QC Configuration Options for CORDEX

H.-D. Hollweg

DKRZ, Hamburg

Date: 2013-05-29

Outline

- Introduction
- Best Practise
- Rules

  SELECT, LOCK
- List of Options
- Description

Introduction

The QC processing is controlled by configuration options supplied in one or more files and on the command-line. Configuration options are given by

```
Syntax:  KEY-WORD
Syntax:  KEY-WORD  = value[,value ...]
Syntax:  KEY-WORD += value[,value ...]
```

The syntax of options *SELECT* and *LOCK* is slightly different.

```
Syntax:  KEY-WORD    [path[=]][variable]
Syntax:  KEY-WORD  = [path[=]][variable]
Syntax:  KEY-WORD += [path[=]][variable]
```

Please, note that underlined items provide navigation in this paper. Thus, clicking on this will jump to the explanation of syntax rules.

Configuration options may be added to the call of the `qcManager` script on the command-line, given within a configuration files specified on the command line by `-f conf-file` (synonym: `task file`), and by option *QC_CONF=conf-file* embedded in the task file (there is no restriction in nesting depth). The precedence of configuration options is given below from highest to lowest.

- on the command-line
- in the `-f task-file`
- *QC_CONF* assignments embedded in the `task file`
- successively found in a chain of files specified by embedded *QC_CONF* assignments.

A list of available configuration options is given here. A click on every option leads to a description.

Best Practise

- Use configuration options on the command-line only for testing or re-checking a small set of data.

- The configuration options provided by the `task file` should contain directives frequently modified, e.g. choosing only sub-sets of available data.

- Specify embedded *QC_CONF* assignments for long-term options.

- *PROJECT_DATA* should only contain a site specific path to the data set,
  e.g. *PROJECT_DATA=/hdh/data/CORDEX*.

- Use the *SELECT* option for partitioning the amount of data to be checked. E.g. *SELECT AFR-44* would check the data of every model available for the given domain of CORDEX data. On the other hand, *SELECT AFR-44/INST/MODEL/EXPERIMENT/.*/fx=orog* would find any orography file for a specified institute, model, and experiment within every version available.

- Option EXP_PATH_INDEX aids to partition the QC results. Please, choose unique names among the total path components to data. E.g. let the path be
  /path$_{10}$/AFR-44$_{9}$/SMHI$_{8}$/CCCma-CanESM2$_{7}$/historical$_{6}$/r1i1p1$_{5}$/SMHI-RCA4$_{4}$/v1$_{3}$/day$_{2}$/clh$_{1}$, index is given by subscripts only for illustration, and have *EXP_PATH_INDEX=9,8,7,6,5,3*, then the resulting name would be *AFR-44_/SMHI_CCCma-CanESM2_historical_r1i1p1_v1*.

- Execution in the background is recommended, perhaps additionally with `nohup`. If the script is run in the foreground, then command-line option *-m* may be helpful by showing the current filename under investigation on a status-line below the script call.

- Examine the QC results in folder *QC_RESULTS/.../check_logs*. Annotations issued by the QC are put in the particular sub-folder *_Notes/-a-name-based-on-EXP_PATH_INDEX-* within files *annotation_outline.txt* and *annotation_summary.txt*.

Before starting to check data, please make sure that everything was set properly:

- `/your-path/QC-0.4/scripts/qcManager -f file -E_SHOW_CONF`
  Inspect the configuration options displayed on the screen.

- `/your-path/QC-0.4/scripts/qcManager -f file -E_SHOW_EXP`
  Path and and filename of every SELECTed item will be displayed below the executed command-line call. Searching the data base may take a rather long time, depending on the amount of data.

- `/your-path/QC-0.4/scripts/qcManager -f file -E_NEXT`
  Only the first path-variable resulting from the *SELECT* evaluation will be checked. When everything looks fine in folder *QC_RESULTS/. . ./check_logs*, then restart the call without `-E_NEXT`. This will resume the session. Execution in the background is recommended, additionally with `nohup`.

- `/your-path/QC-0.4/scripts/qcManager -f file -E_CLEAR=note,lock`
  Whenever a check was accomplished, some annotations occurred, and the provider of data applied any fixes, then a check is redone only for variable-frequency files that had been locked before or an annotation had been issued.

Rules

KEY-WORDS regulating the QC run.

```
Syntax:  KEY-WORD
Syntax:  KEY-WORD  = value[,value ldots]
Syntax:  KEY-WORD += value[,value ldots]
```

- `KEY-WORD`
  Consisting only of alphanumeric characters and underscore; only upper-case with alphabetic begin. `KEY-WORD` and following = or += may be separated by spaces.

- `KEY-WORD`
  Set alone enables a corresponding feature. Equivalent to `KEY-WORD=t` .

- `KEY-WORD=value`
  Make a simple assignment; any previous assignment is overwritten; also by `KEY-WORD` alone.

- `KEY-WORD=value1,value2, ...`
  Assign a comma-separated-expression. This will be expanded to an array.

- `KEY-WORD+=value`
  Add another item to a comma-separated expression (or just assign).

- Configuration options may be supplied on the command line with the prefix `-E_`.

- Options `SELECT` and `LOCK` have a syntax of their own. Additionally, `-S` param and `-L` param may be given alternatively on the command-line.

- Invalid/misspelled `KEY-WORD` is discarded without notice. The effect of invalid assignments is trapped sometimes.

SELECT, LOCK, -S and -L:

```
Syntax:  KEY-WORD    [path[=]][variable]
Syntax:  KEY-WORD  = [path[=]][variable]
Syntax:  KEY-WORD += [path[=]][variable]
```

Spaces between `KEY-WORD` and the assignment is equivalent to = assignment, i.e. overriding. `path` may not necessarily point to a directory where specified variables, i.e. netcf files, reside. If a path is incomplete, then all sub-paths are searched for occurrences of filenames beginning with `variable`. Command-line options `-S` ands `-L`, respectively, may be used for convenience (same syntax).

Specification for `variable` may be ambiguous. For instance, `tas` would result in selecting files begining with *tas*, *tasmin*, and *tasmax*, etc. . An unambiguous selection of *tas* alone requires `tas_`.

Examples:

- *path=var*
  a single path where to look for a file beginning with variable name `var`
- *path1,path2=var*
  two paths to look for the same variable.
- *path1,path2=*
  two paths with every variable, equivalent to path1,path2=.*
- *path1=var1,var2*
  one path with two variables
- *token1[,token2 . . .]0*
  a) no '/' in any token –> selecting variables
  b)'/' anywhwere –> only paths; no trapping of errors
- relative path without a '/' must have '=' appended
- *path1=,path2=var . . .*
  This is an ERROR

List of Options

Options are grouped in sections. Please, note that a click on an option for would lead to a detailed description.

Section: Customisation

PROJECT_DATA

QC_RESULTS

SELECT

LOCK

QC_BIN

Section: Files and Tables

APPLY_MAXIMUM_DATE_RANGE

CF_STANDARD_NAMES

CHECK_LIST

COMPLAIN_EMPTY_FILE

DEREFERENCE_SYM_LINKS

DISTRIBUTED_FS

FORCE_STND_TABLE

HIDDEN_DIRECTORY_DESCENT

IGNORE_BROKEN_LINKS

IGNORE_LOCK_FILES

PROJECT_TABLE_PREFIX

PT_PATH_INDEX

SYNC_FILE_AMBIGUITY_CHECK

TABLE_REQUIREMENTS

TABLE_STANDARD

TABLE_UPDATE_ALWAYS

TIME_TABLE

Section: Command-line used Options

CLEAR

CHECK_TOOLS

DEBUG_C

DEBUG_E

DEBUG_M

DEBUG_X[=script]

FLOW_TRACE

NEXT[=num]

NEXT_RECORDS

ONLY_SUMMARY

SHOW_CALL

SHOW_CONF

SHOW_EXP

SHOW_NEXT[=num]

REUSE_PATH_LIST

Section: Processing

EMAIL_SUMMARY

EMAIL_TO

ENABLE_DISK_SPACE_INQ

EXP_PATH_INDEX

GROUP_NAME

HARD_SLEEP_PERIOD

IGNORE_TEMP_FILES

INQ_DISK_SPACE

NEVER_BREAK_SESSION

NICE

NIGHT_SHIFT

NUM_EXEC_THREADS

QC_CONF

QC_EXEC_HOSTS

REATTEMPT_LIMIT

WORK_AT_LOW_LOAD

Section: Miscellaneous

ARITHMETIC_MEAN

CASE_INSENSITIVE_VARIABLE_NAME

CHECK_MODE

CHECKSUM

CHECKSUM_DIR

CHECKSUM_FILE

DISABLE_INF_NAN

EXCLUDE_VARIABLE

IGNORE_REF_DATE_ACROSS_EXP

IGNORE_REFERENCE_DATE

NON_REGULAR_TIME_STEP

NOTE

NOTE_ALWAYS

NOTE_LEVEL_LIMIT

OUTLIER_TEST

PROJECT

QC_VERSION

REPLICATED_RECORDS

TIME_LIMIT

TRACKING_ID_ONLY

Section: Frequency Distribution

FREQ_DIST

FD_TIME_PART

FD_PROPERTY_PATH

FD_EXPLICIT_PROPS

FD_PLAIN

FD_BARS

Section: HLEV

HLEVD

HLEVD_DELETE_FILE

HLEVD_DETREND

HLEVD_KEEP_THRESHOLD_FILE

HLEVD_THRESHOLD

Description

Descriptions of options are extracted by script. Values in brackets following '#' indicate a default for the given option.

Section: Customisation

PROJECT_DATA Path to the root of directory tree(s) with netCDF data files. Note: the contained sub-directory tree(s) are appended to QC_RESULTS. Note: this option is mandatory.
Example: PROJECT_DATA=/work/ik0555/cordex/archive/CORDEX

QC_RESULTS Path to QC results and logs. Note that results will be inegrated into the path to QC-0.4 by default (not recommended).
Example: QC_RESULTS=/work/bm0021/hdh/QC_results/CORDEX/CCLM

SELECT Selection of paths and variables (RegExp of the 'expr' command, i.e. full specification from the beginning of the word). Paths and variables are given by '[path1[,path2,..]][=][var1[,var2,..]]'.

- The '=' character is required when both path and variable items are specified. It is also required for paths without any '/'.
- Omission of path components selects variables in all sub-paths cointained in PROJECT_DATA.
- Omission of variable component(s) selects all variables in the path(s).
- By default, all files in PROJECT_DATA (and sub-trees) are selected.

Full syntax rules and examples at the end of this file.
Example: SELECT path=var
Example: SELECT AFR-44/CLMcom/MPI-ESM-LR/historical/r1i1p1/CCLM4-8-17

LOCK Locking of variables (takes precedence over SELECT). See SELECT for a description.
Example: LOCK .*/Z= # [disabled]

QC_BIN Purpose: If the QC package is installed on a file system accessed by machines of different architectures and if more than one is dedicated to run the QC, then these require different names for the directories of residing executables. By default: your-path/QC-0.4/bin . The bin-directories must have been specified during the installation process. If the path is relative, the directory has to be a sub-dir of QC-0.4. If directories are prefixed by the hostname(s) of computer(s) separated by ':' from the directory, then multiple directories may be specified as comma-separated-list. The name(s) of the computer(s) must be available in the list of hosts, i.e. QC_EXEC_HOSTS. A name without a prefix is dedicated for the host on which qcManager is running (but could also have a prefix). You have to compile all executables into all bin-directories specified.
Example: QC_BIN=bin # surge:bin,another-host:bin2 #[bin]

Section: Files and Tables

APPLY_MAXIMUM_DATE_RANGE This option controls whether the second time-stamp in the filename expresses the sharp end (deadline) of the period or whether the period in the stamp has to be extended to the end, e.g. 2000-03 <=> 2000-03-31 for extended. The difference is explained best by examples.

- Extended: 2000 - 2001 <=> 2000-01-01T00:00:00 - 2001-12-31T24:00:00

If there is a mix for the components of the stamps, then use this syntax: 'yx-mx-dx-Hx-Mx-Sx' with x=s or x=e for sharp and extended respectively. Omission of any component sets individually x=s, i.e. ye-ms is equivalent to ye-ms-ds-Hs-Ms-Ss. Sharp is applied as default.
Example: APPLY_MAXIMUM_DATE_RANGE=ye-ms # [disabled]

CF_STANDARD_NAMES The XML table with CF standard names and units is automatically converted into csv format, but only if more recent. If no path is provided, the table is supposed to reside in QC-0.4/tables. Default: no check against CF standard names.
Example: CF_STANDARD_NAMES=cf-standard-name-table.xml

CHECK_LIST This table is telling the QC-program about notification codings, discarding certain test, issuing email, and grade of severeness of findings.
Example: CHECK_LIST=CORDEX_check-list.conf

COMPLAIN_EMPTY_FILE The QC does not complain about an empty data file by default, because the QC could run on-the-fly during the creation of data files. If it is clear that there shall never be an empty file in the data directory tree, then issuing a message in the log-files can be enabled.

DEREFERENCE_SYM_LINKS The original directory tree structure of the source data is preserved, i.e. a symbolic link in the data causes making a corresponding symbolic link in the directory tree of the QC results, too. Enable this option if symbolic links are to be dereferenced, i.e. genuine data are to replace the symbolic links in the QC result tree.

DISTRIBUTED_FS If distributed computing is also involved with distributed file system, then enable this option. Bi-directional communication and data transfer is done by ssh and scp, respectively. IMPORTANT NOTE: this is not installed and not projected for CMIP5 or CORDEX. But, the QC may have anyway distributed computing with a shared file system.

FORCE_STND_TABLE Use always the standard table even if there is an entry in the project table.

HIDDEN_DIRECTORY_DESCENT By default, paths to data are found ignoring hidden directories.

IGNORE_BROKEN_LINKS Skip broken links in the directory tree of data, i.e. files pointing to a non-existing target. By default, a warning is issued.

IGNORE_LOCK_FILES Occurrences rising a notification of level L>2 cuase the creation of a qc_lock_<filename>.txt file, which will prevent any further checking of files of the given variable. This option ignores any blocking of following temporal subsets. Note: In case that a blocking file was found but no qc_<filename>.nc file, then the blocking holds anyway.

PROJECT_TABLE_PREFIX Standard and project tables are used to check consistency and continuity of parted atomic-data sets and experiments. Each sub-temporal file will be checked against the project table. The assignment will be prefixed to the string indicated by PT_DIR_INDEX. If the latter is not specified, then prefix is the name of the project table.
Example: PROJECT_TABLE_PREFIX=pt

PT_PATH_INDEX Purpose: automatic determination of the project table name. Usage is the same as for EXP_PATH_INDEX. Note that meta-data of each file will be checked against the project table. So, components representing a different layout explicitely, e.g. different driving-models (index 7) could have different calendars. Note that the example provides index in [], which are not part of the name. /path[10]/AFR-44[9]/SMHI[8]/CCCma-CanESM2[7]/historical[6]/r1i1p1[5]/SMHI-RCA4[4]/v1[3]/d
Example: PT_PATH_INDEX=9,8,7,4

SYNC_FILE_AMBIGUITY_CHECK By default, the QC examines the begin and end dates specified by the time records of all sub-temporal files and finds by synchronisation the right file for proceeding (or starting) a check. The purpose of this option: a) check the times in the filenames as well as in the files and search for overlapping periods in the set of sub-temporal files of a given variable. b) check the modification times of the sub-temporal files of a given variable. Issue a warning, if the modification time of the set of the sorted files according to a) is not ascending. This is disabled by assigning 'no_mod'. c) ignore filename embedded times by assigning 'mixed' (csl accepted)
Example: SYNC_FILE_AMBIGUITY_CHECK=no_mod,mixed # [disabled]

TABLE_REQUIREMENTS Specification of requirements. Given in QC-0.4/tables/SVN_defaults
Example: TABLE_REQUIREMENTS=CORDEX_archive_design_110628.csv

TABLE_STANDARD An 'officially' standardised table. Format: comma-separated-values Note: if TABLE_UPDATE_ALWAYS is enabled, then the value of TABLE_STANDARD will be overruled.
Example: TABLE_STANDARD=CORDEX_variables_requirement_table_130218.csv

TABLE_UPDATE_ALWAYS Use always up-to-date CORDEX_variables_requirement_table. If a particular version should be kept in stead of an updated one, then specify a name in a qc-config file and keep TABLE_UPDATE_ALWAYS disabled in every config/task file. If TABLE_UPDATE_ALWAYS is enabled, then TABLE_STANDARD is synchronised with the svn repository. If disabled and TABLE_STANDARD is not specified, then the file with the highest revision number is selected.

TIME_TABLE By default, the record of time is tested on the basis of regular steps and for consistency across files and experiments. However the entire set of CMIP5 experiments includes also irregularities and distinct periods. Time information was extracted from the CMIP5 standard table and is given in QC/tables/time_table.csv which provides a default. Please have a copy of your own with a different name in case of individual settings; the file 'time_table.csv' is part of svn updates. Please have a very close look at the file, because this table is certainly not free of errors. If you detect any, please give a short note. The default falls back to 'regular' when no time table file is found in QC/tables.
Example: TIME_TABLE=... [time_table.csv]

Section: Command-line used Options

CLEAR Remove former results corresponding to the paths and variables selected before re-doing the QC. However, corresponding logfiles are not affected. Assignments:

- CLEAR=only, do a clearing, but no QC will take place.
- CLEAR=lock for files qc_lock_<filename>.txt .
- CLEAR=note for qc_note_<filename>.txt.
- CLEAR='resume', then qc_lock_<filename>.txt files will be removed before processing any further.
- CLEAR=follow_links. When targets are pointed to by symbolic links, then both target and link will be removed. Note: symbolic links are removed by default, but not the targets.
- CLEAR=mark, when a file 'clear.mark' or a file <filename_root>.clear is present.

Multiple options by a comma-separated-list; the order doesn't matter. Note: CLEAR=lock,note or mark enables option FORCE_STND_TABLE for the respective variable.
Example: CLEAR=only
Example: CLEAR=lock,note

CHECK_TOOLS Check and show availability of tools and exit. Note that this is done at the beginning of each session.

DEBUG_C Log execution of qcConfigurator processing to STDERR.

DEBUG_E Log execution of qcExecutor processing to files corresponding to variable names (in the directory where qcManager was started).

DEBUG_M Log execution of the qcManager to STDERR.

DEBUG_X[=script] Enable debugging of external scripts invoked in qcManager; default: every. If the name of a script is specified, then this applied only to this one.

FLOW_TRACE Flow trace analysis of the qcManager main-loop. Note that specific commenting-in is required for qcManager and qcExecutor. e.g. : sed -i 's/#trace/trace/' qcManager

NEXT[=num] Process the next num variable-frequency occurrences.

NEXT_RECORDS Only progress by the given number of records at each check. Purpose: probably debugging
Example: NEXT_RECORDS=50

ONLY_SUMMARY Force creation of annotation files in QC_RESULTS/check_logs/_Notes. Only for testing, because this is usually performed at the end of a run.

SHOW_CALL Show the qC_main.x call; no execution.

SHOW_CONF Show configuration and exit.

SHOW_EXP Show names of paths with scheduled variable-frequency occurrences.

SHOW_NEXT[=num] Show the next num variable-frequency occurrences scheduled for processing.

REUSE_PATH_LIST Use a path-list left previously for a resumed session. Note that this never scans for any changes in the data directory tree!

Section: Processing

EMAIL_SUMMARY Send a summary of notifications and checked periods of variables to this (list of) email recipients after each finished QC session. Default: no email is sent. Note: this is different from EMAIL_TO.
Example: EMAIL_SUMMARY=name@site.dom [disabled]

EMAIL_TO In case of error: send e-mails to this comma-separated-list.
Example: EMAIL_TO=name@site.dom #[disabled]

ENABLE_DISK_SPACE_INQ If the QC stores results in a files system with occasional 'no space left on device', then you could enable a disk space inquiry before write operations are started for each variable.

EXP_PATH_INDEX The purpose is just to tag a larger set of various checks with a name that contains a certain volume of netCDF files. This is done by specifying a comma-separated sequence of directory path components to be connected to a underscore separated string, i.e. EXP_NAME. The index of the components is counted from the end of the path. Note that the example provides index in [], which are not part of the name. For instance, the sequence 9,6,8,7 applied to the path /path[10]/AFR-44[9]/SMHI[8]/CCCma-CanESM2[7]/historical[6]/r1i1p1[5]/SMHI-RCA4[4]/v1[3]/day[2]/clh[1] results in AFR-44_historical_SMHI_CCCma-CanESM2, by default EXP_NAME=unkownExp. Note: this description is also valid for PT_PATH_INDEX.
Example: EXP_PATH_INDEX=9,8,7,6,5

GROUP_NAME Set group name. This is necessary, if the QC is also operated by users who are not in the default group of the user who initially checked out from the svn repository. This will

automatically also set the SGID-bit, i.e. grant permissions to all group members. The current setting of user permissions of each file is duplicated to the respective group permissions. Note: this can only be done by the owner of the QC directory.
Example: GROUP_NAME= # [none]

HARD_SLEEP_PERIOD In order to enable trapping signals, long sleeping period are subdivided into smaller intervals of consecutive sleep commands.
Example: HARD_SLEEP_PERIOD=10 # [10]s

IGNORE_TEMP_FILES If there are (temporary) files failing the right filename syntax required for temporal sub-files AND it is obvious that the syntax rules are always followed, then ignore these in the syncFile process. Exception: files below the 'fx' directory in the data directory tree.

INQ_DISK_SPACE Available disk space is checked before a file is processed. If you know that there will be always space enough, then enable this option. NOTE: If the compiler complains about structure 'statvfs64', please try the macro -DSTATVFS parameter for compilation of the program 'diskUsage.c'.

NEVER_BREAK_SESSION Notifications of level 4 as well as segmentation fault will terminate a running session immediately. This behaviour may be cancelled. Note that Ctrl C, i.e. TERM, terminates nevertheless always.

NICE Apply linux 'nice' command to executables on guest machines (see 'man nice').
Example: NICE=15 # [0]

NIGHT_SHIFT Suspend processes between 8 - 19 o'clock on guest hosts and allow only a single process on the qcManager host.

NUM_EXEC_THREADS List of number(s) of simultaneous execution processes per host. Note: there is always a single qcManager process. Special: If a single number is specified, this is assigned to all hosts in the QC_EXEC_HOSTS list. Fine tuning: each positional number in the list corresponds to a position in the QC_EXEC_HOSTS list. If less positions are given than in QC_EXEC_HOSTS,

then the last position is assigned to the omitted positions.
Example: NUM_EXEC_THREADS=2 #1,0 # [1]

QC_CONF More configuration options in another file. These options have a lower priority than those given in the current file. Note that there is no nesting depths of additional configuration files.
Example: QC_CONF=another-configuration-file

QC_EXEC_HOSTS The list of machines executing jobs for this session. Usually disabled causing the current computer to be the only one.
Example: QC_EXEC_HOSTS=comma-separated-list-of-hosts # [HOSTNAME of qcManager]

REATTEMPT_LIMIT If a process cannot be run, because the data base is not available or a server is absent for instance, then retry after a sleep for the specified number.
Example: REATTEMPT_LIMIT=5 # [5]

WORK_AT_LOW_LOAD Sleep period if waiting for a server or anything else. Note: particular ordinary processes have a sleep time of their own to distinguish them in the output of 'ps'. Start the analysis on a guest machine, if the average load from the uptime command is < 1.5 for 5 minutes and < 1 for 15 min. This option is not available at present.

Section: Miscellaneous

ARITHMETIC_MEAN By default, calculation of cell averages over a geographical grid is performed by weighting with the corresponding grid-cell area. In particular for ocean basin cross-section, this is very time consuming due to the irregular shape of the basins taken into account by varying filling values (also for regular grids with varying effective areas, e.g. T of snow). If the precise value of the average is less important, this option reduces calculation time.

CASE_INSENSITIVE_VARIABLE_NAME Considers variable name as case-insensitive; sensitive by default.

CHECK_MODE The QC checks meta data, time data and data (of variables), where for the latter two the records are inspected. Modes can be enabled by key-words: meta, time, data. Note: data includes meta and time; time includes meta.
Example: CHECK_MODE=time #[data]

CHECKSUM Calculate checksum of every file that passes successfully the QC. If md5 or sha1 are assigned, then a separate text file is generated where the checksum is stored with appended extension md5 or sha1 respectively, to the filename, i.e. filename.nc.md5. Else, a named script/program is executed. Such a script/program must be available in the user's search path for commands or the path has to be absolute. The function is not part of the QC package. The calling qcExecutor script provides the following parameters:

- filename
- path to the root of the CMIP5 directory tree
- path to the data file in the CMIP5 directory tree
- path to the session-log file
- name of the experiment as defined by EXP_PATH_INDEX

If the script/program outputs two elements to standard output (suppose: filename and a kind-of-checksum), then these are written to a table in the project path named according to the experiment-log-name prefixed by 'cs_'. Select the checksum algorithm (md5 | sha1 | function/program)
Example: CHECKSUM=checkSum # [disabled]

CHECKSUM_DIR An additional path to a redundant storage of checksum (albeit differently). By default in QC_RESULTS/checksums.
Example: CHECKSUM_DIR=CS_CORDEX

CHECKSUM_FILE No checksum-file will be created by default containing the checksum of a corresponding file.

DISABLE_INF_NAN Disable testing for Inf or Nan. Note that this may also be disabled by setting D[iscard] in the check-list file for flag R6400.

EXCLUDE_VARIABLE Exclude specific variables globally from a check.
Example: EXCLUDE_VARIABLE=average_T1,average_T2,average_DT

IGNORE_REF_DATE_ACROSS_EXP By default, the reference date of variable 'time' written to the project table is checked for each file, even across experiments. There is nothing wrong with different reference dates in principle thus throwing warnings can be disabled. Please, notice the difference to the option IGNORE_REFERENCE_DATE, which disables any check for differing reference dates, even across files.

IGNORE_REFERENCE_DATE This option disables any check for differing reference dates, even across files (includes option IGNORE_REF_DATE_ACROSS_EXP).

NON_REGULAR_TIME_STEP This disables any check related to a constant time step within a file, across files, or across experiments. In fact, the only property of time to be checked is that time marches on and that time bounds don't overlap.

NOTE User-defined handling of exceptions, i.e. how to proceed with notification of QC check results. Explanations in file QC-0.4/Project_tables/CORDEX_check-list.conf. Keys: Flag, L[1-4], D, PT, ST, variable names. Please note that any misspelling is considered the name of a variable and would throw a flag in most cases.
Example: NOTE=o2,PT,58_2,L1

NOTE_ALWAYS Apply a unique annotation handling to all entries of the check-list.
Example: NOTE_ALWAYS=L1,EM

NOTE_LEVEL_LIMIT Limit maximum annotation level to all entries of the check-list.
Example: NOTE_LEVEL_LIMIT=L1

OUTLIER_TEST Min/max of data are tested whether they could be out of scope of reasonable value ranges. Peaks/breaks are considered such that a constant number of data points fulfils M-times the condition abs(value - average) > stdDev + n * stdDev / 4 with increment n= 1,2,3,... Average, standard deviation, and range for the statistics of global minimum and maximum, respectively. But, if more than P % of all the data points were found to be outside, then these are considered regular. In that case, outliers are still detected when exceeding orders of magnitude L. Alternatively a fixed number N, which must not be exceeded, may be provided. Both N and P are accepted and taken into account simultaneously. Default setting: M=5/P=1/L=5, disabled N); M,P, and/or L may be assigned to the OUTLIER_TEST option. Keyword ignore suppresses email notification as usual. Comma-sep-list of variables may be specified. Groups of parameters may be indicated by braces, which may be nested. Rules for nested braces:

- The outmost level does not require braces.
- The parameters within matching braces are not applied to the variables in the same level.
- When there are no variables specified for the current level, then the parameters are applied to all groups nested in the current level.

Thus, if you want to establish a new default for all variables, place the parameters in the outmost level without any variable name. Note: spaces will be removed internally. Example: var0, ignore, var1, p0, discard, var2, p1,p2,p3. Test for var2 will be discarded (thus, p1 has no effect). Var1 gets p0 and from the outer level 'ignore' and p2. Var0 gets p3. Multiple lines are accepted, even with unclosed braces. All lines will be concatenated in the order of appearance. Matching pairs of ’’ and ’’ are checked. Compatibility: ignorevar_1,... ,var_n still accepted (same for discard). Key-word 'POST' will postpone tests to be done after the regular QC completed. Enabling FORCE_POST will redo the test anyway.
Example: OUTLIER_TEST=POST,N10,P0.01,M3,L3

PROJECT The name of the project. Supported at present: CMIP5, LUCID, CORDEX Disabling suppresses related meta-data checks.
Example: PROJECT=CORDEX # [CMIP5]

QC_VERSION Write the svn version-numbers of the current working directory to the current session log-file. Location: QC_RESULTS/session_logs Note: if there is no svn, then names and modification dates of each file are written.

REPLICATED_RECORDS Find occurrence of multiple identical data records. A checksum of all data is calculated for each record and compared to each other. Special assignments (without ”):

- no_0-D: ignore for 0-D variables.
- only_groups: omit isolated replicated records, report only groups. The size of the group may be specified by appending '=num'; by default num=2.
- clear_bitsN: clear the N least significant bits of data values with default of N=2.

Example: REPLICATED_RECORDS=pr,tas
Example: REPLICATED_RECORDS=no_0-D,only_groups_12,clear_bits
Example: REPLICATED_RECORDS=no_0-D,only_groups=12

TIME_LIMIT Stop the QC at a specific time value. Use ISO-8601 format (yyyy-mm-ddThh:mm:ss). Truncations are accepted (1955-06 would process the records before 1955-06-01 00:00:00). If the parameter gets a leading 't' or 'T' character, then a time value (a pure number) is accepted that must correspond to the reference date in the data file.
Example: TIME_LIMIT=1955-08

TRACKING_ID_ONLY Scan through the list of SELECTED netCDF files for the tracking_id and create an appropriate tracking_id files. Nothing else.

Section: Frequency Distribution

FREQ_DIST Calculate the frequency distribution of the entire domain Note: generally, you will need post-processing with 'hist.x' to assemble the final frequency distribution, except you select specific debug options. Rules for selecting class widths and starting point (e.g. centering): Firstly, a fd-build-file of the same variable at destination has precedence. Second choice is a fd-build-file or fd-prop-file (the header of a build-file) in a specified path. Thirdly, explicit statement of properties. Default: automatic determination of fd properties.

FD_TIME_PART Partitioning the total time span of an experiment into smaller ranges. FreqDists will be calculated separately for each window. Note: the residual time interval may be shorter. Note: if no unit-designator is appended (y, mo, d, h, mi, s) the unit from the time variable is used.
Example: FD_TIME_PART=5y # 30y [default: the entire experiment]

FD_PROPERTY_PATH By default, frequency distributions adjust automatically the bin (class) width while processing the first time window (or the total time span), which is then used in subsequent time windows for the entire experiment. Specifying this path enables to retain the properties of the same variable (if found) of another experiment, e.g. control run. Such a file must have extension .build or .prop
Example: FD_PROPERTY_PATH= # [no path]

FD_EXPLICIT_PROPS Explicit properties (two quantities separated by '/'): class-width (number) / init value (number), e.g. 1/0 will centre a bin (class) of the freq.dist around zero. Note: The alignment of the fd-bin borders is done automatically. It is not possible, do have a class boundary exactly at say 273.15 with class-width of 1. This option is only reasonable for selecting a particular common type of variable, e.g. temperature, in a separate run. See priority rules.
Example: FD_EXPLICIT_PROPS=1.E-05/3.5E-05

FD_PLAIN Output of the Freq Dist as ready-to-use. By default, the output will be in a format to be re-read later in order to resume a previous session (the output file gets extension '.build'). File extension in plain mode: '.hist'. Attention: this starts and completes a FD calculation. Makes little sense for multi-sessions. Purpose: debugging

FD_BARS Like FREQ_DIST_PLAIN, but output of the Freq Dist shows complete shapes of bars. Purpose: debugging

Section: HLEV

HLEVD The Hans-Luthardt-Extreme-Value-Detection (see code in qcExecutor). The HL extreme value is expressed as deviation range normalised to the units of standard deviation: (max-min)/stdDev, i.e. is dimensionless for each grid-cell. If a threshold is exceeded, then by default, a notification is written to the SESSION_LOG. Makes not really sense for applying to 3D. Derived by and therefore requires CDO.

HLEVD_DELETE_FILE Delete netCDF file with HLEV in any case; only notification takes place.

HLEVD_DETREND To detrend a time series takes some time. If a trend of a variable does not boost exponentially detrending should not affect this kind of extreme value detection (which is anyway rather vague).

HLEVD_KEEP_THRESHOLD_FILE Keep netCDF file with HLEV if threshold was exceeded.

HLEVD_THRESHOLD Threshold: (max-min)/stdDev > threshold
Example: HLEVD_THRESHOLD=9 # [1]