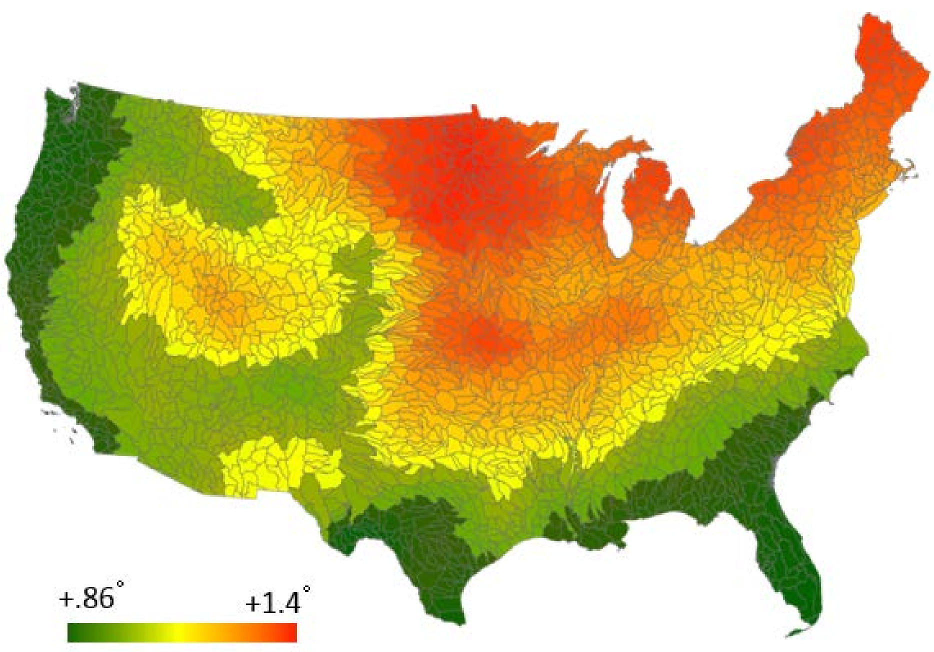
RECLAMATION

Managing Water in the West

Design and Development of a Prototype Tool for Integrated Climate Downscaling and Streamflow Prediction using Open Source GIS Software

Research and Development Office
Science and Technology Program
Final Report ST-2015-X9449-1



U.S. Department of the Interior
Bureau of Reclamation
Research and Development Office

March 2016

Mission Statements

The U.S. Department of the Interior protects America's natural resources and heritage, honors our cultures and tribal communities, and supplies the energy to power our future.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
T1. REPORT DATE March 17, 2016	**T2. REPORT TYPE** Research		**T3. DATES COVERED** 2013-2016
T4. TITLE AND SUBTITLE **Climate Analysis Toolkit** **Design and Development of a Prototype Tool for Integrated Climate Downscaling and Streamflow Prediction using Open Source GIS Software**			**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** R13AC10030 **5c. PROGRAM ELEMENT NUMBER** 1541 (S&T)
6. AUTHOR(S) Dustin Woodbury Dr. Daniel Ames Gregory Gault			**5d. PROJECT NUMBER** X9449 **5e. TASK NUMBER** **5f. WORK UNIT NUMBER** PN-3910
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Gregory Gault, BORGIS System Manager | Architect, Lands Resources Management Group, Pacific Northwest Region Office, Boise, ID			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Research and Development Office U.S. Department of the Interior, Bureau of Reclamation, PO Box 25007, Denver CO 80225-0007			**10. SPONSOR/MONITOR'S ACRONYM(S)** R&D: Research and Development Office BOR/USBR: Bureau of Reclamation DOI: Department of the Interior **11. SPONSOR/MONITOR'S REPORT NUMBER(S)** ST-2015-X9449-1

12. DISTRIBUTION / AVAILABILITY STATEMENT
Final report can be downloaded from Reclamation's website: https://www.usbr.gov/research/

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*
This research project developed prototype tools for acquiring, processing and managing global climate model data using an open source GIS platform. The project identified software requirements of people with the most immediate need for accessing and processing downscaled climate data, and then developed a suite of tools to: 1) assist users acquire global climate model data for user-defined areas of interest, 2) assist users select ensembles from numerous projection scenarios using statistical methods, and 3) generate forcing file outputs for selected hydrologic models.

15. SUBJECT TERMS: global climate model, GIS, open source, streamflow projection

16. SECURITY CLASSIFICATION OF: U			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Gregory Gault
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	U	72	**19b. TELEPHONE NUMBER** 303-445-2125

S Standard Form 298 (Rev. 8/98)
P Prescribed by ANSI Std. 239-18

PEER REVIEW DOCUMENTATION

Project and Document Information

Project Name: Design and Development of a Prototype Tool for Integrated Climate Downscaling and Streamflow Prediction using Open Source GIS Software WOID X9449

Document Final Report Document Author(s) Gregory Gault Document date March 3, 2016

Peer Reviewer Jennifer Cuhaciyan

Review Certification

Peer Reviewer: I have reviewed the assigned items/sections(s) noted for the above document and believe them to be in accordance with the project requirements, standards of the profession, and Reclamation policy.

Reviewer Jennifer Cuhaciyan Date reviewed 8/1/16
(Signature)

Disclaimer

The USDI Bureau of Reclamation (Reclamation) provides this report "as-is" without warranty of any kind, expressed or implied, including but not limited to, any warranties of fitness for a particular purpose. The burden for determining fitness for use lies entirely with the user. The user assumes all responsibility for accuracy, completeness, validity and appropriateness. In no event will the producers of Reclamation research be liable for any damages arising from the use of, or the inability to use, these research. All software and data associated with research provided by Reclamation are public domain, and the recipient may not assert any proprietary rights thereto nor represent data sets as other than government-produced data.

Acknowledgements

The Bureau of Reclamation acknowledges the contributions of the Dr. Daniel Ames, Associate Professor at the Department of Civil and Environmental Engineering at Brigham Young University, and his graduate students.

Acronyms and Abbreviations

CMIP3	Coupled Model Intercomparison Project – Phase 3
CMIP5	Coupled Model Intercomparison Project – Phase 5
GCM	Global Climate Model
GIS	Geographic Information System
LLNL	Lawrence Livermore National Laboratory
DHSVM	Distributed Hydrology Soil Vegetation Model
VIC	Variable Infiltration Capacity model

Executive Summary

The goal of this research project is to develop prototype tools for acquiring, processing and managing global climate model data using an open source GIS platform. Climate data is inherently geospatial in nature, with global climate models and their many scenarios and projections tied to 2-degree grid cells. A number of federal agencies, including Reclamation, have developed and applied downscaling methodologies resulting in vast libraries of climate data. Reclamation, in partnership with Lawrence Livermore National Laboratory (LLNL), has processed and now stores downscaled and bias-corrected climate model data for CMIP3 and CMIP5 projections. Users of these climate data libraries currently use somewhat complex online access tools to select and download the data they need to perform more detailed climate and hydrologic modeling.

The first task in this research project is to identify the software requirements of the people with the most immediate need for accessing and processing downscaled climate data. In general, this audience is climate and hydrologic modelers. A major finding of the requirements analysis is the need to output forcing files for projected changes in temperature and precipitation in formats ready to use in hydrologic models, in particular: the Variable Infiltration Capacity (VIC) and the Distributed Hydrology Soil Vegetation Model (DHSVM).

Armed with an understanding of user requirements, the major task of this research project is to develop a suite of prototype tools to: 1) assist users of global climate model data with statistical methods for selecting ensembles from numerous projection scenarios, 2) assist users in the selection and downloading of global climate model data for a user-defined area of interest, and 3) generate forcing file outputs for selected hydrologic models.

Contents

Introduction

This research focused on the three major capabilities found to be of greatest interest in the requirements analysis:

- Acquire downscaled, bias-corrected global climate model data
- Apply statistical methods to assist selecting ensembles of global climate model projection scenarios
- Generate output forcing files for selected ensembles of global climate model projection scenarios

Methods

This section describes: 1) the user requirements analysis, 2) the selected open source GIS platform, 3) the development of an open source software project, and 4) pre-processing of Global Climate Model data.

User Requirements Analysis

The first task of this research project was to better understand the needs of users of Global Climate Model data. This was accomplished through a user requirements analysis conducted in the Fall 2013. The methods and results of the user requirements analysis are reported Appendix II of this report.

Select an Open Source GIS Platform

The software requirements analysis evaluated a number of open source GIS platforms and selected HydroDesktop as the best candidate for developing the desired software tools. Refer to Appendix II of this report for details.

HydroDesktop

The built-in extension framework in HydroDesktop is particularly well suited to developing and integrating software tools. A partnership was initiated with Dr. Daniel Ames at the Department of Civil and Environmental Engineering at Brigham Young University, who is one of the main contributors to the HydroDesktop platform. Dr. Ames and graduate student staff leveraged the extension framework of HydroDesktop to build the Climate Analysis Toolkit, the main product of this research project. The user interface for this extension will use the extension framework, and the associated look and feel of HydroDesktop.

This allowed this research project to focus exclusively on functionality and workflows for climate data analysis and processing tools.

Open Source Software Development

The central part of this research project is the development of an open source software with the key capabilities desired by users of Global Climate Model data. Open source software solutions were sought to enable broader use and to inspire others to extend and expand capabilities over time.

Climate Analysis Toolkit

The Climate Analysis Toolkit, was initially established as an open source code project on the public site CodePlex by the project partner. The open source project was later moved to the Department of Interior's outlet for open source software projects on GitHub - https://github.com/usbr/ClimateAnalysisToolkit. Details are provided in the Appendix I of this report.

Data Acquisition and Processing

One of the main issues of users of Global Climate Model data is selecting and downloading the appropriate grids for their area of interest.

Global Climate Models

Global Climate Model data are available from a number of repositories around the world. Reclamation is a principal partner in a major repository of climate model data hosted at the Department of Energy's Lawrence Livermore National Laboratory (LLNL). LLNL has processed and serves downscaled CMIP3 and CMIP5 climate and hydrology projections. Data are available for two downscaling methodologies: Bias-Correction Spatial Disaggregation (BCSD) and Bias-Correction Constructed Analogues (BCCA). These data (LLNL, 2014) cover the period 1950-2099:

- 234 monthly BCSD CMIP5 projections of precipitation and monthly means of daily-average, daily maximum and daily minimum temperature
- 134 monthly BCCA CMIP5 projections of precipitation and daily maximum and daily minimum temperature
- 112 monthly BCSC CMIP3 projections of precipitation and monthly means of daily-average, daily maximum and daily minimum temperature

- 58 projections of daily BCCA CMIP3 projections of precipitation and daily maximum and daily minimum temperature

Access to these data is provided through a web application that allows users to initiate a request to select climate data grid files based on projection, geographic area, time period, and several other parameters. The request process is asynchronous. Once the request is submitted, the system cues the process, then sends the user an email with a link when the request processing is complete. Data files must then be downloaded manually, stored locally, and further processed to suit the user's purposes.

One of the major challenges of this research project grew out of a user requirement to simplify and automate the process of selecting and acquiring Global Climate Model data for a user-defined area of interest, typically a watershed. Refer to the Results section of this report describing how GCM data were pre-processed and made available for simplified discovery and download.

Characterizing and Selecting Climate Projection Scenarios

Users of Global Climate Model data defined a requirement for two statistical methods for characterizing climate projection scenarios—Delta or Hybrid Delta. These methods allow users to set one or more future periods to be evaluated, and select the collection of climate projection scenarios to be evaluated for the selected period(s).

Existing code created by a Reclamation hydrologic engineer using the open source R statistical software was leveraged as a base. This logic was re-written in C# programming language with some refinements and was incorporated into the code base of the Climate Analysis Toolkit. Refer to the report in Appendix I.

References

USBR, 2013, Climate Analysis Tools – Software Requirements Report, Bureau of Reclamation, Boise, Idaho, 2013 (see Appendix II of this report)

LLNL, 2014, Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections (website), http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/dcpInterface.html

Data Sets that support the final report

If there are any data sets with your research, please note:

- Datasets described in this report are located can be accessed using the Climate Analysis Toolkit in HydroDesktop, or can be accessed directly at:

  http://worldwater.byu.edu/climate/services/

- Point of Contact: Dr. Daniel Ames, dan.ames@byu.edu, 801-422-3620
- Description: CMIP5 Global Climate Models organized by hydrologic units, subbasins, or HUC8
- Keywords: global climate model, projections, CMIP5
- Approximate total size of all files: varies by geographic location

Appendix I

Final Project Report

Design and Development of a Prototype Tool for Integrated Climate Downscaling and Streamflow Prediction using Open Source GIS Software

Final Project Report

U.S. Bureau of Reclamation, Agreement Number R13AC10030 and R13AC10040

March 3, 2016

Dustin Woodbury, Graduate Research Assistant, Department of Civil and Environmental Engineering, Brigham Young University, Provo, Utah

Daniel P. Ames, Associate Professor, Department of Civil and Environmental Engineering, Brigham Young University, Provo, Utah

Gregory Gault, BORGIS System Manager | Architect, Bureau of Reclamation, Pacific Northwest Regional Office, Boise, Idaho

Background

Climate change models have the potential to significantly inform regional-scale watershed runoff simulations of future scenarios for the purposes of improved short and long-term water resources management. A number of General Circulation Models (GCMs) have been produced and proven useful in predicting trends in future climate conditions and are of increasing interest to hydrologists and water resource managers for planning purposes.

Given the growing interest in climate change research, the development and testing of GCMs continues to be a major focus of the watershed modeling community. GCMs are capable of providing projections of ground level air temperatures and precipitation values at various locations given specified atmospheric conditions. Using a sophisticated application of the Navier-Stokes equation, climate models are able to predict vector paths of hot air masses around the globe. By estimating atmospheric compositions, and the corresponding thermal irradiance captured by the atmosphere, GCMs predict air temperature and precipitation conditions for long-term forecasting over the next century. However, GCMs vary based on the preferred modeling approach of each parent research institution and how atmospheric composition is estimated.

In 1995 the Coupled Model Intercomparison Project (CMIP) was established under the World Climate Research Programme (WRCP) as a standard experimental protocol for atmosphere-ocean general circulation models. This collection includes atmosphere-ocean general circulation models from most major research institutions in the international climate modeling community. CMIP has collected model outputs from the pre-industrial climate simulations ("control runs") and 1% per year increasing-CO2 simulations. The latest collection is phase 5 (CMIP5) and includes 234 different projections for historical and future climate data at a spatial resolution of between 1/2° and 4° grid cells across the globe. Daily average air temperature and precipitation values for each month are available for these cells from 1950 through 2099.

Climate change across the western United States will impact the way water is stored, managed, and used over the next century and beyond. Therefore, the Bureau of Reclamation has authorized that steps be taken to develop tools that most effectively leverage the latest climate data in order to fulfill its mission to manage, develop, and protect water resources in an environmentally sound manner. In 2013 the Pacific Northwest Regional Office of the Bureau of Reclamation awarded a two-year grant to the Hydroinformatics Laboratory at Brigham Young University (BYU) to develop software tools to interactively identify watershed drainage areas related to critical management points in a stream network, select ensembles of downscaled and bias-corrected GCM projections for selected an area of interest, and generate forcing files for use in hydrologic models. It was also expected that supporting documentation be developed as well as a peer reviewed technical manuscript. These products were to be made available for public use through internet download.

The BYU Hydroinformatics Laboratory under the direction of Dr. Daniel Ames has recently completed this work and delivered the requested products. Additional components have also been delivered to meet the needs and feedback of Bureau of Reclamation users throughout the process. These changes have been documented. This work has been the primary research focus of civil engineering graduate student Dustin Woodbury and has received ancillary support from other students including Guy Werner who served as a software developer on the project. This report summarizes the work products delivered.

Climate Analysis Toolkit

Two principle challenges facing hydrologists when attempting to analyze climate models are which of the many models will provide useful scenarios for analysis and how to use this data to force hydrologic models in a statistically reliable way. In the past, this has been done by identifying changes in means when comparing two 30 year spans of time. The changes in each model for temperature and precipitation can then be sorted into clusters representing climate “scenarios” by percentile. These clusters are then used to create a composite change factor which is in turn used to force a watershed model. Previously, this has been an arduous task. To streamline this process and offer a quick workflow to perform these functions the BYU team has developed a toolkit within a graphical user interface that provides user defined inputs and customization.

The Climate Analysis Toolkit is a tool for integrated climate model analysis for streamflow prediction. It is powered by the open source GIS software package HydroDekstop. The toolkit allows users to download or manually input CMIP3 or CMIP5 projections and perform the following functions:

- Retrieve downscaled data from USGS server via HydroDesktop’s GIS interface
- Generate cluster graph and hybrid delta ensembles of useful climate scenarios
- Allow for input of VIC or DHSVM hydrologic models which can be adjusted within the toolkit

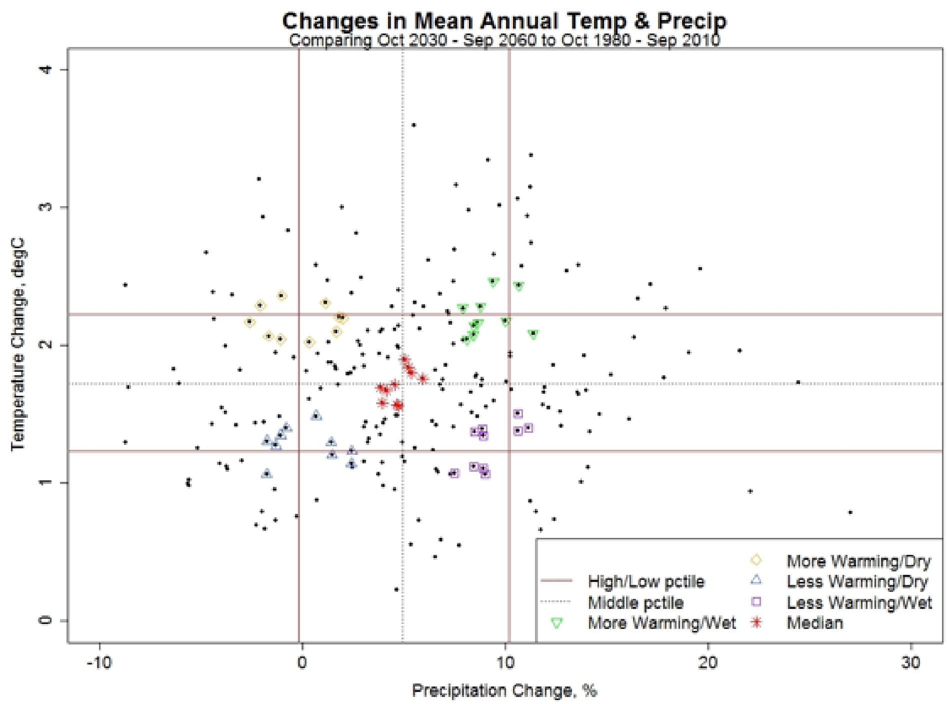


Figure 1 Scatter Plot of All 234 CMIP5 Models for One Watershed Produced in the Climate Analysis Toolkit

The Climate Analysis Toolkit offers a simple yet powerful tool for hydrologists and engineers to study climate change. Deploying the toolkit as an extension to HydroDesktop increases its accessibility by leveraging its familiar graphical user interface and by making it universally available as part of an open source software platform.

Recently the Pacific Northwest Regional Office of the Bureau of Reclamation has implemented the Climate Analysis Toolkit as part of the Columbia River Basin Climate Impacts Assessment. This Program is aimed to address Section 9503 of the SECURE Water Act (SECURE) which authorizes Reclamation to assess the risks to water supplies and demands posed by climate change, including changes in snowpack, changes in timing and quantity of runoff, changes in groundwater recharge and discharge, as well as changes in demands and consumptive usage within major river basins in the Western United States. The Assessment quantitatively evaluates the impacts associated with climate change as they relate to the mission of Reclamation including risks to water supplies.

The study utilizes Variable Infiltration Capacity (VIC) macroscale hydrologic models. In predicting future trends of water availability and climate behavior, the meteorological inputs (forcing files) to the VIC models need to be adjusted by change factors generated using hybrid delta ensembles of downscaled Global Circulation Models (GCM's). Using the Climate Analysis Toolkit, hydrologists were able to input downscaled GCMs, successfully generate hybrid delta ensemble change factors, and create climate adjusted forcing files for the VIC models used in this study.

In using the Climate Analysis Toolkit as part of this study, Bureau of Reclamation hydrologists provided useful feedback of how to make the software more useful for this study and other hydrologists interested in forcing file adjustment. The BYU team were able to make these suggested changes in the software and subsequently published a new release that was used in the aforementioned study. These changes include:

- Allowing for multiple forcing files to be adjusted at once
- Manual model selection and identification: users can now view all 234 models on the cluster graph and readily identify each model's name on the graph. Users may also opt to bypass the use of ensembles in selecting which model(s) to use and manually select models either from the graph or from a drop down list.

- Additional supported forcing file types: hydrologic models, like the Variable Infiltration Capacity model, can now be input not only as comma separated value files but as NetCDF files as well. These file types represent common formats of climate data.
- Pisces database: upon adjusting forcing files, users now have the ability to select a check box which prompts the program to automatically upload adjusted files to the Pisces database.

The adjusted files are now being used for climate change model runs for the Columbia River Basin and appear to be working successfully.

HUC-8 Based CMIP-5 Climate Model Dataset

To complement the Climate Analysis Toolkit the HUC-8 Based CMIP-5 Dataset was created in an effort to provide easy access to climate models for Hydrologic study. Despite the emergence of downscaled climate models, these data remain difficult to use at the primary spatial area of interest to hydrologists – the watershed. Currently available downscaled climate models are available at 15km grid cells across the United States and require processing to be re-scaled to the spatial area of a watershed. While these processing methods may be familiar to experts in the field of geoinformatics, practicing water resources engineers may be unable to invest the requisite time or resources. It is therefore relatively difficult to quickly investigate climate models at the spatial level of local watersheds of various size and shape.

To address this need, the Hydroinformatics Laboratory at Brigham Young University has developed and deployed the Climate Forecast Models for USGS HUC8 Watersheds. This effort required collaboration with the United States Geological Survey's Center for Integrated Data Analytics (CIDA). BYU worked closely with CIDA to process the existing downscaled CMIP-5 dataset to the spatial level of USGS HUC-8 watersheds through a process of weighted averaging. This process leveraged CIDA's Geo Data Portal (GDP). The GDP offers a web service that is accessible via Python or other scripting languages and is intended to streamline the time-consuming and resource-intensive tasks associated with data access and manipulation that can inhibit the sharing and use of interdisciplinary environmental science data. The collaborative work with CIDA through use of the Geo Data Portal resulted in the creation of a new dataset which includes 2,106 virtual observation "sites" (watershed centroids) each with 698 associated time series representing average monthly temperature and precipitation between 1950 and 2099 based on 234 unique climate model simulations.

In addition to the dataset, BYU created tools to store and host the newly created dataset. The resulting database was installed on a web server and made available online through a CUAHSI HydroServer and distributed using a WaterOneFlow web service. The dataset and a graphical user interface are published on the server of Brigham Young University in Provo, Utah as part of the World Water Project (http://worldwater.byu.edu). The web service was also registered on the CUAHSI Water Data Center (WDC) catalog (hiscentral.cuahsi.org) which makes the database widely available through all CUASHI products and services. The resulting database has been made available for wide distribution to hydrologists and water resources managers and educators via the WaterML format. The address of the web services end point is:

http://worldwater.byu.edu/climate/services/

In addition to the web service interface to the new climate database, it is also accessible through the CUAHSI network including the open source GIS-based software package HydroDesktop which allows users to access the datasets in time series and tabular form as well as to download the datasets locally. The WaterML R package can also be used to access the data for local analysis within the R statistical software system.

For quick access to climate model predictions for a specific HUC-8 watershed, the web-based HydroServer user interface on http://worldwater.byu.edu/climate/ui/ can also be used. Retrieving the data using this interface has two steps. First the user selects the HUC-8 mid-point on a Google map (Figure 2) following which the site details page can be used to select the climate model and time range, and to download the data in CSV format (Figure 3).

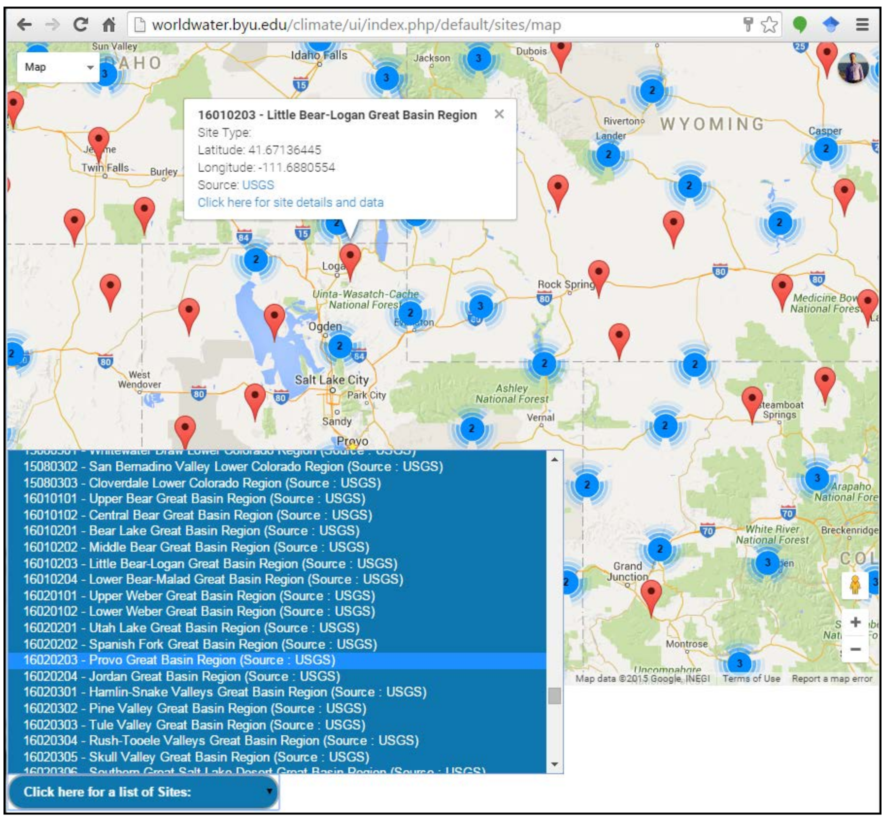


Figure 2 HydroServer Google Map User Interface for selecting the HUC-8 Watershed of Interest

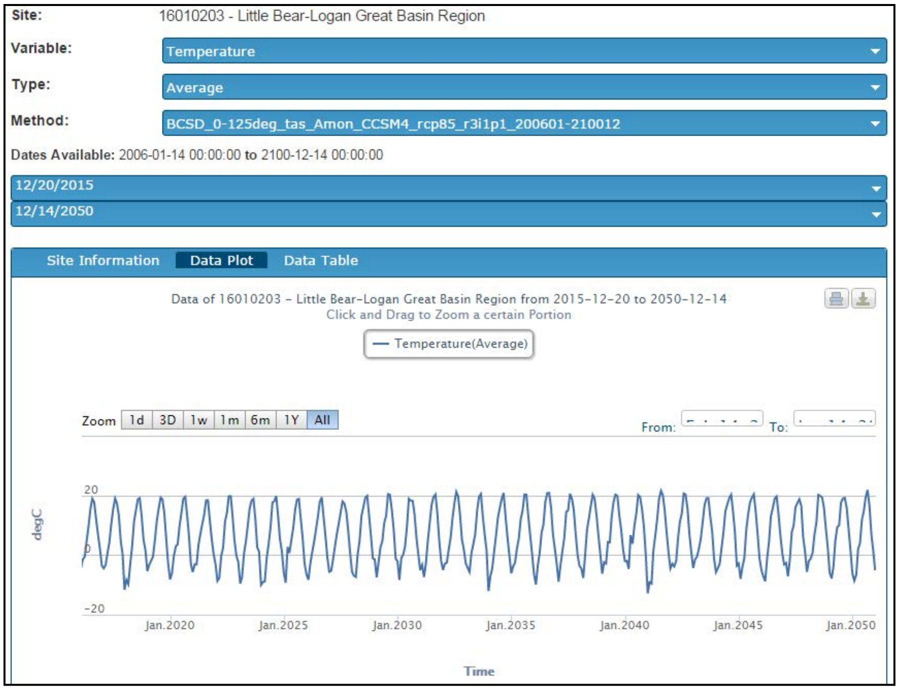


Figure 3 Site Details User Interface on HydroServer Lite Where Climate Data can be accessed

This effort has resulted in CMIP5 GCM data being scaled to the level of HUC-8 watersheds and made openly available for water resource engineers and managers. Through a variety of interfaces users can access this data from any computer with a web browser. Figure 4 represents a workflow diagram showing the progression of this work from GCMs to web services.

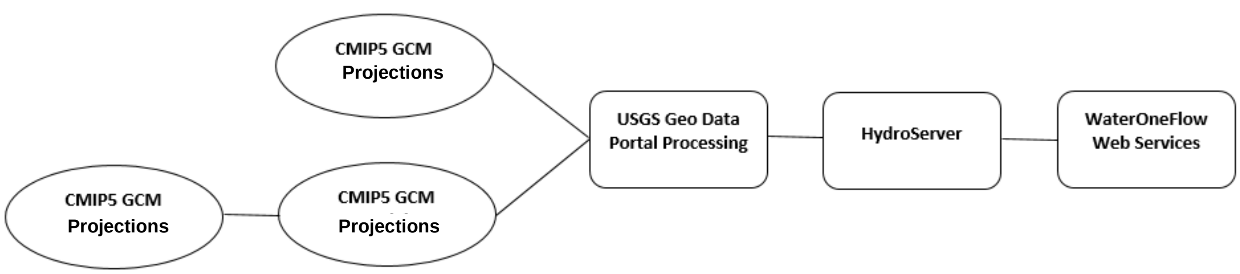


Figure 4 Work Flow Diagram of Process Translating CMIP5 Data to Openly Available HUC-8 Scaled Data

Promotion and Publication

Both the Climate Analysis Toolkit and the Climate Forecast Models for USGS HUC8 Watersheds database have been designed to be freely available through using open source software and by implementing data standards that are universally familiar across disciplines. A professional or amateur user with access to a computer and a web browser can easily download HUC-8 scaled climate models as well as the Climate Analysis Toolkit to analyze them.

In addition to promoting these tools through their design, several forums have been used as part of an effort to publicize these new climate analysis tools and encourage their use throughout the hydrologic science and water resource engineering communities. Currently, an information science paper is under review for publication in the Journal of the American Water Resources Association as part of a special collection for the Open Water Data Initiative (OWDI). This paper is co-authored by individuals at BYU and the Bureau of Reclamation involved in this work and outlines the need for, and development of, the HUC-8 Based CMIP-5 Database. Several use cases are presented therein, as well as information regarding where these tools can be accessed. It is intended that this published paper will make these tools known throughout the water resources and scientific community.

In November, 2015 the American Water Resources Association held their annual conference in Denver, Colorado. This conference brings professionals throughout the water resource community together to learn and share ideas about trends in water resource management and hydrologic research. The 2015 conference focused specifically on climate change and tools to help prepare for its effects. Two presentations were given at this conference by Dustin Woodbury on the climate analysis tools developed as part of this project. These discussions presented the Climate Analysis Toolkit and well as the Climate Forecast Models for USGS

HUC8 Watersheds database to leaders in government agencies and industry who will hopefully implement them more broadly.

To support these products, information has been published on GitHub regarding both the toolkits and the database. Here users can find instructions regarding their use, background information, the source code, and links to download the products. These sites for the Climate Analysis Toolkit and the Climate Forecast Models for USGS HUC8 Watersheds database can be accessed at https://github.com/usbr/ClimateAnalysisToolkit and https://github.com/CMIP5/HUC8Climate respectively.

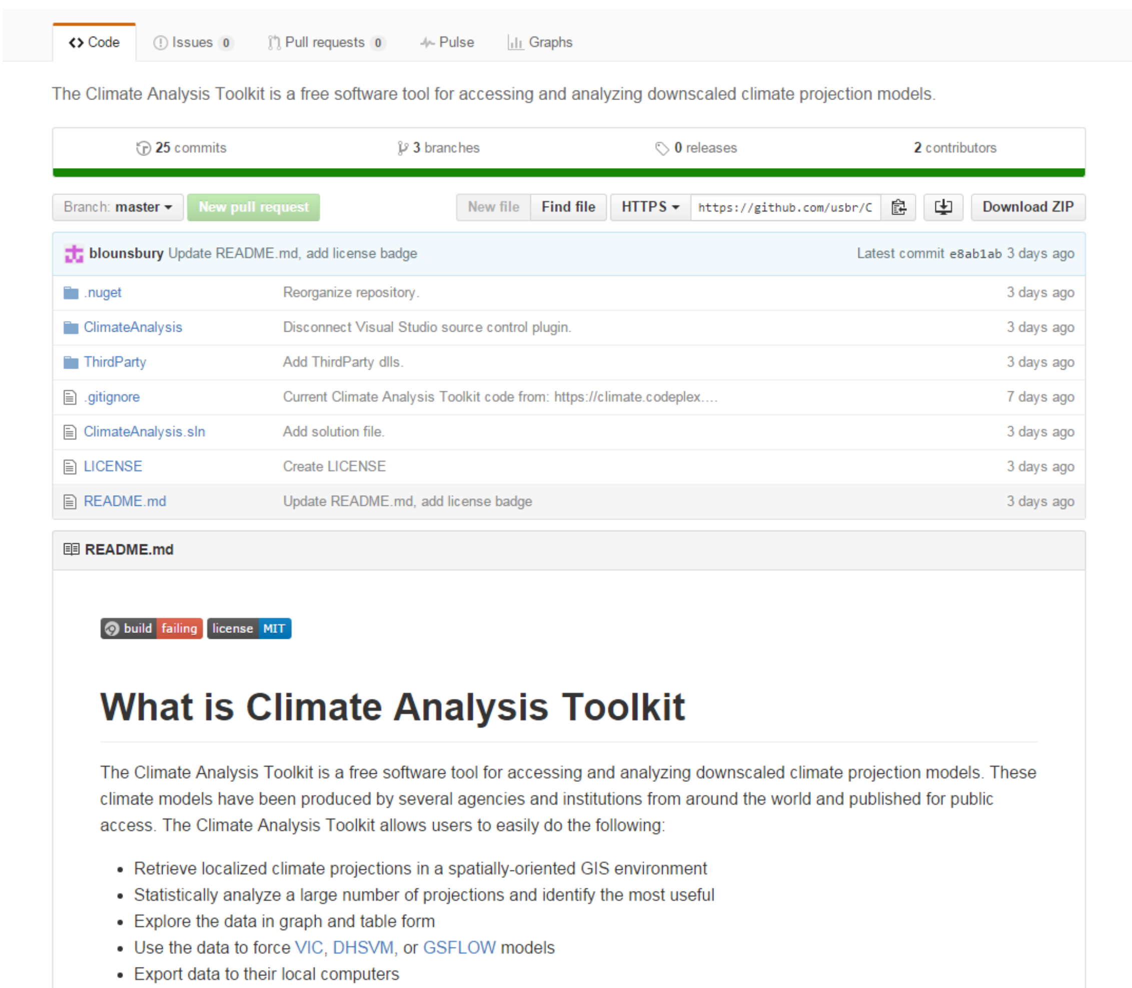


Figure 5 GitHub Site for Climate Analysis Toolkit

Summary

The results of this project are a first-of-its-kind toolkit for statistical analysis of climate projections with watershed model forcing capability, a first-of-its-kind database of HUC-8 watershed scaled climate models (including a host of tools for online access), and promotion and publication of these products through the hydrologic science and water resource engineering community. Hydrologists now have easy access to climate modeling data at the level of the watershed and a toolkit for their analysis and translation to streamflow modeling. Great care has been taken to make these tools openly available to the public and to ensure their design is as user friendly as possible. In developing and deploying these tools in this way, it is our hope that the most valuable result of this work will be the further and enhanced implementation of climate studies in water resource planning throughout the Bureau of Reclamation and the greater water resources community.

Appendix

We submitted the following Journal Article to the Journal of the American Water Resources Association special issue on the Open Water Data Initiative. The paper has been reviewed, revised, and resubmitted and is currently under re-review.

The citation is:

Dustin H. Woodbury, Daniel P. Ames, Jiří Kadlec, Stephen Duncan, Greg Gault. “A New Open-Access HUC-8 Based Downscaled CMIP-5 Climate Model Forecast Dataset for the Conterminous United States.” *Journal of the American Water Resources Association*. Revision submitted December 2015.

A New Open-Access HUC-8 Based Downscaled CMIP-5 Climate Model Forecast Dataset for the Conterminous United States

[1]Dustin H. Woodbury, [2]Daniel P. Ames, [1]Jiří Kadlec, [1]Stephen Duncan, [3]Greg Gault

[1]Graduate Assistant, Department of Civil and Environmental Engineering, Brigham Young University, Provo, Utah 84602

[2]Professor, Department of Civil and Environmental Engineering, Brigham Young University, 242 J Clyde Building Provo, Utah 84602, dan.ames@byu.edu

[3] BORGIS System Manager | Architect – United States Bureau of Reclamation, Boise, Idaho, 83706

Abstract

Watershed-scale hydrologic simulation models generally require climate data inputs including precipitation and temperature. These climate inputs can be supplied by downscaled global climate simulations which has introduced the potential to project runoff at the level of local watersheds. While a simulation designed to drive a watershed model would ideally be constructed at a relatively small scale, global climate simulations are, by definition, large-scale and coarsely gridded in arbitrarily determined rectangular spatial grid patterns. This paper addresses the technical challenge of making climate simulation model results readily available in the form of downscaled data sets that can be used for watershed scale models. Specifically, we present the development and deployment of a new Coupled Model Intercomparison Project phase 5 (CMIP5) based database which has been prepared through a scaling and weighted averaging process for use at the level of USGS HUC-8 watersheds (roughly 1,800 square-km). The resulting dataset includes 2,106 virtual observation "sites" (watershed centroids) each with 698 associated time series data sets representing average monthly temperature and precipitation between 1950 and 2099 based on 234 unique climate model simulations. The new database is deployed on a Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) HydroServer and distributed using WaterOneFlow web services in the WaterML format. Two example use cases for the data set are also presented.

Key Terms: Climate model, CMIP5, GCM, Downscaling, HUC-8 Watersheds, database, HydroServer

Introduction

Downscaled climate change models have the potential to inform regional-scale watershed runoff simulations of future scenarios for the purposes of improved short and long-term water resources management (Fonseca et al. 2014; LaFontaine et al. 2015; Pradhanang et al. 2013; Stöckle et al. 2014). A number of General Circulation Models (GCMs) have been produced and proven useful in predicting trends in future climate conditions and are of increasing interest to hydrologists and water resource managers for planning purposes (Flato 2013; Knutti and Sedláček 2013; Taylor et al. 2012). Downscaling the spatial resolution of GCMs allows hydrologists to visualize climate behavior at a more relevant spatial resolution leading to better projections of streamflow and water resource availability (Blodgett et al. 2013; Hwang and Graham 2014; Wooten et al. 2014).

Given the growing interest in climate change research, the development and testing of GCMs continues to be a major focus of the climate modeling community (Ford et al. 2011). GCMs are capable of providing projections of ground level air temperatures and precipitation values at various locations given specified atmospheric conditions. Using a sophisticated application of the Navier-Stokes equation, climate modelers are able to predict vector paths of hot air masses around the globe (Zhang and Moore 2014). By estimating atmospheric compositions, and the corresponding thermal irradiance captured by the atmosphere, GCMs predict air temperature and precipitation conditions for long-term forecasting over the next century. GCMs vary based on the preferred modeling approach of each parent research institution and how atmospheric composition is estimated. These variations in the GCMs themselves result in a wide variety of predicted future scenarios for any specific region of interest.

In 1995 the Coupled Model Intercomparison Project (CMIP) was established under the World Climate Research Programme (WRCP) as a standard experimental protocol for atmosphere-ocean general circulation models (Covey et al. 1996). This collection includes atmosphere-ocean general circulation models from most major research institutions in the international climate modeling community. CMIP has collected model outputs from the pre-industrial climate simulations ("control runs") and 1% per year increasing-CO2 simulations (Andrews et al. 2012). The latest collection is phase 5 (CMIP5) and includes 234 different projections for historical and future climate data at a spatial resolution of between 1/2° and 4° grid cells across the globe. Daily average air temperature and precipitation values for each month are available for these cells from 1950 through 2099 (Taylor et al. 2012).

To increase the spatial resolution of climate projections and provide more localized data, a collaborative initiative led by the Lawrence Livermore National Laboratory (LLNL) was completed in 2013 that resulted in the downscaling of general circulation models using a method known as bias-correction spatial disaggregation (BCSD). BCSD is a process for statistically downscaling spatial data that relies on mapping GCM values by quantile onto historically observed data wherein each quantile value has received its own adjustment (Girvetz, 2013). This method, as well as the downscaling work by LLNL, has been published and shown to be consistent with historical observed data (Cavazos and Arriaga-Ramírez 2012; Hayhoe 2006; Werner 2011). This previously performed work by LLNL resulted in a collection of climate projection data at a spatial resolution of 1/8° (~12 kilometers on a side) grid cells over the contiguous United States and parts of Mexico and Canada (Wood et al., 2002 and 2004; Reclamation, 2013). This downscaling increased the resolution from approximately 190 grid cells in the original CMIP3 and CMIP5 data set to approximately 53,220 grid cells. This downscaled collection offers all 234 GCM projections over virtually any surface of the contiguous United States at a much finer resolution than the original CMIP5 collection and brings climate data one step closer to practical application for hydrologists. A table listing the CMIP5 models used in the LLNL downscaling work as well as their parent institutions IDs can be found at http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5_modeling_groups.pdf.

Despite the emergence of downscaled BCSD circulation models, these data remain difficult to use at the primary spatial area of interest to hydrologists – the watershed. BCSD data are mapped to 1/8° grid cells across the United States and require processing to be re-scaled to the spatial area of a watershed. While these processing methods may be familiar to experts in the field of geoinformatics, practicing water resources engineers may be unable to invest the requisite time or resources. It is therefore relatively difficult to quickly investigate BCSD models at the spatial level of local watersheds of varying size and shape. While downscaled GCM data is available via the LLNL collection, the water resources engineering community has a need for a readily available database of GCM's pre-processed to the size and shape of commonly-used watershed study areas.

This paper describes the development and deployment of a new open-access database of downscaled climate model forecasts. This database has been prepared through a downscaling and weighted averaging process for use at the level of USGS HUC-8 watersheds (between 500 and 2500 square-km). The resulting dataset includes 2,106 virtual observation sites represented by watershed mid-points that contain monthly average temperature and precipitation between 1950 and 2099 based on 234 unique climate model simulations.

Methods

Defining Practical Watersheds

The USGS standard Hydrologic Unit Code (HUC) level 8 watersheds were used as a representative modeling unit for this study. Smaller watersheds (e.g. level 12 watersheds) could be processed in a similar manner as presented here. Although the size of HUC-8 watersheds are somewhat variable, HUC-8 watersheds have an average surface area of 1,800 square kilometers. The HUC system is widely used and universally recognized within the hydrologic community. Figure 1 depicts all HUC-8 watersheds within the contiguous United States and represents all of the watersheds represented in the new dataset.

Figure 1 USGS HUC-8 Watersheds Representative of Spatial Regions Used in Data Processing

Rescaling using the USGS Geo Data Portal

To rescale the BCSD data to the spatial level of HUC-8 watersheds we employed the USGS Geo Data Portal (GDP) developed by the Center for Integrated Data Analytics (CIDA). The GDP offers a web service that is accessible via Python scripting language and is intended to "streamline the time-consuming and resource-intensive tasks associated with data access and manipulation that can inhibit the sharing and use of interdisciplinary environmental science data." (Blodgett 2012). This process allows for the upload of HUC-8 watershed polygons which can then be matched with gridded climate data and processed to produce weighted average precipitation and temperature projections.

The GDP returns a weighted mean for the downscaled GCM across a given spatial area by using weighted grid statistical processing. In this case the given spatial areas are HUC-8 watersheds. Figure 2 demonstrates the weighted-averaging process. The dashed lines represent the downscaled GCM grid boundaries, the polygon represents a watershed, and the dots in the middle of each grid cell represent their respective centroids.

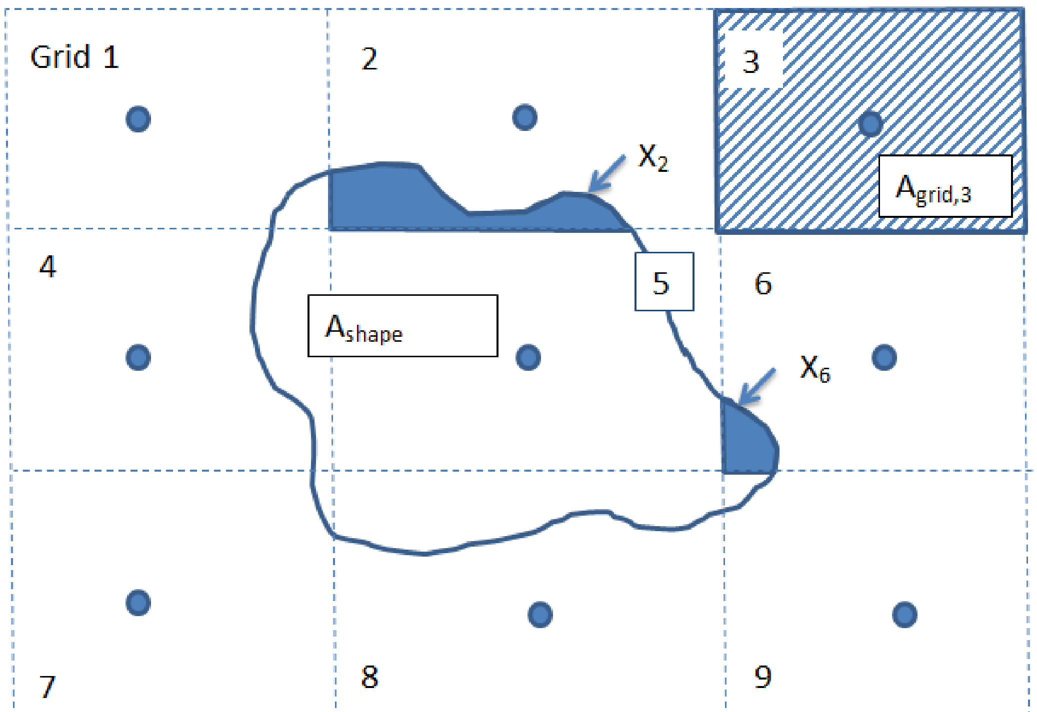


Figure 1 Example of Weighted Grid Averaging Used in Scaling of Climate Data to HUC-8 Polygons (adapted from http://cida.usgs.gov/enddat/UserGuide.jsp)

Equation 1 is used to calculate the weighted mean (μ). Each grid box has an area represented by $A_{grid,i}$ and a reported value (representing monthly average temperature or precipitation) represented as P_i. The area of the watershed polygon that lies within each grid is represented by X_i and the total area of the watershed polygon is represented by A_{shape}.

Equation 1 Weighted Mean Calculation

$$\mu = \frac{\sum_{i=1}^{9} P_i \frac{X_i}{A_{grid,i}}}{\sum_{i=1}^{9} \frac{X_i}{A_{grid,i}}}$$

A polygon shapefile of the contiguous United States was used as the processing bounds for the GDP. This polygon included all HUC-8 watersheds as polygon features. The HUC-8 watershed boundaries were simplified by removing large numbers of excessive vertices that do not affect the polygon area or overall shape. Because the GDP must calculate distance between each set of vertices along a watershed boundary, the natural watershed boundary was found to require an unacceptable amount of processing time. To reduce this processing time while maintaining the spatial integrity of the watershed boundary, the number of vertices of each HUC-8 polygon was reduced in the shapefile. This process used the Ramer-Douglas-Peuker algorithm which removes vertices from a geometry. For the purpose of intersection with grid cells, the simplification of the HUC-8 polygons resulted in a negligible difference in the results. Figure 3 depicts a natural HUC-8 watershed boundary as well as the same boundary that has simplified vertices. Further information regarding the Geo Data Portal can be found at http://pubs.usgs.gov/of/2011/1157/pdf/ofr2011-1157.pdf.

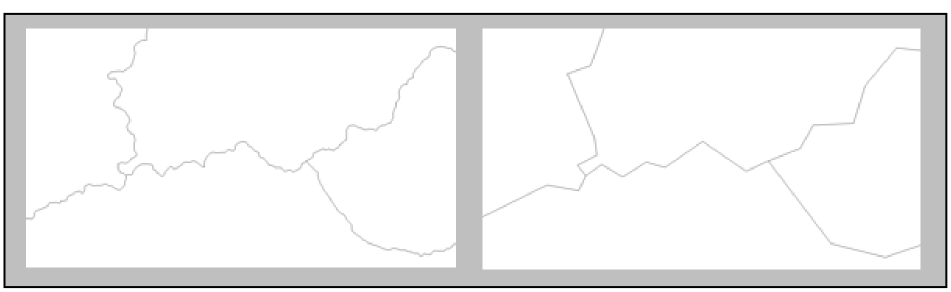


Figure 2 Natural (Left) and Simplified (Right) Watershed Boundary of a HUC-8 Polygon

Processing

The GDP web service is accessible via a Python package published by the USGS called pyGDP. The package is openly available at the following internet address: https://github.com/USGS-CIDA/pyGDP.

To use the pyGDP package, a simple Python script was constructed which imports the package, declares needed variables, and calls the "submitFeatureWeightedGridStatistics" method. This function call is central to the operation of the GDP and is provided here for reference purposes:

```
def submitFeatureWeightedGridStatistics(self, geoType,
dataSetURI, varID, startTime, endTime, attribute='the_geom',
value=None, gmlIDs=None, verbose=None, coverage=True,
delim='COMMA', stat='MEAN', grpby='STATISTIC', timeStep=False,
summAttr=False, weighted=True)
```

The arguments which the method accepts are defined in Table 1. The method returns a delimited text file that contains the requested processed data.

Table 1 Inputs of submitFeatureWeightedGridStatistics Method

Argument	Title	Explanation
geoType	WFS Feature Collection	A feature collection encoded as a WFS request or one of the supported GML profiles. This would be the simplified HUC-8 shapefile containing all watersheds as feature polygons.
datSetURI	Dataset URI	The base data web service URI for the dataset of interest.
varID	Dataset Identifier	The unique identifier for the data type or variable of interest. This would be the downscaled model time-series of interest.
startTime	Time Start	The date to begin analysis. All available dates were collected.
endTime	Time End	The date to end analysis. All available dates were collected.
Attribute	Feature Attribute Name	The attribute that will be used to label column headers in processing output.
Delim	Delimiter	The delimiter that will be used to separate columns in the processing output (i.e. a comma).
Stat	Statistics	Statistics that will be returned for each feature in the processing output, such as “MEAN” for the average value of the weighted grid statistic calculation.
Grpby	Group By	If multiple features and statistics are selected, this will change whether the processing output columns are sorted according to statistics or feature attributes..
timeStep	Summarize Time Step	If selected, processing output will include columns with summarized statistics for all feature attributes for each time step.
summAttr	Summarize Feature Attribute	If selected, processing output will include a final row of statistics summarizing all time steps for each feature attribute value.
weighted	Weighting	If selected, processing output will include weighting of grids who partially intersect with watershed polygon. This is desired in processing downscaled models.

The GDP divides the 234 BCSD models into different time-series for time period and variables. The historical time-series spans the years 1950 through 2006 while the future time-series spans the years 2006 through 2099. Each BCSD model contains a time-series for historical precipitation, historical temperature, future temperature, and future precipitation. A historical time-series is considered a run for a particular model while a future time-series for the same model may contain multiple runs. As a result, many models contain multiple future time-series and only one historical time-series. This results in 698 time-series needed to collect a complete database for all CMIP5 models. The 698 time series are represented as follows:

- Historical Time Series (1950-2006)
    - Precipitation: **115**
    - Temperature: **115**
- Future Time Series (2006-2099)
    - Precipitation: **234**
    - Temperature: **234**
- Total Time Series: **698**

While a batch of polygons (HUC-8 watersheds) may be processed within a single request to the GDP without additional processing time, each time-series must be processed individually. Requesting more than one time series per request is beyond the limits of the web service. Therefore, each request processes all 2,106 HUC-8 watersheds for a single time series.

Requests to the GDP take up to 15 minutes to process and return a time-series. Therefore, the processing of the entire collection demands a significant amount of time. To manage this process, a program was written in the C# programming language to dispatch requests to the GDP as well as to retrieve and save the files returned. This program uses IronPython, an open-source implementation of the Python language and is integrated into the Microsoft .NET Framework (Foord, 2009). This allows for the creation of variables and file paths in C# that can be used directly in a Python script called from C#. Additionally, variable and file paths created in a Python script can be passed back to be used in the original C# script. Thus the Python scripting to communicate to the GDP can be managed within the overlying C# program. This code as well as others used in this work can be found at the following repository: https://github.com/CMIP5/HUC8Climate.

This program initially reads a comma separated value file (CSV) containing the names of all time-series which are then stored within an array. These data types will later be passed into the Python script and included in each request to the GDP. The program then calls a GDP request using the first data type in the array and waits for the GDP request to finish before continuing. Once the request has finished, an internet address where the time-series can be downloaded is returned. The program uses the address to retrieve the resulting file and stores it as a comma separated value file in the specified location. The process is then reiterated until all 698 time series are processed.

Database Implementation

To reduce the processing time required to retrieve the BCSD climate model time series for any HUC-8 watershed and any time period, we stored the processed HUC-8 level time series data in an online database compatible with the Hydrologic Information System (HIS) of the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI). This particular database presented unique challenges in how it would be organized and hosted due to its large size. The final database contained approximately 1.5 billion data values as well as the associated metadata. A traditional CUAHSI Observations Data Model (ODM) database was originally created to host this data and required over 500 GB of disk space. This was concerning as it would not only require a significant amount of storage to host the data, but the large size of the database led to performance issues in trying to retrieve time series.

To address this issue a non-traditional file-based approach was tested and ultimately implemented successfully. This approach leverages a table-based structure to store all metadata and in so doing retains the ease of access and query that ODM offers. The difference is that the data values are not stored within a table, but are instead stored as comma separated value (CSV) files in a special folder on the same server as the metadata. The series catalog table contains links to the associated CSV file. Each CSV file contains a time series associated with one HUC-8 watershed, one variable, and one climate model run. This results in 1.5 million files. Compared with the previous 1.5 billion values that were assigned using the traditional approach, the disk storage size was reduced to 40 GB or by a factor of 10. The performance was also improved by using this new approach. The naming of the CSV files follows a predefined convention:

[HUC8 code]-[Method]_[Start year][Start month]-[End year][End month].csv

For example in the file name:

01010001-BCSD_0-125deg_pr_Amon_HadGEM2-CC_historical_r1i1p1_195001-200511.csv,

01010001 is the HUC-8 code number, 0-125deg_pr_Amon_HadGEM2-CC_historical_r1i1p1 is the method, 1950 is the start year, 01 is the start month, 2005 is the end year, and 11 is the end month. The method identifier contains the downscaling method name, variable, climate model name, and model run. The methods table in the database contains the method identifiers that are used to locate the corresponding CSV file.

The resulting database was installed on a web server and made available online through a CUAHSI HydroServer. Specifically we adapted the HydroServer Lite software using a MySQL database engine and Linux operating system. HydroServer Lite is open source software written in the PHP programming language and is available for download from http://hydroserverlite.codeplex.com (Conner et al. 2013; Kadlec and Ames 2012). The design of HydroServer Lite uses a Model View Controller (MVC) architecture pattern to separate the persistent storage (model), user interface design (view), and application logic (controller). Enabling the new CSV file data storage in HydroServer Lite required modifying the Method and DataValue classes in the model. The HydroServer with a WaterOneFlow web service (Kadlec et al. 2015; OGC 2012) and a graphical user interface was published on the server of Brigham Young University in Provo, Utah as part of the World Water Project (http://worldwater.byu.edu). The web service was also registered on the CUAHSI Water Data Center (WDC) catalog (hiscentral.cuahsi.org) which makes the database widely available through all CUASHI products and services. The resulting database has been made available for wide distribution to hydrologists and water resources managers and educators via the WaterML format and standard CUAHSI data clients such as HydroDesktop (Ames et al. 2012) and the WaterML R package (Kadlec et al. 2015).

The address of the web services end point is:

http://worldwater.byu.edu/climate/services/ and has the following methods to query the climate data (Table 2):

Table 2 Climate Services Feature Table

Method	Parameter	Description
GetSites	Geographic bounding rectangle (optional)	Show coordinates and details about all HUC-8 watersheds
GetVariables	(none)	Show metadata about the variables (units, time units)
GetSiteInfo	HUC-8 code	Show what projections and time range of data is available for a given HUC
GetValues	Variable code (tas or pr), HUC-8 code, Start date (optional), End data (optional), Model code (optional)	Get the climate projection values for a given HUC, variable, time range and model

In addition to the web service interface to our new climate database, it is accessible through the CUAHSI network including the open source GIS-based software package HydroDesktop (Ames et al. 2012) which allows users to access the datasets in time series and tabular form as well as to download the datasets locally. The WaterML R package can also be used to access the data for local analysis within the R statistical software system (Kadlec et al. 2015). For quick access to climate model predictions for a specific HUC-8 watershed, the web-based HydroServer user interface on http://worldwater.byu.edu/climate/ui/ can also be used. Retrieving the data using this interface has two steps. First the user selects the HUC-8 mid-point on a Google map (Figure 4), following which the site details page can be used to select the climate model and time range, and to download the data in CSV format (Figure 5).

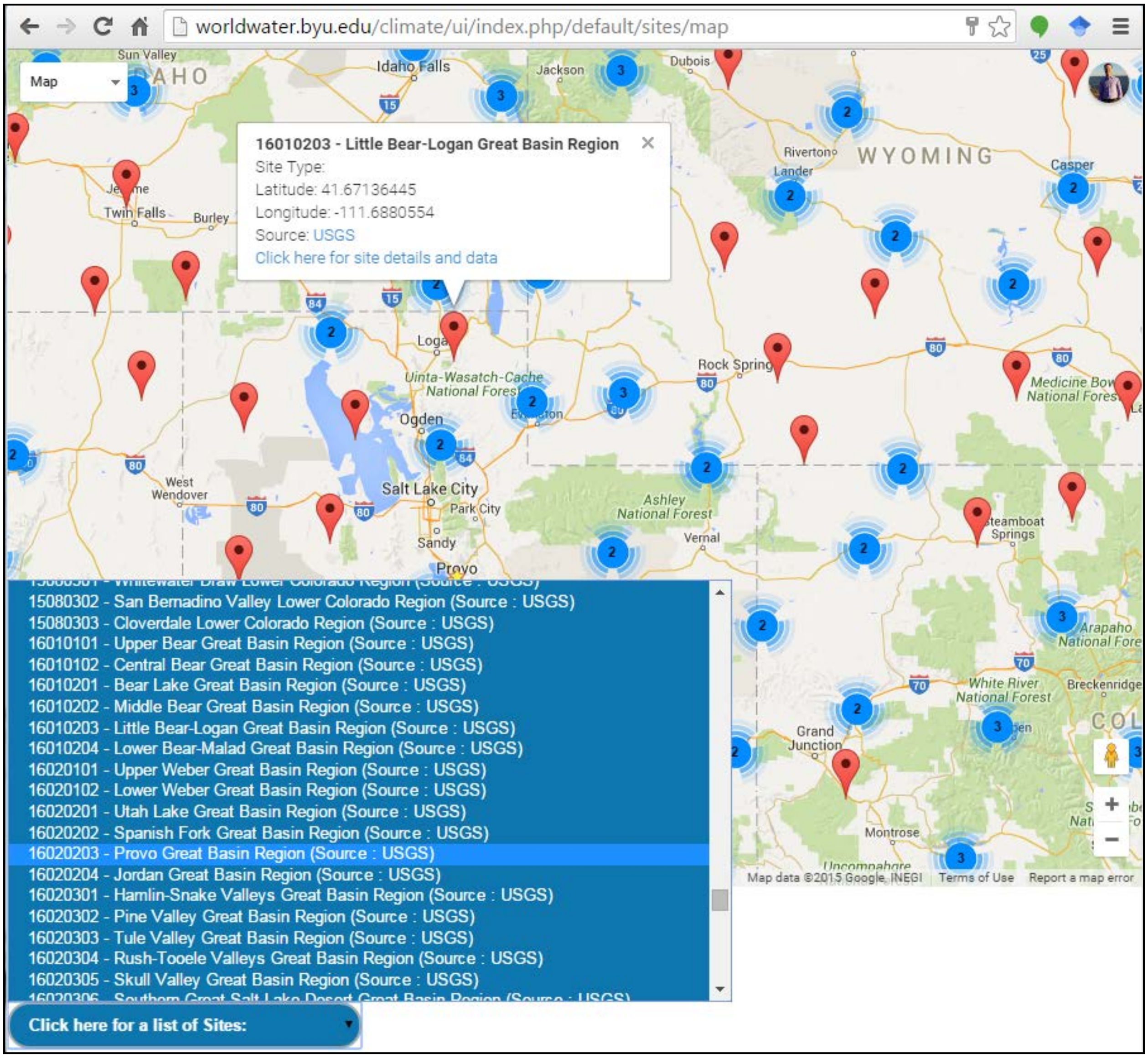


Figure 4 HydroServer Google Map User Interface for selecting the HUC-8 Watershed of Ineterst

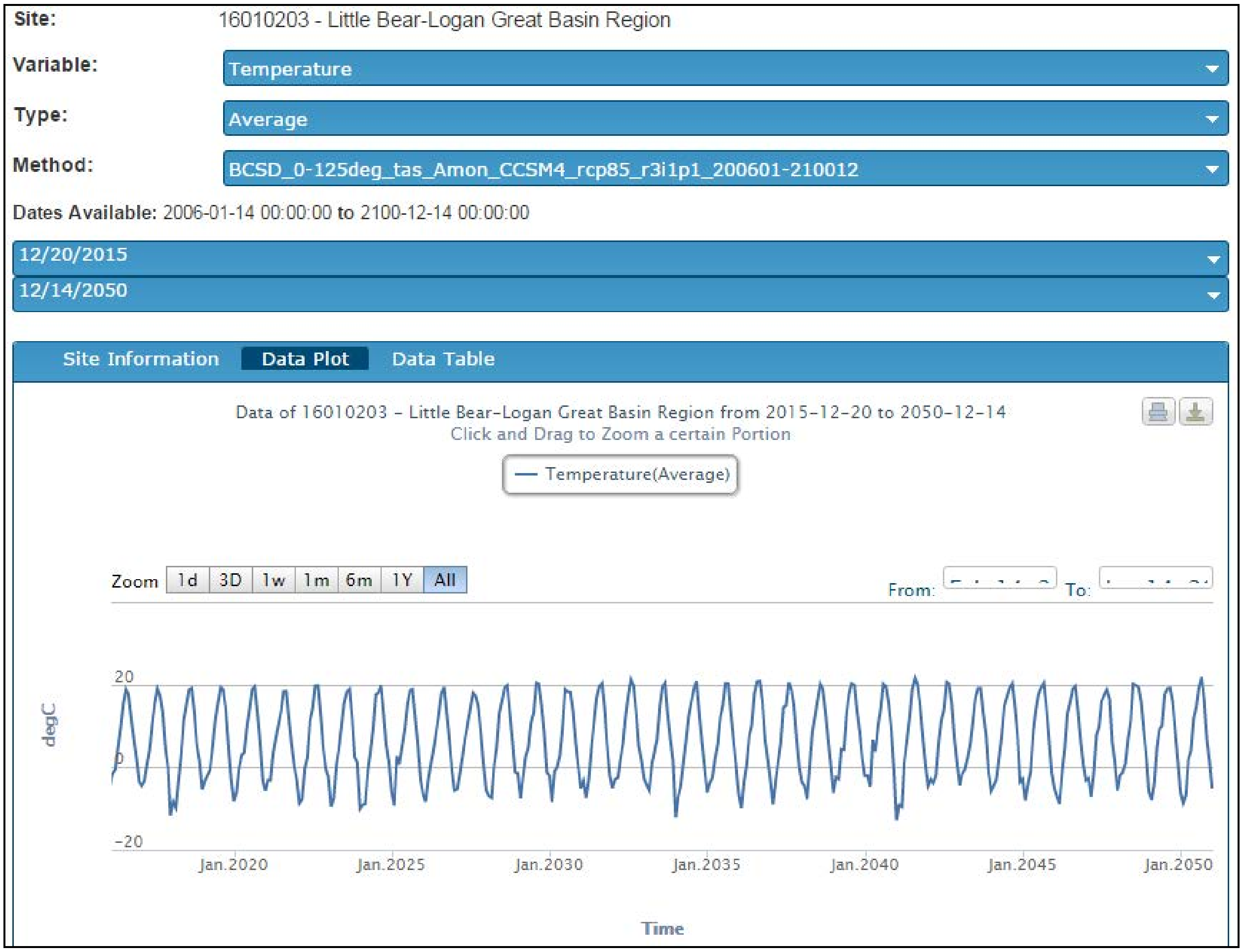


Figure 5 Site Details User Interface on HydroServer Lite Where Climate Data can be accessed

Results

Discussion

This effort has resulted in CMIP5 GCM data being scaled to the level of HUC-8 watersheds and made openly available for water resource engineers and managers. Through a variety of interfaces users can access this data from any computer with a web browser. Figure 6 represents a workflow diagram showing the progression of this work from GCMs to web services.

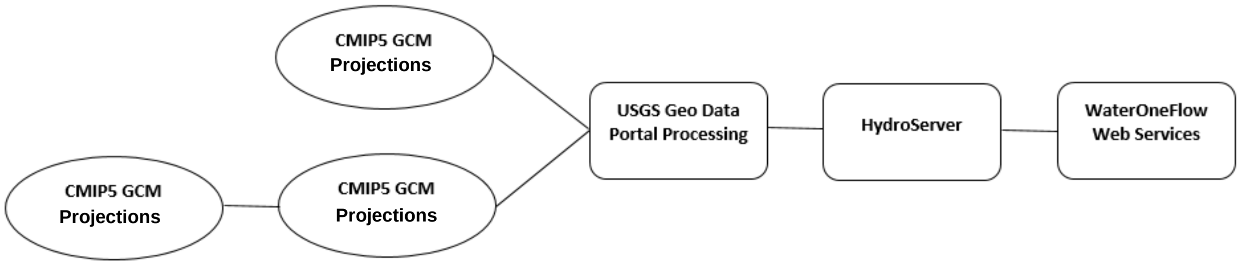


Figure 6 Work Flow Diagram of Process Translating CMIP5 Data to Openly Available HUC-8 Scaled Data

Use Case 1: Examining Climate Models

The database has been designed for use with the Climate Analysis Toolkit (an extension to HydroDesktop) which allows for the statistical analysis of downscaled CMIP3 and CMIP5 datasets via hybrid-delta ensemble analysis (Hamlet et al. 2010) and for the processing of meteorological inputs for macro-scaled hydrologic models. With the addition of the HUC-8 downscaled climate model database, users can use the tool to compare model predictions for any HUC-8 watershed.

Given that the CMIP-5 collection contains 234 models, it is often useful to compare the anticipated changes in temperature and precipitation that each model provides. Using the Climate Analysis Toolkit we have compared all models for a particular HUC-8 watershed; the data for which was downloaded from the database. Figure 7 demonstrates a scatterplot of all 234 climate models.

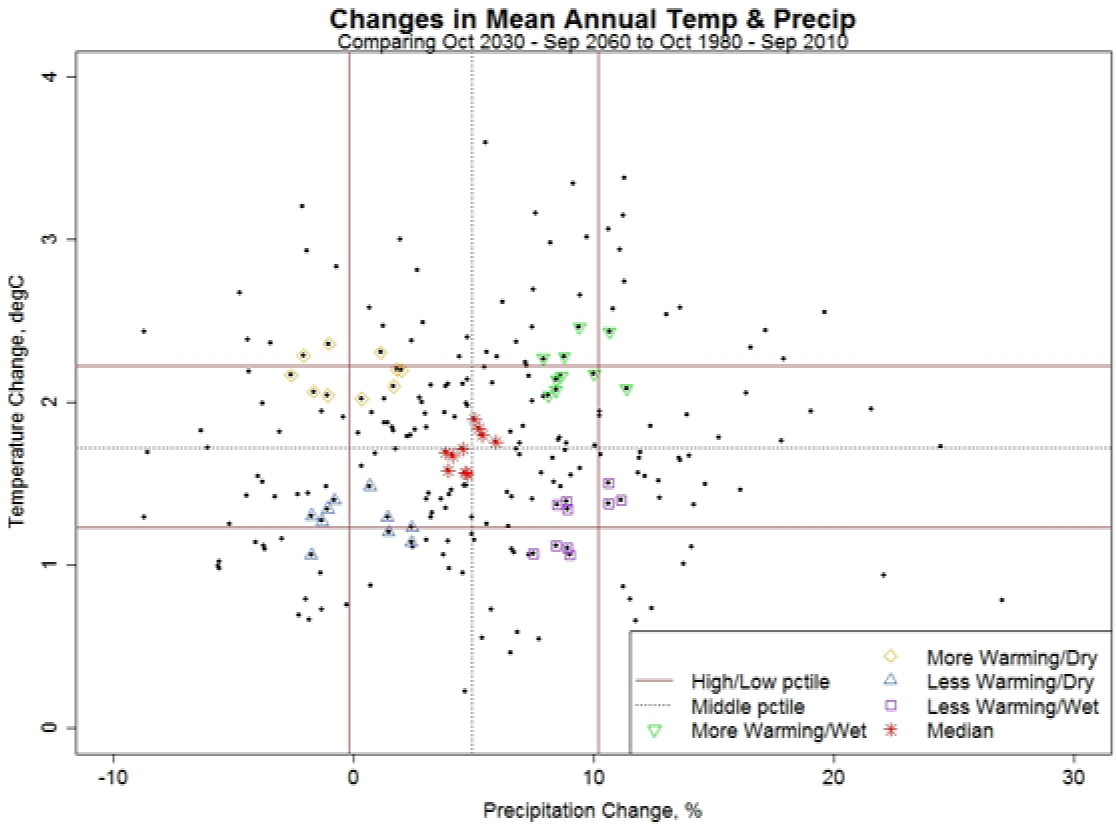


Figure 7 Scatter Plot of All 234 CMIP5 Models for One Watershed Produced in HydroDesktop's Climate Analysis Toolkit

These figures represent the temperature change and percent precipitation change when comparing two thirty-year spans across the projected dataset (1980-2010 and 2030-2060). The changes are then plotted on the above graph and grouped into ensembles that represent scenarios. The scenarios are determined by quantile and represent the scenarios shown in Table 3.

Table 3 Common Climate Change Scenarios

Climate Change Scenario	Percentile in terms of temperature change	Percentile in terms of precipitation change
More Warming – Drier	80	20
More Warming – Wetter	80	80
Median	50	50
Less Warming – Drier	20	20
Less Warming – Wetter	20	80

These scenarios represent a range of potential changes in climate for a particular HUC-8 watershed and can be helpful in water resource planning and study.

Use Case 2: Comparing Watersheds

The database can also be used to quickly compare a single model's forecast across watersheds in the United States. To demonstrate this potential use case, we downloaded one model for temperature projections across all HUC-8 watersheds in the United States from the newly created database. This data was downloaded in a comma separated value format which allowed us to display it in Microsoft Excel. Within Excel average temperatures for each watershed were then calculated for a 30 year period at the beginning of the time series (2010-2040) and the end of the time series (2069-2099). The change in average temperature between these two time periods was then calculated for each watershed. Within ESRI ArcMap a polygon shapefile of simplified HUC-8s was added as a layer. By accessing the attribute table we were able to add an average temperature attribute and then add the data from Excel matching each HUC-8 code in the attribute table to its corresponding calculated change in average temperature. Finally, the polygon's properties were adjusted to assign a color to each watershed feature based on the intensity of the average temperature change. The result of this exercise is shown in Figure 8 as a simple HUC-8 based view of temperature as a heat map. This map represents average temperature change when comparing two periods in time and graphically illustrates temperature change according to the model "BCSD_0-125deg_tas_Amon_ACCESS1-0_rcp45_r1i1p1_200601-210012".

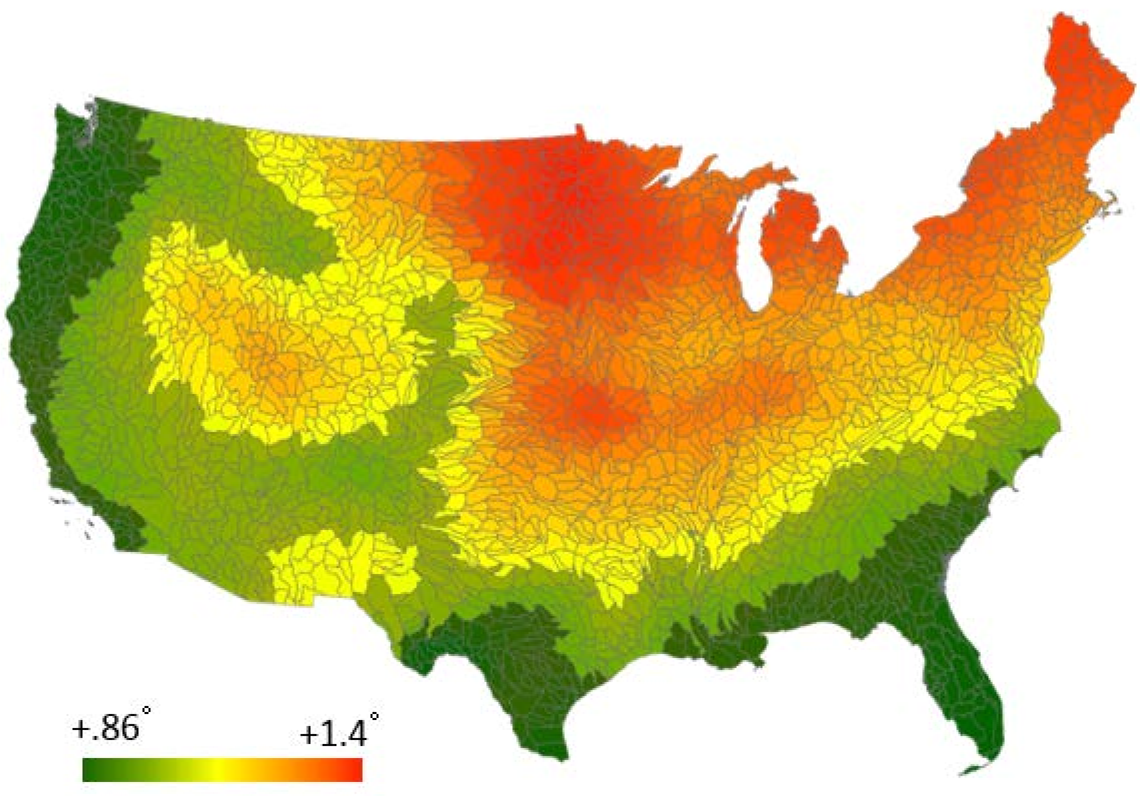


Figure 8 Temperature Change of HUC-8 Watershed Regions as Described by a Model From Database

Planned implementation

Section 9503 of the SECURE Water Act (SECURE) authorizes the United States Bureau of Reclamation (Reclamation) to assess the risks to water supplies and demands posed by climate change within major river basins in the Western United States. Reclamation and other use macroscale hydrologic models to predict future trends of water availability and climate behavior. The meteorological inputs (forcing files) to the hydrologic models are adjusted by change factors generated using hybrid delta ensembles of downscaled GCM projections. Using the newly created Climate Forecast Models for USGS HUC8 Watersheds database, Reclamation hydrologists will be able to input downscaled GCM projections and successfully generate watershed model forcing files to use as inputs in future studies.

Future Development

This database is limited to HUC-8 watersheds within the contiguous United States; however, the methods employed can be applied to other areas (e.g. outside the conterminous United States) as rescaling services for these areas become available. Further development of downscaled GCM gridded data would make this a possibility. Also, if individuals need to develop downscaled GCM results for specific drainage areas smaller than a HUC-8, they can access the GDP directly. The methods presented in this paper can be adapted to such a need.

Other future work may include using the GDP to scale GCM projections to an even finer spatial resolution. In keeping congruent with the use of USGS HUC watersheds, a future database could be processed through the same algorithm published herein but with HUC-12 or even HUC-16 watershed polygons. While this would employ a greater number of polygons and therefore take more processing time, it would nonetheless be feasible.

Conclusions

By pre-processing climate models through the GDP a significant amount of effort has been completed for users. Indeed, the processing time for all 234 climate models did in fact take 116 hours. By making these pre-processed model datasets available via CUAHSI HIS, hydrologists, engineers and other users can now access this information in a few seconds from any web browser, resulting in significant savings of both time and money.

The HUC-8 Based Downscaled CMIP-5 Climate Model Forecast Dataset is the first to offer downscaled global circulation models at the scale of nationally predefined watersheds. Through this database water resource managers and engineers have access to hundreds of climate projections for each HUC-8 basin through the year 2099, allowing for the study of, and preparation for, climate change.

Acknowledgments

This project was primarily funded by the United States Bureau of Reclamation under Cooperative Agreement, R13AC10030. We acknowledge Guy Werner and the BYU Hydroinformatics laboratory students for their insights in developing scripts for this project. We also acknowledge the contributions of David Blodgett in using and understanding the Geo Data Portal.

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's

Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Literature Cited

Ames, D. P., J. S. Horsburgh, Y. Cao, J. Kadlec, T. Whiteaker & D. Valentine, 2012. HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis. Environmental Modelling & Software 37:146-156.

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. 2012. Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, *39*(9).

Blodgett, D., 2012. Description of the US Geological Survey Geo Data Portal Data Integration Framework. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of:1687=1691.

Blodgett, D., J. Walker, I. Suftin, M. Warren & T. Kunicki, Tools for Interdisciplinary Data Assimilation and Sharing in Support of Hydrologic Science. In: AGU Fall Meeting Abstracts, 2013. vol 1. p 1203.

Cavazos, T. & S. Arriaga-Ramírez, 2012. Downscaled climate change scenarios for baja california and the north american monsoon during the twenty-first century. Journal of Climate 25(17):5904-5915.

Conner, L. G., D. P. Ames & R. A. Gill, 2013. HydroServer Lite as an open source solution for archiving and sharing environmental data for independent university labs. Ecological Informatics 18(0):171-177 doi:http://dx.doi.org/10.1016/j.ecoinf.2013.08.006.

Covey, C., B. D. Santer & E. Cohen-Solal, 1996. CMIP: a study of climate models and natural climate variability. Lawrence Livermore National Lab., CA (United States).

Flato, G., 2013. Evaluation of Climate Models. IPCC, 768-769.

Fonseca, A., D. P. Ames, P. Yang, C. Botelho, R. Boaventura & V. Vilar, 2014. Watershed model parameter estimation and uncertainty in data-limited environments. Environmental Modelling & Software 51:84-93.

Ford, J., L. Berrang-Ford & J. Paterson, 2011. A systematic review of observed climate change adaptation in developed nations. Climatic Change 106(2):327-336 doi:10.1007/s10584-011-0045-5.

Foord, M., & Muirhead, C. (2009). *IronPython in action*. Manning Publications Co..

Girvetz, E. H., Maurer, E. P., Duffy, P. B., Ruesch, A., Thrasher, B., & Zganjar, C. 2013. *Making climate data relevant to decision making: the important details of spatial and temporal downscaling*. The World Bank.

Hamlet, A., P. Carrasco, J. Deems, M. Elsner, T. Kamstra, C. Lee, S. Lee, G. Mauger, E. Salathe & I. Tohver, 2010. Final Report for the Columbia Basin Climate Change Scenarios Project. University of Washington, Climate Impacts Group, Seattle, Washington, DC.
Hayhoe, K., 2006. . Past and future changes in climate and hydrological indicators in the US Northeast. . Climate Dynamics:381-407.
Hwang, S. & W. D. Graham, 2014. Assessment of alternative methods for statistically downscaling daily GCM precipitation outputs to simulate regional streamflow. JAWRA Journal of the American Water Resources Association 50(4):1010-1032.
Kadlec, J. & D. P. Ames, Development of a Lightweight Hydroserver and Hydrologic Data Hosting Website. In: Proceedings of the AWRA Spring Specialty Conference on GIS and Water Resources, New Orleans, 2012.
Kadlec, J., B. StClair, D. P. Ames & R. A. Gill, 2015. WaterML R Package for Managing Ecological Experiment Data on a CUAHSI HydroServer. Ecological Informatics 28:19-28.
Knutti, R. & J. Sedláček, 2013. Robustness and uncertainties in the new CMIP5 climate model projections. Nature Climate Change 3(4):369-373.
LaFontaine, J. H., L. E. Hay, R. J. Viger, R. S. Regan & S. L. Markstrom, 2015. Effects of Climate and Land Cover on Hydrology in the Southeastern US: Potential Impacts on Watershed Planning. JAWRA Journal of the American Water Resources Association.
OGC, 2012. OGC® WaterML 2.0: Part 1- Timeseries. In: Taylor, P. (ed).
Pradhanang, S. M., R. Mukundan, E. M. Schneiderman, M. S. Zion, A. Anandhi, D. C. Pierson, A. Frei, Z. M. Easton, D. Fuka & T. S. Steenhuis, 2013. Streamflow responses to climate change: Analysis of hydrologic indicators in a New York City water supply watershed. JAWRA Journal of the American Water Resources Association 49(6):1308-1326.
Stöckle, C. O., A. R. Kemanian, R. L. Nelson, J. C. Adam, R. Sommer & B. Carlson, 2014. CropSyst model evolution: from field to regional to global scales and from research to decision support systems. Environmental Modelling & Software 62:361-369.
Taylor, K. E., R. J. Stouffer & G. A. Meehl, 2012. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93(4):485-498.
Werner, A., 2011. BCSD downscaled transient climate projections for eight select GCMs over British Columbia, Canada. Pacific Climate Impacts Consortium, University of Victoria, Victoria, BC, 63pp.
Wooten, A., K. Smith, R. Boyles, A. J. Terando, L. Stefanova, V. Misra, T. Smith, D. L. Blodgett & F. Semazzi, 2014. Downscaled climate projections for the Southeast United States: evaluation and use for ecological applications. US Geological Survey.
Zhang, Z. & J. C. Moore, 2014. Mathematical and physical fundamentals of climate change. Elsevier.

Appendix II

Software Requirements Report

RECLAMATION

Managing Water in the West

Climate Analysis Tools – Software Requirements Report

Research Project: Design and Development of a Prototype Tool for Integrated Climate Downscaling and Streamflow Prediction using Open Source GIS Software

Mission Statements

The mission of the Department of the Interior is to protect and provide access to our Nation's natural and cultural heritage and honor our trust responsibilities to Indian Tribes and our commitments to island communities.

The mission of the Bureau of Reclamation is to manage, develop, and protect water and related resources in an environmentally and economically sound manner in the interest of the American public.

Contents

Page

Executive Summary

The goal of this research project is to develop some prototype tools for acquiring, processing and managing global climate model data using an open source GIS platform. Climate data is inherently geospatial in nature with global climate models and their many scenarios and projections tied to 2-degree grid cells. A number of federal agencies, including Reclamation, have developed and applied downscaling methodologies resulting in vast libraries of climate data. Reclamation, in partnership with Lawrence Livermore National Laboratory, have processed and now store downscaled and bias-corrected climate model data for CMIP3 and CMIP5 scenarios and projections. Users of these climate data libraries currently use somewhat complex online access tools to select and download the data they need to perform more detailed climate and hydrologic modeling.

The first task in this research project is to identify the software requirements of the people with the most immediate need for accessing and processing downscaled climate data. In general, this audience is the climate and hydrologic modelers.

Introduction

The process of identifying and documenting software user requirements for a largely unknown audience of potential users proved to be a major challenge because climate data and is of growing interest across many disciplines. For this reason, the scope of this report was purposefully narrowed to identify the software requirements of the audiences with the most immediate needs for tools to assist their work in considering the impacts of future climate change.

Open Source GIS Platform

A scoping effort for this research project explored and reviewed candidate open source GIS platforms that would be suitable for extension or modification to support a prototype climate tool(s). The basic criteria for selecting an open source GIS platform were:

- established development community with long-term support
- mainstream programming environment (e.g., C, C#, Java, Python)
- use of common open source geospatial data formats
- support for time series data

Several candidate open source GIS platforms were reviewed that meet the above criteria, including:

- QGIS
- HydroDesktop
- GRASS
- MapWindow

All were found to be viable candidates; however, one possessed a specific focus on acquiring, processing and analyzing hydrologic data. The data acquisition and handling functionality have built-in support for time series data, including acquisition, storage, and charting. These features make HydroDesktop, a near ideal platform. Further, HydroDesktop is an open source desktop GIS application supported by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI). The platform is designed with extensibility in mind.

This research project was developed in partnership with one of the main contributors for HydroDesktop, Dr. Daniel Ames, Brigham Young University.
Through a cooperative agreement, Dr. Ames and graduate student staff leveraged the extension framework to build Climate Analysis Tools, the main product of this research project. The user interface for this extension will use the framework, and the associated look and feel. This decision allowed the software requirements task to focus exclusively on functionality and workflow.

Methodology

A series of interviews and conversations were conducted with climate and hydrologic modelers, mostly in Reclamation's Pacific Northwest Region. Initial results were used to develop a first-generation prototype tool that was used for demonstration to help the target audience refine their requirements.

Interviews with Practitioners

Interviews and conversations were conducted beginning in the Fall 2013 with Reclamation staff in the Pacific Northwest Region office, Boise, Idaho. Several were found to already be working with downloading process global climate data. One modeler had developed a couple scripts to automate repetitive processing tasks working with gridded climate data.

Although the group of interviewed users is small, they are reasonably representative because several of them regularly collaborate and participate in projects with peers across Reclamation.

The interviews revealed the following requirements:

- select climate data grids based on a watershed
- download desired grid files for multiple projections, variables, and emissions scenarios
- perform statistical analysis on downscaled climate data
- calculate adjustment factors based on statistical analysis
- apply adjustment factors to observed climate data
- generate climate adjusted input files for hydrologic modeling
- parameterize and run selected hydrologic models
- display routed streamflow on a map

These requirements are described in further detail in the next section.

Software Requirements

The Climate Analysis Tools extension is developed as an extension to HydroDesktop. The software requirements described below are framed in the context of this development platform.

Acquiring Climate Model Data

Reclamation is a principal partner in a major repository (archive) of climate model data hosted at DOE's Lawrence Livermore National Laboratory (LLNL). LLNL has processed and serves downscaled CMIP3 and CMIP climate and hydrology projections. Data are available for two downscaling methodologies: Bias-Correction Spatial Disaggregation (BCSD) and Bias-Correction Constructed Analogues (BCCA). These data (LLNL, 2014) cover the period 1950-2099:

- 234 monthly BCSD CMIP5 projections of precipitation and monthly means of daily-average, daily maximum and daily minimum temperature
- 134 projections of daily BCCA CMIP5 projections of precipitation and daily maximum and daily minimum temperature

Access to these data is provided through a web application where users can initiate a request to select climate data grid files based on projection, geographic area, time period, and several other parameters. The request process is asynchronous. Once the request is submitted, the system cues the process, then sends the user an email with a link when the request processing is complete. Data files must then be downloaded manually, stored locally, and further processed to suit the user's purposes.

Initial investigation into developing some interoperation with the existing capability to support acquiring climate data using Climate Analysis Tools was found to be beyond the scope of the current research. The requirements described in the remainder of this section are currently met using the web application hosted by LLNL.

Interviewed users generally found this to be acceptable, but were interested in a more streamlined approach. Refer to the section Requirements for Future Research in this document.

Select a Global Climate Projection Context

The first requirement is the ability to select either CMIP3 or CMIP5 projections. This requirement is currently met by using the LLNL web application.

Selecting Grids Based on a Watershed

The requirement described by users is the capability to select appropriate climate data grid files based on a geographic area of interest. Typically this is a watershed of some scale ranging from a large river basin like the Columbia River Basin to a watershed generated from pour point and a terrain surface.

Although HydroDesktop has the built-in functionality to generate a user-defined watershed, there is currently no method for passing this geometry to the LLNL web application. Although feasible, creating a data acquisition tool to work with the existing LLNL repository capabilities is outside the scope of this research.

Downloading Desired Grid Files for Multiple Projections, Variables, and Emissions Scenarios

The second step described by users is selecting the desired projection sets, the associated variables, and the model runs by emissions scenario. Again, the LLNL web application provides a user interface that meets this requirement.

Like the previous requirement, it is feasible to develop an interface in the Climate Analysis Tools extension to pass selections to the LLNL web application; however, it is outside the scope of this research. Refer to the section Requirements for Future Research in this document.

Process Downscaled Climate Model Data

Characterize Future Climate and Select Method

For this requirement, users described the need to be able to set one or more future periods to be evaluated, and select the method—Delta, Hybrid Delta, or transient. These selections directly feed the next requirement identified by users.

Select Future Climate Change Scenarios

This requirement is driven by the West-wide Climate Risk Assessments (WWCRA) group to reflect decisions made about how change scenarios are to be bracketed into groups or ensembles where changes in future temperature and precipitation can be represented by the following change scenarios:

- More Warming/Wetter
- More Warming/Drier
- Less Warming/Wetter
- Less Warming/Drier

In this requirement, users need to be able to select percentiles that lead to the above change scenarios. The percentiles are:

- 10%, 50%, 90%
- 25%, 50%, 75%

The requirement element may be implemented with a sliding scale of values from 0 to 100 percent, with the default values preconfigured at 10-50-90.

Select a Single Projection or an Ensemble

This requirement is also driven by a WWCRA decision. The software needs to provide users with the ability to select a single projection or an ensemble of projections to be used in the calculation of adjustment factors (WWCRA, 2014). WWCRA decided ensembles of 10 should be used. The software can provide an option to select the number of members in ensembles. The default value should be 10.

Perform Statistical Analysis on Downscaled Climate Data

For this requirement, the software needs to perform some statistical analysis on the desired downscaled climate data (acquired from LLNL). This requirement entails analyzing processing all the grid files for the area of interest using a Delta, Hybrid Delta, or transient methodology.

Users described this requirement as the capability to browse to climate data (acquired from LLNL), analyze the data, and generate a scatter plot of climate models displaying ensembles based on one or more user-defined future period(s), percentiles defining ensembles, and number of models per ensemble.

One of the interviewed modelers had partially automated this requirement in a script using the open source R statistical software. This code will be used to compare an equivalent capability in the Climate Analysis Tools extension.

Calculate Adjustment Factors Based on Statistical Analysis

In addition to performing statistical analysis and displaying the results, users require the calculation of adjustment factors for temperature and precipitation. These adjustment factors are an interim product needed to adjust forcing files generated from historical observed climate data for modeled future predicted climate conditions.

Apply Adjustment Factors to Observed Climate Data

This requirement entails applying the adjustment factors to forcing file generated from historical observed climate data (downloaded by the user). Users reported a desire to be able to obtain the historical observed data for the same area of interest when acquiring climate model data. Although HydroDesktop has a built-in capability to acquire historical observed hydrologic and weather data from multiple sources, users reported the observed data needs to be from the same source.

Generate Input Files for Selected Hydrologic Models

The final data analysis and processing requirement is generating climate adjusted forcing file to be applied to historical observed hydrologic and weather data for temperature and precipitation in the specific file format of a hydrologic model selected by a user. Although the predominant requirement is to support the VIC model, interviewed modelers also suggested this requirement include other models used in Reclamation. The most frequently mentioned models included the following:

- VIC
- DSHVM
- PRMS
- GSFLOW

Users require the software to output ready-to-use input files for the above models. The scope of this project will likely only be able to implement one or two of these in the prototype Climate Analysis Tools.

Requirements for Future Development

During the process of identifying software requirements, several requirements were found to be beyond the scope of this research project, but are deemed important for consideration in future research and development in this area.

Acquiring Climate Data using Climate Analysis Tools

There are a number of repositories of global climate model data that are likely of interest to climate practitioners. As new downscaling and bias correction methodologies are developed, the need to provide easy access to these climate data products will become increasing important.

There appears to be a growing need to develop a solution similar to CUAHI's Hydrologic Information System (HIS) for acquiring downscaled, bias-corrected climate data. Such a common framework for query and access of climate data across user-selected repositories would save considerable time and allow practitioners to use multiple data sources with greater ease.

Parameterize and Run Selected Hydrologic Models

Several users suggested exploring the feasibility of creating an interface to configure model parameters and run hydrologic models. This would likely require the development of a user interface component for each supported model. A future research direction might be the development of a model interface framework and common model description specification to map model requirements (inputs, data formats, parameters, etc.).

Display Routed Streamflow on a Map

In addition to running hydrologic models, several users suggested displaying the results of model runs on a map as routed streamflow. Because this requirement is dependent on an interface between the model and the HydroDesktop platform, it cannot be directly incorporated in the current project. However, it should be could be considered a requirement of future development involving running hydrologic models and displaying their output in a GIS platform.

References

WWCRA, June 2014, Decision Points for Climate Change (draft).

Interviews with Modelers, conducted during a series of meetings held in Pacific Northwest Region Office during late 2013 into mid-2014.

LLNL, 2014, Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections (website), http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/dcpInterface.html