An Adaptive Predictor for Efficient Lossless Coding of FTS Interferograms

Lee A. Belfore II

Old Dominion University, Norfolk, Virginia, USA

ABSTRACT

Interferograms obtained from Fourier transform spectrometers (FTS) can be used to find high resolution optical spectra. Accompanying high resolution spectra is the corresponding large data volumes. Interferograms generally often contain informational redundancies that can be exploited in lossless coding. Inspecting interferograms shows regions that show short term periodicity. The sort term periodicity is derived by taking the Fourier transform of the points leading up to the desired prediction point. Prediction residuals are then losslessly coded with the Rice entropy coder. This paper describes efforts directed at defining a predictor that will result in efficient lossless coding of interferograms. Compared with interferograms compressed with the Rice coding algorithm, the proposed predictor shows an average improved compression of 7% across sixteen different interferograms.

Keywords: FTS, interferogram, lossless coding, predictive coding

1. INTRODUCTION

A Fourier transform spectrometer (FTS) instrument is a Michelson interferometer with a moving mirror. An interferogram is the series of light intensity measurements as a function of mirror position. It can be shown that the interferogram is the Fourier transform of the optical spectrum [1]. FTS instruments provide high spectral resolution generating in some cases providing thousands of spectral channels. The successes of this type of instrument has led to a wide range of applications that include medicine, geology, atmospheric science, remote sensing, and meteorology (find citations at Langley). While the transformed spectra are customarily used to derive the desired inferences, compression of the interferogram is appropriate where it is important to save all aspects of the measurement for further scientific analysis, analysis validation, and archive.

The Fourier transform spectrometer operates by measuring the interference pattern between two beams of light that are split from an incident light beam and that follow different length paths. The measured interferogram can be shown, under ideal circumstances, to be the Fourier transform of spectrum of the incident light. The interferogram structure has a wide dynamic range with a peak magnitude at the point of zero path difference (ZPD). In the interferogram tails, away from ZPD, the interferogram has a small magnitude compared with the magnitude at ZPD. In addition, the tails of the interferogram shows localized intervals of periodicity of varying periods.

Due to the non-ideal instrument characteristics, lossless compression of interferograms is usually desired to ensure proper retrieval of spectral content. Because the greatest magnitudes occur within a handful of measurements around ZPD, bit trimming is one compression approach that can be employed. Bit trimming is a heuristic approach whereby the number of bits allocated to an interferogram is a function of the displacement from ZPD. If the bit allocations are static, inefficiencies can result when the number of allocated bits are greater than are necessary from an information theoretic perspective. The Rice compression algorithm implemented in the Universal Source Encoder for Space [7] provides an adaptive entropy compression approach that delivers excellent compression in many applications, including FTS applications. The Rice compression often includes a one step differential predictor that frequently improves compression performance. In addition to the standard differential predictor, the USES chip allows one to input a sequence that is processed by an arbitrary predictor.

In this work, a predictor has been developed using a short period Fourier transform. The predictor is formed by extrapolating the sinusoid formed by the fundamental (greatest magnitude) frequency component from the short period Fourier transform that passes through the last known point. The next predicted point is the point on the sinusoid corresponding to the next sample. The residual, the difference between the next sample and the prediction. The

Further author information: L.A.B.: E-mail: lbelfore@odu.edu, Telephone: 757-683-3746

sequence of residuals are then coded using the Rice compression without a predictor. Experiments were performed with various lengths of Fourier transform and it was found that a length of four was found to give the best performance for the interferograms tested. The effective compression performance was found to improve over the basic Rice compression with differential prediction in all cases. The average reduction in compressed data volume was found to be 7Since the length of the short period Fourier transform was found to be short, the computational performance of the predictor is very reasonable.

This paper is organized into six sections including an introduction, a presentation of background information on FTS instruments and compression, a discussion of interferogram compression, a derivation of the predictor algorithm, results, and a summary.

2. BACKGROUND

The reader requires background in two areas to provide a suitable foundation. These areas include a description of imaging Fourier transform instruments and data compression.

2.1. Fourier Transform Spectrometer (FTS) Instruments.

Fourier transform spectrometer (FTS) instruments enable the collection of hyper-spectral data at high spectral resolution. The operating principle of an FTS instrument is based on the Michelson interferometer used in the previous century in their attempt to detect the presence of ether. Several good sources exist that describe the operating principles of the an FTS instrument (for example [1]). Here, we will only describe the very basic operational characteristics of such an instrument. The interferometer input is optical. A beam of incident light is split into two components by a beam splitter. The different beams follow different optical paths and are then recombined at the beam splitter. The recombined beam represents the interference of the two paths of the beam. By changing the path length on one beam, an interference pattern results that can be shown to be the Fourier transform of the incident light spectrum [1]. The resulting fringe pattern is converted to electrical signals at the detector and sampled according to a scheme that is appropriate for the measurement. Sampling and measuring the intensity of the light at different positions of the mirror results in a measurement, termed the interferogram, that typically has the appearance of a sinc pattern. The interferogram has a wide dynamic range peaking at a point termed zero path difference (ZPD) where the light that following the different travel exactly the same length, resulting in constructive interference of all contributing frequencies. An ideal interferogram is perfectly symmetric with respect to ZPD. Non-ideal features come from several sources including photon noise, noise in the electronics, non-ideal optics, and sampling jitter. As a result, the interferogram is no longer symmetric, and the derived spectrum is complex.

Some noise reduction can be achieved by averaging several interferograms for the same observation (termed co-adding). After co-adding, the result is quantized to a sufficient number of bits to retain the proper precision and made available to subsequent processing. From the standpoint of data compression, this interferogram is compressed.

2.2. Data Compression.

Data compression is the transformation of binary data representative representing a measurement in such a way that fewer bits are are required compared with the original information [2]. Data compression methods can be lossless or lossy. In lossless data compression, data is processed in such a way that an exact replica of the original data is produced upon decompression. Lossy compression, on the other hand, results in a loss of information upon decompression. In lossy compression, the goal is to lose in such a way that the fidelity of the desired information is not impacted. Since a lossless data compression algorithm is presented in this paper, lossless data compression is described further in general terms.

Lossless data compression either implicitly assumes a model that represents the probability of a particular symbol or empirically determines a model. From the model, variable length code words are assigned to symbols in such a way that the number of bits in the compressed data are as small as possible. For example, assuming two bit symbols, if the symbol 00 occurred with probability $p_{00} = 0.9$, the symbol 01 $p_{01} = 0.05$, the symbol 10 $p_{10} = 0.025$, and the symbol 11 $p_{10} = 0.025$, a more compact data stream can be achieved by encoding 00 with one bit and the remaining symbols with more bits based on their expected probabilities. A code book relates the symbols to their encoded versions. Because science data includes measurement error and other non-idealities, encoding symbols based on empirical probabilities often results in poor compression performance. For many types of science data, the Rice compression algorithm provides excellent compression performance. [3–5]. The Rice compression algorithm partitions the data into blocks of from 8 to 16 consecutive measurements and optimally splits measurements in such a way that the most significant bits are compressed and the least significant bits are unchanged. Furthermore, the compression of the most significant bits is performed assuming optimal Huffman coding without a code book, i.e. symbols numerically are assumed to have a Gaussian distribution. The optimal split is that point that results in the fewest number of compressed bits for a particular block. As a result of the generally recognized performance of the Rice algorithm, compression performance will be measured against it. In addition, where appropriate, the Rice compression algorithm can be a phase in the compression process. in the results presented in this paper, the Rice compression algorithm is applied where it is suitable to do so.

following duplicate??
The USES compression algorithm is based on Huffman coding but processes the data in blocks that are typically 8, 10, or 16 measurements. Within a block, the sample splitting point is selected by examining different points where the most significant portion is compressed with an implied Huffman code and the least significant bits are not modified. By blocking the data, the USES compression can adapt to changing properties of the data. Signals may have clear structures that are handled poorly by Huffman coding and USES compression. For example, one can construct a sinusoidal signal that would compress poorly using Huffman coding because the measurements are not concentrated at specific values. Basic USES compression with be the comparison of any new compression approaches. Our experience is that USES compression provides excellent compression when the data appears nearly Gaussian. As a result, the compression approach described in this paper will employ USES compression to compress residuals or modeling errors.

3. INTERFEROGRAM STRUCTURE AND IMPACT ON COMPRESSION

Examination of the interferogram suggests some opportunities and challenges for lossless coding. Interferograms can contain information redundancies that can be exploited in the context of data compression. First, spectra are band limited before being measured by the FTS instrument to eliminate aliasing of the optical frequencies [1]. As a result, an interferogram will have more points than in the final form of the data. Second, interferograms are occasionally recorded in a double sided fashion. Recording double sided interferograms doubles the number of points that must be recorded with the benefit being a better quality, lower noise spectrum [6] after calibration. As noted above, interferograms have non-ideal features that are present in the measurement. Directly taking non-ideal features is beyond the scope of this work and will likely result in reduced coding efficient when exhibited.

Figure 1 shows an interferogram that was obtained from the National Optical Astronomy Observatory (NOAO) site. The signal is characterized by a large peak at the point of ZPD and tails that have relatively small amplitude. The dynamic range of the signal is determined by the extent of ZPD. Furthermore, much of remainder of the interferogram is confined to small amplitudes. As a result, the interferogram tails have smaller entropies. Exploiting correlations between successive samples can further reduce coding entropies. Various processing and data collection effects can affect data compression however we simply accept the processed information. For example, an ideal interferogram will be symmetric about ZPD [1]. Closer examination of Figure 1.b. shows that interferogram points around ZPD are not symmetric. One source of asymmetry can result from not sampling exactly at ZPD.

As suggested above, interferometric data displays data that appears to show local periodicity. If the period can be identified, future samples can be predicted, presumably with smaller residuals. Because the data is noisy to some extent, it is impossible to achieve perfect prediction. Under the most optimistic circumstances, if the interferogram is perfectly predicted, the residual consists only of noise. For discussion, the an interferogram was retrieved from the National Optical Astronomy Observatory (NOAO) designated by 870507I0.028. The length of the interferogram over 1 million points. Fundamentally, all interferograms show the same properties including a wide dynamic range, a region around the zero path distance containing large optical intensities, and extended tails with relatively small amplitudes. Before preparing the plots the interferogram components were scaled down by a factor of four and then rounded. This processing results in an interferogram with a 16 bit dynamic range which is the maximum dynamic range allowed by the Rice software compression provided by UNM. In this form, the interferogram requires over 2 million bytes. If the device were an imaging FTS instrument, this number would be multiplied by the number of pixels that were images. Figure 1.a. shows the full interferogram. Note the characteristic center burst at ZPD and the long tails of confined to nearly uniform amplitude bounds. Figure 1.c and 1.d show the tail of the interferogram.

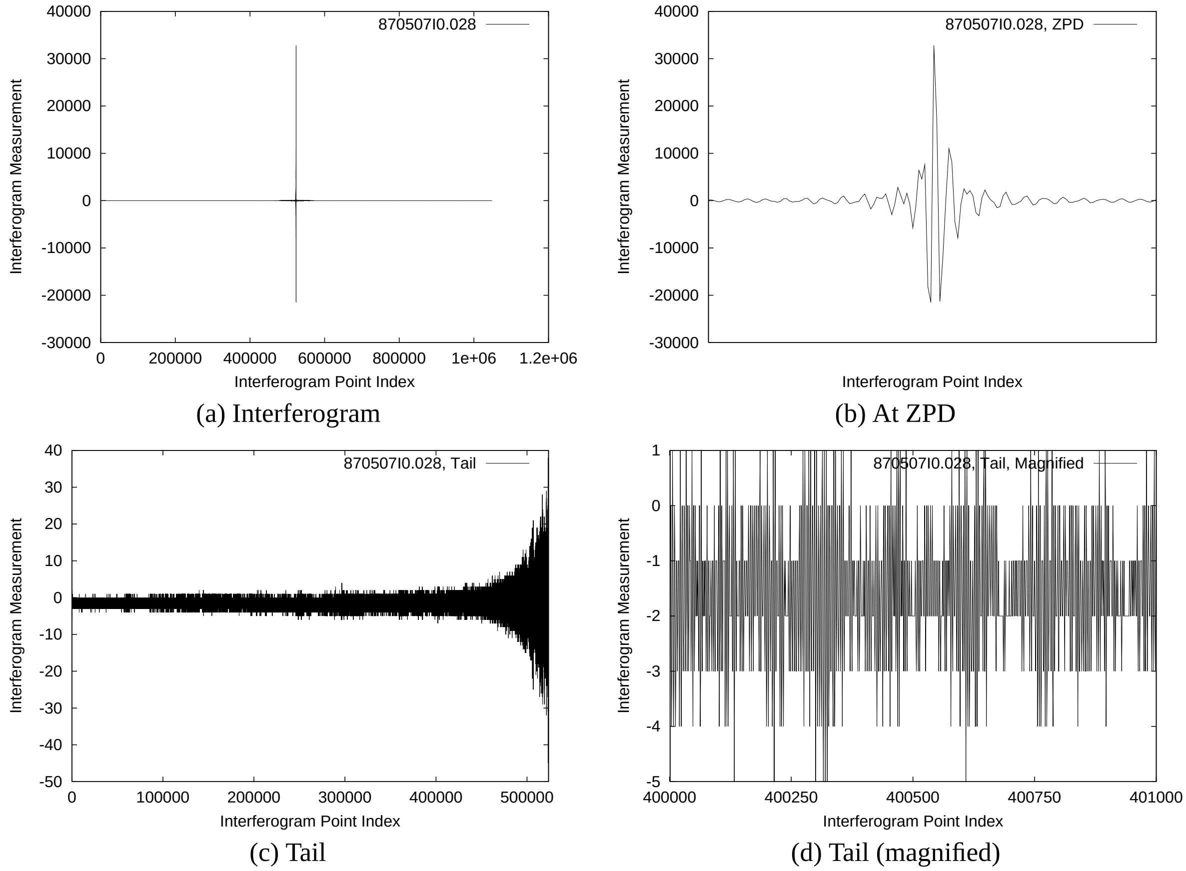

Figure 1. Different Interferogram Features (870507I0.028)

4. PREDICTION USING PSEUDO-PERIODICITY

Since the interferogram shows some structure that in some cases appears sinusoidal, our predictor model will be based on identifying representative sinusoidal components that approximate the signal locally. This frequency component is identified by taking the Fourier transform for the window of measurements leading up to the desired prediction point. From the Fourier transform, the greatest magnitude frequency component, we will call this the fundamental, is selected as the predictor. Phase information for the fundamental is taken directly from the phase for that frequency computed in the Fourier transform. As a result, the prediction can be expressed as

$$P_i = \text{round}\left(\frac{|c_f|}{M}\cos(\Omega_f\delta + \phi) + c_0\right) \tag{1}$$

where c_f is the magnitude of the fundamental frequency, M is the prediction window, Ω_f is the fundamental frequency, ϕ is the phase of the fundamental, c_0 is the average for the window and δ is the prediction time. The predictor process resulting from these discussions is shown in Figure 2. Note that the prediction for samples with indices $i \leq M$ is zero. This will have little impact on the compression process for two reasons. First, the length of the interferogram is much larger than the prediction window, $N >> M$. If poor compression results for $i \leq M$, the impact on overall compression will be small. Second, it is expected the initial set of samples will be from the tail of the interferogram. Some compression of these samples is likely as can be seen from results in Table 1.

The overall lossless coding process is shown in in Figure 3. The residual is the prediction error and is the quantity that is compressed. A good prediction will result in a small prediction error and is expected to appear nearly Gaussian. Under this assumption, the best performance results when the residual is entropy coded.

```
inputs:
    X sequence of data
    N length of data sequence
    M length of prediction

output:
    P predictions for measurements M + 1 to N

    // initialization
    δ = 2π/M // sampling interval

    for i = 1 to N
        if i ≤ M // insufficient samples to make prediction
            P_i=0
        else
            b = X(i − M to i − 1) // select the basis for prediction
            c_0 = b̄ // determine bias (zero frequency) term
            c=DFT(b − c_0) // calculate the discrete Fourier transform
            c+=positive frequency terms from c
            select c_f, the greatest fundamental component from c+
            Ω_f is the discrete frequency associated with c_f

            // determine the phase of this frequency component
            φ=arg(c_f)
            if the fundamental frequency is Ω_f zero
                P_i=round(c_0) // use the mean as the prediction
            else
                P_i=round( (|c_f|/M) cos(Ω_f δ + φ) + c_0 )
            end //if
        end //if
    end // for
```

Figure 2. Overview of Prediction Algorithm

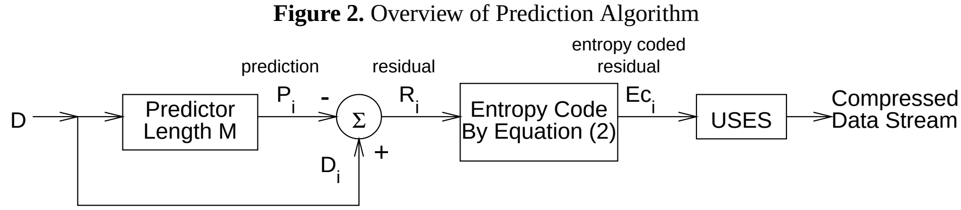

Figure 3. The Compression Process

5. COMPRESSION EXPERIMENT AND RESULTS

In this section, we describe the data that we have used to test the compression approach described in this papers. Nominally, we are comparing our coding results with obtained from the USES compression. Two variants of the USES compression form the basis for the comparison. The core capabilities of the USES algorithm is provided by the entropy coding of the input values. Once the data is in the proper form, USES compresses it directly without any additional processing. The compression performance is a function of collection of values within a block that is compressed and no correlations between values is used.

5.1. Data and Data Preparation

Several interferograms were obtained from the NOAO FTS data repository and were selected to cover a range of interferogram lengths and data collection dates. The data, itself, is in the FITS format. Individual samples in the interferograms are floating point quantities that must be converted to integral values before the compression experiment.

The USES algorithm obtained from the New Mexico Microelectronics Research Center can compress data that can be represented by, at most, 16 bits. Some of the interferometric data requires more bits. In order to make the data fit, the interferometric data is rescaled to fit within 16 bits. In cases where the USES entropy coding option is used, signed data is transformed in the following way [7]

$$ec = \begin{cases} 2d & \text{if} \quad d \geq 0 \\ -2d+1 & \text{if} \quad d < 0 \end{cases} . \quad (2)$$

The transform in Equation (2) maps a two's complement integer to an unsigned integer in such a way that small positive and negative values are transformed to small positive values. The USES compression performed this internally when compressing two's complement numbers.

5.2. USES Compression

USES compression is performed on all of the interferograms to form a baseline for comparison. Both USES entropy coding and delta prediction modes are used for the comparison. A summary of all results are presented in Table 1. Inspecting Table tb:summary reveals that 25% of the interferograms compress better in EC-mode. Furthermore, many of the remaining 75% of the interferograms compress much worse in EC-mode.

5.3. Predictor Compression Experiments

In order to analyze the performance of the proposed compression algorithm, predictors of differing lengths are run. The results can be stated in terms of a generic predictor length, or in terms of the best predictor length for a particular interferogram. This can be studied both in terms of a specific predictor length and also a predictor length chosen for a particular interferogram. Because the localized periodicity are likely related to the instrument line function and the sampling strategy, it is not expected that one predictor length will perform best across all interferograms.

Figure 4 presents the compression results as a function of predictor length. The metric of comparison is compressed number of bytes for predictor lengths from four to 64. This provides a gross estimate of performance since for the small number of interferograms in the test set, longer interferograms may unfairly bias the result. We believe that this is the likely source of the periodic pattern in compression performance. The plot does show that no one predictor length provides a coding performance that is largely worse than the USES compression, with the exception of a length 20 predictor.

As previously noted, we expect that each interferogram will have a specific predictor length that gives the best performance. Table 1 includes a summary of the best performance The results of applying the proposed predictor are presented in Table 1. The predictor performance is compared with USES compression performance in both entropy coding mode (EC-mode) and using a one step predictor, using the previous sample to predict the current sample, and then coding the residual. Comparison with the EC-mode was performed to determine how well the interferogram, itself, compiled with the USES compression. If the interferogram were a random noise sequence, EC-mode would compress better than the USES one step predictor because the predictor residual would have an increased variance compared with the original sequence.

5.4. Discussion

Residual Overflow Overflow can result if the prediction residual is too large to be represented in the number of bits available. In interferograms, this can happen in the vicinity of ZPD. In this work, we will assume that the offending sample is noted and transmitted separately to ensure proper reconstruction. In terms of the compressed data volume, the additional marking information will add little to the data volume. Among the data compression results above, the number of overflows For the interferograms we have used in this study, the number of overflows did not exceed 0.003% of residuals for any interferogram above. The worst case residual is a simple overflow, so knowledge of the position of the overflow is sufficient to reconstruct the proper value.

Predictor Length Selection The optimal predictor lengths were determined inductively, by performing compression experiments for each different predictor length. This provides the best opportunity for providing good compression performance. The disadvantage of this approach is that significant CPU resources are necessary in order to make this determination. Future work can consider developing computationally inexpensive techniques for selecting the best predictor length. In some circumstances, an instrument tuned for a particular set of observing circumstances may require that only a limited set of predictor lengths be tested.

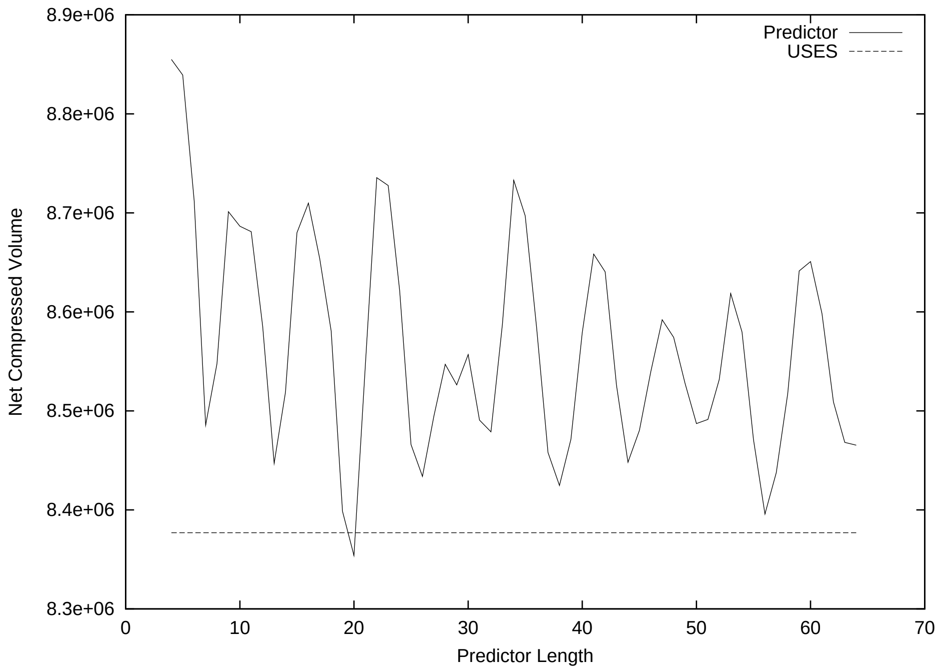

Figure 4. Compression performance versus predictor length

Cases that Compress Poorly For one interferogram, the compression performance provided marginal improvements over USES while a second showed significantly worse compression. In these cases, the proposed algorithm is clearly unsuitable. The compression performance as a function of predictor length is examined for interferogram 930205I0.015 in Figure 5. This interferogram is the longest of the interferograms used in our analysis, although we do not expect this to be detrimental to compression performance. Note the compression performance various as a function of predictor length but also has a periodic component of period 26. We interpret this as meaning that the predictor length is a best fit for the data when the compression performance is at its best. We also observe that the trend for compression performance, for the predictor lengths shown, is slowly improving with increased predictor length.

6. SUMMARY

In this paper, we have presented an adaptive predictor that exploits localized periodicities in interferograms. The predictor algorithm was presented and was compared with compression results using the Rice compression algorithm. The predictor showed an average reduction in data volume of 7% compares with a Rice coder with a delta predictor. Future work will consider issues related to improving the predictor performance and also developing methods for determining a good or optimal predictor length.

ACKNOWLEDGMENTS

This work was funded by NASA Langley Research center Contract number L70541D. NSO/Kitt Peak FTS data used here were produced by NSF/NOAO.

REFERENCES

1. R. Beer, *Remote Sensing by Fourier Transform Spectrometry*, John Wiley & Sons, 1992.
2. K. Sayood, *Introduction to Data Compression*, Morgan Kaufmann Publishers, 1996.
3. CCSDS, “Report concerning lossless data compression,” Tech. Rep. CCSDS 120.0-G-1, Consultative Committee for Space Data Systems, May 1997. Green Book.
4. CCSDS, “Recommendation for lossless data compression,” Tech. Rep. CCSDS 121.0-B-1, Consultative Committee for Space Data Systems, May 1997. Blue Book.

Table 1. Summary of best compression results

Interferogram	Uncompressed size (bytes)	USES Entropy Compression	USES with Delta Predictor	Predictor Compression	Predictor Length	Percentage Improvement
840210I0.001	2,096,544	1,195,153	1,169,971	1,098,737	10	6.5
840214I0.001	97,696	40,889	43,058	39,284	5	9.6
850725I0.035	349,600	255,288	256,472	235,785	7	8.8
850903I0.001	1,144,224	889,850	942,236	873,913	4	7.8
870507I0.028	2,098,592	504,085	379,799	303,571	5	25.1
930205I0.015	3,900,960	1,612,974	1,348,800	1,590,105	78	-14.0
981125I0.011	799,200	214,711	140,819	130,635	28	7.8
990129I0.025	236,160	94,908	87,692	83,073	7	5.6
990406I0.001	1,175,040	433,199	415,601	366,535	7	13.4
990518I0.011	280,800	86,219	82,763	78,508	10	5.4
990713I0.001	2,999,520	2,540,918	2,546,410	2,390,586	7	6.5
990811I0.001	370,080	97,443	91,868	85,097	6	8.0
991122I0.001	799,200	140,950	126,799	119,976	28	5.7
000125I0.007	711,360	200,525	182,453	165,132	6	10.5
000203I0.033	2,024,640	532,961	476,094	475,212	6	0.2
000310I0.013	280,800	89,563	86,101	81,694	10	5.4
Cumulative	18,437,943	8,929,636	8,376,936	8,096,539		3.5
Median					7	7.1
Average					14.0	7.0

5. J. Venbrux, P.-S. Yeh, and M. N. Liu, "A VLSI chip set for high-speed lossless data compression," *IEEE Trans. on Cir. and Sys. for Vid. Tech.* **2**, pp. 381–391, Dec 1992.
6. H. E. Revercomb, H. Buijs, H. B. Howell, D. D. LaPorte, W. L. Smith, and L. A. Sromovsky, "Radiometric calibration of IR Fourier transform spectrometers: Solution to a problem with the High-Resolution Interferometer Sounder," *Applied Optics* , 1988.
7. Institute of Advanced Microelectronics, Albuquerque, New Mexico, *Universal Source Encoder for Space (USES)*, November 1996.

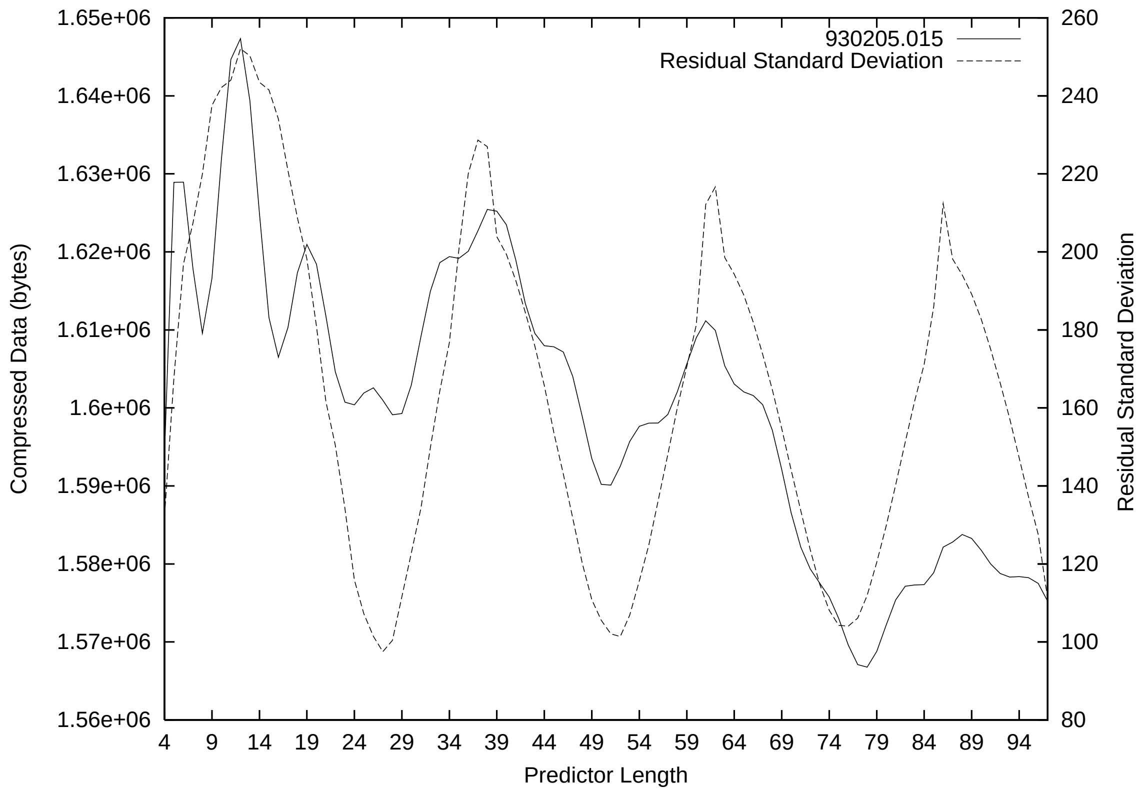

Figure 5. Compression Performance for 930205I0.015 versus Compression Performance and Residual Standard Deviation