

---

## **TBD2 Software User Tutorial**

---

This document contains the flex\_extract v7.0.3 software user tutorial for the on-demand configuration.

### Document history

Version	Date	Author	Description
0.1	14 July 2003	Alexander Beck	Initial draft
2.0	27 January 2007	Leopold Haimberger	Applies to ECMWFDATA v2.0
3.0	24 March 2009	Leopold Haimberger	Applies to ECMWFDATA v3.0
4.0	30 November 2010	Leopold Haimberger	Applies to ECMWFDATA v4.0
5.0	28 June 2012	Leopold Haimberger	Applies to ECMWFDATA v5.0
6.0	20 December 2013	Leopold Haimberger	Applies to ECMWFDATA v6.0
7.0	1 December 2015	Leopold Haimberger Anne Philipp	Applies to ECMWFDATA v7.0
7.0.3	13 June 2018	Leopold Haimberger Anne Philipp	Applies to flex_extract_v7.0.3

## Contents

1. Scope.....	4
1.1. Identification.....	4
1.2. System overview.....	4
1.3. Document overview.....	4
2. Top-level scripts: Submit.py.....	4
2.1. Specifications within the script.....	5
2.1.1. Hard-coded specifications.....	7
2.1.2. Compilation of Fortran90 programmes.....	7
2.2. Retrieval of model data from MARS archive.....	7
2.3. De-accumulation of flux data.....	7
2.4. Calculation of vertical velocity and preparation of EN-files.....	8
2.4.1. Calculation of vertical velocity from horizontal wind using the continuity equation 8	
2.4.2. Calculation of vertical velocity from precalculated MARS parameter 77.....	9
2.5. Checking of EN-files and communication to IDC.....	9
3. job preparation and submission.....	10
3.1. Job preparation and submission on remote host.....	10
3.2. Job submission on local host.....	10
4. Example.....	11
4.1. Data retrieval from the operational archive.....	11
5. monitoring and ERROR handling.....	11
Glossary.....	12
Abbreviations.....	12
References.....	13

## 1. SCOPE

### 1.1. Identification

This document applies to the flex\_extract version 7.0.3

### 1.2. System overview

Flex\_extract v7.0.3 is a software to supply the IDC ATM system with input data describing the state of the atmosphere and thus the current transport conditions. The system retrieves the data from the European Centre for Medium-Range Weather Forecasts (ECMWF). The system complements the NCEPDATA software already applied at IDC as another source for input data for ATM modelling. Flex\_extract v7.0.3 is run in two distinct configurations denoted “operational” and “on-demand”. This document describes the functionality of the on-demand configuration. For the operational configuration an individual tutorial is provided.

### 1.3. Document overview

This document defines the flex\_extract version 7.0.3 software user tutorial for the **on-demand configuration**.

This document is intended for users of the on-demand configuration of flex\_extract to be run remotely at the ECMWF computer facilities via the EAccess software or WebMARS. Thus the user is assumed to have access to ECMWF computer system.

Flex\_extract version 7.0.3 is a complete rewrite of previous versions in the python programming language. Compared to flex\_extract version 6.0 the functionality is expanded towards retrieval of long forecast runs and the possibility to use WebMARS.

All retrieved fields can be converted to GRIB2 if needed and also to a new binary .fp format containing data interpolated to height levels. FLEXPART can run much faster if input data are provided in this format.

For description of the operational configuration please consult the SUT\_operational document. The installation process is described in the Software Installation Protocol.

This document is compliant with the IDC Software Documentation Framework (2002) and the CTBTO Editorial Manual (2002).

## 2. FLEX\_EXTRACT DIRECTORY STRUCTURE

It is always helpful to see in which directory structure the scripts and auxiliary files are stored and what the working directories are. In the optimal case flex\_extract resides in the FLEXPART directory structure.

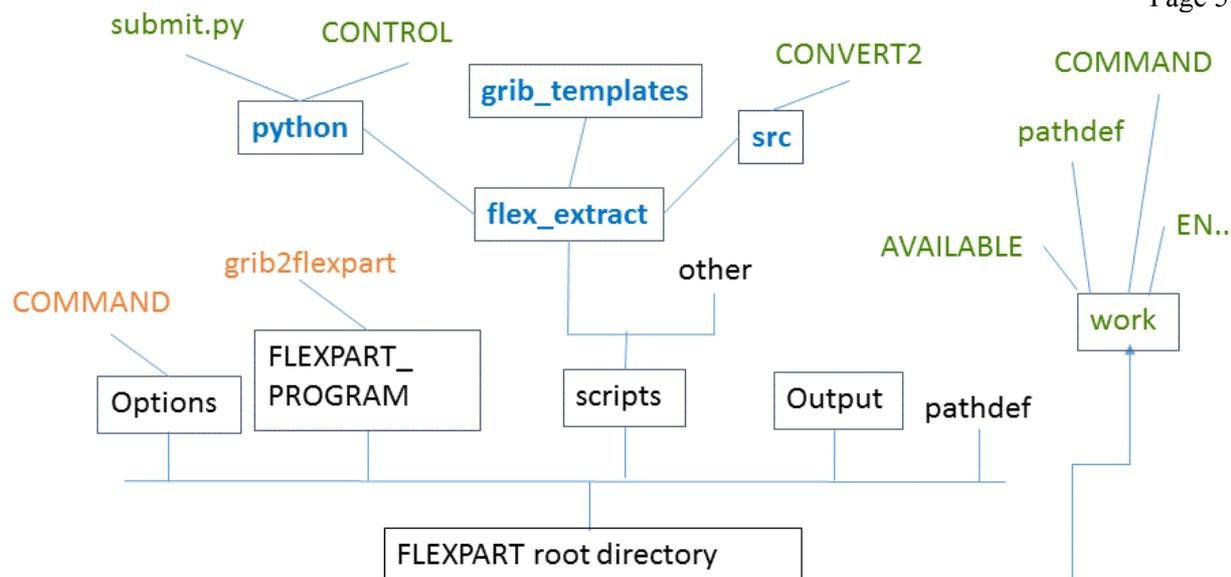


Fig. 1: FLEXPART directory structure (black) with directories in rectangles, files as text. Flex\_extract software should reside in scripts directory. In order to create input files on height levels (fp files) flex\_extract will need to have access to COMMAND file and grib2flexpart executable (red). Flex\_extract software will create directory work and place all output files there. Usually the work directory will reside on a separate large file system.

### 3. TOP-LEVEL SCRIPTS: SUBMIT.PY

Flex\_extract “on-demand” is implemented as a top-level python-script. It reads command line parameters and a CONTROL file. Based on this input information it either retrieves the FLEXPART input data either on the local machine via WebMARS or it submits a job to ecgata that retrieves the data and sends them to the local machine.

The necessary python scripts and programs reside in a directory tree that has its root in the parent directory of the script **submit.py**. The directory tree should also contain directories grib\_table and src, otherwise it would not work.

The scripts has two sources of information about which data to retrieve:

- A file with control parameters (we call it CONTROL file). The default CONTROL file is named CONTROL.temp and resides in the flex\_extract\_v7.0.3/python directory.
- Command line parameters

Before we discuss those the user should change into directory flex\_extract\_v7.0.3/python and issue

```
submit.py -start_date=YYYYMMDD
```

This should initiate a retrieval of global fields from ERA-Interim via the ECMWF Web API. YYYYMMDD should be some valid date between 2000 and 2014. If this command fails please consult the Software Installation Protocol (SIP) or your system administrator. The most common error at this stage is that the ECMWF WebAPI certificate has not yet been downloaded by the user. The other common error is that the user has not yet compiled the program for calculating the hybrid coordinate vertical velocity (CONVERT2 fails or cannot be found). In this case as well please consult the SIP on how to compile CONVERT2.

### 3.1. The CONTROL file

For the version 7.0 of the software, the following parameters can be specified in the CONTROL file:

**UPDATE:** In order to address different datasets in version 7.0.3 the specification of type, time and step for accumulated data has been separated from the normal data specification. Therefore new parameter ACCTYPE, ACCTIME and ACCMAXSTEP were defined. See below for their meaning.

DAY1/START_DATE		Format YYYYMMDD
DAY2/END_DATE		Format YYYYMMDD – can be same as DAY1
DTIME	3	Time interval – set to 1 for hourly resolution
DATE_CHUNK	3	Maximum number of days retrieved at once in a MARS request. This can be set larger for reanalysis data but may be too large for high resolution retrievals.
BASETIME		Can be set to 00 and 12. This parameter is intended for retrievals during operational ECMWF dissemination. It should not be used in on demand retrievals. Only half a day will be retrieved.
[M_]TYPE	24 values, e.g. (AN, FC, CV)	Type of Field at each hour of day. Analyses (type AN) are available at 00/06/12/18. For better temporal resolution, the space between the AN can be filled with forecasts (FC). Note that the M_ prefix is no longer necessary for any parameter in version 7.0 although it is accepted for backward compatibility.
[M_]TIME	24 values, e.g. 00, 01	Time of start of forecast/analysis at each hour. To get fields between 1 and 11 hours, forecasts starting at 00 are normally used. To get fields between 13 and 23h, forecasts starting at 12 are used. Analyses are normally used at 00/06/12/18
M_STEP	24 values, e.g. 00, 01	Forecast step at each hour. To get, e.g. a field valid for 21GMT, one specifies M_TIME=12, M_STEP=9
ACCTYPE	TYPE[2]	The type of field for accumulated data extraction. For downward compatibility, the default value is the type at the second position of the original TYPE parameter TYPE[2]. Note: This is very important as soon as you change the original type parameter from FC to AN for example.
ACCTIME	00/12	Start times of accumulated forecast fields (for fluxes). This is in general an optional parameter. For downward compatibility, the default value is 00/12 and appropriate for most datasets. The ACCTIME parameter is needed as soon as the specified dataset has different start times, e.g ERA5 has start times 06/18 and CERA20C has only 18.
ACCMAXSTEP	12	This specifies the maximum step for a specific accumulated forecast start time. For example if you want to retrieve Era-Interim with starting times 00 and 12 you should retrieve the accumulation fields for steps

		1 to 12 corresponding to 1,2,3,4,5,6,7,8,9,10,11,12 UTC and 13,14,15,16,17,18,19,20,21,22,23,24 UTC. If you want to retrieve ERA5 data you have to set this parameter to 24 since there is only one forecast starting time for this dataset per day.
M_STREAM	OPER	Normally the operational stream. Use ENFO for ensemble forecasts
DATASET	"	This keyword has to be defined as soon as you want to retrieve public data. This is a different server and available for everyone after registration at ECMWF (see installation plan). Available public dataset can be found here: <a href="https://software.ecmwf.int/wiki/display/WEBAPI/Available+ECMWF+Public+Datasets">https://software.ecmwf.int/wiki/display/WEBAPI/Available+ECMWF+Public+Datasets</a>
M_CLASS	OD	Use EI for ERA-Interim, E4 for ERA-40, EA for ERA5
M_EXPVER	1	Experiment number, necessary for RD experiments, E-suites
M_NUMBER	0	0 in most cases but necessary for access to individual ensemble members of ensemble forecasts. 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. Right now only one ensemble member per retrieval can be extracted.
M_GRID	1000	Horizontal resolution of Lat/Lon grid in 1/1000 degree
M_RESOL	159	Horizontal resolution of spectral fields
M_LEVEL	137	Number of vertical levels
M_LEVELIST	ALL	List of vertical levels for MARS request
M_FORMAT	GRIB1	Output format (either GRIB1 or GRIB2). Use GRIB2 only when using FLEXPART versions >9.2
M_PREFIX	EN	Prefix of output files
GRIB2FLEXPART		Path to grib2flexpart program. If it is set, the script will try to convert the EN files into .fp format using grib2flexpart
MAXSTEP	12	This is an important new parameter. It allows to retrieve data from forecasts longer than 24 hours. Since MAXSTEP>24 the forecasts from different days overlap, the naming scheme of the EN files changes from ENYYMMDD to ENYYMMDD.HH.SSS where HH is the hour of the start of the forecast and SSS is the forecast step in hours.
M_ADDPAR		Additional optional surface parameters (27/28/173/186/187/188/235/139/39)
M_WRF	0	If set to 1, additional parameters needed to run WRF are extracted. This is experimental.
M_CWC	0	If set to 1, cloud water content (cloud liquid water and cloud ice) is retrieved and included in the EN files.
M_UPPER		Latitude of upper right corner of grid area to be retrieved in 1/1000 degree (e.g. 37500 for 37.5N)

M_LOWER		Latitude of lower left corner of grid area in 1/1000 degree
M_LEFT		Longitude of lower left corner of grid area in 1/1000 degree
M_RIGHT		Longitude of upper right corner of grid area in 1/1000 degree. For cyclic (global) grids, use e.g. M_LEFT=-180000+M_GRID, M_RIGHT=180000. For noncyclic grids crossing the dateline (180W), M_RIGHT may be smaller than M_LEFT.
M_ACCURACY	24	Accuracy of GRIB data in bits
M_GAUSS	1	Calculation of etadot on Lat/Lon grid (0) or on Gaussian grid (1)
M_OMEGA	0	0/1 (0) Retrieve Omega from MARS and put it to file OMEGAyymmddhh
M_OMEGADIFF	0	0/1 (0) Calculate Omega and Dps/Dt from continuity equation for diagnostic purposes and include it in file OMEGAyymmddhh
M_ETA	1	0/1 (1) Read etadot precalculated by ECMWF and multiply it with dp/deta to be compatible with etadot calculated from continuity equation Etadot calculation from continuity equation either on Gaussian or lat/lon grid is disabled unless M_ETADIFF is set to 1 as well. Precalculated etadot is operationally available from September 2008 onwards. Setting M_ETA=1 for earlier dates or for ERA-40/Interim data leads to failure of the extraction script.
M_ETADIFF	0	0/1 (0) Calculate etadot and Dps/Dt from continuity equation for diagnostic purposes and include it in file ETAYymmddhh. Expensive option, only for debugging purposes.
M_ETAPAR	77	GRIB parameter number for ETADOT DP/DETA (77)
M_SMOOTH	0	Spectral truncation of etadot after calculation on Gaussian grid
M_DPDETA	1	Multiply etadot with dpdeta – this is the default. In some future version this may change.
GATEWAY		Name of eaccess gateway server (ctbto4.ctbto.org). Normally it is not needed since GATEWAY is set during installation.
DESTINATION		Name of destination at ECaccess gateway server (atmops@ops) Normally it is not needed since DESTINATION is set during installation
ECSTORAGE	1	Store EN-files in ECFS file system 0/1 (1)
ECTRANS	0	Transfer EN-files to gateway server 0/1 (0)
ECFSDIR		Destination directory on ECFS file system (ectmp:/cba/ecops/)
COMPRESSION		Not implemented in flex_extract v7.0.3
MAILOPS		List of email addresses for operational log-files (cba)
MAILFAIL		List of email addresses for failure log-files (cba)

EXEDIR		Location of CONVERT2, default is in ../src relative to the flex_extract_v7.0.3/python directory. This parameter is no longer needed
SOURCECODE		This parameter is no longer needed.

The format of the CONTROL file is very simple: The first string of each line is the parameter name, the following string(s) separated by spaces is (are) the parameter values. For most parameters default values are available. If the default value is appropriate the parameter can be omitted (although this is not recommended). Some of the parameters may be overruled by command line parameters.

The parameters DAY1, DAY2 (or alternatively START\_DATE and END\_DATE have to be specified in the form YYYYMMDD for consistency with MARS conventions. The naming of the resulting EN-files is of the form ENYYMMDD in accordance with the SRS, i.e. the century is omitted.

Note that the parameters ECSTORAGE, ECTRANS, GATEWAY, DESTINATION, SOURCECODE are no longer necessary for local retrievals. Only if the retrievals are done via ecgate or the HPC they are used. ECTRANS is set to 0 as default to avoid unintended triggering of file transfers when testing the scripts. The users should activate ECTRANS only if they have verified that copying from ECMWF with ectrans works.

### **3.1.1. Hard-coded specifications**

The hard coded specifications for special datasets that were necessary in versions 1-4 to access e.g. ERA-40 data have been removed. Access to these datasets can be specified with parameters M\_TYPE-M\_CLASS above in version 7.0

At the end of the scripts there are commented `rm -r *` lines that do a cleanup. Leaving the working directory has the advantage that debugging is simplified. It may be necessary to uncomment these lines in case of disk space limitations.

### **3.1.2. Compilation of Fortran90 programmes**

The script submit.py does not trigger any compilation of Fortran90 programmes. These programmes are compiled once during installation using the script install.py. The script submit.py assumes that the program CONVERT2 is available in the flex\_extract\_v7.0.3/src directory.

## **3.2. Command line parameters**

Submit.py can retrieve all necessary information from the control file. However, it is possible to override a few settings in the control file with command line parameters. This helps reducing the number of control files.

```
usage: submit.py [-h] [--start_date START_DATE] [--end_date END_DATE]
                [--basetime BASETIME] [--step STEP] [--levelist LEVELIST]
                [--area AREA] [--inputdir INPUTDIR] [--outputdir OUTPUTDIR]
                [--flexpart_root_scripts FLEXPART_ROOT_SCRIPTS] [--ppid PPID]
                [--job_template JOB_TEMPLATE] [--queue QUEUE]
                [--controlfile CONTROLFILE] [--public PUBLIC] [--debug DEBUG]
```

The meaning of the parameters is as follows

- start\_date START\_DATE**  
start date YYYYMMDD (default: None). This is a parameter that is often specified in practice.
- end\_date END\_DATE**  
end\_date YYYYMMDD (default: None). If not specified it is set to START\_DATE, so that one day is retrieved.
- basetime BASETIME**  
base such as 00/12 (for half day retrievals) (default: None). This parameter is normally set only for operational twice daily retrievals for dissemination. Deaccumulation of fluxes is less accurate if this parameter is set, so it should be used only if needed.
- step STEP**  
steps such as 00/to/48 (default: None). It is sufficient, however, to specify only the last step to override MAXSTEP in the CONTROL file
- levelist LEVELIST**  
Vertical levels to be retrieved, e.g. 30/to/60 (default: None).
- area AREA**  
area defined as north/west/south/east (default: None). Values are given as float or in thousandths of a degree, so 90./-179./-90/180. is valid. If set it overrides the values UPPER, LEFT, LOWER, RIGHT in the parameter file
- inputdir INPUTDIR**  
directory for storing intermediate files (default: None). If not set as argument or in the control file, flex\_extract\_v7.0.3/work is used. This directory should be on a large file system.
- outputdir OUTPUTDIR**  
directory for storing output files (default: None). If not set the EN files are stored in INPUTDIR
- flexpart\_root\_scripts FLEXPART\_ROOT\_SCRIPTS**  
This specifies the path to the FLEXPART\_ROOT/scripts. FLEXPART\_ROOT is the directory where e.g. the pathdef is found (see Fig. 1). The program grib2flexpart is expected to be in the FLEXPART\_ROOT\_SCRIPTS/./FLEXPART\_PROGRAM directory. (default: None) This parameter is necessary only if .fp files are to be produced.
- ppid PPID**  
Specify parent process id for rerun of prepareFLEXPART. It should not be used with submit.py (default: None).
- job\_template JOB\_TEMPLATE**  
job template file for submission to ECMWF (default: job.temp). This parameter may be useful e.g. if other queues on ecgate are desired (e.g. large instead of normal).
- queue QUEUE**  
queue used by eaccess\_job\_submit, typically ecgate or cca (default: None). If this parameter is not set, a retrieval on the local machine will be attempted.
- controlfile CONTROLFILE**  
full path of file with control parameters (default: CONTROL.temp). Typically there are different control files for different data classes (ei, e20, oper...) or stream or different areas.
- public PUBLIC**  
defines which server is used for retrieving data. By default its the memerb-state server which has limited access (default: False). If you would like to retrieve data from the public server please set this parameter to "True".
- debug DEBUG**

If DEBUG is set to something != 0 all the temporary files during the extraction are stored. The default value is 0 which means that all temporary files are deleted at the end. Only the final FLEXPART prepared GRIB files are left.

### 3.3. Retrieval of model data from MARS archive

Submit.py calls getMARSdata.py, which sends MARS-requests either on the local machine or on ecgate, that produce GRIB files named e.g.

```
FCGG__SL.20131107.13761.38723.grb
FCOG__ML.20131107.13761.38723.grb
FCSH__ML.20131107.13761.38723.grb
FCOG_acc_SL.20131106.13761.38723.grb
FCOG__SL.20131107.13761.38723.grb
FCSH__SL.20131107.13761.38723.grb
```

The first two letters specify the field type, e.g. FC for forecasts, AN or 4V for analyses, CV for ensemble control forecasts.

The second two letters denote the grid type, e.g. OG for output grid (typically lat/lon), GG for Gaussian grid, SH for spherical harmonics

\_\_ means instantaneous fields

\_acc means accumulated fields

SL/ML stands for surface fields or model level fields

Then follows the date, the process number of the parent process who called submit.py and the process number of submit.py. The process ids should avoid mixing of fields if several submit.py jobs are performed in parallel (which is, however, not recommended).

### 3.4. De-accumulation of flux data

After successful completion of the MARS requests, the fluxes are deaccumulated. In earlier versions of the retrieval scripts, this was done with a Fortran program (FLXACC2). In this version it is done within python since it is computationally not very demanding.

For precipitation the accumulated values are just divided by the number of hours (i.e., 1, 3 or 6). The accumulated values for the other variables are first divided by the number of hours and then interpolated to the exact times X using a polynomial fitting.

In order to be able to carry out the polynomial interpolation procedure proposed by Paul James, additional flux data is retrieved automatically for one day at the beginning and one day at the end of the period specified. Thus, data for flux computation will be requested for the period DAY1-1 to DAY2+1. Note that these (additional) dates are used only for interpolation within the script and are not communicated to IDC. The flux deaccumulation produces files named fluxYYYYMMDDHH.

Note also that for operational retrievals (BASETIME set to 00 or 12) forecast fluxes are only available until BASETIME, so that no polynomial interpolation is possible in the last two time intervals. This is the reason why setting BASETIME is not recommended for on demand scripts.

**NOTE:** For ERA5 data the deaccumulation works slightly different. Usually, the fields are summed up over all forecast steps. Therefore they are subtracted from each other so there is

only the data between the last step and the current one. In ERA5 data, this is different because they are not summed up anymore. This is automatically considered in the program.

### 3.5. Calculation of vertical velocity and preparation of EN-files

Flex\_extract version 7.0.3 has two ways to calculate the vertical velocity for FLEXTRA/FLEXPART: (i) from the horizontal wind field, (ii) from the new MARS parameter 77, which is available for operational forecasts and analyses since September 2008.

Especially for high resolution data, use of the new MARS parameter 77 is recommended, since the computational cost (measured in ECMWF HPC units) is reduced by 90-95% at T799. The extraction time, which depends heavily also on the performance of MARS, is generally reduced by 50% as well. The MARS parameter 77 is then multiplied by  $dp/deta$  to give a vertical velocity in Pa/s as needed by FLEXPART.

Calculation from the horizontal wind field is still required for historical case studies using ERA-40, ERA-Interim or operational data prior to September 2008

#### 3.5.1. Calculation of vertical velocity from horizontal wind using the continuity equation

The vertical velocity is computed by the FORTRAN90 program CONVERT2 in the ECMWF vertical coordinate system applying the equation of continuity thereby ensuring mass-consistent 3D wind fields. A detailed description of CONVERT2 can be found in the documents [v20\\_update\\_protocol.pdf](#), [V30\\_update\\_protocol.pdf](#) and [V40\\_update\\_protocol.pdf](#). The computational demand and accuracy of CONVERT2 is highly dependent on the specification of parameters M\_GAUSS, M\_RESOL and M\_SMOOTH. The following guidance can be given for choosing the right parameters:

- For very fine output grids (0.25 degree or finer) the full resolution T799 or even T1279 of the operational model is required (M\_RESOL=799, M\_SMOOTH=0). The highest available resolution (and the calculation of vertical velocity on the Gaussian grid (M\_GAUSS=1) is, however, rather demanding and feasible only for resolutions up to T799. Higher resolutions are achievable on the HPC. If data retrieval at T1279 needs to be performed on ecgate, the computation of the vertical velocity is feasible only on the lat/lon grid (M\_GAUSS=0), which also yields very good results. Please read document [v20\\_update\\_protocol.pdf-v60\\_update\\_protocol.pdf](#) to see if the errors incurred are acceptable for the planned application.
- For lower resolution (often global) output grids, calculation of vertical velocities with lower than operational spectral resolution is recommended. For global grids the following settings appear optimal:
  - For 1.0 degree grids: M\_GAUSS=1, M\_RESOL=255, M\_SMOOTH=179
  - For 0.5 degree grids: M\_GAUSS=1, M\_RESOL=399, M\_SMOOTH=359
  - Calculation on the lat/lon grid is not recommended for less than the operational (T1279) resolution.
  - If M\_GAUSS is set to 1, only the following choices are possible for M\_RESOL on ecgate: 159,255,319,399,511,799, (on the HPC also 1279, 2047 in future models). This choice is restricted because a reduced Gaussian grid is defined in then ECMWF EMOSLIB only for these spectral resolutions. For M\_GAUSS=0, M\_RESOL can be any value below the operational resolution.

- For M\_SMOOTH any resolution lower than M\_RESOL is possible. If no smoothing is desired, M\_SMOOTH=0 should be chosen. M\_SMOOTH has no effect if vertical velocity is calculated on lat/lon grid (M\_GAUSS=0).
- The on demand scripts send an error message for settings where M\_SMOOTH (if set) and M\_RESOL are larger than  $360000/M\_GRID/2$ , since in this case, the output grid cannot resolve the highest wave numbers. The scripts continue operations, however.
- Regional grids are not cyclic in zonal directions, but global grids are. The software assumes a cyclic grid if M\_RIGHT-M\_LEFT is equal to M\_GRID or is equal to M\_GRID-360000. Note that units for these parameters are 1/1000 degrees.
- Finally, model and flux data as well as the vertical velocity computed are written to files ENyymmddhh for application in ATM modelling. If the parameters M\_OMEGA or M\_OMEGADIFF are set, also files OMEGAyymmddhh are created, containing the pressure vertical velocity (omega) and the difference between omega from MARS and the surface pressure tendency. M\_OMEGADIFF should be zero except for debugging, since it triggers expensive calculations on the Gaussian grid.

### 3.5.2. Calculation of vertical velocity from precalculated MARS parameter 77

Since November 2008, the parameter 77 ( $d\eta/dt$ ) is stored in MARS on full model levels. FLEXTRA in its current version requires  $d\eta/dt$  on model half levels, multiplied by  $dp/d\eta$ . In flex\_extract version 7.0.3, the program CONVERT2 assumes that this parameter is available if the environment variable M\_ETA is set to 1. This is the default.

*It is recommended to use the precalculated parameter 77 by setting M\_ETA to 1 whenever possible.*

Setting parameter M\_ETA to 1 normally disables calculation of vertical velocity from the horizontal wind field, which saves a lot of computer time. Note, however that the calculation on the Gaussian grid are avoided only if both M\_GAUSS and M\_ETADIFF is set to 0. Please set M\_ETADIFF to 1 only if you are really need it for debugging since this is a very expensive option. In this case ETAYymmddhh files are produced that contain the vertical velocity from horizontal winds and the difference to the precalculated vertical velocity.

The parameters M\_RESOL, M\_GRID, M\_UPPER, M\_LOWER, M\_LEFT, M\_RIGHT still apply. As for calculations on the Gaussian grid, the spectral resolution parameter M\_RESOL should be compatible with the grid resolution (see previous subsection).

### 3.6. Creation of EN-files and communication to IDC

After completion of CONVERT2 the EN-files are created by concatenating GRIB files output by CONVERT2 as well as flux files and OROLSM files.

Model level fields are typically in GRIB2 format whereas surface level fields are still in GRIB1 format. They can be converted into GRIB2 if parameter [M\_]FORMAT is set to GRIB2. Please note, however, that older versions of FLEXPART may have difficulties reading pure GRIB2 files since some parameterIDs change.

FLEXPART can work with the EN files but needs to interpolate them from model levels to height levels, which is computationally demanding. This interpolation step can be done

already during the retrieval if FLEXPART is installed on the system and the program `grib2flexpart` is available. If the control parameter `GRIB2FLEXPART` is set to the path to `grib2flexpart`, the retrieval script will convert the `ENYYMMDDHH` files into `ENYYMMDDHH.fp` files. Please note that this step is recommended only for retrievals on the local machines since the `.fp` files are several times larger than the normal `EN` files. If executed on `ecgate` that would also increase the data transfer time by the same factor.

If the retrieval is executed remotely at ECMWF, the `EN`-files can be communicated to IDC via the `Ectrans` utility if the parameter `ECTRANS` is set to 1 and the parameters `GATEWAY`, `DESTINATION` have been set properly during installation. The status of the transfer can be checked with the command `ecaccess-ectrans-list` (on the local system). If the script is executed locally the progress of the script can be followed with the usual Linux tools.

## 4. JOB PREPARATION AND SUBMISSION

### 4.1. Job preparation and submission on remote host

While most users may prefer data retrieval on the local host, now since the WebAPI is in place, for more demanding tasks it is still advisable to retrieve and convert data on ECMWF machines and to download only the `EN` files. For a first try one could use

```
submit.py --queue=ecgate --start_date=YYYYMMDD
```

`submit.py` then reads the files `job.temp` and `CONTROL.temp` and digests the command line parameters. This information is merged and written into the file `job.ksh`. This file is then submitted to `ecgate`. It contains the necessary `SLURM` or `PBS` headers, the control files and the calls to `submit.py` on `ecgate`. Note that during installation a copy of the retrieval scripts must have been installed on `ecgate` (see `SIP` for details).

`CONTROL.temp` contains parameters for retrieving ERA-Interim data with a coarse resolution. As such it has a fast turnaround but may not be suitable for the needs of the user. The user may adapt `CONTROL.temp` to her/his needs or may create a copy that is then modified. The new control file will be used if the command line parameter `--controlfile` points to it.

A few example control files can be found in the `flex_extract_v7.0.3/python` directory that could be used as templates. The more accurately the control file describes the retrieval needed, the fewer command line parameters are needed when calling `submit.py`.

Once submitted one can check the progress of the submitted job using `ecaccess-job-list`. The user should get an email after the job is finished.

In case the job fails one can check on `ecgate` in the `$SCRATCH` -directory for debugging information.

```
cd $SCRATCH
cs -lrt
```

Lists the most recent job logs and temporary retrieval directories (usually `python??????`, where `??????` is the process id).

If the job was submitted to the HPC ( `--queue=cca` ) one may login to the HPC and look into the directory `/scratch/ms/ECGID/ECUID/.ecaccess_do_not_remove` for job logs. The working directories are deleted after job failure and thus normally cannot be accessed.

If `ectrans` is used to transfer the EN files to the local machine one may also check with `ecaccess-ectrans-list` if a transfer is ongoing or has already finished.

## 4.2. Job submission on local host

Due to the user-friendly interface and simplicity of the WebMARS interface it has become much easier to retrieve the FLEXPART input data directly to the local machine, without the need of a gateway server and without the need to submit a job to `ecgate` or `cca`.

If `submit.py` is called without the `--queue` argument, it tries to retrieve data via WebMARS. A typical call will be

```
submit.py --start_date=YYYYMMDD --inputdir=$SCRATCH/work
```

In this case `submit.py` reads the parameters from file `CONTROL.temp` and retrieves all data into `$SCRATCH/work`. If the directory does not exist, it will be created. The EN files will also be in `$SCRATCH/work` unless `--outputdir` is set.

While job submission on the local host is convenient and easy to monitor, there are a few caveats with this option:

1. There is a maximum size of 20GB for single retrieval via WebMARS. Normally this is not a problem but for global fields with T1279 resolution and hourly time steps the limit may already apply.
2. If the retrieved MARS files are large but the EN files are relative small (small local domain) then the retrieval to the local host may be inefficient since all data must be transferred via the Internet. This scenario applies most notably if `etadot` has to be calculated via the continuity equation as this requires global fields even if the domain is local. In this case job submission via `ecgate` might be a better choice. It really depends on the use patterns and also on the internet connection speed.

## 5. EXAMPLE

The following subsections provides a step-by-step example for the retrieval of historical data sets.

### 5.1. Data retrieval (analyses + short term forecasts) from the operational archive

```
cd PATH_TO_flex_extract_v7.0.3/flex_extract_v7.0.3/python
submit.py --controlfile=CONTROL_OPS_V6.0 --inputdir=$SCRATCH/work
```

### 5.2. Data retrieval (longer term forecasts) from the operational archive

```
cd PATH_TO_flex_extract_v7.0.3/flex_extract_v7.0.3/python
```

## **6. MONITORING AND ERROR HANDLING**

Monitoring of the job has already been described above.

### **Glossary**

#### **Abbreviations**

CTBTO	Comprehensive Nuclear-Test-Ban Treaty Organization
IDC	International Data Centre
PTS	Provisional Technical Secretariat
HPC	High Performance Computing

#### **REFERENCES**

Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) (2002). Editorial Manual.

International Data Centre (IDC) (2002). IDC Software Documentation Framework.