SCI-WMS: A Python based Web Mapping Service for Visualizing Geospatial Data

Brandon A. Mayer, Brian McKenna, Alexander Crosby and Kelly Knee

Abstract SCI-WMS [1] is an open-source Python implementation of an OGC WMS service [2] for qualitatively assessing society-critical atmospheric and oceanagraphic applications including forecasting, risk assessment, model comparison and algorithmic/parameter selection. The modular cross-platform implementation of SCI-WMS allows the service to keep pace with the rapid developments in the geospatial data science community and to produce visualizations for numerous types of model outputs with transparent support for both structured and unstructured geo-referenced topologies. This article outlines the implementation and technology stack for visualizing geospatial Climatological and Forecasting (CF) data using SCI-WMS and details the deployment of SCI-WMS for visualizing model data and simulations within the scope of the U.S. IOOS Coastal and Ocean Modeling Testbed project [3].

1 Motivation

Due to the explosion in the amount of atmospheric, oceanographic, climate and weather data either recorded *in-situ* or generated by modeling, inference and prediction algorithms, it is no longer feasible for a single institution to host and maintain a centralized database of information. Modern data management has been shifting hosting and maintenance responsibilities of large datasets to multiple participating institutions unified by a catalogue to provide a single, unified view of the distributed data to end users and analysts.

Brandon A. Mayer
Brown University, 182 Hope Street Providence RI 02906 e-mail: brandon_mayer@brown.edu

Brian McKenna, Kelly Knee
RPS-ASA e-mail: {BMcKenna,KKnee}@asascience.com

Alexander Crosby
5 River Rd. Suite 1 Cos Cob, CT 06807 e-mail: alexc@oceanweather.com

A common practice is for data producing organizations to host their own data and provide metadata information to a centralized aggregation service called a catalogue. The institution responsible for maintaining the catalogue provides a unified view of the data to end users by compiling meta data provided by other members. Such federated datasets may span petabytes of geospatial information, be composed of millions of files in different formats generated and hosted by vastly different systems located across the globe. By interacting with the catalogue the end user (such as an analyst or scientist) can navigate and search through the aggregated meta data and interact with particular data of interest, agnostic to distributed nature of the database.

While a decentralized approach to data management offers many advantages, data reduction and analysis tools have been slow to adapt to the distributed framework. For example, there are an abundance of applications for visualizing cartographic data on a single computer or computer clusters, yet many such programs are designed to deal with data on a local system, requiring analysts to download a local copy datasets and potentially reformat the data into the appropriate file format before processing and analysis can begin. Even if an end user has access sufficient resources to download and process a dataset, which is not always the case, tools designed for local or centralized data increase project costs in terms of bandwidth usage, time and storage. Additionally, applying centralized processing with decentralized storage the risk that different analysts working with identical local copies of data obtained from the same federation may use different local programs to generate incompatible visualizations or reach conflicting conclusions. Normalizing these results introduces a potential point of error and likewise increases the costs of analysis in terms of time and accuracy.

By maintaining a list of web accessible data endpoints, SCI-WMS is able to transparently produce consistent visualizations of web-hosted data. While SCI-WMS implements the OGC WMS [2] protocol, it is augmented with services for interacting with standard meta-data catalogues such as CSW [4], allowing SCI-WMS to autonomously track dynamic catalogues. SCI-WMS works in conjunction with NcML [5] (NetCDF Markup Language) to implement a data abstraction layer allowing data hosting agencies to maintain their own file formats which are exposed to SCI-WMS without writing custom i/o software, data duplication or reformatting.

SCI-WMS embraces distributed database principles and promotes the separation of concerns software and project management practice. For example, SCI-WMS may be deployed by a member of a much larger project who's sole responsibility is providing a visualization platform, promoting quality through specialization and saving time and costs for data analysis projects by providing a simple interface for end users to generate consistent visualizations of federated data.

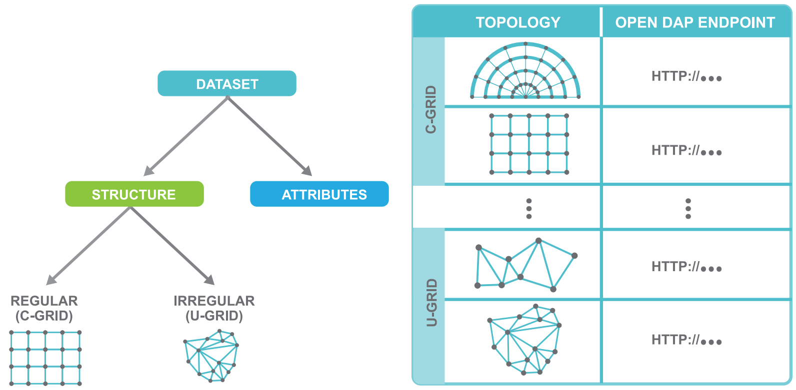
2 SCI-WMS

SCI-WMS is an open-source implementation of the Open Geospatial Consortium's (OGC) Web Map Service (WMS) [2] written in Python using the Django [6] web framework and standard cross-platform numerical software, NumPy [7] and Matplotlib [8] for generating visual content. The WMS specification defines a REST API [9] which responds to standardized HTTP GET requests from a web client for serving rasterized visualizations of geospatial data. A typical WMS request specifies a data layer and region of interest with optional meta data such as rendering style parameters. A WMS response may include additional meta data providing information about data registered with the service or a visualization in the form of a PNG or other standard image format.

While the OGC WMS specification standardizes the client-server communication protocol, WMS implementations vary dramatically in how a particular system fulfills the WMS request and responses. Vital to the efficiency of SCI-WMS is the decomposition of a registered dataset into **structure** (also known as **topology**) and **attributes** as shown in figure 1a and detailed in subsequent sections. SCI-WMS maintains a local topology cache for efficient storage and processing of spatial neighborhoods and subsets with respect to data structure. To minimize redundancy, attributes are not replicated locally but referenced via standard web-services and a database of structure-endpoint pairs is maintained as visualized in figure 1b. As geospatial WMS requests are commonly restricted to a subset of the Earth's surface, SCI-WMS uses the topology cache to compute the subset of numerical attributes needed to fulfill each request prior to retrieving the appropriate data, which are typically external to the server hosting SCI-WMS and accessed via HTTP. Furthermore, by classifying a topology as regular or irregular, efficient algorithms and data structures are exploited to optimize the computation of attribute subsets.

2.1 Data Model

To represent a continuous function without a closed-form representation, a digital computer must store measurements of the function a discrete sample in a particular domain. Yet rendering numerical data typically requires knowledge of the values between samples to produce perceptually continuous images from arbitrary viewpoints. To develop efficient algorithms, visualization tools often decompose data into structure and attributes [10]. Structure encapsulates both the locations and connectivity relations onto which attributes are mapped and connectivity information serves to constrain the interpolation problem. Note that some authors continue the abstraction of structure into topology and geometry [11], however in the context of this research, topology is synonymous with structure. Figure 1a outlines the data model adopted by SCI-WMS. A dataset is composed of attributes with associated structure which is further classified as a regular or irregular, known as **c-grid** and **u-grid** topologies in SCI-WMS documentation.



(a) (b)

Fig. 1: (a) Decomposition hierarchy of the data model. (b) SCI-WMS topology and endpoint data store.

2.2 c-grid Topology

c-grid topologies refer to data with regular structure that can be defined analytically, e.g., rectilinear or curvilinear grids. Storing c-grid topologies amount to storing the closed-form expression. Algorithms for processing c-grids such as finding nearest neighbors or computing points that fall withing a polygonal subset are computed directly using the implicit c-grid representation, incurring minimal computational overhead.

2.3 u-grid Topolgy

An unstructured (u-grid) topology is defined as a topology that is not regular, i.e., does not admit a closed-form representation of sample locations and connectivity. Unstructured topologies offer the highest level of flexibility from a visualization and modeling standpoint as higher sampling frequencies may be used in regions of interest while sparsely sampling low-impact areas to conserve computational resources.

There is a large quantity of CF data associated with structured topologies. However, as storage and processing hardware has become more accessible, complex unstructured representations have become more prevalent. However, most existing visualization technologies in the CF community can render only regularly structured datasets. SCI-WMS is one of the first visualization services to support rendering data assocwith an unstructured topology.

Any NcML file exposing the topology of an externally hosted dataset according to the CF-UGRID sepcification can be processed by SCI-WMS. According the CF-

UGRID standard, a topology is always embedded on the real line, in the plane or space with sample locations, the vertices of the topology, exposed as an array of coordinates in the appropriate ambient space. Vertex connectivity is expressed as an array where each element is an index into the vertex list. The dimension of a topology defines the atomic spatial element created by the connectivity list. The CF-UGRID specification defines topology dimension recursively: a topology with dimension 0 defines a set of disconnected points (no connectivity) called ***nodes***, a 1D topology consists of lines or curved boundaries known as ***edges***, a 2D topology is a set of planes or surfaces enclosed by a set of edges (e.g. triangulation) called ***faces***, and a 3D topology specifies the volume enclosed by a set of faces called ***volumes***.

In contrast with c-grid topologies, u-grid topologies require explicit enumeration of sample locations and connectivity, requiring spatially-aware data structures for optimal storage and processing for performant visualization algorithms. To this end, SCI-WMS maintains a local topology cache, storing u-grid topologies using binary R-tree [12] data structures on disk locally on the deployment server for fast access. The R-tree is created when the dataset is first registered with the SCI-WMS service and if a change in the underlying data is detected at an endpoint associated with a topology cache, the R-tree is rebuilt.

2.4 Distributed Memory Model

Attributes are numerical quantities associated with a topology. For example, common attributes may be vector valued wind directions computed by an atmospheric modeling algorithm at the vertices of a triangulated 2D topology. The same algorithm may have also simulated air temperature, a scalar attribute, at the centroid of cell volumes specified by a 3D topology. Attributes have their own dimensionality which is not necessarily equal to the dimension of the topology.

The local topology cache and external attribute mechanism defines a distributed memory model for datasets registered with SCI-WMS. Given a request for the visualization of a attributes pertaining to a region of interest, the visualization pipeline consists of first computing the sample locations withing the region of interest, using the implicit representation for c-grid and R-trees for u-grid topologies, then fetching the external attributes over HTTP. For rendering, the sample connectivity within the area of interest is constructed from the connectivity array which is used for interpolation.

3 Deploying SCI-WMS for the U.S. IOOS COMT Testbed

The U.S. Integrated Ocean Observing System (IOOS) Coastal and Ocean Modeling Testbed (COMT) was formed to unify otherwise disparate entities in government, academia and industry to leverage the proliferation of oceanographic data and modeling techniques to combat natural and man-made coastal stressors by accelerating the turnaround from research and development to operational application of society-critical applications including: forecasting, model comparison, model skill assessment, and algorithmic/parameterization improvements [3]. A crucial component for the success of the U.S. IOOS COMT mission is a web accessable tool for quickly visualizing and assessing a diverse set of coastal modeling and observational data. While SCI-WMS is a general software solution for geospatial visualization, it is a key component in realizing the U.S. IOOS COMT mission, facilitating qualitative model comparisons and aggregation through a unified visualization framework.

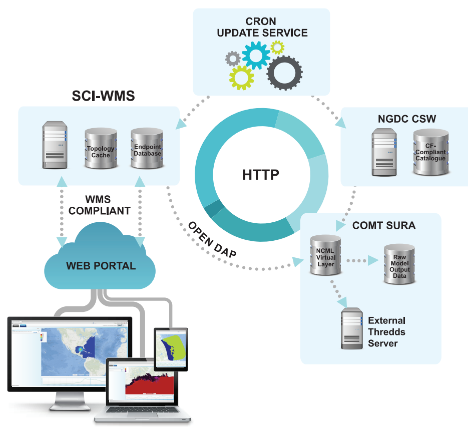


Fig. 2: Overview of the SCI-WMS deployment for the U.S. IOOS COMT project. SCI-WMS updates its topology and endpoint database via a nightly service which queries CF-Compliant datasets cataloged by NGDC. Model data is hosted on an external web server exposed by an NcML facade as a single NetCDF data structure accessable to SCI-WMS via OPeNDAP. SCI-WMS responds to requests by end users interfacing through a custom built web portal.

Figure 2 outlines the cyberinfrastructre behind the deployment of SCI-WMS for the COMT project[1]. The National Oceanic and Atmospheric Administration (NOAA) - National Geophysical Data Center (NGDC) geoportal indexes public geophysical datasets and provides an OGC CSW service to query datasets by their metadata attributes. SCI-WMS queries the NGDC Geoportal at regular intervals updating the local topology cache and structure-endpoint database (figure 1b) with new or modified datasets. Raw coastal data is hosted by the Southeastern Universities Research Association (SURA) on a dedicated server for the COMT project [13]. Each data set may consist of multiple files in different formats, and may be the result of very different models run by various institutions with disparate computing

[1] http://testbedwww.sura.org/explorer/

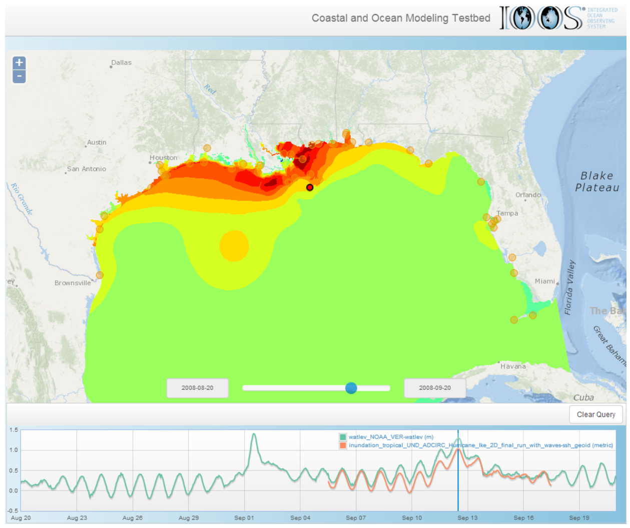


Fig. 3: Comparison of ADCIRC (unstructured topology) model results with observed water levels in the Northern Gulf of Mexico for Hurricane Ike. Verified observed water levels are from NOAA's Station 8760922 (red dot on map). The map shows modeled water levels (in meters above the geoid) at the peak of the storm in southern Louisiana. The time series plot shows both the modeled (green) and observed (orange) water levels. The vertical blue line in the time series plot corresponds to the current time of the map.

resources. However, accompanying each dataset is an NcML virtual layer which exposes each dataset as a single NetCDF [14], OPeNDAP [15] accessible object. Furthermore, the NcML facade presents a consistent set of meta information in accordance to CF-Conventions [16] providing services like SCI-WMS access to the raw data through a uniform interface.

Currently, SCI-WMS is used to visualize data from the first phase groups of IOOS COMT program: *estuarine hypoxia, shelf hypoxia and coastal inundation* [3]. For each modeling group, SCI-WMS successfully generates consistent visualizations of data generated by ADCIRC [17], FVCOM [18], SELFE [19] and SLOSH [20] coastal modeling algorithms and serves as a use-case for how SCI-WMS can be leveraged as a scalable solution for delivering visualizations of scientific data to a diverse community.

SCI-WMS currently supports contour and filled-contour visualizations styles for scalar attributes while 2D flow fields can be shown as arrows or barbs for vector valued attributes. Figure 3 shows a web portal utilizing the SCI-WMS backend to compare ADCIRC model output for Hurricane Ike with water levels observed by NOAA stations and figure 4 visualizes current direction and speed in the Chesapeake Bay area. Figure 5 renders the sea surface wave height computed along the Atlantic coast of South America, the Gulf of Mexico, up to Canada. The topology in this example is unstructured (u-grid), a triangulation containing over 5 million vertices (sample locations). Attributes are fetched from the appropriate external server as needed, rendered, cached for performance, but ultimately discarded after processing to minimize storage redundency. Ongoing development is in progress for SCI-WMS to support emerging geophysical datasets such as ensemble model output and to provide clear visual support for the assessment and quantification of model skill and performance metrics.

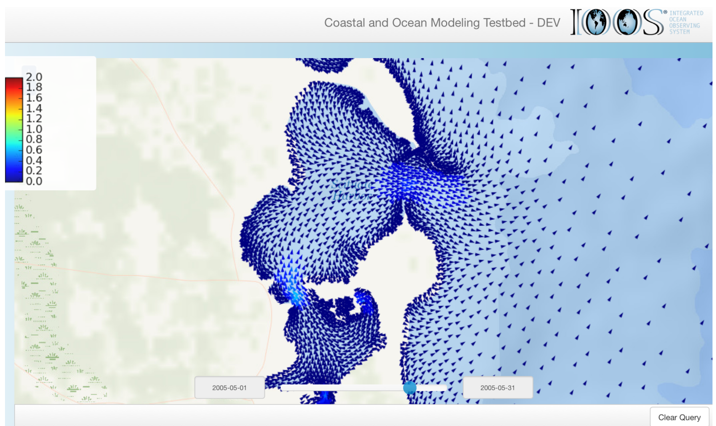


Fig. 4: Visualizing Current direction and speed in the Chesapeake Bay area.

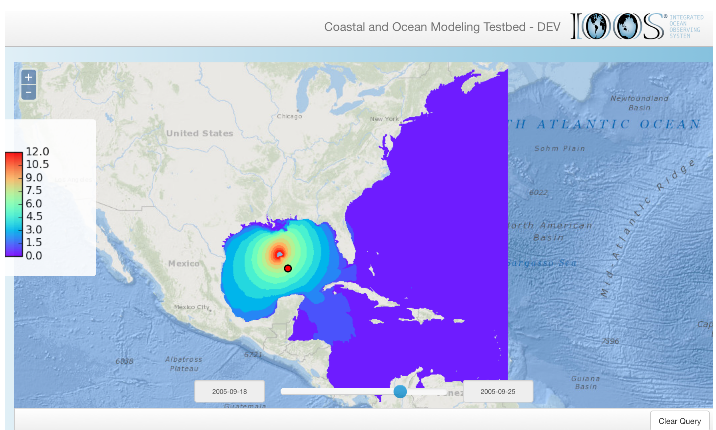


Fig. 5: Visualizing significant sea surface wave height along the eastern coast of the United States.

References

1. ASA-RPS https://github.com/asascience-open/sci-wms, “SCI-WMS.”
2. Open Geospatial Consortium http://www.opengeospatial.org/standards/wms [Accessed: 2014-07-24], “OpenGIS Web Map Service.”
3. R. A. Luettich, L. D. Wright, R. Signell, C. Friedrichs, M. Friedrichs, J. Harding, K. Fennel, E. Howlett, S. Graves, E. Smith, G. Crane, and R. Baltes, “Introduction to special section on the U.S. IOOS Coastal and Ocean Modeling Testbed,” *Journal of Geophysical Research: Oceans*, vol. 118, no. 12, pp. 6319–6328, 2013.
4. Open Geospatial Consortium http://www.opengeospatial.org/standards/cat, “OpenGIS catalogue services specification.”
5. R. B. Jerard and O. Ryou, “Ncml: a data exchange format for internet-based machining,” *Int. J. Comput. Appl. Technol.*, vol. 26, pp. 75–82, June 2006.
6. Django [Computer Software]. Retrieved from https://djangoproject.com.
7. S. V. D. Walt, S. C. Colbert, and G. Varoquaux, “The numpy array: A structure for efficient numerical computation,” *Computing in Science & Engineering*, vol. 13, no. 2, pp. 22–30, 2011.
8. J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
9. R. T. Fielding and R. N. Taylor, “Principled design of the modern web architecture,” *ACM Trans. Internet Technol.*, vol. 2, pp. 115–150, May 2002.
10. B. L. Will Schroeder, Ken Martin, *The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*. Kitware Inc., 4 ed., 2006.
11. K. J. Weiler, *Topological Structures For Geometric Modeling*. PhD thesis, Rensselear Polytechnic Institute, 1986.
12. A. Guttman, “R-trees: A dynamic index structure for spatial searching,” in *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data*, SIGMOD ’84, (New York, NY, USA), pp. 47–57, ACM, 1984.
13. R. A. Luettich, L. D. Wright, and S. Elizabeth, “SURA Final Report: A super-regional testbed to improve models of environmental processes on the U.S. atlantic and gulf of mexico coasts,” tech. rep., SURA, May 2012.
14. R. Rew and G. Davis, “Netcdf: an interface for scientific data access,” *Computer Graphics and Applications, IEEE*, vol. 10, pp. 76–82, July 1990.
15. P. Cornillon, J. Gallagher, and T. Sgouros, “Opendap: Accessing data in a distributed, heterogeneous environment,” *Data Science Journal*, vol. 2, pp. 164–174, 2003.
16. Climate & Forcast (CF) Metadata Conventions, http://cfconventions.org/latest.html.
17. R. Luettich and J. Westernick, “Formulation and numerical implementation of the 2D/3D ADCIRC finite element model version 44.xx,” tech. rep., University of North Carolina at Chapel Hill, 2004.
18. C. Chen, R. C. Beardsley, and G. Cowles, “An unstructured grid, finite-volume coastal ocean model (fvcom) system,” *Oceanography*, vol. 19, pp. 78–89, 2006.
19. Y. Zhang and A. M. Baptista, “SELFE: A semi-implicit Eulerian-Lagrangian finite-element model for cross-scale ocean circulation,” *Ocean Modelling*, vol. 21, pp. 71–96, 2008.
20. J. Chen, W. Shaffer, and A. Gilad, “SLOSH–A hurricane storm surge forecasting model,” *Preprints, Oceans*, vol. 81, pp. 314–317, 1984.