Effects of Tunable Data Compression on Geophysical Products Retrieved from Surface Radar Observations with Applications to Spaceborne Meteorological Radars

PHILIP M. GABRIEL
Colorado State University, Fort Collins, Colorado, USA

PENSHU YEH AND SI-CHEE TSAY
NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Submitted to *Journal of Atmospheric and Oceanic Technology* on

December 9, 2013

Keywords: Radar observation; Data compression; Cloud properties; CCSDS

Corresponding Author:

Dr. Philip M. Gabriel
Colorado State University
Fort Collins, CO 80523
Tel: +1-970-491-8670
Fax: +1-970-491-8449
E-mail address: gabriel@atmos.colostate.edu

ABSTRACT

This paper presents results and analyses of applying an international space data compression standard to weather radar measurements that can easily span 8 orders of magnitude and typically require a large storage capacity as well as significant bandwidth for transmission. By varying the degree of the data compression, we analyzed the non-linear response of models that relate measured radar reflectivity and/or Doppler spectra to the moments and properties of the particle size distribution characterizing clouds and precipitation. Preliminary results for the meteorologically important phenomena of clouds and light rain indicate that for a ±0.5 dB calibration uncertainty, typical for the ground-based pulsed-Doppler 94 GHz (or 3.2 mm, W-band) weather radar used as a proxy for spaceborne radar in this study, a lossless compression ratio of only 1.2 is achievable. However, further analyses of the non-linear response of various models of rainfall rate, liquid water content and median volume diameter show that a lossy data compression ratio exceeding 15 is realizable. The exploratory analyses presented are relevant to future satellite missions, where the transmission bandwidth is premium and storage requirements of vast volumes of data, potentially problematic.

1. Introduction

Observations of atmospheric processes for the purpose of understanding, diagnosing, predicting and projecting weather and climate rely increasingly on the analysis of data from a host of instruments that include surface-based, suborbital and spaceborne radars, lidars as well as imaging spectrometers. Undoubtedly, employment of suites of instruments on either space/airborne or ground platforms will generate vast volumes of data that can quickly overwhelm data storage and transmission bandwidths. To alleviate data congestion, various approaches to data processing, editing and compression techniques have been studied. However, the most relevant question is "*if and how does the processing technique affect the end products used in understanding and predicting weather and climate*?" To address this question, we will first investigate the effects of data compression, using the Consultative Committee for Space Data Systems (CCSDS, 2005) "*Image Data Compression*" standard on ground-based, (inherently noisy) weather radar signals. Studies connected to the applications of this standard to spectroscopic observations (which span a much smaller dynamic range) have been performed (e.g., Barrie *et al*., 2009; García-Vílchez and Serra-Sagristà, 2009). However, to the best of the authors' knowledge, studies characterizing the effects of the CCSDS data compression algorithm to radar data and its derived products have not been conducted. As such, the results presented here are timely in that they demonstrate the achievable onboard compression for selected applications while underscoring the benefits of such analyses. Our characterization will provide crucial information for current (e.g., Earth Observing System, 1999) and future missions (e.g., Decadal Survey and Venture Class missions in NASA Strategic Plan, 2011).

The space data compression standard algorithm used in this study was derived by the CCSDS body composed of major international space agencies with NASA as a major partner (www.ccsds.org). Commonly known compression techniques generally fall into either the fully lossless, or the lossy categories (Sayood, 2012). The lossless technique preserves data fidelity with very limited data reduction performance while the lossy techniques with good performance require much sophisticated computation as in JPEG2000 (Taubman *et al.*, 2002). The CCSDS standard addresses space implementation constraints such as power, computation resources and a relatively high required throughput with excellent performance. Additionally it provides user precisely selectable data reduction ratio from highly lossy to full lossless, i.e., *tunable*. This feature allows flexibility in spacecraft downlink rate allocation amongst multiple science instruments. The former guarantees the restored data identical to the original; the latter generally furnishes higher compression ratios but introduces some level of distortion in the reconstructed data. This algorithm allows a user to directly control the compressed data volume or the fidelity with which the data can be reconstructed. The higher fidelity required by lossless compression results in a higher volume of compressed data for a given source data set. The compression ratio (CR) is defined as the ratio of the number of bits per sample before compression to that of the encoded data. With larger CR, the total data volume that needs to be transmitted is much reduced. For example, at CR=24, the volume is 1/12th of the volume obtained at CR=2. A larger CR not only requires less onboard storage (if needed), it is less demanding in terms of either narrower bandwidth for transmission within a fixed time frame, or a much reduced transmission time period given a fixed bandwidth. However, increasing the CR introduces increasing reconstruction noise in the decompressed data.

While lossless compression is mandatory for many types of data (e.g. complied computer codes), measurements with inherent noise need not be kept perfectly intact for transmission or storage provided the introduced distortions are below the inherent noise levels. The pressing needs for yielding higher compression ratios for certain types of applications, formulated in terms of the previously posed question is the major drive for the current study. We contend that onboard data compression of spaceborne radar, lidar and spectroscopic observations of the Earth-atmosphere system must advance in lockstep and eventually unite in an indistinguishable fashion. We envision a future in which archives of these suites of instruments output will not be monstrous dumps of data, but rather, the information mined from these data, occupying a fraction of the volume and coded in a format that is more useful to the scientific communities and to policy makers. In the meantime, it is necessary to evaluate the existing lossy compression algorithm developed for use in spaceborne platforms, applied here, to radar observations.

Because W-band radars differ in several respects from those operating at lower frequencies, we provide a brief background on their salient characteristics that are exploited in spaceborne observations of the Earth's atmosphere. W-band pulsed-Doppler radars are employed since they exhibit great sensitivity arising from the proportionality of the backscattering cross-section in the Rayleigh regime ($D \ll \lambda$) to $1/\lambda^4$, where D is the particle diameter and λ, the wavelength. Such radars are capable of detecting particles with diameters of tens of microns, typically found in clouds and in light precipitation. In addition, they can be configured to have excellent temporal and spatial resolution and can operate with physically small antennas that have a very narrow beamwidth, resulting in sampling volumes that are very small compared with those of longer wavelength radars. This reduced sampling volume decreases the effects of the Doppler spectrum broadening due to turbulence. These features of W-band radars, compounded with their portability and their ability to measure range-resolved velocities of particles, make them powerful tools for studying the macrophysics/microphysics of frequently occurring boundary-layer stratocumulus and widespread high-altitude cirrus clouds.

According to Lhermitte (1988), the deep Mie backscattering oscillations occurring in the raindrop particle size range make W-band radars an attractive choice for vertical air motion and particle size distribution measurements, particularly when used in conjunction with an S-band (e.g., 2-4 GHz) or an X-band (e.g., 8-12 GHz) radar. Furthermore, when W-band radars are used with longer wavelength radars, estimation of cloud liquid/ice water content in precipitating clouds is possible (e.g., Gaussiat *et al.*, 2003). Already, some of the stated advantages of W-band radar are being realized by the CloudSat mission (Stephens *et al*., 2002), even though in the radar employed, velocity measurement capability by the Doppler effect is absent. However, the spaceborne W-band radar to be used in the upcoming, European-Japanese EarthCare mission (Bézy *et al*., 2005) will include Doppler processing. For the reasons just discussed, it is expected that future spaceborne observation platforms will incorporate multi-frequency radars (as well as lidars and other passive instruments such as spectrometers); hence the critical need for advanced data compression techniques. Before proceeding, we acknowledge that there are significant differences between surface and spaceborne radars. The latter move at a high velocity and consequently, smear the scene below. This work addresses only the effects of data compression and not effects attributed to motion of the radar platform. We contend that as long as the complexity of the meteorological scenes is comparable, the results we obtain are transferable.

To begin to address the crucial question posed earlier, the paper is divided as follows: Section 2 describes the methodology. Here, a description is provided of the preparation of the data to be compressed using the CCSDS standard. Also included is a brief overview of the compression standard. Data products depending non-linearly on the radar reflectivity are taken from the literature (e.g., the rainfall rate, liquid water concentration and median volume diameter) and the procedure used to evaluate the effect of the standard is given. Attenuation by gaseous absorption and precipitation, as they impact the W-band radar are discussed. Section 3 shows the effects of data compression on the aforementioned products. Concluding remarks and future work are given in Section 4.

2. Methodology

2.1 Data Source

Pulsed-Doppler W-band radar signals, provided by SMARTLabs/ACHIEVE (*cf.* http://smartlabs.gsfc.nasa.gov/) mobile laboratory pictured in Fig. 1a, were acquired using a commercial receiver. The output from the receiver front-end (i.e., from the in-phase, *I*, and the quadrature components, *Q*) was digitized using a pair of 16 bits·sample^{-1} analog to digital converters running at a data rate of 50×10^6 samples·sec^{-1} and converted to double-precision reflectivity data, whose minimum discernible value is -55 dBZ at 1 km. For this study, W-band radar reflectivity measurements of a complex weather system occurring over GSFC on 8 May 2012 were used to demonstrate the performance of lossless and lossy data compression. The test-bed data shown in Fig. 1b were obtained when the W-band radar, running at a pulse repletion frequency (PRF) of 5482 Hz was zenith pointing with vertical resolution set to 24 meters in 524 range bins, for a maximum range of 12.576 km. The total observation time of 1,800 seconds is comprised of 7,709 dwell time intervals, each interval spanning 0.233 seconds; hence, 4,039,516 points constitute the reflectivity image. Furthermore, as depicted in Fig. 1b, a large dynamic range of reflectivity measurements was acquired within the duration of 30 minutes, starting at 18:27:24UTC. Retrieved cloud products (e.g., cloud top temperature, height, etc.) inferred from the overpass of MODIS sensors onboard EOS/Aqua at 18:05 UTC, indicated the presence of a large multi-layer, multi-phase (ice/melting/liquid) cloud rain system. The corresponding W-band linear depolarization ratio (LDR), shown in Fig. 1c can differentiate ice (~ -20dB), melting (~ -10dB, ice coated with water, peaking at ~3.5 km range in Fig. 1b) and water (~ -30dB) cloud phases. The mean fall velocity shown in Fig. 1d is also indicative of drizzle/rain reaching the radar site, occurring within an elapsed time of ~7.5 minutes.

Reflectivity is a measure of a radar target's efficiency in intercepting and returning radiofrequency energy that depends upon the size, shape, orientation, and dielectric properties of the target. In the meteorological context, reflectivity finds prolific use in inferring characteristics of clouds and precipitation that are fundamental, such as the particle size distribution of clouds, liquid/ice water content and rainfall/snowfall rates. While this multi-parameter radar is capable of displaying LDR and hydrometeor velocity profiles, attention here has been restricted to reflectivity data only, since such data exhibits the greatest dynamic range. It is expected that quantities characterized by a smaller dynamic range such as those shown in Figs. 1c and 1d can be compressed using larger compression ratios. Hence, the motivation of this study is to understand how perturbations introduced by lossy data compression affect derived products.

2.2 Compression Technique

The CCSDS tunable *Image Data Compression* standard employs a 2-D discrete wavelet transform (DWT) to decompose input image into wavelet coefficients. These coefficients are then selected according to their energy levels through the use of a bit plane encoder (BPE), which codes them at each bit plane. With this algorithm, users can easily, after analyzing the raw image data, make decisions on the desirable compression ratio for the image under consideration. In fact, the reallocation of such desirable compression ratio can be applied after the image has been compressed at a lower than desired compression ratio if pre-compression data analysis is un-suitable (e.g., due to unavailability of onboard processing power). The selected final higher compression ratio can be applied by simply truncating the previously compressed bit stream because of the nature of the "*embedded bit stream*" property of the algorithm. Such property guarantees that the highly compressed image information is located at the front part of the coded bit stream, followed by bit stream which improves the fidelity of the compression but lowers compression ratio. Figure 2 describes how the compression standard can be applied to facilitate optimal onboard resource utilization when data from multiple instruments have to be adjusted for downlink rate allocation. For use in the compression algorithm, the entire data set was first offset by 55 dB so that the smallest datum is 0 dBZ. A scaling factor, k, of 2^{18} was chosen to reproduce the dynamic range spanned by the reflectivity. The choice of the number of bits (here 18) is determined by the intrinsic variance of the reflectivity. However, the latter is difficult to compute because: (1) the reflectivity is not strictly a function of PRF, since the sample values from pulse-to-pulse are not independent, (2) the variance of the reflectivity is a complex function of the velocity spectral width of the cloud, (3) the variance depends on the PRF, the radar wavelength, the fast Fourier Transform (FFT) length and the number of FFT's averaged to create the power spectrum, and (4) the variance also depends on the moment estimation algorithm used to extract the signal power from the noisy power spectrum. In view of these difficulties, we have approached the problem of estimating the number of bits heuristically, using the expression in Bringi and Chandrasekar (2001) to approximate the standard deviation of the mean power ($\hat{\hat{P}}$) of a sample of N_p correlated pulses, $\sigma[\hat{\hat{P}}(dB)]$:

$$\sigma[\hat{\hat{P}}(dB)] = 10Log_{10}\Big(1+\frac{1}{\sqrt{N_p}}\sqrt{\sum_{j=-(N_p-1)}^{(N_p-1)}(1-\frac{|j|}{N_p})\exp(-\frac{16\pi^2\sigma_v^2 j^2 T_s^2}{\lambda^2})}\Big).$$

Here, $T_s=PRF^{-1}$ is the pulse repetition interval and σ_v is the standard deviation of the velocity. For a dwell time τ_d=0.233 seconds, $N_p=\tau_d*PRF$=1280 samples. The result is $\sigma[\hat{\hat{P}}(dB)]$~0.12 dB for λ=0.0032 m and σ_v=±2.5 m·sec^{-1}. The standard deviation of mean power in dB is the same as the reflectivity factor Z_e in (dBZ). To separate the effect of compression noise from the quantization noise, we introduced a noise threshold ($T\delta$) that was set to 1% of $\sigma[\hat{\hat{P}}(dB)]$, or $T\delta$ =0.0012 dB. Then, the number of bits is given by:

$b = Round\ (Log_2(DR/T\delta))$=17, where the dynamic range of the radar, DR=80 dB. We used 18 bits to guard against the possibility of clipping. Also, the resulting integer data are considered as representative of the raw, integer receiver counts.

There are two types of DWT to choose from: an integer DWT and a floating point DWT, to be noted as float DWT for brevity. Fully lossless compression can only be achieved with the integer DWT while the float DWT generally provides higher performance in tunable (i.e., lossy) applications. After applying a 2-D wavelet transform to the data, the bit-plane encoder is employed for accurate compression rate control in the lossy mode. The CCSDS algorithm has demonstrated excellent performance when applied to various types of images. However, the performance of the algorithm would degrade in the presence of large amounts of random noise. The CCSDS standard was chosen for evaluation for several reasons: first, ground-based radars can be considered as proxies for those employed in spaceborne observation platforms; secondly, the standard was created to process space instrument data with onboard processing constraints that include limited processing power and memory, as well as other effects arising from the data packetization scheme, etc. Furthermore, radiation-tolerant hardware has already been developed (e.g., Winterrowd *et al*., 2011) and integrated into NASA's mission, greatly reducing the risk and cost for future applications involving radar instruments. The results of this study therefore can serve as indicators of the expected levels of performance of data compression for spaceborne radars attainable by this algorithm.

2.3 Data Products

The SMARTLabs/ACHIEVE radar provides measurements of the horizontal and vertical components of the reflectivity, moments of the hydrometeor velocity and the linear depolarization ratio from which various meteorological products can be derived that characterize clouds and precipitation. This study focuses on the reflectivity field data product since it exhibits the largest variability and dynamic range, making it ideally suited for evaluating the data compression algorithm. The approach taken here is to use the uncompressed and compressed reflectivity fields Z ($mm^6 \cdot m^{-3}$), to derive: rainfall rate R

(mm·hr^{-1}), liquid water content W (g·m^{-3}), and median particle size D_o (cm). By comparing the results, we can investigate how non-linearities propagate error introduced by the compression/decompression process and affect the derived microphysical parameters in a way that is more insightful than merely subtracting the compressed and uncompressed reflectivity fields. To attain this objective it is necessary to introduce a set of working assumptions and to propose a model. Regarding the former, the analysis will be based on an input/output relationship $X_k=\Phi_k(Z_{\{u,c\}})$ where X_k is the derived field of interest (i.e., X=R, W, or D_o), Φ_k is the non-linear function that accepts the uncompressed or compressed reflectivity Z_u or Z_c respectively and k is a field identifying index that assumes $\{R, W, D\}$. Because interest is centered on investigating the effects of non-linearity, the function Φ_k can in principle be arbitrary. However, such arbitrariness can easily either grossly amplify the compression error, or underrepresent its effects, hence the need of a physically-based model to introduce constraints. To model the electromagnetic scattering, we use the well-known fact that radar echoes from hydrometeors depend on the moments of the particle size distribution (PSD). Knowing, the PSD allows the derivation of other products from the same PSD such as R, W, or D_o. To this end, we referred to the PSD in the seminal work of Ulbrich (1981) and Rosenfeld and Ulbrich (2003) who made significant progress in addressing the longstanding question of the connections between raindrop-size distributions and radar reflectivity-rainfall rate *(Z-R)* relationships.

The PSD we employed is the gamma distribution given as:

$$N(D) = N_o D^{\mu} Exp(-\Lambda D) \quad 0 \le D \le D_{\max}, \tag{1}$$

where D is the equivolume spherical diameter of the particles and N_o, the number concentration (m^{-3}·cm^{-1}). The slope parameter is designated by Λ (cm^{-1}), and the shape parameter, μ (dimensionless), is an exponent that can have positive or negative values. The diameter varies from zero to a maximum of D_{max}. This PSD has been considered adequate (e.g., Ulbrich, 1981; Rosenfield and Ulbrich, 2003) in characterizing precipitation since it yields simple expressions for its moments in the limit of $D_{max} \rightarrow \infty$. Table 1 illustrates how the rainfall rate (R) is related to the median particle size (D_o) and the liquid water content (W) via the reflectivity (Z) as given by Ulbrich (1981), derived from Eq. (1).

The values of the parameters: N_o, μ, b, δ, κ, A, ε, and ζ, required by the formulae in Table 1, were compiled by Ulbrich (1981) who references 23 investigations extending from (1953–2002) that characterize precipitation ranging from stratiform to convective in the form of power-law Z-R relationships. The aforementioned parameters were inferred from S-, C- and X-band radar measurements and whose values define the model parameter space used in our analyses. In this study, W-band reflectivity data were used to calculate rainfall rate, liquid water content and median volume diameter fields. The rainfall rate from each model was first computed and then propagated to calculate the liquid water content and median volume diameter, according to Table 1. These fields were then compared to those calculated from the uncompressed data. The microwave frequencies employed by the authors cited by Ulbrich (1981) differ from the W-band. However, the analyses presented are nevertheless useful in evaluating the lossy compression algorithm, considering the uncertainties introduced by the non-uniqueness of the PSD and the largest measured amplitude of the W-band reflectivity.

A search of the W-band data reveals the largest reflectivity to be 0.678 dBZ, a value that suggests the presence of light drizzle (<0.2 mm·hr^{-1}), which is consistent also with our visual observations of the event. In the absence of rain or when light drizzle is present, Rayleigh scattering by the cloud/water particles dominates as it does at the longer wavelengths used by the investigators referenced by Ulbrich (1981). The relatively small reflectivity is significant because otherwise, at the nominal frequency of 94 GHz, heavy precipitation characterized by large reflectivity would give rise to Mie scattering and be strongly attenuating, further exacerbating uncertainties in the interpretation of the rainfall rate. This contrasts with radars operating at longer wavelengths (e.g., S-, C- and X-band) where attenuation by heavy precipitation is significantly reduced.

Finally, the W-band is not an atmospheric clear window, since water vapor and oxygen are actively absorbing gases in this region of the spectrum, the former dominating the latter. Thus, the reflectivity must normally be corrected for the 2-way attenuation by the absorbing gases and by cloud/precipitation particles. Using attenuation models given by Liebe (1993), the one-way attenuation rate on the day of the measurement, by water vapor and drizzle, at the surface was calculated to be 1.5 $dB \cdot km^{-1}$ and decreasing with increasing altitude. As can be seen from Fig. 1b, light drizzle extends to approximately 1 km, thus eliminating the need for this correction. The uncertainties just described are much greater than those produced by compression noise, as will be seen. To summarize, the purpose of these analyses using 23 models of the PSD is to explore the impact of data compression noise inherent in the decompressed data on the meteorological fields previously discussed and not in accurate retrievals of the parameters characterizing an assumed PSD. To carry out this objective we employed mathematical models characterized by frequently employed power-law non-linearities (e.g., Lohmeier *et al*., 1997; Uijlenhoet, 2001) over a broad range of exponents, with particular interest in the amplification of error in R, W and D_0. The pervasiveness of power-law relationships is evident in the literature; these have even been developed to relate precipitation in the form of snowfall rate to radar reflectivity at W-band (e.g., Matrosov *et al*., 2008).

3. Results

After applying various degrees of lossy compression on the digital counts, the reconstructed reflectivity values were first compared to the original values directly from the radar. A *root-mean-square error* (*RMSE*) criterion was employed to determine the maximum lossy compression in terms of *compression ratio* (CR) corresponding to ±0.5 dB uncertainty in the radar reflectivity as it is commensurate with that introduced by the radar calibration procedure. This statistic measures the difference between reflectivity values compressed/decompressed by the CCSDS algorithm and the reflectivity values actually observed. It can also be used as a measure of error in products that are derived from the compressed reflectivities as described below. Figures 3a-3d, computed by subtracting the compressed reflectivities from the uncompressed reflectivities, show the noise introduced by float DWT and integer DWT modes of compression for different values of CR. Such pixel differences are aggregated by the *RMSE* into a single, global measure of error attributed to compression noise introduced by the CCSDS algorithm. The *RMSE* of the compressed variable Z_{CCSDS} is defined as the square root of the mean squared error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(Z_{i,j}^{obs} - Z_{i,j}^{CCSDS})^2}{NM}}, \qquad (2)$$

where $Z_{i,j}^{obs}$ is the observed reflectivity, $Z_{i,j}^{CCSDS}$ is the compressed reflectivity at the same pixel location (i, j) in the profile and NM, the number of pixel elements. The calculated *RMSE* values conveniently have the same units as the residuals. This error criterion, by virtue of the squaring process gives disproportionate weight to large errors by comparison to either the *mean absolute error (MAE)* or *mean error (ME)* that employ the size of the residual, not its square. The *ME* statistic yields a signed measure of the error and is indicative of positive/negative bias. The *MAE* criterion yields results similar in magnitude, but smaller than the *RMSE*. The reason for employing the *RMSE* is that because it is sensitive to outliers, it can be used as a diagnostic to identify the location(s) in an image where the CCSDS introduces large compression noise errors and thus gain some insight as to what properties of the image cause undesirable algorithmic behavior. We shall describe and illustrate a simple, *local* measure of bias later in this section, where it can be associated with measurements taken at a particular time. We note here that since the compression algorithm introduces bias and variance, these components are combined in the mean squared error.

From the *RMSE* curves shown in Fig. 4, it is seen that for the radar reflectivity distortion range of ±0.5 dB, a data CR of 15 is achievable. As expected, the *RMSE* increases monotonically with increasing CR and the float DWT performs slightly better than the integer DWT. The effects of the different compression ratios on other radar data products are illustrated in Figs. 5a-5c. These figures were calculated by using Eq. (2) but with $Z_{i,j}^{obs}$ and $Z_{i,j}^{CCSDS}$ now replaced by the meteorological fields of interest, derived from the unperturbed and perturbed reflectivities, respectively. In particular, for every product computed from the 23 different meteorological models, the product with the largest *RMSE*, i.e., the worst compression result, was selected for presentation. A second search was also performed to locate the minimum *RMSE*. The *RMSE* for the different products are not necessarily from the same models. In all cases, the *RMSE* errors increase with increasing CR and the results of using integer DWT are in excellent agreement with those calculated using float DWT. Taking the logarithm of the *RMSE* permits visualizing the maximum and minimum curves on the same plot.

Figures 6a and 6b addresses the question of the distribution of bias across the image at all acquisition times for CR=15. To calculate this local bias, a series of linear, least-squares regressions was performed at the different measurement times, between reflectivity values in the original and compressed image. Thus, 7,709 independent regressions were calculated using 524 points per regression. In the absence of compression noise, the resulting line must have unity slope and zero intercept. The latter is the desired measure of bias. However, as can be seen in the histogram of Fig. 6a, the CCSDS algorithm introduces a bias (for lossy compression), whose largest value of 0.225 dB falls within the ±0.5 dB imposed requirement. Figure 6b illustrates the slope and intercept at different measurement times. For all regressions, the slope is nearly unity, suggesting that the CCSDS algorithm does *not* introduce nonlinearities, further attested by the fact that the minimum correlation coefficient found is 0.995 and by the observation that variances of the uncompressed and compressed reflectivities at all measurement times lie on a 45º line as shown in Fig. 6c. Analyses of the spread of points about this line, depicted by Fig. 6d indicate that the largest difference in the standard deviations between the compressed and uncompressed reflectivities is 0.44 dB which also falls within the ±0.5 dB uncertainty in calibration. The calculations used to produce Fig. 6d account for the correlations between the compressed and uncompressed reflectivities at all sampling times.

The methods just described only produce convenient, two-point summary statistics and cannot provide information about the shape of the error distribution. Shape information can be obtained from the error histogram, but plotting such figures for reflectivities at all observation times and for all derived products is impractical. However, global plots of errors are feasible as shown in Figs. 7a-7d. The figures display histograms of the differences in Z, R, W and D_0 between uncompressed and compressed data over the image, for the model that exhibits the largest *RMSE* in these fields at a CR value of 15 using float DWT. Biases are indicated by symmetric histograms not centered at the origin or by highly asymmetric histograms that include the origin. For example, it is seen that the reflectivity exhibits a small bias, considering that out of a total of 4,039,516 points, about 600,000 are without error and that the error mass pedestal is nearly symmetric. The bias is located slightly to the left of the origin and the dynamic range of uncertainty at the base of the histogram extends from -1 to +1 dBZ. It is also seen that R and W are relatively insensitive to compression noise in the reflectivity and do not exhibit undesirable bias since the error distribution is essentially symmetrical, centered about the origin. In lossless compression, all the plots would be delta functions centered at the origin. Finally, the error distribution in the median volume diameter, D_0 exhibits a small asymmetry, with a mean of -2×10^{-5} mm and standard deviation of 2.97×10^{-4} mm, suggesting that it is more sensitive to compression noise than R or W.

4. Concluding Remarks and Future Work

From this preliminary study, it can be seen that a lossy compression ratio of at least 15 can be achieved (depending on the meteorological situation) with an acceptable radar reflectivity noise margin of ±0.5 dB. For this value of the compression ratio, the derived products incur insignificant error. When rain rate, liquid water content and median volume diameter fields are computed from the reflectivity data using 23 different models, the worst RMSE is below 10^{-3} over the full range of tested compression ratios from 2 to

24. This is significant in that for a fully lossless compression, on the contrary, a compression ratio of 1.2 is observed instead. The implication is that no appreciable data reduction can be achieved if a fully lossless compression technique is employed, and such low compression is attributed to the inherent noisy characteristics of radar signals. As long as the compression technique introduces noise in the reflectivity that is below the noise margin set by the calibration, derived products dependent on the reflectivity will be negligibly perturbed. Furthermore, the analyses presented have tacitly assumed that the radar calibration does not change during the observation period(s). Our study was performed on one set of data acquired in light drizzle and rain. To fully characterize the effects of compression on weather radar signals, extensive tests will be needed for data acquired under different weather conditions. The analyses of this dataset are not limited to reflectivity but can include polarimetric variables such as the linear depolarization ratio and the differential reflectivity. To further probe the effects of compression on meteorological products, tests will be conducted using a numerical retrieval technique to infer profiles of parameters that define the PSD in clouds and precipitation. The analyses presented have focused on a complex cloud system from which a compression ratio (i.e., 15) was derived. In the future, more comprehensive analyses will be performed for nominal terrestrial cloud systems; in turn, higher compression ratios can be expected. We note in closing that lossy data compression has not yet been fully adopted by the remote sensing community. The current perception is that employing compressed images (or data) may ultimately affect the results of posterior processing (e.g., image classification and retrieved products), potentially hindering the attainment of science goals. However, future satellite missions will certainly require the use of a suite of passive and active instru-

ments, raising the specter of bandwidth limitations and storage of unprecedented volumes of data. Thus, lossy compression may provide an effective means to mitigate these difficulties. Our approach to evaluating the effects of such compression, though preliminary, is insightful, providing a rational basis of addressing these issues.

Acknowledgements. The authors thank the continuous support of SMARTLabs deployments, as part of NASA *Radiation Sciences Program* managed by Dr. Hal B. Maring. We also thank the *Mission Planning and Technology Development* and the *Earth Observing System Project Science Office*, managed respectively by Drs. Lisa W. Callahan and Steven E. Platnick, at the NASA/GSFC Earth Sciences Division to provide partial funds for analyzing radar data. We are grateful for the insightful conversations on radar statistics with Dr. James B. Mead, president of ProSensing Inc. and to Dr. Cuong M. Nguyen at Colorado State University for the Matlab code used to perform insightful statistical experiments. Finally we extend our appreciation to the anonymous reviewers whose suggestions and comments helped us clarify the presentation and content of this paper.

5. References

Barrie, A., M. L. Adrian, P.-S. Yeh, G. Winkert, J. Lobell, A. F. Vinas, and D. G. Simpson, 2009: Fast Plasma Instrument for MMS: Data Compression Simulation Results. *AGU Fall Meeting*, SM11B-1157, San Francisco, CA.

Bézy, J.-L., W. Leibrandt, A. Hélière, P. Silvestrin, C.-C. Lin, P. Ingmann, T. Kimura, and H. Kumagai, 2005: The ESA Earth Explorer EarthCare Mission. *Proc. SPIE*, **5882**, 143–154.

Bringi, V.N. and V. Chandrasekar, 2001: *Polarimetric Doppler Weather Radar Principles and Applications*, published by Cambridge University Press.

Consultative Committee for Space Data Systems, 2005: Image Data Compression, CCSDS 122.0-B-1, Blue Book. Washington D.C., available from www.ccsds.org.

Earth Observing System, 1999: *EOS Science Plan. The State of Science in the EOS Program*. M. D. King (ed.), NASA NP-1998-12-069-GSFC, 397pp.

García-Vílchez, F., and J. Serra-Sagristà, 2009: Extending the CCSDS Recommendation for Image Data Compression for Remote Sensing Scenarios. *IEEE Trans. Geosci. Remote Sens.*, **47**, 3431–3445.

Gaussiat, N., H. Sauvageot, and A. J. Illingworth, 2003: Cloud Liquid Water and Ice Content Retrieval by Multi-wavelength Radar. *J. Atmos. Oceanic Technol*., **20**, 1264–1275.

Lhermitte, R. M., 1988: Observations of Rain at Vertical Incidence with a 94 GHz Doppler radar: An Insight of Mie Scattering. *Geophys. Res. Lett*., **15**, 1125–1128.

Liebe, H. J., G. A., Hufford, and M. G., Cotton, 1993: Propagation Modeling of Moist Air and Suspended Water/Ice Particles at Frequencies below 1000 GHz. *Proc. NATO/AGARD Wave Propagation Panel*, 52nd meeting, No. 3/1–10, Mallorca, Spain.

Lohmeier, S. P., S. M. Sekelsky, J. M. Firda, G. A. Sadowy, and R. E. McIntosh, 1997: Classification of Particles in Stratiform Clouds Using the 33 and 95 GHz Polarimetric Cloud Profiling Radar System (CPRS). *IEEE Trans. Geos. Remote Sens*., **35**, 256–270.

Matrosov, S. Y., M. D. Shupe, and I. V. Djalalova, 2008: Snowfall Retrievals Using Millimeter-Wavelength Cloud Radars. *J. Appl. Meteor. Climatol*., **47**, 769–777.

NASA Strategic Plan, 2011, NP-2011-01-699-HQ, NASA Headquarters, Washington, DC 20546 (http://www.nasa.gov/pdf/516579main_NASA2011StrategicPlan.pdf).

Rosenfield D., and W. C. Ulbrich, 2003: Cloud Microphysical Properties, Processes, and Rainfall Estimation Opportunities. *Meteor. Monographs*, **10**, 217–227.

Sayood, K., 2012: *Introduction to Data Compression*, 4th edition, published by Morgan Kaufmann, 225 Wyman St., Waltham, MA 02451, USA.

Stephens, G. L., D. G., Vane, R. J. Boain, G. G. Mace, K. Sassen, Z. Wang, A. J. Illingworth, E. J. O'Connor, W. B. Rossow, S. L. Durden, S. D. Miller, R. T. Austin, A. Benedetti, C. Mitrescu, and the Science Team, 2002: The CloudSat Mission and the A-Train: A New Dimension of Space-Based Observations of Clouds and Precipitation. *Bull. Amer. Meteor. Soc*., **83**, 1771–1790.

Taubman, D. S., and M. W. Marcellin, 2002: *JPEG2000: Image Compression Fundamentals, Standards and Practice*, published by Kluwer Academic Publishers, 101 Philip Drive, Assinippi Par, Norwell, MA 02061, USA.

Uijlenhoet, R., 2001: Raindrop Size Distributions and Radar reflectivity-Rain Rate Relationships for Radar Hydrology. *Hydrol. Earth Sys. Sci.*, **5**, 615–627.

Ulbrich, W. C., 1981: Natural Variations in the Analytical Form of the Raindrop Size Distribution. *J. Climate Appl. Met.*, **22**, 1764–1775.

Winterrowd, P., C. Orbe, J. Venbrux, S. Whitaker, E. Cameron, R. Nelson, G. Maki, D. Fisher, and P. Yeh, 2011: A 320 Mbps Flexible Image Data Compressor for Space Applications. *Proc. IEEE Aerospace Conference*, Big Sky, Montana.

Figure Captions

Figure 1. (a) Instrumentation setup of SMARTLabs-ACHIEVE (Aerosol-Cloud-Humidity Interaction Exploring & Validating Enterprise) mobile laboratory, shown a W-band cloud (94 GHz, pulsed) and X-band rain (10 GHz, FM-CW) radar mounted on a heavy-duty pedestal, with a zenith pointing K-band drizzle (24 GHz, FM-CW) radar and supplementary measurements by a ceilometer (910 nm, cloud base) mounted on the sidewall and an all-sky imager (cloud coverage), (b) an example of time series of W-band radar reflectivity collected on 8 May 2012 at NASA/GSFC, depicting drizzle and light-rain by a complex weather system passing overhead, (c) the corresponding linear depolarization ratio differentiating ice, melting, and water cloud phases, and (d) mean fall velocity indicating strongest rain occurred at ~7.5 minutes elapsed time.

Figure 2. Optimal application of the CCSDS *Image Data Compression* standard onboard: CR can be determined for instrument *i* when compression is executed with input from Path 1; or the final CR can be assigned after rate optimization is performed on multiple instruments via the Path 2 input to the down link processor (DLP). Onboard solid-state recorder (SSR) holds the coded bit stream until down link is scheduled. Readjusting CR at DLP is simply achieved by truncating the coded bit stream appropriately for each instrument.

Figure 3. The distortion of radar reflectivity produced (a) using a CR value of 2 and float DWT compression mode; (b) as in (a), except for using integer DWT compression mode; (c) as in (a), except for CR=24; and (d) as in (b), except for CR=24.

Figure 4. The root-mean-square error as a function of CR under both float and integer DWT compression modes.

Figure 5. Response of (a) rainfall, (b) liquid water content and (c) median volume diameter models to data compression noise produced by integer DWT and float DWT, illustrating the minimum and maximum root-mean-square errors. Both the upper and lower pairs of curves are practically indistinguishable; this shows that these derived products are insensitive to the chosen mode of data compression. These errors are not necessarily computed from the same model. See text for details.

Figure 6. (a) Histogram of bias variations in the radar reflectivity image for CR=15, (b) the scatter plot of their slopes at all observation times, calculated from a linear regression model for CR=15 and float DWT, (c) plot of variance comparisons computed for compressed and uncompressed reflectivities at corresponding columns of the reflectivity images, and (d) plots of variance differences computed from the compressed and uncompressed reflectivity images at corresponding column locations and accounting for correlations.

Figure 7. Global error distribution of (a) radar reflectivity, (b) rainfall rate, (c) liquid water content and (d) median volume diameter, computed for CR=15.

TABLE 1. Relationships between radar reflectivity and rainfall rate, median particle diameter and liquid water content

$Z = AR^{b}$	$A = \dfrac{10^{6}\Gamma(7+\mu)N_{o}^{-2.33/(4.67+\mu)}}{[33.31\Gamma(4.67+1)]^{(7+\mu)/(4.67+\mu)}}$	$b = \dfrac{7+\mu}{4.67+\mu}$
$D_{o} = \varepsilon R^{\delta}$	$\varepsilon = \dfrac{3.67+\mu}{[33.31 N_{o}\Gamma(4.67+1)]^{1/(4.67+\mu)}}$	$\delta = \dfrac{1}{4.67+\mu}$
$W = \zeta R^{\kappa}$	$\zeta = \dfrac{\pi\Gamma(4+\mu)N_{o}^{0.67/(4.67+\mu)}}{6[33.31\Gamma(4.67+\mu)]^{(4+\mu)/(4.67+\mu)}}$	$\kappa = \dfrac{4+\mu}{4.67+\mu}$

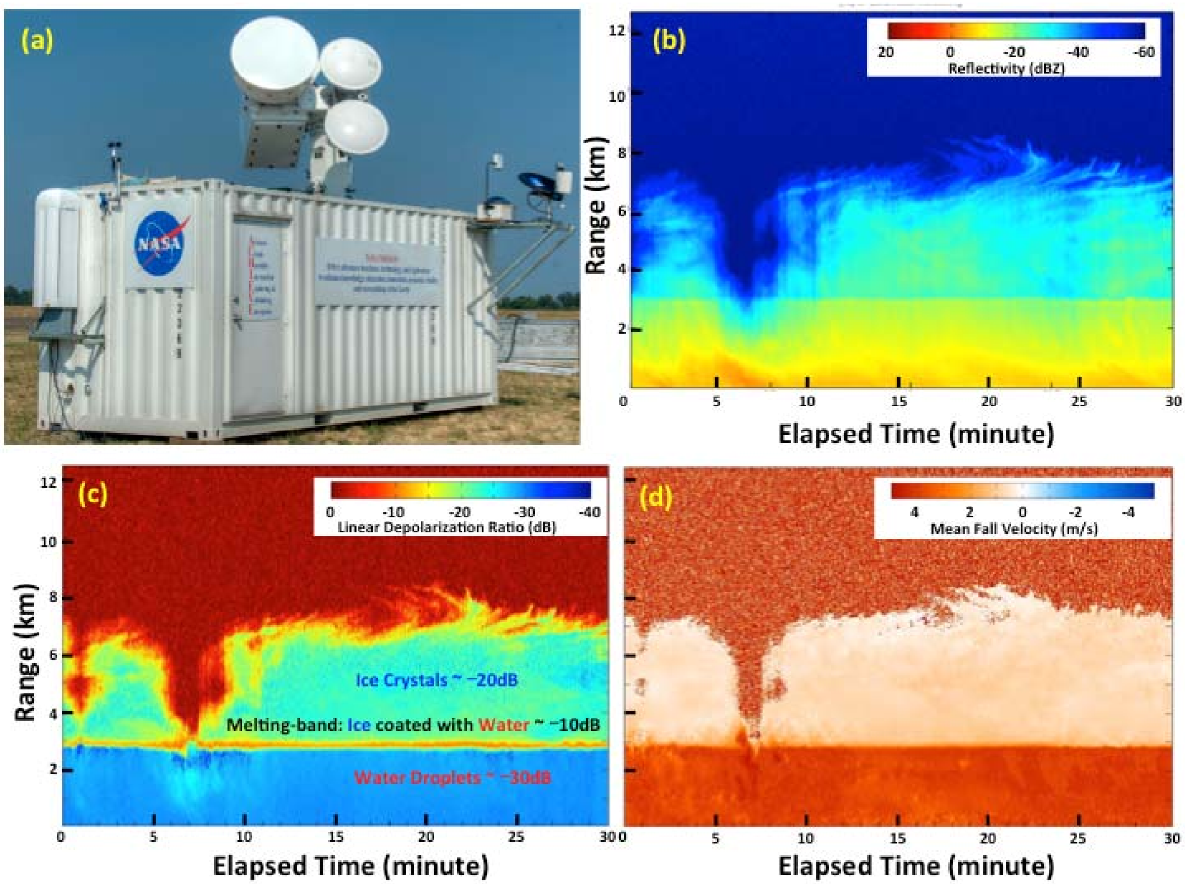

Figure 1. (a) Instrumentation setup of SMARTLabs-ACHIEVE (Aerosol-Cloud-Humidity Interaction Exploring & Validating Enterprise) mobile laboratory, shown a W-band cloud (94 GHz, pulsed) and X-band rain (10 GHz, FM-CW) radar mounted on a heavy-duty pedestal, with a zenith pointing K-band drizzle (24 GHz, FM-CW) radar and supplementary measurements by a ceilometer (910 nm, cloud base) mounted on the sidewall and an all-sky imager (cloud coverage), (b) an example of time series of W-band radar reflectivity collected on 8 May 2012 at NASA/GSFC, depicting drizzle and light-rain by a complex weather system passing overhead, (c) the corresponding linear depolarization ratio differentiating ice, melting, and water cloud phases, and (d) mean fall velocity indicating strongest rain occurred at ~7.5 minutes elapsed time.

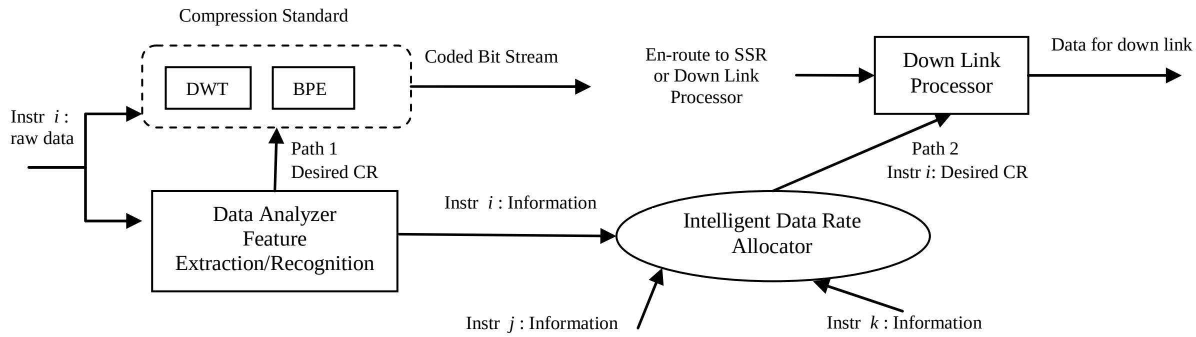

Figure 2. Optimal application of the CCSDS *Image Data Compression* standard onboard: CR can be determined for instrument *i* when compression is executed with input from Path 1; or the final CR can be assigned after rate optimization is performed on multiple instruments via the Path 2 input to the down link processor (DLP). Onboard solid-state recorder (SSR) holds the coded bit stream until down link is scheduled. Readjusting CR at DLP is simply achieved by truncating the coded bit stream appropriately for each instrument.

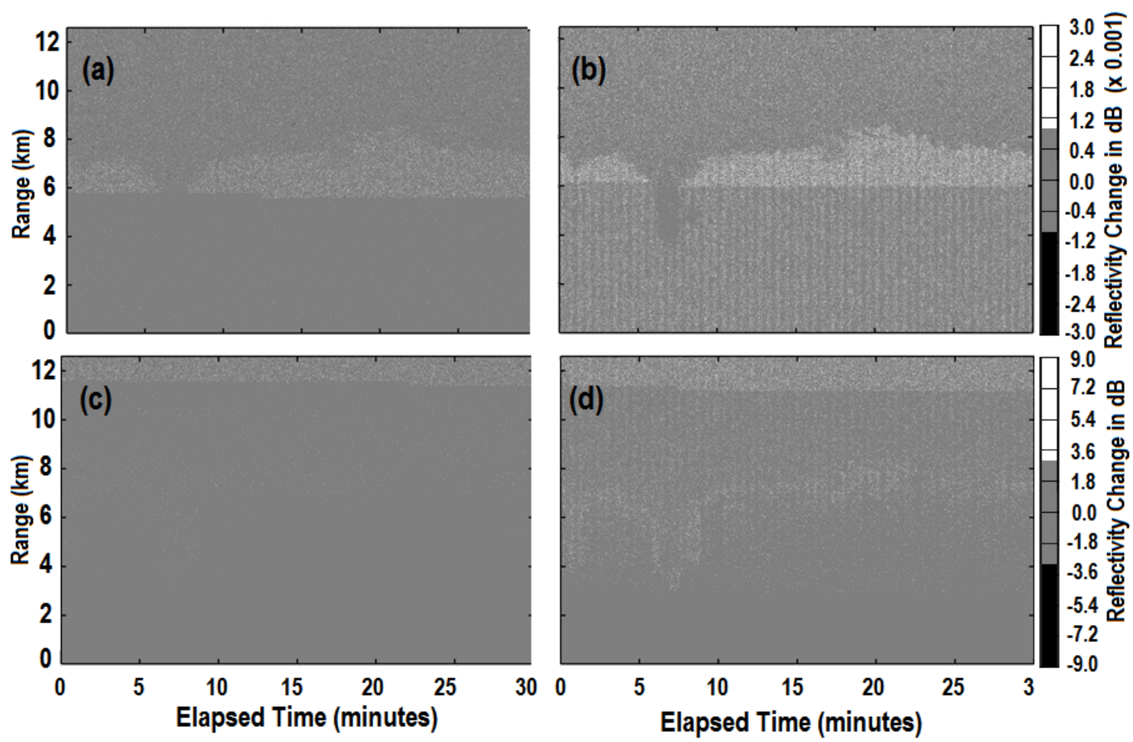

Figure 3. The distortion of radar reflectivity produced (a) using a CR value of 2 and float DWT compression mode; (b) as in (a), except for using integer DWT compression mode; (c) as in (a), except for CR=24; and (d) as in (b), except for CR=24. A low gray-level resolution clearly highlights the aforementioned distortions.

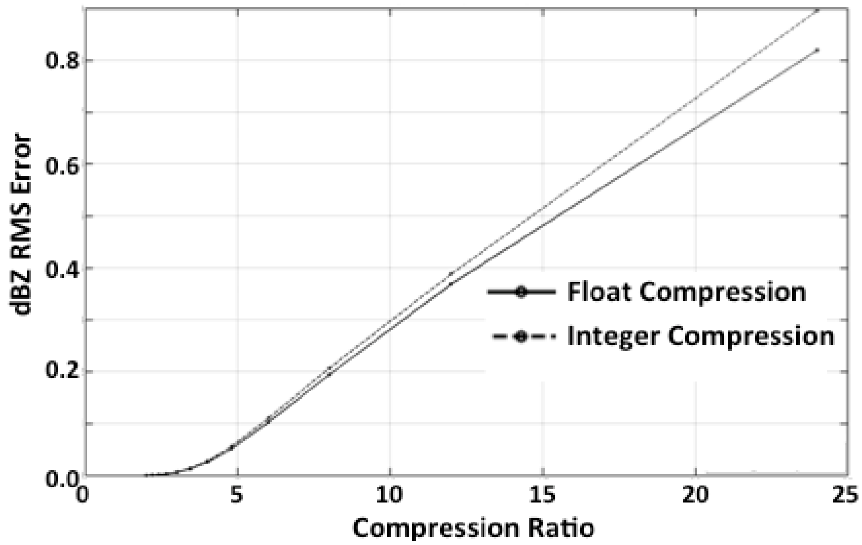

Figure 4. The root-mean-square error as a function of CR under both float and integer DWT compression modes.

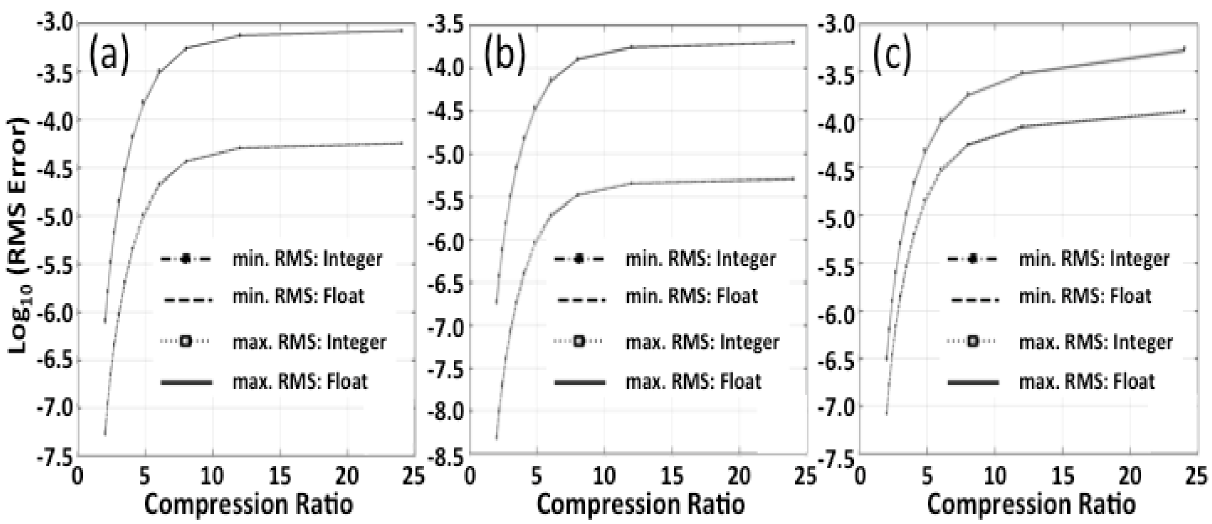

Figure 5. Response of (a) rainfall, (b) liquid water content and (c) median volume diameter models to data compression noise produced by integer DWT and float DWT, illustrating the minimum and maximum root-mean-square errors. Both the upper and lower pairs of curves are practically indistinguishable; this shows that these derived products are insensitive to the chosen mode of data compression. These errors are not necessarily computed from the same model. See text for details.

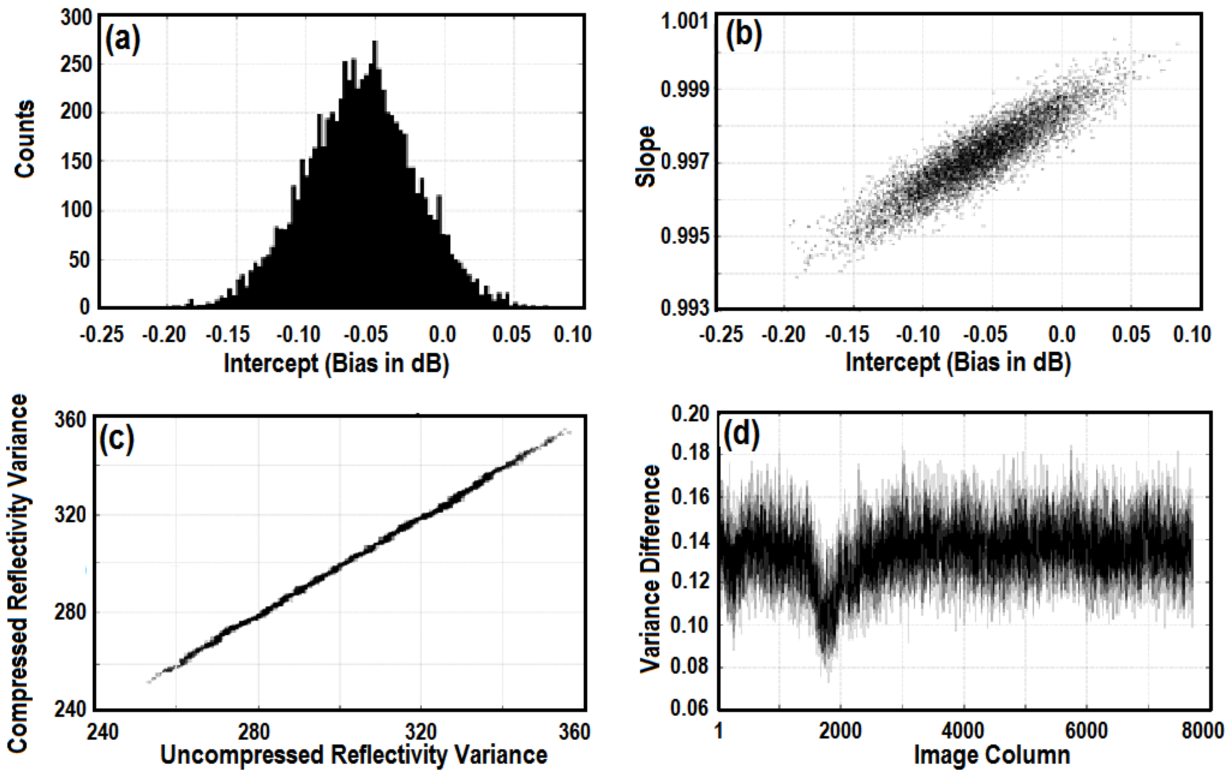

Figure 6. (a) Histogram of bias variations in the radar reflectivity for CR=15, (b) the scatter plot of their slopes at all observation times, calculated from a linear regression model for CR=15 and float DWT, (c) plot of variance comparisons computed for compressed and uncompressed reflectivities at corresponding columns of the reflectivity images, and (d) plot of variance differences computed from the compressed and uncompressed reflectivity images at corresponding column locations and accounting for correlations.

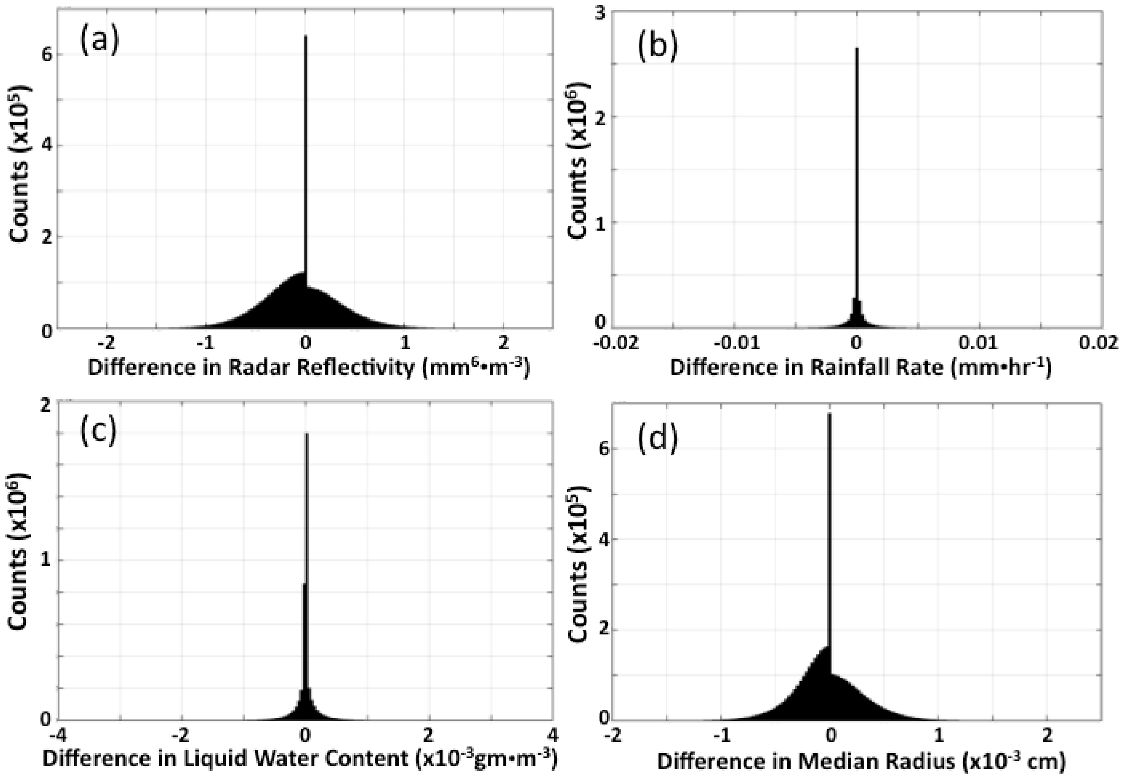

Figure 7. Global error distribution of (a) radar reflectivity, (b) rainfall rate, (c) liquid water content and (d) median volume diameter, computed for CR=15.