Freshwater Network

Water Quality Application Data Analysis and Application Development Methods

Introduction

The Nature Conservancy was awarded a Natural Resource Conservation Service (NRCS) Conservation Innovation Grant to create a GIS based decision support application that would aid NRCS in identifying potential "hot spots" of poor water quality in Louisiana's watersheds. We agreed that the basic assessment unit for this project would be U.S.G.S 8-digit Hydrologic Unit Code sub-basins. Working with NRCS staff in Louisiana, we developed three parallel workflows within the application that a user would be use to explore various aspects of water quality. The spatial analysis workflow allows users to view "hot spot" maps for each water quality characteristic within each sub-basin for every year for which there were enough data. The temporal trends workflow allows users to graph monthly and yearly mean values for each characteristic at each sampling station. The Impaired Watershed workflow is a visualization of LDEQ's 303b Impaired Watershed list and allows users to explore, impairments, causes and sources in an interactive mapping environment. The Impaired Watershed workflow uses LDEQ basin sub-segments rather than U.S.G.S. sub-basins.

Methods

Data Acquisition

We downloaded water quality sampling records from the EPA STORET Data Warehouse (https://ofmpub.epa.gov/storpubl/dw_pages.querycriteria) for all 57 HUC 8 Sub-basins that have all or part of their area within Louisiana. After reviewing these data, we determined that there are additional samples with the Louisiana Department of Environmental Quality (LDEQ) database that are not included in the STORET database. Additional samples for all LDEQ sub-segments were acquired from LDEQ's Ambient Water Quality Monitoring Data portal (http://www.deq.louisiana.gov/portal/DIVISIONS/WaterPermits/WaterQualityStandardsAssessment/AmbientWaterQualityMonitoringData.aspx) for years spanning from 1995 to 2015.

Data Preparation

All data from STORET was received as comma separated text files that can be read by Microsoft Excel. LDEQ data was received as Excel files. There were many differences in attribute format that needed to be addressed before the datasets could be merged (Table 1). The following differences in field names for sample attributes had to be rectified before a merge could be implemented: "Site#" from LDEQ and "Station ID" from STORET store the station id codes for each sample station. The "Site#" and "Station ID" fields were merged into a new "Station_ID" field. The "Result" field from LDEQ contained both the result value from the sample and the units in which the sample was recorded as a text string. The result value and units needed to be split into separate fields and the value converted to a number. The converted result number was merged with "Result Value as Number" from STORET to form the "result_num" field. The units from LDEQ "Result" were merged with the "Units" field from STORET to form the new "SamUnits" field. Both "Collection Date" from LDEQ and "Activity Start" from STORET are date fields and were merged to form the "samDate" field. Both the "Parameter" field from LDEQ and "Characteristic Name" from STORET contained the water quality characteristic name for that sample. These fields were merged into the "CharName" field and re-formatted to bring the characteristic name from each dataset into alignment. See more about characteristic formatting in the Selected Characteristics section below. Lastly, we converted the Latitude and Longitude coordinates from the STORET dataset to UTM coordinates and merge with the UTM fields from LDEQ to form the "UTM_N" and "UTM_E" fields.

Table 1. Merged Field Names

LDEQ Field Name	**STORET Field Name**	**New Field Name**
Site#	Station ID	Station_ID
Collection Date	Activity Start	samDate
Result (includes result value and units)	Result Value as Number	result_num
Parameter	Characteristic Name	CharName
Result (includes result value and units)	Units	SamUnits
Site	none	Site
UTM_N (acquired from LDEQ_Ambient_WQN_Sites shapefile)	Station Latitude	UTM_N
UTM_E (acquired from LDEQ_Ambient_WQN_Sites shapefile)	Station Longitude	UTM_E

The statewide LDEQ and STORET files were then merged into a single ESRI geodatabase feature class based on the merge rules in the table above. Two more fields, “samMonth” (the month in which the sample was taken) and “samYr” (the year in which the sample was taken) were added to the statewide feature class. Both of these fields were derived from the “samDate” field.

Data Scrubbing

In the course of reviewing the statewide feature class for errors, a number of records were found to have the result “NONDETECT”. All results marked “NONDETECT” or which the result field was left blank were removed from the merged dataset. With the exception of turbidity, which were consistently reported as NTU, sample units were represented as parts per million (ppm), milligrams per liter (mg/L), grams per liter (g/L) or micrograms per liter (ug/L). All units for all characteristics except turbidity were calibrated to mg/L.

Selected Characteristics

We worked with NRCS staff to develop a list of water quality characteristics for assessment. NRCS staff expressed interest in analyzing dissolved oxygen, turbidity, total suspended solids, total dissolved solids, nitrogen and phosphorus. There were multiple forms of nitrogen and phosphorus available in the dataset so these other variants were assessed as well. When possible, multiple related characteristics were aggregated and analyzed as a single characteristic (table 2). The “AMMONIA NITROGEN” characteristic from the LDEQ dataset was combined with “Ammonia-nitrogen” and “Nitrogen, ammonium (NH4) as NH4” from the STORET dataset and analyzed as “Ammonia”. “Inorganic nitrogen (nitrate and nitrite) as N” and “Inorganic nitrogen (nitrate and nitrite)” from the STORET dataset were combined and analyzed as “Inorganic Nitrogen”. “NITRATE+NITRITE NITROGEN” and “NITRATE NO3 (AS N)” from LDEQ were combined with “Nitrate” and “Nitrogen, Nitrate (NO3) as NO3” from STORET and were analyzed as “Nitrate”. “Orthophosphate”, “Phosphate-phosphorus” and “Phosphate-phosphorus as P” from the STORET were combined and analyzed as “Orthophosphate”. This final combined dataset was used for both the spatial analysis and temporal trends analysis.

Table 2. Aggregated Characteristics

LDEQ	STORET	Combined Characteristic
AMMONIA NITROGEN	Ammonia-nitrogen; Nitrogen, ammonium (NH4) as NH4	Ammonia
DISSOLVED OXYGEN	Dissolved oxygen (DO)	Dissolved Oxygen

	Inorganic nitrogen (nitrate and nitrite) as N; Inorganic nitrogen (nitrate and nitrite)	Inorganic Nitrogen
NITRATE+NITRITE NITROGEN; NITRATE NO3 (AS N)	Nitrate; Nitrogen, Nitrate (NO3) as NO3	Nitrate
	Orthophosphate; Phosphate-phosphorus; Phosphate-phosphorus as P	Orthophosphate
PHOSPHORUS (AS P)	Phosphorus	Total Phosphorus
NITROGEN, KJELDAHL		Total Nitrogen
TOTAL DISSOVLED SOLIDS	Total dissolved solids; Solids, Dissolved	Total Dissolved Solids
TOTAL SUSPENDED SOLIDS	Total suspended solids; Solids, Total Suspended (TSS)	Total Suspended Solids
TURBIDITY	Turbidity	Turbidity

Spatial Analysis

The purpose of the spatial analysis workflow is to use existing water quality sampling data to identify "hot spots" of poor water quality. One approach for conducting such an analysis would be to average all values of a particular characteristic for each station and then create hot spot maps for each sub-basin from the averaged data. The limitation with this approach is that since values may have changed over time, averaging could overly smooth out variations between sampling station which would decrease the predictive power of the heat map. Using a single year's data for a watershed also would only provide a snapshot of spatial distribution of values within that year which could be misleading if that year was atypical in some way.

The limitations of the two above mentioned approaches led us to develop "hot spot" maps for each year from 1995 to the preset. This provides users with yearly snapshots but also allows the user to toggle through multiple years to investigate whether the hot areas are just a temporary anomaly or if they persist year after year. It is important to note that stations are not always sampled every year. This means that the suite of stations available for analysis varies year to year, which means the size and shape of the heat map varies as well. This is entirely dependent on the location of the samples in that year.

Sample Selection

Samples were selected for each interpolation run by using a series of nested queries to account for every combination of sub-basin, characteristic and year values found within the combined dataset. First all of the samples for a sub-basin would be selected, then from that query, all samples of a given characteristic, then finally from that query, all samples from a particular year. With 57 sub-basins, 10 characteristics and 21 sample years (1995 - present), there were 11,970 possible sample sub-sets that could be used for interpolation. In reality, not every characteristic is represented in every sub-basin for every year so the resulting number of sample sub-sets is somewhat less than the possible total.

Once each sub-set was identified, we checked to determine that there were samples from at least four stations before continuing to the interpolation process. Once the minimum station criterion was applied, at total of 3,498 sample sub-sets were identified for interpolation. The interpolation method required one value per geographic location, so in cases where there were more than one sample for a single characteristic at a single station during the course of a year, the mean value was calculated for that year. For dissolved oxygen interpolation, only samples from taken from May to October were used since dissolved oxygen levels typically are at their lowest in warm months.

Interpolation

Heat maps of water quality characteristics were generated by Inverse Distance Weighting (IDW). IDW creates a two dimensional raster surface by interpolating the value of each cell based on the values of nearby sample points. The shorter the distance a sample point is from a given cell, the greater the “weight” of that sample in determining the value of that cell while farther away samples have less influence. For this exercise an inverse distance squared weighted interpolation was used meaning that a power of 2 was applied to the distance weighting increasing the influence of near samples and decreasing the influence of far samples.

The number of samples used to interpolate the value of each cell was determined by the total number of stations available for that sub-basin, characteristic and year. If there were 12 or more stations, the nearest one third of stations was used to determine the value of each cell. If the number of stations was greater than or equal to 8 and less than 12, the 4 nearest stations were used. If the total number of stations was less than 8, the nearest 3 stations were used. U.S.G.S. 8 digit HUC boundaries were used as bounding lines for the interpolation so that the output raster was confined to the sub-basin.

Temporal Trends

The spatial workflow provides a powerful and instantly understandable representation of spatial variability of water quality characteristics within watersheds. There are a few limitations that make the temporal trend workflow an important compliment to the spatial one. First, the yearly mean values used to make the heat maps removed any seasonal variation that may have occurred at each sampling station. Secondly, it may be difficult to appreciate whether the on-the-ground conditions are improving through time by looking at the heat maps alone.

In order to explore the overall temporal trends for characteristics within sub-basins we elected to calculate monthly means for each characteristic at each station. We then developed a graphing application that would allow users to click on stations on the map and have the graph display monthly means on a trend line and simultaneously show bar graphs representing the yearly means. This technique provides an understanding of the seasonal variability of the characteristic throughout a given year as well as the general trend over the course of the study period.

Impaired Watersheds

We downloaded LDEQ's 2014 Water Quality Inventory Integrated Report Appendix G: Louisiana's Final 2014 Section 303(d) List as an Excel file. This table cataloged listed impairments for basin sub-segments within Louisiana and the causes, sources and priorities for each. Many of the sub-segments had multiple impairments and therefore had multiple rows in the table. We aggregated all impairments for each sub-segment into a single row which made it possible to display the impairments as a grid. We then joined the re-formatted table to an LDEQ sub-segment GIS layer.