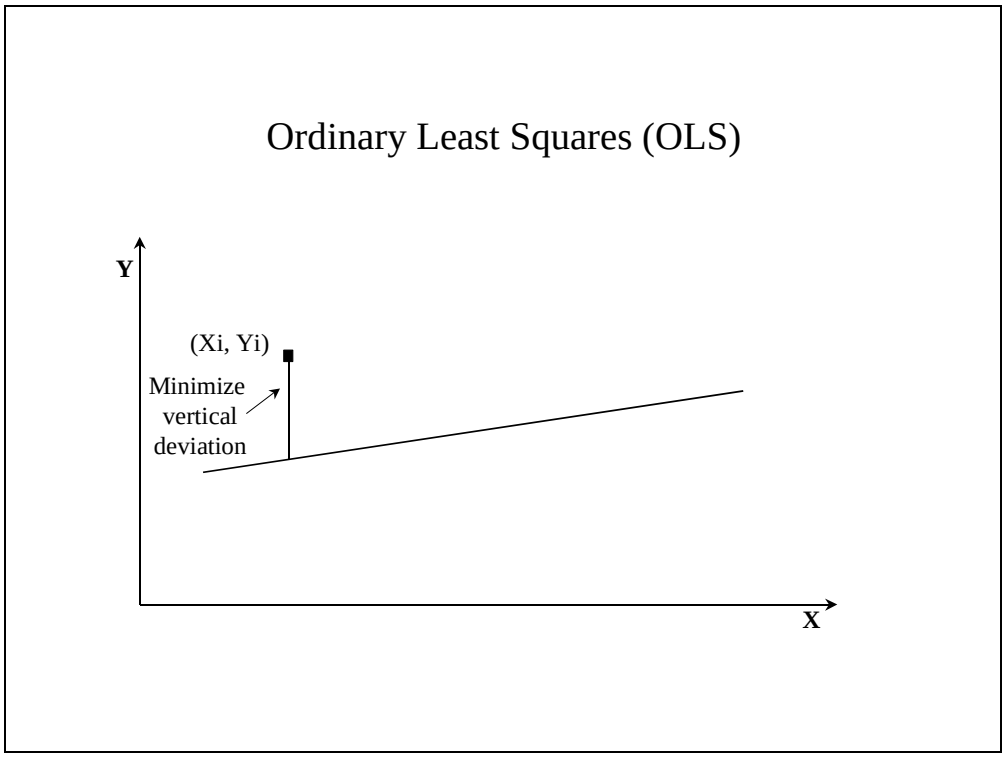
Command Reference: fillRegression()

Fill Missing Time Series Data Using Ordinary Least Squares Regression

Version 06.10.05, 2005-08-05, Color, Acrobat Distiller

The `fillRegression()` command fills missing data in a time series using ordinary least squares (OLS) regression (see also the `fillMOVE2()` command). Prior to TSTool version 05.00.00, regression was performed using `regress*()` commands, of which there were several variations. These commands have been replaced with a single `fillRegression()` command, which uses parameters to indicate the type of regression. Regression can be applied only to regular interval time series. The first time series selected (dependent time series) will be filled using the other selected time series (independent time series). The periods of record and output period for the time series should be verified to make sure that the time series periods overlap sufficiently. The ***Results…Graph - XY-Scatter*** is a useful tool for reviewing data. Regression relationships are developed using the analysis period for the time series and are applied to the fill period. Refer to the log file and time series properties for analysis details.

In OLS regression, the vertical distance from the data point to the regression line is minimized. OLS regression provides the minimum-variance estimate for a single value or observation. However, if an ensemble of points is estimated from OLS regression, the estimated values will have lesser variability than the true values.



The following OLS equation is used to estimate values for the dependent time series from the independent time series:

$$Y_i = \overline{Y}_1 + r \frac{S_{y1}}{S_{x1}} \left[X_i - \overline{X}_1 \right]$$

or

$$Y_i = a + bX_i$$

where

N_1 = concurrent or overlapping period of record

$\overline{X}_1$ = mean for independent variable for N_1 years

$\overline{Y}_1$ = mean for dependent variable for N_1 years

S_{y1} = standard deviation for N_1 years

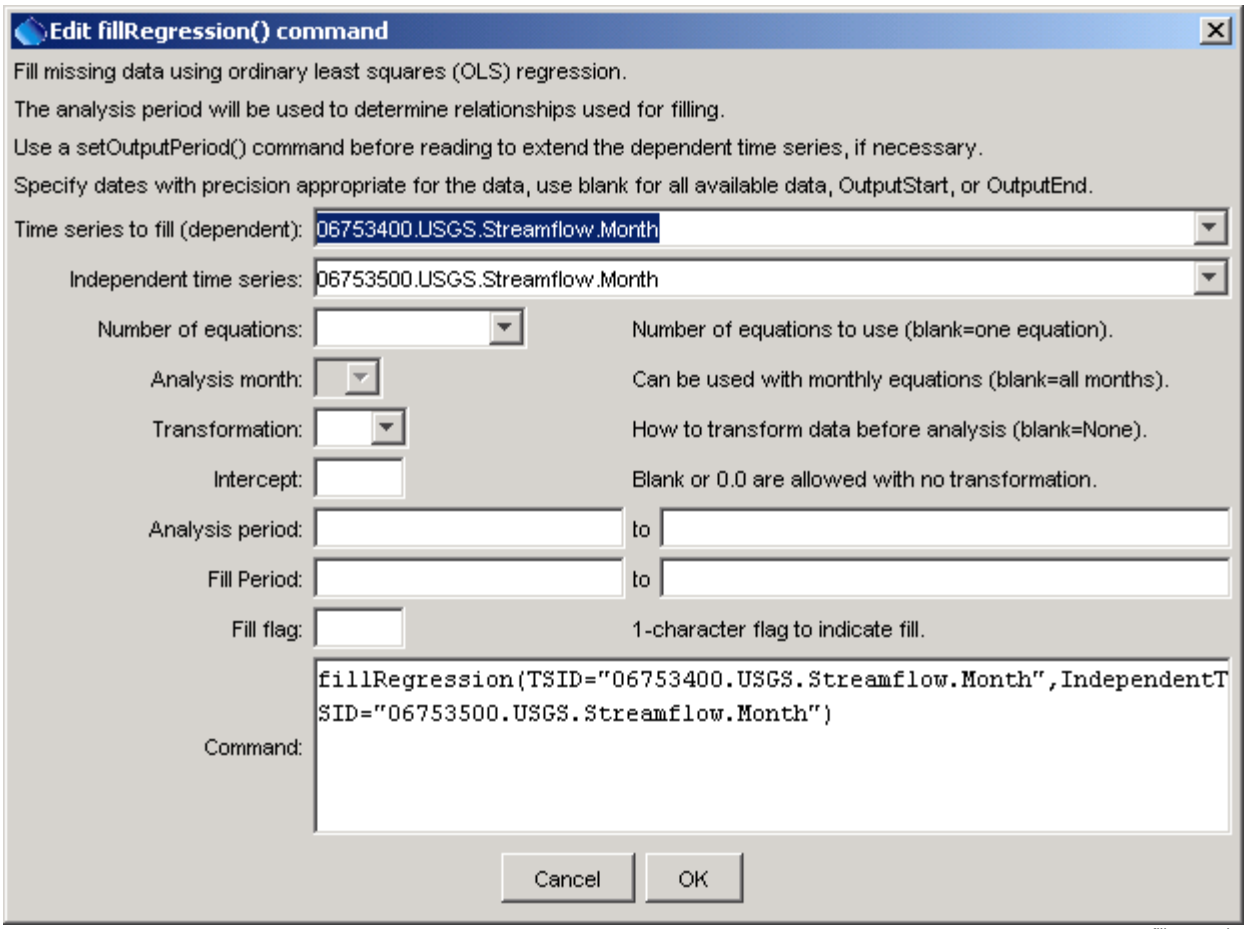
S_{x1} = standard deviation for N_1 years

$b = r \frac{S_{y1}}{S_{x1}}$, r = correlation coefficient

$a = \overline{Y}_1 - b\overline{X}_1$

Note that the correlation coefficient, r, is used to compute the slope, *b*, of the line.

The following dialog is used to edit the command and illustrates the syntax of the command:



fillRegression

fillRegression() Command Editor

The command syntax is as follows:

```
fillRegression(param=value,…)
```

Command Parameters

Parameter	Description	Default
TSID	The time series identifier or alias for the time series to be filled.	None – must be specified.
Independent TSID	The time series identifier or alias for the independent time series. Right-clicking on a time series allows it to be converted to a temporary time series (TEMPTS).	None – must be specified.
NumberOf Equations	The number of equations to use for the analysis: OneEquation or MonthlyEquations.	OneEquation
AnalysisMonth	Indicate the month to process when using monthly equations. Currently only a single month can be specified.	Process all months.
Transformation	Indicates how to transform the data before analyzing. Specify as None (previously Linear) or Log (for Log_{10}). If the Log option is used, zero and negative values are set to .001 (missing data values are ignored in the analysis).	None (no transformation).
Intercept	Specify as 0 to force the intercept of the best-fit line through the origin (not available for log transformation).	Parameter is optional and if specified the default is to not force the intercept through zero.
AnalysisStart	The date/time to start the analysis – use to focus on only a period appropriate from analysis. For example specify the unregulated period for streamflow.	Analyze the full period.
AnalysisEnd	The date/time to end the analysis – use to focus on only a period appropriate from analysis.	Analyze the full period.
FillStart	The date/time to start filling, if other than the full time series period.	Fill the full period.
FillEnd	The date/time to end filling, if other than the full time series period.	Fill the full period.
FillFlag	A single character that will be used to flag filled data.	Filled values will not be flagged.

A sample commands file is as follows (the fillRegression() command would be on one line in TSTool):

```
# 06753400 - LONETREE CREEK AT CARR, CO.
06753400.USGS.Streamflow.Month~HydroBase
# 06753500 - LONETREE CREEK NEAR NUNN, CO.
06753500.USGS.Streamflow.Month~HydroBase
fillRegression(TSID="06753400.USGS.Streamflow.Month",
IndependentTSID="06753500.USGS.Streamflow.Month")
```