# RFC: Nagg Option to Specify Minimum Time of Aggregation

Larry Knox
Albert Cheng
Elena Pourmal

## 1. Terms

Full aggregation: set of granules in a file containing the number of granules specified by –n or calculated from -A for each product in the file.

Partial aggregation: set of granules in a file containing less that the specified number of granules for each product. Partial aggregations are produced by nagg when the first or last input granules are not aligned with a predetermined bucket boundary, since leading and trailing fill granules are not produced.

Aggregation span: the time difference between the start of the first granule in an aggregation and the end of the last granule in the aggregation.

Granule duration: the time difference between the beginning and ending time of a granule.

Aggregation number: the number of granules in an aggregation or the number of granules of each product in a file.

## 2. Purpose

This RFC describes the nagg command line option for specifying the desired minimum time for a full aggregation.

## 3. Introduction

The aggregation number is an integer number of granules for each included product specified in **nagg** either directly or calculated according to a specified length of time. The –n option specifies the number of granules per file and was implemented in the initial nagg prototype. The –A option proposed in this document determines the number of granules per file according to the requested minimum time span of a full aggregation. The actual aggregation 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.

## 4. Requirements

The basic requirement for the aggregation time option is to determine the correct number of granules to provide the requested time for each aggregation. The other three requirements are to check for correct command line syntax and provide proper feedback when it is not correct.

- The Aggregation Time option, -A seconds, directs the Nagg tool to aggregate just enough granules of a single product so that the aggregation span for a full aggregation is at least "seconds" long. The tool will set the number of granules for a full aggregation to the smallest number needed to provide the requested aggregation span.
- Seconds must be a positive integer. The tool will reject entries for the –A parameter that are less than 1 or are not integers.
- Aggregation size may be specified by –n or –A, which are mutually exclusive. If both are supplied, the tool will return an error.
- -A is optional. The tool will determine the number of granules per aggregation by another method if it is absent.

## 5. 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.

**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.

**Proposed entry**

-**A** seconds

The minimum number of seconds to be included in each aggregate file. The actual number of seconds per full aggregation = ceiling( seconds / granule_duration).

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.

## 6. User Guide additions

Data granules in JPSS files have a fixed 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 the "nagg –h" command.

Each full aggregation will have the same number of granules and the same time span. Partial aggregations will have fewer granules and may be produced for beginning and ending aggregations, since nagg does not produce leading and trailing fill granules. Partial aggregations 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.

## 7. Test Specifications

1. -A and -n should not both be given. "nagg -A 2 -n 3 ..." should fail.
2. -A is optional. "nagg -n 3 ..." should succeed and produce files with 3 granules each.
3. -A must have a parameter. "nagg -A ..." should fail.
4. The -A parameter must be a positive integer. "-A 0", " -A -2" , or "-A 319.97" should all fail.
5. The number of granules determined by nagg must be just sufficient to provide an aggregation span that is not less than the time requested. Tests should have enough input granules to produce at least three output files. The first and last files should be ignored since they may be partial aggregations.
    a. “nagg –A 159 –t REDRO <input files>” should produce intermediate files with 4 REDRO granules each.
    b. “nagg –A 160 –t REDRO <input files>” should produce intermediate files with 5 REDRO granules each, as should any –A from 160 through 191.
6. The – A specification may appear anywhere in the list of options before the list of input files. “nagg –A seconds -t <products> -g 0 <input list>” and “nagg –g0 –t <products> -A seconds <input list> should both succeed for valid products and inputs.