



Argo FileChecker

ADMT-18 – Hamburg
November 2017

Mark Ignaszewski
FNMOC



Overview

FileChecker examines every file before it is allowed on the (dac/ directory) of the GDAC

Two types of checks:

1) Format Checks

2) Data Consistency Checks

Format Checks:

- Compares the format of every file with the standard in the Users Manual.
- Checks: Global attributes, Dimensions, Variables, Variable attributes (names and settings)
 - Ensures <PARAM> variables only in correct file type (core- / bio-)



Overview

Data Consistency Checks: *---Sanity checks---*

- Values of date/time variables
- Values of all “reference table variables” are checked against the official table.
- Variable inter-comparisons
 - Example: <PARAM> and <PARAM>_QC
DATA_MODE, <PARAM> and <PARAM>_ADJUSTED
- Additional D-mode profile file checks defined by the DM group.
- Does **NOT** do any cross-file checks (ie, meta-data file compared to profile file)

Applied only to v3.1 files only.



Overview

Three Levels of Results:

- **ACCEPTED:** No problems were detected
File is distributed on the GDAC
- **WARNING:** One or more non-critical problems were detected
File is distributed on the GDAC
Warning e-mail is sent to DAC
- **REJECTION:** One or more critical problem was detected
File is rejected --- not distributed on the GDAC
Error e-mail is sent to DAC



Overview

Documentation of FileChecker processing is available on the ADMT website.

- latest update posted on the ADMT site prior to this meeting

The executable system, specification files (reference tables), source code, and documents are available on the US GDAC (including an “installation and user’s guide”)

<http://usgodae.org/pub/outgoing/argo/etc/FileChecker/>

History of all updates available



Developments during past year

- New/modified checks:
 - TEMP: No longer required in trajectory files; still required in profile files (July 2017)
 - <PARAM>_QC: Cannot be “unchecked” (0) in core-profile files (July 2017)
 - CONFIG_MISSION_NUMBER: (October 2017)
 - Cannot be 0 (zero)
 - *** Currently: This test only generates warnings. Will cause rejections starting ...
 - Can be FillValue in R-mode profile and trajectory files
 - Cannot be FillValue in D-mode files
 - Meta-data files: FillValue generates warnings (but is allowed)
 - DATA_TYPE: Non-standard values are rejected (v3.1 only) (June 2017)
- Implemented STATUS column of reference tables (October 2017)
 - Discussed at ADMT-17 (Active, Publication underway, Creation underway, Deprecated, Obsolete)
 - Allowed: Active, Publication underway, or blank (no status)
 - Rejected: Creation underway, Obsolete, Refused
 - Warning: Deprecated ← Supports “transition periods” during updates to new settings
- Allow “templates” in reference tables (Jan 2017)
 - For instance, to allow a “sensor number” to be appended SENSOR settings



Developments during past year

- Reference table updates:
 - Add new float types to reference tables (Oct 1026, Jan, May, Sept, Oct 2017)
 - New/updated SENSOR and SENSOR_MODEL (January 2017)
 - Allow sensor number to be appended to SENSOR (September 2017)
 - SENSOR_MODEL (27) contain deprecated values. Transition to obsolete ...
 - Configuration and Technical parameter names (January, March, April 2017)
 - Physical parameter names (October 2017)
- Latest version of code (v2.5) promoted to operations (Oct 2016)
- Testing of full Trajectory consistency checks on all v3.1 files (January 2017)
 - Duplicate all submitted v3.1 trajectory files in a “test area”
 - Errors and Warnings shared with DACs



Current Operational Status

Operational at both GDACs

Format checks performed on every incoming file

Data consistency checks performed on every v3.1 file

...full Trajectory check testing in progress...

Trajectory File Data Consistency Checks



Current operational version includes:

- Meta-data checks
 - DATA_CENTRE, PLATFORM_NUMBER, PLATFORM_TYPE, etc
- Reasonable date checks:
 - Similar to other file types
 - DATE_CREATION, DATE_UPDATE, etc

Current TEST version adds:

- N_MEASUREMENT variable checks
- N_CYCLE variable checks
- Inter-dependence between these

(By far the most complicated of the data consistency checks)

Trajectory File Data Consistency Checks



Results of Trajectory files submitted 1-24 November 2017

	Submitted	Accepted	% Accepted	Rejected
v3.1 Files	19156	17019	88.8%	2137
AOML	9177	9125	99.4%	52
BODC	118	118	100%	0
Coriolis	3140	3077	98.0%	63
<i>CSIO*</i>	<i>109</i>	<i>0</i>	<i>0%</i>	<i>109</i>
CSIRO	5888	4255	72.3%	1633
INCOIS*	63	50	79.4%	13
JMA	428	373	87.1%	55
<i>KMA*</i>	<i>56</i>	<i>0</i>	<i>0%</i>	<i>56</i>
KORDI	0	0	0%	0
MEDS	342	21	6.1%	321
<i>NMDIS*</i>	<i>0</i>	<i>0</i>	<i>0%</i>	<i>0</i>

* Was not correctly capturing submitted files. Numbers not complete.



Future Developments

- Check of config and tech parameter names that contain “templates”
 - For example: “<short_sensor_name>”: There is an approved list of these values.
- Check TECHNICAL_PARAMETER_VALUES to ensure their setting agrees with the specified “data type”.
 - Test is “proposed” in the current document
- Format checks: Validate that attribute settings match expected type
- Cross-file checks: Compare information in a file against its Metadata file



N_MEASUREMENT Variable Checks

- **CYCLE_NUMBER / CYCLE_NUMBER_ADJUSTED**
 - Launch cycle (-1) only in first value; all others ≥ 0
 - Real-time file (all cycles in R-mode)
 - Cannot be fill value
 - **CYCLE_NUMBER_ADJUSTED** must be FillValue (*core-file only*)
 - Delayed-mode file (at least one D-mode cycle; may contain R-mode cycles)
 - R/A-mode cycles: **CYCLE_NUMBER** must be set
 - **CYCLE_NUMBER_ADJUSTED** (*core-file only*)
 - D-mode cycle: must be set
 - R/A-mode cycle: must be FillValue
- **MEASUREMENT_CODE**
 - Cannot validate that codes are applied correctly, only that they are valid values
 - Cannot be FillValue
 - One of:
 - Primary/Secondary codes (MC): multiple of 50
 - Relative special codes: within range MC-15 to MC-1
 - Specific special codes: list of defined special codes



N_MEASUREMENT Variable Checks

- JULD / JULD_STATUS / JULD_QC:
 - JULD_QC: Reference table 2 or blank
 - JULD_STATUS: Reference table 19 or blank
 - JULD_QC / JULD_STATUS
 - If JULD_QC = blank, JULD_STATUS = blank
 - If JULD_QC = '9', JULD_STATUS = '9'
 - JULD
 - If FillValue, JULD_QC must be blank or '9'
 - If set, JULD_QC must *not* be blank or '9'
 - If JULD_QC indicates “good”
 - Date must be after 1997-01-01
 - Must be before DATE_UPDATE (with a 2 day buffer to allow for clock drift)
- JULD_ADJUSTED / JULD_ADJUSTED_STATUS / JULD_ADJUSTED_QC:
 - *core-file only*
 - Very similar to above, plus
 - If JULD_ADJUSTED set
 - If associated JULD is missing, this represents “an estimation” and DATA_MODE for this cycle must be 'A' or 'D'



N_MEASUREMENT Variable Checks

- **LATITUDE / LONGITUDE / POSITION_QC / POSITION_ACCURACY**
 - POSITION_QC: Reference Table 2 or blank
 - POSITION_ACCURACY: Reference table 5 or blank
 - Where POSITION_QC = blank or '9', LATITUDE/LONGITUDE set to FillValue
 - Otherwise, LATITUDE/LONGITUDE must not be FillValue
- **<PARAM> / <PARAM>_QC**
 - <PARAM>_QC: Reference Table 2 or blank
 - Where <PARAM>_QC = blank or '9', <PARAM> must be FillValue
 - Otherwise, <PARAM> must not be FillValue
- **<PARAM>_ADJUSTED / <PARAM>_ADJUSTED_QC / <PARAM>_ADJUSTED_ERROR**
 - Data Mode = 'R'
 - <PARAM>_ADJUSTED / <PARAM>_ADJUSTED_ERROR: All FillValue
 - <PARAM>_ADJUSTED_QC: All blank or '9'



N_MEASUREMENT Variable Checks

- $\langle \text{PARAM} \rangle_ADJUSTED / \langle \text{PARAM} \rangle_ADJUSTED_QC / \langle \text{PARAM} \rangle_ADJUSTED_ERROR$ (continued)
 - Data Mode = ‘A’
 - $\langle \text{PARAM} \rangle_ADJUSTED_QC$: Reference Table 2 or blank
 - Where $\langle \text{PARAM} \rangle_ADJUSTED_QC = “ “$ (not measured)
 - $\langle \text{PARAM} \rangle_QC = “ “$ and $\langle \text{PARAM} \rangle_ADJUSTED = \text{FillValue}$
 - Core-parameters: $\langle \text{PARAM} \rangle_ADJUSTED_ERROR$ set to FillValue ($\langle \text{PARAM} \rangle_ADJUSTED_ERROR$ may be set for bio-parameters)
 - Where $\langle \text{PARAM} \rangle$ is FillValue, $\langle \text{PARAM} \rangle_ADJUSTED, *_QC$ is FillValue
 - DATA_MODE = ‘D’
 - $\langle \text{PARAM} \rangle_ADJUSTED_QC$: Reference Table 2 or blank
 - Where $\langle \text{PARAM} \rangle_ADJUSTED_QC = “ “$ (not measured)
 - $\langle \text{PARAM} \rangle_QC = “ “$ and $\langle \text{PARAM} \rangle_ADJUSTED = \text{FillValue}$
 - Where $\langle \text{PARAM} \rangle$ is not FillValue and $\langle \text{PARAM} \rangle_ADJUSTED$ is FillValue
 - $\langle \text{PARAM} \rangle_ADJUSTED_QC = ‘4’$ or ‘9’
 - Where $\langle \text{PARAM} \rangle$ is not FillValue and $\langle \text{PARAM} \rangle_ADJUSTED$ is not FillValue
 - $\langle \text{PARAM} \rangle_ADJUSTED_QC \neq ‘4’$ or ‘9’
 - Where $\langle \text{PARAM} \rangle_ADJUSTED_QC \neq ‘4’$ or ‘9’: $\langle \text{PARAM} \rangle_ADJUSTED_ERROR$ is not FillValue



N_CYCLE Variable Checks

- DATA_MODE: Set to 'R', 'A', or 'D'
- CYCLE_NUMBER_INDEX / CYCLE_NUMBER_INDEX_ADJUSTED
 - Real-time file:
 - CYCLE_NUMBER_INDEX (*Core- and bio- files*)
 - Cannot be FillValue
 - Must be ≥ 0 (Cycle -1 not allowed)
 - No duplicates
 - CYCLE_NUMBER_INDEX_ADJUSTED (*Core-files only*)
 - Must be FillValue



N_CYCLE Variable Checks

- CYCLE_NUMBER_INDEX / CYCLE_NUMBER_INDEX_ADJUSTED
 - Delayed-mode file (may also contain R/A-mode cycles):
 - CYCLE_NUMBER_INDEX (*Core- and bio- files*)
 - R/A-mode cycle: Cannot be FillValue
 - Must be ≥ 0 (or FillValue)
 - No duplicates (among set values)
 - CYCLE_NUMBER_INDEX_ADJUSTED (*Core-files only*)
 - D-mode cycle: Must be set
 - R/A-mode cycle: Must be FillValue
 - Must be ≥ 0 (Cycle -1 not allowed)
 - No duplicates (among set values)



N_CYCLE Variable Checks

- JULD_* / JULD*_STATUS: (*Core-files only*)
(eg, JULD_DESCENT_START, JULD_PARK_START, JULD_FIRST_MESSAGE, etc)
 - JULD_*
 - If FillValue, JULD*_STATUS must be blank or ‘9’
 - If set, JULD*_STATUS must *not* be blank or ‘9’
 - JULD*_STATUS: Reference table 19 or blank
- GROUNDED: Reference table 20 (*Core-files only*)
- CONFIG_MISSION_NUMBER:
 - R/A-mode: Not 0 (zero)
 - D-mode: Not 0 or FillValue
 - Except CYCLE_NUMBER = 0, which may be FillValue

N_MEASUREMENT / N_CYCLE

Variable Inter-dependence



JULD[N_MEASUREMENT] / JULD_* [N_CYCLE]

The JULD_*[N_CYCLE] variables correspond to significant events in the JULD[N_MEASUREMENT] variable. Cycle number and measurement code “map” JULD values to the specific JULD_* variables.

- For example, MC = 100 maps to JULD_DESCENT_START for a particular cycle

The following checks are performed:

- Every JULD and JULD_STATUS value from one of the significant measurement codes is compared to the associated JULD_* value on a cycle-by-cycle basis.
 - If they are not the same, the file is rejected.
 - Code (703) is used for a series of values within a single cycle. The associated JULD_* variables (JULD_FIRST/LAST_LOCATION) correspond to the first and last value within a single cycle. For the purposes of the checks, “first” and “last” are determined by the sequential position within the JULD variable.
- Every value within a JULD_*[N_CYCLE] variable that is not associated with a JULD value must be set to FillValue (STATUS = ‘ ‘ or ‘9’).

N_MEASUREMENT / N_CYCLE

Variable Inter-dependence



- CYCLE_NUMBER / CYCLE_NUMBER_INDEX:
 - Every cycle number in CYCLE_NUMBER (except “-1”) must be in CYCLE_NUMBER_INDEX
 - Every cycle number in CYCLE_NUMBER_INDEX must be in CYCLE_NUMBER



JULD ordering

This test has been proposed. It has not been coded or tested in any way.

The events in the N_MEASUREMENT arrays are required to be in the order that they occurred. Due to clock drift, you cannot simply compare JULD values for “different types of events” (real-time versus adjusted versus delayed mode; float versus satellite) to determine they are in the correct order.

The proposal: Test for proper ordering by ensuring that successive JULD values are ascending (or the same) within a given “type of event”, as determined by measurement codes. Examples of the “types” being:

Real-time file:

- Float-measured, R-mode times
- Float-measured, A-mode times
- Satellite-measured, R-mode times
- Satellite-measured, A-mode times

Delayed-mode file:

- D-mode cycle: Float-measured
- D-mode cycle: Satellite-measured
- Non-D-mode cycles will be ignored

The QC values and whether you are comparing JULD or JULD_ADJUSTED will all have to be taken into account.