The Thesis Committee for Fernando Renzo Salas
Certifies that this is the approved version of the following thesis:

Design and Development of a Hydrometeorological Information System

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

____
David R. Maidment

____
Daene McKinney

Design and Development of a Hydrometeorological Information System

by

Fernando Renzo Salas, B.S.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

May 2010

Dedication

I would like to dedicate this thesis to my family and friends especially my parents, Mami y Papi, for their unconditional love and support. Thank you!

Acknowledgements

I would like to first and foremost thank my advisor Dr. David Maidment for his continuous guidance, support, energy and inspiration throughout this work.

I would also like to thank the Consortium of Universities for the Advancement of Hydrologic Science Inc. for supporting this work. I appreciate the opportunity to advance the field of hydrologic science and I look forward to continuing this work in the future and achieving the vision I have for this topic.

I would like to thank Dr. Xingong Li for providing a significant amount of the data that is mentioned in this thesis.

Finally, I would like to thank my colleagues, past and present, at the Center for Research in Water Resources for all their help and support throughout this process.

May 2010

Abstract

Design and Development of a Hydrometeorological Information System

Fernando R. Salas, M.S.E.

The University of Texas at Austin, 2010

Supervisor: David R. Maidment

Work has been successfully performed by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) to synthesize the nation's hydrologic data. Through the building of a national Hydrologic Information System, the organization has demonstrated a successful structure which promotes data sharing. While the access to and synthesis of observations data has been improved, the same cannot be said for spatiotemporal data. Data available on grids (e.g. meteorological data) has yet to be organized in a manner that can easily be shared. The development of a national Hydrometeorological Information System which incorporates temporal, spatial and spatiotemporal data is necessary for the improved synthesis of hydrologic phenomena which inherently is a coupled process that integrates both principles in hydrologic science and meteorology.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

"Web users ultimately want to get at data quickly and easily. They don't care as much about attractive sites and pretty design."

~ Tim Berners-Lee

Data is everywhere. It is the coffee you drink, it is the time at which you drink it, it is the location from which you purchase it, it is the amount of time you spend drinking it. Everywhere you look, there is data. As new technology emerges, more sophisticated methods for gathering, storing and analyzing data, from the data-sphere in which we inhabit, become available. Smart devices, processors, sensors etc. are all being utilized and imbedded into the everyday objects we interact with. Every single second, minute, hour and day, we learn more about society, the environment and the world than any other generation could ever fathom. It is becoming evident that data is changing the world; it is changing the manner in which we do the old and enabling us to strive for the new.

As a society, it has always been in our second nature to collect data. It is the means by which we learn and understand the processes which occur around us and the means by which we gain insight, reach conclusions and undergo revelations. In the last 20 years, the internet has evolved into one of the main vehicles through which humans exchange information. It has transformed from a simple repository of information into a communal platform for participation (Hendler et al., 2010). Social networks like Wikipedia, Facebook, YouTube and Twitter have all flourished because of this transformation. The notion that individuals feel a part of a community and want to share their information with that community has empowered society.

Individuals have always utilized data to make decisions that affect their everyday lives. Whether the decisions involve choosing which football game to watch or how much money to tip the waiter, data is used to make decisions. In science, the process of decision making, in short, is no different. Scientists, researchers, engineers etc. all collect and analyze data to benefit society. In the world of water, data is available all around us: water is in the air we breathe, the lakes we swim and the plants we grow. Groups and individuals have been collecting data pertaining to water since the beginning of time: flood levels for the Nile River near Roda can be dated back to 3,000 B.C. (Lyons, 1906). There is no doubt that there is an abundance of water data available on earth, however the task of actually collecting and sharing that information is a rigorous and tedious process. In a study performed by the Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI), it was found that researchers spend a significant part of their time searching, acquiring and preparing data (Maidment, 2005). Approximately 35% of researchers surveyed spend more than 25% of their time acquiring and preparing data while 12% spend more than 50% of their time (Maidment, 2005). The scientific community needs to improve upon these statistics. Data needs to become more transparent and widely shared. Organizations and agencies need to publish their data, make their data searchable and provide the means through which to access it. CUAHSI embarked on this venture in July of 2001 and developed the Hydrologic Information System (HIS). The HIS is a system that provides the infrastructure that facilitates data publication, discovery and accessibility. Although this technology has improved the state of hydrologic information in the scientific community, the same cannot be said for hydrometeorological information. This thesis will begin to address the issues and develop the framework for sharing hydrometeorological information.

1.1 What is Hydrometeorology and Why is it Important?

Hydrometeorology is the science that integrates and studies the physical processes that characterize the movement of water on earth, the movement of water in the atmosphere and the movement of water across the earth and atmospheric boundary layer (Pukh Raj Rakhecha, 2009). In the last few years, hydrometeorology has evolved into a new scientific field which has focused on understanding how atmospheric patterns, such as water vapor, clouds and precipitation, influence the hydrologic cycle: processes such as runoff, evaporation, and snowmelt (see Figure 1.1). In the past, the inherent natural coupling of the hydrology on the ground and science in the atmosphere has not traditionally been emphasized however in hydrometeorology; the coupled system is studied simultaneously.

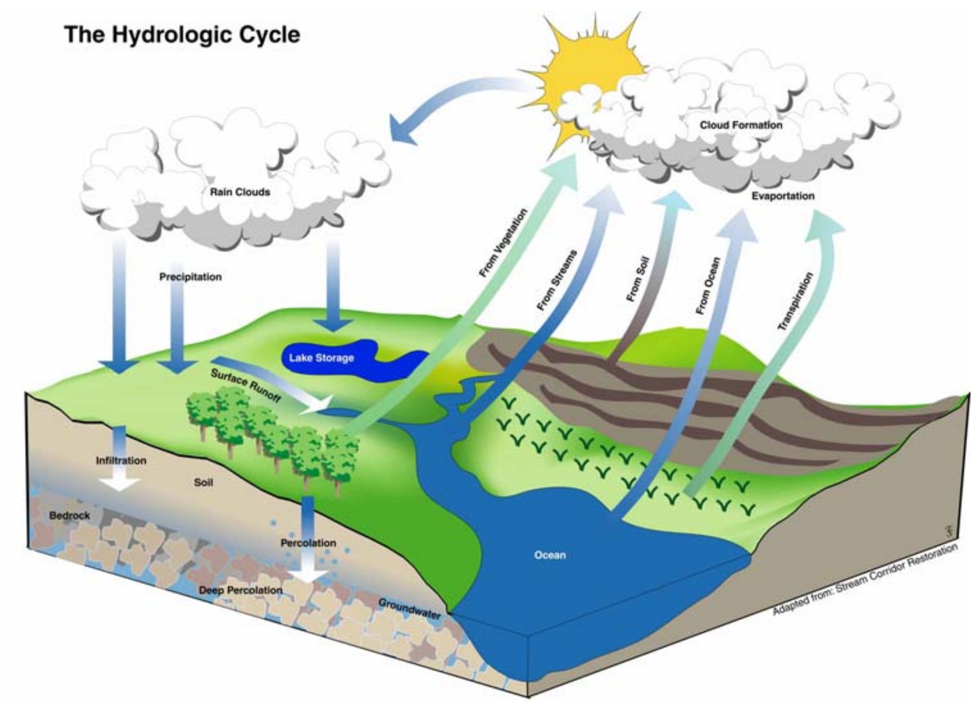


Figure 1.1 Hydrologic Cycle (University, 2010)

One of the important areas in hydrometeorology is the study of how energy is transferred between atmospheric water and the earth's surface. As water changes phases between liquid and gas, energy is absorbed onto the earth and released as heat into the atmosphere. Although at any given time, atmospheric water represents a small fraction of the earth's total water volume, approximately .001 %, a significant amount of water moves through the earth's atmosphere during the water vaporization and precipitation process (Chow et al., 1988). Because water travels through the atmosphere at such a high rate, it is difficult to characterize and model the movement of water both on the surface and in the atmosphere. This makes it difficult to model, forecast and predict critical hydrometeorological phenomena affecting the water resources that communities rely on daily.

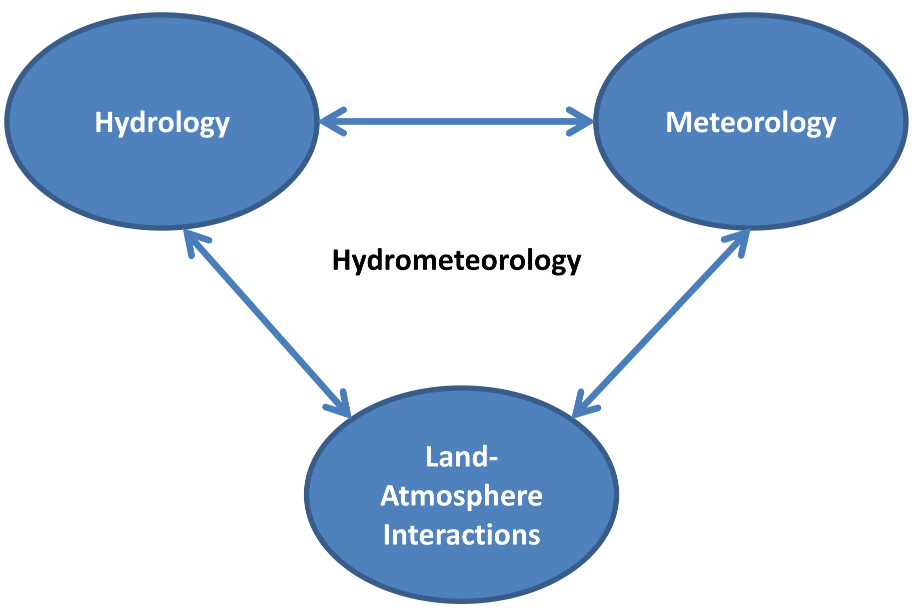


Figure 1.2 Scientific Areas Pertaining to Hydrometeorology

The blending of hydrology and meteorology into a connected field is important for studying storms, floods, droughts and other hydrometeorological events both on a global, national and regional scale. As the attention being paid to climate change escalates, scientists rely more and more on models to describe the behavior and distribution of water on earth. Historically, models have focused either strictly on land processes or atmospheric processes even though it has been understood that feedback mechanisms between the land surface and atmosphere exist. Climate change will play an important role in the policy that shapes the future. Changes in temperature and precipitation, will undoubtedly affect the amount and timing of water that ends up flowing in the world's river networks. To completely and accurately describe the physical processes that shape the water cycle, scientists need to encompass three major areas: hydrology, meteorology and land-atmosphere interactions.

Hydrometeorology has focused on improving the understanding of all water processes in the hydrologic cycle as they interact with one another. There is a clear disconnect between meteorological agencies like the National Weather Service (NWS) and hydrological agencies like the U.S. Geological Survey (USGS). In order to produce forecasts of various hydrometeorological events, such as floods, experts need to have an interdisciplinary understanding of the physics of the water cycle. Hydrometeorology is important and incorporates critical aspects of the hydrologic cycle. With a better understanding of the physics of both the land and atmosphere, scientists will become more fit to model, predict, and forecast events.

1.2 WHAT IS A HYDROMETEOROLOGICAL INFORMATION SYSTEM?

A hydrometeorological information system (HMIS) is a system that aids in the collection, management, distribution and synthesis of hydrometeorological information. An HMIS is not a single program or a set of computers but instead a network and a set of services that allow for efficient communication between data providers and data users.

An HMIS provides the tools necessary to share, discover and access information pertaining to hydrometeorology. With the wealth of information that is collected by groups, individuals and machines on a daily basis, a HMIS helps to extract information locked away in databases, servers and hard drives by maneuvering past the prices, passwords and software needed to access them.

There is no doubt that the world is transforming. It is morphing into a virtual world where social networks dominate the manner in which individuals communicate and share information. Although services like Facebook and YouTube have thrived in this new age of technology, the scientific community has yet to overcome the burden of sharing and collecting information amongst its peers. For centuries, information has been stored in databases managed by groups and individuals all across the globe, however accessing that information has been a tedious process.

An HMIS is a reliable solution to solving this problem. By utilizing an HMIS, data providers are able to quickly store their information on servers and share their data with the rest of the world through the use of web services. On one side, data providers can publish their data quickly and in a standardized format which allows for querying of their data, while on the other side data users can easily search catalogs full of metadata and discover and access new information in which they might be interested. An HMIS strives to bridge the gap between data providers and data users. Hydrometeorological information is becoming increasingly important with the trends of climate change. Without a system in place to facilitate the sharing of information, scientists, researchers and engineers will be hindered in their quest to make useful decisions to benefit society.

1.3 WHERE DO WE FIND HYDROMETEOROLOGICAL INFORMATION?

A plethora of agencies and organizations collect, store, publish and share hydrometeorological information. Groups like the National Oceanic and Atmospheric Administration (NOAA), National Weather Service (NWS), National Climatic Data Center (NCDC), and University Corporation for Atmospheric Research (UCAR), all archive enormous amounts of hydrometeorological data and make it available through the internet. Unlike hydrologic information, which typically consists of observational time series at specific locations, hydrometeorological data is available in many different forms. Because the processes involved in hydrometeorology are on a continuous spatial domain, hydrometeorological data is represented in a grid or a set of grids. Furthermore, hydrometeorological data can represent many different types of information. Some of it represents data collected by radar and satellite (e.g. Next Generation Radar, or NEXRAD), while other data are computed by models (e.g. North American Regional Reanalysis, or NARR).

The internet has simplified the process for discovering and visualizing hydrometeorological information, however the process is still in need of a well-defined infrastructure. Currently, many of the groups that publish hydrometeorological information have disparate methods for storing and sharing data; some groups use ftp sites and catalogs while others use servers and web services. Regardless of the method, there is a lack of standardization for accessing data across different data providers.

The Unidata project, developed within UCAR, has been working on standardizing the manner in which hydrometeorological information is openly shared across the web. Unidata has worked to integrate new and valuable data sources into their system and facilitate data discovery through the use of digital libraries. Two tools that Unidata has developed are the Thematic Realtime Environmental Data Distributed Data Services

(THREDDS) server and Network Common Data Form (NetCDF). THREDDS is a software server for cataloging, browsing and accessing remote data and metadata. NetCDF on the other hand is a data structure for storing grids in an array-oriented structure than can be ingested into other analytical and visualization applications.

1.4 Objectives

It is becoming apparent that our world is on the cusp of entering a new era. An era so profound, it mimics the one sparked by the Industrial Revolution. With the creation of the internet, the world entered into the "Data Revolution" where oceans of information could be stored and shared in the "cloud". Recognizing this transition, the U.S. government has named its first Chief Information Officer, and created a website called data.gov to share with the public countless amounts of public data and records in hopes of creating new ideas and more transparent government processes. One important concept to remember is that although data is available everywhere, it cannot be useful unless it is provided with some context or metadata. Data represented with relationships connects more ideas. Either way, one notion is quickly becoming fact; data will rule the world.

This thesis begins to answer some of the questions surrounding the aggregation, standardization, and distribution of hydrometeorological data, specifically data which varies in both space and time. The following questions are addressed:

- What type of performance standards can we expect from WaterOneFlow web services?
- Can we thematically organize, manage and distribute data available through CUAHSI web services?

- How can we organize, manage and distribute data which varies in both space and time?
- What type of utility can we gain from representing spatially and temporally varying data as events or entities?
- Using the "event" philosophy for sharing hydrometeorological information, can we characterize phenomena over spatial and temporal domains?

The structure of this thesis follows the outline provided by the questions presented above. In Chapter 2, a review of some of the current and latest technology available pertaining to Hydrometeorological Information Systems is discussed and presented. In Chapter 3, performance standards for sharing data using the latest technology (i.e. web services) are addressed followed by an organization scheme for managing observations data. Chapter 4 focuses on implementing some of the technology discussed in Chapter 2 as well as the efficient data management scheme presented in Chapter 3. Finally, Chapter 5 reviews the contents of the work that has been accomplished; provides an assessment of the current state of Hydrometeorological Information Systems and presents a list of recommendations and future goals.

CHAPTER 2: LITERATURE REVIEW

2.1 CUAHSI-HIS

The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUASHI) was developed with the support of the Earth Sciences Division of the National Science Foundation (NSF) to build infrastructure and services to advance hydrologic science in the United States (Maidment, 2008). Through the cooperation of 122 universities (see Figure 2.1), the project aims to enhance the appreciation and understanding of hydrologic science as well as increase the utility of hydrologic information through vehicles such as academics and technology (CUAHSI, 2010). One component of CUAHSI is the Hydrologic Information System (HIS) project which was created to improve the accessibility to hydrologic information across the United States. In order to reach this goal, CUAHSI-HIS has been developing a set of tools, services, standards and procedures through which “large volumes of high quality hydrologic data” can be obtained efficiently in a structured and organized manner. (Maidment, 2008).

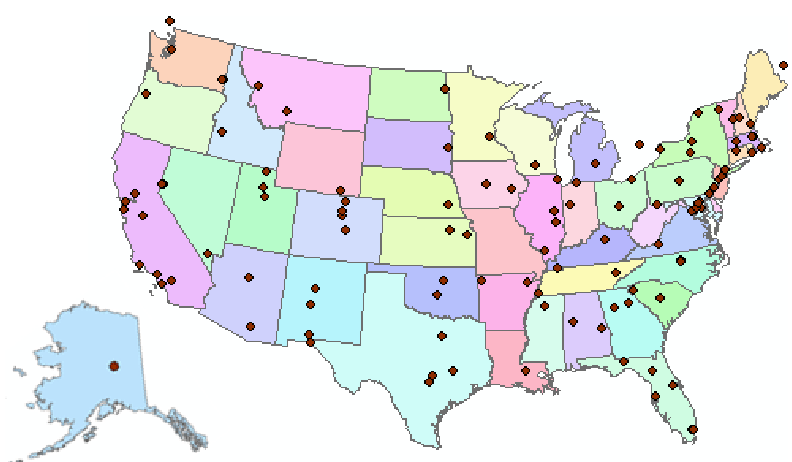


Figure 2.1 Locations of 122 Universities in Cooperation with CUAHSI

Many of today's advancements in science, specifically hydrologic science, rely on assimilating and synthesizing copious amounts of data. In the academic and research community, it is well accepted that vast quantities of time are spent searching for, collecting and preparing data for more insightful analysis (Maidment, 2008). For decades now, data users have yearned for simple methods to discover and download ample amounts of high quality data. The CUAHSI-HIS project has shown that an effective approach to solving this problem is through the perspective of connecting data providers using web services. By developing a collaborative framework through which numerous groups and agencies can archive and publish data via the tools available on the web, data users simultaneously gain the ability to collect data efficiently in a format that can be readily manipulated (Maidment, 2008). The framework developed by the CUAHSI-HIS project highlights the following four initiatives:

Data Storage in an *Observations Data Model* (ODM)

Data Delivery through internet-based *Water Data Web Services*

Data Discovery through a *National Water Metadata Catalog* AND

Data Access and Analysis through a *Client Application (*HydroDesktop)

With the implementation of the following components, CUAHSI-HIS provides a medium that effectively enhances the user experience from both the data provider's and data user's perspective with minimal costs. CUAHSI-HIS has been working to make hydrologic information universally available in a timesaving and functional format that allows for visualization, synthesis and evaluation (Whiteaker et al, 2008).

2.1.1 Hydrologic Information Systems

Hydrologic Information Systems (HIS) can be defined as a collection of components which work together to store, index, access and distribute hydrologic information (Maidment, 2009). The system contains servers, catalogs, applications etc. which communicate with one another through a common standardized language. In the case of the CUAHSI-HIS project, the individual components utilize water web services to relay information from one component to the other. CUAHSI-HIS has developed a specific set of web services coined WaterOneFlow web services which provide a systematic protocol through which hydrologic information can be shared and transferred. WaterOneFlow web services primarily serve two functions: to provide metadata services and data services (Maidment, 2009). In conjunction, WaterOneFlow web services and HIS support the infrastructure needed to share hydrologic information in a useful, efficient and complete manner.

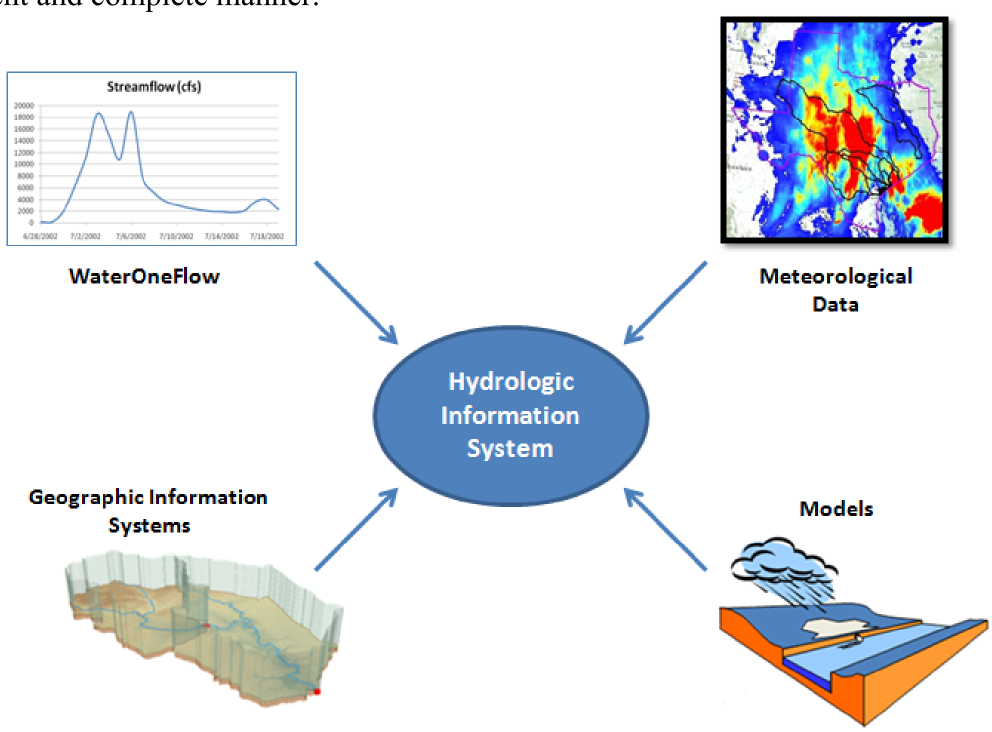


Figure 2.2 Hydrologic Information Systems

The particular feature of HIS is in the extended functionality it provides for working with hydrologic data which can be dynamic in space (groundwater), time (streamflow) or both, space and time (precipitation). See Figure 2.2 above. Furthermore, the application of HIS at a variety of different spatial scales provides users with the facility of managing only the hydrologic information of interest without the hassle of data mining and/or data filtering. See Figure 2.3 below.

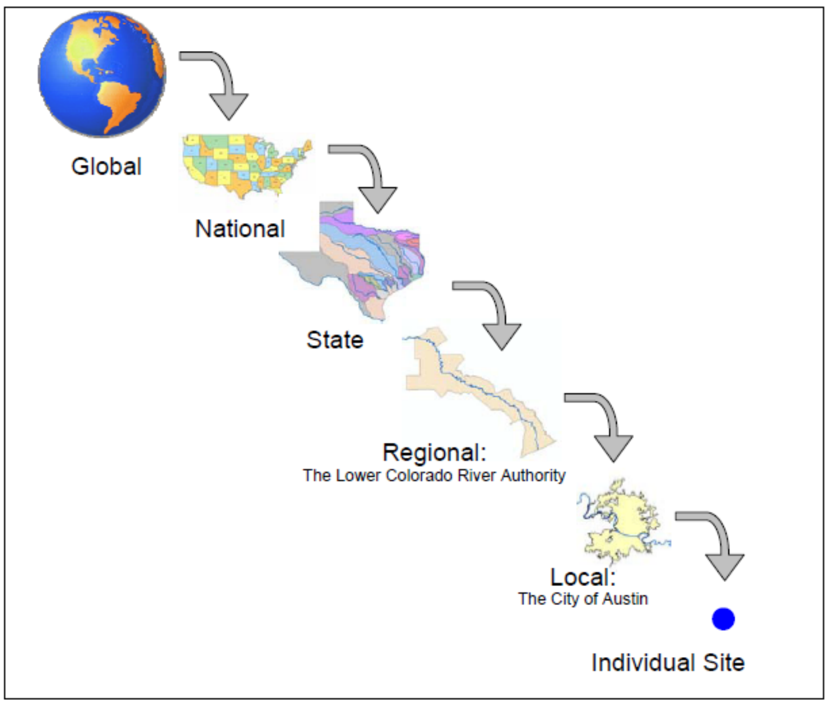


Figure 2.3 Different Scales for a HIS (Jantzen, 2007)

Until now, Geographic Information Systems (GIS) have been the leading analytical tool for working with geospatial data, a subset of hydrologic data (Maidment, 2005). However, for the better part of its existence, hydrology has traditionally been a temporal science. Synthesis, analysis and evaluation of hydrologic information have primarily been based on records which inherently change over time, time series. Time series can be analyzed at one site or at a collection of sites. With the development of HIS, hydrologists are now able to integrate both GIS and HIS to manage and organize temporal and spatial information. Although this collaborative effort has helped hydrologists tackle many of the complications revolving around the data collection process, many issues dealing with the visualization of and integration of analytical tools for hydrologic information, remain unresolved.

2.1.2 Services Oriented Architecture and the HIS Model

As part of the discussion of HIS, the need for an organized system of connected components was identified as a key technology to providing a successful infrastructure to support the sharing of hydrologic information. CUAHSI-HIS has designed a Services Oriented Architecture (SOA), which establishes the fundamental relationships between the components of HIS and allows for communication among them. Similar to the architecture defined for the internet and accessing web pages where users search Google to retrieve information and display it in their browser, hydrologic information users can seek out data in catalogs (HIS Central), retrieve data from servers (HIS Servers) and ingest data into applications and software suitable for hydrologic analysis (HydroDesktop) (see Figure 2.4). As mentioned in section 2.1.1, it is the WaterOneFlow web services which allow for communication between each individual component of HIS.

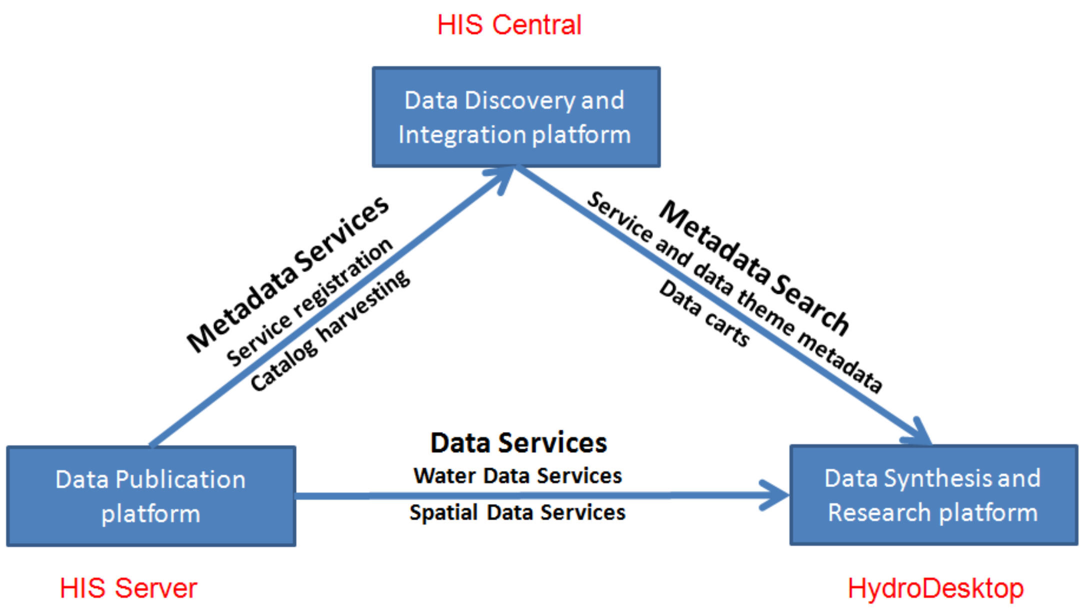


Figure 2.4 Services Oriented Architecture for a Hydrologic Information System (Maidment, et al., 2009)

The individual components of HIS each serve an important role in the data fetching process. HIS servers function as the principle locations for storing large volumes of hydrologic information, specifically time series. Within the server itself, data is managed in a database which can be accessed by users remotely through various web services. CUAHSI publishes information housed on HIS servers as WaterOneFlow web services. Because servers are typically managed locally, data can be appended or modified at any point in time without involving the user. This system provides an efficient structure for sharing data. In the case where data is sensitive to public access, servers can protect their information and limit the accessibility to them. In HIS, HIS servers are the primary repositories for hydrologic information (Maidment, 2009).

Another component of HIS is HIS Central or the hydrologic metadata catalog. This component facilitates the discovery of hydrologic information across data which has already been published on HIS Servers. HIS Central provides an interface where users can search all registered HIS servers with keywords and metadata which describe the hydrologic data of interest. HIS Central is like a Google for hydrologic information. Data publishers can register their data on HIS Central and provide brief descriptions of the datasets they want to share. In addition, data publishers can provide metadata and keywords that relate to the data being shared. This is an important aspect of HIS because it allows for data to be organized and discovered in an efficient and methodical process.

Finally, HydroDesktop is the component of HIS which allows for the harvesting of hydrologic information at the locality of one's own computer or analytical system. HydroDesktop is a platform located on the user's machine and communicates with both HIS Servers and HIS Central. Users can directly download hydrologic information from HIS Servers if they already know of their existence or can search HIS Central for data that they might not know about. Once the data of interest has been discovered, users can download the information onto their local databases; this is called data harvesting. With the information readily and locally available, users can take data they have harvested and combine it with other data already available on their machine and use it to perform insightful analysis and/or modeling. HydroDesktop is intended to synthesize hydrologic information in an environment that supports both time series and geographic visualization (Maidment, 2009). With this unique structure, HydroDesktop provides a unique method for users to efficiently manage and work with hydrologic information.

2.1.3 Observations Data Model

Given the model presented above, CUAHSI has developed the infrastructure and technology to support the storing, discovery and sharing of hydrologic information efficiently via web services and the internet. Because HIS have a component-based infrastructure, communication among the components is of utmost importance. CUAHSI has succeeded in providing an efficient method for communication through WaterOneFlow web services. From the data provider's perspective, WaterOneFlow services aid in sharing data through the standardization and representation of water information. Hydrologic information is based in large part on data measured at multiple point locations: streamflow, precipitation, evaporation etc. In order to support the storage of this data, CUAHSI has established an Observations Data Model (ODM) for sharing hydrologic information on HIS Servers.

Simply put, the ODM is a relational database which facilitates the organization and storage of time series recorded at point locations. Its fundamental design relies on a collection of tables which can be linked together through key data fields. These links are defined as relationships and underline the conceptual principles used to represent observations data. The ODM is very practical because it provides sufficient data fields for describing in complete detail all the information necessary for working with observations data. Information about the data itself, the type of data, how the data was collected, the quality of the data and more can all be entered into the ODM for complete and accurate representation for the data user to see and use. Because data values themselves do not provide information about their heritage, the ODM is a vital element of HIS and the data sharing experience. The ODM is essential for ensuring the quality of shared hydrologic information. Figure 2.5 below shows the complete schema for the ODM.

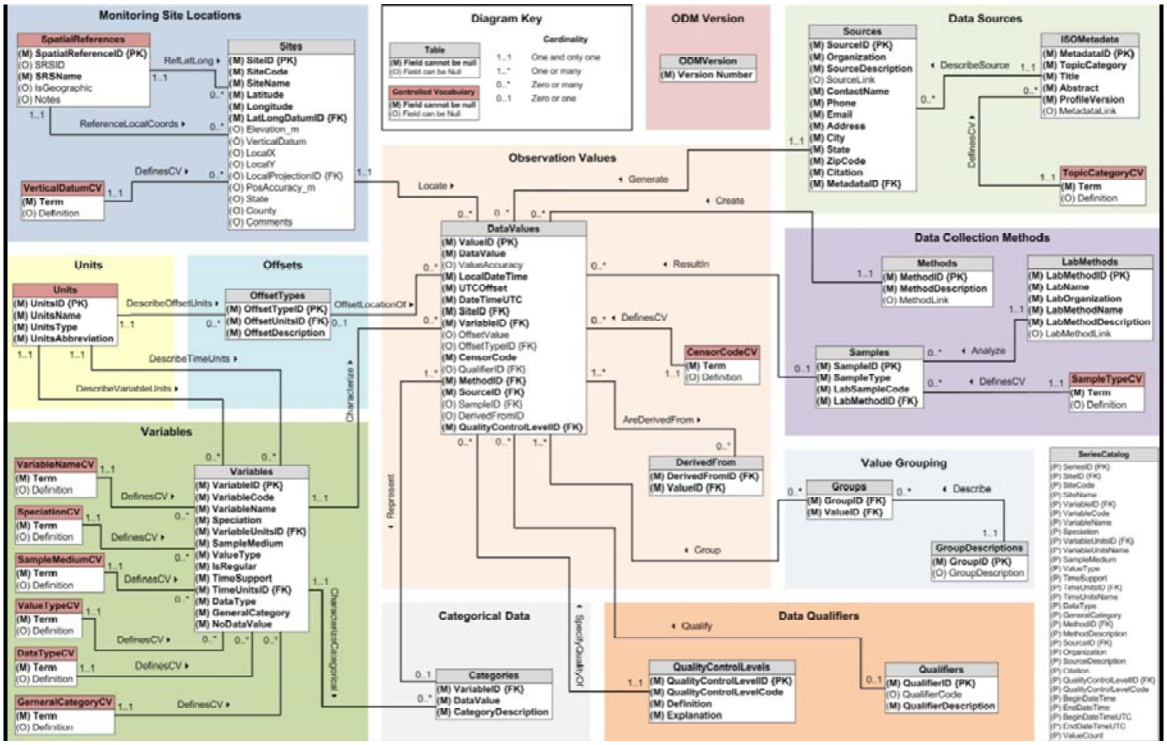


Figure 2.5 Observations Data Model

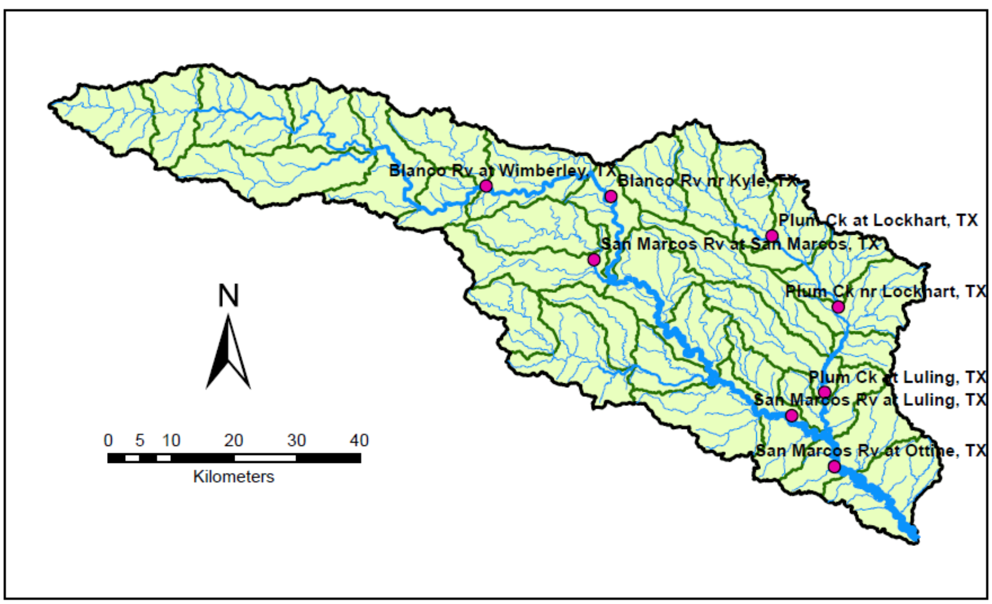


Figure 2.6 Network of Observations Data

HIS can be used to build a network of observations data as depicted in Figure 2.6. Observation sites can record values for a variety of different variables over any selected period of time. This foundation has led to the development of three primary tables that are used to represent hydrologic information through the use of the ODM: Sites Table, Variables Table, and Time Series Table. The Sites Table stores information related to the location of all the sites in the network while the Variables Table stores information related to all the variables being recorded throughout the whole observations network. The Time Series table acts as the repository for all the time series values available in the observations network. Every time series value has a code which is linked to the Variables Table and the Sites Table to represent the value. This inherent model is represented as the "Data Cube" model for representing observations data. The "Data Cube" model can be seen below in Figure 2.7.

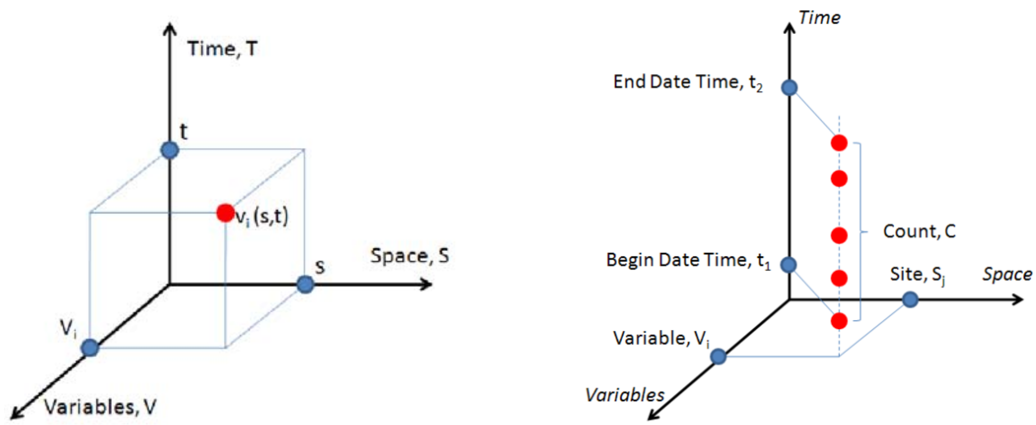


Figure 2.7 Data Cube Representation for Observations Data and Time Series

It is important when storing data that metadata be collected as well to facilitate the management of hydrologic information. With the structure that the ODM provides, a Series Catalog Table can also be created. The Series Catalog is a collection of all the Time Series available in a given network. Each row in the Series Catalog Table denotes a single time series and stores the geographic location and variable, among other descriptors, being represented by that set of values. Enhancing the utility of the Series

Catalog table, each row also stores the Uniform Resource Identifier (URI) of the web service storing the specified data. The Series Catalog is essentially a metadata catalog for the observations network and is beneficial for the discovery of data within a given network.

2.2 WaterOneFlow Web Services

A web service is a single web application programming interface (API) which users can access over a network to perform pre-defined functions on a remote system or server. In the case of CUAHSI, web services have provided the vehicle through which users can discover and download hydrologic information. When a data provider offers access to an HIS Server, a WaterOneFlow web service is created. This web service aids the user in seeking out and collecting data of interest without the complexities of web surfing and data mining. Through web services, users utilize query functions that access the data provider's database and retrieve desired information. Figure 2.8 below demonstrates the architecture defined for web services and the interaction between the data user or service broker and the data provider or service provider.

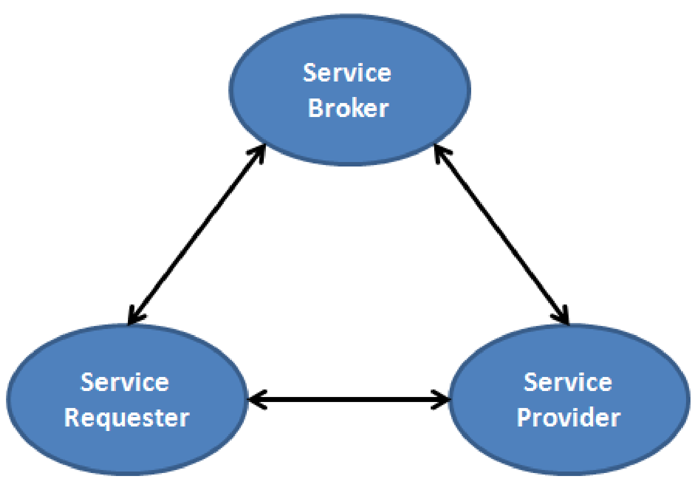


Figure 2.8 Web Services Architecture

An important aspect of the web services architecture is the Web Services Descriptive Language (WSDL) which allows for machine to machine interaction over a given network. A WSDL is defined at the time a web service is created and provides the URI or location of the web service. When utilizing a WSDL, a user can call a function described in the web service and connect to the underlying database holding the sought out information. In HIS, functions for WaterOneFlow services query the database of HIS Servers by either requesting site, variable, or value information. Below is a list of the functionality provided by WaterOneFlow web services:

GetSites: Returns the list of observation sites in a given network.

GetSiteInfo: Returns predefined site information for an observation site including the list of variables measured, period of record and number of values.

GetVariableInfo: Returns details about variables measured in a given network such as their description, name and units of measurement.

GetValues: Returns the list of recorded values at an observation site for a given variable.

2.2.1 SOAP and REST Services

CUAHSI web services typically utilize two common forms of communication to exchange information in a standardized format: Simple Object Access Protocol (SOAP) and Representational State Transfer (REST). These methods of communication utilize a variety of languages, such as eXtensible Markup Language (XML), as the message format and Hypertext Transfer Protocol (HTTP) as the message transmission method. In the past, SOAP was the more common method of communication for web services however because of its complexity and inefficiency in relaying information, REST has become the preferred option. Unlike SOAP which adds an extra layer between the service requester and the service provider, REST conforms to a variety of different message formats by supplying service providers with architecture for relaying information without adding additional transmission layers; SOAP only utilizes XML as the protocol for transferring information (Al-Zoubi and Wainer, 2009).

The World Wide Web Consortium (W3C) has defined the above mentioned sets of standards for accessing, relaying and visualizing information via the internet in an efficient manner. CUAHSI has taken the initiative to utilize these standards as means of sharing hydrologic information with the world. The utilization of the internet and its enormous storage and transmission capabilities has made CUAHSI successful in implementing its vision for HIS.

2.2.2 WaterML

Hydrologic information can be provided from many different organizations, agencies, groups, universities etc. and typically, providers already have their own method in which to store and transfer information. Through the use of an ODM and WaterOneFlow web services, hydrologic data is standardized in an optimal format that allows data to be exchanged directly and efficiently. When data is requested through a WaterOneFlow service, the information is always requested and returned in the same format whether the data are provided by the USGS, the EPA or a university source. Standardization of data transfer across multiple data providers is very beneficial because it reduces the time allocated to searching for data and transforming data into a format which is suitable for the user. WaterOneFlow web services use a language developed by CUAHSI called WaterML to communicate between a server and client (Whiteaker, 2008). WaterML synthesizes the common data and metadata elements across different data providers and transfers the information in a consistent format without altering the raw information stored in the database (Whiteaker, 2008).

2.2.3 Open Geospatial Consortium

Similar to CUAHSI and the standards developed for sharing observations data, the Open Geospatial Consortium (OGC) is a collective group of voluntary government, nonprofit and research organizations which define standards for geospatial data and services. With the inherent spatial representation of observation data networks, the OGC provides a standardized method through which hydrologic information can be shared and visualized through the internet. As an example, GIS features representing observation locations can be published and made available for users to access readily. Using a customization of XML, the OGC has developed a standardized language called Geographic Markup Language (GML), which enables geospatial content to be transferred in a consistent format. CUAHSI has used several of the services defined by the OGC and is currently sharing geospatial data as part of the CUAHSI-HIS project.

CUAHSI uses three main services developed by the OGC: Web Mapping Services (WMS), Web Feature Services (WFS) and Web Coverage Services (WCS). WMS are used to transfer map images that are available on a map server using information from a GIS database. Similarly, WFS are used to transfer GIS features while WCS are used to transfer GIS coverages or rasters. With these standardized services and protocols in place, CUAHSI has been able to extend its capabilities by sharing spatial information in conjunction with temporal information. This technology has now given hydrologists the ability to construct new spatial-temporal relationships from information provided by geospatial information and observational time series.

2.3 Hydrometeorological Information

As was introduced in chapter 1, the science of hydrometeorology is concerned with the movement of water on the surface, in the lower atmosphere and across the surface-atmospheric boundary. During the 20^{th} century, observations data were the primary means through which researchers collected and analyzed hydrometeorological information. Although this has provided the global community with significant advancements in hydrometeorologic science, a need for higher quality synthesis of physical processes that vary both spatially and temporally is of utmost importance if methods in analyzing, forecasting and modeling hydrometeorological systems are to improve.

Gridded data has become the leading vehicle through which hydrometeorological information is shared. Because of the spatiotemporal variability of hydrometeorological phenomena, gridded data provide researchers with the ability to analyze a higher quantity and quality of data over small spatial and temporal scales. One of the relatively recent developments in gridded technology has been satellite and radar imagery. Although efforts have been made to readily distribute satellite and radar information, improvements in data collection and data management are still a primary concern for researchers as we move further into the 21^{st} century.

In addition to satellite and radar imagery, global and regional hydrometeorological models have also been developed and made available in gridded format. These models combine volumes of historical observations data and physical models to reproduce products that represent numerous hydrometeorological variables across various regions of space. Like the information available through satellite and radar technology, the improvement of distributing and managing information available through hydrometeorological models is crucial.

There are a large number of government agencies and research communities which have focused on the collection, management and distribution of gridded hydrometeorological data or coverages however; many of the efforts have proceeded independently from one another without a collaborative standardization system through which to share data. Table 2.1 provides a list of a few of the large agencies and groups which have contributed significant amounts of hydrometeorological information to the research and public communities. Up until now, providers have either been sharing data through repositories linked to their websites (i.e. File Transfer Protocol or FTP sites) or through remote servers and catalogs (i.e. THREDDS server). Data has been made available in many different forms ranging from GIS features to raster images however with the large amount of hydrometeorological variables available and plethora of data providers, it has become a tedious and arduous task to discover and download hydrometeorological information efficiently.

Table 2.1 Hydrometeorological Information Data Providers

Agency	**URL**
National Oceanic and Atmospheric Administration	http://www.noaa.gov/
National Climatic Data Center	http://www.ncdc.noaa.gov/oa/ncdc.html
National Weather Service	http://www.weather.gov/
National Center for Atmospheric Research	http://www.ncar.ucar.edu/
University Corporation for Atmospheric Research	http://www2.ucar.edu/
Earth Systems Research Laboratory	http://www.esrl.noaa.gov/
National Severe Storm Laboratory	http://www.nssl.noaa.gov/
National Resources Conservation Service	http://www.wcc.nrcs.usda.gov/snow/
National Aeronautics and Space Administration	http://gcmd.nasa.gov/

The groups listed above have largely been involved in providing and sharing a wide array of hydrometeorological information. Variables such as precipitation, evaporation, runoff and infiltration are some of the key processes which are of interest to hydrometeorological information clients and users. Because many of these physical processes are related to the incoming solar radiation, energy variables such as sensible heat and latent heat flux are equally as important and of particular interest to hydrometeorologists.

Precipitation has become one of the most sought after pieces of information in the 21st century both because of its relevance to water resources and water management but also because of its importance to the common individual's day to day lifestyle. There are multiple datasets which provide precipitation data some of which are observation based (i.e. NEXRAD) and some of which are remote sensing based. In addition, models like the North American Regional Reanalysis (NARR) and the North American Mesoscale (NAM) have been developed to provide a historical archive of hydrometeorological information. Although all these datasets supply data at different spatial and temporal scales, it is clear that gridded data has become the common format in which to synthesize hydrometeorological information. Table 2.2 below lists some of the prominent hydrometeorological datasets currently available.

Table 2.2 Hydrometeorological Information Datasets

Dataset	**URL**
CFSR	http://nomads.ncdc.noaa.gov/thredds/catalog/cfsrr.html
NARR	http://nomads.ncdc.noaa.gov/thredds/catalog/narr/catalog.html
NAM	http://nomads.ncdc.noaa.gov/thredds/catalog/nam/catalog.html
GFS	http://nomads.ncdc.noaa.gov/thredds/catalog/gfs4/catalog.html
RUC	http://nomads.ncdc.noaa.gov/thredds/catalog/ruc/catalog.html
NEXRAD	http://www.ncdc.noaa.gov/thredds/catalog/radar/StIV/catalog.html

2.3.1 Next Generation Radar

Next Generation Radar (NEXRAD) is a network of 159 doppler radar towers, see Figure 2.9 below, that are operated by the National Weather Service (NWS) and measure both precipitation and atmospheric movement or wind. NEXRAD towers are located throughout the United States and data from their sites is synthesized to produce spatially continuous maps that represent the movement of moisture in the atmosphere. NEXRAD utilizes electromagnetic waves to bounce signals off air and water parcels. The reflectivity, measured in dBz, is then recorded and used to calculate precipitation rates. NEXRAD towers have the ability to record reflectivity at a variety of different elevations and angles thus are able to depict a thorough representation of the atmospheric column at any point in the landscape.

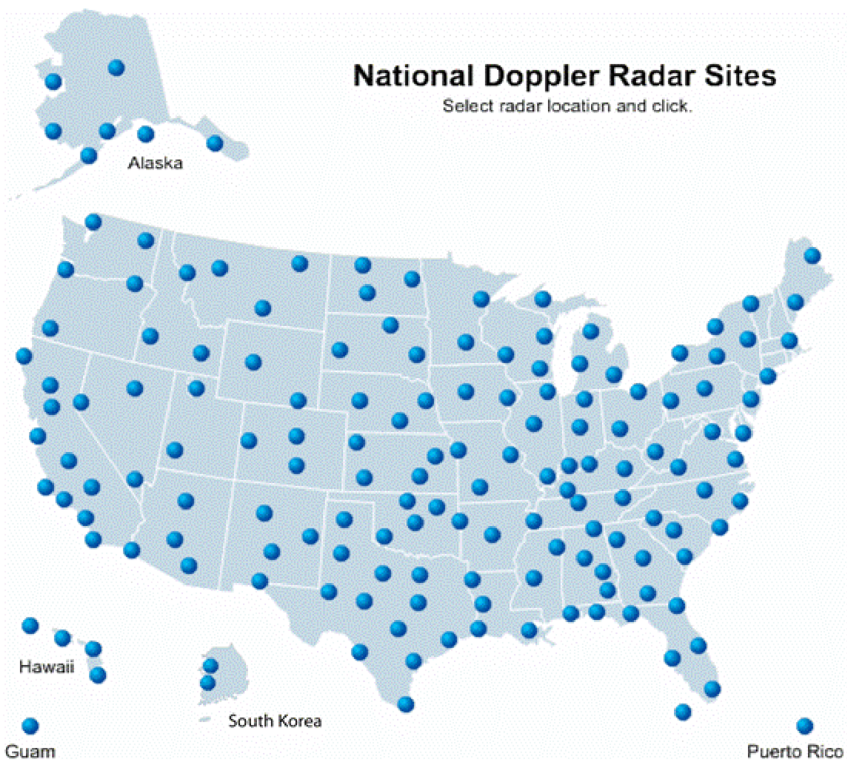


Figure 2.9 Location of NEXRAD Towers (NOAA, 2010)

NEXRAD products utilize the Z-R relationship to determine how much precipitation is actually present in the atmosphere. Because the reflectivity is a function of the size of the water particle, different relationships are used to determine precipitation rates. The basic Z-R relationship is of the form:

$$Z = AR^B$$

where Z is the reflectivity measured in dBz and R is the precipitation rate measured in millimeters per hour. There have been many empirical studies performed studying the relationship between reflectivity and precipitation rate however only a few have been widely accepted. The equations seen in Table 2.3 below are the relationships recommended by NOAA for different storm types.

Table 2.3 Z-R Relationships for NEXRAD (NOAA, 2010)

Relationship	**Optimum for:**
Marshall-Palmer ($Z = 200R^{1.6}$)	General Stratiform Precipitation
East-Cool Stratiform ($Z = 130R^{2.0}$)	Winter Stratiform Precipitation – East of Continental Divide
West-Cool Stratiform ($Z = 75R^{2.0}$)	Winter Stratiform Precipitation – West of Continental Divide
WSR-88D Convective ($Z = 300R^{1.4}$)	Summer Deep Convection
Rosenfeld Tropical ($Z = 250R^{1.2}$)	Tropical Convective Systems

NEXRAD has gone through much scrutiny because of its deviation from observed rain gauge measurements. In addition, NEXRAD has had difficulty in differentiating precipitation from snow and hail. Because of the nature of the technology, the NWS has developed numerous post-processing schemes to improve the accuracy of NEXRAD measurements. Currently, stage IV NEXRAD products are available through various sources. Stage IV NEXRAD data combines radar measurements, satellite measurements and rain gauge measurements to methodically stitch together a continuous high quality precipitation product over the majority of the contiguous United States. Figure 2.10 below demonstrates and example of a NEXRAD product visualized in ArcGIS.

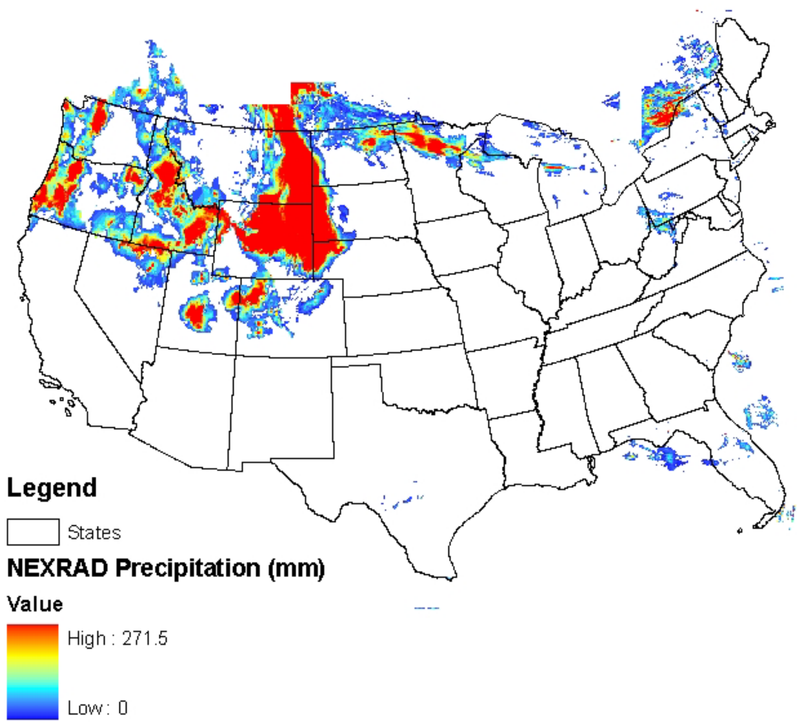


Figure 2.10 NEXRAD Data Visualized in ArcGIS

From the direct reflectivity measurement to the final post-processed product for NEXRAD data, it is inevitable that errors for estimating actual precipitation values will occur. As was mentioned above, radar reflectivity is not very sensitive to the type of precipitation occurring in the atmosphere. Because of this inherent error in the measurement, NEXRAD data has been shown on occasion to overestimate precipitation by a factor of 2 and underestimate precipitation by a factor of 10 (NWS, 2002). Improvements to NEXRAD products continually are being made to reduce the error between recorded measurements and observed values. At present, NEXRAD is the only measured continuous spatially distributed precipitation database in the United States.

2.3.2 North American Regional Reanalysis

The North American Regional Reanalysis (NARR) is a data assimilation model developed by the National Centers for Environmental Prediction (NCEP) in conjunction with other research entities. NARR is a compilation of a variety of observational datasets and land surface models that span the entire North American region at a resolution of 32 km^2. Input data for NARR is gathered from multiple sources including the preceded NCEP/NCAR Global Reanalysis project. The NARR provides data for a variety of hydrometeorological variables at three hour intervals from 1979 to the present day. Until recently, NARR has been considered the best archive for hydrometeorological data for the entire North American region (see Figure 2.11). At the beginning of 2010, NCEP publically released a new dataset which supersedes the coverage and quality of the NARR project. The NCEP Climate Forecast System Reanalysis (CFSR) is a global data assimilation model and is published at a scale of 38 km^2. CFSR includes all conventional satellite data in its model, and similar to NARR, provides records dating back to 1979. CFSR provides information every 6 hours whereas NARR provides data every 3 hours. Both models have been utilized significantly in scientific research over the last decade.

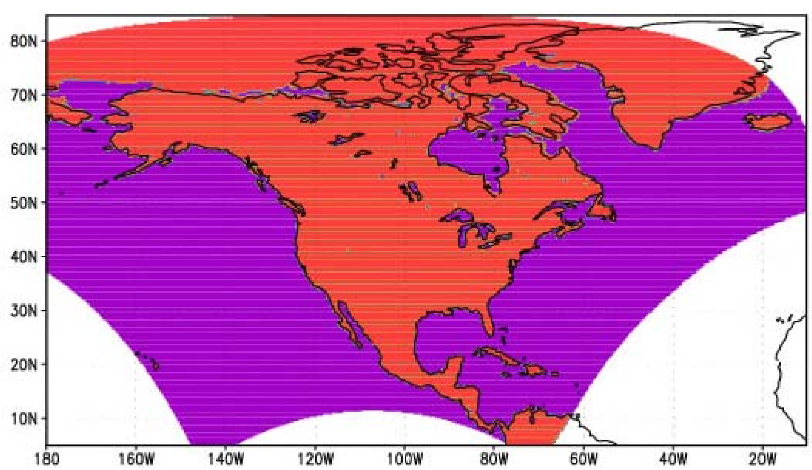


Figure 2.11 Geographic Extent for NARR (Environment Canada, 2010)

2.4 Sharing Hydrometeorological Information

As indicated above, gridded data has become the preferred method for sharing hydrometeorological information among the scientific community. One of the most recognized and accepted tools for sharing gridded data is the Network Common Data Form (NetCDF) model developed by the Unidata program at the University Corporation for Atmospheric Research (UCAR). NetCDF is a model that incorporates a set of interfaces, libraries and standardized formats that support the creation, access and sharing of gridded scientific data (Domenico, 2009). Studies have shown that the array-oriented structure of NetCDF files have provided the most efficient form of storing and retrieving gridded time series (Doraiswamy et al., 1999) In addition, NetCDF provides the functionality to access multidimensional data at a point, along a cross section and over an area. Moreover, netCDF allows data to be visualized using GIS software which has become a leading technological and analytical platform through which hydrometeorological studies are performed. The efficient structure of NetCDF has become very desirable because it allows for access to small subsets of large data sets which in hydrometeorology can be very beneficial.

2.5 Space and Time

Many of the processes involved in hydrologic and hydrometeorologic systems are variable in both space and time. Dynamic processes such as precipitation, runoff and evaporation all change on varying spatial and temporal scales. The science of hydrology has been focused on studying how phenomena change over various scales of time. Statistical methods for analyzing time series have in large part dominated the direction of hydrologic analysis. Because of this focus, hydrologists have predominantly been interested in collecting observational data at various locations in space. Although this methodology has provided numerous insightful analyses, it cannot be overlooked that these data are very simplified representations of spatiotemporal processes.

It is not yet fully understood how to integrate the analysis of physical processes that are variable over a spatiotemporal domain. Analytical software has for the most part been difficult in providing a structural support system or framework for working with data of this nature. GIS is a critical component and has provided a framework for working with geospatial data however this has yet to be fully extended to a spatiotemporal scale. In recent years, GIS has broadened its analytical capabilities to the temporal regime with the implementation of the ArcHydro data model developed by David Maidment at the University of Texas-Austin and others. ArcHydro has been an integral part in integrating time series support into a traditionally static analytical toolset.

Recent research encompassing the ArcHydro data model has focused on designing a framework for working with spatiotemporal data types using GIS. Depending on the type of analysis that is required, researchers can utilize the standard data types to manage, visualize and synthesize spatiotemporal information (Goodall et al., 2004). The four panel diagram below, see Figure 2.12, is a representation of how data in space and time can be represented in ArcHydro.

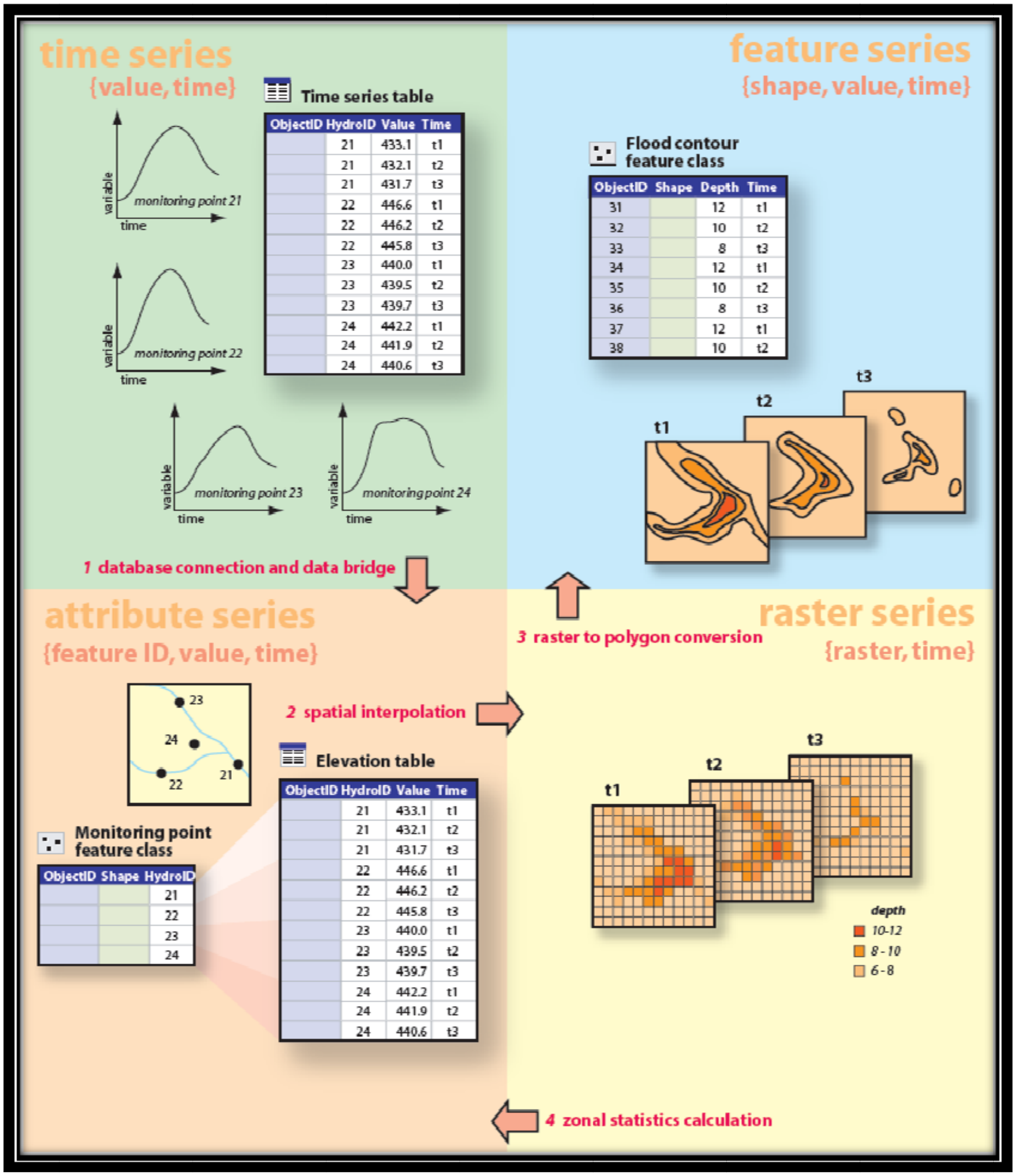


Figure 2.12 Four Panel Diagram for Spatiotemporal Data (ESRI, 2009)

Time Series: A sequence of {value, time} pairs.

Attribute Series: A sequence of {feature ID, value, time} triples indexing time series recorded at a set of spatial features with fixed locations and shapes.

Raster Series: Raster datasets of a spatially-continuous phenomena indexed by time.

Feature Series: A sequence of {shape, value, time} triples for vector description of values where shapes vary in space and time.

2.5.1 Precipitation in Space and Time

Precipitation has always been one of the most desirable and important sets of observational data in both hydrology and hydrometeorology. Precipitation data provides an input to many hydrometeorologic models and a historic record for flood and drought frequency analysis. Because precipitation is such an important factor in forecasting hydrometeorological events, having access to high quality precipitation data is of utmost importance to hydrology.

Precipitation events vary in both space and time at very fine scales. Understanding the spatial and temporal variability of precipitation is important for providing accurate and meaningful characterizations and predictions for the distribution of the water over a variety of different spatial and temporal scales (see Figure 2.13). For example, it is not sufficient to say it rained 2 inches yesterday because that precipitation event occurred over an area with a specified orientation and over a specified amount of time with a certain distribution. Water planners and managers use precipitation data to get an idea of how to regulate, operate and distribute water to communities. With the increasing concern for climate change, understanding the variability of precipitation on a spatial-temporal scale has become even more crucial.

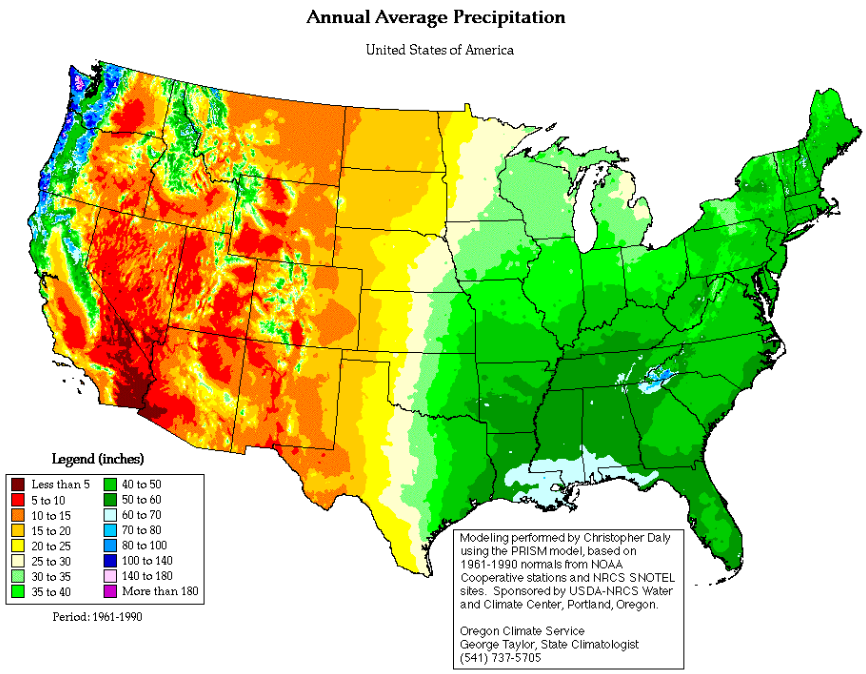


Figure 2.13 Spatial and Temporal Variability of Annual Precipitation over the United States (PRISM, 2010)

2.5.2 Storm Events

Precipitation patterns along scales of space and time can be described as a collection of contiguous objects or storm events which have associated characteristics or attributes. Storms have discrete start times, end times, durations, areas, intensities, and frequencies. Although storms may originate on a local scale, storms have the potential to travel across long distances and affect large areas at the regional, national and global scale. Storms move in conjunction with the large air masses or weather systems they are generated within. Consequently, storm events occur in different forms and thus vary in physical dynamics (see Figure 2.14).

There are three common types of storm events: convective, orographic and frontal or cyclonic. Distributed along each of these types, the type of precipitation that occurs within a storm event can also vary in form. Types of precipitation include mist, rain, sleet, snow, and hail. The physics involved in characterizing storms is very difficult to model and thus difficult to predict and forecast however it has become increasingly important to study storm events in order to characterize how water is distributed over particular areas.

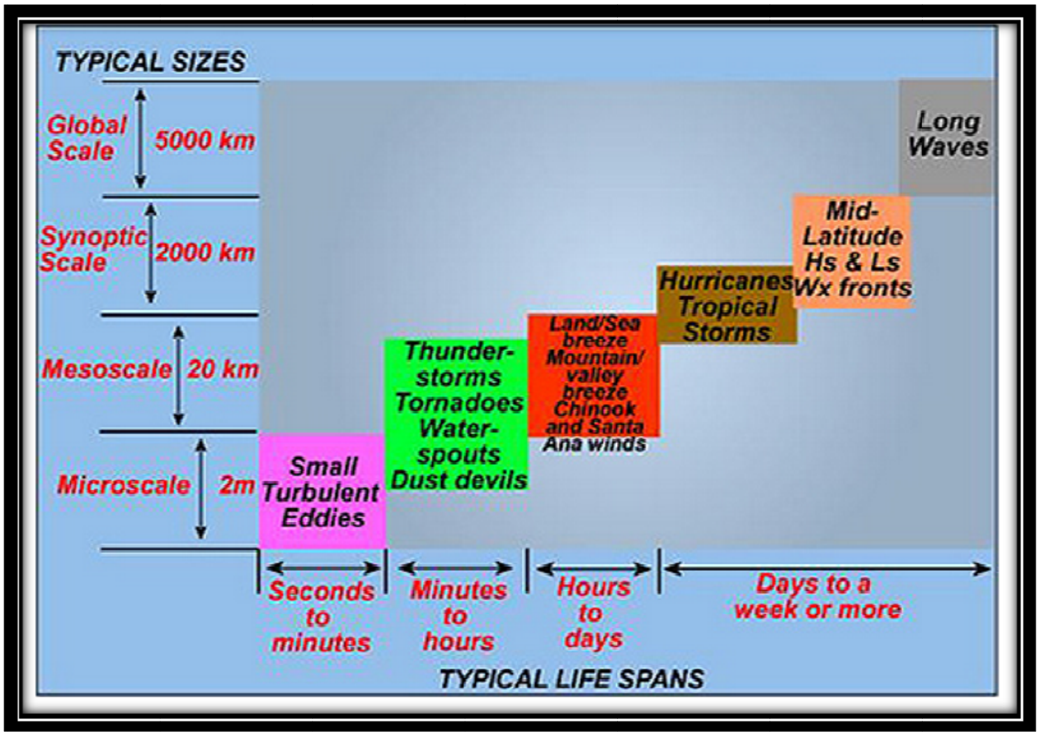


Figure 2.14 Time and Space Scale of Storm Events (NASA, 2010)

Storm analysis is an integral part of understanding how water is distributed over a specified area like a watershed. Analysis can focus on modeling the direct runoff into the stream networks within a given area or the associated damage costs from a storm induced flooding event. Regardless of the output, characterizing storm events as they vary in space and time is important to understanding the hydrometeorology of a system.

Storm events have traditionally been measured using a collection of discrete rain gauges distributed over an area. Although this provides a basic understanding of a storm's progression throughout its lifetime, it is still inadequate to fully describe the complexity of a storm event. Storms travel across areas in many different directions, orientation and at many different speeds. Without higher quantity and quality data, much of this information is lost or at the very least is interpreted incorrectly.

2.5.3 Design Storms and Depth Area Duration

Storm events can be characterized in many different ways. Relationships between area, intensity, frequency, duration have all been compiled for different regions of the United States. For water resources planners and manages, characterizing the storm events for a given area is of utmost importance. Design storms are typically used during the design of a hydrologic system. Design storms can be generated from historical data at a location or from the general characteristics of precipitation events surrounding the entire area.

One of the most common relationships used to describe storm events over a specified region are the depth-area-duration relationships or DAD curves formulated from a historical database. A DAD curve is used to determine either the maximum or average rainfall depth of a storm event for a given area and given duration. These relationships change over different regions depending on the meteorological conditions defining that area. DAD curves are important for the design of hydraulic structures because they relate the maximum precipitation depth calculated from a frequency analysis to a design area and duration.

CHAPTER 3: RESEARCH METHODS

3.1 Data Retrieval Through WaterOneFlow Web Services

As society quickly conforms to Web 2.0, where web applications facilitate interactive information sharing, web services have become increasingly popular. The phenomenon has expanded the perception of computers from a resource that can only create and manage information to a resource that can simultaneously create, manage and share information. Web services have aided in this movement by providing a middle ground through which clients and providers can communicate. In recent years, vast amounts of data have become freely available through the web.

Following this trend, large volumes of hydrologic and hydrometeorologic information have also become available. Despite having access to this information through an abundance of agencies, research groups and persona, hydrologists have often ignored this information because of the tedious steps required to actually organize and process the data. Web services have created a platform through which information from all these entities can be shared using standard protocols and procedures. Through web services, hydrologists can utilize standardized functions to query information across numerous data providers. Recognizing its benefit, CUAHSI has embraced the notion of web services and deemed it a crucial component to sharing observations data freely and efficiently with the world.

The best web services expose the fewest details about how the raw data or information is actually represented. They instead exchange information through simple methods and display information in common formats easily understood by the client. Figure 3.1, demonstrates the basic layered architecture of a web service. The application layer provides the raw information and is accessed through a WSDL. The WSDL is then queried by a SOAP service which retrieves the information in XML format. Following the standards defined by the W3C and the internet, this information is then accessed through HTTP, FTP, SMTP etc. protocols and transported to the client on the transport layer though which the information is finally displayed. The WSDL, SOAP, and HTTP standards all encompass what we call a web service. The web service layer retrieves and transforms data for the client. Because this layer performs a set of functions, it requires time, which in the computer science world, is called overhead (Tian et al., 2004). Depending on the service and the size of the request, overhead varies in length due to the parsing of XML code (Tian et al., 2004). Once the code is transformed, data is transferred at a constant rate. Again, depending on the service, the transfer rate varies however it is not typical for the rate to fluctuate by orders of magnitude.

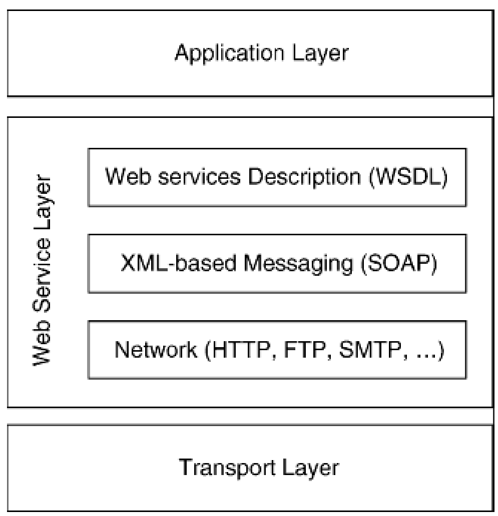
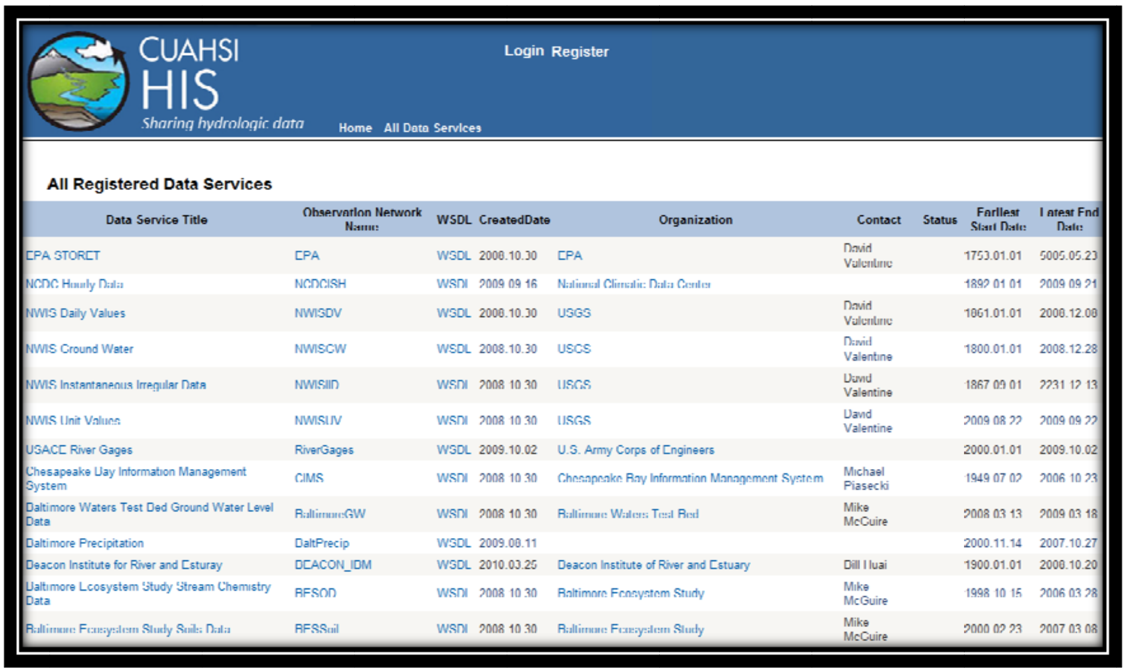


Figure 3.1 Web Service Architecture (Tian et al., 2004)

Part of the CUAHSI-HIS program is to assess the quality of and set standards for WaterOneFlow web services as they perform small to large data transfers from HIS servers to HIS applications via the internet. WaterOneFlow web services were designed to be generic and autonomous so that data can be accessed across multiple providers in a simple manner. CUAHSI-HIS has created a repository, HIS Central, through which all WaterOneFlow web services can be registered to aid clients in data discovery. HIS Central, seen in Figure 3.2, currently has 47 web services registered. Many of the web services are stored in virtual servers and libraries located all across the United States however HIS Central provides the means through which clients can discover data and utilize it by accessing the location provided by the WSDL.



Data Service Title	Observation Network Name	WSDL	CreatedDate	Organization	Contact	Status	Earliest Start Date	Latest End Date
EPA STORET	EPA	WSDL	2008.10.30	EPA	David Valentine		1753.01.01	5005.05.23
NCDC Hourly Data	NCDCISH	WSDL	2009.09.16	National Climatic Data Center			1892.01.01	2009.09.21
NWIS Daily Values	NWISDV	WSDL	2008.10.30	USGS	David Valentine		1861.01.01	2008.12.08
NWIS Ground Water	NWISGW	WSDL	2008.10.30	USGS	David Valentine		1800.01.01	2008.12.28
NWIS Instantaneous Irregular Data	NWISIID	WSDL	2008.10.30	USGS	David Valentine		1867.09.01	2231.12.13
NWIS Unit Values	NWISUV	WSDL	2008.10.30	USGS	David Valentine		2009.08.22	2009.09.22
USACE River Gages	RiverGages	WSDL	2009.10.02	U.S. Army Corps of Engineers			2000.01.01	2009.10.02
Chesapeake Bay Information Management System	CIMS	WSDL	2008.10.30	Chesapeake Bay Information Management System	Michael Piasecki		1949.07.02	2006.10.23
Baltimore Waters Test Bed Ground Water Level Data	BaltimoreGW	WSDL	2008.10.30	Baltimore Waters Test Bed	Mike McGuire		2008.03.13	2009.03.18
Baltimore Precipitation	BaltPrecip	WSDL	2009.08.11				2000.11.14	2007.10.27
Beacon Institute for River and Esturay	BEACON_IDM	WSDL	2010.03.25	Beacon Institute of River and Estuary	Bill Huai		1900.01.01	2008.10.20
Baltimore Ecosystem Study Stream Chemistry Data	BESOD	WSDL	2008.10.30	Baltimore Ecosystem Study	Mike McGuire		1998.10.15	2006.03.28
Baltimore Ecosystem Study Soils Data	BESSoil	WSDL	2008.10.30	Baltimore Ecosystem Study	Mike McGuire		2000.02.23	2007.03.08

Figure 3.2 Snapshot of HIS Central

Between the dates of September 1, 2008 and January 1, 2009, the author tested 26 WaterOneFlow web services for performance and speed. The goal was to evaluate the response time created by the web service layer during the data collection process. Table 3.1 below, identifies the names of the 26 services which were tested along with the organization that provided the data. Data services were selected from the HIS Central

Network Registry. All public data services registered as of 1/1/2009 were tested. Registered data services can be found at http://hiscentral.cuahsi.org/pub_services.aspx.

The performance of each web service can be characterized by a simple linear regression model that singles out the overhead time and the data retrieval rate. The model takes the form of the equation below:

$$N = \alpha(T - \beta)$$

where N = Number of Values returned and T = Retrieval Time. The parameters α and β represent the data retrieval rate (in values per second) and the overhead time (in seconds), respectively.

Table 3.1 WaterOneFlow Web Services Tested for Performance

Data Service Name	Data Service Name
Australian_ODM	MudLake
Baltimore Waters Test Bed Ground Water Level Data	NADP
CCBay	NCDC
CIMS	NWISDailyValues
COTCsnow	NWISUnitValues
DCEW	RCEW
EPA	SDRPF
IIHRNexrad	SFe_MICORWAVECITRA
IIHRTippingB	SFe_YSI
IIHRWQ	SNOTEL
Little Bear River	SRBHOS
MODMON	TCEW_TRACS
MPE	TWDB

Monitoring the performance of web services is an important aspect of an information system. In HIS, WaterOneFlow web services are what allow communication between the key components in the information system's SOA. Without swift performance standards, the service provided by an information system can become obsolete. Clients want to seek out information and access it quickly for synthesis, analysis or visualization. It is the efficiency in the data collection process which has allowed information systems like the CUAHSI-HIS to provide a superior alternative to managing hydrologic information.

3.1.1 WaterOneFlow Web Service Performance Testing

One of the primary concerns for implementing a web service layer between client applications and information systems is the overhead time required to call a web service. Previous studies on WaterOneFlow web services have shown that for data requests yielding less than 1000 values, the overhead cost is minimal (Goodall et al., 2008). For data requests in excess of 1000 values, overhead costs become more significant. (Goodall et al., 2008) Much of the overhead costs are hypothesized to be attributed to the verbosity of the XML therefore it has become a future goal of CUAHSI to assess the advantages of offering larger data requests in compressed formats. CUAHSI-HIS is interested in quantifying and standardizing the time required to collect data across multiple data providers. The CUAHSI-HIS project currently has three types of web services registered on HIS Central:

Local Services: Request made from a machine that houses both the WaterOneFlow web service and ODM database.

Remote Services: Request made from a machine that does not house the WaterOneFlow web service or the ODM database however the request is directed at another machine that does house the WaterOneFlow web service and ODM database.

Hybrid Services: Request made on machine that does not house the WaterOneFlow web service or the ODM database however the request is directed at another machine that does house the WaterOneFlow web service but not the ODM database.

Figure 3.3 below shows an illustration of the tree types of CUAHSI web services. Measuring the time required to call a web service for each of these types is essential to setting standards for web services in HIS.

The author has developed a procedure for testing the performance of WaterOneFlow web services. Utilizing the open source web service testing tool, soapUI, CUAHSI-HIS performed load tests for accessing a wide range of values for differing web services. Tests were performed using a Dell Precision PW 390, Intel Dual Core Processor, with 2.4 GHz and 2.00 GB of RAM. The internet connection speed was 100 mbps. In addition, a modified version of HydroExcel, a CUAHSI-HIS application, was used to display the results of the performance tests.

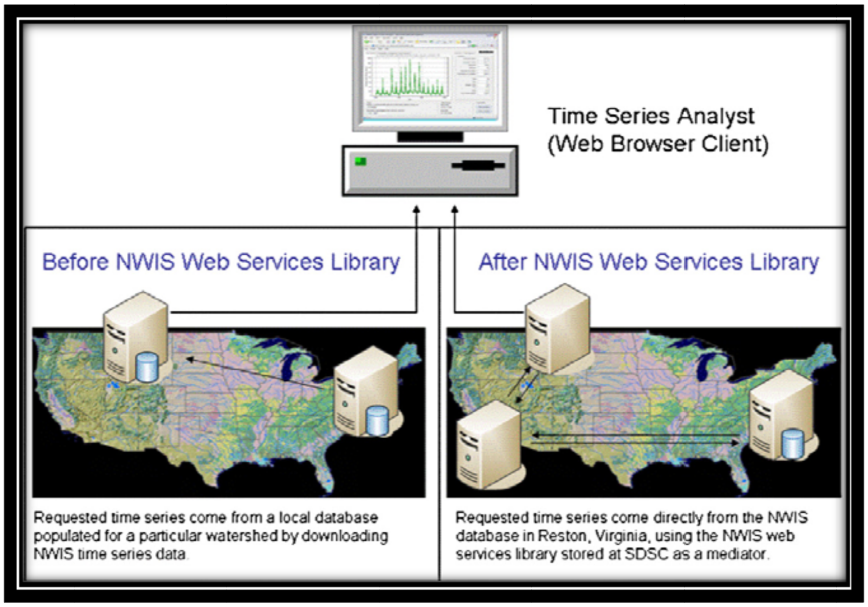


Figure 3.3 Types of CUAHSI-HIS Services (Goodall et al., 2007)

The following describes the testing procedure utilized for evaluating the performance speed of WaterOneFlow web services. Figure 3.4 shows a screenshot of HydroExcelTest during a performance test.

1. Using the HIS Central Network Registry, WSDLs for desired data services were acquired.
2. WSDLs were loaded into HydroExcelTest and used to populate the "SiteCatalog" worksheet.
3. Within the "SiteCatalog" worksheet, TimeSeries were selected to run performance tests. TimeSeries were chosen at random given that the TimeSeries contained a substantial number of values (> 500 values). If number of values in TimeSeries were insufficient, then TimeSeries within SiteCatalog containing the most values were chosen for testing.

4. Given the TimeSeries *Start Date* and *End Date* (obtained from SiteCatalog), TimeSeries were tested for response time using the SOAPUI utility within HydroExcelTest. The number of values tested, for a given TimeSeries, varied by changing the *Start Date* and *End Date* cells in the "HydroExcelTest" worksheet. For each data service, the number of values tested ranged from 0 to 100 percent of the total values contained within a given TimeSeries. Standard amounts of values tested were 100%, 60%, 50%, 30%, 25%, 5% and 2% of the total values.

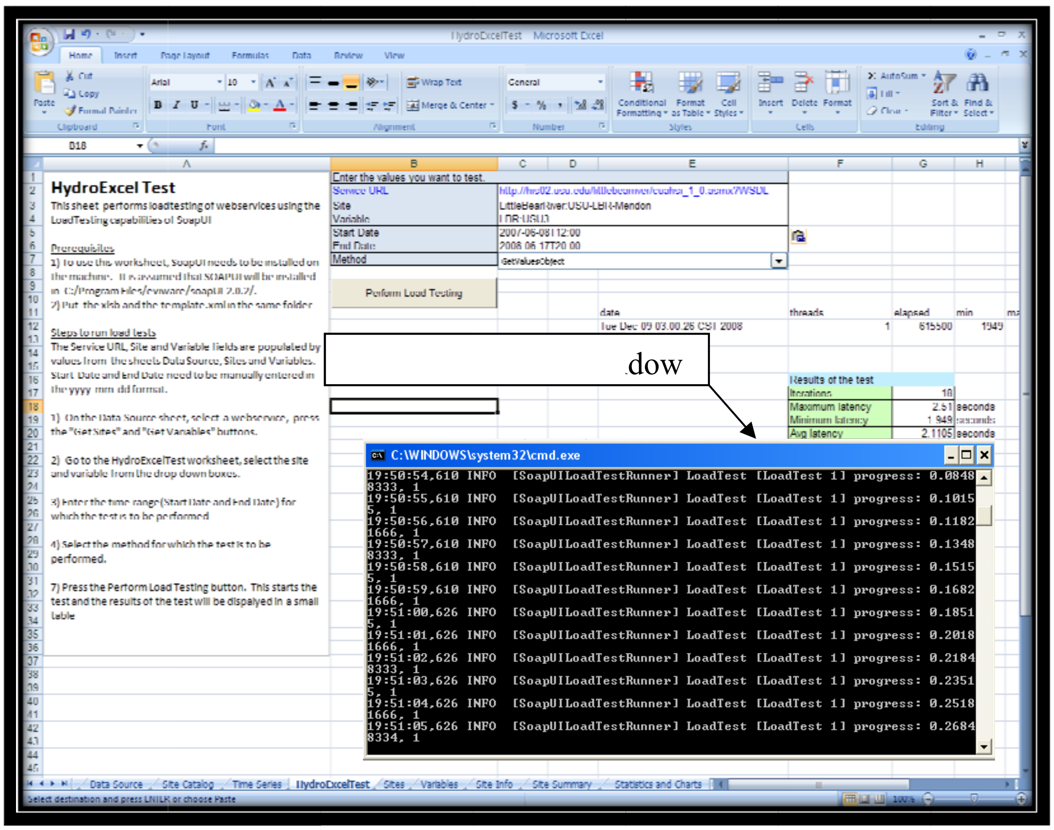


Figure 3.4 HydroExcel and soapUI

3.1.2 Performance Standards

Table 3.2 and Figure 3.5 below report the results of the performance testing executed on the aforementioned WaterOneFlow web services. It is worth noting that the TWDB web service is a local service and the NWISUnitValues, NWISDailyValues and EPA web services are hybrid services. All other web services listed are university services.

Table 3.2 Results of Performance Testing of WaterOneFlow Web Services

	Parameter	
Web Service Name	**a**	**b**
TWDB	31402	0.021
SDRPF	10083	0.158
SRBHOS	7259	0.484
CCBay	6702	0.087
Little Bear River	6511	0.554
IIHRWQ	6389	0.307
Baltimore Waters Test Bed Ground Water Level Data	6329	0.278
IIHRTippingB	6237	0.411
MudLake	6032	0.396
SFe_MICORWAVECITRA	5618	0.166
NADP	5339	0.495
IIHRNexrad	4647	0.563
SNOTEL	4612	1.034
DCEW	3270	1.417
SFe_YSI	2427	0.135
COTCsnow	2320	0.325
NWISUnitValues	2199	2.485
RCEW	2154	6.552
MPE	1969	0.300
CIMS	1753	0.351
TCEW_TRACS	1431	0.701
MODMON	1370	0.197
NWISDailyValues	909.5	0.517
NCDC	764.4	0.838
Australian_ODM	762.8	0.156
EPA	35.74	2.966

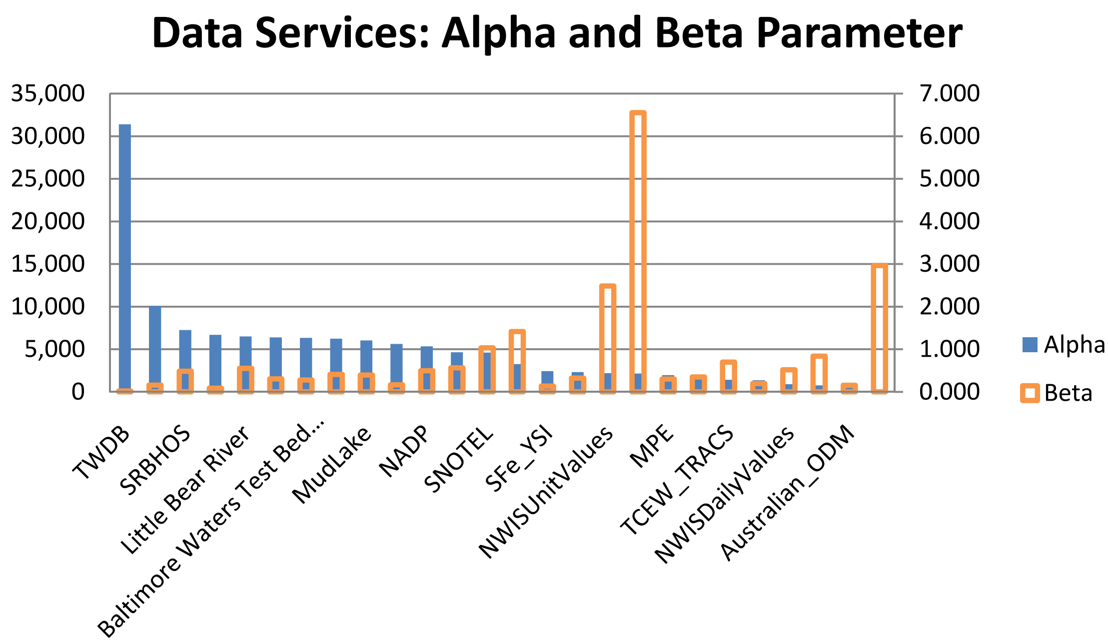


Figure 3.5 Alpha and Beta Parameters for WaterOneFlow Web Services

3.1.3 Local Services vs. Remote Services

The results presented in Table **3.2** suggest an inherited lag time from using proxy services to download time series data from ODM databases. Furthermore, services which are located on local networks have significantly shorter response times than services published on remote networks.

The Texas Water Development Board (TWDB) ODM database and WaterOneFlow web service is contained on a local network at the University of Texas-Austin Center for Research and Water Resources (CRWR). The results show almost an instantaneous response to a query with a download rate of approximately 30,000 values per second. In contrast, a query on the Little Bear River web service and ODM database, located at Utah State University, takes a fraction of a second to respond before delivering a rate of 5,000 values per second. Proxy services like the ones setup for the USGS at the San Diego Supercomputing Center (SDSC) demonstrate a response time between 0.5

seconds and 2.5 seconds with a download rate between 1,000 and 2,000 values per second. In the case of the USGS services, a query is made to the web service located at the SDSC and then from there a call is made to the ODM database contained in Reston, Virginia.

One result to note is the performance values reported by the EPA web service. The EPA service demonstrated a very slow response time and download rate relative to the other hybrid services. The EPA service takes approximately 3 seconds to respond before delivering a download rate of approximately 36 values per second. Both these values are significantly lower than the values reported by the other CUAHSI web services. The EPA uses Water Quality Exchange (WQX) as the standardized format for transferring information through the web, therefore when users make a request to the EPA via the CUAHSI web service, a transformation from WQX to WaterML must occur. It is this transformation that is responsible for the slow response.

3.1.4 Stored Database vs. Virtual Database

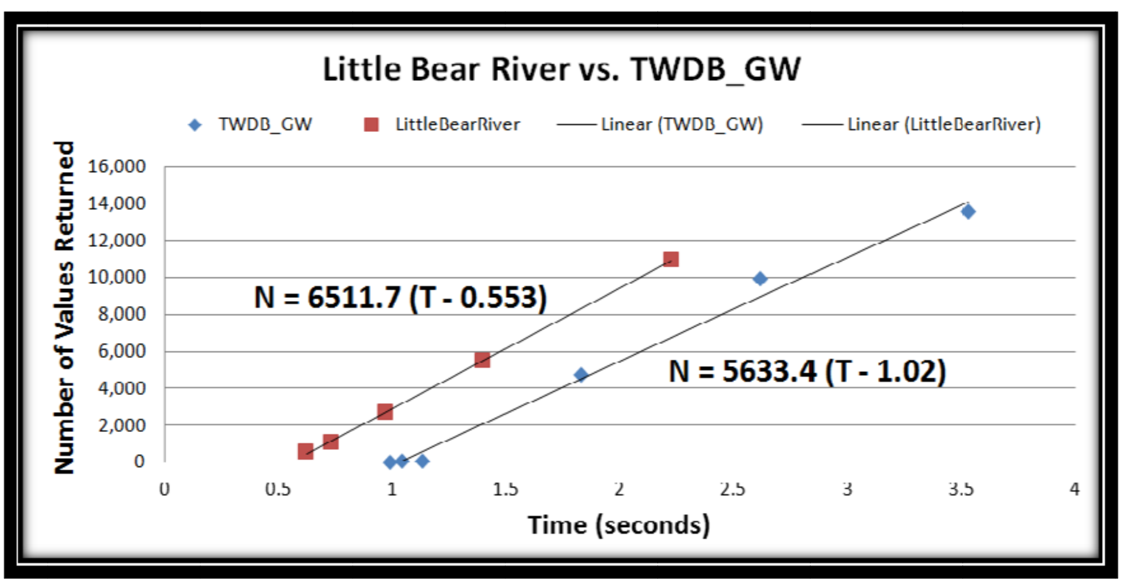


Figure 3.6 Performance Test for TWDB_GW and Little Bear Web Service

The TWDB_GW web service presented in Figure 3.6 represents a special adaptation of a CUAHSI WaterOneFlow web service. Generally, CUAHSI web services directly perform queries upon ODM databases which are managed by the data provider however the TWDB_GW service does not have an ODM database stored. Instead, the raw data is stored in a database in its original format and then transformed into an ODM database virtually through a Table View. An ODM Table View is a virtual table that is created from the original raw data table but is only created when the web service is called. Because of this extra transformation in the calling of the web service, the TWDG_GW performance test shows a lag time of 1.0 second, approximately double of the other CUAHSI services and then delivers a download rate of 6,000 values per second which is comparable to the other CUAHSI web services.

3.2 PUBLISHING THEMATIC INFORMATION AND DATASETS

HIS have provided the framework for data providers to share information with clients in an efficient manner. By successfully integrating web services with an observations data model, CUAHSI has developed the infrastructure through which data users can discover, harvest and manage hydrologic observations data. One of the keys to developing successful HIS are the WaterOneFlow web services which mask each of the datasets data providers want to publish. In the section above, WaterOneFlow web services have been shown to provide a swift and reliable method for users to collect hydrologic information. After discovering data through a metadata catalog like HIS Central, users download time series data, through the web service of a single provider, which represent a set of measurements for one variable at a specific site over a selected period of time. Although this offers a simple and organized scheme for obtaining information, there has been a push to thematically organize and publish hydrologic data available over numerous data providers instead of a single one.

A GIS allows for the management and organization of geospatial information by representing data in themes or layers of different classes of information that describe landscape (Maidment, 2008). Similarly, hydrologic and hydrometeorological information can also be grouped into themes or thematic datasets that describe specified variables or events. Thematic datasets organize and present observations data in a catalog that provides the user with sufficient descriptive information to directly access time series data from multiple different services provided that the time series falls within the description of the theme. Representing data in this format allows users to integrate information across multiple data sources and allows for synthesis along a specified thematic classification. Before the notion of a thematic dataset existed, users would only query information from a single service. Now that one can aggregate information from multiple services and share the data as a package, users can more efficiently download time series data.

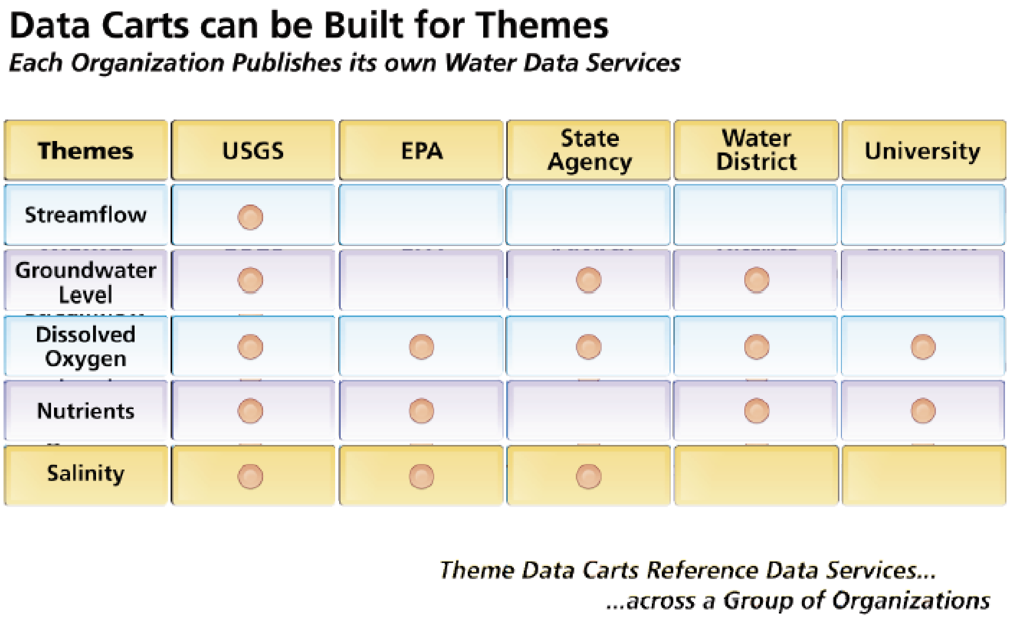


Figure 3.7 Organizing Data as Themes (Dangermond and Maidment, 2010)

There are two components for sharing observations data as a theme. There is (1) the thematic datacart or thematic dataset catalog and (2) the thematic dataset itself. The thematic datacart is the collection of all the time series metadata available across all networks that pertain to the specified theme. This does not include the actual time series values. The thematic dataset on the other hand provides the metadata along with the actual values in a relational geodatabase that supports the Arc Hydro time series data structure. Because this method of presenting data inherently combines GIS features along with time series, thematic datasets are shared using the OGC standards for sharing geospatial information over the web. Web Feature Services (WFS) are the preferred method for sharing themes using OGC services. Recently, ESRI has released a social network, ArcGIS Online, that allows users to share geospatial data quickly as files. With these two technologies available, CUAHSI has been working on compiling thematic datasets and sharing them with the hydrologic and hydrometeorological communities.

3.2.1 Building a DataCart

The thematic datacart is the fundamental object that allows for efficient management of thematic information. A datacart is in essence an aggregated catalog or list of all the time series available across all services contained within HIS. A datacart is presented in a table where each row represents a single time series for a single variable from a specific service. Figure 3.8 below shows an example of a thematic datacart for salinity from the HIS for Texas.

Attributes of SalinityDataCart

ServCode	SiteName	VarName	ValueCount	StartDate	EndDate	SiteCode	Latitude	Lon
TCEQ	GUADALUPE	SALINITY,#OFMEASUREMENTSIN24-HRS	14	5/2/2002 12:00:00 PM	8/6/2006 2:00:00 PM	12577	28.476667	-96
TCEQ	GUADALUPE	SALINITY,#OFMEASUREMENTSIN24-HRS	1	6/15/2003 11:00:00 AM	6/15/2003 11:00:00 AM	12577	28.476667	-96
TCEQ	GUADALUPE	SALINITY,24-HR,AVERAGE,PPT	14	5/2/2002 12:00:00 PM	8/6/2006 2:00:00 PM	12577	28.476667	-96
TCEQ	GUADALUPE	SALINITY,24-HR,AVERAGE,PPT	1	6/15/2003 11:00:00 AM	6/15/2003 11:00:00 AM	12577	28.476667	-96
TCEQ	GUADALUPE	SALINITY,24-HR,MINIMUM,PPT	14	5/2/2002 12:00:00 PM	8/6/2006 2:00:00 PM	12577	28.476667	-96
TCEQ	GUADALUPE	SALINITY,24-HR,MINIMUM,PPT	1	6/15/2003 11:00:00 AM	6/15/2003 11:00:00 AM	12577	28.476667	-96
TCEQ	GUADALUPE	SALINITY,24-HR,MAXIMUM,PPT	14	5/2/2002 12:00:00 PM	8/6/2006 2:00:00 PM	12577	28.476667	-96
TCEQ	GUADALUPE	SALINITY,24-HR,MAXIMUM,PPT	1	6/15/2003 11:00:00 AM	6/15/2003 11:00:00 AM	12577	28.476667	-96
TCEQ	GUADALUPE	SALINITY-PARTSPERTHOUSAND	169	6/22/1980 6:30:00 PM	7/9/2006 3:40:00 PM	12577	28.476667	-96
TCEQ	GUADALUPE	SALINITY-PARTSPERTHOUSAND	8	6/6/2005 2:00:00 PM	8/6/2006 2:00:00 PM	12577	28.476667	-96
TCEQ	ARROYO CO	SALINITY,#OFMEASUREMENTSIN24-HRS	1	5/20/2002 1:35:00 PM	5/20/2002 1:35:00 PM	13073	26.264999	-97
TCEQ	ARROYO CO	SALINITY,24-HR,AVERAGE,PPT	1	8/16/2005 9:01:00 AM	8/16/2005 9:01:00 AM	13073	26.264999	-97
TCEQ	ARROYO CO	SALINITY,24-HR,AVERAGE,PPT	1	5/20/2002 1:35:00 PM	5/20/2002 1:35:00 PM	13073	26.264999	-97
TCEQ	ARROYO CO	SALINITY,24-HR,MINIMUM,PPT	1	8/16/2005 9:01:00 AM	8/16/2005 9:01:00 AM	13073	26.264999	-97
TCEQ	ARROYO CO	SALINITY,24-HR,MINIMUM,PPT	1	5/20/2002 1:35:00 PM	5/20/2002 1:35:00 PM	13073	26.264999	-97
TCEQ	ARROYO CO	SALINITY,24-HR,MAXIMUM,PPT	1	8/16/2005 9:01:00 AM	8/16/2005 9:01:00 AM	13073	26.264999	-97
TCEQ	ARROYO CO	SALINITY,24-HR,MAXIMUM,PPT	1	5/20/2002 1:35:00 PM	5/20/2002 1:35:00 PM	13073	26.264999	-97
TCEQ	ARROYO CO	SALINITY-PARTSPERTHOUSAND	99	12/9/2000 11:36:00 AM	8/10/2002 10:45:00 AM	13073	26.264999	-97
TCEQ	ARROYO CO	SALINITY-PARTSPERTHOUSAND	105	11/30/1997 11:30:00 AM	5/6/2006 1:45:00 PM	13073	26.264999	-97
TCEQ	STAR LAKE	SALINITY,#OFMEASUREMENTSIN24-HRS	5	9/9/2003 11:00:00 AM	3/15/2006 11:00:00 AM	10485	29.979166	-93
TCEQ	STAR LAKE	SALINITY,24-HR,AVERAGE,PPT	5	9/9/2003 11:00:00 AM	3/15/2006 11:00:00 AM	10485	29.979166	-93
TCEQ	STAR LAKE	SALINITY,24-HR,MINIMUM,PPT	5	9/9/2003 11:00:00 AM	3/15/2006 11:00:00 AM	10485	29.979166	-93
TCEQ	STAR LAKE	SALINITY,24-HR,MAXIMUM,PPT	5	9/9/2003 11:00:00 AM	3/15/2006 11:00:00 AM	10485	29.979166	-93
TCEQ	STAR LAKE	SALINITY-PARTSPERTHOUSAND	145	10/15/1989 12:00:00 PM	7/3/2006 10:26:00 AM	10485	29.979166	-93
TCEQ	NUECES RIV	SALINITY,#OFMEASUREMENTSIN24-HRS	4	8/21/2002 12:35:00 PM	7/12/2006 1:30:00 PM	12960	27.846666	-97
TCEQ	NUECES RIV	SALINITY,24-HR,AVERAGE,PPT	4	8/21/2002 12:35:00 PM	7/12/2006 1:30:00 PM	12960	27.846666	-97
TCEQ	NUECES RIV	SALINITY,24-HR,MINIMUM,PPT	4	8/21/2002 12:35:00 PM	7/12/2006 1:30:00 PM	12960	27.846666	-97
TCEQ	NUECES RIV	SALINITY,24-HR,MAXIMUM,PPT	4	8/21/2002 12:35:00 PM	7/12/2006 1:30:00 PM	12960	27.846666	-97

Record: 1 Show: All Selected Records (0 out of *2000 Selected) Options

Figure 3.8 Thematic DataCart for Salinity in Texas

Using the Texas HIS, it important to understand the process involved in building a datacart. In the Texas HIS, numerous agencies provide observations data using the ODM and publish their information using WaterOneFlow web services. However, from one agency to the other there is no standard or protocol for naming variables. Each service has its own unique variable naming scheme. For example TWDB calls their salinity data “Total_Salinity” while TPWD calls their salinity data “Salinity”. Because there is a disconnect amongst agencies, it is important to implement a well-developed ontology system in order to build thematic datacarts efficiently. CUAHSI has built an ontology and is currently working to implementing it across all of its data services. The system classifies variables into broader categories called “concepts” and applies them to every variable provided by a data provider. Concepts were developed using various standards including the EPA substance registry system. With a successful ontology in place, thematic datacarts can now be created on the fly using the concept field in the ODM. This means that “Total Salinity” and “Salinity” variables can all be classified as “Salinity” and included in the creation of a datacart.

The design of a thematic datacart table incorporates metadata that completely describes the origin of each time series included in the table. Each time series in the datacart has the location and name of the variable of which it is representing and the name of the service that is providing the data. In addition, the row has the URI of the time series so that the user can call the web service to access the data. The complete design of the datacart can be seen below in Table 3.3.

Table 3.3 DataCart Design

Field Name (Field Type)	Definition	Example
ServCode (Text)	Network prefix for site codes used by the WaterOneFlow service, giving the context within which the site code applies	CCBay
SiteName (Text)	Name of a site	Hypoxia_1
VarName (Text)	Name of a variable	Dissolved Oxygen Concentration
ValueCount (Integer)	Number of time series values for the variable at the site for the given time period	270
StartDate (Date)	Start date and time for the time period of the variable at the site	5/3/94 8:40 AM
EndDate (Date)	End date and time for the time period of the variable at the site	8/31/06 11:26 AM
SiteCode (Text)	Unique text identifier for a site within a given WaterOneFlow service	H1
Latitude (Decimal)	Latitude of the site location in decimal degrees	27.814
Longitude (Decimal)	Longitude of the site location in decimal degrees	-97.141
Vocabulary (Text)	Vocabulary prefix for variable codes giving the context within which the code applies	CCBay
VarCode (Text)	Unique text identifier for a variable within a given WaterOneFlow service	DOC
Concept (Text)	Leaf concept from the HIS ontology to which this variable applies	dissolvedOxygen
VarUnits (Text)	Units of measure for the variable	milligrams per liter
IsRegular (Boolean)	TRUE if variable is measured/calculated regularly in time; FALSE otherwise	FALSE
TimeUnits (Text)	For regular data, the time step and time units give the length of time between measurements, e.g., 1 day, 6.5 hrs, 1 month	Day
TimeStep (Decimal)	For regular data, the time step and time units give the length of time between measurements, e.g., 1 day, 6.5 hrs, 1 month	1

Table 3.3 Continued

DataType (Text)	Type of data	Value Average Maximum Minimum StandardDeviation
Medium (Text)	Medium in which the variable applies	Surface Water
MethodID (Integer)	Unique ID within a WaterOneFlow service for the method used to measure the variable	1
Method (Text)	Description of the method used to measure the variable	Multiprobe measurement
QCLevelID (Integer)	Unique ID within a WaterOneFlow service for the quality control level of the time series	0
QCLevel (Text)	Description of the quality control level of the time series	Raw Data
SourceID (Integer)	Unique ID within a WaterOneFlow service for the original source of the data	1
SourceName (Text)	Name of the original source of the data	Texas A&M University Corpus Christi
ServType (Text)	Type of service -- indicates how the Location parameter of a WaterOneFlow.GetValues call should be formatted	SiteCode LatLongBox LatLongPoint
XLL (Decimal)	For point data, Longitude of the point. For data defined by a lat/lon box, western longitude of the box	-97.141
YLL (Decimal)	For point data, Latitude of the point. For data defined by a lat/lon box, southern latitude of the box	27.814
XUR (Decimal)	For data defined by a lat/lon box, eastern longitude of the box	-93.5
YUR (Decimal)	For data defined by a lat/lon box, northern latitude of the box	30.2
Location (Text)	Properly formatted location parameter to pass to WaterOneFlow.GetValues	CCBay:Hypoxia_1 GEOM:BOX(-97.141 27.814,-93.5 30.2) GEOM:POINT(-97.141 27.814)

Table 3.3 Continued

Variable (Text)	Properly formatted variable parameter to pass to WaterOneFlow.GetValues	CCBay:DOC
AuthToken (Text)	Authentication token for a WaterOneFlow service (typically left blank)	
WaterMLURI (Text)	URI of WaterOneFlow service WSDL	http://data.com/WoF/ /cuahsi_1_0.asmx?WSDL
WebSiteURI (Text)	URI of web site describing the data or project	http://data.com /
WFSURI (Text)	URI of web feature service showing site locations	http://data.com /WFSServer
WMSURI (Text)	URI of web mapping service related to the data	http://data.com /WMSServer
DAccessURI (Text)	URI of Data Access Service, which provides REST querying capabilities for WaterOneFlow, user management, data cart management, and more	http://data.com/DataService

3.2.2 Creating a Thematic Dataset

Although thematic information can be shared using the table design of the thematic datacart, thematic information can also be shared through the creation of a thematic dataset. Thematic datasets provide a more complete method for sharing information as a theme. Thematic datasets organize information as a collection of tables and shapefiles all packaged within a relational geodatabase. A thematic dataset includes the thematic datacart itself along with tables that contain site information, variable information and the actual time series values. By establishing relationships among key fields in the tables, relationships can be formed to enhance the management of the information within a GIS. Figure 3.9 below shows a screenshot of a thematic dataset being utilized in ArcGIS.

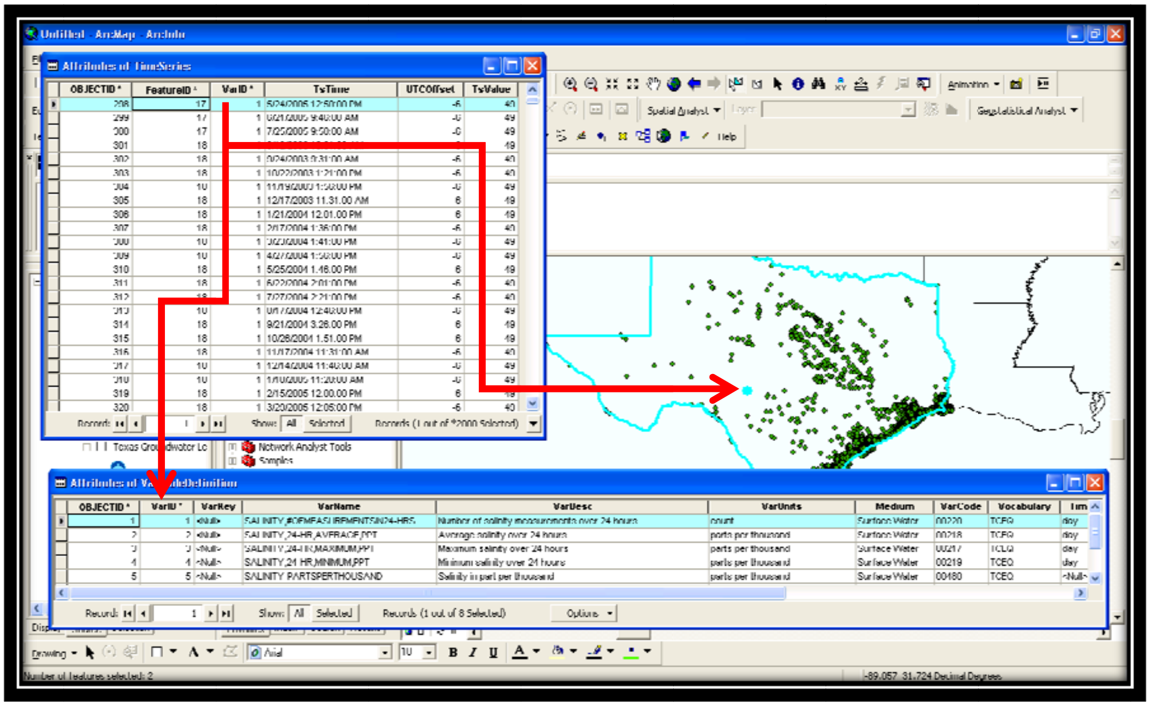


Figure 3.9 Thematic Dataset for Salinity in Texas

3.2.3 Serving and Accessing Thematic Data

There are currently two prototypical methods for sharing thematic data via the web: utilizing (1) HydroPortal and (2) ArcGIS Online. HydroPortal is an online metadata repository and platform for sharing geospatial data. Data providers can create an account, register their data and publish it in three steps. HydroPortal allows data providers to submit detailed information describing their dataset and post data access links for users to follow to retrieve data. Access links on HydroPortal are primarily used for GIS services such as MapServices and WFSServices. Once users access the information through their local system, users can visualize and synthesize thematic data using GIS.

One advantage to utilizing the HydroPortal as a vehicle for sharing thematic information is the interoperability and flexibility it provides both the data provider and data user. Instead of creating redundancy by storing volumes of data on a server, HydroPortal simply stores the metadata and directs individuals to the location of the data. HydroPortal works as a catalog service which facilitates the administration, publishing and discovery of geospatial information. See Figure 3.10.

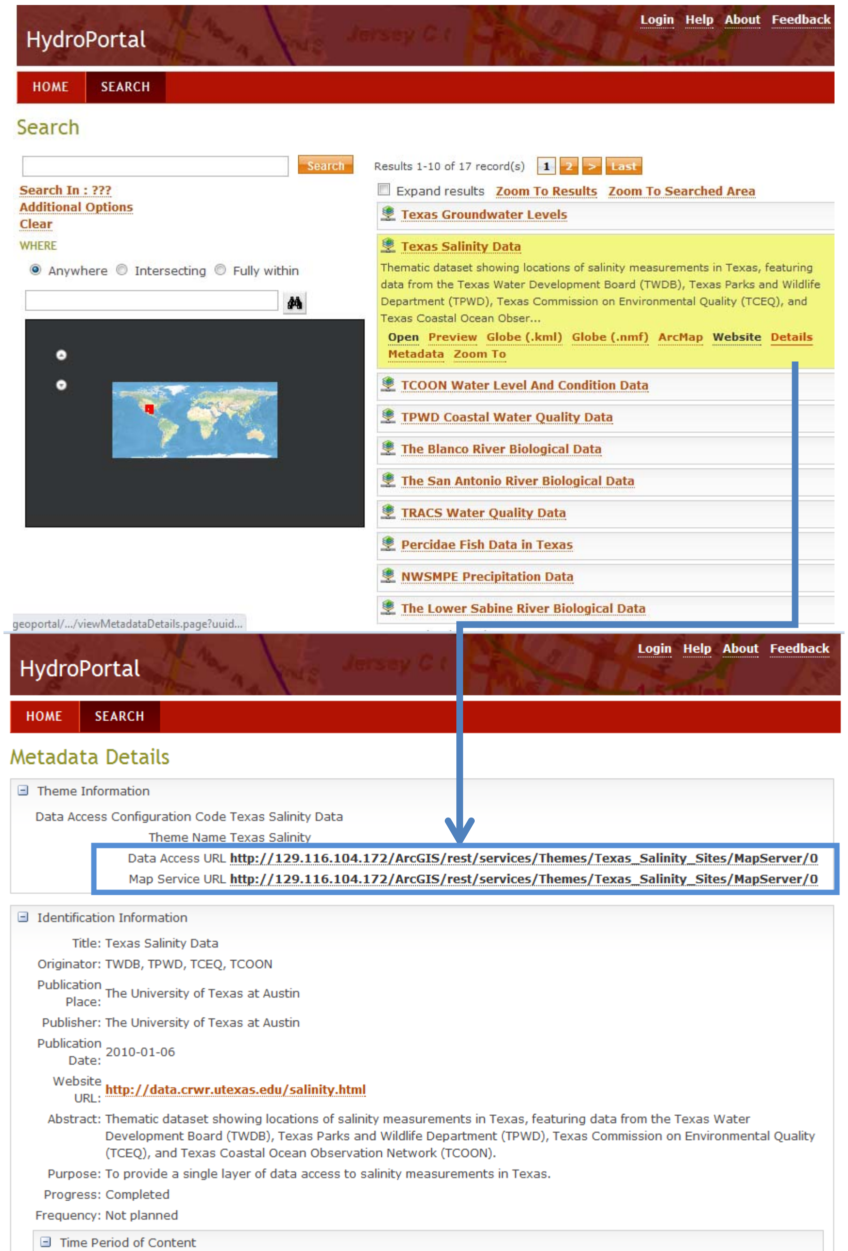


Figure 3.10 Accessing Thematic Information via HydroPortal

The Environmental Systems Research Institute (ESRI) has recently released ArcGIS Online, a platform that enables users to share geospatial data efficiently via the web. Similar to social networks like Facebook, ArcGIS Online enables users to create profiles and join different groups depending on their data interests. Users can then ingest GIS data directly into ArcGIS Desktop for further evaluation. Just as easily, users can upload data onto ArcGIS Online by providing GIS Layer Package which is simply a special compressed format for serving GIS data.

ArcGIS Online is a suitable resource for sharing thematic information because it provides visualization tools as well as analytical tools. Because of the inherent spatiotemporal domain of thematic data, it is important to provide a manner in which data can be served and accessed while preserving its spatial and temporal integrity.

Figure 3.11 Utizling ArcGIS Online (Dangermond and Maidment, 2010)

3.3 Building a Framework for Hydrometeorological Information Systems

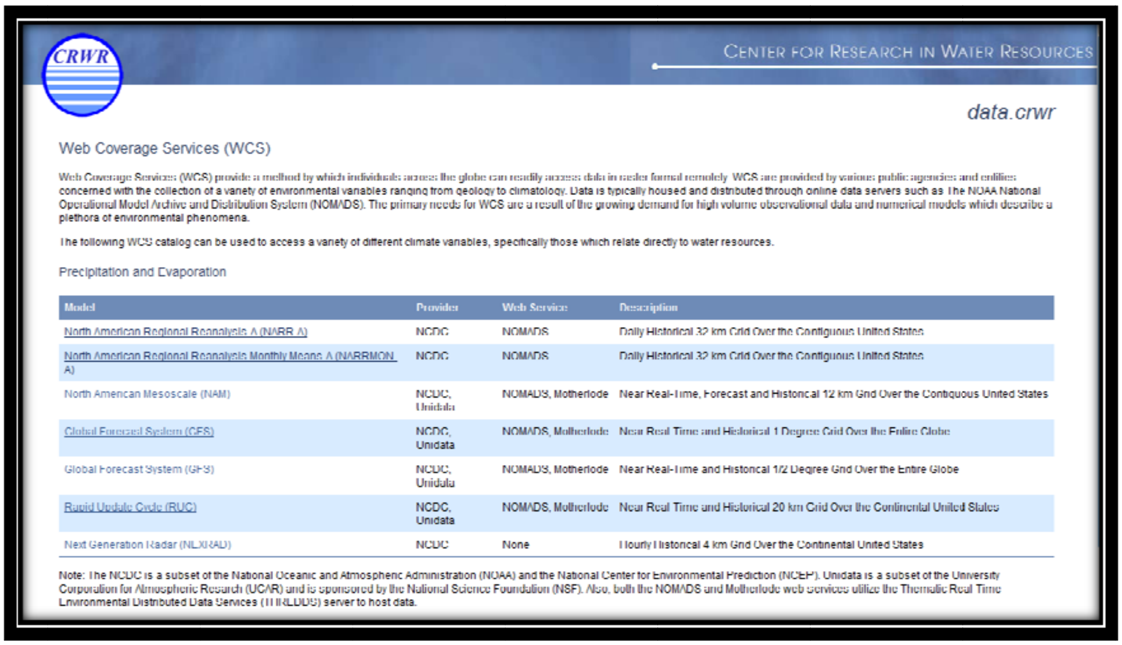
As has been described in the sections above, information systems provide an efficient and methodical approach to managing, synthesizing and sharing geospatial and temporal information via the web. In hydrology, the goal has been to integrate point observations data collected by multiple sources and utilize tools and models to analyze the information. In hydrometeorology however, data from grids typically come from one source not several, therefore, the goal is not to integrate datasets together but instead to separate them out.

Information pertaining to water resources covers a wide spectrum of data types ranging from areas, lines to points (i.e. watersheds, streams and gages). The development of WaterML and WaterOneFlow web services by CUAHSI has allowed for the efficient management and integration of observations data with geospatial information. Water resources however, encompasses copious amounts of information which isn't collected at specific points. A plethora of data is now being collected using satellites and remote sensing technologies which are best represented on a series of grids. Following this trend, information systems need to adapt and incorporate new methods to manage, share and integrate gridded data with the already integrated geospatial and temporal data so that improved methods in forecasting and modeling can be applied efficiently without the hassle of collecting data.

3.3.1 Hydrometeorological Information Catalog

Similar to HIS Central within the CUAHS-HIS project, an essential component to building a successful HMIS is the Hydrometeorological Information Metadata Catalog. This catalog serves as an integrated database that manages the numerous data sources and providers of hydrometeorological data. Combined with a standardized ontology for hydrometeorological information, the metadata catalog documents the metadata for each data service and provides a platform through which data users can index, query and access spatiotemporal data efficiently.

One example of a Hydrometeorological Information Metadata Catalog is the Web Coverage Services (WCS) catalog available on the data.crwr server (see Figure 3.12 below). Here, a list of WCS have been compiled to facilitate the sharing of hydrometeorological information. Data links pertaining to precipitation and evaporation variables are documented in conjunction with suitable metadata describing the dataset. Metadata includes information about the variables available, the spatial and temporal extent, variable units as well as data type.



CRWR

Center for Research in Water Resources

data.crwr

Web Coverage Services (WCS)

Web Coverage Services (WCS) provide a method by which individuals across the globe can readily access data in raster format remotely. WCS are provided by various public agencies and entities concerned with the collection of a variety of environmental variables ranging from geology to climatology. Data is typically housed and distributed through online data servers such as The NOAA National Operational Model Archive and Distribution System (NOMADS). The primary needs for WCS are a result of the growing demand for high volume observational data and numerical models which describe a plethora of environmental phenomena.

The following WCS catalog can be used to access a variety of different climate variables, specifically those which relate directly to water resources.

Precipitation and Evaporation

Model	Provider	Web Service	Description
North American Regional Reanalysis A (NARR A)	NCDC	NOMADS	Daily Historical 32 km Grid Over the Contiguous United States
North American Regional Reanalysis Monthly Means A (NARRMON A)	NCDC	NOMADS	Daily Historical 32 km Grid Over the Contiguous United States
North American Mesoscale (NAM)	NCDC, Unidata	NOMADS, Motherlode	Near Real-Time, Forecast and Historical 12 km Grid Over the Contiguous United States
Global Forecast System (GFS)	NCDC, Unidata	NOMADS, Motherlode	Near Real Time and Historical 1 Degree Grid Over the Entire Globe
Global Forecast System (GFS)	NCDC, Unidata	NOMADS, Motherlode	Near Real-Time and Historical 1/2 Degree Grid Over the Entire Globe
Rapid Update Cycle (RUC)	NCDC, Unidata	NOMADS, Motherlode	Near Real Time and Historical 20 km Grid Over the Continental United States
Next Generation Radar (NEXRAD)	NCDC	None	Hourly Historical 4 km Grid Over the Continental United States

Note: The NCDC is a subset of the National Oceanic and Atmospheric Administration (NOAA) and the National Center for Environmental Prediction (NCEP). Unidata is a subset of the University Corporation for Atmospheric Research (UCAR) and is sponsored by the National Science Foundation (NSF). Also, both the NOMADS and Motherlode web services utilize the Thematic Real Time Environmental Distributed Data Services (THREDDS) server to host data.

Figure 3.12 Metadata Catalog for Web Coverage Services (http://data.crwr.utexas.edu)

Similar to the CUAHSI WaterOneFlow web services developed for observations data, Unidata has developed the THREDDS server for storing and accessing multidimensional arrays or grids provided by multiple different data sources (Maidment, 2008). The THREDDS server acts as the intermediary between the data provider and data user by standardizing the format in which gridded data is made accessible regardless of the format the underlying data is stored in. Because THREDDS allows the data user to access subsets of large datasets (i.e. data over a specified time domain T or data over a specified region A or data for a specified group of variables V), THREDDS has become the most robust tool for facilitating the interoperability and integration between spatiotemporal data, geospatial data and temporal data. Because of its wide range of capabilities, hydrometeorological information catalogs should be created around data published on the THREDDS server.

3.3.2 Visualizing Hydrometeorological Information

Despite being the leading technological advancement for visualizing and analyzing geospatial data, GIS still lacks the tools to analyze spatiotemporal data in an efficient and meaningful manner. GIS provides a consistent framework and set of spatial analysis tools that allow users to gain insight on data that is static over long periods of time. Studies pertaining to land use, soil type and elevation all fall under this category. The analysis of hydrometeorological information requires the integration of hydrology (temporal data), meteorology (spatiotemporal data) and GIS (geospatial data) however a framework for assimilating these three areas of science has yet to be developed.

There are two basic frameworks that have historically been used for analyzing spatiotemporal data: the Eulerian frame of reference and Lagrangian frame of reference (Goodall et al., 2008). The Eulerian framework refers to observations at a stationary location while the Lagrangian framework refers to observations along the path of motion of an entity. In hydrometeorological systems, the Eulerian framework can be applied to a watershed where gridded data can be transformed into measurements of areal precipitation recorded over a period of time; this can be referred to as a geospatial time series. For certain modeling applications, this type of information can provide sufficient data for characterizing the response of a watershed system given some input, in this case precipitation. In other modeling applications however, it may be necessary to analyze how precipitation events develop as they move across a region (i.e. watershed, state etc.). The Lagrangian framework in this case can be used to either track the location of a storm or study the intensity distribution of a storm throughout its duration. Regardless of the application, visualizing gridded data is best represented when one can group large amounts of data together either by a region of space (geospatial time series, average etc.) or through a duration of time (daily average, monthly average etc.) or both (areal precipitation as function of space and time). These capabilities are within the limits of GIS however the framework applies to how the data is collected as well as the tools available for analysis.

Various properties and relationships can be measured and created through the utility of a well-designed functional framework and a set of tools for analyzing spatiotemporal data. Cumulative distributions, probability distributions, frequency analysis, stochastic modeling etc. can all be modeled through the extensive use of both the lagrangian and eulerian frameworks.

3.3.3 Managing Hydrometeorological Information

One of the main difficulties in modeling spatiotemporal data lies in the organization and management of data that varies in space and time. It has been accepted that grids are the best solution for providing a high quantity and quality of data to describe various hydrometeorological phenomena however it can be shown that data lacking relationships and context are just as inefficient as having no data at all.

Arc Hydro has provided a reliable method for managing spatiotemporal data in a relational database. The Arc Hydro data model originally was developed to extend the capabilities of GIS into the temporal domain, however new developments have given GIS the ability to further extend its analytical potential. This model for spatiotemporal data was presented in Chapter 2 however is important for understanding how best to manage and visualize gridded information.

One of the key aspects stressed in the Arc Hydro data model is the notion of the relational database. The idea that one feature can have relationships with one or many other features helps researchers organize and filter out information efficiently. This idea can be extended to a vast amount of hydrometeorological information or any type of data for that matter. One idea is to define events in space and time and create relationships between all the critical pieces of information that characterize and describe that event. For example, a storm event can be related to the watersheds it is moving across. In addition, the watersheds can be related to the stream network within its domain. The stream network can be related to the various observational gauges along its network etc. Similar to the layered approach and framework for GIS, events can be defined by the layers that describe its very essence (i.e. meteorology, hydrology, geology, stream network, observations etc.) Relationships are what provide context to data. By developing a framework through which relationships between different datasets can be created, managing and sharing information over a spatiotemporal domain will improve the quality of scientific information in the United States.

CHAPTER 4: RESULTS

4.1 Utilizing Hydrometeorological Information Systems

The sections and chapters above have provided detailed information about some of the key components that are critical to providing the necessary infrastructure for building HMIS. In contrast to HIS, where time series information is collected at multiple point locations, hydrometeorological information is generally collected over a continuous domain in both space and time. It is not yet fully understood how information varying in space and time can be integrated with geospatial and temporal data because many of the variables that relate to hydrometeorology have different applications and are visualized, analyzed and synthesized differently. The following case study is a demonstration of how HMIS can be useful in performing hydrometeorological analysis.

4.1.1 Precipitation

One of the least understood processes in hydrometeorology is precipitation. It is a complex phenomenon that couples both physical processes on the surface and in the atmosphere and incorporates dynamics in many different scientific disciplines. Many of us have had experiences where we have had to rely on the local weather forecast to predict whether or not it will rain the next day. Many of these experiences have resulted in the weakening of our faith in our local meteorologist. There is a reason why weather forecasts report the “chance of precipitation” instead of the actual amount. It is, simply put, difficult to predict precipitation.

The NWS and the US Army Corps of Engineers studied precipitation events until the 1950s (Tomlinson, 2009). Since then there have been very few national studies performed. If one looks at the archive of reports used as references for developing depth-area-duration and probable-maximum-precipitation curves for the entire United States, the majority of these reports date back to the 1970s.

Recent developments in remote sensing technology and satellite imagery give reason to believe that these analyses can be improved upon. For the greater part of the last century, precipitation studies have primarily been based on point precipitation gages however we know that precipitation events occur over an area. Because we have lacked the ability to measure areal precipitation directly, many studies over the last few decades have been performed estimating this very quantity (Olivera, 2006).

It is widely accepted that large watersheds are less likely to experience high intensity storms over the area as whole than smaller watersheds (Siriwardena, et al., 1996). It is also understood that precipitation events can have high variability over small scales of space and time. Because of this uncertainty in spatial and temporal scales, it is important to study the characteristics of precipitation events as they relate to the watersheds that govern the distribution of water throughout hydrologic systems. Some of the most important characteristics that describe a precipitation event are: depth, area, duration, frequency, velocity and orientation. Depending on the combination of all these variables, hydrologic systems will respond differently to different precipitation events. As forecasting and modeling become more important with the effects of climate change, characterizing precipitation events over specified areas also becomes important.

4.1.1.1 Storm Events as Entities

Precipitation patterns are variable in both space and time. At any given moment, the precipitation rate at one location can be significantly higher or lower than the precipitation rate at another location even if the distance between them is relatively small. Similarly, at any given location, the precipitation rate from one moment to the next can be significantly higher or lower. Because of this variability along both scales of space and time, precipitation data has been difficult to assimilate and manage for information sharing. Contrary to observations data where time series at specified locations can be retrieved, the process for accessing precipitation data on grids has yet to be formalized.

One solution for sharing precipitation data focuses on the notion that precipitation data can be represented as a collection of storm events or entities. This is analogous to representing stream flow as a collection of time series however storm entities convey information that varies in space and time not just time. This abstraction is important to conveying the idea of a storm because precipitation events occur over an area not just a location. Storms also move in space as a function of time, thus the total precipitation in a storm entity fluctuates as it moves in space.

Managing precipitation as a collection of storm events is useful for integrating hydrometeorology with GIS. By representing a storm event as a feature, storms can have properties and attributes which describe the event throughout its duration. The attributes associated with a storm event can convey and provide all the information needed to perform insightful analysis of a hydrometeorologic system. Linking GIS with the Feature Series in the Arc Hydro data structure can provide an alternative method for managing and sharing hydrometeorological information.

4.1.1.2 Storm Tracking and Extraction

As was discussed in the previous chapter, storms can be tracked as a function of both space and time following the Lagrangian framework. Using NEXRAD precipitation data, it is possible to extract single storm events using a spatial and temporal algorithm that tracks the connectivity between one storm cell to the next (Tucker and Li).

A study at the University of Kansas-Lawrence has developed an algorithm for extracting storm events from NEXRAD datasets. In this study a storm event is defined as a contiguous precipitation object in space and time (Tucker and Li). Storm events are delineated by selecting a connected set of precipitation cells from a collection of stacked hourly raster layers. Figure 4.1 shows an example of an extracted storm event in space and time. The following sections in this chapter utilize the database generated from this storm extraction study.

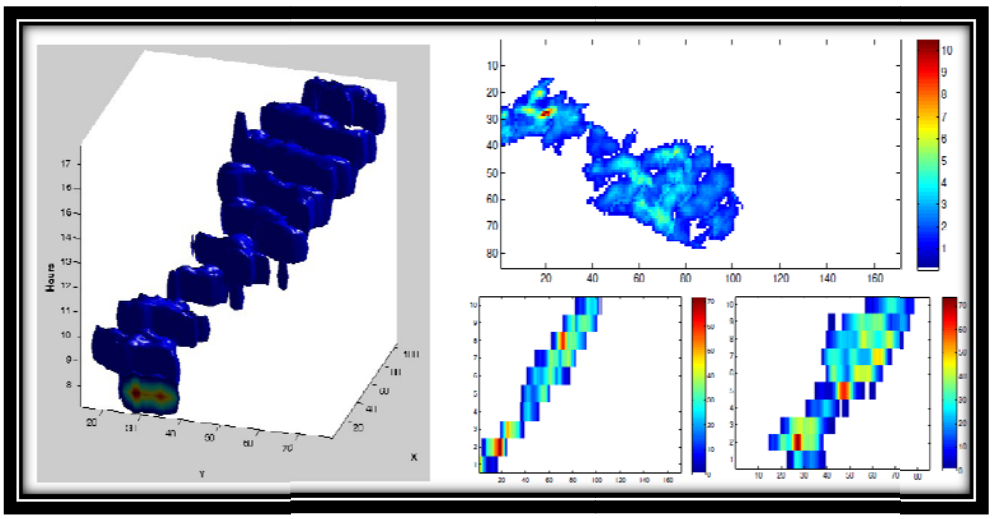


Figure 4.1 Sample Extracted Storm Event from NEXRAD

4.1.2 Visualizing Storm Data

Storm events can be visualized using GIS. Figure 4.2 shows an example of a collection of storm events depicted as features. Attached is also the attribute table associated with the storm events. The attribute table contains descriptive information about each storm event.

StormID: Unique identifier for a storm event.

StartTime: Start date and time of the storm event.

EndTime: End date and time of the storm event.

TotalPrecip: Total precipitation depth over the entire storm event.

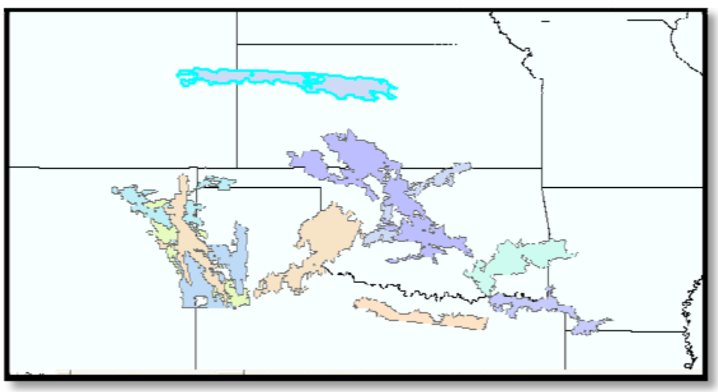
Duration: Duration of the storm event.

Area: Area of the storm event.

MeanPrecip: Mean precipitation depth over the entire storm event.

CentroidX: Geographic coordinate of the centroid of the storm event. Longitude.

CentroidY: Geographic coordinate of the centroid of the storm event. Latitude.



Attributes of Sheet3$

StormID	StartTime	EndTime	TotalPrecip	Duration	StormCells	FootprintSize	Area (sq# mi)	MeanPrecip (mm)
57302	12/18/1995 11:00:00 PM	12/19/1995 8:00:00 AM	45969.54	10	5932	3711	32879.179138	12.387373
65971	5/10/1996 1:00:00 AM	5/10/1996 10:00:00 AM	20732.67	10	2251	1823	16151.642029	11.37283
64923	5/21/1996 11:00:00 PM	5/22/1996 8:00:00 AM	13587.49	10	1312	799	7079.079529	17.00562
36810	6/3/1996 6:00:00 PM	6/4/1996 3:00:00 AM	12093.49	10	870	766	6786.702026	15.787846
62630	6/6/1996	6/6/1996 9:00:00 AM	34208.51	10	3450	3079	27279.70697	11.110266
62634	6/14/1996 11:00:00 PM	6/15/1996 8:00:00 AM	58302.28	10	4050	2260	20023.428955	25.797469
24808	6/17/1996 7:00:00 PM	6/18/1996 4:00:00 AM	24858.69	10	2042	1261	11172.364563	19.713473
24966	6/27/1996 1:00:00 AM	6/27/1996 10:00:00 AM	17261.04	10	1537	1108	9816.796143	15.578556
46272	6/30/1996 7:00:00 PM	7/1/1996 4:00:00 AM	9195.13	10	768	550	4872.958374	16.718418
58262	7/7/1996 4:00:00 AM	7/7/1996 1:00:00 PM	19347.75	10	1877	1412	12510.213135	13.702373

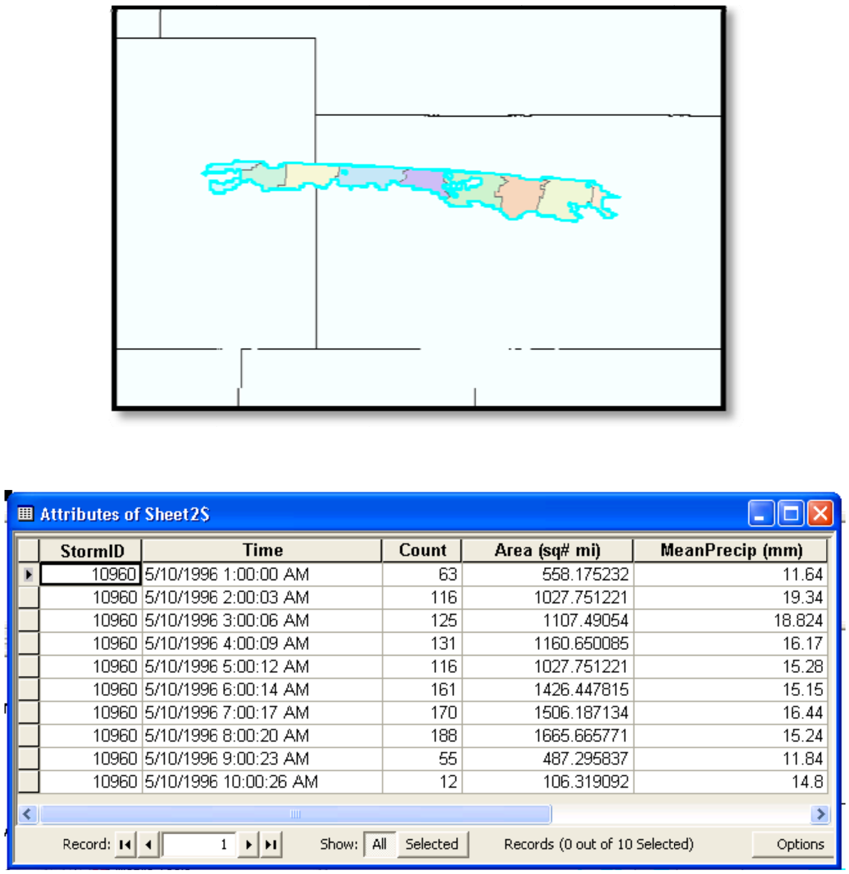
Record: 1 Show: All Selected Records (1 out of 10 Selected) Options

Figure 4.2 Collection of Storm Events over the Study Area

4.1.3 Managing Storm Data

Utilizing this unique method of sharing precipitation data, data users can efficiently assemble hydrometeorological information for a variety of different analyses. In addition to providing the areal footprint for a storm event, the NEXRAD database contains the areal footprint for each individual hour of the storm. As can be seen in

Figure 4.3 below, the feature series for a storm event can be visualized using GIS. With this information, researchers can study how the distribution of precipitation within a storm changes as the storm event develops over time. Combining GIS features with raster images in the representation of a storm event can provide all the necessary information to describe a storm event in complete detail.



StormID	Time	Count	Area (sq# mi)	MeanPrecip (mm)
10960	5/10/1996 1:00:00 AM	63	558.175232	11.64
10960	5/10/1996 2:00:03 AM	116	1027.751221	19.34
10960	5/10/1996 3:00:06 AM	125	1107.49054	18.824
10960	5/10/1996 4:00:09 AM	131	1160.650085	16.17
10960	5/10/1996 5:00:12 AM	116	1027.751221	15.28
10960	5/10/1996 6:00:14 AM	161	1426.447815	15.15
10960	5/10/1996 7:00:17 AM	170	1506.187134	16.44
10960	5/10/1996 8:00:20 AM	188	1665.665771	15.24
10960	5/10/1996 9:00:23 AM	55	487.295837	11.84
10960	5/10/1996 10:00:26 AM	12	106.319092	14.8

Figure 4.3 Example Feature Series for a Storm Event

4.2 CHARACTERIZING HYDROMETEOROLOGICAL INFORMATION – STORM EVENTS

It can be demonstrated through a simple analysis and characterization of storm events that sharing hydrometeorological information as GIS features as opposed to raster images can be very beneficial and efficient in the study of hydrometeorological phenomena.

4.2.1 Arkansas-Red River Basin

The geographic extent for this study covers the Arkansas-Red River basin located along the great plains of the central United States. The area extends through five states all of which share a fairly flat topography. See Figure 4.4 for details. The database created for this study spans 11 years of stage III NEXRAD data from January 1995 to December 2006.

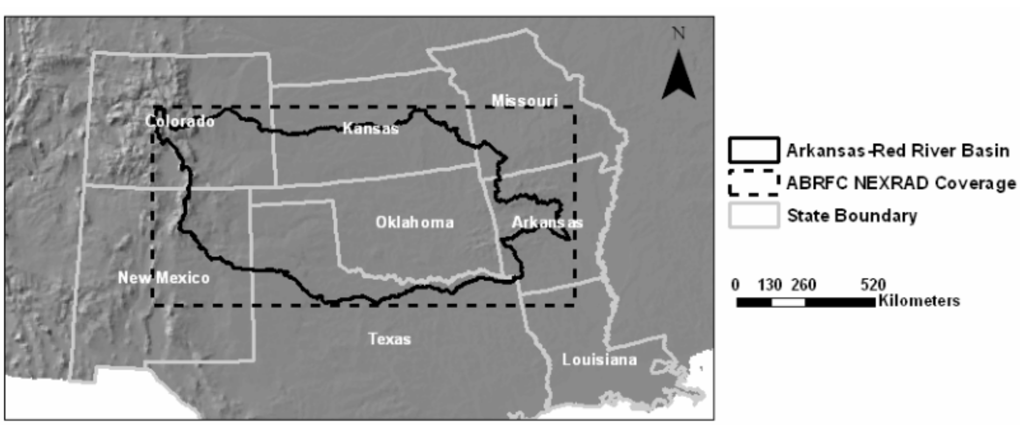


Figure 4.4 Arkansas-Red River Basin

Approximately 73,000 storms were extracted by Tucker and Li using the spatiotemporal storm extraction algorithm mentioned above. Storms ranging from 2 hours to 20 hours in duration were collected and stored in a relational geodatabase. Figure 4.5 and 4.6 demonstrate some of the general characteristics of the entire storm collection. Figure 4.5 depicts the number of storms extracted for each storm duration length while Figure 4.6 depicts the average areal precipitation intensity for varying storm durations. As is typical for watersheds located in the central plains, Figure 4.5

demonstrates that the Arkansas-Red River basin encounters significantly more short duration storms that long duration. Approximately 95% of the storms that pass through the basin last less than 5 hours. This is characteristic of convective storms that have rapidly changing rainfall intensity. To support this claim, one can see from Figure 4.6 that the intensity for short duration storms have a much greater variability than do storms with longer durations.

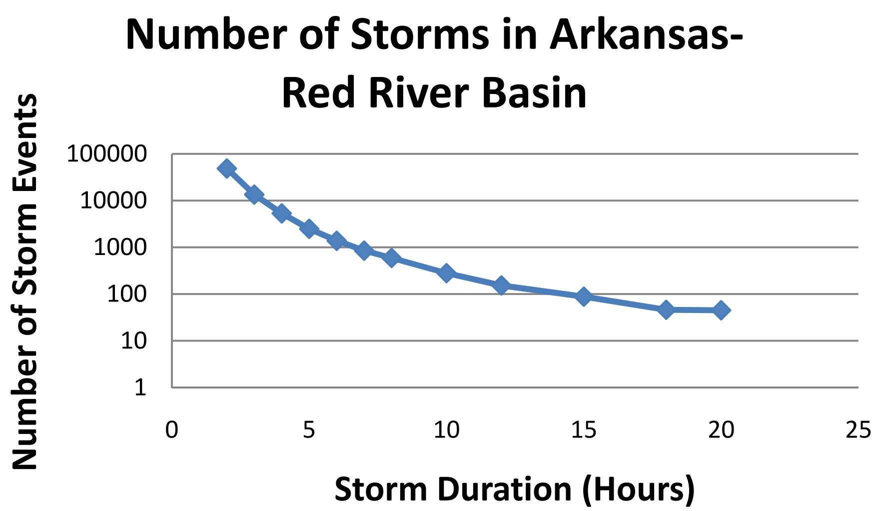
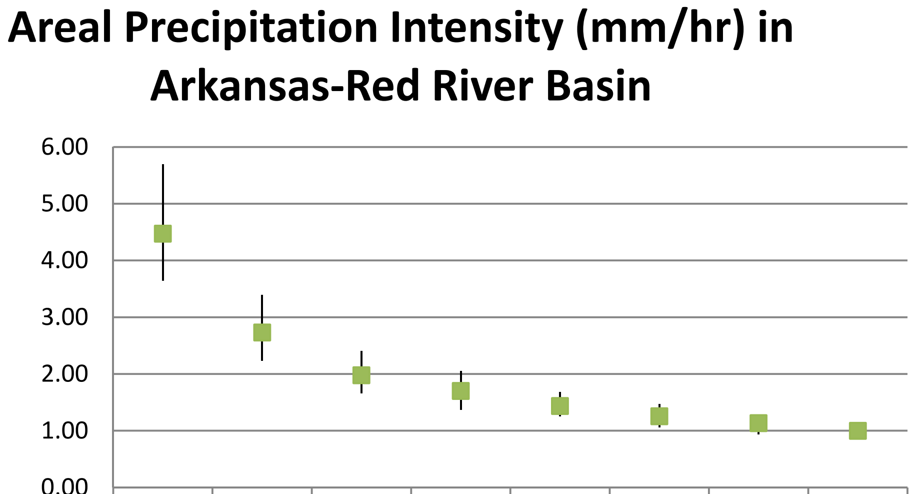


Figure 4.5 Number of Storms for Varied Duration in Arkansas-Red River Basin



	2 Hour Storms	5 Hour Storms	8 Hour Storms	10 Hour Storms	12 Hour Storms	15 Hour Storms	18 Hour Storms	20 Hour Storms
75th Percentile	5.69	3.39	2.40	2.06	1.68	1.47	1.24	1.15
25th Percentile	3.64	2.23	1.66	1.37	1.25	1.06	0.94	0.85
Median	4.47	2.73	1.98	1.70	1.43	1.26	1.13	1.00

Figure 4.6 Average Areal Precipitation Intensity in Arkansas-Red River Basin

4.2.2 Distribution of Storm Events

Although it can be said that the majority of the storms crossing the Arkansas-Red River basin are short duration convective storms, it is important to further characterize the short duration storms by how intense they are. As was saw in Figure 4.6, short duration storms are highly variable in magnitude. Figure 4.7 shows the cumulative distribution function for storms of different durations. One can see that approximately 70% of the 2 hour storms have an average areal precipitation of 10 mm which is approximately 5 mm/hr. Conversely, there is only a fraction of 20 hour storms that have an average precipitation of 10 mm; equivalent to approximately 0.5 mm/hr. This demonstrates that the majority of long duration storms have an average intensity rate of 0.5 mm/hr and thus can sustain a significant amount of precipitation over a larger area.

Short duration storms, however, cannot sustain over a long duration because they contribute a significant amount of rain onto the landscape below at a rapid rate.

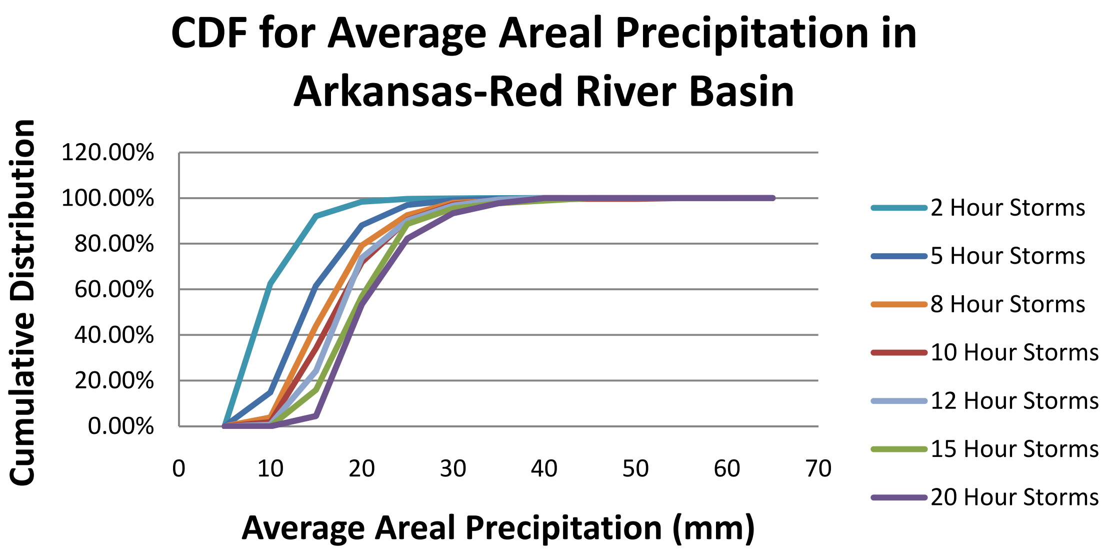


Figure 4.7 CDF for Areal Precipitation in Arkansas-Red River Basin

In addition to characterizing the strength of a storm, it is also important to characterize storms by the area they affect. Similar to the chart generated in Figure 4.7, Figure 4.8 demonstrates the variability in the footprint of a storm as a function of duration. One can see that as the duration of the storm decreases, the more likely it is for the storm to resort to a characteristic area.

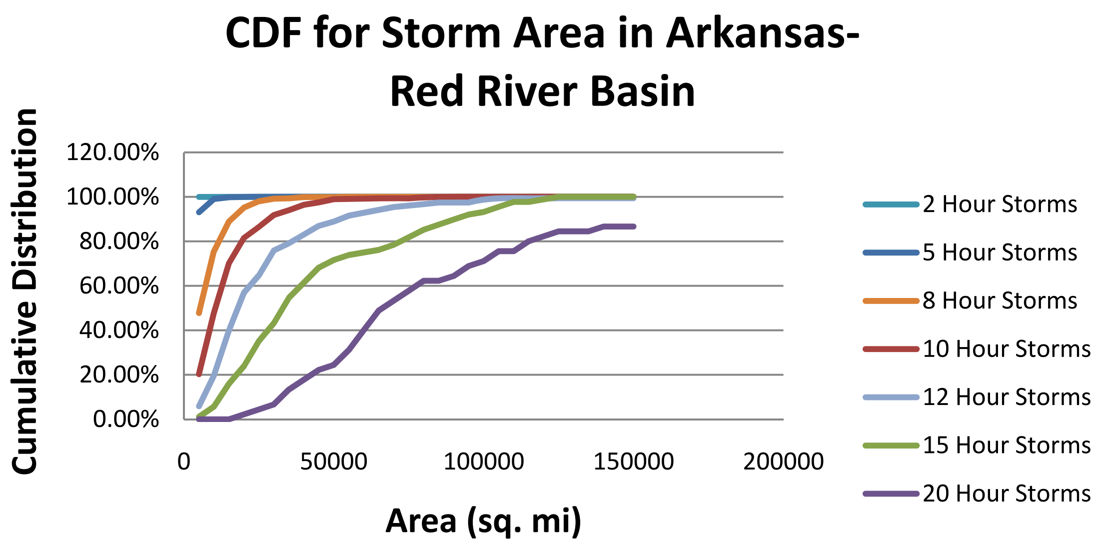


Figure 4.8 CDF for Storm Area in Arkansas-Red River Basin

Although one can study the average areal precipitation and area for numerous storm events, one can also perform frequency analysis on the storm events in the database. Because a storm entity is represented as a feature series in the Arc Hydro data model, each storm has an associated timestamp attached to each event. In this case, each storm event has a beginning and end date. Figure 4.9 shows the storm events with the maximum precipitation rate between 1995 and 2006. Again, as the duration of the storm decreases, there is a noticeable shift to increased storm intensity.

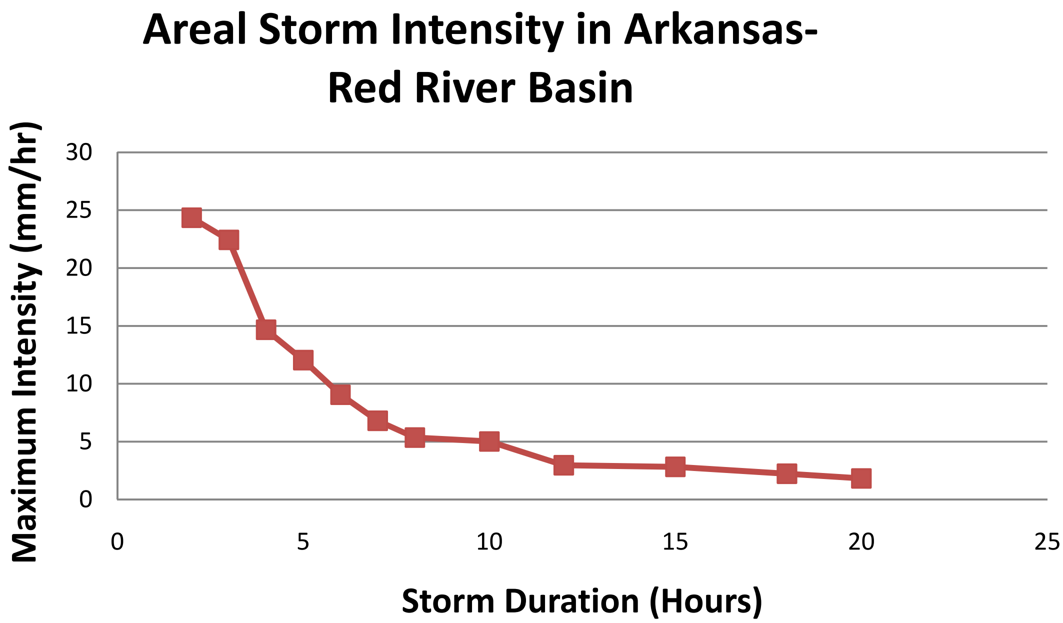


Figure 4.9 Maximum Precipitation Intensity

4.2.3 Intensity Area Duration Frequency Curves

After performing frequency analyses, hydrometeorologists typically use relationships between storm precipitation depth and area to fit design storms to various hydrometeorological systems. If one plots the average area of a storm for a given duration against the average areal precipitation, an exponential relationship emerges (see Figure 4.10). As average precipitation increases, so does area. However, if one plots the area for a given storm against its equivalent areal intensity, the exponential relationship vanishes and instead a collection of clusters emerge. What one can see from Figure 4.11 is that as the duration of the storms increase the clusters come together. This implies that longer duration storms are more likely to share like characteristics whereas short duration storms appear in all different shapes and sizes (typical of convective storm events). This diagram also suggests that storms should be evaluated as multi-dimensionally distributed random variables.

One noticeable property of the diagram in Figure 4.11 is that there are distinct envelopes and thresholds to characterize storm events. One can see that storm events with specified durations have a theoretical minimum and maximum intensity. The same can be said for the affected areas of each storm however the distinctions between one group of storms and the other is not as well defined.

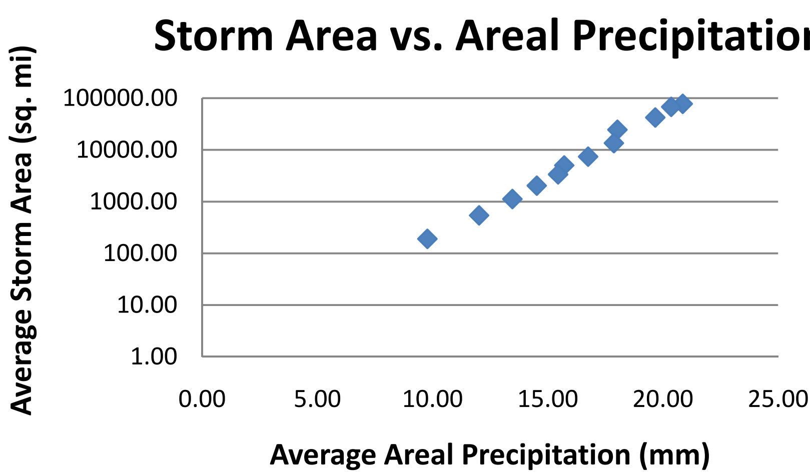


Figure 4.10 Storm Areas vs. Areal Precipitation

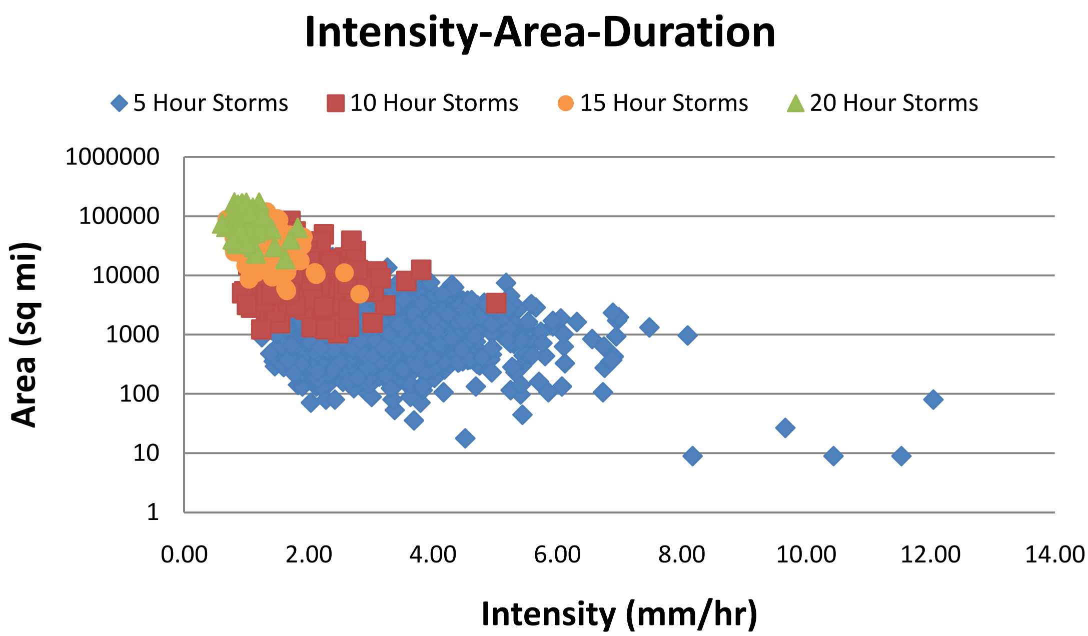
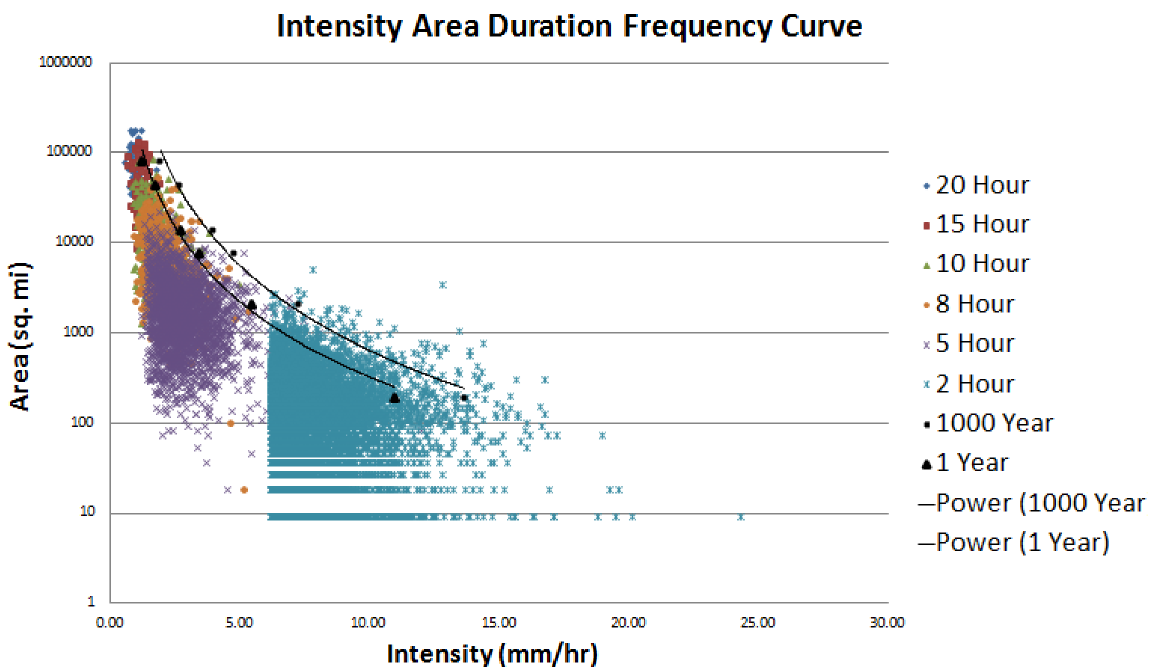


Figure 4.11 Intensity-Area-Duration Curves

Finally, extending the application of the Intensity-Area-Duration diagram, one can superimpose contours that represent theoretical return periods for storms of specified duration, area and intensity. In Figure 4.12, one can see that storms of different durations are multi-dimensionally distributed in clusters along different portions of the diagram. Characterizing each group of storms, one can identify the probability distribution of a likely event occurring and characterize that event as one with a duration, area, intensity and frequency.



4.12 Intensity-Area-Duration-Frequency Chart

CHAPTER 5: CONCLUSIONS

In today's world, data is being captured at a rate far greater than ever imaginable. Data about global markets, national infrastructures, environmental systems etc. are all being collected using sensors and computers on a network that store, manage and distribute information. There is no doubt that the world currently is equipped with the technology to ingest oceans of data however there is still a struggle to use the resources we have to distribute information to individuals in a format that can be used to make prompt, smart and progressive decisions.

In the scientific world, collecting, managing and utilizing the volumes of data that are freely available across the web is a problem that has persisted for decades. Organizations and agencies have locked data away in databases with frustrating passwords and puzzling schemas that have hindered potential scientific advancements that could otherwise be achieved if data were to be shared more readily and more importantly with context.

In an effort to improve the accessibility to scientific information, the CUAHSI-HIS project has paved the road to unlocking the boundless mounds of hydrologic information housed in databases all across the world. CUAHSI has implemented database schemas and integrated web data services as part of an information system to efficiently facilitate how individuals store, share, discover and access hydrologic observations data in the form of time series.

With the increasing need for improvements in modeling and forecasting in the field of water resources, there has been a push to integrate the fields of hydrology and meteorology into a science called hydrometeorology. Because the water cycle is a coupled system that incorporates areas of hydrology, meteorology and modeling, the improvements for sharing information need to be extended to hydrometeorological systems.

Advancements in technology have led to the development of satellites and radar that have increased the quantity and quality of hydrometeorological information that is available. Because hydrometeorological systems are highly variable across both small and large scales of space and time, hydrometeorological information has primarily been shared in the form of a grid or a collection of grids (multi-dimensional arrays). To aid in the storing, sharing, discovering and accessing of hydrometeorological information, hydrometeorological information systems (HMIS) can be developed as the medium or tool through which data providers and data users communicate.

5.1 BENEFITS

HMIS are crucial to advancing the science of hydrometeorology. With data being stored by a plethora of sources in a variety of different formats, HMIS are useful in that they provide a central location through which one can discover and access data provided by numerous different agencies, organizations and groups.

One of the most important aspects of creating a vehicle, through which information accessed, is providing a service that will meet the efficiency needs for the data user. Users simply want to access information quickly therefore in order to provide a robust data service, the performance and speed of transferring data need to be monitored and standardized. Web services, specifically WaterOneFlow web services which can be utilized to access hydrometeorological information, have been shown to provide a quick and reliable service for users to retrieve data. WaterOneFlow services take approximately 0.5 seconds to initiate and then transfer data at a minimal rate of 1,000 data values per second depending on the whether or not proxy services are utilized.

With an efficient and reliable system for transferring data in place, the facilitating of data discovery becomes a primary goal. Data users would like the ability to discover and access hydrometeorological information from multiple data sources efficiently. One approach is to manage and share information thematically through the utility of a thematic datacart or thematic dataset. In the past, accessing information related to one variable was restricted to querying one data source at a time. However by organizing information in a table across multiple data sources, users can now access data pertaining to one variable by querying all the data sources in the system at once. This method of organizing information increases the utility and efficiency of the data collection process.

Similar to the HIS that are currently in place for collecting hydrologic information, HMIS need to be developed to facilitate the collection of hydrometeorological data. Many of the same components in HIS can be utilized when creating HMIS. Metadata catalogs need to be managed so users can quickly discover and query information pertaining to hydrometeorology. Servers need to be created to store the volumes of information on grids and finally, tools need to be developed to ingest, access and synthesize hydrometeorological information quickly. Currently, the most efficient tools for accessing hydrometeorological information are provided by Unidata. Unidata has created the netCDF data structure and THREDDS server. The THREDDS server provides an organized framework for storing gridded data from a variety of different models and data providers whereas netCDF is a tool that allows for the management and storing of multidimensional arrays or grids.

Although HMIS will be primarily utilized to share information on grids, some hydrometeorological information can best be shared using features in GIS. Utilizing advancements in radar technology (NEXRAD) and the Arc Hydro data model, precipitation information can be collected and distributed as a collection of storm events.

Each event can be represented as a feature with an associated attribute table that will store all the information necessary to providing context for the entity. In addition, creating relationships in a relational database can be helpful in organizing storm events with the appropriate objects that relate to the hydrometeorological system they are a part of (watershed, stream network etc.). This technique for sharing precipitation data can be utilized for characterizing storm events that pass through specified areas. Relationships like Depth-Area-Duration, Probable-Maximum-Precipitation, and Intensity-Duration-Frequency curves can all derived using this data sharing technique.

5.2 FURTHER RESEARCH AND RECOMMENDATIONS

Although there have been significant advancements made in sharing scientific information readily through the web, improvements still need to be made if scientists are to utilize these systems for accessing data efficiently. Currently, one of the most promising tools available for sharing hydrometeorological data on grids is the THREDDS server. The THREDDS server organizes information in folders that specify the time period for which data is being provided. In addition, the THREDDS server allows for data to be accessed in a standardized manner regardless of the format the underlying data is being stored in. The most versatile and robust method of accessing information on grids is netCDF. NetCDF allows numerous grids to be stored in one file as a multi-dimensional array and can be easily visualized in analytical software applications like GIS.

One of the common uses of gridded data is in hydrologic modeling. With the increasing concern for climate change, hydrologic models are being improved and developed. One of the many inputs to hydrologic models, is gridded hydrometeorological information that has been spatially and temporally averaged (i.e. precipitation, evaporation, humidity etc.) Currently, there is no framework in place that can facilitate the collection of data in this format. While, the THREDDS server, netCDF and GIS provide the tools necessary to make data collection for hydrologic modeling efficient, there is yet to be a system that integrates all of these tools together.

What one can envision as solution for the future, is a system that allows a data user to navigate through multiple different hydrometeorological data sources, select the appropriate variables of interest over a specified geographic location (i.e. a group of watersheds, catchments etc.) and then automatically create subsets of data that are either spatially averaged (i.e. arithmetic average, weighted average etc.) and/or temporal averaged (daily mean, monthly mean etc.) for the area and time of interest. This currently cannot be accomplished because gridded datasets on the THREDDS server lack a common ontology (e.g. Climate and Forecast Conventions) and data structure. For example, one dataset will have precipitation data, called precip, organized by hour so that the grids are only accessible in netCDF format one hour at a time while another dataset will have precipitation data, called prate, organized by year so that multiple grids are accessible in netCDF format. Nevertheless, when data is ingested in GIS, grids need to be manually extracted one by one so that they can be visualized and analyzed. This process is currently very inefficient for compiling and aggregating gridded data.

In addition to providing an improved framework for working with gridded data, improved methods for sharing data that is variable in space and time can be very beneficial in hydrologic design. Variables like precipitation can be organized as events and shared as GIS features that contain all the necessary information to describe the event in complete detail. Because processing gridded data can be computationally time consuming, it may be beneficial to include "Analysis" servers as part of the HMIS to complete these very tasks. This would also allow for events to be stored in a database that can be queried by geographic location, temporal extent, area, intensity etc.

This thesis has provided some insight and prompted some discussion for how to improve the sharing of scientific information, specifically hydrometeorological data, utilizing the extended capabilities of the internet.

Appendix A: Addendum to Exercise 5

The following addendum is an extension to Exercise 5 in the GIS and Water Resources class taught by Dr. David Maidment at the University of Texas-Austin during the Fall of 2009. This exercise demonstrates the utility of using spatiotemporal tools to aid in the representation of data that is variable in both space and time.

Automating Zonal Statistics for NEXRAD Precipitation over a Specified Feature Layer

GIS in Water Resources
Fall 2009

Prepared by Fernando R. Salas

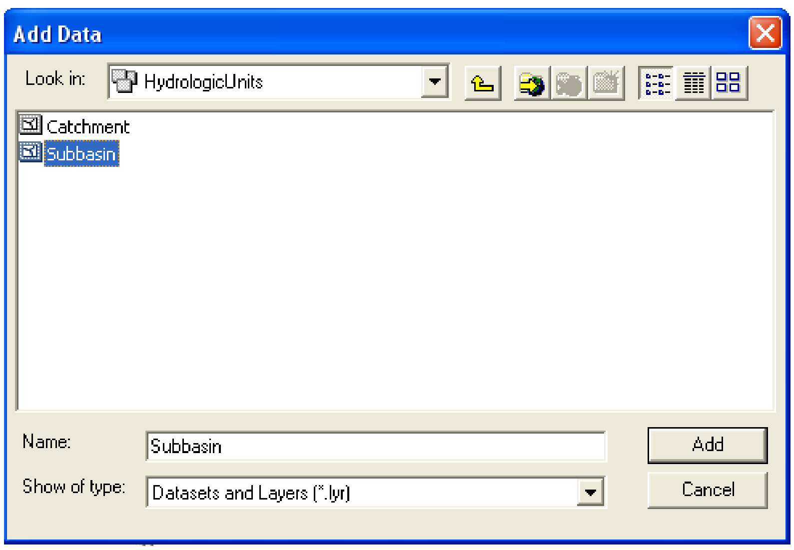
In **Exercise 5**, we used the **Animation** tool in conjunction with the **ZonalStatistics** tool to create multiple tables (one for each TimeStamp in the NEXRAD netcdf layer) which contained some basic statistical information over the San Marcos basin. To speed things up, we have created a tool which automates this very process. This tool has been generalized so that one can compute the **ZonalStatistics** of any NEXRAD Precipitation netcdf file given a specified feature layer. Below are step by step instructions on how to perform this operation specifically the operation performed in Exercise 5.

First, **download** the Ex5Addendum.zip file from the class website.
http://www.ce.utexas.edu/prof/maidment/giswr2009/giswr2009/Ex5/Ex5Addendum.zip

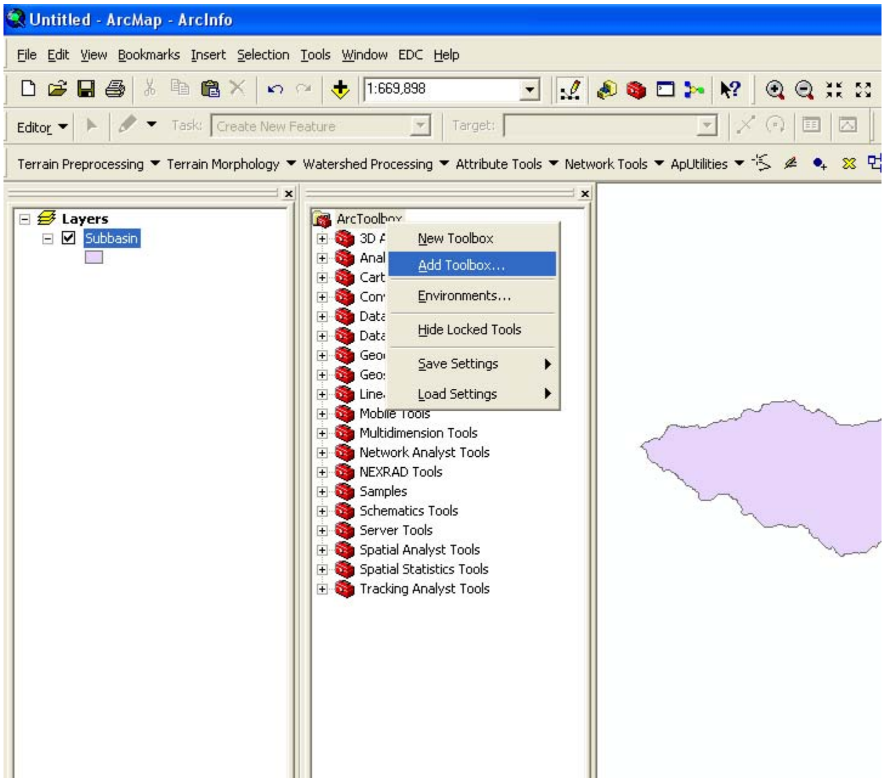
Extract the zip file to a folder of your choosing. Inside the zip file you will find an ArcMap Toolbox and a Python script file. Python is a general purpose programming language which ArcMap can compile and run. Python is very helpful in ArcMap because one can automate tedious tasks such as the one performed in Exercise 5.

Note: It is possible to use **Model Builder** to automate tasks in ArcMap as well however sometimes building a model can be just as tedious as the task itself.

Open ArcMap and **Add** the SanMarcos basin from the **Ex5.zip** file to a **new** map document.

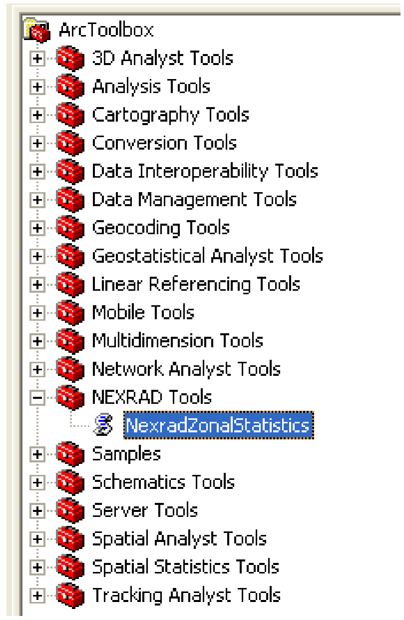


Next rick-click on ArcToolbox and click on AddToolbox… Navigate to the unzipped Ex5Addendum folder, highlight the toolbox and click open.

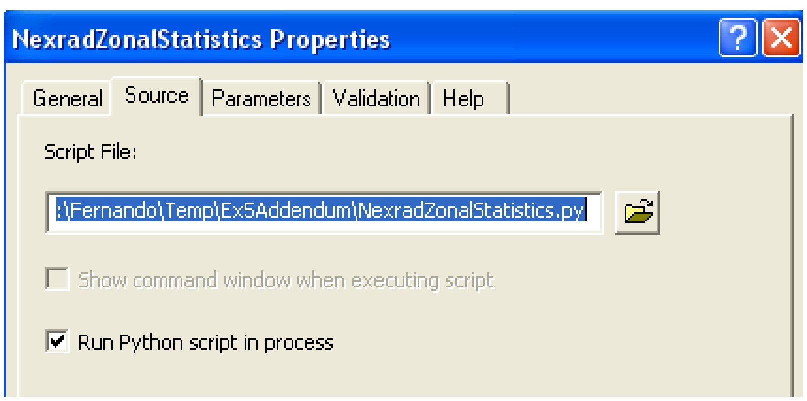


Now **add** the ST4_Agg_2002-TEST.nc file from the Ex5.zip file using the same procedure indicated in Exercise 5 (i.e. Multidimension Tools → Make Netcdf Raster Layer).

Expand the NEXRAD Tools toolbox and **double-click** on **NexradZonalStatistics**.



Right-click on the **NexradZonalStatistics** script and **click** on **Properties**. **Go** to the **source** tab and make sure that the "Script File" being referenced is the script file contained within the Ex5Addendum.zip file.



Click OK and now **double-click** on the **NexradZonalStatistics** script. **Fill** the inputs as seen below.

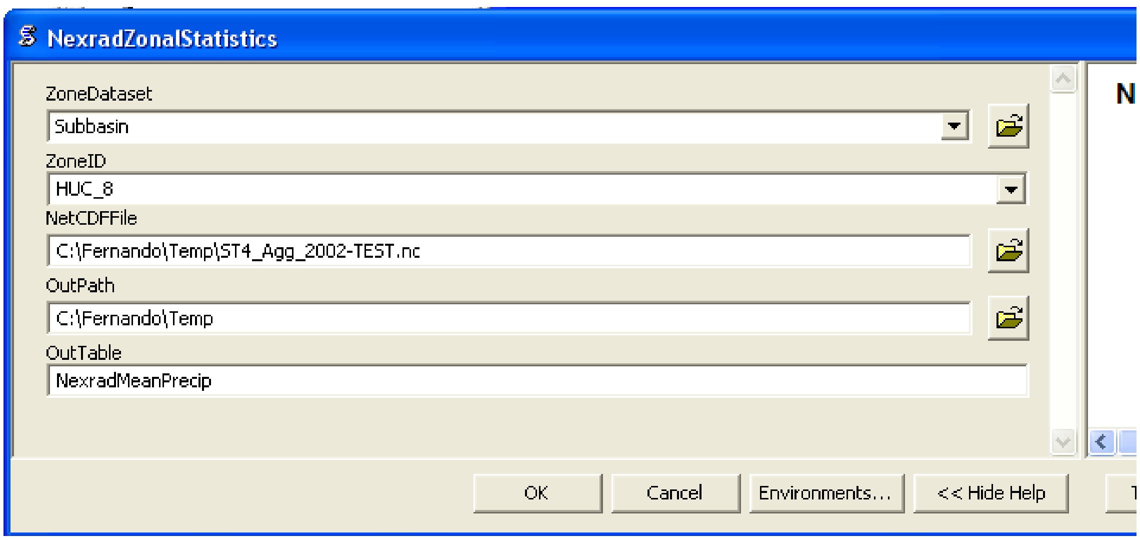
ZoneDataset: Feature layer to perform zonal statistics on.

ZoneID: Feature ID to attach to the output table for feature reference.

NetCDFFile: Netcdf file used for zonal statistics.

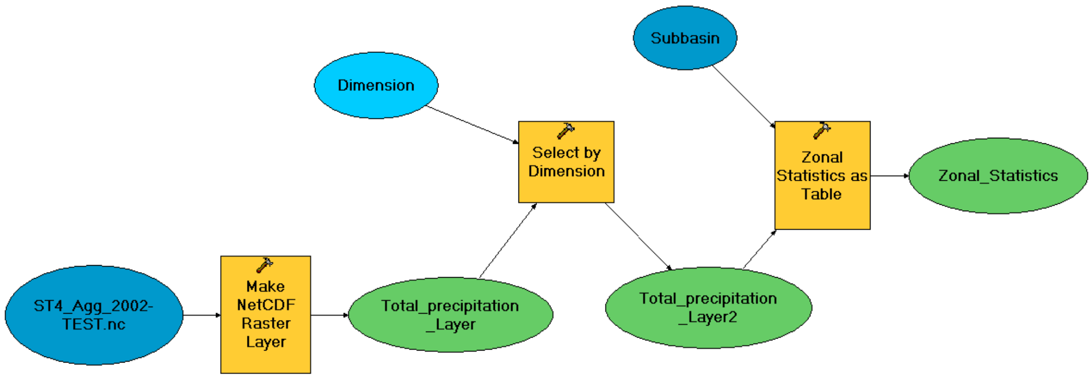
OutPath: Folder in which to save output table.

OutTable: Name of output table.



Click OK to run the script.

The script will run and output a DBF file containing the zonal statistics for each polygon contained in the feature layer at each timestamp within the netcdf file. Below is the basic layout/flowchart of the script being run as it would be displayed in model builder. The actual script has some modifications in it which are not seen in the model below.



Note: It is not necessary to add the NEXRAD netcdf file to the map in order to run the python script however it is good to do this anyway so you can visualize the data contained within the netcdf file.

Below is the output table you should get after running the script. You can display it either in ArcMap by adding the table to your map or opening the output file in Excel.

Attributes of NexradMeanPrecip

OID	HUC_8	ZONE_CODE	COUNT	AREA	MIN	MAX	RANGE	MEAN	STD	SUM	TIMEINDEX	TIME
0	12100203	1	246	5578.4502	0	6.1	6.1	0.077642	0.486263	19.1	0	6/28/2002
1	12100203	0	246	5578.4502	0	53.5	53.5	3.60976	8.4997	888	1	6/29/2002
2	12100203	0	246	5578.4502	30.299999	148.39999	118.1	75.733704	26.021	18630.5	2	6/30/2002
3	12100203	0	246	5578.4502	2.6	268.29999	265.70001	49.306499	53.816399	12129.4	3	7/1/2002
4	12100203	0	246	5578.4502	17	205.89999	188.89999	79.5728	39.136101	19574.9	4	7/2/2002
5	12100203	0	246	5578.4502	26.799999	247.89999	221.10001	72.980103	44.377201	17953.1	5	7/3/2002
6	12100203	0	246	5578.4502	0	129.8	129.8	16.698	27.226601	4107.7002	6	7/4/2002
7	12100203	0	246	5578.4502	1.5	147.8	146.3	36.355701	38.214199	8943.5	7	7/5/2002
8	12100203	0	246	5578.4502	0	32.799999	32.799999	7.94756	7.40609	1955.1	8	7/6/2002
9	12100203	0	246	5578.4502	0	60.299999	60.299999	3.67846	9.49665	904.90002	9	7/7/2002
10	12100203	0	246	5578.4502	0	42.099998	42.099998	4.37846	7.43229	1077.1	10	7/8/2002
11	12100203	0	246	5578.4502	0	23.5	23.5	2.19146	3.37493	539.09998	11	7/9/2002
12	12100203	0	246	5578.4502	0	18.299999	18.299999	2.30935	2.6192	568.09998	12	7/10/2002
13	12100203	0	246	5578.4502	0	8.4	8.4	0.146748	0.686787	36.099998	13	7/11/2002
14	12100203	0	246	5578.4502	0	26.9	26.9	0.948374	3.40475	233.3	14	7/12/2002
15	12100203	0	246	5578.4502	0	21	21	1.57073	3.68109	386.39999	15	7/13/2002
16	12100203	0	246	5578.4502	0	47.099998	47.099998	10.4577	9.47836	2572.6001	16	7/14/2002
17	12100203	0	246	5578.4502	0	52.400002	52.400002	8.7561	10.4534	2154	17	7/15/2002
18	12100203	0	246	5578.4502	0.3	39.299999	39	12.7276	8.84541	3131	18	7/16/2002
19	12100203	0	246	5578.4502	3.4	77.300003	73.900002	22.8211	11.4863	5614	19	7/17/2002
20	12100203	0	246	5578.4502	0	49.599998	49.599998	14.8252	9.49944	3647	20	7/18/2002
21	12100203	0	246	5578.4502	0	10.6	10.6	0.991057	1.90928	243.8	21	7/19/2002
22	12100203	0	246	5578.4502	0	0	0	0	0	0	22	7/20/2002

Record: 1 Show: All Selected Records (0 out of 23 Selected) Options

References

A Comparison of NEXRAD WSR-88D Radar Estimates of Rain Accumulation with Gauge Measurements for High- and Low-Reflectivity Horizontal Gradient Precipitation Events. **Gerrard E. Klazura, Jessica M. Thomale, D. Scott Kelly, Paul Jendrowski. 1999.** s.l. : Journal of Atmospheric and Oceanic Technology, 1999, Vol. 16.

A First Approach to Web Services for National Water Information System. **Jonathan L. Goodall, Jeffery S. Horsburgh, Timothy L. Whiteaker, David R. Maidment, Ilys Zaslavsky. 2008.** s.l. : Environmental Modeling & Software, 2008, Vol. 23.

A Spatiotemporal Data Model for River Basin-Scale Hydrologic Systems. **Jonathon L. Goodall, David R. Maidment. 2008.** 2, s.l. : International Journal of Geographical Information Science, 2008, Vol. 23.

Canada, Environment. Canada Network Climate Change Scenarios Quebec Region. *Data Access Integration.* [Online] [Cited: April 28, 2010.] http://loki.qc.ec.gc.ca/DAI/reanalysis-e.html.

CUAHSI. 2010. CUAHSI, Universities Allied for Water Research. *Mission.* [Online] 2010. http://www.cuahsi.org/.

CUAHSI Web Services for Ground Water Data Retrieval. **Whiteaker, Tim and To, Ernest. 2008.** s.l. : GroundWater, 2008, Vol. 46.

Domenico, Ben. 2009. *OGC Encoding Specification: CF-netCDF.* s.l. : Open Geospatial Consortium, 2009.

Donna F. Tucker, Xingong Li. *Characteristics of Warm Season Precipitating Storms in the Arkansas-Red River Basin.* Lawrence, Kansas : University of Kansas.

Francisco Olivera, Dongkyum Kim, Janghwoan Choi, Ming-Han Li. 2006. *Calculation of Areal Reduction Factors Using Nexrad Precipitation Estimates.* College Station, Texas : Texas Transportation Institute, 2006.

From the Semantic Web to social machines: A research challenge for AI on the World Wide Web. **Jim Hendler, Tim Berners-Lee. 2010.** s.l. : Artificial Intelligence, 2010, Vol. 174.

Integrating Water Resources Information Using GIS and The Web. **Jack Dangermond, David R. Maidment. 2010.** Orlando, Florida : AWRA 2010 Spring Specialy Conference, 2010.

Jantzen, Tyler. 2007. *Implementation of a State Hydrologic Information System.* Austin : Center for Research in Water Resources, 2007.

Lyons, H. G. 1906. *The Physiography of the River Nile and its Basin.* Cairo : National Printing Department, 1906.

Maidment, David R. 2009. *CUAHSI Hydrologic Information System: 2009 Status Report.* Center for Research in Water Resources, The University of Texas-Austin. Austin : CUAHSI, 2009. Status Report.

Maidment, David R. 2008. *CUAHSI Hydrologic Information System: Overview of Version 1.1.* Center for Research in Water Resources, The Unversity of Texas-Austin. Austin : CUAHSI, 2008. Status Report.

Maidment, David R. 2005. *Hydrologic Information System Status Report.* s.l. : CUAHSI, 2005.

Maidment, David, et al. 2009. Development of a Community Hydrologic Information System. s.l. : CUAHSI-HIS, 2009.

NASA. 2010. The Water Planet- Meteorological, Oceanographic and Hydrologic Applications of Remote Sensing. *Precipitation, Storms and Other Weather Phenomena.* [Online] 2010. [Cited: April 20, 2010.] http://rst.gsfc.nasa.gov/Sect14/Sect14_1d.html.

NOAA Satelite and Information Service. 2010. NCDC NEXRAD Data Inventory Search. *National Doppler Radar Sites.* [Online] March 2010. [Cited: March 30, 2010.] http://www.ncdc.noaa.gov/nexradinv/.

NWS. 2002. About Stage III Data. [Online] 2002. [Cited: April 20, 2010.] http://www.nws.noaa.gov/oh/hrl/dmip/stageiii_info.htm.

Performance Considerations for Mobile Web Services. **M. Tian, T. Voigt, T. Naumowicz, H. Ritter, J. Schiller. 2004.** Berlin : Computer Communications, 2004, Vol. 27.

PRISM. PRISM Climate Group. [Online] Oregon State University. [Cited: May 5, 2010.] http://www.prism.oregonstate.edu/.

Pukh Raj Rakhecha, Vijay P. Singh. 2009. *Applied Hydrometeorology.* New Delhi : Capital Publishing Company, 2009.

Representation of Spatial and Temporal Data in ArcGIS. **John L. Goodall, David R. Maidment, J. Sorenson. 2004.** Nashville, TN : AWRA GIS and Water Resources III Conference, 2004.

Siriwardena, L and Weinmann, P.E. 1996. *Derivation of Areal Reduction Factors for Design Rainfalls in Victoria.* s.l. : Cooperative Research Center for Catchment Hydrology, 1996.

Techniques for Methods of Collection, Database Management and Distribution of Agrometeorological Data. **P.C. Doraiswamy, P.A. Pasteris, K.C. Jones, R.P. Motha, P. Nejedlik. 1999.** s.l. : Agricultural and Forest Meteorology, 1999, Vol. 103.

Tomlinson, Edward. 2009. *Storm Precipitation Analysis System (SPAS).* 2009.

University, Iowa State. 2010. Healthy Land, Healthy Streams, Riparian Management Systems. *Photo Gallery.* [Online] 2010. http://www.buffer.forestry.iastate.edu/.

Using REST Web-Services Architecture for Distributed Simulation. **Khaldoon Al-Zoubi, Gabriel Wainer. 2009.** Washington DC : IEEE Computer Society, 2009.

Ven Te Chow, David R. Maidment, Larry W. Mays. 1988. *Applied Hydrology.* New York : McGraw-Hill, 1988.

Voormann, H. 2010. Web Services. *Wikipedia.* [Online] March 2010. [Cited: March 29, 2010.] http://en.wikipedia.org/wiki/Web_service.

Working with Temporal Data in ArcGIS. **David Kaiser, Morakot Pilouk. 2009.** San Diego : ESRI, 2009.

Vita

Fernando Renzo Salas was born in Fort Collins, Colorado to parents Jose Daniel Salas and Gabriela Maria Cecilia Salas on August 18th, 1986. He attended primary school in Fort Collins, Colorado until graduation from Fort Collins High School in 2004. After graduating high school, Fernando attended Cornell University where he graduated with a Bachelor of Science in Civil Engineering in 2008. Upon graduation from Cornell, Fernando attended the University of Texas at Austin with the intent of obtaining a Master of Science in Engineering in Environmental and Water Resources Engineering. Fernando has chosen to stay at the University of Texas at Austin to pursue a doctorate degree within the Environmental and Water Resources Engineering department. Throughout his career, Fernando has been an active member of the American Society of Civil Engineers and the Society of Hispanic Professional Engineers. In addition, Fernando has had the opportunity to travel the world extensively while living in Italy and Switzerland for extended periods of time.

Permanent address: 2721 Autumn Harvest Way, Fort Collins, Colorado 80528

This thesis was typed by the author.