

# The U.S. Geological Survey Climate Geo Data Portal: An Integrated Broker for Climate and Geospatial Data

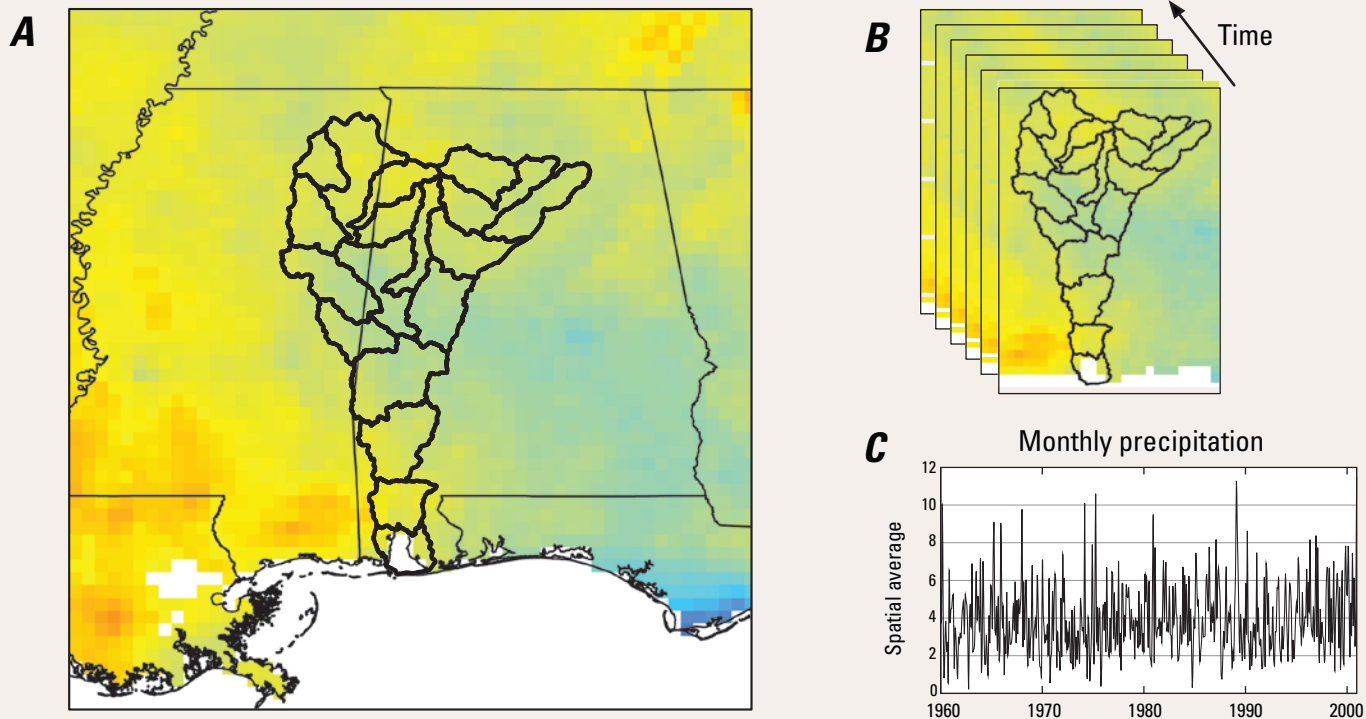
## Introduction

The increasing availability of downscaled climate projections and other data products that summarize or predict climate conditions, is making climate data use more common in research and management. Scientists and decisionmakers often need to construct ensembles and compare climate hindcasts and future projections for particular spatial areas. These tasks generally require an investigator to procure all datasets of interest en masse, integrate the various data formats and representations into commonly accessible and comparable formats, and then extract the subsets of the datasets that are actually of interest. This process can be challenging and time intensive due to data-transfer, -storage, and(or) -processing limits, or unfamiliarity with methods of accessing climate data. Data management for modeling and assessing the impacts of future climate conditions is also becoming increasingly expensive due to the size of the datasets. The Climate Geo Data Portal (<http://cida.usgs.gov/climate/gdp/>) addresses these limitations, making access to numerous climate datasets for particular areas of interest a simple and efficient task.

The portal allows a user to process data by:

1. selecting areas of interest by uploading a polygon shapefile or drawing on a map,
2. choosing from a searchable list of historical and future climate datasets,
3. selecting the variables and time period of interest from the dataset, and
4. configuring processing to return raw data or spatially averaged summaries.

Datasets available in the portal are hosted by the U.S. Geological Survey (USGS) or outside partners using standard data distribution protocols. These protocols and the high bandwidth connectivity of Geo Data Portal processing servers eliminate the need for users to download and manage large collections of data. The Geo Data Portal processing servers act as a data broker, conducting all integration and subsetting automatically, thus removing the need for computing hardware, software, and expertise generally required for these data processing tasks.



**Figure 1.** Schematic representation of Climate Geo Data Portal functions using the Mobile Bay watershed as an example of: *A*, grid of regional climate data with watershed polygons; *B*, subsets of regional grids for a specified time period as output by OPeNDAP Subset algorithm; *C*, gridded time series spatially averaged for a subwatershed as output by Area Grid Statistics (Weighted) algorithm.

## Algorithms

The Climate Geo Data Portal offers two processing algorithms for accessing regional or national scale gridded time-series data (fig. 1A). *Area Grid Statistics (Weighted)* attributes spatial statistics of gridded time-series datasets to any polygon or set of polygons. *OPeNDAP Subset* provides a subset of raw data for the area and time period specified. A user interface to interact with these algorithms is available at <http://cida.usgs.gov/climate/gdp/>.

### Area Grid Statistics (Weighted)

The Area Grid Statistics (Weighted) algorithm is the primary processing tool made available by the Climate Geo Data Portal. The algorithm calculates the mean, minimum, maximum, standard deviation, count, and sum of contributing cell weights of a gridded dataset for polygons in any commonly used coordinate reference system. Spatial statistics are calculated using an area weighted method where the fraction of a grid cell actually intersecting a polygon is used to weight the contribution from that cell. This is best illustrated by contrasting the count and sum of weights statistics returned by the algorithm. Count is the total number of cells sampled to construct the statistics for a given polygon. Sum of weights is the sum of grid cell weights, or the fractional number of cells, contributing to the statistics for a given polygon.

One time series of statistics is returned for each polygon selected from an uploaded shapefile. The coordinate reference system of the polygons provided, as opposed to the gridded climate data, is used for spatial intersection calculations. Outputs of these calculations are returned in a simple delimited text format to be plotted, used in a model, or applied to an analysis of climate. Figure 1C shows a schematic representation of this functionality in which a time series for a particular polygon is plotted to look at trend.

### OPeNDAP Subset

The standard data transfer protocol used by the Climate Geo Data Portal to access gridded data is called the OpenSource Protocol for a Network Data Access Protocol (OPeNDAP). This algorithm constructs a file containing a subset of a selected climate dataset derived from the larger (national, regional, etc.) dataset distributed using the OPeNDAP protocol (fig. 1B).

The OPeNDAP subset algorithm writes data into a format known as Network Common Data Format Version 3 (NetCDF3). This format is compatible with a wide array of geospatial and climate data access and manipulation software, but has some limitations. The technical details (and limitations) are beyond the scope of this document. For more detail, see Blodgett, 2011 and documentation available on the portal.

## Newly Introduced Datasets

It is important to note that the Climate Geo Data Portal is not limited to the datasets available from its built-in catalog. The address of a compatible data service can be given to the portal and it will work with the newly introduced dataset.

## Example Uses

The potential uses of the Climate Geo Data Portal described below illustrate how it can assist scientific research and resource management.

### Hydrologic Modeling

Interest in the impacts of future climate conditions on water resources is growing. Spatially continuous representations of historical climate time-series datasets are required for calibrating hydrologic models. After calibration, climate projections can be applied to the model, and model outputs like soil moisture and streamflow can be produced for projected future conditions. The portal can be used to assist in generating time series of climate data, past and future, attributed to watershed polygons or other hydrologic modeling units.

### Ecosystem Response to Changes in Climate

Understanding how ecosystems respond to changing climate conditions is one of the largest areas of study for ecosystem management and restoration. A scientist may have a set of field observations to be correlated with climate data and derived climate indices. The scientist can use the portal to access archives of climate data for the area near observation sites or for an ecoregion of interest, and can then use the historical climate data to construct correlations and models of ecosystem response. Thus climate projections can be accessed and used to make predictions about future climate impacts on a given ecosystem.

### Regional Adaptation Planning

Investigators can use the portal to access any data published via common open standards. For example, scientists and decisionmakers may be interested in assessing the range of possible climate impacts on drinking water supply or stormwater management infrastructure. The portal can be used to gain access to a particular subset of projections from many scenarios and projection techniques. By accessing such subsets, scientists and decisionmakers can form projection ensembles tailored for their municipalities or other areas by using the portal's area grid statistics. Customized projections can assist in understanding the range of predictions of future climate conditions for adaptation planning.

## Reference

Blodgett, D.L., Booth, N.L., Kunicki, T.C., Walker, J.L., and Viger, R.J., 2011, Description and testing of the Geo Data Portal: Data integration framework and Web processing services for environmental science collaboration: U.S. Geological Survey Open-File Report 2011-1157, 9 p.

## For more information contact:

David L. Blodgett  
USGS Center for Integrated Data Analytics  
8505 Research Way  
Middleton, WI 53562  
Phone: 608-821-3899  
email: [dblodgett@usgs.gov](mailto:dblodgett@usgs.gov)