GRIB naming in CDM

John Caron and Ethan Davis

4/2/2012

Outline

- GRIB background and issues
- Questions
  - What is best naming scheme for GRIB / netCDF library ?
  - What is best way for applications to present variable selections to users ?
  - What to do about backwards compatibility?
- Options

GRIB Background

- WMO standard for gridded meteorological data
- NCEP uses exclusively for transmitting model output
- All IDD model data is in GRIB
- GEMPAK converts to GEMPAK format with hand-maintained table (NCEP only?)
- CDM aspires to be general purpose GRIB reader
- IDV reads GRIB through the CDM library

GRIB in CDM

- Problems in GRIB discovered 2010 (CDM 4.2.4)
  - Time interval coordinates – affected 25% NCEP
  - NCEP local tables were always used (GRIB2)
  - Many errors in local tables use (esp GRIB1)
  - Mistakes in standard WMO tables
  - Variable naming algorithm was flawed
  - Etc.
- NCDC $$ for serving large collection of GRIB
  - eg Climate Forcast System Reanalysis (CFSR)
  - eg hpr-ts45 contains 1.2M files, 250M records
- Complete rewrite of GRIB for TDS 4.3
  - Complete review of all things GRIB

GRIB Issues Summary

1. GRIB does not encode the “dataset schema”
   – No unique identifier for variables
2. GRIB tables are serious problem
   – No canonical GRIB tables
   – Inconsistent use of local tables
   – No foolproof way of knowing which tables were used when writing the GRIB file
   – GRIB parameter names are not required to be unique, short or stable.

No “dataset schema”

- GRIB data model is an unordered collection of 2D (horiz) slices. Each GRIB record stands alone.
  - There is no way for a data provider to describe the dataset schema = “ncdump –h” (show netCDF header)
- To create netCDF multidimensional data model:
  - Decide which records belong in a variable
  - Construct time, vert, ensemble coordinates

No unique variable identifier

- A GRIB record has a collection of attributes
  - Parameter (discipline / category / number)
  - Level Type (pressure, surface, pressure layers, etc)
  - Level Value(s)
  - Base Time (typically the model run time)
  - Forecast Time type (instantaneous or interval)
  - Forecast Time value(s)
  - Background Generating Process, Forecast Generating Process, Ensemble derived type, Probability type, …
  - Etc.
- GRIB2 has ~30 PDS templates, each with 10-20 attributes
- To create netCDF data model
  - Decide which attributes from which templates are used to create unique variables
  - See if that works on as many datasets as possible

GRIB names in GFS (partial list)

Latent heat net flux (Mixed_intervals Average) @ Ground or water surface	64,361,720	time1,lat,lon
Sensible heat net flux (Mixed_intervals Average) @ Ground or water surfa...	64,361,720	time1,lat,lon
Specific humidity @ Specified height level above ground	65,1,361,720	time,height_above_grou
Specific humidity @ Level at specified pressure difference from ground to...	65,1,361,720	time,pressure_differenc
Relative humidity @ Level of 0°C isotherm	65,361,720	time,lat,lon
Relative humidity @ Isobaric surface	65,25,361,720	time,pressure3,lat,lon
Relative humidity @ Specified height level above ground	65,1,361,720	time,height_above_grou
Relative humidity @ Sigma level layer	65,4,361,720	time,sigma_layer,lat,lon
Relative humidity @ Sigma level	65,1,361,720	time,sigma,lat,lon
Relative humidity @ Level at specified pressure difference from ground t...	65,1,361,720	time,pressure_differenc
Relative humidity @ Entire atmosphere layer	65,361,720	time,lat,lon
Relative humidity @ Highest tropospheric freezing level	65,361,720	time,lat,lon
Precipitable water @ Entire atmosphere layer	65,361,720	time,lat,lon
Precipitation rate (Mixed_intervals Average) @ Ground or water surface	64,361,720	time1,lat,lon
Total precipitation (Mixed_intervals Accumulation) @ Ground or water sur...	65,361,720	time2,lat,lon
Convective precipitation (Mixed_intervals Accumulation) @ Ground or wat...	65,361,720	time2,lat,lon
Water equivalent of accumulated snow depth @ Ground or water surface	65,361,720	time,lat,lon
Cloud mixing ratio @ Isobaric surface	65,21,361,720	time,pressure1,lat,lon

GRIB Parameter Tables

Parameter == (discipline / category / number bytes)

- Look up in an external table, either WMO standard table or a local table

No canonical machine-readable GRIB parameter tables

- WMO publishes in MS Word format (recently also started publishing GRIB2 tables in XML)
- Some mistakes and inconsistencies in standard
- Other mistakes and variations from hand-transcribing
- There are no 2 identical copies of WMO tables anywhere

Inconsistent use of local tables

No foolproof way of knowing which tables were used when writing the GRIB file

On the suitability of BUFR and GRIB for archiving data

http://www.unidata.ucar.edu/staff/caron/papers/GRIBarchivals.pdf

Official GRIB-2 tables (pdf)

Product Discipline 2: Land surface products, Parameter Category 0: Vegetation/Biomass

Number	Parameter	Units
0	Land cover (1=land, 0=sea)	Proportion
1	Surface roughness	m
2	Soil temperature	K
3	Soil moisture content*	$kg\ m^{-2}$
4	Vegetation	%
5	Water runoff	$kg\ m^{-2}$
6	Evapotranspiration	$kg^{-2}\ s^{-1}$
7	Model terrain height	m
8	Land use	code table (4.212)
9	Volumetric soil moisture content**	Proportion
10	Ground heat flux*	$W\ m^{-2}$
11	Moisture availability	%
12	Exchange coefficient	$kg\ m^{-2}\ s^{-1}$
13	Plant canopy surface water	$kg\ m^{-2}$
14	Blackadar's mixing length scale	m
15	Canopy conductance	$m\ s^{-1}$
16	Minimal stomatal resistance	$s\ m^{-1}$
17	Wilting point**	Proportion
18	Solar parameter in canopy conductance	Proportion
19	Temperature parameter in canopy conductance	Proportion
20	Soil moisture parameter in canopy conductance	Proportion
21	Humidity parameter in canopy conductance	Proportion
22	Soil moisture	$kg\ m^{-3}$
23	Column-integrated soil water	$kg\ m^{-2}$
24	Heat flux	$W\ m^{-2}$

center_id	subcenter_id	version	center
0	-1	-1	WMO standard table
7	-1	128	US National Weather Service - National Centres for Environmental Prediction (NCEP)
7	-1	129	US National Weather Service - National Centres for Environmental Prediction (NCEP)
7	-1	130	US National Weather Service - National Centres for Environmental Prediction (NCEP)
7	-1	131	US National Weather Service - National Centres for Environmental Prediction (NCEP)
7	-1	133	US National Weather Service - National Centres for Environmental Prediction (NCEP)
7	-1	140	US National Weather Service - National Centres for Environmental Prediction (NCEP)
7	-1	141	US National Weather Service - National Centres for Environmental Prediction (NCEP)
7	-1	-1	US National Weather Service - National Centres for Environmental Prediction (NCEP)
9	-1	128	US National Weather Service - Other
57	-1	2	US Air Force - Air Force Global Weather Central
58	42	2	Fleet Numerical Meteorology and Oceanography Center, Monterey, CA, USA
58	-1	-1	Fleet Numerical Meteorology and Oceanography Center, Monterey, CA, USA
60	255	2	United States National Center for Atmospheric Research (NCAR)
74	-1	174	UK Meteorological Office - Exeter (RSMC)
74	-1	175	UK Meteorological Office - Exeter (RSMC)
80	-1	201	Rome (RSMC)
85	-1	128	Toulouse (RSMC)
98	-1	128	European Centre for Medium-Range Weather Forecasts (ECMWF) (RSMC)
98	-1	129	European Centre for Medium-Range Weather Forecasts (ECMWF) (RSMC)
98	-1	130	European Centre for Medium-Range Weather Forecasts (ECMWF) (RSMC)
98	-1	132	European Centre for Medium-Range Weather Forecasts (ECMWF) (RSMC)
98	-1	133	European Centre for Medium-Range Weather Forecasts (ECMWF) (RSMC)
98	-1	140	European Centre for Medium-Range Weather Forecasts (ECMWF) (RSMC)
98	-1	150	European Centre for Medium-Range Weather Forecasts (ECMWF) (RSMC)

number	WMOdesc	description	
55	Vapor pressure	Vapor pressure	water_vapor_p
56	Saturation deficit	Saturation deficit	water_vapor_s
57	Evaporation	Evaporation	water_evapora
58	Cloud ice	Cloud ice	atmosphere_c
59	Precipitation rate	Precipitation rate	precipitation_fl
60	Thunderstorm probability	Thunderstorm probability	thunderstorm_
61	Total precipitation	Total precipitation	precipitation_a
62	Large scale precipitation	Large scale precipitation	large_scale_p
63	Convective precipitation	Convective precipitation	convective_pre
64	Snowfall rate water equivalent	Snowfall rate water equivalent	snowfall_flux
65	Water equivalent of accumulated snow ...	Water equivalent of accumulated snow depth	surface_snow_
66	Snow depth	Snow depth	surface_snow_
67	Mixed layer depth	Mixed layer depth	ocean_mixed_
68	Transient thermocline depth	Transient thermocline depth	
69	Main thermocline depth	Main thermocline depth	
70	Main thermocline anomaly	Main thermocline anomaly	
71	Total cloud cover	Total cloud cover	cloud_area_fra
72	Convective cloud cover	Convective cloud cover	convective_clo
73	Low cloud cover	Low cloud cover	
74	Medium cloud cover	Medium cloud cover	
75	High cloud cover	High cloud cover	
76	Cloud water	Cloud water	atmosphere_c
77	Best lifted index to 500 hPa	Best lifted index to 500 hPa	
78	Convective snow	Convective snow	convective_sn

```
number=65, name='null', description='Water equivalent of accumulated snow depth', unit='kg.m-2'} from WMO_GRIB1.xml (-1
number=65, name='WEASD', description='Water equiv of accum snow depth', unit='kg/m2'} from ncepGrib1-2.xml (-920240846)
number=65, name='ttl_snow', description='The estimated depth of the snow water equivalent that accumulated over a fixed
ccumulation period ends a snap-shot is taken grid made and then the accumulation is set to zero the bucket tip The accu
me for each grid Grids are available at forecast hours that are multiples of the accumulation period One or more accumu
grids for each accumulation period for that forecast hour If period attribute is specified the response will contain da
lation period is every 3 hours For NOGAPSEnsemble the accumulation period is every 6 hours', unit='kg/m2'} from US058MM
number=65, name='None', description='convective mass flux', unit='kg.s-1.m-2'} from local_table_2_version_201 (-1599351
number=65, name='-', description='Skin temperature difference', unit='K'} from ECMWF local table 2: Version Number 128
number=65, name='-', description='Skin temperature difference', unit='K'} from ECMWF local table 2: Version Number 129
number=65, name='TPRL1', description='Total precipitation rate less than 1 mmday', unit='%'} from ECMWF local table 2:
number=65, name='-', description='Skin temperature difference', unit='K'} from ECMWF local table 2: Version Number 171
number=65, name='-', description='Skin temperature difference', unit='K'} from ECMWF local table 2: Version Number 200
number=65, name='-', description='convective mass flux', unit='kg.s-1.m-2'} from ECMWF local table 2: Version Number 20
number=65, name='TCCH4', description='Total column Methane', unit='kg.m-2'} from ECMWF local table 2: Version Number 21
number=65, name='SF', description='Water equivalent of accumulated snow depth', unit='kg.m-2'} from WMO standard table
number=65, name='SF', description='Water equivalent of accumulated snow depth', unit='kg.m-2'} from WMO standard table
number=65, name='SF', description='Water equivalent of accumulated snow depth', unit='kg.m-2'} from WMO standard table
number=65, name='SF', description='Water equivalentof accumulated snow depth', unit='kg.m-2'} from WMO standard table 2
number=65, name='SF', description='Water equivalentof accumulated snow depth', unit='kg.m-2'} from Table 2 standard OMM
number=65, name='WENV', description='WAT EQUIV ACC SNOW DEPTH', unit='kg/m2'} from cptec_254_gtb.h (363285314)
number=65, name='W_SNOW', description='water equivalent of accumulated snow depth', unit='kg/(m2)'} from dwd_002_gtb.h
number=65, name='MASS_FL_CO', description='convective mass flux', unit='kg/(s.m2)'} from dwd_201_gtb.h (-993837869)
number=65, name='VIO3', description='total vertically integrated ozone content', unit='Pa'} from dwd_202_gtb.h (1272743
number=65, name='SN_DEPWE_S', description='water equivalent of accum snow depth smoothed', unit='kg/(m2)'} from dwd_206
number=65, name='SN_DEPWE_C', description='water equivalent of accum snow depth calibrated', unit='kg/(m2)'} from dwd_2
number=65, name='', description='Skin temperature difference', unit='K'} from ecmwf_128_gtb.h (531848967)
number=65, name='', description='Skin temperature difference', unit='K'} from ecmwf_129_gtb.h (-1581303727)
number=65, name='TPRL1', description='Total precipitation rate less than 1 mm per day', unit='%'} from ecmwf_131_gtb.h
number=65, name='HCCPG30', description='High Cloud Cover probability greater than 30%', unit='%'} from ecmwf_133_gtb.h
number=65, name='VIEMF', description='Vertical integral of eastward mass flux', unit='kg.m-1.s-1'} from ecmwf_162_gtb.h
number=65, name='', description='Skin temperature difference', unit='K'} from ecmwf_171_gtb.h (-1943412381)
number=65, name='', description='Skin temperature difference', unit='K'} from ecmwf_200_gtb.h (70217622)
number=65, name='', description='convective mass flux', unit='kg.s-1.m-2'} from ecmwf_201_gtb.h (-1272748388)
number=65, name='TCCH4', description='Total column Methane', unit='kg.m-2'} from ecmwf_210_gtb.h (1167124613)
number=65, name='WEASD', description='water equivalent of accumulated snow depth', unit='kg/m/m'} from fnmoc_gtb.h (596
number=65, name='WEASD', description='Water equivalent of accumulated snow depth', unit='kg/m2'} from fsl0_gtb.h (-1162
number=65, name='WEASD', description='Water equivalent of accumulated snow depth', unit='kg/m2'} from fsl1_gtb.h (-1168
number=65, name='WEASD', description='Water equivalent of accumulated snow depth', unit='kg/m2'} from fsl2_gtb.h (-1168
number=65, name='SNWE', description='Water equivalent snow depth', unit='m'} from jma_3_gtb.h (290430085)
```

Proposed BUFR/GRIB Table registration

- Registered users can upload BUFR/GRIB tables
  - Unique id is assigned (MD5 16 byte checksum?)
  - Convince producers to include the id into the data – unambiguous which table was used
  - Anyone can download.
- Reference GRIB and BUFR Decoding
  - Using CDM – find bugs !
- Could be Unidata developed web service
  - Turn over to WMO if they want it
- Survival of Human Race is at stake here

Question: What is best variable naming scheme for a general GRIB reader?

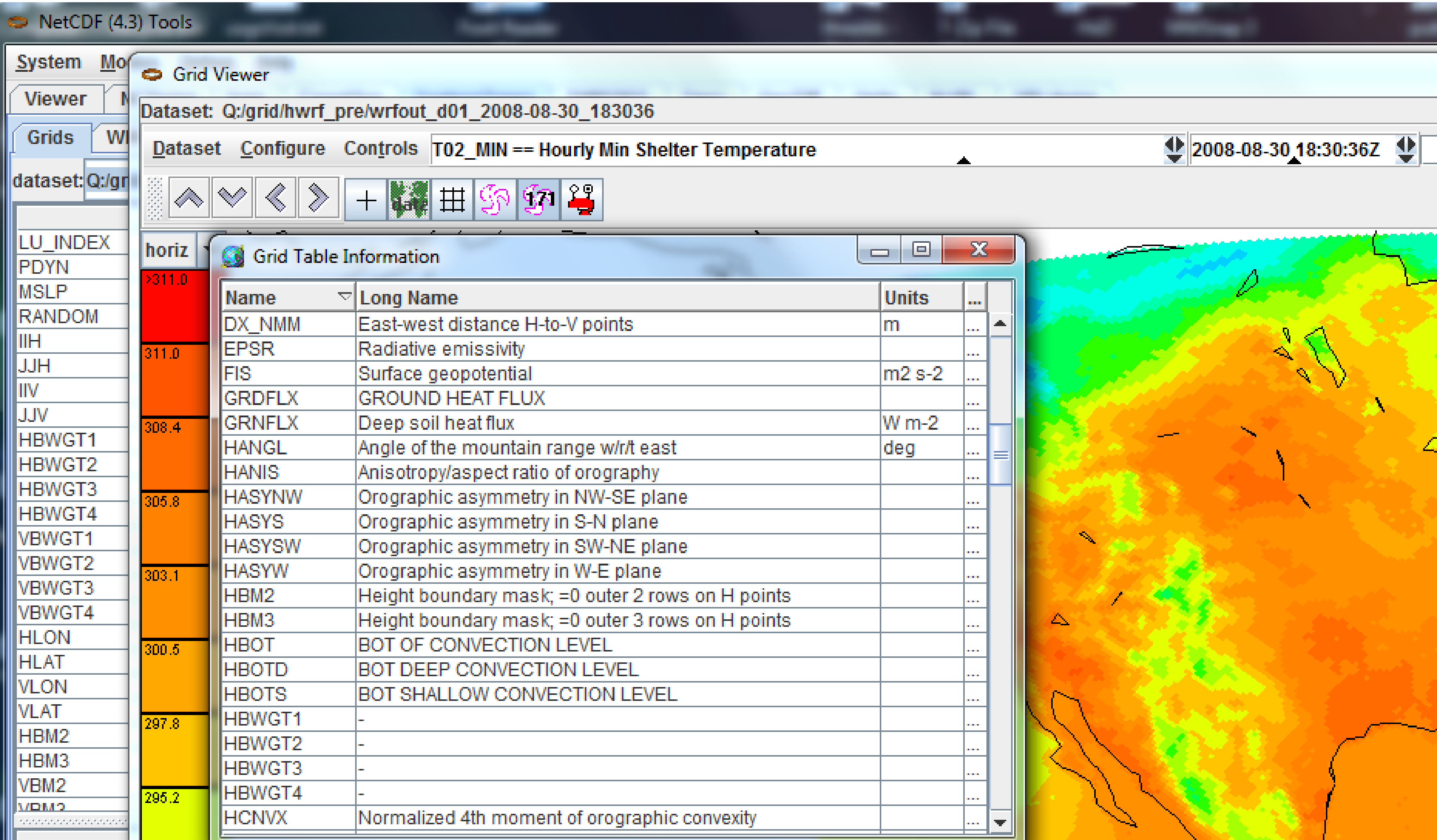
- Variable names have to be unique, not too long, and stable
- GRIB parameter tables are not
- Option: hand maintained tables
  - Doesn’t scale, could only be done for a subset, eg NCEP IDD model data
- Option: seperate variable names from descriptions
  - Generate variable names from just the records, not the external tables
  - Generate descriptions from the external tables
  - NCL has chosen a similar path to this solution

e	name	description	
	time3_bounds	bounds for time3	1,2
	VAR_0-0-0_L1	Temperature @ Ground or water surface	1,2
	VAR_0-0-0_L7	Temperature @ Tropopause	1,2
	VAR_0-0-0_L100	Temperature @ Isobaric surface	1,3
	VAR_0-0-0_L103	Temperature @ Specified height level above ground	1,2
	VAR_0-0-0_L105	Temperature @ Hybrid level	1,1
	VAR_0-0-0_L108_layer	Temperature @ Level at specified pressure difference from ground to lev...	1,6
	VAR_0-0-3_L1	Pseudo-adiabatic potential temperature or equivalent potential temperatu...	1,2
	VAR_0-0-6_L103	Dew-point temperature @ Specified height level above ground	1,1
	VAR_0-1-0_L103	Specific humidity @ Specified height level above ground	1,2
	VAR_0-1-0_L105	Specific humidity @ Hybrid level	1,1
	VAR_0-1-1_L4	Relative humidity @ Level of 0°C isotherm	1,2
	VAR_0-1-1_L100	Relative humidity @ Isobaric surface	1,3
	VAR_0-1-1_L103	Relative humidity @ Specified height level above ground	1,1
	VAR_0-1-1_L108_layer	Relative humidity @ Level at specified pressure difference from ground to...	1,6
	VAR_0-1-1_L204	Relative humidity @ Highest tropospheric freezing level	1,2
	VAR_0-1-3_L200	Precipitable water @ Entire atmosphere layer	1,2
	VAR_0-1-7_L1	Precipitation rate @ Ground or water surface	1,2
	VAR_0-1-9_L1_Imixed_S1	Large-scale precipitation (non-convective) (Mixed_intervals Accumulation)...	3,2
	VAR_0-1-10_L1_Imixed_S1	Convective precipitation (Mixed_intervals Accumulation) @ Ground or wat...	3,2
	VAR_0-1-11_L1	Snow depth @ Ground or water surface	1,2
	VAR_0-1-13_L1	Water equivalent of accumulated snow depth @ Ground or water surface	1,2
	VAR_0-1-13_L1_I2_Hour_S1	Water equivalent of accumulated snow depth (2_Hour Accumulation) @ G...	1,2
	VAR_0-1-22_L100	Cloud mixing ratio @ Isobaric surface	1,3
	VAR_0-1-22_L105	Cloud mixing ratio @ Hybrid level	1,1
	VAR_0-1-24_L100	Rain mixing ratio @ Isobaric surface	1,3
	VAR_0-1-25_L100	Snow mixing ratio @ Isobaric surface	1,3
	VAR_0-1-32_L100	Graupel (snow pellets) @ Isobaric surface	1,3
	VAR_0-1-192_L1	Categorical Rain @ Ground or water surface	1,2

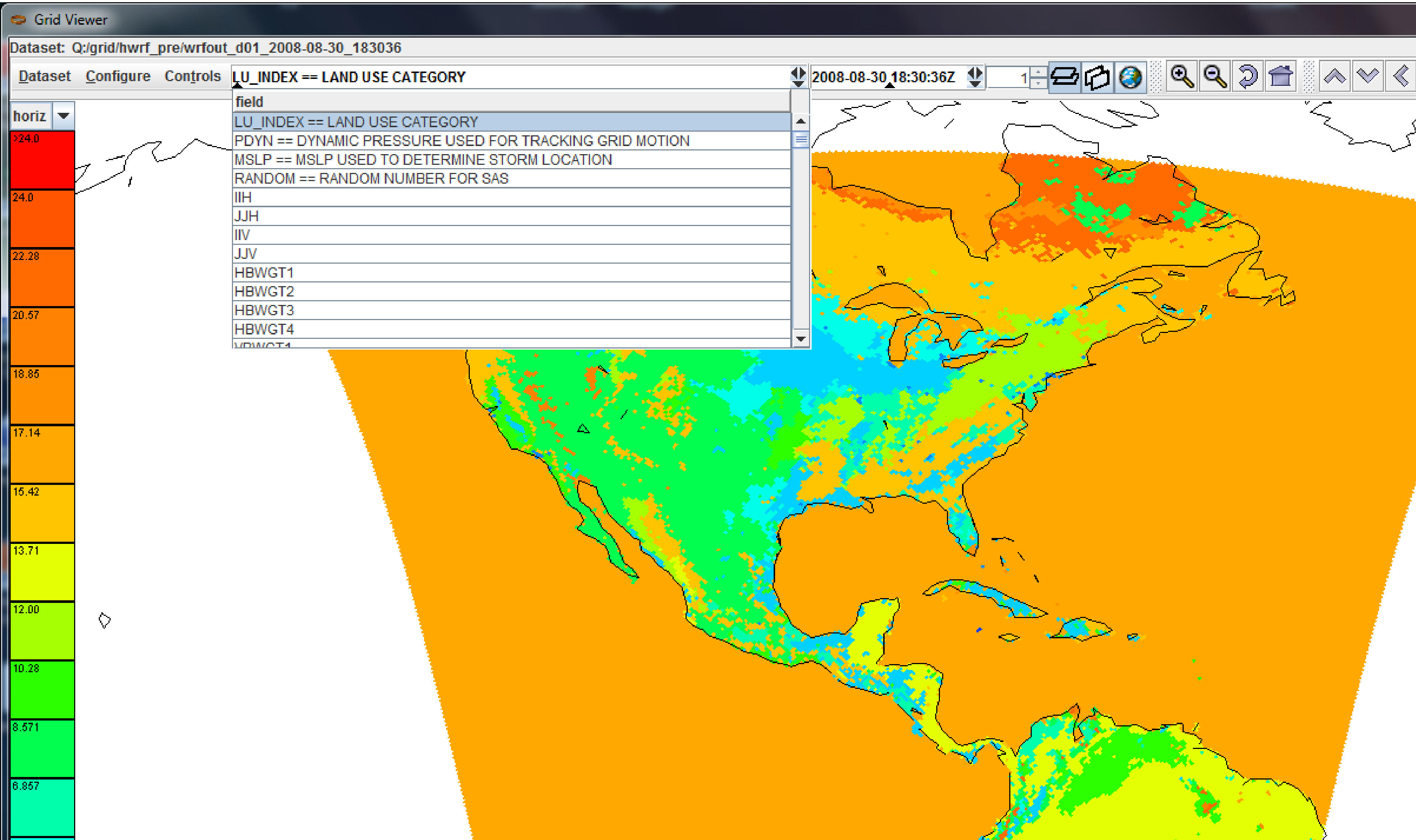
Mistake in CDM 4.2 variable naming

Temperature	Temperature @ pressure	65,26,361,720
Temperature_altitude_above_msl	Temperature @ altitude_above_msl	65,3,361,720
Temperature_depth_below_surface_la...	Temperature @ depth_below_surface_layer	65,4,361,720
Temperature_height_above_ground	Temperature @ height_above_ground	65,1,361,720
Temperature_high_cloud_top	Temperature_Average (Average for Mixed Intervals) @ high_cloud_top	64,361,720
Temperature_low_cloud_top	Temperature_Average (Average for Mixed Intervals) @ low_cloud_top	64,361,720
Temperature_maximum_wind	Temperature @ maximum_wind	65,361,720
Temperature_middle_cloud_top	Temperature_Average (Average for Mixed Intervals) @ middle_cloud_top	64,361,720
Temperature_potential_vorticity_surface	Temperature @ potential_vorticity_surface	65,2,361,720
Temperature_pressure_difference_layer	Temperature @ pressure_difference_layer	65,1,361,720
Temperature_sigma	Temperature @ sigma	65,1,361,720
Temperature_surface	Temperature @ surface	65,361,720
Temperature_tropopause	Temperature @ tropopause	65,361,720
Total_cloud_cover	Total_cloud_cover_Average (Average for Mixed Intervals) @ boundary_la...	64,361,720
Total_cloud_cover_convective_cloud	Total_cloud_cover @ convective_cloud	65,361,720
Total_cloud_cover_entire_atmosphere	Total_cloud_cover_Average (Average for Mixed Intervals) @ entire_atmo...	64,361,720
Total_cloud_cover_high_cloud	Total_cloud_cover_Average (Average for Mixed Intervals) @ high_cloud	64,361,720
Total_cloud_cover_low_cloud	Total_cloud_cover_Average (Average for Mixed Intervals) @ low_cloud	64,361,720
Total_cloud_cover_middle_cloud	Total_cloud_cover_Average (Average for Mixed Intervals) @ middle_cloud	64,361,720
Total_ozone	Total_ozone @ entire_atmosphere	65,361,720
Total_precipitation	Total_precipitation_Accumulation (Accumulation for Mixed Intervals) @ s...	65,361,720
U-Component_Storm_Motion	U-Component_Storm_Motion @ height_above_ground_layer	65,1,361,720
U-component_of_wind	U-component_of_wind @ pressure	65,26,361,720
U-component_of_wind_altitude_above...	U-component_of_wind @ altitude_above_msl	65,3,361,720

Question : What is best way for applications to present variable selections to users?



Answer : Both variable name and description must be used



Question: What to do about backwards compatibility?

- ~ 20 % of variable names have to change in order to fix the “too clever” naming algorithm
- Option: break 20%, create maps to the old names and do a translation, hand maintain tables so nothing ever changes
- Option: break everything at once, create tools to translate bundles (etc) to new names once

Reality Check

- Variable names (GRIB parameter names, WRF model output, etc) will continue to change in the future
- Applications have to be able to gracefully deal with change (especially applications that use web resources)
- Can't depend on variable names being meaningful in netCDF files

Technical Debt

“Shipping code is like going into debt. A little debt speeds development so long as it is paid back promptly with a rewrite...

“The danger occurs when the debt is not repaid. Every minute spent on not-quite-right code counts as interest on that debt.

“Entire engineering organizations can be brought to a stand-still under the debt load of an unconsolidated implementation”

Ward Cunningham

Technical Debt at Unidata

- Code is difficult to maintain/change except by the original programmers.
  - *Bring new people on, give them ownership, refactor*
- Build is brittle, cannot easily be replicated on another machine
  - *Switching to maven for standard builds*
- Bundles (etc) cant tolerate changes in the referenced datasets (URLs, names, etc)
  - *Create tools to gracefully transition bundles*

"all software dies when it becomes impossible to change without breaking something"

Conclusion

- Use of variables' names from GRIB records alone is ugly but are stable, short and unique
- Put information from GRIB tables into variable's descriptions
- Applications must use both names and descriptions when presenting selections to users
- Creating tools to help IDV bundles change gracefully would be a real benefit now and in the future, and would be part of a program of paying down Unidata technical debt