An Interactive Parallel Coordinates Technique Applied to a Tropical Cyclone Climate Analysis

Chad A. Steed [a,*] Patrick J. Fitzpatrick [c] T.J. Jankun-Kelly [b]

Amber N. Yancey [d] J. Edward Swan II [b]

[a] *1005 Balch Blvd., Naval Research Laboratory, Stennis Space Center, MS, USA, (voice) 228-688-4558, (fax) 228-688-4853*

[b] *Department of Computer Science, Mississippi State University, MS, USA*

[c] *Northern Gulf Institute, Mississippi State University, Stennis Space Center, MS, USA*

[d] *Department of Physics and Astronomy, Mississippi State University, MS, USA*

Abstract

A highly interactive visual analysis system is presented that is based on an enhanced variant of parallel coordinates – a multivariate information visualization technique. The system combines many variations of previously described visual interaction techniques such as dynamic axis scaling, conjunctive visual queries, statistical indicators, and aerial perspective shading. The system capabilities are demonstrated on a hurricane climate data set. This climate study corroborates the notion that enhanced visual analysis with parallel coordinates provides are deeper understanding when used in conjunction with traditional multiple regression analysis.

Key words: parallel coordinates, hurricane, climate study, multivariate information visualization, geovisualization

1 Introduction

In climate studies, scientists are interested in discovering which environmental factors influence significant weather phenomena. A prominent weather feature is a *tropical cyclone*, defined as a warm-core non-frontal synoptic-scale cyclone, originating over tropical or subtropical waters, with organized thunderstorms and a closed surface wind circulation. Tropical cyclones begin as a tropical depression, with sustained 10-meter winds less than 17 ms^{-1}. Most intensify into *tropical storms* (sustained winds between 17 and 32 ms^{-1}). 56% of tropical cyclones reach winds of at least 33 ms^{-1}, and are then designated with regional terms such as *hurricanes* in the Atlantic basin, and *typhoons* in the Western North Pacific Ocean. When sustained 10-meter winds reach 49 ms^{-1}, they are called *intense hurricanes* in the Atlantic.

Tropical cyclone activity in each ocean basin can vary on a yearly scale as well as a multidecadal scale due to large-scale atmospheric influences and climate forcing. As a result, scientists are developing procedures to forecast whether an upcoming tropical cyclone season will be active, normal, or below normal. Others are studying causes of multidecadal cycles, and whether anthropogenic global warming is also an influence (Landsea, 2005). Recent destructive tropical cyclones seasons have escalated these research efforts.

* Corresponding author.
Email addresses: csteed@nrlssc.navy.mil (Chad A. Steed), fitz@ngi.msstate.edu (Patrick J. Fitzpatrick), tjk@acm.org (T.J. Jankun-Kelly), anb130@msstate.edu (Amber N. Yancey), swan@acm.org (J. Edward Swan II).

Several atmospheric and climate variables impact the intensity and frequency of seasonal storm activity. Identifying the most critical environmental variables help scientists generate more accurate seasonal forecasts which, in turn, improve the preparedness of the general public and emergency agencies. One useful method for predicting and understanding the seasonal variability in tropical cyclones is multiple regression. Predictors are chosen from historical tropical cyclone data (Vitart, 2004), and provide an ordered list of the most important predictors for the dynamic parameters.

[Fig. 1 about here.]

In conjunction with statistical analysis, researchers have relied on simple scatter plots and histograms which require several separate plots or layered plots to analyze multiple variables. Using separate plots, however, is not an optimal approach in this type of analysis due to perceptual issues such as change blindness (a phenomenon described Rensink (2002)), especially when searching for combinations of conditions. The scatter plot matrix is a more useful technique employed by statisticians to uncover patterns in multivariate data that contains all the pairwise scatter plots of the variables on a single display in a matrix configuration; but it requires a large amount of screen space and forming a multidimensional association from a set of two-dimensional displays is mentally challenging. Although layered plots condense the information into a single display, there are significant issues due to occlusion and interference as demonstrated by Healey et al. (2004). Furthermore, the geographically-encoded data used in climate studies are usually displayed in the context of a geographical map; although certain important patterns (those directly related to geographic position) may be recognized in this context, additional information may be discovered more rapidly using non-geographical information visualization techniques. Due to the multivariate nature of climate study data, researchers need interactive visualization techniques that can accommodate the simultaneous display of many variables.

[Table 1 about here.]

This paper discusses the application of a popular multivariate information visualization technique, parallel coordinates, to a tropical cyclone climate study and regression analysis. With parallel coordinates, n-dimensional data is represented as a polyline where its n-points are connected in n parallel y-axes. The resulting visualization provides a compact two-dimensional representation of even large multivariate data sets (Siirtola, 2000). In this research, several previously introduced interactive parallel coordinate extensions have been combined into a unique application for climate analysis. This paper also discusses how these techniques increase the scientists' ability to discover the relationships between dependent and independent variables. Using a climate study data set that consists of several seasonal tropical cyclone predictors, it is shown that parallel coordinates provides a useful representation of multiple regression analysis. The results suggest that parallel coordinates can be used as an alternative method for finding relationships among a set of variables, and the technique can be used in conjunction with stepwise regression to enhance and speed up the relationship discovery process.

2 Related Work

The parallel coordinates visualization technique was first introduced by Inselberg (1985) to represent hyper-dimensional geometries. Later, Wegman (1990) applied the technique to the analysis of multivariate relationships in data.

Since then, several innovative extensions to the technique have been described in the visualization research literature.

The system described in this paper implements a dynamic axis re-ordering capability, axis inversion, and some details-on-demand features similar to those described by Hauser et al. (2002). In addition, some interactive visual query and frequency representation (histogram) capabilities described by Siirtola and Räihä (2006) are included, as well as a variant of the interactive aerial perspective shading technique described by Jankun-Kelly and Waters (2006). The system also includes a focus+context technique for axis scaling that is similar to the capabilities described by Fua et al. (1999), Artero et al. (2004), Johansson et al. (2005), and Novotńy and Hauser (2006).

The system also provides dynamic query capabilities based on the double slider concept of Ahlberg and Shneiderman (1994). The PCP axes also display important frequency information between the double sliders in a manner similar to the Influence Explorer described by Tweedie et al. (1996). More recently, Siirtola and Räihä (2006) implemented these visual query mechanisms with parallel coordinates.

The visual analysis software described in this paper provides a unique parallel coordinate based interface by fusing variants of the above mentioned capabilities. Moreover, this research describes one of the most in-depth validations of enhanced parallel coordinate plots for use in climate analysis

Multiple regression traditionally has been used to identify statistically significant variables from multivariate data sets, including tropical cyclones data sets. Klotzbach et al. (2006a) use this technique to determine the most important variables for predicting the frequency of tropical cyclone activity for the

North Atlantic basin. Similarly, Fitzpatrick applied stepwise regression analysis to the prediction of tropical cyclone intensity (Fitzpatrick, 1996, 1997). It will be shown that multiple regression and interactive parallel coordinates can complement each other, with the regression identifying the relevant associations and the interactive software highlighting additional features of the variables.

3 Climate Study Data Set

This research analyzes a data set containing potential environmental predictors for a tropical cyclone climate study. This data set was provided by the Tropical Meteorology Project at Colorado State University (Klotzbach, 2007), and is used to predict the frequency of Atlantic tropical cyclones for the upcoming hurricane season by categories. These categories include: 1] number named storms (winds 33 ms^{-1} or more, at which tropical cyclones receive a "name"); 2] number of hurricanes; and 3] number of intense hurricanes. These variables have known relationships to Atlantic tropical cyclone activity. For example, the North Atlantic basin has fewer tropical cyclones during El Niño Southern Oscillation (ENSO) years, and active seasons in La Niña years (Chu, 2004). Because of this relationship, scientists use ENSO signals as some predictors of seasonal storm activity. Scientists at the Tropical Meteorology Project issue six forecast reports based on statistically significant predictors from this data set.

[Table 2 about here.]

Table 2 lists 16 potential environmental predictors from the data set along with their geographical region. In the remainder of this section, the physical relationships of these climate variables to Atlantic tropical cyclone activity are discussed.

3.1 El Niño Variables

In a normal year, air rises in the western tropical Pacific (where the water is the warmest as well as slightly elevated) and sinks in the eastern tropical Pacific which is a phenomenon known as the Walker Circulation. During an El Niño event, the easterly surface trade winds that cause this water bulge in the western Pacific weaken, and the warm water travels eastward. Furthermore, El Niño conditions shift the upward portion of the Walker Circulation to the eastern Pacific, creating upper-level westerly winds in the Atlantic Ocean as well as subsidence. Both of these factors inhibit tropical cyclone formation and intensification in this region. Opposite conditions (abnormally strong trade winds and colder than normal eastern Pacific water) are called La Niña. La Niña years are associated with weak wind shear and little subsidence in the Atlantic, typically producing active tropical cyclone activity in this basin.

El Niño events are characterized by several possible variables. The *June–July Niño 3* (1) variable represents sea surface temperature (SST) anomalies of the eastern equatorial tropical Pacific Ocean. Positive values of this variable indicate an El Niño event, and negative represents a La Niña event. *May SST in the eastern equatorial Pacific* (2) represents a similar relationship. The first clues of an impending El Niño can be detected in February by observing three variables. Upper-level westerly (zonal) wind anomalies off the northeast coast of South America imply that the upward branch of the Walker Circulation associated with ENSO remains in the western Pacific and that El Niño conditions are likely to be present in the eastern equatorial Pacific for the next

4-6 months. This situation is measured by the *February 200-mb zonal wind (U) in equatorial East Brazil* (3). Likewise, anomalous late winter meridional (north) winds at 200-mb in the South Indian Ocean are also associated with El Niño conditions (*February–March 200-mb V in the South Indian Ocean* (4)). Finally, sea level pressure (SLP) in the eastern Pacific south of the equator is a measure of the trade winds whereby weak trade winds (or westerly surface winds) are associated with lower SLP and, therefore, El Niño conditions, while the opposite is correlated to La Niña conditions. Therefore, *February SLP in the eastern South Pacific* (5) is a possible variable. Some Fall variables are also correlated to El Niño conditions, such as the *October–November SLP in the Gulf of Alaska* (6), *September 500-mb Geopotential Height in western North America* (7), and *November SLP in the subtropical northeast Pacific* (8).

3.2 Sea Level Pressure Variables

Pressure in the Atlantic Ocean is also inversely related to tropical cyclone activity, and seems to contain both monthly as well as longer term relationships. Low SLP in the tropical Atlantic implies increased atmospheric instability, moisture, and ascent (more favorable for the genesis of tropical cyclones), and weaker trade winds (which correspond to less wind shear that can tear up the thunderstorms in tropical cyclones). Low SLP in the spring tends to persist through the summer and fall. Therefore, potential variables include *March–April SLP in the eastern tropical Atlantic* (9), *June–July SLP in the tropical Atlantic* (10), and *September–November SLP in the southeast Gulf of Mexico* (11).

3.3 Teleconnection Variables

The atmosphere is characterized by long-term oscillations which impact global wind patterns, known as teleconnections. Two of these are the Arctic Oscillation and the North Atlantic Oscillation. When these oscillations are in one phase, they cause more ridges in the Atlantic, which corresponds to less wind shear. Also, on decadal timescales, weaker zonal winds in the sub-polar areas are indicative of a relatively strong thermohaline circulation and therefore a warmer Atlantic Ocean. A variable which measures this oscillation is the *November 500-mb Geopotential Height in the North Atlantic* (12).

3.4 Quasi-Biennial Oscillation Variable

Research has also shown that the Quasi-Biennial Oscillation (QBO) is correlated to tropical cyclone activity. The QBO is a stratospheric (16 to 35 km altitude) oscillation of equatorial east-west winds which vary with a period of about 26 to 30 months or roughly 2 years. These winds typically blow for 12-16 months from the east, then reverse and blow 12-16 months from the west, then back to easterly again. The west phase of the QBO has been shown to provide favorable conditions for development of tropical cyclones, possibly because it reduces wind shear. A variable which measures the QBO is the *July 50-mb Equatorial Wind (U) around the globe* (13).

3.5 Atlantic Sea Surface Temperature Variables

The Atlantic SST is another major influence on tropical cyclone activity in that basin. Like SLP, winter and spring anomalies tend to persist throughout the season. Therefore, *February SST off the northwest European Coast* (14), *April–May SST off the northwest European Coast* (15), and *June–July SST in the northeast subtropical Atlantic* (16) are potential predictors. In addition, warm SST anomalies also tend to correlate with low SLP.

4 A Dynamic Interactive Parallel Coordinates Application

To facilitate a deeper understanding of the climate data, a parallel coordinates application has been developed that fuses several previously introduced interactive extensions. In addition to fundamental PCP capabilities such as relocatable axes, axis inversion, and details-on-demand, this application provides several intuitive interaction capabilities such as axis scaling, aerial perspective shading, and dynamic visual queries. Since these individual capabilities are derived (with minor variations) from earlier research publications, the main contribution of this application lies in its collective capabilities and its application to climate analysis.

4.1 Dynamic Visual Queries

[Fig. 2 about here.]

[Fig. 3 about here.]

Since the viewer is often interested in grouping subsets of data, a method to select lines using double-ended sliders is provided for each axis (Siirtola and Räihä, 2006; Ahlberg and Shneiderman, 1994). As shown in Fig. 2, each axis has a pair of sliders (the large black triangles on each axis) which define the top and bottom range for the query area. Using the mouse cursor, the viewer can drag these sliders to dynamically adjust which lines are highlighted. Lines within the query area of every axis are rendered with a more prominent, dark color while the remaining lines are rendered with a less prominent, lighter shade of gray. An example of a conjunctive query using the sliders is shown in Fig. 3. In this image, the sliders show only two storm seasons had an above average number of named storms but a below average number of intense hurricanes. In other words, when many named storms are observed, there tends to be an average or above average number of intense hurricanes as well.

4.2 Axis Scaling (Focus+Context)

[Fig. 4 about here.]

[Fig. 5 about here.]

In displays where many relation lines are shown, it is often desirable to interactively tunnel through the relations until a smaller subset of the original data set is in focus. This application allows the user to modify the minimum and maximum values of the axes using the mouse wheel movement – a unique variation of previous axis scaling approaches (Fua et al., 1999; Artero et al., 2004; Johansson et al., 2005; Novotńy and Hauser, 2006).

On the axis bar, there are three distinct areas delineated by horizontal tick marks (Fig. 4) that are important to the axis scaling capability: the central focus area, and the top and bottom context areas. When the mouse is hovering over the focus area, an upward mouse wheel motion expands the display of the focus area outward and pushes outliers to the context areas (Fig. 5). A downward mouse wheel motion causes the inverse effect: focus region compression. Alternatively, the user may use the mouse wheel over either of the two context areas to alter the minimum or maximum values separately. The scaling capability frees space and reduces line clutter, thereby making it easier to analyze relation lines of interest.

4.3 Aerial Perspective

[Fig. 6 about here.]

The system also provides an innovative line shading scheme that is useful for quickly monitoring trends due to the similarity of data values over multiple dimensions (Jankun-Kelly and Waters, 2006). This shading scheme simulates the human perception of aerial perspective whereby objects in the distance appear faded while objects nearer to the viewer seem more vivid. In this implementation, aerial perspective shading can be used in either a discrete or a continuous mode. In the discrete mode, the lines are colored according to the axis region that they intersect which is similar to the technique described by Siirtola and Räihä (2006). If any point of a relation line is in the context (non-focus) area of at least one axis, the line is shaded with a light gray color and drawn beneath the non-context lines (Fig. 5). If all the points on a relation line fall within the query area of each axis (the area between the two query sliders), the line is colored using a dark gray value that attracts the viewer's attention (Fig. 6). The remaining lines (non-query and non-context) are colored a shade of gray that is slightly darker than the context lines but lighter than the query lines.

In the continuous mode, non-context lines go through an additional step to encode the distance of the line from the mouse cursor in a manner similar to the approach described by Jankun-Kelly and Waters (2006). Query lines that are nearest to the mouse cursor are shaded with the darkest gray color while lines furthest from the mouse cursor are shaded with a lighter gray. The other query lines are shaded according to a non-linear fall-off function that yields a gradient of gray colors between extremes. Consequently, the lines that are nearest to the mouse cursor are more prominent to the viewer due to the more drastic color contrast and depth ordering treatments (Fig. 6) giving the viewer the ability to effectively use the mouse to perform rapid, visual queries.

4.4 Descriptive Statistical Indicators

To support the interactive analysis capabilities of the system, each axis offers visual representations of key descriptive statistics that are identified in Fig. 2 (Siirtola and Räihä, 2006; Hauser et al., 2002). The mean, standard deviation range, and the frequency information are calculated for the data in the focus area of each axis. Alternatively, the user can configure the system to display the median and interquartile range. All plots and analysis in this paper utilize the mean and standard deviation display mode. These central tendency and variability measures provide a numerical value that indicates the typical value and how "spread out" the samples are in the distribution, respectively. The axis box plots represent the descriptive statistics for all the samples within the focus area of the axis. In each axis interior, the frequency information is also displayed by representing histogram bins as small rectangles with gray values that are indicative of the number of lines that pass through the bin's region (see Fig. 2). That is, the darkest bins have the most lines passing through while lighter bins have less lines. In Fig. 5, the histogram display is illustrated during an axis scaling operation.

5 Parallel Coordinates Validation: North Atlantic Case Study

As discussed previously, regression analysis is often employed to identify the most relevant climate relationships for tropical cyclone activity. Such techniques are effective in screening data and providing quantitative associations. However, multivariate analysis can be difficult. This section will outline how stepwise regression and parallel coordinates can compliment each other in such an analysis.

Stepwise regression with a "backwards glance" is used which selects the optimum number of most important variables using a predefined significance value (90% in this study). Stepwise regression can compliment parallel coordinate visualization by isolating the significant variables in a quantitative fashion. An interactive parallel coordinates visualization can then be used to develop a deeper understanding of the complex relationships between the variables.

An extra step is taken to ensure the proper selection of variables. The initially chosen variables are examined for multicollinearity; if any variables are correlated with each other by more than 0.5, one is removed and the code rerun. In this way, the chosen variables are truly independent of each other.

A normalization procedure is also executed for equal comparison between the variables. Denoting σ as the standard deviation of a variable, y as the dependent variable (named storms, hurricanes, or intense hurricanes in this study), $\overline{x}$ as the predictor mean, and $\overline{y}$ as the dependent variable mean, a number k of statistically significant predictors are normalized by the following regression:

$$(y - \overline{y})/\sigma_y = \sum_{i=1}^{k} b_i(x_i - \overline{x}_i)/\sigma_i \tag{1}$$

The advantage of this approach is that the importance of a predictor may be assessed by comparing regression coefficients b_i between different variables, and that the y-intercept becomes zero.

In addition, $\overline{x}_i$ may be interpreted (to a first approximation) as a "threshold" value which distinguishes between positive and negative contributions (for $b_i > 0$), and the opposite for negative b_i. Years when independent variables contain large deviations from the mean could be associated with very active or inactive years, and require closer examination. As will be seen, the parallel coordinates technique facilitates the examination of active and quiet Atlantic hurricane seasons.

[Table 3 about here.]

The 16 potential variables listed in Table 2 are examined in the stepwise regression, yielding several independent variables for each dependent variable. These results show that several climate factors impact tropical cyclone activity. The chosen predictors are shown in Table 3, along with their normalized regression coefficient and sample mean. The explained variance (R^2) is shown in the 3 table headings.

The stepwise regression shows only one significant El Niño variable (late winter South Indian Ocean 200-mb meridional winds (4)) impacts total number of storms; it is the second most influential predictor. Late winter northwest coastal European SST (14) is the leading predictor. The North Atlantic Oscillation (manifested by 500-mb geopotential height in the North Atlantic (12)) ranks third, and is also the only variable seen in all three tables. This suggests that the presence of a ridge in the Atlantic is conducive to an above average tropical cyclone season. Finally, low SLP in the southeast Gulf of Mexico (11) also encourages the formation of tropical cyclones. Note that the coefficient has a negative sign, showing that the lower the pressure, the better the chance of tropical cyclone activity.

For number of hurricanes, the analysis surprisingly shows that October–November SLP in the Gulf of Alaska (6) is the most important predictor. The physical role is not clear, although scientists know it is correlated to El Niño activity. Northeast subtropical Atlantic SST (16) and North Atlantic 500-mb geopotential height (12) are tied for second, and southeast Gulf SLP again ranks fourth (11). The explained variance is 42% — more than the 34% for named storms. This suggests stronger predictor relationships for number of hurricanes.

For intense hurricanes, the variance increases to 54%. In this case, the North Atlantic November 500-mb height variable (12) is the strongest predictor. Early summer tropical Atlantic SLP (10) ranks number two, followed by September 500-mb geopotential height in western North America (7) and February SST off northwest coastal Europe (14). The higher variance and distinctly different chosen predictors suggests different environmental influences are required for intense hurricanes. This analysis correlates the presence of high pressure in the western U.S. and over the Atlantic, low summer Atlantic SLP, and warm SST as necessary conditions for intense hurricanes.

Because there is unexplained variance and several predictors, can parallel coordinates glean any more information? To answer this question, the data sets are stratified into below normal, normal, and above normal seasons using the software's interactive capabilities, and the significant predictors identified by the stepwise regression are analyzed visually. Using the axis box plots (drawn using the standard deviation and mean), the below normal, normal, and above normal seasons are determined by moving the query sliders for the axis of interest to encapsulate the lines above the standard deviation range, within the standard deviation range, and below the standard deviation range, respectively. After setting the query sliders, the aerial perspective shading highlights the relationships of interest, thus enabling rapid visual analysis of the variables.

[Fig. 7 about here.]

[Fig. 8 about here.]

[Fig. 9 about here.]

[Fig. 10 about here.]

[Fig. 11 about here.]

[Fig. 12 about here.]

Figure 7 shows a plot for seasons with below normal named storms (sample size of 16). Even though the regression shows February Atlantic SST (14) as the most important overall predictor, it is not as effective for discerning inactive seasons. The plot shows considerable scatter, and with only 6 years of significantly below average SST. The dynamic query capabilities of this parallel coordinates application make these combined queries and subsample analysis an intuitive exercise.

September–November Gulf of Mexico SLP (11) also exhibits much scatter, with a slight majority of years with above normal pressure. However, February–March 200-mb South Indian Ocean meridional winds (4) — a surrogate measurement of El Niño, shows 15 seasons (94%) of strong north winds, tightly clustered in the plots. *This suggests El Niño is the major contributor to inac-*

tive Atlantic tropical cyclone seasons. Note also that below normal November North Atlantic 500-mb geopotential heights (12) plays a pivotal role for quiet seasons. Fourteen seasons (87%) contain lower geopotential heights in November, suggesting the presence of upper-level troughs which can shear tropical cyclones. However, this signal is not as strong as the El Niño predictor. Additionally, many unshaded lines exist for positive 200-mb V, showing that other factors besides El Niño contribute to normal and active seasons. In fact, a similar parallel coordinates stratification analysis shows that November North Atlantic 500-mb geopotential heights (12) and September–November Gulf of Mexico SLP (11) tend to be the critical players for an active tropical cyclone season (not shown).

Figure 8 shows seasons with below normal hurricane activity (19 seasons). El Niño again tends to dominate the signal through the fall Gulf of Alaska SLP (6) term. However, in contrast to number of named storms, Atlantic SST (16) becomes important for number of hurricanes. This suggests that when water temperature is below normal, tropical storms will have difficulty reaching hurricane status. For above normal hurricane activity (Fig. 9), June–July Atlantic SST (16), November North Atlantic 500-mb geopotential height (12), and Gulf of Mexico SLP (11) tend to exert dominant roles, with El Niño a secondary factor.

Intense hurricanes warrant special consideration, since they cause 80% of the economic damage from tropical cyclones. Figure 10 shows that cold February Atlantic SSTs (14) and high Atlantic June–July SLP (10) tend to reduce the number of intense hurricanes, with November North Atlantic 500-mb geopotential heights (12) playing a secondary role and September 500-mb geopotential heights in western North America (7) contributing no role. In contrast, all four predictors have tightly clustered lines showing they all play dominant roles in seasons with above normal intense hurricane activity (Fig. 11). These terms are associated with the presence of ridges in the western U.S. and the Atlantic, below average Atlantic SLP, and warm wintertime Atlantic SST off the northwestern European Coast. Ridges are low shear environments, showing that the lack of upper level troughs is an important factor for seasons with many intense hurricanes. Low SLP indicates minimal subsidence. Sinking air suppresses cloud growth and also dries the lower atmosphere, both of which are not conducive to the formation and development of tropical cyclones. Low SLP also could indicate better organized tropical waves (from which many Atlantic tropical cyclones form). Warm wintertime northeast Atlantic water also is a good precursor for above average intense hurricane activity.

This parallel coordinates application can also investigate the differences between the extremely busy 2005 season and the slightly below average 2006 season. Figure 12 shows the 2005 and 2006 seasons along with the chosen predictors from all three categories (named storms, hurricanes, and intense hurricanes) listed in Table 3. This plot reveals that most of the terms are nearly the same except for October–November SLP in the Gulf of Alaska (6) (above average in 2005, below average in 2006) and June–July SLP in the tropical Atlantic (10) (below average in 2005, above average in 2006). Klotzbach et al. (2006b) and Bell et al. (2007) show that the tropical Atlantic was quite dry through most of the 2006 hurricane season due to subsidence associated with the onset of an unusually late ENSO event (indicated by the Gulf of Alaska SLP), as well as frequent outbreaks of African dust storms that year.

6 Conclusion

This research has shown that a visual analysis system based on interactive parallel coordinates can be used to confirm and clarify the results of stepwise regression in climate analysis. The effectiveness of the system concepts are demonstrated via a real-world case study to identify the most significant predictors for seasonal tropical cyclone statistics. While multiple regression provides an ordering of the most significant variables, the visual analysis using the PCP system facilitates a deeper understanding of the environmental causes for above average and below average hurricane seasons.

Acknowledgements

This research is sponsored by the Naval Research Laboratory's Long-Term Training Program, by the National Oceanographic and Atmospheric Administration (NOAA) with grants NA060AR4600181 and NA050AR4601145, and through the Northern Gulf Institute funded by grant NA06OAR4320264. This particular project was initiated in the Information Visualization course taught at Mississippi State University by Dr. T.J. Jankun-Kelly. The authors wish to thank Dr. Phil Klotzbach of Colorado State University's Tropical Meteorology Project for providing the Atlantic tropical cyclone data set.

References

Ahlberg, C., Shneiderman, B., 1994. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In: Proceedings of

Human Factors in Computing Systems. ACM, Boston, MA, pp. 313–317, 479–480.

Artero, A. O., de Oliveira, M. C. F., Levkowitz, H., Oct. 2004. Uncovering clusters in crowded parallel coordinates visualization. In: IEEE Symposium on Information Visualization. IEEE Computer Society, Austin, Texas, pp. 81–88.

Bell, G. D., Blake, E., Landsea, C. W., Chelliah, M., Pasch, R., Mo, K. C., Goldenberg, S. B., 2007. The tropics – Atlantic basin. In: Arguez, A. (Ed.), State of the Climate in 2006. Vol. 88. Bulletin of the American Meteorological Society, pp. S48–S51.

Chu, P.-S., 2004. ENSO and tropical cyclone activity. In: Murnane, R. J., Liu, K.-B. (Eds.), Hurricanes and Typhoons: Past, Present, and Future. Columbia University Press, pp. 297–332.

Fitzpatrick, P. J., 1996. Understanding and forecasting tropical cyclone intensity change. Ph.D. thesis, Department of Atmospheric Sciences, Colorado State University, Fort Collins, CO.

Fitzpatrick, P. J., 1997. Understanding and forecasting tropical cyclone intensity change with the typhoon intensity prediction scheme (TIPS). Weather and Forecasting 12 (4), 826–846.

Fua, Y.-H., Ward, M. O., Rundensteiner, E. A., Oct. 1999. Hierarchical parallel coordinates for exploration of large datasets. In: Proceedings of IEEE Visualization. IEEE Computer Society, San Francisco, California, pp. 43–50.

Hauser, H., Ledermann, F., Doleisch, H., 2002. Angular brushing of extended parallel coordinates. In: Proceedings of IEEE Symposium on Information Visualization 2002. IEEE Computer Society, Boston, MA, pp. 127–130.

Healey, C. G., Tateosian, L., Enns, J. T., Remple, M., 2004. Perceptually-based brush strokes for nonphotorealistic visualization. ACM Transactions on Graphics 23 (1), 64–96.

Inselberg, A., 1985. The plane with parallel coordinates. The Visual Computer 1 (4), 69–91.

Jankun-Kelly, T. J., Waters, C., 2006. Illustrative rendering for information visualization. In: Posters Compendium: IEEE Visualization 2006. IEEE Computer Society, Baltimore, MD, pp. 42–43.

Johansson, J., Ljung, P., Jern, M., Cooper, M., Oct. 2005. Revealing structure within clustered parallel coordinates displays. In: IEEE Symposium on Information Visualization. IEEE Computer Society, Minneapolis, Minnesota, pp. 125–132.

Klotzbach, P. J., Jan. 2007. personal communication.

Klotzbach, P. J., Gray, W. M., Thorson, W., 2006a. Extended range forecast of Atlantic seasonal hurricane activity and U.S. landfall strike probability for 2007. Tech. rep., http://tropical.atmos.colostate.edu/Forecasts/2006/dec2006/ (current 25 Oct. 2007).

Klotzbach, P. J., Gray, W. M., Thorson, W., 2006b. Summary of 2006 Atlantic tropical cyclone activity and verification of author's seasonal and monthly forecasts. Tech. rep., http://tropical.atmos.colostate.edu/Forecasts/2006/nov2006/ (current 25 Oct. 2007).

Landsea, C. W., 2005. Hurricanes and global warming. EOS 438, E11–E13.

Novotný, M., Hauser, H., 2006. Outlier-preserving focus+context visualization in parallel coordinates. IEEE Transactions on Visualization and Computer Graphics 12 (5), 893–900.

Rensink, R. A., 2002. Change detection. Annual Review of Psychology 53, 245–577.

Siirtola, H., 2000. Direct manipulation of parallel coordinates. In: Proceedings of the International Conference on Information Visualisation. IEEE Computer Society, London, England, pp. 373–378.

Siirtola, H., Räihä, K.-J., Dec. 2006. Interacting with parallel coordinates. Interacting with Computers 18 (6), 1278–1309.

Tweedie, L., Spence, R., Dawkes, H., Su, H., 1996. Externalising abstract mathematical models. In: Proceedings of the Conference on Human Factors in Computing Systems. ACM, Vancouver, British Columbia, Canada, pp. 406–412.

Vitart, F., 2004. Dynamical seasonal forecasts of tropical storm statistics. In: Murnane, R. J., Liu, K.-B. (Eds.), Hurricanes and Typhoons: Past, Present, and Future. Columbia University Press, pp. 354–392.

Wegman, E. J., 1990. Hyperdimensional data analysis using parallel coordinates. Journal of the American Statistical Association 85 (411), 664–675.

List of Figures

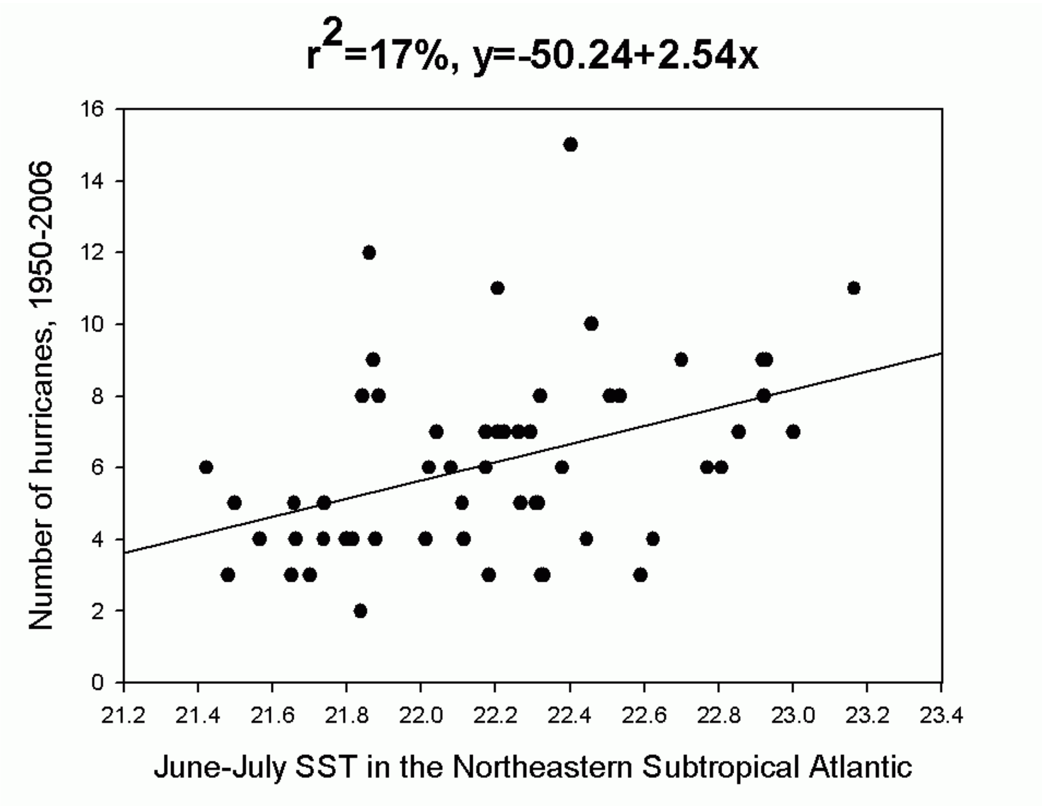


Fig. 1. A common visualization technique used in climate studies is the scatter plot overlaid with a linear regression line. This example shows the linear relationship between June–July SST (16) in the northeastern subtropical Atlantic Ocean, and the number of hurricanes from 1950 to 2006. The explained variance is 17%.

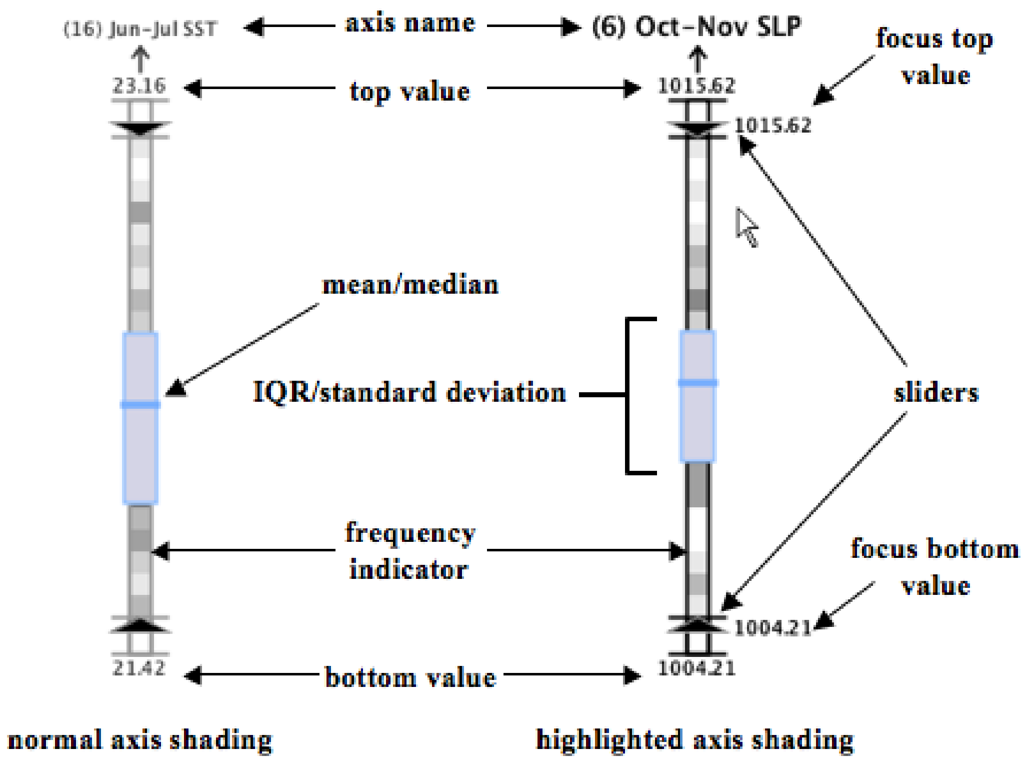


Fig. 2. An annotated view of the parallel coordinate axis display widget. Normally, an axis is displayed using a muted color scheme (left). However, when the mouse moves into an axis space, the axis is displayed with the highlighted color scheme and focus area limits are shown (right).

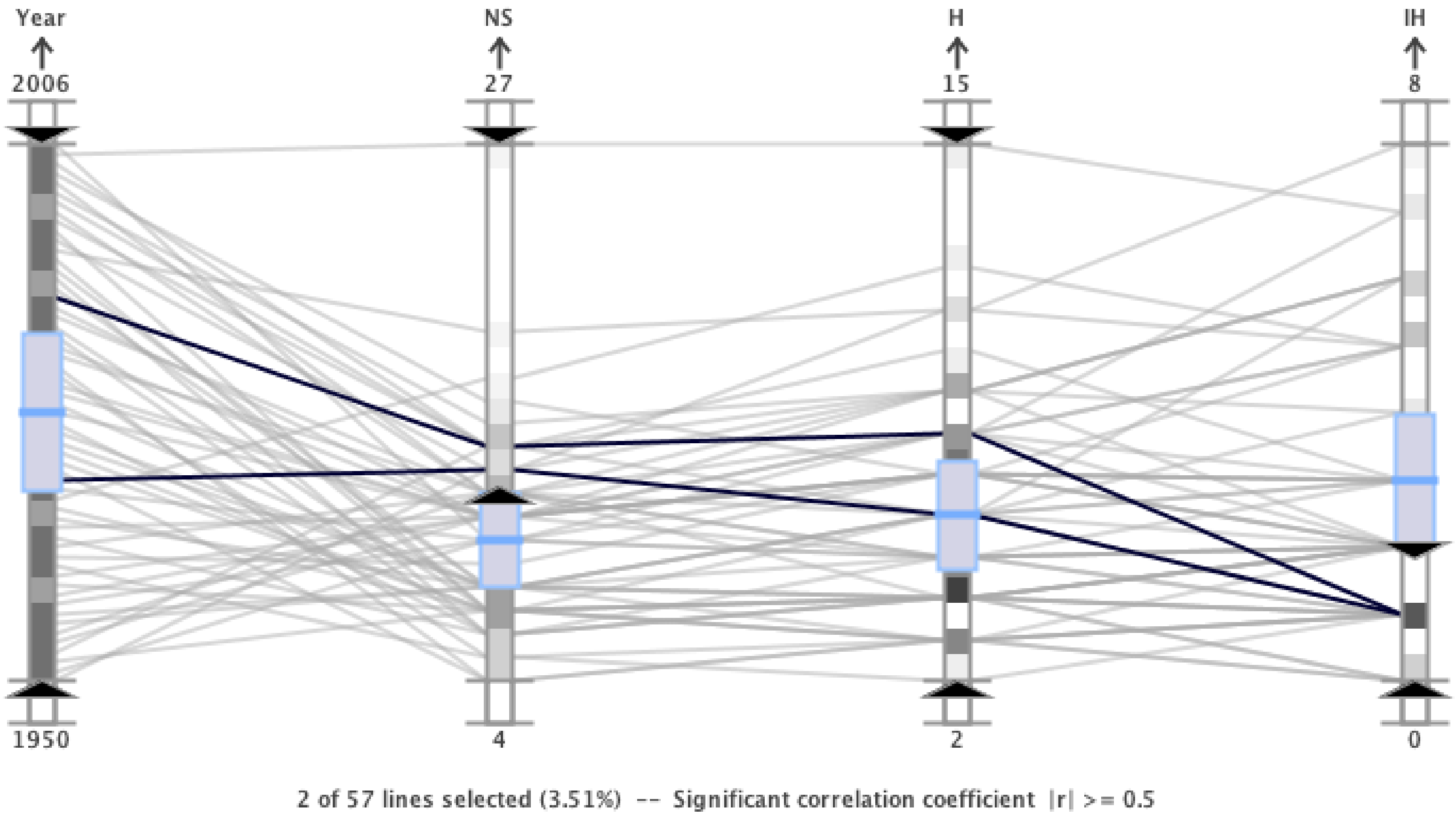


Fig. 3. An example of the conjunctive query capability using the dynamic query sliders for multiple axes. In this example, the sliders are set for the above average range of the Named Storms (NS) axis and the below average range of the Intense Hurricanes (IH) axis for data between years 1950 and 2006. This query reveals that only 2 storm seasons fulfilled this criteria.

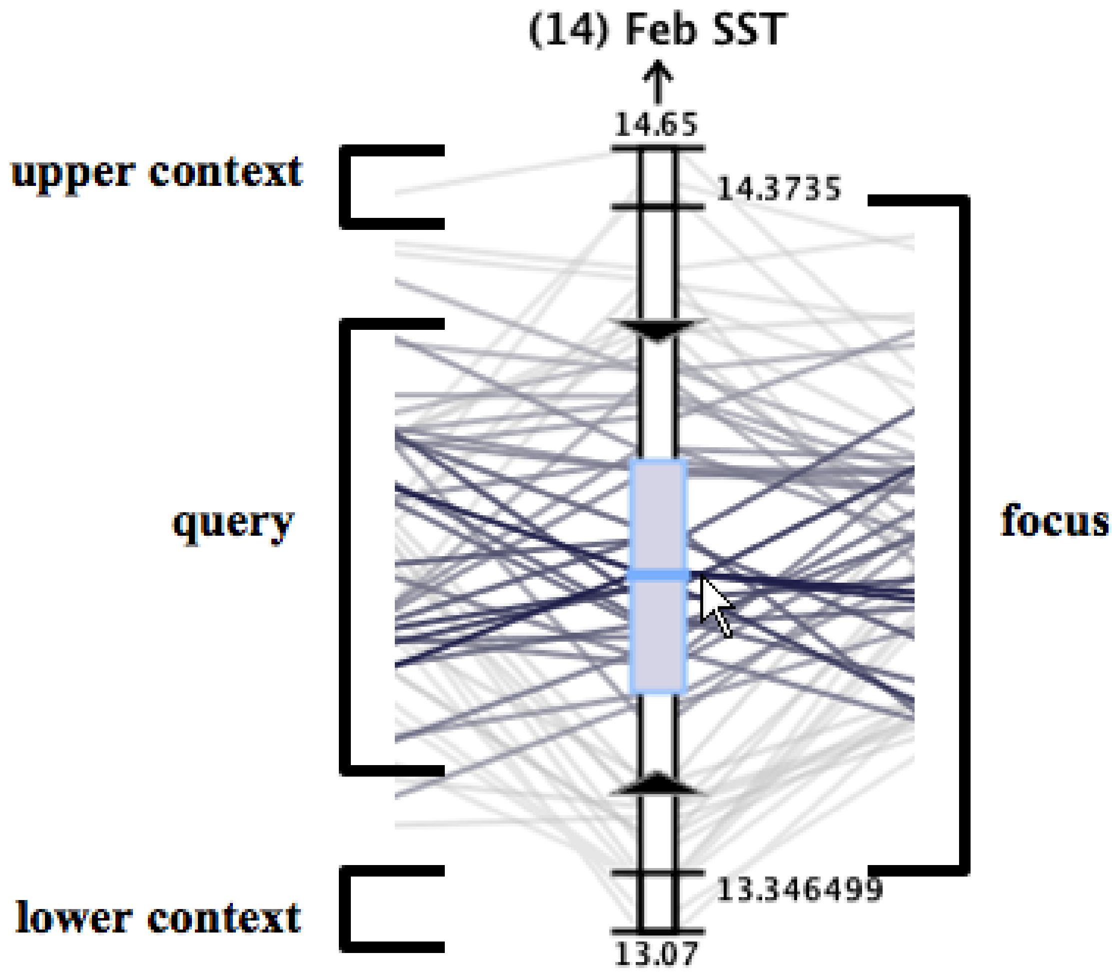
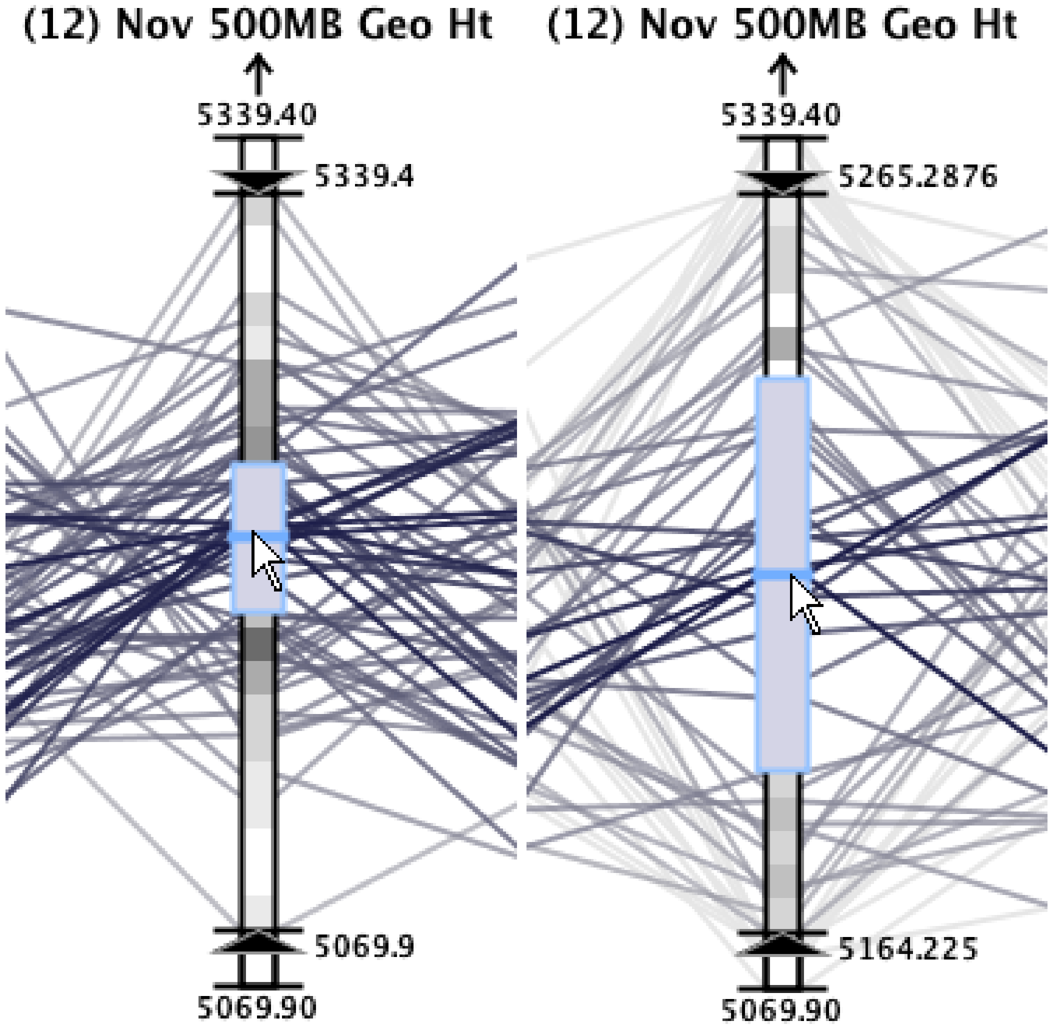
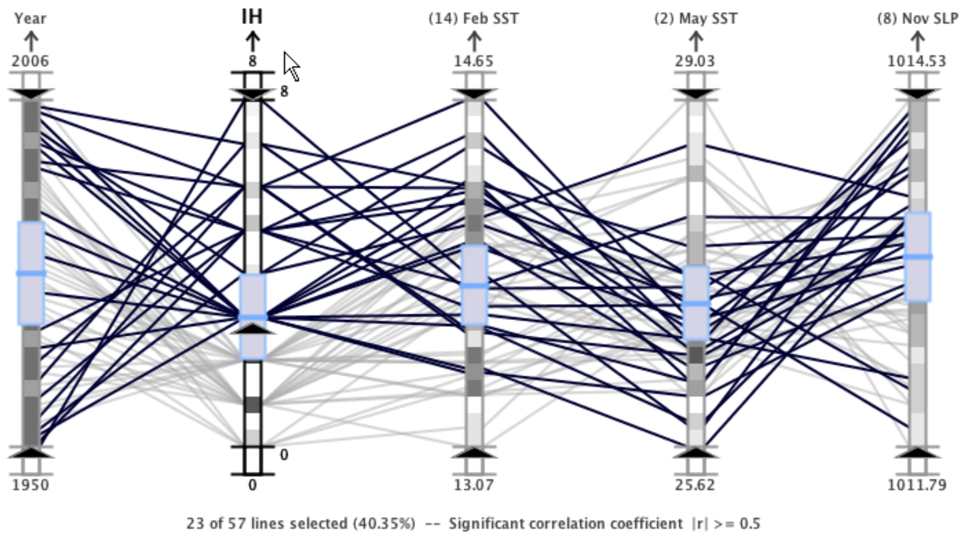


Fig. 4. The axis bar is segmented into four distinct areas: the query area, the focus area, and an upper and lower context area.

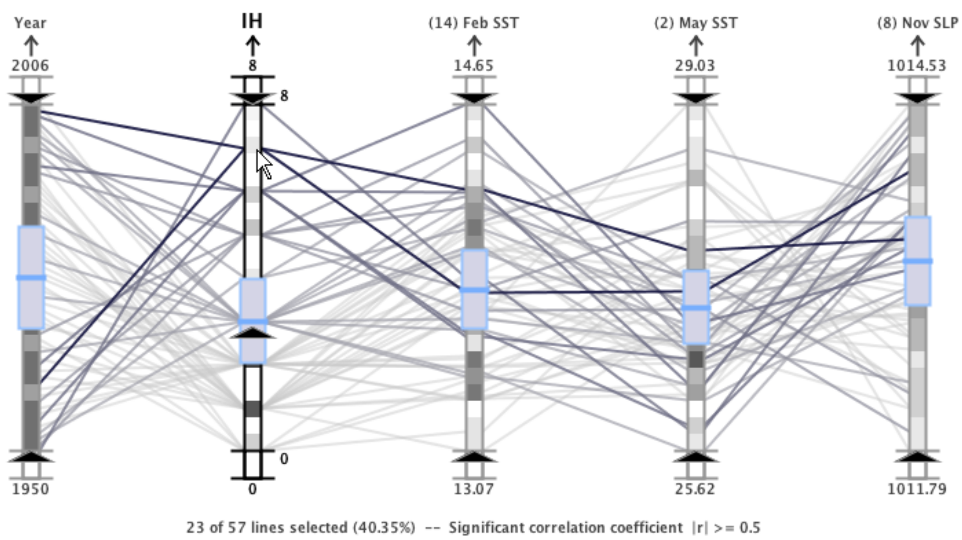


(a) (b)

Fig. 5. A screen shot of the parallel coordinates application before (a) and after (b) scaling has been performed. In this example, scaling occurs by performing an upward mouse wheel function in the focus area of the axis which moves the values for the top and bottom closer together, effectively stretching the display upward and downward (with the base of the display fixed).



(a) Discrete aerial perspective shading.



(b) Continuous aerial perspective shading.

Fig. 6. A screen shot of the aerial perspective shading capability which can be used in either discrete (a) or continuous (b) shading mode. The line colors are determined based on the location of the line with respect to the context, focus, and query areas of the axes and, in continuous mode, the distance from the mouse cursor is encoded with color value. In the above examples, the mouse cursor is positioned at the top of the second axis (the IH axis) which highlights the storm seasons with above average intense hurricane activity. The continuous shading mode gives more emphasis to the lines representing the most active seasons.

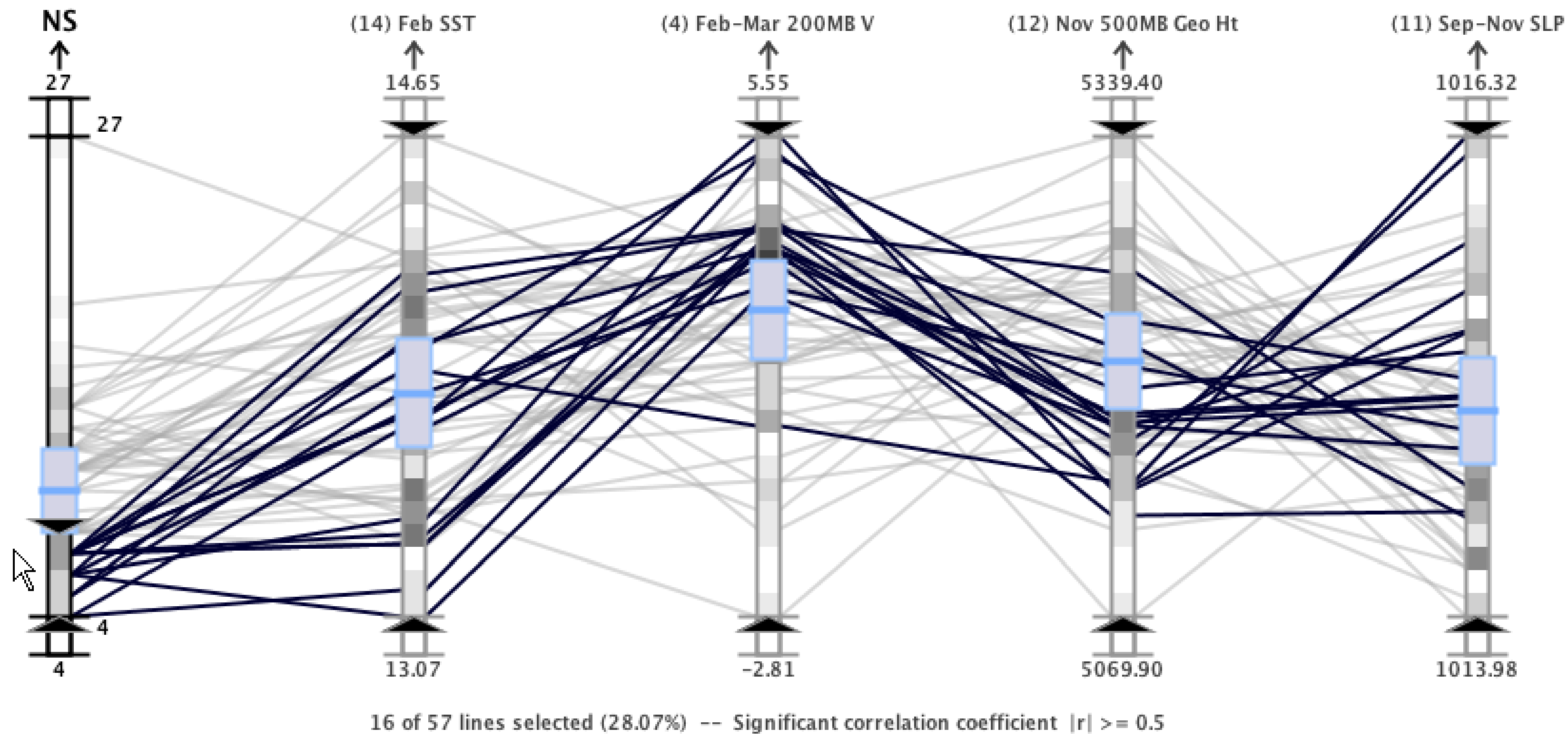


Fig. 7. A plot of the variables that the regression analysis selected as the most influential factors for the number of named storms in a season (1950 to 2006). The below average seasons are highlighted. The tighter clustering of lines for February–March 200-mb South Indian Ocean meridional winds (4) and November North Atlantic 500-mb geopotential heights (12) suggest they are the most influential contributors to quiet Atlantic tropical cyclone seasons.

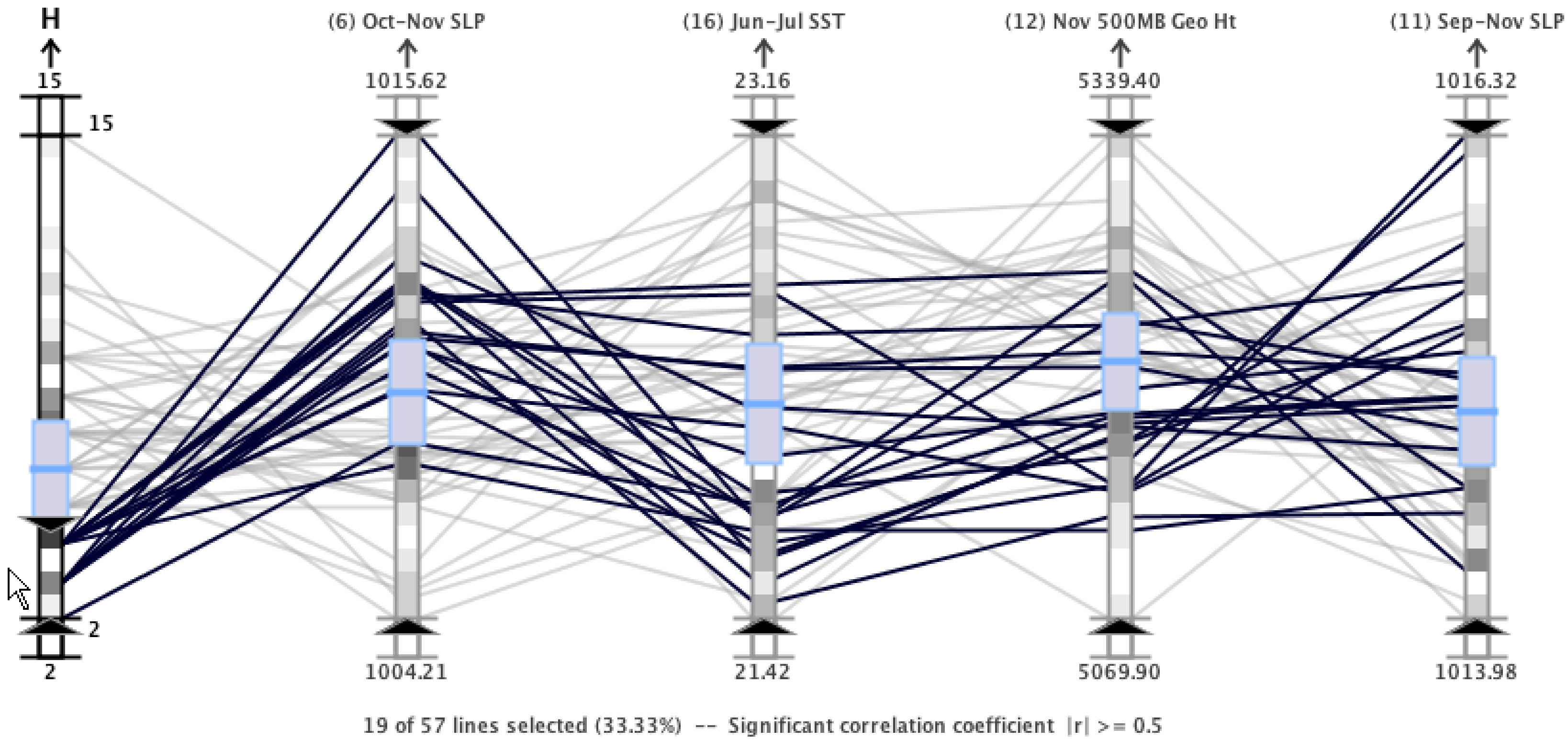


Fig. 8. A plot of the variables that the regression analysis selected as the most influential factors for the number of hurricanes in a season (1950 to 2006). The below average seasons are highlighted. El Niño dominates the signal with the October–November Gulf of Alaska SLP (6) term, and the June–July northeast subtropical Atlantic SST (16) becomes important.

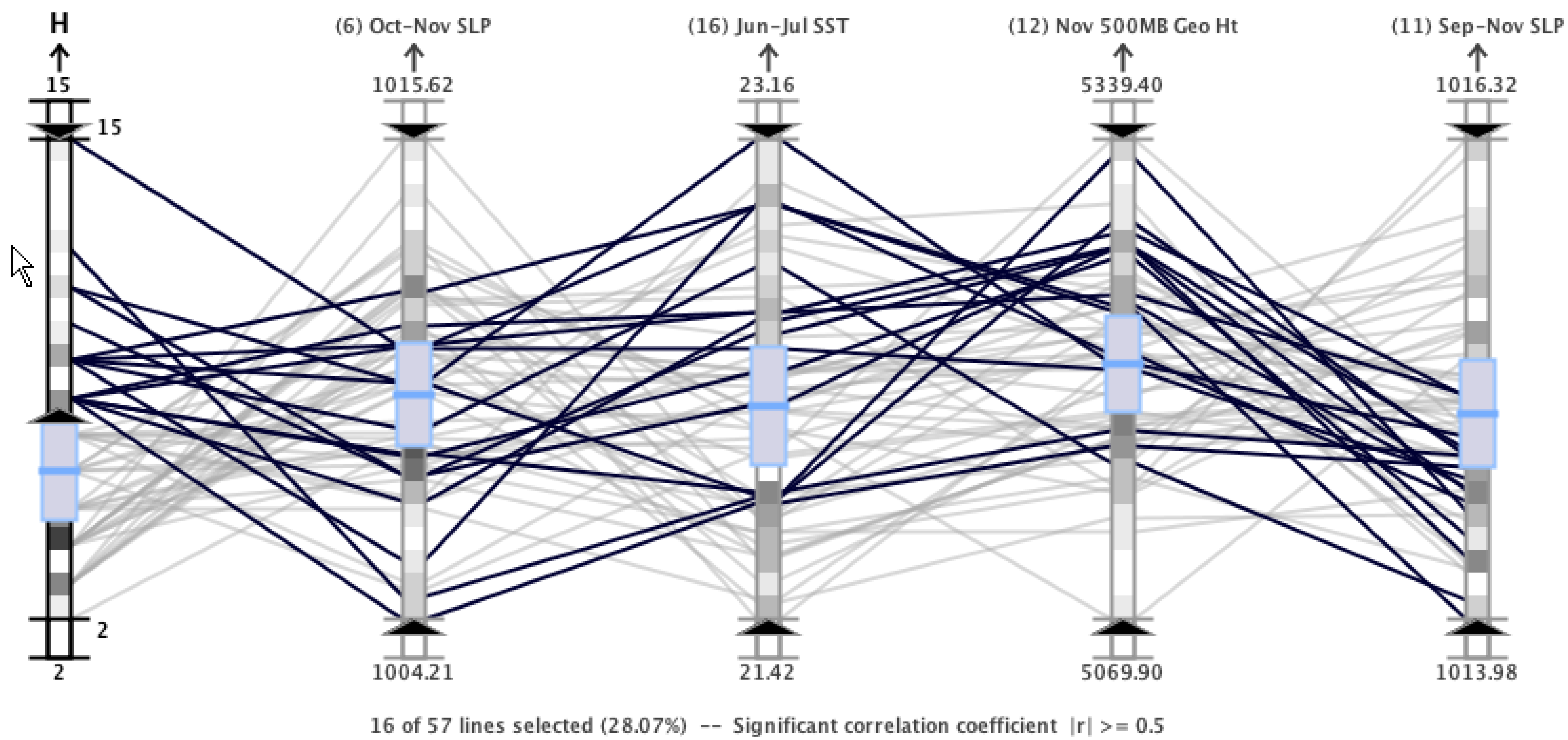


Fig. 9. A plot of the variables that the regression analysis selected as the most influential factors for the number of hurricanes in a season (1950 to 2006). The above average seasons are highlighted. This plot suggests that the El Niño term (Gulf of Alaska October–November SLP (6)) is a secondary factor to the other three terms.

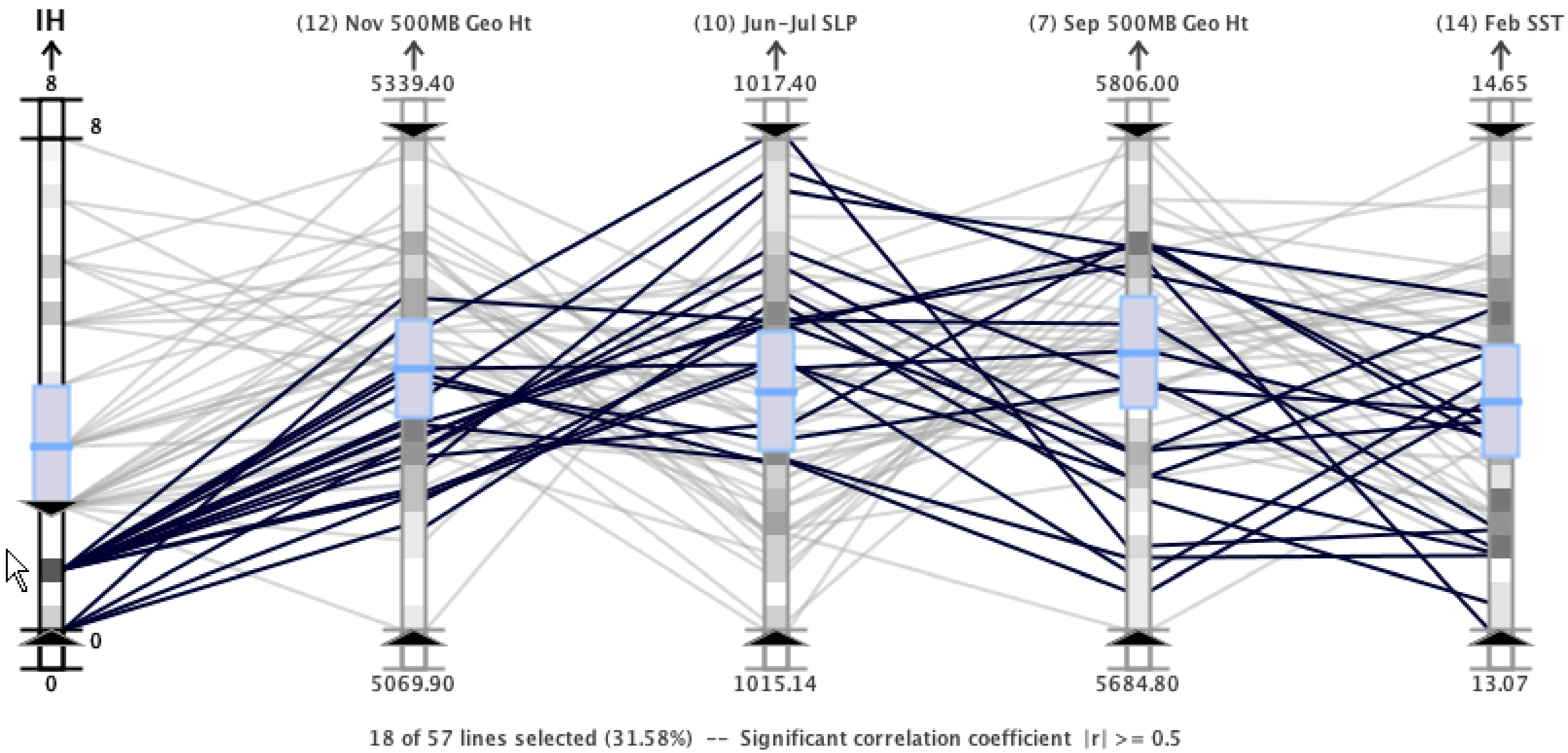


Fig. 10. A plot of the variables that the regression analysis selected as the most influential factors for the number of intense hurricanes in a season (1950 to 2006). The below average seasons are highlighted. The plots shows that cold February coastal Europe SST (14) and high June–July tropical Atlantic SLP (10) tend to reduce the number of intense hurricanes. November 500-mb North Atlantic geopotential height (12) also plays a secondary role.

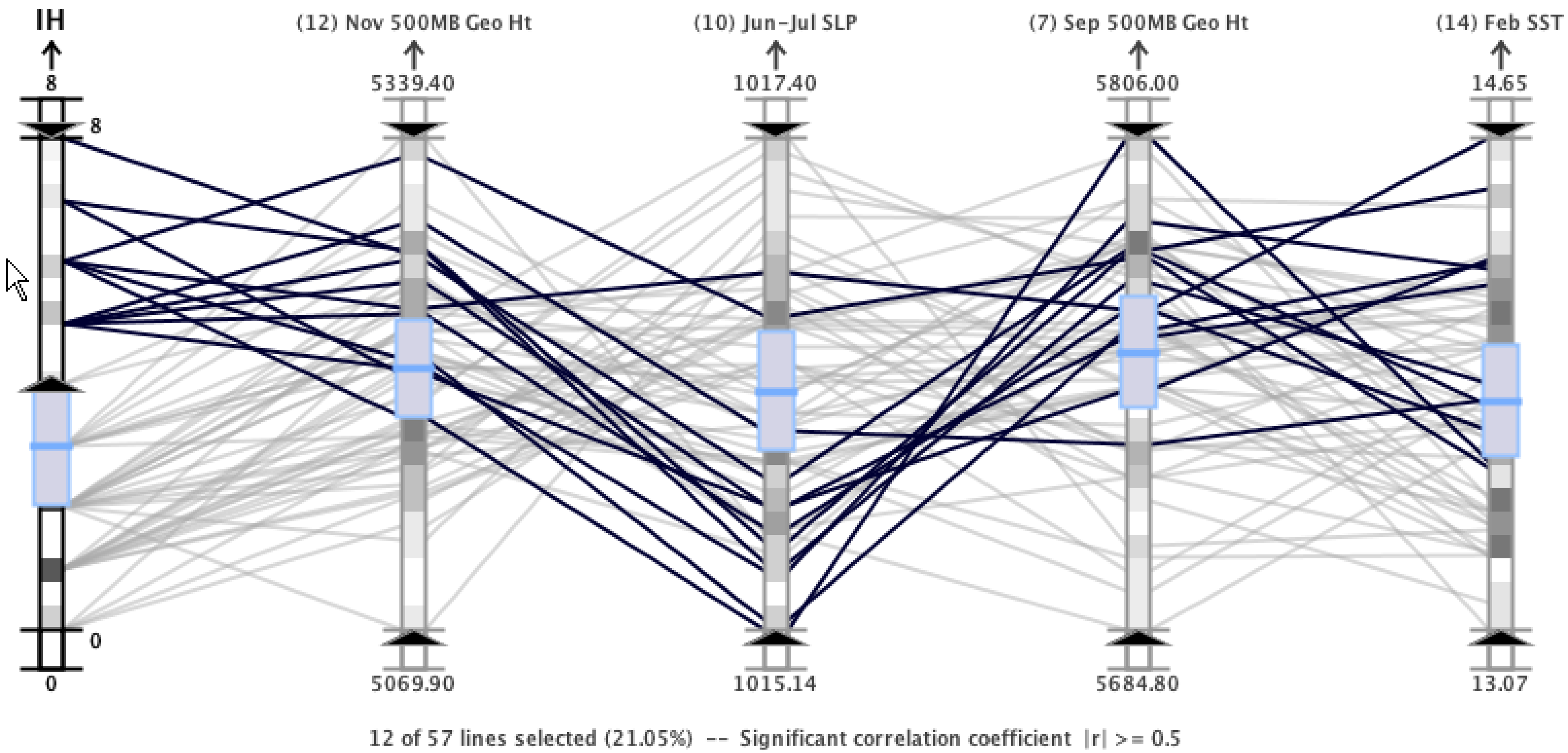


Fig. 11. A plot of the variables that the regression analysis selected as the most influential factors for the number of intense hurricanes in a season (1950 to 2006). The above average seasons are highlighted. In this plot all four predictors have tightly clustered lines suggesting they all play dominant roles in seasons with high intense hurricane activity.

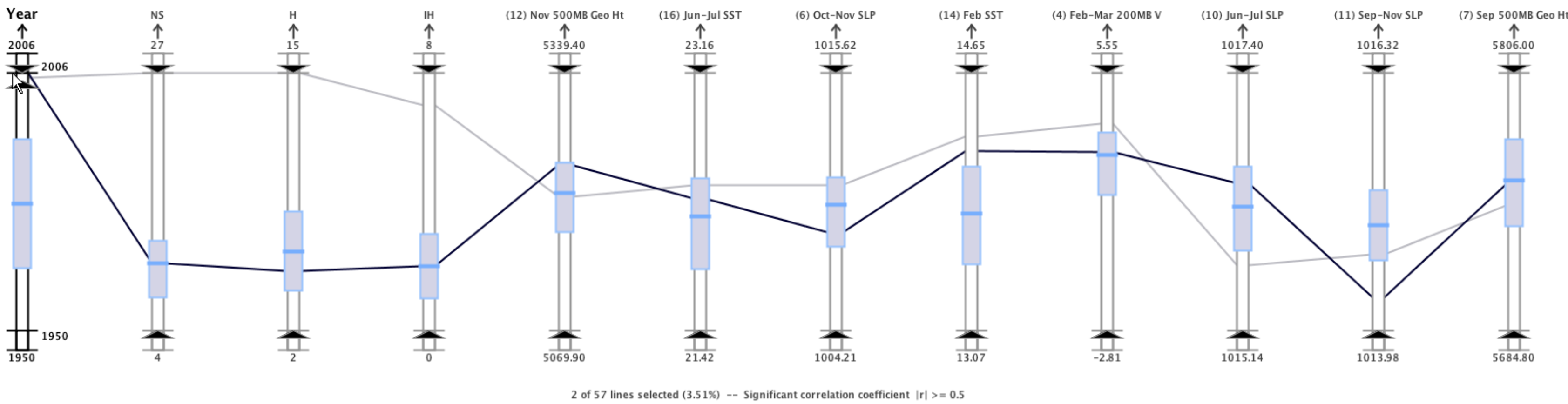


Fig. 12. A plot of all the influential variables that the regression analysis selected for the number of named storms, hurricanes, and intense hurricanes in a season (1950 to 2006). The very busy 2005 and slightly below average 2006 seasons are highlighted. Continuous aerial perspective shading is used to highlight the 2006 season polyline with a darker shade of gray. The plot suggests that October–November Gulf of Alaska SLP (6) and June–July tropical Atlantic SLP (10) were the biggest differences between these seasons.

List of Tables

Table 1
Interaction and representation features included in the parallel coordinates based visualization system developed in this research.

Focus+Context	Interactively scales an axis and zooms into a subset of relations for that axis.
Aerial Perspective	Facilitates visual queries by shading lines based on proximity to the mouse cursor using a shading scheme that mimics human perception.
Dynamic Visual Query	Explores multidimensional relationships with double-sided sliders.
Statistical Indicators	Indicates statistical quantities to support interaction model.
Relocatable Axes	Reorganizes the axes by dragging with the mouse to observe the correlation between variables.
Axis inversion	Inverts the axis display scale by swapping the top and bottom values.
Details-on-demand	Shows additional details for the highlighted axis, and displays the value on the axis scale under the mouse by clicking on the axis with the middle mouse button.
Customizable Display	Modifies the display (statistics display, color schemes, tick marks) via a pop-up menu interface.

Table 2. Environmental tropical cyclone climate variables evaluated as predictors in the multiple regression procedure.

	Variable Name	Geographical Region
(1)	June–July Niño 3	5S-5N, 90-150W (eastern equatorial tropical Pacific Ocean)
(2)	May SST	5S-5N, 90-150W (eastern equatorial tropical Pacific Ocean)
(3)	February 200-mb U	5S-10N, 35-55W (equatorial East Brazil)
(4)	February–March 200-mb V	35-62.5S, 70-95E (South Indian Ocean)
(5)	February SLP	0-45S, 90-180W (eastern South Pacific Ocean)
(6)	October–November SLP	45-60N, 120-160W (Gulf of Alaska)
(7)	Sept. 500-mb Geopotential Height	35-55N, 100-120W (western North America)
(8)	November SLP	7.5-22.5N, 125-175W (subtropical northeast Pacific Ocean)
(9)	March–April SLP	0-20N, 0-40W (eastern tropical Atlantic Ocean)
(10)	June–July SLP	10-25N, 10-60W (tropical Atlantic Ocean)
(11)	September–November SLP	15-35N, 75-97W (southeast Gulf of Mexico)
(12)	Nov. 500-mb Geopotential Height	67.5-85N, 50W-10E (North Atlantic Ocean)
(13)	July 50-mb U	5S-5N, 0-360 (equatorial globe)
(14)	February SST	35-50N, 10-30W (northwest European Coast)
(15)	April–May SST	30-45N, 10-30W (northwest European Coast)
(16)	June–July SST	20-40N, 15-35W (northeast subtropical Atlantic Ocean)

Table 3
Significant climate variables chosen from Table 2 by the stepwise regression for number of named storms, hurricanes, and intense hurricanes in 1950-2006. Also shown is the explained variance R^2, the normalized coefficients b, and the sample mean.

Number of Named Storms (NS)

(R^2 is 34%)

Chosen Variables	Normalized Coefficients c	Sample Mean
Feb. SST (14)	0.302	13.8
Feb.–Mar. 200-mb V (4)	–0.244	2.5
Nov. 500-mb Geopot. Ht. (12)	0.232	5213
Sep.–Nov. SLP (11)	–0.175	1015.0

Number of Hurricanes (H)

(R^2 is 42%)

Chosen Variables	Normalized Coefficients c	Sample Mean
Oct.–Nov. SLP (6)	–0.284	1009.6
June–July SST (16)	0.259	22.2
Nov. 500-mb Geopot. Ht. (12)	0.258	5213
Sep.–Nov. SLP (11)	–0.208	1015.0

Number of Intense Hurricanes (IH)

(R^2 is 54%)

Chosen Variables	Normalized Coefficients c	Sample Mean
Nov. 500-mb Geopot. Ht. (12)	0.345	5213
June-July SLP (10)	–0.315	1016.2
Sep. 500-mb Geopot. Ht. (7)	0.292	5753.3
Feb. SST (14)	0.235	13.8