

Tools for Beach Health Data Management, Data Processing, and Predictive Model Implementation

The overall mission of this work is to provide science-based information and methods that will allow beach managers to more accurately make beach closure and advisory decisions, understand the sources and physical processes affecting beach contaminants, and understand how science-based information can be used to mitigate and restore beaches and protect the public.

The U.S. Geological Survey (USGS), in collaboration with many Federal, State, and local agencies and universities, has conducted research on beach health issues in the Great Lakes Region for more than a decade. The work consists of four science elements that align with the USGS Beach Health Initiative Mission: real-time assessments of water quality; coastal processes; pathogens and source tracking; and data analysis, interpretation, and communication. The ongoing or completed research for the data analysis, interpretation, and communication topic is described in this fact sheet.

Background

Poor recreational water quality can adversely affect human health through exposure to pathogens (disease-causing organisms). To address this risk to public health, the U.S. Environmental Protection Agency (EPA) Beach Act was founded with specific aims to establish water-quality monitoring programs, improve understanding of beach health, clean up U.S. coastal recreational waters, and put forth a criterion for posting advisories or closing beaches during periods of poor water quality (U.S. Environmental Protection Agency, 2000). Beach managers face several challenges with respect to issuing these advisories and closures:

- *Traditional water-quality testing is slow.* Most beach managers collect water samples and analyze them for fecal indicator bacteria (FIB), such as *Escherichia coli* (*E. coli*) or Enterococcus, which is a measure of the water quality at the beach. This analysis can take up to 24 hours for accurate results, and therefore may not provide adequate results for timely decisions. This may result in beach advisories that are issued too late to protect visitor health, or unnecessarily issued the next day, after the water quality has improved.
- *Monitoring is expensive.* Resources of local agencies that are responsible for monitoring are limited, making it difficult at times to implement adequate monitoring programs for assessing beach water quality.

- *It is difficult to efficiently access data.* Environmental data can come from a wide variety of sources and in a variety of formats. Finding and compiling the available data necessary for studying and predicting beach water quality can be a very time-consuming task.

USGS scientists have developed empirical models in order to predict *E. coli* concentrations (a surrogate for water quality) in nearshore waters based on environmental data (rainfall, water current, turbidity or water clarity, and temperature). These predictive models of beach water quality can address some of the challenges faced by beach managers with respect to issuing advisories and closures. To perform predictive modeling efficiently, data must be collected from multiple sources, aggregated, and processed before being used in a model. Real-time or near real-time monitoring data is required for use in these models, so that beach managers can assess the water quality of their beach as close as possible to the time they need to make a decision about a closure or advisory. Many modelers are moving toward model implementations that predict beach water quality based on environmental data from external sources.

This fact sheet describes USGS tools created for efficient data management (data entry, access, aggregation, and processing) and predictive modeling to assist beach managers in making decisions to protect public health. The Environmental Data Discovery and Transformation

(EnDDaT) Web service accesses, identifies, compiles, and sorts environmental data from a variety of sources. The Great Lakes Beach Health Database (GLBH-DB), which is comprised of both a relational database and a Web application, is a tool designed for flexible input, export, and storage of beach water-quality and sanitary-survey monitoring data. A real-time predictive modeling strategy has been implemented that incorporates both tools. USGS scientists also have developed a concept for a web-accessible geographic information system (GIS)-based tool for beach data visualization and summary.

Environmental Data Discovery and Transformation

Efficient data discovery, aggregation, and processing can be a barrier to sophisticated environmental modeling efforts. As environmental aquatic models have become more intricate and comprehensive, the amount of data necessary to build and run the models has increased significantly. An increase in available environmental data could lead to more reliable models, but only if scientists are aware of the data and can gather and make efficient use of the data. EnDDaT (<http://cida.usgs.gov/enddat/>) is a Web tool used to identify, access, process, and compile data from our natural environment. It is a multi-disciplinary data integration framework currently implemented for projects and models that require real-time or historical climate, hydrology,

and coastal hydrodynamics data. The end result is that environmental data from multiple sources are sorted into a single table (see “EnDDaT Options” section). EnDDaT was developed with the capabilities of retrieving publicly available data resources through standard Web services, aggregating the disparate data sources, and processing the data through a single Web-accessible user interface. In addition, the tool provides a variety of output formats and data visualization tools.

The initial EnDDaT application was built to assist Great Lakes beach managers in retrieving historical data for development of statistical models, as well as executing predictive models of beach water quality using real-time data. It was used successfully during summer 2012 in running 17 operational models, which assisted beach managers in posting swimming advisories at Great Lakes beaches owing to exceedances of EPA recreational water-quality criteria.

User Interface

The initial EnDDaT data discovery page has an interactive map, which allows users to create a ‘Project Location.’ From that choice in location, the user can specify a bounding box (in miles) to search for data, and identify which data sources are available. Once those parameters are set, locations where data are available are displayed on the map, and available parameters and temporal extents are listed on the sidebar (fig. 1).

On subsequent pages of EnDDaT, users can refine their requests, specify the dates and times for which they want data, and request post-processing (see “EnDDaT Processing” section). If a user only wants data with the same sample dates and times as a field monitoring sample, or as close as possible, they also may upload these dates and times to get datasets with matching time stamps. There are several options for using EnDDaT data, including export of data in comma or tab delimited data tables, visualization in interactive graphics, and a Web service accessed through a Web address or uniform resource locator (URL), which may be called or executed manually or automatically by custom programming scripts.

The USGS has worked with the EPA to facilitate use of EnDDaT data in Virtual Beach, which is a software package for building and evaluating beach water-quality predictive models (U.S. Environmental Protection Agency, 2012; <http://www.epa.gov/ceampubl/swater/vb2/>). This includes importing an EnDDaT-formatted file into Virtual Beach version 2.0 and earlier or direct import from EnDDaT into Virtual Beach version 3.0 (see “Predictive Modeling” section).

Data visualization (fig. 2) can be useful for data exploration and simple analysis. There are options on the EnDDaT visualization page to refine the plotted data, zoom and pan, and create linear or logarithmic y-axis scaling. The customized image can be saved (using the ‘Create Image’ button), or the data can be downloaded directly into a text file.

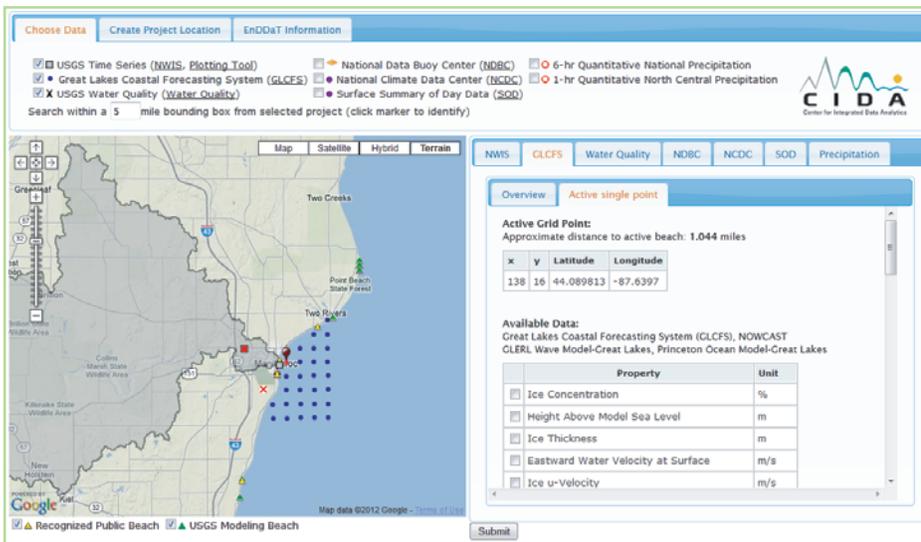


Figure 1. Screenshot of the data discovery page in the Environmental Data Discovery and Transformation (EnDDaT) Web service. The user may choose data to be retrieved from tabs on the right.

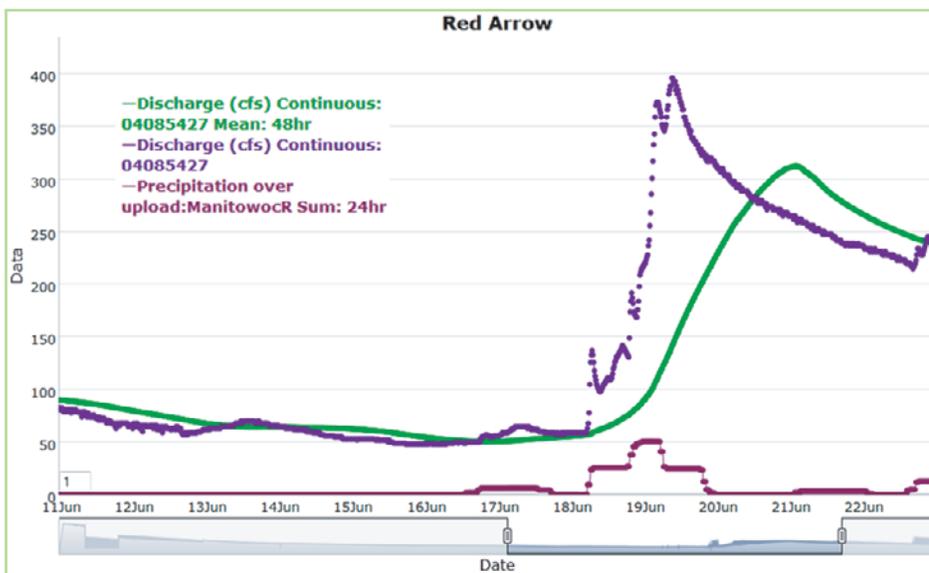


Figure 2. Screenshot of the data visualization options in the Environmental Data Discovery and Transformation (EnDDaT) Web service. (cfs, cubic feet per second)

EnDDaT Data Resources

The EnDDaT framework is designed to access multiple and diverse datasets and to compile them into one data table. EnDDaT connects to various data sources, merges them together, and provides instantaneous post-processing on chosen data.

The following data currently are available in EnDDaT; additional data sources can be added in the future:

- National Water Information System (NWIS): USGS time-series data.
- Water Quality Portal data (cooperation between USGS and EPA): Discrete water-quality sample data.
- Great Lakes Coastal Forecasting System (GLCFS): National Oceanic and Atmospheric Administration (NOAA) model output data for wind, waves, currents, temperatures, and ice.
- National River Forecasting Centers Precipitation Data: Radar indicated rainfall.

EnDDaT Options

The following processing tools currently are provided (fig. 3):

- Time-Windowed Statistics: Mean, minimum, maximum, summation, difference, maximum difference, and standard deviation over a user-specified time range on a data source can be provided as output. This not only provides a method to combine data from sources that have very different time resolutions (for example, 5-minute data combined with hourly or daily data), but also allows the user to choose the most relevant time period for each individual parameter requested.
- Beach Orientation Calculator: EnDDaT offers an interactive tool to calculate the angle of beach orientation for computation of parallel and perpendicular components of directional vectors. The user can adjust markers on a map to automatically calculate the offset of the beach alignment from north.
- Vector Processing: Using the orientation calculated by EnDDaT, or a user-specified angle, data that are provided as directional vectors such as currents and wind can be converted to components that are parallel and perpendicular to the beach.

For further information regarding EnDDaT, contact Laura De Cicco (ldecicco@usgs.gov).

The Great Lakes Beach Health Database

The GLBH-DB was developed to meet the need for a common and flexible system for input, export, and storage of beach water-quality and sanitary-survey monitoring data. The sanitary survey established by the EPA is a method for investigating contamination at a beach and the surrounding watershed, which contains a standard set of data collected. Data commonly collected for the sanitary survey include physical water-quality parameters such as turbidity and water temperature, as well as other parameters that may have an influence on water quality such as the number of birds at the beach, and amount of algae on the beach. Many beach managers have conducted sanitary surveys at their beaches to assess the baseline water quality and potential contamination sources.

The adaptable nature of the system allows it to be used across States and agencies, regardless of which parameters are collected or which monitoring methods are used. Users are able to create customized input forms, if desired, for each of their beaches each season with multiple monitoring sites and (or) tributaries. The system currently stores more than 5,600 sanitary-survey records from beaches in Wisconsin, New York, Michigan, and Ohio, and is currently available for use by any USGS scientist or collaborator monitoring Great Lakes beaches. Users also are able to view and edit records, and perform custom exports from the system by filtering based on date, monitoring location (beach, monitoring point, or tributary), export format, and file type. Lastly, the system has been integrated with EnDDaT, in order to allow beach water-quality prediction models to be run in real-time, based on collected sanitary-survey data and environmental data from external sources gathered by EnDDaT.

The screenshot displays the 'Beach Orientation Calculator' interface. It is divided into several sections:

- Choose Data and Data Processes:** A header bar with the tool name.
- Available Data:** A list of data sources including 'Precipitation (in) Continuous: 04085427', 'Discharge (cfs) Continuous: 04085427', 'E v_water surface: [16,138]', 'N v_water - surface: [16,138]', 'E v_air: [16,138]', 'N v_air: [16,138]', 'Cloud cover: [16,138]', 'Angle v_water (surface): [16,138]', 'Magnitude v_water (surface): [16,138]', 'Angle v_air [16,138]', 'Magnitude v_air: [16,138]', and 'Precipitation over upload: ManitowocR'. Buttons for 'Add Selected', 'Add All', and 'Refresh' are present.
- Requested Data:** A list of selected data including '24hr Σ: Precipitation (in) Continuous: 04085427', '48hr μ: Discharge (cfs) Continuous: 04085427', 'Ghr μ: v_air, < 17.534°, [16,138]', 'Ghr μ: v_water (surface), < 17.534°, [16,138]', 'Ghr Σ: Precipitation over upload:ManitowocR', '74hr Σ: Precipitation over upload:ManitowocR', 'Ghr μ: Cloud cover: [16,138]', and 'Discharge (cfs) Continuous: 04085427'. Buttons for 'Remove Selected', 'Remove All', and 'Refresh' are present.
- Data Processing*:** A section with a 'Process' table and an 'Over:' dropdown. The 'Process' table has columns for 'Process' and 'Over:'. The 'Over:' dropdown is set to 'Hours'. The 'Process' table includes: Mean (μ) [checked], Minimum (Min) [unchecked], Maximum (Max) [unchecked], Summation (Σ) [unchecked], Difference (Δ) [unchecked], and St. Dev. (σ) [unchecked]. The 'Over:' dropdown options are: 1 hour, 2 hours, 6 hours, 12 hours, and 24 hours.
- Resolve Velocity Vectors with Beach Orientation:** A section with two columns: 'Direction' and 'Orientation'. The 'Direction' column has radio buttons for '90 deg' and '17.534 deg'. The 'Orientation' column has checkboxes for 'Parallel (||)' and 'Perpendicular(⊥)'. Both are checked.
- Output Style:** A dropdown menu set to 'Tab Delimited Text' and a checkbox for 'As Download'.
- Missing Value Fill:** A text input field set to 'NaN'.
- Buttons:** 'Get Data' and 'Generate URL' buttons are at the bottom.

Figure 3. Screenshot of processing options in the Environmental Data Discovery and Transformation (EnDDaT) Web service showing how users can have the data delivered as reported, or processed with rolling means, minimums, maximums, etc., over a user-specified time period.

User Interface

Beach managers are able to create a new customized data entry form for each of their beaches before the season starts (fig. 4). The system allows managers to add, remove, reorder, and organize fields into subsections. This often is useful as many beach managers like to organize their digital data-entry form to mimic their field forms. This flexibility allows different groups to monitor completely

different sets of data and use a form that has only the fields they need. The system has the flexibility to establish a set of monitoring points and tributaries for a beach. It provides a form that allows for entry of monitoring data for the entire beach, for each monitoring point, and for each tributary as necessary. Furthermore, the data are stored in a way that relates the beach data to its respective monitoring point and tributary-monitoring data.

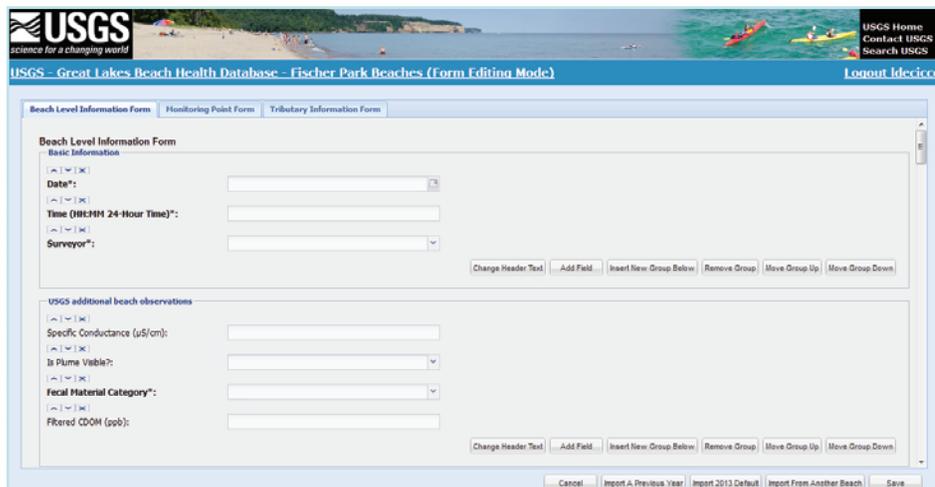


Figure 4. Screenshot of the U.S. Geological Survey Great Lakes Beach Health Database (GLBH-DB) edit-form page.

Great Lakes Beach Health Database Export and Reporting

Detailed data reporting as well as summary reports are available from the GLBH-DB.

Summary reports give beach managers the ability to get an overview of the data records completed for their beaches. Available summary information includes the following:

- Counts of completed records per beach
- Date and time the first and last record was entered or edited for that season
- Data-export options were designed to provide the necessary flexibility for management, quality assurance, and scientific purposes, including the following:
 - Data Format: Users can choose to download data organized per survey or per parameter

- Date Range: Users can optionally choose to filter data by date
- Datasets: Users may choose overall beach data, data from individual monitoring points, or associated tributary data
- Parameter Group: The parameters in the form were classified thematically into groups, and users are able to limit their download to data from one or more of these parameter groups (e.g., Algae, FIB, Physical conditions)
- File Format: Microsoft Excel or text files formatted as eXtensible Markup Language (XML), Comma Separated Value (CSV), or Tab Separated Value (TSV) output

For further information regarding the GLBH-DB, contact Jessica Lucido (jlucido@usgs.gov).

Real-Time Model Implementation

The GLBH-DB was integrated with EnDDaT to run predictive models by automated means. Predictive models, such as those developed using Virtual Beach or other statistical software, draw upon numerous sources of data to predict water quality and assist beach managers with beach-advisory decisions. These statistical models can be incorporated into the GLBH-DB (see “Predictive Modeling” section). Data used to run these models include data collected daily from the beach, such as turbidity and wave height, in combination with hydrological data, climatic data, and hydrodynamic data that are retrieved online using the EnDDaT system. After data are entered into the database using the Great Lakes Beach Web Application, an “execute model” button is available to generate model predictions. Upon execution of the predictive model, model input and resulting predictions are archived in the database and used to evaluate model effectiveness.

Understanding Beach Health Using Geographic Information Systems

The 2005 Great Lakes Beach Health Needs Workshop (Great Lakes Beach Association, 2006) identified improved Great-Lakes-wide data analysis, interpretation, and communication as critical research needs. Specifically recommended was a means to reach beyond political boundaries to map beaches based on their physical and biological characteristics. USGS scientists developed a prototype for a Web-accessible GIS-based tool for beach data visualization and summary. The prototype tool was shared with numerous partners and collaborators, and provided concepts and information to aid in the development of new and more advanced beach data Web-access portals, such as the EPA Beach Advisory and Closing On-line Notification system (<http://watersgeo.epa.gov/beacon2/>) and EnDDaT. Web-accessible GIS maps remain a goal for future development and could encompass multiple scales of analysis—watershed,

shoreline, and beach. Furthermore, these systems could incorporate driving factors for beach pollution, such as tributaries, stormwater outfalls, or land-use analysis, into spatial models of beach-bacteria dynamics. For example, GIS has been used in previous research to relate water-associated illness to spatial patterns and demographics (Turbow and others, 2008) and, in conjunction with antibiotic-resistance patterns, to determine sources of fecal pollution in an estuary (Kelsey and others, 2003). Research by USGS scientists has shown that temporal patterns of FIB may be similar among beaches along adjacent stretches of coastline (Whitman and Nevers, 2008). Recently, USGS scientists reported that this also may be the case for some bacterial pathogens (Haack and others, 2012).

Predictive Modeling

There currently are three methods available to execute predictive models from the tools described in this paper.

Great Lakes Beach Health Database

Beaches that have been configured for data entry and storage within the GLBH-DB can be set up for model execution through a link between the GLBH-DB and EnDDaT (fig. 5).

- Predictive models from Virtual Beach or other linear model-development methods are entered into this system,
- Daily sanitary survey data are collected and entered into the GLBH-DB,
- The “Execute model” option is chosen,
- The GLBH-DB contacts EnDDaT and requests the most current model parameters,
- EnDDaT runs the model, returns predictions to a specified user list via email, and logs results in the database.

For further information regarding linking these tools to predictive modeling, contact Steve Corsi (srcorsi@usgs.gov).

Virtual Beach

Virtual Beach 3.0 offers an option for direct import of EnDDaT data (fig. 6). Using the URL generated by EnDDaT, data can be imported to obtain a prediction result (U.S. Environmental Protection Agency, 2012). Virtual Beach 3.0 should be publically available in early summer 2013.

Web Service to External Models

Using a custom URL to retrieve model-specific parameters, a table can be downloaded from EnDDaT and imported to predictive beach models (including older versions of Virtual Beach) and any other model that can make use of data that are accessible from EnDDaT.

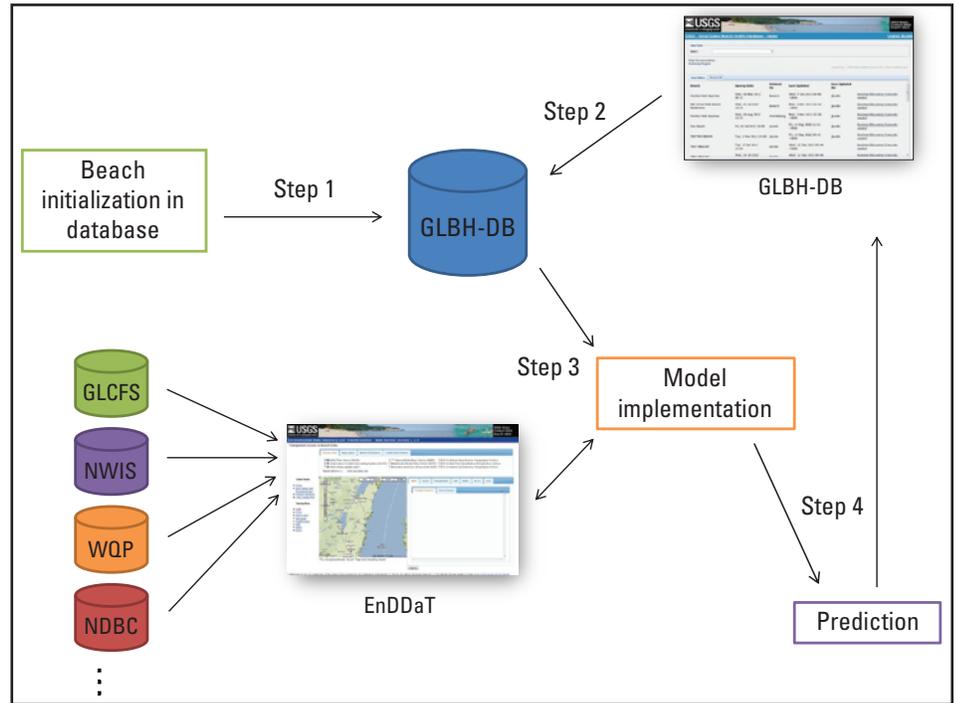


Figure 5. Schematic of data flow for predictive model implementation through the Great Lakes Beach Health Database (GLBH-DB) and the Environmental Data Discovery and Transformation (EnDDaT) Web service. (GLCFS, Great Lakes Coastal Forecasting System; NWIS, National Water Information System; WQP, water-quality portal; NDBC, National Data Buoy Center)

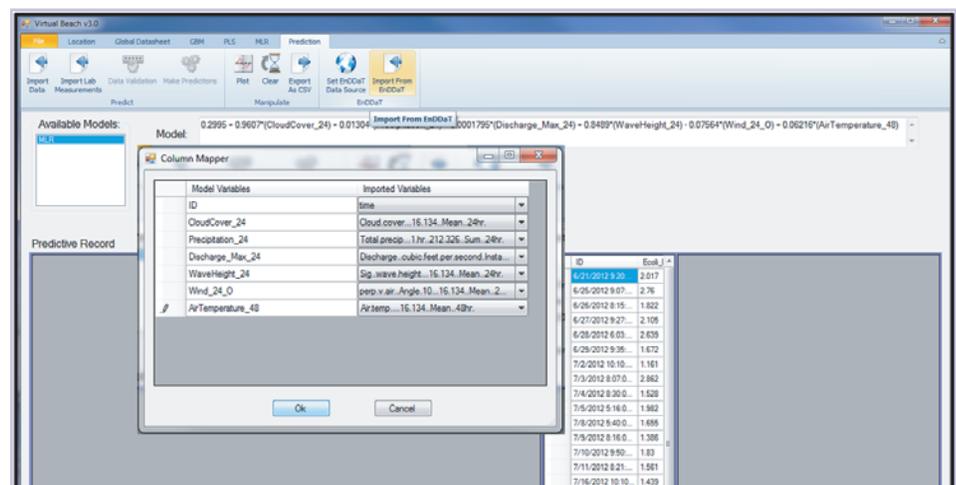


Figure 6. Screenshot of the Virtual Beach 3.0 software tool for direct import of Environmental Data Discovery and Transformation (EnDDaT) Web service data.

Future Work

Plans for future work include continuation of the use of EnDDaT and the GLBH-DB for predictive modeling and water-quality research at Great Lakes beaches. Three methods for incorporation of these tools in predictive modeling will be used: custom web services to retrieve the latest model-parameter values through EnDDaT to be uploaded to models; direct import of model parameters into Virtual Beach from EnDDaT; and execution of predictive models that have been uploaded directly into the EnDDaT system. Enhancements of the EnDDaT system will be made to include additional sources of data for predictive-modeling purposes at beaches and for other environmental research as demand evolves. Other research conducted as part of the same USGS initiative indicates that the geographic setting of Great Lakes beaches has a significant influence on FIB, bacterial-pathogen abundance, and temporal patterns. Functionality to define geographic settings of beaches, such as that of the prototype GIS-based tool, will be important for future studies.

References

- Great Lakes Beach Association, 2006, Great Lakes beach health research needs—Workshop summary: National Oceanic and Atmospheric Administration Technical Memorandum GLERL-138, accessed March 28, 2013, at http://www.glerl.noaa.gov/fip/publications/tech_reports/glerl-138/tm-138.pdf.
- Haack, S.K., Francy, D.S., Fogarty, L.R., Brennan, A.K., Stelzer, Erin, Isaacs-Cosgrove, N., Kephart, C.M., Johnson, H.E., Spencer, C., Fuller, L.M., and Duris, J.W., 2012, Factors influencing bacterial pathogen genes at Great Lakes beaches, in 12th Annual Great Lakes Beach Association Conference, Oct. 16–18, 2012, Mackinac Island, Mich., accessed June 6, 2013, at www.glin.net/glba/pdf/2012conf/Haack-USGSMich.pdf.
- Kelsey, R.H., Scott, G.I., Porter, D.E., Thompson, Brian, and Webster, Laura, 2003, Using multiple antibiotic resistance and land use characteristics to determine sources of fecal coliform bacterial pollution: Environmental Monitoring and Assessment, v. 81, nos. 1–3, p. 337–348.
- Turbow, D.J., Kent, E.E., and Jiang, S.C., 2008, Web-based investigation of water associated illness in marine bathers: Environmental Research, v. 106, no. 1, p. 101–109.
- U.S. Environmental Protection Agency, 2000, Beaches Environmental Assessment and Coastal Health Act of 2000, accessed May 15, 2013, at <http://water.epa.gov/lawsregs/lawsguidance/beachrules/act.cfm>.
- U.S. Environmental Protection Agency, 2012, Exposure Assessment Models—Virtual Beach, accessed March 15, 2013, at <http://www.epa.gov/ceampubl/swater/vb2/>.
- Whitman, R.L., and Nevers, M.B., 2008, Summer *E. coli* patterns and responses along 23 Chicago beaches: Environmental Science & Technology, v. 42, no. 24, p. 9217–9224.

Program Information

Funding for USGS beach projects and research in the Great Lakes comes from a variety of sources including the Ocean Research Priority Plan, the USGS, the Great Lakes Restoration Initiative, the EPA, the NOAA Center for Great Lakes and Human Health, and many State and local partner agencies and organizations throughout the region.



Contact

James R. Morris, Director
USGS Michigan and Ohio Water Science Centers
614-430-7700
<http://greatlakesbeaches.usgs.gov/>