Dataimages And Other Graphical Displays For Directional Data

Bill Morphet
ATK Thiokol Inc., Brigham City, Utah, 84302

Dr. Juergen Symanzik
Utah State University, Logan, Utah, 84322

Author's Footnote

Bill Morphet is an engineer at ATK Thiokol Inc., http://www.atk.com/CorporateOverview/corpover_thiokol.asp, supporting identification, documentation, and resolution of Space Shuttle solid rocket motor anomalies (Bill.Morphet2@ATK.com), and a PhD statistics student at Utah State University.

Abstract

Vectors, axes, and periodic phenomena have direction. Directional variation can be expressed as points on a unit circle and is the subject of circular statistics, a relatively new application of statistics. An overview of existing methods for the display of directional data is given. The dataimage for linear variables is reviewed, then extended to directional variables by displaying direction using a color scale composed of a sequence of four or more color gradients with continuity between sequences and ordered intuitively in a color wheel such that the color of the 0° angle is the same as the color of the 360° angle. Cross over, which arose in automating the summarization of historical wind data, and color discontinuity resulting from the use a single color gradient in computational fluid dynamics visualization are eliminated. The new method provides for simultaneous resolution of detail on a small scale and overall structure on a large scale. Example circular dataimages are given of a global view of average wind direction of El Niño periods, computed rocket motor internal combustion flow, a global view of direction of the horizontal component of earth's main magnetic field on 9/15/2004, and Space Shuttle solid rocket motor nozzle vectoring.

Key Words

Vector Field Visualization/ Geophysics/ Meteorology/ Oceanography/ Circular Statistics/ Color Histogram/ Color Wheel/ Choropleth

1. INTRODUCTION

Directional random variates include types circular, vectorial, axial, and spherical. Circular random variates, or random points on a unit circle, have the total probability of all possible directions represented distributed on the arc of a unit circle. Vectorial random variates in a plane have both random direction and random magnitude. Axial random variates describe random axis direction where there is no reason to distinguish a direction from the opposite direction, e.g., crack or crystal axis orientation. Spherical random variates, or random points on a unit sphere, have the total probability of all possible directions distributed on the surface of the unit sphere. Some applications of directional statistics include: creature departure and vanishing directions after treatment and release (Biology), direction of fault lines or crystal axis orientation (Geology), magnetic field direction (Geophysics), response to a treatment and time of treatment (Chronotherapeutics), wind direction (Meteorology), and ocean currents (Oceanography). Recent books on circular statistics include Fisher (1993), Mardia and Jupp (2000), and Jammalamadaka and Sengupta (2001).

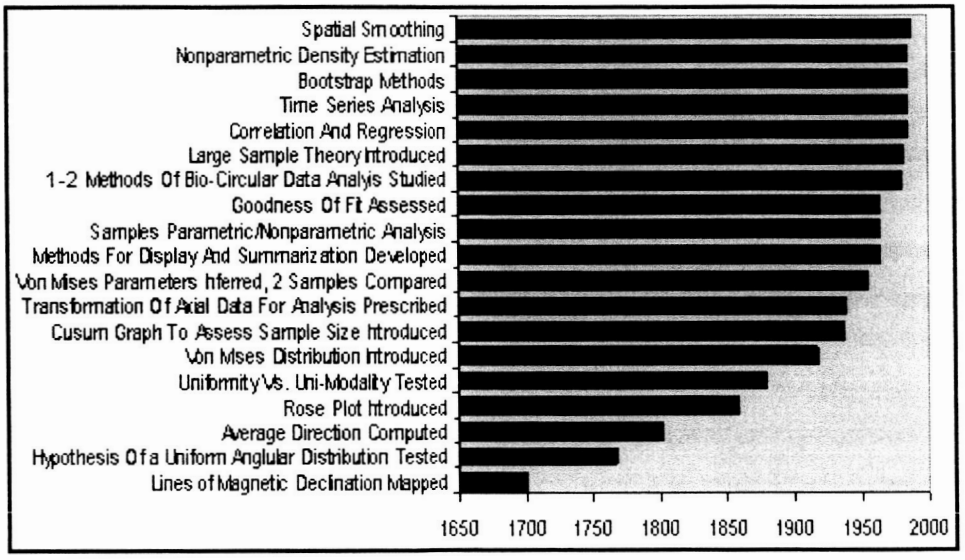


Figure 1, Some Milestones in the History of Circular Statistics. Many of the significant theoretical developments are recent.

Figure 1 summarizes some milestones in circular statistics, as extracted from Fisher (1993, ch. 1). In 1767, John Mitchell, FRS (Fellow of the Royal Society) tested the hypothesis that the distribution of angular separations of stars is uniform. He determined that the number of close stars were too many to support this hypothesis. In 1802, John Playfair noted that directional data should be analyzed differently from linear data, recommending that average direction be the direction of the vector resultant. In 1858, Florence Nightingale, chief nurse in the British Army during the Crimean War, created the rose diagram displaying the effect of sanitation vs. month of year, saving thousands of lives in military hospitals. In 1880, Lord Rayleigh created a statistical test for the hypothesis of the uniform circular distribution vs. the unimodal alternative. In 1918, Von Mises defined the circular normal, or Von Mises distribution, which is a basis of parametric statistical inference for circular data. In 1938, Reiche introduced the Cusum chart, which plots cumulative vector direction and

average magnitude to determine when a sufficient amount of data have been acquired. When the cumulative vector direction and average magnitude of a stable process are consistent, the sample size is adequate. In 1939, Krumbein introduced the transformation of axial to vectorial data for analysis, and back transformation to axial results. The paper by Watson and Williams (1956) about statistical inference for the mean and dispersion of a sample from the Von Mises distribution, and methods for comparing two or more samples, began a period of significant theoretical development. Thus, most of the theoretical developments in the statistics of circular/axial random variation are relatively recent.

This paper is organized as follows: **Section 2** describes the data and software used to generate graphical displays for directional data. **Section 3** is an overview of existing methods for the display of directional and related data. Included are plots for raw directional data, plots to summarize frequency of direction in intervals of a circular-angular scale, smoothed representations of frequency vs. angle, and plots for vector-spatial data. Next, the dataimage for linear variables, which will be extended to directional data, is reviewed. **Section 4** describes a new graphical method providing a high resolution continuous display for directional data. This method employs a color scale composed of a sequence of four or more color gradients with continuity between sequences and ordered intuitively in a color wheel such that the color of the 0° angle is the same as the color of the 360° angle. An example and algorithm for color wheel generation in red green blue (RGB) space is given. **Section 5** provides example circular dataimages including a global view of average wind direction of El Niño periods, simulated rocket motor internal combustion flow, direction of the horizontal component of earth's main magnetic field estimated for 9/15/2004, and Space Shuttle solid rocket motor nozzle vectoring direction angle. **Section 6** concludes this paper finishes with an outline of future work.

2. DATA AND SOFTWARE USED

2.1 Software

Most of the figures were generated in R, version 2.0.1 (R Development Core Team 2004). R was originally created by Ihaka and Gentleman (1996) and is now a collaborative world wide effort. The binary distributions of R and R contributor libraries are freely downloadable from http://www.r-project.org/, and supported on Windows (NT, 95 and later) and in some versions which run on other operating systems. **Figure 2(b)** (enhanced) and the rose plot in **Figure 3(a)** were produced by CircStats, R software for circular statistics (Lund and Agostinelli 2003). **Figures 3(b)-(c)** were produced by Oriana 2, software for analysis and display of circular random variables (Kovach Computing). Oriana is at http://www.kovcomp.co.uk/oriana/oribroc.html and runs on PC/Windows 98/NT or later. **Figure 5** was generated using Fields, R software for analysis of spatial data (Nychka, 2004). **Figure 6** was produced by dataimage, R software for imaging high dimensional data, which is freely downloadable at http://math.usu.edu/~minnotte/research/pubs.html (Minnotte, 1998). **Figures 2(a), 4,** and **7-11** were generated by the author using R.

2.2 Data

Comprehensive Ocean Atmosphere Data Set (COADS) began in 1981 as a cooperative project of the National Climatic Data Center (NCDC), the Environmental Research Laboratories, the cooperative Institute for Research in Environmental Sciences, and the National Center for Atmospheric Research. The objective of COADS is to provide a consistent and easily used historical record of surface marine data from 1854. 70 million unique reports from ships of opportunity and ocean buoys containing 28 variables were organized, and cleaned of outliers and data over land. Trimmed monthly summaries give mean and median statistics for observed air and sea surface temperatures, wind east and north components in meters/second, sea level pressure, humidity, cloudiness, and derived variables. The wind data used in **Figures 2 - 5** were freely extracted from COADS at http://iridl.ldeo.columbia.edu/SOURCES/.COADS/.mean/ for the period from 1980 - 1990, December - March, and in 2° increments for the area of latitude -3° N to +3° N and longitude -93° E to -87° E (+87° W). The extracted data comprises 1934 observations of six variables: month, year, longitude, latitude, and east and north components of wind velocity.

The automotive data used in **Figure 6** are 32 1973 - 1974 model cars extracted from the 1974 *Motor Trend* US magazine (Henderson and Velleman 1981). The dataset comprises 32 observations of 11 variables including weight, displacement, horse power, and miles per gallon.

COADS was renamed International COADS (ICOADS) in 2002 to recognize extensive international collaboration. The wind data used in **Figure 8** were freely extracted from the ICOADS dataset at http://dss.ucar.edu/pub/coads/forms/msg/msga.form.html for El Niño years 1972, 1976, 1982, 1987, 1991, 1994, and 1997, January through April, and in 1° increments for the area of latitude -60° N to +60° N and longitude 0° E to +359° E. The dataset comprises 130,269 observations of the six variables in the COADS dataset with wind speed in units of 0.01 m/sec.

The flow data imaged in **Figure 9** were computed in FLUENT, http://www.fluent.com/, for the Space Shuttle solid rocket motor nozzle at about half way through ignition. The dataset comprises 30,351 observations of five variables (axial and radial coordinates in meters (m), and axial and radial velocity components in m/sec).

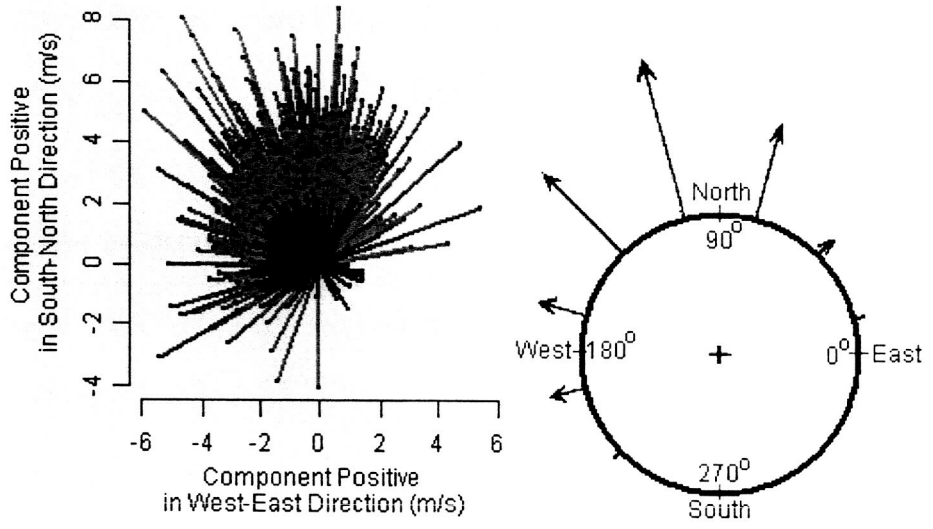
The scientific domain of the National Geophysical Data Center (NGDC) spans the distance from the bottom of the sea to the surface of the sun, providing data describing the marine, solid earth, and solar-terrestrial environments, e.g. earth's magnetic field. The geomagnetic field at any point on the Earth's surface is a combination of several magnetic fields generated by various sources. The main field, which generates more than 90% of the total field, is generated in Earth's outer core. See Frequently Asked Questions are at http://www.ngdc.noaa.gov/seg/geomag/faqgeom.shtml. The vector direction of the north and east components of the main field is shown in **Figure 10**. It was estimated using the International Geomagnetic Reference Field (IGRF) Estimated Values of Magnetic Field Properties at http://www.ngdc.noaa.gov/seg/geomag/jsp/IGRFGrid.jsp for 9/15/2004, elevation 0 km., and in 1° increments for

the area -180° E to +180° E and -90° N to +90° N. The dataset comprises 65,340 observations of four variables (longitude, latitude, and east and north components of the magnetic field in nano Tesla, nT).

The data used in **Figure 11** show the Space Shuttle solid rocket motor nozzle vectoring direction angle, which is the direction in degrees a nozzle is pointing to in the plane perpendicular to its motor axis. The dataset comprises about 425 observations by time from 6 - 22 seconds of motor number, motor side, nozzle direction angle, orbital altitude in nautical miles (nm), and angle in degrees of the motor trajectory relative to the equatorial plane for each solid rocket motor in a subset of the Space Shuttle flights. This figure has been cleared for publication by ATK Thiokol Inc. and National Aeronautics and Space Administration (NASA).

3. OVERVIEW OF EXISTING GRAPHICAL METHODS

3.1 Raw Directional Data



(a) (b)

Figure 2, Vector Plots of the Direction Ocean Wind Blows Toward, from 1980 to 1990, December to March, in the Area Latitude -3° N to +3° N and Longitude -93° E to -87° E. Figure (a) shows both vector magnitude (m/s) and direction of raw data. For clarity, increasing quartiles of vector magnitude are colored blue, green, red, and violet, respectively, and dots replace arrow heads. Figure (b) summarizes the direction wind blows toward by stacking data points in arc bins, and on outside of a circle with degree location on the inside. Arrows are added to emphasize that this is the direction wind was blowing toward.

Existing graphical displays for directional data include arrow or vector plots, circular data plots, the rose plot, circular and linearized-circular histograms, and smoothed circular and linearized-circular histograms. Figures 2 – 5 show wind data from the COADS dataset. Figure 2(a) is a vector plot of the wind data which are used in Figures 2(b) and 3 – 5. It shows magnitude and direction, but not location. For clarity, the longer vectors were plotted first to reduce covering the shorter vectors, increasing quartiles of vector magnitude were colored blue, green, red, and violet, respectively, and the arrow heads of a vectors were replaced with dots. Figure 2(b)

summarizes the circular data of Figure 2(a). The circular data plot is constructed similar to a histogram. The arc bin origin is 0°, bin width is 30°, and number of bins is 12. One point for each observation is stacked on the outside of a circle at the center of its arc bin (Points have also been stacked on the inside of the circle). Analysis: The direction wind is blowing toward is approximately symmetric with mode at about 105° counterclockwise (CCW) from 0° E. Alternatively, the modal direction wind is blowing from is about 285° or 15° from the south toward the east.

3.2 Histograms For Circular Data

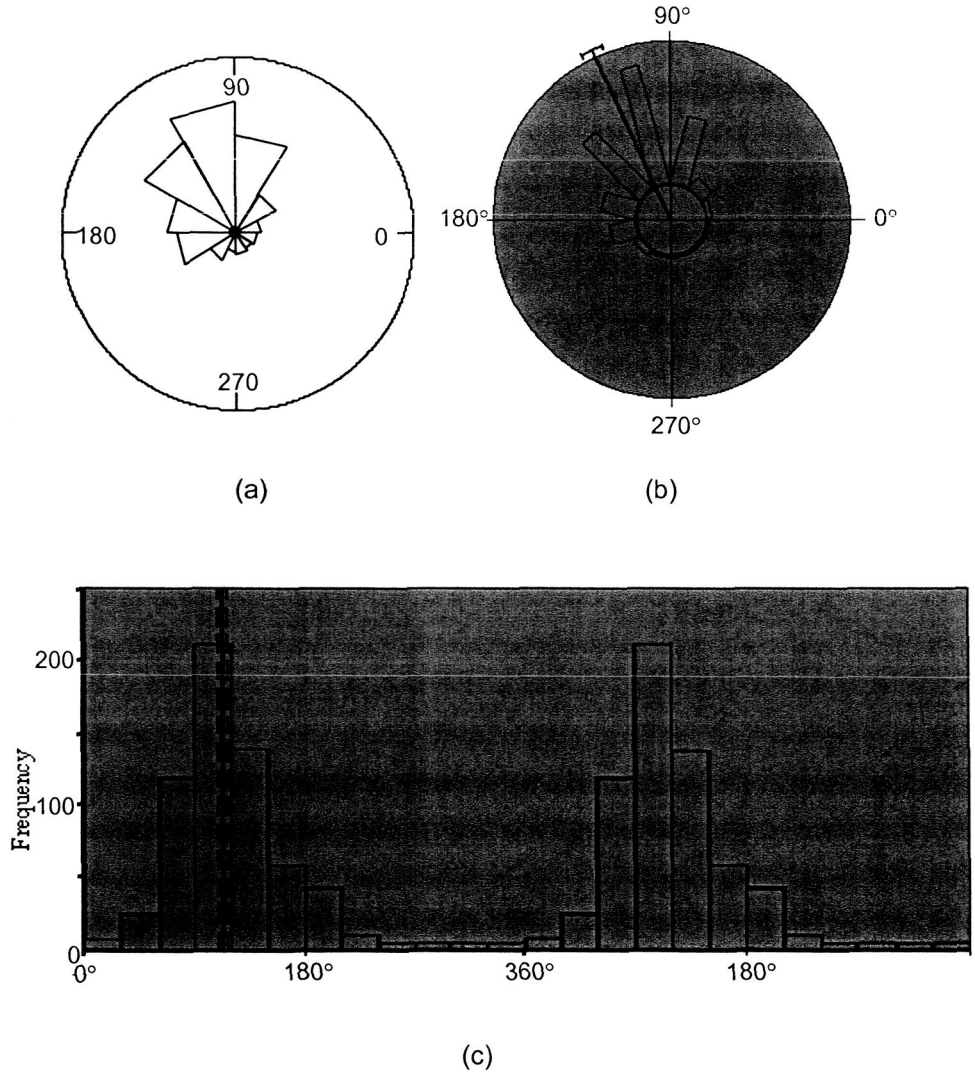


Figure 3, Circular Histograms of the Direction Ocean Wind Blows Toward (0° = East, 90°=North), from 1980 to 1990, December to March, in the Area Latitude -3° N to +3° N and Longitude -93° E to -87° E. The rose plot in Figure (a), circular histogram in (b), and linearized-circular histogram in (c) are types of histograms for summarizing circular data. The linearized histogram in (c) repeats one period of data, and preserves the shape of the distribution at the crossover point of 360°.

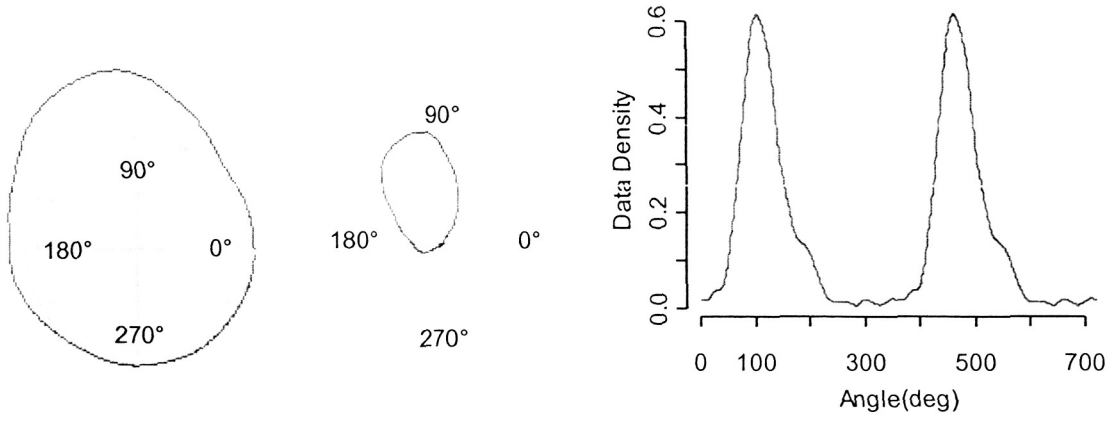
The histogram is a common method of summarizing data. The data are grouped and counted in intervals. A rectangular bar is constructed of area proportional to the count in the corresponding interval and centered above the midpoints of the intervals. The vertical axis of the histogram provides a scale for bar height. With the histogram, we can see the frequency with which data occur relative to value, whether frequency is consistent over the range, has some structure, e.g., or is concentrated at some value. The histogram for circular data is constructed similar to the histogram for linear data. An arc bin origin and arc bin width (e.g. 5°, 10°, or 20°) are selected, and the data grouped and counted. The next step is to wrap the horizontal axis of a the histogram around a unit circle. The frequency bars are aligned with the circle center and arc interval midpoint, plotted on the outside of the circle, and have length or area proportional to their frequency.

The circular histograms in Figure 3 have bin origin 0° and bin width 30°. They summarize the same data as was shown in Figure 2. Figure 3(a) shows a rose plot of wind direction. The angle of the wedge is the bin width and the area of a wedge corresponds to bin frequency. Figure 3(b) shows a circular histogram with a radial bar area corresponding to frequency. Figure 3(c) unwraps the circular histogram of Figure 3(b) onto a linear scale, and repeats one period, which facilitates visual extraction of period, mode count, and prevents breakup of interesting features occurring near the cross over point of 0°. Additionally, Figures 3(b) – (c) show a 95% confidence interval about the mean of (104.8°,107.8°) from east (east=0°, north=90°). These confidence intervals are based on the Von Mises distribution (Fisher 1993, pp. 88 - 89). In Figure 3(b), the mean direction is indicated by a thick black radial line from the center with the confidence interval indicated by a thick black arc slightly outside the large circle. In Figure 3(c), the mean is indicated by a black vertical line enclosed in confidence limits displayed as vertical dashed lines. In Oriana (Kovach Computing, 2004), the confidence interval changes to red when the estimate of the confidence interval is unreliable.

3.3 Nonparametric Density Estimates For Circular Data

The smoothed histograms in Figure 4, following, are based on the same data as was used in Figures 2 - 3. Circular histograms of Figure 3, like histograms for linear variables, can distort the structural information in the sample about the number, sizes, and locations of modes through an arbitrary choice of bin origin and width. The nonparametric smoothed density estimate replaces the bin edge and origin decisions with an easier smoothing band width decision. In nonparametric smoothing, a symmetric unimodal function is centered on each observation. It effectively spreads out the mass of an observation with maximum value at the location of the observation, mimicing the stochastic process where the observation is an instance of random direction and could have occurred at other locations in neighborhood of the observed direction. Increasing the bandwidth increases smoothness and decreases noise. As bandwidth is decreased, more detail is exposed. The smoothed histograms of Figure 4 implement the method given in Fisher (1993, p. 26). Different visual effects result from plotting the estimated data density on a radii of one and zero. Figure 4(a) plots the estimated data density on a unit circle, drawing your attention to regions of high data density (Fisher 1993, p. 25). Figure 4(b) is a polar plot of the same estimated density. It reveals the variation in data density with direction. Figure 4(c) shows the density of Figure 4(b) unwrapped onto a linear scale with an extra period. The latter method removes the issues associated with choice of plotting radius, and like the linearized circular histogram with an extra

period, eliminates breakup of interesting features at the crossover point of 0° and makes it easier to assess structure. Analysis: The linearized histogram in Figure 4(c) indicates that there is one mode at about 103° from east, or 13° from the north toward the west.



(a) (b) (c)

Figure 4, Smoothed Density Plots of the Direction Ocean Wind Blows To, from 1980 to 1990, December to March, in the Area Latitude -3° N to +3° N and Longitude -93° E to -87° E. In Figure (a), the density of wind direction is plotted on the outside of a unit circle, and shows that the wind in this region of space and time tends to blow toward 105°. Figure (b) is a polar plot of the same data density. The effect of plotting radius is eliminated by unwrapping the density plot in Figure 4(b) onto a linear scale in (c). The extra period facilitates assessment of structure.

3.4 Arrow Plots Of Vector-Spatial Data

In arrow plots, vectors of magnitude and direction are represented by an arrow superimposed on a coordinate grid, contour, or surface. The observations of wind data in the arrow plot of Figure 5, following, are the same observations as were used in Figures 2 – 3, and used to construct Figure 4, but Figure 5, shows location in addition to direction and magnitude. Average magnitude and direction are computed from the averages of the horizontal and vertical wind velocity components, respectively, for each location. The raw data are plotted as black arrows and the average direction as red arrows. In this region, the structure of the average wind direction is somewhat continuous and consistent. In the sampled area, the mean wind direction rotates clockwise as latitude increases in the S to N direction, but is approximately constant as longitude increases in the W to E direction.

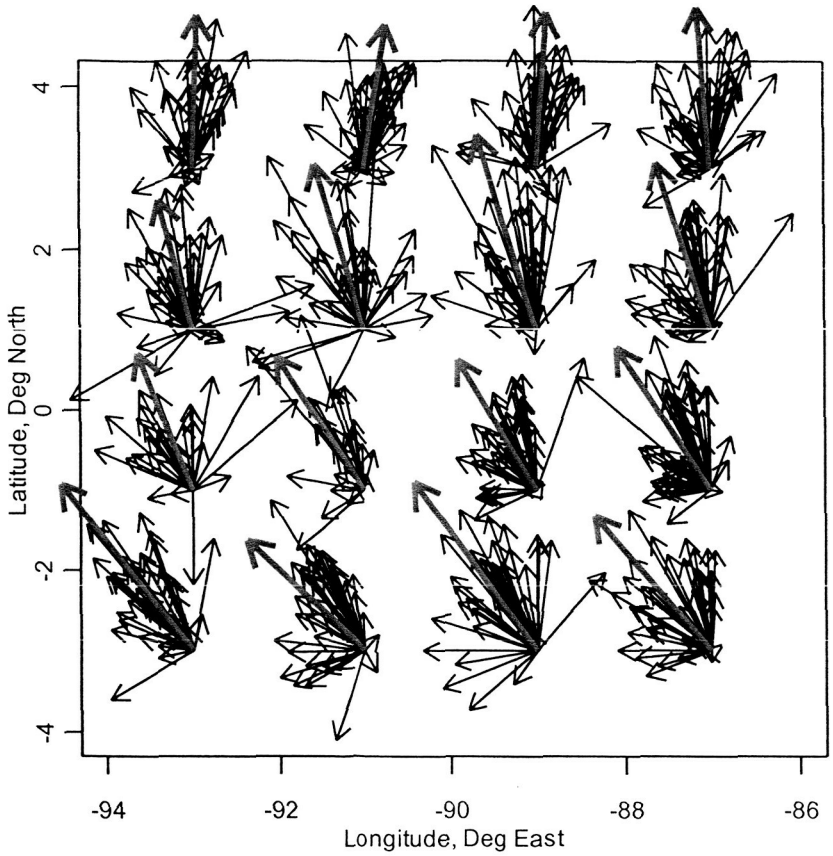


Figure 5, Vector Plot of Vector Magnitude and Direction Ocean Wind Blows Toward by Location, for 1980 to 1990, December to March, in Area Latitude -3° N to +3° N and Longitude -93° E to -87° E. At each sampling location, the raw data are indicated by black arrows and the vector averages by red arrows. The continuity of vector average direction indicates that spatial correlation is greater in the west-east direcition than in the south-north direction in this region, time, and season.

3.5 Dataimages For Linear Data

The dataimage for circular variables, which will be introduced in Section 4, is an extension of the dataimage for linear variables (Minnotte and West 1998), which in turn, evolved from the color histogram of Wegman (1990). The color histogram is a relatively simple plot for visualizing hypervariate structure (variable dimension greater than three). With variables plotted along one axis, observations along the second axis, and the graph area at the intersection of observation and variable colored according to a binned color scale, hundreds of variables and observations can be displayed simultaneously. By sorting the variables, positive and negative associations among variables are easily seen as patterns of color indicating increasing and decreasing trends in values.

In the dataimage for linear variables, the units of observation are plotted on the horizontal axis and variables are plotted on the vertical axis. The capability to image hypervariate structure is enhanced using a linear-color gradient for variable level, and by sorting variables and/or observations using an N-dimensional distance metric. In the red-cyan linear-color gradient, as the level of a variable increases, the amount of red in a pixel decreases

linearly while the level of green and blue increase linearly. Hence, the minimum of a variable is indicated by pure red at maximum saturation and the maximum by cyan, which is green and blue each at maximum saturation. Figure 6 shows a dataimage with a red-cyan linear-color gradient for 1973 - 1974 model car weight (wt), displacement (disp), horse power (hp), and miles per gallon (mpg). The Maserati Bora has the greatest horse power, the Toyota Corolla has the greatest miles per gallon, and the Chrysler Imperial, Lincoln Continental, and Cadillac Fleetwood have the largest weight, displacement, and lowest miles per gallon. Sorting based on all variables and/or observations simultaneously reveals correlation and clusters. This image shows an inverse correlation between miles per gallon and {horse power, displacement, weight}. The two vertical bands indicate two major subgroups of cars: {low weight, low displacement, low horse power, high miles per gallon} at the left, and {high weight, high displacement, high horse power, low miles per gallon} at the right. Further information on dataimages for linear variables such as S-PLUS code and a Java applet may be accessed at http://math.usu.edu/~minnotte/research/pubs.html.

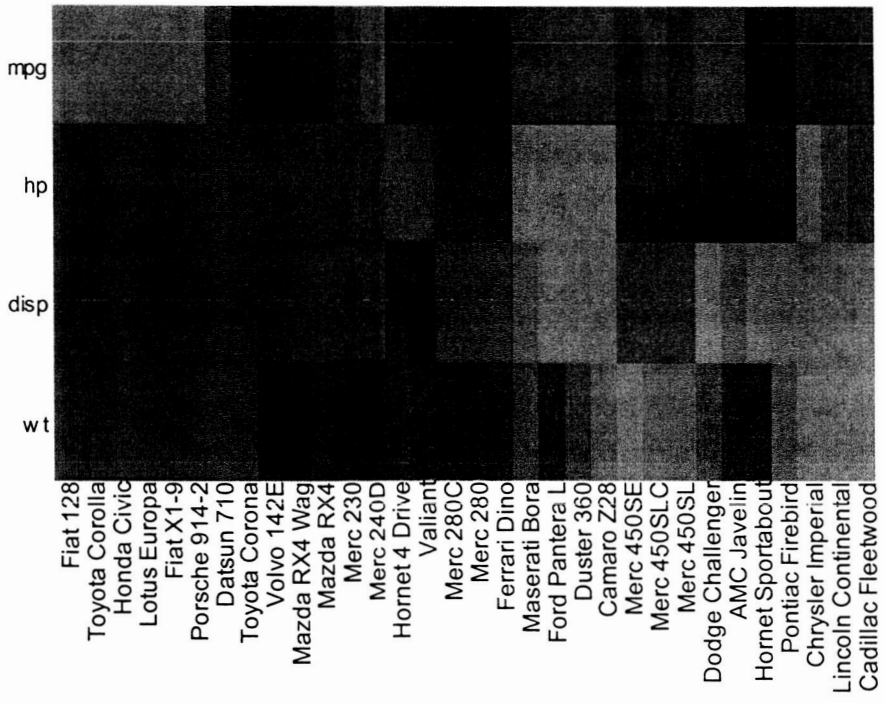


Figure 6, Sorted Dataimage of Variables Mpg, Hp, Disp, and Wt for 32 1973 - 1974 Model Cars Using Red-Cyan Color Gradient (Red=Low, Cyan=High). Two subgroups are shown: large heavy cars with low mileage and high horsepower, and small light cars with high mileage and low horsepower. Hp, Disp, and Wt are are positively correlated, and Mpg is negatively correlated with Hp, Disp, and Wt.

4. A NEW GRAPHICAL METHOD FOR CIRCULAR DATA

4.1 Cross Over Problem, and Image Continuity and Resolution

The average direction of 1° and 359° is not the arithmetic mean of 180°. This problem arose in automating the summarization of wind data and is called the cross over problem. The correct average of 0° or 360° is obtained using vector addition. Cross over also occurs when plotting angle on a linear scale vs. time. The plotted angle can vanish at the bottom of the scale at 0° and instantly reappear at the top at 360° resulting in a discontinuous plot. In a computational fluid dynamics (CFD) visualization, when flow direction from -180° to +180° is colorized with a one color gradient, image discontinuity occurs with variation around 180°. To avoid this discontinuity, users examine vector plots and partial cross sections oriented so as to avoid the cross over point. In turn, this decreases the ability to resolve overall structure and detail simultaneously.

4.2 The Color Wheel For Directional Data, A Periodic Scale

The dataimage of Minnotte and West for linear variables is extended to circular data by displaying the value of direction as a color at the corresponding location on color wheel such as shown in Figure 7, following. The color wheel for directional data is composed of a sequence of four or more color gradients with the same color between gradients, and ending at the beginning color. Using this color scheme, the dataimage of circular data is color continuous through any crossover point, whether approached in the clockwise and counterclockwise direction. This eliminates visual cross over and image discontinuity. The color wheel can also be used to image periodic data. To determine if this method had been previously implemented, CFD visualization literature and software for analysis of circular data were examined: ACUITIV, FLUENT, FIELDVIEW, and CEI, Axis (Pisces Conservation Ltd), CircStats (an R contributor program), Oriana 2 (Kovach Computing), SAS, Surfer 8 (Golden Software), and Vector Rose 3.0 (Paz Software). No examples the display of directional data using the color wheel for circular data have been found to date.

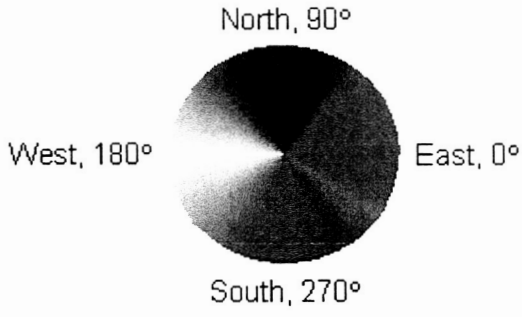


Figure 7, A GBYR Color Wheel Giving Direction as the Color at the Corresponding Degree Location or Geographic Direction. The color wheel for directional data is composed of a sequence of four or more color gradients with the same color joining sequences, and the end color equal to the beginning color. GBYR is an acronym for the sequence Green Blue Yellow Red in CCW order.

4.3 Example Of Color Wheel Computation In RGB Space

The use of color in computer graphics is described in Foley, Van Dam, Feiner, and Huges (1992). Color spaces include RGB (integer levels of red, green, and blue), RGB (levels of red, green, and blue expressed in a hexadecimal value), CMYK (cyan, magenta, yellow, black), HSV (hue, saturation, and value), and CIELAB and CIELUV. CIELAB and CIELUV are efforts to linearize the visual extraction of color differences. R (R Development Core Team) has basic support in HSV and RGB space. The dataimages of Minnotte and West (1998) used a red - cyan linear color gradient. The color wheels in this paper employ four linear-color gradients in RGB space.

Figure 7 is called a GBYR (from first letters of green, blue, yellow, and red) color wheel, because it is composed of the color gradients green → blue, blue → yellow, yellow → red, and red → green in CCW order beginning at the origin, which is East or 0° in Figure 7. This color scheme can be remembered by association with green forests in the east, cold blue sky to the north, yellow deserts to the west, and red hot temperatures to the south. In the wind image in Figure 8, following, wind blowing from south to north is associated with warmer rising air and colored red, and wind blowing from the north to the south is associated with cold falling air and colored blue. In a geographic context, these four colors seem to be more useful and intuitive than three colors, or spectral or divergent color sequences. However, use colors which are intuitive in their application to avoid repetitive reference to a color legend.

Table 1 shows how the GBYR color wheel is implementation in the RGB space:

Table 1, GBYR Color Wheel Implementation in RGB Space

Angle (°)	Color Gradient	Level of Red	Level of Green	Level of Blue
[0,90)	green to blue	0	1 -angle/90	angle/90
[90,180)	blue to yellow	(angle -90)/90	(angle -90)/90	1 -(angle -90)/90
[180, 270)	yellow to red	1	1 -(angle -180)/90	0
[270, 360)	red to green	1 -(angle -270)/90	(angle -270)/90	0

For example, as direction goes CCW from 0° to 90°, the corresponding color is obtained by decreasing the amount of green and increasing the amount of blue linearly.

4.4 Color Considerations

Different numbers of different colors, color orders, nonlinear gradients, and/or color spaces can be more effective. Brewer (1997, p. 210) summarized effective color schemes for scientific visualization including diverging color, which is useful to focus attention on a band of directions, and a spectral sequence of red purple, red, orange, yellow, green, blue, and purple. This sequence has repeating color grading from light to dark within each hue band increasing the number of distinguishable color bins. Adjacent darkest and lightest colors mark hue changes and form visually prominent boundaries through the color distribution. The modified sequence purple, red, orange, yellow, green, blue, and purple divides the angular range into 60° bins while enjoying most of the benefits of the parent sequence. The spectral sequence yellow, orange, red, magenta, blue, green, and

yellow also divides the angular range into 60° bins. However, the combination of red and green cannot be distinguished by about 99% of people with color impairment. Colors for the color impaired are described in http://www.toledo-bend.com/colorblind/index.html. Brewer (1997) recommends the spectral sequence red, orange, yellow, blue green, blue, and purple blue for all red-green color impairments. The Color Brewer for maps, which accommodates linear variables, is at http://www.personal.psu.edu/faculty/c/a/cab38/ColorBrewerBeta.html.

The RGB and HSV color spaces are said to be perceptually nonuniform in scale. CIEluv is a transformation of these color spaces such that distances in the transformed space are perceived to be equal. Ihaka is examining the CIEluv color space as indicated at http://www.stat.auckland.ac.nz/~ihaka/Graphics/index.html. Thus, the CIEluv color space may increase the accuracy with which directions are visually extracted.

5. EXAMPLES OF CIRCULAR DATAIMAGES

5.1 Global Wind Direction

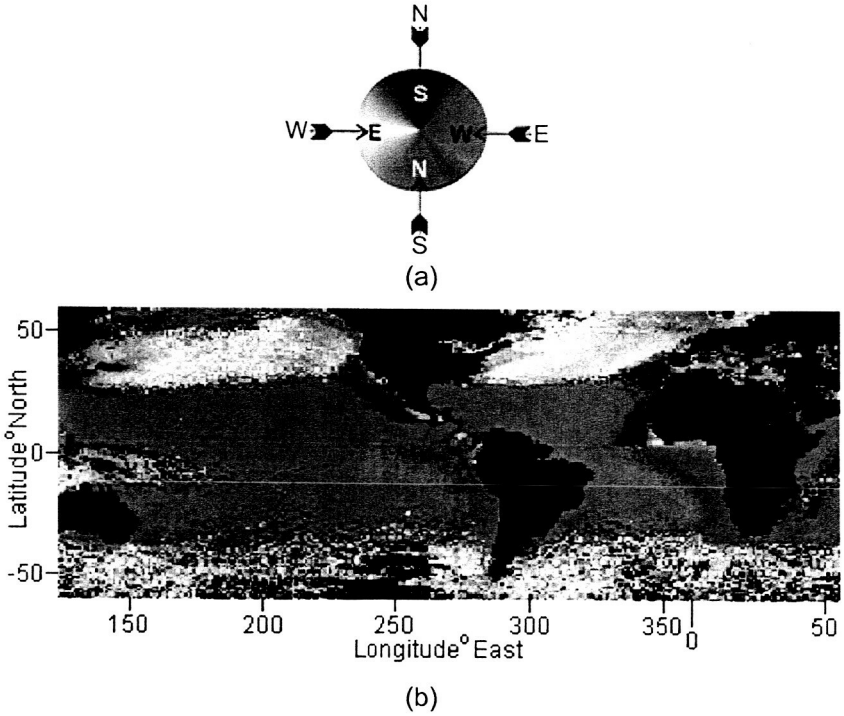


Figure 8(a) GBYR Color Wheel with Arrows Indicating the Direction Wind is Blowing. Figure 8(b) Average Direction Ocean Wind Blows for El Niño Years '72, '76, '82, '87, '91, '94 and '97, January - April. Direction was colored using the YRGB color wheel, but interpreted using Figure 8(a). The predominant average wind direction is east to west inside the equatorial region, and west to east outside. On the west side of both the Americas and Africa, wind blows toward the equator.

Figure 8(a) shows a GBYR color wheel of the direction wind blows used to interpret Figure 8(b). Figure 8(b) shows the average direction ocean wind blows for El Niño years 1972, 1976, 1982, 1987, 1991, 1994, and 1997, January through April, and in 1° increments for area of latitude -60° N to +60° N. El Niño (the child) refers

to the Christmas season when changes in Pacific ocean currents usually begin. These changes are often accompanied by severe climate disruptions to countries in and adjacent to the Pacific with the most severe events having occurred in 1982 and 1997. These data were selected from ICOADS with the intent of using interesting and homogeneous data. Average wind direction was computed as the inverse tangent of the average vertical component of wind velocity divided by the average horizontal component. The resultant direction was actually colorized by a YRGB (Yellow Red Green Blue going CCW) color wheel, but is easily interpreted with the color wheel in Figure 8(a). This was done so that wind blowing from the south of the equator is on the same side of the color wheel. Swapping yellow and green may improve the perception of detail in the latitudes around the equator. Areas of missing data or land mass are colored black. Analysis: Areas at the bottom of Figure 8(b) with little or no data have more variation in the averages, but overall, the wind blows west to east. An ocean and continent away, the pattern of direction on the west side of the Americas is similar to that on the west side of Africa: wind flows toward the equator from both the north and the south, then turns to the west. This suggest that this flow is not random, but a systematic process involving interaction with the continents.

5.2 Rocket Motor Internal Flow

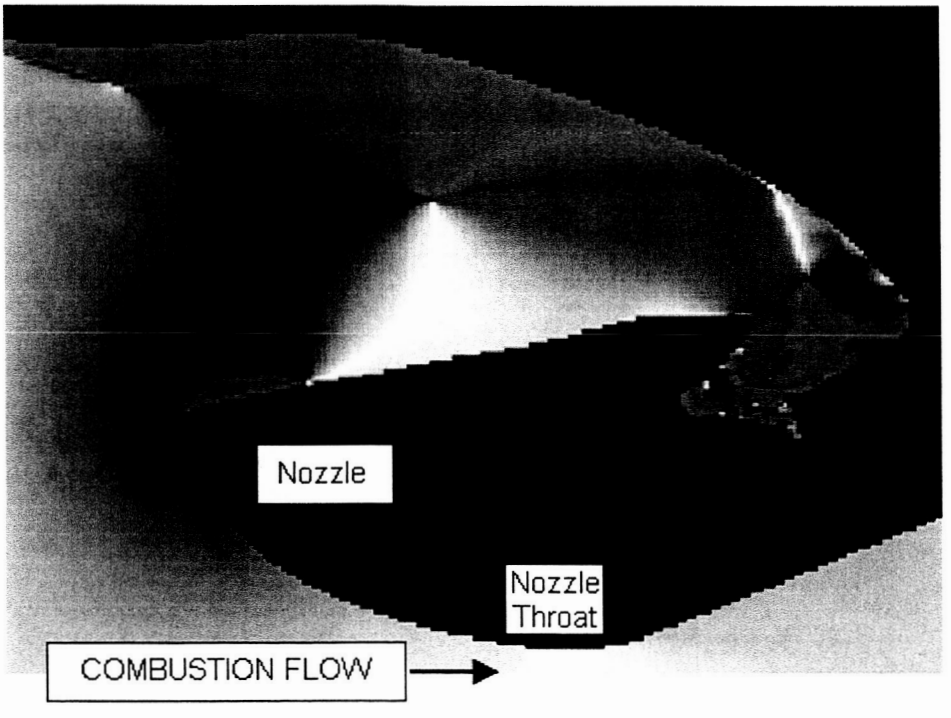


Figure 9, Solid Rocket Motor Axisymmetric Internal Combustion Flow in a Plane through the Nozzle of the Space Shuttle Solid Rocket Motor. The flow vectors were computed by FLUENT, software for computational fluid dynamic analysis. Direction of flow is interpreted with the color wheel in Figure 8(a), the arrows giving the direction of flow. This image has two counter-rotating vortices.

Figure 9 shows the contour of the Space Shuttle solid rocket motor nozzle. Combustion products entering the nozzle compress with maximum compression occurring at the throat where nozzle diameter is minimum. Aft of the throat (to the right), nozzle diameter increases. Gases exiting the throat expand, increase in velocity, and produce thrust. With downward flow = blue, upward = red, forward =green, and aft =yellow, the YRGB CCW pattern in the cavity above the nozzle indicates a CCW flow. The smaller pattern at the right end of the cavity indicates a CW flow. These two vortices mesh like oppositely rotating gears.

5.3 Direction of Horizontal Component of Earth Main Magnetic Field On 9/15/2004

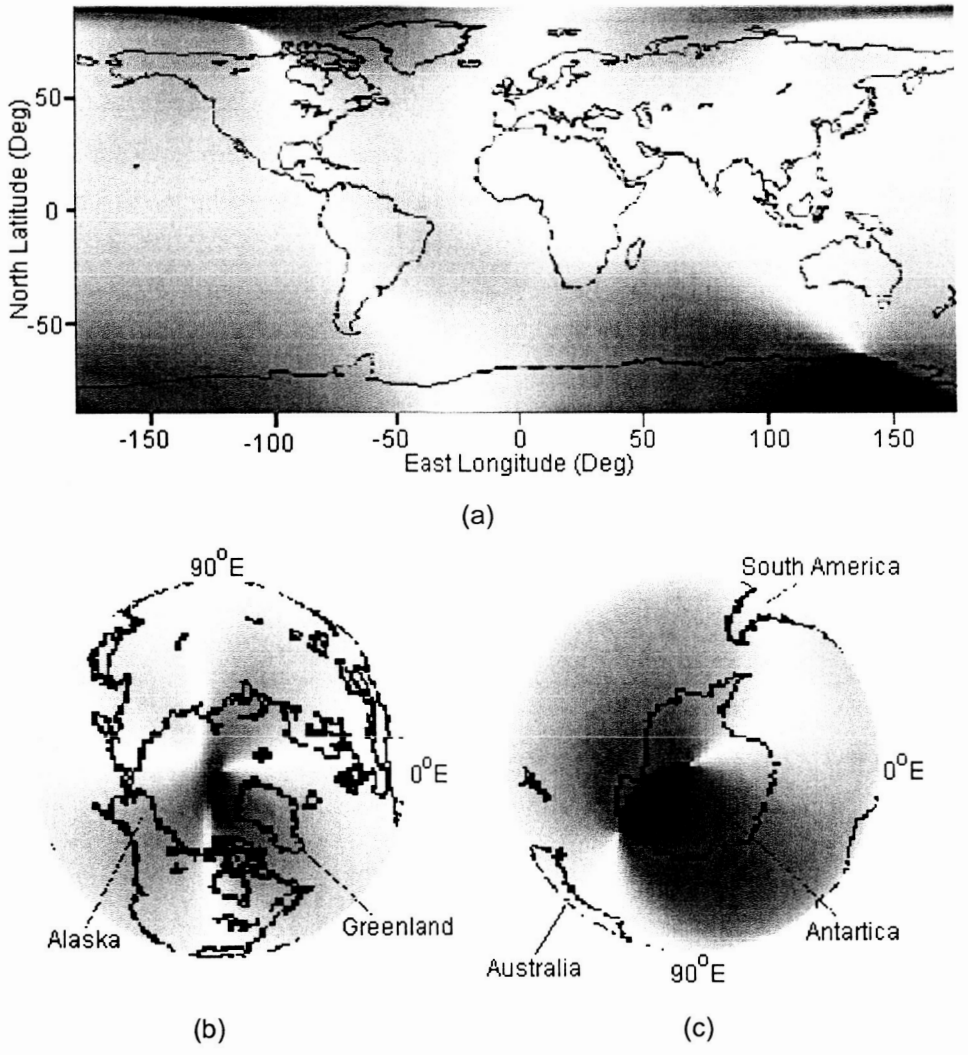


Figure 10, Direction of Earth Main Magnetic Field Using East and North Components Estimated by IGRF Model for Elev. 0 km on 9/15/2004 Using a GYRB Color Wheel (Green=East=0°, Yellow = North=90°, Red=West=180°, Blue=South=270°). Figure (a) shows direction on a planar grid of resolution 1° longitude x 1° latitude. Because the data are defined on a sphere, pattern distortion in (a) increases with distance from the equator. Figures (b)-(c) are perspectives of the circular dataimage on a sphere. Figure (b) is the view looking from the north geographic pole perpendicular to equatorial plane. Figure (c) is the view looking from the south geographic pole perpendicular to equatorial plane.

Figures 10(a) - (c) show the vector direction of the east and north components of earth's main magnetic field, to which a compass needle aligns, using a GYRB color wheel. Yellow was chosen for the dominant north color because it seems easier to focus on yellow and discern subtle color variations than if blue were used. The choice of east=green and west=red results in light green and red color shadings around yellow.

The rectangular representation on a grid of longitude and latitude in Figure 10(a) distorts the pattern of direction, especially at the poles. Figures 10(b) - (c) corrects this distortion by mapping the color of the magnetic field direction onto a sphere and displaying it in perspective. This is a spherical dataimage of a directional data. Figure 10(b) shows the northern hemisphere as viewed from the north geographic pole perpendicular to the equatorial plane. Starting at the center and going outward, latitude decreases from +90° N to zero at the equator. Figure 10(c) shows the southern hemisphere as viewed from the south geographic pole perpendicular to the equatorial plane. Continuing at the outside of Figure 10(c) and going toward the center, latitude decreases from 0° N to -90° N.

The spherical dataimage of directional data provides a new way to image remotely sensed terrestrial and extraterrestrial vector fields. The radius of the surface at a point (longitude, latitude) can be used for a linear covariate, e.g. magnitude of radial or orthogonal component of earth's main magnetic field.

5.3 Space Shuttle Solid Rocket Motor Nozzle Direction Angle

The Space Shuttle solid rocket motors are the largest in the world at 126 feet long and 12 feet in diameter. Direction angle is the angle a solid rocket nozzle is pointing to in the plane perpendicular to the motor axis. Figure 12, following, shows direction angle colorized using the RGBY color wheel in Figure 11.

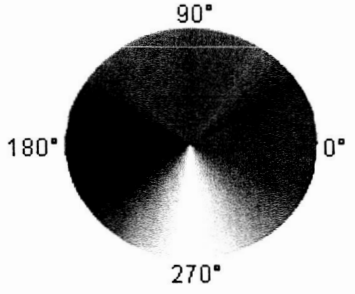


Figure 11, RGBY Color Wheel for Space Shuttle Solid Rocket Motor Nozzle Vectoring in Figure 12.

In Figure 12, each narrow horizontal band of colors represents a circular time series of direction angle with time increasing left to right time=6 sec to time=22 sec of ignition. Like the dataimage for linear variables of motor trend cars in Figure 6, the observations were sorted to show structure hidden by ordering of the data. The horizontal strips (units of observation, motors) were vertically sorted: first by location of the motor (left side of Space Shuttle, then right side), second by angle of the Space Shuttle to the plane through the equator (inclination), and last by orbital altitude. The resulting prominent red horizontal band shows that the left side rocket nozzle tends to vary about the 0° location (In Figure 11, red=0°), and the equally prominent blue

horizontal band above the red band shows that the right side rocket nozzle tends to vary about the 180° location. Given the assembly orientation of the motors, this means simply that the two nozzles are pointed toward earth. The nonrandom diagonal structures between about 14 and 22 seconds, which were revealed by sorting, reflects a relationship between the linear variables of inclination and altitude, and the circular variable of nozzle direction angle. The vertical bands centered at about 10 seconds reflect turning of the rocket nozzles to initiate the Space Shuttle roll maneuver as illustrated in Figure 13, following. As given by A.Brown at http://www.aerospaceweb.org/question/spacecraft/q0127a.shtml,
the roll maneuver orients the cargo bay towards the Earth to satisfy scientific and Space Shuttle engineering requirements, and provides the astronauts a spectacular view of earth.

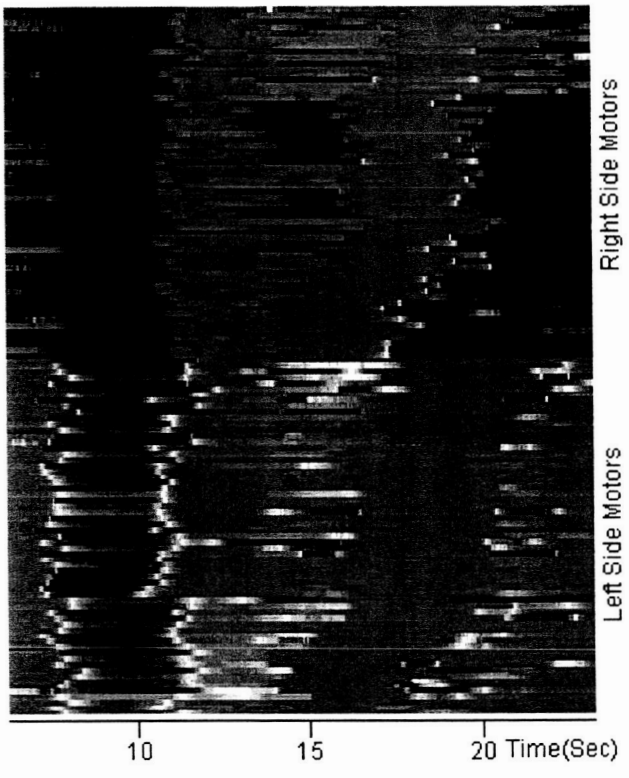


Figure 12, Space Shuttle Solid Rocket Motor Nozzle Direction Angle vs. Time. Each horizontal strip is the nozzle direction angle vs. time of a motor. Nozzle direction was colorized using the RGBY color wheel of figure 11. The horizontal strips are ordered by side (left, then right), second level sort by inclination, and third level sort by altitude. The vertical blue pattern over about 10 seconds is the turning of the left nozzle to initiate roll maneuver as illustrated in Figure 13. The diagonal structures over 14 - 22 seconds indicate a relationship between the circular variable of nozzle direction, and the linear variables of orbital altitude and angle of the trajectory relative to the equatorial plane.

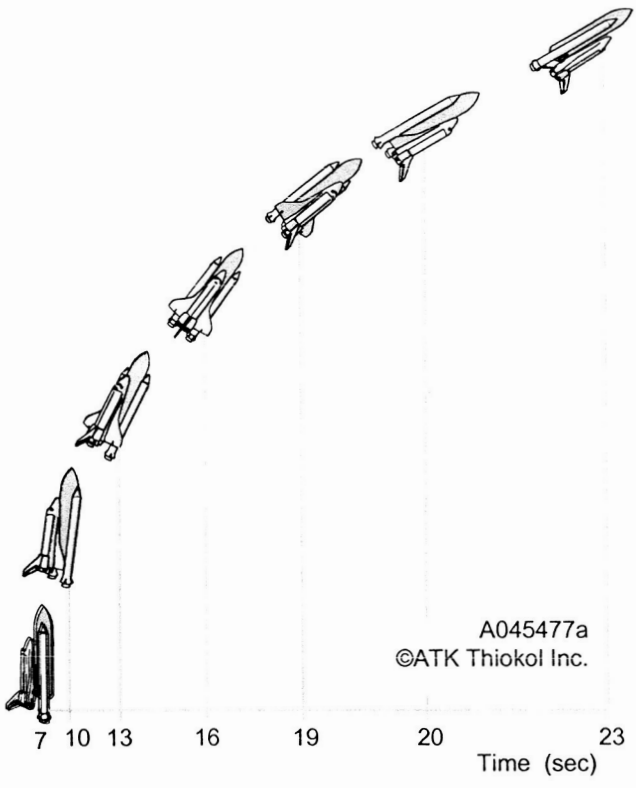


Figure 13, Illustration of Space Shuttle Roll Maneuver vs. Time from Ignition. Printed by permission of ATK Thiokol Inc.

6. SUMMARY AND FUTURE WORK

The dataimage for linear variables was extended to circular variables by representing direction as a color from the corresponding angle of a color wheel constructed of a sequence of four continuous color gradients such that the last color is also the first. This eliminates the cross over, and provides for a continuous image of circular data at high resolution. On a large scale, both detail and overall structure can be visually extracted. Examples included earth views of wind and direction of east and north components of the main magnetic field, rocket motor internal combustion flow, and a circular time series of the direction of rocket nozzle vectoring. Future work includes further development of color wheels, development of a dynamic-interactive graphic user interface (GUI) with linked views, brushing, interrogation of the data, and application of methods for analysis of circular-spherical spatial data. Analysis projects include animation of images of earth main magnetic field for a sequence of years, and statistical analysis of Pacific ocean and wind currents for La Nina and El Niño periods with emphasis on the vector correlation of wind and ocean currents.

ACKNOWLEDGEMENTS

The author acknowledges and expresses thanks to **ATK Thiokol Inc.** and **NASA** for authorization (date TBD) to display computed flow in the Space Shuttle solid rocket motor nozzle in Figure 9 and for authorization (11/19/2004) to display nozzle vectoring data in Figure 12, and to **Technical Artist Alan Eaton** (#A045477a, ATK Thiokol Inc, copyright © 2005) for Figure 13 showing the Space Shuttle roll maneuver.

REFERENCES

Brewer, C. (1997), "Spectral Schemes: Controversial Color Use on Maps," *Cartography and Geographic Information Systems,* 24 (4), pp. 203 - 220.

Fisher, N. (1993), *Statistical analysis of circular data*, Cambridge: Cambridge University Press.

Foley, J., Van Dam, A., Feiner, S., and Huges, J. (1992), *Computer graphics : principles and practice*, Reading, MA : Addison-Wesley Pub. Co.

Henderson, H., and Velleman, P. (1981), "Building Multiple Regression Models Interactively," *Biometrics,* 37, pp. 391 - 411.

Ihaka, R., and Gentleman, R. (1996), "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics,* 5 (3), pp. 299 - 314.

Kovach Computing (2004), Oriana 2 at http://www.kovcomp.co.uk/oriana/oribroc.html.

Lund, U., and Agostinelli, C. (2003), CircStats: Circular Statistics. R package version 0.1-8. S-plus original at http://statweb.calpoly.edu/lund/.S-plus original by Ulric Lund and R port by Claudio Agostinelli.

Jammalamadaka, S., and Sengupta, A. (2001), *Topics in circular statistics*, Singapore : World Scientific.

Mardia, K., and Jupp, P. (2000), *Statistics Of Directional Data*, New York: John Wiley & Sons.

Minnotte, M. (1998), dataimage: software for imaging high dimensional data at http://math.usu.edu/~minnotte/research/pubs.html.

Minnotte, M., and West, R. (1998), "The Data Image: A Tool for Exploring High Dimensional Data Sets," American Statistical Association 1998 Proceedings of the Section on Statistical Graphics, pp. 25 - 33.

Nychka, D. (2004), fields: Tools for spatial data. R package version 1.7.2. at http://www.cgd.ucar.edu/stats/Software/Fields.

R Development Core Team (2004), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Watson, G., and Williams, E. (1956), "On the Construction of Significance Tests on the Circle and the Sphere," *Biometrika,* 43, pp. 344 - 352.

Wegman, E. (1990), "Hyper-dimensional Data Analysis Using Parallel Coordinates," Journal of the American Statistical Association, 85, pp. 664 - 675.

DATA SOURCES

COADS data extraction at http://iridl.ldeo.columbia.edu/SOURCES/.COADS/.mean/
ICOADS data extraction at http://dss.ucar.edu/pub/coads/forms/msg/msga.form.html
IGRF web site at http://www.ngdc.noaa.gov/IAGA/vmod/