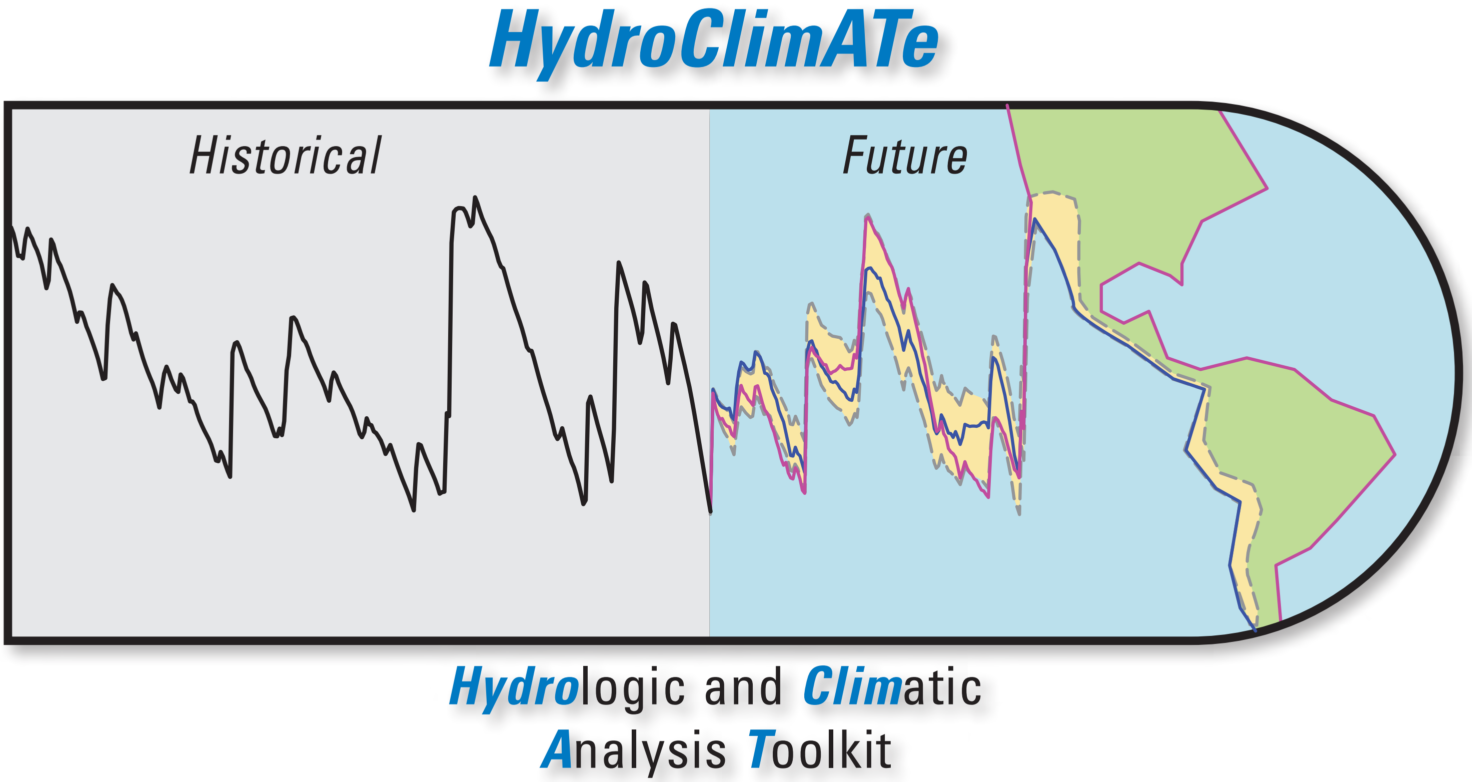
USGS
science for a changing world

Groundwater Resources Program

HydroClimATe—Hydrologic and Climatic Analysis Toolkit

Chapter 9 of
Section A, Statistical Analysis
Book 4, Hydrologic Analysis and Interpretation



Techniques and Methods 4–A9

U.S. Department of the Interior
U.S. Geological Survey

Cover. HydroClimATe—Hydrologic and Climatic Analysis Toolkit

HydroClimATe—Hydrologic and Climatic Analysis Toolkit

By Jesse E. Dickinson, Randall T. Hanson, and Steven K. Predmore

Chapter 9 of
Section A, Statistical Analysis
Book 4, Hydrologic Analysis and Interpretation

Groundwater Resources Program

Techniques and Methods 4–A9

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2014

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications,
visit http://www.usgs.gov/pubprod

To order this and other USGS information products, visit http://store.usgs.gov

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Suggested citation:
Dickinson, J.E., Hanson, R.T., and Predmore, S.K., 2014, HydroClimATe—Hydrologic and climatic analysis toolkit: U.S. Geological Survey Techniques and Methods 4–A9, 49 p, *http://dx.doi.org/10.3133/tm4a9*.

ISSN (online) 2328-7055

Contents

Contents—Continued

Figures

Figures—Continued

Table

Abbreviations

ACF	autocorrelation function
AIC	Akaike information criteria
AICc	corrected Akaike information criteria
AR	autoregression
DFT	discrete Fourier transform
CCF	cross-correlation function
d.o.f.	degrees of freedom
ENSO	El Niño Southern Oscillation
EOF	empirical orthogonal function
FPE	Akaike final prediction error
GIS	Geographic Information System
HydroClimATe	Hydrologic and Climatic Analysis Toolkit
MEI	Multivariate ENSO Index
MEM	maximum entropy method
MTM	multi-taper method
NWIS	National Water Inventory System
PCs	principal components
RCs	reconstructed components
SPI	standardized precipitation index
SSA	singular spectrum analysis
T-EOFs	temporal empirical orthogonal functions
T-PCs	temporal principal components
WOSA	Welch overlapping segment analysis

Acknowledgments

The authors thank Dr. Jason Gurdak at San Francisco State University and Jesus Gomez of the U.S. Geological Survey for their reviews of the software and documentation. The example problems are based on homework sets for the Objective Analysis in the Atmospheric and Related Sciences course presented by Dr. Christopher Castro in the Department of Atmospheric Sciences at the University of Arizona. The synthetic time-series data in the harmonic analysis example were created by Dr. David W. J. Thompson at Colorado State University. Course notes by Dr. Christopher Castro and Dr. Dennis Hartman at the University of Washington and Dr. David Meko at the University of Arizona were especially helpful in the programming of the analysis routines and in the preparation of this report.

Brad Garner of the U.S. Geological Survey Arizona Water Science Center provided instrumental guidance on designing the software and the user interface.

HydroClimATe—Hydrologic and Climatic Analysis Toolkit

By Jesse E. Dickinson, Randall T. Hanson, and Steven K. Predmore

Abstract

The potential consequences of climate variability and climate change have been identified as major issues for the sustainability and availability of the worldwide water resources. Unlike global climate change, climate variability represents deviations from the long-term state of the climate over periods of a few years to several decades. Currently, rich hydrologic time-series data are available, but the combination of data preparation and statistical methods developed by the U.S. Geological Survey as part of the Groundwater Resources Program is relatively unavailable to hydrologists and engineers who could benefit from estimates of climate variability and its effects on periodic recharge and water-resource availability. This report documents HydroClimATe, a computer program for assessing the relations between variable climatic and hydrologic time-series data. HydroClimATe was developed for a Windows operating system. The software includes statistical tools for (1) time-series preprocessing, (2) spectral analysis, (3) spatial and temporal analysis, (4) correlation analysis, and (5) projections. The time-series preprocessing tools include spline fitting, standardization using a normal or gamma distribution, and transformation by a cumulative departure. The spectral analysis tools include discrete Fourier transform, maximum entropy method, and singular spectrum analysis. The spatial and temporal analysis tool is empirical orthogonal function analysis. The correlation analysis tools are linear regression and lag correlation. The projection tools include autoregressive time-series modeling and generation of many realizations. These tools are demonstrated in four examples that use stream-flow discharge data, groundwater-level records, gridded time series of precipitation data, and the Multivariate ENSO Index.

Introduction

The potential response of water resources to climate variability is one of the most vital issues for sustainability in the United States (Gleick and Adams, 2000; Lins and others, 2010) and around the world (Green and others, 2011; Taylor and others, 2012; Treidel and others, 2012). As water resources become fully allocated, the responses of water availability to climate variability become especially important for long-term planning and resource management (Lins and others, 2010; Hanson and others, 2012). Climate variability represents reversible and periodic changes in the global weather systems that occur over periods of a few years to several decades or longer. Oceanic-atmospheric phenomena, such as the El Niño-Southern Oscillation (ENSO), are important predictors of precipitation in many regions around the globe (Ropelewski and Halpert, 1987; Dai and Wigley, 2000). For example, previous studies have identified evidence that the warm (El Niño) phase of ENSO is related to increased winter precipitation in the southwestern U.S. and decreased precipitation in the northwestern U.S. (Redmond and Koch, 1991). The opposite pattern characterizes the cool (La Niña) phase of ENSO; precipitation is lower in the Southwest and greater in the Northwest (Redmond and Koch, 1991; Livezey and others, 1997). The exploration of how these cycles affect the use and movement of water through the hydrosphere can provide fundamental insight concerning how to manage and sustain the resources.

Climate variability influences the timing and volumes of inflow and outflow components of hydrologic budgets (for example Hanson and others, 2003; Gurdak and others, 2007, 2009; Faunt, 2009; Hanson and others, 2009, 2012; Campbell and Coes, 2010; Heilweil and Brooks, 2011). Influences on surface water include the distribution, timing, and amount of runoff-producing precipitation, which leads to changes in surface-water deliveries in agricultural and municipal areas, as well as streamflow for ecosystems. The influences on groundwater include natural changes in recharge, discharge, as well as variability in groundwater withdrawals that can coincide with the availability of surface-water deliveries.

One challenge facing water resource managers is to predict how groundwater systems will recover after a periodic stress, such as drought, so that they can properly assess the vulnerability of human water-supply systems and riparian habitats to climate variability and change. Knowledge of how climatic variability can influence hydrologic inflows and outflows could prove to be essential for achieving sustainable water resources, especially in regions of limited water availability. Hydrologic models that are commonly used for assessing the effects of water-management strategies have often used long-term trends in inflow and outflow components. Simulations of long-term trends in these components can depict the general state and movement of water, but variations in the inflows and outflows from mean conditions can result in extremely dry and wet conditions. Temporal variability about the long-term trends can exacerbate the effects of extremely dry periods, leading to disconnected groundwater and surface water to the extent that ecosystems potentially are not able to recover. High frequency, short-term increases in precipitation can produce flashy runoff, but infiltration from short-term events does not always result in significant or lasting changes in subsurface water storage (Bakker and Nieber, 2009). Conversely, persistent periods of flashy runoff can be captured and stored in reservoirs or artificial recharge facilities. Because groundwater is often used to reduce the effects of drought, temporal variability in precipitation in a long-term declining trend and patterns of human use could combine to have dramatic consequences for human as well as natural systems (Alley and others, 1999).

Recently, hydrologic models have included more accurate representations of interannual and interdecadal variations in the inflow and outflow components, as well as more complete coupling of atmospheric and hydrologic processes (for example, Faunt, 2009; Clark and others, 2011; Hanson and others, 2012), which could improve the accuracy of the simulated effects of short- and long-term water management strategies. Such assessments become increasingly important as water resources are fully allocated and variations in inflows or outflows from the long-term conditions stress the water supplies for human uses and ecosystem functions.

The assessment of the effects of climate variability on hydrologic systems often begins with an analysis of long-term hydrologic records for features such as temporal trends and variability (Hanson and others, 2004; Kumar and Duffy, 2009; Lins and others, 2010). Often, such assessments are based on an objective statistical analysis of these records to identify relations among indicators of climate variability and hydrologic conditions. A systematic approach for identifying such features in hydrologic time-series data was established as part of the U.S. Geological Survey Southwest Ground Water Resource Project (Leake and others, 2000; Hanson and others, 2004; Hanson and others, 2006). Relations between climatic variability and multiple types of hydrologic time series (groundwater levels, streamflow, precipitation, and tree-ring data) in the southwestern United States (Hanson and others, 2006), as well as a statistical basis for delineating the climatic and anthropogenic variations in hydrologic systems, lead to methods for estimating climate-controlled recharge to alluvial aquifer systems (Dickinson and others, 2004). Projections of future hydrologic conditions based on objective analysis of long-term records were used to simulate wet and dry periods in streamflow and precipitation in simulations of groundwater flow in the Santa Clara—Calleguas Basin in central California (Hanson and others, 2003; Hanson and Dettinger, 2005).

Purpose and Scope

The purpose of this report is to document how to use the program Hydrologic and Climatic Analysis Toolkit (HydroClimATe), which automates the use of several objective methods for assessing relations among one or more sets of data that vary in time and space with climate variability and variability in hydrologic time-series data. Although the software was written for the purpose of identifying relations between hydrologic and climatic time series, the available tools and methods can be applied to many other types of time series generated by physical, biological, economic, or social processes. The methods include standardization, detrending, regression and correlation, Fourier analysis, maximum entropy method (MEM), singular spectrum analysis (SSA), empirical orthogonal function (EOF) analysis, and autoregressive time series modeling. These tools have been used extensively to identify relations among climatic indicators and meteorological and hydrologic data to hydrologic conditions in previous HydroClimATeic investigations (for example, Dettinger and others, 1995a; Dettinger and Diaz, 2000; Dickinson and others, 2004; Hanson and others, 2004, 2006, 2009; Gochis and others, 2007a, b; Kumar and Duffy, 2009). The software combines these methods in order to provide a set of analysis tools that can be readily implemented for processing hydrologic time-series data. The software includes all of the methods for assessing relations between climatic and hydrologic time series described by Hanson and others (2004) and implemented by Hanson and others (2003). This report describes how to use the tools in the software, but it is recommended that users become familiar with each technique by review of the literature on each technique. This report is intended to describe how to operate the software and provide some context of where and when each technique can be useful, and is not intended to serve as a comprehensive review on these techniques. The implementation of the techniques in the software is described in appendix A.

HydroClimATe includes tools for (1) identifying responses of hydrologic systems to climate variability; (2) quantifying statistical relations between multiple time series and climate indices, such as the Multivariate ENSO Index (MEI; Wolter and Timlin, 2011), and (3) projecting hydrologic time-series data by using time-series models and spectral analysis. The software consists of a graphical user interface that is executable in Windows operating system with the .NET Framework version 4.0. Software inputs can be any long-term time-series data, such as groundwater levels, streamflow, precipitation, tree-ring data, air temperature, and climate indices. Other types of time-series data, such as economic data, could be applicable. However, the methods of analyses demonstrated in this report are for climatic and hydrologic time-series data. Software output can be exported to files that are read by a text editor, Microsoft Excel®, or geographic information system (GIS) software. Example analyses are presented for assessing relations between global climate indices and hydrologic time series for sites in the southwestern U.S.

System Requirements

HydroClimATe requires the Microsoft Windows XP, Windows Vista, Windows 7, or Windows 8 operating systems. The software may run on future versions of Microsoft Windows products. Installation of the Microsoft .NET Framework Version 4.0 that is distributed with Microsoft Windows 7 is required in order to run HydroClimATe. If Microsoft .NET Framework Version 4.0 is not installed, it is free and can be obtained from Microsoft at *http://www.microsoft.com/downloads*. An internet connection is required to import data from the USGS National Water Inventory System (NWIS).

Installation

HydroClimATe does not require installation. However, the installation of the Microsoft .NET Framework Version 4.0 is required.

Approach

The software automates the methods of processing and analyzing climatic and hydrologic time series described by Hanson and others (2004). The toolkit includes the following steps of analysis:

1. Data acquisition and preprocessing.
    a. Retrieval from an ASCII file, Excel® file or the USGS NWIS database.
    b. Interpolation, standardization, and transformation by a cumulative departure.
2. Analyses.
    a. Trend removal by differencing and curve fitting.
    b. Frequency analysis by using Fourier analysis, MEM, and SSA.
    c. Principal Component Reconstructions including inverse transformations.
    d. Spatial/temporal modal decomposition by using EOF analysis.
    e. Statistical Estimation using correlation, lag, covariance, and other multi-series relations.
3. Projections of time series with autoregressive (AR) time series models.

How to Use this Report

This report describes how to use the software features that complete the sets of analysis described above. This report describes each of the tools in the software in separate sections. The steps of analysis can vary depending on the final goal of the analysis and the groups of time series that are being analyzed together. Some types of data require preprocessing before they can be analyzed by the spectral or spatial and temporal analysis tools, or can be projected in time. For example, precipitation, streamflow, tree-ring data, and climate indices could need to be transformed into cumulative departure to be compatible with analyses that include changes in groundwater levels. Data often require some standardization in order to make comparisons of different data types or of data from locations at which the physical driving processes differ. For example, the Standardized Precipitation Index (SPI; McKee and others, 1993) is useful for identifying wet and dry periods from different regional climates and for comparing data that are not normally distributed, such as streamflow and precipitation data. Standardization is also helpful for identifying anomalies, which can be analyzed by the spectral analysis and the EOF tools. The user can proceed to the section of the report that pertains to the tool and step of analysis that is of interest.

The data for the examples described in this report are also provided in the release package so that the user can perform each analysis step-by-step to recreate the results presented from the analysis with HydroClimATe. The example sections provide specific instructions on how to import data, the tools and options that were used to generate the results, and a summary of the important features in the results. The example problems provide a small subset of the potential uses of the software. Other examples of how these tools have been used in hydrologic and climatic studies are referenced throughout this report.

Following the introduction, which describes the overall purpose and scope of the report and the software, this report describes the user interface of the software, the computer requirements for running the software, and how to import and export vector and matrix data from and to ASCII and Excel® formats. The names of labels in the user interface are emphasized in the text of the report by bold lettering to help the user identify different parts of the user interface. The next section describes how to operate perform different statistical methods available in the user interface. The example section describes four different examples, how to replicate each example, and a summary of the important results from each example. The appendix includes descriptions of the mathematical techniques used by the software.

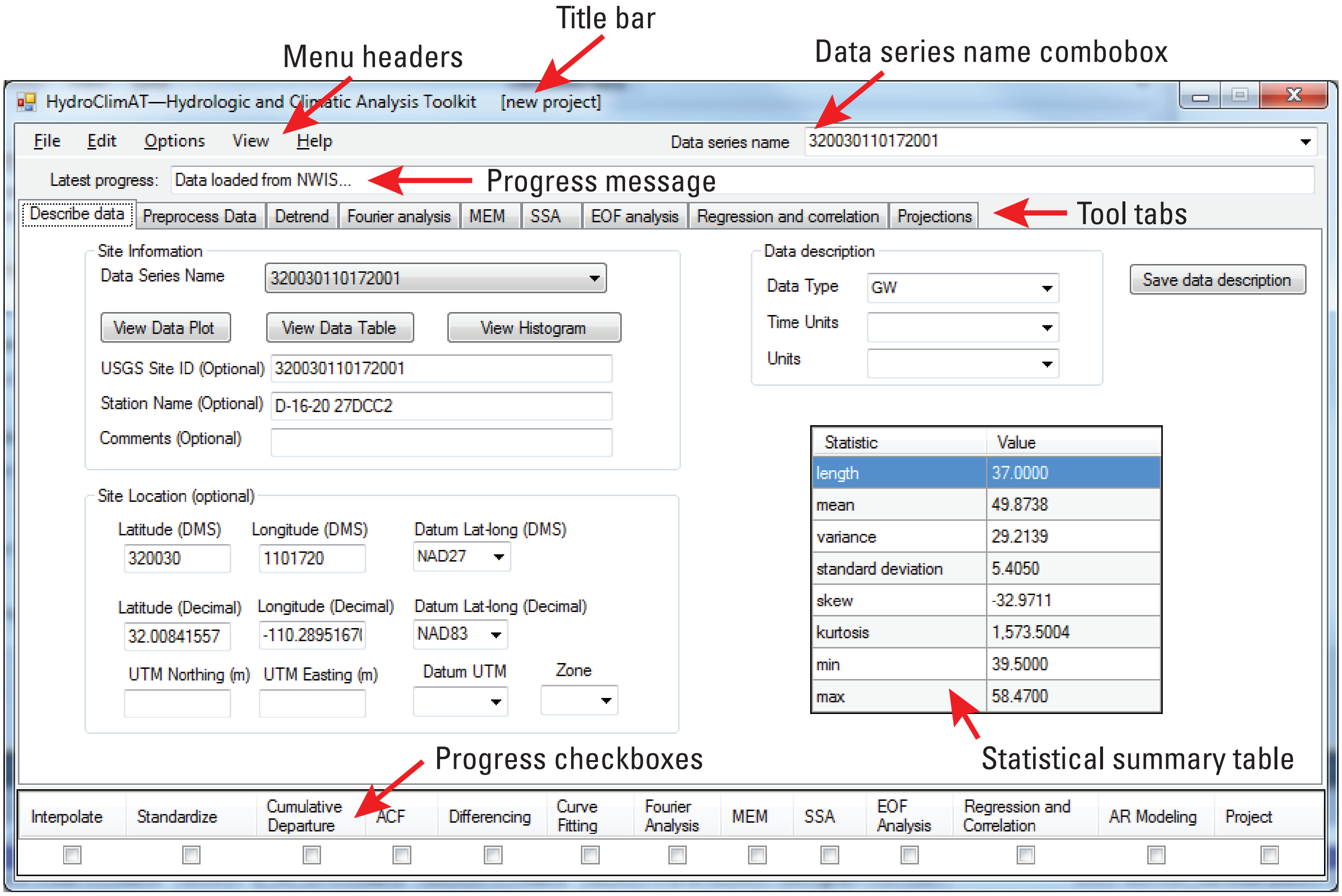
User Interface

Each of the steps of analysis is completed by the selection of options that are displayed on a user interface (fig. 1). The user interface consists of a main form that contains tabs (**Describe data**, **Preprocess data**, **Detrend**, **Fourier Analysis**, **MEM**, **SSA**, **EOF analysis**, **Regression and correlation**, **Projections)** that are labeled for the different groups of tools or steps of analysis. The tools for each step of analysis are displayed in the main form by selecting a tab. In general, the most basic steps of analysis are displayed by selecting the tabs that are arranged in the left half of the main form. Other steps that could require some preprocessing of the data are organized in the tabs toward the right side of the main form. At the top of the main form are menu headers for various options. The tabs are organized so that the user can begin with the tabs on the left half and sequentially use the tabs to the right. A list of progress checkboxes at the bottom of the user interface indicates which tools, or steps of processing, have been completed for a data series. The progress checkboxes are intended to give the user a quick summary or review of which steps were performed in an analysis. Additional forms are displayed when the user performs certain actions, such as importing and exporting data or generating plots.

The software is designed to have multiple time series available to the user to compare different data sets through regression and correlation analyses. Only one time series is active at any time while performing an analysis, however. The exception is that two time series can be active during the regression and correlation analysis. The active time series is selected by using the drop-down list at the top of the main form that is labeled **Data series name**. The drop-down list **Data series name** contains a list of all of the data series that the user has imported and, by default, will display the name of the most-recently imported time series.

Most of the tools require some input from the user, and an output time series is produced by pressing a button that is typically labeled **Calculate**, followed by the name of the tool. In general, the **Calculate** button and other buttons for displaying results are not enabled until all required inputs have been provided by the user. Whenever the **Calculate** button is clicked, by default, the program displays a plot of the results, unless the user deactivates **Automatically generate plots** under the menu header **Options** and **Plotting**.

A list of the time series that are stored in the program can be viewed in the **Workspace** form (fig. 2) by selecting the menu header **View** (fig. 1) and **Workspace**. **Workspace** contains a list of the series names that are subsets of a data series.



Data series name 320030110172001

Latest progress: Data loaded from NWIS...

Site Information

Data Series Name 320030110172001

USGS Site ID (Optional) 320030110172001

Station Name (Optional) D-16-20 27DCC2

Comments (Optional)

Site Location (optional)

Latitude (DMS) 320030

Longitude (DMS) 1101720

Datum Lat-long (DMS) NAD27

Latitude (Decimal) 32.00841557

Longitude (Decimal) -110.2895167(

Datum Lat-long (Decimal) NAD83

UTM Northing (m)

UTM Easting (m)

Datum UTM

Zone

Data description

Data Type GW

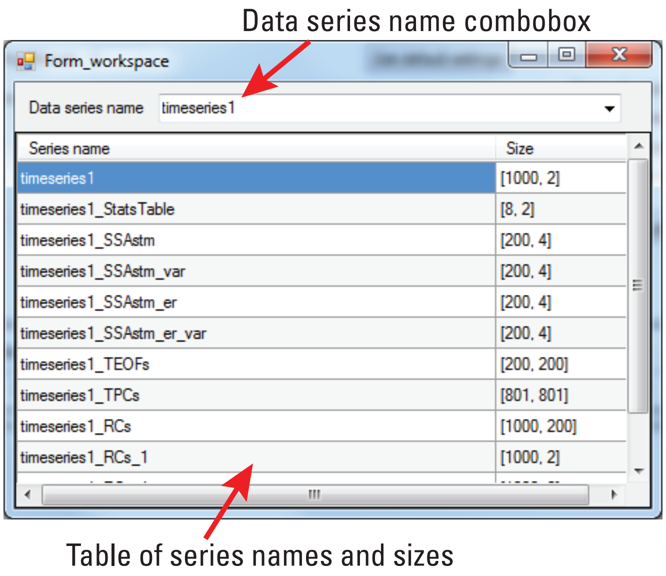
Time Units

Units

Statistic	Value
length	37.0000
mean	49.8738
variance	29.2139
standard deviation	5.4050
skew	-32.9711
kurtosis	1,573.5004
min	39.5000
max	58.4700

Interpolate	Standardize	Cumulative Departure	ACF	Differencing	Curve Fitting	Fourier Analysis	MEM	SSA	EOF Analysis	Regression and Correlation	AR Modeling	Project
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Figure 1. User interface showing tabs and menus.



Data series name combobox

Form_workspace

Data series name: timeseries1

Series name	Size
timeseries1	[1000, 2]
timeseries1_StatsTable	[8, 2]
timeseries1_SSAstm	[200, 4]
timeseries1_SSAstm_var	[200, 4]
timeseries1_SSAstm_er	[200, 4]
timeseries1_SSAstm_er_var	[200, 4]
timeseries1_TEOFs	[200, 200]
timeseries1_TPCs	[801, 801]
timeseries1_RCs	[1000, 200]
timeseries1_RCs_1	[1000, 2]

Table of series names and sizes

Figure 2. Workspace form showing the series names, sizes of the stored data arrays that are subsets of the selected data series, and the root name and extensions of the data series for several outputs.

The data series is selected by a drop-down list labeled **Data series name** at the top of the **Workspace** form. Information for other data series and subset series can be displayed in the **Workspace** form by selecting another data series in the drop-down list.

Each step of analysis calculates a new time series or table that is stored within the software. The name of each new time series or table is based on a root name plus the character “_” and an extension (fig. 2). The root name of each newly-generated time series or table is equal to the name of the active time series. An extension is a short sequence of characters that are appended to the root and separated from the root by the character “_”. By default, an extension is a shortened version of the name of the tool that was just used, but the extension can be modified by the user. For example, the extension for the output from the standardizing tool is “std” and is “trd” for a trend. A new extension is appended to the series name each time a tool is used because the default extension of the most recent tool is appended to the previous extension. For example, a trend of a standardized series can be “root_std_trd” if the root name of the series is “root.” If a new time series is generated that has the name of an existing time series, the previous time series will be overwritten by the new time series. The previous time series can be retained by giving the new time series a different extension.

The time series and spectrum plots include a feature that allows the user to zoom into areas of interest (fig. 3). To zoom in, click on one part of the plot, which is one extent of the zoomed area, and drag and release at the other extent. The plot can be unzoomed by clicking on the circle at the lower left corner of the plot (fig. 3).

Data Requirements

Time-series data are required to be in a vector or matrix format. The vector format can represent a time series in a single dimension or a matrix where two or three dimensional data are looped along a single dimension. The file containing the data either can be in an ASCII file or an Excel® file.

Vector Format for a Single Dimension

An example of a vector of a single dimension is a list of water levels at different times. A vector containing a single dimension is required to have two columns and n rows, where n is the number of values in the time series. The first column contains the time t of the value and the second column contains the value of the time series $a(t)$ at time t. The time values must be a real number. An example is the decimal year 1980.0, which represents the month of January 1980, or 1980.92, which represents the month of December 1980. (fig. 4).

Vector Format for Two or Three Dimensions

An example of a vector of two or three dimensions is a series of maps of precipitation at different times, in which each row is stacked along a single dimension (fig. 5). A vector of two or three dimensions contains a single column that contains the value of the time series $a(x,y)$ or $a(x,y,t)$ at spatial coordinates x and y and map or time t. The matrix format consists of a two-dimensional matrix, usually of spatial values, that repeats n times, where n is the number of maps or time steps. An example is a series of two dimensional matrices that contain precipitation values that repeat for each time step.

Importing Data And Exporting Data

Vector or matrix (gridded) time series data can be imported from a local file or by retrieval from the USGS NWIS. Data can be exported and saved to a local file.

Importing Vector or Matrix Time Series Data from a File

Data in a vector or matrix form can be imported by selecting the name of a local file. To import a vector time series, select the **File** menu and click “Import time series from file.” Then, select either **ASCII file** or **Excel® file** on the menu that appears to the right. This opens a separate form that has options for selecting a file, the separator between values, the encoding of the file, and if the file has column names in a header above each column (figs. 4 and 6). The time-series data must be assigned a name, which, by default, is the name of the loaded file without the file extension. The name can be changed by editing the name in the text box at the bottom of the form.

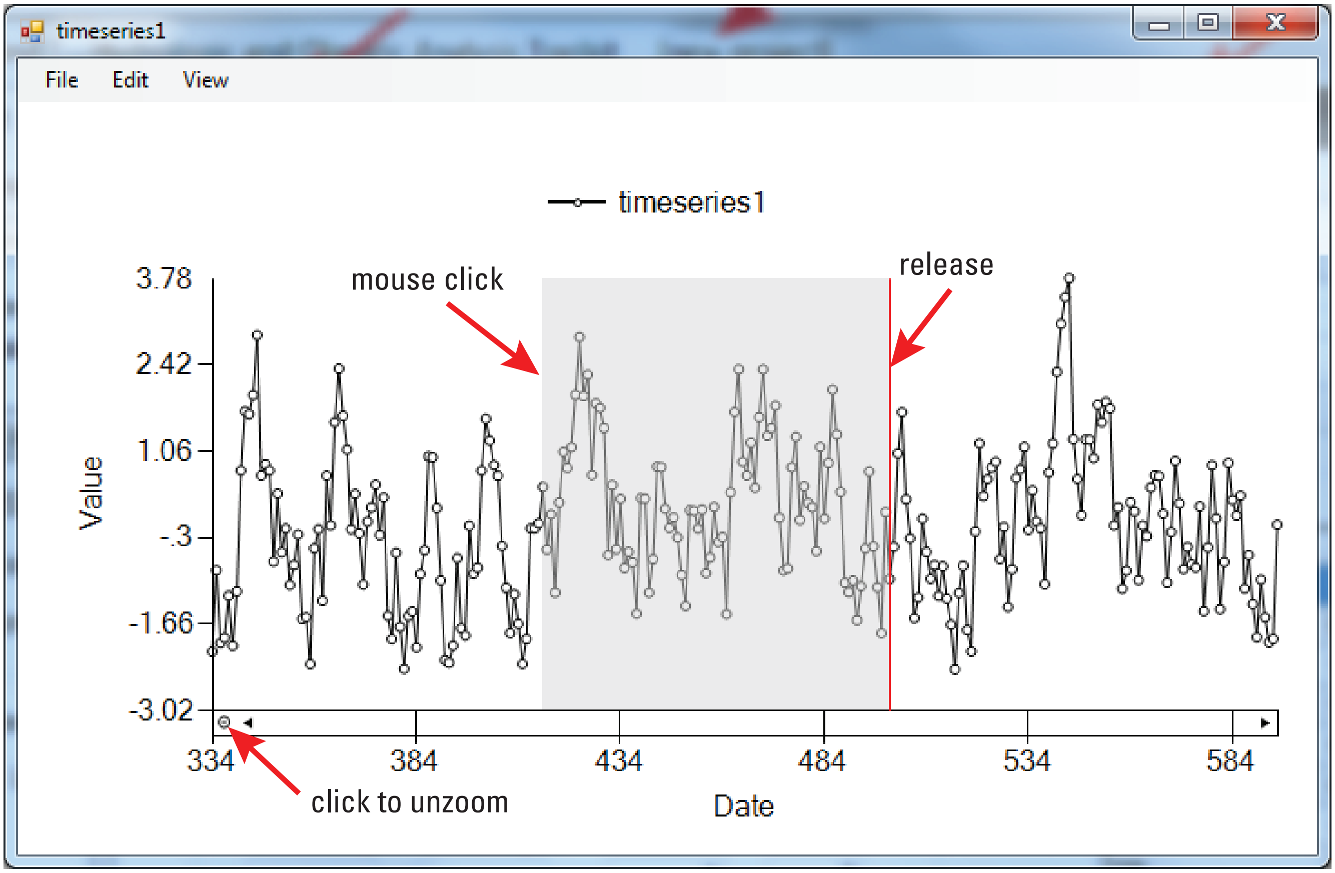
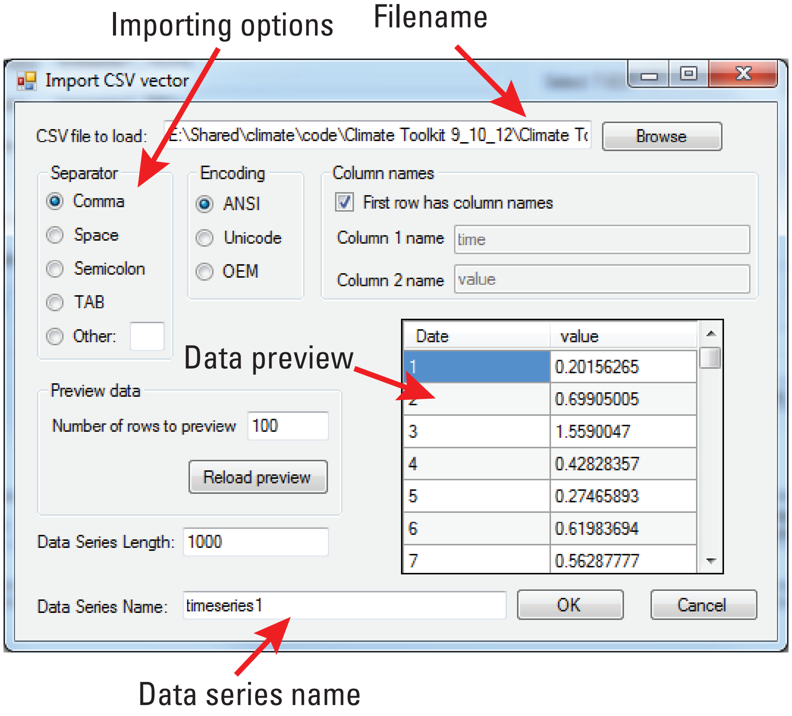


Figure 3. Procedure for zooming in and zooming out of plots.



CSV file to load: E:\Shared\climate\code\Climate Toolkit 9_10_12\Climate Tc

Separator
● Comma
○ Space
○ Semicolon
○ TAB
○ Other:

Encoding
● ANSI
○ Unicode
○ OEM

Column names
☑ First row has column names
Column 1 name: time
Column 2 name: value

Preview data
Number of rows to preview: 100

Data Series Length: 1000

Date	value
1	0.20156265
2	0.69905005
3	1.5590047
4	0.42828357
5	0.27465893
6	0.61983694
7	0.56287777

Data Series Name: timeseries1

Figure 4. Options for importing data from a file in vector format.

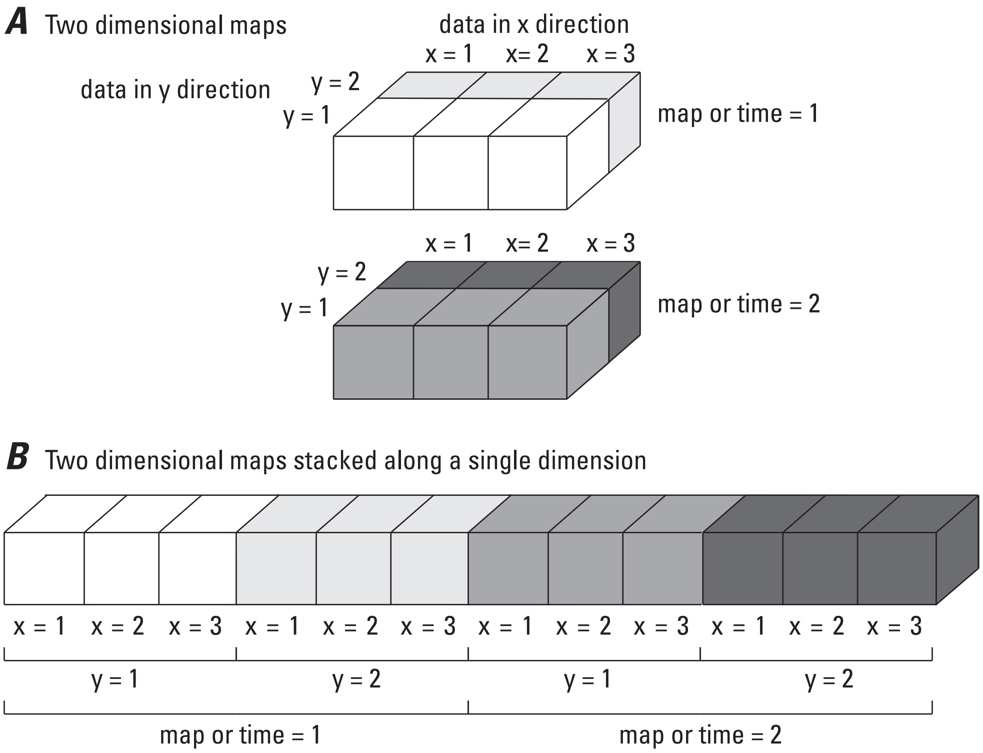
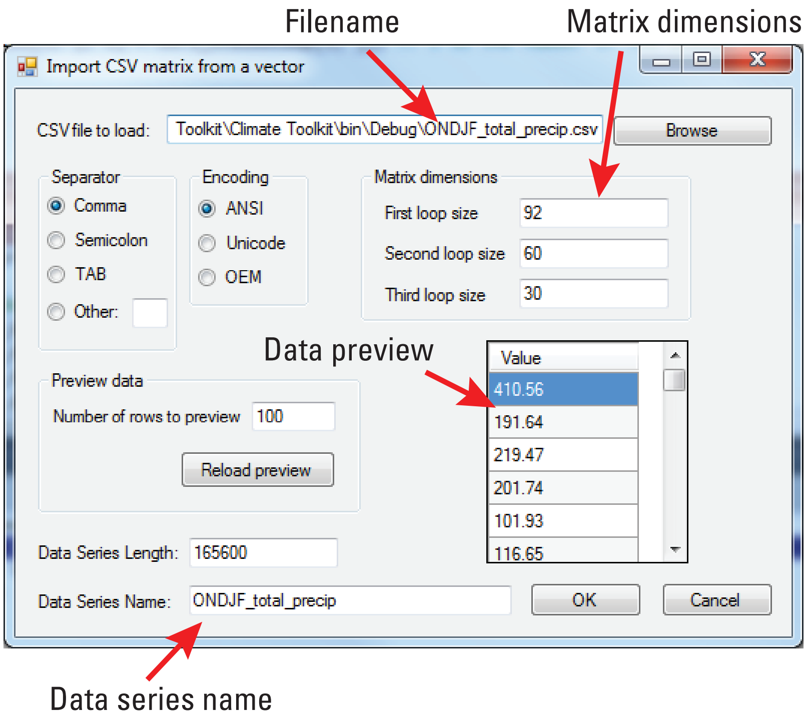


Figure 5. Example of stacking a series of two-dimensional maps into a vector of a single dimension.



Import CSV matrix from a vector

CSV file to load: Toolkit\Climate Toolkit\bin\Debug\ONDJF_total_precip.csv

Separator: ● Comma ○ Semicolon ○ TAB ○ Other:

Encoding: ● ANSI ○ Unicode ○ OEM

Matrix dimensions

First loop size: 92

Second loop size: 60

Third loop size: 30

Preview data

Number of rows to preview: 100

Value
410.56
191.64
219.47
201.74
101.93
116.65

Data Series Length: 165600

Data Series Name: ONDJF_total_precip

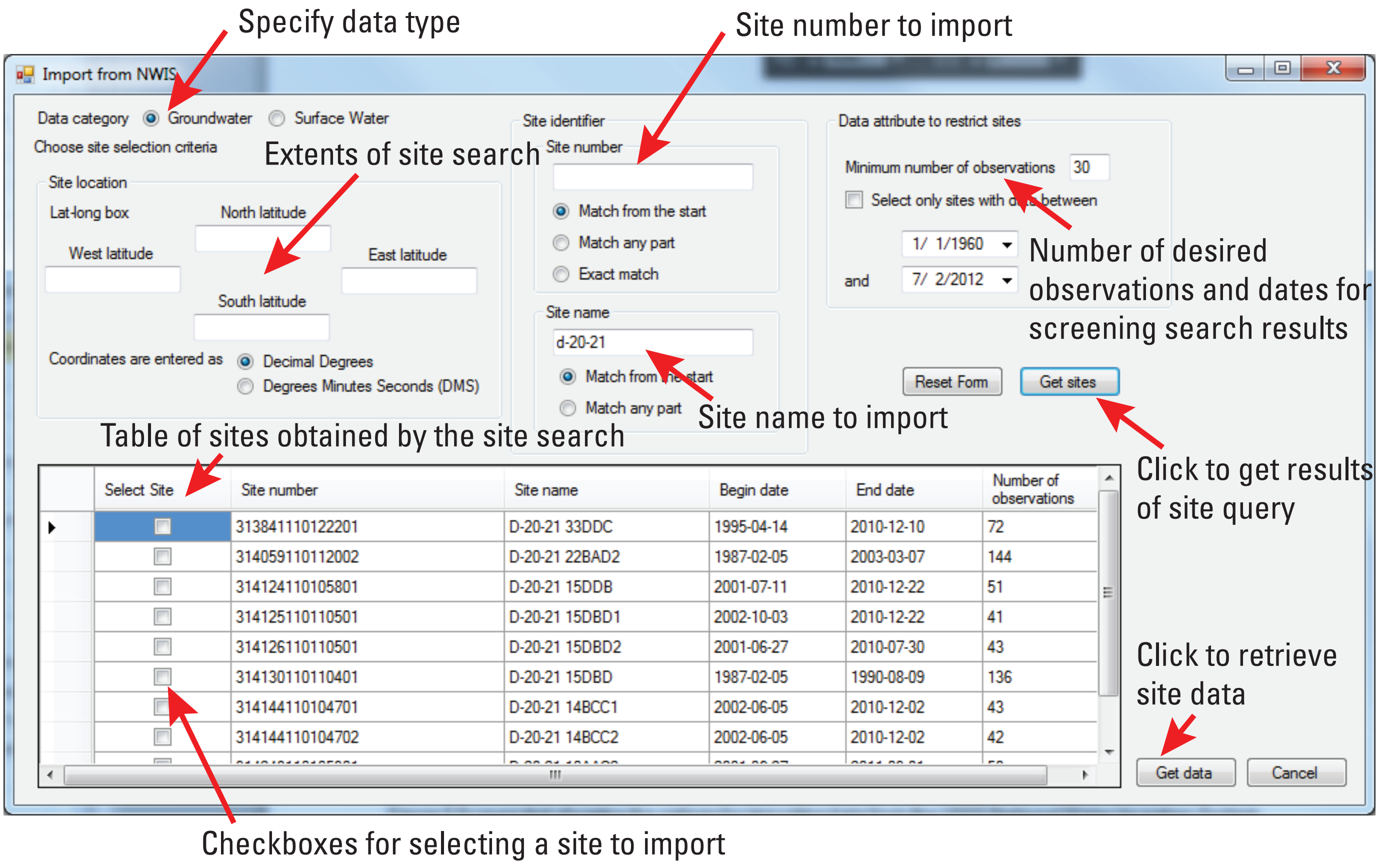
Figure 6. Options for importing data from a file in matrix format.

Importing Data from the National Water Inventory System (NWIS)

Groundwater-level data can be imported from the NWIS by using search criteria, such as site ID, site name, location, and length of record. To import data from NWIS, on the **File** menu, click **Import time series from NWIS**. This opens a separate form that has search criteria of site location, site identifier, and data attribute (fig. 7). After entering the search criteria, select the button **Get sites**, and the table on the form will fill with sites from NWIS that match the search criteria. To select a site, click the checkbox in the table that corresponds to the site. Finally, to import the data, select the button **Get data** at the bottom of the form. Future versions of this software are planned to include search options for surface water, water quality, and tree-ring data.

Exporting Data

Vector or matrix (gridded) time-series data can be exported and saved to a local file. The saved file can be ASCII, Excel®, or raster format. To save data, on the **File** menu, click either **Export time series** or **Export grid**, and select the file format from the options that appear to the right (figs. 8 and 9). A separate form appears that has options for the name of the time series to save and the name of the file to save to. Matrix data can be saved in a vector or matrix format. For the vector format, the data are saved in a vector that contains a repeating list of data in a single column for each increment along each dimension. In the matrix format, the data are saved as two-dimensional matrices that repeat for each time step.



Data category ● Groundwater ○ Surface Water

Choose site selection criteria

Site location

Lat-long box

North latitude

West latitude

East latitude

South latitude

Coordinates are entered as ● Decimal Degrees ○ Degrees Minutes Seconds (DMS)

Site identifier

Site number

● Match from the start
○ Match any part
○ Exact match

Site name: d-20-21

● Match from the start
○ Match any part

Data attribute to restrict sites

Minimum number of observations: 30

□ Select only sites with data between 1/ 1/1960 and 7/ 2/2012

Select Site	Site number	Site name	Begin date	End date	Number of observations
□	313841110122201	D-20-21 33DDC	1995-04-14	2010-12-10	72
□	314059110112002	D-20-21 22BAD2	1987-02-05	2003-03-07	144
□	314124110105801	D-20-21 15DDB	2001-07-11	2010-12-22	51
□	314125110110501	D-20-21 15DBD1	2002-10-03	2010-12-22	41
□	314126110110501	D-20-21 15DBD2	2001-06-27	2010-07-30	43
□	314130110110401	D-20-21 15DBD	1987-02-05	1990-08-09	136
□	314144110104701	D-20-21 14BCC1	2002-06-05	2010-12-02	43
□	314144110104702	D-20-21 14BCC2	2002-06-05	2010-12-02	42

Specify data type

Site number to import

Extents of site search

Number of desired observations and dates for screening search results

Site name to import

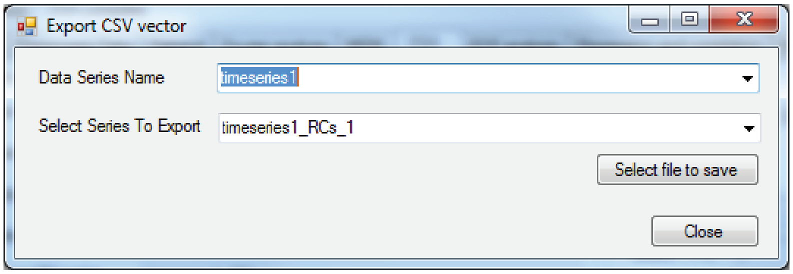
Table of sites obtained by the site search

Click to get results of site query

Click to retrieve site data

Checkboxes for selecting a site to import

Figure 7. Options for importing data from the U.S. Geological Survey National Water Inventory System.

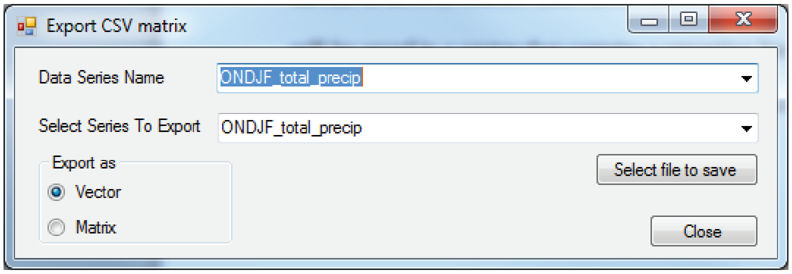


Export CSV vector

Data Series Name: timeseries1

Select Series To Export: timeseries1_RCs_1

Figure 8. Options for exporting data from a vector to an ASCII or Excel file.



Export CSV matrix

Data Series Name: ONDJF_total_precip

Select Series To Export: ONDJF_total_precip

Export as

● Vector

○ Matrix

Figure 9. Options for exporting data from a matrix to an ASCII or Excel file.

Describe Data

Data descriptions can be provided in order to document the time-series data (fig. 1). These descriptions are optional and are not included in any of the steps of analysis. The data descriptions can be assigned on the tab **Describe data** on the main form. The types of descriptors include the site information, data type and time units, and the location of the site where the data were collected. These descriptors are only saved if the button “Save data description” is pressed. Some of these fields are automatically populated if the data series is imported from NWIS. The **Describe Data** tab displays a statistical summary table for the imported data series.

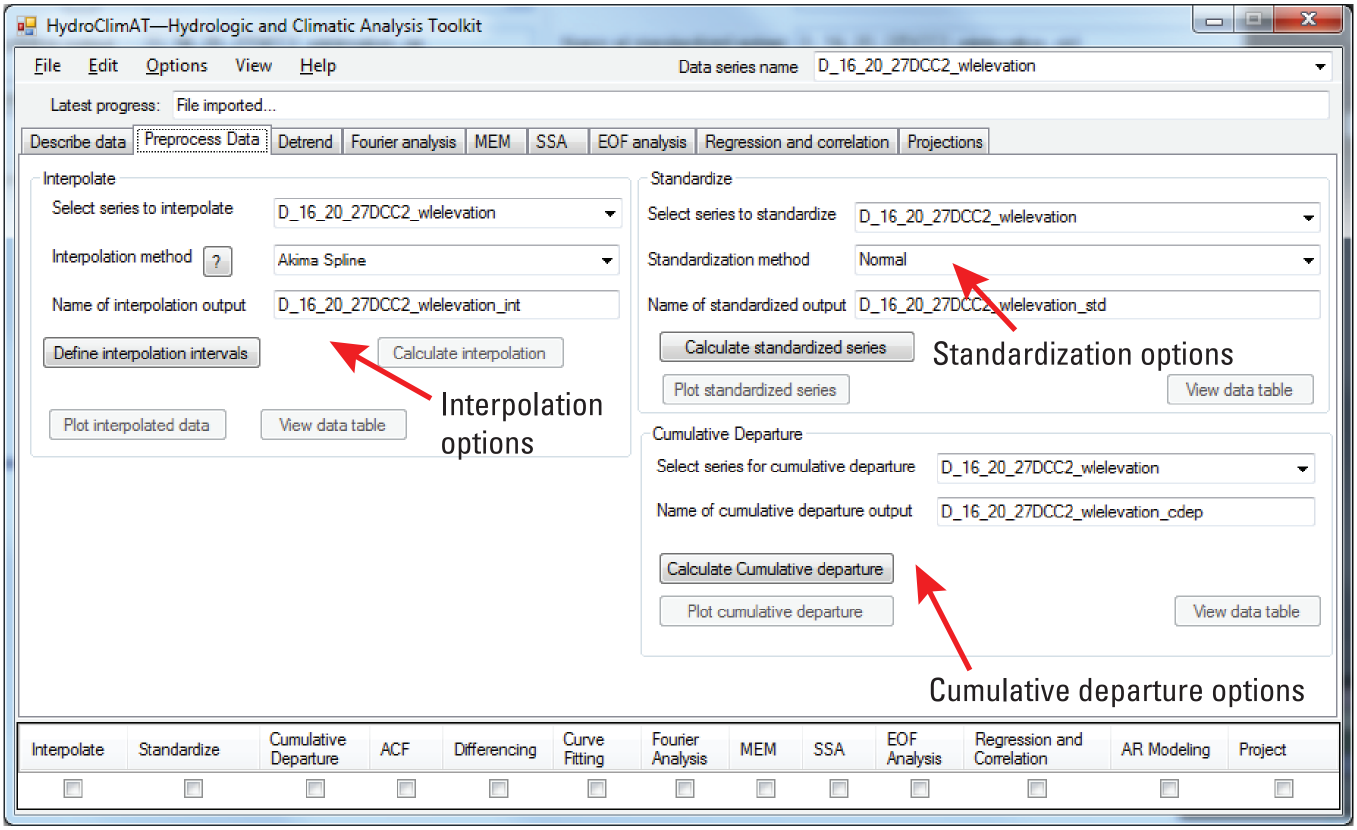
Preprocess Data

Several preprocessing steps that are available on the tab **Preprocess Data** could be necessary prior to using some of the other tools that are available on the other tabs. The preprocessing steps are separated into different group boxes labeled **Interpolate**, **Standardize**, and **Cumulative Departure** (fig. 10). Interpolation can be used in some situations for estimating missing values. Standardization is often used, but not necessary prior to the SSA, EOF, and regression analysis. Cumulative departure is useful for removing values of zero that are common in some time series, such as precipitation records in arid environments, or making the data consistent with another data type, such as groundwater levels that represent a cumulative departure of changes in aquifer storage.

Interpolation for Missing Data

Hydrologic time series often include measurements at somewhat irregular intervals, which can range from daily to annual values. Hanson and others (2004) used interpolation to estimate missing values and to create a uniform time interval between the values in the time series. A uniform interval is generally necessary for the Fourier and SSA analysis tools. Interpolation can be required for measured data but could be unnecessary when analyzing simulated data from a model that are already available at uniform time intervals.

Tools for interpolating missing values in a vector time series are available in the groupbox labeled **Interpolate**. A vector time series can be selected from the drop-down list that is labeled **Select series to interpolate**. The interpolation method can be selected from the drop-down list labeled **Interpolation method**. The available methods are **linear interpolation**, **cubic spline**, and **Akima spline**. The text box labeled **Name of interpolation output** is used to assign a series name to the output from the interpolation. By default, the series name has the extension **_int** added to the root name of the data series. The times at which interpolation is performed are defined by clicking on the button **Define interpolation intervals**, which opens a separate form that allows the user to define the number of points between the beginning and ending date or by loading a separate file that has a list of times for interpolation (fig. 11). After defining the interpolation times, interpolation can be calculated by clicking the button **Calculate interpolation**. Outputs can be viewed by clicking the buttons **Plot interpolated data** or **View data table.**



Data series name: D_16_20_27DCC2_wlelevation

Latest progress: File imported...

Select series to interpolate: D_16_20_27DCC2_wlelevation

Interpolation method ?: Akima Spline

Name of interpolation output: D_16_20_27DCC2_wlelevation_int

Select series to standardize: D_16_20_27DCC2_wlelevation

Standardization method: Normal

Name of standardized output: D_16_20_27DCC2_wlelevation_std

Select series for cumulative departure: D_16_20_27DCC2_wlelevation

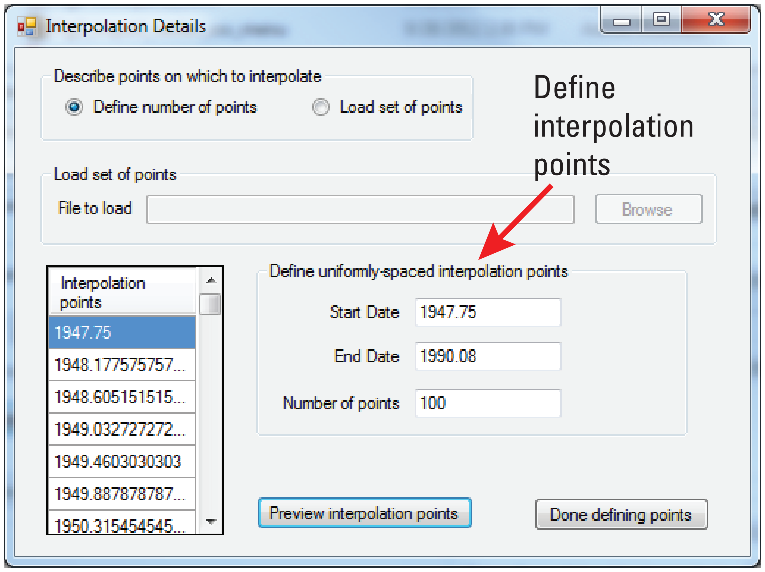
Name of cumulative departure output: D_16_20_27DCC2_wlelevation_cdep

Interpolation options

Standardization options

Cumulative departure options

Figure 10. Options for interpolating, standardizing data, and calculating cumulative departure.


Interpolation Details
Describe points on which to interpolate
● Define number of points
○ Load set of points
Load set of points
Browse
Define uniformly-spaced interpolation points
Preview interpolation points
Done defining points

File to load:

Interpolation points
1947.75
1948.177575757...
1948.605151515...
1949.032727272...
1949.4603030303
1949.887878787...
1950.315454545...

Start Date: 1947.75

End Date: 1990.08

Number of points: 100

Define interpolation points

Figure 11. Options for defining the times at which the time series is interpolated.

Standardize Data

Data can be standardized to a normal distribution or a gamma-to-normal distribution by using the Standardized Precipitation Index (SPI) procedure described by McKee and others (1993) and Edwards and McKee (1997). Standardization allows for comparisons across data types, such as precipitation, streamflow discharge, and groundwater levels. Standardization by using SPI also allows for comparisons of data across regions having different climates. SPI is typically used for precipitation because these values often are not normally distributed, such as in arid and semiarid regions. SPI standardization is useful for analyzing continental-scale patterns of precipitation on interannual and interdecadal times scales because climates often vary spatially within continental scales (Castro and others, 2009).

Tools for standardizing data in a time series are available in the groupbox labeled **Standardize**. Either a vector or a matrix time series can selected from the drop-down list that is labeled **Select series to standardize**. The standardization method (**Normal** or **Gamma to Normal (SPI)**) can be selected from the drop-down list labeled **Standardization method**. The text box labeled **Name of standardized output** is used to assign a series name to the output from the standardization. By default, the series name has the extension **_std** added to the root name of the data series. The data can be standardized by clicking the button **Calculate standardized series**. Outputs can be viewed by clicking the buttons **Plot standardized series** or **View data table.**

Cumulative Departure

The cumulative departure transformation provides serial correlation for intermittent temporal processes, such as precipitation and ephemeral streamflow. This allows for comparison to many geophysical time series that have persistence between subsequent values, such as groundwater-level data (Hanson and others, 2004). On the cumulative departure curve, an interval of time with a positive slope generally indicates that the short-term mean of the period is greater than the overall mean, and a negative slope generally indicates that the short-term mean is less than the overall mean.

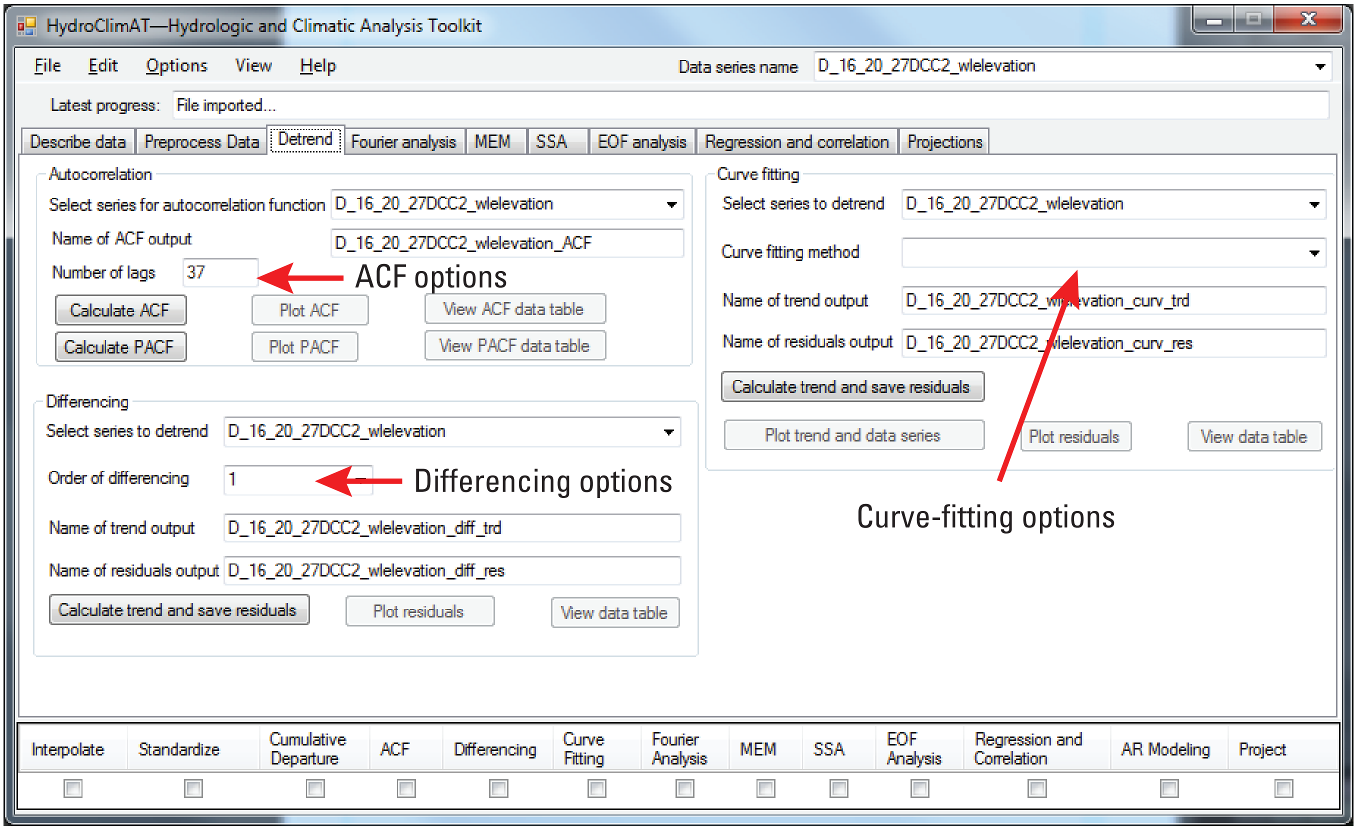
The cumulative departure tool is available in the groupbox labeled **Cumulative Departure**. A vector time series can be selected from the drop-down list that is labeled **Select series for cumulative departure**. The text box labeled **Name of cumulative departure output** is used to assign a series name to the output. By default, the series name has the extension **_cdep** added to the root name of the data series. The cumulative departure can be calculated by clicking the button **Calculate Cumulative departure**. Outputs can be viewed by clicking the buttons **Plot cumulative departure** or **View data table.**

Detrend

A trend in a time series is generally a gradual change in the values of the series. Here, detrending is a mathematical operation that removes a trend from a time series. Detrending is often used to remove a characteristic that distorts features of the time series that are of interest. The methods for evaluating time series described by Hanson and others (2004) include the identification and removal of non-stationary elements and any low-frequency cycles in order to prepare the time series for techniques that assume stationarity. In hydrologic data, non-stationary elements are trends in the mean or variance that are typically caused by (1) human activity or changes in watershed characteristics, such as urbanization or geomorphic changes to a stream channel, or (2) changes in climate that available information cannot identify as a repeating cycle. Non-stationary trends can follow a linear pattern, curvilinear pattern, a step-change pattern, or other pattern. Low-frequency cycles can be removed because the Fourier analysis and MEM tools cannot identify low-frequency cycles that are longer than half the period of record. The autocorrelation tool can identify persistence in time series that can be attributed to a trend or low-frequency cycle. The trend or low-frequency cycle can be removed by differencing or by using a least-squares fit of a linear or polynomial fit to the trend or cycle. The Fourier analysis, MEM, and SSA tools can then use the residuals from the fitted relation. The approaches to detrending that are available in the software are differencing and curve fitting. Tools for these methods are available on the tab **Detrend** and are separated into groupboxes labeled **Differencing** and **Curve fitting** (fig. 12).

Autocorrelation

In a general sense, a time series is assumed to be stationary if its statistical properties are the same after it has been shifted through time. The autocorrelation function (ACF) is a tool that can be used to identify any persistence between subsequent values in a time series. The ACF tool produces a plot of the correlation between a time series and the same series after being shifted by an integer number of lags. A time series with persistence will have a small autocorrelation at any lag other than zero. By definition, the autocorrelation at a lag of zero is equal to one. The ACF tool is available on the tab labeled **Detrend**. The maximum number of lags to be considered can be assigned by entering an integer value in the text box labeled **Number of lags**. The text box labeled **Name of ACF output** is used to assign a series name to the output. By default, the series name has the extension **_ACF** added to the root name of the data series. The ACF is calculated by clicking the button **Calculate ACF**, and the output can be viewed by clicking the buttons **Plot ACF** or **View ACF data table**.



Data series name: D_16_20_27DCC2_wlelevation

Latest progress: File imported...

Select series for autocorrelation function: D_16_20_27DCC2_wlelevation

Name of ACF output: D_16_20_27DCC2_wlelevation_ACF

Number of lags: 37

Select series to detrend: D_16_20_27DCC2_wlelevation

Order of differencing: 1

Name of trend output: D_16_20_27DCC2_wlelevation_diff_trd

Name of residuals output: D_16_20_27DCC2_wlelevation_diff_res

Select series to detrend: D_16_20_27DCC2_wlelevation

Curve fitting method:

Name of trend output: D_16_20_27DCC2_wlelevation_curv_trd

Name of residuals output: D_16_20_27DCC2_wlelevation_curv_res

□ □ □ □ □ □ □ □ □ □ □ □ □

Figure 12. Options for assessing for autocorrelation and for detrending by using differencing or curve fitting.

Differencing

Differencing (Brockwell and Davis, 2002) can be used to remove non-stationarity in the mean of a time series. The differencing tool is available in the groupbox labeled **Differencing** on the tabbed page **Detrend**. A vector time series can be selected from the drop-down list that is labeled **Select series to detrend**. The order of the differencing (from one to three) can be selected from a drop-down list labeled **Order of differencing**. The order indicates the number of times that a first difference is calculated (Brockwell and Davis, 2002). The text box labeled **Name of trend output** is used to assign a series name of the output. By default, the series name has the extension **_trd** added to the root name of the data series. The results from the differencing can be calculated by clicking the button **Calculate trend and save residuals**. Outputs can be viewed by clicking the buttons **Plot residuals** or **View data table.**

Curve Fitting

Curve fitting uses a fitted function of time to represent the trend. The residuals of the time series from the trend can represent the time series with the trend removed. The curve-fitting tool is available in the groupbox labeled **Curve fitting** on the tabbed page **Detrend**. A vector time series can be selected from the drop-down list that is labeled **Select series to detrend**. The curve-fitting method (linear fit, 2nd order polynomial fit, 3rd order polynomial fit, and 4th order polynomial fit) can be selected from a drop-down list labeled **Curve fitting method**. The shape of the fitted polynomial is controlled by the order. Hanson and others (2004) used polynomial fits to remove persistent changes that appeared to be trends, such as long-term declines in water levels that they attributed to groundwater withdrawals. The text box labeled **Name of trend output** is used to assign a series name to the output. By default, the series name has the extension **_trd** added to the root name of the data series. The fitted curve (the trend) and the residuals from the trend can be calculated by clicking the button **Calculate trend and save residuals**. Outputs can be viewed by clicking the buttons **Plot trend and data series** to see how well the fitted curve represents the changes in the time series, **Plot residuals** to view a plot of the residuals, or **View data table** to view the values of the residuals in a table.

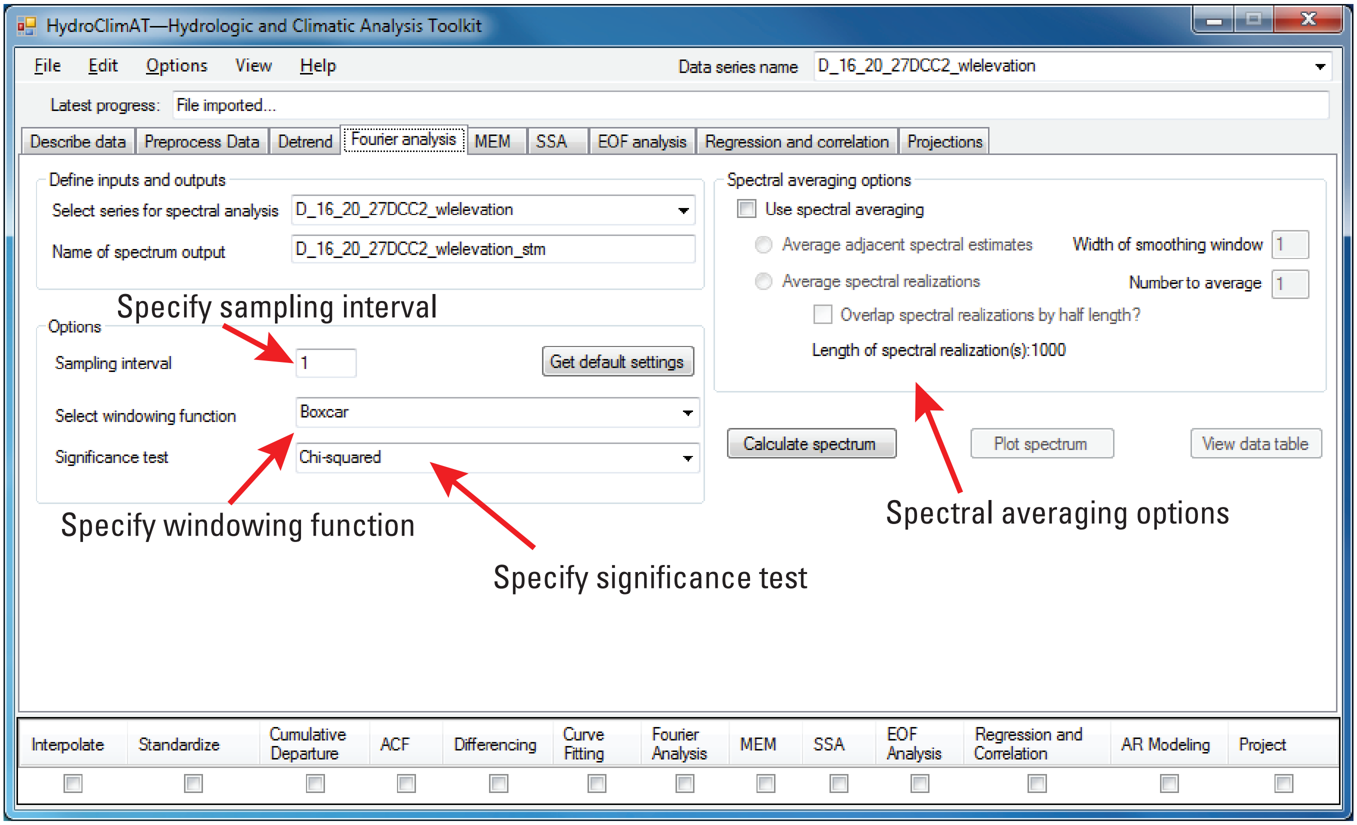
Fourier Analysis

Fourier analysis is useful for interpreting a time or space series as a superposition of harmonic functions that have a characteristic time or space scale. The goal is to obtain a plot of the variance of a time series as a function of wave number, frequency, or period of fitted periodic functions, which is called a spectrum. The discrete Fourier transform (DFT) and spectral averaging tools can be used to compute the spectrum of a time series. Fourier analysis is useful because many time series include variability that can be caused by reoccurring physical mechanisms, and an analysis of the variability could provide insight on the relation to these mechanisms. Spectral averaging is useful for increasing the number of degrees of freedom of the spectral estimates, which increases the reliability of each spectral estimate. Approaches to spectral averaging include averaging adjacent spectral estimates and averaging separate realizations of spectra. Windowing is useful for addressing the phenomena known as “spectral leakage,” which results from analyzing a discrete time series with a beginning and end instead of the continuous time series that is theoretically required by DFT. Other approaches for obtaining the spectrum (Blackman-Tukey method, Multi-taper method, and maximum entropy method) are available in the freely-available software SSA-MTM Toolkit available at *http://www.atmos.ucla.edu/tcd/ssa/* (Dettinger and others, 1995b).

Tools for DFT and spectral averaging are available in the tabbed page labeled **Fourier analysis** (fig. 13). A vector time series can be selected from the drop-down list that is labeled **Select series for spectral analysis**. The selected time series can contain raw values, but is often the output from the preprocessing or detrending tools. The text box labeled **Name of spectrum output** is used to assign a series name of the spectrum. By default, the series name has the extension **_stm** added to the root name of the data series. The sampling interval, which is the interval between values in the time series used in the Fourier analysis, is specified by entering an integer in the textbox labeled **Sampling interval**. For example, a value of one indicates all values are used, and a value of two indicates every other value is used. A windowing function (for example, boxcar, Hann, Hamming, or Parzen), which is applied to the time series prior to the DFT calculation, is selected from a down-down list labeled **Select windowing function**. The importance of selecting a windowing function for addressing spectral leakage is addressed in many texts on time series analysis (for example, Otnes and Enochson, 1978). The significance of the spectrum in relation to a red-noise null hypothesis can be evaluated by either a Chi-squared or F test, which is chosen from the drop-down list labeled **Significance test**. Default values can be assigned by clicking on the button **Get default settings**, which resets the sampling interval to one, the windowing function to Hann, the significance test to Chi-squared, and turns off the spectral averaging options. The Hann window is considered to be the most commonly used window in meteorological applications. The spectral averaging option is turned off because the user could want to begin with no averaging and test the effects of different averaging approaches.

The reproducibility of the spectrum can be improved by increasing the degrees of freedom at each specified frequency. The degrees of freedom can be increased by increasing the bandwidth, or range of frequencies, for each spectral estimate through spectral averaging. This means that the power of a component is obtained for a range of frequencies. This approach could be reasonable in many applications because geophysical phenomena are typically not strictly periodic at a single frequency, but often operate within a range of frequencies. Increasing the bandwidth reduces the resolution of the computed power spectrum, however. Without spectral averaging, a spectrum estimates the power spectrum at $N/2$ frequencies, and the estimate of the power at each frequency only has 2 degrees of freedom, which is not reproducible in many situations. The degrees of freedom for each spectral estimate can generally equal N/M^*, where N is the total number of values in the time series, and M^* is the total number of degrees of freedom in the spectrum. For example, if a time series has 1,000 values and the spectrum has 500 independent spectral estimates, each spectral estimate has 2 degrees of freedom. Adjacent spectral estimates can be averaged (for example, a moving average) to increase the degrees of freedom. If 10 adjacent spectral estimates within a single spectrum are averaged, then the total number of spectral estimates is 100, and each spectral estimate has 10 degrees of freedom. The degrees of freedom can also be increased by averaging spectral realizations. If the original time series is split into separate segments of equal length, a spectral realization, or separate spectrum, for each segment can be calculated. If the spectra for each segment are averaged at each frequency, then the degrees of freedom for each spectral estimate is approximately $2N/M_{ch}$, where M_{ch} is the number of values in each segment. For example, if a times series of 1,000 values is separated into 10 segments (each has 100 values), each spectral estimate has 20 degrees of freedom.

Spectral averaging options are enabled by clicking the checkbox labeled **Use spectral averaging**. The two spectral averaging options are **Average adjacent spectral estimates** and **Average spectral realizations**. The **Average adjacent spectral estimates** method uses a moving average to smooth the spectrum, and the width of the smoothing window is specified by the textbox labeled **Width of smoothing window**. The **Average spectral realizations** method splits the time series into separate segments of the original time series and computes a spectrum for each segment. The final spectrum is an average of the spectra for each frequency. The number of separate segments used is specified by entering an integer in the textbox labeled **Number to average**. An option for averaging spectral realizations is to use the Welch Overlapping Segment Analysis (WOSA; Welch, 1967), in which half of the



Data series name: D_16_20_27DCC2_wlelevation

Latest progress: File imported...

Define inputs and outputs

Select series for spectral analysis: D_16_20_27DCC2_wlelevation

Name of spectrum output: D_16_20_27DCC2_wlelevation_stm

Options

Sampling interval: 1

Select windowing function: Boxcar

Significance test: Chi-squared

Spectral averaging options

□ Use spectral averaging

○ Average adjacent spectral estimates — Width of smoothing window: 1

○ Average spectral realizations — Number to average: 1

□ Overlap spectral realizations by half length?

Length of spectral realization(s):1000

Specify sampling interval

Specify windowing function

Specify significance test

Spectral averaging options

Figure 13. Options for harmonic analysis by using a discrete Fourier transform and spectral averaging.

length of a time-series segment overlaps with the adjacent by half of the length. To use WOSA, click the checkbox labeled **Overlap spectral realization by half length**. The spectrum is calculated by clicking the button labeled **Calculate spectrum**. Outputs can be viewed by clicking the buttons **Plot spectrum** and **View data table.**

The plot of the spectrum includes options for displaying confidence intervals and for displaying each component as the wave number, frequency, or period. The values on the x-axis on the plots range from zero to the Nyquist frequency, which is the highest frequency that can be resolved. Frequency is equal to the number of cycles completed over the length of the time series. The x-axis can be changed by clicking on **View** on the menu bar and selecting **Wave number**, **Frequency**, or **Period** under the menu **Change x axis**. Significance tests can be plotted by clicking on **View** and selecting **Red noise null hypothesis**, **90%**, **95%**, or **99%** on the menu **Show sig. test**. The scale of the x-axis and y-axis can be modified by clicking **View** and **Log scale**, and selecting **X-axis linear**, **X-axis base e**, **X-axis base 10**, **Y-axis linear**, **Y-axis base e**, or **Y-axis base 10**.

Maximum Entropy Method

The maximum entropy method (MEM; Burg, 1967; Childers, 1978) can be used to estimate a spectrum of a time series. MEM is efficient for identifying frequencies that contribute most of the variance of a stationary time series. MEM obtains a spectrum by identifying an autoregressive (AR) model that is similar to the original time series. The coefficients of the AR model correspond to the location and width of the peaks in the spectrum. The output spectrum is dependent on the number of the coefficients, which is referred to as the number of poles, or order *M,* of the AR model. Greater values of *M* provide more resolution and identification of more peaks in the spectrum, but some of the peaks can be spurious. Lesser values of *M* produce a smoother spectrum, but the peaks of interest sometimes are not identified (Press and others, 1988). In practice, *M* should be several times greater than the total number of sharp spectral peaks that are desired in the spectrum (Press and others, 1988).

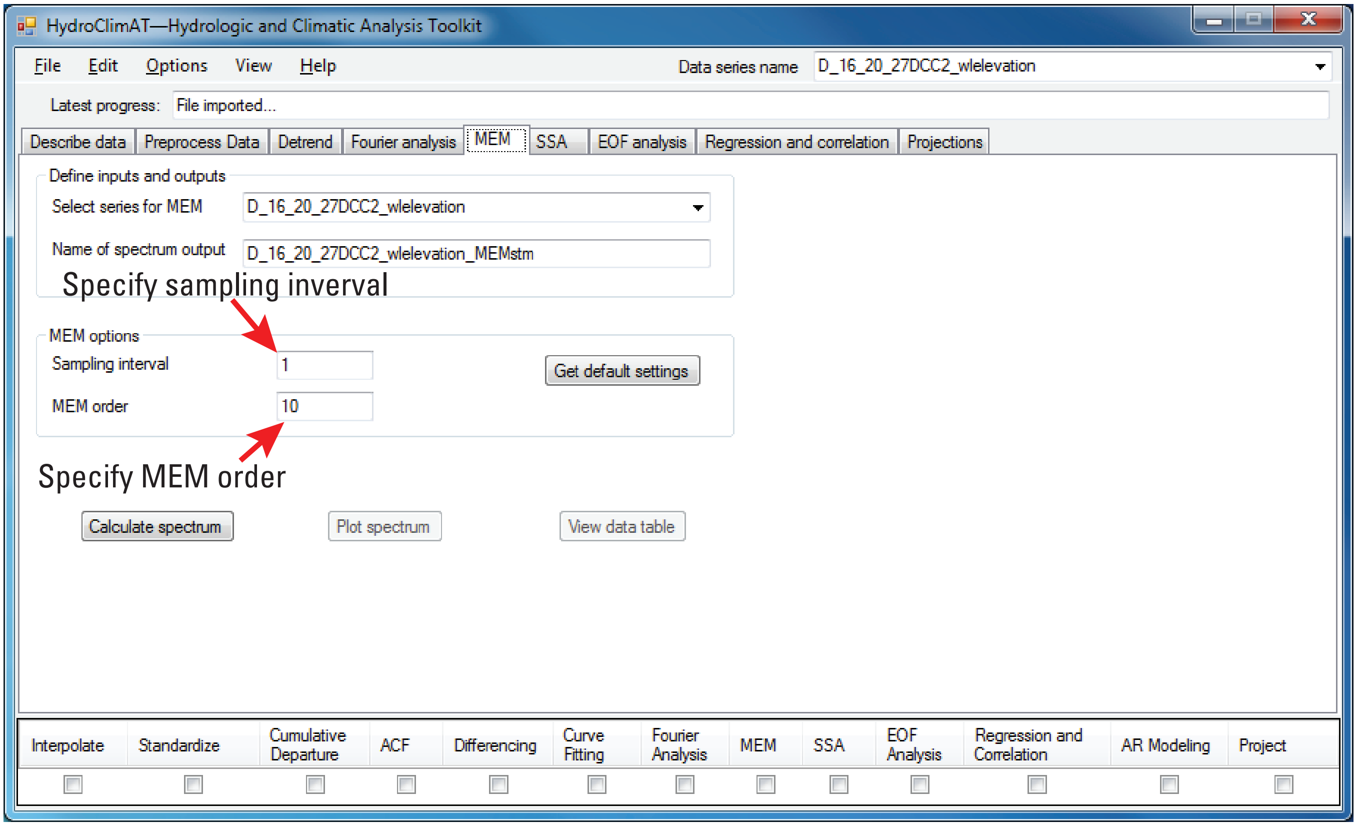
Tools for maximum entropy method are available on the tab **MEM** (fig. 14). This page has options for defining the input time series and naming output spectrum, as well as the sampling interval and the MEM order *M*. By default, the spectrum name has the extension _MEMstm added to the root name of the data series. Default values can be assigned by clicking on the button **Get default settings**. This automatically and arbitrarily sets the sampling interval to 1 and *M* to 10. The spectrum is computed by clicking the button labeled **Calculate spectrum**. To view the output, click the buttons labeled **Plot spectrum** or **View data table**.

Singular Spectrum Analysis

Periodic or quasi-periodic components can be identified by using the singular spectrum analysis (SSA) tool. SSA is a form of empirical orthogonal analysis (or principal component analysis) of a lagged covariance matrix (Broomhead and King, 1986; Vautard and others, 1992) and is useful for identifying oscillatory signals in short and noisy time series. This approach has an advantage over harmonic analysis through a Fourier transform because the fitted functions are not defined a priori, but are based on structures determined through eigenanalysis. Another advantage is that detrending is not required—the trend can be identified as a structure.

The goal of SSA is to obtain temporal structures that explain the maximum possible amount of covariance in time through an eigenanalysis of the lagged covariance matrix. The structures are explained by the eigenvectors, and the explained covariance per structure is obtained from its corresponding eigenvalue. The structures are often called the “temporal empirical orthogonal functions” (T-EOFs; Dettinger and others, 1995a), and the manner in which the T-EOFs change through time is described by the “temporal principal components” (T-PCs). The reconstructed structures in real time are called “reconstructed components” (RCs). The review paper by Ghil and others (2002) and the documentation for the SSA-MTM Toolkit (Dettinger and others, 1995b) provide extensive details of SSA. Further details on the T-EOFs and T-PCs are provided in the section “Empirical Orthogonal Function analysis.”

Tools for SSA are available in the tabbed page labeled **SSA** (fig. 15). This page has options for defining the inputs, SSA options, and options for viewing and exporting outputs. A vector time series can be selected from the drop-down list that is labeled **Select series for SSA**. Text boxes labeled **Name of spectrum output**, **Name of output T-EOF matrix**, **Name of output T-PC matrix**, and **Name of output RC matrix** are used to assign output names to be stored internally by the software. By default, the series names have the extensions _SSAstm, _TEOFs, _TPCs, and _RCs added to the root name of the data series accordingly.



Data series name: D_16_20_27DCC2_wlelevation

Latest progress: File imported...

Select series for MEM: D_16_20_27DCC2_wlelevation

Name of spectrum output: D_16_20_27DCC2_wlelevation_MEMstm

Specify sampling inverval

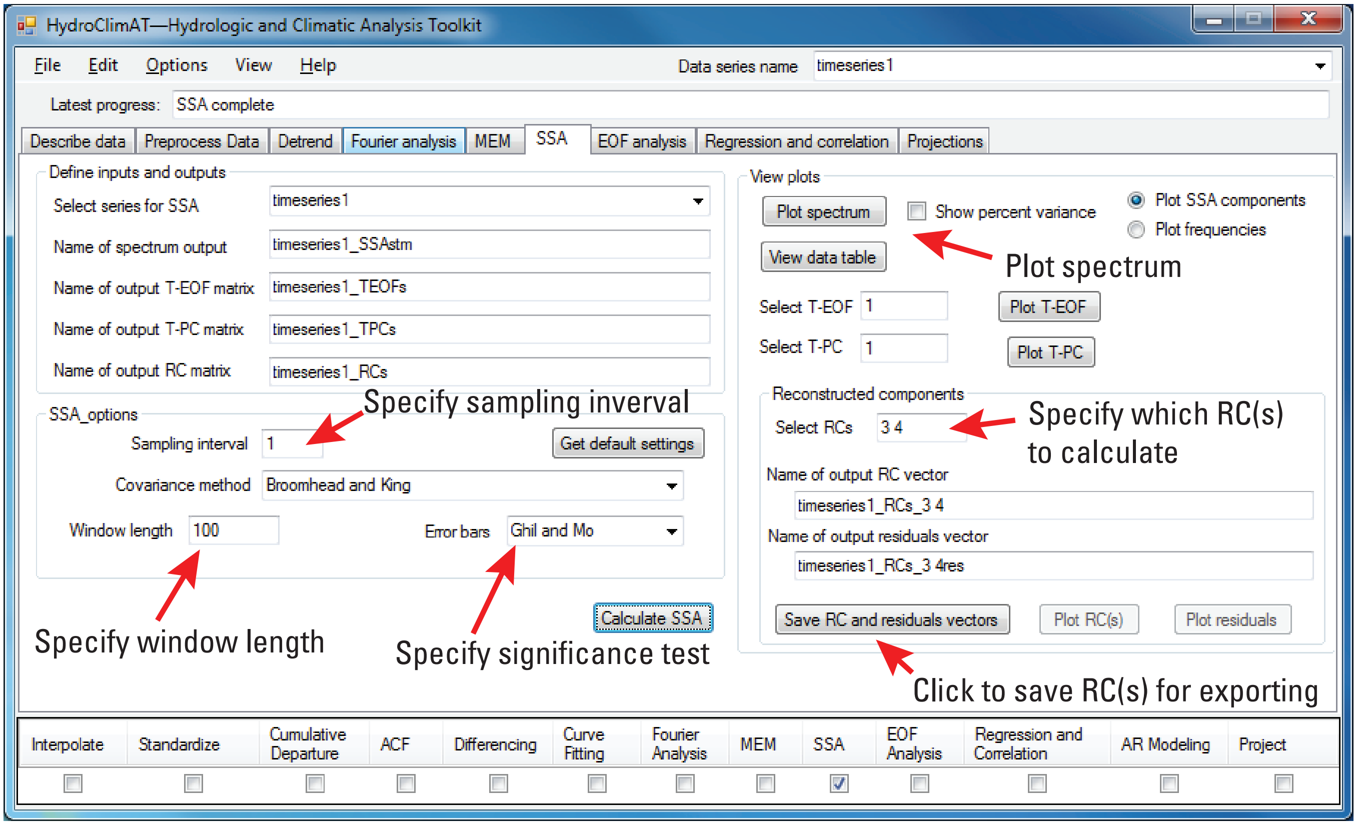
Sampling interval: 1

MEM order: 10

Specify MEM order

Interpolate	Standardize	Cumulative Departure	ACF	Differencing	Curve Fitting	Fourier Analysis	MEM	SSA	EOF Analysis	Regression and Correlation	AR Modeling	Project
□	□	□	□	□	□	□	□	□	□	□	□	□

Figure 14. Options for maximum entropy analysis.



Data series name: timeseries1

Latest progress: SSA complete

Select series for SSA: timeseries1

Name of spectrum output: timeseries1_SSAstm

Name of output T-EOF matrix: timeseries1_TEOFs

Name of output T-PC matrix: timeseries1_TPCs

Name of output RC matrix: timeseries1_RCs

Sampling interval: 1

Covariance method: Broomhead and King

Window length: 100

Error bars: Ghil and Mo

☐ Show percent variance

● Plot SSA components

○ Plot frequencies

Select T-EOF: 1

Select T-PC: 1

Select RCs: 3 4

Name of output RC vector: timeseries1_RCs_3 4

Name of output residuals vector: timeseries1_RCs_3 4res

Interpolate	Standardize	Cumulative Departure	ACF	Differencing	Curve Fitting	Fourier Analysis	MEM	SSA	EOF Analysis	Regression and Correlation	AR Modeling	Project
☐	☐	☐	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐

Specify sampling inverval

Plot spectrum

Specify which RC(s) to calculate

Specify window length

Specify significance test

Click to save RC(s) for exporting

Figure 15. Options for singular spectrum analysis.

The options for SSA are the sampling interval, window length, and method of calculating error bars of the spectral estimates. The window length needs to be wide enough to contain sufficient data over the interval of the oscillatory component that is of interest. For example, if components that vary from 20 to 30 years are of interest, and the samples are available at 12 points per year, then the window length needs to be at least 360. Vautard and others (1992) suggest that the window length be less than one-fifth of the total number of points in the time series. By default, the program sets the window length to be one-tenth of the total number of points. The spectrum is computed by clicking the button labeled **Calculate SSA**.

Ad hoc significance tests for the spectrum proposed by Ghil and Mo (1991), Vautard and Ghil (1989), and Unal and Ghil (1995) are selected from a drop-down list labeled **Error bars**. Significance of the components can be assessed by visual inspection of the spectrum and the error bars. Significant components contribute more variance than that from noise background and tend to be separated from the components on the flatter, right side of the spectrum by the length of the error bars. Ghil and others (2002) provide extensive details on assessing the significance of the components.

Plotting options for SSA are available for viewing the spectrum, T-EOFs, T-PCs, and reconstructed components (RCs). The spectrum can be plotted with the number of the SSA components, or frequency along the x-axis, and the variance, or percentage of variance along the y-axis, for each component or frequency. A time series of the RCs can be plotted for each individual RC or as a sum of selected RCs. A vector of the time series for the RCs can be stored internally by specifying a name of a series for exporting. Click the button labeled **Save RC vector** to store the vector. The amount of variance explained by the modes can be viewed by clicking on the button labeled **View data table** on the tab **SSA**. A sum of multiple RCs can be generated by entering multiple RC numbers, such as “1 2” for RC1 and RC2, in the text box labeled **Select RCs** and saved by clicking on the button **Save RCs and residuals vector**. A single RC or a sum of RCs can be plotted by clicking the button **Plot RC(s)**, and the residuals of the RCs from analyzed time series can be plotted by clicking the button **Plot residuals**.

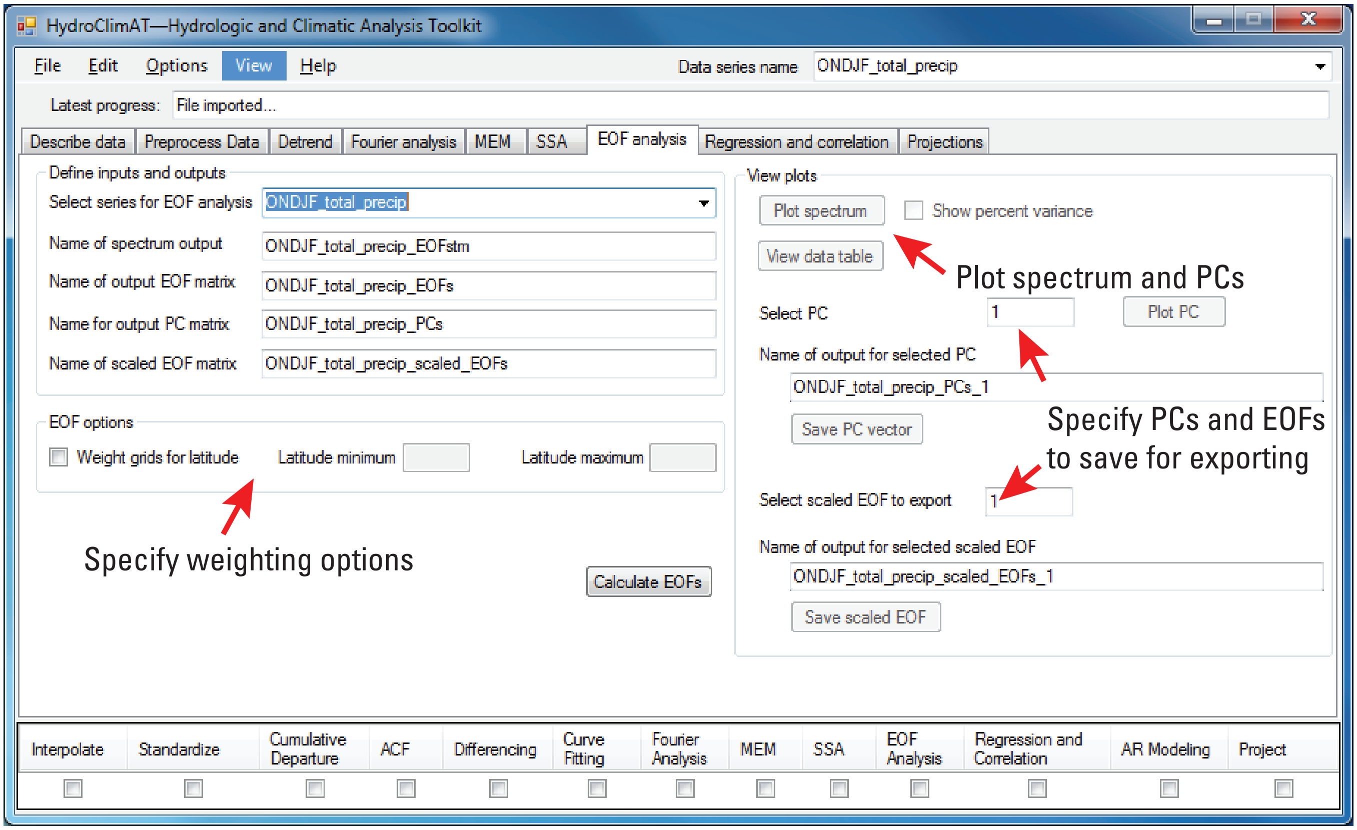
Empirical Orthogonal Function Analysis

Empirical orthogonal function (EOF) analysis is useful for identifying spatial and temporal structures that explain the most variance in two-dimensional data sets. EOF analysis is a form of principal component analysis that is commonly used in the atmospheric sciences to identify patterns in a time series of gridded spatial data. One dimension of the data set often contains the physical values in which structures are to be found, and the other dimension contains realizations of the physical values. An example of such a data set is spatially distributed precipitation along the dimension of physical values and the time for the precipitation along the dimension of the realizations. For this example, EOF analysis produces structures in the spatially distributed precipitation, which are called "empirical orthogonal functions." EOF analysis also produces structures in the realization, or time, dimension called "principal components," or "PCs," which explain how each "EOF" pattern varies through time. An application of EOF analysis to precipitation data in the southwestern U.S. is described in example 2.

Tools for EOF analysis are available on the tab **EOF analysis** (fig. 16). This page has options for defining the input series and naming the spectrum, EOF matrix, PC matrix, and the scaled EOF matrix. For large spatial datasets that span a wide range of latitudes and are gridded by using a non-equal-area projection, the values at higher latitudes are representative of smaller spatial areas and will dominate the results of the EOF analysis. In order to reduce the influence of the values at higher latitudes, the gridded values of the anomaly can be scaled by area at particular latitudes by selecting the checkbox labeled **Weight grids for latitude**. If this option is used, the user is required to specify the minimum and maximum latitudes for the data in the textboxes labeled **Latitude minimum** and **Latitude maximum**. The EOF analysis is computed by clicking the button labeled **Calculate EOFs**.

The resulting spectrum of the eigenvalues for each EOF and PC pair can be displayed by clicking the button **Plot spectrum**. To plot the spectrum as the variance explained by each EOF/PC pair, click the checkbox **Show percent variance** and click on the button **Plot spectrum**. A table of the values in the spectrum can be displayed by clicking the button **View data table**.

Individual PC time series and scaled EOF matrix outputs must be saved before the software can export the output. The individual PC time series and scaled EOF matrices represent the temporal and spatial patterns, respectively, that are associated with each eigenvalue. For example, if a dataset includes gridded precipitation data at 30 different times, the output from the EOF analysis includes 30 PCs and 30 EOFs. The scaled EOF matrices are scaled to be in the same units as the input data. In order to save a PC for analysis in the software or in order to export it, the PC must be saved by specifying a series name for the PC in the textbox labeled



Data series name: ONDJF_total_precip

Latest progress: File imported...

Select series for EOF analysis: ONDJF_total_precip

Name of spectrum output: ONDJF_total_precip_EOFstm

Name of output EOF matrix: ONDJF_total_precip_EOFs

Name for output PC matrix: ONDJF_total_precip_PCs

Name of scaled EOF matrix: ONDJF_total_precip_scaled_EOFs

□ Weight grids for latitude Latitude minimum: Latitude maximum:

□ Show percent variance

Select PC: 1

Name of output for selected PC: ONDJF_total_precip_PCs_1

Select scaled EOF to export: 1

Name of output for selected scaled EOF: ONDJF_total_precip_scaled_EOFs_1

□ Interpolate □ Standardize □ Cumulative Departure □ ACF □ Differencing □ Curve Fitting □ Fourier Analysis □ MEM □ SSA □ EOF Analysis □ Regression and Correlation □ AR Modeling □ Project

Plot spectrum and PCs

Specify PCs and EOFs to save for exporting

Specify weighting options

Figure 16. Options for empirical orthogonal function (EOF) analysis.

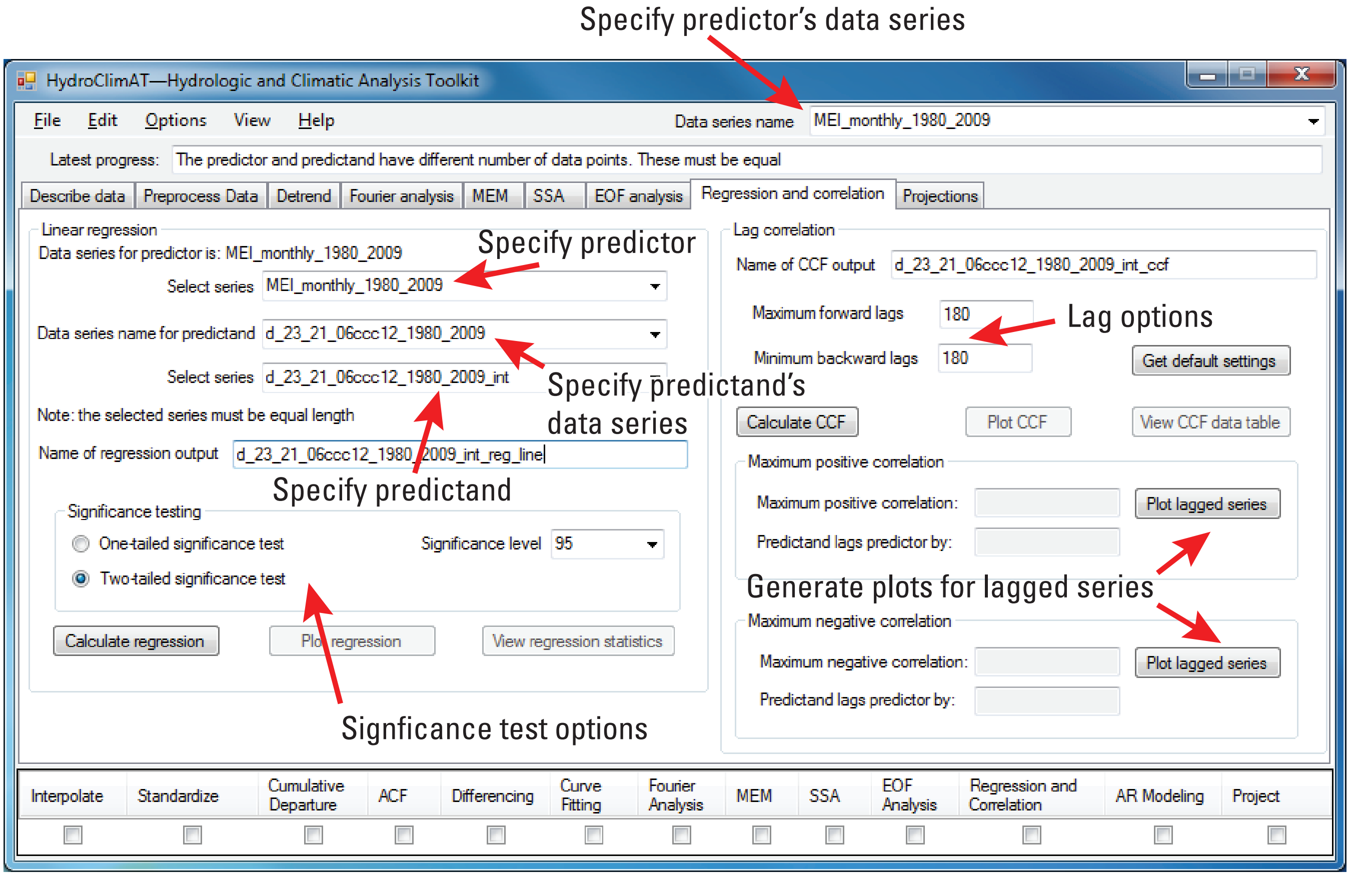
Name of output for selected PC and by clicking the button **Save PC vector**. A plot of any of the PC time series can be displayed by selecting a PC number in the textbox labeled **Select PC** and clicking the button **Plot PC**. A gridded plot of the EOFs cannot currently be generated by the software, but the gridded data can be exported to an ASCII raster file that can be imported and displayed in GIS software.

Linear Regression And Correlation

Linear regression and evaluation of the correlation between data series are useful for exploring linear relations between time series or for modeling one series (the predictand) as a function of the other series (the predictor). A lag correlation is useful for investigating the phase shift (for example, months) of the dependent time series that results in the strongest correlation.

Tools for regression and correlation analysis are available on the tab **Regression and correlation** (fig. 17). To specify the predictor, first, select a series in the drop-down list in the upper right corner that is labeled **Data series name** and, then, select a series name for the predictor in the drop-down list labeled **Select series**. The series that are available are subsets of the data series that is selected in the drop-down list at the upper right corner. To select a predictand, first, select a data series name in the drop-down list **Data series name for predictand** and, then, select a series name in the underlying drop-down list **Select series**. The series names will be subsets of the data series for the predictand. The series for the predictor and predictand must have an equal number of records. Either a one-tailed or two-tailed significance test using a *t* test at the 90, 95, or 99 percent confidence level can be selected to evaluate the significance of the correlation coefficient. These are calculated after clicking the button **Calculate regression**, and the results of the *t* test can be viewed by clicking **View regression statistics**.

The results of the lag correlation can be plotted as a cross correlation plot, which shows the correlation coefficient as a function of the lag, or time shift, between two time series. The two time series can be plotted together, where one series is lagged by the amount that has either the greatest positive correlation, or, when the series are expected to be negatively correlated, the series can be lagged by the amount that has the greatest negative correlation. The maximum forward and backward lags can also be specified, which is useful if there is an *a priori* expectation that a lag cannot be greater or less than a certain amount. The lag correlation can be performed by clicking the button **Calculate CCF**, and a plot of the lagged series at either the maximum positive or negative correlation can be generated by clicking the buttons **Plot lagged series**.


Specify predictor's data series

Specify predictor

Lag options

Specify predictand's data series

Specify predictand

Generate plots for lagged series

Signficance test options

Data series name: MEI_monthly_1980_2009

Latest progress: The predictor and predictand have different number of data points. These must be equal

Data series for predictor is: MEI_monthly_1980_2009

Select series: MEI_monthly_1980_2009

Data series name for predictand: d_23_21_06ccc12_1980_2009

Select series: d_23_21_06ccc12_1980_2009_int

Note: the selected series must be equal length

Name of regression output: d_23_21_06ccc12_1980_2009_int_reg_line

○ One-tailed significance test

● Two-tailed significance test

Significance level: 95

Name of CCF output: d_23_21_06ccc12_1980_2009_int_ccf

Maximum forward lags: 180

Minimum backward lags: 180

Maximum positive correlation:

Predictand lags predictor by:

Maximum negative correlation:

Predictand lags predictor by:

Interpolate	Standardize	Cumulative Departure	ACF	Differencing	Curve Fitting	Fourier Analysis	MEM	SSA	EOF Analysis	Regression and Correlation	AR Modeling	Project
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

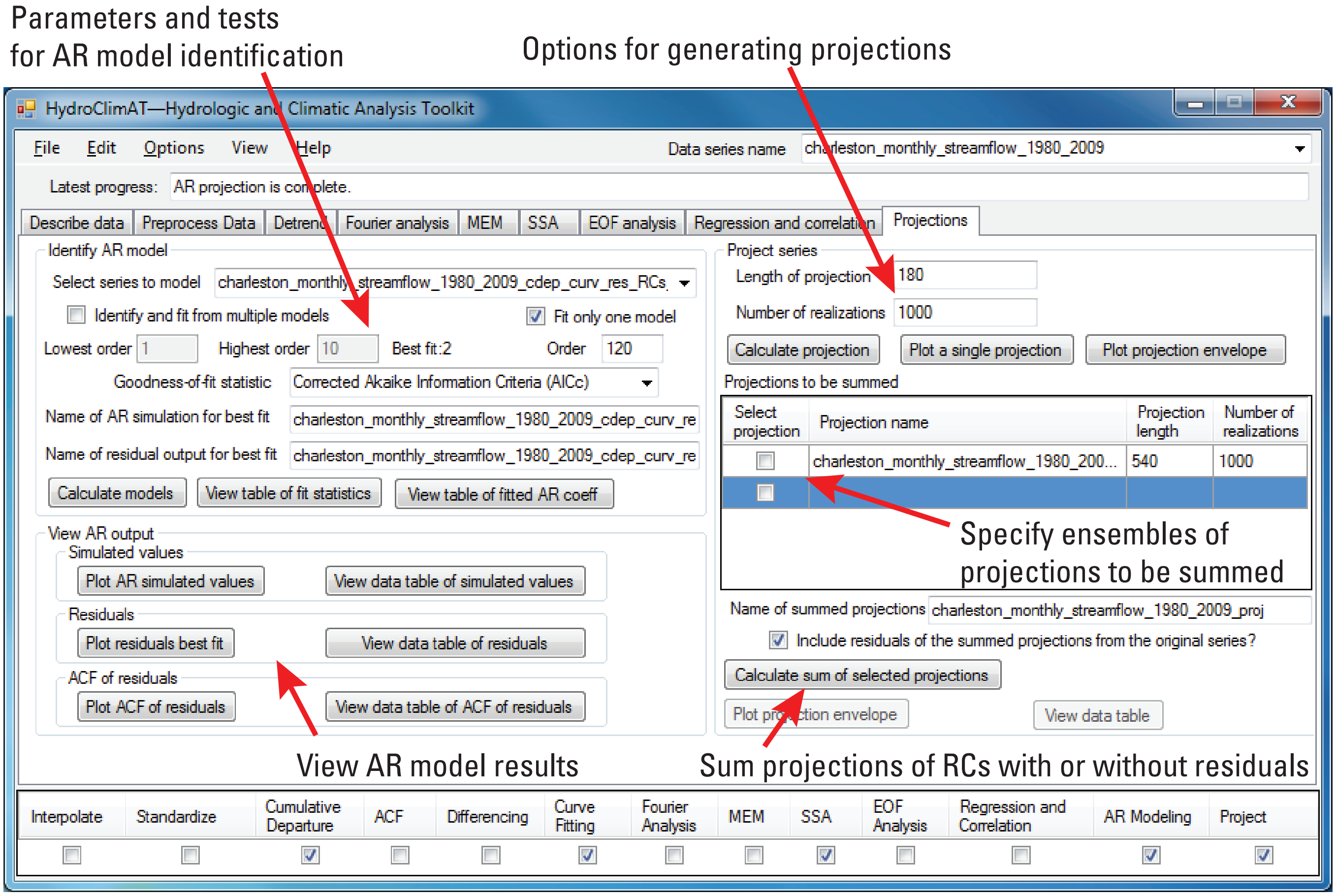
Figure 17. Options for assessing relations through regression and correlation analyses.

Projections

Time-series projections can be generated by using a combination of time series modeling, random-number synthesis, and extrapolation of the time series as described by Keppenne and Ghil (1992) and Jiang and others (1995) and implemented by Hanson and others (2003). An autoregressive (AR) mathematical model is the method used in the software to represent the persistence, or autocorrelation, in a time series. The persistence is used to extrapolate a time series under the assumption that the behavior in the past is useful for generating values at a future time.

The tools for AR modeling and generating projections are available on the tab **Projections** (fig. 18). To create an AR model of order *p*, the series to be modeled can be selected in the drop-down list **Select series to model**, and several options are available for selecting the order *p*. The order can be selected by checking **Identify and fit from multiple models** or **Fit only one model**. If multiple models are used, the software will calculate AR models having orders from 1 to the highest specified order. The order indicates how many values in the past will be used to generate a new value. A single order is specified if only one model is selected. In either case, the corrected Akaike information criteria (AICc), Akaike information criteria (AIC), or Akaike's final prediction error (FPE) tests can be used to select a model. These tests identify a model that minimizes the error between the data and model while accounting for parsimony. A model that has an order that minimizes the goodness-of-fit statistic, or, if only one model is desired, that has a specified order, is identified by clicking the button **Calculate models**. Additional results are plotted in tables by clicking the buttons **View table of fit statistics** and **View table of fitted AR coefficients**. The time series and the simulated values can be displayed by clicking **Plot AR simulated values** or **View data table of simulated values**. The residuals of the time series from the simulated values can be displayed by clicking **Plot residuals best fit** and **View data table of residuals**. The autocorrelation function for the residuals can be shown by clicking **Plot ACF of residuals** and **View data table of ACF of residuals**. The autocorrelation function plot is useful for evaluating whether the residuals are random, in which case the plotted values are near zero.

The fitted AR(*p*) model can be used to generate future values by entering values in the textbox **Length of projection** and **Number of realizations** (fig. 18). The length of the projection is an integer number of values that the time series is stepped forward past the last value in the time series. The number of realizations is the number of different projections (realizations of the stochastic process) with random



Parameters and tests for AR model identification

Options for generating projections

HydroClimAT—Hydrologic and Climatic Analysis Toolkit

Data series name: charleston_monthly_streamflow_1980_2009

Latest progress: AR projection is complete.

Identify AR model

Select series to model: charleston_monthly_streamflow_1980_2009_cdep_curv_res_RCs

☐ Identify and fit from multiple models ☑ Fit only one model

Lowest order: 1 Highest order: 10 Best fit:2 Order: 120

Goodness-of-fit statistic: Corrected Akaike Information Criteria (AICc)

Name of AR simulation for best fit: charleston_monthly_streamflow_1980_2009_cdep_curv_re

Name of residual output for best fit: charleston_monthly_streamflow_1980_2009_cdep_curv_re

Calculate models
View table of fit statistics
View table of fitted AR coeff
View AR output
Simulated values
Plot AR simulated values
View data table of simulated values
Residuals
Plot residuals best fit
View data table of residuals
ACF of residuals
Plot ACF of residuals
View data table of ACF of residuals

Project series

Length of projection: 180

Number of realizations: 1000

Projections to be summed

Select projection	Projection name	Projection length	Number of realizations
☐	charleston_monthly_streamflow_1980_200...	540	1000
☐			

Specify ensembles of projections to be summed

Name of summed projections: charleston_monthly_streamflow_1980_2009_proj

☑ Include residuals of the summed projections from the original series?

View AR model results

Sum projections of RCs with or without residuals

Interpolate	Standardize	Cumulative Departure	ACF	Differencing	Curve Fitting	Fourier Analysis	MEM	SSA	EOF Analysis	Regression and Correlation	AR Modeling	Project
☐	☐	☑	☐	☐	☑	☐	☐	☑	☐	☐	☑	☑

Figure 18. Options for fitting autoregressive models and generating projections.

components that are generated past the last value in the time series. A large set of realizations is useful for generating a range of possible values of the projection. The projection is created by clicking the button **Calculate projection**, and the results can be displayed by clicking **Plot a single projection** or **Plot projection envelope**. If plotting a single projection, the plotted series is selected randomly from the set of realizations. The projection envelope includes the mean of the realizations at each time, the upper and lower values of an envelope that contains 90 percent of the realizations, and three randomly selected realizations.

Multiple projections can be summed in order to generate a single projection by selecting more than one projection in the table **Projections to be summed** (fig. 18). This can be useful when a time series is decomposed into several reconstructed components by SSA, and each reconstructed component is projected separately. If the time series is reconstructed components, residuals from the sum of the reconstructed components from the original time series can be added into the sum of the projections by checking **Include residuals of the summed projections from the original series**. The final projection is created by clicking **Calculate sum of selected projections**, in which the selected projections are checked in the table, and the results can be displayed by clicking **Plot projection envelope** and **View data table**.

Examples

The following examples demonstrate several ways of using the tools in HydroClimATe to assess relations between hydrologic and climatic time series and for generating time-series projections.

Example 1—Spectral Analysis of a Synthetic Time Series

Spectral analysis of time series is a powerful tool for identifying repeating, frequency-dependent variability that could be related to a causal physical mechanism. Example 1 demonstrates how to use the spectral analysis tools and provides some guidance for interpreting the computed spectrum. The spectrum is a plot of the variance of a time series as a function of wave number, frequency, or period of fitted periodic functions. The spectrum can be used to identify an underlying pattern that explains much of the variability in a time series and can provide some insight into what physical processes could cause the variability. Examples of how spectral analysis has been used extensively in tree-ring research include the identification of relations between precipitation and temperature on growth rates of trees (LaMarche, 1974), and between tree-rings and solar cycles (La Marche and Fritts, 1972). This example demonstrates the following tools:

1. Discrete Fourier analysis.

2. Maximum entropy method.

3. Singular spectrum analysis.

This example demonstrates that the choice of the spectral method can produce differences in the power spectrum. Discrete Fourier analysis, maximum entropy method (MEM), and singular spectrum analysis (SSA) are demonstrated on a synthetic time series (fig. 19*A*, *B*) having several periodic components and noise that was created by Professor David W. J. Thompson at Colorado State University. Two of the main periodic components that compose most of the variance in the time series are reconstructed by using a convolution of the T-EOFs and T-PCs obtained from SSA. The first part of the example demonstrates how to perform the analysis using the software, and the last part describes the output and the consequences of selecting different options.

Several of the discrete Fourier analysis options are used to generate four different spectra, each having different amounts of noise in the spectra and significant spectral peaks. The options include the application of windowing functions and spectral averaging. The purpose of using these options is

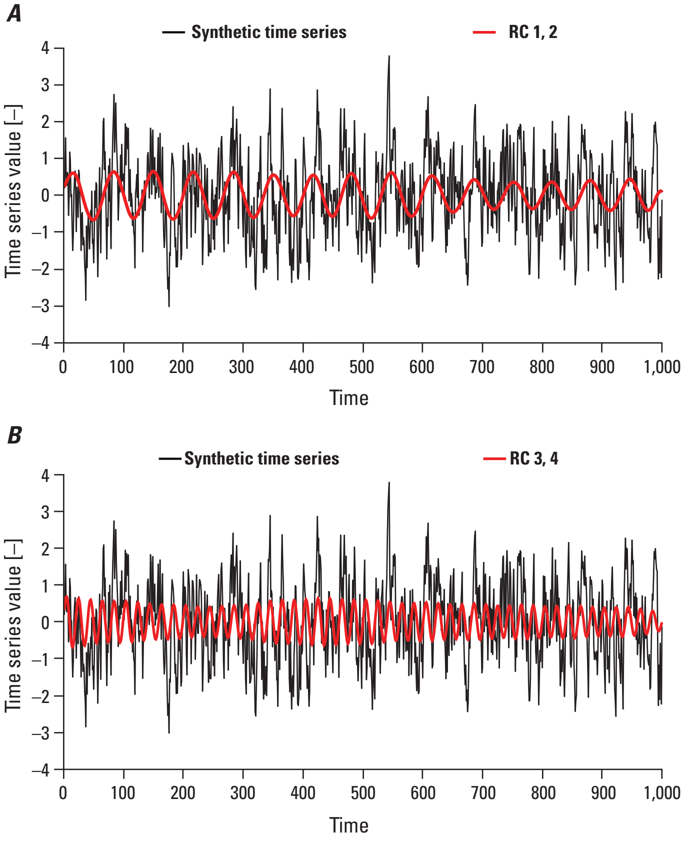


Figure 19. The synthetic time series used in example 1: *A*, the sum of the first and second reconstructed components; *B*, the sum of the third and fourth reconstructed components.

to increase the number of degrees of freedom (d.o.f.) of the spectral estimates and, therefore, to increase the repeatability of the spectral estimates. This can reduce some of the noise in the spectrum, but comes with a loss of resolution.

The first step in this analysis is to import the time series by selecting the **File** menu, clicking **Import time series from file**, and selecting **ASCII file** on the menu that appears to the right. The file "timeseries1.csv" is selected, the checkbox **First row has column names** is selected to skip the headers on each column, and it is named "timeseries1" in the textbox labeled **Data Series Name**. Click **OK** to close this window and import the data. The discrete Fourier analysis is performed by selecting the tab labeled **Fourier analysis** and selecting "timeseries1" in the drop-down list **Select series for spectral analysis**. The window is selected in the drop-down list **Select windowing function** (boxcar and Hann are used in this example). Spectral averaging options are selected by clicking the checkbox **Use spectral averaging**, entering the smoothing window width, the total number of spectral realizations to average, and then clicking the checkbox **Overlap spectral realizations by half length**. MEM is performed by selecting the tab labeled **MEM** and selecting "timeseries1" in the drop-down list **Select series for MEM**, specifying an order of 40, which is arbitrary and assumed to provide a reasonable spectral resolution, and then clicking **Calculate spectrum**. SSA is performed by selecting the tab labeled **SSA**, specifying a sampling interval of 1 and window length of 200, which is one-fifth of the total number of points in the time series, as recommended by Ghil and others (2002). Ghil and Mo error bars were selected in order to assess the statistical significance of the eigenvalues.

Discrete Fourier Analysis Using a Boxcar Window and No Spectral Averaging or Smoothing

Spectral estimates of the synthetic time series by the discrete Fourier analysis include a relatively noisy spectrum when using only boxcar windowing function and no spectral averaging (fig. 20*A*). This spectrum indicates two main significant spectral peaks at 99 percent confidence level (using a chi-squared test) at frequencies of 0.015 and 0.05.

Discrete Fourier Analysis Using a Boxcar Window and Spectral Smoothing

The spectrum is smoothed by taking the mean of the spectrum within a moving window of width 5 (fig. 20*B*). The smoothing of the spectrum increases the d.o.f., but reduces the resolution of the spectral peaks. The loss of resolution is accompanied by a larger bandwidth for any spectral power estimate. It also appears that the spectral power is decreased for the two peaks that were obtained without smoothing (fig. 20*A*), and additional peaks that are next to the two significant peaks are now significant because the power associated with the two peaks in figure 20*A* are "spread out." The spectral smoothing increases the d.o.f. because the five-point window uses five data points to calculate the power for each wave number. In this case, the d.o.f. equals 2 d.o.f. per estimate times 5 points per estimate equals 10 d.o.f. per averaged power estimate.

Discrete Fourier Analysis Using a Boxcar Window and Spectral Averaging

The spectral averaging tool subdivides the original time series into separate time "chunks," runs the discrete Fourier transform on each time chunk to obtain a spectral realization, and then takes the mean of the spectral realizations at each frequency. This example averages 10 spectral realizations and does not use the overlapping option. The spectral averaging (fig. 20*C*) increases the d.o.f., but drops the information about the lowest frequencies in the data. The lowest frequencies are lost because the discrete Fourier transform algorithm obtains N/2 spectral estimates, so as N (total number of points) decreases, so does the number of spectral power estimates. The lowest frequencies are obtained by fitting the longest sine and cosine functions to the time series, and the lowest resolved frequency is 2π per N/10, instead of per N, as is the case for the whole dataset N.

After applying spectral averaging, only one significant peak remains at the lower frequencies. The spectral averaging increases the d.o.f. because the 10 subsets use 10 data points to calculate the power for each frequency. In this case, the d.o.f. equals 2 d.o.f. per estimate times 10 points per estimate, which equals 20 d.o.f. for each averaged power estimate.

Discrete Fourier Analysis Using a Hann Window and Spectral Averaging with Overlap

The Hann window tapers the ends of the time series, which partially cancels out the negative side lobes of the rectangular response function and lessens spectral leakage (for example Otnes and Enochson, 1978). The disadvantage of the Hann window is that it smooths and broadens the central lobe, which means that the spectrum will be slightly smoothed compared to a rectangular boxcar window (fig. 20*D*). The spectrum is averaged by overlapping the windows by exactly one half of the chunk length, so each datum point is given the same weight in the resulting spectrum. This counteracts how the Hann window broadens the central lobe and weakens the pinched out ends; by moving the window by half of the chunk length, each section is broadened and weakened at least once. The exceptions are for the first and last points in the original time series. Because the analysis uses 9 chunks of time series, the d.o.f. is equal to 2 d.o.f. per spectral line times 9, which is equal to 18 for the averaged spectrum.

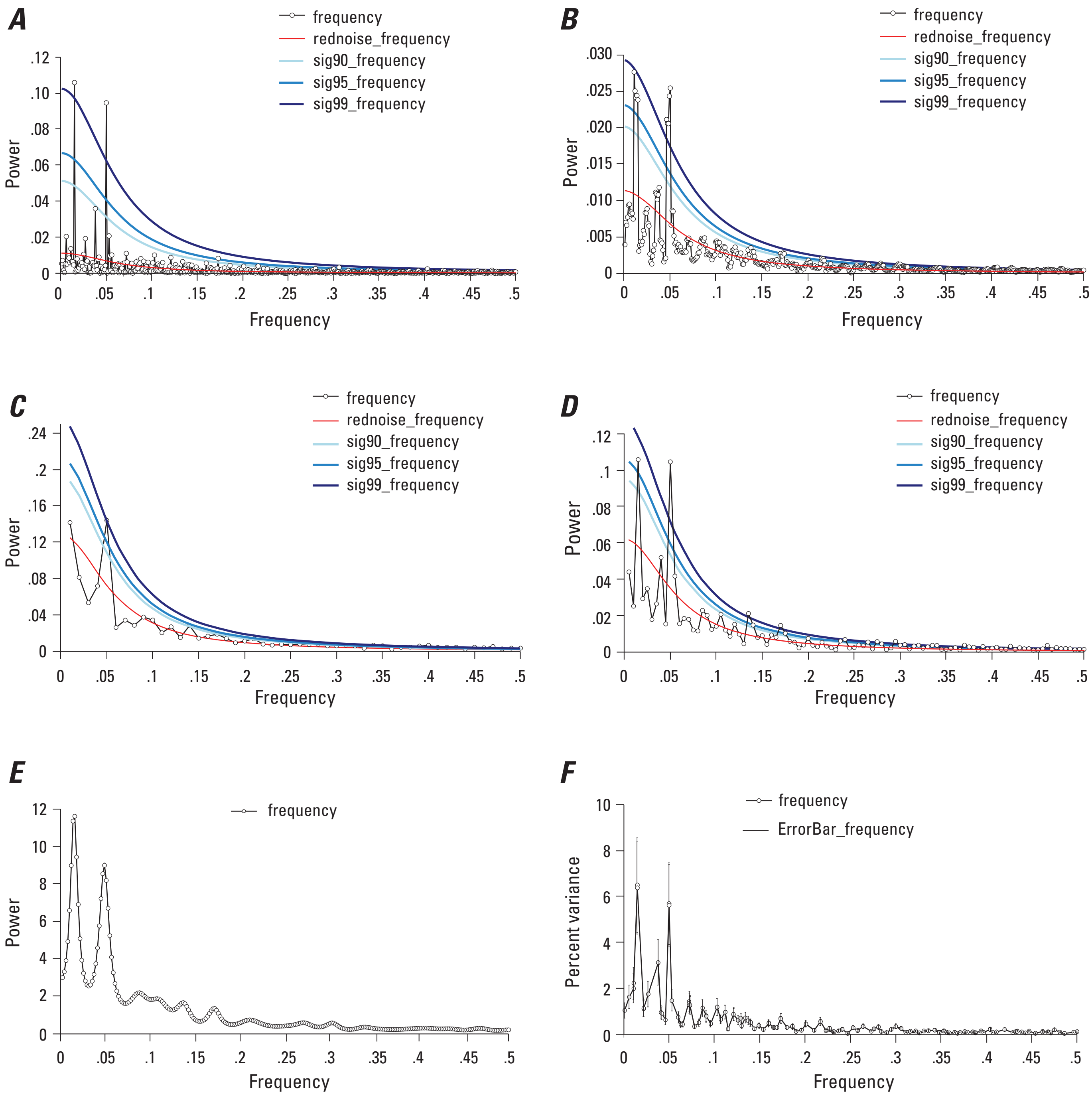


Figure 20. Spectra for a synthetic time series obtained by different methods: *A*, Discrete Fourier transform using a boxcar window and no spectral averaging; *B*, Discrete Fourier transform using a boxcar window and averaging adjacent spectral estimates; *C*, Discrete Fourier transform using a boxcar window and averaging of 10 spectral realizations; *D*, Discrete Fourier transform using a Hann window, averaging of nine overlapping spectral realizations; *E*, Maximum entropy method; *F*, Singular spectrum analysis.

The two main significant peaks obtained without using spectral averaging (fig. 20*A*) have returned, and the overall pattern of the spectrum resembles the spectrum obtained by using only spectral smoothing and a boxcar window (fig. 20*B*). Overall, the spectrum resembles the spectrum obtained without smoothing or averaging (fig. 20*A*), but has more degrees of freedom because of the averaging. This case also has two additional significant peaks at the 99 percent confidence level (using a chi-squared test) at frequencies of 0.135 and 0.17 that were not significant in the other tests.

Maximum Entropy Method

The resulting power spectrum indicates two main spectral peaks at different frequencies—the first one is centered near frequencies of 0.014 and 0.016 and the second one is centered near frequencies of 0.047, 0.048, and 0.042 (fig. 20*E*). The overall shape of the spectrum is very similar to the spectrum obtained by the discrete Fourier analysis.

Singular Spectrum Analysis

The spectrum obtained by SSA (fig. 20*F*) contains two main spectral peaks that were also identified by the discrete Fourier analysis and MEM. The first peak is at a frequency of 0.015 and the second is at 0.05. The first and second eigenvalues, as well as the third and fourth eigenvalues are nearly equal pairs. The sum of the reconstructed components for the first and second T-EOFs and T-PCs, which correspond to the first peak, explain approximately 13 percent of the total variance (fig. 19*A*). The second peak is related to the third and fourth T-EOFs and T-PCs and their reconstructed components (fig. 19*B*). A visual comparison between the spectra for the synthetic time series (fig. 19*A, B*) obtained by SSA, discrete Fourier analysis, and MEM indicate that, in general, two main oscillatory features dominate both the spectral power and percentage of variance. Differences among the spectra occur because of the use of different spectral estimation and averaging methods.

Example 2—Correlations Between Spatial and Temporal Modes in Winter Precipitation in the Southwestern United States to Winter Multivariate ENSO Index from 1980 to 2009

This example demonstrates how to use EOF analysis to extract spatial and temporal modes (patterns) in a gridded time series of precipitation data and how to use correlation to identify relations between the precipitation modes to ENSO. This type of analysis can be useful in other applications in which a physical process could be contributing to much of the variability in a gridded time series. Other examples of possible relations that can be assessed with these tools include (1) streamflow and remotely sensed vegetation data, (2) streamflow and gridded temperature data, and (3) groundwater withdrawals and remotely sensed land subsidence. This example uses the following tools:

1. Standardization by using the gamma to normal (SPI) tool.
2. Empirical orthogonal function analysis.
3. Linear regression.

Relations between winter precipitation in the southwestern U.S. and winter values of multivariate ENSO Index (MEI) are briefly explored for the period of 1980–2009. Spatial and temporal modes (patterns) in precipitation are extracted by using empirical orthogonal function analysis of precipitation that is first normalized using the standardized precipitation index (SPI; McKee and others, 1993). The relations between MEI and the principal components (PC) obtained by the EOF analysis are assessed by using linear regression and by evaluation of statistically significant correlations.

Winter precipitation for the months of October through February are compiled for the southwestern U.S. for the spatial extent of 43°N to 30.5°N and 121°W to 102°W and the temporal period of the water year 1980 (beginning October 1979) to 2009 (ending September 2009). Precipitation data are extracted from the Parameter-Elevation Regressions on Independent Slopes Model PRISM dataset (*http://www.prism.oregonstate.edu/*, accessed November 29, 2011). These data from PRISM are available at month and year intervals at 4-km resolution (0.042 degrees). For this example, precipitation data are resampled to a coarser 0.208 degree resolution (92 rows and 60 columns) because of computer memory limitations that can occur during the EOF analysis.

Prior to importing the data into the software, the precipitation values for each month are summed to represent the winter precipitation for a single year. The winter sum for each year includes the October, November, and December values for the preceding year. For example, the winter value for 1980 is the sum of the precipitation from October, November, and December of 1979, and the precipitation for January and February of 1980.

The data are saved into a single ASCII file, which has a single column (vector format) and 165,600 rows, and are organized by the pattern shown in figure 5 so that the software reads the data by using three loops. The first loop reads data in the eastward direction, the second loop reads the data in a northward direction, and the third loop reads the data for successive years. The first value in the file corresponds to the most westward and most southern point at the earliest time. The last value in the file corresponds to the most eastward and the most northern point at the latest time. For this example, the sizes of the loops are 92, 60, and 30 for the first, second, and third loop, respectively. The sizes of these loops correspond to the size of the 30 grids (one for each year) of precipitation data that are resampled to 92 rows and 60 columns. These loop sizes are entered into the form **Import matrix in vector format from ASCII file** at the textboxes labeled **First loop size**, **Second loop size**, and **Third loop size**. The data are assigned the data series name "ONDJF_total_precip" in the **Data Series Name** textbox.

The first processing step is to normalize the winter precipitation by using the standardized precipitation index (McKee and others, 1993) to characterize anomalous precipitation over the 30 years. The data are standardized by selecting "ONDJF_total _precip" in the drop-down list **Select series to standardize** and selecting the **Gamma to Normal (SPI)** standardization method on the **Preprocess Data** tab. The standardized data are saved to the output "ONDJF_total_precip_std." A time series of the standardized precipitation from a randomly selected spatial location can be plotted by clicking **Plot standardized series**. The plot can be used to check if the standardization produced reasonable values and if the data are imported correctly.

The EOF analysis is performed by selecting the standardized series "ONDJF_total_precip_std" in the drop-down list **Select series for EOF analysis** on the **EOF analysis** tab. The checkbox **Weight grids for latitude** is selected, 30.5 was entered in the textbox **Latitude minimum,** and 43 is entered in the textbox **Latitude maximum.** The EOFs are calculated by clicking the button **Calculate EOFs**.

The outputs from the EOF analysis include a spectrum of the spectral power and the percentage of variance explained for each EOF/PC pair, time series of the PCs, and scaled EOFs. The eigenvalue spectrum displays the spectral power and percentage of the variance explained by each EOF/PC pair (fig. 21). The North test (North and others, 1982) is used to assess the statistical significance of the EOF/PC pairs at the 95 percent level. For the North test, the sampling error (Δλ) is calculated for each eigenvalue and plotted as a length above and below the eigenvalue on the spectrum. Each year of data is assumed to be statistically independent, and the effective sample size is equal to 30. For the North test, an EOF/PC pair is significant if the error bars of its eigenvalue do not overlap vertically with the error bars of the neighboring eigenvalues. In this example, only the first two eigenvalues are separated from each other, and none of the other eigenvalues are separated vertically by their error bars. A null hypothesis of red noise could explain the EOFs if the percentage of explained variance by each eigenvalue decreased exponentially. A null hypothesis of white noise could explain the EOFs if the eigenvalues plot along a flat line on the spectrum. The North test indicates that the first two eigenvalues do not decrease exponentially, thus the null hypothesis that these are attributable to red noise is rejected at the 95 percent confidence level.

The first two combined EOFs for normalized Oct–Feb precipitation (figs. 22 and 23) explain approximately 60 percent of the total variance, and the remaining variance is explained by the higher modes. The first EOF for Oct–Feb and its PC explains 40 percent of the variance (fig. 22) and is negative everywhere in the map. The most negative areas are in the central part of the map, and the values become less negative toward the edges of the map. The second EOF for Oct–Feb captures 20 percent of the variance (fig. 23) and includes steep gradients in areas of high topography, such as in the Rocky Mountains of Colorado and Utah. Another pattern in the second EOF that reflects topography is a triangular area over the Mojave Desert area of southern California. In this second mode, the values are positive for much of the southern and eastern part of the study area and negative for the northwestern corner and in areas of high topographic areas in Colorado and Utah. Because this mode also has smooth gradients in areas of high topography, such as in New Mexico, influences other than orography are likely contributing to the patterns in this EOF map. The sign (positive or negative) of the patterns in the EOF maps and the PC time series is arbitrary in the EOF analysis, meaning that there is no physical reason for the sign. The magnitude, whether the value is positive or negative, however, is useful for assessing the strength of a spatial or temporal pattern.

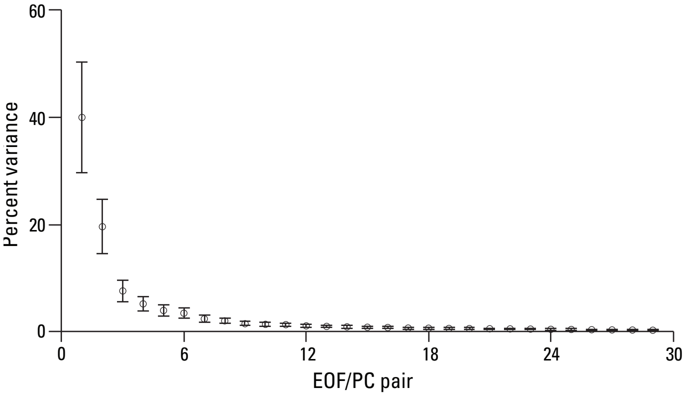


Figure 21. Panel showing spectrum of the eigenvalues from Empirical Orthogonal Function (EOF) analysis of Standardized Precipitation Index (SPI) normalized October through February precipitation in the southwestern U.S. from 1980 to 2009. PC, principal component.

The plots of the PCs of the first and second leading modes indicate how the strength of the first and second EOFs varies through time (fig. 24). A significant correlation of –0.41 at the 90 percent level (two tailed) between the first PC (composing 40 percent of the variance) with MEI for November through April suggest a teleconnection among winter atmospheric and oceanic conditions and precipitation for winter precipitation in the southwestern U.S. The negative correlation coefficient does not suggest that there is a negative correlation between a physical process that is identified by the first EOF and MEI. Because the sign of the EOF and PC is arbitrarily positive or negative, the correlation coefficient is also arbitrarily positive or negative. Previous investigators have identified positive correlation between precipitation and ENSO in the southwestern U.S. (for example, Redmond and Koch, 1991), and the results of the EOF analysis need to be interpreted in the context of physical processes. The sign of the values of the scaled EOFs and the time series values in the

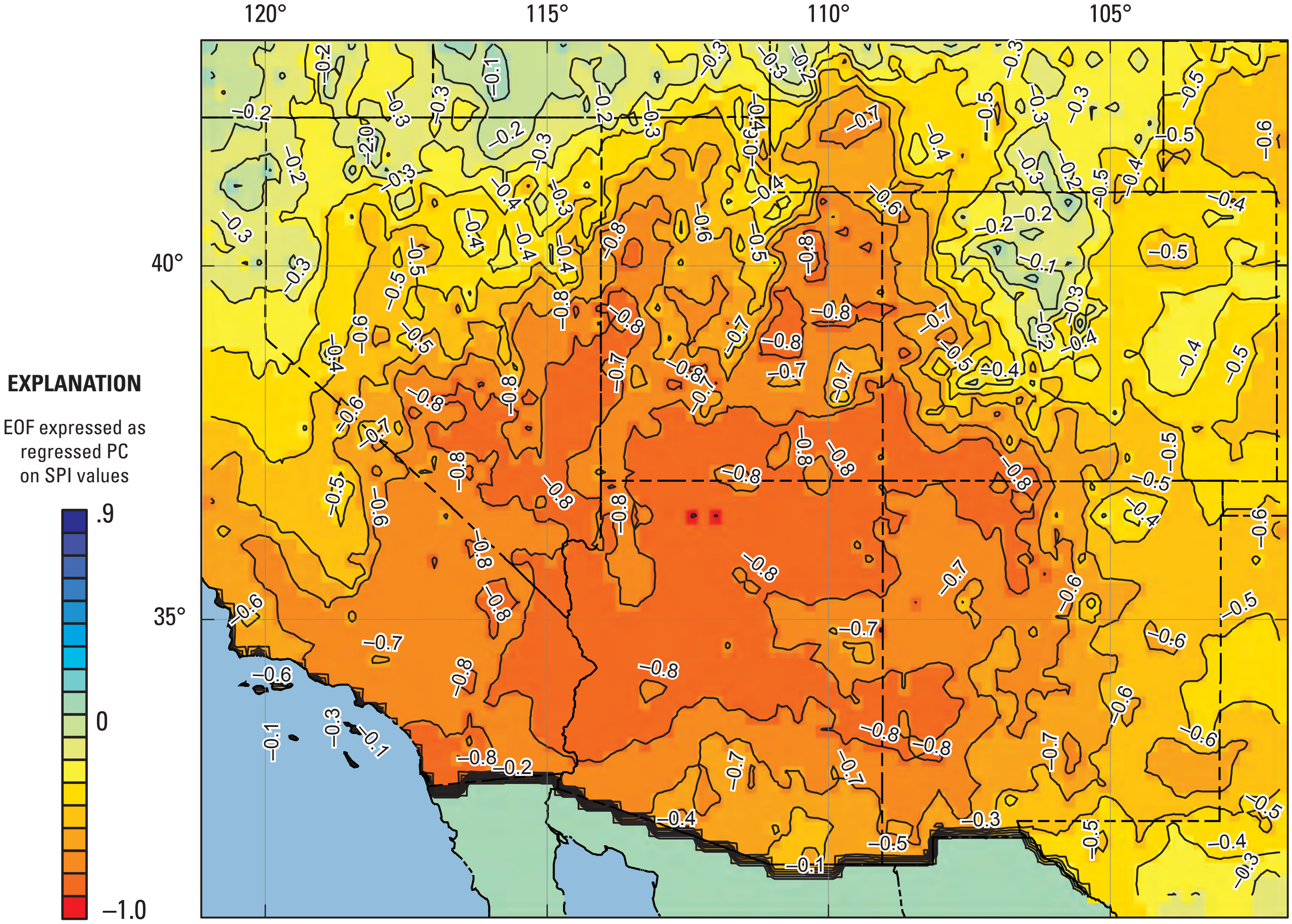


Figure 22. First scaled empirical orthogonal function (EOF) from analysis of standardized precipitation index (SPI) normalized October through February precipitation in the southwestern U.S. from 1980 to 2009. PC, principal component.

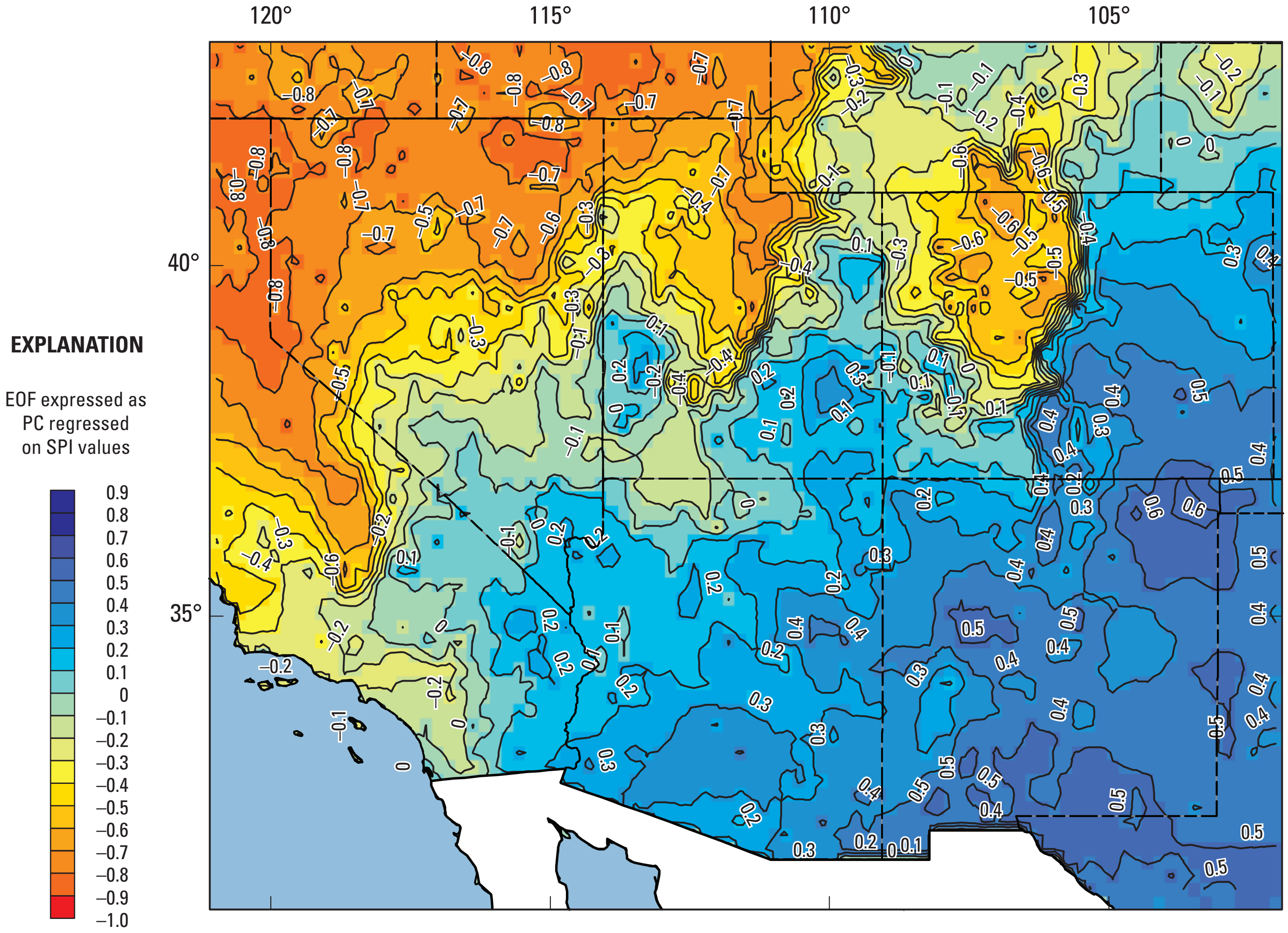


Figure 23. Second scaled empirical orthogonal function (EOF) from analysis of standardized precipitation index (SPI) normalized October through February precipitation in the southwestern U.S. from 1980 to 2009. PC, principal component.

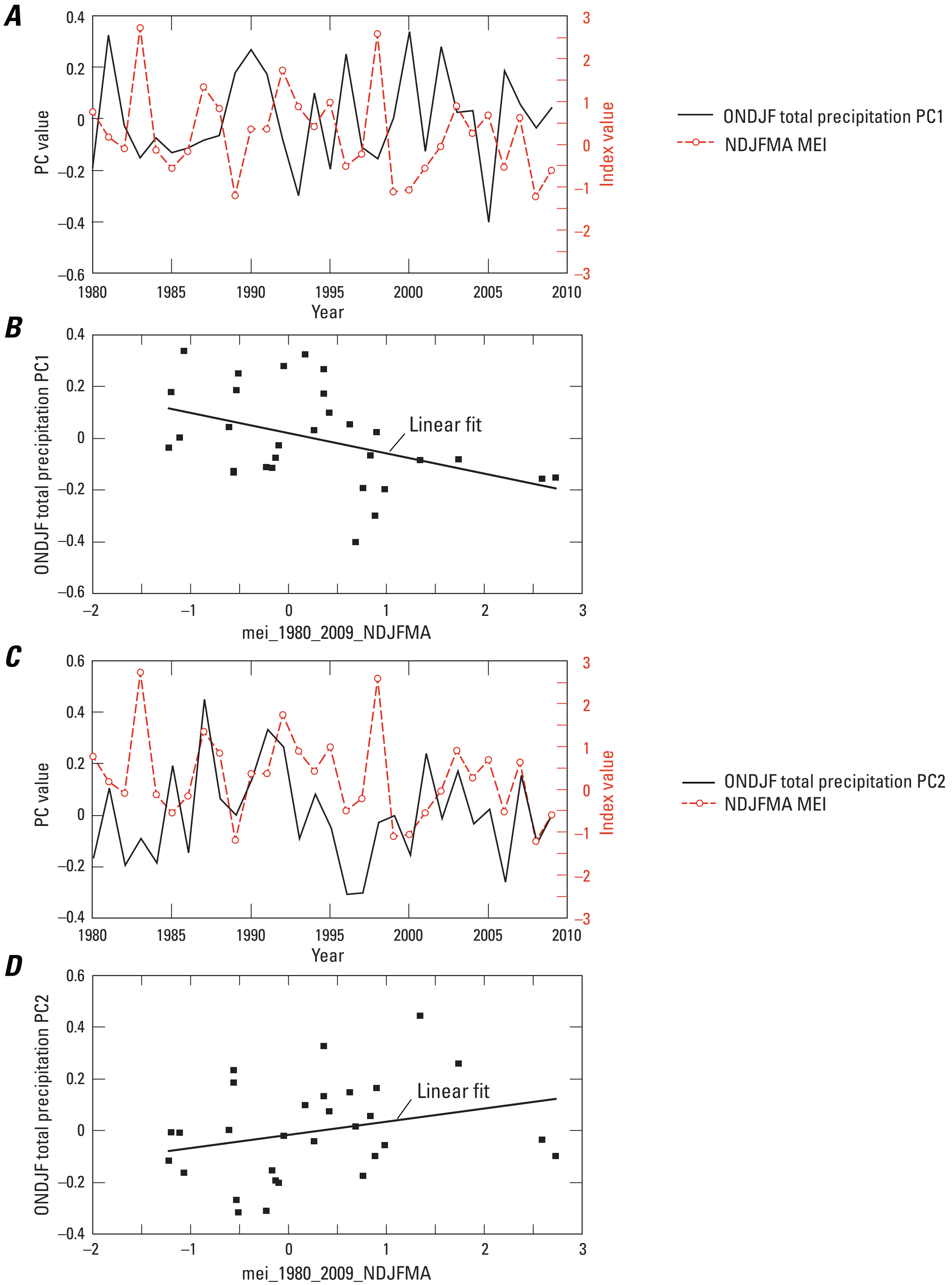


Figure 24. Results of the empirical orthogonal function (EOF) analysis of the standardized precipitation index (SPI) normalized October through February precipitation from 1980 to 2009 as *A*, the first principal components (PC) time series; *B*, a scatter plot and linear fit of the first PC and Oct–Feb values of Multivariate ENSO Index (MEI); *C*, the second PC time series; *D*, a scatter plot and linear fit of the second PC and values of MEI.

first EOF and PC pair can be switched so that these match the expected relation between this pattern and ENSO. The correlation coefficient of 0.29 between the second PC and MEI was not significant at the 90 percent level (two tailed), suggesting that the physical process identified by the second EOF and PC pair is not physically related to MEI.

Example 3—Assessing Temporal Relations Between the Multivariate ENSO Index and Hydrologic Time Series in the Upper San Pedro Basin, Southeastern Arizona

This example demonstrates how to use lag correlation analysis to identify a temporal relation between streamflow discharge and groundwater levels to ENSO. Lag correlation analysis is useful for determining the time lag between a potentially causal physical process and an observed response. In many complex hydrologic systems, the time lags for such causal relations are generally undetermined, and this type of analysis provides insight into a system without a dynamic model. Other examples of possible relations that could be assessed with these tools include the lag between (1) changes in streamflow and precipitation or groundwater withdrawals, (2) temperature and vegetation responses, and (3) runoff events and changes in water quality.

This example uses of the following tools:

1. Transformation by using the cumulative departure tool.
2. Detrending by using curve fitting.
3. Standardization by using the normal distribution.
4. Lag correlation.

Relations between streamflow and groundwater levels in the Upper San Pedro basin and ocean or atmosphere phenomena, as indicated by ENSO events, are briefly explored here by using a lag correlation analysis. The Multivariate ENSO Index (MEI; Wolter and Timlin, 2011), which is an index of six variables (sea-level pressure, surface air and sea-surface temperatures, zonal and meridional wind components, cloudiness fraction) in the tropical Pacific, is correlated, at different monthly lags, to monthly streamflow and groundwater-level data in the San Pedro River from 1980 to 2009. The monthly discharge time series at the San Pedro River at Charleston, Arizona (09471000; fig. 25*A*) contains long periods of low flow that are less than 20 cubic feet per second (cfs). These low flows are primarily groundwater discharge to the stream channel as base flow. Large discharge events (from 20 cfs to 650 cfs) generally occur after runoff-producing precipitation from frontal systems or cutoff, low-pressure systems in the winter and from monsoonal convective systems in the summer.

The monthly discharge time series at the San Pedro River at Charleston, Arizona (09471000) (fig. 25*A*), contains long periods of low values near zero that are separated by large flow events. For this example, the low values are removed by transforming the time series into a monthly cumulative departure (fig. 25*B*). The cumulative departure transformation also adds persistence to the time series, which can indicate intra-annual or interannual periods in which discharge was either greater than or less than the average value. In general, wet climatic periods are indicated by the rising limb, and dry climatic periods by the falling limb, of the cumulative departure curve.

A trend in the discharge time series could be related to anthropogenic effects, or part of a low-frequency climatic forcing that cannot be resolved within the 30-year time series. A trend is estimated by fitting a cubic polynomial to the cumulative departure time series (fig. 25*C*). Differences in the time series from the trend are obtained by subtracting the cumulative departure series from the polynomial fit (fig. 25*D*). The residuals are standardized using a normal distribution to allow for comparison between the series and the normalized MEI index (fig. 25*E*).

Well D-23-21 06CCC2 is within 30 meters of an ephemeral channel that routes runoff from the nearby Huachuca Mountains to the San Pedro River. The groundwater levels appear to fluctuate in response to time-varying recharge rates that could be related to variable precipitation (fig. 26*A*). The temporal relation between MEI to both discharge and the water levels in the well are assessed by comparing the correlation coefficients at the different lags. This identifies the lags having the greatest correlation, which can indicate how much time passes before the oceanic and atmospheric processes that drive responses in MEI could produce responses in streamflow and groundwater levels.

Several preprocessing steps are completed before the lag correlation analysis between MEI and two hydrologic time series. First, the number of values in the two time series used in the lag correlation analysis must be equal. Often, both time series are chosen to be coincident in time at a lag of zero, and each lag is the amount of time that the time series are shifted away from being coincident. If the time series are measured over the same interval of time, then the values are often obtained at some uniform interval, such as the case of daily, monthly, or annual values. This would ensure that the two time series have the same number of values. However, a time series of water levels in wells are often measured at irregular times and have non-uniform time intervals between measurements. In order to allow for correlation with another time series with uniform intervals, interpolation can be used to generate values at some uniform interval that is coincident with the other time series. For this example, the water-level records for wells D-23-21 06CCC1 and D-23-21 06CCC2 (fig. 26*A*), which are

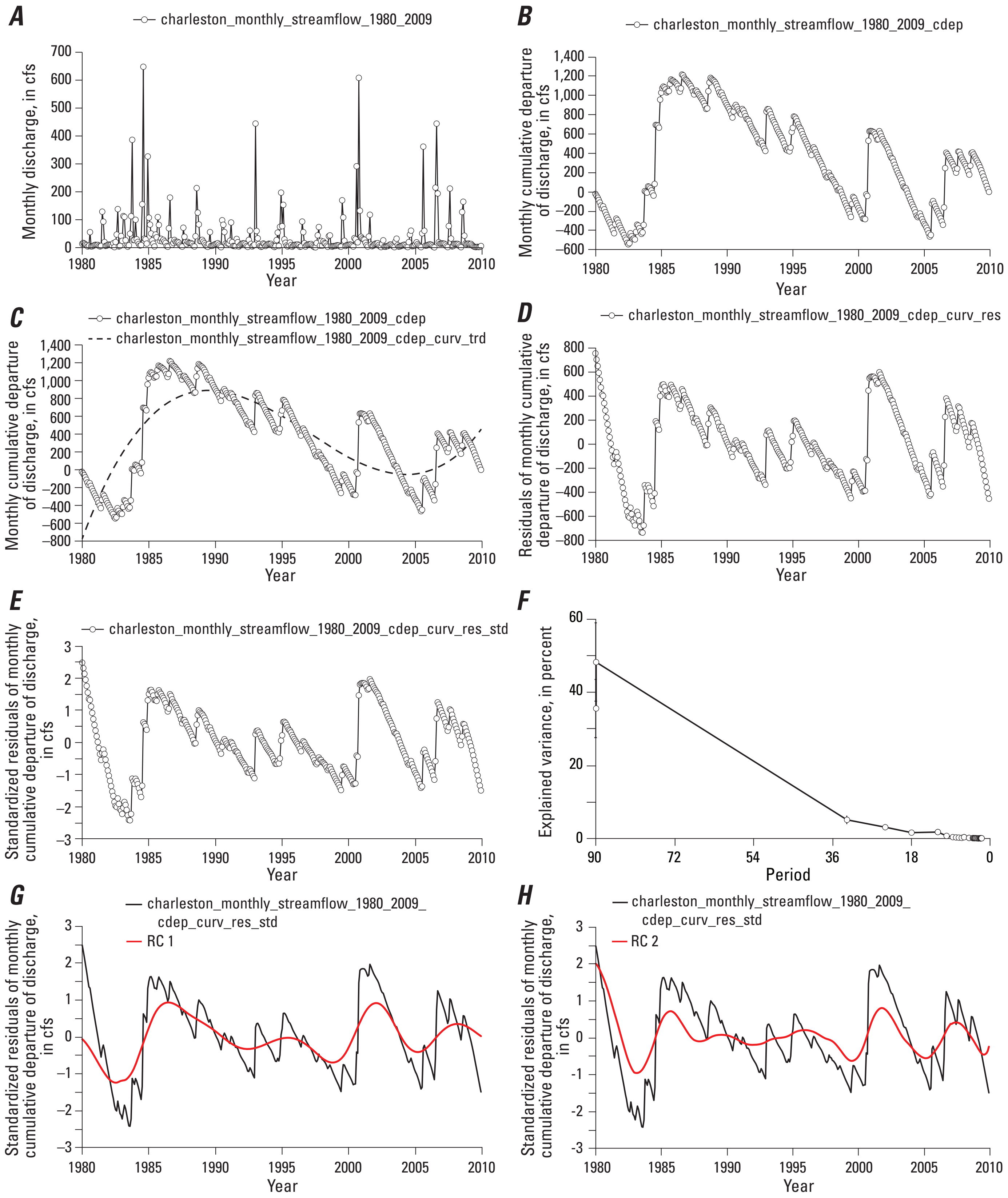


Figure 25. The steps for processing the discharge record at San Pedro River at Charleston, Arizona (09471000), from 1980 to 2009: *A*, the monthly discharge values; *B*, the monthly cumulative departure of the time series; *C*, a fitted cubic polynomial to the cumulative departure; *D*, the residuals from the fitted cubic polynomial; *E*, the residuals after standardization using a normal distribution; *F*, the power spectrum of the standardized residuals obtained by SSA; and *G* and *H*, the first and second reconstructed components obtained by SSA.

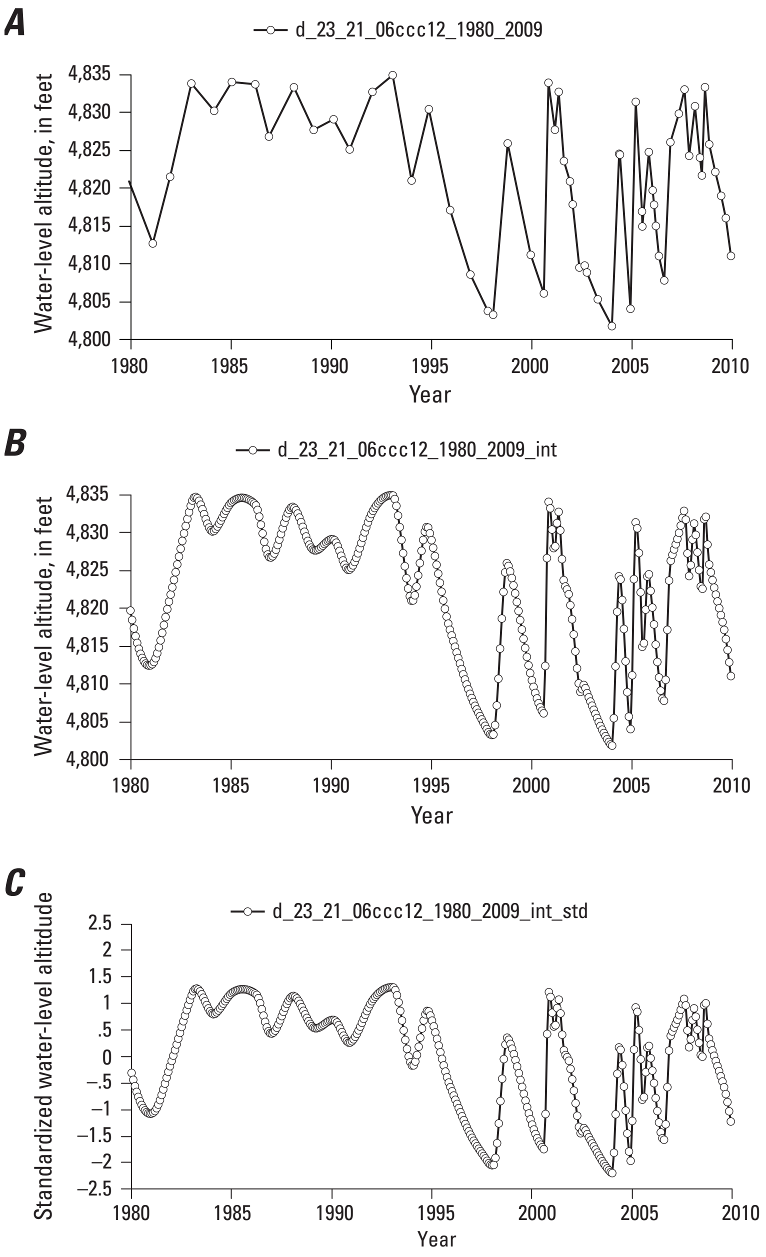


Figure 26. The steps for processing the water level record at well D-23-21 06CCC1,2: *A*, the water level time series in elevation above NGVD29; *B*, the interpolated time series; and *C*, the standardized time series using a normal distribution.

in close proximity, were combined into a single time series for the period of 1980–2009. The Akima spline tool is used to generate monthly values from 1980 to 2009, and the resulting series 360 values (fig. 26*B*). The splined time series is standardized using a normal distribution to obtain a series that can be compared visually to the normalized MEI index (fig. 25*C*).

The cross-correlation function (CCF) for MEI and cumulative departure of streamflow at Charleston (fig. 27*A*) and for MEI and water levels at well D-23-21 CCC1,2 is used to explore how the correlation varies as a function of lag. The CCF is also used to identify the lag, from 0 to 60 months, at which the positive correlation coefficient is greatest. The positive correlation is of interest because of a prior expectation that increased precipitation during periods of positive MEI will produce larger discharge values. For cumulative departure of streamflow, the correlation coefficient is negative from 0 to 32 months, and is significant at the 95 percent confidence interval up to 31 lags. The correlation is near zero for lags 32 and 33, then becomes positive until a lag of 57, after which it is near zero. The maximum positive correlation at a lag of 43 months (3.6 years) suggests that long-term changes in the cumulative departure of streamflow do not occur rapidly (within several months) with atmospheric conditions, but could be influenced by the storage properties in the groundwater system. Pool (2005) reported that in southeastern Arizona, variations in groundwater recharge are related to ENSO events, in which greater amounts of precipitation generally correspond to El Niño conditions, while lesser precipitation conditions generally correspond to La Niña conditions. As a result of the increased precipitation, groundwater recharge rates are three times greater during the period of frequent El Niño conditions (1977–98) than during a period of frequent La Niña conditions (1941–57; Pool, 2005). The lag at the maximum positive correlation could be related to the time required for the stored groundwater from increased recharge to propagate through the aquifer system before it discharges as base flow in the San Pedro River. Several major features of the lagged cumulative departure series appear to coincide with rapid changes in the MEI index, such as a large increase in both series in 1983 and 1998 (fig. 27*B*).

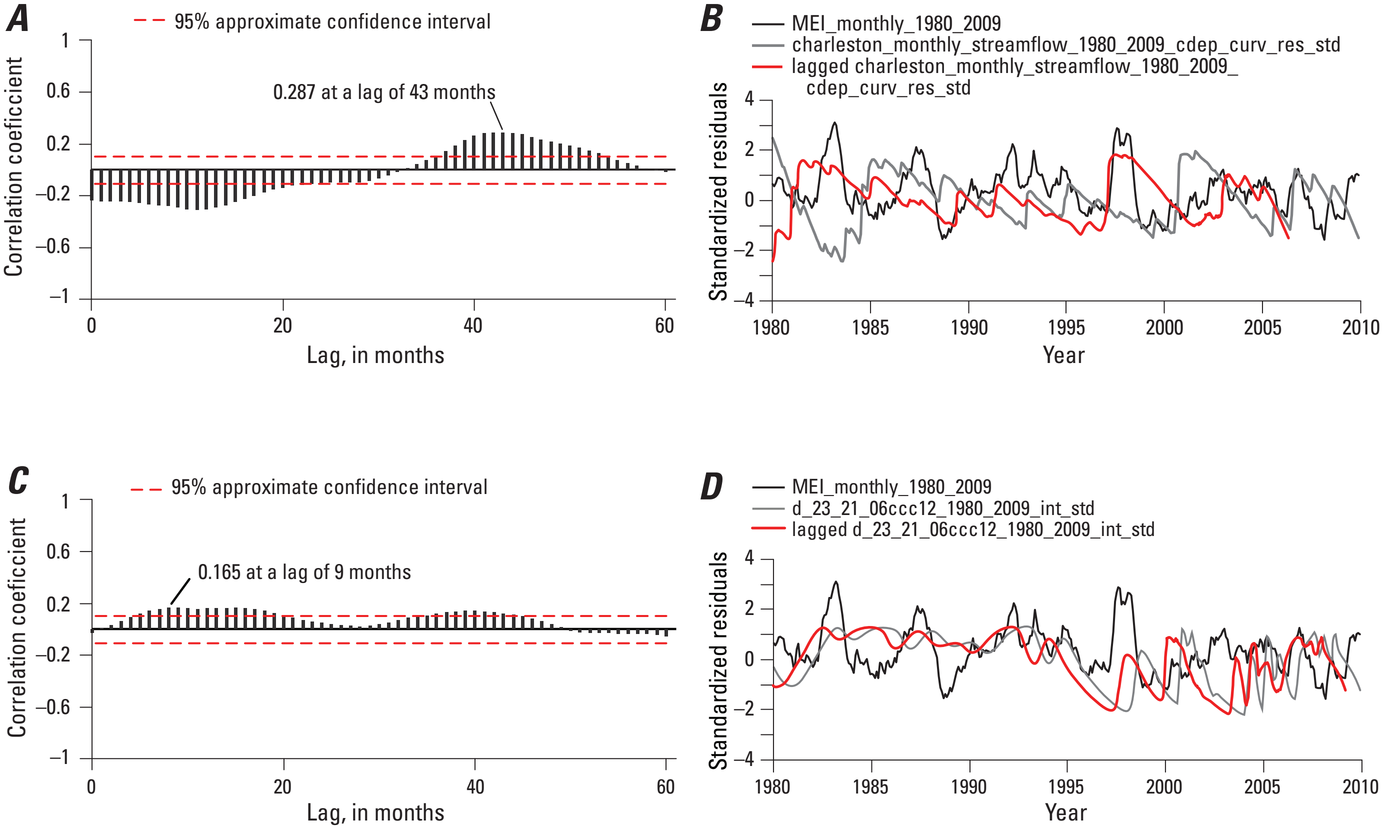


Figure 27. *A*, The cross correlation function plot between monthly values of normalized residuals of cumulative departure of discharge at San Pedro River at Charleston and the multivariate ENSO index (MEI) from 1980 to 2009; *B*, the discharge record at San Pedro River at Charleston, MEI, and the discharge record lagged 43 months; *C*, the cross correlation plot between monthly values of normalized residuals of water levels at well D-23-21 06CCC1,2 and MEI; *D*, the water level record at D-23-21 06CCC1,2, MEI, and the water level record lagged 9 months.

The CCF for MEI and the water levels in well D-23-21 06CCC2 indicates that the correlation is generally positive from lags of 0 to 60 months (fig. 27*C*). The lag at the maximum positive correlation is 9 months (fig. 27D), which is shorter in time than the lag for streamflow at Charleston. This shorter lag could indicate that infiltration and recharge of runoff-producing precipitation related to ENSO occurs on a scale of several months near the well. The well is within 30 meters of a large ephemeral wash that routes runoff from the Huachuca Mountains. ENSO events in 1983 and 1998 also correspond with increases in the lagged groundwater-level time series.

Example 4—Projecting Streamflow in the San Pedro River, Southeastern Arizona

This example demonstrates a method for generating a projection of streamflow at the San Pedro River at Charleston, Arizona (09471000). The projection is based on an autoregressive time series model that was fitted to the observations over part of the record. Projections of time series are useful for generating inputs and boundary conditions in predictive models. These methods are based on the projection methods described by Keppenne and Ghil (1992) and demonstrated for the Southern Oscillation Index, and as applied by Hanson and others (2003) for precipitation in the Santa Clara-Calleguas Basin, California. While the range of data types that can be projected is practically unlimited, potential hydrologic examples include (1) water quality data, (2) vegetation indices, and (3) groundwater withdrawals. This example uses the following tools:

1. Singular spectrum analysis (SSA).
2. Autoregressive modeling (AR).
3. Projection of autoregressive models.

The time series at Charleston from 1980 to 2010 was used to project streamflow for 15 years from 2011 to 2025 by using several steps of transformation of the streamflow time series, autoregressive (AR) time-series modeling, and extrapolation of two reconstructed components obtained by SSA. This streamflow time series was used in example 3. The steps for generating the cumulative departure time series and obtaining the residuals from the trend are also explained in the description for example 3. These processing steps were completed in order to obtain and project the long-term changes in the cumulative departure and to remove low frequency components that typically explain most of the variance of the series and that dominate the RCs obtained through SSA.

The first step in the projection was to obtain quasi-periodic modes from the residuals by using SSA. The modes are useful for generating a projection because the modes usually vary regularly within a narrow frequency range and are generally more predictable than the original time series. On the tab **SSA**, the time series of residuals named "Charleston_monthly_streamflow_1980_2009_cdep_curv_res_std" was selected in the drop-down list **Select series for SSA**. Using a default window length of 36 (one tenth of the residual time series length), 83 percent of the total variance could be explained by two oscillatory modes—the first mode explains 48 percent of the variance, the second mode explains 35 percent, and the remaining 15 percent of the variance is explained by modes that are not directly used in the projection. The reconstructed components of the first and second modes (RC1 and RC2) have the same length as the original time series, and both oscillated at approximately a period of 90 months (7.5 years). RC1 and RC2 were summed to create a single series (RC12) having a period of 90 months (fig. 28*A*). These are summed by entering "1 2" in the text box **Select RCs**.

The next step in the analysis is to obtain an AR(p) time-series model to RC12 that can be used to generate future values of the time series. A first step in creating an AR(p) model is to estimate the order p of the model, and then to fit the coefficients of the model by using a least squares minimization of the residuals between the simulated and observed values. In this analysis, models having orders of 1 up to 180 (half the length of the time series) are evaluated by using the corrected Akaike information criteria (AICc), which calculates a goodness-of-fit statistic that is penalized by the complexity of the model. Another test of the fit is that the residuals of the time series from the AR model are random, and the residuals are not autocorrelated in time. On the tab **Projections**, the time series "RC12" is selected in the drop-down list **Select series to model.** The lowest AICc statistic is produced when using an order of 141. However, the simulated values are either less or greater than the observations over continuous intervals of the time series (fig. 29*A*). Instead, an order of 120 is selected so that the projection is based on a considerable number of values from the previous 10 years and so that the simulated values have better agreement with the observations (fig. 29*B*). For the final 10 years, the standardized residuals of cumulative departure of streamflow contains two periods of increasing values and one period of low values. The autocorrelation function plot of the residuals indicates that the residuals become less autocorrelated within a smaller lag distance from a lag of zero for the AR(120) model (fig. 29*C*) than the AR(141) model (fig. 29*D*).

After selecting the order of the AR model, the model can be projected for additional time steps by using a combination of stepping forward in time and random number generation. Numerous projections can also be done to create an envelope of realizations, and each one is equally as plausible. The model is stepped forward in time by selecting a starting time $t - 1$. The software automatically selects $t - 1$ to be the final value of the simulated time series. In this example, $t - 1 = 2009.92$ is the decimal year for December 2009, and $t = 2010.00$ is the decimal year for January 2010. A random number is sampled from a normal distribution, in which the mean and variance parameters are equal to the sample mean and sample variance of the residuals between RC12 and the simulated time series.

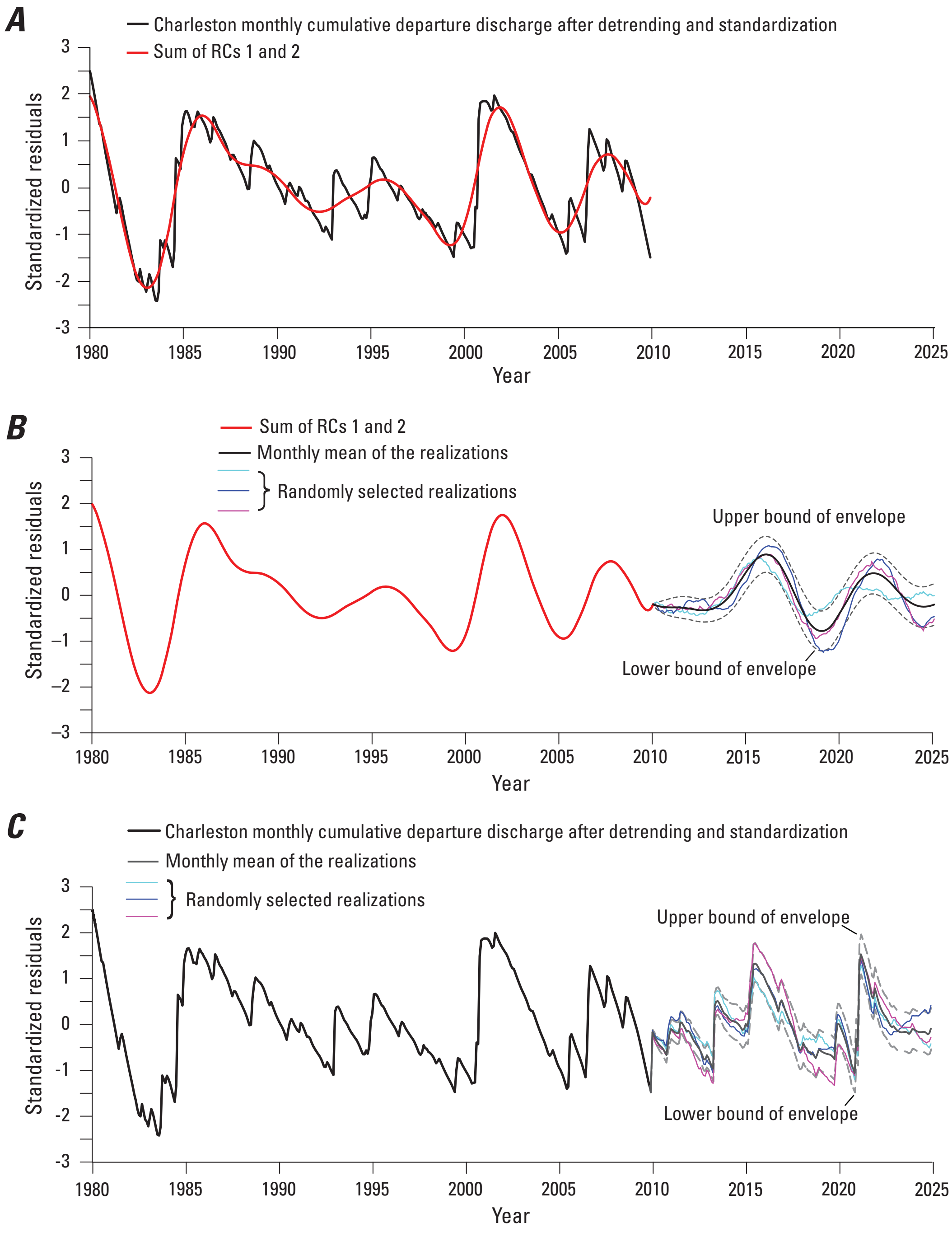


Figure 28. *A*, The processed monthly discharge at San Pedro River at Charleston described in example 3, and the sum of the first two reconstructed components (RC); *B*, projected discharge from 2010 to 2025 without adding the residuals between the processed monthly discharge and the sum of the first two reconstructed components; *C*, the projected discharge after adding the residuals.

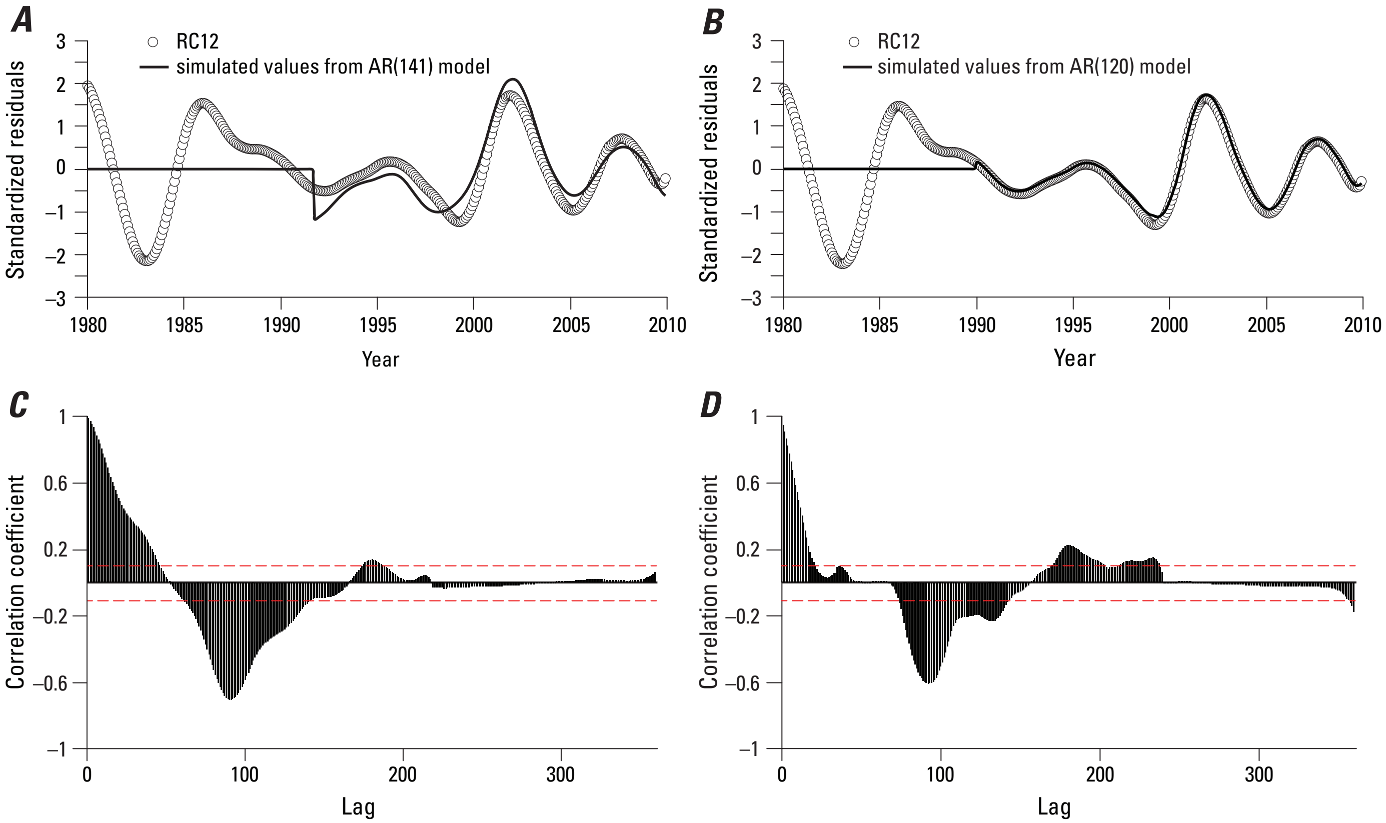


Figure 29. *A*, The sum of the first two RCs and the simulated values from an autoregressive AR(141) model; *B*, the sum of the first two RCs and the simulated values from an AR(120) model; *C*, the autocorrelation function of the residuals shown in *A*; and *D*, the autocorrelation function of the residuals shown in .

For each projected time, a new random number is sampled and added to the series. Additionally, 1,000 separate AR models were projected to create a range of possible conditions.

The projection of streamflow cumulative departure based on the 1,000 different realizations contains quasi-periodic characteristics of RC12, including a period of approximately 7.5 years and similar amplitude (fig. 28*B*). The projection of RC12 does not contain all of the variability contained in the cumulative departure time series of streamflow, however. Following the method described by Hanson and others (2003), the remaining variance is added by creating a separate time series of residuals of the cumulative departure time series from RC12. The starting time for the time series of added residuals is selected randomly from between the start and end date of the modeled time series and restarts at the beginning of the series if the time extends past the end date of the modeled series. The resulting projection of streamflow cumulative departure (fig. 28*C*) contains similar features as the original modeled time series, including a continuation of wetting and drying that repeat at approximately 7–8 years. The addition of residuals adds the remaining variance that was not explained by RC1 and RC2, and the total variability of the projection overall appears to be plausible.

Evaluation of Coded Procedures

The coded procedures in the software were evaluated by comparison to analytical solutions when these were available, a statistical test to ensure that the results have a required statistical property, or to other published software that perform similar types of analysis. The following sections describe how each method was evaluated.

Standardization

Standardized series were evaluated by checking for a mean of zero and a standard deviation of one.

Interpolation

The interpolation tool was evaluated by visual inspection of interpolated values along linear and nonlinear time series.

Cumulative Departure

Visual comparison was used to ensure that periods of greater-than-average values in a time series coincided with periods during which the cumulative departure had a positive slope. Similarly, periods of less-than-average values in the time series were also evaluated for coincident periods of a negative slope in the cumulative departure series. The tool was also evaluated to ensure that the final value was always zero, which is required in a cumulative departure time series.

Differencing

The differencing tool was evaluated by comparison with results produced by software by Dr. David Meko at the University of Arizona. The comparison used the same datasets to ensure that the results produced by both programs were reasonably similar.

Curve Fitting

The curve fitting tool was evaluated by comparison with results produced by the software Excel®. The comparison used the same datasets to ensure that the results produced by both programs were reasonably similar.

Discrete Fourier Transform

The discrete Fourier transform tool was evaluated by comparing the output spectrum to the spectrum produced by the SSA-MTM toolkit (Dettinger and others, 1995b). The time series used for the evaluation comprised a sum of sinusoidal functions having different frequency and amplitude. The spectra from the SSA-MTM toolkit were produced by using MEM and SSA. Because the time series was simple, the resulting spectra were nearly identical. The windowing and averaging features are not available in the SSA-MTM toolkit, and the testing of the results from these features was limited to a visual comparison of the output spectra to the spectra produced by the SSA-MTM toolkit.

Maximum Entropy Method

The maximum entropy method tool was evaluated by comparing the output spectrum to the spectrum produced by the SSA-MTM toolkit (Dettinger and others, 1995b). The time series was identical to the time series used for evaluating the discrete Fourier transform tool and the time series in example 1. The spectrum from the SSA-MTM toolkit was also produced by MEM, and the results for both tested time series were identical.

Singular Spectrum Analysis

The singular spectrum analysis tool was evaluated by comparing the output spectrum to the spectrum produced by the SSA-MTM toolkit (Dettinger and others, 1995b). The tested time series was identical to the time series used for evaluating the discrete Fourier transform tool and the time series in example 1. The spectrum from the SSA-MTM toolkit was also produced by SSA, and the results for both tested time series were identical.

Empirical Orthogonal Function Analysis

The empirical orthogonal function analysis tool was evaluated by comparing the output EOF maps to published maps in previous investigations (for example, Wallace and Gutzler, 1981). Results of the EOF tool were also compared to the results of homework sets for the “Objective Analysis in the Atmospheric and Related Sciences” course presented by Dr. Christopher Castro in the Department of Atmospheric Sciences at the University of Arizona. The authors of this report are not aware of any commercially available software packages that perform EOF analysis that could have been used to evaluate the results of the EOF tool.

Linear Regression and Correlation

The linear regression and correlation tools were evaluated by comparing the fitted linear models from example 3 to the fitted linear models produced by the software Excel©. HydroClimATe and Excel© produced reasonably similar results.

Projections

The autoregressive model (AR) generation and selection procedures were evaluated by comparison to the results from software by Dr. David Meko at the University of Arizona. The generated projections produced by the AR models were not evaluated by comparing with independently-produced results. Instead, the projections were evaluated by ensuring that the mean and variance of the projected series were preserved in the projected values.

Summary

This report documents the software package HydroClimATe (Hydrologic and Climatic Analysis Toolkit), which automates the use of several objective methods for assessing relations between climate variability and variability in hydrologic time series. The methods include standardization, detrending, regression and correlation, Fourier analysis, maximum entropy method, singular spectrum analysis, Empirical Orthogonal Function analysis, and autoregressive time series modeling. These tools have been used extensively to identify relations between climatic indicators and meteorological and hydrologic data to hydrologic conditions in previous HydroClimATeic investigations (Dettinger and others, 1995a; Dettinger and Diaz, 2000; Dickinson and others, 2004; Hanson and others, 2004, 2006; Gochis and others, 2007a, b; Kumar and Duffy, 2009). A possible advantage of this software is that it presents these methods in a sequential order that can be useful for evaluating relations between hydrologic and climatic time series. The software includes all of the methods for assessing relations between climatic and hydrologic time series described by Hanson and others (2004) and implemented by Hanson and others (2003).

HydroClimATe includes tools for (1) identifying responses of hydrologic systems to climate variability; (2) quantifying statistical relations between multiple time series and climate indices, such as the Multivariate Enso Index (Wolter and Timlin, 2011); and (3) projecting hydrologic time series by using time-series models and spectral analysis. The software consists of a graphical user interface that is executable in Windows operating system with the .NET Framework version 4.0. Software inputs can be any long-term time-series data, such as groundwater levels, streamflow, precipitation, tree-ring data, air temperature, and climate indices. Other types of time-series data, such as economic data, could be applicable. The methods of analyses are demonstrated in this report for climatic and hydrologic time series, however. Software output can be exported to files that are read by a text editor, Microsoft Excel®, or geographic information system (GIS) software. Example analyses are presented for assessing relations between global climate indices and hydrologic time series for sites in the southwestern U.S.

The results from the tools were mainly evaluated by comparison to results from other available software packages or for simple analytical examples where the result is already known. The SSA-MTM toolkit (Dettinger and others, 1995b) was used extensively to evaluate the results of the discrete Fourier transform, maximum entropy method, and singular spectrum analysis tools. In some cases, other software packages for these tools were not available and the tools were tested by comparing the results to the results that were expected for certain data sets.

References Cited

Bakker, M. and Nieber, J.L., 2009, Damping of sinusoidal surface flux fluctuations with soil depth: Vadose Zone Journal, v. 8, no. 1, p. 119–126.

Brockwell, P.J. and Davis, R.A., 2002, Introduction to time series and forecasting (2d ed.): New York, Springer, 434 p.

Broomhead, D.S. and King, G.P., 1986, Extracting qualitative dynamics from experimental data, Physica D: nonlinear phenomena, 20, p. 217–236.

Burg, J. P, 1967, Maximum entropy spectral analysis, paper presented at the 37th Annual International Meeting of the Society of Exploration Geophysicists, Oklahoma City, Okla., 1967 *in* Modern Spectrum Analysis, 1978, Childers, D.G., ed., Piscataway, N. J., IEEE Press, p. 42–48

Campbell, B.G., and Coes, A.L., eds., 2010, Groundwater availability in the Atlantic Coastal Plain of North and South Carolina: U.S. Geological Survey Professional Paper 1773, 241 p., 7 pls.

Castro, C.L., Beltrán-Przekurat, A.B. and Pielke, R.A., Sr., 2009. Spatiotemporal variability of precipitation, modeled soil moisture, and vegetation greenness in North America within the recent observational record: J. Hydrometeor., v. 10, p. 1355–1378.

Childers, D. G. (Ed.) 1978, Modern Spectrum Analysis: Piscataway, N. J., IEEE Press, 331 p.

Clark, B.R., Hart, R.M., and Gurdak, J.J., 2011, Groundwater availability of the Mississippi Embayment: U.S. Geological Survey Professional Paper 1785, Reston, Va, 62 p., *http://pubs.usgs.gov/pp/1785/*

Dai, A., and Wigley, T.M.L., 2000, Global patterns of ENSO-induced precipitation: Geophysical Research Letters, v. 27, no. 9, p. 1283–1286.

Dettinger, M.D, and Diaz, H.F, 2000, Global characteristics of stream flow seasonality and variability: Journal of Hydrometeorology, v. 1, p. 289–310.

Dettinger, M.D., Ghil, M., and Keppenne, C.L., 1995a, Interannual and interdecadal variability in United States surface-air temperatures, 1910–1987: Climatic Change, v. 31, p. 35–66.

Dettinger, M.D., Ghil, M., Strong, C.M., Weibel, W., and Yiou, P., 1995b, Software Expedites Singular-Spectrum Analysis of Noisy Time Series: EOS, Transactions of the American Geophysical Union, v. 76, 2 p.

Dickinson, J.E., Hanson, R.T., Ferré, T.P.A., and Leake, S.A., 2004, Inferring time-varying recharge from inverse analysis of long-term water levels: Water Resources Research, v. 40, no. 7, p. 1–15.

Edwards, D.C., and McKee, T.B., 1997, Characteristics of 20th century drought in the United States at multiple time scales: Fort Collins, Colorado, MS Thesis, Colorado State University, 155 p.

Faunt, C.C., ed., 2009, Groundwater Availability of the Central Valley Aquifer, California: U.S. Geological Survey Professional Paper 1766, 225 p.

Ghil, M., and Mo, K.C., 1991, Intraseasonal oscillations in the global atmosphere. Part I: Northern hemisphere and tropics: *J. Atmos. Sci., v. 48,* no. 5, p. 752–779.

Ghil, M., Allen, M.R., Dettinger, M.D., Ide, K., Kondrashov, D., Mann, M.E., Robertson, A.W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P., 2002, Advanced spectral methods for climatic time series: Reviews of Geophysics, v. 40, no. 1, p. 3-1–3-41.

Gleick, P.H., and Adams, D.B., 2000, Water: The Potential Consequences of Climate Variability and Change for the Water Resources of the United States: The report of the Water Sector Team of the National Assessment of the Potential Consequences of Climate Variability and Change for the U.S. Global Research Program, Pacific Institute for studies in Development, Environment, and Security, Washington D.C., 151 p.

Gochis, D.J, Brito-Castillo, L., and Shuttleworth, W.J., 2007a, Correlations between sea-surface temperatures and warm season streamflow in northwest Mexico: International Journal of Climatology, v. 27 p. 883–901.

Gochis, D.J, Brito-Castillo, L., and Shuttleworth, W.J., 2007b, Hydroclimatology of the North American Monsoon region in northwest Mexico: Journal of Hydrology, v. 316, p. 53–70.

Green, T., Taniguchi, M., Kooi, H., Gurdak, J.J., Hiscock, K., Allen, D., Treidel, H., and Aurelia, A., 2011, Beneath the surface of global change: Impacts of climate change on groundwater: *Journal of Hydrology, v.* 405, p. 532–560, doi:10.1016/j.jhydrol2011.05.002.

Gurdak, J.J., Hanson, R.T., McMahon, P.B., Bruce, B.W., McCray, J.E., Thyne, G.D., and R.C. Reedy, 2007. Climate variability controls on unsaturated water and chemical movement, High Plains aquifer, USA: *Vadose Zone Journal, v.* 6, no.2, p. 533–547, doi: 10.2136/vzj/2006.0087.

Gurdak, J.J., Hanson, R.T., and Green, T.T., 2009, Effects of Climate Variability and Change on Groundwater Resources of the United States: U.S. Geological Survey Fact Sheet FS09–3074, 4 p.

Hanson, R.T., Martin, P.and Koczot, K.M., 2003, Simulation of Ground Water/Surface Water Flow in the Santa Clara-Calleguas Basin, Ventura County, California: U.S. Geological Survey Water-Resources Investigation 02–4136, 214 p.

Hanson, R.T., Newhouse, M.W., and Dettinger, M.D. 2004, A methodology to assess relations between climate variability and variations in hydrologic time series in the southwestern United States: Journal of Hydrology, v. 287, p. 253–270.

Hanson, R.T., and Dettinger, M.D., 2005, Ground-water/surface-water responses to ensembles of global climate simulations, Santa Clara-Calleguas basin, Ventura County, California, 1950-93: Journal of the American Water Resources Association, 41, 517–536.

Hanson, R.T., Dettinger, M.D., Newhouse, M.W., 2006, Relations between climatic variability and hydrologic time series from four alluvial basins across the southwestern United States: Hydrogeology Journal, v. 14, p. 1122–1146.

Hanson, R.T., Izbicki, J.A., Reichard, E.G., Edwards, B.E., Land, M.T., and Martin, P., 2009, Comparison of groundwater flow in southern California coastal aquifers, Chapter 5.3 *in* Earth science in the urban ocean: The Southern California Continental Borderland, Lee, H.J. and Normark, B., eds., GSA Special Volume 454, p. 345–373

Hanson, R. T., Flint, L. E., Flint, A. L., Dettinger, M. D., Faunt, C. C. Cayan, D., and Schmid, W., 2012, A method for physically based model analysis of conjunctive use in response to potential climate changes: Water Resources Research, v. 48, W00L08, doi:10.1029/2011WR010774.

Heilweil, V.M., and Brooks, L.E., eds., 2011, Conceptual model of the Great Basin carbonate and alluvial aquifer system: U.S. Geological Survey Scientific Investigations Report 2010–5193, 191 p.

Jiang, N., Neelin, J.D., and Ghil, M., 1995, Quasi-quadrennial and quasi-biennial variability in the equatorial Pacific: Climate Dynamics, v. 12, no. 2, p. 101–112.

Keppenne, C., and Ghil, M., 1992, Adaptive filtering and prediction of the Southern Oscillation Index: Journal of geophysical research, v. 97, no. D18, p. 20449–20454.

Kumar, M, and Duffy, C. J., 2009, Detecting hydroclimatic change using spatio-temporal analysis of time series in Colorado River Basin: Journal of Hydrology, v. 374, p. 1–15.

LaMarche, V.C., 1974, Frequency-dependent relationships between tree-ring series along an ecological gradient and some dendroclimatic implications: Tree-Ring Bulletin, v. 34, 1–20.

LaMarche, V. C., and Fritts, H.C., 1972, Tree-rings and sunspot numbers: Tree-Ring Bulletin, v. 32, p. 19–33.

Leake, S. A., Konieczki, A. D., and Rees, J. A. H., 2000, Ground-water resources for the future—Desert basins of the southwest: U.S. Geological Survey Fact Sheet, 086–00.

Lins, Harry F., Hirsch, Robert M., and Kiang, Julie, 2010, Water—the Nation's Fundamental Climate Issue: A White Paper on the U.S. Geological Survey Role and Capabilities: U.S. Geological Survey Circular 1347, 9 p., available at http://pubs.usgs.gov/circ/1347/.

North, G.R., Bell, T.L., Cahalan, R.F., and Moeng, F.J., 1982, Sampling erros in the estimation of empirical orthogonal functions: Monthly Weather Review, v. 110, no. 7, p. 699–706.

McKee, T.B., Doesken, N.J., and Kleist, J., 1993,The relationship of drought frequency and duration to time scales *in* Proceedings of the Eighth Conference on Applied Climatology: Boston, American Meteorological Society, p. 179–184.

Otnes, Robert K., and Enochson, Loren, 1978, Applied time series analysis: New York, John Wiley and Sons, 449 p.

Parzen, E., 1962, On estimation of a probability density function and mode: The annals of mathematical statistics, v. 33, no. 3, p. 1065–1076.

Pool, D.R., 2005, Variations in climate and ephemeral channel recharge in southeastern Arizona, United States: Water Resour. Res., v. 41, no. 11, p. W11403.

Press, W. H., Flannery, B. P., Teukolski, S. A., Vettering, W. T., 1988, Numerical Recipes: The art of scientific computing: Cambridge University Press, 818 p.

Redmond, K.T., and Koch, R.W., 1991, Surface climate and streamflow variability in the Western United States and their relationship to large-scale circulation indices: Water Resour. Res., v. 27, no. 9, p. 2381–2399.

Ropelewski, C.F., and Halpert, M.S., 1987, Global and regional scale precipitation patterns associated with the El Nino/Southern Oscillation, Mon. Wea. Rev., 115, p. 1606–1626.

Taylor, R.G., Scanlon, B., Doll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.F., Holman, I., and Treidel, H., 2012, Ground water and climate change: Nature Clim. Change, v. 3, no. 4, p. 322–329.

Treidel, H., Martin-Bordes, J.J., and Gurdak, J.J., eds., 2012, Climate change effects on groundwater resources: A global synthesis of findings and recommendations, International Association of Hydrogeologists (IAH) - International Contributions to Hydrogeology: Taylor & Francis publishing, 414 p., ISBN 978-0415689366, *http://www.crcpress.com/product/isbn/9780415689366*

Unal, Y. S., and Ghil, M., 1995, Interannual and interdecadal oscillation patterns in sea level: *Clim. Dyn., v. 11,* p. 255–278.

Vautard R., and Ghil, M., 1989 Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series: *Physica* v. *D35,* p. 395–424.

Vautard, R., Yiou, P., and Ghil, M., 1992, Singular-spectrum analysis: A toolkit for short, noisy chaotic signals: Physica D: Nonlinear Phenomena, v. 58, no. 1–4, p. 95–126.

Wallace, J.M., and Gutzler, D.S., 1981, Teleconnections in the geopotential height field during the northern hemisphere winter: Monthly Weather Review, v. 109, no. 4, p. 784–812.

Welch, P.D., 1967, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms: IEEE Transactions on Audio Electroacoustics, AU-15, p. 70–73.

Wolter, K. and Timlin, M. S., 2011, El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext): Int. J. Climatol., v. 31 p. 1074–1087.

Appendix

Appendix

Table A1. Definition of symbols used in the equations in the appendix.

Symbol	Definition
x	Observed time series
$\overline{x}$	Mean of x
x'	Difference between x and $\bar{x}$
$\overline{x'^2}$	Sample variance of x
t	Time of observation
N	Number of values in a time series
z	Standardized value in a time series
q	Probability of a zero value in a time series
w_1	Value of the first differenced time series
u_2	Value of the second differenced time series
y	Temporal signal regressed against harmonic functions
A_k	Regression coefficient
B_k	Regression coefficient
C_k	Regression coefficient
k	Wave number
T	Length of a period of record
Δt	Temporal spacing between records
Φ_{red}	Continuous theoretical red noise power spectrum
Φ_{sig}	Continuous significant power spectrum
ω	Angular frequency
T_e	e-folding time
r_1	Lag-1 autocorrelation
r_n	Lag-n autocorrelation
χ^2	Chi-squared statistic
v	Degrees of freedom
M^*	Number of spectral estimates
f_ω	Factor relating to smoothing of spectrum by a window
$w(t)$	Windowing function
$P_x(\omega)$	Power spectrum from MEM
a_k	Autoregressive coefficient at MEM pole k
M_p	Order (number of poles) for MEM
N_T	Number of columns in a trajectory matrix
M_T	Embedding dimension
$\boldsymbol{D}$	M_T by N_T trajectory matrix
$\boldsymbol{C}$	M_T by M_T covariance matrix
$\boldsymbol{E}$	M_T by M_T matrix of eigenvectors (T-EOFs)
λ	Vector of eigenvalues of length M_T or M_{eof}
$\boldsymbol{A}$	M_T by N_T matrix of principal components (T-PCs)
$R_\kappa(t)$	Reconstructed component for a set κ
κ	Set of components
u_t	Normalization factor
L_t	Lower bound of summation
U_t	Upper bound of summation
M_{eof}	Temporal dimension in an EOF analysis
N_{eof}	Spatial dimension in an EOF analysis
$\boldsymbol{X}$	M_{eof} by N_{eof} matrix of gridded time series observations
$\boldsymbol{C}_{EOF}$	M_{eof} by M_{eof} covariance matrix
$\boldsymbol{A}_{EOF}$	M_{eof} by M_{eof} matrix of eigenvectors (T-PCs)
λ_{EOF}	Vector of eigenvalues of length M_{eof}
$\boldsymbol{E}_{EOF}$	N_{eof} by M_{eof} matrix of the EOFs (S-EOFs)
$\boldsymbol{D}_{EOF}$	N_{eof} by M_{eof} matrix of scaled EOFs
n_s	Sample size in EOF analysis
$\boldsymbol{u}_i$	i^{th} vector in matrix $\boldsymbol{U}$
$\boldsymbol{U}$	M_{eof} by M_{eof} matrix of normalized time series of the PCs
N^*	Effective sample size
$\Delta\lambda$	North test sampling error
c	Predictand
d	Predictor
$\hat{c}$	Estimate of c
β_0	Fitted regression coefficient
β_1	Fitted regression coefficient
ε	Regression error term
$\bar{c}$	Mean of c
$\bar{d}$	Mean of d
c'	Difference between c and $\bar{c}$
d'	Difference between d and $\bar{d}$
r	Correlation coefficient
σ_c	Standard deviation of c
σ_d	Standard deviation of d
n^*	Effective sample size for linear regression
$r_{1,c}$	First-order autocorrelation coefficient for c
$r_{1,d}$	First-order autocorrelation coefficient for d
a_i	Autoregressive model coefficient
p	Order of the autoregressive model
y_{t-i}	Time series value at lag t–i
np	Number of autoregressive model parameters
V	Variance of the autoregressive model residuals

Standardization to Normal Distribution

The standardization tool transforms a set of normally distributed variables *xi* to a new variable z_i that is also normally distributed and has a sample mean $\overline{x}$ equal to zero and sample standard deviation *s* equal to one. The sample mean $\overline{x}$ of x_i is calculated as follows:

$$\overline{x} = \frac{1}{N}\sum\nolimits_{i=1}^{N} x_i \qquad (1)$$

where

N is the number of x values.

The sample standard deviation *s* is calculated as follows:

$$s = \sqrt{\overline{x'^2}} \qquad (2)$$

where

x' is the difference between x_i and $\overline{x}$; and

$\overline{x'^2}$ is sample variance, which is the mean of the squared values of x'.

The variable *x* can be normalized, or transformed, into a new variable *z*:

$$z_i = \frac{x_i - \overline{x}}{s} \qquad (3)$$

Standardized Precipitation Index

The Standardized Precipitation Index described by McKee and others (1993) and Edwards and McKee (1997) fits a gamma probability density function *g*(*x*) to a frequency distribution (data histogram). The cumulative probability *G*(*x*) of each value is transformed to a standard normal variable z with a mean of zero and variance of 1. *G*(*x*) is modified in order to account for zero values that are common in precipitation records because the gamma function is undefined for *x* = 0. To account for zero values, a modified cumulative probability *H*(*x*) is calculated:

$$H(x) = q + (1 - q)G(x) \qquad (4)$$

where

q is the probability of a zero value.

H(*x*) is transformed to the standard normal random variable *Z*, which is equal to the value of *SPI*:

$$z = SPI = -\left(t - \frac{c_0 - c_1 t + c_2 t^2}{1 + d_1 t + c_2 t^2 + d_3 t^3} \right) \qquad (5)$$

$$t = \sqrt{\ln\left(\frac{1}{\left(H(x)\right)^2} \right)} \qquad (6)$$

in which $0 < H(x) \leq 0.5$ and

$$z = SPI = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + c_2 t^2 + d_3 t^3} \right) \qquad (7)$$

$$t = \sqrt{\ln\left(\frac{1}{\left(1.0 - H(x)\right)^2} \right)} \qquad (8)$$

in which $0.5 < H(x) < 1.0$, as described by Edwards and McKee (1997) and Abramowitz and Stegan (1965). The coefficients are given as follows:

$$c_0 = 2.515517 \qquad (9)$$

$$c_1 = 0.802853 \qquad (10)$$

$$c_2 = 0.010328 \qquad (11)$$

$$d_1 = 1.432788 \qquad (12)$$

$$d_2 = 0.189269 \qquad (13)$$

$$d_3 = 0.001308 \qquad (14)$$

Cumulative Departure

Cumulative departure is calculated as the sum of the differences between consecutive values in a time series and the mean of the series:

$$\Sigma\ (x_i - \overline{x}) \qquad (15)$$

where

x_i is the value at time *i* and

$\overline{x}$ is the mean of the time series.

Interpolation

The interpolation tool uses linear interpolation, natural cubic spline, or Akima spline. These are calculated by using the mathematical routines available in the ALGLIB (*www.alglib.net*) library. Please see *www.alglib.net* for details.

Differencing

Non-stationarity in the mean of a time series (for example, a trend in the mean) can be removed by taking the first difference (Brockwell and Davis, 2002), which is calculated as follows:

$$w_1(t) = x(t) - x(t-1) \tag{16}$$

where

$w_1(t)$ is the value of the first difference at time t;
$x(t)$ is the value of the time series at time t; and
$x(t-1)$ is the value of the time series x at time $t - 1$.

Sometimes the trend in the mean is also changing. This can be removed by second-order differencing, which is the first difference of the first difference:

$$u_2(t) = w_1(t) - w_1(t-1) \tag{17}$$

where

$u_2(t)$ is the value of the second difference at time t, and
$w_1(t-1)$ is the value of the first difference at time $t - 1$.

A third-order differencing is then the first difference of the second difference.

Curve Fitting

The curve fitting tool fits polynomials using a barycentric form using the mathematical routines available in the ALGLIB (*www.alglib.net*) library. Please see *www.alglib.net* for details. The residuals for a time series $x(t)$ are calculated as the difference between the fitted polynomial and the time series at time t.

Discrete Fourier Transform, Windowing, and Spectral Smoothing

The power spectrum of a continuous series can be evaluated by using a discrete Fourier transform (DFT). The DFT uses a least-squares procedure to find the coefficients of the expansion:

$$y(t) = A_0 + \sum_{k=1}^{N/2}\left(A_k \cos\left(2\pi k \frac{t}{T}\right) + B_k \sin\left(2\pi k \frac{t}{T}\right)\right) \tag{18}$$

where

$y(t)$ is a continuous function of t;
T is the length of the period of record;
N is the number of grid points or time steps;
k is the wave number, which equals 1 to $N/2 - 1$; and
A_k and B_k are regression coefficients for each wave number k.

The solutions for the A_k and B_k coefficients are as follows:

$$A_k = \frac{2}{N}\sum_{i=1}^{N} y_i \cos\left(2\pi k \frac{i\Delta t}{T}\right) \tag{19}$$

$$B_k = \frac{2}{N}\sum_{i=1}^{N} y_i \cos\left(2\pi k \frac{i\Delta t}{T}\right) \tag{20}$$

where

Δt is the temporal spacing between points, and

$$A_{N/2} = \frac{1}{N}\sum_{i=1}^{N} y_i \cos\left(\pi N \frac{\Delta t}{T}\right) \tag{21}$$

$$\mathrm{B}_{N/2} = 0. \tag{22}$$

The variance explained by each wave number k is as follows:

$$\frac{C_k^2}{2} = \frac{A_k^2 + B_k^2}{2} \tag{23}$$

where

$\frac{C_k^2}{2}$ is the variance explained for wave number k, for $k = 1$ to $N/2$.

The red noise spectrum (null hypothesis) is constructed using the following relation:

$$\Phi(\omega)_{red} = \frac{2T}{1+\omega^2 T^2} \qquad (24)$$

where

Φ_{red} is the red noise spectrum,
ω is angular frequency, and
T_e is the e-folding time.

T_e is calculated as follows:

$$T_e = -\Delta t/\ln(r_1) \qquad (25)$$

where

r_1 is the lag-1 autocorrelation.

The lag-n autocorrelation r_n is calculated as follows:

$$r_n = \frac{\sum_{i=1}^{N-n}\left(x_i - \bar{x}\right)\left(x_{i+n} - \bar{x}\right)}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} \qquad (26)$$

Significance Testing

The statistical significance of the ratio of the power spectrum to the red noise spectrum can be evaluated by using either a chi-squared or F test. The null hypothesis is that the time series is not periodic and is simply red noise. The null hypothesis is rejected if the spectral peak is greater in amplitude than the critical value at a specified level of significance. The amplitude $\Phi(\omega)_{sig}$ of the spectral peak for a level of significance at frequency ω is compared to the amplitude $\Phi(\omega)_{red}$ of a peak that would be produced for red noise.

If the chi-squared test is used, the critical values of the spectrum at frequency ω are calculated as follows:

$$\Phi(\omega)_{sig} = \Phi(\omega)_{red} * \chi^2/v \qquad (27)$$

where

χ^2 is the chi-squared statistic with parameter v at a specified significance level, and
v is the number of degrees of freedom, calculated as follows:

$$v = \frac{N}{M^*} f_\omega \qquad (28)$$

where

N is the total sample size;
M^* is the number of spectral estimates;
f_ω is a factor related to smoothing by a windowing function, which is specified to be 1.2 if a Hamming window is used, or 1.0 if Boxcar, Hann, or Parzen windows are used.

If the F test is used, the critical values of the spectrum at frequency ω are calculated as follows:

$$\Phi(\omega)_{sig} = \Phi(\omega)_{red} * \mathrm{F} \qquad (29)$$

where

F is the F statistic with parameters v and infinity (degrees of freedom for red noise spectrum) at the specified significance level.

Windowing

Windowing modifies the original data series before performing spectral analysis. Windowing is necessary because the continuous Fourier transform presumes the time series extends from t = –∞ to ∞ and that the true spectrum can be calculated exactly by an analytical function. In reality, a time series is observed through a "window" in time (for example, from the beginning to the end of an observed time series), and the window of observation has a finite length (Otnes and Enochson, 1978). A main purpose of windowing is to reduce the discontinuities at the beginning and end of a finite time series (Otnes and Enochson, 1978). "Windowing" here means to apply a function to a time series of data, which can reduce the errors of applying the analytical Fourier transform. Windowing also can enhance some feature of the spectrum if a function is applied to a moving window that is chosen to be shorter than the time-series length.

Boxcar

The boxcar window, also known as a square or rectangular window, is calculated as follows:

$$w(t) = \begin{Bmatrix} 1/T & 0 \le |t| \le T \\ 0 & |t| > T \end{Bmatrix} \tag{30}$$

where

$w(t)$ is the transformed value in the time series at time t, and

T is the length of the period of record.

Hann

The Hann (or Hanning) window (Otnes and Enochson, 1978) tapers the ends of the time series by applying a cosine-shaped bell curve to the time series and is calculated as follows

$$w(t) = \begin{Bmatrix} \frac{1}{2}\left(1-\cos\frac{2\pi t}{T}\right) & 0 \le |t| \le T \\ 0 & |t| > T \end{Bmatrix} \tag{31}$$

Hamming

The Hamming window (Otnes and Enochson, 1978) is a slight modification of the Hann window:

$$w(t) = \begin{Bmatrix} \left(0.54+0.46\cos\frac{\pi t}{T}\right) & 0 \le |t| \le T \\ 0 & |t| > T \end{Bmatrix} \tag{32}$$

Parzen

The Parzen window is calculated as follows:

$$w(t) = \begin{Bmatrix} \left[1-\frac{|t|^m}{T}\right] & 0 \le |t| \le T \\ 0 & |t| > T \end{Bmatrix} \tag{33}$$

Maximum Entropy Method

The maximum entropy method (MEM; Burg, 1967; Parzen, 1968; Ghil and others, 2002) tool is based on the implemented in the SSA-Toolkit (Dettinger and others, 1995b). According to the Wiener-Khinchin theorem, a wide-sense-stationary random process has a power spectrum that is equal to the Fourier transform of its autocorrelation function. MEM fits an autoregressive process of order M having correlation coefficients φX that mimics a stationary time series X. The power spectrum is then identified as follows (Press and others, 1988):

$$P_x(\omega) = \frac{a_0}{\left|1+\sum_{k=1}^{M_p-1} a_k e^{ik\omega}\right|^2} \tag{34}$$

where

$P_x(\omega)$ is the power spectrum,

ω is the frequency,

a_0 is the variance of the time series,

a_k are autoregression coefficients at pole k, and

M_p is the order (number of poles) of the autoregressive process.

Singular Spectrum Analysis

The singular spectrum analysis (SSA) tool is based on the methods for extracting information from short and noisy time series described by Broomhead and King (1986), Vautard and Ghil (1998), and Ghil and others (2002). The software implements the approach proposed by Broomhead and King (1986), which utilizes a trajectory matrix $\boldsymbol{X}$ that is composed of a series of windows of the time series that are of length M. The dimensions of $\boldsymbol{X}$ are M_T by N_T,

where

N_T is equal to N–M+1,

N is the number of time steps in the time series, and

M_T is the embedding dimension of $\boldsymbol{X}$.

The second step is the construction of the covariance matrix $\boldsymbol{C}$:

$$\boldsymbol{C} = \frac{\boldsymbol{DD^T}}{N_T} \tag{35}$$

where

$\boldsymbol{C}$ is an M_T by M_T covariance matrix,

$\boldsymbol{D}$ is an M_T by N_T trajectory matrix, and

$\boldsymbol{D^T}$ is the transpose of $\boldsymbol{X}$.

The eigenvectors and eigenvalues of $\boldsymbol{C}$ are obtained by an eigenanalysis of $\boldsymbol{C}$. Broomhead and King (1986) obtained the eigenvectors and eigenvalues by performing singular value decomposition for $\boldsymbol{C}$ (SVD; see Golub and Van Loan, 1996), which provides equivalent results. Eigenanalysis is used instead of SVD because the computer memory requirements are lower, which permits the analysis of larger datasets. The eigenanalysis of $\boldsymbol{C}$ takes the following form:

$$\boldsymbol{CE} = \lambda\mathbf{E} \tag{36}$$

where

$\boldsymbol{E}$ is an M_T by M_T matrix of the eigenvectors, and

λ is the vector of eigenvalues of length M_T.

The eigenvectors are commonly referred to as the T-EOFs. A matrix of the principal components $\boldsymbol{A}$, also called the T-PCs, is obtained by projecting the eigenvectors $\boldsymbol{E}$ onto the trajectory matrix $\boldsymbol{D}$**,** as described by Ghil and others (2002) and Wilks (2011):

$$\mathbf{A} = \mathbf{E}^T\mathbf{D} \tag{37}$$

where

$\boldsymbol{E}$ is an M_T by M_T matrix of the eigenvectors, and
$\boldsymbol{A}$ is an M_T by N_T matrix of the principal components.

The reconstructed components (RCs) are formed by convolution of the principal components with the eigenvectors as described by Ghil and others (2002):

$$R_{\mathrm{K}}(t) = \frac{1}{M_t}\sum\nolimits_{\kappa\in\mathrm{K}}\sum\nolimits_{j=L_t}^{U_t} A_{\kappa}(t-j+1)E_{\kappa}(j) \tag{38}$$

where

K is the set of eigenvectors that are used in the reconstruction,
M_t is a normalization factor,
L_t is a bound of summation, and
U_t is a bound of summation.

The values of M_t , L_t , and U_t vary depending on interval within the times series:

$$(M_t, L_t, U_t) = \left\{ \begin{array}{ll} \left(\frac{1}{t}, 1, t\right), & 1 \le t \le M_T - 1 \\ \left(\frac{1}{M_T}, 1, M_t\right), & M_T \le t \le N_T \\ \left(\frac{1}{N-t+1},\ t - N + M_T,\ M_T\right), & N_T + 1 \le t \le N \end{array} \right\}$$

Empirical Orthogonal Function Analysis

The empirical orthogonal function (EOF) tool performs a principal component analysis for spatial and temporal patterns in gridded time series data as described by Wallace and Gutzler (1981), Dettinger and others (1998), and Wilks (2011). The EOF tool uses an eigenanalysis of a covariance matrix to obtain the EOFs (spatial patterns), PCs (temporal patterns), and the eigenvalues. An example of a hydrological dataset commonly used in EOF analysis is a series of maps of gridded precipitation values at monthly intervals.

Several processing steps are automatically completed prior to the eigenanalysis. First, the gridded time series are converted to a space-by-space-by-time matrix. An optional step is to weight each value in the gridded time series by the square root of the cosine of latitude to account for smaller area within the grids with increasing latitude. Next, the space-by-space-by-time matrix is condensed to a time by space matrix $\boldsymbol{X}$ of dimensions M_{eof} by $N_{eof,}$

where

M_{eof} is the number of time steps, and
N_{eof} is the spatial dimension that is equal to the total number of grid points in each map.

The second step is the construction of the covariance matrix $\boldsymbol{C}$:

$$\boldsymbol{C}_{EOF} = \frac{\boldsymbol{XX}^{\boldsymbol{T}}}{n_s} \tag{39}$$

where

$\boldsymbol{C}_{EOF}$ is an M_{eof} by M_{eof} matrix,
$\boldsymbol{X}$ is an M_{eof} by N_{eof} matrix,
$\boldsymbol{X}^{\boldsymbol{T}}$ is the transpose of $\boldsymbol{X}$, and
n_s is the sample size.

$\boldsymbol{C}$ is usually smaller if it is calculated by using $\boldsymbol{XX}^{\boldsymbol{T}}$, instead of $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}$, because M_{eof} is typically smaller than N_{eof}. The eigenvectors and eigenvalues of $\boldsymbol{C}$ are obtained by an eigenanalysis of $\boldsymbol{C}$, which takes the following form:

$$\boldsymbol{C}_{EOF}\boldsymbol{A}_{EOF} = \lambda_{EOF}\boldsymbol{A}_{EOF} \tag{40}$$

where

$\boldsymbol{A}_{EOF}$ is an M_{eof} by M_{eof} matrix of the eigenvectors, and
λ_{EOF} is the vector of eigenvalues of length M_{eof}.

Since $\boldsymbol{C}_{EOF}$ is obtained by $\boldsymbol{XX}^{\boldsymbol{T}}$, the eigenvectors are the PCs or S-PCs. A matrix of the EOFs $\boldsymbol{E}_{EOF}$, also called the S-EOFs, is obtained by projecting the eigenvectors $\boldsymbol{A}_{EOF}$ onto $\boldsymbol{X}$, as described by Wilks (2011):

$$\boldsymbol{E}_{EOF} = \boldsymbol{A}_{EOF}{}^{\boldsymbol{T}}\boldsymbol{X} \tag{41}$$

where

$\boldsymbol{E}_{EOF}$ is an N_{eof} by M_{eof} matrix of the EOFs.

The scaled EOFs $\boldsymbol{D}_{EOF}$ are calculated by regressing $\boldsymbol{X}$ (the original unweighted data in an M by N matrix) onto the normalized $\boldsymbol{U}^T$:

$$\tilde{u}_i = \frac{u_{ij} - \bar{u}_i}{\sigma_{ui}} \tag{42}$$

$$\boldsymbol{D}_{EOF} = \frac{\boldsymbol{A}\tilde{\boldsymbol{U}}^T}{NN_{eof}} \tag{43}$$

where

$\boldsymbol{D}_{EOF}$ is an N_{eof} by M_{eof} matrix of scaled EOFs,

$\boldsymbol{U}$ is a matrix M_{eof} by M_{eof} matrix of normalized time series of the PCs,

$\boldsymbol{U}^{\mathrm{T}}$ is the transpose of $\boldsymbol{U}$, and

$\boldsymbol{u}_i$ is the ith vector in the matrix $\boldsymbol{U}$.

The eigenvalue spectrum displays the percentage of the variance explained by each EOF. The North test (North and others, 1982) is used to evaluate the statistical significance at the 95 percent level. For the North test, the sampling error ($\Delta\lambda$) is calculated for each eigenvalue and is plotted as a length above and below the eigenvalue on the spectrum. A significant EOF is separated from the others within a sampling error of its eigenvalue. The sampling error is calculated as follows:

$$\Delta\lambda = \lambda_{EOF}\sqrt{\frac{2}{N^*}} \tag{44}$$

where

$\Delta\lambda$ is the sampling error obtained from the North test (North and others, 1982), and

N^* is the effective sample size.

In practice, the effective sample size is difficult to quantify. If the data were dominated by red noise, the eigenvalue spectrum would decrease slowly and exponentially. A null hypothesis of red noise can be rejected if the eigenvalue spectrum decreases faster than exponentially, in which the eigenvalues along the left side of the spectrum are typically much greater than the values along the flatter right side. This typically occurs if the eigenvalues are either above or below the error bars for the adjacent eigenvalue. If the data were mainly white noise, then the eigenvalue spectrum would plot along a flat line. A null hypothesis of white noise can also be rejected for the eigenvalues if the consecutive values do not plot along a flat line.

Linear Regression And Correlation

Linear regression analysis approximates the relation between a predictor x and a predictand y using the following expression:

$$\hat{c} = \beta_0 + \beta_1 d + \varepsilon \tag{45}$$

where

$\hat{c}$ is the estimate of predictand c,

d is a predictor,

β_0 and β_1 are fitted coefficients, and

ε is the normally-distributed error term.

The error is minimized by using the method of least squares. The solutions for the coefficients are as follows:

$$\beta_1 = \frac{d'c'}{d'^2} \tag{46}$$

$$\beta_0 = \bar{y} - \beta_1\bar{x} \tag{47}$$

where

$\bar{d}$ is the mean of d, and

$\bar{c}$ is the mean of c.

d' and c' are calculated as follows:

$$d' = d - \bar{d} \tag{48}$$

$$c' = c - \bar{c} \tag{49}$$

β_1 is also equal to the correlation coefficient r multiplied by the ratio of the standard deviations:

$$\beta_1 = r\frac{\sigma_c}{\sigma_d} \tag{50}$$

β_1 indicates how c and d change relative to each other, whereas the correlation coefficient r is a measure of the fit of the regression line and is determined as follows:

$$r = \frac{\overline{d'c'}}{\sigma_d\sigma_c} \tag{51}$$

The statistical significance of the sample correlation is evaluated by using the *t* statistic, which accounts for the sample size and the size of *r*. In order to evaluate the significance using the *t* statistic, several assumptions are made:

- The samples are from populations that are distributed normally.
- The samples are drawn completely randomly from the population.
- The correlation coefficient ρ of the population is zero.

The significance of the correlation coefficient can be evaluated by comparing the computed t statistic for the regression coefficient to user-specified upper and lower limits for the t distribution. The first step is to select a confidence level and either a one-tailed or two-tailed test. The one-tailed test can be used when either positive or negative correlations are of interest. The one-tailed test considers the following hypotheses:

H0: the correlation coefficient is zero.

H1: the correlation coefficient is significantly greater than zero (for a positive test) or less than zero (for a negative test).

The two-tailed test can be used when the correlation coefficient that is different than zero is of interest. The two-tailed test considers these hypotheses:

H0: the correlation coefficient is zero.

H1: the correlation coefficient is significantly different than zero.

The t statistic is calculated as follows:

$$t = \frac{r\sqrt{N'-2}}{\sqrt{1-r^2}} \tag{52}$$

where

n^* is the effective sample size, and
r is the correlation coefficient.

The effective sample size is calculated as follows:

$$Nn^* = N\frac{1-r_{1,x}r_{1,y}}{1+r_{1,x}r_{1,y}} \tag{53}$$

where

N is the number of samples,
$r_{1,c}$ is the first-order autocorrelation coefficient for *c*, and
$r_{1,d}$ is the first-order autocorrelation coefficient for *d*.

Autoregressive Model And Projections

For a stationary time series *x*, the autoregressive (AR) model tool represents a value x_t as a linear function of its previous values (Brockwell and Davis, 2002). The order *p* of the AR(*p*) model is the number of the most recent values that are included in the model.

The general form of an AR(*p*) model is as follows:

$$x_t = \sum\nolimits_{i=1}^{p} a_i y_{t-i} + \varepsilon_t \tag{54}$$

where

a_i is an autoregressive model coefficient,
y_{t-i} is a time series value at lag $t-i$;
p is the order of the AR model, and
ε_t is an error term.

The values of coefficients a_i are estimated by minimizing the residuals ε_t between the AR(*p*) predictions and observations of x_t by using the Yule Walker equations. The selection of the order of the model is not straightforward, and higher orders can give lower residuals at the expense of a more complex model. The Akaike's final prediction error (FPE), Akaike information criterion (AIC), and AIC with correction (AICc) can be used to evaluate the relative goodness of fit of the model while balancing model accuracy and complexity.

FPE is calculated as follows (Ljung, 1999):

$$FPE = \frac{1+np/N}{1-np/N}V \tag{55}$$

where

np is the number of parameters in the model,
N is the length of the time series, and
V is the variance of the model residuals.

AIC is calculated as follows (Ljung, 1999):

$$AIC = \log\left[V\left(1+\frac{2np}{N}\right)\right] \tag{56}$$

AIC$_C$ is calculated as follows:

$$AIC_C = AIC + \frac{2np\left(np+1\right)}{N-np-1} \tag{57}$$

A projection of a time series that has the same properties as the observations can be simulated by using a fitted AR model. The projection tool generates projections by manipulating an ensemble of randomly generated time series from a fitted AR model. The projections are randomly generated because a new error e_t is obtained for each time step *t* and for each projection in the ensemble. Hanson and others (2002) used this approach to develop projections of precipitation by AR models of oscillatory reconstructed components obtained by singular spectrum analysis.

References Cited

Abramowitz, M., and Stegun, I. A., 1964, Handbook of mathematical functions with formulas, graphs, and mathematical tables, National Bureau of Standards Applied Mathematics Series 55: U.S. Department of Commerce, 1,046 p.

Brockwell, P. J. and Davis, R. A., 2002, Introduction to time series and forecasting (2nd ed.): New York, Springer, 434 p.

Broomhead, D.S. and King, G.P., 1986, Extracting qualitative dynamics from experimental data: Physica D: nonlinear phenomena, 20, p. 217–236.

Burg, J. P, 1967, Maximum entropy spectral analysis, paper presented at the 37th Annual International Meeting of the Society of Exploration Geophysicists, Oklahoma City, Okla., 1967. (Published in Modern Spectrum Analysis, edited by D. G. Childers, pp. 42–48, IEEE Press, Piscataway, N. J., 1978.)

Dettinger, M.D., Ghil, M., Strong, C.M., Weibel, W., and Yiou, P., 1995b, Software Expedites Singular-Spectrum Analysis of Noisy Time Series: EOS, Transactions of the American Geophysical Union, v. 76, 2 p.

Dettinger, M.D., Cayan, D.R., Diaz, H.F., and Meko, D.M., 1998, North-South precipitation patterns in western North America on interannual-to-decadal timescales: Journal of Climate, v. 11, no. 12, p. 3095–3111.

Edwards, D.C., and McKee, T.B., 1997, Characteristics of 20th century drought in the United States at multiple time scales: Fort Collins, Colorado, MS Thesis, Colorado State University, 155 p.

Ghil, M., Allen, M.R., Dettinger, M.D., Ide, K., Kondrashov, D., Mann, M.E., Robertson, A.W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P., 2002, Advanced spectral methods for climatic time series: Reviews of Geophysics, v. 40, no. 1, p. 3-1–3-41.

Golub, G.H., and Van Loan, C.F., 1996, Matrix computations (3rd ed.): Baltimore, MD, The Johns Hopkins University Press, 694 p.

Ljung, L., 1999, System identification: theory for the user (2nd ed.): Englewood Cliffs, New Jersey, Prentice Hall, 672 p.

McKee, T.B., Doesken, N.J., and Kleist, J., 1993, The relationship of drought frequency and duration to time scales, in Proceedings of the Eighth Conference on Applied Climatology: Boston, American Meteorological Society, p. 179–184.

Otnes, Robert K., and Enochson, Loren, 1978, Applied time series analysis: New York, John Wiley and Sons, 449 p.

Parzen, E., 1962, On estimation of a probability density function and mode: The annals of mathematical statistics, v. 33, no. 3, p. 1065–1076.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T., 1988, Numerical Recipes in C: Cambridge, UK, Cambridge University Press, 735 p.

Vautard R., and Ghil, M., 1989 Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series: *Physica* v. *D35,* p. 395–424.

Wallace, J.M., and Gutzler, D.S., 1981, Teleconnections in the geopotential height field during the northern hemisphere winter: Monthly Weather Review, v. 109, no. 4, p. 784–812.

Wilks, D.S., 2011, Statistical methods in the atmospheric sciences (3rd ed.): Oxford, UK, Academic Press, 676 p.

Prepared by the Sacramento Publishing Service Center.

For more information concerning this report, contact:

Director
U.S. Geological Survey
California Water Science Center
6000 J Street, Placer Hall
Sacramento, CA 95819
dc_ca@usgs.gov

or visit our Web site at:
http://ca.water.usgs.gov

Dickinson and others—**HydroClimATe—Hydrologic and Climatic Analysis Toolkit**—Techniques and Methods 4–A9

ISSN (online) 2328-7055

Printed on recycled paper

http://dx.doi.org/10.3133/tm4a9