Using the ecCodes GRIB Tools: comparison and modification tools

Computer User Training Course 2016

Paul Dando

User Support Section

advisory@ecmwf.int

ECMWF

Contents

- Comparing messages
- Copying messages
- Setting key / value pairs
- Converting from GRIB to NetCDF

grib_compare – compare GRIB messages

- Use grib_compare to compare the GRIB messages contained in two files
- By default, messages are compared in the same order, bit-by-bit and with floating point values compared exactly
  - Tolerances for data values can be specified based on the absolute, relative or packing error
  - Default tolerance is absolute error = 0
- If differences are found grib_compare
  - switches to a key-based mode to find out which coded keys are different
  - fails returning a non-zero exit code
- Options are available to compare only specific keys or sets of keys

grib_compare – basic usage

```
grib_compare [options] grib_file grib_file
```

- Options

`-b key1,key2,…`	All keys in this list are skipped when comparing files
`-c key1,namespace2:n…`	Compare these keys only
`-H`	Compare message headers only
`-e`	Edition-independent compare
`-w key1=val1,key2!=val2,…`	Where option
`-f`	Do *not* fail on error
`-r`	Messages not in the same order
`-v`	Verbose
`...`	

grib_compare – a simple example

- Two GRIB messages in f1.grib1 and f2.grib1 contain the land-sea mask at different forecast time steps

```
> grib_compare f1.grib1 f2.grib1
-- GRIB #1 -- shortName=lsm paramId=172 stepRange=3
   levelType=sfc level=0 packingType=grid_simple
   gridType=reduced_gg --
long [P1]: [3] != [6]

> echo $?
1
```

- The exit code is set to 1 because the comparison failed

grib_compare – a simple example

- If we blacklist the key P1 and compare the files again

```
> grib_compare -b P1 f1.grib1 f2.grib1

> echo $?
0
```

- The exit code is set to 0 because the comparison is successful according to the blacklist

grib_compare – verbose output

- The verbose option shows details of all keys being compared

```
> grib_compare -v f1.grib1 f2.grib1
f1.grib1
    comparing totalLength as long
    comparing editionNumber as long
    comparing section1Length as long
    comparing table2Version as long
    comparing centre as string
    comparing generatingProcessIdentifier as long
    comparing gridDefinition as long
         ...
    comparing P1 as long
-- GRIB #1 -- shortName=lsm paramId=172 stepRange=3 levelType=sfc
evel=0 packingType=grid_simple gridType=reduced_gg --
long [P1]: [3] != [6]
    comparing P2 as long
    ...
```

grib_compare – limit the keys compared

- The `-c` option can be used to compare only specific keys

```
> grib_compare -c dataDate f1.grib1 f2.grib1
-- GRIB #1 -- shortName=lsm paramId=172 stepRange=3
   levelType=sfc level=0 packingType=grid_simple
   gridType=reduced_gg --
long [dataDate]: [20160224] != [20160225]
```

- Or a set of keys in a particular namespace

```
> grib_compare -c time:n  f1.grib1 f2.grib1
-- GRIB #1 -- shortName=lsm paramId=172 stepRange=3
   levelType=sfc level=0 packingType=grid_simple
   gridType=reduced_gg --
long [dataDate]: [20160224] != [20160225]
long [validityDate]: [20160224] != [20160225]
```

grib_compare – compare headers only

- To compare only the headers of two GRIB messages use the `-H` option

```
> grib_compare -H  f1.grib1 f2.grib1
-- GRIB #1 -- shortName=lsm paramId=172 stepRange=3
   levelType=sfc level=0 packingType= gridType= --
long [day]: [24] != [25]
```

- The **`-H`** option cannot be used with the **`-c`** option

grib_compare – edition-independent

- Two GRIB messages are very different if they are encoded with different editions

```
> grib_compare sp.grib1 sp.grib2
-- GRIB #1 -- shortName=sp paramId=134 stepRange=0 levelType=sfc
   level=0 packingType=grid_simple gridType=reduced_gg --
long [totalLength]: [4284072] != [4284160]
long [editionNumber]: [1] != [2]
long [section1Length]: [52] != [21]
[table2Version] not found in 2nd field
[gridDefinition] not found in 2nd field
[indicatorOfParameter] not found in 2nd field
[indicatorOfTypeOfLevel] not found in 2nd field
[yearOfCentury] not found in 2nd field
[unitOfTimeRange] not found in 2nd field
[P1] not found in 2nd field
[P2] not found in 2nd field
...
```

grib_compare – edition-independent

- Using the `-e` option grib_compare compares only the higher level information common to the two messages

```
> grib_compare -e sp.grib1 sp.grib2
-- GRIB #1 -- shortName=sp paramId=134 stepRange=0
   levelType=sfc level=0 packingType=grid_simple
   gridType=reduced_gg --
string [param]: [134.128] != [134]
```

- The two messages contain the same information encoded in two different ways
- Only the MARS param is different

grib_compare – summary of differences

- When files contain several fields and some keys are different, it is useful to have a summary report

```
> grib_compare -f f1.grib1 f2.grib1
-- GRIB #1 -- shortName=z paramId=129 stepRange=0 levelType=pl
   level=1000 packingType=spectral_complex gridType=sh --
long [marsType]: [an] != [fc]

-- GRIB #3 -- shortName=z paramId=129 stepRange=0 levelType=pl
   level=850 packingType=spectral_complex gridType=sh --
long [marsType]: [an] != [fc]
...
## ERRORS SUMMARY #######
##
## Summary of different key values
## marsType ( 6 different )
##
## 6 different messages out of 12
```

grib_compare – comparing data values

- By default floating point values are compared exactly
- Different tolerances can be provided using one of the following options

`-A absolute_error`	Use absolute error as tolerance
`-R key=rel_error,…`	Use relative error as tolerance for **`key`**
`-P`	Use packing error as tolerance
`-T factor`	Compare data values using tolerance specified in options **`-A`**, **`-R`**, **`-P`** multiplied by an integer **`factor`**

grib_compare – setting the tolerance

- Comparison of the data values in two files shows that one of the seven values is different with the default absolute error tolerance of zero

```
> grib_compare -c data:n f1.grib1 f2.grib1
-- GRIB #1 -- shortName=2t paramId=167 stepRange=0 levelType=sfc level=0
   packingType=grid_simple gridType=reduced_gg --
double [packedValues]: 1 out of 7 different
 max absolute diff. = 2.0000000000000000e+00, relative diff. = 0.4
        max diff. element 2: 3.0000000000000000000e+00
   5.0000000000000000000e+00
        tolerance=0.0000000000000000e+00 packingError: [0.0625005]
   [0.0625005]
        values max= [70]  [70]         min= [1] [1]
```

- Set the absolute error tolerance to 2.0 and the comparison is successful

```
> grib_compare -A 2.0 -c data:n  f1.grib1 f2.grib1
```

grib_compare – setting the tolerance

- We can also set a relative error as tolerance for each key

```
> grib_compare -c data:n f1.grib1 f2.grib1
-- GRIB #1 -- shortName=2t paramId=167 stepRange=0 levelType=sfc level=0
   packingType=grid_simple gridType=reduced_gg --
double [packedValues]: 1 out of 7 different
 max absolute diff. = 2.0000000000000000e+00, relative diff. = 0.4
        max diff. element 2: 3.0000000000000000000e+00
   5.0000000000000000000e+00
        tolerance=0.0000000000000000e+00 packingError: [0.0625005]
   [0.0625005]
        values max= [70]  [70]          min= [1] [1]
values max= [70]  [70]          min= [1] [1]
```

- Set a relative error of 0.4 as the tolerance for packedValues

```
> grib_compare -R packedValues=0.4 -c data:n  f1.grib1 f2.grib1
```

- The comparison is successful because the relative tolerance is greater than the relative difference

grib_compare – setting the tolerance

- Different packing precision can give different data values

```
> grib_compare -c data:n f1.grib1 f3.grib1
-- GRIB #1 -- shortName=2t paramId=167 stepRange=0 levelType=sfc level=0
   packingType=grid_simple gridType=reduced_gg --
double [packedValues]: 1 out of 7 different
 max absolute diff. = 5.0000000000000000e-01, relative diff. = 0.166667
        max diff. element 1: 2.5000000000000000000e+00
   3.0000000000000000000e+00
        tolerance=0.0000000000000000e+00 packingError: [0.0625005] [0.5]
        values max= [70]  [70]         min= [1] [1]
values max= [70]  [70]         min= [1] [1]
```

- Here we can use the packing error as the tolerance

```
> grib_compare -P -c data:n  f1.grib1 f3.grib1
```

- The comparison is successful because the maximum absolute difference is within the larger of the two packing errors – only the packing precision has changed

Practicals

- Work in your $SCRATCH

```
cd $SCRATCH
```

- Make a copy of the practicals directory in your $SCRATCH

```
tar -xvf /home/ectrain/trx/ecCodes/grib_tools2.tar
```

- There is a sub-directory for each practical:

```
ls $SCRATCH/grib_tools2
```

```
compare filter modify
```

Practical: using grib_compare

1. Use grib_compare to compare the GRIB messages contained in the files file1.grib and file2.grib
   - Which keys does grib_compare report as different ? What is the exit code returned ?
2. Now use the `-b` option to 'black list' the keys that you know are different and use grib_compare to compare the messages again
   - Are any keys reported as different ? What is the exit code ?
3. Compare the data namespaces (use `"grib_compare -c data:n …"`) for file1.grib and file2.grib. What values need to be set for the absolute (with `-A`) and relative (with `-R`) error tolerances for the comparison to be successful ?
4. Repeat question 3 with file2.grib and file3.grib [More difficult]

grib_copy – copy contents of GRIB files

- Use grib_copy to copy selected messages from GRIB files optionally printing some key values
- Without options grib_copy prints no key information
- Options exist to specify the set of keys to print
  - Use verbose option (`-v`) to print keys
- Output can be ordered
  - E.g. order by ascending or descending step
- Key values can be used to specify the output file names
- grib_copy fails if a key is not found
  - Use the `-f` option to force grib_copy not to fail on error

grib_copy – usage

```
grib_copy [options] grib_file grib_file … out_grib_file
```

- Options

`-p key1,key2,…`	Keys to print (only with `-v`)
`-w key=val1,key2!=val2,…`	Where option
`-B "key1 asc, key2 desc"`	Order by: "step asc, centre desc"
`-v`	Verbose
`-f`	Do *not* fail on error
`...`	

grib_copy – examples

- To copy only the analysis fields from a file

```
> grib_copy -w dataType=an in.grib1 out.grib1
```

- To copy only those fields that are not analysis fields

```
> grib_copy -w dataType!=an in.grib1 out.grib1
```

- Information can be output using the –v and –p options

```
> grib_copy -v -p shortName in.grib1 out.grib1
in.grib1
shortName
t
1 of 1 grib messages in in.grib1
1 of 1 total grib messages in 1 files
```

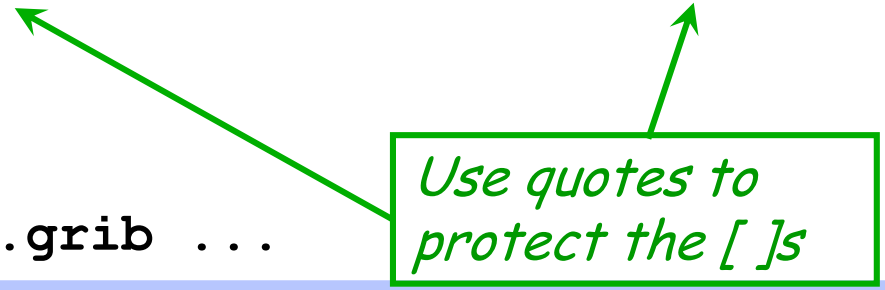
grib_copy – using key values in output file

- Key values can be used to specify the output file name

```
> grib_copy in.grib "out_[shortName].grib"

> ls out_*

out_2t.grib  out_msl.grib ...
```



- This provides a convenient way to filter GRIB messages into separate files

grib_set – set key / value pairs

- Use grib_set to
  - Set key / value pairs in the input GRIB file(s)
  - Make simple changes to key / value pairs in the input GRIB file(s)
- Each GRIB message is written to the output file
  - By default this includes messages for which no keys are changed
  - With `-s` (strict) option only messages matching all constraints in the where option are copied
- An option exists to repack data
  - Sometimes after setting some keys involving properties of the packing algorithm the data needs to be repacked
- grib_set fails when an error occurs
  - e.g. when a key is not found

grib_set – usage

```
grib_set [options] grib_file grib_file … out_grib_file
```

- Options

`-s key1=val1,key2=val2,…`	List of key / values to set
`-p key1,key2,,…`	Keys to print (only with `-v`)
`-w key1=val1,key2!=val2,…`	Where option
`-d value`	Set all data values to `value`
`-f`	Do *not* fail on error
`-v`	Verbose
`-S`	Strict
`-r`	Repack data
`...`	

grib_set – examples

- To set the parameter value of a field to 10m wind speed (10si)

```
> grib_set -s shortName=10si in.grib1 out.grib1
```

- This changes e.g.
  - shortName to 10si
  - paramId to 207
  - name / parameterName to ‘10 metre wind speed’
  - units / parameterUnits to ‘m s ** -1’

  - indicatorOfParameter to 207 GRIB edition dependent !
  - marsParam to 207.128

grib_set – examples

- Some keys are read-only and cannot be changed directly

```
> grib_set -s name="10 metre wind speed" in.grib1
  out.grib1

ECCODES ERROR   :  grib_set_values[0] name (3)
  failed: Value is read only
```

- The read-only keys can be set only by setting one of the other keys, e.g.
  - shortName=10si
  - paramId=207
  - indicatorOfParameter=207 GRIB edition dependent !

grib_set – set key values to missing

- When a key is not used all the bits of its value should be set to 1 to indicate that it is ‘missing’
- Different keys have different lengths so the value that needs to be coded for missing keys is not unique
- To set a key to missing a string "missing" or "MISSING" is accepted by grib_set

```
> grib_set -s Ni=missing in.grib2 out.grib2
```

- Note that some values cannot be set to “missing” !

```
> grib_set -s dataDate=missing file1.grib2 file2.grib2
ECCODES ERROR   :  unable to set dataDate=missing (Value cannot be missing)
ECCODES ERROR   :  grib_set_values[0] dataDate (7) failed: Value cannot be missing
```

grib_set – changing decimal precision

- To pack a temperature expressed in Kelvin with 1 digit of precision after the decimal point we can set changeDecimalPrecision=1
  - N.B. this is different to setting the number of significant digits !

```
> grib_set -s changeDecimalPrecision=1 T.grib1 T1.grib1
```

- Use grib_compare to see the differences

```
> grib_compare -c data:n T.grib1 T1.grib1
-- GRIB #1 -- shortName=2t paramId=167 stepRange=0 levelType=sfc level=0
   packingType=grid_simple gridType=reduced_gg --
double [packedValues]: 2132215 out of 2140702 different
 max absolute diff. = 5.0000000000011369e-02, relative diff. = 0.000207239
        max diff. element 17: 2.4121679687500000000e+02 2.4126679687500011369e+02
        tolerance=0.0000000000000000e+00 packingError: [0.000984192] [0.0500122]
        values max= [315.447]  [315.467]         min= [216.967] [216.967]
```

grib_set – modify data values

- An offset can be added to all data values in a GRIB message by setting the key offsetValuesBy

```
> grib_get -F "%.5f" -p max,min,average TK.grib
315.44727 216.96680 286.34257

> grib_set -s offsetValuesBy=-273.15 TK.grib TC.grib

> grib_get -F "%.5f" -p max,min,average TC.grib
42.29726 -56.18321 13.19257
```

grib_set – modify data values

- The data values in a GRIB message can be multiplied by a factor by setting the key scaleValuesBy

```
> grib_get -F "%.2f" -p max,min,average Z.grib
65035.92 -3626.08 2286.30

> grib_set -s scaleValuesBy=0.102 Z.grib1 orog.grib1

> grib_get -F "%.2f" -p max,min,average orog.grib1
6633.64 -369.86 233.20
```

grib_set – using key values in output file

- Key values can be used to specify the output file name

```
> grib_set -s time=0000 in.grib "out_[shortName].grib"


> ls out_*
out_2t.grib  out_msl.grib ...
```

- Remember: Use quotes to protect the [ ]s !

What cannot be done with grib_set

- grib_set cannot be used for making transformations to the data representation
  - It cannot be used to transform data from spectral to grid-point representation (and vice-versa)
- grib_set cannot be used transform data from one grid representation to another
  - It cannot be used to transform data from regular or reduced Gaussian grids to regular latitude-longitude grids
- grib_set cannot be used to select sub-areas of data
  - It will change the value of, e.g. latitudeOfFirstGridPointInDegrees etc, but the data will still be defined on the original grid

- The GRIB tools cannot be used to interpolate the data

grib_to_netcdf – convert to NetCDF

- Use grib_to_netcdf to convert GRIB messages to NetCDF
- Input GRIB fields must be on a regular grid
  - typeOfGrid=regular_ll or regular_gg
- Options allow user to specify
  - the NetCDF data type:
    - NC_BYTE, NC_SHORT, NC_INT, NC_FLOAT or NC_DOUBLE
    - NC_SHORT is the default
  - either classic (NetCDF 3) or NetCDF 4 file format
  - the reference date
    - default is 19000101
- Used in the MARS web interface and the public Data Servers to provide data in NetCDF files

grib_to_netcdf – usage

```
grib_to_netcdf [options] grib_file grib_file …
```

Options

-o output_file	Output netCDF file
-R YYYYMMDD	Use **YYYYMMDD** as reference date
-D NC_DATATYPE	NetCDF data type
-k kind	Kind of file to be created: 1 → netCDF classic file format 2 → netCDF 64 bit classic file format (Default) 3 → netCDF-4 file format 4 → netCDF-4 classic model file format
-T	Do not use time of validity.
-u dimension	Set **dimension** to be an unlimited dimension
-f	Do *not* fail on error
...	

grib_to_netcdf – examples

- To convert the fields in file.grib1 to NetCDF

```
> grib_to_netcdf -o out.nc file.grib1
grib_to_netcdf: Version 0.13.0
grib_to_netcdf: Processing input file 'file.grib1'.
grib_to_netcdf: Found 1 GRIB field in 1 file.
grib_to_netcdf: Ignoring key(s): method, type, stream, refdate, hdate
grib_to_netcdf: Creating netcdf file 'out.nc'
grib_to_netcdf: NetCDF library version: 4.3.2 of Oct 14 2014 14:34:41 $
grib_to_netcdf: Creating large (64 bit) file format.
grib_to_netcdf: Defining variable 't2m'.
grib_to_netcdf: Done.

> ls --size out.nc
160 out.nc
```

grib_to_netcdf – examples

- To convert the fields in file.grib1 to NetCDF with data type set to NC_FLOAT

```
> grib_to_netcdf -D NC_FLOAT -o out.nc file.grib1
grib_to_netcdf: Version 0.13.0
grib_to_netcdf: Processing input file 'surface.grib1'.
grib_to_netcdf: Found 1 GRIB field in 1 file.
grib_to_netcdf: Ignoring key(s): method, type, stream, refdate, hdate
grib_to_netcdf: Creating netcdf file 'out.nc'
grib_to_netcdf: NetCDF library version: 4.3.2 of Oct 14 2014 14:34:41 $
grib_to_netcdf: Creating large (64 bit) file format.
grib_to_netcdf: Defining variable 't2m'.
grib_to_netcdf: Done.

> ls --size out.nc
316 out.nc
```

Output NetCDF file is about twice the size

Practical: modifying GRIB messages

1. The file file1.grib contains parameters T and Z on five pressure levels.
   - Use grib_copy to create two files, one containing all the pressure levels for parameter T, the other for Z. Check the content of the new files with grib_ls
   - Repeat but output the messages so the levels in the new files are in increasing numerical order
2. Use grib_set to change the date and time to 00UTC on 1 March 2016 for all messages in file1.grib
   - Repeat but change the date and time for T at 500hPa only
   - Repeat so that T at 500hPa only is written to the output file
3. Use grib_to_netcdf to convert the GRIB messages in file2.grib to NetCDF.
   - Try with both the default data type (NC_SHORT) and NC_FLOAT.
   - Check the data values in each case with ncdump.
   - Repeat but set the Reference date to 28 February 2016 and compare with previous results
4. Use grib_to_netcdf to convert the GRIB messages in file3.grib to NetCDF.
   - What happens … and why ?

Extra practicals

5. Use grib_copy to split file1.grib into separate files for each parameter/level combination
   - Create files named t_500.grib1, z_500,grib1, etc
6. An SST field has been created by masking the Soil Temperature at Level 1 (STL1) with the Land-Sea Mask and is included with other messages in the file surface.grib
   - Use grib_set to change the parameter for the field from STL1 to SST and level type to ‘surface’
   - Be careful not to change the other parameters !
   - Repeat with each different message output to a separate file