

[Aggregation Series Slots in RiverWare 6.3](#)

Initial Analysis / Phil Weinstein / 5-24-2012

- Status: Initial writing in progress

Contents

- (1.0) Overview
- (2.0) Related Existing Capabilities in RiverWare and Supporting Applications
 - (2.1) RDF Annualizer -- Stand-alone application and Study Manager plugin
 - (2.2) RiverWare SCT Aggregated Views
 - (2.3) RiverWare Statistical Table Slot
 - (2.4) RiverWare RPL Expression Series and Scalar Slots
- (3.0) Design Analysis
 - (3.1) Supported Aggregation Periods
 - (3.2) Issue: Time aggregations defined by period start-time or end-time?
 - (3.3) Issue: Series Slots Series' timesteps have fixed alignment
 - (3.4) Supported Aggregation Period Start Time Precision (Specificity)
 - (3.5) In RiverWare, Weekly Timesteps Start on Tuesday at Midnight (Wednesday Morning)
- (4.0) Design

(1.0) Overview:

Aggregated Series Slots will be a new type of computed series slot in RiverWare which represents a temporal aggregation of some other series slot. These will be created by the user as slots on a Data Object. The configuration of an Aggregated Series Slot will include:

1. A **reference** to another series slot in the model, picked by the user.
2. A standard **timestep size** ("**aggregation period**"), larger than that of the referenced series slot.
3. A particular **periodic start time** within the period. The "precision" of the periodic start time is the timestep size of the referenced slot.*
4. A "**summary function**" for the aggregation computation, possibly including mean, median, sum, minimum, maximum, first, last.
5. A setting to indicated whether NaNs in the referenced series slot are treated as zero (or a non-value, in the case of mean and median), or reported as an error (upon evaluation).
6. An **evaluation time** -- when the Aggregation Series Slot is evaluated -- similar to the support for Rpl Expression Slots.

***-note:** The implementation of RiverWare series slot series' has a limitation which doesn't ideally support the semantics of a "non-calendar" start time for aggregation periods. For example, technically, monthly timesteps always begin on the first day of the month, and annual timesteps always begin on the first day of the year. The Aggregated Series Slot aggregation computation WILL support a user-specified periodic start time, but uses of Aggregation Series Slots in other computations within RiverWare (involving value lookup at a date/time) will have some limitations. This is explored further in this document.

(2.0) Related Existing Capabilities in RiverWare and Supporting Applications

The following sections discuss relevant capabilities currently provided by the following components:

1. RDF Annualizer -- Stand-alone application and Study Manager plugin.
2. RiverWare SCT Aggregated Views
3. RiverWare Statistical Table Slot
4. RiverWare RPL Expression Series and Scalar Slots

(2.1) RDF Annualizer -- Stand-alone application and Study Manager plugin

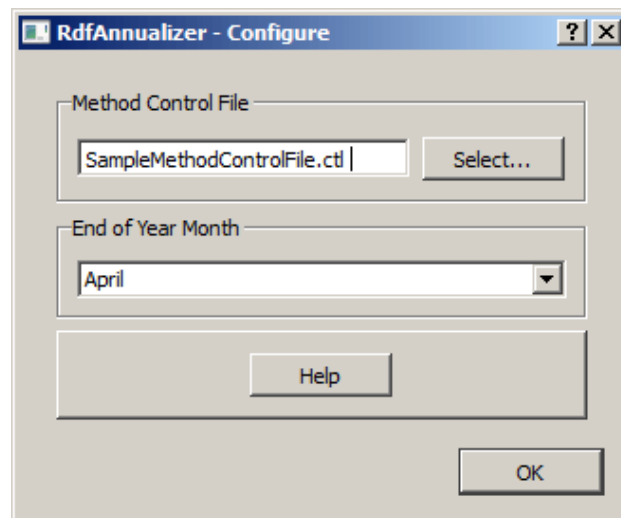
- Help File: [Stand-Alone Application](#)
- Help File: [Study Manager Plugin](#)

These tools can operate on multiple time series (which is not relevant for the new RiverWare Aggregated Series Slots). They aggregate values only into full years, with years *ending* on a specified calendar month.

The only aggregation operation configuration directly supported by the GUI (the *plugin* GUI is shown here) is the starting calendar month for annual aggregation. Aggregation *functions* (methods) are selected within a referenced "method control file". The following methods are supported:

- Sum (adds all of the values for the year)
- Average (average of all the values for the year)
- Max (maximum of all the values for the year)
- Min (minimum of all the values for the year)
- End (last value in the year)
- Begin (first value in the year)

There are actually two versions of each of the aggregation methods listed above: (1) ignore NaNs, and (2) fail on encountering a NaN, in the input series.



(2.2) RiverWare SCT Aggregated Views

The SCT displays multiple series slots with the "time" dimension along either axis. In both axis orientations, a special timestep aggregation mode is supported -- for data display only.

All series slots displayed within an SCT much have the same timestep size. An SCT's timestep aggregation can be configured with a timestep size larger than (or trivially, the same size as) the SCT's series timestep size. Also, the starting timestep for each aggregation can be configured -- for example (shown here), daily series slots aggregated to weeks starting on Saturday.

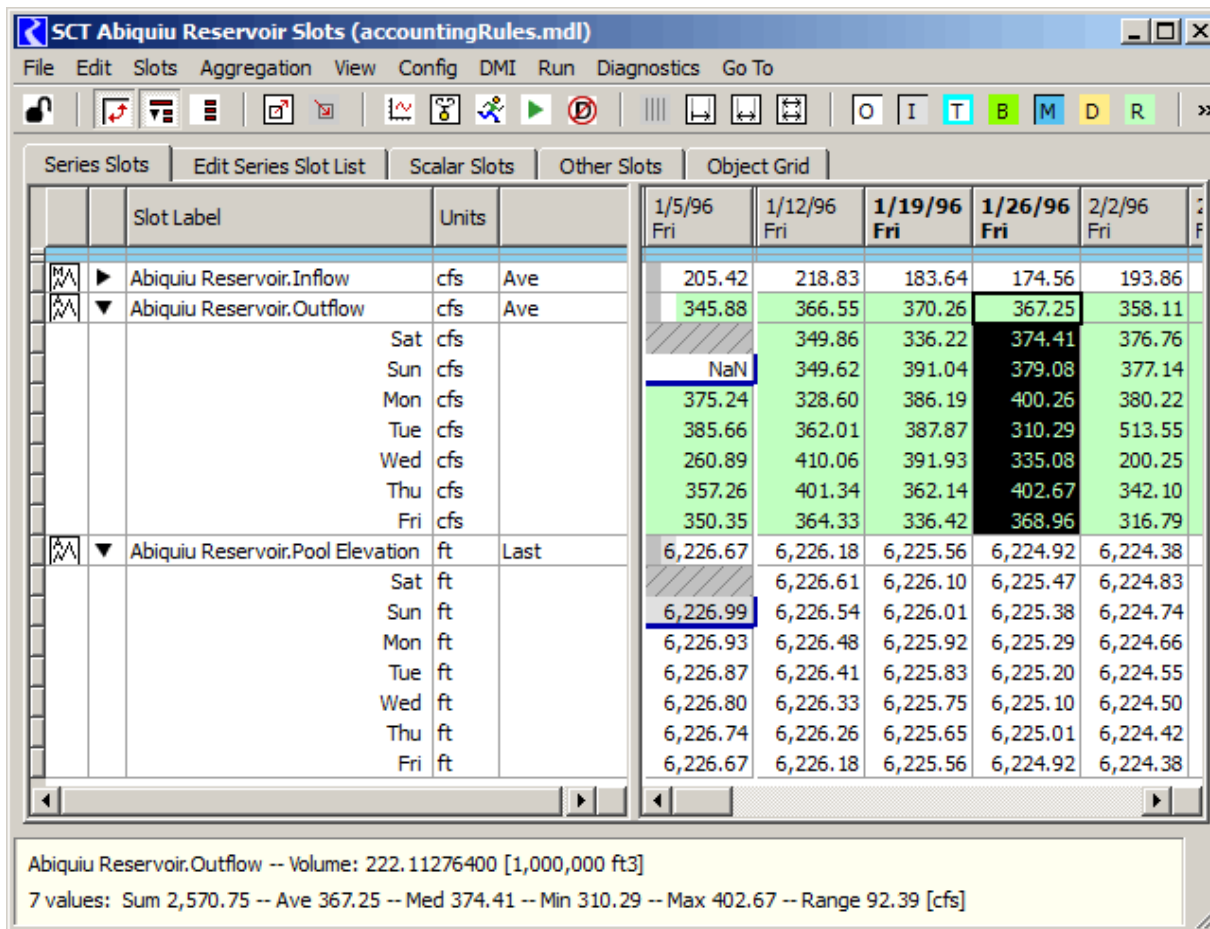
Model Timestep: 1 Day

☒ Summarize Timesteps for Regular Interval

1 Week Start At:

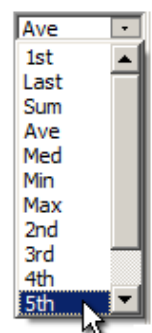
Hour: 0:00 Week Day: Sat

Month Day: 1 Month: Jan



The following aggregation functions are supported, and can be set independently for each series slot. (The SCT supports several ways of setting the series slots' aggregation functions):

- 1st value
- Nth value* (where N is an ordinal number up to the number of series timesteps in each aggregation, which is sometimes variable, e.g. in monthly aggregations of daily or smaller timesteps).
- Last value
- Sum
- Average value (mean)
- Median value* (average of middle-two values when there are an even number of values).
- Minimum value
- Maximum value



*-note: The RDF Annualizer does not support the "Nth" and "mean" aggregation functions (methods). Each of these

two functions have their own complexities, either computationally or in GUI support, so we might consider omitting these two from the Aggregation Series Slot, *at least for the initial development*.

For any particular referenced series timestep size, the SCT constrains the aggregation period to values which result in *reasonable display presentations*, being that both the original series values *and* the aggregated values are shown together, within the same display. For example, the SCT's display aggregation capability does not support annual aggregation of hourly series slots. These display-motivated constraints would not generally apply to Aggregated Series Slots.

Below are three SCT aggregation configurations supported for daily series slots.

- Notice that the weekly aggregation start day is specified as a weekday, but the monthly aggregation start day is specified as a month day [1..31].
- The SCT's supported aggregation periods do not have to be timestep sizes supported for actual series slots in RiverWare. For example, the SCT supports a quarterly ("3 Months") aggregation (which can start on any day of the year). However in the Aggregation Series Slot, only timestep sizes supported for actual series slots can be used for the Aggregation Period.
- Minor: In these three illustrated cases, the "Hour" control is not operable, and should be shown as disabled.

SCT Timestep Aggregation Conf... ? X

Model Timestep: 1 Day

☐ Fixed Timesteps per Summary

7

☒ Summarize Timesteps for Regular Interval

1 Week Start At:

Hour: 0:00

Week Day: Sun

Month Day: 1

Month: Jan

OK Cancel

SCT Timestep Aggregation Conf... ? X

Model Timestep: 1 Day

☐ Fixed Timesteps per Summary

7

☒ Summarize Timesteps for Regular Interval

1 Month Start At:

Hour: 0:00

Week Day: Sun

Month Day: 1

Month: Jan

OK Cancel

SCT Timestep Aggregation Conf... ? X

Model Timestep: 1 Day

☐ Fixed Timesteps per Summary

7

☒ Summarize Timesteps for Regular Interval

3 Months Start At:

Hour: 0:00

Week Day: Sun

Month Day: 1

Month: Jan

OK Cancel

(2.3) RiverWare Statistical Table Slot

There are two provisions in the RiverWare Statistical Table Slot where are relevant for, and could be implemented within the Open Slot Dialog for the new Aggregation Series Slot.

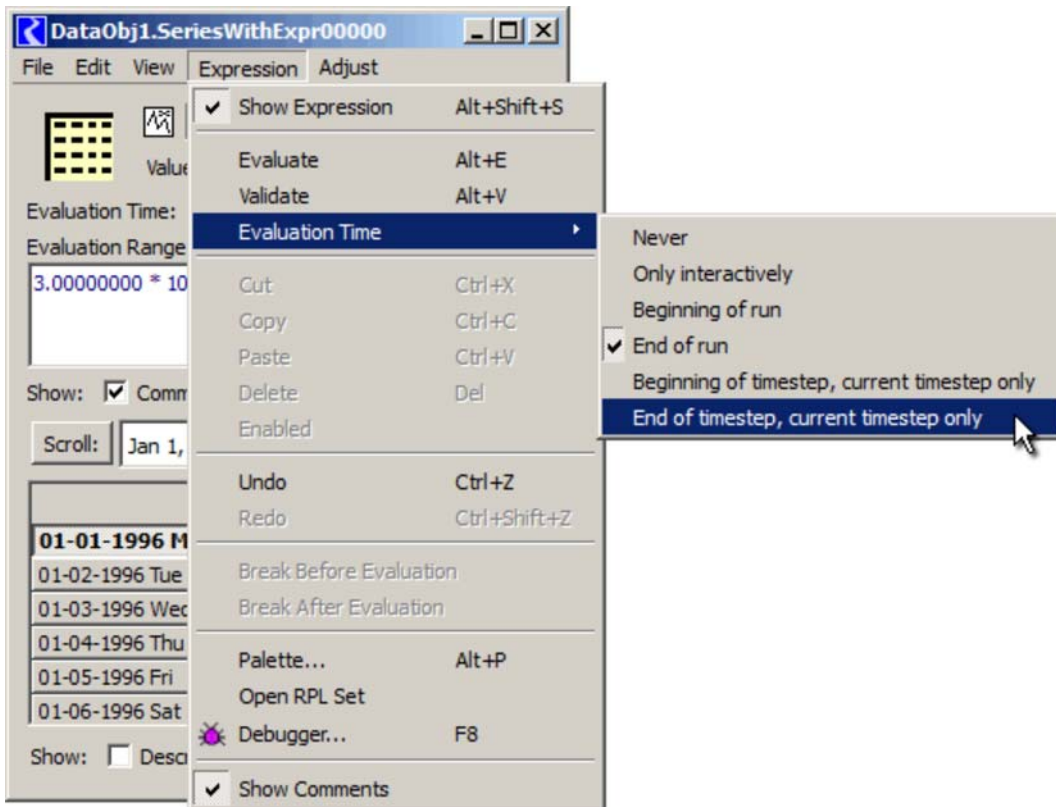
1. The Statistical Table Slot configuration includes a reference to a Series Slot in the model.
2. The choice of one particular Function is presented, in a "combo box".

The evaluation of Statistical Table Slots is strictly manual, on command by the user.

	Percent of time equaled or exceeded (percent)	Inflow (cfs)	Rank (NONE)	Date (FullDateTime)
1	0.365	1,568.06	1	24:00 April 27, 1996
2	0.730	1,388.17	2	24:00 April 26, 1996
3	1.095	1,335.90	3	24:00 April 28, 1996
4	1.460	1,315.73	4	24:00 May 12, 1996

(2.4) RiverWare RPL Expression Series and Scalar Slots

The RPL Expression Slots are relevant because they have a user-configurable "execution time", including several automatic execution settings. See image below. Of the various "Execution Times", perhaps only "Interactive" and "End of Run" and are relevant. Probably "Beginning of Run" is not relevant because the Aggregation Series Slot will be used to summarize the values of some other slot, and won't itself be the input to any other model processing. So, just an "Evaluate automatically at end of run" checkbox will be sufficient for the Aggregation Series Slot configuration.



(3.0) Design Analysis

(3.1) Supported Aggegation Periods

The valid timestep sizes for Aggregation Series Slots will depend on the timestep size of the series slot being aggregated. The table below defines the valid timestep size combinations. Notice that series slots having a timestep size smaller than one week can be aggregated into a weekly series slot, but a weekly series slot cannot be aggregated into a monthly or annual series slot.

QUESTION: Should we bother with aggregations to 6-Hour and 12-Hour timesteps? (i.e. the first two columns). Omitting those may simplify the presentation of this feature.

Input Series Time Step	Computed Aggregated Series Time Step					
	6 Hour	12 Hour	Daily	Weekly	Monthly	Annual
1 Hour	✓	✓	✓	✓	✓	✓
6 Hours		✓	✓	✓	✓	✓
12 Hours			✓	✓	✓	✓
Daily				✓	✓	✓
Weekly						
Monthly						✓
Annual						

(3.2) Issue: Time aggregations defined by period *start-time* or *end-time*?

Aggregation periods not aligned with the calendar -- e.g. a water year beginning on October 1 -- could be specified either by their start time or end time (as a design decision). That is, for the stated example, the annual aggregations would either:

- Start on October 1 at 0:00
- End on September 30 at 24:00

The SCT's aggregation specification uses the first option (start time). The RDF Annualizer uses the second option (ending month). **I recommend going with the "start" semantics**, as that's easier to understand. (Notice that the RDF Annualizer's display is a little ambiguous). And specifying March 1 (0:00) as a start time is technically difficult to state as an end date.

Of course the start-time (or end-time) can't be more precise than the timestep size of the series slot being aggregated. That is, the start-time (or end-time) needs to be aligned with that timestep size.

(3.3) Issue: Series Slots Series' timesteps have fixed alignment

This limitation was stated in the overview:

The implementation of RiverWare series slot series' has a limitation which doesn't ideally support the semantics of a "non-calendar" start time for aggregation periods. For example, technically, monthly timesteps always begin on the first day of the month, and annual timesteps always begin on the first day of the year. The Aggregated Series Slot aggregation computation WILL support a user-specified periodic start time, but uses of Aggregation Series Slots in other computations within RiverWare (involving value

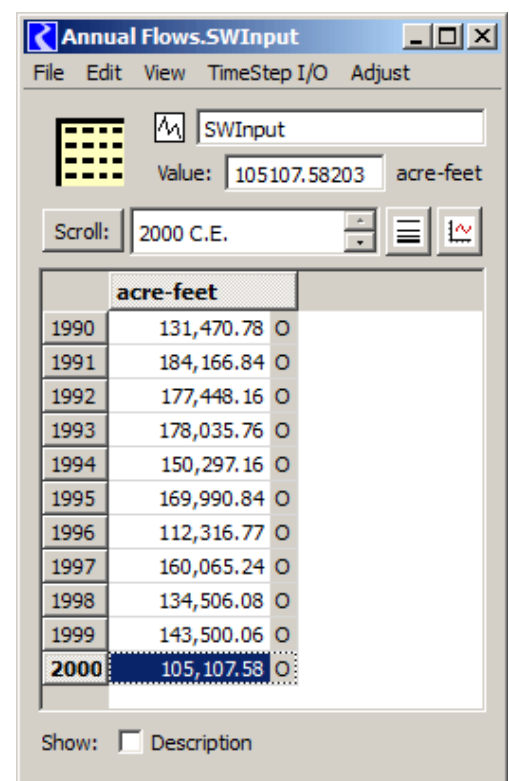
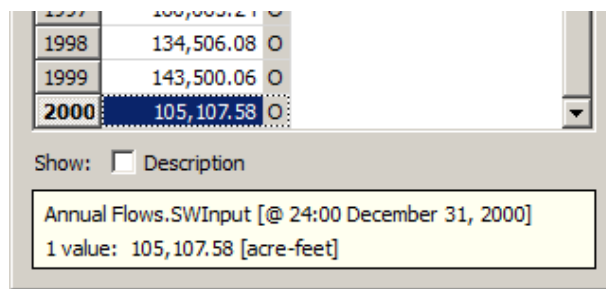
lookup at a date/time) will have some limitations.

It is a requirement that the Aggregation Series Slot support the sorts of (annual) aggregation provided by the RDF Annualizer. That is, for each year, the ability to aggregate all of the values for a whole year from, for example, October through the following September. However, the individual values within an Annual Series Slot in RiverWare are for the whole calendar year -- January through December. Similarly issues exist for smaller timeframes -- for example, monthly aggregations could start on the 10th day of each month, but timestep values in a monthly series slot are for calendar months.

For the purpose of computing the values of an Aggregation Series Slot, there is no problem with assigning a computed aggregation value to a corresponding timestep within an annual series slot. We just have to be clear about whether the nominal date of the beginning or ending of an aggregation period is used for that association -- e.g. is October 1994 through September 1995 the "1994" year or the "1995" year? (I will propose that this be an explicit user option).

Of the various display provisions for timestep date/times, e.g. in the Open Series Slot dialog, some are insufficiently precise to show the problem, but some will show the problem. For example, the row labels in the Open Series Slot dialog for an annual series slot (see **image to the right**) don't reveal that the end-time for each row is, nominally, the end of the calendar year.

However, when a single timestep cell is selected, the selection status text indicates that the timestep ends at midnight on December 31st (see **below**). For Aggregation Series Slots computed using a non-calendar period start-time, this would be incorrect.



Similar problems may exist for value lookups based on times within Aggregation Series Slots -- from RPL or other internal mechanisms. I recommend that we just document this as a known limitation of Aggregation Series Slots using non-standard start times.

POSSIBLE MODEST SERIES SLOT ENHANCEMENT: Short of implementing correct time-based value-lookup algorithms and timestep time labeling for Series Slots having non-standard timestep boundaries (e.g. months that start on the 10th day of the month), we could consider adding to the SeriesSlot class an optional specification for the **minimum precision with which timestep times are displayed**, e.g. "never show a timestep time more precise than years (or months, days) for this series slot". This would initially be applied to the Open Series Slot dialog for the support of Aggregation Series Slots.

(3.4) Supported Aggregation Period Start Time Precision (Specificity)

As illustrated in the SCT's timestep aggregation support (above), whether exact hour, weekday, month-day, and/or

month is specified in the periodic start time depends on the timestep size of the referenced series slot and of the Aggregated Series Slot. This is summarized in the following table.

Input Series Time Step	Computed Aggregated Series Time Step					
	6 Hour	12 Hour	Daily	Weekly	Monthly	Annual
1 Hour	hour	hour	hour	hour w-day	hour m-day	hour m-day month
6 Hours		6-hour	6-hour	6-hour w-day	6-hour m-day	6-hour m-day month
12 Hours			12-hour	12-hour w-day	12-hour m-day	12-hour m-day month
Daily				w-day	m-day	m-day month
Weekly						
Monthly						month
Annual						

Specifier	Valid Values
hour*	0..5 or 0..11 or 0..23
6-hour*	0, 6, 12, 18
12-hour*	0, 12
w-day	Sun, Mon, Tue, ... Sun
m-day	1..31

*These would be the hourly values which would be used if time aggregations are defined with respect to a periodic *start* time. In, instead, aggregations are defined iwth respect to a periodic *end* time, then supported hour values will be 1 to 24.

(3.5) In RiverWare, Weekly Timesteps Start on Tuesday at Midnight (Wednesday Morning)

Our internal "epic" base is January 1, 1800 (Wednesday). In our end-of-timestep venacular, that's 24:00 December 31, 1799 (Tuesday). Our weekly timesteps are based on the number of seconds since that epic base moment, divided by the number of seconds in a week.

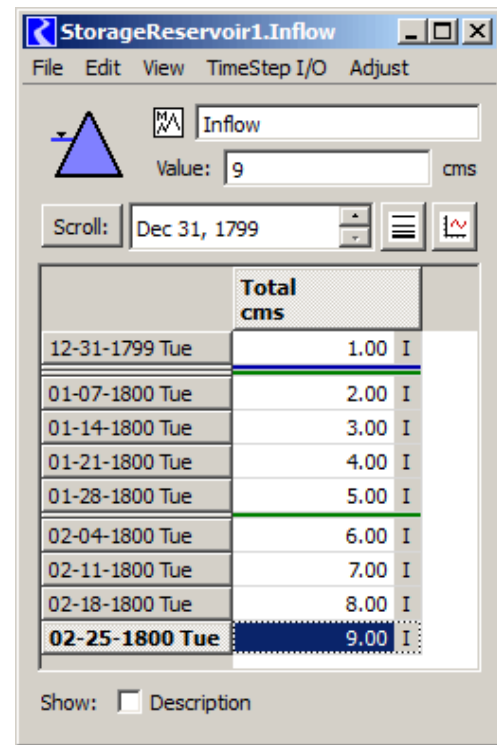
This means that:

1. Weekly aggregations will typically use a non-standard periodic start time.
2. Timestep labels in the Open Slot Dialogs for weekly Aggregation Series Slots **will generally be wrong**.

(4.0) Design

(not yet written).

--- (end) ---



Calendar for January 1800

Sun	Mon	Tue	Wed	Thu	Fri	Sat
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	