

---

# Command Reference: FillMixedStation()

**Fill missing data in dependent time series using the best fit from 1+ independent time series, using OLS regression or MOVE2, data transforms, one/monthly equations**

Version 09.06.03, 2009-09-01

The `FillMixedStation()` command fills missing data in a time series where one or more independent time series is used to sequentially fill missing data. This approach has been developed to automate analysis of regression filling (see **Mixed Station Analysis Tool** below) and to facilitate batch filling of many related time series. Typically, the time series involved in the analysis are related, such as being from nearby locations in a region.

For each dependent time series being filled, the Mixed Station Analysis (MSA) selects the independent time series and parameters that result in the best filling results, considering combinations of the following:

- The list of independent time series being considered can be constrained to a subset of available time series.
- Filling methods include ordinary least squares (OLS) regression (see the `FillRegression()` command for details) and MOVE2 (see the `FillMOVE2()` command for details).
- One equation or monthly equations can be used. However, both options cannot be evaluated together due to the complexity of ranking and reporting results.
- The data can be transformed using  $\log_{10}$ , or no transformation can be applied.
- A minimum number of overlapping data points ( $N1$ ) can be specified to indicate a valid relationship.
- A minimum correlation coefficient  $r$  can be specified to indicate a valid relationship.
- The best fit indicator can be the correlation coefficient ( $R$ ) or the standard error of prediction (SEP, described below).

The interactive MSA tool is available to facilitate configuration of the `FillMixedStation()` command. The primary uses of the MSA tool and `FillMixedStation()` command are:

- Use the MSA tool to evaluate filling options before finalizing the fill commands. For example, run the MSA tool and review the output report to confirm the filling methodology and corresponding command parameters. A `FillMixedStation()` command can then be created and passed to the TSTool command list.
- If no evaluation is needed, use the or `FillMixedStation()` command editor to create a single `FillMixedStation()` analysis command, which when run evaluates the best filling parameters for filling and performs the filling. A single command can be used to fill many time series.

Because extensive analysis may be necessary to evaluate all the combinations of parameters, the `FillMixedStation()` command will be slower than other commands that specifically indicate how to perform the filling. Performance can be increased by using the Mixed Station Analysis tool to determine time series that result in the best fit, and excluding all other time series in the fill command. The number of combinations can also be limited by reducing the number of parameter options and using stricter limitations on the number of points and correlation coefficient that are required for a good regression result.

The full MSA process is as follows:

1. For each dependent time series, perform a regression analysis using a unique combination of parameters (e.g., use an independent time series, OLS regression with one equation, no data transform). This results in 1+ regression results for each dependent time series.
2. Qualifying results (those that meet the requirements of minimum number of overlapping points and correlation coefficient) are retained in a list for the dependent time series, for processing in the next step.
3. The qualifying results are ranked according to the best fit indicator (e.g.,  $R$  or standard error of prediction SEP, as described below). If a monthly analysis is performed, the results for each month are ranked.
4. Missing data in the dependent time series are filled using the regression results. If missing values remain, the next highest ranking regression result is used until all missing values are filled (or no additional qualifying regression results are available). Monthly filling occurs on each of the 12 months.

## Best Fit Indicators

Best fit indicators that are available include:

1. Correlation coefficient,  $R$ , defined as:

$$R = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2] \cdot [N \sum Y^2 - (\sum Y)^2]}}$$

where  $X$  is the independent and  $Y$  is the dependent for all overlapping data points.

2. Standard Error of Prediction (SEP), defined as:

$$SEP = \sqrt{\frac{\sum (Y - Y')^2}{N}}$$

where  $Y$  is the observed value and  $Y'$  and is the estimated value using the relationship determined for filling, for all original points in the independent time series.

In the future, additional indicators may be added, such as the Nash-Sutcliffe efficiency coefficient.

## Mixed Station Analysis Tool

The Mixed Station Analysis tool is started after time series results are generated in TSTool, for example as the result of reading a list of time series. The following dialog illustrates the parameters of the tool, which are essentially the same as for the `FillMixedStation()` command that is described in the next section.

**Mixed Station Analysis**

This tool finds the best fit to fill the dependent time series with data from the independent time series, and generates commands to perform the filling.  
 The dependent and independent time series can be selected using the TS list parameters:  
 The working directory is: C:\Develop\TSTool\_SourceBuild\TSTool\test\regression\commands\general\FillMixedStation

**Edit Mixed Station Analysis Parameters**

Dependent TS list:  Optional - indicates the time series to process (default=AllTS).

Dependent TSID (for TSList=AllMatchingTSID):

Independent TS list:  Optional - time series used to fill (default=AllTS).

Independent TSID (for TSList=AllMatchingTSID):

Best Fit:  Required - best fit indicator, for ranking output.

Analysis method(s): ☐ MOVE2 ☐ OLSRegression Optional - methods to use in analysis (default=OLSRegression).

Number of equations:  OneEquation Optional - number of equations to use in the analysis (default=OneEquation).

Transformation(s): ☐ Log ☐ None Optional - transformations to use in analysis (default=None).

Intercept:  Optional - 0.0 is allowed with Transformation=None (default=no fixed intercept).

Analysis period:  to

Fill period:  to

Minimum data count:  Optional - minimum number of overlapping points required for analysis (default=10).

Minimum R:  Optional - minimum correlation coefficient R required for a best fit (default = 0.5).

Output file:

**Perform Mixed Station Analysis and Review Results**

**Copy Command(s) to TSTool**

Fill command(s): 

```
FillMixedStation(BestFitIndicator=SEP,NumberOfEquations=OneEquation)
```

Ready

tool\_MixedStationAnalysis

**Mixed Station Analysis Tool after Initial Display**

The MSA tool is used as follows:

1. The tool will initially display default parameter values, but these values can be changed using the interface. Select parameters as appropriate for the analysis. The corresponding `FillMixedStation()` command is shown at the bottom of the window.
2. Press the **Analyze** button to perform the analysis. The analysis may run for several minutes.
3. When the analysis is complete, press the **View Output File** button to view the analysis results.
4. If satisfied with the results, meaning that reasonable relationships have been determined, go to the following step. Otherwise adjust the analysis parameters and run the analysis again.
5. Use the **Copy Command to TSTool** button to copy the `FillMixedStation()` command to TSTool. The command will be inserted as if it were edited from the **Commands** menu.

6. The MSA tool can then be closed and commands run as usual to automate data processing.

Subsequent edits of the command can occur using the normal command editor. The following section provides an example of the standard command editor and a description of all of the parameters.

## Command Editing

The following dialog is used to edit the `FillMixedStation()` command and illustrates the syntax of the command:

**Edit FillMixedStation() Command**

This command finds the best fit to fill the dependent time series with data from the dependent time series, and performs the filling.  
 The dependent and independent time series can be selected using the TS list parameters:  
 The working directory is: C:\Develop\TSTool\_SourceBuild\TSTool\test\regression\commands\general\FillMixedStation

Dependent TS list:  Optional - indicates the time series to process (default=AllTS).

Dependent TSID (for TSList=AllMatchingTSID):

Independent TS list:  Optional - time series used to fill (default=AllTS).

Independent TSID (for TSList=AllMatchingTSID):

Best Fit:  Required - best fit indicator, for ranking output.

Analysis method(s): ☒ MOVE2 ☒ OLSRegression Optional - methods to use in analysis (default=OLSRegression).

Number of equations:  MonthlyEquations Optional - number of equations to use in the analysis (default=OneEquation).

Transformation(s): ☒ Log ☒ None Optional - transformations to use in analysis (default=None).

Intercept:  Optional - 0.0 is allowed with Transformation=None (default=no fixed intercept).

Analysis period:  to

Fill period:  to

Minimum data count:  Optional - minimum number of overlapping points required for analysis (default=10).

Minimum R:  .7 Optional - minimum correlation coefficient R required for a best fit (default = 0.5).

Output file:  Results\Test\_FillMixedStation\_gunnv.xbg\_MonthlyEquations\_out.txt

Fill command(s):

```
FillMixedStation(BestFitIndicator=SEP,AnalysisMethod="MOVE2,OLSRegression",NumberOfEquations=MonthlyEquations,Transformation="Log,None",MinimumR=.7,OutputFile="Results/Test_FillMixedStation_gunnv.xbg_MonthlyEquations_out.txt")
```

FillMixedStation

## FillMixedStation() Command Editor

The command syntax is as follows:

```
FillMixedStation(Parameter=value,...)
```

### Command Parameters

Parameter	Description	Default
DependentTSList	Indicates the list of independent time series to be processed, one of: <ul style="list-style-type: none"> <li>• AllMatchingTSID – all time series that match the TSID (single TSID or TSID with wildcards) will be processed.</li> <li>• AllTS – all time series before the command will be processed.</li> <li>• EnsembleID – all time series in the ensemble will be processed.</li> <li>• FirstMatchingTSID – the first time series that matches the TSID (single TSID or TSID with wildcards) will be processed.</li> <li>• LastMatchingTSID – the last time series that matches the TSID (single TSID or TSID with wildcards) will be processed.</li> <li>▪ SelectedTS – the time series selected with the SelectTimeSeries() command will be processed.</li> </ul>	None – must be specified.
DependentTSID	The time series identifier or alias for the dependent time series to be processed, using the * wildcard character to match multiple time series.	Required if DependentTSList=*TSID.
IndependentTSList	Indicates the list of independent time series to be considered for each dependent time series, one of: <ul style="list-style-type: none"> <li>• AllMatchingTSID – all time series that match the TSID (single TSID or TSID with wildcards) will be processed.</li> <li>• AllTS – all time series before the command will be processed.</li> <li>• EnsembleID – all time series in the ensemble will be processed.</li> <li>• FirstMatchingTSID – the first time series that matches the TSID (single TSID or TSID with wildcards) will be processed.</li> <li>• LastMatchingTSID – the last time series that matches the TSID (single TSID or TSID with wildcards) will be processed.</li> <li>• SelectedTS – the time series selected</li> </ul>	None – must be specified.

Parameter	Description	Default
	with the <code>SelectTimeSeries()</code> command will be processed.	
<code>IndependentTSID</code>	The time series identifier or alias for the independent time series to be compared, using the * wildcard character to match multiple time series.	Required if <code>IndependentTSList=</code> *TSID.
<code>BestFitIndicator</code>	Specifies the indicator to use when determining the best fit, one of: <ul style="list-style-type: none"> <li>▪ <code>R</code> (correlation coefficient).</li> <li>▪ <code>SEP</code> (Standard Error of Prediction), defined as the square root of the sum of differences between the known dependent value, and the value determined from the equation of best fit at the same point.</li> <li>▪ <code>SEPTotal</code>, when used with one equation, it is the same as <code>SEP</code>. When used with monthly equations, it is the average error considering all months.</li> </ul>	<code>SEP</code> .
<code>AnalysisMethod</code>	Specify the method(s) to analyze the data, in order to determine the best fit, including <code>OLSRegression</code> and/or <code>MOVE2</code> . If multiple methods are specified, separate with commas and surround with double quotes.	<code>OLSRegression</code>
<code>NumberOfEquations</code>	The number of equations to use for the analysis: <code>OneEquation</code> or <code>MonthlyEquations</code> . Only one may be chosen. If necessary, use more than one command to use different parameter combinations for different groups of time series.	None – must be specified.
<code>Transformation</code>	Indicates how to transform the data before analyzing. Specify as <code>None</code> (no transformation) or <code>Log</code> (for $\log_{10}$ ). If the <code>Log</code> option is used, zero and negative values in data are set to .001. Missing data are ignored. If multiple values are selected, separate with a comma and surround with double quotes.	None (no transformation)
<code>Intercept</code>	Specify as 0 to force the intercept of the best-fit line through the origin. This is made available only for <code>OLS</code> regression analysis on untransformed data, to be consistent with the <code>FillRegression()</code> command.	Do not force the intercept through zero.
<code>AnalysisStart</code>	The date/time to start the analysis, to focus on a period appropriate for analysis. For example, specify the unregulated period for streamflow.	If blank, analyze the full period.
<code>AnalysisEnd</code>	The date/time to end the analysis.	If blank, analyze the full period.

Parameter	Description	Default
FillStart	The date/time to start filling, if other than the full time series period.	If blank, fill the full period.
FillEnd	The date/time to end filling, if other than the full time series period.	If blank, fill the full period.
MinimumDataCount	The minimum number of overlapping data points that are required for a valid analysis (N1 in FillRegression() and FillMOVE2() documentation). If the minimum count is not met, then the independent time series is ignored for the specific combination of parameters. For example, if monthly equations are used, the independent time series may be ignored for the specific month; however, it may still be analyzed for other months.	10
MinimumR	The minimum correlation coefficient required for a best fit. If the minimum is not met, then the results are not considered in the best fit ranking or filling.	0.5
OutputFile	Output file for the results, either as a file name to be written to the working directory, or a full path.	If not specified, partial results of the analysis may be available in the log file.

The following example command file fills natural flow time series from a StaeMod file using one equation (not monthly):

```
# Test filling the gunnison monthly baseflow time series with
# Mixed Station Analysis (all combinations for one equation)
StartLog(LogFile="fill-baseflow.log")
ReadStateMod(InputFile="gunnv.xbg")
FillMixedStation(BestFitIndicator=SEP,AnalysisMethod="MOVE2,OLSRegression",
NumberOfEquations=OneEquation,
Transformation="Log,None",OutputFile="Results.txt")
# Check for missing data - all should be filled
CheckTimeSeries(CheckCriteria="Missing",MaxWarnings=10)
# Check for negative flows - should not be any
CheckTimeSeries(CheckCriteria="<",Value1=0,MaxWarnings=10)
```

This page is intentionally blank.