TOWARDS THE VISUALIZATION OF MULTI-DIMENSIONAL STOCHASTIC DISTRIBUTION DATA

Kristin Potter

Jens Krüger

Christopher Johnson

Scientific Computing and Imaging Institute
University of Utah

ABSTRACT

Uncertainty information is an important characteristic associated with much of the data scientists encounter. While such uncertainty information is often available, incorporating uncertainty into visualization techniques has proved challenging. This paper presents novel visualization approaches for a class of uncertainty data that is generated from the sensitivity analysis of electrical conductivity within a model of bioelectric fields from the heart. The data can be characterized as a set of probability density functions (PDFs) defined across a triangular mesh; however, we are also interested in the relationship between input and output parameters of the sensitivity analysis. This increases the complexity of the data set and motivates a visualization approach that provides for exploration of the data set.

KEYWORDS

Uncertainty, sensitivity analysis, probability density function.

1. INTRODUCTION

The estimation and visualization of uncertainty information is an important research problem in both simulation [DeVolder 2002] and visualization [Johnson 2004]. Uncertainty is a term used to describe the error, confidence, and variation of a dataset in order to allow a scientist to understand the accuracy not only of the data but also of the processes used to explore the data. One such technique, sensitivity analysis, helps the scientist to understand the effects of perturbing parameters of a function. Small perturbations of the input parameters that create large perturbations in the output results can indicate areas of the function that are highly dependent on the input parameters and may be interpreted as unstable or possibly wrong. Sensitivity analysis techniques can be used not only to explore the mathematical models used to generate uncertainty data but also to better understand the effects of input parameters of visualization techniques. While this paper focuses on uncertainty data generated from the sensitivity analysis of a mathematical model reconstructing a biological experiment, the methods presented are applicable to many other datasets as well.

The National Institute of Standards and Technology [Taylor et al., 1994] defines the uncertainty of a measured result to be the standard deviation of the collection of data samples approximating the measurand. A measurand is a particular quantity to be measured, such as the temperature at a certain location and time of day. This measurand is estimated by a series of measurements; the true value of the measurand is expected to lie within the set of measurements. Uncertainty can then be derived as the standard deviation of the set of measurements, and this uncertainty value can be used to describe the confidence that the actual value of the measurand lies within a stated range. Uncertainty information must accompany a set of data samples for the set to be considered a complete result. Thus, uncertainty is becoming a more important value as the quantity of generated data grows with the increase in the sophistication of scientific simulation and measurement devices. Visualizing these data sets can become complicated, and incorporating the additional data parameter of uncertainty into visualizations is challenging both because adding uncertainty increases the complexity and information content of the visualization and because there is no agreed upon way to visually represent uncertainty for three-dimensional (and higher) problems.

Thus far, uncertainty visualization approaches represent uncertainty as a scalar, vector, or tensor value. This paper explores the underlying probability distribution function that was used to create the individual measures of uncertainty. Mean and standard deviation (or variance) are easy and robust ways to quickly reduce the amount of data into an understandable quantity. If the underlying distribution is Gaussian (normal), the mean and standard deviation completely describe the underlying distribution. However, not all distributions are exactly normal, and the deviation from normal is often the more interesting information.

The data used in this paper can be looked at in several ways, of which reducing it to mean and standard deviation is the simplest. However, since there is a functional relationship that such an approach ignores, we explore a variety of techniques with the goal of learning about the relationship between input parameters and output results.

The remainder of this paper is structured as follows: In Section 2 we describe more fully the data we are using and present the traditional mean and standard deviation methods for visualizing the data. In Section 3 we examine previous work in the field. In the following Section 4 we explain our approach to visualizing the data and conclude with discussion and future work.

2. APPLICATION DATA

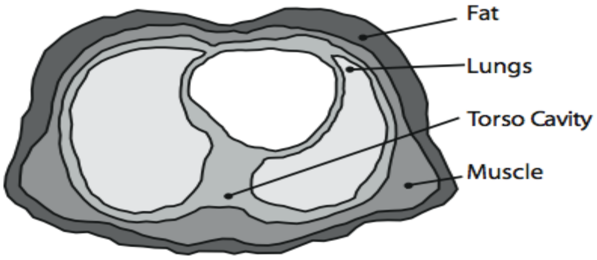

Figure 1. The classified human torso. *Image courtesy Sarah Geneser.*

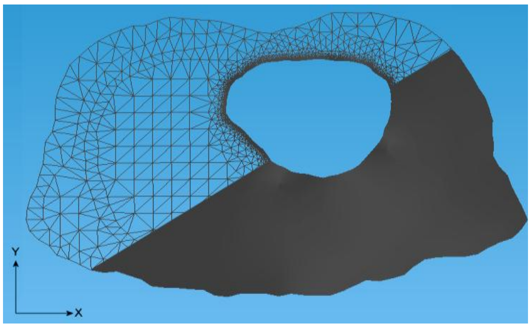

Figure 2. The torso mesh space (x,y)

The data used in this paper comes from the sensitivity analysis electrical conductivity within computational models of bioelectric fields of the heart. Such an approach quantifies the sensitivity of the electrocardiographic forward problem by creating a mathematical model to reconstruct a biological experiment in which the voltages on the human torso are estimated based on the input electrical conductivities [Geneser et al. 2005]. The simulation stochastically varies the input conductivities of different tissues such as fat, lungs, or muscle and examines the resulting changes in potential across the torso. Figure 1(a) depicts the torso classified by tissue type, and (b) shows the triangular domain on which the simulation is computed. The dataset visualized in this paper stochastically varies the lung conductivity ±50% from the reference lung conductivity. For each individual variation of input conductivity, κ, a set of heart voltages is generated at each mesh point of the torso. This pairing of input conductivity and the set of output voltages is called a realization; our dataset consists of 10,000 such realizations. The most straightforward way to think about this data is as a volume of slices (realizations) with axes (x, y, κ), called the κ-volume.

The dataset resulting from numerous runs of the simulation describes the sensitivity of voltages across the torso to small changes in the input conductivity, thus describing the uncertainty associated with the mean voltage at each mesh point. The main challenge in visualizing this data is in its large size; not only is it hard to visually understand 10,000 2D slices, it is unclear how to present this information in a way that leads to a proper interpretation of the sensitivity. A simple approach is to summarize the data, thus reducing the dimensionality of the visualization problem. This can be seen in Figure 3, in which mean (a) and standard deviation (b) are displayed through a color-mapped flood contour. This presentation is straightforward, however it discards a large amount of data, and decouples the input conductivity parameters from the output voltages, a relationship this works strives to preserve.

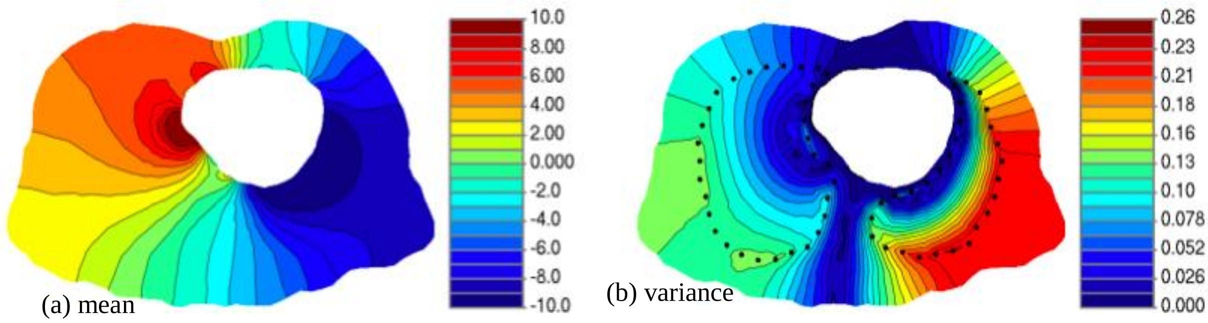

Figure 3. The (a) mean and (b) variance of stochastic variation of lung conductivity (±50% from the reference lung conductivity). *Images courtesy of Sarah Geneser.*

3. PREVIOUS WORK

The data used in this paper can be examined in numerous ways including by looking solely at summary quantities, either as a collection of probability density functions (PDFs) at each mesh point or as a 3D volume of 2D slices in κ-space. Visualization techniques exist to address the first two interpretations; visualizing the third approach is not straightforward.

Much uncertainty visualization has focused on adding scalar or vector quantities to already existing visualization techniques. These values are often available alongside a dataset as tables, graphs, and charts [Spear 1952, Tukey 1977], but they must be incorporated into existing visualization techniques if we are to portray the data completely and accurately [Wittenbrink 1995]. Pang, Wittenbrink, and Lohda, 1997, have created a taxonomy of visualization approaches based on the type of uncertainty data. Common approaches include color mapping [Pang et al. 1997], glyphs [Wittenbrink et al. 1997, Grigoryan and Rheingans 2004], iso-surfacing [Jospeh et al. 1999, Rhodes et al. 2003], volume rendering [Djurcilov et al. 2002], and annotation [Cedilnik and Rheingans 2000].

While the data used herein can be summarized through mean and standard deviation, a true representation of the data expresses the underlying distribution. Potter et al., 2007, use visual summaries to describe individual distributions, but this approach is problematic for higher dimensional distributions as well as for large collections of distributions with spatial locality resulting from a high amount of visual clutter.

Other methods filter down the amount of data and provide a user interface for exploration of the dataset. Ehlschlaeger et al., 1997, presents a method to smoothly animate between realizations of surface elevation. Bordoloi et al., 2004, use clustering techniques to reduce the amount of data while providing ways to find features of the datasets such as outliers. Kao, 2002, uses a slicing approach to reduce the dimensionality by projecting distributions onto cutting planes. Streamlines and volume rendering have been used by Luo et al., 2003, to show distributions mapped over 2 or 3 dimensions. Finally, case studies of EOS Satellite and Lidar data have been performed by Kao, 2001 and 2004.

This small body of work on visualizing probability density functions is an interesting starting point for this research. However, the aim of this paper is to avoid the data reduction employed in these techniques, and instead preserve the correspondence between input and output parameters. This additional directive requires a visualization paradigm that can appropriately handle large datasets with high dimensionality.

4. OUR APPROACH

The main challenges to visualizing this dataset are its large size and data the meaning of which is not intuitive. While data reduction is not the goal of this work, getting an initial understanding of the data through summarization is beneficial. Figure 3 summarizes the results of the simulation, presenting them in two images. Figure 4 combines the summaries into a single visualization by displaying the mean (a) as a height field using the (x.y) mesh space and voltages as the z axis, and color mapping standard deviation (b) using a rainbow color map. The simplicity of this approach is the strongest argument for displaying the data in this way. However, the reduction of data allows for very little further exploration, and this visualization still does not capture the dependencies between the input conductivities and the output voltages. To overcome these limitations, we have to look at more than just these two basic stochastic properties while still presenting these quantities in a clear fashion.

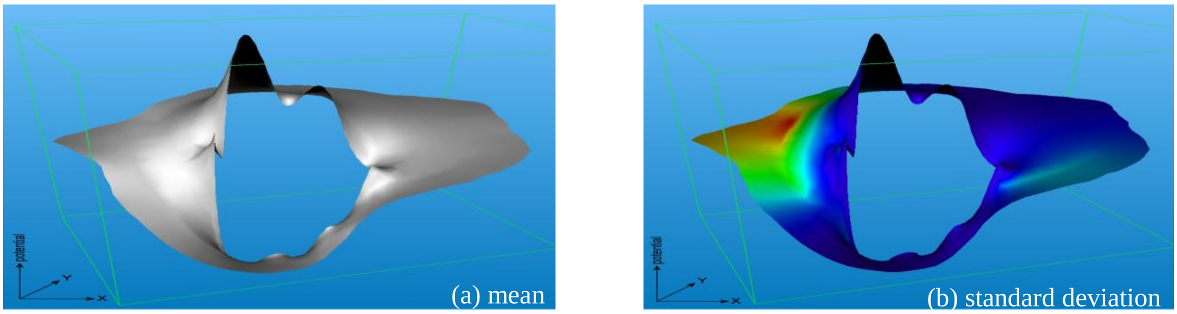

Figure 4. Combining mean and standard deviation into one visualization. (a) Mean is represented as height. (b) Standard deviation is color mapped onto the mean height field with high deviation in red, low in blue.

Rather than simply summarizing the data, we need to consider the entire volume of data, which stores all realizations of the voltages in 2D space for every conductivity, κ. To create the κ -volume, we stack the 2D datasets for each realization into a 3D volume with z-axis associated as κ (in contrast to the (x, y, voltages) axes for the mean height field). For this particular dataset, the κ-volume would result in a volume with 10,000 slices in the κ direction. Since this would lead to a very large dataset, we have chosen to subsample the volume to 512 slices, computing the mean of roughly 20 voltages for every voxel. This scalar volume can be visualized using direct volume rendering or iso-surface raycasting. Figure 5(a) shows the κ-volume superimposed onto the mean/variance height field. While direct volume rendering allows the user to see the entire volume, this technique is problematic for this data because the voltages occlude each other, or – if the transparency is reduced – are indistinguishable. Therefore, we have chosen to focus on iso-surface raycasting. Figure 5 (b-c) shows two different iso-surfaces of the potential. For our dataset, the structure of the iso-surface rather than the actual iso-value is of interest. An iso-surface that falls straight down through κ-space (red box in Figure 5c) indicates that the potential at this point in 2D space does not vary when changing; thus, it is considered independent at that point. A bending iso-surface (yellow box in Figure 5c) indicates areas of high dependence on the input conductivity. As can be seen in comparison with Figures 3 and 4, the bend in the iso-surface corresponds to an area of high variance; however, in our visualization, the actual reason for this high variance becomes clear. We can also see that even in the region with the highest variance only small κ's (lowest values on the κ-axis) result in potential changes. This relationship becomes visible only in our new visualization system.

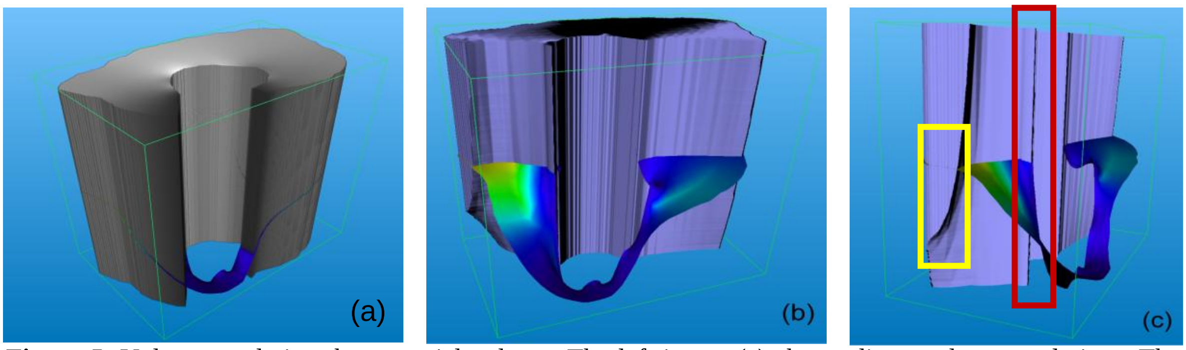

Figure 5. Volume rendering the potential volume. The left image (a) shows direct volume rendering. The other two images (b-c) show iso-surfaces of the voltages in κ-volume. Bends in the iso-surfaces indicate high dependence on the input conductivity (κ) (highlighted by the yellow bow), while straight iso-surfaces (red box) indicate stochastic independence.

While the iso-surface rendering qualitatively depicts stochastic dependency and presents a quick method to get a global overview of potential distribution, a more quantitative measurement is also desirable. To this end, we consider the gradient volume of the potential. In this vector-valued dataset, we perform streamline tracing to visualize the potential gradients. The physical interpretation of these streamlines is as follows: If a streamline's course is primarily horizontal, meaning that its tangents (i.e. the gradient in the potential field) mainly point in x and y direction but not into κ-direction, the points along the streamline have a potential that is locally independent of the realizations of the input conductivities. This in turn means that a streamline pointing up- or downwards indicates a high dependency on κ. In fact, the length of the streamline – influenced by the gradient-magnitude – is a quantitative measure of these dependencies (see Figures 6 and 7). To further emphasize the gradient-magnitude, we also seed particles into the gradient volume. These particles not only convey the direction of the gradient but also allow gradient magnitude to be derived easily form their speed.

5. CONCLUSION AND FUTURE WORK

In this paper, we present a novel method to approach multi-dimensional stochastic distribution data. Our technique allows for the exploration of these datasets, not only presenting a very limited overview by displaying the mean and variance as previous visualizations do, but also allowing for both a global qualitative as well as a local quantitative view. As our approach makes use of an existing GPU visualization system [Krüger 2005], the exploration of these datasets can be performed interactively on commodity PC hardware. For a demonstration, please see the video at ftp://datex.sci.utah.edu/Videos/CGV08.

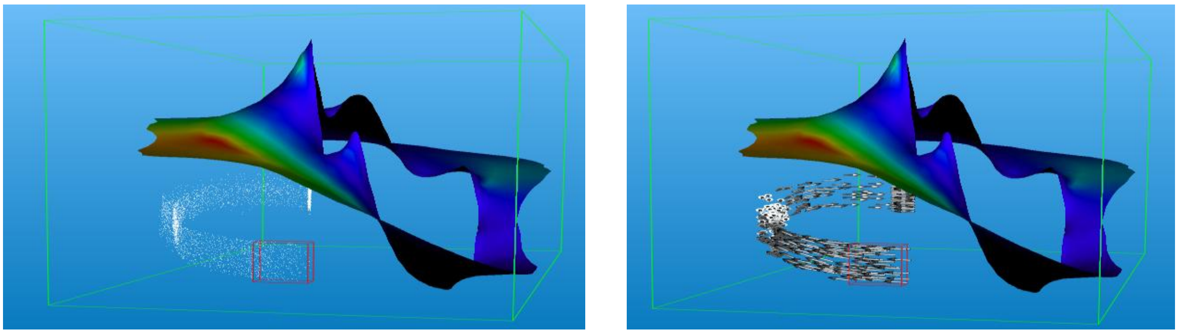

Figure 7. Particle tracing, the left image shows simple points as particles while in the right image arrow glyphs are rendered. In still images, arrows glyphs add more information to the image.

As the sources of uncertainty data become more diverse, understanding the meaning of the generated images becomes more difficult. While many of the traditional visualization techniques will still be applicable, new and better techniques for understanding complicated data sets will also be required. It is our hope that this work serves as a starting point for the visualization of complex data and uncertainty.

Further work in this area includes using data sets with higher-order input parameters (i.e. vector, tensor, etc.), and data distributions with higher-dimensional spatial localities. In addition, information visualization techniques can be used to increase the ways a user can explore the data, both globally and locally. Such a system would provide ways to answer known questions as well as to pose and test new hypotheses.

REFERENCES

Bordoloi, U. et al, 2004. Visualization Techniques for Spatial Probability Density Function Data. *In Data Science Journal*, Vol. 3, pp. 153-162.

Cedilnik, A. and Rheingans, P., 2000. Procedural Annotation of Uncertain Information. *In Proceedings of the IEEE Conference on Visualization.* Salt Lake City, UT, pp. 77-83.

Djurcilov, S. et al, 2002. Visualizing Scalar Volumetric Data with Uncertainty. *In Computers and Graphics*, Vol 26, pp. 239-248.

DeVolder, B., Glimm, J., Grove, J., Kang, Y., Lee, Y., Ye, K. Uncertainty Quantification for Multiscale Simulations. *Journal of Fluids Engineering*, March 2002, Volume 124, Issue 1, pp. 29-41.

Ehlschlaeger, C. et al., 1997. Visualizing Spatial Data Uncertainty Using Animation. *In Computers in GeoSciences*, Vol 23, No. 4, pp. 387-395.

Grigoryan, G. and Rheingans, P., 2004. Point-Based Probabilistic Surfaces to Show Surface Uncertainty. *In IEEE Transactions on Visualization and Computer Graphics*, Vol. 10, No. 5, pp. 546-573.

Geneser, S. et al., 2005. Sensitivity Analysis of Cardiac Electrophysiological Models Using Polynomial Chaos. *In Proceedings of the IEEE Engineering in Medicine and Biology Annual Conference*, Shanghai, China, pp. 4042-404.

Geneser S. et al., 2005. 2D Stochastic Finite Element Study of the Influence of Organ Conductivity in ECG Forward Modeling. *In International Journal of Bioelectromagnetism,* Vol. 7, No. 1, pp. 321-324.

Interrante, V., 2000. Harnessing Natural Textures for Multivariate Visualization. *In IEEE Computer Graphics and Applications*, Vol. 20, No. 6, pp. 6-11.

Johnson, C., 2004. Top Scientific Visualization Research Problems. *In IEEE Computer Graphics and Applications*, Vol. 24, No. 4, pp.13-17.

Johnson, C. and Sanderson, A., 2003. A Next Step: Visualizing Errors and Uncertainty. *In IEEE Computer Graphics and Applications,* Vol. 23, No. 5, pp. 6-10.

Johnson, C., 2004. Top Scientific Visualization Research Problems. *In IEEE Computer Graphics and Applications,* Vol. 24, No. 4, pp. 13-17.

Jospeh, A. et al, 1999. UISURF: Visualizing Uncertainty in Iso-surfaces. *In Proceedings of Computer Graphics and Imaging.* Palm Springs, CA, pp. 184-191.

Kao, D et al, 2001. Visualizing 2D Probability Distributions from EOS Satellite Image-Derived Data Sets: A Case Study. *In Proceedings IEEE Visualization*. San Diego, CA.

Kao, D. et al, 2002, Visualizing Spatially Varying Distribution Data. *In Proceedings of the 6th International Conference on Information Visualization.* London, England, pp. 219-225.

Kao, D. et al, 2004. Visualizing Distributions from Multi-Return Lidar Data to Understand Forest Structure. *In Geoinformatics*, Gavle Sweden.

Krüger J., Kipfer P., Kondratieva P., Westermann R., 2005, A Particle System for Interactive Visualization of 3D Flows *IEEE Transactions on Visualization and Computer Graphics Vol. 11, No. 6*

Luo, A. et al, 2003. Visualizing Spatial Distribution Data Sets. *In Joint EUROGRAPHICS-IEEE TCVG Symposium on Visualization*. Grenoble, France, pp. 29-38.

Pang, A. et al, 1997. Approaches to Uncertainty Visualization. *The Visual Computer*, Vol. 13, No. 8, pp. 370-390.

Potter, K. et al, 2007.Visual Summary Statistics. *University of Utah Tech Report #UUCS-07-004*.

Rhodes, P. et al, 2003. Uncertainty Visualization Methods in Iso-surface Rendering. *In EUROGRAPHICS 2003 Short Papers*. Granada, Spain, pp.83-88.

Spear, M., 1952. *Charting Statistics*. McGraw-Hill.

Taylor, B. and Kuyatt, C., 1994. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. *NIST Technical Note 1297.*

Tukey, J., 1977. *Exploratory Data Analysis*. Addison-Wesley.

Wittenbrink, C. et al, 1996. Glyphs for Visualizing Uncertainty in Vector Fields. *In IEEE Transactions on Visualization and Computer Graphics,* Vol. 2, No. 3, pp. 266-279.

Wittenbrink, C. et al, 1995. *Verity Visualization: Visual Mappings*. Technical Report, University of California, Santa Cruz.