WaterVis: GeoVisual Analytics for Exploring Hydrological Data

Mahshid Marbouti, Rahul Bhaskar, Zahra Shakeri, Craig Anslow, Leland Jackson[1], Frank Maurer
Department of Computer Science
Department of Biological Sciences[1]
University of Calgary, Canada.
Email: {mmarbout, rbhaskar, zshakeri, canslow, ljackson, fmaurer}@ucalgary.ca

***Abstract*—Visualizing and analyzing large amounts of environmental and hydrological data on maps is difficult. Interaction and manipulation of data is crucial for decision making during natural disasters like floods. In this paper we present WaterVis: a geovisual Big Data analytics application to help analysts explore large amounts of hydrological data to create early flood warnings, and make strategic decisions in critical situations. The implications of WaterVis can help inform the design of future Big Data analytics applications.**

I. Introduction

In 2013, there was a century frequency flood in the Bow River basin (Alberta, Canada) that caused an estimated $5 billion worth of damage. To prevent such damage in the future the Calgary City Council requires better tools for effectively monitoring water operations and flood forecasting.

Collecting and analyzing relevant data are fundamental for describing hydrological phenomena and climate change [1]. To understand these environmental issues a significant amount of hydrological and weather data are collected from different stations and instruments. There are many weather stations that provide hourly data such as precipitation, snowfall, temperature, and wind. The data is useful to model and predict events in critical situations.

In recent years, some visual analytics prototypes have been developed to solve problems in the hydrological domain. The prototypes include visualization of water data for specific geographic areas [2], novel graphical representations of spatio-temporal data [3], and interactive mechanisms for exploring data within visualizations [4]. Our work identifies relationships between water and environmental parameters specifically for the Bow River valley. We combine modern web-based visualization and data mining techniques to explore large datasets. Moreover we emphasize on building an application which enables analysts to perform tasks interactively using maps.

This paper, presents WaterVis which is a geovisual analytics web-based application designed to support monitoring and managing hydrological and environmental resources which helps to investigate parameters that can cause a change in the water levels of rivers, and forecast future values. In the rest of the paper, we will explain our implementation approach (i.e. design and user interaction scenarios) and suggestions on how we can help inform the design and development of future geovisual systems in hydrological domain.

II. Related Work

In recent years considerable research has been focused on water management and specifically flood prediction. To address these issues, numerous studies from Hydrology [5], environment [6], data mining [7]–[9], and data visualization [4], [10], [11] have been conducted on the geovisual and predictive analytics aspects of flood and environmental data.

GeoVisual Analytics. To perform integrated flood management, Qi and Altinakar [10] developed a GIS-based decision support system that utilizes the ArcGIS framework to support two-dimensional numerical simulation. An effective user interface and visualization of results based on user requests are key features of their system.

Vondrak et al. [11] developed a modular system for a variety of users: citizens, governments, and specialists to provide information about emergent situations. Emergent flood prediction is one of the important modules in their system that simulates flood by modelling spatial (catchment and river channels schematization) and time series (Hydrometeorological Data) data. Dykes and Mountain [3] describe novel graphical representations of spatio-temporal data to improve the assessment of geographic relevance.

Demir and Krajewski [4] have developed an integrated flood information system at the Iowa Flood Centre (IFC) that enables users to have access to flood data, perform analysis of available data and interact with data visualizations. The main features of this system that help analysts to make appropriate decisions in a time efficient manner are access to flood maps, real time flood and rainfall conditions, flood forecast, community-based watersheds and visualizing flood scenarios, historical and real-time data of water level and gauge heights.

Merchant et al. [2] implemented a water management system for Bangalore (India) that provides a dashboard view that integrates visualization of water data (flow levels, ground level reservoirs) on the city map.

Predictive Analytics. Imrie et al., [8] used Artificial Neural Networks (ANN) to predict the river flow by modelling non-linear relationships. Etoth et al. [9] applied ANN and Auto-Regressive Moving Average (ARMA) models to compare short-term rainfall prediction models with real-time flood forecasting. Damle and Yalcin [7] proposed an approach for flood prediction by applying Time Series Data Mining (TSDM) techniques and have combined chaos theory and data mining techniques to model non-linear hydrological data sets. Li et al [12] performed an analysis of 400-years historical flood data to identify the extreme flood signals along the flood events. Tehnary et al [13] provided an approach to spatialy predict flood by using rule-based decision tree (DT). Based on the results of this study, applying DT is a reliable method to recognize flood prone areas (87% success rate and 82% prediction rate).

In our work, we combine visualization and data mining techniques to identify relationships between different environmental parameters in addition to factors that contribute to different levels of water in the Bow river. Moreover, our visual web-based application enables users to have interaction with environmental data visualizations and analyze each parameter with regard to the value of other parameters.

mds

III. WaterVis

WaterVis provides interactive visualizations to help analysts gain and discover insight into hydrological data. As a geo-visual analytics application, WaterVis integrates geospatial, hydrological, and weather data. It facilitates exploration of water resources by utilizing different types of visualizations.

A. Design

The following requirements were considered in the design of WaterVis:

Environmental Data Exploration: data includes hydrological and weather data that are associated with discrete stations. Analysts retrieve and compare these data by selecting different stations on the map.

Map Interaction: Due to significant amounts of varied geospatial data, WaterVis was designed to be a GIS application that can represent geospatial data on web based maps. This helps analysts gain a geospatial awareness of environmental data. Moreover, analysts need to interact with geospatial data on the map and WaterVis allows analysts to interact with different datasets by separating the geospatial data into different layers (e.g. stream layers, flood plains, and watersheds).

Web Enabled: The web make WaterVis accessible from different locations and devices. To design WaterVis, the following issues were considered:

> Data Acquisition: How can we collect, prepare and store all the necessary data for monitoring hydrological and environmental events?
> Representation: What visualization techniques are most effective to help analysts understand and discover hidden trends in the data?
> Interaction: What kind of map interaction is effective to explore geospatial data?
> Prediction: What analytical and data mining methods are effective to predict future floods?

B. Data

We acquired environmental and geospatial data for the Bow River basin from different publicly available sources. Environmental data included hydrological data (gauging stations which have parameters like water level, discharge) and weather station data (temperature, rainfall, waterfall, and snowfall). WaterVis uses these data sets, processes them and generates different visualizations.

WaterVis displays geospatial data of the Bow River basin geographical structure including location of different stations on the river, watersheds, streams and floodplain data. Therefore, multiple geospatial datasets were collected from the different sources to provide geospatial visualizations.

C. User Interface

Figure 1 shows the user interface of WaterVis, which contains features to visualize and analyze spatially explicit environmental data.

Geospatial Features: Initially, when an analyst opens WaterVis, a map is centered on the Bow River basin (label D). The map display contains different geospatial information like the DEM (Digital Elevation Model, a digital topographic map), rivers and lakes, watersheds (an area where all water drains into one place), weather stations, and gauging stations. There are multiple widget panels on the top right and top left of WaterVis. The panels at the top of the screen provide features to interact with the geospatial data and update according to an analyst´s requirements. The 'switch basemap' (label C) widget provides a feature to change the base map. Default base map is the street view and analysts can switch between nine available basemaps. The default behavior of WaterVis is that it displays all layers at launch. To change the visibility of layers an analyst can use the switch layers widget (label B) from the top and toggle a layer's visibility on or off. Weather stations are represented by blue icons. Water gauging stations are represented by varying shades of green, which helps analysts see information about data ranges of a station (0-20, 20-40, 40-60 and 60+ years).

Analytical Features: WaterVis uses time series and multidimensional graphs for analytical features (label H). This feature can be viewed by clicking on a weather station (label E) or gauging station (label F). When an analyst selects any station a window is displayed on the screen and shows different visualizations in different tabs (label H). For weather stations there are two tabs showing different analysis results. Similarly, gauging stations use four tabs to show analysis results.

D. Visualizations

Exploring information from a complex dataset is difficult. Structuring and visualizing the data could help to summarize and extract patterns from datasets [14]. The visualizations provide analysts the ability to recognize patterns, trends, and outliers in data. WaterVis supports different types of visualizations including maps, line graphs, dual x axis line graphs, and box-plots.

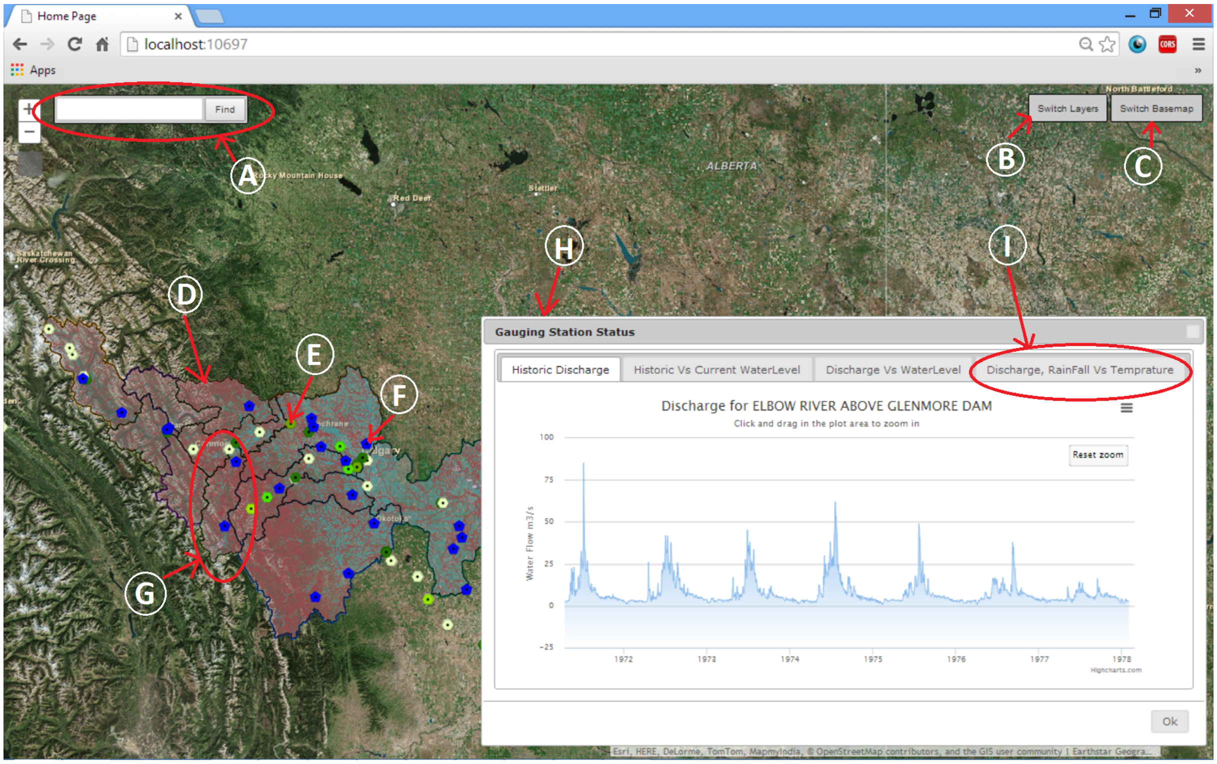
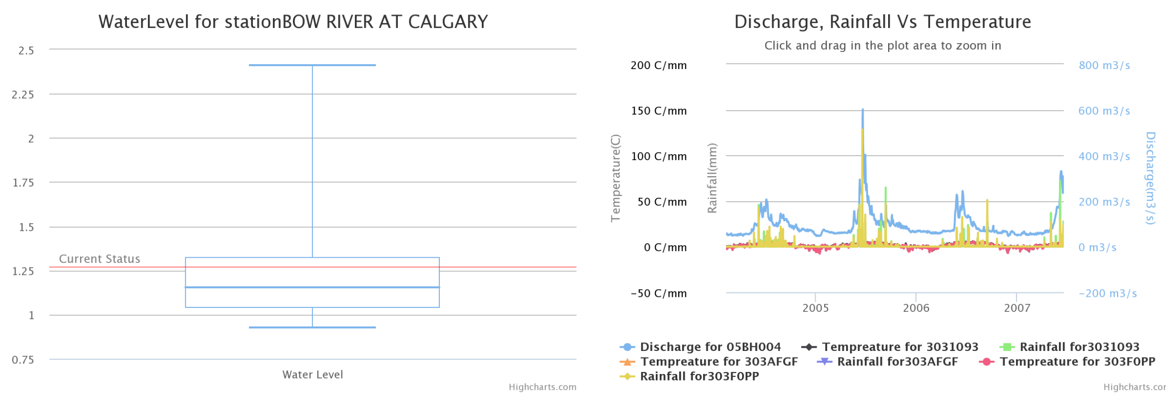


Fig. 1. User Interface of WaterVis. (A. Search Box, B. Switch Layer Widget, C. Switch Basemap Widget D. Bow River Basin, E. Weather Station (blue icons), F. Gauging Station (green icons), G. Watershed, H. Data Visualization screen for station, I.Tabs for multiple visualizations.)



(a) Current water level in comparison with historic quartile values.

(b) Discharge, Rainfall Vs Temperature graph.

Fig. 2. WaterVis - Gauging Station Visualizations.

Geospatial Visualizations: Geospatial visualizations help display data like the DEM, lakes, rivers, weather and gauging station locations which are hard to visualize using temporal or multi-dimensional visualizations. WaterVis displays world base maps with different layers. Layers display the Bow River basin geospatial data and can show small sub-areas (i.e. watershed), which help analysts see geographic regions in which they might be particularly interested. These watershed areas contain multiple markers that represent stations in each watershed area. This geospatial reference provides analysts the ability to navigate from one station to another and check the status of each station quickly.

Visualizations for Gauging Stations: WaterVis contains multiple visualizations for gauging stations. WaterVis uses line graphs to display historic trends of water flow for a particular station (label H in Figure 1). Box plots show the current status of water level in comparison to the historical high and low water levels and can be used to generate alarms when water levels are above a certain threshold (Figure 2(a)). WaterVis has multiple y-axis line graphs called 'discharge, rainfall vs temperature' to show comparative analysis between discharge and different environmental factors like rainfall and temperature (Figure 2(b)).

Visualizations for Weather Station: Weather stations contain current as well as the future weather forecast for each station and they use temporal charts for historic rainfall and snowfall data, which helps analysts understand climate trends at each station location.

E. Hydrological Data Exploration Scenario

Suppose an analyst seeks information about a gauging station within the Bow River basin. The analyst wants to find past trends and determine which months extreme events like sudden increase and decrease in water levels occurred.

The analyst begins by navigating to WaterVis from a web browser. Once WaterVis is loaded the analyst will locate a station. Searching for stations in WaterVis can be accomplished by visual or manual search. A visual search is done by locating a station on the map by looking at the map. A manual search is performed by using the search box feature in which the analyst can type the station name or number. If the station exists, WaterVis will highlight the station on the map by showing a red square over the station on the map.

Once the analyst has found the location of a station on the map and clicked on the gauging station, a pop up is displayed that shows the station number and name. To retrieve the analysis result, the analyst will then click on the detail button to see different tabs. The first tab on the visualization labelled historic discharge (label H in Figure 1) and can be used to see the historic trend of water levels, mostly occur between late June and July in this example.

An analyst may want to check that the current water level is within some defined limits, to determine if there is an alert situation. For this task the analyst switches to the second tab 'current vs historic' (Figure 2(a)) where they can see a comparative analysis between the current water level and the historic maximum and minimum values.

To check which parameter has significant impact on the flow of rivers, an analyst would navigate to the fourth tab (Figure 2(b)) of the gauging station to observe combinations of different parameters from multiple weather stations like snowfall, rainfall, temperature imposed over each gauge station discharge value in the same visualization. This helps find the parameter that has significant correlations with discharge. According to the graph (Figure 2(b)) it is apparent that temperature has a major impact on discharge for the selected station as the change in the temperature (i.e. yellow line) also changes discharge, but it is not the same with the rainfall. Moreover, this graph contains data of all weather stations within the watershed for the selected gauging station.

To find which weather stations have impact on the particular gauging station, the analyst can check the watershed (label G in (Figure 1)) to find weather stations in the same watershed as the gauging station.

F. Architecture and Implementation

The architecture of WaterVis has three layers. First a data layer is used to store different kinds of geospatial and environmental data in a relational database. Second a server layer is used to process data queries and deliver geospatial maps. Third a user interface layer displays the geospatial and environmental visualizations. Data was collected in different formats, then parsed for errors, and manually loaded into the database. The server layer focuses on data preparation and filtering. The database provides input for the ArcGIS server for geospatial processing and an analytics API for analytical processing. The user interface layer visualizes the processed data on the web using different web APIs (i.e. Highchart JS, ArcGIS JavaScript API).

IV. Discussion

Evaluation: We focused the evaluation on our domain expert. We conducted preliminary user studies with the domain expert to evaluate our application. We had bi-weekly meetings with our domain expert and at each meeting we requested him to perform user scenarios and confirm if features aligned with his requirements. This iterative feedback and hints helped us to develop an application based on real requirements. After developing the applications we conducted a demo in which we gathered feedback from a group of students and business analysts. The feedback can be categorized as follows:

UI Enhancements: How can the design of the web interface be more effective and usable for interacting with the hydrological datasets?

Representation Power: How can representation of hydrological elements on a map be more expressive? For example after visualizing different gauging stations on the map, the domain expert asked us to colour code gauges (from dark green to light green) to show the range of data that is available for each gauge.

Hydrological Knowledge: The bi-weekly meetings with our customer helped us to gain important hydrological knowledge we needed to understand the domain for data gathering and analysis.

Performance: One of the technical challenges that we faced was performance. The challenge was regarding the delay in rendering visualizations of long term temporal data. In the UI, each station contains different charts about environmental temporal data and each chart was related to different parameters that were stored in different database tables. Hence, to display a temporal chart, the following steps were necessary: querying database and load data, filtering fetched data and transferring it from the server layer to the user interface layer and processing and then displaying the transferred data on the screen. Running these steps for large amounts of data (i.e.100 years of daily temporal data) was slow. The future plan to overcome this issue is to display the data at multiple levels and let users interactively select and compare different time intervals .

Predictive Analytics: One of the design considerations for developing WaterVis was to provide easy access to geospatial and hydrological data for analysts. Our further consideration was to provide early flood warnings. To predict what might happen in critical situations is important. To move forward in this direction, we focused on predictive analytics. We needed to predict water level and discharge to help estimate what areas are going to be covered during a flood. As a first step, we chose Neural Networks (NN) – a verified Artificial Intelligence (AI) algorithm-which is one of the most suitable methods to predict discharge according to other researchers [8], [9]. We used Matlab for the implementation of the NN method. The next step is to integrate the Matlab code with WaterVis to provide prediction regarding endangered areas in flood time.

Data Acquisition: We had a large variety of data distributed across many resources and not all the data we needed was publicly available so we needed to contact various organizations and explore different resources to gather the data. We are confident that we collected a valuable dataset of hydrological, weather, and geospatial data for the Bow River basin. Acquiring data sets from other flood prone areas is possible and something which we would like to explore in the future.

V. Summary

Exploring large amounts of flood related data on a map is challenging. In July 2013 there was a major flood in Alberta that caused huge damage. The late warnings during the flood demonstrated the need for an effective application to help hydrologists predict such disasters. To help hydrological analysts, we developed WaterVis, a domain specific geovisual analytics web-based application that helps analysts to explore and monitor environmental data with respect to climate change. In the future we plan to extend WaterVis by integrating our predictive analytics code and conduct usability studies with hydrological analysts to evaluate the effectiveness of the application.

References

[1] J. Guzman, D. Moriasi, M. Chu, P. Starks, J. Steiner, and P. Gowda, "A tool for mapping and spatio-temporal analysis of hydrological data," *Environmental Modelling & Software*, vol. 48, pp. 163–170, 2013.

[2] A. Merchant, M. Kumar, P. Ravindra, P. Vyas, and U. Manohar, "Analytics driven water management system for bangalore city," *Procedia Engineering*, vol. 70, pp. 1137–1146, 2014.

[3] J. Dykes and D. Mountain, "Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications," *Computational Statistics & Data Analysis*, vol. 43, no. 4, pp. 581–603, 2003, data Visualization. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167947302002943

[4] I. Demir and W. F. Krajewski, "Towards an integrated flood information system: Centralized data access, analysis, and visualization," *Environmental Modelling and Software*, vol. 50, pp. 77–84, 2013.

[5] D. Sui and R. Maggio, "Integrating GIS with hydrological modeling: practices, problems, and prospects," *Computers, environment and urban systems*, vol. 23, no. 1, pp. 33–51, 1999.

[6] M. Acreman, "Principles of water management for people and the environment." 1998.

[7] C. Damle and A. Yalcin, "Flood prediction using time series data mining," *Journal of Hydrology*, vol. 333, no. 2, pp. 305–316, 2007.

[8] C. Imrie, S. Durucan, and A. Korre, "River flow prediction using artificial neural networks: generalisation beyond the calibration range," *Journal of Hydrology*, vol. 233, no. 1, pp. 138–153, 2000.

[9] E. Toth, A. Brath, and A. Montanari, "Comparison of short-term rainfall prediction models for real-time flood forecasting," *Journal of Hydrology*, vol. 239, no. 1, pp. 132–147, 2000.

[10] H. Qi and M. Altinakar, "A GIS-based decision support system for integrated flood management under uncertainty with two dimensional numerical simulations," *Environmental Modelling and Software*, vol. 26, no. 6, pp. 817–821, 2011.

[11] I. Vondrak, J. Martinovic, J. Kozusznik, S. Stolfa, T. Kozubek, P. Kubicek, V. Vondrak, and J. Unucka, "A description of a highly modular system for the emergent flood prediction," in *Computer Information Systems and Industrial Management Applications (CISIM)*. IEEE, 2008, pp. 219–224.

[12] Y. Li, Y. Guo, and G. Yu, "An analysis of extreme flood events during the past 400 years at Taihu Lake, China," *Journal of Hydrology*, vol. 500, pp. 217–225, 2013.

[13] M. S. Tehrany, B. Pradhan, and M. N. Jebur, "Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS," *Journal of Hydrology*, vol. 504, pp. 69–79, 2013.

[14] C. Ware, *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., 2004.