National Lake Assessment 2012
Survey Design

Target population

All lakes, reservoirs, and ponds within the 48 contiguous United States greater than 1 hectare in surface area that are permanent waterbodies. The word “lake” in the remainder of this document includes lakes, reservoirs and ponds. Lakes that are saline are excluded as are those used for aquaculture, disposal-tailings, sewage treatment, evaporation, or other unspecified disposal use.

Sample Frame

The sample frame was derived from the National Hydrography Dataset (NHD).

Once the initial shapefile that included all lake objects in NHD was prepared additional attributes were created to identify lakes included in the sample frame and other properties used to construct the survey design.

Lakes included in the sample frame were those lakes with DES_FYTPE values equal to:

- Lake/Pond
- Lake/Pond: Hydrographic Category = Perennial
- Lake/Pond: Hydrographic Category = Perennial; Stage = Average WaterElevation
- Lake/Pond: Hydrographic Category = Perennial; Stage = Normal Pool
- Reservoir
- Reservoir: Reservoir Type = Water Storage
- Reservoir: Reservoir Type = Water Storage; Hydrographic Category = Perennial

Lakes excluded in the sample frame were those lakes with DES_FYTPE values equal to:

- Lake/Pond: Hydrographic Category = Intermittent
- Lake/Pond: Hydrographic Category = Intermittent; Stage = Date of Photography
- Lake/Pond: Hydrographic Category = Intermittent; Stage = High Water Elevation
- Playa
- Reservoir: Reservoir Type = Aquaculture
- Reservoir: Reservoir Type = Cooling Pond
- Reservoir: Reservoir Type = Disposal
- Reservoir: Reservoir Type = Evaporator
- Reservoir: Reservoir Type = Tailings Pond
- Reservoir; Reservoir Type = Treatment
- Swamp/Marsh

Next lakes were excluded that were evaluated during the NLA 2007 and were identified as lakes that did not meet definition of a lake for NLA 2012. These were lakes with evaluation codes of Lake_Saline, Lake_Shallow, Lake_Special_Purpose, Lake_Vegetated, Non_Target, or Not_Lake".

Finally, lakes that were less than or equal to 1 hectare were excluded.

Survey Design

A Generalized Random Tessellation Stratified (GRTS) survey design for a finite resource was used with stratification and unequal probability of selection. The design includes reverse hierarchical ordering of the selected lakes.

Stratification

The survey design was stratified by state and by NLA12_CLS. NLA12_CLS has three classes: (1) NLA07RVT – defined as all NLA 2007 lakes that were target and sampled, (2) NLA12NEW – remaining lakes in NHD-Plus that are included in the sample frame, and (3) Exclude – lakes in NHD-Plus that are excluded from the sample frame (see Sample Frame section above). Each state design has two strata, ST_ NLA07RVT and ST_ NLA12NEW (where ST is replaced by two letter state abbreviation. The total number of strata is 96 (two for each state).

Unequal Probability Categories

The 48 state strata for lakes from the NLA 2007 that would be visited again in 2012 was an equal probability design within each stratum. The 48 state strata NLA12NEW was an unequal probability design within each state stratum. The unequal probability categories were defined based on lake area: 1 to 4 ha, 4 to 10 ha, 10 to 20 ha, 20 to 50 ha and greater than 50 ha.

Panels

The survey design has four panels: NLA07RVT – identifies lakes from NLA 2007 that will be visited in 2012, NLA12NAT – identifies new lakes that will be sampled along the lakes in panel NLA07RVT as part of the NLA2012 national survey design, NLA12ST – identifies additional lakes that a state may sample to achieve a total sample size of 50 lakes for the state, and OverSamp – identifies lakes to be used to replace lakes that cannot be sampled for some reason (not a lake, denied access, physically inaccessible, etc).

The national survey design includes all lakes within a state that are in either panels NLA07RVT or NLA12NEW.

A state survey design includes all lakes within a state that are either in panels NLA07RVT, NLA12NEW or NLA12ST.

Expected Sample Size

The expected sample size depends on the strata, panels and lake area category. For the NLA07RVT strata, the objective was to resample 400 of the NLA 2007 lakes out of the 1028 lakes that were sampled in 2007, i.e., approximately 38% of the lakes. The sample size for each state in the strata was proportional to the number of lakes sampled in the state in 2007. Exceptions were made when a state implemented a state-level design in 2007. These sample sizes are given in accompanying spreadsheet in sheet named

"#Lakes State". A total sample size of 1000 lakes was desired for the national design. The sample size for each state was proportional (approximately 60%) to the state's sample size in NLA 2007 (columns D & E). The minimum number of lakes for a state was set at 8 and the maximum at 43 (column F). Although aggregated ecoregions were not explicitly used in the survey design or setting sample sizes, they are implicitly used since the NLA 2007 allocated sample sizes using aggregated ecoregions. Once these two sample sizes were set for a state, an additional sample size (column I) was allocated to a state so that the total number of sites in a state would be 50 lakes.

Lakes in the NLA 2007 Revisit stratum were selected with equal probability and did not depend on lake area. Note that the NLA 2007 did depend on lake area. New lakes in the design were selected with unequal probability based on five lake area categories. The total number of lakes for a state in this strata (sum columns D, E, and I) was divided by five and that sample size (approximately) was assigned to the "(10,20]" lake area category. Sample sizes for lake area categories "(20,50]" and ">50" were decreased successively by one and for lake area categories "(4,10]" and "(1,4]" were increased successively by one. This process was adjusted to meet the total sample size requirement for the stratum. The rationale for this assignment of sample sizes is based on experience that smaller lakes are more likely not to be lakes or be inaccessible than larger lakes. When lakes are replaced, the process is expected to more likely result in a equal number of lakes sampled by lake area category.

Lake Use and Replacement

Each lake selected to be sampled is given unique site identification (siteID). Site numbers consist of NLA12_ST-XXX where ST is two letter state abbreviation and XXX is a number between 101 and 999. It is critical this siteID be used in its entirety to make sure that the lakes are correctly identified. Within each state, lakes evaluated for potential sampling must have all siteIDs from the largest to the lowest number evaluated. For example, if NLA12_AL-0155 is the largest siteID evaluated within Alabama, then all siteID that are lower than "0155" within the state must be evaluated. Even more critical is that if NLA12_AL-0155 is the largest siteID that is actually sampled in the field, then all lower siteIDs within the state that are evaluated to be a target lake and are accessible must be sampled in the field.

National Design Lakes and Replacement Process

Lakes scheduled to be sampled for the national design are identified by the attribute "panel" having values of "NLA07RVT2", "NLA07RVT", "NLA12RVT" and "NLA12NAT". These lakes must all be evaluated and sampled if they meet the definition of a lake and are accessible. If one of these lakes cannot be sampled for any reason, then the replacement lakes are taken within the state from the lakes where "panel" is equal to "NLA12ST". The first replacement lake will be the lake within that list that has the lowest siteID; the second will have then next lowest siteID; etc. If all the lakes with "panel" equal to "NLA12ST" have been evaluated and additional lakes are still required, then lakes are used from lakes identified as "OverSamp" in "panel" are used in

siteID order (i.e. lowest siteID within the OverSamp lakes within the state). The national design includes all lakes > 1 hectare.

Lakes designated "NLA07RVT2" and "NLA12RVT" are lakes that are to be sampled twice in 2012. If a lake designated "NLA07RVT2" cannot be revisited, then the next lake designated as "NLA07RVT" should be sampled twice. If a lake designated "NLA12RVT" cannot be revisited, then the next lake designated as "NLA12NAT" should be sampled twice. In each case the same process is used if the second lake can not be sampled.

State Design Lakes and Replacement Process

If a state implements a state level design with 50 lakes, then the lakes scheduled to be sampled for the state design are identified by the attribute "panel" having values of "NLA07RVT2" , "NLA07RVT", "NLA12RVT" , "NLA12NAT" and "NLA12ST". These lakes must all be evaluated and sampled if they meet the definition of a lake and are accessible. If one of these lakes cannot be sampled for any reason, then the replacement lakes are taken within the state from the lakes where "panel" is equal to "OverSamp". The first replacement lake will be the lake within that list that has the lowest siteID; the second will have then next lowest siteID; etc.

If a state implements a state level design and elects not to include 1-4 hectare lakes in the state design, then the state must first follow the procedures for the national design, which includes 1-4 hectare lakes. After the required number of lakes for the national design is sampled, then the remaining lakes for the state design may exclude 1-4 hectare lakes during the lake evaluation process. Otherwise the lake replacement process follows the procedure described in the previous paragraph.

Sample Frame Summary

`See accompanying spreadsheet NLA2012 Design Summary 20110320.xlsx`

Site Selection Summary

`See accompanying spreadsheet NLA2012 Design Summary 20110320.xlsx`

Description of Sample Design Output:

The dbf file for the shapefile ("NLA Lake Sites Final") has the following variable definitions:

Variable Name	Description
siteID	Unique identification label for each lake in the sample
Lon_DD	Lake location longitude in decimal degrees coordinate (see projection below for datum)
Lat_DD	Lake location latitude in decimal degrees coordinate (see projection information below)

xcoord	x-coordinate of lake centroid (see projection information below)
ycoord	y-coordinate of lake centroid (see Albers projection information below)
AREA_CAT6	Lake area category used for survey design based on 6 area categories
AREA_CAT6	Lake area category based on 10 area categories
NLA12_SF	Include if Lake included in NLA 2012 sample frame; Exclude otherwise
COMP_SF	Identifies lakes that are in both NLA 2007 and NLA 2012; in NLA 2007 but not NLA 2012; not in NLA 2007 but in NLA 2012; and lakes not present in 2007 but now in 2012 sample frame
NLA12_CLS	NLA 2012 class: NLA07RVT – lake sampled in NLA 2007; NLA12NEW – lake not sampled in NLA 2007; Exclude – lake excluded from sample frame
NLA07_EVLP	Identifies probability sample lakes that were or were not evaluated in NLA 2007
NLA12_STRA	Strata used for NLA2012 survey design based on state and NLA12_CLS
mdcaty	Categories used for unequal probability selection within a stratum.
wgt	Weight (number of lakes) to be used in the statistical analyses. It is the inverse of the inclusion probability
stratum	Strata used in design. All equal to NLALake
panel	Panel_1 identifies the 1000 lakes in the base design. Oversamp identifies lakes to be used as replacements as necessary.
EvalStatus	Placeholder to record the results of the lake recon evaluation (see below)
EvalReason	Placeholder to record reason for the evalution result.
COMID	From original NHD lake frame shapefile
FCODE	NHD fcode field
FTYPE	NHD ftype field
DES_FTYPE	Design Ftype based on NHD Ftype
REACHCODE	NHD Reach code
NLA07_RCHC	NHD reach code from NLA 2007 sample frame
COMID2007	NHD COMID from NLA 2007 sample frame (primary)
COMIDs2007	NHD COMIDs from NLA 2007 sample frame when multiple polygons from 2007 sample frame were combined for 2012 sample frame
GNIS_ID	GNIS identification
GNIS_Name	GNIS name
LAT_DD_N83	Latitude in decimal degrees from NHD
LON_DD_N83	Longitude in decimal degrees from NHD
X_ALBERS	x-coordinate from Albers projection for latitude above
Y_ALBERS	y-coordinate from Albers projection for longitude above
STATECTY	FIPS state and county code

ST	State two-letter codes for all states lake polygon occurs in
ST_NLA2012	State lake is assigned to for NLA 2012
STATE_PCNT	Percent of lake area that occurs in
BORD_LAK	State border lake = Yes; state non-border lake = No
CNTYNAME	County name
OWNSHP	Land ownership
EPA_REG	EPA Region
URBAN	Yes – Urban lake; No- non-urban lake
AREA_HA	Lake area in hectares (NHD)
ELEVATION	Lake elevation from NED
US_L4CODE	Omernik Level IV ecoregion code
US_L4NAME	Omernik Level IV ecoregion name
US_L3CODE	Omernik Level III ecoregion code
US_L3NAME	Omernik Level III ecoregion name
NA_L3CODE	CEC North American Level III ecoregion code
NA_L2CODE	CEC North American Level III ecoregion code
NA_L2NAME	CEC North American Level II ecoregion name
NA_L1CODE	CEC North American Level I ecoregion code
NA_L1NAME	CEC North American Level I ecoregion name
WSA3	WSA 3 aggregated Omernik ecoregions
WSA9	WSA 9 aggregated Omernik ecoregions
HUC2	HUC2 Region coded
Region	HUC2 Region name
HUC_8	Hydrologic unit code 8-digit
CU_NAME	HUC name
NLA07_SF	Include – Lake included in NLA 2007
NLA07_STRA	NLA 2007 Strata
NLA07_NAME	Lake name in 2007
NLA07_REF	Identifies reference lakes in 2007
NLA07_MDC	NLA 2007 multi-density category
NLA07_EVAL	NLA 2007 lake evaluation status
NLA07_WGT	NLA 2007 final weight
NES_LAKE	NES lake from 1970s survey
NESLAKE_ID	NES lake identification number
STORETNUM	Storet number for NES Lake

Projection Information

PROJCS["USA_Contiguous_Albers_Equal_Area_Conic",
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",37.5],
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 http://www.epa.gov/nheerl/arm. 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