Vision, Modeling, and Visualization (2024)
L. Linsen and J. Thies (Eds.)

Region-based Visualization in Hierarchically Clustered Ensemble Volumes

H. Rave[†1], M. Evers[2], T. Gerrits[3], L. Linsen[1]

[1]University of Münster, Germany [2]University of Stuttgart, Germany [3]RWTH Aachen University, Germany

Abstract

Ensembles of simulations are generated to capture uncertainties in the simulation model and its initialization. When simulating 3D spatial phenomena, the value distributions may vary from region to region. Therefore, visualization methods need to adapt to different types and shapes of statistical distributions across regions. In the case of normal distribution, a region is well represented and visualized by the means and standard deviations. In the case of multi-modal distributions, the ensemble can be subdivided to investigate whether sub-ensembles exhibit uni-modal distributions in that region. We, therefore, propose an interactive visual analysis approach for region-based visualization within a hierarchy of sub-ensembles. The hierarchy of sub-ensembles is created using hierarchical clustering, while regions can be defined using parallel coordinates of statistical properties. The identified regions are rendered in a hierarchy of interactive volume renderers. We apply our approach to two real-world simulation ensembles to show its usability.

1. Introduction

Volume data ensembles are generated when models of 3D phenomena are applied to simulate respective fields, examples being climate simulations [SD10], oceanographic simulations [TZG*17], or simulating medical processes [RPHL14]. The ensembles are created by varying initial configurations or parameters to the model to capture the *uncertainty* in the outcome. This uncertainty introduced into the data can lead to a more conclusive understanding of the underlying phenomenon, but it also makes the analysis of the data much more difficult. Several studies conclude that developing new visual analysis methods for such uncertain ensemble data are an important challenge in visualization [WSJ*12; GSWS21]. In this paper, we focus on 3D single- or *multi-field ensembles*, i.e., the ensemble members consist of one or multiple volumetric scalar fields.

Direct volume rendering [DCH88] is one of the main approaches to visualizing 3D scalar fields. It allows for an interactive analysis of the scalar field based on the configuration of a globally defined *transfer function*. When analyzing 3D scalar field ensembles, a common strategy is to visualize the volume based on a statistical model such as assuming normal distributions of the scalar values per voxel and analyzing derived fields of mean and standard deviation. Such an analysis is based on the assumption that the same statistical model applies to all voxels. However, the value distributions per voxel may vary substantially and a single statistical model cannot match all distributions.

We, therefore, propose to visualize ensembles based on regions with similar distributions. While a region with normal distributions can be well represented (and visualized) using the means and standard deviation of its voxels, the representation of multi-modal distributions is more complex. We propose to subdivide the ensemble into *sub-ensembles* (subsets of the ensemble) in regions with multi-modal distributions to investigate whether the sub-ensembles exhibit distributions with simpler statistical models. This process can be iterated by making use of a *hierarchical clustering* approach of ensemble members. We propose and evaluate three different ways of creating a hierarchy of sub-ensembles (see Section 3).

To define regions of similar distributions, we use several common *statistical properties* such as mean and standard deviation as well as derived measures that can capture the characteristics of uncertainty within spatial regions and enable the analysis of statistical distributions (see Section 4). These measures form a multi-dimensional data space, whose visualization and interactive selection we support by employing *parallel coordinates* [ID90].

The hierarchy of sub-ensembles can then be directly integrated into the visual analysis process by translating it into a modifiable hierarchy of direct volume renderers (see Section 5). This facilitates the visual analysis of ensembles by allowing the simultaneous analysis of multiple sub-ensembles. It also quantifies the probability of different modes occuring within the ensemble as larger sub-ensembles represent more likely events. Additionally, the hierarchical structure also lends itself well to a proposed top-down analysis approach.

Since the ensemble may exhibit different patterns in different

† hennes.rave@uni-muenster.de

DOI: 10.2312/vmv.20241206

regions of the volume, the user can define 1D and 2D transfer functions for each of these regions individually, which are then combined into a single rendering. Various *brushing* interactions are available to the user to select regions of interest such as axis brushing in parallel coordinates, region brushing in the volume renderers, and region brushing in the scatter plot of the 2D transfer function editor.

Our main contributions center around building a hierarchy of sub-ensembles by identifying regions with similar statistical distributions and creating a hierarchical visualization of volume renderers, all encapsulated within an integrated open-source tool (`https://github.com/hennesrave/reghievis`). The tool is aimed at helping simulation analysts find and explore multimodality within ensemble data.

2. Related Work

While *direct volume rendering* works well for individual volumes, it lacks a straightforward generalization to ensemble data. Rendering a high number of ensemble members individually can lead to mental overload when visualizing them sequentially, visual clutter when visualizing them in superposition, and bad scalability when visualizing them in juxtaposition. To solve this problem, one can create a statistical model of the underlying data distributions, for example, by calculating the mean and standard deviation in each voxel. Kniss et al. [KKH02] showed that *multi-dimensional transfer functions* can be used effectively when multiple features are of interest in a given dataset, e.g., the original data value and the gradient magnitude, which can also be applied to the mean and standard deviation for ensembles.

Fitting a single Gaussian distribution to the data by computing the mean and standard deviation results in a loss of detail in most cases. Liu et al. [LLBP12] chose *Gaussian mixture models* (GMM) to represent the input data because they allow complex data to be fitted with relatively few parameters. When rendering the volume, the GMM can then be sampled every frame to create a fuzzy rendering, where narrower distributions produce a more stable image. Since this requires animation over time, which may be undesirable, Liu et al. also suggest using *Monte Carlo sampling* in a single frame to create still images. However, these approaches perform worse when the underlying data cannot be accurately represented using a GMM. Athawale et al. [AMS*21] describe an approach to direct volume rendering for cases in which the data uncertainty is characterized by non-parametric distributions, following up on earlier work [SE17]. First, the authors estimate the probability density function at each voxel and divide it into quantiles. During the rendering process, they interpolate the quantile representations of neighboring voxels and integrate the values. The color is then applied using a multi-dimensional transfer function. However, these approaches do not target the analysis of regionally different distributions, which is what we propose. We make use of parallel coordinates, which have previously been used in combination with volume visualizations to distinguish between regions of interest [GXY12; RFA*22].

A recent survey on ensemble data visualization [WHLS19] presents many approaches operating on *2D visualizations*. These reduce the problem of occlusion but do not represent the volumetric shapes well. Some approaches include a linearization of the volume, e.g. along a space-filling curve, to accomplish such 2D visualizations [DDW14; WFG*19]. An alternative volume visualization approach are *isosurface renderings* which have been adapted to include uncertainty [RLBS03; PWH11; ASJ21]. Confidence isosurfaces [ZWK10] have been generalized to cover uncertainty in multi-variate data [SAJ21]. Again, none of these approaches allows for a detailed investigation of the different statistical distributions.

Hierarchical clustering in feature space has been used for multivariate data [LVRR08; LVR09; ML18; DLL11]. However, we do not cluster different features, but ensemble members. This has been proposed as a first step in several approaches [EL22; FKRW17; KBL19], but they do not make use of the hierarchy to analyze the different sub-ensembles directly, e.g., by using a hierarchy of volume renderers. Similar to He et al. [HTWL18], one could use bicluster analysis to identify subsets of voxels that have a similar value pattern among multiple ensemble members. However, we are mainly looking into identifying regions where voxels have a unimodal distribution, independent of the order of ensemble members, which is a less restrictive criterion. Also, their approach does not yield disjoint sets, making it more difficult to assess the likelihood of certain patterns to emerge just by looking at the resulting visualizations.

3. Hierarchical Sub-ensemble Analysis

Given an ensemble of volumes, our main goal is to identify modes and variations within spatial regions of the ensemble and sub-ensembles thereof. The main idea behind clustering the ensemble members is to find smaller groups of members that belong to the same mode, making them easier to analyze, but still providing a comprehensive picture of the whole ensemble when combined.

Since it is not a priori known how granular the clustering of ensemble members should be, hierarchical clustering is a good choice for the clustering algorithm. It allows the user to follow a top-down approach by starting with the ensemble as a whole and then gradually breaking it down into smaller sub-ensembles. The goal in each iteration is to find a sub-ensemble in which all members belong to the same mode in a large region of the volume. We propose three methods for creating such a hierarchy of sub-ensembles: First, we try complete-linkage agglomerative hierarchical clustering using a suitable similarity measure between members, in our case the *field similarity* [FL19]. We use complete-linkage to minimize variance within our clusters since it increases the likelihood of finding a sub-ensemble that forms a single mode of the ensemble. Second, we test divisive hierarchical clustering using k-means clustering with $k = 2$ (to generate a binary tree) and the Euclidean distance as a distance metric. Third, we sort the values in each voxel and compute the average rank of each ensemble member. Let $A = (a_1, ..., a_m)$ be the ensemble members ordered by this average rank. We then split the ensemble at index

$$j = arg \max_{1 \leq i \leq m} i \cdot q(\{a_1, ..., a_i\}) + (m - i) \cdot q(\{a_{i+1}, ..., a_m\}),$$

where q is defined as the number of voxels where the Shapiro-Wilk normality test [SW65] yields a non-significant result ($p \geq 0.05$)

for values of the given ensemble members. Note that we weight the quality of a subset by its size to penalize small sub-ensembles. The rationale behind this third approach is as follows: We want to split the ensemble such that for both sub-ensembles, the number of voxels exhibiting a normal distribution is maximized. Those voxels can then be easily understood and visualized using, for example, the mean and standard deviation by mapping them to color and opacity respectively.

Since some spatial regions may have already been fully analyzed, the calculation of all of these methods may be restricted to a user-selected subset of voxels to increase clustering quality. In Section 6.5, we evaluate the effectiveness of these three approaches.

The result of such a hierarchical clustering can be visualized in a *dendrogram* as shown in Figure 1. Since this dendrogram may become unwieldy for a large number of ensemble members, we provide several mechanisms to ease interaction. There is a compressed view, which saves horizontal screen space, the user can collapse individual nodes to hide entire sub-trees and, if clustering based on field similarity was performed, there is also a slider to set a similarity threshold, such that all nodes whose similarity is higher than the threshold are automatically hidden. Interactions with the dendrogram are shown in the accompanying video.

The main purpose of the hierarchical clustering is to help the user to perform a top-down analysis. Therefore, the user can directly transfer the current tree into a hierarchy of volume renderers (cf. Section 5). A node in the tree corresponds to a volume renderer that displays the corresponding sub-ensemble of the underlying tree node. In this way, the user can analyze different levels of the ensemble simultaneously, identify differences between sub-ensembles, and estimate the likelihood of key patterns in the ensemble.

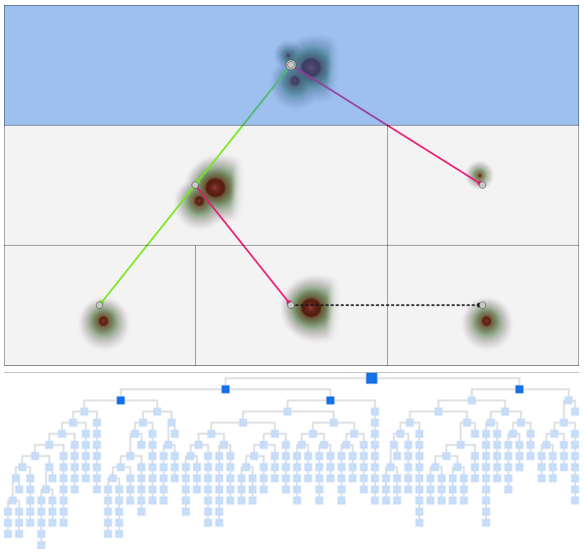

Figure 1: *Directed graph of volume renderers (top) and corresponding dendrogram (bottom) for a dataset representing three spheres with 150 ensemble members. Green lines are edges of type left child, red lines of type right child, and the dashed black line of type sibling. The corresponding nodes are highlighted in dark blue in the dendrogram. The node of the currently selected volume renderer at the top is displayed larger.*

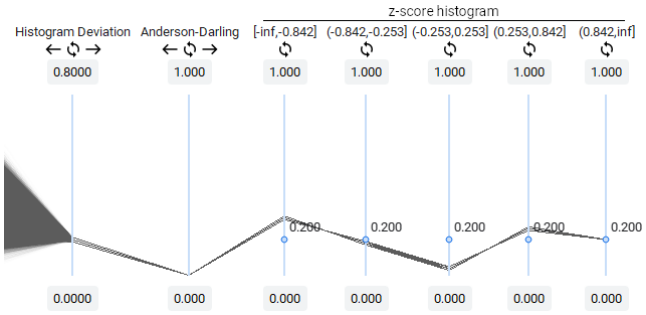

Figure 2: *Parallel coordinates of several derived volumes. The values of the Anderson-Darling test are all close to 0 and the z-score histogram shows no straight lines, indicating that the values in those voxels do not follow a normal distribution.*

4. Region-based Analysis

Parallel coordinates [ID90] are a common tool for multidimensional data analysis. Here, they are used to identify correlations between different derived quantities or to define regions inside the volume that meet certain criteria such as a specific statistical data distribution. Hence, they can be used to identify regions of voxels with similar distributions. When the user selects a sub-ensemble, various derived volumes can be calculated to observe statistical properties or distributions. For each of these quantities, a parallel coordinates axis is available. Given fields of multi-field ensembles are added as further parallel coordinate axes whenever the user selects them. Having configured the parallel coordinates axes, commonly used brushing operations support the definition of regions. A region is represented as a binary volume, where the value of a voxel is 1, if it is selected, and 0, otherwise. In this way, existing regions can be combined using common *Boolean operations* to form new regions. These regions can then be used, edited, and refined in the 1D and 2D transfer function editors and in the volume renderers (cf. Section 5).

Several derived volumes are available to the user for use as parallel coordinates axes, see Figure 2. In the following, we will explain which features are available and how they are computed. All features are computed per voxel and we chose them because they are commonly used statistics that provide information about the data distribution. This list is not exhaustive and more measures can always be added if required by the specific analysis task.

Minimum and Maximum. The derived volumes *min* and *max* store at each voxel x the minimum or maximum value of all ensemble members for voxel x, respectively.

Mean and Standard Deviation. The derived volumes *mean* and *stddev* store at each voxel x the mean and standard deviation of all ensemble members for voxel x, respectively.

Gradient Magnitude. The derived volume *grad* stores the magni-

tude of the gradient of the *mean* volume at each voxel. The gradient is computed using finite differencing and its magnitude using the L^2-norm.

Anderson-Darling Test. Determining whether the values of a specified sub-ensemble conform to a normal distribution is an important analysis task in our approach. In the case of normal distribution, the corresponding voxels can be fully understood using mean and standard deviation only. We do not use the Shapiro-Wilk test from before because the computation of derived volumes happens on demand and must therefore be interactive. That is why the *Anderson-Darling test* [AD54] is used to test for normality. Although Razali et al. [MY11] showed that the Shapiro-Wilk test is better for testing normality, the Anderson-Darling test was ranked second in terms of statistical power (cf. Section 6 for a demonstration of its effectiveness) and its computation is much faster.

Z-Score Histogram. Another tool to determine normality is the *z-score* histogram. The z-values are counted in five bins. For normal distributions, each bin should contain exactly 20% of the values. Therefore, we chose the following intervals for the bins: $[-\infty,-0.842]$, $(0.842,-0.253]$, $(-0.253,0.253]$, $(0.253,0.842]$, and $(0.842,\infty)$. The derived volumes $hist_1,\ldots,hist_5$ represent the percentage of ensemble members that fall into the five bins. A voxel whose values are normally distributed is then displayed as a straight horizontal line along the five parallel coordinate axes for the z-scores $hist_1,\ldots,hist_5$. While brushing is more intuitive on the single axis of the Anderson-Darling test, we included the z-scores in addition, as they provide more detailed insight into the distribution of the data.

Z-Score Deviation. The maximum deviation from the 20% z-score among $hist_1,\ldots,hist_5$ provides another normality measure and is computed for each voxel x by $dev(x) = \max_{1\leq i\leq 5}(|hist_i(x)-0.2|)$.

Principal Components. *Principal component analysis* is performed on a data matrix where each row contains the values of a single voxel across all ensemble members. Then, each voxel is projected into a 2D space using the first two principal components. The result is mapped to the unit square $[0,1]^2$ using only translation and uniform scaling, thus, maintaining the shape of the projection. The two principal components per voxel are used as two derived volumes. Their main purpose is to use the resulting scatter plot to define a 2D transfer function, but the volumes can also be used as parallel coordinates axes.

5. Region-based Sub-ensemble Volume Visualization

We are now in the position to describe the rendering of sub-ensembles as defined in Section 3 for selected regions as defined in Section 4 using transfer functions. The rendering process uses a modified version of volume ray casting, to create a *hierarchy of volume renderers* for the simultaneous analysis of multiple sub-ensembles.

Region-based Ray Casting. Instead of using a global transfer function for the entire volume, the user can select different transfer functions for different regions within the volume. If several regions overlap, the user can assign priorities to the regions. Then, the region with the highest priority that evaluates to an opaque color will be used to color the corresponding sample. Additionally, the user can select a separate 1D transfer function for the alpha channel, separating the meanings of color and opacity.

The direct volume rendering follows the concept of ray casting, where the colors of samples along each ray are lit using the *Phong reflection model* for a point light source with no attenuation. The gradient of alpha values is calculated by sampling around the current position using central differences. The negative gradient is then used as surface normal. The material properties in the form of the reflection coefficients can be set by the user. Four common composition schemes are supported: *First hit*, *maximum intensity*, *first local maximum*, and *front-to-back* composition.

The volume renderer also supports region brushing. By using a lasso selection, voxels that are within the selection (when projected to the screen) are added to or removed from the current region (depending on which mouse button is used).

Directed Graph of Volume Renderers. Using multiple volume renderers for simultaneously rendering multiple sub-ensembles can improve the insight into the structure of the ensemble. The sub-ensembles can interactively be selected from the dendrogram (cf. Section 3). The user can transfer the current dendrogram directly into a hierarchy of volume renderers, where collapsed nodes in the dendrogram are ignored. The layout of the volume renderers is then calculated automatically based on the number of leaves and the level of the corresponding nodes. For fine-grained control over the visualization, the respective volume renderers can be freely rearranged and resized inside a grid of user-defined size. In particular, the user can create edges between volume renderers manually to form a hierarchy. Three types of edges can be added: *left child*, *right child*, and *sibling*, which directly correspond to the relationships of nodes in the dendrogram. Hence, the user can create a *directed graph*, where the nodes represent volume renderers of sub-ensembles corresponding to the dendrogram, see Figure 1. Volume renderers without an inbound edge are called *root volume renderers*. To define root volume renderers, the user must manually select a node in the dendrogram. All other volume renderers are then automatically assigned their corresponding sub-ensemble by following the corresponding edges in the dendrogram. If an edge in the volume renderer graph is not present in the dendrogram, the respective volume renderer is disabled.

All visualization parameters, such as illumination parameters, composition modes, camera position, and viewing direction, are shared by all volume renderers to allow for direct comparisons. The only differences between the various volume renderers in the directed graph are the input volumes and (when desired) the transfer functions. Note that two volume renderers when applied to different sub-ensembles can display differently shaped regions, since the selection of regions by brushing on parallel coordinates axes is evaluated individually for each volume renderer.

While the juxtaposed volume renderers already allow for visual comparisons, we also support the explicit encoding of differences by computing *difference volumes* of each node to its parent node (if existent). Then, the difference volumes instead of the given volumes are rendered in the directed graph of volume renderers, which allows for detecting regions where the sub-ensembles mostly differ.

6. Evaluation and Results

In the following, we will first describe the general analytical workflow, before presenting results for two real-world datasets (*Red Sea* and *Radiofrequency Ablation*). Our goal within this analysis is to explore the uncertainty by identifying modes and estimating the likelihood of their appearance in the ensemble. Finally, we compare the three methods proposed for the hierarchical clustering of sub-ensembles. In the supplementary material, we provide further results including synthetic datasets (*Teardrop* and *Tangle*). Finally, we discuss the performance and scalability of our implementation.

6.1. Analytical Workflow

The presented methods are integrated into an interactive visual analysis tool. A top-down analysis approach is recommended, which starts by analyzing the ensemble as a whole, e.g. by searching for patterns in the parallel coordinates widget. A common step is also to look for regions where the values are normally distributed, using the Anderson-Darling axis or the z-score histogram. The histogram can also help to find regions that follow a bi-modal distribution. Once such regions of interest are identified, the users typically define one or more transfer functions that meet their analytical needs. This may be a simple 1D transfer function, e.g., using the mean, or a more complex 2D transfer function, e.g., using the first two principal components. If a deeper understanding of a region is required, the user may recalculate the dendrogram using only that region. As a next step, the ensemble can be split into smaller sub-ensembles – creating a hierarchy of volume renderers. Thus, the problem can be broken down into smaller parts that are easier to understand. By combining the insights from the visual analyses of sub-ensembles, the user can gain a comprehensive understanding of the ensemble. If a region is of particular interest, investigations using the slice viewer provide exact values for a deeper understanding. Of course, this general analytical workflow is not fixed and the user can deviate from it if desired.

6.2. Red Sea Dataset

The Red Sea dataset [Hot] is a multi-field ensemble with 50 members, 60 timesteps, and volumes of size $50 \times 500 \times 500$. We analyze the first timestep for all 50 members and the two fields *temperature* (in °C) and *salinity* (in $\mathrm{g\,m^{-3}}$) to demonstrate the advantage of hierarchical volume rendering. When looking at the whole ensemble, regions of high variance, and thus uncertainty, can be identified by visualizing the standard deviation volume of the salinity field, see Figure 3. While it is easy to identify regions of high variance, there is no way to determine the various factors that contribute to this variance. Using hierarchical volume rendering, we visualize multiple sub-ensembles simultaneously, which gives insight into the key differences between ensemble members in these regions, see Figure 4. We conclude that this central region exhibits multiple modes across the ensemble while surrounding regions are much more stable. These regional differences are caused by the large number of unstable vortices in the center of the Red Sea and could not be identified by looking at a volume rendering of the whole ensemble.

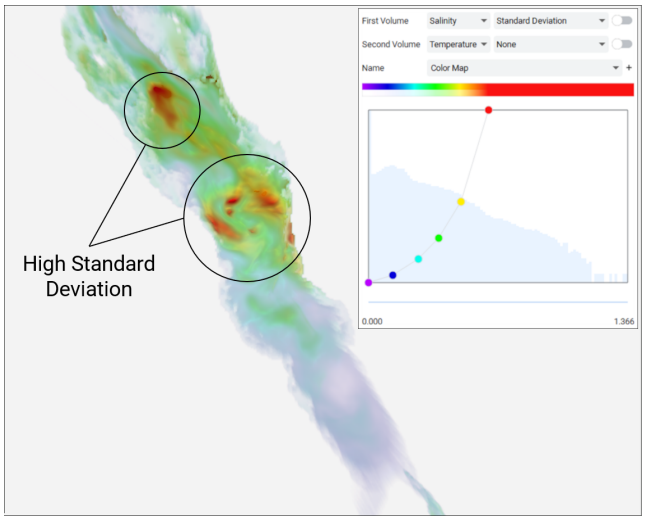

Figure 3: *Volume rendering of the Red Sea dataset using the standard deviation volume of the salinity field of all members. The regions of high variance are clearly visible. However, the source of this variance is unknown – no factors leading to the differences among members can be identified from this visualization.*

6.3. Radiofrequency Ablation Dataset

The radiofrequency ablation (RFA) dataset is an ensemble with 510 members and volumes of size $92 \times 92 \times 92$ courtesy of Sandeep Gyawali, David Sinden and Tobias Preusser (Fraunhofer MEVIS, Jacobs University Bremen). The volumes contain temperature values in Kelvin that are the result of simulations of radiofrequency ablation [HEG*22] of a liver tumor. The analysis begins by brushing the parallel coordinates axis of the Anderson-Darling test to find regions where the values follow a normal distribution. A volume hierarchy with a field similarity threshold of 0.9995 is created, see supplementary material. Then, the mean difference volume is used to see the differences between the sub-ensembles using a diverging color map in the transfer function editor, see Figure 5. The differences are mostly located inside the tumor region, where differences between the upper tumor region, where the tip of the ablation needle is located, and the bottom tumor region, where the shaft of the ablation needle is located, can be seen. It can be observed that the first sub-ensemble on the left, which contains 275 volumes, has lower values on average (blue colors), while the second sub-ensemble containing the remaining 235 volumes has higher values (red colors). This is confirmed by looking at the 1D histograms in the 1D transfer function editor for these two sub-ensembles. Looking only at voxels close to the tumor, our approach revealed a level of bi-modality within the ensemble. Our collaboration partners were surprised by this finding, stating that they were not expecting this bi-modality since the underlying parameter space was sampled uniformly for this ensemble.

6.4. Performance and Scalability

We performed several performance measurements to determine the scalability of our implementation. For the computation of derived

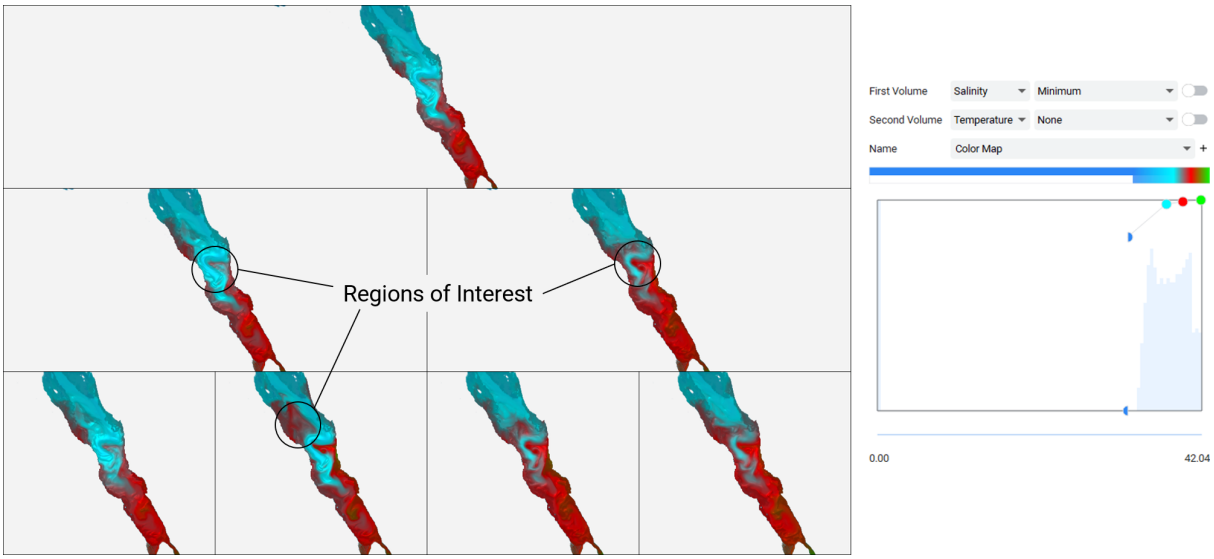

Figure 4: *Hierarchical volume rendering of the Red Sea dataset using the minimum volume of the salinity field of all members and the transfer function on the right. The first three levels of the dendrogram were translated into a hierarchy of volume renderers. Clear differences between sub-ensembles can be seen in regions previously identified as highly uncertain (see Figure 3). Salinity in the middle parts of the Red Sea appears to differ significantly among ensemble members.*

volumes and similarity measures, we tested different volume sizes (100^3, 200^3, and 300^3) and different numbers of ensemble members (from 20 to 750). All tests were performed using an AMD Ryzen 5 5600X CPU, an NVIDIA RTX 2060 graphics card, and 16 GB of main memory. Detailed results are included in the supplementary material. Our results show that all derived volumes can be computed on demand, with the *Anderson-Darling* volume taking the most time with up to 11.37 seconds at volumes of size 200^3 for 200 ensemble members. The computation of the hierarchical clustering takes by far the most time. For *k*-means clustering, it takes less than a minute to compute, for field similarity-based clustering, it can take a couple of minutes. Clustering using the average rank of ensemble members can take a few hours. These times are not optimized and can still be greatly improved. However, clustering is usually carried out once at the beginning as a pre-processing step, so that these times do not hinder the interactive analysis later on.

For rendering times, we tested different volume sizes (100^3, 250^3, and 500^3) and different numbers of volume renderers (1, 3, 7, and 15) on a 2560×1440 screen with a ray sampling rate of 3 samples per voxel. As the parallel coordinates and volume renderers work on derived volumes, their rendering times grow with an increasing number of voxels, but they do not depend on the number of members. For parallel coordinates, the performance can always be improved to interactive rates by only rendering a selected number of random voxels. For volume rendering, the worst time of 59.44 ms still yields an interactive 16 frames per second and was measured using a volume of size 500^3 and 15 volume renderers. Note that the rendering time does not scale linearly with the number of volume renderers, since they occupy the same total screen space, so the number of rays cast remains roughly constant.

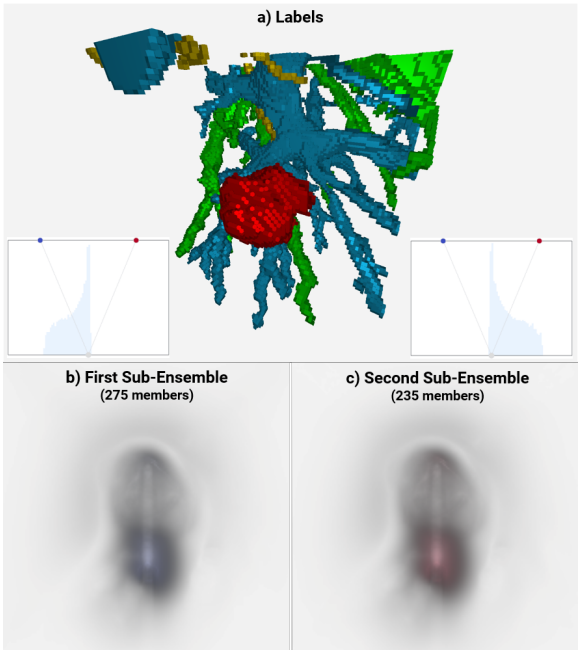

Figure 5: *(a) Volume rendering of the label volume of the RFA dataset: tumor (red), hepatic venous system (green), portal vein (blue), and hepatic artery (yellow). (b, c) Volume renderings for the two sub-ensembles after splitting the ensemble once. The difference in mean to the whole ensemble is color-coded. The first sub-ensemble (b) has cooler temperatures compared to the whole ensemble while the second one (c) has higher temperatures. The 1D histograms (insets) confirm this observation.*

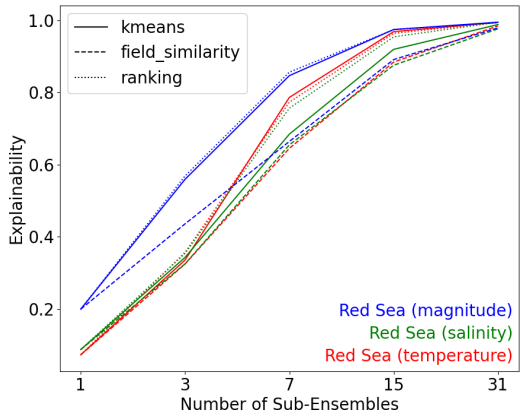

Figure 6: *Explainability of the sub-ensemble hierarchy for the temperature, salinity, and velocity magnitude fields of the Red Sea dataset. The explainability increases with the number of allowed splits in the hierarchy. The three methods, kmeans, field_similarity, and ranking perform very similarly.*

6.5. Hierarchical Clustering

We wanted to compare the quality of our three hierarchical clustering methods. Therefore, we evaluate how well the clustering methods allow for dividing the ensemble into sub-ensembles with many normally distributed voxels. We used the temperature, salinity, and velocity magnitude fields of the Red Sea dataset.

For each dataset, we started with the whole ensemble and performed the Shapiro-Wilk normality test for each voxel. If the p-value is not significant ($p \geq 0.05$), we mark the voxel as completed. We then split the ensemble in two using the three different methods, only considering non-completed voxels. This procedure is repeated recursively until a sub-ensemble consists of less than six members or until a maximum of four splits. We determined the quality of a sub-ensemble clustering as follows: For each voxel, we count the number of ensemble members in each sub-ensemble of our hierarchy where its values are normally distributed. This definition makes sense from a visual analysis standpoint, as volume rendering using the mean and standard deviation becomes an effective tool in this case. Note that this number can never exceed the total number of ensemble members because our clustering is disjoint and once normally distributed, voxels are marked as completed and therefore ignored on the next tree level. We refer to the ratio between this number and the total number of ensemble members as *explainability*, which is a value in $[0, 1]$. Computing the mean over all voxels gives us explainability for the entire sub-ensemble hierarchy, see Figure 6. This measure essentially represents the percentage of the ensemble that we can model using normal distributions, thus making it more easily explainable.

We can see that the explainability increases with the number of allowed splits. The three methods perform mostly the same. Thus, given the vastly different computation times described previously, the k-means clustering seems to be a good compromise between quality and speed and is used in the application.

7. Discussion and Conclusion

We presented an interactive visual analysis approach that is based on defining a hierarchy of sub-ensembles and analyzing them simultaneously using a directed graph of volume renderers. The analysis is driven by deriving several statistical fields from the given scalar field(s) to obtain insights into statistical distributions. By brushing in parallel coordinates on those derived fields, one can define regions of different behaviors such as regions with voxels that exhibit a uni-modal normal distribution and those that exhibit multi-modal distributions. The properties of regions can be visually analyzed using 1D and 2D transfer function editors.

We proposed multiple methods for the construction of such sub-ensemble hierarchies and demonstrated their effectiveness in increasing the explainability of ensemble datasets. We applied our approach to real-world data and obtained meaningful results. Of course, the hierarchy of sub-ensembles is only meaningful, if clusters of similar sub-ensembles exist. If the statistical distribution is much simpler, it may not be possible to define meaningful sub-ensembles, but then the ensemble as a whole can be analyzed easily using a single volume renderer in conjunction with the parallel coordinates plot and the transfer function editors.

A potential pitfall for our approach could be that regions of similar statistical distributions do not form connected regions in 3D space, but are scattered across the volume. Such regions would not allow for a meaningful analysis. However, our tool provides enough interaction mechanisms to exclude not meaningful regions from further analysis steps.

The computation of the hierarchical clustering is the most expensive part but can be performed in a pre-processing step. If recomputing the clustering on selected regions during the interactive analysis is desired, this can be performed significantly faster. The computation of derived volumes is fast enough to be performed on demand. All other operations are fully interactive including interacting with the volume renderer hierarchy, the transfer function editors, and the parallel coordinates.

Many ensembles describing real-world phenomena, such as the Red Sea dataset [Hot], contain time-varying data. Supporting the analysis of such ensembles, for example, by providing a similarity measure for the hierarchical clustering that incorporates time, could be a subject of future research. Also, although the analysis of multi-field ensembles is supported, similarity is calculated using only the currently specified field. It may be beneficial to use a multi-field similarity measure [FL19] instead.

References

[AD54] ANDERSON, T. W. and DARLING, D. A. "A Test of Goodness of Fit". *Journal of the American Statistical Association* 49.268 (1954), 765–769. DOI: 10.1080/01621459.1954.10501232 4.

[AMS*21] ATHAWALE, TUSHAR M., MA, BO, SAKHAEE, ELHAM, et al. "Direct Volume Rendering with Nonparametric Models of Uncertainty". *IEEE TVCG* 27.2 (2021), 1797–1807. DOI: 10.1109/TVCG.2020.3030394 2.

[ASJ21] ATHAWALE, TUSHAR M., SANE, SUDHANSHU, and JOHNSON, CHRIS R. "Uncertainty Visualization of the Marching Squares and Marching Cubes Topology Cases". *2021 IEEE VIS*. 2021, 106–110. DOI: 10.1109/VIS49827.2021.9623267 2.

[DCH88] DREBIN, ROBERT A., CARPENTER, LOREN, and HANRAHAN, PAT. "Volume rendering". *SIGGRAPH Comput. Graph.* 22.4 (June 1988), 65–74. ISSN: 0097-8930. DOI: 10.1145/378456.378484. URL: https://doi.org/10.1145/378456.378484 1.

[DDW14] DEMIR, ISMAIL, DICK, CHRISTIAN, and WESTERMANN, RÜDIGER. "Multi-Charts for Comparative 3D Ensemble Visualization". *IEEE TVCG* 20.12 (2014), 2694–2703. DOI: 10.1109/TVCG.2014.2346448 2.

[DLL11] DOBREV, PETAR, LONG, TRAN VAN, and LINSEN, LARS. "A Cluster Hierarchy-based Volume Rendering Approach for Interactive Visual Exploration of Multi-variate Volume Data". *Vision, Modeling, and Visualization (2011)*. The Eurographics Association, 2011. ISBN: 978-3-905673-85-2. DOI: /10.2312/PE/VMV/VMV11/137-144 2.

[EL22] EVERS, MARINA and LINSEN, LARS. "Multi-dimensional parameter-space partitioning of spatio-temporal simulation ensembles". *Computers & Graphics* 104 (2022), 140–151. ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2022.04.005 2.

[FKRW17] FERSTL, FLORIAN, KANZLER, MATHIAS, RAUTENHAUS, MARC, and WESTERMANN, RÜDIGER. "Time-Hierarchical Clustering and Visualization of Weather Forecast Ensembles". *IEEE TVCG* 23.1 (2017), 831–840. DOI: 10.1109/TVCG.2016.2598868 2.

[FL19] FOFONOV, A. and LINSEN, L. "Projected Field Similarity for Comparative Visualization of Multi-Run Multi-Field Time-Varying Spatial Data". *CG Forum* 38.1 (2019), 286–299. DOI: https://doi.org/10.1111/cgf.13531 2, 7.

[GSWS21] GILLMANN, CHRISTINA, SAUR, DOROTHEE, WISCHGOLL, THOMAS, and SCHEUERMANN, GERIK. "Uncertainty-aware Visualization in Medical Imaging - A Survey". *CG Forum* 40.3 (2021), 665–689. DOI: https://doi.org/10.1111/cgf.14333 1.

[GXY12] GUO, HANQI, XIAO, HE, and YUAN, XIAORU. "Scalable Multivariate Volume Visualization and Analysis Based on Dimension Projection and Parallel Coordinates". *IEEE TVCG* 18.9 (2012), 1397–1410. DOI: 10.1109/TVCG.2012.80 2.

[HEG*22] HEIMES, KARL, EVERS, MARINA, GERRITS, TIM, et al. "Studying the Effect of Tissue Properties on Radiofrequency Ablation by Visual Simulation Ensemble Analysis". *Eurographics VCBM*. The Eurographics Association, 2022. ISBN: 978-3-03868-177-9. DOI: 10.2312/vcbm.20221187 5.

[Hot] HOTEIT, IBRAHIM. *Red Sea Data Courtesy of Red Sea Modeling and Prediction Group (PI Prof. Ibrahim Hoteit), KAUST*. https://kaust-vislab.github.io/SciVis2020/data.html. accessed June 19, 2024 5, 7.

[HTWL18] HE, XIANGYANG, TAO, YUBO, WANG, QIRUI, and LIN, HAI. "Biclusters Based Visual Exploration of Multivariate Scientific Data". *2018 IEEE Scientific Visualization Conference (SciVis)*. 2018, 77–81. DOI: 10.1109/SciVis.2018.8823605 2.

[ID90] INSELBERG, A. and DIMSDALE, B. "Parallel coordinates: a tool for visualizing multi-dimensional geometry". *Proceedings of the First IEEE Conference on Visualization: Visualization '90*. 1990, 361–378. DOI: 10.1109/VISUAL.1990.146402 1, 3.

[KBL19] KAPPE, C.P., BÖTTINGER, M., and LEITTE, H. "Analysis of Decadal Climate Predictions with User-guided Hierarchical Ensemble Clustering". *CG Forum* 38.3 (2019), 505–515. DOI: https://doi.org/10.1111/cgf.13706 2.

[KKH02] KNISS, J., KINDLMANN, G., and HANSEN, C. "Multidimensional transfer functions for interactive volume rendering". *IEEE TVCG* 8.3 (2002), 270–285. DOI: 10.1109/TVCG.2002.1021579 2.

[LLBP12] LIU, SHUSEN, LEVINE, JOSHUA A., BREMER, PEER-TIMO, and PASCUCCI, VALERIO. "Gaussian mixture model based volume visualization". *IEEE Symposium on Large Data Analysis and Visualization (LDAV)*. 2012, 73–77. DOI: 10.1109/LDAV.2012.6378978 2.

[LVR09] LINSEN, LARS, VAN LONG, TRAN, and ROSENTHAL, PAUL. "Linking Multidimensional Feature Space Cluster Visualization to Multifield Surface Extraction". *IEEE CG&A* 29.3 (2009), 85–89. DOI: 10.1109/MCG.2009.45 2.

[LVRR08] LINSEN, LARS, VAN LONG, TRAN, ROSENTHAL, PAUL, and ROSSWOG, STEPHAN. "Surface Extraction from Multi-field Particle Volume Data Using Multi-dimensional Cluster Visualization". *IEEE TVCG* 14.6 (2008), 1483–1490. DOI: 10.1109/TVCG.2008.167 2.

[ML18] MOLCHANOV, VLADIMIR and LINSEN, LARS. "Upsampling for Improved Multidimensional Attribute Space Clustering of Multifield Data". *Information* 9.7 (2018). ISSN: 2078-2489. DOI: 10.3390/info9070156 2.

[MY11] MOHD RAZALI, NORNADIAH and YAP, BEE. "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests". *J. Stat. Model. Analytics* 2 (Jan. 2011) 4.

[PWH11] PÖTHKOW, KAI, WEBER, BRITTA, and HEGE, HANS-CHRISTIAN. "Probabilistic Marching Cubes". *CG Forum* 30.3 (2011), 931–940. DOI: https://doi.org/10.1111/j.1467-8659.2011.01942.x 2.

[RFA*22] RAVE, H., FINCKE, J., AVERKAMP, S., et al. "Multifaceted Visual Analysis of Oceanographic Simulation Ensemble Data". *IEEE CG&A* 42.04 (July 2022), 80–88. ISSN: 1558-1756. DOI: 10.1109/MCG.2021.3098096 2.

[RLBS03] RHODES, PHILIP J., LARAMEE, ROBERT S., BERGERON, R. DANIEL, and SPARR, TED M. "Uncertainty Visualization Methods in Isosurface Rendering". *Eurographics 2003 - Short Presentations*. Eurographics Association, 2003. DOI: /10.2312/egs.20031054 2.

[RPHL14] RISTOVSKI, GORDAN, PREUSSER, TOBIAS, HAHN, HORST K., and LINSEN, LARS. "Uncertainty in medical visualization: Towards a taxonomy". *Computers & Graphics* 39 (2014), 60–73. ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2013.10.015 1.

[SAJ21] SANE, SUDHANSHU, ATHAWALE, TUSHAR M., and JOHNSON, CHRIS R. "Visualization of Uncertain Multivariate Data via Feature Confidence Level-Sets". *EuroVis 2021 - Short Papers*. The Eurographics Association, 2021. ISBN: 978-3-03868-143-4. DOI: 10.2312/evs.20211053 2.

[SD10] SCHUMACHER, RUSS S. and DAVIS, CHRISTOPHER A. "Ensemble-Based Forecast Uncertainty Analysis of Diverse Heavy Rainfall Events". *Weather and Forecasting* 25.4 (2010), 1103–1122. DOI: 10.1175/2010WAF2222378.1 1.

[SE17] SAKHAEE, ELHAM and ENTEZARI, ALIREZA. "A Statistical Direct Volume Rendering Framework for Visualization of Uncertain Data". *IEEE TVCG* 23.12 (2017), 2509–2520. DOI: 10.1109/TVCG.2016.2637333 2.

[SW65] SHAPIRO, S. S. and WILK, M. B. "An analysis of variance test for normality (complete samples)". *Biometrika* 52.3-4 (Dec. 1965), 591–611. ISSN: 0006-3444. DOI: 10.1093/biomet/52.3-4.591 2.

[TZG*17] TOYE, HABIB, ZHAN, PENG, GOPALAKRISHNAN, GANESH, et al. "Ensemble Data Assimilation in the Red Sea: Sensitivity to Ensemble Selection and Atmospheric Forcing". *Ocean Dynamics* 67.7 (2017), 915–933. DOI: 10.1007/s10236-017-1064-1 1.

[WFG*19] WEISSENBÖCK, JOHANNES, FRÖHLER, BERNHARD, GRÖLLER, EDUARD, et al. "Dynamic Volume Lines: Visual Comparison of 3D Volumes through Space-filling Curves". *IEEE TVCG* 25.1 (2019), 1040–1049. DOI: 10.1109/TVCG.2018.2864510 2.

[WHLS19] WANG, JUNPENG, HAZARIKA, SUBHASHIS, LI, CHENG, and SHEN, HAN-WEI. "Visualization and Visual Analysis of Ensemble Data: A Survey". *IEEE TVCG* 25.9 (2019), 2853–2872. DOI: 10.1109/TVCG.2018.2853721 2.

[WSJ*12] WONG, PAK CHUNG, SHEN, HAN-WEI, JOHNSON, CHRISTOPHER R., et al. "The Top 10 Challenges in Extreme-Scale Visual Analytics". *IEEE CG&A* 32.4 (2012), 63–67. DOI: 10.1109/MCG.2012.87 1.

[ZWK10] ZEHNER, BJÖRN, WATANABE, NORIHIRO, and KOLDITZ, OLAF. "Visualization of gridded scalar data with uncertainty in geosciences". *Comput. Geosci.* 36.10 (Oct. 2010), 1268–1275. ISSN: 0098-3004. DOI: 10.1016/j.cageo.2010.02.010 2.