Visualizing the Variations of Ensemble of Isosurfaces

Subhashis Hazarika*   Soumya Dutta†   Han-Wei Shen‡

The Ohio State University, Columbus, Ohio, USA.

ABSTRACT

Visualizing the similarities and differences among an ensemble of isosurfaces is a challenging problem mainly because the isosurfaces cannot be displayed together at the same time. For ensemble of isosurfaces, visualizing these spatial differences among the surfaces is essential to get useful insights as to how the individual ensemble simulations affect different isosurfaces. We propose a scheme to visualize the spatial variations of isosurfaces with respect to statistically significant isosurfaces within the ensemble. Understanding such variations among ensemble of isosurfaces at different spatial regions is helpful in analyzing the influence of different ensemble runs over the spatial domain. In this regard, we propose an isosurface-entropy based clustering scheme to divide the spatial domain into regions of high and low isosurface variation. We demonstrate the efficacy of our method by successfully applying it on real-world ensemble data sets from ocean simulation experiments and weather forecasts.

Index Terms: Ensemble Visualization, Ensemble Isosurface , Isosurface Variation Analysis, Parallel Coordinates.

1 INTRODUCTION

In the field of feature based ensemble analysis [14], scientists often investigate how much variation exists among the features across the different members of the ensemble. Oftentimes these features can be represented by the salient isocontours [2]. Visualization methods like spaghetti plots created by rendering all the isocontour lines coupled with glyphs [15] have been shown to be effective in comparing how ensemble members *agree/disagree* to a given scalar value in 2D data. However, for 3D data, showing such comparative isosurface variations across the ensemble of isosurfaces is a non-trivial problem because of problems like occlusion and clutter. Alabi et al. [1] have approached this problem by slicing the isosurfaces and placing them side-by-side to bring out the differences in an ensemble of isosurfaces but their approach is effective only for ensemble of relatively smaller sizes. Apart from visualizing the spatial variations of multiple surfaces, users often want to identify which spatial regions exhibit maximum variation of features across the ensemble.

To address such issues we propose a new scheme to visualize the surface variations among ensembles of isosurfaces. Instead of finding variations between each pair of isosurfaces we first sort the isosurfaces of an isovalue according to their center-outwards ordering and then find variation among the isosurfaces by using this order information. Center-outwards ordering has been exploited by Whitaker et al. [17] in a previous work, where it was used to create a boxplot equivalent visualization for 2D isocontours. After extending this center-outwards ordering scheme to isosurfaces we take advantage of this ordering to assign a *statistical significance* to

---

*e-mail: hazarika.3@osu.edu
†e-mail:dutta.33@osu.edu
‡e-mail:hwshen@cse.ohio-state.edu

individual ensemble isosurfaces based on their centrality. Using the isosurface order information of a range of isovalues we advocate the following visualization techniques:

- Visualization of the local surface variation of individual isosurfaces of any ensemble member using a novel ensemble distance field. It is to be noted that the proposed scheme not only enables us to measure the distance from the central-most isosurface, but also from a definite confidence interval of isosurfaces as well (e.g: from the top 30% centrally located isosurfaces or 30% quantile)
- Automatic detection and exploration of regions in space where the features from the ensemble members form a consensus or tend to disagree with each other.
- Use of a multivariate information visualization tool (PCP) to display the spatial positioning/ordering of ensemble isosurfaces, which otherwise would be difficult to display using the traditional rendering of surfaces.
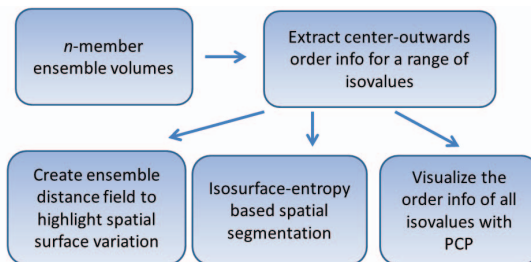

Figure 1: Overview of the steps involved.

2 ORDERING ENSEMBLE OF CONTOURS

Center-outwards ordering [11, 10] of contours was first proposed by Whitaker et al. [17], which was used to create statistical boxplot equivalent for isocontours(called *Contour boxplot*). Order/rank of a contours is also referred to as its *depth*, as it signifies how central or deep the instance is within a distribution. Contour depth is a generalization of the concept of *band-depth* proposed by Sun et al. [16] in statistical literature to quantify and visualize ensemble of functions. Band-depth computation shows sensitivity to the shape as well as the position of functions, hence its generalization finds application in the field of shape analysis [5] as well. The ordering is established by the band-depth(BD) value that gets assigned to each ensemble member. The higher the BD value the more central it is within the given distribution of ensemble.

But applying the concept of band depth to contours/shapes requires us to first create equivalent functional representation of the shapes by using indicator functions [17, 5], i.e, binary functions that are 1 inside and 0 outside of a shape. Once the shapes are converted to binary functions calculating band depth reduces to a sequence of set union and intersection operations. Consider the binary function for the underlying 3D function $F(x,y,z)$

$$f_{bin}(x,y,z) = \begin{cases} 1 : F(x,y,z) \geq v \\ 0 : otherwise \end{cases}$$

This converts the contour for isovalue v to a functional form. The result of this binary function is vectorized and stored as set, say S. Thus, an ensemble of shapes/features/contours can now be represented as an ensemble of sets, say $E_S = \{S_1, ..., S_n\}$. A member of the ensemble, S_i belongs to the band of a random subset of size j (i.e, $\{S_1, ..., S_j\} \in E_S$) if it is confined within the union and the intersection of this subset, i.e:

$$\bigcap_{k=1}^{j} S_k \subset S_i \subset \bigcup_{k=1}^{j} S_k \qquad (1)$$

Set band-depth (sBD) of an ensemble member is thus defined as the probability of the set lying in the band of a random selection of sets from the E_S,

$$sBD_J(S_i) = \sum_{j=2}^{J} P(S_i \in \{S_1, ..., S_j\}). \qquad (2)$$

This application of band depth for contours is coined contour band depth(cBD) [17]. However in practice if there are too many variation of shapes of the individual function or contour and if the size of ensemble is too small the BD value calculated may turn out to be very low or zero which is not going to help establish the relative ordering of the functions among themselves. Hence, a modified approach [12], which is less sensitive to shape variation was proposed by Whitaker et al. [17]. Accordingly a relaxed definition of subset was used that allows a small percentage of each set to violate the classical subset definition. They defined an epsilon subset $\subset_\varepsilon$ where ε is the relaxation parameter.

$$A \subset_\varepsilon B \Longleftrightarrow |A| = 0 \quad or \quad \frac{|A - B|}{|A|} < \varepsilon. \qquad (3)$$

The value of ε should be kept as small as possible such that it still reveals the relative ordering of the shapes.

3 VISUAL-ANALYTIC EXPLORATION OF ENSEMBLE ISO-SURFACE VARIATIONS

The goal of our scheme is not only to convey the isosurface ordering information for a single isovalue as presented in section 2, but also to highlight and utilize the ordering trend for a range of isovalues. In order to achieve this, we start by finding the common scalar data-range for all the ensemble member volumes. The common scalar range is then discretized into m discrete partitions (bins) $\{i_0, i_1, ..., i_m\}$. We then extract the band-depth order information for all the m discrete partitions $\{i_0, i_1, ..., i_m\}$ by taking the bin center as the representative isovalue. For n ensemble members this gives us an m x n matrix, BD_{mn}. An element of this matrix $BD_{mn}[i][j]$ stores the band-depth value of the isosurface for isovalue i of the jth ensemble member and each row of this matrix forms a sequence of band-depth values for isovalue i.

3.1 Visualizing Specific Isosurface Variations using Ensemble Distance Fields

In order to visualize the variation of an isosurface with respect to the ensemble of isosurfaces of the same isovalue, we propose the creation of a novel distance field for ensemble data. This special distance field stores the minimum distance information from a spatial point to the *statistically significant central isosurface(s)* of the data-value at that point. The *statistically significant isosurface(s)* can be strictly the highest band-depth value isosurface i.e, median isosurface or it can be a quantile of isosurfaces i.e, the set of say top 30% centrally located isosurfaces. We create such distance fields for individual volumes of the ensemble data which is then used to color-map any isosurface extracted from that volume.

Distance field [7] of an isosurface is a field where each grid point is assigned the minimum distance from the point to the isosurface. If IS_θ is the isosurface for isovalue θ, the distance field value at a point p is given by,

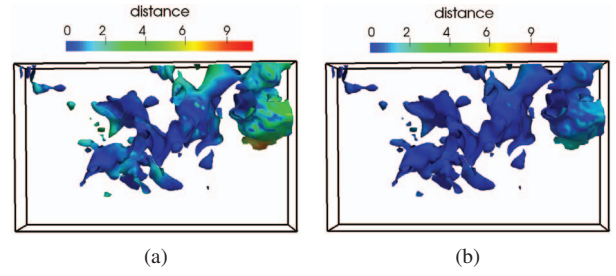$$D_\theta(p) = \min_{\forall q \in IS_\theta} dist(p, q) \qquad (4)$$

Figure 2: Showing the variation of an isosurface from the statistically significant isosurface(s). (a) from median isosurface (b) from the 30% quantile(centrally located isosurfaces). The blue regions are closer to the median isosurface than the greener regions.

where, $dist(p, q)$ is the Euclidean distance between points p and q (in our case p and $q \in \mathbf{R}^3$).

For our new distance field each grid point is assigned the minimum distance from that point to a *statistically significant isosurface(s)*. Let us denote this set of *statistically significant isosurface(s)* for isovalue θ as SSS_θ. SSS_θ can be a set with just one isosurface (just the median) or it can contain multiple isosurfaces (like 30% quantile).

As we have computed the band-depth ordering of all the m discrete isovalues $\{i_0, i_1, ..., i_m\}$, we can easily query for the *statistically significant isosurface(s)* from the matrix BD_{mn} given an isovalue. Let $\delta(v)$ return the discrete isovalue (bin) in which the continuous scalar value v falls in. Then the ensemble isosurface distance field at a grid point p with scalar value v can be formulated as,

$$D_v(p) = \min_{\forall q \in SSS_{\delta(v)}} dist(p, q) \qquad (5)$$

Such an ensemble distance field allows us to quantify the far-off-ness of any isosurface from the central isosurface at every grid point in the spatial domain. Therefore, given this distance field, users can easily explore the local spatial variation of any isosurface. In order to visualize the variation, we apply this distance field as the coloring attribute for the isosurfaces. As seen in Fig. 2 the green regions of the surface is far-off from the statistically significant isosurface(s) compared to the blue region. Thus, with the prior statistical knowledge of the surfaces (i.e, center-outwards ordering) we are able to visualize the individual isosurface variation among multiple surfaces without rendering them together.

3.2 Segmenting Spatial Regions based on Ensemble Isosurface Variation

To allow users to quickly locate the regions that exhibit low or high feature(isosurface) variation we segment the spatial domain of an ensemble member volume based on entropy of isosurfaces variation.

Let, IS_θ^j be the median isosurface for isovalue θ and j signifies that it is from the jth ensemble member. Now for any given ensemble member, i, we compute the minimum distance (Euclidean distance) of all the surface points of the isosurface of θ(i.e, IS_θ^i) to IS_θ^j. The distribution of this distance over all the surface points for IS_θ^i gives us a hint at the credibility/variation of IS_θ^i. Ideally, the distribution should only have non-zero probability at distance 0 as it should be well aligned with the median isosurface (IS_θ^j). A wider distribution tells that the isosurface is far-off and dissimilar to the median isosurface. We call this the *Credibility Distribution* of the isosurface, to quantify this credibility we calculate the Shannon entropy for the sample distribution. The distribution is usually sampled in a discrete form using a histogram. To compute the entropy of a histogram we have to first compute the discrete probabil-
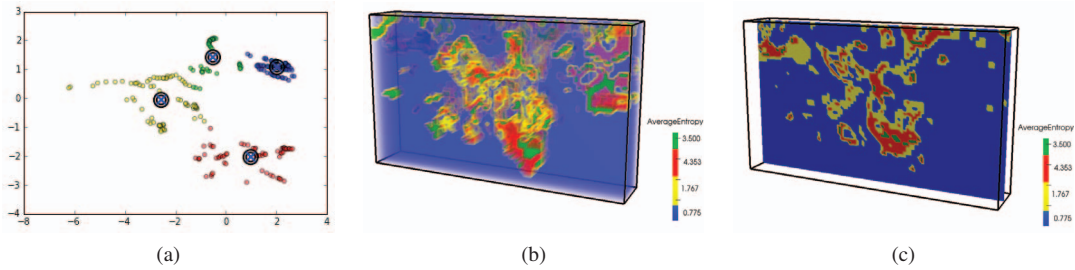
Figure 3: Clustering the ensemble volumes based on isosurface variations. (a) scatter plot of the 2 principal components colored with the k-means ids. (b) one of the ensemble volumes being segmented to regions of high and low isosurface variation. (c) a slicer interface to view different parts of the segmented volume.

ity density function of the histogram, say $p_i(x)$. The entropy of the isosurface of the ith member for isovalue θ is given by,

$$H_\theta^i(X) = - \sum_{l=0}^{bin-1} p_i(x_l) \log p_i(x_l) \qquad (6)$$

We compute this entropy for all the n ensemble members at each of the m discrete θ (i.e, $\{i_0, i_1, ..., i_m\}$). This leaves us with m number of n-dimensional vectors. To cluster the data-values into groups based on their entropy, we perform a k-means clustering of these n-dimensional vectors. The overall variation of a group/cluster is quantified by the average entropy of the ensemble isosurfaces for all the cluster values. When these clusters of data-values are visualized in the spatial domain using volume rendering, we get to see the regions in space with different ensemble isosurfaces variations (Fig. 3b) . In order to visualize these clusters in the value-domain, we do a dimensionality reduction of the vectors via Principal Component Analysis (PCA) and show the scatter plot diagram of the two principal components as in Fig. 3a. We color the points in the plot with the cluster ids of k-means algorithm. The value of k (number of clusters) can be decided by the user by checking the scatter plot to get an idea of how well the points get clustered.

The average entropy of each of this clusters gives an idea of how disturbed are the isosurfaces in that cluster are compared to the other clusters. The segmented volume is another way of looking into ensemble isosurface variation, as it highlights the spatial region for an ensemble member where the isosurface variation is high or low. To help the user explore through different sections of the volume, we also provide an interactive slicer based visualization that helps to slice the volume at different spatial planes to perform detailed variation analysis as shown in Fig. 3c.

3.3 Comparative Visualization of Spatial Ordering of Ensemble Isosurface

In order to provide a comprehensive view of how the relative spatial ordering of the ensemble isosurfaces are affected in different ensemble runs we propose to layout the band-depth order information for multiple isovalues using *Parallel Coordinates Plot* (PCP). PCP is a popular multivariate information visualization tool commonly used for visual exploration of patterns and relationships in high-dimensional/multivariate data [4, 6].

The vertical axes of our PCP represent individual ensemble members and each polyline in the plot is mapped to a discretized isovalue. The values on the vertical axes are the band-depth values for the corresponding isovalues. A high band-depth value signifies that the isosurface is located closer towards the center of the distribution of surfaces for that isovalue. Each line in the PCP conveys the order information for a given isovalue. In Fig. 4b we show the relative order information for the isovalue of 101. It is evident from the plot that, the isosurface from the ninth ensemble member is the median isosurface. Instead of rendering the individual isosurfaces
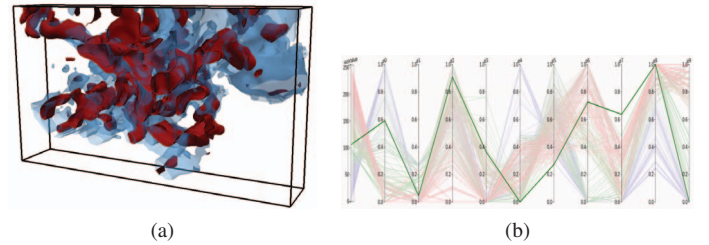

Figure 4: Ensemble isosurfaces for single isovalue. (a) shows the 10 isosurfaces rendered at the same time(the central surface is color red). (b) highlights the order information in PCP for the same 10 isosurfaces. The median isosurface is from the ninth ensemble member as it has the highest band-depth value for axes $e8$

as in Fig. 4a we get a better idea of the surface ordering from the PCP. PCP allows us to convey the order information for not just one isovalue but for a range as well. This in-turn leads to a means for exploring how the isosurface spatial ordering changes for different ensemble runs. Interactive capabilities such as brushing and filtering in the PCP helps users to refine and query for order information for specific data-range as shown in Fig. 5b. The polylines are colored based on the isovalues that they represent, blue corresponds to lower isovalues while red to the higher isovalues.

4 APPLICATION AND RESULTS

We apply our visualization scheme on two different ensemble 3D data sets to demonstrate the effectiveness of our technique.

Our first data set is a multivariate ensemble data set covering the region from the Massachusetts Bay to the Cape Cod area of the US east coast [13, 9, 8]. This data set is divided into 53x90 grid with 16 depth levels. The results shown so far in the paper are generated from this data set.

We have selected the temperature variable and 10 ensemble members to perform the experimentation. Fig. 5 shows the PCP plot of the ordering information of the temperature isosurfaces for 256 discrete isovalues. Fig. 5a shows that the lower isovalues (with blue line) have high BD values at the first axis, which means that the isosurfaces from the first ensemble member tend to be more centrally located for this value range. Moreover, the interactive brushing and selection helps filter out results for certain range of value (Fig. 5b), thus giving the users more flexibility to explore the order information.

Fig. 2a shows an isosurface from the ninth ensemble member. The color on the isosurface reflects the distance of the surface from the median isosurface for that isovalue. The blue parts are closer to the median compared to the green region. Fig. 2b presents the distance of the same surface from the top 30% centrally located iso-
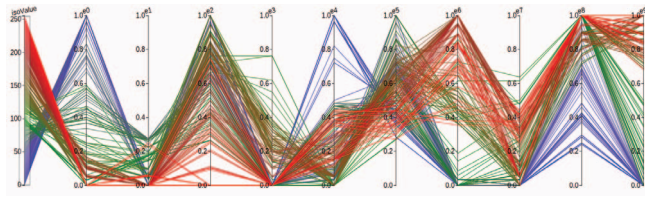
(a) Center-outwards ordering information of ensemble isosurfaces for all iso-values using PCP
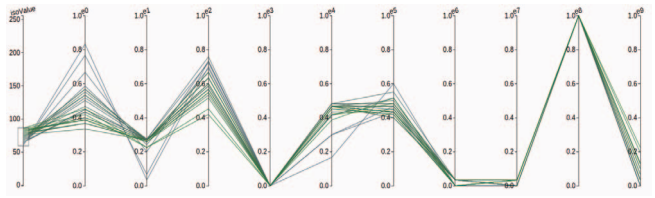
(b) Use of brushing to select specific isovalue range.

Figure 5: A PCP view of the band-depth(ordering) information for all temperature isovalues for The Massachusetts Bay Ensemble.
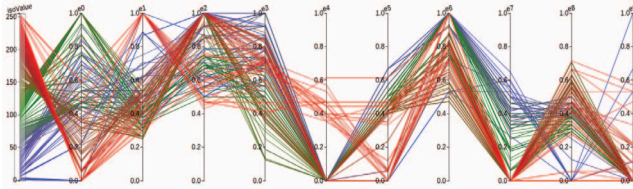

Figure 6: A PCP view of the Great Lakes WRF ensemble data

surfaces. Fig. 3b demonstrates the second ensemble volume with its spatial regions segmented based on the isosurface variations. The blue region(mostly the landmass) shows less isosurface variation compared to the red and the green regions. Fig. 3c shows a slice-view in the z-direction. The scatter-plot of Fig. 3a shows how the entropy of the isosurfaces are grouped in a low dimensional space.
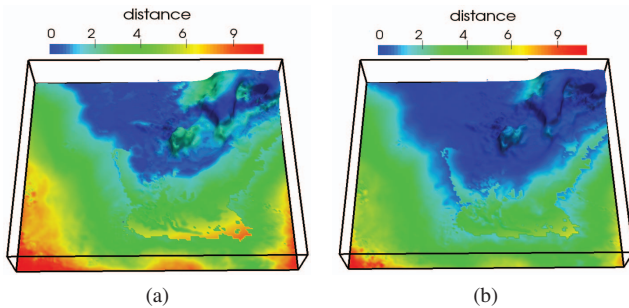

Figure 7: Isosurface of one of the Great Lakes ensemble member (a) shows the distance of different parts of the isosurface from the median isosurface (b) shows the distance from 30% quantile.
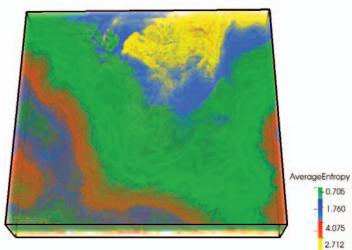

Figure 8: Clustering of regions based on isosurface entropy for Great Lakes data. The segmented volume of one of the ensemble members showing regions of feature variation.

Our second data set is the Great Lakes WRF ensemble [3]. This is a 10-member ensemble of numerical weather forecasts across

the Great Lakes and Upper Midwest using the WRF-ARW forecast model. The data set covers $167x151$ grid with 30 vertical levels, 10kms apart. We have used the pressure variable over the domain to perform our analysis.

The complete center-outwards ordering is revealed in Fig. 6. Fig. 7a shows an isosurface from one of the ensemble members color-mapped to the distance from the corresponding median iso-surface whereas Fig. 7b shows the distance of the same surface from the top 30% centrally located isosurfaces. To identify the spatial regions with high and low feature variation we have performed the k-means clustering of the isosurface entropy values for all the iso-values. Fig. 8 shows one of the ensemble volumes segmented based on the clustering results into regions of high and low feature varia-tions. The yellow and the blue regions tend to cover most parts of the four lakes in the great lakes region.

5 PERFORMANCE

The computation of the center-outwards ordering was done as a preprocessing step in C++ on a standard workstation PC (Intel i7 at 3.40GHz and RAM of 15.7GB). The Massachusetts dataset took about 30 minutes to evaluate the order information for entire range of isovalues while the Great Lakes dataset took about 45 minutes. Once the order information are available the computation time of the *ensemble distance field* per member and *spatial segmentation* is shown in the following table.

Table 1: Average CPU Time performance per ensemble member

Data Sets	Dist. Field(mins.)	Cluster Volume(mins.)
Mass. Bay	7.67	5.67
Great Lakes	12.20	11.31

6 CONCLUSION AND FUTURE WORK

In this work, we have proposed new ways of visualizing ensem-ble isosurface variations using a new distance field measure, an isosurface-entropy based spatial clustering and an information vi-sualization tool(PCP) based way to provide surface variation clue to users. This work is not a solution to the problem of visual clut-ter and occlusion while rendering multiple surfaces but rather an alternative means to understand the surface variation effects in en-semble study despite those inherent problems. In future, we plan to get some domain expert opinions and try to look for alternative means of ordering the isosurfaces since the current approach is a performance bottleneck when used for a range of isovalues.

REFERENCES

[1] O. S. Alabi, X. Wu, J. M. Harter, M. Phadke, L. Pinto, H. Petersen, S. Bass, M. Keifer, S. Zhong, C. Healey, et al. Comparative visualization of ensembles using ensemble surface slicing. In *IS&T/SPIE Electronic Imaging*, pages 82940U–82940U. International Society for Optics and Photonics, 2012.

[2] A. Biswas, S. Dutta, H.-W. Shen, and J. Woodring. An information-aware framework for exploring multivariate data sets. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2683–2692, 2013.

[3] C. Evans. *http://derecho.math.uwm.edu/wrf/*.

[4] J. Heinrich and D. Weiskopf. State of the art of parallel coordinates. *STAR Proceedings of Eurographics*, 2013:95–116, 2013.

[5] Y. Hong, Y. Gao, M. Niethammer, and S. Bouix. Depth-based shape-analysis. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pages 17–24. Springer, 2014.

[6] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.

[7] M. W. Jones, J. A. Bærentzen, and M. Sramek. 3d distance fields: A survey of techniques and applications. *Visualization and Computer Graphics, IEEE Transactions on*, 12(4):581–599, 2006.

[8] P. F. Lermusiaux. Uncertainty estimation and prediction for inter-disciplinary ocean dynamics. *Journal of Computational Physics*, 217(1):176–199, 2006.

[9] P. F. Lermusiaux, C.-S. Chiu, G. G. Gawarkiewicz, P. Abbot, A. R. Robinson, R. N. Miller, P. J. Haley, W. G. Leslie, S. J. Majumdar, A. Pang, et al. Quantifying uncertainties in ocean predictions. Technical report, DTIC Document, 2006.

[10] R. Y. Liu. On a notion of data depth based on random simplices. *Ann. Statist.*, 18(1):405–414, 03 1990.

[11] R. Y. Liu, J. M. Parelius, and K. Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by liu and singh). *Ann. Statist.*, 27(3):783–858, 06 1999.

[12] S. López-Pintado and J. Romo. On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734, 2009.

[13] A. L. Love, A. Pang, and D. L. Kao. Visualizing spatial multivalue data. *Computer Graphics and Applications, IEEE*, 25(3):69–79, 2005.

[14] H. Obermaier, K. Joy, et al. Future challenges for ensemble visualization. *Computer Graphics and Applications, IEEE*, 34(3):8–11, 2014.

[15] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. J. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1421–1430, 2010.

[16] Y. Sun and M. G. Genton. Functional boxplots for complex data visualization, 2009.

[17] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2713–2722, 2013.