Cramer Fish Sciences

Montana Lakes Assessment

March 2012

Lead Investigator:

Mark Teply
Cramer Fish Sciences, Lacey Office
677 Woodland Sq. Lp SE, Suite D6
Lacey, WA 98503
(360) 456-4621
markt@fishsciences.net

CRAMER FISH SCIENCES®
Oregon • California • Washington • Idaho • Alaska

CORPORATE OFFICE
Gresham, Oregon

OAKDALE OFFICE
Oakdale, California

IDAHO OFFICE
Moscow, Idaho

AUBURN OFFICE
Auburn, California

COASTAL OFFICE
Coos Bay, Oregon

LACEY OFFICE
Lacey, Washington

Table of Contents

Introduction 1
Methods 1
Results 5
  Biological Indicators 6
  Chemical Stressors 9
  Physical Stressors 10
  Recreational Suitability 11
  Trophic State 12
Discussion 13

Appendix A: Lakes Sampled in Montana in the NLA
Appendix B: EPA Procedures for Calculating Population Estimates

List of Figures

Figure 1. Location of lakes sampled in Montana in the NLA.
Figure 2. Frequency of lakes sampled in Montana in the NLA by size class and ecoregion category.
Figure 3. Reference condition thresholds used for good, fair, and poor assessment.
Figure 4. NLA findings for all Montana lakes. Bars show the percentage of lakes within a condition class for a given indicator. For Recreational Chlorophyll risk and Cyanobacteria risk, the percentage numbers indicate the risk of exposure to algal toxins associated with the presence of chlorophyll-*a* and cyanobacteria, not the risk of exposure to chlorophyll-a and cyanobacteria *per se*.
Figure 5. NLA findings for the Western Mountains. Bars show the percentage of lakes within a condition class for a given indicator. For Recreational Chlorophyll risk and Cyanobacteria risk, the percentage numbers indicate the risk of exposure to algal toxins associated with the presence of chlorophyll-*a* and cyanobacteria, not the risk of exposure to chlorophyll-a and cyanobacteria *per se*.
Figure 6. NLA findings for the Northern Plains. Bars show the percentage of lakes within a condition class for a given indicator. For Recreational Chlorophyll risk and Cyanobacteria risk, the percentage numbers indicate the risk of exposure to algal toxins associated with the presence of chlorophyll-*a* and cyanobacteria, not the risk of exposure to chlorophyll-a and cyanobacteria *per se*.

List of Tables

Table 1. Summary of Carlson’s TSI for Montana lakes by ecoregion.

Introduction

Under the Clean Water Act (CWA), the U.S. Environmental Protection Agency (EPA) must periodically report on the condition of the nation's water resources by summarizing water quality information provided by the states. In this report, we summarize data collected in 2007 from 40 lakes, ponds, and reservoirs in the state of Montana as part of the National Lakes Assessment (NLA) and interpret these results to assess water quality, biological condition, habitat condition, and recreational suitability. The NLA is the first statistical survey of over 1,000 lakes. Survey results represent the state of natural and man-made lakes that are greater than 10 acres and over one meter deep.

Author's Note: Given the intended use of this report, and for consistency, we have incorporated many descriptions of methods and indicators verbatim *from the National Lakes Assessment Report (http://water.epa.gov/type/lakes/lakessurvey_index.cfm). Where appropriate, we have added detail relevant to Montana. Otherwise, the summaries of NLA results and their interpretation remain our original work product.*

Methods

EPA selected lakes for the NLA using a probability-based sampling design. Rules were developed to ensure that the design yielded a set of lakes that would support statistically valid conclusions. With input from the states, the following framework guided the national sample selection process. Information specific to Montana is noted.

- The National Hydrography Dataset (NHD) was used to derive a list of lakes for potential inclusion in the survey. For NHD summaries in this report, we downloaded the NHD data set available on November 7, 2011 from the Montana State Library, Natural Resource Information System (NRIS).
- For purposes of this survey, "lakes" refers to natural and manmade freshwater lakes, ponds, and reservoirs greater than 10 acres (4 hectares). In Montana, 5,547 lakes meet this criterion within the NHD data set downloaded from NRIS.

- The sample size was set to include 1,000 lake sampling events nationally. An 'oversample' of additional lakes was also done so that any state wishing to conduct a state scale survey could be accommodated. Nationally, the result was the selection of 909 lakes, with 91 scheduled for revisits. In Montana, 40 lakes were selected; of the 40 lakes, 5 were re-visited for verification sampling.
- The sample design was constructed to include a representative subset of the lakes that were included in the National Lake Eutrophication Study (NES), conducted by EPA in the 1970s. This allows for investigation of changes in trophic state. In Montana, 8 NES lakes were sampled by the NLA.
- Lake selection for the NLA survey covered 5 size class categories, as well as spatial distribution across the lower 48 states and 9 aggregated Omernik Level 3 ecoregions. In Montana, two aggregated ecoregions were encountered—Western Mountains and Northern Plains—and all five size classes were sampled.

Appendix A lists the NLA lakes sampled in Montana. Figure 1 shows the location of these lakes. Figure 2 shows the distribution by size class category and aggregated ecoregions. Nineteen lakes were natural; 21 were man-made. Nationally, the site selection process ensured that EPA can make unbiased estimates concerning the health of the target population with statistical confidence. Population estimates analysis weight each sample lake according to its probability of selection. We used an automated procedure provided by EPA (see Appendix B) to calculate the proportion of lakes in each condition class category being analyzed. Results are as reported by the EPA procedure and, because of rounding, some may not sum to 100% when tallied across categories. In some instances, results may not sum to 100% when some lakes were not assessed for a particular condition class. The margin of error for the Montana sample is displayed as thin lines on either side of the bars in the graphs throughout this report. These represent the 95% confidence interval.

Each lake was sampled in summer 2007. Samples were collected at the deepest point of the lake and at ten stations equidistant along the shore. Mid-lake sampling included physical parameters along a depth profile, collection of single grab samples for nutrients and zooplankton, and a sediment core. Along the shore, physical characteristics in the riparian and littoral zone were observed, the littoral zone was sampled for benthic macroinvertebrates, and the water was sampled for pathogens.

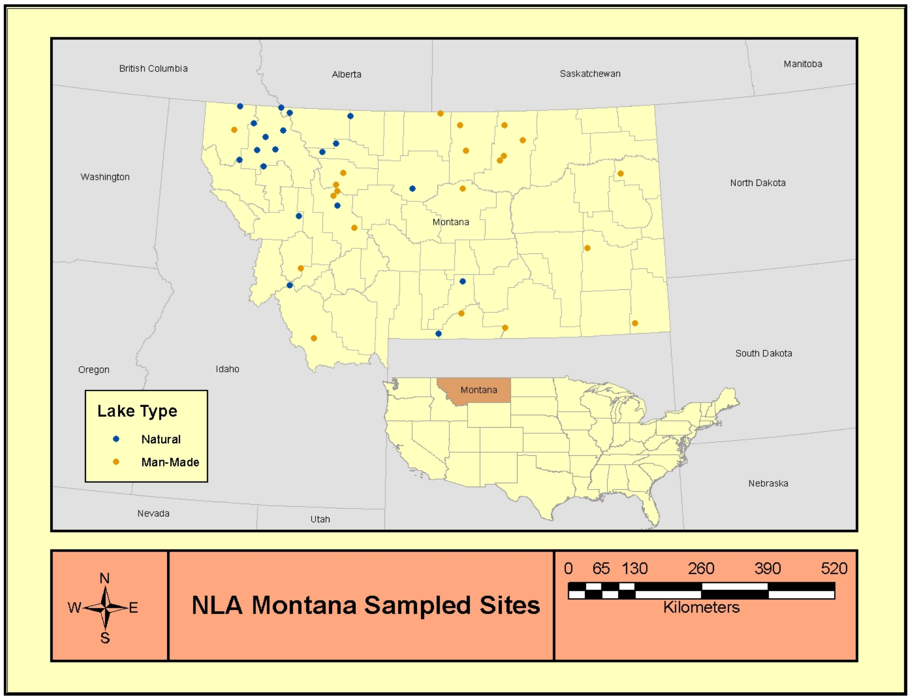

Figure 1. Location of NLA lakes sampled in Montana.

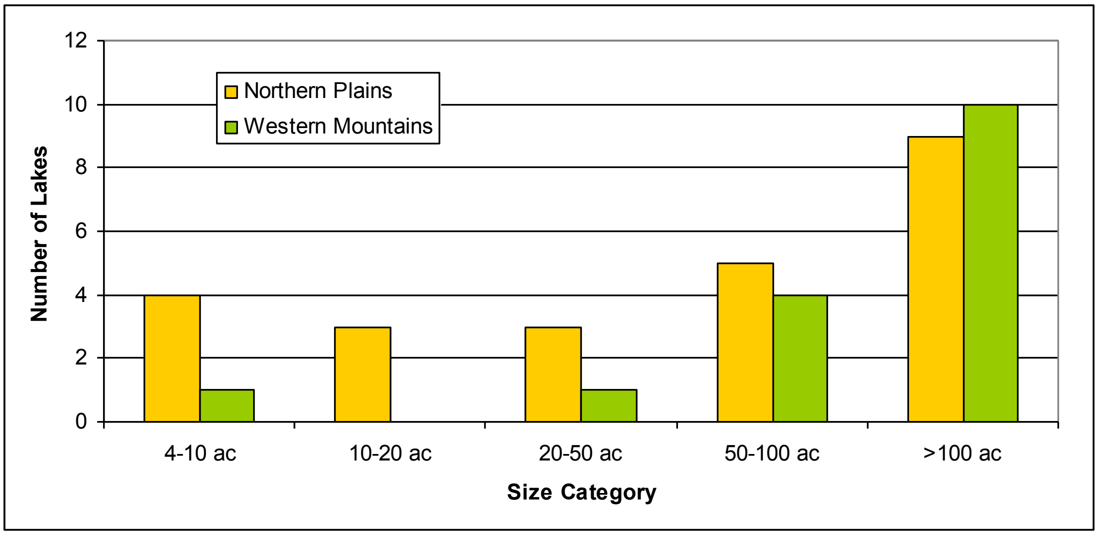

Figure 2. Frequency of NLA lakes sampled in Montana by size class and ecoregion category.

A suite of chemical, physical and biological indicators were chosen to assess biological integrity, trophic state, recreational suitability, and key stressors impacting the biological quality of lakes. NLA analysts decided that the results of the phytoplankton and zooplankton assessment would serve as the primary biological indicator. To address recreational/human health related concerns, the NLA looked at actual levels of the algal toxin microcystin, along with cyanobacterial cell counts and chlorophyll-*a* concentrations, as indicators of the potential for algal toxins. The presence and concentration of microcystin were used as the primary indicators for recreational condition. Chlorophyll-*a* was used as the primary indicator of trophic status.

Both physical and chemical stressor indicators were measured. Shoreline conditions were used given their effect on biological communities, such as providing food and shelter for aquatic wildlife, and moderating the magnitude, timing, and pathways of water, sediment, and nutrient inputs. Shorelines also buffer the lake from human activities. Water quality characteristics were measured—such as nutrient levels and dissolved oxygen—given their influence on environments essential for aquatic organisms to survive and grow. At the bottom of the lake, sediment diatoms allow for the examination of current water quality conditions such as phosphorus levels, as well as for a determination of historical conditions.

Two types of assessment thresholds were used in the NLA. Fixed thresholds are based on longstanding accepted values from the peer reviewed scientific literature. They are well established, and widely and consistently used. An example of this is standard chlorophyll-*a* thresholds, which are used to classify lakes into the different trophic categories. Reference-based thresholds are based on the distribution (i.e., the range of values) of a particular indicator derived from the reference lakes data. Only four reference sites were sampled in Montana; this was not a sufficient sample size for deriving assessment thresholds. Therefore, we adopted reference-based thresholds provided in the NLA database http://water.epa.gov/type/lakes/web_data.cfm. The derivation of threshold values from reference lakes is described below.

In the NLA, each indicator was classified as either “good,” “fair,” or “poor” relative to the conditions found in reference lakes. “Good” denotes an indicator value similar to that found in reference lakes, “fair” indicates conditions on the borderline of reference conditions, and “poor” denotes conditions definitely different from reference conditions. Specifically, these reference-based thresholds are then applied to the results from the target lakes and are classified as follows: lake results above 25% of the reference range values are considered “good,” those between 25% and 5% are “fair,” and those below 5% of the reference range value are “poor” (Figure 3).

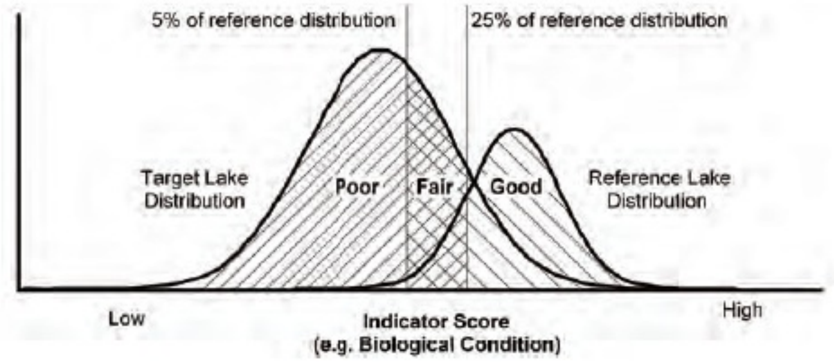

Figure 3. Reference condition thresholds used for good, fair, and poor assessment.

Results

Figures 4 through 6 provide a comprehensive summary of chemical, physical and biological indicators that were chosen to assess biological integrity, trophic state, recreational suitability, and key stressors impacting the biological quality of lakes. Results are shown for all Montana lakes and for Montana lakes by the two aggregated ecoregions encountered—Western Mountains and Northern Plains. The Western Mountains ecoregion occurs in the western portion of the state and includes the Bitterroot and Rocky Mountains. Ninety-eight percent of lakes in the Western Mountains are natural. Lakes in this ecoregion are relatively lower in nutrients and generally have lower productivity. The Northern Plains ecoregion occurs in the central eastern portion of the state. The Missouri River is the major river system draining this area. Seventy-five percent of the lakes in the Northern Plains are man-made.

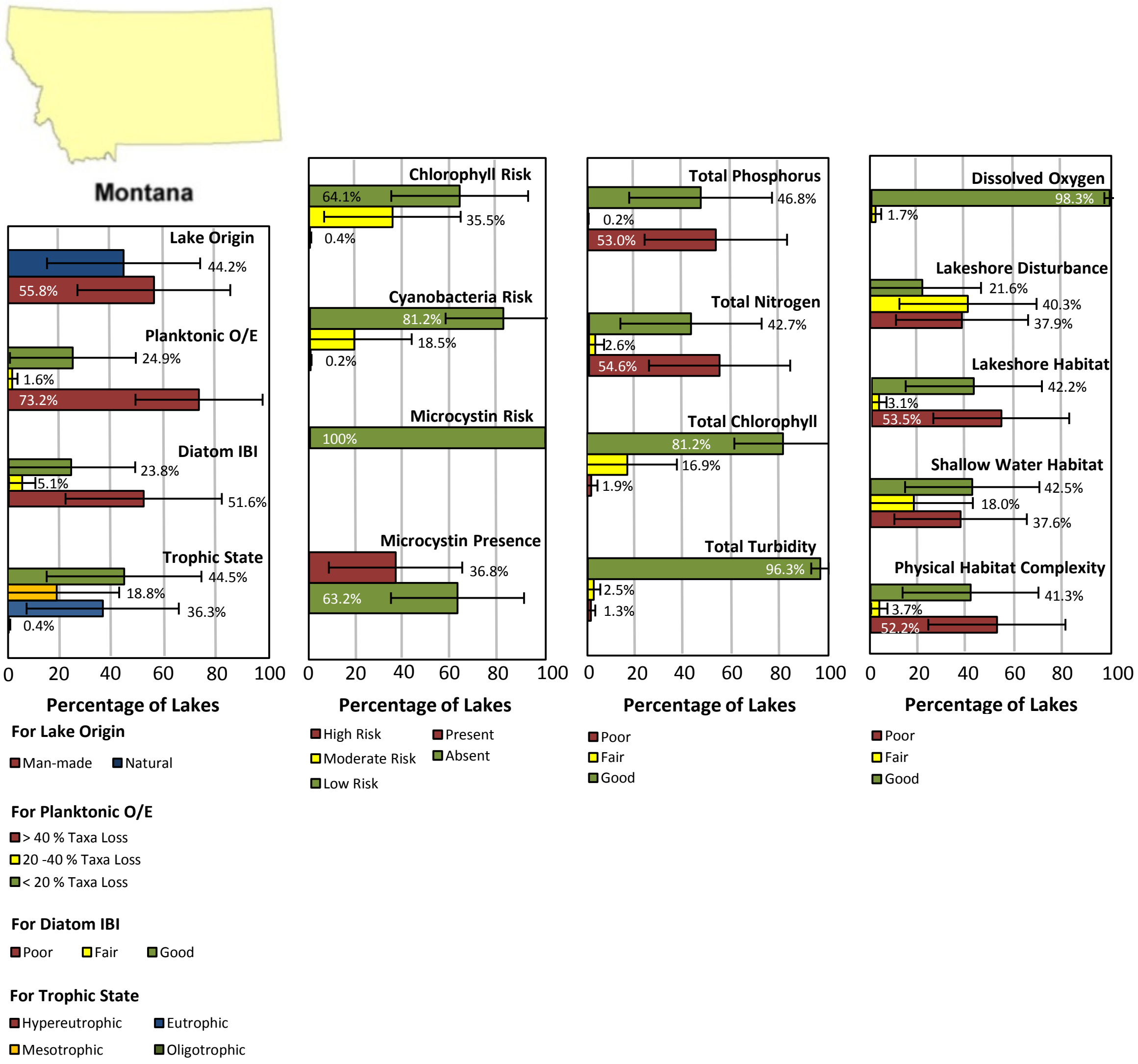

Figure 4. NLA findings for all Montana lakes. Bars show the percentage of lakes within a condition class for a given indicator. For Recreational Chlorophyll risk and Cyanobacteria risk, the percentage numbers indicate the risk of exposure to algal toxins associated with the presence of chlorophyll-*a* and cyanobacteria, not the risk of exposure to chlorophyll-a and cyanobacteria *per se*.

Biological Indicators

Two indices were used to assess biological condition of Montana lakes. NLA analysts used the planktonic O/E taxa loss model to assess the condition of the planktonic community, combining data from both phytoplankton and zooplankton. The O/E model evaluates whether or not organisms (taxa) one would expect to find based on reference

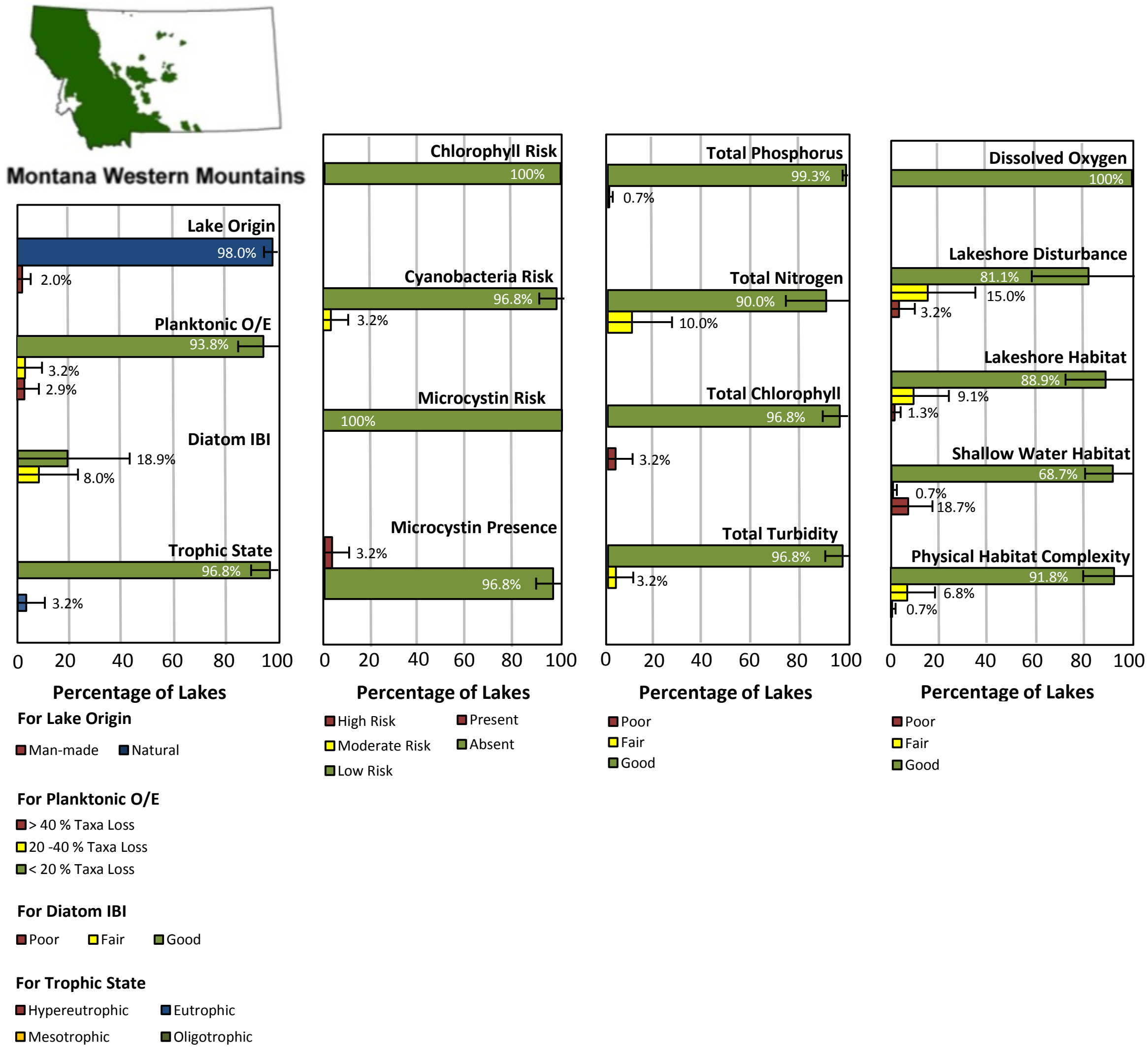

Figure 5. NLA findings for the Western Mountains. Bars show the percentage of lakes within a condition class for a given indicator. For Recreational Chlorophyll risk and Cyanobacteria risk, the percentage numbers indicate the risk of exposure to algal toxins associated with the presence of chlorophyll-*a* and cyanobacteria, not the risk of exposure to chlorophyll-a and cyanobacteria *per se*.

lakes are, in fact, present. Typically O/E values are interpreted as the percentage of the expected taxa present. Each tenth of a point less than 1 represents a 10% loss of taxa at the site; the higher the percentage, the healthier the lake. For the phytoplankton and zooplankton data, NLA analysts developed regionally-specific O/E models to predict the extent of taxa loss. They defined three categories of plankton taxa loss: good (<20% taxa loss), fair (20-40% taxa loss), and poor (>40% taxa loss).

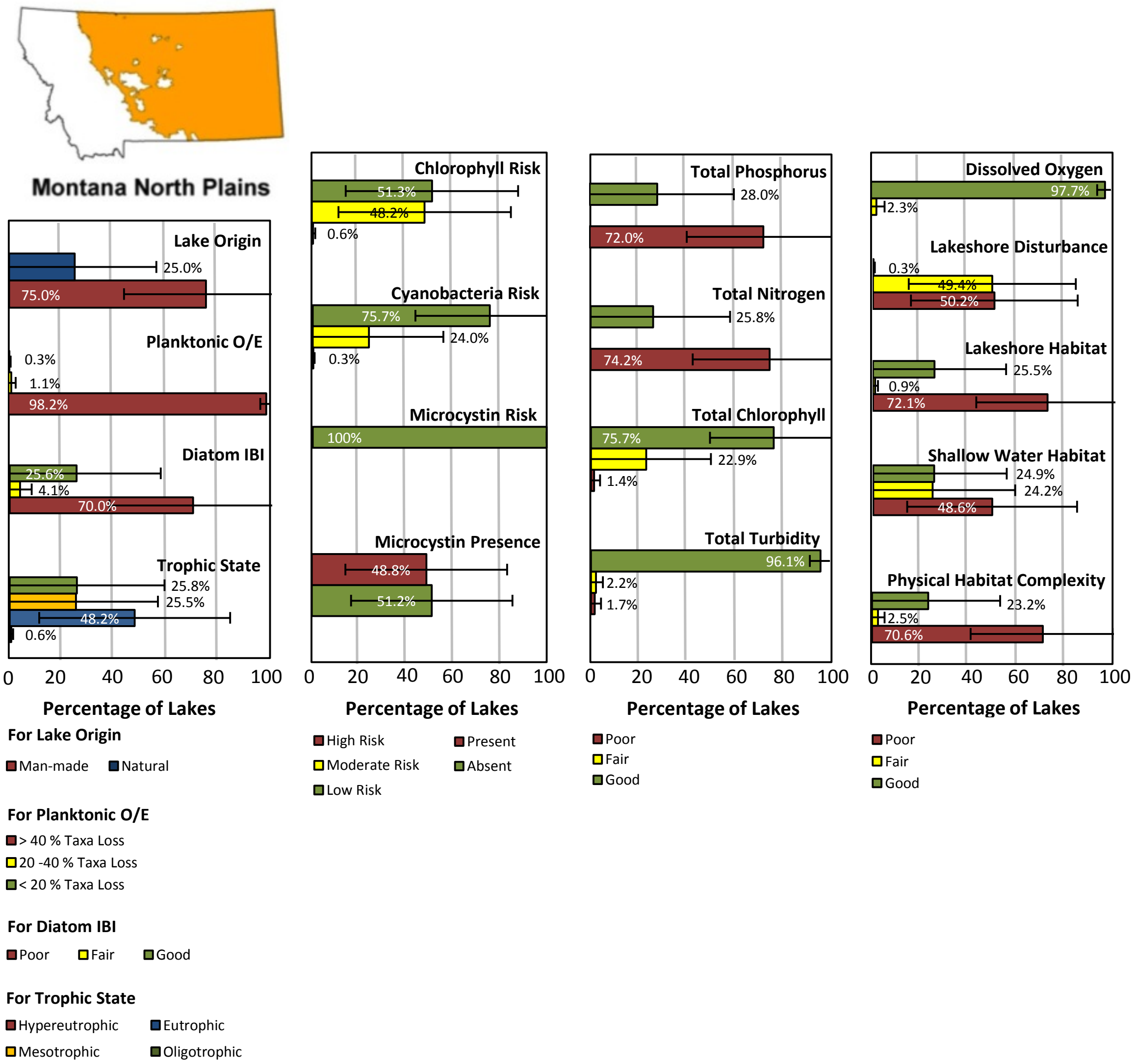

Figure 6. NLA findings for the Northern Plains. Bars show the percentage of lakes within a condition class for a given indicator. For Recreational Chlorophyll risk and Cyanobacteria risk, the percentage numbers indicate the risk of exposure to algal toxins associated with the presence of chlorophyll-*a* and cyanobacteria, not the risk of exposure to chlorophyll-a and cyanobacteria *per se*.

The Lake Diatom Condition Index (LDCI)—or the Diatom IBI—is based on five diatom assemblage characteristics: 1) Taxonomic richness, which represents the number of distinct taxa, or groups of organisms, identified within a sample; 2) Taxonomic composition, which assesses the relative abundance of organisms across taxa groups; 3) Taxonomic diversity, which considers the distribution of organisms among these groups; 4) Morphology, which classifies organisms based on their adaptations, including how they move and where they live; and, 5) Pollution tolerance, which evaluates the sensitivity of taxa found in the sample to chemical contamination—diverse assemblages of intolerant taxa generally indicate healthier biological conditions. NLA analysts calculated regionally-specific thresholds that were based on percentages of reference lake distributions of LDCI values.

Using the planktonic O/E metric, 25% of Montana lakes are in good condition, 2% are in fair condition, and 73% are in poor condition. The LCDI shows 24% in good condition, 5% in fair condition, 52% in poor condition; about 19% of the lakes were not assessed. Throughout Montana, this indicates that less than half of the state's lakes are in good condition while the remainder are experiencing some level of stress that is negatively affecting the aquatic biological communities. In the Western Mountains, most of the lakes that were assessed are in good condition, whereas in the Northern Plains, most are in fair or poor condition according to these metrics. Based on these metrics, the biological condition of Montana lakes is lower than that found nationally; about half of the lakes nation-wide are rated in good condition.

Chemical Stressors

Four chemical indicators of lake stress were evaluated: total phosphorus concentration, total nitrogen concentration, turbidity, and dissolved oxygen concentration. Phosphorus and nitrogen can be limiting to primary productivity. Modest increases in either of them can cause very rapid increases in algal growth. This can lead to lake eutrophication which has a negative impact on the biological community and other impacts (e.g., recreation). Turbidity measures the murkiness or clarity of the lake water. High turbidity can be caused by high concentrations of suspended sediment and/or high levels of algal cells. Increased turbidity often results in habitat alteration and changing algal growth, either of which can affect biological and recreational conditions. Dissolved oxygen is an indicator of the lake's ability to support aquatic life, since low DO levels can be limiting to aquatic life. Assessment thresholds are regionally-specific based on conditions in the NLA reference lakes.

Throughout Montana, slightly less than half of the lakes are in good condition with respect to phosphorus (47%) and nitrogen (43%). Almost all lakes are in good condition for turbidity (96%) and dissolved oxygen (98%). In the Western Mountains, 99% have good phosphorus conditions and 90% have good nitrogen conditions, whereas in the Northern Plains, 28% have good phosphorus conditions and 26% have good nitrogen conditions. There was less difference between ecoregions for other parameters—nearly all lakes in either had good turbidity and DO conditions. Generally, Montana lakes are in slightly poorer condition for nutrients than found in lakes nationally. Turbidity and dissolved oxygen conditions are better than those found nationally.

Physical Stressors

Physical habitat was assessed based on four indicators: 1) Human disturbance, which quantifies the extent and intensity of human activity along the lakeshore—human disturbance can be a physical stressor to aquatic life; 2) Lakeshore habitat, which examines the amount and type of shoreline vegetation—generally, lakeshores are in better condition when shoreline vegetation is full developed; 3) Shallow water habitat, which assesses the presence of living and non-living features such as overhanging vegetation, macrophytes, wood, boulders, and ledges—lakes with greater and more varied shallow water habitat are typically better able to support complex aquatic communities; and, 4) Physical habitat complexity, which combines information on the lakeshore and shallow water interface—greater complexity generally supports more complex biological communities. Thresholds are regionally specific based on conditions at NLA reference lakes.

Most Montana lakes exhibit some level of disturbance or habitat degradation that could affect biological communities. Based on lakeshore disturbance, over 75% of the lakes in Montana are in fair or poor condition. Slightly less than half exhibit good habitat conditions: 42% of lakes exhibit good lakeshore habitat condition, 3% are in fair condition, and 53% are in poor condition. For shallow water habitat indicators, a similar trend is evident: 43% are in good condition, 18% are fair, and 38% are in poor condition.

A similar trend is also evident for physical habitat complexity: 41% were found to be in good condition, 4% in fair condition, and 52% are in poor condition. In the Western Mountains, about 80% of the lakes have low (good) levels of lakeshore disturbance and about 90% exhibit good habitat conditions. In the Northern Plains, nearly all lakes have fair or poor lakeshore disturbance ratings. About 25% of the lakes exhibit good habitat condition ratings. State-wide, habitat conditions are poorer than those found nationally.

Recreational Suitability

Biological data were assessed in terms of suitability and safety for recreational use. Three indicators were assessed: microcystin, cyanobacteria, and chlorophyll-*a*. Blue-green algae (cyanobacteria) are part of all freshwater ecosystems. Eutrophication in lakes often results in conditions that can favor their growth. Cyanobacterial blooms can be unsightly and also have the potential to produce cyanotoxins. Algal density—measured by cyanobacteria or chlorophyll-*a*—serves as a proxy for the actual presence of these toxins. Cyanotoxins can pose multiple health risks, including tumors and death. Microcystin is a cyanotoxin produced by *Microcystis sp.* and is currently believed to be the most common algae in lakes, nationally. The NLA adopted World Health Organization thresholds for the assessment of these metrics.

Most lakes in Montana are considered to be at low risk based on these three indicators. Nearly all lakes in the Western Mountains are at low risk based on cyanobacteria and chlorophyll-*a* indicators. In comparison, 48% of the lakes in the Northern Plains have chlorophyll-*a* levels that pose a moderate or high risk, and 24% have cyanobacteria levels indicating a moderate or high risk. Microcystin levels for all lakes encountered in this project indicate low risk to human health. It is important to note that although microcystin was found at low levels, it was present in37% of the lakes. The highest presence rate was among lakes in the Northern Plains (49%). This, along with several other aspects of the sampling approach, as well as the reliability of these indicators, suggests that caution should be used in the interpretation of the results. The risk to human health could, in fact, be higher.

Trophic State

Trophic state depicts the biological productivity in lakes. Lakes with high nutrient levels, high plant production rates, and abundant plant life are eutrophic. Conversely, those that have low nutrient levels, low productivity, and low biomass are termed oligotrophic. Lakes that fall in between are mesotrophic. Lakes naturally fall into all of these categories, but human influences that increase the amount of nutrients in lakes can accelerate eutrophication and lead to undesirable effects including nuisance algae, excessive plant growth, lower clarity, odor, and fish kills. For the NLA, trophic state is characterized using nationally consistent chlorophyll-*a* concentrations. Alternative classifications based on Secchi transparency, nitrogen, and phosphorus can also be used. However, NLA analysts chose to use chlorophyll-*a* concentrations.

Using only chlorophyll-*a* concentrations (data not shown in this report), 44% of lakes are oligotrophic, 19% are mesotrophic, 36% are eutrophic, and <1% are hypereutrophic. Almost all lakes in the Western Mountains are oligotrophic (97%), whereas in the Northern Plains, 26% are oligotrophic. Nearly half of the lakes in the Northern Plains are eutrophic or hypereutrophic. Compared to national findings, Montana lakes show relatively lower levels of eutrophication across the state. This is somewhat inconsistent with the observations we made for chemical stressors, where we found that Montana lakes are generally in poorer condition for nutrients—factors in eutrophication. Conversely, turbidity and DO—also factors in eutrophication—were slightly better.

We also calculated Carlson's Trophic State Index (TSI), following procedures on the EPA website http://www.epa.gov/bioiweb1/aquatic/carlson.html. Carlson's TSI uses Secchi disk values as a measure of algal mass. Carlson's TSI can also be calculated based on chlorophyll-*a* and total phosphorus, which can be closely correlated with Secchi disk readings. Trophic state can be interpreted from Carlson's TSI values following thresholds at http://www.secchidipin.org/tsi.htm. Lower values of Carlson's TSI indicate oligotrophic conditions; higher values indicate eutrophication. Table 1 summarizes Carlson's TSI values for all Montana lakes and by ecoregion. Carlson's

TSI indicates a slightly higher eutrophication rate than that indicated by using the NLA chlorophyll-*a*. TSI based on Secchi disk readings and total phosphorus showed an even greater rate of eutrophication. For all indices, ecoregional differences are still evident. Lakes in the Northern Plains show greater rates of eutrophication compared to lakes in the Western Mountains.

Table 1. Summary of Carlson's TSI for Montana lakes by ecoregion.

Region	**TSI Parameter**	**Oligotrophic 0-30**	**Mesotrophic 30-50**	**Eutrophic 50-70**	**Hypereutrophic 70+**
Montana	Secchi Disk	2%	30%	65%	3%
	Chlorophyll-*a*	23%	40%	36%	0%
	Total Phosphorus	9%	19%	19%	53%
Western Mountains	Secchi Disk	8%	92%	0%	0%
	Chlorophyll-*a*	80%	17%	3%	0%
	Total Phosphorus	31%	69%	0%	0%
Northern Plains	Secchi Disk	0%	3%	93%	4%
	Chlorophyll-*a*	3%	48%	48%	1%
	Total Phosphorus	1%	1%	26%	73%

Discussion

According to the NLA biological indices, about one-quarter of Montana lakes are in good biological condition. Indicators in the Western Mountains show mostly good conditions; whereas in the Northern Plains, they indicate mostly fair or poor conditions. Poor conditions in the Western Mountains occur less frequently, and are mostly attributable to physical stressors (lakeshore disturbance, lakeshore habitat, shallow water habitat, and physical habitat complexity). In the Northern Plains poor conditions can be attributed to physical stressors as well as nutrient stressors (total phosphorus and total nitrogen). Good conditions predominate for turbidity and DO in both ecoregions.

Montana lakes show lower rates of eutrophication than found nationally. Eutrophication was higher in the Northern Plains than in the Western Mountains. The relatively higher levels of eutrophication we found in Northern Plains lakes are generally consistent with the higher incidence of total nitrogen and total phosphorus stress encountered in these lakes. In comparison, nearly all lakes in the Western Mountains ecoregion were oligotrophic. Nutrient stress was not common in the Western Mountains. This reflects the generally low levels of nutrients found in these lakes. Trophic state measured using Carlson's TSI indicates slightly higher levels of eutrophication. Some of this difference could simply be attributable to different thresholds used to interpret the metrics. However, using TSI, eutrophication still occurred at a higher rate in the North Plains.

Between 1972 and 1976, EPA conducted the National Eutrophication Survey (NES). The NES was designed to assess the trophic condition of lakes influenced by waste water treatment plants. Eight of the lakes sampled in Montana in the NLA are NES lakes. We found sampling results for seven of these lakes on EPA's Storage and Retrieval Data Warehouse (STORET). We used results in STORET and in the NLA to evaluate trends in Carlson's TSI over time. We used summer observations from STORET that were comparable with the NLA. Based on Secchi disk observations, TSI decreased in 6 of 7 NES lakes. TSI based on chlorophyll-*a* decreased in 5 of 7 lakes. Using total phosphorus results, TSI decreased in all 7 NES lakes for which data were available. All lakes showed improvement in at least two of these TSI metrics. Generally, this indicates that nutrient levels and algal growth are decreasing; the trophic status of these lakes is better than observed in the mid-1970s.

Appendix A: Lakes Sampled in Montana in the NLA

SITE_ID	LAKENAME	LAT_DD	LON_DD	REPEAT	NESLAKE	AREA_CAT7	WSA_ECO9	LAKE_ORIGIN
NLA06608-0001	Lake Wurdeman	48.97902855	-114.0218399			(50,100]	WMT	NATURAL
NLA06608-0064	Fitzpatrick Lake	48.8940613	-112.1680165	YES		(20,50]	NPL	NATURAL
NLA06608-0065	Sophie Lake	48.96243381	-115.116608	YES		(50,100]	WMT	NATURAL
NLA06608-0128	Twin Reservoir	48.75289143	-109.2416928	YES		(10,20]	NPL	MAN-MADE
NLA06608-0129	Ashley Lake	48.21240982	-114.5994561	YES		>100	WMT	NATURAL
NLA06608-0225	Nilan Reservoir	47.48039976	-112.5389845	YES		>100	NPL	MAN-MADE
NLA06608-0240	Nelson Reservoir	48.47756954	-107.5902123		NESLake	>100	NPL	MAN-MADE
NLA06608-0257	Koocanusa Reservoir	48.54139505	-115.2253416		NESLake	>100	WMT	MAN-MADE
NLA06608-0378	Willow Creek Reservoir	47.55716773	-112.4475051			>100	NPL	MAN-MADE
NLA06608-0385	Green Lake	48.24355	-112.8672712			(20,50]	NPL	NATURAL
NLA06608-0496		48.75171319	-108.0524686			(10,20]	NPL	MAN-MADE
NLA06608-0570	Cooney Reservoir	45.44213549	-109.2242066			>100	NPL	MAN-MADE
NLA06608-0833	Glenns Lake	48.90066286	-113.79482			>100	WMT	NATURAL
NLA06608-0929	Georgetown Lake	46.17967552	-113.2957121		NESLake	>100	WMT	MAN-MADE
NLA06608-0944	West Alkali Reservoir	48.20653603	-108.0879027			(50,100]	NPL	MAN-MADE
NLA06608-1002		47.64355769	-109.1837945			(4,10]	NPL	MAN-MADE
NLA06608-1008	Eureka Reservoir	47.88008342	-112.3069538			>100	NPL	MAN-MADE
NLA06608-1089	Lake Mary Ronan	47.92615582	-114.4032969		NESLake	>100	WMT	NATURAL
NLA06608-1153	Lake McDonald	48.57775125	-113.9310021		NESLake	>100	WMT	NATURAL
NLA06608-1338	Hailstone Lake	46.01155947	-109.1798402			>100	NPL	NATURAL
NLA06608-1344		48.39247752	-112.5192316			(4,10]	NPL	NATURAL
NLA06608-1354	Bighorn Lake	45.18177616	-108.1327104		NESLake	>100	NPL	MAN-MADE
NLA06608-1377	Mystic Lake	45.86796274	-113.5499319			(4,10]	WMT	NATURAL
NLA06608-1462		46.53835681	-106.0071156			(10,20]	NPL	MAN-MADE
NLA06608-1510		47.82117438	-105.0491133			(4,10]	NPL	MAN-MADE
NLA06608-1578		45.18292402	-104.9019583			(4,10]	NPL	MAN-MADE
NLA06608-1600	McCann Reservoir	48.31132355	-109.0848391			(20,50]	NPL	MAN-MADE
NLA06608-1610	Fossil Lake	45.08626125	-109.7887823			(50,100]	WMT	NATURAL
NLA06608-1633	Clark Canyon Reservoir	44.95420653	-112.8818444		NESLake	>100	WMT	MAN-MADE
NLA06608-1856	Creedman Reservoir	48.96606399	-109.75936			(50,100]	NPL	MAN-MADE
NLA06608-1857	Lake Blaine	48.2461517	-114.1200149			>100	WMT	NATURAL
NLA06608-1953	Salmon Lake	47.09945654	-113.4075253			>100	WMT	NATURAL
NLA06608-1968	Wild Horse Reservoir	48.13399568	-108.1899505			(50,100]	NPL	MAN-MADE
NLA06608-2049	Lower Thompson Lake	48.02025253	-115.0345388			(50,100]	WMT	NATURAL
NLA06608-2170	Bean Lake	47.30587695	-112.4310378			(50,100]	NPL	NATURAL
NLA06608-2177	Bull Lake	48.671943	-114.7318229			(20,50]	WMT	NATURAL
NLA06608-2426	Lost Lake	47.6347722	-110.4842973			(50,100]	NPL	NATURAL
NLA06608-2800	Pishkun Reservoir	47.67808029	-112.4771506			>100	NPL	MAN-MADE
NLA06608-2874	Holter Lake	46.93083483	-111.9632243			>100	NPL	MAN-MADE
NLA06608-2881	Whitefish Lake	48.45509278	-114.388		NESLake	>100	WMT	NATURAL

SITE_ID	ID Assigned to Each Site
LAKENAME	Lake name (from field forms)
LAT_DD	Latitude (decimal degrees) obtained from NHD (NAD83)
LON_DD	Longitude (decimal degrees) obtained from NHD (NAD83)
REPEAT	Repeat visit lake (YES/blank)
NESLAKE	NESLake-Lake was included in 1970s National Eutrophication Survey
AREA_CAT7	Lake area unequal probability category (7 acreage categories)
WSA_ECO9	Wadeable Stream Assessment nine aggregated Omernik level 3 ecoregions
LAKE_ORIGIN	Lake origin (MAN-MADE, NATURAL [which includes natural lakes augmented by dams])

Appendix B: EPA Procedures for Calculating Population Estimates

Condition Class Estimates for an Individual State

Thomas Kincaid

May 18, 2011

Contents

1 Preliminaries **1**

2 Read the chemical condition class file **2**

3 Analysis of proportion and size for the number of lakes in each condition class category **5**

4 Analysis of proportion and size for the area of lakes in each condition class category **7**

1 Preliminaries

This document presents an example analysis of condition class variables for an individual state using data from the National Lakes Assessment (NLA) for 2007. The state chosen for the analysis is Oregon. Condition class variables are chosen from the set of chemical condition class variables that are included in the data. The analysis will include calculation of two types of population estimates: (1) estimation of proportion and size for the number of lakes in each condition class category; and (2) estimation of proportion and size for the area of lakes in each condition class category. R code for the analyses included in this document is contained in the R script file named ”Condition_Estimates.R”.

Prior to executing the R script file, it is necessary to install the current version of the R software (version 2.13.0). Once R is installed, current versions of the spsurvey library (version 2.2) and sp library (version 0.9-81) will need to be installed. The sp library should be installed prior to the spsurvey library. The Aquatic Resources Monitoring (ARM) web page, `http://www.epa.gov/nheerl/arm/`, includes the installation files for R and the spsurvey and sp libraries in addition to installation instructions for the files and tips for using R. Click on the ”Download Software” tab to access the installation and instruction files.

The initial step is to use the library function to load the spsurvey package. After the spsurvey packages is loaded, a message is printed to the R console indicating that the spsurvey package was loaded successfully.

Load the spsurvey package

```
> library(spsurvey)
```

```
Version 2.2 of the spsurvey package was loaded successfully.
```

2 Read the chemical condition class file

The next step is to read the chemical variables condition class file. The read.delim function is used to read the tab-delimited file and assign it to a data frame named condition. In order to use one of the other sets of condition variables, say trophic state condition class, change the name of the file to the appropriate value. The data frame is then restricted to include only sites for which the ADJWGT_CAT variable equals ”OR”, the VISIT_NO variable equals ”1”, the SITE_TYPE variable equals ”PROB_Lake”, and the LAKE_SAMP variable equals ”Target_Sampled”. In order to retain the desired sites, the first step is to create a logical variable named ”keep” that contains values ”TRUE” or ”FALSE” depending on whether the conditions are met. The second step is to retains only those rows in the condition data frame for which keep equals ”TRUE”.

Read the chemical variables condition class file and retain desired sites

```
> condition <- read.delim("nla_chem_cond_20091123.txt")
> keep <- condition$ADJWGT_CAT == "OR" &
+         condition$INDXSAMP_CHEM == "YES" &
+         condition$SITE_TYPE == "PROB_Lake" &
+         condition$LAKE_SAMP == "Target_Sampled"
> condition <- condition[keep,]
```

In order to select sites for a state other than Oregon, use the code appropriate for the desired state in the logical equality involving the ADJWGT_CAT variable, e.g., replace ”OR” with ”OK” for Oklahoma. For several of the condition class files, the INDXSAMP_CHEM variable must be placed in the code used to create the keep variable. Use INDXSAMP_LDC for lake diatom variables. Use INDXSAMP_OE5 for O/E variables. Use INDXSAMP_MICR for recreation variables. Use INDXSAMP_INF for diatom-inferred chemistry variables. In addition, for two sets of condition class variables, the code used to create the keep variable must be modified. For dissolved oxygen and physical habitat condition variables, use the following code:

```
keep <- condition$ADJWGT_CAT == "OR" &
        condition$VISIT_NO == "1" &
        condition$SITE_TYPE == "PROB_Lake" &
        condition$LAKE_SAMP == "Target_Sampled"
condition <- condition[keep,]
```

The local mean variance estimator is used to calculate variance of the condition class estimates. This variance estimator requires calculation of distance between sample points, which means that locations for sample points must be expressed in a coordinate system appropriate for calculation of distance. The geodalbers function is used to convert latitude and longitude to the Albers projection.

Use the geodalbers function to convert latitude and longitude to the Albers projection.

```
> temp <- geodalbers(condition$LON_DD, condition$LAT_DD, sph = "GRS80")
> condition$xAlbers <- temp$xcoord
> condition$yAlbers <- temp$ycoord
```

Finally, the initial six lines and the final six lines in the condition data frame are printed using the head and tail functions, respectively.

Display the initial six lines in the chemical condition class data frame.

```
> head(condition)
```

```
          SITE_ID VISIT_NO SITE_TYPE      LAKE_SAMP    TNT   LAT_DD    LON_DD
64  NLA06608-0049        1 PROB_Lake Target_Sampled Target 45.06233 -117.1534
244 NLA06608-0290        1 PROB_Lake Target_Sampled Target 42.19565 -120.5257
254 NLA06608-0306        1 PROB_Lake Target_Sampled Target 43.41788 -119.4133
301 NLA06608-0402        1 PROB_Lake Target_Sampled Target 42.88975 -124.0781
305 NLA06608-0406        1 PROB_Lake Target_Sampled Target 45.18038 -121.7044
412 NLA06608-0614        1 PROB_Lake Target_Sampled Target 43.37864 -123.2686
    ST   EPA_REG AREA_CAT7 NESLAKE STRATUM   PANEL        DSGN_CAT     MDCATY
64  OR Region_10   (10,20]         NLALake Panel_1  WMT_OR_(10,20] 0.01349449
244 OR Region_10  (50,100]         NLALake Panel_1 WMT_OR_(50,100] 0.05533142
254 OR Region_10      >100         NLALake Panel_1     XER_OR_>100 0.05417501
301 OR Region_10    (4,10]         NLALake Panel_1          (4,10] 0.00131289
305 OR Region_10      >100         NLALake Panel_1     WMT_OR_>100 0.05025013
412 OR Region_10  (50,100]         NLALake Panel_1 WMT_OR_(50,100] 0.05533142
          WGT    WGT_NLA ADJWGT_CAT URBAN WSA_ECO3 WSA_ECO9 ECO_LEV_3 NUT_REG
64   73.39093  16.398147         OR    NO    WMTNS      WMT        11      II
244  17.89893   3.999259         OR    NO    WMTNS      WMT         9      II
254  18.28100   4.084627         OR    NO    WMTNS      XER        80     III
301 754.34539 168.547626         OR    NO    WMTNS      WMT         1      II
305  19.70887   4.403664         OR    NO    WMTNS      WMT         4      II
412  17.89893   3.999259         OR   YES    WMTNS      WMT        78      II
                   NUTREG_NAME ECO_NUTA LAKE_ORIGIN  ECO3_X_ORIGIN REF_CLUSTER
64  Western Forested Mountains       II    MAN-MADE WMTNS MAN-MADE           F
244 Western Forested Mountains       II    MAN-MADE WMTNS MAN-MADE           F
254                 Xeric West      III    MAN-MADE WMTNS MAN-MADE           G
301 Western Forested Mountains       II    MAN-MADE WMTNS MAN-MADE           F
305 Western Forested Mountains       II     NATURAL  WMTNS NATURAL           F
412 Western Forested Mountains       II    MAN-MADE WMTNS MAN-MADE           F
    RT_NLA HUC_2    HUC_8 FLAG_INFO COMMENT_INFO SAMPLED SAMPLED_CHEM
64   SO-SO    17 17050201                            YES          YES
244  TRASH    18 18020001                            YES          YES
254  TRASH    17 17120004                            YES          YES
301  TRASH    17 17100305                            YES          YES
305  SO-SO    17 17070306                            YES          YES
412  SO-SO    17 17100301                            YES          YES
    INDXSAMP_CHEM PTL  NTL   TURB    ANC  DOC COND SAMPLED_CHLA INDXSAMP_CHLA
64            YES  15  256  1.550  829.8 1.88   92          YES           YES
244           YES 178  273 46.700 1404.9 5.54  142          YES           YES
254           YES 271 1525 96.300 3413.5 9.64  330          YES           YES
301           YES 176 2287 10.200  829.7 8.43  101          YES           YES
305           YES   4  172  0.792  202.9 2.34   24          YES           YES
412           YES   8  259  0.825  705.7 2.82  219          YES           YES
     CHLA          PTL_COND                   NTL_COND         CHLA_COND
64   1.14 1:LEAST DISTURBED          1:LEAST DISTURBED 1:LEAST DISTURBED
244  4.30  3:MOST DISTURBED          1:LEAST DISTURBED  3:MOST DISTURBED
254 38.73  3:MOST DISTURBED 2:INTERMEDIATE DISTURBANCE  3:MOST DISTURBED
301 73.73  3:MOST DISTURBED           3:MOST DISTURBED  3:MOST DISTURBED
305  1.23 1:LEAST DISTURBED          1:LEAST DISTURBED 1:LEAST DISTURBED
```

```
412  1.22 1:LEAST DISTURBED          1:LEAST DISTURBED 1:LEAST DISTURBED
                     TURB_COND          ANC_COND     SALINITY_COND  xAlbers
64  2:INTERMEDIATE DISTURBANCE 1:LEAST DISTURBED 1:LEAST DISTURBED -1650431
244            3:MOST DISTURBED 1:LEAST DISTURBED 1:LEAST DISTURBED -1989729
254            3:MOST DISTURBED 1:LEAST DISTURBED 1:LEAST DISTURBED -1868057
301            3:MOST DISTURBED 1:LEAST DISTURBED 1:LEAST DISTURBED -2247455
305           1:LEAST DISTURBED 1:LEAST DISTURBED 1:LEAST DISTURBED -1994097
412           1:LEAST DISTURBED 1:LEAST DISTURBED 1:LEAST DISTURBED -2169024
    yAlbers
64  2637102
244 2390980
254 2500657
301 2544833
305 2737125
412 2578329
```

Display the final six lines in the chemical condition class data frame.

```
> tail(condition)
```

```
           SITE_ID VISIT_NO SITE_TYPE      LAKE_SAMP    TNT   LAT_DD    LON_DD
984  NLA06608-2082        1 PROB_Lake Target_Sampled Target 42.15140 -122.6008
1029 NLA06608-2438        1 PROB_Lake Target_Sampled Target 42.06681 -119.5634
1030 NLA06608-2450        1 PROB_Lake Target_Sampled Target 43.63164 -124.1790
1035 NLA06608-2481        1 PROB_Lake Target_Sampled Target 43.98863 -119.3928
1052 NLA06608-2673        1 PROB_Lake Target_Sampled Target 44.30658 -118.6850
1059 NLA06608-2726        1 PROB_Lake Target_Sampled Target 43.71379 -121.7665
     ST   EPA_REG AREA_CAT7 NESLAKE STRATUM    PANEL       DSGN_CAT     MDCATY
984  OR Region_10      >100         NLALake OverSamp    WMT_OR_>100 0.05025013
1029 OR Region_10   (20,50]         NLALake OverSamp XER_OR_(20,50] 0.02202713
1030 OR Region_10   (10,20]         NLALake OverSamp WMT_OR_(10,20] 0.01349449
1035 OR Region_10    (4,10]         NLALake OverSamp         (4,10] 0.00131289
1052 OR Region_10   (10,20]         NLALake OverSamp WMT_OR_(10,20] 0.01349449
1059 OR Region_10   (20,50]         NLALake OverSamp WMT_OR_(20,50] 0.02043131
           WGT    WGT_NLA ADJWGT_CAT URBAN WSA_ECO3 WSA_ECO9 ECO_LEV_3 NUT_REG
984   19.70887   4.403664         OR    NO    WMTNS      WMT        78      II
1029  44.96153  10.046007         OR    NO    WMTNS      XER        80     III
1030  73.39093  16.398147         OR    NO    WMTNS      WMT         1      II
1035 754.34539 168.547626         OR    NO    WMTNS      WMT        11      II
1052  73.39093  16.398147         OR    NO    WMTNS      WMT        11      II
1059  48.47333  10.830667         OR    NO    WMTNS      WMT         9      II
                   NUTREG_NAME ECO_NUTA LAKE_ORIGIN  ECO3_X_ORIGIN REF_CLUSTER
984  Western Forested Mountains       II    MAN-MADE WMTNS MAN-MADE           F
1029                Xeric West      III    MAN-MADE WMTNS MAN-MADE           F
1030 Western Forested Mountains       II     NATURAL  WMTNS NATURAL           F
1035 Western Forested Mountains       II    MAN-MADE WMTNS MAN-MADE           F
1052 Western Forested Mountains       II     NATURAL  WMTNS NATURAL           F
1059 Western Forested Mountains       II     NATURAL  WMTNS NATURAL           F
```

```
     RT_NLA HUC_2    HUC_8 FLAG_INFO COMMENT_INFO SAMPLED SAMPLED_CHEM
984   SO-SO    17 17100308                            YES          YES
1029  TRASH    17 17120008                            YES          YES
1030  SO-SO    17 17100304                            YES          YES
1035  SO-SO    17 17070201                            YES          YES
1052  SO-SO    17 17070201                            YES          YES
1059  SO-SO    17 17070301                            YES          YES
     INDXSAMP_CHEM PTL  NTL    TURB    ANC   DOC COND SAMPLED_CHLA
984            YES  17  496   3.350 1070.3  3.07  118          YES
1029           YES 636 1674 152.000 1510.6 13.00  152          YES
1030           YES   3  148   0.817  219.1  2.28   87          YES
1035           YES  36  421   8.750 2006.5  4.79  216          YES
1052           YES  72  423   0.530  407.9  2.31   45          YES
1059           YES   5  191   0.610 1485.2  2.45  152          YES
     INDXSAMP_CHLA CHLA                   PTL_COND                   NTL_COND
984            YES 3.57 2:INTERMEDIATE DISTURBANCE           3:MOST DISTURBED
1029           YES 4.64           3:MOST DISTURBED 2:INTERMEDIATE DISTURBANCE
1030           YES 0.90          1:LEAST DISTURBED          1:LEAST DISTURBED
1035           YES 0.95           3:MOST DISTURBED           3:MOST DISTURBED
1052           YES 1.33           3:MOST DISTURBED           3:MOST DISTURBED
1059           YES 1.18          1:LEAST DISTURBED          1:LEAST DISTURBED
             CHLA_COND                  TURB_COND          ANC_COND
984   3:MOST DISTURBED 2:INTERMEDIATE DISTURBANCE 1:LEAST DISTURBED
1029 1:LEAST DISTURBED           3:MOST DISTURBED 1:LEAST DISTURBED
1030 1:LEAST DISTURBED          1:LEAST DISTURBED 1:LEAST DISTURBED
1035 1:LEAST DISTURBED           3:MOST DISTURBED 1:LEAST DISTURBED
1052 1:LEAST DISTURBED          1:LEAST DISTURBED 1:LEAST DISTURBED
1059 1:LEAST DISTURBED          1:LEAST DISTURBED 1:LEAST DISTURBED
         SALINITY_COND  xAlbers yAlbers
984  1:LEAST DISTURBED -2155207 2431464
1029 1:LEAST DISTURBED -1916837 2357249
1030 1:LEAST DISTURBED -2231092 2626452
1035 1:LEAST DISTURBED -1850930 2562015
1052 1:LEAST DISTURBED -1787635 2582877
1059 1:LEAST DISTURBED -2042558 2581004
```

3 Analysis of proportion and size for the number of lakes in each condition class category

The first analysis that will be examined is estimation of proportion and size for the number of lakes in each chemical condition class category. The chemical condition class variables that will be examined are: (1) PTL_COND, which classifies lakes by categories of total phosphorus; (2) NTL_COND, which classifies lakes by categories of total nitrogen; (3) CHLA_COND, which classifies lakes by categories of chlorophyll a; (4) TURB_COND, which classifies lakes by categories of turbidity; (5) ANC_COND, which classifies lakes by categories of acid neutralizing capacity; and (6) SALINITY_COND, which classifies lakes by categories of salinity). The cat.analysis function will be used to calculate condition class estimates. Four data frames constitute the primary input to the cat.analysis function. The first column (variable) in the four data frames provides the unique identifier (site ID) for each sample site and is used to connect records among the data frames. The SITE_ID variable in the condition data frame is assigned to the siteID variable in the data frames. The four data frames that will be created are named as follows: sites, subpop, design, and data.cat.

The sites data frame identifies sites to use in the analysis and contains two variables: (1) siteID - site ID values and (2) Use - a logical vector indicating which sites to use in the analysis. The rep (repeat) function is used to assign the value TRUE to each element of the Use variable.

Create the sites data frame.

```
> sites <- data.frame(siteID=condition$SITE_ID,
+                     Use=rep(TRUE, nrow(condition)))
```

The subpop data frame defines populations and, optionally, subpopulations for which estimates are desired. Unlike the sites and design data frames, the subpop data frame can contain an arbitrary number of columns. The first variable in the subpop data frame identifies site ID values and each subsequent variable identifies a type of population, where the variable name is used to identify type. A type variable identifies each site with a character value. If the number of unique values for a type variable is greater than one, then the set of values represent subpopulations of that type. When a type variable consists of a single unique value, then the type does not contain subpopulations. For this analysis, the subpop data frame contains two variables: (1) siteID - site ID values and (2) StateWide - which will be used to calculate estimates for all of the lakes in Oregon.

Create the subpop data frame.

```
> subpop <- data.frame(siteID=condition$SITE_ID,
+                      StateWide=rep("StateWide", nrow(condition)))
```

The design data frame consists of survey design variables. For the analysis under consideration, the design data frame contains the following variables: (1) siteID - site ID values; (2) wgt - the adjusted survey design weights; (3) xcoord - x-coordinates for location; and (4) ycoord - y-coordinates for location. The wgt, xcoord, and ycoord variables in the design data frame are assigned values using corresponding variables in the condition data frame.

Create the design data frame.

```
> design <- data.frame(siteID=condition$SITE_ID,
+                      wgt=condition$WGT_NLA,
+                      xcoord=condition$xAlbers,
+                      ycoord=condition$yAlbers)
```

Like the subpop data frame, the data.cat data frame can contain an arbitrary number of columns. The first variable in the data.cat data frame identifies site ID values and each subsequent variable identifies a response variable. The response variables are the six chemical condition class variables. Missing data (NA) is allowed for the response variables, which are the only variables in the input data frames for which NA values are allowed.

Create the data.cat data frame.

```
> data.cat <- data.frame(siteID=condition$SITE_ID,
+                        PTL_COND=condition$PTL_COND,
+                        NTL_COND=condition$NTL_COND,
+                        CHLA_COND=condition$CHLA_COND,
+                        TURB_COND=condition$TURB_COND,
+                        ANC_COND=condition$ANC_COND,
+                        SALINITY_COND=condition$SALINITY_COND)
```

Use the cat.analysis function to calculate estimates for the condition class variables.

```
> Condition_Estimates <- cat.analysis(sites, subpop, design, data.cat)
```

The object produced by cat.analysis is a data frame containing thirteen columns. The first five columns identify the population (Type), subpopulation (Subpopulation), response variable (Indicator), levels of the response variable (Category), and number of values in a category (NResp). A category labeled ”Total” is included for each combination of population, subpopulation, and response variable. The next four columns in cat.analysis provide results for the proportion estimates: the proportion estimate ((Estimate.P), standard error of the estimate (StdError.P), lower confidence bound (LCB95Pct.P), and upper confidence bound (UCB95Pct.P). Argument conf for the cat.analysis function allows control of the confidence bound level. The default value for conf is 95, hence the column names for confidence bounds contain the value 95. Supplying a different value to the conf argument will be reflected in the confidence bound names. For example, to obtain 90% confidence bounds, insert conf=90 into the argument list for the cat.analysis function. Confidence bounds are obtained using the standard error and the Normal distribution multiplier corresponding to the confidence level. The final four columns in cat.analysis provide results for the size (units) estimates: the units estimate (Estimate.U), standard error of the estimate (StdError.U), lower confidence bound (LCB95Pct.U), and upper confidence bound (UCB95Pct.U).

The write.table function is used to store the estimates as a comma-separated value (csv) file. Files in csv format can be read by programs such as Microsoft Excel.

Use the write.table function to write the condition estimates as a csv file.

```
> write.table(Condition_Estimates, file="Condition_Estimates_number.csv",
+             sep=",", row.names=FALSE)
```

4 Analysis of proportion and size for the area of lakes in each condition class category

The second analysis that will be examined is estimation of proportion and size for the area of lakes in each chemical condition class category. This type of analysis is called a size-weighted analysis, where the area for each lake is its size-weight. For this analysis, the sites, subpop, and data.cat data frames from the first analysis can be used. the design data frame must be modified to include the size-weights. Since the condition data frame does not include a variable containing lake area, the variable must be obtained from the NLA site information file. The read.delim function is used to read the site information file into a data frame named siteinfo. the match function is then used to match site ID values between the condition and siteinfo data frames. Lastly, a lake area variable is created in the condition data frame.

Read the lake information file and add a lake area variable to the condition data frame.

```
> siteinfo <- read.delim("nla_lakeinfo_sampled_20091113.txt")
> temp <- match(condition$SITE_ID, siteinfo$SITE_ID, nomatch = 0)
> condition$AREA_HA <- siteinfo$AREA_HA[temp]
```

Create the design data frame.

```
> design <- data.frame(siteID=condition$SITE_ID,
+                      wgt=condition$WGT_NLA,
```

```
+                       swgt=condition$AREA_HA,
+                       xcoord=condition$xAlbers,
+                       ycoord=condition$yAlbers)
```

Use the cat.analysis function to calculate estimates for the condition class variables.

```
> Condition_Estimates <- cat.analysis(sites, subpop, design, data.cat,
+                                     sizeweight=TRUE)
```

Use the write.table function to write the condition estimates as a csv file.

```
> write.table(Condition_Estimates, file="Condition_Estimates_area.csv", sep=",",
+             row.names=FALSE)
```

```
# File: Condition_Estimates.R
# Purpose: Calculate condition class estimates for an individual state using
#          data from the National Lakes Assessment (NLA) survey
# Programmer: Tom Kincaid
# Date: May 18, 2011


# Load the spsurvey package
library(spsurvey)

# Read the chemical variables condition class file and retain desired sites
# Note: Change condition$ADJWGT_CAT == "OR" to the desired state code
# Note: For condition class data files other than the chemical variables file
#       and the trophic state variables file, replace condition$INDXSAMP_CHEM == "YES"
#       as follows to match the data file being used:
#          condition$VISIT_NO == "1" for the dissolved oxygen variable
#          condition$VISIT_NO == "1" for physical habitat variables
#          condition$INDXSAMP_LDC  == "YES" for lake diatom variables
#          condition$INDXSAMP_OE5  == "YES" for O/E variables
#          condition$INDXSAMP_MICR   == "YES" for recreation variables
#          condition$INDXSAMP_INF  == "YES" for diatom-inferred chemistry variables
condition <- read.delim("nla_chem_cond_20091123.txt")
keep <- condition$ADJWGT_CAT == "OR" &
        condition$INDXSAMP_CHEM == "YES" &
        condition$SITE_TYPE == "PROB_Lake" &
        condition$LAKE_SAMP == "Target_Sampled"
condition <- condition[keep,]

# Convert latitude and longitude to the Albers projection
temp <- geodalbers(condition$LON_DD, condition$LAT_DD, sph="GRS80")
condition$xAlbers <- temp$xcoord
condition$yAlbers <- temp$ycoord

# Display the initial six lines in the chemical condition class data frame
head(condition)

# Display the final six lines in the chemical condition class data frame
tail(condition)

#
# Conduct an analysis of chemical condition class for number of lakes
#

# Create the sites data frame, which identifies sites to use in the analysis
sites <- data.frame(siteID=condition$SITE_ID,
                    Use=rep(TRUE, nrow(condition)))

# Create the subpop data frame, which defines populations and subpopulations for
# which estimates are desired
subpop <- data.frame(siteID=condition$SITE_ID,
                     StateWide=rep("StateWide", nrow(condition)))

# Create the design data frame, which identifies the stratum code, weight,
#    x-coordinate, and y-coordinate for each site ID
design <- data.frame(siteID=condition$SITE_ID,
                     wgt=condition$WGT_NLA,
                     xcoord=condition$xAlbers,
                     ycoord=condition$yAlbers)

# Create the data.cat data frame, which specifies the variables to use in the
```

```
# analysis
data.cat <- data.frame(siteID=condition$SITE_ID,
                       PTL_COND=condition$PTL_COND,
                       NTL_COND=condition$NTL_COND,
                       CHLA_COND=condition$CHLA_COND,
                       TURB_COND=condition$TURB_COND,
                       ANC_COND=condition$ANC_COND,
                       SALINITY_COND=condition$SALINITY_COND)

# Calculate estimates for the condition class variables
Condition_Estimates <- cat.analysis(sites, subpop, design, data.cat)

# Write results as a comma-separated value (csv) file
write.table(Condition_Estimates, file="Condition_Estimates_number.csv", sep=",",
            row.names=FALSE)

#
# Conduct an analysis of chemical condition class for area of lakes
#

# Read the lake information file and add a lake area variable to the condition
# data frame
siteinfo <- read.delim("nla_lakeinfo_sampled_20091113.txt")
temp <- match(condition$SITE_ID, siteinfo$SITE_ID, nomatch=0)
condition$AREA_HA <- siteinfo$AREA_HA[temp]

# Create the design data frame
# Note that the existing sites, subpop, and data.cat data frames can be reused
design <- data.frame(siteID=condition$SITE_ID,
                     wgt=condition$WGT_NLA,
                     swgt=condition$AREA_HA,
                     xcoord=condition$xAlbers,
                     ycoord=condition$yAlbers)


# Calculate estimates for the condition class variables
Condition_Estimates <- cat.analysis(sites, subpop, design, data.cat,
                                    sizeweight=TRUE)

# Write results as a csv file
write.table(Condition_Estimates, file="Condition_Estimates_area.csv", sep=",",
            row.names=FALSE)
```