U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
OFFICE OF SYSTEMS DEVELOPMENT
TECHNIQUES DEVELOPMENT LABORATORY

TDL OFFICE NOTE 97-1

SATELLITE, GRIDPOINT, AND VECTOR DATA PACKING

Harry R. Glahn

July 1997

SATELLITE, GRIDPOINT, AND VECTOR DATA PACKING

Harry R. Glahn

1. INTRODUCTION

By far, the largest components of data to be transmitted by the Advanced Weather Interactive Processing System (AWIPS) are satellite data and gridpoint data. These data are characterized as being of a single "type" (e.g., 1000-mb temperature) in a "message," and therefore the units of the data are the same and the general magnitude does not generally vary greatly within the message. Other aggregates of data, such as the forecasts of, say, cloud amount or height at 500 to 1000 locations (not gridpoints) also have those characteristics. There is one difference, though; once the parameters of the grid over which the gridpoint or satellite data are transmitted are specified, the "gridpoint" or "pixel" locations are known in reference to the earth, but for the data not related to a grid, the locations are not automatically specified without reference other information. This reference could be to another message, another section within the same message, or to a table or tables. These data are called single-element vector data (or just vector data) in this note.

Also, non-gridpoint (call them station) data may be grouped by location so that several different "kinds" or "types" of data are in one message (e.g., different weather elements). Such data can be formatted, and compressed if that feature is used, in BUFR (Binary Universal Form for the Representation of meteorological data) (OFCM 1995). While a non-trivial amount of such data needs to be transmitted by AWIPS (e.g., MOS forecasts and upper air observations), the volume is small compared to gridpoint and satellite data.

The vast quantities of data slated for transmission by AWIPS to try to accommodate the insatiable thirst for more and better products almost mandates some form of packing.[1]

Today, for transmission over the AWIPS Satellite Broadcast Network (SBN), the simplest form of GRIB (GRIdded Binary) (WMO 1988; Dey 1996) is used for gridpoint data by the National Centers for Environmental Prediction (NCEP), and satellite data are not packed at all, except that only one byte is used per datum. GRIB, in this simple form, does basically two things: (1) the overall minimum value is subtracted from the field (the values at the gridpoints), making all values positive and having no larger magnitude than necessary, and (2) only as many bits per value are used as the largest value in the field (after subtraction of the overall minimum) requires, both after the desired decimal and/or binary scaling is done. For a 12-h precipitation amount field, one gridpoint could have a value of 300, meaning 3 inches scaled to hundredths of inches, which would require nine bits to pack. Therefore, each and every point would require nine bits even though most points would be zero or quite close to zero.

---

[1]No distinction is made here between "packing" and "compression"--both terms just mean sending the same information in fewer bits.

The standard World Meteorological Organization (WMO) GRIB provides for second-order packing, which the WMO calls the "complex" method (WMO 1988). This feature allows for groups of "adjacent" values to be transmitted with only the number of bits required for that group after the minimum for the group is subtracted. The definition of the groups is left to the originating organization, and the GRIB code provides enough information so that the unpacker can function properly. The groups can be of constant size (such as defined by grid row or column, which for fields covering a large area is not very useful) or the group size can vary. Unfortunately, the way GRIB carries the group sizes is quite inefficient (see Glahn 1992, 1993), but not enough so to offset the advantage of "grouping." The trick is to pick appropriate groups, but an algorithm for this exists (Glahn 1994).

However, packing is not free; computer central processor (cp) cycles are consumed, and the packing and unpacking process can be complicated. Fortunately, software for packing and unpacking can be written once and (almost) for all, provided each user doesn't insist on writing his/her own code. The cp time can be reduced to a minimum by optimization, but is still not trivial. However, the volume of transmission over each of the various circuits and local area networks (lans) will be significantly reduced, as well as possibly the storage, depending on the database and retrieval design.[2]

It would be advantageous if one efficient form of packing could be used in AWIPS for all forms of binary data when the "message" is composed of the same type of data (ruling out several types of data for one location). Since a "standard" for satellite data does not exist, an efficient form could be specified. Also, the general format of GRIB can be used for defining the "field" to be transmitted.

For instance, as a possibility, why not instead of transmitting 500-mb heights, transmit second-order (spatial) differences of height as defined by the Office of the Federal Coordinator (OFCM 1990, p. 10-2)[3]? Enough latitude is provided in GRIB to individual Centers (of which NCEP is one) that this could be done and still fit roughly within the WMO GRIB framework. While certain products would undoubtedly need to be produced in standard GRIB for international exchange and to fulfill the National Weather Service's part of the weather data backup bargain,[4] the vast quantities of data to be carried by

---

[2]There is some indication that a product retrieved at random (so that caching is not generally possible) can be returned and unpacked quicker than the same unpacked information could be retrieved. Even so, either method could be better, depending on the bottlenecks in the system.

[3]This has been done with some products on the Automation of Field Operations and Services (AFOS) system in order to get a modicum of gridpoint data to users of that system.

[4]In fact, most international exchange of data and data for other U.S. Centers for backup purposes is on the so-called WAFS grids (Dey 1996, Table B), which are used in AWIPS only for model data from other countries or for backup purposes in case NCEP is not operative. That is, there is little overlap between the voluminous AWIPS SBN gridpoint data and the much lesser

(continued...)

the AWIPS SBN might not need to be tied to those restrictions. Efficient packing becomes even more important to the extent the AWIPS Wide Area Network (WAN) is used for backup when the SBN is inoperable or for product restoration at individual stations.

This office note reports on a packing scheme, called TDLPACK, appropriate for both gridpoint and satellite data as well as for single-element vector data. The formats of GRIB and TDLPACK are discussed in Section 2, and examples of their performance in terms of storage and the related software cp time is shown in Section 3. A summary is provided in Section 4.

2. THE GRIB AND TDLPACK MESSAGE FORMATS

GRIB is essentially a self-describing code; enough information is sent with the "message"[5] to fully identify the data. This requires that the format of these "describing" data be specified. The TDLPACK data format generally follows the GRIB format, but important differences exist. This data packing and description format has been adopted by the Techniques Development Laboratory (TDL) for archival and developmental purposes. Several kinds of data are stored in this format and used by both developmental and operational software.

A complete description of the TDLPACK format is contained in Glahn and Dallavalle (1997, Chapter 5) and is summarized below. Basically, the format consists of five Sections, but the numbering scheme of GRIB is followed, so the description is in terms of six Sections, zero through five, with one section, Section 3, not being used.

A. Section 0

This section always consists of eight bytes in three groups, the first group of four bytes containing four characters (TDLP for TDLPACK, but could easily be any four characters, such as GRIB or AWIP), the second group of three bytes containing the total message length in bytes[6], and the last byte containing the edition number (a number describing the version of rules that the packed message follows). This section precisely follows the GRIB format.

B. Section 1

This is the product description section (PDS) of GRIB. It has been tailored to TDL needs, but the exact content can be modified to fulfill the purpose at hand. To use this format for gridpoint data, the exact GRIB definition could

---

[4](...continued)
amounts of data for international exchange.

[5]"Message" is used in this note, and the term is consistent with the usage in the WFO document, but it is understood that the full code can also be used as a storage format.

[6]Note that this allows a maximum message size of $2^{24} - 1 = 16,777,215$ bytes, which seems adequate.

be used.[7] For satellite data, the GRIB convention could be followed, although no arrangement has been made for the "definition" of satellite products in a table such as Table 2 (Dey 1996), so some accommodation would be in order. Alternatively, the description of the satellite data could be the same as is now transmitted in the humongous (as compared to necessity) 512-byte header for satellite products.

C. Section 2

This is the grid description section. Again, it has been tailored to TDL needs, but could conform to the GRIB standard. For satellite data, there is no standard, except what is in the current 512-byte header. For vector data, this section is not needed, as such, but provision must be made for defining where the points in the vector are. This can be a separate message, or the PDS can point to a table of values defining the locations of the data points, much as BUFR does (or can do) in referencing locations.[8]

D. Section 3

This section is not used in TDLPACK. In GRIB, it contains the so-called (primary) bit-map which, when necessary, defines the location of the "missing" data points. That is, when there are gridpoint values that are missing, the location of the point with data (no data) can be represented by a 1 (0) in the map. This scheme is inefficient, especially if only a few points are missing, because one bit is required for each and every point. GRIB does provide for a referenced bit map that is not explicitly included--if used, another complexity for the user and a degradation of the "self-defining" aspect of the GRIB message. Generally, for gridpoint data, no bit map is required. However, it turns out that the NCEP Eta model (Rogers, et al. 1996) messages require a bit map, because forecasts are not made at all points on the AWIPS grid being used to transmit the data.

---

[7]There are some problems with the GRIB PDS. For instance, only one byte (No. 9) has been reserved for the parameter and its units (Table 2, Dey 1996). These 255 definitions have already been used, so definition of new products requires some imagination. For instance, the NWS has been allocated three Center numbers (byte No. 5), and each Center can have Subcenters (byte No. 26). By implication, each Subcenter can have its own Table 2. Not a pretty picture in receiving and sorting out, from various tables that could be needed, what the product actually is.

[8]In TDL's archive, the first record of each data set, and interspersed if needed following a "trailer" record, is a record defining the location by (what else!) station call letters. The call letters are then entry points into a station directory containing station locations as well as other information.

E. Section 4

Section 4 is the data section. In both GRIB and TDLPACK, the information for actually unpacking the data are contained here.[9] In GRIB, if second-order (complex) packing were used, Section 4 would contain a secondary bit map (distinguished from the possible bit map in Section 3) such that for every point a 1 would indicate the start of a group, and all other bits would be zero.[10] While the purpose of Section 4 is the same in both GRIB and TDLPACK, the exact layout of the data is considerably different. This should not present a problem for satellite data as there is no international standard and GRIB is not currently being used.

The primary difference between GRIB and what is proposed here for satellite data--and what could be used for AWIPS transmission of gridded data--is that instead of transmitting the actual data values, second-order spatial differences are transmitted according to the boustrophedonic scheme described by OFCM (1990). To use second-order differences, the first actual value is sent, the first first-order difference, and then N-2 second-order differences, where N is the total number of points. The second-order differences are calculated starting at the lower left, proceeding along the first row to the end, then jumping up to the next higher row and proceeding back along the row. In this way, the second-order differences are minimized.[11] Finally, the second-order packing (complex, in WMO's terminology) is applied.

For the current 8-bit satellite data products on the SBN, the largest value of the byte necessary for sending the data is reserved for the missing value. Another important difference between GRIB and TDLPACK is that an extension of this scheme is used for missing data in TDLPACK; whatever the number of bits required to pack the values in a group, the largest possible value is reserved for a missing value. If the field has no missing value, no reservation is necessary. TDLPACK goes one step further; the next lower value is reserved for a "secondary" missing value, (only) when there is such a secondary missing value present. This was included primarily because certain MOS forecasts cannot be made, and it is necessary to distinguish between the case when there is just no forecast made for some reason (probably because of missing data), and when a forecast cannot be made (e.g., the probability of the lowest category of visibility for a particular station, because not enough data were available to develop a statistical relationship). This is also handy for an "unlimited" height of clouds, to be distinguished from a missing report. Other uses will undoubtedly crop up. This reservation of one value (or two in the case of a secondary as well as a primary missing value) is much more

---

[9]This is not quite true in GRIB; Section 1 contains the decimal scale factor and Section 4 contains the binary scale factor.

[10]If there is a Section 3, only the non-missing points as indicated by this primary bit map have a corresponding value in the secondary bit map.

[11]Second-order differences are not always advantageous (e.g., precipitation amounts). An algorithm, part of the TDLPACK software suite, differentiates appropriately.

efficient than the use of a bit map, which, incidently, doesn't provide for a secondary missing value.[12]

F. Section 5

Section 5 is the trailer. It consists of four bytes, each containing an ASCII character 7. TDLPACK and GRIB are identical in this respect.

3. PERFORMANCE

An example of the use of TDLPACK is given here for each of the data types, satellite, gridpoint, and vector. Consistent with the earlier report by Glahn (1995), the adjustable parameters MINPK and INC were used as 14 and 1, respectively (see Glahn 1994). For determining processing times, a Hewlett Packard (HP) 755[13] was used that was otherwise only performing "housekeeping" chores with about 1 to 2 percent of its cp cycles. The timing software available was quite precise, but gives clock time, not actual cp time. Timing results were quite consistent, as shown by replication, but differences of only 1 or 2 percent between values are in the noise level.

A. Satellite Data

Two 4-km water vapor products and two 1-km visible products were packed; the results are shown in Table 1 and discussed below.

Table 1. Statistics associated with packing satellite data. The packed message sizes are the average of the two products. The times for the visible are per scan line (message), the way the data are currently sent.

Product	Original Product Size (bytes)	New Product Size (bytes)	Size New/Old (%)	Packing Time Per Message (sec)	Unpacking Time Per Message (sec)
Water Vapor					
8-bit data	243,561	144,980	0.60	0.83	0.19
10-bit data	304,323	208,216	0.68	0.84	0.19
Visible					
8-bit data	26,220,032	15,528,244	0.59	0.012	0.003
10-bit data	32,773,632	22,435,032	0.68	0.013	0.003

[12]One <u>could</u> use a two-bit map--about what one would be worth.

[13]No endorsement of specific equipment or companies is expressed or implied in this document.

Water Vapor

Each of these products contained 243,049 points (on a 493 X 493 grid). The data were for February 21, 1997, at 2345 UTC and February 22, 1997, at 0045 UTC. The "input" message consisted of a 512-byte header, making the product 243,561 bytes in length. After packing by the method described in this note (TDLPACK), one product was reduced to 60% of original size, and the other to 59%. This equates to 4.80 and 4.74 bits per point, respectively, rather than the original 8 bits per point (it is not surprising that the values for the two products are similar, because the satellite images were only 1 hour apart). It is noted that if one were to send 10-bit data, these two samples would have each packed in 6.85 bits or less per point--less than the original 8 bits per point for the 8-bit data. The 10-bit data were manufactured by randomly adding 2 bits, one at a time, to the 8-bit data. This is a worst-case scenario for second-order differencing and second-order packing.

The time necessary to pack (unpack) each 8-bit field was about 0.83 (0.19) seconds on an HP 755 almost entirely devoted to this process as described earlier. These results were very nearly the same for 10-bit data.

Visible

Each of these products contained 26,214,400 points (on a 5,120 X 5,120 grid), plus a 512-byte header and a 5,120-byte trailer record, making the total product size 26,220,032 bytes. The data were for April 22, 1996, at 1702 UTC, and January 30, 1997, at 1815 UTC. The data are sent line by line, each of 5,120 points, as the data are created by the National Environmental Satellite and Data Service (NESDIS) as soon as they are available from the satellite. Therefore, to be consistent, each line was packed separately.

As with water vapor, the size of one packed product was 60% of its original size, and the other 59%. Obviously, this also equates to about 4.7 or 4.8 bits per point. It was also determined that 10-bit visible data could pack in about 6.8 bits per point, a reduction of about 32% for 10-bit data.[14] The total time for packing (unpacking) was 63 (15) seconds. Note that this is only 0.012 seconds (0.003) seconds per line, the way the messages are sent and received. A question here is whether cp cycles equivalent to 0.012 seconds on an HP 755 should be used to reduce the size of the product by 40%. Undoubtedly, the cp time on AWIPS for unpacking--equating to 0.003 seconds per line (message) on an HP 755 as it is received--would be available, and the throughput with the smaller message size might even be greater.

[14]It was assumed that for 10-bit data, the "original" message would contain 32,768,000 bytes of data plus a 512-byte header, plus a 5,120-byte trailer, although the trailer would probably not remain at that size if the data were not byte-oriented. Present procedures might require two full bytes to be sent to accommodate the extra two bits. If that is the case, the reduction would be 58% instead of 32%.

B. Gridpoint Data

The experiments reported here are similar to those reported earlier (Glahn 1995). However, different data were used, a new method of packing is included (TDLPACK), and more efficient routines for packing, unpacking, and determining group sizes were used.

The data used were from the NCEP Eta model for April 7, 1997, at 0300 UTC. Both 40-km and 20-km data were available. These grids were downloaded from the NWS Office of Systems Operations file server, deGRIBbed and written as a binary file, then GRIBbed and deGRIBbed by both simple and complex standard GRIB methods and by TDLPACK. The packing precision was in all cases the same as that used by NCEP; that is, the binary and decimal scale factors were those with the grids. These factors were in many instances different from those used earlier (see Glahn 1995). Generally, the decimal factors were zero, and the binary scaling was used to regulate the precision. Evidently, these factors now generally follow the discussion advanced by Petersen.[15]

40-km data

The results for 40-km data are shown in Table 2. For these fields, the decimal scaling was in all cases zero. For height, the binary scaling was -3; for temperature, it was -3, except for 400 mb at 9 hours and 100 mb at 30 hours it was -2; for relative humidity, it was 0; for wind components, it was -3, except for one component at 500 mb it was -2 for 21 hours; for precipitation, it was -3, except it was -2 for one field at 27 hours; and for vertical velocity, it was -5, except it was -4 for 150 mb at 18 hours and 1000 mb at hours 3 and 21. Also, in keeping with past practice, precipitation values of 0 were really -0.25.

From Table 2, it can be seen that complex GRIB improves on simple GRIB by 20% overall, ranging from 9% for relative humidity to 64% for precipitation amount. Processing times go up for complex by a factor of about 54% and 33% for GRIBbing and deGRIBbing, respectively. For TDLPACK, the packing time further increases, although for unpacking, the time is actually less than for even simple GRIB. This unexpected result is due to the fact that the grid contains missing values. The missing values are indicated in GRIB by a bit map (see Section 2.D), which is not only expensive in size but also in processing, because each bit in the product has to be unpacked and dealt with. TDLPACK gains further on complex GRIB because GRIB uses another (!) bit map to indicate where the groups start and stop--another expensive processing job requiring stuffing and unstuffing a bit for each gridpoint again. Without missing data, the unpacking time for the simple method is about the same as that for TDLPACK (see last entries in the table).[16]

---

[15]Private communication.

[16]The missing gridpoints were filled in with a constant value equal to the last non-missing value. This probably gives a slight advantage to the complex method, but not necessarily so for TDLPACK, because the series of missing data are relatively short, and a hiatus occurs between the filled in values and the next non-missing value, making adjacent second-order differences there relatively large and opposite in sign.

Table 2. Statistics associated with packing 40-km Eta model fields by three different methods. The two numbers in parentheses in the field definition are from the WMO Tables 2 and 3, respectively (WMO 1988).

Field Definition/ Packing Method	No. Fields	Message Length (byte)	Bits/ Point	Size New/Old (%)	Points/ Group	GRIB Time Per Field (sec)	DeGRIB Time Per Field (sec)
Geopotential Height (7, 100)							
Simple	324	40,338	13.53	--	--	0.047	0.025
Complex	324	32,008	10.73	79	17.0	0.074	0.033
TDLPACK	324	15,499	5.20	38	20.4	0.087	0.020
Temperature (11, 100)							
Simple	324	29,672	9.61	--	--	0.046	0.024
Complex	324	21,050	7.06	71	17.4	0.071	0.032
TDLPACK	324	10,492	3.52	35	24.8	0.089	0.020
Relative Humidity (52, 100)							
Simple	264	23,656	7.93	--	--	0.045	0.024
Complex	264	21,513	7.21	91	18.1	0.070	0.031
TDLPACK	264	12,215	4.09	52	21.9	0.090	0.020
U-, V-Wind Components (33 & 34, 100)							
Simple	648	30,056	10.08	--	--	0.046	0.025
Complex	648	24,865	8.34	83	17.6	0.072	0.032
TDLPACK	648	12,785	4.29	43	21.8	0.087	0.020
Vertical Velocity (39, 100)							
Simple	324	22,586	7.57	--	--	0.047	0.024
Complex	324	17,013	5.70	75	19.2	0.068	0.032
TDLPACK	324	11,603	3.89	51	23.0	0.090	0.020
Precipitation (61 & 63, 1)							
Simple	24	23,289	7.81	--	--	0.050	0.025
Complex	24	8,482	2.84	36	51.3	0.063	0.023
TDLPACK	24	5,936	1.99	25	49.6	0.092	0.018
Overall							
Simple	1908	29,328	9.83	--	--	0.046	0.024
Complex	1908	23,427	7.85	80	17.9	0.071	0.032
TDLPACK	1908	12,491	4.19	43	22.4	0.087	0.020
Overall (No missing points)							
Simple	1908	26,733	8.96	--	--	0.037	0.018
Complex	1908	20,653	6.92	77	17.7	0.063	0.025
TDLPACK	1908	12,331	4.13	46	21.3	0.070	0.017

A major finding here is that, overall, the TDLPACK product size is only 43% of that of the original, simple GRIB product. This is achieved at the expense of packing time of 0.087 - 0.046 = 0.041 sec per product, but with an improvement of 0.004 sec per product in unpacking.

In comparing values in Table 2 with results in Glahn (1995), one can note:

- The original message sizes are somewhat smaller for this set of data than for the one used previously, being on the average 9.83 bits/point rather than 10.20--still considerably over the Appendix K (U.S. Government 1989) expectation of 8 bits/point.

- The improvement of complex over simple GRIB is about 20% for both samples.

- GRIBbing times are considerably smaller--0.046 versus 0.113 sec/message for simple and 0.071 versus 0.188 for complex--due to improvements in algorithms and program structure.[17]

- DeGRIBbing times were also down from 0.055 sec/message to 0.024 for simple GRIB and from 0.085 sec/message to 0.032 for complex GRIB.

20-km data

The results for the 20-km fields are shown in Table 3. For these fields, the decimal scaling was 0 except for absolute vorticity for which it was 5. For temperature and wind components, the binary scaling was -4; for absolute vorticity, it was 0; and for precipitation, it was -3.

Table 3 reveals similar information to that in Table 2. Generally, the improvement of complex GRIB over simple GRIB is greater at 20 km than at 40 km (25% versus 20%). Although this was expected from the earlier report (Glahn 1995), it may be somewhat an artifact of different fields being present at 40 km than at 20 km. However, where there were matching fields and the precision remained constant (precipitation amount), the improvement increased from 64% to 67%; where the precision increased by 1 bit for 20-km data (wind components), the improvement increased from 17% to 21%; and where the precision increased by 2 bits (temperature), the improvement was approximately the same (29% versus 28%).

Also, the improvement of TDLPACK over simple GRIB is greater at 20 km than at 40 km (58% versus 57%), but the difference is small and not significant. Where there were matching fields and the precision remained constant (precipitation amount), the improvement increased from 75% to 78%; where the precision increased by 1 bit for 20-km data (wind components), the improvement remained constant at 57%; and where the precision increased by two bits (temperature), the improvement decreased from 65% to 59%.

[17]Improvements included removing computations involving powers of 2, using different bit manipulation routines, removing calls to bit manipulation routines in loops by using more in-line code, and improving the basic algorithm for finding groups. Each of these brought significant improvement.

Table 3. Statistics associated with packing 20-km Eta model fields by three different methods. The two numbers in parentheses in the field definition are from the WMO Tables 2 and 3, respectively (WMO 1988).

Field Definition/ Packing Method	No. Fields	Message Length (byte)	Bits/ Point	Size New/Old (%)	Points/ Group	GRIB Time Per Field (sec)	DeGRIB Time Per Field (sec)
Temperature							
(11, 105 & 108)							
Simple	72	128,994	10.88	--	--	0.188	0.102
Complex	72	92,909	7.84	72	17.2	0.292	0.133
TDLPACK	72	53,368	4.50	41	21.7	0.357	0.086
Relative Humidity							
(52, 105 & 108)							
Simple	72	93,770	7.91	--	--	0.184	0.097
Complex	72	73,826	6.23	79	18.0	0.281	0.127
TDLPACK	72	41,724	3.52	44	24.1	0.367	0.083
U-, V-Wind Components							
(33 & 34, 105 & 108)							
Simple	144	126,478	10.67	--	--	0.190	0.102
Complex	144	100,400	8.47	79	17.4	0.296	0.133
TDLPACK	144	54,912	4.63	43	21.0	0.358	0.085
Absolute Vorticity							
(41, 100)							
Simple	60	110,910	9.36	--	--	0.182	0.097
Complex	60	84,733	7.15	76	17.9	0.284	0.130
TDLPACK	60	47,203	3.98	43	24.8	0.367	0.083
Precipitation							
(61 & 63, 1)							
Simple	24	94,256	7.95	--	--	0.182	0.098
Complex	24	31,092	2.62	33	72.6	0.266	0.098
TDLPACK	24	20,875	1.76	22	66.8	0.396	0.077
Overall							
Simple	372	116,045	9.79	--	--	0.186	0.100
Complex	372	86,808	7.32	75	18.5	0.288	0.129
TDLPACK	372	48,621	4.10	42	23.4	0.363	0.084
Overall (No missing points)							
Simple	372	105,666	8.91	--	--	0.147	0.068

For 20-km grids, TDLPACK achieved an average product size of only 42% of the original simple GRIB at the expense in packing of 0.177 sec/grid, but with an improvement in unpacking of 0.016 sec/grid.

In comparing values in Table 3 with results in Glahn (1995), one can note:

- o The original message sizes were very similar--10.34 and 10.24 bits/point, a difference easily attributable to differences in precision or just the different sample.

- o The improvement in complex over simple GRIB is slightly smaller in the current sample--25% versus 29%.

- o GRIBbing times have decreased commensurate with those for 40-km fields, being down from 0.481 sec/grid to 0.186 for simple GRIB and from 0.710 sec/grid to 0.288 for complex GRIB.

- o DeGRIBbing times are down from 0.219 to 0.100 sec/grid for simple GRIB and from 0.344 to 0.129 sec/grid for complex GRIB

From both Tables 2 and 3, it can be noted that second-order spatial differences, employed by TDLPACK, not only provides smaller numbers to pack (slightly over 4 bits/point for TDLPACK versus between 7 and 8 bits/point for complex GRIB), but larger strings of similar values can be found--about 23 versus 18 points/group.

It is interesting that while it is almost always true that the unpacking of complex GRIB took more computer time than the unpacking of simple GRIB, this is not true for precipitation amount. How can this be? While this phenomenon was not investigated, it is likely due to (1) the long strings of similar values (values of -0.25 representing zero amount), and (2) the small number of bits per value for complex packing (less than three) as compared to simple packing. A small number of groups would contribute somewhat to less difference in unpacking time, and especially when all values in a group are identical, because then no value is sent for each point, only the one value (the group minimum) is needed. Also, when two computer words must be accessed to unpack one value (the value spans a word boundary), more machine instructions are necessary. The likelihood of a word boundary being spanned is considerably less when only three bits or less are required for a value than when about eight are required.

C. Vector Data

TDL archives surface observed hourly data in TDLPACK for use in development. The number of stations for which data are available varies considerably from hour to hour. For January 1996, there was a total of 21,468,152 "reports," a report being a value for each weather element for each station that would normally have a value for a specific hour (e.g., ceiling height should be reported each hour, but precipitation amount would not). For four of the weather elements, an 888 is used for "unlimited." There are also many missing values, each of which is initially (before packing) given the value of 9999. This month of data was packed with both primary and secondary missing values specified, with only a primary missing value specified, and without either a primary or secondary missing value specified. The statistics are shown in Table 4. A station call letters record is present with each hour because, generally, the stations reporting vary from hour to hour and for this rather raw archive no attempt is made to keep a constant set of stations. Call letters are not packed and the values discussed do not include them.

Table 4. Statistics associated with packing one month of hourly data in TDLPACK.

Packing Definition	Total Volume (byte)	Packing Time (sec)	Unpacking Time (sec)
Simple (1 Group)			
w/o primary missing	40,310,168	34.3	14.1
w/o secondary missing	22,361,064	34.2	14.8
with both primary and secondary missing	21,930,608	36.3	15.5
Second-Order			
w/o primary missing	31,178,976	54.4	13.7
w/o secondary missing	19,824,064	55.1	16.2
with both primary and secondary missing	19,006,024	55.3	16.9

In the table, "Simple" means no second-order grouping. "Second-Order" in this table is similar to Complex GRIB in Tables 2 and 3. That is, TDLPACK with no second-order grouping is much like the Simple GRIB--only the overall minimum is subtracted before packing. The first row "w/o Primary Missing" means the values of 9999 for missing and 888 for unlimited cloud height were packed just like any other value. These are large compared to the other data values, and it took at least 14 bits to pack each datum. The second row represents packing 9999 as missing (reserving the largest value possible in a group) and packing values of 888 the same as other values. The reduction was very substantial. The third row shows some additional reduction when the next lower value was reserved for 888 for those few fields in which 888 occurred. While the additional reduction is not great for the value of 888, use of a secondary missing value is more important when values of 9997 can occur.[18] In other words, with one or more values of 9997 in a field, the bits to pack a field would go back up to 14 as shown in the first row unless a secondary missing value were provided for. Without any packing at all, four bytes would be used for each value (on the 32-bit HP 755), and the total volume would be nearly 86 megabytes rather than the 19 megabytes shown in the last row of Table 4.

[18]In TDL's system of Model Output Statistics software, the value of 9997 is reserved for forecasts which cannot actually be made because insufficient data were present to obtain the necessary statistical relationship.

For second-order packing, the times to pack were about 55 to 60 percent larger than those for simple packing, but the unpacking times were only slightly higher. Second-order reduces in each case the bytes required for no grouping, whether missing values are considered or not. The overall differences for using a secondary missing value are not great, partly because the 888 is not particularly large (requires 10 bits to pack), but mostly because most fields had no values of 888.

SUMMARY

Large quantities of satellite and gridpoint data are being transmitted over the AWIPS SBN and it may on occasion be necessary to send some of these data over the terrestrial WAN. The satellite data are not compressed, except to the extent that only 8 bits of information are retained, and the gridpoint data are packed by the simplest of the GRIB packing options. The satellite data can be packed to occupy only about 60% of its original volume. For the visible, this can be done line by line so that the timeliness of the product is not compromised. No "standard" WMO format exists, so the packing scheme presented in this note could be adopted. The coding and decoding software exists; is documented, non-proprietary government property; and can be made available to any user.

The gridpoint data can be reduced to about 75 or 80% of its original volume by using the WMO standard GRIB "complex" method when some gridpoints are missing (requiring a primary bit map). When points are not missing, the ratio of new size to old is nearer to 71 to 77%. If one were to adopt the more efficient TDLPACK scheme discussed in this note (the same one mentioned in the paragraph above for satellite data), the volume of gridpoint data could be reduced to 40 or 45% of its present volume even with missing data points. These results are based on a sample of Eta model 20- and 40-km data. As the resolution of the grid is increased, the saving becomes even larger.

The packing of vector data is really no different from packing satellite data and is appropriate when data of only one type (e.g., satellite visible or surface temperature) are being packaged together. Reduction in volume of hourly data is about 50% of what would be required when values that can indicate a missing value (9999) are packed along the lines of the "simple" WMO GRIB scheme. While BUFR is being used to transmit MOS forecasts packaged by station, the data could be sent in vector format and packed as described here.

While the packing with more complex schemes than the simple GRIB take some computer time, it is really trivial on modern main frames. Also, the unpacking is well within the capability of workstation-class machines, and in fact, the method discussed in this note takes less unpacking time than the "simple" GRIB when missing points are present, because the latter requires a bit map that must also be "unpacked." It may be argued that the times given by this set of software are not representative of other existing or future software performing the same functions. This is true, but a substantial effort was put into optimizing the routines and making the algorithms similar in all testing done. If different algorithms are found to be better, they can be applied to all packing/unpacking schemes discussed here, and the <u>relative</u> results would not likely change substantially.

REFERENCES

Dey, C. H., 1996: GRIB, Edition 1, the WMO format for the storage of weather product information and the exchange of weather product messages in gridded binary form. National Centers for Environmental Prediction Office Note 388, National Weather Service, NOAA, U.S. Department of Commerce, 5 sections and an appendix.

Glahn, H. R., 1992: On the packing of gridpoint data for efficient transmission. TDL Office Note 92-11, National Weather Service, NOAA, U.S. Department of Commerce, 32 pp.

_____, 1993: An analysis of some features of the GRIB code. TDL Office Note 93-2, National Weather Service, NOAA, U.S. Department of Commerce, 42 pp.

_____, 1994: An improved algorithm for determining the groups in GRIB second-order packing. TDL Office Note 94-1, National Weather Service, NOAA, U.S. Department of Commerce, 25 pp.

_____, 1995: GRIB product sizes for AWIPS. TDL Office Note 95-3, National Weather Service, NOAA, U.S. Department of Commerce, 29 pp.

_____, and J. P. Dallavalle, 1997: MOS-2000. TDL Office Note 97-x, National Weather Service, NOAA, U.S. Department of Commerce, in preparation.

OFCM, 1990: Standard Formats for Weather Data Exchange among Automated Weather Information Systems FCM-S2-1990. Office of the Federal Coordinator, NOAA, U.S. Department of Commerce, 10 chapters and appendices.

_____, 1995: A Guide to WMO Code Form FM 94 BUFR FCM-16-1995. Office of the Federal Coordinator, NOAA, U.S. Department of Commerce, 6 chapters and appendices.

Rogers, E., T. L. Black, D. G. Deaven, G. J. DiMego, Q. Zhao, N. W. Junker, and Y. Lin, 1996: Changes to the operational "Early" Eta analysis/forecast system at the National Centers for Environmental Prediction. Wea. Forecasting 11, 392-423.

U.S. Government, 1989: AWIPS-90 System Requirements Specification, Vol. I, Appendix K (7/1/94 version).

WMO, 1988: Manual on Codes, Vol. 1, Part B--Binary Codes. WMO No. 306, World Meteorological Organization, Geneva.