EuroVis Workshop on Trustworthy Visualization (TrustVis) (2019)
R. Kosara, K. Lawonn, L. Linsen, and N. Smit (Editors)

Uni- and Multi-modal Uncertainty Visualization in 2D Scalar Field Ensembles

Eike Gebauer and Lars Linsen

Westfälische Wilhelms-Universität Münster, Germany

Abstract

The aim of uncertainty-aware scalar field visualization is to convey the most likely case, but also the uncertainty associated with it. In scientific simulations, uncertainty can be modeled using an ensemble approach. Statistics are then drawn from the ensemble outcome to compute the most likely case and its uncertainty. However, the statistical distributions do not necessarily need to be uni-modal. We present an approach to visualize uncertain 2D scalar fields that extends existing uni-modal distributions based on colored heightfields and 2D glyphs to multi-modal ones. We compare the approaches by conducting user experiments for both the uni- and multi-modal case.

1. Introduction

Uncertainty in physical simulations is commonly captured by generating multiple simulation runs forming an ensemble. In particular, when using a Monte Carlo approach one tries to sample the uncertainty in the input parameters to map the uncertainty to the simulation outcome. Then, statistics can be computed from the ensemble to estimate the uncertainties in the outcome. Early approaches for visualizing uncertainties from ensemble data proposed to use colored heightfields to visualize mean and standard deviation of 2D scalar field ensembles from climate simulations [PWB*09, SZB*09]. Colored heightfield visualizations are simple and, thus, were considered to be intuitive and easy to use.

Many approaches have been proposed since to tackle uncertainty visualization in 3D scalar fields, e.g., [PRH10, PH10, PRW11], but most effective visual encodings still operate on 2D slices. Recently, Ristovski et al. [RGH*19] proposed to use circular 2D glyphs to encode uncertainty in 2D slices of 3D scalar fields. They also derive the uncertainties as mean and standard deviation from simulation ensembles, but in a medical context. Such slice-based approaches have the advantage over heightfields that no occlusion and depth perception issues occur, i.e., no interaction is necessary to observe all visualized data.

For this paper, we re-implemented in a first step the approaches of Sanyal et al. [SZB*09] and Ristovski et al. [RGH*19], see Section 3, and performed a user study to evaluate their effectiveness and interpretability, see Section 4. While both approaches restrict themselves to uni-modal uncertainty distributions, we propose extensions of both approaches to multi-modal uncertainty distributions, see Section 5. We also evaluate the extensions in a comparative user study, see Section 6. For the evaluation, we apply our methods to climate simulation ensemble data.

2. Related Work

Early attempts to visualize uncertainty in 2D scalar fields are based on color coding mean and standard deviation by heights and/or color. Potter et al. [PWB*09] visualized the mean by color and standard deviation by color in a juxtaposed view or by isocontours or heights in the same view. Sanyal et al. [SZB*09] encoded mean as height and standard deviation by 1D glyphs or color. They performed a comparative study, but there was no clear winner. Recently, Ristovski et al. [RGH*19] proposed to use 2D glyphs that encode both mean and standard deviation. They compared their approach to juxtaposed views of color-encoded mean and standard deviation [PWB*09] to document advantages of the glyphs. Juxtaposed views have the issue that it is difficult to precisely investigate corresponding locations in both views. Ristovski et al. [RGH*19] ruled out heighfield visualizations [PWB*09, SZB*09] arguing that necessary interactions are not desirable in the clinical context to which they applied their tool. In general though, it is worth investigating how their 2D glyphs compare to heightfields, which we investigate in this paper. Since the study by Sanyal et al. [SZB*09] had no clear winner, we decided to use color-coded heightfields, which was also proposed by Potter et al. [PWB*09]. Moreover, all these approaches assume uni-modal distributions of function values within the ensemble. We propose novel visualizations that extend both approaches (colored heightfields and 2D glyphs) to multi-modal distributions.

In this paper, we restrict ourselves to 2D scalar fields. Such 2D scalar fields often occur in climate simulations, which we use as an application field. Obviously, extensions to 3D scalar fields are desirable. However, we argue that effective uncertainty visualizations in volume visualization, that allow for quantitative and not only qualitative estimations of uncertainty (cf. [SE17]), operate in 2D by using slices or clipping planes through the 3D vol-

DOI: 10.2312/trvis.20191184

ume. Ristovski et al. [RGH*19] follow this argumentation and only use 2D views, as quantification is important for medical applications. Praßni et al. [PRH10] used 3D uncertainty visualizations, but decided to combine them with 2D views for quantitative assessments. Pöthkow et al. [PH10] and Pfaffelmoser et al. [PRW11] also used 3D visualizations, but they are most effective when combining them with clipping planes, i.e., again operating on 2D views. Also, approaches that visualize ensembles by bands such as functional [SG11], curve [MWK14], or contour box plots [WMK13] are most effective on 2D slices. Thus, we argue that visualizing uncertainty in 2D scalar fields should still be regarded as an important research task, as it is still the most effective way for visualizing 3D scalar field uncertainty.

3. Uncertainty Visualizations for Uni-modal Distributions

Given an ensemble of 2D scalar fields

$$S := \{f_i : \mathbb{R}^2 \to \mathbb{R}, i = 1, \ldots, n\} .$$

Then, the distribution of the ensemble's function values at any position $\mathbf{x} \in \mathbb{R}^2$ can be summarized by its statistical moments. In particular, when assuming a uni-modal non-skewed normal distribution, the first two statistical moments (mean and variance) describe the distribution. We denote the mean of the ensemble S's function values in $\mathbf{x}$ by

$$\mu_{\mathbf{x}} := \frac{1}{n} \sum_{i=1}^{n} f_i(\mathbf{x}) .$$

Instead of the variance, one can also compute its square root, i.e., the standard deviation in $\mathbf{x}$ from the mean $\mu_{\mathbf{x}}$, which we denote by

$$\sigma_{\mathbf{x}} := \sqrt{\frac{1}{n} \sum_{i=1}^{n} (f_i(\mathbf{x}) - \mu_{\mathbf{x}})^2} .$$

Assuming that the ensemble approximates the uni-modal non-skewed normal distribution of an uncertain phenomenon, the mean $\mu_{\mathbf{x}}$ represents the expected value in $\mathbf{x}$, while the standard deviation $\sigma_{\mathbf{x}}$ is a measure for the uncertainty in the expected value. In particular, if all ensemble members agree, i.e., $f_i(\mathbf{x}) = \mu_{\mathbf{x}}$ for all i, the standard deviation vanishes and there is no uncertainty.

Colored Heightfields. To visualize an uncertain scalar field, one is primarily interested in observing the expected field value, i.e., the ensemble's mean for all points $\mathbf{x} \in \mathbb{R}^2$ within the domain of the scalar field. An uncertainty-aware visualization then, in addition, depicts the uncertainty in the form of the standard deviation at all points $\mathbf{x} \in \mathbb{R}^2$. We follow the colored heightfield approach presented by Sanyal et al. [SZB*09], where the mean is encoded by height and standard deviation by color. For color map generation, we follow the guidelines by Kovesi [Kov15]. Figure 1 shows an example. Obviously, due to occlusion and depth perception issues multiple views are necessary to fully observe the encoded information.

2D Glyphs. The 2D glyphs propsed by Ristovski et al. [RGH*19] have a circular shape (isotropic encoding) and use three areas (inner circle, ring, outside the ring) for each glyph at a position $\mathbf{x}$, see Figure 2 (left). Instead of encoding mean $\mu_{\mathbf{x}}$ and standard deviation $\sigma_{\mathbf{x}}$, the three areas encode $\mu_{\mathbf{x}} - \sigma_{\mathbf{x}}$ (inner circle), $\mu_{\mathbf{x}}$ (ring), and $\mu_{\mathbf{x}} + \sigma_{\mathbf{x}}$ (outside the ring) by applying a perceptually uniform color map. The idea behind this design is that the color variation within a glyph represents the uncertainty, while the three colors mix perceptually when zooming out and one perceives the mean color encoding $\mu_{\mathbf{x}}$, see Figure 2 (mid). Figure 2 (right) shows a zoomed in version, where the glyphs become visible.

4. Evaluation of Uni-modal Uncertainty Visualizations

In a first experiment, we want to compare the two visual encodings presented in Section 3 assuming uni-modal distributions of function values.

Hypotheses. When providing scales for height values and for the color map legend (not shown in the figures), our hypothesis is that quantification is more precise when reading heights than when reading values off a color map. Hence, we assume that means are more precisely read when working with heightfields, while standard deviation errors should not differ for the two approaches. Moreover, we hypothesize that the interactions with the heightfield would require more time to fulfill the analysis tasks and that the interactions would make the users feel less confident that they found the right answer.

Tasks. Common visual analysis tasks can be described as localization, quantification, and comparison. We have developed respective tasks for both mean and standard deviation. More precisely, the six tasks can be listed as: (1/2) Localize the largest mean/standard deviation. (3/4) Read mean/standard deviation at a selected position. (5/6) Choose the larger mean/standard deviation among two selected positions.

Experiments. We selected 15 participants (4 female, age 22–26 years with 2 outliers at age 57, no or corrected visual deficiencies) with no experience in uncertainty/ensemble visualization (7 participants had experience with navigating in 2D/3D environments). We used two climate simulation data sets, where a sufficient number of positions were pre-selected. Given the relatively low amount of participants, we performed a within-subject study, i.e., all participants performed all tasks on both visualization methods. A counter-balanced design was chosen for the order of the visualization methods. We assured that no task was performed on the same data twice. For data sets we used multiple time steps of a climate simulation ensemble with $n = 10$ simulation runs. Before starting the experiments, the participants were given some time to familiarize themselves with the visualizations and the analysis tasks and to ask questions. The participants were told that we estimate errors and record task execution times, where the primary goal was to fulfill the tasks correctly. The task completion error is computed as the absolute value of the differences between the correct and the chosen mean/standard deviation (for tasks 1–4) or as binary correctness values (tasks 5–6). Finally, the participants were also asked to fill in a questionnaire after the tasks to rate their confidence (on a 5-step Likert scale) in the reported results.

Analysis. To statistically analyze the outcome of the experiments, we first test the distributions for normality using the Shapiro-Wilk

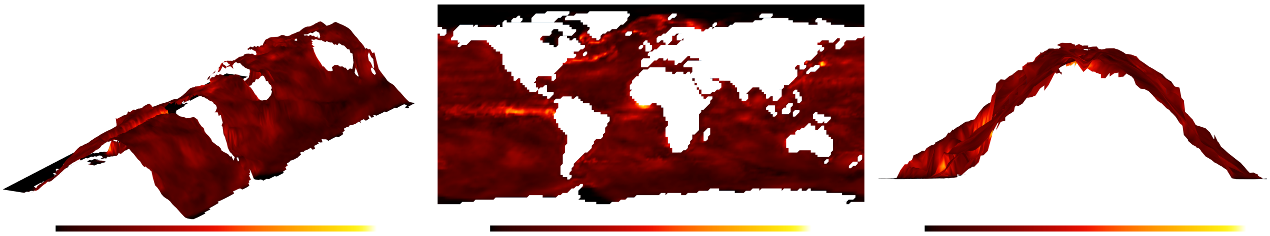

Figure 1: *Multiple views on colored heightmaps (height encodes mean, color encodes standard deviation) for uni-modal uncertainty-aware 2D scalar field visualization. White areas represent land and have no data.*

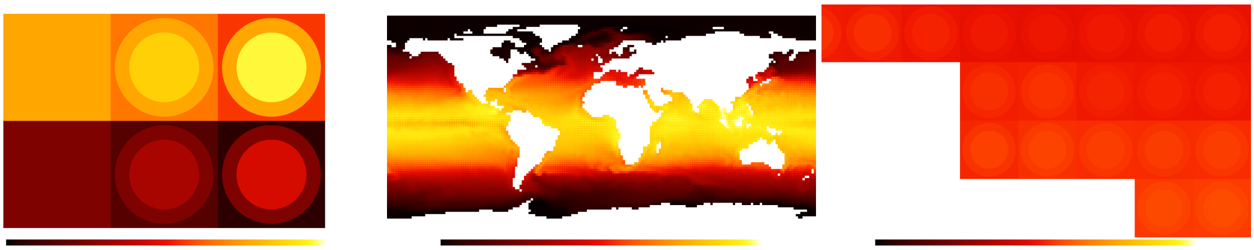

Figure 2: *(left) 2D glyph design illustrated for* $\mu_{\mathbf{x}} = 0.75$ *in top row and* $\mu_{\mathbf{x}} = 0.25$ *in bottom row with* $\sigma_{\mathbf{x}} = 0$*, i.e., no uncertainty, in left column,* $\sigma_{\mathbf{x}} = 0.1$ *in mid column, and* $\sigma_{\mathbf{x}} = 0.2$ *in right column. (mid) Zoomed-out and (right) zoomed-in view of 2D glyphs for uni-modal uncertainty-aware 2D scalar field visualization. White areas represent land and have no data.*

test. In case of normality, we perform a one-way ANOVA test for statistical significance of the null hypothesis against p-value level 0.05. Otherwise, we apply the Wilcoxon signed rank test instead.

Results and Discussion. We first present the error analysis results for the tasks involving the mean. Surprisingly, there was no statistical significance for task 1 (p-value 0.76) or for task 3 (p-value 0.40). Only for task 5, the error is significantly lower for the heightfield visualization (p-value 0.0016). For other tasks, we did not find support for our hypothesis that heights are easier to read than colors.

Next, we look into the error analysis results for the tasks involving the standard deviation. Surprisingly at first glance, the error was significantly higher for all tasks when using the 2D glyphs (task 2: p-value 0.019, task 4: p-value 0.000005, task 6: p-value 0.0016). Hence, we reject the hypothesis that both visualizations would perform equally well for reading standard deviation values. We assume that the reason for this outcome is that the 2D glyphs are not explicitly encoding the standard deviation. Instead, the standard deviation needs to be estimated as the difference between two colors, which apparently is much more difficult than just reading one color value.

Finally, we look into task completion time and confidence averaged over all tasks. Indeed, the confidence levels were much higher when using the 2D glyphs (p-value 0.009) such that we can conclude that the necessary interactions make the user less secure (hypothesis confirmed). On the other hand, for the task completion times there was no significant difference (p-value 0.71, hypothesis not confirmed). We believe that the reason may be that the zooming interactions for the 2D glyph visualization (task 1/2) and the color difference estimation (tasks 2/4/6) took some additional time.

5. Uncertainty Visualizations for Multi-modal Distributions

The visualizations presented above assume a uni-modal non-skewed normal distribution of the function values $f_i(\mathbf{x})$ at a given location $\mathbf{x}$. Hence, it assumes that the mean is indeed the most likely case. It could, however, also be that the function values $f_i(\mathbf{x})$ are high for half of the simulation runs and low for the other half. The mean $\mu_{\mathbf{x}}$ would then be in between and would represent a value that has never occurred in any of the simulation runs. Hence, the mean would not be suitable to represent the ensemble in $\mathbf{x}$. One could use the median instead of the mean to account for skewness, but the median would also not represent the ensemble well. Thus, a better solution is to use a mixture model, i.e., the distribution is a mixture of uni-variate distributions. Assuming as before non-skewed normal distributions, we use a Gaussian mixture model: Let $N(\mu_j, \sigma_j^2)$ be a normal distribution with mean μ_j and standard deviation σ_j, then a Gaussian mixture model with k modes is given by

$$\sum_{j=1}^{k} \alpha_j N(\mu_j, \sigma_j^2)$$

with weights α_j. Given a distribution of function values $f_i(\mathbf{x})$ for $i = 1, \ldots, n$, we fit a Gaussian mixture model using the expectation maximization algorithm, cf. [LLBP12]. The expectation maximization algorithms assumes as an input the number of modes k, which is typically unknown. A common strategy is to execute the algorithm for different k and evaluate the quality of the fit using an information criterion such as the Bayesian or the Akaike information criterion. Despite the fact that only the Bayesian information criterion considers the number of samples n, the Akaike information criterion worked better in our experiments.

Colored Heightfields. The design space for extending the colored heightfield approach to multi-modal distributions is limited. Due to the issue with juxtaposed views of finding corresponding locations, we rejected this option. Drawing multiple colored heightfields with transparency in one view was considered an option, but it quickly turned out to be impossible to read color values from the blended surface renderings. Hence, the only option that remained was to follow the idea of using animations, which was already proposed in the early work by Lundström et al. [LPLY07] and picked up by Liu et al. [LLBP12]. While these approaches show animations of ensemble members respectively Gaussian mixture modes in direct volume rendering, our goal is to explicitly encode mean and standard deviation using the color heightfields. Hence, we alternate between the different modes for each location $\mathbf{x}$ independently. Assuming a Gaussian mixture with k modes and weights α_j, $j = 1,\ldots,k$, for location $\mathbf{x}$, we alternate through all modes and render the j^{th} mode for a duration of $\alpha_j C$ seconds, where C is the duration of one animation cycle. Obviously, if our algorithm only detects one mode for location $\mathbf{x}$, i.e., if $k = 1$, there is no animation.

2D Glyphs. When extending the 2D glyphs for multi-modal distributions, we do not need to consider juxtaposition, animation, or overlay. Instead, we can adjust the glyph design. We follow two design choices. Since the weights α_j sum up to one, they represent percentages and we can generate a pie-chart design making use of a visual encoding familiar even to novices. Thus, we cut the 2D glyphs into slices of size α_j and fill each slice independently with the colors of the respective mode values $\mu_j - \sigma_j$, μ_j, and $\mu_j + \sigma_j$ as in the uni-modal case, see Figure 3 (left) for a bi-modal case.

Because the pie-chart design is not an isotropic encoding, we consider the alternative design choice of using concentric circles. Since the modes are sorted by their mean, we can render the color-coded means in that order using k rings. The width of the rings is proportional to the weights α_j. The circle in the middle shows the color for value $\mu_1 - \sigma_1$, while the area outside the rings shows the color for value $\mu_k + \sigma_k$, see Figure 3 (right). Thus, we produce an isotropic encoding at the expense of not showing the standard deviations of all modes. We consider this a suitable choice, as in the case of multiple modes, the means already reflect the distribution well. On the other hand, introducing more rings for standard deviation encoding would have made the glyphs rather complex, one may have confused means with means $\pm$ standard deviations, and colors may not be sorted anymore due to overlap of neighboring modes.

6. Evaluation of Multi-modal Uncertainty Visualizations

The evaluation of the multi-modal visualizations follows the same approach as the one for uni-modal visualizations. Our **hypotheses** are that animations induce a higher cognitive challenge, which makes the task completion more cumbersome, leading to higher errors, larger timings, and less confidence. Moreover, we do not expect any significant difference in the results for the two glyph designs. The **tasks** were adjusted to the multi-modal settings. Localization tasks: (7/8) Localize the smallest/largest mean among all modes within a region of interest. Quantification tasks: (9/10/11) Read all means/largest mode/standard deviation of first and last mode at a selected position. Comparison task: (12) Choose the largest mean among all modes of two selected positions. The **experiments** were conducted as before with the same participants, same data sets, same set-up, and same protocol. The **analysis** was also conducted as before, only that now we have to compare three conditions. For more than two conditions, we applied the Friedman test in case of no normal distribution. If it reported statistical significance, the Bonferroni-Dunn test was used as a pairwise post-hoc significance test.

Figure 3: *2D glyph designs for multi-modal distributions: (left) pie-chart glyph and (right) isotropic glyph for $k = 2$.*

The **results** show that the errors for the three tested conditions (animated heightfield, two glyph designs) were approximately equal among all tasks. No statistical difference was reported (p-values between 0.19 and 0.95). However, for both average confidence (p-value 0.00000003) and average task completion time (p-value 0.00002) the differences were strongly statistically significant in favor of the glyph designs. The pairwise comparison between animated heightfields and any of the two glyph designs also reported strong statistical significance, whereas there was no significant difference between the two glyph designs. Thus, we can conclude that the hypotheses that animated heightfields lead to lower confidence and higher timings were both confirmed, while the hypothesis that the animated heightfields also produce higher errors was not supported.

7. Conclusion

We have extended existing approaches for uncertainty visualizations of scalar field ensembles from uni-modal to multi-modal distributions. We evaluated the approaches for both uni-modal and multi-modal cases. We have seen that heightfields and glyphs perform generally equal for the uni-modal case, where the glyph design had the disadvantage of not explicitly encoding standard deviation (which can be considered a task formulation issue) while the heightfields led to lower confidence and higher completion times presumably due to the interactions. For the multi-modal case, there was no difference in errors, but the animated heightfields led to even lower confidence and even higher completion times. No difference in the performance of the two glyph designs was found such that it depends on the specific application, which to pick, i.e., whether isotropy is more important or the encoding of all standard deviations.

References

[Kov15] KOVESI P.: Good colour maps: How to design them. *CoRR abs/1509.03700* (2015). 2

[LLBP12] LIU S., LEVINE J. A., BREMER P., PASCUCCI V.: Gaussian mixture model based volume visualization. In *IEEE Symposium on Large Data Analysis and Visualization (LDAV)* (2012), pp. 73–77. doi:10.1109/LDAV.2012.6378978. 3, 4

[LPLY07] LUNDSTRÖM C., PERSSON A., LJUNG P., YNNERMAN A.: Uncertainty visualization in medical volume rendering using probabilistic animation. *IEEE Transactions on Visualization and Computer Graphics 13*, 06 (Nov. 2007), 1648–1655. doi:10.1109/TVCG.2007.70608. 4

[MWK14] MIRZARGAR M., WHITAKER R. T., KIRBY R. M.: Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 2654–2663. URL: doi.ieeecomputersociety.org/10.1109/TVCG.2014.2346455, doi:10.1109/TVCG.2014.2346455. 2

[PH10] PÖTHKOW K., HEGE H.-C.: Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Transactions on Visualization and Computer Graphics*, 99 (2010), 1–15. 1, 2

[PRH10] PRASSNI J.-S., ROPINSKI T., HINRICHS K. H.: Uncertainty-aware guided volume segmentation. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Dec. 2010), 1358–1365. 1, 2

[PRW11] PFAFFELMOSER T., REITINGER M., WESTERMANN R.: Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. *Computer Graphics Forum 30*, 3 (2011), 951–960. 1, 2

[PWB*09] POTTER K., WILSON A., BREMER P.-T., WILLIAMS D., DOUTRIAUX C., PASCUCCI V., JOHNSON C.: Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated ViSUS-CDAT systems. *Journal of Physics: Conference Series 180* (jul 2009), 012089. doi:10.1088/1742-6596/180/1/012089. 1

[RGH*19] RISTOVSKI G., GARBERS N., HAHN H. K., PREUSSER T., LINSEN L.: Uncertainty-aware visual analysis of radiofrequency ablation simulations. *Computers & Graphics 79* (2019), 24 – 35. doi:10.1016/j.cag.2018.12.005. 1, 2

[SE17] SAKHAEE E., ENTEZARI A.: A statistical direct volume rendering framework for visualization of uncertain data. *IEEE Transactions on Visualization and Computer Graphics 23*, 12 (Dec 2017), 2509–2520. doi:10.1109/TVCG.2016.2637333. 1

[SG11] SUN Y., GENTON M. G.: Functional boxplots. *Journal of Computational and Graphical Statistics 20*, 2 (2011), 316–334. doi:10.1198/jcgs.2011.09224. 2

[SZB*09] SANYAL J., ZHANG S., BHATTACHARYA G., AMBURN P., MOORHEAD R. J.: A user study to compare four uncertainty visualization methods for 1D and 2D datasets. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1209–1218. 1, 2

[WMK13] WHITAKER R. T., MIRZARGAR M., KIRBY R. M.: Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2713–2722. doi:10.1109/TVCG.2013.143. 2