ELSEVIER

Computers & Geosciences 31 (2005) 403–413

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

# A relational database for the monitoring and analysis of watershed hydrologic functions: II. Data manipulation and retrieval programs

Christian J. Carleton[a], Randy A. Dahlgren[a,*], Kenneth W. Tate[b]

[a]*Department of Land, Air, and Water Resources, University of California, Davis, CA 95616, USA*
[b]*Department of Agronomy and Range Science, University of California, Davis, CA 95616, USA*

Received 8 December 2003; received in revised form 24 September 2004; accepted 6 October 2004

**Abstract**

With increasing hydrologic and water quality data being collected for monitoring and research purposes, it has become evident that storing data in traditional spreadsheet formats are no longer a practical means of managing these large datasets. The Watershed Monitoring and Analysis Database (WMAD) was developed to ease the burden of handling watershed-scale hydrology, water quality, and meteorology data. WMAD is a relational database designed to make the management of large hydrologic datasets more efficient and less prone to manipulation errors. However, this increased efficiency comes at the cost of needing to understand relational database engines, including the ability to write queries to retrieve data. To facilitate the use of WMAD, we have applied the WMAD structure using the Microsoft Access database application. Separate WMAD applications were developed within the Access environment that allow users to access and manage data without having to understand details of the database. These applications are graphically based and allow the user to perform a series of tasks including: data import, data management, basic calculations, and data export. A constituent flux or load calculation was developed to perform flux calculations using a time-weighted algorithm. This paper highlights these applications and their tasks to assist new WMAD users in quickly becoming proficient in using WMAD for data management.


*Keywords:* Hydrology; Non-point source pollution; Relational database; TMDL; Water quality

## 1. Introduction

The need for efficient storage and data manipulation tools for hydrologic/water quality data has dramatically increased in recent years. Increased data collection is due to a renewed focus on non-point source water pollution and the ability to acquire vast amounts of temporal data with automated instrumentation. For example, the Environmental Protection Agency's (EPA's) Total Maximum Daily Loads (TMDL) program requires ambient monitoring across the nation to identify impaired water bodies. States have already identified about 21,000 polluted river segments, lakes, and estuaries affecting over 480,000 river and shore kilometers and 2 million lake hectares (EPA, 1997). It is estimated that more than 40,000 TMDLs must be developed during the next 10–15 years to address these impaired water bodies (National Research Council, NRC, 2001). Each TMDL requires extensive pre- and

*Corresponding author.
*E-mail address:* radahlgren@ucdavis.edu (R.A. Dahlgren).

0098-3004/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2004.10.008

post-TMDL development monitoring to identify pollutant sources and the effectiveness of best-management practices and watershed restoration projects to mitigate pollutant loadings. These monitoring programs require simultaneous collection of data on several meteorological, hydrological and water quality parameters. In many cases, automated equipment is utilized to collect data at a time-step ranging from seconds to hours. The combined effect of multiple parameters and intensive temporal data collection results in large data sets. These data must be stored in a format that can be efficiently retrieved and manipulated to calculate parameters of interest (e.g., daily pollutant loadings) and the relationships between water quality variables. Data management problems can quickly become overwhelming when dealing with these large data sets, thereby hindering the data analysis process.

We developed the Watershed Monitoring and Analysis Database (WMAD) to deal with hydrologic/water quality data management issues resulting from large and complex data sets. WMAD is a relational database design that allows for the management and manipulation of meteorological, hydrological and water quality data (Carleton et al., 2005, this issue). WMAD uniformly stores all three types of data within a single file, allowing the database engine to organize and manage the data. This saves storage space and reduces data retrieval and manipulation times. The database structure also minimizes errors associated with manual spreadsheet calculations, such as cutting and pasting of data formulas. However, there are some problems associated with managing data in a relational database. Data retrieval and manipulation is achieved through the use of Structured Query Language (SQL), which requires knowledge of both SQL programming and overall database design. Initially, this can make retrieving raw data or performing calculations difficult. In addition, some calculations cannot be performed within a query and require a non-SQL solution, such as the calculation of pollutant fluxes or loads and statistical analysis.

To help facilitate data retrieval and expedite data analysis, we developed data accessing programs that work with a Microsoft Access version of the WMAD database. These applications were written using Visual Basic for Applications, the built in Access development environment. They provide a graphical user interface that prompts for required information, queries the desired data, and then returns results for viewing or export to external files for statistical analysis, graphing or web-based applications. These applications eliminate the need to understand SQL and the database design.

Applications were also developed using algorithms to perform various tasks not easily performed within a relational database environment. These include an import process and a pollutant flux or load calculation program. The import program takes a flat or spreadsheet file, and converts it into the WMAD relational format allowing it to be imported. A pollutant flux application was developed to calculate pollutant loads (mg) from pollutant concentrations (mg/L) and stream discharge (L/s). Fluxes are calculated for user-defined time intervals (e.g., daily or annual fluxes) over a specified time period (e.g., storm event). All data manipulation programs were developed using the Data Access Object (DAO) model (Haught and Ferguson, 1997).

The primary objectives of this paper are to provide an overview of the basic WMAD applications and to provide an in depth discussion of how the pollutant flux algorithm applies the time-weighted approach to calculating constituent loads. Calculation of constituent loads is a key issue in TMDL development and monitoring. Thus, the WMAD database will provide an important tool for data management and manipulations in TMDL monitoring.

## 2. Applications

The WMAD database design was applied using the Microsoft Access relational database program. Database accessing applications were developed using the built in Access development environment based upon Visual Basic for Applications (VBA) and using the DAO model (Haught and Ferguson, 1997). These applications dynamically create the queries covered by the previous paper (Carleton et al., 2005, this issue). Each individual user interface screen or form prompts for the desired information needed to construct that specific query, such a location and date range. The query is then dynamically constructed based upon the provided information, executed, and the data are returned for the user to view. These results can be exported in a comma delimited text format or as a Microsoft Excel spreadsheet for input into a statistical or graphical application. This allows one to easily extract only the necessary data from the database required to perform a desired analysis.

There are four separately developed applications: (1) References, (2) Data Access, (3) Quality Assurance/Quality Control (QA/QC), and (4) Stream Fluxes. Each of these applications consist of multiple forms that were bundled together with the end analysis in mind, grouping together forms that returned similar or related data set types. The tasks performed by each of these forms within the four applications can be grouped into one of three categories: (1) Data Management, (2) Data Manipulation and (3) Data Calculations. The forms and their functions will be described in this paper based upon their task grouping and not their parent application.