

National Coastal Assessment 2010

Great Lakes Survey Design

Description of Sample Design

Target population: Near shore waters of the Great Lakes of the United States and Canada. Near shore zone is defined as region from shoreline to 30m depth constrained to a maximum of 5 km from shoreline. Great Lakes include Lake Superior, Lake Michigan, Lake Huron, Lake Erie, and Lake Ontario. The NARS Great Lakes survey will be restricted to the United States portion.

Sample Frame: The sample frame was obtained from Jack Kelly.

Survey Design: A Generalized Random Tessellation Stratified (GRTS) survey design for an area resource is used. The survey design is stratified by Lake and country with unequal probability of selection based on state shoreline length within each stratum.

Stratification: Stratification is based on Great Lake and country.

Multi-density categories: Unequal probability categories are states or province within each Great Lake based on proportion of state shoreline length within each stratum.

Panels: Design uses three panels: Revisit-sites are visited twice during the study, Base-sites are visited once during the study, OverSamp-over sample sites to be used if a Revisit of Base site can not be sampled. The over sample sites were selected independent of the Base and Revisit panels.

Expected sample size: Expected sample size of 45 sites in Shallow NearShore zone for each Great Lake and country combination for a total of 405 sites. Sample sizes were allocated proportional to shoreline length by state within each Great Lake.

Over sample: Over sample size of 405 (100%) was selected to provide replacement sites that either are not part of the target population or could not be sampled. The over sample sites were selected independently of the base design.

Site Use: Sites should be used in SiteID order within each state within each Great Lake. For example, if site NCAGL10-1001 (Lake Superior site in Minnesota) can not be sampled, then the first over sample site (NCAGL10-2002) for Lake Superior within Minnesota should be used. If a Revisit site can not be sampled, then the next Base site within the same Great Lake that is also within the same state will be the revisit site. For example, since site NCAGL10-1001 is a Revisit site and can not be sampled, the next Base site in Lake Superior that is in Minnesota will be revisited (site NCAGL10-1011).

Sample Frame Summary

	Lake Erie	Lake Huron	Lake Michigan	Lake Ontario	Lake Superior	Sum
Illinois	0.0	0.0	466.5	0.0	0.0	466.5
Indiana	0.0	0.0	325.9	0.0	0.0	325.9
Michigan	274.2	3448.5	4464.8	0.0	2315.4	10502.8
Minnesota	0.0	0.0	0.0	0.0	268.3	268.3
New York	541.3	0.0	0.0	1359.7	0.0	1901.0
Ohio	1653.9	0.0	0.0	0.0	0.0	1653.9
Ontario	2515.7	4609.6	0.0	2110.5	2091.3	11327.1
Pennsylvania	394.6	0.0	0.0	0.0	0.0	394.6
Wisconsin	0.0	0.0	2611.3	0.0	756.2	3367.5
Sum	5379.7	8058.1	7868.4	3470.2	5431.2	30207.5

Site Selection Summary

Number of sites by state/province for base sample:

	Lake Erie	Lake Huron	Lake Michigan	Lake Ontario	Lake Superior	Sum
Illinois	0	0	3	0	0	3
Indiana	0	0	2	0	0	2
Michigan	6	45	26	0	30	107
Minnesota	0	0	0	0	7	7
New York	10	0	0	45	0	55
Ohio	25	0	0	0	0	25
Ontario	45	45	0	45	45	180
Pennsylvania	4	0	0	0	0	4
Wisconsin	0	0	14	0	8	22
Sum	90	90	45	90	90	405

Description of Sample Design Output:

The output is provided as a shapefile for the sites. Note that the ".dbf" file may be read in Excel. The attributes are as follows:

Variable Name	Description
SiteID	Unique site identification (character)
LON_DD	Longitude in decimal degrees (NAD 1983)
LAT_DD	Latitude in decimal degrees (NAD 1983)
xcoord	x-coordinate from map projection (see below)
ycoord	y-coordinate from map projection (see below)
mdcaty	Multi-density categories used for unequal probability selection
weight	Weight (in sq km), inverse of inclusion probability, to be used in statistical analyses
stratum	Strata used in the survey design
panel	Identifies base sample by panel name and Oversample by OverSamp
EvalStatus	Site evaluation decision for site: TS: target and sampled, LD: landowner denied access, etc (see below)
EvalReason	Site evaluation text comment
auxiliary variables	Remaining columns are from the sample frame provided

Projection Information

ROJCS["NAD_1983_Albers",
 GEOGCS["GCS_North_American_1983",
 DATUM["D_North_American_1983",
 SPHEROID["GRS_1980",6378137.0,298.257222101]],

```

PRIMEM["Greenwich",0.0],
UNIT["Degree",0.0174532925199433]],
PROJECTION["Albers"],
PARAMETER["False_Easting",0.0],
PARAMETER["False_Northing",0.0],
PARAMETER["Central_Meridian",-96.0],
PARAMETER["Standard_Parallel_1",29.5],
PARAMETER["Standard_Parallel_2",45.5],
PARAMETER["Latitude_Of_Origin",23.0],
UNIT["Meter",1.0]]

```

Evaluation Process

The survey design weights that are given in the design file assume that the survey design is implemented as designed. Typically, users prefer to replace sites that can not be sampled with other sites to achieve the sample size planned. The site replacement process is described above. When sites are replaced, the survey design weights are no longer correct and must be adjusted. The weight adjustment requires knowing what happened to each site in the base design and the over sample sites. EvalStatus is initially set to "NotEval" to indicate that the site has yet to be evaluated for sampling. When a site is evaluated for sampling, then the EvalStatus for the site must be changed.

Recommended codes are:

EvalStatus Code	Name	Meaning
TS	Target Sampled	site is a member of the target population and was sampled
LD	Landowner Denial	landowner denied access to the site
PB	Physical Barrier	physical barrier prevented access to the site
NT	Non-Target	site is not a member of the target population
NN	Not Needed	site is a member of the over sample and was not evaluated for sampling
Other codes		Many times useful to have other codes. For example, rather than use NT, may use specific codes indicating why the site was non-target.

Statistical Analysis

Any statistical analysis of data must incorporate information about the monitoring survey design. In particular, when estimates of characteristics for the entire target population are computed, the statistical analysis must account for any stratification or unequal probability selection in the design. Procedures for doing this are available from the Aquatic Resource Monitoring web page given in the bibliography. A statistical analysis library of functions is available from the web page to do common population estimates in the statistical software environment R.

For further information, contact

Anthony (Tony) R. Olsen

USEPA NHEERL
Western Ecology Division
200 S.W. 35th Street
Corvallis, OR 97333
Voice: (541) 754-4790
Fax: (541) 754-4716
email: Olsen.Tony@epa.gov

Bibliography:

Diaz-Ramos, S., D. L. Stevens, Jr, and A. R. Olsen. 1996. EMAP Statistical Methods Manual. EPA/620/R-96/002, U.S. Environmental Protection Agency, Office of Research and Development, NHEERL-Western Ecology Division, Corvallis, Oregon.

Stevens, D.L., Jr. 1997. Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics*, 8:167-95.

Stevens, D.L., Jr. and Olsen, A.R. 1999. Spatially restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics*, 4:415-428

Stevens, D. L., Jr., and A. R. Olsen. 2003. Variance estimation for spatially balanced samples of environmental resources. *Environmetrics* **14**:593-610.

Stevens, D. L., Jr., and A. R. Olsen. 2004. Spatially-balanced sampling of natural resources in the presence of frame imperfections. *Journal of American Statistical Association*:99:262-278.

Horn, C.R. and Grayman, W.M. (1993) Water-quality modeling with EPA reach file system. *Journal of Water Resources Planning and Management*, 119, 262-74.

Strahler, A.N. 1957. Quantitative Analysis of Watershed Geomorphology. *Trans. Am. Geophys. Un.* 38,913-920.

Web Page: <http://www.epa.gov/nheerl/arm>