**A VisTrails Platform for the USGS Monthly Water Balance Model**

NCCSC Final Report

Andrew Bock1, Andrew Reimanis2, Lauren Hay3, Steven Markstrom3, Roland Viger3, Laura Perry4,5, and Patrick Shafroth5

1USGS Colorado Water Science Center

2Computer Science, Colorado State University

3USGS MoWS National Research Program

4Biology, Colorado State University

5USGS Fort Collins Science Center

Administrative

Recipient: Dr. Lauren Hay

USGS Modeling of Watershed Systems (MoWS), National Research Program (NRP)

West 6th Avenue & Kipling Street, DFC Building 53

Lakewood, Colorado, 80225

Telephone: (303) 236-7279

Fax: (303) 236-5034

E-mail: lhay@usgs.gov

Project Title: Developing a VisTrails platform for modeling streamflow hydrology and projecting climate change effects on streamflow

Agreement Number: 2013 Change of Allocation

Date of Report: January 15, 2015

Period of Time Covered by the Report: July 31, 2013 to December 31, 2014

Actual Total Cost: $95,000

Public Summary

Hydrologic models of streamflow are used throughout the world to forecast and simulate streamflow, inform water management, municipal planning, and ecosystem conservation, and investigate potential effects of climate and land-use change on water availability. Generally, hydrologic models are created as needed for specific watersheds or regions, with different model structures, calibration methods, and data sources used for different locations. This piece-meal approach limits the utility of existing hydrological models for reliable comparisons among different watersheds or regions and for larger-scale regional or national decision-making.

To facilitate consistency in streamflow modeling, we developed an automated modeling platform for the USGS Monthly Water Balance Model (MWBM) using VisTrails, a scientific workflow and provenance management system (www.vistrails.org). Our MWBM–VisTrails platform is built to simulate historic or future streamflow at USGS streamflow gaging stations across the conterminous U.S. The platform is simple to run and includes mechanisms for importing and adjusting climate and landscape data inputs, calibrating and evaluating the model, and visualizing model projections. The automated data processing, calibration, and record-keeping steps save time, reduce error, and can be used to ensure that modeling methods are consistent across different users or regions. The platform allows water managers and research hydrologists to model monthly streamflow more easily and consistently, and thus may facilitate efforts to understand current and historic flow regimes and predict future streamflows under climate change across the U.S.

Technical Summary

Hydrologic models are used throughout the world to forecast and simulate streamflow, inform water management, municipal planning, and ecosystem conservation, and investigate potential effects of climate and land-use change on hydrology. Generally, hydrologic models are created as needed for specific basins or regions, with different model structures, calibration methods, and data sources used for different locations. This piece-meal approach limits the utility of existing hydrological models for reliable comparisons among different basins or regions and for larger-scale regional or national projections and decision-making.

To facilitate consistency in streamflow modeling, we used NCCSC funds to develop an automated modeling platform for the USGS Monthly Water Balance model (MWBM) ([McCabe and Markstrom, 2007](#_ENREF_8)) using VisTrails, a scientific workflow and provenance management system (www.vistrails.org). Our MWBM–VisTrails platform is built to simulate historic or future monthly streamflow at USGS streamflow gaging stations across the conterminous U.S. The platform is simple to run and includes mechanisms for importing and adjusting climate and landscape data inputs, calibrating and evaluating the model, and visualizing model projections. The automated data processing, calibration, and record-keeping steps save time, reduce error, and can be used to ensure that modeling methods are consistent across different users or regions. Because the MWBM–VisTrails platform allows water managers and research hydrologists to model monthly streamflow more easily and consistently, it may facilitate efforts to understand current and historic flow regimes and predict future streamflows under climate change across the U.S.

Purpose and Objectives

The purpose of this project was to develop a tool to make hydrologic modeling easier, more consistent, and better organized. Hydrologic models are used throughout the world to forecast and simulate streamflow, inform water management, municipal planning, and ecosystem conservation, and investigate potential effects of climate and land-use change on hydrology. However, differences in model structures, data sources, data processing, and calibration methods often make it challenging to compare streamflow simulations or projections among studies performed by different researchers for different basins or regions, and thus hinder efforts to make large-scale projections to inform regional and national decision-making. To address this issue, the USGS Modeling of Watershed Systems (MoWS) group is currently developing the infrastructure for a National Hydrologic Model (NHM) to support consistent hydrologic model development and application for the conterminous U.S., using the USGS Monthly Water Balance model (MWBM) and the USGS Precipitation-Runoff Modeling System (PRMS).

Developing a NHM requires tools to ensure that modeling is performed with consistent methods by separate researchers working in different regions, and to organize and maintain detailed records of the specific methods used for each basin. Based on these needs, our objective was to build an automated modeling platform for the MWBM using VisTrails, a scientific workflow and provenance management system that supports data exploration and visualization (www.vistrails.org). VisTrails is designed to integrate, automate and track the use of multiple discrete programs involved in multi-step, complex processes such as streamflow modeling. Our goal was to create a MWBM–VisTrails platform that would integrate programs for importing and adjusting climate and landscape data inputs, calibrating and evaluating MWBM models, and visualizing model projections. In addition, our secondary objective was to use development of the MWBM–VisTrails platform as a test case to evaluate the feasibility and utility of developing a more complex VisTrails platform to model daily streamflow hydrology with the USGS Precipitation-Runoff Modeling System (PRMS) ([Leavesley et al., 1983](#_ENREF_7)).

We met both of our objectives. With assistance from Colin Talbert at the NCCSC, we developed an operational MWBM–VisTrails platform that incorporates climate and landscape data inputs, model calibration and evaluation, and graphical output. Further, we gained useful insights into the potential benefits and challenges of developing a VisTrails platform for PRMS (see Conclusions). However, our MWBM-VisTrails platform does not include all the capabilities we initially expected it to include. For one, we typically use STATSGO soil moisture data for the MWBM, but the MWBM-VisTrails platform uses a hard-coded value of 150 mm instead, because STATSGO data are not yet in netCDF format and therefore cannot be accessed and summarized by the USGS Geo Data Portal. We expect to be able to correct this limitation soon. In addition, more work may be required to ensure that the platform takes full advantage of VisTrails capabilities, particularly with regard to maintaining and organizing records of modeling efforts, use of widgets, and other tools. Our limited experience with VisTrails programming made this aspect of the project more difficult.

Organization and Approach

We hired Andy Reimanis, an undergraduate in the CSU Computer Sciences Department, to do the bulk of the programming for the MWBM–VisTrails platform. Because Andy Reimanis was still training to become a programmer, he required support to accomplish this task. In particular, Andy Bock at the USGS Denver Water Science Center and Colin Talbert at the NCCSC provided Andy Reimanis with considerable programming training and guidance.

Initially, we planned to run the MWBM in VisTrails using a shell script with command lines. Thus, our first main step was to help Andy Reimanis learn to program in Python. However, Jeff Morisette, the NCCSC USGS director, pointed out the utility of using existing R code for some of the platform components. Therefore, our second main step was to help Andy Reimanis learn to program in R.

For some components of the MWBM-VisTrails platform, Andy Reimanis wrote Python or R code from scratch to perform the necessary tasks, with help from Andy Bock and/or Colin Talbert. In other cases, we were able to find and adjust existing R code or Python code to perform necessary tasks. We used only open-source libraries for R code and Python code, to ensure that the MWBM-VisTrails platform could be made fully publically available. In particular, open-source libraries of R code were useful for programming methods to interface with GIS software.

We developed the MWBM–VisTrails platform as a combination of five modules: (1) general model inputs, (2) USGS NWIS streamflow data downloads, (3) USGS Geo Data Portal downloads, (4) MWBM calibration with current conditions, and (5) MWBM simulations for future conditions. For each module, Python and/or R code was written or found to perform the required tasks and revised until operational. Once all five modules were fully operational, we incorporated them into VisTrails. Before his position with us ended, Andy Reimanis tested the resulting MWBM–VisTrails platform for 20 different basins across the conterminous U.S., using different periods of record and different climate datasets (Figure 1). Then, Andy Bock edited the code for each module, removing unnecessary and vestigial components and adding explanatory text. Finally, Andy Bock re-evaluated the minor changes made to the final platform using the same 20 basins for current and projected future conditions.

Project Results

The resulting MWBM–VisTrails platform worked well for modeling monthly streamflow hydrology. The platform successfully ran the MWBM with no glitches for all tested basins, periods of record and climate datasets. Model output for calibration for 20 USGS stream gage example basins is listed in Table 1. Additional examples of model output and graphics are provided in Appendix I.

Analysis and Findings

We successfully developed a MWBM–VisTrails platform that can be used to facilitate simulations of historic and future streamflows at USGS gaging stations across the conterminous U.S. The platform is designed to be simple to run. To model monthly streamflow at a gaging station, the user only needs to input (1) a shapefile of the upstream contributing area, (2) the station identification of the USGS stream gage, and (3) the desired climate dataset and period of record. The platform includes the MWBM structure, which computes the allocation of available water among different hydrologic components of the basin using an accounting procedure that incorporates climatic water supply and demand, seasonality in climatic water supply and demand, snow accumulation and melt, and soil-moisture storage, and has been used successfully in numerous contexts to evaluate effects of climate change on streamflow hydrology (e.g., [Gray and McCabe, 2010](#_ENREF_2); [Hay and McCabe, 2010](#_ENREF_4)). Gridded climate data available at the USGS Geo Data Portal (http://cida.usgs.gov/gdp/) are used as input, and the model is calibrated using measured streamflow at the selected stream gage.

A full description of the MWBM–VisTrails platform is provided in Appendix I. Briefly, in the first module (Model Input), the user inputs the streamflow gage ID, a shapefile of the basin upstream of the gage, start and end dates of the desired simulation period, and the desired location of the working file directory. The module then sets up the file directory, documents basin parameters such as basin surface area and latitude, and uploads the basin shapefile to the USGS Geo Data Portal. In the second module (NWIS Streamflow), the module retrieves measured streamflow from USGS NWIS for the selected gage and period of record, reformats the units and timestep of the retrieved data, and produces a hydrograph for the period of record. In the third module (USGS Geo Data Portal), the user inputs the desired climate dataset, and, for future projections, the desired emissions scenarios and start and end dates of the simulation period. Then the module retrieves monthly precipitation and mean temperatures for the basin over the requested period and reformats the data for the MWBM. Available historic climate datasets include PRISM (www.prism.oregonstate.edu), Daymet (daymet.ornl.gov), and GSD (Maurer et al., 2002), and future climate datasets include bias-corrected, statistically-downscaled CMIP3 projections (gdo-dcp.ucllnl.org). In the fourth module (MWBM Current Conditions), the user inputs the start and end dates for the calibration period. Then the module calibrates the MWBM to the downloaded streamflow and climate data, using the Shuffled Complex Evolution algorithm to optimize model fit based on the monthly and annual Nash-Sutcliffe model efficiency coefficients and the monthly correlation coefficient, The MWBM Current Conditions module also simulates streamflow for the period of record with the calibrated parameters, and creates a series of output graphs illustrating measured and simulated streamflow. Finally, in the fifth module (MWBM Future Conditions), the module uses the calibrated parameters to simulate future streamflow based on downloaded future climate data.

The MWBM–VisTrails platform will allow water managers and researchers to model monthly streamflow more easily, more consistently, and with better organization, and thus may improve understanding of current and historic flow regimes across the country and facilitate projections of future streamflow under climate change. The main advantage of the platform is that the data processing and formatting steps for the MWBM are automated, which saves the user considerable time and reduces the potential for errors in manually-formatted data. In addition, the platform will make it easier for multiple water managers or researchers to model streamflow using truly consistent methods, which in turn will make it easier to compare results from different regions.

Conclusions and Recommendations

With the help of NCCSC staff and funding, we created a MWBM–Vistrails platform that works well for modeling historic and future monthly streamflow across the U.S., is easy to use, and includes all the capabilities we anticipated in an initial MWBM–Vistrails platform. It is worth noting, however, that this effort was more difficult and labor-intensive than we expected. There is a steep learning curve involved in VisTrails programming, as well as in writing Python code for the undergraduate student we employed. It was a major undertaking for the undergraduate student to write most of the code for the project, and it took us considerable time and effort to complete his unfinished tasks. In future efforts to build VisTrails platforms for hydrologic models or other applications, we recommend employing an experienced VisTrails and Python programmer who could make an effective product more efficiently.

A number of “next steps” could be taken to improve or build upon our MWBM–Vistrails platform. The soil moisture holding capacity parameter for each basin is fixed at 150 mm. This is the default setting for previous continental and global applications of the MWBM (McCabe and Wolock, 2011). Efforts are currently underway to make STATSGO data available in netCDF format, after which we will be able to estimate soil moisture holding capacity locally for each basin. Second, the retrieval and summary of CMIP5 data from the GDP is currently not operational, but can be integrated quickly in the future. Third, an experienced VisTrails programmer may be able to improve VisTrails functionality within the platform, particularly with regard to how records of model outputs and provenance are maintained and organized. Fourth, the fourth module could be expanded to include a mechanism to evaluate model fit for a validation period independent of the calibration period. Finally, the fourth and fifth modules could be expanded to calculate commonly used flow metrics to characterize simulated streamflow.

A much larger next step would be to expand the MWBM–VisTrails platform into a more complex platform for modeling daily streamflow with PRMS. However, given the challenges we encountered developing a VisTrails platform for the much simpler MWBM, we are uncertain that the benefits of a PRMS-VisTrails platform would warrant the time, effort, and expense required to create it. Alternatively, other researchers could use our MWBM–VisTrails platform as the streamflow component for more complex VisTrails platforms aimed at modeling vegetation, wildlife, ecosystem, or landscape responses to climate or land-use change.

Outreach

We do not anticipate any publications or presentations on the MWBM-VisTrails platform. However, the platform will be hosted at multiple public sites, including GitHub and USGS ScienceBase.

Literature Cited

Leavesley, G.H., Lichty, R.W., Troutman, B.M., and Saindon, L.G., 1983. Precipitation-runoff modeling system - User's manual. U.S. Geological Survey Water-Resources Investigation Report 83-4238.

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P., and Nijssen, B. 2002. A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. Journal of Climate 15:3237-3251.

McCabe, G.J., and Markstrom, S.L., 2007. A monthly water-balance model driven by a graphical user interface. U.S. Geological Survey Open File Report 2007-1088, 6 pp.

McCabe, G.J., and Wolock, D.M., 2011. Independent effects of temperature and precipitation on modeled runoff in the conterminous United States. Water Resources Research, 47:11.

Table 1. USGS streamflow gages and time periods used to test the MWBM-VisTrails platform, and three statistics indicating calibrated model fit to observed streamflow data. Monthly climate data for the calibration period were PRISM Climate Group data (www.prism.oregonstate.edu) hosted at the USGS Geo Data Portal.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| USGS Gage | River | State | Calibration  Period | NSEm | NSEa | COR |
| 01033500 | Pleasant River | ME | 1930-1950 | 0.79 | 0.55 | 0.90 |
| 01165500 | Moss Brook | MA | 1940-1955 | 0.80 | 0.82 | 0.90 |
| 01411300 | Tuckahoe River | NJ | 1975-1987 | 0.85 | 0.85 | 0.93 |
| 01644000 | Goose Creek | VA | 1930-1960 | 0.76 | 0.82 | 0.88 |
| 02059500 | Goose Creek | VA | 1960-1980 | 0.72 | 0.71 | 0.87 |
| 02464146 | Turkey Creek | AL | 1985-1992 | 0.59 | 0.49 | 0.80 |
| 03065000 | Dry Fork | WV | 1950-1965 | 0.79 | 0.97 | 0.90 |
| 03235500 | Tar Hollow Creek | OH | 1960-1967 | 0.36 | 0.89 | 0.66 |
| 03450000 | Beetree Creek | NC | 1940-1970 | 0.79 | 0.87 | 0.90 |
| 04055000 | Manistique River | MI | 1940-1955 | 0.60 | -0.80 | 0.88 |
| 04213075 | Brandy Run | PA | 1990-2000 | 0.69 | 0.74 | 0.86 |
| 05399500 | Big Eau Pleine River | WI | 1940-1965 | 0.63 | 0.63 | 0.80 |
| 05559000 | Gimlet Creek | IL | 1950-1965 | 0.46 | 0.77 | 0.68 |
| 06335000 | Little Beaver Creek | ND | 1940-1955 | -0.07 | -0.18 | 0.04 |
| 07052120 | South Creek | MO | 2000-2005 | -0.02 | -0.15 | 0.05 |
| 07249413 | Poteau River | OK | 1990-2000 | 0.77 | 0.98 | 0.88 |
| 07362100 | Smackover Creek | AR | 1950-1970 | 0.63 | -0.12 | 0.84 |
| 08218500 | Goose Creek | CO | 1960-1975 | -0.27 | 0.15 | 0.06 |
| 09037500 | Williams Fork | CO | 1940-1970 | 0.83 | 0.53 | 0.93 |
| 09361000 | Hermosa Creek | CO | 1930-1950 | 0.24 | 0.58 | 0.55 |
| 12035000 | Satsop River | WA | 1930-1950 | 0.92 | 0.68 | 0.97 |

NSEm – Monthly Nash-Sutcliffe Coefficient of Model Efficiency

NSEA – Annual Nash-Sutcliffe Coefficient of Model Efficiency

COR – correlation coefficient

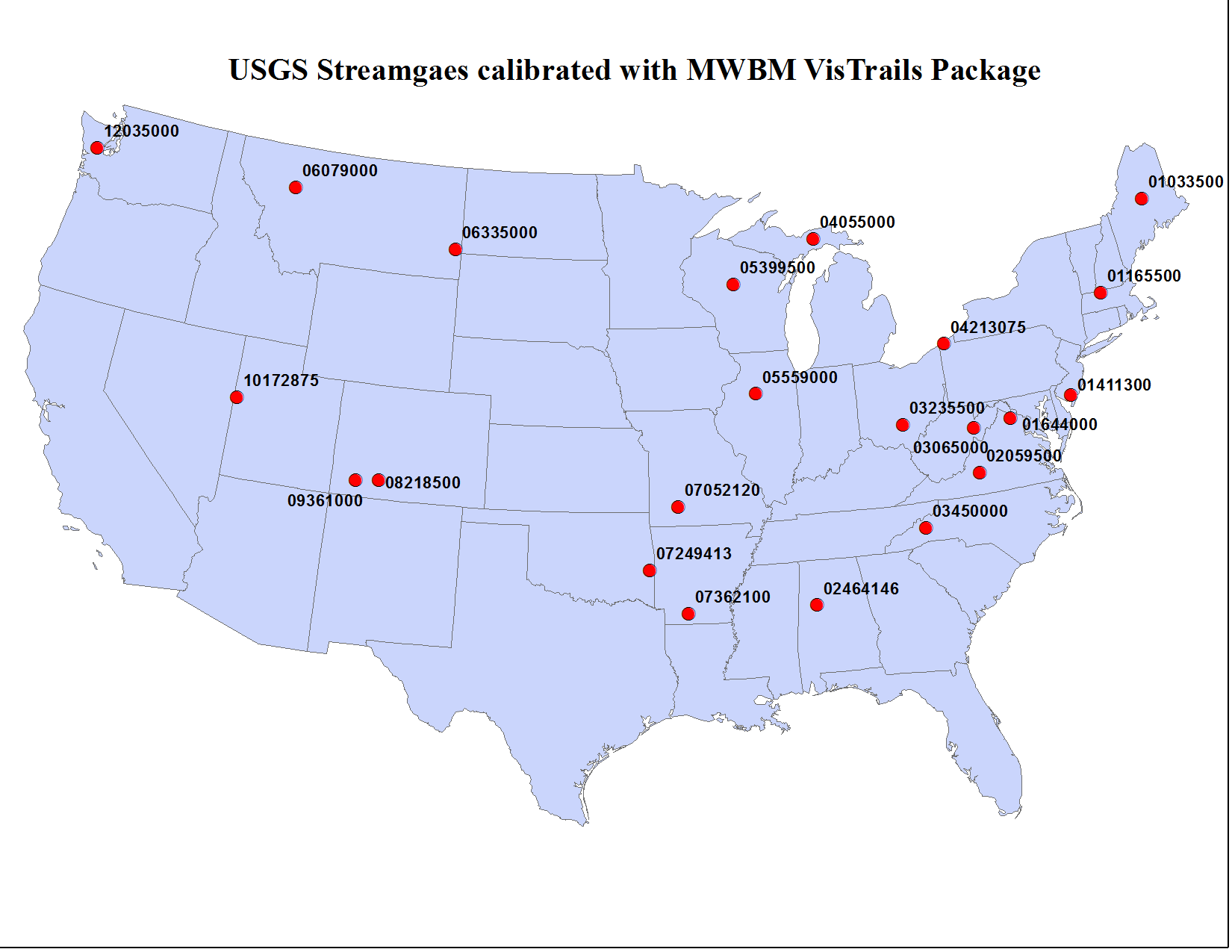
****

Figure 1. USGS streamflow gages used to test the MWBM-VisTrails platform.

Appendix I. Guide to the MWBM-VisTrails Platform

The VisTrails MWBM Tool contains five modules. Only a handful of user inputs and arguments are necessary. The outline below explains how each module operates and the necessary information to run the entire workflow.

**Model\_Input –** Sets up the model working directory, creates documentation for shapefile and geographic model input parameters for basin, uploads shapefile of basin to the USGS GeoData Portal.

**NWIS\_Streamflow** – Retrieves streamflow at daily and monthly timesteps for the streamgage identified in model input.

**pyGDP** – Retrieves and summarizes climate data to drive the MWBM for the specific basin using the USGS GeoData Portal.

**WBM CC** – Calibrates and simulates the Monthly Water Balance Model for station-based datasets for historic and current conditions.

**WBM FC** – Calibrates the Monthly Water Balance Model for current conditions using the Maurer Gridded Station-based Data (GSD) and simulates the Monthly Water Balance Model for future conditions using Bias-corrected statistically-downscaled data (BCSD).

**Model Input**

Input Ports (Figure I):

* *Simulation Start Date (CC):* The start month and year of calibration of MWBM for current conditions (string, M-YYYY).
* *Simulation End Date (CC):* The end month and year of calibration of MWBM for current conditions (string, M-YYYY).
* *GageID:* The 8-12 digit USGS station ID (STAID) for the streamgage to be used for MWBMcalibration/comparison purposes (string).
* *Shapefile:* Shape file to be uploaded to the USGS GeoData Portal for purposes of summarizing climate information and providing geographic model parameters (file location). The shape file should be a single feature and have a standard projection system.
* *ID Field*: Attribute field within the shape file that provides unique ID for basin for data queries (string).
* *Working Directory:* Folder location where the working directory will be set up. A unique folder will be set up for each unique basin (directory).

Output Ports (Figure 1):

* *Model\_info*: The python dictionary with the model run file information. In this module, the GageID, Shapefile, ID field of shapefile, working directory, and the POR for simulation for current conditions are added to the dictionary.

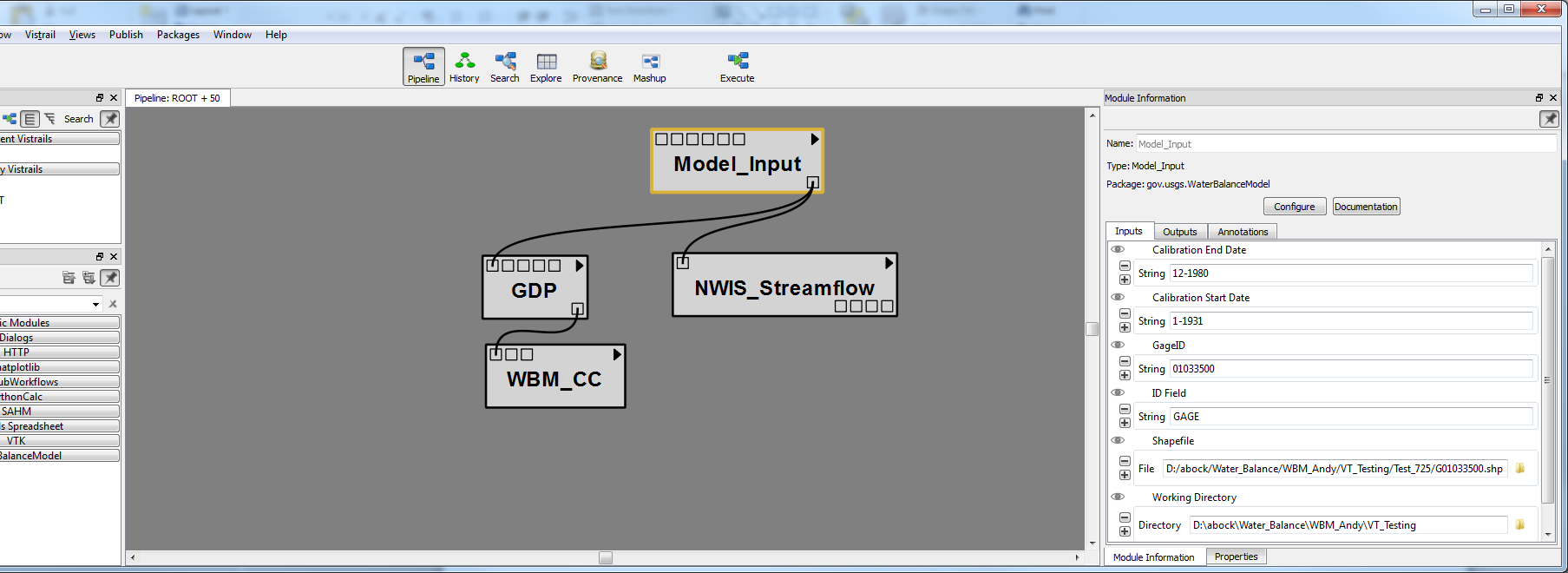


Figure I. Input Port Arguments for ‘Model\_Input’

This module initiates the project. A working directory is created in the specification provided by the user in the input port, which is named for the STAID with a G appended to the beginning of the name to avoid naming directories beginning with 0 (Figure II). Three sub-directories are created:

* *BASE* – Holds files that document information and specifications of the model run, as well as streamflow data used to calibrated the MWBM.
* *CC* – Folder directory for climate data represented by station-based gridded datasets that span current and historic conditions.
* *FC* – Folder directory for downscaled datasets that span current and future conditions.

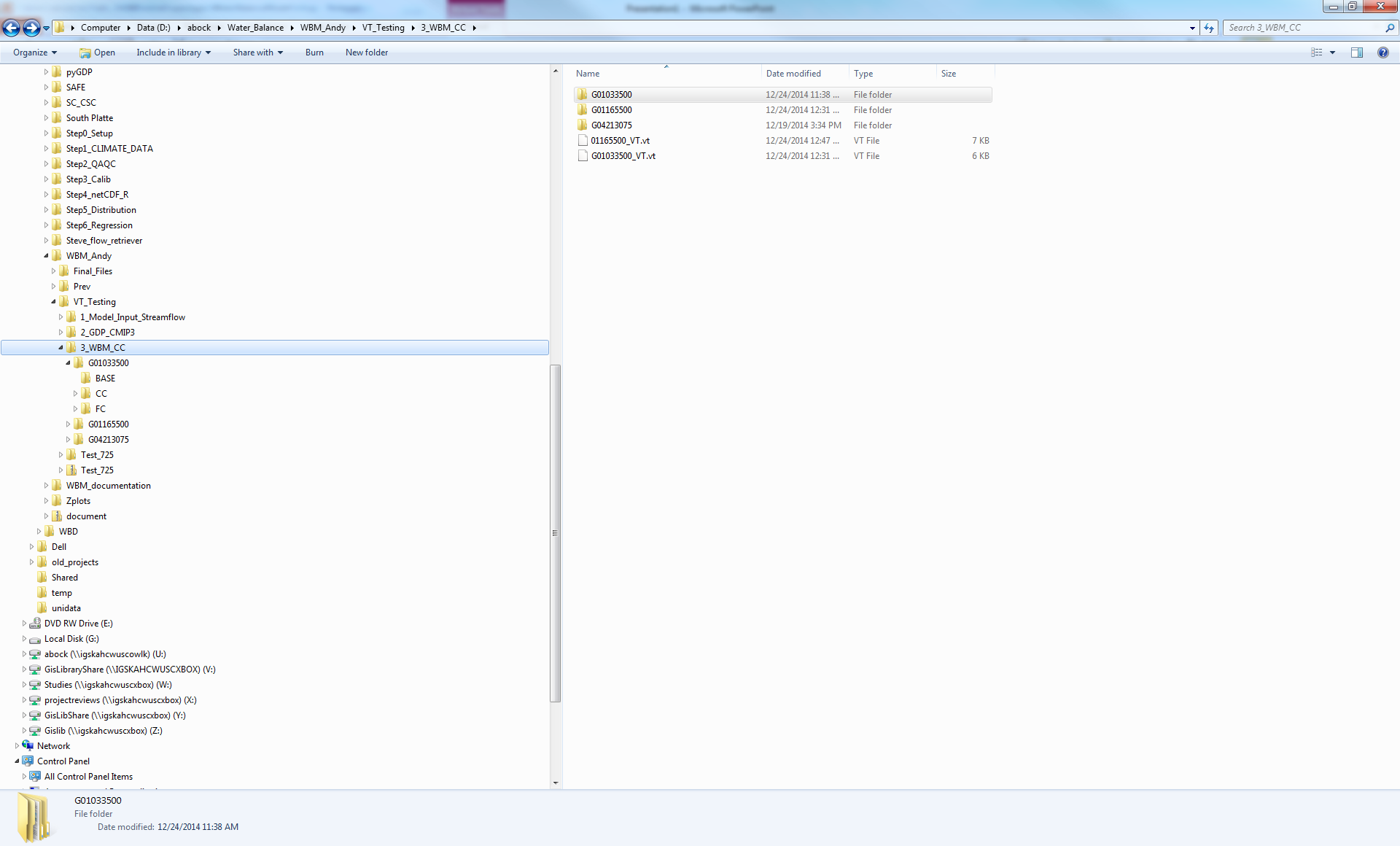


Figure II. MWBM working directory.

Within the *BASE* folder the program creates a text file identifying the unique parameters and inputs chosen by the user for each MWBM simulation (Figure III). Each file is given a unique name that is a combination of the USGS STAID, date, and time of simulation. Information, results, and data from subsequent modules are appended to the file. The contents of this file are also contained digitally in a python dictionary called “model\_info” that is passed as a global argument to other modules.

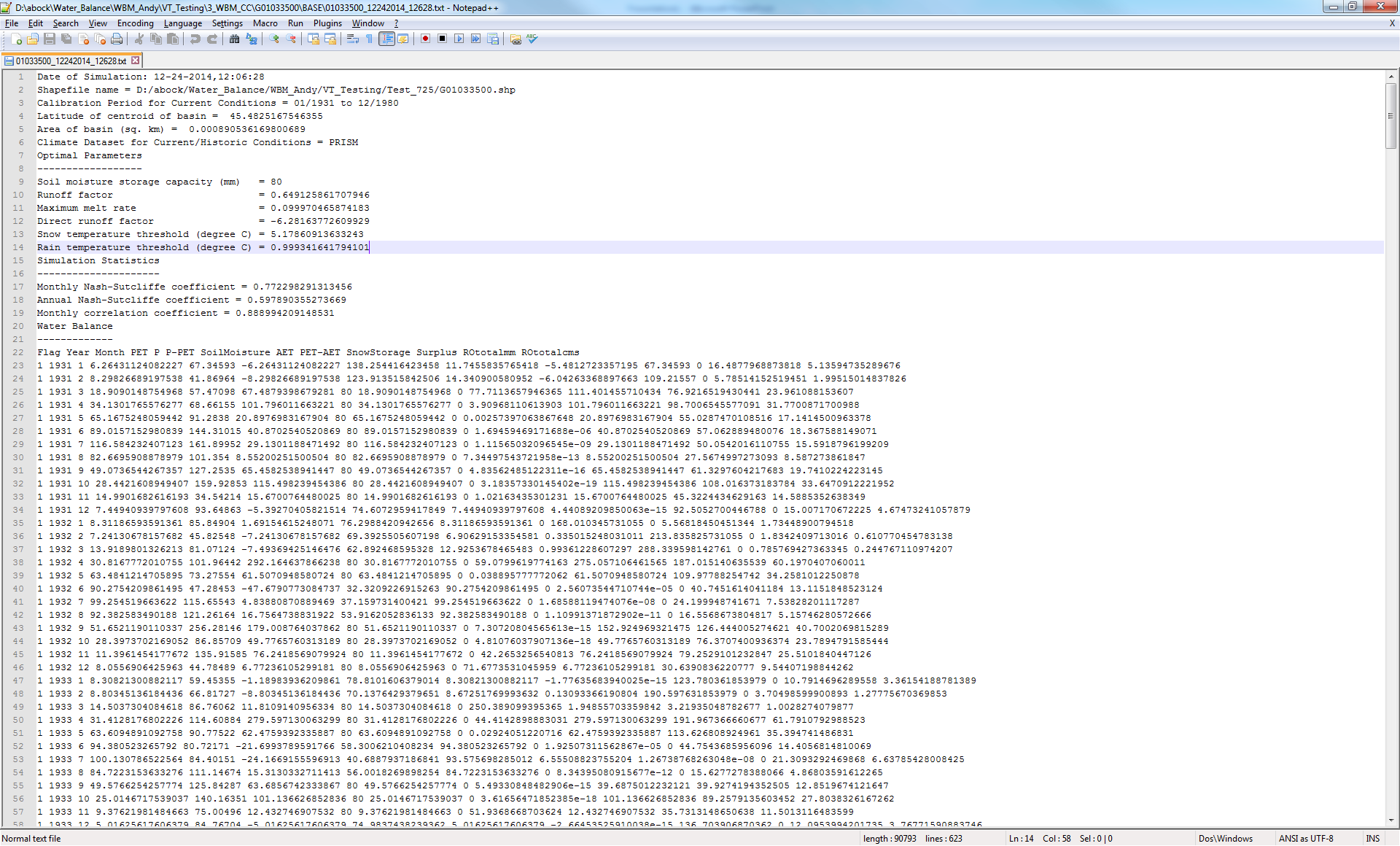


Figure III. Header of model run information textfile stored in the *BASE* directory.

Three pieces of geographic information from the basin are MWBM input parameters: basin area, latitude of the basin, and the basin soil moisture storage capacity. Two of these parameters are determined in this module: latitude of the basin centroid (decimal degrees) and basin area (square kilometers). Both parameters are written to the model info file and python dictionary. The third component is the Soil Moisture Storage Capacity. This is currently hard-coded as 150 mm. The last step of Model\_Input is to compress the shapefile components into a single zip file, which is then uploaded to the USGS GeoData Portal for summarization of climate data for the basin based on user inputs specified in the GDP Module.

**NWIS Streamflow**

Input Ports :

* *Model\_info*: The python dictionary with the model run file information.

Output Ports:

* *Daily Streamflow*: File of measured streamflow for the designated streamgage for the period of record identified in units of cubic feet per second.
* *Monthly Streamflow*: Streamflow aggregated to the monthly timestep and converted to units of average cubic meters per second for use with MWBM.
* *Hydrograph*: Hydrograph of measured daily streamflow for the period of record (Figure IV).

The NWIS Streamflow module retrieves the measured streamflow data for the designated streamgage from the model input module for the calibration period designated. Daily streamflow in cubic feet per second is written to ‘Flows\_daily.txt’, and monthly streamflow in cubic meters per second is written to ‘Flows\_monthly.txt’. Both files are saved to the *BASE* folder. If no streamflow data is retrieved, a message box will pop up informing the user, and will offer the user a chance to start the MWBM workflow over (Figure Va). If less than 5 years of streamflow data is retrieved, another message with the same functionality will pop up (Figure Vb). At least five full years of data is recommended for reliable monthly calibration. Otherwise, the message box will inform the user of the number of years of streamflow data retrieved (Figure Vc). After successful extraction of streamflow data, a png file of the daily streamflow for the designated period of record is also saved to the *BASE* folder (Figure IV).

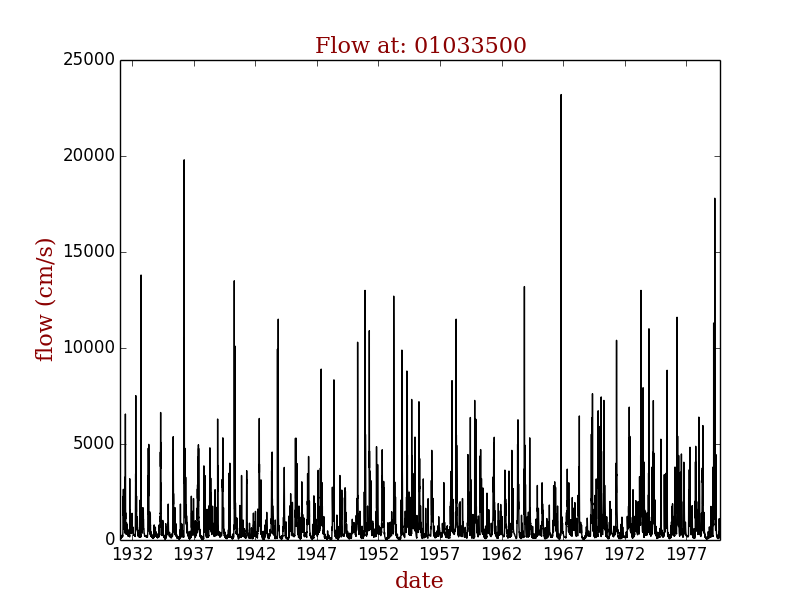
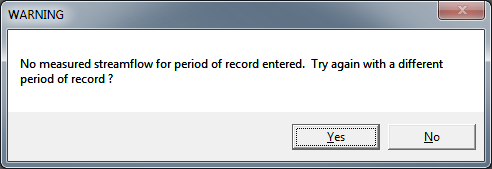
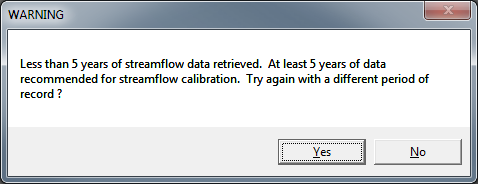


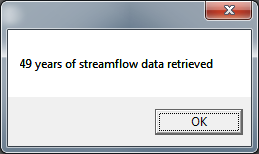
Figure IV. Hydrograph produced from the NWIS streamflow plot.



a.



b.



c.

Figure Va-c. Messages generated by the NWIS\_Streamflow module.

**GDP**

Input Ports (Figure VI):

* *Model\_info*: The python dictionary with the model run file information.
* *Climate Dataset*: The climate dataset for which to summarize information (String; “PRISM”, “GSD”,”DAYMET”, “CMIP3”,”CMIP5”)
* *Scenarios*: Scenarios to retrieve climate data for when climate dataset is ‘CMIP3’ or ‘CMIP5’ (String; “a1b”, “a2”, “b1”, “rcp2.6”, “rcp4.5”, “rcp6.0”, “rcp8.5”).
* *Simulation Start Date (FC):* Beginning of period of record for downscaled data for future conditions (String, M-YYYY).
* *Simulation End Date (FC)*: End of period of record for downscaled data for future conditions (String, M-YYYY).

Output Ports:

* *Model\_info*: The python dictionary with the model run file information. In this step the module appends climate data dictionary to the model\_info dictionary, and includes the climate dataset name, timestep, and a number of additional properties that were used to summarize the dataset with the pyGDP.

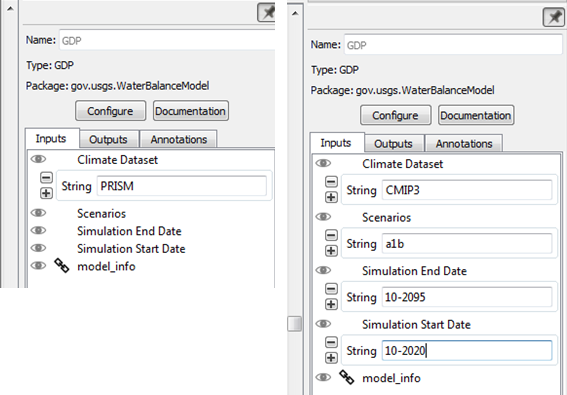


Figure VI. Input Port Arguments for the GDP module for station-based datasets on the left, and the bias-corrected statistically downscaled data on the right.

The GDP module summarizes the climate dataset selected by the user for the basin of the streamgage selected by the user with climate datasets on the USGS Geodata Portal (GDP). The GDP utilizes a web-based processing service to provide consistent data access, analysis, and formatting for a variety of georeferenced climate and environmental datasets, including CONUS-extent climate data for current and future conditions. The GDP web-processing service allows gridded climate data to be summarized for many forms of georeferenced vector data such as feature classes shapefiles created by desktop GIS software. The shapefiles can then be loaded to the GDP through the GDP website, called from a hosting location in the USGS Science Base catalog, or loaded programmatically. The GDP allows users to customize the climate dataset, variables, and period of record they want summarized for a feature class of their given area of interest. For more efficient and automated access to the GDP, pyGDP, a python library utilizing open source capabilities, was developed for the python environment. For more details on the GDP see Blodgett et al. (2011, 2013).

The GDP module utilizes the pyGDP python library to access the GDP. A zipped shapefile of the basin was uploaded to the GDP in the model input module. A number of different arguments are required by users to run the pyGDP web processing functions. The GDP module streamlines these arguments to include just the simulation start and end dates specified in model input modules, and the name of the climate dataset specified in the GDP module. Three historic and current gridded station-based datasets are available to the user for MWBM calibration for historic and future conditions: PRISM (PRISM, 2004), GSD (Maurer et al., 2002), and Daymet (Thornton et al., 2014). For future conditions, the user can choose dynamically-downscaled climate data from the CMIP3 and CMIP5 model families (U.S. Bureau of Reclamation 2011, 2014). To select a dataset, the user types in the name of the dataset (“PRISM”, “GSD”, “DAYMET” for the current conditions station-based datasets, “CMIP3” or “CMIP5” for the statistically-downscaled datasets) to download in the “Climate Dataset” box. If one of the three station-based datasets is selected, it will retrieve data for the date for the simulation start and end dates entered in “Model Input” module. If “CMIP3” or “CMIP5” is selected, the simulation start and end dates entered in model input will be considered the range of dates to be considered for WBM calibration, and the user can enter separate simulation start and end dates for simulating CMIP3/5 for future conditions with the Simulation Start Date and Simulation End Date arguments in the GDP module. For the downscaled datasets, the user enters the scenarios they are interested in summarizing in the “Scenarios” box (“a1b”, “a2”, “b1” for “CMIP3”, “rcp2.6”, “rcp4.5”, “rcp6.0”, and “rcp8.5” for “CMIP5”). The module is capable of obtaining any combination of scenarios for either the CMIP3 or the CMIP5 dataset.

The climate data is summarized using the GDP, formatted into MWBM-ready input “PPT\_month.txt”, “TAVE\_month.txt”), and written to its own folder. Data summaries for the three station-based datasets are written under the *CC* directory in the project workspace, and the downscaled datasets are written under the *FC* directory. If the user entered PRISM, GSD, or DAYMET in the climate argument, the GDP module will retrieve climate data for the simulation period indicated in the model setup module (“Simulation Start Date”, “Simulation End Date”). If the user summarized climate data for CMIP3 or CMIP5, GSD will also be retrieved for the simulation period indicated in the model setup module. Both “CMIP3” and “CMIP5” are given their own folder space under the *FC* directory, with sub-directories containing the scenarios, the GCMs, and the initial GCM simulations (Figures VII, VIII).

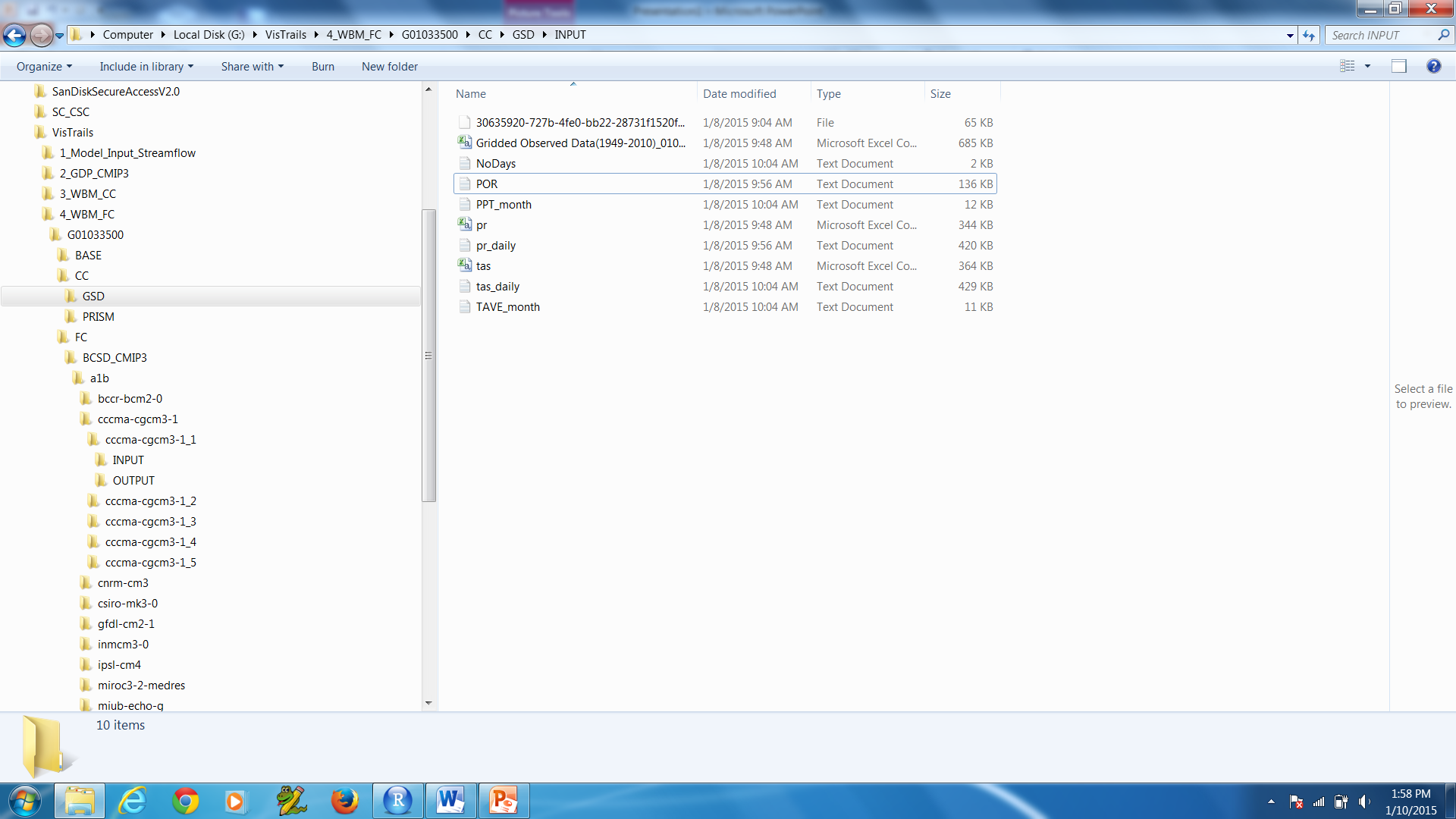


Figure VII. Structure of the *CC* and *FC* directories

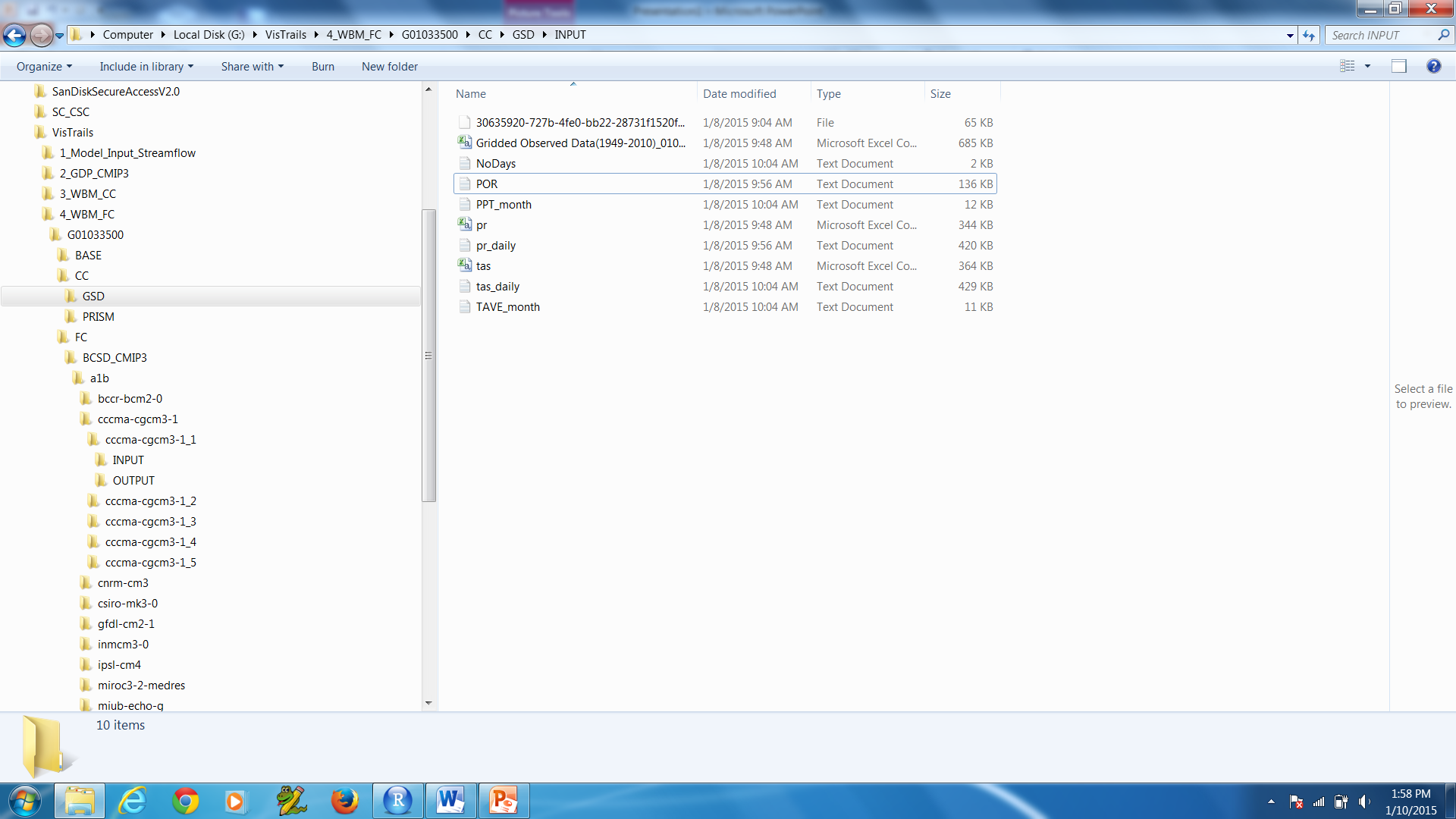


Figure VIII. Contents of Input folder for “GSD”

**Water Balance Model (Current Conditions)**

Input Ports (Figure IX):

* *Calibration POR:* Beginning of period of record for calibration for current conditions (String,YYYY,YYYY).
* *Evaluation POR*: End of period of record for evaluation for current conditions (String, YYYY,YYYY).
* *Model\_info*: The python dictionary with the model run file information.

Output Ports:

* *None –* The module writes calibration output to several places. Performance statistics, and outputs are written to the model run information. Model outputs and summary graphs are written to the *OUTPUT* folder of the climate dataset being calibrated.

The MWBM\_CC module calibrates the Monthly Water Balance Model (McCabe and Markstrom, 2007) for current conditions using the station-based dataset indicated by the user in the GDP module and the measured streamflow obtained from the NWIS module. If the user specified “CMIP3” or “CMIP5” in the GDP module, the model will be calibrated using the GSD dataset. The module leans on a water balance model and calibration package written in R (Elsner, 2013). The five parameters calibrated are: the *direct runoff factor*, which determines the proportion of monthly precipitation that becomes runoff, the *surplus runoff*, which determines the proportion of monthly surplus that becomes runoff, *tsnow*, which is the temperature below which all precipitation falls as snow, *train*, which is the temperature above which all precipitation falls as rain, and *meltcoef*, which determines snowmelt (Table I).

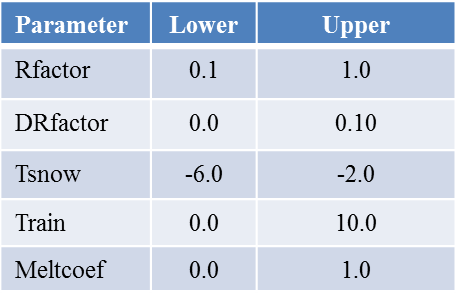
**

Table I. Model parameters and their defined ranges.

The module has three input ports: Calibration POR, Evaluation POR, and model\_info (Figure VIII). The Calibration POR is the period of record for which the user wishes to calibrate the MWBM. This argument is two years (the beginning and ending year) separated by a comma. The user should rely on the hydrograph image (“*BASE/hydrograph.png*”) and knowledge of the streamgage to pick a reliable calibration period. The calibration period should be at least five years long. The Evaluation POR is a second POR that the user can select to test the robustness of the optimized parameters derived from the MWBM calibration. This should also be a length of five years or greater, and, if possible, should be a POR independent of the calibration POR. The module uses a number of items from the model\_info python dictionary file, including the working directory, the basin ID, latitude, area, and water holding capacity of the basin, the simulation start and end dates, and the filename holding the simulation info.

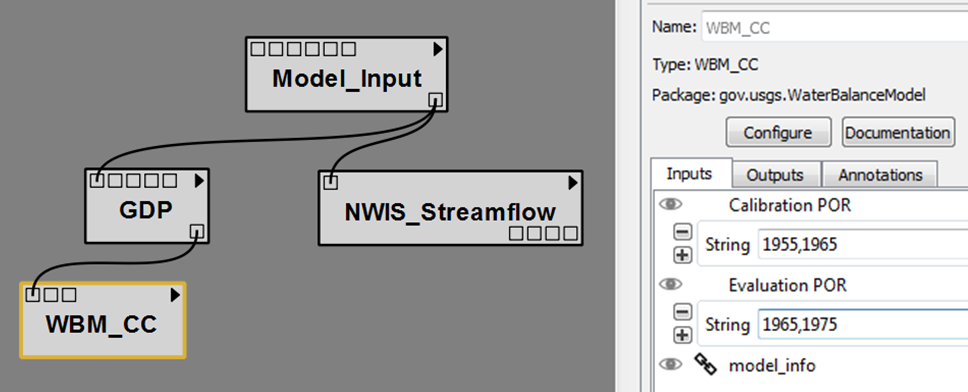


Figure IX. WBM\_CC input ports.

The calibration uses an objective function that is equally weighted for three different statistics: the monthly Nash-Sutcliffe model efficiency coefficient (mNSE), the annual Nash-Sutcliffe model efficiency coefficient (aNSE), and the monthly correlation coefficient (R). The NSE measures the predictive power of hydrologic models, and ranges from - ∞ to 1, with 1 being a perfect prediction and values less than 0 providing less predictive power than the streamflow mean (Nash and Sutcliffe, 1970). The correlation coefficient measures the strength and direction of the relationship between two datasets, and ranges from -1 to 1 with -1 meaning there is a perfect negative correlation between the two datasets, 0 indicating no correlation, and 1 indicating a perfect positive correlation between the two datasets. The Shuffled Complex Evolution algorithm (SCE) is used as the optimization algorithm. For more information on SCE see Duan et al. (1994). The optimization algorithm determines the set of parameter values for the five MWBM parameters that simulate streamflow from the MWBM that best matches the observed streamflow from the designated streamgage. The five optimized parameters are written to the model run information, and the MWBM is simulated for the entire simulation period entered from the Model Input module. After a successful calibration and simulation, the module produces a number of plots of the simulated and measured streamflow aggregated at different timesteps.

The outputs are written to a number of different locations. The performance statistics and model outputs for variables and streamflow in several different units is written to the model run information in the *Base* folder (Figure (X). All model outputs are in millimeters, with Streamflow summarized for centimeter per second (cm/s). Straight model output is also written to the *Output* folder of the climate data being calibrated. Several graphs summarizing model output aggregated at different timesteps are also written to the same *Output* folder (Figures XI-XV).

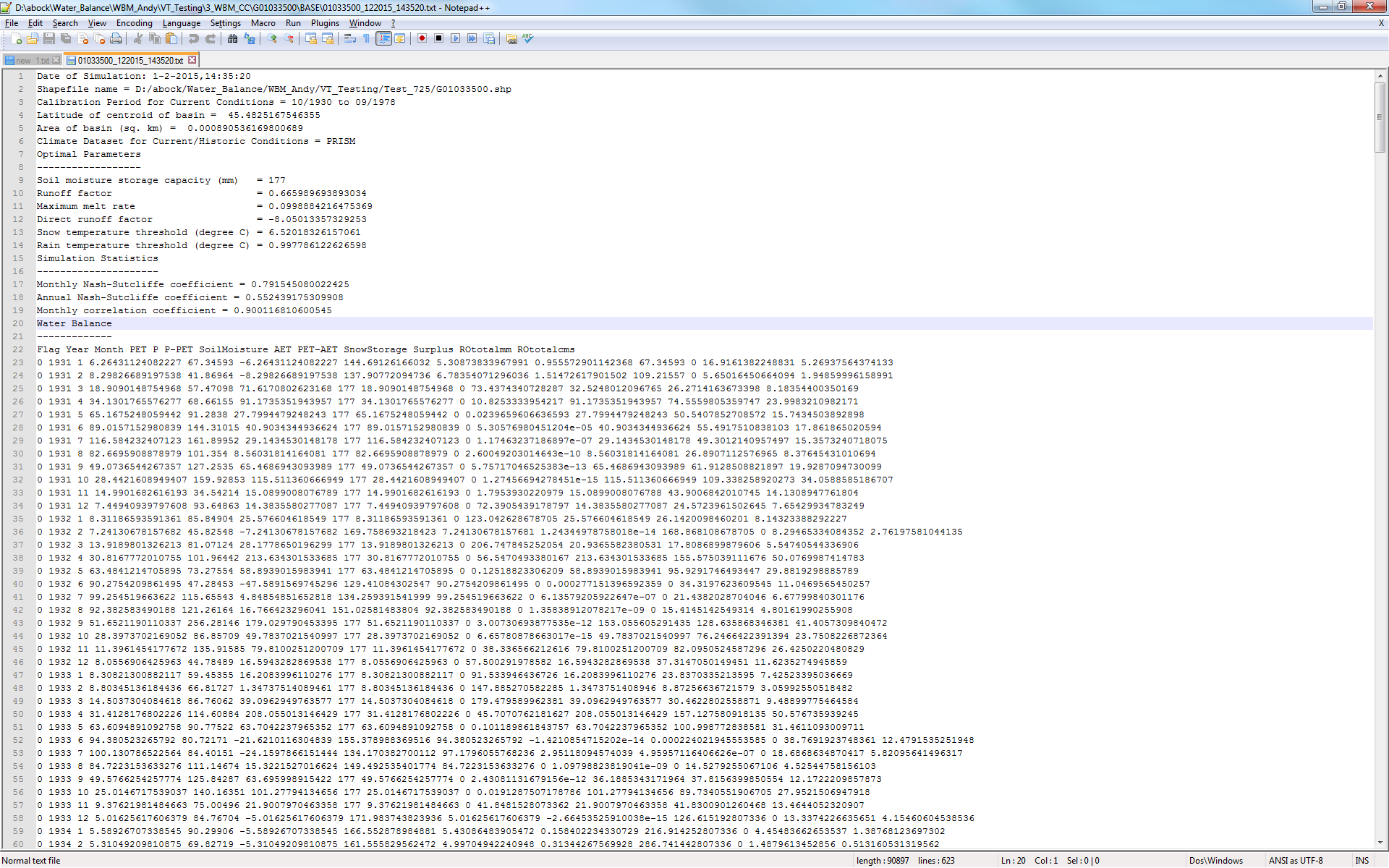


Figure X. Model Output and calibration statistics written to the model info file in the *Base* directory.

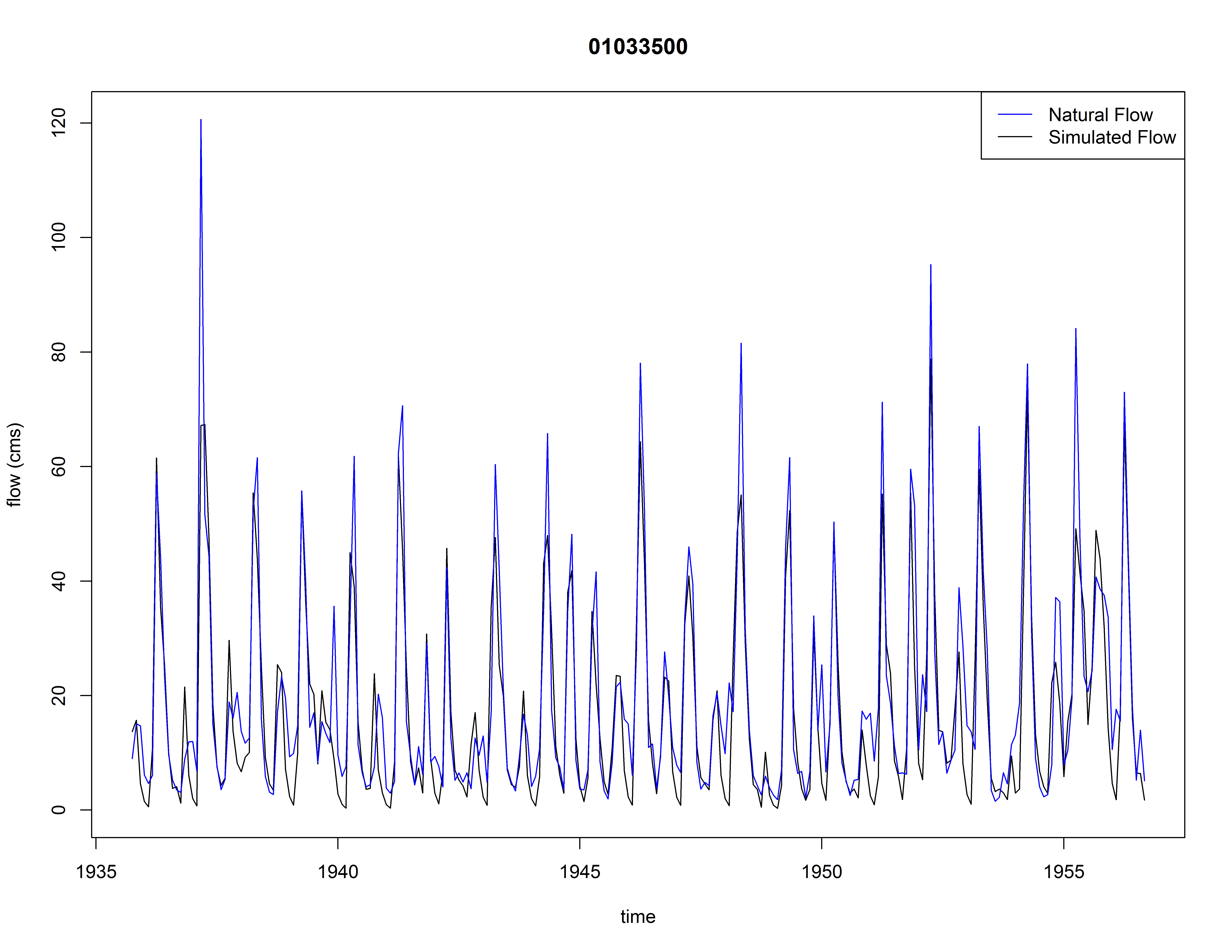


Figure XI. Monthly simulated and measured streamflow for the streamgage 01033500 for the period of record (1930-1980).

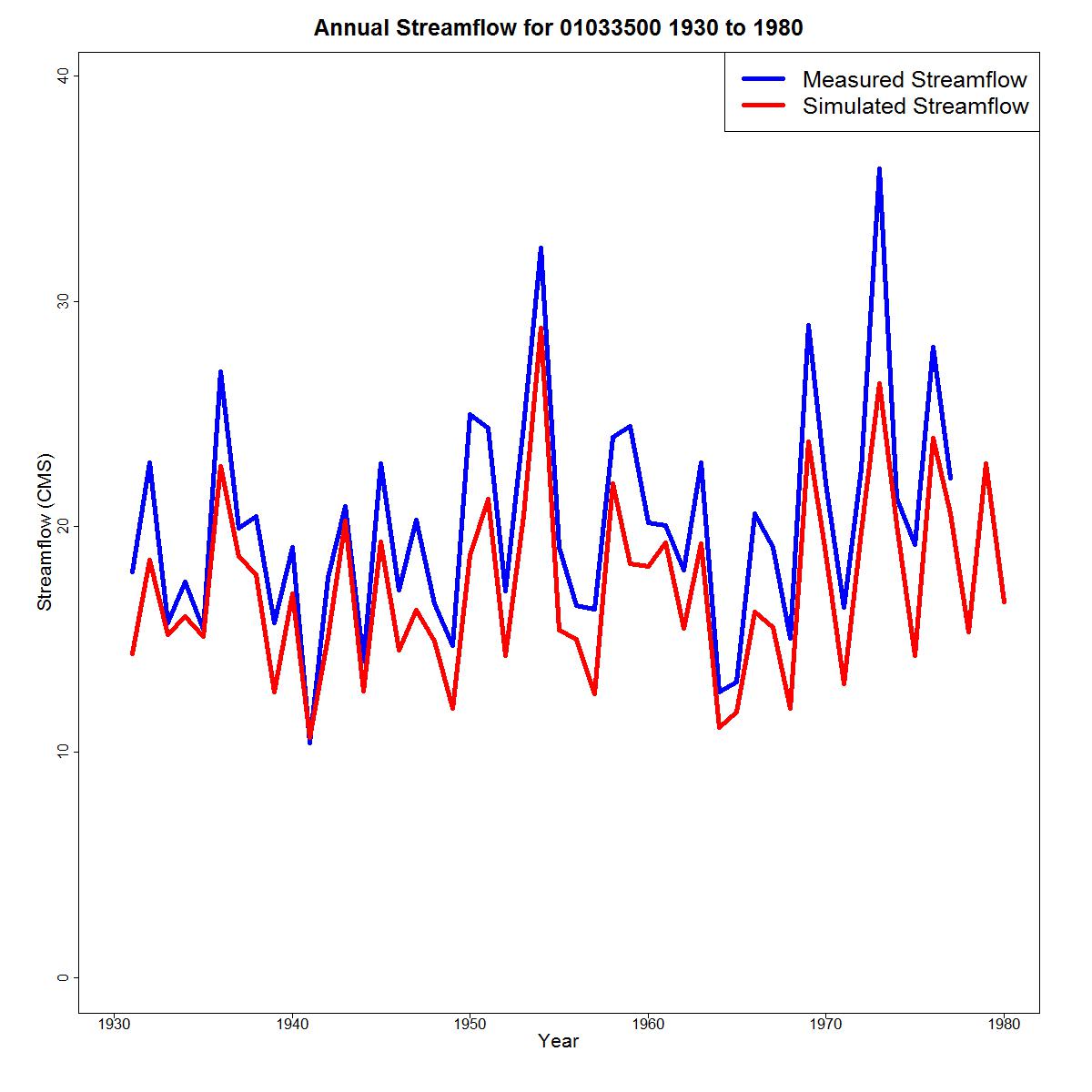
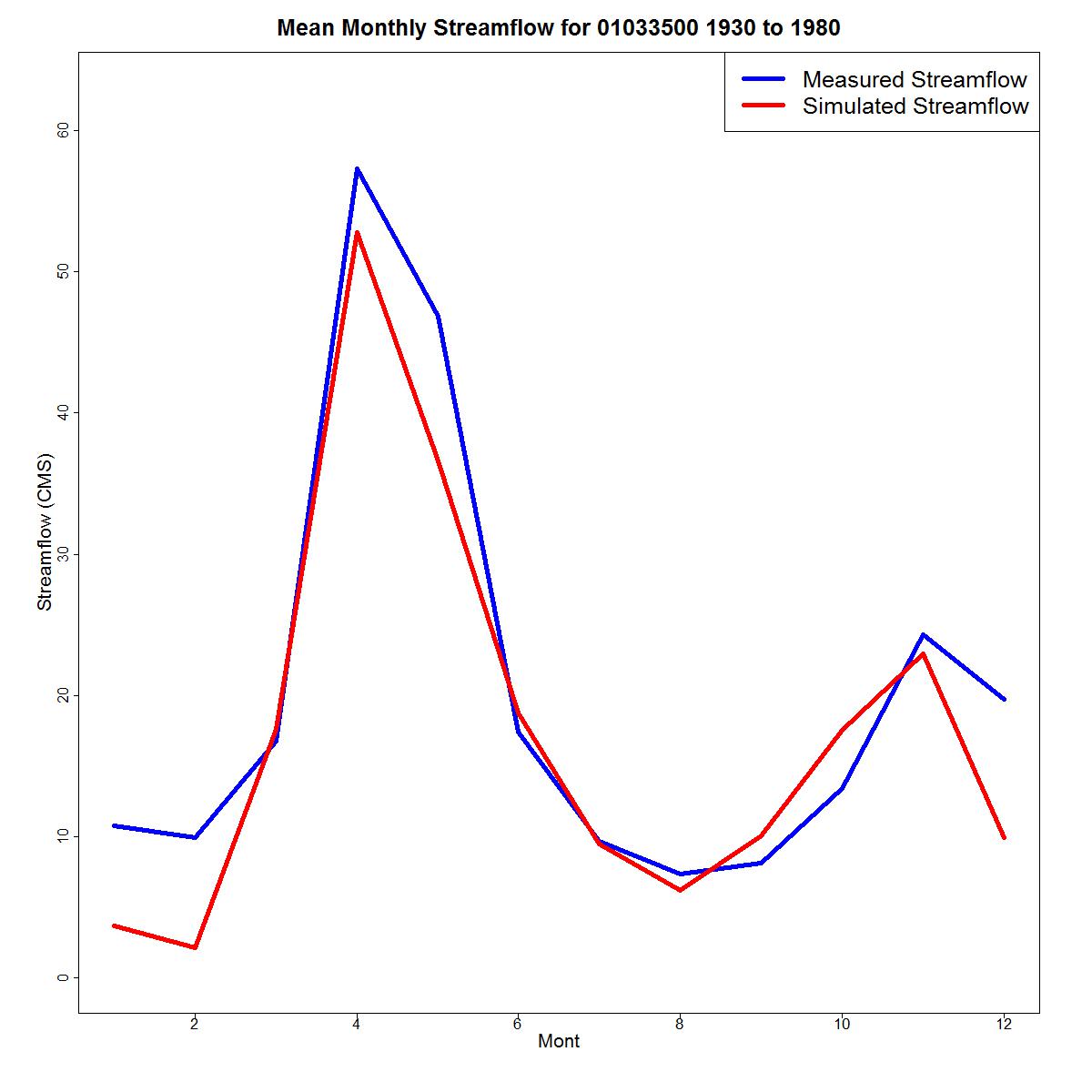


Figure XII. Annual simulated and measured streamflow for the streamgage 01033500 for the period of record (1930-1980).

Figure XIII. Mean monthly simulated and measured streamflow for the streamgage 01033500 for the period of record (1930-1980).

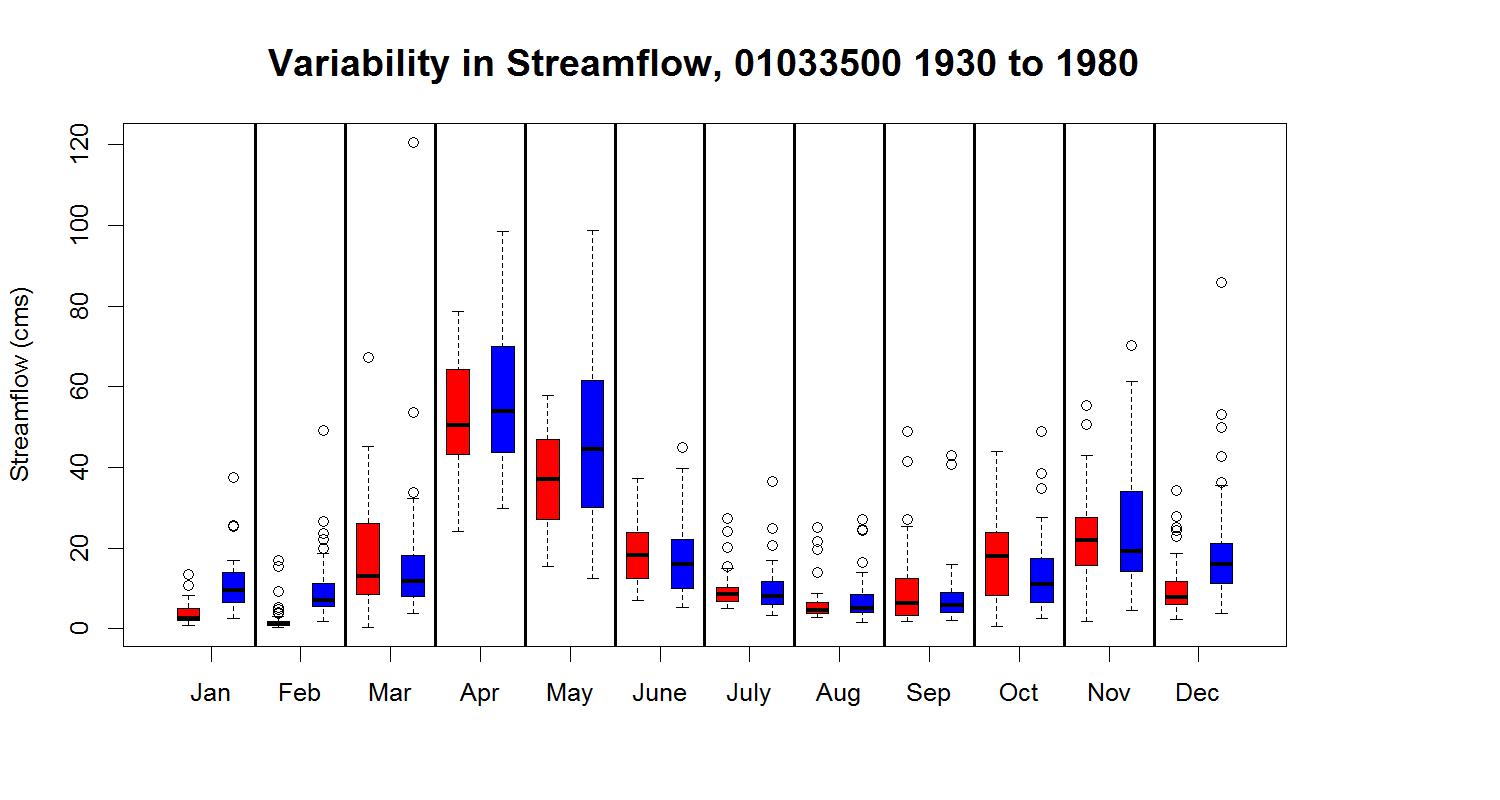


Figure XIV. Boxplots for mean monthly simulated and measured streamflow for the streamgage 01033500 for the period of record (1930-1980).

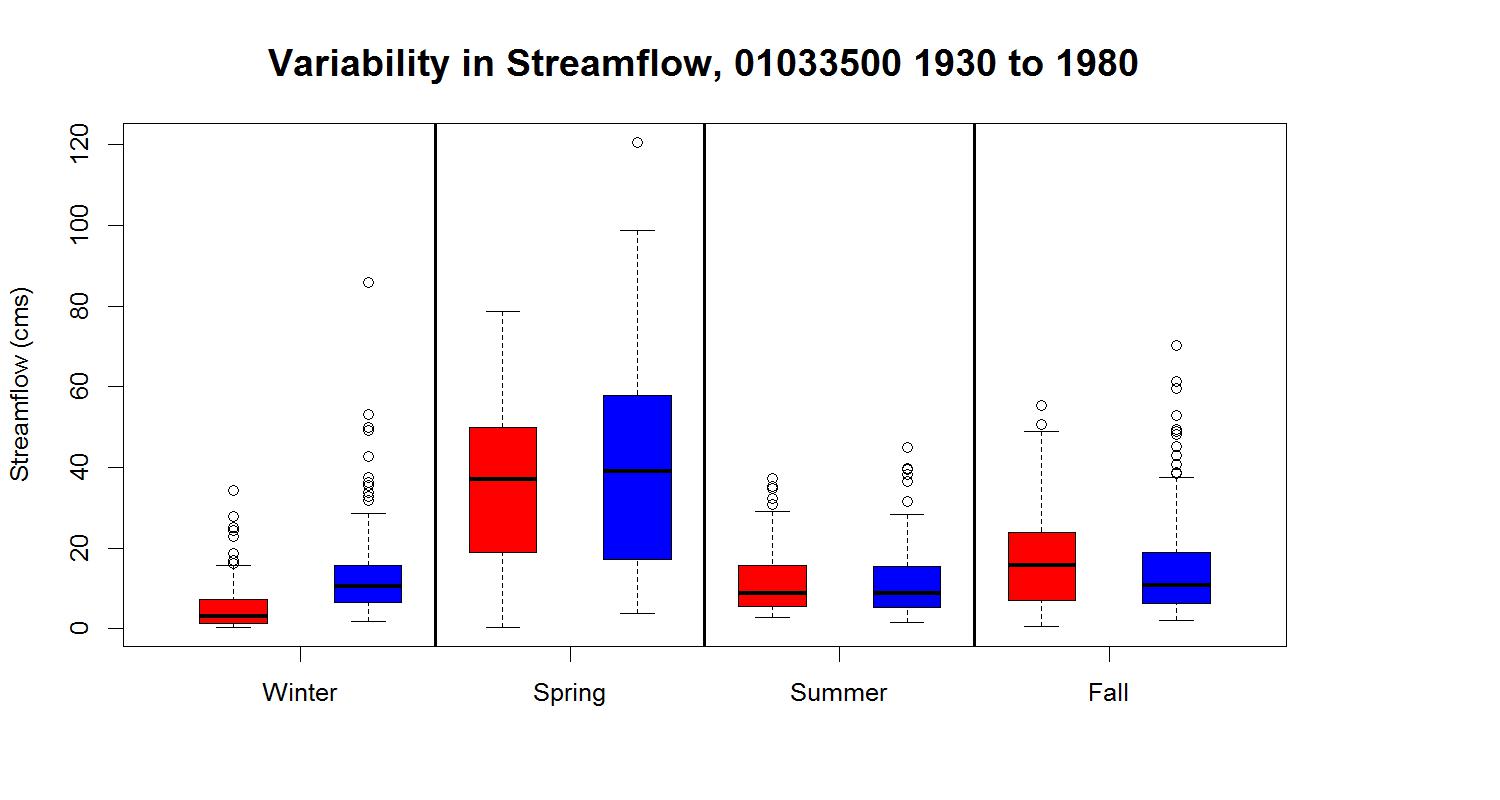


Figure XV. Boxplots for mean seasonal simulated and measured streamflow for the streamgage 01033500 for the period of record (1930-1980).

**Water Balance Model (Future Conditions)**

Input Ports (Figure XVI):

* *Calibration POR:* Beginning of period of record for calibration for current conditions (String,YYYY,YYYY).
* *Evaluation POR*: End of period of record for evaluation for current conditions (String, YYYY,YYYY).
* *Model\_info*: The python dictionary with the model run file information.

Output Ports:

* *None –* The module writes calibration outputs to several places. Performance statistics, and outputs are written to the model run file. For current conditions data used for model calibration, model results and calibration statistics are written to the model info file in the *Base* folder, with graphical summaries and output also written to the Output folder for the Maurer GSD dataset. For downscaled data simulated with the parameters from the Maurer GSD calibration, MWBM outputs for each individual GCM/run are written to each GCM/run Output folder. Each scenario is summarized in graphics in the scenario folder, with graphics from multiple scenarios written to the CMIP3 or CMIP5 folder.

This module is analogous to the MWBM Current Conditions module. The MWBM is calibrated for current conditions using the Maurer GSD for calibration periods entered by the user, and simulated for the evaluation periods using the calibrated parameters. The calibrated parameter set is used to simulate the MWBM for future conditions with the datasets retrieved from the GDP. A text file of model output is written to the *OUTPUT* folder of each individual GCM/run. The results are summarized graphically at the scenario folder (Figures XVII-XX) and dataset folder (CMIP3 or CMIP5) (Figures XXI-XXII).

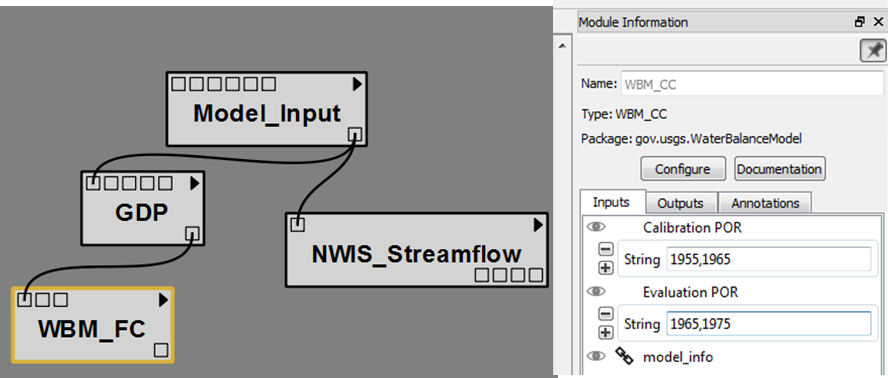


Figure XVI. Input Ports for the WBM\_FC module.

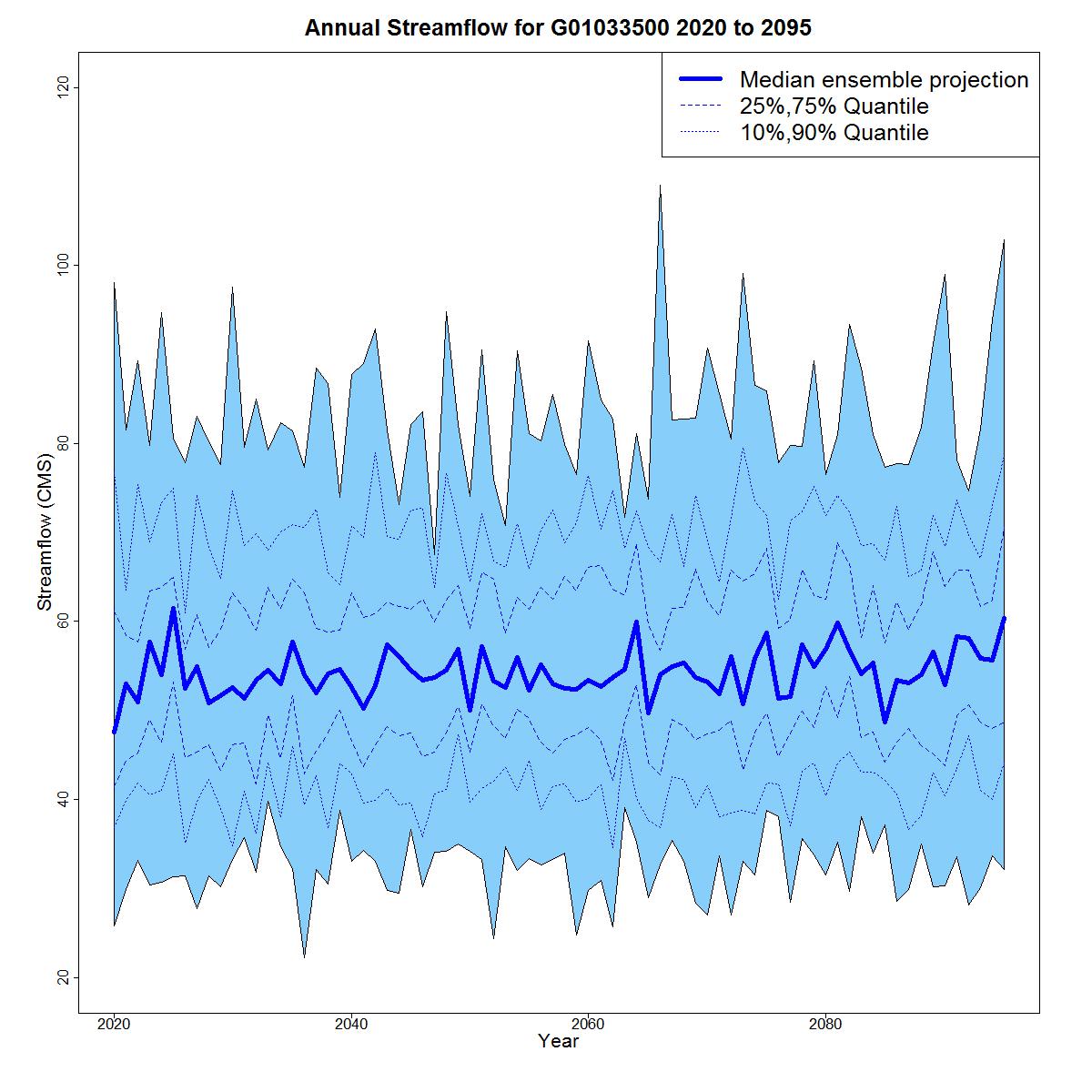


Figure XVII. Annual simulated streamflow for the A2 scenario, 2020-2095, for USGS stream gage 01033500.

.

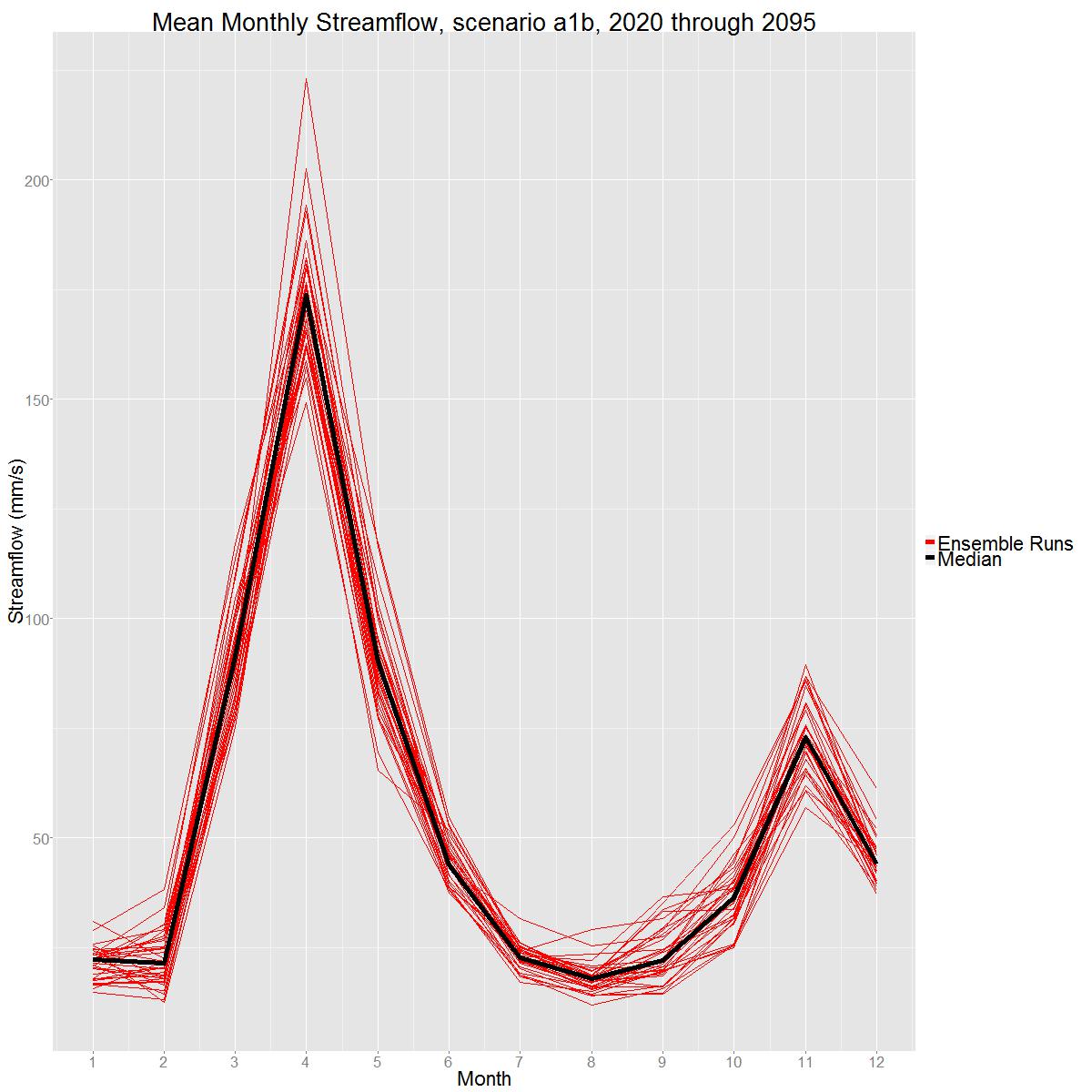


Figure XVIII. Mean monthly simulated streamflow for the A1B scenario, 2020-2095, for USGS stream gage 01033500.

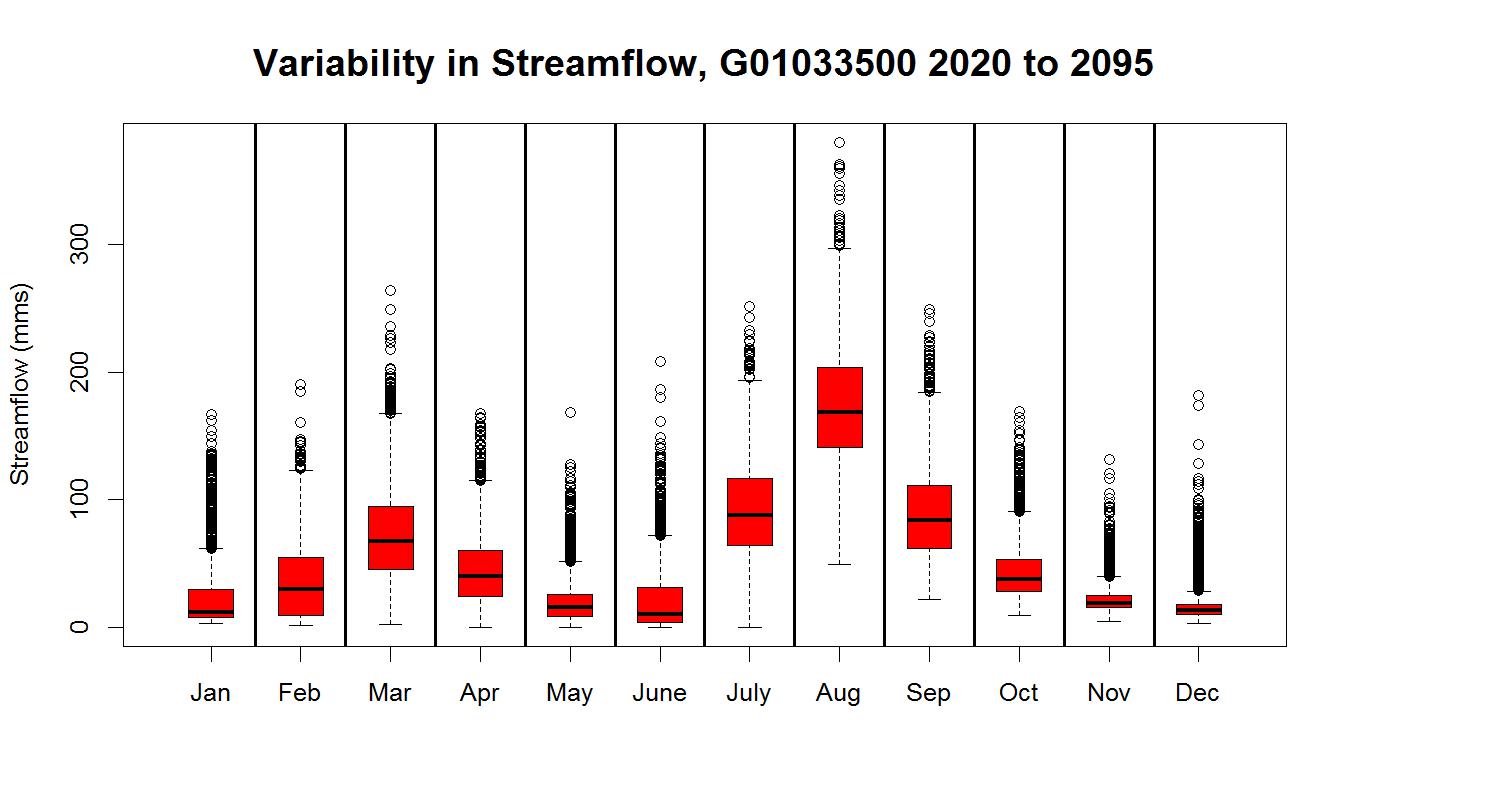


Figure XIX. Variability in monthly and streamflow for the A1B scenario, 2020-2095, for USGS stream gage 01033500.

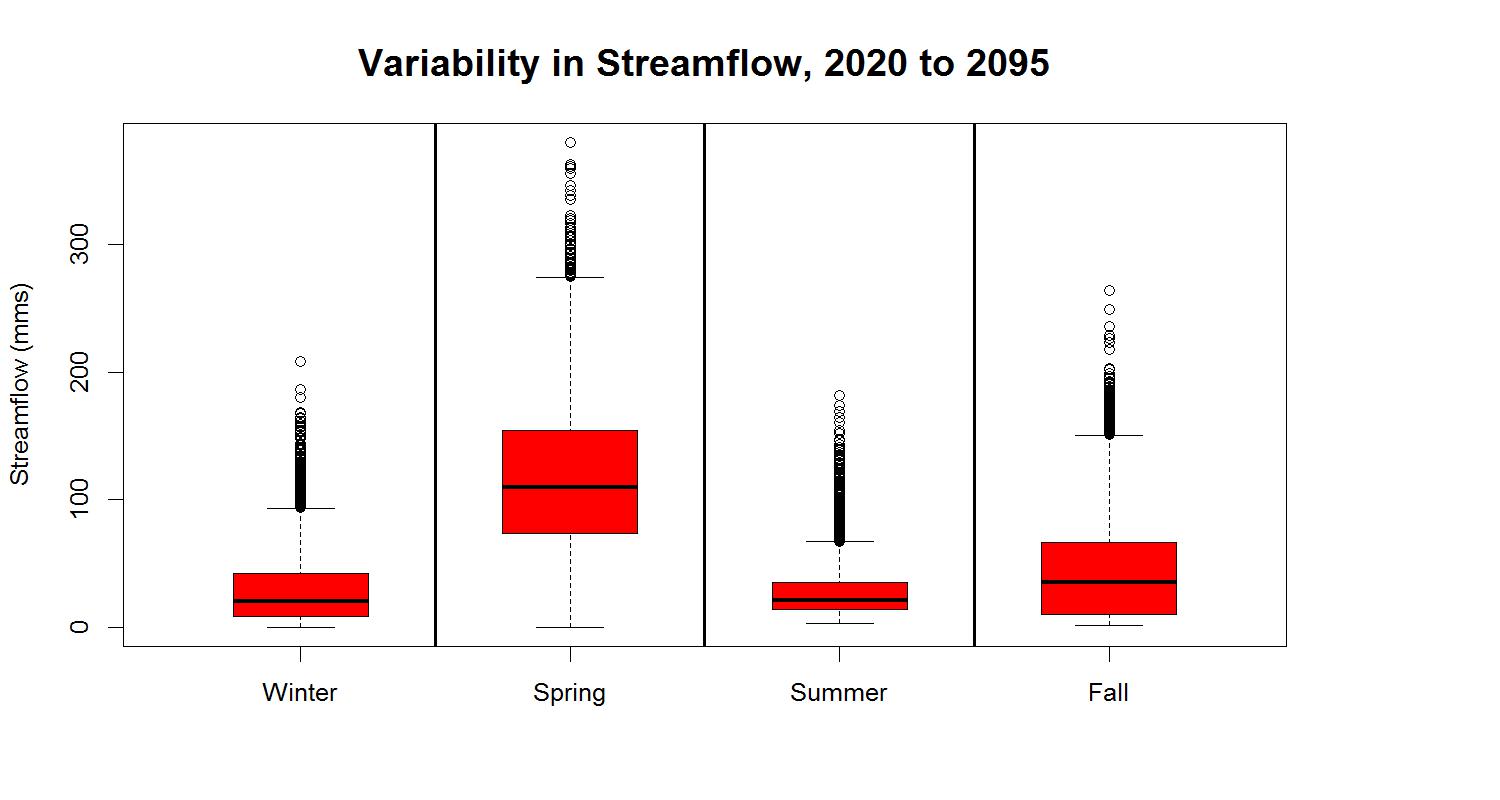


Figure XX. Variability in seasonal streamflow for the A1B scenario, 2020-2095, for USGS stream gage 01033500.

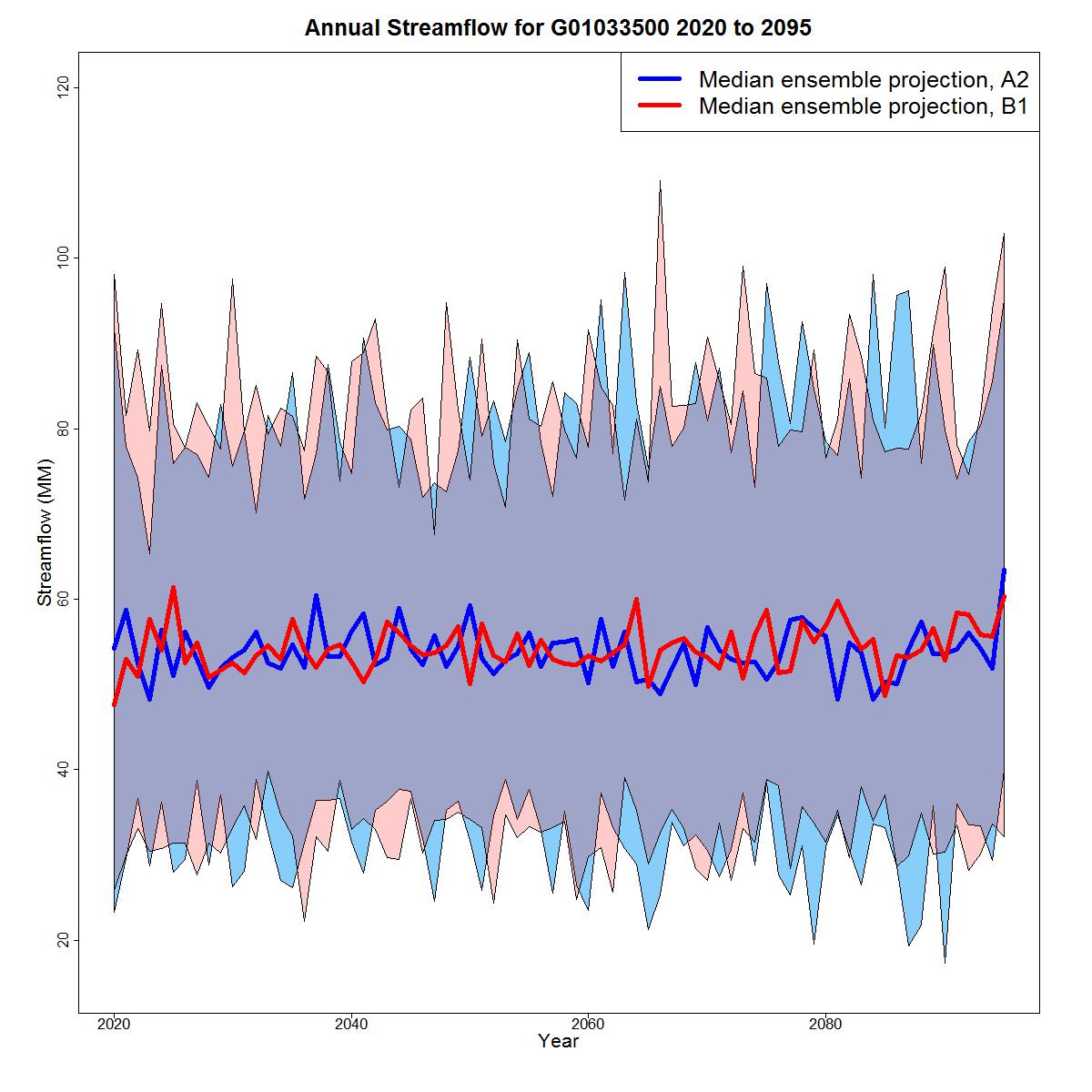


Figure XXI. Annual simulated streamflow for the A2 and B1 scenarios, 2020-2095, for USGS stream gage 01033500.

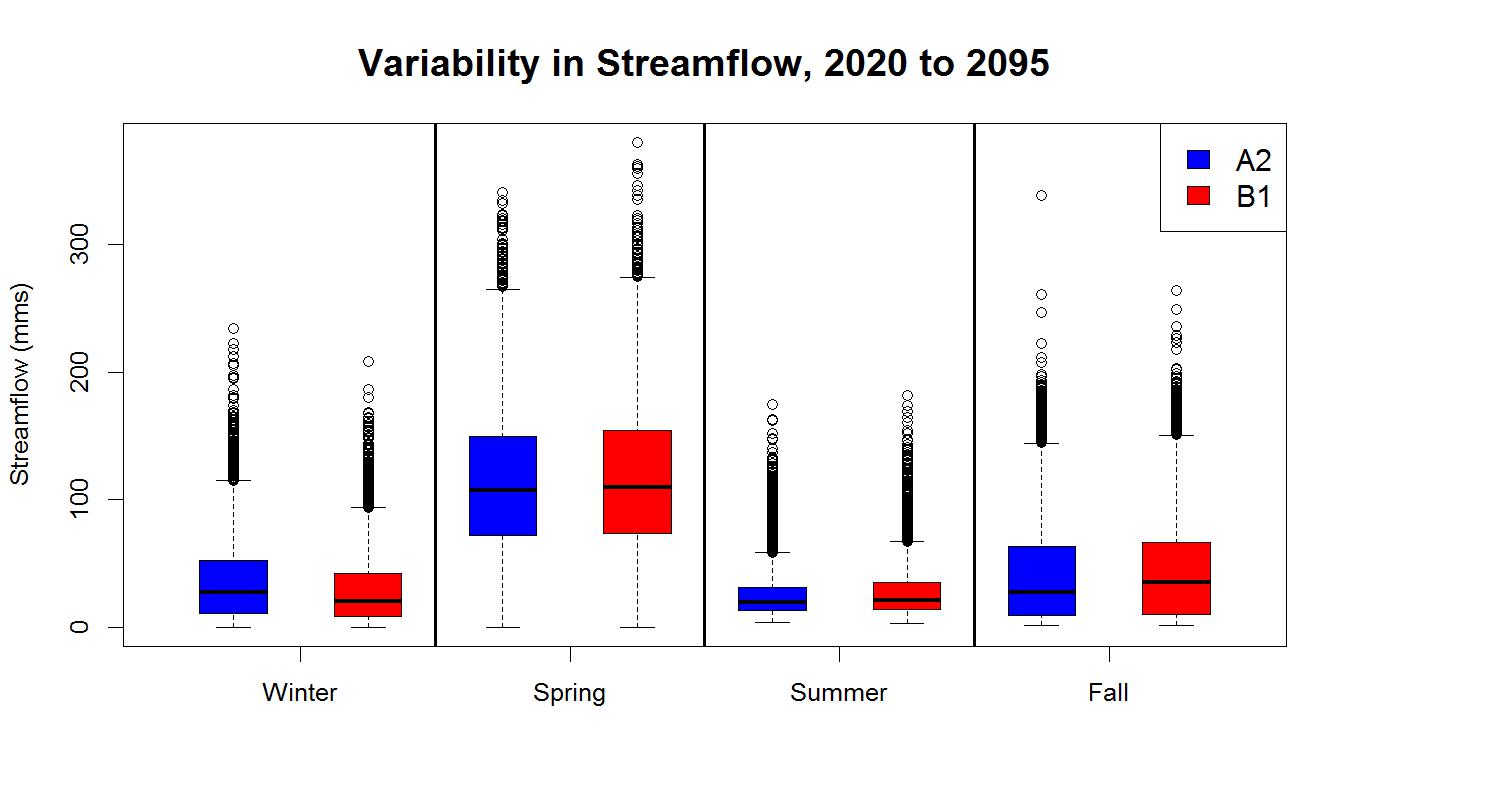


Figure XXII. Variability in seasonal streamflow for the A2 and B1 scenarios, 2020-2095 for USGS stream gage 01033500.

**Appendix Literature Cited**

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

Blodgett, D.L., 2013. The U.S. Geological Survey Climate Geo Data Portal: an integrated broker for climate and geospatial data. U.S. Geological Survey Fact Sheet 2013–3019, 2 p., http://pubs.usgs.gov/fs/2013/3019.

Duan, Q., Sorooshian., S., and Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology* 158:265–284.

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P., and Nijssen, B. 2002. A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States, *Journal of Climate*15:3237-3251.

McCabe, G.J., and Markstrom, S.L., 2007. A monthly water-balance model driven by a graphical user interface. U.S. Geological Survey Open File Report 2007-1088, 6 p.

Nash, J. E., and Sutcliffe, J.V., 1970. [River flow forecasting through conceptual models, part I: A discussion of principles](http://dx.doi.org/10.1016/0022-1694(70)90255-6). [*Journal of Hydrology*](http://en.wikipedia.org/wiki/Journal_of_Hydrology) 10:282–290.

PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 4 Feb 2004.

Thornton, P.E., Thornton, M.M, Mayer, B.W., Wilhelmi, N., Wei, Y., Devarakonda, R., and Cook, R.B., 2014, Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2. Data set. Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. <http://dx.doi.org/10.3334/ORNLDAAC/1219>.

U.S. Bureau of Reclamation, 2011. West-Wide Climate Risk Assessments: Bias-Corrected and Spatially Downscaled Surface Water Projections. Technical Memorandum No. 86-68210-2011-01, prepared by the U.S. Department of the Interior, Bureau of Reclamation, Technical Services Center, Denver, Colorado. 138 pp.

U.S. Bureau of Reclamation, 2014. Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections: Release of Hydrology Projections, Comparison with preceding Information, and Summary of User Needs. Prepared by the U.S. Department of the Interior, Bureau of Reclamation, Technical Services Center, Denver, Colorado. 110 pp.