GIS and integrated coastal ocean forecasting

Gabrielle Allen\textsuperscript{1,2,*}, Philip Bogden\textsuperscript{3,4}, Gerald Creager\textsuperscript{5}, Chirag Dekate\textsuperscript{1,2}, Carola Jesch\textsuperscript{1,6}, Hartmut Kaiser\textsuperscript{1}, Jon MacLaren\textsuperscript{1}, Will Perrie\textsuperscript{7}, Gregory Stone\textsuperscript{8}, Xiongping Zhang\textsuperscript{8}

\textsuperscript{1} Center for Computation & Technology, Louisiana State University, Baton Rouge, LA 70803, USA, \textsuperscript{2} Department of Computer Science, Louisiana State University, Baton Rouge, LA 70803, USA, \textsuperscript{3} Southeastern Universities Research Association, 1201 New York Avenue, Washington, D.C 20005, USA \textsuperscript{4} GoMOOS, 350 Commercial Street, Portland, ME 04101, USA \textsuperscript{5} Academy for Advanced Telecommunications, 3139 TAMU, Texas A&M University, College Station, TX 77843-3139, USA \textsuperscript{6} Wetland Biogeochemistry Institute, Department of Oceanography and Coastal Science, Louisiana State University, Baton Rouge, LA 70803, USA \textsuperscript{7} Fisheries and Oceans, Canada, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada \textsuperscript{8} Coastal Studies Institute, Louisiana State University, Baton Rouge, LA 70803, USA

SUMMARY

The SURA Coastal Ocean Observing and Prediction (SCOOP) program is using GIS technologies to visualize and integrate distributed data sources from across the United States and Canada. Hydrodynamic models are run at different sites on a developing multi-institutional Computational Grid. Some of these predictive simulations of storm surge and wind waves are triggered by tropical and subtropical cyclones in the Atlantic and the Gulf of Mexico. Model predictions and observational data need to be merged and visualized in a geospatial context for a variety of analyses and applications. A data archive at LSU aggregates the model outputs from multiple sources, and an automated workflow triggers remotely performed conversion of a subset of model predictions to georeferenced data sets, which are then delivered to a Web Map Service located at Texas A&M University. Other nodes in the distributed system aggregate the observational data. This paper describes the current use of GIS within the SCOOP program, along with details of the automated distributed dataflow and workflow which results in geospatial products. We also focus on future plans related to the complimentary use of GIS and

*Correspondence to: Center for Computation & Technology, 302 Johnston Hall, Louisiana State University, Baton Rouge, LA 70803, USA.

Contract/grant sponsor: Office of Naval Research, Award; contract/grant number: N00014-04-1-0721

Contract/grant sponsor: National Oceanic and Atmospheric Administration; contract/grant number: NA04NOS4730254

Copyright © 2000 John Wiley & Sons, Ltd.
Grid technologies in the SCOOP program, through which we hope to provide a wider range of tools that can enhance the tools and capabilities of earth science research and hazard planning.

KEY WORDS: Grid Computing, GIS, Geographic Information Systems, Coastal Modeling, SCOOP, SURA, Open Geospatial Consortium

1. INTRODUCTION

Geographical Information System (GIS) technologies are becoming increasingly important in coastal sciences, due largely to recent advances in web mapping. The most novel applications leverage open standards for web services from the Open Geospatial Consortium (OGC). This paper describes how GIS is being used for a large collaborative project in coastal modeling and prediction, and how emerging technologies and ideas from the Grid community could be used to enhance the functionality, usability and accessibility of GIS.

Web-based GIS and Grid computing are both relatively new and emerging fields that share some common needs. They both depend on community standards to enable interoperability at various levels. In Grid computing, the Global Grid Forum (GGF) provides the venue for advancing the standards and tools. The Open Geospatial Consortium (OGC) [14] plays that role for web-based GIS. Both fields are looking at web services to provide the foundation for interoperability. In the case of Grid computing, new web services definitions are being investigated to address the requirements of the HPC community. Although the more popular OGC specifications relate directly to Web-mapping, the OGC is advancing specifications for geospatial and location-based services of all kinds, including sensor networks.

The members of the SURA Coastal Ocean Observing & Prediction (SCOOP) program are a diverse collaboration of coastal modelers, computer scientists and government agencies who are seeking to create an open integrated network of distributed sensors, data and computer models for the southeastern coastal region of the United States (US). SCOOP activities are driven by the need for improved forecasts and real-time information for severe storm events, such as tropical storms and hurricanes. The recent catastrophes in the southeast US following the triad of hurricanes Katrina, Rita and Wilma have highlighted the pressing need for timely and accurate forecasts as well as improved coordination and information transfer between domain experts, policy makers and emergency responders. Much of this information transfer necessarily takes place using map-based graphics. News agencies, bloggers, and concerned colleagues, made use of Google Earth during the Katrina aftermath, providing a small indication of how powerful a standard integrated system of geographically tagged information could be. Exploitation and standardization of geographical information, and its integration with new fields such as Grid computing, will be important for the prediction and management of severe storm events, as well as for many other scenarios involving geographical locations.

SCOOP covers a wide range of activities with the central aim of providing a service-oriented cyberinfrastructure for the community, to be achieved by modularizing critical components, providing standard interfaces and data descriptions, and leveraging new Grid technologies. This cyberinfrastructure will include components for data archiving, integration, translation
and transport, model coupling and workflow, event notification and resource brokering. Rather than developing a single community model or toolkit, using framework approaches such as the Earth System Modeling Framework (ESMF) [1] or the Cactus Code [19, 16], SCOOP is building interfaces to allow existing models to communicate with each other with coarse-grained connectivity.

A key component of the SCOOP vision is the real time delivery and visualization of integrated information products. Despite the wide array of data formats and storage conventions in use by the various disciplines, all of the data sources share the need for a geospatial context that enables integration. GIS technologies provide a critical path to the solution for three reasons. First, the terrestrial scientists that developed GIS have already settled on standards that overcome many of the challenges to interoperability, including georeferencing, map projections, data formats, etc. Second, a huge wealth of legacy data needed for earth science applications already exist in GIS-compatible data formats, including coastal topology, road networks, etc. And third, there is a huge market for GIS applications that can leverage and support advances in earth science provided the results can be integrated into the legacy GIS frameworks. The OpenIOOS interoperability testbed [5] (see Figure 1), is one such end product which demonstrates GIS data integration from varied sources including the SCOOP community. The OpenIOOS testbed server includes model results, in situ observations and satellite imagery. The data collected by OpenIOOS comes from a multitude of distributed contributors, who provide both real-time and historical information. The interfaces used by OpenIOOS are OGC-compliant web services, and most of these interfaces are invoked on-demand by the website users for real-time up-to-date information gathering and presentation. The OpenIOOS contributors includes several federal agencies and more than a dozen academic research institutions across the U.S. The activity is supported by the Office of Naval Research and the NOAA Coastal Services Center.

A different product which used SCOOP data and GIS technologies, along with other data sources, was a 3-D visualization combining multiple data sets to show the build up of Hurricane Katrina as it bore down on the city of New Orleans. This visualization was put together for the Supercomputing 2005 conference to explore and motivate research and development in advanced real time visualization techniques in this field. Figure 2 shows an image from the visualization, additional details are available at [2] along with a downloadable movie.

This paper describes the current integration and use of GIS in the developing service-oriented architecture for SCOOP as well as a vision for how GIS technologies can become a core enabler for future scenarios. In Section 2 we provide an overview of the different use-cases for the modeling activities in SCOOP, and Section 3 describe the model2gis software which converts model output to GIS format files. Section 4 describes how model2gis is currently integrated into the current SCOOP dataflow and Section 5 details the experiences and lessons learnt deploying these components operationally during the 2005 hurricane season. Finally Section 6 describes related work in combining Grid technologies and GIS and Section 7 proposes future directions for the integration and exploitation of these combined fields.
2. COASTAL MODELING IN SCOOP

The SCOOP community currently engages in distributed coastal modeling across the southeastern US, including both the Atlantic and Gulf of Mexico coasts. Various coastal hydrodynamic models are run on an operational (24/7/365) basis to study physical phenomena such as wave dynamics, storm surge and current flow. The computational models include Wave Watch 3 (WW3), Wave Model (WAM), Simulating Waves Nearshore (SWAN), ADvanced CIRCulation (ADCIRC) model, ElCIRC, and CH3D [8]. Atmospheric model results from models such as NAM, NOGAPS, COAMPS, and analytical models provide the wind forcing which feed into the coastal hydrodynamical models. Most SCOOP models are run on an
Figure 2. An image from a visualization of Katrina which integrates different fields such as wind vectors, surge height and air temperature, and different scales. Such visualizations can help researchers interpret simulation data, and the general public better comprehend the impact on the region. [Visualization: W. Benger, S. Venkataraman (CCT)].

operational basis at least once a day and generate 72 hour forecasts. The models are currently run at specific sites with existing local expertise, although the project is implementing a Grid-based infrastructure for future coordinated distributed deployment.

In addition to continuous operational modeling, extreme events such as hurricanes or tropical storms initiate a series of event-driven workflows. Advisories from the National Hurricane Center (NHC) about impending tropical storms or hurricanes are used to trigger automated workflows that start with the generation of high resolution wind fields around the center of the energetic event. These wind fields then initiate hydrodynamic models at different sites. The current scenario involves running of ADCIRC (at University of North Carolina), CH3D (at University of Florida), and ELCIRC (at Virginia Institute of Marine Sciences), WW3 and WAM will be added in the near future. Figure 3 illustrates the different regions for which the SCOOP models provide forecasts in their present implementations.

The resulting data fields obtained from both the operational and storm event-driven scenarios are distributed to the SCOOP partners for local visualization and further analysis, and are also archived as described in Section 4.
Figure 3. The different geographical regions over which the SCOOP model set is currently deployed.

3. **model2gis**: CONVERTING MODEL RESULTS TO GIS FORMAT

GIS data formats usually contain descriptive metadata along with sufficient information for unambiguous georeferencing. The data itself consists of vector data (points and polygons) or raster data (e.g., 2-dimensional images or regular meshes of data values).

Currently the output files from each of the models (listed in Table I) are converted to GIS data formats using the *model2gis* program. This section describes the current mechanisms for this conversion, and Section 4 describes the dataflows involved to deliver model files for conversion, and dispatch the resulting GIS data files.

A software component, called *model2gis*, was developed to perform the data conversion. This software, written in C++, reads in a model data and configuration file, and outputs GIS shape files which store non-topological geometry and attribute information for the spatial features in a data set. The shapefile format is a widely adopted GIS format and supported by many GIS applications (Table I lists the sizes of the different data and GIS files). Specific to a particular model, the configuration file contains information needed to read and interpret model data...
<table>
<thead>
<tr>
<th>Model Name</th>
<th>Number of Simulations</th>
<th>Raw File Size (MB) per Simulation</th>
<th>Total Size (MB) (incl. shp dbf dbx)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH3D (UFl)</td>
<td>5</td>
<td>7.9 (NetCDF)</td>
<td>72</td>
</tr>
<tr>
<td>ADCIRC (UNC)</td>
<td>6</td>
<td>45 (NetCDF)</td>
<td>138</td>
</tr>
<tr>
<td>ElCIRC (VIMS)</td>
<td>1</td>
<td>128 (NetCDF)</td>
<td>159</td>
</tr>
<tr>
<td>WW3 (BIO)</td>
<td>1</td>
<td>0.500 (ASCII)</td>
<td>101</td>
</tr>
<tr>
<td>ADCIRC (UMIAMI)</td>
<td>1</td>
<td>60 (NetCDF)</td>
<td>411</td>
</tr>
</tbody>
</table>

Table I. Description of model output files and converted GIS files produced by current SCOOP deployment each day.

files. `model2gis` runs on data four times each day, currently at 02:30, 08:30, 14:30 and 20:30 CST, corresponding to times when new model data is expected to be available.

A key difference between models is the type of computational mesh used. ADCIRC, ElCIRC and CH3D use non-uniform meshes, whereas WW3, WAM, SWAN all use uniform meshes. The output files for all of these models contain data values that are valid at the mesh nodes. However, each model uses its own data format and description for these files. In addition, the data files do not always contain explicit coordinate locations of data points. For example, in the case of ADCIRC, water levels and velocities at each node are stored in two different files, while a third file specifies the mapping from any node to a corresponding longitude and latitude.

Presently in SCOOP, a sequence of steps is used to process output files for visualizing the vector \((u, v)\) components of water current direction, water elevation (i.e., storm surge), wave heights and wave directions on a two-dimensional sea surface. The different steps in the conversion procedure are:

1. From the output filename, following the SCOOP filenaming convention, `model2gis` identifies the model and reads the appropriate configuration file.
2. Scalar fields representing different contributions to the water surface are constructed depending on the mesh type and model:

   2a. For non-uniform mesh models (ADCIRC, ElCIRC or CH3D) the water elevation for any element is calculated by averaging the water level values at each of the associated nodes (the elements are either triangular or quadratic). This simple spatial interpolation scheme is applied to the entire computational mesh.
2b. For uniform mesh models (WW3, WAM and SWAN) a virtual box with the same dimensions as the element is constructed around each node, with the node at the center. The scalar value of the node’s wave height is applied across the virtual box.

3. Vector fields representing the water velocity are constructed depending on the mesh type. For ADCIRC, ELCIRC and CH3D, to generate the velocity vector representing the water current, the two dimensional velocity components, \((u, v)\), at each node are used. The vector amplitude, \(A\), and direction, \(\theta\) at each node are calculated using the standard formulae, \(A = \sqrt{u^2 + v^2}\), \(\theta = 180 \arctan(u/v) / \pi\), which are applied across the mesh to determine the amplitude and direction at each node.

4. Vector fields representing wave heights and direction, are computed explicitly within WW3, WAM and SWAN. Group velocity of the wave spectra are also computed at each model grid point, but are not usually output. However, waver currents can impact wave estimates, and waves can seriously impact estimates of water currents in explosively developing storms. Dynamical coupling of wave models and ocean circulation models is planned for 2006, including a web-based demo, as a partner activity to SCOOP.

5. Each interpolated velocity or wave field or water level is written to disk in GIS shape file format with file names following the SCOOP file naming convention.

6. Once all model output files are converted by \textit{model2gis} the resulting shape files are placed on a Samba staging area which is accessible by SCOOP data management services.

One of the greatest challenges facing the SCOOP program concerns the diversity of output formats, especially for those models using irregular meshes. At present, with irregular meshes we have the choice of interpolating to a regular mesh or converting to a polygon representation of isolines. There are drawbacks with each approach. Each approach involves some amount of interpolation that degrades the precision used in the numerical algorithm. Moreover, although there is a plan to adopt standard data-storage formats in SCOOP, each of the coastal models currently generates results in a different formats. For example, ADCIRC uses binary NetCDF, and WW3 uses ASCII. Model diversity also creates challenges due to different representations of the time dimension and the forecast intervals. For instance, in one 72 forecast, ADCIRC generally outputs forecasts at one hour intervals, whereas the WW3 model produces one forecast every three hours. These differences are mainly due to model configuration issues and the time/computational constraints for running the model. To provide a consistent time interval across all the models, linear temporal interpolation algorithms are applied in \textit{model2gis}. In the future, SCOOP partners may implement a common output time interval.

4. CURRENT IMPLEMENTATION

In working towards efficient and open sharing of data and results, the SCOOP community have agreed on using standard input and output formats (e.g. NetCDF, GRIB), establishing a common file naming convention, and archiving model results in SCOOP archives located at TAMU and LSU [22]. The Unidata Local Data Manager (LDM) [7], a collection of cooperating programs that select, capture, manage, and distribute arbitrary data products, was selected as a common data transport mechanism. In the current version of SCOOP, metadata is encoded
in the file name. Future implementations will move towards metadata catalogs. Output from various atmospheric and hydrodynamic models are published by SCOOP partner sites around the US and Canada using LDM. The subscriber-based LDM system delivers the files to registered sites, including the LSU SCOOP archive. Upon arrival at the archive, hydrodynamic model files are automatically detected and forwarded to the LSU Coastal Studies Institute (CSI) for conversion to the GIS format.

Using the model2gis software, the results of the various computational models can be converted to the appropriate GIS data format. However, in the distributed SCOOP production environment, complex dataflows are workflows that are needed to aggregate data across the collaboration, initiate the model to GIS conversion, and deliver the GIS data to the appropriate location in a timely manner. The SCOOP collaboration has built much of their current infrastructure around existing technologies such as LDM, ArcIMS, Mapserver [12] and Globus 3.2.1. However, some SCOOP tools such as the LSU Archive [22] and client tools [20] use grid application toolkits which allow underlying technologies to be easily adapt to different underlying Grid middlewares. ArcIMS is the ESRI solution for delivering dynamic maps and GIS data and services via the Web. MapServer is an open source counterpart to ArcIMS that is relatively widely used in the SCOOP program. Both MapServer and ArcIMS use OGC-compliant web services.

The current data flow is the result of integrating several different systems with existing constraints, and Section 5 identifies problems with this system. The data flow, illustrated in Figure 4, consists of the following steps:

1. Data from the models (ADCIRC, WW3, CH3D ELCIRC) is pushed by the data producers and intermediaries to the LSU SCOOP archive via LDM. The system is designed for event-driven data distribution.

2. When the model results arrive at the archive, two actions are performed. First the datasets are archived in a specific physical location and appropriate entries are added in a logical file service (Globus RLS). Each file is exposed for download via multiple protocols including HTTPS and GridFTP (NetCDF format files are also made available through the OpenDAF protocol). Secondly the archive matches the metadata patterns of the ingested files to a configuration file that allows the archive to perform subsequent actions based on the ingested files, such as notification, running of external scripts etc. In the case of this GIS workflow, the archive initiates scripts that move the files from the SCOOP archive to a Samba-based staging area.

3. A Windows machine that schedules and manages the model2gis program, accesses the model data sets from the linux buffer machine using the SMB protocol.

4. The GIS processing resource runs the model2gis utility to convert all the available model datasets to GIS formats as described in Section 3 according to a predefined schedule (at 0230, 0830 1430 2030 CST).

5. GIS files are made available for distribution: (a.) The GIS files generated by model2gis are stored on the Samba share; (b.) The Converted GIS files are also made available using an ArcIMS Web Map service interface.
6. crontab based scripts transfer (via GridFTP) the GIS files to the scoop archive and utilize the LSU SCOOP Archive’s upload utility to push the files to the SCOOP archive based on predefined schedule (at 0245, 0845 1445 2045 CST).

7. The LSU SCOOP archive currently does not store these results locally, due to the large storage requirements (Table I). Instead the LSU Archive instantaneously relays these files over to the TAMU archive via LDM. This scenario might change in the future as LSU is partnering with the SDSC DataCenter to utilize the Storage Resource Broker to expand the LSU storage archive capacity by several terabytes.

8. The TAMU site publishes the GIS files from LSU using a Mapserver based WMS.

Figure 4. Current dataflow for GIS shapefile creation in SCOOP which was implemented for the 2005 hurricane season.
9. The OpenIOOS server displays the results from the LSU ArcIMS based WMS. The TAMU based Mapserver solution was deployed on an experimental basis, and will replace the ArcIMS solution.

5. RESULTS

During the official 2005 hurricane season (June 1st to November 30th) the system described in Section 4 was deployed as a production capability. The OpenIOOS interface provided end users with the ability to create customized views showing the results of both the operational (24/7/365) and hurricane event-driven modeling scenarios (e.g. Figure 1). One use of GIS data formats was to simplify the calculation of maximum of maximum (MOM) storm surges at the computational nodes across ADCIRC ensembles. This calculation was achieved by extracting the maximum water elevation values at each computational node across the multiple datasets. Using such capabilities, researchers were able to study the maximum possible storm surge over a predicted forecast track. GIS converted data formats also enabled skill assessment of the coastal models, by facilitating comparison of model predicted results with observed measurements from buoys, and water level gauges. Such studies are important to identify the strengths and weaknesses of the computational models and thereby lead to model improvements.

In deploying the GIS data tools in a production setting, several immediate problems and limitations were identified which are now described. On the server side, processing WMS requests introduced delays following client requests to ArcIMS or MapServer for the visualization of model outputs. One source of delay, particularly for large areas, is the processing of output files. For example, a request to view the ADCIRC model results for the entire south east modeling domain required a typical processing time of 20-25 seconds. In its current implementation, for each request from a client the requested GIS shape file is reloaded, in other words multiple client requests for the same data set results in redundant processing. Smart algorithms to provide advanced functionalities such as server side caching or preemptive loading of datasets are needed.

Several of the steps shown in Figure 4 involve time choreographed events. In the current implementation steps (4.) and (6.) are scheduled to run four times each day using a crontab or windows task scheduler. These steps add time dependencies and restrictions to the process and assume that data sets are available and complete when the different steps are executed. This leads to a workflow which is difficult to monitor, manage, or to make reliable and fault tolerant. Any delay in data arrival can make the process unreliable, by delaying timely processing of results. Some of the model results although available early, had to wait until the model2gis process could be triggered by the scheduler. If model2gis were to be triggered by the arrival of model results, the dependence on time schedules could be eliminated and the entire process made more reliable.

Currently the model2gis code only runs on the Microsoft Windows operating system due to its reliance on certain libraries. Once the code is ported to Linux, the dataflow will be improved by hosting the model2gis executable on the SCOOP archive. Since the SCOOP archive stores the model output files, model2gis could be triggered at the archive resource as and when output
files become available in the archive. Such an archive triggered \textit{model2gis} mechanism would help eliminate dependence on time schedules.

The mechanisms described in Section 4 involve numerous data transactions across different administrative domains. These transactions are rarely monitored making it difficult to audit the data flow. The lack of auditing information leads to difficulties in isolating problems or investigating usage and performance. An inventory of datasets at nodes managing the LDM data transport mechanisms would provide the ability to better track the data flow. Plans are currently under way to provide such tracking mechanisms based on information collected by transport mechanisms including LDM.

The storage of the converted GIS shape files for different models (listed in Table 1) provided significant challenges. For the 2005 hurricane season 2.4TB of GIS files were created. For the current range of scenarios around 162 files are produced each day. In the next versions of SCOOP, additional scenarios are being considered, that would significantly increase the storage and metadata requirements. To accommodate such scenarios, services that generate accurate metadata, high availability catalog services, high availability multi-terabyte storage will be needed for archival of the GIS files.

6. RELATED WORK

Several earth science projects that utilize GIS services in conjunction with Grid middlewares. Linked Environments for Atmospheric Discovery (LEAD) \([4]\) is one such project providing Grid based dynamic forecasting of mesoscale atmospheric models, and comparison of model results to sensor observations to perform skill assessment of models. Data requirements for the LEAD project are representative of many earth science related projects including SCOOP, with respect to both storage and metadata management. LEAD project has operational portlets that allow for selection of geospatial regions over which different computational models need to be run. The Solid Earth Research Virtual Observatory (ServoGrid) \([6]\) project utilizes GIS and Grid technologies to integrate data from seismic simulations and sensors, providing relevant information such as time series analysis comparisons for model skill assessment. Several projects including LEAD, ServoGRID, GEON incorporate GIS based solutions utilizing Grid infrastructure to provide solutions, to analyze computational model results with real time observation from sensors, to varying degrees. With the emergence of Grids and collaborative modeling in earth sciences community, the integration of GIS and Grid technologies at various levels will help drive innovative solutions for the domain scientists.

7. FUTURE DIRECTIONS

The standards under development by the community-driven Open Geospatial Consortium (OGC) are leading to a new era for GIS related research and technologies. The OGC OpenGIS Web service definitions such as WMS for maps, or WFS for features allow geospatial data from many different sources to be aggregated and analyzed. Other standards define unified data formats and data interfaces between applications. Geospatial data sources can be geographically distributed and varied, providing data from scanners, remote sensing devices,
GPS receiver, etc. Different communities can for the first time truly share data, results and expertise, jointly work on a rich spectrum of problems, broadly advancing science and understanding. As these developments take place, we are also able to contemplate new scenarios where geospatial information becomes a fundamental part of simulation codes and advanced simulations, leveraging ideas and technologies from distributed and Grid computing. GIS-aware steering and interaction of simulation codes could lead to a complete new way of interactive and responsive research and investigation. The emerging need to investigate the integration of GIS and Grids led to a workshop on *Building Geographical Information System Grids* at the Grid standards meeting, GGF15 [9].

**Distributed and Grid Computing**

One common definition of Grid computing is *coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations* [17]. Grid computing goes beyond traditional distributed computing in the support of distributed, diverse and transient collections of people (virtual organizations) and its focus on standard, general purpose protocols and interfaces, along with tools which provide significant quality of service. The SCOOP community forms one such virtual organization, with common needs for model deployment, data location and transfer, security, etc.

The emerging standards in GIS, along with an explosion in the collection of environmental and geospatial data, means that data management is an increasingly important consideration. It is no longer feasible to store the data needed for a community on a single machine, for example SCOOP relies on data archives distributed across LSU, TAMU and more recently at SDSC. Further, the need for real time data integration from different administrative domains and of different formats requires sophisticated support for collaboration from the underlying software infrastructure. Such an infrastructure would have to provide several components including, but not limited to: consistent access permissions to the data archives [15]; information repositories to allow location and retrieval of specific data sets using diverse data sharing protocols [10, 23]; and the ability to accommodate diverse software stacks across administrative domains.

Some of these issues have been addressed using Grid computing technologies developed in the context of Computational and Data Grids. However, it is important to point out the Grid computing is often mis-interpreted as being able to miraculously provide an application community with software and data interoperability at all levels. While the standards, technologies and experiences being developed in this field can help guide the move towards interoperability, most of the burden of providing appropriate software interfaces, data and metadata descriptions and importantly architectural requirements requires this expertise of the application community themselves. Within the Grid community, there is still a lack of useful standards in key areas, although there are efforts such as the Open Grid Services Architecture, or Simple API for Grid Applications being produced by working groups in the Global Grid Forum [11]. The integration of GIS and Grid technologies has been demonstrated in preliminary forms in the context of SCOOP, but can be further improved in some areas as described below.
Data Grid Technologies  In the Grid community, data management technologies are being driven by the needs of applications such as high energy physics, astronomy, and gravitational wave detection, where experimental devices are already generating data up to petabytes each year. Many of the tools being developed and the experiences gathered are applicable to GIS data, including distributed storage and database systems (for instance [13]), replica catalogs [3, 24], data placement management (for instance Stork [21]). GIS applications that are in vogue are yet to take full advantage of these technologies. The ability to integrate these GIS and Grid technologies will be one of the key future challenges for the development of spatial systems. GIS components when exposed in service oriented architectures could allow the implementation of new parallel and distributed algorithms and analysis methods without having to convert between different data formats.

Grid-based Information Providers: Using computational and data grid technologies involves maintaining information about the available resources and data. This includes meta data describing the semantics of available data sets and information about computational resources and data storage locations. Since GIS generally maintain semantically rich data, many of the existing systems already support management of meta data describing the actual data in a project. However a major weak point in today’s systems is the inability to locate and access remote data in an efficient manner. The infrastructure would also need to account for ability to aggregate information from legacy data transport mechanisms such as LDM. Using information repositories from Grid related projects (such as [18]) can help to overcome these limitations.

GIS and Simulations

GIS for Model Steering: GIS are known today to have very powerful tools providing users with the ability to analyze, integrate and homogenize different data from a wide range of data providers including static data such as maps and satellite images and dynamic data such as real time sensor data or modeling results. In the context of SCOOP we consider the integration of GIS related technologies in the process of the preparation of input data for modeling tasks as an important step towards faster turnaround times of simulations and improved simulation accuracy, which are major problems. Hurricane path prediction ideally involves the simulation of different possible hurricane paths, which is best done interactively based on the analysis of previous simulations and other related data.

Integration of GIS and Simulation Codes: A step further in this direction will be the direct integration of GIS components with existing simulation codes (as for instance Cactus and the ADCIRC model). This allows to get information about the internal state of a running simulation from the very beginning - a major precondition for timely simulation parameter steering. Overlaying and combined analysis of the in situ simulation data with other related spatial and attributive information creates a unique decision making environment for the people interested in the simulation results.

Comparison with available in situ and remotely sensed data requires GIS-based data-point extraction software. We need to integrate the point extraction facilities into the operational forecast models. Thus, we would have the ability on how to extract the model variable that
corresponds to in situ and remotely sensed data, in order to monitor the accuracy of forecast predictions. The procedure must: a) Extract predicted forecast results given the position of observed data, b) Compute and visualize comparisons between predicted and observed data, c) Deploy comparisons between predicted and observed waves to the Web.

The image from the visualization of data related to hurricane Katrina (see Figure 2) was created using various existing tools, for instance GIS for the data preparation, the Amira package for 3D visualization using data sets generated by different simulation codes in varied data formats. The construction of the Katrina visualization was a time consuming and manual process, identifying and locating appropriate data sets, converting data sets to GIS formats, and developing Amira networks to provide optimal information. In the future, technologies such as the system being developed in SCOOP, should be developed to automatically generate and archive GIS processed data, and provide the workflows for automated visualizations. Extending this concept to generate GIS data in realtime as a simulation runs, either as an integral part of a simulation framework or as an accompanying component, will allow for visualization windows into the running simulation itself, which would enable many different scenarios for new research and information dissemination to scientists, policy makers and the public.

ACKNOWLEDGEMENTS

This work was carried out as a component of the SURA Coastal Ocean Observing and Prediction (SCOOP) Program, an initiative of the Southeastern Universities Research Association (SURA). Funding support for SCOOP has been provided by the Office of Naval Research, Award N00014-04-1-0721 and by the National Oceanic and Atmospheric Administration’s NOAA Ocean Service, Award NA04NOS4730254. Part of this work was supported by the CLEAR project funded by the Department of Natural Resources of Louisiana, Interagency Agreement No. 2511-02-24. The authors would like to thank Andre Merzky, Vrije Universiteit, for providing assistance towards this paper.

REFERENCES


Copyright © 2000 John Wiley & Sons, Ltd.


