Visual Analysis of Uncertainties in Ocean Forecasts for Planning and Operation of Off-Shore Structures

Thomas Höllt*
King Abdullah University
of Science and Technology

Ahmed Magdy*
King Abdullah University
of Science and Technology

Guoning Chen†
University of Houston

Ganesh Gopalakrishnan‡
University of California, San Diego

Ibrahim Hoteit*
King Abdullah University
of Science and Technology

Charles D. Hansen§
SCI Institute and School of Computing
University of Utah

Markus Hadwiger*
King Abdullah University
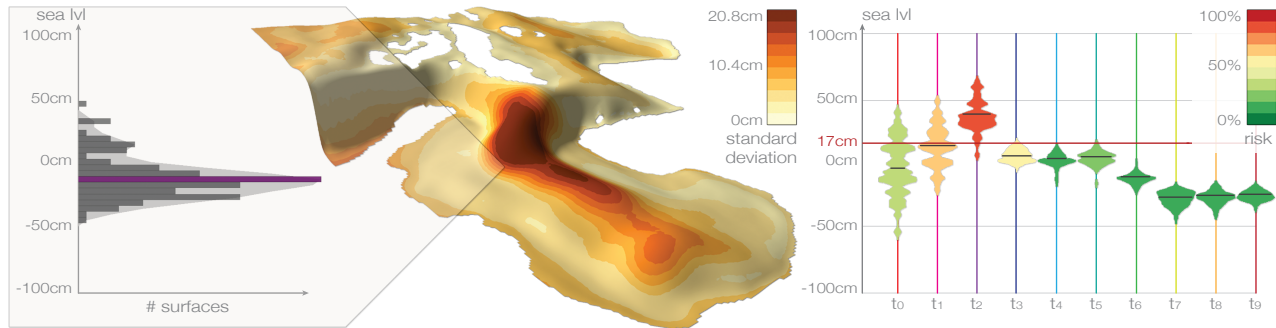of Science and Technology

Figure 1: A selected time step of an ocean forecasting ensemble computed for the Gulf of Mexico is visualized via the mean surface of the ensemble (center). For a more detailed inspection of the entire distribution of the surfaces comprising the ensemble, we provide two linked views. The first linked view (left) shows a histogram of depth positions of the surfaces at a selected spatial position and time step. The second linked view (right) is a time-series view that depicts a glyph for each time step at the selected position. The horizontal line corresponds to a chosen critical sea level, where each glyph's color depicts the risk corresponding to how much of the distribution is above that critical level.

ABSTRACT

We present a novel integrated visualization system that enables interactive visual analysis of ensemble simulations used in ocean forecasting, i.e, simulations of sea surface elevation. Our system enables the interactive planning of both the placement and operation of off-shore structures. We illustrate this using a real-world simulation of the Gulf of Mexico. Off-shore structures, such as those used for oil exploration, are vulnerable to hazards caused by strong loop currents. The oil and gas industry therefore relies on accurate ocean forecasting systems for planning their operations. Nowadays, these forecasts are based on multiple spatio-temporal simulations resulting in multidimensional, multivariate and multi-valued data, so-called ensemble data. Changes in sea surface elevation are a good indicator for the movement of loop current eddies, and our visualization approach enables their interactive exploration and analysis. We enable analysis of the spatial domain, for planning the placement of structures, as well as detailed exploration of the temporal evolution at any chosen position, for the prediction of critical ocean states that require the shutdown of rig operations.

Keywords: Uncertainty, Ensemble Simulation, Risk Estimate.

Index Terms: I.3.8 [Computing Methodologies]: Computer Graphics—Applications; J.2 [Computer Applications]: Physical Sciences and Engineering—Earth and Atmospheric Sciences

---

*e-mail:{thomas.hollt|ahmed.magdy|ibrahim.hoteit|markus.hadwiger}
@kaust.edu.sa
†e-mail:chengu@cs.uh.edu
‡e-mail:ggopalak@ucsd.edu
§e-mail:hansen@cs.utah.edu

1 INTRODUCTION

Oil exploration in the deep Gulf of Mexico is vulnerable to hazards due to strong currents at the fronts of highly non-linear warm-core eddies [38]. The dynamics in the Gulf of Mexico are indeed dominated by the powerful northward Yucatan Current flowing into a semi-enclosed basin. This current forms a loop, called the Loop Current, that exits through the Florida Straits, and in turn merges with the Gulf Stream. At irregular intervals, the loop current sheds large eddies that propagate westward across the Gulf of Mexico. This eddy shedding involves a rapid growth of non-linear instabilities [3], and the occasional eddy detachment and reattachment make it very difficult to clearly define, identify, monitor, and forecast an eddy shedding event [2, 4, 11].

The predictability of loop current shedding events in the Gulf of Mexico poses a major challenge for the oil and gas industry operating in the Gulf. The presence of these strong loop currents potentially causes serious problems and safety concerns for the rig operators. Millions of dollars are lost every year due to drilling downtime caused by these powerful currents. As oil production moves further into deeper waters, the costs related to strong current hazards are increasing accordingly, and accurate 3D forecasts of currents are needed. These can help rig operators to avoid some of these losses through better planning, and avoid potentially dangerous scenarios. A 3D ocean forecasting system for the Gulf of Mexico therefore becomes crucial and highly desired by the oil and gas industry, where accurate loop current forecasts over a time frame of one to two weeks provide a reasonable time window for planning the drilling operations.

Developing efficient tools to visualize and clearly disseminate forecast outputs and results is becoming a very important part of the forecasting process. Such tools have to be conceived in a way that allows users to easily extract and clearly identify the necessary information from large ensembles and the associated statistics representing the forecast and its uncertainties. In this paper, we present
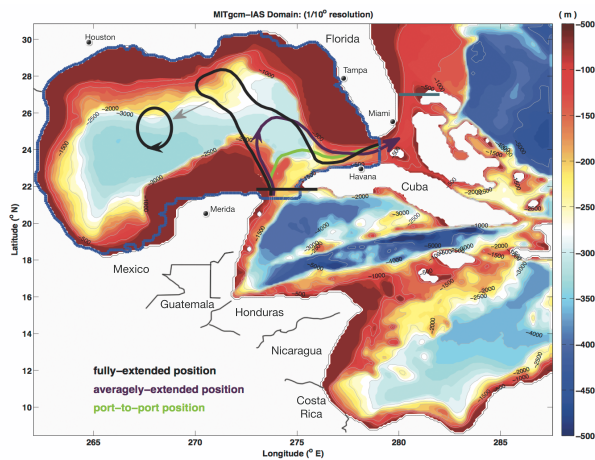
Figure 2: **The Gulf of Mexico Simulation Area** covered by the presented dataset. The colors denote water depth in meters [11].

the first system for the visual exploration and analysis of these kinds of forecasts. Our system handles multivalued ensembles of height-fields comprising multiple time steps. A set of statistical properties is derived from the ensemble and can be explored in multiple linked views, while the complete ensemble is always available for detailed inspection on demand. Our system enables domain experts to efficiently plan the placement and operation of off-shore structures, such as oil platforms.

1.1 Ocean Forecast Simulation

The development of a reliable ocean forecasting system requires models capable of simulating ocean circulation and an efficient assimilation scheme that, given enough observations, provides accurate initial conditions for forecasting. High-resolution 3D general circulation ocean models are necessary to reproduce complex mesoscale dynamics like in the Gulf of Mexico [2]. However, such models cannot provide accurate forecasts of mesoscale variability, such as eddy shedding events, without data assimilation. A general circulation ocean model is subject to several sources of uncertainties, not only from the poorly known inputs such as the initial state, and atmospheric and lateral boundary conditions, but also from the use of approximate parameterization schemes of sub-grid physics and ocean mixing dynamics. Data assimilation methods address this issue by constraining model outputs with incoming data.

The important role of uncertainties is now increasingly recognized in the ocean prediction community for proper decision making and efficient risk management.

New assimilation methods based on Bayesian filtering theory have been recently developed by the ocean and atmospheric communities for efficient propagation and quantification of uncertainties [6, 27, 12, 13, 11]. These methods, known as ensemble Kalman filter methods, follow a Monte Carlo approach to represent the uncertainties on a state estimate by an ensemble of model states. These are then integrated forward in time with the general circulation ocean model to quantify uncertainties in the forecast. The estimated forecast uncertainties are then combined with the observation uncertainties to assimilate the new incoming data using a Kalman filter correction step [6], before a new forecast cycle begins. Developing and implementing efficient ensemble Kalman filters with state-of-the-art ocean and atmospheric models is a very active area of research.

With the fast-growing high performance computing resources, the implementation of ensemble Kalman filters with large ensemble members is now practically feasible using highly sophisticated general circulation ocean models. When a filter's ensemble is available, it is customary to calculate various statistical measures of the ensemble spread as indicators of the uncertainties and of their evolution in space and time, which are then used in decision making.

Recently, Hoteit et al. [11] developed an ensemble forecasting system for the Gulf of Mexico circulation based on the Massachusetts Institute of Technology General Circulation Model (MITgcm) [20], and the Data Assimilation Research Testbed (DART) [13]. This system is capable of assimilating various sets of satellite and in-situ ocean observations. We use this system as a real-world scenario that illustrates the new capabilities for analysis and exploration provided by our visualization approach. Figure 2 gives an overview of the area covered by the forecasting system.

1.2 Visualization Contributions

We present a GPU-based interactive visualization system for the exploration and analysis of ensemble heightfield data, with a focus on the specific requirements of ocean forecasts. Based on an efficient GPU pipeline we perform on-the-fly statistical analysis of the ensemble data, allowing interactive parameter exploration. We present a novel workflow for planning the placement and operation of off-shore structures needed by the oil and gas industry. While we focus on the visualization and analysis of ocean forecast data, the presented approach could also be used for the exploration of heightfield ensembles from other areas, such as weather forecasting or climate simulation.

2 RELATED WORK

Uncertainty and ensemble visualization are widely recognized as important topics in the field of visualization, which has resulted in a large body of related work in recent years. In the following overview, we restrict ourselves to key publications in uncertainty and ensemble visualization, as well as selected publications from other areas related to the techniques presented in this paper.

Uncertainty Visualization. A good introduction to uncertainty visualization is provided by Pang et al. [24], who present a detailed classification of uncertainty, as well as numerous visualization techniques. Johnson and Sanderson [14] give a good overview of uncertainty visualization techniques for 2D and 3D scientific visualization, including uncertainty in surfaces. For a definition of the basic concepts of uncertainty and another overview of visualization techniques we refer to Griethe and Schumann [7]. Riveiro [34] provides an evaluation of different uncertainty visualization techniques for information fusion. Rhodes et al. [33] present the use of color and texture to visualize uncertainty of iso-surfaces. Brown [1] employs animation for the same task. Grigoryan and Rheingans [8] present a combination of surface and point based rendering to visualize uncertainty in tumor growth. There, uncertainty information is provided by rendering point clouds in areas of large uncertainty, as opposed to crisp surfaces in certain areas.

Recently, Pöthkow et al. [29, 30] as well as Pfaffelmoser et al. [25] presented techniques to extract and visualize uncertainty in probabilistic iso-surfaces. Pfaffelmoser and Westermann [26] describe a technique for the visualization of correlation structures in uncertain 2D scalar fields. They use spatial clustering based on the degree of dependency of a random variable and its neighborhood.

A system which models and visualizes uncertainty in segmentation data based on a priori shape and appearance knowledge has been presented by Saad et al. [35].

Ensemble Visualization. Early work on visualization of ensemble data was conducted by Pang, Kao and colleagues [15, 19, 16, 18]. While the authors did not use the term ensemble, these works deal with the visualization of what they call spatial distribution data, which they define as a collection of n values for a single variable in m dimensions. These are essentially ensemble data. The authors adapt standard visualization techniques to visualize these data gathered from various sensors, e.g. satellite imaging or mutli-return Lidar. Frameworks for visualization of ensemble data gained
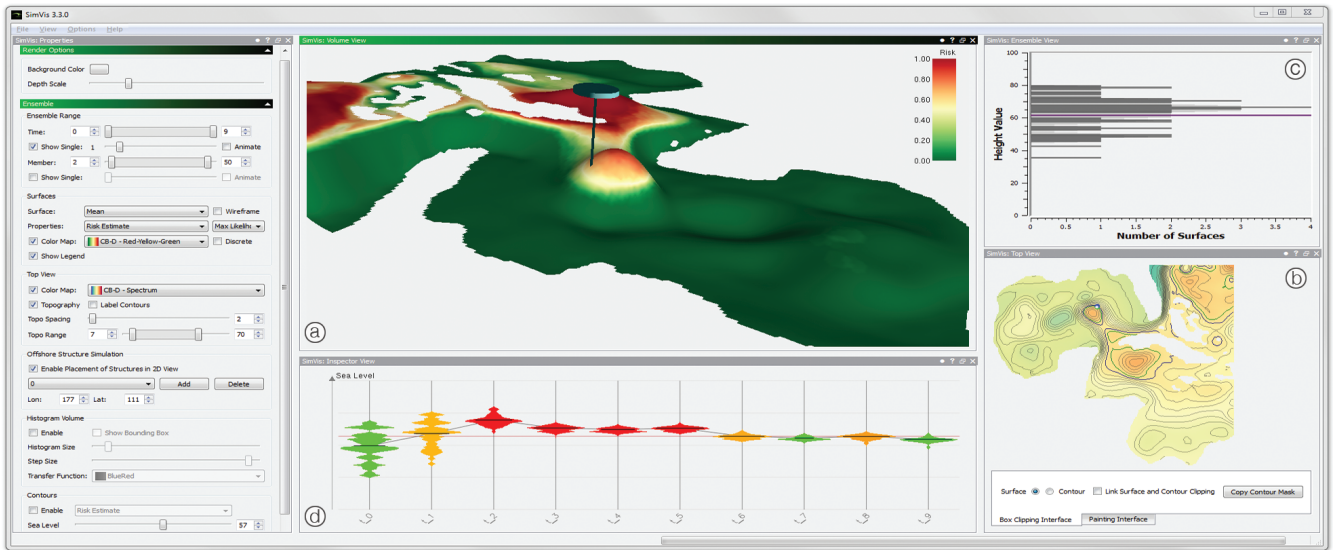
Figure 3: **Application Overview.** Our application for exploration of ocean forecast ensembles consists of four main views. The simulated ocean surface, or a derived version, like the mean surface for a time step, can be shown in 3D or 2D (ⓐ and ⓑ). The histogram view (ⓒ) shows the complete distribution of the ensemble at a selected position, while the time-series view (ⓓ) shows the distribution and the resulting operational risk at a selected position for multiple time steps.

from weather simulations include *Ensemble-Vis* by Potter et al. [32] and *Noodles* by Sanyal et al. [36]. These papers describe fully featured applications focused on the specific needs for analyzing weather simulation data. They implement multiple linked views to visualize a complete set of multidimensional, multivariate and multivalued ensemble members. While these frameworks provide tools for visualizing complete simulation ensembles including multiple dimensions, to solve the problem presented in this work we focus on 2.5D surface, i.e. heightfield ensemble data.

Matković et al. [21] present a framework for visual analysis of families of surfaces by projecting the surface data into lower dimensional spaces. Piringer et al. [28] describe a system for comparative analysis of 2D function ensembles used in the development process of powertrain systems. Their design focuses on comparison of 2D functions at multiple levels of detail. Healey and Snoeyink [9] present a similar approach for visualizing error in terrain representation. There, the error, which can be introduced by sensors, data processing or data representation, is modeled as the difference between the active model and a given ground truth.

Several published extensions of box plots have inspired our time-series view. Hintze and Nelson [10] introduce violin plots to give an indication of the distribution using the sides of the box. Esty and Banfield [5] combine box and percentile plots to add the complete distribution to the plot while keeping the simplicity of box plots. Potter et al. [31] combine quartile, moment and density plots, based on the histogram, to create summary plots. The density of curves in 1D function plots can be visualized effectively using kernel density estimation [17]. Our histogram view that shows the distribution of surfaces embedded in 3D passing through each (x, y) position is similar in spirit to such approaches, but for primitives of one dimension higher.

3 VISUAL OCEAN FORECAST EXPLORATION

Our system targets the interpretation of forecasts from the planning phase of an off-shore structure to its operation. Since the different phases have different requirements, we provide a set of four main views, which are used in different combinations depending on the application scenario. Figure 3 shows our application with the main views plus a unified settings panel. The views are two spatial views showing the surface data themselves, one in 3D ⓐ, the other one in 2D ⓑ, a linked histogram view ⓒ as well as a time-series view ⓓ.

While the accessibility of an existing reservoir is the key factor when planning an oil platform, ocean forecasts can provide valuable additional information. Modern drilling techniques to some extent allow flexible paths and thus considerable flexibility for the actual placement of a platform. However, the complexity of the path has implications on the cost of drilling. On the other hand, slight changes of the position might move a platform from an area that is strongly affected by eddy shedding, which leads to long downtimes, to a less affected area, overall resulting in more efficient operations. In the planning phase, the interaction mainly happens in the two spatial views.

2D View. The simple 2D top-down view shown in Figure 3ⓑ is a common tool for visualizing heightfield data and familiar to domain scientists. In the standard setting, this view provides a general overview. It shows the mean surface of a specific time step, using iso contours and pseudo-coloring for the heightfield values. In the first step, domain scientists can use this view to select a region of interest. Regions that are not suitable for placement in the first place, i.e., regions from which no reservoir can be reached, can be marked in this view. For this task, we provide both a simple rectangular clipping interface, as well as allowing the user to paint a mask directly inside the view. The latter enables arbitrary free-form selections. Once the region of interest is defined, a first overview of the data can be gained using the 2D view. In addition to the mean heightfield, any 2D scalar field resulting from a statistical analysis can also be visualized in this view. In particular, our domain scientists are interested in two main properties: the variance, and a simple risk estimate (see below). Both pseudo-coloring and iso contours can be used to visualize any of these properties. However, one is usually reserved for the mean sea level surface in order to provide context. Once the area of interest is defined and the user has a general overview of the data, the main task is to find a position at which downtimes of the platform, caused by dangerous ocean conditions, will be minimal. Here, these conditions are mostly defined by the sea level. We allow users to specify a critical height value, whose iso contour derived from the currently selected surface is then highlighted. Pseudo-coloring the variance can give an idea of the uncertainty of the contour. To provide insight on spatial variations, all values can be modified interactively. Often, slight variations of the critical height result in large variations of the actually affected area. By simultaneously showing the iso con-
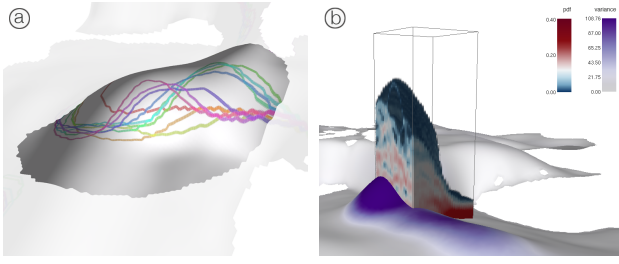
Figure 4: **3D View Detail.** ⓐ shows iso contours of the mean surfaces for all time steps blended over the current surface. The area of interest is rendered with full opacity, while the context is preserved by rendering the remaining parts semi-transparently. ⓑ shows a volume rendering of the pdf at a user-selected position. The surface is color-mapped with the variance. The large spread in areas of high variance is clearly visible in the volume rendering.
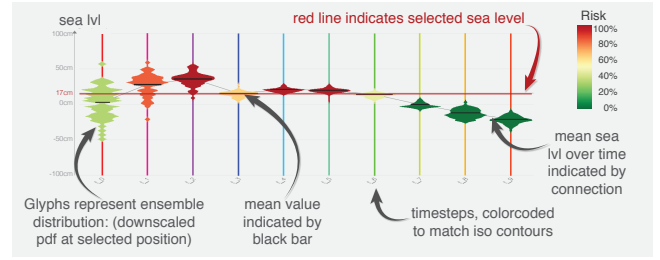

Figure 5: The **Time-Series View** in detail. The y axis corresponds to the sea level, the x axis to time. Each glyph shows the distribution of the selected position for one time step. Glyphs are colored using the associated risk, i.e., the fraction of the distribution above the selected critical sea level. Each time step corresponds to a unique color, as indicated by the vertical lines, which is also used for rendering the iso contours of multiple time steps (Figure 4ⓐ).

tours for the defined height of the mean surfaces of all time steps at once, experts can easily identify the most affected areas. However, uncertainty information is hard to include in this case. On the one hand, color-coding the surface allows showing only one time step at a time. On the other hand, a rubber-band approach for several contours results in a large amount of clutter. We avoid this by visualizing a risk estimate.

Risk Estimate. We define and visualize a simple risk estimate as the percentage of ensemble members above the defined critical height. This value is computed for every (x,y)-position for every time step. The surface in Figure 3ⓐ is colored with such a risk estimate. We can use color-coding only for a single time step, but by using the iso contour for an acceptable risk (e.g. 10%), the iso contours for all time steps can be overlaid, as described above. The user can also interactively modify both parameters, the critical height and the acceptable risk, to iteratively find an acceptable compromise and define a set of possible positions for further inspection.

3D View. The linked 3D view (Figure 3ⓐ) provides all the features described for the 2D view plus several additional tools for a more detailed spatial and temporal inspection. Typically, in the 3D view the height values of the displayed surface are mapped to the third dimension, freeing pseudo-coloring and iso contours for additional information. Figure 4ⓐ shows an example, where the surface is pseudo-colored using the risk estimate, and iso contours corresponding to the critical sea level for all time steps are shown as well. An additional benefit of the 3D view is that it is possible to use volume rendering for showing details of the distribution of the ensemble. Similar to approaches presented by Pöthkow et al. [29, 30], as well as Pfaffelmoser et al. [25], we depict the actual distribution of the ensemble as a volume around the surface. Instead of using a parametric representation of the data based on mean and variance, we allow rendering the full probability density function (pdf) of the distribution, to allow detailed inspection of the actual data. However, since at this point the user usually has picked a set of points of interest, to avoid unnecessary occlusion, we do not render the complete volume, but a small subset of adjustable size, which essentially works like a volumetric cursor (see Figure 4ⓑ). The user can simply probe the data by hovering with the mouse over a position of interest, and the probability density volume is then rendered around the picked position.

Histogram View. Another way to inspect the distribution in detail is the histogram view shown in Figure 3ⓒ. This view shows the histogram over the values of the heightfield as well as the probability density function (pdf) for a selected (x,y)-position. Similar to the volumetric representation in the 3D view, the position is defined by picking directly in any of the spatial views. When the user moves the mouse over the surface, the histogram view is updated on the fly to show the histogram at the current mouse position.

At this point, after defining the area of interest, narrowing down using the risk estimate, and finally inspecting the distribution in detail at a small number of positions, the user can use the time-series view to inspect the data at the selected (x,y)-positions over all time steps. This last step in the planning phase is very similar to the planning in the operational phase, with the distinction that the position is not yet defined. Positions of interest can either be defined interactively as described above, or loaded from file.

Time-Series View. This view always shows the complete time-series for a single position. Once defined, or loaded, our application caches several positions of off-shore structures, from which one can be selected using a drop-down menu. Using the time-series view (Figures 3ⓓ and 5), domain experts can then easily identify the points in time where operations should be halted. For each time step, a glyph similar to a violin plot [10] is displayed. Here the probability density function (pdf) is used for the outline and the mean value is indicated by a horizontal line at the appropriate position. To provide spatial context, the plots are arranged at their original depth positions and as such can be compared directly. In addition, the critical height is indicated by a horizontal line. To help identify the critical time steps, the glyphs are color-coded according to the risk estimate. Without ensemble forecasts, rigs were operated based on a single simulation. However, even when ensemble forecasts are available, a visual exploration approach is necessary. Without our visualization system, our domain experts would define the safeness by simply looking at the mean and variance values of the ensemble. In each time step, we indicate the mean value of the distribution by a bold black bar, providing the same information as before, for all time steps in a single view. The user can immediately identify critical time steps, looking at the color and position of each glyph, and possibly order an unavoidable shutdown of operations. If the situation is unclear, the glyph provides the complete distribution to enable the expert to make a decision.

4 ANALYSIS AND VISUALIZATION PIPELINE

For efficient exploration and analysis of the data an elaborate statistical analysis as described in Section 4.1 is performed. To allow interactive updates, we have implemented a GPU-based analysis and visualization pipeline presented in Section 4.2.

4.1 Statistical Analysis

The basis for visual statistical analysis of the input data is a 3D spatial distribution histogram. We define the axes of this histogram such that the x and y axes correspond to the domain of the heightfield, and the z-axis corresponds to its values. This results in a volume with the same x and y extents as the input surface data, and the z extent depending on the range and sampling of the image of the input function. The histogram shows the number of surfaces passing through the (x,y,z)-position, corresponding to each bin.
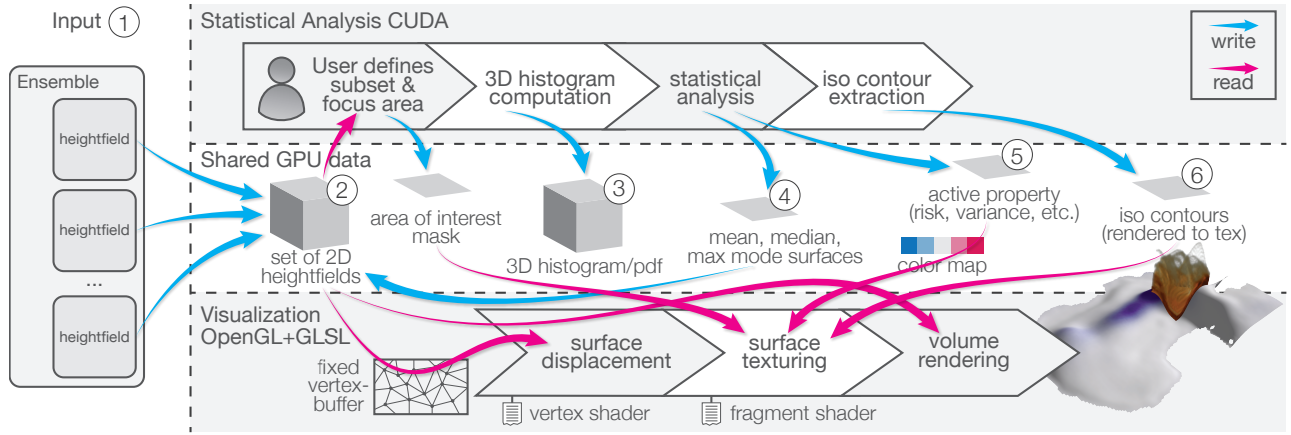
Figure 6: Pipeline Overview. The pipeline is divided into two major blocks: The statistical analysis part at the top, and the rendering part shown at the bottom. Both parts are entirely GPU-based, and all data (middle) are shared by both parts in GPU memory.

Since there is only one value per (x,y)-position per heightfield we can derive two important properties; First, we can interpret the 3D histogram as a set of 1D histograms, one for each (x,y)-position. This means that the statistical analysis can be carried out for each (x,y)-position separately. We use this to parallelize the computation as shown in Section 4.2. Second, each of these 1D histograms can be interpreted directly as a probability distribution of the surfaces at the corresponding (x,y)-position by normalizing the value of each bin by dividing by the total number of surfaces. In addition to this simple probability measure, we also compute a kernel density estimation to approximate the continuous probability density function at each (x,y)-position. These 1D histograms and probability density functions are used for the glyphs in the time series view and also for the histogram view. The complete 3D volume is the basis for the volume rendering depicted in the 3D view.

From the 1D histograms, a number of statistical properties including range, mean, median, maximum mode, standard deviation, variance, skewness, kurtosis and the risk estimate described in Section 3 are computed for each (x,y)-position. While mean, median and maximum mode are added to the ensemble as surfaces, the other properties are added as meta information, for example to color-code the surfaces accordingly.

Additionally, if enabled, iso contours are extracted from selected properties, most importantly from the active surface (mean, median or maximum mode), as well as the risk estimate for all time steps.

To explore the parameter space, for example to look at the influence of a certain parameter or to remove outliers, the statistical analysis, as well as iso contour extraction, can be carried out either for the complete ensemble, or for any user-defined subset of the ensemble. In the standard setting for oceanography, time is mapped to one parameter, while the simulation starting condition is mapped to a second one. In addition, the time parameter's range is set to a single time-step, while the second parameter's range covers all starting conditions. This results in the statistical analysis being carried out for a single time step only. However, both parameters can be adjusted causing the analysis to be carried out for the adjusted range on the fly.

To allow interactive exploration of the parameter space, all updates of the statistical analysis must be computed in real time, or at least at interactive rates. For this reason, we employ a pipeline that is entirely GPU-based, which is presented in the next section.

4.2 GPU-Based Analysis and Visualization Pipeline

Our GPU-based analysis and visualization pipeline is illustrated in Figure 6. In the remainder of this section, circled numbers refer to this figure. The pipeline is divided into two main parts: The statistical analysis and iso surface extraction is carried out using CUDA, while the visualization is based on OpenGL and GLSL shaders. All data are shared between the two parts of the pipeline, so that after the initial upload of the ensemble onto the GPU no expensive bus transfer is necessary. Since usually only a small part of the ensemble is required by the visualization, a streaming approach would be possible for datasets that are larger than GPU memory, but we currently assume that the dataset fits into GPU memory.

Input. The input ① to our system is a set of heightfields. These can be part of a simulation ensemble, e.g. from ocean or weather forecasts, a time series of some sort, or the results of a parameterized segmentation. Even though we focus on heightfields in this work, the concepts can also be applied to surfaces in n dimensions as long as the correspondences between all surfaces in the dataset are known for every nD-datapoint. In our framework, we assume the 2D spatial (x,y)-coordinate to be the correspondence between the surfaces.

Data Representation. Before computation of statistics or visualization, the ensemble is converted into a 3D texture ② and loaded onto the GPU. Every heightfield of the ensemble will be represented by one slice in this texture. Additionally, space for the mean, median and maximum mode heightfield will also be reserved in this texture. The surfaces are indexed using the original parametrization. If there is only a single parameter, for example the time steps in a time series, the surface ID corresponds to the texture index. For higher-dimensional parameter spaces, e.g. ensemble ID plus time, the linear texture index is computed from the original parameters. This allows the user to define subranges for each parameter separately, for example to examine the complete ensemble at a single time step.

Statistical Analysis. The first step in the statistical analysis is the creation of the 3D histogram ③. Changes in the parameter range trigger an update of the 3D histogram and subsequently of the representative surface and property texture. Since each ensemble member provides exactly one entry to the histogram per (x,y)-position, rather than using a thread for each member, we use one thread per (x,y)-position. Each thread then loops over all selected surfaces and inserts the corresponding height values into the histogram. This way, write conflicts can be avoided and no critical sections or atomic operations are needed. The kernels for the derived properties are set up in a similar fashion. The desired statistical property is computed by one thread per (x,y)-position. The main difference to the histogram computation is that this results in a single scalar per thread, all of which are then assembled into a 2D texture. While mean, median and maximum mode ④ are attached to the 3D heightfield texture to be used as representative surfaces, the other properties ⑤ are copied into a 2D texture available to the visualization pipeline for texturing the surface. Exploiting

the parallelism of the GPU and eliminating costly bus transfers between CPU and GPU allows interactive modification of the parameter range even for ensembles containing several hundred surfaces. Section 4.3 provides a detailed performance analysis.

Iso Contouring. We have implemented marching squares using CUDA, based on the marching cubes example from the CUDA SDK. However, on our test system this implementation performs worse than a CPU version (compare Section 4.3). The CUDA version also requires considerable communication between the CPU and GPU meaning the advantage of a completely GPU based pipeline gets lost when extracting the iso contours. Hence we use the faster CPU version. We keep the initial geometric representation of the contours, for example for use in the 2D view, but for overlaying the contours onto the 3D surfaces we render the contours into an offscreen buffer ⑥, which is then used for texturing.

Rendering. The rendering pipeline takes advantage of the fact that all ensemble data are already stored in GPU memory, which facilitates efficient surface rendering. Instead of creating new surface geometry every time a different surface of the ensemble is rendered, a single generic vertex buffer of fixed size is created. This buffer covers the entire (x, y)-domain, but does not contain any height information. The z-value of each vertex is set later in the vertex shader. Before transforming the vertex coordinates into view space, the object space (x, y)-coordinates of the vertex in combination with the ID of the active surface are used to look up the z-value of the current vertex in the ensemble texture. At this point, the desired surface geometry is available. In order to be able to visualize the results of the statistical analysis, the object space coordinates are attached to each vertex as texture coordinates (x and y are sufficient). In the fragment shader, this information can then be used to look up the active statistical property in the 2D texture. This texture contains the raw information from the statistical analysis, which is then converted to the fragment color by a look up in a 1D color map. We provide a selection of several continuous, diverging cool-to-warm color maps, as presented by Moreland [22], but also allow the creation of custom color maps. These color maps minimally interfere with shading, which is very important in this case, as shading is an important feature to judge the shape of a surface. During testing we realized that using the continuous version made it very hard to relate an actual value to a color in the rendering so we decided to optionally provide a discrete version with ten steps. After the surface geometry has been rendered, a surrounding volume, for example the 3D probability density function, can be rendered as well. This is done in a second rendering pass in order to guarantee correct visibility [37].

Interaction. With the described pipeline in place, a number of features can be implemented very easily and efficiently. If desired, the user can choose to render any surface from the ensemble. This requires no data transfer to or from the GPU, except for the ID of the surface in the ensemble to render. In addition, it is possible to automatically animate all surfaces in a predefined range. In the presented application this can be useful in two ways; As shown by Brown [1] animation is a powerful tool for visualizing uncertainty. The user can choose to animate through all members of a single time step to get an impression of the surface distribution. Secondly animating the mean surfaces over the time domain can show the behavior of the loop currents.

The described visualization techniques can give a very good impression of the quantitative variation in the data. Detailed information on the surface distribution can be gained by animating through or manually selecting individual surfaces from the ensemble. However, it is hard to compare more than two surfaces this way. We therefore provide an additional view showing the histogram and probability distribution for a selected position. The position to investigate can be picked directly in the 3D view. All information that is required for picking is already available in our rendering pipeline:

We use the same vertex shader as described before for rendering the surface into an off-screen buffer of the same size as the frame buffer. Instead of using the object space coordinates to look up the scalar values in the fragment shader, we use the coordinates directly as the vertex color. This way, we can look up the current mouse position directly in the downloaded off-screen buffer. With the (x, y)-part of the resulting volume position, we can then directly look up the histogram and probability density distribution for this position. To facilitate easy comparison, we color the bin corresponding to the current representative surface differently than the remaining bins.

4.3 Performance

The performance of the statistical analysis is crucial for interactive exploration of the parameter space. We used the dataset described in Section 5 for a performance analysis. The dataset consists of a total of 500 surfaces spread over ten time steps. Since usually one time step is investigated at a time we compare performance for a single time step, consisting of 50 surfaces, as well as the complete dataset. Table 1 shows the resulting computation times.

The computations were performed using an NVIDIA GeForce GTX 580 with 1.5GB of graphics memory. The timings were averaged over 1000 kernel executions. As all data stays on the GPU, no bus transfer has to be considered. For comparison, we also show computation times of a single time step on the CPU. The computations were carried out on a workstation with two six-core Xeons (12 physical cores plus hyper threading) clocked at 3.33GHz and 48GB of main memory. The CPU computations were parallelized using OpenMP, utilizing 24 threads.

In general, it can be seen in Table 1 that using the GPU even for 500 surfaces, the slowest update including skewness and all dependencies plus the probability density function (which needs to be computed for the histogram and time series views) still allows for interactive update rates. Compared to the CPU version, we achieved a speedup of roughly $5\times$ for all tasks when considering the dependencies.

The histogram, range, mean, variance, kurtosis and the risk estimate are calculated directly from the ensemble and as such the complexity relies solely on the number of surfaces and valid data points per surface. We would expect the computation time for these values to scale linearly with the number of surfaces/valid data points, which seems to be in line with the measured numbers. For even larger datasets, however, it would make sense to compute range, mean, variance, kurtosis and the risk estimate using the histogram. This would result in constant time, only depending on the size of the histogram. For the datasets here, however, the histogram computation is the limiting factor. The probability density function, median and mode are looked up using the histogram, and therefore there is no difference between the small and the large data set. Standard deviation and skewness are implemented as linear combinations of other surface properties, and thus computation times are also independent of the number of surfaces. With the dependencies precomputed, the computation of both properties is trivial, which results in very short computation times.

5 APPLICATION SCENARIOS

We illustrate our approach using two different scenarios for a real-world Gulf of Mexico ocean forecast dataset. The dataset covers the Gulf of Mexico basin between $8.5°$ N and $31°$ N, and $262°$ E and $287.5°$ E on a $1/10° \times 1/10°$ grid with 40 vertical layers. Forecasting experiments were performed over a six-month period in 1999 between May and October during which a strong loop current event occurred (Eddy "Juggernaut") [23]. The resulting dataset consists of ten time steps, each consisting of 50 ensemble members. The lateral dimensions are represented by a grid consisting of 275×325 samples.

Table 1: **Computation times for all properties.** All times are in milliseconds. The first column shows ID and name of the property. The second column lists the IDs that are required to compute the corresponding property. The columns titled *w/o dep* and *w dep* show the computation for just this property, and the property plus all dependencies, respectivley. The last two columns show the speedup from CPU to GPU. For iso contour extraction (12) one contour per time step was extracted, resulting in one and ten contours respectively.

	Property	Depends on Property	50 Surfaces CUDA w/o dep	w dep	500 Surfaces CUDA w/o dep	w dep	50 Surfaces CPU w/o dep	w dep	CPU/GPU Speedup w/o dep	w dep
1	Histogram	-	3.23	3.23	38.56	38.56	19.24	19.24	6.0x	6.0x
2	PDF	1	12.93	16.16	12.78	51.34	45.70	64.94	3.5x	4.0x
3	Range	-	0.71	0.71	11.09	11.09	3.45	3.45	4.9x	4.9x
4	Mean	-	0.71	0.71	10.89	10.89	3.48	3.48	4.9x	4.9x
5	Median	1	0.70	3.93	0.70	39.26	8.78	28.02	12.5x	7.1x
6	Mode	1	1.40	4.63	1.41	39.97	4.65	23.89	3.3x	5.2x
7	Variance	4	0.72	1.43	10.87	21.76	3.85	7.33	5.3x	5.1x
8	Std Dev	4, 7	0.02	1.45	0.02	32.78	0.14	7.47	7.0x	5.2x
9	Skewness	1, 4, 6, 7, 8	0.05	6.13	0.05	72.80	0.16	31.42	3.2x	5.1x
10	Kurtosis	4, 7	0.74	2.17	10.76	32.52	4.05	11.38	5.5x	5.2x
11	Risk	-	1.70	1.70	21.00	21.00	27.93	27.93	16.4x	16.4x
12	Iso Contour	any of 3 - 11	5.20	n/a*	23.90	n/a*	1.40	n/a*	0.27x	n/a*

*Computation time with dependencies varies, depending on the property used for iso contouring.

5.1 Scenario I: Planning Phase

Planning the placement of an off-shore structure demands a complete overview of the ensemble in the spatial domain, but also over all available time steps. Figure 7 outlines all necessary steps. First, the user defines the area of interest (defined by factors not available in the ocean forecast, like reservoir reachability) in the 2D view (Figure 7a) either by a simple bounding rectangle, or completely free by painting directly on the map. In Figure 7b, the sea level of the mean surface for a single time step is mapped to the third dimension. The standard deviation is used for pseudo-coloring in the 3D view. By animating all time steps, the user can now get an overview of the mean sea level at the selected area of interest, as well as the corresponding uncertainties. Besides the 3D view, animation can also be used in the 2D view, showing the sea level using iso contours and pseudo-coloring (inset). While the animation is very effective

(a) Area of Interest Definition.   (b) Sea Level and Std. Deviation.

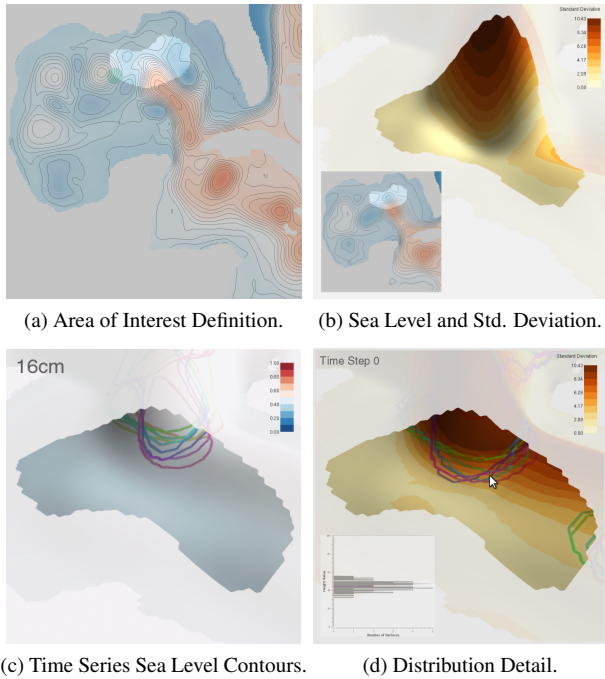(c) Time Series Sea Level Contours.   (d) Distribution Detail.

Figure 7: **Spatial Exploration** for placement planning consists of four main steps: Definition of the area of interest based for example on reservoir reachability (a), general overview (b), time series analysis (c) and detailed analysis for verification (d).
Please use Adobe Reader $\geq$ 9 to enable animations.

to give a first impression of the changing sea level, it is challenging to derive qualitative results. Therefore, in the next step, the user can look at iso contours from the mean surfaces, or risk estimates of multiple time steps in a single view. The contour for a single selected sea level and maximum allowed risk is extracted for all time steps and rendered on the mean surface. The selected sea level, as well as the maximum risk, can be changed on the fly (compare the animation in Figure 7c). Starting with a low sea level and zero risk, the user can gradually approach a suitable compromise of available positions, critical sea surface height and resulting risk, to narrow down the area of interest to a few points. Once a compromise is found, the ensemble distribution can be probed interactively at the interesting positions, to verify the results using the histogram view (inset Figure 7d). At this point the potential placement is narrowed down to a few positions. A detailed analysis of all time steps, identical to the analysis for operations (Section 5.2) can be performed.

5.2 Scenario II: Operational Phase

Most of the ensemble analysis for planning operations and unavoidable downtimes is carried out in the time-series view shown in Figure 5. For a detailed explanation of the view, see Section 3. After definition of a set of positions corresponding to the managed rigs, one position can be selected at a time from a drop-down box. This location is then depicted in the 2D and 3D views for spatial context. The critical sea level, as well as the acceptable risk, can be defined from the user interface. We provide a set of standard color maps for coloring the glyphs. The color map is also freely customizable, most importantly to adapt to the acceptable risk. A good color map should highlight three cases based on the risk estimate: Time steps which are safe for operation with a high certainty, time steps where the rig needs to be shut down with large certainty, and finally uncertain time steps. We found the green to yellow to red diverging color map, as used in Figure 5 to be a good fit, with the green and red mapping to the percentages which indicate safe operations and a high risk, respectively, and the yellow to percentages indicating the need for additional inspection.

The actual operation planning is a recurring process with only a few future time steps available at a time. The parameters like position, critical sea level, acceptable risk and the corresponding color map, however, typically do not change. Hence all these settings can be loaded from a state file alongside new forecast data. Assuming a color map as described, after loading the data the user can immediately identify safe and unsafe time steps from the color of the corresponding glyphs. Only uncertain time steps need further investigation. The main factor to consider for these cases is the spread or uncertainty of the distribution. A compact glyph corresponds to a distribution with little uncertainty. Here, the risk estimate can im-

mediately be used for making a decision to shut down the rig. A large glyph in general indicates large uncertainty. Here, the user must carefully weigh several properties: Are ensemble members in the critical range close to the critical sea level or far above, is the distribution skewed to either side, etc. While in general this information can be derived from the glyph, the user can also access the raw results from the statistical analysis at this point before making a final decision.

6 CONCLUSION

In this work we present an interactive system for the visualization, exploration and analysis of heightfield ensemble data. The core of our framework, which consists of statistical analysis and rendering, is implemented in an efficient GPU-based pipeline. We show the utility of our framework for ocean forecasting. We have received very promising feedback from our domain expert collaborators, and are planning a formal user study in the future.

REFERENCES

[1] R. A. Brown. Animated visual vibrations as an uncertainty visualisation technique. In *International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, pages 84–89, 2004.

[2] E. P. Chassignet, H. E. Hurlburt, O. M. Smedstad, C. N. Barron, D. S. Ko, R. C. Rhodes, J. F. Shriver, A. J. Wallcraft, and R. A. Arnone. Assessment of data assimilative ocean models in the gulf of mexico using ocean color. *Circulation in the Gulf of Mexico: Observations and Models*, 161:87–100, 2005.

[3] L. M. Cherubin, W. Sturges, and E. Chassignet. Deep flow variability in the vicinity of the yucatan straits from a high resolution micom simulation. *Journal of Geophysical Research*, 110:20–72, 2005.

[4] F. Counillon and L. Bertino. High-resolution ensemble forecasting for the gulf of mexico eddies and fronts. *Ocean Dynamics*, 59:83–95, 2009.

[5] W. W. Esty and J. D. Baneld. The box-percentile plot. *Journal of Statistical Software*, 8, 2003.

[6] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2006.

[7] H. Griethe and H. Schumann. The visualization of uncertain data: Methods and problems. In *Proceedings of SimVis '06*, 2006.

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

[9] C. G. Healey and J. Snoeyink. Vistre: A visualization tool to evaluate errors in terrain representation. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 1056–1063, 2006.

[10] J. L. Hintze and R. D. Nelson. Violin plots: A box plot-density trace synergism. *The American Statistician*, 52(2):181–184, 1998.

[11] I. Hoteit, T. Hoar, G. Gopalakrishnan, J. Anderson, N. Collins, B. Cornuelle, A. Kohl, and P. Heimbach. A mitgcm/dart ensemble analysis and prediction system with application to the gulf of mexico. *Dynamics of Atmospheres and Oceans*, 2012.

[12] I. Hoteit, D. T. Pham, and J. Blum. A simplified reduced order kalman filtering and application to altimetric data assimilation in tropical pacific. *Journal of Marine Systems*, 36:101–127, 2002.

[13] J. A. Jeffrey, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano. The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, 90:1283–1296, 2009.

[14] C. R. Johnson and A. R. Sanderson. A next step: Visualizing errors and uncertainty. *IEEE Computer Graphics and Applications*, 23(5):6–10, 2003.

[15] D. Kao, J. Dungan, and A. Pang. Visualizing 2d probability distributions from eos satellite image-derived data sets: a case study. In *Proceedings of the conference on Visualization '01. VIS '01.*, pages 457–560, 2001.

[16] D. Kao, M. Kramer, A. Love, J. Dungan, and A. Pang. Visualizing distributions from multi-return lidar data to understand forest structure. In *In Geoinformatics, Gavle Sweden*, 2004.

[17] O. D. Lampe and H. Hauser. Curve density estimates. *Computer Graphics Forum*, 30(3):633–642, 2011.

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

[19] A. Luo, D. Kao, and A. Pang. Visualizing spatial distribution data sets. In *VISSYM '03: Proceedings of the Symposium on Data Visualisation 2003*, pages 29–38, 2003.

[20] J. Marshall, A. Adcroft, C. Hill, L. Perelman, and C. Heisey. A finite-volume, incompressible navier stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research*, 102:5735–5766, 1997.

[21] K. Matkovic, D. Gracanin, B. Klarin, and H. Hauser. Interactive visual analysis of complex scientific data as families of data surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1351–1358, 2009.

[22] K. Moreland. Diverging color maps for scientific visualization. In *Proceedings of the 5th International Symposium on Visual Computing*, pages 92–103, 2009.

[23] L. Oey, T. Ezer, and H. Lee. Loop current, rings and related circulation in the gulf of mexico: A review of numerical models and future challenges. *Geophysical Monograph-American Geophysical Union*, 161:31, 2005.

[24] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13:370–390, 1997.

[25] T. Pfaffelmoser, M. Reitinger, and R. Westermann. Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. *Computer Graphics Forum*, 30(3):951–960, 2011.

[26] T. Pfaffelmoser and R. Westermann. Visualization of global correlation structures in uncertain 2d scalar fields. *Computer Graphics Forum*, 31(3):1025–1034, 2012.

[27] D. T. Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129:1194–1207, 2001.

[28] H. Piringer, S. Pajer, W. Berger, and H. Teichmann. Comparative visual analysis of 2d function ensembles. *Computer Graphics Forum*, 31(3):1195–1204, 2012.

[29] K. Pöthkow and H.-C. Hege. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1393–1406, 2011.

[30] K. Pöthkow, B. Weber, and H.-C. Hege. Probabilistic marching cubes. *Computer Graphics Forum*, 30(3):931–940, 2011.

[31] K. Potter, J. Kniss, R. Riesenfeld, and C. R. Johnson. Visualizing summary statistics and uncertainty. *Computer Graphics Forum*, 29(3):823–832, 2010.

[32] K. Potter, A. Wilson, P.-T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. R. Johhson. Ensemble-vis: A framework for the statistical visualization of ensemble data. In *IEEE Workshop on Knowledge Discovery from Climate Data: Prediction, Extremes.*, pages 233–240, 2009.

[33] P. J. Rhodes, R. S. Laramee, R. D. Bergeron, and T. M. Sparr. Uncertainty visualization methods in isosurface rendering. In *EUROGRAPHICS 2003 Short Papers*, pages 83–88, 2003.

[34] M. Riveiro. Evaluation of uncertainty visualization techniques for information fusion. In *Information Fusion, 2007 10th International Conference on*, pages 1–8, 2007.

[35] A. Saad, G. Hamarneh, and T. Möller. Exploration and visualization of segmentation uncertainty using shape and appearance prior information. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1366–1375, 2010.

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

[37] H. Scharsach, M. Hadwiger, A. Neubauer, and K. Bühler. Perspective isosurface and direct volume rendering for virtual endoscopy applications. In *Eurovis 2006*, pages 315–322, 2006.

[38] F. M. Vukovich. An updated evaluation of the loop current's eddy-shedding frequency. *Journal of Geophysical Research*, 100:8655–8659, 1995.