

An Extendable Experiment with GIS and ICT to make Environmental Data and Modelling User-Friendly and Accessible

Saurav Kumar¹, Adil N. Godrej², Thomas J. Grizzard³ and Harold E. Post⁴

¹Post-doctoral Associate, (saurav@vt.edu),

²Associate Director and Research Associate Professor,

³Director and Professor

⁴Research Associate

Occoquan Watershed Monitoring Laboratory (OWML), Department of Civil and Environmental Engineering, Virginia Tech, Manassas, Virginia, USA

Abstract: Occoquan Watershed Monitoring Laboratory (OWML), has been monitoring the Occoquan Reservoir and its tributary watershed in Washington DC suburb for over 40 years. OWML has also developed and maintains a state-of-the-art water quality and quantity model for the region based on seven HSPF and two CE-QUAL-W2 applications. To make the data and modelling tools easily accessible to stakeholders, OWML has embarked on an experiment to use modern Information Communication Technology (ICT) along with web-mapping systems. As a result the web-portal <https://wqdata.owml.vt.edu> has been developed to provide easy access to the real-time and historic water resources data for the Occoquan watershed region. Another portal is under development that will allow users from the internet to delineate land use changes through their web browsers and then analyse the impact of these land use changes by executing calibrated water resources models for the region. The technologies used to develop these portals are open-source and have a wide user-base, permitting for easy extension to other regions with water resources data or model to share. Due to reliance on the modern ICT there were some hurdles in development and maintenance of this system and several problems still exist, such as need for constant updates to keep the system relevant. Based on experience in OWML, this paper presents the design of the two portals along with some lessons learnt that may be applicable to the wider community. We have also discussed possible extensions of the system beyond Occoquan Watershed based on cloud computing and collaborative software development.

Keywords: Watershed Modelling; Data Visualization; Decision Support System; Occoquan Watershed.

1 INTRODUCTION

Fuelled by advances in the Internet and the World-Wide-Web (web) such as adoption of HTML5, new advanced web-mapping APIs, better JavaScript (JS) engines, and better and more capable web browsers, there has been tremendous growth in web-based data visualization and modelling efforts in the environmental sector and water sector in particular. A search for the term “web” on the Environmental Modelling & Software journal returned 55, 66, 54,108, and 70 results for years 2010, 2011, 2012, 2013, and first 3 months of 2014; clearly suggesting that the environmental modelling community has started to leverage the internet for delivering modelling tools. Web-based Simulations (WBS) have become popular due to various advantages, identified by Byrne et al. (2010) in their review of web-based simulation and supporting tools, such as ease-of-use, collaboration, wide availability, controlled access, and cross-platform capabilities. To harness similar potential benefits the Occoquan Watershed Monitoring Laboratory (OWML) has started an experiment to develop tools and improve stakeholder accessibility to their water resources data and modelling system.

2 ABOUT OWML

Occoquan Reservoir and its tributary watershed (Figure 1) are located in Northern Virginia near Washington DC. The Occoquan Reservoir is one of the first large-scale water supply reservoirs in the US to implement indirect potable water reuse for the purpose of improving drinking water yield. This novel, for the time, water reuse effort also placed great emphasis on monitoring the water quality of the Occoquan Reservoir. In 1972, Virginia Tech was commissioned to perform an independent monitoring program for the reservoir and watershed system resulting in establishment of the OWML. OWML has since maintained a comprehensive database of the water quality in the reservoir and the tributary watershed (Occoquan system) and also maintains the complexly-linked water resources modelling tool for the Occoquan system based on widely accepted and open source computer modelling software.

2.1 Water Resources Data Collection

OWML operates a network of monitoring stations for collecting water samples manually and automatically through auto samplers. Some of these stations are shown in Figure 1. OWML also performs regular (weekly, bi-weekly) onsite water quality measurement, laboratory water quality testing, along with other periodic and less frequent activities such as bathymetric surveys. Most of the water quality testing is done in-house, using established accredited methods.

Stream sampling stations in the watershed are equipped to perform automatic flow-weighted storm sampling for all storm events using an on-site microcomputer that is integrated with a level-sensing device (see Kumar et al. (2013) for more details about the procedure). Apart from storm events, weekly grab samples are collected from the streams. The frequency of sampling is reduced to bi-weekly in winter. From each station, two separate water samples are collected for analysis of parameters, including OP, TP, TSP, NH₃-N, SKN, TKN, Ox-N, COD, Chlorophyll a, Turbidity, Total Suspended Solids, Total Dissolved Solids, Total Hardness, and selected cations and anions. Weekly stream sampling is augmented by quarterly water sampling for metals (including Cu, Pb, and Zn) and Synthetic Organic Compounds. The reservoir sampling also takes place on a biweekly schedule in winter and weekly for the rest of the year. In the reservoir, water samples are taken from the surface and bottom for analysis. Depth profiles are also constructed, using spot measurement at various depths, of directly sensed water quality parameters such as pH, DO, Temperature, ORP and NO₃-N. This extensive sampling regime aids in high degree of characterization of the Occoquan system and thus enabling better decision support.

2.2 Water Resources modeling

The Occoquan system model is a complexly-linked water quality model, designed by connecting seven implementations of the Hydrological Simulation Program–Fortran (HSPF) (Bicknell et al., 2001) and two implementations of CE-QUAL-W2 (Cole and Wells, 2008) software executed in a serial configuration (Figure 1). Such complexly-linked model applications are expected to mimic the naturally-connected systems better (Xu et al., 2007). Seven HSPF models are used to simulate seven sub-watersheds, one HSPF model representing each sub-watershed, and two CE-QUAL-W2 models are used to simulate the two major waterbodies—the Lake Manassas, and the Occoquan reservoir. Models are executed in series and models that rely on output of the upstream models are executed after all upstream model dependencies have finished execution. A sample Occoquan system model execution will combine the results obtained from HSPF runs for the Upper Broad Run and the Middle Broad Run sub-basins and use it as input for simulating water quality in Lake Manassas with the 2-D reservoir model CE-QUAL-W2. The output from CE-QUAL-W2 for Lake Manassas and HSPF execution for Cedar Run sub-basin is used as input for Lower Broad Run sub-basin HSPF execution. Finally, the Occoquan Reservoir is simulated by CE-QUAL-W2 using results of HSPF runs for the Lower Occoquan, the Middle Broad run, and the two Bull Run sub-basins as input (Figure 1). The choice of models, HSPF and CE-QUAL-W2, was based on several criteria such as use of an open source software, public maintenance and support of models, wide acceptance of the models, publically-defendable modeling scheme, and state-of-the-art science.

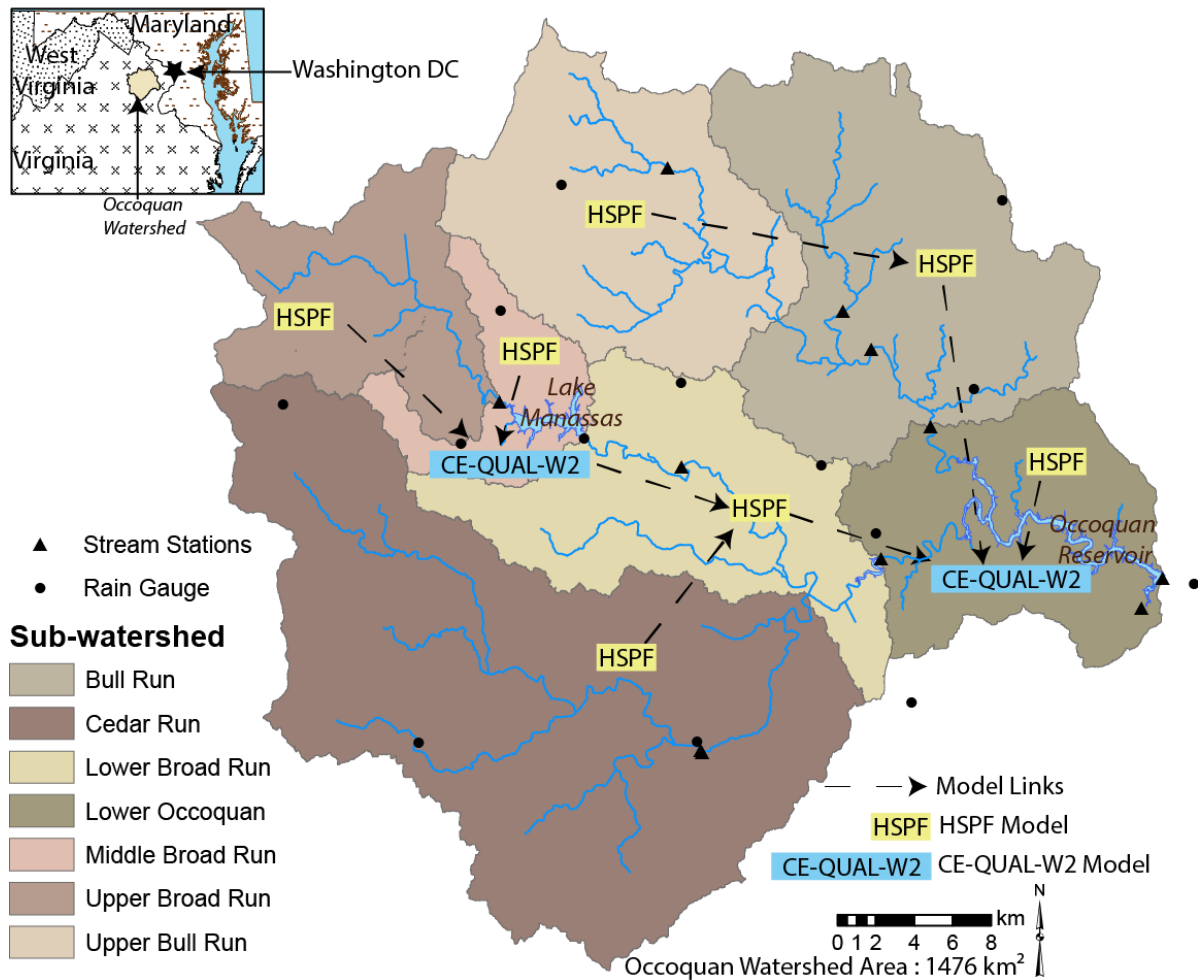


Figure 1. Occoquan watershed and various connection for the water quality model required to simulate the Occoquan System. Fifteen currently active stations on the Occoquan reservoir and the Lake Manassas are not shown.

2.3 Water Resources Data Portal (<https://wqdata.owml.vt.edu>)

The portal is designed using a hybrid web-simulation model, where most of the simulation, data mining, and processing is done on the server and visualization is done on the client's browser using JS and several HTML5 elements. This model was found to be optimum to enable interactivity for the client without slowing down the client machine with extensive computations. Figure 2 show several basic operations performed by the portal, from data acquisition to visualization. There are two main view sets available through the portal: the user views and the admin views.

As the name suggest, the admin views lets the administrator perform several configuration tasks for the portal along with data curation. In the portal, some checks for automatically sensed high frequency data are built in. For example, all water resources data on the web database are checked by hard max-min limits, defined for the station and parameter, and data outside the limit is not displayed unless an administrator adjusts the limits. Further, extreme readings (tentatively defined as >98 percentile and <2 percentile) are flagged for verification. However they are displayed in real-time. Several spatio-temporal outlier detection methods have been identified in the literature (Gupta et al. 2013) and are been investigated to be added to the data curation system for the portal. The portal also informs the administrators (via Twitter and email) if a field station has not reported data for long time, which could be due to equipment failure or communication error. Finally, the user management view (part of the admin views) allows the administrator to control and limit the data access for each individual user or group. For example by default any new user may only get access to one month of data, however, other users who may need longer term data may be allowed longer-term access. The user views provide

methods for the registered users to interact with the portal and analyse the static water resources data available for the region. Note that at present no forecasting and other such modelling is incorporated in the design, a separate portal (described in section 2.4) is used to interact with Occoquan system model. Map view (Figure 3) and Graph View (Figure 4) are the main views that provide interactive data visualization. We invite the readers to explore the site and learn more about user views.

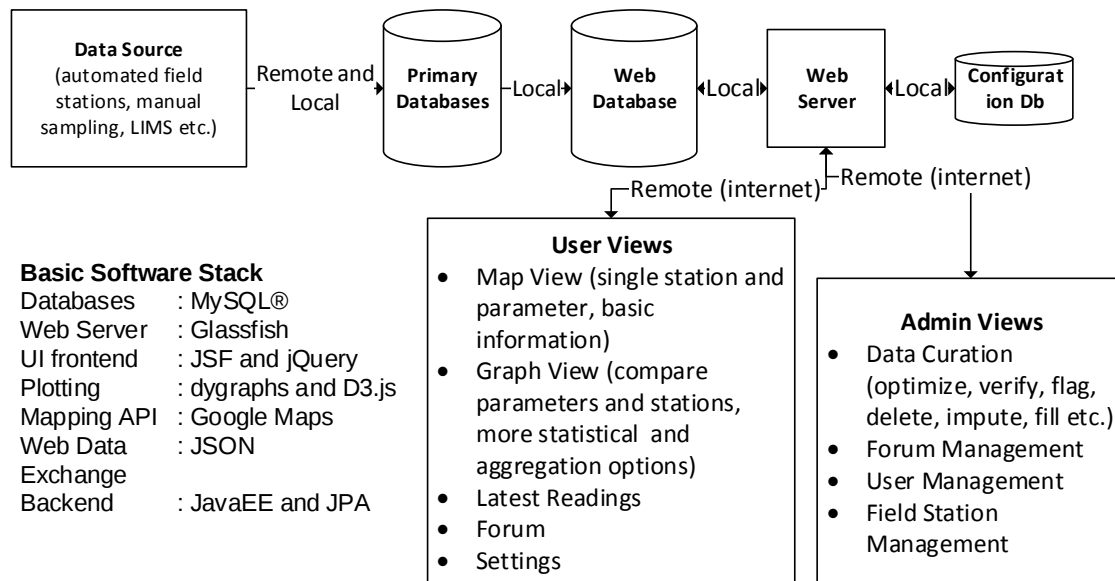


Figure 2. Schematic of data portal design

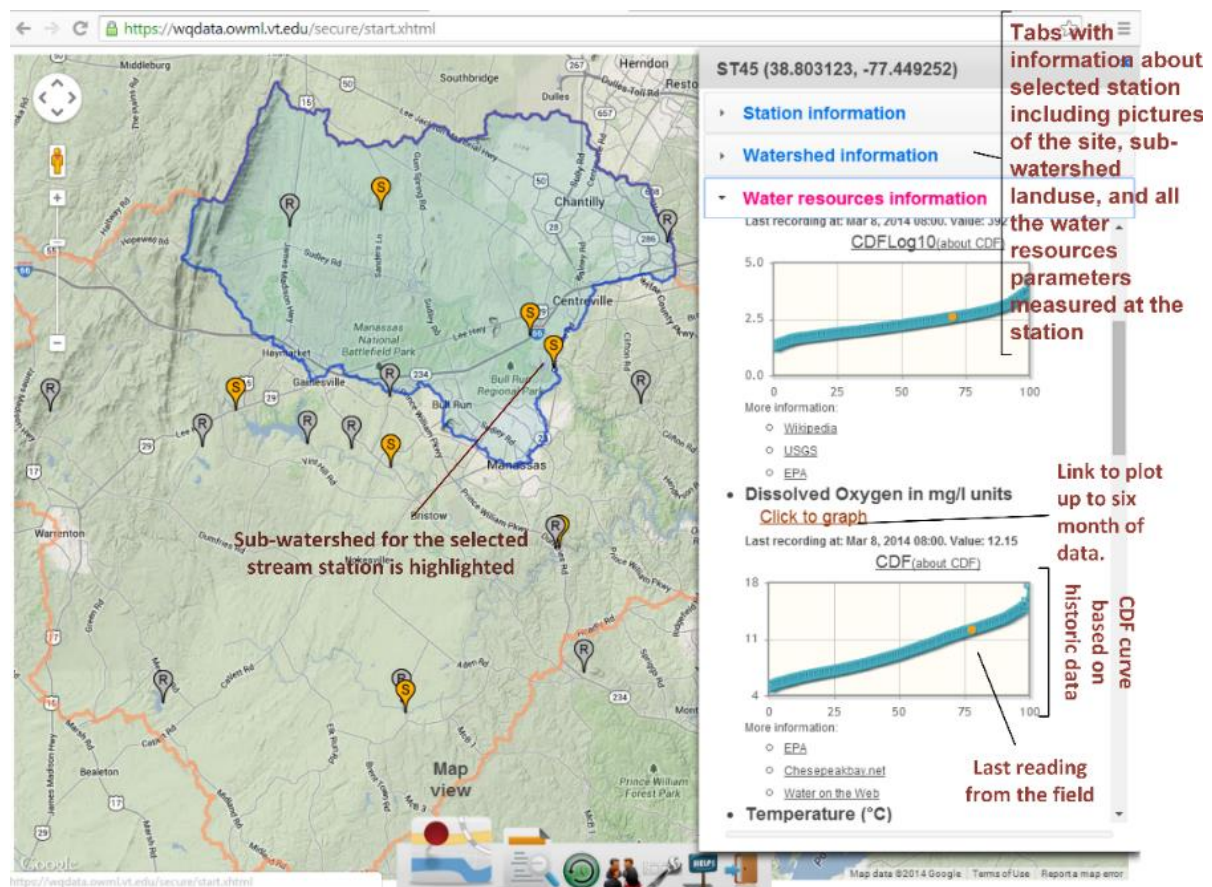


Figure 3. Screenshot of the Map View. Note that stream stations are marked "S" and rain gauges "R".

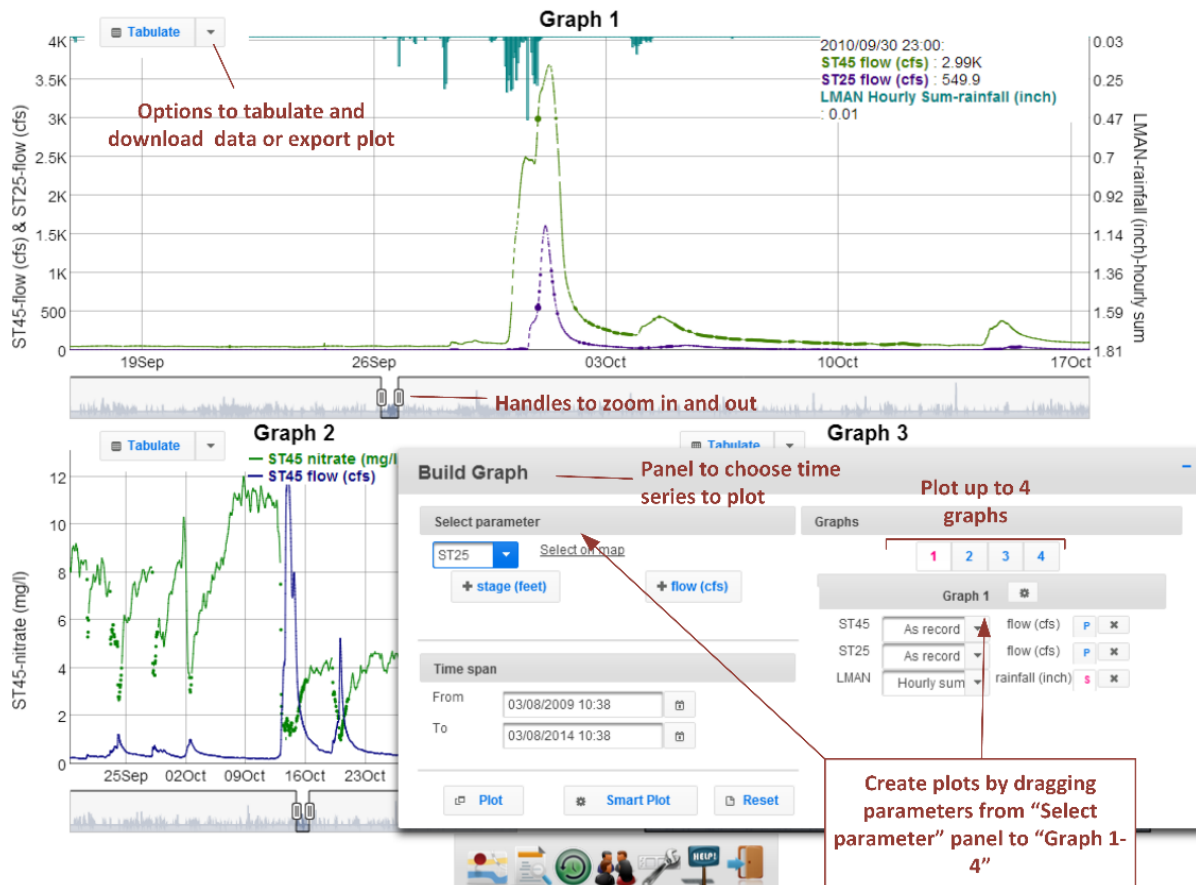


Figure 4. Screenshot of Graph View.

2.4 Land Use Change Simulation Portal (under development)

The Land Use Change Simulation (LUCS) portal is being designed to expose the calibrated water quality model for the Occoquan system to the stakeholders without losing the scientific rigor and benefits of the underlying calibrated models. LUCS has three major components: the Web-Server, the Bridge, and the Distributed Model Runner (DMR). Figure 5 shows a schematic with key purposes of each component listed. Note that the communication between each component is buffered by using the databases.

The web-server is independent of the underlying water quality models and enables users to design land use changes for model simulations by delineating land use modifications from their web-browser over a map of existing land use (Figure 6). In addition, users may also utilize this web-interface to group related land use modifications for easier comparison, share their land use modifications with other users of the LUCS, and analyse results obtained by executions of the water quality models for different modified land uses to aid decision making. The web-server organizes all the individual land use modifications (referred to here as individual *Runs*), available in the database, into *Scenarios* for easier analysis. Results are presented by this system as percent and absolute change from current land use in some water quality and quantity parameters such as DO, TN, TP, and low flow. Note that it is beyond the scope of this paper to explain the functionalities of the modelling system. Please refer to Kumar (2012) for further details of a similar system.

The Bridge module analyzes the *Run* submitted by users, which are essentially land use edits to the base land use condition made through the web-interface (described earlier), for execution. It uses the area and location of land use modifications to determine if any similar *Run* exists in the database. If a similar *Run* is found, then the results for that *Run* are used for analysis. Otherwise, the Occoquan system modeling software is executed, and new results are added to the database. Only the models simulating the portion of the watershed downstream of the area where changes were made are

executed, as no water quality changes are expected in the upstream region. In situations where some (but not all) of the models are similar, all the models downstream to the non-similar sub-basin model are executed.

The DMR-module executes water quality models as required by the bridge module on a server cluster. The server cluster utilized here is a connection of disparate machines (different hardware and software) available on the local network, connected to work together as a single system. The cluster also has the ability to dynamically add or drop machines as they go online or offline, respectively. Kumar (2012) describe the design and operations of the server cluster in detail. The bridge-module acts as a client to the DME model and requests all model execution.

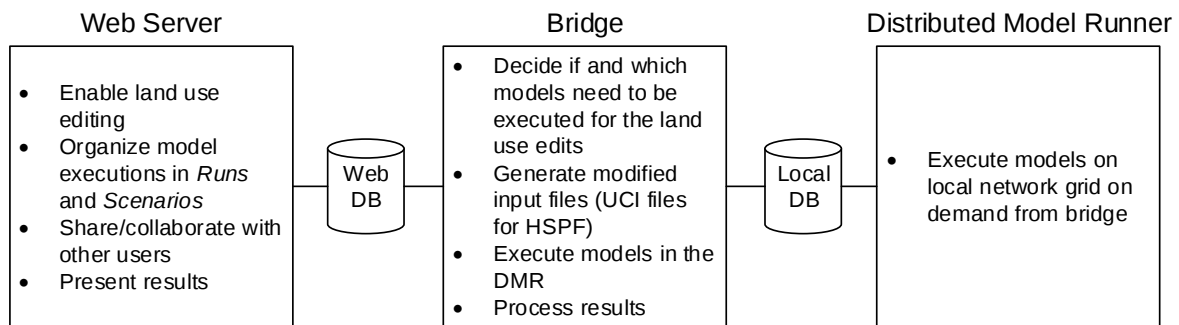


Figure 5. Schematic of the LUCS

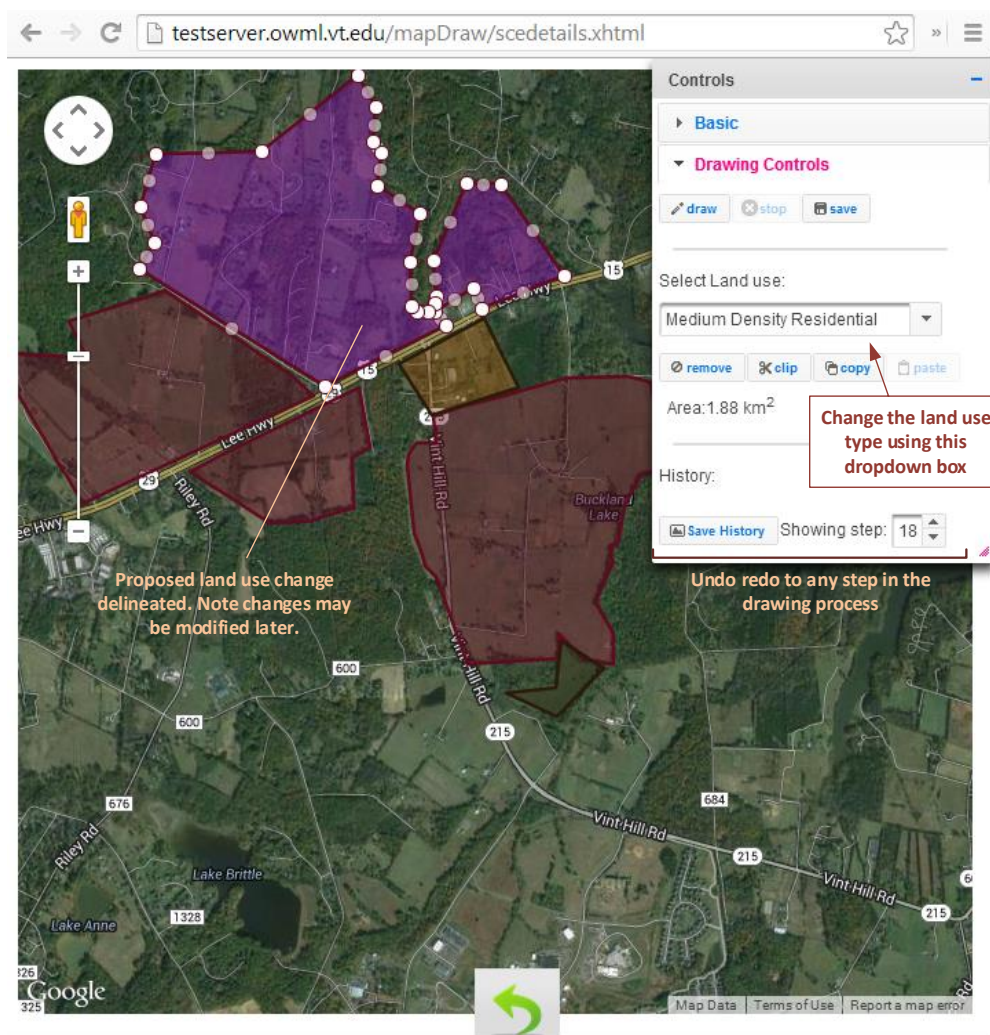


Figure 6. Screenshot of land use edit and delineation page.

3 SOME LESSONS APPLICABLE TO WIDER COMMUNITY

There are several advantages demonstrated in the literature for providing a wider audience access to often expert domain tools and data in an interactive and user-friendly manner, stemming from the participatory nature of decision making and education through such systems (Kumar 2012). However, as we have experienced while developing the systems described above, there are significant hurdles, too. Some of key lessons are listed below.

Consolidate Data Streams:

It is essential at the early stage of design to create a data acquisition and storage policy along with the hardware and software infrastructure required. OWML has been collecting data for the watershed and reservoirs in the Occoquan System for over 40 years, in that time the data collection methodology and storage has changed several times (from paper, to flat files, and then to relational databases). Also, several data sources have different methods of data acquisition. For example, data from stream stations is collected by dialing in to the station data logger at regular intervals, data from rain gauges are posted to an online data center in real time, and data from lab analysis is available through the in-house Laboratory Information Management System (LIMS). A significant task before the actual design of portals was to collect all the data together in a central database and optimize the database for web retrieval. This was achieved by using Java programs that connect to the various data streams and post the data to the central database and only the data required (visualized on the portal) is stored in a secondary web database.

We obtained significant data curation benefits, especially for high-frequency sensor data, by consolidating the data from different streams in a database. Once consolidated, several established spatio-temporal outlier detection mechanisms (Gupta et al. 2013) may be applied to the data.

Anticipate Expertise required:

As is obvious from the design, to develop and maintain these portals some knowledge and understanding in several domains such as software development, computer programming, web server administration, database management, GIS, water resources modelling, and environmental data management is required. Also, being available on the web necessitate that some expertise in maintaining web-security along with hardware and software administration. Use of cloud service providers, especially those that provide Platform as a Service such as Google App Engine, may ease some of these burdens. For a sustainable solution other software management tools such as versioning and collaboration tools (e.g. GIT) may also be desirable along with practices that allow internationalization and localization.

Develop Web-based Visualization Techniques:

JS based toolkits such as D3.js and Dygraphs along with web-mapping APIs have made time series and other interactive visualization easier, however, plotting environmental data, especially censored data, rain data, multiple data sets, and depth profile may not be straightforward. We customized dygraphs to create plots for censored time series data and inverted rain axis. We found that specialized visualization libraries for water resources data (e.g. NADA in R) provide a good starting point for development of scientifically rigorous visualization which may be translated to JS. Further, for model output which may include several water quality parameters at different location and time, local water quality indicators that utilize the dynamic nature of the web may need to be designed.

Use Open Source Software:

There are several open source, community supported, non-proprietary solutions available that were very well suited for these portals and help to keep the cost of development down. We started development with several proprietary solutions, such as Microsoft SQL Server, and ESRI ArcGIS Server, due to familiarity with these systems but were able to migrate out of proprietary software without losing much ease-of-use and developed in-house expertise in the process. However, we still use ArcGIS Engine for some geo-processing on the server as the cost-benefit trade-off favored using proprietary software.

4 DISCUSSION

These experiments have contributed to make the water resources data and modelling tools accessible to the stakeholders in a user-friendly manner for environmental education, and decision making. The portals were developed and are being updated with both experts and lay-stakeholders in mind. For example, data visualization and other analysis have option such as custom normalization for experts to utilize, at the same time also have default settings (or smart settings) to let lay stakeholders use the portal too. We also intend to integrate the two portals described and add functionalities to analyse anticipated climate change impacts, which should further enhance the decision making advantages. Another indirect contribution is in data curation of large datasets, which was obtained by bringing all the available dataset in one database and applying spatial-temporal and surrogate based outlier discovery mechanisms. Finally, we also intend to make the software tools developed in this experiment available to the larger community of water resources modellers, software developers, and other scientists by using a collaborate development framework.

Look-and-feel, speed, and interactivity for the majority of popular web services, such as search, email, and social networks, delivered through a multitude of devices has advanced rapidly and will likely keep advancing. It is thus essential that water resources community should keep pace with the changing nature of the web and the user expectations while interacting with web services/portals. It is also clear that significant resources and technical know-how may be required to develop and update these platforms. Nevertheless, experiments at the OWML and others in the recent literature have clearly shown that it is possible to use the modern web technologies and make the water resources data and modelling tools available to stakeholders. Since several of the data stream generally collected for water resources monitoring, such as high frequency field sensor data and water chemistry data, are similar between different entities collecting the data and several standards such as OGC's WaterML and Observation Data Model already exist, it may be possible to use the collaborative software development model to develop and maintain such data presentation systems. There are several examples of notable collaborative software development, often open source, such as GRASS GIS and Firefox, in various sectors and framework of such projects may be used in this case too. Further, rise and establishment of software collaboration portals such as GitHub have made the task of collaborative coding easier. However, it is obvious that a critical mass of users, academics, and programmers will be needed to make the collaborative process a sustainable success.

Development of a common water resources simulation tool may be harder because several models (including proprietary models) are being actively used and have different Input/Output (IO) methods. Adoption of efforts to standardize model IO may eventually help solve this problem, however, other problems also exist. For example, based on the purpose of modelling (exploration, education, forecasting, understanding the system etc.) the design requirements for a web-based modelling tool may be different.

5 REFERENCES

- Bicknell, B.R., Imhoff, J.C., Kittle Jr, J.L., Jobes, T.H., Donigian Jr, A.S., Johanson, R.C. (2001). Hydrological Simulation Program–Fortran: HSPF, Version 12 User's Manual. Aqua Terra, Mountain View, CA.
- Byrne, J., Heavey, C., Byrne, P.J. (2010). A review of Web-based simulation and supporting tools. *Simulation Modelling Practice and Theory* 18, 253-276.
- Cole, T., Wells, S. (2008). CE-QUAL-W2: A Two-Dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model, Version 3.6 U.S. Army Corps of Engineers, Washington D.C.
- Gupta, M., Gao, J., Aggarwal, C., Han, J. (2013). Outlier Detection for Temporal Data: A Survey. *Knowledge and Data Engineering, IEEE Transactions on PP*, 1-1.
- Kumar, S., (2012). An Environmental Decision Support System to Facilitate Stakeholder Interaction with Water Quality Models, Civil Engineering. Virginia Tech, Manassas, Virginia.
- Kumar, S., Godrej, A.N., Grizzard, T.J. (2013). Watershed size effects on applicability of regression-based methods for fluvial loads estimation. *Water Resources Research* 49, 7698-7710.
- Xu, Z., Godrej, A.N., Grizzard, T.J. (2007). The Hydrological Calibration and Validation of a Complexly-Linked Watershed-Reservoir Model for the Occoquan Watershed, Virginia. *Journal of Hydrology* 345, 167-183.