Eurographics Conference on Visualization (EuroVis) (2013)
M. Hlawitschka and T. Weinkauf (Editors)

Visualizing Contour Distributions in 2D Ensemble Data

Tobias Pfaffelmoser and Rüdiger Westermann

Computer Graphics and Visualization Group, Technische Universität München, Germany

Abstract

Overlaid plots of iso-contours in individual members of a scalar ensemble field are a popular concept to visualize the data uncertainty. However, such plots do not allow inferring on the spatial cumulative probability distribution of the iso-contours, and they cannot reveal distribution characteristics like spread and topology for very large amounts of contours. In this paper, we propose a new visualization technique for iso-contours in ensemble data sets to overcome these limitations. Our technique makes no assumption about a stochastic uncertainty model, rendering it suitable for arbitrary ensemble distributions. It computes a statistical summary of the ensemble over the spatial domain, including probability density values for arbitrary domain points. From this information, the uncertainty and topology of iso-contours can be determined, as well as the variations in gradient magnitude around these contours. Since the visualization is carried out on the GPU, our approach allows analyzing even very large ensemble data sets at interactive rates.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms, Viewing algorithms

1. Introduction

Analyzing ensemble data sets is important, because they give insight about the uncertainty in the data. Such insight is often derived by assuming a multivariate Gaussian probability distribution. In this case, a number of techniques can effectively visualize the possible variations of specific data features like iso-contours. For arbitrary distributions, however, a methodology is necessary to estimate this variability.

The variability of iso-contours is often visualized via so-called *spaghetti-plots*. They show the iso-contours for an iso-value θ in all members of a scalar 2D ensemble simultaneously. See Fig. 1 (a) and (b) for an example, in which the plots do not reveal any difference between the two ensembles. In Fig. 1 (c) and (d), a member of each ensemble is shown. The data is color-coded from blue ($< \theta$) over white ($= \theta$) to red ($> \theta$). As indicated, while in one ensemble the iso-contour is positioned *either* in the left *or* the right branch, in the other ensemble it occurs simultaneously in both branches. These examples clarify that visualizing features in individual ensemble members in one image can be greatly misleading and does not allow for a reliable estimation of the feature uncertainty in general.

This paper presents a study to shed light on the visualization of iso-contour distributions in scalar ensemble fields. It makes no assumption about a stochastic uncertainty model, yet, by computing a statistical summary of the ensemble over the spatial domain, it generates point-wise measures for the positional likelihood of occurrence of iso-contours.

2. Related Work

An overview and taxonomy of uncertainty visualization techniques is given in [PWL97, JS03, THM*05, GS06, Pot]. A variety of general techniques for visually representing data uncertainty have been proposed in [WPL02, DKLP02, RLB*03, LLPY07]. Other approaches address the visualization of the positional variations of specific features in scalar and vector fields [PWL97, ZWK10, KWTM03, GR04, Bro04, PRW11, PWH11, PH10, PPH12, OT12, OGT11b, OGT11a, OGHT10]. Structural variations of salient features in scalar fields have been addressed in [PW12b, PW12a] by visualizing positive and inverse global and local correlation structures. Underlying all these approaches is a Gaussian uncertainty model. The visualization of uncertain iso-contours in 2D ensemble data sets was explicitly addressed in [PWB*09] using spaghetti-plots. The use of spaghetti-plots in combination with glyphs and confidence ribbons was proposed in [SZD*10].

DOI: 10.2312/PE.EuroVisShort.EuroVisShort2013.055-059

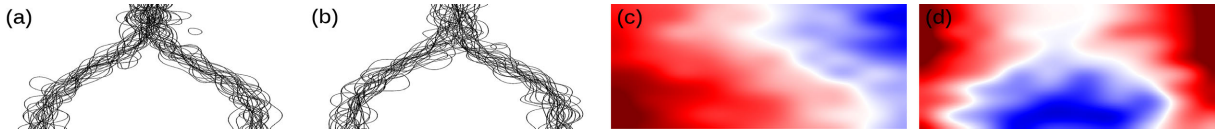

Figure 1: *(a), (b) Spaghetti plots of iso-contours in two different ensembles. (c), (d) Members of the respective ensembles.*

3. Spatial Probability Distribution

We first develop stochastic distribution functions for characterizing the *spatial variability of iso-contours* in uncertain scalar fields, i.e., the *spatial cumulative distribution function* (CDF) and the *spatial probability density function* (PDF) of iso-contours. We assume a discrete sampling of a 2D domain on a Cartesian grid structure with grid points $\mathbb{S}_{a,b} = \{\mathbf{x}_{i,j} : 1 \leq i \leq a, 1 \leq j \leq b\}$. An ensemble has n members, the k-th member containing scalar values $y_k(\mathbf{x}_{i,j})$ and gradients $\nabla y_k(\mathbf{x}_{i,j})$. At each grid point, the *data uncertainty* is given by the variation of the data values. The data uncertainty can be modeled by assigning a *random variable* $Y(\mathbf{x}_{i,j})$ to each grid point $\mathbf{x}_{i,j}$.

To analyze the *variability in position of iso-contours* for an iso-value θ in the ensemble members, our goal is to determine for every grid point the probability that a contour is located exactly at this point. At a point $\mathbf{x}_{i,j}$, the probability that the data takes on the value θ is $P(Y(\mathbf{x}_{i,j}) = \theta)$. Unfortunately, this probability vanishes if the data is real-valued and smooth. This is because a 1D contour (or 1-manifold) in a 2D domain has a zero Lebesgue measure, i.e., its area is zero. Since a non-zero probability of the occurrence of a contour is only possible across a non-zero area, it must hold that $P(Y(\mathbf{x}_{i,j}) = \theta) = 0$. It is thus not possible to directly compute probabilities of the occurrence of iso-contours at particular domain points.

Instead of interpreting an iso-contour as a 1-manifold embedded into the 2D domain, however, we regard it as the boundary between the regions containing all points with scalar values above and below the iso-value. These regions are called the *superlevel* and *sublevel sets*, respectively, and have non-zero Lebesgue measures in general. They can thus be used to derive probability measures.

For this purpose we introduce the spatial CDF

$$\Psi_\theta(\mathbf{x}_{i,j}) := P(Y(\mathbf{x}_{i,j}) \geq \theta) = \frac{1}{n}\sum_{k=1}^{n} P_k(Y(\mathbf{x}_{i,j}) \geq \theta), \quad (1)$$

where P_k is a boolean indicator function $\mathbb{1}_k$ for each ensemble member k:

$$P_k(Y(\mathbf{x}_{i,j}) \geq \theta) = \mathbb{1}_k(\mathbf{x}_{i,j}) := \begin{cases} 0 & \text{if } \; y_k(\mathbf{x}_{i,j}) < \theta \\ 1 & \text{if } \; y_k(\mathbf{x}_{i,j}) \geq \theta \end{cases}. \quad (2)$$

The spatial CDF expresses the probability that the scalar value at a particular grid point is greater than θ with respect to the data uncertainty at that point. It is defined for any number n of ensemble members. In Fig. 2 (b), Equ. (2) was applied to the single ensemble member shown in (a). A color table was used to map values in the interval $[0,1]$ to a color range from blue to red.

Since a CDF is obtained by integrating the respective PDF, we obtain the spatial PDF via differentiation as

$$\psi_\theta(\mathbf{x}_{i,j}) := \left\| \nabla \Psi_\theta(\mathbf{x}_{i,j}) \right\|. \quad (3)$$

In the following we will show for the spatial PDF that a) it can be computed at each grid point by using only the ensemble data at that point, and b) can be used directly to encode the positional uncertainty of the iso-contours. Note that the spatial CDF and PDF only model the distribution of the iso-contour along the so-called *normal curves* of the CDF field (cf. [PRW11]). Both functions are defined on a 2D *spatial domain* and not on a 2D *data* or *parameter domain* like traditional 2D distribution functions.

3.1. Gaussian Contour Representation

The spatial PDF expresses the probability density of the iso-contours in the 2D domain. However, it cannot be evaluated directly, because the indicator function in Equ. (2) is neither continuous nor differentiable. Thus, we replace the binary indicator function by a function that generates a smooth and differentiable transition between the sublevel and superlevel sets. We choose a function with a closed-form first derivative, such that a closed form of Equ. (3) can be obtained.

One possible choice is the well-known CDF Φ of the standard normal distribution, yielding

$$P_k(Y(\mathbf{x}_{i,j}) \geq \theta) = \Phi\left(\frac{y_k(\mathbf{x}_{i,j}) - \theta}{\sigma_s}\right). \quad (4)$$

Since Φ is computed as an integral of the Gaussian PDF, it serves our requirements. Fig. 2 (c) shows, for the ensemble member in (a), the values of P_k for a small positive *sharpness parameter* σ_s.

Due to the chosen smoothing function, with increasing/decreasing values $y_k(\mathbf{x}_{i,j})$ above/below the iso-value in one ensemble member k, we obtain $P_k(Y(\mathbf{x}_{i,j}) \geq \theta) \to 1$ and $P_k(Y(\mathbf{x}_{i,j}) \geq \theta) \to 0$, respectively. The transition rate between the superlevel and sublevel region is controlled by the sharpness parameter. For a fixed sharpness parameter, the spatial CDF has a strong gradient magnitude in those regions along the contour where the data gradient is strong, too.

By substituting the smoothing function (Equ. (4)) into Equ. (1), and using the resulting CDF in Equ. (3), we arrive

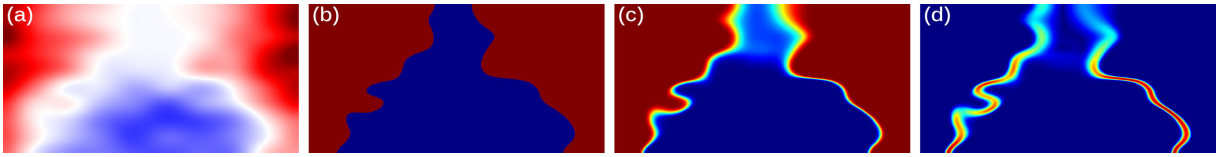

Figure 2: *(a) Member of the ensemble in Fig. 1 (b). (b) Spatial CDF for single ensemble member in (a) using binary indicator transition function. (c) Spatial CDF using Gaussian transition function. (d) Spatial PDF as derivative of spatial CDF.*

at the spatial PDF for the selected member:

$$p_k\left(\mathbf{x}_{i,j}\right) = \phi\left(\frac{y_k(\mathbf{x}_{i,j}) - \theta}{\sigma_s}\right)\frac{\left\|\nabla y_k(\mathbf{x}_{i,j})\right\|}{\sigma_s}. \qquad (5)$$

Here, ϕ is the bell-shaped PDF obtained by differentiating the standard normal distribution function. Note that for each member k, Equ. (5) models a Gaussian "uncertainty region" around each single iso-contour, but not for the distribution of the set of iso-contours. The degree of this "uncertainty" is directly related to the data gradient and can be interpreted as a *condition indicator* (cf. visual condition analysis of iso-contours in [PH10]). In Fig. 2 (d), for the ensemble member in (a), the values of p_k are first transformed to $[0,1]$ via $1-\exp\left(-p_k(\mathbf{x}_{i,j})\right)$, and then to color. The color transition between blue and red allows clearly distinguishing between low and strong gradient regions along the iso-contour. The spatial PDF modeling the probability density of the contours in the whole ensemble can now be written as

$$\psi_\theta(\mathbf{x}_{i,j}) = \left\|\frac{1}{n}\sum_{k=1}^{n}\phi\left(\frac{y_k(\mathbf{x}_{i,j}) - \theta}{\sigma_s}\right)\frac{\nabla y_k(\mathbf{x}_{i,j})}{\sigma_s}\right\|. \qquad (6)$$

The spatial PDF gives rise to a quantitative assessment of the local probability density of iso-contours, which is caused by their positional variation in the ensemble. The distribution puts into relation the strength of the spatial variations. It is worth noting, however, that the spatial PDF covers only the positional variation of an oriented boundary contour. Variations in topology, e.g., flip of the superlevel and sublevel regions across one and the same contour, are not taken into account. However, this is not the case for most ensemble data sets, affected by moderate uncertainty.

3.2. Visualization

The spatial CDF computes for every grid point the probability that this point belongs to the region in which the scalar values are greater than the iso-value. The probabilities range from 0 to 1, and they are mapped linearly to gray scales from black to white. The resulting colors serve as background colors C_b, which indicate the sublevel and superlevel regions, as well as the transition zone in between.

Next, we display the color-coded spatial PDFs of all ensemble members simultaneously. The color-coded contours appear wider and more diffuse in regions showing low gradients, and more narrow and sharper in regions with strong gradients. By drawing all spatial PDFs simultaneously, wider regions could completely hide narrower ones. Therefore, instead of averaging the spatial PDF values of all members (cf. Equ. (6)), we always select the largest value at each domain point:

$$\psi_\theta^{\max}(\mathbf{x}_{i,j}) = \max_{k=1,2,\ldots,n} p_k\left(\mathbf{x}_{i,j}\right). \qquad (7)$$

Note that both ψ_θ and $\psi_\theta^{\max}$ range between 0 and ∞. These values are first mapped to $[0,1]$ using the transfer function $\alpha(\mathbf{x}_{i,j}) = 1-\exp\left(-\tau\cdot \mathrm{T}(\mathbf{x}_{i,j})\right)$, where $\mathrm{T}(\cdot)$ is either replaced by ψ_θ or $\psi_\theta^{\max}$. The scaling parameter τ is used to control the color contrast.

In addition to the background color C_b, we further construct two foreground colors: The first one, *lower color* C_l, is obtained by linearly mapping Ψ ($\in[0,1]$) to the color map [yellow $\rightarrow$ green $\rightarrow$ cyan]. The second one, the *upper color* C_u, is constructed by mapping Ψ to [red $\rightarrow$ magenta $\rightarrow$ blue]. The final color at each grid point is obtained by linearly mapping α from $[0 \rightarrow 0.5 \rightarrow 1]$ to $[C_b \rightarrow C_l \rightarrow C_u]$. The terms *lower color* and *upper color* indicate relatively low and high gradients along the iso-contours.

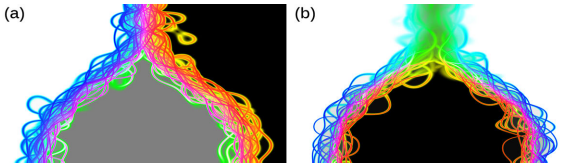

Figure 3: *Spatial CDF/PDF plots for ensembles in Fig. 1.*

The color scheme allows the simultaneous encoding of the values of the spatial CDF—shows the transition between the sublevel and superlevel sets—and the spatial maximum PDF—indicates regions with high and low gradients. In Fig. 3, the color scheme was applied to the ensembles in Fig. 1. In (a), regions colored white/black contain those points that belong to the superlevel/sublevel region in all ensemble members. A gray value of 0.5 indicates an equal number of ensemble members in which the respective point is in the superlevel or sublevel set. Thus, from the location of the gray region it can be concluded on a multi-modal distribution of the iso-contours, i.e., the contours are positioned either left or right of the gray region. The color represents the spatial PDF: Strong presence of C_u indicates almost constant gradient magnitude along the iso-contours and no contrast be-

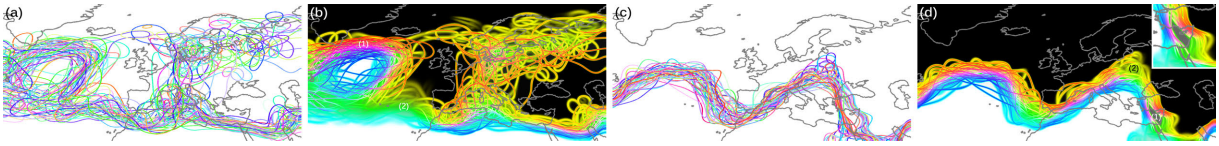

Figure 4: *Two ECMWF temperature ensemble data sets: (a),(c) Spaghetti plots. (b),(d) Spatial CDF and PDF.*

tween low and high gradient regions. In (b), the background color indicates that there are no larger regions belonging *either* to the superlevel *or* sublevel regions. Compared to (a), it cannot be concluded on a strong multi-modality in the iso-contour representation; black and white regions reveal a clear separation between the superlevel and sublevel region. The presence of C_l in the upper part of the domain indicates a much lower gradient strength than in the lower part.

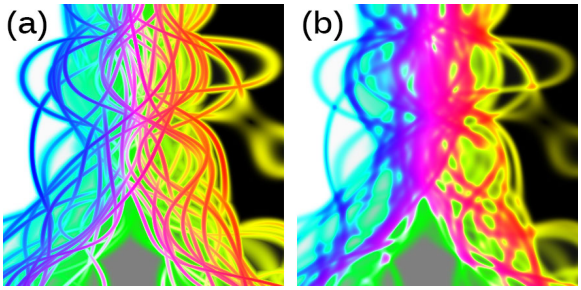

Figure 5: *Spatial PDF values for ψ_θ^{max} (a) and ψ_θ (b).*

In Fig. 5, we illustrate the differences between using $\psi_\theta^{\max}$ and ψ_θ for visualizing the contour distributions. In (a) and (b) we show the upper domain part of the data set in Fig. 1 (a), with $\psi_\theta^{\max}$ and ψ_θ as probability measures, respectively. The individual contours can be visualized much more effectively in (a), and the visual focus is always put on the contours along which the gradients are most prominent. Due to the averaging of values in (b), contour points where several contours intersect each other receive higher values than points where no crossing occurs. This results in a shift of the visual focus from the contours to the intersection regions, and an increasing loss of the contours' shapes.

4. Results

Our approach has been applied to two different temperature ensembles, each comprising 50 members. The ensembles have been generated by the European Center for Medium-Range Weather Forecast (ECMWF) for two different forecast periods and pressure levels above Europe. Since all computations are carried out for every domain point in parallel on the GPU, the user can interactively monitor the changes caused by selecting different iso-values θ, sharpness parameters σ_s and color contrasts τ.

In Fig. 4 (a), iso-contours are visualized via spaghetti-plots, with different colors assigned to contours in different ensemble members. Although only 50 members are plotted, the visualization quickly becomes cluttered, and the limitations of spaghetti-plots prohibit a detailed statistical analysis. In (b), the spatial CDF and the maximum values of the spatial PDF are shown. The gray-valued background allows clearly segmenting the domain into regions with temperature values above (white) and below (black) the selected iso-value. The "sharpness" of the iso-contours and the presence of color C_u in region (1) identify a sharp temperature transition with low gradient uncertainty towards the Greenland border. In region (2), the iso-contours have a smooth appearance and do not show a clear preferential direction. Together with the presence of the lower color C_l, this indicates lower gradients and a much smoother temperature transition. These observations cannot be made by just looking at the spaghetti-plot in (a).

Image (c) shows the spaghetti-plot for a different iso-value and pressure level in the second ensemble. Compared to the first ensemble, the visualization in (d) reveals a new statistical feature. In region (1) (also shown in the small image), the iso-contours split up into two branches, which indicates a bi-modal distribution. Both branches enclose a gray-valued background area. In addition, the presence of the upper color C_u indicates strong gradients in this region, compared to, for instance, region (2).

The visualization techniques presented in this paper can be implemented very efficiently on the GPU. For a high screen resolution pixel raster of 1860×1040 and a 2D scalar ensemble data set with a resolution of 1060×460 and 50 members the spatial CDF/PDF computation takes below 20 ms on a NVIDIA GeForce GTX 660 Ti graphics adapter.

5. Conclusion

To enable an uncertainty analysis of iso-contours in 2D scalar ensemble fields, we have derived probability distributions for iso-contours in such fields, and proposed a method to visually convey this information. We did not make any assumption on the underlying uncertainty model, but computed statistical summaries and generated continuous distribution functions thereof.

In the future, we will look into ways to extend our approach towards an uncertainty analysis of other features in scalar ensembles. Furthermore, we aim to extend our method to 3D. Here, an approach similar to the 2D case should work, yet one has to investigate adequate mapping strategies to visually convey the spatial distribution values.

References

[Bro04] BROWN R.: Animated visual vibrations as an uncertainty visualisation technique. In *GRAPHITE* (2004), ACM, pp. 84–89. 1

[DKLP02] DJURCILOV S., KIM K., LERMUSIAUX P., PANG A.: Visualizing scalar volumetric data with uncertainty. *Computers & Graphics 26*, 2 (2002), 239–248. 1

[GR04] GRIGORYAN G., RHEINGANS P.: Point-based probabilistic surfaces to show surface uncertainty. *Visualization and Computer Graphics, IEEE Transactions on 10*, 5 (2004), 564–573. 1

[GS06] GRIETHE H., SCHUMANN H.: The visualization of uncertain data: Methods and problems. In *Proceedings of SimVis 2006* (2006), pp. 143–156. 1

[JS03] JOHNSON C., SANDERSON A.: A next step: Visualizing errors and uncertainty. *Computer Graphics and Applications, IEEE 23*, 5 (2003), 6–10. 1

[KWTM03] KINDLMANN G., WHITAKER R., TASDIZEN T., MOLLER T.: Curvature-based transfer functions for direct volume rendering: Methods and applications. In *Visualization, 2003. VIS 2003. IEEE* (2003), IEEE, pp. 513–520. 1

[LLPY07] LUNDSTROM C., LJUNG P., PERSSON A., YNNERMAN A.: Uncertainty visualization in medical volume rendering using probabilistic animation. *Visualization and Computer Graphics, IEEE Transactions on 13*, 6 (2007), 1648–1655. 1

[OGHT10] OTTO M., GERMER T., HEGE H., THEISEL H.: Uncertain 2d vector field topology. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 347–356. 1

[OGT11a] OTTO M., GERMER T., THEISEL H.: Closed stream lines in uncertain vector fields. In *Proc. Spring Conference on Computer Graphics (SCCG)* (2011), vol. 2. 1

[OGT11b] OTTO M., GERMER T., THEISEL H.: Uncertain topology of 3d vector fields. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE* (2011), IEEE, pp. 67–74. 1

[OT12] OTTO M., THEISEL H.: Vortex analysis in uncertain vector fields. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 1035–1044. 1

[PH10] PÖTHKOW K., HEGE H.: Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *Visualization and Computer Graphics, IEEE Transactions on*, 99 (2010), 1–1. 1, 3

[Pot] POTTER K.:. http://www.sci.utah.edu/~kpotter/Library/Catalogs/uncertaintyVis/. 1

[PPH12] PETZ C., PÖTHKOW K., HEGE H.: Probabilistic local features in uncertain vector fields with spatial correlation. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 1045–1054. 1

[PRW11] PFAFFELMOSER T., REITINGER M., WESTERMANN R.: Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 951–960. 1, 2

[PW12a] PFAFFELMOSER T., WESTERMANN R.: Correlation visualization for structural uncertainty analysis. *Int. J. Uncertain. Quantification* (2012). 1

[PW12b] PFAFFELMOSER T., WESTERMANN R.: Visualization of global correlation structures in uncertain 2d scalar fields. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 1025–1034. 1

[PWB*09] POTTER K., WILSON A., BREMER P., WILLIAMS D., DOUTRIAUX C., PASCUCCI V., JOHNSON C.: Ensemble-vis: A framework for the statistical visualization of ensemble data. In *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on* (2009), IEEE, pp. 233–240. 1

[PWH11] PÖTHKOW K., WEBER B., HEGE H.: Probabilistic marching cubes. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 931–940. 1

[PWL97] PANG A., WITTENBRINK C., LODHA S.: Approaches to uncertainty visualization. *The Visual Computer 13*, 8 (1997), 370–390. 1

[RLB*03] RHODES P., LARAMEE R., BERGERON R., SPARR T., ET AL.: Uncertainty visualization methods in isosurface rendering. In *Eurographics* (2003), vol. 2003, pp. 83–88. 1

[SZD*10] SANYAL J., ZHANG S., DYER J., MERCER A., AMBURN P., MOORHEAD R.: Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *Visualization and Computer Graphics, IEEE Transactions on 16*, 6 (2010), 1421–1430. 1

[THM*05] THOMSON J., HETZLER E., MACEACHREN A., GAHEGAN M., PAVEL M.: A typology for visualizing uncertainty. In *Proc. SPIE* (2005), vol. 5669, Citeseer, pp. 146–157. 1

[WPL02] WITTENBRINK C., PANG A., LODHA S.: Glyphs for visualizing uncertainty in vector fields. *Visualization and Computer Graphics, IEEE Transactions on 2*, 3 (2002), 266–279. 1

[ZWK10] ZEHNER B., WATANABE N., KOLDITZ O.: Visualization of gridded scalar data with uncertainty in geosciences. *Computers & Geosciences* (2010). 1