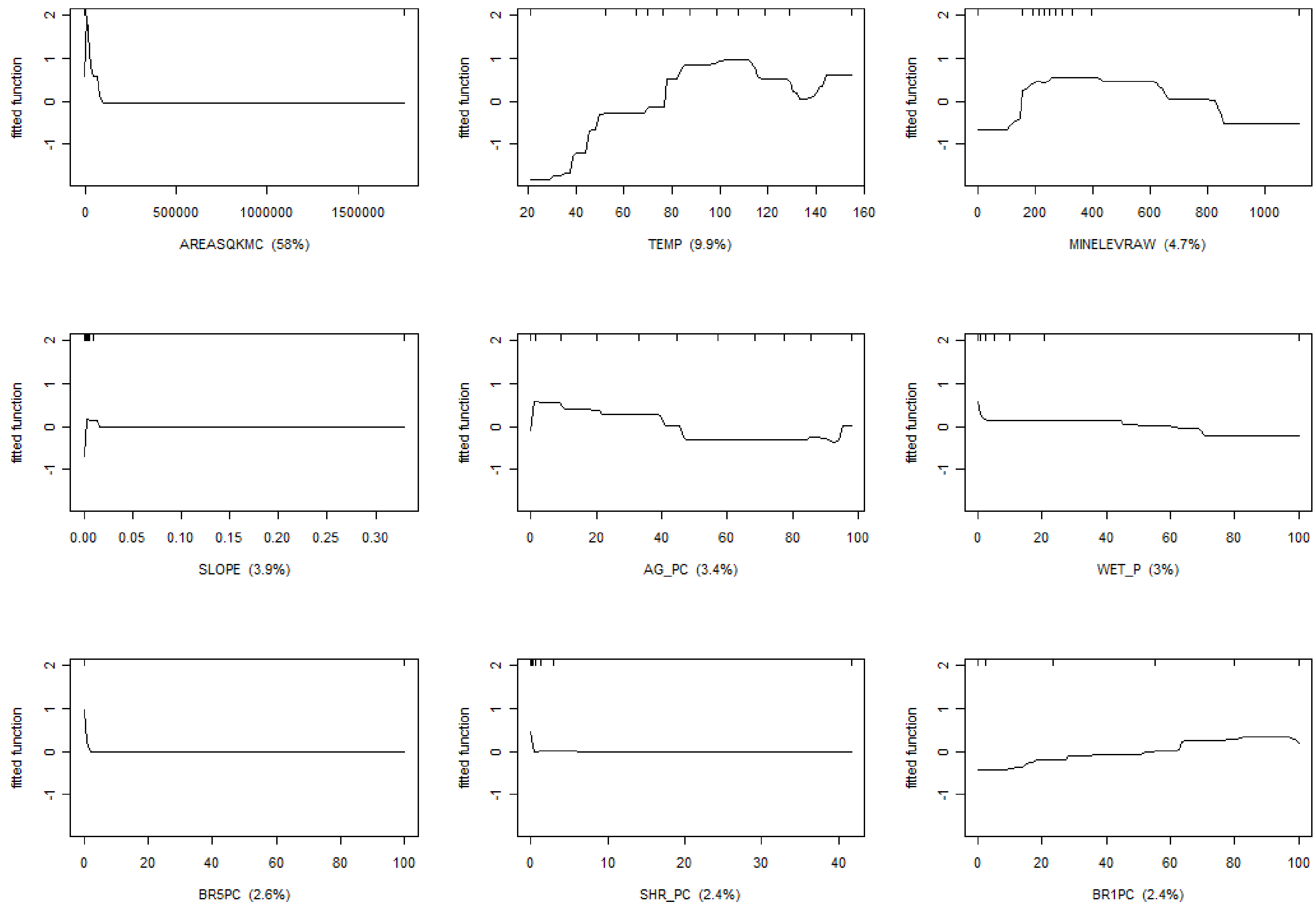
Appendix B: FUNCTIONAL RESPONSE PLOTS

Coldwater

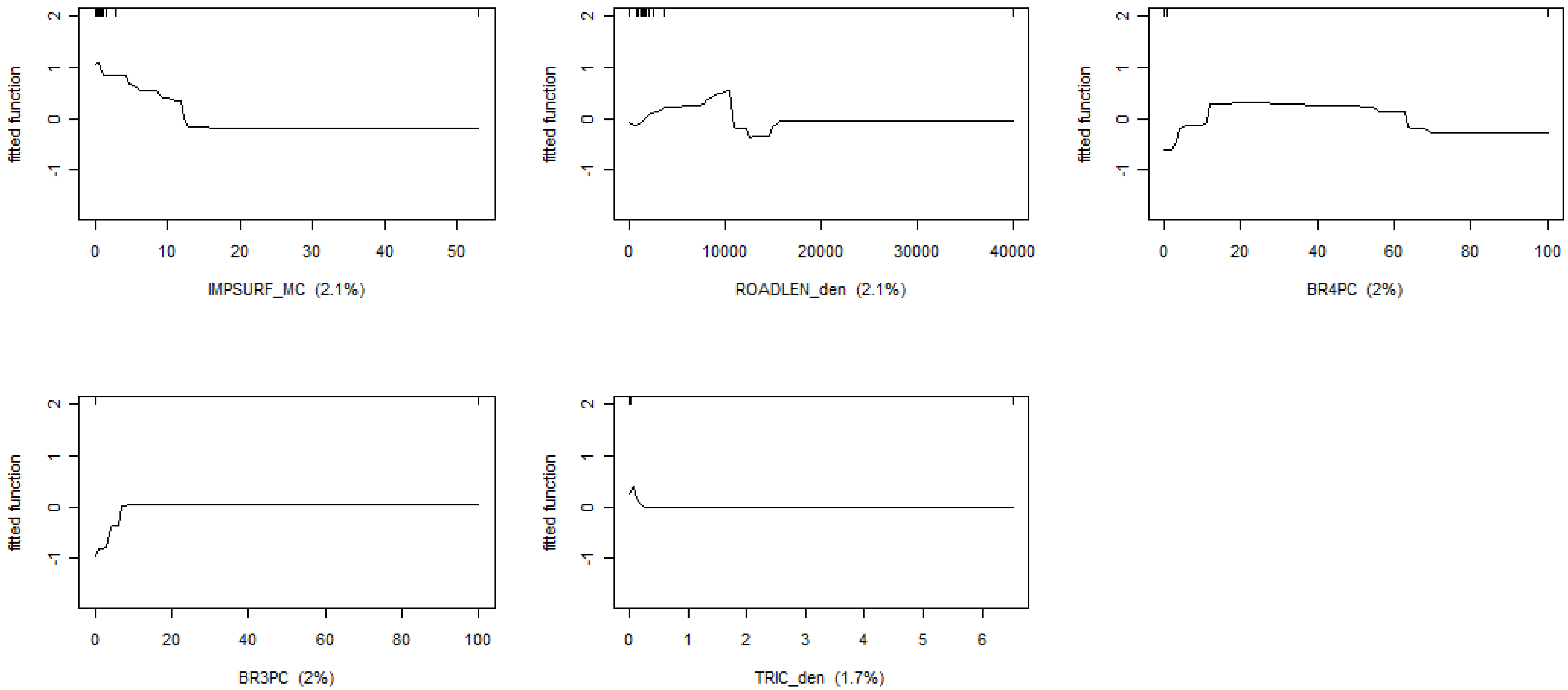



fitted function
-2
-1
0
1
2
0
5000
10000
15000
20000
25000
30000
WATER_SWC (1.3%)

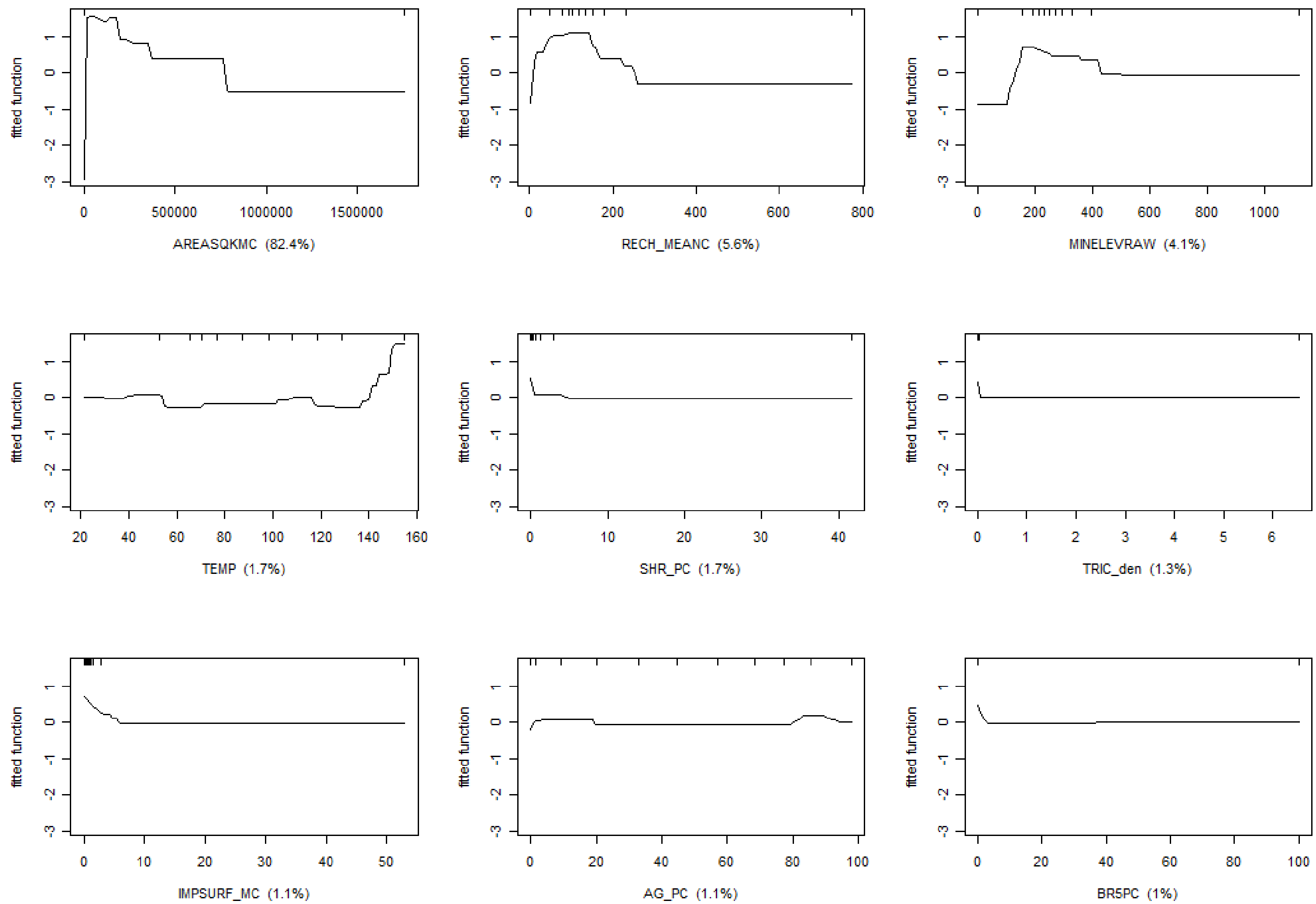
Coolwater



fitted function
AREASQKMC (58%)
TEMP (9.9%)
MINELEVRAW (4.7%)
SLOPE (3.9%)
AG_PC (3.4%)
WET_P (3%)
BR5PC (2.6%)
SHR_PC (2.4%)
BR1PC (2.4%)


fitted function
IMPSURF_MC (2.1%)
ROADLEN_den (2.1%)
BR4PC (2%)
BR3PC (2%)
TRIC_den (1.7%)

Warmwater


fitted function
AREASQKMC (82.4%)
RECH_MEANC (5.6%)
MINELEVRAW (4.1%)
TEMP (1.7%)
SHR_PC (1.7%)
TRIC_den (1.3%)
IMPSURF_MC (1.1%)
AG_PC (1.1%)
BR5PC (1%)

[i] http://www2.research.att.com/~phillips/pdf/Elith_et_al_ecography.pdf

[ii] http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2656.2008.01390.x/pdf