Uncertainty-Awareness in Open Source Visualization Solutions

Christina Gillmann*
University of Kaiserslautern

Thomas Wischgoll †
Wright State University

Hans Hagen‡
University of Kaiserslautern

Abstract

The popularity of open source tools is constantly increasing, as they offer the possibility to quickly create and use visualizations of arbitrary data sources. As the positive effects of uncertainty communication to all kinds of visualizations were discussed over the past years in the academic world, this work examines the uncertainty-awareness of open source solutions. Through a categorization and classification of available tools, this paper identifies the problems in uncertainty-awareness of available open source solutions. To solve this problem, a new paradigm of data handling that extends raw datasets by its uncertainty is suggested.

Keywords: Uncertainty-awareness, Open Source, Data Management

1 Introduction

Uncertain data is defined as data that may contain values within some margin of error, which deviates the measured point from the intended or correct result. Studies have shown that these variations need to be addressed to any kind of data in order to help the analysts fulfill their tasks [4]. The main source of uncertainty is the data acquisition process itself, which always contains uncertainty, due to incomplete data, sampling size and mapping errors. Along a data processing pipeline this data is transformed using a model which itself introduces additional uncertainty. In total, the acquisition uncertainty and the model uncertainty accumulate and propagate along the data processing pipeline as Figure 2 shows.

As easy-to-access tools became increasing popular in the last years diverse open source visualization tools are available that help scientists and users from different applications create powerful visualizations of their data. In order to communicate the uncertainty of the used datasets, a proper uncertainty handling in open source tools is beneficial. To handle these problems, various researchers address uncertainty in their work. In fact, uncertainty-aware visualization methods can be found for all large groups of data types, covering information-, volume-, geospatial-, timeline- and graph visualizations (Figure 1).

To identify the level of uncertainty awareness of open source solutions, this study performs an analysis that investigates the uncertainty-awareness of available open source tools. For each of these tools, the level of uncertainty awareness is determined according to the work of Sacha et al. [7]. As a result, this study shows a potential for improvement in the open source visualization culture regarding uncertainty awareness. Therefore, the U-Dataset, that extends raw datasets by its uncertainty quantification, is presented. In case of data transformation tasks, the U-Dataset can be utilized to aggregate and propagate uncertainty through the data processing pipeline, by additionally transforming the uncertainty quantification of the U-Dataset through the data transformation model.

*c_gillma@cs.uni-kl.de
†thomas.wischgoll@wright.edu
‡hagen@cs.uni-kl.de

2 Uncertainty-awareness criteria

In the work of Sacha et al. [7], the role of uncertainty awareness in human decision making is discussed. The work suggested a list of requirements, that should be fulfilled in order to enable a proper communication of uncertainty in visualization. The requirements are not independent from each other, which means, that some have to be fulfilled to enable another requirement. Therefore, this paper sorts the listed requirements by their dependencies, resulting in the following stages of uncertainty-awareness:

- **(C1)** Quantify uncertainty in each component
- **(C2)** Visualize uncertainty information
- **(C3)** Enable interactive uncertainty exploration
- **(C4)** Propagate and aggregate uncertainty

In order to rate the uncertainty level of the found open source visualizations, they need to be divided by their ability. In this work two different kinds of open source solutions are discussed:

- **(A1)** Pure visualization
- **(A2)** Data-transformation and visualization

This separation is important, as there is no need to propagate and aggregate the uncertainty in the case of A1 open source tools, as they do not perform data transformation. In order to identify the level of uncertainty both open source types obtain a scale describing the amount of uncertainty-awareness. Therefore, this paper presents scales to measure the level of uncertainty-awareness of A1 and A2 open source visualization tools as shown in Figure 3, which reach from *no uncertainty awareness* to *full uncertainty awareness*. Due to the fact that A2 open source visualization tools do not need to propagate and aggregate uncertainty, their scale contains one step less than the scale for A2 open source visualization tools.

3 Open Source Solution Analysis

Open source visualization tools can be found for information-, volume-, geospatial-, timeline-, graph-, and hybrid visualization tasks. In this paper, only free solutions are considered. The determination of their level of uncertainty-awareness is based on the online portfolio of the solutions. This means, that the listed features and presented examples available on the tools' websites where examined for the classification as well as the estimation of the level of uncertainty-awareness according to the defined scales. The reviewed websites can be found in Table 5 and Table 5. Resulting from that, 28 open source tools could be identified as type A1 visualization tools, whereas 6 tools count to the A2 visualization group.

The amount of uncertainty-awareness of group A1 visualization tools can be found in Figure 5. It can be observed, that information visualization is the best covered area in terms of uncertainty awareness. In contrast to that, open source tools in the area of volume-, geospatial, timeline- and graph visualization, do not cover uncertainty awareness in their scope. Instead, most of the hybrid open source visualization tools offer a high level of uncertainty awareness.

Considering the group of A2 open source visualizations, the field of volume visualization still provides no options for uncertainty visualization. In contrast to that, there exist a tool which is able to perform partially geospatial visualization under uncertainty. The group of hybrid visualization tools offers the largest amount of uncertainty awareness, as can be seen in Table 5. Although this is a good starting point, a complete coverage of uncertainty awareness is not provided.

It can be observed, that type A2 tools and especially hybrid visualization tools are making the most use of uncertainty visualization. A reason could be, that hybrid visualization tools have more possibilities to visualize data and it is therefore easier to incorporate uncertainty awareness in their visualizations. Although all open source tools offer a large variety of useful visualization possibilities and are therefore wide spread, uncertainty awareness is an ability, that can be further improved upon.

4 Future Directions

Although the previous section showed, that uncertainty awareness of open source tools holds an potential for improvement, this does not mean, that they have to be redesigned in their entirety. On the contrary, they offer a good possibility to help scientists visualize their data. Still, uncertainty awareness is a desired feature for open source visualization tools, that can be incorporated in an existing open source tool using the following concepts.

4.1 U-Dataset

The U-Dataset describes the extension of the raw dataset by expressing its uncertainty. This means, that each datapoint needs to hold an additional value describing its amount of uncertainty. For example an image (as shown in Figure 6 needs to provide a quantification of uncertainty for each of its pixels.

Therefore, two data acquisition pathways in open source tools are possible, as Figure 4 shows: The first version (orange pathway) is related to the raw data acquisition process. Therefore, the analyst can pick one of the various available uncertainty measures and apply them to the raw dataset in order to generate the U-Dataset. To allow a backwards compatibility in open source solutions to datasets that contain no uncertainty quantification, the open source tool should be able to add a basic statistical uncertainty quantification to generate a U-Dataset from a classic dataset (shown in the red pathway).

As shown in the examples of Figure 1 each type of data and its uncertainty can be visualized. Open source tools should incorporate these methods in their scope of available visualizations to allow a proper visualization of U-Datasets.

4.2 U-Pipeline

The open source solution analysis showed, that type A2 tools in general lack the ability to propagate and aggregate uncertainty along the data processing pipeline. In order to solve this problem, the U-Dataset can be used. In contrast to the classic data processing pipeline, this paper suggest to use the data transformation pipeline to additionally transform and aggregate the uncertainty of a U-Dataset. Therefore, it is not necessary to create a new data processing pipeline. Instead, the existing pipeline of an arbitrary open source tool can be used without be re-engineered. As Figure 5 shows, the data generation and transformation itself is still performed in the same manner with the only difference being, that the uncertainty of a dataset is manipulated by the data transformation model as well. The uncertainty Δx of each data point x in a U-Dataset which is transformed through a model function f can be transformed by [1]:

$$f(x + \Delta x) = \frac{\Delta f}{\Delta x} * \Delta x \tag{1}$$

The equation is widely used in physics to propagate errors and uncertainty values and can be applied to all datatypes considered in this paper. With this equation it is possible to extend an arbitrary data processing pipeline in an open source visualization tool to generate the uncertainty in each U-Dataset, based on the given model f. Therefore, the uncertainty of each intermediate step can be aggregated and propagated through a processing pipeline. To visualize a U-Dataset at an arbitrary point in the data transformation pipeline, the open source tool can provide either new visualization methodologies to visualize a dataset and its uncertainty simultaneously or simply use its provided visualizations to visualize the dataset and its uncertainty uncoupled.

Figure 6 shows an example for the transformation of an U-Dataset and its visualization, in an uncoupled manner. Therefore, an input image (upper left) was examined by an uncertainty measure (lower left). The input image is transformed by applying a scale (1.5) and shift (30.0) operation (upper right). According to the formula given in Equation 1, the uncertainty can be transformed as well. Resulting from that, the U-Dataset can be transformed but the given data transformation model.

5 Conclusion

In summary, this paper shows that the uncertainty awareness of open source solutions hold a potential for improvement. Therefore, this paper suggests a new paradigm of data handling, that extends raw datasets by its uncertainty and shows how these uncertainty can be used to extend the processing pipeline to achieve a propagation and aggregation of uncertainty.

As a future work an open source tool to generate U-Datasets using user selected uncertainty measures based on raw datasets is planned.

References

[1] R. J. Barlow. Statistics: A guide to the use of statistical methods in the physical sciences. *Wiley*, 1, Feb. 2013.

[2] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. *Overview and State-of-the-Art of Uncertainty Visualization*. Springer London, 2014.

[3] S. Djurcilov, K. Kim, P. F. J. Lermusiaux, and A. Pang. Volume rendering data with uncertainty information. In *Data Visualization 2001: Proceedings of the Joint Eurographics - IEEE TCVG Symposium on Visualization*, pages 243–252, 2001.

[4] G. B. Folland and A. Sitaram. The uncertainty principle: A mathematical survey. *Journal of Fourier Analysis and Applications*, 3(3):207–238, 1997.

[5] H. Guo, J. Huang, and D. H. Laidlaw. Representing uncertainty in graph edges: An evaluation of paired visual variables. 21(10):1173–1186, 2015.

[6] A. M. Maceachren, A. Robinson, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler. Visualizing geospatial information uncertainty: What we know and what we need to know. cartography and geographic. *Information Science*, 32:160, 2005.

[7] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of the Visual Analytics Science and Technology)*, 22(01):240–249, Jan. 2016.

[8] T. Zuk, S. Carpendale, and W. D. Glanzman. Visualizing temporal uncertainty in 3d virtual reconstructions. In *Proceedings of the 6th International Conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, pages 99–106. Eurographics Association, 2005.

Acknowledgements

This research was funded by the ”German Research Foundation” (DFG) within the IRTG 2057 ”Physical Modeling for Virtual Manufacturing Systems and Processes”.

Datatype	Tool Name	Website	Ability
Information Visualization	Charts.js	http://www.chartjs.org/	
	Raw	https://github.com/densitydesign/raw/	
	Dygraph	http://dygraphs.com/	
	Visualize free	https://visualizefree.com/	
	Fusion Chart	http://www.fusioncharts.com/	
	jqPlot	http://www.jqplot.com/	
	jqGraph	http://jpgraph.net/	
	OpenOffice	https://www.openoffice.org/	
	Processing	https://processing.org/	
	iCharts	http://icharts.net/	
	Flot	http://www.flotcharts.org/	
Volume Visualization	ImageVis	http://www.sci.utah.edu/software/imagevis3d.html	
	VolView	http://www.kitware.com/opensource/volview.html	
Geospatial Visualization	Leaflet	http://leafletjs.com/	
	Polymaps	http://polymaps.org/	
	Open Layers 3	http://openlayers.org/	
	Kartograph	http://www.open-bigdata.com/about/	
Timeline Visualization	Timeline	http://www.readwritethink.org/	
	dipity	http://www.dipity.com/	
	Timeline.js	https://timeline.knightlab.com/	
Graph Visualization	Gephi	https://gephi.org/	
	GraphVis	http://www.graphviz.org/	
	Cytoscape	http://www.cytoscape.org/	
	Neo4j	https://neo4j.com/	
Hybrid Visualization	Exhibit	http://www.simile-widgets.org/exhibit/	
	JS InfoVis TK	http://philogb.github.io/	
	High Charts	http://www.highcharts.com/	
	Google Charts	https://developers.google.com/chart/	

Table 1: A1 open source visualization tools and their uncertainty awareness according to the scale of Figure 3.

Datatype	Tool Name	Website	Ability
Volume Visualization	Voreen	”http://www.voreen.org/”	
Geospatial Visualization	Instant Atlas	http://www.instantatlas.com/de/	
	Modest Maps	http://modestmaps.com/	
Hybrid Visualization	Wolfram Alpha	https://www.wolframalpha.com/	
	D3s	https://d3js.org/	
	ParaView	http://www.paraview.org/	

Table 2: A2 open source Visualization tools and their uncertainty awareness according to the scale of Figure 3.

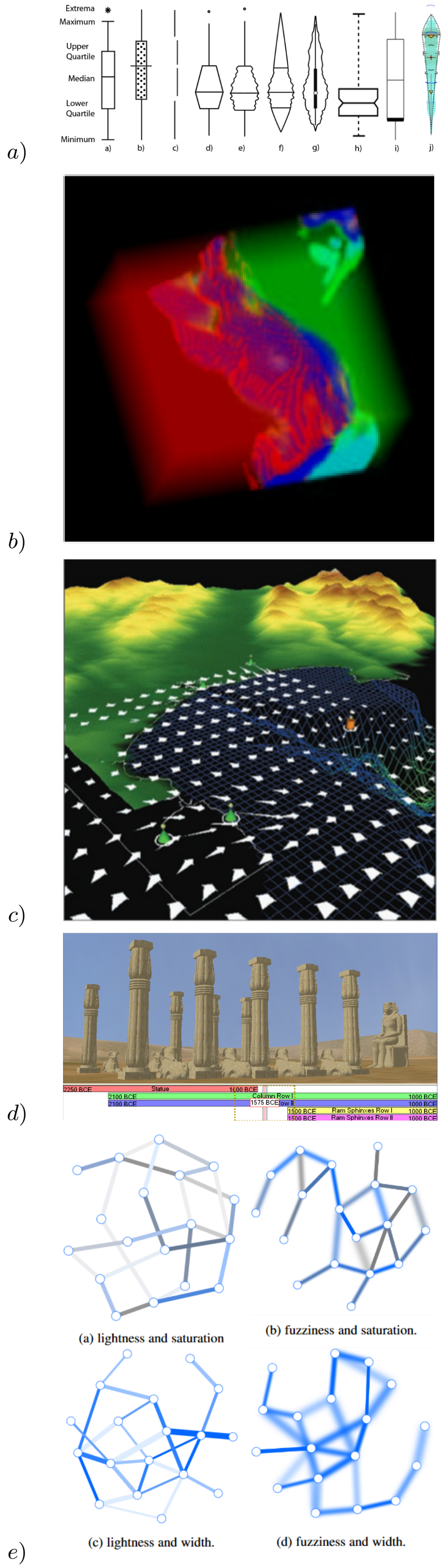

Figure 1: Example uncertainty visualization for different types of data. a) Boxplot (information visualization) [5]. b)Volume rendering with different alpha blending [8]. c) Geopatial visualization with texture overlay [6]. d) Timeline visualization with confidence interval [3]. e) Graph visualization with different uncertainty encoding strategies [2].

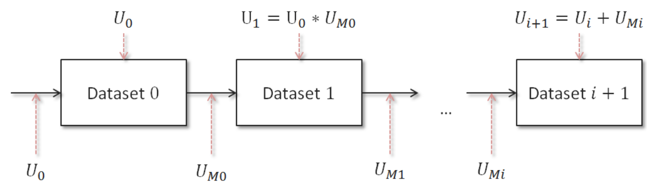

Figure 2: Uncertainty and its sources: Each dataset contains uncertainty through the data acquisition process which is accumulated along the data processing pipeline with the model's uncertainty.

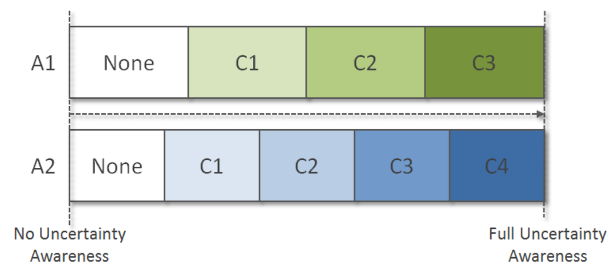

Figure 3: Scales for the level of uncertainty awareness of A1 open source solutions (top) and A1 solutions (bottom). The darker the color, the higher is the uncertainty-awareness of the tool.

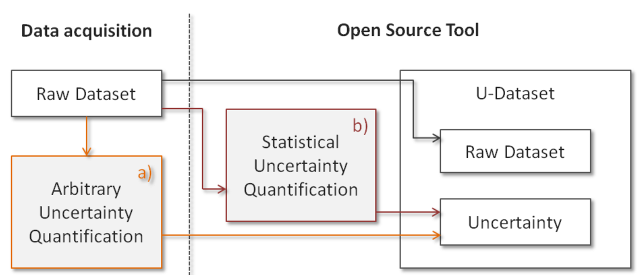

Figure 4: U-Dataset and its generation. A U-Dataset holds a raw dataset as well as its uncertainty. The generation of the uncertainty quantification is possible via two pathways: Orange) Uncertainty estimation through the data analyst before using the open source tool. Red) Backup solution of the open source tool in case the data analyst does not provide an uncertainty quantification.

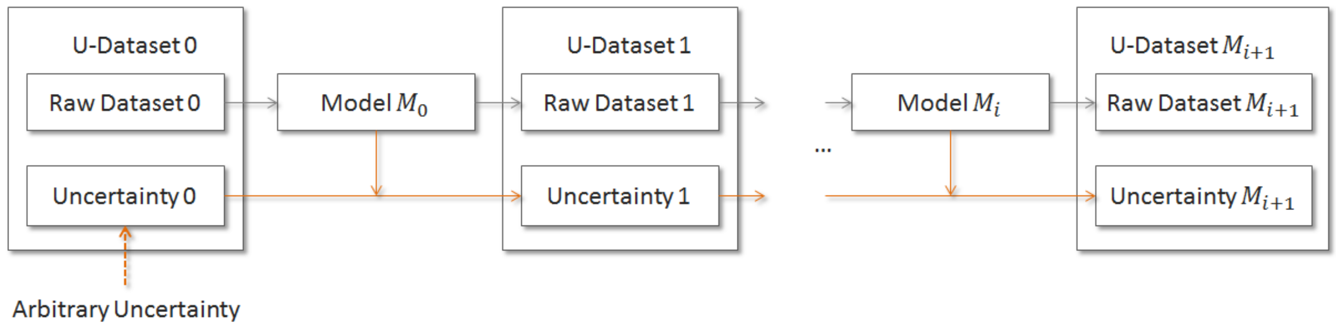

Figure 5: U-Pipeline: The U-Pipeline extends the existing data processing pipeline by inserting a second path aggregating and propagating the uncertainty of the U-Datasets.

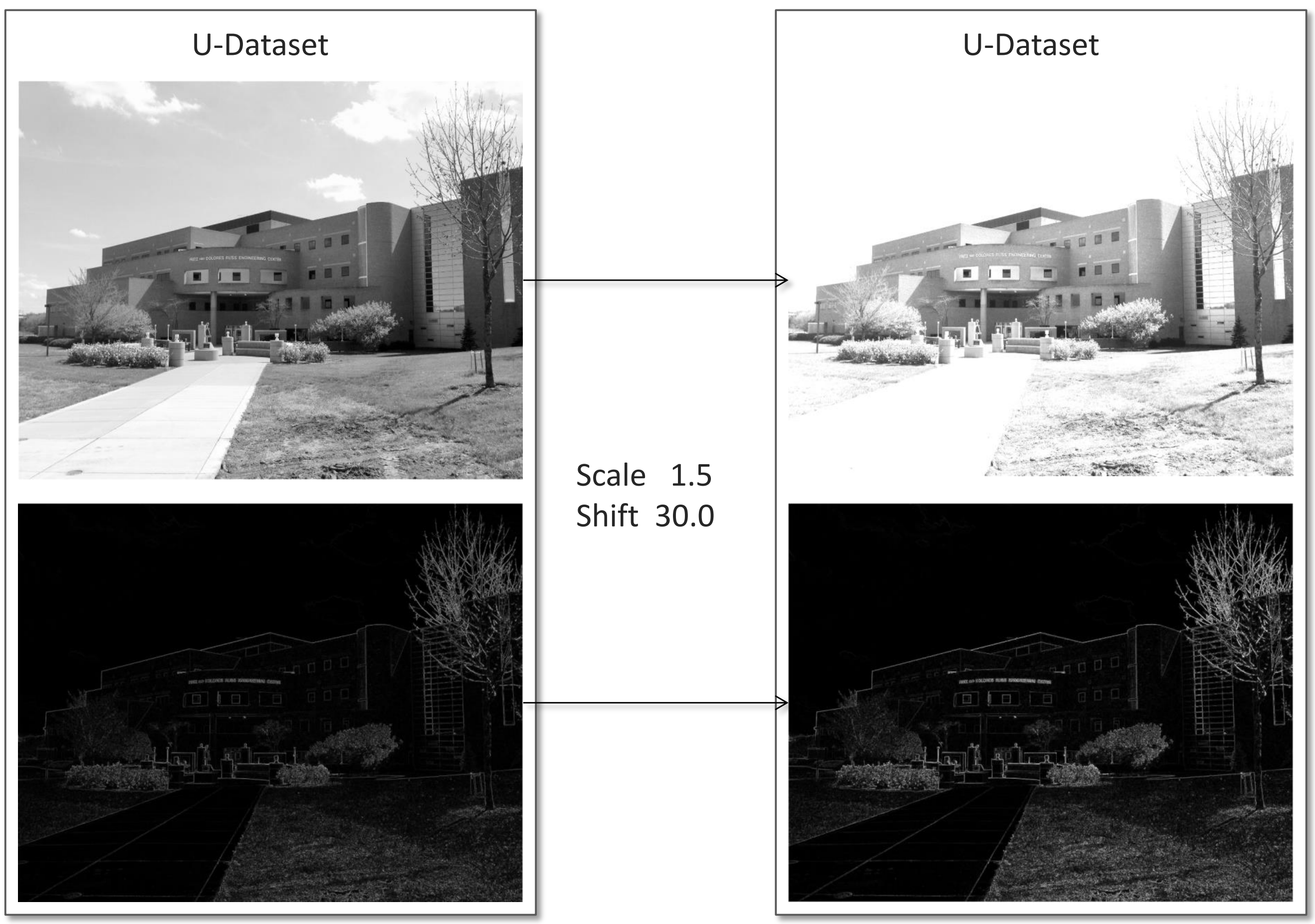

Figure 6: Example of a real world U-Dataset and its transformation through the U-Pipeline. Left: Original dataset (photography) and its uncertainty. Right: Dataset transformed through a scale(1.5) and shift(30.0). The uncertainty of the dataset is also transformed through the transformation model.