ACCESSING, BROWSING AND VISUALIZING MULTITHEMATIC GLOBAL DATABASES: TOOLS FOR THE 1990S

A. M. Hittelman, D. M. Clark, and R. E. Habermann
U.S. Department of Commerce, National Oceanic and Atmospheric Administration (NOAA)
National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80303 U.S.A.
FAX: +303-497-6513; Email: info@ngdc.noaa.gov

Abstract

A state-of-the-art approach to meeting the data and information requirements of environmental science researchers has been developed at NOAA's National Geophysical Data Center. This approach spans a full spectrum of data management, including assimilation, validation, documentation, distribution and access using modern browse and extraction methods. The system developed at NGDC employs graphical user interface technology which is applicable to CD-ROMs as well as on-line network systems like MOSAIC.

GeoVu is a windows based, multiplatform, data browse, access and visualization tool developed as part of this system. It helps users identify data of interest and export data to an application of their choosing such as GIS or image processing software. Using GeoVu, researchers are able to view numerous CD-ROMs, as well as their own data collection, with the same user interface. NOAA's National Geophysical Data Center offers GeoVu to interested individuals at no charge. The software for Microsoft Windows, UNIX and Macintosh as well as a data sampler (which can be used to explore the various features of GeoVu), are available over Internet: ftp.ngdc.noaa.gov; gopher.ngdc.noaa.gov; or http:/www.ngdc.noaa.gov/ngdc.html.

Several NGDC CD-ROM data products have been supported with GeoVu technology. Some of these will be demonstrated, including: Global View, Global Ecosystem Database, Digital Elevation Models, Coastal Change Analysis Program, Geologic Hazards Photos, and others.

Introduction

The two most revolutionary data management techniques for the environmental researcher in the 1990's will almost assuredly be the utilization of CD-ROM technology to provide large volume, integrated thematic data sets which describe the environment, and the provision of such data via a high speed global network and related systems such as Internet, the World Wide Web and Mosaic. Both aspects are changing the way that many researchers are conducting their day to day activities in investigating the spatial and temporal relationships which describe the Earth processes.

The National Geophysical Data Center (NGDC), which has been involved with environmental data management for thirty years, has recently compiled a suite of integrated, some which were peered reviewed, global environmental databases (Global View) and developed a pioneering software system (GeoVu) which addresses a wide range of environmental databases. GeoVu can browse, access, import external data and export subsets of database for implementation in to native systems for research applications. Using innovative object oriented programming and data description language, it can access many of NGDC's CD-ROMs which use a wide range of formats. Finally, these data, along with the GeoVu software, are currently being implemented for on-line access on the Internet via the World Wide Web, using Mosaic.

A systematic approach to servicing the data and information needs of global researchers has been developed at NOAA's National Geophysical Data Center. This approach spans a full spectrum of data management, including: assimilation, validation, documentation, distribution and access using modern browse and extraction methods. Providers can consistently describe their data and users can consistently access data of interest. The approach employs portable graphical user interface technology which is applicable to CD-ROMs, as well as on-line network systems.

GeoVu is a Windows-based, platform independent, data browse, access and visualization tool. It helps users identify data of interest and export them to an application of their choosing, such as GIS packages. Using GeoVu, researchers are able to view numerous CD-ROMs, as well as their own data collection, with the same user interface. NOAA's National Geophysical Data Center offers GeoVu to interested individuals at no charge. The software -- for Microsoft Windows, Unix and Macintosh -- as well as a data sampler (which can be used to explore the various features of GeoVu), are available over Internet:
ftp.ngdc.noaa.gov; gopher.ngdc.noaa.gov; http:/www.ngdc.noaa.gov/ngdc.html.

Several CD-ROM data products have been supported with GeoVu technology. Some of these will be discussed and demonstrated, including: Global View, Global Ecosystem Database, TerrainBase, Coastal Change Analysis Program, Geologic Hazards Photos, and Spitak Earthquake of 1988.

1. The Global View Project

The integration of global multithematic environmental data sets for the use in environmental modeling has been an area of discussion and research ever since global models have been developed (Hastings et al, 1992 and Clark et al, 1992). They are also needed for regional and continental scale Earth process studies and environmental assessments. In order to address the needs of global environmental modelers, NGDC developed a series of integrated global data sets which culminated in the product Global View.

Global View features a collection of global data sets previously published by NOAA's National Geophysical Data Center (NGDC) on CD-ROMs. Common access software (known as GeoVu) is included and facilitates data browse, selection and visualization. The software operates in multiple windows environments including Microsoft Windows on PC-compatibles, X-Windows (on SUN and SGI), or Macintosh.

The data included in this collection come from previously published CD-ROMs, including:

A. **NGDC's Ecosystem and Global Change Program**: Global Ecosystem—Disc A (Kineman, 1992); Global Ecosystem—Disc B (Kineman, 1993)
B. **NOAA's Coastal Ocean Program**: Coastal Change Analysis Project: Chesapeake Bay Region (Hittelman et al., 1994)
C. **NGDC's Global Remote Sensing Program**: Experimental Calibrated Vegetation Index (Hastings and Gallo, 1994)
D. **NGDC's Terrain Model Program**: TerrainBase: Worldwide Digital Terrain Data (Row and Hastings, 1994)

Some of these items have had wide distribution as individual products, but most represent data sets that are still in a developmental stage. NGDC combined these data together with a common user interface to improve data integration and to demonstrate their usefulness in integration of multi-thematic spatial data, which is used in global and regional assessments, process modeling, and research.

The four compact discs contains approximately 1.5 gigabytes of data, organized as several dozen thematic data sets. The primarily gridded data included in this collection were contributed by many national and

international organizations, from both the academic and government sectors. It is impossible to cite all contributors since there are often many organizations that have participated in some small way in the development of a single data set. Major contributing institutions include:

> Dept. of Energy; Dept. of Defense; NOAA; NASA; U.S. Environmental Protection Agency; U.S. Geologic Survey; Centro Internacional de Agricultura Tropical, Columbia; International Institute for Applied Systems Analysis, Austria; National Center for Atmospheric Research; United Nations Environmental Program; World Conservation Monitoring Center; Australian National University; Chiba University, Japan; Columbia University; Cornell University; Technishe Universitat Graz, Austria; Technishe Universitaeit Delft, Netherlands; University of Liverpool; University of Miami; University of Oklahoma; Macquarie University; Washington University; and many others.

GeoVu, the access software provided with this package, is designed to help users identify data of interest and export them to an application of their choosing such as Geographic Information System (GIS) packages. Using GeoVu, researchers are able to view numerous CD-ROMs, as well as their own data collections, with the same intuitive user-interface.

The GeoVu menu system allows the user to:

- select a specific data set
- identify the geographic area-of-interest
- output data from the compact disc onto hard disk or diskette in any of several output formats
- view the geographic location of the data observations

GeoVu was developed with funding support from NOAA's Climate and Global Change (C&GC) Program and from the NOAA's Earth System Data and Information Management (ESDIM) Program.

1.1 Global Ecosystems Database

The Global Ecosystems Database (GED) was developed (in part) as an interagency project between NGDC and the EPA's Corvallis Environmental Research Laboratory (ERL).

This database program is complementary to global change modeling activities, particularly those dealing with the interaction between the climate and biosphere.

The goal of the ecosystem project is to provide an integrated set of modern global and continental scale data needed by the global change research community. The project concentrates on modern observational data, including remotely sensed data and data from other insitu observational programs.

A prototype of this database was distributed on CD-ROM in 1991 to reviewers for quality evaluation, scientific review, and experimentation. The Global Ecosystem Database currently consists of three CD-ROMs:

- Disc A—released in 1992
- Disc B—in beta-test review (1994) and scheduled for public release in 1995
- Disc C—in development (scheduled for release in late 1995)

The Global Ecosystem Database CD-ROM incorporated in *Global View* includes data sets from the published Disc A and Disc B.

The Global Ecosystem Database includes complementary multi-thematic data sets on compatible grids, registered to a common origin and projection. These data have been structured to be operable with several existing Geographic Information Systems (GIS), so that a complete analytical package can be available to reviewers and other scientists for evaluation, experimentation, and further development.

While data for this project were structured for use by IDRISI (a GIS developed at Clark University), the database is also designed to be easily loaded into other GISs (such as GRASS). The database structure has been kept as system-independent as possible, so that data can be ported relatively easily to other systems.

1.2 Coastal Change Analysis Project

One of the most useful observation methods for studying global change are data collected from Earth observation satellites. Changes in Earth processes can be derived from an array of multispectral sensors currently operating under the flags of many nations. Changes in environmental indicators such as vegetation, snow cover, and land use are but a few of the areas currently being investigated. As an example of the utility of these methods to study change in baseline conditions, the Coastal Change Analysis Project data was included in Global View. While the CD-ROM does not contain a global data set, it demonstrates methods, using Landsat data, which will be useful by scientists studying global processes.

The Coastal Change Analysis Project (CCAP) CD-ROM was prepared by NGDC, in cooperation with the National Oceanographic Data Center and the National Marine Fisheries Service of NOAA. This CD-ROM (Hittelman et al., 1994) contains satellite-derived (Landsat) land cover classification.

The Coastal Change Analysis Project's purpose, under the Estuarine Habitat Program component of NOAA's Coastal Ocean Program, is to "improve understanding of the relationships among land cover changes, changes in aquatic habitat, and changes in fisheries" (Dobson and Bright, 1991 and 1993). One means of accomplishing this is to monitor the change of U.S. coastal habitat using remote sensing. However, to monitor land cover changes successfully, "a classification system for coastal uplands, wetlands, and submerged ecosystems was required" (Klemas et al., 1993).

Through a series of regional workshops, CCAP has worked with other Federal and State agencies, and academia to establish a CCAP protocol, including a suitable classification system, for performing land cover analysis (Dobson et al., in press). A primary objective is to periodically assess land cover changes by analyzing remotely-sensed data for specified coastal regions.

The Chesapeake Bay area (on the east coast of the U.S.) is the first in a series of land cover change products to be released by CCAP. Other CCAP land cover products are expected over the next year for selected coastal areas of the following states: Maine, Connecticut, Rhode Island, North Carolina, Washington, Oregon, and Alaska.

The Chesapeake Bay product consists of three parts: (1) a 1984 land cover analysis, (2) a 1988–89 land cover analysis, and (3) a land cover change analysis, all in raster format. Land cover data on this CD-ROM are stored in a raster format using a Universal Transverse Mercator (UTM) coordinate system. Each of the three components has 8,958 pixels per row with 12,150 rows, for a total of nearly 109 megabytes in each data file. Each of the three land cover files were derived from Landsat Thematic Mapper scenes. The Landsat scenes were used to define the 14 land cover classes of the Chesapeake Bay area for 1984 and 1988–1989. Changes in land cover from 1984 to 1988–1989 were also identified and put into 81 classes.

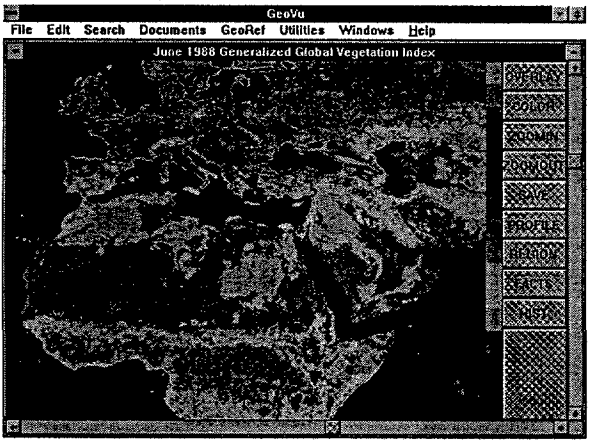
Land Cover Classes

Class	Class Description	Class	Class Description	Class	Class Description
1	Developed—High Intensity	6	Evergreen Forest	11	Palustrine Emergent Wetland
2	Developed—Low Intensity	7	Mixed Forest	12	Tidal Flats
3	Cropland	8	Mixed Shrub/Scrub	13	Exposed Land
4	Grassland	9	Palustrine Forest	14	Water
5	Deciduous Forest	10	Estuarine Emergent Wetland		

GeoVu uses color palettes to highlight the nature of change. Once an area of interest is displayed, the researcher can easily alternate between specialized palettes. For example, to view only changes with respect to developed land, the available specialized palette represents all other classes as black with only developed land in colors -- these colors differentiate the classification with respect to their class prior to becoming developed. Since GeoVu does not re-read the data to alter color presentations, the displays tend to appear quickly.

1.3 Experimental Calibrated Global Vegetation Index from NOAA's AVHRR

The Advanced Very High Resolution Radiometer (AVHRR) on NOAA's Polar Orbiting Environmental Satellites is a source of several widely available environmental parameters, including NOAA's operational Sea Surface Temperature, Aerosol Optical Thickness, and Global Vegetation Index (GVI) products. NOAA's operational GVI has been produced since 1982; since that time several proposals for improved GVIs, with additional corrections for sensor performance, have been made.



One product has been NGDC's own experimental Monthly Generalized GVI, included in the Global Ecosystems Database. Another product was an effort by Kevin P. Gallo of NESDIS's Office of Research and Applications, which added pre-launch sensor calibration, plus additional screening for data drop-outs, low solar zenith angle, and additional temporal compositing and cloud screening in the computation of biweekly GVIs from NOAA's operational weekly GVIs. The resultant GVIs have reduced contamination from clouds, use pre-launch calibrated albedos rather than uncalibrated digital counts, and provide flags where the data do not meet specified quality control criteria.

In the Gallo data biweekly indices are compiled from averages of weekly temporal composites of AVHRR data. The data cover 75°N to 55°S and 180°W to 180°E for the period of April 1985 to December 1991.

An experimental normalized difference vegetation index (NDVI) was developed and produced during 1988 through 1990, from weekly visible and near-infrared AVHRR channel data (Kidwell, 1991). The data are produced for the region between 75° North latitude and 55° South latitude. Data resolution in the Mercator projection varies from 19.6 km pixel size at the equator to 15 km at 40° (North or South). The reflectance values of the visible and near-IR data were computed from pre-launch calibration coefficients. The NDVI was computed as:

$$NDVI = (nearIR - visible)/(nearIR + visible)$$

The calibrated visible and near-IR data, and solar zenith angle data included on the NOAA GVI product were utilized to screen the NDVI data for cloud contamination and low (less than 15 degrees) solar elevation at the time of data acquisition. Data were also screened for data drops.

NGDC's monthly maxima GVIs are computed from the biweekly averages for each month between April 1985 and December 1990. The data have the same spatial and temporal coverage as the biweekly averages. These data are also reprojected to a 10-minute latitude-longitude grid.

Monthly maxima of Gallo's experimental GVIs were computed by taking the maximum values of biweekly GVIs for each month, then reprojecting the original Mercator-projected data to latitude-longitude projection. The compilation was produced from Gallo's data by the U.S. Geological Survey's EROS Data Center.

Inspection of these data at NGDC showed that the computed data were internally consistent to within one grid cell (the locational accuracy usually attributed to the NOAA Polar Orbiting Environmental Satellites that house the AVHRR sensor). However, the data were misregistered to the Earth by 1 grid cell (to the south). After independent the data were reregistered by removing the northernmost row of data, and in-

serting a new row at the bottom of each data file. As these rows contained no GVI values, no data were lost in the process.

1.4. TerrainBase: Worldwide Digital Terrain Data

The most requested data type for studying the Earth is hypsography or the vertical relief of the solid portion of the globe. The uses for these data are numerous and include a boundary conditions for General Circulation Models, integrated with remotely sensed data to produce spatially distributed geomorphologic features, combined with hydrology to derive regional watershed models, and so on. NGDC has had a long history of compiling these types of data, both on the land surface and the seafloor. The latest effort is Terrainbase, an ongoing project conducted by NGDC and World Data Center-A for Solid Earth Geophysics (NGDC/WDC-A) to develop a research-quality collection of digital terrain models (DTMs). The collection consists of data models that provide land elevation and ocean depth values for areas of regional and global coverage.

All of the data acquired through this project are periodically compiled onto CD-ROMs and made available. Collected data are also used to produce the TerrainBase Global Terrain Model which provides a complete matrix of land and ocean values for the entire world. Improvements to the global model are made systematically as new source data become available.

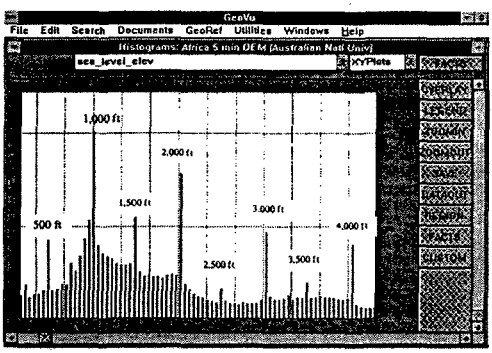
The terrain data contained in the Global View product are referred to as *digital terrain models*, or DTMs. These models consist of digital height values that represent the elevation of Earth's surface relative to mean sea level. All of the models used in TerrainBase are in a raster format, meaning that the height values are distributed at regular spatial intervals to form a grid of elevations covering a specified geographic region. Distances are expressed as minutes or seconds of arc. The global model has a grid cell spacing of five arc-minutes, while the regional models have grid spacings ranging from 30 arc-seconds to 10 arc-minutes.

All grids are referenced to the Cartesian coordinate system; however, the grid interval spacings vary from model to model. Smaller grid intervals produce detailed models, while larger grid intervals produce very generalized models. The accuracy of digital terrain models depends on the quality and spatial density of source data as well as the procedures used to convert these data into a gridded model. It is incorrect to assume that digital terrain models provide more accurate information than other sources of terrain data.

The digital nature of the models is misleading and is often misinterpreted by users that assume that the digital values represent exact measurements. On the contrary, most terrain models are derived from sources, such as paper maps, that are generally interpretive in nature. In areas where source data are abundant, the terrain models tend to be relatively more reliable, and in areas where source data is scarce, the models tend to be products of generalized interpolation and are thus less reliable. Even in the best cases, however, the source data are mostly interpretive.

NGDC conducted a series of quality analyses on each terrain data set to aid users in determining the suitability of each model for various applications. The analyses are only meant to give the user a general idea of the quality of each data set by identifying for the user as many significant errors and artifacts as possible. Problems were not documented unless there is potential for significant adverse affects in end-user applications.

Users should devise their own criteria for quality standards based on their specific applications and test each selected data set to ensure that they meet these standards. Every terrain model contains imperfections that will affect certain applications. It is important that the user identify and understand the potential effects of these imperfections before using any terrain model. In the adjacent diagram, GeoVu displays a data distribution histogram of one of the digital elevation models. Note the data contour interval artifacts which are present in the DTM.



2. Natural Hazard Data Collections

Living on our planet is not always easy, and we must adjust to catastrophic natural processes such as those caused by earthquakes, volcanoes and tsunamis. Understanding these events and planning for their aftermath often can save lives and mitigate the consequences. In the discipline of natural hazards, popular forms of data include photographs, maps and diagrams. Two recent CD-ROM products which support these formats are the *Geologic Hazards Photos* (a 2 volume set) and the *Spitak Earthquake (7 Dec. 1988).*

The *Geologic Hazards Photos* CD-ROM was developed from a popular series of hazard slide sets, which focuses on destructive events caused by earthquakes, volcanoes and tsunamis. For years the National Geophysical Data Center has been documenting significant events with sets of 20 well documented 35-mm slides. These pictures were scanned as part of a data preservation effort and images in both 24-bit TIF and 8-bit PCX formats were created. GeoVu software helps viewers scan the PCX images as a continuous slide show or through a hiearchical menu. All photos have descriptive text that can be pulled up using the 'facts" button.

The *Spitak Earthquake (7 Dec. 1988)* CD-ROM was developed as a cooperative effort between centers in Russia and the United States to provide researchers with a complete set of data for the devastating 1988 earthquake that occurred in Armenia. The International Decade of Natural Disaster Reduction Commission supported the development of this database which includes over 2,000 files -- the majority of which are descriptive text, maps, charts, photos, and other graphic images. GeoVu software supports MS-Windows access to these data, and a native DOS access system (developed by Russian colleagues) supports additional access.

3. The National Geophysical Data Center

NOAA collects, manages, and disseminates scientific data that result from studies into the environment. NGDC, one of the several data management centers of NOAA, is responsible for data activities in the fields

of topography, solid earth geophysics, paleoclimatology, marine geology and geophysics, and solar-terrestrial physics.

NGDC data support many scientific and engineering endeavors, such as the compilation of global baseline data sets, the assessment and mitigation of geologic hazards (for example, earthquakes and tsunamis) and in the exploration for minerals and petroleum. Our data also support basic and applied research in many of the Earth science and engineering disciplines. Customers include the general public, data managers, engineers, and scientists in private industry, academia, and federal and state governments.
Typical services include:

- Selective searches of data bases
- Customized data retrieval and output
- Computer-generated data products—including formatted listings, data summaries, map plots
- Technical reports, catalogs of data inventories
- Publication of selected reports for other organizations (through World Data Center-A)
- International data exchange (through World Data Center-A)
- Data systems for managing all types of geoscience data
- Data management and archival for special projects

World Data Centers

Five World Data Centers are hosted by NGDC, under the auspices of the National Academy of Sciences, including: Solid Earth Geophysics, Marine Geology and Geophysics, Solar Terrestrial Geophysics, Paleoclimatology, and (in cooperation with the University of Colrado) Glaciology, Snow and Ice. These World Data Centers (designated WDC-A) exchange data internationally, and assists scientists in obtaining copies of original or calibrated data held by other scientists or other world data centers. WDC-A also publishes data reports, maps, and compilations (including selected manuscripts) for international organizations. Both our national and world data centers encourage data exchange agreements with other organizations.

References

Clark, D. M., Ruttenberg, S., Hastings, D.A., Kineman, J.J., (1992) The integration of multithematic global databases, GIS and environmental modeling: A tool for the study of global change. In: Glazier, P., and Ruttenberg, S., (eds.) Proceedings of the 13th International CODATA Conference: Data for Global Change. CODATA Secretariat, Paris, pp. 15-22

Dobson, J.E., and E.A. Bright, 1991: "Coastwatch—Detecting Change in Coastal Wetlands," *GeoInfo Systems*, Jan/Feb 1991, pp. 36–40.

Dobson, J.E., and E.A. Bright, 1993: "Large-Area Change Analysis: The Coastwatch Change Analysis Project (CCAP)," Unpublished manuscript prepared for *Proceedings of the Pecora 12 Conference, August 1993.*

Dobson, J.E., R.L. Ferguson, D.W. Field, L.L. Wood, K.D. Haddad, H. Iredale III, J. Jensen, V.V. Klemas, R.J. Orth, and J.P. Thomas [in press]: "NOAA Coastal Change Analysis Project Guidance for Regional Implementation," Version 1.0–1994, NOAA Technical Report.

Hastings, D.A., D.M. Clark, and J.J. Kineman, 1992: "Integrated Global Environmental (Global Change) Data Activities Support Researchers and Educators," *Proceedings of the Central Symposium of the*

International Space Year Conference, 30 March–4 April 1992, Munich, Germany, ESA SP-341 (7/92), pp. 1379–1382.

Hastings, D.A. and K. Gallo, eds., 1992: *Experimental Calibrated Global Vegetation Index from NOAA's AVHRR, 1985–1991*, Key to Geophysical Records Documentation 28, National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO. Revised, 1994.

Hittelman, A.M., L.W. Row, J.J. Kineman, R.E. Habermann, and D.A. Hastings, 1994: *Global View CD-ROMs and User's Manual (1994 Edition)*. National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, Colorado.

Hittelman, A.M., R.E. Habermann, L. Di, and D. Sinay, 1993: "GeoVu: A Platform Independent, Data Browse and Selection Application," *Proceedings, 2nd International Conference on Integrating GIS and Environmental Modeling, Sept. 1993*, Breckenridge, CO.

Hittelman, A.M., H. Iredale, M. Hollinger, and D. Field, 1994: *Coastwatch Change Analysis Project: Chesapeake Bay Region*, National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO.

Kidwell, Katherine B., ed., 1991: *Global Vegetation Index User's Guide*, National Oceanic and Atmospheric Administration, National Climatic Data Center, Washington, DC.

Kineman, J.J., 1992: *Global Ecosystem Database, Version 1.0 on CD-ROM, User's Guide*, Key to Geophysical Records Documentation 26, National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO.

Kineman, J.J. and M.A. Ohrenschall, eds., 1992: *Global Ecosystem Database, Version 1.0 on CD-ROM, Documentation Manual*, Key to Geophysical Records Documentation 27, National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO.

Klemas, V.V., J.E. Dobson, R.L. Ferguson, K.D. Haddad, 1993: "A Coastal Land Cover Classification System for the NOAA Coastwatch Change Analysis Project," *Journal of Coastal Research*, 9(3), pp 862–872.

Row, Lee W., III, and David A. Hastings, 1994: *TerrainBase: Worldwide Digital Terrain Data: Documentation Manual*, Key to Geophysical Records Documentation 30, Beta Test Release 1.0, National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO.