

## Overview

ProjectTS facilitates the structured management and analysis of multiple time-series datasets applicable to a project.

Typically, MS Excel has been used as a both a container and analysis tool for time series data. An Excel dataset, however, is fragile and vulnerable to errors because each cell is naturally independent of all other cells, so special actions and attention are needed to maintain and use a table composed of cells. While ranges provide a means of treating the cells as a unit, the range is inherently general, less convenient and not specific to time series.

ProjectTS encapsulates and hides the numerous, routine tasks in data management. In the Project TS environment, a series is treated as a single, ordered, integrated object composed of one or more measurements of the same kind over time. All data is stored in a single database in a common format. The user avoids the numerous, inherently error-prone tasks of managing rows and columns of data and can focus on analysis.

ProjectTS features include

- Imports data from files originated with typical sources, such as WDL, CDEC, USGS
- Data can be pasted from a spreadsheet
- Data can be exported to a spreadsheet or text file
- Any time series can be plotted and multiple series can be plotted together

Special functions

- Combine series into one
- Compute daily averages of series
- Add, subtract, multiply or divide by a constant
- Merge daily time series in order of priority
- Add, subtract, multiply or divide two daily time series where dates match
- Show basic statistics (number of points, mean, minimum, maximum, standard deviation, median)
- Compute a linear regression between two series (not implemented)

ProjectTS was funded by the (California) State Water Project Contractors Authority (SWPCA). Elaine Archibald was the project manager. Dennis Huff was the author.

## Recent Changes

See Cumulative Change Log at the end of this document for prior changes.

March 2015	<ul style="list-style-type: none"><li>• Enabled ability to paste multiple columns from a spreadsheet (tab-delimited) source.</li><li>• Enabled edits to time series values using a data grid on a special form.</li></ul>
April 2015	<ul style="list-style-type: none"><li>• Sums by month can now be computed and plotted for a series in the workspace in the same way that monthly averages are computed and plotted.</li><li>• To support consistent sizing of the chart area, the current chart width and height are shown on the status bar. Also, the form opens to the size it was when last closed.</li><li>• When two or more series are plotted, and the first series is selected to use the right axis, the left axis label is updated to take the label from the next series.</li><li>• Root nodes were added to the Data Explorer tree view, one for the open database, if any, and one for the open import file. These nodes remain in place while their respective files are open.</li><li>• Icons were added to the Data Explorer nodes.</li><li>• Numeric axis grouping and decimal places set automatically. User can adjust decimal lplaces.</li><li>• Time series axis grid line and tick mark intervals can be adjusted by the user.</li><li>• Added a title to the graphic workspace. The title can be edited.</li></ul>

## Data Container

ProjectTS uses a single SQL Server Compact Edition (SQL CE, with an sdf extension) file to store time series data, typically for a single project. A SQL CE file holds up to 80 Gb of data. It can be copied, deleted and so on like any other file.

New project files can be created as desired. Each new SQL CE file is created from a template that is installed in the same folder as the application executable.

## Data Structure

Two tables are used for the data.

## TimeSeries

The table TimeSeries contains the actual data observations, tagged by time (date + time), in chronological order, one value per time. Each observation includes the following data:

ObsTime	The date and time of the observation
RawValue	A text copy of the source. Pertinent only to imported data only.
Value	The value interpreted from RawValue.

## Series Metadata

The SeriesMetadata table contains information about the series, such as station, analyte, units and source.

GroupName	The group to which the series is assigned. Defaults to station for imported data. Used to group series at the top level in the data explorer.
NameTag	A name assigned to the series. Defaults to analyte for imported data. User as a name for the series in the data explorer.
Tags	Tags assigned to the series to aid in filtering and searching. Not currently used..
Station	The official, public name of the station, typically assigned by the data collecting agency.
Analyte	The full name of the analyte values reported in the series. "Analyte" is used here in a broad sense and could be applied to data values that are not quality constituents.
Analyte Description	A full description of the analyte as provided in the import source.e
Analyte Code	The code for the analyte used by the collecting agency.
Units	Units of the analyte observations
Source	An identifier for the source, typically an agency or agency sub unit, such as "USGS" or "WDL" (for DWR's Water Data Library).
SourceURL	A file path or web address.
Notes	Miscellaneous, user-defined.
Duration	A code noting the duration. If IsAverage (below) is true denotes the averaging period. "I" means instantaneous (IsAverage is false). "D" means daily (IsAverage is true).
IsAverage	Indicates whether the data are average values for some period.
nValues	Number of values in the series.
DateEntered	The date the series was created, typically by importing or calculating.

## Installation

Installation places the application in the user's program files folder along with a database template and a USGS parameter codes lookup SQL CE database.

### From DWR Intranet

Install the application from <\\mrsbmapp20932\db\projectts\setup.exe>

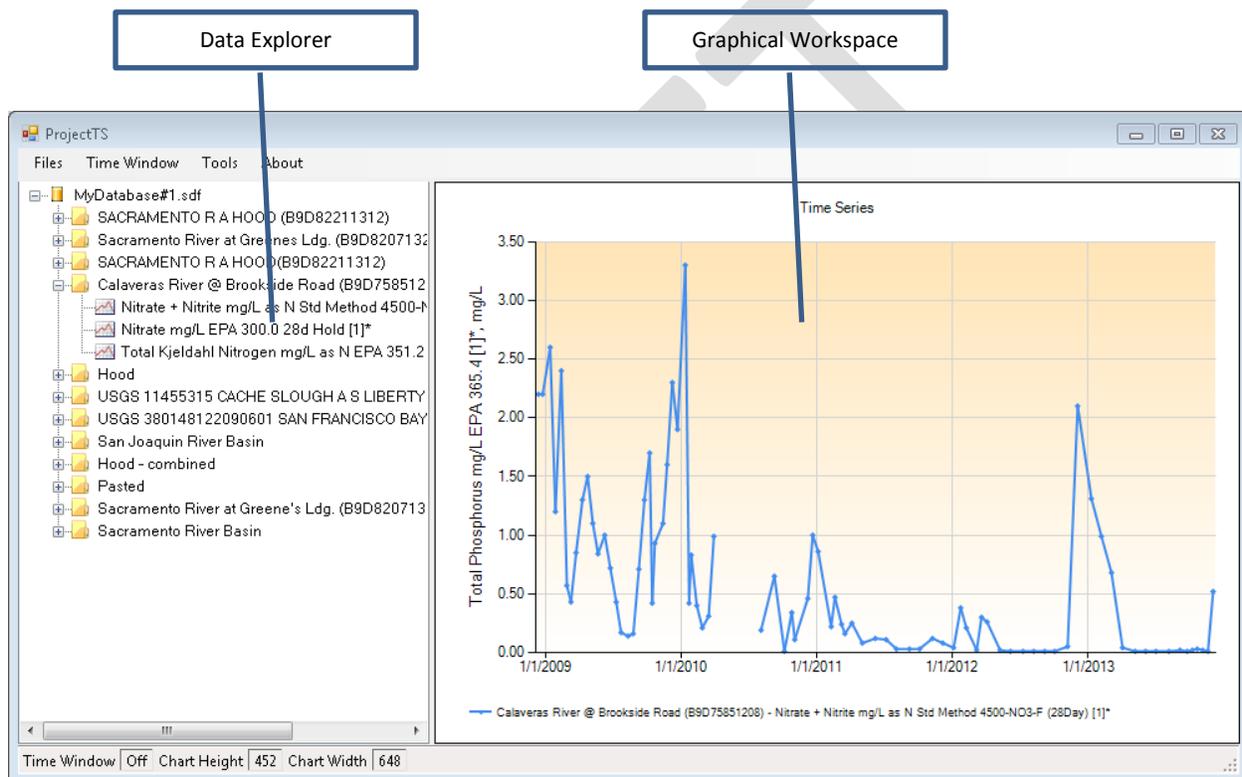
## From Files

Two files contain the application and related files. They must be installed from the same folder:

- ProjectTS.msi
- Setup.exe

Run setup.exe to launch the installation.

## Using the Application



## Main Form

The main form consists of

- A menu bar
- A data explorer pane (tree view) that lists the data in the selected source. The default source is the database. The contents of text and csv files and Excel workbook import files are shown instead when they are open for examination.
- A graphical workspace pane. The workspace displays data selected for analysis. Most operations are performed on data plotted in the workspace.
- A status bar

- Pop-up menus (activated by a mouse right click) are associated with the data explorer, graphical workspace, and individual series, either by clicking on the line or the legend.

The explorer and graphical workspace are separated by a moveable splitter. Simply click on the splitter and move it as desired when the appropriate icon is displayed.

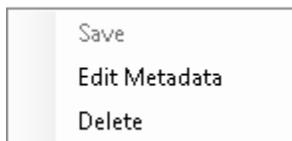
The form may be resized by dragging the right or bottom border or the right-bottom corner. The explorer and graphical workspace expand to fill the adjusted space.

The width and height of the graphical workspace is shown on the status bar.

## Data Explorer

The data explorer (tree view, hierarchic list) lists the data in the selected source database by group name and then by name tags for member series of the group. The default group name and name tag, used when importing data, are the station name and the analyte. It is typically more useful and effective to organize and manage other data by setting the group and name tag with values that are more concise and descriptive.

Series metadata is selected for viewing and editing and deletion by right clicking on the name tag.



The Save operation applies when the data is from a file or the internet and has not been saved in the database.

To plot or select a series for analysis, double (left) click on its name tag.

## Workspace

Time series data for the analytes selected using the tree view can be plotted in the workspace. A time window may be used to filter the data brought into the workspace from a database file.

Statistical analyses and most data computations or operations are performed on data plotted in the workspace. Order-dependent operations are documented under "Tools".

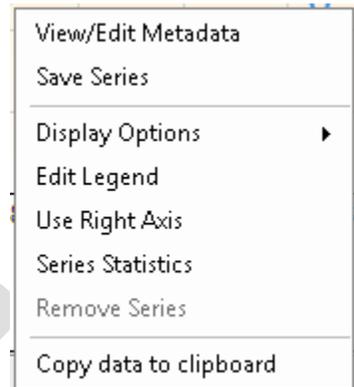
The y axis labels may be changed by right-clicking on it and entering a new label at the prompt.

The number of decimal places used for the y axis can be changed by right clicking on the axis.

Two popup menus are used in the workspace.

## Series Pop Up Menu

A series-related pop up appears by right clicking on a series or on a legend. It provides functionality applicable to a series, such as the display of series statistics described below.



- Display options:

The supported styles are line, points, line with points, stepped line and columns.

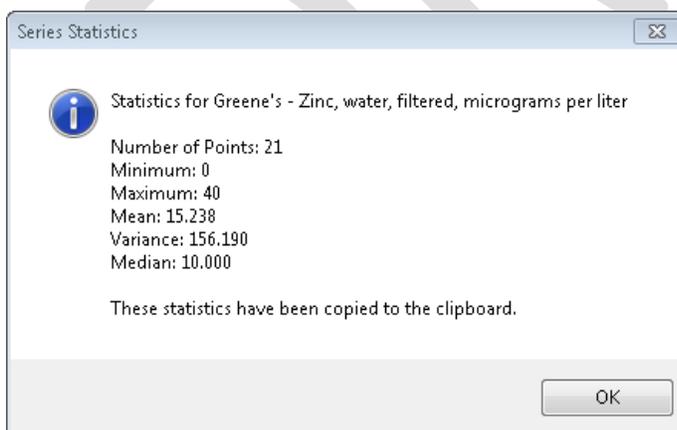
Line widths are expressed as pixels. A value of two (2) is a typical value.

Column widths are expressed as a fraction, in one-tenth increments, of the potential width of the column, roughly, the total length of the axis divided by the number of data values. The default is 0.6 (i.e., 60%). The maximum is 1.0.



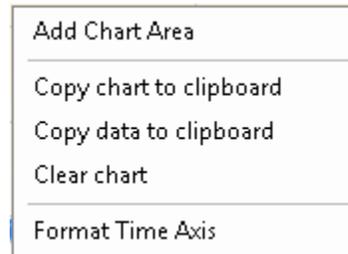
- Series Statistics

Series statistics are shown on a simple dialog form as shown below. These statistics are also copied to the clipboard for pasting into other applications.



## General Workspace Pop Up Menu

A general workspace pop up menu appears by right clicking anywhere else. It provides general functionality for the workspace.



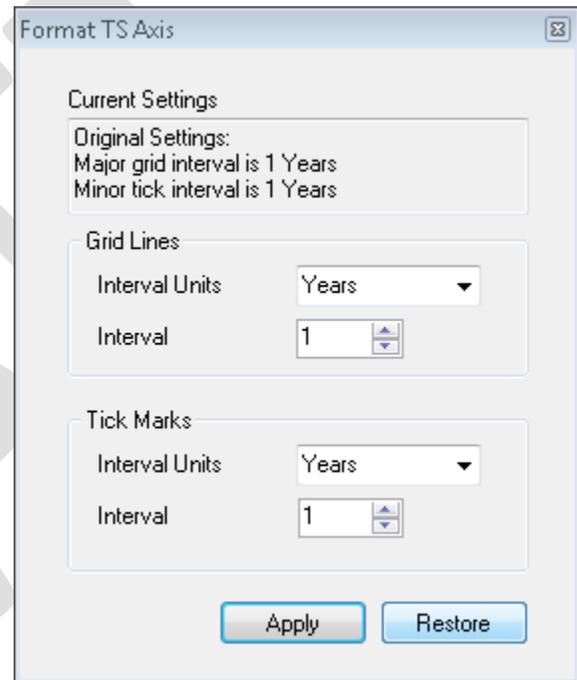
The built-in grid and tick interval behavior of the time series axis is often sufficient. However, grid and tick mark intervals can be adjusted using the Format Time Axis tool from the workspace pop up menu.

When the tool is selected, the current settings are displayed and saved.

New grid line and tick mark intervals can be specified and testing using the *Apply* button.

The *Restore* button changes the settings back to their original values when the form was opened.

In some cases, built-in behaviors still affect the time series grid with strange results. The chances of this is lessened by choosing new settings that are even multiples of the existing settings..

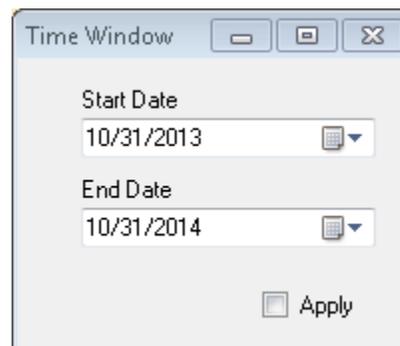


## Time Window

A time window may be used to filter the data loaded into the workspace. A time window is specified, activated or deactivated by selecting *Time Window* on the menu bar and making the selections on the form shown nearby.

Clicking *Apply* makes activates the specified time window and closes the form.

The state of the time window filter (on or off) is shown on



the status bar at the bottom of the main application form.

## Series Metadata

Series metadata may be viewed or edited from the tree view list or from the workspace legend.

The form below is used for viewing and editing series metadata. To save edits, use the “Apply” button. To exit without saving, simply close the form.

The screenshot shows a dialog box titled "View/Edit Series Metadata". It contains the following fields and values:

- ID: 39
- Date Entered: 6/15/2014
- Group: USGS 380148122090601 SAN FRANCISCO BAY A MARTINEZ
- Name Tag: Temperature, water, degrees Celsius
- Tags: (empty)
- Station: USGS 380148122090601 SAN FRANCISCO BAY A MARTINEZ
- Analyte: Temperature, water, degrees Celsius
- Analyte Description: Temperature, water, degrees Celsius
- Units: (empty)
- Notes: (empty)
- Duration: 1
- # Values: 9
- Is Average:
- Source: USGS
- Analyte Code: p00010
- Source URL: C:\Projects\ProjectTS\USGS\_Archive\_Martinez.txt
- Command: (empty)

An "Apply" button is located at the bottom right of the dialog box.

## Connecting To a Database File

### New Database File

Use Files → Create Database to create a new file.

A Windows open file dialog will appear with the title “New ProjectTS Database File”. Navigate to the desired folder, enter a name for the new file, and click on the Save button.

### Existing Database File

Use Files → Open Database on the menu bar to open (connect with) a database file. Database files have the extension “sdf”. A standard windows file open dialog is used to navigate to and select a file.

## File Data Sources

### General

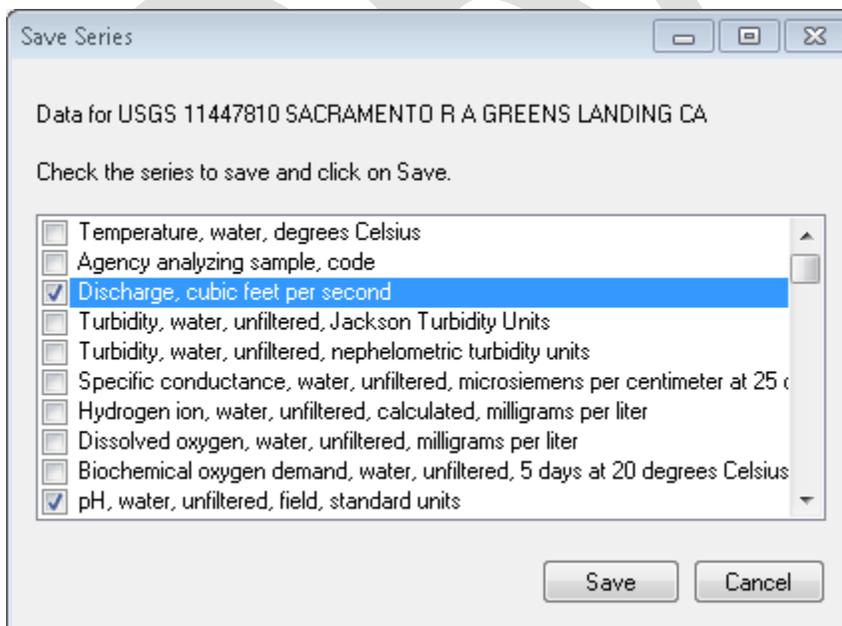
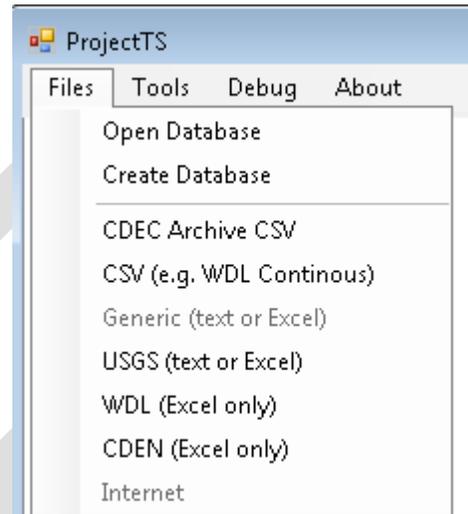
Files →<source> from the menu bar.

Where Excel workbooks with multiple worksheets are encountered, the user is asked to select the appropriate worksheet.

The file source data can be plotted and tabulated if desired without being saved. A database file need not be opened. This allows the application to be used to view and extract data from a file source independently.

The file’s data needs to be imported into (saved in) the database for use in operations, combinations and so on, however. To save the contents of the file, right click on the top level name (usually the file name) in the data explorer, then click Save on the pop-up menu.

The next step is to select the data to be saved. Often, or typically, the downloaded file contains data that is not needed for a particular analysis. A list with check boxes is presented as shown below. Check the data to be saved, then click on the Save button.



The import file path is recorded in the series metadata for reference.

DRAFT

## USGS WDIS

Can be either a text files or Excel workbook for a single station. Both the real-time and archive formats are recognized.

Cell values beginning with "<" are assigned the value given. For example, "< 10" is taken as a value of 10.

## WDL Discrete Samples

WDL data may be imported from an Excel workbook in the crosstab format for a single station or in a continuous table of data with a station and analyte specified for each row.

Data in a crosstab format (beginning with "<<<Laboratory Results>>>" in row 1, cell 1) may also be copied from the workbook and pasted into the workspace.

The analyte measurement method is described in the analyte name and description along with the units.

Units are based on the analyte name: mg/L is assumed. This can be changed by editing the metadata.

In the crosstab form:

- The string "N/A" in a column is taken to mean "no data" and is therefore ignored.
- Some values require special handling:

<i><b>Value Reported</b></i>	<i><b>Interpretation</b></i>
<0.1	0.1
0.3,0.4**	0.35
1.2,1.3**	1.25
<1.0,1.3**	1.15

## Generic CSV Data

The assumed format consists of at least two columns, date and value, separated by a comma. If the first column is not a date, the line is ignored.

## CDEC csv archive format

CDEC archive csv files begin like the following.

```
Title: "FPT.csv"  
599,PST,'FLOW, RIVER DISCHARGE (cfs) '  
20140601,2300,4413
```

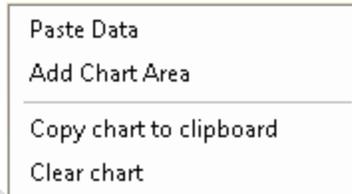
The station name is taken from the title. It would be “FPT” from the above example

The analyte name, analyte definition and units are taken from the second line. “Flow, River Discharge” would become the analyte and analyte definition and “cfs” would become the units.

## Pasting Data From A Spreadsheet

Data copied from a spreadsheet (tab-delimited) can be pasted into the graphic workspace and saved if desired.

First, select and copy the data from a worksheet. Then select “Paste Data” on the workspace popup menu (right click on the workspace). The data series is then imported into ProjectTS and plotted. Metadata for the series may be edited and the series saved through the series popup menu.



## Generic Case with One or More Data Columns

The data must consist of adjacent columns of data having dates and times in the first column and values in the other columns. The first row is assumed to contain labels. The analyte name is taken from the label for the values column. Cells with empty or textual data value cells are ignored.

The data must be in chronological order. Dates and times must be in a standard MS Windows date/time format explicitly specifying year, month and day and, if applicable, hours and minutes, like *3/1/14 16:52*.

The series are displayed on the workspace, but are not saved automatically.

## WDL Crosstab Data

The contents of a WDL crosstab worksheet may be pasted rather than imported as described above under “WDL Discrete Samples”. The data must be copied from Excel; csv files cannot be imported reliably.

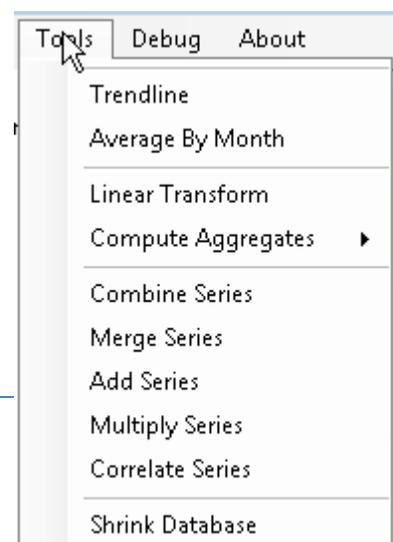
The data are pasted into the tree view catalog of data but not saved automatically. At this point, the series are treated as if they were imported from a file.

## Time Series Tools

Time series operations are performed on data in the workspace, i.e., the plotted data.

To perform an operation, use the Tools item on the menu bar and select the tool (operation).

Some operations result in a new time series. The new series will need to be saved if it is to be used again in subsequent operations or

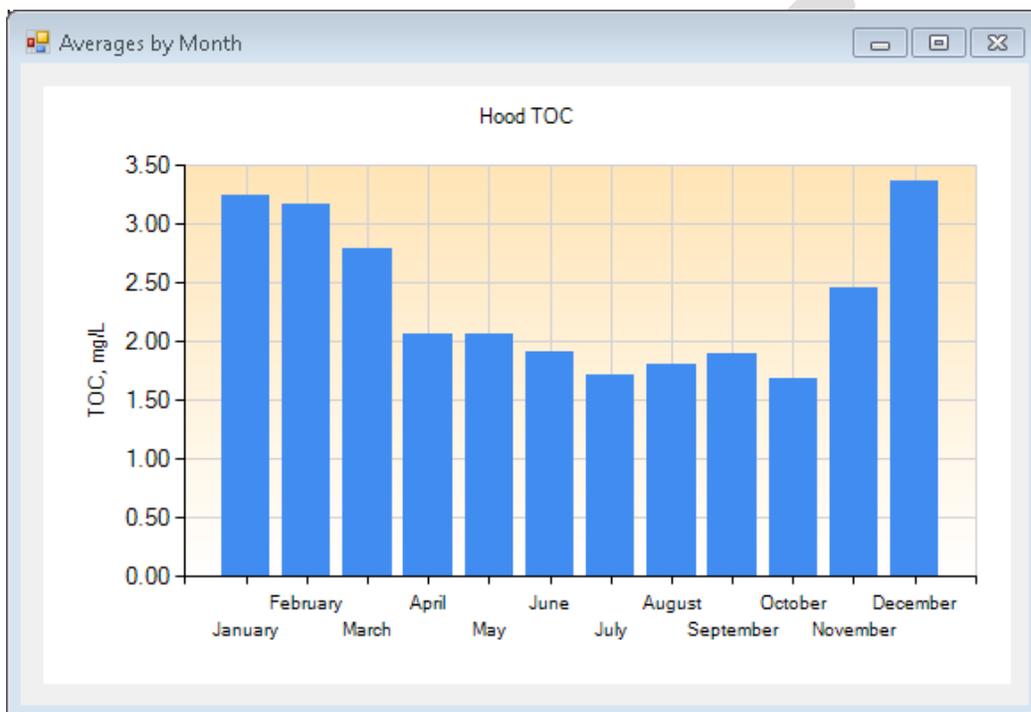


if is simply an end product that may be needed later. The metadata for the new series should be edited as appropriate.

Default values are assigned to computed series in order to avoid duplicate records, but aren't complete or necessarily accurate.

### Single Series

The following tools modify a single, selected series, creating a modified version. If no series is selected, the last series in the workspace is modified.



### *Averages by Month*

Compute averages for each month and display them in a graph.

### *Sums by Month*

Compute totals for each month and display them in a graph.

### *Linear Transform*

Create a new series as the linear transformation on the last selected series.

$$y = ax + b$$

Note that simple addition or subtraction of a constant requires that  $a=1$  (the default value).

### *Trendline*

Compute a trendline for the selected series and display a graph.



### Compute Aggregates

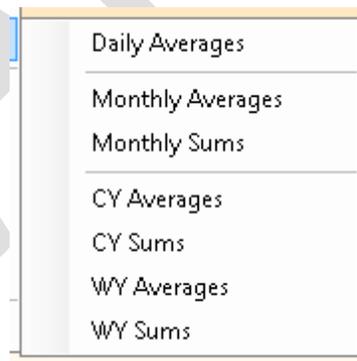
The items on the *Compute Aggregates* submenu are shown on the right.

#### **Daily Averages**

Compute the daily averages of a series from the source (input) data.

#### **Monthly Average, Sums**

Compute monthly aggregates (average or sum) from the input data. The date assigned is the first day of the month.



#### **CY, WY Averages, Sums**

Compute annual aggregates (average or sum) from the input data. The date assigned is the first day of the year. CY computes calendar year averages. WY computes water year averages.

### Multiple Series

#### **Combine**

All series in the workspace will be combined in a new series of average daily values.

Multiple values for a date will be averaged.

## ***Merge***

All series in the workspace will be merged to create a new series of average daily values.

Values from each series brought in to be merged will be added to the new series only if a value does not already exist from the previous merge.

The order of merging is first to last in the workspace.

Average daily values will be converted for each input series if they are not already average daily values.

## ***Add or Multiply***

A value from each series in the workspace for a given date and time is required for computing a result. In other words, each resulting value reflects values from each series in the workspace.

## ***Correlate Series***

A regression is computed from simultaneous values (i.e., matching data and time) in the last two series in the workspace. The last series is taken as the x series, the next-to-last series becomes the y series. If the y series is an instantaneous series, matching x values are interpolated in time from the x series.

The correlation is performed and the results displayed by the Correlation and Trendlines Form described below.

An example of a series correlation is shown on the following page. Average daily chloride for Banks is correlated with the average daily EC.

## **Correlation and Trendlines Form**

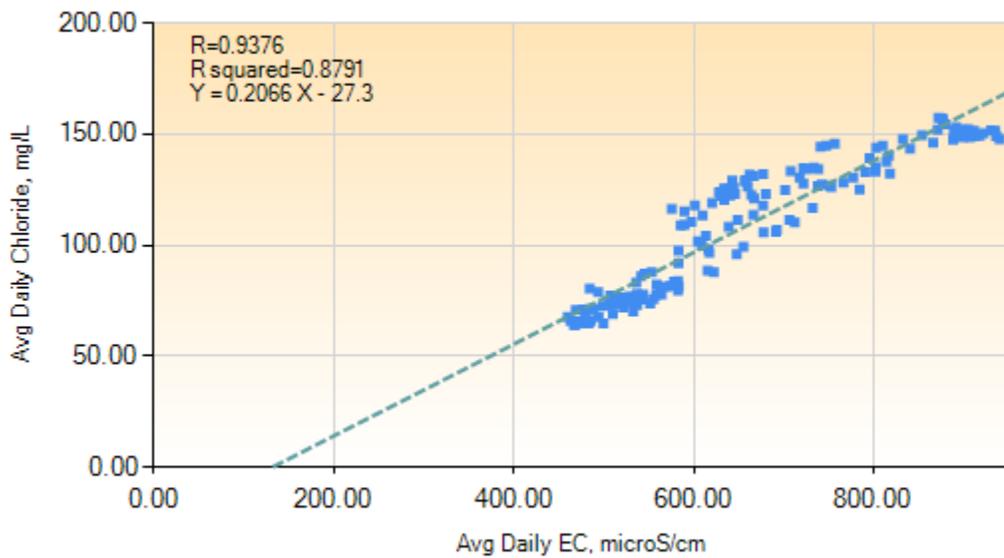
The statistics shown on the correlation and trend line charts may be moved or edited.

The chart title and axis legends of the correlation and trend line charts can be edited by right clicking on the existing text.

The correlation and trend line charts may be copied to the clipboard by right-clicking on the chart and selecting *Copy to the clipboard* from the pop-up menu.



Banks Daily Chloride vs EC 2014



## Cumulative Change Log

These changes were made prior to those listed under “Recent Changes” above.

August 4, 2014	<ul style="list-style-type: none"> <li>Revised Paste Data to ignore empty value cells.</li> <li>Enabled ability to select lines or points or both for a series.</li> </ul>
August 13, 2014	<ul style="list-style-type: none"> <li>The metadata variable AnalyteDefinition is copied from a pertinent series to the series resulting from an operation. Previously, the newly created variable was empty.</li> <li>Date format changed to full date and time when chart series are copied to the clipboard for export.</li> <li>Recognizes USGS samples below the non-detect limit and sets the value to the limit. This is consistent with how WDL data is handled.</li> <li>Enabled tool for computation of monthly averages. May need to record the number of observations involved for use in computing loads.</li> </ul>
August 15, 2014	<ul style="list-style-type: none"> <li>Added a pop-up menu option to clear imported data that is not saved.</li> </ul>
August 18, 2014	<ul style="list-style-type: none"> <li>Added a tool to compute regressions between two series.</li> </ul>
August 20, 2014	<ul style="list-style-type: none"> <li>Updated the tree view handler to drop empty top level nodes. These occur when the group name for metadata is changed for the one remaining series in a group.</li> </ul>
August 24, 2014	<ul style="list-style-type: none"> <li>Added a trendline analysis tool</li> <li>Modified the regression tool chart.</li> </ul>
September 7, 2014	<ul style="list-style-type: none"> <li>Added tools to compute annual average series by both calendar year and water year. Water year data is assigned the same date as calendar year (Jan 1), but spans the 12 months beginning with the previous October.</li> <li>Add a tool to compute summary averages by month.</li> <li>Added tools to adjust series display, including line width, chart type (line/points, stepped line and column).</li> </ul>
September 8, 2014	<ul style="list-style-type: none"> <li>Revised regression coefficients to use scientific notation in displaying very small or large numbers.</li> </ul>
September 11, 2014	<ul style="list-style-type: none"> <li>Added value of R to correlation display.</li> <li>Enabled creation of new database without the need for template. The necessary tables are added by the application.</li> <li>Added a tool to compute daily averages.</li> <li>Added internal logic to interpolate a matching x series for an instantaneous y series.</li> </ul>

September 15, 2014	<ul style="list-style-type: none"><li>• Enabled import of CDEN spreadsheet data</li></ul>
September 18, 2014	<ul style="list-style-type: none"><li>• Enabled editing rather than replacement of titles and axis labels</li></ul>
September 19, 2014	<ul style="list-style-type: none"><li>• Fixed problems with plots of regressions and trendlines.</li></ul>
September 24, 2014	<ul style="list-style-type: none"><li>• Enabled import of WDL “discrete sample data” from an Excel workbook. Choose <i>WDL (Excel Only)</i> on the imports menu.</li></ul>
October 29, 2013	<ul style="list-style-type: none"><li>• Enabled use of right axis for a series</li><li>• Enabled ability to set the width of data columns</li><li>• Enabled time window for filtering data added to the workspace</li><li>• Enabled automatic detection and display of data gaps</li><li>• Enabled computation of monthly and yearly sums of data points. Similar to the computation of monthly and yearly averages. Includes both calendar year and water year options.</li></ul>
February 12, 2015	<ul style="list-style-type: none"><li>• Enabled copy of a single selected series to the clipboard.</li><li>• Added information about the program’s origin and sponsor to the Overview.</li><li>• Documented the ability to paste from a WDL crosstab spreadsheet.</li><li>• Updated the description of WDL file imports.</li></ul>