Int. J. Human-Computer Studies (2000) **53,** 827–850
doi:10.1006/ijhc.2000.0419
Available online at http://www.idealibrary.com on IDEAL®

Turning pictures into numbers: extracting and generating information from complex visualizations

J. GREGORY TRAFTON

HCI Laboratory, Naval Research Laboratory, Code 5513, 4555 Overlook Av. S.W., Washington, DC 20375-5337, USA. email:trafton@itd.nrl.navy.mil.

SUSAN S. KIRSCHENBAUM

Naval Undersea Warfare Center Division, Newport, RI, USA

TED L. TSUI

Marine Meteorology Division, Naval Research Laboratory, Monterey, CA, USA

ROBERT T. MIYAMOTO

Applied Physics Laboratory, University of Washington, Seattle, WA, USA

JAMES A. BALLAS

HCI Laboratory, Naval Research Laboratory, Code 5513, 4555 Overlook Av. S.W., Washington, DC 20375-5337, *USA*

PAULA D. RAYMOND

Center for Applied Research, Great Falls, VA, USA

(Received 14 February 2000, and accepted in revised form 31 May 2000)

We present a study of complex visualization usage by expert meteorological forecasters. We performed a protocol analysis and examined the types of visualizations they examined. We present evidence for how experts are able to make use of complex visualizations. Our findings suggest that users of complex visualizations create qualitative mental models from which they can then generate quantitative information. In order to build their qualitative mental models, forecasters integrated information across multiple visualizations and extracted primarily qualitative information from visualizations in a goal-directed manner. We discuss both theoretical and practical implications of this study.

1. Introduction

Everywhere you look, newspapers and banners declare this the "Information Age." Scientists, engineers and weather forecasters (just to name a few) examine extremely large amounts of data on a daily basis. The visualizations that they examine have so much data that it is obvious that they cannot understand all the data all the time. How do experts deal with this large amount of information? This paper presents a study showing how expert weather forecasters extract information from complex visualizations, integrate that information into a qualitative mental model, and then use that information to make a (mostly) quantitative forecast.

How do people comprehend and use complex visualizations? Very little theoretical or empirical work has been done on complex visualization comprehension, but there is a sizable research body on graph comprehension. Bertin (1983) presented a task analysis of how graphs are comprehended. A great deal of the past research on graph comprehension, however, has focused on the encoding and perception of the graph. Cleveland and McGill, for example, have examined the psychophysical aspects of graphical perception (Cleveland & McGill, 1984, 1986). Similarly, Pinker's theory of graph comprehension, while quite broad, focuses on the encoding and understanding of graphs (Lewandowsky & Spence, 1990; Pinker, 1990). Kosslyn's work emphasizes the cognitive processes that make a graph more or less difficult to read, while focusing on the encoding and perception of graphs (Kosslyn, 1989).

Unfortunately, the generalizability of past research on graph comprehension to more complex visualizations is unclear. Most of the past research has focused on artificial situations and artificial domains where the only usable information available to participants is the graph itself. For example, a traditional task given to participants in graph comprehension experiments is to read information directly off of a graph of a made-up company or product (Carter, 1947; Sparrow, 1989; Pinker, 1990; Lohse, 1993). In contrast, real-world users of complex visualizations not only have a great deal of domain knowledge but have to decide how and what information to visualize, what information to extract and what to do with that knowledge. Users of complex visualizations also have particular reasons for extracting information. They use that information for deducing, generalizing, extrapolating and predicting (to name just a few goals).

Additionally, most studies have focused on data spaces that are purposely limited to 2 or 3 variables (Simkin & Hastie, 1987; Casali & Gaylin, 1988; Sparrow, 1989). Experimenters limit their data spaces to 2 or 3 variables to retain experimental control. This experimental control is not only necessary for performing good science, but it is critical to understanding how encoding and perception of graphs occurs. If complex graphs and visualizations were used, the standard participants in experiments (undergraduates), would probably be very confused. In contrast, many scientific and engineering disciplines use complex data spaces that have many variables over a wide range of scales and across time. For example, it is not uncommon for a typical meteorological forecasting chart (shown in Figure 1) to contain five or more variables. Figure 1, for example, represents at least six variables: (1) The latitude/longitude, (2) the geographical region (southern California), (3) wind speed, (4) wind direction, (5) temperature and (6) barometric pressure. If this information is examined across time or multiple altitudes are examined (as happened in this study), the number of variable greatly increases.

To study how complex visualizations are comprehended and used, we believe that it is necessary to study use of complex visualizations "in the wild" (e.g. Hutchins, 1995) or "*in vivo*" (e.g. Dunbar, 1995, 1996; Trickett, Trafton & Schunn, 2000; Trickett, Fu, Schunn & Trafton, 2000) in addition to studying situations where there is a great deal of laboratory control (e.g. Lohse, 1993; Shah & Carpenter, 1995; Carpenter & Shah, 1998).

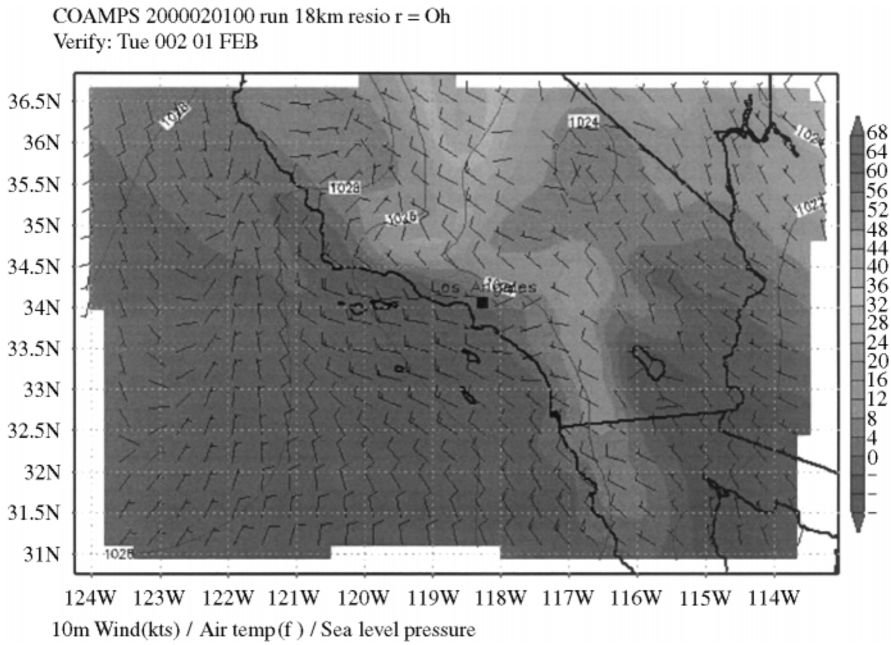

FIGURE 1. A sample screenshot of a typical visualization that forecasters used. This visualization was made using a COAMPS model run and shows wind speed and direction, temperature and pressure at the surface. The original picture is in color.

We propose that users of complex visualizations use heuristics to help them deal with the large amount of data they have access to. We believe that three of these heuristics are (1) attending to information in a goal-driven manner; (2) integrating information across visualizations and (3) building a qualitative mental model from which they can abstract quantitative information.

Weather forecasting is an appropriate domain in which to study these phenomena for several reasons. First, weather forecasting is extremely complex and the visualizations that forecasters use often have at least five different variables in addition to spatial and temporal information. However, forecasters are among the best decision makers, making accurate forecasts a majority of the time (e.g. it rains 70% of the times that they forecast a 70% chance of rain; Murphy & Winkler, 1992). There are computational weather models that make forecasts, but their interpretation requires a great deal of training, knowledge and information. In addition, the weather models need to be updated with current data frequently and re-run. Weather models rarely use data that is as up to date as forecasters would like.

2. Method

The experiment was an exploratory investigation with some of the characteristics of a field study. We were unable to control the actual weather and the occasional computer problems such as World Wide Web (WWW) sites being down, computer crashes and the like. These are realistic problems normally encountered by METOC (METeorological and OCeanographic) forecasters.

2.1. TASK

The task assigned to the forecasters was to prepare a written information brief for an airplane flown from an aircraft carrier at a given offshore location 12 h in the future. They were given flight path, altitudes, destinations, alternate airports and expected times of arrival. The destination was Whidbey Island, Washington, DC. The brief was to cover the entire round trip with primary emphasis over the destination. The intended audience was the flight coordinator and the pilots, several hours before takeoff. The forecasters were requested to provide specific data for departure, enroute, destination, tanker refueling (if applicable), recovery and alternate airfield. Forecasters also provided astronomical data for solar and lunar conditions. Thus, the forecasters were told to determine detailed qualitative and quantitative information on what the weather would be 12 h in the future and to write that information in a standard briefing package. This type of task is routine for Navy and Marine forecasters; they were not being asked to do anything out of the ordinary, and all forecasters felt comfortable with their task. It should also be noted that not every forecaster filled in every piece of information (due to time constraints). Table 1 shows the information requested as well as a sample of the kind of information that was written down on the brief.

2.2. PARTICIPANTS

The participant sample was representative of the range of expertise and training within the population. Participants were selected on the basis of scheduling availability. Two individuals participated in each session. All forecasters were Naval or Marine Corps forecasters and forecasters-in-training. All had completed at least the first-level weather school. They ranged in forecasting experience from 1 to 16 years with the more experienced individual in each session serving as lead forecaster and the less experienced serving as technician. All forecasters had significant operational experience. Table 2 shows the qualifications of the forecasters and technicians who participated in this study. Because of the forecasters' training and experience, we categorized them as experts (Hayes, 1985; Ericsson, Krampe & Tesch-Roemer, 1993). In addition, two senior METOC Officers played the role of the regional center and of an intelligence officer. The regional center is typically staffed with experienced forecasters; it is their job to help out the local forecasters on the ship and provide advanced and/or specialized visualizations. In this case, the center was available to provide this help to the forecasters in the study. The intelligence officer provided detailed information about other military-related questions.

Both the forecaster and the technician provided talk-aloud protocols (Ericsson & Simon, 1993). Participants were asked to talk aloud as they were working through their forecasting task. If they needed to speak to someone else (usually their partner), they stopped the talk-aloud protocol; when they went back to working on their own, they continued their verbal protocol. Three pairs of forecasters performed the forecasting and briefing tasks.

Table 1

A partial description of weather information that participants were asked to provide for a brief to pilots. This information comes from several different briefs

Area	Information to fill in	Example
Destination weather	Weather synopsis Wind direction and speed (altitude to source) Temperature max/min Cloud bases/heights/amounts	Mostly clear 080: 24 010 050: 28 005 (etc.) Max: 17°/Min: 11° 0
Astronomical data	Sunrise Sunset Sun angle Azimuth Moonrise Moonset Illumination	0544 2055 – 20° 340° 2112 0556 99%
Departure weather	Skies Source wind Visibility Turbulence Icing weather warnings/advisory	Mostly cloudy; bases 010–020′ 290/04 KTS Unrestricted/07 NIL NIL MWA27C
Enroute weather	Flight level 050 Flight level 100 Flight level 150 Flight level 200 Flight level 250 Flight level 300 Turbulence Icing	300/07 KTS 250/26 KTS 240/30 KTS 260/19 KTS Lgt sfc-140 NIL
Tanker weather	Synopsis	Bases: BKN010 Tops: 025
Recovery weather	Skies Winds Visibility	BKN-OVC010/015 CASE III 32 005 KTS 5
Alternate airfield weather	Synopsis	Mostly clear
Water survivability	With survival suit Without survival suit	5.3 h 1.3 h
General	Remarks	

TABLE 2
Forecaster and technician qualifications

Session	Job	Years METOC experience	No. of briefs/years
1	Forecaster	9	100
1	Technician	3	50
2	Forecaster	11	180
2	Technician	5	180
3	Forecaster	16	600
3	Technician	1	200

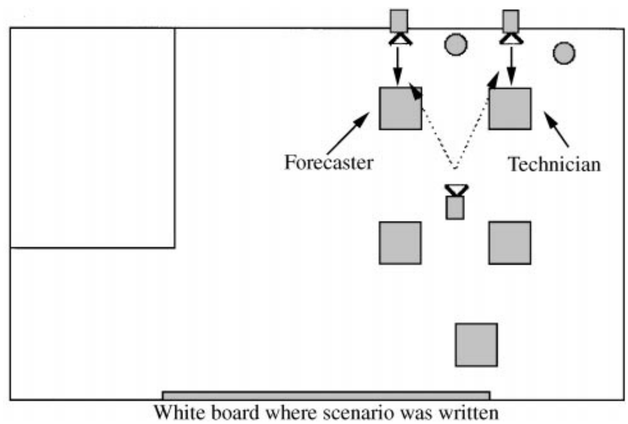

FIGURE 2. A simple schematic of how the room where data were collected was arranged. Both the forecaster's and the technician's monitors could display all visualizations. [video camera icon], video camera; [square icon], computer terminal; [circle icon], SME with notepad.

2.3. SETTING AND APPARATUS

The experimental sessions took place in a large room with tables across from each other. On one side of the room two Windowns-NT workstations were arranged side by side. The forecaster and technician worked on these two computers. In addition, there was one subject matter expert (SME) standing behind the forecaster and one SME standing behind the technician. The two SMEs took notes about what they thought the pair was doing. On the opposite side of the room were three workstations which were used as the regional center. One of the senior METOC officers was in charge of the regional center. One of the center's computers was connected to the forecaster's workstation by the secure military link called SIPERNET. The other workstations were (1) an unclassified workstation to draw information from the unclassified Internet and (2) a workstation with the meso-scale forecasting tool, TAMS/RT loaded on it. TAMS/RT is described in more detail below. All communications between the simulated shipboard and the simulated Regional Center were carried out over a chat tool familiar to all, called IRC Chat. Figure 2 shows a simple schematic of this layout.

2.4. FORECASTING BACKGROUND

There were two primary programs that forecasters used: Joint METOC Viewer (JMV) and various satellite images viewed over the World Wide Web. JMV allows weather model data to be visualized in a variety of ways. There are several types of weather model data. The two weather models that were used most often by the forecasters in this study were COAMPS and NOGAPS. COAMPS and NOGAPS describe and predict different atmospheric features in terms of their scales. NOGAPS resolves large-scale weather features (e.g. cold fronts and warm fronts), while COAMPS can resolve fine scale (or meso-scale) weather features (e.g. thunderstorms). Earth terrain, of course, affects the weather significantly. Local terrain features can influence the meso-scale weather features, and large mountain ranges can modify the large-scale weather events. It typically takes hours to run each of these weather models, so no model is perfectly up to date, since it will be based on data that is several hours old (at least).

After the type of weather model data has been downloaded, JMV provides a graphical user interface (GUI) to inspect, examine and animate the data. A user may select up to five variables at a time (wind speed and direction, pressure or temperature at various heights, etc.) and view those variables over a particular geographical area. It is also possible to extract exact information on those variables at a particular latitude and longitude by moving the mouse over a location and reading off text information that appears over the visualization. JMV also allows users to cycle through time to see how different variables change over time. Figure 1 shows a screen snapshot of a typical JMV visualization. Figures 3 and 4 show two kinds of visualizations that forecasters examined to compare the differences between winds at different heights.

Satellite pictures are available in three different kinds: visual (essentially a big camera looking down on the earth), Infrared radiation (IR) and 6.75 μm channel (the amount of water vapor in the atmosphere). It should be emphasized that JMV provides weather model data from which predictions of the future weather are to be made. However, JMV is only as good as the inputs and the weather model being run. Sometimes, of course, the weather models can be dramatically wrong. Satellite images show “truth” as it was some time in the recent past, but there is no capability for prediction in satellite visualizations.

In addition to JMV and various satellite pictures, a number of special-purpose tools were available to the forecasters.

- EOTDA, winEOTDA and TAWS, used to determine the Electro-Optical ranges for a variety of sensors.
- AREPS, an electro-magnetic (EM) effects decision aid.
- Solar/Lunar Almanac Program (SLAP) for sunrise, sunset and illumination calculations.
- TAMS/RT, the on-scene meso-scale numerical weather prediction system. It is capable of resolving meso-scale features and to be run locally at a very fine resolution.

Video recorders were positioned to capture the forecaster’s and the technician’s computer screens. A third video recorder captured interactions between the two forecasters.

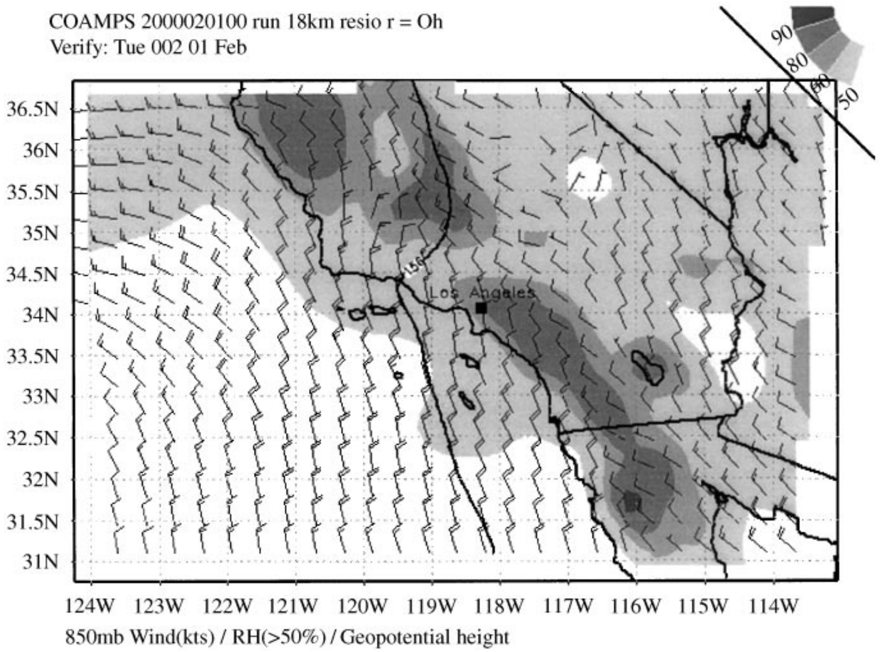

FIGURE 3. A sample screenshot of a typical visualization that forecasters used. This visualization was made using a COAMPS model run and shows wind speed and direction and relative humidity at 850 mb. The original picture is in color.

2.5. PROCEDURE

Each trial was run with slightly different task parameters. The task differed slightly because the daily weather conditions at the destination location differed because real weather data and weather changes were used (i.e. a forecast was made using current weather, not a simulation or past data).

Each session began with a description of the task by the Intelligence Officer, giving destination, times and other information as needed. This information was also recorded on a whiteboard that was visible to both the forecaster and the technician for reference when needed. Forecasters created briefs using either PowerPoint or simply wrote down the information on paper.

The forecaster served as leader during the session, requesting information from the technician as needed. As the technicians differed widely in experience, the forecasters sometimes gave detailed instructions on where to find specific information, which tools to use and how to use the tools. In other groups, the forecaster and technician worked relatively independently. This type of interaction was in keeping with the way that METOC officers operate ship-board.

For example, all forecasters would ask their technician to perform simple tasks like finding a good satellite image that showed the Whidbey Island area. The technician would work on that task for a while, and then after a good satellite image had been found, the forecaster would look over the technician's shoulder and discuss different

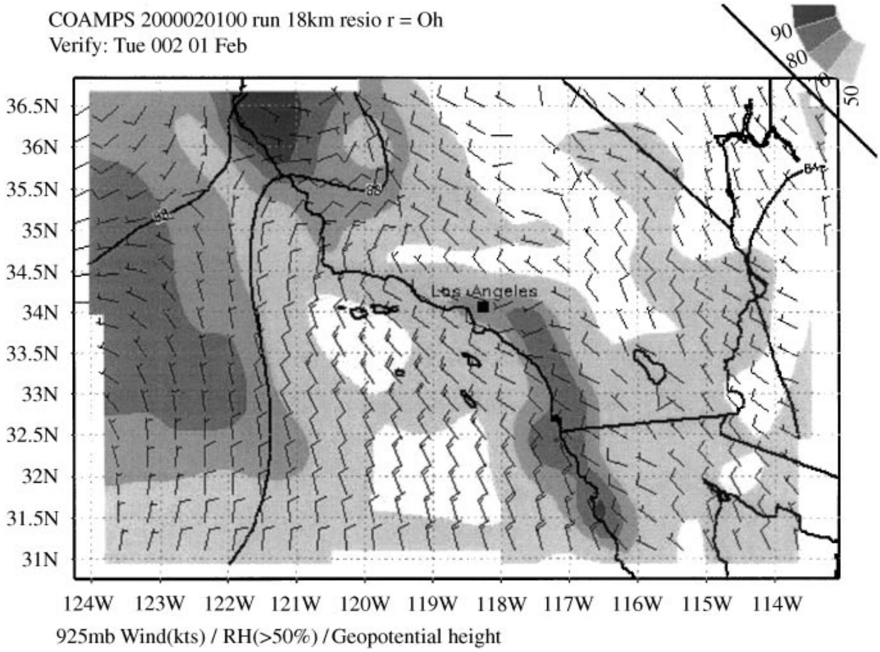

FIGURE 4. A sample screenshot of a typical visualization that forecasters used. This visualization was made using a COAMPS model run and shows wind speed and direction and relative humidity at 925 mb. The original picture is in color.

features of the image or forecasting task. The communicative aspects of this task are discussed more fully in a separate report (McFarlane & Trafton, in preparation).

Forecasters were given 2 h to complete their forecast and brief. Everyone completed the task within the time allotted. When time allowed, the forecaster gave his brief at the end of the session. All sessions were concluded with a debrief during which the experimenters had an opportunity to ask questions and the participants had an opportunity to give feedback to the experimenters.

2.6. COGNITIVE TASK ANALYSIS

A cognitive task analysis (CTA) was performed. The overall importance of the structure of the task is reflected in later analyses: all later analyses will be structured around the CTA.

For the sake of brevity, only a high-level description will be provided.

Forecasters went through four stages to complete their task: (1) initialize their understanding of the large scale (global, whole US, etc.) weather; (2) build a qualitative mental model (QMM) of the weather; (3) verify and adjust their QMM and (d) write the brief.

A brief description of this process follows.

2.6.1. Initialization of large-scale weather picture. First, the forecaster looks at the large-scale weather to see if there are any fronts, anomalies, etc. that they should be particularly aware of. This allows the forecaster to place the details in context and observe any fronts or other features that may impact their forecast. For example, a forecaster at this stage may say something like "I'm going to look at global satellite pictures."

2.6.2. Building the QMM of the weather. Second, the forecaster generates a qualitative mental model of the weather status and trends in the area. He does this by examining the main aspects of the weather in the specific area he wants to forecast at a rather detailed level. For example, the forecaster will tend to look at pressure changes and wind changes in the destination area with respect to time and height. This is the most difficult and time-consuming aspect of the entire forecasting process. When this stage is finished, the forecaster will have a qualitative mental model (QMM) of the main features of the weather around the area of interest. Most of their knowledge will be qualitative (though they will have looked at a lot of quantitative information), and they will be able to generate quantitative information from their QMM. This type of QMM has been discussed by other researchers in the domain of meteorology (Hoffman, 1991; Pliske, Crandall & Klein, in press) and in other domains studied by cognitive scientists (Gentner & Stevens, 1983; Johnson-Laird, 1983; Taylor & Tversky, 1992).

For example, a forecaster attempting to build a QMM of the weather around the destination may say "I'm going to interpolate between these reports...there may be an upper level low in this area" or the forecaster may look at wind direction and speeds across an area and say "There could be some mixing in this area."

2.6.3. Verify and adjust QMM. Third, they will verify and adjust their qualitative mental model. They will do this by comparing their QMM to other weather information (i.e. another source). The other source could be another forecaster, a different satellite image (at comparable times), or another model forecast. They may need to make adjustments to their QMM at this stage if need be. This stage is necessary because, as mentioned before, the numerical weather prediction models are not perfect and are always at least several hours out of date.

For example, a forecaster attempting to verify and adjust his QMM may look at a satellite image ("truth" sometime in the recent past) and say "The low on the satellite image and the low on the JMV image don't quite match up, so I'm going to have to [mentally] adjust the wind speed...". By doing this, the forecaster is explicitly comparing the weather model results to the current status and adjusting his QMM.

2.6.4. Brief writing. Finally, they will fill out the brief and make the official forecast. The majority (over 85%) of the information the forecasters were requested to fill out was quantitative (or numerical, see Table 1). The specific forecast information may come from their QMM (i.e. their own mental extrapolation) and/or tools they have at hand (i.e. various visualizations like JMV).

This cognitive task analysis was presented as if forecasters went through each stage in order, but there is some iteration between the steps, particularly when building the QMM and verifying and adjusting it with another source.

There was extensive evidence for these different stages in the data and general agreement by the project's SMEs and other forecasters, but this report will not discuss these details, instead focusing on how information is extracted and used in this complex domain.

2.7. CODING SCHEME

The protocol and visualizations of the forecaster and the technician (when the forecaster looked over the shoulder of the technician for one of the stages above) were coded using the MacSHAPA software package (Sanderson Scott, Johnston, Mainzer, Watanabe & James, 1994). MacSHAPA is a protocol analysis tool that allows the coder to associate categories, notes and timestamps with events on tape.

The notes that the two SMEs took were used to facilitate coding and category membership.

The first thing that was done was to categorize the protocol (both visualization and protocol) depending on what cognitive task analysis stage the forecaster was at. All following coding was therefore associated with a particular stage.

2.7.1. Visualization coding. Multiple data sources were used by the forecasters. JMV visualizations and satellite images were the most commonly used, but Skew-T's (temperature profiles at a particular geographic location up through the atmosphere), Horizontal Weather Descriptions (general weather information) as well as several other visualizations were examined. Each time the forecaster examined a visualization (whether on his computer screen or the technician's), it was recorded and identified which visualization it was (JMV, IR satellite image, etc.). Each visualization was also categorized as either a *picture* (actual data like a satellite image) or a *chart* (which displayed representations of data like a JMV image that used weather model data) or a *graph* (like a bar graph of a temperature profile). Forecasters also used paper forecasts called Terminal Aerodrome Forecasts (TAFs). These were classified as *text*.

2.7.2. Usage coding. We also categorized every utterance that participants made in terms of how they used the visualizations. Several categories were used: a *goal* was coded when the forecaster described a plan for the future. An *extract* was coded any time the forecaster extracted some information from a visualization or made an inference based on that visualization. A *brief-writing* code was given any time the forecaster wrote down something for the brief. Table 3 shows samples of these categories.

Many of the goals seemed to be organized in an opportunistic manner by the forecaster, based on either memory or the display (i.e. Altmann & Trafton, 1999*a*, 1999*b*, under review). The goals were not further coded, but were referenced by subsequent events (see below for examples).

Data that were coded as Extract were further coded as either *Quantitative* or *Qualitative*, whether there was a *goal* to extract that information or if the data were *opportunistically noticed* and whether the extracted information was being *integrated* with another visualization or if there was *no integration*. Table 4 shows examples of how this coding scheme was implemented.

Data that were coded as Brief-Writing were further categorized as either *Quantitative* or *Qualitative* and whether the information came from the user's mental model (QMM) or if it came from a visualization or from notes they made while looking at a

TABLE 3
Sample coding scheme of the METOC protocol

Category	Example
Goal	I need to look at surface pressure and winds at the 500 millibar height
Extract	There's an area of low pressure off of northern California
Brief-writing	Flight level winds over destination at 050 are 30010

TABLE 4
Example of how the Extract category was coded

Example	Visualization(s)	Type of data	Goal	Integrated
The low is at latitude 33.5 N. and longitude 120 W.	JMV	Quantitative	Goal	No integration
The low is right on the money on JMV compared to the satellite image	JMV and satellite	Qualitative	Goal	Integrated (visualization)
I can assume the JMV model run is handling the system well	JMV and satellite	Qualitative	Goal	Integrated (visualization)
It [JMV] shows a bit more funneling there (compared to his expectations)	JMV	Qualitative	Noticed	Integrated (QMM)
Temperature here is 17°	JMV	Quantitative	Goal	No integration
The numbers are not here so I am going to have to interpolate mentally at 15°	Skew-T	Quantitative	Goal	No integration

visualization. We were able to determine where information came from by examining where the forecaster was looking just before the information was written down and by their protocol. If the forecaster was looking at a visualization and extracting relevant information, that Brief-Writing code was categorized as coming from a visualization. If the forecaster was not looking at a visualization or if the contents of the computer screen was not relevant to the Brief-Writing task (i.e. the computer screen showed a blank IRC chat screen), the information was coded as coming from the forecaster's QMM. In addition, the notes that the SMEs took were used to facilitate this coding (i.e. sometimes the SME's notes said something like "the wind speed the forecaster wrote down came out of the forecaster's head"). Table 5 shows examples of the Brief-Writing coding scheme.

TABLE 5
Example of the Brief-Writing coding scheme

Example	Type of data	Source
Wind speed is 20 knots	Quantitative	Visualization (JMV)
Skies are mostly cloudy	Qualitative	Visualization (satellite)
In route weather, 30 000 foot level, winds will be about 25 (knots)	Quantitative	QMM

The difference between the extraction coding during the Brief stage of the CTA and the Brief-Writing coding was that the extraction coding during the Brief stage contained extraction events (using JMV, etc.), while the Brief-Writing coding contained information they wrote down as part of their brief.

3. Results and Discussion

How are complex visualizations used, given the large amount of data they contain? How do experts keep track of all the information, and how is that information attended to? This results section attempts to answer these questions.

3.1. OVERVIEW

The three forecasters extracted information from various visualization sources an average of 40 times and worked on the brief (Brief-Writing) an average of 30 times. The forecasters looked at an average of 58 different data sources. These numbers are probably an underestimate of the amount of information extracted because of the frequent conversations with the technician (extractions were not coded when the forecaster was tutoring or helping the technician). All further analyses will use the combined data set.

3.2. CODING THE STAGES OF THE CTA

Each Extraction and Brief-Writing code from above was categorized as one of the four stages of the cognitive task analysis presented above. For the most part, the forecasters followed the description above: they initialized their knowledge of global weather, built their Qualitative Mental Model, verified and adjusted their QMM and wrote the brief. There was, of course, some iteration between building and verifying/adjusting the QMM, but it was a straightforward task to put each extraction event into one of the task analysis stages.

Interestingly, there were no explicit extraction events during the initialization stage, though there was evidence that forecasters were in an Initialization stage by making comments like "I'm just going to look at a couple of satellite pictures first to get a sense of what's happening globally." After the satellite picture was examined, however, no explicit information was extracted. These initialization episodes were quite short (no more than

4 min). In a focus group (unreported in this study), expert forecasters reported that they carry a general mental model of weather conditions with them at all times. Expert forecasters have an excellent knowledge of the physical attributes of "normal" weather. For example, the physical attributes of a cold front include pressure and temperature changes, wind shift, cloud distribution and precipitation. In order to make a prediction, however, forecasters have to gather "current" information by examining different visualizations and considering the impact of the Earth's terrain. Access to the latest satellite pictures thus updates the general mental model and initializes the focus on a specific time and place.

Not surprisingly, the majority of the information extraction occurred while the forecaster was building his qualitative mental model, $\chi^2(3) = 70.9$, $p < 0.001$, Bonferroni adjusted χ^2's significant at $p < 0.05$. Table 6 shows the emphasis on the QMM.

Was there any relation between the type of visualization used and the stage of the CTA? To investigate this question, we coded the types of visualization that the forecaster examined. Recall that pictures were depictions of actual data (e.g. a satellite image), charts displayed representations of data (like a JMV image), graphs were more traditional displays (like a bar graph) and the text category contained textual forecasts (TAFs).

As expected, the expert forecasters did use different types of visualizations depending on the stage they were in (see Figure 5). Specificially, forecasters seemed to use primarily chart visualizations (e.g. JMV) while building their QMM, $\chi^2(3) = 94.9$, $p < 0.001$, Bonferroni adjusted χ^2's significant at $p < 0.05$ and primarily pictures (e.g. satellite images) while verifying and adjusting their QMM, $\chi^2(3) = 18.0$, $p < 0.001$, Bonferroni adjusted χ^2's significant at $p < 0.05$. Interestingly, forecasters seemed to draw from all visualization types during the Brief-Writing stage, $\chi^2(3) = 5.2$, $p = 0.16$. This analysis lends further support to the different stages in the CTA and shows the forecasters' sensitivity to using different visualizations to the type of task they were engaged in, at least while building and verifying/adjusting their QMM.

3.3. QUALITATIVE MENTAL MODEL

How and when did the qualitative mental model get built? And to what extent was it qualitative? Recall that we categorized each extraction event as quantitative or

TABLE 6
Cognitive task analysis breakdown

CTA stage	No. of extraction events
Initialization	0
QMM	63
Verify and adjust	26
Brief-writing	25

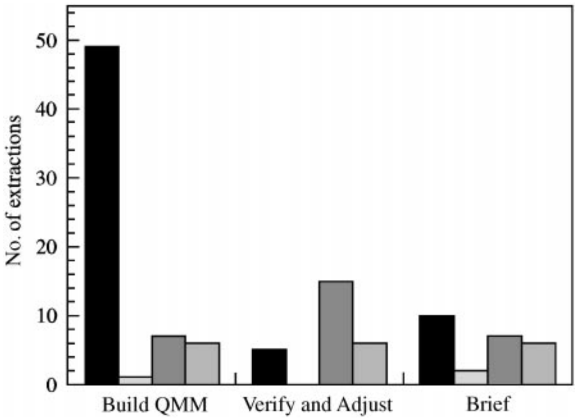

FIGURE 5. The relation between the stage of the CTA and the type of visualization used by the forecasters: ■, chart; ■, graph; ■, picture; ■, text.

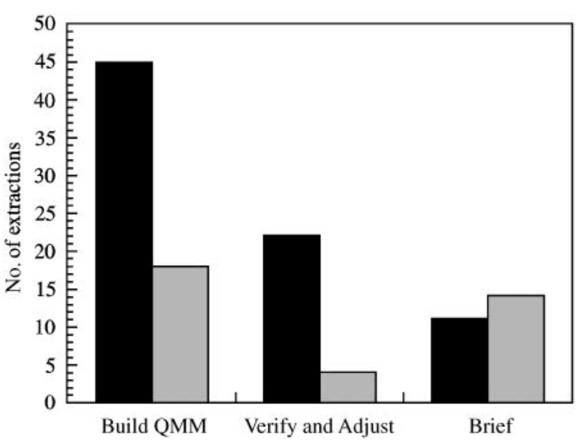

FIGURE 6. Type of information extracted by the forecasters: ■, qualitative; ■, quantitative.

qualitative. As Figure 6 shows, qualitative information is the most common type of information extracted at every stage except for the Brief stage, $\chi^2(5) = 24.4$, $p < 0.001$, Bonferroni adjusted χ^2's significant at $p < 0.05$. This finding is somewhat surprising, since the forecasters' main task was to create a mostly quantitative brief (see Table 1). How forecasters were able to turn their QMM into a numerical brief is discussed below. Note that this figure (and the next two figures) shows the amount and type of information extracted during the Brief stage (i.e. they looked at a JMV visualization to extract a specific number), not information that was written down for the Brief-Writing stage.

This analysis also suggests that the qualitative mental model is aptly named. The forecasters did extract quantitative information, but they extracted far more qualitative information. The idea of a qualitative mental model is in keeping with other research

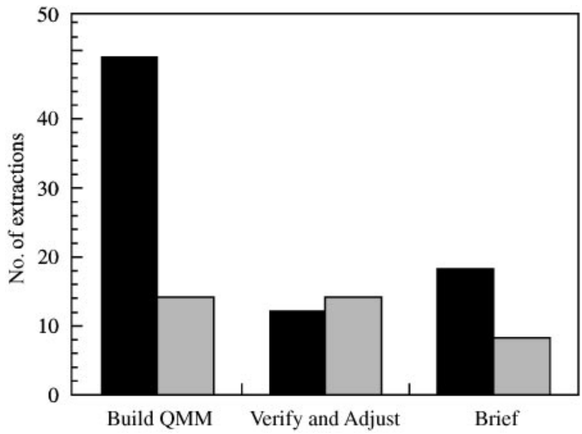

FIGURE 7. How information was attended to by the forecasters: ■, goal; ■, noticed.

implying that mental models are at a "lower resolution" than reality (e.g. Chambers & Reisberg, 1985; Tabachneck-Schijf, Leonardo & Simon, 1997). It should be noted that the tools themselves did not prevent forecasters from extracting quantitative information; the primary tool that forecasters used (JMV) provided quantitative information as part of the graphical user interface.

3.4. DIRECTED EXTRACTION

Given the large amount of information that each of these complex visualizations displayed, how did forecasters extract information? One possibility is that when presented with complex visualizations, users would not only inspect the visualization hoping to see something they found interesting but also directly search for specific information they wanted to extract. As Figure 7 shows, forecasters were very goal-directed while they were building their QMM, $\chi^2(1) = 19.4$, $p < 0.001$. They did sometimes notice features of interest (like more funneling in one visualization than they had seen in a previous visualization), but the strong emphasis was to have a particular goal in mind as they extracted information (cf., Trickett *et al.*, 2000).

3.5. INTEGRATION OF VISUALIZATIONS

The forecasters looked at a very large number of different visualizations. How did the forecasters keep track of the large amount of information from the visualizations? We propose that they built a consistent qualitative mental model by integrating information from multiple visualizations.

Overall, there was a majority of simply extracting information without reference to another visualization (57% vs. 43%). However, as Figure 8 shows, this effect is driven entirely by the lack of integration during the Brief stage. Interestingly, when forecasters were building their QMM, there was a great deal of integration with another visualization (56%). Likewise, during the Verification and Adjustment stage, forecasters spent

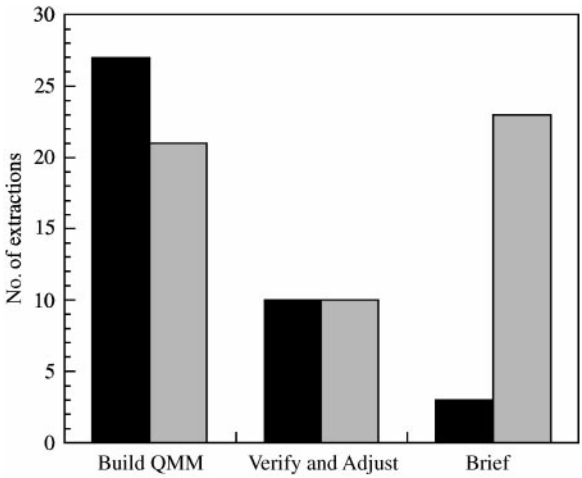

FIGURE 8. The types of integration performed by the forecasters: ■, visualization integration; ■, no integration.

a large amount of effort comparing different visualizations. In contrast, there was very little integration during the Brief stage. Figure 8 shows this interaction between integration and stage of the cognitive task analysis.

3.6. USING THE QMM

Forecasters have spent a great deal of time and effort building up a detailed qualitative mental model of the weather. They are primarily goal-directed when extracting information, they integrate information across multiple visualizations, and they extract qualitative information from the complex visualizations. How, then, is this QMM used, particularly when the brief they are writing requires a majority of quantitative information? To answer this question, the knowledge source was coded when the forecasters were filling out their briefs.

As Figure 9 shows, a great deal of information was generated from the forecaster's qualitative mental model as he filled out his brief. What's more, the majority of the information that comes from the QMM is quantitative (73% vs. 27%, $\chi^2(1) = 7.8$, $p < 0.01$), suggesting that even though forecasters extract primarily qualitative information from the different visualizations, they are able to generate quantitative information from their own mental representation. In support of this idea, in post-study interviews, forecasters said that they thought of their QMM (though they did not refer to it as such, of course) as visual or pictorial, but they could extract quantitative information from it if needed.

To show the interaction between extraction and use of information more clearly, a graph of the type of information extracted during the QMM stage and the type of information that was generated from the QMM and written down for the brief is shown in Figure 10. As the left half of Figure 10 shows, while building the QMM, forecasters extracted primarily qualitative information. This information was used to help them

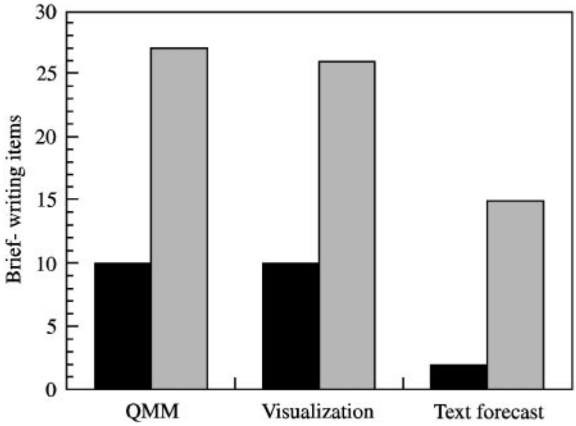

FIGURE 9. The sources that were used by forecasters to write their brief. ■, qualitative; ■, quantitative.

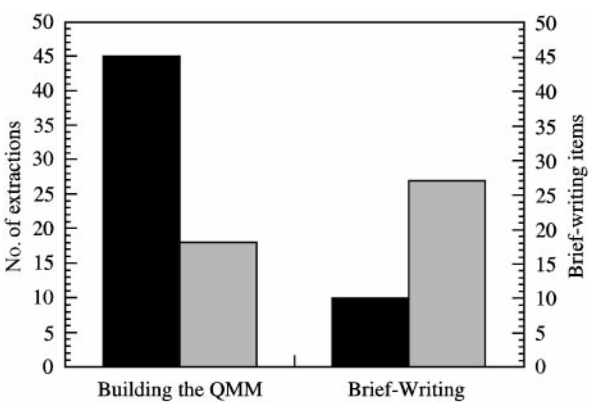

FIGURE 10. The left bars shows the number and type of extractions while building the QMM. The right bars show the type number and type of items generated from the QMM during the Brief-Writing stage. ■, qualitative; ■, quantitative.

construct their QMM. Later, when they actually put their QMM to use by writing down information (Brief-Writing), the majority of information they wrote down was quantitative as the right half of Figure 10 shows. Thus, forecasters extracted qualitative information and constructed their QMM. They then generated numbers for the brief using their QMM.

4. General discussion

How did forecasters deal with the large amount of data they had at their disposal to predict the weather? This study suggests that they used several strategies and heuristics to simplify their job. First, they did not attempt to extract every piece of information from the visualization. Instead, they extracted primarily qualitative information in a goal-directed manner. They also integrated information across visualizations by comparing and contrasting different sources of information. These two heuristics—(1) look at data in a goal-directed way and (2) integrate information from other sources—helped the forecasters keep track of the data and build a qualitative mental model (QMM).

This study suggests that experts look at complex visualizations to extract primarily qualitative information. This qualitative information is then used to build a qualitative mental model. The QMM can then be used to generate both numbers (as it was in this study), qualitative information, and what–if scenarios. Thus, forecasters are able to turn complex visualizations on the screen into numbers by their use of a QMM.

A simple example may help show this more clearly. One forecaster used a COAMPS run on JMV to look at several wind speeds across several different altitudes. While looking at the visualizations (which contained 4–5 different variables), he extracted rather general qualitative information like "There is a lot of mixing at the 50 mb level" and "The low winds are coming up quickly at the destination." This forecaster never extracted specific quantitative information about the wind speed that he would need later for his brief. During the Brief-Writing stage, he did not go back to a visualization but instead said "The winds over the destination at 100 mb will be about 30 knots." This example shows a path of how one forecaster turned a complex visualization into a final number; the evidence presented in this paper suggests that the forecasters used a QMM to accomplish this task.

The QMM itself can be used to extract both qualitative and quantitative information. This is one of the primary purposes of the QMM because forecasters need to be able to manipulate and discuss with others both qualitative and quantitative information about their forecast. Forecasters spent a great deal of time creating a qualitative mental model and then used that qualitative information to generate quantitative predictions. Why did they do it this way, relying on their own cognitions instead of relying on the more precise and potentially more accurate computerized weather models? We believe that forecasters relied on their own QMM because the large amount of quantitative information was simply too much to remember. It is easier to construct and use a QMM to create quantitative information than to remember the multitudes of numbers that are derived from the weather models. In essence, the mind is a much better averaging (and editing) device then the multiple visualizations and weather models (perhaps because each of the weather models has different strengths and weaknesses).

Another reason that forecasters rely on their QMM is that is the way that the information is typically communicated to another forecaster. For example, when discussing the weather, forecasters seem to talk more about the qualitative aspects of the weather (e.g. "There's a low in southern California") than the quantitative aspects of the weather (e.g. wind speed at 950 mb).

Additionally, the verification stage of weather forecasting is critical. During this stage, forecasters match up their QMM with another source (a satellite image, another forecaster, a paper forecast, etc.). They then modify their own QMM based on this information. More often than not, the weather model data (i.e. COAMPS viewed through JMV) will be wrong in some way, so the numeric information that is produced by that weather model will be wrong also. Instead of examining those numbers and then adjusting them, we found that the forecasters instead preferred to adjust their QMM internally and generate their numbers directly from their own mental representation.

4.1. THEORETICAL IMPLICATIONS

4.1.1. Qualitative mental models. How is the forecaster's QMM represented, and what are its features? We hypothesize that the forecaster's QMM is imagerial (e.g. Kosslyn, 1980, 1994), and other researchers have presented evidence that forecasters' representation of the weather is pictorial (Hoffman, 1991; Pliske *et al.*, in press). We feel that the QMM is very similar to Tversky and her colleagues' description of a spatial mental model (Taylor & Tversky, 1992; Tversky, 1991). We believe that the QMM, like a spatial mental model, is perspective free but allows many different perspectives (e.g. either large-scale or meso-scale level) and time scales. However, unlike the spatial mental model, which focuses primarily on relations between landmark objects, a QMM represents different physical properties (like wind speed) and the interaction between them (i.e. how pressure interacts with wind speed at multiple heights).

The QMM is not infallible nor a perfect image, of course. We expect that the same types of errors that are made with other visual stimuli (cognitive maps, for example) would also be made with qualitative mental models. These errors could include alignment problems (e.g. Tversky, 1981), rotation errors (e.g. Chase, 1983), and general metric errors (Tversky, 1993). How these types of errors manifest themselves in the type of rich QMM studied here is a topic of future interest and research.

4.1.2. Generalizability of the QMM. This study suggests that a qualitative mental model is one of the ways that forecasters understand the vast amount of qualitative and quantitative information available. Is this type of QMM generalizable to other domains? We believe that there are several characteristics of this type of QMM that should generalize to other domains. First, many other domains contain too much multivariate data to comprehend all at once (or even after an extended examination). These domains include many areas of scientific visualization (e.g. astronomy and fMRI visualization, Trickett *et al.*, 2000*a*,*b*) and the scientists working in these domains probably have some form of QMM.

Also, practitioners in complex domains that emphasize prediction probably use some form of QMM. For example, a key component in the hazardous materials domain is prediction. In the HazMat domain, a truck containing toxic chemicals crashes along an interstate highway (for example) and the spill must be cleaned up in a safe and timely manner. Experts in this domain need to consider when or if a fire may start, the rate of spread of the spill, how dangerous the spill and fire may be and other related issues (Iba, Gervasio, Langley & Sage, 1998; Iba & Gervasio, 1999). Though no empirical studies of experts in this domain have been performed, it is easy to imagine that experts would build a QMM from many data sources that would be very similar to the QMM that weather forecasters use.

Thus, the type of QMM discussed in this paper is probably common in a number of domains, including a variety of scientific visualization domains and domains where prediction plays a prominent role.

4.1.3. Integration between different visualizations. Past research has shown that users integrate elements within a graph. For example, Carpenter and Shah (1998) used an eye-tracker to examine participants' eye movements as they were comprehending a graph. They found that participants read and reread the information contained on the axes and labels of graphs, interspersed with the graph itself (though less time overall was spent focusing on the graph itself). They showed clearly that participants integrated information across different features of the same graph. In contrast, this study showed the importance of integrating between different visualizations. We found that, particularly when the forecaster was building, verifying and adjusting his QMM, there was a great deal of integration between different visualizations. This finding suggests that in multivariate and multidimensional domains, practitioners need to look at a very large number of graphs and perhaps integrate information not only within a specific graph as Carpenter and Shah (1998) claim, but also between different graphs.

4.2. PRACTICAL IMPLICATIONS

There are several practical implications that arise from this work. First, the four stages that arose from the cognitive task analysis present an orienting framework that could be used to help meteorological visualizations. Using this approach, each stage could be examined and improved. For example, the Initialization stage is where the forecasters initialize their large-scale understanding of the weather. This stage is critical because forecasters use this information to constrain later stages of the process (i.e. building the QMM, verifying and adjusting the QMM, etc.). Thus, the visualizations that are examined during the Initialization stage function as an anchor where later adjustments can be made while building and verifying and adjusting the QMM (e.g. Slovic, Fischhoff & Lichtenstein, 1980; Tversky & Kahneman, 1974). One way to improve the Initialization stage, then, is to make sure that the original anchor that is used is a very recent satellite image, rather than an old satellite image or (possibly inaccurate) JMV weather model visualizations.

Another way to use the stages of the CTA to improve meteorological forecasting would be to add automation to the type and time that different visualizations are presented. For example, satellite images could be queued up for the verification and adjustment stage. Thus, if a forecaster was looking at a specific area of interest, an "intelligent agent" could find, download and retrieve satellite images to make the verification and adjustment stage more complete.

It would also be possible to improve the way that different visualizations are integrated. Recall that forecasters spent a great deal of time and effort integrating information between different visualizations. One way to decrease "cognitive load" or spatial workload (e.g. spatial transformations, Trafton, in preparation) would be to simplify the comparison and integration process between different visualizations. Probably, the best way of simplifying these cognitive operations would be to overlay or combine visualizations that were frequently integrated. For example, creating a single visualization that combined a JMV visualization and a satellite image would greatly increase the ability of a forecaster to compare, contrast and integrate information from both visualizations. This suggestion is currently being implemented in the next version of JMV. Other visualizations could be integrated by presenting them side by side or so they are both visible at once.

In summary, this paper has shown that meterologists build a qualitative mental model in order to comprehend the vast amount of information available to them. They build their QMM by integrating information across different visualizations and extracting information in a goal-directed manner. This QMM is then used to generate quantitative information.

This research was supported in part by grants 55-7188-00 from the Office of Naval Research and by 55-7850-00 to the first author from the Office of Naval Research. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the US Navy.

We would like to thank Erik Altmann, Chris Schunn and Susan Trickett for comments and suggestions on this research. We would also like to thank two anonymous reviewers for comments on a previous draft.

We would especially like to thank Capt. Chris Gunderson, Kim Curry and Kurt Nelson for providing facilities, expertise, computer equipment and discussions on this research.

References

Altmann, E. M. & Trafton, J. G. (1999*a*). Functional encoding in memory for goals. *Proceedings of the 6th Annual ACT-R Workshop*. George Mason University.

Altmann, E. M. & Trafton, J. G. (1999*b*). Memory for goals: an architectural perspective. In M. Hahn & S. C. Stoness, Eds. *The Proceedings of the 21st Annual Conference of the Cognitive Science Society*, p. 19–24. Simon Fraser University: Lawrence Erlbaum Association.

Altmann, E. M. & Trafton, J. G. Memory for goals in means ends behavior. *Cognitive Science*. (under review).

Bertin, J. (1983). *Semiology of Graphs*. Madison, WI: University of Wisconsin Press.

Carpenter, P. A. & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, **4,** 75–100.

Carter, L. F. (1947). An experiment on the design of tables and graphs used for presenting numerical data. *Journal of Applied Psychology*, **31,** 640–650.

Casali, J. G. & Gaylin, K. B. (1988). Selected graph design variables in four interpretation tasks: a microcomputer-based pilot study. *Behaviour and Information Technology*, **7,** 31–49.

Chambers, D. & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance*, **11,** 317–328.

Chase, W. G. (1983). Spatial representations of taxi drivers. In R. Rogers & J. A. Sloboda, Eds., *Acquisition of Symbolic Skills*. New York: Plenum Press.

Cleveland, W. S. & McGill, R. (1984). Graphical perception: theory, experimentation, and application to the development of graphical method. *Journal of the American Statistical Association*, **79,** 531–553.

Cleveland, W. S. & McGill, R. (1986). An experiment in graphical perception. *International Journal of Man–Machine Studies*, **25,** 491–500.

Dunbar, K. (1995). How scientists really reason: scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson, Eds. *Mechanisms of Insights*, pp. 365–395. Cambridge, MA: MIT Press.

Dunbar, K. (1996). How scientists think: online creativity and conceptual change in science. In T. B. Ward, S. M. Smith & S. Vaid, Eds. *Conceptual Structures and Processes: Emergence, Discovery and Change*. Washington, DC: APA Press.

Ericsson, K. A., Krampe, R. T. & Tesch-Roemer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, **100,** 363–406.

ERICSSON, K. A. & SIMON, H. A. (1993). *Protocol Analysis: Verbal Reports as Data* (Revised edn). Cambridge, MA: MIT Press.

GENTNER, D. & STEVENS, A. L. Eds. (1983). *Mental Models*. Hillsdale, NJ: Erlbaum.

HAYES, J. R. (1985). Three problems in teaching general skills. In S. CHIPMAN, J. W. SEGAL & R. GLASER, Eds. *Thinking and Learning Skills*, Vol. 2, p. 391–406. Hillsdale, NJ: Erlbaum.

HOFFMAN, R. R. (1991). Human factors psychology in the support of forecasting: The design of advanced meteorological workstations. *Weather and Forecasting*, **6,** 98–110.

HUTCHINS, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

IBA, W. & GERVASIO, M. (1999). Adapting to user preferences in crisis response. In *Proceedings of the 1999 Conference on Intelligent User Interfaces*. Redondo Beach, CA.

IBA, W., GERVASIO, M., LANGLEY, P. & SAGE, S. (1998). Evaluating computational assistance for crisis response. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. Madison, WI.

JOHNSON-LAIRD, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard Press.

KOSSLYN, S. M. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

KOSSLYN, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, **3,** 185–226.

KOSSLYN, S. M. (1994). *Image and Brain*. Cambridge, MA: MIT Press.

LEWANDOWSKY, S. & SPENCE, I. (1990). The perception of statistical graphs. *Sociological Methods and Research*, **18,** 200–242.

LOHSE, G. L. (1993). A cognitive model for understanding graphical perception. *Human Computer Interaction*, **8,** 353–388.

MCFARLANE, D. & TRAFTON, J. G. (in preparation). The give and take of communication requests in a complex visualization task.

MURPHY, A. H. & WINKLER, R. L. (1992). Approach verification of probability forecasts. *International Journal of Forecasting*, **7,** 435–455.

PINKER, S. (1990). A theory of graph comprehension. In R. FREEDLE, Ed. *Artificial Intelligence and the Future of Testing*, pp. 73–126. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

PLISKE, R. M., CRANDALL, B. & KLEIN, G. (in press). Competence in weather forecasting. In J. SHANTEAU, P. JOHNSON & K. SMITH, Eds. *Psychological Investigations of Competent Decision Making*. Cambridge, MA: Cambridge University Press.

SANDERSON, P. M., SCOTT, J. J. P., JOHNSTON, T., MAINZER, J., WATANABE, L. M. & JAMES, J. M. (1994). Macshapa and the enterprise of exploratory sequential data analysis (ESDA). *International Journal of Human-Computer Studies*, **41,** 633–681.

SHAH, P. & CARPENTER, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, **124,** 43–62.

SIMKIN, D. & HASTIE, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, **82,** 454–465.

SLOVIC, P., FISCHHOFF, B. & LICHTENSTEIN, S. (1980). Fact versus fears: understanding perceived risks. In R. SCHWING & W. A. ALBERS, Eds, *Societal Risk Assessment: How Safe is Safe Enough?* New York: Plenum Press.

SPARROW, J. A. (1989). Graphical displays in information systems: some data properties influencing the effectiveness of alternative forms. *Behavior and Information Technology*, **8,** 43–56.

TABACHNECK-SCHIJF, H. J. M., LEONARDO, A. M. & SIMON, H. A. (1997). CaMeRa: a computational model of multiple representations. *Cognitive Science*, **21,** 305–350.

TAYLOR, H. A. & TVERSKY, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, **31,** 261–292.

TRAFTON, J. G., *Spatial Transformations*, (in preparation).

TRICKETT, S. B., FU, W., SCHUNN, C. D. & TRAFTON, J. G. (2000). From dipsy-doodles to streaming motions: changes in representation in the analysis of visual scientific data. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*.

TRICKETT, S. B., TRAFTON, J. G. & SCHUNN, C. D. (2000). Blobs, dispy-doodles and other funky things: Framework anomalies in exploratory data analysis. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*.

TVERSKY, A. & KAHNEMAN, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, **185,** 1124–1131.

TVERSKY, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, **13,** 407–433.

TVERSKY, B. (1991). Spatial mental models. In *The Psychology of Learning and Motivation*, Vol. 27, pp. 109–145. New York: Academic Press.

TVERSKY, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. FRANK & I. CAMPARI, Eds. *Spatial Information Theory: A Theoretical Basis for GIS*, pp. 14–24. New York: Springer-Verlag.