Eurographics/ IEEE-VGTC Symposium on Visualization 2008
A. Vilanova, A. Telea, G. Scheuermann, and T. Möller
(Guest Editors)

Volume 27 (2008), Number 3

Results of a User Study on 2D Hurricane Visualization

Joel P. Martin[1], J. Edward Swan II[2], Robert J. Moorhead II[1], Zhanping Liu[1], and Shangshu Cai[1]

[1]GeoResources Institute, Mississippi State University, USA
[2]Department of Computer Science and Engineering, Mississippi State University, USA

Abstract

We present the results from a user study looking at the ability of observers to mentally integrate wind direction and magnitude over a vector field. The data set chosen for the study is an MM5 (PSU/NCAR Mesoscale Model) simulation of Hurricane Lili over the Gulf of Mexico as it approaches the southeastern United States. Nine observers participated in the study. This study investigates the effect of layering on the observer's ability to determine the magnitude and direction of a vector field. We found a tendency for observers to underestimate the magnitude of the vectors and a counter-clockwise bias when determining the average direction of a vector field. We completed an additional study with two observers to try to uncover the source of the counter-clockwise bias. These results have direct implications to atmospheric scientists, but may also be able to be applied to other fields that use 2D vector fields.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: Evaluation/methodology, I.3.6 [COMPUTER GRAPHICS]: Methodology and Techniques

1. Introduction

In this paper, we report the design and results of a user study that investigated the current visualization methods used in the weather modeling community and determined their efficacy. The study used model output simulating Hurricane Lili (2002). This study concentrated on the ability of observers to integrate both the magnitude and the direction of vector fields over an area and to determine the effects of layering in a 2D vector field.

2. Background

Hurricane Lili began forming on the west coast of Africa on September 16, 2002 and became a hurricane on September 30 over Cayman Brac and Little Cayman Islands. As Lili approached the southeastern United States, it intensified, reaching a maximum wind speed of 125 knots (category four). However, the storm unexpectedly weakened in the 13 hours before landfall, becoming a category two hurricane. Lili made landfall in the United States on October 3 near Intracoastal City, LA with a maximum wind speed of 80 knots [Law03]. Lili's path as it approached the U.S. can be seen in Figure 1.

The Hurricane Lili data used in this study were generated by Zhang et al. [ZXPF07] using the Fifth-Generation NCAR / Penn State Mesoscale Model (MM5). They attribute the rapid weakening of Hurricane Lili to a dry air infusion that can be seen in the MM5 model when output from the *Aqua* Moderate Resolution Imaging Spectroradiometer (MODIS) satellite is included in the model run. To study

Figure 1: *Hurricane Lili's track in Google Earth. Track data is from the National Hurricane Center [Law03]. Image credit: Mahnas Jean Mohammadi-Aragh.*

their simulation results, these domain experts created imagery using Read/Interpolate/Plot Version 4 (RIP) [Sto06], a visual analysis package that is a de facto standard in that community. RIP uses a domain-specific glyph to visualize flow.

The past two decades have seen a wide variety of new visualization techniques. Texture-based methods [LHD*04] are able to produce high-resolution and visually pleasing results. However, geometry-based techniques such as streamlines and hedgehogs remain the most common visualization techniques used to investigate real-world flow phenomena [War08]. For the most part, there is only anecdotal or assumptive understanding of the effectiveness of these techniques. The visualization research community is beginning to accept that the efficacy of most visualization techniques need to be verified by user studies [JMM*06].

Some flow visualization user studies have been reported. Laidlaw et al. [LKJ*05] dealt with detecting and identifying critical points. They also studied an advection task in which the observer estimated where a particle in a flow field would move. This was done using several common flow visualization techniques including arrow glyphs and line-integral convolution (LIC). They found that arrows on a regular grid were generally less effective than the other techniques they tested. Andrysco [And05] only dealt with the advection of a particle, but included streamlines and pathlines. Also, the observers were asked to advect a particle back to its origin. Unfortunately, Andrysco was unable to find statistical significance, which was contributed to the test methodology and observers that did not take the tasks seriously.

This weather and flow visualization study is focused on domain-specific 2D glyphs, a simple and intuitive technique still in widespread use. While some weather researchers have started to consider 3D visualization approaches, many continue to rely on 2D methods. In this study, we are interested in investigating the overall performance of these glyphs coupled with one or more layers of scalar data. This will pave the way for future work on other flow visualization techniques.

3. Experimental Task and Setting

To encourage weather researchers to start using visualization methods that we felt had greater efficacy, we designed a user study to test some potential improvements to their approach. We were particularly interested in how image layering affected an observer's performance. For this study, we defined layering as the number of different techniques (contours, color map, state boundaries) that are added to a base image. Our hypothesis was that as more layers are added, the observer's ability to interpret the base image will be degraded. Additionally, we were interested in an observer's ability to mentally compute the average direction of vectors over an area.

3.1. Experimental Design and Procedure

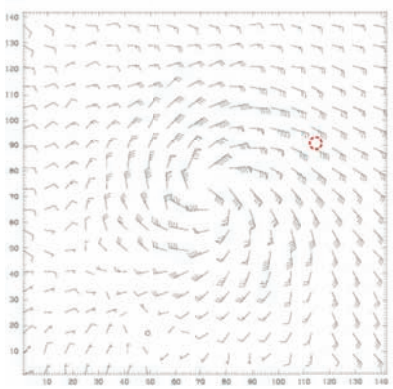

Figure 2: *Point selection circle that the observer would see. The selected point is at the center of the red/white circle.*

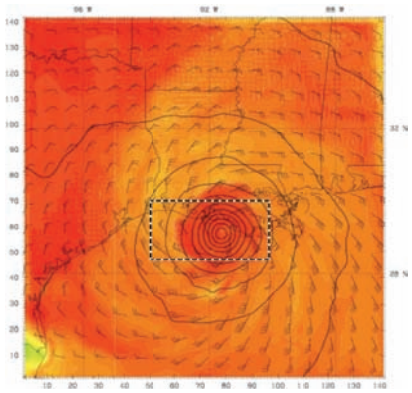

Figure 3: *Area selection box that the observer would see. The selected area is the area included in the black/white box.*

The prevailing wind magnitude and direction is an important aspect of weather understanding and forecasting. In this experiment, we presented observers with vector fields denoted by glyphs (Figure 2) that indicated both wind magnitude and wind direction. While not common in the visualization community, we used the vector glyphs that were used in [ZXPF07] and are common in the weather modeling community (e.g. of the nine articles in the January 2008 issue of Monthly Weather Review that contained glyphs for wind, seven of them used the glyph style shown in Figure 2) [Sch08]. All of the data sets we used showed the Gulf of Mexico region during the 24 hours prior to the landfall of Hurricane Lili, hence all of the data sets contained a spiral wind field and eye structure (e.g., Figure 2). Note this is the period when Lili rapidly weakened from a category four to a category one hurricane on the Saffir-Simpson Scale. We examined observer performance on two different sub-tasks:

Magnitude Estimation: We asked observers to estimate the

wind speed at either a point (Figure 2) or over a rectangular area (Figure 3). The glyphs encoded the wind speed by the number and type of extensions drawn on the right-hand side of the glyph; observers always saw this encoding given in the legend shown in Figure 4. In the point case, the observer's task was to estimate the magnitude at the specified location using the glyphs immediately surrounding it. In the area case, the observer's task was to average the wind speed of all of the glyphs contained within the selection area.

Direction Estimation: We asked observers to estimate the wind direction at either a point (Figure 2) or over a rectangular area (Figure 3). The glyphs encoded the wind direction using a weather-vane metaphor: we imagine that we mount the tip of the glyph on a rigidly-fixed, rotating point; the wind catches the 'flags' along the back of the glyph and rotates the glyph to point along with wind direction (Figure 4). As with all low pressure systems in the northern hemisphere (hurricanes included), the wind field rotates in a counter-clockwise direction. This can also be seen in Figure 4. In the point case, the observer's task was to estimate the wind direction at the specified point using the surrounding glyphs. In the area case, the observer's task was to visually integrate the directions of all of the glyphs contained in the selection area and report the average direction. Two glyphs of equal magnitude that point in opposite directions would cancel each other out.

The motivation behind these tasks is that it is important for atmospheric scientists to understand the magnitude and direction of wind, both at discrete points, as well as over extended areas. This is different from the tasks presented by Laidlaw et al. and Andrysco because they tested the observer's ability to determine where a single particle would go rather than determining the observer's understanding over an area [LKJ*05, And05]. While in the most general case atmospheric scientists might be interested in various types of extended areas, for tractability in this experiment we only examined rectangular areas of certain sizes. Imagery for this study was created using Read/Interpolate/Plot Version 4 (RIP) [Sto06] and was based off the images included in [ZXPF07].

During the training and test, observers were in a conference room with only the test administrator(s). The observer was asked to sign a consent form and answer a general questionnaire, as well as given an incentive for completing the study ($10). The observer was also asked to turn off their cell phone or set it to silent.

The test was administered on an Apple PowerBook laptop with a 15.2 inch screen running at the screen's native resolution of 1280x854. Observers were free to move the laptop to a comfortable position. Most observers placed the laptop within an arm's length, but no specific measurements were taken of this. A Logitech V270 Bluetooth mouse was placed beside the computer. The touchpad was still active for observers that wanted to use it. As many potential distractions were removed from the screen as possible (e.g., the screen saver was turned off, icons removed, etc.).

Observers were given training on the testing software based on a bulleted training script. There were 15 specific points that the administrator covered with the observer during training. The test administrator walked the observer through four training images (two point and two area). These locations and images were randomly selected at the same time the images for the trials were selected. The observer was taught the meaning of the glyphs, instructed on the use of the two sliders (magnitude and direction), and instructed to only use the glyphs in determining their answers (i.e. they were not to use contours, color map or state lines to answer the questions). Observers were given opportunities to ask questions before beginning the test.

The observer then began the test. The test administrator watched as the observers closely during the first few questions to see if their answers were reasonable. Some continued to ask questions leading to response time outliers, but magnitude and direction were generally unaffected by this. Had an observer's responses been significantly different than expected, a new training file could have been generated and the test restarted. This, however, was not necessary.

The observer was given an opportunity for a break after every 10 questions. There was no set length for these breaks and the observer could continue at any point. The magnitude and direction the observer entered for each task was recorded, along with the time it took to complete each subtask. The test administrator recorded notes about events that might affect results (e.g. interruptions or additional questions). Each observer completed 80 tasks, where one task was answering the magnitude and direction for one image.

3.1.1. GUI and Software

To implement the study, several scripts and a GUI were written. The scripts were written in Perl; the GUI is a Cocoa application. The hurricane images presented to the observer are 607x607 pixels. The full GUI occupies 1269x716 pixels.

To generate the list of tests an observer would encounter, a script randomly selects a data set, time step, and region of interest. The region selected is restricted so that it is not within 100 pixels of the edge of the data (in screen-space). The test conditions (point/area, color map, state lines, pressure contours) are randomly ordered, but balanced so that there is an equal number of each test condition for each observer. A second script determines the ground truth for each selected location by examining the input data for the imagery.

During a test, the GUI (Figure 4) displays the predetermined tests to the observer. It records the magnitude and direction that the observer selects, along with the time it takes the observer to make their decisions. The output is cleaned and formatted before inputting it into the statistics packages Minitab, SPSS, and R. Minitab was used for publication-quality graphics and statistical tests, SPSS for

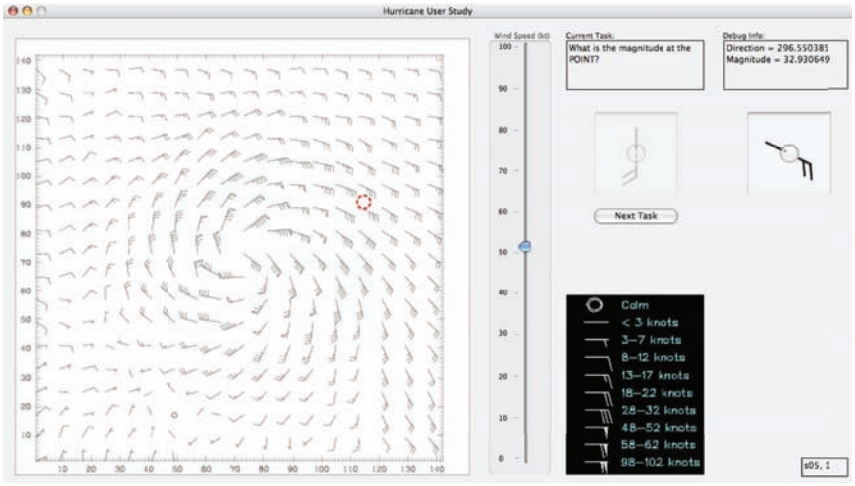

Figure 4: *The user study GUI in training mode. Note the correct answer on the right side of the screen. This was removed once the test began.*

statistical tests, and R for exploratory and preliminary analysis graphics.

3.2. Independent Variables

Layers: All of the images contained wind glyphs. Additionally, the images contained zero to three additional layers which we hypothesized would reduce the observer's ability to determine the magnitude and direction of the glyphs. The layers were pressure contour lines, state boundary lines, and a color map of relative humidity, for a total of eight different layer combinations. Figure 5 shows the four base images.

Selection: The selection for each question was either a point or an area (Figures 2 and 3). For a point, we told observers to use the four vectors surrounding the designated point to determine its value. For an area, we told observers to mentally integrate all of the vectors surrounded by the box. For black and white images, a red/white indicator was used; for color images, a black/white indicator was used. These can also be seen in figures 2 and 3. These colors were picked because they provided good contrast when they were applied.

Shapes: For area selections, three shapes were chosen: square, horizontal rectangle, and vertical rectangle. Two sizes of each shape were presented to the observer (sizes are in screen-pixels):

- Square: 100x100 and 175x175
- Vertical rectangle: 50x150 and 100x200
- Horizontal rectangle: 150x50 200x100

3.3. Dependent Variables

There were two main measured quantities: wind magnitude (velocity in knots) and wind direction (in degrees). For each, we have a ground truth measurement and the observer's measurement, as well as the observer's response time. From these, we calculated four dependent measures: (1) magnitude error, (2) magnitude error response time, (3) direction error, and (4) direction error response time.

Magnitude Error: The magnitude error is computed as the difference in the ground-truth-magnitude and the observer-magnitude. A magnitude error of 0 means the observer has chosen the ground truth wind speed. A positive magnitude error means the observer has overestimated the ground truth wind speed, while a negative magnitude error means the observer has underestimated the ground truth wind speed.

Direction Error: The direction is collected in degrees; 0° / 360° is due north; degrees are left-handed, meaning they increase clockwise. For processing purposes, the direction is converted to $-180^\circ \leq$ direction $\leq +180^\circ$, where 0° is still north. The direction error is the difference in the ground-truth direction and the observer-direction. A direction error of 0° means the observer has chosen the correct ground truth direction. A positive direction error means the observer has overestimated the ground truth direction in a positive left-handed sense; i.e. the observer's direction is clockwise from the ground truth direction. A negative direction error means the observer has underestimated the ground truth direction in a negative left-handed sense; i.e. the observer's direction is counter-clockwise from the ground truth direction.

Response Time: Response time is defined as the number of microseconds from the time the observer is shown the question until the observer presses the 'next task' button. Separate values are collected for both magnitude and direction.

3.4. Observers

Initially, we had planned to recruit observers who were Mississippi State University students and faculty in the Broadcast and Operational Meteorology programs. However, as the study design progressed, we determined that a strong

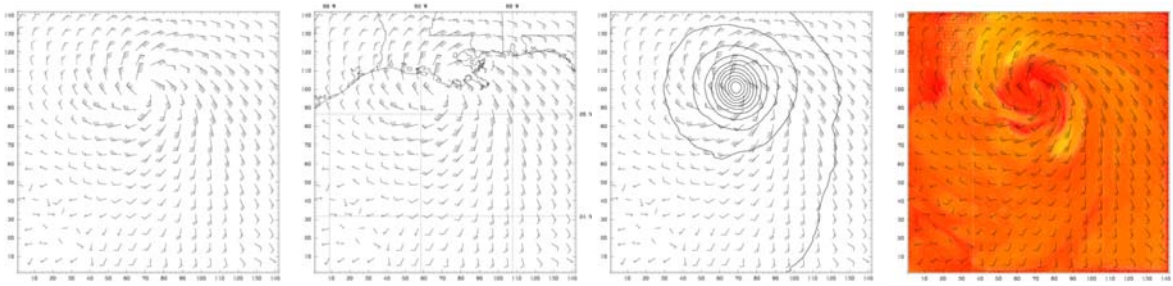

Figure 5: *Each of the four basic types of images. Combinations of these four formed the eight image types used in the study. From left, (1) glyphs only (2) glyphs + state lines (3) glyphs + pressure contours (4) glyphs + color map.*

background in weather sciences was not necessary to complete the study. Researchers from various fields commonly look at vector visualizations and should be capable of answering the questions we were asking. As a result, the observers recruited for this study include students and faculty with weather, visualization, or computational fluid dynamics backgrounds. At this point, nine observers have participated in the main study, comprising six males and three females. In addition, as described in Section 4 (Results and Discussion), we recruited two additional male observers to study reversed glyphs. Thus a total of eleven observers have participated, but unless otherwise specified, all of the results are based on the first nine observers. None of the observers reported color blindness in the questionnaire. Since observers were not interpreting the colors in the images, we did not perform tests to verify their responses.

4. Results and Discussion

Each of the nine observers who participated in the main study completed 80 trials, for a total of 720 completed trials. Each trial produced a value for magnitude error, magnitude response time, directional error, and directional response time. We analyzed the data using standard error plots and univariate analysis of variance (ANOVA). For the ANOVA, we modeled our experiment as a repeated-measures design that considers *observer* a random variable and all other independent variables as fixed. The distributions on which ANOVA analysis is based assume that, for each tested effect, the data is normally distributed and the variance is homogenous. For repeated-measures designs such as the ones we report here, these two assumptions are usually violated [How02]. Therefore, following the recommendations of Howell (p. 486), for each tested effect we applied the Huynh and Feldt correction ε; when the F-test is conducted, the degrees of freedom are multiplied by ε. This results in a more conservative test that corrects for the degree to which the ANOVA assumptions are violated. For our analysis, we applied the Huynh and Feldt correction whenever the data was completely balanced (and thus it was possible to calculate ε using SPSS). However, some of our F-tests were over unbalanced data, and thus we were not able to calculate this correction for every F-test. Therefore, we do not report Huynh and Feldt-corrected F-tests in this section. However, whenever we were able to calculate the more conservative Huynh and Feldt correction we did so, and we did not find any cases where the Huynh and Feldt correction changed the outcome of an F-test from significant to non-significant.

We processed outliers in the data with the procedure described by Barnett and Lewis [BL94]. We determined outliers by examining histograms that summarized each dependent measure; for magnitude error and directional error the histograms showed symmetric normal distributions, while for response times the histograms showed skewed normal distributions. We determined outliers on a case-by-case basis, by examining the tails of the distributions and noting values that appeared after conspicuous gaps in the histogram. Each outlier was replaced by the median of the remaining values in the experimental cell. Given that outliers are considered mistaken values, this procedure improves the calculation of means, standard errors, and the sums-of-squares terms used in ANOVA, which would otherwise be inappropriately influenced by the outlying values.

4.1. Wind Magnitude

In the wind magnitude data, there were 31 outliers that needed to be processed. This is 31 of 720 responses (4.3% of the values). While this is quite a few outliers, it reflects the difficulty of determining the magnitude, particularly over areas. There is negative bias as all observers tended to underestimate wind speed by an average of 4.0 knots; this underestimation is significantly different from zero ($F(1,8) = 2629, p < .000$). Figure 6 shows the magnitude error results for each observer. All observers underestimated the magnitude, from an average of 1.8 knots for observer 2 to an average of 5.7 knots for observer 5.

Figure 7 shows the main results for magnitude error. In the results, we encoded *layers* as a three-digit binary number where the digits indicate, from left to right, the presence or absence of contour lines, color map, and state lines ("1" indicates presence and "0" indicates absence).

Observers were considerably more accurate with points than with areas ($F(1,8) = 151.2, p < .000$). This is expected

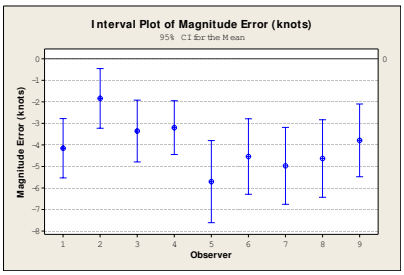

Figure 6: *Magnitude Error vs. Observers.*

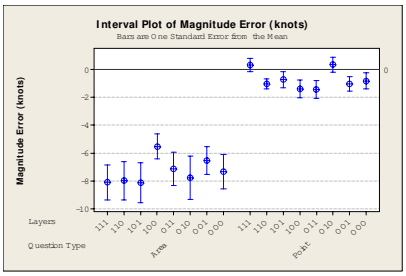

Figure 7: *Magnitude Error vs. Question Type, Layers.*

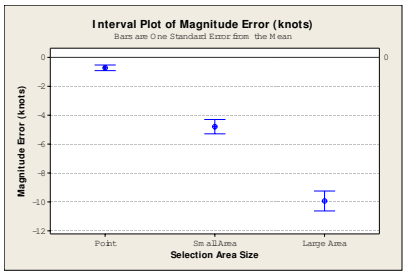

Figure 8: *Magnitude Error vs. Selection Area Size.*

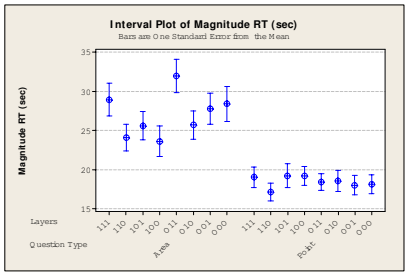

Figure 9: *Magnitude Response Time vs. Question Type, Layers.*

because integrating over areas should be more difficult than determining the magnitude at a single point. It also shows that the overall underestimation trend comes mostly from area tasks. Because the standard error bars overlap within both area and point questions, there is no evidence that the number of layers make a systematic difference in observers' ability to determine the magnitude; analysis showed no effect of layers ($F(7,56) = .392, p = .903$), nor any interaction with question type ($F(7,56) = .894, p = .518$). For us, these are somewhat negative results, as they do not support our initial hypothesis that increasing the number of layers in the imagery would make the task more difficult. It is possible that the specific portion of the data set queried, and perhaps the queried value itself, make a much larger difference than the number of layers.

Figure 8 shows that the task gets harder as the size of the selection area increases ($F(2,16) = 84.5, p < .000$). Observers increasingly underestimate the magnitude for larger squares or rectangles. We did not find any effect of selection area shape; observers gave equivalent results for squares, horizontal rectangles, and vertical rectangles.

We also recoded and analyzed response times for each magnitude trial. Figure 9 shows that the average response time for areas was greater than the time for points ($F(1,8) = 26.8, p = .001$), which is further evidence that the area questions were harder than the point questions. We found an overall main effect of layer ($F(7,56) = 2.12, p = .057$), which is further analyzed below. We found a trend towards an interaction between layer and question type ($F(7,56) = 1.9, p = .090$). Figure 10 shows that observers took longer to answer with larger areas ($F(2,16) = 25.4, p < .000$), which is further evidence that larger areas were more difficult. We found no response time differences for different area shapes.

Figure 9 suggests a layer effect for areas, but not for points. Indeed, there was a main effect of layers for the area trials ($F(7,56) = 2.75, p = .016$). Figure 11 shows the response times for the absence ("0") and presence ("1") of each layer for the 360 area trials. We found main effects for the presence or absence of contour lines ($F(1,8) = 12.3, p = .008$), color maps ($F(1,8) = 4.74, p = .061$), and state lines ($F(1,8) = 18.1, p = .003$), but no interaction effects. Contrary to our hypothesis, observers were *faster* in the presence of contour lines. We conjecture that the contour lines helped observers estimate the wind speed because they frame the wind fields (Figure 5). As we expected, observers were slower when color maps and especially state lines were

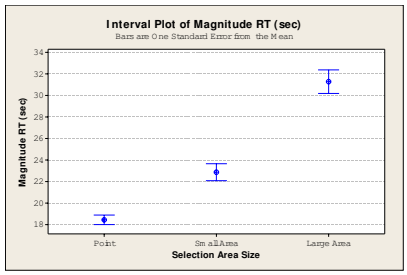

Figure 10: *Magnitude Response Time vs. Selection Area Size.*

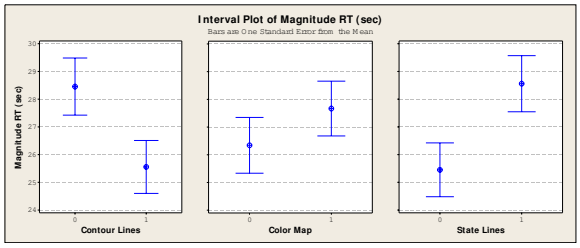

Figure 11: *Magnitude Response Time vs. Contour Lines, Color Map, State Lines for the area questions (360 trials).*

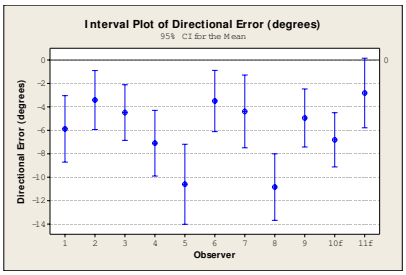

Figure 12: *Directional Error vs. Observers (880 trials). Observers 10 and 11 saw reversed glyphs (Figure 14).*

present, which indicates that these layers made the task more difficult. Interestingly, although color maps are much more visually salient than the state lines (Figure 5), the magnitude of the effect for color maps (1.33 seconds) was much smaller than for state lines (3.11 seconds). This may occur because the relative humidity shown by the color maps also follows the wind field, while the state lines are completely arbitrary with respect to the wind field. In addition, note that these results are only for response time — the lack of magnitude error results indicates that observers' accuracy was not effected by the layers.

4.2. Wind Direction

The second dependent measure in this study was wind direction. Out of 720 wind direction responses, 27 (2.9%) outliers were removed. The directional error shows a negative (counter-clockwise) bias of -6.1°, which is significantly different from zero ($F(1,8) = 41.8, p < .000$). Figure 12 shows the directional error results for each observer; observers 10 and 11 saw "flipped" glyphs and are discussed in more detail later. All observers showed a negative bias, which ranged from an average of -3.41° for observer 2 to an average of -10.84° for observer 8.

Figure 13 shows that the magnitude of the negative bias increased as the size of the shape increased ($F(2,16) = 16.9, p < .000$). We did not find any effects of area shape on directional error.

We also analyzed the response time for directional error. Our only response time finding was that it took observers

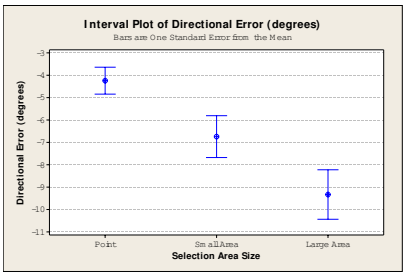

Figure 13: *Directional Error vs. Selection Area Size.*

Figure 14: *The glyph that is produced by RIP is shown on the left. In order to prove that the counter-clockwise bias is caused by the shape of the glyph, we implemented a small study (2 observers) where the glyph shape is flipped, as shown on the right.*

longer to enter the direction for area questions (11.3 seconds) than for point questions (10.0 seconds) ($F(1,8) = 6.08, p = .039$).

4.3. Directional Error Study

Considering that these glyphs are commonly used by the weather science community, it is interesting that observers consistently (1) underestimated the wind magnitude shown by the glyphs, and (2) showed a consistent counter-clockwise bias when estimating wind direction. We hypothesized that the counter-clockwise bias was likely due to the glyphs' asymmetric visual design, where the wind-speed flags are always on the left (Figure 14). To quickly test this hypothesis, we flipped the glyph orientation (Figure 14), and ran two additional observers through exactly the same protocol. We expected that this would give us a clockwise bias with a similar magnitude as the RIP's regular glyphs did. Figures 12 and 15 show the results. Contrary to our hypothesis, our two observers still displayed a counter-clockwise bias with the flipped glyphs, and the magnitude of the bias is comparable to what we found with our first nine observers. The difference between the groups is not significant ($F(1,878) = 1.41, p = .235$). Further study will be required to determine what, if any, role glyph shape plays in this directional bias.

5. Conclusions and Future Work

Our experiment has empirically verified some expected results, as well as revealed some surprising and unexpected results. Among the expected results is that determining the magnitude and direction of wind speed over an area is a

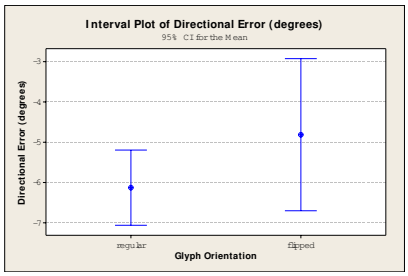

Figure 15: *Directional Error vs. Glyph Orientation (880 trials).*

harder task than at a point. Another unsurprising result is that the larger the area that needs to be mentally integrated, the harder the task and the longer it takes.

This experiment revealed three surprising findings. First, although some data layers (state lines) increased the response time for wind speed estimation over areas by as much as 3.11 seconds, overall we did not find that additional data layers made the glyphs more difficult to read. Second, observers underestimated wind speed by an average 4.0 knots; and their underestimation became worse as the area over which they were estimating increased. Third, observers showed an average counter-clockwise bias of –6.1° in wind direction, and this bias also became worse as the area over which they were estimating increased. The wind direction bias cannot be entirely explained by the asymmetric nature of the glyph shape.

As discussed in the introduction, the glyphs studied here are widely used in the weather science community. Our unexpected positive finding is that these glyphs can be reliably read in the presence of additional data layers. However, our unexpected negative finding is that observers reliably underestimated wind speed and showed a bias in estimating wind direction. Because understanding wind speed and direction are fundamental tasks in weather data analysis, this is a potentially serious finding.

5.1. Future Work

It is well-known in the visualization field that the design space for any glyph-based vector field visualization technique is very large. A fruitful goal for future work would be to study related glyph techniques to see if they exhibit the same sorts of biases we found here. Another fruitful goal would be to tweak the parameters of the current glyphs, which are widely used, to see if they can be improved. For example, perhaps the wind speed underestimation could be reduced if the flags that denote wind speed were made larger (Figure 4).

Finally, while a strength of the current study is that it used real-world data, it is possible that the nature of the dataset influenced the results. In particular, we suspect that the directional bias might have arisen because every dataset showed a counter-clockwise circular wind field around a northern hemisphere tropical cyclone. We would like to repeat the study with a southern hemisphere tropical cyclone data set, where the wind field would rotate in a clockwise direction.

6. Acknowledgements

This work was supported by the High Performance Computing and Visualization Initiative (HPCVI) funding provided by the Department of Defense (DoD) High Performance Computer Modernization Program (HPCMP) through the Army Corps of Engineers Engineering Research and Development Center (ERDC) and Jackson State University (JSU).

References

[And05] ANDRYSCO N.: A User Study Contrasting 2D Unsteady Vector Field Visualization Techniques. Master's thesis, The Ohio State University, 2005.

[BL94] BARNETT V., LEWIS T.: Outliers in Statistical Data, 3rd ed. John Wiley and Sons, 1994.

[How02] HOWELL D. C.: Statistical Methods for Psychology, 5th ed. Duxbury, Pacific Grove, CA, 2002.

[JMM*06] JOHNSON C. R., MOORHEAD II R. J., MUNZNER T., PFISTER H., RHEINGANS P., YOO T. S. (Eds.): NIH-NSF Visualization Research Challenges Report, 1st ed. IEEE Press, Los Alamitos, CA, USA, 2006.

[Law03] LAWRENCE M. B.: Tropical cyclone report: Hurricane Lili, Apr. 2003.

[LHD*04] LARAMEE R., HAUSER H., DOLEISCH H., VROLIJK B., POST F., WEISKOPF D.: The State of the Art in Flow Visualization: Dense and Texture-Based Techniques. Computer Graphics Forum 23, 2 (2004), 203–221.

[LKJ*05] LAIDLAW D. H., KIRBY R. M., JACKSON C. D., DAVIDSON J. S., MILLER T. S., DA SILVA M., WARREN W. H., TARR M. J.: Comparing 2D vector field visualization methods: A user study. IEEE Transactions on Visualization and Computer Graphics 11, 1 (Jan./Feb. 2005), 59–70.

[Sch08] SCHULTZ D. M. (Ed.): Monthly Weather Review, vol. 136. American Meteorological Society, 2008.

[Sto06] STOELINGA M. T.: Users' Guide to RIP Version 4: A Program for Visualizing Mesoscale Model Output. University of Washington, Dec. 2006.

[War08] WARE C.: Toward a perceptual theory of flow visualization. Computer Graphics and Applications 28, 2 (March/April 2008), 6–11.

[ZXPF07] ZHANG X., XIAO Q., P.J. FITZPATRICK .: The impact of multisatellite data on the initialization and simulation of hurricane Lili's (2002) rapid weakening phase. Monthly Weather Review 135, 2 (Feb. 2007), 526–548.