Software Module for Statistical Analysis in Climate-Related Impact

Gardar Johannesson
(925) 422-3901
johannesson1@llnl.gov

Projected climate change will impact all aspects of everyday life and has the potential to alter the world as we know it. To carry out studies of the potential impact of climate change requires working with massive output data from global climate models (GCMs). These models are inherently uncertain about their future climate projections; they operate on various spatial and temporal scales; and their output is not in general directly suited for use in local/regional climate impact studies.

The starting point of a regional climate impact study is to spatially downscale one or more GCM projections. This can be done by either a finer-resolution regional climate model (bounded by the output from a given GCM) or a statistical model (trained using historical data). In either case, additional uncertainty is introduced in downscaling a given GCM projection.

The aim of this effort is to create statistical methods and tools to assist with local/regional climate impact studies, in particular in terms of quantifying climate-change uncertainty. This involves gathering relevant data from multiple GCMs, both for past and future climate, along with various observation-based datasets for the region of interest. The data-acquiring step is then followed by a statistical analysis of the projected climate change, both at the spatial scale of the GCMs and on statistically downscaled projections.

Project Goals

The goals of this effort are to 1) set up a web-access to a static (already established) database of downscaled results from multiple GCMs; 2) build software tools to gather and process climate data from multiple online databases; and 3) implement statistical analysis tools for both large- and fine-scale analysis of the regionalized climate data.

Relevance to LLNL Mission

Climate change has been identified as a major national and global security threat, as highlighted in a recent report by the National Intelligence Council. This is reflected in LLNL's Science and Technology Plan, which identifies regional climate-change prediction, mitigation, and adaptation as one of its primary thrust areas.

FY2008 Accomplishments and Results

A static database of downscaled temperature and precipitation projections from multiple GCMs is now hosted at LLNL Green Data Oasis with a public web-access interface: http://gdo-dcp.ucllnl.org/downscaled_cmip3_projections/dcpInterface.html.

The downscaled projections cover North America at a one-eighth degree (~ 12 km) spatial resolution and 1-month temporal resolution, which might be sufficient in some cases, but quite often finer-resolution data is required, along with a quantification of uncertainty in the projected changes.

A Python module was created under this effort to gather and postprocess climate-related data. This module leverages the Climate Data Analysis Tools (CDAT) developed by the Program for Climate Model Diagnosis and Intercomparison (PCMDI) at LLNL to extract and postprocess GCM data from the archive at PCMDI. It also fetches large-scale reanalysis data (NCEP2 data) and local weather station data from the online Climate Data Library of the International Research Institution (IRI) of Climate and Society. Hence, given a region of interest such as the Livermore Valley, the Python module is used to build a local database of 1) GCM variables of interest for past and future climate relevant to that region of interest; 2) large-scale observational-based data; and 3) local weather station data. This typically yields 1 to 2 Gb of data (Fig. 1).

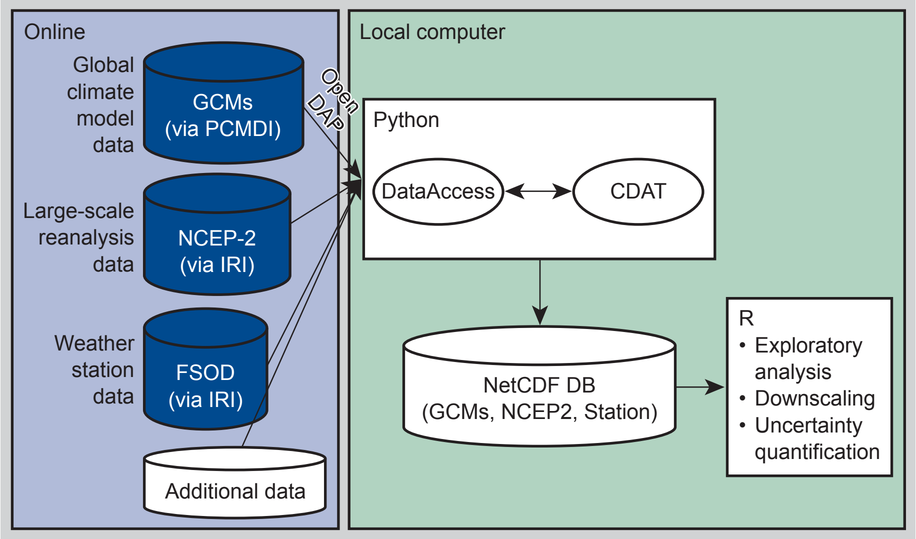

Figure 1. The Python climate-data access (DataAccess) module. The module runs on a local computer and is used to build a local regional climate database by extracting the relevant data from online sources. The local climate database is then analyzed using a library of tools written in the R statistical language.

A library of statistical analysis tools was developed in the R statistical computing language to interact with the regionalized climate database and carry out exploratory data analysis, statistical downscaling of future GCM projections, and uncertainty quantification.

An example of an application of these tools is given in Figs. 2, 3 and 4 for two weather stations located at the Ross and the Diablo Dams in the North Cascade of the state of Washington, an important hydroelectricity region that might be severely impacted by climate change. The location of the dams, approximately 3.5 mi apart, is shown in Fig. 2. Figure 3 summarizes changes in daily average temperature at the GCM grid-scale, while Fig. 4 summarizes changes in daily maximum temperature at the two sites using statistical downscaling. Similar analysis can be carried out for other climate variables of interest, such as precipitation, at the same sites.

Project Summary

This project produced an initial version of a Python module to gather and process climate-related data for a regional impact study. It also yielded a suite of statistical tools in R to carry out analysis of the resulting data. The analysis tools are being refined and extended under LLNL's current Climate Initiative.

Related References

1. Christensen, J. H., *et al.*, "Regional Climate Projections," Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2007.
2. Collins and Knight, Eds., *Phil. Trans. of the Royal Soc. Series A, Theme Issue Ensembles and Probabilities: A New Era in the Prediction of Climate Change*, **365**, pp. 1957–2191, 2007.
3. Hastie, T., R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag, 2003.

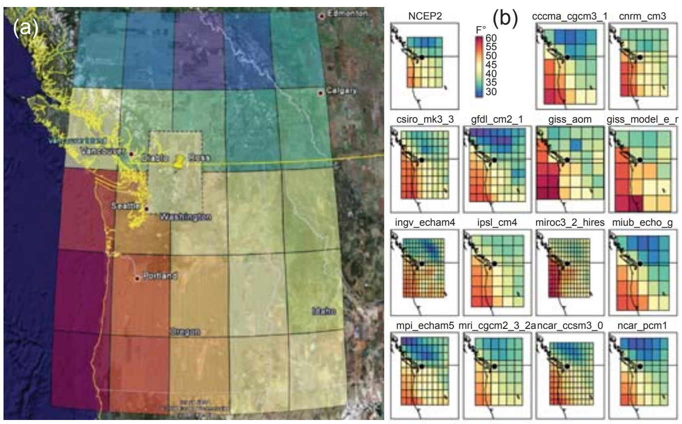

Figure 2. (a) Location of the Diablo and the Ross dams in Washington (yellow pin). The map is overlaid by a 5-x-5 grid showing the 1980 to 2000 average temperature as given by the large-scale NCEP2 reanalysis data. (b) Smaller maps contrast the spatial resolution of the NCEP2 data (upper-left panel) to that of 14 different GCMs in the PCMDI archive.

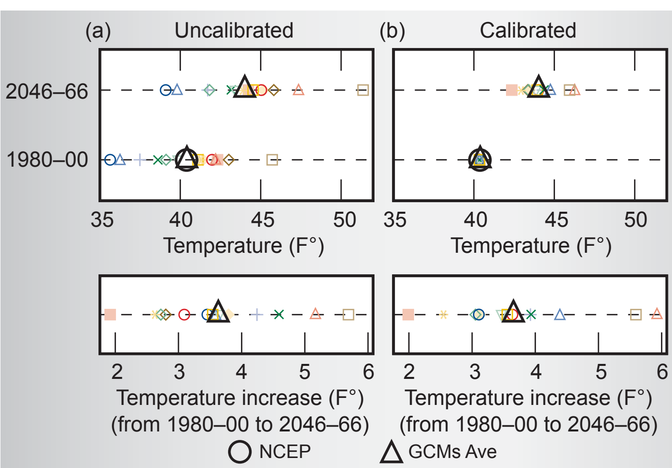

Figure 3. Changes in the average daily temperature between 1980 and 2000 and 2046 and 2066 as projected by 14 GCMs in a single NCEP2-size grid-cell centered on the sites of interest (Fig. 2a). The difference between the uncalibrated plots in (a) and the calibrated plots in (b) is that the data from each GCM is calibrated (shifted and scaled) to yield the same mean and variance as the NCEP2 data for 1980 to 2000. This results in a visible difference in the uncertainty in the projected temperature range (the two top plots), but less so in the temperature increase (the two bottom plots).

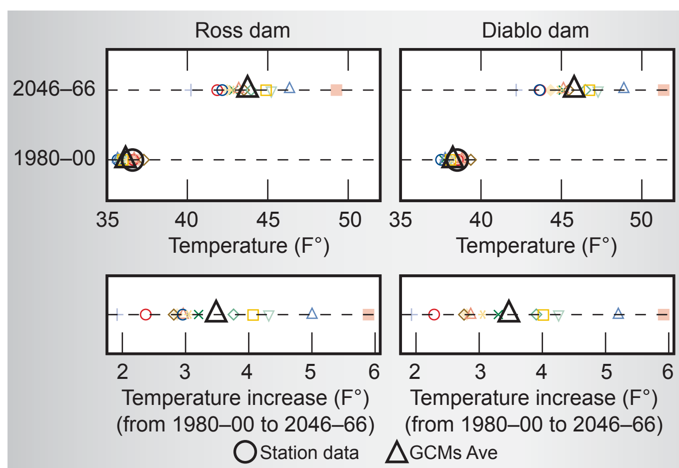

Figure 4. Changes in the daily maximum temperature from 1980 to 2000 and from 2046 to 2066 at the two sites of interest as projected by 13 GCMs. The GCMs data is statistically downscaled using Multivariate Adaptive Regression Splines (MARS) to the two sites using daily temperature and mean sea level data in a 3-x-3 NCEP2-sized grid, where the center grid-point covers the two sites.