North Atlantic Hurricane Trend Analysis using Parallel Coordinates and Statistical Techniques

Chad A. Steed[a], Patrick J. Fitzpatrick[b], T.J. Jankun-Kelly[c], and J. Edward Swan II[c]

[a]Naval Research Laboratory, Mapping, Charting, and Geodesy Branch, Stennis Space Center, MS, USA, 39529;
[b]Northern Gulf Institute, Mississippi State University, Stennis Space Center, MS, 39529;
[c]Department of Computer Science and Engineering, Mississippi State University, Mississippi State, MS, 39762.

One of the most challenging tasks in multidimensional multivariate data analysis is to identify and quantify the associations between a set of interrelated variables. In real-world climate studies, this task is even more daunting due to the uncertainty and complexity of dynamic, environmental data sets. Notwithstanding the difficulty, the variability and destructiveness of recent hurricane seasons has invigorated efforts by weather scientists to identify environmental variables that have the greatest impact on the intensity and frequency of seasonal hurricane activity. In general, the goal of such efforts is to improve the accuracy of seasonal forecasts which should, in turn, improve preparedness and reduce the impact of these devastating natural disasters.

One particularly useful method for predicting seasonal hurricane variability is based on the idea that there are predictors of the main dynamic parameters that affect storm activity which can be observed up to a year in advance. Using historical data, their importance is estimated using statistical regression techniques similar to those described by Vitart [1]. Klotzbach et al. [2] used these technique to determine the most important variables for predicting the frequency of North Atlantic tropical cyclone activity. Similarly, Fitzpatrick [3] applied stepwise regression analysis to the prediction of tropical cyclone intensity. Although sometimes complicated to establish, these techniques provide an ordered list of the most important predictors for the dynamic parameters. Scientists gain additional insight in these studies by evaluating descriptive statistics and performing correlation analysis.

Over the years, the volume and sophistication of climate analysis have progressed rapidly; but methods to display and scrutinize the information in useful and meaningful ways have not kept pace. For example, researchers have relied on simple scatter plots and histograms 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 by Ware [4]), especially when searching for combinations of conditions. 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. [5]. 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. What's more, few multivariate visualization techniques have been integrated with the statistical analysis techniques (e.g. multiple regression and correlation) that are commonly utilized in climate studies. To compensate for these deficiencies and enable more effective climate analysis, new solutions based on geovisual analytics are needed that intelligently integrate statistical processes and accommodate the simultaneous display of multidimensional multivariate, environmental data.

The geovisual analytics system developed in this research was built around an enhanced display panel that is primarily based on a popular multidimensional multivariate information visualization technique called the parallel coordinate plot (PCP). The PCP was first introduced by Inselberg [6] to represent hyper-dimensional geometries and later applied to the analysis of multivariate data relationships by Wegman [7]. Since then, several innovative PCP extensions have been introduced.

This geovisual analytics system provides a unique PCP-based interface by fusing together variations of several PCP interaction techniques. Fundamental PCP extensions described in prior publications [8,9] such as frequency

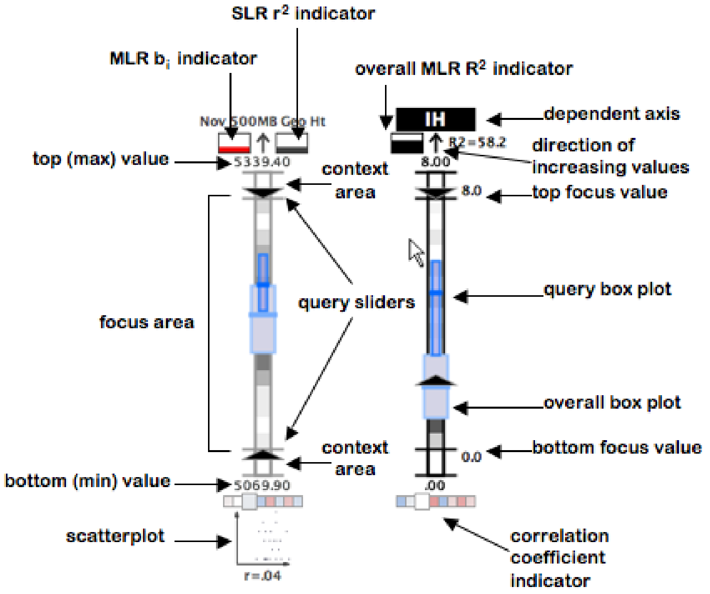


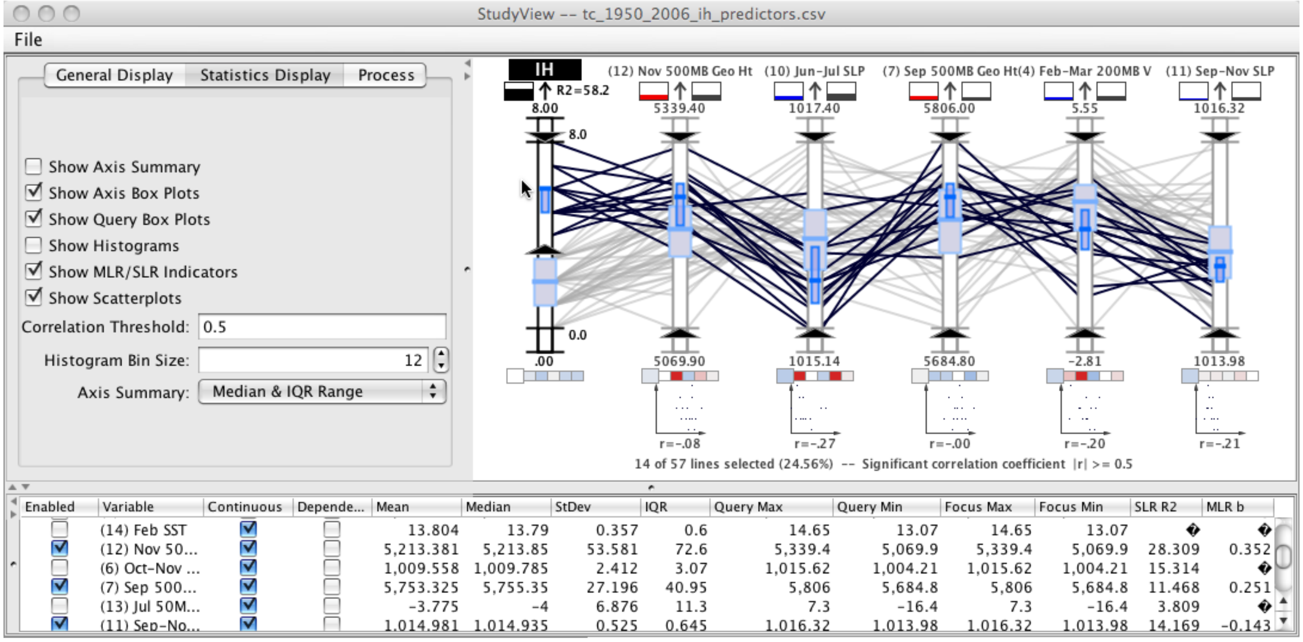
Figure 1: An annotated view of the PCP axis display widget for the system highlighting the visual interaction components and statistical indicators. The axis shown on the left illustrates the normal axis shading while the axis on the right illustrates a highlighted, dependent axis shading.

information display, dynamic axis re-ordering, axis inversion, and details-on-demand are provided in this system. The system also provides several more innovative PCP capabilities derived from recent research publications. The double-ended sliders on each PCP axis (see Fig. 1) facilitate the investigation of subsets of data [9]. The system also provides an axis scaling capability, inspired by earlier approaches [10–13], that captures the mouse wheel movement to allow the user to interactively tunnel through the data until a smaller subset of the original data is in focus. This focus+context technique pushes outliers into context regions of the axis display thereby reducing clutter in the central focus area. Also the system uses aerial perspective shading [14] to encode the proximity of polylines to the mouse cursor which allows the viewer to quickly investigate multidimensional associations in the data set.

These visualization features are coupled with statistical indicators and automated statistical analysis capabilities resulting in a unique system tailored for advanced climate study. To support the interactive analysis capabilities of the system, each axis graphically represents key descriptive statistics (see Fig. 1). In addition to representing frequency information, the axes represent central tendency and variability statistics in the form of box plots for the data in the central focus area of each axis. To facilitate correlation analysis, correlation indicators that represent the correlation between axis pairs are represented beneath each axis in a color-coded set of blocks. The correlation representation was derived from an exploded correlation matrix plot. Furthermore, small scatterplots are displayed for each axis when an axis is highlighted yielding a more detailed view of the variable associations such as nonlinear relationships.

In addition to simple linear regression (SLR), the system offers stepwise multiple linear regression (MLR) capabilities with a backwards glance which selects the optimum number of the most important variables using a predefined significance level. These regression capabilities complement the visual analysis capabilities of the system by isolating the significant variables in a quantitative fashion. Although a full description of the regression interactions and indicators is beyond the scope of this abstract, Fig. 1 and Fig. 2 show the system representation of regression results in the color-filled boxes beneath the axis name which are filled bottom to top according to the magnitude of the resulting MLR coefficient and SLR r^2 value.

The application of PCP-based geovisual analytics to the study of climate predictors is an important contribu-



☐ Show Axis Summary
☑ Show Axis Box Plots
☑ Show Query Box Plots
☐ Show Histograms
☑ Show MLR/SLR Indicators
☑ Show Scatterplots
Correlation Threshold: 0.5
Histogram Bin Size: 12
Axis Summary: Median & IQR Range

14 of 57 lines selected (24.56%) -- Significant correlation coefficient |r| >= 0.5

Enabled	Variable	Continuous	Depende...	Mean	Median	StDev	IQR	Query Max	Query Min	Focus Max	Focus Min	SLR R2	MLR b
☐	(14) Feb SST	☑	☐	13.804	13.79	0.357	0.6	14.65	13.07	14.65	13.07	�	�
☑	(12) Nov 50...	☑	☐	5,213.381	5,213.85	53.581	72.6	5,339.4	5,069.9	5,339.4	5,069.9	28.309	0.352
☐	(6) Oct-Nov ...	☑	☐	1,009.558	1,009.785	2.412	3.07	1,015.62	1,004.21	1,015.62	1,004.21	15.314	�
☑	(7) Sep 500...	☑	☐	5,753.325	5,755.35	27.196	40.95	5,806	5,684.8	5,806	5,684.8	11.468	0.251
☐	(13) Jul 50M...	☑	☐	-3.775	-4	6.876	11.3	7.3	-16.4	7.3	-16.4	3.809	�
☑	(11) Sep-No...	☑	☐	1,014.981	1,014.935	0.525	0.645	1,016.32	1,013.98	1,016.32	1,013.98	14.169	-0.143

Figure 2: The geovisual analytics system developed in this research is composed of a settings panel (upper left), parallel coordinates plot panel (upper right), and table view panel (lower). The statistical indicators, correlation/regression indicators, dynamic query, and discrete aerial perspective line shading features on hurricane trend data.

tion from this research. During the system development and testing, a systematic workflow for using the system in climate analysis was developed to guide the scientist through the knowledge discovery process. Moreover, the effectiveness of the system was evaluated in a hurricane climate study where the primary objective was to find the most important predictors for seasonal intense hurricane (*IH*) activity in the North Atlantic to improve forecasting skill. The secondary objective of the case study is to discover additional temporal associations among the predictors in the data.

In the climate study, a data set containing sixteen potential environmental predictors observed annually for fifty-seven years was analyzed. The data set was provided by Dr. Phil Klotzbach of Colorado State University, where it was used operationally to predict hurricane activity for storm seasons by categories. The focus of this study was on the *IH* activity in a hurricane season. Initially, the data set was scrutinized for general and temporal associations using the interaction capabilities and descriptive statistical indicators. Later, several interesting patterns were discovered using the correlation and regression analysis features. Although a full discussion of the findings from this study are beyond the scope of this paper, the identification of the most important predictors reveals the potential of this system for enhanced climate analysis.

In Fig. 2, a plot from the multiple and single regression analysis is shown. Here the predictors selected by the MLR model are arranged according to the magnitude of the regression coefficient using the dependent variable, *IH*. Using the interquartile range, the predictors have been further characterized by examining their behavior in years with above normal and below normal activity. In Fig. 2, years with above normal *IH* activity are highlighted. From this analysis, the predictor ranges for above normal years were recorded. Then, confirmatory analysis revealed that using these predictor ranges to predict the *IH* activity in a season would have resulted in successfully identifying 11 of the 14 seasons (74%) that had a high number of intense hurricanes between 1950 and 2006. Minor adjustments were applied to the predictor ranges to capture all 14 seasons with above normal *IH* activity. These numerical predictor ranges can be used to predict the *IH* activity of future topical cyclone seasons. In addition to this exploratory analysis, the physical relationships of the selected predictors were ascertained to demonstrate the validity of their selection from a weather science perspective. The full description of this case study represents one of the most comprehensive validations of geovisual analytics in the weather science domain.

The geovisual analytics system developed in this research provides a comprehensive environment for multivariate analysis by combining an enhanced PCP with automated statistical analysis. Using this system, a systematic workflow was formulated to explore climate data. The system has also been evaluated via a real-world case study where the primary objective was to identify and quantify the most important seasonal predictors for *IH* activity. Using traditional climate analysis techniques alone in this case study would require the examination of about 136 scatterplots to observe the associations that are efficiently captured by one frame of the system presented in this paper. The results of this practical evaluation suggest that geovisual analytics can be used in conjunction with statistical processes to more efficiently conduct real-world, multivariate data analysis on complex environmental data sets, particularly those related to climate analysis. Furthermore, this research effort addresses the NIH/NSF Visualization Challenges Report recommendation that visualization researchers "collaborate closely with domain experts who have driving tasks in data-rich fields to produce tools and techniques that solve clear real-world needs [15]" through the inclusion of a hurricane expert throughout the design and evaluation of the system.

Acknowledgments

This research is sponsored by the Naval Research Laboratory's Long-Term Training Program and the National Oceanographic and Atmospheric Administration (NOAA) with grants NA060AR4600181 and NA050AR4601145, and through the Northern Gulf Institute funded by grant NA06OAR4320264. 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

[1] F. Vitart, "Dynamical seasonal forecasts of tropical storm statistics," in *Hurricanes and Typhoons: Past, Present, and Future* (R. J. Murnane and K.-B. Liu, eds.), pp. 354–392, Columbia University Press, Dec. 2004.

[2] P. J. Klotzbach, W. M. Gray, and W. Thorson, "Extended range forecast of Atlantic seasonal hurricane activity and U.S. landfall strike probability for 2007," tech. rep., 2006. http://tropical.atmos.colostate.edu/Forecasts/2006/dec2006/ (current 31 Mar. 2008).

[3] P. J. Fitzpatrick, *Understanding and Forecasting Tropical Cyclone Intensity Change.* PhD thesis, Department of Atmospheric Sciences, Colorado State University, Fort Collins, CO, 1996.

[4] C. Ware, *Information Visualization: Perception for Design.* San Francisco, CA: Morgan Kaufmann, 2nd. ed., 2004.

[5] C. G. Healey, L. Tateosian, J. T. Enns, and M. Remple, "Perceptually-based brush strokes for nonphotorealistic visualization," *ACM Transactions on Graphics*, vol. 23, no. 1, pp. 64–96, 2004.

[6] A. Inselberg, "The plane with parallel coordinates," *The Visual Computer*, vol. 1, no. 4, pp. 69–91, 1985.

[7] E. J. Wegman, "Hyperdimensional data analysis using parallel coordinates," *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 664–675, 1990.

[8] H. Hauser, F. Ledermann, and H. Doleisch, "Angular brushing of extended parallel coordinates," in *Proceedings of IEEE Symposium on Information Visualization 2002*, (Boston, MA), pp. 127–130, IEEE Computer Society, 2002.

[9] H. Siirtola, "Direct manipulation of parallel coordinates," in *Proceedings of the International Conference on Information Visualisation*, (London, England), pp. 373–378, IEEE Computer Society, 2000.

[10] Y.-H. Fua, M. O. Ward, and E. A. Rundensteiner, "Hierarchical parallel coordinates for exploration of large datasets," in *Proceedings of IEEE Visualization*, (San Francisco, California), pp. 43–50, IEEE Computer Society, Oct. 1999.

[11] A. O. Artero, M. C. F. de Oliveira, and H. Levkowitz, "Uncovering clusters in crowded parallel coordinates visualization," in *IEEE Symposium on Information Visualization*, (Austin, Texas), pp. 81–88, IEEE Computer Society, Oct. 2004.

[12] J. Johansson, P. Ljung, M. Jern, and M. Cooper, "Revealing structure within clustered parallel coordinates displays," in *IEEE Symposium on Information Visualization*, (Minneapolis, Minnesota), pp. 125–132, IEEE Computer Society, Oct. 2005.

[13] M. Novotńy and H. Hauser, "Outlier-preserving focus+context visualization in parallel coordinates," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, pp. 893–900, 2006.

[14] T. J. Jankun-Kelly and C. Waters, "Illustrative rendering for information visualization," in *Posters Compendium: IEEE Visualization 2006*, (Baltimore, MD), pp. 42–43, IEEE Computer Society, 2006.

[15] C. Johnson, R. Moorhead, T. Munzner, H. Pfister, P. Rheingans, and T. S. Yoo, eds., *NIH/NSF Visualization Reserach Challenges.* IEEE Press, 2006. http://tab.computer.org/vgtc/vrc/index.html (current 31 Mar. 2008).