UPDATED ANALYSIS OF LOSSLESS COMPRESSION TECHNIQUES FOR THE GOES-R REBROADCAST (GRB) SUB-SYSTEM

Peter Finocchio, Bobby Braswell, Yuguang He, David B. Hogan, Daniel Hunt, and T. Scott Zaccheo

AER, Inc, Lexington, MA 02421-3126

1. INTRODUCTION

This work describes an in-depth analysis of potential lossless data compression techniques for use in the design, development and deployment of the GOES-R ReBroadcast (GRB) sub-system. GRB is the next generation transponder-based communication link integrated into the GOES-R system, and is designed to provide the NOAA Satellite Operations Facility (NSOF) and other real-time users with GOES-R Level 1b data and Level 2+ Lightning products. The GRB data stream will include calibrated and geo-located Level 1b imagery from the Advanced Baseline Imager (ABI), Level 1b Solar Ultraviolet Imager (SUVI), Extreme Ultraviolet/X-Ray Irradiance Sensor (EXIS), Magnetometer (MAG) and Space Environment In Situ Suite (SEISS) data, and Level 2+ Geostationary Lightning Mapper (GLM) products. In order to ensure that the GRB bandwidth constraints are met at all times, lossless data compression must be employed. We assess the potential use of szip and jpeg2000 to meet these constraints.

This study focused on ABI L1b data, since ABI L1b data is the main driver for GRB volume. The ABI proxy data were created using currently available multi-spectral level 1b data. After applying the compression algorithms, the compression ratios over the proxy data are used to develop a relationship between compressibility and solar zenith angle. The relationship can then be integrated into an ABI fixed grid model to predict the compressed data rate for any GRB block. Additional studies addressing the latency of the compression algorithms, the optimal size and configuration of blocks within the ABI proxy datasets, the dampening of data rate peaks, and the compressibility of non-ABI data (SUVI, GLM) will supplement the results from the current analysis.

2. DATA

Spectral radiances from 270 MODerate Resolution Imagery Spectrometer (MODIS) Aqua granules are used to create an ABI proxy dataset. Due to significant striping observed in the Aqua near infrared band 6 (ABI band 5), we plan to substitute data from the Terra satellite for this band. Each MODIS granule was divided into 8 square blocks prior to compression.

Due to the limited geographic and temporal coverage of the MODIS data, we also plan to use 26 MSG-SEVIRI full disk images spanning 24 hours to create a global proxy dataset.

3. METHODS

In this section we present the methodology for applying compression algorithms to ABI-like data in order to obtain compressed GRB data rates. There are several other ongoing or planned future activities relating to the GRB compression analysis which are briefly described in section 5.

3.1 Creating ABI Proxy Data Sets

Evaluating the compressibility of ABI-like data requires the generation of an ABI proxy data set. The first step involved in generating ABI proxy datasets was scaling spectral radiances from the MODIS sensor to match the ABI quantization step size and dynamic range. In order to perform this scaling, we mapped the MODIS bands to the 16 ABI bands. Because we will eventually perform a similar compression analysis using MSG-SEVIRI data as the basis of an ABI proxy dataset, the mapping to SEVIRI bands is also provided in Table 1 below.

Table 1: ABI bands and central wavelengths (μm) mapped to corresponding MODIS and SEVIRI bands

ABI Band (μm)	MODIS Band (um)	SEVIRI Band (um)
1 (0.47)	3 (.47)	1 (0.64)
2 (0.64)	1 (.65)	1 (0.64)
3 (0.87)	2 (.86)	2 (0.81)
4 (1.38)	26 (1.38)	3 (1.64)
5 (1.61)	6 (1.63)	3 (1.64)
6 (2.25)	7 (2.11)	3 (1.64)
7 (3.9)	22 (3.96)	4 (3.92)
8 (6.19)	27 (6.75)	5 (6.25)
9 (6.95)	27 (6.75)	5 (6.25)
10 (7.34)	28 (7.33)	6 (7.35)
11 (8.5)	29 (8.52)	7 (8.70)

ABI Band (μm)	MODIS Band (um)	SEVIRI Band (um)
12 (9.61)	30 (9.74)	8 (9.66)
13 (10.35)	31 (11.02)	9 (10.80)
14 (11.2)	31 (11.02)	9 (10.80)
15 (12.3)	32 (12.03)	10 (12.00)
16 (13.3)	33 (13.36)	11 (13.40)

Quantization step size for ABI ($Q_{\nu_{ABI}}^{ABI}$) was computed using the dynamic range and bit depth for each ABI band and has units $mW/(m^2 \cdot sr \cdot cm^{-1})$. In creating the proxy datasets, we divided the measured MODIS radiance in each pixel by the ABI quantization step size in the corresponding band. Given a spectral radiance R_λ at MODIS band λ in units of $mW/(m^2 \cdot sr \cdot \mu m)$ and the ABI quantization step size, the quantized proxy ABI value at the same band ($D_{\nu_{ABI}}^{ABI\text{-}Proxy}$) was calculated as,

$$D_{\nu_{ABI}}^{ABI\text{-}Proxy} = \text{round}\left(R_\lambda^{MODIS} \cdot \left(\frac{\lambda^2}{10^4} \right) \Big/ Q_{\nu_{ABI}}^{ABI}(T) \right)$$

The factor $\lambda^2/10^4$ converts units from $mW/(m^2 \cdot sr \cdot \mu m)$ to $mW/(m^2 \cdot sr \cdot cm^{-1})$. As a result, D is a dimensionless quantity.

Before implementing the compression algorithms on the ABI proxy dataset, some additional steps were taken to exclude sensor-specific distortions from the proxy dataset. The land surface area within a MODIS pixel increases toward the edge of each swath, giving swaths the appearance of a bowtie (Gomez-Landsea et al., 2004). In order to overcome the bowtie effect, we ignored 150 pixels closest to each edge of the scan. Additional pixels were excluded in fitting two 506-km square blocks in the across-track direction and four 506-km square blocks in the along-track direction. The resulting MODIS granule data considered in all subsequent analyses is depicted as the dark gray region in Figure 1.

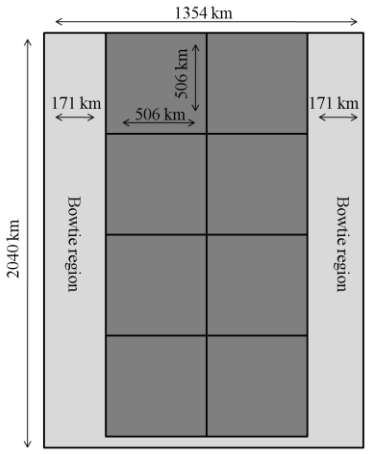


Figure 1: Depicts the portion of the MODIS granule that is used to create the ABI proxy dataset (dark gray region), and that which is trimmed due to the bowtie effect. The portion of the granule that is utilized is divided into eight square blocks.

3.2 Applying Lossless Compression

Lossless szip and jpeg2000 compression algorithms were applied to each block within the ABI proxy scenes. The resulting output files contain compression ratios (uncompressed size/compressed size) for each block of data, as well as the solar zenith angle at the center pixel of each block.

Dividing the proxy data into blocks has multiple advantages, one of which is the decreased size of ABI packets. Transmitting smaller packets through GRB allows for increased parallelization and decreased transmission latency. In addition, it was found that the compression algorithms achieved higher average compression ratios when implemented on blocks of MODIS data as opposed to entire MODIS granules.

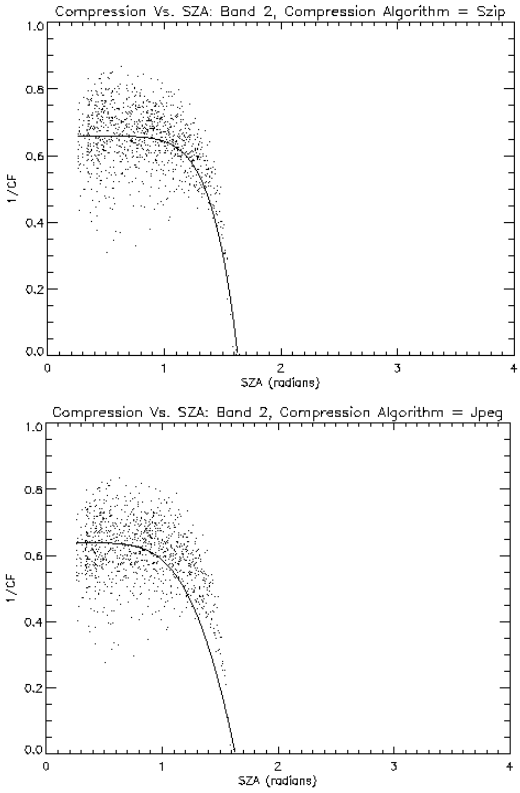
3.3 Compression vs. SZA

Attached to each block of ABI proxy data is a corresponding solar zenith angle (SZA) determined from the geolocation of the block and the time stamp of the MODIS granule. In order to develop a relationship between compression ratio and SZA, we fit a curve to scatter plots of compression vs. SZA. The following function was used in the curve fitting analysis:

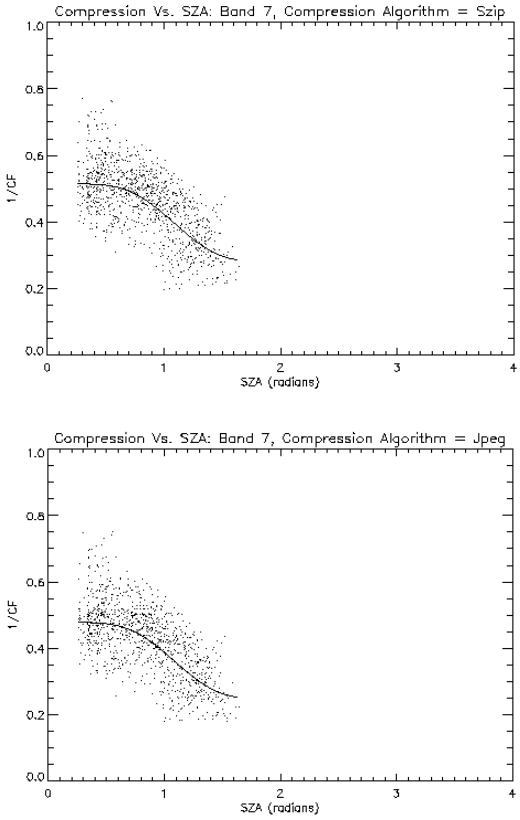
$$1/C = A_0 \cdot SZA^8 \cdot e^{A_1 \cdot SZA} + A_2$$

C is the compression ratio, *SZA* is the solar zenith angle (in radians), and A_0, A_1, and A_2 are a unique set of coefficients generated for each band that define the

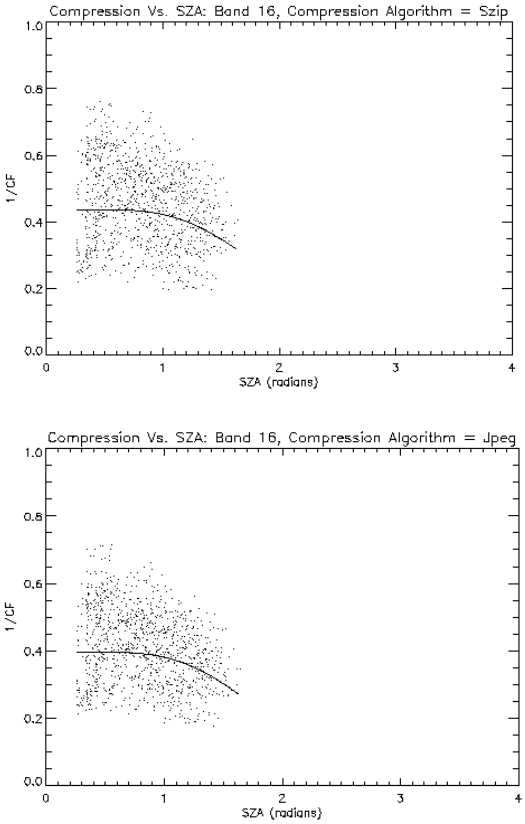
curve. In Figures 2-4 below, the resulting curves for each compression algorithm are presented for three bands: the visible band (2) with central wavelength 0.64µm, the SWIR band (7) with central wavelength 3.9µm, and an IR window/CO_2 band with central wavelength 13.3µm.



Figures 2a (top) and 2b (bottom): Scatter plots of ABI band 2 (0.64µm) inverse compression ratios vs. SZA for szip (top) and jpeg2000 (bottom) compression algorithms. The curve fit to the data defines the SZA-compression relationship



Figures 3a (top) and 3b (bottom): Scatter plots of ABI band 7 (3.9µm) inverse compression ratios vs. SZA for szip (top) and jpeg2000 (bottom) compression algorithms. The curve fit to the data defines the SZA-compression relationship



Figures 4a (top) and 4b (bottom): Scatter plots of ABI band 16 (13.3µm) inverse compression ratios vs. SZA for szip (top) and jpeg2000 (bottom) compression algorithms. The curve fit to the data defines the SZA-compression relationship

As expected, compression in visible and SWIR bands increases at larger SZA due to a decrease in the dynamic range of the image in low illumination conditions. There is also a slight increase in compression observed in the IR bands. One possible explanation for this is as land cools at night, the temperature disparity between the land surface and cloud tops decreases. As a result, clouds become less discernable as thermal features, thereby decreasing the structure of the image and increasing its compressibility.

3.4 The ABI Fixed Grid Block Model

Once a function relating compression ratio to SZA in each band is established, GRB data rate can be computed for blocks defined in an ABI fixed grid block model. The block model is a regular grid of square blocks superimposed on the ABI fixed grid. Blocks are arranged in 22 rows corresponding to the 22 ABI full disk swaths. The model assumes a 17.4-degree earth disk, with a "buffer" zone extending out to 17.76 degrees that contains earth atmosphere pixels (the narrow ring-shaped region surrounding the disk in the figure below).

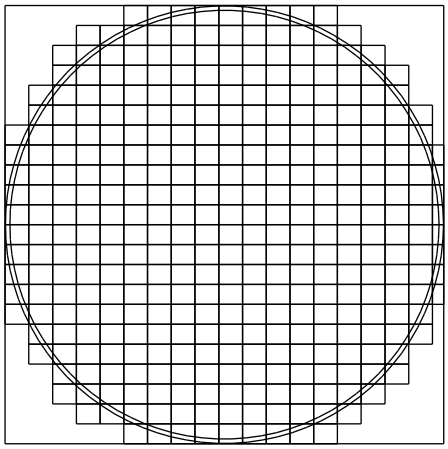


Figure 5: The ABI fixed grid block model overlays an even grid of square blocks on the full disk. Block dimensions are such that each row of blocks roughly aligns with each of the 22 ABI swaths.

In addition to defining the block configuration, the model contains a variety of critical information for each block, including (1) location identification info (ABI swath number, block number, references to pixel columns in the ABI fixed grid, and mean latitude/longitude); (2) total number of earth, fill (space), and atmosphere pixels; and (3) mean SZA.

3.5 Compressed Data Size and Rate

Different data sizing approaches were applied for each pixel type (earth/space/atmosphere) due to the variable compressibility among pixel types. Space pixels and ABI Level 1b QC bits were assumed to compress 10:1, while atmosphere pixels (in the region between 17.4 and 17.76 degrees) were assumed to compress 5:1. As described above, the earth pixel compression ratios are a function of SZA. The mean SZA for each block in the block model was input into the each band's Compressibility-SZA function to determine the earth pixel compression ratio. Using the compression information and the pixel counts for each pixel type, the following equation was used to determine total data size per block:

$$Size^{n} = \sum_{ch=1}^{16} (N^{n}_{space} \cdot B_{ch} / C_{space}) + (N^{n}_{atmos} \cdot B_{ch} / C_{atmos}) + (N^{n}_{Earth} \cdot B_{ch} / C^{n}_{earth}) + (N^{n}_{total} \cdot B_{QC} / C_{QC})$$

n is the block index, *ch* is the ABI band index, *N* is the number of each type of pixel in a block, B_{ch} is the band bit depth defined by the instrument vendor

(ITT), B_c is the ABI L1b QC bit depth of 2, C is the compression ratio per pixel type (C_{QC} is the assumed QC bit compression ratio of 10.0). The ABI data rate per block can then be easily calculated by dividing the size per block by the time it takes to scan that block, assuming a constant instrument scan rate.

Due to the large volume of data required to be sent through GRB, the L-band transmission is dual-polarized so as to double its capacity (Mazur et al, 2008). ABI data from the 16 bands are partitioned among the two GRB polarizations (Left-Hand Circular Polarization, or LHCP, and Right-Hand Circular Polarization, or RHCP), while all of the non-ABI data is sent through RHCP. Ultimately, all GOES-R data must fit within a 31 Mbps GRB bandwidth (15.5 Mbps per polarization channel).

Table 2: The data included in the left and right polarization channels during GRB transmission

LHCP	RHCP
ABI Bands 2, 7, 8, 10, 14, 15, 16 + QC	ABI Bands 1, 3, 4, 5, 6, 9, 11, 12, 13 + QC
	GLM
	SUVI
	SEISS/EXIS/MAG
	GRB Info Packets

In order to derive non-ABI data rates, we apply a set of assumptions based on preliminary analyses of compressed data rate for each GOES-R instrument. A compressed level 2 GLM data rate of 1.431 Mbps corresponds to a peak event rate of 30,000 events per second. Although the number of events per group and groups per flash will vary, analysis of data from current lightning detection platforms indicates that on average there are 10 groups per flash and 4 events per group (Mach et. al, 2007). Based on a preliminary compression analysis, we applied compression ratios of 2.4, 2.1, and 1.6 for events, groups and flashes, respectively. The compressed SUVI data rate of 1.209 Mbps assumes one solar image is transmitted every 10 seconds, where a solar image contains 1280*1280 pixels with 14 science bits plus 5 QC bits per pixel. Again, based on preliminary compression analysis, we compressed SUVI science bits 2:1 and QC bits 10:1. The uncompressed data rate contribution from SEISS/EXIS/MAG and GRB information packets was minimal (~0.143 Mbps combined), so compression is not required for these data. The total compressed non-ABI data rate of 2.772 Mbps was added to RHCP data rate for each block. In the future, a thorough compression analysis will be performed for GLM and SUVI data. Consequently, we expect to periodically update the non-ABI data rates as we learn more about the nature and compressibility of these data.

4. RESULTS AND ANALYSIS

In this section, we present the GRB data rates resulting from applying two industry standard lossless compression algorithms to the ABI proxy dataset. Each pair of plots (one for jpeg2000 and one for szip) reflects the solar conditions on the 75th Julian day, roughly corresponding to the vernal equinox. Plot times are 13Z, 17Z, and 22Z, representing morning, noon and evening full disk illumination conditions, respectively, for a GOES-East full disk centered on 75°W. Each data rate peak corresponds to an ABI swath. The data rate is variable within each swath and among different swaths for two reasons: First, the number of highly compressible space and earth atmosphere pixels is greater on the eastern and western swath edges, and in swaths near the poles. In addition, the data rate dependency on SZA results in a variable data rate across swaths: For the morning cases, data rate increases toward the eastern edge of swaths as these are the sunlit portions of the disk at that time. Peaks become more symmetric for the local noon plots, and shift to the western edge of swaths for the evening plots. These sunlight patterns are clearly evident in the shape of the data rate peaks over the course of a Mode 4 full disk scan depicted in Figures 6a-f.

Figures 6a-f: Compressed GRB data rate in LHCP (top panel) and RHCP (bottom panel) over a single 300-second full disk scan in ABI Mode 4. The horizontal line extending across each plot indicates the 15.5 Mbps channel bandwidth.

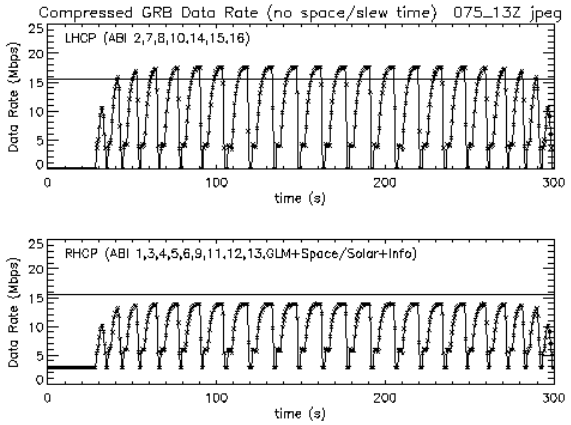


Figure 6a: jpeg2000 compression algorithm at 13Z on Julian day 75

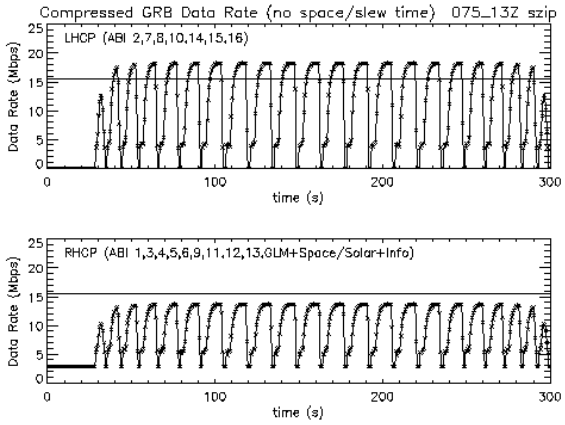


Figure 6b: szip compression algorithm at 13Z on Julian day 75

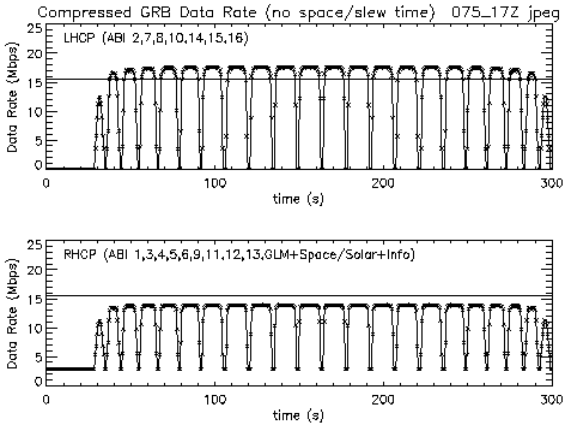


Figure 6c: jpeg2000 compression algorithm at 17Z on Julian day 75

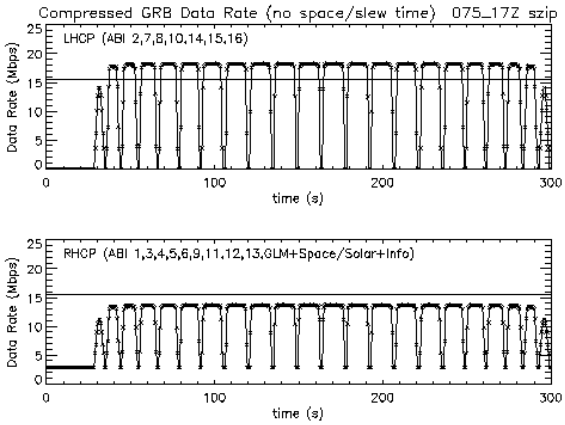


Figure 6d: szip compression algorithm at 17Z on Julian day 75

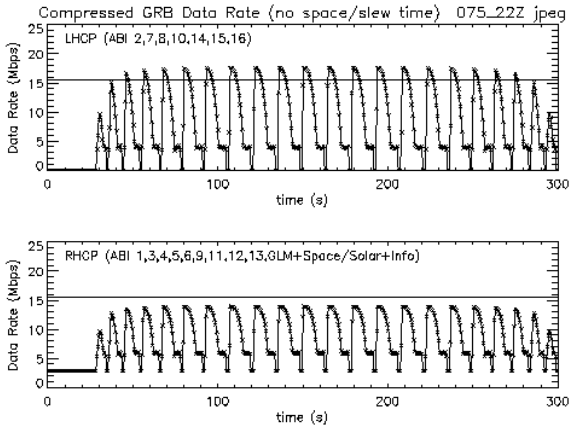


Figure 6e: jpeg2000 compression algorithm at 22Z on Julian day 75

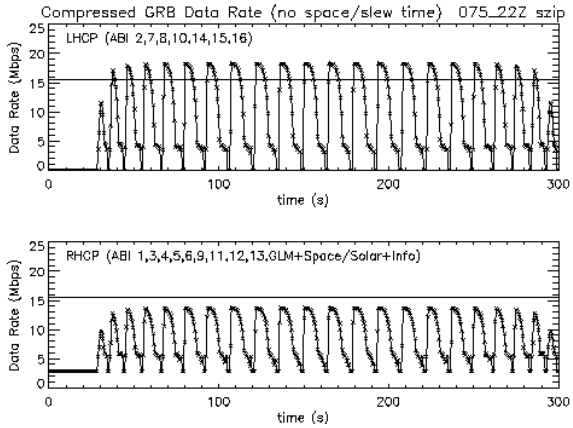


Figure 6f: szip compression algorithm at 22Z on Julian day 75

In general, the jpeg2000 algorithm compressed slightly better than szip. This was expected based on the findings of Gottipati et al. (2007), in which jpeg2000 exhibited superior compression of MODIS imagery. Regardless of the compression algorithm, the data rate peaks frequently exceeded the maximum of 15.5 Mbps in LHCP. This is primarily attributed to inclusion of the 0.5 km ABI band 2 in LHCP. Fortunately, there are many avenues for alleviating the over-peak data rates, one of which involves sending data through GRB during non-scan times (e.g. instrument slew time between full disk swaths) which would decrease the data rate peaks. Another option is to fill a data buffer with the over-peak data and transmit it through GRB during non-peak intervals in the scan cycle. This buffering activity and its impacts on meeting L1b latency requirements is part of a planned future study.

5. ONGOING AND FUTURE WORK

Section 4 summarized the results of a preliminary study involving the compressibility of GOES-R sensor data transmitted through the GOES-Rebroadcast. This section aims to summarize the extensive ongoing and planned future work involved in the GRB data compression study.

In order to more closely simulate a GOES-R full disk, we have begun a similar compression analysis using geostationary MSG-SEVIRI data. One disadvantage of the MODIS imagery is its temporal and spatial limitation. In using the compression statistics from the proxy data derived from MODIS imagery, we are essentially generalizing the compressibility of small scale image structures within the 270 MODIS granules to a full disk. In reality, compression results will hinge on global cloud cover and illumination conditions and their diurnal and seasonal variations. These patterns are not captured using a MODIS dataset, which is why the SEVIRI data may provide a more realistic basis for an ABI proxy dataset.

One additional step involved in using MSG-SEVIRI data is the addition of noise. Due in part to their coarser resolution, SEVIRI images tend to have lower entropy and thus exhibit higher compression than what is anticipated from ABI images. Theoretically, by adding noise and thereby increasing the entropy of the ABI proxy scenes created from SEVIRI data, we can more closely match the expected entropy of ABI data. Similar to MODIS, the SEVIRI images are divided into a grid of 22 by 22 blocks in order that the height of each block roughly corresponds to the expected ABI swath height. Use of the SEVIRI data is ongoing and requires further investigation.

We have also begun exploring the dependence of compression algorithm latency on block size and configuration. The goal of this study is to determine the optimal block size and shape for minimizing compression algorithm latency while maximizing block compressibility.

Future studies involve performing a similar compression analysis on the GLM and SUVI data, as well as incorporating an auto-regressive model into the ABI compression analysis that better captures spatial and temporal variations in compression. We also plan to pursue the use of a data buffer in order to dampen data rate peaks exceeding the GRB bandwidth.

6. ACKNOWLEDGEMENTS

The authors would like the thank Lizzie Lundgren, Christian Alcala and Lucas Finn at AER, as well as J. D. Warren at Harris Corporation.

7. REFERENCES

Gomez-Landsea, E., A. Rango, M. Bleiweiss (2004), An Algorithm to Address the MODIS Bowtie Effect, Can. J. Remote Sensing, 30: 644-650. <http://usda-ars.nmsu.edu/biblio/pdf/04-055.pdf>

Gottipati, Srikanth, Jamal Goddard, Michael Grossberg and Irina Gladkova, "A comparative study of lossless compression algorithms on MODIS data", Proc. SPIE 6683, 66830F (2007); doi:10.1117/12.736771

Mach, D. M., H. J. Christian, R. J. Blakeslee, D. J. Boccippio, S. J. Goodman, and W. L. Boeck (2007), Performance assessment of the Optical Transient Detector and Lightning Imaging Sensor, J. Geophys. Res., 112, D09210, doi:10.1029/2006JD007787.

Mazur, W. E., R. A. Peters, E. P. Ekelman (2008), GOES-R Direct Readout Implications, Presented at Satellite Direct Readout Conference, 11 December 2008.