Picturing Data With Uncertainty

David Kao[1], Alison Love[2], Jennifer L. Dungan[1], and Alex Pang[2]
[1]NASA Ames Research Center [2]Computer Science Department, UCSC

NASA is in the business of creating maps for scientific purposes to represent important biophysical or geophysical quantities over space and time. For example, maps of surface temperature over the globe tell scientists where and when the Earth is heating up; regional maps of the greenness of vegetation tell scientists where and when plants are photosynthesizing. There is always uncertainty associated with each value in any such map due to various factors. For example, exactly where on Earth's surface does a given pixel represent? There could also be cloud and atmospheric contamination in the images taken from satellites. When uncertainty is fully modeled, instead of a single value at each map location, there is a distribution expressing a set of possible outcomes at each location. We consider such distribution data as multi-valued data since it consists of a collection of values about a single variable. Thus, a multi-valued data represents both the data and its uncertainty. We have been working on ways to visualize spatial multi-valued data sets effectively for fields with regularly spaced units or grid cells such as those in NASA's Earth science applications. Successful uncertainty visualization would improve the understanding of climate change, one of the goals in NASA's mission to understand and protect our home planet.

Presently, the most common approach to picturing uncertainty is to plot a histogram representing the distribution of possible values for a chosen grid location. However, this approach does not scale to support spatial multi-valued data sets where there is a distribution at every spatial location. A scientist would need to probe the data at many locations to begin to understand how uncertainty varies spatially. Another common approach to visualizing uncertainty is to depict the spread or shape of the distribution using statistical summaries such as standard deviation, skewness and kurtosis at each grid cell. However, statistical summaries are only meaningful if the distribution is well-behaved e.g. the distribution is unimodal and can be adequately approximated by a Gaussian. However, distributions from most real-world applications are often not well behaved.

A new way to display distributions at multiple grid locations is to project the distributions from an individual row, column or other user-selectable straight transect from the 2D domain. First, at each grid cell in a given slice (row, column or transect), we compute a smooth density estimate from the underlying data. Such a density estimate for the probability density function (PDF) is generally more useful than a histogram, which is a classic density estimate. Then, the collection of PDFs along a given slice are presented vertically above the slice and form a wall. To minimize occlusion of intersecting slices, the corresponding walls are positioned at the far edges of the boundary. The height of the wall is determined by the number of evaluation points of the density estimate. The magnitude of the density estimate is then used to displace points on the wall. This yields a "PDF wall" (Figure 1). The PDF wall depicts the shapes of the distributions very clearly since peaks represent the modes (or bumps) in the PDFs. We've defined roughness as the number of peaks in the distribution [Kao et al. 2002]. Roughness is another useful summary information for multimodal distributions. The uncertainty of the multi-valued data can also be interpreted by the number of peaks and the widths of the peaks as shown by the PDF walls. "Rough" sections of the wall, where both the number and widths of the peaks are higher, implies regions of higher uncertainty. This is important since it give some idea of the confidence levels in the data.

The rendering of the PDF walls can be enhanced by using custom-designed color maps. Statistical summaries shown simultaneously with the walls will allow even better visualization of uncertainty in the domain. For a given distribution, the central tendencies often show what the most likely values are and the spread of a distribution reveals its variability or uncertainty. Other statistical measures describe distribution shape such as its flatness (kurtosis) or asymmetry (skewness). We pseudocolor the walls using an HSV color model (Figure 1) where hue is mapped to frequency of the density estimate, saturation is mapped to the absolute value of skewness, and value is mapped to standard deviation. Hence, a distribution that is very lopsided (high skewness) would have low saturation and a symmetric distribution would have full saturation in color. And when the standard deviation is high (implying more variability or uncertainty), the color of the wall approaches black. Additional information about this work can be found in www.soe.ucsc.edu/research/avis/nasa_is/

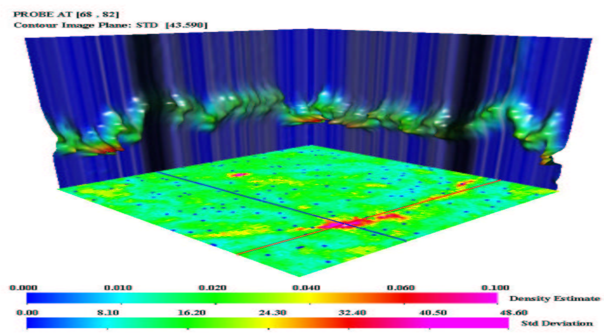

Figure 1: Visualization of a satellite-image derived map, where there is a multi-valued data, for the percentage of forest cover, at each location. The horizontal plane shows the standard deviation of forest cover at each grid cell. The PDF walls are colored by the density estimate, showing mostly unimodal distributions. The left and right walls depict distributions along slices marked by the red and blue lines on the horizontal plane, respectively. The walls are pseudo-colored using our HSV color model. Blue indicates low frequencies and red indicates high frequencies in the density estimates. Dark regions in the walls indicate high uncertainty. This means that the PDFs with the highest standard deviation (greatest uncertainty) are almost invisible. This is quite effective for uncertainty visualization since in the regions where we have the least knowledge (most uncertainty), the walls look dark or blank.

References

JOHNSON, C., AND SANDERSON, A. 2003. A next step: Visualizing errors and uncertainty. In *Computer Graphics and Applications 2003*, IEE Computer Press, 6–9.

KAO, D., LUO, A., DUNGAN, J., AND PANG, A. 2002. Visualizing spatially varying distribution data. In *Proceedings of the 6th International Conference on Information Visualization '02*, IEEE Computer Society, 219–225.