SCI-WMS: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 python implementation of an OGC WMS [1] service 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[1] 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 [2].

1 Motivation

Due to the explosion of the amount of atmospheric, oceanagraphic, 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 datastore. Modern data management has been shifting hosting and maintenance responsibilities of large datasets to multiple participating institutions unified by a catalogue service to provide a single, unified view of the distributed data to the end user or analyst.

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

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

A common practice is for data producing organizations to host their own data to be exposed to the catalogue service via a particular communication protocol. The institution responsible for maintaining the catalogue then provides a unified view of the aggregated dataset to end users by compiling meta data associated with the external datum. Such federated datasets may span petabytes of geospatial information, be composed of millions of files in different formats all generated and hosted by vastly different systems located across the globe. Additionally, the catalogue can grow or shrink as new datasets or participants are added and removed from the federation. By interacting with the catalogue, the end user (such as an analyst) can navigate and search through the aggregated meta data and interact with particular datasets of interest, agnostic to distributed nature of the database.

While a decentralized approach to data management offers many advantages such as robustness to failure (a failure at any one organization only effects the data associated with that organization), data reduction and analysis tools have been slower to adapt to the new framework. For example, there are an abundance of applications for visualizing cartographic data on a single machine. However, many such programs are designed to deal with the output of specific modeling applications or data saved in a particular file format. To use these tools, analysts must download local copies of datasets to process a particular dataset with their local tools. This increases the projects cost in terms of bandwidth usage and time needed to wait for downloads to finish. Additionally, such a decentralized-local analytical workflow increases 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 comparisons of the same data. Normalizing comparisons and visualizations made with different software introduces another potential point of error and increases the costs of analysis in terms of time and accuracy.

SCI-WMS is a web service designed to solve many of the aforementioned problems. By maintaining a list of web accessible endpoints, SCI-WMS is able to transparently produce consistent visualizations for federated data. While SCI-WMS implements the OGC WMS protocol, it is augmented with services for reading standard data-aggregation catalogues such as CSW [**?**], allowing SCI-WMS to autonomously track dynamic federations.

SCI-WMS uses NcML (NetCDF Markup Language) as a data abstraction layer. This allows data hosting agencies can run their modeling software and host it in their chosen environment which they expose to SCI-WMS via the NcML file. This avoids writing custom i/o software or reformat each dataset to a standard format. Additionally, SCI-WMS is CF-Compliant [3], offering consistent views of endpoints which adhering to the CF-Metadata conventions embedded in the NcML file.

SCI-WMS saves costs by providing a simple interface for analysts to generate consistent visualizations of federated datasets. Perhaps more importantly, by the introduction of a single web-based visualization service for federated data ensures qualitative assessments and conclusions are made on equal footing regardless of analyst or data origin.

2 SCI-WMS

SCI-WMS is an open-source implementation of the Open Geospatial Consortium's (OGC) Web Map Service (WMS) standard which specifies an HTTP interface for generating rasterized visualizations of geospatial data [1]. SCI-WMS is implemented in Python using the Django[2] web framework and standard cross-platform numerical software, NumPy [4] and Matplotlib [5] for generating visual content. Additionally, the open-source python implementation provides a cross-platform WMS solution which can leverage the suite of tools developed by the geospatial data analysis community such as pyugrid[3] to maintain pace with the latest geospatial software and standards developments such as unstructured grid support and CF-UGrid Compliance[4].

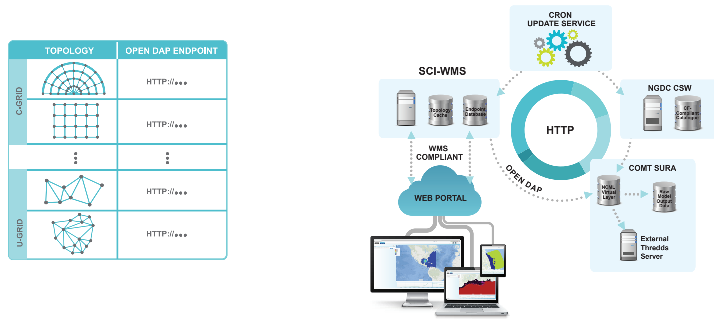
Visualization tools often decompose data into structure and attributes. Structure encapsulates both the locations (where) and geometry (connectivity) relations between attributes and constrain the visualization task. Similarly, SCI-WMS decomposes an externally hosted dataset into a structure (topology), defined as a geo-referenced spatial set of locations and geometries (connectivity of locations), and the underlying data layer. For example, atmospheric and oceanagrphic models typically defined a fixed topology over a spatial extent of the globe. The geometry of the topology defines locations where variables of interest are to be computed and the geometry of the topolgy governs how information propagates through the model. In the oceanagraphic domain, the model may predict a sea-surface-height attribute as well as wind and current directions and magnitudes for a particular region on the earth. When an analyst generates a visualization, a particular region and data layer is specified, such as Eastern Coast of the U.S. and sea-surface-height.

Vital to the efficiency of SCI-WMS is the abstraction of a dataset into two entities: a topology, defined as a geo-referenced spatial structure and numerical model output as visualized in Figure 1a. SCI-WMS creates a local topology cache for efficiently computing spatial neighborhoods with respect to topology structure. For storage efficiency, model output is not replicated locally but referenced via OGC compliant web-services. Because geospatial WMS requests are commonly restricted to a subset of the Earth's surface, SCI-WMS uses the local topology to compute the subset of model data needed to fulfill each request prior to accessing the external data. Furthermore, by classifying topologies as either regular or unstructured, efficient algorithms and data structures are exploited to optimize the computation of relevant model data subsets.

[2] https://www.djangoproject.com/

[3] https://github.com/pyugrid/pyugrid

[4] https://github.com/ugrid-conventions/ugrid-conventions/blob/v0.9.0/ugrid-conventions.md



(a) SCI-WMS topology and endpoint data store.

(b) Overview of the SCI-WMS deployment for the U.S. IOOS COMT project.

Fig. 1: SCI-WMS architecture.

2.1 Local Topology Cache

Visualization tools often decompose data into structure and attributes. Structure encapsulates both the locations (where) and geometry (connectivity) relations between attributes and constrain the visualization task. Similarly, SCI-WMS decomposes an externally hosted dataset into a structure (topology), defined as a geo-referenced spatial set of locations and geometries (connectivity of locations), and the underlying data layer.

For example, atmospheric and oceanagrphic models typically define a fixed topology covering a particular spatial extent of the earth. A model will extimate attributes of interest such as sea-surface-height, wind or current magnitudes and directions. The topology of the model referrs to positions and connectivity of the locations for which attributes are to be computed. While a model produces data associated with every location specified by the topology, visualizations are typically generated for a subset of the topology (a region of interest) and a single attribute such as current direction. It is therefore paramount to the efficiency of visualization software to represent topologies in such a way as to optimize topology storage and reduction to facilitate efficient attribute retrieval.

Topologies can be further classified into one of two categories, structured and unstructured, referred to as **c-grid** and **u-grid** topologies respectively in SCI-WMS nomenclature. Structured topologies can be represented using analytically while unstructured topologies require explicit enumeration. These differences admit different optimal data structures and algorithms for storage and processing.

To this end, when an dataset endpoint is submitted to SCI-WMS, the topology of the underlying endpoint is stored locally to SCI-WMS and a database of topology-endpoint associations are maintained as visualized in Figure 1a.

2.2 Structured (c-grid) Topologies

Structured or **c-grid** topologies refer geo-referenced locations and geometries that can be analytically specified, e.g., rectilinear or curvilinear grids. When an analyst requests a view of a particular region, the subseting

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

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. Figure 1b outlines the cyberinfrastructre behind the deployment of SCI-WMS for the COMT 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 oceanagraphic 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 [2]. Key to the U.S. IOOS COMT mission is an extensible and universally available tool for quickly visualizing and assessing a diverse set of coastal modeling data. SCI-WMS is a general OGC WMS solution for serving rasterized visualizations of geospatial data which has been deployed for the COMT project to provide visualizations of a wide range of scientific data.

The NOAA[5]–NGDC[6] Geoportal indexes public geophysical datasets and provides an OGC Catalogue Web Service (CSW) to query datasets by their metadata attributes. SCI-WMS queries the NGDC Geoportal at regular intervals updating both the topologies and OPeNDAP links for all new or modified datasets.

Raw coastal data is hosted by the Southeastern Universities Research Association (SURA) on a dedicated server for the COMT project [6]. 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 each dataset is an NcML virtual layer which exposes each dataset as a single NetCDF object which may be accessed via OPeNDAP[7]. Furthermore, the NcML facade presents a consistent set of meta information in accordance to CF-Conventions [3] providing services like SCI-WMS access to the raw data through a uniform interface.

[5] National Oceanic and Atmospheric Administration

[6] National Geophysical Data Center

[7] http://www.opendap.org/

Currently, SCI-WMS is used to visualize data from the first phase groups of IOOS COMT program: *estuarine hypoxia, shelf hypoxia and coastal inundation* [2]. For each modeling group, SCI-WMS successfully generates visualizations from ADCIRC, FVCOM, SELFE and SLOSH models and serves as a use-case of how SCI-WMS can be leveraged as a scalable solution for delivering consistent visualizations of scientific data to a diverse community.

4 Results

Figure 2 shows a web portal utilizing the SCI-WMS backend to compare ADCIRC model output for Hurricane Ike with water levels observed by NOAA stations. 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.

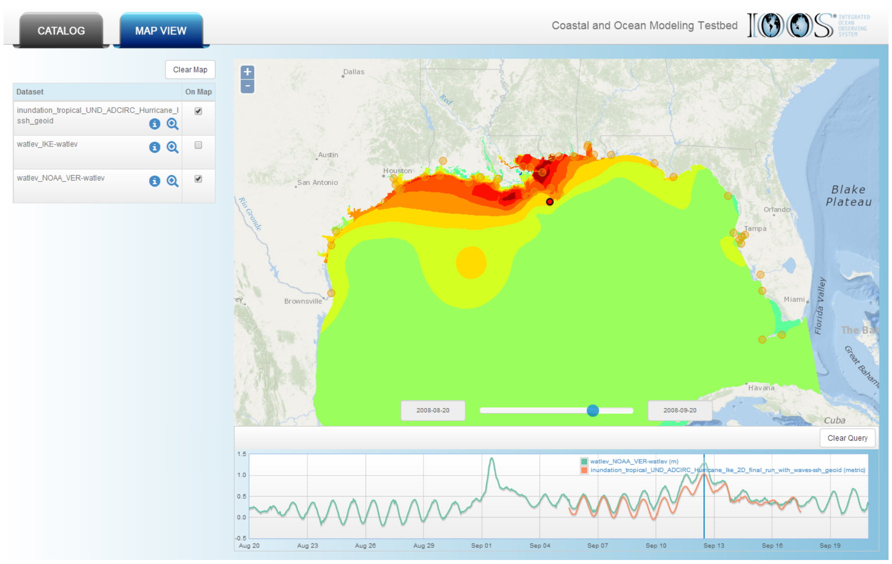


Fig. 2: Comparison of ADCIRC (unstructured topology) model results with observed water levels in the Northern Gulf of Mexico for Hurricane Ike. Verified observed water levels are from NOAA's Station 8760922 (red dot on map). The map shows modeled water levels (in meters above the geoid) at the peak of 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.

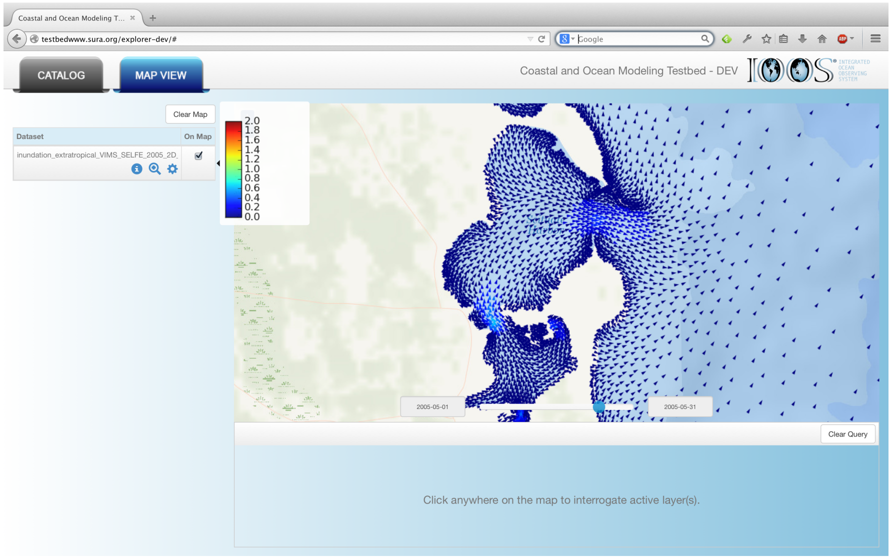


Fig. 3: SELFE model of current direction and speed in the Chesapeake Bay area.

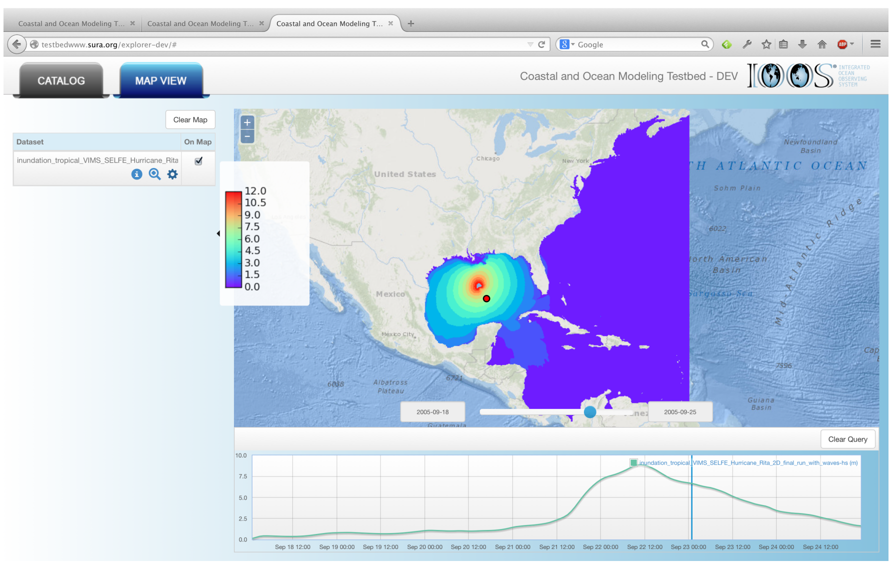


Fig. 4: Visualizing SELFE model of significant sea surface wave height along the eastern coast of the United States. The underlying topology is an unstructured grid with over 5 million nodes which SCI-WMS can handle in real time.

References

1. Open Geospatial Consortium http://www.opengeospatial.org/standards/wms [Accessed: 2014-

07-24], "OpenGIS Web Map Service."
2. 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.
3. CF Conventions Document, http://cfconventions.org/latest.html.
4. 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.
5. J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
6. 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.