National Flowing Waters Assessment Survey Design: 2008-2009

Contact:

Susan Holdsworth
Office of Water
OWOW Monitoring Branch
USEPA (4503T)
1200 Pennsylvania Ave. NW
Washington, DC 20460
(202) 566-1187
Email: Holdsworth.Susan@epa.gov

Description of Sample Design

Objectives:

The objectives, or design requirements, for the National Flowing Waters Assessment are to produce:

1. Estimates of the 2008-2009 status of flowing waters nationally and regionally (9 aggregated Omernik ecoregions),
2. Estimates of the 2008-2009 status of wadeable streams and non-wadeable rivers nationally and regionally (9 aggregated Omernik ecoregions),
3. Estimates of the 2008-2009 status or urban flowing waters nationally,
4. Estimates of the change in status in wadeable streams between 2008-2009 and 2004, nationally and regionally (9 aggregated Omernik ecoregions).

A secondary objective is to have each state sample approximately an equal number of sites (37-38).

Target population: The target populations consists of all streams and rivers within the 48 contiguous states that have flowing water during the study index period excluding portions of tidal rivers up to head of salt. The study index period extends from April/May to September and is generally characterized by low flow conditions. The target population includes the Great Rivers. Run-of-the-river ponds and pools are included while reservoirs are excluded.

Sample Frame: The sample frame was derived from the National Hydrography Dataset (NHD), in particular NHD-Plus. Attributes from NHD-Plus and additional attributes added to the sample frame that are used in the survey design include: (1) state, (2) EPA Region, (3) NAWQA Mega Region, (4) Omernik Ecoregion Level 3 (NACEC version), (4) WSA aggregated ecoregions (nine and three regions), (5) Strahler order, (6) Strahler order categories (1st, 2nd, …, 7th, and 8th+), (6) FCode, (7) Urban, and (8) Frame07.

The version of NHD-Plus used includes two separate Strahler order calculations, one that is included on the publicly available NHD-Plus version. The other Strahler order calculation (SO attribute name) more accurately reflects the true Strahler order and is used for the survey design. The StrahCat attribute collapses 8th, 9th, and 10th order rivers in to a single category.

The Urban attribute was created by intersecting a modified version of the Census Bureau national urban boundary GIS coverage with NHD-Plus. The Census Bureau's boundaries were buffered 100 meters to include a majority of stream features intersecting and coincident with urban areas. Where this buffer did not completely gather all the river features within the urban areas (rivers intersecting cities are excluded from the Census Bureau's urban areas), the NHD-Plus river area (polygon) features were clipped at a three kilometer buffer around the urban areas and combined with the buffered urban area to create the modified urban database. If a stream or river segment was within this boundary, it is designated as "Urban"; otherwise as "NonUrban".

FCODE is directly from NHD-Plus and is used to identify which segments in NHD were included in the sample frame. The attribute Frame07 identifies each segment as either "Include" or "Exclude". Frame07 was created so that segments included in the sample frame could be easily identified. FCODE values included in the GIS shapefile:
Included in FW08 sample frame (Frame07='Include'):
    33400   Connector
    33600   Canal/Ditch
    42801   Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = At or Near
    46000   Stream/River
    46003   Stream/River (Intermittent)
    46006   Stream/River (Perennial)
    58000   Artificial Path (removed from dataset if coded through Lake/Pond and Reservoirs)
Excluded in FW08 sample frame (Frame07='Exclude')
    42800   Pipeline
    42802   Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = Elevated
    42803   Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = Underground
    42804   Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = Underwater
    42806   Pipeline: Pipeline Type = General Case; Relationship to Surface = Elevated
    4280   Pipeline: Pipeline Type = General Case; Relationship to Surface = Underground
    42809   Pipeline: Pipeline Type = Penstock; Relationship to Surface = At or Near
    42811   Pipeline: Pipeline Type = Penstock; Relationship to Surface = Underground
    42813   Pipeline: Pipeline Type = Siphon
    56600   Coastline

Rivers that had Strahler order greater than or equal to 5th order and had FCODE equal to 46003 (intermittent) were included in the FW08 sample frame for all states west of 96 degrees longitude (North Dakota to Texas and states west). This was done to ensure that all large rivers in the more arid west were included regardless of NHD-Plus intermittent code.

Survey Design: The survey design consists of two major components in order to address the dual objectives of (1) estimating current status for all flowing waters and (2) estimating change in status for wadeable streams from the 2004 Wadeable Stream Assessment. These two components are termed: (1) NFW design and (2) WSA_Revisit design. A Generalized Random Tessellation Stratified (GRTS) survey design for a linear resource is used for the

NFW design and a GRTS survey design for a finite resource is used for the WSA_Revisit design. The design includes reverse hierarchical ordering of the selected sites.

Stratification: The survey design is explicitly stratified by state for the NFW design. The original WSA design had several strata (EMAP West, New England, Virginia, Iowa, and remaining eastern states combined). The WSA_Revisit design ignores these strata in the selection of the subset of sites from the WSA to be revisited as part of the current Flowing Water design.

Multi-density categories: A complex unequal probability selection process was used in each of the two components of the survey design. They are described separately.

NFW design: Unequal probability categories are defined separately for wadeable streams (1st to 4th order) and non-wadeable rivers (5th to 10th order). Note wadeable and non-wadeable are used to designate Strahler order classes and do not imply that the streams will actually be wadeable or non-wadeable. The expected sample size is 450 for wadeable streams and 900 for non-wadeable rivers.

For wadeable stream category, within each state unequal selection probabilities were defined for 1st, 2nd, 3rd, and 4th order streams so that an equal number of sites would occur for each order. Then these unequal selection probabilities were adjusted by WSA nine aggregated ecoregion categories so that an equal number of sites would occur in each WSA nine aggregated ecoregion category.

For non-wadeable river category, unequal selection probabilities were defined for 5th, 6th, 7th, and 8th+ order rivers so that the expected number of sites would be 350, 275, 175, and 100 sites, respectively. Then these unequal selection probabilities were adjusted by WSA nine aggregated ecoregion categories so that an equal number of sites would occur in each WSA nine aggregated ecoregion category.

Given these initial selection probabilities, the expected number of urban and non-urban sites was calculated to determine if at least 150 urban sites would be selected. Over 150 urban sites were expected so no additional adjustment was required to satisfy the urban design requirement.

The final adjustment of the selection probabilities was to adjust them to minimize the range in the number of sites across the 48 states while still meeting the other design requirements. Given a total of 1350 sites for the NFW design, each state would sample 28 sites. This could not be achieved, although the range was able to be decreased.

WSA_Revisit Design: The Wadeable Stream Assessment sampled 1390 sites between 2000 and 2004. To estimate change, 450 of these sites will be revisited as part of the 2008-9 Flowing Waters assessment. The revisit design selects the 450 sites using unequal selection probabilities. Initially, all sites were assigned an equal selection probability of 1.

First, four intensification study regions were sampled as part of the WSA. These regions are the Wenatchee Watershed in Washington, Lower John Day and Deschutes watersheds in Oregon, Northern California coastal watersheds, and southern California coastal watersheds. The selection probabilities for sites within these regions were adjusted (lowered) to give the expected number of sites within a study region if a state-wide survey design was done without intensification.

Second, the density of sites sampled for the EMAP-West portion of the WSA was greater than for the 36 eastern states. The selection probabilities were reduced for EMAP-West states to adjust for this. The Southern Appalachian aggregated ecoregion sites in WSA had high weights and when the initial selection probabilities were increased for these sites as well to ensure that the final weights for these sites were not extreme.

Third, the selection probabilities developed above were then adjusted to achieve approximately an equal number of sites across all nine WSA aggregated ecoregions.

Fourth, the overall weight, inverse of selection probability, was calculated by multiplying the original WSA weight by the inverse of the above selection probability. This accounts for the fact that the WSA_Revisit design is a two-stage sample of wadeable streams.

WSA_Revisit design weights and NFW design weights associated with wadeable streams will have to be adjusted to account for fact that they are two independent survey designs of wadeable streams for the 48 states. This will be done after the sites are evaluated and sampled.

State Designs: For states that have a current, compatible state-wide probability design that cover all flowing waters, an option is provided to use their sites instead of the flowing water design sites. Whether the option is exercised for a state, requires that (1) their state design be a probability survey design, (2) their target population of streams and rivers includes the target population for the national Flowing Waters Assessment, (3) their sample frame includes the national Flowing Waters Assessment sample frame, and (4) their design is implemented state-wide in 2008-2009. The state must also agree to measure all the indicators included in the national Flowing Waters Assessment using the national field and laboratory protocols.

Panels: Design uses a single panel (Base).

Expected sample size: Expected sample size is 1800 flowing water sites, 450 sites revisited from the WSA, 450 new sites from 1st to 4th order, and 900 new sites from 5th to 10th order.

Over sample: No over sample sites were selected for the WSA_Revisit design. The expectation is that all, or almost all, of the 450 sites selected will be sampled given they were sampled previously. For the NWW design, the over sample is nine times the expected sample size within each state. The large over sample size was done to accommodate those states who may want to increase the number of sites sampled within their state for a state-level design.

Site Use: Each stream/river selected to be sampled is given unique site identification (siteID) that consists of two parts: (1) NFW08 that identifies the sites as part of the 2008-9 National Flowing Waters Assessment and (2) the two-letter state FIPS code followed by a number between 001 and 999 within each state. It critical this siteID be used in its entirety to make sure that the stream and rivers sites are correctly identified.

Sites are organized to be used within each state. If a stream or river site is evaluated and determined that it can not be sampled, then it is to replaced by another site within the state. Sites that are coded as 1st, 2nd, 3rd and 4th are to be replaced by over sample sites that are coded 1st, 2nd, 3rd or 4th, ignoring order within this range. For example, a 2nd order would be replaced by either a 1st, 2nd, 3rd or 4th order stream. Sites that are coded as 5th, 6th, 7th, 8th, 9th, or 10th order are to be replaced by over sample sites that are coded 5th, 6th, 7th, 8th, 9th, or 10th order, ignoring order within this range. For example, a 5th order river would be replaced by a 5th, 6th, 7th, 8th, 9th, or 10th order river. In each case the next lowest siteID that is within the Strahler order set is used for the replacement.

Site Revisits: The monitoring design includes 200 site revisits to obtain estimates of local measurement variability and to satisfy quality assurance needs. Each state will complete four (4) site revisits: two wadeable streams and two nonwadable rivers. These sites will be the first two (in siteID order) wadeable and nonwadeable river sites selected that are sampleable.

Change Estimation

The design objectives include estimating the change in status in wadeable streams between 2008-2009 and 2004, nationally and regionally (9 aggregated Omernik ecoregions). The survey design for accomplishing this is a split-panel design. The split-panel design for change estimation includes 1390 wadeable sites visited as part of the Wadeable Stream Assessment in 2000-2004 and an expected 900 wadeable sites visited in 2008-2009 of which 450 are repeat visits of a subset of the 1390 sites. This design enables two different types of change estimates to be made: net change and gross change. The difference in these can be illustrated by two tables. The Without Revisits table illustrates that it is possible to estimate net change between NFWA and WSA, in this case 33% impaired in WSA versus 22.2% in NFWA. The Using Revisits table illustrates that it is not only possible to estimate net change between the two time periods but also possible to estimate gross change, e.g., 60/150 = 40% of WSA Impaired are no longer impaired or that 10/300 = 3.3% of WSA Not Impaired is now Impaired in NFWA. A split-panel design enables gross change estimates to be made while also providing opportunity to visit new stream locations.

Without Revisits	NFWA Impaired	NFWA Not Impaired	Total
WSA Impaired			150
WSA Not Impaired			300
Total	100	350	450

Using Revisits	NFWA Impaired	NFWA Not Impaired	Total
WSA Impaired	90	60	150
WSA Not Impaired	10	290	300
Total	100	350	450

Sample Frame Summary

The sample frame is available as a shapefile. This shapefile was also summarized by all the unique combinations of attribute variables that could be used in determining the selection probabilities. A few summaries are given below in km.

```
FCODE Exclude Exclude-E Include     Sum
33400       0        2    7007    7009
33600       0      425  276461  276886
33601     526        0       0     526
42800      66        0       0      66
42801    2334        2       6    2342
42802      26        0       0      26
42803    7239        2       0    7241
42804       1        0       0       1
42806       2        0       0       2
42807      39        0       0      39
42809      73        0       0      73
42811      15        0       0      15
42813     149        0       0     149
46000       0        0     126     126
46003 2298650      130  909432 3208212
46006       0     1625 1864018 1865643
55800       0    15372  122120  137492
56600   20384    23483       0   43867
Sum   2329504    41041 3179170 5549715


Frame07=Include only
Order NonUrban Urban     Sum
0       292695  27499  320194
1      1465222  66972 1532194
2       517101  26378  543479
3       317161  16441  333602
4       193440  10739  204179
5       123682   9891  133573
6        57788   8071   65859
7        25140   4511   29651
8         8527   2396   10923
9         2711    848    3559
10        1470    486    1956
Sum   3004937 174232 3179169



              Strahler Order
         0       1      2      3      4      5      6     7     8    9   10     Sum
CPL 134111  355821 115083  61419  35768  19757  10636  7979  1485  223 1655  743937
NAP   5188  138080  41830  21806  11786   5735   2469   475     0    0    0  227369
NPL   8664    5880   7456  10429  11584  10847   7181  1801   807  404    0   65053
SAP   6987  406914 116267  67142  39233  18214   9458  2249  1445  256    0  668165
```

```
SPL  15512   10695  15403  17932  16928  15553  8034  4966   868    0    0  105891
TPL  24836  271573  91433  56266  32920  19909  7908  3595  1714 1747  302  512203
UMW  14483  116764  41908  24116  12817   6528  2202   122   697    0    0  219637
WMT  24706  200340  89421  52706  26633  13800  6510  1692   608  277    0  416693
XER  85707   26127  24676  21787  16512  23232 11464  6772  3299  653    0  220229
Sum 320194 1532194 543477 333603 204181 133575 65862 29651 10923 3560 1957 3179177
```

Site Selection Summary

Number of sites for the National Flowing Waters Design:

By Aggregated Ecoregion and Wadeable (1st to 4th) and Non-wadeable (5th to 10th)
Base sites

```
       Wade NonWade   Sum
  CPL   136     125   261
  NAP   100     104   204
  NPL    68     101   169
  SAP   147     131   278
  SPL    74      95   169
  TPL   104      85   189
  UMW    83      75   158
  WMT   101      63   164
  XER    91     117   208
  Sum   904     896  1800
```

```
Over Sample Sites
       Wade NonWade   Sum
  CPL   831    1098  1929
  NAP   457     929  1386
  NPL   310     741  1051
  SAP   636    1226  1862
  SPL   353     794  1147
  TPL   475     862  1337
  UMW   335     656   991
  WMT   375     693  1068
  XER   387     992  1379
  Sum  4159    7991 12150
```

```
By Urban and Non Urban
            Base OverSamp    Sum
  NonUrban  1591    10475  12066
  Urban      209     1675   1884
  Sum       1800    12150  13950
```

```
By EPA Region
      Base OverSamp   Sum
  1    115      828   943
  2     61      386   447
  3    155     1121  1276
  4    261     1884  2145
  5    266     1772  2038
  6    214     1422  1636
  7    169     1194  1363
  8    331     2032  2363
```

```
 9     120      755    875
10     108      756    864
Sum   1800    12150  13950
```

```
      Base OverSamp   Sum
AL     36      225    261
AR     37      252    289
AZ     36      252    288
CA     44      252    296
CO     48      252    300
CT     22      180    202
DE     21      180    201
FL     29      252    281
GA     27      216    243
IA     46      252    298
ID     35      252    287
IL     35      252    287
IN     33      252    285
KS     44      342    386
KY     40      252    292
LA     38      252    290
MA     19      162    181
MD     28      207    235
ME     21       90    111
MI     52      333    385
MN     48      333    381
MO     36      252    288
MS     33      252    285
MT     62      387    449
NC     41      252    293
ND     61      387    448
NE     44      342    386
NH     16      117    133
NJ     15      117    132
NM     31      252    283
NV     40      252    292
NY     46      270    316
OH     36      252    288
OK     52      342    394
OR     37      252    289
PA     39      243    282
RI     14      117    131
SC     24      207    231
SD     65      387    452
TN     33      252    285
TX     56      324    380
UT     41      252    293
VA     38      252    290
VT     23      162    185
WA     36      252    288
WI     55      315    370
WV     32      252    284
WY     55      369    424
Sum  1800    12150  13950
```

By Aggregated Ecoregion and Strahler Order

	0	1	2	3	4	5	6	7	8	9	10	Sum

```
CPL   16    57    31    17    15    38    33    36     9     1     8    261
NAP    0    37    34    14    15    60    35     9     0     0     0    204
NPL    4     0    13    21    30    47    40     8     4     2     0    169
SAP    1    57    40    33    16    51    59    14     6     1     0    278
SPL    5     6    18    23    22    38    28    26     3     0     0    169
TPL    1    35    28    20    20    37    22    11     6     9     0    189
UMW    3    28    32    12     8    45    16     2    12     0     0    158
WMT    2    31    33    27     8    32    23     5     3     0     0    164
XER   18     7    21    28    17    42    35    26    12     2     0    208
Sum   50   258   250   195   151   390   291   137    55    15     8   1800
```

Description of Sample Design Output:

The dbf file for the shapefile (“*NFWA* Sites Final”) has the following variable definitions:

Variable Name	Description
siteID	Unique identification label for each stream/river in the sample
xalb83	x-coordinate of stream/river site (see Albers projection information below)
Yalb83	y-coordinate of stream/river site (see Albers projection information below)
Lon_dd83	Longitude in decimal degrees of stream/river site
Lat_dd83	Latitude in decimal degrees of stream/river site
mdcaty	inclusion probability used for the unequal probability selection of the sample stream/river
wgt	Weight (in km) to be used in the statistical analyses. It is the inverse of the inclusion probability
stratum	Strata used in design
panel	Base identifies the 1800 stream and river sites in the base design. Oversamp identifies sites to be used as replacements as necessary.
EvalStatus	Placeholder to record the results of the stream/river recon evaluation (see below)
EvalReason	Placeholder to record reason for the evelution result.
COMID	NHD-PLUS unique segment identifier
GNIS_NAME	GNIS feature name
FCODE	NHD feature type name
SO	NHD-Plus new Strahler order
StrahCat	Strahler order categories used in survey design
STATE	FIPS state two letter code
EPA_REG	EPA region
FW_ECO9	Nine Omernik aggregated ecoregions
FW_ECO3	Three Omernik aggregated ecoregions
NACEC_ECO	Omernik ecoregion names based on North American Ecoregions
NAWQA_REG	NAWQA Mega Regions
Urban	Urban if site is within urban boundary. NonUrban otherwise

NFWACat	For NFW design, unique categories used to determine inclusion probabilities
WSA_SITEID	SiteID from the 2004 WSA design
WSA_STRATU	Strata used in the 2004 WSA design
WGT_WSA	WSA weight (weight for the first stage of two-stage WSA revisit design)
WSA_ECO9	WSA aggregated ecoregion (included since ecoregion boundaries changed)
WGT_RESAMP	Combined two-stage weight for WSA Revisit design
WSA_ORD	Strahler order based on RF3 as used in WSA design

Projection Information

PROJCS["USA_Contiguous_Albers_Equal_Area_Conic_USGS_version",
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]],
VERTCS["NAD_1983",
DATUM["D_North_American_1983",
SPHEROID["GRS_1980",6378137.0,298.257222101]],
PARAMETER["Vertical_Shift",0.0],
PARAMETER["Direction",1.0],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 Pages:
        US EPA Aquatic Monitoring Research: http://www.epa.gov/nheerl/arm
        NHD Plus: http://www.horizon-systems.com/nhdplus