A Web-based Data Visualization and Analysis System for Watershed Management

Yufeng Kou, Changtien Lu
Dept. of Computer Science
Virginia Polytechnic Institute
and State University
7054 Haycock Road
Falls Church, VA 22043
{ykou,ctlu}@vt.edu

Adil Godrej, Thomas Grizzard, Harold Post
Occoquan Watershed Management Lab
9408 Prince William Street
Manassas, VA 20110
{agodrej,grizzard,hpost}@vt.edu

June 15, 2005

Abstract

In recent years, the combination of efficient databases, data analysis, and Internet technologies has greatly enriched the functionalities of GIS systems and facilitated the dissemination of geospatial information. In this paper, we propose a web-based data visualization and analysis system for watershed management of the Occoquan Basin in Northern Virginia. This system distinguishes itself from other systems in the following respects. First, it enables near real-time data collection and dissemination. Second, data query tools are embedded in the system to support effective data analysis and visualization functionalities. In addition to regular queries, aggregation queries have been developed to support summarization views of historical data with various granularities. Third, the visualization and analysis tools can be accessed via the Internet. WWW technologies, such as active web pages and Java programming, provide user-friendly interface for browsing, plotting, comparing, and downloading information of interest, without the need for dedicated GIS software. This system can render information support for many applications, including pollution control and bio-chemical detection.

1 Introduction

Traditional GIS systems appear as stand alone tools deployed on dedicated computers with sophisticated graphical display equipment. In recent years, however, GIS has been changing its role to become an integrated system which combines geo-spatial applications with databases, data mining, and Internet technology. Databases, especially spatial databases, play an important role in storing huge amounts of spatial and non-spatial data. Dedicated GIS databases can provide powerful data processing and map management functionalities, such as efficient data accessing, flexible queries, and spatial data visualization. However, they come with expensive price tags and require high performance computers. Their maintenance also demands highly professional engineers. In addition, traditional GIS systems are not designed for non-professional users, which limits their wide adoption. Since 1990, Internet technology has produced a revolutionary change in data access and dissemination. It has simplified access to geographic data and now provides a vehicle for disseminating information to a multitude of users across

various hardware and operating systems. The development of web technologies, such as CGI(Common Gateway Interface), scripting languages, Java Applet, JSP(Java Server Page), and Java Servlet, has provided an efficient and effective means of delivering spatial data products using the Internet.

Data visualization is an important area of data analysis, where the collected data is summarized and presented in visual form to help grasp the minute details and relationships of data sets for more effective decision-making. Advances in visualization technology have revolutionized fields such as architecture, medicine, etc. Data visualization is essential for complex data analysis. In complex data, the structure, features, patterns, trends, anomalies and relationships are not easily detectable. Visualization supports the extraction of hidden patterns by presenting the data in various visual forms. Visualization not only provides a qualitative overview of large and complex data sets, it can also assist in identifying regions of interest and parameters for further quantitative analysis. Along with the wide spread of the World Wide Web, visualization on the Internet has become the trend. Users do not need purchase any specific hardware or install any dedicated graphic software. As long as they have access to internet, they can view a particular data set in either graphical or tabular formats. In this paper, we present a web-based GIS system which can collect data from distributed sensors and provide a highly interactive and intuitive graphical user interface for data monitoring and data analysis.

The Occoquan Watershed Monitoring Laboratory (OWML), established by the Virginia Polytechnic Institute Department of Civil Engineering, began its on-site operations in 1972 and has conducted comprehensive studies of receiving water quality and the effects of the AWT(Advanced Waste-water Treatment) discharges to the present time. In the course of its operation, OWML has developed a comprehensive database of water quality in the Occoquan Basin and has been instrumental in making determinations in a number of areas which have proven to be critical to the ongoing management of water quality, for example, providing receiving water data for use in decision on AWT plant expansions and rendering information on water quality effects and cost-effective control of nonpoint sources of pollution. In this paper, we present a web-based visualization system that can be used to collect water data in near real-time and disseminate it to the public in an efficient way.

Our system is a hydrologic and water quality data acquisition and analysis system designed by the Occoquan Watershed Monitoring Laboratory. The Occoquan Watershed is located in northern Virginia and is situated on the southwestern periphery of the Virginia suburbs of the City of Washington, D.C. The basin encompasses six political subdivisions of the Commonwealth of Virginia, including portions of four counties, and the entire land area of two independent cities. The watershed lies to the south and west of the U.S. National Capital, Washington, D.C. It is bounded by the Potomac Estuary to the east and Bull Run Mountain to the west. The northern and southern boundaries lie in the Counties of Fairfax and Fauquier-Prince William, respectively. At the location of the Occoquan High Dam, the watershed drains 570 square miles.

Contributions: In this paper, we propose a web-based online GIS data visualization tool to monitor and analyze the surface water resource of the Occoquan Watershed Basin in Northern Virginia. This system can render information support for many applications, including pollution control and bio-chemical detection. This system distinguishes itself from other systems in the following aspects. First, it enables near real-time data collection and dissemination. The readings of sensors can be collected and published on the Internet within minutes. Second, data query tools are embedded in the system to support effective data analysis functionalities. In addition to regular queries, aggregation queries have been developed to support summarization views of historical data with various granularities. Third, the new information and analysis tools can be accessed via the Internet. WWW technologies, such as active web page and Java programming, provide user-friendly interfaces with which to browse, plot, compare, and download

information of interest without the need for dedicated GIS software. Fourth, we provide easy-to-use visualization components. The water data can be visualized with tables, JPEG figures and PDF files. Both linear scale and log scale are supported for displaying the data curve. By exploring the data cube concept [5], an aggregation view of the water data can be provided with different time granularities.

Paper Organization: The remainder of this paper is organized as follows. In Section 2, related work is surveyed. Section 3 describes the hardware architecture and software components of our system. The system functionalities and visualization components are demonstrated in Section 4. Section 5 discusses key technical issues in distributed GIS system visualization and identifies several directions for future research. Finally, we summarize our work in Section 6.

2 Related Work

Web-based GIS has been rapidly evolving with the development of WWW and Internet technologies. Generally, there are two major kinds of Web-based GIS, server-side and client-side [10]. Server-side GIS performs all processing on the server, with only a "thin" client (Web Browser) installed on the user's machine. This method maximally simplifies user's responsibility for installation and maintenance. However, when the number of users increases, the server load increases and system performance degrades dramatically. Client-side GIS is based on client computing technologies, such as Active X, plug-ins, and Java Applet. It can cache necessary data on the user machine and support a more interactive Graphic User Interface(GUI). The price is that more software components need to be installed on the client machine. The common dilemma of Web-GIS systems is the balance between large data sets and limited network bandwidth. Much research has been conducted to promote the efficiency of Web-GIS [6, 10]. In [12], a GIS web service is implemented to reduce the user-perceived response time and support user self-guided navigation. The core idea is to decompose GIS data sets into small blocks on the server side, and use locality-based caching and pre-fetching on the client side. Wei et al. proposed a new method of efficient spatial data transmission for client-side Web-based GIS [15]. Their system first divides a large-size map into several parts, called "Tiles", according to appropriate granularity. A tile will be sent by server only if it overlaps with the requested region. Since data is only transmitted on demand, the bandwidth utilization is optimized. 3D techniques are increasingly used to represent GIS data in a more intuitive way, especially for urban planning. In [4], a data model was proposed to support 3D topology and 3D geometry. Coors and Flick developed a web-based 3D-GIS based on Java and VRML [2]. To promote transmission and display efficiency, their system also supports multiple levels of detail for visualization.

There are various Web-based GIS systems that have been used for regional natural resource analysis [14], mobile device positioning [11], or moving objects tracking [3]. However, these GIS are heterogeneous, with different data representations and query languages. Thus, it has become an interesting problem to develop methods that will provide an integrated view of the data supplied by all sources and to devise a geographical query language for accessing and manipulating this integrated data. Tu, Ratcliff and their coworkers proposed a J2EE-based approach which can support the exchange of both spatial vector data and raster image data [13]. Boucelma et al. introduced a multi-tier client-server architecture which uses standard wrappers to access data and uses "derived wrappers" to capture additional query capabilities [1].

Han et al. proposed a spatial data mining protocol, DBMine, for knowledge discovery in relational databases and data warehouses [7]. They extended DBMiner to GIS applications and developed a spatial

data mining tool, GeoMiner [8,9]. GeoMiner is capable of identifying various rules in spatial data, such as characteristic rules, comparison rules, and association rules.

There are several online water data query systems, for example, "Water On the Web" by the U.S. Geological Survey(USGS). However, they suffer from shortcomings in both visualization and query performance. First, they cannot support aggregation views of the data. Second, they can not provide near real-time data dissemination. Even the "real-time water data" from the USGS can only display data that is two hour old. In this paper, we propose a web-based data visualization system which can support both near real-time data collection and flexible data visualization using the data cube concept, log scale plotting and JPEG/PDF/text format displays.

3 Concept and System Architecture

OWML manages a dozen monitoring stations, distributed over a large area of the Occoquan Watershed. Figure 1(a) shows the locations of these stations on a map of northern Virginia. The stations are set along rivers or runs, collecting various readings with automatic gauging devices. Figure 1(b) shows the picture of station ST01, which is on the upstream and southern side of the Occoquan dam. Based on these stations, our system performs the systematic collection, recording, evaluation, and dissemination of information about water quality, flow characteristics, and chemical dissolution. As shown in Figure 2, the collected data can be used for environmental quality evaluation, fish and wild life services, natural resources conservation, and weather and climate research. This system has made it possible for the local governments of northern Virginia to successfully deal with the competing uses of urban development, attendant waste water discharges and urban runoff, and public water supply in a critical watershed-impoundment system.

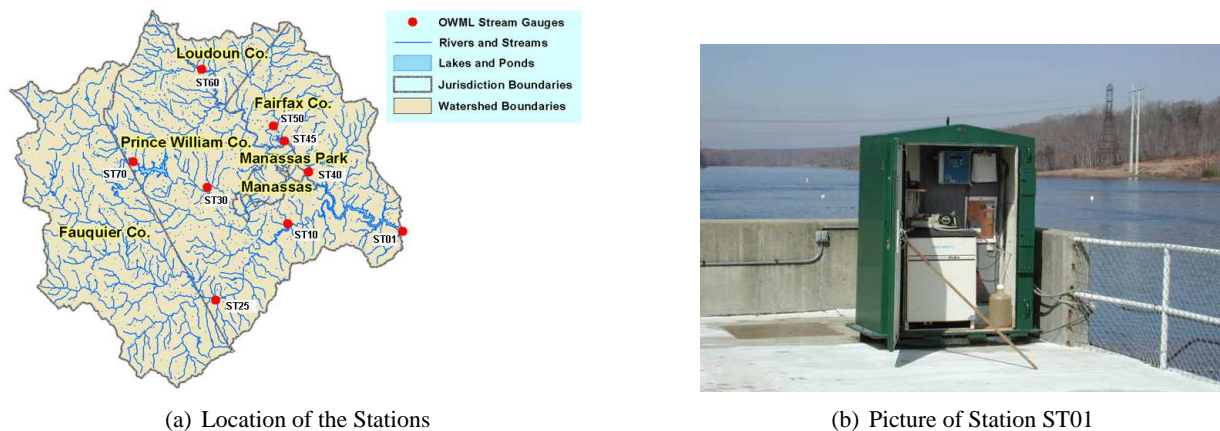(a) Location of the Stations

(b) Picture of Station ST01

Figure 1: The Monitoring Stations

This system implements two main functionalities: data collection and data visualization. Data collection controls the communication between monitoring stations and the database. It collects the readings from distributed sensors, converts them to an appropriate format, and inserts records into the database. Visualization is a web-based GUI, used to accept a user's request and display the query result in textual or graphical format. Data visualization can also help users perform basic data analysis, such as presenting aggregation views and statistical summaries.
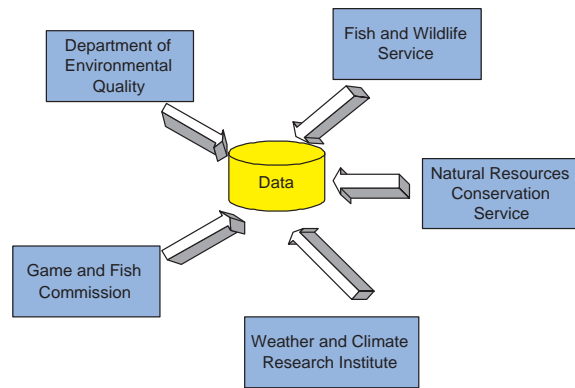
4

Figure 2: The Applications of Water Monitoring Data

The hardware architecture of our system is illustrated in Figure 4. The monitoring stations are connected to the OWML data center using phone lines, and the monitoring data are transmitted every 15 minutes. There are two databases in our system. The master database is located at OWML where it receives data from monitoring stations. The engineers can modify the master database if the data contain errors or noise. The slave database resides on the web server at Northern Virginia Center of Virginia Tech, which contains a subset of the master database. The slave database is synchronized with the master database and is only used for data visualization on the Internet. Duplicate databases are maintained for reason of security. The data in the master database can only be operated by specific engineers at OWML, as some of the data are confidential and need to be filtered before available to the public. The master database and slave database are connected with a Virtual Local Area Network(VLAN), which provides secure communication. Data requests are initiated by web users, checked by the firewall, and then replied by the web server.

Figure 5 shows the software components and their relationship. Our system is a standard three-tier architecture, consisting of user interfaces, a web server and databases. The three-tier architecture has two major advantages. The first is "thin client," which means that only a web browser needs to be installed on the user side and all the computation and maintenance is performed on the web server and database side. This will simplify client operations and reduce the maintenance costs. High performance is another advantage. In our system, most of the user requests are data centric, requiring large amounts of data but comparatively little CPU time. Thus, data processing should reside on the database side to avoid the need to transmit large data packages through the network. The software components in our system can be grouped into three parts according to their locations: User Interface, Information Publication Center, and Data Collection Center.

- **User Interface:** All the query requests and results are displayed through a web browser. Two data display formats are supported, textual format and graphical format. In textual format, data is listed in a table, which can be downloaded as a text file. Graphical format displays changes in the data with figures in the form of curves or color maps. Users can choose to save the figure as either a JPEG or PDF file. Figure 3(a) shows the locations of the available monitoring stations and a table of their current readings. Users can click the map or the station list to browse the data from a particular station. Figure 3(b) shows the query interface. Users can select a particular station, a time period, a set of attributes of interest, and the display format which they prefer to browse the
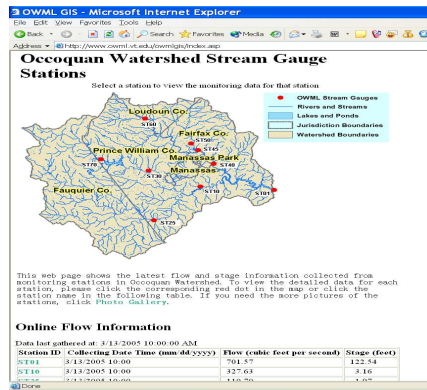
(a) Station Index
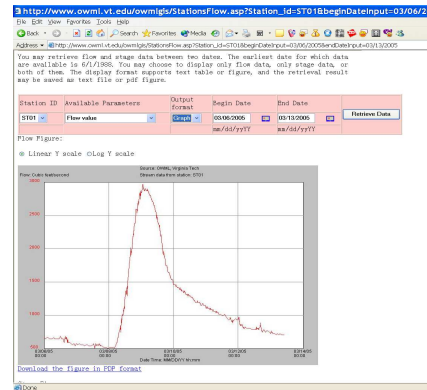
(b) Information for a Particular Station

Figure 3: User Interface

water data.

- **Information Publication Center:** The information publication center consists of a web server and the slave database. The web server is the container for static web pages, active web pages, and Java Servlets. The HTML files are used to display static information, such as the website introduction and project description. The active pages and Java Servlets are responsible for receiving user requests, converting them to SQL, communicating with the slave database, and translating the query results to the various display formats. The query result can be saved as downloadable files in text, JPEG or PDF format. JDBC is used as an open interface between Java Servlets and the database management system. The main task of the slave database is to synchronize with the master database. An FTP client programmed with Java is in charge of the data synchronization. Since the size of master database is large, we employ an "incremental update" scheme to accomplish synchronization. Once mater database receives a batch of new data, it will generate a flat file containing these new data. An FTP client Java program on the web server will check frequently to see if the flat file on the master database server is new. If the file is new, it builds an FTP connection, fetches the file, and calls a "Data Appender" program to write the new data to the slave database. Since only increments of data are transmitted, this does not consume much bandwidth.

- **Data Collection Center:** The data collection center is located at OWML and controls the data collection from the monitoring stations. The master database obtains readings from monitoring stations periodically, integrates and transforms them to the desired format, and builds an index for them. Finally the new created data will be output as a flat file for synchronization with the slave database. There is always a possibility that some malfunction of the sensors or communication equipment may cause errors in the data. Engineers at OWML can manually correct these data and generate a data file. This file contains all the data having been modified, to be inserted, and to be deleted. As the new data is synchronized, these corrected data update the slave database automatically.
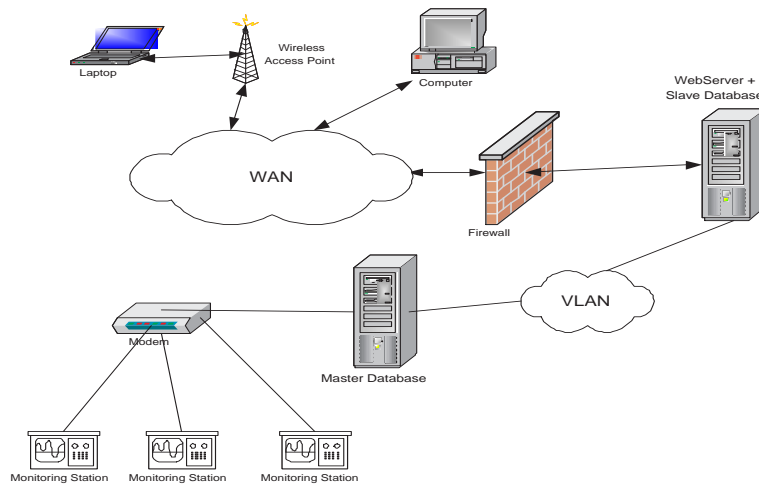
6

Figure 4: The Hardware Architecture of OWML-GIS

4 System Demonstration

Our system can dynamically display both real-time and historical surface water monitoring data, such as flow and stage, upon request. These data are collected at a frequency of every 15 minutes. For rapidly changing situations, the data can be collected as often as every 5 minutes. We have accumulated more than 30-year data for the Occoquan Watershed, from which some interesting patterns can be identified by using various visualization methods. Our system can monitor the flow and stage information in both tabular and graphical formats. By comparing the data collected from different stations, some associations can be identified. In addition, statistics extracted from the historical data, such as minimum, average or maximum flow, can be provided to help identify time periods when abnormal weather patterns occur. Moreover, aggregation queries can be supported in order to present data summarization in various granularities.

The functionalities of our system are demonstrated as follows, using several examples.

Flow and Stage Illustration: Figure 6 shows the flow and stage fluctuations detected by station ST01 between Sept 15, 2004 and Sept 22, 2005. The X-axis denotes the time period and the Y-Axis denotes the flow/stage value. There is a peak around September 20th, lasting from 1AM September 19 to 18PM September 21. During this period, Hurricane Isabel just passed close by the Occoquan watershed. The rainfall that accompanied it contributed significantly to the high flow value.

Data Comparison Between Different Stations: For stations deployed along the same river, the flow information collected by the upstream stations may have a close relationship with readings of the downstream stations. For example, Figure 7 shows a flow comparison between station ST60 and station ST45. ST60 is located upstream and ST45 is located downstream of Bull Run. The X-axis denotes the time period and the Y-Axis denotes the flow value. We can clearly observe that the peak flow detected by ST60 occurs at around 7AM Jan 14, 2005, whereas the peak flow detected by ST45 occurs at around 5PM Jan 14, 2005, a 9 hour gap. This relationship is useful for flood predictions. By collecting the comparison data from different stations along the river, we can develop a model to predict future flow of the downstream and whether there will be a flood.

Log Scale vs Linear Scale: For water data over a long period, it is difficult to observe many details,
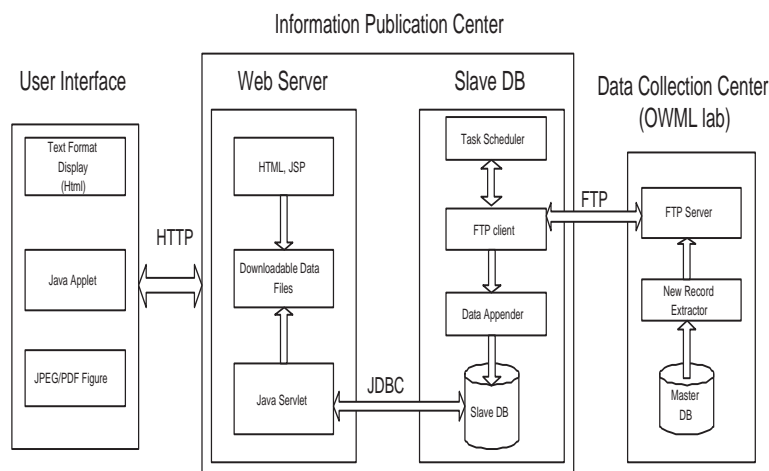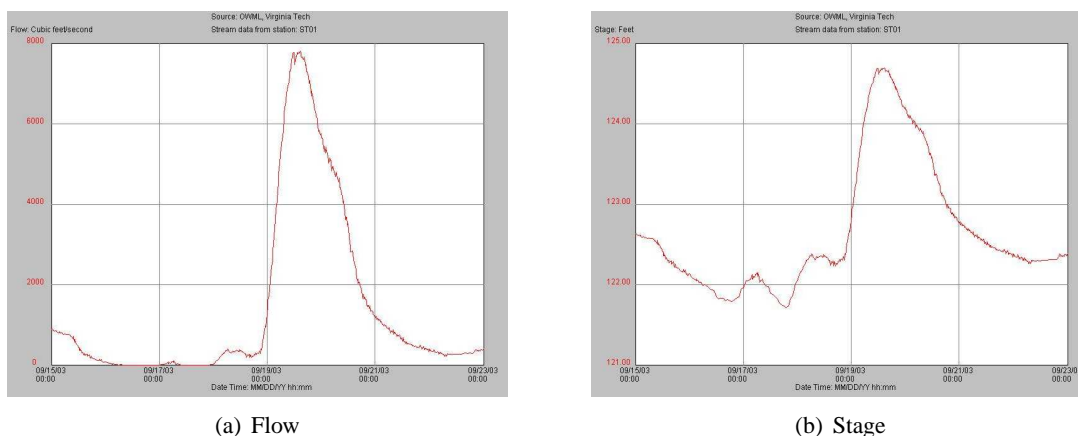
Figure 5: The Software Architecture of OWML-GIS

(a) Flow

(b) Stage

Figure 6: The Flow and Stage Information for ST01

since the difference between the maximum value and minimum value is so large that the data curve is overly compressed. To address this problem, an option of log scale is provided. Figure 8 shows the comparison between linear scale and log scale for flow data collected by station ST25 between Jan 1, 2004 and Feb 17, 2005. The X-axis denotes the time period and the Y-Axis denotes the original flow value(for linear scale) or the logarithm of the flow value(for log scale). With the linear scale, only the largest data values can be clearly read, while the intermediate data are displayed in a small region and cannot be clearly differentiated. With a log scale, the data values are more evenly distributed in the figure, thus revealing more details. For example, for the data values between 1AM June 29 2004 and 1AM August 28 2004, Figure 8(b) can show considerably more detail than Figure 8(a).

Illustration of Data Statistics: In some cases, statistics from the historical data can help identify interesting patterns, for example, which years were unusually dry and which years were particularly rainy. Figure 9 shows the 30-year statistics of the flow information collected by station ST70, including the minimum flow, the average flow, and the maximum flow. We can clearly observe that for northern Prince William County, where ST70 is located, 1980 was a dry year, since its average flow was very low,

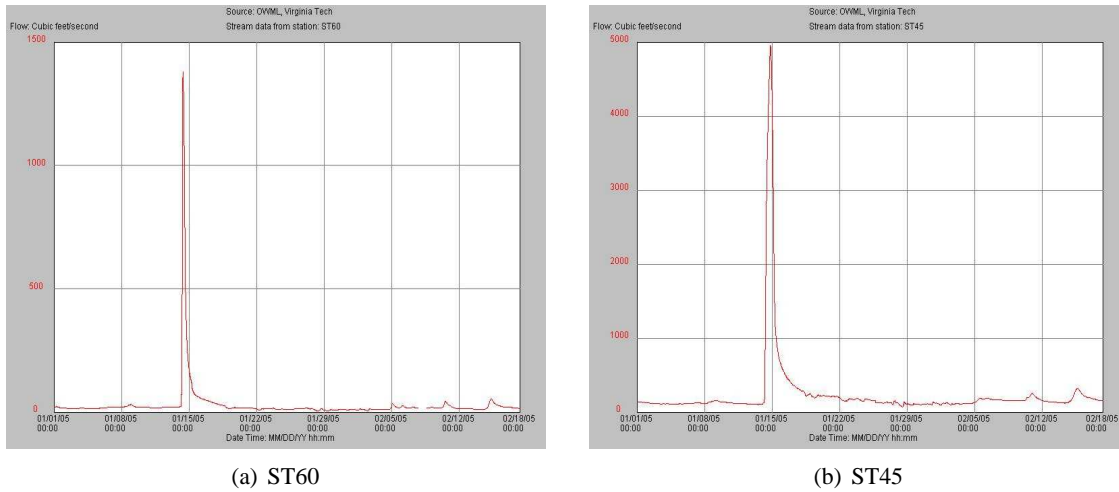|	
(a) ST60	(b) ST45

Figure 7: The Flow Comparison Between ST60 And ST45

as well as minimum flow and maximum flow. 1973 was a rainy year considering its high flow values of minimum, average, and maximum flow statistics.

Data Cube: By integrating the data cube concept with an interactive graphic user interface, the spatial patterns and temporal trends can be identified easily. The concept of the data cube is the engine behind our system. A data cube is used to generate the union of a set of alpha-numeric summary tables corresponding to a given hierarchy. For example, the flow data of a specific station can be aggregated to "Time of Day", T_{TD}, "Day of Month", T_{DM}, and "Month of Year", T_{MY}. Also, the different monitoring stations can be viewed as another dimension, S. By combining S with different time dimensions, we can see the data change for multiple stations simultaneously. Since different stations represent different geographical areas, the temporal aggregated data for various stations may reveal interesting spatial-temporal associations. Via roll-up or drill-down operations, users can observe data visualization on different levels of the aggregation hierarchy. For example, aggregating the historical data to different granularities and displaying them in a single figure can help reveal interesting correlations among data from different time periods. Several data cube examples are described as follows.

- ST_{DM}**(Stations vs. Day of Month):** Figure 10 shows the relationship between different stations and different days in a particular month. The X-axis denotes the days of a month and the Y-axis denotes stations, from ST01 to ST70. The sub-figures (a), (b), and (c) illustrate the data cubes for April, May, and June in 2004, respectively. As shown in Figure 10(a), high flow values were detected by ST01, ST25 and ST45 on April 13th and 14th. For May 2004, Figure 10(b) shows that high flow values cluster in 2 periods, May 3th and 4th, May 8th and 9th. These high values were detected by stations ST01 and ST10. Figure 10(c) shows that on June 17th and 18th, flow was very high for station ST01 and ST10, whereas other stations did not report high flow. On June 5th, a fairly high flow was detected by all the stations.

- $T_{TD}T_{DM}$**(Time of Day vs. Day of Month):** Figure 11 shows the flow information for station ST70 in two dimensions: Time of Day and Day of Month. The X-axis denotes the time of a day, and the Y-axis denotes the days of the month. The data cubes for April, May and June in 2004 are illustrated in sub-figures (a), (b) and (c), respectively. Figure 11(a) shows that in April 2004,
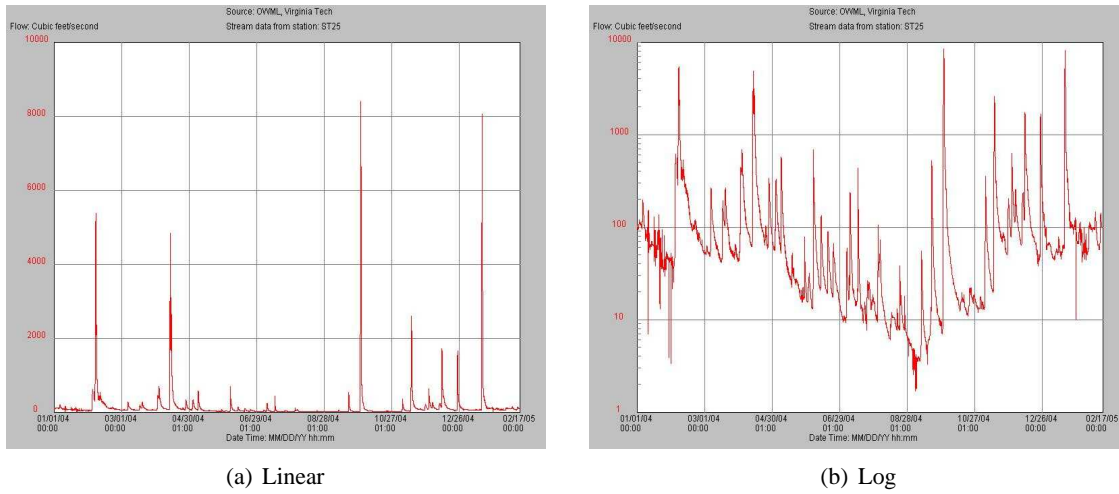
9

(a) Linear

(b) Log

Figure 8: The Linear and Log Scale Display

(a) Minimum Flow
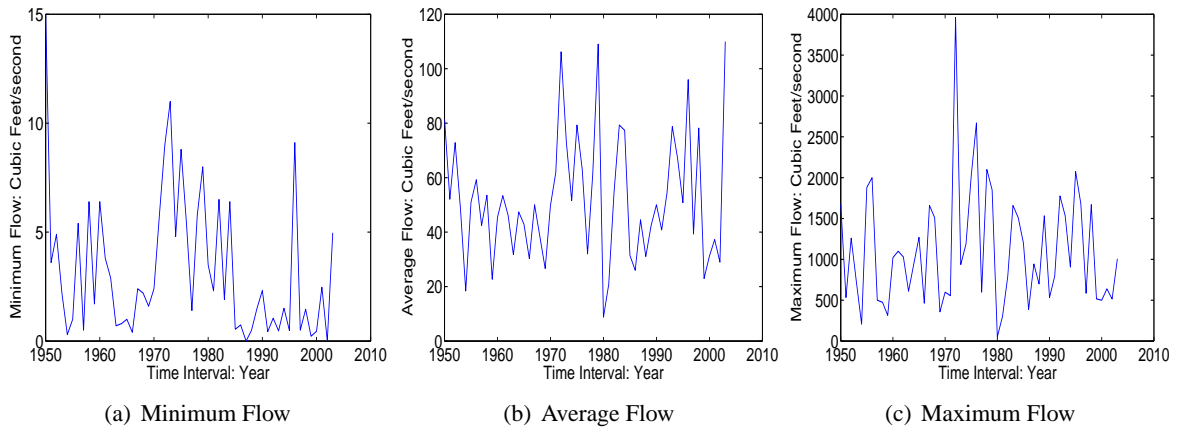
(b) Average Flow

(c) Maximum Flow

Figure 9: The Minimum, Average, and Maximum Flow for Station ST70

ST70 detected comparatively high flow on three days, the 12th, 13th and 14th. On April 12th, the high flow occurred between 20PM and 24PM. On April 13th, the high flow occurred over two time periods, 0-3AM and 19-24PM. On April 14th, the high flow values were evenly distributed between 0AM and 12AM. In Figure 11(b), it can be seen that high flow values cluster at two time periods, 20-24PM on May 7th and 0-5AM on May 8th. Figure 11(c) shows that a high flow was identified on June 6th, 2004, which started from 6AM, reached a peak at 12AM, and returned to a low level by 24PM. In addition, this figure reveals that in June 2004, there were four days with high flow values, the 5th, 6th, 12th, and 18th.

- $T_Y T_{DY}$ **(Years vs. Day of Year):** Figure 12 shows a range of years(1950-2003) and the daily flow in each year for ST70. The X-axis shows the 365 days of a year and the Y-axis lists the years. In this figure, we can easily identify that on which day in which year there was a particularly high flow. For example, near the 170th day in year 1973, there was an extremely high flow which is represented by dark color.
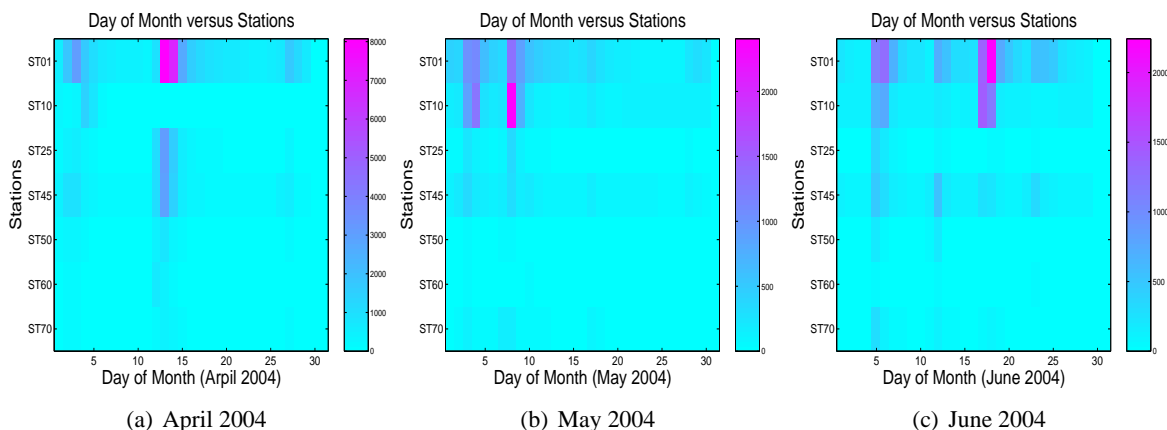
10

(a) April 2004   (b) May 2004   (c) June 2004

Figure 10: The Data Cube: Stations versus Day of Month

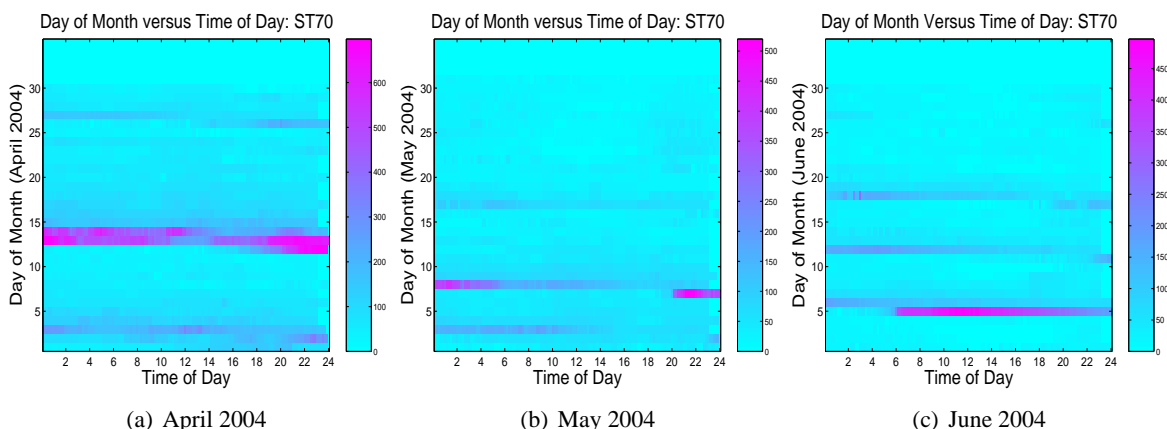(a) April 2004   (b) May 2004   (c) June 2004

Figure 11: The Data Cube: Time of Day versus Day of Month (ST70)

The data cube has obvious advantages over ordinary visualization methods, e.g. data curves, especially when we need to combine two different dimensions to obtain aggregated views of the original data. There are many possible combinations of dimensions, for example, two different time granularities, or the spatial dimension versus the temporal dimension. With the help of data cubes, interesting patterns or relationships can be identified very effectively.

5 Discussion

Real-time data processing systems must collect and analyze information in a very short period of time. The quick response of a Web-GIS system depends on two key factors: network speed and database processing speed. As the number of users increases, the performance of the network and database server will degrade dramatically. To reduce system delay effectively, users' requests need to first be analyzed so that different strategies can be used to handle different tasks. If the request involves large amounts of data but the query result is small in size, server-side scheme is the best choice. This minimize the cost of transmitting huge amounts of data through the Internet, although it requires some computation
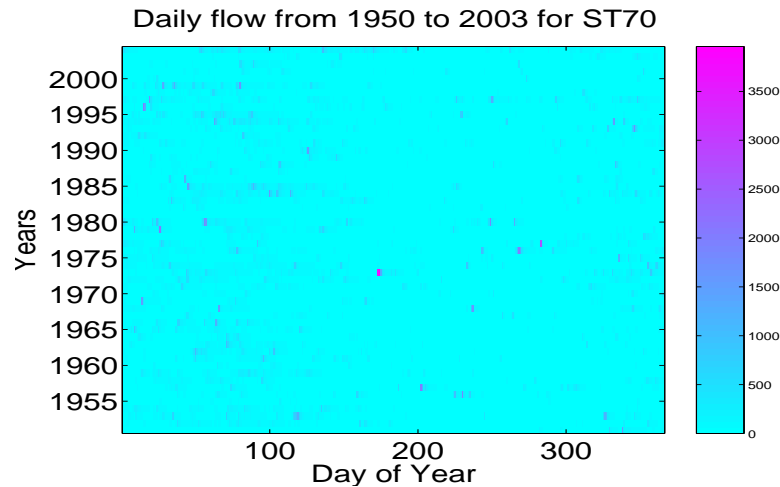
Figure 12: The Data Cube: Years versus Day of Year

resources on the server side. If the request is "computation intensive", requiring small amounts of data but long CPU times, a client-side scheme will be preferred. For example, as the demand for 3D visualization grows, we may need to transfer more data processing to the client side to reduce the server's burden. Our future plans include the provision of a dynamic request classification method in order to decide which scheme to use, "client-side" or "server side". This can be achieved by analyzing the user request, for example, the requested data size and desired output format. The importance of real-time data processing lies not only in online monitoring and visualization, but also in trend analysis and prediction. It is important to be able to build a prediction model which can explore the correlations among multiple stations based on historical data, and extract practical association rules to predict if a flood will occur or the impact of a pollution event. Spatial data mining techniques such as spatial outlier detection and spatial association rule identification will be employed to analyze the relationship between spatial attributes (location, altitude, etc.) and non-spatial attributes(e.g., flow, Oxygen density). Also, the weather data set should be integrated into the prediction model, since precipitation also plays an important role in flood prediction.

GIS data is much more complex than traditional transaction data. It contains not only nonspatial attributes such as numbers and characters but also spatial properties such as location, size, and outline. Thus efficient visualization methods are in great demand for accurate and intuitive data representation. Currently, our system provides only 1D and 2D data visualization. An effort is underway to develop 3-D representations, for example, a data cube for TimeOfDay-DayOfMonth-MonthOfYear. Also, pie charts and histograms will be supported to illustrate the statistics of various attributes. A zoom functionality is currently under development, in order to provide multiple-resolution views for different purposes. For online pollution monitoring, detailed data may be required to be collected every 5 minutes, whereas to see the flow pattern of a particular year, we need only daily flow data. A more intelligent visualization system should be developed so that the accuracy levels of data can be dynamically adjusted according to the user's query request.

In addition to the current stations, more monitoring equipment can be installed and more useful attributes can be collected for analysis and visualization. A multi-parameter water quality monitor will shortly be installed at one of the stations on Bull Run. This monitor will measure dissolved oxygen,

water temperature, specific conductance, pH, and nitrate concentrations in Bull Run. In addition a buoy-mounted multi-parameter monitor may be installed in the Occoquan Reservoir in the near future. The buoy mounted system will be capable of making measurements through the water column to provide near real-time water quality depth profiles in the reservoir. Lastly, the OWML maintains a network of twelve recording rain gages. Most of these are located in remote locations where phone lines are not available. With the appropriate wireless communication equipment, these gages could be queried by the data collection server and then their information displayed on the web.

6   Conclusion

In this paper, we propose a web-based data visualization and analysis system for watershed management of the Occoquan Watershed Basin in Northern Virginia. This system monitors real-time watershed data through a distributed sensor network, visualizes the information in textual or graphic format on the Internet, and provides data analysis based on the data from the past 50 years. This system distinguishes itself from other systems in three aspects. First, it enables near real-time data collection and publication. Second, it supports web-based data visualization and query. Third, we employ the concept of the "data cube" to present aggregation views of historical data with various degrees of granularity. The system has been deployed by the Occoquan Watershed Management Lab(OWML) and is successfully providing online data monitoring functionalities. In the near future, more data analysis functionalities and visualization formats will be incrementally added into the system, such as 3-D data aggregation and its visualization.

References

[1] O. Boucelma, J.-Y. Garinet, and Z. Lacroix. The virgis wfs-based spatial mediation system. In *CIKM '03: Proceedings of the twelfth international conference on information and knowledge management*, pages 370–374. ACM Press, 2003.

[2] V. Coors and S. Flick. Integrating levels of detail in a web-based 3d-gis. In *GIS '98: Proceedings of the 6th ACM international symposium on advances in geographic information systems*, pages 40–45. ACM Press, 1998.

[3] C. du Mouza and P. Rigaux. Web architectures for scalable moving object servers. In *GIS '02: Proceedings of the 10th ACM international symposium on advances in geographic information systems*, pages 17–22. ACM Press, 2002.

[4] S. Flick. An object-oriented framework for the realization of 3d geographic information systems. In *JEC-GI '96: Proceedings of the second joint European conference & exhibition on geographical information (Vol. 1) : from research to application through cooperation*, pages 187–196. IOS Press, 1996.

[5] J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, and H. Pirahesh. Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. *Data Min. Knowl. Discov.*, 1(1):29–53, 1997.

[6] R. H. Guting. An introduction to spatial database systems. *The VLDB Journal*, 3(4):357–399, 1994.

[7] J. Han, J. Y. Chiang, S. Chee, J. Chen, Q. Chen, S. Cheng, W. Gong, M. Kamber, K. Koperski, G. Liu, Y. Lu, N. Stefanovic, L. Winstone, B. B. Xia, O. R. Zaiane, S. Zhang, and H. Zhu. Dbminer: a system for data mining in relational databases and data warehouses. In *CASCON '97: Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research*, page 8. IBM Press, 1997.

[8] J. Han, K. Koperski, and N. Stefanovic. Geominer: a system prototype for spatial data mining. In *SIGMOD '97: Proceedings of the 1997 ACM SIGMOD international conference on management of data*, pages 553–556. ACM Press, 1997.

[9] K. Koperski and J. Han. Discovery of spatial association rules in geographic information databases. In *SSD '95: Proceedings of the 4th International Symposium on Advances in Spatial Databases*, pages 47–66. Springer-Verlag, 1995.

[10] R. Peng. An assessment of the development of internet gis. 1997.

[11] A. Stockus, A. Bouju, F. Bertrand, and P. Boursier. Integrating gps data within embedded internet gis. In *GIS '99: Proceedings of the 7th ACM international symposium on advances in geographic information systems*, pages 134–139. ACM Press, 1999.

[12] S. Tu, X. He, X. Li, and J. J. Ratcliff. A systematic approach to reduction of user-perceived response time for gis web services. In *GIS '01: Proceedings of the 9th ACM international symposium on advances in geographic information systems*, pages 47–52. ACM Press, 2001.

[13] S. Tu, L. Xu, M. Abdelguerfi, and J. J. Ratcliff. Achieving interoperability for integration of heterogeneous cots geographic information systems. In *GIS '02: Proceedings of the 10th ACM international symposium on advances in geographic information systems*, pages 162–167. ACM Press, 2002.

[14] R. R. Vatsavai, T. E. Burk, B. T. Wilson, and S. Shekhar. A web-based browsing and spatial analysis system for regional natural resource analysis and mapping. In *GIS '00: Proceedings of the 8th ACM international symposium on advances in geographic information systems*, pages 95–101. ACM Press, 2000.

[15] Z.-K. Wei, Y.-H. Oh, J.-D. Lee, J.-H. Kim, D.-S. Park, Y.-G. Lee, and H.-Y. Bae. Efficient spatial data transmission in web-based gis. In *WIDM '99: Proceedings of the 2nd international workshop on Web information and data management*, pages 38–42. ACM Press, 1999.