Uncertainty Visualization by Representative Sampling from Prediction Ensembles

Le Liu, *Student Member, IEEE*, Alexander P. Boone, Ian T. Ruginski, Lace Padilla, Mary Hegarty, Sarah H. Creem-Regehr, William B. Thompson, Cem Yuksel, and Donald H. House, *Member, IEEE*

Abstract—Data ensembles are often used to infer statistics to be used for a summary display of an uncertain prediction. In a spatial context, these summary displays have the drawback that when uncertainty is encoded via a spatial spread, display glyph area increases in size with prediction uncertainty. This increase can be easily confounded with an increase in the size, strength or other attribute of the phenomenon being presented. We argue that by directly displaying a carefully chosen subset of a prediction ensemble, so that uncertainty is conveyed implicitly, such misinterpretations can be avoided. Since such a display does not require uncertainty annotation, an information channel remains available for encoding additional information about the prediction. We demonstrate these points in the context of hurricane prediction visualizations, showing how we avoid occlusion of selected ensemble elements while preserving the spatial statistics of the original ensemble, and how an explicit encoding of uncertainty can also be constructed from such a selection. We conclude with the results of a cognitive experiment demonstrating that the approach can be used to construct storm prediction displays that significantly reduce the confounding of uncertainty with storm size, and thus improve viewers' ability to estimate potential for storm damage.

Index Terms—Implicit uncertainty presentation, ensembles, ensemble visualization, sampling, uncertainty, hurricane prediction

1 Introduction

Simulation models have become a primary tool in the generation of predictions, but projections from these models often contain a high degree of uncertainty. This uncertainty can have many sources. The unavoidable one occurs when the system being modeled is governed by nonlinear dynamics that are sensitively dependent on initial and boundary conditions. Other sources of uncertainty include assumptions and approximations made in modeling the real system, parameter estimation, and the accumulation of numerical errors [1].

Ensembles are one of the key tools for sampling the space of projections that may be produced by a model containing uncertainty. The use of models in weather prediction is a good example. Typically, one or more weather models are run multiple times, slightly varying initial conditions or parameters for each run [2]. This results in an ensemble of individual model-based projections, from which meteorologists must determine an aggregate prediction to be presented to the general public. Normally, this will include both the predicted weather outcome, and a measure of the certainty or confidence in the prediction.

Although ensembles are an essential tool for making predictions, they are difficult to use to create effective visualizations. Therefore, summary displays, attempting to convey the ensemble's statistics, are typically preferred, especially when presenting the data to the general public [3]. On the other hand, recent studies show that summary displays, portraying the prediction along with its uncertainty, can lead to inaccurate perception of the underlying data [4]. In addition, any summary display requires an explanation and legend, placing cognitive load on the viewer. In this paper we explore approaches for making effective ensemble displays that avoid the limitations of both summary displays and direct visualizations of the original data.

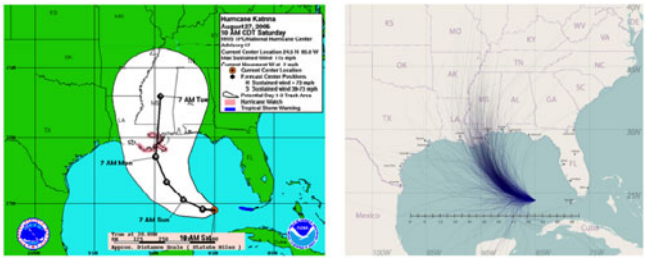
1.1 Summary Displays

Summary visualizations attempt to show at least the mean or median of the ensemble, as well as some indication of the spread of the data. For example Stephenson and Doblas-Reyes used contour lines, over maps of the earth's surface, to show the location and spread of atmospheric pressure predictions [2]. Whitaker et al. [5] developed the idea of contour boxplots to show the median, spread, and outliers in ensembles of spatial contours. Mirzargar et al. [6] extended the contour boxplot idea to handle more general spatial curves. Most recently, Liu et al. [7] demonstrated an approach to viewing hurricane predictions that summarizes an ensemble of storm spatial positions via a set of concentric confidence intervals.

When spatial spread is used as an indicator of uncertainty, it can lead the viewer to incorrect conclusions. A case in point

- *L. Liu and D.H. House are with Clemson University, 100 McAdams Hall, Clemson, SC 29672. E-mail: lel@g.clemson.edu, dhouse@clemson.edu.*
- *A. P. Boone and M. Hegarty are with the Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA 93106. E-mail: {alexander.boone, mary.hegarty}@psych.ucsb.edu.*
- *I.T. Ruginski, L. Padilla, and S.H. Creem-Regehr are with the Department of Psychology, University of Utah380 S 1530 E Beh S 502, Salt Lake City, UT 84112. E-mail: {ian.ruginski, sarah.creem}@psych.utah.edu, lace.m.k.padilla@gmail.com.*
- *W.B. Thompson and C. Yuksel are with the School of Computing, University of Utah50 S. Central Campus Dr, 3190, Salt Lake City, UT 84112. E-mail: thompson@cs.utah.edu, cem@cemyuksel.com.*

Manuscript received 7 June 2016; revised 2 Sept. 2016; accepted 5 Sept. 2016. Date of publication 8 Sept. 2016; date of current version 2 Aug. 2017. Recommended for acceptance by N. Elmqvist. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TVCG.2016.2607204



(a) Uncertainty cone (b) Ensemble (reprinted with permission from [11])

Fig. 1. The National Hurricane Center's uncertainty cone versus a direct ensemble display.

is the cone used by the US National Hurricane Center (NHC) to display the predicted path of a hurricane and its uncertainty, as seen in Fig. 1a. The NHC website [8] states:

> *The cone represents the probable track of the center of a tropical cyclone, and is formed by enclosing the area swept out by a set of circles (not shown) along the forecast track (at 12, 24, 36 hours, etc). The size of each circle is set so that two-thirds of historical official forecast errors over a 5-year sample fall within the circle.*

Therefore, the width of the cone represents the 66 percent confidence region. There is strong experiential evidence that this well-known display leads to misperceptions about the size of a storm-viewers tend to understand that the storm is increasing in size with time. Further, viewers' perceptions may be biased by the presence of the centerline, and the display's binary inside-outside view of uncertainty can lead to overestimation of likelihood within the cone, and underestimation outside [9]. These are not problems that can be fixed by using methods like varying color and opacity to reveal uncertainty. A study, using eye tracking and psychophysiological measures, showed that using such methods to refine the presentation of the uncertainty cone did not result in any significant differences in viewer response [10]-apparently the problems are inherent in the nature of this summary display.

1.2 Direct Ensemble Displays

While directly displaying the ensemble used to make a prediction has the advantage of making all of the data visually available, including its spatial distribution, it is fraught with visualization problems. The most vexing of these is the potential for the display to become a confused jumble, often referred to as "spaghetti plots". In working with point ensembles, if individual points are too close together, display glyphs will overlap, making them difficult to interpret.

The work, presented here, builds on a mounting body of evidence that a well constructed direct display of a prediction ensemble can be superior to a summary display in both conveying the spatial distribution associated with prediction uncertainty, and in minimizing the confounding of spatial attributes of a prediction with its uncertainty. A further potential advantage of direct ensemble displays is that they reduce the dimensionality of the display elements. Instead of an ensemble of points being displayed using a summary area or volumetric display, the points remain points in 2D or 3D. Likewise, paths remain line segments. This has the advantage, for point displays, that a glyph can be displayed at each point, and used to convey additional information about the ensemble member. For example, in a hurricane position display, each ensemble member could be represented by a glyph encoding hurricane category (i.e., maximum windspeed).

If an ensemble of predictions is sampled in time, it becomes a set of points. Directly displaying these points as glyphs will typically create a layout problem, due to visual clutter and overdraw. One sampling approach to avoid these problems is through blue noise sampling, which generates a set of samples that are randomly located but remain spatially well separated. For instance, Wei [12] developed a multi-class blue noise sampling technique, and more recently, Chen et al. [13] utilized this technique to develop a visualization framework for multi-class scatter-plot data. This sampling approach detects collisions of points using a matrix encoding of the inter-class and the intra-class minimum distances in such a way that both the mixed samples and the samples of individual classes are uniformly distributed. The approach is extendable to support adaptive sampling by designing spatially-varying functions, and constructing a distance matrix at each sample using these functions.

Even though using this approach can successfully mitigate the visual clutter, there is no guarantee that the sampled distribution is nearly identical to the original distribution. For our work, control of the distributions is essential, as our goal is to use them to communicate the uncertainty in the prediction implicitly. Our review of the sampling literature revealed no existing technique that can generate a subset of points that both accurately preserves the spatial distribution of the full set, and effectively addresses the overdraw problem.

1.3 Evaluation of Ensemble versus Cone Displays

The earliest attempt at using geospatially displayed path ensembles to represent uncertainty in National Hurricane Center predictions was by Cox et al. [11]. An example of their visualization is shown in Fig. 1b. They conducted a user study that showed that these visualizations lead viewers to make estimates of the spatial spread of uncertainty that are either indistinguishable from estimates made from the cone display, or in some cases show a tendency to improve users' understanding of the spread of likelihood outside of the cone boundaries. This was achieved without the need for either explanation or reference to legends.

Recently, a between-subjects study by Ruginski et al. [4] was conducted, requiring respondents to estimate potential damage to an oil platform, when presented with one of five hurricane prediction visualization styles: NHC cone with centerline, centerline only, cone without centerline, cone with opacity gradient, and path ensemble. Each respondent was presented with one of these styles, seeing presentations with various combinations of platform position and storm path, and asked to make damage estimates at either 24 or 48 hours into the prediction.

Using the NHC cone with centerline as the base case, the hypothesis was that there would be an effect of visualization type on how damage judgment relates to distance from the prediction centerline. No effect was attributable to the cone or fuzzy cone, but at the 48 hour time point the

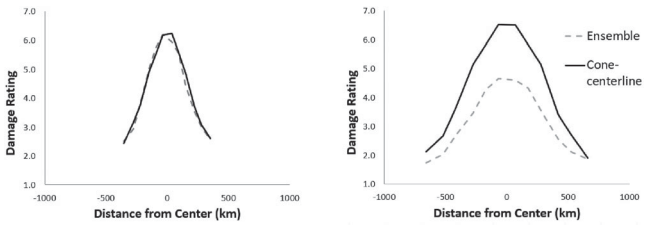


Fig. 2. Mean hurricane damage estimates as a function of distance from center centerline, for ensemble visualization (dotted lines), and the NHC cone (solid lines) for the 24-hour and 48-hour time points. Data from the study by Ruginski et al. [4].

ensemble display produced peak damage estimates that were significantly lower than all of the other cone styles ($p < 0.001$), the curve of damage with distance from the centerline was shallower, and it reached a much lower peak. This can be seen dramatically in Fig. 2. Differences in the curves at 48 hours invite the interpretation that subjects thought that the storm was significantly larger when looking at the cone display. Confirming this, a "Think Aloud" session after the experiment, in which respondents were asked to describe their ranking strategy, revealed a definite perceived size-of-storm effect with all versions of the cone, but almost none with the ensemble. A questionnaire administered following the trials showed that circa 70 percent of respondents reported seeing the storm getting larger with the three cone versions, versus only 31 percent with the ensemble.

1.4 Contributions

The goal of this paper is to explore approaches to making effective displays from prediction ensembles when the ensemble is a set of points. Its main contribution is the elaboration of a methodological framework for either selecting a subset from the original ensemble, or resampling to create a new set that supports a variety of visualization approaches. The resulting subset must have the following properties:

- It must correctly convey the key statistical spatial properties of the original ensemble.
- It must be representative of the full ensemble if used as a basis for scattered data interpolation.
- Points must be spatially well separated to minimize occlusion.

The key idea, in our approach, is first to construct a space in which ensemble members are uniformly distributed, and then to conduct the selection or resampling processes in this space before projecting back to the original space for display.

Fig. 3 shows an example of our methodology applied to an ensemble of predicted hurricane paths supplied to us by the NHC. The original ensemble (Fig. 3a) is sampled at 36 hours into the prediction (Fig. 3b) to show possible hurricane positions in the prediction data. These are displayed using standard glyphs that the NHC uses to indicate hurricane geospatial position and strength on their maps. We construct a spatial warp (Fig. 3c) to transform the set of geospatial samples into a space where they are uniformly distributed (Fig. 3d) that we call the *UD space*. We select a subset of the

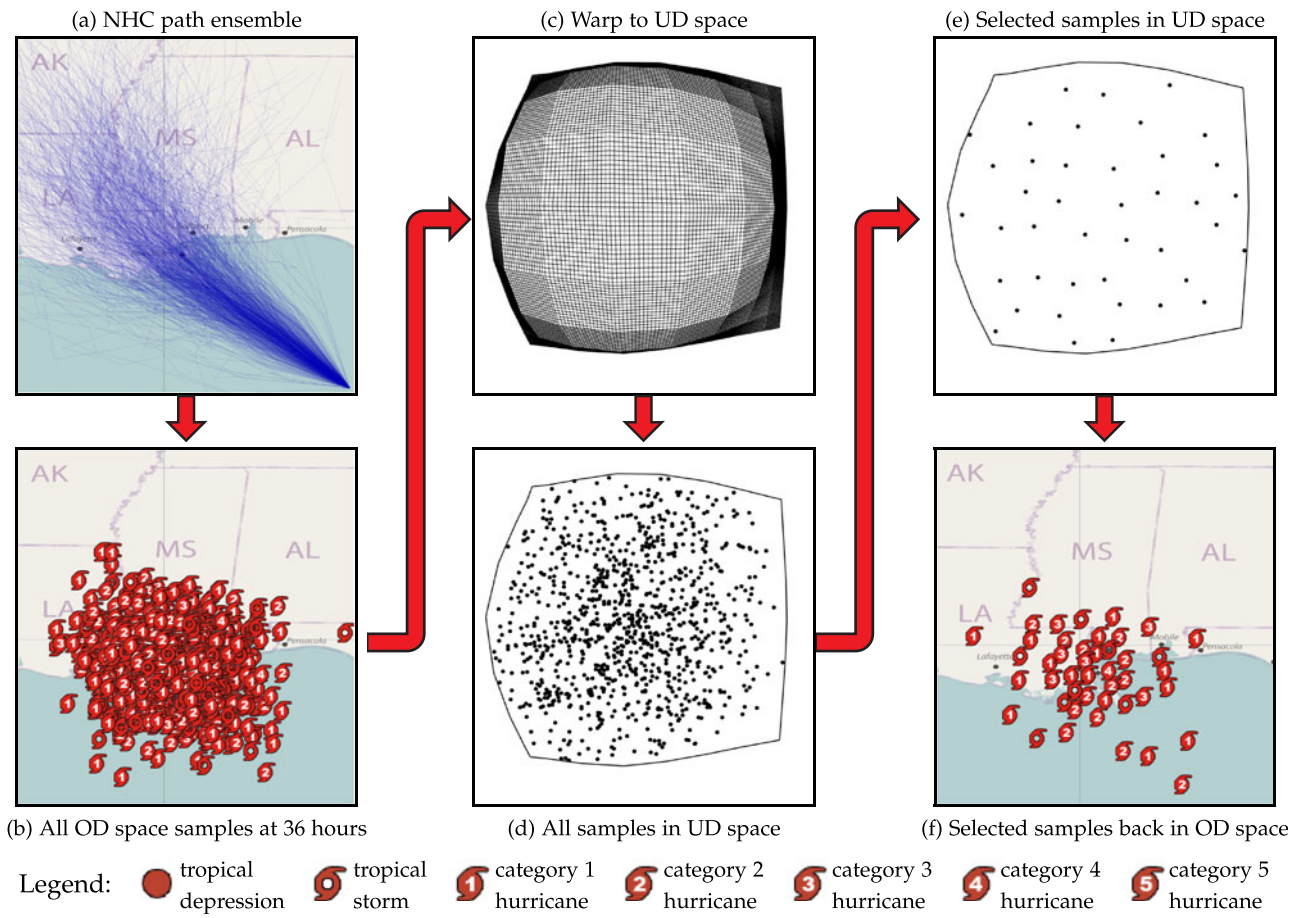


Fig. 3. Our pipeline for sampling from a path ensemble to produce a well structured time-specific ensemble visualization.

samples (Fig. 3e) for achieving a Poisson disk distribution in UD space. Finally, the selected points are transformed back into the original space (that we abbreviate *OD space*) and displayed geospatially using the NHC glyphs (Fig. 3f).

The primary contributions of this paper are:

- Development of a sampling technique that extracts a representative subset of an ensemble of points that not only accurately preserves the original distribution, but also maintains good spatial separation among subset members.
- Development of two dynamic visualizations of a time-sampled hurricane prediction ensemble. One of them conveys the hurricane intensity by glyphs at sample points, while the other conveys storm size by circles whose radii are determined by the storm size.
- Improvement of a previous summary display by Liu et al. [7], showing how our sampling technique can be used to support smooth and accurate interpolation.
- Reporting on the results of a cognitive experiment to compare participants' interpretations of ensemble visualizations to those of existing summary displays.

2 Background

Here we provide a brief mathematical and algorithmic foundation for our approach to finding representative subsets from point and path ensembles.

2.1 Radial Basis Function (RBF) Interpolation

Radial basis function techniques [14] provide a powerful way to interpolate scattered data [15]. RBF's are used to build a continuous function from a set of samples using radially symmetric kernel functions centered on the samples. In the RBF work reported in this paper, we use the Gaussian kernel

$$\phi_i(\mathbf{x}) = \exp - \frac{(\mathbf{x} - \mathbf{x}_0)^2}{2r^2},$$

where r is the kernel spread parameter. This kernel has the desired property of extrapolating to 0 away from the data, and has infinite support, which is desirable for handling broadly spread data. Many other kernel choices are possible, including some of finite support [16].

2.2 Subset Selection via Orthogonal Least Squares (OLS)

Simply using RBF interpolation to form a field from a large set of point samples is problematic for several reasons. Most importantly, each one of the point samples must be considered to contain some uncertainty, so fitting a surface exactly to these point samples will result in overfitting that will not generalize the data based on its statistics. The *Orthogonal Least Squares* method [17] comes out of the machine learning community, and is an approach to solving RBF interpolation problems by selecting a small representative subset to use as a reduced basis for interpolation [17], [18]. OLS starts with an empty subset to which one point sample is added at a time, by selecting the sample that minimizes the sum squared RBF interpolation error (SSE) at the original sample points. The *forward selection algorithm* [19] is a fast implementation of this algorithm.

2.3 Subset Selection via Weighted Sample Elimination (WSE)

A completely different approach to choosing a subset of point samples from the original set is to optimize for spatial layout, rather than mean squared interpolation error. Choosing a subset of an ensemble that maintains the Poisson disk property (i.e., that all randomly located points are tightly packed but lie beyond a chosen Poisson disk radius r from each other) will assure that the space under consideration is being sampled in an optimal way, in the sense of signal processing theory [20]. A number of techniques have been developed to generate Poisson disk samplings over a spatial domain, including acceleration methods using spatial data structures [21], and varying the disk radius over a domain using importance measures [22].

Quite recently, Yuksel [23] developed the weighted sample elimination approach, which starts with an existing sample set, and returns a radius and a subset of a specified size that has the Poisson disk property for this radius. This approach is ideal for our purposes, which is to determine a representative subset of a spatially distributed ensemble. The approach uses a kd-tree structure to determine a set of neighbors for each point, and a heap data structure organized by a sum of weights representing the accumulated distances of each point from its neighbors. The algorithm iteratively discards the point with the highest weight, recomputes the summed weights, and reorganizes the heap until the desired number of samples is reached. The Poisson disk radius is then the minimum distance between the remaining points.

3 Methodology

The individual steps of our approach for selecting a representative subset of N points from an ensemble of M points is given below:

1) *Determine the bounding region of the M points in OD space.*
   All work is done with respect to the bounding region of the ensemble points in the original density (OD) space.
2) *Compute a continuous density field over the bounding region.*
   Within this bounding region, we use ensemble points to estimate a continuous density field, as explained below in Section 3.1.
3) *Construct a warp from OD to UD space.*
   We then use this field to construct a warping function that maps the points in OD space to a space in which the density per unit area is constant, i.e., the uniform density (UD) space. We discuss the construction of this mapping function in Section 3.2.
4) *Map each point from OD space to UD space.*
   This warping function is used to map each point (x_i, y_i) in OD space to (u_i, v_i) in UD space, such that samples in UD space are uniformly distributed.
5) *Select a set of $N < M$ points in UD space.*
   A point selection algorithm is then used in UD space to select a subset of samples, as explained in Section 3.3.
6) *Project the N selected points back to OD space and display.*

Finally, the selected points in UD space are mapped back to OD space for display, as discussed in Section 3.4.

As long as the selected subset is uniformly distributed in UD space, we have a guarantee that the corresponding sample set in OD space is representative of the statistical properties of the original ensemble. Note, that this approach is insensitive to whether or not the original distribution is unimodal or multimodal. In either case, the distribution we end up with in UD space is uniform, while the inverse map back to OD space restores the original modality.

3.1 Computing the Continuous Density Field

We compute the density field over the sample space using radial basis function interpolation, after assigning a local data density value to each sample point. We avoid using a grid-based discretization for the density estimation, because the input data from the original ensemble may be distributed in a highly non-uniform way, which can lead to sampling issues. Instead, we directly use the samples of the original ensemble utilizing a k nearest neighbors (kNN) approach [24].

Given a sample point i with position $\mathbf{x}_i = (x_i, y_i)$, from a data set of size M, the kNN density σ_i at this point is defined as

$$\sigma_i = \frac{k}{M\pi {\rho_i}^2},$$

where ρ_i is the radius of the circle with center at $\mathbf{x}_i$ that minimally encloses the k nearest neighbors of the point i. A kd-tree can be used for rapid determination of the k nearest neighbors of each sample point.

However, interpolating the density field from all of the points in the ensemble would result in overfitting. Therefore, we use the forward selection implementation of the OLS algorithm [19] to select a subset that minimizes mean squared interpolation error. For each sample point i we compute an RBF spread parameter

$$r_i = \beta \frac{w}{\sqrt{\sigma_i}}, \tag{1}$$

that adapts to the local density measure. Here, w is the largest dimension of the bounding region in OD space, and β is a user settable constant. In this way, the RBF interpolation algorithm can handle scattered data with widely varying density.

It is possible to use the subset of the original points obtained with this procedure as the subset for displaying the ensemble as well. However, by design, this subset is more uniformly distributed than the original data set. Therefore, it does not provide a good representation of the statistical properties of the original ensemble. Furthermore, this subset can also contain points that are too close together, resulting in glyph occlusion. These two problems can be clearly seen in Fig. 4. Comparing the original ensemble of time-specific hurricane position predictions (Fig. 4a) to the subset of 100 sample points chosen by the OLS algorithm (Fig. 4b), it is apparent that the resulting subset is too uniform and some points are too close together. These are limitations of not only the OLS approach, but any sampling algorithm selecting a subset of points in the space they are originally living in. To address these limitations, we perform our sample selection in UD space.

(a) entire set (b) 100 selected samples

Fig. 4. An optimal subset chosen using OLS in the original space, is too uniformly distributed.

3.2 Mapping from OD to UD Space

There are a number of ways to construct a mapping

$$OD, UD \subseteq \mathbb{R}^2, \quad f : UD \rightarrow OD, \quad (x, y) \mapsto (u, v),$$

but not all of these result in a warp suited to our purposes. Our warp function should minimize both shear and non-uniform scaling, since the selection process in UD space will be based on euclidean distances between the points. Locally, the mapping should be as close as possible to uniform scaling.

It is possible to define a mapping from a non-uniform distribution to a uniform distribution using a cumulative density function. In $\mathbb{R}^2$, the probability density p at position (x, y) can be written as

$$p(x, y) = p_x(x) p_y(y|x),$$

where

$$p_x(x) = \int_{-\infty}^{y} p(x, y) dy \quad \text{and} \quad p_y(y|x) = p(x, y)/p_x(x),$$

yielding

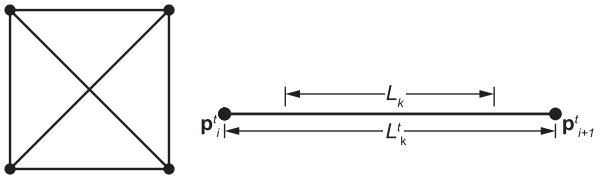
$$u(x, y) = \int_{-\infty}^{x} p_x(x) dx, \quad \text{and} \quad v(x, y) = \int_{-\infty}^{y} p_y(y|x) dy.$$

Although this mapping results in a uniform distribution, it introduces a great deal of shear along the v axis, such that points that are distant from each other in OD space can be moved near to each other in UD space, and vice versa. Therefore, the Euclidian distances between points in UD space will not be a good metric for selecting a subset.

The approach that we have found to be most useful is a Gauss-Seidel style relaxation process. We begin with splitting the OD space into a number of grid cells with uniform sizes. In UD space, this grid should be deformed, such that the area of each grid cell should roughly correspond to the density within the grid cell in OD space. The deformation of this grid represents our warping function. A point (x_i, y_i) in a grid cell in OD space is mapped to (u_i, v_i) in the corresponding cell in UD space, using matching barycentric coordinates.

3.2.1 Relaxation for Grid-Based Warping Function

Without loss of generality, let us assume that the initial grid is a square with dimensions $S \times S$. Since UD space has uniform density, each deformed grid cell in UD space should have the same average density



(a) cell model (b) edge model

Fig. 5. Model used for a grid cell and its edges.

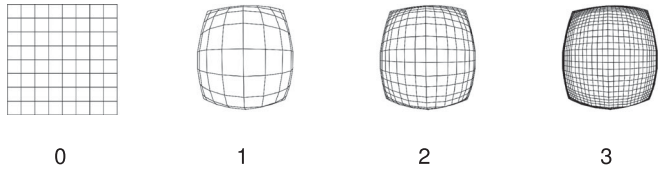


Fig. 6. Initial, and three levels of hierarchical refinement for computing the warping function.

$$d_a = \frac{\sum_{j=1}^{S^2} d_j}{S^2 A} \ ,$$

where d_j is the initial density of the cell j and A is the initial area of a grid cell. In practice, we estimate the initial density d_j of a grid cell by sampling at the center of the cell, using the kNN density field described in Section 3.1. The target area of a cell A_j can be written as

$$A_j = A\frac{d_j}{d_a}.$$

For computing the deformed grid, we assign a target length to each edge of the grid using the A_j values of the neighboring cells. The target length of an edge k on the perimeter of the grid that has a single cell neighbor j is

$$L_k = L\sqrt{\frac{A_j}{A}},$$

where L is the initial length of a grid edge. For other edges that are shared between two cells, we use the average of the two target length values computed for each face. We also construct two diagonal edges for each cell to minimize shear during the relaxation. The target lengths of the diagonal edges are computed similarly, using the target areas of their cells. The structure of a single cell is shown in Fig. 5a.

Our relaxation process changes the length of each edge k towards the target length L_k, at each step t, by moving its vertices along the edge direction. Let $\mathbf{p}^t_i$ and $\mathbf{p}^t_{i+1}$ be the positions of the two vertices of an edge k at relaxation step t, as shown in Fig. 5b. The updated position of the vertex $\mathbf{p}^{t+1}_i$ due to the update operation for edge k can be written as

$$\mathbf{p}^{t+1}_i = \mathbf{p}^t_i + \alpha\left(\frac{L^t_k - L_k}{2}\right)\left(\frac{\mathbf{p}^t_{i+1} - \mathbf{p}^t_i}{L^t_k}\right), \tag{2}$$

where $L^t_k = \left|\mathbf{p}^t_{i+1} - \mathbf{p}^t_i\right|$ is the length of the edge at relaxation step t and $0 < \alpha \leq 1$ is a user adjustable acceleration factor, such that if $\alpha = 1$, $L^{t+1}_k = L_k$ at the end of the update for edge k.

3.2.2 Hierarchical Relaxation

For achieving a high quality warping function, we need to have a high-resolution grid. But, when the displacement of the grid vertices due to the deformation of the grid is much larger than the length of a grid edge, the relaxation process can result in crossed edges, inverting cells and thus changing the grid's topology, leading to instability. To avoid this, we use a hierarchical progressive-refinement approach, as illustrated in Fig. 6. We first use relaxation to compute a low-resolution deformed grid. Then, we subdivide the cells and continue the relaxation process for the subdivided grid with higher resolution. This way, large deformations of the grid are handled using lower resolution grids and fine details of the deformation are introduced using relaxation of higher resolution grids.

The relaxation process that we use to obtain a solution at a level in the hierarchy has two stopping conditions. First, we monitor the sum of squared errors (SSE), defined as the sum of squared differences between desired areas and current areas of all grid cells

$$\mathrm{SSE} = \sum_{p=1}^{S_k^2} (A_p^D - A_p)^2,$$

where S_k denotes the grid dimension at the kth level in the hierarchy. We terminate the relaxation if the SSE is less than a small threshold. The second stopping condition is the detection of an inverted cell. When a target cell area is very small, the relaxation process can still produce inversion. If an inverted cell is detected, we restore the system to its previous state and terminate the relaxation for that level of the hierarchy and continue the relaxation at the next lower level.

At each stage of the refinement process, we try to adjust the area of each grid cell so that its density equals the average density of the OD domain. In our implementation, we set the grid dimensions as powers of two, so we associate the finest grid dimensions $S \times S$ with M samples using

$$S = \lfloor \log_2 \gamma M \rfloor, \tag{3}$$

where γ is a user adjustable fraction between 0 and 1. If S_0 is the dimension of the grid at the coarsest level, our method requires

$$T = 1 + \log_2 S - \log_2 S_0 \tag{4}$$

hierarchical levels to compute the highest resolution deformed grid.

3.3 Selecting a Representative Subset

In order to select a representative subset, we first transform all samples in OD space to UD space. Subsequently, we choose a representative subset of the samples in UD space using either the Orthogonal Least Squares algorithm explained in Section 2.2 or the Weighted Sample Elimination method presented in Section 2.3. These two methods select subsets with different features.

The OLS method is a selection approach based on a construction of an RBF system. To avoid the high computational cost of the naive OLS algorithm, we utilize Orr's forward selection algorithm [19], which monitors the reduction in SSE of the RBF system, aimed at selecting the minimal subset satisfying a specific error criterion. While building the RBF, the righthand side of the linear system is assigned by computing simplicial depth values [25] of the

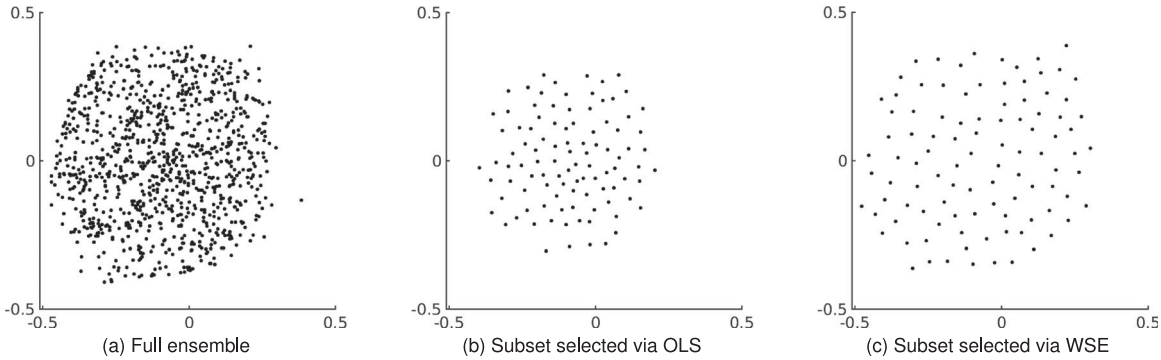


Fig. 7. Comparison of samples selected via OLS and WSE algorithm in UD space.

original samples. Simplicial depth measures the centrality of a point within a point cloud. A sample with bigger simplicial depth value is considered to be more representative, and thus will tend to have a higher chance to be selected by the OLS algorithm. Fig. 7a shows the entire ensemble, and Fig. 7b shows the subset selected by the OLS algorithm. This subset is concentrated in the central region of UD space, and thus is tightly clustered in OD space.

On the other hand, the WSE algorithm attempts to select a subset that maintains the Poisson disk property. Therefore, the samples in the chosen subset should be uniformly spread over the entire UD space, and thus are more broadly spread in OD space, as can be seen in Fig. 7c. In fact, any algorithm that can produce Poisson disk sample sets can be integrated into our framework. However, we chose WSE because it is the most efficient published algorithm that can produce a Poisson disk subset from an existing set. Other Poisson disk algorithms, e.g., dart throwing, generate new samples rather than selecting from an existing set. We demonstrate our approach using WSE because it performs as well as other alternatives, is fast, and easy to implement.

In the WSE algorithm, the sum of weights from neighbors of a sample determines the sample's depth in a heap data structure. Because samples near the boundary of UD space have no neighbors outside of the boundary, their weights are biased, making them more likely to be kept in the selected samples. To address this problem, we extract the concave hull of the points collection, and randomly generate extra samples in the region outside of this hull, so that the density of the boundary region matches the average density of UD space. These generated samples are considered while computing the weights of the original samples but are never selected by the algorithm.

3.4 Inverse Mapping from UD to OD Space

Inverse mapping of the original points from UD space to OD space is trivial, since they simply map back to their original coordinates in OD space. However, if it is desired to resample in UD space to create new points, an inverse mapping is required to find the corresponding positions in OD space. The inverse mapping can be defined similarly to the forward map, but starting with the deformed grid. Given a point (u_i, v_i) in UD space, we first find the deformed grid cell that contains this point. Since our relaxation procedure ensures that no grid cell is either inverted or has crossing edges, there exists a unique deformed grid cell for any point in UD space. A space partitioning structure can be utilized here for quickly finding the corresponding cell. Then, barycentric coordinates can be used to project the point to the corresponding undeformed OD space cell.

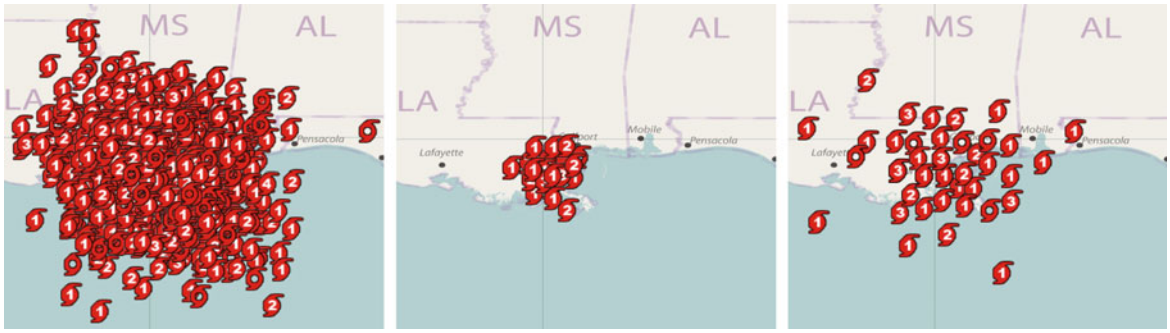
4 Validation and Visualization Design

In this section, we show experimental results demonstrating the utility of our proposed technique to choose a representative subset of time-specific hurricane positions, from a prediction path ensemble. We first demonstrate that these chosen subsets closely reproduce the spatial distribution encoded by the full point ensemble. We then show that our approach can be used to support two forms of representative visualization superimposed on a map: a point ensemble display using glyphs to show a set of predicted storm locations annotated with additional storm information, and a summary display showing three levels of risk region, similar to the displays developed by Liu et al. [7].

We demonstrate our approach by using the NHC's ensemble of 1,000 predicted paths for Hurricane Isaac. The prediction was made about 36 hours before the hurricane made landfall in Louisiana, on August 29, 2012. Since the paths can be sampled in time and encode predicted storm characteristics, such as wind speed and storm size at regular time intervals, they can be used to produce spatiotemporal visualizations. We have already shown an example of our complete process, using this NHC prediction, in Fig. 3.

There are five user settable parameters in our approach. In all of our results, shown below, these parameters are set as follows. To determine the number of nearest neighbors for the kNN density calculation in Equation (3.1), we let $k = \lfloor 0.01M \rfloor$, where M is the data set size. In Equation (1), determining radial basis function spread for the kNN density calculation, we set the scaling parameter $\beta = 0.01$. In determining S, the grid dimension at the finest level of the hierarchy, we set $\gamma = 0.2$ in Equation (3). Thus, given 1,000 predicted hurricane locations, our final grid resolution is $S = 128$. We chose the coarsest grid resolution to be $S_0 = 8$, so there are $T = 5$ levels in our hierarchy (Equation (4)). The acceleration constant used in performing the relaxation algorithm, defined by Equation (2), was set to $\alpha = 0.066$. For all of our demonstration images, the OD space is measured geospatially, by latitude and longitude. The UD space is centered at the origin, with dimensions of one unit by one unit: $[-0.5, 0.5]$.

Fig. 8 shows an example of how the choice of the OLS versus the WSE selection algorithm affects the layout of the selected samples. We assign the sizes of the chosen subsets based on a heuristic that attempts to control the tightness of



(a) Full ensemble (b) Subset selected via OLS (c) Subset selected via WSE

Fig. 8. Comparison of samples selected via OLS and WSE algorithms, Hurricane Isaac prediction at 36 hours.

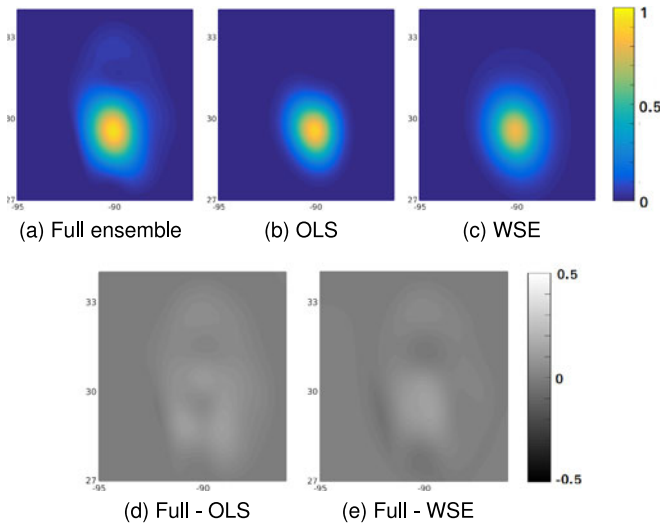
spacing. Given the area of a bounding region B of the original set of samples, the size of the subset is determined as $N = \delta B$, where δ denotes the desired number of samples per unit area. In our implementation, δ is set to be 0.72 for the OLS algorithm and 1.45 for the WSE algorithm, reflecting the relative tightness of spacing between the two approaches. Comparing Fig. 8b, which has 21 points, with Fig. 8c, which has 42, the OLS algorithm tends to produce a subset that is more concentrated around the most likely position of the hurricane, while the WSE algorithm tends to produce a subset that is more widely spread, indicating not only the most likely storm position but the possibility of outliers. The reasons for these differences have already been discussed in Section 3.3.

It is important to note that while the original ensemble in Fig. 8 appears to spread over an even larger area than both of the subsets generated by OLS and WSE, these outlier samples away from the center actually appear with very low probability and the density of the ensemble samples near the center is completely imperceivable due to occlusion. Therefore, the visualization of the full ensemble can be misleading.

While the two selection algorithms produce subsets with different layouts, both subsets are representative of the original ensemble. One way to verify this is to test if a selected subset can be used to reproduce a distribution supported by the original ensemble. To make this comparison, we built simplicial depth fields from both the original ensemble and each of the selected subsets. Since simplicial depth provides a measurement of the centrality of a point in a collection, by comparing two simplicial depth fields, we are able to evaluate the representativeness of the subsets. Simplicial depth values are normalized to the range $[0, 1]$ and color coded for display. Fig. 9a shows a simplicial depth field interpolated from the original ensemble, Fig. 9b from the OLS subset, and Fig. 9c from the WSE subset. To make a visual comparison, we subtract the grayscale value of the OLS and WSE images from that of the original ensemble image, with the differences displayed in Fig. 9d and Fig. 9e. Both of those subtractions give a variation of simplicial depth values between $[-0.05, +0.05]$ out of a possible range of $[-1, +1]$. Thus, the maximum variation is only 5 percent of the range of simplicial depth values. Therefore, we can conclude that both of the two selected subsets are highly representative of the original sample set, and the choice of the selection algorithm can be left to the user in order to satisfy their visualization requirements.

Although this is not the main thrust of this paper, because the subsets selected by our approach are highly representative of the original ensemble, they can be used to smoothly interpolate spatio-temporal information carried by the original ensemble, and used to generate summary visualizations. By interpolating the simplicial depth field of a selected subset, we can produce a visualization showing the geospatial spread of hurricane strike likelihood at a particular point in time, in the form of concentric ellipsoidal confidence intervals. In a recent paper, Liu et al. [7] produced confidence interval visualizations by fitting ellipses to the simplicial depth field computed from the full ensemble. The top row of Fig. 10 shows a time sequence produced utilizing their proposed approach. To make their method frame-to-frame coherent, they required the use of a non-linear stabilizing filter to reduce temporal instability in the ellipse orientation. Such a filter could introduce arbitrary errors in the visualization. Using our approach, with the OLS selection algorithm, we are able to produce very similar visualizations directly from the interpolated simplicial depth field, without resorting to ellipse fitting. The results are not only faster to compute but remain stable, without the use of a filter, when viewed as a time sequence. The bottom row of Fig. 10 shows such a time sequence produced utilizing our algorithm.

The main goal of our work was to produce uncluttered, nicely spaced, direct ensemble displays. One attribute of



(a) Full ensemble (b) OLS (c) WSE

(d) Full - OLS (e) Full - WSE

Fig. 9. Comparison of simplicial depth fields of original full ensemble with subsets selected via the OLS and WSE algorithms.

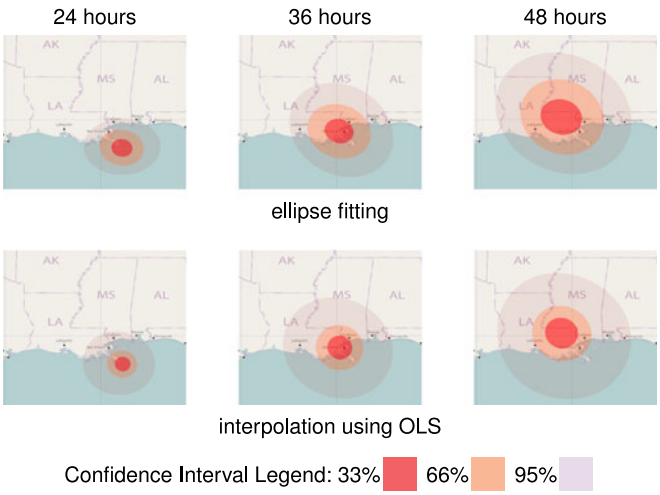


Fig. 10. Simplicial depth field displays over time, comparing Liu et al.'s ellipse fitting [7] with interpolation from an optimal subset.

this style of display is that glyphs carrying information, in addition to position, can be placed at the location of each sample without concern for occlusion. To demonstrate that our proposed technique has great potential to support this type of visualization, we created a direct ensemble display by placing glyphs indicating predicted storm intensity at each of the predicted locations. Fig. 11 shows two such displays at 36, and 48 hours into the prediction. These displays are intended to show the most likely storm position, while indicating that outliers are more and more likely as time progresses into the prediction. Because we wanted to clearly show outliers, we chose the WSE algorithm to obtain the subsets used in this particular case.

We demonstrated these visualizations to meteorologists at the NHC and one of their important critiques was that this visualization did not have enough coverage of the area likely to experience hurricane force winds. A second concern was that individual elements of the ensemble are highly salient, potentially misleading viewers into paying too much attention to certain glyphs instead of the overall distribution.

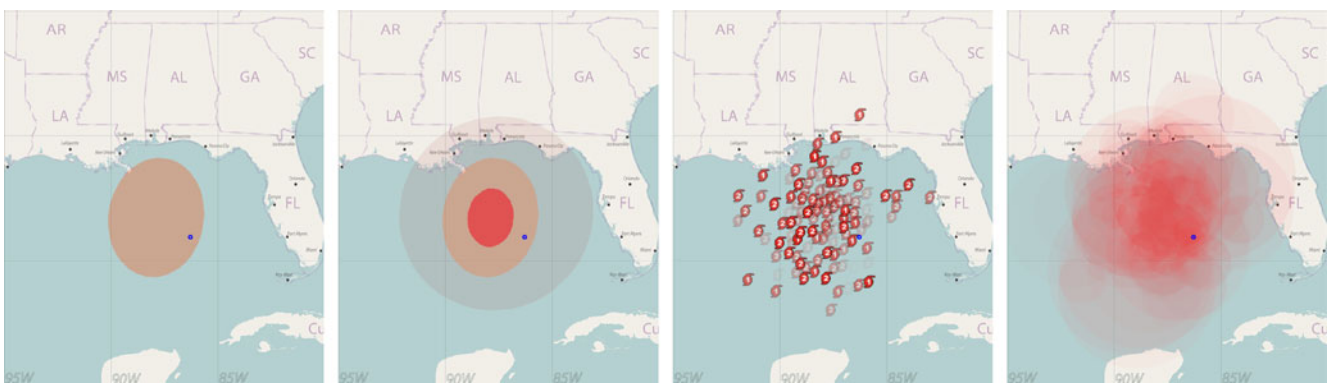
To address these concerns, we created an animated visualization that continuously adds new ensemble members to the display while slowly fading out glyphs that have been on the screen for a while. Over time this allows many more ensemble members to be displayed, without creating clutter, and, since they are always fading away,

Fig. 11. Glyph displays of the prediction, over time, using ensemble subsets from the WSE algorithm.

also deemphasizes any particular glyph. Within this visualization, areas with a high concentration of samples tend to be more opaque than those with low concentration. Thus, more opacity indicates a higher level of certainty. This was implemented by initially randomly selecting opacities for all of the samples. Samples are divided into two groups, one group with non-zero opacity that is on the display, and one group with zero opacity that is not displayed. Each time the opacity of a displayed sample is decreased to zero, we pick a new sample from the zero opacity group whose simplicial depth value is closest to that sample, and display it at maximum opacity. This ensures that the distribution at each time frame is as close to the distribution of the full ensemble as possible. Fig. 12c shows a snapshot of such a visualization.

Another critique from the NHC meteorologists was that even though our glyphs depict a set of predicted positions of the center of the storm annotated by storm strength, they do not show the extent of the area potentially affected by hurricane force winds. To address this concern, we developed a visualization that uses circles entered at the selected samples whose radii correspond to predicted distance from the center at which 50 kn winds are predicted. We employ the same pipeline used for drawing strength glyphs to create a dynamically updating display. A snapshot of this style of visualization is shown in Fig. 12d.

We are confident that our approach can be easily implemented to operate in real time. We did not conduct timing studies, since our research software was not fully integrated and used Matlab for some calculations. Even with this approach, all computations are done in under a second.



(a) 66% CI (b) 33%, 66%, and 92% CI Bands (c) Strength (Icons) (d) Sizes (Splats)

Fig. 12. The four visualization styles studied in the cognitive experiment. The blue dot indicates the position of an oil rig platform.

5 Experimental Evaluation

In order to evaluate the effectiveness of our visualizations in helping users to estimate hurricane risk, we conducted a cognitive experiment comparing each of the three new visualizations described above with a visualization simply showing the 66 percent confidence interval for the expected storm center. We chose the 66 percent CI as the standard for comparison because it is most similar to current practice in communicating hurricane forecasts (i.e., the NHC uncertainty cone, see Fig. 1a).

We started from historical NHC predictions of five US Gulf Coast hurricanes. Three of these were translated slightly to assure that the predicted storm center was offshore at 48 hours from the time of the forecast. Four visualizations were generated for each storm, taken at 48 hours. Two of these visualizations were static summaries of hurricane locations, the 66 percent confidence interval (66 CI) (Fig. 12a), and the 33, 66, and 92 percent concentric confidence intervals (CI Bands) (Fig. 12b). The other two were animated displays of ensembles of points. Both ensemble visualizations represented the storm location, with one using glyphs to show the storm strength (Icons) (Fig. 12c) and the other representing the storm size (Splats) (Fig. 12d). In the experiment, participants were informed of the contents of each type of display by showing brief descriptions. For instance, before seeing the splats condition, they were told "You will see a series of circles that will appear and disappear. Each circle represents a possible location and size of the storm".

5.1 Experimental Methodology

133 participants from the University of California, Santa Barbara and the University of Utah were first randomly assigned to a visualization type, and provided with a description of that visualization and instructions for the task. Modeled after the experiment by Ruginski et al. [4], participants were shown one of the forecast display techniques that additionally included a superimposed blue dot indicating the location of an offshore oil platform (see Fig. 12), using E-Prime 2.0.10.353 software [26]. Displays were presented on either a Dell U2412MB monitor or an Asus VG248 monitor in sRGB color mode with pixel resolutions of $1{,}920 \times 1{,}080$. The offshore oil platform was located at one of eight locations defined relative to the center of the forecast. The participants were given a description of an oil platform and instructed to estimate the level of damage that the platform would incur based on the likelihood of the storm affecting the platform and the strength of the storm in the affected region. Below the display was a Likert scale ranging from 1 (no damage) to 9 (severe damage), and the participants entered their responses on a keyboard. This scale was chosen to provide participants with a means of hurricane interpretation that would simulate a real world decision task. Each participant completed 40 trials in which the locations of both the platform and hurricane were randomly presented, for their visualization type only. Finally, participants were asked a series of seven true/false questions to assess their understanding of the displays, including misconceptions.

A multilevel model (MLM) was fit to the damage judgment data using Hierarchical Linear Modeling (HLM 7.0) software and restricted maximum likelihood estimation procedures [27], [28]. Multilevel modeling is a generalized form of linear regression that is used to analyze variance in experimental outcomes on both individual (within-participants) and group (between-participants) levels. A MLM was appropriate for modeling our data and testing our hypotheses for two major reasons: 1) MLM allows for the inclusion of interactions between continuous variables (in our case, distance) and categorical predictors (in our case, the type of visualization); 2) MLM uses estimation procedures appropriate for partitioning variance and error structures in mixed and nested designs (repeated measures nested within individuals in this case).

Level 1 of our multilevel model is described by

$$\text{Damage}_{ij} = \beta_{0j} + \beta_{1j} \times \text{Distance}_{ij} + r_{ij},$$

and level 2 by

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01} \times \text{Icons}_j + \gamma_{02} \times \text{Splats}_j \\ &\quad + \gamma_{03} \times \text{CIBands}_j + u_{0j}, \quad \text{and} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} \times \text{Icons}_j + \gamma_{12} \times \text{Splats}_j \\ &\quad + \gamma_{13} \times \text{CIBands}_j,\end{aligned}$$

where i represents trials, j represents individuals, and the γ terms are the regression coefficients. The error term r_{ij} indicates the variance in the outcome variable on a per trial basis, and u_{0j} on a per person basis [29].

Damage rating, although an ordinal variable by definition, was treated as continuous in the model because it contained over five response categories [30]. For the distance variable, we analyzed the absolute value of oil rig distances, regardless of which side of the hurricane forecast they were on, as none of our hypotheses related to whether oil rigs were located on a particular side. We divided distance by 10 prior to analysis so that the estimated model coefficient would correspond to 10 km changes (rather than a 1 km change). The type of visualization was dummy coded such that the 66 percent CI was the reference group, allowing for comparison of each visualization to the 66 percent CI. The analysis was collapsed over the five hurricane forecasts.

The mixed two-level regression model tested whether the effect of distance from the center of forecasts (level 1), varied as a function of visualization (level 2). We hypothesized that individuals who viewed the CI Band, Icon, and Splat visualizations would make damage judgments that more gradually change as a result of distance compared to the 66 percent CI, demonstrating interpretations more in line with an uncertainty distribution than the 66 percent CI.

5.2 Damage Judgment Results

The analysis indicated that average damage ratings were around the maximum possible (9.13) at the center of the 66 percent CI as shown in Table 1. The intercept terms in Table 1 reflect the average value for the 66 percent CI whereas the additional coefficients represent the difference from this value. Similarly, hypothesis tests for the intercept term reflect whether the average value for the 66 percent CI is meaningfully different from zero; hypothesis tests for other coefficients represent whether the difference between the specified group and the 66 percent is meaningfully

TABLE 1
Intercept Level 1, β_0

Fixed Effect	Coeff.	Std Err	t-ratio	DOF	p-value	95% CI
Intercept 2, γ_{00}	9.129	0.177	51.68	119	<0.001	(8.78,9.48)
Icons, γ_{01}	−2.525	0.371	−6.79	119	<0.001	(−3.26, −1.79)
Splats, γ_{02}	0.159	0.221	0.72	119	0.473	(−0.27, 0.59)
CI Bands, γ_{03}	−1.433	0.353	−4.06	119	<0.001	(−2.12, −0.74)

Average damage judgments made at the center of the forecasts. Main effect (intercept) references 66 percent CI visualization. The Table summarizes the comparison of each visualization to the 66 percent CI.

TABLE 2
Slope Level 1, β_1

Fixed Effect	Coeff	Std Err	t-ratio	DOF	p-value	95% CI
Intercept 2, γ_{10}	−0.210	0.006	−32.36	4793	<0.001	(−0.22,-0.19)
Icons, γ_{11}	0.094	0.015	6.31	4793	<0.001	(0.06, 0.12)
Splats, γ_{12}	−0.003	0.009	−0.36	4793	0.720	(−0.02, 0.01)
CI Bands, γ_{13}	0.053	0.018	2.94	4793	0.003	(0.02, 0.09)

Effect of a 10 km change in distance on damage ratings. Main effect (intercept) references 66 percent CI visualization. The table summarizes the comparison of each visualization to the 66 percent CI.

different from zero. The average damage rating value is estimated above the top of the scale since individuals never made a damage judgment exactly at the center of the hurricane. Compared to the 66 percent CI, damage judgments made using the CI Bands at the center of the hurricane forecast were 1.43 lower on average, and judgments made using the Icon were 2.53 lower on average. Interestingly, average damage ratings made at the center of the forecast using the Splat visualization were not meaningfully different from those made using the 66 percent CI.

Second, our analysis revealed a significant association between distance and damage rating for the 66 percent CI visualization. There was a 0.21 decrease in damage judgment per 10 km on average as shown in Table 2. This was expected, given that individuals had little information aside from distance when making damage judgments using this visualization. More importantly, both the CI Bands visualization (0.157 decrease per 10 km, see Fig. 13a) and the Icon visualization (0.116 decrease per 10 km, see Fig. 13b) demonstrated a significantly less strong association between distance and damage rating than the 66 percent CI. Lastly, the relationship between distance-from-center and damage rating for the Splats visualization did not differ significantly from the 66 percent CI visualization (see Fig. 13c).

5.3 Post-task Questionnaire Results

In a post-task survey, participants were asked a series of seven true/false questions (given in Table 3). We compared each of the other visualization conditions to the 66 percent CI condition using chi-squared tests of independence.

Responses to Questions 2, 3, and 4 did not differ significantly, indicating that across all conditions, most participants understood that the displays indicate 1) that the area being shown has a chance of being damaged, 2) where the center of the storm is likely to be, and 3) that the forecasters are uncertain of the storm's location.

Responses to Question 1 indicated that participants in the CI Bands group, $\chi^2(1, N = 65) = 4.99, p = 0.03$, and participants in the Splats condition, $\chi^2(1, N = 62) = 10.09, p = 0.001$, were more likely to endorse the statement that the forecast shows a distribution of possible locations.

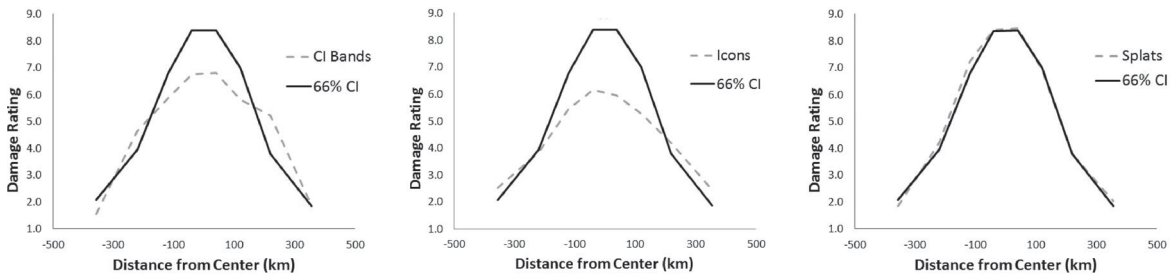
Importantly, although only the Splats condition showed the size of the hurricane, there was no significant difference between conditions in endorsement of Statement 5 that the forecast visualization indicates the size of the hurricane. Participants in the 66 percent CI condition in particular, had very similar endorsement rates of this statement to participants in the Splats condition, indicating a strong misconception that size of glyph indicates size of hurricane.

There does not seem to be strong misconception that the size of the glyph in the 66 percent CI condition indicates intensity (see responses to Statement 6). Optimistically, most participants in the Icon display condition correctly indicated that this display showed intensity and were significantly different from the 66 percent CI condition, $\chi^2(1, N = 62) = 47.23$, $p < 0.001$. However the CI Bands condition, $\chi^2(1, N = 65) = 10.11, p = 0.001$, and the Splats condition, $\chi^2(1, N = 62) = 25.06, p < 0.001$, were also more likely (than the 66 percent confidence interval group) to endorse the statement that the displays showed intensity of the hurricane, although neither of these displays actually shows intensity.

Finally, although none of these displays shows the passage of time, participants in the two animation conditions (Splats, $\chi^2(1, N = 62) = 16.93, p < .001$; Icon, $\chi^2(1, N = 62) = 16.93$, $p < 0.001$) were more likely to endorse Statement 7 compared to the 66 percent CI condition. This result suggests a possible misconception that change of time in a display indicates change over time in the forecast.

5.4 Experiment Summary

Overall, our analysis revealed that the CI Bands and Icon visualizations were interpreted as changing more gradually as a



(a) 66% CI v.s CI Bands (b) 66% CI v.s Icons (c) 66% CI vs Splats

Fig. 13. Mean damage ratings as a function of distance from the center of the storm. Each visualization is compared to the 66 percent CI.

TABLE 3
Percentage of Participants in Agreement (Out of Total Answering)

Statement	66% CI	CI Bands	Splats	Icons
1) The forecast shows a distribution of possible hurricane locations	63.6%	87.5%	96.6%	82.8%
2) The forecast shows the area that has a chance of being damaged	78.8%	87.5%	93.1%	92.9%
3) The forecast shows where the center of the hurricane has a chance of being	87.9%	87.5%	93.1%	92.9%
4) The forecast visualization indicates that the forecasters are not certain about the location	66.7%	68.8%	62.1%	44.8%
5) The forecast visualization indicates the size of the hurricane	57.6%	40.6%	58.6%	37.9%
6) The forecast visualization indicates the intensity of the hurricane	9.1 %	43.8%	71.4%	96.6%
7) The forecast visualization indicates the passing of time	0.0 %	6.3 %	41.4%	41.4%

Roughly 1/3 of the respondents were in each visualization category.

function of distance than the 66 percent CI visualization. This suggests that the former two were interpreted more like an uncertainty distribution than the 66 percent CI, and is similar to the effect seen with the path ensemble displays used in the earlier experiment by Ruginksi et al. [4]. However, the Splats visualization was interpreted similarly to the 66 percent CI, both in terms of average overall damage rating provided across all trials and the change in damage ratings that occurred as a function of distance-from-center. Although the CI bands and Icon visualizations were interpreted as more like an uncertainty visualization, and alleviated some misconceptions, answers to the questionnaire showed that other misconceptions about these displays remained. For example, many viewers saw the animated displays as indicating the passage of time. An interesting and unexpected result was that, even though they were not visually similar, the 92 percent CI and Icon displays elicited remarkably similar judgments. This indicates that decisions are likely driven by the spatial properties, not just the ensemble or summary nature of the display, and suggests a direction for future study.

6 Conclusion

We first made the case that the direct display of an ensemble of predictions can be superior to a statistical summary display for conveying the uncertainty in a prediction. The problem is that such displays can be cluttered and difficult to interpret if the original ensemble is too large. We then presented a sampling approach that extracts a representative subset of an ensemble of points while maintaining the statistical spatial distribution of the full ensemble, and showed how such a subset can be used to construct an ensemble-based visualization, supporting the superimposition of additional information on each sample via glyphs. As a byproduct, we also showed how the subset can be used as a basis for interpolation to create nicely-structured summary displays.

We have demonstrated our approach using hurricane predictions as an example, showing how our ensemble approach can be used to make time-specific visualizations of hurricane position, strength, and size. None of the currently produced visualizations from the National Hurricane Center are either time-specific, or able to simultaneously show the distribution of multiple storm attributes.

Additionally, we conducted a cognitive experiment to compare our ensemble visualizations to summary displays by examining participants' estimations of potential storm damage. The results indicate that an ensemble display using icons showing storm intensity, based on a representative subset of samples, can be effective in helping users to more correctly estimate the risk-which inherently includes estimating both the likelihood of a storm hitting and the intensity of the storm. This result is consistent with earlier experiments showing that ensemble displays are effective in reducing confusion between uncertainty and storm size.

In choosing subsets of an ensemble to support our visualizations, we explored both the OLS and WSE selection algorithms. These produce subsets with different spatial characteristics. OLS subsets are tightly distributed, with few outliers. WSE subsets, on the other hand, are less tightly distributed and include some outliers. For the hurricane prediction visualization problem, the tight distribution of OLS is good at showing the most likely location of the storm, and is useful for supporting interpolation-based summary displays. The broader distribution of WSE has the advantage of making the viewer aware of the high uncertainty in the prediction, and is most useful for supporting ensemble-based displays. For future work, we would like to explore a hybrid of the two algorithms, choosing a subset that emphasizes layout centrality, while preserving heightened awareness of outliers.

While our subset selection method preserves spatial distribution, there is no guarantee that it preserves the distribution of ancillary information, such as storm strength and size. We are examining approaches to preserving the distributions of all information, while still supporting good spatial layout.

Acknowledgments

The authors would like to thank Ricardo Gutierrez-Osuna of Texas A&M University, Edward Rappaport and Mark DeMaria of the National Hurricane Center, and Matthew Green of FEMA for key discussions and advice. The work reported in this paper was supported in part by the US National Science Foundation under Grant Nos. IIS-1212501, IIS-1212577, and IIS-1212806.

References

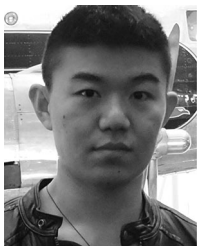
[1] K. Potter, P. Rosen, and C. R. Johnson, "From quantification to visualization: A taxonomy of uncertainty visualization approaches," *IFIP Adv. Inf. Commun. Technol. Series*, 2012, pp. 226–249.

[2] D. Stephenson and F. Doblas-Reyes, "Statistical methods for interpreting Monte Carlo ensemble forecasts," *Tellus*, vol. 52A, pp. 300–322, 2000.

[3] A. Pang, "Visualizing uncertainty in natural hazards," in *Risk Assessment, Modeling and Decision Support*, vol. 14. Berlin, Germany: Springer, 2008, pp. 261–294. [Online]. Available: http://dx.doi.org/10.1007/978–3-540-71158-2_12

[4] I. T. Ruginski, et al., "Non-expert interpretations of hurricane forecast uncertainty visualizations," *Spatial Cognition Comput.*, vol. 16, no. 2, pp. 154–172, 2016.

[5] R. Whitaker, M. Mirzargar, and R. Kirby, "Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2713–2722, Dec. 2013.
[6] M. Mirzargar, R. Whitaker, and R. Kirby, "Curve boxplot: Generalization of boxplot for ensembles of curves," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2654–2663, Dec. 2014.
[7] L. Liu, M. Mirzargar, R. Kirby, R. Whitaker, and D. House, "Visualizing time-specific hurricane predictions, with uncertainty, from storm path ensembles," *Comput. Graph. Forum*, vol. 34, no. 3, pp. 371–380, 2015.
[8] N. H. C. NOAA, "Definition of the NHC track forecast cone," Dec. 2014. [Online]. Available: http://www.nhc.noaa.gov/aboutcone.shtml
[9] K. Broad, A. Leiserowitz, J. Weinkle, and M. Steketee, "Misinterpretations of the "cone of uncertainty" in Florida during the 2004 hurricane season," *Bulletin Amer. Meteorological Soc.*, vol. 88, no. 5, pp. 651–667, 2007.
[10] L. Gedminas, "Evaluating hurricane advisories using eye-tracking and biometric data," Master's thesis, Faculty Dept. Geography, East Carolina Univ., Greenville, NC, Jun. 2011.
[11] J. Cox, D. House, and M. Lindell, "Visualizing uncertainty in predicted hurricane tracks," *Int. J. Uncertainty Quantification*, vol. 3, no. 2, pp. 143–156, 2013.
[12] L.-Y. Wei, "Multi-class blue noise sampling," in *Proc. ACM SIGGRAPH*, 2010, pp. 79:1–79:8. [Online]. Available: http://doi.acm.org/10.1145/1833349.1778816
[13] H. Chen, et al., "Visual abstraction and exploration of multi-class scatterplots," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1683–1692, Dec. 2014.
[14] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Syst.*, vol. 2, pp. 321–355, 1988.
[15] J. C. Carr, et al., "Reconstruction and representation of 3d objects with radial basis functions," in *Proc. 28th Annu. Conf. Comput. Graph. Interactive Techn.*, 2001, pp. 67–76.
[16] M. D. Buhman, *Radial Basis Functions: Theory and Implementations*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[17] S. Chen, C. Cowan, and P. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.
[18] S. Chen, E. Chng, and K. Alkadhimi, "Regularized orthogonal least squares algorithm for constructing radial basis function networks," *Int. J. Control*, vol. 64, no. 5, pp. 829–837, 1996.
[19] M. J. L. Orr, *Introduction to Radial Basis Function Networks*, 1996. [Online]. Available: http://www.anc.ed.ac.uk/rbf/papers/intro.ps, Accessed on: Feb. 2015.
[20] R. L. Cook, "Stochastic sampling in computer graphics," *ACM Trans. Graph.*, vol. 5, no. 1, pp. 51–72, Jan. 1986. [Online]. Available: http://doi.acm.org/10.1145/7529.8927
[21] D. Dunbar and G. Humphreys, "A spatial data structure for fast Poisson-disk sample generation," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 503–508, 2006.
[22] M. McCool and E. Fiume, "Hierarchical Poisson Disk sampling distributions," in *Proc. Graph. Interface*, 1992, pp. 94–105.
[23] C. Yuksel, "Sample elimination for generating Poisson disk sample sets," *Comput. Graph. Forum*, vol. 34, no. 2, pp. 25–32, 2015.
[24] E. Fix and J. Hodges, Discriminatory analysis, nonparametric discrimination: Consistency properties, U.S. Air Force School Aviation Medicine, Randolph Field, TX, USA, *Project 21-49-004, Rep. 4*, Feb. 1951.
[25] R. Y. Liu, "On a notion of data depth based on random simplices," *Ann. Statist.*, vol. 18, no. 1, pp. 405–414, 1990.
[26] W. Schneider, A. Eschman, and A. Zuccolotto, *E-Prime user's guide*. Pittsburgh, Pa, USA: Psychology Softw. Tools, 2012.
[27] S. Raudenbush, Y. C. A.S. Bryk, R. Congdon, and M. du Toit, *Hierarchical Linear Modeling 7*. Lincolnwood, IL, USA: Scientific Softw. Int. Inc., 2011.
[28] S. Raudenbush and A. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*, vol. 1. Newbury Park, CA, USA: Sage, 2002.
[29] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Evanston, IL, USA: Routledge, 2013.
[30] D. Bauer and S. Sterba, "Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation," *Psychol Methods*, vol. 16, pp. 373–90, 2011.



Le Liu received the BS degree in computer science from Chongqing University, and the MS degree in computer science from Clemson University. He is working toward the PhD degree in the School of Computing, Clemson University. His research interests include visualization and computer graphics. He is currently working on developing novel techniques for visualizing hurricane forecasts from path prediction ensembles. He is a student member of the IEEE Computer Society.

Alexander P. Boone received the BA degrees in psychology and history from the University of Missouri, in 2010. He is now working toward the PhD degree in the Department of Psychological and Brain Sciences, University of California, Santa Barbara. He is widely interested in individual differences in spatial cognition of small and large scale phenomena, particularly where it concerns issues of human navigation.

Ian T. Ruginski received the BA degree in cognitive science and religious studies from Vassar College and the MS degree in psychology from the University of Utah. He is currently working toward the PhD degree in the Department of Psychology, University of Utah, specializing in visual perception and spatial cognition. His research interests include applying cognitive theory to uncertainty visualization design and evaluation, as well as the influence of emotional and social factors on perception and performance.

Lace Padilla is working toward the PhD degree in the Cognitive Neural Science Department, University of Utah. She is a member of the Visual Perception Spatial Cognition Research Group directed by Sarah Creem-Regehr, PhD, Jeanine Stefanucci, PhD, and William Thompson, PhD. Her work focuses on graphical cognition, decision-making with visualizations, and visual perception. She works on large interdisciplinary projects with visualization scientists and anthropologists.

Mary Hegarty received the BA and MA degrees from University College Dublin, Ireland. She received the PhD degree in psychology, in 1988. She worked as a research assistant for three years with the Irish National Educational Research Centre before attending Carnegie Mellon. She is a professor in the Department of Psychological and Brain Sciences, University of California, Santa Barbara. She has been on the faculty with University of California, Santa Barbara, since then. The author of more than 100 articles and chapters on spatial cognition, diagrammatic reasoning, and individual differences. She is an associate editor of the *Journal of Experimental Psychology: Applied and TopiCS in Cognitive Science* and is on the editorial board of the *Learning and Individual Differences* and *Spatial Cognition and Computation*. She is a fellow of the American Psychological Society, a former Spencer postdoctoral fellow, and the former chair of the governing board of the Cognitive Science Society.

Sarah H. Creem-Regehr received the MA and PhD degrees in psychology from the University of Virginia. She is a professor in the Psychology Department, University of Utah. Her research serves joint goals of developing theories of perception-action processing mechanisms and applying these theories to relevant real-world problems in order to facilitate observers' understanding of their spatial environments. In particular, her interests include space perception, spatial cognition, spatial transformations, embodied cognition, and virtual environments. She co-authored the book *Visual Perception from a Computer Graphics Perspective*, previously was an associate editor of the journal *Psychonomic Bulletin & Review*, and currently serves as an associate editor of the *Journal of Experimental Psychology: Human Perception and Performance*.

William B. Thompson received the ScB degree in physics from Brown University and the PhD degree in computer science from the University of Southern California. He is a professor in the School of Computing, University of Utah. Prior to moving to Utah, he was on the faculty of the Computer Science Department, University of Minnesota. His current research lies at the intersection of computer graphics and visual perception, with the dual aims of making computer graphics more effective at conveying information and using computer graphics as an aid in investigating human perception. This is an intrinsically multi-disciplinary effort involving aspects of computer science, perceptual psychology, and computational vision. He has also made significant contributions in the areas of visual motion perception and in the integration of vision and maps for navigation. He has been an associate editor of the *ACM Transactions on Applied Perception* since the founding of that journal. He is the lead author of the *Visual Perception from a Computer Graphics Perspective*.

Cem Yuksel received the PhD degree in computer science from Texas A&M University, advised by John Keyser and Donald House. He is an assistant professor in the School of Computing, University of Utah, and the founder of Cyber Radiance LLC, a computer graphics software company. Previously, he was a postdoctoral fellow with Cornell University, under the guidance of Doug James. His research interests include computer graphics, including physically-based simulations for both offline and real-time applications, rendering techniques, global illumination, sampling, GPU algorithms, modeling complex structures, and hair modeling, animation, and rendering.

Donald H. House received the BS degree in mathematics from Union College, the MS degree in electrical engineering from Rensselaer, and the PhD degree in computer science from the University of Massachusetts, Amherst. He is a professor and chair of visual computing in the School of Computing, Clemson University. Previously he was a founding member in the Department of Computer Science, Williams College, and the Department of Visualization, Texas A&M University. His research spans computer graphics and visualization, with special interests in physically based animation, and perceptual and cognitive issues in visualization. He is coeditor of the seminal book on cloth simulation, *Cloth and Clothing in Computer Graphics*, and coauthor of *Foundations of Physically Based Modeling and Animation*, which is currently in press. He is a student member of the IEEE Computer Society.