Toward a Near-Lossless Image Compression Strategy for the NASA/USGS Landsat Next Mission

Rehman S. Eon[a,*], Craig De Groot[b], Jeffrey A. Pedelty[c], Aaron Gerace[a], Matthew Montanaro[a], Richard K. Covington[d], Amy S. DeLisa[e], Wen-Ting Hsieh[g], Joy M. Henegar-leon[f], Douglas J. Daniels[d], Christopher Engebretson[h], Christopher J. Crawford[h], Thomas R.H. Holmes[g], Philip Dabney[g], Bruce D. Cook[g]

[a] *Rochester Institute of Technology, NY 14623 USA*
[b] *KBR Inc., Contractor to USGS EROS, SD 57198 USA*
[c] *Bay Engineering Innovations, MD 20771 USA*
[d] *Aerospace Corporation, MD 20771 USA*
[e] *FTS International, LLC., VA 20151 USA*
[f] *McCallie Associates Inc., MD 20771 USA*
[g] *NASA Goddard Space Flight Center, MD 20771 USA*
[h] *U.S. Geological Survey EROS, SD 57198 USA*

Abstract

As orbiting Earth imaging platforms carry more complex and capable instruments, efficient methods are needed to reduce the time and cost associated with storing and downlinking greater volumes of image data. The upcoming NASA/USGS Landsat Next mission, with an increase in spatial and spectral resolution over previous Landsat missions, is no exception. Landsat Next will produce nearly six times the amount of image data per day over either of the current Landsat 8 or Landsat 9 observatories. Near-lossless compression, where the image after compression is not identical to the original image, allows for the efficient storage and transmission of all image data while meeting the mission's global coverage, temporal revisit frequency, and science measurement and performance requirements. Although the Landsat user community is understandably cautious about lossy compression, it is possible to constrain the maximum loss, or error, introduced during compression, ensuring that any added error remains within the intrinsic noise level of the instrument. The Consultative Committee for Space Data Systems image compres-

*Corresponding Author
Email address: eoncis@rit.edu (Rehman S. Eon)

sion standard, CCSDS 123.0-B-2, was chosen for the Landsat Next mission because it is an internationally supported standard suited for integration with space hardware, and it allows control over the magnitude and distribution of compression error. Using several proxy datasets as a surrogate for Landsat Next image data, an investigation was performed to determine a preliminary set of parameter values that would keep the added compression error within acceptable limits. The results of these studies demonstrate that near-lossless image compression can be utilized by the Landsat Next instruments to store and downlink all science data without compromising image quality or mission requirements.

Keywords: Landsat Next, LandIS, near-lossless image compression, CCSDS 123.0-B-2

1. Introduction

Landsat Next will be the follow-on mission to the joint National Aeronautics and Space Administration and United States Geological Survey (NASA-USGS) Landsat series of satellites, and will continue the now 50+ year legacy of continuously acquiring medium resolution, multispectral imagery of the Earth's surface since the program began in 1972 [1, 2, 3, 4]. At the time of this writing, a Landsat Next observatory will carry a single imaging sensor known as the Landsat Next Instrument Suite (LandIS), which will provide a significant increase in the number of spectral bands and spatial resolution compared to previous Landsat missions (Table 1). LandIS will carry ten Landsat "heritage" multispectral bands along with five synergistic bands in common with European Sentinel 2 observatories to improve Harmonized Landsat Sentinel-2 products (HLS) [5, 6], plus an additional 11 bands to support new science applications [7], for a total of 26 spectral channels. The ground sample distance (GSD) will improve to 10 m and 20 m for the VNIR/SWIR surface reflectance bands, and to 60 m for the TIR surface temperature bands, which is a notable increase over the 30 m (VNIR/SWIR) and 100 m (TIR) bands for Landsat 8 and 9. Additionally, the bit depth of a particular band will range between 12 to 15 bits to accommodate the dynamic range for that band, which is another request of

the science community over previous Landsat data.

With the increased spatial resolution and number of spectral channels, each LandIS instrument is expected to generate up to 1.8 Gbps of raw science data, excluding embedded telemetry and metadata. Given the mission's primary goal of capturing imagery of all land surfaces and inland water bodies, which collectively cover approximately 30% of the Earth's surface, each LandIS instrument could produce nearly 6 TB of image data per day. The management of this large volume of data, particularly the storage and transmission to ground stations, presents a significant challenge for the mission architecture. Historically, the Landsat Ground Network has supported X-band frequencies with roughly a 384 Mbps downlink rate to transmit science data [8, 9]. However, with approximately 105 minutes of contact time per day, the available data transmission volume would be insufficient to fully downlink the expected daily generated data volume, leaving a significant portion of the science image data undelivered. To address this issue, a shift to higher downlink rates using Ka-band frequencies is necessary. With potential Ka-band downlink rates of 2.6 Gbps, and the same 105 minutes of contact time per day, the transmission of more than 2 TB of data per day would be possible. Although this represents a notable improvement over X-band, this Ka-band system still cannot accommodate the full volume of data generated by the LandIS instruments. While even higher bandwidth systems, such as optical communication, are being explored, they are not mature for current mission designs.

Consequently, it is essential to compress the image data from the LandIS sensors before storage and transmission. Landsat 8 and 9 employ lossless compression using the CCSDS 121.0-B-1 standard [10], which typically achieves a compression ratio of approximately 1.5-to-1 for the Operational Land Imager (OLI) data that comprise over 95% of the total mission data volume. The Thermal Infrared Sensor (TIRS) data are not compressed. Even with this level of compression, however, the data volume for the LandIS instruments would still exceed downlink capabilities. Therefore, to meet the demands of the mission, a higher compression ratio is required, necessitating the use of near-lossless compression methods.

This paper outlines the image compression methodology chosen for the Landsat

Table 1: LandIS characteristics for each spectral band. The highlighted bands indicate the current spectral bands on Landsats 8 and 9.

Band Number	Band Name	Center Wavelength $[nm]$	GSD $[m]$	Minimum Bit Depth	L_{ref} $[W/m^2/sr/\mu m]$	L_{max} $[W/m^2/sr/\mu m]$	Signal-to-Noise Ratio (SNR)
1	Violet	412	60	14	45	501	300
2	Coastal Aerosol	443	20	14	129	551	330
3	Blue	490	10	13	128	578	293
4	Green	560	10	13	128	535	315
5	Yellow	600	20	12	118	489	300
6	Orange	620	20	14	21.5	397	200
7	Red 1	650	20	14	23	380	200
8	Red 2	665	10	13	108	443	260
9	Red Edge 1	705	20	13	74.5	417	253
10	Red Edge 2	740	20	12	68	378	220
11	NIR Broad	842	10	12	103	299	270
12	NIR 1	865	20	14	52.4	282	357
13	Water Vapor	945	60	15	9	239	227
14	Liquid Water	985	20	12	18	117	130
15	Snow/Ice 1	1035	20	13	16	177	130
16	Snow/Ice 2	1090	20	13	13	142	130
17	Cirrus	1375	60	14	6	107	200
18	SWIR 1	1610	10	13	4	72.4	87
19	SWIR 2a	2038	20	13	1.6	29.7	116
20	SWIR 2b	2108	20	13	1.5	24.6	116
21	SWIR 2c	2211	20	13	1.4	22.6	116
22	TIR 1	8300	60	12	9.38	24.7	260
23	TIR 2	8600	60	12	9.62	24.5	270
24	TIR 3	9100	60	12	9.87	23.9	283
25	TIR 4	11300	60	12	9.41	19.4	350
26	TIR 5	12000	60	12	8.96	17.9	370

Next mission. The key objectives of this work involved: determining if a set of algorithm parameter values could be derived that would limit the added compression noise to within science requirements; applying the settled upon parameter values to a sufficiently diverse dataset to produce statistics to extrapolate to expected LandIS performance; and determining if the resulting compression ratios would allow the system to download all science data within the available downlink rates. An overview of the compression algorithm is discussed along with the approach taken to limit the maximum compression-induced error to meet the mission's scientific requirements. A preliminary assessment of the compression algorithm was conducted using three different proxy LandIS datasets, and the results are presented.

2. Theoretical Development

Image compression reduces file size by eliminating redundant information within an image. Key performance metrics for evaluating compression include the compression ratio (i.e., the ratio between the original and the compressed data size),

and the compression error, which quantifies the difference in pixel values between the original and the reconstructed image. The objective is to achieve a high compression ratio with minimal to no compression error. Compression with zero added error is termed *lossless*, while compression that allows error is referred to as *lossy*. Compression that constrains error to a low-level threshold approximating instrument noise is often referred to as *near lossless*.

The compression ratio for a given image depends on both the algorithm itself, which encodes the image data based on specified parameters, and on the distribution of spatial frequencies within the image. Generally, areas of low entropy (i.e., low spatial variability) in an image will result in higher compression efficiency than areas with high entropy (i.e., high spatial variability). Although image content cannot be controlled and is only partially predictable for any location on Earth, the choice of compression algorithm and its operating parameters can be optimized to achieve the desired average performance across diverse Earth scenes.

For the Landsat Next mission, the requirement for a higher compression ratio necessitates the use of near-lossless compression, where the reconstructed image may not be identical to the original at a bit level. Near-lossless compression is a departure from the typical lossless compression implemented on previous NASA science missions such as Landsat 8 and 9. The European Sentinel 2 observatories utilize a near-lossless compression standard that adjusts the added noise in real time to maintain a constant bit rate for an acceptable compression ratio for each band [11]. However, Landsat mission needs require that the added noise be known, even at the expense of a variable bit rate (i.e., variable compression ratio). The main considerations for compression for Landsat are the ability to control the maximum allowable error during compression, the feasibility of implementing the algorithm in onboard hardware, and the requirement for the algorithm to be open source to the Landsat user community.

Preliminary efforts to investigate advanced image compression strategies for Landsat missions began with the NASA/USGS Sustainable Land Imaging Architectural Study Team in 2019. Several different compression algorithms, including the well-known JPEG2000 [12] and JPEG-LS [13] (both lossless and lossy) stan-

dards, were evaluated using Landsat-like data. The CCSDS 123.0-B-2 algorithm garnered considerable interest thereafter when USGS began processing Landsat 9 datasets using an implementation of the standard known as the fast lossless extended (FLEX) code developed by NASA/Jet Propulsion Laboratory (JPL) [14, 15]. The Landsat project office ultimately adopted the CCSDS 123.0-B-2 compression standard due to its ability to meet the previously stated mission requirements regarding error control, hardware compatibility, and public availability. The FLEX implementation of the standard is currently used by the NASA Earth Surface Mineral Dust Source Investigation (EMIT) instrument onboard the ISS in lossless mode [16, 17]. Additionally, the HYPerspectral Small Satellite for Ocean Observation (HYPSO)-1 cubesat mission upgraded to CCSDS 123.0-B-2 to take advantage of the enhanced error control and compression throughput [18]. Landsat Next, however, will be the first global mapping mission to require this compression standard be implemented in near-lossless mode to maintain science integrity of its imaging data.

2.1. CCSDS 123.0-B-2 Standard

The CCSDS 123.0-B-2 standard, published by the Consultative Committee for Space Data Systems (CCSDS), provides a method for both lossless and near-lossless compression of hyperspectral and multispectral images [15, 19, 20]. A comprehensive discussion of the CCSDS 123.0-B-2 standard is beyond the scope of this paper; however, several key advantages warrant mention. The standard is specifically designed for hyperspectral data cubes to reduce image size for storage and transmission. During compression, a prediction stage uses the local 2D or 3D spatial distribution of pixel values to predict the next pixel value in the image array which improves upon the CCSDS 121.0-B-1 standard used for Landsat 8 and 9 data. Once the predicted values are calculated, an encoding stage is performed that quantizes the difference between the pixel value and the predicted pixel value. A key feature is its hybrid entropy encoder, which is designed to improve compression efficiency by switching between low- and high-entropy coding schemes based on data complexity [19, 20, 15]. These qualities make the CCSDS 123.0-B-2 compression standard attractive for highly variable Earth images.

For decompression, the algorithm reverses the encoding process to reconstruct the original image using only information embedded in the compressed data file. This self-contained feature is particularly advantageous as each compressed file includes all necessary information for image reconstruction without requiring additional information.

The algorithm has many user-controllable options that weight the predictor and quantization to affect the compression efficiency and data loss. However, the primary advantage of the algorithm is the ability to set the maximum error between the original image and the reconstructed image.

2.2. Absolute and Relative Error Modes

The standard introduces two main error modes for controlling the precision of reconstructed images during near-lossless compression: *absolute* error mode and *relative* error mode [19]. In absolute error mode, the user specifies a maximum difference, in units of digital counts, between the original image and the reconstructed image. In relative error mode, the user specifies the maximum relative error so that the ratio between error and predicted pixel value is maintained over the dynamic range (i.e., smaller predicted pixel values will have lower magnitude error while higher predicted pixel values will have higher magnitude error to maintain the same maximum percent error over all pixel values).

The absolute error mode ensures that the maximum reconstruction error in digital counts for any pixel, $m_z(t)$, does not exceed the user-specified threshold, known as the absolute error limit, a_z, as given by Equation 1,

$$m_z(t) = a_z \qquad [DN], \tag{1}$$

where the subscript z is the spectral band index and t is the pixel index and defines the permissible deviation between the original sample value and the reconstructed value after compression. The algorithm calculates the prediction residual between the original value, $s_z(t)$, and the predicted value, $\hat{s}_z(t)$, which is then quantized using a uniform quantizer, centered around the predicted value, ensuring the reconstructed value has an error no larger than a_z for any pixel [19, 15].

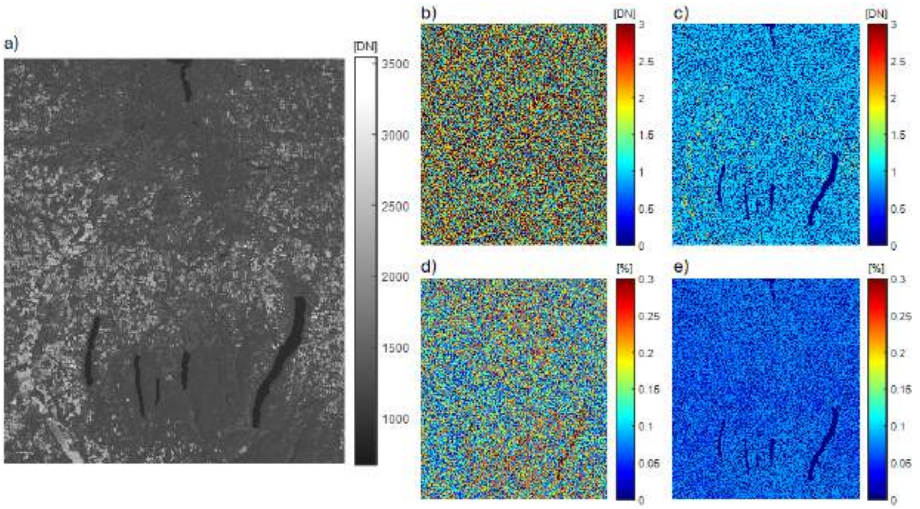

Figure 1: (a) A simulated level-0 DN image at 560 nm containing water, urban, agricultural, and rural ground cover types in the Finger Lakes region of western New York state (70 x 60 km). (b) DN error from compression using absolute error mode, and (c) relative error mode, (d) percent error from compression using absolute error mode, and (e) relative error mode.

The relative error mode similarly limits the residual between the original and reconstructed sample values except that it adjusts the allowable error based on the magnitude of the predicted pixel value. This mode is particularly advantageous when compressing images with a wide range of pixel intensities, as smaller values can be preserved with greater fidelity. The relative error limit parameter, r_z, scales the error according to the magnitude of the predicted pixel value, $\hat{s}_z(t)$, and the dynamic range in bits, D, of the original sample, ensuring that smaller values are compressed with higher accuracy, as shown in Equation 2 [21, 15, 19],

$$m_z(t) = \left\lfloor r_z \cdot \frac{|\hat{s}_z(t)|}{2^D} \right\rfloor \qquad [DN]. \tag{2}$$

The error distribution in the decompressed image varies significantly between the two error modes, as illustrated in Figure 1. Here, a simulated image at 560 nm, depicting rural, urban, and water land cover types (Figure 1a), was compressed using the two different error methods. In absolute error mode, errors are distributed randomly across the image (Figure 1b). In relative error mode, although the maxi-

mum DN error magnitude remains the same as in absolute mode, the distribution is such that lower pixel values tend to exhibit little to no error, while higher pixel values exhibit greater error in the reconstructed image (Figure 1c). The advantage of relative error mode becomes apparent when errors are expressed as percentages rather than absolute counts. Since error magnitudes are independent of the pixel value in the absolute error mode, relative (percent) errors can be quite high in certain cases, particularly over low-DN regions such as water bodies (Figure 1d). In contrast, the relative error mode adjusts the error in the reconstructed image based on the pixel value, which results in percent errors that are uniformly distributed throughout the image and are also lower in magnitude (Figure 1e). Two difference metrics were used to further quantify the performance of the two error modes. The Structural Similarity Index (SSIM) [22] and the mean squared error (MSE) were computed between the original and reconstructed images in Figure 1. Both absolute and relative error modes yielded an SSIM of 0.999, indicating virtually indistinguishable visual fidelity. The MSE values of 4.0 DN for absolute error mode and 0.7 DN for relative error mode highlight the superior pixel-level accuracy of the relative error approach.

The error magnitudes just described are illustrated as a function of pixel value in Figure 2. In absolute error mode, the errors are uniformly distributed across the dynamic range of the pixel values, leading to disproportionately large percentage errors for pixels with lower value (Figure 2a, b, c). In relative error mode, the error distribution scales with the pixel value, resulting in a more consistent percentage error across the dynamic range of the image (Figure 2d, e, f). This aspect of relative error mode minimizes the distribution of added compression error and makes it an ideal choice for performing near-lossless compression for Landsat Next.

In both error modes, the maximum allowable error can be defined. For absolute error mode, as described in Equation 1, users specify the maximum permissible deviation, a_z, in digital numbers (DN). In relative error mode, as outlined in Equation 2, users provide a relative error value, r_z, which determines the allowable DN error based on the predicted pixel value, $\hat{s}_z(t)$, and the bit depth, D.

To determine a suitable value for the r_z parameter, the error tolerance, $m_z(t)$,

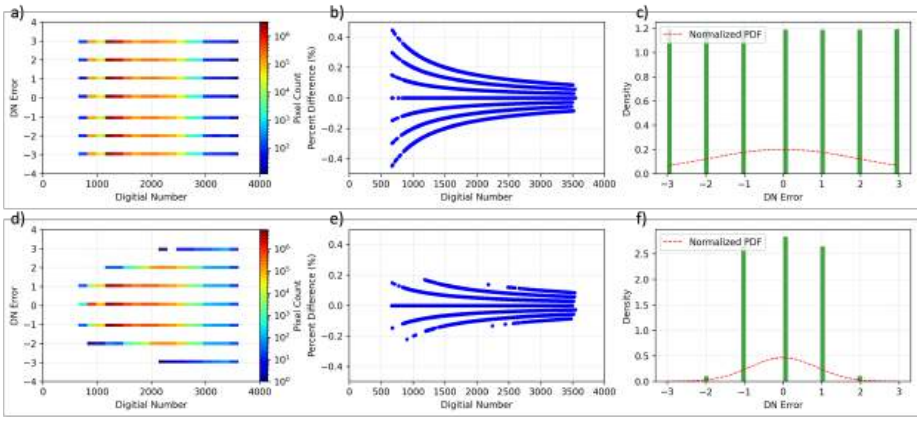

Figure 2: Top row: (a) Distribution of DN error with pixel value when compression is performed using absolute error mode, with the color indicating the frequency of each DN error value. Each data point represents one pixel in the image. (b) Distribution of percentage difference between the original and reconstructed image using absolute error mode. (c) Histogram of DN error density for absolute error. Bottom row: (d) Distribution of DN error with pixel value when compression is performed using relative error mode. (e) Distribution of percentage difference between the original and reconstructed image using relative error mode. (f) Histogram of DN error density for relative error mode.

for a given predicted value, $\hat{s}_z(t)$, must be specified. Given the scientific mission of Landsat Next, the aim is to align the added error, or noise, due to compression with the radiometric requirements for LandIS. The dynamic range of each spectral band is defined by the reference radiance, L_{ref}, the maximum radiance L_{max}, and the signal-to-noise ratio, SNR (see Table 1). The system noise, or noise equivalent change in radiance (NEdL), is calculated as the ratio of L_{ref} to the SNR (i.e., L_{ref} / SNR). When determining the necessary bit depth for each band, it was decided to follow the convention of representing the 1σ noise with 3 digital counts, where 1 DN corresponds to $1/3$ of the NEdL. For compression purposes, allowing an error of 3 counts at L_{ref} would introduce noise similar to the inherent quantization error. In relative error mode, this means setting $m_z(t) = 3$ DN and assuming that $\hat{s}_z(t) = L_{\text{ref}}$ (in units of DN) in Equation 2, resulting in a predicted error (PE) of,

$$\text{PE} = \frac{m_z(t)}{\hat{s}_z(t)} = \frac{3}{L_{\text{ref}}} \qquad [unitless]. \tag{3}$$

This assumption holds in uniform regions but may not apply near edges in the image, where the predicted pixel value $\hat{s}_z(t)$ diverges from the actual pixel value

$s_z(t)$. For the example illustrated in Figure 3, compression of $m = 3$ with a predicted error of 0.15% was applied to a Landsat 9 image of an urban area with numerous features and edges (path 190, row 026). As shown in Figure 3a, compression error remains within the specified 0.15% for nearly all pixels (99.7%). The few pixels that exceed this threshold are primarily located along edges in the image as highlighted in yellow in Figures 3b and c (left). The absolute percent differences between the original and reconstructed images are shown in Figures 3b and c (right), and illustrate that the edge pixel errors are not significantly different than non-edge pixel errors.

The CCSDS 123.0-B-2 compression standard involves many tunable parameters to allow the user to adjust the compression performance to the image dataset. As described in the next section, the algorithm was tested on a diverse set of proxy image data to derive compression performance statistics on LandIS-like data. A single set of parameter values was applied to all image sets since it is not practical to constantly change the algorithm parameters in flight for each individual image. Monte Carlo simulations determined the sensitivity of each parameter and results showed that the error mode limit parameter (either absolute or relative) had the most significant impact on compression performance. This study focused on investigating the effect of the relative error limit parameter and held all other parameters constant at their default values. The algorithm was executed on proxy LandIS datasets using the relative error mode with values of $m = 0, 1, 2, 3, 4$ and the results showed the compression error, along with compression ratio, increased as a function of m value, as expected. However, in keeping with the strategy of limiting the added compression noise to within the overall system noise (i.e., 3 DN at L_{ref}), only the results for the $m = 3$ case will be presented here. The choice of $m = 3$ is preliminary and all compression parameters are designed to be adjustable in-flight and can be fine-tuned once actual Earth data are acquired on orbit.

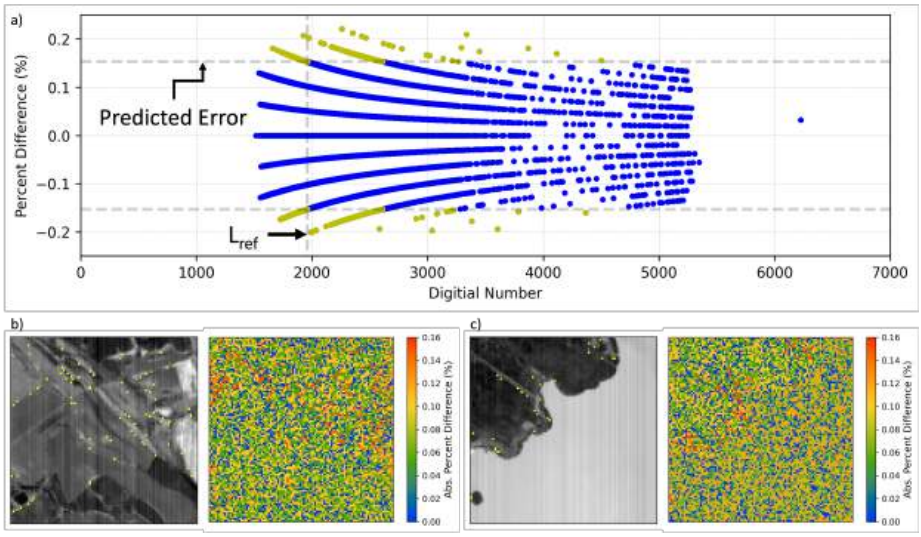

Figure 3: (a) Percentage difference between the original and reconstructed Landsat 9 image as a function of pixel value, with pixels exceeding the predicted error highlighted in yellow; (b) and (c) (left) display the locations of these higher-error pixels, shown in yellow, which primarily occur along edges. (b) and (c) (right) display the difference between the zoomed original and reconstructed images to emphasize that edge pixels are not grossly different than non-edge pixels even though edge pixels may have a slightly higher error than non-edge pixels.

3. LandIS Proxy Datasets

The expected Landsat Next dataset is unique among previous Landsat missions and other Earth science missions with 26 spectral bands spanning the visible wavelengths through the near and short wave infrared to thermal infrared at 10, 20, and 60 m GSDs and various bit depths. Proxy datasets were employed to understand the impact of near-lossless compression due to the CCSDS 123.0-B-2 standard, specifically using the relative error mode. To comprehensively evaluate compression performance across multiple scenarios, it was essential to use a large and varied dataset covering different biomes and cloud conditions. To address these requirements, the study utilized Earth image datasets from the Landsat 9 (OLI/TIRS) and EO-1 (Hyperion) missions, along with synthetic imagery from the Digital Imaging and Remote Sensing Image Generation (DIRSIG) 3D radiative transfer model. Each dataset has its own strengths and weaknesses in terms of approximating actual LandIS image data. While none of the datasets perfectly represent every condi-

tion anticipated by LandIS, their combined diversity affords broad coverage of expected scenes. As will be shown in Section 4, the compression performance across all three datasets exhibit similar statistical properties reinforcing the idea that the three datasets together adequately represent the variation expected in real LandIS data.

3.1. Landsat 9 OLI/TIRS

The current Landsat 9 (L9) bands exhibit similar range of SNR and radiometric performance characteristics to the LandIS bands. L9 captures data across 11 spectral bands ranging from 400 to 12000 nm, with a GSD of 30 m (plus a 15 m panchromatic band) for the OLI instrument and 100 m for the two TIRS bands, with a data quantization of 14 bits for OLI and 12 bits for TIRS.

The L9 archive provides a substantial source of sample data covering a wide range of spatio-temporal conditions. A key advantage of the Landsat archive is the availability of Level 0 (L0) data, which contains raw focal plane data in digital numbers in a rasterized format. The odd/even detector offsets in the L0 data were removed to spatially align the pixel data in two dimensions to make it suitable for compression using the CCSDS 123.0-B-2 standard. The limitations to using L9 imagery as a proxy for LandIS are mainly differences in sensor resolution, data quantization, and the absence of certain spectral bands.

Given the high availability of L9 Level 0 data, it is the largest proxy dataset utilized in this study. The primary L9 dataset include Worldwide Reference System 2 (WRS-2) paths 14, 109, 111, 112 and 178 over an entire calendar year (i.e., all seasons). These paths were chosen specifically to represent locations of both high and low entropy, along with the longest continuous land path to capture a range of entropy variations. The total volume of L9 data is 1780 WRS-2 scenes with a ground track width of 185 km. These images cover a wide variety of North American, European, African, and Australian landscapes, including deserts, vegetation, rain forests, urban areas, and other terrain types. The diverse range of biomes, atmospheric conditions, and cloud coverage were intended to help better understand the impact of near-lossless compression over the range of Earth conditions.

The large Landsat archive also enabled an estimation of global compression ratios across different seasons.

3.2. EO-1 Hyperion

The Earth Observing-1 (EO-1) Hyperion sensor captures hyperspectral data in 220 spectral bands, covering wavelengths from 400 to 2500 nm, with a spectral full-width at half-maximum (FWHM) of 10 nm. Hyperion operates as a pushbroom sensor following the WRS-2 path/row system producing data with a GSD of 30 m per pixel. The dataset includes Level 1A radiance images that are radiometrically calibrated and orthorectified, offering a swath width of 7.7 km [23].

The creation of proxy LandIS data from Hyperion involved several key steps. First, L1A Hyperion radiance data at a 30 m GSD is spectrally sampled to match the LandIS bands, with the relative spectral response for each band modeled as a RECT function based on the required LandIS band center and FWHM. The spectrally sampled Hyperion imagery is spatially upsampled to 10 m GSD using pan-sharpening with the 10 m EO-1/Advanced Land Imager (ALI) pan-band, to accurately represent the higher resolution of the LandIS bands. To simulate Level 0 raw DN data, the imagery is rotated according to the satellite's orbital inclination angle (98.2°), ensuring that any zero padding is removed without affecting the radiance values. The spectrally adjusted data is then converted to DN to simulate the L0 raw data as captured by the LandIS sensor, following the dynamic range parameters for each band specified in Table 1 (including L_{max}, L_{ref}, $NEdL$, and D). Finally, the imagery is spatially resampled to the LandIS GSDs using bilinear interpolation.

Bands 2 – 21 of LandIS fall completely within the spectral coverage of Hyperion and can thus be simulated as proxy data. However, it is important to note that using Hyperion to simulate proxy LandIS data has some shortcomings, including the lack of data for the thermal bands and low SNR in certain spectral regions [24]. Some LandIS bands, specifically the atmospheric bands 13 and 17, fall within low SNR regions for Hyperion and were excluded for this compression study.

The selection of Hyperion image data was closely matched to the geographic locations of L9 Paths 14, 109, 111, 112, and 178. This was done to compare the

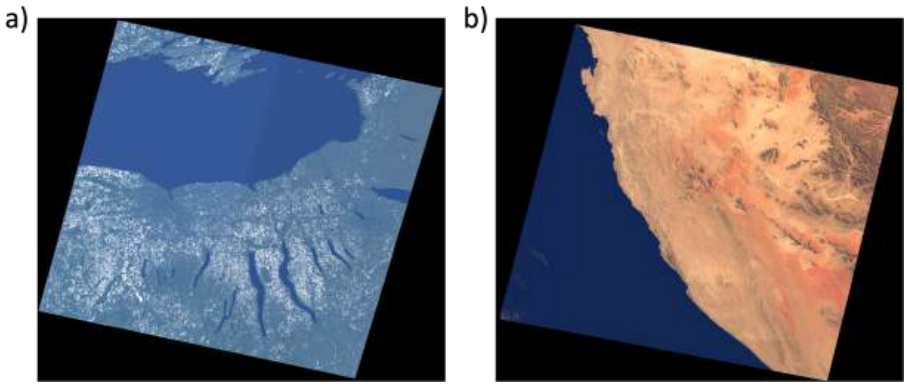

Figure 4: DIRSIG renderings (RGB) of (a) Western, NY USA (Path/Row 016/030) and (b) Namibia, Africa (Path/Row 178/079)

compression performance of Hyperion proxy-LandIS and L9 imagery over the same geographic areas. Additionally, the Hyperion time series runs from 2002 to 2011 to capture longer-term surface dynamics and conditions.

3.3. DIRSIG

The Digital Imaging and Remote Sensing Image Generation (DIRSIG) model, developed by the Rochester Institute of Technology (RIT), has been previously used to perform sensor trade-studies for Landsat 8 [25, 26, 27]. DIRSIG is a physics-based image generation tool that employs 3D radiative ray tracing to simulate image data for airborne and spaceborne platforms across the reflective and emissive portions of the EM-spectrum. DIRSIG requires three major inputs that define the simulation: a synthetic landscape, atmospheric conditions, and a sensor model. To define the scene for these compression studies, DIRSIG ingests facetized digital elevation model (DEM) data and material spectral and bidirectional reflectance characteristics for each facet. The MODTRAN radiative transfer model is used to introduce atmospheric effects into the simulation [28]. Finally, DIRSIG has built-in sensor models that can be used to image the scene or allows the user to define more complicated data-driven sensor models when focal plane measurements, such as detector line-of-sight vectors or relative spectral response functions, are available.

In this work, two large-scale simulated landscapes (185x185-km) were devel-

oped for WRS-2 Path/Row 016/030 & 178/079 (see Figure 4). One arc-second Shuttle Radar Topography Mission (SRTM) DEM data were facetized to define the digital terrain for each scene. Four Sentinel-2 10-m bands (Blue, Green, Red, NIR) were used as texture (class) maps, which function as a classification mechanism to sample from a spectral database of materials for the pixels of interest. Although the exact LandIS sensor architecture was not yet well-defined at the time of this study, the DIRSIG sensor model incorporated required LandIS sensor performance values for spectral response, edge response, and system noise as stated in the Landsat Next instrument requirements document (see Table 1) to accurately model the type of data expected from the LandIS sensor. The DIRSIG sensor model used a generic sensor model (2D framing array) as the baseline to image an approximately 60 x 60 km field of view for each of these scenes. Gaussian functions were used for relative spectral responses, where band placement (i.e., band centers and FWHMs) was defined based on the LandIS requirements. Similarly, a 2D Gaussian point spread function (PSF), based on the LandIS edge response requirements for each band, was applied to the image data to provide a realistic spatial response for the simulated data. System noise was also added to the image data based on the SNR requirements for each LandIS band. Finally, atmospheric effects were included using the MODTRAN atmosphere radiative transfer model. For Path/Row 016/030, a simple nominal summer atmospheric scenario was simulated; a mid-latitude profile with rural aerosols (default 23 km visibility) at 11:00 local time on the summer solstice. To better understand the impact of atmosphere and solar angle on compression ratios, several simulations were conducted for Path/Row 178/079. A mid-latitude profile and rural aerosols were used but the visibility parameter varied between 10, 20, and 30 km. Additionally, 11:00 local time was used for solar angle but the day-of-year (DOY) parameter was varied between DOY 1, 90, and 180. In total, nine variations of aerosol loading and DOY were simulated.

Although the DIRSIG model has been validated previously, such as the participation in the radiation transfer model inter-comparison (RAMI) campaigns [29, 30, 31], a comparison was performed between the DIRSIG renderings and real Landsat 9 data to provide confidence that DIRSIG is producing valid Landsat-like pixel

value ranges and spatial variability. These results along with detailed information on the simulated scene construction for the Landsat Next program will be published in a companion publication.

4. Compression Performance

The compression algorithm, using the hybrid encoder and relative error mode (with $m = 3$ error at L_{ref}), was applied to the three proxy datasets and performance statistics were recorded. As noted, the actual compression performance (compression ratio and compression error) will vary depending on the input image but a statistical analysis on a large number of images allows for overall trends to be studied. Additionally, since each of the three proxy datasets has its own strengths and weaknesses in regard to approximating LandIS imagery, observing the compression performance over the three very different datasets minimizes bias that may affect one proxy dataset over the others. In total, 994 Hyperion images, 1790 Landsat 9 images, and 10 DIRSIG images were used in this analysis. An implementation of the CCSDS 123.0-B-2 standard by the Centre National D'études Spatiales (CNES) was obtained and tested on the proxy datasets. The CNES implementation, freely available to non-profit organizations, provides binaries for compression and decompression and allows users full control over all parameters in the CCSDS 123.0-B-2 standard. Although the standard was originally designed for 3D compression of hyperspectral imagery, the compression was performed on a per-band (2D) basis since the LandIS spectral bands were tailored to be mostly uncorrelated to each other for typical Earth land cover types.

The compression results are summarized in Figure 5, which illustrates the compression ratio for each spectral band after compression in relative error mode with $m = 3$ error. The compression ratio (CR) was calculated by,

$$CR = \frac{N_X \cdot N_Y \cdot N_Z \cdot D}{\text{compressed data size [in bits]}} \qquad [unitless] \tag{4}$$

where N_X and N_Y represent the image dimensions in the across-track and along-track directions, respectively. N_Z is the spectral dimension ($N_Z = 1$, in this case),

and D is the dynamic range in bits. The resulting ratios are represented by box plots for each dataset independently where the center line of the box plot represents the median (50th percentile) value of the data and the lower and upper edge of the box are the first (25th percentile) and third (75th percentile) quartiles of the data, respectively. The top and bottom whiskers are the maximum and minimum values, respectively, when outliers are removed. Outliers are defined by data points beyond 1.5 times the inter-quartile range (all removed outliers were very high compression ratios). As mentioned in Section 3, DIRSIG was the only dataset with all 26 LandIS bands while the L9 and Hyperion datasets contain a subset of the LandIS bands.

Several key observations can be noted in these results. The spread of compression ratios were roughly consistent among the three datasets, suggesting that although each dataset was different, the compression performance was roughly similar. Secondly, the box plots (i.e., histograms) were skewed toward higher compression ratios. This is advantageous to the Landsat Next mission as higher compression ratios will add margin to the data downlink architecture. Lastly, there does not appear to be a strong correlation between compression ratio and GSD. Initially, there were concerns that higher spatial resolution might lead to lower compression ratios, as the algorithm would encounter greater pixel-to-pixel differences at higher resolutions, potentially reducing the efficiency of the predictor step in the compression process. This aspect will continue to be tracked as more datasets are incorporated and analyzed.

To further illustrate the variable nature of compression, Figure 6 illustrates the compression ratios along a single WRS-2 path with multiple revisits over one year for the L9 panchromatic band. Each data point represents one WRS-2 image and the black curve represents the average value over all instances for the particular WRS-2 row (i.e., latitude). Although the same physical location on Earth is revisited, the compression performance varies due to the changing scene content and atmospheric conditions. High altitude, rugged terrain contain high spatial frequency content that results in lower compression ratios. Other locations are more variable due to seasonally changing ground cover (snow, vegetation health, etc.) and atmospheric conditions (presence of clouds and aerosols). This dataset demonstrates

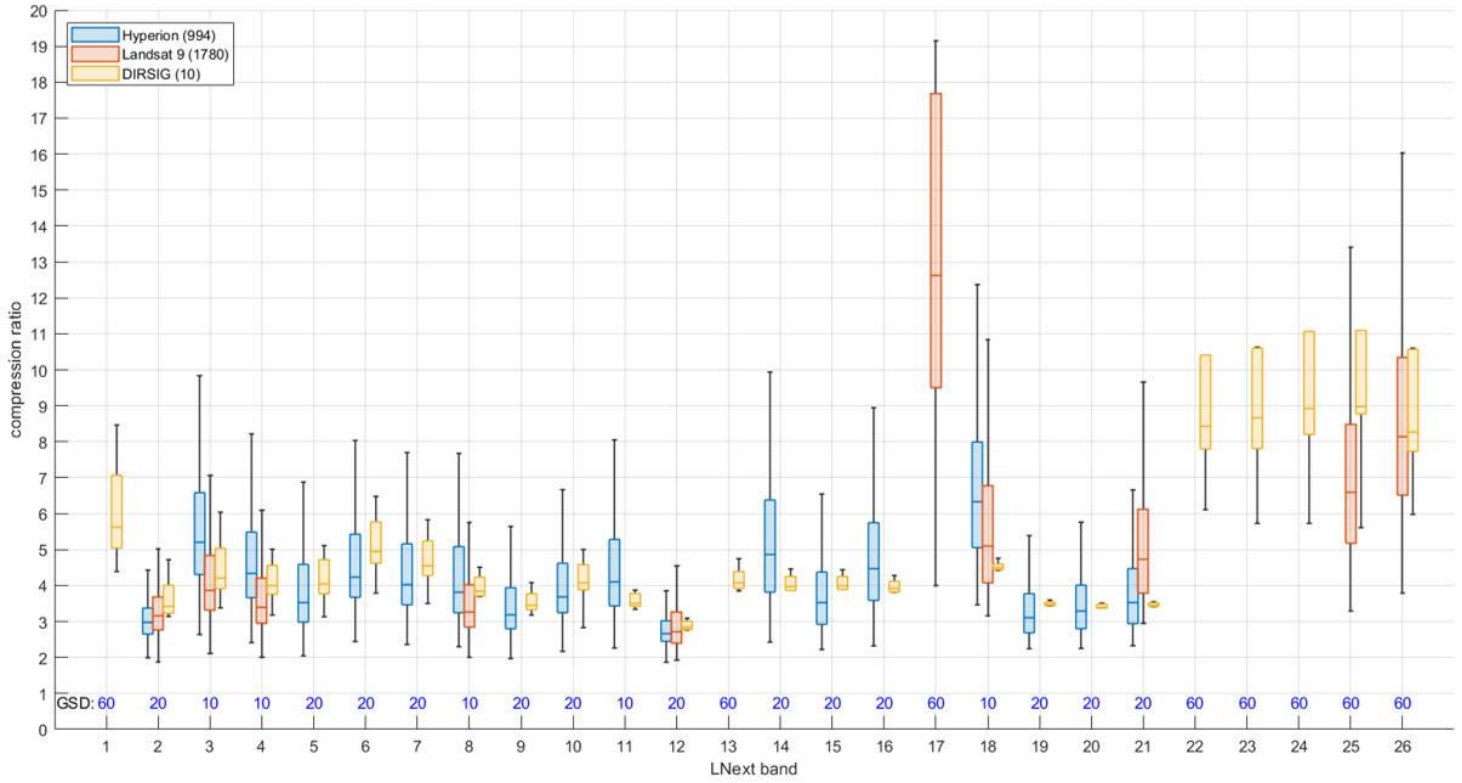

Figure 5: Compression ratio statistics for the three different LandIS proxy datasets separated by spectral band. The number of images for each dataset is noted in the legend. The compression algorithm was executed in relative error mode, with $m = 3$ error at L_{ref}. Each box plot represents the 25th, 50th (median), and 75th percentiles along with the minimum and maximum values when outliers are removed. The GSD of each band is shown for reference. Note that band 17, the cirrus band, experiences a wide range of high compression ratios since a cloud-free scene would contain low, uniform pixel values for this band and hence compress efficiently.

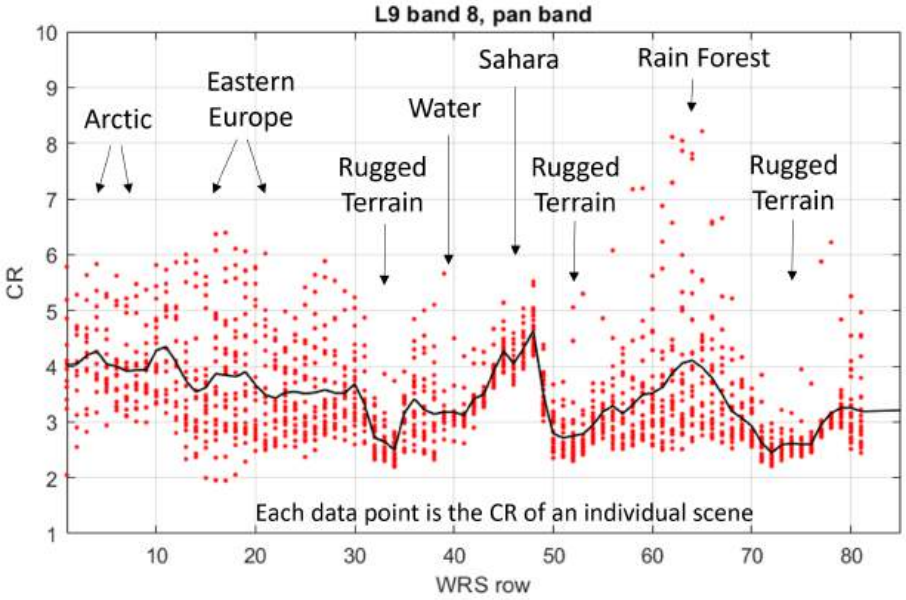

Figure 6: Compression ratio along a single WRS-2 Path (178) with multiple revisits over a year for the L9 panchromatic band. Each dot represents the CR for one image. The black curve is the average of all images for the particular WRS-2 Row.

the difficult nature of predicting exact compression performance and why statistical analyses must be relied on to model actual LandIS performance.

Although the results of Figures 5 and 6 represent a limited sampling of the Earth's surface with imperfect proxy datasets, they provide an estimate of the compression performance that is reasonably achievable with the CCSDS 123.0-B-2 standard on Landsat-like data. These datasets, however, were not meant to provide a global average compression, and the statistics only reflect the variability of test scenes evaluated. To expand on these datasets, the Landsat 9 archive (which contains a global set of image data repeated every 16 days) was leveraged to study how compression performance varies across the globe and throughout the seasons. Figure 7 shows the global compression ratios for the green band (560 nm) from L9 across four northern seasons in 2024: (a) winter (DOY 044-059), (b) spring (DOY 092-107), (c) early summer (DOY 153-169), and (d) late summer (DOY 214-229). The median global CRs for these seasons are 3.2, 3.5, 3.3, and 3.4, respectively. Note

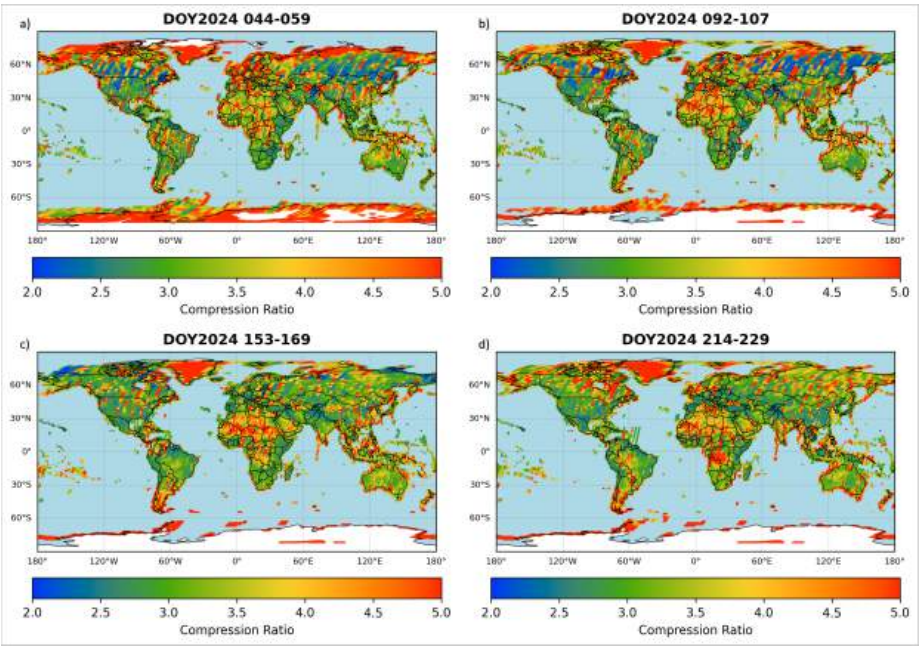

Figure 7: Global compression ratios for the green band (560 nm) from Landsat 9 across four northern seasons in 2024: (a) winter (DOY 044-059), (b) spring (DOY 092-107), (c) early summer (DOY 153-169), and (d) late summer (DOY 214-229). The color scales are clipped so that any CR at or above 5.0 is red.

that the colorbar is capped at a CR of 5, so values displayed as 5 represent ratios of 5 or greater. Overall, compression ratios are higher at the poles as expected, where the terrain is predominantly covered in snow and uniform, and over the Sahara, which is primarily desert and similarly uniform. The CRs show more variability in geographic regions with seasonal changes in vegetation and atmospheric conditions, as also illustrated in Figure 6. Globally, average compression ratios of 3-to-1 appear achievable with the specified algorithm parameters. Furthermore, there do not seem to be geographical regions with consistently low CRs that could have skewed the compression statistics observed in the narrower datasets shown in Figure 5.

5. Conclusion & Summary

Given the increased number of spectral bands and higher spatial resolution over legacy Landsat observatories, the Landsat Next mission must utilize near-lossless

image compression to ensure all acquired science image data can be successfully downlinked to ground stations. Although other popular compression algorithms are available, the CCSDS 123.0-B-2 compression standard was chosen for Landsat Next as it is open source, easily compatible with flight hardware, and allows control over the magnitude and distribution of added compression error in the image. The algorithm was applied to several proxy LandIS datasets to investigate the compression ratio and added error as a function of algorithm parameter values. The preliminary consensus is to operate the algorithm in relative error mode, which distributes the compression error based on pixel value, ensuring that the added percent error remains approximately constant across the dynamic range. The maximum compression error was limited to not exceed the system noise, NEdL (i.e., 3 digital counts at L_{ref}). Since the instrument performance requirements were designed to meet the vast majority of user needs (based on the user study performed by Wu *et al.* (2019) [7]), the added compression error determined here should be well within acceptable limits for users. By limiting compression-induced error to below the intrinsic system noise level, it is expected that the impact on Level 2 and Level 3 science products to be negligible.

Furthermore, the results of this study using proxy datasets demonstrated that median compression ratios of approximately 3-to-1 or higher are achievable, with the added compression error remaining within expected limits for nearly all pixels in the image. Coincidentally, this compression ratio is approximately the performance that is needed to fit the science data into the Ka-band downlink rate. This feedback directly influences overall mission design in terms of expected onboard storage requirements and number of ground stations.

The preliminary studies with the CCSDS 123.0-B-2 compression standard on proxy datasets have provided confidence that Landsat Next can meet its mission objectives of imaging all land area and downlinking the data to ground stations, without compromising image quality beyond science requirements. As the actual performance characteristics of the LandIS instruments become available during pre-flight characterization testing (e.g., system noise, edge response, etc.), the parameter values can be adjusted to maximize algorithm performance. The next stage

of efforts will focus on collaborating with the LandIS instrument vendor as the on-board FPGA architecture design is matured to incorporate the CCSDS 123.0-B-2 algorithm. Test datasets run through the hardware implementation of the compression standard will be compared to results presented here to ensure similar performance results between the software and hardware versions. Equally as important, the algorithm parameters will be required to be commandable (changeable) in flight so that compression performance tuning can be finalized during pre-flight testing and on-orbit commissioning using actual Earth image data from the LandIS instrument.

Funding

This research was funded in part by NASA Landsat Next contract award numbers 80NSSC19K1441, 80NSSC23K1014 and 80NSSC23K1015.

Acknowledgments

The authors acknowledge personnel at USGS EROS for their initiation of the compression studies for the Landsat Next mission that led to the selection of the CCSDS 123.B.2 algorithm, including Saeed Arab and Esad Micijevic.

References

[1] M. A. Wulder, J. C. White, T. R. Loveland, C. E. Woodcock, A. S. Belward, W. B. Cohen, E. A. Fosnight, J. Shaw, J. G. Masek, D. P. Roy, The global landsat archive: Status, consolidation, and direction, Remote Sensing of Environment 185 (2016) 271–283.

[2] M. A. Wulder, T. R. Loveland, D. P. Roy, C. J. Crawford, J. G. Masek, C. E. Woodcock, R. G. Allen, M. C. Anderson, A. S. Belward, W. B. Cohen, et al., Current status of landsat program, science, and applications, Remote sensing of environment 225 (2019) 127–147.

[3] M. A. Wulder, D. P. Roy, V. C. Radeloff, T. R. Loveland, M. C. Anderson, D. M. Johnson, S. Healey, Z. Zhu, T. A. Scambos, N. Pahlevan, et al., Fifty years of landsat science and impacts, Remote Sensing of Environment 280 (2022) 113195.

[4] C. J. Crawford, D. P. Roy, S. Arab, C. Barnes, E. Vermote, G. Hulley, A. Gerace, M. Choate, C. Engebretson, E. Micijevic, et al., The 50-year landsat collection 2 archive, Science of Remote Sensing 8 (2023) 100103.

[5] M. Claverie, J. Ju, J. G. Masek, J. L. Dungan, E. F. Vermote, J.-C. Roger, S. V. Skakun, C. Justice, The harmonized landsat and sentinel-2 surface reflectance data set, Remote sensing of environment 219 (2018) 145–161.

[6] S. Saunier, B. Pflug, I. M. Lobos, B. Franch, J. Louis, R. De Los Reyes, V. Debaecker, E. G. Cadau, V. Boccia, F. Gascon, et al., Sen2like: Paving the way towards harmonization and fusion of optical data, Remote Sensing 14 (16) (2022) 3855.

[7] Z. Wu, G. Snyder, C. Vadnais, R. Arora, M. Babcock, G. Stensaas, P. Doucette, T. Newman, User needs for future landsat missions, Remote Sensing of Environment 231 (2019) 111214.

[8] J. G. Masek, M. A. Wulder, B. Markham, J. McCorkel, C. J. Crawford, J. Storey, D. T. Jenstrom, Landsat 9: Empowering open science and applications through continuity, Remote Sensing of Environment 248 (2020) 111968.

[9] J. R. Irons, J. L. Dwyer, An overview of the landsat data continuity mission, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI 7695 (2010) 58–64.

[10] B. Book, Lossless data compression, recommended standard ccsds 121.0-b-1, Tech. rep., CCSDS (1997).

[11] A. Meygret, S. Baillarin, F. Gascon, E. Hillairet, C. Dechoz, S. Lacherade, P. Martimort, F. Spoto, P. Henry, R. Duca, SENTINEL-2 Image Quality and

Level 1 Processing, in: Earth Observing Systems XIV, Vol. 7452, SPIE, 2009. doi:10.1117/12.826184.

[12] A. Skodras, C. Christopoulos, T. Ebrahimi, The jpeg 2000 still image compression standard, IEEE Signal processing magazine 18 (5) (2001) 36–58.

[13] M. J. Weinberger, G. Seroussi, G. Sapiro, The loco-i lossless image compression algorithm: Principles and standardization into jpeg-ls, IEEE Transactions on Image processing 9 (8) (2000) 1309–1324.

[14] A. B. Kiely, M. Klimesh, I. Blanes, J. Ligo, E. Magli, N. Aranki, M. Burl, R. Camarero, M. Cheng, S. Dolinar, et al., The new ccsds standard for low-complexity lossless and near-lossless multispectral and hyperspectral image compression, in: International Workshop on On-Board Payload Data Compression, no. GSFC-E-DAA-TN61202, 2018.

[15] M. Hernández-Cabronero, A. B. Kiely, M. Klimesh, I. Blanes, J. Ligo, E. Magli, J. Serra-Sagrista, The ccsds 123.0-b-2 "low-complexity lossless and near-lossless multispectral and hyperspectral image compression" standard: A comprehensive review, IEEE Geoscience and Remote Sensing Magazine 9 (4) (2021) 102–119.

[16] R. O. Green, N. Mahowald, C. Ung, D. R. Thompson, L. Bator, M. Bennet, M. Bernas, N. Blackway, C. Bradley, J. Cha, et al., The earth surface mineral dust source investigation: An earth science imaging spectroscopy mission, in: 2020 IEEE aerospace conference, IEEE, 2020, pp. 1–15.

[17] D. K. T. Pham, M. Klimesh, G. Allen, G. Flesch, R. Valencia, H. Xie, A. Kiely, D. Dolman, K. Roth, K. Crocker, et al., High-performance embedded system-on-a-chip for space imaging spectrometer, in: 2023 IEEE Aerospace conference, IEEE, 2023, pp. 1–10.

[18] D. Vorhaug, S. Boyle, S. Eine, M. Orlandic, Development and integration of ccsds 123.0-b-2 fpga accelerator for hypso-1, in: Proceedings of the 9th International Workshop on On-Board Payload Data Compression (OBPDC), 2024.

[19] B. Book, Low-complexity lossless and near-lossless multispectral and hyperspectral image compression, ccsds 123.0-b-2, Tech. rep., CCSDS (2019).

[20] I. Blanes, A. Kiely, M. Hernández-Cabronero, J. Serra-Sagristà, Performance impact of parameter tuning on the ccsds-123.0-b-2 low-complexity lossless and near-lossless multispectral and hyperspectral image compression standard, Remote Sensing 11 (11) (2019) 1390.

[21] Y. Barrios, A. Sánchez, R. Guerra, R. Sarmiento, Hardware implementation of the ccsds 123.0-b-2 near-lossless compression standard following an hls design methodology, Remote Sensing 13 (21) (2021) 4388.

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE transactions on image processing 13 (4) (2004) 600–612.

[23] S. G. Ungar, J. S. Pearlman, J. A. Mendenhall, D. Reuter, Overview of the earth observing one (eo-1) mission, IEEE Transactions on Geoscience and Remote Sensing 41 (6) (2003) 1149–1159.

[24] R. O. Green, B. E. Pavri, T. G. Chrien, On-orbit radiometric and spectral calibration characteristics of eo-1 hyperion derived with an underflight of aviris and in situ measurements at salar de arizaro, argentina, IEEE Transactions on Geoscience and Remote Sensing 41 (6) (2003) 1194–1203.

[25] J. Schott, A. Gerace, S. Brown, M. Gartley, M. Montanaro, D. C. Reuter, Simulation of image performance characteristics of the landsat data continuity mission (ldcm) thermal infrared sensor (tirs), Remote Sensing 4 (8) (2012) 2477–2491. doi:10.3390/rs4082477.
URL https://www.mdpi.com/2072-4292/4/8/2477

[26] A. Gerace, J. Schott, M. Gartley, M. Montanaro, An analysis of the side slither on-orbit calibration technique using the dirsig model, Remote Sensing 6 (11) (2014) 10523–10545. doi:10.3390/rs61110523.
URL https://www.mdpi.com/2072-4292/6/11/10523

[27] A. Gerace, M. Gartley, J. Schott, N. Raqueño, R. Raqueño, Data-driven simulations of the Landsat Data Continuity Mission (LDCM) platform, in: S. S. Shen, P. E. Lewis (Eds.), Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVII, Vol. 8048, International Society for Optics and Photonics, SPIE, 2011, p. 804815. doi:10.1117/12.885561.
URL https://doi.org/10.1117/12.885561

[28] A. Berk, P. Conforti, R. Kennett, T. Perkins, F. Hawes, J. van den Bosch, MODTRAN6: a major upgrade of the MODTRAN radiative transfer code, in: M. Velez-Reyes, F. A. Kruse (Eds.), Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX, Vol. 9088, International Society for Optics and Photonics, SPIE, 2014, pp. 113 – 119. doi:10.1117/12.2050433.
URL https://doi.org/10.1117/12.2050433

[29] J.-L. Widlowski, B. Pinty, M. Lopatka, C. Atzberger, D. Buzica, M. Chelle, M. Disney, J.-P. Gastellu-Etchegorry, M. Gerboles, N. Gobron, et al., The fourth radiation transfer model intercomparison (rami-iv): Proficiency testing of canopy reflectance models with iso-13528, Journal of Geophysical Research: Atmospheres 118 (13) (2013) 6869–6890.

[30] A. A. Goodenough, S. D. Brown, Development of land surface reflectance models based on multiscale simulation, in: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXI, Vol. 9472, SPIE, 2015, pp. 132–139.

[31] B. D. Roth, A. A. Goodenough, S. D. Brown, J. A. van Aardt, M. G. Saunders, K. Krause, Simulations of leaf bsdf effects on lidar waveforms, Remote Sensing 12 (18) (2020) 2909.