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 is an open-source web service for the visualization and qualitative assessment of distributed geospatial data. The modular cross-platform Python implementation of SCI-WMS allows the service to keep pace with the rapid developments in the geospatial data science community 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 data using SCI-WMS and details the deployment of SCI-WMS for visualizing model data and simulations within the scope of the U.S. Integrated Ocean Observing System (IOOS) Coastal and Ocean Modeling Testbed (COMT) project [15].

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 catalog service which provides a single view of the distributed data to end users [15, 25, 4]. The institution responsible for maintaining the catalog compiles metadata regarding externally hosted data, exposed to the catalog by registered data producers. Such federated datasets potentially span petabytes of

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

information, may be composed of millions of files in different formats and are typically generated and hosted by vastly different systems located across the globe. End users (such as analysts or scientists) interface with the catalog to search through the aggregated metadata and interact with particular data of interest, agnostic to distributed nature of the database.

Although a decentralized approach to data management offers many advantages, data reduction and analysis tools have been slow to adapt to the distributed framework. There exists an abundance of applications for visualizing cartographic data on local computing resources, requiring analysts to download local copies of datasets and potentially reformat the data into the appropriate file format before processing and analysis can begin. Even if an end user has access to sufficient resources to download and process a dataset of interest, tools designed for centralized local systems increase project costs in terms of bandwidth usage, time and storage. Additionally, coupling centralized processing with decentralized storage introduces 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 and reach conflicting conclusions. Normalizing these results introduces a potential point of error and likewise increases project costs in terms of time and accuracy.

The Open Geospatial Consotrium (OGC) defines the Web Mapping Service (WMS) [16] standard that specifies how a compliant visualization server responds to HTTP requests from a OGC-WMS compliant client application. SCI-WMS is an implementation of an OGC-WMS server which responds to requests from clients returning metadata, data or geo-registered visualizations. While there exists other OGC-WMS compliant solutions including ncWMS [1], MapServer [12], and GeoServer [18], SCI-WMS and ncWMS are the only platforms which support NetCDF [19], a community standard file format. Additionally, SCI-WMS is the only OGC-WMS service to support modern model outputs which associate data with unstructured geo-registered topologies as outlined in section 2.3.

SCI-WMS is designed in such a way as to fill significant gaps in cooperative and distributed geoscientific computing. Data redundancy is minimized by avoiding dataset replication by fetching only the minimal amount of information from a distributed datastore to fulfill each OGC-WMS request. SCI-WMS enables end users to generate scientific visualization using OGC-WMS compatible clients, which may be simple web browser applications, facilitating rapid and consistent algorithmic and parametric comparisons. Furthermore, because SCI-WMS may be deployed on a server with dedicated hardware for storage, processing and visualization, SCI-WMS lowers costs and barriers to entry for analysts who would otherwise have to invest in the necessary local cyberinfrastructure to download and visualize local copies of distributed data.

2 SCI-WMS

SCI-WMS[1] is an open-source Python implementation of the Open Geospatial Consortium (OGC) Web Mapping Service (WMS) protocol [16] using standard cross-platform numerical software, NumPy [22], Matplotlib [10] and the Django [6] web framework for generating and serving visual content. The OGC-WMS specification defines a Representational State Transfer (REST) API [8] which responds to standardized HTTP GET requests from a WMS client for serving rasterized visualizations of geospatial data. A typical WMS request specifies a data layer and region of interest with optional metadata such as rendering style parameters. A WMS response may include additional information regarding the selected data or a visualization in the form of a PNG or other standard image format. A base WMS client has been developed in JavaScript which gives analysts the ability to generate and interact with visualizations using only a web browser [20].

While SCI-WMS is OGC-WMS compliant, it is augmented with services for automatically interacting with standard metadata catalogs such as the OGC Catalog Service for the Web (CSW) [17], allowing SCI-WMS to autonomously track dynamic distributed datasets. Data to be visualized by SCI-WMS should be exposed by data producers in one of the many community standard formats for geoscientific gridded data such as NetCDF, HDF/HDF5 [21], or GRIB/GRIB2 [26] with accompanying metadata adhering to the CF (Climate and Forecast) metadata conventions [7].

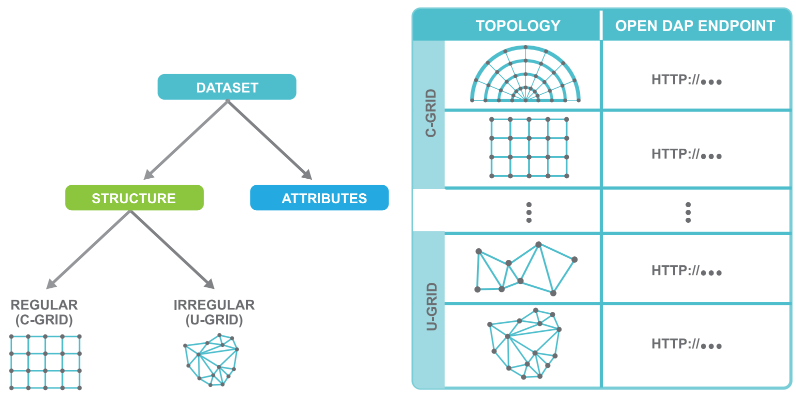
Though the OGC-WMS specification standardizes client-server communication, implementations vary dramatically in how a particular system fulfills the a WMS request. Vital to the efficiency of SCI-WMS is the decomposition of a registered dataset into **structure** (also known as **topology**) and **attributes** 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, typically via HTTP request. 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 devoid of a closed-form representation, a digital computer must store measurements of the function at discrete samples taken in

[1] https://github.com/brandonmayer/sci-wms

a given 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 [24]. 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 [23], 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*

A c-grid (also known as **regular** or **structured**) topology refer to structures that can be defined analytically, e.g., rectilinear or curvilinear grids. Storing a c-grid topology amounts 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

A u-grid (also called **irregular** or **unstructured**) topology is defined as a set of sample locations with connectivity relations that do not admit a closed-form representation. 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, but have larger storage and processing requirements compared to regular topologies. As storage and processing hardware has become more accessible, unstructured data has become more prevalent due to the flexibility of the representation. However, most existing visualization technologies in the geospatial community can only render regularly structured datasets. SCI-WMS is one of the first visualization services to support rendering irregularly structured data.

Any NcML [11] 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 as binary R-tree [9] 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. Another algorithm may have 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 define 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 within the region of interest, using the implicit representation for c-grid and R-trees for u-grid topologies, then fetching the corresponding external attributes. For rendering, the sample connectivity within the area of interest is reconstructed from the connectivity array which is utilized for interpolation.

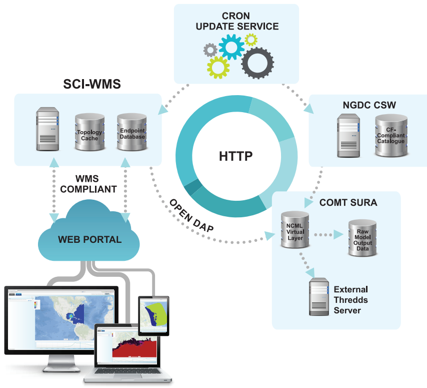
3 SCI-WMS Deployment: U.S. IOOS COMT Testbed Project

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 [15]. 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 of distributed data with a unified visualization framework.

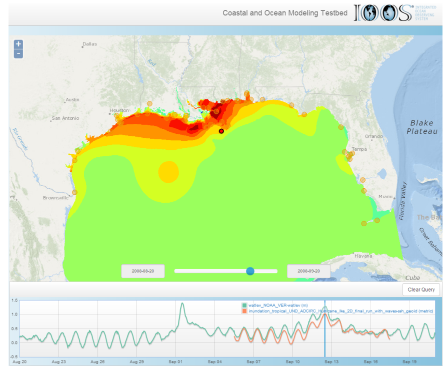
Figure 2a outlines the cyberinfrastructre behind the deployment of SCI-WMS for the COMT project[2]. The National Oceanic and Atmospheric Administration (NOAA) - National Geophysical Data Center (NGDC) geoportal indexes public geophysical datasets and provides an OGC CSW service to query datasets by their 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 [14]. 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 resources. However, accompanying the raw data is an NcML virtual layer which exposes each dataset as a single NetCDF [19], OPeNDAP [5] accessible object. Furthermore, the NcML facade presents a consistent set of meta information in accordance to CF-Conventions [7] 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 inunda-*

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

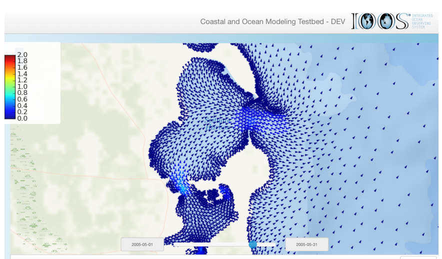


(a) Overview of the SCI-WMS deployment for the U.S. IOOS COMT project. SCI-WMS updates its topology and endpoint database via a nightly service which queries CF-Compliant datasets cataloged by NGDC. Model data is hosted on an external web server exposed by an NcML facade as a single NetCDF data structure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

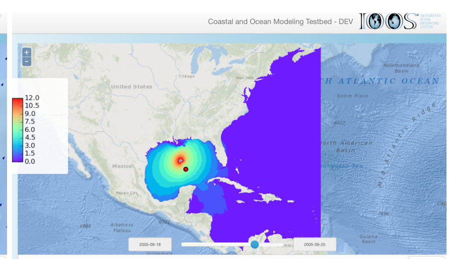


(b) Comparison of ADCIRC (unstructured topology) model results with observed water levels in the Northern Gulf of Mexico for Hurricane Ike. Verified observed water levels are from NOAA's Station 8760922 (red dot on map). The map shows modeled water levels (in meters above the geoid) at the peak of the 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.

tion [15]. For each modeling group, SCI-WMS successfully generates consistent visualizations of data generated by ADCIRC [13], FVCOM [2], SELFE [27] and SLOSH [3] 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.



(a)



(b)

Fig. 3: (a) Visualizing Current direction and speed in the Chesapeake Bay area. (b) Visualizing significant sea surface wave height along the eastern coast of the United States.

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 2b 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 3a visualizes current direction and speed in the Chesapeake Bay area. Figure 3b 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.

References

[1] Blower J, Gemmell A, Griffiths G, Haines K, Santokhee A, Yang X (2013) A web map service implementation for the visualization of multidimensional gridded environmental data. Environmental Modelling & Software 47(0):218 – 224
[2] Chen C, Beardsley RC, Cowles G (2006) An unstructured grid, finite-volume coastal ocean model (fvcom) system. Oceanography 19:78–89
[3] Chen J, Shaffer W, Gilad A (1984) SLOSH–A hurricane storm surge forecasting model. Preprints, Oceans 81:314–317
[4] Cherenak A, Foster I, Kesselman C, Salisbury C, Tuecke S (2000) The data grid: Towards an architecture for the distributed management and analysis of large scientific datasets. Journal of Network and Computer Applications 23(3):187 – 200
[5] Cornillon P, Gallagher J, Sgouros T (2003) Opendap: Accessing data in a distributed, heterogeneous environment. Data Science Journal 2:164–174
[6] Django (2014) [Computer Software] https://djangoproject.com
[7] Eaton B, Gregory J, Drach B, Taylor K, Hankin S (2014) NetCDF Climate and Forecast (CF) Metadata Conventions. URL http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/build/cf-conventions.pdf
[8] Fielding RT, Taylor RN (2002) Principled design of the modern web architecture. ACM Trans Internet Technol 2(2):115–150, DOI 10.1145/514183.514185, URL http://doi.acm.org/10.1145/514183.514185
[9] Guttman A (1984) R-trees: A dynamic index structure for spatial searching. In: Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, SIGMOD '84, pp 47–57, DOI 10.1145/602259.602266, URL http://doi.acm.org/10.1145/602259.602266

[10] Hunter JD (2007) Matplotlib: A 2d graphics environment. Computing In Science & Engineering 9(3):90–95
[11] Jerard RB, Ryou O (2006) Ncml; a data exchange format for internet based machining. Int J Comput Appl Technol 26(1/2):75–82
[12] Lime, Stephen (2014) MapServer. URL http://www.mapserver.org/
[13] Luettich R, Westernick J (2004) Formulation and numerical implementation of the 2D/3D ADCIRC finite element model version 44.xx. Tech. rep., University of North Carolina at Chapel Hill
[14] Luettich RA, Wright LD, Elizabeth S (2012) 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
[15] Luettich RA, Wright LD, Signell R, Friedrichs C, Friedrichs M, Harding J, Fennel K, Howlett E, Graves S, Smith E, Crane G, Baltes R (2013) Introduction to special section on The U.S. IOOS Coastal and Ocean Modeling Testbed. Journal of Geophysical Research: Oceans 118(12):6319–6328
[16] Open Geospatial Consortium Inc (2006) OpenGIS Web Map Server Implementation Specification. URL http://www.opengeospatial.org/standards/wmsa
[17] Open Geospatial Consortium Inc (2007) OpenGIS Catalogue Services Specification. URL http://www.opengeospatial.org/standards/cat
[18] OpenGeo (2014) GeoServer. URL http://geoserver.org/
[19] Rew R, Davis G (1990) Netcdf: an interface for scientific data access. Computer Graphics and Applications, IEEE 10(4):76–82, DOI 10.1109/38.56302
[20] RPS-ASA (2014) COMT-UI: U.S. IOOS Coastal and Ocean Modeling Testbed (COMT) User Interface. Computer Software, https://github.com/asascience-open/comt-ui
[21] The HDF Group (1997-NNNN) Hierarchical Data Format, version 5. Http://www.hdfgroup.org/HDF5/
[22] Walt SVD, Colbert SC, Varoquaux G (2011) The numpy array: A structure for efficient numerical computation. Computing in Science & Engineering 13(2):22–30
[23] Weiler KJ (1986) Topological structures for geometric modeling. PhD thesis, Rensselear Polytechnic Institute
[24] Will Schroeder BL Ken Martin (2006) The Visualization Toolkit: An Object-Oriented Approach to 3D Graphics, 4th edn. Kitware Inc.
[25] Williams DN, et al (2009) The earth system grid: Enabling access to multi-model climate simulation data. Bulletin of the American Meteorological Society 90(2):195–205
[26] World Meteorological Organization (WMO) Commission for Basic Systems (2003) Fm 92 grib edition 2. URL http://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/GRIB2_062006.pdf
[27] Zhang Y, Baptista AM (2008) SELFE: A semi-implicit Eulerian-Lagrangian finite-element model for cross-scale ocean circulation. Ocean Modelling 21:71–96, DOI 10.1016/j.ocemod.2007.11.005