International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

AMADEUS: A System for Monitoring Water Quality Parameters and Predicting Contaminant Paths

Abdeltawab M Hendawi[1], David Hazel[1], Joel Larson[1], YiRu Li[1], Dwaine Trummert[1], Mohamed Ali[2], Ankur Teredesai[1]
[1]*Institute of Technology, University of Washington, Tacoma, WA, USA*
{hendawi, dhazel, jilarson, yiruli, dwainet, ankurt}@uw.edu
[2]*Microsoft Corporation, One Microsoft Way, Redmond WA 98052*
mali@microsoft.com

Abstract: Managing the water quality in an urban environment is extremely challenging. While it flows, the water picks up pollutants such as lawn care chemicals, oil, and pet waste bacteria. In fact, topography plays a factor in where water runoff goes. However, there are many other factors, such as urban density, impermeable surface coverage, weather events and tidal patterns which all have the potential to impact not only the final destination of a particular pollutant but also the rate of travel along the route. In this paper, we propose a system, named AMADEUS (Azure Marketplace of Applications for Diverse Environmental Use as a Service), which is an interactive, self-service framework that allows end users to explore, analyse, and visualize the environmental data within the context of their applications.

As a case study, we present a sample application on AMADEUS which aims to identify contaminant sources in the Puget Sound region. AMADEUS integrates chemical spill data, meteorological data, Puget Sound buoy data, and water runoff models to perform pollutant path tracking and prediction. More specifically, given a water fall location, AMADEUS is able to identify the runoff path, compute the impact of environmental factors. For example, it can trace back the pollutant to its source, and predict the final destination of the pollutant. In addition, AMADEUS provides user friendly visualization to demonstrate the tracking and prediction of pollutants' routes.

***Keywords**: GIS and Environmental Science; World Wide Telescope; Microsoft StreamInsight and Windows Azure; Water Quality Monitoring.*

1 Introduction

Over the past few decades, there has been a massive research contribution on monitoring and managing water quality indicators [Zhang et al 2011, Huang et al. 2001]. To build a system that can accurately and efficiently measure and interpret those indicators; we need to correlate a wide range of relevant data sources. These sources include but are not limited to: pollution origins, water quality readings, tide level, land topography, i.e., digital elevation model (DEM), and land use [Wang 2001]. However, the heterogeneity and unavailability of these sources inside a single database are challenges to water quality and ecosystems researchers. Collecting essential data from different related sectors empowers the study and the analysis of environmental science, in general, and water quality, in specific.

One of the powerful tools that has been widely used in this domain is the GIS (Geographic Information Systems) platform [Martin et al. 2004, Thomas et al. 2005]. The usage of GIS tools enables the visual inspection of possible correlations among relevant data sources, e.g., water network and contaminant locations [Ivanov, et al 2008]. Moreover, users can explore the distribution of contaminant sources over the map. Consequently, they can visually deduce existing relationships among contamination events and link them back to their possible origins. In addition, users can study their possible effects on neighbouring places on the map. Hence, suitable decisions are taken [Maciejerski et al. 2011]. Decisions can be a plan for resource allocation to clean up those areas and preventative actions to avoid the case originally.

In this paper, we propose the AMADEUS (Azure Marketplace of Applications for Diverse Environmental Use as a Service) system. The main services AMADEUS provide are as follows:

(1) The ability to consume data from a wide variety of environmental data sources, and to store them in one integrated data store. To guarantee the availability and accessibility of this data, AMADEUS hosts all data in the *Microsoft* cloud platform, i.e., *Windows Azure*.

(2) The luxury to query and analyse the underlying data sets by writing SQL queries and/or through the system's user-friendly graphical user interface.

(3) The ability to study the correlation between multiple data sets using the algorithms and data structures natively offered within the framework. For example, the integration between chemical spill data and land surface topology allows the prediction of the possible paths pollutants would follow.

(4) The visualization capabilities inherited by adopting the World Wide Telescope, a visualization engine developed by Microsoft research, is employed as a visualization front end for AMADEUS.

As an example problem of what is being addressed under its umbrella, AMADEUS identifies the runoff path of a water fall, computes the impact of environmental factors along the path, and traces back pollutants along the path to their origins. The main idea is to leverage the reachability tree [Abdeltawab 2013] to obtain the locations that are most likely to be reached by a moving flow of contaminants. These reachable locations depend on various factors such as (a) the elevation of the surrounding land surface, and (b) the source, speed and direction of this flow. A probability value will be assigned to each possible location according to its reachability by the source of contamination, e.g., a chemical spill. Based on that probability, users are provided with a visualization of the pollutant's anticipated spread.

In the rest of the paper, we give an overview of the AMADEUS system, highlight its modules, and describe the underlying technologies. Then, we present the case study of monitoring contaminant sources in the Puget Sound region.

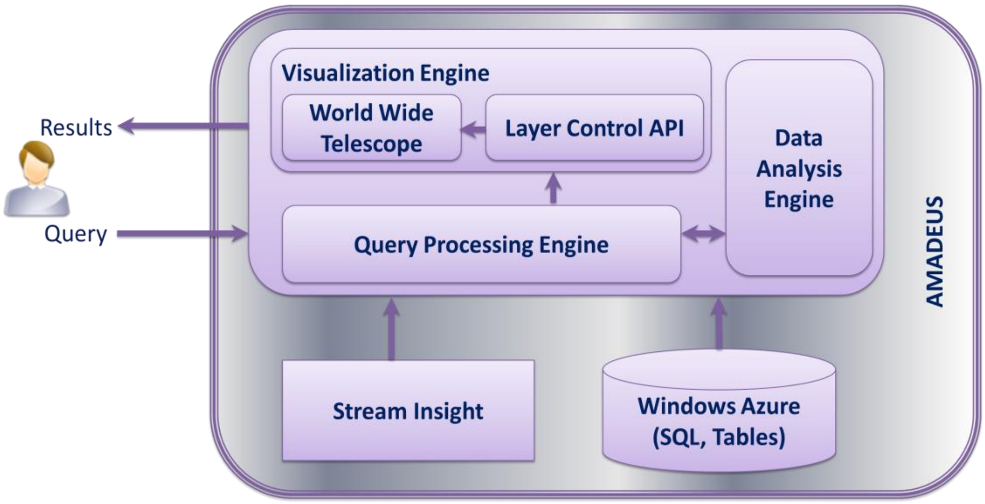


Figure 1: The System Architecture of AMADEUS

2 System Overview

AMADEUS is an interactive and self-service framework that allows end users to consume and understand environmental data within the context of their world. Through the same interface, users choose relevant data sets from a variety of data sources, and utilize a library of available tools to intelligently query, explore and understand the hidden correlations among these data sets. Then, users utilize these consolidated data sets to drive powerful visualizations that are rendered and explored in a number of ways. These visualization capabilities help the user understand the data within the context of external factors. Due to the heterogeneity of Environmental Data sources and formats, it is often difficult for researchers to gather and combine these sources into a usable form. AMADEUS handles this data burden, so that the end user focuses on deriving actionable insights from the data exploration.

The system architecture of AMADEUS (Figure 1) consists of three main modules, namely, the query processing engine, the data analysis engine, and the visualization engine. In addition, AMADEUS accesses two types of data management frameworks in the cloud, specifically, (a) Windows Azure and SQL Azure for large volume look-up tables and traditional database tables, respectively, and (b) Microsoft StreamInsight for real time streaming data. In the following sections we describe each one of these modules.

2.1 Data Access

Many data sources have already uploaded (or are currently streaming) their data to the Azure cloud which is considered the AMDEUS main persistent storage. These wide varieties of environmental data sets are collected from multiple data providers, e.g., Encyclopaedia of Puget Sound, Pierce County, City of Tacoma and Puget Sound buoys. These data sources are described in Section 3. In addition, AMADEUS consumes data streams obtained from different sensors on a real time basis, e.g., water level sensors and temperature sensors. These streams are continuously extracted via the Microsoft StreamInsight engine that is described towards the end of this section. This sensor data is provided by GIS division at the Pierce County. Through our research partners and collaborators, we have assembled a wide array of proprietary as well as well as publically available data sources. These data sets range from stream flow data, water quality data, weather data, chemical spill data, geological survey data, and others. Because aggregating, migrating and cleaning data sets can be cumbersome and time consuming, a sub goal of this project is to integrate, clean and upload these data sources to the cloud. Then, we give access to interested researchers to a wealth of pre-processed data sets.

In fact, in many instances, the data is dirty, outdated and sparse with many missing values in the time series. As part of the data migration effort we do the following:

- Given the noise that is due to the age of many of these data sets, we perform a data cleaning step to improve the data quality.
- Combine disparate data sets and unify the same data from multiple vendors or partners taking into consideration all necessary transformations and mappings along the unification process.

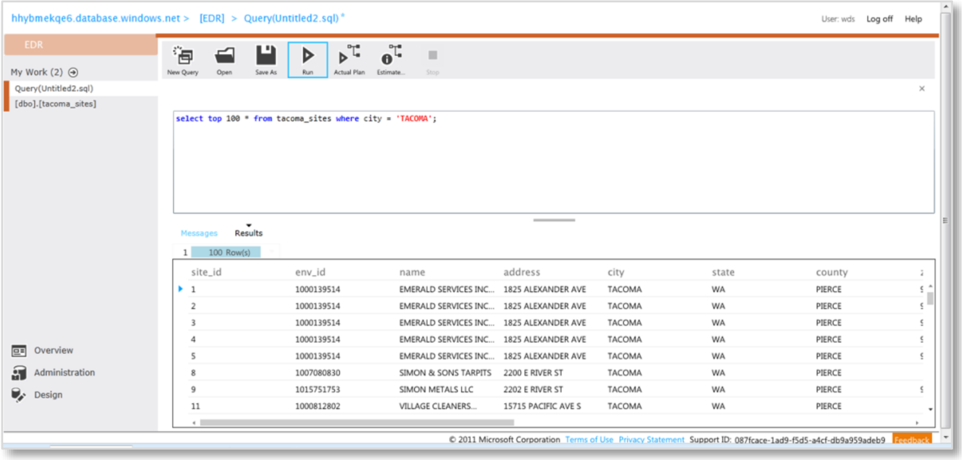
AMADEUS utilizes two infrastructures to acquire and store data:

Window Azure is a cloud computing platform and infrastructure, provided by Microsoft, for building, deploying and managing applications and services through a global network of Microsoft-managed data centers. It includes a number of separate features with corresponding developer services which can be used individually or together, e.g., compute, data, networking and app services. Data Services provide the ability to store, modify and report on data in Windows Azure. Benefits include manageability, high availability, high scalability, and a familiar development model. While Azure tables provide a large scale lookup table repository for key/value pairs, SQL Azure is a structured relational database management system in the cloud. Azure tables and SQL Azure complement each other and blend the scalability and manageability of the data into a cloud based system.

StreamInsight is a high-throughput stream processing platform that is based on the Microsoft .NET Framework. StreamInsight enables developers to quickly implement robust and highly efficient event processing applications. StreamInsight facilitates the development of complex event processing applications, derives immediate business value from high input rate raw data, and reduces the cost of extracting, exploring, analysing and correlating streams of data. Also, it allows users to monitor, process, search, and mine the data for conditions, opportunities, and defects almost in real time [Ali et al. 2010].

2.2 Query Processing Module

The Query Processing Module receives user queries that are composed either by writing SQL queries or through the AMADEUS graphical interface. Then, the query processor accesses the referenced data source at Azure, optimizes and executes the queries, and generates the results as illustrated in Figure 2. Once a query is executed successfully, users have the options to either download the query result for their own purpose, or/and they visually inspect the result using the AMADEUS visualization module.



```
select top 100 * from tacoma_sites where city = 'TACOMA';
```

1 100 Row(s)

site_id	env_id	name	address	city	state	county
1	1000139514	EMERALD SERVICES INC...	1825 ALEXANDER AVE	TACOMA	WA	PIERCE
2	1000139514	EMERALD SERVICES INC...	1825 ALEXANDER AVE	TACOMA	WA	PIERCE
3	1000139514	EMERALD SERVICES INC...	1825 ALEXANDER AVE	TACOMA	WA	PIERCE
4	1000139514	EMERALD SERVICES INC...	1825 ALEXANDER AVE	TACOMA	WA	PIERCE
5	1000139514	EMERALD SERVICES INC...	1825 ALEXANDER AVE	TACOMA	WA	PIERCE
8	1007080830	SIMON & SONS TARPITS	2200 E RIVER ST	TACOMA	WA	PIERCE
9	1015751753	SIMON METALS LLC	2202 E RIVER ST	TACOMA	WA	PIERCE
11	1000812802	VILLAGE CLEANERS...	15715 PACIFIC AVE S	TACOMA	WA	PIERCE

Figure 2: Querying the chemical spills data in Windows Azure

2.3 Data Analysis Module

The Data Analysis Module assists users to search inside the data for hidden patterns using a set of built-in tools and algorithms. For example, users are able to combine two or more data sets and study the correlation among these data sets. As an example, users dispatch prediction algorithms to anticipate the future path of a given source of water. As will be shown in the case study, AMADEUS can join the watersheds, rivers and land surfaces networks with the chemical spill data sets in attempt to predict the future possible destinations of the pollutants coming from these spills. Similarly, the data analysis module can help users trace back the contaminant to their most likely sources. In this specific example, the algorithm used in the prediction of a chemical spill path is based on leveraging the reachability tree data structure [Abdeltawab et al. 2012, 2013].

2.4 Visualization Module

The front end data visualization in AMADEUS adopts the World Wide Telescope (WWT). Users can leverage the WWT to visualize their queries results through adding data layers to the place of focus on the earth. The back end of the visualization module employs the layer control APIs (LCAPI) to customize the behaviour of the WWT such that a user can easily add, update and delete a layer. Below, we give a general overview about WWT and its LCAPIs.

World Wide Telescope is powerful visualization and media authoring kit that allows end users to explore the earth, the planets, the solar system and the sky. It serves as a tool and valuable resource for both astronomical and educational communities, as well as a way to make science interesting to the lay person. The WWT application brings together imagery from the best ground- and space-based observatories across the world. WWT uses a high-performance Microsoft Visual Experience Engine to allow seamless and interactive panning and zooming over terabytes of stitched-together imagery data. Images are loaded on demand in order to converse network bandwidth, with several optimizations to enhance the viewing experience. A particularly interesting use of WWT is as a viewer of a 3-D model of earth, similar to Microsoft Virtual Earth and Google Earth. This application uses Virtual Earth's satellite and map images, but is instead powered by the WWT engine [Ali et al. 2011].

Data Source	Purpose / Utilization
Puget Sound Buoys	Buoy data is currently supplied in 15-minute or 1-hour cycles to a website for manual consumer consumption. The infrequency of this data means that averages must be used for periods of time. Access to a web service can be utilized with StreamInsight to provide real-time data to the analysis and layering model to provide accurate water quality data.
Chemical Spills	Chemical Spill information is consolidated from multiple sources by ERD. This information provides spill report date, geo-location, type and quantity of spill. Frequency and size of known spills can be used to help with predictive models for predicting spill source when joined with other sources such as, tidal information and stream flow networks.
City of Tacoma Infrastructure	City of Tacoma provides mapping of each sanitary and storm water segment and data to connect path from the first segment to the last, depending on if it ends at a treatment plant (sanitary) or empties directly into a local waterway (storm water). This data can be used in conjunction with County data to identify pipe network segments and specifically how water is directed from a spill site to a treatment plant or waterway. County data currently does not track routing of piped system other than source and destination.
Pierce County	In addition to the information provided by the City of Tacoma, Pierce County has data defining each watershed in the county and the stream flow network of where that water will flow. This brings together the sanitary and storm water information with the stream flow and watershed information to give a complete picture of where a chemical would travel after a spill.
LIDAR Network	LIDAR data is detailed mapping by satellite that also includes elevation. In conjunction with other data, LIDAR can help to provide a three-dimensional look at water flow.

Table 1, Description of underlying data sources

Layer Control API provides an extensive range of functions to transmit and receive data to and from World Wide Telescope. In general, Layer Control API (LCAPI) enables the customization of data and the development of customized interfaces through an API, such as importing data into the application. For a large volume of data, LCAPI appropriately controls the size of buffers for data retrieval and visualization. For a smaller amount of data the entire data chunk could be loaded in a single session.

2.4 System Users

The environmental protection agencies (EPA) is one strong candidate user for our system. They are aware about the importance of predicting the areas that have high likelihood to be affected by chemical pollutants. AMADEUS will help to identify those areas, so right decision can be taken. Also, the city can use this system to know who in charge for those discovered contaminated areas of the land. In addition, for educational awareness, AMADEUS, can be used to show the severe effect of a chemical spill on the surrounding areas.

3 Case Study: Chemical Spills in Puget Sound

Managing the water quality in an urban environment is extremely challenging, this is especially true in the Pacific Northwest where much of our storm water runoff flows into drains and ditches, emptying directly into our lakes, rivers, streams and ultimately the Puget Sound. As it flows, the water picks up pollutants such as lawn care chemicals, oil, grease, car wash soap, and pet waste bacteria. Topography plays a factor in where this runoff goes, but there are many other factors, such as urban density, impermeable surface coverage, infiltration and saturation rates, weather events and tidal patterns which all have the potential to impact not only the final destination of a particular pollutant but also the rate of travel, and dilution rates along the route.

In this case study, through the intelligent integration of chemical spills data, water runoffs modelling, and Puget Sound buoys data, we build a prediction model inside AMADEUS that can, for a given environmental location, visually demonstrate the runoff path. Figure 3 gives an example for the Puget Sound buoy ORCA (Ocean Remote Chemical Analyser) used to collect water measurements and send them as data streams.

Additionally, we demonstrate the ability of the proposed model, given an identified event such as pollutant in a water region, to trace back that pollutant to its possible sources. Initially, we obtained a number of the relevant data sources, and we successfully hosted them into Windows Azure. Table 1 provides a brief description for those data sources.

The core of the AMADEUS logic is implemented as a windows client application using C# in addition to the LCAPI library to contact with the WWT. Through this application, users can interact with the AMADEUS storage on the Azure cloud by writing SQL queries to get specific data that matches their interest. Then they can set the parameters that control the behaviour of the WWT to customize the results visualization. For example, a user can write a query to access the chemical spills data for the Puget Sound area.

The returned results for that query are visualized in different layers on the WWT. As depicted in Figure 4, the selected chemical spills in that area are represented as red dots. The size of each dot indicates the volume of the chemicals come from that spill. By combining the water follow prediction model with the chemical spills data, users can also see the possible routes that spills might run to. By doing so, users can have an anticipation for the areas that might be effected by the pollutants coming from those spills. The use of World Wide Telescope allows the addition of temporal data to show the effects over time. Thus, this 4-D modelling allows for the tracking and prediction of contaminants forward and back. In turns, this allows response teams to know when and where to place mitigation devices as well as to trace back to a potential source spill.

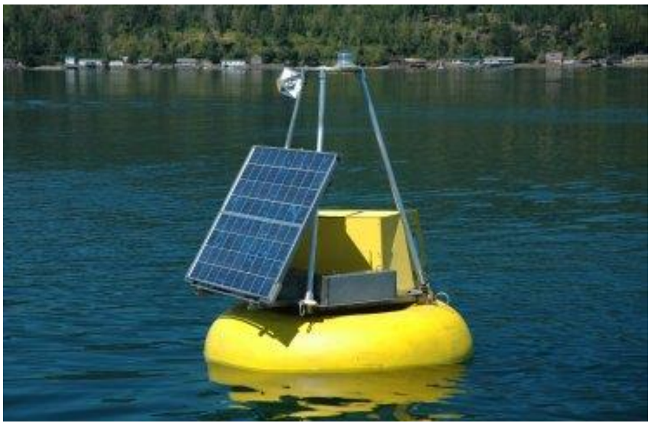
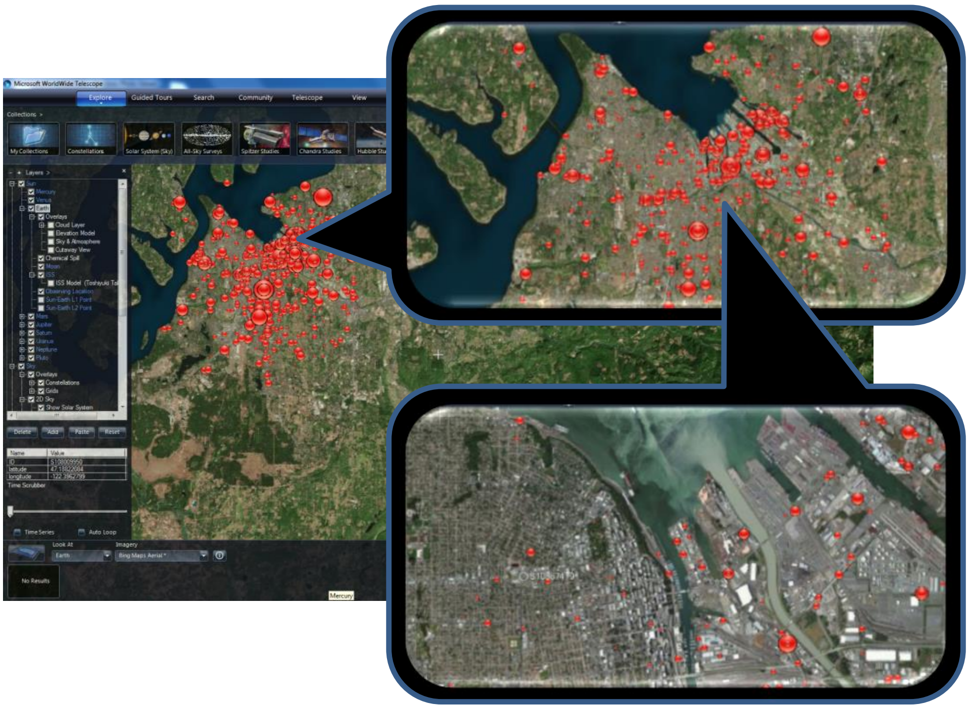


Figure 3, Puget Sound ORCA (Ocean Remote Chemical Analyser)

.

4 Conclusion and Future work

In this paper, we introduced the AMADEUS (Azure Marketplace of Applications for Diverse Environmental Use as a Service) system which aims at integrating a wide variety of environmental data sources along with providing querying tools and analysis algorithms. Users interact with AMADEUS in different ways to perform data manipulation, analysis and visualization. Users can write SQL queries to fetch their data of interest, e.g., chemical spills data, run prediction modules to predict the future path of a certain pollutant and/or trace back its source. In addition, AMADEUS leverages the World Wide Telescope as the front-end visualization engine to give users the ability to have 2D, 3D, and 4D view for the results. We also provided a real world case study as a sample application with AMADEUS to identify contaminant paths within the Puget Sound region.



Name	Value
ID	5108009950
latitude	47.18822084
longitude	-122.3962799

Look At: Earth

Imagery: Bing Maps Aerial *

Figure 4, AMDEUS front end visualization using WWT

ACKNOWLEDGMENTS

This research is partially supported by a research grant from Microsoft Research Connections. This grant provided cloud computing and storage resources on Windows Azure "http://research.microsoft.com/en-us/projects/azure/projects.aspx". Many thanks to our partners, (City of Tacoma, Center for Urban Waters, Encyclopedia of Puget Sound, Pierce County GIS Division, EDR) who devoted their time to provide us with the required data and knowledge.

REFERENCES

Abdeltawab M. H., Mohamed F. M., 2012. "Panda: A Predictive Spatio-Temporal Query Processor". In Proceedings of the 20th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL GIS, Redondo Beach, CA, USA.

Abdeltawab F. H., Jie B., Mohamed F. M., 2013." iRoad: A Framework For Scalable Predictive Query Processing On Road Networks ". In Proceedings of the International Conference on Very Large Databases, VLDB, Riva Del Garda, Italy.

Huang, G. H., Xia, J., 2001. Barriers to sustainable Water-quality Management. Journal of Environmental Management, vol. 61, pp. 1-23.

Ivanov, A. Yu., Zatyagalova, V. V., 2008. A GIS Approach to Mapping Oil Spill in a Marine Environment. International Journal of Remote Sensing, 29:21, pp. 6297-6313.

Maciejewski, R, Hafen, R., Rudolph, S., Larew, S. G., Mitchell, M. A., Cleveland, W. S., Ebert, D. S., 2011. Forecasting Hotspots – A Predictive Analytics Approach. IEEE Tansactions on Visualization and Computer Graphics, Vol. 17, No. 4, pp. 440-453

Martin, P.H., LeBoeuf, E. J., Daniel, E. B., Dobbins, J. P., Abkowitz M. D., 2004. Development of a GIS-based Spill Management Information System. Journal of Hazardous Materials B112, pp. 239-252.

Mohamed Ali, Badrish Chandramouli, Balan S. Raman, Ed Katibah, 2010. Real-Time Spatio-Temporal Analytics using Microsoft StreamInsight. In Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS) San Jose, CA, USA.

Mohamed Ali, Badrish Chandramouli, Jonathan Fay, Curtis Wong, Steven Drucker,, Balan Sethu Raman, 2011. Online Visualization of Geospatial Stream Data using the WorldWide Telescope. In Proceedings of the International Conference on Very Large Data Bases (VLDB).

Thomas, J.J., Cook, K. A., 2005. Illuminating the Path: The R&D Agenda for Visual Analytics. IEEE Press.

Wang, X. 2001. Integrating Water-quality Management and Land-use Planning in a Watershed Context. Journal of Environmental Management, vol. 61, pp. 25-36.

Zhang H., Huang G.H., 2011. Assessment of non-point source pollution using a spatial multicriteria analysis approach, Ecological Modelling, Volume 222, Issue 2, 24 January 2011, Pages 313-321.