

RFC: nagg -A <seconds> option to specify aggregation size

Larry Knox

Aggregation size is an integer number of granules specified in **nagg** either directly, or by length of time depending on user preference. The `-n` option directly specifies the number of granules per file and was implemented in the initial nagg prototype. The `-A` option determines the number of granules per file according to the requested minimum time span of an aggregation. The actual aggregation time span for full aggregations will never be smaller than the specified time, but will often be larger, since there must be an integral number of granules in an aggregation.

1 Purpose

This RFC describes the `-A` option for specifying the desired minimum aggregation time span to be produced by nagg.

2 Requirements

- For aggregation size specified by `-A seconds`, the number of granules per HDF5 output file is given by: $\text{number} = \text{ceiling}(\frac{\text{seconds}}{\text{granule time span}})$. The effective aggregate time span is the granule time span (duration) times number which may be greater than the user specified seconds.
- Specification of the aggregation size with `-A seconds` will result in an aggregation size with the smallest integer number of granules that provides the specified seconds of data. Partial aggregations produced at the beginning or end may have fewer granules.
- The actual number of seconds of data will be at least the specified number but will not exceed the specified number by more than the granule duration for the product.
- Aggregation size may be specified by `-n` or `-A`. If both are supplied, the tool should return an error. If neither are supplied the tool defaults to `-n 1`. Interactions with a future `-l` option are not yet determined.

3 Update for Reference Manual

These are the current and proposed reference manual entries for the `-A` option. The `-l` option and making the output like the first file encountered when `-l`, `-t`, `-n` and `-A` are omitted are not yet specified or implemented.

3.1 Current entry

-A seconds

(To be supported in future implementation.)

The number of seconds in each aggregate file. Aggregation size is a fixed number of granules per file equal to seconds divided by the period of the granule rounded up to the next integer number of granules

If `-l`, `-t`, `-n` and `-A` are omitted, then the first NPP data product file encountered will be used to determine the `-t` list and `-n` number.

3.2 Proposed entry

-A seconds

The minimum number of seconds to be included in each aggregate file. Aggregation Size is a fixed number of granules per file equal to seconds divided by the period of the granule rounded up to the next integer number of granules. The actual number of seconds of data per aggregation will be the next integer number of granules times the duration of a granule for the product.

If `-l`, `-t`, `-n` and `-A` are omitted, then the first NPP data product file encountered will be used to determine the `-t` list and `-n` number.

4 User Guide additions

Data granules in JPSS files have a fixed size, or duration for each product. The data products may be obtained de-aggregated with one granule per file or in standard aggregations with some standard number of granules in each file. If the original number of granules, or aggregation size is not well suited for a particular application, nagg will re-aggregate the granules from the input files specified into new files with the number of granules specified by either directly with `-n <number>` or as calculated from `-A <seconds>` according to the granule duration for the product.

The `-A` option allows for specifying a minimum aggregation time span without knowing the granule duration or calculating the required number of granules. The nagg tool rounds the number of granules up to the next integer number required to provide the requested time span. Consequently, the actual time span may exceed the requested time span by almost as much as the granule duration. If this difference might be an issue, the granule duration for a product in microseconds can be found in Appendix 3 of `doc/NPP-Aggregation-Tool-Components.docx` or by running `"nagg -h"`.

Beginning and ending aggregations may be partial, since nagg does not produce leading and trailing fill granules. Those aggregations may have fewer granules and a smaller time span than for a full aggregation. Partial aggregations are determined by the alignment of the input files and the pre-determined bucket boundaries for the requested aggregation size.

There are currently 3 granule durations for JPSS data products. A request for a 5 minute (300 seconds) aggregation size should result in the following aggregate time spans.

For products (Cris, ATMS) with granule duration 31.997 seconds: 5 minutes, 19.97 seconds.
 For products (OMPS) with granule duration 37.405 seconds: 5 minutes, 36.645 seconds.
 For products (VIIRS) with granule duration 85.35 seconds: 5 minutes, 41.4 seconds.

The difference between the actual and requested aggregation times is cumulative. Therefore for a `-A 300` (5 minutes) request for a VIIRS product with input files spanning an hour, 10.54 aggregations will be created, where the initial expectation may have been 12 aggregations for an hour's data.

In the case for the OMPS product above the aggregation time requested results in an actual aggregation time that is nearly the entire granule duration larger than the requested time. If the user prefers the actual time of one less granule, the granule number can be specified directly with `-n`. For the OMPS product `-n 8` results in an aggregation time span of 4 minutes, 59.24 seconds.

Example Command

The full hour of single granule SVM05 files for the command below includes 42 SVM05 files, and 42 matching GMODO files with geo-location information. The output includes one file at the beginning with one granule for SVM05 and one for GMODO, 10 files with 4 granules of each product, and one file at the end with a single granule of each product. The partial files are a result of the alignment of the bucket boundaries with this particular hour.

```
nagg -A 300 -t SVM05 -S SVM05_npp_d20120623_t11*.h5
```

Example 4.1	Total granules	Granules per aggregation for each product	Files (containing full and partial aggregations)	Hyperslabs in each raw data dataset for a full aggregation
Input aggregations	42	1	42	1
Output aggregations	42	4	12 (10 full, 2 partial)	4

This command produces packaged GMODO-SVM05... files. Adding the `-S` option would produce separate files for each product.

5 Test Specification

1. `-A` and `-n` should not appear together: should return error
2. With no `-A`, `-n` is used; if no `-n`, `n = 1`
3. Aggregate time span is not less than request time and aggregate time span - request is less than granule duration.