13A.5 A New Method for Compressing Quality-Controlled Weather Radar Data by Exploiting Blankout Markers Due to Thresholding Operations

W. David Pan*
Dept. of Electrical & Computer Engineering, University of Alabama in Huntsville, Alabama

Paul R. Harasti, Michael Frost, Qingyun Zhao and John Cook
Marine Meteorology Division, Naval Research Laboratory, Monterey, California

Timothy Maese
Basic Commerce and Industries, Moorestown, New Jersey

Lee J. Wagner, Claude P. Hattan and Bryan T. Akagi
Atmospheric Propagation Branch, SPAWAR Systems Center, San Diego, California

1. INTRODUCTION

A meteorological radar data assimilation system has been developed at the Marine Meteorology Division of the Naval Research Laboratory (NRL) to provide environmental information to enhance the safety of ship and aircraft operations. Radar observations are assimilated into the Navy's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS®) to improve the forecasts of hazardous weather (Zhao et al. 2006; 2008) and to provide decision makers with timely products to help exploit or mitigate those predictions. The system will take advantage of a selected group of Navy ships that are to have weather processors installed for their tactical radars (e.g., SPS-48E/G: Hazardous Weather Detection and Display Capability (HWDDC); SPY-1 Tactical Environmental Processor). This group of ships will be able to digitally generate full-resolution, volumetric weather radar data every minute, and also archive them to Universal Format (UF) files (Barnes 1980). There are plans to transmit three UF files per hour in near-real-time to Fleet Numerical Meteorology and Oceanography Center (FNMOC) where the data assimilation into COAMPS® is conducted.

UF file sizes range from ~5 MB (SPS-48E) to ~13 MB (SPY-1), which would be too large a load on the operational bandwidth of the ships' communication systems. To overcome this obstacle, we have developed a novel UF file compressor that typically reduces UF file sizes by a factor of forty, thus permitting their transmission from a ship to FNMOC (Pan et al. 2009). NRL has delivered the UF file compressor to Basic Commerce and Industries, Moorestown, NJ, who is the developer of the HWDDC under contract for Space and Naval Warfare Systems Command (SPAWAR) Systems Center, Pacific, San Diego, CA. From May to August 2009, BCI and SPAWAR conducted tests using the land-based SPS-48E at Navy facilities in Dam Neck, VA on the HWDDC. One of these tests was the successful transfer of compressed UF files from the HWDDC to NRL in real-time (Harasti et al. 2009). In June 2010, SPAWAR Systems Center, Pacific, conducted further tests at their Integration Test Facility using archived and compressed SPS-48E/HWDDC UF files. These tests measured the bandwidth load during the transmission of a series of compressed UF files from a HWDDC to a simulated FNMOC Linux Server. The simulation was conducted continuously over a 24 hour period and successfully met the bandwidth requirements. Having passed all the required tests, the HWDDC has since been installed on the USS Boxer (LHD-4) and the USS Reagan (CVN-76), and ten more ships are scheduled for installations by October 2012.

Since UF files are typically analyzed in ~5 minute intervals, temporal correlations are expected to exist between radar data in neighboring UF files. Given that the current UF file compressor achieved compression by exploiting only intra-UF file correlations, we have investigated inter-UF file coding methods to exploit temporal correlations using motion estimation and compensation techniques (Pan et al. 2010). Although we developed inter-file differential compression method that is completely lossless and more than meets the 1 MB maximum file size requirement imposed by the operational bandwidth, it cannot compete with the existing intra-file data reduction and compression technique (Pan et al. 2009), in terms of overall file size reduction. On the other hand, the inter-file compression method tends to have considerably larger computational complexity than the intra-file compression method. In contrast, the work in this paper focuses on low-complexity, intra-file compression methods, with an attempt to further improve the compression efficiency of the current UF file compressor (Pan et al. 2009).

As determined by some appropriate quality control requirements on weather radar applications, thresholding operations could be employed to blank out data entries whose values are below certain thresholds and thus considered less critical. For example, as shown in Pan et al. 2009, if a raw reflectivity (DZ) data entry has a value below 5 dBZ, then this data entry would be blanked out and the corresponding data location would be labeled with a marker with a certain value. As a result of such a thresholding operation, over

*Corresponding author address: Dr. W. David Pan, Dept. of Electrical and Computer Engineering, University of Alabama in Huntsville, Huntsville, AL 35899. Email: dwpan@eng.uah.edu.

COAMPS® is a registered trademark of the Naval Research Laboratory.

40% of the DZ data entries were found to be blanked out and thus their locations were labeled with markers with identical values. While the current UF file compressor achieved a high efficiency, it did not directly exploit these markers that comprise a significantly large proportion of the data to be compressed.

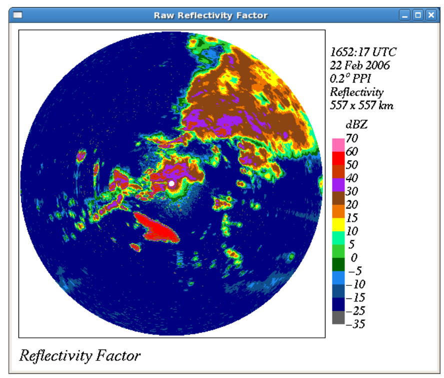
This work investigates a new data compression method, which divides the thresholded data entries into two categories: blankout markers and critical data entries, and compresses these two types of data entries separately. While markers were introduced by the thresholding operation to replace less critical data, their locations convey important spatial information that must be retained. To this end, a lossless approach is proposed to compress the spatial information and perfectly reconstruct the information. On the other hand, with the markers in the original thresholded data being skipped, the critical data entries can be accumulated and reorganized into a new chunk of data, which tend to have smaller variations in value than the original thresholded data where blankout markers were intermixed with critical data entries, thereby making higher compression on the original data possible.

We analyzed archived SPS-48E UF data obtained from an at-sea experiment onboard the USS PELELIU (LHA5) in February 2006 (Harasti et al. 2006; Maese et al. 2007). This data set contains a wide range of precipitation echoes spanning 22 hours of observations. The UF files contain full-resolution (915 m in range; 1° in azimuth) reflectivity factor (DZ), raw radial velocity (VE), de-aliased radial velocity (VD), spectrum width (SW), signal to noise ratio (SN) and valid radial velocity indicator (VV) data. These data are available within PPI scans at 22 different elevation angles ranging from 0.2° to 24° out to 275 km range, except for VE, VD, SW and VV, which are only available from the first three elevation tilts (up to 1.6° elevation) out to 52 km range. In this study, we focused on compressing the DZ data, which account for a great majority of the data entries in the UF files.

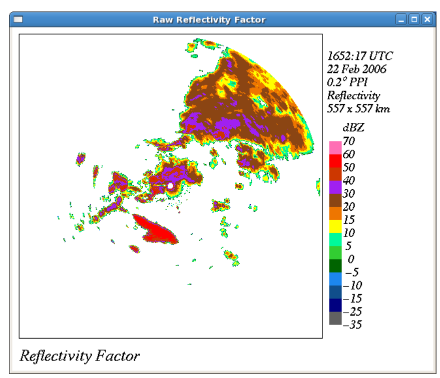
2. DATA REDUCTION WITH THRESHOLDING OPERATIONS

Several data reduction methods were employed by the UF file compressor we developed (Pan et al. 2009). For example, the original data (tape values) in the UF files were divided by the scaling factors indicated in field headers to obtain the actual data in meteorological units (Barnes 1980). The scaled down values were then rounded to the nearest integers to facilitate compression. Furthermore, thresholding operations were conducted in order to blank out data entries whose values are below certain thresholds as determined by the appropriate quality control requirements. For example, if a location has either DZ ≤ 5 dBZ or SN ≤ 10 dB, then all three types of co-located data, including DZ (reflectivity), VE (radial velocity), and SW (spectrum width), if available, will be blanked out and then labeled with an identical marker (with a value of -100). Thresholding operations create potentially long runs of identical values for the blankout markers.

After markers were introduced by the thresholding operation to replace less critical data, they were intertwined with the remaining data entries (referred to as critical data entries in this paper) in the thresholded data, which as a whole were compressed using *bzip2* by the UF file compressor. Alternatively, the markers could be separated from the critical data entries, so that one could compress these markers and the remaining critical data entries separately. The question arises as to which scheme would provide higher compression on the thresholded data (including markers and critical data entries). Our previous study on the test data set (Pan et al. 2009) showed that thresholding operations resulted in a significantly large number of blankout markers in the thresholded data. For example, Fig. 1 illustrates the greatly reduced DZ data entries after the thresholding operation for a particular UF file at 16:52:17 UTC. Fig. 2 shows that on average over 40% of the DZ data entries were blanked out due to thresholding operations.



(a)



(b)

Fig. 1. Reflectivity (DZ) data at the lowest elevation taken at 16:52:17 UTC. (a) Original data. (b) DZ data after the thresholding operation (with about 70% of the original data entries being blanked out as a result of thresholding operation).

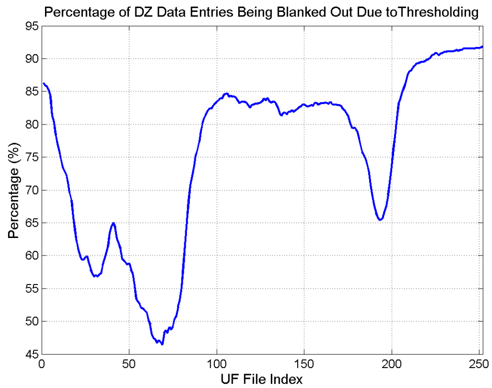


Fig. 2. Percentage of the DZ data entries that were blanked out due to the thresholding operation.

Blankout markers have the identical value of -100, which is far below the dynamic range of the critical DZ data entries. The critical data entries, after being separated from the markers, tend to be easier to compress, since getting rid of the markers would smooth out the otherwise large variations of data entry values. On the other hand, while makers carry the same value and thus carry not much information, they are not always contiguously located. Hence, their locations convey important spatial information that must be retained and compressed losslessly, if they were to be compressed separately from the critical data entries. This motivated us to develop a new method to compress the thesholded data by efficiently exploiting the blankout markers.

3. THE NEW DATA COMPRESSION METHOD

In this new method, we divide the thresholded data entries into two categories: blankout markers and critical data entries. We compress these two types of data entries separately at the encoder (see Fig. 3).

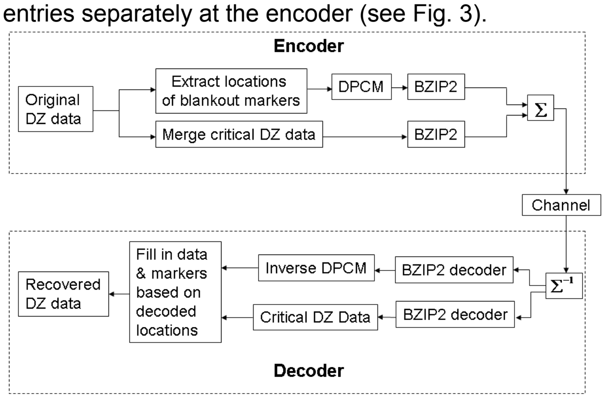


Fig. 3. The block diagrams of the proposed UF data compression algorithm.

To better explain the proposed algorithm, let us take UF file at 16:52:17 UTC as an example to illustrate the distribution of blankout markers and critical DZ data entries in the thresholded data. Fig. 4 shows the locations of these two types of data entries as bi-level images for the bottom two elevations. A one-dimensional data sequence, such as the one shown in Fig. 5, could be obtained from the original 3-D data volume, if we scan the DZ data entries in a UF file, first along the azimuth direction, then along the radial direction, and then repeat the scan at the next higher elevation, until all the elevations have been exhausted. Such a data sequence could be compressed by *bzip2* to 155,422 bytes. Rather than compressing this data sequence directly, the proposed algorithm merges the critical DZ data entries in the original sequence, after expelling the blankout markers. Consequently, a shortened data sequence free of blankout markers is derived (see Fig. 6). Note the much reduced dynamic range of this new sequence compared to the original one in Fig. 5. As expected, the critical-data only sequence can be compressed to a smaller size than the original sequence – it was compressed by *bzip2* down to 120,843 bytes.

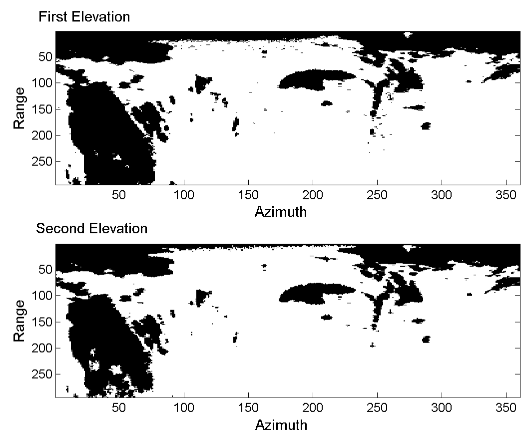


Fig. 4 Binary images showing the locations of the markers (in white) and the critical DZ data entries (in dark) at the bottom two elevations at 16:52:17 UTC.

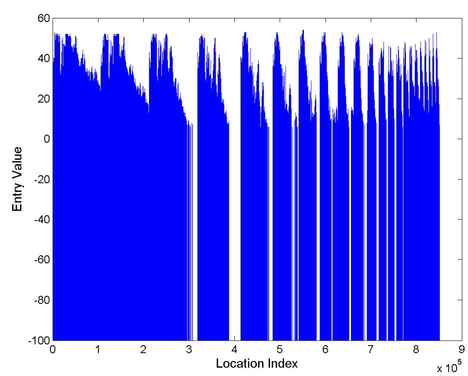


Fig. 5 The sequence of the DZ data (including critical DZ data and markers) at 16:52:17 UTC. It can be compressed to 155,422 bytes by *bzip2*.

On the other hand, with the value of the blankout markers being know *a priori*, the encoder only needs to detect and keep track of the spatial locations (or indices in the sequence) of the blankout markers when the encoder scans through the original data sequence.

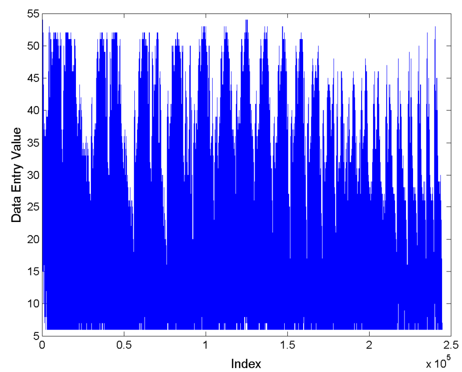


Fig. 6 . The sequence of the merged critical DZ data at 16:52:17 UTC (after markers were removed from the sequence in Fig. 5). This sequence can be compressed to 120,843 bytes by *bzip2*.

It should be noted that the locations of the critical DZ data entries need not be saved, since their locations could be inferred by the decoder, given the locations of the blankout markers. For example, in Fig. 4, the locations of the dark pixels (representing the critical DZ data entries) in the black-and-white image can be readily figured out after the coverage of the white background (corresponding to markers) is known.

In order to store or transmit the location information efficiently, a lossless approach based on an approach generally known as DPCM (Differential Pulse Code Modulation) is proposed, which compresses the location information and perfectly reconstructs the location information (Fig. 3). As an example, the sequence of blankout marker locations in the original thresholded data is shown in Fig. 7 (upper plot). It turned out to be difficult to directly compress such a sequence to a sufficiently small size that would justify the proposed approach of separately compressing the data and markers. The reason is that these sequences consist of positive integers with monotonically increasing (hence non-repeating) values that indicate the locations of blankout markers; therefore, the sequences tend to have large entropies (an information theoretic measure), making them not conducive to compression.

The redundancies existing in the sequence could be better exploited by the DPCM approach, where one codes the increments of the location indices, which tend to have much lower entropies than the original sequence, thereby making highly efficient compression possible. More specifically, given two immediately neighboring blankout markers with indices, M_i and M_{i+1}, where the i-th marker is followed by the (i+1)-th marker when one scans through the original thresholded data, the increment (D_i) of their location indices is given as $D_i = M_{i+1} - M_i$, for $i > 1$. For the first location index, we have $D_1 = M_1$. One such sequence of increments (differential location indices) is shown in Fig. 7 (lower plot), which could be efficiently compressed to 20,174 bytes by *bzip2*.

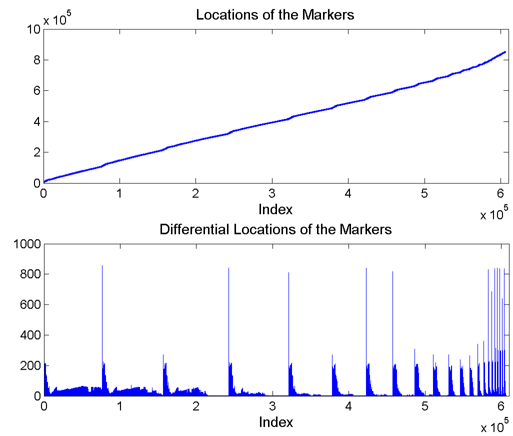


Fig. 7 . The sequence of location indices of the blankout markers (16:52:17 UTC), and the resulting location index increments after applying DPCM on the marker location indices. The increment sequence can be compressed to 20,174 bytes by *bzip2*.

At the decoder, the inverse DPCM amounts to reconstructing the location information (M_{i+1}) by recursively applying the following operation: $M_{i+1} = M_i + D_i$, for $i \geq 1$.

In summary, for this particular UF file at 16:52:17 UTC, the proposed method can compress the original thresholded DZ data to 141,017 bytes (120,843 bytes for merged critical DZ data, plus 20,174 bytes required for spatial information of the blankout markers), representing (155,422 – 141,017) / 155,422 ≈ 9% improvement over the method adopted in the UF file compressor, which directly compressed the original thresholded data where the blankout markers are intermixed with the critical data entries.

4. SIMULATION RESULTS AND DISCUSSION

We applied the proposed method on the UF radar data set and compared its compression efficiencies with the original method described in Pan et al. 2009, which directly compressed the thresholded data without separating the blankout markers from the critical DZ data. Simulations results are summarized in Fig. 8 and Fig. 9. On average, the new method achieves an improvement of approximately 3% over the original method for all UF files in the data set, with a maximal improvement of over 13%.

Fig. 9 shows that the new method achieved higher compression than the original method on approximately 80% of the UF files in the data set. The new method is slightly (approximately 1% on average) less efficient than the original method for the remaining 20% of the UF files, on which the original method employed by the current UF file compressor already achieved exceptionally high compression (to sizes well below 100 Kbytes for DZ data as shown in Fig. 8). Hence, one can argue that the degradations are negligible in terms of meeting the maximum file size requirement imposed by the operational bandwidth.

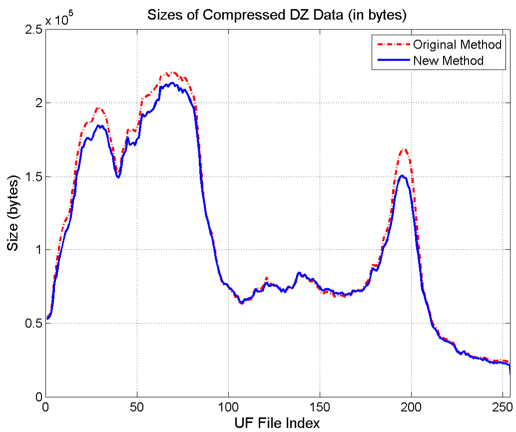


Fig. 8. Comparison of the sizes of the compressed DZ data.

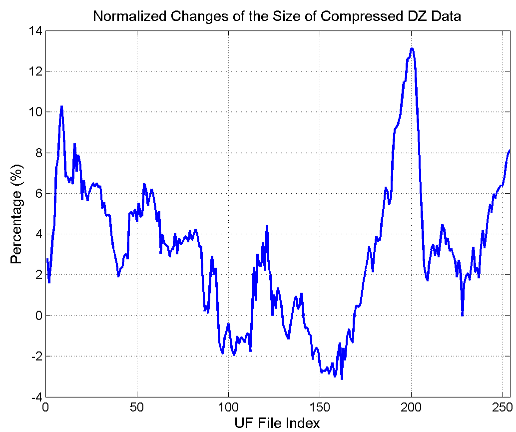


Fig. 9. Normalized size changes of the compressed DZ data, expressed in percentage as (size of original method – size of new method) / (size of original method) × 100. Positive changes mean the new method outperforms the original method, whereas negatives mean the opposite.

On the other hand, Fig. 2 shows that over 80% percent of the DZ data entries were blanked out due to the thresholding operations for the UF files in this latter group. When such a great majority of the entries in the thresholded data are blankout markers, the *bzip2* algorithm in the original method seemed to be capable of very efficiently capturing long runs of blankout markers with identical value, thereby eliminating the need to separate these predominant markers out from the original sequence. Therefore, it would be straightforward to devise an adaptive algorithm that would outperform the original method consistently, by switching between the two data compression methods, based on a concentration threshold (in terms of percentage) of the resultant blankout markers after the DZ data are thresholded.

In terms of computational efficiency, the new method is not significantly more complex than the original method, with additional cost associated with simple operations such as marker/data separation, data merging, and subtractions in DPCM operations at the encoder (Fig. 3). The decoder is slightly less complex than the encoder without the need for marker/data separation.

5. SUMMARY AND FUTURE WORK

We propose a new data compression method, which compresses the blankout markers separately from the critical data entries in the UF data files after thresholding operations. Simulations showed that the new method could provide further improvement on the compression efficiency of the current UF file compressor.

We will investigate other methods of compressing the blankout marker locations efficiently. For example, more sophisticated lossless compression methods such as those adopted in the JBIG2 standard could be considered for compressing the bi-level images in Fig. 4, with the challenge of maintaining a low overall computational complexity of the algorithm.

6. ACKNOWLEGEMENTS

The research was supported by the NRL Grant N00173091G005.

7. REFERENCES

Barnes, S. L., 1980: Report on a meeting to establish a common Doppler radar data exchange format. *Bull. Amer. Meteor. Soc.*, **61**, 1401–1404.

Harasti, P. R., Pan, W. D., M. Frost, Q. Zhao, J. Cook, L. J. Wagner, T. Maese, and R. Owens, 2009: On the development of a weather radar data assimilation system for the US Navy. *Proceedings of the 34th Conference on Radar Meteorology*, Williamsburg, VA 5-9 October, 2009, Amer. Meteor. Soc., 5 pp.

Harasti, P. R., J. Cook, Zhao, Q., L. J. Wagner, C. Kessinger, D. Megenhardt, J. Pinto, and B. Hendrickson, 2006: Extracting weather information from the SPS-48E for US Navy NOWCAST. *Proceedings of the Fourth European Conference on Radar in Meteorology and Hydrology,* Barcelona, Spain, 18-22 September 2006, 314-317.

Maese, T., J. Hunziker, H. Owen, M. Harven, L. Wagner, R. Wilcox, K. Koehler and CDR Gerald Cavalieri, 2007: Harzardous weather detection and display capability for US Navy Ships. *23rd Conference on Interactive Information and Processing Systems (IIPS)*, San Antonio, Texas, 14-18 January 2007, Amer. Meteor. Soc., 7pp.

Pan, W. D., P. R. Harasti, M. Frost, Q. Zhao, J. Cook, T. Maese, and L. J. Wagner, 2009: Efficient reduction and compression of weather radar data in universal format. *25th Conference on International Interactive*

Information and Processing Systems for Meteorology, Oceanography and Hydrology, *AMS 89th Annual Meeting*, Phoenix, Arizona, 10-16 January.

Pan, W. D., P. R. Harasti, M. Frost, Q. Zhao, J. Cook, T. Maese, and L. J. Wagner, 2010: Lossless differential compression of weather radar data in universal format using motion estimation and compensation. *26th Conference on International Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, *AMS 89th Annual Meeting*, Atlanta, Georgia, 17-21 January.

Zhao, Q., J. Cook, Q. Xu and P. R. Harasti, 2006: Using radar wind observations to improve Mesoscale Numerical Weather Prediction, *Weather and Forecasting*, **21**, 502-522.

Zhao, Q., J. Cook, Q. Xu and P. R. Harasti, 2008: Improving Short-Term Storm Predictions by Assimilating both Radar Radial-Wind and Reflectivity Observations, *Weather and Forecasting*, **23**, 373-391.