Techniques for methods of collection, database management and distribution of agrometeorological data

P.C. Doraiswamy¹,*, P.A. Pasteris², K.C. Jones³, R.P. Motha⁴, P. Nejedlik⁵

¹ United States Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA
² United States Department of Agriculture, Natural Resources Conservation Service, National Water and Climate Center, Portland, OR, USA
³ United States Department of Agriculture, World Agricultural Outlook Board, Joint Agricultural Weather Facility, Washington, DC, USA
⁴ Slovak Hydrometeorological Institute, Kosice, Slovak Republic

Received 8 October 1998; received in revised form 2 September 1999; accepted 7 September 1999

Abstract

The major concerns for the availability of climate and agrometeorological data as we move into the 21st century continue to be in areas of data collection and database management. New technologies have advanced our ability to address these issues but the solutions may require the commitment of resources that may yet be outside the reach of developing countries. This paper briefly reviews new and existing technologies in the areas of data collection methods with some emphasis on remote sensing methods. An example for a conceptual data base management system adapted by the US Department of Agriculture is presented as a framework for the acquisition, maintenance and distribution of climate and agrometeorological data for the 21st century. The access of data to the national and international community can be resolved with the standardization of data base management and electronic accessibility. Databases in the US have been upgraded for access through the World Wide Web, and include data archived from other countries that have co-operative programs with US institutions. The operational needs of national agricultural producers, researchers, and agricultural weather and crop forecasters are the focus of this paper although it also serves some of the needs of the international research community. © 2000 Published by Elsevier Science B.V. All rights reserved.

Keywords: Agrometeorological database; Climate data collection; Climate data management; UCAN; WMO

1. Introduction

The availability of a proper meteorological and agrometeorological data base is a major prerequisite for studying and managing the processes of agricultural and forest production. In an examination of the ‘Agrometeorological Needs and Perspectives in the 21st Century’, the building of a database of meteorological, phenological, soil and agronomic information is inevitably a major priority. The acquisition of pertinent climate and agrometeorological data, processing, quality control, archive, timely access and database management are important components that will make the information valuable and used in agricultural research and operational programs.

Agrometeorological observations are focused mostly to the important food-producing areas around the world. Because of wide territorial scope, the basic national databases that include meteorological, climatological and crop phenological data are typically
kept and maintained by the National Meteorological Services. The building of these databases usually are dependant on the availability of technical personnel, and the hardware and software resources at the local level. The recent development in personal computer technology, database software, and the adoption of advanced interpretative systems like Geographic Information System (GIS), have brought large data volumes closer to the end users. As the use of agrometeorological data depends on the defined purposes, the accuracy and the management of the data can be different and should be defined accordingly.

Great strides in the automation age have offered new ways of making data products available to the user community, particularly in industrialized countries. The introduction of electronic transfer of data files via Internet using File Protocol Transfer (FTP) and the World Wide Web (WWW) has advanced this information transfer process to a new level. The WWW allows users to access text, images and even sound files that are linked together electronically. The attributes of WWW include the flexible to handle a wide range of data presentation methods and the popularity to reach a large audience. The development and access to this type of electronic databases may currently be limited to industrialized countries.

The WWW interface handles the communication between a remote server, where the data files reside, and the user’s display capability. This interface is controlled by a series of scripts and programs, which execute commands to read, process, and display data products. The data can be stored in a self-describing file format. The scripts and programs to generate menus for the WWW interface can read the information that describes the content of the data file (i.e., metadata). This modernization of data observing, processing, archiving, and display have made available worldwide large quantities of environmental data. The increasing challenge is to manage these growing resources of data so that users can efficiently extract needed information for their specific applications.

The global climate community and specifically the agricultural climate community has great interest in incorporating these new information technologies into a systematic design for agrometeorological data management. The goal is to ensure that the climate and agrometeorological data bases needed by the agricultural sector are collected, quality controlled, archived, and disseminated in a timely manner. Any such design needs to facilitate the collection, organization, analysis, and dissemination of data from a national agricultural weather and climate station data network for global, national, regional, state, and local program management and policy decisions and for dissemination to the public for use in the preparation of value-added services.

A recent attempt has been made in the US to promote the creation of such an agricultural climate database system. The National Agricultural Weather Information System (NAWIS) was designed to be a partnership between National Weather Service (NWS), US Department of Agriculture (USDA), and other public and private sector users. The goal is to facilitate the collection, organization, analysis, and dissemination of data from a national agricultural weather and climate station data network. The principal users are Federal, State, and local program management, policy decision-makers, and the public for use in the preparation of value-added services. The goals of the program are to provide (a). A nationally coordinated agricultural weather information system, based on the participation of Federal and State agencies, colleges and universities, and the private sector and aimed at meeting domestic agricultural weather and climate information needs. (b). Data for research and education, aimed at improving the quality and quantity of weather and climate information available to agricultural producers, including research on data analysis of weather and climate, particularly as they affect ecosystems.

2. Conceptual database

The developing of a climate information delivery system requires a basic foundation with the ability to expand and adapt with future needs. The cornerstones and building blocks of a conceptual system are presented. The concept, shown in Fig. 1, relies on a series of sequential and parallel operations, all leading to a free flowing system of climate information. The heart of the design is the database from which all information is obtained, while the body is the functions and processes, working and interacting to address the operational needs of agricultural producers, researchers, and agrometeorological forecasting. In this concept,
the initial database is the essential building block for a diverse set of functions that depend on the user needs. There are a wide variety of users, some requiring only basic general information for their decisions, others demand more detailed and specific data products for further analysis.

The various elements in the conceptual design will be investigated in some depth, drawing on the experience and opportunities that now exist in the technologies of personal computers and the WWW. The intent of this paper is not to be a ‘how to’ cookbook, but rather a review of basic concepts of design. The concept of a database system for the 21st century is derived from USDA’s experiences in the set up of a Unified Climate Access Network (UCAN). The UCAN consists of a consortium of State and Federal organizations, which record, manage and archive meteorological data (Pasteris et al., 1997). The primary goal is to provide access to nationally, regionally and locally-managed climate data to a clientele of researchers, decision makers, and other users worldwide via the Internet. Each participating organization maintains a computer that serves its climate data to the UCAN. This ‘node’ functions in concert with other UCAN nodes to perform distributed searches, retrievals and analyses of requested sets of meteorological observations. The various UCAN data servers involved in the process of searching for, assembling, and distributing climate data sets to the public are coordinated through an efficient catalog of all meteorological observing stations from which data are archived and disseminated by the UCAN.

3. Data collection

3.1. Operational data needs of interest to agriculture

Agricultural weather and climate data systems are necessary to expedite generation of products, analyses and forecasts that affect agricultural cropping and management decisions, irrigation scheduling, commodity trading and markets, fire weather management and other preparedness for calamities, and ecosystem conservation and management. Basic data collection activities should ensure that information on weather and climate, and their impact on crops, livestock, water and soil resources, and forest management can be processed, analyzed and distributed in the most efficient and timely manner. WMO publications 100 and 134 (WMO, 1983, 1986) should serve as guides to the spatial and temporal array of parameters necessary for successful operations.

3.2. Methods of data observation

Agrometeorology as an interdisciplinary branch takes data and provides the information of relatively
wide scope. Agrometeorological station networks are designed to observe the data of meteorological and biological phenomena together with supplementary data as disasters and crop damages occur. Meteorological data can be considered as typically physical elements that can be measured with relatively high accuracy while other types of observation are in many cases much more subjective and vary with the observer. In collecting, managing and analyzing the data for agrometeorological purposes, the source of data and the methods of observation define their character and possible management. The number of parameters observed differ with the method of data collection.

The method of observation can be categorized into two major classes, manually observed and automated data collection stations. A third source for agrometeorological data that is gaining recognition for its complementary nature to the traditional methods is satellite remote sensing technology. An introduction to this topic is presented only as a source of data and no attempts will be made to prescribe a database system for remote sensing data.

3.2.1. Manually observed data

Manually observed data mainly form the basis of agrometeorological information not only in the planning but also in the operative use. While the system of manually observed meteorological data collection is well established and standardized, the definitions of agrometeorological data are not so specific and the collection and control of these data are much less unique. The typical observations are air temperature, relative humidity, sunshine duration, wind run, precipitation and pan evaporation. Though the methodologies of agronomic data observation are quite well defined from the general point of view, the practice varies greatly even on the regional level based on variety and species. Observations are generally dates for planting, a few key stages of crop development and harvest. Plant height and crop water stress conditions may be observed for selective crops and situations. Following are some basic considerations for agronomic measurements:

- Observations are from a plant community and not single plants.
- There are much more descriptive elements in the observation.
- The frequency of the observation is different and can vary during the growing season.
- In Europe (WMO, 1988), up to 40% of collected phenological data and used it operationally.
- 30% of countries collecting phenological data do not include it data in their database.

The collection of manually observed data is highly time consuming with transmission done by a variety of methods and frequencies. But, as stated in WMO (1997), in many countries the main mode of transmitting data from observing sites to the national climate data centers is by radiotelephones supplemented by postal returns at designed time intervals. It is a slow system, but the indications are that it will remain in place for a long time to come. These data are then typically entered into a computer data file. The implementation of the climate computing (CLICOM) (http://www.gsf.de/UNEP/wcdmp.html) project by WMO has afforded many countries a much higher level of data processing and data control, however, the CLICOM system has not been accepted as a universal standard and many countries have adopted other integrated systems of data processing and archiving. The type and data collection method then, plays the decisive role in data processing and management.

3.2.2. Automatic weather station data (AWS)

AWS is becoming the ideal method for collection of climate and agrometeorological data, and is a necessity for many real-time operational programs. The data are now automatically collected from electronic media and automatically transmitted over the radio or telephone systems. The initial cost is high to establish the network but in the long run it is a more efficient method for simplified data processing and archive. This technology continues to evolve and the decline in costs makes it within reach especially in developed countries. The maintenance of these stations requires more technical skills and can be centralized to minimize costs.

The AWS can collect data at a desired range of frequencies from minutes to 24 h averages, depending on the need and data storage capacity. The data logger can be programmed to provide daily summaries in addition to the regular acquisition. The number of parameters collected can be numerous, but the standard station may collect air temperature, relative humidity, wind speed, barometric pressure, solar radiation,
rainfall, soil temperature (multiple depths), pan evaporation, leaf wetness and soil moisture over a limited range. Manual and AWS data need to be archived at designated centers if possible. In the US, for example, data are archived at six regional data centers and at offices of the state climatologists.

3.2.3. Remotely sensed data

Remotely sensed data and AWS systems provide in many ways an enhanced and very feasible alternative to manual observation for a very short time span between data collection and transmission. In certain countries where only few stations are in operation as in northern Turkmenistan (Seitnazarov, 1999), remotely sensed data can improve information on crop conditions for an early warning system. We have over two decades of experience in the use of data acquired from geostationary and polar orbiting satellites than any other forms of remotely collected data. Geostationary orbiting environmental satellites (GOES) can measure frequent surface temperature, solar radiation (cloud cover) and relative rainfall. Access to the other geostationary satellites spanning the globe should be possible through international cooperation. Unlike the polar orbiting satellites, the GOES satellites can provide continuous monitoring of the Earth’s atmosphere and surface over a large region of the western hemisphere in a geosynchronous orbit. These satellites monitor potential severe weather conditions, such as tornadoes, flash floods, hail storms, and hurricanes. The GOES Imager instrument consisting of five channels ranging from the visible to the infrared wavelengths, has a ground resolutions of 1 and 4 km at nadir.

The GOES sounder carries 18 thermal infrared channels. Each of the GOES satellites scans predetermined areas of the earth from the Mid Pacific region to the eastern Atlantic region. During routine mode, observations are taken four times every hour, but when severe weather threatens, the GOES Imager is capable of 1 min interval scanning over a smaller area. A variety of products from the sounder and imager are created operationally. These products are archived at the National Climatic Data Center (NCDC) (http://www.ncdc.noaa.gov/ol/satellite/satellite products.html)

Polar orbiting satellites such as NOAA’s advanced very high resolution radiometer (AVHRR) sensor can collect daily surface temperature and vegetative growth and condition (relative crop water deficit). The ground resolution of observation is 1.1 km for NOAA AVHRR data covering most parts of the globe and is complete. Kogan (1994, 1995) developed the Global Vegetation Index (GVI) product, which provides information for the area between 75°N and 55°S and produces weekly assessment of (16 km resolution) relative changes in seasonal land cover by comparing the data with normal dynamics of vegetation. The magnitude of the GVI is calibrated to delineate drought areas when a major drought is in progress.

Doraissamy et al. (1997) demonstrated the use of the Normalized Difference Vegetation Index (NDVI) to monitor biweekly changes in crop growth and development at 1 km spatial resolution. In an operational mode, the spatial and temporal information derived from NOAA AVHRR data can provide an early warning to potential drought conditions. The integration of remote sensing data with ground-based climate data has been demonstrated for monitoring evapotranspiration and vegetation growth at watershed scales (Kustas et al., 1994; Maas and Doraissamy, 1996). Doraissamy et al. (1998) integrated remote sensing data with crop simulation models to monitor crop condition and yields at regional scales in the US. Moussa (1999) discusses an approach where NDVI is integrated with yield monitoring and forecasting systems in Niger.

There are other numerous remote sensing satellites that are in operation, but only systems that are directly pertinent and have been in use for agrometeorological observations are presented. These remote sensing methods of data collection are based on measurement of reflected (short-wave) and emitted (thermal) electromagnetic radiation from the surface of the earth. The data processing can be rigorous, however, most processing software are standardized for retrieval of basic information that is useful for agrometeorological applications. Proper calibration and atmospheric correction of imagery data is still a major concern in data processing especially for application in retrieval of crop growth parameters.

Some developing countries are beginning to set up their own basic remote sensing centers that should be able to handle the required level of processing. Regional centers are set up to service surrounding countries such as in the African continent. The Food and Agriculture Organization (FAO) has been very
instrumental in promoting and establishing these centers in developing countries. Programs such as the Famine Early Warning System (FEWS) sponsored by the US State Department rely on remotely sensed data as a surrogate for agrometeorological data when no ground data is available.

Newer technologies on the horizon such as those on NASA’s Tropical Rainfall Measuring Mission (TRMM) satellite is being investigated for their accuracy in estimating the magnitude and spatial distribution of rainfall. The TRMM is a joint space project with Japan. TRMM is designed to measure rainfall in the tropical belt. TRMM’s unique combination of sensor wavelengths, coverage, and resolving capabilities together with a low-altitude, non-sun-synchronous orbit provides a sampling capability that should yield monthly precipitation amounts to a reasonable accuracy over a 500 km x 500 km grid.

The NASA Goddard Distributive Active Archive Center (DAAC) has been archiving TRMM data since the November 1997 launch has been released to the public. The data are useful for forecast model research, disaster mitigation, climatological studies, agricultural predictions, and many other applications.

The need for supplementing the measurement of such a highly spatial variable parameter as rainfall was clearly recognized at the earliest conception of TRMM (Thiele, 1987). A corollary objective involves a substantial supporting research component to pursue process studies for better understanding and characterization of various rain systems and to improve surface-based measurements of rainfall by advancing radar and rain gauge technology. Characteristics of rainfall cover a wide range of temporal and spatial scales, affecting both in situ validation and remote sensing measurements.

Because of the high spatial and temporal variability of tropical rainfall intensity, estimation of rainfall patterns by in situ networks and remote sensing platforms require special attention to spatial and temporal statistical properties of rainfall. One of the goals of this mission is to measure rain rate intensity distributions, and spatial and temporal statistics of rainfall variability. TRMM is effective only in the tropical belt and may be the best solution in the long run. These methods have been investigated (Rosenfeld et al., 1993, 1995) and will continue to be evaluated by cooperators from around the world.

4. Data transmission

New technologies allow many entities to collect climate data, however, proper management, processing and database establishment of the information are the keys to success. In many cases, the data must be shared to order to ensure a representative spatial depiction of the climate for a specific area. Each network has internal protocols to observe and a database of the information collected. It is this processing of the information that requires significant coordination to ensure that stations are correctly identified and the data are properly associated with the station.

4.1. Manual data forwarding

The basic options for an observer to transmit climate information once it has been collected have expanded dramatically. The following techniques may be used, some very effectively, if funds are not available to invest in an automated climate station.

- Mail (postal service)
- Telephone (voice w/recorder and subsequent transcription)
- Touch-tone electronic entry directly to a data collection computer system
- Internet browser entry directly to a data collection computer system

The Internet option can provide the observer with the tools needed to interactively enter, quality control and view not only their observations, but from others who have reported their observations in both a text and graphic environment.

4.2. Automated data forwarding

Once data are collected by an automated climate station, a mixture of both old and new technologies can be accommodated for data transmission. The technologies are summarized as follows:

- Telephone (a climate station dials-out or is dialed-up by a computer system)
5. Data exchange formats

The electronic exchange of climate data has been made easier thanks to the high-speed computer networks of the Internet. However, the need for common exchange formats has grown more critical. A wide variety of exchange formats exists and has been adopted for use. This presentation focuses on a human and machine-readable format in wide use within the US as an example.

5.1. SHEF — standard hydrometeorological exchange format

The standard hydrometeorological exchange format (SHEF) (Pasteris et al., 1996) is a documented set of rules for coding of data in a form for both visual and computer recognition. It is designed specifically for real-time use and is not designed for historical or archival data transfer. All the critical elements for identification of data are covered. Station identifiers, parameter descriptors, time encoding conventions, unit and scale conventions, and comment fields are all part of the code.

SHEF was designed for interagency sharing of data, visual and machine readability, and compatibility with anticipated receiving databases. The widespread implementation of SHEF allows the same decoding software to process data from various agencies. New data sources can easily be added as they become available. The visual nature of SHEF allows users quickly to become familiar with it. SHEF fully qualifies the data so that a receiving database has all the necessary information to describe the data. Additional information about SHEF can be found at the following URL (http://hsp.nws.noaa.gov/hrl/shef/index.htm).


There are three formats that make up SHEF. Through the use of parameter code characters to identify the data, these three basic message formats have the flexibility to transmit a wide range of hydrometeorological information. The formats are as follows:

- .A — single station, multiple parameter
- .B — multiple station, multiple parameter, header driven
- .E — single station, single parameter, evenly spaced time series

5.2.1. SHEF .A format

The .A format is designed for the transmission of one or more hydrometeorological parameters observed at various times for a single station. The .A format consists of positional fields and the data string. The format can be used for stations that report several different types of hydrometeorological data or report data with uneven time spacing.

5.2.2. SHEF .B format

The .B format can be used for the transmission of one or more hydrometeorological parameters from several stations for which many or all of the parameters are the same and are observed at corresponding times. The .B format consists of three basic parts: header, body, and terminator. The header consists of the positional fields and the parameter control string. The body contains station identifiers and data with optional date/data overrides. The terminator ends the entire .B format message. The format is useful for a routine morning roundup of precipitation and river data for a group of stations.

The .B header provides all pertinent date-time and parameter code information needed to decode the data contained in the body of the message. The order of the parameter list is flexible and can vary from message to message. The data values in the body are associated with the order of the parameter codes supplied in the header line. Any parameter that can be reported in the .A format can be reported in the .B format; however, if
more than three stations are to be transmitted routinely, the .B format should be more convenient.

5.2.3. SHEF .E format
The .E format can be used for the transmission of several values of a single hydrometeorological parameter at an evenly spaced time interval for a single station. As shown further, the F format consists of positional fields and the data string. This format is useful in the transmission of any evenly spaced time series of observational, processed, or forecast data.

The .E format is very similar in structure to the .A format except for two minor differences. First, in order to avoid ambiguity, the .E format accommodates only one hydrometeorological parameter while the .A format can handle several. Second, the .E format requires the explicit specification of a time interval (incremental or decremental) between data values. SHEF has been adopted as a NWS standard transmission format and has also been accepted as a Office of Federal Coordinator for Meteorology (OFCM) standard for hydromet data transmission. A complete document describing SHEF can be found on the WWW at the following URL: http://hsp.nws.noaa.gov/hrl/shef/index.htm

6. Data management
6.1. Introduction
The management of climatic data in this electronic/information age has become easier, faster, and more efficient. Yet, management of the data is perhaps one of the most critical processes in any design concept. There are several key areas where attention needs to be directed as one considers their particular design requirements. These areas include data collection, data processing, quality control, archiving, data analysis and product generation, and product delivery as shown in Fig. 2. Establishing a database for remotely sensed data can be overwhelming and requires

---

**Fig. 2.** Conceptual design of data acquisition, processing, data base establishment and accessibility of data by users.
large disk storage capacity. Potential users are assumed to have access to personal computers that have 8mm tape storage that may be slow to retrieve. The read-write CD systems offer a better alternative for storage and access to the data. Volumes of remotely sensed and AWS data are so overwhelming that the collection and processing of these data should be fully automated. Potential users of remotely sensed data are fully dependent on these data management facilities for their studies, as relatively fewer users have the resources that allow for the analysis of remote sensed information as highlighted in WMO (1996).

6.2. Data processing requirements

The design of a data processing system is directed by the users needs such as the type of data, frequency of observation, source of the data, and the format it is delivered.

Obviously if the data are being delivered via high-speed satellite downlink the process is much different than that being manually entered. Yet both these extremes have a common data management requirement, they have to be accepted into the system. In other words, an information system has to have multiple data entry points, requiring multiple user/machine interfaces, processing multiple types of information.

The PC environment (or the networked workstation) offers flexible machine/human interfaces while the WWW adds dynamic and creative solutions not restricted to local computer systems or databases. In any event, the information that enters a database maintained in a computer and the products retrieved by users are their responsibility. It is critical that a long-term relationship be established with both the data suppliers and product users to ensure that hydroclimatic information is of the highest quality to properly guide decisions made by natural resource decision managers. This activity requires constant attention to detail. The importance of the time devoted to basic data entry design cannot be overemphasized.

6.3. Quality control requirements

The essence of quality control is the assessment and improvement of imperfect information, by making use of other imperfect information. The following highlight some of the reasons for data quality control:

- To insure that the exact information needed is being properly generated.
- Inconsistent data need to be identified or will lead to incorrect decision inputs.
- For operational needs to detect problems that require immediate maintenance attention.
- For some purposes, continuous data are required, even if such data are estimated.
- Maintain an audit trail on data for future needs to update documentation.
- Data are often important for legal or design considerations.
- An established QC process attaches individual data values with flags that allow software to select data tolerance levels for product generation.

Quality control should be thought of as an end-to-end process, which is integral to the successful acquisition, temporary storage, permanent archival, and subsequent retrieval, use and interpretation of the data and information produced by a network.

Since there can be literally thousands of values to be examined (depending on a network’s size), there may be a need for automation and/or an intelligent expert system approach. Although human intervention is needed, good automated processes will reduce the need for intervention, and provide supplementary information to help the human make a decision. Examples of quality control screening of data are found in most all-domestic networks in the US. The NCDC (http://www.ncdc.noaa.gov) performs extensive quality control on the US NOAA cooperative network. These include both manual (double key entry of data manuscripts) and automated (temporal and spatial procedures). Many automated networks, including the NWS automated surface observation system (ASOS) (http://www.nws.noaa.gov/) also have built in QC filters. Whatever the process, record keeping is perhaps the most important. The establishment of QC flags for climate data is well documented for many networks. Individual datasets archived at a variety of climate centers describe the data flags used to identify the level of QC (http://www4.ncdc.noaa.gov/ol/documentlibrary/datasets.html).

Today’s technologies of electronic processing actually make the documentation problem less troublesome, and hence should not be avoided as you
develop your network QC process. The American society for agricultural engineers (ASAE) has published draft guidelines reviewing the measurement and reporting practices for automatic agricultural weather stations (Ley et al., 1994). It is an excellent source of technical standards for data management issues.

7. Metadata concepts

7.1. Station identification and history

In managing metadata it is critical to maintain all known names and aliases and identification numbers for stations. The accurate identification of data collection stations, an obviously fundamental component of metadata, is more complex than it first appears. Geographic ‘place’ names are the easiest and most user friendly means of station identification. The following station description elements (Lazar et al., 1999) will be incorporated into the UCAN, on line, relational metadatabase:

- Station identity (site name, aliases, and all identification numbers including network identification)
- Location (coordinates, elevation, geopolitical placement, topography, etc.)
- Equipment/instrumentation and its exposure
- Data observing and dissemination practices (network membership)
- Data inventory, and
- Temporal changes to any of this information.

The station history data will detail the changes in the station over time. This includes, but is not limited to, the location, naming and equipment. Such documentation include the following:

- Station location changes (latitude, longitude, and elevation), station name changes,
- Time of observation for each element and dates of any changes,
- Beginning and ending dates for each reported element,
- Addition or removal of element or sensor,
- For hourly precipitation stations, type of rain gauge,
- Type of recording equipment and dates of changes, and
- Observers names and dates of service.

7.2. Metadatabase management

The UCAN metadata database design employs a relational database, which incorporates the four rules discussed earlier in order to maintain consistency, insure flexibility, and yet accommodate the inconsistencies that are so common in climate station history information. The database design begins with a central table that defines a single station with a unique, internal station identification number. This content-free internal station identifier is purely for database management purposes (a primary key variable) and is not for distribution to data and metadata users.

Multiple station identifiers, names and aliases, networks and all other station history information are maintained in normalized tables linked by many relationships in the internal station (ID). Tables documenting station attributes that change through time includes time variables (begin and end dates). The station network table contains a specific type of station identifier that is associated with a network (for example, the NCDC TD3200 network uses the NWS Coop station number). The final consideration in this metadata management system solution is a rule for the depiction of climate data in report products.

7.3. Database management

A wide variety of database choices face the climate user community. The increased use of relational databases to store significant amounts of data provides a tempting opportunity to create a climate database that comes with a significant amount of database management software. These databases work well if you have a relatively small database, run the database in a central location and do not plan to exchange large amounts of data on a regular basis. There is usually a licensing agreement attached to using a commercial database, which can vary significantly based on the computer platform running the database.

8. NetCDF (Network common data form)

For the UCAN project, the climate databases would be very large (over one billion data values),
would be distributed among eight regional and federal climate centers, and would need to support rapid access and exchange of data files between users throughout the US. It was determined through prototype testing that NetCDF’s stored and retrieved time series information more efficiently than any other relational database available. This, coupled with the freely available software tools and ease in data exchange via the Internet, made NetCDF a logical choice. A full description of the entire UCAN Project can be found at the following URL, (http://www.srcc.lsu.edu/ucan.net/UCAN.html). The URL also supports a demonstration of the UCAN data access methodology.

NetCDF is an interface for array-oriented data access and a library that provides an implementation of the interface. The netCDF library also defines a machine-independent format for representing scientific data. Together, the interface, library, and format support the creation, access, and sharing of scientific data. The netCDF software was developed at the Unidata Program Center in Boulder, Colorado. The freely available source is at (ftp://ftp.unidata.ucar.edu/pub/netcdf/) or from other mirror sites. NetCDF, is software for storing and retrieving scientific data. More than a data format, the netCDF package is a set of programming interfaces that can be used with widely varying scientific data sets and by machines of widely varying architecture.

Multidimensional data may be accessed one point at a time, in cross sections, or all at once. Data are directly accessible, permitting efficient access to small subsets of large data sets.

8.1. Climate element identification

The variables and their associated dimensions are named. An example of the TD 3200 — summary of the day (includes TD 3210) are described in Table 1. Information about the data, such as what units are used and what the valid range of data values is, can be stored in attributes associated with each variable. The processing history of a data set can be stored with the data.

8.2. NetCDF transportability

The netCDF format renders netCDF files machine independent. The netCDF package is particularly useful at sites with a mix of computers connected by a network. Data stored on one computer may be read directly from another without explicit conversion. The netCDF software has been used successfully on a broad range of computers, from PCs to supercomputers. The netCDF library can be invoked from C, C++, FORTRAN, or PERL programming languages. Data stored using programs in one language may be retrieved with programs in another language.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name</th>
<th>Source</th>
<th>Units</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMAX</td>
<td>Daily maximum temperature</td>
<td>3200</td>
<td>°F</td>
<td></td>
</tr>
<tr>
<td>TMIN</td>
<td>Daily minimum temperature</td>
<td>3200</td>
<td>°F</td>
<td></td>
</tr>
<tr>
<td>TOBS</td>
<td>Temperature at observations time</td>
<td>3200</td>
<td>°F</td>
<td></td>
</tr>
<tr>
<td>PRCP</td>
<td>Daily precipitation</td>
<td>3200</td>
<td>0.1 in.</td>
<td></td>
</tr>
<tr>
<td>EVAP</td>
<td>Daily evaporation</td>
<td>3200</td>
<td>0.1 in.</td>
<td></td>
</tr>
<tr>
<td>MNPN</td>
<td>Daily minimum pan evaporation temperature</td>
<td>3200</td>
<td>°F</td>
<td></td>
</tr>
<tr>
<td>MXPN</td>
<td>Daily maximum pan evaporation temperature</td>
<td>3200</td>
<td>°F</td>
<td></td>
</tr>
<tr>
<td>SNOW</td>
<td>Daily snowfall</td>
<td>3200</td>
<td>0.1 in.</td>
<td></td>
</tr>
<tr>
<td>SNWD</td>
<td>Snow depth at observations time</td>
<td>3200</td>
<td>0.1 in.</td>
<td></td>
</tr>
<tr>
<td>WTEQ</td>
<td>Water equivalent of snow depth</td>
<td>3200</td>
<td>0.1 in.</td>
<td></td>
</tr>
<tr>
<td>WDMV</td>
<td>24 h wind movement</td>
<td>3200</td>
<td>Miles</td>
<td></td>
</tr>
<tr>
<td>SNtnm</td>
<td>Daily minimum soil temperature</td>
<td>3200</td>
<td>°F</td>
<td></td>
</tr>
<tr>
<td>SXtnm</td>
<td>Daily maximum soil temperature</td>
<td>3200</td>
<td>°F</td>
<td></td>
</tr>
<tr>
<td>SOtnm</td>
<td>Soil temperature at observations time</td>
<td>3200</td>
<td>°F</td>
<td></td>
</tr>
<tr>
<td>DYSW</td>
<td>Daily occurrence of weather</td>
<td>3200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Many groups and organizations have adopted netCDF for their scientific-data-access needs. Numerous freely available, licensed, and commercial packages for data analysis and visualization can display and manipulate netCDF data.

Unidata’s purpose in creating the netCDF library is to generalize access to scientific data so that the methods used for storing and accessing data are independent of the computer architecture and the applications being used. In addition, the library minimizes the fraction of development effort devoted to dealing with data formats.

Standardized data access facilitates the sharing of data. Since the netCDF package is quite general, a wide variety of analysis and display applications can use it. The netCDF library is suitable, for example, for use with satellite images, surface observations, upper-air soundings, and grids. By using the netCDF package, researchers in one academic discipline can access and use data generated in another discipline.

The Unidata netCDF package is available free of charge. It is also one component of a suite of software tools that Unidata distributes to universities. With these additional tools, universities can capture data, store them in netCDF files, and display and analyze them as desired. The netCDF package contains C-language source code for the netCDF data-access library; source code for the C++ and FORTRAN interfaces; documentation of the netCDF library and utilities; and test programs to verify the implementation.

Unidata provides software revisions and upgrades to all its software tools, including the netCDF package. The netCDF software and documentation may be obtained from the netCDF WWW site at (http://www.unidata.ucar.edu/packages/netcdf/)

9. Product generation and access

9.1. Formats and products

One of the exciting features of PCs is the ability to produce quality products, formatted for easy reading and presentation, generated through simple word processors, databases, or spread sheet applications. However, some thought needs to be given to what the product looks like and what it contains, before your database delivery design is finalized. The greatest difficulty many be encountered when producing a critical report and it is realized that key data or information attributes are missing, especially when climatic data are combined with agricultural, and human resource information from separate databases. WMO (1986) is a good source for data formats and distribution.

9.2. Graphics

The presentation of agrometeorological data in graphical form has also been made easier via today’s computer applications. While some of us may not care about a particular software package ‘graphic style’, it may be important to have access to a suitable package to graph information that can be cut and pasted in a report or presentations. A wide variety of software now exists to plot both time series and spatial information. Information about time series plotting can be found at (http://www.wrcc.sage.dri.edu) and information on spatial plotting can be found at (http://grads.iges.org/grads/head.htm). Fig. 3 is an example of a time series plot of NRCS SNOTEL data collected at a test site in Oregon, USA. An example of a spatial plot is shown in Fig. 4. Shades of color are used to show gradients in air temperature and also display contour lines for dew point temperature distribution.

9.3. Local processing versus remote processing

In general most computer systems now have the central processing unit (CPU) power to run a variety of applications locally. However, access to large databases and CPUs using Telnet software and the Internet can make remote processing very attractive. This is especially true if your data center does not have the database resources to support a spatially and temporally relevant database. Graphics and products shown previously can be run remotely and images and products downloaded to your computer via FTP. If you contribute observations to a central database, and have this computer access, you would benefit from the processing of information not collected by your agency, but available from the central facility. The Internet is clearly a new communications tool that must be considered when designing a database and climatic data processing operation.
Fig. 3. Time Series Plot of NRCS Snotel Data. Courtesy Western Regional Climate Center.

Fig. 4. Spatial plot of RAWS Temperature and Dew Point. Courtesy Western Region Climate Center using free GrADS software package. Temperature Field — Shaded: Station Temperatures (Red Values); Dew Point Field — Contoured.
9.4. Data and statistical computation

It is not our intent to cover the statistical techniques appropriate for application to climatic datasets which are thoroughly discussed in WMO (1983). Considerations in database and interface design requirements to aid in subsequent analysis are documented. Issues of missing data, extreme values, non-numerical values, or out-of-range values are common in data analysis. However, simple presentation of datasets to analytical software packages can be a challenge. New methods of data analyses have been developed using artificial neural network analysis techniques (Michaelides, 1999). These methods have been used for estimating missing data and completing time series and many other applications. Consider carefully your ability to extract precise datasets from your database and making them available to your software package.

Common formats acceptable to most statistical packages and/or spread sheets include: comma or tab delimited ASCII text file structures, LOTUS 8, Microsoft Excel 8, SAS 8, or Quattro Pro 8 file structures. The netCDF specialized structure and format has the potential of streamlining statistical data analysis and include the following:

- Raw daily lister
- Compact daily lister
- Monthly time series
- Degree day summaries
- Normal lister
- Generalized climatology
- Daily normals
- Daily percentiles
- Temperature summaries
- Frost-last and first frost dates
- Growing season length
- Extreme daily values for the month
- Monthly/annual model data
- Frequency distribution and probability.

10. Summary

The importance for climate and agrometeorological data in the operational needs of agricultural producers, researchers, and agricultural weather and crop forecasters are gaining attentions in the domestic and the international research community. Databases have been developed in many countries although there are still quite some countries where there is a need for establishing databases that are accessible for their citizens and the international community. Modernization of data acquisition, processing and archiving can be a costly proposition but steps need to be taken, to ensure that such data are archived in a standardized manner that can be useful.

This paper outlines the current methods of data collection and particularly introduces remote sensing techniques that may be a solution in countries that are considering upgrading their current system. Remote sensing data provides a spatial component of crop growth and development that cannot be collected in ground-based measurements.

Computer technologies are advancing rapidly, and therefore it is unrealistic to try and articulate all potential solutions to agrometeorological information systems for there are surely more technology solutions currently under development. It is hoped however, that the reader has been exposed to some of the key elements of what should be considered in designing and implementing a computerized climate information delivery system.

An example for a conceptual data base management system adapted by the US Department of Agriculture is presented as a framework for the acquisition, maintenance and distribution of climate and agrometeorological data for the 21st century. The access of data to the national and international community can be resolved with the standardization of data base management and electronic accessibility.

Databases in the US have been upgraded for access through the world wide web, and include data archived from other countries that have co-operative programs with US institutions. Perhaps the work currently being done with UCAN in the US Department of Agriculture can be expanded into the international climate community.

The inclusion of remotely sensed data in the climate and agrometeorological database will take a major effort but the operational and research community is currently using a number of pertinent parameters. The Technical Commission for Agricultural Meteorology of the WMO could move towards identifying suitable parameters and standardizing the format of remote sensing data in the climate and agrometeorological database.
Acknowledgements

Contributions for this chapter were received from S.C. Michaelides, L. Moussa and I. Seitnazarov is greatly appreciated. The authors wish to acknowledge the excellent technical review and recommendations provided by Professor C.J. Stigter, Dr. Richard Coe and the other reviewers in preparation of the manuscript.

References


