An expandable web-based platform for visually analyzing basin-scale hydro-climate time series data

Joeseph P. Smith [a,*], Timothy S. Hunter [b], Anne H. Clites [b], Craig A. Stow [b], Tad Slawecki [c], Glenn C. Muhr [b], Andrew D. Gronewold [b,d]

[a] Cooperative Institute for Limnology and Ecosystems Research, University of Michigan, Ann Arbor, MI 48109, USA
[b] Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, Ann Arbor, MI 48108, USA
[c] LimnoTech, Ann Arbor, MI 48108, USA
[d] Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, USA

ABSTRACT

Growing demand from the general public for centralized points of data access and analytics tools coincides with similar, well-documented needs of regional and international hydrology research and resource management communities. To address this need within the Laurentian Great Lakes region, we introduce the Great Lakes Dashboard (GLD), a dynamic web data visualization platform that brings multiple time series data sets together for visual analysis and download. The platform's adaptable, robust, and expandable Time Series Core Object Model (GLD-TSCOM) separates the growing complexity and size of Great Lakes data sets from the web application interface. Although the GLD-TSCOM is currently applied exclusively to Great Lakes data sets, the concepts and methods discussed here can be applied in other geographical and topical areas of interest.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Software availability

Names of software product(s): The Great Lakes Dashboard, The Great Lakes Water Level Dashboard, and The Great Lakes Hydro-Climate Dashboard
Developers: National Oceanic and Atmospheric Administration, Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan, USA and Cooperative Institute for Limnology and Ecosystems Research, University of Michigan, Ann Arbor, Michigan, USA
Hardware required: Adobe Flash capable computer with modern system specifications
Software required: Internet browser (Mozilla Firefox, Google Chrome, Microsoft Internet Explorer, etc.), Adobe Flash Plugin
Program languages: MXML and ActionScript under the Apache Flex Framework, compiled under the Adobe Flash Builder and JetBrains IntelliJ IDEA, HTML, JavaScript with jQuery and Dygraphs packages

Availability: All Adobe Flash based products are freely available at the following sites: http://www.glerl.noaa.gov/data/gldb, http://www.glerl.noaa.gov/data/wldb, and http://www.glerl.noaa.gov/data/hcdb
An HTML5 draft version of the Great Lakes Dashboard is available at: http://www.glerl.noaa.gov/data/dashboard/GLD_HTML5.html
A portal to all of these products and to source code for the dashboards is available at: http://www.glerl.noaa.gov/data/dbportal

1. Introduction

The North American Laurentian Great Lakes comprise the largest system of freshwater lakes in the world. In response to changes in the Great Lakes ecosystem, natural resource availability, and hydrology (Morris and Gill, 1994; Ricciardi and Rasmussen, 1998; Wilcox et al., 2002; Gronewold and Stow, 2014), and amidst growing regional and international awareness of global climate change (Stocker et al., 2013) and urban development (Martine et al., 2008), research and monitoring on the Great Lakes system has intensified over the past couple of decades. Consequently, large amounts of data have been and continue to be

* Corresponding author.
E-mail address: joeseph@umich.edu (J.P. Smith).

produced for this region (Bulatewicz et al., 2014). For instance, the National Oceanic and Atmospheric Administration (NOAA), the Canadian Hydrographic Service (CHS), and the Detroit District of the United States Army Corps of Engineers (USACE) are three, among many other, agencies that develop or maintain large amounts of Great Lakes regional data. Developed or maintained data include water level and water budget model simulations, observations, and forecasts, as described in Gronewold et al. (2011) and Deacu et al. (2012). Given the recent acceleration of regional research on the Great Lakes system (e.g. projects funded by the Great Lakes Restoration Initiative), we have observed a coincident rise in demand for these agencies to make their data and model output readily available to the public.

In and outside of the Great Lakes research community, these needs have been documented and attempts have been made to fulfill them − providing access to, and visual analysis of, multi-dimensional data (Jeong et al., 2006). Two notable efforts include the Consortium of Universities for the Advancement of Hydrologic Science, Inc. Hydrologic Information System (CUAHSI-HIS, Horsburgh et al. (2009)) and the Great Lakes Observing System (GLOS, Read et al. (2010)). Both projects focus on providing access to and documentation for point observation data, as opposed to aggregate (e.g. averages − over temporal, spatial domains, or both) data. Access is realized through web services, with CUAHSI HIS implementing a propietary Observations Data Model (ODM, Horsburgh et al. (2008)) via a Relational Database Management System (RDBMS), and GLOS utilizing tools compliant with Open Geospatial Consortium (OGC, see Web References section) standards.

Beyond the need for basic hydrological data management and dissemination, there is a strong demand for simple, inexpensive or free, and accessible tools facilitating exploratory analysis of regional environmental and aggregate data. Such applications could be used to assess initial quality, identify candidate supporting information, and provide decision support (Jakeman et al., 2008). Demand for these tools is enhanced by the existence of rapidly advancing and increasingly sophisticated software packages, such as those employed by the user accessible WeatherSpark.com, as well as Aquatic Informatics' AQUARIUS and KISTERS' WISKI (see Web References section). These web and commercial software applications, respectively, streamline the data discovery, fetching, visualization, and analysis process, making a wide variety of tools and data easily available to users. While CUAHSI HIS, GLOS, and other existing distribution platforms (for example, see Hendler et al., 2012) serve data, they often do not demonstrate that data are readily available and easy to access and analyze. For instance, while web services may provide data in standard formats to users, data insights are not realized without expertise in web scraping, programming, or other technical skill.

In this paper, we introduce the Great Lakes Dashboard (GLD, Fig. 1), a free web application that provides user-friendly tools to explore aggregated time series data, model output, and forecasts for multiple variables potentially affecting the Great Lakes along with background information. Produced by NOAA's Great Lakes Environmental Research Laboratory (NOAA-GLERL) and the Cooperative Institute for Limnology and Ecosystems Research (CILER), the GLD is one of very few, if any, available tools that aggregate multiple time series data sets on the Great Lakes in a basin-wide spatial domain and present them in an accessible interface.

The GLD is driven by a robust, expandable Time-Series Core Object Model (GLD-TSCOM). As a generic framework and programming model similar to MapReduce (Dean and Ghemawat, 2008), programmers can develop applications similar to the dashboard using the GLD-TSCOM. In that sense, it can also be adapted to new requests from users, applied to regions other than the Great

Lakes or other topical areas of interest, and be framed around advances in available technology. As an example, instead of focusing on sub-basins within the Great Lakes (see top of Fig. 2), the GLD-TSCOM could work with hydrological inputs, outputs, and storage within subregions of California, defined by the United States Geological Survey (USGS), currently experiencing extreme drought conditions (Swain et al., 2014).

2. Methods

2.1. Predecessors

The GLD and the GLD-TSCOM were preceded by the Great Lakes Water Level Dashboard, or GLWLD (Gronewold et al., 2013) and the Great Lakes Hydro-Climate Dashboard, or GLHCD (Clites et al., 2014). Prior to their release, many previously available data services advertised as Great Lakes data or containing Great Lakes data were exclusive to the United States, leaving out Canadian data, which are invaluable for Great Lakes basin-wide data and model output. Other previously available applications focused strictly on point measurements, most on land and not over water. Lastly, a subset of project stakeholders desired an application to quickly generate plots for presentations through various media. The GLWLD and GLHCD were designed to begin filling in those gaps in available software applications and web data services.

The GLWLD and GLHCD were developed with a focus on ease of use and ability of the general public to understand data being presented without being overwhelmed (Kelleher and Wagener, 2011; Spiegelhalter et al., 2011). WeatherSpark.com's graphical user interface design was, therefore, a good model for development. The GLWLD and GLHCD were used to test certain features as well as communicate to the public general characteristics of Great Lakes hydrology. Programming methodology for these was originally functional and linear, with the side effect of the data model being tightly interwoven into the code. As a result, the addition of data sets required production of several lines of programming, and as data and improvement requests increased, this development paradigm quickly became unsustainable. The code written to drive these initial dashboards is traditionally called "spaghetti code" (Mikkonen and Taivalsaari, 2007). "Spaghetti code" is programming that is not well organized, contains processes with many unnecessary routes to completion, and is difficult to break up into independent, functional parts for reuse and modification. While such code yields timely results and is useful for prototyping and trial releases, creating a sustainable and expandable application requires careful programming analysis and redesign.

Due to the above issues and additional requests to include better communication of displayed data, an Object Oriented Programming (OOP) approach was adopted, and the GLD and GLD-TSCOM were developed amidst initial success with and widespread adoption of the GLWLD and GLHCD. In addition, the GLD-TSCOM was implemented in new versions of the GLWLD and GLHCD to increase their sustainability and demonstrate the model's applicability to more than one web application.

2.2. GLD-TSCOM

The GLD and its predecessors were built using the free, open-source Apache Flex (or Flex) framework (Apache Software Foundation − Forest Hill, Maryland, USA) under the Adobe Flash Builder (Adobe Systems − San Jose, California, USA) and JetBrains' Intellij IDEA (Czech Republic). Flex is a high level, event-driven (Etzion and Niblett, 2011) framework providing graphical user interface (GUI) objects or components and other behind-the-scenes services (i.e. data, web connectivity) which can be implemented at a high level
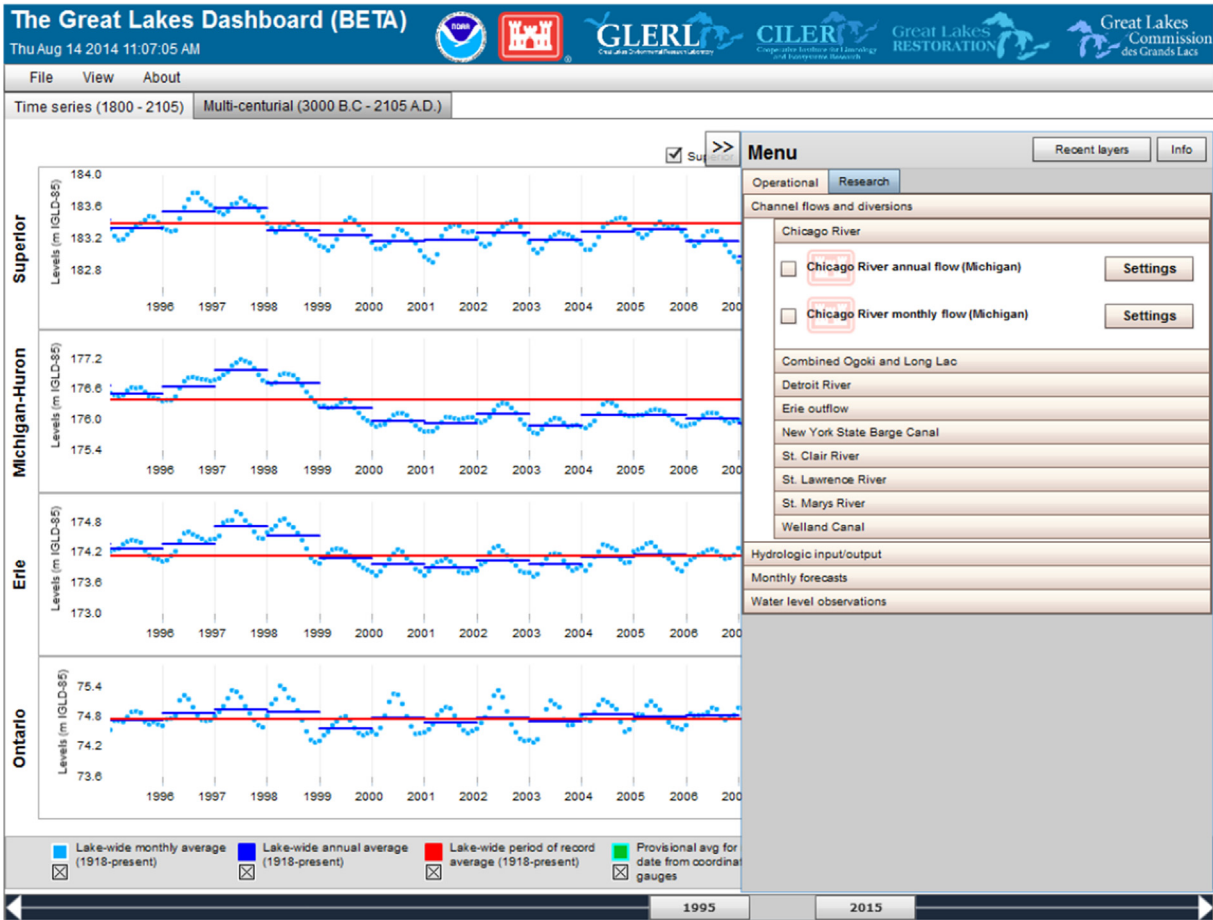
**Fig. 1.** Screenshot of the Great Lakes Dashboard on initial load.

of programming abstraction. Event-driven frameworks, analogous to a series of dominoes set up to fall once one domino is tipped towards others, can be very efficient and fast depending on the quality of engineering. The front-end user interface can be expressed in MXML - Macromedia Extensible Markup Language, a modification to XML — the Extensible Markup Language (Bray et al., 2006), with similar behavior to HTML for web pages. Programmatically-generated user interface components along with efficient responses to user actions are expressed in the object-oriented ActionScript (Gamma et al., 2002; Smith, 2011). Together, packages of MXML and ActionScript in a Flex project are compiled, producing a Small Web Format (SWF) file for display on the web through the client-side Adobe Flash Player, which is available for all popular desktop computer operating systems - a major target platform for the dashboards.

We will describe at a high level the GLD-TSCOM's mechanics. This model may be used as a template for other data series formats (e.g. categorical series). We note that the design of the GLD-TSCOM is heavily influenced by the underlying Flex framework. Manager objects described below and depicted in Fig. 2 manipulate framework-provided chart, series, and axis objects. The properties and behavior of those and other framework-provided objects influenced the design of the managers and ultimately the GLD-TSCOM. Objects in this context are groups of variables, functions, and data structures that are utilized for a single general task.

When a dashboard loads, at least one time series data inventory is read in. Each inventory item is stored as an individual metadata object. Once an inventory is read in completely, the dashboard is notified of that event, and the collection of metadata is broken up by basins, currently defined as Superior, Michigan-Huron, St. Clair, Erie, Ontario, and the entire Great Lakes basin. A regional chart manager object is created for each region, and a single primary chart manager object is given domain over all the regional managers, facilitating communication between the user and the dashboard platform. This allows for administration of user requests, from visible time scale changes to data set selection. User events can be processed via a single call to the primary chart manager, which calls the regional managers to process the request.

The regional chart manager objects contain the metadata for its assigned region, the chart component to display the region's data, a series manager object for handling data sets, a queue for processing data series into the series manager as series objects, and an axis manager object for managing the vertical axes that describe data sets of various units. The series queue keeps an archive of previously processed data in a user session so that costly data processing does not repeatedly occur. Only data initially set to 'on' are processed when the dashboards load. Additionally, as the number of series available can be expanded, the axis manager can be configured to handle many vertical axis types. As of this publication, axes are configured for water level data, percentage data such as ice cover over a lake's surface, hydrologic input and output data, flow rate data for channel flows and diversions to and from the lakes, temperature data, velocity data such as wind speeds, and global climatological pattern or oscillation indices such as the Multivariate El Niño Southern Oscillation (ENSO) Index.
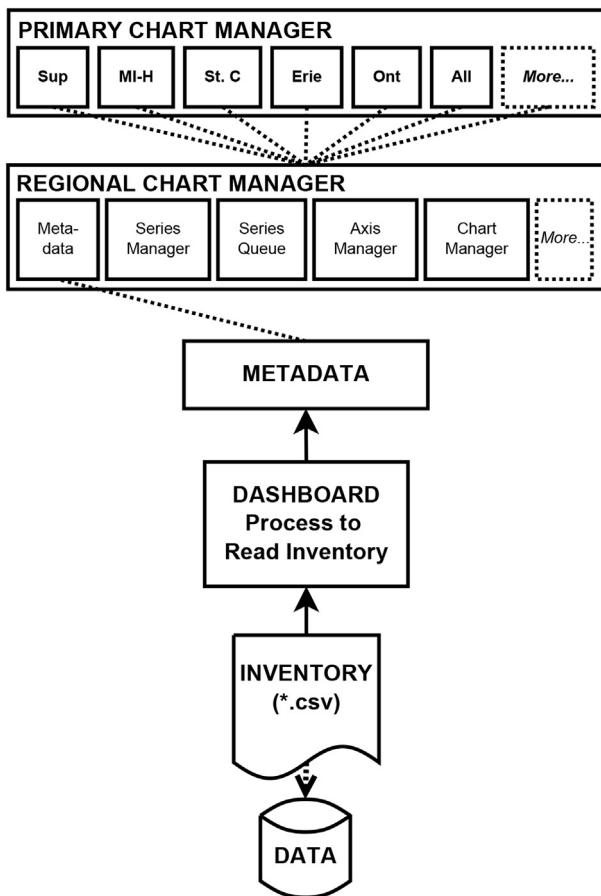
Fig. 2. The GLD-TSCOM and its data management model. The dashboards read comma-delimited inventories, splitting them by their inventory items and transforming those into metadata objects. The metadata objects are then grouped by region and sent to a regional chart manager object, each of which has a set of objects for configuring the charts and the series displayed on them. A primary chart manager has domain over all regional chart managers created, allowing for administration of user requests and facilitating communication between the user and the dashboard.

2.3. Data management

Creating a virtual wall of separation to isolate the growing amount of available data sets and the dashboards is critical to ensuring sustainability for the software. We have, therefore, developed a comprehensive data management model that isolates the growth and complexity of Great Lakes data from the web application interfaces. In this section, we begin by describing the formats of data parsed and displayed by the dashboards, followed by a description of their inventories, ending with discussion of a taxonomic organization scheme employed by the GLD's inventories. An overview of data organization schemes in addition to those described here can be found in Garshol (2004).

2.3.1. Data formats

Time-series data are copied from their source, processed, and stored in the 'data' directory of the dashboard as delimited text files in a variety of formats, with the common delimiter being a comma (a Comma Separated Values file, or CSV). As a majority of the data appropriate for the dashboard can be expressed in tabular format, delimited text files are sufficient and simple, compared to formats more appropriate for non-tabular structured data such as XML or JSON — JavaScript Object Notation (Bray, 2014). Delimited text data can easily be used in a variety of ways by users who download the

data as there is a wide selection of text editors, spreadsheet applications, and data analysis tools available to work with the data. String parsing methods in a dashboard data factory are written to handle the delimited formats, and more methods can be written to handle other formats, including other delimiter formats (tabbed, for example). Currently the following formats are supported:

1. Singleton: a number, usually representing a standard or statistic, followed by the starting year for which that number applies. We assume that the single number applies up to the present day
2. Per month records and averages: 12 lines, with a number, and if they are not averages, the year the extreme record was set (maximum or minimum)
3. Monthly data grid: line for every year of data available, starting with the year, followed by 12 data points for data from January through December
4. Time series: multiple lines which have either a year or date in MM/DD/YYYY format, followed by a single data point
5. Time series of ranges: multiple lines which have either a year or date in MM/DD/YYYY format, followed by two data points, used often for forecast ranges
6. Time series deviation from average: multiple lines which have either a year or date in MM/DD/YYYY format, followed by three data points - the observation for that date, the period of record average, and the deviation of the observation from that average
7. Time series high, low, open, close (HLOC): multiple lines which likely have a year, followed by four data points: the starting data for that year, ending data, the maximum data point for that year, and the minimum for that year. These data are often based off of a set of monthly averages

Data format 6, time series deviation from average, allows viewing of monthly scale hydrological input/output total data with the same data at an annual scale. Because the ranges significantly differ where the monthly total data would be essentially hidden under the full scale of the plain annual total data, it works conveniently to use the deviation from average statistic for annual total data.

2.3.2. Inventories

Time series data inventories consist of line items that represent individual data sets and even data points (i.e. the period of record average water level for a given lake). Each line item contains the following elements, most of which follow case-sensitive controlled vocabularies for machine readability and processing into essential dashboard components:

1. Lake, basin, or region covered by the data
2. General descriptor which can act as an identifier of the data in memory
3. Label for the data tip a user will see when they hover over a data point with their mouse cursor
4. Web directory location (URL) of the delimited text data set
5. Format of the delimited text data set
6. Number of lines to prime the data when reading (to ignore header info)
7. Number of decimal places to display to the user (i.e. for data tips)
8. Metric units of the data
9. Target vertical axis for the data
10. Default display state when the dashboard initially loads
11. How the data set is to be plotted (dots, step-wise, etc.)
12. Plot default fill and/or stroke color and opacity
13. Plot point radius and/or stroke weight

We note that the web directory location can either be specified as a URL relative to the dashboard directory or an absolute URL. This implies that not only can the inventories point to data processed into the dashboard 'data' directory, but they can also point to data external to the dashboard directory. Pointing to external data would require either 1) the data be in an already established format, or 2) the development of a new method to parse a common format that the data are in.

Inventories act as high level tools to control how data are communicated to the public. The dashboards' programming and inventories are intrinsically flexible enough that we can expand these line items to include other information if additional communication is necessary. For example, the GLD inventories' line items have an element for a watermark to display behind a data series item in the menu, indicating the origin of the data (i.e. NOAA, USACE, etc.). Lastly, delimited text files are currently sufficient for inventory data. Spreadsheet applications and text editors are widely available to maintain the inventory, and using additional formats would create an unnecessary conceptual barrier to future maintenance. However, a switch to more structured formats such as XML or JSON could easily be done.

2.3.3. Data taxonomy

To support a wide range of Great Lakes data, the GLD features a three-tier, alphabetically-organized menu generated from inventory items. By clarifying whether a data set is operational or produced in research, and what the data describe, the GLD's menu aids in improving communication and organization as well. It is thus useful for the GLD to expand the general descriptors in inventory line items to include information about data's operational status, category, and potential subcategory. This organization scheme can be visualized as a taxonomy for the data (Fig. 3), whose four levels are used in varying combinations to establish data categories and sub-categories:

1. Status: operational or research
2. Lake basin or region
3. Variable of interest
4. Temporal scale

For example, in the Operational (level 1) tab of the GLD data menu, under the main category of 'Water level observations' (level 3), there are subcategories for annual and monthly average data (level 4), along with a subcategory for statistics and standards over the period of record (level 3). Level 2 is used in placing the data on a specific lake's chart, and not often attached to a single set of data's label. We note that basins or regions can be defined in a custom fashion (thus the 'Other' subcategory in Fig. 3) as done, for example, by the Great Lakes — St. Lawrence River Basin Water Resources Compact and Great Lakes St. Lawrence River Basin Sustainable Water Resources Agreement (see Web References section). They combine the sub-basins of St. Clair and Erie, modify the Ontario basin, and have the St. Lawrence River basin as an additional focus region.

If certain data are not given a sub-category in the inventory, they fall under a generic sub-category labeled 'General' in the GLD menu.

3. Results

The expandable GLD-TSCOM makes many options and features available on the GLD interface (Fig. 4). The GLD's capability to interchange axes of different types facilitates time series visualizations for a wide variety of variables of interest on the Great Lakes, unlike its predecessors. If multiple axes are required, they will appear on both the left and right sides of the dashboard, with the right-hand axes visible when users fold in the data menu (Fig. 4C) using the '≫' button on its top left corner.

Here, we demonstrate the GLD's ability to allow users to explore data sets from two distinct studies.

3.1. Visualizing the impacts of water temperature and over-lake evaporation on water levels

The water levels of the Great Lakes rise and fall in regular seasonal patterns corresponding with rainfall, snow melt, and evaporation. Unlike a typical lake, the surface area of each of these massive inland seas comprises about a third of its drainage area, causing total evaporation to be comparable in magnitude to total precipitation on the lake surface. When the upper Great Lakes (Superior, Michigan, and Huron) experienced a precipitous drop in water levels beginning in 1997, hydrologists sought to explain the sudden decline (Gronewold and Stow, 2014). These lakes stayed below their monthly averages, at times significantly below, for a period of 15 years, including an all-time record low set in January 2013 on Lakes Michigan and Huron.

The GLD can be used (Fig. 5) to compare annual average water levels from 1950 to 2013 for these upper lakes. Lakes Michigan and Huron are shown on one panel because they are considered one lake in terms of water level, joined at the north by the Straits of Mackinac. Annual average modeled surface water temperature and annual over-lake simulated evaporation's deviation from average, from 1950 to 2013, are also displayed. One interesting parallel is the correlation between water temperature and over-lake evaporation, which is especially apparent for Lake Superior. The late 1990's and early 2000's were periods with high water temperatures, dramatically high over-lake evaporation, and consistently low water levels on the northern Great Lakes — particularly for Lakes Michigan and Huron. The GLD makes this time series analysis straight forward.

3.2. Relating a major atmospheric oscillation to regional climate response using the GLD

Climatologists have suggested that the decline in water levels in 1997 and 1998 were caused by a particularly strong El Niño coincident with a warm winter in those years (Van Cleave et al., 2014). The GLD makes it easy to overlay any major climate oscillation with, for example (Fig. 6), Lake Superior water temperature and over-lake evaporation estimates. The timing of both the large El Niño and significant Pacific Decadal Oscillation (PDO, plotted), corresponds with the transition from low to high water temperatures, low to high evaporation, and average to low water levels.

4. Conclusions

Amidst growth in collected data, establishment of large, central data discovery and distribution services, and demand for simple access and visualization of aggregate data, we have developed applications based on a single model that fulfill needs within the Laurentian Great Lakes data user communities. The GLD and its predecessors are flexible, powerful, expandable applications for easy visual analysis of aggregated time series data. We have considered and incorporated the demands of many stakeholders, researchers, and managers in the Great Lakes community. These applications represent the leading edge of big data efforts (Vitolo et al., 2015) — aggregating and visualizing measurements and model outputs delivered from data catalogs and services. Instead of accessing data through scripts or data services of variable accessibility, users of the GLD can visualize and download data through a few clicks and drags of a computer mouse.

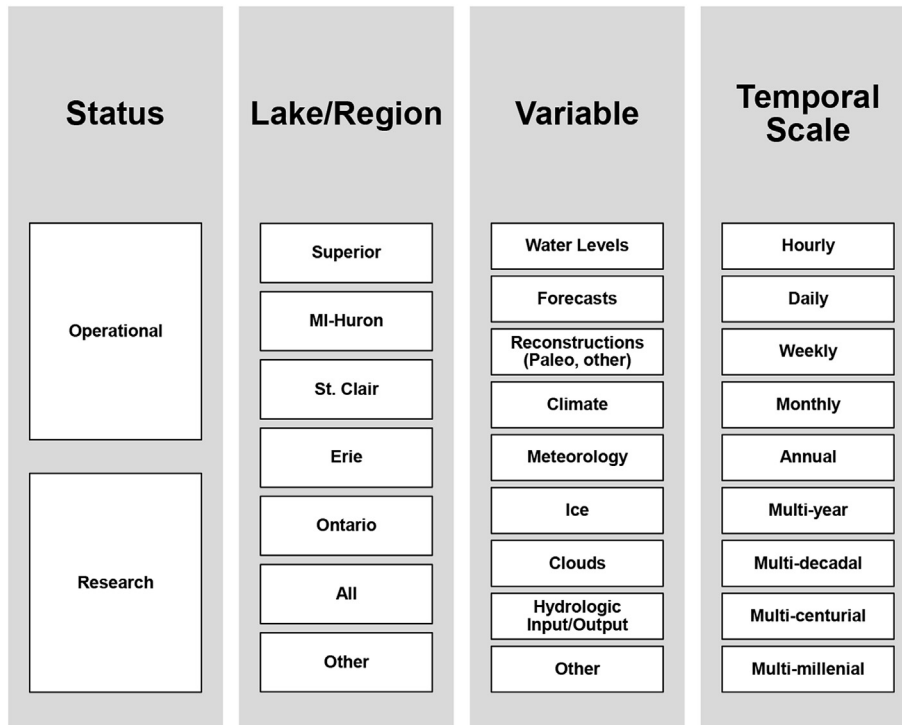Adopting an Agile development approach (Ahalt et al., 2014), the

Fig. 3. Taxonomic structure used to help define primary and secondary categories in the GLD data inventory.
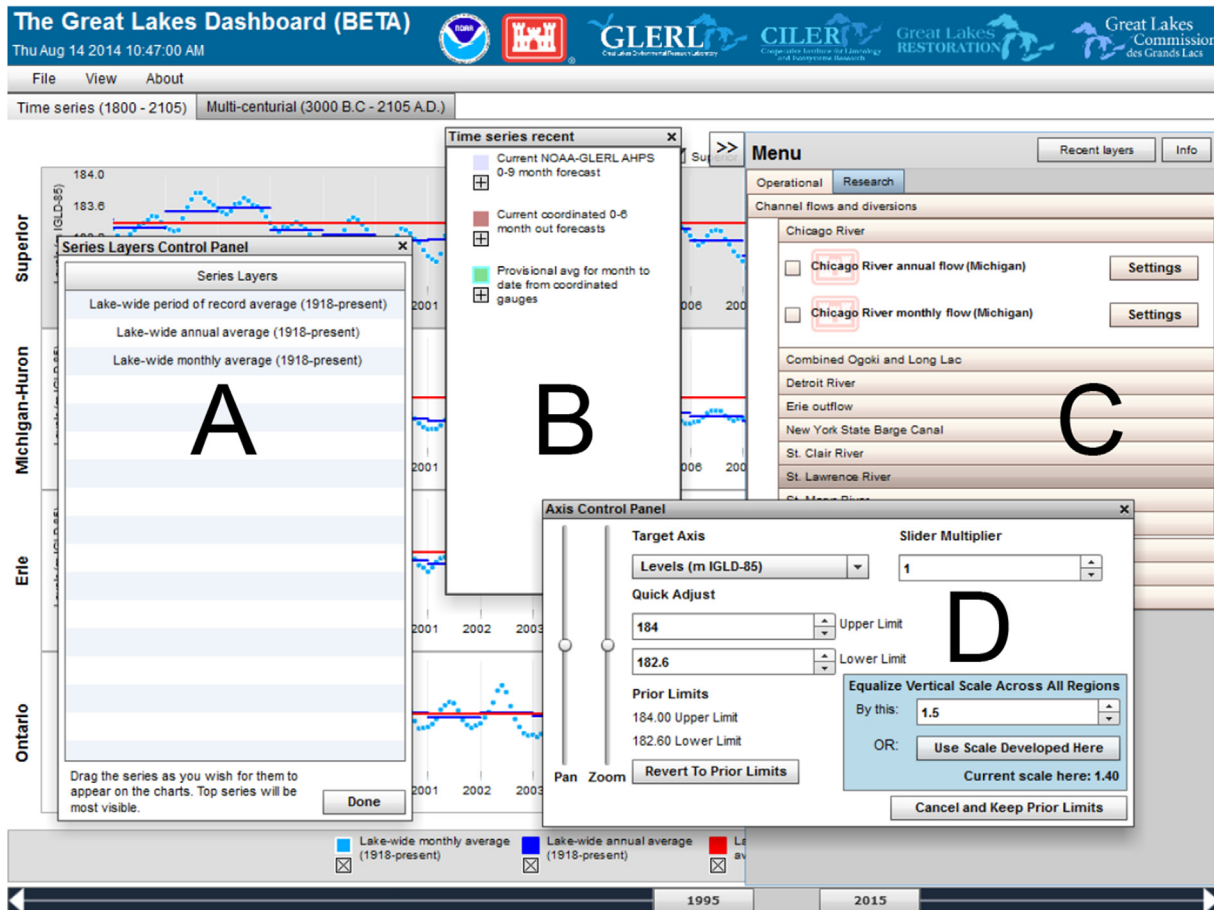

Fig. 4. Visual synopsis of features for the new GLD. A.) The series layers control panel enables users to drag and drop series by their label to configure the order in which they are overlayed. This is aided by the Series Manager in the GLD-TSCOM. B.) Users may open a window to reactivate series which were previously visible in the session. C.) Three-tier, alphabetically organized data menu generated from dashboard metadata fed in via inventories D.) By clicking on a basin's chart, users can adjust the vertical axis scale for active axes via the GLD-TSCOM's Axis Manager. The chart of focus is shaded in gray, which in this case is the chart that represents data from Lake Superior.
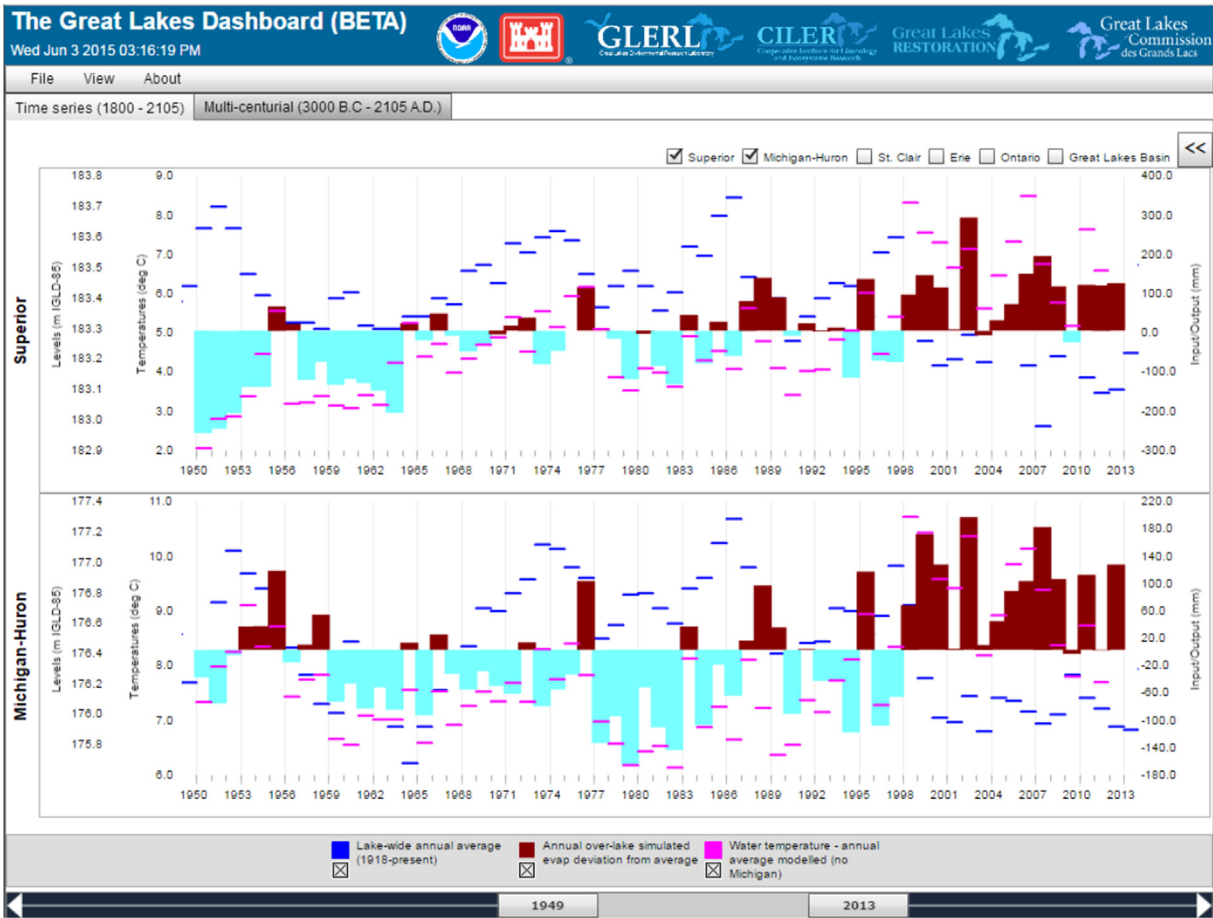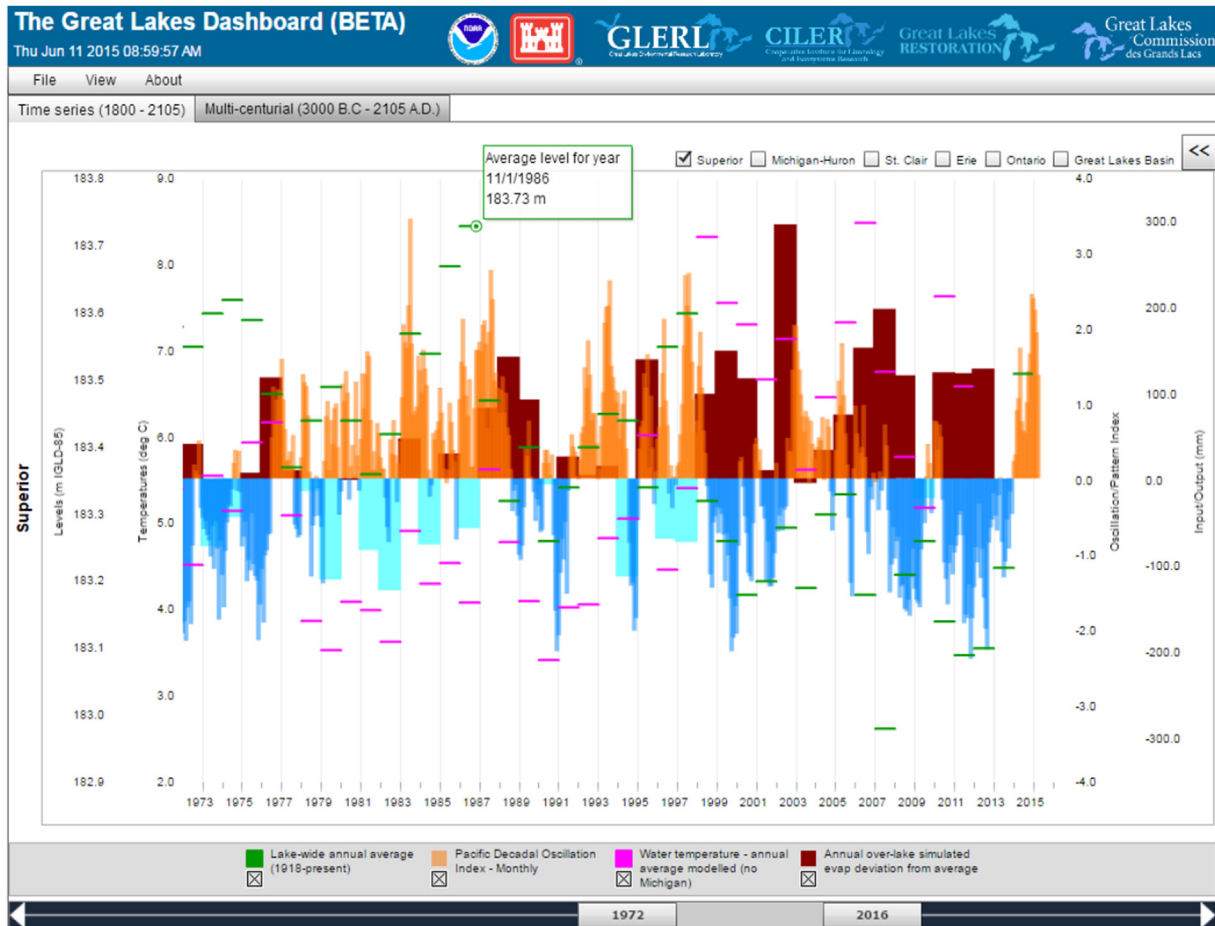
Fig. 5. The GLD displaying datasets — including lake-wide water levels, surface water temperature and evaporation — found in Gronewold and Stow (2014).

GLD-TSCOM can grow in its scope of data and in its capabilities to serve users. This web framework's expandability and its resulting graphical front-end functionalities are useful in light of recent implementations of adaptive management practices (Linkov et al., 2006), such as NOAA's Next Generation Strategic Plan. The flexibility built into the platform leaves open the possibility of utilizing large data services such as CUAHSI-HIS and GLOS — displaying aggregate versions of their data on the dashboards. Additionally, these tools are helpful in the participatory modeling process between stakeholders and modellers as discussed by Voinov and Bousquet (2010), especially when it comes to model analysis, discussion, and planning improvements to models. This platform provides a means by which users can gain perspective and insights that could not be achieved without the juxtaposition of potentially related data.

During the time that the dashboard was developed, available server capabilities and resources were limited both in terms of installed software and support staff. Consequentially, text-based data and inventories were the best option, as updating and maintaining them does not require a specific skillset beyond spreadsheet manipulation. Recent developments, including the instantiation of a PostgreSQL service, may present opportunities to migrate the inventories and data to a more structured, more scalable, and robust system. Such developments would make options, such as the ability to download only the actively displayed data, simple to implement. However, adopting server-entrenched technologies may render the dashboard less transferable. Thus, as of this publication, we have a more lightweight, portable, and easily deployable data system than data infrastructures such as the CUAHSI-HIS.

While the dashboards are built under the Flash platform due to 1.) the availability of powerful, easy to use Integrated Development Environments for the platform and 2.) initial developer experience with application programming, there is a trend towards doing web development under the relatively new HTML5 platform. This is driven by growth in mobile platform adoption, Flash's general incompatibility with mobile devices, and recent exposure of critical and severe security issues in the platform (see the Symantec link in the Web References section). Early in the research and development process, we considered the JavaScript package Flot. While we previously drafted version of the GLWLD using jQuery and Dan Vanderkam's Dygraphs, the resulting performance was insufficient and unable to match what is achieved with the Flex framework. However, using the GLD-TSCOM as a guide, we have developed a draft HTML5 version of the GLD using some of the pseudo object-oriented features of the HTML5 platform (see the Software Availability section). The draft dashboard has a limited feature set since the platform composed of JavaScript and HTML5 is not as innately rich as Flash. Additionally, as there are limitations to the Dygraphs package, we employ a different visualization paradigm - users can view multiple variables' data for a single Great Lakes basin with one chart per variable, or view a single variable type for all the Great Lakes basins with one chart per basin. The variable types available are the same as the axis types employed in

Fig. 6. The GLD displaying datasets also used in Van Cleave et al. (2014), including lake-wide water levels, surface temperature, evaporation, and Pacific Decadal Oscillation (PDO). The "data tip", or small box at the top of the plot, is shown to clarify the green lines are for annual average water level.

the Flash GLD. Despite the limitations, the HTML5 dashboard is capable of performing the similar analyses that its Flash counterparts can, and do them on mobile platforms such as Apple Inc.'s iOS (Cupertino, California, USA).

Acknowledgments

References

Ahalt, S., Band, L., Christopherson, L., Idaszak, R., Lenhardt, C., Minsker, B., Palmer, M., Shelley, M., Tiemann, M., Zimmerman, A., 2014. Water Science Software Institute: Agile and open source scientific software development. Comput. Sci. Eng. 16 (3), 18–26.

Bray, T., 2014. The JavaScript Object Notation (JSON) Data Interchange Format. Available at: http://tools.ietf.org/html/rfc7159 (accessed 13.06.14.).

Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E., Yergeau, F., Cowan, J., 2006. Extensible Markup Language (XML) 1.1, second ed. Available at: http://www.w3.org/TR/2006/REC-xml11-20060816 (accessed 13.06.14.).

Bulatewicz, T., Adresen, D., Auvenshine, S., Peterson, J., Steward, D., 2014. A distributed data component for the open modeling interface. Environ. Model. Softw. 57, 138–151.

Clites, A.H., Smith, J.P., Hunter, T.S., Gronewold, A.D., 2014. Visualizing relationships between hydrology, climate, and water level fluctuations on Earth's largest system of lakes. J. Gt. Lakes. Res. 40 (3).

Deacu, D., Fortin, V., Klyszejko, E., Spence, C., Blanken, P.D., 2012. Predicting the net basin supply to the Great Lakes with a hydrometeorological model. J. Hydrometeorol. 13 (6), 1739–1759.

Dean, J., Ghemawat, S., Jan. 2008. MapReduce: Simplified data processing on large clusters. Commun. ACM 51 (1), 107–113.

Etzion, O., Niblett, P., 2011. Event Processing In Action. Manning, Stamford, CT.

Gamma, E., Helm, R., Johnson, R., Vlissides, J., 2002. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, Boston, MA.

Garshol, L.M., 2004. Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. J. Inf. Sci. 30 (4), 378–391.

Gronewold, A.D., Clites, A.H., Hunter, T.S., Stow, C.A., 2011. An appraisal of the Great Lakes advanced hydrologic prediction system. J. Gt. Lakes. Res. 37 (3), 577–583.

Gronewold, A.D., Clites, A.H., Smith, J.P., Hunter, T.S., 2013. A dynamic graphical interface for visualizing projected, measured, and reconstructed surface water

elevations on the Earth's largest lakes. Environ. Model. Softw. 49, 34–39.

Gronewold, A.D., Stow, C.A., 2014. Water loss from the Great Lakes. Science 343 (6175), 1084–1085.

Hendler, J., Holm, J., Musialek, C., Thomas, G., 2012. US government linked open data: semantic.data.gov. IEEE Intell. Syst. 27 (3), 25–31.

Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I., 2008. A relational model for environmental and water resources data. Water Resour. Res. 44 (5), w05406.

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D., Whitenack, T., 2009. An integrated system for publishing environmental observations data. Environ. Model. Softw. 24 (8), 879–888.

Jakeman, A.A., Voinov, A.A., Rizzoli, A.E., Chen, S.H., 2008. Environmental Modelling, Software, and Decision Support – State of the Art and New Perspectives. Elsevier, Amsterdam, Netherlands.

Jeong, S., Liang, Y., Liang, X., 2006. Design of an integrated data retrieval, analysis, and visualization system: application in the hydrology domain. Environ. Model. Softw. 21 (12), 1722–1740.

Kelleher, C., Wagener, T., 2011. Ten guidelines for effective data visualization in scientific publications. Environ. Model. Softw. 26 (6), 822–827.

Linkov, I., Satterstrom, F., Kiker, G., Batchelor, C., Bridges, T., Ferguson, E., 2006. From comparative risk assessment to multi-criteria decision analysis and adaptive management: recent developments and applications. Environ. Int. 32 (8), 1072–1093.

Martine, G., McGranahan, G., Montgomery, M., Fernandez-Castilla, R., 2008. The New Global Frontier: Urbanization, Poverty, and Environment in the 21st Century. Routledge Earthscan, New York, NY.

Mikkonen, T., Taivalsaari, A., 2007. Web Applications – Spaghetti Code for the 21st Century. Tech. Rep.. Sun Microsystems.

Morris, C.S., Gill, S.K., 1994. Variation of Great Lakes water levels derived from GEOSAT altimetry. Water Resour. Res. 30 (4), 1009–1017.

Read, J., Klump, V., Johengen, T., Schwab, D., Paige, K., Eddy, S., Anderson, E., Manninen, C., 2010. Working in freshwater: the Great Lakes observing system contributions to regional and national observations, data infrastructure, and decision support. Mar. Technol. Soc. J. 44 (6), 84–98.

Ricciardi, A., Rasmussen, J.B., 1998. Predicting the identity and impact of future biological invaders: a priority for aquatic resource management. Can. J. Fish Aquatic Sci. 55 (7), 1759–1765.

Smith, B., 2011. AdvancED ActionScript 3.0: Design Patterns. Apress, New York, NY.

Spiegelhalter, D., Pearson, M., Short, I., 2011. Visualizing uncertainty about the future. Science 333 (1393), 1393–1400.

Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M.M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., 2013. Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York, NY.

Swain, D.L., Tsiang, M., Haugen, M., Singh, D., Charland, A., Rajaratnam, B., Diffenbaugh, N.S., 2014. The extraordinary California drought of 2013/2014: character, context, and the role of climate change. Bull. Am. Meteorol. Soc. 95 (9), S3.

Van Cleave, K., Lenters, J.D., Wang, J., Verhamme, E.M., 2014. A regime shift in lake superior ice cover, evaporation, and water temperature following the warm El Nino winter of 1997–1998. Limnol. Oceanogr. 59 (6), 1889–1898.

Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C.J., Buytaert, W., 2015. Web technologies for environmental big data. Environ. Model. Softw. 63 (0), 185–198.

Voinov, A., Bousquet, F., 2010. Modelling with stakeholders. Environ. Model. Softw. 25 (11), 1268–1281.

Wilcox, D.A., Meeker, J.E., Hudson, P.L., Armitage, B.J., Black, M.G., Uzarski, D.G., 2002. Hydrologic variability and the application of index of biotic integrity metrics to wetlands: a Great Lakes evaluation. Wetlands 22 (3), 588–615.

Web references

Aquatic Informatics AQUARIUS – http://aquaticinformatics.com/products/.

CUAHSI HIS – http://his.cuahsi.org/.

Dygraphs JavaScript charting library – http://dygraphs.com/.

Flot JavaScript Charts – http://www.flotcharts.org/.

Great Lakes Restoration Initiative – http://glri.us/.

Great Lakes – St. Lawrence River Basin Water Resources Compact and Agreement - http://www.glslcompactcouncil.org/Agreements.aspx.

JQuery – http://jquery.com/.

KISTERS WISKI – http://www.kisters.net/wiski.html.

NOAA Next Generation Strategic Plan – http://www.ppi.noaa.gov/ngsp/.

Open Geospatial Consortium – http://www.opengeospatial.org/.

Python – https://www.python.org/.

Symantec: Third Adobe Flash zero-day exploit (CVE-2015-5123) leaked from Hacking Team cache – http://www.symantec.com/connect/blogs/third-adobe-flash-zero-day-exploit-cve-2015-5123-leaked-hacking-team-cache/.

WeatherSpark – http://weatherspark.com/.