

---

## Itasca County Lake Data Compilation and Clean-up

### Final Report

---

To: Itasca Soil and Water Conservation District and Itasca Water Legacy Partnership  
From: Moriya Rufer, RMB Environmental Laboratories, Inc  
Date: February 8, 2018  
Regarding: Itasca County Lake Data Compilation and Clean Up

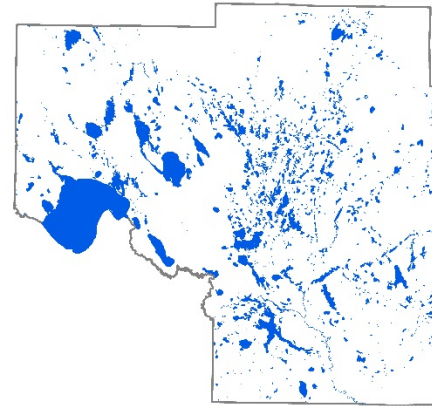
---

### Project Background

Itasca County has a long history of lake monitoring. With over 1,000 lakes in the county and some of the best water quality in the nation, it is important to continually monitor the water quality of these lakes to make sure they are protected for future generations.

In 2015, Itasca County Soil and Water Conservation District (SWCD) began working with RMB Environmental Laboratories (RMB) to update their county water plan. While using the public database housed at the Minnesota Pollution Control Agency (MPCA), they discovered that not all the data collected on Itasca County Lakes had made its way to the state database (EQuIS). At the local level, lake associations knew of other data or other years that data had been collected. Driven by local concern, lake associations had created their own volunteer monitoring programs and used different labs throughout the state to process the samples. Some of that data had stayed in files kept locally.

The Itasca Water Legacy Partnership (IWLP) sponsored a project with RMB, who also donated staff hours to the project, to contact all the lake associations and compile the missing data. The end product was a one-page report for each lake summarizing the data and water quality characteristics of that lake. In addition, accumulated data, which met metadata requirements, were able to be submitted to the MPCA for EQuIS. The details for compiled data sets are listed in the Results section of this report.



*Figure 1. Itasca County lakes.*

## Methods

The project began by receiving a download of all the Itasca County lake data from the MPCA database, EQUIS. Then a call for lake data was disseminated within Itasca County through IWLP, the Itasca County Coalition of Lake Associations (ICOLA), the SWCD, and the local newspaper (*Grand Rapids Herald Review*). RMB received data in many different formats including paper hard copy, *Excel* worksheets, and *Word* files. These data were all compiled into one large database called RMB\_raw (Figure 2, Table 1).

As data was compiled, it became evident that there were many anomalies from all the different data sources, as well as records that had missing information, so a process was designed in an attempt to clean up the compiled data. Figure 2, Table 1 describes the process used to best accomplish the task of bringing all available Itasca County data together, document anomalies, and provide an end product made available to the state for future assessments.

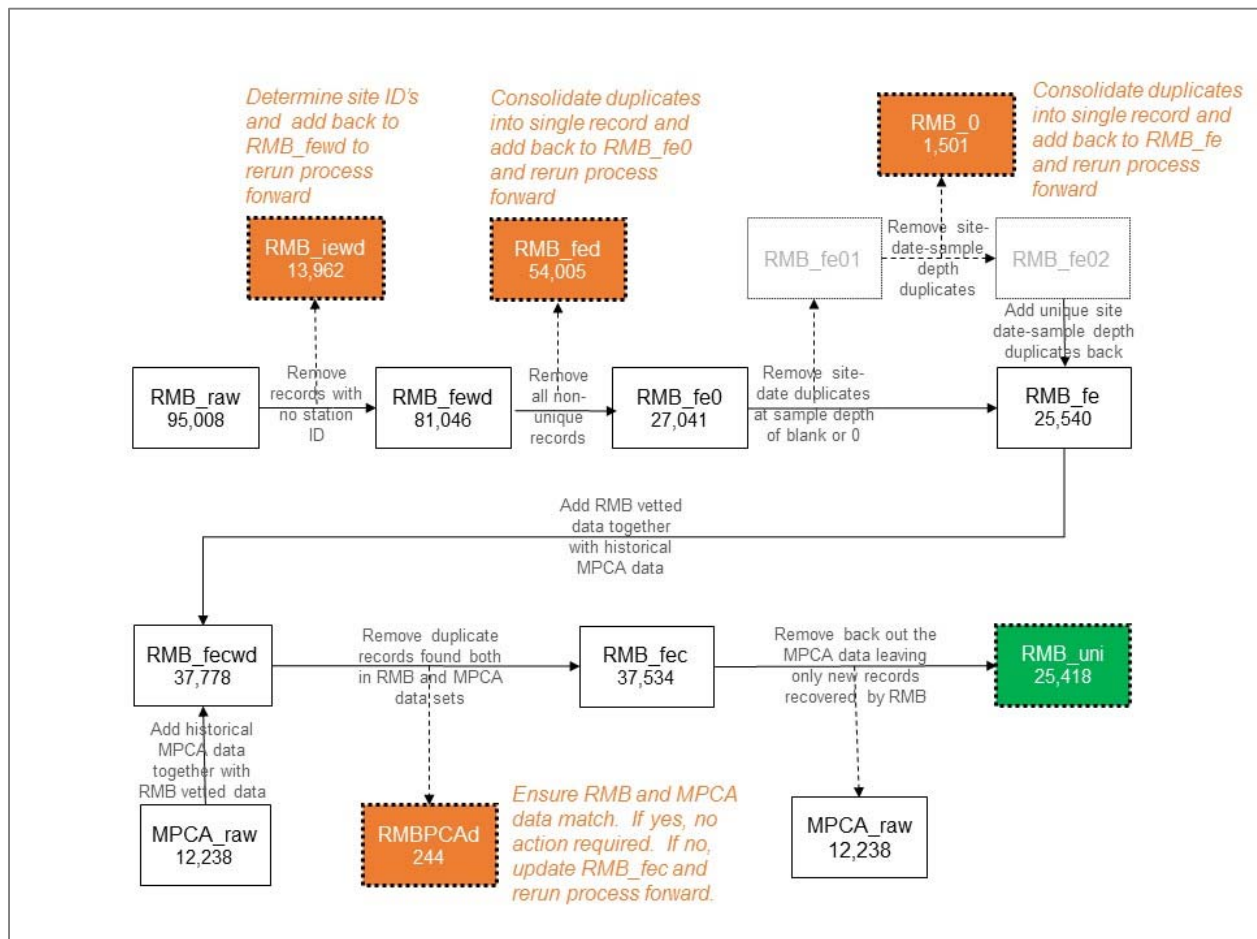


Figure 2. Data clean-up process for this project.

Table 1. Explanation of all the files and processes in Figure 1. “RMB” data means data that was compiled by RMB Staff from numerous sources including lake associations and the SWCD.

<b>RMB_raw</b>	All data collected and combined by RMBEL staff that would later be compared to data already in existence with the MPCA. 95,008 records
<b>RMB_iewd</b>	RMBEL collected data that have an <i>incomplete EQuIS with duplicates</i> (iewd). For these data to be of further use, duplicates need to be removed and the EQuIS_Location_ID needs to be expanded to include the site number. Most of these only contain a lake number but no site number. 13,962 records
<b>RMB_fewd</b>	RMBEL collected data that have <i>full EQuIS with duplicates</i> (fewd). Records with incomplete EQuIS_Location_IDs have been deleted and stored in RMB_iewd, leaving records only with complete EQuIS_Location_ID numbers. 81,046 records
<b>RMB_fed</b>	RMBEL collected data that are <i>full EQuIS duplicates</i> (fed). These data have full EQuIS station ID’s but contain duplicate records for collections on same site, same day, and same sampling depth. For these data to be of further use, all duplicate X_SampleID values need to be consolidated to a single record. It will require manually moving and copying data from multiple incomplete duplicate records to make one complete record. 54,005 records
<b>RMB_fe0</b>	RMBEL collected data that have <i>full EQuIS 0</i> (fe0). The zero-blank problem occurs where some records have “0” to represent the Sample_Depth_Upper_m but other samples taken at the same site on the same day have no value (i.e., blank) but likely represents “0”. This creates extra duplicates that are not “seen” using a duplicate search on site-day-depth identifiers. To ID these problem samples, we did a duplicate search only on site-day and created an identifier in the table called XSample_ID2. Because some of the legitimate records would show up as duplicates given they were part of a depth profile we could not just delete them all. We were only interested in extracting the marked duplicates that had a Sample_Depth_Upper_m = blank or zero. Subsequently, of all of the identified duplicates we filtered only for Sample_Depth_Upper_m = blank or zero and held these RMB_fe01. We repeated the process to find duplicates in RMB_fe01 and saved those in a separate table called RMB_0. What remained in RMB_fe02 were unique records that were then added back to RMB_fe0 to create a new table called RMB_fe. Tables RMB_fe01 and RMB_fe02 were intermediates and not saved. Records = 27,041
<b>RMB_fe</b>	RMBEL collected data that have <i>full EQuIS</i> (fe). The table represents the cleaned up version after removing duplicates and records without station IDs. All unique records and full EQuIS station IDs are listed here. This table can then be compared to the historical MPCA data to look for further matches. Records = 25,540
<b>RMB_0</b>	RMBEL collected data that includes the zero-blank problems (0). For these data to be of further use, each duplicate pair of records that are matched will need to be consolidated into a single record. Records = 1,501
<b>MPCA_raw</b>	MPCA collected data that were reformatted to compare directly to the RMBEL collected data. A preliminary screen showed no duplicates or missing station sites internally. The data were used “as is”. Records = 53,502
<b>RMB_fecwd</b>	RMBEL collected data that have <i>full EQuIS combined with duplicates</i> (fecwd). MPCA_raw was appended to RMB_fe. Any duplicates between MPCA and RMB remain. Records = 79,042
<b>RMBPCAd</b>	Duplicate values found between RMBEL and MPCA historical data. For these data to be of further use, the duplicate records between MPCA and RMBEL need to be compared. If the data are identical, no action is required. If consolidation is required, the “Source” should be changed to RMB and added to RMB_fec. Records = 3,475
<b>RMB_fec</b>	RMBEL collected data that have full EQuIS combined with MPCA data. All duplicate data between RMBEL and MPCA have been extracted. Records = 75,567
<b>MPCA_uni</b>	MPCA unique (uni) records. This table contains the raw MPCA data after removing data that were duplicates with RMB collected records. Records = 50,950
<b>RMB_uni</b>	RMBEL unique (uni) records. This table is the culmination of the project and represents new data that were obtained by RMBEL not previously part of the MPCA database. Records = 24,617

One of the original goals of this project was to submit the additional compiled data to the MPCA's EQuIS database. In order for data to be accepted by the MPCA, it needs a valid sample site location ID, sample depth metadata, sample date and time records, and laboratory metadata. Through the data compilation process, numerous issues were discovered, including missing site IDs, missing lab metadata, and data quality issues. Many of these issues are specified in Appendix II.

The data that had these issues will take additional time to sort out and are therefore outside the scope of the original project.

## Results

This project resulted in numerous data sets for Itasca County. The details about the current state of these data sets are spelled out below. Some items to note as far as data quality include reporting limits. Many of the Itasca County lakes are oligotrophic, meaning they have very low phosphorus and chlorophyll *a* concentrations. Some of the labs used in the past had reporting limits of 5 ug/L for chlorophyll *a*, which lead to many <5 ug/L results that cannot be used for statistical analysis. Past reporting limits for phosphorus were as high as 10 ug/L, which lead to many <10 ug/L results that cannot be used for statistical analysis. Having such a vast number of different data sources (laboratory service providers and project coordinators) over the years (Appendix I) creates datasets which prove difficult to perform long-term statistical analysis and assessments due to variables including monitoring program design, laboratory and sampling procedures, analytical methodology, reporting formats, reporting limits, associated metadata or lack thereof.

### **The final deliverable products from this project include:**

- One-page Lake Fact Sheets for 72 lakes.
- RMB\_uni.xlsx spreadsheet. This spreadsheet contains the Itasca County data that is not already in the MPCA database. It does still include anomalies mentioned in Appendix II. Data anomalies are marked in yellow.
- Full Itasca County dataset (includes MPCA data). This spreadsheet does still include anomalies mentioned in Appendix II.

### **Additional deliverable products that were part of the 2016-2018 SWCD BWSR grant but used the data compiled in this data compilation project include:**

- Complete lake reports with trend analysis for 72 lakes (Appendix III), delivered to Itasca SWCD for the Itasca County Water Plan.
- Lake Prioritization and Protection Planning Document summarizing all the data and prioritizing the 72 lakes for the Itasca County Water Plan.

**Data sets submitted to MPCA for data review and entry into EQUIS include:**

1. These data sets had all necessary information for submittal. The MPCA had the metadata for ICC Lab on hand from those years.

- Wabana Chain of Lakes: 2010-2011, 2014; ICC Lab.

2. These datasets have all information for submittal except for valid site location. The MPCA will try and determine if they have past records of these sites, otherwise they would be stored as a generic site “100”. Or the samplers could be contacted for sample location information.

- Pokegama Lake: 2007-2008; Era Lab.
- Deer Lake: 1992-1993, 2001-2002, 2004-2009; NRRI Lab
- Horseshoe Lake: 2001, 2003-2007, 2010; NRRI Lab

3. These data sets that are intact, but lab metadata is needed. The MPCA will determine if they have it in their historical records or if the corresponding lab needs to be contacted to submit the data.

- Deer Lake, 2010-2014; A.W. Research Labs
- WCOLA, 2005 and 2008; Instrumental Research
- WCOLA, 1991, 1999, and 2003; Minnesota Chippewa Tribe Lab
- Jessie Lake: 2003 (missing sample times); NRRI
- Numerous lakes: 2010 Dissolved Organic Carbon (DOC) results; Iowa State University Limnology Lab.
- Numerous lakes: 2010 Alkalinity and DOC results; Northeast Technical Services, Inc.

## Recommendations and Next Steps

If this data compilation and cleanup project was to be continued, the following items could be completed to further this project.

1. If the MPCA is not able to locate the lab metadata and site information for the lakes listed on page 5:
  - a. work directly with the labs listed to acquire the metadata.
  - b. work directly with the lake volunteers to determine the location that they collected their lake samples.
2. Go through the “RMB\_uni” data spreadsheet and address each of the issues highlighted in yellow and noted in Appendix II. At times, this will require referring back to the original data source or laboratory to verify correct numbers.
3. Quality check SWCD field data for submittal to the MPCA. Results need to be verified (see Appendix II for specific details).
4. Quality check USFS field data for submittal to the MPCA. Results need to be verified (see Appendix II for specific details).

## Appendix I. Itasca County Monitoring Programs and Associated Laboratories

MPCA Project Code	Project Name	Laboratory Name
PRJ00063	DNR Fisheries Routine Water Quality Database	Northeast Technical Services
PRJ00064	DNR Shallow Lakes Monitoring Program	Minnesota Department of Agriculture Lab
PRJ00259	Itasca County Lake Model Calibration Data Set	Natural Resources Research Institute, Minnesota Department of Agriculture
PRJ00260	Itasca County Lake Assessment (includes Jessie L CWP)	Minnesota Department of Health, Natural Resources Research Institute, RMB Environmental Laboratories
PRJ00261	Bigfork River Watershed Priority Waters Assessment	Itasca Community College Water Quality Laboratory, Pace Analytical Services - Virginia MN, Iowa State University Limnology Laboratory, University of Missouri Limnology Laboratory
PRJ00270	Lake Monitoring Program	Minnesota Department of Health, Minnesota Valley Testing Laboratories
PRJ00272	Lake Trend Monitoring	Minnesota Department of Health
PRJ00273	Lake Assessment Projects	Minnesota Department of Health
PRJ00277	Atmospheric and Nonpoint Trends in MN Lakes LCMR study	Minnesota Department of Health, Natural Resources Research Institute
PRJ00293	MPCA Lake Monitoring Program Project	Minnesota Department of Health, Minnesota Valley Testing Laboratories, U.S. Geological Survey Mercury Research Laboratory, Montgomery Watson Harza (formerly JMM), Barr Engineering Laboratory, Interpoll Laboratories, A.W. Research Laboratories, Eco Agri Laboratories, RMB Environmental Laboratories, Magic Water Testing Lab, Natural Resources Research Institute, Albert Lea Technical College Lab, Eau Galle Aquatic Ecology Lab (Limnology, Army Corps of Eng), Microbiologics (formerly Environmental Protection Laboratory), U of MN St. Paul campus Soil Testing & Research Analytical
PRJ00338	Assess of Chemicals of Potential Concern in MN Lk Sed & Fish	Minnesota Department of Health
PRJ00350	Mercury Trends	Minnesota Department of Health, Frontier Geosciences
PRJ00415	National Lake Assessment	Minnesota Department of Health
PRJ01067	1st WQ Inventory of Itasca County Lakes. 2010: Assessment Itasca County's Small-Medium Sized Lakes	Northeast Technical Services, Natural Resources Research Institute, Itasca Community College Water Quality Laboratory, Pace Analytical Services - Virginia, MN, Iowa State University Limnology Laboratory, University of Missouri Limnology Laboratory
PRJ01961	Pokegama Lake Watershed Project	U.S. Geological Survey Mercury Research Laboratory
PRJ05432	Jessie Lake Watershed TMDL Project	Natural Resources Research Institute
PRJ07081	Clean Water Legacy Surface Water Monitoring	Minnesota Department of Health, RMB Environmental Laboratories
PRJ07138	Citizen Lake Monitoring Program	Minnesota Department of Health, Minnesota Valley Testing Laboratories, U.S. Geological Survey Mercury Research Laboratory
PRJ07181	Deer Creek and Pokegama Lakes: A Diagnostic Study	Pace Analytical - Frontenac KS, Itasca Community College Water Quality Laboratory, Pace Analytical Services - Virginia MN

PRJ07269	PCA Special Studies: lakes in outstate MN	AXYS Analytical Services
PRJ07296	RMB Environmental Laboratory Monitoring Program	RMB Environmental Laboratories
PRJ07375	Itasca County Mississippi River Grand Rapids SWAG	Pace Analytical - Frontenac KS, RMB Environmental Laboratories, Itasca Community College Water Quality Laboratory, Pace Analytical Services - Virginia, MN
PRJ07840	Wild Rice - Sulfate Project	GUSTAVUS, Minnesota Department of Health, U of MN St. Paul campus Soil Testing & Research Analytical, SMM
PRJ07911	Wild Rice and Sulfate Assessment	U.S. Geological Survey Mercury Research Laboratory, Braun Intertec Corporation (lab), Minnesota Department of Agriculture, Northeast Technical Services, Legend Technical Services, RMB Environmental Laboratories, Pace Analytical Services - Virginia, MN
PRJ07916	Leech Lake Band of Ojibwa	U.S. Geological Survey Mercury Research Laboratory
PRJ07938	A.W. Research Summer TSI	Minnesota Valley Testing Laboratories, A.W. Research Laboratories



## Appendix II. Data Anomalies

Below is a summary of some of the data issues/anomalies found in the Itasca County data set. This list includes those issues that were noticeable from a few days of data review and appear to be routine issues in some instances. It is not a complete exhaustive list. This list could be used as a guide for further data clean-up if desired. These additional clean-up efforts would require line by line review, noting the original raw data set contained 95,008 records. This initial review attempted to automate the screening process (Figure 1, Table 1). Additional data clean-up would require significant staff hours and additional financial support.

1. There are duplicate dates at certain lakes where one Chl-a value is exactly half of the other, usually 5 ug/L and 2.5 ug/L. These samples with unusual Chl-a values are from 2010, which include: Three Island, Little Wabana, Amen, Boy, Burrows, L. Long, White Swan, and Moose. As a note, these Chl-a duplicate values do not have a correlation with the TP duplicates. These values are from the Itasca Community College Laboratory, and we think that the reporting limit for Chl-a was 5, so the values should be <5 ug/L.
2. For Moose, 1992 Chl-a and TP are very unusual and don't seem to match up. Also these data do not have a site number. These data were not used in the reports.
3. The smaller of the duplicate values in 2010 Chl-a duplicates for East Smith and Trestle are linked to much higher TP values.
4. Guile Lake, 6/18/1998, there are 3 different CHLA results on the same day at the same site. All 3 results are slightly different. The middle one was used in the reports.
5. Moose Lake (31-0722-00-102) the Phosphorus ppb result doesn't match the Phosphorus ppm result.
6. In 2001, the phosphorus data for Clearwater (31-0214-00) is much higher than other years and could be inaccurate.
7. Little Flower Lake Specific Conductance values are in the 1,000s and field turbidity is negative numbers on 9/3/2009, which doesn't look right.
8. Buck Lake samples on 2/25/2010 it looks like the pH and Conductivity are switched.
9. Clear Lake (31-0209-00) samples in 2006 it looks like pH and Conductivity are switched.
10. Sand Lake (31-0826-00) samples in 2002 it looks like pH and Conductivity are switched.
11. Trout (31-0216-00), and Snaptail have dissolved oxygen readings that do not look accurate (for example 90 mg/L). They are highlighted in yellow.
12. Loon (31-0571-00) and Crooked (31-0809-00) lakes have negative dissolved oxygen results. They are highlighted in yellow in the spreadsheet.
13. Numerous samples are missing sample times. They are highlighted in yellow in the spreadsheet.
14. Buck Lake and Sand Lake 2006 samples from Bigfork River Project have phosphorus results of -99, which doesn't look right. Lab was NRRI.
15. Little Fowler Lake 10/14/2004 and Moss Lake 7/12/2006 have negative chlorophyll a results (-99 and -33 respectively).

16. Field turbidity is negative for many records, highlighted in yellow.
17. Many of the records in the column “TP\_ppm” look like they should be in ppb (ug/l). They are highlighted in yellow.
18. Ice Lake (31-0372-00), 9/21/2005, the Nitrate + Nitrite result is -99.
19. Crum Lake in 1989-1998 it looks like the “TP\_ppb” and “TN\_ppb” should be ppm (mg/L). They are highlighted in yellow.
20. Reporting limit issues – reporting limits from old data appear to be 5 ug/L for Chlorophyll-a and either 10 or 11 ug/L for total phosphorus. These are not low enough to truly reflect Itasca County’s water quality since many lakes are oligotrophic and below those levels.

## Appendix III. Lakes that had reports generated for them for the county water plan.

Table 2. Lakes assessed in the 2015-2016 lakes assessments.

Lake Name	Lake ID	Lake Size (acres)	Lake Name	Lake ID	Lake Size (acres)
Adele	31-0642-00	22	Napoleon	31-0290-00	138
Amen	31-0597-00	215	Natures	31-0877-00	2,250
Balsam	31-0259-00	714	North Star	31-0653-00	821
Bass	31-0576-00	2,765	Pickerel	31-0339-00	241
Battle	31-0197-00	243	Pokegama	31-0532-00	1,123
Beatrice	31-0058-00	124	Rice	31-0717-00	863
Beaver	31-0638-00	13	Round (Clear)	31-0209-00	126
Bello	31-0726-00	530	Round	31-0896-00	2,860
Bluewater	31-0395-00	359	Sand (0438)	31-0438-00	195
Boy	31-0623-00	43	Sand (0826)	31-0826-00	4,225
Buck	31-0069-00	495	Scraper	31-0345-00	172
Burnt Shanty	31-0424-00	198	Shallow	31-0084-00	539
Burrows	31-0413-00	306	Siseebakwet	31-0554-00	1,210
Caribou	31-0620-00	247	Snaptail	31-0255-00	177
Clearwater (Round)	31-0214-00	132	South Sugar	31-0555-00	91
Crum	31-0171-00	19	Swan	31-0067-00	2,116
Cut Foot Sioux	31-0857-01	2,378	Three Island	31-0542-00	250
Deer (0334)	31-0334-00	1,853	Trestle	31-0127-00	88
Deer (0719)	31-0719-00	4,163	Trout	31-0410-00	1,736
Dixon	31-0921-00	622	Trout	31-0216-00	1862
Dora	31-0882-00	430	Turtle	31-0725-00	2,156
Dunning	31-0221-00	67	Wabana	31-0392-00	2,221
Eagle	31-0454-00	285	White Swan	31-0260-00	165
East Smith	31-0616-00	152	Winnibigoshish	11-0147-00	53,425
Erskine	31-0311-00	40			
Fifth Chain	31-0497-00	104			
Five Island	31-0183-00	214			
Grave	31-0624-00	525			
Guile	31-0569-00	88			
Gum	31-0492-00	32			
Gunn	31-0452-00	108			
Hale	31-0361-00	126			
Hale	31-0373-00	130			
Hart	31-0020-00	328			
Horseshoe	31-0696-00	260			
Island	31-0913-00	3,108			
Jack the Horse	31-0657-02	260			
Jessie	31-0786-00	1,740			
Little Bowstring	31-0758-00	327			
Little Dead Horse	31-0621-00	79			
Little Jessie	31-0784-00	628			
Little Long	31-0613-00	305			
Little Trout	31-0394-00	86			
Little Wabana	31-0399-00	116			
Loon	31-0571-00	231			
Maki	31-0759-00	16			
McGuire	31-0078-00	79			
Moose	31-0722-00	1,274			

## Appendix IV. In-kind time donation to this project from RMB Environmental Laboratories, Inc.

In all, RMB Staff spent 428 hours on this lake data compilation and cleanup project. At our normal billable rates for each staff member who worked on the project, the total value of time was \$19,751. This time was in addition to the contracted deliverables of the lake reports for the county water plan as part of the 2016-2018 SWCD BWSR grant. IWLP direct sponsored \$2,000 towards RMB's efforts in the data compilation and cleanup project. A remaining portion of these data compilation and cleanup hours were donated by RMB to Itasca SWCD as in-kind time towards the 2016-2018 SWCD BWSR grant project.



Moriya Rufer  
Environmental Services Director  
RMB Environmental Laboratories, Inc  
22796 County Highway 6  
Detroit Lakes, MN 56501  
218-846-1465  
[Moriya.Rufer@rmbel.info](mailto:Moriya.Rufer@rmbel.info)  
<https://www.rmbel.info/>