# A Software Architecture for Distributed Geospatial Decision Support Systems*

Steve Goddard, Shifeng Zhang, and William J. Waltman
Computer Science & Engineering, University of Nebraska–Lincoln, Lincoln, NE
goddard@cse.unl.edu
shzhang@cse.unl.edu
wwaltman@unlnotes.unl.edu

Dennis Lytle
USDA National Food and Agriculture Council
Beltsville, MD
dennis.lytle@nfac.usda.gov

Seavey Anthony
USDA Risk Management Agency
Kansas City, MO
seavey_anthony@rm.fcic.usda.gov

**Project URL: http://nadss.unl.edu/**

## Abstract

*The National Agricultural Decision Support System (NADSS) is a joint project with the U.S. Department of Agriculture and the University of Nebraska-Lincoln. The initial focus of the NADSS project is to improve the quality and accessibility of drought related knowledge, information, and data for drought risk management. At the core of the NADSS project is a software architecture for distributed geospatial decision support systems that consists of four distributed layers: the data layer, the information layer, the knowledge layer, and the presentation layer. This work presents that architecture as well as the benefits and impact of the architecture and the NADSS project.*

## 1. Introduction

The United States Department of Agriculture (USDA) processes an enormous amount of data (both historical and current) related to all aspects of agriculture. The data are collected by many different agencies and stored in just as many different formats. Many of these organizations are beginning to explore the use of Geographic Information Systems (GIS) in their decision-making processes by generating maps that convey information gleaned from their respective databases. However, as they try to use data from other government or scientific agencies, they increasingly encounter problems accessing and interpreting the data. For example, the USDA Risk Management Agency (RMA) uses crop yield data from the USDA National Agricultural Statistics Service, crop-planting data from the 3000+ USDA Farm Service Agency field offices, and its own crop insurance policy and risk assessment databases to set crop insurance underwriting terms and policies. Even though USDA RMA employees understand the data collected from the other USDA agencies, they have tremendous difficulty in accessing and integrating the data in a manner consistent with their existing tools and processes. This is known as the *data-interoperability problem*. The USDA RMA would like to integrate climatic, topographic, geologic, and pedologic (soil survey) data in their spatial analysis of crop losses (indemnities) and risk assessment, as well as decision-making, but often the application of these large multidisciplinary databases requires a cadre of domain specialists to extract meaningful interpretations (assuming they could overcome the data-interoperability problem.) We call this problem, the *data-interpretation problem*, which commonly arises when data sources, resolution and temporal context, and scale become disconnected to the underlying people, places, and processes.

Unfortunately, the USDA RMA's problems are not unique. Data-interoperability is a well-known problem in the field of Geographic Information Sciences [11] and even the database community [1]. A primary goal of the Interoperability Program of the OpenGIS Consortium (OGC) [10] is to develop a standard, open geodata specification that promotes data interoperability. However, data availability and access is not enough. Few people know how to use data generated outside of their domain, even when they can access

* This work supported, in part, by grants from the National Science Foundation (EIA-0091530) and the Nebraska Research Initiative.

the data. This data-interpretation problem results in shallow application of available data and often leads to a misunderstanding of what the data can and cannot convey. We posit that the ultimate goal of geospatial analysis for USDA is to develop relationships explaining the landscape and cultural processes that describe agriculture and transform these relationships into knowledge that can be used to make sound decisions.

Solving the data-interoperability problem is a very important step to being able to access geospatial data. However, data interoperability is only one of many issues that need to be resolved in the area of geospatial decision support systems (GDSS). A GDSS is a collection of tools that can be used by analysts to *assist* in the decision-making process. In our view of a GDSS, tools are made available by domain experts to combine data into pieces of information that can lead to domain knowledge useable by non-experts. Thus, a GDSS is fundamental to solving the data-interpretation problem.

In this work, we present an extensible GDSS software architecture and demonstrate an instance of the architecture used by researchers, analysts, and public officials working on drought risk management, drought education, drought impact assessment, and drought vulnerability. The GDSS relies upon a suite of drought indices that can quantitatively describe intensity, duration, and magnitude of events and multiple windows of resolution—county-level to the community and farm scales. An important component to this project is the merging of the USDA RMA Policy Database, the NASS Historical Data and Census of Agriculture, NRCS State Soil Geographic Database, and Census2000 to derive new interpretations about rural communities and their potential impacts from drought events.

## 2. GDSS Architecture

We have developed a software architecture for distributed geospatial decision support systems that assumes data-interoperability is only one part of the problem in using geodata in the decision-making process. The decision-making process begins by combining and organizing data into pieces of information. Multiple pieces of information are then examined and combined to discover or create knowledge, which is the basis upon which a decision should be made.

The abstract, four-layer architecture of Figure 1 shows the software architecture we have created to support this view of the decision-making process. Each of the three lower layers (data, information, and knowledge) is associated with a cache for performance reasons. Strictly speaking, the cache is not needed. However, we have found that for interactive distributed systems, building the cache into the architecture provides performance benefits that outweigh the complexity it brings (e.g., cache coherency). Usually a layered architecture implies that each layer interacts only with adjacent layers. We have found, however, that higher layers occasionally need to access lower layers that are not adjacent. Thus, the large vertical interface arrow at the right of the figure is meant to

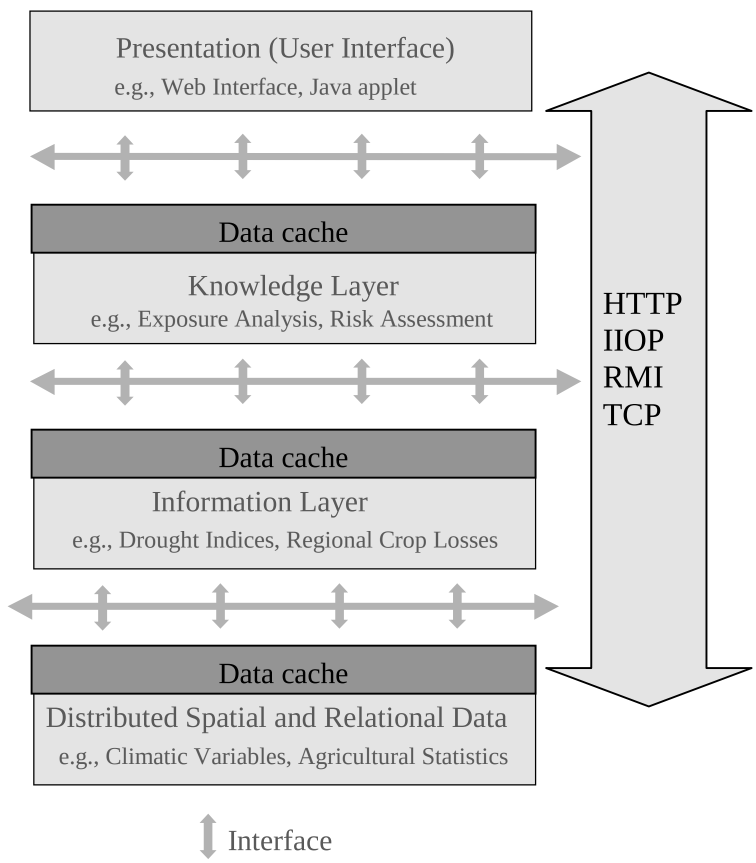



**Figure 1**. A four-layer software architecture for distributed geospatial decision support systems.

represent the ability of high-order layers to make requests to non-adjacent, low-order layers.

**Data Layer**. The data layer contains distributed spatial, constraint, and relational databases. The purpose of this layer is to provide transparent access to either local or remote data without concern for data formats. We do not attempt to solve the data-interoperability problem. Rather, we build on solutions to this problem by providing a mechanism to encapsulate existing solutions such as IBM's DiscoveryLink [6], CORBA-based or DCOM-based OGC objects, or data access via the Open Geographic Datastore Interface (OGDI).

The interface for this layer can be hosted locally or remotely. (Our current implementation is based on servers that can be accessed either via TCP/IP, RPC, or CORBA connections.) Assuming a cache exists, the local cache is checked once the data request is received. If the request is found, a reply is immediately sent. If the data request is not in the local cache (i.e., a cache miss), the data layer consults a meta-data repository to locate the data and the request is sent there. The reply is sent back to the cache and then forwarded back to the client. If a cache is not used, the reply can be sent back directly to the requesting client.

**Information Layer**. Data are combined and organized into information by tools in the information layer. Like each of the other layers in the architecture, the information layer is a distributed layer that may or may not have a local component beyond an information cache. The information layer is organized around a collection of domain-specific servers that aggregate data into information. Each request is either serviced via a local cache or sent to the appropriate server for processing.

Examples of servers in this layer are data interpolation servers and map servers, which may be either domain independent (e.g., spline interpolation) or domain specific (e.g., terrain regression). Depending on the domain, other servers can be added to this layer. For example, for the GDSS we are developing with the USDA RMA, we have developed drought index servers that process current and historical climate data from a weather station. The resulting index reflects how dry or wet a site is for a given period of time relative to its historical record. Thus the drought index is domain specific information developed from climate data.

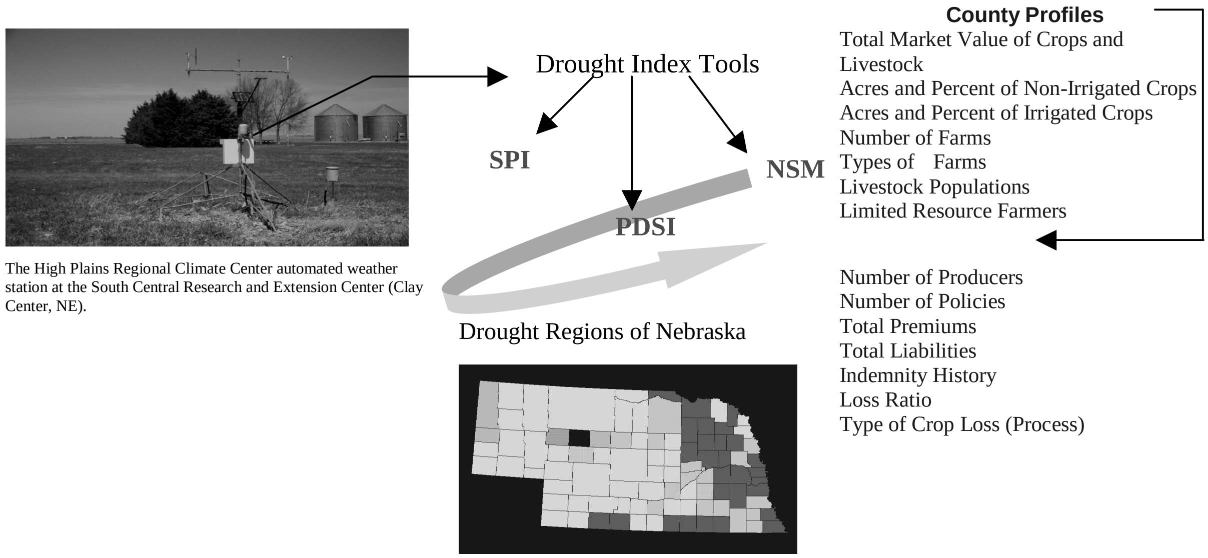
**Knowledge Layer**. Knowledge is created or discovered by combining information in new ways. Servers that provide or discover domain-specific knowledge are implemented in the knowledge layer. Examples might include data mining and knowledge discovery algorithms as well as simulation models. Servers at this level, might also provide more traditional domain-specific regression analysis of information (or data) generated (stored) at lower levels. The intent is that decision-makers will interact with this layer, via the User Presentation interface, to build and gather domain-specific knowledge. The tools in the knowledge layer do not make decisions, rather they contribute and organize knowledge that is used in the decision-making process. (An additional layer could be placed on top of the knowledge layer to create an expert system, but that is not the goal of this work.)

**Presentation Layer.** Decision-makers interact with the GDSS via the Presentation Layer. The user interface can take many forms and we are experimenting with both thin and thick clients. The simplest interface is developed using Web pages that interact with the lower layers via CGI requests. In the case of a Web interface, there must exist at least one Web server that provides an interface to the various GDSS layers and users must know the name of that server. We have also developed Java applets and CORBA clients that discover services via a CORBA Object Request Broker (ORB) by browsing a registry of services available.

## 3. National Agricultural Decision Support System

An instance of the architecture described in Section 2 is being developed in cooperation with the USDA as part of the National Agricultural Decision Support System (NADSS) project. The initial focus of the NADSS project is to improve the quality and accessibility of drought related knowledge, information, and spatial analysis for drought risk management.

**Figure 2.** The National Agricultural Decision Support System (NADSS) project combines data from weather stations and various geospatial databases to compute drought indices, risk assessment and exposure analysis.




Drought is a complex phenomenon that lacks a precise definition. Even though most people understand the concept of a drought and relate it to a deficit in precipitation over a period of time, it is often difficult to identify the beginning and ending of such an event. Moreover, a given quantity of precipitation over an interval of time may represent a drought at one location while the same quantity of precipitation over the same time interval may represent surplus moisture at another location. To assist in the analysis of droughts, climatologists have developed drought indices that attempt to describe the intensity, duration, and spatial extent of droughts. The drought indices rely upon the length of record at a location and quantify a drought in historical terms to derive drought interpretations (information). For example, a –3 Standardized Precipitation Index (SPI) [8] for an interval of time means that the amount of precipitation received by the site is three standard deviations below the mean of a gamma distribution of the site's historical record for the same interval of time. Thus, a drought index represents information developed from entire length of weather station record (climate data).

A set of drought indices derived for a population of weather station sites in a region for the same time period can be represented through a drought assessment map to quickly convey the spatial extent and severity of a drought. An example of this process is shown in Figure 2. Drought information can be analyzed and combined with historical crop yields, crop models, and economic information to expose the economic impact of a drought on a region or to determine the vulnerability of a region to drought. This knowledge may then be used to make decisions about disaster assistance to farmers or potential incoming or liabilities.

The NADSS software architecture is based on the four-layer architecture introduced in Section 2, as shown in Figure 3. Boxes with dashed lines represent components that are not yet implemented. The User Presentation layer is currently implemented with Web pages (http://nadss.unl.edu/). We are in the process of developing Java clients that conform to the OpenGiS Simple Feature specification to demonstrate interoperability with OpenGis tools. The Knowledge layer contains data mining tools that discover relationships between various drought, climate, and meteorological indices and crop yields [4,5]. (The knowledge tools are not yet available to the general public via the project Web page.) We are in the process of developing new tools that can evaluate and predict the impact of a drought. The Information layer currently contains three drought index programs (SPI [8], Palmer Drought Severity Index (PDSI) [12] and

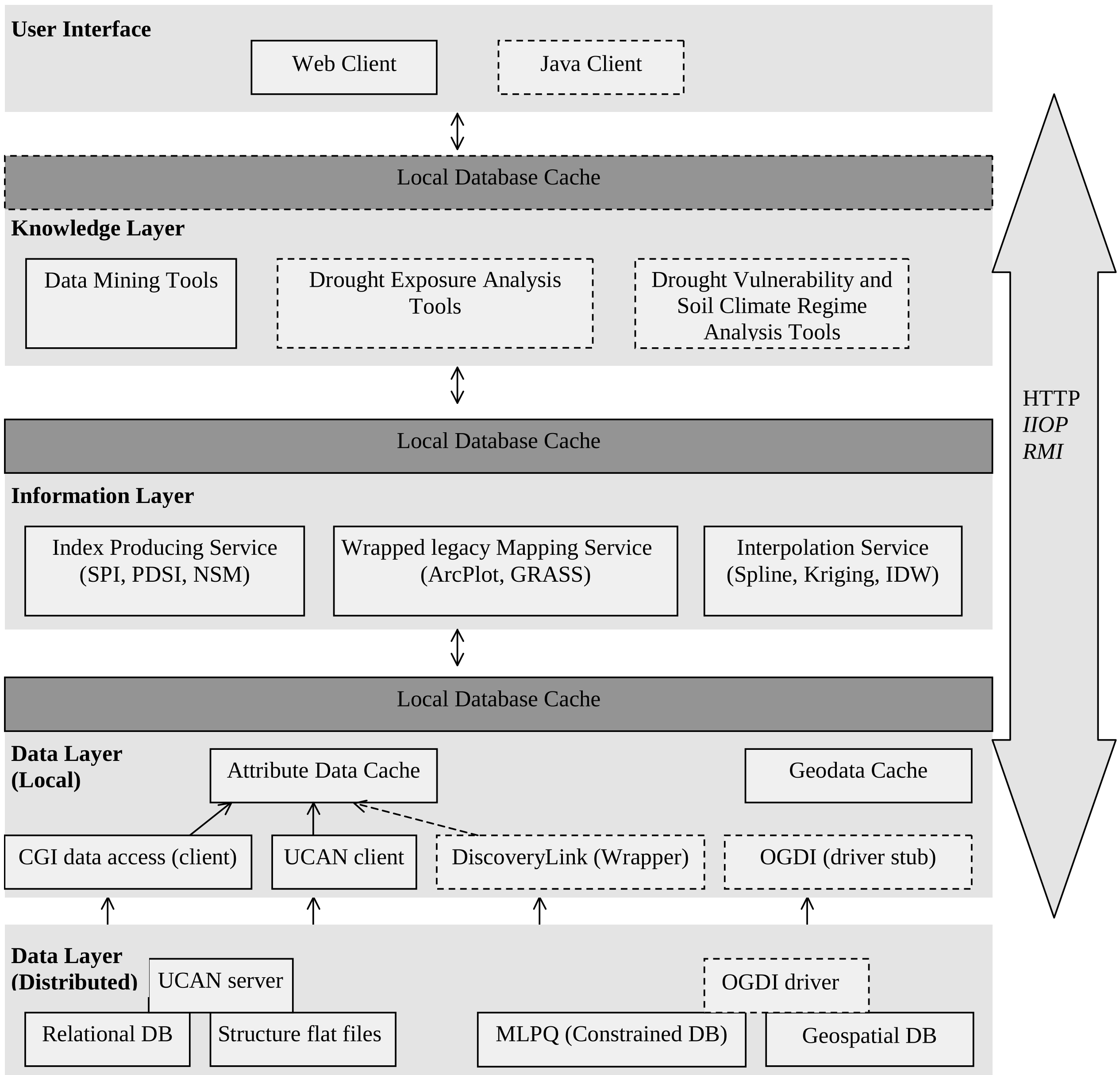



**Figure 3.** The four-layer distributed GDSS software architecture developed for the NADSS Project.

the Newhall Simulation Model (NSM) [9,14,15,16]), two map servers (ArcPlot [2] and GRASS [3]), and a data interpolation server. The data layer has been implemented as a local layer on the machine hosting the project Web site that caches recently accessed data and hosts clients capable of retrieving data from remote databases in the case of a cache miss. The remote databases conceptually form a distributed data layer that consists of relational databases and structured flat files accessed via a Unified Climate Access Network (UCAN) [13] server, a constrained database called MLPQ [7], and geo-spatial databases for ArcInfo and GRASS. We are working on developing an OGDI interface to the MLPQ and Geospatial database interfaces.

The interface between each layer should be open and well specified so that other organizations can either access the existing decision support tools or contribute new tools to the system. We are developing an omniORB communication backbone and Java 2 clients to reduce complexity and to improve interoperability with the OpenGIS and UCAN communities.

## 4. Impact and Benefits

Given the complexity of drought, where the impacts from a drought can accumulate gradually over time and vary widely across many sectors [17], a well-designed decision support system is critical to effectively respond to drought events. In many cases, the linkage between drought indicators and potential impacts have not been well developed because of the complexity of databases involved (data-interoperability problems) and the difficulty of building interdisciplinary systems that can address agricultural infrastructure, climatology, and the human dimensions of drought (data-interpretation problems).

We are developing a comprehensive geospatial decision support system for drought monitoring and mitigation that government policy makers, local communities, farmers/ranchers, and natural resource managers can readily access for current conditions and historical contexts tailored to specific localities. USDA program managers, local communities, natural resource managers, and farmers will be able to access new decision support tools that provide greater temporal and spatial resolution without encountering data-interoperability or data-interpretation problems. Thus, tribal, local, state, and federal agencies will be able to:

- respond to drought events more effectively, clearly illustrating the spatial extent of drought, and its severity,
- provide historical comparisons of the current drought event to past events, farm/community behavior and impacts, and
- develop more quantitative and objective risk assessments involving drought.

In addition, the USDA National Agricultural Statistics Service and Risk Management Agency's county-level databases were coupled with climatic characteristics to derive new relationships for estimating crop yields and identifying growing environments favorable to corn, soybeans, sorghum, and wheat. These geospatial databases can be used to characterize shifts in growing environments through time. As part of the Exposure Analysis, thematic maps can be generated to describe the agricultural infrastructure at the county level to compare irrigated and nonirrigated yield trends, yield ratios (corn:soybeans) to identify favored environments, shifts in crop acreages reflecting past climatic events and improved genetics, and dominant "cause-of-loss" processes for specific crops. New agronomic and livestock metrics will be presented to develop county-level profiles and regions of similar agricultural behavior. Changes in the agroecology and behavior of farmers in Nebraska over time will be explored with respect to El Nino/La Nina events.

We are already identifying innovative drought mitigation strategies being coordinated across USDA agencies, Nebraska's Climate Assessment and Response Committee (CARC), and through UNL's Cooperative Extension Service's outreach. As the system becomes fully operational and more widely accessible, we anticipate the similar benefits to be seen nationwide.

## 5. Conclusion and Future Work

We conjecture that the software architecture being developed for the NADSS project is applicable to many other distributed geospatial decision support systems. However, its applicability remains an open question. It was developed because of our negative experiences in accessing and interpreting geospatial data from existing Web sites. An open architecture is needed that goes beyond addressing data-interoperability. The architecture must also promote processing interoperability: the goal of the OpenGIS Consortium (OGC). However, a decision support system must also help resolve the data-interpretation problem. Since there is clearly an overlap with the goals of the OGC and our goals for an open, distributed, geospatial decision support system, we are in the process of supporting OGC processing-interoperability by developing and implementing OGC conforming Simple Feature Interface specifications.

USDA staff, research scientists, extension educators, local government decision makers, agribusiness, and individual farmers are already taking advantage of the drought index tools available in the Information Layer of our GDSS architecture. In the next year, we plan to make available data mining, exposure analysis, and drought vulnerability tools in Knowledge Layer. We are also working to improve our implementation of the data layer to provide more transparency and to make available more data sources from locations across the nation.

## References


[1] Chomicki, J., and P.Z. Revesz, "Constraint-Based Interoperability of Spatiotemporal Databases," *Geoinformatica*, Vol. 3, No. 3, pp. 211-243, 1999.

[2] ESRI, http://www.esri.com/software/arcgis/arcinfo/overview.html, accessed March, 2002.

[3] Geographic Resources Analysis Support System (GRASS), http://www.geog.uni-hannover.de/grass/, accessed March, 2002.

[4] Harms, S.K., J. Deogun, J. Saquer, and T. Tadesse, "Discovering Representative Episodal Association Rules from Event Sequences using Frequent Closed Episode Sets and Event Constraints," *Proceedings 2001 IEEE International Conference on Data Mining (ICDM '01)*, San Jose, California, November 2001.

[5] Harms, S.K., S. Goddard, S. Reichenbach, W.J. Waltman, and T. Tadesse, "Data Mining in a Geospatial Support System for Drought Risk Management," *Proceedings of the First National Conference on Digital Government Research*, Los Angeles, California, pp. 9-16, May, 2001.

[6] Haas, L.M., P.M.Schwarz, P.Kodali, E.Kotlar, J.E.Rice, and W.C.Swope, "DiscoveryLink: A system for integrated access to life sciences data sources," *IBM SYSTEMS Journal*. Vol. 40, No. 2, 2001.

[7] Kanjamala, P., Revesz, P.Z., Wang, Y., "MLPQ/GIS: A Geographic Information System using Linear Constraint Databases," *Proc. Ninth International Conference on Management of Data*, Hyderabad, India, pp. 389-392, December 1998.

[8] McKee, T. B., N. J. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," *Preprints, 8th Conference on Applied Climatology*, 17-22 January, Anaheim, CA, pp. 179-184, 1993.

[9] Newhall, F., and C.R. Berdanier., "Calculation of soil moisture regimes from the climatic record," Soil Survey Investigations Report No. 46, National Soil Survey Center, Natural Resources Conservation Service, Lincoln, NE, 1996.

[10] Open GIS Consortium, Inc., http://www.opengis.org, accessed March, 2002.

[11] Open GIS Consortium, Inc., *The Open Geodata Interoperability Specification Part I: The OpenGIS Guide An Introduction to Interoperable Geoprocessing*, OGIS TC Document 96-001, Wayland, Massachusetts, USA, 1996.

[12] Palmer, W. C., "Meteorological Drought," Research Paper No. 45, U.S. Department of Commerce Weather Bureau, Washington, D.C., 1965.

[13] Robbins, K., Perot, C., "Unified Climate Access Network (UCAN)," *Sixth International Conference on Computers in Agriculture*, Cancun, Mexico. June, pp. 1135-1142, 1996.

[14] Van Wambeke, A., P. Hastings, and M. Tolomeo, Newhall Simulation Model--A BASIC Program for the IBM PC (DOS 2.0 or later). Dept. of Agronomy, Cornell University, Ithaca, NY, 1992.

[15] Waltman, W.J., E.J.Ciolkosz, M.J. Mausbach, M.D. Svoboda, D.A. Miller, and P.J. Kolb, "Soil climate regimes of Pennsylvania," Penn State Agricultural Experiment Station Bull. 873, The Pennsylvania State University, University Park, PA 16802, 1997.

[16] Waltman, W.J. , H.R. Sinclair, and S.W. Waltman, "Characterizing and mapping soil climate in the Northern Plains region," Climate Change Impacts on the Great Plains, Workshop Report 89-96, Colorado State University, Ft. Collins, CO, 1997.

[17] Wilhite, D.A., Drought as a Natural Hazard: Concepts and Definitions. In: *Drought: A Global Assessment*, Vol. 1, D. Wilhite, ed., Routledge, London, p 10, 2000.