ECMWFDATA V 2.0

A software system for retrieving and storing meteorological data for use in atmospheric transport modelling.

Univ. Ass. Dr. Leopold Haimberger
Institut fuer Meteorologie und Geophysik
Universitaet Wien

1. Installation 3
1.1. Installation procedures 3
1.1.1. Installation using a gateway server 3
1.1.2. Interactive installation on ecgate without a gateway server 4
1.2. Installation testing 5
1.2.1. Success criteria 5
1.2.2. Installation checklist 5
1.3. Rollback procedures 6
2. Retrieval of ECMWF data 6
2.1. Specifications within the script 7
2.1.1. Hard-coded specifications 8
2.1.2. Polynomial interpolation of flux data 8
2.1.3. Compilation of Fortran90 programmes 8
2.2. Retrieval of model data from MARS archive 8
2.3. De-accumulation of flux data 9
2.4. Calculation of vertical velocity and preparation of EN-files 9
2.5. Checking of EN-files and communication to local system 10
2.6. Job preparation and submission 10
2.7. Example 10
2.7.1. Data retrieval from the operational archive 10
2.8. Monitoring and ERROR handling 11

ECMWFDATA V 2.0 is a software system that retrieves and stores meteorological data for use in atmospheric transport modelling (ATM). The ECMWFDATA software system consists of one subsystem that runs on the ECMWF server, one subsystem that runs on the local ECaccess gateway server, and one subsystem that runs on the local operational server. The first subsystem will retrieve data from the ECMWF Mars database and send it to the local ECaccess gateway server. The second subsystem will perform data processing and quality control.

The system will be interoperable with the existing FLEXPART and HYSPLIT ATM models. The system will operate continuously for long periods and will report system status via emails and log files. The user will control the system through text configuration files. Data will be regularly deleted from the operational archive and put on the mass storage system.

1. INSTALLATION

ECMWFDATA needs a UNIX/LINUX environment with an implementation of the Korn-shell. The installation requires a working FORTRAN90 compatible compiler as well as the availability of the ECMWF EMOS library.

1.1. Installation procedures

The ECMWFDATA v2.0 software is distributed as one tar-file `flex_extract_ecgate_V2.tar` containing korn shell (ksh)-scripts, FORTRAN90 source code files and three sample ksh scripts are provided for on demand retrievals.

The tar file contains a directory `flex_extract_ecgate_V2` with the following files:

- *update_script.ksh* copies all necessary files to the ECMWF member state server (ecgate) into the directory `$HOME/flex_extract_ecgate_V2`. $HOME is the user's home directory on ecgate, e.g. /home/ms/xxx/xxx.
- *upload_source* copies only the file `Source.tar` s to the directory `$HOME/flex_extract_ecgate_V2` on ecgate
- *source.tar* contains FORTRAN90 source files and Makefiles
- The files *flex_ecmwf_???.tmp* are sample on demand scripts, which contain more options. During the installation some user specific settings are changed and they are copied to files without the .tmp suffix.The configuration file is generated in the script. They are normally submitted by the user.

The recommended installation procedure depends on the availability of a gateway server. If one exists, the software can be installed without the need to log in interactively on ecgate.

1.1.1. Installation using a gateway server

If a local gateway server exists or if there is no firewall between the ECMWF standard gateway server ecaccess.ecmwf.int and the local server, one can install the ECMWFDATA v2.0 software without the need to login interactively on ecgate.

On the local system (suitable directory) execute:

```
tar xvf flex_extract_ecgate_V2.tar
```

If necessary renew the ECACCESS certificate:

```
eccert
```
Enter `ecmwf_userid` and `passcode`

Then enter
```
setenv ECUID ecmwf_user_id
setenv ECPATH ecmwf_group path
setenv GATEWAY gatewayserver
setenv DESTINATION directory
```

Then enter:
```
cd flex_extract_ecgate_V2
ksh update_script.ksh
```

update_script.ksh sets the user specific lines in the scripts Then it sends the updated scripts to ecgate. It also submits a test script (flex_ecmwf_91_global_1.0_ecgate automatically. The user should receive email notification that the scripts have started after a few minutes. The scripts run at least 20 minutes but can take longer depending on the load on ecgate and on the Meteorological Archival and Retrieval System (MARS) server.

The others can be submitted manually if needed. For example one can try out:
```
ecjput ecgate flex_ecmwf_91_global_1.0_ecgate
```
This retrieves the input files on a 1.0/1.0 degree global grid. If there is enough disk quota on ecgate, one can also execute:
```
ecjput ecgate flex_ecmwf_91_finegrid_ecgate
```
If access to the HPC is available, one can also try
```
ecjput hpce flex_ecmwf_91_finegrid_ecgate
```

The example scripts send emails to `$ECUID`, notifying success and failure of the scripts. The script protocols are archived as well in directories `ectmp:/$ECUID/econdemand` or `ectmp:/$ECUID/econdemand.`

1.1.2. Interactive installation on ecgate without a gateway server

If the local server is behind a firewall and no local gateway server exists, the software needs to be installed interactively on ecgate.

After the file flex_extract_ecgate.tar is copied into the $HOME-directory on ecgate, the installation steps on ecgate are as follows:

```
tar xvf flex_extract_ecgate_V2.tar
```

Then enter
```
setenv ECUID ecmwf_user_id
setenv ECPATH ecmwf_group path
```

Since no automatic transfer of data to the local server via ecaccess is possible, the parameter `ECTRANS` in the CONTROL_ERA configuration files should be set to 0. Instead, one should replace the ectrans commands in the scripts with scp commands to the local machine.

Then continue with
```
cd flex_extract_ecgate_V2
ksh update_script.ksh
```

update_script.ksh sets the user specific lines in the scripts. It also submits two (relatively fast) test scripts automatically. One can monitor progress of the scripts via the command
llq | grep $USER

The other scripts can be submitted manually if needed, e.g. with the command
`llsubmit flex_ecmwf_91_global_1.0_ecgate`
If enough disk space is available, one could also try
`llsubmit flex_ecmwf_91_finegrid_ecgate`
If access to the HPC is available, one can login to hpce and try
`llsubmit`
`$WSHOME/flex_extract_ecgate_V2/flex_ecmwf_91_finegrid_ecgate`

The scripts send emails to `$ECUID`, notifying success and failure of the scripts. The script protocols are archived as well in directories `ectmp:/$ECUID/econdemand` or `ectmp:/$ECUID/econdemand.`

1.2.Installation testing

1.2.1. Success criteria

After installation, the directory $HOME/flex_extract_ecgate_V2 with the scripts and with the other files mentioned above must exist on ecgate. In the scripts, reference to the right ECMWF userID should be made. The command

```
fgrep $USER  $HOME/flex_ecmwf_91_finegrid_ecgate_V2/*
```

on ecgate should output several lines where reference is made to the ECMWF userID in the scripts. If no such references are found, the references to user lh0 in the scripts and in the control files must be changed to $USER or the installation process must be repeated with the ECUID variable set (see above).

The installation was successful if the example scripts (for ecgate) produce ENyymmddhh files and the CHECK program does not fail (see protocol files and email notification).
The ENyymmddhh files are stored by default on ectmp:/$ECUID/econdemand and if ECTRANS is set to 1, they are also sent to the local system.

The execution of the scripts after submission can be followed by issuing
`llq | grep $USER`
on ecgate or by the Web access facilities provided by ECMWF (see www.ecmwf.int for documentation of these). An email notifying the successful completion of the script should be sent by the script to the specified user, who submitted the jobs. Note that the scripts involve MARS requests which can take from a few minutes up to a few hours, depending on the load of the MARS system.

If the scripts fail, the log files can be found on ecgate in the directory $SCRATCH for the on demand scripts and in the directory $SCRATCH/ms_sms_output_V2 for the operational scripts.

1.2.2. Installation checklist

The following checks should be performed

- Is there a directory flex_extract_ecgate_V2 on ecgate and does it contain scripts and source code? If no, repeat steps 3.2.1 or 3.2.2.

- Submit some of the example scripts (flex_ecmwf_91_global_1.0_ecgate is recommended, since they take least time). After submission:
    - Are the scripts running? (see 3.3.1 how to check this)
    - Is there a new log file in $SCRATCH or $SCRATCH/ms_sms_output_V2?
- After the scripts have been completed:
    - Did you get a notification via email? All parameters searched by program CHECK should be OK.
        - If not, check the email and the job log files for error messages.
    - Are there new files ENyymmddhh at the places noted in 3.3.1?
        - If not check the paths set in the in CONTROL_ERA configuration file
    - If ECTRANS is set to 1: Are there new files ENyymmddhh on the local gateway?
        - If not check the paths set in the in the CONTROL_ERA configuration file

1.3.Rollback procedures

The software installation procedure should not in any case cause failures of the overall system. The software can be removed completely by executing the command

```
\rm -r $USER/flex_extract_ecgate_V2
```

on ecgate.

2. RETRIEVAL OF ECMWF DATA

ECMWFDATA "on-demand" is implemented as a top-level ksh-script. Three example scripts are provided:

- flex_ecmwf_91_global_1.0_ecgate retrieves data on a 1.0/1.0 degree global grid at resolution T179 and calculates the vertical velocity on the Gaussian grid
- flex_ecmwf_91_finegrid_ecgate retrieves data on a 0.2/0.2 degree grid over Europe at resolution T799. Vertical velocities are calculated on lat/lon grid.
- flex_ecmwf_91_finegrid_hpce retrieves data on a 0.2/0.2 degree grid over Europe at resolution T399. Vertical velocities are calculated on Gaussian grid. Full accuracy can be achieved by setting resolution to T799 (see below)

These scripts are intended as templates and can be changed as appropriate. In most cases at least the date (DAY1, DAY2) has to be changed. Note that several settings, e.g. the ecaccess gateway, ECMWF userID should have been set already during installation. It should not be necessary to change these.

The submission of the script on ecgate is usually carried out vie the ECaccess software package using command ecjput. Upon successful completion of the job the output data (i.e., the EN-files) are communicated to the local system via the ECtrans software that is part of the ECaccess package. The user definable parameters are changed within the script itself. No further input files are necessary. The FORTRAN90 source code is assumed to be available on the remote system (i.e., ecgate). As in the operational configuration, the script subsequently

- Compiles the executables on ecgate from the provided source code
- Retrieves model output from the ECMWF MARS archive
- De-accumulates the atmospheric flux data (FORTRAN90 program FLXACC2)
- Calculates the vertical velocity from the equation of continuity and prepares the input data (EN-files) for ATM modelling (FORTRAN90 program CONVERT2)
- Checks the resulting EN-files for completeness and internal consistency (FORTRAN90 program CHECK) and communicates them to the local system via the ECtrans utility

2.1. Specifications within the script

For the version 2.0 of the software, the following parameters may be changed by the user (defaults are given in brackets):

FORCE_E40	0	Retrieve data from class E4 (ERA-40)
FORCE_EI	0	Retrieve data from class EI (ERA-Interim)
FORCE_EPS	0	Retrieve data from ensemble prediction system
M_STREAM	OPER	Allows to retrieve data from IFS experiment
M_EXPVER	1	Experiment number, necessary for ERA-Interim
M_NUMBER	1	Forecast leg, necessary for EPS. 1=T399 forecasts, 2=T255
M_GRID	100	Horizontal resolution of Lat/Lon grid in 1/100 degree
M_RESOL	159	Horizontal resolution of spectral fields
M_LEVEL	91	Number of vertical levels
M_LEVELIST	ALL	List of vertical levels for MARS request
M_ADDPAR		Additional surface parameters (/27/28/173/186/187/188/235/139/39)
M_UPPER		Latitude of upper right corner of grid area to be retrieved in Hundredth of a degree (e.g. 3750 for 37.5N)
M_LOWER		Latitude of lower left corner of grid area in 1/100 degree
M_LEFT		Longitude of lower left corner of grid area in 1/100 degree
M_RIGHT		Longitude of upper right corner of grid area in 1/100 degree.
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	Retrieve Omega from MARS and put it to file OMEGAyymmddhh
M_OMEGADIFF	0	Calculate Omega and Dps/Dt from continuity equation for diagnostic purposes and include it in file OMEGAyymmddhh
M_SMOOTH	0	Spectral truncation of etadot after calculation on Gaussian grid
GATEWAY		Name of ECaccess gateway server
DESTINATION		Name of destination at ECaccess gateway server
ECSTORAGE	1	Store EN-files in ECFS file system 0/1 (1)
ECTRANS	1	Transfer EN-files to gateway server 0/1 (1)

ECFSDIR		Destination directory on ECFS file system
MAILOPS		List of email addresses for operational log-files
MAILFAIL		List of email addresses for failure log-files
EXEDIR	./	Location of FORTRAN executables
SOURCECODE		Location of file source.tar containing source code and makefiles ($HOME/flex_extract_ecgate_V2)

Specification of the time-interval is mandatory! All other parameters are optional with default values as indicated in the table. Note that the dates (DAY1, DAY2) have to be specified in the form yyyymmdd for consistency with MARS conventions. The naming of the resulting EN-files on the other hand is of the form ENyymmdd in accordance with the SRS.

2.1.1. Hard-coded specifications

A number of specifications have been hard-coded according to the requirements. The following table summarizes these specifications:

Specification	Purpose	Date
ERA40START	Starting date for ERA40 data	19570902
ERA40END	Ending date for ERA40 data (one month overlap)	20001013
VAR12START	Starting date for 12h 4D-Var	20000913
RES511	Horizontal resolution changed to T511	20001121

The MARS requests are tailor-made according to the choice of DAY1 and DAY2 and the specifications impose above. Note that between ERA40START and ERA40END reanalyses data is retrieved whereas from VAR12START onwards operational 4D-Var data is used.

The switch FORCE_E40 overrides these specifications thus requesting ERA40 data independent of the date specification. Note that no consistency checks whatsoever are made in this case. The same refers to switches FORCE_EPS and FORCE_EI. It is evident that not more than one of the three switches can be set to 1. It is the user's responsibility to check whether the job finishes successfully or fails due to MARS data requested being unavailable.

2.1.2. Polynomial interpolation of flux data

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 local system.

2.1.3. Compilation of Fortran90 programmes

The script automatically extracts the file SOURCECODE and compiles it in the temporary directory on the $SCRATCH filesystem on ecgate. The source code is normally installed during the installation process in its default place (ecgate:flex_extract_ecgate_V2).

2.2. Retrieval of model data from MARS archive

MARS requests for field and flux data are created within the ksh-script. These requests are subsequently forwarded to the MARS system.

2.3. De-accumulation of flux data

For precipitation the accumulated values are just divided by the number of hours (i.e., 3 or 6). The accumulated values for the other variables are first divided by the number of hours and than interpolated to the exact times X using a polynomial fitting. These computations are carried out within the FORTRAN program FLXACC2 adapted from code provided by A. Stohl, G. Wotawa, and P. James.

2.4. Calculation of vertical velocity and preparation of EN-files

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 document v20_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 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) yield the most accurate results. This is, however, rather demanding and feasible only on the HPC. If data retrieval needs to be performed on ecgate, the computation of the vertical velocity is feasible only on the lat/lon grid (M_GAUSS=0). Please read document v20_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 since it incurs the largest errors (see v20_update_protocol.pdf).
    - If M_GAUSS is set to 1, only the following choices are possible for M_RESOL: 159,255,319,399,511,799, (1023,2047 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 currently reject settings where M_SMOOTH (if set) and M_RESOL are larger than 36000/M_GRID/2, since in this case, the output grid cannot resolve the highest wave numbers.
- 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-36000. Note that units for these parameters are 1/100 degrees.
- The user can choose between input data from the operational 4DVAR-System (the default), from ERA-40 (FORCE_E40=1), from the EPS control forecasts (FORCE_EPS=1) or from the ERA-interim reanalysis (FORCE_EI=1). The user must

take care that the required input data are available in MARS. For some applications the EPS control forecasts may be an interesting option, since they are consistent at lower resolutions (T399L62 or T255L62) than the operational forecasts (T799L91)

- 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.

2.5. Checking of EN-files and communication to local system

The checking is carried out by the FORTRAN program CHECK. The EN-files are checked for completeness in terms of variables and vertical levels. If M_OMEGA or M_OMEGADIFF are set to 1, the completeness of the OMEGAyymmddhh files is also checked. After the successful check of the EN-files the data is communicated to local system via the ECtrans utility. The status of the transfer can be check with the command *ectls* (on the local system).

2.6. Job preparation and submission

Use of ECMWFDATA in "on-demand" configuration is relatively simple: Assuming a working installation of the FORTRAN executables and availability of the on demand scripts
flex_ecmwf_91_global_1.0_ecgate,
flex_ecmwf_91_finegrid_ecgate or
flex_ecmwf_91_finegrid_hpce
locally the following steps are necessary for the retrieval of ATM data:

- Create a personal copy of the script to avoid overwriting
- Edit the specifications section, in particular, the starting and ending dates
- Submit the script remotely to the ECMWF server via *ecjpu, e.g.*
  `ecjput ecgate flex_ecmwf_91_global_1.0_ecgate`

Note that e*cjput* requires the explicit specification of a (remote). See ECACCESS documentation and the examples below.

2.7. Example

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

2.7.1. Data retrieval from the operational archive

To retrieve data from the operational forecasts, the following steps are recommended:

- If ECMWFDATA should run on the HPC, open script flex_ecmwf_91_finegrid_hpce in an editor of your choice. If ECMWFDATA should run on ecgate, open script flex_ecmwf_91_finegrid_ecgate (for regional grid) or script flex_ecmwf_91_global_1.0_ecgate (for global grid).

- In the scripts the file CONTROL_ERA is created, which contains the control parameters. This is different to the operational scripts, where the control parameter file must be provided externally. Make changes as appropriate. Check especially the settings for DAY1, DAY2, M_GAUSS, and of the output grid.

Submit the scripts either on ecgate using llsubmit, e.g.:
```
llsubmit flex_ecmwf_91_global_1.0_ecgate
```
or on the local machine via the ecaccess server:
```
ecjput ecgate flex_ecmwf_91_global_1.0_ecgate
```

2.8. Monitoring and ERROR handling

Monitoring of the job is possible with the command *ecjls* that is part of the ECaccess package. Upon successful completion the script automatically sends its log-file as email to the addresses specified in MAILOPS. Additionally, possible non-critical warnings are reported.

In case of aborting the log-file is sent to the addresses specified in MAILFAIL indicating the cause of the abort. Note that in case of an abort the working directory under $SCRATCH is not deleted automatically and thus has to be deleted manually by the user to avoid flooding of the scratch space.