Command Reference: RunningStatisticTimeSeries()

Create a new time series containing running statistics computed from input

Version 10.24.00, 2013-09-13

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

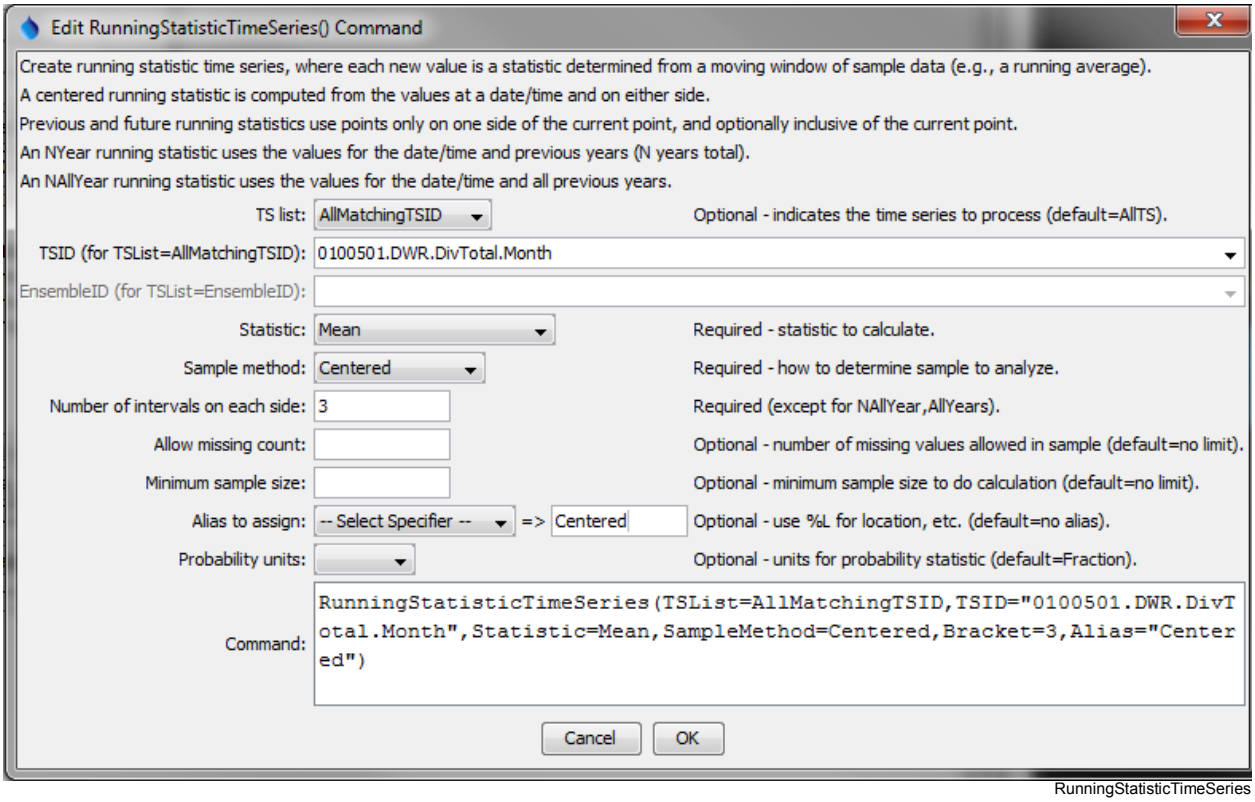
1. Compute a running statistic around a moving point, in order to smooth the time series, for example to focus on underlying short-term forcings rather than variability or noise
2. Compute a statistic by using values from the historical period, for example to illustrate how a daily value compares to historical values for the same day

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 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 statistic requires that the number of intervals prior to or after the current point be specified.
- The N-year running statistic 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 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.

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



RunningStatisticTimeSeries

RunningStatisticTimeSeries() Command Editor for Centered Running Average

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: 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	`AllTS`
`TSID`	The time series identifier or alias for the time series to	Required if

Parameter	Description	Default
	be processed, using the * wildcard character to match multiple time series.	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, one of: • ExceedanceProbability – the probability that the value will be exceeded, best-suited for the N* sample methods (see discussion below about how statistics are computed) • GeometricMean – geometric mean value • Lag-1AutoCorrelation – the autocorrelation between values and the those that follow in the next time step, given by: $r_k = \frac{\Sigma_{i=1}^{N-k}(Y_i - Y_{mean})(Y_{i+k} - Y_{mean})}{\Sigma_{i=1}^{N}(Y_i - Y_{mean})^2}$ • Max – maximum value • Mean – arithmetic mean value • Median – median value • Min – minimum value • NonexceedanceProbability – 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) • PercentOfMax – percent of the Max statistic output • PercentOfMean – percent of the Mean statistic output • PercentOfMedian – percent of the Median statistic output • PercentOfMin – percent of the Min statistic output • Skew – skew coefficient, as follows: $Cs = \frac{N\Sigma_{i=1}^{N}(Y_i - Y_{mean})^3}{(n-1)(n-2)s^3}$ where s = standard deviation • StdDev – standard deviation • Total – sum of values • Variance – variance	None – must be specified.
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 statistics	None – must be specified.

Parameter	Description	Default
	• `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
`MinimumSampleSize`	The minimum sample size is checked with `SampleMethod=AllYears` and `SampleMethod=NAllYear` because `Bracket` and `AllowMissingCount` do not control the sample size.	`1`
`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.
`ProbabilityUnits`	Units to use for calculated probability statistics.	`Fraction` (0 – 1).

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

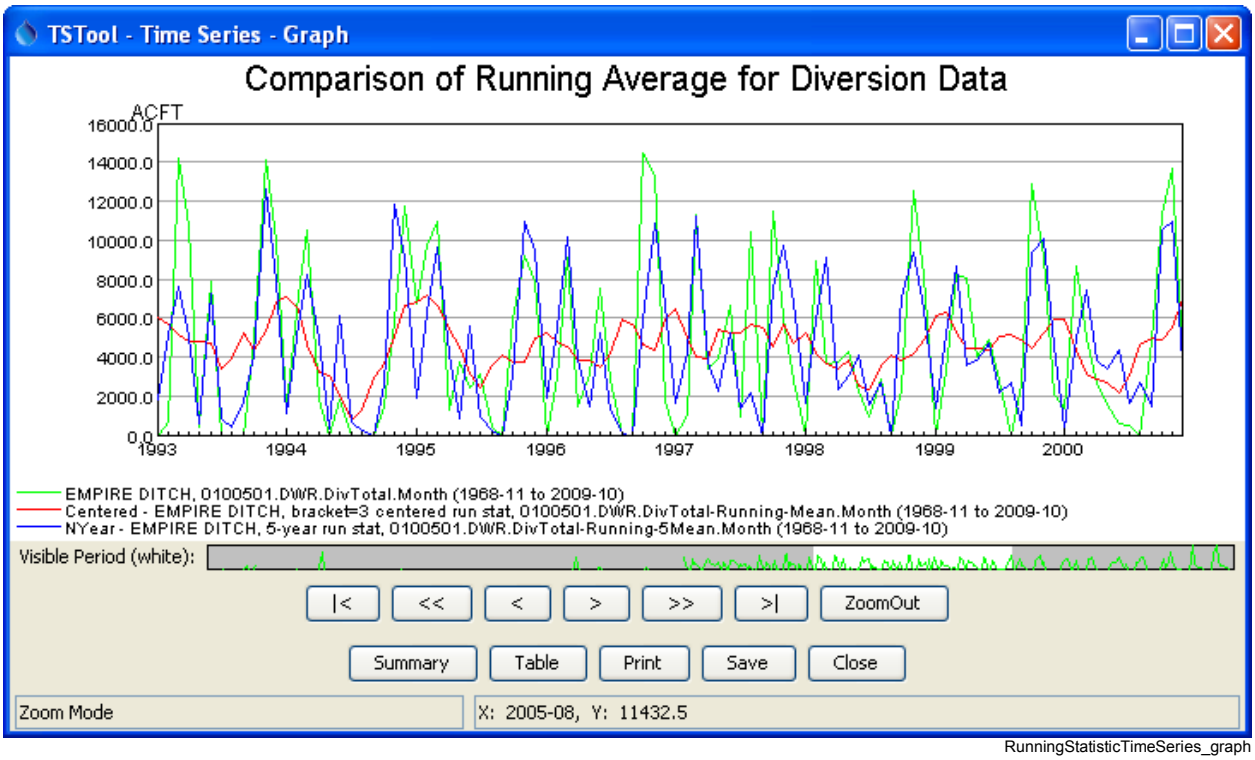
Statistic Computation Details

Statistic	Computation Details
`Exceedance Probability`	1. Rank the values in the sample from highest to lowest. Duplicate values are retained in the sample 2. Search the list of ranked values, starting from the largest: a. If the value 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 as *i/(n + 1)*, where *n* is the sample size. b. If the value is outside any values in the sample (e.g., for `Future` and `Previous` sample methods), 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): 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. 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.

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