CFSRR Known Issues

Technical Document – August 2011

By NCDC-NOMADS Team ~ verified by Wesley Ebisuzaki and Suranjana Saha (NCEP)

Various issues have been identified with the CFS Reanalysis and Reforecast. Proper interpretation of the GRIB2 metadata can be tricky in some cases. The purpose of this document is to clarify all known issues and provide insight on how to correctly interpret/workaround this metadata.

Notations and acronyms meanings:

$[t_i]$: Initialization time axis / run time / cycle / reference time
$[t_f]$: Forecast time / various representations depending on the context
$[t_v]$: Verification time −> the value $[t_i + t_f]$
CFSR: Climate Forecast System **Reanalysis** component
CFSRR: Climate Forecast System **Reforecast** component
MM: Monthly Mean

CFS Reanalysis Monthly Means

Metadata: Actual

PDS section 4 uses GRIB2 template 4.8 including local extensions, which may be misinterpreted by decoders unfamiliar with them. Ensure your decoder(s) are as up to date as possible when working with CFSR data.

The t_v encoded in Section 4 Octets 35-41 are zeroes.

The num_in_avg, or the count of the number of points (6hr cycles) averaged within the respective monthly average, is encoded in Section 4 Octet 53, possibly 32 bits: 50-53.

Section 4 Octet 58 (possibly 32 bits: 55-58) is the length of the averaging period per unit. For cycle fractions, this is 24, for complete monthly averages, it is 6.

The product of this and the num_in_avg {above} should always equal the total number of hours in a respective month.

Section 4 Octet 65 is the hours skipped between each calculation component.

The original form of these files, as output from NCEP computers, is GRIB1 format, many of these issues cropped up when the data was converted from GRIB1 to GRIB2 for considerable storage and transmission savings. Regardless of being in GRIB2 format, they still suffer deficiencies of GRIB1 file metadata.

Metadata: Producer Intention

The format of these filenames: <type>**AA**.gdas.<yyyymm>(.**BB**Z)?.grb2

AA: which can be nl or 00-06, is the averaging portion for each 6 hour cycle in the time series. The difference between nl and 00 is a few minutes. The 06 would represent a complete average. The intermediate hourly values (f01-f05) for averaged quantities should not be used as if they were instantaneous values at those hours. To obtain the hourly averages, for example the 2-3 hour average, it is necessary to subtract the 0-3 hr forecast values multiplied by 3 from the 0-2 hour forecast averages multiplied by 2, and so on. Only accumulation type fields can be reduced using straight subtraction.

BB: This, (if present), is the cycle fraction average, or the cycles involved in the **statistical** process.   If it is not present, then all 4 cycles per day are involved. If it is 06Z, for example, then the statistical data in the file are from ONLY the 06 UTC cycles during that month. These are mainly used for studying processes that occur during a portion of a normal diurnal cycle.

In summary, both the t_i and t_v are the month in the filename. The t_f is always 0. The length of the statistical processing is always 1 month. The form of the statistical process for that month can be determined using the AA and BB described above.

Data Section

There is a very subtle issue verified by Wesley Ebisuzaki in 2010 that exists in the reanalysis monthly means, for fields with bits per pixel (Section 5 Octet 20

under JPEG2000 compression template) greater than 16 bit precision. This fortunately is a very small percentage of the records. In the affected cases studied, the damage was limited to 1 point in the entire x/y field. While not a crippling problem, it needs to be taken into consideration as bad points can have a dramatic affect on average/min/max calculations. Be vigilant and check for unreasonable values.

CFS Reanalysis High Priority Time Series

#1)
This affected **only** data downloaded from NCDC-NOMADS prior to 22:00 UTC February 18, 2010.

A GRIB format conversion issue was verified affecting the metadata in non-instantaneous CFSR time series files (28 of the 73 file types affected are listed below). The issue, triggered during the conversion process between GRIB1 to GRIB2 by NCEP (to facilitate transmission and storage), caused the non-instantaneous fields (0-1, 1-2, 2-3 hr statistical) to appear instantaneous (0-0, 1-1, 2-2 hr statistical) when decoded. At 22:00 UTC February 18 2010, NCDC will begin bringing corrected versions of these files online. Users who have already downloaded these files are encouraged to re-download
these files on/after Friday, February 19th, otherwise be aware of
these facts when using your existing copy. If you improvise the metadata, the data itself is valid.

The issue affects the following file types (28):
dlwsfc dswsfc gflux lhtfl ocndt20c ocnheat ocnmld ocnsal15
ocnsal5 ocnslh ocnsst ocnt15 ocnu15 ocnu5 ocnv15 ocnv5
ocnvv55 prate runoff shtfl snohf tmax tmin ulwsfc ulwtoa
uswsfc uswtoa wndstrs

#2)

In the original February 01, 2010 distribution of the CFSR Time Series, at NCEP's request, NCDC-NOMADS released a copy of the CFSR with the f00 records removed. Afterwards we discovered issues with this removal process and some records which should remain in these files were removed. As of 15:15 UTC on February 08, 2010, this dataset has been replaced with the original unmodified dataset sent by NCEP. The forecast at the first time step (f00) of several minutes constitutes a spin up of the model physics, and extreme care should be taken when using it as a proxy of any type of validation.
IT IS NOT THE ANALYSIS.

For more details, see the pdf from NCEP:
http://nomads.ncdc.noaa.gov/docs/CFSR-Timeseries.pdf

CFS Reanalysis Hourly Files : Particular Files/GRIB Records

http://nomads.ncdc.noaa.gov/modeldata/   (Aug 2011)

 There are temperature at the surface (tmpsfc) records in the ipv*nl files. These records should not be used, more appropriate records can be found in the flx* files.

The ocnh* and ocnf* group contains a variable named "Geometric depth below sea surface" which **may** have their isothermal level mislabeled as all zeroes for all the level values. In cases where this is encountered, the zeroes correspond to (in record order): 2.5C, 5.0C, 10.0C, 15.0C, 20.0C, 25.0C, and 28.0C isotherms.

CFS Reforecast Monthly Means (MM daily avgs. & MM time series)

Metadata : Actual

PDS Octet 35-41 which supposedly holds the verification time is set incorrectly to t_i values.
Any program or human interpreting this header strictly as the t_v will be mislead.

Metadata : <u>Producer Intention</u>

The first forecast step (t_f = 0) should range from t_i to the end of that month, then all future t_f are whole months (day 1 through 28-31), not exact increments of t_i + X month.

For the first forecast step, GRIB2 Section 4 Octet 18, Indicator of unit of time range – should be set to 2 (days) with the corresponding number of days to close at the end of the t_i month. For other steps, it may be alright to keep the time frame in whole months assuming the t_v was properly set to the end of the month.

CFS Reforecast High Priority sets

The issue stated above in the CFS Reanalysis hourly files section regarding the tmpsfc in the ipv and the ocn isothermal levels also applies to the reforecast high priority sets where these file types exist.

Similar results are documented and can be found at the
CISL Research Data Archive (08/2011)

http://dss.ucar.edu/datasets/ds093.0/docs/FAQs_6hrly.html

Appendix A
GRIB2 Template

For reference purposes, this is a copy of the GRIB2 template used for statistically processed fields used in reanalysis and reforecast monthly means.

Product Definition Template 4.8 - Average, accumulation, and/or extreme values or other statistically processed values at a horizontal level or in a horizontal layer in a continuous or non-continuous time interval

Octet No.	Contents
10	Parameter category (see Code Table 4.1)
11	Parameter number (see Code Table 4.2)
12	Type of generating process (see Code Table 4.3)
13	Background generating process identifier (defined by originating center)
14	Analysis or forecast generating processes identifier (defined by originating center)
15-16	Hours after reference time of data cutoff (see Note 1, below)
17	Minutes after reference time of data cutoff
18	Indicator of unit of time range (see Code Table 4.4)
19-22	Forecast time in units defined by octet 18
23	Type of first fixed surface (see Code Table 4.5)
24	Scale factor of first fixed surface
25-28	Scaled value of first fixed surface
29	Type of second fixed surface (see Code Table 4.5)
30	Scale factor of second fixed surface
31-34	Scaled value of second fixed surface
35-36	Year of time of end of overall time interval
37	Month of time of end of overall time interval
38	Day of time of end of overall time interval
39	Hour of time of end of overall time interval
40	Minute of time of end of overall time interval
41	Second of time of end of overall time interval
42	n - Number of time range specifications describing the time intervals used to calculate the statistically processed field

43-46	Total number of data values missing in statistical process

47-58 Specification of the outermost (or only) time range over which statistical processing is done

47	Statistical process used to calculate the processed field from the field at each time increment during the time range (see Code Table 4.10)
48	Type of time increment between successive fields used in the statistical processing (see Code Table 4.11)
49	Indicator of unit of time for time range over which statistical processing is done (see Code Table 4.4)
50-53	Length of the time range over which statistical processing is done, in units defined by the previous octet
54	Indicator of unit of time for the increment between the successive fields used (see Code Table 4.4)
55-58	Time increment between success fields, in units defined by the previous octet (see Notes 3 and 4, below)

*59-nn These octets are included only if n > 1, where nn = 46 + 12 * n*

59-70	As octets 47 to 58, next innermost step of processing
71-nn	Additional time range specifications, included in accordance with the value of n. Contents as octets 47 to 58, repeated as necessary

Notes:

(1) Hours greater than 65534 will be coded as 65534.
(2) The reference time in section 1 and the forecast time together define the beginning of the overall time interval.
(3) An increment of zero means that the statistical processing the result of continuous (or near continuous) process, not the processing of a number of discrete samples. Examples of such continuous process are the temperatures measured by analogue maximum and minimum thermometers or thermographs and the rainfall measured by a rain gauge.
(4) The reference and forecast times are successively set to their initial plus or minus the increment, as defined by the type of time increment (one of octets 48,60,72...). For all but the innermost (last) time range, the next inner range is then processed using these reference and forecast times as the initial reference and forecast time.