Internet-Based Climate Analysis Software Applications

Phillip A. Pasteris* USDA-NRCS, National Water and Climate Center, Portland, Oregon

> Keith L. Eggleston and William Noon Northeast Regional Climate Center, Ithaca, New York

1. INTRODUCTION

There is an ever-growing reliance on Internetaccessible climate data and products to monitor drought, manage water resources, support agriculture, and determine energy strategies. Distributed climate databases linked to software applications allow users to generate specific products needed to dynamically and efficiently manage activities affected by climate without the burden of managing large, local databases. This paper describes an Internet-based system containing several software applications that process historical and near real-time climate information to produce a wide variety of user-oriented products.

2. RECOGNIZING THE NEED FOR TIMELY CLIMATE INFORMATION

During the past two years a variety of studies and hearings have been conducted to determine the current status of the United States Climate Program and options to guide its future. Distilling the broad spectrum of climate services into a clear definition proved illusive and challenging before a consensus could be achieved.

In "A Climate Services Vision - first steps toward the future," the National Research Council report stated that "after reviewing the various meanings used in the past, the National Research Council's Board on Atmospheric Sciences (BASC) took a broad view and defined climate services as the timely production and delivery of useful climate data, information and knowledge to decision makers." (The National Research Council, 2001),

Rapid and potentially extreme climate variations were a basis for congressional testimony before the Subcommittee on Environment, Technology and Standards Committee on Science, U.S. House of Representatives. Dr Eric Barron stated that "the demand for climate services is growing and we must anticipate that these services will expand into a variety of climate-related areas including human health, water availability and guality, air guality, agricultural forecasting, and the stewardship of ecosystems." (Barron, 2001).

The Drought Commission Report recognized the need for climate data and a product delivery system and established a goal to "improve collaboration among scientists and managers to enhance the effectiveness of observation networks, monitoring, prediction, information delivery, and applied research and to foster public understanding of and preparedness for drought."

The Commission also recommended that "the President should direct the appropriate federal agencies to develop an effective drought information delivery system such as the Unified Climate Access Network (UCAN) to communicate drought conditions and impacts to decision makers at the federal, regional, state, tribal, and local levels and to the private sector and general public." (Drought Policy Commission, 2000)

A common theme in all of these studies is the need to identify existing climate networks, organize the data into Internet-accessible databases and allow users to produce the climate products necessary to meet their needs.

3. CRITICAL ROLE OF CLIMATE IN NATURAL RESOURCE ACTIVITIES

3.1 The Role of Climate in Monitoring Drought

Drought monitoring requires reliable climate data to create reliable products for the end user. The end users come from many sectors, but, for all of their differences, they have one thing in common: all are affected by drought. In the past, existing networks did not meet the needs of states and other institutions, leading to the establishment of various networks to provide weather and climate data. Often, the mission of the agency operating the network was narrow and the budget insufficient to widen the purpose of the network.

To whatever form they have evolved, climate networks play an irreplaceable role in monitoring the three dimensional nature of drought in the U.S. The data are used to create a wide variety of products (e.g. Standardized Precipitation Index, Palmer

^{*} Corresponding author address: Phil Pasteris, USDA-NRCS, National Water and Climate 101 SW Main St., Suite 1600, Portland OR 97204-3224; e-mail: ppasteris@wcc.nrcs.usda.gov

Drought Severity Index, Surface Water Supply Index, Soil Moisture Indexes).

3.2 The Role of Climate Information in Agriculture

The United States Department of Agriculture (USDA, 1999) requires Internet access and software applications to process a variety of surface data, upper air data and satellite imagery in its cropweather monitoring and assessment work. Weather plays an important role in agricultural production, not only in a crop's development but also with respect to planting and harvesting activities. The type of crop, as well as the particular stage of plant development, has different weather requirements for optimum growth. Knowledge of these requirements is essential in crop-weather monitoring.

3.3 The Role of Climate Information in Water Resources

Up to 80% of the West's water comes from the seasonally variable snowpack that melts in the spring and summer. Competition and demand for this valuable resource--the lifeblood of this semi-arid and arid region--is growing at a dramatic pace along with western populations. These problems are increasing in other parts of the country as well. Water supply information plays a prominent role in processes designed to provide early-season forecasts that facilitate planning among a wide variety of water uses: urban, agricultural, industrial, hydropower, navigation, recreation, tourism, species preservation, and international treaty compliance.

All facets of water resource management and use are affected by climate information. For example, inland waterways serve as transportation corridors, and drought can reduce river levels to the point that barge traffic is eliminated (nearly 40% of all grain is transported by barge). Management of discharge from reservoirs may extend the barge season, but a reduced shipping season may force shippers to seek additional transport from the railroad and trucking industry, creating an opportunity and at the same time a major challenge associated with redirecting available railroad cars and/or trucks to meet the need.

3.4 The Role of Climate Information in Energy

Lower river flows lead to reduced hydroelectric power generation and in some cases insufficient water for cooling at nuclear power stations. Reservoir management can serve as a tool to maintain river flow, but drought duration and intensity may reduce the storage below the point where this approach can be used. Summer drought is often accompanied by high temperatures, which lead to greater air conditioning loads that further exacerbate the problem.

4. INTERNET ACCESS TO CLIMATE DATA AND PRODUCTS

The widespread adoption of Internet protocols and standards made the access to a wide variety of climatic information possible with the implementation of the Online Store and Climate Data Online systems at NCDC in 1998 (Lott, 2000). It is being expanded as other products are placed online through the NESDIS National Virtual Data System (NVDS). The system will include Unified Climate Access Network (UCAN) and Regional Climate Center (RCC) products.

UCAN (Robbins and Perot 1996; Pasteris, Reinhardt, Robbins 1997) is a collaborative effort between six Regional Climate Centers (RCC), the USDA World Agricultural Outlook Board, NRCS National Water and Climate Center (NWCC) and the USDA Forest Service to build a software interface to process climate data remotely through the Internet.

UCAN has been implemented at the Northeast Regional Climate Center (NRCC) as shown in Figure 1. Additional federal and state cooperators scheduled for the spring of 2002. Phased UCAN access for Tribal Nations, local officials, the public, and the private sector is scheduled for the remainder of 2002.

4.1 System Goals

UCAN's goal is to provide seamless web-based access to distributed archives of national and regional climatic data, information and sanctioned interactive software to produce user-specified products. Most of the data warehoused at these centers consists of national climate archives comprising more than 25,000 daily and 2,500 hourly observation sites, with numerous sites composed of observations dating back into the 1800s.

The UCAN data structures are designed to store data collected by agencies, states and others. Example observation networks include the Natural Resources Conservation Service (NRCS) SNOTEL (SNOw TELemetry) and SCAN (Soil Climate Analysis Network) systems, U.S. Forest Service RAWS (Remote Automated Weather Stations) system, National Weather Service ALERT (Automated Local Evaluation in Real Time) system and state MESONETS. Several thousand additional automated observation sites are estimated to be available through these networks. Many of which provide hourly observations.

UCAN design requirements can be encapsulated within several distinct functional components. These include: 1) climate datasets, 2) climate metadata, 3) CORBA (Common Object Request Broker Architecture) interfaces between components, 4) software to generate user-oriented products dynamically and 5) a flexible user interface.

4.2 Dataset Description

The climate data archived at the RCCs, USDA, NCDC and others form a unique set of climatic information for the entire United States. For the most part, these data can be described as time-series observations of climatic elements collected at spatially distinct locations. The basic observational elements include, but are not limited to, air temperature, dew point, relative humidity, wind speed, wind direction, horizontal visibility, occurrence of weather, cloud height and coverage, precipitation amount, snow depth, soil temperature, evaporation, and solar radiation. The actual elements contained within a dataset depend on the observational network to which a station belongs. Some of these time-series observational records began in the late 1800s and are still operational.

Datasets are usually categorized by observational networks and administering organizations. They are also categorized by the time resolution (hourly, daily, monthly, annual, etc.) over which an observed or derived data element is applicable. National-scale datasets are collected by national agencies such as the National Weather Service, U.S. Forest Service, U.S. Geological Survey, U.S. Corps of Engineers, and Bureau of Land Management. Regional and statelevel datasets are operated by state and federal agencies, universities, municipalities, private corporations, and even individuals.

Within the U.S. government, the National Atmospheric and Oceanic Administration (NOAA) National Environmental Satellite, Data, and Information Service (NESDIS) mission is to provide and ensure timely access to global environmental data from satellites and other sources to promote, protect, and enhance the Nation's economy, security, environment, and quality of life. To fulfill its responsibilities, NESDIS acquires and manages the Nation's operational environmental satellites, provides data and information services, and conducts related research.

Most of the national datasets are maintained and archived by NESDIS at the NCDC, but others at the national level are maintained, solely, by the managing agency. Indeed, many regional and state-level datasets are maintained, almost exclusively, by the managing agencies. The UCAN participants recognize the need to collaborate with these agencies to collect and manage regional and local datasets for use in end-user applications as part of a larger National Virtual Data System.

4.3 Data Storage Concepts

Operational data storage for the UCAN system at the RCCs and the NRCS utilizes either a netCDF (Network Common Data Form) file, a RDBMS

(Relational Database Management System), or proprietary systems that employ indexed binary files. Since different systems are employed to store similar datasets, UCAN is designed to operate using any underlying data storage structure. Low-level data access within the UCAN system is defined as a local implementation consideration. This is a design decision that allows for incorporation of legacy datasets as well as new data centers with proprietary data storage systems.

Most UCAN data centers use netCDF as their basic dataset storage methodology. The netCDF software was developed at the Unidata Program Center (UPC) in Boulder, Colorado, and consists of a library interface and machine-independent binary file format for array-oriented access to scientific data. It is perfectly suited for the storage of time-series climate data. The netCDF interface has been enhanced for the UCAN project by the addition of libraries that provide dynamic file compression and a high-level PYTHON language interface.

As mentioned previously, all datasets belong to a distributed data archive. As a primary organizational method, each RCC archives data for its own region of responsibility. Specialized datasets, such as the NRCS SNOTEL network, are maintained by the managing agency and may cross regional boundaries. However, this strategy of single-source data management is not conducive to a stable archive system. Hardware or software failures could cause a data center or network link to fail, resulting in loss of an entire portion of the data archive. Therefore, data storage redundancy is implemented such that each dataset is duplicated, at least once, at different data centers. Thus, barring a network-wide communication failure, data accessibility should be high. This concept is facilitated by using a redundant array of inexpensive servers.

4.4 UCAN Metadata Management

Over the span of a station's history, a site may experience relocation or instrument redefinition that requires documentation of station location, instrumentation changes, and observational activity or inactivity. Relocation alone accounts for more than 100,000 station descriptions for the national station archive. The RCCs also warehouse extensive climatic information for local, state, and federal climate networks that do not appear in the national archive, further increasing the magnitude of inventory information. Management of these data inventories is made possible using distributed relational database management systems (RDBMS) synchronized to a master RDBMS.

The climate data resources archived in the UCAN system consist of thousands of individual "stations" (data collection sites) with different instrumentation, data collection frequencies, reporting intervals, and

management agencies. Many of these stations have changed fundamental characteristics during a history spanning more than 100 years. These stations form the basis of a long-term climate data record used to design large-scale engineered systems, establish insurance rates for weather-related natural disasters, and assess the magnitude of global climate change. It should not be surprising that metadata is essential for proper interpretation of the climate information obtained from these stations. Furthermore, metadata is at the heart of the UCAN strategy for locating climate data to satisfy user requests for specific climate information and data products.

4.5 Climate Products Concepts

A product, as defined in UCAN, is any organized collection of climatic information returned to a user to satisfy a specific request or query. A product could be a tabular listing of raw observations, an analysis of recurrence intervals for flood-producing rainfall within a hydrologic region, a graphical depiction of daily average temperature for a specific location, or a single number that represents the maximum observed wind gust in Washington, D.C. Thus, products vary in complexity and may involve single or multiple datasets or simple or complex climatic analysis tools; they may even require interaction between multiple data centers.

Critical design considerations for all UCAN products include accuracy, consistency, and the ability to satisfy user information needs. Accuracy includes factors such as adherence to accepted practices for scientific analysis and correct parsing and interpretation of user inputs so that returned results reflect user-intended results. Consistency refers to the ability to accurately regenerate stored elements derived from observed elements, one-to-one mapping of a returned product to a unique user request, and exact correlation of results obtained from different data centers when given identical input requests. Finally, products must be designed with sufficient analysis and formatting options to satisfy user needs for climatic information.

These design requirements are met by defining UCAN Standard Products that are installed at each data center. The analysis routines that operate on the data archives must be the same at all centers to ensure consistency of results. The calculation modules (statistics, data quality rules, sorting, etc.) must also be consistent between products at a single data center to ensure consistency of inter-product results. Thus, a modular design philosophy using object-oriented programming (OOP) was adopted with tightly defined modularity and a software design philosophy that is centered on data elements.

4.6 Internet User Interface

The system is designed to support any Internet browser or system-level software to seek climate information or run an application. Specifically, the user interface to the UCAN system is designed as three separate entities: a forms-based HTML document, a custom graphical user interface (GUI), and an embedded programmatic interface. The GUI and programmatic interface have received less development attention, so far. A prototype GUI was built which allowed for selection of station groups and returned data into a JAVA mapping application. The programmatic interface has been specified but not implemented. It is intended to make network calls to distributed metadata and data resources from applications such as simulation and decision models. The general access method that most users will experience is the forms-based interface. We have chosen this as our primary interface because it is a recognizable metaphor for most users, it is a welltested methodology embedded in a graphical browser, and it is cost-effective.

A form-based interface offers other practical design advantages (Figures 2 and 3). They are easy to modify and can be enhanced with JAVA applets to increase functionality. Ease of configuration also makes it possible to customize the UCAN interface to meet the agency-specific needs for climate data access. For instance, an agency can define an interface that restricts access to a subset of UCAN information that satisfies specific agency needs. A simpler interface could be designed to reduce complex options, perform input error checking, restrict output formats to agency-wide standards, and reduce the overall costs associated with obtaining climate data. Since the interface is implemented on an Internet browser, it can be implemented and maintained by the local agency and linked to UCAN brokers (we'll mention these later) to complete the request.

Information passed from the form, GUI, or programmatic interface is formatted as name-value pairs. Names represent registered entities associated with climate data or UCAN product metadata. These name-value pairs are parsed by the UCAN broker and passed to appropriate routines for processing. Adherence to a standard name-value interface construction simplifies the implementation of the CORBA brokers.

4.7 Current and Scheduled Climate Products

All climate products undergo extensive testing to ensure scientifically sound results. The following climate products are available from the NRCC webpage:

Single station climate summaries

- Temperature, precipitation and snowfall observations, normals and records for a day.
- Daily temperature, precipitation and snowfall data for a month
- Daily degree day listing
- Monthly time series summary
- Precipitation summary and historical context
- Monthly and seasonal temperature and precipitation distributions

The following climate products have been evaluated and are scheduled for webpage implementation during March 2002:

Single station climate summaries

- Activity Planner (one day for a number of years)
- Frost Statistics
- Frost Summary Statistics
- Snow Statistics
- Threshold / Exceedance Summary
- Station Summary (General Climatology)
- Daily Lister
- Normals Lister
- Consecutive Days summary

Multi-station climate summaries

 Daily Date for a Time Interval for all stations in county or Climate Division

Interactively generated products are shown in Figures 4 through 6. Specific examples include monthly precipitation with summary statistics (Figure 4), last and first frost date analysis (Figure 5) and number of consecutive days analysis (Figure 6).

5. CONCLUSIONS

Internet technology has facilitated an explosion of opportunities to fulfill the need to acquire, manage and deliver climate information. The climate community has responded by developing robust tools that meet a critical part of the climate services framework described in the literature referenced by this paper. Collaborative efforts, as part of a larger National Virtual Data System, continue to ensure that NCDC and UCAN products will be integrated within the overall NVDS system for data, product and information access to meet agency needs and serve the public both efficiently and effectively.

6. REFERENCES

Barron, E., "Climate Science and Services Challenges for the National Oceanic and Atmospheric Administration," *Testimony Before the Subcommittee on Environment, Technology and Standards Committee on Science, U.S. House of Representatives*, May 9, 2001 Lott, Neal and Dee Dee Anders. NNDC Climate Data Online for Use in Research, Applications, and Education. Twelfth Conference on Applied Climatology. Pages 36-39. American Meteorological Society, May 8-11, 2000, Asheville, NC.

National Drought Policy Commission "Preparing for Drought in the 21st Century," U.S. Department of Agriculture's Office of Communications and the National Drought Policy Commission, May 2000

- National Research Council, "A Climate Services Vision - first Steps Toward the Future," *Board on Atmospheric Sciences and Climate*, National Academy Press, 2001
- Pasteris, P., Reinhardt, R., Robbins, K., Perot, C., 1997: "UCAN - Climate Information <u>Now</u> for the Next Century," *First Symposium on Integrated Observing Systems,* American Meteorological Society, Long Beach, California.

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

U.S. Department of Agriculture, "Operational Meteorological Data Requirements," USDA Interagency Report, March 1999



Figure 1 Current UCAN Internet Interface, Northeast Regional Climate Center http://climod.nrcc.cornell.edu

Monthly Time Series	
Idaho 💌 by name 💌 <u>Search help</u>	
Station Name: rich	
Starting year (0000 for beginning of record): 1999 Ending y	/ear (9999 for end of record): 9999
Analysis Selection Additional Selections	
C Äverage Maximum Temperature	Total Precipitation
⊂ Average Minimum Temperature	Greatest One-Day Precipitation
← Average Mean Temperature	 Total Snowfall
⊂ Highest Maximum Temperature	⊂ Greatest One-Day Snowfall
⊂ Lowest Minimum Temperature	C Maximum Snow Depth
 Heating Degree Days (Base 65) 	Growing Degree Days (Base 50)
← Cooling Degree Days (Base 65)	← Growing Degree Days (86/50 Method)
Type of Output: HTML Table Missing Value Code: -999 Update Search	▼
Coop ID Station Name Period of Record Select 107673 RICHFIELD 8/1/1948 - current Figure 2. Basic Forms-Based Interface fo	or Station Selection and Element Analysis

Monthly Time Series	
Idaho 💌 by name 💌 Search help	
Station Name: rich	
Starting year (0000 for beginning of record): 1999 Ending year (9999 for end of record): 9999	
Analysis Selection Previous Selections	
⊂ Number of days exceeding threshold: Max Temp 💽 ᠵ 🔽	
C Number of days between thresholds: MaxTemp ▼ >= ▼ and <= ▼	
C Lowest Maximum Temperature	
C Highest Minimum Temperature	
Greatest Daily Temperature Range	
C Total Evaporation	
⊂ Greatest One-Day Snow Water Equivalent	
Type of Output: HTML Table 💌 Missing Value Code: -999 💌	
Update Search	
Coop ID Station Name Period of Record select 107673 RICHFIELD 8/1/1948 - current	
Figure 3. Enhance Forms-Based Interface for Station Selection and Element Analysis	

Station: R] Latitude: 4	ICHFIEL 13.05 de	_D grees	Longitud	le: -114	.16 degr	St ees Eld	ate: I evation) • 4282	feet	ID: 10	7673		
Station per	iod of r	ecord:	08/01/	1948 - 0	3/09/2	002							
CLIMOD Pr	oduct:	Monthly	Time S	eries		Cr	eation	Time: (03/10/2	:002 17:	51 EST		
Element: Pi	recipita	tion	Units: i	nch		An	alysis:	Sum					
Max allowa	ble miss	ing day	s: 3										
Lowest Acc	eptable	Quality	y of Da	ta: Raw	data								
Yeor(s)	Jon	Feb	Mor	Apr	Mav	Jun	Jul	Αυα	Seo	Oct	Nov	Dec	Annual
1999	3.20a	2.48	0.36	0.74	1.36	0.69	0.00	0.00	0.00	0.27	0.43	0.59	10.12
2000	1.88	1.75	0.43	0.46	0.71	0.00	0.00	0.00	0.53	1.18	0.47	0.72	8.13
2001	0.58	0.46a	0.42	1.24	0.16	0.03	0.33	0.00	0.62	0.17	1.60a	2.89	8.50
2002	0.87c	0.11c	-999v	-999z	-999z	-999z	-999z	-999z	-999z	-999z	-999z	-999z	-999j
Mean	1.63	1.20	0.40	0.81	0.74	0.24	0.11	0.00	0.38	0.54	0.83	1.40	8.92
Median	1.38	1.11	0.42	0.74	0.71	0.03	0.00	0.00	0.53	0.27	0.47	0.72	8.50
Max value	3.20	2.48	0.43	1.24	1.36	0.69	0.33	0.00	0.62	1.18	1.60	2.89	10.12
Min value	0.58	0.11	0.36	0.46	0.16	0.00	0.00	0.00	0.00	0.17	0.43	0.59	8.13
# years	4	4	3	3	3	3	3	3	3	3	3	3	3
Flags:	_												
a = 1, b = 2, A = Accumu	, c = 3, Ilation ov	., or z = /er mor:	26 or m e than oi	ore miss ne dav, S	sing days 5 = Subs	in a moi equent	nth or r	nissing r	nonths i	n a year			

Long-term means based on columns. Thus, the sum (or average) of the monthly values may not equal the annual value.

Figure 4. Results from Data Query - Richfield, Idaho Monthly Precipitation

Frest Summary

Station: DE KALE	8			State: 1	L	1	b: 112223							
Latitude: 41.93 d	legrees Long	itude: -68	.78 degrees	Elevation	: 873 feet									
Station period of	record: 03/	01/1966 -	03/09/200	2										
CLIMOD Product: Frost Statistics Summary Spring/Fall Cutoff Date: August 1					Creation Time: 03/10/2002 17:57 EST									
					Requested years: 1971-2000 Distribution: Normal									
			Probat	ility of La	ter Date in	Spring The	an Indicated	4						
Temperature	Earliest	.90	.80	.70	.60	.50	.40	.30	.20	.10	Latest			
36	04/23	04/26	05/01	05/05	05/08	05/12	05/15	05/18	05/22	05/28	06/10			
32	03/28	04/11	04/17	04/21	04/25	04/28	05/02	05/06	05/10	05/16	05/27			
28	03/24	04/01	04/06	04/10	04/13	04/16	04/19	04/22	04/26	05/01	05/11			
24	03/13	03/23	03/28	04/01	04/04	04/07	04/10	04/13	04/16	04/22	05/07			
20	02/27	03/09	03/14	03/18	03/21	03/24	03/27	03/31	04/03	04/09	04/13			
ló	02/17	02/28	03/05	03/09	03/12	03/15	03/19	03/22	03/26	03/31	04/08			
			Proba	bility of Ec	rlier Date	in Fall Tha	n Indicated							
Temperature	Earliest	.10	.20	.30	.40	.50	.60	.70	.80	.90	Latest			
36	09/06	09/17	09/22	09/25	09/28	10/01	10/03	10/06	10/09	10/14	10/16			
32	09/20	09/25	09/30	10/03	10/06	10/09	10/11	10/14	10/18	10/22	11/01			
28	09/27	10/03	10/10	10/15	10/19	10/22	10/26	10/30	11/04	11/11	11/20			
24	10/08	10/18	10/23	10/27	10/31	11/03	11/06	11/09	11/13	11/18	1VZZ			
20	10/19	10/27	11/01	11/05	11/08	11/11	11/14	11/17	11/20	11/25	11/30			
16	10/30	11/05	11/11	11/15	11/19	11/22	11/26	11/30	12/04	12/10	12/22			

Station: DE KAL	.B		State: IL	ID: 112223
Latitude: 41.93	degrees	Longitude: -88.78 degrees	s Elevation: 873 fe	et
Station period o	of record: 03/01/1966	- 03/09/2002		
CLIMOD Produc	t: Thresholds		Creation Time: 0	3/10/2002 18:08 EST
Search Years: 1	965 - 2002		Season: October -	September
Search Element	: Maximum Temperature	e Unit : degree F	Threshold : >=90	Consecutive Days 9
11 straight days:	07/12/1983 - 07/22/1	1983 O days of missing data		
	07/28/1988 - 08/05/1	1988 Odays of missing data		