Proceedings of the *Open source GIS − GRASS user conference 2002* − Trento, Italy, 11−13 September 2002

Support of WMO binary format (BUFR and GRIB)

Jean Claude Bergès

PRODIG, Université Paris 1, 191 rue St Jacques, 75005 Paris, France, E−mail:zebulon@univ−paris1.fr

1 Introduction

Meteorological phenomenon ignores political boundaries and open exchange of observation data is a main issue for meteorological community. In 1951, the newly created World Meteorological Organisation (WMO) established the Global Telecommunication System (GTS) a world−wide network dedicated to environmental data exchange [1]. This first network used 50 bps telegraphic lines and Baudot (5 bits) character codes. Associated formats were defined to code various observation parameters (ground pressure, temperature, wind speed, etc.). Most of these formats are still used, although Baudot has been replaced by ASCII and telegraph by Internet links.

New formats have been added to match with evolution of observation system. As long as only ground observations were transmitted, message size was small and formats were tailored for human interpretation. But with the broadcast of satellite and weather forecast numeric model data, data flow volume has increased and binary formats have been defined to optimise data transmission. These large data sets are important for climatological studies as unlike station observations, they are space homogeneous and quality controlled.

It must be stressed that, even for satellite imager, WMO formats are not image formats as they contain the whole information necessary for data interpretation. Binary format specifications are very large containing rare or unused features. Furthermore there is no warranty that the full set of specifications is coherent and local extensions are authorised. Thus, a realistic analyse and interpretation program will be designed only to support the most useful and widespread data sources.

2 Why ingesting binary formatted data in GIS ?

As broadcasted data flow was becoming more important and more complex, software tools were developed to help meteorologist to integrate different type of data. These tools often running on dedicated forecaster workstation are in fact meteorology specialised GIS. They integrate different information layers, perform some thermodynamic computation and also support specific meteorological graphical convention (frontal lines, wind array etc.).

Many forecaster workstation systems are available under different licences: *Magic* (ECMWF), *Synergie* (Météo−France), *GRADS* (IGES) [2] and *MacIdas* (University of Wisconsin). We have to mention Unidata consortium, which delivers free code for decoding programs [4].

As long as we are only using meteorological data these software packages are obviously the best choice. But should we merge these data with any land related information, we cannot avoid ingestion in a general use GIS. A typical situation is damage assessment in front of natural disaster. Meteorology inform us about environmental parameters. But to perform a fine impact evaluation, land use information has to be introduced.

Although data coded in WMO format could be converted in standard image format by a software like GRADS and then ingested in a GIS, it will be far most efficient to directly support these formats as GIS input. User interface will be more simple and some conversion errors will be avoided.

3 GRIB format

GRIB (**GRI**dded **B**inary) has been primarily designed to broadcast grid generated by numerical weather forecast model, but it is also used for meteorological satellite images [5].

A GRIB message is a sequence of six parts (or sections):
1. Start of message (section 0)
2. A product definition section which describes content of message (date and time, source, parameter and unit, etc.)
3. A grid definition section specifying data geometry (projection, grid size, extension). Although a lot of projections are listed in GRIB tables, latitude−longitude projection seems to be the only one used for forecast model output. To avoid oversampling on poles, grids are often thinned: the number of grid nodes on a parallel is latitude dependent.
4. An optional bit map section used to mask missing or not relevant data
5. The data section itself. Various compression algorithms are proposed: the most used relies on integer increments from a minimum value. Number of bits for each value can be freely specified on the beginning of this section.
6. End of messages

In the most simple situation (a satellite image without compression) GRIB format is only an header before raw raster data.

4 BUFR format

BUFR format (**B**inary **U**niversal **F**orm of **R**epresentation of meteorological data) is in fact a very general format which could support any type of data assuming we could get the corresponding tables.

A BUFR message is also defined as a sequence of sections:
1. Start of message
2. A product definition section indicating mainly originating centre, date and time
3. A data definition section
4. The data as defined by the previous section
5. End of message

The data definition section is composed of 16 bit words. Each word specifies one data field according to predefined tables or a sequence repetition. Sequence repetition factor may be part of the data section. Thus, it is possible to define vector objects (jet−streams or fronts) with a variable size.

Whereas writing a general BUFR decoder is possible, actual data interpretations depend on reference table which defines parameter, unit and storage size. As data section storage unit size is variable, without any requirement of word or byte alignment, message decoding supposes that reference tables have been properly updated.

Main use of BUFR is broadcasting significant weather charts. Formally this information was graphically coded as fax format (T4). But this representation is not appropriate for automatic merging with other type of information (satellite, ground observation, gridded data). Moreover transmitting objects rather than their representation decreases dramatically message size.

Although BUFR may code any type of observation data, former formats are still in use, mainly because some messages have to be readable without any dedicated software. Anyway its use is expanding and NOAA is planning to broadcast wind profiler data in this format. An other frequent use is dissemination of precipitation radar data.

5 Flow analysis issue

Life should be easier if all produced information were packed in one message for each run of a numerical model. But to be compliant with message switching system specifications, message size is limited (around 15K). Compression algorithms are not sufficient and information is split in different messages to match with this requirement. We cannot assume a sequential reception of these messages and not even a complete one.

A common option in forecaster workstation software is to decode incoming messages and to ingest them in a database: Oracle (Synergie) or NetCdf (Unidata). Thus user requests will be faster as data access will be optimised. Nevertheless this option has two major drawbacks:

1. Data compression is lost and generated data base size is much more important than those of original messages.
2. Software complexity is increasing as we require installation of an extra data base package.

There is a trade−off between speed and simplicity. Dealing with ingestion inside a general oriented GIS, we consider appropriate to use a two steps analysis process:

1. messages extraction and identification
2. information retrieval on user request

During first step input data flow is segmented in GRIB or BUFR messages according to section 0 information. Then section 1 is analysed and file names are generated according to section 1 content. If we get data from an ftp server (as NCEP reanalysis) one file will contain only one message, but if we get our data through satellite receiver (Retim or Sadis), we will have not only to split data flow, but also to check message correctness and to discard incomplete ones.

When retrieving data, the program will analyse all the files, whose names match with the user request. First step analysis of section 1 acts as a filter to increase second step performances.

File name pattern will depend on the type of encoded information. For GRIB encoded weather forecast model, information. supplied will be:

- Model identification and originating centre
- Vertical co−ordinate value
- Parameter
- Date and time
- Period

Whereas for BUFR chart file name will indicate:

- Originating centre
- Date and time
- Data category

In this design, the data base manager is fully replaced by the operating system file manager. An alternative design should be to associate an information file with GRIB messages. But as file name maximum size is far beyond our requirements, there is no need to introduce this extra file and corresponding update process. Moreover we can presume that hash coded directory scanning should be much faster than text file analysis.

6 Data internal representation

Data representation of BUFR coded significant weather chart is rather straightforward as the type of spatial object (point, line, surface) is specified in a message. One should only care that spatial sampling of vector objects is rather coarse and have to be completed by a spline algorithm to be properly represented.

Weather forecast model data are broadcasted on a regular grid of a geographical coordinate . In many cases, region size will be small compared with model global coverage and analyser will extract site data. So, various interpolation and geostatistic methods could be carried out on these data.

When satellite images are coded in GRIB format, data type will be obviously raster. Sometimes, these data are not georeferenced in a projection but directly broadcasted in satellite view. As geostationary satellite projection is not supported by GRASS, may be the simpler method should be to convert it in lat−lon projection.

7 Unit conversion and index computation

For some requests, processing GRIB messages can be rather easy. Extracting a supplied parameter on an isobaric level will require at most a mosaic of different messages without any data interpretation. But assuming that received messages will always contain information on resquested form is unrealistic. It could be necessary to convert one parameter into an other (for example dew point temperature to relative humidity) or compute multilevel indicators (as various instability index) [3]. To be efficient a GRIB support software should include a moderate amount of these computation modules. Unfortunately they are highly dependent upon selected parameters in model output and user investigation. A proper specification of these functions is an open issue.

Interpolation has to be done carefully. As most of thermodynamic relations are non linear, it should be inaccurate to perform associated computation on generated fields rather than on grid nodes.

8 Organisation and status of decoding software

WMO format support software is composed of:
- A flow analysis command
- An information command
- A model−grib decoding function
- A satellite−grib decoding function
- A *sigwet−bufr* decoding function.

The flow analyser creates an element called *wmo_messages* to store individual wmo messages. As quoted before file names indicate their content.

The information command lets the user know which information he actually has.

Decoding function for weather forecast models use the following input parameters:
– Geographical window coordinates
– Originating centre
– Date, time and period
– Isobaric layer (if any)
– Parameter identification

Parameter is identified according to GRIB table. Values above 256 indicate derivative index locally defined. The function checks first, if this parameter is present in the message directory. If not, it searches for a derivative relation (e.g. wind direction can be deduced from two components of wind speed) and checks again for new input parameters.

On output, it returns the number of grid nodes inside the window and three arrays (latitude, longitude and value).

Satellite grib function does not perform any unit computation. It returns a raster image extracted on the selected window. If necessary, it projects data from the satellite view to latitude and longitude.

Sigwet−bufr function is much more simple as it does not filter on layers and parameters. It returns all the vectors which are intersecting with window area.

Test are carried out on Sadis and Retim satellite data flow. Beta version can be downloaded from:

ftp://expedito.univ−paris1.fr/grass_wmo

A lot has still to be done to properly support GRIB and BUFR in GRASS and mainly writing a user interface for the extraction functions.

9 And what about other WMO messages ?

As we quoted before, a lot of meteorological information is transmitted in alphanumeric messages: synoptic station or drifting buoy data, airport bulletins, atmospheric sounding plane measurements. Each kind of message is associated with a specific format. Corresponding analyser software is available on the web. Unfortunately, coding is not homogeneous between different formats and writing a general alphanumeric decoder is a rather difficult task. Moreover, as they are designed for human coding and interpretation, an error detection module has to be implemented.

A major change is going on with introduction of CREX (**C**haracter form for **R**epresentation and **EX**change of data) , a new format defined by WMO in order to replace the various alphanumeric formats. In many ways, this format is very similar with BUFR. A data description section has to be included before actual data and managing CREX messages should require only minor changes from BUFR decoder. But unlike BUFR, CREX is an alphanumeric format and generated messages are human readable. Last version of CREX table has been released in November 2001 and GTS conversion process is beginning. We can expect in next future that significant amounts of meteorological information will be transmitted in unified format.

References

[1] Fenix J.L.R. A history in communication technology evolution. NOAA on line library: http://www.history.noaa.gov/stories/gateway.html

[2] IGES on line ressources: http://grads.iges.org/home.html/

[3] Triplet J.P, Roche G. Météorologie générale. Météo−France, Paris, 1996.

[4] Unidata on line software ressources: http://www.unidata.ucar.edu/software.html

[5] WMO. International codes: volumes I.2. Wmo publications. On line version: http://www.wmo.ch/web/catalogue