Command Reference: RunningStatisticTimeSeries()

Create a new time series containing running statistics computed from input

Version 10.30.00, 2014-06-10

The `RunningStatisticTimeSeries()` command uses a sample of values from a time series to compute a running statistic, resulting in new time series. The main purposes of the command are:

1. Compute a running statistic around a moving point, in order to smooth the time series, for example to focus on underlying short-term forcings rather than variability or noise
2. Compute a statistic by using values from the historical period, for example to illustrate how a daily value compares to historical values for the same day of year
3. Compute a statistic by comparing a value to a statistic computed for a "normal" period, such as a standard 30-year climate period

The sample is computed relative to a date/time in the time series and consequently the resulting statistic may vary at each date/time in the time series. The resulting time series will have a time series identifier (TSID) that is the same as the original, with "`-Running-`" and the statistic appended to the data type (an alias can be assigned to customize the identifier that is used for processing). There are several approaches to determining the sample for the running statistic (as specified by the `SampleMethod` command parameter):

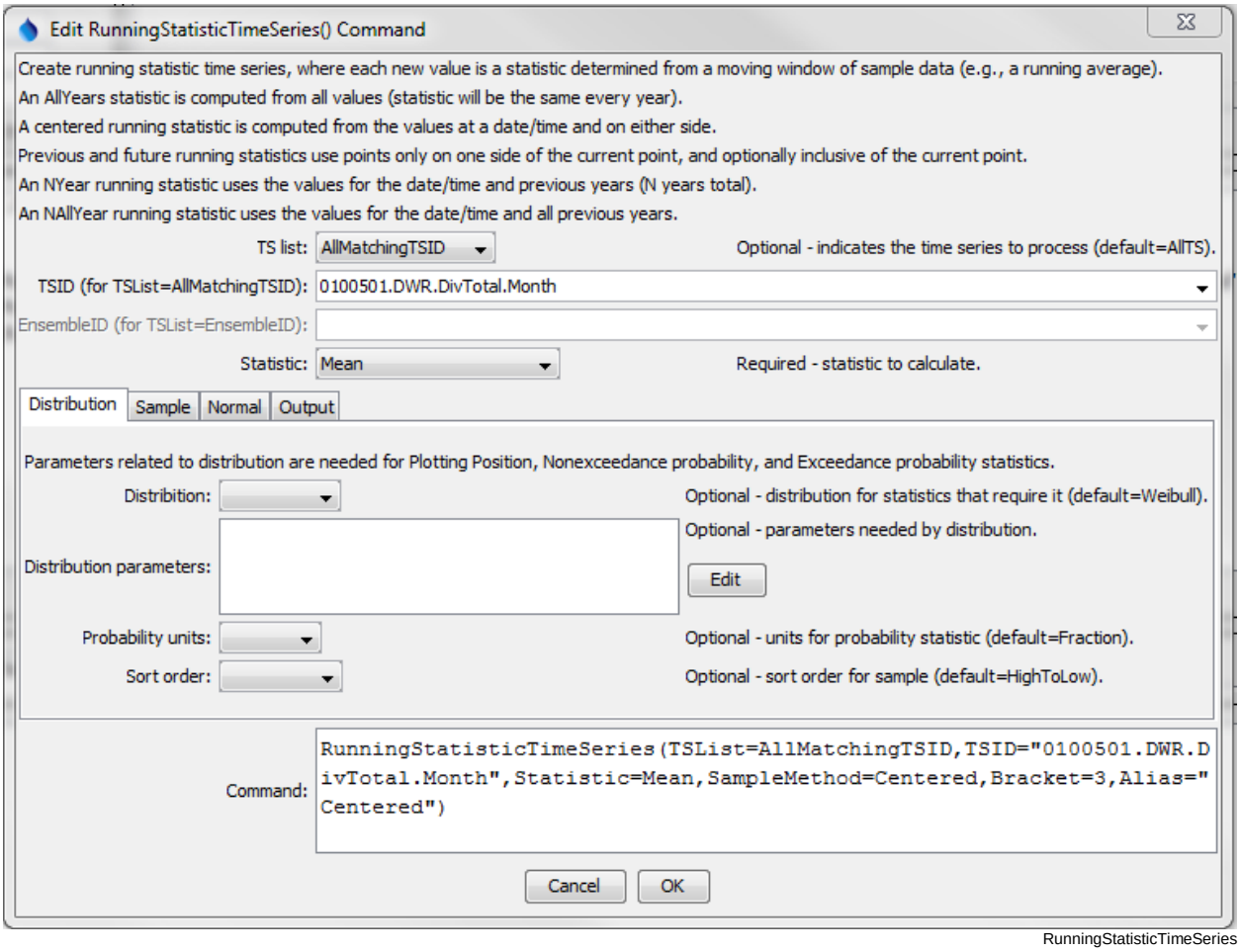
- The centered running statistic (`SampleMethod=Centered`) requires that the number intervals on each site of a point be specified (e.g., specifying 1 will use 3 values at each point).
- The previous/future running statistics (`SampleMethod=Future`, `FutureInclusive`, `Previous`, `PreviousInclusive`) require that the number of intervals prior to or after the current point be specified.
- The N-year running statistic (`SampleMethod=NYear`) is computed by processing the current year and N - 1 values from previous years, for a specific date. A resulting value is produced only if N non-missing values are available. Currently N-year running statistic values for Feb 29 for daily or finer data will always be missing because a sufficient number of values will not be found – an option may be added in the future to allow Feb 29 values to be computed based on fewer than N values.
- A special case of the N-year running statistic (`SampleMethod=NAllYear`) is to use all previous years' and the current value.
- Use `SampleMethod=AllYears` to use data from the full analysis period. In this case some statistics may have the same value for the full period. This sample method is used with `NormalStart` and `NormalEnd` and `PercentOf*` statistics to indicate how values compare to a normal period.

Statistics may be calculated directly from the sample or may be derived from an additional calculation. For example, the `Mean` statistic is computed by computing the mean of the values in the sample, and is assigned as the output time series value for the date/time that defines the sample. However, the `PercentOfMean` statistic is computed first by computing the `Mean` statistic and then dividing the original time series value by the mean, for each date/time in the time series. Derived statistics could be computed for many statistics but are provided only for cases that have common use.

Some statistics require the specification require that a distribution be specified. These statistics include `ExceedanceProbability`, `NonexceedanceProbability`, and `PlottingPosition`. See the **Statistics Summary** table for more information. The distributions are specified using the `Distribution`, `DistributionParameters`, and `ProbabilityUnits` parameters. The see **Distribution Summary** table below.

The `SortOrder` parameter is used for the `Rank` statistic and may in the future be used for statistics that use a distribution. Currently, the above statistics that use a distribution always sort data so that the largest data value is in rank position 1. The `Rank` statistic can be calculated as a simple statistic and will consider the `SortOrder` parameter.

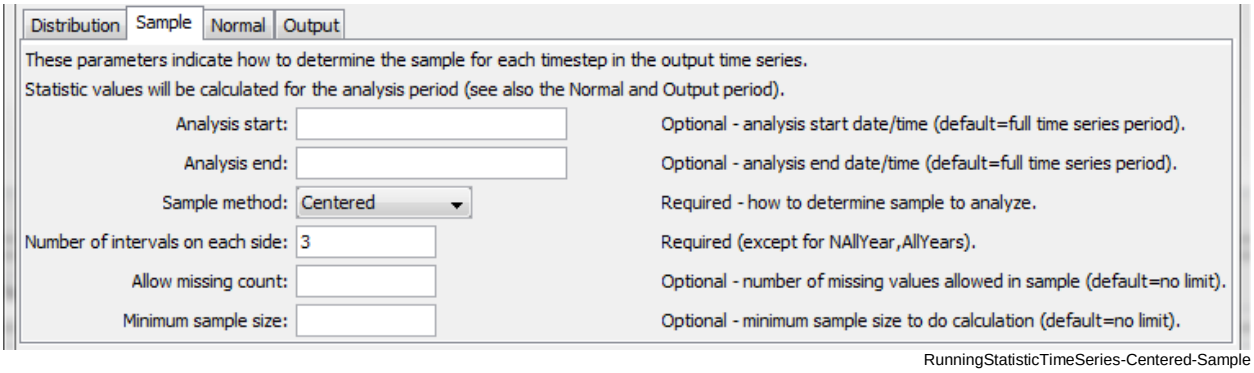
The following dialog is used to edit the command and illustrates the centered running average command syntax. In this case the distribution does not need to be specified. Refer to the sequence of figures to see input fields for all parameters.



RunningStatisticTimeSeries

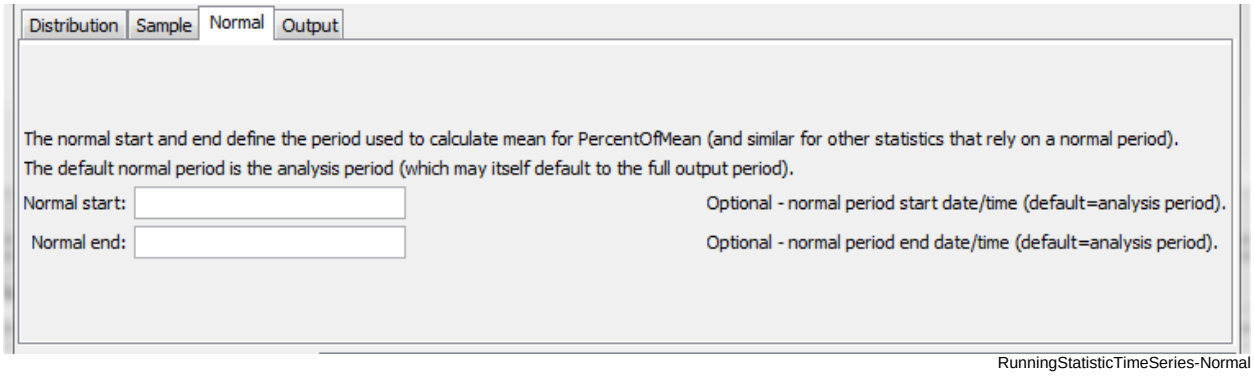
RunningStatisticTimeSeries() Command Editor – Distribution Parameters

The following dialog is used to edit the command and illustrates the centered running average command syntax.



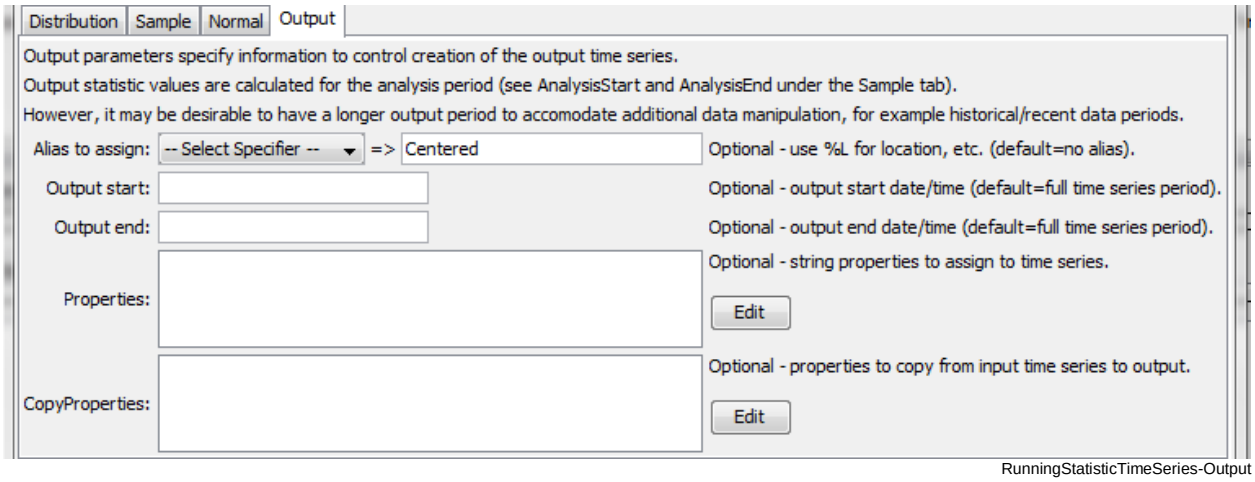
RunningStatisticTimeSeries-Centered-Sample

RunningStatisticTimeSeries() Command Editor – Sample Parameters for Centered Running Average



RunningStatisticTimeSeries-Normal

RunningStatisticTimeSeries() Command Editor – Normal Period Parameters



RunningStatisticTimeSeries-Output

RunningStatisticTimeSeries() Command Editor – Output Parameters

The command syntax is as follows:

```
RunningStatisticTimeSeries(Parameter=Value,…)
```

Command Parameters

Parameter	Description	Default
`TSList`	Indicates the list of time series to be processed, one of:	`AllTS`

Parameter	Description	Default
	1. `AllMatchingTSID` – all time series that match the `TSID` (single TSID or TSID with wildcards) 2. `AllTS` – all time series generated before the command 3. `EnsembleID` – all time series in the ensemble 4. `FirstMatchingTSID` – the first time series that matches the `TSID` (single TSID or TSID with wildcards) 5. `LastMatchingTSID` – the last time series that matches the `TSID` (single TSID or TSID with wildcards) 6. `SelectedTS` – the time series selected with the `SelectTimeSeries()` command	
`TSID`	The time series identifier or alias for the time series to be processed, using the `*` wildcard character to match multiple time series.	Required if `TSList=*TSID`.
`EnsembleID`	The ensemble to be processed, if processing an ensemble.	Required if `TSList=EnsembleID`.
`Statistic`	The statistic to compute for each point in the created time series. See the **Statistics Summary** below. Some statistics require additional input, as noted in the table.	None – must be specified.
`Distribution`	Indicates the distribution, needed for certain statistics (see **Statistics Summary** table below for indication or statistics that need distribution information). See the **Distribution Summary** table below for information about distributions.	
`Distribution Parameters`	Additional parameters needed to specify a distribution. See the **Distribution Summary** table below.	
`Probability Units`	Units to use for calculated probability statistics: • `Fraction` • `Percent` or `%`	`Fraction` (0 – 1).
`SortOrder`	Order to sort the sample, used with exceedance probability, plotting position and rank: • `LowToHigh` – rank 1 in plotting position is smallest value • `HighToLow` – rank 1 in plotting position is largest value	`HighToLow` for `Exceedance Probability`, `Nonexceedance Probability`, and `PlottingPosition`.
`AnalysisStart`	Start of period to analyze. A value will be computed at each time step in the analysis period.	Analyze the full (output) period.
`AnalysisEnd`	End of period to analyze	Analyze the full period.
`SampleMethod`	The method used to determine the data sample for each statistic calculation, one of: • `AllYears` – the sample is taken from all the years in the period (one value per year), appropriate for `PercentOfMean` and similar	None – must be specified.

Parameter	Description	Default
	statistics. See also `NormalStart` and `NormalEnd`. • `Centered` – N (bracket) values on each side of a date/time and the center value • `Future` – average the next N (bracket) values but do not include the current value • `FutureInclusive` – average the next N (bracket) values and also include the current value • `NYear` – values for the current year and (N – 1) preceding years, for the same date/time in each year • `NAllYear` – values for the current year and all preceding years, for the same date/time in each year (missing values are allowed) • `Previous` – the previous N (bracket) values but do not include the current value • `PreviousInclusive` – the previous N (bracket) values and also include the current value If a sample method such as `NAllYear` is desired, but including previous, current, and future values, then the `NewStatisticTimeSeries()` command can be used.	
`Bracket`	For centered `SampleMethod`, the bracket is the number of points on each side of the current point (therefore a value of 1 will average 3 data values). For future and previous `SampleMethod`, the bracket is the number of previous or future values. For N-year `SampleMethod`, the bracket is the total number of years to process, including the current year. The bracket is not used with sample method `NAllYear` and `AllYears`.	None – must be specified.
`AllowMissing Count`	The number of values allowed to be missing in the sample and still compute the statistic. Care should be taken to specify a value that is relatively small for the sample size.	0 – no missing values are allowed in the sample
`Minimum SampleSize`	The minimum sample size is checked with `SampleMethod=AllYears` and `SampleMethod=NAllYear` because `Bracket` and `AllowMissingCount` do not control the sample size.	`1`
`NormalStart`	Start of normal period. The normal period is used to compute an intermediate statistic such as `Mean`, which is then used in the final statistic (e.g., `Statistic=PercentOfMean`). The normal period is used for the initial calculation and the analysis period specified by `AnalysisStart` and `AnalysisEnd` are is used for the final calculation.	Analyze the full (output) period.

Parameter	Description	Default
NormalEnd	End of normal period.	Analyze the full period.
Alias	The alias to assign to the time series, as a literal string or using the special formatting characters listed by the command editor. The alias is a short identifier used by other commands to locate time series for processing, as an alternative to the time series identifier (TSID).	None – must be specified.
OutputStart	Start of the output period, use to size the output time series.	Input time series start.
OutputEnd	End of normal period.	Input time series end.
Properties	String properties to be assigned to the time series using syntax PropertyName1:Value1, PropertyName2:Value2 Use the syntax %L to specify standard time series properties as per alias specification and ${ts:Property} for user-assigned properties. The properties will be taken from the input time series.	
Copy Properties	Properties to copy from the input time series to the output time series using syntax PropertyName1:NewPropertyName1, PropertyName2:NewPropertyName2 The new property name can be specified as * to keep the old name or specify a new property name.	

The following table lists available statistics.

Statistic Summary

Statistic	Description	Needed Input
Exceedance Probability	The probability that the value will be exceeded, best-suited for the N* sample methods (see discussion below about how statistic is computed).	Requires distribution parameters.
GeometricMean	Geometric mean value.	
Lag-1Auto Correlation	The autocorrelation between values and the those that follow in the next time step, given by: $r_k = \frac{\sum_{i=1}^{N-k}(Y_i - Y_{mean})(Y_{i+k} - Y_{mean})}{\sum_{i=1}^{N}(Y_i - Y_{mean})^2}$	
Max	Maximum value.	
Mean	Arithmetic mean of values.	
Median	Median value.	
Min	Minimum value.	
Nonexceedance Probability	The probability that the value will not be exceeded, 1-ExceedanceProbability, best-suited for the N* sample methods (see discussion below about how statistics are computed).	Requires distribution parameters.
PercentOfMax	Percent of the Max statistic output.	
PercentOfMean	Percent of the Mean statistic output.	
PercentOfMedian	Percent of the Median statistic output.	

Statistic	Description	Needed Input
PercentOfMin	Percent of the Min statistic output.	
PlottingPosition	Plotting position for distribution (see ExceedanceProbability calculation explation in Statistic Computation Details table.	Requires distribution parameters.
Rank	Rank order, based on SortOrder command parameter. Duplicate values are each assigned a rank that is the average of the ranks for the duplicate values. This is necessary because selecting one of the ranks would be arbitrary. A new command parameter may be added to allow control of this behavior.	
Skew	Skew coefficient, as follows: $Cs = \frac{N \sum_{i=1}^{N}(Y_i - Y_{mean})^3}{(n-1)(n-2)s^3}$ where s = standard deviation.	
StdDev	Sample standard deviation.	
Total	Sum of values.	
Variance	Sample variance.	

The following table provides additional information about how some statistics are computed.

Statistic Computation Details

Statistic	Computation Details
Exceedance Probability, Plotting Position	1. Rank the values in the sample from highest to lowest. The largest value in the sample will be in position 1. Duplicate values are retained in the sample. 2. Search the list of ranked values: a. If the value being examined exactly matches a value in the sample: i. The matched value has a position *i* (where the largest value is in position *i=1*). ii. The exceedance probability is calculated based on the distribution: Weibull: *i/(n + 1)*, where *n* is the sample size. Gringorten: *(i – a)/(n + 1 -2a)*, where the coefficient *a* is provided with the DistributionParameters command parameter. A value of .4 is often used for hydrology data. b. If the value is outside any values in the sample (e.g., for Future and Previous sample methods where the sample array does not include the current value), then the exceedance value is not calculated and warnings are generated. In this case a different sample method should be used. c. If the value does not exactly match a value in the sample (e.g., for Future and Previous sample methods where the sample array does not include the current value): i. Find the ranked values that bound the value. ii. The exceedance probability for each bounding value is calculated as *i/(n + 1)*, where *i* is the list position (1 for the largest value) and *n* is the sample size.

Statistic	Computation Details
	iii. The exceedance probability for the specific value is interpolated from the bounding values. Note that the exceedance probability is not recomputed by adding the value to the sample. If this is desired, use the `FutureInclusive` or `PreviousInclusive` sample methods. Duplicate values are handled by using the first value found in the sequence of duplicates. This may be refined in the future similar to the Rank statistic behavior.
`Nonexceedance Probability`	`1 - ExceedanceProbability` (see notes above for `ExceedanceProbability`)

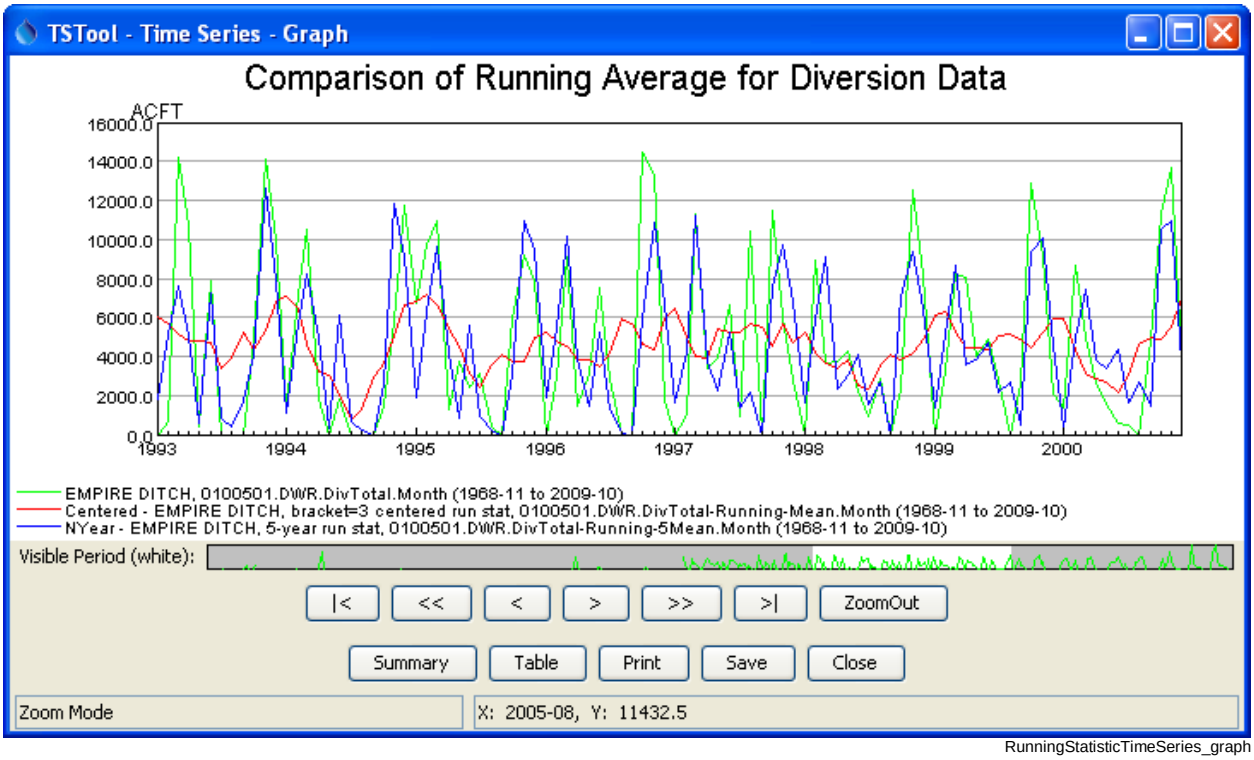
Distribution Summary

Distribution	Description
`Gringorten`	The Gringorten distribution uses plotting positions for exceedance probability: $(i - a)/(n + 1 - 2a)$ where i is the rank position for data sorted from large to small (largest value is rank 1) an a is a coefficient. Specify the coefficient using the `DistributionParameters` command parameter with `a:aValue`.
`Weibull`	The Weibull distribution uses plotting positions for exceedance probability: $i/(n + 1)$ where i is the rank position for data sorted from large to small (largest value is rank 1). No additional parameters are needed for the distribution.

A sample command file to convert State of Colorado HydroBase diversion time series to running averages is as follows:

```
# SetInputPeriod(InputStart="1993-01",InputEnd="2000-12")
# 0100501 - EMPIRE DITCH
0100501.DWR.DivTotal.Month~HydroBase
RunningStatisticTimeSeries(TSList=AllMatchingTSID,
    TSID="0100501.DWR.DivTotal.Month",Statistic=Mean,SampleMethod=Centered,
    Bracket=3,Alias="Centered")
RunningStatisticTimeSeries(TSList=AllMatchingTSID,
    TSID="0100501.DWR.DivTotal.Month",Statistic=Mean,SampleMethod=NYear,
    Bracket=5,Alias="NYear")
ProcessTSProduct(TSProductFile="Test_RunningStatisticTimeSeries_Example.tsp")
```

The resulting graph is as follows:



RunningStatisticTimeSeries_graph

Results from RunningStatisticTimeSeries() Commands

This page is intentionally blank.