

# Assessment of Expert Interaction with Multivariate Time Series ‘Big Data’

Susan Stevens Adams<sup>(✉)</sup>, Michael J. Haass, Laura E. Matzen, and Saskia King

Sandia National Laboratories, Albuquerque, USA  
{smsteve,mjhaass,lematze,shking}@sandia.gov

**Abstract.** ‘Big data’ is a phrase that has gained much traction recently. It has been defined as ‘a broad term for data sets so large or complex that traditional data processing applications are inadequate and there are challenges with analysis, searching and visualization’ [1]. Many domains struggle with providing experts accurate visualizations of massive data sets so that the experts can understand and make decisions about the data e.g., [2, 3, 4, 5].

Abductive reasoning is the process of forming a conclusion that best explains observed facts and this type of reasoning plays an important role in process and product engineering. Throughout a production lifecycle, engineers will test subsystems for critical functions and use the test results to diagnose and improve production processes.

This paper describes a value-driven evaluation study [7] for expert analyst interactions with big data for a complex visual abductive reasoning task. Participants were asked to perform different tasks using a new tool, while eye tracking data of their interactions with the tool was collected. The participants were also asked to give their feedback and assessments regarding the usability of the tool. The results showed that the interactive nature of the new tool allowed the participants to gain new insights into their data sets, and all participants indicated that they would begin using the tool in its current state.

**Keywords:** Big data · Eye tracking · Evaluation study · Knowledge elicitation

## 1 Introduction

‘Big data’ is a phrase that has gained much traction recently. It has been defined as ‘a broad term for data sets so large or complex that traditional data processing applications are inadequate and there are challenges with analysis, searching and visualization’ [1]. Many domains struggle with providing experts accurate visualizations of massive data sets so that the experts can understand and make decisions about the data e.g., [2, 3, 4, 5]. One such domain includes tasks requiring abductive reasoning. Abductive reasoning is the process of forming a conclusion that best explains observed facts. This type of reasoning plays an important role in fields such as scientific research, economics and medicine. A common example of abductive reasoning is medical diagnosis. Given a set of symptoms, a doctor determines a diagnosis that best explains the combination of symptoms. Abductive reasoning is also important in process and product engineering. Throughout a

production lifecycle, engineers will test subsystems for critical functions and use the test results to diagnose and improve production processes.

This paper describes an evaluation study for expert analyst interactions with big data for a complex visual abductive reasoning task. The experts in our study use multivariate time series data to diagnose device performance throughout a production lifecycle and are tasked with determining whether there are failures or anomalies in these complex data sets. The current tools available to these analysts do not fully support interaction with this type of data. As such, our research team developed a new tool with the goal of allowing these analysts to explore, interact and better understand this ‘big data’ associated with task and their decision making process.

## 2 Visualization Evaluation with Experts

Visualization of data and information is growing in popularity and results in impressive images and pictures. But how well do these visualizations allow experts to perform their tasks and solve the problems they need to solve? Previous work has suggested that reviews with experts are a valuable way to evaluate visualizations [6]. As such, we performed an evaluation of the Dial-A-Cluster (DAC) tool using the analyst experts, and followed the recommended steps laid out by [6]. For example, we choose a set of experts who were most familiar with the analysis, had the experts work independently on the tasks and took copious notes.

We also resonated with the idea of value-driven evaluations [7]. This work argues that the value of a visualization goes beyond the ability to simply answer questions about the data (which is common in typical usability studies) but should provide a broader, more holistic, “bigger picture” understanding of the data set. The author explains that the value of a visualization includes the total time required to answer a variety of questions about the data, a visualization’s ability to incite and discover insights or insightful questions about the data, a visualization’s ability to convey overall essence or take-away sense of the data and a visualization’s ability to generate confidence, knowledge and trust about the data [7]. Effective visualizations excel at presenting a set of heterogeneous data attributes in parallel, allowing a person to make inferences about the data set, allowing a person to gain a broad, total sense of a large data set beyond what can be gained from each individual data case, and allowing a person to learn and understand more than just the raw information contained within the data. The tool development and our expert evaluation study used a value-driven approach.

## 3 Analyst Task and Tool

The analysts in our study use complex, multivariate time series data to diagnose device performance throughout the production lifecycle. As we found in our previous work [8], these analysts made decisions by looking at trends across many different types of waveforms. The current analyst tool for analyzing these waveforms presents the waveforms one at a time, which does not allow the analyst to assess trends among the waveforms. As such, the team developed a new tool, termed Dial-a-Cluster (DAC), which allows

the analysts to visualize and inspect multiple waveforms at a time as well as view other important metadata.

The DAC tool [9] uses multidimensional scaling to provide a visualization of the data points depending on distance measures provided for each time series. The analyst can interactively adjust (dial) the relative influence of each time series to change the visualization (and resulting clusters). Additional computations are provided which optimize the visualization according to metadata of interest and rank time series measurements according to their influence on analyst selected clusters. The tool was created to allow the analyst to pull in different types of information and to visualize many different waveforms at once. See [9] for a complete description of the DAC tool. Figure 1 displays the DAC interface.



Fig. 1. Dial-a-Cluster interface

We performed a value-driven evaluation study of the DAC tool for complex, multi-variate time series ‘big data’ with the expert analysts. We asked the participants to perform different tasks using the tool, while collecting eye tracking data of their interactions with the tool. We also collected their feedback and assessments regarding the usability of the tool.

## 4 Evaluation Study

### 4.1 Participants

Seven participants at Sandia National Laboratories volunteered to participate in our study. Six of the participants in the study were classified as experts; that is, they diagnosed device performance using the multivariate time series data as part of their daily job. These experts had an average of 10 years’ experience performing this type of activity

(range 5–14). One participant was categorized as a novice, with less than one year of experience in this domain.

## 4.2 Procedure

The participants completed the study individually. In the work domain studied, access to experts was limited due to their senior roles spanning multiple engineering teams; therefore usability sessions had to be as brief as possible, while still being thorough enough to acquire all data relevant to the work and the expert's reasoning processes. Many of the same participants from our first study (see [8]) participated in this usability study. If the participant had not previously participated, he/she first read through and signed the study consent form and asked any questions he/she had about the study.

The experimenter then calibrated the FaceLAB 5 Standard System Eye Tracker with two miniature digital cameras and one infrared illumination pod. Eye tracking was collected during both training and the actual study trials; the experimenters anticipated that eye tracking data during the training session would shed light on how the participants learned how to use the tool and could improve future training on the tool. The participants then received training on the DAC tool. The experimenter explained the functions of the DAC tool buttons and panes using weather data and walked through a series of practice tasks using the different buttons and capabilities of the tool. The participant was encouraged to ask questions throughout the training session and to experiment with the tool and the weather data. Training lasted about 20 min.

After the training session was over, the participant completed a series of trials using the tool. Each trial contained multivariate time series data from multiple device tests. For each trial, the experts were presented with 100 tests, 11 different waveforms and 14 columns of metadata. This was in stark comparison to the existing tool which displayed fewer than 10 tests, presented one waveform at a time and did not have metadata as readily accessible. The participant was asked to classify the data as anomalous or normal. If the participant indicated that any of the tests was anomalous, he/she was asked to indicate the type of anomaly. Eye tracking data and response times were recorded while the subject worked with the DAC tool for each trial. There were a total of ten trials that participants could complete, although no participant completed more than five trials during the time allotted for the experiment session. The participants completed the trials in the same order.

After the determination was made (and response time was collected) for each trial, participants were encouraged to explain their thought process to better understand how they reached their decision and to understand how they interacted with the tool to make their determination. Also, any comments made by the participants during the study trial were noted by the experimenters.

At the end of the study, participants completed a questionnaire assessing their satisfaction with the tool. Participants were asked what they liked best and least about the tool, to provide suggestions for improving the usefulness of the tool and whether they would actually use the tool to complete their regular analysis tasks.

### 5 Analysis and Results

The amount of time it took the participants to complete each trial varied widely. Two participants completed five trials during the experimental session, one completed four trials, two completed three trials, and two completed only two trials. The novice participant completed only two trials and did not identify any anomalous data in either trial. The more experienced participants identified several anomalies, averaging between one and four anomalies per trial. This difference in performance highlighted the interplay between domain expertise and tool usability and was informative to the team in terms of future tool training.

In general, the participants completed the trials more quickly as the experiment progressed and they became more familiar with the tool. The duration of each trial for each participant is shown in Fig. 2. The expert participants are labeled E1-E6 and the novice participant is N1. Some trials were more difficult than others in terms of how readily the anomalous data “popped out” in the DAC tool. On Trial 3, a relatively easy trial, most participants found the answer in less than five minutes. On Trial 4, a more difficult trial, the average response time was closer to ten minutes.

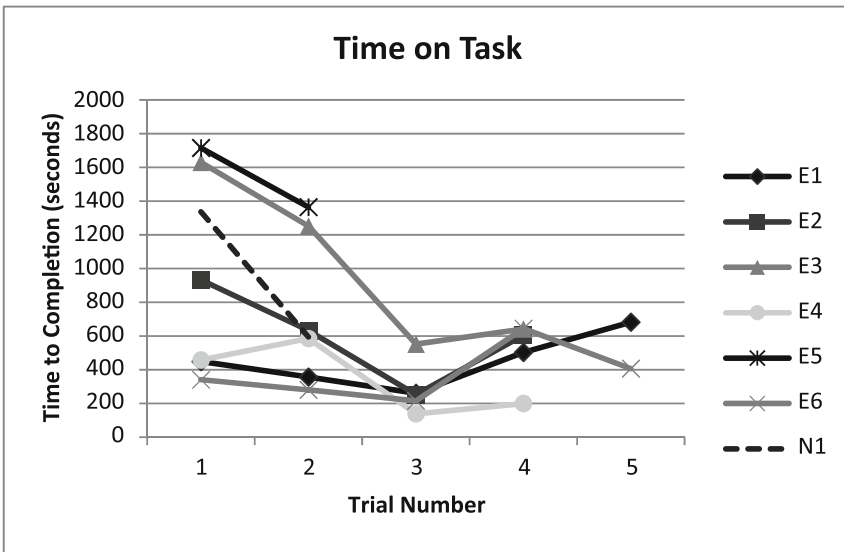
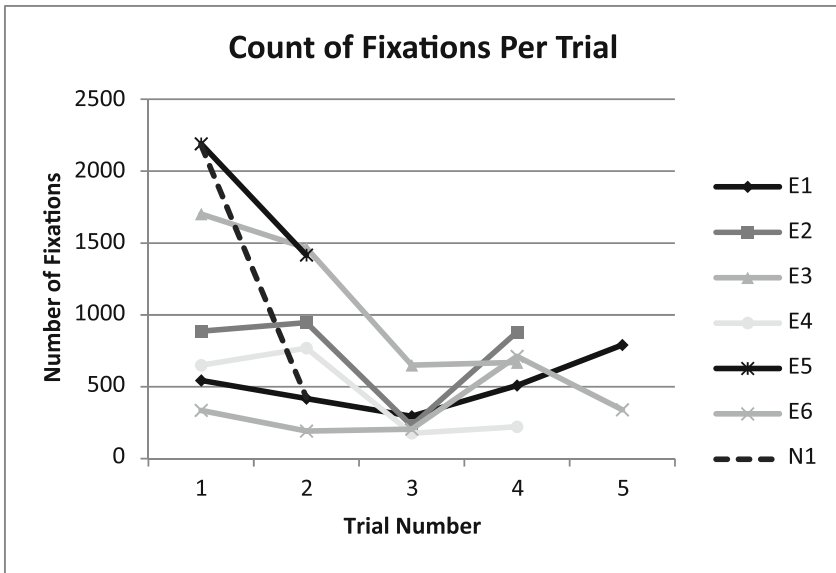


Fig. 2. Duration of each trial in seconds

Eye tracking data were analyzed using EyeWorks software, Eye Tracking Inc., Solana Beach, CA. The number of fixations per trial mirrored the time-on-task data, as shown in Fig. 3.



**Fig. 3.** Count of fixations on each trial for each participant

To analyze how the participants were using the DAC tool, the tool was divided into several regions of interest (ROIs), including the cluster pane, the graph pane, the slider pane, and the metadata pane. Figure 4 shows how the ROIs related to the DAC interface. The ROI analysis was conducted only for the first four trials, since few of the participants completed the fifth trial. On average, as the experiment progressed, participants spent more time viewing the cluster and graph panes and less time viewing the slider and metadata panes, as shown in Fig. 5. This pattern could indicate that as the participants became more comfortable with the tool, they spend more of their time focused on the data visualizations. Once the participants were familiar with the types of information in the slider and metadata panes, they would only need to consult that information when adjusting the way the data were displayed in the cluster panel or when investigating specific data points in the metadata.

We further subdivided the graph pane ROI to better understand how participants were using the data visualizations. The participants could display variables of their choice in the three graphs, or they could use a differencing tool that automatically set the graphs to show the variables that contributed most to the difference between two selections in the Cluster Pane. Early in the experiment, participants fixated on the top and middle graphs almost equally often. Surprisingly, as the experiment progressed, participant's average proportion of fixations increased for the middle graph, as shown in Fig. 6. This change could indicate that participants were developing strategies for how best to organize the information within the DAC tool in order to find the anomalies. A qualitative analysis of the participants' strategies, based on the observational notes taking during the sessions, indicated that most participants chose to display one key variable in the top graph. Their interactions with the cluster pane and the other



Fig. 4. Dial-a-Cluster interface divided into ROIs for the eye tracking analysis. A is the Cluster Pane, B is the Graph Pane, C is the Slider Pane, and D is the Metadata Pane.

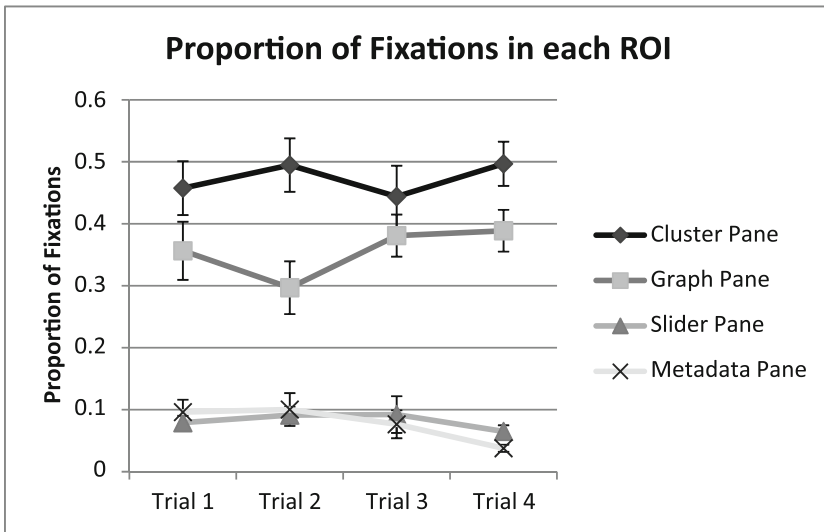
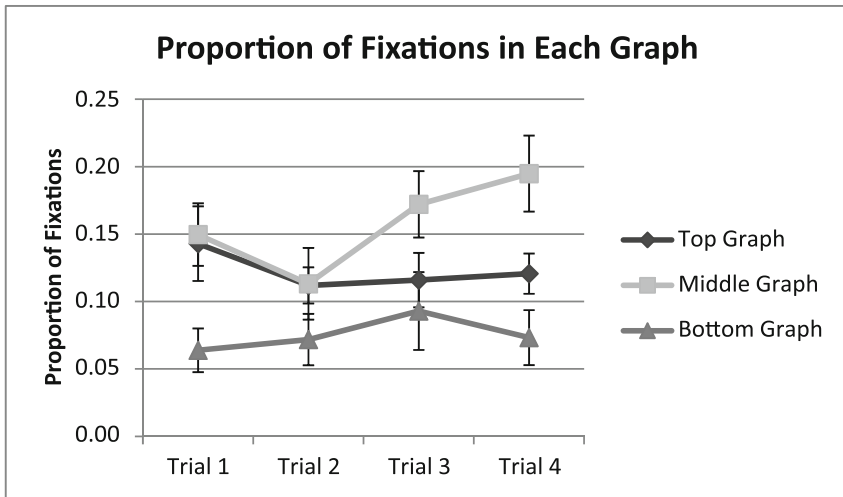


Fig. 5. Proportion of fixations in each ROI for each of the first four trials

graphs were largely focused on determining how other variables related to the variable of interest.

From a qualitative perspective, the participants responded positively to the tool. In their verbal and written assessments of the tool, they indicated that the interactivity and the linked visualizations were the key features that supported gaining insight into the data sets. Two of the participants (E5 and N1) revealed through their written feedback



**Fig. 6.** Proportion of fixations in each graph within the Graph Pane ROI

that they viewed the information in the cluster pane as a correlation, rather than two-dimensional projection of multi-dimensional data. This misinterpretation may have slowed their analyses, as these were also the only two participants who completed only two trials during the experiment session. Identifying a potential source of confusion for future users of the DAC tool was a valuable outcome of the study.

In summary, this evaluation showed that users were readily able to adopt a new tool for performing abductive reasoning with large, complex data sets. The DAC tool provided users with a new way to view types of data that they work with frequently, allowing them to assess larger data sets and to perform new types of analyses in order to identify trends and outliers in the data. The interactive nature of the tool allowed the users to gain new insights into their data sets, and all seven participants indicated that they would begin using the tool in its current state. The value-driven evaluation approach, using multiple types of analysis (behavioral, eye tracking, and qualitative), pointed toward trends in how participants used the tool as they became familiar with it. It also revealed some of the strategies that participants adopted, as well as potential pitfalls where a misunderstanding of the data visualizations could lead to confusion. This information will be used to further refine and improve the DAC tool.

**Acknowledgements.** Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2016-1724C. This work was funded by Laboratory Directed Research and Development (LDRD).



## References

1. Wikipedia, Accessed on Dec 2015
2. Marx, V.: Biology: the big challenges of big data. *Nature* **498**, 255–260 (2013)
3. Ayhan, S., Pesce, J., Comitz, P., Sweet, D., Bliesner, S., Gerberick, G.: Predictive analytics with aviation big data. In: *Integrated Communications, Navigation and Surveillance Conference (ICNS)*, pp. 1–13. IEEE (2013)
4. Fan, J., Han, F., Liu, H.: Challenges of big data analysis. *Nat. Sci. Rev.* **1**(2), 293–314 (2013)
5. Katal, A., Wazid, M., Goudar, R.H.: Big data: issues, challenges, tools and good practices. In: *IEEE Sixth International Conference on Contemporary Computing (IC3)*, pp. 404–409 (2013)
6. Tory, M., Möller, T.: Evaluating visualizations: do expert reviews work? *IEEE Comput. Graph. Appl.* **25**(5), 8–11 (2005)
7. Stasko, J.: Value-driven evaluation of visualizations. In: *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pp. 46–53. ACM, November 2014
8. Haass, M.J., Matzen, L.E., Stevens-Adams, S.M., Roach, A.R.: Methodology for knowledge elicitation in visual abductive reasoning tasks. In: Schmorow, D.D., Fidopiastis, C.M. (eds.) *AC 2015. LNCS*, vol. 9183, pp. 401–409. Springer, Heidelberg (2015)
9. Martin, S., Quach, T-T.: Interactive visualization of multivariate time series data. In: *HCI 2016 Proceedings* (2016)