CUAHSI Subsetting Service Functional Specification

Version 0.1

Last Updated: 09-20-2018

Contents

1 Project Overview

Large scale surface water and groundwater models are an essential tool for improving our understanding of the dynamic interaction between the water cycle and human activity. This is especially true when investigating human impacts, extreme hydrologic events, and future water resource availability. Results from these models advance hydrologic science and inform neighboring research disciplines—for example drought and flooding forecasts influence research in biology, sociology, and economics. As a result, community intellectual contributions to the physics, configuration, and validation of continental-scale models is essential to improving the usefulness and adoption of these models within the academic community. Due to the computational-scale of these models, they often require specialized computing hardware and vast amounts of domain and forcing data, making it difficult for the broader water science community to directly engage in development efforts. The Consortium of Universities for the Advancement of Hydrologic Science, Inc and the National Center for Atmospheric Research have recognized this challenge and are collaborating to improve the accessibility of model domain data to lower the barrier of entry for using and applying these models and engage a wide variety scientists and a diverse spectrum of expertise.

The purpose of this application is to introduce a collaborative effort for preparing, publishing, and sharing subsets of the National Water Model input data at watershed scales. With a combination of modern cyberinfrastructure techniques and state-of-the-science modeling tools, researchers will have access to subsets of National Water Model information that would otherwise require extensive computational resources. This work will provide the foundation onto which similar efforts can be applied to other large-scale model simulations and input data.

1.1 Project Goals

The goals specific to version 0.1 of the CUAHSI subsetting service are general and rudimentary within the long term context of the project. The goal is to provide a minimal effort proof-of-concept tool, to evaluate interface usability and integration with NCAR code. Users will have the ability to define a bounding box using specified coordinates (WGS84). These coordinates will be supplied to various R and Python scripts used for subsetting the National Water Model domain input files. Domain files will be collected from the National Centers for Environmental Prediction (NCEP) publicly hosted archive of NWM v1.2.2 data (http://www.nco.ncep.noaa.gov/pmb/codes/nwprod/nwm.v1.2.2/). The tool will consist of minimal consideration for user experience and will focus almost entirely on establishing the foundation onto future versions will build on.

1.2 Major Features

Version 0.1

- All NWM version 1.2.2 domain data
- API endpoint for submitting jobs using bounding box coordinates
- API endpoint for checking job status
- API endpoint for downloading subsetted data

2 Scenarios

The current release of this software is designed to meet the requirements of the following scenarios:

2.1 Scenario: Research Scientist - Jose

Jose is a research scientist at a US university who spends most of his time exploring and applying new hydrologic models to predict flooding in his local watershed. Out of necessity, he's learned various programming languages, geographic information system technologies, and spends most of his time working in Linux. While talking with colleagues at the latest academic gathering, he was informed of the efforts of the National Weather Service to forecast flooding across the continental United States (CONUS) using a variant of the WRF-Hydro model called the National Water Model (NWM). Shortly thereafter he decided to compare his existing models to the NWM for predicting river discharge at his local watershed, without considering the technological difficulties in doing so. When he got back to his office and read through the theoretical and technical documentation, he quickly became overwhelmed, and began frantically searching for small scale applications of the NWM. Minutes before giving up on this endeavor, he stumbled upon a CUAHSI service that extracts the NWM domain for small bounding boxes in the CONUS. Given his years of experience using GIS, it was trivial for him to define a bounding box for his watershed using WGS 1984 coordinates. Within minutes he had downloaded all of the domain files for his watershed. Ecstatic that this just saved him weeks of work, he decided to go on vacation for the remainder of the month.

3 Non Goals

This current version of this software will *not* support the following features:

3.1 Release Non Goals

- There will not be public access to the subsetting algorithms implemented in this version of the software. This is because version 0.1 represents a prototype of the subsetting architecture that could break at any time. Future versions will be made publically available.

- This software release is limited to catchment scale subsetting. We have not evaluated the performance of this service, and therefore have arbitrarily decide that this version of the application should only be applied at HUC-12 scales, however this limit will not be currently enforced. Future releases will explicitly state and enforce the bounding box subsetting limits in terms of square kilometers.

- Since the focus of this work is to reproduce NWM analysis simulations, nudging will not be included in the subset. Future versions will investigate the possibility of including these data.

- This version of the software will reuse as much existing code as possible. As a result, it will leverage multiple programming languages, wrapped in Python 3.6. Future versions will migrate R code to Python as necessary.

- The current version of the software will perform naive bounding box subsetting, which will not use stream tracing and therefore may not capture the entire upstream catchment in the subsetted domain. Future versions will evaluate how this can be done.

- This version will not include model restart files because we currently do not have access to the operational restart files.

3.2 Project Non Goals

- Due to the sensitive nature of the CONUS lake and reservoir datasets, they will not be included in any version of this service.

- Subsetting WRF-Hydro forcing data is beyond the scope of this work and is not be included in the current release or future releases. Instead CUAHSI will continue to work

with researchers at Brigham Young University to develop the technology for subsetting model forcing data.

4 Requirement Specifications

4.1 User Requirements

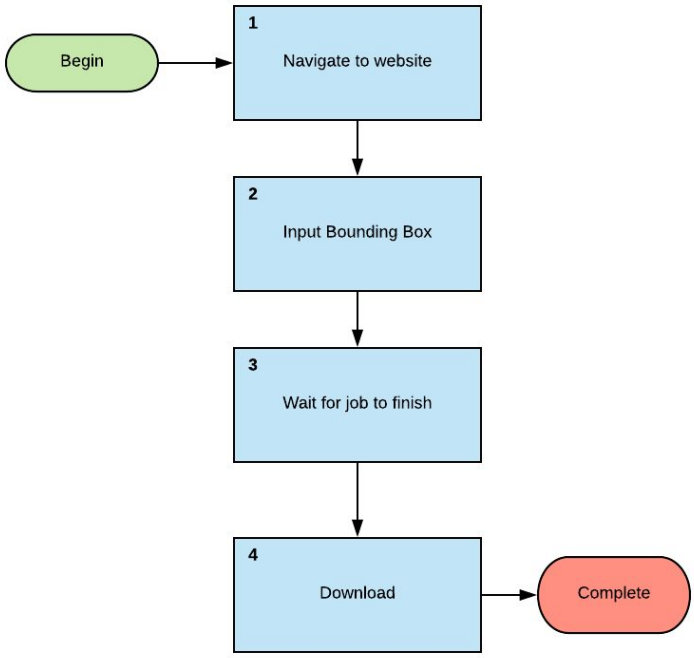

Figure 1. Application-level flowchart for user in scenario 2.1.

The actions of user scenario 2.1 are outlined in Figure 1, and consists of the following:

- User navigates to the CUAHSI domain subsetting service and is presented with a basic web page to enter input coordinates.
- User specifies four coordinates that define a bounding box for their domain and submits the job. This bounding box should be on the scale of a typical HUC-12, however sizes will not be enforced. Coordinates must be in the WGS84 coordinate system since it is common and will enable this service to support other subsetting applications in the future.
- User is redirected to a page that displays the status of their job.

- A link to provided to a *.tar.gz archive containing the following files specific to their domain. These files will enable them to run a subsetted NWM v1.2.2 configuration of WRF-Hydro:
    -

4.3 Data Requirements

- The service must provide the official NWM v1.2.2 domain datasets hosted by the National Centers for Environmental Prediction: (http://www.nco.ncep.noaa.gov/pmb/codes/nwprod/nwm.v1.2.2/)

- Subsetting capabilities must be available for all regions within the contiguous 48 United States.

- National Water Model subset output must contain ONLY the following files:
    - Fulldom_hires.nc
    - Geo_em.d01.nc
    - GEOGRID_LDASOUT_Spatial_Metadata.nc
    - GWBUCKPARM.nc
    - Hydro2dtbl.nc
    - Route_Link.nc
    - Soil_properties.nc
    - Spatialweights.nc
    - Wrfinput_d01.nc
    - README.md

4.2 Operational Requirements

- There are no operational requirements for version 0.1 of this service.

5 Technical Details

5.1 System Architecture

This software architecture for this service consists of two primary components: a tornado web server (Python 3.6) and subsetting code (R, provided by NCAR). The general architecture is outline in the Figure 2. Future versions of this service will include, among other things, decoupling web service endpoints from the web site as well as load balancing and a reverse proxy (e.g. Nginx), and expanding the job scheduler to work with a variety of subsetting algorithms.

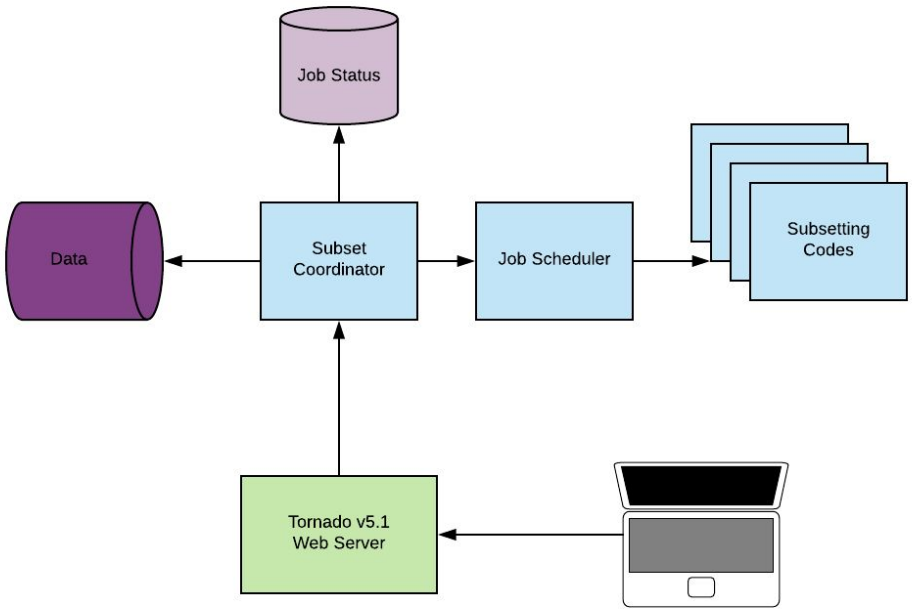

Figure 2: The general architecture of the subsetting service

5.2 API Endpoints

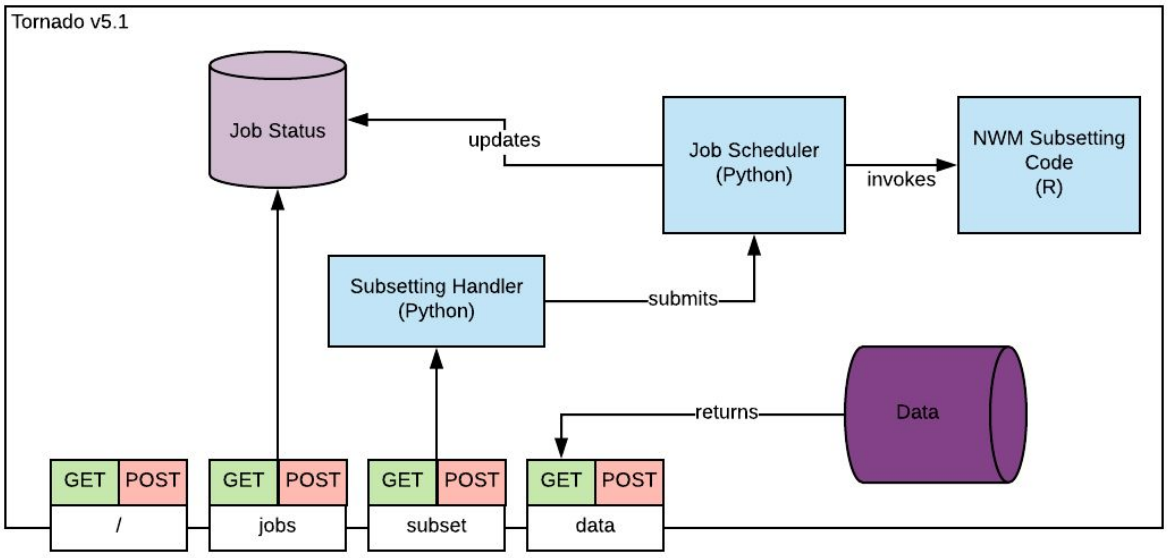

Figure 3: Overview of the subsetting service REST endpoints.

There are three REST endpoints in version 0.1 of the subsetting tool: *subset*, *jobs*, and *data* (Figure 3). The *subset* endpoint accepts bounding box coordinates in the WGS84 spatial reference system, and uses these data to (1) validate the bounding box, (2) convert into the coordinate system used by NWM (an Albers Conformal Conic variant), and (3) submit the job to run as a background task. The *jobs* endpoint is used to lookup the status of any given job via unique identifier. Finally, the *data* endpoint is used to download the subsetted domain data via unique identifier. Technical details for each endpoint outlined below.

GET - /jobs

- Description: gets the status of all jobs
- Returns: json string of all known jobs on the server
- Parameters: None

GET - /jobs/{id}

- Description: gets the status of the job {id} in json
- Returns: a json string for the status of job {id}
- Parameters:
    - id - the unique identifier for the job

GET - /subset?{params}

- Description: submits a new subsetting job
- Returns: a json string containing the a unique job identifier
- Parameters:

    - llat = the lower latitude of the bounding box in WGS84
    - llon = the lower longitude of the bounding box in WGS84
    - ulat = the upper latitude of the bounding box in WGS84
    - ulon = the upper longitude of the bounding box in WGS84

GET - /data/{id}

- Description: gets the file created by the subsetting algorithm
- Returns: the location of the file for {id}
- Parameters:
    - id - the unique identifier for the job

6 Disclaimer

This specification is a work in progress and will be updated as necessary throughout the development lifecycle. Although this project is a collaboration between CUAHSI and the WRF-Hydro team at NCAR, this undertaking is CUAHSI funded and the NCAR team will be under no obligation to support user inquiry, tool development, etc., relating to this service. This tool provides input domain files used in the analysis configuration of the National Water Model, version 1.2.2. With that said, this is not a complete operational configuration of the National Water Model and the results should be interpreted and communicated accordingly.