LinkWinds, A Visual Data Analysis System And Its Application To Remote Sensed Data

Allan S. Jacobson and Andrew L. Berkin
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109
(818) 354-0693 (ASJ), (818) 354-0728 (ALB), (818) 354-0966 (FAX)
budj@apex.jpl.nasa.gov, berkin@krazy.jpl.nasa.gov

The Linked Windows interactive Data System (LinkWinds) is a prototype visual data exploration system resulting from a NASA /JPL program of research into applying graphical methods for rapidly accessing, displaying and analyzing large multivariate multidisciplinary data sets. Running under UNIX, it is an integrated multi-application execution environment allowing the dynamic interconnection and control, through a data-linking paradigm, of multiple windows containing a variety of displays and manipulators. This paradigm, resulting in a system much like a graphical spreadsheet, is not only a powerful method for organizing large amounts of data for analysis, but leads to a highly intuitive, easy-to-learn user interface. It provides great flexibility in rapidly interacting with large masses of complex data to detect trends, correlations and anomalies. The system, containing an expanding suite of non-domain specific applications, provides for the ingestion of a variety of database formats, and hard copy output of all displays. Remote networked workstations running LinkWinds may be interconnected, providing a Multi-User Science Environment (MUSE) for collaborative data exploration by a distributed science team. The system is being developed in close collaboration with investigators in a variety of science disciplines using both archived and real-time data. The application of LinkWinds to optical and infrared multispectral geological images with accompanying elevation data will be discussed.

1. INTRODUCTION

Recent advances in remote sensing capabilities are providing unprecedented ability to study our world. These improvements are also producing an ever increasing flood of data which must be gathered, transported, stored, and analyzed to be fully utilized. This paper reports on a system capable of facilitating the rapid visual analysis of large masses of data, and discusses its application to remote sensing, especially geology. The first portions of this paper review the properties of this system; a more thorough discussion is contained in Jacobson *et al* (1994). Applications to remote sensed atmospheric chemical concentrations (Jacobson and Berkin, 1994) and temperature data (Botts and Spencer, 1994) are found elsewhere. This system grew out of a NASA project at the Jet Propulsion Laboratory to study the application of computer graphics to the problems of quickly and interactively exploring and analyzing very large amounts of scientific data. The objectives of the program are: (1) to develop a software environment which will support the rapid prototyping of visual data analysis applications, while at the same time maintaining the high level of performance necessary for interactively manipulating graphical displays; (2) to develop a user interface that is truly intuitive, allowing quick access to the software for the novice as well as the advanced user; (3) to provide a suite of sample applications which are useful across a variety of scientific disciplines; and (4) to provide tools to support user development of applications for this environment.

2. LINKWINDS

The Linked Windows Interactive Data System, or LinkWinds, is a prototype product of this research effort. In compliance with the research objectives, LinkWinds, an integrated multi-application execution environment with a full graphical user interface (GUI), is a visual data analysis and exploration system designed to rapidly and interactively investigate large multivariate and multidisciplinary data sets to detect trends, correlations and anomalies. The system, operating under UNIX, is based on an object-oriented programming model and is implemented in the C language. It draws upon the Silicon Graphics Inc. (SGI) GL library for its GUI and graphics support software and presently runs only on workstations supporting this library. This includes all SGI workstations, and those of other manufacturers who have licensed and support the GL library. Third party software and the advent of OpenGL as a standard graphics library will greatly increase the portability of LinkWinds in the near future.

Data sets and individual tools for display or control of the data are coded as objects, each occupying a window on the LinkWinds screen, and communicating with other objects through a message passing protocol. The objects or windows can be linked or unlinked at the discretion of the user. Linking the windows sets up one-way message paths between objects. This data-linking paradigm makes the system perform much like a graphics spreadsheet, and as in a spreadsheet, is a powerful way of organizing the data for analysis while providing a natural and intuitive interface. Data-linking, and its user interface implications, are discussed below.

Messages generated by LinkWinds objects are recorded as program statements in an underlying language called Lynx. The message passing characteristics are the basis for three key LinkWinds functions. The first is the maintenance of an internal journal of all user originated commands executed by the environment. This file can be sad at any time through a menu option. The record can then be replayed at the initiation of subsequent LinkWinds sessions, allowing the user to draw upon a previous layout of Link Winds applications and links, or repeat a full analysis session. Lynx also is the basis for a macro capability, in which a series of steps may be stored for re-execution.

The third function based upon the Lynx message passing protocol is the Mu]t,i-User Science Environment (MUSE), which provides a method for multiple Link Winds systems to communicate via networks. Using menu options, users remotely separated can connect to one another, and by also establishing a telephone voice connection, can cooperatively view and manipulate their data. A successful connection requires that each user be executing LinkWinds and that each has access to the data. sets being analyzed. This is normally arranged by transporting the data sets prior to the collaborative session. Because only the messages arc sent, and not the actual data, a very low bandwidth is required, making for quick and efficient communication. The MUSE capability is also used to give tutorials over the network to new users and to allow users to demonstrate recommendations for application changes or to point out bugs.

II ard copy of the LinkWinds displays are provided by function keys on the keyboard. Placing the cursor in a window and pressing F1 produces an image of a window's contents; pressing F2 saves the complete window and frame; and F3 saves the full screen. The figures shown were obtained in this manner. Text files containing statistical information and data table listings arc available via buttons 011 the appropriate applications.

3. DATA-LINKING AND THE USER INTERFACE

111 addition to the normal GUI functions provided by the windowing environment, dynamic manipulation of graphs and images is facilitated through the data-linking paradigm. Data-linking can be understood in the context of a spreadsheet, where cells containing numbers arc linked to other cells. Formulae arc associated with each cell, so that when a number changes, all cells linked to the changed cell recalculate their values. LinkWinds dots the same thing, but in a graphics environment where the rigid grid structure gives way to free form, and a ccl] can translate, for instance, into sliders or large scale number arrays such as images.

This data-linking paradigm is one of the most distinguishing features of LinkWinds, and evolved from a desire to create a truly easy-tc)-lcar)l and intuitive user interface. A guiding principle is that users arc impatient and want to get started 011 productive work as quickly as possible. Large manuals only discourage them (Rettig, 199]). Therefore, an interface was needed which can be learned by exploration, and which conforms to expectations as the user works.

Data-linking is effected through two icons. The link icon is a button displaying; two interlocking rings, while the unlink icon displays two rings that arc separated. Objects on the screen may have a single link button, the full set of link and unlink buttons, or 110 buttons. The presence of a single link button indicates a data object, while the presence of the pair indicates applications with control functions. A window with 110 buttons is an application with only display capabilities. To perform a link, the cursor is placed on the appropriate button, and a "rubber band" is dragged out and dropped into the application to be linked. To break the link, the same thing is done using the unlink button. The rubber bands signifying the links may be displayed at any time during the session. There arc two simple rules to follow in applying the linking paradigm: (1) When as a result of menu selections an empty window appears on the screen, put data in it. This is done by linking a data object into the window. (2) When an object with the pair of link symbols appears, exercise its control function by linking it into any application object.

Our experiences with users has confirmed the intuitiveness of this paradigm. Scientists very quickly become familiar with the manipulations that control LinkWinds, often grabbing the mouse out of our hands during demonstrations to try themselves. New users typically require about 30 minutes of demonstration to understand the system wc]] enough to embark upon productive work.

While the LinkWinds scheme of linking together objects on the screen is reminiscent of data-flow systems, the similarity is only superficial. LinkWinds is a multi-application execution environment and therefore has significant architectural advantages. Any data transformations are done within high level applications which access the data through pointers to shared memory, negating the need to store intermediate copies of the data. This results in more efficient use of memory and better performance. LinkWinds is optimized for execution rather than programming, and with no need for linguistic completeness is less complex and therefore more easily learned.

4. DATABASE INTERFACE

The current version of LinkWinds interfaces with both archived and real-ti)llc data. The archived mode accepts data in a variety of standard formats. These arc:

- Raw binary data in either byte or floating-point format. The data must be arranged as a sequence of images, each in stall-lillc order. 'A n ancillary data description file must also be provided.
- The Silicon Graphics, inc.., native RGB format.
- The Hierarchical Data Format (HDF) created by the National Center for Supercomputing Applications.
- The Common Data Format (CDF) originated at the Goddard Space Flight Center.
- NetCDF, derived from CDF and supported by the National Center for Atmospheric Research.
- The Planetary Data System (PDS) originated at NASA/JPL.
- The Flexible Image Transport System (FITS) widely used by the astrophysics comm nuni ty.

Data sets may be subset upon initial load or subsequently during the session by including contiguous regions of interest. A wider variety of subsetting modes, file searching capabilities, and the import of additional data formats, including the Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) format, arc provided by DataHub (Handley and Rubin, 1993), a system being developed in close collaboration with LinkWinds.

Data sets Lo be ingested by LinkWinds arc listed in a text file and appear in the top level "Databases" menu. Wild cards may be used in this text file, allowing whole classes of data files Lo be accessible. New data files created during a session by DataHub or Link Winds' subsetting tool are automatically added to the menu. Metadata needed to translate data and axis values to meaningful numbers, such as the number of axes and their names, is obtained from data files in the standard formats. A special metadata file exists for the raw byte data or to provide additional information not contained in the standard formats.

The real-time mode of LinkWinds is a recent development (Jacobson and Berkin, 19{13), and must be exercised through the creation of a server tailored to the format of the input data stream. The data description file contains the name of the server, as well as the usual auxiliary metadata information. The server is connected to LinkWinds through UNIX sockets. An interrupt system within LinkWinds senses the arrival of data and notifies the pertinent applications. In addition 10 the actual measured data, clock times and engineering information about the status of the sending devices may be rind. Incoming data is saved in HDF 8-bit format, and may then be replayed by requesting a replay server in the metadata file. Specific data files and time intervals may be chosen for replay. As a test demonstration, LinkWinds was used to monitor and analyze data from the University of Iowa's Plasma Wave Subsystem aboard the Galileo spacecraft during the Earth 2 encounter in December 1992.

5. APPLICATION TO GEOLOGIC DATA

A suite of applications useful across many disciplines has been developed for the LinkWinds environment (sew '1'able 1). Figure 1 shows a typical session exploring TM (Thematic Mapper) data of the Deadman's Butte area of Wyoming. The data is three dimensional, giving the radiance as it varies with ground position and wavelength. The images are 512 by 512 pixels, with each pixel representing 28.5 meters. There are seven wavelength channels, ranging from the visible to the middle infrared. Coregistered elevation data obtained from topographic maps also exists.

The LinkWinds top-level menu appears in the upper left of the figure, From this menu the databases, tools and system options are selected. Data objects, with their single link buttons, are just below the menu. The window entitled P1ane.1 displays a three dimensional representation of the data. Channels 4 (red), 3 (green), and 7 (blue) have been used to make a composite RGB (red, green, blue) false color image, which has been overlaid 011 the elevation data. The height gain is adjustable by the user via the rotary dial on the left, while Pan-Zoom (not shown) and 2-Axis Rotator controls have been linked to the Plane, allowing it to be interactively positioned as desired.

The channels used to make this composite were selected by the three sliders on the LinePlot application at the lower left, which has been linked to the Plane. The LinePlot also displays a spectrum at any point. In this case, the Combine application at the far right has been linked to the LinePlot, which thus displays the spectrum at the location of Combine's crosshair. The Histogram tool at center shows the distribution of radiance values for the three slices used in the composite. Linear stretching has been performed on the values between the white sliders to sharpen the colors.

Photogeologic maps may be constructed from interpretation of images such as this. Geomorphology, determined from both elevation data and shadowing, aids greatly in this interpretation. Different strata are recognized by their colors, corresponding to the different spectra of their lithological components. A full discussion of lithologic information contained in such pictures is discussed by Lang *et al* (1994).

The Combine tool allows mathematical manipulations to be performed on slices of data. The slices are selected from sliders, in this case from the LinePlot. The calculator is called from a button on Combine, and takes both slices and constants as input. In this case, channel 3 has been subtracted from channel 4, and the difference divided by the sum of the two. The result was then rebinned using the embedded histogram, to enhance contrast and focus on values of interest. This manipulation gives the normalized difference vegetation index. Areas with highly vigorous vegetation appear as white, while places with little vegetation are black. Dividing by the sum removes shadow and atmospheric effects.

Figure 2 shows a different LinkWinds session examining both TM and TIMS (Thermal infrared Multispectral Scanner) data of the same region. The TM data object at lower left has been put into the long form by clicking on the "more" button. Using the topmost slider, the data set has been subset to the lowest four channels, as the last three channels were not used during this session.

The TM data is displayed in Image1 at the upper right. Using the LinePlot and Histogram tools as described above, an RGB composite image has again been made, this time with channels 4 (red), 3 (green), and 2 (blue). This particular combination of channels is standard amoung geologists using TM data. Vegetation is displayed in red, while red beds are shown in yellow-gold, and grey rocks appear grey. This color combination was used as the basis for photogeologic mapping of the Deadman's Butte region, as described in Lang *et* al (1987). The Profile tool at the lower left displays the radiance of the red channel (4) along the cyan line on Image. This line is called from Image's menu, and is drawn by the user with any length and angle.

The Plane application to the right of the main LinkWinds menu shows TIMS data rendered on the elevation data. Channels 5 (red), 3 (green) and 1 (blue) have been composited, and their colors stretched using a Histogram tool no longer shown. Lithologic information about the different strata is obtained by variations in the color (Lang *et al*, 1987). The Pan-Zoom and 2-Axis Rotator tools below were used to orient Plane as desired.

In another application of LinkWinds to remote sensed geologic data, a geologist is studying a process called potasium metasomatism using AVIRIS data. In this process, an alteration of the rock occurs, with potassium from groundwater replacing sodium. Potassium metasomatism may thus provide a good indication of faults which are carrying the water. However, the changes are visually subtle, and thus hard to detect in the field. Using LinkWinds, the geologist has been able to interactively scan through a large combination of channels and their ratios, and believes that certain AVIRIS wavelenghts allow the remote detection of potassium metasomatism.

Amoung the tools in Table **1, two** which deserve special comment facilitate the creation of animations for immediate display on the screen or subsequent recording on video or film. One animator is frame based, with the user selecting starting and ending control values, and the number of frames desired in the animation. The other animator is time based. The user sets any number of control positions, each with associated key times graphically selected by moving the hands on a clock. The Animator then interpolates between these set positions to easily make animations with great flexibility in the control. The desired frame rate is selected from the animator menu. Rates include those for film, video, and a screen display mode which makes the animation in real time.

The availability of remote sensed multispectral geologic data, combined with visualization and analysis systems such as Link Winds, thus enables detailed study of areas which may be inaccessible, with a great, saving of time and expense. The advent of instruments with enhanced spatial and spectral resolution, such as those on the Earth Observing System (EOS), and increased computational power will make the application of techniques described above even more important in the future.

6. FUTURE PLANS

Several developments are planned for the future to significantly improve the usefulness of LinkWinds. As in the past, we will continue to develop applications in collaboration with science users seeking to solve real problems. Where rel eva nt, these applications will make use of modern rendering techniques which can successfully be applied in an interactive environment.

A major impedance to the use of any visualization tool is the difficulty users have in inputting their data. Developments of LinkWinds and DataHub will continue to make this process as seamless and automatic as possible. A related issue is the types of data addressed by LinkWinds. Available tools for visual data analysis are generally confined to relatively well-1~c])avd and rectangularly gridded data sets. Link Winds' first venture from the common mold was into the realm o f real-time data, creating a capability to ingest such data and building interactive applications for monitoring and analyzing it. '1'here arc other major neglected categories of data that are quite common in scientific research, and badly in need of tools to support their exploration and analysis. Several problem arms to be addressed in the future arc (]) data sets in which there are significant sources of error, either statistical and/or systematic; (2) data sets which are ungridded samples, either sparse or numerous, from which the user desires to construct gridded data sets over extended regions; and (3) disparate sired data sets from a variety of instruments which must be warped and/or co-registered for overlay or comparison.

In the future, we intend to pursue the development of a users' applications generator for LinkWinds. Currently, the layout of the objects, or widgets, in all of the windows is determined by a text file. The user can re-configure these windows either by editing this file or interactively from a menu-selectable "redesign" mode. We intend to expand this tool kit approach to further allow users to throw widgets away, or add new widgets from a provided catalog. in conjunction, LinkWinds will generate a C code source module to make these widgets work. This code will be suitable for use as a template for the development of a full application. As experience is gained with this approach, and a planned conversion of the code to C++ is accomplished, we anticipate that the rendering and display processes will also lend themselves to a limited catalog of processes selectable by the user.

Aware that the only way to develop useful tools is in conjunction with research on meaningful problems, our policy has been to encourage users and potential users to contact us concerning LinkWinds' changes and needs. We have responded and will continue to do so to the limit of our resources. LinkWinds is currently in use at more than ten institutions, being applied to problems in both remote-sensed and field geology, atmospheric physics and chemistry, meteorology, oceanography, chemical spectroscopy, space plasmas, genetics and cellular biology. As in the past, we will continue to develop applications in collaboration with science users seeking to solve real problems.

Acknowledgments. In addition to the authors, the current LinkWinds development team includes Martin Orton. Alumni who made important contributions are Brian Beckman, Leo Blume, Bonnie Boyd, Jim McLeod, Philip Mercurio, Mark Rubin and Mite.11 Wade. Kathi Beratan and Harold Lang provided useful discussions on remotes sensed geology. We extend our thanks to Joe Bredekamp and his staff at NASA, whose sponsorship makes this work possible. '1 'he research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

Botts, M., and Spencer, R. "The use of LinkWinds for validation] and analysis of 13 years of MSU daily global temperature anomaly data," in Applications of Advanced Data handling and Visualization Tools to Complex Problems in Space and Atmospheric Sciences, 1994 (to appear).

Handley, T.H Jr., and Rubin, M.R. "DataHub: intelligent distributed science data management to support interactive exploratory analysis, in Computing in Aerospace 9, 1993, pp. 91-101

Jacobson, A. S., and A.L. Berkin, A.],. "LinkWinds, a system for interactive scientific data analysis and visualization," in High Performance Computing, 1993, pp. 214-21{}.

Jacobson, A. S., and A.L. Berkin, A.],. "LinkWinds, a visual data analysis system and its application to the atmospheric ozone depletion problem," in Applications of Advanced Data Handling and Visualization Tools to Complex Problems in Space and Atmospheric Sciences, 1994 (to appear).

Jacobson, A.S., Berkin, A.],. , and Orton, M.N. "LinkWinds: interactive scientific data analysis and visualization.') *Comm. ACM* 37 (1994): 42-52.

Lang, 11. R., Adams, S.L., Conel, J.E., McGuffie, B.A., Paylor, E.D., and Walker, R.E. "Multispectral remote sensing as strat igraphic and structural tool, Wind River Basin and Big Horn Basin areas, Wyoming." *AAPG Bulletin* 71 (1987):

389-402.

Lang, H.R., and Paylor, E.D. "Spectral stratigraphy: remote sensing lithostratigraphic procedures for basin analysis, central Wyoming examples." *J. Nonren. Resources* 3 (19{) 4): 25-45.

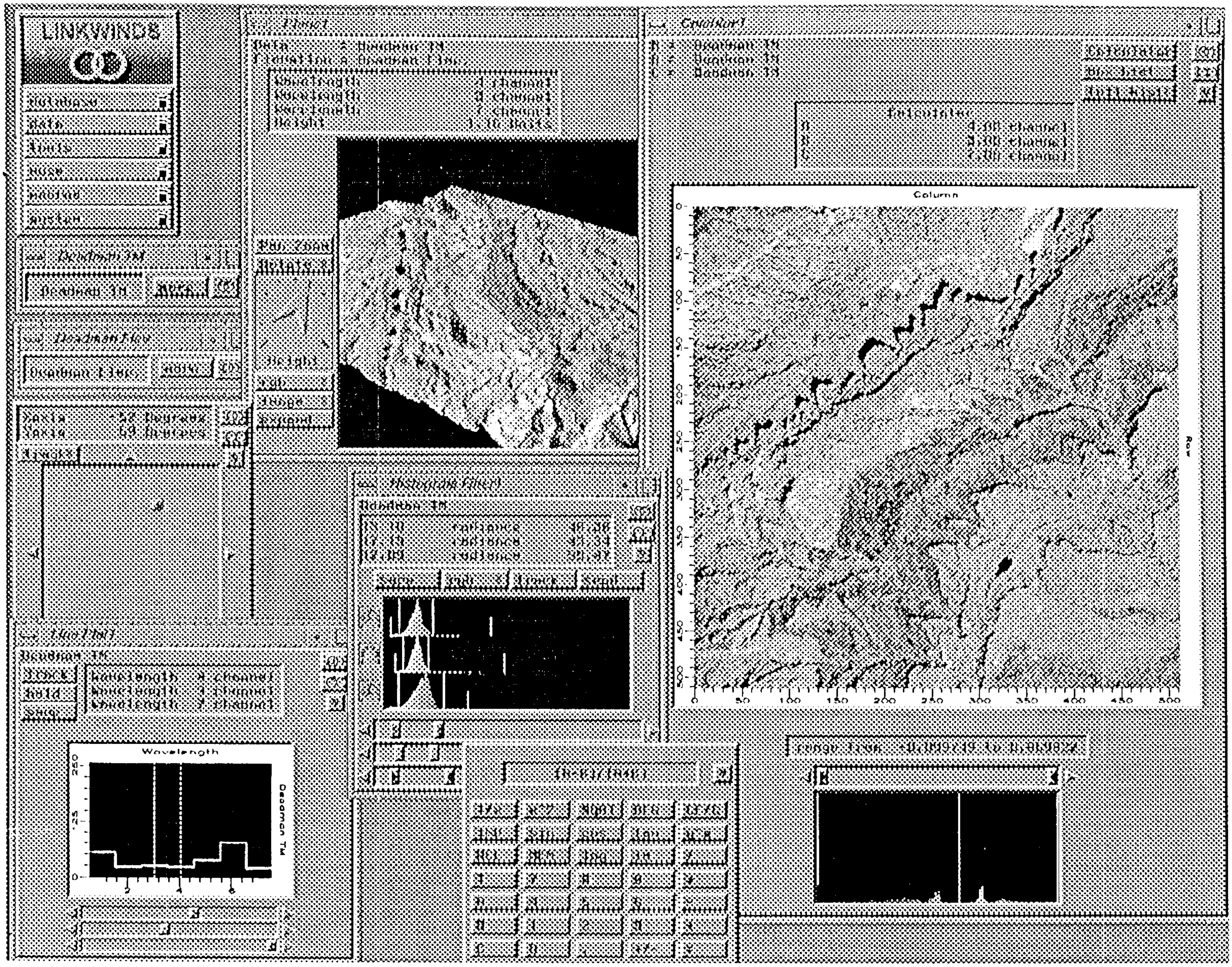
Rettig, M. "Nobody reads documentation." *Comm ACM* 34 (1991): 19-24.

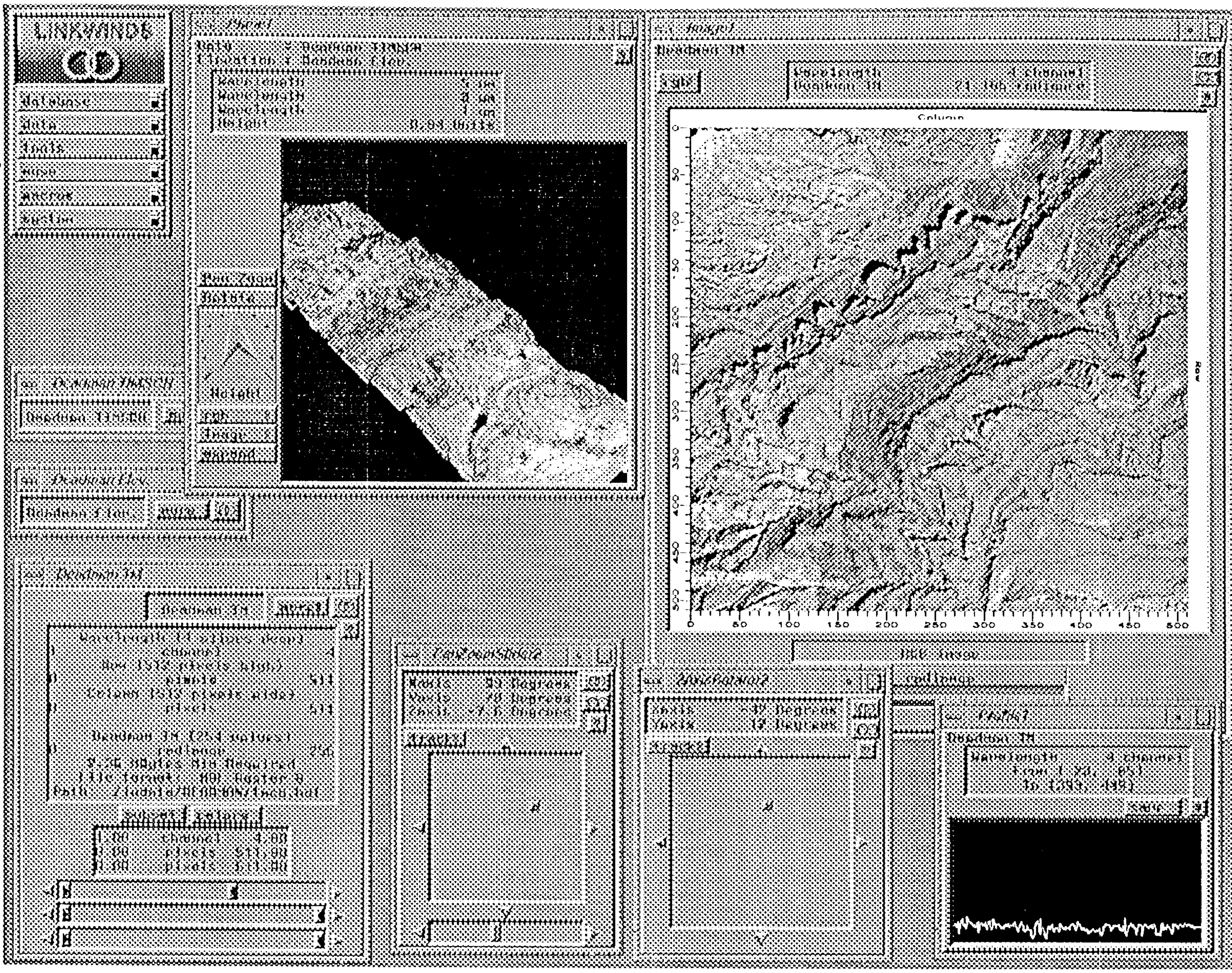
FIGURES AND TABLES

Figure 1. LinkWinds session to explore TM data of Deadman's Butte.

Figure 2. Examining other features of both TM and TIMS data of Deadman's Butte.

Table 1. Current Suite of LinkWinds Applications.


LINKWINDS
Column
Row
Wavelength


LINKWINDS
Column
Row
0 50 100 150 200 250 300 350 400 450 500

	Control
Slider	Controls which slice of data is displayed along any of three orthogonal axes
3-Axis Slider	Three sliders control the slice of data displayed along the three orthogonal axe
3-Axis Rotator	Three sliders, one for each axis, rotate three dimensional applications
2-Axis Rotator	One slider moving in a two-dimensional area rotates three dimensional applications
Pan-Zoom Slider	Controls three dimensional applications by changing the viewpoint
Combine Slider	Six sliders determine three slices of data and their three constant offsets
Animator	Provides time based animations of any application with an arbitrary number of key frames, with a variety of frame rates
Frame Animator	Provides frame based animations of any application, with only the start and stop control values being set
Contour	Computes isovalue contours for rendering on applications which display slices of data
Color '1'001	Interactive data set palette manipulation allowing color editing, data ranges being ramped in color, or substitution of pre-defined palettes
	Display/Control
Combine	Up to three slices of data are combined using standard mathematical functions entered in an embedded calculator
Compare	The functional behavior of each point in a data set is compared with a reference point using a variety of mathematical functions
Data Subset	"Allows the user to interactively save portions of the displayed data into "11 1)F
LinePlot	Plots the values along a straight line going completely through a data set parallel to any axis, and also functions as a slider to select three slices
Histogram	Displays the distribution of values in the 256 data channels for up to three slices, and provides filtering and color stretching
Image	Displays a single slice of data or a composite RGB image of three dices with embedded crosshair, bounding box, and line controls
	Display
Plane	Polygonally renders an image in perspective relief, with an optional accompanying height field, of either a single slice or RGB three slice composite
Globe	Polygonally renders an image on a globe, with an optional accompanying height field, of either a single slice or RGB three slice composite
Polar	Polygonally renders an image on polar projections, with accompanying height field, of either a single slice or RGB three slice composite
Orthoview	Displays in a three dimensional point by point rendering all the values in a data set between two limits
Profile	Displays the data values along a line drawn on the Image, Combine, or Compare tools
2D Scatter Plot	At every location in a slice, plots the values of one data set against the other to show the correlation
3D Scatter Plot	At every location in a slice, plots the values 'of three data sets in three dimensions to show their correlation
TrackPixel	Gives numerical information about the data in Image, Combine, or Compare, both at a point and averaged over a bounding box
	Real-time
StreamPlot	A stripe.]l:irt rcc.order displaying data as either color or line plots, as a function of channel number vertically and time horizontally
StreamLine	Plots data value versus channel number, updating in real time with options of saving a spectrum and averaging in time
StreamPlane	Functions much like StreamPlot, but with the data polygonally rendered in relief using the data values for height and color
StreamClock	Provides timing information for the current data, giving date, day of year, time of day, and internal spacecraft, time
ChannelSlider	Controls the range of channels to be viewed, allowing "" concentration on features of interest