

Calamari Reports Documentation

Edit

New Page

[Jump to bottom](#)

sbodorkos edited this page on Aug 30, 2016 · 9 revisions

Calamari produces six reports as '.csv' files from the Prawn XML file selected, each named using the first eight letters of the XML file followed by an underscore (e.g. '100142_G_'), followed by either 'SBM_' or 'NOSBM_', followed by either 'LINREG_' or 'SPOTAV_', followed by one of the headings listed below that begin with either 'Check_' or 'SQUID_'.

As an example, a report from processing the demo XML '100142_G6147_10111109.43.xml' file using SBM and SPOTAV parameter choices, yields the filename '100142_G__SBM_SPOTAV_Check_01_IonIntegrations_PerScan.csv'.

All six files are included inside another folder named '[YYYYMMDD-HH24MISS]_X[Y]', where the first part is the system time-stamp for folder creation (unidirectional downward from year to second, and with HH24 = hour in 24-hour format), [X] is either 'SBM' or 'NOSBM', and [Y] is either 'LINREG' or 'SPOTAV'.

In turn, this folder is included inside another named the same as the Prawn XML file, without the '.csv' extension.

All report folders are included inside the user's choice of a CalamariReports folder, which defaults to a folder names 'CalamariReports_v[CalamariVersionNumber]' in the same folder as the Calamari '.jar' file that is executing the analysis.

The six report formats follow:

Check_01_IonIntegrations_PerScan

This CSV is a simple and direct extract from the Prawn XML file, with one row per scan, and 11 columns per species (comprising one for 'count_time_sec', and one for each of the 10 measured ion-count 'integrations'). The output was initially used to verify the function of the biweight algorithm with tuning 9 (in the higher count-rate nuclides) and replicate SQUID's Poisson outlier identification and rejection (in the lower count-rate nuclides).

It has 5 'left-hand' columns, which are used to uniquely identify the rows and sort them:

- **Title** = analysis-specific text-string read from XML
- **Date** = analysis-specific date read from XML, expressed as YYYY-MM-DD HH24:MI:SS
- **Scan** = integer, starting at 1 within each analysis
- **Type** = 'ref mat' or 'unknown'; analyses with prefix 'T' labelled 'ref mat', all others 'unknown'
- **Dead_time_ns** = analysis-specific integer read from XML

These are followed by 11 columns for each species (i.e. 110 columns for the demo XML '100142...'):

- **[entry-label].count_time_sec** = analysis-specific integer read from XML
- **[entry-label].1** = integer value corresponding to the first of 10 'integrations' within tags ' ' for the specified combination of analysis, scan and species
- **[entry-label].2** = integer value corresponding to the second of 10 'integrations' within tags ' ' for the specified combination of analysis, scan and species
- ...
- **[entry-label].10** = integer value corresponding to the tenth of 10 'integrations' within tags ' ' for the specified combination of analysis, scan and species

Row Sorting: Primary criterion = Date (ascending), secondary criterion = Scan (ascending).

For the 10-peak demo XML ('100142...'), the array has 684 rows of data (114 analyses x 6 scans), and 115 columns (5 'left-hand' columns, then for each of the 10 measured species, 11 columns comprising count_time_sec and the integer values of the 10 integrations).

Check_02_SBMIntegrations_PerScan

This CSV is a simple and direct extract from the Prawn XML file, with one row per scan, and 11 columns per species (comprising one for 'count_time_sec', and one for each of the 10 measured SBM 'integrations'). The output was initially used to verify the function of the biweight algorithm with tuning 6. The format is similar to Check_01... in many respects, so points of difference are indicated in bold:

It has 5 'left-hand' columns, which are used to uniquely identify the rows and sort them:

- **Title** = analysis-specific text-string read from XML
- **Date** = analysis-specific date read from XML, expressed as YYYY-MM-DD HH24:MI:SS
- **Scan** = integer, starting at 1 within each analysis
- **Type** = 'ref mat' or 'unknown'; analyses with prefix 'T' labelled 'ref mat', all others 'unknown'
- **SBM_zero_cps** = analysis-specific integer read from XML

These are followed by 11 columns for each species (i.e. 110 columns for the demo XML '100142...'):

- **[entry-label].count_time_sec** = analysis-specific integer read from XML
- **[entry-label].SBM.1** = integer value corresponding to the first of 10 'integrations' within tags ' ' for the specified combination of analysis, scan and species
- **[entry-label].SBM.2** = integer value corresponding to the first of 10 'integrations' within tags ' ' for the specified combination of analysis, scan and species
- ...
- **[entry-label].SBM.10** = integer value corresponding to the first of 10 'integrations' within tags ' ' for the specified combination of analysis, scan and species

Row Sorting: Primary criterion = Date (ascending), secondary criterion = Scan (ascending).

For the 10-peak demo XML ('100142...'), the array has 684 rows of data (114 analyses x 6 scans), and 115 columns (5 'left-hand' columns, then for each of the 10 measured species, 11 columns comprising count_time_sec and the integer values of the 10 integrations).

SQUID_01_TotalCounts_IonsAndSBM_PerScan

This CSV is an array matching as closely as possible the 'condensed PD/XML' worksheet of a SQUID-workbook (which usually takes the name of the machine-data file created by the SHRIMP control software). The calculations incorporated into this array are described here:

https://github.com/CIRDLES/ET_Redux/wiki/SHRIMP:-Step-1. The output has one row per scan, and 5 columns per species (one for each key attribute of 'total counts at peak' as described below).

It has 4 'left-hand' columns, which are used to uniquely identify the rows and sort them:

- **Title** = analysis-specific text-string read from XML
- **Date** = analysis-specific date read from XML, expressed as YYYY-MM-DD HH24:MI:SS
- **Scan** = integer, starting at 1 within each analysis
- **Type** = 'ref mat' or 'unknown'; analyses with prefix 'T' labelled 'ref mat', all others 'unknown'

These are followed by 5 columns for each species (i.e. 50 columns for the demo XML '100142...'):

- **[entry-label].Time** = integer 'time_stamp_sec' read from XML for the specified combination of analysis, scan and species
- **[entry-label].TotalCounts** = calculated decimal value for 'total counts at mass', for the specified combination of analysis, scan and species

- **[entry-label].1SigmaAbs** = calculated decimal value for '+/-1sigma at mass', for the specified combination of analysis, scan and species
- **[entry-label].TotalSBM** = calculated decimal value for 'total SBM counts', for the specified combination of analysis, scan and species
- **[entry-label].TrimMass** = decimal 'trim_mass_amu' read from XML for the specified combination of analysis, scan and species

Row Sorting: Primary criterion = Date (ascending), secondary criterion = Scan (ascending).

For the 10-peak demo XML ('100142...'), the array has 684 rows of data (114 analyses x 6 scans), and 54 columns (4 'left-hand' columns, then for each of the 10 measured species, 5 columns as specified above).

SQUID_02_NuclideCPS_PerSpot

This CSV is an array matching the contents of the 'total [mass] cts/sec' columns within the StandardData and SampleData worksheets of a SQUID-workbook. The calculations incorporated into this array are described here: https://github.com/CIRDLES/ET_Redux/wiki/SHRIMP:-Step-2. The output has one row per analysis, and one column per species.

It has 3 'left-hand' columns, which are used to uniquely identify the rows and sort them:

- **Title** = analysis-specific text-string read from XML
- **Date** = analysis-specific date read from XML, expressed as YYYY-MM-DD HH24:MI:SS
- **Type** = 'ref mat' or 'unknown'; analyses with prefix 'T' labelled 'ref mat', all others 'unknown'

These are followed by 1 column for each species (i.e. 10 columns for the demo XML '100142...'):

- **[entry-label].TotalCps** = calculated decimal value for 'total counts per second, for the specified combination of analysis and species

Row Sorting: Primary criterion = Type (alphabetical ascending), secondary criterion = Date (ascending).

For the 10-peak demo XML ('100142...'), the array has 114 rows of data (114 analyses), and 13 columns (3 'left-hand' columns, then one for each of the 10 measured species).

SQUID_03_WithinSpotRatios_PerScanMinus1

This CSV is an array matching as closely as possible the 'Within-Spot Ratios' worksheet of a SQUID-workbook. The calculations incorporated into this array require a user choice regarding whether the data are to be SBM-normalised or not, and are described here:

https://github.com/CIRDLES/ET_Redux/wiki/SHRIMP:-Step-3. The output has one row per 'interpolated time' (i.e. 'Ndod', where $Ndod = Nscans - 1$), and 3 columns per predefined 'ratio of interest' (comprising interpolated time, 'ratio of interest' value, and 1sigma absolute error in the ratio-value).

It has 4 'left-hand' columns, which are used to uniquely identify the rows and sort them:

- **Title** = analysis-specific text-string read from XML
- **Date** = analysis-specific date read from XML, expressed as YYYY-MM-DD HH24:MI:SS
- **Ndod** = integer, starting at 1 within each analysis
- **Type** = 'ref mat' or 'unknown'; analyses with prefix 'T' labelled 'ref mat', all others 'unknown'

These are followed by 3 columns for each of the 10 predetermined 'ratios of interest', each of which has a label of the form '[nominal mass of numerator]/[nominal mass of denominator]', abbreviated below as [NUM/DEN]:

- **[NUM/DEN].InterpTime** = calculated decimal value of 'RatEqTime', for the specified combination of analysis, bracketing scans, and 'ratio of interest'
- **[NUM/DEN].Value** = calculated decimal value of 'RatEqVal' (labelled by looking up a list of the 'ratios of interest'), for the specified combination of analysis, bracketing scans, and 'ratio of interest'
- **[NUM.DEN].1SigmaAbs** = calculated decimal value for 'RatEqErr', for the specified combination of analysis, bracketing scans, and 'ratio of interest'

Row Sorting : Primary criterion = Type (alphabetical ascending), secondary criterion = Date (ascending), tertiary criterion = Ndod (ascending).

For the 10-peak demo XML ('100142...'), there are 10 predefined 'ratios of interest', so the array has 570 rows of data (114 analyses x 5 Ndod), and 34 columns (4 'left-hand' columns, then for each of the 10 'ratios of interest', the 3 columns specified above).

SQUID_04_MeanRatios_PerSpot

This CSV is an array matching the contents of the '[NUM]/[DEN]' and associated '%err' columns within the StandardData and SampleData worksheets of a SQUID-workbook, but also incorporates some other metadata. The calculations incorporated into this array require a user choice regarding whether 'spot mean' isotopic ratios are to be calculated from their constituent Within-Spot Ratios as time-invariant 'spot averages', or linear regression of the Within-Spot Ratios vs time, interpolated to the time-midpoint of the analysis. The calculations are described here: https://github.com/CIRDLES/ET_Redux/wiki/SHRIMP:-Step-4 and at the associated URLs documenting the four subroutines invoked during the course of the calculation. The output has one row per analysis, and 3 columns per predefined 'ratio of interest' (comprising index of identified outlier [if any], 'ratio of interest' mean value, and 1sigma percentage error in the ratio-mean value).

It has 3 'left-hand' columns, which are used to uniquely identify the rows and sort them:

- **Title** = analysis-specific text-string read from XML
- **Date** = analysis-specific date read from XML, expressed as YYYY-MM-DD HH24:MI:SS
- **Type** = 'ref mat' or 'unknown'; analyses with prefix 'T' labelled 'ref mat', all others 'unknown'

These are followed by 3 columns for each of the 10 predetermined 'ratios of interest', each of which has a label of the form '[nominal mass of numerator]/[nominal mass of denominator]', abbreviated below as [NUM/DEN]:

- **[NUM/DEN].MinIndex** = integer defining the index number of any Dodson point meeting the criteria for definition as an 'outlier', for the specified combination of analysis and 'ratio of interest'. Values of 0 indicate that no Dodson point qualifies as an outlier; values of -2 indicate 'ratios of interest' that do not qualify for the Dodson interpolation routine (usually because either the numerator or the denominator is characterised by very low count-rates)
- **[NUM/DEN].Value** = calculated decimal value of 'RatioMean' (labelled by looking up a list of the 'ratios of interest'), for the specified combination of analysis and 'ratio of interest'
- **[NUM.DEN].1SigmaPct** = calculated percentage value of 'RatioMeanSig', for the specified combination of analysis and 'ratio of interest'

Row Sorting: Primary criterion = Type (alphabetical ascending), secondary criterion = Date (ascending).

For the 10-peak demo XML ('100142...'), the array has 114 rows of data (114 analyses), and 33 columns (3 'left-hand' columns, then for each of the 10 'ratios of interest', the 3 columns specified above).