Guidelines for the setup of a minimum-impact grib2 I/O in COSMO model

---

Introduction

A conservative approach is proposed: the variable table in `src_setup_vartab.f90` is kept as close as possible to its current status, i.e. containing information in grib1 form. In this way all the modifications are mainly confined to `src_input.f90` and `src_output.f90`, where some special grib2 procedures, similarly to the ones for grib1 and netcdf, will have to be added.

In the grib2 procedures, all the information to be coded in grib2 output will be derived from the grib1 information received, according to the guidelines indicated below, the opposite will happen for grib2 input.

In a second phase, a more native grib2 approach could be desirable, but this also depends on how long we will have to support grib1 I/O. In this sense ECMWF grib_api could help since it tries to hide, where possible, the differences between grib1 and grib2, thus reducing code duplication...

...however, no assumptions are made here about the grib2 library to be used.

Sections are numbered following grib2 documentation.

Section 1 - identification

Some additional information may be provided in the grib2 message:

- production status of grib data (code table 1.3) [operational | test | research], should be specified by namelist.
- type of processed data (code table 1.4) [analysis | forecast], could be implicitly derived from run configuration.

Section 3 - grid definition

The definitions of rotated grid (grid definition template 3.0), scanning mode, etc. do not change significantly. Are there plans for supporting other type of projections in COSMO?

Grid extremes and increments now can be coded with 10^{-6}° precision, which should be enough for most COSMO applications, so increments now should be given; hopefully the "basic angle of production domain" and "subdivision of basic angle" will not be used, they are mainly useful for specifying angles such as (1/60)° with infinite precision.

We have to keep in mind that, in grib2, longitudes are in the interval 0-360°, this may involve some extra computations mainly for grib2 input, output is quite straightforward.

Section 4 - product definition

Parameter

Conversion of grib1 table 2 version and parameter to grib2 discipline (section 0), parameter category and parameter number is too complex to be discussed here; hopefully a 1-1 mapping could be set up (with some exceptions indicated later in this section); a unit conversion is probably needed in this mapping process.

For satellite data, discipline 3 "space products", category 1 "quantitative products" appears more suitable than discipline 0 "meteorological products", category 4/5 "short wave radiation" and "long wave radiation", in order to emphasize the satellite perspective of the data, however the former category is quite empty at the moment.

Type of generating process

The entry "type of generating process" (code table 4.3) seems to contain some redundant information already specified in section 1 "type of processed data" ([analysis | forecast | observation]), however it should be possible to derive it implicitly from run configuration. This may be useful to encode the information that synthetic radar or satellite data are simulated and not observed.

Time range

Grib1 time range indicator translates into a product definition template (e.g. analysis/forecast at a fixed point in time or statistically processed value in time) + type of statistical processing performed:

grib1 time range indicator	grib2 product definition template	grib 2 type of statistical processing (code table 4.10)
0 (instantaneous value)	4.0	none
2 (value valid over an interval, used for max and min)	4.8	2 or 3 for max and min respectively
3 (value averaged over an interval)	4.8	0
4 (value accumulated over an interval)	4.8	1

In case of `TMAX_2M`, `TMIN_2M`, `VMAX_10M` (and possibly other) variables, these should be coded simply as temperature or wind speed, with the proper type of statistical processing (2 or 3), so some kind of ad-hoc dirty trick, involving also a change of parameter, is required here.

When using template 4.8, redundant information has to be coded in grib2: reference time (section 1, same as for grib1), forecast time (beginning of the statistical processing period = P1 for grib1), time of end of statistical processing period (= reference time + P2 for grib1) and duration of the statistical processing

period (= P2 - P1 for grib1).

Time range indicator 13 (nudging analysis, DWD extension) could possibly be replaced by using the correct template (4.0 or 4.8) and setting properly the "type of generating process" (code table 4.3) e.g. analysis, initialization, forecast, here 192-254 are reserved for local use, so we could think of some local COSMO convention, e.g. nudging, LHnudging, 1D-var, KENDA and their possible combinations.

In the case of analysis output, there is still the problem to correctly code the accumulated/averaged fields, since in grib2, as in grib1, it is not possible to specify a time interval starting in the past (i.e. a forecast time <0).

Additional time information can be provided in section 4:

- hours/minutes, after reference time, of data cutoff (end of AOF observation interval - start of run)
- time increment between successive fields used for statistical processing (model time step)

Vertical level

Grib 1 level type translates into two (one of which possibly missing) types of fixed surface:

grib1 level type	grib2 type of fixed surface(s) (code table 4.5)
1 (Earth surface)	1
2 (cloud base)	2
3 (cloud top)	3
4 (0° isotherm)	4
8 (top of atmosphere)	8
100 (isobaric level)	100
102 (mean sea level)	101
103 (height above mean sea level)	102
105 (height above ground)	103
109 (hybrid level)	105
110 (layer between two hybrid levels)	105 (x2)
111 (depth below land surface)	106
112 (layer between two levels below land surface)	106 (x2)
200 (entire atmosphere)	10
222 (satellite, DWD extension)	8 (top of atmosphere)?

Values should be properly rescaled and can assume virtually any reasonable order of magnitude and precision.

For moisture in multi layer soil model, `W_SO`, layers instead of levels can be used with any depth (if I well understand this is desirable), so they could be changed to layers (112) in `src_setup_vartab.f90`, in order to

keep the whole information, and an ad-hoc conversion to level could be done for grib1.

Synthetic satellite data

For satellite data simulated with RTTOV package, unfortunately the product definition template 4.31 "satellite product", where sensor information could be encoded, is not suitable because it does not contain the necessary time range information encodable in template 4.0.

Should we propose WMO to add a "simulated satellite data" template containing a combination of this information? Is any action already going on? It would be desirable to avoid using local extensions, since COSMO is not the only model which can produce synthetic satellite images.

Something similar holds for simulated radar data, if we want to encode the characteristics and location of the simulated radar antenna along with forecast information, template 4.20 is not complete (how is this done now with grib1?).

Vertical coordinate parameters

Vertical coordinate parameters do not change significantly, they are located in section 4 as "coordinate values" after the product definition template.

Section 5 - data representation

Template 5.0 "grid point data simple packing" or 5.40 "grid point data jpeg format" or 5.41 "grid point data png format" could be used in output, possibly chosen through namelist.

Conclusion

- Implementation of grib2 in COSMO is possible with a reasonable effort and limited drawbacks on the model stability.
- Grib2 could solve some issues which we currently face in grib1 messages with COSMO, e.g. precision of grid definition, use of non standard time range indicator.
- Some standard-conformance issues still remain in the coding of synthetic satellite and radar fields and in the coding of "analysed" precipitation and other accumulated or averaged quantities.
