Universitat Autònoma de Barcelona

Pyramidal regression-based coding for remote sensing data

Sara Álvarez Cortés

Universitat Autònoma de Barcelona

Departament d'Enginyeria de la Informació i de les Comunicacions

Pyramidal regression-based coding for remote sensing data

Submitted to Universitat Autònoma de Barcelona in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science

by Sara Álvarez Cortés

Bellaterra, May 2019

Supervisor:

Dr. Joan Serra Sagristà

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Bellaterra, May 2019

Dr. Joan Serra Sagristà

Committee:

Dr. Enrico Magli
Dr. Ian Blanes
Dr. Anton Bardera
(substitute) Dr. Diego Valsesia
(substitute) Dr. Francesc Aulí Llinàs
(substitute) Dr. Jordi Portell i de Mora

Abstract

Remote sensing hyperspectral data have hundreds or thousands of spectral components from very similar wavelengths. To store and transmit it entails excessive demands on bandwidth and on on-board memory resources, which are already strongly restricted. This leads to stop capturing data or to discard some of the already recorded information without further processing. To alleviate these limitations, data compression techniques are applied. Besides, sensors' technology is continuously evolving, acquiring higher dimensional data. Consequently, in order to not jeopardize future space mission's performance, more competitive compression methods are required.

Regression Wavelet Analysis (RWA) is the state-of-the-art lossless compression method regarding the trade-off between computational complexity and coding performance. RWA is introduced as a lossless spectral transform followed by JPEG 2000. It applies a Haar Discrete Wavelet Transform (DWT) decomposition and sequentially a regression operation. Several regression models (Maximum, Restricted and Parsimonious) and variants (only for the Maximum model) have been proposed.

With the motivation of outperforming the latest compression techniques for remote sensing data, we began focusing on improving the coding performance and/or the computational complexity of RWA. First, we conducted an exhaustive research of the influence of replacing the underlying wavelet filter of RWA by more competitive Integer Wavelet Transforms (in terms of energy compaction). To this end, we reformulated the Restricted model, reducing the execution time, increasing the compression ratio, and preserving some degree of component-scalability. Besides, we showed that the regression variants are also feasible to apply to other models, decreasing their computational complexity while scarcely penalizing the coding performance. As compared to other lowest- and highest-complex techniques, our new configurations provide, respectively, better or similar compression ratios.

After gaining a comprehensive understanding of the behavior of each operation block, we described the impact of applying a Predictive Weighting Scheme (PWS) in the Progressive Lossy-to-Lossless (PLL) compression performance. PLL decoding is possible thanks to the use of the rate control system of JPEG 2000. Applying this PWS to all the regression models and variants of RWA coupled by JPEG 2000 (PWS-RWA + JPEG 2000) produces superior outcomes, even for multi-class digital classification. From experimentation, we concluded that improved coding performance does not necessarily entail better classification outcomes. Indeed, in comparison with other widespread

techniques that obtain better rate-distortion results, PWS-RWA + JPEG 2000 yields better classification outcomes when the distortion in the recovered scene is high. Moreover, the weighted framework presents far more stable classification versus bitrate trade-off.

JPEG 2000 may be too computationally expensive for on-board computation. In order to obtain a cheaper implementation, we render results for RWA followed by another coder amenable for on-board operation. This framework includes the operation of a smart and simple criterion aiming at the lowest bitrates. This final pipeline outperforms the original RWA + JPEG 2000 and other state-of-the-art lossless techniques by obtaining average coding gains between 0.10 to 1.35 bits-per-pixel-per-component.

Finally, we present the first lossless/near-lossless compression technique based on regression in a pyramidal multiresolution scheme. It expands RWA by introducing quantization and a feedback loop to control independently the quantization error in each decomposition level, while preserving the computational complexity. To this end, we provide a mathematical formulation that limits the maximum admissible absolute error in reconstruction. Moreover, we tackle the inconvenience of proving the huge number of possible quantization steps combinations by establishing a quantization steps-allocation definition. Our approach, named NLRWA, attains competitive coding performance and superior scene's quality retrieval. In addition, when coupled with a bitplane entropy encoder, NLRWA supports progressive lossy-to-lossless/near-lossless compression and some degree of embeddedness.

Resumen

Los datos hiperespectrales capturados por teledetección cuentan con cientos o miles de componentes espectrales de similares longitudes de onda. Almacenarlos y transmitirlos conlleva una demanda excesiva en ancho de banda y memoria, ya de por sí bastante limitados, que pueden dar lugar a descartar información ya capturada o a dejar de capturarla. Para paliar estas limitaciones, se aplican algoritmos de compresión. Además, la tecnología de los sensores evoluciona continuamente, pudiéndose adquirir datos con mayores dimensiones. De ahí que, para no penalizar el funcionamiento y rendimiento de futuras misiones espaciales, se necesitan desarrollar métodos de compresión más competitivos.

Regression Wavelet Analysis (RWA) es el método de compresión sin pérdidas más eficiente en relación a la complejidad computacional y al rendimiento de codificación. RWA se describe como una transformada espectral sin pérdida seguida de JPEG 2000. Ésta aplica un nivel de descomposición de la transformada discreta de onda Haar y una regresión. Hay varios modelos de regresión (Maximum, Restricted y Parsimonious) y variantes (solo para Maximum).

Inicialmente, nos centramos en aumentar el rendimiento de codificación y/o reducir la complejidad computacional de RWA para diseñar técnicas de compresión más competitivas. Primero, investigamos en profundidad la influencia que tiene el reemplazar el filtro de RWA por transformadas más eficientes en cuanto a la compactación de energía. Para ello, redefinimos el modelo Restricted, reduciendo el tiempo de ejecución, incrementando el ratio de compresión, y preservando un cierto grado de escalabilidad por componente. Además, mostramos que las variantes de regresión se pueden aplicar a todos los modelos de regresión, disminuyendo así su complejidad computacional sin apenas penalizar el rendimiento de codificación. Nuestras nuevas configuraciones proporcionan ratios de compresión mayores o bastante competitivos con respecto a otras técnicas de menor y mayor complejidad.

Tras ello, describimos el impacto que tiene el aplicar un esquema de pesos predictivo (PWS) en el rendimiento de compresión cuando se decodifica de forma progresiva desde con-pérdida hasta sin-pérdida (PLL). La aplicación de estos pesos a todos los modelos de regresión y variantes de RWA con JPEG 2000 (PWS-RWA + JPEG 2000) mejora los resultados del esquema original (RWA + JPEG 2000). Por otro lado, vemos que un mejor rendimiento de la codificación no implica necesariamente mejores clasificaciones. De hecho, en comparación con otras técnicas que recuperan la escena con mayor calidad, PWS-RWA + JPEG 2000 provee de mejores clasificaciones cuando la distorsión en

la recuperación es elevada.

Para obtener una implementación de más baja complejidad computacional, presentamos resultados de RWA acompañada de un codificador que se puede ejecutar a bordo. Además, con un sencillo criterio de decisión conseguimos tasas de bits más bajas, mejorando al esquema original y otras técnicas de compresión sin pérdidas al obtener ganancias de codificación promedio entre 0,10 y 1,35 bits-por-píxel-por-componente.

Finalmente, presentamos la primera técnica de compresión sin-pérdida/casi-sin-pérdida basada en un sistema piramidal que aplica regresión. Para ello, ampliamos RWA introduciendo cuantización y un algoritmo de retroalimentación para controlar independientemente el error de cuantificación en cada nivel de descomposición, al mismo tiempo que preservamos la complejidad computacional. Proporcionamos también una ecuación que limita el máximo error en valor absoluto admisible en la reconstrucción. A su vez, evitamos probar la gran cantidad de combinaciones posibles de pasos de cuantificación mediante el desarrollo de un esquema de asignación de pasos. Nuestra propuesta, llamada NLRWA, logra obtener un rendimiento de codificación muy competitivo y recuperar la escena con mayor fidelidad. Por último, cuando el codificador por entropía se basa en planos de bits, NLRWA puede proporcionar una compresión progresiva desde con-pérdida hasta sin-pérdida/casi-sin-pérdida y cierto grado de integrabilidad.

Acknowledgements

First of all, I would like to express my most sincere gratitude to Joan Serra for his patient guidance, continuous encouragement and wise advice. They have made possible the elaboration of this Ph.D. dissertation. Particularly, a well-deserved thanks for his time and commitment.

I am very thankful to have had the opportunity to spend three months under the supervision of Prof. Michael Marcellin. He always gave me valuable and fruitful insights and inputs. It was a pleasure working with him.

To my parents and sister, to whom I have so much to be grateful for. I feel words are not enough to describe their unconditional love and support. If all the people were like them, this world would be a much more wonderful place.

To my friends, whom I chose to be my second family. I am so proud of you all. Special thanks to Nuria, Marina, Clara, Jorge, Bea, Adrián, Carmen, Borja, Alan, Neri, David, José A., Emejota, Loreto, Ana, and Vanessa. You are totally **amazing**.

Lastly, but not least, many thanks go to my fellow doctoral colleagues for all the memorable moments together. Specially to Iván, Carlos V., Roland and Carlos D.C. for their support and for being so *notorious*.

¡Gracias a todos!

Contents

Chapter 1

Introduction

1.1 Remote sensing data

Remote sensing refers to the non-physical contact acquisition of information, specially from a large distance. Nowadays, satellite's imaging sensors collect hyperspectral scenes (images) with high 2D spatial resolutions. These enormous volumes have hundreds or thousands of very correlated spectral components (bands), providing a wealth of spectral information very convenient for Earth Observation's applications, such as appraisals of climate changes, for farming or military purposes, or to give support when natural disasters strike.

1.2 Bandwidth and on-board storage capacity

Spaceborne and sounding interferometer sensors' technology evolves continuously to improve nowadays applications' performance and meet users' needs. This places excessive demands on bandwidth and on on-board storage capacity, e.g.,

1. Met-Op satellite can only transmit data to ground stations on Earth at a rate of 1.5 Mb/s, while the recording rate is of at most 45 Mb/s [1];
2. IASI sensor on the Met-Op satellite acquires close to 20 GB daily [2], while the

on-board storage capacity on this satellite is 24 Gbit.

Although downlink transmission rates and on-board memory resources have increased during the years, they are still strongly constrained. This leads to stop capturing data, or discarding already recorded data without further processing. Data compression becomes therefore imperative to alleviate these limitations and fulfill the space missions' requirements. Note that compression is also beneficial for on-the-ground storage and dissemination.

1.3 Remote sensing data compression techniques

Due to the large spectral redundancy in hyperspectral scenes, remote sensing data compression methods that decorrelate the spectral dimension have proven to yield the most efficient coding performances [3]. These techniques aim at lossless, near-lossless or lossy recovery according to the loss of fidelity permitted in reconstruction.

- Lossless compression ensures perfect recovery at the price of low compression ratios, close to 2:1.
- Near-lossless coding aims at higher compression ratios than lossless methods by allowing some loss of fidelity in reconstruction. They bound the l_∞-norm -equivalently, the peak absolute error (PAE) or maximum absolute distortion (MAD)- via setting an error tolerance value Λ, guaranteeing thus some image quality control. This user-specified parameter Λ sets the maximum admissible absolute error so that PAE $\leq \Lambda$.
- Lossy compression enables high compression ratios at the expense of allowing uncontrolled loss in decoding.

Several compression techniques for remote sensing data can be found in the literature. However, the technical evolution of the optical and sounding interferometer instruments, and the constantly limited downlink bandwidth unveil an insufficient on-board storage capacity, and the on-the-ground dissemination is also contested. Under such a scenario, more innovative and updated compression techniques are required.

This thesis stems from the aim of improving the coding performance of nowadays compression techniques for remote sensing data, without barely penalizing or even decreasing the computational complexity. With this motivation, first an exhaustive study of the state-of-the-art compression techniques is conducted.

1.3.1 State-of-the-art review

To compress an image a transform- or a prediction-based approach is normally pursued. The first one compacts the energy of a scene into a few components, while the second computes the estimation of a pixel's value from previous processed pixels, and entropy-encodes the prediction error afterwards. We first concentrated on the lossless compression techniques.

For lossless compression, these methods usually perform a Lifting Scheme [4] or some rounding/calculation to operate with integer values. Some lossless coding proposals apply a 3D spatio-spectral transform [5, 6, 7]. Others use first a 1D spectral transform to later decorrelate the 2D spatial domain [7, 8]. Discrete Wavelet Transforms (DWTs) are the most widespread transforms due to their minor computational cost, short execution time, and component-scalability, i.e., to retrieve a scene component, only some, but not all, of the remaining transformed components are needed. Despite all these strengths, they do not remove the statistical dependencies in the spectral domain as Karhunen-Loève Transform (KLT) does. KLT is optimal for decorrelating Gaussian sources [9, 10], and usually yields high compression ratios. However, it is data-dependent and requires every transformed component to retrieve any single scene coefficient. Moreover, and in contrast to common DWTs, KLT focuses on the 2^{nd} order moments, letting aside higher orders that may be meaningful, is computationally high demanding, and entails a non-negligible side information, in particular for scenes with a large number of spectral bands. Consequently, for hyperspectral scenes KLT results to be computationally unattainable and inefficiently-performing. Several approaches have come to light to overcome some of its drawbacks, which also try to minimize the coding performance decline [11]: some conduct a spectral sub-

sampling method [7], others extract the transformed coefficients from a single scene to apply them afterwards to the remaining scenes of the same corpus [12], and finally, another group of techniques employ a divide-and-conquer strategy, e.g., through multilevel clusterization [13] or through multiple applications of pairwise KLT [14]. The last technique accomplishes to reduce the KLT computational cost and large memory requirements, while procuring competitive compression ratios and a reasonable component-scalability.

Since higher degree of statistical independence usually means higher efficiency in coding performance, wavelet transforms may give rise to inefficient results. Nowadays, several techniques that obtain competitive coding outcomes at considerably lower computational complexity than KLT have been published. An examples is Multiband Context-based Adaptive Lossless Image Coding (M-CALIC) [15] which is one of the most renowned methods for lossless hyperspectral data compression. It incorporates a multiband spectral predictor, and optimized parameters and quantization thresholds into CALIC [16] and its 3D extension (3D-CALIC) [17], improving their performance. In its turn, M-CALIC utilizes CALIC as spatial compression engine. CALIC uses a nonlinear adaptive predictor, context-templates, which consider only encoded pixels from the two previous immediately preceded rows, quantization, and an entropy encoder conditioned on a large number of contexts, instead of on conditional error probabilities. 3D-CALIC includes a simple spectral predictor that considers one band as reference to estimate the current one. It is very efficient compressing multispectral data but not hyperspectral scenes. To enhance its coding performance, a more sophisticated spectral predictor is required, as the included in M-CALIC.

Another technique that entails minor complexity is the one proposed by the Consultive Committee for Space Data Systems (CCSDS) in 2012, the standard CCSDS-123.0-B-1 [18]. It can be computed on-board and is formed by an adaptive linear predictor, a mapper function, and an entropy encoder. It exploits the redundancy within the nearby 3-D spatio-spectral neighbor pixels. Years later, a modification

of CCSDS-123.0-B-1 was published [19], henceforth named as CCSDS-123-AC. It deploys the CCSDS-123.0-B-1's predictor and mapper, and includes a lightweight contextual arithmetic encoder. This coder defines a context model and computes the probabilities that will be later used by a fixed-length arithmetic encoder. CCSDS-123-AC improves the performance of CCSDS-123.0-B-1 and of M-CALIC.

Up to now, other prediction-based techniques, in particular based upon recursive least-squares, [2, 20, 21, 22, 23] have been proposed. Among them, references [20, 21, 22, 23] perform extensive and time-consuming algorithms to aim at competitive estimations. Results of these techniques have only been reported for scenes recorded by the Airbone Visible/Infrared Imaging Spectrometer (AVIRIS) instrument. For this corpus, they outperform the coding performance of CCSDS-123-AC at the expense of higher complexity. The remaining technique, Regression Wavelet Analysis (RWA) coupled with JPEG 2000 [2] (where JPEG is the acronym of the Joint Photographic Experts Group), yields compression ratios close to those of [20, 21, 22, 23] at lower computational resources consumption. RWA is straightforward and fast-computing, partly thanks to avoiding any adaptive-iterative scheme application. RWA also attains superior coding performance than CCSDS-123-AC, DWT-based, and also KLT-based techniques, becoming the state-of-the-art lossless prediction-based technique regarding the trade-off between computational cost and coding performance.

RWA benefits from its minor side information requirements and low complexity. It is a lossless transform that applies a DWT and solves an ordinary least-squares [24] problem per DWT decomposition level. The first operation is conducted by the integer Haar S-Transform, whereas the second predicts the wavelet details from the wavelet approximations of the same decomposition level. This procedure is iteratively computed until the highest possible wavelet level. After that, the prediction error is encoded and stored/transmitted, significantly reducing the statistical relationships the DWT can not remove. In all the publications related to RWA [2, 25, 26, 27], this second stage is carried out through three regression models: Maximum, Parsimoniuos and Restricted, and two variants of the Maximum model: Fast and Exogenous.

Though RWA has always been coupled with the compression standard JPEG 2000 [28], other coders could be used. JPEG 2000 is well-known for obtaining excellent lossless and lossy performance in terms of Mean Squared Error (MSE), but is not always appropriate for on-board computation. In spite of this, it can gradually refine the quality retrieval of the reconstructed scene until no loss in recovery exists. This type of coding is known as Progressive Lossy-to-Lossless (PLL) compression. Generally, lossy pipelines apply quantization, an entropy encoder on the quantized signal, and a rate control allocation stage afterwards. The PLL compression is possible by using the rate control allocation stage. In fact, [25] provides rate-distortion curves of RWA when followed by JPEG 2000. On the other hand, applying uniform scalar quantization is one of the simplest manners to introduce loss when encoding a signal, and therefore, to increase the compression ratio. This scheme maps each element value into a particular value within a delimited subset of disjoint intervals. For near-lossless compression, applying quantization in a feedback loop allows for controlling the error via a maximum error tolerance value Λ. There are three near-lossless coding modalities: prediction-based techniques followed by quantization and lossless compression, two-stage near-lossless coders, and prequantization before lossless coding.

In *short, remote sensing compression techniques decorrelate the spectral dimension of hyperspectral data. They are capable of yielding lossless, near-lossless and lossy recovery depending on the error allowed in reconstruction. Normally, they are transform- or prediction-based methods. The first group is greatly influenced by the trade-off between efficiency and complexity. The second usually yields better coding performance for lossless compression. Regression Wavelet Analysis is a state-of-the-art lossless transform that combines both modalities, i.e., it computes first a discrete wavelet transform, and then a prediction of the wavelet details. Several publications of RWA can be found in the literature. However, results are only reported for lossless or progressive lossy-to-lossless coding, when using Haar-DWT at the highest decomposition level, and when followed by JPEG 2000.*

1.4 Contributions of the thesis

At the beginning of the thesis, our main focus consisted in providing novel coding techniques that outperform the lossless state-of-the-art compression field. To this end, we extended RWA to improve its coding performance, reduce its computational cost, and prove its applicability in nowadays remote sensing applications. Later, we expanded our research beyond our initial aspiration, and succeeded in developing the first near-lossless compression technique based on regression, specially in a multiresolution scheme. Therefrom, this technique gives rise to a new near-lossless compression category: near-lossless compression based on regression in a pyramidal framework. The contributions can be found published in several journals on the remote sensing data compression area:

- Sara Álvarez-Cortés, Naoufal Amrani, and Joan Serra-Sagristà, **"Low complexity regression wavelet analysis variants for hyperspectral data lossless compression,"** International Journal of Remote Sensing, vol. 39, no. 7, pp. 1971–2000, Sep. 2017, DOI: 10.1080/01431161.2017.1375617 [29].

- Sara Álvarez-Cortés, Naoufal Amrani, Miguel Hernández-Cabronero, and Joan Serra-Sagristà, **"Progressive lossy-to-lossless coding of hyperspectral images through regression wavelet analysis,"** International Journal of Remote Sensing, vol. 39, no. 7, pp. 2001–2021, Jul. 2017, DOI: 10.1080/01431161.2017.1343515 [30].

- Sara Álvarez-Cortés, Joan Bartrina-Rapesta, and Joan Serra-Sagristà, **"Multilevel split regression wavelet analysis for lossless compression of remote sensing data,"** IEEE Geoscience and Remote Sensing Letters, no. 99, pp. 1–5, Jul. 2018, DOI: 10.1109/LGRS.2018.2850938 [31].

- Sara Álvarez-Cortés, Joan Serra-Sagristà, Joan Bartrina-Rapesta, and Michael

Marcellin, **"Regression wavelet analysis for near-lossless remote sensing data compression,"** SUBMITTED in Apr. 2019 to *IEEE Transactions on Geoscience and Remote Sensing*.

1.5 Organization of the thesis

THE organization of the thesis is briefly described next:

Chapter 2 discusses publication [29]. In this manuscript, we present an exhaustive evaluation when replacing Haar-DWT by usually more competitive, but a bit more complex transforms, the reversible IWT 5/3 and IWT 9/7M. To this effect, we change the cubic polynomial formulation that defines the Restricted model of RWA to a first-order equation, preserving some degree of component-scalability. In order to decrease even more the computational cost, while barely penalizing the coding performance, we propose new variants for the Restricted and Parsimonious regression models. Computational complexity is assessed in Floating-point Operations (FLOPs) and execution time. We also study the influence of varying the number of RWA decomposition levels in the coding performance. Moreover, we disclose dynamic range, precision, and entropy distributions in bits-per-spectral-component (bpppc) of a scene in the transform domain, and appraise the coding performance of our approaches in bpppc, Shannon entropy, Pearson's correlation, mutual information and energy terms. Finally, an extensive comparison against state-of-the-art techniques, based upon recursive least-squares, and other widespread methods, has been supplied too, highlighting the competitive performance of RWA.

Chapter 3 presents our paper [30] in which an in-depth assessment of the Progressive Lossy-to-Lossless RWA (PLL-RWA) coding process is conducted. We cover the beneficial use of applying a Predictive Weighting Scheme (PWS) [25] not only for Maximum and Exogenous-Maximum as in [25], but also for the Restricted model and the different variants presented in [29]. For comparison purposes, we provide results of M-CALIC and competitive lossless spectral transforms followed by JPEG 2000.

We also evaluate the effect of varying the wavelet transform (Haar-DWT, IWT 5/3 and IWT 9/7M) of the first operation block in the PLL-RWA scheme. The benefits of the PWS are explained investigating the bitrate distributions per transformed spectral component according to a global target bitrate. Finally, PWS-RWA + JPEG 2000 [25] is assessed for multi-class digital classification, suggesting that this framework yields superior performance.

Chapter 4 describes publication [31]. RWA is coupled for the first time with a coder other than JPEG 2000. Here, we describe a novel technique. It consists of applying RWA followed by the on-board affordable coder CCSDS-123-AC. As a side remark, RWA + JPEG 2000 achieves its highest coding performance when applying the highest number of RWA decomposition levels [29]. In contrast, computing the highest level in our approach, RWA + CCSDS-123-AC, does not necessarily give rise to the best coding performance, helping to decrease the computational complexity. To further improve the coding gain, we introduce a strategy, named Multi-Level Split RWA (MLS-RWA), to select the optimal number of RWA levels to apply. Our proposal MLS-RWA + CCSDS-123-AC obtains coding gains even higher than 1 bpppc with respect to nowadays highly-performing techniques.

Chapter 5 includes our last manuscript, submitted to IEEE Transactions on Geoscience and Remote Sensing (TGRS) journal in April 2019. It introduces the first near-lossless compression technique based on regression, specially in the pyramidal multiresolution RWA. Besides, our novel technique, named near-lossless RWA (NLRWA), provides exactly the same lossless outcomes as the original RWA. We also propose a smart criterion that is independent of the image to process. It selects a unique quantization steps combination, avoiding thus to solve a slow multivariate optimization problem that iteratively tries every possible combination. Apart from that, NLRWA can be coupled with any entropy coder. Here, NLRWA is presented followed by JPEG 2000. NLRWA + JPEG 2000 can afford progressive lossy-to-lossless/near-lossless compression, and produces competitive compression ratios and superior reconstructed scene's quality.

Chapter 6 synthesizes the results presented in previous chapters.

Chapter 7 brings forward the conclusions of the contributions of this thesis. Finally, some insights of the future work are also introduced.

Chapter 2

Low complexity regression wavelet analysis variants for hyperspectral data lossless compression

Low Complexity Regression Wavelet Analysis Variants for Hyperspectral Data Lossless Compression

Sara Álvarez-Cortés, Naoufal Amrani, and Joan Serra-Sagristà

Universitat Autònoma de Barcelona

***Abstract*—The evolution of the optical and of the sounding interferometer instruments along with the increase of the spaceborne storage capacity allow for the acquisition of large data volumes. However, the strongly limited downlink bandwidth unveils an insufficient on-board storage capacity, and the on-the-ground storage and dissemination are also contested. In these scenarios, data compression techniques are demanded. We discuss here the Regression Wavelet Analysis (RWA) spectral transform, introducing novel variants that lead to an improved lossless coding performance. A comprehensive comparison with state-of-the-art remote sensing data compression techniques shows the competitive behavior of RWA in terms of lossless coding performance (yielding lower bit-rates), computational complexity (requesting lower execution time) and other signal measurements (decreasing energy, mutual information and entropy). Experimental results are performed on uncalibrated and calibrated data from Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), from Hyperion instrument and from Infrared Atmospheric Sounding Interferometer (IASI).**

***Index Terms*—Lossless compression, Hyperspectral data compression, Regression Wavelet Analysis, Regression models, Discrete Wavelet Transforms.**

I. Introduction

Remote-sensing refers to the non-physical contact acquisition of information. Thanks to the enhancement of the spaceborne optical and sounding interferometer instruments and to the increase of the on-board storage capacity, data with high spectral sampling and high spatial resolution can be recorded by on-board satellites. As an example, the Infrared Atmospheric Sounding Interferometer (IASI) on the Met-Op satellite captures orbits with 8,359 spectral components, generating close to 20 GB daily. These hyperspectral data provide a wealth of spectral information, leading to a better understanding and a more precise characterisation and awareness of short and long condition changes of the environment, atmosphere and global surface. However, the transmission of this large amount of data poses a challenging problem, as the satellite download channel capacity is very restricted; for instance, the Met-Op satellite records data at a rate of 45 Mb s^{-1}, but can only transmit to Earth ground stations at 1.5 Mb s^{-1}. Moreover, although the on-board storage capacity has increased during the last decades, it is not enough to store all the information satellite instruments retrieve. Data compression techniques are crucial for an efficient on-board storage and downlink transmission. Compression is also beneficial for on-the-ground storage and dissemination of remote sensing data.

Emails: {Sara.Alvarez, Naoufal.Amrani, Joan.Serra}@uab.cat

Hyperspectral data scenes have hundreds or thousands of spectral components of similar wavelengths, resulting in scenes with high spectral redundancy. The decorrelation along the spectral dimension is then essential for a competitive coding performance. Some compression techniques exploit the data-redundancy by applying a 3D spatio-spectral transform (Tang and Pearlman, 2006; Fowler and Rucker, 2007). Other techniques apply first a 1D spectral transform followed by a 2D spatial transform, using the most suitable transform for each dimension (Penna et al., 2007; Zhang, Fowler, and Liu, 2008). For the first step, Discrete Wavelet Transforms (DWTs) are commonly used due to their computational simplicity and short execution time. However, they do not yield statistical independence in the spectral domain as Karhunen-Loève Transform (KLT) does. KLT is optimal for decorrelating Gaussian sources (Jolliffe, 2002; Effros, Feng, and Zeger, 2004); it even compacts much better the energy. Nevertheless, it is a data-dependent transform (requiring a unique computation for every scene) that entails a non-negligible side information and a high computational complexity. In fact, for hyperspectral scenes with a very large number of spectral components, it is computationally unfeasible to resort to KLT as spectral transform. The high complexity stems from the computation of the covariance matrix, the eigenvalues calculation and the matrix factorization. Additionally, KLT is not a component-scalable transform, i.e., to retrieve a single data component, all the remaining transformed components are required.

On one hand, KLT focuses on the covariance matrix, letting aside higher order moments, while DWTs assess higher order moments. On the other hand, hyperspectral scenes are not necessarily Gaussian signals neither in the spectral nor in the spatial dimension (Camps-Valls et al., 2011). On this basis, higher order moments may be meaningful. However, DWTs do not aim at the largest statistical independence among coefficients and therefore some dependence remains in the transformed domain. Since higher degree of statistical independence usually means higher efficiency in coding performance, wavelet transforms may give rise to inefficient results.

To overcome these problems, Regression Wavelet Analysis (RWA) has been recently proposed (Amrani et al., 2016).

This method provides a coding performance superior to widespread lossless techniques such as the transform-based CCSDS-122.1 (CCSDS, 2011), the predictive-based CCSDS-123.0 (CCSDS, 2012), both developed by the Consultative Committee for Space Data Systems (CCSDS), and also KLT-based techniques. It is based on applying first the suboptimal but simple Haar wavelet filter to decorrelate the spectral components, and then a regression analysis to significantly reduce the remaining statistical relationships in the wavelet domain. This method works as a predictive scheme that shares some DWT properties like component-scalability and low complexity. Besides, it reduces the energy and entropy of the original signal, leading to a representation without dynamic range expansion. Lossless compression is achieved by means of a reversible integer mapping of all the transformed components.

Departing from the original RWA proposal, further improvements can be deployed. Amrani, Serra-Sagristà, and Marcellin (2017) proposed a variation of the Restricted RWA model, where the prediction process is modified by computing a linear mathematical formulation instead of a cubic equation. However, this variation was proposed for progressive-lossy-to-lossless compression.

II. Proposed spectral transform

This section briefly surveys the main basis of the overall Regression Wavelet Analysis spectral transform. It relies on applying in sequence a wavelet transform and a regression operation. This second step significantly reduces the dependence statistical relations among the wavelet transformed components. To clarify the functioning of the algorithm, we first proceed to explain a general multi-resolution DWT decomposition of a scene.

A. Discrete wavelet transform

Let $\mathbf{V}^0 \in \mathbb{R}^{m\times z}$ be a multi-component scene with z spectral components and m spatial samples per component, i.e., $\mathbf{V}^0 = \left[\boldsymbol{V}_1^0, \ldots, \boldsymbol{V}_z^0\right]$, with $\boldsymbol{V}_i^0 \in \mathbb{R}^{m\times 1}$. A 1D-DWT with one level decomposition applied along the spectral dimension splits $\mathbf{V}^{j-1}$ into the approximation components $\mathbf{V}^j$ and the detail components $\mathbf{W}^j$ for $1 \le j \le J$, where $J = \lceil \log_2(z) \rceil$:

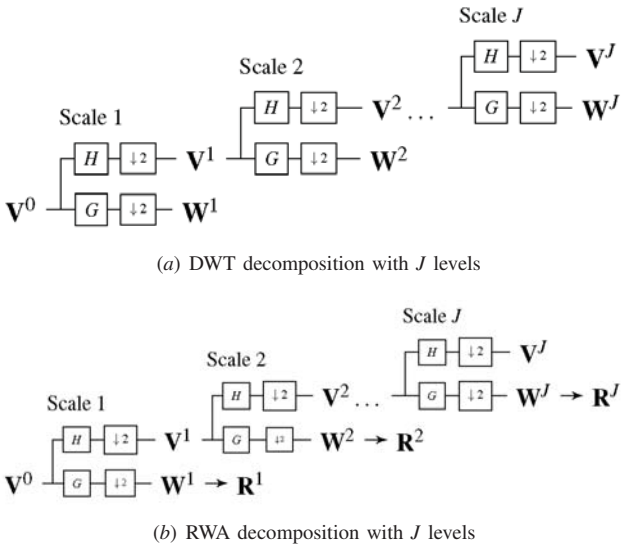
$$(\mathrm{DWT})(\mathbf{V}^{j-1}, 1) = \left(\mathbf{V}^j, \mathbf{W}^j\right). \tag{1}$$

The wavelet-transformed scene at the maximum level J is denoted as:

$$(\mathrm{DWT})(\mathbf{V}^0, J) = \left(\mathbf{V}^J, (\mathbf{W}^j)^{1\le j\le J}\right). \tag{2}$$

Figure 1(a) shows the cascade decomposition from $\mathbf{V}^0$ to the final transformed scene at level J, $(\mathrm{DWT})(\mathbf{V}^0, J) = \left[\mathbf{V}^J, \mathbf{W}^J, \mathbf{W}^{J-1}, \ldots, \mathbf{W}^1\right]$. As can be seen, a DWT is a double convolution process of the signal $\mathbf{V}^0$ through the low H and the high G pass filters to obtain the approximation and the detail components, respectively.

At each level j, the signal $\mathbf{V}^{j-1} \in \mathbb{R}^{m\times(z\cdot 2^{-j+1})}$ is decomposed in two spectral-volume halves, the approximation components $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ and the detail components $\mathbf{W}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$. The approximation signal $\mathbf{V}^j$ includes a low-pass version of the previous signal $\mathbf{V}^{j-1}$, while the detail components is the information difference between $\mathbf{V}^j$ and $\mathbf{V}^{j-1}$. The low and high decomposition process can be iteratively repeated over the approximation signal $\mathbf{V}^j$, until having only one approximation component and $z-1$ detail components (when the maximum number of decomposition levels $J = \lceil \log_2(z) \rceil$ is applied).



(*a*) DWT decomposition with J levels

(*b*) RWA decomposition with J levels

Figure 1 Spectral transforms decomposition schemes.

B. Regression wavelet analysis transform

RWA applies first a DWT followed by a regression operation to exploit the remaining spectral redundancy that the DWT can not remove (see Figure 3 to appreciate the lower correlation and mutual information of the transformed components after applying a RWA transform, as compared to applying only a wavelet transform). As RWA was designed for lossless coding, a lifting procedure is deployed (Calderbank et al., 1997), producing integer coefficients $\left(\mathbf{V}^j, \mathbf{W}^j\right)$ based on a reversible rounding operation.

Each detail component $\boldsymbol{W}_i^j \in \mathbb{R}^{m\times 1}$ at level j is estimated from the approximation components at the same level $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$. This prediction is done through a mathematical model f_i:

$$\widehat{\boldsymbol{W}}_i^j = f_i(\mathbf{V}^j). \tag{3}$$

RWA removes then the estimation, and has to store/transmit the residual (usually much cheaper) to provide lossless coding:

$$\mathbf{R}^j = \mathbf{W}^j - \widehat{\mathbf{W}}^j. \tag{4}$$

Analogous to Equations 1 and 2, the one level RWA decomposition can be expressed as $(\mathrm{RWA})(\mathbf{V}^{j-1}, 1) = \left(\mathbf{V}^j, \mathbf{R}^j\right)$ and the RWA at level J as $(\mathrm{RWA})(\mathbf{V}^0, J) = \left(\mathbf{V}^J, (\mathbf{R}^j)^{1\le j\le J}\right)$. Figure 1($b$) illustrates how this regression model affects only the detail components at each level, leaving the approximation components unchanged.

Function f_i predicts the conditional mean of each $\boldsymbol{W}_i^j \in \mathbb{R}^{m\times 1}$ from some or all the approximation components $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$.

In the inverse process, the approximation component at level $j-1$ is obtained by reverting Equation 4 and Equation 1 respectively.

III. Regression wavelet analysis transform features

This section discusses several properties and characteristics of different models and variants of the Regression Wavelet Analysis transform.

A. Wavelet filters

In the original proposal, RWA employs Haar-DWT as its first operation block because of its inexpensive computational complexity and high component-scalability. However, it yields suboptimal energy compaction when compared to other DWTs that employ more components in the transform. Here we assess the suitability of exchanging Haar-DWT by other more efficient DWTs like the reversible IWT 5/3 or IWT 9/7M (integer version of the biorthogonal CDF 9/7).

To produce one approximation and one detail component in the transformed signal, Haar-DWT uses only 2 coefficients (scene components), while IWT 5/3 uses, respectively, 5 and 3 coefficients for the low-pass and the high-pass filters, and IWT 9/7M uses, respectively, 9 and 7 coefficients.

B. Models and variants

To predict each detail component $\widehat{\boldsymbol{W}}_i^j \in \mathbb{R}^{m\times 1}$, $i \in \mathrm{I} = \{1,\ldots,k{=}z{\cdot}2^{-j}\}$ at level j, Equation 3 is differently formulated for the different regression models and variants: Maximum model, Restricted model, Fast-Maximum variant, Fast-Restricted variant, Exogenous-Maximum variant and Exogenous-Restricted variant.

(*i*) **Maximum model:** includes all the approximation components $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ in the prediction of each detail component $\boldsymbol{W}_i^j \in \mathbb{R}^{m\times 1}$. It is a linear model and the most general one. It is described as follows:

$$\widehat{\boldsymbol{W}}_i^j = \beta_{i,0}^j + \beta_{i,1}^j \boldsymbol{V}_1^j + \cdots + \beta_{i,k}^j \boldsymbol{V}_k^j, \ \left(\boldsymbol{V}_i^j \in \mathbb{R}^{m\times 1}\right). \quad (5)$$

This model generally corresponds to the most accurate one in terms of prediction error.

(*ii*) **Restricted model:** includes in the prediction of $\widehat{\boldsymbol{W}}_i^j$ a subset of approximation components $\boldsymbol{V}_{u\in\mathfrak{J}}^j$, where $\mathfrak{J} \subset \{1,\ldots,n\}$. This subset of approximation components is selected in order to maintain the component-scalability of the DWT filter. Here scalability refers to the number of transformed components required to reconstruct a single scene component; the smallest the number, the largest the scalability of the transform.
When Haar-DWT is the chosen wavelet filter, Restricted model may be computed as:

$$\widehat{\boldsymbol{W}}_i^j = \beta_{i,0}^j + \beta_{i,1}^j \boldsymbol{V}_i^j + \beta_{i,2}^j (\boldsymbol{V}_i^j)^2 + \beta_{i,3}^j (\boldsymbol{V}_i^j)^3. \quad (6)$$

For this filter, only one approximation component $\boldsymbol{V}_i^j$ is involved in the prediction of $\widehat{\boldsymbol{W}}_i^j$, but second and third-order terms are also used to increase the prediction accuracy without affecting the scalability.
Restricted model when using IWT 5/3 or 9/7M is respectively defined as follows:

$$\widehat{\boldsymbol{W}}_i^j = \beta_{i,0}^j + \beta_{i,1}^j \mathbf{V}_{i-1}^j + \beta_{i,2}^j \mathbf{V}_i^j + \beta_{i,3}^j \mathbf{V}_{i+1}^j + \beta_{i,4}^j \mathbf{V}_{i+2}^j \quad (7)$$

and

$$\widehat{\boldsymbol{W}}_i^j = \beta_{i,0}^j + \beta_{i,1}^j \mathbf{V}_{i-3}^j + \ldots + \beta_{i,4}^j \mathbf{V}_i^j + \ldots + \beta_{i,8}^j \mathbf{V}_{i+4}^j. \quad (8)$$

The regression parameters $\boldsymbol{\beta}^j$ for Equations 5, 6, 7 and 8 are computed through a least squares method to minimize the squared residual length:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\mathbf{W} - \boldsymbol{\beta}\mathbf{V}\|_2. \quad (9)$$

The parameters $\boldsymbol{\beta}^j$ need to be stored/transmitted as side information for the decoding procedure.

(*iii*) **Fast variant:** uses only a subset of $m'{=}\rho m$ spatial samples ($m' \ll m$) to compute the least squares parameters $\boldsymbol{\beta}^j$. The computational complexity is significantly reduced, specially for scenes with a large spatial resolution. In general, this spatial sub-sampling barely affects the quality of the estimation step and considerably reduces the cost of computing the regression parameters $\boldsymbol{\beta}$.

- **Fast-Maximum:** applies the aforementioned subsampling to the components included in the Maximum model (Equation 5).
- **Fast-Restricted:** applies the aforementioned subsampling to the components included in the Restricted model (Equations 6, 7 and 8).

(*iv*) **Exogenous variant:** uses fixed regression parameters $\overline{\boldsymbol{\beta}}^j$, obtained by training a single scene from an instrument corpus. This can be done since hyperspectral scenes recorded by the same instrument may have similar statistical relationships. Coefficients $\overline{\boldsymbol{\beta}}^j$ are computed only once, for a single scene, and are then used for coding the other scenes in the corpus. Hence, the side information from the training procedure does not need to be stored. This variant gives rise to lower computational complexity and a significant execution time reduction, because the learning process is performed only once. In addition, it possibly yields higher compression ratios, because side information is spared.

- **Exogenous-Maximum:** relies on applying Equation 5, but by using fixed parameters to any new scene recorded by the same instrument. The formulation remains as follows:

$$\widehat{\boldsymbol{W}}_i^j = \overline{\beta}_{i,0}^j + \overline{\beta}_{i,1}^j \boldsymbol{V}_1^j + \ldots + \overline{\beta}_{i,k}^j \boldsymbol{V}_k^j. \quad (10)$$

- **Exogenous-Restricted:** applies Equations 6, 7 and 8 using fixed parameters for the estimation of $\widehat{\boldsymbol{W}}_i^j$. Note that the computation of the fixed regression parameters is done through the Restricted model, yielding a reduced computational cost.

Table I Restricted model's component-scalability when different wavelet filters are employed as the basis of the RWA first operation.

Haar-RWA	IWT 5/3-RWA	IWT 9/7M-RWA
$\left\{ \boldsymbol{V}_{2i}^{j-1} \right\} \xleftarrow{\text{Reconst.}} \boldsymbol{V}_{i}^{j}$	$\left\{ \boldsymbol{V}_{2i}^{j-1} \right\} \xleftarrow{\text{Reconst.}} \left\{ \boldsymbol{V}_{u}^{j} \right\}_{u=i-1}^{i+2}$	$\left\{ \boldsymbol{V}_{2i}^{j-1} \right\} \xleftarrow{\text{Reconst.}} \left\{ \boldsymbol{V}_{u}^{j} \right\}_{u=i-3}^{i+4}$

C. Component-scalability

The component-scalability depends on the regression model and on the wavelet filter used as the basis of the RWA first operation.

Concerning the regression model, Maximum model has the worst component-scalability, as all approximation components in a given wavelet decomposition level are employed for predicting a particular detail component, meaning that all the approximation components would be required to invert the spectral transform. On the other hand, the Restricted model preserves the highest scalability. For Restricted model using Haar-DWT, only one low and one high-pass components from level j are needed to reconstruct one approximation coefficient at level $j-1$ (Equation 6).

Concerning the wavelet filter, for a given filter, the number of approximation $\boldsymbol{V}_{i\in \mathrm{I}_r}^{j}$ and details $\boldsymbol{W}_{i\in \mathrm{I}_q}^{j}$ components used to retrieve an approximation component at a lower level, $\boldsymbol{V}_{r_i}^{j-1}$, depends on the number of components employed from the forward $j-1$ level to the j level. Here, I_r and I_q are sets of components for, respectively, the low and high-pass. By extension, a set of approximation components from a higher level ($\boldsymbol{V}_{i\in \mathrm{I}_{r_i}}^{j}$) is needed to reconstruct the approximation component $\boldsymbol{V}_{r_i}^{j-1}$:

$$\begin{cases} \left\{ \boldsymbol{V}_{r_1}^{j-1} \right\} \xleftarrow{\text{Reconst.}} \boldsymbol{V}_{i\in \mathrm{I}_{r_1}}^{j} \\ \qquad \vdots \qquad\qquad\quad \vdots \\ \left\{ \boldsymbol{V}_{r_t}^{j-1} \right\} \xleftarrow{\text{Reconst.}} \boldsymbol{V}_{i\in \mathrm{I}_{r_t}}^{j}. \end{cases}$$

The scalability relationships for the three reversible wavelet filters when the Restricted model is applied are given in Table I.

The reconstruction of $\boldsymbol{V}_{2i}^{j-1}$ depends also on several detail components from higher transform levels. Here, we have only considered the low-pass components, since the regression model in RWA selects the regressors from the approximation.

D. Computational complexity

The computational cost of RWA is displayed in Figures 2(a)-2(c) in Floating-Point Operations (FLOPs). These figures compare the computational complexity of different RWA regression models and variants for, respectively, uncalibrated Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Hyperion and Infrared Atmospheric Sounding Interferometer (IASI) scenes. For comparison purposes, we include the computational complexity of rKLT and Haar-DWT, IWT 5/3 and IWT 9/7M. In the figures, the cost of RWA transform is plotted in bars with different colors depending of the wavelet filter used as the basis of RWA: blue, yellow and green are used for, respectively, IWT 9/7M, IWT 5/3 and Haar-DWT.

Although rKLT usually attains very competitive compression ratios, it is the most expensive spectral transform. RWA with the Maximum model is the second most expensive spectral transform, significantly lower than rKLT. As expected, Restricted model has a much lower complexity compared to Maximum model, since only a small set of components are used as regressors. Fast-Maximum and Fast-Restricted variants considerably reduce the cost of Maximum and Restricted model respectively, since the computation and application of least squares parameters is significantly alleviated. Exogenous variant offers a slightly lower computational complexity than Fast, as the computation of the regression parameters is conducted off-line.

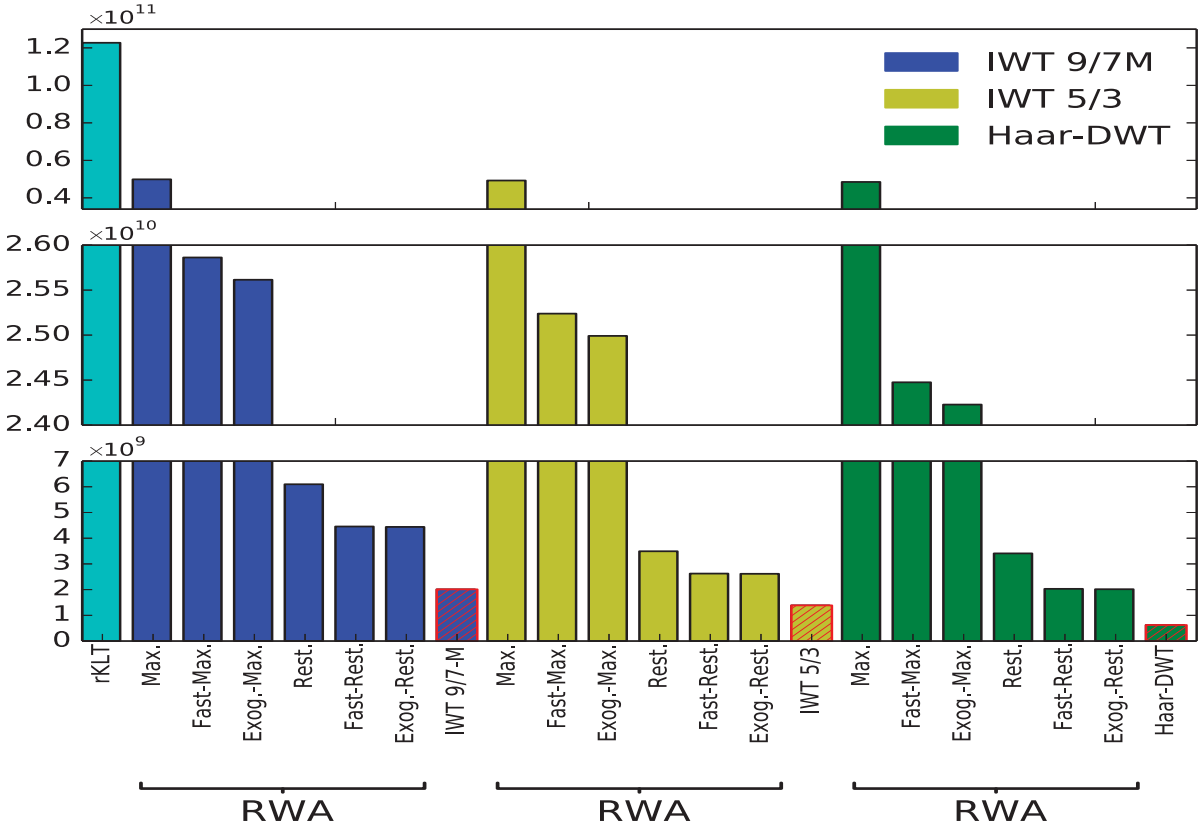
The leverage of the wavelet filter used as the basis of RWA is also apparent, being due to the number of coefficients needed to apply the transform. Haar-DWT entails a lower cost than IWT 5/3, which, in turn, is cheaper than IWT 9/7M.

E. Execution time

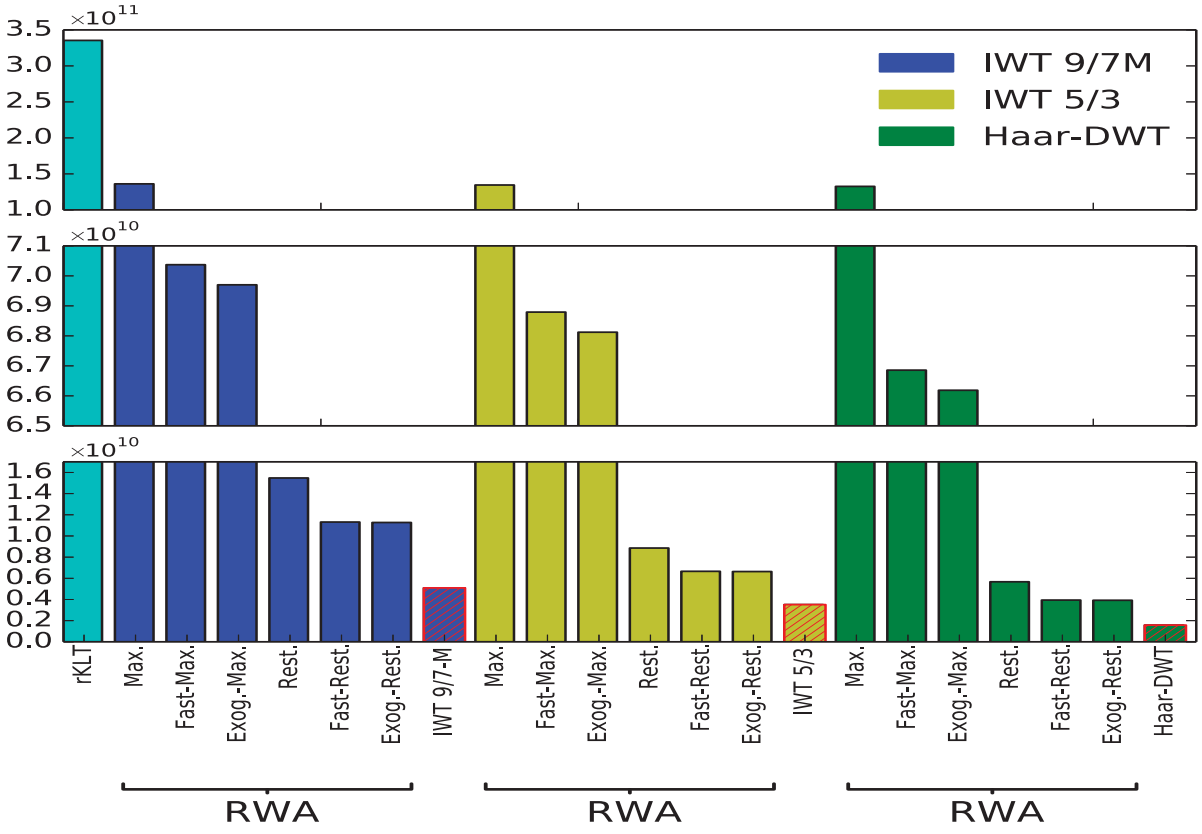
Still referring to the computational cost, Table II reports average execution time when executing all RWA regression models and variants on an Intel(R) Xeon(R) CPU E5520 @ 2.27GHz processor with a single thread. Results for Haar-DWT, IWT 5/3 and IWT 9/7M as wavelet transforms for the first RWA operation block are included for all models and variants.

Exogenous-Maximum variant usually provides the fastest execution, because the computation of the regression parameters is performed off-line. It would be expected that Exogenous-Restricted variant entailed even lower execution time, but this behavior does not happen because, for each detail estimation, different approximation components have to be selected, and this sequential selection prevents a faster matrix operation.

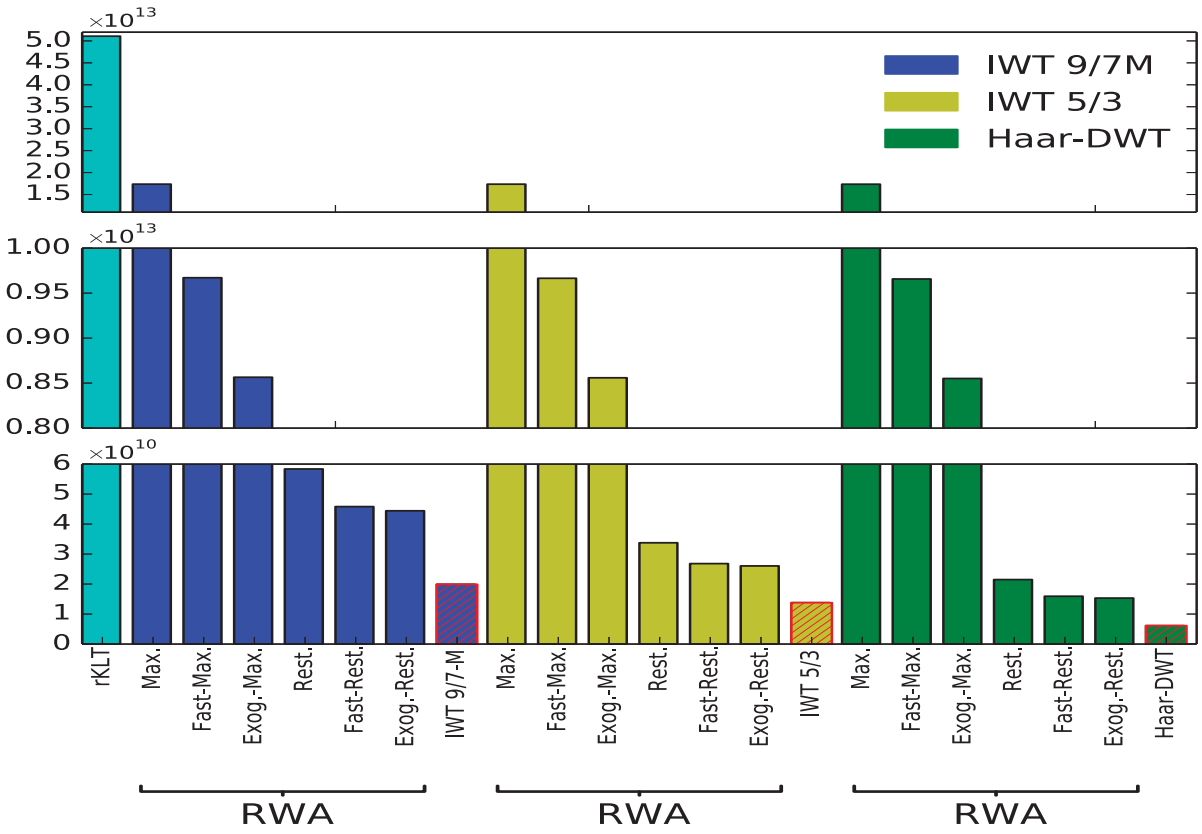
For Maximum model and its variants, which employ all components in a wavelet decomposition level to estimate a detail component, the choice of the wavelet filter dictates the complexity, with Haar-DWT showing a faster execution time than IWT 5/3 and IWT 9/7M. For Restricted model and its variants, IWT 5/3 yields the fastest execution time. IWT 9/7M is more expensive because it employs more coefficients for the wavelet decomposition and more components are needed for the regression. Haar-DWT is here more expensive because the regression model involves a polynomial of order 3, requiring to compute the second and third power of an approximation component.



(*a*) AVIRIS scene (224 spectral components and 512 x 680 spatial samples) and $\rho = 0.01$.



(*b*) Hyperion scene (242 spectral components and 256 x 3187 spatial samples) and $\rho = 0.01$.



(*c*) IASI scene (8359 spectral components and 764 x 30 x 4 spatial samples) and $\rho = 0.1$.

Figure 2 Computational cost comparison in FLOPs for different spectral transforms and for three uncalibrated scenes. The abbreviated terms Max., Rest. and Exog. correspond respectively to the Maximum and Restricted models and the Exogenous variant.

IV. Signal metrics for the performance evaluation

Several measurements are considered to appraise the performance of the different regression models and variants when used as spectral transform, including average bit-rates, coefficient of determination or squared Pearson's correlation coefficient r^2, energy percentage $\mathscr{E}(\%)$, Shannon entropy $\mathscr{H}$ and mutual information.

The squared correlation coefficient r^2 of two signals $\mathbf{X}$ and $\mathbf{Y}$ is defined as:

$$r^2 = \left(\frac{\mathrm{Cov}(\mathbf{X},\mathbf{Y})}{\sigma_X \sigma_Y} \right)^2, \tag{11}$$

where $\mathrm{Cov}(\mathbf{X},\mathbf{Y})$ is the covariance matrix of the signals $\mathbf{X}$ and $\mathbf{Y}$, and σ_X, σ_Y are their standard deviations. The correlation matrix for a scene with z spectral components is a square matrix $z \times z$ that depicts the correlation in terms of r^2 between each pair of spectral components.

The energy percentage is given by:

$$\mathscr{E}(\%) = \frac{\sum_{k=1}^{z} \sum_{r=1}^{m} ((\mathrm{RWA})\,(\mathbf{V}^0, J))^2}{\sum_{k=1}^{z} \sum_{r=1}^{m} (\mathbf{V}^0)^2}. \tag{12}$$

The Shannon entropy, measured in bits, is given by:

$$\mathscr{H} = -\sum_{i=1}^{n} P(v_i) \log_2 P(v_i), \tag{13}$$

where $P(v_i)$ defines the probability mass function of v_i, and $\boldsymbol{V} = [v_1, .., v_n]$ is the vectorization of the spatial samples of each spectral component of the scene and $n{=}zm$.

Mutual information can describe linear and non-linear statistical relationships between variables. It indeed includes correlation and non-Gaussianity. Therefore, a change in the mutual information analysis can come from a variation of the decorrelation and/or the statistical dependencies between components. Reducing then the correlation usually means a decrease in the mutual information. In its turn, correlation only takes into account linear variations. The mutual information between two variables $\mathbf{X}$ and $\mathbf{Y}$ is defined as follows:

$$\mathscr{MI}(\mathbf{X},\mathbf{Y}) = \mathscr{H}(\mathbf{X},\mathbf{Y}) - \mathscr{H}(\mathbf{X}|\mathbf{Y}) - \mathscr{H}(\mathbf{Y}|\mathbf{X}), \tag{14}$$

where $\mathscr{H}(\mathbf{X},\mathbf{Y})$ is the joint entropy, and $\mathscr{H}(\mathbf{X}|\mathbf{Y})$ and $\mathscr{H}(\mathbf{Y}|\mathbf{X})$ are the conditional entropies. A Matlab implementation can be found in Chen (2010).

To illustrate the impact of the regression, Figure 3 displays the mutual information and the squared correlation coefficient

Table II Average execution time (in seconds) for several RWA models and variants. The largest number of wavelet decomposition levels has been applied in all cases.

	Average execution time (s)						
	AVIRIS			Hyperion		IASI	
	16-bit calibrated scenes	16-bit uncalibrated scenes	12-bit uncalibrated scenes	16-bit calibrated scenes	16-bit uncalibrated scenes	16-bit calibrated scenes	16-bit uncalibrated scenes
Maximum Haar-RWA	69.36	78.19	63.75	198.04	193.75	4413.73	4185.43
Maximum IWT 5/3-RWA	69.68	78.23	63.67	198.88	193.25	4436.06	4188.35
Maximum IWT 9/7M-RWA	73.45	81.89	67.15	209.02	202.09	4561.62	4324.99
Fast-Maximum Haar-RWA	44.96	53.12	39.93	133.12	123.92	1990.50	1819.67
Fast-Maximum IWT 5/3-RWA	45.21	53.38	40.65	133.53	123.21	1999.37	1827.24
Fast-Maximum IWT 9/7M-RWA	48.77	56.83	43.40	142.93	132.76	2139.09	1952.73
Exog.-Maximum Haar-RWA	45.06	53.06	35.57	130.34	124.06	1524.87	1409.92
Exog.-Maximum IWT 5/3-RWA	45.24	53.13	36.27	135.83	129.08	1534.57	1412.13
Exog.-Maximum IWT 9/7M-RWA	46.47	54.05	39.66	135.82	129.01	1670.52	1414.51
Restricted Haar-RWA	90.52	100.19	84.56	229.84	229.62	1136.31	958.21
Restricted IWT 5/3-RWA	73.64	81.74	67.57	197.90	188.12	930.74	784.36
Restricted IWT 9/7M-RWA	102.86	112.17	96.19	286.29	280.28	1317.67	1182.80
Fast-Restricted Haar-RWA	82.63	92.04	75.95	204.21	202.99	1066.73	898.57
Fast-Restricted IWT 5/3-RWA	61.45	69.68	55.63	166.41	153.87	833.97	689.54
Fast-Restricted IWT 9/7M-RWA	76.87	84.67	70.93	215.33	206.08	1142.22	987.39
Exog.-Restricted Haar-RWA	73.58	81.89	66.88	178.56	174.39	1031.30	822.84
Exog.-Restricted IWT 5/3-RWA	67.19	75.84	48.23	200.13	199.42	811.72	824.62
Exog.-Restricted IWT 9/7M-RWA	71.82	79.79	63.20	179.44	176.71	1104.43	829.29

matrices and their average values after transforming uncalibrated AVIRIS Yellowstone03 scene. In all subfigures, the blue-to-red colormap specifies the low-to-high statistical relationships (colorbars) and the diagonal coefficients represent the relation between each component of the scene with itself. The figure shows that Haar-DWT provides worse average mutual information and squared correlation than the other two IWT filters. Although a slight relationship between components can still be noticed when applying RWA, particularly for the first and more significant transformed components, the figure points out a clear benefit over traditional wavelet transforms and even over the reversible KLT (rKLT). RWA reaches a better average mutual information measurement and very similar average squared correlation to rKLT, with a difference of only 6×10^{-4}. Outcomes for the other signal measurements are reported in next section.

V. Experimental results

In this section we present a comprehensive assessment of the proposed RWA models and variants when used as spectral transforms in a coding system for lossless compression.

Figure 4 displays the two main blocks of our coding system. In the first stage, a 1D transform along the spectral dimension is applied. In the second stage, JPEG 2000 standard (Taubman and Marcellin , 2001) is used. Several 1D spectral transforms are assessed: rKLT, POT, wavelets and RWA with Haar-DWT, IWT 5/3 and IWT 9/7M as underlying wavelet filters. For the second block, JPEG 2000 applies a 2D spatial IWT 5/3 with 5 levels for POT, Haar-DWT, RWA Restricted and Parsimonious Restricted models and their variants and for RWA when transforming Hyperion scenes. This configuration has been found to produce the best performance in compression ratio and energy compaction.

To account for reproducibility, Matlab implementations of rKLT, Haar-DWT, IWT 5/3 and RWA are publicly available at Amrani and Serra-Sagristà (2015). IWT 9/7M definition can be found in CCSDS (2011). The employed software is Matlab 64-bits R2014b for RWA computation and Kakadu v6.0 for JPEG 2000.

Experimental results are reported for six hyperspectral scenes corpora from three well-known instruments: airborne AVIRIS and satellite-borne hyperspectral Hyperion and ultraspectral IASI. AVIRIS instrument stores scenes with a bit-depth of 16 bits per pixel per component (bpppc), except for two 12 bpppc scenes, Hawaii and Maine. Hyperion and IASI instruments store scenes with 16 bpppc. AVIRIS, Hyperion and IASI instruments record scenes with, respectively, 224, 242 and 8,359 (uncalibrated) or 8,461 (calibrated) spectral components. Table III provides some information of the employed scenes.

A. Influence of wavelet filters in RWA

Tables IV and VI provide the average bit per pixel per component (bpppc) when encoding the scenes for each in-

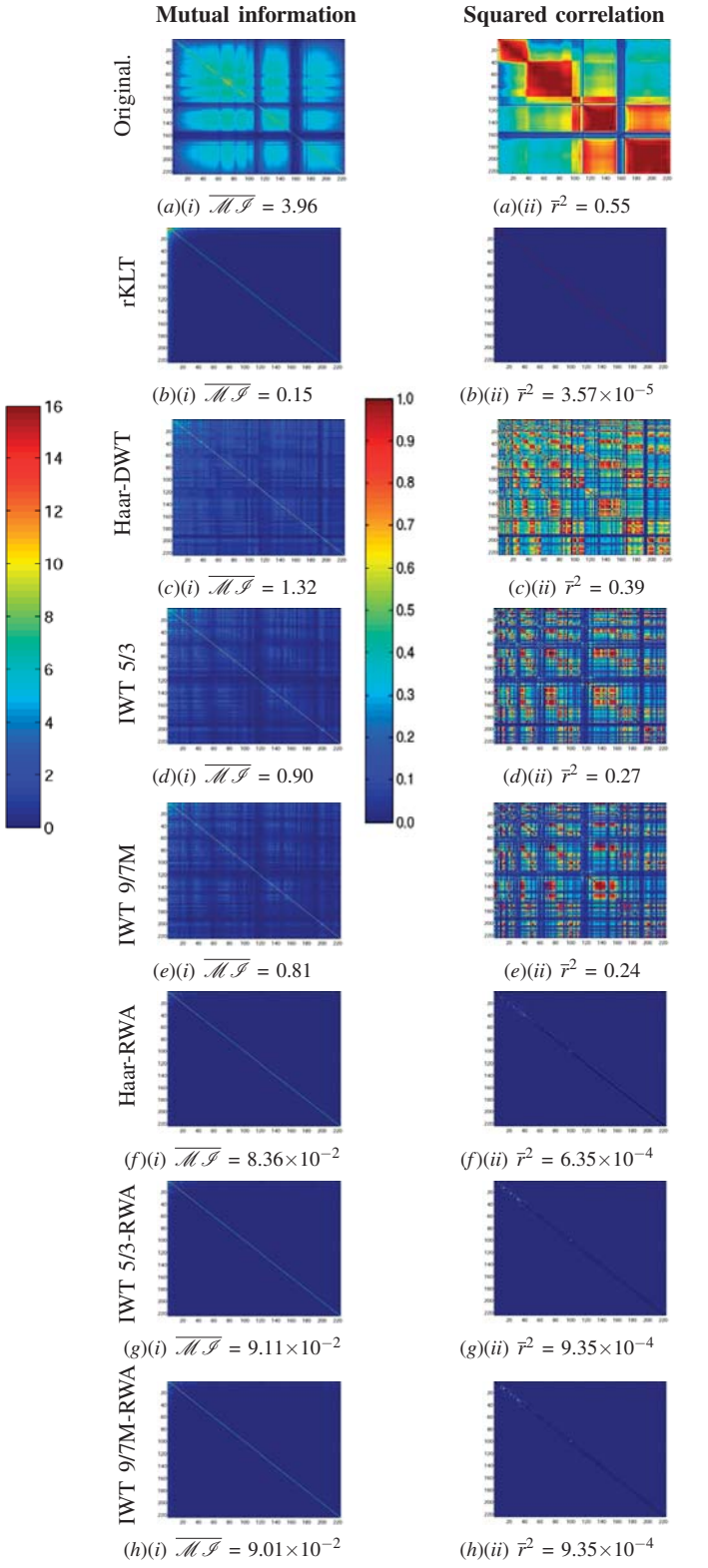


Figure 3 Mutual information and squared correlation matrices when transforming the uncalibrated AVIRIS Yellowstone03 scene. For RWA, the Maximum model with the largest number of wavelet decomposition levels is applied. Colorbars indicating the low-to-high range values of mutual information (average given by $\overline{\mathscr{MI}}$) and correlation (average given by $\bar{r}^2$) are included.

Figure 4 Pipeline of the coding system: 1D spectral transform followed by JPEG 2000.

Table III Characteristics of scenes from AVIRIS, Hyperion and IASI instruments used in the experiments. z is the number of spectral components, y is the height and x is the width.

Corpus	Scenes
AVIRIS (Uncalibrated) z=224, y=512, x=680	Yellowstone, sc: 00, 03, 10, 11, 18 Hawaii (x=614) and Maine
AVIRIS (Calibrated) z=224, y=512, x=680	Yellowstone, sc: 00, 03, 10, 11, 18
Hyperion (Uncalibrated) z=242 x=256	ErtaAle (y=3187) Lake Monona (y=3176) Mt. St. Helens (y=3242)
Hyperion (Calibrated) z=242 x=256	Agriculture (y=3129) Coral Reef (y=3127) Urban (y=2905)
IASI Level 0 (Uncalibrated) z=8359 y=1528 x=60	L0 1: 20091007093900Z L0 2: 20091007143900Z L0 3: 20100319050300S6 L0 4: 20120718075700Z
IASI Level 1 (Calibrated) z=8461, y=1530, x=60	L1 1: 20130816230553Z L1 2: 20130817004753Z

vestigated regression model and variant under different RWA settings. Table VIII presents bit-rate values of Parsimonious Haar-Restricted (Amrani, Serra-Sagristà, and Marcellin, 2017) and for the novel variants Parsimonious Fast-Restricted and Parsimonious Exogenous-Restricted. Only average values are reported due to space constraints; however, all scenes within a particular corpus give rise to similar performance.

Comparing the results for RWA Maximum model with reversible Haar-DWT, IWT 5/3 or 9/7M wavelet filters as the underlying transform, we can conclude that the coding performance seems not to depend much on the wavelet filter. Even though IWT 5/3 and 9/7M remove the dependencies better than Haar-DWT (Figure 3), the regression operation is able to deal with the dependence statistical relationships regardless of the wavelet filter. As an example, for calibrated AVIRIS corpus, the average bit-rate when applying 8 spectral decomposition levels is 3.52, 3.53 and 3.54 for, respectively, Haar, IWT 5/3 and IWT 9/7M (Table VI).

On the other hand, regarding the Restricted model and its variants, the wavelet filter used as the basis of the RWA first operation meaningfully affects the coding performance (Tables IV and VI). For these cases, IWT 9/7M-RWA obtains better outcomes than Haar-RWA for uncalibrated and calibrated AVIRIS and calibrated IASI scenes. For uncalibrated IASI scenes, Restricted IWT 9/7M-RWA and variants provide similar coding performance than Haar-RWA. Presented com-

Table IV Average RWA bit-rate (in bpppc) for uncalibrated AVIRIS corpus. The training process for the Exogenous variants has been done for, respectively, Yellowstone00 and Hawaii scene for Yellowstone scenes and Maine scene.

Wavelet decomposition levels	Spectral wavelet filter	Mean bpppc Maximum		Mean bpppc Fast ($\rho = 0.01$)		Mean bpppc Restricted		Mean bpppc Fast-Restricted ($\rho = 0.01$)		Mean bpppc Exog.-Maximum		Mean bpppc Exog.-Restricted	
		Y. 00-18	Hawaii Maine	Y. 00-18	Hawaii Maine	Y. 00-18	Hawaii Maine	Y. 00-18	Hawaii Maine	Y. 03-18	Maine	Y. 03-18	Maine
0	—	10.45	6.52	10.45	6.52	8.88	4.58	8.88	4.58	—	—	—	—
1	Haar	7.86	4.34	7.88	4.35	7.54	3.77	7.53	3.77	7.77	4.37	7.44	3.79
	IWT 5/3	7.78	4.27	7.80	4.28	7.29	3.65	7.29	3.65	7.69	4.30	7.20	3.69
	IWT 9/7M	7.81	4.29	7.82	4.31	7.29	3.67	7.29	3.67	7.69	4.31	7.21	3.70
5	Haar	5.87	2.67	5.89	2.69	6.44	3.02	6.42	3.02	5.88	2.86	6.37	3.05
	IWT 5/3	5.87	2.70	5.89	2.72	6.00	2.89	6.00	2.89	5.87	2.90	5.97	2.93
	IWT 9/7M	5.88	2.73	5.90	2.75	5.93	2.87	5.93	2.87	5.87	2.91	5.92	2.92
8	Haar	5.83	2.61	5.85	2.63	6.44	3.01	6.42	3.01	5.84	2.82	6.37	3.05
	IWT 5/3	5.85	2.66	5.87	2.68	6.00	2.88	6.00	2.88	5.85	2.88	5.98	2.92
	IWT 9/7M	5.85	2.68	5.87	2.70	5.94	2.87	5.94	2.87	5.85	2.89	5.92	2.91

Table V Signal measurements when applying RWA Maximum and Fast-Restricted to uncalibrated AVIRIS corpus.

Wavelet decomposition levels	Spectral wavelet filter	Maximum						Fast-Restricted					
		Mean r^2		Mean $\mathscr{E}(\%)$		Mean $\mathscr{H}$		Mean r^2		Mean $\mathscr{E}(\%)$		Mean $\mathscr{H}$	
		Y. 00-18	Hawaii Maine	Y. 00-18	Hawaii Maine	Y. 00-18	Hawaii Maine	Y. 00-18	Hawaii Maine	Y. 00-18	Hawaii Maine	Y. 00-18	Hawaii Maine
0	—	0.61	0.46	100	100	12.16	8.82	0.61	0.46	100	100	12.16	8.82
1	Haar	0.15	0.11	49.90	49.39	9.87	6.70	0.17	0.14	49.91	49.39	10.39	7.02
	IWT 5/3	0.15	0.11	50.01	50.18	9.79	6.63	0.15	0.11	50.01	50.18	9.88	6.67
	IWT 9/7M	0.15	0.11	49.93	49.89	9.81	6.65	0.15	0.11	49.93	49.89	9.87	6.68
5	Haar	1.27×10^{-3}	8.51×10^{-4}	2.96	2.79	6.17	2.99	0.06	0.05	3.00	2.82	7.43	3.92
	IWT 5/3	1.14×10^{-3}	8.65×10^{-4}	3.15	2.88	6.18	3.01	5.93×10^{-3}	3.33×10^{-3}	3.16	2.88	6.46	3.20
	IWT 9/7M	3.12×10^{-3}	9.68×10^{-4}	3.11	2.79	6.18	3.03	3.03×10^{-3}	1.84×10^{-3}	3.11	2.79	6.30	3.10
8	Haar	7.53×10^{-4}	3.79×10^{-4}	0.37	0.31	6.02	2.81	0.06	0.05	0.42	0.35	7.27	3.75
	IWT 5/3	8.31×10^{-4}	5.81×10^{-4}	0.48	0.40	6.05	2.86	5.92×10^{-3}	3.29×10^{-3}	0.49	0.40	6.35	3.05
	IWT 9/7M	2.65×10^{-3}	5.83×10^{-4}	0.47	0.40	6.05	2.87	2.71×10^{-3}	1.56×10^{-3}	0.47	0.40	6.17	2.94

Table VI Average RWA bit-rate (in bpppc) for the different corpora. The training process for the Exogenous variants has been conducted on, respectively, Yellowstone00, Erta Ale, Agriculture, IASI L0 1 and IASI L1 1. ρ parameter is 0.01 for AVIRIS and Hyperion corpora and 0.1 for IASI corpus.

Wavelet decomp. levels	Spectral wavelet filter	Mean bpppc Maximum	Mean bpppc Fast	Mean bpppc Restricted	Mean bpppc Fast-Rest.	Mean bpppc Exog.-Max.	Mean bpppc Exog.-Rest.
		AVIRIS (calibrated)					
0	—	7.08	7.08	6.55	6.55	—	—
1	Haar	5.27	5.29	5.25	5.25	5.16	5.17
	IWT 5/3	5.19	5.20	5.03	5.03	5.08	4.94
	IWT 9/7M	5.22	5.23	5.03	5.03	5.11	4.95
5	Haar	3.56	3.58	4.20	4.19	3.56	4.17
	IWT 5/3	3.55	3.57	3.76	3.76	3.55	3.74
	IWT 9/7M	3.57	3.59	3.72	3.72	3.57	3.70
8	Haar	3.52	3.54	4.19	4.18	3.52	4.16
	IWT 5/3	3.53	3.55	3.75	3.75	3.53	3.73
	IWT 9/7M	3.54	3.56	3.71	3.71	3.54	3.69
		Hyperion (uncalibrated)					
0	—	5.11	5.11	5.11	5.11	—	—
1	Haar	4.82	4.82	4.66	4.65	4.92	4.70
	IWT 5/3	4.80	4.80	4.73	4.73	4.89	4.77
	IWT 9/7M	4.81	4.82	4.75	4.75	4.91	4.80
5	Haar	4.54	4.55	4.38	4.38	4.75	4.44
	IWT 5/3	4.56	4.57	4.47	4.47	4.77	4.56
	IWT 9/7M	4.57	4.58	4.49	4.49	4.78	4.62
8	Haar	4.53	4.54	4.37	4.37	4.75	4.43
	IWT 5/3	4.56	4.57	4.47	4.47	4.77	4.56
	IWT 9/7M	4.57	4.58	4.49	4.49	4.78	4.62
		Hyperion (calibrated)					
0	—	6.42	6.42	6.42	6.42	—	—
1	Haar	5.74	5.74	5.77	5.77	5.67	5.68
	IWT 5/3	5.77	5.77	5.82	5.81	5.70	5.74
	IWT 9/7M	5.86	5.86	5.87	5.86	5.79	5.79
5	Haar	5.32	5.33	5.43	5.38	5.32	5.40
	IWT 5/3	5.41	5.42	5.50	5.58	5.42	5.57
	IWT 9/7M	5.49	5.50	5.58	5.58	5.50	5.58
8	Haar	5.33	5.34	5.44	5.37	5.34	5.41
	IWT 5/3	5.42	5.42	5.59	5.58	5.43	5.58
	IWT 9/7M	5.49	5.50	5.58	5.58	5.50	5.59
		IASI L0 (uncalibrated)					
0	—	5.65	5.65	4.18	4.18	—	—
1	Haar	3.83 + 0.68	4.04 + 0.68	$3.60 + 7.92\times10^{-4}$	$3.60 + 7.92\times10^{-4}$	3.86	3.60
	IWT 5/3	3.77 + 0.68	3.98 + 0.68	$3.54 + 9.46\times10^{-4}$	$3.55 + 9.46\times10^{-4}$	3.80	3.55
	IWT 9/7M	3.80 + 0.68	4.01 + 0.68	$3.55 + 1.61\times10^{-3}$	$3.55 + 1.62\times10^{-3}$	3.83	3.55
5	Haar	2.49 + 0.91	2.77 + 0.91	$2.95 + 1.53\times10^{-3}$	$2.96 + 1.53\times10^{-3}$	2.54	2.95
	IWT 5/3	2.55 + 0.91	2.81 + 0.91	$2.96 + 1.83\times10^{-3}$	$2.97 + 1.83\times10^{-3}$	2.60	2.96
	IWT 9/7M	2.58 + 0.91	2.85 + 0.91	$2.95 + 3.13\times10^{-3}$	$2.95 + 3.13\times10^{-3}$	2.63	2.95
14	Haar	2.40 + 0.91	2.67 + 0.91	$2.91 + 1.58\times10^{-3}$	$2.91 + 1.58\times10^{-3}$	2.44	2.90
	IWT 5/3	2.46 + 0.91	2.73 + 0.91	$2.92 + 1.89\times10^{-3}$	$2.92 + 1.89\times10^{-3}$	2.51	2.92
	IWT 9/7M	2.49 + 0.91	2.76 + 0.91	$2.90 + 3.23\times10^{-3}$	$2.91 + 3.23\times10^{-3}$	2.54	2.90
		IASI L1 (calibrated)					
0	—	10.10	10.10	9.26	9.25	—	—
1	Haar	8.14 + 0.70	8.37 + 0.70	$8.11 + 7.92\times10^{-4}$	$8.10 + 7.93\times10^{-4}$	8.17	8.11
	IWT 5/3	7.95 + 0.70	8.18 + 0.70	$7.72 + 9.37\times10^{-4}$	$7.71 + 9.38\times10^{-4}$	7.99	7.72
	IWT 9/7M	7.97 + 0.70	8.20 + 0.70	$7.70 + 1.62\times10^{-3}$	$7.70 + 1.62\times10^{-3}$	8.01	7.70
5	Haar	6.55 + 0.93	6.84 + 0.93	$7.18 + 1.53\times10^{-3}$	$7.18 + 1.53\times10^{-3}$	6.60	7.18
	IWT 5/3	6.47+ 0.93	6.76 + 0.93	$6.82 + 1.82\times10^{-3}$	$6.82 + 1.83\times10^{-3}$	6.52	6.82
	IWT 9/7M	6.47 + 0.93	6.76 + 0.93	$6.74 + 3.14\times10^{-3}$	$6.75 + 3.14\times10^{-3}$	6.52	6.74
14	Haar	6.41 + 0.93	6.70 + 0.93	$7.11 + 1.57\times10^{-3}$	$7.11 + 1.58\times10^{-3}$	6.46	7.11
	IWT 5/3	6.35 + 0.93	6.64 + 0.93	$6.75 + 1.88\times10^{-3}$	$6.75 + 1.88\times10^{-3}$	6.40	6.74
	IWT 9/7M	6.35 + 0.93	6.64 + 0.93	$6.67 + 3.23\times10^{-3}$	$6.67 + 3.24\times10^{-3}$	6.40	6.66

pression ratios are also higher than those reported in Amrani et al. (2016) and in Amrani, Serra-Sagristà, and Marcellin (2017).

B. Coding performance of RWA models, variants and decomposition levels

For all regression models and variants, and for all wavelet filters, increasing the number of wavelet decomposition levels produces an improvement in the coding performance of RWA approaches, even though the number of regressors at higher decomposition levels keeps decreasing. Results reported in Tables IV and VI suggest that Maximum Haar-RWA model at the highest decomposition level usually gives rise to the lowest bit-rates at a moderate computational complexity. This is the case for uncalibrated and calibrated AVIRIS, for uncalibrated and calibrated Hyperion (5 levels is slightly better than 8 levels for the calibrated corpus), and for uncalibrated IASI (if side information is not taken into account). For calibrated IASI, IWT 9/7M-RWA always outperforms the other wavelet filters.

For uncalibrated Hyperion corpus, the Restricted model and its variant Fast-Restricted –with $\rho = 0.01$, based on Haar, and with the maximum number of decomposition levels–, yield the lowest bit-rate at an even lower computational complexity and providing excellent component-scalability. Even the variant Exogenous-Restricted generally performs better than Maximum. Different to other corpora, the correlation of uncalibrated Hyperion scenes is higher in local regions (Figure 5). In addition, Restricted, Fast-Restricted and Exogenous-Restricted benefit more from the 2D IWT 5/3 spatial transform than Maximum. The difference in bit-rate when applying or not this spatial transform is higher for Restricted than for Maximum model.

For IASI instrument, when the side information is taken into account, Exogenous variant provides the best coding performance, as the side information can weight up to 0.9 bpppc. In general, the Exogenous variant applied on any regression model and with any wavelet filter provides a very similar performance to that obtained by applying the optimal regression parameters computed on-line. The computational complexity is also low and no side information is required.

Overall, applying 5 wavelet decomposition levels yields very similar performance to the case of applying the largest number of decomposition levels, $J = \lceil \log_2(z) \rceil$, in terms of computational complexity and compression ratio. For instance, for 12-bit uncalibrated AVIRIS scenes, the entropy is reduced from 8.82 bits to 2.99 and 2.81 (Table V) for, respectively, 5 and 8 decomposition levels for Haar-RWA. For 16-bit uncalibrated Hyperion, the bit-rate is 4.54 and 4.53 when applying, respectively, 5 and 8 levels of Haar-RWA Maximum model (Table VI).

Beyond coding performance, Figure 5 displays the squared correlation of different scenes from four of the six scene corpora. The different nature of the data according to the instrument and the calibration process, reported in the first row, is clear. As expected, the correlation coefficient decreases significantly after the scenes have been spectrally transformed. RWA Maximum model and Fast variant are able to remove almost all the correlation, while Restricted model is not as efficient, due to the reduced number of components employed for the regression.

When analysing the influence of the calibration process, for AVIRIS instrument, the calibrated scenes yield lower bit-rates, while for Hyperion and IASI instruments, the uncalibrated scenes yield lower bit-rates. Similar observations are found in Amrani et al. (2016) and Blanes and Serra-Sagristà (2011).

C. Signal measurements results

Tables V and VII present the squared Pearson's correlation coefficient, the energy percentage and the Shannon entropy. Again, only average values are presented, as a similar pattern is observed for all scenes of a given corpus.

The tables show that these measurements significantly decrease when spectrally transforming the scenes at a higher number of wavelet decomposition levels. Indeed, the energy percentage may dramatically decrease at one half or even one quarter when applying only one transform level for, respectively, AVIRIS and Hyperion or IASI instruments. It even reaches a value lower than 0.4% for the uncalibrated Hyperion scenes when decorrelating them until only one approximation component remains. This value is even smaller, 0.006%, when transforming the uncalibrated IASI corpus for the Maximum model at 14 Haar-DWT decomposition levels.

For all the scene corpora, the average entropy approximately decreases by 3 to 6 bpppc when applying RWA at the highest decomposition level with respect to the original scene.

The squared Pearson's correlation coefficient measurement dramatically decreases at the order of four decimal digits for AVIRIS and calibrated Hyperion and to 0.001 for uncalibrated Hyperion scenes. For the uncalibrated and calibrated IASI corpora, it even reaches such negligible values as, respectively, 7.43×10^{-6} and 4.45×10^{-5}.

Notice that for these signal statistics, the three wavelet filters yield similar behaviour too. On the other side, although these measurements for Fast-Restricted are worse than those achieved by Maximum model, they tend to decrease in a similar manner. For Fast-Restricted, IWT 9/7M generally produces the best results because $\widehat{\boldsymbol{W}}_i^j$ is no longer being estimated from a polynomial function considering only a single approximation component of the same level $\boldsymbol{V}_i^j$, as happens for Haar-DWT, but from a linear combination of several components $\left\{\boldsymbol{V}_u^j\right\}_{u=i-3}^{i+4}$, yielding more accurate estimations. The modest complexity of Haar is here replaced by IWT 9/7M for better compression outcomes.

Figure 6 plots the squared correlation of the uncalibrated AVIRIS Yellowstone03 scene for different Restricted RWA configurations: original Haar-RWA (Amrani et al., 2016), its novel variants Fast and Exogenous, with different wavelet filters as their first operation block, and for original Parsimonious model (Amrani, Serra-Sagristà, and Marcellin, 2017), and its novel variants Fast and Exogenous, also with different wavelet filters.

The original Parsimonious model is based on a Haar-DWT and a regression through the Restrictred model, but it employs three approximation components for computing the regression

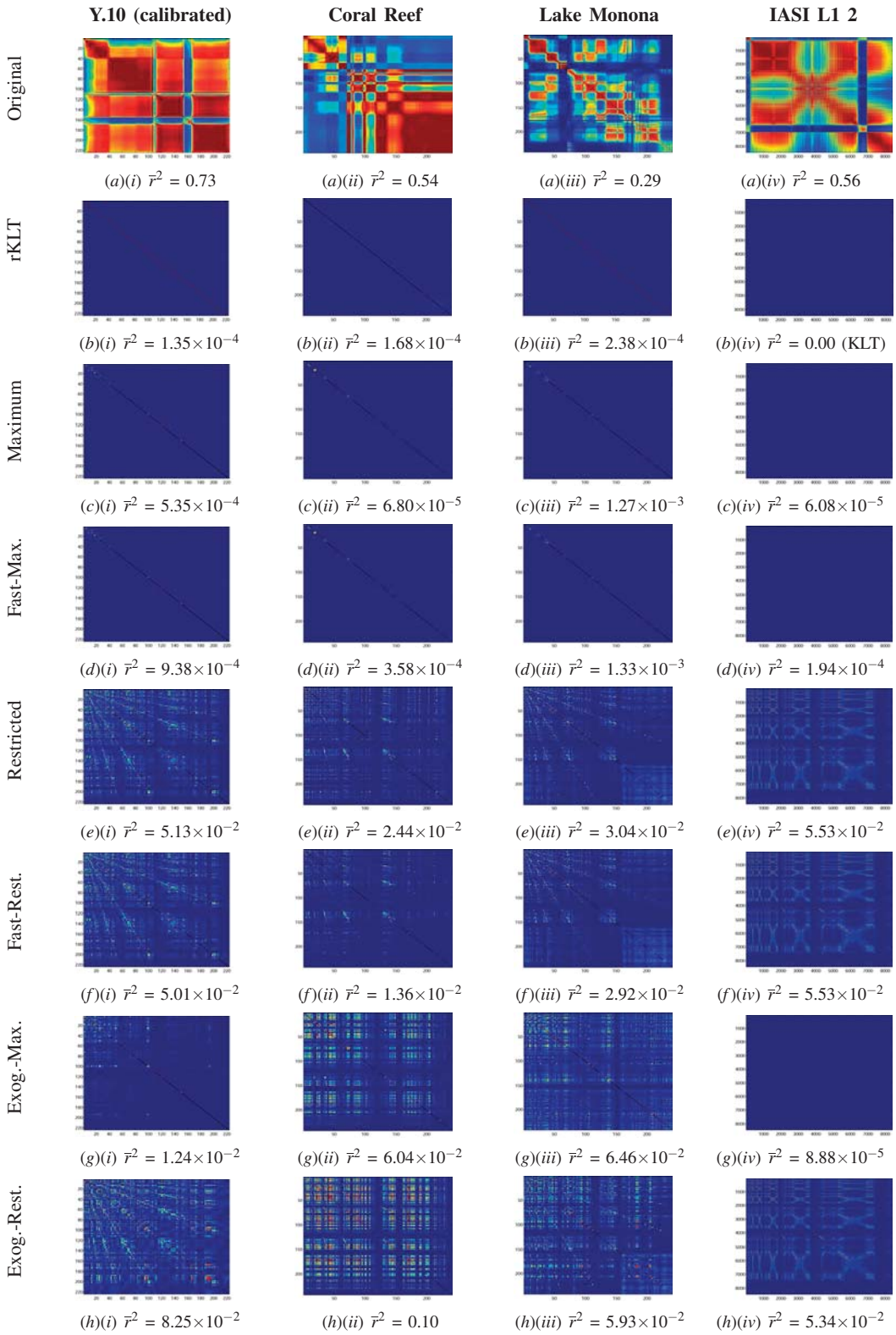
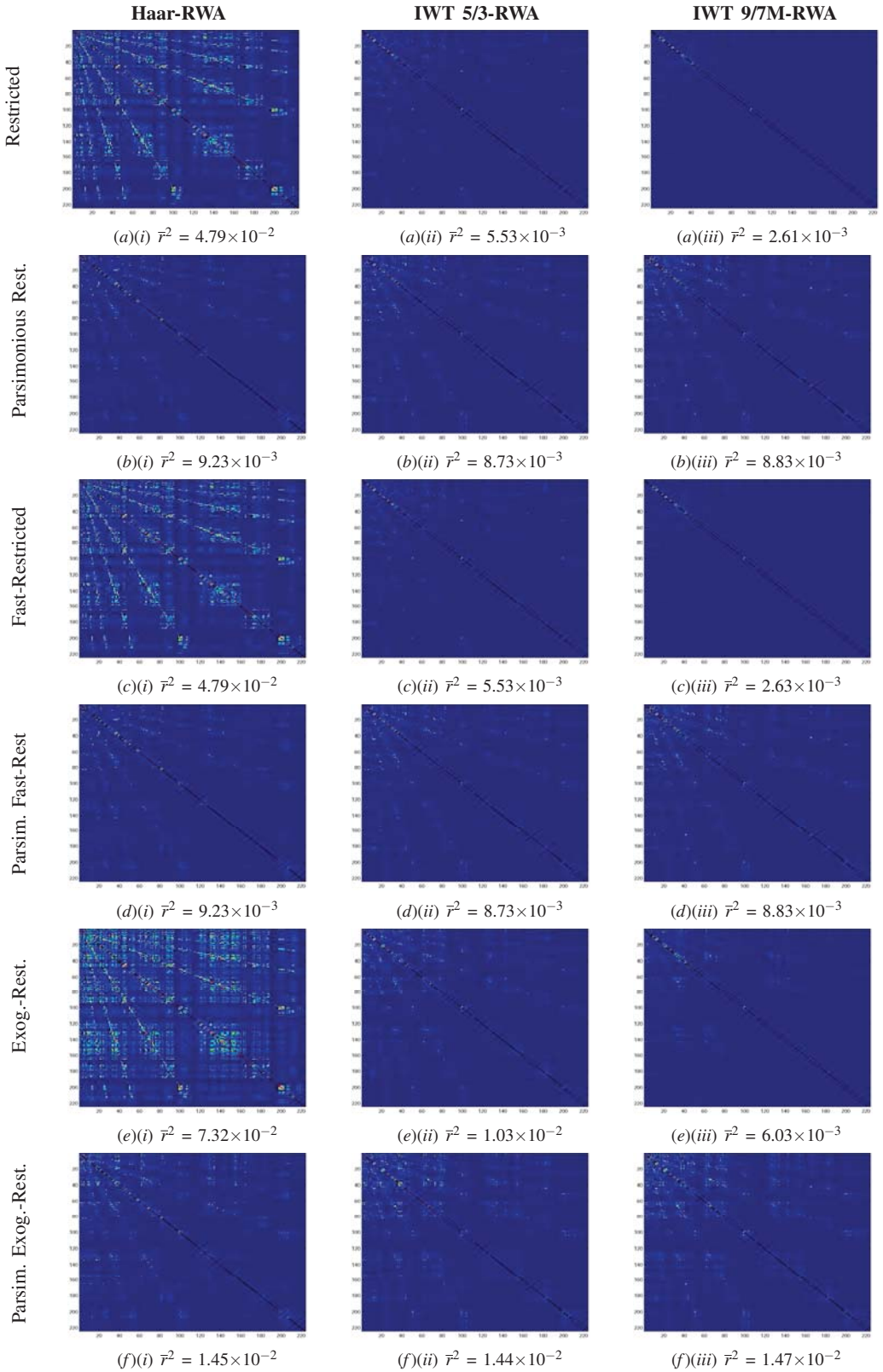


Figure 5 Squared correlation coefficient matrices when applying rKLT and Haar-RWA at the highest decomposition level. Results reported for four scenes of three different instruments. For the Fast variants, spatial sub-sampling parameter ρ is 0.01 for AVIRIS and Hyperion scenes and 0.1 for IASI corpora. The average correlation is denoted with $\bar{r}^2$.



(*a*)(*i*) $\bar{r}^2 = 4.79\times10^{-2}$ (*a*)(*ii*) $\bar{r}^2 = 5.53\times10^{-3}$ (*a*)(*iii*) $\bar{r}^2 = 2.61\times10^{-3}$

(*b*)(*i*) $\bar{r}^2 = 9.23\times10^{-3}$ (*b*)(*ii*) $\bar{r}^2 = 8.73\times10^{-3}$ (*b*)(*iii*) $\bar{r}^2 = 8.83\times10^{-3}$

(*c*)(*i*) $\bar{r}^2 = 4.79\times10^{-2}$ (*c*)(*ii*) $\bar{r}^2 = 5.53\times10^{-3}$ (*c*)(*iii*) $\bar{r}^2 = 2.63\times10^{-3}$

(*d*)(*i*) $\bar{r}^2 = 9.23\times10^{-3}$ (*d*)(*ii*) $\bar{r}^2 = 8.73\times10^{-3}$ (*d*)(*iii*) $\bar{r}^2 = 8.83\times10^{-3}$

(*e*)(*i*) $\bar{r}^2 = 7.32\times10^{-2}$ (*e*)(*ii*) $\bar{r}^2 = 1.03\times10^{-2}$ (*e*)(*iii*) $\bar{r}^2 = 6.03\times10^{-3}$

(*f*)(*i*) $\bar{r}^2 = 1.45\times10^{-2}$ (*f*)(*ii*) $\bar{r}^2 = 1.44\times10^{-2}$ (*f*)(*iii*) $\bar{r}^2 = 1.47\times10^{-2}$

Figure 6 Squared correlation coefficient matrices when applying the Restricted and Parsimonious Restricted models and variants (Fast and Exogenous) with different wavelet transforms as underlying filters at the highest decomposition level. Results are reported for the uncalibrated AVIRIS Yellowstone03 scene. Note that for the Parsimonious model, each prediction is conducted taking into account 3 approximation components, and that, for the Fast variants, spatial sub-sampling parameter ρ is 0.01. The average correlation is denoted with $\bar{r}^2$.

Table VII Signal measurements when applying RWA Maximum and Fast-Restricted to the different scene corpora.

Wavelet decomposition levels	Spectral wavelet filter	Maximum			Fast-Restricted		
		Mean r^2	Mean $\mathscr{E}(\%)$	Mean $\mathscr{H}$	Mean r^2	Mean $\mathscr{E}(\%)$	Mean $\mathscr{H}$
			AVIRIS (calibrated)				
0	—	0.69	100	9.77	0.69	100	9.77
1	Haar	0.17	47.28	7.40	0.19	49.87	7.77
	IWT 5/3	0.16	47.75	7.31	0.17	50.25	7.39
	IWT 9/7M	0.17	47.63	7.33	0.17	50.13	7.38
8	Haar	7.23×10^{-4}	0.21	3.87	0.07	0.33	5.00
	IWT 5/3	7.45×10^{-4}	0.58	3.91	7.34×10^{-3}	0.60	4.17
	IWT 9/7M	7.41×10^{-4}	0.54	3.91	3.01×10^{-3}	0.55	4.03
			Hyperion (uncalibrated)				
0	—	0.27	100	9.57	0.27	100	9.57
1	Haar	0.07	49.89	8.52	0.09	50.10	8.83
	IWT 5/3	0.06	50.01	8.46	0.08	50.03	8.71
	IWT 9/7M	0.07	49.89	8.48	0.08	49.90	8.68
8	Haar	1.20×10^{-3}	0.39	5.52	0.03	0.91	6.28
	IWT 5/3	1.34×10^{-3}	0.37	5.55	0.01	0.40	5.89
	IWT 9/7M	1.38×10^{-3}	0.37	5.55	8.22×10^{-3}	0.39	5.80
			Hyperion (calibrated)				
0	—	0.54	100	9.50	0.54	100	9.50
1	Haar	0.13	44.75	8.12	0.13	49.51	8.27
	IWT 5/3	0.13	45.43	8.18	0.13	50.31	8.27
	IWT 9/7M	0.13	45.12	8.28	0.13	49.97	8.31
8	Haar	1.01×10^{-4}	0.37	6.28	0.01	0.84	6.71
	IWT 5/3	4.01×10^{-4}	0.56	6.29	3.90×10^{-3}	0.81	6.65
	IWT 9/7M	8.01×10^{-4}	0.52	6.30	2.33×10^{-3}	0.59	6.49
			IASI L0 (uncalibrated)				
0	—	0.57	100	8.09	0.57	100	8.09
1	Haar	0.14	26.27	6.37	0.16	49.74	6.83
	IWT 5/3	0.14	26.49	6.29	0.14	50.15	6.50
	IWT 9/7M	0.14	26.40	6.32	0.14	49.89	6.46
14	Haar	7.43×10^{-6}	6.08×10^{-3}	2.43	0.03	0.10	3.69
	IWT 5/3	1.55×10^{-5}	0.01	2.44	5.05×10^{-3}	0.02	2.97
	IWT 9/7M	1.53×10^{-5}	0.01	2.47	1.99×10^{-3}	0.02	2.81
			IASI L1 (calibrated)				
0	—	0.57	100	12.83	0.57	100	12.83
1	Haar	0.14	37.07	10.58	0.16	49.96	11.18
	IWT 5/3	0.14	37.16	10.40	0.14	50.07	10.55
	IWT 9/7M	0.14	37.12	10.42	0.14	50.02	10.51
14	Haar	5.97×10^{-5}	0.02	6.77	0.05	0.08	8.12
	IWT 5/3	4.66×10^{-5}	0.02	6.75	0.04	0.03	7.17
	IWT 9/7M	4.45×10^{-5}	0.02	6.74	1.00×10^{-3}	0.03	6.99

parameters in each detail estimation instead of only one component (for which second and third-order terms were also used), thus providing a more accurate estimation at a lower computational complexity.

D. Dynamic range, precision and entropy

Figure 7 provides the dynamic range, precision and entropy per spectral component for uncalibrated AVIRIS Yellowstone03 scene. Measurements are reported for the original scene and for the Haar-DWT and RWA transforms when applied along the spectral dimension (left column). The plots display the higher performance of RWA over traditional wavelet transforms and the equivalent performance of RWA approach regardless of the wavelet filter used in the first stage of the transform (right column). RWA does not expand the dynamic range, requires less bits of precision and yields a lower entropy (as also reported in Tables V and VII). Behaviour for all the other employed hyperspectral scenes is similar.

E. Coding performance comparison against state-of-the-art techniques

For comparison purposes, bit-rates and execution time of several state-of-the-art coding techniques are reported (Table VIII). Following a chronological order, a brief description of each of these techniques is depicted next:

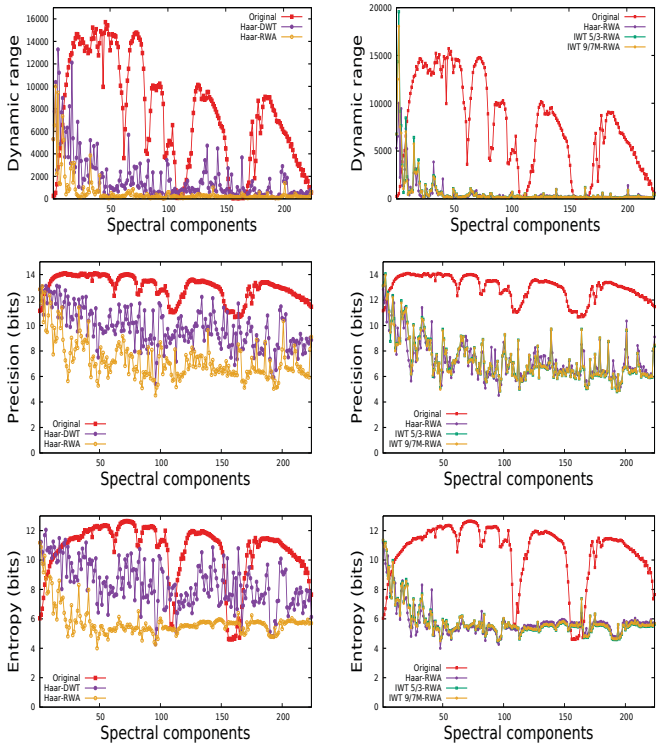


Figure 7 From top to bottom: dynamic range, precision and entropy of the uncalibrated AVIRIS Yellowstone03 scene and of its transformed versions when applying different wavelet filters as the basis of Maximum RWA and applying the Haar-DWT.

(i) Reversible Karhunen-Loève Transform (rKLT) diagonalizes the covariance matrix. It formulates a stochastic process as an infinite linear combination of orthogonal functions.

(ii) Haar-DWT is one of the most elementary DWT. It pyramidally decorrelates the spectral dimension of hyperspectral scenes by applying a simple Haar wavelet filter.

(iii) Multiband Context-based Adaptive Lossless/Near-Lossless Image Coding (M-CALIC) (Magli, Olmo, and Quacchio , 2004) applies an efficient spectral decorrelation. It is a predictive-based technique that allows for near-lossless and lossless reconstruction.

(iv) Pairwise Orthogonal Transform (POT) (Blanes and Serra-Sagristà , 2011) is a transform-based approach inspired on the KLT but aiming at competitive coding performance without the disadvantages the KLT introduces: high computational cost, lack of component-scalability, large memory requirements, and difficult implementation. POT solves these problems by generating multiple compositions of two-component instances of the KLT.

(v) The standar CCSDS-123.0 (CCSDS, 2012) tackles onboard lossless coding of multi-spectral and hyperspectral scenes. It is based on a causal context-based prediction and an entropy-coding operations. This entropy-coder utilizes Golomb power-of-two codes.

(vi) Clustered Differential Pulse Code Modulation with Adaptive Prediction Length (C-DPCM-APL) (Mielikainen and Huang , 2012) exploits the use of an adaptive prediction length method in clustered differential pulse code modulation (C-DPCM) for lossless compression. The differences between the predicted and the original values are sent to the entropy encoder that uses an adaptive range coder for each cluster.

(vii) Lossless Recursive Least-Squares filtering (RLS) (Song, Zang, and Chen , 2013) calculates the subtraction of the average value of 4 neighbour pixels and the current pixel value. The local differences between the co-locate pixels with the current pixel in previous components conform the input signal to the RLS filter. The RLS filter predicts the value of the current local difference and encodes the prediction residual by means of an adaptive arithmetic encoder.

(viii) Clustered Differential Pulse Code Modulation with Removal of Local Spectral Outliers (C-DPCM-RLSO) (Wu et al. , 2015) removes the local spectral outliers by enhancing the prediction accuracy of the high-order regression predictor and decreasing the residual values between the prediction and the original values.

(ix) Lossless Conventional Recursive Least-Squares predictor (CRLS) (Gao and Guo , 2016) is a spectral domain predictor that quantifies the optimal number of prediction components twice. CRLS uses 24 neighbour pixels instead of 4. The technique is adaptive in that it looks for the best prediction components (A-CRLS). This adaptation is based on an exhaustive search of the number of components that minimizes the prediction residual.

(x) RWA (Amrani et al., 2016) introduces Maximum and Restricted RWA models with Haar-DWT as the underlying wavelet transform.

(xi) Parsimonious RWA (Amrani, Serra-Sagristà, and Marcellin, 2017) simplifies Maximum RWA model by selecting a finite subset of approximation components when computing the regression parameters (Restricted model). It selects a subset of neighbour approximation components to predict a detail component. The original version employs Haar-DWT as the first operation block; here we provide results also for Parsimonious IWT 5/3-RWA and Parsimonious IWT 9/7M-RWA. In addition, novel Fast-Parsimonious and Exogenous-Parsimonious lossless variants have been investigated.

POT is a line-based transform that generates side information for each input scene row. Since both Hyperion and IASI instruments capture tall and narrow scenes, the side information is considerably reduced when rotating the scenes by 90°. To provide the best performance possible, the uncalibrated Hyperion corpus has been rotated 90° for all the techniques but M-CALIC. The calibrated Hyperion corpus is rotated for all the techniques but M-CALIC, CCSDS-123.0 and RWA. For POT, IASI corpus has been also rotated.

The C++ software implementation of M-CALIC does not handle transformed signals with more than 15 bpppc. Therefore, outcomes for uncalibrated AVIRIS Yellowstone11 and 18, for the calibrated Hyperion corpora and for the uncali-

Table VIII Average bit-rates (in bpppc) and execution time (in seconds) for different lossless compression techniques. For all RWA variants (shown in italics), the underlying wavelet filter is Haar-DWT at the highest decomposition level. For all cases where only the spectral transform is indicated, JPEG 2000 has been applied as the coding technique. Execution time includes the whole coding process. Techniques are ordered from faster to slower for the AVIRIS instrument.

	Average bit-rates							Average execution time (s)						
	AVIRIS			Hyperion		IASI		AVIRIS			Hyperion		IASI	
	16-bit calibrated scenes	16-bit uncalibrated scenes	12-bit uncalibrated scenes	16-bit calibrated scenes	16-bit uncalibrated scenes	16-bit calibrated scenes	16-bit uncalibrated scenes	16-bit calibrated scenes	16-bit uncalibrated scenes	12-bit uncalibrated scenes	16-bit calibrated scenes	16-bit uncalibrated scenes	16-bit calibrated scenes	16-bit uncalibrated scenes
Haar-DWT	4.76	7.12	3.44	6.01	4.49	7.37	3.00	44.09	53.29	35.44	118.10	101.91	604.68	433.94
Exog.-Maximum	3.53	5.84	2.82	5.38	4.76	6.46	2.49	45.06	53.06	35.57	130.34	124.06	1524.87	1409.92
Parsimonious Exog.-Restricted	3.77	5.98	2.90	5.49	4.52	6.87	2.91	46.68	54.31	37.23	129.95	120.15	699.16	556.10
POT	4.19	6.36	3.03	5.78	4.31	7.12	2.87	47.14	56.88	39.62	121.40	111.51	630.48	443.78
Parsimonious Fast-Restricted	3.81	6.02	2.88	5.51	4.42	6.88	2.93	48.53	56.30	43.04	134.66	121.63	710.51	566.99
Parsimonious Restricted	3.81	6.02	2.88	5.51	4.42	6.88	2.92	58.83	66.66	52.91	163.47	153.44	793.66	647.32
Maximum	3.52	5.83	2.61	5.33	4.53	7.34	3.31	69.36	78.19	63.75	198.04	193.75	4413.73	4185.43
RLS	3.58	5.79	2.60	—	—	—	—	$\approx$ 96.00	$\approx$ 97.00	$\approx$ 93.50	—	—	—	—
CCSDS-123.0	3.70	6.15	2.73	5.65	4.29	6.60	2.89	163.77	163.93	154.47	489.60	351.87	1465.27	1465.24
rKLT	3.69	5.81	3.21	5.76	4.55	9.53	—	193.70	204.79	182.59	476.03	487.48	—	—
M-CALIC	3.86	6.06	2.86	—	4.83	6.87	2.88	258.90	272.79	242.59	—	660.24	5111.70	4748.20
CRLS	3.50	5.71	2.51	—	—	—	—	1366.60	1369.20	1047.50	—	—	—	—
C-DPCM-RLSO (order 85)	3.27	5.57	2.44	—	—	—	—	3518.00	3365.60	2714.50	—	—	—	—
A-CRLS	3.29	5.57	2.44	—	—	—	—	5819.00	5673.00	4526.50	—	—	—	—
C-DPCM-APL	3.29	5.56	2.43	—	—	—	—	32961.00	31901.00	26835.00	—	—	—	—

brated IASI L0 1 and 2 scenes are not available.

For the assessment of average execution time, rKLT, Haar-DWT, POT, M-CALIC, CCDSDS-123.0, RWA and variants, and Parsimonious RWA and variants are computed on an Intel(R) Xeon(R) CPU E5520 @ 2.27GHz processor with a single thread.

The average execution time for AVIRIS corpus for recursive least-squares (RLS) filtering, C-DPCM-APL, C-DPCM-RLSO, CLRS and A-CLRS are taken from the original papers Wu et al. (2015); Gao and Guo (2016) and Song, Zang, and Chen (2013). Note that for A-CRLS, the uncalibrated and calibrated AVIRIS Yellowstone00 scenes have not been included when computing the average execution time, because they are used to fix the optimal prediction components for coding the rest of the scenes in the corpus.

1) AVIRIS scenes: A-CRLS, C-DPCM-APL and C-DPCM-RLSO (order 85) give rise to the highest coding performance, however, they demand significant computational resources and their execution time is too large, requiring, respectively, 5819, 32961 and 3518 seconds for compressing a 16-bit calibrated AVIRIS scene. Haar-DWT and POT require a short execution time at the expense of worsening the coding performance. Maximum RWA outperforms the other techniques for calibrated AVIRIS and for 12-bit uncalibrated scenes. Its performance lies very close to rKLT for uncalibrated AVIRIS scenes with a difference of 0.02 bpppc. However, rKLT demands higher computational complexity. RLS and CRLS obtain similar bit-rates than RWA but are more expensive. Maximum and Exogenous-Maximum Haar-RWA yield a highly competitive coding gain at a very low execution time. The Exogenous variant secures bit-rates only 0.26, 0.27 and 0.38 higher than those provided by C-DPCM-RLSO (order 85), the best performing coding technique, at considerably lower execution time, namely, respectively, 78.07, 63.43 and 76.31 times faster. Parsimonious RWA, M-CALIC and CCSDS-123.0 achieve slightly higher bit-rates at a reasonable computational time.

2) Hyperion and IASI scenes: For calibrated Hyperion scenes, Maximum RWA outperforms all the state-of-the-art techniques at a very competitive execution time. For uncalibrated Hyperion, CCSDS-123.0 provides superior performance, close to that achieved by POT and Haar-RWA Fast-Restricted (Table VI), but demands larger computational complexity.

The RWA Exogenous-Maximum variant encodes the uncalibrated and calibrated IASI corpora with the lowest bit-rates at considerably lower execution time than the rest of the techniques, becoming the best performing technique for scenes with thousands of spectral components. For these scenes, side information plays a relevant role in the coding performance, and Exogenous-Maximum benefits from not having to store/-transmit it.

VI. Conclusions

Regression Wavelet Analysis (RWA) is a state-of-the-art hyperspectral scenes lossless coding method which enables different regression models. Maximum model yields the finest prediction accuracy and competitive coding performance, with a computational cost lower than the cost of reversible KLT, which uses to provide the best performing coding results. The Restricted model largely reduces Maximum model computational complexity with an almost equivalent coding performance.

To further reduce the computational cost of RWA, we present two approaches than can be coupled with the Restricted regression model. Fast-Restricted variant reduces the complexity by applying a spatial sub-sampling (using only 1% of the pixels for learning for AVIRIS and Hyperion scenes and 10% for IASI data). Exogenous-Restricted variant fully avoids the regression computation by employing fixed parameters – learned over a training scene– for the other scenes of the same instrument. It also reduces the Exogenous-Maximum complexity by employing less approximation components in the generation of the predicted details. This last variant may even imply a better coding performance than Maximum for scenes with large spectral dimension, for which the side information related to the regression parameters is not negligible.

Three different choices of reversible wavelet transform have been investigated: Haar-DWT, IWT 5/3 and 9/7M. Experimental results suggest that the coding performance of Maximum and Parsimonious RWA for lossless compression is barely affected by the underlying wavelet transform, indicating that regression is able to predict the details from the approximation components regardless of the wavelet filter.

However, when analysing the Restricted model and its variants, the wavelet filter used as first RWA operation block meaningfully affects the coding performance. IWT-RWA may achieve higher compression ratios than those presented in Amrani et al. (2016) at lower execution times. For instance, IWT 5/3-RWA and IWT 9/7M-RWA outperform Haar-RWA for uncalibrated and calibrated AVIRIS and calibrated IASI instruments, at a very similar computational complexity. IWT-RWA may obtain lower bit-rates than the Parsimonious model (Amrani, Serra-Sagristà, and Marcellin, 2017) at similar complexity too. This is because more approximation components are considered in each detail prediction. Moreover, the estimation formulation changes from a cubic equation to a polynomial of first order. The same happens for their variants Fast and Exogenous.

To summarize, RWA allows a wide range of variations, either in the wavelet transform or in the regression model, that yield an accurate prediction with a reduced computational complexity and bearable execution time. This technique provides superior coding performance than other recent widespread techniques such as, from faster to slower, POT, RLS, CCSDS-123.0, rKLT and M-CALIC for lossless compression of hyperspectral scenes. RWA also aims at very competitive compression ratios at considerably lower execution time than other state-of-the-art techniques such as CRLS, A-CRLS, C-DPCM-RLSO (order 85) and C-DPCM-APL. RWA is therefore among the most competitive approaches.

VII. Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments.

An earlier version of this work was presented at the ESA CNES On-Board Payload Data Compression Workshop 2016.

VIII. Funding

This work was supported in part by the Spanish Ministry of Economy and Competitiveness (MINECO); European Regional Development Fund (FEDER);[TIN2015-71126-R]; Catalan Government; [2014SGR-691]; Universitat Autònoma de Barcelona; [UAB-PIF-472/2012 and UAB-PIF-472/2015].

References

Amrani, N. and J. Joan Serra-Sagristà. 2015. "RWA coding toolbox." http://gici.uab.cat/GiciWebPage/downloads.php#RWA.

Amrani, N., J. Serra-Sagristà, V. Laparra, M. Marcellin, and J. Malo. 2016. "Regression wavelet analysis for lossless coding of remote-sensing data." *IEEE Transactions on Geoscience and Remote Sensing.* 54 (9): 5616-5627. ISSN: 0196-2892. doi: 10.1109/TGRS.2016.2569485.

Amrani, N., J. Serra-Sagristà,and M. Marcellin. 2017. "Low Complexity Prediction Model for Coding Remote-Sensing Data with Regression Wavelet Analysis." *Data Compression Conference DCC. IEEE.* 112-121. ISSN: 2375-0359. doi: 10.1109/DCC.2017.61.

Blanes, I. and J. Serra-Sagristà. 2011. "Pairwise Orthogonal Transform for Spectral Image Coding." *IEEE Transactions on Geoscience and Remote Sensing.* 49 (3): 961–972. ISSN: 0196-2892. doi: 10.1109/TGRS.2010.2071880.

Calderbank, A.R., I. Daubechies, W. Sweldens and Y. Boon-Lock. 1997. "Lossless image compression using integer to integer wavelet transforms." *Image Processing, 1997. Proceedings., International Conference on.* 1: 596–599. doi: 10.1109/ICIP.1997.647983.

Camps-Valls, G., D. Tuia, L. Gomez-Chova, S. Jimenez and J. Malo. 2011. "Remote Sensing Image Processing." *Synthesis Lectures on Image, Video and Multimedia Proc. Morgan & Claypool Publ. Austin, TX, USA*. doi: 10.2200/S00392ED1V01Y201107IVM012.

Chen, M. 2010. "Normalized Mutual Information." *The Chinese University of Hong Kong. https://es.mathworks.com/matlabcentral/fileexchange/29047-normalized-mutual-information/content/nmi/nmi.m.*

Consultative Committee for Space Data Systems (CCSDS). 2011. "Image Data Compression CCSDS 122.0-B-1." *CCSDS. Blue Book. Accessed 3 September*. https://public.ccsds.org/Pubs/122x0b1c3.pdf.

Consultative Committee for Space Data Systems (CCSDS). 2012. "Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1." *CCSDS. Blue Book*. https://public.ccsds.org/Pubs/123x0b1ec1.pdf.

Effros, M., H. Feng and K. Zeger. 2004. "Suboptimality of the Karhunen-Loeve Transform for transform coding." *IEEE Transactions on Information Theory.* 50 (8): 1605–1619. doi: 10.1109/TIT.2004.831787.

Fowler, J.E. and J.T. Rucker. 2007. "Hyperspectral Data Exploitation: Theory and Applications. 3D wavelet-Based Compression of Hyperspectral Imager." *John Wiley & Sons Inc. Hoboken, NJ, USA*. 379–407.

Gao, F. and S. Guo. 2016. "Lossless compression of hyperspectral images using conventional recursive least-squares predictor with adaptive prediction bands." *Journal of Applied Remote Sensing.* 10(1): 015010-015010. doi: 10.1117/1.JRS.10.015010.

Jolliffe, I. T. 2002. "Principal Component Analysis." *Springer Verlag, Berlin, Germany.* 487.

Magli, E., G. Olmo and E. Quacchio. 2004. "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC." *IEEE Geoscience and Remote Sensing Letters.* 1 (1): 21–25. ISSN: 1545-598X. doi: 10.1109/LGRS.2003.822312.

Mielikainen, J., and B. Huang. 2012. "Lossless compression of hyperspectral images using clustered linear prediction with adaptive prediction length." *IEEE geoscience and remote sensing letters.* 9(6): 1118–1121. doi: 10.1109/LGRS.2012.2191531.

Penna, B., T. Tillo, E. Magli, and G. Olmo. 2007. "Transform Coding Techniques for Lossy Hyperspectral Data Compression." *IEEE Transactions on Geoscience and Remote Sensing.* 45 (5): 1408–1421. doi: 10.1109/TGRS.2007.894565.

Song, J., Z. Zhang, and X. Chen. 2013. "Lossless compression of hyperspectral imagery via RLS filter." *Electronics Letters,* 49, no. 16 (2013): 992-994. ISSN 0013-5194. doi: 10.1049/el.2013.1315.

Tang, X. and W.A. Pearlman. 2006. "Three-dimensional wavelet-based compression of hyperspectral images." *Hyperspectral Data Compression. US. Springer.* 273–308.

Taubman, D., and M. Marcellin. 2001. "JPEG 2000: Image compression fundamentals, practices and standards" *New York. Springer*.

Wu, J., W. Kong, J. Mielikainen, and B. Huang. 2015. "Lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers." *IEEE Signal Processing Letters,* 22(2): 2194-2198. doi: 10.1109/LSP.2015.2443913.

Zhang, J., J.E. Fowler and G.Liu. 2008. "Lossy-to-lossless compression of hyperspectral imagery using three-dimensional TCE and an integer KLT." *IEEE Geoscience and Remote Sensing Letters* 5: 814–818. doi: 10.1109/LGRS.2008.2006571.

Chapter 3

Progressive lossy-to-lossless coding of hyperspectral images through regression wavelet analysis

Progressive Lossy-to-Lossless Coding of Hyperspectral Images through Regression Wavelet Analysis

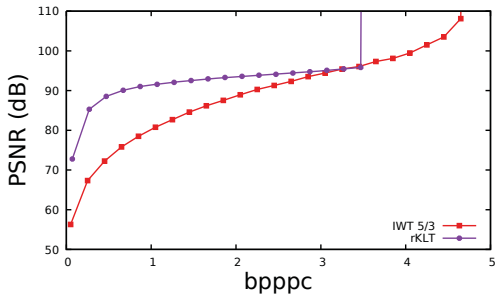
Sara Álvarez-Cortés[+], Naoufal Amrani[+], Miguel Hernández-Cabronero[*] and Joan Serra-Sagristà[+]

[+] Universitat Autònoma de Barcelona

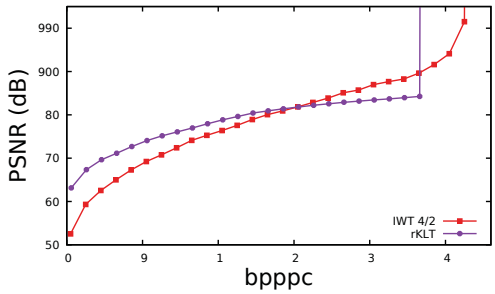
[*] University of Arizona

***Abstract*—Progressive Lossy-to-Lossless (PLL) coding techniques enable a gradual quality improvement of the recovered images, starting from a coarse approximation up to a perfect reconstruction. PLL is becoming a widespread approach in several scenarios, in particular, for compression of hyperspectral images. In this paper we assess the suitability of Regression Wavelet Analysis (RWA) for hyperspectral image progressive lossy-to-lossless coding. RWA is a recent spectral transform that combines a wavelet transform with a regression stage, providing excellent coding performance for lossless compression. When coupled with a pyramidal predictive weighting scheme, RWA also yields very competitive coding results for PLL at a low computational cost. Coding performance is assessed within the framework of JPEG 2000 standard, comparing RWA against state-of-the-art spectral transforms, including reversible Karhunen-Loève Transform (rKLT) and Pairwise Orthogonal Transform (POT). Comparison with respect to Multiband Context-based Adaptive Lossless/Near-Lossless Image Coding (M-CALIC) technique is also provided. Experiments are conducted on uncalibrated and calibrated hyperspectral images from Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), satellite-borne Hyperion and Infrared Atmospheric Sounding Interferometer (IASI) sensors. Discussion embraces rate-distortion performance, bit-per-pixel-per-component rate distribution and classification outcome.**

***Index Terms*—Progressive lossy-to-lossless compression, regression wavelet analysis, predictive weighting scheme, hyperspectral image compression, image quality recovery.**



(*a*) Calibrated AVIRIS-Yellowstone03



(*b*) Uncalibrated Hyperion-Lake Monona

Figure 1 IWT 5/3 and reversible KLT Rate-Distortion curves.

I. Introduction

Nowadays sensors record enormous volumes of information and their data-capture capacity is only further increasing. Handling this large amount of information poses several challenging problems, for instance, how to transmit the information from the satellite to the ground stations on Earth through a downlink channel with restricted bandwidth. This limitation may give rise to an insufficient on-board storage capacity and lead to discarding some of the just acquired remote-sensing data. Compression techniques can mitigate this problem and enable efficient on-board and on-land data storage and transmission.

Several lossless, near-lossless and lossy coding techniques have been recently introduced. Lossless compression techniques allow for perfect data recovery, usually providing compression ratios close to 2:1. Near-lossless compression techniques restrict the –peak-absolute– distortion and enable larger compression ratios. Lossy compression techniques allow an arbitrary compression ratio at the cost of an increased distortion in the recovered image. Progressive Lossy-to-Lossless (PLL) is a particular type of lossy compression, where the recovered quality is gradually refined as the code-stream is decoded. It is lately attracting a large interest, specially for hyperspectral image coding. Reviews of PLL coding for remote-sensing can be found in Ginesu, Giusto, and Pearlman (2004) and Zhang, Fowler, and Liu (2008).

Remote-sensing data has large spectral redundancy and removing the statistical relationships along the spectral dimension is required for achieving high compression ratios. Significant differences in coding performance are achieved for lossless, near-lossless and lossy compression when the spectral redundancy is exploited (Blanes, Magli, and Serra-

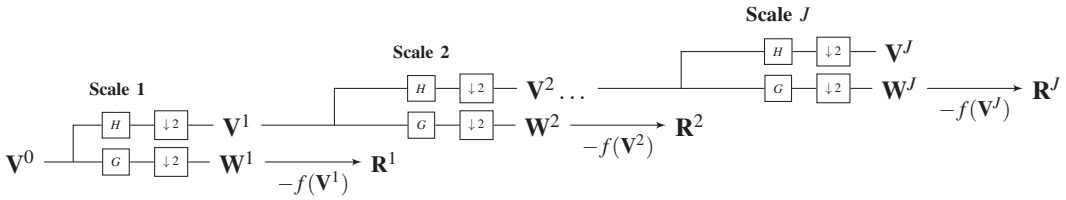


Figure 2 RWA representation with J levels.

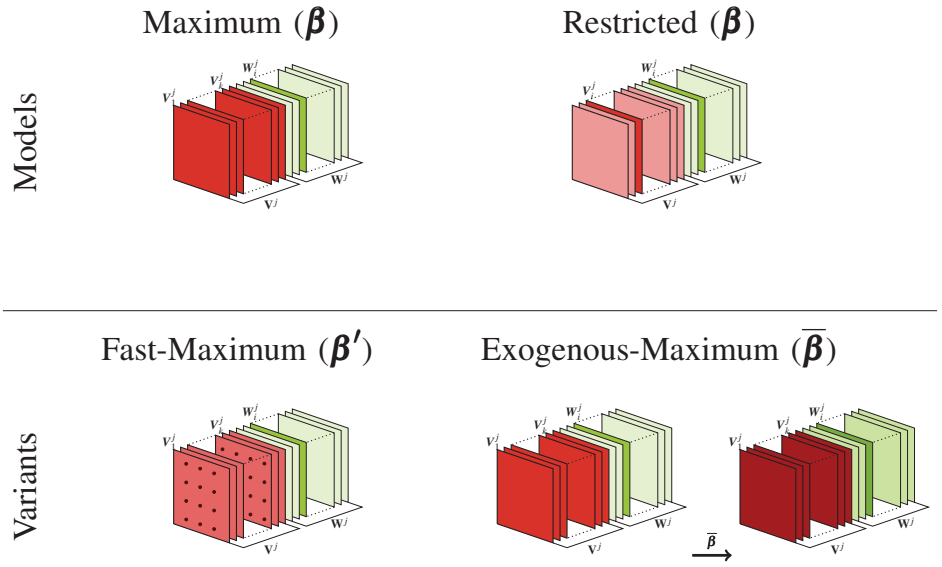


Figure 3 Lossless RWA regression models and variants schemes.

Sagristà , 2014). Different compression techniques based on spectral-components decorrelation have thus been designed. Karhunen-Loève Transform (KLT) is the optimal decorrelating transform for Gaussian sources (Jolliffe , 2002; Effros, Feng, and Zeger , 2004). Although neither 1D-spectral nor 2D-spatial hyperspectral signals are Gaussian (Camps-Valls et al., 2011), KLT usually yields also very efficient coding performance when used as a spectral transform. However, it is a linear data-dependent transform and requires large computational resources, such so that it even becomes unfeasible for images with a large number of spectral components (such as images recorded by IASI sensor, which have more than 8,000 spectral bands). Several approaches for reducing KLT's computational complexity have been proposed, which also try to minimize the coding performance decline. Some of them are based on sub-sampling the dataset in the spectral and/or the spatial dimension, thus computing the covariance matrix on a reduced subset of coefficients and as a result decreasing the computational complexity (Penna et al. , 2007). Other approaches reduce KLT complexity by learning the transform coefficients on a single image from a particular sensor and applying later the same learned coefficients (an exogenous transform) to other images from the same sensor (Barret, Gutzwiller, and Harati, 2011).

Figure 1 displays the coding performance of applying a reversible KLT (rKLT) and an Integer Wavelet Transform (IWT) 5/3 on a calibrated and on an uncalibrated image. The coding performance of rKLT is clearly above that of IWT 5/3 in all bit-rates (except for a short range for Figure 1(b), where the lifting scheme introduced in rKLT penalizes its efficiency). The goal of several subsequent coding approaches is to approximate the coding performance of rKLT at a reduced computational cost.

Regression Wavelet Analysis (Amrani et al., 2016a) is one such approach. It has been recently proposed for lossless coding of hyperspectral images. It achieves lossless coding performance superior to other widespread methods such as CCSDS-122.1 (CCSDS , 2012a), based on a spectral transform followed by a spatial Discrete Wavelet Transform (DWT), the predictive-based CCSDS-123.0 (CCSDS, 2012b), both developed by the Consultative Committee for Space Data Systems (CCSDS), and other KLT-based approximations. RWA applies an inexpensive Haar-DWT S-transform followed by a regression analysis to account for the 1D spectral correlation. The regression analysis removes almost all the remaining statistical relationships in the wavelet domain. RWA works

Table I Lossless RWA formulation.

Forward	Backward
$(\text{RWA})(\mathbf{V}^{j-1},1)=\left(\mathbf{V}^{j},\ \underbrace{\mathbf{R}^{j}}_{\mathbf{W}^{j}-\widehat{\mathbf{W}}^{j}}\right)$	$\mathbf{V}^{j-1}=(\text{DWT})^{-1}\left((\mathbf{V}^{j},\mathbf{R}^{j}+\widehat{\mathbf{W}}^{j}),1\right)$

as a predictive scheme preserving some of the DWT features as component scalability and low complexity.

RWA can be adapted to progressive lossy-to-lossless recovery (PLL-RWA) through a pyramidal Predictive Weighting Scheme (PWS) (Amrani et al. , 2016b). PWS takes into consideration the predictive contribution of each transformed spectral component into JPEG 2000 rate-allocation process. To this end, it progressively attributes higher weights to the most significant components in the decoding process, while securing an error propagation minimization. These weights improve the PLL-RWA recovered quality data. It yields similar rate-distortion performance to pure lossy techniques at low bit-rates, and the same coding performance as RWA in the lossless regime.

In this paper we present an in-depth discussion of the PLL-RWA coding process. While in Amrani et al. (2016b) a predictive weighting scheme is proposed for aiming at better rate-allocation performances and for avoiding the non-monotonic growth of the Maximum RWA Rate-Distortion curves, this article presents further experiments and demonstrates its feasibility in common nowadays applications. In our approach we cover the beneficial use of applying this PWS not only for the Maximum regression model but also for the Restricted model and the different variants: Fast-Maximum, Fast-Restricted, Exogenous-Maximum and Exogenous-Restricted. For comparison purposes we provide performance assessment against different widespread state-of-the-art spectral transforms such as reversible KLT (rKLT) and Pairwise Orthogonal Transform (POT) (Blanes and Serra-Sagristà , 2011) and the near-lossless and lossless technique M-CALIC (Magli, Olmo, and Quacchio , 2004). We also assess the effect of varying the wavelet transform (Haar-DWT, IWT 5/3 and IWT 9/7M) as the basis of the first operation in the progressive lossy-to-lossless RWA scheme. The benefits of the predictive weighting scheme are explained investigating the bit-rate distributions per transformed spectral component according to a global target bit-rate. PWS entails that most important transformed components are sent first, enabling a better image quality recovery for progressive lossy-to-lossless transmission. This improved reconstruction capability is illustrated for digital classification, suggesting that PWS-RWA yields more competitive performance than reversible KLT and Pairwise Orthogonal Transform, specially for low bit-rates.

II. Progressive Lossy-to-Lossless Coding

This section briefly surveys the basis of RWA and the predictive weighting scheme. Notation and rate allocation across the transformed components in the final code-stream are also introduced.

A. Regression Wavelet Analysis

RWA applies first a DWT and in sequence a regression operation. This second stage exploits the remaining spectral redundancy the DWT can not remove. This lossless coding method deploys a lifting scheme (Calderbank et al. , 1997), dealing with integer coefficients $\left(\mathbf{V}^{j},\mathbf{W}^{j}\right)$ thanks to a reversible rounding operation.

Figure 2 shows the RWA scheme, with H and G being, respectively, the low and high-pass filters. As can be noticed, the approximation components remain unchanged during the whole estimation process. The forward and backward formulation of the one level RWA decomposition of the approximation components $\mathbf{V}^{j-1}$ are given in Table I. When decoding, the inverse DWT ($(\text{DWT})^{-1}$) is applied once the detail components $\mathbf{W}^{j}$ are estimated and combined with the residuals variables, $\mathbf{R}^{j}$. In general, the detail components $\boldsymbol{W}_{i}^{j}\in\mathbb{R}^{m\times 1}$, $i\in \mathrm{I}=\{1,...,k{=}z{\cdot}2^{-j}\}$, are estimated from some or all of the approximation components at level j, $\mathbf{V}^{j}\in\mathbb{R}^{m\times(z\cdot 2^{-j})}$. These estimations $\widehat{\mathbf{W}}^{j}$ are computed through a least-squares method (Nocedal, and Wright, 2006) and are involved in the regression coefficients, $\boldsymbol{\beta}$, computation. It minimizes the sum of squares of the distances between the DWT decomposition details and the estimated ones, i.e.:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}}\|\mathbf{W}-\boldsymbol{\beta}\mathbf{V}\|_{2}. \tag{1}$$

The parameters $\boldsymbol{\beta}^{j}$ are the regression coefficients. For all the regression models and variants –but Exogenous– they require to be stored as side information for the reverse process.

Table II reports the computational cost, in Floating-point operations (FLOPs), of the different operations involved in a lossless RWA process.

The prediction of each detail component $\widehat{\boldsymbol{W}}_{i}^{j}$ at level j, $\widehat{\boldsymbol{W}}_{i}^{j}=f_{i}(\mathbf{V}^{j})$, is differently formulated depending on the used RWA regression model or variant. Function f_{i} defines the employed regression model as specified in Table III, both for computing the regression coefficients and for generating the estimation details in the RWA reverse procedure. The Maximum model uses all the approximation components from $\mathbf{V}^{j}$. It is the most accurate model and stores the regression coefficients as side information for the decoding procedure. The Restricted model employs a subset of approximation components. This subset is selected to maintain the component-scalability (number of transformed components required to reconstruct an image component) of the original DWT. The Fast variant randomly chooses a subset of $m'{=}\rho\cdot m$ spatial samples ($m'{\ll}m$), where ρ corresponds to the subsampling percentage. Only these samples are used for the least-squares method and, therefore, for the regression coefficients calculation. Finally, Exogenous variant uses fixed regression coefficients obtained through training a single image from a corpus to estimate the

Table II Computational complexity of the lossless RWA operations in Floating-point Operations (FLOPs). z and m are the number of spectral components and spatial samples per component. l corresponds to the wavelet decomposition level and k_i to the number of approximation components involved in each prediction level i.

Operation	FLOPs
Haar S-Transform (C_{HT})	$8(1-\frac{1}{2^l})mz$
Regression coefficients calculation (C_{RC}): $\boldsymbol{\beta}=(\mathbf{V}^\intercal\mathbf{V})^{-1}\mathbf{V}^\intercal\mathbf{W}$	$\sum_{i=1}^{l}(2m-1)(k_i+1)^2+(k_i+1)^3+(\frac{z}{2^i})(k_i+1)\,[(2m-1)+(2k_i+1)]$
Estimation generation (C_{EG}): $\widehat{\mathbf{W}}=\boldsymbol{\beta}\mathbf{V}$	$2\sum_{i=1}^{l}(2k_i-1)m\frac{z}{2^i}$
Apply and Remove (C_{AR})	$2m(z-1)$

Table III Details estimation and component-scalability preservation of the different Haar-based RWA regression models and variants. Please refer to Table II and to Figure 3 for notation. j is the number of applied wavelet decomposition levels.

	Regression type	Details estimation	Component-scalability
Models	Maximum	$\widehat{\boldsymbol{W}}_i^j=\beta_{i,0}^j+\beta_{i,1}^j\boldsymbol{V}_1^j+\cdots+\beta_{i,k}^j\boldsymbol{V}_k^j$	$z(1-\frac{1}{2^j})$
	Restricted	$\widehat{\boldsymbol{W}}_i^j=\beta_{i,0}^j+\beta_{i,1}^j\boldsymbol{V}_i^j+\beta_{i,2}^j(\boldsymbol{V}_i^j)^2+\beta_{i,3}^j(\boldsymbol{V}_i^j)^3$	$j+1$
Variants	Fast-Maximum	$\widehat{\boldsymbol{W}}_i^j=\beta'^j_{i,0}+\beta'^j_{i,1}\boldsymbol{V}_1^j+\cdots+\beta'^j_{i,k}\boldsymbol{V}_k^j$	$z(1-\frac{1}{2^j})$
	Fast-Restricted	$\widehat{\boldsymbol{W}}_i^j=\beta'^j_{i,0}+\beta'^j_{i,1}\boldsymbol{V}_i^j+\beta'^j_{i,2}(\boldsymbol{V}_i^j)^2+\beta'^j_{i,3}(\boldsymbol{V}_i^j)^3$	$j+1$
	Exogenous-Max.	$\widehat{\boldsymbol{W}}_i^j=\overline{\beta}_{i,0}^j+\overline{\beta}_{i,1}^j\boldsymbol{V}_1^j+\ldots+\overline{\beta}_{i,k}^j\boldsymbol{V}_k^j$	$z(1-\frac{1}{2^j})$
	Exogenous-Rest.	$\widehat{\boldsymbol{W}}_i^j=\overline{\beta}_{i,0}^j+\overline{\beta}_{i,1}^j\boldsymbol{V}_i^j+\overline{\beta}_{i,2}^j(\boldsymbol{V}_i^j)^2+\overline{\beta}_{i,3}^j(\boldsymbol{V}_i^j)^3$	$j+1$

Table IV Computational cost and side information of the different Haar-based RWA regression models and variants. Please refer to Table II for the notation. ρ is the spatial subsampling percentage.

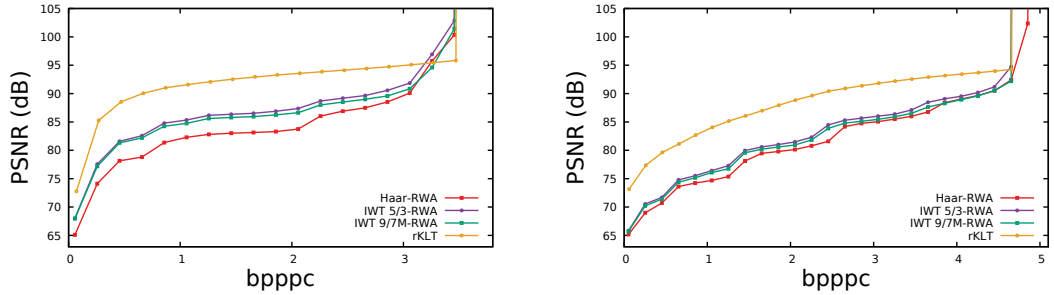
	Regression type	Computational cost	Side information
Models	Maximum	$C_{\mathrm{HT}}+C_{\mathrm{RC}}+C_{\mathrm{EG}}+C_{\mathrm{AR}}$	$\frac{z^2}{3}(1-\frac{1}{2^{2j}})+z(1-\frac{1}{2^j})$
	Restricted	$C_{\mathrm{HT}}+4\Big(\sum_{i=1}^{l}4(2m-1)+16+\frac{z}{2^{i-1}}(m+3)\Big)+10mz\left(1-(\frac{1}{2})^l\right)+C_{\mathrm{AR}}$	$2z(1-\frac{1}{2^j})$
Variants	Fast-Maximum	$\simeq C_{\mathrm{HT}}+\rho C_{\mathrm{RC}}+(1-\rho)\left[(k_i+1)^3+\frac{z}{2^{i-1}}(k_i+1)k_i\right]+C_{\mathrm{EG}}+C_{\mathrm{AR}}$	$\frac{z^2}{3}(1-\frac{1}{2^{2j}})+z(1-\frac{1}{2^j})$
	Fast-Restricted	$\simeq C_{\mathrm{HT}}+\rho C_{\mathrm{RC}}+4\,(1-\rho)\,(16+\frac{3z}{2^{i-1}})+10\rho mz\left(1-(\frac{1}{2})^l\right)+C_{\mathrm{AR}}$	$2z(1-\frac{1}{2^j})$
	Exogenous-Max.	$C_{\mathrm{HT}}+0+C_{\mathrm{EG}}+C_{\mathrm{AR}}$	—
	Exogenous-Rest.	$C_{\mathrm{HT}}+0+C_{\mathrm{EG}}+C_{\mathrm{AR}}$	—

detail components of the other images captured by the same sensor. The regression coefficients are computed only once, off-line, and are thus not stored as side information. Tables III and IV and Figure 3 illustrate these different RWA models and variants.

Figure 4 displays the coding performance of RWA. For lossless coding, it achieves perfect data-recovery at bit-rates close to those obtained by rKLT. However, when adapting it into a PLL framework (Matlab implementation of RWA plus JPEG 2000 bit-plane and entropy coders), a staircase shape appears in the Rate-Distortion curves and the performance worsens. To procure smooth and monotonously increasing Rate-Distortion curves, the predictive weighting scheme is introduced.

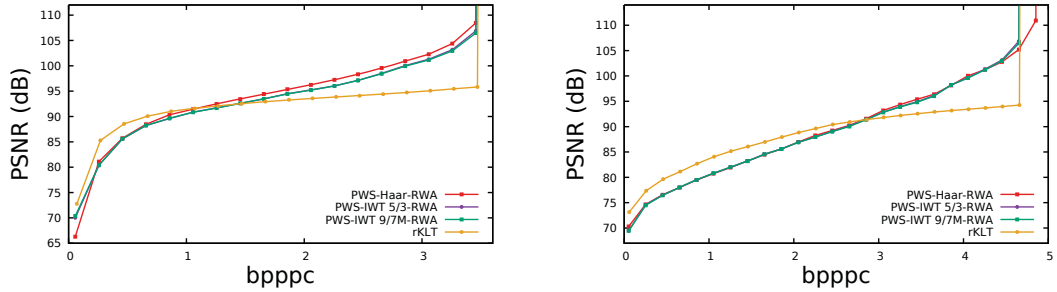
B. Prediction Weighting Scheme

The predictive weighting scheme conveys the contribution of each transformed component to the prediction process when coupled with a rate-distortion optimization (as conducted within JPEG 2000 standard). For the data-recovery process, the existing error in the residuals $\mathbf{R}^j$ is propagated to the detail components of the same scale j, ($\mathbf{W}^j=\mathbf{R}^j+\widehat{\mathbf{W}^j}$). Since $\mathbf{W}^j$ depends on the information contained in $\mathbf{V}^{j-1}$, ($\boldsymbol{W}_i^j=\boldsymbol{V}_{2i}^{j-1}-\boldsymbol{V}_{2i-1}^{j-1}$), the error is also spread to the detail predictions at the previous level $\widehat{\mathbf{W}}^{j-1}$, ($\widehat{\boldsymbol{W}}_i^j=f_i(\mathbf{V}^j)$). To minimize the error propagation, the prediction weights enable JPEG 2000 to allocate more bit-rate to the most significant transformed components, i.e., those of the highest decomposition level, because of the degree of influence of the error propagation in the reconstruction. The prediction weights are employed in the



(*a*) Calibrated AVIRIS-Yellowstone03 (*b*) Uncalibrated Hyperion-Lake Monona

Figure 4 Rate-Distortion curves assessing the performance of several spectral transforms: rKLT and RWA built upon Haar-DWT, IWT 5/3 and IWT 9/7M.



(*a*) Calibrated AVIRIS-Yellowstone03 (*b*) Uncalibrated Hyperion-Lake Monona

Figure 5 Rate-Distortion curves assessing the performance of several spectral transforms: rKLT and RWA (Maximum) built upon Haar-DWT, IWT 5/3 and IWT 9/7M when the Predictive Weighting Scheme is applied.

Post Compression Rate Distortion (PCRD) optimisation within JPEG 2000, where a convex hull process is applied assessing the contribution to the distortion reduction of each coefficient against the bit-rate needed for that coefficient. Assigning larger weights to most significant transformed components leads to transmitting first these significant transformed components, thus better contributing to an improved reconstruction.

From the backward equation in Table I, the prediction of the details at level $j-1$ in the inverse RWA process is computed as:

$$\widehat{\boldsymbol{W}}_i^{j-1} = f_i\left[\mathbf{V}^{j-1}\right] = f_i\left[(\texttt{DWT})^{-1}\left(\mathbf{V}^j, \widehat{\mathbf{W}}^j + \mathbf{R}^j\right)\right]. \quad (2)$$

By extension to level $j-2$ and by replacing $\widehat{\boldsymbol{W}}_i^{j-1}$ (Equation 2), the following formulation is derived:

$$\widehat{\boldsymbol{W}}_i^{j-2} = f_i\left[(\texttt{DWT})^{-1}\left(\mathbf{V}^{j-1}, \widehat{\mathbf{W}}^{j-1} + \mathbf{R}^{j-1}\right)\right] = f_i\left[(\texttt{DWT})^{-1}\left(\mathbf{V}^{j-1}, f_i\left[(\texttt{DWT})^{-1}\left(\mathbf{V}^j, \widehat{\mathbf{W}}^j + \mathbf{R}^j\right)\right] + \mathbf{R}^{j-1}\right)\right]. \quad (3)$$

Reproducing this procedure iteratively until the first decomposition level, the dependency of $\widehat{\boldsymbol{W}}_i^j$ with respect to all the residuals from higher decomposition levels $\left(\mathbf{R}^k\right)^{j+1\leq k\leq J}$ becomes clear. To clarify this, the level one estimated details are given by:

$$\begin{aligned}\widehat{\boldsymbol{W}}_i^1 = f_i\Big[(\texttt{DWT})^{-1}\Big(\mathbf{V}^2, f_i\Big[(\texttt{DWT})^{-1}\left(\mathbf{V}^3, \cdots\cdots\right. \\ \ldots f_i\big[\,(\texttt{DWT})^{-1}\left(\mathbf{V}^{J-1}, f_i\left[(\texttt{DWT})^{-1}\left(\mathbf{V}^J, \widehat{\mathbf{W}}^J + \mathbf{R}^J\right)\right]\cdots\right. \\ \left.\left.\cdots \mathbf{R}^3\,\right)\big] + \mathbf{R}^2\,\Big)\Big]. \end{aligned} \quad (4)$$

Each residual $\mathbf{R}^j$ with $j>1$ is involved into all the $(\widehat{\mathbf{W}}^k)^{1\leq k\leq j-1}$ predictions. Notice that the predicted detail components at level k has $z\cdot 2^{-k}$ components, $\widehat{\mathbf{W}}^k \in \mathbb{R}^{m\times(z\cdot 2^{-k})}$. Hence, the weighting function takes into consideration the number of predictions in which $\mathbf{R}^j$ contributes, and their relative proportion, $u(\mathbf{R}^j)$, with respect to the z components of the image:

$$u(\mathbf{R}^j) = \begin{cases} \sum_{i=1}^{j-1} 2^{-i} & ,\text{if } j>1 \\ 0 & ,\text{if } j=1 \end{cases}. \quad (5)$$

The Predictive Weighting Scheme assigns a different weight to each residual component according to the predictive significance of each component with respect to the ones at the previous scales; it is defined as follows:

$$\mathscr{W}(\mathbf{R}^j) = \frac{1}{u(\mathbf{R}^j) - u(\mathbf{R}^{j-1})} = 2^{j-1}. \quad (6)$$

Finally, the approximation component at the highest RWA decomposition level J, $\mathbf{V}^J \in \mathbb{R}^{m\times 1}$, is assigned weight $\mathscr{W}(\mathbf{V}^J) = 2^J$.

Figure 5 reflects the benefits of applying this PWS. Now, smoothly increasing Rate-Distortion curves are obtained. Moreover, PWS-RWA may yield superior coding performance than rKLT at a much lower computational complexity.

III. Experiments and Results

This section discloses the Rate-Distortion curves, bit-per-pixel-per-component (bpppc) distributions according to a fixed global target bit-rate and different classification measurements for several spectral transforms: rKLT, POT, PWS-Haar-DWT and RWA. The near-lossless technique M-CALIC is also considered.

Figure 6 Pipeline of our coding system: 1D spectral transform followed by JPEG 2000.

Table V Information of the images from AVIRIS, Hyperion and IASI sensors sets. z is the number of spectral components, y is the height and x is the width.

Corpus	Images
AVIRIS (uncalibrated) z=224, y=512, x=680	Yellowstone, sc: 0, 3, 10, 11, 18 Hawaii (x=614) and Maine
AVIRIS (calibrated) z=224, y=512, x=680	Yellowstone, sc: 0, 3, 10, 11, 18
Hyperion (uncalibrated) z=242 x=256	ErtaAle (y=3187) Lake Monona (y=3176) Mt. St. Helens (y=3242)
Hyperion (calibrated) z=242 x=256	Agriculture (y=3129) Coral Reef (y=3127) Urban (y=2905)
IASI Level 0 (uncalibrated) z=8359, y=1528, x=60	L0 1: 20091007093900Z L0 2: 20091007143900Z
IASI Level 1 (calibrated) z=8461, y=1530, x=60	L1 1: 20130816230553Z L1 2: 20130817004753Z

Figure 6 displays the two main blocks that compose our system. The first block corresponds to the 1D transform along the spectral dimension. In our case, this operation is conducted by either rKLT, POT, PWS-Haar-DWT or RWA. The second and last operation stage is performed by a plain JPEG 2000 (Taubman, and Marcellin , 2001), including a bit-plane and an entropy coding step. The Predictive Weighting Scheme is inserted within the JPEG 2000 framework through parameter **Cweight** in Kakadu implementation. Note that no spatial transform has been applied.

The employed software is Matlab 64-bits R2014b for the spectral transforms and Kakadu v6.0 for JPEG 2000. For reproducibility, the rKLT, Haar-DWT, IWT 5/3 and RWA codes are publicly available at Amrani and Serra-Sagristà (2015). IWT 9/7M definition is provided in CCSDS (2012a).

Results of our experiments are reported for six hyperspectral images corpus from three well-known uncalibrated and calibrated sensors: airborne hyperspectral AVIRIS, satellite-borne hyperspectral Hyperion and ultraspectral IASI sensors. For AVIRIS sensor, five scenes with a bit-depth of 16 bits per pixel per component (bpppc), Yellowstone 00, 03, 10, 11 and 18, and two 12 bpppc images, Hawaii and Maine, are used. These two sets have 224 spectral components. The Hyperion sensor records images with 242 spectral components. IASI sensor records images with 8,359 (uncalibrated) and 8,461 (calibrated) spectral components. Both sensors store images with 16 bpppc. Table V gives a description of the dimensions of these images.

A. Predictive Weighting Scheme for Progressive Lossy-to-Lossless coding

Figure 7 displays the Rate-Distortion curves for different RWA regression models and variants (Maximum, Restricted, Exogenous-Maximum and Exogenous-Restricted), illustrating the benefits of applying the predictive weighting scheme. These benefits are achieved for all RWA regression models and variants, and for all images. Comparing the performance of rKLT and PWS-RWA, the former is superior at low to medium bit-rates while the latter is superior at medium to high bit-rates.

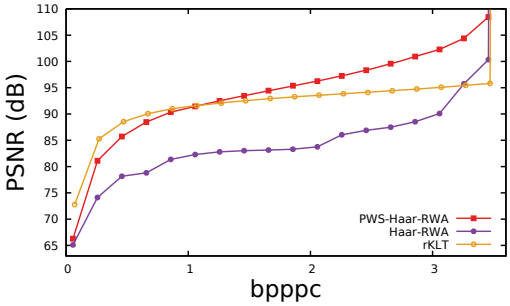
Results for Fast-Maximum and Fast-Restricted are not reported for the sake of paper length and because the performance in these cases is very similar to, respectively, Maximum and Restricted models. Fast-Maximum and Fast-Restricted employ 1% of the total spatial samples for AVIRIS and Hyperion sensors and 10% for IASI sensor for the regression, reducing considerably the computational complexity.

The Restricted model performance slightly falls behind for AVIRIS and IASI corpus. However, for uncalibrated Hyperion images, which are affected by streaking artefacts, the Restricted model achieves better image quality values at lower bpppc (i.e., higher compression ratios), because Haar-RWA Restricted model applies a polynomial function for the details prediction while Maximum model relies on a linear combination of all approximation components.

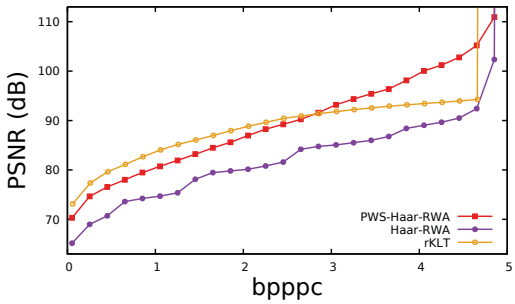
Figure 8 displays the bit-rate distribution per spectral component at different global target bit-rates. These graphics present the assigned bit-rate to each spectral component in the transform domain after compressing by fixing an overall bit-rate. Bit-rate distributions for Maximum and Restricted models are provided. The influence of the predictive weighting scheme in the JPEG 2000 rate-distortion allocation is apparent: comparing the first and second rows, it is noticeable that PWS entails the assignment of larger bit-rates to the most significant components, regardless of the regression method. When no weights are employed, a more uniform bit-rate distribution among all components is found. The third row reports the difference in bpppc between applying or not the PWS. For the two global target bit-rates, the curves depict positive differences for the first and most significant components, and negative differences for the rest. The fourth row plots the accumulation of encoded bits percentage; again, the consequences of applying the PWS become obvious.

B. Comparison among state-of-the-art coding techniques

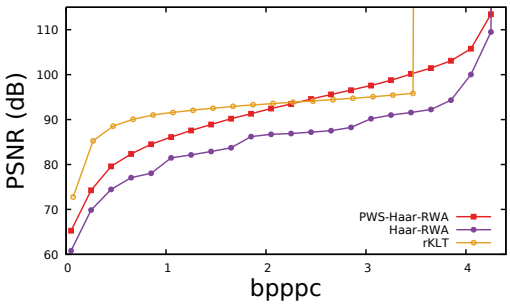
Figure 9 displays the performance produced by two different coding approaches: the first one is a progressive lossy-to-



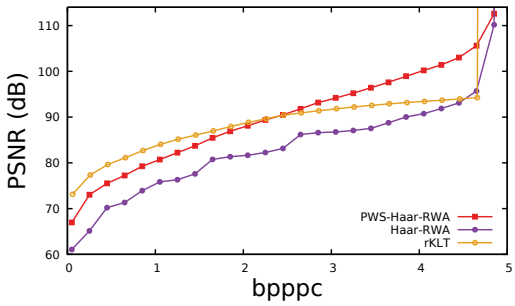
(*a*)(*i*) Maximum (calibrated AVIRIS-Yellowstone03)



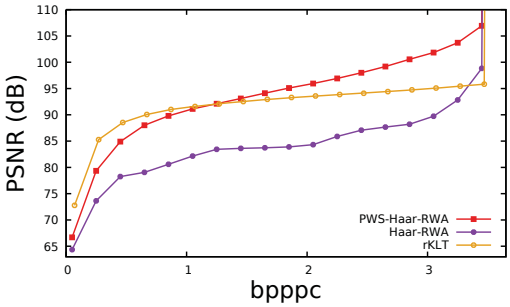
(*a*)(*ii*) Maximum (uncalibrated Hyperion-Lake Monona)



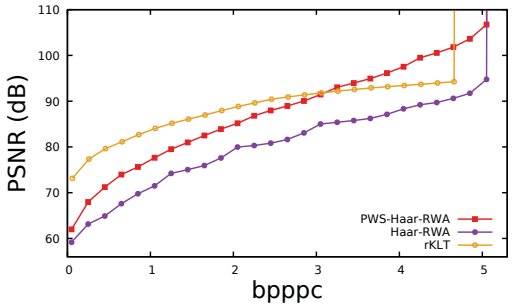
(*b*)(*i*) Restricted (calibrated AVIRIS-Yellowstone03)



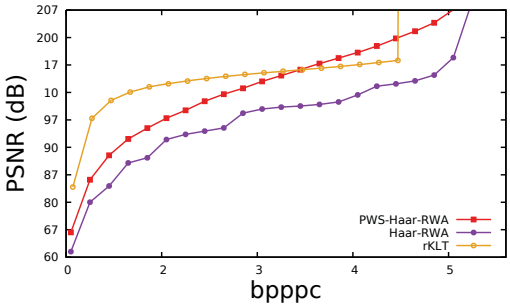
(*b*)(*ii*) Restricted (uncalibrated Hyperion-Lake Monona)



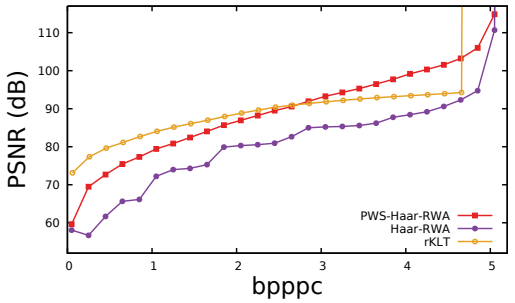
(*c*)(*i*) Exog.-Max. (calibrated AVIRIS-Yellowstone03)



(*c*)(*ii*) Exog.-Max. (uncalibrated Hyperion-Lake Monona)

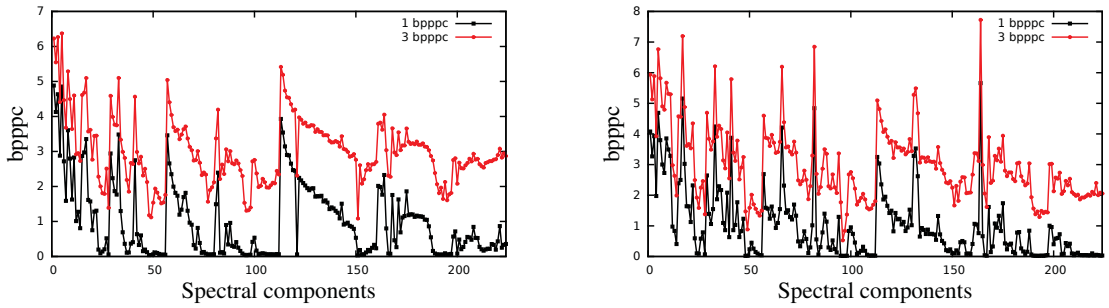


(*d*)(*i*) Exog.-Rest. (calibrated AVIRIS-Yellowstone03)



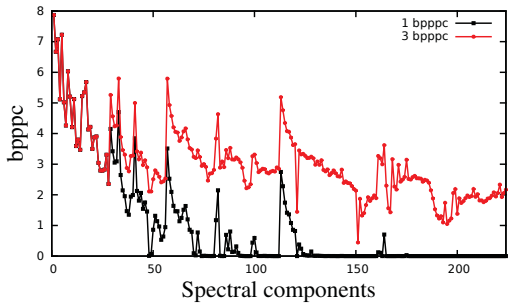
(*d*)(*ii*) Exog.-Rest. (uncalibrated Hyperion-Lake Monona)

Figure 7 Rate-Distortion curves for different RWA regression models and variants when applying or not the Predictive Weighting Scheme. rKLT performance is included for comparison purposes.

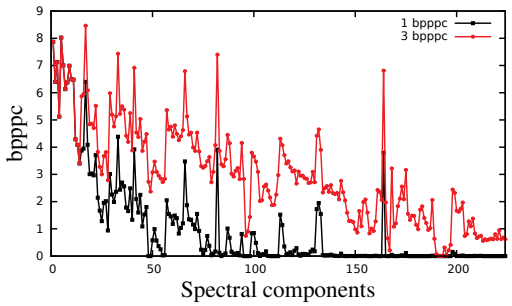


(*a*)(*i*) Maximum Haar-RWA

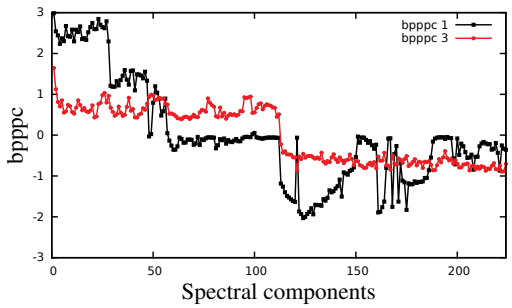
(*a*)(*ii*) Restricted Haar-RWA



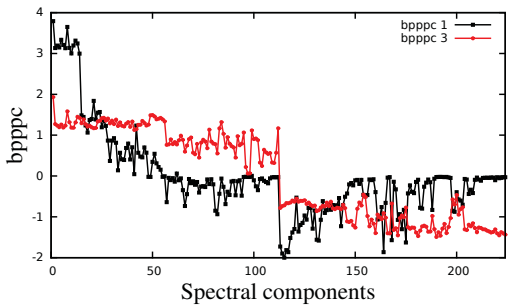
(*b*)(*i*) Maximum Haar-PWS-RWA



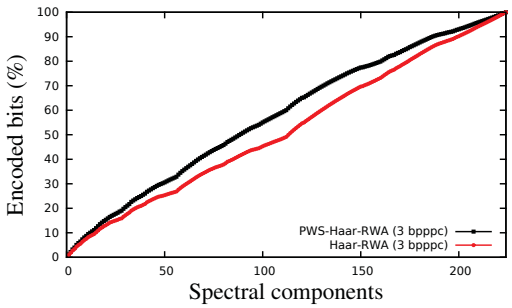
(*b*)(*ii*) Restricted Haar-PWS-RWA



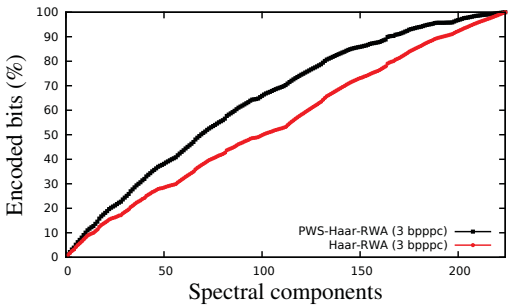
(*c*)(*i*) Differences (Maximum)



(*c*)(*ii*) Differences (Restricted)



(*d*)(*i*) Bits accumulation (Maximum)



(*d*)(*ii*) Bits accumulation (Restricted)

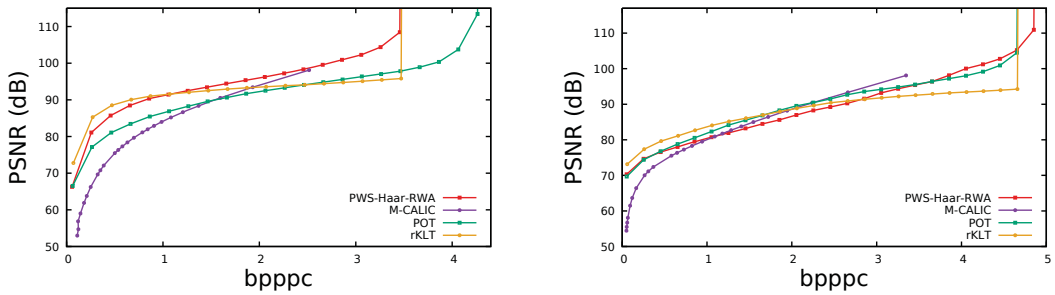
Figure 8 Bit-rate distribution in the transform domain for two different global target bit-rates. First row: distribution when PWS is not applied. Second row: distribution when PWS is applied. Third row: difference in bit-rate distribution between applying the PWS and not applying the PWS. Fourth row: accumulative percentage (%) of the encoded bits per spectral component. Results for Maximum and Restricted regression models over calibrated AVIRIS Yellowstone03 image are reported.

Table VI Time measurements for different lossless compression methods. Note that here the basis DWT for RWA is the Haar-DWT.

		RWA (Max.) + JPEG 2000	RWA (Exog.-Max.) + JPEG 2000	rKLT + JPEG 2000	POT + JPEG 2000	M-CALIC
Calibrated AVIRIS Yellowstone03	Forward	0:00:50.91’	0:00:10.79’	0:02:11.78’	0:00:12.43’	0:04:19.26’
	JPEG 2000	0:00:20.94’	0:00:21.11’	0:00:23.17’	0:00:22.86’	
	Backward	0:00:13.58’	0:00:10.50’	0:00:59.90’	0:00:10.46’	
Uncalibrated Hyperion Lake Monona	Forward	0:02:12.55’	0:00:22.10’	0:05:31.37’	0:00:29.10’	0:11:10.94’
	JPEG 2000	0:01:08.15’	0:01:07.30’	0:01:08.80’	0:01:01.84’	
	Backward	0:00:35.36’	0:00:25.35’	0:03:04.71’	0:00:24.54’	
Calibrated IASI L1 2	Forward	1:37:44.72’	0:05:26.38’	—	0:03:04.49’	1:27:25.34’
	JPEG 2000	0:05:49.45’	0:05:45.81’	—	0:06:25.36’	
	Backward	0:14:37.80’	0:06:02.17’	—	0:01:50.84’	

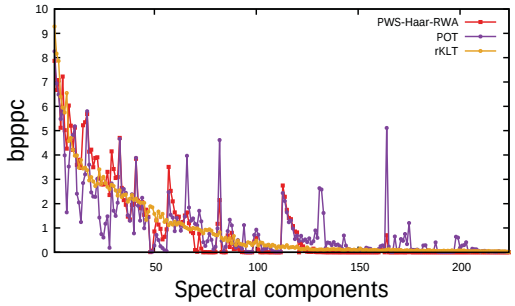
Table VII Indian Pines and Salinas ground truth labeling.

	Indian Pines		Salinas	
Labels	Class	Number of samples	Class	Number of samples
0	Background	10776	Background	56975
1	Alfalfa	46	Brocoli (green weeds 1)	2009
2	Corn-notill	1428	Brocoli (green weeds 2)	3726
3	Corn-mintill	830	Fallow	1976
4	Corn	237	Fallow (rough plow)	1394
5	Grass-pasture	483	Fallow smooth	2678
6	Grass-trees	730	Stubble	3959
7	Grass-pasture-mowed	28	Celery	3579
8	Hay-windrowed	478	Grapes (untrained)	11271
9	Oats	20	Soil (vinyard develop)	6203
10	Soybean-notill	972	Corn (senesced green weeds)	3278
11	Soybean-mintill	2455	Lettuce (romaine 4wk)	1068
12	Soybean-clean	593	Lettuce (romaine 5wk)	1927
13	Wheat	205	Lettuce (romaine 6wk)	916
14	Woods	1265	Lettuce (romaine 7wk)	1070
15	Buildings-Grass-Trees-Drives	386	Vinyard (untrained)	7268
16	Stone-Steel-Towers	93	Vinyard (vertical trellis)	1807

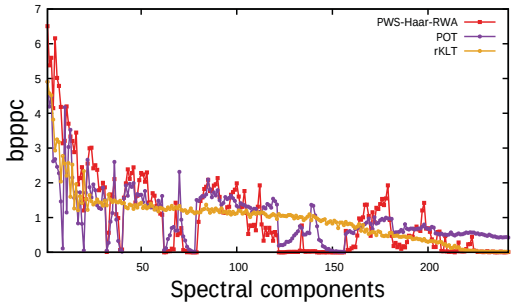


(*a*) Calibrated AVIRIS-Yellowstone03 (*b*) Uncalibrated Hyperion-Lake Monona

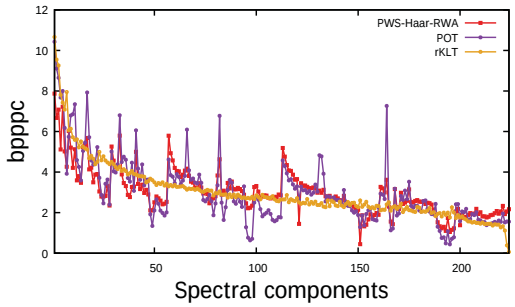
Figure 9 Rate-Distortion curves assessing the performance of several spectral transforms: rKLT, POT and PWS-Haar-RWA (Maximum). The near-lossless method M-CALIC performance is also reported.



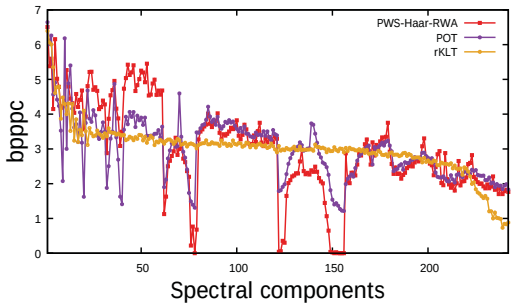
(*a*)(*i*) Global target bit-rate: 1 bpppc (calibrated AVIRIS-Yellowstone03)



(*a*)(*ii*) Global target bit-rate: 1 bpppc (uncalibrated Hyperion-Lake Monona)

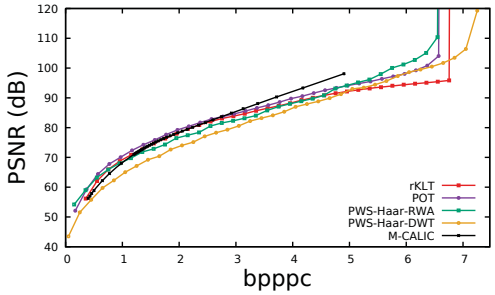


(*b*)(*i*) Global target bit-rate: 3 bpppc (calibrated AVIRIS-Yellowstone03)

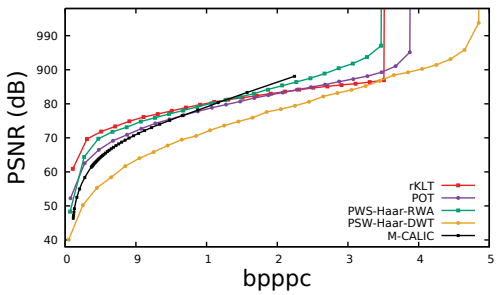


(*b*)(*ii*) Global target bit-rate: 3 bpppc (uncalibrated Hyperion-Lake Monona)

Figure 10 Bit-rate distribution for two global target bit-rates and three spectral transforms: rKLT, POT and RWA (Maximum).



(*a*) AVIRIS-Indian Pines



(*b*) AVIRIS-Salinas

Figure 11 Rate-Distortion of M-CALIC, POT, rKLT, PWS-Haar-DWT and PWS-Haar-RWA when applying the Maximum model.

lossless approach, where either rKLT, POT or PWS-Haar-RWA are applied as spectral transform, and JPEG 2000 standard is performed afterwards. The second approach consists of near-lossless M-CALIC coding technique.

(i) rKLT is a reversible version of KLT. It depicts stochastic processes as infinite linear combinations of orthogonal functions combinations and applies a lifting scheme.
(ii) POT is a low complexity version of KLT. It achieves very competitive performance at lower computational complexity and memory requirements, and it offers a high degree of component-scalability.
(iii) M-CALIC is a lossless and near-lossless adapted version of CALIC coding technique specifically intended for hyperspectral images, conducting an efficient spectral decorrelation in the first stage. M-CALIC is not a progressive lossy-to-lossless technique and no Rate-Distortion curves can be properly rendered; however, by fixing a progressive maximum absolute error per pixel value, we can obtain the image quality at different compression ratios.

For calibrated AVIRIS Yellowstone image, PWS-RWA outperforms POT and M-CALIC at all compression ratios. For Hyperion Lake Monona, PWS-RWA achieves a very competitive performance. For high bit-rates, i.e., close to 2.4 bpppc and higher, PWS-RWA gives rise to higher Peak Signal-to-Noise Ratio (PSNR) values than rKLT at considerably

lower computational complexity. For higher bit-rates, it also outperforms POT. In other words, PWS-RWA undertakes slightly better the global target bit-rate relationship for high bit-rates. For low bit-rates, rKLT outperforms the rest of the methods. Note that here M-CALIC is computed with its near-lossless configuration, where the lowest fixed maximum error per pixel is one. M-CALIC provides perfect data-recovery at 3.98 and 4.92 bpppc for, respectively, the Yellowstone03 and Lake Monona images (lossless compression, or, equivalently, maximum allowed error is zero).

By analising their bpppc distribution according to a global target bit-rate (Figure 10), all the techniques generally assign more bit-rate to the most significant components. It can also be noticed that PWS-RWA globally allocates less bit-rate to the less significant components than POT. However and in contrast to the rKLT method that distributes the bit-rate in a monotonic decreasing form, both POT and PWS-RWA show significant peaks.

Finally, referring to the computational cost, Table VI reports the running times required for the compared methods when executed on an Intel(R) Xeon(R) CPU E5520 @ 2.27GHz processor with a single thread. Measured times correspond to encoding and decoding images from the three different corpus: AVIRIS, Hyperion and IASI. Some caution has to be taken when analysing these execution times: rKLT is implemented in Java, POT and M-CALIC in C++, and RWA in Matlab; Java and C++ are known to be considerably faster than Matlab. In spite of Matlab inefficiency, Exogenous-Maximum RWA is significant less expensive than other methods such as rKLT and M-CALIC. Exogenous-Maximum RWA is comparable to POT for AVIRIS and Hyperion images. For IASI images, the fastest spectral transform is POT, whose implementation has been carefully optimised, while rKLT is not able to transform them.

C. Digital classification

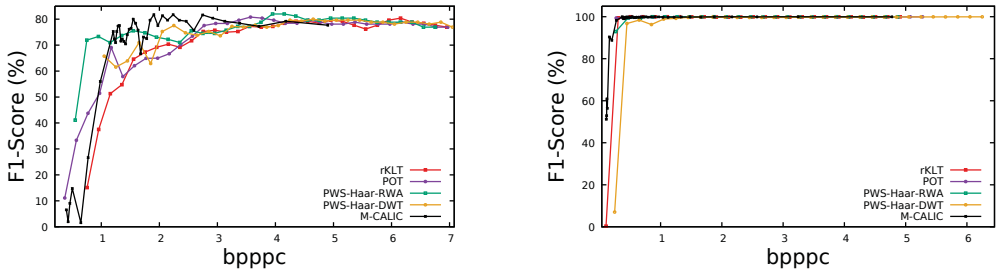
To illustrate the influence the image quality loss has over widely used applications, we introduce in this section results for digital land classification. To this end, two AVIRIS images, Indian Pines (Purdue University , 1992) and Salinas (NASA , 1998), are classified after they have been compressed through different PLL coding techniques. Two different classifiers have been employed: Linear Discriminant Analysis (LDA) (Dwinnell, and Sevis , 2010) and linear Support Vector Machines (SVM) (Neuburger , 2012). Indian Pines represents a land portion of North-Western Indiana. It has a size of 145 x 145 spatial samples. Here we consider 200 spectral reflectance bands, as 24 water absorption bands were removed from the original 224 bands image. This scene is composed by two-thirds agriculture and one-third forest and other types of natural vegetation. Some low density housing, buildings, small roads and a rail line are also visible. Salinas has been taken over Salinas Valley (California) from a NOAA Twin Otter flight at low altitude. It has 512 lines by 217 samples at high spatial resolution (3.7 meter pixels). 20 water absorption spectral bands were also discarded from the 224 spectral components. This image covers land with vegetables, bare soils and vineyard fields. Table VII enumerates the 16 classes and background with which the ground truths have been labeled. Notice that for the classification results, the background samples have not been considered.

First, Figure 11 reports the Rate-Distortion curves when applying the coding methods based on rKLT, POT, M-CALIC, PWS-Haar-DWT, and PWS-Haar-RWA (Maximum) to the images Indian Pines and Salinas. In general, all techniques show a similar behaviour, with PWS-Haar-RWA yielding, on average, the best coding performance. However, for digital classification, improved coding performance might not necessarily imply improved classification performance.

For the multi-class classification application we use the Linear Discriminant Analysis (LDA) and several linear Support Vector Machines (SVM) algorithms. LDA is commonly used in feature and pattern recognition for separating several classes, objects or events by means of features linear combinations. It is close to KLT in that they both attempt to classify the data according to linear combinations. Nonetheless, LDA is more focused on modeling the differences between classes of a certain data. On the other side, the multi-class SVM implementation is based on applying as many binary SVMs as the number of labeled classes for the training process. To this end, the employed strategy corresponds to one vs. all. After this, the best label region candidate from training is assigned to each test features sample set (Neuburger , 2012). SVMs are supervised learning models with associated learning algorithms. In this article, both classifiers are trained using 20% of the original image samples of each class.

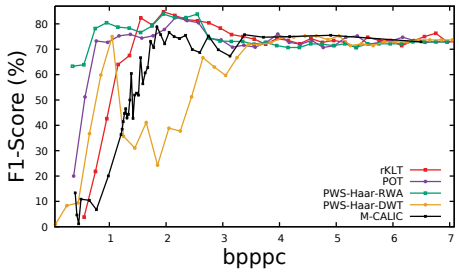
For multi-class classification, the confusion matrices are usually assessed and presented. In our case, and for both images, several samples from different classes are very similar. For instance, for the Indian Pines image, classes 10, 11 and 12 correspond to some variation of soybean region. For the Salinas image, there are four different classes that cover lettuce. This may conduct to poorer or richer classification results when analysing each class separately. Besides, specially for the Indian Pines image, some labeling imbalance effect is strongly registered. Its class 9 (Oats) has only 20 spatial samples but other classes such as class 11 (Soybean-mintill) contains 2,455 samples. Because of this, the confusion matrices do not give rise to a clear conclusion of the classification performance and we can not deduce which coding technique influences more the classification labeling outcomes. For a more general assessment and for better understanding purposes, the F1-Score and the hits percentage are therefore provided. F1-Score corresponds to the harmonic mean of precision and recall. The hits percentage is defined as the total number of true positives with respect to all the classification outputs. The best value for these metrics is 100 % and the worst is 0 %.

Figures 12 and 13 plot the micro-average F1-Score and the hits percentage with respect to all the testing classification results percentages per image, when encoded at different target bit-rates. In our case, Indian Pines and Salinas are composed by classes from respectively 20 to 2,455 and from 916 to 11,271 different number of samples (Table VII). We compute these metrics once all the total number of true positives, true negatives, false positives and false negatives for

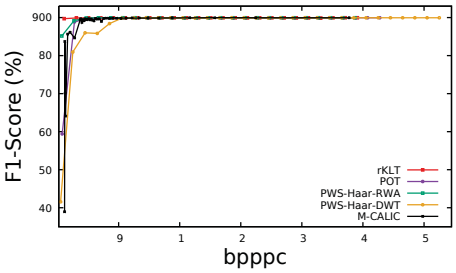


(*a*)(*i*) LDA (AVIRIS-Indian Pines)

(*a*)(*ii*) LDA (AVIRIS-Salinas)

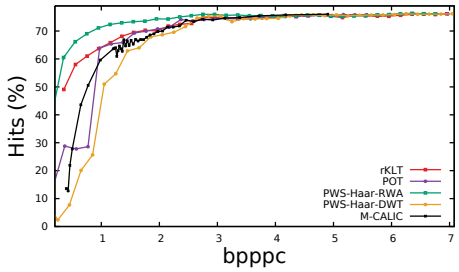


(*b*)(*i*) SVM (AVIRIS-Indian Pines)

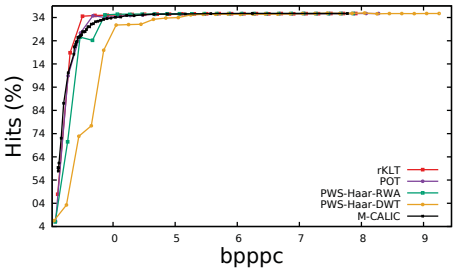


(*b*)(*ii*) SVM (AVIRIS-Salinas)

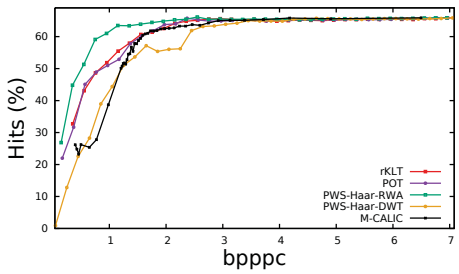
Figure 12 F1-Score for two classifiers: LDA and linear SVM. The assessed coding methods are rKLT, POT, PWS-Haar-DWT, PWS-Haar-RWA(Maximum) and M-CALIC.



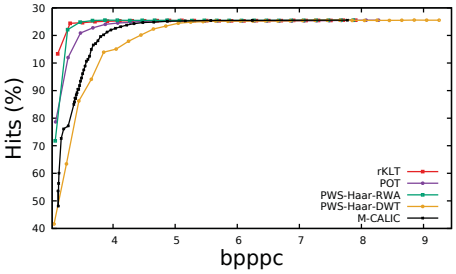
(*a*)(*i*) LDA (AVIRIS-Indian Pines)



(*a*)(*ii*) LDA (AVIRIS-Salinas)



(*b*)(*i*) SVM (AVIRIS-Indian Pines)



(*b*)(*ii*) SVM (AVIRIS-Salinas)

Figure 13 Hits percentage for two classifiers: LDA and linear SVM. The assessed coding methods are rKLT, POT, PWS-Haar-DWT, PWS-Haar-RWA(Maximum) and M-CALIC.

all the predicted classes have been added together. This global measure tackles better the label imbalance than computing these measurements per class. Both figures depict a strongly transitory period, specially for M-CALIC, POT, rKLT and PWS-Haar-DWT for very low bit-rate (i.e., for reconstructed images with a large distortion). At low to medium bit-rates, these techniques achieve overall worse classification results than PWS-Haar-RWA. This latter technique achieves a relatively good performance even at bit-rates below 1 bpppc. At high bit-rates, above 3 bpppc, all coding techniques yield equivalent classification performance.

Concerning the F1-Score and hits percentage curves, a sawtooth shape can be appreciated in Figures 12 and 13, mostly for Indian Pines image. For F1-Score, PWS-Haar-RWA presents the most stable curves evolution, i.e., the F1-Score variance corresponds to 0.0051 and 0.0023 for respectively the LDA and the multi-SVM classifiers; in comparison, M-CALIC compression method leads to variances of, respectively, 0.0576 and 0.0547, and the coding approach based on POT spectral transform yields variances of, respectively, 0.0235 and 0.0108 . For the hits percentage, PWS-Haar-RWA achieves the best percentage at all bit-rates. For Salinas image, the performance for the different approaches is quite similar, with rKLT and PWS-Haar-RWA yielding, on average, the best outcome.

IV. Conclusions

The Regression Wavelet Analysis (RWA) adaptation to a progressive lossy-to-lossless coding technique for hyperspectral images gives rise to non-monotonic Rate-Distortion curves. A predictive scheme is therefore required to achieve smooth and monotonously increasing curves. This Predictive Weighting Scheme (PWS) assigns different weights according to the pyramidal significance of the RWA spectral bands, thus better driving the JPEG 2000 rate-allocation process. It considerably improves the coding gain, becoming the most competitive progressive lossy-to-lossless framework published to date, while minimizing the error propagation effect.

In this paper, we show that PWS-RWA concentrates the larger amount of bit-rate into the most significant spectral components in the recovery process, yielding steady quality rate-distortion performance, better than other lossy techniques at low bit-rates and the same coding behaviour in the lossless regime. PWS-RWA is superior to rKLT spectral transform at high bit-rates at a significantly lower computational cost. At low bit-rates, it can outperform other spectral transforms such as POT and near-lossless coding technique M-CALIC.

The Predictive Weighting Scheme enhances the PLL-RWA framework regardless of the regression model or variant. For all of them, Maximum, Restricted, Fast-Maximum, Fast-Restricted, Exogenous-Maximum or Exogenous-Restricted, PWS-RWA substantially outperforms the unweighted RWA approach at all bit-rates.

The effects of lossy compression on nowadays popular applications such as digital land classification have also been investigated. Coding techniques such as near-lossless M-CALIC and other progressive lossy-to-lossless techniques based on PWS-RWA, rKLT, POT and PWS-Haar-DWT spectral transforms have been proven to be suitable for multi-class classification, obtaining successful results at medium and high bit-rates, where the loss in image quality is not abnormally high. In the case of PWS-RWA, the classification outcomes are overall better than for the other methods at low bit-rates and with a lower computational complexity.

To summarize, applying a Predictive Weighting Scheme to the adapted progressive lossy-to-lossless RWA approach provides smooth and steady image quality recovery. At the same time, it outperforms other state-of-the-art methods at lower computational complexity. PWS-RWA has proven to be the most competitive compression method for achieving good classification results at low to medium bit-rates, providing a smoother performance for both F1-Score and hits percentage than the other approaches.

V. Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments.

This work was supported in part by the Spanish Ministry of Economy and Competitiveness (MINECO), by the European Regional Development Fund (FEDER) and by the European Union (EU) under Grant TIN2015-71126-R, by the Catalan Government under Grant 2014SGR-691 and by Universitat Autònoma de Barcelona under Grants UAB-PIF-472/2012 and UAB-PIF-472/2015.

An earlier version of this work was presented at the ESA CNES On-Board Payload Data Compression Workshop 2016.

References

Amrani, N. and J. Joan Serra-Sagristà. 2015. “RWA coding toolbox.” http://gici.uab.es/GiciWebPage/downloads.php

Amrani, N., J. Serra-Sagristà, V. Laparra, M. Marcellin, and J. Malo. 2016. “Regression wavelet analysis for lossless coding of remote-sensing data.” *IEEE Transactions on Geoscience and Remote Sensing* 54 (9): 5616-5627, ISSN: 0196-2892, DOI: 10.1109/TGRS.2016.2569485.

Amrani, N., J. Serra-Sagristà, M. Hernández-Cabronero, and M. Marcellin. 2016. “Regression wavelet analysis for progressive-lossy-to-lossless coding of remote-sensing data.” *Data Compression Conference DCC. IEEE.* 121-130.

Barret M., J.L. Gutzwiller, and M. Hariti. 2011. “Low-Complexity Hyperspectral Image Coding Using Exogenous Orthogonal Optimal Spectral Transform (OrthOST) and Degree-2 Zerotrees.” *IEEE Transactions on Geoscience and Remote Sensing.* 49 (5): 1557–1566, DOI: 10.1109/T-GRS.2010.2083671.

Blanes, I. and J. Serra-Sagristà. 2011. “Pairwise Orthogonal Transform for Spectral Image Coding.” *IEEE Transactions on Geoscience and Remote Sensing* 49 (3): 961–972, ISSN: 0196-2892, DOI: 10.1109/TGRS.2010.2071880.

Blanes, I., E. Magli, and J. Serra-Sagristà. 2014. “A Tutorial on Image Compression for Optical Space Imaging Systems.” *IEEE Geoscience and Remote Sensing Magazine* 2(3): 8-26.

Calderbank, A.R., I. Daubechies, W. Sweldens and Y. Boon-Lock. 1997. “Lossless image compression using integer to integer wavelet transforms.” *Image Processing, 1997.*

Proceedings., International Conference on. 1: 596–599, DOI: 10.1109/ICIP.1997.647983.

Camps-Valls, G., D. Tuia, L. Gomez-Chova, S. Jimenez and J. Malo. 2011. "Remote Sensing Image Processing." *Synthesis Lectures on Image, Video and Multimedia Proc. Morgan & Claypool Publ. Austin, TX, USA*.

Consultative Committee for Space Data Systems (CCSDS). 2012. "Image Data Compression CCSDS 122.0-B-1." *CCSDS. Blue Book.* https://public.ccsds.org/Pubs/122x0b1c3.pdf

Consultative Committee for Space Data Systems (CCSDS). 2012. "Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1." *CCSDS. Blue Book.* https://public.ccsds.org/Pubs/123x0b1ec1.pdf

DIS, I. 10918-1. 1991. "Digital Compression and Coding of Continuous-tone Still Images (JPEG)." *CCITT Recommendation T.* 81.

Dwinnell, W. and D. Sevis. 2010. "LDA: Linear Discriminant Analysis." *Senior Data Scientist. Statistically rigorous predictive models codes. Springleaf Financial Services, Barclaycard USA*. https://es.mathworks.com/matlabcentral/fileexchange/29673-lda-linear-discriminant-analysis/content/LDA.m?requestedDomain=www.mathworks.com

Effros, M., H. Feng and K. Zeger. 2004. "Suboptimality of the Karhunen-Loeve Transform for transform coding." *IEEE Transactions on Information Theory.* 50 (8): 1605–1619.

Ginesu, G., D. D. Giusto, and W. A. Pearlman. 2004. "Lossy to lossless SPIHT-based volumetric image compression." *In Acoustics, Speech, and Signal Processing (ICASSP'04) Proceedings. IEEE International Conference on.* 3: iii-693.

Jolliffe, I. T. 2002. "Principal Component Analysis." *Springer Verlag, Berlin, Germany.* 487.

Magli, E., G. Olmo and E. Quacchio. 2004. "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC." *IEEE Geoscience and Remote Sensing Letters.* 1 (1): 21–25, ISSN: 1545-598X.

Nasa. 1998. "AVIRIS Data Products." *California Institute of Technology (Jet Propulsion Laboratory - JPL).* http://aviris.jpl.nasa.gov/

Neuburger, C. 2012. "Multi-SVM implementation." *Florida Atlantic University.* http://es.mathworks.com/matlabcentral/fileexchange/39352-multi-class-svm?focused=3776247&tab=function

Nocedal, J. and S.J. Wright. 2006. "Least-Squares Problems." *Springer.*

Penna, B., T. Tillo, E. Magli and G. Olmo. 2006. "Transform Coding Techniques for Lossy Hyperspectral Data Compression." *IEEE Transactions on Geoscience and Remote Sensing.* 45 (5): 1408–1421.

Purdue University Research Repository (PURR). 1992. "Freeware multispectral image data analysis system." https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html

Tang, X. and W.A. Pearlman. 2006. "Three-dimensional wavelet-based compression of hyperspectral images." *Hyperspectral Data Compression. Springer.* 273–308.

Taubman, D., and M. Marcellin. 2001. "JPEG 2000: Image compression fundamentals, practices and standards" *Springer.*

Zhang, J., J. E. Fowler, and G. Liu. 2008. "Lossy-to-lossless compression of hyperspectral imagery using three-dimensional TCE and an integer KLT." *IEEE Geoscience and Remote Sensing Letters.* 5(4): 814-818.

Chapter 4

Multilevel split regression wavelet analysis for lossless compression of remote sensing data

Multi-Level Split Regression Wavelet Analysis for Lossless Compression of Remote Sensing Data

Sara Álvarez-Cortés, Joan Bartrina-Rapesta, and Joan Serra-Sagristà

***Abstract*—Spectral redundancy is a key element to be exploited in compression of remote sensing data. Combined with an entropy encoder, it can achieve competitive lossless coding performance. One of the latest techniques to decorrelate the spectral signal is Regression Wavelet Analysis (RWA). RWA applies a wavelet transform in the spectral domain and estimates the detail coefficients through the approximation coefficients using linear regression. RWA was originally coupled with JPEG 2000. This letter introduces a novel coding approach where RWA is coupled with the predictor of CCSDS-123.0-B-1 standard and a lightweight contextual arithmetic coder. In addition, we also propose a smart strategy to select the number of RWA decomposition levels that maximizes the coding performance. Experimental results indicate that, on average, the obtained coding gains vary between 0.1 to 1.35 bits-per-pixel-per-component compared with other state-of-the-art coding techniques.**

***Index Terms*—Spectral decorrelation, predictive coding, lossless coding.**

I. Introduction

The data acquired by on-board remote sensing sensors is an unvaluable tool for governments, rescue teams, and aid organizations to manage infrastructure and natural resources, to appraise climate changes, or to give support when natural disasters strike. The data produced by these sensors is increasing unprecedently in each mission and, therefore, low-complexity and high-performing lossless compression techniques are of paramount importance.

Lossless coding techniques exploit the redundancy in the spatial and spectral dimensions of the scenes, allowing perfect reconstruction. Several transforms, such as Discrete Wavelet Transform, Principal Component Analysis (PCA) [1], Pairwise Orthogonal Transform (POT) [2] or the most recent Regression Wavelet Analysis (RWA) [3], can be employed to take advantage of the spectral redundancy. RWA is a low-complexity transform that reduces the redundancy still remaining after the computation of a wavelet transform along the spectral dimension. In [3] it is shown that RWA followed by JPEG 2000 [4] (RWA+JPEG2000) usually achieves higher compression ratios than widespread state-of-the-art lossless coding techniques such as PCA+JPEG2000, POT+JPEG2000, M-CALIC [5] and CCSDS-123.0-B-1 [6]. Additionally, in [7], the performance of RWA+JPEG2000 at several decomposition levels for different regression models and variants is evaluated, concluding that, when using all the regressors and the highest number of decomposition levels, RWA gives rise to the best coding gain on remote sensing data.

Although JPEG 2000 can achieve high lossless compression ratios, it is too computationally demanding for remote sensing sensors. In this regard, the Consultive Committee for Space Data Systems (CCSDS) published in 2012 the CCSDS-123.0-B-1 standard (CCSDS-123 in what follows), intended for lossless coding of multispectral and hyperspectral data and based on prediction and two different entropy encoders. Spatial and spectral redundancy is exploited through an adaptive linear prediction method that makes use of its nearby samples within a small three-dimensional neighborhood. The prediction is sequentially conducted in a single pass, entailing minor computational complexity.

Despite the affordable computational complexity of CCSDS-123, it allows to achieve high lossless compression ratios. Lately, several contributions [8]–[11] have been presented that surpass its coding performance at the expense of a higher computational cost. In particular, [10], [11] use a conventional recursive least-squares technique to predict the current sample and compute the residual, then the residual is entropy encoded by an arithmetic encoder. Thus, both contributions are a trade-off between coding performance and computational complexity. In 2017, Bartrina-Rapesta et al. [12] presented a compression technique based on the CCSDS-123 predictor and followed by a low-complexity arithmetic coder (CCSDS-123-AC) using a novel and simple context model that only assesses causal adjacent samples and inexpensive low-cost bitwise operations for the symbol probability estimations; on average, it improves CCSDS-123 and M-CALIC by 0.1 and 0.86 bits-per-pixel-per-component (bpppc), respectively.

This letter presents a novel compression technique based on RWA that obtains higher compression ratios with lower computational cost than the original [3]. In our proposal, the computational cost is reduced by employing a smart strategy that keeps applying further RWA decomposition levels only if it provides coding gain. After the smart-RWA is conducted, the decorrelated signal is encoded with the low complexity CCSDS-123-AC.

This paper is structured as follows: Section II reviews RWA transform. Section III introduces our proposed coding technique. Section IV provides experimental results and discussion of the achieved results as compared to state-of-the-art techniques for data captured by different hyperspectral sensors. Section V brings forward our conclusions.

Sara Álvarez-Cortés, Joan Bartrina-Rapesta, and Joan Serra-Sagristà are with the Department of Information and Communications Engineering, Building Q, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona, Spain. Tel. + 34 93 581 1861.

This work was supported in part by the Spanish Ministry of Economy and Competitiveness (MINECO) and by the European Regional Development Fund (FEDER) under Grant TIN2015-71126-R, by the Catalan Government under Grant 2017SGR-463 and by Universitat Autònoma de Barcelona under Grant UAB-PIF-472/2015.

II. Regression Wavelet Analysis Overview

RWA [3] is composed by two sequential operations: a discrete wavelet transform (DWT) –commonly the S-Transform– and a prediction stage -performed through a linear regression-.

A. Discrete Wavelet Transforms

The DWT operation comprises a pyramidal multi-resolution decomposition along the 1-D spectral dimension. Considering a multi-component scene with z spectral components and m spatial samples per component, let $\mathbf{V}^0 \in \mathbb{R}^{m\times z}$ be

$$\mathbf{V}^0 = \left[V^0(1), \ldots, V^0(z)\right], \quad \text{with} \quad V^0(i) = V_i^0 \in \mathbb{R}^{m\times 1},$$

where $i=\{1, ..., z \cdot 2^{-j}\}$ and $j=\{1, ..., J\}$, $J = \lceil \log_2(z) \rceil$. The original signal $\mathbf{V}^0$ is decomposed into the half-resolution approximation $\mathbf{V}^1$ and the half-resolution detail $\mathbf{W}^1$ signals. This process is repeated for $\mathbf{V}^{j-1}$, yielding $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ and $\mathbf{W}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ signals. $\mathbf{V}^j$ contains a reduced version of $\mathbf{V}^{j-1}$ information and $\mathbf{W}^j$ retains the detail information, namely, the difference between $\mathbf{V}^{j-1}$ and $\mathbf{V}^j$.

The one level DWT decomposition of each $\mathbf{V}^{j-1}$ can be expressed as follows:

$$(\text{DWT})(\mathbf{V}^{j-1}, 1) = \left(\mathbf{V}^j, \mathbf{W}^j\right). \tag{1}$$

This decomposition is recursively performed in a pyramid form over $\mathbf{V}^j$, while the detail components $\mathbf{W}^{1\leq j}$ remain unchanged. By induction, from $\mathbf{V}^0$, the wavelet representation with J levels is described as follows:

$$(\text{DWT})(\mathbf{V}^0, J) = \left(\mathbf{V}^J, (\mathbf{W}^j)^{1\leq j\leq J}\right). \tag{2}$$

B. Ordinary Least-Squares Method

The second sequential operation of the RWA consists of a light linear regression aimed to remove the spectral redundancy that still remains in $\mathbf{W}^{1\leq j\leq J}$.

At each level j, the regression model estimates each detail component $\left(\widehat{\mathbf{W}}_i^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}\right)$ from $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ employing an ordinary least-squares (OLS) method [13].

At level j, the estimations can be computed as:

$$\widehat{\mathbf{W}}_i^j = f_i\left[\mathbf{V}_i^j\right] = \beta_{i,0}^j + \beta_{i,1}^j \boldsymbol{V}_1^j + \cdots + \beta_{i,k}^j \boldsymbol{V}_k^j, \tag{3}$$

where $\boldsymbol{\beta}^j$ correspond to the regression coefficients at level j, $\boldsymbol{\beta}_i^j \in \mathbb{R}^{(k+1)\times 1}$, $\mathbf{V}_i^j \in \mathbb{R}^{m\times 1}$ and $k = z \cdot 2^{-j}$ is the number of approximation components employed by the OLS method. Note that $\boldsymbol{\beta}^j$ values must be stored as side information for decoding purposes. The size of the side information amounts to $\frac{z^2}{3}(1-\frac{1}{2^{2j}}) + z(1-\frac{1}{2^j})$ parameters.

Once the predicted signal $\widehat{\mathbf{W}}^j$ is computed, the residuals are obtained by subtraction:

$$\mathbf{R}^j = \mathbf{W}^j - \widehat{\mathbf{W}}^j. \tag{4}$$

Then, assuming that the highest number of decomposition levels J has been applied, $\mathbf{V}^J$ and $\mathbf{R}^{1\leq j\leq J}$ are losslessly entropy encoded together. At the decoder side, to recover the signal $\mathbf{V}^{J-1}$, the approximations $\mathbf{V}^J$ and the residuals $\mathbf{R}^J$ are employed. This process is computed in cascade form (for $j = J-1, \ldots, 1$). For a more in-detail explanation of RWA reverse process, see [3].

C. RWA Variants

RWA can be executed on different modes to compute the estimation $\widehat{\mathbf{W}}^j$ of the detail components $\mathbf{W}^j$ based on the approximation components $\mathbf{V}^j$: *Maximum*, *Restricted*, *Fast* and *Exogenous* variants. The maximum model utilizes all approximation components from $\mathbf{V}^j$ for the prediction of each detail component $\mathbf{W}_i^j$. The restricted model only employs a subset of components from $\mathbf{V}^j$ during the prediction to preserve the component scalability of the original DWT. The Fast variant applies a spatial sub-sampling of the approximation components only for the regression operation computation. Finally, as scenes captured by the same hyperspectral sensor may have similar statistical relationships among their components, the exogenous variant computes the regression coefficients only once, for a giving training scene, and employs these predictions for other scenes captured by the same sensor. This regression coefficients operation is performed off-line, not on-board, saving significant computational resources and execution time. This is a convenient strategy for those scenes corpora with a very large number of spectral components z.

III. Proposed Low-Complexity Coding technique

A. RWA with CCSDS-123-AC

Up to now, the RWA output signals $\mathbf{V}^J$ and $\mathbf{R}^{1\leq j\leq J}$ are losslessly entropy encoded with JPEG 2000 bit-plane and entropy coder machinery [3], [7]. Here we substitute JPEG 2000 by CCSDS-123-AC [12], which is a coding technique based on the predictor of CCSDS-123 followed by a lightweight contextual arithmetic encoder. The performance of this approach is discussed next.

All the results reported in this letter have been obtained from 27 hyperspectral scenes[1] captured by three different sensors: the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), the Infrared Atmospheric Sounding Interferometer (IASI) and the Atmospheric Infrared Sounder (AIRS). These instruments store scenes with a bit-depth of 16 bpppc. Table I depicts the dimensional information of these scenes together with the sensor's names abbreviations, the RWA regression model or variant and the CCSDS-123-AC's predictor configurations used per each sensor. The average order-0 entropy is also disclosed.

Regarding the RWA configurations: maximum regression model is applied for UA and CA scenes and exogenous variant for UI, CI and AG scenes (Table I). Though the maximum model attains better prediction, it also needs an important amount of side information to recover the original signal. For instance, UI, CI and AG sensors need, respectively, 0.918, 0.939 and 1.249 bpppc for the side information if maximum variant is employed. The expensive storage of the side information yields to a more competitive result if using the exogenous variant. The scenes UI-1, CI-1 and AG-129 have been employed only for the training procedure. Results are therefore not provided for these three scenes.

Leaning on results of [12] and after conducting an extensive evaluation, experimental results are produced for the CCSDS-123 predictor's parameters reported in Table I -columns 6 and 7-.

[1]The scenes are available at http://cwe.ccsds.org/sls/docs/sls-dc/123.0-B-Info/TestData.

Table I: SUMMARY OF DATA EMPLOYED FOR THE EXPERIMENTAL RESULTS. FROM LEFT TO RIGHT: THE SENSOR NAME TOGETHER WITH THE DIMENSIONS OF THEIR RECORDED SCENES, SENSOR'S NAMES ABBREVIATIONS, USED SCENES, SCENES AVERAGE ORDER-0 ENTROPY, RWA REGRESSION MODEL OR VARIANT, AND MODE AND LOCAL SUM USED BY THE CCSDS-123-AC PREDICTOR. z IS THE NUMBER OF SPECTRAL COMPONENTS, y IS THE HEIGHT AND x IS THE WIDTH.

Sensors & Dimensions	Abbreviation	Scenes	Order-0 entropy	RWA Model or Variant	Predictor Mode	Predictor local sum
Uncalibrated AVIRIS z=224, y=512, x=680	UA	Yellowstone, sc: 00, 03, 10, 11, 18	12.16	Maximum	Neighbor Oriented	Full Mode
Calibrated AVIRIS z=224, y=512, x=680	CA	Yellowstone, sc: 00, 03, 10, 11, 18	9.77	Maximum	Neighbor Oriented	Full Mode
Uncalibrated IASI Level 0 z=8359 y=1528 x=60	UI	L0 1: 20091007093900Z L0 2: 20091007143900Z L0 3: 2010031905030OS6 L0 4: 20120718075700Z	8.12	Exogenous	Neighbor Oriented	Full Mode
Calibrated IASI Level 1 z=8461 y=1530 x=60	CI	L1 1: 20130816230553Z L1 2: 20130817004753Z L1 3: 20130817041457Z L1 4: 20130817055657Z L1 5: 20130817073857Z	12.89	Exogenous	Neighbor Oriented	Full Mode
AIRS Gran z=1501, y=135, x=90	AG	sc: 9, 16, 60, 126, 129, 151, 182, 193	11.39	Exogenous	Neighbor Oriented	Reduced Mode

Table II reports average bit-rates for RWA at 1, 5 and the highest number of decomposition levels followed by JPEG 2000 standard (RWA+JPEG2000) and CCSDS-123-AC (RWA+CCSDS-123-AC). RWA+CCSDS-123-AC coding gain values with respect to RWA+JPEG2000 are also disclosed in column 6. It is worth noting that, after RWA computation, two very different types of data must be processed, the approximations $\mathbf{V}^j$ and the residuals $\mathbf{R}^{1\leq j}$. As in JPEG 2000 each component is encoded independently, the approximations and residuals can be considered a single signal to be encoded. However with CCSDS-123 and after computing the RWA, this consideration does not hold, since, to estimate data from a certain component, the CCSDS-123 predictor can employ information from other components or even decomposition levels. This leads to an incorrect prediction that may penalize the encoding performance. Consequently, for CCSDS-123-AC, each RWA decomposed signal (approximation and details) is separately encoded. In this case, the best results per each sensor are enhanced in bold. From these results we can see that: 1) RWA+CCSDS-123-AC almost always outperforms RWA+JPEG2000 at all levels; 2) the best coding performance for RWA+JPEG2000 is always obtained when the highest number of decomposition levels is applied; and 3) RWA+CCSDS-123-AC at the highest decomposition level does not always yield the lowest bit-rates.

Table II: AVERAGE BIT-RATES (IN BPPPC) WHEN ENCODING DIFFERENT SCENES CORPORA (TABLE I) WITH JPEG 2000 AND CCSDS-123-AC AFTER RWA AT 1, 5 AND THE HIGHEST NUMBER OF DECOMPOSITION LEVELS. P_{R} CORRESPONDS TO THE NUMBER OF PREVIOUS COMPONENTS THAT THE CCSDS-123 PREDICTOR USES WHEN ESTIMATING THE RWA RESIDUAL COMPONENTS. THE BEST RESULTS PER EACH SENSOR ARE ENHANCED IN BOLD.

	RWA levels	RWA + JPEG2000	RWA + CCSDS-123-AC $P_{\mathrm{R}}=0$	RWA + CCSDS-123-AC $P_{\mathrm{R}}=3$	Coding gains
UA	1	7.86	5.79	5.79	2.07
	5	5.87	5.71	**5.68**	0.19
	8	5.83	5.72	5.69	0.14
CA	1	5.27	3.64	3.69	1.63
	5	3.56	**3.52**	3.63	0.04
	8	3.52	3.53	3.63	-0.01
UI	1	3.86	2.66	2.77	1.20
	5	2.54	**2.51**	2.75	0.03
	14	2.44	**2.51**	2.77	-0.07
CI	1	8.17	6.54	6.41	1.76
	5	6.60	6.38	6.24	0.36
	14	6.46	6.37	**6.23**	0.23
AG	1	5.62	4.15	4.23	1.47
	5	4.25	**3.99**	4.16	0.26
	11	4.18	4.01	4.17	0.17

B. Multi-Level Split RWA

As seen above, fixing the number of RWA levels is not the best strategy when CCSDS-123-AC is harnessed. In order to obtain the best coding performance, we automatically select the adequate number of decomposition levels for RWA+CCSDS-123-AC, and we name this proposal as MLS-RWA+CCSDS-123-AC.

Fig. 1 renders our proposed coding technique when applying RWA for any regression model or variant. The main insight of MLS-RWA is to process a new RWA decomposition level over $\mathbf{V}^j$ only if the sum of bit-rates of approximations, residuals and regression coefficients does not exceed the bit-rate required to losslessly compress the approximations of the lower level, i.e., $\mathrm{BR}(\mathbf{V}^j) + \mathrm{BR}(\mathbf{R}^j) + \mathrm{BR}(\boldsymbol{\beta}^j) < \mathrm{BR}(\mathbf{V}^{j-1})$. Iteratively, further RWA decomposition levels are applied until no coding gain exists. The code-streams from the last applied level are discarded and the process stops. In the figure, CS and the BR refer respectively to the code-stream term and bit-rate operation. $\mathrm{BR}(\mathbf{V}^0)$ has been previously obtained after encoding $\mathbf{V}^0$ with CCSDS-123-AC. We use LZMA [14] as entropy coder for encoding the regression coefficients. However, another lossless coding technique could be employed. All the residual and regression coefficients code-streams, and the approximation coefficients code-stream at the highest decomposition level that satisfy the condition are stored together into the final code-stream. For exogenous variant, the regression coefficients would have been fixed beforehand and would not be stored into the final code-stream.

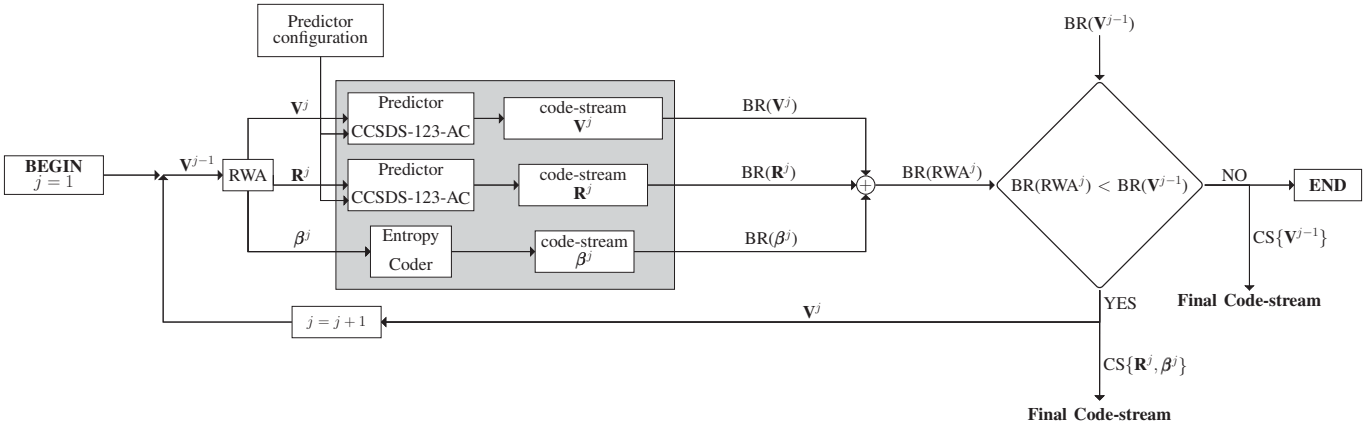


Fig. 1: Proposed coding scheme.

Table III: AVERAGE BIT-RATES (IN BPPPC) FOR PCA+JPEG2000, RWA+JPEG2000, M-CALIC, CCSDS-123, CCSDS-123-AC AND MLS-RWA+CCSDS-123-AC FOR SCENES OF TABLE I. THE BEST CODING PERFORMANCES OF OUR PROPOSAL ARE BOLD ENHANCED. THE CODING GAINS WITH RESPECT TO OUR BEST PROPOSAL'S OUTCOMES ARE ALSO INCLUDED IN PARENTHESES. P_{R} CORRESPONDS TO THE NUMBER OF PREVIOUS COMPONENTS EMPLOYED BY CCSDS-123-AC'S PREDICTOR STAGE TO ESTIMATE THE RWA RESIDUAL COMPONENTS.

	MLS-RWA levels	**PCA+JPEG2000**	**RWA+JPEG2000**	**M-CALIC**	**CCSDS-123**	**CCSDS-123-AC**	**MLS-RWA+CCSDS-123-AC**	
							$P_{\mathrm{R}}=0$	$P_{\mathrm{R}}=3$
UA	4	5.81 (0.14)	5.83 (0.16)	6.06 (0.39)	5.98 (0.31)	5.92 (0.25)	5.69	**5.67**
CA	4	3.69 (0.17)	3.52 (0.00)	3.86 (0.34)	3.72 (0.20)	3.76 (0.24)	**3.52**	3.62
UI	6	—	2.44 (-0.07)	2.88 (0.37)	2.88 (0.37)	2.82 (0.31)	**2.51**	2.75
CI	7	—	6.46 (0.23)	6.88 (0.65)	6.59 (0.36)	6.53 (0.30)	6.36	**6.23**
AG	4 (sc:16) 5 (sc:126,193) 6 (sc:9,60,151,182)	7.69 (3.70)	4.18 (0.19)	4.34 (0.35)	4.33 (0.34)	4.29 (0.30)	**3.99**	4.14
Average		5.73 (1.35)	4.48 (0.10)	4.80 (0.42)	4.70 (0.32)	4.66 (0.28)	4.38	

MLS-RWA decreases the computational cost by reducing the number of RWA levels to apply and by replacing JPEG 2000 with CCSDS-123-AC lower complexity encoder. By applying a lower number of RWA decomposition levels, we decrease not only the number of floating-point operations, but also the number of regression coefficients that have to be computed on-board and stored as side information. In this work, neither spatial sub-sampling nor a spectral components selection are proposed to alleviate the complexity of the regression operation block as explained in [3] and [7], although MLS-RWA could apply them too. Also, MLS-RWA could be combined with any other coding system.

IV. DISCUSSION AND RESULTS

Table III reports the lossless coding performance in average bit-rate for PCA+JPEG2000, RWA+JPEG2000, CCSDS-123, M-CALIC, CCSDS-123-AC coding techniques -which are competitive coding techniques-, and our approach MLS-RWA+CCSDS-123-AC. For CCSDS-123 and CCSDS-123-AC coding techniques, 3 previous components are used in the prediction process. The entropy encoder used for CCSDS-123 is the sample adaptive, which is the entropy coder that attains higher lossless compression ratios. For CCSDS-123-AC and when encoding approximation coefficients with our approach, the number of prediction components is set to 3 too, because, as with the original image components, the approximation components are still highly correlated in the spectral domain. For MLS-RWA+CCSDS-123-AC, the best performance of the predictor is achieved with different configurations depending on the sensor, as higher or lower statistical relationships can still remain within the residual components after RWA computation. Therefore, and only for our proposal, we report outcomes when 0 and 3 previous components are employed for predicting the residuals. The best coding performance is enhanced in bold. Our proposal's best coding gains with respect to each coding technique are reported within parenthesis. The last row reports the average of each coding technique; MLS-RWA+CCSDS-123-AC average value is computed only using the best results.

Comparing with coding techniques based on transforms, such as PCA+JPEG2000 and RWA+JPEG2000, we can observe that our proposal yields significant lower bit-rates for all the sensors, except for CA and UI scenes when applying RWA+JPEG2000. For CA scenes, no improvement is achieved, while an average loss of only 0.07 bpppc is produced for UI sensor. As mentioned, MLS-RWA+CCSDS-123-AC entails also a decrement of the computational cost. For CA and UI, our proposal only computes 6 and 14 RWA decomposition levels, letting aside 130 and 14 components for which neither a Haar-DWT nor a regression operation to estimate the details need be performed. On average, code-stream size reductions of, respectively, 23.56% and 2.23% between

our proposal and PCA+JPEG2000 and RWA+JPEG2000 are achieved. Results for PCA+JPEG2000 are not provided for UI and CI scenes due to their large number of spectral components, which makes unfeasible the PCA computation, as the covariance matrix and eigenvalues calculation, the matrix factorization and the amount of side information become prohibitive. For AG sensor, PCA's average side information requires 3.71 bpppc, yielding a final bit-rate of 7.69 bpppc.

M-CALIC and CCSDS-123 achieve quite competitive results at low computational complexity too, for instance, CCSDS-123 is, on average, about 0.1 bpppc worse than CCSDS-123-AC. However, our low complexity proposal MLS-RWA+CCSDS-123-AC improves, on average, by about 8.75%, 6.80%, and 6.00% as compared to M-CALIC, CCSDS-123 and CCSDS-123-AC, respectively.

Comparing with recently published coding techniques based on conventional recursive least-squares (and adaptations thereof) for uncalibrated and calibrated AVIRIS scenes, the best results reported in [10] are, respectively, 5.57 and 3.29 bpppc, while the best results reported in [11] are, respectively, 5.55 and 3.31 bpppc. These results are, respectively, about 0.1 and 0.2 bpppc better than those of MLS-RWA+CCSDS-123-AC, which is reasonable due to the trade-off between coding performance and computational complexity of [10] and [11].

V. Conclusions

This paper introduces a lossless coding technique that provides superior coding performance than state-of-the-art coding techniques at low computational complexity. It is based on Regression Wavelet Analysis (RWA) followed by a recently presented Lightweight Contextual Arithmetic encoder prepended by CCSDS-123 prediction (CCSDS-123-AC). RWA removes the spectral redundancy by computing a Discrete Wavelet Transform and, in sequence, a light regression operation to estimate the detail coefficients from the approximation coefficients.

Different to what happens when coding RWA code-streams with JPEG 2000, applying the highest number of RWA decomposition levels with CCSDS-123-AC does not imply the best coding performance. To provide lower bit-rates, we developed a smart strategy to properly select the number of RWA decomposition levels that affords the highest coding gain, named Multi-Level Split RWA (MLS-RWA). By decreasing the number of levels and by replacing JPEG 2000 by a light arithmetic coder, our proposal also reduces the computational cost of the original RWA+JPEG200 approach.

Extensive experimental results over 27 hyperspectral scenes from 5 corpus sensors have been performed, indicating that MLS-RWA followed by CCSDS-123-AC outperforms CCSDS-123.0, M-CALIC, PCA+JPEG2000, CCSDS-123-AC and RWA+JPEG2000, and provides significant average coding gains of 0.32, 0.42, 1.35, 0.28 and 0.10 bits-per-pixel-per-component (bpppc), respectively, for lossless compression. Our approach yields very competitive results even when RWA is computed only for a single decomposition level. In this case, an average coding gain of at least 1.74 bpppc over RWA+JPEG2000 is obtained.

References

[1] I. T. Jolliffe, "Principal Component Analysis," *Springer Verlag, Berlin, Germany,* pp. 487, 2002.

[2] I. Blanes and J. Serra-Sagristà, "Pairwise Orthogonal Transform for Spectral Image Coding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no 3, pp. 961-972, Oct. 11, 2010, ISSN: 0196-2892, DOI: 10.1109/TGRS.2010.2071880.

[3] N. Amrani, J. Serra-Sagristà, V. Laparra, M. W. Marcellin and J. Malo, "Regression wavelet analysis for lossless coding of remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no 9, pp. 5616-5627, Jun. 8, 2016, ISSN: 0196-2892, DOI: 10.1109/TGRS.2016.2569485.

[4] D. Taubman and M. W. Marcellin, "JPEG 2000: Image compression fundamentals, practices and standards," *Springer, New York*, 2001.

[5] E. Magli, G. Olmo and E. Quacchio. "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC," *IEEE Geoscience and Remote Sensing Letters,* vol. 1, no 1, pp. 21-25, Feb. 19, 2004, ISSN: 1545-598X, DOI: 10.1109/LGRS.2003.822312.

[6] Consultative Committee for Space Data Systems (CCSDS), 'Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1," *CCSDS. Blue Book*, https://public.ccsds.org/Pubs/123x0b1ec1.pdf, May 2012.

[7] S. Álvarez-Cortés, N. Amrani and J. Serra-Sagristà, "Low complexity regression wavelet analysis variants for hyperspectral data lossless compression," *International Journal of Remote Sensing*, Taylor & Francis, pp. 1–30, Sep. 19, 2017, DOI: 10.1080/01431161.2017.1375617.

[8] J. Song, Z. Zhang and X. Chen, "Lossless compression of hyperspectral imagery via RLS filter," *Electronics Letters,* vol. 49, no 16, pp. 992-994, Aug. 1, 2013. ISSN 0013-5194. ISSN: 0013-5194, DOI: 10.1049/el.2013.1315.

[9] J. Wu, W. Kong, J. Mielikainen and B. Huang, "Lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers," *IEEE Signal Processing Letters,* vol. 22, no 2, pp. 2194-2198, Jun. 10, 2015. ISSN: 1558-2361, DOI: 10.1109/LSP.2015.2443913.

[10] F. Gao and S. Guo, "Lossless compression of hyperspectral images using conventional recursive least-squares predictor with adaptive prediction bands," *Journal of Applied Remote Sensing,* vol. 10, no 1, pp. 015010, Feb. 2016. DOI: 10.1117/1.JRS.10.015010.

[11] A. C. Karaca and M. K. Güllü, "Lossless hyperspectral image compression using bimodal conventional recursive least-squares," *Remote Sensing Letters*, vol. 9, no 1 pp. 31-40, Jan. 2018. DOI: 10.1080/2150704X.2017.1375612.

[12] J. Bartrina-Rapesta, I. Blanes, F. Aulí-Llinàs, J. Serra-Sagristà, V. Sànchez and M. W. Marcellin, "A lightweight contextual arithmetic coder for on-board remote sensing data compression," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no 8, pp. 4825-4835, May 30, 2017, ISSN: 0196-2892, DOI: 10.1109/TGRS.2017.2701837.

[13] J. Nocedal and S. J. Wright, "Least-Squares Problems," *Numerical optimization*, Springer, pp. 245-269, 2006.

[14] I. Pavlov, "Lzma sdk (software development kit)", https://www.7-zip.org/sdk.html, 2007.

Chapter 5

Regression wavelet analysis for near-lossless remote sensing data compression

Regression Wavelet Analysis for Near-lossless Remote Sensing Data Compression

Sara Álvarez-Cortés, Joan Serra-Sagristà, *Senior Member, IEEE*, Joan Bartrina-Rapesta, Michael Marcellin, *Fellow, IEEE*

***Abstract*—Regression Wavelet Analysis (RWA) is one of the current state-of-the-art lossless compression techniques for remote sensing data. This paper presents the first regression-based near-lossless compression method. It is built upon RWA, a quantizer, and a feedback loop to compensate the quantization error. Our near-lossless RWA (NLRWA) proposal can be followed by any entropy coding technique. Here, NLRWA is coupled with a bitplane-based coder that supports progressive decoding. This successfully enables gradual quality refinement and lossless and near-lossless recovery. A smart strategy for selecting the NLRWA quantization steps is also included. Experimental results show that the proposed scheme outperforms state-of-the-art lossless and near-lossless compression methods in terms of compression ratios and quality retrieval.**

***Index Terms*—lossless and near-lossless compression, progressive lossy-to-lossless/near-lossless coding, regression wavelet analysis, remote sensing data compression.**

I. Introduction

SATELLITES carry on-board hyperspectral sensors that collect enormous volumes of data, with large spectral and spatial resolutions. Recording this information places excessive demands on bandwidth and on on-board storage capacity, meaning that part of the data could go uncaptured, or that part of the acquired data could be immediately discarded without further processing. Data compression has proven to be a convenient means to mitigate these issues, and meet the requirements of space missions. Different remote sensing data compression techniques provide lossless, lossy, and/or near-lossless recovery.

Lossless coding ensures perfect reconstruction at the price of low compression ratios. Multiband Context-based Adaptive Lossless Image Coding (M-CALIC) [1] is one of the most renowned methods. It is a context-based adaptive system that uses a nonlinear predictor. In 2012, the Consultive Committee for Space Data Systems (CCSDS) proposed the standard CCSDS-123.0-B-1 [2], which is formed by a predictor, a mapper function, and an entropy encoder. It entails minor computational cost and exploits the redundancy within 3-D spatial and spectral neighborhoods of pixels. During the last decade, techniques based upon recursive least squares methods have been presented [3]–[7]. Among them, Regression Wavelet Analysis (RWA) [3] yields the best coding performance trade-off concerning compression ratio and computational cost.

Lossy techniques enable high compression ratios at the expense of allowing loss in decoding. Commonly, lossy pipelines apply quantization prior to an entropy encoder, and a rate control allocation stage afterwards. Lossy wavelet-based techniques, such as JPEG 2000 [8], are well-known for attaining excellent performance in terms of mean squared error (MSE). Unfortunately, it does not provide any guarantees on the error incurred by individual pixels. Other major lossy contributions [9]–[11] extend the CCSDS-123.0-B-1 framework. The fast and lightweight rate control algorithm of Valsesia et al. [9] achieves comparable or better coding performance than [10] and [11], while decreasing the computational complexity.

Near-lossless compression aims at higher compression ratios than lossless methods by allowing some loss of fidelity in reconstruction. They bound the l_∞-norm -equivalently, the peak absolute error (PAE) or maximum absolute distortion (MAD)- via setting an error tolerance value Λ. This user-specified parameter sets the maximum admissible absolute error so that PAE $\leq \Lambda$ provides a guaranteed bound on the error incurred by individual pixels. Near-lossless compression is used in remote sensing applications such as appraisals of climate changes, natural resources and disasters, and also for farming and military purposes.

Near-lossless techniques can be classified into: prediction-based coding followed by quantization; and two-stage near-lossless coders.

Prediction-based followed by quantization techniques compute first a prediction of a pixel's value from previously encoded pixels. They provide near-lossless compression by introducing a quantization feedback loop and by including the corresponding reconstruction function in the coder. M-CALIC provides near-lossless compression and is one of the most relevant techniques in this category. Other prominent techniques are the two near-lossless adaptations of the lossless compression standard CCSDS-123.0-B-1, henceforth referred to as NLCCSDS-123 [12], and CCSDS-123-AC [13]. Both NLCCSDS-123 and CCSDS-123-AC rely on the predictor and mapper of CCSDS-123.0-B-1. CCSDS-123-AC includes a lightweight contextual arithmetic encoder that defines a context model and computes the probabilities that will be used by a fixed-length arithmetic encoder. CCSDS-123-AC improves the performance of both NLCCSDS-123 and M-

Manuscript received April 25, 2019. This work was supported in part by the Spanish Ministry of Economy and Competitiveness (MINECO) and by the European Regional Development Fund (FEDER) under Grant RTI2018-095287-B-I00, by the Catalan Government under Grant 2017SGR-463 and by Universitat Autònoma de Barcelona under Grant UAB-PIF-472/2015.

S. Álvarez-Cortés, J. Serra-Sagristà and J. Bartrina-Rapesta are with the Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona, Spain (e-mail: salvarez@deic.uab.cat). M. W. Marcellin is with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721-0104, USA.

CALIC, thus becoming the state-of-the-art in this compression modality.

Two-stage near-lossless coders generally include a first stage that generates a lossy reconstructed scene (image), whereas the second stage quantizes the difference between the reconstructed and the original scene and finally encodes the quantized signal with an entropy encoder. The performance of this strategy strongly depends on the distortion introduced in the lossy stage. Methods such as [14] and [15] do not include any selection criterion to determine the lossy bitrate. Other methods such as [16], [17] and [18] achieve competitive coding performance after using computationally expensive iterative approaches to identify the optimal lossy bitrate. The embedded two-stage near-lossless coder [18] yields the state-of-the-art compression performance in this category.

This paper presents a low-complexity approach that constitutes the first near-lossless technique based on regression in a pyramidal multiresolution scheme. It is based on RWA and 1) yields the same lossless performance as the original RWA; 2) employs a novel and low-cost strategy to select the quantization steps for near-lossless reconstruction; and 3) can be followed by any entropy coder. Here, we report results when employing JPEG 2000. This allows progressive lossy-to-lossless/near-lossless transmission, while minimizing the error propagation and optimizing both signal-to-noise ratio (SNR) and PAE performance.

The paper is organized as follows: Section II introduces our novel near-lossless scheme. Section III describes the mathematical derivation that allows us to control the PAE, and puts forward a smart criterion for selecting the quantization steps. Section IV presents experimental results and provides comparison to other state-of-the-art techniques. Finally, Section V brings forward our conclusions.

II. Near-lossless Regression Wavelet Analysis

Regression Wavelet Analysis [3] exploits the correlation of a scene in the spectral dimension. It is composed of two sequential blocks: a simple integer Haar Discrete Wavelet Transform (DWT) followed by a regression operation. This second block is performed through an Ordinary Least-Square (OLS) method which predicts the Haar wavelet details from the Haar approximations of the same decomposition level. The two blocks are repeated pyramidally, level by level, from the first to the highest possible decomposition level $L = \lceil log_2(z) \rceil$, where z denotes the number of spectral components of the original scene.

As explained in detail below, our near-lossless RWA (NL-RWA) adaptation begins by applying Haar-DWT at the highest decomposition level (first operation). The approximation and detail components at this level constitute the input signal for the regression (second operation). After the regression/prediction computation, the difference between the original Haar wavelet details and their predictions is obtained. The result is known as the residuals. The residuals are then quantized (third operation) and then dequantized and passed through one level of inverse Haar wavelet transform (fourth operation). The resulting approximations form the input to the linear regression for the next lower level. Computation proceeds from the second to the fourth operation iteratively from level $L-1$ down to level 1. A detailed explanation is provided next.

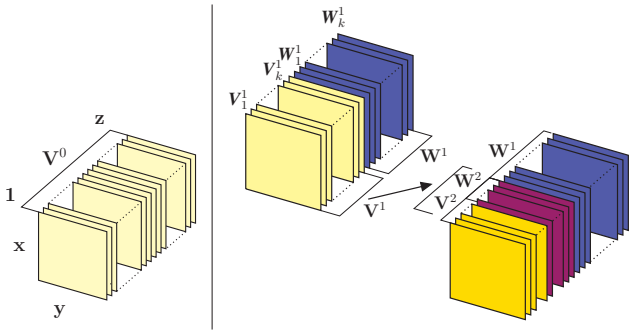


Figure 1: *Left:* original scene $\mathbf{V}^0$ with z spectral components, and $x \cdot y$ spatial samples. *Right:* first and second DWT decomposition levels, i.e., respectively $\mathbf{DWT}(\mathbf{V}^0, 1) = \left(\mathbf{V}^1, \mathbf{W}^1\right)$, and $\mathbf{DWT}(\mathbf{V}^1, 1) = \left(\mathbf{V}^2, \mathbf{W}^2\right)$ or, equivalently, $\mathbf{DWT}(\mathbf{V}^0, 2) = \left(\mathbf{V}^2, \mathbf{W}^2, \mathbf{W}^1\right)$, where $k = z \cdot 2^{-1}$.

A. First operation: Integer Haar Discrete Wavelet Transform

Let us consider a scene $\mathbf{V}^0 \in \mathbb{R}^{m \times z}$ with z spectral components and $m = x \cdot y$ spatial samples, where z is a power of two (a suitable boundary handling procedure is used otherwise), $\mathbf{V}^0 = \left[\mathbf{V}^0(1), \ldots, \mathbf{V}^0(z)\right]$ and $\mathbf{V}^0(i) = \mathbf{V}^0_i \in \mathbb{R}^{m \times 1}$. An integer Discrete Wavelet Transform (DWT) decomposition level on the original scene is denoted as

$$\mathbf{DWT}(\mathbf{V}^0, 1) = \left(\mathbf{V}^1, \mathbf{W}^1\right). \tag{1}$$

Here, $\mathbf{V}^1 \in \mathbb{R}^{m \times (z \cdot 2^{-1})}$ and $\mathbf{W}^1 \in \mathbb{R}^{m \times (z \cdot 2^{-1})}$ refer, respectively, to the half-resolution DWT approximation and detail components at the first decomposition level. A second decomposition level can be computed on the approximation components $\mathbf{V}^1$, maintaining the details $\mathbf{W}^1$ unchanged. After the second decomposition level, the transformed scene is composed of the approximations $\mathbf{V}^2$ at the second level and the details from the first $\mathbf{W}^1$ and second $\mathbf{W}^2$ levels. This process is iteratively applied until level L. The application of L decomposition levels to the original scene $\mathbf{V}^0$ is denoted as

$$\mathbf{DWT}(\mathbf{V}^0, L) = \left(\mathbf{V}^L, (\mathbf{W}^j)^{1 \le j \le L}\right). \tag{2}$$

See Fig. 1 for a graphical explanation.

In order to secure perfect reconstruction, a lifting scheme [19] is employed. The Haar-DWT that considers this scheme corresponds to the Haar S-Transform, and its forward equations at level j are

$$\textit{Forward:} \begin{cases} \mathbf{W}^j_i = \mathbf{V}^{j-1}_{2i} - \mathbf{V}^{j-1}_{2i-1} \\ \mathbf{V}^j_i = \mathbf{V}^{j-1}_{2i-1} + \lfloor \frac{1}{2} \mathbf{W}^j_i \rfloor, \end{cases} \tag{3}$$

where $i \in \mathbb{I} = \{1, .., z \cdot 2^{-j}\}$ corresponds to the spectral component.

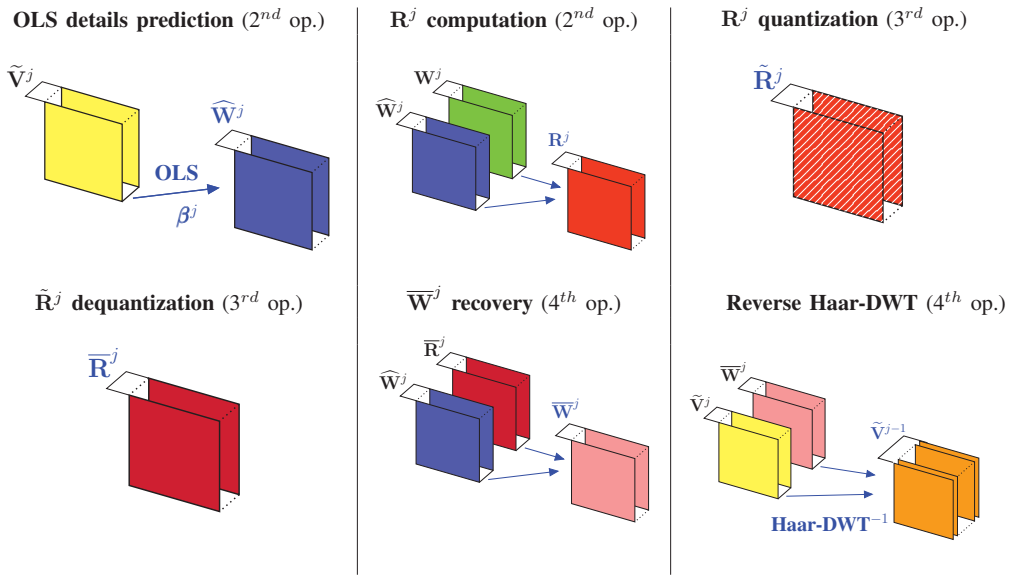


Figure 2: From *left* to *right* and from *top* to *bottom*, the NLRWA coding process is rendered for the decomposition level j and after the Haar S-Transform application, i.e., from the second operation to the fourth. All these stages are computed by the coder, which obtains the details prediction signal $\widehat{\mathbf{W}}^j$ in the second operation and employs it in the fourth operation.

B. Second operation: Ordinary Least-Squares Method

An Ordinary Least-Squares [20] problem (OLS) is solved to obtain the prediction of the details at level j, $\widehat{\mathbf{W}}^j$. It consists of a regression operation that minimizes the sum of the squares of the distances between the original and the predicted details:

$$\underset{\boldsymbol{\beta}^j}{\operatorname{argmin}} \|\mathbf{W}^j - \widehat{\mathbf{W}}^j\|_2. \tag{4}$$

The OLS is solved by applying a regression model. Two regression models are considered here: Maximum and Parsimonious.

The *Maximum* model [3] delivers the most accurate predictions, but uses all the $k{=}z{\cdot}2^{-j}$ approximation components at each level j to predict each detail component. The predicted detail component i at level j, $\widehat{\mathbf{W}}_i^j$, is computed as

$$\widehat{\mathbf{W}}_i^j = \beta_{i,0}^j + \beta_{i,1}^j \mathbf{V}_1^j + \cdots + \beta_{i,k}^j \mathbf{V}_k^j. \tag{5}$$

The regression coefficients $\boldsymbol{\beta}^j$ need to be stored and transmitted as side information (SI), which amounts to $\frac{z^2}{3}(1-\frac{1}{2^{2j}}) + z(1-\frac{1}{2^j})$ [3].

The *Parsimonious* model [21] does not employ all the $z{\cdot}2^{-j}$ approximation components at level j to obtain the regression coefficients, but, at most, $2r+1$. r is a natural number and specifies the number of previous and subsequent spectral neighbors of the component i considered by the OLS operation. The SI size is thus smaller. Specifically, the number of $\boldsymbol{\beta}_i^j$ involved in each prediction equates to $\min\{(2r+2), (z{\cdot}2^{-j}+1)\}$.

For both models, the regression coefficients $\boldsymbol{\beta}^j$ are computed for each individual scene.

Next, the prediction residuals are computed as

$$\mathbf{R}^j = \mathbf{W}^j - \mathit{round}(\widehat{\mathbf{W}}^j). \tag{6}$$

The predictions $\widehat{\mathbf{W}}^j$ are rounded to operate with integer values.

C. Third operation: USDZ Quantization

Each component i of the residuals at each level j, $\mathbf{R}_i^j$, is quantized with a Uniform Scalar Dead-Zone Quantizer (USDZQ), which delivers a symmetric behavior around 0. Let $\mathbf{c}_i^j$ be a coefficient of component i of the residuals at level j. The quantized coefficient is obtained as follows:

$$\tilde{\mathbf{c}}_i^j = \mathrm{USDZQ}(\mathbf{c}_i^j) = \mathrm{sign}(\mathbf{c}_i^j) \left\lfloor \frac{|\mathbf{c}_i^j|}{\Delta_i^j} \right\rfloor. \tag{7}$$

Δ_i^j refers to the quantization step for component $\mathbf{R}_i^j$, and *sign* is a function that extracts the sign value of coefficient $\mathbf{c}_i^j$.

Analogously, the dequantized coefficient can be described as $\overline{\mathbf{c}}_i^j$. The dequantization is computed as

$$\overline{\mathbf{c}}_i^j = \mathrm{USDZQ}^{-1}(\tilde{\mathbf{c}}_i^j) = \Delta_i^j \ \tilde{\mathbf{c}}_i^j. \tag{8}$$

Let $\overline{\mathbf{R}}_i^j$ be the dequantized residual component i at level j and let $\overline{\mathbf{R}}^j$ be the set of dequantized residual components at level j.

D. Fourth operation: Reconstruction of Details

Let the same quantization step be applied for all the components within a decomposition level j, i.e., $\forall i,\ \Delta_i^j = \Delta^j$. If it is greater than 1, the dequantized residual components will contain errors due to the quantization process.

Since only the dequantized residuals $\overline{\mathbf{R}}_i^j$ will be available in the decoder, the reconstructed details $\overline{\mathbf{W}}_i^j$ are computed using $\overline{\mathbf{R}}_i^j$ as

$$\overline{\mathbf{W}}^j = \overline{\mathbf{R}}^j + \mathit{round}(\widehat{\mathbf{W}}^j). \tag{9}$$

The reconstructed details $\overline{\mathbf{W}}^j$ are used both in the encoder and decoder as input signal to the inverse wavelet transform

to derive the approximation components at the next Haar-DWT level $j-1$,

$$\widetilde{\mathbf{V}}^{j-1} = \mathbf{DWT}^{-1}\big((\widetilde{\mathbf{V}}^{j}, \overline{\mathbf{W}}^{j}), 1\big). \tag{10}$$

$\widetilde{\mathbf{V}}^{L-1\leq j\leq 1}$ refers to the reconstructed approximation components of each decomposition level j. These components differ from the original approximations because of the quantization error.

Following the NLRWA notation, the inverse Haar S-Transform equations are:

$$\left\{ \begin{array}{lcl} \widetilde{\mathbf{V}}^{j-1}_{2i-1} & = & \widetilde{\mathbf{V}}^{j}_{i} - \lfloor \frac{1}{2}\overline{\mathbf{W}}^{j}_{i} \rfloor \\ \widetilde{\mathbf{V}}^{j-1}_{2i} & = & \overline{\mathbf{W}}^{j}_{i} + \widetilde{\mathbf{V}}^{j-1}_{2i-1}. \end{array} \right. \tag{11}$$

$\widetilde{\mathbf{V}}^{j-1}$ is fed back into the system to produce the regression coefficients at level $j-1$ (Eq. 4). Fig. 2 graphically represents the whole NLRWA procedure for level j.

III. Reconstruction Error Control

This section introduces how NLRWA controls the peak absolute error. In addition, a smart strategy to select the quantization steps is proposed. By using this criterion, competitive compression ratios and scene quality retrieval are obtained.

A. Peak Absolute Error Restriction

In this section, we show that NLRWA can control the largest absolute error to a tolerance value Λ. Let $\mathrm{v}^0_{p,b}$ represent the pixel of the scene $\mathbf{V}^0$ located at position (p, b), where $1 \leq p \leq m$ and $1 \leq b \leq z$, p corresponds to the spatial sample and b to the spectral component (band) location. Denote the reconstructed scene after applying NLRWA by $\widetilde{\mathbf{V}}^0$, and let $\widetilde{\mathrm{v}}^0_{p,b}$ be the pixel of the recovered scene at the same location. Then, the error is limited to

$$\Lambda \geq \mathrm{PAE} = ||\mathbf{V}^0 - \widetilde{\mathbf{V}}^0||_\infty = \max_{p,b} |\mathrm{v}^0_{p,b} - \widetilde{\mathrm{v}}^0_{p,b}|. \tag{12}$$

Lemma 1. *For NLRWA, Λ is equal to*

$$\Lambda = \sum_{j=1}^{L} \Big\lfloor \frac{1}{2}\ \Delta^j \Big\rfloor. \tag{13}$$

This equation depicts the largest possible cumulative error introduced by the quantization stage in the final reconstruction.

Proof: let the error in the quantized residual be denoted by $\epsilon_{\mathrm{R}}{}^j_i = \overline{\mathbf{R}}^j_i - \mathbf{R}^j_i$. Examination of the USDZQ reveals that $-(\Delta^j - 1) \leq \epsilon_{\mathrm{R}}{}^j_i \leq (\Delta^j - 1)$. Thus,

$$|\epsilon_{\mathrm{R}}{}^j_{\max}| = \max_i |\epsilon_{\mathrm{R}}{}^j_i| = \max_i |\overline{\mathbf{R}}^j_i - \mathbf{R}^j_i| \leq \Delta^j - 1. \tag{14}$$

From Eq. 6, the recovered details at level j (Eq. 9) are then $\overline{\mathbf{W}}^j_i = \mathbf{R}^j_i + \epsilon_{\mathrm{R}}{}^j_i + \mathit{round}(\widehat{\mathbf{W}}^j_i) = \mathbf{W}^j_i + \epsilon_{\mathrm{R}}{}^j_i$.

In the reverse Haar S-Transform, for odd indexed components (Eq. 11), the approximations are reconstructed as $\widetilde{\mathbf{V}}^{j-1}_{2i-1} = \widetilde{\mathbf{V}}^j_i - \Big\lfloor \frac{1}{2}\big(\mathbf{W}^j_i + \epsilon_{\mathrm{R}}{}^j_i\big) \Big\rfloor =$

$$\left\{ \begin{array}{ll} \widetilde{\mathbf{V}}^j_i - \Big(\Big\lfloor \frac{1}{2}\mathbf{W}^j_i \Big\rfloor + \Big\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \Big\rfloor + 1\Big), & \text{if } \mathbf{W}^j_i \text{ and } \epsilon_{\mathrm{R}}{}^j_i \text{ are odd,} \\ \widetilde{\mathbf{V}}^j_i - \Big(\Big\lfloor \frac{1}{2}\mathbf{W}^j_i \Big\rfloor + \Big\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \Big\rfloor\Big), & \text{otherwise.} \end{array} \right. \tag{15}$$

Comparing the original approximation components with the reconstructed approximation components, for odd components, the produced error is

$$\delta^{j-1}_{\mathrm{2i\text{-}1}} = \left\{ \begin{array}{ll} -\Big\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \Big\rfloor - 1, & \text{if } \mathbf{W}^j_i \text{ and } \epsilon_{\mathrm{R}}{}^j_i \text{ are odd,} \\ -\Big\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \Big\rfloor, & \text{otherwise.} \end{array} \right. \tag{16}$$

A similar analysis for even indexed components yields

$$\delta^{j-1}_{\mathrm{2i}} = \left\{ \begin{array}{ll} \Big\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \Big\rfloor + 1, & \text{if } \mathbf{W}^j_i \text{ is even and } \epsilon_{\mathrm{R}}{}^j_i \text{ is odd,} \\ \Big\lfloor \frac{1}{2}\epsilon_{\mathrm{R}}{}^j_i \Big\rfloor, & \text{otherwise.} \end{array} \right. \tag{17}$$

Now, the largest distortion in the reconstructed approximation components happens when the error in the reconstructed residuals is the largest at all levels j. In such a case, for even Δ^j values, $\epsilon_{\mathrm{R}}{}^j_{\max}$ will be odd, and vice versa. In summary, the highest possible absolute error can be expressed as

$$|\delta^{j-1}_{\max}| = \max_i |\delta^{j-1}_i| \leq \left\{ \begin{array}{ll} \Big\lfloor \frac{1}{2}(\Delta^j - 1) \Big\rfloor + 1, & \text{if } \Delta^j \text{ even,} \\ \Big\lfloor \frac{1}{2}(\Delta^j - 1) \Big\rfloor, & \text{if } \Delta^j \text{ odd.} \end{array} \right. \tag{18}$$

The maximum possible error in the residuals at level j contributes to the error as

$$\lambda^j = \Big\lfloor \frac{1}{2}\ \Delta^j \Big\rfloor. \tag{19}$$

By induction, and after applying NLRWA iteratively from level L down to level 1, the PAE in the reconstructed scene is limited by the error tolerance value, derived as

$$\Lambda = \sum_{j=1}^{L} \lambda^j = \sum_{j=1}^{L} \Big\lfloor \frac{1}{2}\ \Delta^j \Big\rfloor. \qquad \blacksquare$$

We note that other integer DWT could be used instead of Haar wavelet transform in the first operation of NLRWA [22]. For such a case, Eq. 13 should be adapted accordingly.

B. Quantization Steps Selection Criterion

The number of combinations of quantization step sizes that fulfill Eq. 13 is given by

$$\mathrm{N} = \sum_{m=1}^{\min\{\Lambda, L\}} 2^m \binom{L}{m} \binom{\Lambda - 1}{m - 1} \tag{20}$$

(see [23]), which depends on the number of the highest decomposition level L and on the error tolerance value Λ. This combinatorial number grows rapidly as L and Λ increase, such

that assessing the coding performance for every combination becomes unattainable.

Now, considering Eq. 13, odd quantization steps yield better performance than even quantization steps. If only odd quantization steps are considered, the possible combinations are reduced to (see [23])

$$\mathrm{N}_{\mathrm{odd}\Delta} = \binom{\Lambda + L - 1}{L - 1}. \tag{21}$$

Despite this reduction, exhaustive search over every possible combination is still prohibitive for reasonable values of L and Λ.

Rather, a heuristic selection criteria is proposed. It prioritizes the introduction of distortion into the residuals of the lowest decomposition levels (the less significant components in reconstruction), achieving therefore higher compression ratios and preserving better the signal's quality retrieval. Recall that the quantization steps of the residuals at level j are the same ($\Delta_i^j = \Delta^j,\ \forall i$). Given Λ fixed by the user, an odd value is assigned to the quantization step at level j according to:

$$\Delta^j = 2\left\lfloor \frac{\Lambda + 1}{2^j} + \frac{1}{2} \right\rfloor - 1. \tag{22}$$

As an example, for a scene where L=8 wavelet decomposition levels have been applied, the quantization steps for Λ=10, $[\mathbf{R}^8, \mathbf{R}^7, \mathbf{R}^6, \mathbf{R}^5, \mathbf{R}^4, \mathbf{R}^3, \mathbf{R}^2, \mathbf{R}^1]$ are [1,1,1,1,3,3,7,11]; for Λ=25, [1,1,1,3,5,7,13,27]; for Λ=50, [1,1,3,5,7,13,27,51].

The proposed strategy may not necessarily secure the best rate-distortion results. However, it attains a very competitive coding performance, is independent of the processed scene, and is computationally efficient.

IV. Experimental Results

Experimental results of our embedded coding framework, NLRWA coupled with JPEG 2000, are reported in comparison to state-of-the-art prediction-based followed by quantization techniques, in comparison to the state-of-the-art two-stage near-lossless coding technique, and in comparison to the best performing rate control algorithm that enables lossy and near-lossless recovery. Scenes used in the experiments are available at [24].

A. Coding Pipeline

Our proposed coding method applies NLRWA through a Matlab implementation, and then JPEG 2000 compresses the NLRWA transformed data through Kakadu software. Although our framework uses JPEG 2000, other coders could be used [25]. Here, JPEG 2000 is employed mostly because of its scalability capability, competitive lossy and lossless performance, and capacity of providing progressive refinement of the scene's quality retrieval.

To obtain smooth and steady increasing rate-distortion curves with JPEG 2000, the predictive weighting scheme (PWS) [26] is applied. It attributes pyramidal weights according to the significance of the NLRWA spectral components in the reconstruction [27].

For our NLRWA approach, Maximum or Parsimonious regression model is selected depending on the number of spectral components of the original scene: Maximum model for scenes where $L \leq 8$, and Parsimonious model otherwise. Notwithstanding, other models, such as Restricted, or variants, e.g., Fast-Maximum or Fast-Restricted, can be selected. For a more detailed description, see [22].

The side information for NLRWA can be encoded by any entropy coder. Results are reported when applying LZMA [28].

B. Prediction-based followed by Quantization Coders Comparison

Table I presents an extended study of NLRWA + JPEG 2000 coding performance in comparison with CCSDS-123-AC, NLCCSDS-123 and M-CALIC, for different Λ values.

Experimental results are presented for 26 scenes from different hyper- and ultraspectral sensors with a bit-depth of 16 bpppc (bits-per-pixel-per-component): calibrated and uncalibrated AVIRIS (referred to as CA and UA, respectively), calibrated Hyperion (CH), filtered uncalibrated Hyperion (FUH), uncalibrated IASI (UI), and AIRS Gran (AG). The scenes from the uncalibrated Hyperion (UH) corpus are filtered [29], [30] to remove the streaking artifacts along one of the spatial dimensions [31]. These artifacts appear because of the pushbroom sensor nature, and they should be dealt with for a better scene information assessment and visualization.

NLRWA applies the Maximum regression model for AVIRIS and Hyperion scenes and Parsimonious for the rest. Here, NLCCSDS-123 employs its sample-adaptive encoder for lossless and near-lossless coding, when $\Lambda = 1$, and block-adaptive for the rest of near-lossless results. The predictor of both NLCCSDS-123 and CCSDS-123-AC considers 3 previous spectral components, and their mode and local sum are selected depending on the corpus. The neighbor oriented mode has been used for AVIRIS, AIRS and IASI sensors, and the column oriented predictor configuration for Hyperion sensor. The local sum applied in prediction is full mode only when processing the AVIRIS and IASI corpus scenes. The reduced mode is set for the rest.

The C++ M-CALIC software implementation does not handle transformed signals with more than 15 bpppc. Due to the dynamic range extension, no results can be obtained for the CH corpus. In contrast, NLRWA induces a dynamic range expansion of only, at most, 1 bit in the detail components. This avoids severe inconveniences in systems that support only a limited bit-depth [32].

Table I reports coding performance measured in bitrate and quality. For lossless coding ($\Lambda = 0$), our approach is superior. That is, it yields the lowest rate of any of the compared methods. For near-lossless coding, the best coding performance in terms of bitrate is provided by CCSDS-123-AC, while our approach is competitive for most error tolerance values Λ and for most sensors. Concerning quality, as measured by SNR, our approach is always the best performing, with increasingly larger differences as Λ grows.

Fig. 3 depicts a crop of a component of a CA scene. For a fair visual comparison, first we encode the scene for different

Table I: Average lossless ($\Lambda = 0$) and near-lossless ($\Lambda > 0$) compression results for several prediction-based techniques and our proposal, NLRWA + JPEG 2000. For all of them, USDZQ has been used for near-lossless. The best results are enhanced in bold. The coding gains of our method with respect to the other three techniques are included within parentheses. A positive difference means that our approach is better.

					Average bitrates (bpppc)				**Average SNR (dB)**		
Sensor abbreviation	Number of scenes	Number of spectral components	Average order-0 entropy (bpppc)	Λ values	CCSDS-123-AC	NLCCSDS-123	M-CALIC	NLRWA + JPEG 2000	CCSDS-123-AC & NLCCSDS-123	M-CALIC	NLRWA + JPEG 2000
CA	5	224	9.77	$\Lambda = 0$	3.66 (0.13)	3.73 (0.20)	4.03 (0.50)	**3.53**	—	—	—
				$\Lambda = 1$	**2.45** (-0.03)	2.54 (0.06)	2.87 (0.39)	2.48	59.29 (0.37)	59.29 (0.37)	**59.66**
				$\Lambda = 10$	**0.58** (-0.13)	0.94 (0.23)	0.88 (0.17)	0.71	42.94 (7.20)	41.38 (8.76)	**50.14**
				$\Lambda = 20$	**0.35** (-0.13)	0.72 (0.24)	0.53 (0.05)	0.48	37.46 (10.35)	35.38 (12.43)	**47.81**
				$\Lambda = 30$	**0.26** (-0.13)	0.63 (0.23)	0.40 (0.00)	0.40	34.15 (12.32)	31.84 (14.63)	**46.47**
UA	3	224	12.13	$\Lambda = 0$	5.87 (0.05)	5.95 (0.13)	6.13 (0.31)	**5.82**	—	—	—
				$\Lambda = 1$	**4.80** (-0.12)	4.89 (0.03)	5.05 (0.13)	4.92	74.73 (0.06)	74.74 (0.05)	**74.79**
				$\Lambda = 10$	**1.96** (-0.56)	2.24 (-0.28)	2.22 (-0.30)	2.52	57.09 (4.19)	57.07 (4.21)	**61.28**
				$\Lambda = 20$	**1.14** (-0.41)	1.46 (-0.09)	1.40 (-0.15)	1.55	51.95 (4.93)	51.65 (5.23)	**56.88**
				$\Lambda = 30$	**0.83** (-0.27)	1.18 (0.08)	1.05 (-0.05)	1.10	48.90 (5.90)	48.20 (6.60)	**54.80**
CH	3	242	9.50	$\Lambda = 0$	**5.36** (-0.02)	5.61 (0.23)	—	5.38	—	—	—
				$\Lambda = 1$	**4.50** (-0.14)	4.75 (0.11)	—	4.64	65.63 (0.78)	—	**66.41**
				$\Lambda = 10$	**2.25** (-0.52)	2.54 (-0.23)	—	2.77	47.87 (4.82)	—	**52.69**
				$\Lambda = 20$	**1.44** (-0.57)	1.77 (-0.24)	—	2.01	42.42 (5.58)	—	**48.00**
				$\Lambda = 30$	**1.05** (-0.53)	1.40 (-0.18)	—	1.58	39.27 (5.97)	—	**45.24**
FUH	3	242	9.42	$\Lambda = 0$	4.26 (0.21)	4.37 (0.32)	4.28 (0.23)	**4.05**	—	—	—
				$\Lambda = 1$	3.08 (0.05)	3.19 (0.16)	3.10 (0.07)	**3.03**	59.67 (0.20)	59.67 (0.20)	**59.87**
				$\Lambda = 10$	0.73 (-0.20)	1.14 (0.21)	**0.61** (-0.32)	0.93	42.73 (5.84)	42.53 (6.07)	**48.57**
				$\Lambda = 20$	0.38 (-0.12)	0.79 (0.29)	**0.32** (-0.18)	0.50	37.38 (8.46)	36.24 (9.60)	**45.84**
				$\Lambda = 30$	0.26 (-0.07)	0.66 (0.33)	**0.24** (-0.09)	0.33	34.15 (10.35)	32.23 (12.27)	**44.50**
UI	4	8359	8.12	$\Lambda = 0$	2.82 (0.27)	2.89 (0.34)	2.94 (0.39)	**2.55**	—	—	—
				$\Lambda = 1$	1.53 (0.05)	1.68 (0.20)	1.74 (0.26)	**1.48**	46.95 (0.65)	46.95 (0.65)	**47.60**
				$\Lambda = 10$	**0.13** (-0.02)	0.49 (0.34)	0.41 (0.26)	0.15	31.52 (9.06)	28.34 (12.24)	**40.58**
				$\Lambda = 20$	**0.06** (-0.01)	0.42 (0.35)	0.29 (0.22)	0.07	26.26 (13.38)	21.89 (17.75)	**39.64**
				$\Lambda = 30$	**0.03** (-0.03)	0.40 (0.34)	0.26 (0.20)	0.06	23.17 (16.26)	18.36 (21.07)	**39.43**
AG	8	1501	11.39	$\Lambda = 0$	4.25 (0.28)	4.31 (0.34)	4.38 (0.41)	**3.97**	—	—	—
				$\Lambda = 1$	3.08 (0.12)	3.13 (0.17)	3.21 (0.25)	**2.96**	70.73 (0.17)	70.73 (0.17)	**70.90**
				$\Lambda = 10$	**0.61** (-0.09)	0.98 (0.28)	0.70 (0.00)	0.70	54.16 (5.49)	53.75 (5.90)	**59.65**
				$\Lambda = 20$	**0.29** (0.03)	0.66 (0.34)	0.36 (0.04)	0.32	49.04 (8.59)	47.97 (9.66)	**57.63**
				$\Lambda = 30$	**0.20** (-0.04)	0.57 (0.33)	0.26 (0.02)	0.24	45.77 (11.32)	44.27 (12.82)	**57.09**

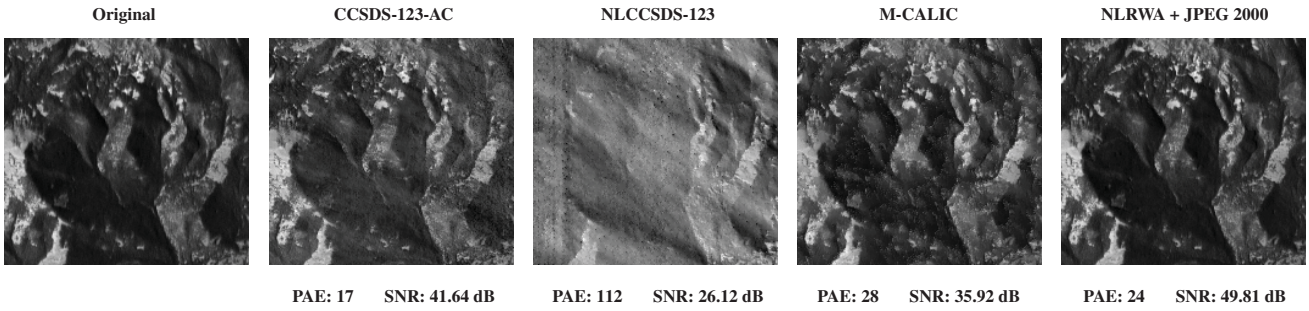


Figure 3: Crops of the spectral component 107 of the calibrated AVIRIS Yellowstone scene 18 (CA-Yellowstone sc18), and its reconstruction after applying CCSDS-123-AC, NLCCSDS-123, M-CALIC, and our proposal NLRWA + JPEG 2000. In all cases, the bitrate is about 0.5 bpppc.

error tolerance values and then we choose the encoded scene that requires a bitrate as close as possible to 0.5 bpppc. It can be noticed that the crop corresponding to NLRWA + JPEG 2000 conserves better the details and is less noisy.

C. Two-Stage Near-lossless Coders Comparison

Table II reports bitrates and quality (SNR) results for the 9 scenes presented in the original work of Beerten et al. [18] in comparison with our approach. UA12 refers to the UA sensor that has stored the scenes with a bit-depth of 12 bpppc.

Both [18] and our approach are embedded techniques that enable progressive lossy-to-lossless/near-lossless transmission. Concerning the quality of the reconstructed scenes, given any error tolerance value Λ, our proposal is always superior, as happened for the case of comparing against prediction-based coding techniques.

Concerning the bitrate performance, a cursory glance to Table II shows the good behaviour of [18] (with the exception of results for the CH corpus in the lossless regime). However, [18] applies a Karhunen-Loève-Transform (KLT) before JPEG 2000 lossy layer. KLT efficiently decorrelates the spectral dimension of a scene, but entails a high computational complexity and a non-negligible side information. When processing scenes with a very large number of spectral components, e.g., the scenes recorded by AIRS or IASI sensors, KLT results to be computationally untenable. Contrary to this, our proposal not only provides competitive bitrates at significantly lower computational complexity, but also benefits

Table II: Bitrate and SNR results of Fully Embedded Two-Stage Coder [18] in comparison with our proposal, NLRWA + JPEG 2000, at several Λ values. Again, bold font indicates highest coding performance.

	Bitrate (bpppc)								**SNR (dB)**					
	[18]				NLRWA + JPEG 2000				[18]			NLRWA + JPEG 2000		
Scene \ Λ	*0*	*1*	*5*	*32*	*0*	*1*	*5*	*32*	*1*	*5*	*32*	*1*	*5*	*32*
UA12-Hawaii	**2.45**	**1.00**	**0.17**	**0.07**	2.54	1.47	0.46	0.15	51.50	47.44	41.95	**53.07**	**48.19**	**44.22**
UA12-Maine	**2.61**	**1.16**	**0.23**	**0.07**	2.69	1.61	0.58	0.22	54.87	49.80	40.00	**56.46**	**51.22**	**46.41**
CA-Yellowstone sc00	3.76	**2.24**	**0.70**	**0.17**	**3.74**	2.70	1.30	0.42	59.04	50.57	42.77	**60.56**	**53.22**	**46.65**
UA-Yellowstone sc00	**5.95**	**4.37**	**2.54**	**0.53**	6.08	5.19	3.74	1.25	75.00	63.24	52.62	**76.30**	**67.18**	**55.57**
CH-Agricultural	6.11	**4.53**	**2.77**	**0.84**	**5.49**	4.76	3.60	1.61	65.43	53.83	42.23	**67.60**	**58.42**	**45.84**
CH-Coral Reef	5.80	**4.21**	**2.40**	**0.60**	**5.12**	4.37	3.19	1.31	61.06	49.87	38.52	**63.25**	**54.16**	**42.29**
CH-Urban	6.14	**4.55**	**2.80**	**0.88**	**5.50**	4.77	3.60	1.66	66.20	54.60	43.12	**68.37**	**59.21**	**46.76**
UH-Erta Ale	**4.54**	**3.00**	**1.26**	**0.11**	4.79	3.86	2.44	0.60	57.92	46.75	39.01	**59.25**	**50.42**	**39.65**
UH-Lake Monona	**4.64**	**3.08**	**1.37**	**0.16**	4.99	4.07	2.66	0.73	59.55	48.33	40.50	**60.88**	**51.97**	**40.74**

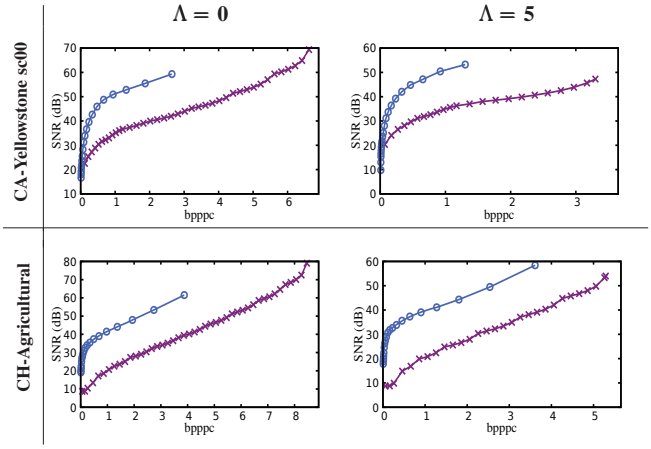


Figure 4: *Blue* with *circles* and *purple* with *crosses* curves represent the rate-distortion performance of, respectively, NLRWA and [18] with KLT replaced by Haar-DWT.

from its minor side information requirements, so that NLRWA is applicable also for scenes with more than 8,000 spectral components.

A fair comparison between NLRWA and the state-of-the-art method [18] in terms of computational complexity is provided in Fig. 4, where the rate-distortion curves for both techniques employ the Haar-DWT. In this case, our approach not only provides again better quality, but also better bitrate.

D. Lossy Rate Control-based Coders Comparison

Table III depicts bitrate and quality (SNR and PAE) values of our proposal and of the rate control presented by Valsesia et al. [9] for all the scenes presented in this last paper.

Notice that [9] only provides end-to-end rate-controlled outcomes given a prefixed target bitrate, while our approach, thanks to JPEG 2000's rate control algorithm, permits progressive decoding of a codestream. In this experiment, several Λ values have been tested in our approach, choosing those that provide the closest bitrate to that reported in [9].

NLRWA considers all (Maximum) or several neighbor (Parsimonious) spectral components per spatial sample as regressors in the prediction stage. Therefrom, it is very suited to apply when processing scenes with a large number of spectral components, i.e., for the CA Yellowstone scene 00 (CA-Yellowstone sc00), the AIRS Gran scene 09 (AG sc09) and

Table III: Bitrate, SNR and PAE results of the rate control presented by Valsesia et al. [9] and our near-lossless proposal. The mean of the squared Pearson correlation coefficients of each scene, $\overline{r}^2$, is included. z refers to the number of spectral components.

Scene	z	$\overline{r}^2$	[9]			NLRWA + JPEG 2000		
			Bitrate	SNR	PAE	Bitrate	SNR	PAE
CA-Yellowstone sc00	224	0.65	1.00	44.60	255	0.91	51.28	8
			2.00	57.79	25	1.73	55.37	3
			3.00	64.10	25	2.69	60.56	1
			4.00	71.13	3	3.74	∞	0
AG sc09	1501	0.65	1.00	53.86	18	0.96	60.40	7
			2.00	63.04	4	1.86	64.53	3
			2.99	67.28	3	2.87	70.15	1
			4.02	79.25	1	3.90	∞	0
CASI-T0477F06-NUC	72	0.44	1.00	41.69	49	0.96	45.80	24
			2.00	50.92	7	2.06	51.64	7
			3.00	57.64	4	3.05	56.56	3
			4.00	62.03	3	4.02	62.45	1
CRISM-sc167-NUC	545	0.40	1.00	44.14	48	1.02	41.07	39
			2.00	52.59	7	2.00	48.25	14
			3.00	59.37	3	2.97	54.43	6
			4.00	64.25	2	4.06	61.61	2
LANDSAT MOUNTAIN	6	0.53	1.00	27.13	10	1.00	21.20	28
			2.00	34.17	3	2.03	32.54	4
			3.00	39.37	3	3.03	38.74	1
			3.73	∞	0	4.06	∞	0
MODIS-MOD01DAY	14	0.65	1.00	38.70	230	1.11	28.27	491
			2.00	49.59	134	2.00	38.09	139
			3.00	58.66	17	2.99	46.11	52
			4.00	63.90	7	4.02	53.00	20

CASI-T0477F06-NUC. For them, NLRWA + JPEG 2000 is superior in PAE-bpppc and PAE-SNR performance. Conversely, LANDSAT MOUNTAIN and MODIS-MOD01DAY scenes have a low number of spectral components. This may lead to higher prediction errors, penalizing the performance of our approach. For these two scenes, the quantization steps criterion of Subsection III-B has been modified, allowing the assignment of large quantization step values to high decomposition levels.

The CRISM sensor is affected by common artifacts present in pushbroom sensors. Although a non uniformity calibration (NUC) has been applied for the CRISM scene, it contains groups of spectral components still strongly compromised, as shown by the low mean of the squared Pearson correlation coefficient (0.40). For this sensor, [9], whose predictor takes into consideration the correlation between adjacent spatial samples, outperforms our proposal.

V. Conclusion

This manuscript introduces NLRWA, the first near-lossless compression technique based on regression, in particular, on the pyramidal multiresolution regression wavelet analysis. It expands the state-of-the-art lossless compression technique RWA by introducing quantization and a feedback loop to

compensate the quantization error. We also provide a smart criterion, which is independent of the scene, to select a unique quantization steps combination. This criterion helps avoiding iterative computations while producing very competitive coding performances in terms of compression ratio and quality of the reconstruced scene. NLRWA can be followed by any entropy coder. Here, NLRWA is coupled with JPEG 2000, enabling progressive lossy-to-lossless/near-lossless decoding.

Experimental results indicate that NLRWA + JPEG 2000 considerably outperforms rate control-based algorithms in both bitrate and scene quality reconstruction for scenes with a large number of spectral components. When comparing with two-stage near-lossless coders, our approach always yields superior quality retrieval, and achieves competitive compression ratios at significantly lower computational cost. With respect to state-of-the-art prediction-based followed by quantization techniques such as CCSDS-123-AC, NLCCSDS-123 and M-CALIC, our approach NLRWA + JPEG 2000 always yields reconstructed scenes with the highest quality, and obtains outstanding compression ratios while offering progressive decoding and some degree of embeddedness.

REFERENCES

[1] E. Magli, G. Olmo, and E. Quacchio, "Optimized Onboard Lossless and Near-lossless Compression of Hyperspectral Data Using CALIC," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 1, pp. 21–25, Feb. 2004, DOI: 10.1109/LGRS.2003.822312.

[2] Consultative Committee for Space Data Systems (CCSDS), "Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1." *CCSDS. Blue Book*, 2012. https://public.ccsds.org/Pubs/123x0b1ec1.pdf

[3] N. Amrani, J. Serra-Sagristà, V. Laparra, M. W. Marcellin, and J. Malo, "Regression Wavelet Analysis for Lossless Coding of Remote-Sensing Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5616–5627, Jun. 2016, DOI: 10.1109/TGRS.2016.2569485.

[4] J. Song, Z. Zhang, and X. Chen, "Lossless Compression of Hyperspectral Imagery Via RLS Filter," *IEEE Electronics Letters*, vol. 49, no. 16, pp. 992–994, Aug. 2013, DOI: 10.1049/el.2013.1315.

[5] W. Jiaji, K. Wanqiu, M. Jarno, and H. Bormin, "Lossless Compression of Hyperspectral Imagery via Clustered Differential Pulse Code Modulation with Removal of Local Spectral Outliers," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2194–2198, Jun. 2015, DOI: 10.1109/LSP.2015.2443913.

[6] F. Gao, and S. Guo, "Lossless Compression of Hyperspectral Images Using Conventional Recursive Least-Squares Predictor with Adaptive Prediction Bands," *SPIE Journal of Applied Remote Sensing*, vol. 10, no. 1, pp. 015010, Feb. 2016, DOI: 10.1117/1.JRS.10.015010.

[7] J. Mielikainen, and B. Huang, "Lossless Compression of Hyperspectral Images Using Clustered Linear Prediction with Adaptive Prediction Length," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 6, pp. 1118–1121, Apr. 2012, DOI: 10.1109/LGRS.2012.2191531.

[8] D. S. Taubman, and M. W. Marcellin, "JPEG 2000 Image Compression Fundamentals, Standards and Practice," *Norwell, Massachusetts 02061 USA: Kluwer Academic Publishers*, 2002.

[9] D. Valsesia, and E. Magli, "Fast and Lightweight Rate Control for Onboard Predictive Coding of Hyperspectral Images,", *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 3, pp. 394–398, Jan. 2017, DOI: 10.1109/LGRS.2016.2644726.

[10] D. Valsesia, and E. Magli, "A Novel Rate Control Algorithm for Onboard Predictive Coding of Multispectral and Hyperspectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6341–6355, Jan. 2014, DOI: 10.1109/TGRS.2013.2296329.

[11] M. Conoscenti, R. Coppola, and E. Magli, "Constant SNR, Rate Control, and Entropy Coding for Predictive Lossy Hyperspectral Image Compression", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7431–7441, Sep. 2016, DOI: 10.1109/TGRS.2016.2603998.

[12] I. Blanes, E. Magli, and J. Serra-Sagristà, "A Tutorial on Image Compression for Optical Space Imaging Systems," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 3, pp. 8–26, Oct. 2014, DOI: 10.1109/MGRS.2014.2352465.

[13] J. Bartrina-Rapesta, I. Blanes, F. Aulí-Llinàs, J. Serra-Sagristà, V. Sanchez, and M. W. Marcellin, "A Lightweight Contextual Arithmetic Coder for On-Board Remote Sensing Data Compression", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4825–4835, May 2017, DOI: 10.1109/TGRS.2017.2701837.

[14] G. Carvajal, B. Penna, and E. Magli, "Unified Lossy and Near-lossless Hyperspectral Image Compression based on JPEG 2000," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 593–597, Nov. 2008, DOI: 10.1109/LGRS.2008.2000651.

[15] C. W. Chen, T. C. Lin, S. H. Chen, and T. K. Truong, "A Near-lossless Wavelet-based Compression Scheme for Satellite Images," in *Computer Science and Information Engineering, WRI World Congress on*, vol. 6, pp. 528–532, Jul. 2009, DOI: 10.1109/CSIE.2009.933.

[16] R. Ansari, N. Memon, and E. Ceran, "Near-lossless Image Compression Techniques", *Journal of Electronic Imaging*, vol. 7, no. 3, pp. 486–494, Jul. 1998, DOI: 10.1117/1.482591.

[17] X. Wu, and P. Bao, 'Near-lossless image compression by combining wavelets and CALIC", *IEEE Conference Record of the 31 Asilomar Conference on Signals, Systems and Computers (Cat. No. 97CB36136)*, vol. 2, pp. 1427–1431, Nov. 1997, DOI: 10.1109/ACSSC.1997.679139.

[18] J. Beerten, I. Blanes, and J. Serra-Sagristà, "A Fully Embedded Two-Stage Coder for Hyperspectral Near-lossless Compression", *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1775–1779, May 2015, DOI: 10.1109/LGRS.2015.2425548.

[19] A. Calderbank, I. Daubechies, W. Sweldens, and B. L. Yeo, "Wavelet Transforms that Map Integers to Integers", *Applied and computational harmonic analysis,* vol. 5, no. 3, pp. 332–369, Aug. 1996, DOI: 10.1006/acha.1997.0238.

[20] J. Nocedal, and S. J. Wright, "Least-Squares Problems," *Numerical optimization, Springer,* pp. 245–269, 2006, DOI: 10.1007/978-0-387-74759-0_329.

[21] N. Amrani, J. Serra-Sagristà, and M. W. Marcellin, "Low Complexity Prediction Model for Coding Remote-Sensing Data with Regression Wavelet Analysis," *In Data Compression Conference (DCC), IEEE*, p. 112–121, May 2017, DOI: 10.1109/DCC.2017.61.

[22] S. Álvarez-Cortés, N. Amrani, and J. Serra-Sagristà, "Low Complexity Regression Wavelet Analysis Variants for Hyperspectral Data Lossless Compression," *International Journal of Remote Sensing*, Taylor & Francis, p. 1–30, Sep. 19, 2017, DOI: 10.1080/01431161.2017.1375617.

[23] J. Serra-Sagristà, "Enumeration of Lattice Points in l1 Norm," *Information processing letters*, vol. 76, no. 1-2, pp. 39–44, Nov. 2000, DOI: 10.1016/S0020-0190(00)00119-8.

[24] Consultative Committee for Space Data Systems (CCSDS), "123.0-B-Info TestData," 2015. http://cwe.ccsds.org/sls/docs/SLS-DC/123.0-B-Info/TestData

[25] S. Álvarez-Cortés, J. Bartrina-Rapesta, and J. Serra-Sagristà, "Multilevel Split Regression Wavelet Analysis for Lossless Compression of Remote Sensing Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 10, pp. 15400–1544, Jul. 2018, DOI 10.1109/LGRS.2018.2850938.

[26] N. Amrani, J. Serra-Sagristà, M. Hernández-Cabronero, and M. Marcellin. 2016. "Regression Wavelet Analysis for Progressive-Lossy-to-Lossless Coding of Remote-Sensing Data." *Data Compression Conference DCC. IEEE.* pp. 121–130, Dec. 2016, DOI: 10.1109/DCC.2016.43.

[27] S. Álvarez-Cortés, N. Amrani, M. Hernández-Cabronero, and J. Serra-Sagristà, "Progressive Lossy-to-Lossless Coding of Hyperspectral Images through Regression Wavelet Analysis," *International Journal of Remote Sensing*, vol. 39, no. 7, pp. 2001–2021, Jul. 2017, DOI: 10.1080/01431161.2017.1343515.

[28] I. Pavlov, "Lzma sdk (software development kit)", https://www.7-zip.org/sdk.html, 2007.

[29] I. Blanes, and J. Serra-Sagristà, "Pairwise Orthogonal Transform for Spectral Image Coding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, p. 961–972, Oct. 2010, DOI: 10.1109/TGRS.2010.2071880.

[30] Consultative Committee for Space Data Systems (CCSDS). 2015. "Lossless Multispectral & Hyperspectral Image Compression CCSDS 120.2-G-1." *CCSDS. Green Book. https://public.ccsds.org/Pubs/120x2g1.pdf*

[31] U.S. Geological Survey and NASA, "Earth Observing 1, Hyperion Website."http://eo1.usgs.gov/hyperion.php.

[32] I. Blanes, M. Hernández-Cabronero, F. Aulí-Llinas, J. Serra-Sagristà, and M. W. Marcellin, "Isorange Pairwise Orthogonal Transform", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, p. 3361–3372, Dec. 2014, DOI: 10.1109/TGRS.2014.2374473.

Chapter 6

Results summary

In this chapter, we sum up the results of our proposed variations of Regression Wavelet Analysis (RWA) for lossless [29, 31] and progressive lossy-to-lossless [30] compression, and of our near-lossless RWA (NLRWA) proposal. For this last implementation, we provide image quality retrieval outcomes. Our proposals are thoroughly compared with state-of-the-art and widespread coding techniques: rKLT + JPEG 2000, Haar-DWT + JPEG 2000, NLCCSDS-123 [3], POT [14] + JPEG 2000, M-CALIC [15], CCSDS-123.0-B-1 [18], CCSDS-123-AC [19], RLS [20], C-DPCM-RLSO [21], CRLS and A-CRLS [22], C-DPCM-APL [23], the embedded two-stage near-lossless coder [32], and the fast and lightweight rate control version of Valsesia et al. [33]. We justify our outcomes through coding performance, computational complexity and signal measurements, provide lossy bitrate distributions per spectral component for RWA, and prove its feasibility and excellent performance for nowadays remote sensing applications, such as digital classification.

6.1 Datasets

Experimental results of this thesis are resumed in this chapter for different hyperspectral scenes that are available at [34]. They are recorded by several sensors: calibrated and uncalibrated Airborne Visible/Infrared Imaging Spectrometer -AVIRIS- (referred to as CA and UA, respectively), calibrated and uncalibrated satellite-borne

Table 6.1: Description of the sensors corpora used in experimentation. y and x refer respectively to the height and width of the scene.

Sensor denotation	**Number of spectral components**	**Number of spatial samples**	**Processed scenes**	**Avg. order-0 entropy (bpppc)**
CA	224	512 × 677	Yellowstone, sc: 00, 03, 10, 11, 18	9.77
UA	224	512 × 680	Yellowstone, sc: 00, 03, 10, 11, 18	12.16
UA12	224	512 × x	Hawaii (H) ($x = 614$), Maine (M) ($x = 680$)	8.82
UA-200	200	145 × 145	Indian Pines	11.79
UA-204	204	217 × 512	Salinas	11.59
CH	242	y × 256	Agriculture (A) ($y = 3129$) Coral Reef (CR) ($y = 3127$) Urban (U) ($y = 2905$)	9.50
UH	242	y × 256	Lake Monona (LM) ($y = 3176$) Erta Ale (EA) ($y = 3187$) Mt. St. Helens (MSH) ($y = 3242$)	9.57
FUH	242	y × 256	Lake Monona (LM) ($y = 3176$) Erta Ale (EA) ($y = 3187$) Mt. St. Helens (MSH) ($y = 3242$)	9.42
CI	8461	1530 × 60	L1 1: 20130816230553Z L1 2: 20130817004753Z L1 3: 20130817041457Z L1 4: 20130817055657Z L1 5: 20130817073857Z	12.89
UI	8359	1528 × 60	L0 1: 20091007093900Z L0 2: 20091007143900Z L0 3: 2010031905030OS6 L0 4: 20120718075700Z	8.12
AG	1501	135 × 90	sc: 9, 16, 60, 126, 129, 151, 182, 193	11.39
CASI	72	1225 × 406	CASI-T0477F06-NUC	10.65
CRISM	545	510 × 640	CRISM-sc167-NUC	10.54
Landsat	6	1024 × 1024	LANDSAT-MOUNTAIN	6.33
MODIS	14	2030 × 1354	MODIS-MOD01DAY	7.99

Hyperion (CH and UH, respectively), filtered uncalibrated Hyperion (FUH), calibrated and uncalibrated Infrared Atmospheric Sounding Interferometer -IASI- (CI and UI, respectively), Atmospheric Infrared Sounder -AIRS Gran- (AG), Compact Reconnaissance Imaging Spectrometer for Mars -CRISM-, Moderate Resolution Imaging Spectroradiometer -MODIS-, Compact Airborne Spectrographic Imager -CASI-, and a scene from the Landsat's multispectral sensor. The scenes are stored with a bit-depth of 16 bits-per-pixel-per-component (bpppc), with the exception of Hawaii, Maine, MODIS-MOD01$_{\text{DAY}}$, and the Landsat's scene; uncalibrated AVIRIS Hawaii and Maine scenes (UA12) and MODIS employ 12 bpppc, and the Landsat's scene is stored with 8 bpppc. Table 6.1 briefly describes all these scenes. In this table, the term NUC refers to Non Uniformity Calibration.

We depict results for the original UH corpus for a proper comparison with the original lossless RWA work. For our NLRWA approach, experiment results are also provided for this UH corpus after filtering [14, 35]. This procedure removes the streaking artifacts along one of the spatial dimensions [36]. These artifacts appear because of the pushbroom sensor nature, and they should be dealt with for a better scene information assessment and visualization.

To illustrate the influence image quality loss has over widely use applications, such as digital classification, we evaluate two AVIRIS images: Indian Pines and Salinas [37], stored with a bit-depth of 16 bpppc. Both images cover land with natural vegetation, bare soils, and vineyard fields. For the first image we consider 200 spectral reflectance bands, as 24 water absorption components are removed from the original 224 spectral bands. For Salinas, 20 water absorption spectral components are also discarded from again a total of 224. The abbreviation name of the sensor is, for each image, UA-200 and UA-204.

6.2 General pipeline

The two main blocks common in all of our coding systems are displayed in Figure 6.1. The first block applies a 1D transform along the spectral dimension. The second corresponds to the entropy encoder. Several 1D spectral transforms are assessed: reversible KLT (rKLT), Pairwise Orthogonal Transform (POT), DWTs/IWTs, RWA, and NLRWA. For all of them, results are explained when followed by JPEG 2000 as the second stage. For such case, a 2D spatial IWT 5/3 with 5 levels is applied in the lossless regime for POT, Haar-DWT, Restricted and Parsimonious models and their variants, and for RWA when transforming Hyperion scenes. This is because this configuration has been found to produce the best performance in compression ratio and energy compaction. Beyond that, experiments for RWA coupled with CCSDS-123-AC are also evaluated.

Figure 6.1: Pipeline of our coding systems: 1D spectral transform followed by JPEG 2000 or CCSDS-123-AC.

6.3 RWA: regression models and variants

In [2, 25, 26, 27] three regression models (Maximum, Parsimoniuos and Restricted), and two variants of the Maximum model (Fast and Exogenous) are explained. *Maximum* gives rise to the most accurate predictions by using all the wavelet approximations in each detail component estimation. This model generates $\frac{z^2}{3}(1-\frac{1}{2^{2j}})+z(1-\frac{1}{2^j})$ regression coefficients per decomposition level j, where z corresponds to the number of spectral components the original scene has. These coefficients are stored as side information (SI) to be able to reproduce the estimated details in decoding. Maximum's SI size and computational cost are reduced by the Parsimonious model. *Parsimonious* does not employ all the approximation components at level j to obtain the regression coefficients, but, at most, $2r+1$. r is the number of previous and subsequent

spectral neighbors of the component considered in prediction. The SI size is thus smaller, i.e., the number of regression coefficients involved in each prediction equates to $\min\{(2r+2),(z{\cdot}2^{-j}+1)\}$. Conversely, *Restricted* utilizes a unique approximation coefficient in a cube polynomial formulation, entailing the need to store $2z(1-\frac{1}{2^j})$ regression coefficients per level j.

Fast and Exogenous variants are developed in order to decrease the complexity and side information weight of the Maximum model, without barely penalizing its coding performance. *Fast* applies a spatial sub-sampling before each prediction. It alleviates the computational cost, but not the SI requirements, implying the storage of the same number of regression coefficients. On the other hand, *Exogenous* computes the Maximum model off-line to only one scene to extract the regression coefficients. This scene is used therefore as training image. The coefficients are later employed to process the rest of the scenes from the same corpus, avoiding thus assessing further predictions and any side information storage. Thanks to this, Exogenous reforms RWA into a data-independent method. To reduce even more the complexity, side information, and memory resources consumption of the Restricted and Parsimonious models, we assess the variants: Fast-Restricted, Fast-Parsimonious, Exogenous-Restricted and Exogenous-Parsimonious.

In the original proposal [2], the first operation is conducted by an integer version of Haar-DWT, the Haar S-Transform. Generally, Haar-DWT is suboptimal compacting the energy when compared to other more complex wavelet transforms such as the reversible IWT 5/3 and IWT 9/7M (integer version of the biorthogonal CDF 9/7). With the intention of improving the coding performance, in **Chapter 2** we assess the suitability of exchanging the integer Haar wavelet by these two more efficient IWTs. To this end, we change the Restricted regression operation from a cubic polynomial to a linear formulation. Originally, Restricted only used an approximation component of a decomposition level to predict the same component and level detail. In such a case, the second and third-order terms are also assessed to enhance the prediction accuracy. Instead of this, our regression definitions utilize the approximation components

required by each IWT to recover a coefficient of the next lower decomposition level. Notice that for both IWTs, a boundary procedure is applied to the coefficients that are at the border. Here, the regression in our new IWTs-based RWA (IWTs-RWA) coding techniques do not take into consideration this boundary process. In other words, always the same one-degree formulations are applied indiscriminately of the detail component that is estimated, unless the concerned approximations do not exist. This enhances the prediction accuracy for the components at the border, having a higher number of regressors than if the bounding procedure had been taken into account. For the following experimentations, JPEG 2000 encodes the transformed coefficients (scene components) after RWA application.

6.3.1 Component-scalability

Here, component-scalability refers to the number of transformed components or coefficients required to reconstruct a single scene component; the smallest the number, the largest the scalability of the transform. It depends on the regression model and on the wavelet filter used as the basis of the first operation block in the RWA pipeline.

Haar S-Transform employs two approximation coefficients in the forward transform for obtaining a high- (detail) and a low-pass (approximation) components. They are later utilized to revert the decomposition and obtain the mentioned approximations. Conversely, IWT 5/3 uses, respectively, 5 and 3 coefficients, and IWT 9/7M, respectively, 9 and 7 in the forward process. For the inverse transform, IWT 5/3 needs 4 and IWT 9/7M needs 8 approximations.

Concerning the regression model, Maximum has the worst component-scalability independently of the computed wavelet filter. This is because all the approximation components per decomposition level are required for predicting a particular detail component at the same level. Consequently, to recover a transform coefficient of a lower decomposition level, all the approximations of the current level are required. On the contrary, the formulation of Restricted changes according to the wavelet used

in the first operation block, as the degree of component-scalability consequently does. The original Restricted model employs the Haar S-Transform and one approximation component in each prediction. In its turns, the reverse Haar S-Transform uses the same regressor (approximation) and the detail from the same component and level for recovering an approximation coefficient of the lower level, preserving thus the highest scalability. For IWT 5/3 and IWT 9/7M, the recovery of the approximation components of the adjacent lower level depends also on several details from higher transform levels. Since the regression model in RWA selects the regressors from the approximations, in **Chapters 2** and **3** we have only evaluated the approximation components dependency. Consistently, if IWT 5/3 and IWT 9/7M are applied before regression, the OLS problem needs respectively 4 and 8 approximation components, slightly penalizing the component-scalability.

6.3.2 Computational complexity

With the aim of understanding the decrease in computational cost the new proposed variants attain, we include in **Chapter 2** an exhaustive study of the computational complexity of the different stages involved in the lossless RWA process (referred here as Haar-RWA). The analysis is performed by assessing the Floating-point operations (FLOPs). Then, the computational cost of different Haar-RWA and IWTs-RWA regression models and variants for different uncalibrated scenes is rendered in three bar graphs. For comparison purposes, this study is also conducted for other transforms such as rKLT, Haar S-Transform, IWT 5/3, and IWT 9/7M.

As mentioned, rKLT attains very competitive compression ratios. However, it is the most expensive spectral transform, specially when the number of spectral components increases. From our complexity survey, RWA Maximum model entails a computational cost significantly lower than rKLT, but is the second most expensive spectral transform. As expected, Restricted model has a much lower complexity when compared to Maximum. This is because only a small set of components are used as regressors. Fast variant considerably reduces the cost of their underlying regression

model, since the spatial sub-sampling alleviates the computation of least-squares parameters. Exogenous variant offers a slightly lower computational complexity than Fast, as the computation of the regression parameters is not performed. The wavelet used within the RWA pipeline affects the cost because of the number of operations involved in each transform. It is apparent that Haar-RWA has a lower cost than IWT 5/3-RWA, which, in turn, is cheaper than IWT 9/7M-RWA.

6.3.3 Execution time

As an extension of the computational complexity analysis, we also provide execution time when computing all the RWA regression models and variants on an Intel(R) Xeon(R) CPU E5520 @ 2.27GHz processor with a single thread. It is shown that the computation of the Exogenous-Maximum variant is usually the fastest. Contrary to expected, Exogenous-Restricted needs larger execution time. This is because the sequential selection of the approximations involved in each detail prediction prevents a faster matrix operation. For Maximum model and its variants, the wavelet filter used in its first operation block influences the complexity. For this model, Haar-RWA shows faster execution time than IWT 5/3-RWA and IWT 9/7M-RWA. For Restricted model and its variants, IWT 5/3-RWA yields the fastest execution time. On the other hand, IWT 9/7M-RWA is slower because it employs more coefficients for the wavelet decomposition and for the regression operation. Here, Haar-RWA is also slower due to the calculation of the second- and third- power terms.

6.3.4 Coding performance

The average bitrate (bpppc) for Maximum and Restricted regression models and their Fast and Exogenous variants are delivered in **Chapter 2**. Results are reported for six hyperspectral corpora, for different decomposition levels and underlying wavelet filters. Only average values are included due to all scenes within a particular corpus produce similar performance.

For Maximum and its variants very similar coding performance are obtained for

all the underlying wavelet transforms, i.e., it seems not to depend much on the used wavelet filter. On the contrary, for the Restricted model and its variants, the applied wavelet transform meaningfully affects the coding performance. For this model, IWT 9/7M-RWA almost always outperforms the other wavelet filters when processing AVIRIS and IASI corpora. Consequently, presented bitrates are also lower than those reported in [2]. For example, for the Yellowstone scenes recorded by the UA sensor, Restricted Haar-RWA at the highest decomposition level requires 6.44 bpppc, while IWT 9/7M-RWA under the same considerations needs 5.94 bpppc, producing an average coding gain of 0.50 bpppc.

For all the regression and wavelet configurations, applying higher number of wavelet decomposition levels improves the coding performance, even though the number of regressors at higher decomposition levels keeps decreasing. Indeed, Maximum Haar-RWA model at the highest level usually produces the lowest bitrates at a moderate computational complexity. For UH scenes the Restricted model is superior to Maximum in coding performance, computational complexity, and component-scalability. For instance, Maximum Haar-RWA at 1, 5, and 8 decomposition levels requires on average 0.16 bpppc more than Restricted Haar-RWA, and needs all the approximation components to recover a unique transform coefficient. In general, applying 5 wavelet levels yields very similar performance to when the largest number of levels is applied, in terms of both computational cost and compression ratio.

In general, Exogenous variant evaluated to any underlying regression model and with any wavelet filter achieves a very similar performance to those obtained by applying the optimal regression parameters on-line, entailing lower computational complexity and no side information to be stored/transmitted. Since the side information plays a relevant role in the coding performance for scenes with thousands of spectral components (e.g., for the IASI corpus, side information can suppose up to 0.9 bpppc), the Exogenous variant provides the best coding outcomes, as side information is not transmitted.

6.3.5 Signal measurements results

For a more detailed appraisal of the performance of RWA, we evaluate the evolutionary trend of several signal measurements for the same corpora. Results reveal that the application of any regression model and variant, for all the wavelet filters, entails very similar behavior, i.e., the average squared Pearson correlation coefficient, energy percentage, and Shannon entropy dramatically decrease after spectrally transforming the scenes at the highest number of wavelet decomposition levels. Indeed, they are significantly lower after transforming only one level, e.g., for UH corpus the energy percentage decreases to half for all the wavelet filter-based RWA. This value is even very small, 0.006%, after applying Maximum Haar-RWA at the highest level to UI corpora. For the rest of IWTs-RWA, the energy percentage is reduced to 0.01%, which is also very small.

The dynamic range, precision and entropy per spectral component are also assessed after applying RWA with the three wavelet transforms as first operation block. As rendered in **Chapter 2**, the distributions follow the same tendency. It is worth pointing out that RWA induces a dynamic range expansion of only, at most, 1 bit in the detail components. This prevents severe complications in systems that only work properly under a limited bit-depth.

6.3.6 Progressive lossy-to-lossless compression

In [25] a Predictive Weighting Scheme (PWS) is applied only for the Maximum regression model and its Exogenous variant, providing smooth increasing rate-distortion curves. In **Chapter 3** we illustrate the clear benefit of considering this PWS for other models and variants: Fast-Maximum, Fast-Restricted, and Exogenous-Restricted. After rendering the rate-distortion curves, we see that those obtained by the regression variants are very close to those provided by their underlying regression model for AVIRIS and Hyperion corpora. To apply the PWS gives rise to bitrate-SNR gains for all bitrates and for all corpora. Indeed, PWS-RWA is superior to rKLT at medium and high bitrates.

For learning more about the influence of utilizing these weights in the rate-distortion allocation system of JPEG 2000, we display the bitrate assigned to each spectral component after compression. This is accomplished by fixing different global target bitrates. For instance, for global target bitrates of 1 and 3 bpppc, the effects are clearly apparent, i.e., it is noticeable that considering the weights entails the assignment of larger bitrates to the most significant components in reconstruction. When the PWS is not employed, a more uniform bitrate distribution among all components is found. This is also reflected when plotting the difference in bpppc between applying or not the PWS. For the mentioned two global target bitrates, the curves depict positive differences for the first and most significant components, and negative differences for the rest.

6.3.7 Comparison to state-of-the-art techniques

Bitrate and execution time for several state-of-the-art techniques are provided for comparison purposes. For the AVIRIS scenes, the recursive least-squares-based methods C-DPCM-RLSO (order 85), A-CRLS, and C-DPCM-APL outperform the coding performance of RWA. However, they demand significant computational resources. Furthermore, the performance of RWA lies very close to rKLT for UA scenes with a difference of only 0.02 bpppc. Nonetheless, rKLT also attains higher computational complexity. RLS and CRLS obtain similar bitrates than RWA but are again more expensive too. Hence, Maximum and Exogenous-Maximum RWA implementations yield very competitive coding gains at a very low execution time. M-CALIC, CCSDS-123.0-B-1, Parsimonious of RWA with $r = 3$ usually achieve higher bitrates at a reasonable computational time.

Regarding results for CH corpus, Maximum RWA outperforms all the state-of-the-art techniques. On the contrary, for UH CCSDS-123.0-B-1 provides superior performance, but is slower than RWA. Notice that this standard exploits the 2D spatial domain benefiting from the streaking artifacts that strongly affect the UH corpora.

Although the Exogenous-Maximum variant is not the fastest when encoding IASI corpus, it procures the lowest bitrates at considerably low computational complexity, becoming the best performing technique for scenes with thousands of spectral bands, as already mentioned (Subsection 6.3.4).

6.4 Digital classification

Multi-class digital classification is today an attractive field of study. This is shown by the large and extended community of researchers who are currently working on this topic. Digital classification is applied on images recorded by a wide variety of sensors. More specifically, classification of remote sensing data is performed to recovered scenes that have been previously compressed (for a proper management of the downlink channel and storage capacities). As mentioned in the introduction of this thesis, the compression may be lossless, near-lossless and lossy. Lossy recovery of remote sensing data may present blurred regions or other compression artifacts. Thence, we examine how the loss in reconstruction impacts on the classification outcomes in **Chapter 3**. To this end, we utilize commonly-used classifiers, i.e., a Linear Discriminant Analysis (LDA) and linear Support Vector Machines (SVMs). Given a certain data, LDA concentrates more on modeling the differences between classes. Conversely, multi-class SVM focuses on supervised learning models. It applies as many binary SVMs as the number of labeled classes considered in the training process. Here, one versus all is the followed strategy. Finally, both classifiers are trained using 20% of the original image samples of each class.

For experimentation, we analyze results for the AVIRIS Indian Pines and Salinas scenes after being processed by different PLL coding schemes. These frameworks are: rKLT, POT, PWS-Haar-DWT, and PWS-RWA, all of them coupled with JPEG 2000, and the lossless/near-lossless technique M-CALIC. For PWS-RWA, the used wavelet and regression model correspond to Haar S-Transform and Maximum. Our analysis is carried out by measuring the micro-average F1-Score and hits percentage.

F1-Score describes the harmonic mean between precision and recall. We consider micro-averaged metrics for most reliable results, which are less penalized by the labeling imbalance. The hits percentage is defined as the total number of true positives with respect to all the classification outputs. The best value for these metrics is 100% and the worst is 0%.

We demonstrate that superior coding performance might not necessarily imply better classification results. Then, after rendering bitrate-classification performances curves, we see that for very low bitrates, i.e., for reconstructed images with a large distortion, a strongly transitory period may appear. This is less severe for PWS-RWA + JPEG 2000. Indeed, this technique generally overcomes the rest. In terms of classification, at low to medium bitrates, the same method achieves overall better classification. At high bitrates, the distortion is low, obtaining all coding techniques equivalent classification performance.

6.5 Entropy encoder: CCSDS-123-AC

All the publications in the literature that are related to RWA [2, 25, 26, 27, 29, 30] employ JPEG 2000 as entropy coder. JPEG 2000 is sometimes too complex for on-board computation. In such cases, utilizing other coders may alleviate the consumption of computational resources. On this basis, we assess RWA when coupled with the recently proposed and on-board amenable coder CCSDS-123-AC.

In **Chapter 4**, we point out that RWA followed by CCSDS-123-AC achieves lower mean bitrates when compared with the original RWA + JPEG 2000, even when only one decomposition level is applied, e.g., average gains of 2.07, 1.63, 1.20, 1.76, and 1.47 bpppc for, respectively, UA, CA, UI, CI, AG corpora. Coding RWA codestreams with JPEG 2000 usually attains its best coding performance when applying the highest number of RWA decomposition levels. In contrast, RWA at the highest level does not necessarily achieve the best compression ratio when encoded with CCSDS-123-AC. We propose a smart criterion for selecting the optimal number of levels. In other

words, the number that procures the highest coding gains. This strategy is named Multi-Level Split RWA (MLS-RWA). Its main insight is to process a new decomposition level only if the sum of bitrates of approximations, residuals and regression coefficients does not exceed the bitrate required to losslessly compress the approximations of the lower level. This procedure is iteratively applied until no coding gain exists. The codestreams from the last assessed level are discarded and the process stops.

MLS-RWA followed by CCSDS-123-AC outperforms CCSDS-123.0-B-1, M-CALIC, rKLT + JPEG 2000, CCSDS-123-AC and RWA + JPEG 2000, providing lower average bitrates, on average, between 0.10 and 1.35 bpppc. In fact, MLS-RWA + CCSDS-123-AC improves the coding performance of rKLT + JPEG 2000 by, on average, 3.70 bpppc for AG corpus.

6.6 Near-lossless regression wavelet analysis

RWA is a lossless transform that has never been proposed for near-lossless or lossy implementations. In **Chapter 5** we explain our lossless/near-lossless adaptation of RWA, named NLRWA. It introduces into the RWA framework a quantization step, and a feedback loop to compensate the quantization error. In this sense, NLRWA is capable of controlling the peak absolute error (PAE) via a user-specified parameter Λ. This parameter is known as maximum error tolerance value, and is necessary to determine the quantization steps value per NLRWA decomposition level.

NLRWA is followed by JPEG 2000 (NLRWA + JPEG 2000) to successfully enable progressive error minimization, lossless and near-lossless recovery, and some degree of embeddedness. Moreover, we propose a criterion to select a unique NLRWA quantization steps combination to avoid iteratively testing all the possible ones. Our methodology produces very competitive coding performances without penalizing the compression ratio, neither the recovered scene quality. To prove the competitive performance of both, we describe below an extended comparison with state-of-the-art

and widespread near-lossless techniques and lossy rate control-based coders.

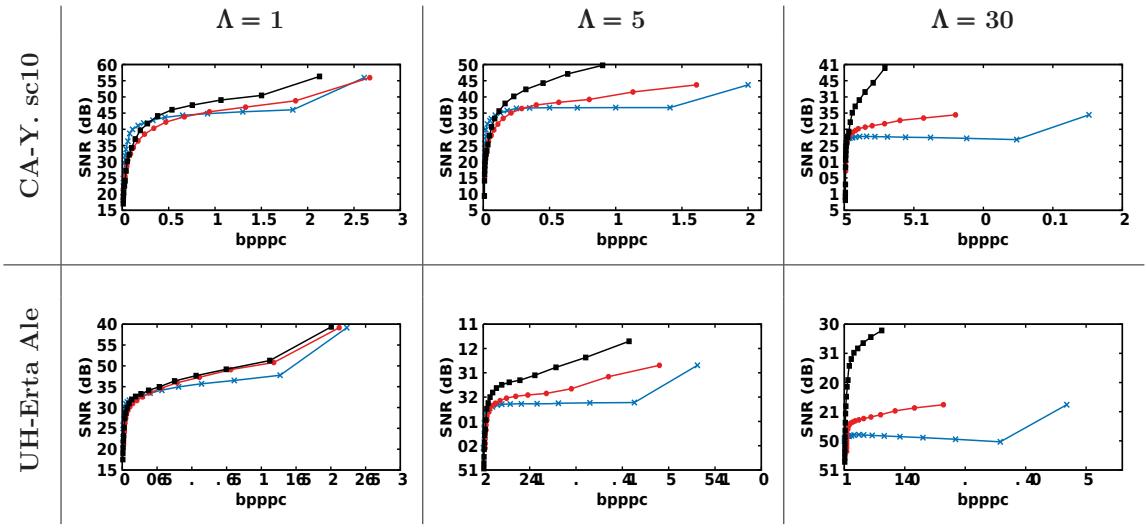


Figure 6.2: PLL rate-distortion curves for different Λ. *Black* with *squares*, *red* with *circles*, and *blue* with *crosses* curves represent results for, respectively, our proposal, Q-POT + JPEG 2000 and Q-rKLT + JPEG 2000.

6.6.1 Prequantization previous lossless compression

Prequantization methods followed by lossless coding are based on computing first a quantization on the original pixels' values. Sequentially, they encode the quantized signal with a lossless compression method. In **Chapter 5** no results for this type of near-lossless compression are described. This is because this strategy is well-known to perform poorly, specially when Λ is large [38].

To expand the experimental comparison, Figure 6.2 plots rate-distortion curves of NLRWA + JPEG 2000 together with progressive pipelines built upon prequantization and a lossless transform, i.e., here rKLT or POT. The transformed components compose the input signal for JPEG 2000. In this subsection, these techniques are called, respectively, as Q-rKLT + JPEG 2000 and Q-POT + JPEG 2000. They are not pure near-lossless methods, but an adaptation of the techniques presented in [30]. In the case of Q-rKLT + JPEG 2000, and for a fair comparison, a weighting scheme

is also applied. As expected and clearly concluded from the figure, for large Λ values, the coding performance of Q-rKLT/POT + JPEG 2000 are far worse than those of our proposal. Their bitrates are higher and the SNR values are lower. Indeed, the SNR results dramatically decrease when Λ increases. This behavior already begins to be apparent at very low Λ, e.g. from Λ equal to 5. For Λ equal to 30, NLRWA + JPEG 2000 attains the high SNR gains of 14.46 and 12.73 dB for, respectively, the CA Yellowstone scene 10 and the UH Erta Ale scene, and with respect to first, the rKLT-, and second, the POT-based systems. Our approach's average bitrate gains are also considerably high, 1.49 and 0.50 bpppc.

6.6.2 Prediction-based near-lossless compression

Prediction-based techniques followed by quantization and lossless coding estimate first a pixel's value from previous encoded pixels. For near-lossless compression they introduce a quantization feedback loop and the corresponding reconstruction function in the coder. Finally, they entropy-encode the quantized prediction error. One of the most popular prediction-based techniques is the near-lossless adaptation of the standard CCSDS-123.0-B-1, referred here to as NLCCSDS-123. Other prominent methods are M-CALIC and CCSDS-123-AC, which apart from lossless coding they can procure near-lossless compression too.

In **Chapter 5**, we compare our proposal with these near-lossless techniques for remote sensing data. In the lossless regime, our approach is superior. For near-lossless compression, the best coding performance in bitrate is usually provided by CCSDS-123-AC, while our approach is still competitive for most error tolerance values Λ and for most sensors. Concerning quality, our approach is always the best performing.

6.6.3 Two-stage near-lossless coders

Two-stage near-lossless compression obtains a lossy reconstructed scene, then quantizes the difference between this reconstructed and the original scene, and sequentially entropy-encodes this quantized signal. The performance of this strategy strongly depends on the distortion introduced in the lossy stage. The embedded two-stage

near-lossless coder of Beerten et al. [32] includes an iterative selection criterion to determine the optimal amount of lossy bitrate, yielding the best compression ratios when compared to the rest of the techniques of this category.

Regarding the bitrate performance, a comparison between this technique and our proposal shows the good behavior of the two-stage coder (with the exception of results for the CH corpus in the lossless regime). However, it applies the very expensive KLT in terms of computational complexity and a non-negligible side information. When processing scenes with a very large number of spectral components, [32] results to be computationally unfeasible. Contrary to this, NLRWA yields competitive bitrates at significantly lower computational complexity and superior scene's quality retrieval.

6.6.4 Lossy rate control-based coders

End-to-end rate control-based systems allow for compressing a codestream to a bitrate the closest possible to a prefixed target bitrate. Different to JPEG 2000, this type of techniques do not enable progressive decoding. In the literature, the state-of-the-art rate control-based technique that provides some guarantees on the error incurred by individual pixels is the fast and lightweight rate control of Valsesia et al. [33]. It extends the prediction-based CCSDS-123.0-B-1 framework, and outperforms previous coding techniques [39, 40] that also permit lossy rate-controlled outcomes while restricting the maximum distortion simultaneously. Indeed, [33] provides comparable or better coding performances, and decreases their computational complexity at the expense of losing some flexibility.

Experimental results give evidence of the superior performance of NLRWA + JPEG 2000 in comparison to the rate control-based algorithm in bitrate, PAE, and scene quality reconstruction for scenes with a large number of spectral components. This occurs specially when statistical relationships are more meaningful in the spectral than in the 2D spatial domain, as commonly occurs in hyperspectral scenes. For scenes with a low number of spectral components, or with a low average of the Pearson

correlation coefficient, [33] usually improves our approach's outcomes.

Chapter 7

Conclusions

7.1 Summary

Hyperspectral scenes (images) acquired by on-board remote sensing sensors have hundreds or thousands of very correlated spectral components. They contain a wealth of spectral information that turns them into a priceless tool for governments, rescue teams, and aid organizations to deal with natural resources. Remote sensing sensors generate an unprecedented amount of data, even increasing in each mission. This poses forthcoming challenges facing the still strongly limited transmission rate and on-board storage. To cope with these limitations, efficient compression methods are of paramount importance. Compression techniques may recover the scene perfectly (lossless), with constrained (near-lossless) or uncontrolled loss (lossy). Compression is usually provided by transform- or prediction-based techniques. In regards to lossless coding, prediction-based usually outperforms transform-based approaches. For near-lossless compression, prediction-based techniques with quantization also produce competitive outcomes.

Evidence of the lossless Regression Wavelet Analysis (RWA) competitive performance remains more than clear [2, 27, 29], becoming the state-of-the-art lossless compression technique for hyperspectral scenes regarding the trade-off between computational cost and compression ratio. It is a spectral transform that results from

the combination of the integer Haar Discrete Wavelet Transform (Haar-DWT) and a prediction stage conducted through a regression operation. The wavelet used as the first operation block of RWA meaningfully affects the coding performance for the Restricted model, but not for Maximum and Parsimonious. In the case of applying these last two models, the regression is able to maintain the predictions' accuracy regardless of the underlying filter. This is because they utilize the same number of regressors indistinctly of the wavelet used in the RWA pipeline. However, its complexity is slightly increased due to more coefficients are employed by the transform.

On the contrary, for the Restricted model applying other Integer Wavelet Transforms (IWTs-RWA) increases the number of regressors in the prediction stage, usually outperforming the coding performance of the original RWA (Haar-RWA), and improving therefore results presented in [2]. Although the cubic formulation of the original Restricted model is simplified for a first-degree equation, avoiding thus any computation of second- and third-order terms, the computational cost is scarcely increased. IWTs employ more coefficients in the transform than Haar-DWT. Besides, IWTs-RWA do not only use one approximation component in the prediction as the original model does. IWT 5/3-RWA generates 5 regression coefficients per first-degree equation prediction while Haar-RWA defines 4 per third-degree formulation, giving rise to close dimensional matrix operations and to improved times, i.e., Restricted IWT 5/3-RWA is faster than Haar-RWA for the same model. In the cases of IWT 5/3- and IWT 9/7M-Parsimonious, the computational complexity falls very close to the one required by our Restricted IWT 5/3- and IWT 9/7M-RWA versions if a low number of neighbor regressors are evaluated, e.g., respectively 2 and 4.

Regarding the component-scalability, it does not change for the Maximum model. This is because all the regressors are required to recover a scene component. For Parsimonious, this scalability depends on the number of regressors evaluated by the prediction stage and on the used wavelet filter. For Restricted, its scalability is given only by the filter, which fixes both the number of approximation components employed in the prediction stage and in the reverse wavelet transform.

Apart from that, and as expected, for all of the models and variants, applying the highest decomposition level gives rise to the best coding performance. This is because the spectral redundancy of a scene is exploited more with every applied RWA level. In each wavelet subband, the number of coefficients to transform is halved, significantly reducing the number of very correlated components in each level. In fact, applying 5 decomposition levels already produces almost the same performance as when applying all decomposition levels, but the computational cost is slightly lower.

Fast and Exogenous variants normally tend to follow the same behavior as their subjacent regression model. In other words, they barely penalize the coding performance, but reduce the execution times and computational cost. There is an exception for the Exogenous variant when processing scenes with thousands of spectral components. For these images, the side information (SI) is usually very heavy. Exogenous does not store this SI. On the contrary, for its underlying model this SI's size contributes into the final bitrate, providing bitrates significantly higher.

When comparing RWA with other state-of-the-art techniques, we show that RWA provides bitrates close to those of the latest least-squares-based methods, being considerably faster. For the rest of widespread techniques, such as M-CALIC or CCSDS-123.0-B-1, RWA is generally superior in execution time and compression ratio.

In the absence of a lossy RWA implementation, RWA benefits from the rate control algorithm of JPEG 2000 to obtain lossy reconstructions and increase the compression ratio. To avoid stepped and non-monotonic rate-distortion curves, the Predictive Weighting Scheme (PWS) [25] is applied (PWS-RWA + JPEG 2000). The bitrate distributions, per spectral component in the transform domain at different global target bitrates, indicate that the PWS reckons on the predictive contribution of each transformed spectral component into the rate allocation system. They show that the most significant transformed components (from a rate-distortion point of view) are assigned a larger bitrate, thus enabling an improved quality in the recovered scene

(in terms of mean squared error). This tendency exists indistinctly of the underlying wavelet filter or regression model, showing that the correction of the stepped evolution of the rate-distortion curves is achievable for every proposed variant of RWA.

Improved coding performance does not imply enhancement in multi-class classification. Indeed, classification metrics suggest that when the distortion in the reconstructed scenes is high, PWS-RWA + JPEG 2000 is superior, and attains a less severe transitory period. This occurs even when it does not always yield the best rate-distortion outcomes. For low distortion, the recovered scenes conserve enough fidelity with respect to the original ones, producing good classification outcomes for all of the compared compression techniques.

The transform-based standard JPEG 2000 might be too expensive for on-board satellites operation. Thus, we report results when Maximum and Exogenous-Maximum RWA are followed by the lightweight coder CCSDS-123-AC, which is a prediction-based coding technique amenable for on-board implementation. By doing so, we do not only decrease the computational requests of RWA + JPEG 2000, but also improve its compression ratios. This is possible due to CCSDS-123-AC uses its predictor to exploit the remaining redundancy in the RWA transformed domain. Conversely, we only employ the bitplane and entropy encoder of JPEG 2000, but not any extra spectral and/or spatial transform.

Unlike what happens with RWA + JPEG 2000, RWA + CCSDS-123-AC at the highest decomposition level does not necessarily ensure the highest compression ratio. After deploying our Multi-Level Split (MLS) criterion to select the number of levels to apply, our technique (MLS-RWA + CCSDS-123-AC) provides average coding gains between 0.1 to 1.35 bpppc when compared with rKLT + JPEG 2000, the original RWA + JPEG 2000, M-CALIC, CCSDS-123.0-B-1, and CCSDS-123-AC. This is achieved by assessing the bitrate contribution of the components to encode in each wavelet subband, and deciding to stop the RWA + CCSDS-123-AC computation if no gain in coding performance is obtained.

We propose the first near-lossless technique based on regression in a pyramidal framework ever-presented. Our approach is named near-lossless RWA (NLRWA), and expands the lossless implementation of RWA. It is capable of generating the same lossless outcomes as RWA, and increases the compression ratio at the expense of allowing some loss of fidelity in reconstruction. We provide a mathematical formulation that limits the maximum absolute distortion in recovery through an error tolerance value. It only depends on the quantization steps value assigned to each NLRWA decomposition level, demonstrating that the quantization error introduced in each level contributes independently into the final peak absolute error.

For NLRWA, the number of quantization steps combinations is huge. We solve the problem of trying every possible combination by developing a quantization steps selection definition. It is independent of the scene to process, and selects a unique combination. This quantization steps-allocation proposal considers the significance of each transformed component in the reconstruction, by introducing a larger distortion in the lowest decomposition levels. Our NLRWA codestream is encoded by JPEG 2000; nonetheless, other coders could be used. NLRWA + JPEG 2000 enables progressive decoding, while minimizing the error propagation and optimizing both signal-to-noise-ratio (SNR) and PAE performance simultaneously. This framework involves some degree of embeddedness. Finally, NLRWA + JPEG 2000 yields competitive coding performance when compared with state-of-the-art near-lossless and lossy rate control-based techniques, being always superior regarding the scene quality recovery.

THROUGHOUT *this thesis, we have performed a thorough study of the lossless compression state-of-the-art for remote sensing data. Thanks to this, we decided to focus on Regression Wavelet Analysis (RWA), which is the best performing method regarding the trade-off between coding performance and computational cost. With the motivation of providing novel techniques that outperform the state-of-the-art, we have proposed several extensions and improvements of RWA in terms of coding perfor-*

mance, computational complexity, and execution time. Moreover, we have proven its excellent performance in multi-class classification for progressive lossy-to-lossless recovery. After analyzing in depth the behavior of varying each block of RWA, we developed its near-lossless adaptation, known as NLRWA. NLRWA is capable of bounding the peak absolute error in reconstruction, while including, for the first time ever, a regression operation in a pyramidal multiresolution-based scheme. Although its coding performance is not always the best, our approach achieves competitive bitrates, while entailing low computational requirements. NLRWA provides superior scene quality retrieval. Our results are always compared with the most high-performing techniques, countering our contributions to the literature.

7.2 Future work

Although this thesis presents improved and novel data compression techniques for remote sensing applications, more extensive researches and proposals can also contribute to this field of study.

The prediction stage of RWA is conducted through an ordinary least-squares (OLS) method. This OLS operation minimizes the sum of the squared distances between a target value and its prediction. Nonetheless, the competitiveness of other metrics solving this problem has never been tested. Regarding this, reference [41] applies clustering to introduce an additive term in the regression of the Restricted model, increasing the prediction's accuracy and improving bitrates results. In spite of this, to provide evidence of other learning features-based methods that reduce the discrepancy between the original and the estimation values is a natural implementation of RWA.

Besides that, results of RWA when applying a spatial sub-sampling before computing the regression coefficients have already been published. However, other strategies for dimensionality reduction may be investigated with the aim of reducing the memory consumption and floating-points operations.

Moreover, a lossy implementation of Regression Wavelet Analysis has never been investigated. RWA is composed by two blocks that can be modified to accept some uncontrolled loss. For instance, the discrete wavelet transform may be replaced by any floating wavelet to stop operating with integer values. Regarding the regression, some of the side information may be selectively discarded before storing, handling the bias in the least-squares problem for successive decomposition levels, or some loss of information may be introduced into the residuals. In such a case and for further developing, it should be interesting to remove first the rounding operation performed on the predictions.

Several high-performing techniques, such as CCSDS-123.0-B-1, exploit both the 2D spatial and 1D spectral resolutions at once (3D-based methods). From our experimentations, we know that for RWA to decorrelate first the 1D spectral dimension to later exploit the spatial information usually turns out not to be the most competitive strategy. RWA focuses only on reducing the statistical relationships of the spectral components and thus, varying the predictor so that the information from neighbor spatial samples are also considered may lead to enhanced coding performance.

Once assessed the feasibility of RWA to progressive lossy-to-lossless/near-lossless coding, new frameworks and type of compression can be investigated. For instance, the quantized residuals of NLRWA from several error tolerance values can be concatenated in a unique codestream. Later, this codestream can be progressively decoded to obtain a rate-distortion curve in which the error incurred by each pixel is bounded for every reconstruction or, equivalently, for all bitrates. That means, NLRWA codestream can be progressively near-lossless-to-lossless/near-lossless decoded. This type of compression has never been proposed before.

Deep learning is a currently attractive research area, which consists in learning features according to some data representation structures. At present, it has become a mainstream field of study, clearly visible through the multiple publications that

have appeared recently. They train specific large datasets of images to obtain the response parameters whereby producing excellent performance when used on testing images. Based on the same idea, the Exogenous variant of RWA operates the OLS problem off-line. Even though there exists a recent publication [42] that contemplates the possibility of conducting the prediction stage of RWA through convolutional neural networks (CNN), there are still some gaps and lacks in this research field. Thus, it would be interesting to see improved deep learning pipelines substituting the OLS method too.

Satellites cover periodically several orbits along the same areas, capturing the same regions at same or similar times. This generates enormous volumes of repeated data. Delving into exploiting the temporal information, in addition to the spectral and spatial resolutions, may be beneficial to reduce the memory consumption. Indeed, this line of study gives rise to a new area of investigation.

Bibliography

[1] European Space Agency (ESA)„ "eoPortal Directory - Sharing Earth Observation Resources", 2000 - 2019. https://directory.eoportal.org/web/eoportal/satellite-missions

[2] N. Amrani, J. Serra-Sagristà, V. Laparra, M. W. Marcellin, and J. Malo, "Regression wavelet analysis for lossless coding of remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5616–5627, Jun. 2016, DOI: 10.1109/TGRS.2016.2569485.

[3] I. Blanes, E. Magli, and J. Serra-Sagristà, "A tutorial on image compression for optical space imaging systems," *IEEE Geoscience and Remote Sensing Magazine*, vol. 2, no. 3, pp. 8–26, Oct. 2014, DOI: 10.1109/MGRS.2014.2352465.

[4] A. Calderbank, I. Daubechies, W. Sweldens, and B. L. Yeo, "Wavelet transforms that map integers to integers", *Applied and computational harmonic analysis,* vol. 5, no. 3, pp. 332–369, Aug. 1996, DOI: 10.1006/acha.1997.0238.

[5] J. E. Fowler, and J. T. Rucker, "Hyperspectral data exploitation: theory and applications," *3D Wavelet-Based Compression of Hyperspectral Imager, John Wiley & Sons Inc.*, Hoboken, NJ, USA, pp. 379–407, 2007.

[6] X. Tang, and W. A. Pearlman, "Three-dimensional wavelet-based compression of hyperspectral images," *Hyperspectral Data Compression*, Springer, USA, pp. 273–308, 2006, DOI: 10.1007/0-387-28600-4_10.

[7] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Transform coding techniques for lossy hyperspectral data compression," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, no. 5, pp. 1408–1421, Apr. 2007, DOI: 10.1109/TGRS.2007.894565.

[8] J. Zhang, J. E. Fowler, and G. Liu, "Lossy-to-lossless compression of hyperspectral imagery using three-dimensional TCE and an integer KLT," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 814–818, Dec. 2008, DOI: 10.1109/LGRS.2008.2006571.

[9] I. T. Jolliffe,"Principal component analysis." *Springer Verlag,* Berlin, Germany, pp. 487, 2002.

[10] M. Effros, H. Feng, and K. Zeger, "Suboptimality of the Karhunen-Loeve Transform for transform coding," *IEEE Transactions on Information Theory,* vol. 50, no. 8, pp. 1605–1619, Jul. 2004, DOI: 10.1109/TIT.2004.831787.

[11] I. Blanes, and J. Serra-Sagristà, "Cost and scalability improvements to the Karhunen-Loève Transform for remote sensing image coding", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2854–2863, Apr. 2010, DOI: 10.1109/TGRS.2010.2042063.

[12] M. Barret, J. L. Gutzwiller, and M. Hariti, "Low-complexity hyperspectral image coding using exogenous Orthogonal Optimal Spectral Transform (OrthOST) and degree-2 zerotrees," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 49, no. 5, pp. 1557–1566, Nov. 2010, DOI: 10.1109/TGRS.2010.2083671.

[13] I. Blanes, and J. Serra-Sagristà, "Clustered reversible-KLT for progressive lossy-to-lossless 3D image coding," *In Data Compression Conference, IEEE*, pp. 233–242, May 2009, DOI: 10.1109/DCC.2009.7.

[14] I. Blanes, and J. Serra-Sagristà, "Pairwise orthogonal transform for spectral image coding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 961–972, Oct. 2010, DOI: 10.1109/TGRS.2010.2071880.

[15] E. Magli, G. Olmo, and E. Quacchio, "Optimized on-board lossless and near-lossless compression of hyperspectral data using CALIC," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 1, pp. 21–25, Feb. 2004, DOI: 10.1109/LGRS.2003.822312.

[16] X. Wu, and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Transactions on Communications*, vol. 45, no. 4, pp. 437–444, Apr. 1997, DOI: 10.1109/26.585919.

[17] X. Wu, and N. Memon, "Context-based lossless interband compression-extending CALIC," *IEEE Transactions on Image Processing*, vol. 9, no. 6, pp. 994–1001, Jun. 2000, DOI: 10.1109/83.846242.

[18] Consultative Committee for Space Data Systems (CCSDS), "Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1." *CCSDS. Blue Book*, May 2012. https://public.ccsds.org/Pubs/123x0b1ec1s.pdf

[19] J. Bartrina-Rapesta, I. Blanes, F. Aulí-Llinàs, J. Serra-Sagristà, V. Sánchez, and M. W. Marcellin, "A lightweight contextual arithmetic coder for on-board remote sensing data compression", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4825–4835, May 2017, DOI: 10.1109/TGRS.2017.2701837.

[20] J. Song, Z. Zhang, and X. Chen, "Lossless compression of hyperspectral imagery via RLS filter," *IEEE Electronics Letters*, vol. 49, no. 16, pp. 992–994, Aug. 2013, DOI: 10.1049/el.2013.1315.

[21] J. Wu, W. Kong, J. Mielikainen, and B. Huang, "Lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers," *IEEE Signal Processing Letters,* vol. 22, no. 2, pp. 2194–2198, Jun. 2015, DOI: 10.1109/LSP.2015.2443913.

[22] F. Gao, and S. Guo, "Lossless compression of hyperspectral images using conventional recursive least-squares predictor with adaptive prediction bands," *SPIE Journal of Applied Remote Sensing*, vol. 10, no. 1, pp. 015010, Feb. 2016, DOI: 10.1117/1.JRS.10.015010.

[23] J. Mielikainen, and B. Huang, "Lossless compression of hyperspectral images using clustered linear prediction with adaptive prediction length," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 6, pp. 1118–1121, Apr. 2012, DOI: 10.1109/LGRS.2012.2191531.

[24] J. Nocedal, and S. J. Wright, "Least-squares problems," *Numerical optimization,* Springer, pp. 245–269, 2006.

[25] N. Amrani, J. Serra-Sagristà, M. Hernández-Cabronero, and M. W. Marcellin, "Regression wavelet analysis for progressive lossy-to-lossless coding of remote sensing data," *Data Compression Conference, IEEE,* pp. 121–130, Dec. 2016, DOI: 10.1109/DCC.2016.43.

[26] N. Amrani, J. Serra-Sagrista, and M. W. Marcellin, "Unbiasedness of regression wavelet analysis for progressive lossy-to-lossless coding," *IEEE Picture Coding Symposium*, pp. 1–5, Dec. 2016, DOI: 10.1109/PCS.2016.7906316.

[27] N. Amrani, J. Serra-Sagrista, and M. W. Marcellin, "Low complexity prediction model for coding remote sensing data with regression wavelet analysis," *Data Compression Conference, IEEE*, pp. 112–121, May 2017, DOI: 10.1109/DCC.2017.61.

[28] D. S. Taubman, and M. W. Marcellin, "JPEG 2000 image compression fundamentals, standards and practice," *Kluwer International Series in Engineering and Computer Science, Secs 642*, Norwell, Massachusetts 02061 USA, 2002.

[29] S. Álvarez-Cortés, N. Amrani, and J. Serra-Sagristà, "Low complexity regression wavelet analysis variants for hyperspectral data lossless compression," *International Journal of Remote Sensing*, Taylor & Francis, vol. 39, no. 7, pp. 1971–2000, Sep. 2017, DOI: 10.1080/01431161.2017.1375617.

[30] S. Álvarez-Cortés, N. Amrani, M. Hernández-Cabronero, and J. Serra-Sagristà, "Progressive lossy-to-lossless coding of hyperspectral images through regression wavelet analysis," *International Journal of Remote Sensing*, Taylor & Francis, vol. 39, no. 7, pp. 2001–2021, Jul. 2017, DOI: 10.1080/01431161.2017.1343515.

[31] S. Álvarez-Cortés, J. Bartrina-Rapesta, and J. Serra-Sagristà, "Multilevel split regression wavelet analysis for lossless compression of remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 10, pp. 15400–1544, Jul. 2018, DOI 10.1109/LGRS.2018.2850938.

[32] J. Beerten, I. Blanes, and J. Serra-Sagristà, "A fully embedded two-stage coder for hyperspectral near-lossless compression", *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1775–1779, May 2015, DOI: 10.1109/LGRS.2015.2425548.

[33] D. Valsesia, and E. Magli, "Fast and lightweight rate control for on-board predictive coding of hyperspectral images,", *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 3, pp. 394–398, Jan. 2017, DOI: 10.1109/LGRS.2016.2644726.

[34] Consultative Committee for Space Data Systems (CCSDS), "123.0-B-Info TestData," 2015. https://cwe.ccsds.org/sls/docs/Forms/AllItems.aspx?RootFolder=%2fsls%2fdocs%2fsls%2ddc%2f123%2e0%2dB%2dInfo%2fTestData&FolderCTID=0x012000439B56FF51847E41B5728F9730D7B55F

[35] Consultative Committee for Space Data Systems (CCSDS). 2015. "Lossless Multispectral & Hyperspectral Image Compression CCSDS 120.2-G-1." *CCSDS. Green Book. https://public.ccsds.org/Pubs/120x2g1.pdf*

[36] U.S. Geological Survey and NASA, "Earth Observing 1, Hyperion Website." http://eo1.usgs.gov/hyperion.php.

[37] Grupo de Inteligencia Computacional (GIC), "Hyperspectral Remote Sensing Scenes", Apr. 2014, http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

[38] R. Ansari, N. Memon, and E. Ceran, "Near-lossless image compression techniques", *Journal of Electronic Imaging*, vol. 7, no. 3, pp. 486–494, Jul. 1998, DOI: 10.1117/1.482591.

[39] D. Valsesia, and E. Magli, "A novel rate control algorithm for on-board predictive coding of multispectral and hyperspectral images," *IEEE Transactions on*

Geoscience and Remote Sensing, vol. 52, no. 10, pp. 6341–6355, Jan. 2014, DOI: 10.1109/TGRS.2013.2296329.

[40] M. Conoscenti, R. Coppola, and E. Magli, "Constant SNR, rate control, and entropy coding for predictive lossy hyperspectral image compression", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7431–7441, Sep. 2016, DOI: 10.1109/TGRS.2016.2603998.

[41] E. Ahanonu, M. W. Marcellin, and A. Bilgin, "Clustering Regression Wavelet Analysis for Lossless Compression of Hyperspectral Imagery", *Data Compression Conference, IEEE*, Mar. 2019, DOI: 10.1109/DCC.2019.00063.

[42] E. Ahanonu, M. W. Marcellin, and A. Bilgin, "Lossless image compression using reversible integer wavelet transforms and convolutional neural networks", *Data Compression Conference, IEEE*, Jul. 2018, DOI: 10.1109/DCC.2018.00048.