Water and Sediment Quality Status and Trends in the Coastal Bend Area – Phase 1: Data Archiving and Publishing

Publication CBBEP – 77
Project Number – 1105
May 15, 2012

Prepared by:

Paul A. Montagna, Ph.D., Principal Investigator
Kevin Nelson, Ph.D. Candidate
Akshara Uppaluri, M.S.
Harte Research Institute for Gulf of Mexico Studies
Texas A&M University-Corpus Christi
6300 Ocean Drive, Corpus Christi, TX 78412
Phone: 361-825-2040
Email: paul.montagna@tamucc.edu

Submitted to:
Coastal Bend Bays & Estuaries Program
1305 N. Shoreline Blvd., Suite 205
Corpus Christi, TX 78401

Water and Sediment Quality Status and Trends in the Coastal Bend Area – Phase 1: Data Archiving and Publishing

Paul A. Montagna, Ph.D., Principal Investigator
Kevin Nelson, Ph.D. Candidate
Akshara Uppaluri, M.S.

Harte Research Institute for Gulf of Mexico Studies
Texas A&M University - Corpus Christi
6300 Ocean Drive, Unit 5869
Corpus Christi, Texas 78412-5869

Phone: 361-825-2040
Email: paul.montagna@tamucc.edu

Final report submitted to:

Coastal Bend Bays & Estuaries Program
1305 N. Shoreline Blvd., Suite 205
Corpus Christi, Texas 78401

CBBEP Project Number 1105

May 15, 2012

Cite as:

Montagna, P.A., K. Nelson and A. Uppaluri. 2012. Water and Sediment Quality Status and Trends in the Coastal Bend Area – Phase 1: Data Archiving And Publishing. Report submitted to the Coastal Bend Bays & Estuaries Program for project 1105. Texas A&M University - Corpus Christi, Harte Research Institute for Gulf of Mexico Studies, 16 pp.

Table of Contents

List of Figures

List of Tables

Water and Sediment Quality Status and Trends in the Coastal Bend Area - Phase 1: Data Archiving and Publishing

Introduction

The original status and trends report for the Coastal Bend area was completed nearly 15 years ago (Ward and Armstrong 1997a, 1997b), and included analyses of water and sediment data through December 1994, 18 years ago. A status and trends project is actually two projects: one to assemble and organize the data base, and one to perform analyses and write the report. This kind of project can require multiple years of effort, and each time the data base has to be re-invented unless permanent data archives exist. In fact, the original report made four data management and eight data archiving recommendations regarding this that have yet to be implemented. Much change has occurred since 1997 in the world of data management, especially with respect to developing access to data over the internet using web services. Therefore, scientists are currently in a position to finally address the data management recommendations made in the 1997 report.

The current report details progress on Phase 1, to collect and archive data. Data have been collected in the Coastal Bend system for many years, by many different agencies and organizations. The variety of data types collected has also been great. Water properties and dynamics, biological measurements in the water column and sediment and environmental variables have been collected to support various studies of estuarine processes. This current project is an effort to compile these data into a database that will facilitate analyses of the current status and trends in the bay system and identify causal mechanisms for variability.

In a future program, Phase 2 of the project will be an analysis of trends. In addition to analysis, the data compiled from this effort will be made available for use by scientists and the public. An industry-standard database management system has been deployed using Harte Research Institute (HRI) computer assets. This database is regularly backed up and securely-housed behind Texas A&M University - Corpus Christi firewalls. Direct access by the public to the database is prohibited, and instead, access is provided via an HRI web server using standardized web services. Desktop computer software applications able to discover and harvest the data have been identified and while still in development, each of the tools has been shown to query the web services, download requested data, and provide a graphical display of the data.

Methods

The data used for this project exists in two forms, electronic and paper records. Paper records were not of sufficient quality to use optical character recognition, so were hand-keyed into Excel spreadsheets prior to import and transformation. For electronic records, a copy of the source database was procured. Many types of source files were represented: Excel spreadsheets,

Access databases, native SAS tables, and text-based extracts of proprietary databases such as those from the Texas Commission on Environmental Quality (TCEQ) and the Texas Parks and Wildlife Department (TPWD). In few cases were the data definitions for the data sources similar. This meant that each source file was treated individually with little opportunity to reuse code.

The database schema used for this project is that developed by the Consortium of Universities for the Advancement of Hydrological Sciences (CUAHSI). Their Observations Data Model (ODM) provides a standard database structure for storing environmental observations and is an integral part for the CUAHSI Hydrologic Information System (HIS) (Tarboton, Horsburgh et al. 2008). The CUAHSI organization provides users the options to find and get data, analyze data, model data, publish data, and develop tools. It includes components such as Web Services, HIS Central, HydroServer, and HydroDesktop (Figure 1).

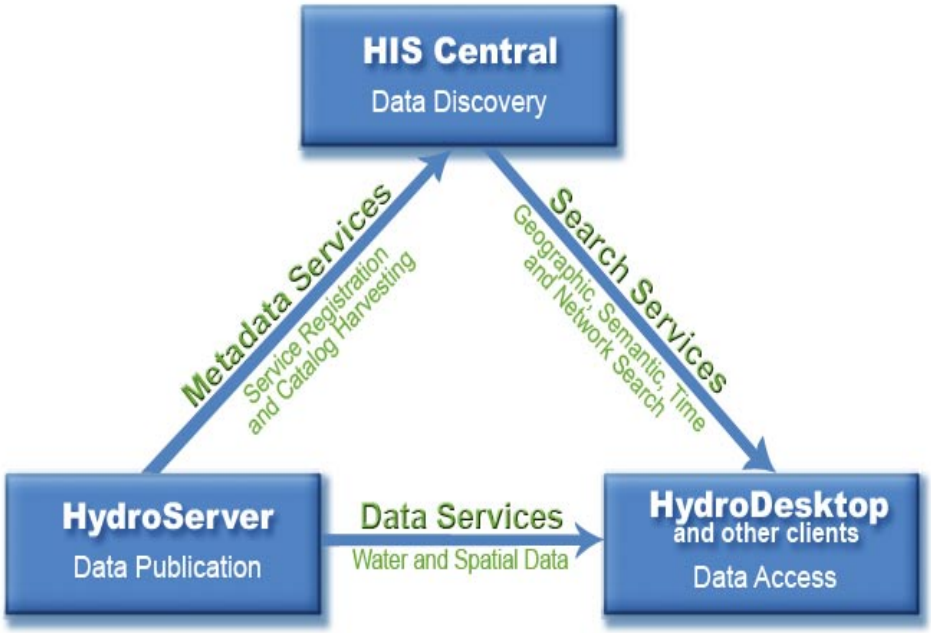

Figure 1 - Components of the CUAHSI HIS.

This database schema was instantiated on Microsoft's SQL Server (MSSQL) database management system. Since SAS is designed to use MSSQL as a data repository, data files were imported into SAS in their native structure using SAS import functions (PROC IMPORT) and the SAS scripting language (SAS Institute 1999), then transformed into the ODM schema using the Microsoft Transact-SQL language. Some limited transformation was accomplished using MS SQL Server Integration Services (Figure 2).

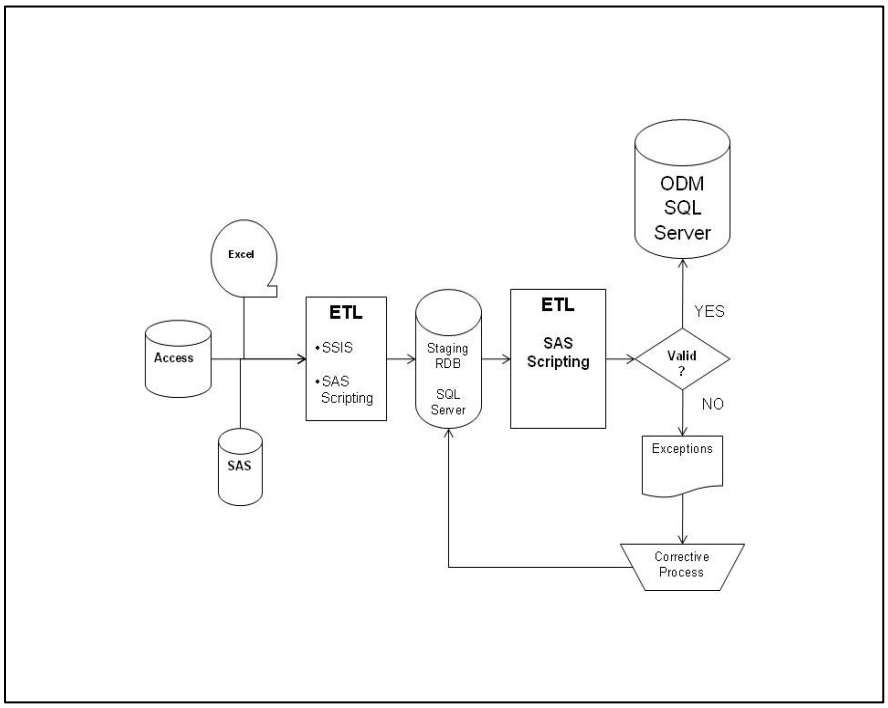

Figure 2 - Extraction, transformation and loading processes from (Montagna and Nelson 2009)

For data security reasons, no direct connections from the Internet to the database server are allowed. Instead, public access to the data is made possible using CUAHSI web services. These services were downloaded from the CUAHSI website: www.cuahsi.org and installed using the user guide provided (Valentine and Whitenack 2008).

Results

Sites

Transformation of the data to date has resulted in 8,938,576 distinct observations, the majority, 7,504,629 from the TCEQ database (Table 1). Observations range in date from 1968 to 2011. In total, there are 8,279 sites in the database. The vast majority of these sites are from the TCEQ database. All sites that have observations of the environmental variables of interest were included in the transformation even if they were not from the Corpus Christi Bay system (Figure 3 and Figure 4).

Table 1 - Listing of observation counts by source description

Source Description	Count
Extraction from TCEQ database	7,504,629
Collection of Continuous sonde observations supporting multiple projects	1,060,998
Extraction from UTMSI Marsh Database	349,318
Collection of nutrient/chlorophyll concentrations for Texas Coastal waters	20,733
Hardcopy text of final report for Texas Water Development Board	2,898

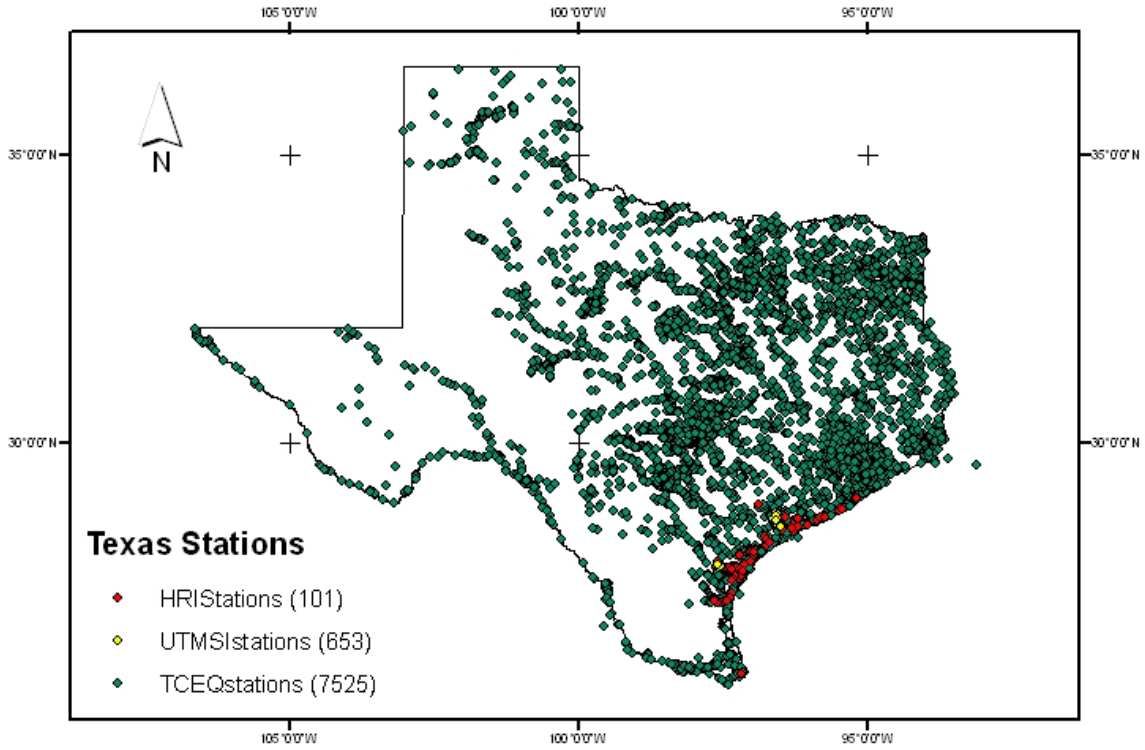

Figure 3 - Sites included in the project

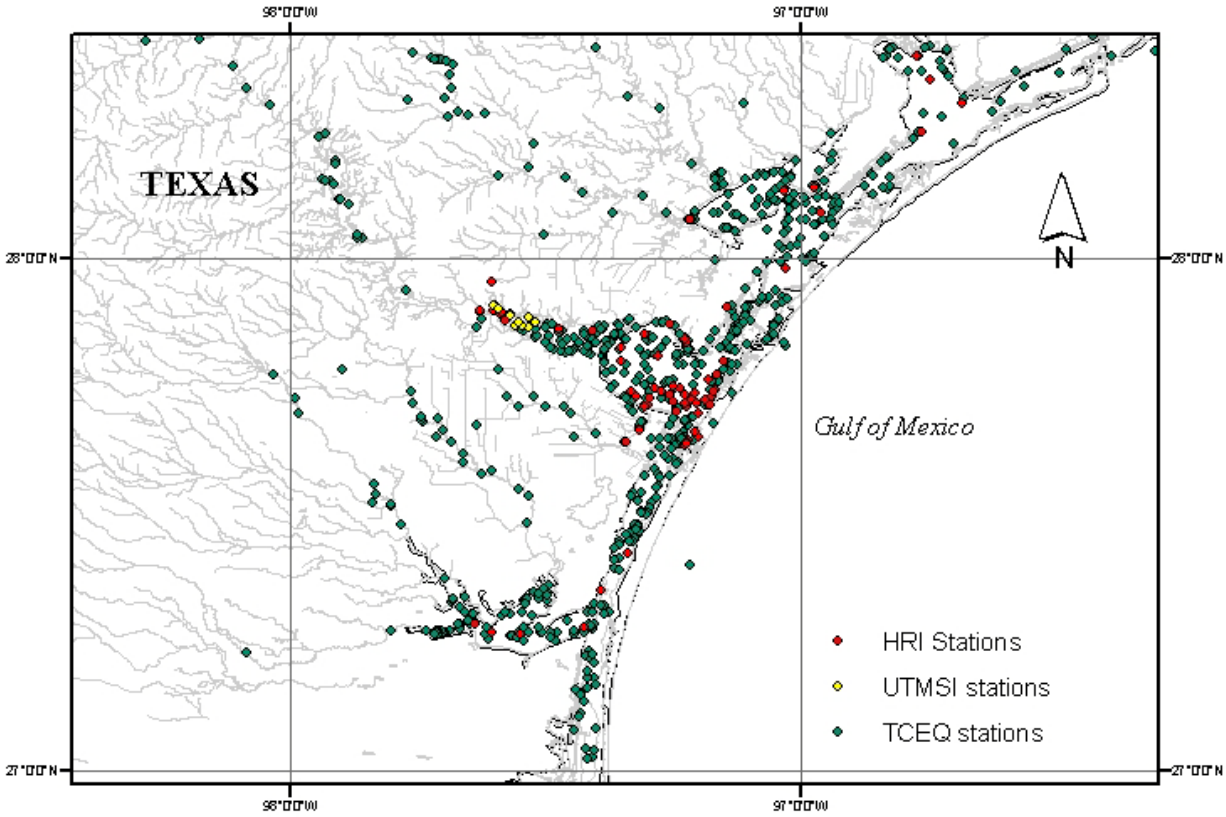

Figure 4 - Coastal Bend Area Stations included in the project

Figure 5 - Example of UT Marsh Database site showing scale and transects.

The UT Marsh Database sites are unique in the project. These sites are arrays of transects consisting of various numbers of sample sites. (Figure 5) The latitude and longitudes for the sites indicate the position of the transect array, not the location of the sample. Interpolation was used to determine the coordinates of each point location on the transects. For this reason, the number of sites listed in Table 1 appears much higher than expected.

Table 2 shows the top 25 sites by the number of observations collected at each site. The list is dominated (16 of the 25 sites) by continuous observations gathered by Dr. Paul Montagna's research on hypoxia (designated as HY) in Corpus Christi Bay. In these instances, 1 to 5 environmental observations were taken at the same time. Because each variable is treated separately in the ODM, each observation is included in the counts even though logically they are part of a larger, multivariable record.

Table 2 - Listing of top 25 Sites by number of observations

Site Name	Count
Site 24 in Bay/Project HY	182,986
Site 19 in Bay/Project HY	100,019
Site 2 in Bay/Project HY	78,931
Site 309 in Bay/Project HY	54,995
Site 41 in Bay/Project HY	53,684
Site 199 in Bay/Project HY	53,679
Lake Travis Near Dam At LCRA Travis County Park	52,881
Site 39 in Bay/Project HY	52,643
Site 310 in Bay/Project HY	51,982
Site 430 in Bay/Project HY	47,362
Site 17781 in Bay/Project HY	40,222
Site 410 in Bay/Project HY	37,592
Site 17787 in Bay/Project HY	36,706
Lake Travis At Arkansas Bend To The West Of Ranch Road 620	36,348
Lake Buchanan Near Buchanan Dam Approx 475 Meters To The West Of Coronado Rd	34,153
Site 17793 In Bay/Project HY	32,716
Site 440 In Bay/Project HY	31,511
Site 18247 In Bay/Project HY	30,944
Site 420 In Bay/Project HY	30,315
Lake Travis Mid Lake Adjacent To Lakeway/To The North Of Corinthian Road	29,257
Lake Lyndon B Johnson Near Alvin Wirtz Dam Approx 658 Meters North Of Fm 2147	26,728
Richland-Chambers Reservoir Chambers Creek Arm Near TCWCID 1 Pump Station 570 M S And 1.16 Km W Of Intersect Of Se 3240 And Se 3250	26,596
Lake Buchanan At Rocky Point Approx 1.3.Km Northwest Of Rocky Ridge	26,254
Lake Travis Mid Lake At Confluence With Cow Creek Arm At Pace Bend Approximately 2.02 Kilometers To The South Of Fm 1431	23,574
Inks Lake Near Inks Dam Approx 161 Meters To The Northeast Of Roy Inks Dam	22,794

Variables

A total of 230 variables have been coded into the database. The vast majority of these are from water quality observations, but variables representing measurements of the biota are also included. While most of the observations in the project so far are from surface water, observations in air, sediment, pore water and tissue are also represented. Table 3 lists the top 25 variables ordered by the number of observations of that variable.

Table 3 - Top 25 sample media variables ordered by number of observations

Variable Name	Count
Temperature	1,056,823
pH	852,000
Oxygen, Dissolved	835,054
Specific Conductance	818,926
Oxygen, Dissolved Percent Of Saturation	286,112
Salinity	270,820
Phosphorus, Total	243,781
Chloride	240,724
Sulfate	230,173
Chlorophyll a	228,330
Streamflow	224,602
Solids, Total Suspended	220,505
Nitrogen, NH3 + NH4	200,485
Coliform, Fecal	195,633
Phosphorus, Orthophosphate Dissolved	195,466
Turbidity	164,002
Solids, Total Dissolved	155,595
Alkalinity, Total	148,108
Carbon, Total Organic	144,101
Secchi Depth	140,694
Nitrogen, Nitrate (NO3)	132,405
Water Depth	129,683
Solids, Volatile Suspended	118,458
Nitrogen, Nitrite (NO2) + Nitrate (NO3)	116,642
Nitrogen, Total Kjeldahl	112,996

Public Access

Access to the data in the project is accomplished via web services. Web services provide a reliable way to discover data, query the database, and display the results. While it is conceivable

that researchers will utilize the web service calls in unique ways, a couple of methods have been developed that are fairly easy to use - HydroDesktop, and HydroExcel.

Web Services

The new web services created for this project can be found at http://ccbay.tamucc.edu/CBBEPDAP_ODWSv11/cuahsi_1_1.asmx. This link can be used by various clients to access the data for this project. Operations provided by the web services are shown in Table 4. Object forms of the operations have a more complicated, but more powerful return. Non-object forms return data streams in WaterML a superset of XML or Extended Markup Language. This list, as well as more information and a service description can be found going to the CBBEP Data Access Project web service address above in any web browser.

Table 4 - CBBEP Web Service Operations

Operation	**Returns**
GetSiteInfo	Site metadata
GetSiteInfoMultipleObjects	Metadata from multiple sites
GetSiteInfoObject	Site metadata (object form)
GetSites	Sites
GetSitesByBoxObject	Sites in a geographical box
GetSitesObject	Sites (object form)
GetValues	Values given site and variable
GetValuesForASiteObject	Values given site and variable(object form)
GetValuesOb ject	Values (object form)
GetVariableInfo	Variable information based on variable code
GetVariableInfoObject	Variable (object form)
GetVariables	Variable
GetVariablesObject	Variable (object form)

HydroDesktop

HydroDesktop is a desktop-based application designed to integrate with the CUAHSI HIS services for data discovery, download, and display (Figure 6). Users can connect to centralized datasets registered with HIS Central, or they can download the observation catalog from any CUAHSI-compatible web service for use with HydroDesktop. Once connected, users can query a geographical region, select multiples sites and variables, and download the selected data series.

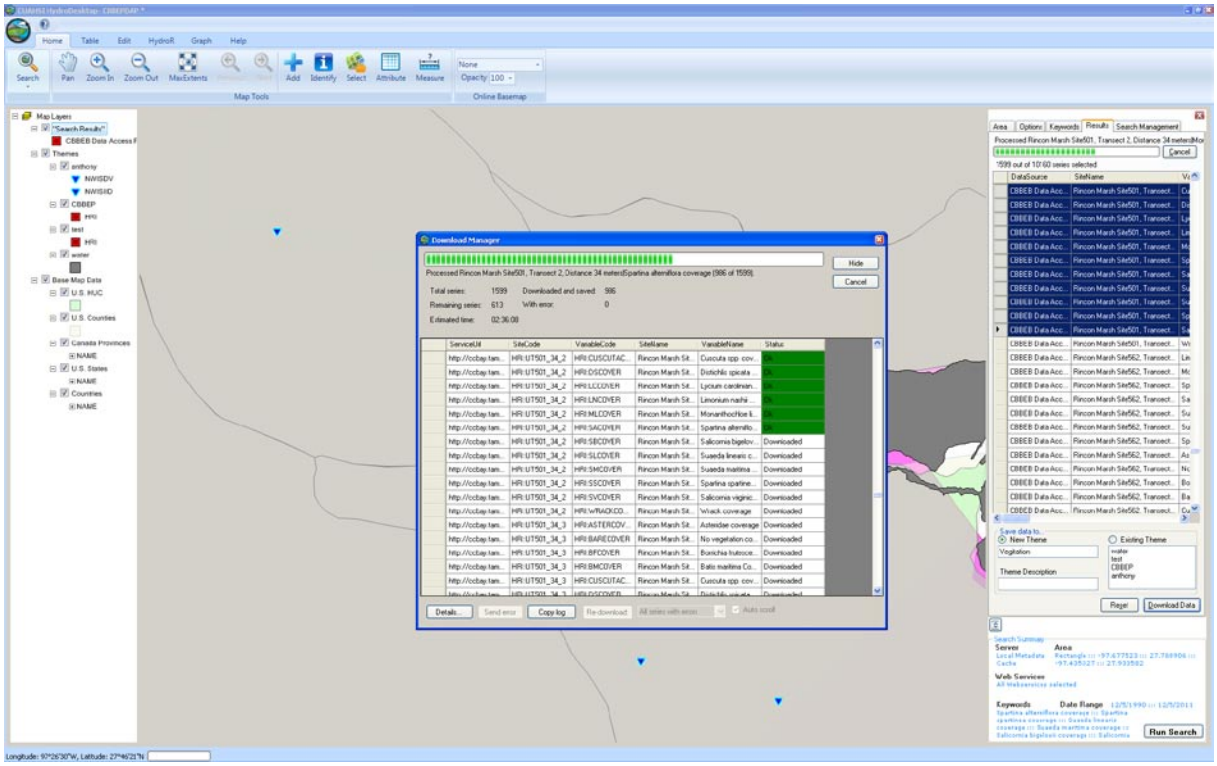

Figure 6 - Screen capture showing selection and download from multiple sites using HydroDesktop

Once data have been selected and downloaded into the user interface, a graphical display of the data can be created. For these screen captures, data from the UT Marsh database were downloaded from the web services. HydroDesktop facilitates the plotting of time series from multiple sources in one graph. For instance, the plot in Figure 7 shows the percent coverage of *Batis maritima* at several points along a transect from the UT Marsh database, overlaid with stream flow data from USGS web services supplied via HIS central.

HydroDesktop can be downloaded free of charge from the CUAHSI website at http://hydrodesktop.codeplex.com/wikipage?title=Getting%20HydroDesktop&referringTitle=Documentation. Thorough documentation for HydroDesktop can be downloaded at http://hydrodesktop.codeplex.com/documentation. Included on this documentation page is a link to a document outlining how to connect unpublished web services to HydroDesktop - Unpublished Web Services Tutorial.pdf. This tutorial outlines the procedure for connecting unpublished services, like the new CBBEP Data Access Project webservices to HydroDesktop. This a slightly more complicated procedure than for HIS Central-published web services, but it enables HydroExcel to connect to any CUAHSI-compliant web services.

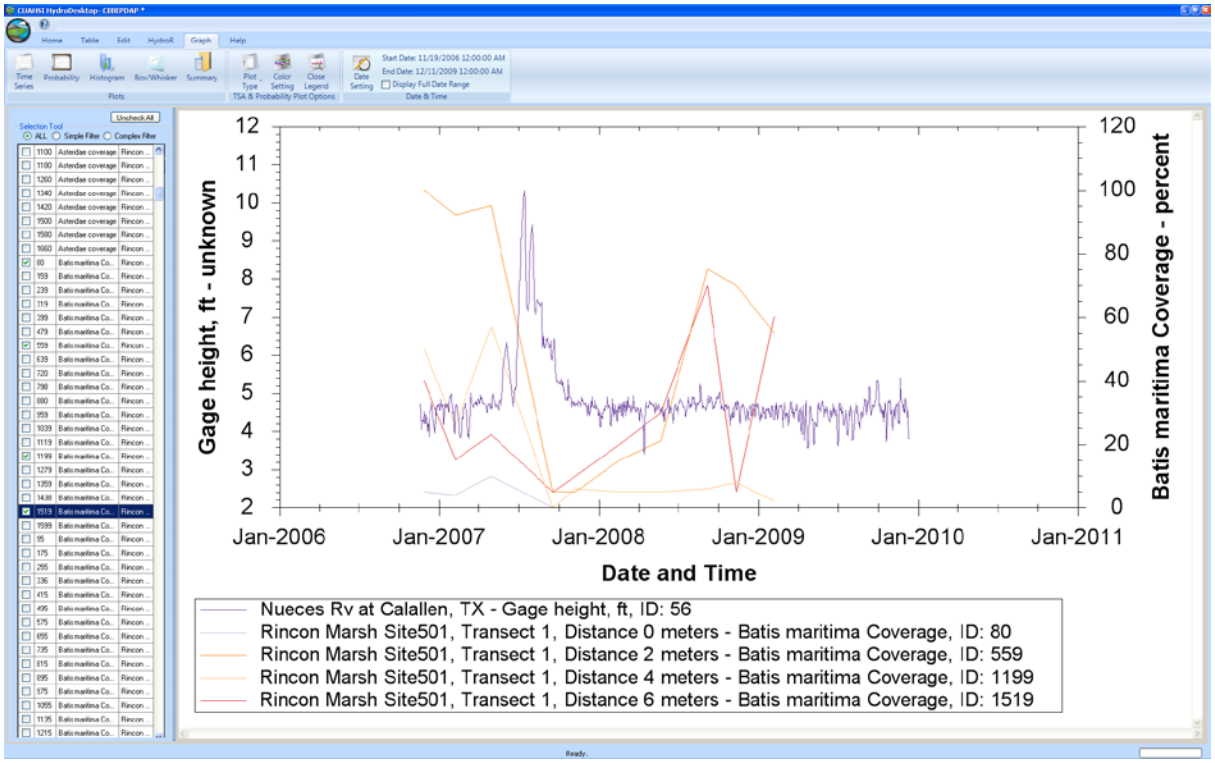

Figure 7 - HydroDesktop screen capture of multiple data source data plot

HydroExcel

HydroExcel is a Microsoft Excel-based desktop application allowing a user to connect to any CUAHSI-compliant web service, and download and display the downloaded data. The functionality is similar to that of HydroDesktop, but not as easy to use and requires more setup. In addition to the necessity of owning a copy of Excel, a separate installation of HydroObjects is required. HydroExcel is provided in either Excel 2003 or Excel 2007 versions and can be downloaded from CUAHSI at http://his.cuahsi.org/hydroexcel.html. Users will also need to download and install HydroObjects from http://his.cuahsi.org/hydroobjects.html. Thorough instructions for use of HydroExcel can be found at http://his.cuahsi.org/documents/HydroExcel_1_1_6_Software_Manual.pdf.

Map-based Web Applications

Web applications designed by CUAHSI, are available for use with HIS Server (Figure 8). These applications remove the requirement of any desktop software except a browser and provide a description of the capabilities of the HydroServer including data regions, and observation and geographic data services (Horsburgh 2011b). Access to query the map application, information about the services, access the time series analysis tool, and data query functions are located at the default website: http://ccbay.tamucc.edu.

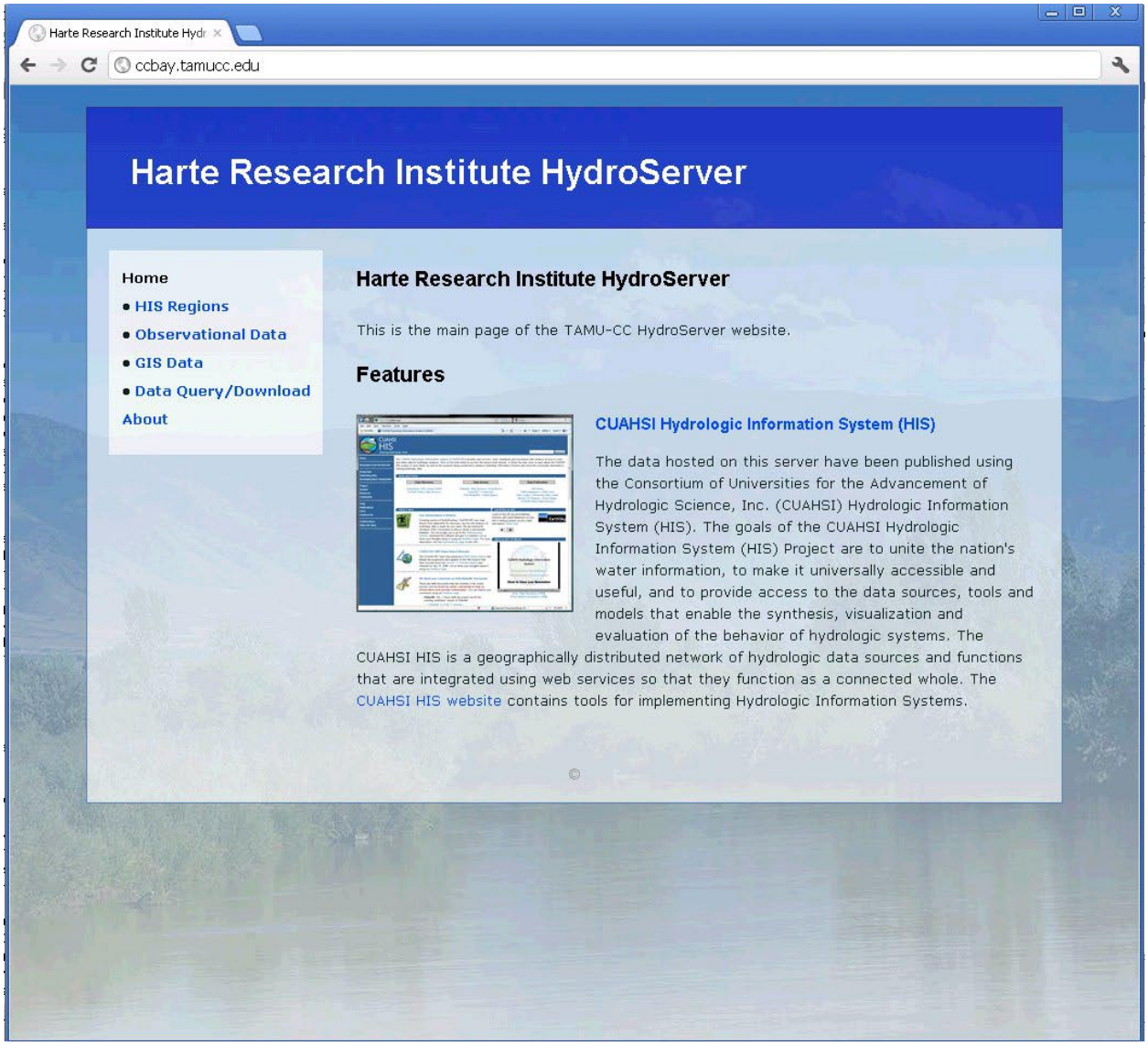

Figure 8 - Default web page for HRI HydroServer Web applications

Selecting a region and launching the map application brings up a web page showing the region and all of the sites available for selection. The base map and site information is delivered by an ArcGIS map service running on the HydroServer and a connection to the ODM database (Horsburgh 2011a). A screen shot of the HydroServer map application displaying Corpus Christi Bay is shown in Figure 9.

Figure 9 - Zoom in of Corpus Christi Bay using the HydroServer map application

The application is fully-functional and allows panning the geographical area and zooming in or out on a region. Once a region of interest has been identified, the user can activate the time series analysis (TSA) tool by clicking the lightning bolt in the top left corner and then selecting a site Figure 10 The TSA will launch with a display of user controls providing site information and a list of variables available at the site.

Once a variable has been selected, clicking <Plot> will create a time series chart of the selected variable as well as summary statistical data. In the example in Figure 10, a chart of continuous values for salinity at HRI site 310 in Corpus Christi Bay are shown. Probability plots, histograms, and box-whisker plots are also available. In the TSA, the user can also view the data series in text form, export it to a local database, or download the metadata for the series into a local spreadsheet. While the TSA is integrated into the map application, it is also a standalone application that can connect to multiple web services on the HydroServer (Horsburgh 2011c).

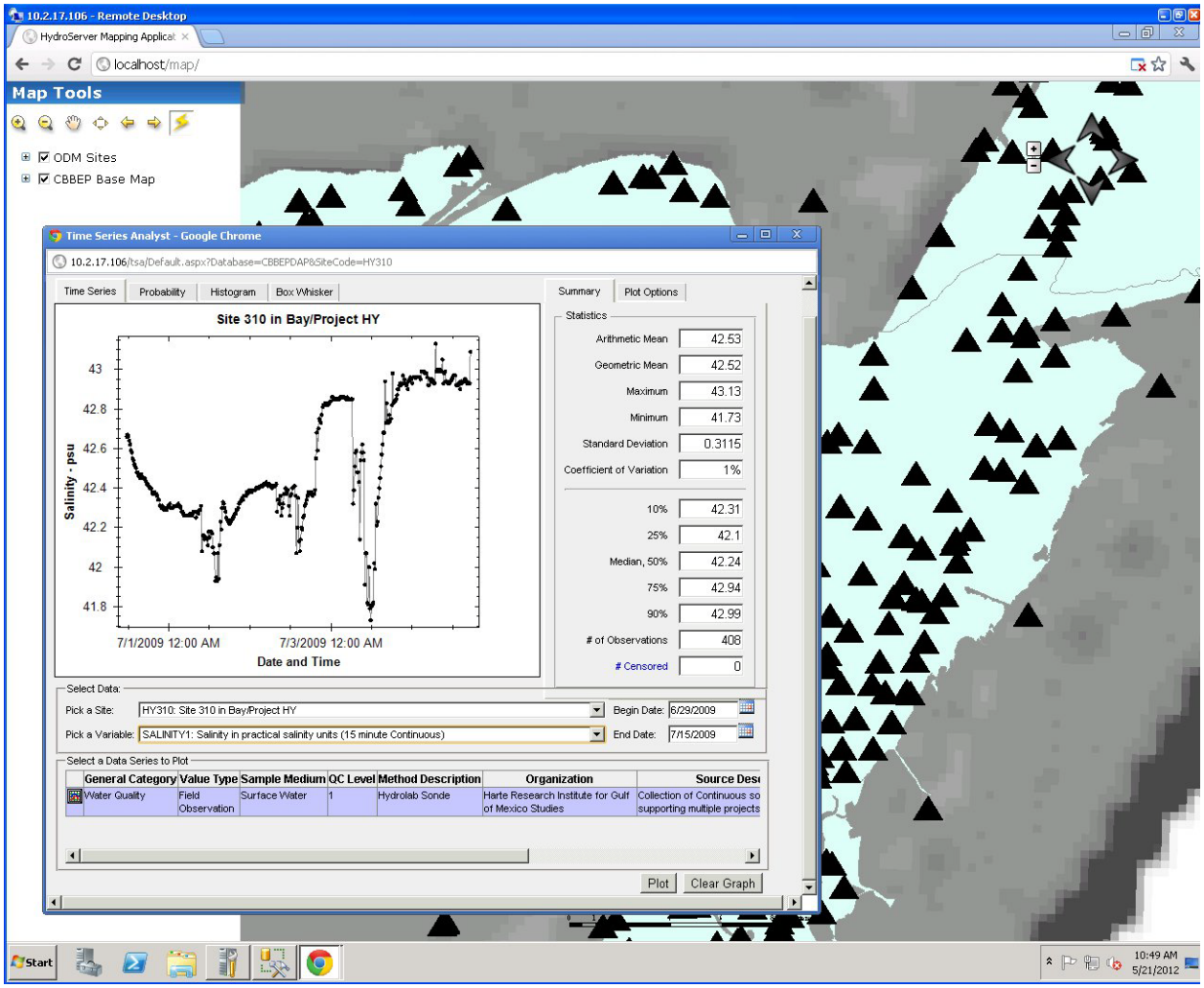

Figure 10 - Time Series Analyst Application showing salinity data from site HY310 in Corpus Christi Bay

The infrastructure utilized for this project also allows for the published web services to be included in CUAHSI's Hydrology Information System (HIS) Central (Figure 11). HIS Central is a data discovery and integration platform where services, like those created for this project, as well as others using this framework can be registered in a centralized catalog. Once registered with HIS Central, the data included in the project would be available in queries by the larger scientific community as part of the entire collection.

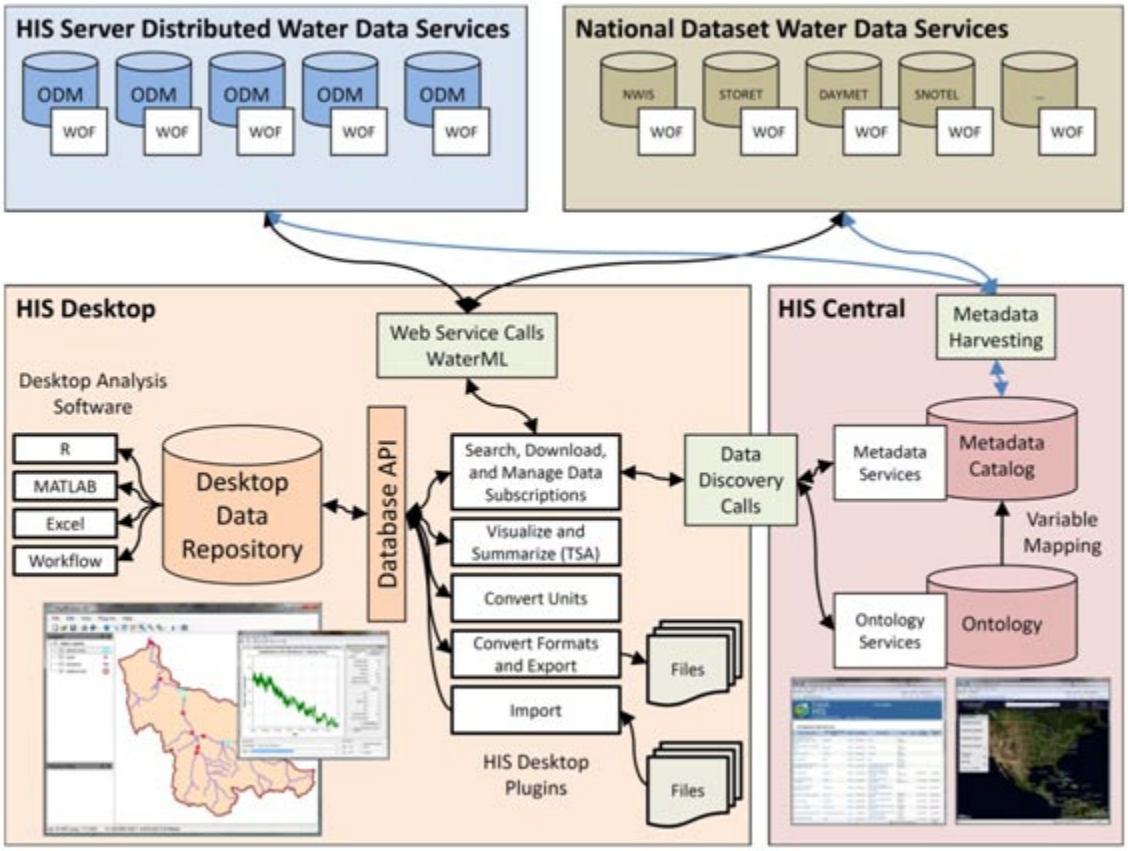

Figure 11 - Development approach for HydroDesktop (http:\\hydrodesktop.codeplex.com)

Citations

Horsburgh, J. S. (2011a). HydroServer Map. A Web Application for Presenting Spatial and Temporal Information on a HydroServer, Utah Water Research Laboratory - Utah State University.

Horsburgh, J. S. (2011b). Hydroserver Website. A Web Application for Presenting the Capabilities of a HydroServer, Utah Water Research Laboratory - Utah State University.

Horsburgh, J. S. (2011c). Time Series Analyst. A Web Application for Visualizing Time Series Data Published on a HydroServer, Utah Water Research Laboratory - Utah State University.

Montagna, P. and K. Nelson (2009). Observation Data Model (ODM) For Rincon Bayou, Nueces Delta. Final Report to the Coastal Bend Bays & Estuaries Program, CBBEP-71, Project Number 8018. http://www.cbbep.org/publicationshome.html

SAS Institute, Inc. (1999). SAS/STAT Users Guide. Cary, N.C., SAS Institute, Inc.

Tarboton, D. G., J. S. Horsburgh, et al. (2008). CUAHSI Community Observations Data Model (ODM) Version 1.1 Design Specifications, Consortium of Universities for the Advancement of Hydrologic Sciences (CUAHSI).

Valentine, D. and T. Whitenack (2008). HIS Document 4: Configuring web services for an observations database (version 1.0). San Diego, CA, San Diego Supercomputer Center, University of California.

Ward, G.H. and N.E. Armstrong (1997a) Current Status and Historical Trends of Ambient Water, Sediment, Fish and Shellfish Tissue Quality in the CCBNEP Study Area. Final Report to the Corpus Christi Bay National Estuary Program, CCBNEP-13, Texas Natural REsource Conservation Commission, Austin, Texas. http://www.cbbep.org/publicationshome.html

Ward, G.H. and N.E. Armstrong (1997b) Ambient Water, Sediment, and Tissue Quality of the Corpus Christi Bay Study Area: Present Status and Historical Trends. Final Report to the Corpus Christi Bay National Estuary Program, CCBNEP-23, Texas Natural REsource Conservation Commission, Austin, Texas. http://www.cbbep.org/publicationshome.html