GAGES II (Geospatial Attributes of Gages for Evaluating Streamflow) summary report

Data · September 2011

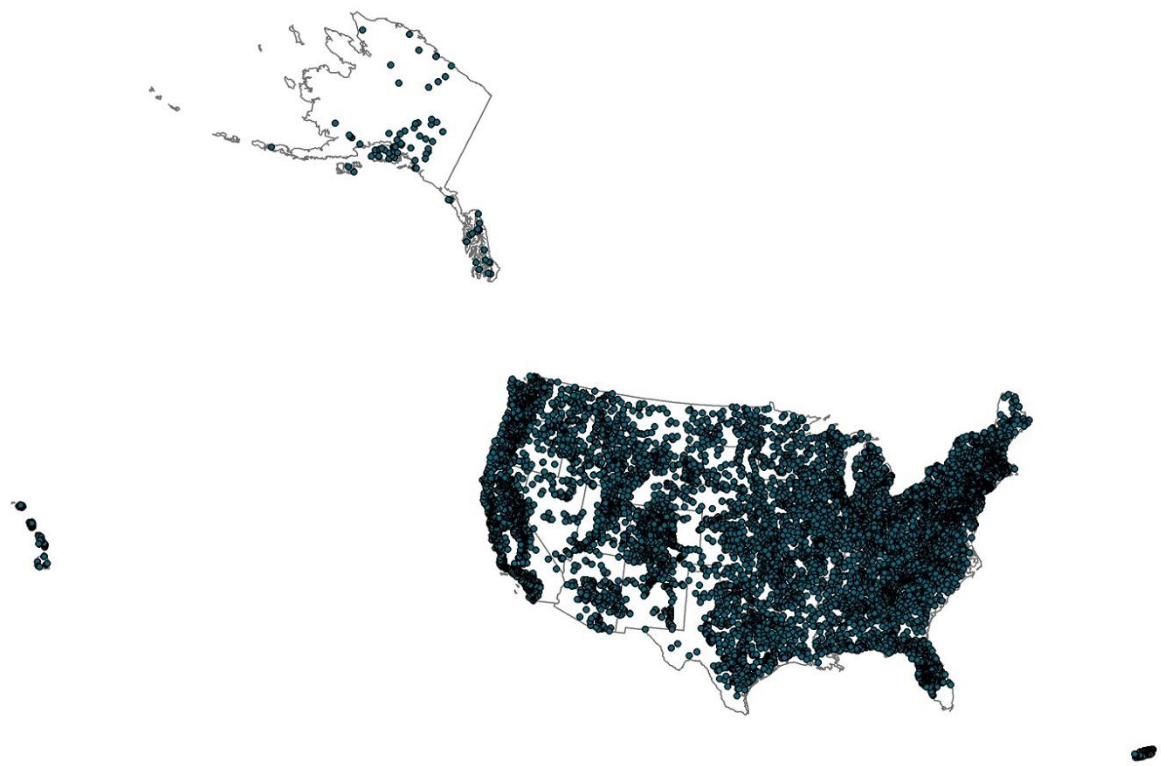
1 author:

James A. Falcone
United States Geological Survey

GAGES II (Geospatial **A**ttributes of **G**ages for **E**valuating **S**treamflow)

summary report

September 2011



1. Summary

This dataset, termed "GAGES II", an acronym for **G**eospatial **A**ttributes of **G**ages for **E**valuating **S**treamflow, version II, provides geospatial data and classifications for 9,322 stream gages maintained by the U.S. Geological Survey (USGS). The geospatial data are published at: http://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml. It is an update to the original GAGES, which was published as a Data Paper on the journal Ecology's website (Falcone and others, 2010b) in 2010. This dataset has two purposes: (1) to provide users with a comprehensive set of geospatial characteristics for a large number of gaged watersheds, particularly for gages with long flow record, and (2) to provide a determination of which of those watersheds represent hydrologic conditions which are least disturbed by human influences ("reference gages"), compared to other watersheds within ecoregions. Identifying reference gages serves important research goals: for example identifying conditions or goals for stream restoration or establishing environmental flow, climate change studies, and more.

The GAGES II dataset consists of gages which have had either 20+ complete years (not necessarily continuous) of discharge record since 1950, or are currently active, as of water year 2009, and whose watersheds lie within the United States, including Alaska, Hawaii, and Puerto Rico. Reference gages were identified based on indicators that they were the least-disturbed watersheds within the framework of broad regions, based on 12 major ecoregions across the United States. Of the 9,322 total sites, 2,057 are classified as reference and 7,265 as non-reference. Of the 2,057 reference sites, 1,633 have (through 2009) 20+ years of record since 1950. Some sites have very long flow records: a number of gages have been in continuous service since 1900 (at least), and have 110 years of complete record (1900-2009) to date.

The geospatial data include several hundred watershed characteristics compiled from national data sources, including environmental features (e.g. climate – including historical precipitation, geology, soils, topography) and anthropogenic influences (e.g. land use, road density, presence of dams, canals, or power plants). The dataset also includes comments from local USGS Water Science Centers, based on Annual Data Reports, pertinent to hydrologic modifications and influences. The data posted also include watershed boundaries in GIS format.

This overall dataset is different in nature to the USGS Hydro-Climatic Data Network (HCDN; Slack and Landwehr 1992), whose data evaluation ended with water year 1988. The HCDN identifies stream gages which at some point in their history had periods which represented natural flow, and the years in which those natural flows occurred were identified (i.e. not all HCDN sites were in reference condition even in 1988, for example, 02353500). The HCDN remains a valuable indication of historic natural streamflow data. However, the goal of this dataset was to identify watersheds which currently have near-natural flow conditions, and the

2,057 reference sites identified here were derived independently of the HCDN. A subset, however, noted in the BasinID worksheet as “HCDN-2009”, has been identified as an updated list of 743 sites for potential hydro-climatic study. The HCDN-2009 sites fulfill all of the following criteria: (a) have complete and continuous flow record for the last 20 years (20 years of complete discharge record for water years 1990-2009), and were thus also currently active as of 2009, (b) are identified as being in current reference condition according to the GAGES-II classification, (c) have < 5% imperviousness as measured from the NLCD 2006, and (d) were not eliminated by a review from participating state Water Science Center evaluators.

2. What's New in GAGES II vs. original GAGES

Although the overall methodology did not change between the original GAGES and this posting of GAGES II (see original report at http://esapubs.org/Archive/ecol/E091/045/metadata.htm), there are a number of additions and improvements, both major and minor. The most significant improvements and changes are given here:

1) The flow criteria for the original GAGES dataset was that a gage had to have at least 20 years of complete flow record since 1950. This set includes those - plus several hundred which were originally not delineated - as well as any gage that was "active" in water year 2009 (being “active” defined here as having at least 50 days of daily discharge data between Oct 1, 2008 and Sept 30, 2009). This dataset therefore has more records (9,322 vs. 6,785 in the original), however not all gages given here necessarily have long-term flow record. However, we have provided the historical flow matrix - which indicates the years for which there was complete flow record, by year since 1900 - so users may select for themselves which gages they wish to use based on their desired period (see FlowRec tab in data spreadsheets).

2) This set contains records for Alaska, Hawaii, and Puerto Rico (n = 255). The method for classifying these records differed slightly from the conterminous US in that we solicited input from the local Water Science Centers for those states (i.e. their judgment about which were the least disturbed watersheds), and added that as additional information when we classified the watersheds. Only a subset of the variables that were calculated for the conterminous US gages were calculated for the AK-HI-PR gages, because it was difficult to identify data layers that were completely consistent for those states and the conterminous US. As much data are provided for those three areas as was feasible; these are primarily indicators of anthropogenic disturbances.

3) The original GAGES dataset contained only basin characteristics and the result of the classification (reference/non-reference). This version additionally includes zip files of the basin boundaries and the mainstem stream lines for each gage.

4) Basin boundaries were corrected for 129 of the original 6,785 gages, and basin characteristics for them re-calculated.

5) Land cover data were updated to the NLCD06.

6) We used a modified version of the Army Corps of Engineers 2009 National Inventory of Dams (NID), as opposed to the 2006 version used previously. It was "modified" in that we updated the locations of 59,000 of the 83,000 records in the 2009 NID from a dams dataset that has been verified and corrected against imagery by the USGS National Map, courtesy of Jeff Simley and colleagues of the USGS. We also eliminated several hundred large duplicate dams. So we believe the dams data source we used is the best that is currently nationally available, to our knowledge. Additionally the method of assigning dates to dams without a date was improved. All of the information in the "HydroMod_Dams" worksheet is derived from this improved data source.

7) Because there is great interest in time series study, we include now mean annual precipitation and air temperature for each gaged watershed for each year between 1950 and 2009 (60 years), based on 4-km PRISM data. These are provided in the "Climate_PptAnnual" and “Climate_TmpAnnual” worksheets.

8) Detailed crop type information is now given, from the USDA's 2009 Cropland Data Layer.

9) A fairly extensive basin boundary QA evaluation is provided, which was not provided previously. This includes the difference of our drainage area to that given in the USGS NWIS system (94% of them are within +/- 10%), the result of a visual check of our basin boundaries against HUC10 watershed boundaries, and an overall quality indicator.

10) In addition to the flow years matrix noted above, an estimate is provided of the quality of the flow records (the variable FLOW_PCT_EST_VALUES in the FlowRec worksheet). This indicator is the percentage of daily discharge values in the entire record for the gage which have been estimated. Although on average only 4% of flow values are estimates, users may decide for themselves if they perhaps wish to exclude specific gages which have a somewhat high percentage of estimated flow values.

11) Several hundred mainstem stream lines from the original dataset were corrected, and mainstem-derived variables re-calculated.

12) A minor modification was made to the Hydrologic Disturbance Index (HYDRO_DISTURB_INDX in the Bas_Classif worksheet), in that the variable ARTFLOW_PCT was dropped as part of one of the index components.

13) Indicators of irrigated agriculture and presence of power plants (PCT_IRRIG_AG and POWER_XXX) are given in the HydroMod_Other worksheet.

14) 35 Western gages which measured the combined flow of a river and parallel canal were dropped. All of these had gages which measured the flow of the river alone.

15) A small number of variables were added, dropped, or updated from more current source data. The basin centroid lat/long were added (Bas_Morph worksheet). The LC_Change92_01 worksheet was dropped, and a variable indicating increase in urban land cover from 2001 to 2006 was added to the Pop_Infrastr worksheet. Population density for both day and night is given from Landscan data (Pop_Infrastr worksheet). Road-stream intersection density is now provided (Pop_Infrastr worksheet). Dominant watershed Potential Natural Vegetation is now provided (Regions worksheet). A small number of corrections were made to values for first and last freeze in the Climate worksheet. Some of the units for a small number of variables were modified to be easier to work with, e.g. distances are given in km instead of meters. A number of variable names changed in order to match StreamStats variable names, e.g. GAGE_ID became STAID, NAME became STANAME.

3. Criteria for Inclusion:

As noted, a description of the methods for this product is provided in Falcone and others, 2010b. Although the process for this dataset is nearly identical, the updated methods are given/summarized here.

The USGS has maintained tens of thousands of stream gages over the years, and at present nearly 9,000 are active in some way. Our criteria for flow record for GAGES II is that a gage must have either 20+ years of complete (365+ days) discharge record since 1950 (and may or may not be currently active), OR it must be currently active (our criteria: at least 50 days of discharge data in water year 2009). A water year runs from Oct 1 through Sept 30.

There are additionally several other criteria for inclusion: (a) watersheds had to be smaller than 50,000 sq km (~19,305 sq miles), (b) at least 95% of the watershed had to be within the United States (because of the unavailability of consistent GIS data sources outside the US), (c) gages which were known to be on canals, flumes, or other manmade channels were excluded, (d) at least some portion of an NHDPlus 100k stream segment had to intersect the watershed, and (e) we had to be able to delineate the watershed with at least a minimal confidence level. In some parts of the United States, for example, South Florida, it is difficult if not impossible to delineate meaningful watersheds, because of the lack of topographic relief and the high degree of manmade drainage.

4. Watershed Delineations

The watershed boundaries in this dataset were assembled or derived over a period of years, and in some cases based on slightly different DEMs (NHDPlus vs. NED) and/or using different tools, so there may be (typically very small) differences in basin boundary alignments for nested or adjacent watersheds. Approximately 8% of the gages in this dataset were previously watersheds under study by the USGS National Water-Quality Assessment Program (NAWQA; USGS 2008a). NAWQA is also the umbrella organization under which this project was executed, so there was incentive to ensure commonality with that program. For those basins we generally used the basin boundary which had already been created and verified by the NAWQA program (these are flagged by the variable NAWQA_SUID in the BasinID worksheet). For most of the remainder we obtained initial basin boundaries from NAWQA colleague Michael Wieczorek, whose delineations were based on NHDPlus 30-m flow direction and flow accumulation grids. A small number of these - approximately 10% - could not be readily done in an automated mode. For these we attempted a delineation (or re-delineation) from a different DEM source - the USGS National Elevation Data (NED, USGS, 2008h) using tools provided by the USGS Elevation Derivatives for National Applications (EDNA, USGS, 2008c) program. If a suitable delineation could still not be performed we attempted yet another delineation by re-attempting each individual delineation against NHDPlus data in ArcToolBox by altering how the "snap to pour point" was performed. If that was still unsuccessful, and for a very small number of basins where it was feasible, we manually digitized the boundary based on HUC10 boundaries and digital topographic maps. If no method seemed feasible for creating a basin boundary we excluded the record. For the Alaska-Hawaii-Puerto Rico records the majority of the basin boundaries were obtained from the local USGS Water Science Center in those states. In all, approximately 83% are from the Wieczorek delineations of NHDPlus data, 8% from NAWQA archives, 6% from EDNA data/tools, 2% obtained from the AK-HI-PR Water Science Centers, and 1% from ArcToolBox/NHDPlus or manual digitization.

5. Basin Boundary Quality Assurance

Because we recognize that automated processes which derive watershed boundaries from DEMS have the potential for some inaccuracy (depending on the spatial resolution and accuracy of the DEMS themselves, the topography of the region, or the prevalence of sinks or peaks, among other factors), we performed a series of evaluations to assess the quality of each boundary. In addition to automated checks, we visually evaluated every one of the 9,322 boundaries at least once, and in some cases two or three times, as described below.

The worksheet BOUND_QA in the data spreadsheets summarizes these evaluations. They were:

(a) Difference of our watershed area to the area reported in the USGS National Water Information System (NWIS; USGS 2008b). NWIS drainage areas, although considered to be an authoritative value, have the potential to be estimates or have some degree of inaccuracy. Additionally, for a small percentage of records - here, 102 of 9,322 - no area is given in NWIS. It is clear, however, that good agreement between our drainage area and the NWIS-reported area would be an excellent indicator of the likely accuracy of our watershed boundary. In our final dataset 87% of watersheds matched the NWIS area within +/- 5%, and 94% matched to within +/- 10%. However, simple agreement of drainage size does not guarantee that our boundary necessarily delineated precisely the correct area: for example, offsetting errors of omission and commission in the same boundary may occur. In that light, we also evaluated other factors, namely:

(b) Visual agreement of the watershed boundary to Watershed Boundary Dataset (WBD, USDA, 2010) HUC10 boundaries. The WBD is a hierarchical set of boundaries for the entire US (including Alaska, Hawaii, and Puerto Rico) that delineate areas ranging from major river basins (2-digit Hydrologic Unit Code, or "HUC02"s) to considerably smaller catchment boundaries (HUC12s). Using HUC10 catchment boundaries as a backdrop in ArcMap, we compared how well each of our watershed boundaries matched the boundaries of the HUC10(s) that the watershed fell in, and made qualitative notes ("good", "reasonable", "poor") about the agreement. Because our watersheds - the work for which began in 2006 - were derived independently of the WBD (WBD HUC12s and HUC10s were completed only in 2009-2010), we reasoned that a match between WDB and our watershed boundaries was another independent indicator of the quality of a boundary. 100% of our watersheds were visually evaluated in this way. The WDB visual evaluations were performed by two individuals (author and Chris Hill). Of the 9,322 watersheds in this dataset, 6,183 (66%) were judged in "good" agreement with the WBD10 boundaries, 2,683 (29%) "reasonable", 281 (3%) "poor", and 175 (2%) could not be judged because they did not share a common boundary (watersheds internal to a HUC10 area). Visual examples are available from the author on request. It should be noted that a "poor" match to HUC10 boundaries does not necessarily mean our boundary was poor (HUC10 boundaries can also have inaccuracy), but rather, that there is more uncertainty.

(c) Distance of gage point location to watershed boundary. It is possible that the point location might be offset somewhat from the basin boundary primarily due to the resolution of the DEM (30-m) and how the gage point is "snapped" to the closest pour point in the DEM when it is created. In the vast majority of cases this is not a problem. However, to judge it, we calculated the straight-line distance of the point to the nearest arc of the watershed boundary. 92% of records in our final dataset are within 100 meters, and the median value is 28 meters. In a small number of cases this calculation indicated a potential problem with the basin boundary (and in a very small number of cases of the lat/long of the gage location). Some gages were eliminated for this reason.

(d) Approximately 2,000 watershed boundaries were visually checked a second or third time, and compared to NHDPlus streamlines, DEMS, and/or digital topographic maps. These included all basins which disagreed with NWIS by more than 10%, those for which NWIS had no drainage area, and a random sample of basins which did match NWIS to within 10%. In a small number of cases watersheds were eliminated because our confidence was very low about their boundaries.

Once all of the above quality assurance information was compiled we assigned an overall BASIN_BOUNDARY_CONFIDENCE rating to each gage, a number which could theoretically range from 2 (low confidence) to 10 (highest confidence). This number was calculated based on points assigned from three of the factors above: (1) agreement with NWIS area: a point value of 6 was given if our boundary matched to within +/- 2% of NWIS, 5 points if +/- 2-5%, 4 points if +/- 5-10%, 3 points if +/- 10-20%, 2 points if +/- 20-30%, and 1 point if +/- 30% agreement to NWIS; (2) HUC10 evaluation: 3 points were given for "good" agreement, 2 points were given for "reasonable" (or "no common boundary"), and 1 point was given if "poor" agreement to HUC10 boundaries, and (3) distance of gage point location to basin boundary: 1 point was given if within 100-m, otherwise 0 points. The point values for the three factors were then summed. So a basin with a confidence rating of 10 matched the NWIS area to within 2%, had a "good" visual match to HUC10 boundaries, and the point location is within 100-m of the basin boundary. Most basins have high confidence: 88% of the 9,322 records have a confidence rating of 8 or above; that is, we are very or extremely confident that our boundaries represent the true drainage area. The confidence of boundaries has a geographic bias: in some areas of the U.S. it is simply more difficult to delineate watersheds (e.g. Florida). For example, in Hawaii, 84 of 91 basins (92%) have a confidence rating of 9 or above. Conversely, of the 284 gages in Florida, only 74 (26%) have a rating of 9 or above. Watersheds are more accurately and easily delineated in Hawaii than Florida.

We re-iterate here our opinion that all of the boundaries in this dataset are adequate for use at the scale of study intended for this dataset (national/regional), and that calculations performed even from lower-confidence watershed boundaries are likely to have a minimal amount of error for the variables in this dataset. We have already excluded those we felt were of unacceptable quality. Our goal is to share the information we have and our observations about varying degrees of certainty regarding the basin delineations.

6. Data Assembly and Organization

The majority of the variables given in the data worksheets for this project are based on watershed, riparian-level (all NHDPlus 100k streams in watershed), or mainstem-level (main channel stream in watershed) GIS calculations. These are also described in Falcone and others 2010b, and the general method, units, source, and citation for each individual variable are given in the variable description spreadsheet. A number of data fields are not the result of GIS calculations, but rather the result of assembling data from other sources - for example the ADR remarks and citations - or as a result of our interpretation - for example the basin classification (reference vs. non-reference). The main data spreadsheet is broken into 27 worksheets (spreadsheet tabs), which represent general types of variables. These are described here:

- BasinID: Basic identification characteristics of the stream gage (e.g., name, drainage area, lat/long, water resources region). Also included are four flags indicating whether the gage is a part of other pertinent streamgage networks: the Hydrologic Benchmark Network (HBN), the old Hydro-Climatic Data Network (HCDN), the new HCDN (HCDN-2009), and the National Streamflow Information Program Sentinel (NSIP-S) network.

- Bas_Classif: Reference/non-reference classification and primary information that went into the classification decision, including the hydro-disturbance index, pertinent Annual Data Report (ADR) remarks regarding flow alteration, and screening comments from NAWQA personnel visual evaluation.

- Bas_Morph: Basin morphology (compactness ratio, centroid location).

- Bound_QA: Information regarding the results of Quality Assurance checks for the basin boundaries.

- Climate: Climate characteristics (e.g., mean precipitation, temperature).

- Climate_Ppt_Annual: Mean basin precipitation, annually from 1950-2009.

- Climate_Tmp_Annual: Mean basin air temperature, annually from 1950-2009.

- FlowRec: Characteristics of the flow record, to include the flow matrix by year since 1900.

- Geology: Geological characteristics (e.g., dominant geology in watershed).

- Hydro: Hydrologic characteristics derived from GIS data (e.g., stream order at the streamgage, base-flow-index, runoff).

- HydroMod_Dams: Information about historical and current dams in the watershed.

- HydroMod_Other: Information about other anthropogenic hydrologic modifications (e.g., percent canals in watershed or on mainstem, presence of permitted pollution discharge sites, estimate of water withdrawal, presence of irrigated agriculture, presence of power plants).

- Landscape_Pat: Landscape pattern metric(s) (e.g., fragmentation of undeveloped land).

- LC06_Basin: Percentages of land cover circa year 2006 in the watershed, from NLCD06. For the Alaska-Hawaii-Puerto Rico records these data are from 2001 (NLCD01), because the 2006 data were not available for those three areas.

- LC06_Mains100: Percentages of land cover classes in 100-m mainstem buffer (approx. 100 m each side of mainstem stream line).

- LC06_Mains800: Percentages of land cover classes in 800-m mainstem buffer (approx. 800 m each side of mainstem streamline).

- LC06_Rip100: Percentages of land cover classes in 100-m riparian buffer (approx. 100 m each side of all streamlines in watershed).

- LC06_Rip800: Percentages of land cover classes in 800-m riparian buffer (approx. 800 m each side of all streamlines in watershed).

- LC_Crops: Percentages of land cover classes in the watershed from the USDA NASS 2009 Crop Data Layer (CDL).

- Nutrient_App: Estimates of nitrogen and phosphorus application in the watershed.

- Pest_App: Estimates of agricultural pesticide application in the watershed.

- Pop_Infrastr: Measures of population density and infrastructure (roads, impervious surfaces) in the watershed. Includes percent increase in urbanization from 2001 to 2006, from NLCD.

- Prot_Areas: Percent area of the watershed in "protected" land cover zones (e.g., National Parks, Wilderness Areas)

- Regions: Site location and percent of watershed area in various regions (e.g., EPA Level II or III ecoregions).

- Soils: Soils characteristics.

- Topo: Topographic characteristics (e.g., mean basin elevation, mean slope).

- X_Region_Names: A crosswalk of codes from the "Regions" worksheet to their names or descriptions.

Each specific variable in the variable description spreadsheet has several pieces of information about it. These are:

VARIABLE_TYPE: Variable type; indicates the type of variable (one of the 27 types above) and in which worksheet the data are contained in (e.g., "Hydro").

VARIABLE_NAME: The variable name (e.g., "STRAHLER_MAX").

DESCRIPTION: A text description of the variable (e.g., "Maximum Strahler stream order in watershed").

EXTENT: Spatial extent of GIS processing for that variable: either "watershed", "site", "riparian", or "mainstem". Unless indicated otherwise "watershed" indicates that the value represents the mean for the watershed for that parameter, "site" represents the value for the point location at the streamgage, "riparian" represents the mean for the riparian zone indicated, and "mainstem" represents the mean for the mainstem zone indicated.

STORAGE TYPE: Storage type: either "character", "integer", "floating point", or "date".

UNITS: Units (e.g., "percent" or "sq km").

TIME PERIOD: Time period of the source data (e.g. "1971-2000").

PROCESSING METHOD: Type of processing: either "grid", "vector polygon", "vector line", "vector point", or "join".

ADDITIONAL NOTE ON PROCESSING OR DATASET: Additional information about unusual or noteworthy aspects of processing, including No Data values, if applicable.

SOURCE DATA: Source of data (e.g., "NHDPlus"); usually the agency, organization, or program that created the data.

NOMINAL RESOLUTION/SCALE OF SOURCE DATA: Spatial resolution or scale (e.g., "1:100,000", or "30-m resolution grid").

CITATION: Reference for source data and/or processing method.

WEB SITE FOR SOURCE DATA: URL for source data, if online.

7. Gage Classifications

A complete description of the gage classification process - assigning a classification of reference or non-reference, is given in Falcone and others, 2010b, however also summarized here.

We judged the degree of human alteration on streamflow for each gage from three sources of information. The first was a GIS-derived hydrologic disturbance index (Falcone and others 2010a, Falcone and others, 2010b) based on seven variables: presence of major dams in the watershed (MAJ_DDENS_2009), change in reservoir storage from 1950-2009 (STOR_2009 - PRE1950_STOR), percentage of streamlines coded canals/ditches/pipelines in the watershed (CANALS_PCT), road density in the watershed (ROADS_KM_SQ_KM), distance of stream gage to nearest major pollutant discharge site (RAW_DIS_NEAREST_MAJ_NPDES), county-level fresh-water withdrawal estimate (WATER_WITHDR), and fragmentation of undeveloped land in the watershed (FRAGUN_BASIN). This index gave a general estimate of which watersheds were most likely to have streamflow alteration. The index scores were a global metric whose calculation was based on the entire universe of records, but interpreted within the context of individual ecoregions (described below).

The second source of information used to judge human alteration was local expert judgment about the site as published in USGS Annual Water Data Reports (U.S. Geological Survey, 2008e), or for Alaska-Hawaii-Puerto Rico from the local Water Science Centers themselves. Most annual data reports (ADRs) provide information about whether and how streamflow at each stream gage is influenced by regulation (e.g., dams) or diversions (e.g., canals, pipelines, or water withdrawals). We searched reports for the most recent water year in both electronic ADRs and paper reports for every stream gage, and we recorded any information about regulation and diversions.

The third source of information used to judge human alteration on streamflow was a visual inspection of each gage on high-resolution imagery (Google Earth; typically 1-2m spatial resolution) and digital topographic maps. This was done primarily to identify evidence of hydrologic alteration and other human activities near the stream gage, the mainstem, and major tributary channels. These images provided evidence of irrigated agriculture (and possible groundwater pumping or streamflow diversions) proximal to or upstream from stream gages. The presence of recently constructed reservoirs also was evident. Further, diversion structures, dams, and constructed channels were often readily visible in high-resolution images. We recorded qualitative notes about each gage based on these observations. It is noted, however, that it is not possible by visual examination of topographic maps and (even high-resolution) imagery to observe every aspect of flow alteration, particularly very small diversions, wells, and elements of groundwater pumping.

Once these three sources of information were assembled, the gages were broken into 12 broad ecoregions: nine in the conterminous US (see Figure 2 in Falcone and others, 2010b, reproduced here), and Alaska, Hawaii, and Puerto Rico. (It is recognized that Alaska spans multiple and very different ecoregions, but for purposes of this project the state was treated as one). Within each ecoregion, we identified those stream gages which we believed to have the least degree of human alteration on streamflow compared to other stream gages within that ecoregion. In general these were approximately the 20% of records with the least disturbance, although that percentage could vary considerably by ecoregion. For example in the agricultural Central Plains ecoregion only 16% of gages were classified as reference gages, whereas in the more pristine Western Mountains 25% of gages were classified as reference.

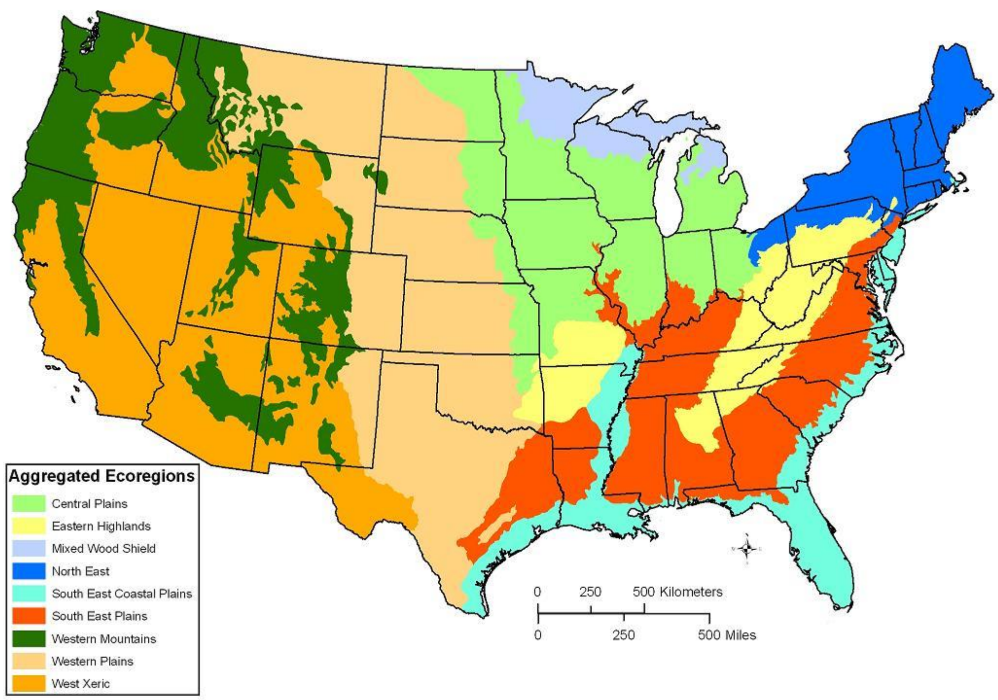


Figure 2 from Falcone and others, 2010b: Aggregated Ecoregions for the conterminous US

Users should note that within these broad regions there may still be variation in topographic, geologic, and other settings, and that for particular scales or applications users may wish to refine or even redefine the reference gages defined in this dataset.

8. Use Notes

It should be noted that the classification assigned in this project (reference or non-reference) was based on our evaluation of the criteria described above, which included both quantitative and qualitative data. Our classification is intended as the best identification of reference-quality sites that we could perform given the limitations of national-scale data, and within the context of the regions we used. Not every reference watershed given here is “pristine”: they were judged to be the best available within those ecoregions. It was believed that the deficiencies of national-scale GIS data, which do not capture small diversions, dams or groundwater pumping particularly well, would be supplemented by an examination of ADR remarks and visual examination of every watershed from high-resolution imagery. Even given that, however, it is possible that users with localized knowledge may be able to perform a better classification for specific areas. Users of these data should be attentive to potential differences between reference and non-reference gages - for example, watershed size (median size of reference gages = 212 sq km; of non-reference gages = 587 sq km). We also note that it would be possible to use the data posted here to identify a different, but equally valid, set of reference-quality stream gages for other purposes, based on other regions, or which place greater emphasis on specific anthropogenic activities or variables (e.g., urbanization), and that the classification presented here is not intended to be definitive.

Acknowledgments

The original conception and rationale for creating GAGES is from Daren Carlisle, David Wolock, and Michael Meador, of the USGS. In its assembly, we gratefully recognize the tremendous work done by Lisa Weber, Kelly Ruhl, Chris Hill, and Michael Wieczorek, as well as the helpful colleague reviews provided by Al Rea, Ken Eng, Silvia Terziotti, and Kerie Hitt, all of the USGS, and of three journal reviewers from the journal Ecology for the original data. We also acknowledge the invaluable input from our colleagues at USGS Water Science Centers: Delwyn Oki in Hawaii, Jackie McIntire and David Meyer in Alaska, and Marilyn Santiago and Heriberto Torres-Sierra in Puerto Rico. We also acknowledge the USGS EDNA program, for assistance and guidance on basin delineation methods, particularly Sandra Poppenga, Susan Greenlee, and Bruce Worstell.

References (for metadata or variable description spreadsheet)

AAPFCO (Association of American Plant Food Control Officials), 2009. Association of American Plant Food Control Officials. Accessed in May 2009 at http://www.aapfco.org/.

Bhaduri, B., Bright, E., Coleman, P., Urban, M.L., 2007. LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. GeoJournal 69 (1-2): 103-117.

Carlisle, D.M., Falcone J., Wolock, D.M., Meador, M.R., and Norris, R.H., 2009. Predicting the natural flow regime: models for assessing hydrological alteration in streams. River Research and Applications, in press, DOI at: http://dx.doi.org/10.1002/rra/1247.

Conservation Biology Institute, 2006. Protected Areas Database (PAD) Version 4. Conservation Biology Institute. Accessed in January 2006 at http://www.consbio.org/.

DiLuzio, M., Johnson, G.L., Daly, C., Eischeid, J.K., and Arnold, J.G., 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. Journal of Applied Meteorology and Climatology 47: 475-497.

Dingman, S.L., 2002. Physical Hydrology, Second Edition. Prentice Hall, Upper Saddle River, NJ, USA.

Falcone, J.A., Carlisle, D.M., and Weber, L.C., 2010a. Quantifying human disturbance in watersheds: variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. Ecological Indicators, 10(2), pp 264-273.

Falcone, J.A., Carlisle, D.M., Wolock, D.M., and Meador, M.R., 2010b. GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States, Ecology, 91 (2), p 621; Data Paper in Ecological Archives E091-045-D1; available online at: http://esapubs.org/Archive/ecol/E091/045/metadata.htm.

GeoLytics, 2001. Census 2000 and street 2000. East Brunswick, NJ, US. GeoLytics, Inc., 2 CDROMS.

Hamon, W.R., 1961. Estimating potential evapotranspiration. Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers 87:107-120.

Holmes, M.G.R., Young, A.R., Gustard, A.G., and Grew, R., 2002. A Region of Influence approach to predicting Flow Duration Curves within ungauged catchments. Hydrology and Earth System Sciences 6(4):721-731.

Hunt, C.D., 1979. National Atlas of the United States of America - Surficial Geology. U.S. Geological Survey. NAC-P-0204-75M-O [map].

Küchler, A.W, 1964. Potential natural vegetation of the conterminous United States (manual and map). Special Publ. 36, 1965 rev. New York: American Geographical Society, 116 p.

Krug, W. R., Gebert, W.A., and Graczyk, D.J., 1989. Preparation of average annual runoff map of the United States, 1951-80. U.S. Geological Survey Open-File Report 87-535.

Markham, Charles G., 1970. Seasonality of precipitation in the United States. Annals: Association of American Geographers 60:593-597.

McCabe, G.J. and Wolock, D.M., 2009. Variability in Northern Hemisphere snow cover. Climatic Change, in press.

Nakagaki, N. and Wolock, D.M., 2005. Estimation of agricultural pesticide use in drainage basins using land cover maps and county pesticide data. U.S. Geological Survey Open-File Report 2005-1188.

Omernik, J.M., 1987. Ecoregions of the conterminous United States. Annals: Association of American Geographers 77(1):118-125.

Orsborn, J.F., 1974. Determining streamflows from geomorphic parameters. Journal of the Irrigation and Drainage Division 100(4):455-475.

Poff, N.L., 1996. A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrological descriptors. Freshwater Biology 36:71-91.

PRISM Group, 2008. Parameter-Elevation Regressions on Independent Slopes Model (PRISM) Home Page. PRISM Group, Oregon State University. Accessed in October 2008 at http://www.prism.oregonstate.edu/.

Reed, J.C. and Bush, C.A., 2005, Generalized geologic map of the United States, Puerto Rico, and the U.S. Virgin Islands, [digital maps], Version 2.0, posted December 2005, accessed September 28, 2010, at http://pubs.usgs.gov/atlas/geologic/.

Ritters, K.H., Wickham, J.D., O'Neill, R., Jones, B., and Smith, E., 2000. Global-scale patterns of forest fragmentation. Conservation Ecology 4:3. [online] http://www.consecol.org/vol4/iss2/art3.

Ruddy, B.C., Lorenz, D.L., and Mueller, D.K., 2006. County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982-2001. U.S. Geological Survey Scientific Investigations Report 2006-5012.

Sanborn, S.C. and Bledsoe, B.P., 2006. Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. Journal of Hydrology 325:241-261.

Schmidt, K.M., J.P. Menakis, C.C .Hardy, W.J.Hann, and D.L. Bunnell. 2002. Development of coarse-scale spatial data for wildland fire and fuel management, General Technical Report RMRS-GTR-87, Fort Collins, CO, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 41 p. Available online at http://www.fs.fed.us/rm/pubs/rmrs_gtr87.html.

SILVIS Lab, 2008. The Wildland-Urban Interface, GIS data library. Forest and Wildlife Ecology, University of Wisconsin-Madison. Accessed in October 2008 at http://silvis.forest.wisc.edu/projects/WUI_Main.asp.

Slack, J.R. and Landwehr, J.M., 1992. Hydro-climatic data network (HCDN): A U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1874-1988. U.S. Geological Survey Open-File Report 92-129.

Soil Survey Staff, 1981. Land resource regions and major land resource areas of the United States. Agriculture Handbook 296. Rev. Ed. United States Department of Agriculture, Soil Conservation Service. Washington, DC.

Thomas, D.M. and Benson, M.A., 1970. Generalization of streamflow characteristics from drainage-watershed characteristics. U.S. Geological Survey Water Supply Paper 1975.

U.S. Army Corps of Engineers, 2010. National Inventory of Dams. U.S. Army Corps of Engineers. Accessed in August 2010 at http://crunch.tec.army.mil/nidpublic/webpages/nid.cfm.

U.S. Census Bureau, 2008. U.S. Census Bureau Cartographic County Boundary Files. U.S. Census Bureau. Accessed in October 2008 at http://www.census.gov/geo/www/cob/index.html.

U.S. Department of Agriculture, 2008. U.S. General Soil Map (STATSGO) NRCS NCGC. Accessed in October 2008 at http://www.soils.usda.gov/survey/geography/statsgo/.

U.S. Department of Agriculture, 2010. Watershed Boundary Dataset (WBD); NRCS. Accessed in July 2010 at http://www.ncgc.nrcs.usda.gov/products/datasets/watershed/.

U.S. Environmental Protection Agency, 1998. National strategy for the development of regional nutrient criteria. 822-R-98-002. U.S. Environmental Protection Agency Office of Water, Washington, DC.

U.S. Environmental Protection Agency, 2006a. Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams. U.S. Environmental Protection Agency; EPA 841-B-06-002 December 2006. Accessed in October 2007 at http://www.epa.gov/owow/streamsurvey/.

U.S. Environmental Protection Agency, 2006b. National Pollutant Discharge Elimination System (NPDES): U.S. Environmental Protection Agency; accessed in June 2006 at http://cfpub.epa.gov/npdes/.

U.S. Environmental Protection Agency, 2008. National Hydrography Dataset Plus (NHDPlus) Home Page. USEPA, USGS, and Horizon Systems Corporation. Accessed in October 2008 at http://www.horizon-systems.com/nhdplus/.

U.S. Geological Survey, 2008a. USGS National Water-Quality Assessment (NAWQA) Program. U.S. Geological Survey. Accessed in October 2008 at http://water.usgs.gov/nawqa/.

U.S. Geological Survey, 2008b. USGS Water Data for the Nation (National Water Information System). U.S. Geological Survey. Accessed in October 2008 at http://waterdata.usgs.gov/nwis/.

U.S. Geological Survey, 2008c. Elevation Derivatives for National Applications (EDNA). U.S. Geological Survey. Accessed in October 2008 at http://edna.usgs.gov/.

U.S. Geological Survey, 2008d. National Land Cover Database. U.S. Geological Survey. Accessed in October 2008 at http://www.mrlc.gov.

U.S. Geological Survey, 2008e. Annual Water Data Reports Home Page. U.S. Geological Survey. Accessed in October 2008 at http://wdr.water.usgs.gov/.

U.S. Geological Survey, 2008g. Water Use in the United States. U.S. Geological Survey. Accessed in October 2008 at http://water.usgs.gov/watuse.

U.S. Geological Survey, 2008h. National Elevation Dataset. U.S. Geological Survey. Accessed in October 2008 at http://ned.usgs.gov/.

U.S. Geological Survey, 2010a. The Hydrologic Benchmark Network. U.S. Geological Survey. Accessed in December 2010 at http://ny.cf.er.usgs.gov/hbn/.

U.S. Geological Survey, 2010b. The National Map: National Hydrography Dataset. U.S. Geological Survey. Accessed in December 2010 at nhd.usgs.gov

Vogel, R.M., Wilson, I., and Daly, C., 1999. Regional regression models of annual streamflow for the United States. Journal of Irrigation and Drainage Engineering 125 (3):148-157.

Wallis, J.R., Lettenmaier, D.P., and Wood, E.F., 1991. A daily hydroclimatological data set for the continental United States. Water Resources Research 27:1657-1663.

Winter, T.C., 2001. The concept of hydrologic landscapes. Journal of the American Water Resources Association 37:335-349.

Wolock, D.M., Hornberger, G.M., Beven, K.J., and Campbell, W.G., 1989. The relationship of catchment topographic and soil hydraulic characteristics to lake alkalinity in the northeastern United States. Water Resources Research 25:829-838.

Wolock, D.M., 1993. Simulating the variable-source-area concept of streamflow generation with the watershed model TOPMODEL. U.S. Geological Survey Water-Resources Investigations Report 93-4124.

Wolock, D.M. and McCabe, G.J., 1995. Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. Water Resources Research 31:1315-1324.

Wolock, D.M. 1997. STATSGO soil characteristics for the conterminous United States. USGS Open-File Report 97-656. Accessed in October 2008 at http://water.usgs.gov/GIS/metadata/usgswrd/XML/muid.xml.

Wolock, D.M. and McCabe, G.J., 1999. Estimates of runoff using water-balance and atmospheric general circulation models. Journal of the American Water Resources Association 35(6):1341-1350.

Wolock, D.M., 2003a. Base-flow index grid for the conterminous United States. U.S. Geological Survey Open-File Report 03-263, digital data set. Accessed in October 2008 at http://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml.

Wolock, D.M., 2003b. Saturated overland flow estimated by TOPMODEL for the conterminous United States. U.S. Geological Survey Open-File Report 03-310, digital data set. Accessed in October 2008 at http://water.usgs.gov/GIS/metadata/usgswrd/XML/satof48.xml.

Wolock, D.M., 2003c. Infiltration-excess overland flow estimated by TOPMODEL for the conterminous United States. U.S. Geological Survey Open-File Report 03-310, digital data set. Accessed in October 2008 at http://water.usgs.gov/GIS/metadata/usgswrd/XML/ieof48.xml.

Wolock, D.M., 2003d. Hydrologic landscape regions of the United States. U.S. Geological Survey Open-File Report 03-145, digital data set. Accessed in October 2008 at http://water.usgs.gov/GIS/metadata/usgswrd/XML/hlrus.xml.

Supplemental Info:

Projection Info:

PROJECTION ALBERS
UNITS meters
SPHEROID GRS1980
DATUM NAD83
PARAMETERS
29 30 0.000 /* 1st standard parallel
45 30 0.000 /* 2nd standard parallel
-96 00 0.000 /* Central meridian
23 00 0.000 /* Latitude of projection's origin
0.0 /* False easting (meters)
0.0 /* False northing (meters)

Contact Info:

James Falcone

jfalcone@usgs.gov

703-648-5008

(September, 2011)