Supplement of Geosci. Model Dev., 13, 5277–5310, 2020
https://doi.org/10.5194/gmd-13-5277-2020-supplement

Geoscientific Model Development
Open Access
EGU

Supplement of

Flex_extract v7.1.2 – a software package to retrieve and prepare ECMWF data for use in FLEXPART

Anne Tipka et al.

Correspondence to: Anne Tipka (anne.tipka@univie.ac.at)

Contents

S1. Execution script parameters

S1.1. `setup.sh` parameters

This shell script defines the command-line parameters for the installation process and executes the Python script `install.py` for the installation of flex_extract. The following parameters can be set:

Name:	Type:	Value range:	Default:
`TARGET`	String	local, ecgate, cca, ccb	None
Description: Defines which flex_extract application mode (location of installation) will be used. Local = local mode; egate = Remote or Gateway mode, cca/ccb = Remote or Gateway mode. Whether the local mode is for public or member state users doesn't matter during the installation process.			

Name:	Type:	Value range:	Default:
`MAKEFILE`	String	any makefile present, according to environment	None
Description: Name of the makefile in `Source/Fortran` directory to be used for compiling the Fortran program. `makefile_ecgate` has the configuration for the ecgate environment and `makefile_cray` has the configuration for the HPC environment. For local server versions there is `makefile_local_gfortran` which has to be adapted (pathes to the ecCodes library) by the user.			

Name:	Type:	Value range:	Default:
ECUID	String	-	None
Description: User id on ECMWF server.			
Name:	Type:	Value range:	Default:
ECGID	String	-	None
Description: User group id on ECMWF server.			
Name:	Type:	Value range:	Default:
DESTINATION	String	<name>@ generic <scope>	None
Description: Ectrans destination which is used to transfer files from ECMWF servers to local gateway. This has to be set up by users on the local gateway. See installation instruction for more information.			
Name:	Type:	Value range:	Default:
GATEWAY	String	IP address / name	None
Description: Name or ip address of member gateway.			
Name:	Type:	Value range:	Default:
INSTALLDIR	String	full path	`$HOME` on ECMWF server and current flex_extract root path on local server
Description: Root path where flex_extract should be installed. It will always be set to `$HOME` on ECMWF servers and on local hosts it will be set to the current flex_extract root path if not set.			
Name:	Type:	Value range:	Default:
JOB_TEMPLAT	String	installscript.template	installscript.template
Description: The rudimentary template file to create a batch job template for submission to ECMWF servers. Should not be changed since it is optimized for ECMWF server. (Remote and Gateway mode)			
Name:	Type:	Value range:	Default:
CONTROLFILE	String	any CONTROL file	CONTROL_ERA5
Description: The file with all CONTROL parameters.			

S1.2. `run.sh` parameters

This shell script defines the command-line parameters and executes the Python script `submit.py` for running flex_extract. The following parameters can be set:

Name:	Type:	Value range:	Default:
QUEUE	String	ecgate; cca; ccb	None
Description: Name of ECMWF server for submission of the job script to its batch system.			

Name:	Type:	Value range:	Default:
START_DATE	String [YYYYMMDD]	depends on data set	None
Description: The first day of the retrieval period. If END_DATE is set, START_DATE must be prior to END_DATE.			

Name:	Type:	Value range:	Default:
END_DATE	String [YYYYMMDD]	depends on data set	None
Description: The last day of the retrieval period. This value is optional and if it is not set, it will be automatically set equal to START_DATE. For a one day retrieval it has to be equal to START_DATE. If set, it has to be greater than or equal to START_DATE.			

Name:	Type:	Value range:	Default:
DATE_CHUNK	Integer	depends on resolution	3
Description: Maximum number of days retrieved within one MARS request. This number is limited due to maximum allowed memory and time limit for one MARS request. Be careful in changing this number. It can be larger for coarse domains and time intervals but may be too large for very high-resolution retrievals.			

Name:	Type:	Value range:	Default:
JOB_CHUNK	Integer	depends on resolution	None
Description: # of days to be retrieved within a single job. Can be selected to start the submission script once and to automatically split the time period into smaller job chunks. For example, it could be very useful if you want to retrieve a month with a resolution of 0.1° and a time resolution of 1 hour. Then only 1 day per job is possible.			

Name:	Type:	Value range:	Default:
BASETIME	Integer	[0;12]	None
Description: This parameter is intended for half-day retrievals. Only half a day will be retrieved starting from BASETIME going back 12 hours. E.g. 20180510 with a BASETIME = 00 would lead to a data retrieval of 20180509 12h until 20180510 00h. Can be set to 00 or 12 only.			

Name:	Type:	Value range:	Default:
LEVELIST	String [start/to/end]	1/to/[60,91,137] depends on data set	None
Description: List of vertical levels. It can be a subset of levels but it has to include the maximum level (end) which is the surface. If full list of levels is needed and parameter LEVEL is set, the LEVELIST parameter is not needed. “end” has to be the maximum number of possible levels and has to be the same as in LEVEL (if LEVEL specified).			

Name:	Type:	Value range:	Default:
AREA	Double [f/f/f/f]	any float within latitude and longitude boundaries	None
Description: Domain defined as north/west/south/east			

Name:	Type:	Value range:	Default:
INPUTDIR	String	any path	None
Description: Path to the temporary directory for the retrieved GRIB files and other processing files. The temporary directory will be created if it does not already exist.			

Name:	Type:	Value range:	Default:
OUTPUTDIR	String	any path	None
Description: Path to the final directory where the final FLEXPART ready input files are stored. The final output directory will be created if it does not already exist.			

Name:	Type:	Value range:	Default:
PPID	Integer	-	None
Description: This is the specific parent process id of a single flex_extract run to identify the files. It is the second number in the GRIB filenames. This is usually only necessary if the GRIB data were retrieved and a rerun of the post-processing has to be done. Then PPID is used to select the files.			

Name:	Type:	Value range:	Default:
JOB_TEMPLATE	String	submitscript.template	submitscript.template
Description: Template file which is used remote and gateway modes to create the korn shell job script to be submitted to the batch system on an ECMWF server.			

Name:	Type:	Value range:	Default:
CONTROLFILE	String	any CONTROL file	CONTROL_ERA5
Description: The file with all CONTROL parameters to define a flex_extract retrieval.			

Name:	Type:	Value range:	Default:
DEBUG	Integer	[0,1]	0
Description: If set to "1" all temporary output files from the MARS requests are kept and some extra information is written out to the log file. Usually all temporary files except the FLEXPART ready input files are deleted at the end of flex_extract.			

Name:	Type:	Value range:	Default:
REQUEST	Integer	[0,1,2]	2
Description: This parameter allows to write out the MARS requests in a separate file. The requests are stored in the file "mars_requests.csv". Possible selections are: 1: normal data retrieval; 2: neglect data retrieval and just writes out the MARS requests to the file; 3: retrieve data and write out the MARS requests.			

S2. `CONTROL` file parameters

S2.1. User section

Name:	Type:	Value range:	Default:
ECUID	String	-	None
Description: User id on ECMWF server.			

Name:	Type:	Value range:	Default:
ECGID	String	-	None
Description: User group id on ECMWF server.			

Name:	Type:	Value range:	Default:
DESTINATION	String	<name>@ generic <scope>	None
Description: Ectrans destination which is used to transfer files from ECMWF servers to local gateway. This has to be set up by users on the local gateway. See installation instruction for more information.			

Name:	Type:	Value range:	Default:
GATEWAY	String	IP address / name	None
Description: Name or ip address of member gateway.			

S2.2. Data section

Name:	Type:	Value range:	Default:
CLASS	String [xx]	EI, E4, EA, EP, OD	None
Description: ECMWF data classification identifier for data sets. EI: ERA-Interim; E4: ERA-40; EA: ERA5; EP: CERA-20C; OD: operational;			

Name:	Type:	Value range:	Default:
DATASET	String	CERA20C, INTERIM	None
Description: This keyword has to be defined for retrievals of public data sets ERA-Interim and CERA-20C with ECMWF Web API. Public data are stored in a different MARS database and are available for everyone after registration at ECMWF (see installation plan). Even though ERA5 is also a public data set, the retrieval works differently (with CDS API) and does not need this parameter.			

Name:	Type:	Value range:	Default:
STREAM	String [xxxx]	OPER, ENFO, ENDO	None
Description: Identifies the forecasting system used to generate the data. Most times the operational data stream OPER is appropriate. Use ENFO for ensemble forecasts and ENDO for CERA data.			

Name:	Type:	Value range:	Default:
EXPVER	Integer	-	1
Description: Experiment version number. Only necessary for R&D experiments or E-suites; otherwise use 1.			

Name:	Type:	Value range:	Default:
NUMBER	String [start/to/end] or Integer	depends on availability	“OFF”
Description: In most cases this can be set to “OFF”; for access to individual ensemble members of an ensemble forecast the individual numbers have to be selected. Note, however, that model level data are not stored in MARS for individual ensemble members except the control run. They exist only for a few days before they are discarded. For retrieving CERA20C data, a number has to be selected explicitly. Select “0” for the control run.			

Name:	Type:	Value range:	Default:
FORMAT	String	[GRIB1, GRIB2]	GRIB1

Description:
Output format (either GRIB1 or GRIB2). Use GRIB2 only when using FLEXPART versions >9.2 or FLEXPART has to be adapted for reading GRIB2. Nowadays, 3D-model level fields are automatically stored on GRIB2 since some GRIB message parameters are not able to be stored in GRIB1 anymore. Surface level fields are always retrieved in GRIB1 and by selecting GRIB2 they are converted to GRIB2 by flex_etract. Be careful with this option, especially if using the ADDPAR parameter to add other fields. Sometimes other surface fields as already defined in flex_etract can't be converted to GRIB2.

S2.3. Data field section

Name:	Type:	Value range:	Default:
TYPE	list of Strings [xx xx ... xx]	[AN, FC, CV, CF, 4V, PF]	None

Description:
A list of field types for each retrieving hour per day. E.g. "AN FC FC FC AN FC FC FC" for a day with 3-hourly retrieval (DTIME=3). At 0 and 12 UTC we retrieve analysis fields and at 3, 6, 9, 15, 18 and 21 UTC forecast fields. So far, flex_extract is using ANalysis (AN), ForeCast (FC), Control Forecast (CF), Perturbed Forecast (PF), Calibration/Validation forecast (CV) and 4D Variational analysis (4V). Other types might be also possible but were not tested. The analysis fields are usually (depending on data set) available at 00/06/12/18 UTC. For better temporal resolution, the time in-between the AN fields can be filled with forecasts (FC). Additionally, it is recommended to use analysis fields only at 00 and 12 UTC and fill the rest of the times with other field types, such as forecasts.

Name:	Type:	Value range:	Default:
TIME	list of Integer [ii ii ... ii]	[00 – 23]	None

Description:
The time of the corresponding field type (TYPE) in hours. It is important to set the correct forecast times, e. g. in the ERA-Interim data set, forecast times of 00 UTC are normally used to obtain forecast fields between +1 and +11 h; to obtain fields between 13 and 23 UTC, a forecast time of 12 UTC is used. In most cases, there are two forecasts starting per day. TIME has to have the same number of values as TYPE! The start times of forecasts can vary from one data set to another, for example CERA20C has only a single forecast per day (18 UTC).

Name:	Type:	Value range:	Default:
STEP	list of Integer [ii ii ... ii]	[00 – max available STEP in data set]	None

Description:
This is the forecast time step in hours for each corresponding field type (TYPE). Counting of the steps starts from the forecast times, e. g. in ERA-Interim, for forecasts verifying at 3, 6, 9 UTC, the STEPS 3, 6 and 9 are used with a forecast TIME 00 UTC. Has to have the same number of values as in TYPE and TIME and has to match with DTIME! For analysis (AN) fields, the STEP has to be 00 always!

Name:	Type:	Value range:	Default:
MAXSTEP	Integer	> 24	None

Description:
This parameter allows to retrieve data from forecasts longer than 24 hours. With MAXSTEP >24, the forecasts from different days overlap, the naming scheme of the output files changes from <PREFIX>YYMMDD to <PREFIX>YYMMDD.HH.SSS where HH is the hour of the start of the forecast and SSS is the forecast step in hours. Optional parameter.

S2.4. Time section

Name:	Type:	Value range:	Default:
START_DATE	String [YYYYMMDD]	depends on data set	None
Description: The first day of the retrieval period. If END_DATE is set, START_DATE must be prior to END_DATE.			

Name:	Type:	Value range:	Default:
END_DATE	String [YYYYMMDD]	depends on data set	None
Description: The last day of the retrieval period. This value is optional and if it is not set, it will be automatically set equal to START_DATE. For a one day retrieval it has to be equal to START_DATE. If set, it has to be greater than or equal to START_DATE.			

Name:	Type:	Value range:	Default:
DTIME	Integer	1,3,6	None
Description: Time step of retrieved data. Detects TYPE, TIME, STEP, ACCTYPE, ACCTIME according to DTIME. Therefore CONTROL file can have more values than needed. Available resolution in time depends on availability in the data set. Coarser resolution can always be selected.			

Name:	Type:	Value range:	Default:
DATE_CHUNK	Integer	depends on resolution	3
Description: Maximum number of days retrieved within one MARS request. This number is limited due to maximum allowed memory and time limit for one MARS request. Be careful in changing this number. It can be larger for coarse domains and time intervals but may be too large for very high resolution retrievals.			

Name:	Type:	Value range:	Default:
BASETIME	Integer	[0;12]	None
Description: This parameter is intended for half-day retrievals. Only half a day will be retrieved starting from BASETIME going back 12 hours. E.g. 20180510 with a BASETIME = 00 would lead to a data retrieval of 20180509 12h until 20180510 00h. Can be set to 00 or 12 only.			

S2.5. General section

Name:	Type:	Value range:	Default:
DEBUG	Integer	[0,1]	0
Description: If set to "1" all temporary output files from the MARS requests are kept and some extra information is written out to the log file. Usually all temporary files except the FLEXPART ready output files are deleted at the end of flex_extract .			

Name:	Type:	Value range:	Default:
REQUEST	Integer	[0,1,2]	2
Description: This parameter allows to write out the MARS requests in a separate file. The requests are stored in the file "mars_requests.csv". Possible selections are: 1: normal data retrieval; 2: neglect data retrieval and just writes out the MARS requests; 3: retrieve data and write out the MARS requests.			

Name:	Type:	Value range:	Default:
PUBLIC	Integer	[0,1]	0
Description: This specifies the selection of the kind of ECMWF Web Api access and therefore the kind of available data sets. Public data sets (1) and Member-state data sets (0). Selecting the public access method, the DATASET parameter has explicitly to be set to select the data set. (CLASS is not enough.) ATTENTION: For public data sets, users have to accept the licence of the data set to be retrieved. See here for available data sets and their licences: https://software.ecmwf.int/wiki/display/WEBAPI/Available+ECMWF+Public+Datasets			
Name:	Type:	Value range:	Default:
OPER	Integer	[0,1]	0
Description: Switch to prepare the operational job script. START_DATE, END_DATE and BASETIME will be prepared with environment variables at ECMWF server. This is only necessary if extraction of half-day retrievals should be done automatically. Specific extra feature which is usually not used by normal flex_extract users.			
Name:	Type:	Value range:	Default:
ECSTORAGE	Integer	[0,1]	0
Description: Switch to store FLEXPART ready input files in the ECFS file system. Mind the data limit.			
Name:	Type:	Value range:	Default:
ECTRANS	Integer	[0,1]	0
Description: Switch to transfer FLEXPART ready input files to the gateway server. The gateway has to be up and running. A destination has to be configured. See installation instructions for more information.			
Name:	Type:	Value range:	Default:
PREFIX	String	anything	EN
Description: Prefix of FLEXPART ready input files. The files usually have the format <PREFIX>YYMMDDHH.			
Name:	Type:	Value range:	Default:
ECFSDIR	String	any path available	ectmp: /$USER/econdemand/
Description: The destination directory on ECFS file system if the retrieved data should be stored on ECMWF servers. This is only used if parameter ECSTORAGE is set to "1".			
Name:	Type:	Value range:	Default:
MAILOPS	list of String [m1, m2, ...]	any number of mail addresses, separated by comma	["$USER"]
Description: Email list for operational log files on ECMWF servers. The email addresses should be separated by a comma. For the ECMWF server it is enough to give $USER as input. On local system an actual email is preferred to operate correctly.			
Name:	Type:	Value range:	Default:
MAILFAIL	list of String [m1, m2, ...]	any number of mail addresses, separated by comma	["$USER"]
Description: Email list for operational log files on ECMWF servers. The email addresses should be separated by a comma. For the ECMWF server it is enough to give $USER as input. On local system an actual email is preferred to operate correctly.			

S2.6. Flux data section

Name:	Type:	Value range:	Default:
ACCTYPE	String [xx]	[FC, CV, CF, 4V, PF]	None
Description: The type of field for accumulated data retrieval. The accumulated data fields are only available from forecast fields. Therefore it is separated from the normal TYPE parameter to allow, for example, retrieval solely of hourly analysis fields in the Era5 data set. For downward compatibility to older versions: if ACCTYPE is not specified, the default value is taken form the second position of the original TYPE parameter (TYPE[2]). NOTE: This is important at the moment the original TYPE parameter is changed from FC to AN for example.			
Name:	Type:	Value range:	Default:
ACCTIME	String	E.g.: Operational, Era-Interim: 00/12 CERA: 18 ERA5: 06/18	None
Description: Forecast start times of accumulated fields (fluxes). The starting times of forecast fields varies between different data sets. For downward compatibility to older versions: if ACCTIME is not specified, the default value is 00/12 for ERA-Interim and operational data, 06/18 for ERA5 data and 18 for CERA-20C data.			
Name:	Type:	Value range:	Default:
ACCMAXSTEP	Integer	12, 24 or larger	None
Description: This parameter specifies the maximum step in hours for a specific accumulated forecast start time. For daily retrievals with one forecast time the step shouldn´t be greater than 24h. For two forecast times the ACCMAXSTEP should be 12. If the parameter MAXSTEP is specified to retrieve forecasts longer than 24 hours, this parameter must have the same value. For downward compatibility to older versions: if ACCMAXSTEP is not specified, the default value is set to 12 for ERA5, Era-Interim and operational data or 24 for CERA-20C data, according to one or two forecast times of the data set.			
Name:	Type:	Value range:	Default:
RRINT	Integer	[0,1]	None
Description: Switch to select method of disaggregation for precipitation fields. Old method (0) with a simple linear disaggregation or new method (1) with 2 additional subsequent intervals per time interval. For more information see article Hittmeir et al. (2018).			

S2.7. Domain section

Name:	Type:	Value range:	Default:
UPPER	String	-90+GRID to 90	None
Description: Latitude of upper right corner of grid area.			
Name:	Type:	Value range:	Default:
LOWER	String	-90+GRID to 90	None
Description: Latitude of lower left corner of grid area.			

Name:	Type:	Value range:	Default:
LEFT	String	-180+GRID to 180	None
Description: Longitude of lower left corner of grid area. For cyclic (global) grids, use e.g. LEFT = - 180 + GRID, RIGHT = 180. For noncyclic grids crossing the dateline (180W), RIGHT may be smaller than LEFT.			

Name:	Type:	Value range:	Default:
RIGHT	String	-180+GRID to 180	None
Description: Longitude of upper right corner of grid area. For cyclic (global) grids, use e.g. LEFT = - 180 + GRID, RIGHT = 180. For noncyclic grids crossing the dateline (180W), RIGHT may be smaller than LEFT.			

Name:	Type:	Value range:	Default:
LEVEL	Integer	60, 91, 137 depends on data set	None
Description: Maximum number of vertical levels. ERA-Interim has 60 levels; ERA5 has 137 levels; CERA-20C has 91 levels; Operational data can have different number of model levels depending on the date. Check upfront in the MARS catalogue. If LEVELIST is set , this parameter is not needed.			

Name:	Type:	Value range:	Default:
LEVELIST	String [start/to/end]	1/to/[60,91,137] depends on data set	None
Description: List of vertical levels. It can be a subset of levels but it has to include the maximum level (end) which is the surface. If full list of levels is needed and parameter LEVEL is set, the LEVELIST parameter is not needed. “end” has to be the maximum number of possible levels and has to be the same as in LEVEL (if LEVEL specified).			

Name:	Type:	Value range:	Default:
GRID	Integer [i/i]	0.1° - appropriate value e.g. 2°	None
Description: Horizontal resolution of Latitude/Longitude grid. Best possible resolution varies for different data sets. E.g in operational data it´s 0.1° whereas in Era-Interim it is 0.75°. It can be specified in tenth degrees (1°) or thousandth degrees (1000 for 1°).			

Name:	Type:	Value range:	Default:
RESOL	String	depends on GRID	None
Description: Horizontal resolution of spectral fields. Specifies the desired triangular truncation of retrieved data, before carrying out any other selected post-processing.			

Name:	Type:	Value range:	Default:
SMOOTH	Integer	approriate number	0
Description: Spectral truncation of ETADOT after calculation on Gaussian grid. For more information see Sardeshmukh and Hoskins (1984).			

S2.8. Vertical wind section

Name:	Type:	Value range:	Default:
GAUSS	Integer	[0,1]	0
Description: A switch to calculate ETADOT from Lat/Lon grid (0) or from Gaussian grid (1).			

Name:	Type:	Value range:	Default:
ACCURACY	Integer	-	24
Description: Specifies the number of bits per value to be used in the generated GRIB coded fields.			

Name:	Type:	Value range:	Default:
OMEGA	Integer	[0,1]	0
Description: Retrieve Omega from MARS and put it to file OMEGAyymmddhh. Only useful for debugging reasons.			

Name:	Type:	Value range:	Default:
OMEGADIFF	Integer	[0,1]	0
Description: Calculate Omega and Dps/Dt from continuity equation for diagnostic purposes and include it in file OMEGAyymmddhh. Only useful for debugging reasons.			

Name:	Type:	Value range:	Default:
ETA	Integer	[0,1]	0
Description: Switch to read ETADOT precalculated by ECMWF and multiply it with DPDETA to be compatible with ETADOT calculated from continuity equation. ETADOT calculation from continuity equation on either Gaussian or lat/lon grid is disabled unless ETADIFF is set to 1 as well. ETADOT is available in ERA5, CERA-20C and operational data sets. Precalculated ETADOT is operationally available from September 2008 onwards. However, it is not available in the ERA-40 and ERA-Interim data set. If ETA is selected in the last two data sets, flex_extract fails.			

Name:	Type:	Value range:	Default:
ETADIFF	Integer	[0,1]	0
Description: Switch to calculate etadot and Dps/Dt from continuity equation for diagnostic purposes and include it in file `ETAyymmddhh`. Expensive option, only for debugging purposes.			

Name:	Type:	Value range:	Default:
DPDETA	Integer	[0,1]	1
Description: Switch to multiply etadot with dpdeta – this is the default. In some future version this may change.			

Name:	Type:	Value range:	Default:
ETAPAR	Integer	77	77
Description: GRIB parameter number for ETADOT/DPDETA.			

S2.9. Additional data section

Name:	Type:	Value range:	Default:
CWC	Integer	[0,1]	0
Description: Switch to retrieve cloud water content (sum of cloud liquid water and cloud ice) (1) or not (0).			

Name:	Type:	Value range:	Default:
ADDPAR	String [p1/p2/../pn]		None
Description: Additional optional surface parameters (2D fields, non-accumulated) Mostly: 27/28/173/186/187/188/235/139/39. Parameters can be specified as the Integer IDs or with the short names.			

Name:	Type:	Value range:	Default:
DOUBLEELDA	Integer	[0,1]	0
Description: Switch to select the calculation of extra ensemble members for the ELDA stream. It doubles the number of retrieved ensemble members. Each ensemble member is used to create a new synthesized ensemble member by subtracting 2 * (current time step value - last time step value) from the current time step value.			

S3. Quality assurance

This section adds additional information to the main manuscript by adding details about the single tests and the code metrics. The corresponding test data is large and therefore not distributed with the release tarball but rather as an supplement. Download links can be found on the community website.

S3.1. Unit tests

We used the `pytest` package which is a part of standard Python as well as the `mock` package which simulates external dependencies or results for the tests solely. This gives the opportunity to test the good and bad paths in a function and usually a function holds as many unit tests as there are different branches. It is a matter of defining all possible results depending on the input states and verify the expected results. The first set of unit tests were applied for functions from the `install` and `tools` modules as well as for the `UIOFiles` and `EcFlexpart` class. The details for each test are not described here; their functionality is obvious from the code.

S3.2. Regression testing for MARS requests

The release comes with a predefined set of `CONTROL` files explicitly for this test as well as with a set of MARS request reference files from the previous version 7.0.4 and version 7.1. The test can compare any number of MARS request files emerging from a set of `CONTROL` files. However, one has to make sure that the reference version contains the request files from the same `CONTROL` files. Results are saved in log files. Instructions on how the test can be conducted are given in a `README.md` file. The comparison between version 7.0.4 and 7.1 only showed expected differences related to a bug fix in the determination of the time period.

S3.3. Regression testing for GRIB files

The current release version 7.1 includes a minimal set of reference data from version 7.0.4 and 7.1, one for each type of data set. There will be more test data in the future which can then be downloaded from the community website to limit the size of the distributed release tarball. The corresponding reference `CONTROL` files are also distributed with the tarball to enable the retrieval of the data with the new version. This has to be done manually followed by placing the resulting GRIB files in a specific path as described in the *README.md* file.

S3.4. Functionality and performance tests for the Fortran code

The code package contains a set of reference outputs, and scripts to create the reference output and to run the actual regression tests. It checks for bitwise identity of the output files (data files and `standard output` written to a log file). A quantitative comparison of the resulting $\dot{\eta}_p$ which would be useful for modifications that affect the results is not yet implemented. The scripts run each test with both the fast and the debug version of the executable. The script for creating the reference also ensures that both yield identical results. In addition, the runtimes are saved to a `csv` file.

S3.5. Generic test using predefined `CONTROL` files

This has been verified for version 7.1 by manually executing the software with all these files, and inspecting the results produced. Note that public users can only use files ending with `.public` except for ERA5; they were tested in local public mode. All other cases were tested both in local and in gateway mode. Since the remote mode does not differ much from the gateway mode, only a subset of files were also tested in this mode. Results were evaluated by inspecting the log files for "success" messages and, where possible, with the regression test for GRIB file comparison (Sect. 3.3). Regarding new features, the files were inspected manually for the expected result.

S3.6. Code Metrics

The *cyclomatic complexity* (CC) is calculated as:

$$\mathrm{CC} = E - N + 2P, \tag{1}$$

where E is the number of edges (or also called links) of the graph, N is the number of nodes of the graph and P is the number of connected components (which are sub-graphs from functions independent of the super-graph) (Lacchia, 2019; Beizer, 1990; Sneed et al., 2010). The nodes represent the conditional branch instructions and program junctions, and edges are the segments between such points. This metric was calculated with the `radon` package (Lacchia, 2019) and provides the CC rank for each function, class and class method. Table 1 shows a summary of all code blocks from version 7.0.4 and 7.1. while Table 2 lists the code blocks which are worse-rated.

The maintainability index (MI), ranking between 0 and 100, is a function of `SLOC`, `CC` and the Halstead volume (`V`) (Lacchia, 2019):

$$\mathrm{MI} = \max\left[0, \frac{100}{171}\left(171 - 5.2\ln\mathrm{V} - 0.23\,\mathrm{CC} - 16.2\ln\mathrm{SLOC} + 50\sin\sqrt{2.4\mathrm{C}}\right)\right] \tag{2}$$

Table 1. Number of code blocks (classes, methods, functions) with a specific rank of cyclomatic complexity and the percentage of the total blocks for version 7.1 (116 in total) and 7.0.4 (45 in total). Determined with the Python package `radon` (Lacchia, 2019).

Rank	Version 7.0.4		Version 7.1	
A	21	44.6 %	76	65.52 %
B	10	22.2 %	28	24.14 %
C	3	6.6 %	9	7.76 %
D	4	8.8 %	1	0.86 %
E	3	6.6 %	1	0.86 %
F	4	8.8 %	1	0.86 %

where C is the fraction of comment lines (converted to radians) (Lacchia, 2019). The Halstead volume is defined as

$$\mathrm{V} = (N_1 + N_2)\log_2(\eta_1 + \eta_2) \tag{3}$$

with η_1 being the number of distinct operators, η_2 being the number of distinct operands, N_1 the total number of operators and N_2 the total number of operands.

The index applies for a complete Python file and Table 3 shows the ranks of version 7.0.4 and 7.1 respectively.

Table 2. Python code blocks and their cyclomatic complexity (CC) which ranks between C and F and their corresponding CC score. The block types are classes (C), class methods (M) and functions (F).

(a) Version 7.0.4

Block	Block type	Rank	CC score
Class methods			
GribTools.setkeys	M	C	11
MARSretrieval.dataRetrieve	M	C	15
Control	C	D	23
EIFlexpart.process_output	M	D	26
EIFlexpart	C	E	31
EIFlexpart.deacc_fluxes	M	E	34
EIFlexpart.retrieve	M	F	43
EIFlexpart.__init__	M	F	49
Control.__init__	M	F	56
EIFlexpart.create	M	F	57
Module functions			
install_args_and_control	F	C	12
getMARSdata	F	D	25
install_via_gateway	F	D	30
interpret_args_and_control	F	E	34

(b) Version 7.1

Block	Block type	Rank	CC score
Class methods			
EcFlexpart	C	C	13
EcFlexpart._create_params	M	C	13
EcFlexpart._prep_new_rrint	M	C	14
EcFlexpart._create_field_types	M	C	15
MarsRetrieval.data_retrieve	M	C	16
ControlFile._read_controlfile	M	C	17
EcFlexpart.retrieve	M	D	25
EcFlexpart.create	M	E	36
EcFlexpart.deacc_fluxes	M	F	57
Module functions			
install.py::check_install_conditions	F	C	11
install.py::mk_tarball	F	C	17
disaggregation.py::IA3	F	C	18

Table 3. Maintainability index in increasing order for the Python files of both versions. This was determined with the Python package `radon` (Lacchia, 2019).

Version 7.0.4

File	Rank	MI score
`Classes/EcFlexpart.py`	B	10.79
`Classes/MarsRetrieval.py`	A	26.92
`Mods/checks.py`	A	26.15
`Classes/MarsRetrieval.py`	A	26.92
`Mods/disaggregation.py`	A	28.55
`Mods/profiling.py`	A	38.10
`Mods/tools.py`	A	38.32
`Mods/get_mars_data.py`	A	44.77
`install.py`	A	47.33
`Mods/prepare_flexpart.py`	A	47.47
`Classes/GribUtil.py`	A	57.07
`submit.py`	A	58.90
`_config.py`	A	77.35
`Classes/UioFiles.py`	A	100.00

Version 7.1

File	Rank	MI score
`FlexpartTools.py`	C	0.00
`opposite.py`	A	45.25
`install.py`	A	48.96
`getMARSdata.py`	A	56.28
`GribTools.py`	A	59.10
`submit.py`	A	67.72
`prepareFLEXPART.py`	A	71.40
`UIOTools.py`	A	85.18

References

Beizer, B.: Software Testing Techniques, Van Nostrand Reinhold, 2nd edn., 1990.

Hittmeir, S., Philipp, A., and Seibert, P.: A conservative reconstruction scheme for the interpolation of extensive quantities in the Lagrangian particle dispersion model FLEXPART, Geoscientific Model Development, 11, 2503–2523, https://doi.org/10.5194/gmd-11-2503-2018, https://www.geosci-model-dev.net/11/2503/2018/, 2018.

Lacchia, M.: radon 4.0.0 - Project description, https://pypi.org/project/radon/, 2019.

Sardeshmukh, P. D. and Hoskins, B. I.: Spatial Smoothing on the Sphere, Monthly Weather Review, 112, 2524–2529, https://doi.org/10.1175/1520-0493(1984)112<2524:SSOTS>2.0.CO;2, https://doi.org/10.1175/1520-0493(1984)112<2524:SSOTS>2.0.CO;2, 1984.

Sneed, H. M., Seidl, R., and Baumgartner, M.: Software in Zahlen, Hanser, 2010.

Spillner, Andreas; Linz, T.: Basiswissen Softwaretest, dpunkt, 5th edn., 2012.

Thénault, S.: Pylint, https://www.pylint.org/, 2001.

van Rossum, G., Warsaw, B., and Coghlan, N.: PEP 8 – Style Guide for Python Code, Online, https://www.python.org/dev/peps/pep-0008/, 2001.

Wolff, E.: Continuous Delivery. Der pragmatische Einstieg, Dpunkt Verlag, Heidelberg, 282 pp., 2014.