Ensemble Visual Analysis Architecture with High Mobility for Large-Scale Critical Infrastructure Simulations

Todd Eaglin and Xiaoyu Wang and William Ribarsky and William Tolone
Charlotte Visualization Center, Department of Computer Science The University of North Carolina at Charlotte, Charlotte, NC 28269, USA

ABSTRACT

Nowhere is the need to understand large heterogeneous datasets more important than in disaster monitoring and emergency response, where critical decisions have to be made in a timely fashion and the discovery of important events requires an understanding of a collection of complex simulations. To gain enough insights for actionable knowledge, the development of models and analysis of modeling results usually requires that models be run many times so that all possibilities can be covered. Central to the goal of our research is, therefore, the use of ensemble visualization of a large scale simulation space to appropriately aid decision makers in reasoning about infrastructure behaviors and vulnerabilities in support of critical infrastructure analysis. This requires the bringing together of computing-driven simulation results with the human decision-making process via interactive visual analysis. We have developed a general critical infrastructure simulation and analysis system for situationally aware emergency response during natural disasters. Our system demonstrates a scalable visual analytics infrastructure with mobile interface for analysis, visualization and interaction with large-scale simulation results in order to better understand their inherent structure and predictive capabilities. To generalize the mobile aspect, we introduce mobility as a design consideration for the system. The utility and efficacy of this research has been evaluated by domain practitioners and disaster response managers.

Keywords: Disaster Forecast, Critical Infrastructure Simulation, Visual Analytics, Mobile Interface

1. INTRODUCTION

Forecasting the destructive impact for a volatile hurricane on a network of vulnerable critical infrastructure network is a central challenge for emergency planners and responders in hurricane-prone areas. After witnessing the devastating destruction from Hurricane Sandy, decision makers in coastal US cities are on high-alert for threats to their critical infrastructures (e.g, power lines, food networks, shelters, etc.); they are requesting more robust simulation analyses to depict the potential impacts from another tropical storm.

Much prior research has focused on using simulations and predictive modeling to anticipate hurricane movement and suggest possible landfall and impact locations.\textsuperscript{1–3} Due to the complexity and scalability of simulation runs, understanding these modeling efforts and their predictive capabilities from large collections of simulation results is challenging.\textsuperscript{4} On the one hand, many of the traditional hurricane modeling approaches depend on a trial-and-error approach that is not always feasible for generating simulations consistently. While this type of approach is widely adopted, each simulation run requires analysts to fine-tune the parameters, which can be very time consuming and less methodological.\textsuperscript{5}

On the other hand, a new simulation approach is based on the idea of data-farming, which prepares many possible simulation outcomes in bulk.\textsuperscript{6} However, this approach is largely limited by the simulation models, for which very few modelers have sufficient computing resources available to do sensitivity studies, validation and verification, effectiveness analysis, and related necessary activities. Exacerbating this challenge is that the large amount of such simulation results is far outpacing decision makers capability to analyze and make use of them.

Further author information: (Send correspondence to Todd Eaglin)
Todd Eaglin: E-mail: teaglin@uncc.edu, Telephone: 1 704 687 7989
Xiaoyu Wang: E-mail: Xiaoyu.Wang@uncc.edu
William Ribarsky: E-mail: ribarsky@uncc.edu
William Tolone: E-mail: William.Tolone@uncc.edu
To meet the challenges of dealing with disaster forecast and preparation, automated simulation methods are essential. However, they are not sufficient. There must be human input and direction, specifically for the interpretation of results and in some cases for directing the simulations. It is important for the decision makers to gain enough actionable knowledge such that they can respond in a timely and correct manner to possible failures of crucial infrastructure.

A key design component for our system is *Ensemble Visualization* of a large simulation space generated as a result of the aforementioned data-farming simulation approach. An ensemble, in this case, is a high-dimensional collection of attributes aggregated from raw simulation results that are centered around a single feature (e.g., Power Station, Airport, or Hospital). Details of the attributes are illustrated in Figure 2.

Another key design consideration is *Mobility*, which, in our case, refers to mobile computing devices or environments that enable analysis in the field. It isn't limited to just personal devices (e.g., iPad), but more broadly includes moveable equipment in general. The demand for mobility has been demonstrated in our previous police and evacuation exercises, where our campus police employed a networked tablet, smart phones, and a mobile command center to provide situationally-aware support.

Central to the goal of our research is, therefore, the use of ensemble visualization of a large scale simulation space to facilitate decision makers reasoning about impact on critical infrastructures with mobility. It is in this spirit that we architected our scalable visual analytics system for analysis, visualization and interaction in order to better understand their inherent structure and forecast capabilities. Our system captures the interplay between cascading simulations of critical infrastructure, ensemble analysis, and a networked mobile visual analytics interface. It aims to balance both human and computer intelligence and provide situational awareness to decision makers in both planning for natural disaster response and taking direct action during a disaster.

### 1.1 Research Procedures and Aims

In line with our design goals, the first activities of our research focused on establishing a simulation space based on computationally traversing possible simulation permutations within a Cascading Infrastructure Simulation (CIS). Specifically, our system relies on massive amount of computer generated simulation results. Access to these results is crucial because a working system requires information about the variance of infrastructure impact within large geospatial areas, which are produced from the cascading model as input parameters are varied.

As detailed in Section 4.1, we have developed a CIS model that runs multiple hurricane paths to generate a raw simulation space with millions of infrastructure failures events. Currently, our simulation includes infrastructures such as electric power, telecommunications, water, and railroad transportation. Large amounts of failures events from these infrastructures provide the fundamentals for ensemble analysis, which enables a multi-scale exploration of the high-dimensional simulation space connecting the complementary insights from global and local analysis of the data. In addition, our simulation space is situational because it depends on the properties of the particular coupled infrastructure and also on the properties of the occurrence (which can vary) that brings stress to the critical infrastructure.

By using an ensemble of simulation runs, we enable users to sample the space of input conditions that is presumed to cover all possible starting conditions in a particular range, and then employ multiple models that provide greater or lesser fidelity in some aspect of the process. The resulting ensemble encompasses the range of plausible outcomes, and the variation within the ensemble members exposes information on simulation uncertainty and the sensitivity of input parameters.

To incorporate human-centered analysis, we further provide an interactive visual analysis interface for exploring the simulation space. The designed visual interface combines a variety of statistical visualization techniques to allow decision makers to quickly identify areas of interest, ask quantitative questions about the ensemble behavior, and explore the uncertainty associated with the data.

To meet the needs for mobility, we have encapsulated the interface into mobile devices, such as an iPad, with multiple coordinated visualizations to provide a cohesive view of the simulations and permit analysis at multiple scales. The coordinated visualizations and interactions are optimized for mobile display, following the design guidelines from Apple. The complexity of the ensemble data will be mitigated by a flexible organization of the
information, and the coordination between views will permit users to focus on the formulation and evaluation of hypotheses. Specifically, our system is designed to help domain users address the follow questions:

- Which facility is more vulnerable than others?
- When will the facility be hit?
- How can responders in the field check buildings along a hurricane path in their local area?
- How is the hurricane’s temporal development affecting the cascading infrastructure failure?

The rest of the paper is structured as following. We first characterize the research domain in Section 2, then describe related work in Ensemble Visualizations in Section 3. Details of our system will be introduced in Section 4. We provide our informal evaluation with disaster prevention planners in Section 5 and provide discussions in Section 6.

2. DOMAIN CHARACTERIZATION AND DESIGN CONSIDERATIONS

2.1 Ensemble Analysis of Simulation Space

The architecture we present in this paper aims to help monitor and adapt to infrastructure changes by addressing the hurdle brought by the sheer size and heterogeneity of the relevant critical infrastructure simulations for the state of North Carolina. Recent disasters have highlighted the great vulnerability of both coastal and inland communities, such as power outages and road blockages caused by Hurricane Irene and Hurricane Isabel (which surged inland through Charlotte). It is important to identify the weak links within the massive infrastructure networks before the next hurricane hits. Road maintenance plans, development policies, hazard mitigation, and emergency response plans depend upon an understanding of the scale and linked cascading effects from hurricane impacts on critical infrastructures (e.g., links between power substations and water services).

Managing critical infrastructure simulations is a complex, multi-stage process. Understanding the impacts of stress on infrastructures requires an effective workflow that includes data acquisition, de-noising, analysis, visualization and interaction. Based on feedback from our emergency response collaborators, one major challenge within the CIS process is that available tools are inadequate for analyzing and managing such large simulation results. Based on a survey conducted by the National Oceanic and Atmospheric Administration (NOAA), 57 out of 198 coastal infrastructure managers showed the need for a decision-support systems that can help them monitor the potential impacts for each infrastructure from all simulation results.9

Our ensemble visualization is, therefore, setup to help understand how changes in the hurricane course and properties alter the stability of infrastructures. Our emphasis is on providing the ability to effectively and (semi-)automatically depict temporal and geospatial changes of coastal infrastructure caused by both historical and simulated natural disasters. Furthermore, we aim to enable users to view and interact with the simulations datasets, as well as the analysis results, in an unencumbered and intuitive fashion.

2.2 Interactive Visual Analytics Interface with Mobility

Mobility is another important aspect in our research effort. Through prior work in evacuation exercises, we have first-hand experience with campus police and DHS emergency planners to design and deliver networked mobile visualizations system using an iPhone.7 Specifically, the campus police chief and his force requires on-the-go analysis while events are occurring. The police chief stressed the benefits to have a mobile app on his iPad that he could use in conjunction with the setup he already employs in the mobile command center. This would greatly improve the police response time and keep them constantly connected while in the field under the coordination of the chief.

Other portable devices like laptops are not as mobile as a tablet and are cumbersome for the first responders as they move around on foot. A tablet is easy to carry around and allows the responder to access what he needs at any time and place. In this regard, the chief even went further and claimed the analysis system with mobility as an invaluable tool to integrate into his current methods.
During the first two years of our collaborations with responders and planners, we explored some of the simulation and data analytics challenges, and considered it critical to identify the most vulnerable infrastructure before a disastrous event, so that natural disaster impacts can be planned for and mitigated. Motivated by the above domain characteristics, our partnership, therefore, focuses on research on ensemble visual analytics over massive amount of infrastructure simulation results, with the goal of providing mobile visual interfaces that support analytic workflows to yield accurate and actionable results.

3. RELATED WORK

Our work is built upon the concept of Ensemble Analysis. A specific research area that deals with ensemble data sets is climate and weather data visualization. An ensemble dataset is defined as a collection of simulation features that are generated by computational simulations of one or more state variables across space and time. As temporal analysis assumes an ordered sequence of functions, a slightly different analysis comes into play when no ordering is imposed on a set of functions. This type of data often arises not from time-series data, but rather from an ensemble of simulations. For example, in our targeted critical infrastructure simulation, stochasticity or randomness within the simulation setup may result in different observations following different runs. Such varying observations form an ensemble of functions.

The variation among the ensembles arises from the use of different numerical models, input conditions, and parameters. The complex nature of ensemble data sets leads to numerous possible approaches to visualization. Multivariate correlation in the spatial domain is a common approach for reducing the complexity of the task of data understanding, as is reducing the data to a hierarchical form which is conducive to 2D plots.

Büerger and Hauser present an overview of techniques for multivariate data and Buja et al. discuss a taxonomy of interaction with high-dimensional data. One important challenge is to understand how the stochasticity or randomness within the simulation design impacts the infrastructure analysis of multiple runs. This could be considered as an uncertainty analysis that explores the relationship between input parameters and the output. One strategy is to derive a statistical description of these ensemble structures: for example, by summarizing each simulation run and then describing the statistical distribution of the entire ensemble simulation space. There has been some recent efforts in conducting statistical ensemble research, such as computing certain statistics on topological summaries. Ensemble analytics and visualization is a new and rapidly growing field where researchers in statistics and computational topology have just begun investigations, and a wealth of questions remain open. Software systems, such as SimEnvVis and Vis5D, are designed to handle atmospheric data formats and include 2D geographical maps with color maps and contours, as well as more sophisticated techniques such as iso-surfacing, volume rendering, and flow visualization. The Noodles system provides the aforementioned capabilities and adds uncertainty contours and glyphs. In addition, Ensemble-Vis uses meteorological data to provide a visual exploration of short-range weather forecasting. It adapts visualization methods common to the domain of meteorology; it also adds indications of uncertainty and enables the user to drill-down to ensemble data.

Given the need to integrate human knowledge in the analysis process, our work also relies on ensemble feature integration. Here, we focus on the combination of interactive visualization and feature analysis that has resulted in a set of feature integration and exploration techniques for analyzing multi-dimensional simulation spaces. Prior work in this area can be grouped into three categories, two-variate visualization, multivariate visualization, and analysis animation, depending on the nature of the simulation data. The utility of these
techniques has been demonstrated in the analysis fields of volumetric data, terrain change detections, and medical imaging. However, these traditional feature integration techniques no longer scale with the increasing size and complexities of the simulation datasets. Our research aims to address several challenges in this research area including feature selection and comparison, interactive exploration and semi-automated annotation.

By integrating knowledge gained from our previous work done for a mobile crowd sourcing application to model 3D buildings, we established our networked visual analytics architecture to build a better way to handle such mobility requirements while supporting data traffic, user interaction, and visualizations. Multiple linked views of data relieve the need to present all data of interest in a single window. Such approaches let the user interactively select regions of interest and reflect those selections in all related mobile views. The resulting mobile interface organizes a collection of views to provide complex investigation of the data under user control.

4. SYSTEM COMPONENTS AND IMPLEMENTATION

The focus of our system design is to aid domain experts and planners in analyzing millions of simulation results in order to identify the vulnerabilities and resiliency of a critical infrastructure. As shown in Figure 1, our networked visual analytics system comprises three integrated components: A cascading CIS model connected to a shared Oracle database that generates the raw simulation space (Section 4.1); an ensemble analysis module that handles statistical computation of ensemble structures and provides an optimization and data retrieval API for the visual analysis module (Section 4.2); and a mobile visual analysis interface that encodes the ensemble results into multiple coordinated visualizations for conducting field examinations and preparations (Section 4.4). The following sections describe each of the components in greater detail.

4.1 CIS: Generating the Raw Simulation Space

To capture the complex, multidimensional, and dynamic effects from a hurricane, we utilize a CIS simulation model that takes into account the interrelationships among critical infrastructures. Built within a rule-based framework for integrating multiple infrastructure components at a high level, our approach allows the user to create a model which represents the users assumptions about the world. This results in a dependency/interdependency ontology. Thus, for example, a breakdown of a power substation would immediately cascade to power loss at points on its distribution network. If a school were a node in the distribution network, it would be switched to backup power that, after a given time, would also shut down.

In addition, our model handles the concept of output requirements which allows the user to specify which inputs are necessary to produce the specified output from the feature. This is especially useful to help users gain focus on the commodities that matters most to their planning or response amid the noise of the overall simulation. For example, a power-grid specialist can select a subset of inputs that only applies to the multi-linked power station within his jurisdiction and run partial simulation around it.

4.1.1 Model Inputs and Simulation Concepts

Our CIS model consists of a wide range of data inputs, such as features, commodities, relationships, networks, and latencies. To keep the model results as realistic as possible, we conducted, over a 6 month period, a thorough collection of infrastructure datasets within the Carolinas.

Specifically, features are the physical infrastructures involved in the model such as communication features (e.g., cell towers), electric network features (e.g., power substations), and services features (e.g., hospitals and restaurants). In our current implementation there are 48 infrastructure including public schools and shelters.

The majority of this data comes from FEMA HAZUS data disks from 2005. The cellular towers, communication facility and TV stations were provided by the FCC. We further generated the switch control data for the communication network based on ESRI’s random point generator. We have acquired or generated a rich set of electrical network features. Not only did we acquire a set of HAZUS data points for power generation facilities, we had also generated a human-annotated network features by employing undergraduates to digitize the transmission power lines and the substations from orthophoto satellite images. Moreover, the service feature data is largely collected through out domestic and international collaborators. The food features where provided by a project partner.
Networks are, therefore, systems of infrastructure features that all share similar connection properties. The electric grid, road, rail, and pipeline system are all modeled as different networks with distinct properties that best reflect their actual applications.

To simulate the cascading effect, our model takes in the concept of a commodity, which is a tangible or conceptual good flowing between selection sets. In total, we have collected a total of 15 commodities such as electric, water, natural gas, manufactured goods, personal relationships, and wealth.

4.1.2 Cascading Relationship for Model Accuracy

This collection of networks and commodities gives us a good foundation for performing realistic cascading critical infrastructure simulations. Within each network, a commodity flows between two features through the creation of a relationship. The relationship establishes the provider and consumer of the commodity, the criticality of the good transfer and the method for establishing relationships. In our current model, we have pre-defined over 80 relationships that aims to make the model results as close to reality as possible.

Thus, according to the definitions above, the interlaced critical infrastructures are captured in a set of networks with each node having a set of properties according to its category with the edges providing a dependency rule according to the category and state of the two connected nodes. For example, a breakdown of a power substation would immediately cascade to power loss at points on its distribution network. If a hospital were a node in the distribution network, it would be switched to backup power that, after a given time, would also shut down. In this way, our CIS model takes cascading events into account.

To enable users to hypothesize different cascading conditions, we have built four methods to create a relationship: explicit, nearest neighbor with/without redundancy, and spatial. The users can directly specify the direct link between two features through an explicit relationship between the origin and destination. While the explicit method specifies a strict point-to-point relationship, the users can also use the nearest neighbor method to create connections between the origin and destination features that are within a specified distance. The redundancy option is built to further narrow down the connectivity between features. A relationship with redundancy means that the destination node is connected to all sources within the distance (as in the case of cell towers), while without redundancy means the destination node is only connected to the closest provider (as in the case of water

<table>
<thead>
<tr>
<th>name</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>Feature ID stored in DB</td>
</tr>
<tr>
<td>Object_Class_Name</td>
<td>Commodity Type (e.g. Fire station)</td>
</tr>
<tr>
<td>Simulation_ID</td>
<td>Index Number for Simulation</td>
</tr>
<tr>
<td>Commodity_Name</td>
<td>Type of the Commodity</td>
</tr>
<tr>
<td>Indirect_Cost</td>
<td>Is the commodity disable indirectly?</td>
</tr>
<tr>
<td>Direct_Cost</td>
<td>Is the commodity disable directly?</td>
</tr>
<tr>
<td>Relative_Time_In_Hours</td>
<td>Time of Disablement</td>
</tr>
<tr>
<td>Number_Hours_Disabled</td>
<td>How long is the commodity disable during the simulation</td>
</tr>
<tr>
<td>Cause_Event_ID</td>
<td>Disable Event ID</td>
</tr>
<tr>
<td>COA_ID</td>
<td>Cause of Action ID</td>
</tr>
<tr>
<td>Model_ID</td>
<td>Which Hurricane Path Model</td>
</tr>
<tr>
<td>Network</td>
<td>Which Infrastructure Network (e.g., Power or Transportation)</td>
</tr>
</tbody>
</table>

Figure 2. Detailed dimension information for each Ensemble.
pipelines). For example, a hospital could connect to any sewer pipe within 1000 meters but it only creates a relationship with the sewer pipe closest to it. Finally, a spatial relationship is an instance where the origin and destination share a relationship based on proximity of their location to the nearby features. In our model, all electric relationships are based on spatial relationships.

4.1.3 Simulation Runs and Results

Our simulation runs with on a set of Courses of Actions (CoAs), which are a pre-determined encapsulation of the above factors to reflect the state changes for particular critical features in the storm path. Through a CoA, a user can specify events that occur during the simulation and inform the CIS of What (selection set instance) undergoes a state change (e.g., intact or disrupted) and When (relative time into simulation) this occurs. The user can make the CoAs permanent or temporary. A permanent event means that the feature will stay at the new state until another CoA event changes the state. A building that has been physically destroyed is disabled and cannot be re-enabled even if all of the buildings inputs are available to it until the user specifies that the building has physically been rebuilt.

Latency is an important part of the CoA, which is designed to represent the delays before a selection set undergoes a state change. Latencies can account for backup electrical power or, in the case of food stores, the time periods until food supplies are exhausted. For example, a hospital can have a disablement latency of 24 hours for blood supply but a 14 day electric disablement latency.

We further constructed six automatic model variations to extend the conditions that are simulated. This set of model variations allows the users to test their assumptions on distance and the importance of redundancy within the system. Specifically, the six model variations only represent changes to these relationships but not to any other aspect of the model. The first model in each set is the base model which represents the users’ best guess on structure and type of relationship between the selection sets. The second model converts all nearest neighbor without redundancy (NN) and converts them to nearest neighbor with redundancy (NNR). The third model doubles all relationship distances. The fourth model takes those new distances and converts any NN to NNR. The fifth model takes the distances in the base model and halves them. The sixth model takes those new distances and converts the NN to NNR. Note, NNR is only applicable to where redundancy makes sense (e.g., not water pipes).

In order to cover the simulation space, we run a large number of simulations with different hurricane paths and different intensities and spreads. In total, we have run the six models across 12 different hurricane paths, including six historic hurricanes paths (see Figures 3) and six notional paths. Doing so allowed the users to gain a comprehensive view of what happens for different storm strengths or by changing our damage radii assumptions. The hurricane paths were created together with our collaborators and give our simulation model a good coverage across the Carolinas. The different intensities and spreads give a comprehensive view of what can happen under different conditions.

With over 50,000 features, so far, we have completed 425 simulations over the 12 hurricane paths and created 14.6 million simulation events. As shown in Table 2, each event is a high-dimensional collection of fields that are centered around a single feature (e.g., Power Station, Airport, or Hospital). All the data is stored in an Oracle database for remote access and analysis.

4.2 Extracting Ensembles from the Simulation Space with Statistical Analysis

While our CIS methods are essential for creating a large simulation result space, this is not sufficient. There must be human input and analysis, specifically for the interpretation of results. It is important for the decision makers to gain enough insights to form actionable knowledge so that they can respond and react effectively to possible failures of crucial infrastructure. This is the primary reason for us to enable interactive ensemble visual analysis of the raw simulation space.

Our ensemble visual analysis is used to abstract and reduce the complex and vast amount of information for each infrastructure. This enables the decision maker to understand the full crisis in its context and to detect potential cascading effects. It further permits the users to select the paths and CoAs most likely to occur for the most likely range of conditions. This will in turn indicate the most likely parts of the infrastructure to be
disrupted and show the likely cascading effects. For example, this gives city emergency planners information on “how likely is my airport going to be hit by this storm and when will this occur?”

We define our ensemble as a high-dimension collection of fields aggregated from raw simulation results that are centered around a single feature. Specifically, we compute an ensemble as a collection of disablements across the entire simulation space for each specific infrastructure. The key value is a probabilistic outage for each infrastructure feature. This value represents the likelihood for a specific infrastructure feature to be disabled at any point during the simulation. We also compute the breakdown of causes for a specific infrastructure feature along with the minimum, median and maximum monetary cost imposed by it being disabled.

In addition, multivariate correlation is used to capture the stochasticity or randomness within the simulation setup and reduce the complexity of the simulation event data. As a result, we are able to abstract and encapsulate the ensemble of every infrastructure. These infrastructures are represented by symbols, such as the glyphs (see Section 4.4.2) for power stations and cellular transmission stations.
Figure 4. This table shows the approximate time in seconds to perform kernel density estimation on varying data sizes in seconds. The spatial region was discretized into approximately 4 million cells. The mobile computation was performed on an iPad air using a CPU parallel implementation. The server computation was done using a Nvidia GeForce GTX 680MX using OpenCL.

4.3 Scalability Optimization for Mobile Analysis

Due to the inherent hardware limitations of commodity mobile devices, creating the needed mobile visual interface for millions of simulation events is both a computation and a visualization challenge. Where desktop computers can perform in-memory operations with large data, a commodity mobile device (e.g., iPad) is quite limited in its computing power. Therefore, a significant part of our research has been devoted to conducting scalable mobile optimization, which is heavily built around computing optimization and visualization scalability.

4.3.1 Offloading Ensemble Encoding to Computing Server

Our first optimization aims to accommodate the need to interactively visualize the entire ensemble space and its temporal changes. In this regard, we have employed a computing server that is designated to perform specific statistical calculations and data retrieval; these operations would otherwise be too time consuming on any mobile device.

Our computing server is connected to the database that contains all the raw simulation results. It handles data requests sent from the mobile device to the server, extracts the necessary data and finally computes the ensemble abstraction and encoding. We use a hash map for data transfer between the mobile device and the server. Each infrastructure has a unique key identifier. Using a hash map allows for quick lookups when doing additional computation on the mobile device.

In order to examine the temporal changes, our server performs statistical analysis on the ensemble through kernel density estimation (KDE). This is done using GPU parallel computing and significantly offloads the demand for performing on-device computing. We use OpenCL to compute each cell of the KDE. This provides exponential speed up, but we have gone a step further since copying the computed data back from the GPU to main memory is usually more time consuming then the computation itself. To minimize data transfer between main memory and GPU memory we take advantage of being able to hand off the data computed from OpenCL to OpenGL for rendering. In the OpenCL kernel we can create geometry and then render it very quickly with OpenGL since it all exists in GPU memory. This allows us to not just create geometry very quickly but also pre-rendered bitmaps.

4.3.2 Utilizing Multi-core and the Graphics Engine for Mobile Rendering

Some processing must be done on the mobile device. These tasks usually involve data loading and retrieval as well as user interaction. Handling data can often be very time consuming but slowing the user interface is not conducive to our goals. When interacting with a mobile device any latency or slowness results in users becoming frustrated and hinders their productivity. Mobile devices often have multiple cores like their desktop counterparts so it is absolutely crucial to utilize any available resources. We utilize parallelism for all our web requests to create asynchronous data pulls. We also load all data from the CPU to the GPU through a background process. We purposely leave the user interface to its own separate process in order to maintain a consistent experience.
We also built a low-level graphics engine specifically for rendering the ensemble simulation space. It enables the device to render vast data sets very quickly. The graphics engine uses a scene graph style data structure for storing elements. It also utilizes graphics techniques like double buffering and constructing texture atlases on the fly for improved rendering.

Moreover, mapping systems included on most mobile devices are not capable of displaying lots of data very quickly and efficiently. They also are not portable between different platforms making cross-platform development difficult. We chose to develop our own mapping system in OpenGL making it very fast but also portable between platforms.

4.3.3 Remote Communications
Our mobile system is dependent upon the data pulled from the computing server. Any latency with this communication will affect the entire system. Due to this dependency we rely extensively on compression and aggressive caching to memory and disk. This limits potentially costly network requests and speeds up the mobile application.

We started with web services to query data on the fly, package the data and send it to the device. This method worked well with moderate datasets; as the data grew, however, the query time and packaging time began to slow down greatly. The response time had reached a certain threshold and at that point the application became unusable.

We therefore created a persistent environment instead of solely relying on a temporary service that queries the database each time and repackages it. This persistent environment caches the packaged data in main memory. Querying the database can be improved through minor tweaks with caching and other features, but what substantially slowed down the entire process was that each time the web service is called a new instance is spun up in main memory to load the queried data in order to package it into a form usable by the mobile device.

We utilize parallelism for all our REST (Representational state transfer) requests to create asynchronous data pulls and provide three possible return types for scalability. Each type is specified by the client based on its needs and capabilities. The first is a list of all points returned from the query stored in a compressed JSON file. Returning all the points allows the client to run additional operations that might not be computationally constrained. The second is a pre-computed kernel density estimation that is computed into indexed triangulated geometry and compressed into binary for transfer. This data type is for a client system that cannot calculate a spatial KDE in sufficient time, but is capable of rendering the triangulated result. The triangulated result is resolution independent in order to provide a higher level of detail for visual analysis. The last type is a pre-rendered bitmap of the kernel density estimation. This is for client systems incapable of handling raw data points or even triangulated geometry.

All these scalability optimizations setup a feasible platform for us to encode and visually represent the ensemble analysis into a mobile visual interface.

4.4 Ensemble Visualization with Mobility
As the first step in our ensemble visualization design, we followed prior visual analytics design studies (e.g., \textsuperscript{34,35}) and collaborated with first responders to understand their workflows and analytics needs. Based on our discussions, it became clear that a key aspect of our effort was to provide them with a visual analytics system that can accommodate their in-field analytics needs. Since, we are focusing on a system with high mobility like a tablet device our visualizations and analysis tools must be simple enough to use with touch gestures and smaller screens.

4.4.1 Geospatial-Temporal Overview for Ensemble Analysis

Customizable Geospatial View: Our simulation data set revolves around geospatial information. While the mobile OS incorporates mapping services they are not specifically designed for displaying large data sets or any complex 2D drawing. For this reason we incorporated a lightweight tile map server specifically for our needs. A tile map server\textsuperscript{36} simply allows a client to fetch pre-rendered geo-spatial map tiles on the fly from a server and display them in an arranged fashion much like Google maps.
Our customized tile map system, as shown in Figure 5 (A), bears a similarity to commercial mapping systems, but also provide additional functionalities that fits our visual analytics needs. Since the map tiles are stored in the cloud and the visualization components are implemented as layers that can be stacked onto the map, our mobile interface remains lightweight. This enables us to completely customize the maps and it also allows us to imbed certain features in the tile map textures to reduce the redundancy of rendering by having all the essential map details baked into pre-rendered tiles.

As shown in the video, we provide two different visualizations. The first is a time-based color coding of each individual infrastructure based on the average time it went out. This allows first responders to see when specific disabling events occurred during a hurricane. The second visualization is a Kernel Density map. The intensity map shows the probability of outages for a particular region. This way users can quickly focus on regions with a high percentage of outages.

**Temporal Visualization:** Understanding temporal behavior of a disaster is another important aspect for emergency planners to isolate time critical infrastructure. We have developed an interactive temporal analysis view that enables the selection of specific time ranges to depict what infrastructure was disabled as the hurricane
Figure 6. This image shows the interactive Temporal View. In this view, users can depict the peak time of the outages and select a range of time to further their investigations.

passed through. As shown in Figure 6, the user can filter the simulation space in time and further examine the outages leading up to a peak or the peak itself.

Our system also enables the users to examine the result of cascading outages and interactively analyze the complex cascading relationship between different infrastructures. This allows the users to visually depict the interdependency of infrastructures and understand the effect of one infrastructure on others. Through the interactions, our collaborators were able to observe hurricanes with two distinct impacts patterns, namely the immediate sparse outages following high peaks of outages as well as the latency outage pattern where an infrastructure feature (e.g., hospital) will stay on with its backup generator for a longer period of time before becoming disabled. These are all key temporal factors for evacuation and planning efforts.

**Animation:** We have further setup animations to help the users more effectively analyze the outage patterns. Previous research has suggested animation plays an excellent role in revealing changes where cascading events take place. This gives an overview of the events that took place and hotspots during the simulation. However,
enabling animation posed a computational challenge given the 50,000 features and millions of events we have in our simulation space. As shown in Figure 4, computing each time segment for visualization exceeds the computation power of any existing commodity mobile device.

To address this challenge, we utilized streaming techniques that offload much of the computation to our server and streams the visualizations in quick succession to a mobile device. As shown in our video, we have created animation of all the simulation events over a thirty day period using spatial KDE.

In addition, the same three data types: data points, geometry, and pre-rendered bitmaps that are available for single queries are also available for animations. A client can request all the raw data points across the entire animation span or the triangulated geometry or pre-rendered bitmaps. This allows the animation to be flexible on a wide range of client needs and scalable even down to mobile devices.

4.4.2 Detailed View for Ensembles and Infrastructures

Each infrastructure in our simulation has a probabilistic chance that it will be disabled sometime across the entire simulation space. When the user zooms in close enough, we display a map glyph with the probability of each infrastructure as a pie chart, as shown in Figure 7. This allows the users to quickly see which infrastructure...
features have a higher probability of being disabled. More importantly each feature might be disabled from multiple events. By selecting an individual infrastructure a secondary pie chart will be displayed that shows individual causes for disablement broken down by percentage. In this way a user can isolate the main cause for the infrastructure being disabled.

The glyphs for each infrastructure are created on demand based on the type of infrastructure and the summation of its potential outage. Figure 8 shows a few examples of different types of glyphs. The inner icon is associated with the type of infrastructure. This allows users to easily discern what specific infrastructure they are looking at secondly, the outer ring displays the probabilistic chance the infrastructure feature is disabled. A full circle would indicate a 100 percent chance of being disabled. This outer ring allows users to quickly identify what features have the greatest likelihood of being disabled.

4.4.3 Interactions

Probe Selection: Probing involves interactions that allow users to visualize a selected region in a free-form manner. As shown in Figure 9, a user can directly draw onto the map with his or her finger, drawing a bounding area around a region. This is an extremely important analysis tool because it allows the user to drive the analysis and focus on what is important to their geographic area. Anything within the bounding region will then be analyzed and returned to the user in a separate window.

When selected, a separate window will appear that initially shows the time distribution for all the features within the region. A user can quickly see any peaks in time in this window. Secondly, the user can then view the breakdown of what feature classes are in the selection and how many were disabled (Figure 7). This allows users to see what features were disabled, how many were disabled and the time distribution of those features, permitting a very fast analysis of regions of interest. It also allows users to pinpoint features that have a high probability of being disabled. Most importantly it allows a user to compare multiple regions for joint analysis.

Each probe selected by the user is copied and saved for additional viewing, as shown in Figure 9(Right). This way users can select multiple probes across many different areas and filters. Users can probe different time segments as well as different hurricane paths and commodities. Along with the selected features a screen shot of the region is attached to the data to give geospatial context to the region that was selected. All this information is presented in a secondary window where each probe is represented as a card with a map image of the selection and a time distribution of the events from the probe.

When the user selects a specific card the view transitions to a detailed breakdown of that specific probe. In this way users can return to further analyze individual events and time distributions from prior selections they have made.

Filtering: In addition, filtering is a key function that allows for the removal of unwanted features or events, as shown in Figure 10. In such a large simulation space, it is crucial for a user to be able to focus on specific commodities and hurricane paths. More particularly the combination of the two together allows for complex
filters. This helps users navigate the simulation space by visualizing different parts of the simulation in a very easy manor.

The first type of filter is a commodity-based filter. Each commodity is separated into specific categories that share a common area. For example all the infrastructure features like ports, train stations, airports are separated into a transportation category. By having high-level categories, users can more easily pick what areas they would like to filter instead of navigating through lots of very specific features. Grouping similar features into categories also allows for easier comparison. We allow users to filter specifically on commodity types like transportation and electricity. This allows for quick comparison between different commodities.

The second type of filtering is hurricane path based. Our model has numerous hurricane paths that create a probabilistic outcome. Users can select specific hurricane paths to see the probabilistic effect of each impact. But, the real core of this feature is the filtering of multiple hurricane paths. By filtering on multiple hurricane paths pertinent to a region a user can get encompassing results for that region or quite specific results (e.g., what hurricane path is most likely to do damage to this particular infrastructure feature in my local emergency response area, and when).

4.4.4 View Coordination on Mobile Device

We took very careful consideration when designing view coordination on the mobile interface. As far as we know, no prior research has been done that studies multiple-coordinated views (MCV) for mobile. Due to the limited screen-space, our view coordination has to occur within the same display area, rather than in a spread out fashion as discussed in the original MCV guideline.38

We ultimately relied on the transparency of each view to make the view coordination apparent, as shown in the video for this paper. Since there is limited screen space on mobile devices the use of overlays and transparency
is necessary in order to maximize every inch of the screen. This allows the user to intuitively make selections and understand the effects of their changes.

Given our focus on the disaster response domain, we choose the geospatial view as the main entry view where all ensembles are presented at the initial stage. The coordinated views for temporal analysis and filtering are split into two transparent overlays. Specifically, the user is able to select either a filter view or a temporal view that contains a transparent overlay in order to see any changes made to the spatial representation. The user can also very quickly hide the overlays with a simple swipe gesture in order to return to spatial navigation.

5. CASE STUDY AND EXPERT FEEDBACK

Evaluating our system and collecting results through a traditional user study is challenging because the focus of our work is centered on both domain experts and responders. Such evaluation requires a group of responders from multiple domains working on a large collaborative disaster response exercise. Due to the limited availability of responders, we have demonstrated our application through simulated scenarios to determine our systems usability, effectiveness, and need for improvements. In this process, we conducted two evaluation sessions with three city
emergency responders and four power grid managers from a large regional power company. We thus are getting feedback from both responders and domain experts.

5.1 Forecasting Events prior to Hurricane Season

Scenario: The center focus of this scenario was Wilmington, NC, a coastal city that is historically vulnerable to hurricanes. Specifically, over 1 million cascading infrastructure events were examined over 8 hurricane paths that had directly passed through the greater NC region. As shown in Figure 9, our mobile ensemble visualization provided users an exploratory environment to identify where and what are the most vulnerable facilities.

In this case, an emergency responder was interested in isolating critical infrastructure that would likely cause financial distress if it were destroyed or rendered inoperable from a hurricane. The city planners need to specifically find these critical points like power substations and water services that could be very costly to the community if they are knocked out. They quickly examined the region using the probing function by drawing a bounding area around the different areas of the city. This action brought up a secondary window with the time distribution of events in that region, as shown in Figure 9. With multiple probing windows opened within the different quadrants of the city, the planners were able to see the aggregate view of the outages over time, but more importantly he was able to compare the peak times of outages side-by-side. Within the same window the planners were able to view the complete break down of events, which contains outages of airports, ports, highway bridges, and train stations. To one planners surprise, the airport in this region shows a significant outage at the early stages of our simulations, suggesting a high chance for this facility to be disabled given the simulated hurricane paths. This is just the sort of unexpected but quite important discovery that our system supports.

The planner hypothesized this may be caused by road blockage and direct impact and navigated through the interface and selected the detailed breakdown for the airport. He was surprised to see the report on what critical events caused the outage, as shown in Figure 7 (B); although it complies with his experience that the outage may be caused by direct impact (44%), there still was an unexpected 51.2% outage caused by Electric Service failure. This provided him more insights and helped him to take action on investigating weak links in power connections to the airport by working with power company participants.

User Feedback: One of our architecture’s advantages is the ability to compute massive amount of physically-based infrastructure simulations ahead of a hurricane season. All collaborators consider this is of great value as they can start their planning process a couple of months ahead and become fairly comfortable with the range of resiliency for their infrastructures. As summarized by one of the emergency planners, “this system gives us the ability to look forward and be prepared. It could help us to find a vulnerable building and plan ahead the scenarios when a hurricane actually hits”.

However, since our analysis is built upon simulations, managers from the energy company suggested further extensions to fuse heterogeneous datasets into the mobile ensemble visualization. In particular, they are interested in learning how we could effectively associate public information, such as census and demographics data.

5.2 Investigating Cascading Effort of Infrastructure Breakdowns

Scenario: Our mobile visual analytics environment enables users to explore and follow the impact caused by the evolution of a natural disaster. A city planner using the system inspected a couple of possible hurricane paths that made a pass through the Wilmington area (Figure 5).

Using the direct filtering methods, the user can quickly narrow down the analysis scope to the paths within close proximity to the city and examine critical areas along it. As a transportation expert, her area of focus is on how to bring back the transportation infrastructure after a catastrophic impact. Hence she further filtered down with transportation infrastructure and noticed an interesting spreading pattern that presents an elongated trailing effect as the effect of the hurricane is still impacting the infrastructure even a month after the initial landfall. She was surprised to see this trailing effect and suspected this may due to a less robust road structure. Indeed, a quick examination of the transportation network near the coast revealed that this may be one of the possible cascading effects due to the less redundant pathways from the outer coastal region to the inland. Once the hurricane passes through, getting relief and assistance back into the city requires outlets for transportation.
Our system allows the user to then apply a specific filter just on transportation to see just those particular events in conjunction with the hurricane paths selected already. In this scenario, the mobile interface permits the user to focus on a specific infrastructure breakdown of her particular analysis interests.

**User Feedback:** One of the benefits that all these users, both electric system experts and emergency planners, see in our architecture is its capability in helping to depict the overall impact on critical infrastructure as well as directly assess individual commodities (e.g., transportation structure). Using this visual interface, they can select specific hurricane paths that would pass through a certain city. Especially to planners, who are responsible for correlating information from various simulation models, the capability to identify and filter the infrastructures from a massive simulations space is of great value. As commented by a power grid manager that “...this is very useful for me to follow and compare my power lines with possible hurricane paths. I can then re-route my power transmissions away from the danger zone.”

Several collaborators mentioned the need to conduct search-by-example functions due to the limited time constraint as a disaster is unfolding. As stated by an emergency response manager that, “it would be great if we can select one [impact] pattern, and the system can automatically suggest other similar ones...this would save a lot of time for me to navigate through the whole dataset.” This requires our architecture to be able to perform comprehensive ensemble structuring, producing a more similar space for all the ensemble structure. His comment is well received, and we are working extensively on researching a quantifiable ensemble structure for ascertaining where attention is needed and resources should be deployed.

It is important to note that these rather complex analyses, building of hypotheses, and checking of the hypotheses were done on the mobile device in both cases. This indicates that the ensemble visualizer can be carried around for impromptu meetings among responders or planners and can also be carried into the field. Being able to look in detail at vulnerability of infrastructure in the field is quite important for both planning and response. Detailed actions can then be taken on-the-spot. In an actual emergency, a response team would be on hand to carry them out.

There is a final important outcome from our studies. We have shown that the mobile interface can be used to derive actionable knowledge and make decisions by both experts and by non-expert users. In the case studies above the experts were electric system managers, who would know a lot about the basis for the underlying critical infrastructure simulations and the non-experts were emergency planners, who know little about these simulations but must still make key decisions based on them. Both types of users were able to use the mobile interface effectively with a small amount of initial training. The details of the complex system outlined in Section 4 were hidden from both types of users. Yet they were available to the experts as needed. Indeed the experts could even control some details of the simulation through an interface that makes this complex activity much more straightforward. This important result shows the broad utility of the system we have set up. It can be used effectively by the range of people who will needs its capabilities in planning for or responding to large scale disasters, and it can support partnership and coordination between them.

### 6. LIMITATIONS AND FUTURE WORK

There are limitations to this research that must be addressed. Although our research was conducted towards the goal of ensemble analysis of simulation results, the complexity, the scalability and the mobility needs of the targeted dataset exacerbated the challenges on a computational-level and visualization-level. Especially, we found it very challenging when designing the multiple-coordinated views on a mobile device. While such research activity is upcoming, there is no gold standard to follow when considering the small screen space in mobile devices. We did attempt to mitigate the challenge a) by following mobile emergency design practices that we have accumulated experience by” and b) by projecting ensemble abstractions to spatial-temporal visualizations that can best utilize the mobile displays. Nevertheless, multiple coordinated visualization in different domains still requires additional design considerations. We acknowledge this challenge and are working closely with domain users to provide a more customized visual interface.

In addition, our scenarios and expert feedback are limited to the available resources within the project. A more in-depth assessment of our system utility can only be derived when it’s deployed in the field with the
responders. Further, in our study the effects of time pressure were not considered, which may have significant influence on decision makers (Section 5.2). We will look into incorporating techniques, such as search-by-example, to reduce user the time pressures.

In future work we would like to run simulations in the cloud that are driven from a mobile device. There are a number of limitations that prevent us from generating new simulations in real time. The first limitation is a computational problem. Running these exhaustive simulations is very time-consuming and would require new more efficient methods to run in real time. Secondly, we would have to develop effective methods for generating a simulation model on a mobile device and then communicate that model to our simulation environment in order to evaluate the simulation. Currently the simulation environment is a closed system that is self contained. Additional modifications and designs would have to be made to allow for two way communication with the simulation environment.

We recognize these limitations and consider the support for disaster forecast and preparation as an important visualization, analytics, and interaction research topic. We expect the presented approach illuminates the role that mobile ensemble interfaces play in such complex problem-solving environments and provides a platform for actual use.

7. CONCLUSION

Understanding large cascading simulation datasets is vital in disaster monitoring and emergency response. But, gaining insight from these simulations requires extensive tools and analysis in order to provide planners, emergency responders, and electric system experts with valuable information. To gain enough insights for actionable knowledge, we utilize ensemble visualization of a large scale simulation space. We have developed a general critical infrastructure simulation and analysis system for situationally aware emergency response during natural disasters. Our system demonstrates a scalable visual analytics infrastructure with mobile interface for analysis, visualization and interaction with large-scale simulation results in order to better understand their inherent structure and forecast capabilities. The utility and efficacy of our research has been evaluated by domain practitioners and disaster response managers.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2009-ST-061-CI0001-06. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

REFERENCES


