Environmental Modelling & Software 78 (2016) 97–105

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

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

ARTICLE INFO

Article history:
Received 16 July 2015
Accepted 5 December 2015
Available online xxx

Keywords:
Data visualization
Great lakes
Data access
Analytics
Software
Object oriented

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.

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).

http://dx.doi.org/10.1016/j.envsoft.2015.12.005

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