Vol. 3, No. 2 ● December 1992

EARTH SYSTEM MONITOR

NOAA workshop promotes agency-wide network integration

A guide to NOAA's data and information services

INSIDE

Linda Matthews Hood
Telecommunications and ADP Security Branch
Office of Administration
NOAA

In recent years elements within the National Oceanic and Atmospheric Administration (NOAA) adopted various networking solutions and strategies to meet their individual needs and requirements. Now efforts to achieve NOAA-wide network integration and development are progressing under the sponsorship of NOAA's Network Advisory and Review Board (NARB).

The NARB was established to provide high-level coordination of telecommunications needs and resources across NOAA programs and elements. In this effort it supports NOAA's Earth System Data and Information (ESDIM) program that seeks to implement an enhanced, upgraded NOAA data management system. The NARB provides a climate that fosters development of a strategy to achieve NOAA's network integration goals.

These goals include:

- interoperability among internal NOAA systems interconnected with telecommunications,
- consistent access to NOAA data and information services for outside users,
- reduced costs,
- optimal capacity to serve agency needs,
- consolidated resources, and
- improved network management services.

One of the NARB's first milestones on the way toward integration was a workshop of key senior Line and Staff Office personnel to assess NOAA's long-term telecommunication plans and objectives. The desired outcome was identification of common integration goals, priorities, constraints, and implementation steps. The NARB intended these to be distilled interactively from the needs expressed by representatives of NOAA's various operational, scientific, and administrative elements.

The NARB and ESDIM co-sponsored the Network Integration Strategy Workshop, August 11-13, 1992, in Washington, DC. The workshop was conducted by the NOAA Office of Administration and facilitated by an outside consulting firm, the Reston Consulting Group, Inc. The workshop convenor was Frank DiGialleonardo, Director of the NOAA Information Systems and Finance Office and NARB Executive Director.

To help attendees focus on the task at hand, the workshop included a day of structured presentations by line organizations. These provided an overview of the comprehensive needs and priorities of the entire Line or Staff Office to quickly enable others unfamiliar with the program to understand its requirements and constraints. NOAA presentations paralleled those by outside experts in network integration.

The strategic goal is to "provide well-defined, standards-based network connectivity among all NOAA computers, resources, and personnel."

The lead presenter was Gregory W. Withee, Deputy Assistant Administrator for Environmental Information Services [as this issue goes to press, now Acting Assistant Administrator for Satellite and Information Services] who set the tone for the workshop. He described the concerns of the ESDIM program and conveyed the urgency of electronic information bonding within NOAA for both programmatic and legislated reasons. He also reported on the results of the ESDIM network workshop held in Denver, May 1992.

With their discussion of standards, strategies, and networking, outside consultants Craig Johnson and Robert Stover of Unified Communications, Inc., helped the group to put the discussions in a broader context. Robert Kahn of the Corporation of National Research Initiatives added more perspective with a discussion of Wide Area Networks.

NOAA Line and Staff Office presenters refined the picture by giving a structured overview of their networks and systems in operation today. Presenters were: William Callicott, National Environmental Satellite, Data, and Information Service; Francis Balint and Richard Zitzmann, National Weather Service; Michael Fraser, Na-

– *continued on page 2*

Telecommunications and ADP Security Branch
NOAA OA124
7600 Sand Point Way NE
Seattle, WA 98115

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration

NOAA network integration, from page 1

tional Marine Fisheries Service; Richard Barazotto, David Enabnit, Peter Grose, and Mark Waters, National Ocean Service; Robert Mahler and Richard Grubb, Office of Oceanic and Atmospheric Research; Bruce Dearbaugh, Office of NOAA Corps Operations; and Robert Swisher, James Pugh, and Joseph Smith, Office of Administration.

After all presentations, attendees worked on connectivity issues in smaller groups. They discussed each one in a structured format and produced succinct descriptions of the main concerns, goals, scope of the effort, forcing functions and mandates, responsibilities and responsible organizations, specific steps for the next 12 months, and benefits.

The working group sessions focussed on networking, network management and security, GOSIP (see box), interoperable applications including E-mail, directory and other services, and user education and training. NARB members and staff from the Reston Consulting Group facilitated each session, and comprehensive notes documenting the deliberations were taken. The work products of each working group were discussed before the larger audience and compiled into a high-level plan for achieving interoperability. This draft document is presently being reviewed in more detail by the attendees.

As it is currently written, the network strategic goal is to "provide well-defined, standards-based network connectivity among all NOAA computers, resources, and personnel." Strategic components include networks, user applications, standards implementation, education, and training. The high-level steps to be taken include:

- Making common use of transport media among NOAA locations wherever possible.
- Designing and implementing a NOAA-wide network management system.
- Establishing a centralized Network Information Center (NIC) and local Network Councils at NOAA campus locations.
- Providing network users with the ability to perform file transfer, centralized calendaring, electronic signatures, Electronic Data Interchange, and remote data exchange, and to access both NOAA and non-NOAA data.
- Providing data security to protect sensitive data.
- Training and Educating users on the services available and how to use or operate a NOAA network.

To move the strategic plan closer to implementation, two additional Network Interoperability Workshops targeted for key technical representatives from each Line and Staff Office are planned for early 1993. The first will be presented in January and will focus on GOSIP education, collecting line and staff office requirements, and coalescing ownership for network integration.

A follow-up workshop in February for the same attendees will be held to develop a GOSIP transition strategy, an agency convergence plan, an initial set of standards, and a consensus mechanism for developing, authorizing, and maintaining the core set of standards.

The NARB is chaired by Helen Wood, Director of the NOAA Office of Satellite Data Processing and Distribution; its Executive Director is Frank DiGialleonardo, NOAA's senior official for Information Resources Management. Other members include representatives from each NOAA Line and Staff Office. NARB meetings and activities are coordinated by the Office of Administration, Information Systems and Finance Office thorugh the Telecommunications and ADP Security Branch, Systems Division.

NOAA's Network Advisory and Review Board meets every 4-6 weeks and will convene again January 12, 1993. NOAA employees are welcome to propose topics for future NARB consideration. Questions may be directed to NARB representatives or to Gary Falk, Chief, Telecommunications and ADP Security Branch, OA124, WSC-5, Room 305. Telephone: 301-443-3692. ■

GOSIP: A Standard for Telecommunications

GOSIP, the Government Open Systems Interconnection Profile, is specified in the Federal Information Processing Standards (FIPS) Publication 146-2 issued by the National Institute of Standards and Technology. It specifies a set of over 40 telecommunications standards that are mandated by the General Services Administration for use in Federal procurements. GOSIP standards apply through the seven layer telecommunications model developed by the International Standards Organization, so that a systems implementor must chose among both "competing" and complementary standards.

EARTH SYSTEM MONITOR

The *Earth System Monitor* is published quarterly by the NOAA Office of Environmental Information Services. If you have any questions, comments, or recommended articles, or if you would like to be placed on the mailing list, please call Richard Abram at 202-606-4561 or write:

Environmental Information Services
NOAA/NESDIS Ex2
Universal Building, Room 506
1825 Connecticut Avenue, NW
Washington, DC 20235

EDITOR
Richard J. Abram

EDITORIAL ASSISTANT
Nancy O'Donnell

U.S. DEPARTMENT OF COMMERCE
Barbara H. Franklin, Secretary

National Oceanic and Atmospheric Administration
John A. Knauss, Under Secretary and Administrator

News briefs

Plan for NOAA data and information management issued

On September 8, 1992, Dr. John A. Knauss, Under Secretary for Oceans and Atmosphere, approved the strategic plan for NOAA-wide data and information management. Titled *NOAA's Data and Information Management Strategy: A Vision for the 1990s and Beyond*, the plan was developed by the Earth System Data and Information Management (ESDIM) Program in conjunction eith an ESDIM team consisting of representatives of the NOAA Line Offices and major cross-cutting programs. Copies of the plan are available from: Environmental Information Services, NOAA/NESDIS Ex2, Universal South Building, Room 506, 1825 Connecticut Avenue, NW, Washington, DC 20235. Telephone: 202-606-4089.

Chinese scientists visit NOAA's national data centers

Wang Wanlin from the Chinese State Meteorological Administration arrived at NOAA's National Climatic Data Center, Asheville, North Carolina, on September 7, 1992, for a stay of approximately one year. During this period he will assist in the development of the Global Historical Climate Network (GHCN) data base and study these data for the Global Climate Perspectives System (GCPS). Data that he brought from China to analyze include monthly maximum and minimum temperature and total precipitation for a 138-station network and 6-hourly synoptic meteorological observations for a smaller network. The synoptic data will support joint efforts of NCDC and other agencies in creating international climatic CD-ROMs and baseline data sets.

Hua Bin of the National Marine Data and Information Service, State Oceanic Adminstration of China, and the World Data Center-D (WDC-D) arrived at the National Geophysical Data Center, Boulder, Colorado, on September 1, 1992, for a six-month stay. Mr. Hua, a group leader for computer networking at the WDC-D, will be studying the data and network structures used by the NGDC to achieve a common foundation for the computer network to be established in China at WDC-D. He will study both networking technology and quality control structures and algorithms used at NGDC for marine cartographic, geophysical, and geological data. This work will facilitate data and information exchange between the WDC-A in Boulder and the WDC-D in China.

On November 11, 1992, Huang Longsheng, a mathematician and computer specialist, and Wang Hong, an oceanographer and project officer, from the Chinese National Marine Data and Information Service arrived at the National Oceanographic Data Center, Washington, DC, for a 6-month stay. Mr. Huang will be working with the Global Temperature Salinity Pilot Project (GTSPP). After learning about the management of GTSPP data, he will develop data products and make it possible for China to play a major role in the GTSPP. Continuing work done at NODC by an earlier visiting scientist from his institution in 1989-1990, Mr. Wang will work on quality control and data products for ocean buoy data.

WDC-A for Paleoclimatology established at NGDC

At a recent international World Data Center meeting, the NOAA Paleoclimatology Program was formally designated as the World Data Center A (WDC-A) for Paleoclimatology. The new WDC-A will be based at the National Geophysical Data Center, Boulder, Colorado, and directed by Dr. Jonathan Overpeck, Head of the NOAA Paleoclimatology Program. The WDC-A will contribute to the International Geosphere-Biosphere Program Past Global Changes core project by coordinating the exchange of paleoclimate data among earth scientists around the world. NGDC also operates the WDC-A's for Marine Geology and Geophysics, Solar-Terrestrial Physics, and Solid Earth Geophysics.

NODC field offices realigned

To reflect changes in ocean science programs and activities, the National Oceanographic Data Center has closed its Liaison Office in Anchorage, Alaska and established a new Liaison Office in Honolulu, Hawaii. Michael Crane, who served in the Alaska Office for 17 years, now heads the Southeast Liaison Office in Miami in place of John Sylvester who retired. Alaska was added to the area of responsibility of Syd Stillwaugh, the Liaison Officer in Seattle, Washington, whose office is now designated as the Northwest and Alaska Liaison Office. The Hawaii Liaison Office is headed by Patrick Caldwell, who is involved in the processing and quality control of sea level data at the TOGA Sea Level Center at the University of Hawaii (see article on page 4).

NOAA Central Library to move to new NOAA campus in 1993

The NOAA Central Library, currently located in Washington Science Center Building 4, Rockville, Maryland, is scheduled to move by mid-1993 to a new facility in the NOAA campus at Silver Spring, Maryland. The library will occupy space in Silver Spring Metro Center III, one of two buildings under contruction at the site that already houses 1,800 NOAA employees in two completed buildings. At the new facility the library's resources will be more readily available to the 4,300 NOAA employees to be housed at the Silver Spring campus. Since it is directly adjacent to the Silver Spring station on the Washington Metro subway Red Line, the new location should also be more convenient for many visitors and non-NOAA users.

The NOAA Central Library holds a collection of over 1 million volumes in the oceanic, atmospheric, and related sciences, including a rare book collection of about 1,000 volumes.

Global change data and information management plan

The *U.S. Global Change Data and Information Management Program Plan* has been published and is available from the National Science Foundation. The report was prepared by the Interagency Working Group on Data Management for Global Change and issued by the Committee on Earth and Environmental Sciences of the Federal Coordinating Council for Science Engineering and Technology (FCCSET). The report provides the framework for the global change data and information management program, describes the programs of the participating agencies, and outlines the next steps to be taken in the program.

Copies of the 94-page report are available from: Committee on Earth and Environmental Sciences, c/o National Science Foundation, Attn: Forms and Publications, 1800 G Street NW, Room 232, Washington, DC 20550. Telephone: 202-357-7861.

Building an archive of tropical sea level data

A cooperative effort of the University of Hawaii and the NODC

Patrick Caldwell
TOGA Sea Level Center and
Hawaii Liaison Office
National Oceanographic Data Center
NOAA/NESDIS

Time series records of sea level heights from around the globe provide a wealth of information on oceanic and atmospheric behavior. High frequency sea level variations with periods from seconds to hours are dominated by wind waves, seiches, tides, storm surges, and occasional tsunamis (Pugh, 1987). Slower movements of sea level height with periods from days to years are mostly associated with transient synoptic scale meteorological features, ocean circulation patterns, and seasonal and other short-term climatic fluctuations (Mitchum and Wyrtki, 1988). The longest sea level records show the influences of local land subsidence, plate tectonic movements, and climate variations (Douglas, 1991).

To increase understanding of ocean-atmosphere coupling and the predictability of climate changes over time scales from months to years, the World Climate Research Program initiated the Tropical Ocean Global Atmosphere (TOGA) Project. One of the key observational elements of the TOGA project is sea level. In 1985, the TOGA Sea Level Center (SLC) was created at the University of Hawaii (UH) under direction of Dr. Klaus Wyrtki to concentrate the efforts of acquiring, processing, and archiving sea level data from the tropics.

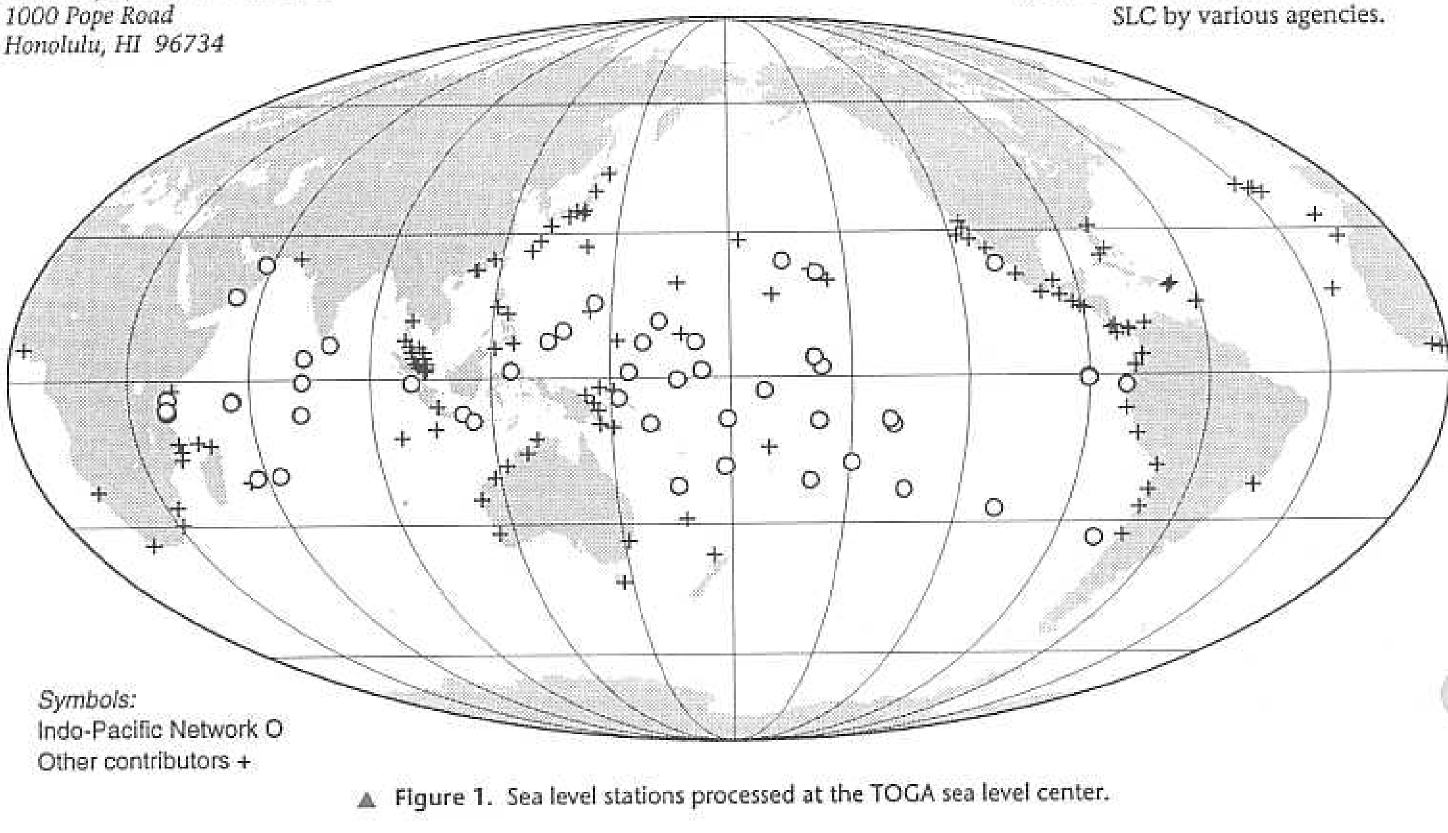
The TOGA program requires daily sea level values from sites identified in the implementation plan (International TOGA Project Office, 1987). Data collecting agencies around the globe have been asked to submit hourly sea level heights, which are necessary for thorough quality control.

As the quantity of data collected by the TOGA SLC increased, data management support was contributed by the U.S. National Oceanographic Data Center (NODC) though the establishment of the Joint Archive for Sea Level (JASL) at UH in 1987. The JASL supports the TOGA SLC in the collection, quality assurance, management, and dissemination of sea level data.

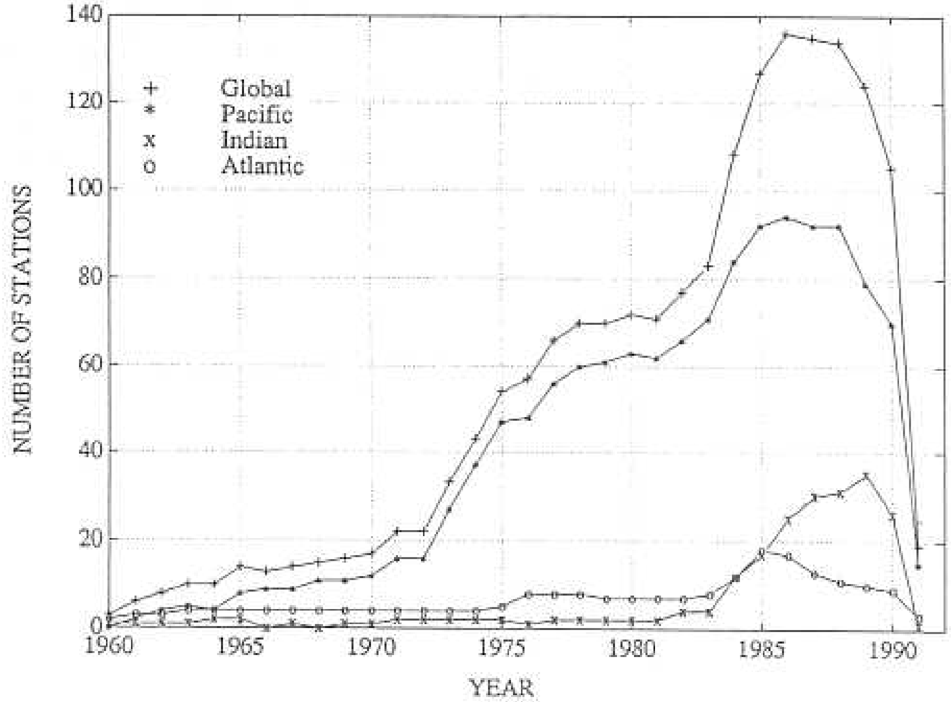
The TOGA SLC prepares scientifically valid, well-documented, standardized sea level data sets of hourly, daily, and monthly values. As of July 1992, 191 stations with 1,772 station-years of quality assured data (Figures 1 and 2) from the tropical Pacific, Indian, and Atlantic Oceans have been archived at the NODC and the collocated World Data Center A (WDC-A) for Oceanography where they are available to users around the globe.

At the TOGA Sea Level Center data processing and quality assurance methodology is applied to two data streams: (1) data from stations within the Indo-Pacific Sea Level Network, which is jointly operated by UH and various international agencies, and (2) data from stations in regional and national networks as contributed to the TOGA SLC by various agencies.

Hawaii Liaison Office
NOAA/NESDIS
University of Hawaii - MSB 316
1000 Pope Road
Honolulu, HI 96734



▲ Figure 1. Sea level stations processed at the TOGA sea level center.



▲ **Figure 2.** Sea level data availability by year and ocean basin.

Data from the Indo-Pacific Network

Through efforts of scientists at the University of Hawaii, the Pacific Tsunami Warning Center, and NOAA , the Pacific Island Sea Level Network (Wyrtki et al., 1988) was established with over 35 sites in the tropics. The success of this network led UH scientists to place additional installations in the Indian Ocean, where presently 17 stations are operational. The TOGA SLC routinely processes data from 52 sites of the Indo-Pacific Sea Level Network, with 31 Pacific and 3 Indian Ocean stations relaying in near-real time via satellite to UH (Kilonsky and Caldwell, 1991). The TOGA SLC also routinely receives via satellite and processes 18 stations in the tropical Pacific from NOAA's Global Sea Level Network of Next Generation Water Level Measuring Systems.

Quality assurance begins at the sea level station. Redundant instrumentation helps reduce data gaps and enhances quality control. Typically, at least two separate float and stilling well configurations are established at each site, with usually more than one recording mechanism for a given well. The station is maintained by a local employee, who is trained by UH technicians and is responsible for tide staff readings, minor repairs, and monthly changes of data storage media. A network of fixed bench marks on land are surveyed to the tide staff. Each site has a datum, the tide staff zero, which is used for referencing the measured sea level heights.

Beginning in 1988, a new feature was added to the installation configuration: a reference level switch. The switch is a triggering mechanism that is mounted at a fixed position on a piling or a stilling well at about the mid-tide level and is surveyed to the tide staff. It is linked to a Handar Data Collection Platform (DCP) that measures the precise time when the sea level vertically passes the trigger in the switch. The switch information supplements the tide staff observations for calibration of the measured sea level heights.

The Handar DCP relays sea level, switch, and technical data in near-real time via satellite to UH, where the data are continuously logged on a dedicated microcomputer and previewed daily to monitor the health of the installation. Spurious signals are pointed out to the UH technicians.

Data quality control relies primarily on the monthly data. A plot is made of the observed data and the residuals, which are the differences between the observed data and the predicted tides. These plots form the core of the quality control and are used for isolating spurious outliers, timing drifts, instrument malfunctions, and datum shifts.

For calibration, the visual tide staff readings are paired with the measured sea surface heights for corresponding times and plotted on a scatter diagram. Normally about 28 pairs are available for analysis each month. A standard deviation and mean difference between the staff and gauge readings are obtained. The standard deviation is typically on the order of 0.1 ft and the mean difference is the preliminary calibration constant.

At the end of the year, the measured sea level heights are linked to the tide staff zero. A final calibration constant is chosen for each extended time period, normally about a year, in which no physical changes occurred to the instrumentation, installation foundation, Handar software, tide staff, or the switch, and such that comparisons between instruments using calibration constants from both the tide staff observations and the switch are internally consistent.

At this point, the data are filtered to hourly values and reviewed as described in the next section. The Indo-Pacific Network data are available for distribution normally about 18 months after the year in which they were collected.

Data from regional and national networks

Hourly data are routinely received at the TOGA SLC from nearly 50 agencies collecting data in nearly 60 countries of the tropical and subtropical Pacific, Indian, and Atlantic oceans. Data holdings and processing techniques have been documented for the TOGA Pacific and Indian Ocean sets (Caldwell et al., 1989 and 1990). After applying checks to ensure the scientific integrity of the data, hourly, daily, and monthly data sets are prepared for the permanent archive with scientific units in millimeters and the time recorded as Greenwich Mean Time.

A linear least squares tidal analysis (Foreman, 1977) is applied to the observed hourly heights to calculate the harmonic constituents, which in turn are used to predict the tides and finally to obtain residuals. Plots of the hourly residuals are a primary quality control tool. For most island stations with deep

– continued on page 6

Tropical sea level data, from page 5

adjacent waters, the tides contain very few high-frequency components, the harmonic analysis reflects the observed tides to a high degree, and residuals have a flat signature. For regions with shallow coastal shelves, influences of rivers, or complex coastal or basin geometry, the tidal analysis may not completely resolve many of the higher frequency components, and the residuals will show the non-linear, shallow water tides. The hourly residuals are inspected on a case by case basis to correct or flag only erroneous features in the observed data.

After inspection of the hourly heights, daily and monthly values are computed using filters. Plots of daily heights occasionally show erroneous spikes caused by incorrect data over a span of hours to a few days. When available, bulletins that provide summaries of storm tracks, such as the *Mariners Weather Log* published by the NODC, are studied to determine if the extreme sea level heights are related to meteorological events.

Plots of daily and monthly values and plots of differences among adjacent stations are used primarily for monitoring the reference level stability. After a jump is identified, attention is returned to the plots of hourly residuals. For non-subtle shifts, the hourly residuals clearly show quasi-step function signatures. When these jumps occur periodically in time, such as at the beginning of the month when the gauge rolls are changed, it is clear that the reference level shifts are erroneous. When a shift is identified, the data contributors are requested to investigate the original data and calibration notes.

After quality control of the data is complete, the hourly values are filtered to daily values with a 119-point convolution filter (Bloomfield, 1976). The daily values in turn are decimated to monthly values with a simple average, which is archived along with a count of the number of days available for the calculation. If more than seven days are missing per month, the monthly value is not calculated and is replaced with a missing data flag. A quality assessment and general information file is maintained, which accompanies the data in the final archive.

The Joint Archive for Sea Level (JASL)

The International TOGA Sea Level database is supported by the Joint Archive for Sea Level, a cooperative effort of the National Oceanographic Data Center and the TOGA Sea Level Center at the University of Hawaii. The database includes hourly, daily, and monthly sea level values and continues to grow as new data are received from existing stations and additional stations are added to the network. After processing and quality control at the TOGA Sea Level Center at the University of Hawaii, data are submitted to the U.S. NODC and the collocated World Data Center-A for Oceanography for inclusion in the archives. New data are added to the archives once a year with about an 18-month lag after the calendar year in which the data were collected. The annual updates may include data previously submitted to NODC that have been reprocessed to improve data quality. Therefore, NODC completely replaces the data time series for that station with a new version that may include both new and reprocessed data.

The data base presently holds 1,772 station-years of data from 191 stations, including 116 in the Pacific Ocean, 26 in the Atlantic Ocean, and 49 in the Indian Ocean. The entire data base or selected subsets can be provided to customers on a variety of media. For further information on data availability, please contact:

National Oceanographic Data Center
User Services Branch
NOAA/NESDIS E/OC21
1825 Connecticut Avenue, NW
Washington, DC 20235

Telephone: 202-606-4561
Fax: 202-606-4586
Omnet: NODC.WDCA
Internet: services@nodc2.nodc.noaa.gov

Conclusion

The TOGA Sea Level Center is supported by the National Oceanographic Data Center through activities of the Joint Archive for Sea Level. The data are received at the TOGA SLC from the UH Indo-Pacific Network, which is mostly motivated through efforts of the TOGA Project, and from many national and international agencies that kindly contribute data and assistance.

Once a year, hourly, daily, and monthly data are submitted to the World Data Center A for Oceanography, which is operated by the U.S. NODC. The WDC-A and NODC maintain a growing sea level data archive and provide data dissemination to the scientific and public communities.

References

Bloomfield, P., 1976. *Fourier Analysis of Time Series: An Introduction.* New York: John Wiley and Sons. pp 129-137.

Caldwell, P., K. Wyrtki and S. Nakahara, 1989. *TOGA Sea Level Center: Data from the Pacific.* Univ. of Hawaii JIMAR-89-202, 34 pp.

Caldwell, P., K. Wyrtki and S. Nakahara, 1991. *TOGA Sea Level Center: Data from the Indian Ocean.* Univ. of Hawaii. JIMAR-91-233, 35 pp.

Douglas, B., 1991. Global sea level rise. *Journal of Geophysical Research.* Vol. 96, NO. C4, pp. 6981-6992.

Foreman, M., 1977. *Manual for tidal height analysis and prediction.* Institute of Oceans Sciences, Patricia Bay, Victoria B. C. Pacific Marine Science Report 77-10. (Unpublished manuscript).

International TOGA Project Office, 1987. *TOGA International Implementation Plan.* Boulder, Colorado. ITPO-No. 1.

Kilonsky, B. J. and P. Caldwell, 1991. In pursuit of high-quality sea level data. *IEEE Oceans Proceedings.* Vol. 2, October 1-3, 1991.

Mitchum, G. T. and K. Wyrtki, 1988. Overview of Pacific sea level variability. *Marine Geodesy,* Volume 12, pp. 235-245.

Pugh, D. T., 1987. *Tides, Surges, and Mean Sea-Level.* Chichester: John Wiley & Sons, 472 pp.

Wyrtki, K., and P. Caldwell, K. Constantine, B. J. Kilonsky, G. T. Mitchum, B. Miyamoto, T. Murphy, and S. Nakahara, 1988. *The Pacific Island Sea Level Network.* Univ. of Hawaii, JIMAR 88-0137, 71 pp. ■

Data storage using 8 mm helical scan tape

Increasing data volumes spur the quest for ever-higher capacity data media

John O. Kinsfather
Chief, Information Services Division
National Geophysical Data Center
NOAA/NESDIS

The 8 mm helical scan tape drive from Exabyte Corporation of Boulder, Colorado has been in use since 1985 and is now a *de facto* standard on workstations for backing up disk files. Small size, low drive cost ($2,500 to $4,000), low media cost ($5-$20), and high volume capacity (2 to 5 gigabytes without compression and 10 to 20+ gigabytes with compression) make this a very attractive tape backup method.

Because of their features, 8 mm helical scan recorded cartridges are being considered as an alternative to both 1/2-inch round tapes and 3480/3490 square cartridge tapes as an archive medium. But, questions about this new medium remain. Can 8 mm tape written on Exabyte* drives serve effectively for long-term data archiving? And what factors determine suitable applications for this data storage alternative?

Data Management Problems

The national data centers within NOAA's National Environmental Satellite, Data, and Information Service currently hold 150,000 reels of 9-track, 1/2-inch tape and 100,000 3480 cartridges totaling approximately 150 terabytes of satellite and *in situ* environmental data. Each of the data centers is converting standard 1/2-inch tapes to 3480 cartridges and also looking at the costs and benefits of switching to 3490 cartridges. Holding 200 megabytes, the 3480 cartridge has twice the capacity of a 6250 bpi, 9-track tape. The 3490 cartridge offers another doubling of this capacity, plus additional compression. Even a doubling of capacity, however, cannot keep up with the volume of new data flowing to the data centers.

*Exabyte is a registered trademark of Exabyte Corporation.

National Geophysical Data Center
NOAA/NESDIS E/GC4
325 Broadway
Boulder, CO 80301

The National Geophysical Data Center (NGDC) at Boulder, Colorado currently archives 300 different types of data. About 1 terabyte is in digital form, and historical growth rates have been about 10 percent per year. NGDC is now starting to receive 5 gigabytes of compressed data per day from the Defense Meteorological Satellite Program (DMSP). Six months after receipt of the first DMSP data, NGDC's digital holdings will have doubled. If just the raw incoming data were copied to 3480 cartridges and a backup copy made for offsite storage, the minimum number of 3480 cartridges required would be 50 per day, 365 days per year. Just making the 3480 copies would require a dedicated workstation and a full-time operator. In addition, many more 3480's would have to be accessed and copied to fill customer requests for data. The media cost of 25 3480's to hold 5 gigabytes is $100 versus $20 maximum for a single 8 mm cartridge that can hold the same number of bytes in 1/50th of the physical space.

The National Climatic Data Center (NCDC) at Asheville, North Carolina has an even larger problem coping with data it has begun receiving from the National Weather Service's WSR88-D advanced weather radar systems. When all 154 stations are operational in 1997, a potential data volume equivalent to 25,000 8 mm cartridges per year may need to be archived. These large data volumes require data managers to look at all factors involved in data storage. These include physical and chemical characteristics of the media for long time archive, data access speeds, I/O transfer rates, tape storage space, labor, and cost.

While there are a number of media storage devices under development or in limited use that could accommodate these high volume data sets, they are not mainstream standards and may never become standard. These devices include the D-1 and D-2 tapes (based on video technology) and CREO (optical tape). Exabyte cartridges are one of the few—perhaps the only—widely used and reasonably priced multigigabyte storage devices that may eventually be considered as a *de facto* standard. As such, the 8 mm media should be investigated for approval as an acceptable long term archive.

Single Sourcing

A major concern with using 8 mm media is the single manufacturing source for the tape drives. Exabyte Corporation and their partner Kubota in Japan are the only sources for the hardware. The mechanical components of an Exabyte drive are manufactured by Sony. Although Exabyte appears to be a sound and stable company, users remain dependent on a single vendor. Even if unforeseen circumstances caused the company to fail, however, it seems likely that another company would step in to continue the business because the large installed base of tape drives (about 500,000 by the end of 1992) makes this an attractive market.

Single sourcing for any equipment is always a concern but with the maturity of the product and the number of drives in use, hardware should not be a problem in using Exabyte products for data archiving. Similar products for recording on VHS media and 4 mm DAT offer Exabyte some competition, but they do not appear to be getting a large enough share of the market to threaten Exabyte's existence. Exabyte tape drives should be available for at least ten years to access data written to 8 mm helical scan recorded media and to migrate those data to a future generation of archive media.

Exabyte Drive Specifications

The typical Exabyte drive is sold in a compact 5.25-inch shoebox configuration. They are also available in a half-height size that can be installed in a PC in the same space as a floppy disk drive. The small size of the drive makes it ideal for use in the field where equipment space is limited. In a computer room environment they have no special space

– *continued on page 8*

8 mm helical scan tape, from page 9

or power requirements and they may actually be used on a desktop in an office.

The normal interface is SCSI (small computer systems interface) so they are easily adaptable to a wide range of open systems including UNIX workstations, Apple Macintoshes, and IBM compatible personal computers. On UNIX workstations they can be written to and read as a standard tape device without the installation of any unique device drivers.

Many system integrators package Exabyte drives with software that writes a proprietary data format. These proprietary formats must be avoided for long term archive data. Standard hardware and software are necessary to insure that the media can be read for many years in the future.

Tape Drive Performance

All tape drives are sequential access devices. Therefore, if small amounts of data scattered across the media are to be accessed quickly and frequently, the data should not be stored on tape. The defining characteristics of an Exabyte tape drive for archive purposes are that it is a relatively slow sequential access device with excellent error detection and correction capabilities and very high volume recording capacities per area of media.

Searching from one end of an 8 mm cartridge to the other may take up to two minutes. While extremely slow compared to milliseconds on an optical or magnetic disk drive, this is still much faster than searching or positioning through an equivalent storage volume of 50 9-track tapes, 25 3480 cartridges or 5 to 10 3490 cartridges.

The I/O (input/output) speed is probably the most critical factor. The Exabyte data transfer rate is equivalent to that of a SCSI 9-track tape drive, but somewhat slower than a mainframe 9-track tape drive. This transfer rate means that it takes over three hours to write or read a full 5-gigabyte cartridge on a drive operating continuously at near peak capacity. Duplication systems for copying 8 mm Exabytes also run at this speed, although with multiple drives 10 or more copies can be made simultaneously.

Storage Capacity

One terabyte of digital data (approximately 1 trillion bytes) written to Exabyte tape cartridges can be stored in a space of less than one cubic foot and a petabyte (10^{15} bytes) of data would fit in 900 cubic feet, the volume of a small office. This very high volume density makes Exabyte recorded 8 mm cartridges one of the most attractive data storage media available. However, it would take 556 hours (almost a month) of continuous operation of a single drive to read one terabyte. For terabyte and petabyte size databases, this becomes critical if the data are to be accessed, exercised, and eventually migrated to new media. Even if the immediate problems of field recording and data storage can be handled with Exabytes, an effective archive management plan imposes future access and migration requirements.

Robotic Access

For accessing an 8 mm archive some type of jukebox or robotic system is needed to make a larger volume of cartridges available unattended in a near-line mode. Exabyte Corporation manufactures both 10- and a 140-cartridge models. Value added retailers market these devices with the necessary software. The 140-cartridge model with multiple drives could provide relatively quick access to a terabyte (or more if the data are highly compressed) for a price in the $100K range. Other vendors also market 8 mm cartridge robotic devices including carousels.

Recording and Media Specifications

Exabyte recorders have rotating heads (helical scan) that write the data on tracks inclined slightly from true linear. The number of tracks per inch is 1,638, providing an areal density of 74 million bits per square inch. The tape itself is approximately 0.3 inches (8 mm) wide versus 0.5 inches for standard 9-track round tapes and 3480 cartridges. In addition to being narrower than other tape media, it is also thinner (10-13 micrometers). The length of a 5-gigabyte capacity 8 mm cartridge is typically 112 meters.

Even if Exabyte access speeds, transfer rates, and recording densities meet a user's requirements, media characteristics and environmental storage requirements must also be considered. Some recent concerns have been voiced that the media has not been formally approved for long term archives and that data stored on 8 mm tape would be unreadable after a few years. It does not have the long history of 9-track tape to state unequivocally that it will absolutely last for 25 years without loss of data. It also does not have the approval of the National Archives and Records Administration (NARA) as an archive medium.

The medium itself has been improved somewhat over the last few years, and archive data should only be written to data grade tape. Although the Sony camcorder mechanics and the Exabyte patented error detection and correction electronics will work with any 8 mm video tape, it is well worth a few extra dollars for the higher quality data grade tape. Data grade media will provide for more data to be recorded with fewer recoverable errors, there will be less deterioration over time, and the Exabyte drives will last longer due to less wear on the tape heads.

Archival Considerations

Storage conditions recommended for 8 mm helical scan recorded cartridges are the same as those recommended by the National Media Labs for other tape media: a constant temperature of 62°-68°F and a constant relative humidity of 35-45%. The 8 mm tape is a metal particle medium and is subject to corrosion by numerous contaminants that also effect other types of tape. But storage requirements for 8 mm Exabyte tapes are no more stringent than those for 9-track and 3480 cartridge tapes.

The substrate of 8 mm helical scan tapes is tensilized (stretched) before the metal particle coating that records the data is applied. Other recording tapes are created in this same manner. There can be some relaxation of the tensilization over time with the end result that the angle of the recorded stripes on the tape will change slightly. This could make it more difficult to read the tapes, but to date it has not been demonstrated to be a serious problem. The sophisticated Exabyte error correction and detection mechanisms handle these

– continued on page 12

The Comprehensive Aerological Reference Data Set

A global upper air database of radiosonde observations

Stephen R. Doty
National Climatic Data Center
NOAA/NESDIS

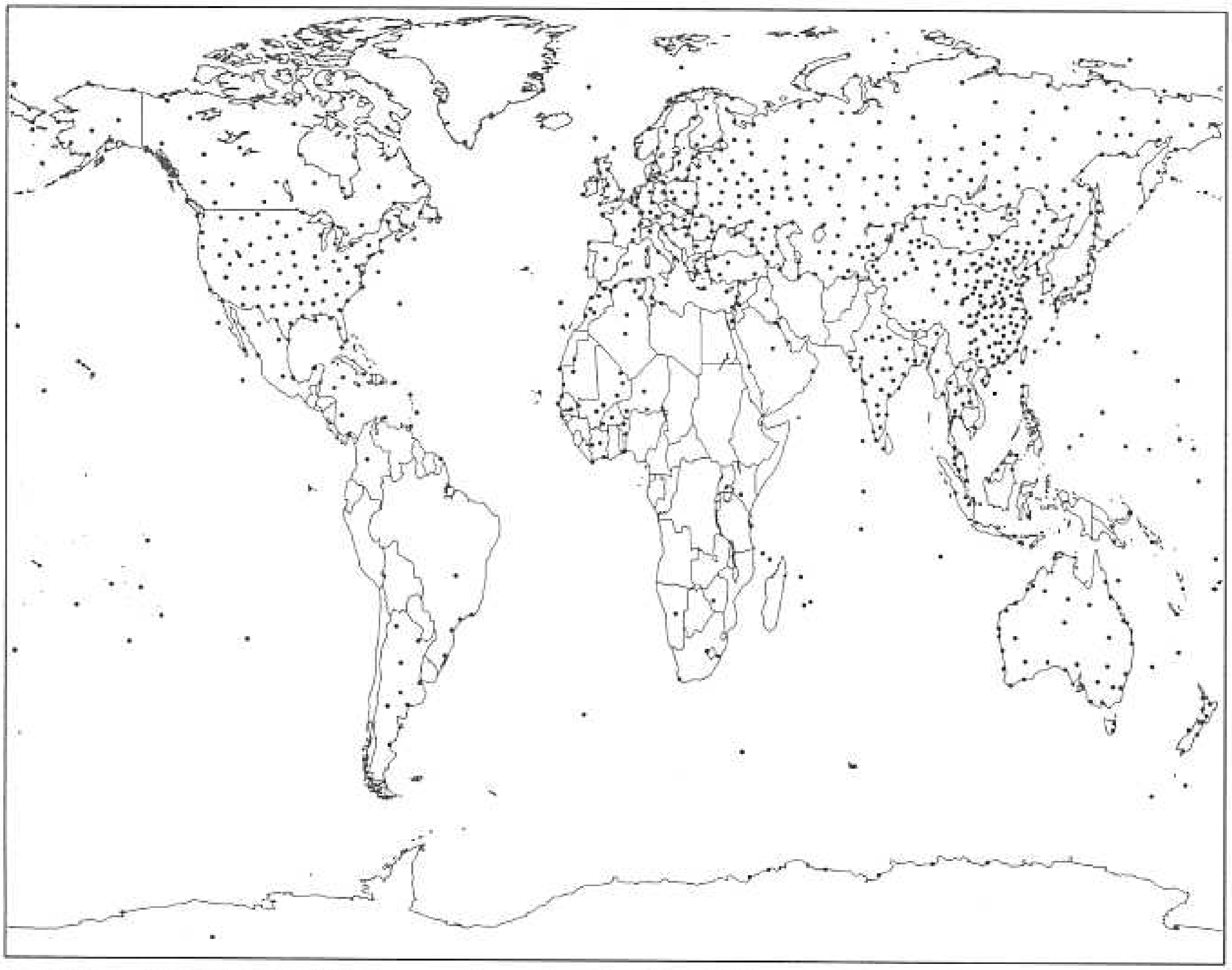
The National Climatic Data Center (NCDC) has a project underway to produce a quality global upper air data set based on daily radiosonde observations (see box on page 10). The Comprehensive Aerological Reference Data Set (CARDS) will be used to evaluate climate models, detect climate change, and support mesoscale research. Although it may sound simple, this project is the first undertaken to produce such a large and complex upper air data set.

National Climatic Data Center
NOAA/NESDIS E/CCx4
Federal Building
Asheville, NC 28801

The CARDS project is funded by the Department of Energy, NOAA's Climate and Global Change Program and the National Climatic Data Center. At NCDC the CARDS Project Manager is Stephen R. Doty, and the Project Scientist is Dr. Robert E. Eskridge. Project activities now underway include:

- assembling the global database from many sources;
- designing a quality control system;
- compiling comprehensive station histories; and
- researching bias adjustment techniques.

Data for over 1,000 global stations are being collected and processed. Figure 1 shows the location of these stations as of July 1985. Even developing countries have been able to maintain reasonably consistent observational programs (see Figure 2), so the database should be of high quality. CARDS will initially build a database consisting of observations from 1970 to 1990. Later the database will be extended back to the 1930s (the beginning of upper air observations using the radiosonde instrument), and it will be updated with more current data on a periodic basis.



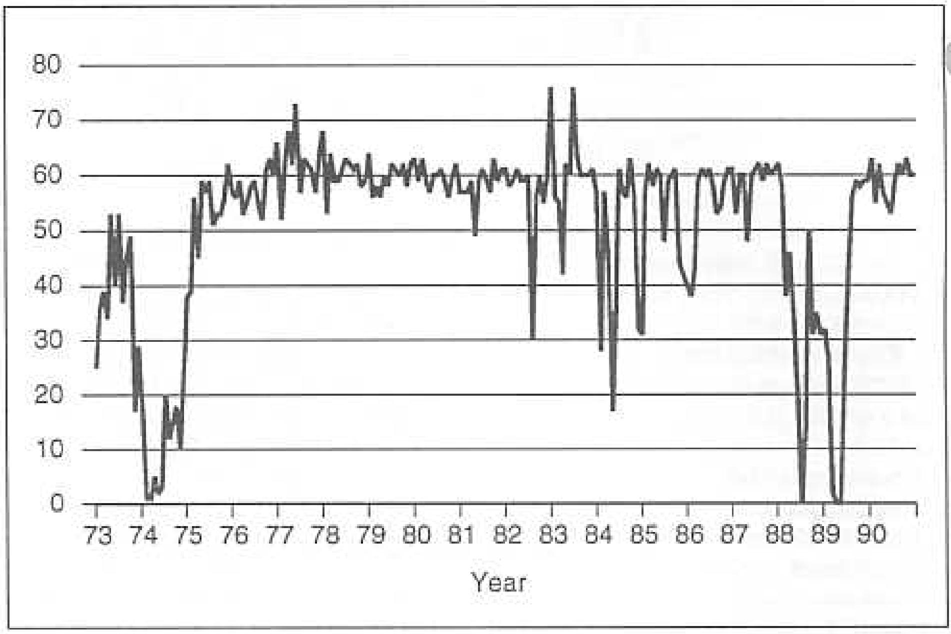
▲ **Figure 1.** Upper air sites with 30 or more observations (as of July 1985).

CARDS, from page 9

Database Development Program

The CARDS project will generate three databases: (1) a "raw" database consisting of all upper air reports with no corrections applied; (2) a quality controlled database where duplicate and incomplete (less than three mandatory levels) observations have been deleted; and (3) a database in which detected biases will be either removed or identified. The original or raw database will reside on magnetic tape while the quality controlled and bias checked databases are being loaded into a Relational DataBase Management System (RDBMS) on a UNIX-based workstation using an 85-gigabyte read-write optical jukebox.

These databases are expected to contain some 20 million observations amounting to 50 gigabytes. As other data sources such as organizations in individual countries are contacted, new data will be added. For example, bilateral agreements may enable a visiting scientist from the People's Republic of China to join the CARDS team in mid-1993, bringing additional data and expertise to the project.



▲ **Figure 2.** Number of upper air observations by month, 1973-1990, for Nairobi, Kenya.

> ### What is a radiosonde?
>
> The radiosonde is a balloon-borne instrument that measures meteorological parameters from the earth's surface up to 20 miles in the atmosphere. Originally developed in the 1930s, the radiosonde measures temperature, pressure, and humidity and transmits or "radios" these data back to earth. Upper air winds are also determined through tracking of the balloon ascent.
>
> Today, as in the last four decades, radiosonde observations are taken generally twice a day (at 0000 and 1200 UTC) around the globe. The data are transmitted to national meterological centers where they form the basis for numerical forecasts. The data are also received for retrospective use at climate centers such as NOAA's National Climatic Data Center. NOAA's National Weather Service operates a network of about 90 radiosonde observing sites located in the United States and its territorities. After the balloon bursts, radiosondes return to earth on a parachute. Approximately 25% are recovered and returned to NWS for reconditioning and reuse.

Quality Control Program

The Comprehensive Hydrostatic Quality Control (CHQC) system, developed at NOAA's National Meteorological Center, has been modified and implemented to perform the initial quality control. The CHQC uses a hydrostatic check to detect and correct errors in reported temperatures and geopotential heights. It will bring a standard QC level against which all data will be judged.

Dr. Oleg Alduchov of the Russian Research Institute of Hydrometeorological Information is working with the CARDS project team to develop more complex and comprehensive quality control algorithms. His new QC procedures will be applied in later stages of the CARDS project.

Station History Program

The CARDS station history has been designed to supplement the official NCDC station history files. Information is being compiled on not only changes in station location over time, but also on instrumentation and processing histories. The current database is being supplemented through the use of questionnaires which have been sent to all U.S. National Weather Service upper air stations.

The World Meteorological Organization also has been helpful in arranging for the CARDS project to receive supplemental information from their member countries. Through personal contacts, additional information such as the preliminary station history for the 205 stations located in countries of the former Soviet Union has begun to arrive.

Bias Removal and Identification Program

Systematic errors or biases in temperature or humidity data from radiosondes can hide true climatic trends. Steps are being taken to detect and correct these biases. This effort is complicated by the fact that every radiosonde instrument design has its own response to environmental conditions. Since the radiosonde was introduced, over 100 different instrument designs have been used around the world.

Dr. Bert Eskridge, NCDC'S Project Scientist, is leading the effort to develop techniques for modifying humidity data taken with VIZ radiosondes during the period 1961 to 1973. Other efforts include developing algorithms to correct temperature errors induced by various thermistor designs and studying statisti-

– continued on page 12

Data products and services

Global Change Data Sets for Research and Education

The National Geophysical Data Center has issued two new CD-ROMs and an educational diskette package that contain data sets related to global change studies. The Global Ecosystems Data CD-ROM, produced jointly by the NGDC and the U.S. EPA Environmental Research Laboratory, Corvallis, Oregon, contains selected global data on ecosystems, land use, wetlands, vegetation, climate, topography, and soils. The data are presented as raster gridded map layers registered to a common latitude-longitude base. The database is designed for use with statistical, image analysis, and geographic information systems.

An experimental CD-ROM from NGDC contains Normalized Difference Vegetation Index (NDVI) based on Advanced Very High Resolution Radiometer (AVHRR) data from NOAA's Polar Orbiting Environmental Satellites. The CD-ROM contains both biweekly and monthly versions of the data. The biweekly data are in Mercator projection, and cover the period April 1985 through December 1991. The monthly data were derived from the two biweekly periods that best fit a given calendar month and were reprojected to a 10-minute latitude-longitude grid.

In collaboration with the International Geosphere-Biosphere Programme (IGBP), the NGDC has released the Global Change Educational Diskette Package. The package includes a set of IBM-compatible diskettes holding (in compressed format) 41 megabytes of environmental data for Africa. The package also includes over 300 pages of documentation decribing methods of production and use of the data, and a 150-page manual of exercises for use in schools and universities, as well as by professionals. The data and exercises are usable in a variety of geographic information systems.
Contact: NGDC

Geosat Altimeter Crossover Differences on CD-ROM

The National Oceanographic Data Center has released a set of eight CD-ROMs containing altimeter crossover differences from the U.S. Navy Geodetic Satellite (Geosat) that were prepared by a group within NOAA's National Ocean Service. At crossover locations, two satellite crossing passes (one ascending and one descending) provide independent sea level measurements at the same place but at different times. The crossover difference data on these CD-ROMs covers the first 2.5 years of the Geosat mission, including both the initial 18-month Geodetic Mission and the first year of the Exact Repeat Mission. This was done to minimize the effect of the 5-week data gap (October 1 to November 8, 1986) when the altimeter was turned off during orbit maneuvers. By computing crossovers for the entire 2.5-year period, continuous sea level time series can be derived. The data on the CD-ROMs are organized by 18 geographic regions and are in a binary format.
Contact: NODC

NODC Inventory of Physical Oceanographic Profiles

In support of "data archaeology" efforts to increase the volume of historical ocean data available to climate and global change researchers, the National Oceanographic Data Center has issued the *NODC Inventory of Physical Oceanographic Profiles: Global Distributions by Year for All Countries.* This publication contains tables and data distribtution plots showing the data holdings in the NODC's major physical oceanographic databases as of June 1992. The plots show the locations of bathythermograph, CTD/STD, and oceanographic station observations in the NODC archive. The aim of this publication is to show researchers and data managers what data are available, help them identify data that are not yet in the archives, and encourage submission of manuscript or digital data that can help fill gaps in the data record. Copies of the publication are free.
Contact: NODC

CONTACT POINTS

For further details and ordering information about any of the NOAA products or services listed here or elsewhere in this issue of the Earth System Monitor, *please contact the appropriate source listed below.*

National Climatic Data Center (NCDC)
Climate Services: 704-259-0682
Satellite Services: 301-763-8399

National Geophysical Data Center (NGDC)
303-497-6958

National Oceanographic Data Center (NODC)
202-606-4549

NOAA Earth System Data Directory
202-606-4548
(Gerald Barton)

NOAA Central Library
Reference Services:
301-443-8330

The Global Historical Climatology Network

The National Climatic Data Center and the Carbon Dioxide Information and Analytsis Center, Oak Ridge National Laboratory, have compiled a new global baseline climate data set called the Global Historical Climatology Network (GHCN), Version 1.0. This data set contains historical time series of four surface climatological parameters: temperature (6,039 stations), precipitation (7,533 stations), sea level pressure (1,883 stations) and station pressure (1,873 stations). All values are monthly and the spatial coverage is of global land areas. In comparison to other major climatological data sets, coverage over Africa, Asia, and South America has been dramatically improved.

The database was compiled from numerous pre-existing digital archives. These include well known data sets such as the temperature database compiled by Jones et al., the World Weather Records, and the World Monthly Surface Station Climatology, as well as previously unpublished archives for nations such as the former Soviet Union, China, and Mexico.

Several data quality checks were performed, including the elimination of duplicate stations and the visual inspection of each time series for gross data processing errors and discontinuities. When possible, all station metadata parameters such as WMO number, latitude, longitude, elevation, and station name were also verified for accuracy.

The GHCN database is available from both the Carbon Dioxide Information Analysis Center and the National Climatic Data Center.
Contact: NCDC

8 mm helical scan tape, from page 9

and other sources of error and provide an extremely low bit error rate.

As with other tape archive media, a clean environment for recording and storing the tapes is essential. The manufacturing process leaves a small amount of particulate debris on new tapes and only after 5-10 passes through an Exabyte drive is the tape debris level reduced to a minimum. While this precleaning may be desirable, it is unreasonable to assume that anyone would have the time to run the tapes through a drive multiple times before recording. A constant temperature of 65°F and humidity of 40% are critical to the long-term storage life of this medium. It is also recommended that the tapes be exercised by being read at normal speeds, once every 6-12 months. Considering that it will take three hours to read each tape, it may be impractical to store very large data holdings on Exabyte tapes.

A general method for detecting changes in tape media over time using check tapes has also been suggested. A few check tapes would be spread throughout the archives. These tapes would be read periodically and checked very closely for errors, deterioration in signal level, contamination, erosion, and other problems. Once signs of deterioration were detected, then all archive tapes of that age would be migrated to new media.

Cost

Many data management decisions are based on cost. Total costs for a data storage alternative include equipment, media, space, and labor for storing, exercising, accessing and migrating an archive data set over the extended life of the archive. For some data sets, Exabyte tape technology may be the best choice, for others it may not be the best and for still others it may be the only affordable option today but this may lead to expensive problems of migration and/or access in the future.

This article is not meant as an endorsement of the products of any commercial vendor. Rather it is intended to provide information that data managers need as they search for effective solutions to specific data storage problems. While perhaps not ideal for all data storage applications, Exabyte technology may be the only affordable option that is available today for some data archives. Data managers will always face the need periodically to migrate archives to new media. Even with their limitations, however, Exabyte hardware and media may be a viable option for archiving some data sets during the next five to ten years. ■

CARDS, from page 10

cal methods to detect biases in time series data or in spatial data when station histories are lacking.

Conclusion

This joint project between NOAA/NCDC and the Department of Energy will make available the most comprehensive upper air data set yet developed. By mid-1993, individual observations from the 1930s to 1990 for over 1,000 sites worldwide will give the climate change research community the reference quality data set required for detailed upper air studies. The station history files being compiled will be the most complete and accurate metadata ever collected for the global upper air network. ■

Contributing to the Earth System Monitor

The *Earth System Monitor* welcomes articles by NOAA authors about agency programs, projects, and activities related to environmental data and information management. A brief sheet of Guidelines for Contributors that gives the publication schedule and other information is available on request.

If you would like to discuss an idea for an article or receive a copy of the *Earth System Monitor* Guidelines for Contributors, please contact the editor, Richard J. Abram:

Telephone:	202-606-4561
Fax:	202-606-4586
Omnet:	NODC.WDCA
Internet:	services @nodc2.nodc.noaa.gov

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
Publication Distribution Facility
Rockville, MD 20852-3809
ATTN: Earth System Monitor

Address Correction Requested
OFFICIAL BUSINESS
Penalty for Private Use $300

Third Class Rate

40