Universitat Autònoma de Barcelona

# Spectral Decorrelation for Coding Remote Sensing Data

**Naoufal Amrani**

**ADVERTIMENT.** L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons: http://cat.creativecommons.org/?page_id=184

**ADVERTENCIA.** El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons: http://es.creativecommons.org/blog/licencias/

**WARNING.** The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license: https://creativecommons.org/licenses/?lang=en

Universitat Autònoma de Barcelona

Departament d'Enginyeria de la Informació i de les Comunicacions

# Spectral Decorrelation for Coding 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 Naoufal Amrani
Bellaterra, December 2016

Supervisor:
Dr. Joan Serra Sagristà

© Copyright 2017 by Naoufal Amrani

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, December 2016

---

Dr. Joan Serra Sagristà

*Committee*:

Dr. Enrico Magli
Dr. Ian Blanes
Dr. Jordi Portell i de Mora
(substitute) Dr. Diego Valsesia
(substitute) Dr. Joan Bartrina Rapesta
(substitute) Dr. Alberto González Villafranca

# Abstract

Today remote sensing is essential for many applications addressed to Earth Observation. The potential capability of remote sensing in providing valuable information enables a better understanding of Earth characteristics and human activities. Recent advances in satellite sensors allow recovering large areas, producing images with unprecedented spatial, spectral and temporal resolution. This amount of data implies a need for efficient compression techniques to improve the capabilities of storage and transmissions. Most of these techniques are dominated by transforms or prediction methods. This thesis aims at deeply analyzing the state-of-the-art techniques and at providing efficient solutions that improve the compression of remote sensing data.

In order to understand the non-linear independence and data compaction of hyperspectral images, we investigate the improvement of Principal Component Analysis (PCA) that provides optimal independence for Gaussian sources. We analyse the lossless coding efficiency of Principal Polynomial Analysis (PPA), which generalizes PCA by removing non-linear relations among components using polynomial regression. Analytically, this approach has proven to provide better performance than PCA. We propose some modifications to the prediction model, using multiple regression in order to increase the prediction accuracy and to reduce the impact of rounding error. We show that principal components are not able to predict each other through polynomial regression, resulting in no improvement of PCA at the cost of higher complexity and larger amount of side information.

This analysis allows us to understand better the concept of prediction in the transform domain for compression purposes. Therefore, rather than using expensive sophisticated transforms like PCA, we focus on theoretically suboptimal but simpler transforms like Discrete Wavelet Transform (DWT). Meanwhile, we adopt predictive techniques to exploit any remaining statistical dependence. Thus, we introduce a novel scheme, called Regression Wavelet Analysis (RWA), to increase the coefficient independence in remote sensing images. The algorithm employs multivariate regression to exploit the relationships among wavelet-transformed components. Specifically, RWA performs a pyramidal prediction based on least squares method in the wavelet domain thus reducing the relations in the residuals and the variance of the representation. We propose several regression models to address different issues concerning compression requirements, although any suitable model can be applied. Moreover, the invertibility of RWA allows reversible integer implementation for lossless coding. Other important advantages are provided, like the low complexity and no dynamic range expansion. Nevertheless, the most important advantage consists of its performance for lossless coding. Extensive

experimental results over a wide range of sensors, such as AVIRIS, IASI and Hyperion, indicate that RWA outperforms the most prominent transforms like PCA and wavelets, and also the best recent coding standard, CCSDS-123.

We extend the benefits of RWA to progressive lossy-to-lossless. We show that RWA can attain a rate-distortion performance superior to those obtained with the state-of-the-art techniques. To this end, we propose a Prediction Weighting Scheme that captures the prediction significance of each transformed components. The reason of using a weighting strategy is that coefficients with similar magnitude can have extremely different impact on the reconstruction quality.

For a deeper analysis, we also investigate the bias in the least squares parameters, when coding with low bitrates. We show that the RWA parameters are unbiased for lossy coding, where the regression models are used not with the original transformed components, but with the recovered ones, which lack some information due to the lossy reconstruction. Also, we show that RWA can use fixed trained parameters, learned over only one image, to accurately encode any other image from the same sensor. Thus, we show that hyperspectral images with large size in the spectral dimension can be coded via RWA without side information and at a lower computational cost.

Finally, we introduce a very low-complexity version of RWA algorithm. Here, the prediction is based on only some few components, while the performance is maintained. When the complexity of RWA is taken to an extremely low level, a careful model selection is necessary. Contrary to expensive selection procedures, we propose a simple and efficient strategy called *neighbor selection* for using small regression models. On a set of well-known and representative hyperspectral images, these small models maintain the excellent coding performance of RWA, while reducing the computational cost by about 90%.

In summary, we have proposed a new efficient algorithm that meets the requirements of remote sensing compression. On-board encoding is highly demanding and asks for efficient and low-complexity approaches. Our proposal provides an excellent performance for both lossy and lossless coding. Moreover, it provides a a wide range of variants that feature a very low-complexity and no need for side information in addition to scalability and no dynamic range expansion.

# Acknowledgements

Whatever has been our progress or everything that we have achieved, we should remember the people who gave us support and understanding. Any acknowledgement falls short.

Firstly, I would like to thank my advisor Joan Serra-Sagristà for the continuous support and motivation since my research was an idea until it became this PhD dissertation. I would like to thank all my co-authors and collaborators for their excellent work. Also, I thank all my colleagues in the department of Information and Communications Engineering (dEIC).

Last but not the least, I express my gratitude to my family and my close friends for their unconditional support.

# Contents

# Chapter 1

# Introduction

## 1.1 Remote Sensing Data

Remote sensing can be defined as the obtention of information or the measurement of objects, areas and phenomenon related to Earth surface, using sensors from a large distance. In particular, optical sensors provide a large and repetitive perception of the Earth surface that significantly contribute to the understanding and study of changes, perhaps due to human activities. Many important applications rely on remote sensing technology, including:

- environment.
- agriculture.
- meteorology.
- water-cycle studies.
- ecology and geology.

## 1.2 Motivation for compression

In recent years, remote sensing technology has developed considerably to meet the different needs of data users and practical applications. As a consequence, recent optical

sensors, deployed on satellites and on airbones, can provide a repetitive and consistent coverage of large geographical areas, producing images with unprecedented spatial, spectral and temporal resolution. For instance, the NASA's instrument Airborne Visible Infrared Imaging Spectrometer (AVIRIS) [1] delivers images of the upwelling spectral radiance in 224 contiguous spectral channels with wavelengths from 400 to 2500 nanometers (nm). On the other hand, the IASI instrument on-board the MetOp satellite captures 8359 spectral channels with a 60 degree field of view, about 1530 lines per orbit, 14 orbits per day, and at an acquisition bit-depth of 16 bits per pixel, producing close to 20 GB daily.

This wealth of data provided by the remote sensing sensors is likely to increase in the future, and that poses a continuing challenge facing modern data storage and transmission. Hence, the need for efficient coding techniques for remote-sensing data becomes more and more imperative to improve the capabilities of storage and transmission.

In general, exploiting the data dependencies along the spectral dimension has proven to play a key role for coding remote sensing images. Thus, the methods that apply efficient spectral models are the ones that yield higher coding performance. This purpose is, in most of the state-of-the-art coding, conducted through either a transform or a prediction approach.

Transform-based methods rely on applying a decorrelation transform that captures most of the image energy in few coefficients. Hence, the data coefficients can be coded independently, while discarding coefficients does not lead to a large error.

Among the popular transforms, the Discrete Wavelete Transform (DWT) belongs to the data-independent family. DWT is widely used as a spectral decorrelator due to its simplicity, component scalability and (very) low complexity, (integrated with JPEG 2000). However, its coding efficiency is fairly limited.

On the other hand, the optimal transform in decorrelating Gaussian sources [2], i.e, principal component analysis (PCA), also known as the Karhunen-Loève Transform (KLT), is also widely applied to multicomponent images. The strength of PCA relies on its excellent coding performance. However, PCA is a data-dependent transform, entailing the need to compute it for each individual image before its application. Fur-

thermore, its computational complexity is substantial due to the covariance matrix calculation, the extraction of eigenvectors, and the matrix factorization/integer mapping (when integer implementation is needed for lossless coding [3]). Also, the PCA is not a component-scalable transform, i.e., the recovery of any single image component depends on every transformed component. In scenarios where the input data consists of a small number of spectral components, such as multispectral images, PCA and its integer implementation can be acceptable. Nevertheless, in scenarios where the input data has a significant size in the spectral dimension, such as IASI images (8359 or 8461 components), PCA and its integer implementation are not feasible.

In light of this fact, a number of approaches have been proposed to reduce the computational complexity of PCA while trying to minimize the degradation in coding performance. One approach is based on sub-sampling the dataset and estimating the covariance matrix using a reduced subset of coefficients spatially and/or spectrally [4, 5, 6]. In this way, the complexity of computing the covariance matrix can be alleviated. However, other complexity sources, such as the eigenvector extraction, the matrix multiplication and the integer mapping for lossless coding remain. A second approach suggests divide-and-conquer strategies to approximate the PCA, while providing reduced computational complexity and some amount of component scalability. These strategies are based mainly on clustered PCA [7, 8] or on multiple pairwise PCA [9]. In most cases, the lossless coding performance of these strategies falls somewhat below that of the full-complexity PCA. Yet a third approach consists in learning the transform on a set of images of one particular sensor in order to obtain an efficient transform that can be applied to new images from the same sensor [10, 11, 12, 13].

Beyond computational issues, a major conceptual limitation of PCA is related to its focus on producing decorrelated (rather than independent) components. In particular, PCA focuses on the covariance matrix (2nd order relations) and neglects higher order moments, which may be relevant in non-Gaussian signals. Remote sensing hyperspectral images have been shown to be non-Gaussian both in the spatial and the spectral dimensions [14, 15, 16]. Since the efficiency of transform coding is attached to the degree of statistical independence achieved [17], approaches focused

on decorrelation may lead to inefficient representations due to higher order statistical relations still being present after decorrelation.

Another drawback of PCA and other transform-based methods is the dynamic range expansion, which is a direct consequence of energy compaction. In some applications, large dynamic range expansion can lead to serious problems, particularly for existing devices or systems that support only a limited bit depth.

Inevitably for transform coding, there is a trade-off between efficiency and complexity. Efficient transforms are expensive, while low complexity transform have a limited performance.

In its turn, the predictive methods usually compute mathematical models to predict data coefficients and only the residual error is encoded [18]. These models can employ spatial, spectral or temporal predictors to decorrelate data. The model parameters can be computed either on-line by using existing data or off-line by a training phase in order to reduce the complexity and the memory requirement. Early hyperspectral coding techniques were based on Differential Pulse Code Modulation (DPCM) [19], which uses pixels in the current and previous bands for prediction, then the residual error is entropy-encoded. In [20] the concept of clustered differential DPCM was introduced. In [21] a coder called Context-based Adaptive Lossless Image Coding (CALIC) was presented. The coding consists of prediction, context modelling, quantization and a conditional entropy coder. A 3D version of CALIC that incorporates interband prediction was proposed in [22] for lossless coding. This extension consists of a simple spectral predictor that employs one reference band to predict the current one. In [23] Magli extended 3D-CALIC to M-CALIC for lossless and near-lossless coding of hyperspectral images. This extension consists of a multiband spectral predictor, along with optimized model parameters and optimization threshold. The proposed M-CALIC outperforms 3D-CALIC for lossless and near-lossless with a limited amount of on-board memory and computational complexity. The Consultative Committee for Space Data Systems (CCSDS) has developed the standard CCSDS-123 [24] for lossless coding of multispectral and hyperspectral images. CCSDS-123 adopts a fast predictive method for lossless coding [25], which relies on a low-complexity adaptive filtering technique.

Summarizing, transform and predictive techniques are the most important and successful approaches for encoding remote sensing images. While transforms allow performing lossless and lossy compression with an efficient quantization of the transformed coefficients, predictive methods usually provide better coding performance for lossless and near-lossless compression [23]. This thesis aims at deeply understanding and exploring the strength and weakness of transform and predictive methods and the potential of their combination, and to introduce novel coding schemes to meet the requirement of remote sensing image compression.

## 1.3 Contributions and Thesis Organization

The contributions of this thesis consist of a compendium of publications addressed to the compression of remote sensing images and published in the relevant journals and conferences on this area:

- N. Amrani, V. Laparra, G. Camps-Valls, J. Serra-Sagrista, and J. Malo, **"Lossless coding of hyperspectral images with principal polynomial analysis,"** in ICIP 2014. Proceedings. International Conference on Image Processing (ICIP), 2014, Oct 2014, pp. 4023–4026 [2].

- N. Amrani, J. Serra-Sagrista, V. Laparra, M. 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, September 2016. [26]

- N. Amrani, J. Serra-Sagrista, M. Hernandez-Cabronero, and M. Marcellin, **"Regression wavelet analysis for progressive-lossy-to-lossless coding of remote-sensing data,"** in Data Compression Conference, March 2016. [27]

- N. Amrani, J. Serra-Sagrista, and M. Marcellin, **"Unbiasedness of regression wavelet analysis for progressive lossy-to-lossless coding"**, in Picture Coding Symposium, December 2016. [28]

- N. Amrani, J. Serra-Sagrista, and M. Marcellin, **"Low Complexity Prediction Model for Coding Remote-Sensing Data with Regression Wavelet Analysis"** in Data Compression Conference, April 2017. [29]

In the first publication [2] discussed in Chapter 2, the idea of predicting transformed coefficients is analyzed through a dimensionality reduction approach called Principal Polynomial Analysis (PPA) [30], which computes polynomial curves in the PCA domain to remove some non-linear dependencies among the principal components. We explore the algorithm for lossless coding of hyperspectral images by applying a reversible implementation with some generalizations to handle the side information and to reduce the impact of the integer mapping. We show that although PPA achieves in theory better higher statistical independence than PCA, in practical hyperspectral coding, exploiting polynomial dependencies among principal components does not necessarily provide a significant improvement for the coding performance. Furthermore, the impact of the integer mapping and the sequential error introduced may penalize the coding gain.

In Chapter 3, the second publication [26] builds upon the previous analysis, consisting of estimating transformed coefficients. However, rather than using sophisticated and expensive data-dependent transforms, we focus on the simpler wavelet transform while adopting regression analysis to exploit the remaining dependencies in the wavelet domain. We introduce Regression Wavelet Analysis (RWA), a predictive scheme that performs a pyramidal prediction based on multiple regression analysis after DWT has been performed. RWA shares the principal DWT properties such as low complexity and spectral scalability while yielding excellent performance for lossless coding. Specifically, RWA yields superior performance to PCA and outperforms the most prominent predictive scheme CCSDS-123.

In Chapter 4, the third publication [27] extends the benefits of RWA to Progressive Lossy-to-Lossless and shows that RWA can attain a rate-distortion performance superior to those obtained by state of the art schemes. For this purpose, a prediction weighting scheme is proposed to capture the prediction significance of each spectral component. The RWA algorithm without component weighting is heavily affected by the error propagation at all bitrates (low, intermediate and high bit-rates), which pro-

duces multiple plateaus in the Rate-Distortion performance. The proposed weighting scheme assigns a weight to each spectral component depending on its contribution to the prediction process. In this regard, it is worth noting that weighting strategies are useful to improve the quality of the recovered data by allocating more bits to the significant coefficients for an algorithm design [31].

In Chapter 5, the fourth publication [28] addresses the problem of bias for lossy coding, when the least squares parameters are used not to the original RWA coefficients, but to the recovered ones, which contain some lack of information due to the lossy reconstruction. This change in the data implies that the regression parameters are suboptimal, but not necessarily biased. To assess the extend of this bias, we adjust the RWA parameters for each bitrate, since the introduced loss is known at the encoder. Based on that, we compare the coding performance before and after adjusting the RWA parameters. We show that the performance sacrifice is negligible (less than 0.2 dB), indicating that the RWA parameters are nearly unbiased for lossy coding.

In Chapter 6, the fifth publication [29] addresses the problem of selecting small prediction models for the RWA scheme to meet the complexity restrictions of on-board satellite compression. Models that include few components as regressors lead to a prediction with low complexity and memory requirement, but usually with a sacrifice in the prediction accuracy. Hence, a strategy to select the best model from a set of candidates ones is needed. However, efficient and sophisticated selection strategies in the state-of-the-art are computationally expensive. We introduce a simple and efficient strategy called *neighbor strategy* that includes some few adjacent spectral components to predict each detail component.

Chapter 7 explains the system coding pipeline, describes the dataset of images and summarizes the coding results. Finally, Chapter 8 closes this thesis with conclusions and provides some insights of the future work.

# Chapter 2

# Lossless Coding of Hyperspectral Images with Principal Polynomial Analysis

# LOSSLESS CODING OF HYPERSPECTRAL IMAGES WITH PRINCIPAL POLYNOMIAL ANALYSIS

*N. Amrani†, V. Laparra‡, G. Camps-Valls‡, J. Serra-Sagristà† and J. Malo‡*

† Dept. Information and Communications Engineering, Universitat Autònoma de Barcelona
‡ Image Processing Laboratory, Universitat de València

## ABSTRACT

The transform in image coding aims to remove redundancy among data coefficients so that they can be independently coded, and to capture most of the image information in few coefficients. While the second goal ensures that discarding coefficients will not lead to large errors, the first goal ensures that simple (point-wise) coding schemes can be applied to the retained coefficients with optimal results. Principal Component Analysis (PCA) provides the best independence and data compaction for Gaussian sources. Yet, non-linear generalizations of PCA may provide better performance for more realistic non-Gaussian sources. Principal Polynomial Analysis (PPA) generalizes PCA by removing the non-linear relations among components using regression, and was analytically proved to perform better than PCA in dimensionality reduction. We explore here the suitability of reversible PPA for lossless compression of hyperspectral images. We found that reversible PPA performs worse than PCA due to the high impact of the rounding operation errors and to the amount of side information. We then propose two generalizations: Backwards PPA, where polynomial estimations are performed in reverse order, and Double-Sided PPA, where more than a single dimension is used in the predictions. Both yield better coding performance than canonical PPA and are comparable to PCA.

***Index Terms***— Principal Component Analysis, Principal Polynomial Analysis, hyperspectral image coding, decorrelation, entropy

## 1. INTRODUCTION

In the last years, a number of techniques have been proposed to exploit spectral and spatial redundancy to encode hyperspectral images. The popular approaches usually combine a 1-D spectral transform followed by a 2-D spatial transform. The spectral decorrelation for multicomponent images has proven to be crucial for compression due to the large amount of inter-component correlation/redundancy, with PCA/KLT [1] transform and its extensions [2] being widely applied due to its high coding efficiency and matrix invertibility.

Recently, several non-linear generalizations of PCA have been proposed [3–5]. These generalizations go beyond linearity and exploit both linear and non-linear correlation between data. Also, their kernel versions [6] have become a powerful tool to extract non-linear features. However, the reconstruction problem has so far restrained their application in compression schemes [7]. Invertibility of the transform is requested for lossless compression, as it leads to perfect reconstruction and provides better understanding of the transform. For lossless coding applications, it is therefore necessary to consider non-linear generalizations of PCA that still satisfy: (1) invertibility to achieve perfect reconstruction, and (2) variance minimization along the considered directions to yield competitive coding performance.

In this paper, we analyze the lossless coding efficiency of an invertible non-linear generalization of PCA, the Principal Polynomial Analysis (PPA) [8, 9], originally proposed for dimensionality reduction. PPA is a deflationary algorithm based on drawing a sequence of Principal Curves that address one dimension at a time [10]. These Principal Curves are analytic and each step in the sequence consists of two basic operations. The first operation is based on PCA, i.e., it is a projection onto the orthogonal direction that maximizes the variance. The second operation estimates and removes the conditional mean of the data in the orthogonal subspace. The estimation is carried out by using a polynomial regression. This provides a better estimation of the conditional mean than the straight line, given the eventual non-linear relation between features.

As opposed to *lossy* compression, direct application of PPA to *lossless* compression may be hampered by the impact of the side information and the necessary lifting scheme applied in each step for integer mapping. In this paper, we specifically analyze these issues: how to handle PPA side information required for the reconstruction, how to reduce the impact of integer mapping error and how to exploit the energy compaction property. Two generalizations of PPA are proposed. The first one, Backwards PPA, works as the original algorithm but in reverse order, i.e., starting from the last components of a PCA, so that the Principal Component, highly responsible for the bitrate budget, is better handled. The second one, Double-Sided PPA, can be seen as a further generalization, a sequential algorithm that simultaneously considers the first and last components of a PCA, uses more dimensions in the prediction, proceeds inwards, and performs less iterations to alleviate the integer mapping penalization.

The paper is organized as follows. Section 2 reviews the elements of PPA transform. Section 3 addresses fundamental issues for the use of PPA in coding: its inverse, the side information, and the proposed generalizations for lossless coding. Section 4 illustrates the application of PPA to hyperspectral images and draws the integer reversible implementation. In section 5 we show the experimental settings and results. Finally, section 6 concludes.

## 2. PRINCIPAL POLYNOMIAL ANALYSIS

Let $\mathbf{X}_0 \in \mathbb{R}^{d\times n}$ be the input data matrix, where rows represent components/dimensions and columns different realizations (or samples). Note that PCA can be seen as a deflationary method (or a sequence of $d-1$ elementary transforms $\mathbf{R}_p$) that maps $\mathbf{X}_0$ to a response domain $\mathcal{R} \subseteq \mathbb{R}^{d\times n}$ . Each elementary transform describes a single curvilin-

This work has been partially funded by the Spanish Government (MINECO), FEDER, the Catalan Government and Universitat Autònoma de Barcelona under grants TIN2012-38102-C03, SGR2009-1224 and 472-03-1/12, respectively.

ear dimension of the input by computing one single component of the response:

$$\begin{pmatrix} \\ \mathbf{X}_0 \\ \\ \end{pmatrix} \xrightarrow{R_1} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \\ \mathbf{X}_1 \end{pmatrix} \xrightarrow{R_2} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \\ \mathbf{X}_2 \end{pmatrix} \cdots \xrightarrow{R_{d-1}} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_{d-1} \\ \mathbf{X}_{d-1} \end{pmatrix}$$

In each step $p$ of the sequence, two basic operations are applied:

$$\boldsymbol{\alpha}_p = \mathbf{e}_p^\top \mathbf{X}_{p-1} \quad (1)$$

$$\mathbf{X}_p^{\text{PCA}} = \mathbf{E}_p^\top \mathbf{X}_{p-1} \quad (2)$$

where (1) is the projection of the data coming from the previous step, $\mathbf{X}_{p-1}$, onto the unit norm vector $\mathbf{e}_p \in \mathbb{R}^{d-p\times 1}$ that maximizes the variance of the projected data and (2) is the projection onto the orthonormal subspace $\mathbf{E}_p \in \mathbb{R}^{(d-p+1)\times(d-p)}$.

PPA follows the same deflationary scheme but it improves the second operation by predicting and removing the conditional mean $\hat{\mathbf{m}}_p$:

$$\boldsymbol{\alpha}_p = \mathbf{e}_p^\top \mathbf{X}_{p-1} \quad (3)$$

$$\mathbf{X}_p^{\text{PPA}} = \mathbf{E}_p^\top \mathbf{X}_{p-1} - \hat{\mathbf{m}}_p \quad (4)$$

The conditional mean is estimated by using polynomial regression to fit the non-linear relationship between the first component $\boldsymbol{\alpha}_p$ and the remaining components $\mathbf{X}_p$ of the projected data coming from the first operation (PCA). In matrix notation, the model can be written as $\hat{\mathbf{m}}_p = \mathbf{W}_p \mathbf{V}_p$:

$$\hat{\mathbf{m}}_p = \begin{pmatrix} w_p^{11} & \dots & w_p^{1(\gamma+1)} \\ \vdots & \vdots & \vdots \\ w_p^{(d-p)1} & \dots & w_p^{(d-p)(\gamma+1)} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \vdots \\ \boldsymbol{\alpha}_p^\gamma \end{pmatrix}$$

where $\mathbf{V}_p$ is the Vandermonde matrix of order $\gamma$ and $\mathbf{W}_p$ is the matrix of polynomial coefficients. The least squares solution for the coefficients is $\mathbf{W}_p = \mathbf{E}_p^\top \mathbf{X}_{p-1} \mathbf{V}_p^\dagger$, where $\mathbf{V}_p^\dagger$ stands for the pseudo-inverse of $\mathbf{V}_p$.

## 3. SIDE INFORMATION AND CODING-ORIENTED GENERALIZATIONS

### 3.1. Inverse and side information

The PPA transform is invertible, thus leading to perfect reconstruction. The inverse transform is, like the forward transform, also a sequence of two elementary inverse operations. The original data $\mathbf{X_0}$ is obtained by applying the following transform recursively on the given transformed data:

$$\mathbf{X}_{p-1} = \begin{pmatrix} \mathbf{e}_p | \mathbf{E}_p \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_p \\ \\ \mathbf{X}_p + \mathbf{W}_p \mathbf{V}_p \end{pmatrix}, \quad p = d-1, \dots, 1$$

In practice, the polynomial coefficients, $\mathbf{W}_p$, and the vector $\mathbf{e}_p$ are necessary side information since they are required at each step of the inversion. On the contrary, there is no need to store $\mathbf{V}_p$ and $\mathbf{E}_p$. On the one hand, $\mathbf{V}_p$ is generated from the data. On the other hand, any method to compute an orthogonal complement from $\mathbf{e}_p$ is fine to obtain $\mathbf{E}_p$ since the reconstruction error does not depend on the selected basis [8, 9].

According to this, the number of elements in the side information corresponding to each elementary transform $\mathbf{R}_p$ is: $(\gamma+1) \times (d-p)$ (from the $\mathbf{W}_p$'s) plus $d-p$ (from the $\mathbf{e}_p$'s). Taking into account that the size of side information for PCA is $\frac{d^2+d}{2}$, PPA side information is $2d^2-d-1$ when using order 2 for the polynomial regression.

### 3.2. Backwards and Double-sided generalizations of PPA

When using dimensionality reduction techniques, the components/coefficients that retain most of the signal energy may have a wide dynamic range. Each PPA step is designed to reduce the dynamic range of the residual signal (prediction error after nonlinear regression). The particular choice of which dimensions are visited in the deflationary scheme will determine the resulting dynamic ranges. While the canonical formulation of PPA follows the *largest-variance-first* criterion (as in PCA), this particular choice may not be appropriate for signal coding.

In order to obtain coefficients with smaller dynamic range, here we propose two variations of the canonical PPA: Backwards PPA and Double-Sided PPA.

In Backwards PPA, the $\mathbf{e}_p$ vector at each step is the one that minimizes the variance of the projected data. This implies retaining the lowest dynamic range projections to reduce the variance of the highest dynamic range ones. The sequence of the Backwards PPA is as follows:

$$\begin{pmatrix} \\ \mathbf{X}_0 \\ \\ \end{pmatrix} \xrightarrow{R_1} \begin{pmatrix} \mathbf{X}_1 \\ \\ \boldsymbol{\alpha}_d \end{pmatrix} \xrightarrow{R_2} \begin{pmatrix} \mathbf{X}_2 \\ \\ \boldsymbol{\alpha}_{d-1} \\ \boldsymbol{\alpha}_d \end{pmatrix} \cdots \xrightarrow{R_{d-1}} \begin{pmatrix} \mathbf{X}_{d-1} \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\alpha}_3 \\ \vdots \\ \boldsymbol{\alpha}_d \end{pmatrix}$$

The side information for the Backwards generalization is essentially the same as in the canonical PPA.

The Double-Sided variation of PPA uses more than one component in each prediction. At each step it uses the largest variance component as in the canonical PPA plus $k$ components from the lowest dynamic range end. This scheme reduces both complexity and side information, since the number of steps needed is $N{=}\lfloor \frac{d}{k+1} \rfloor$ instead of $N{=}(d-1)$. The sequence of the Double-Sided PPA is:

$$\begin{pmatrix} \\ X_0 \\ \\ \end{pmatrix} \xrightarrow{R_1} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ X_1 \\ \boldsymbol{\alpha}_{d-k+1} \\ \vdots \\ \boldsymbol{\alpha}_{d-1} \end{pmatrix} \cdots \xrightarrow{R_i} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ X_{i-1} \\ \vdots \\ \boldsymbol{\alpha}_{d-1} \end{pmatrix}$$

The side information for the Double-Sided generalization is $(\gamma(k+1)+1)\sum_{i=0}^{N}(d-i(k+1))$ due to $\mathbf{W}'_p s$, and $k(\sum_{i=0}^{N-1}(d-i(k+1))$ because of $(\mathbf{e}'_p s)$, where $N = \lfloor \frac{d}{k+1} \rfloor$ is the number of steps.

## 4. PPA FOR HYPERSPECTRAL IMAGE CODING

Hyperspectral imaging leads to 3D spectral-spatial arrays. Even though the $d$-dimensional input to PPA may come from arbitrary arrangements of the hyperspectral arrays, here we will illustrate the use of this transform in a *spectral-first pipeline* (Fig. 1). Current coding standards include PCA in the spectral stage [1, 2]. In our experiments we will stick to this standard scheme and will replace PCA by PPA.

### 4.1. Reversible integer PPA transform

For lossless compression, it is mandatory to consider reversible implementations of the redundancy reduction transforms that produce integer outputs as close as possible to the real coefficients. Hao and Shi [11] proposed a reversible integer KLT that maps integers to integers based on a $\mathbf{PLUS}$ factorization of the transform matrix $\mathbf{A}$. $\mathbf{PLUS}$ is a product of unit Triangular Elementary Reversible Matrices (TERMs), where $\mathbf{P}$ is a permutation matrix, $\mathbf{L}$ and $\mathbf{S}$ are lower TERMs and $\mathbf{U}$ is upper TERM. This factorization is possible, if and only if, $|det(\mathbf{A})| = 1$ [12].

Galli and Salzo [13] improved this factorization by proposing a quasi-complete pivoting. This improvement minimizes the elements' magnitudes of the matrices $\mathbf{L}$ and $\mathbf{U}$, therefore, it reduces the error between the integer implementation and the original transform, which could affect the energy compaction capability.

Going back to the PPA transform and given that it is based on two main operations (PCA projection and conditional mean removal), we propose an integer reversible PPA transform for lossless compression based on two integer reversible operations:

1. Integer reversible PCA/KLT (RPCA/RKLT) based on $\mathbf{PLUS}$ factorization using quasi-complete pivoting.
2. Removal of the quantized conditional mean:

$$\mathbf{X}_p = \mathbf{X}_{p-1} - Q(\hat{\mathbf{m}}_p)$$

The needed conditional mean quantization is performed through a simple rounding operation, $Q(\cdot)$, although any other quantization operation could be applied.

## 5. EXPERIMENTAL SETTINGS AND RESULTS

In this section, we evaluate the proposed reversible integer transforms (PPA, Backwards PPA and Double-Sided PPA) on a set of 5 radiance AVIRIS images: Yellowstone scene 0, Cuprite scene 1, Jasper Ridge scene 2, Moffet Field scene 1 and Low Altitude scene 1, available at *http://aviris.jpl.nasa.gov/data/index.html*. All the images were cropped to 512×512 spatial resolution and 224 components and have a bit-depth of 16 bits per pixel per component (bpppc).

To account for reproducibility, basic Matlab source code for PPA is available online [14], and the proposed generalizations can be found at [15]. The coding system pipeline is shown in Fig. 1 and explains the compression process, where the proposed transforms based on PPA are applied in the spectral domain followed by a DWT 2-D transform 5/3 (5 levels) in the spatial domain and finally a lossless compression using the standard JPEG2000 with the KAKADU software v6.0.

Tables 1–5 provide the lossless coding performance for the proposed transforms as well as for PCA. We report the needed bitrate to encode the images and the size of the requested side information (SI), also measured in bpppc. For Double-Sided PPA with $k$=80, the full matrix $(\mathbf{e}_p\mathbf{E}_p)$ has been used as side information, while for $k < 80$ only the unit norm vector $\mathbf{e}_p$ was used. Coefficients for the side information have used a precision of 32 bits and have been compressed with LZMA. Order 2 has been used for the polynomial regression.

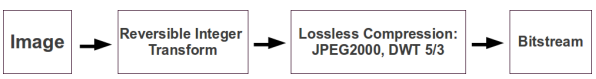



**Fig. 1**: Coding pipeline: 1D transform in the wavelength domain followed by JPEG2000 standard.

**Table 1**: Lossless coding results for Yellowstone scene 00 (bpppc).

| Method | Image | SI $\mathbf{W}_p$'s | SI $\mathbf{e}_p$'s | Total |
|---|---|---|---|---|
| PCA | 3.99 | 0 | 0.025 | 4.02 |
| PPA | 5.01 | 0.038 | 0.012 | 5.06 |
| Backwards | 4.66 | 0.037 | 0.012 | 4.71 |
| D-Sided $k = 1$ | 4.57 | 0.032 | 0.006 | 4.61 |
| D-Sided $k = 7$ | 4.33 | 0.026 | 0.001 | 4.42 |
| D-Sided $k = 80$ | 3.97 | 0.017 | 0.036 | 4.02 |

**Table 2**: Lossless coding results for Cuprite scene 01 (bpppc).

| Method | Image | SI $\mathbf{W}_p$'s | SI $\mathbf{e}_p$'s | Total |
|---|---|---|---|---|
| PCA | 4.99 | 0 | 0.025 | 5.02 |
| PPA | 5.61 | 0.038 | 0.012 | 5.66 |
| Backwards | 5.17 | 0.03 | 0.012 | 5.22 |
| D-Sided $k = 1$ | 5.19 | 0.032 | 0.006 | 5.23 |
| D-Sided $k = 7$ | 5.09 | 0.026 | 0.001 | 5.12 |
| D-Sided $k = 80$ | 4.93 | 0.017 | 0.035 | 4.98 |

**Table 3**: Lossless coding results for Jasper Ridge scene 02 (bpppc).

| Method | Image | SI $\mathbf{W}_p$'s | SI $\mathbf{e}_p$'s | Total |
|---|---|---|---|---|
| PCA | 4.99 | 0 | 0.025 | 5.02 |
| PPA | 5.61 | 0.038 | 0.012 | 5.66 |
| Backwards | 5.17 | 0.037 | 0.012 | 5.22 |
| D-Sided $k = 1$ | 5.19 | 0.032 | 0.006 | 5.23 |
| D-Sided $k = 7$ | 5.11 | 0.028 | 0.003 | 5.14 |
| D-Sided $k = 80$ | 4.91 | 0.017 | 0.035 | 4.96 |

**Table 4**: Lossless coding results for Moffet Field scene 01 (bpppc).

| Method | Image | SI $\mathbf{W}_p$'s | SI $\mathbf{e}_p$'s | Total |
|---|---|---|---|---|
| PCA | 5.01 | 0 | 0.025 | 5.03 |
| PPA | 5.72 | 0.038 | 0.012 | 5.77 |
| Backwards | 5.27 | 0.038 | 0.012 | 5.77 |
| D-Sided $k = 1$ | 5.36 | 0.031 | 0.006 | 5.32 |
| D-Sided $k = 7$ | 5.25 | 0.029 | 0.001 | 5.28 |
| D-Sided $k = 80$ | 5.01 | 0.017 | 0.036 | 5.07 |

**Table 5**: Lossless coding results for Low Altitude scene 01 (bpppc).

| Method (bpppc) | Image | SI $\mathbf{W}_p$'s | SI $\mathbf{e}_p$'s | Total |
|---|---|---|---|---|
| PCA | 5.33 | 0 | 0.025 | 5.35 |
| PPA | 5.91 | 0.038 | 0.012 | 5.96 |
| Backwards | 5.47 | 0.037 | 0.012 | 5.52 |
| D-Sided$k$=1 | 5.58 | 0.031 | 0.006 | 5.62 |
| D-Sided$k$=7 | 5.51 | 0.029 | 0.001 | 5.54 |
| D-Sided$k$=80 | 5.31 | 0.017 | 0.035 | 5.36 |

The reported results indicate that a reversible implementation of PPA does not yield better coding performance than PCA, with a bitrate penalization of at most 0.7 bpppc. This suggests that the sequential rounding error due to the integer mapping on the one hand,

and the conditional mean quantization on the other hand penalize the compression performance. In addition, PPA asks for a larger side information than PCA. Our two PPA-based alternatives are aimed at alleviating these issues.

On the one hand, note that the Backwards PPA is subject to the same rounding process and conditional mean quantization as PPA, and it has the same complexity and side information too; however, it is usually able to improve the results of PPA by about 0.4 bpppc, since in this case the first principal component also has its conditional mean subtracted, decreasing its dynamic range and benefiting the final performance. On the other hand, the Double-Sided PPA outperforms both canonical PPA and Backwards PPA, especially for higher values of parameter $k$ (the number of components from the right-most end that are transformed together with the first component in each iteration). First, the impact of rounding is reduced due to the lower number of iterations. Besides, the predictive power is higher since it uses more than one component ($k+1$), giving rise to a better estimation of the conditional mean. In addition, the computational complexity and side information are smaller in Double-Sided PPA (cf. Section 3.2). According to the reported results, the coding performance of Double-Sided PPA with $k = 80$ is comparable to that of PCA.

## 6. CONCLUSIONS

Lossless compression of hyperspectral images is largely benefited from a 1D spectral transform, which is traditionally performed through a wavelet or a PCA transform, the latter yielding increased performance. PPA is a non-linear generalization of PCA originally proposed in the framework of dimensionality reduction. Here we investigated the suitability of PPA for spectral redundancy reduction in hyperspectral image compression. Coding performance of integer reversible PPA is worse than that of PCA due to the rounding operations error and to the increased size of the side information. For a better adaptation of PPA to the signal at hand, we proposed two generalizations. Backwards PPA reverses the order in which the transform is applied, proceeding from the lower dynamic range hand to the higher dynamic range components. Double-Sided PPA extends canonical PPA by using several components in the regressions and performing less iterations than canonical PPA, thus reducing the penalization of the rounding error. Double-Sided PPA improves PPA performance and reaches PCA performance for lossless compression. These proposals improved the results, and could eventually be used in other nonlinear invertible schemes in the future, not necessarily based on PPA.

Several questions are now open for further research. From a theoretical point of view, it is worth analyzing Double-Sided PPA generalization and investigate whether there is any bound on the number of iterations that still provide some gain. Also, the regression used in PPA to remove the non-linear relations among components has been herein carried out with a simple second-order polynomial regression; other approaches can be devised to improve the estimation and, additionally, to decrease the size of the side information they entail. One such approach could be to apply dedicated polynomial regression in tiles or clustered data, while other *sparse* regression schemes could be explored. It does not escape our notice that PPA could be beneficial compared to PCA in lossy schemes, given that PPA better captures and preserves the relevant image features, and the good results observed previously in dimensionality reduction problems.

## 7. REFERENCES


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

[2] I. Blanes and J. Serra-Sagristà, "Cost and scalability improvements to the Karhunen-Loève transform for remote-sensing image coding," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 7, pp. 2854–2863, Jul. 2010.

[3] M. Schölz, "Validation of nonlinear PCA," *Neural processing letters*, pp. 1–10, 2012.

[4] M. Schölz, F. Kaplan, C.L. Guy, J. Kopka, and J. Selbig, "Non-linear PCA: a missing data approach," *Bioinformatics*, vol. 21, no. 20, pp. 3887–3895, 2005.

[5] J. Venna, J. Peltonen, K. Nybo, H. Aidos, and S. Kaski, "Information retrieval perspective to nonlinear dimensionality reduction for data visualization," *J. Mach. Learn. Res.*, vol. 11, pp. 451–490, 2010.

[6] B. Schölkopf, A. J. Smola, and K-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comp.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[7] Benjamin Huhle, *Kernel PCA for Image Compression*, Ph.D. thesis, Wilhelm-Schickard-Institut für InformatikEberhard Karls Universität Tübingen, 2006.

[8] V. Laparra, D. Tuia, S. Jiménez, G. Camps-Valls, and J. Malo, "Nonlinear data description with principal polynomial analysis," in *IEEE Intl. Workshop Mach. Learn. Sig. Proc.*, 2012, pp. 1–6.

[9] V. Laparra, S. Jiménez, D. Tuia, G. Camps-Valls, and J. Malo, "Principal polynomial analysis," *Intl. Journal of Neural Systems*, 2014.

[10] V. Laparra, S. Jiménez, G. Camps-Valls, and J. Malo, "Nonlinearities and adaptation of color vision from sequential principal curves analysis," *Neural Computation*, vol. 24, no. 10, pp. 2751–88, 2012.

[11] Pengwei Hao and Q. Shi, "Reversible integer KLT for progressive-to-lossless compression of multiple component images," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, 2003, vol. 1, pp. I–633–6 vol.1.

[12] Pengwei Hao and Qingyun Shi, "Matrix factorizations for reversible integer mapping," *Signal Processing, IEEE Transactions on*, vol. 49, no. 10, pp. 2314–2324, 2001.

[13] L. Galli and S. Salzo, "Lossless hyperspectral compression using KLT," in *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, 2004, vol. 1, pp. 313–316.

[14] V. Laparra, G. Camps-Valls, and J. Malo, "PPA toolbox," http://isp.uv.es/ppa.html, 2013.

[15] N. Amrani, V. Laparra, J. Serra, G. Camps-Valls, and J. Malo, "PPA coding toolbox," http://gici.uab.es/GiciWebPage/downloads.php, 2014.

# Chapter 3

# Regression Wavelet Analysis for Lossless Coding of Remote-Sensing Data

# Regression Wavelet Analysis for Lossless Coding of Remote-Sensing Data

Naoufal Amrani, Joan Serra-Sagristà, *Senior Member, IEEE*, Valero Laparra, Michael Marcellin, *Fellow, IEEE*, Jesus Malo

***Abstract*—A novel wavelet-based scheme to increase coefficient independence in hyperspectral images is introduced for lossless coding. The proposed Regression Wavelet Analysis (RWA) uses multivariate regression to exploit the relationships among wavelet-transformed components. It builds on our previous nonlinear schemes that estimate each coefficient from neighbor coefficients [1]–[3]. Specifically, RWA performs a pyramidal estimation in the wavelet domain thus reducing the statistical relations in the residuals and the energy of the representation compared to existing wavelet-based schemes. We propose three regression models to address the issues concerning estimation accuracy, component scalability and computational complexity. Other suitable regression models could be devised for other goals. RWA is invertible, it allows a reversible integer implementation, and it does not expand the dynamic range. Experimental results over a wide range of sensors, such as AVIRIS, Hyperion and IASI, suggest that RWA outperforms not only Principal Component Analysis (PCA) and Wavelets, but also the best and most recent coding standard in remote sensing, CCSDS-123.**

***Index Terms*—Transform coding via regression, wavelet-based transform coding, remote sensing data compression, redundancy in hyperspectral images.**

## I. Introduction

REMOTE-SENSING data has become enormously important for a myriad of applications addressed to the Earth's observation. Recent sensors can cover large geographical areas, producing images of unprecedented spectral and spatial resolution. For instance, the IASI instrument onboard the MetOp satellite captures 8359 spectral channels with a 60 degree field of view, about 1530 lines per orbit, 14 orbits per day, and at an acquisition bit-depth of 16 bits per pixel, producing close to 20 GB daily. Hence, the need for efficient coding techniques for remote-sensing data becomes more and more imperative to improve the capabilities of storage and transmission.

This work was supported in part by the EU (FEDER), the Spanish Government (MINECO), and the Catalan Government, under grants TIN2015-71126-R, TIN2012-38102-C03-00, TIN2013-50520-EXP, BFU2014-58776-R, and 2014SGR-691.

N. Amrani and J. Serra-Sagristà are with Universitat Autònoma de Barcelona, 08193 Barcelona (Spain).

V. Laparra and J. Malo are with Universitat de València, 46980 València (Spain).

M. W. Marcellin is with University of Arizona, Tucson, AZ 85721 (USA).

Most efficient coding techniques for remote-sensing data are based on a redundancy reduction transform that exploits the relations in the spectral and in the spatial dimensions. The problem of appropriate signal representation for transform coding is equivalent to the feature extraction problem in statistical learning [4]. Some hyperspectral coding techniques apply a a discrete wavelet transform in all three dimensions (2 spatial, 1 spectral) [5], [6]. Other techniques apply a 2-D discrete wavelet transform (DWT) spatially, while using a different transform in the spectral dimension [7]–[10]. The transform in the spectral dimension is considered crucial in coding hyperspectral images due to its significant impact on the coding performance [11]. Generally, transforms that provide better exploitation of correlation, resulting in better energy compaction, are the ones that yield larger coding gain [12].

In particular, Principal Components Analysis (PCA), also known as the Karhunen-Loève Transform (KLT), is the optimal decorrelating transform for Gaussian sources [13]–[15]. It is widely applied to multicomponent images because of its excellent performance as a spectral decorrelator. However, PCA is a data-dependent transform, entailing the need to compute it for each individual image before its application. Additionally, its computational complexity is substantial due to the covariance matrix calculation, the extraction of eigenvectors, and the matrix factorization/integer mapping (when integer implementation is needed for lossless coding [16]). Also, the PCA is not a component-scalable transform, i.e., the recovery of any single image component depends on every transformed component. In scenarios where the input data consists of a small number of spectral components, such as multispectral images, PCA and its integer implementation can be acceptable. Nevertheless, in scenarios where the input data has a significant size in the spectral dimension, such as IASI images, PCA and its integer implementation are not feasible.

In light of this fact, a number of approaches have been proposed to reduce the computational complexity of PCA while trying to minimize the degradation in coding performance. One approach is based on sub-sampling the dataset and estimating the covariance matrix using a reduced subset of coefficients spatially and/or spectrally [7], [11], [17]. In this way, the complexity of computing the covariance matrix can be alleviated. However, other complexity sources, such as the

eigenvector extraction, the matrix multiplication and the integer mapping for lossless coding remain. A second approach suggests divide-and-conquer strategies to approximate the PCA, while providing reduced computational complexity and some amount of component scalability. These strategies are based mainly on clustered PCA [18], [19] or on multiple pairwise PCA [20]. In most cases, the lossless coding performance of these strategies falls somewhat below that of the full-complexity PCA. Yet a third approach consists in learning the transform on a set of images of one particular sensor in order to obtain an efficient transform that can be applied to new images from the same sensor [21]–[24].

Beyond computational issues, a major conceptual limitation of PCA is related to its focus on producing decorrelated (rather than independent) components. In particular, PCA focuses on the covariance matrix (2nd order relations) and neglects higher order moments, which may be relevant in non-Gaussian signals. Remote sensing hyperspectral images have been shown to be non-Gaussian both in the spatial and the spectral dimensions [25]–[27]. Since the efficiency of transform coding is attached to the degree of statistical independence achieved [28], approaches focused on decorrelation may lead to inefficient representations due to higher order statistical relations still being present after decorrelation.

The (theoretical) consideration of higher order moments in Independent Component Analysis (ICA) [29], and the (experimental) fact that ICA filters are qualitatively similar to wavelet functions in hyperspectral imagery [25], may be the fundamental reasons for the widespread use of wavelets in transform coding of hyperspectral images [5], [6], [30].

Another limitation of classical PCA (and also classical ICA) is its linear nature. Linear PCA and linear ICA (and their fixed basis approximations, DCT and wavelets) do not completely achieve statistical independence in hyperspectral images. Similar to the case of photographic images [31]–[35], residual dependencies have been reported for hyperspectral imagery in these representations [26], [27]. Thus, despite the simplicity and efficiency that linearity provides, *in principle*, non-linear methods could provide performance improvements by better exploiting dependencies present in the data.

The discussion above suggests the pursuit of nonlinear generalizations of PCA, ICA and their fixed-basis versions, DCT and wavelets, respectively. In [1], [3], the following families of nonlinear generalizations of feature extraction transforms were reviewed: (i) kernel and spectral techniques such as kernel-PCA, kernel-ICA or Local Linear Embedding (LLE) [36]–[38], (ii) neural networks and autoencoders [39]–[42], and (iii) techniques based on curvilinear features [1], [3], [13], [34], [43]–[45]. In the adaptation of these feature extraction ideas to image coding, there are two major considerations of importance: first, the transform must be invertible; and second, the computational complexity and the memory consumption should be reasonable. These requirements eliminate a large number of candidate approaches. For example, many non-linear generalizations of ICA are not invertible (for instance [34], [43], [46]). With this in mind, the most promising approaches seem to lie within the invertible techniques of families (ii) and (iii).

In [2] we explored lossless hyperspectral image coding using curvilinear techniques (family iii) based on Principal Polynomial Analysis (PPA) [1]. PPA exploits regression to remove non-linear dependencies that remain after linear feature extraction (e.g., after classical PCA). In [1] it was shown that PPA achieves higher energy compaction and statistical independence than PCA. However, in practical hyperspectral coding, appropriate handling of side information and the sequential error introduced by the integer mapping dramatically penalize the coding gain. Additionally, the computational complexity and memory requirements of these non-linear transforms are even larger than those of the original PCA, which is already demanding [13].

As a result, rather than using sophisticated non-linear data dependent feature extraction, we focus on theoretically suboptimal but simpler traditional transforms (such as the Discrete Wavelet Transform) while adopting predictive techniques [47], [48] to exploit any remaining *post transform* statistical dependence. This idea has been applied to encode residual errors in the spatial domain after exploiting smoothness [49] in photographic images. In work more closely related to that proposed here, predictive schemes have been employed in the wavelet domain based on the known relations of image coefficients [32], [50].

Building upon this, this paper introduces the Regression Wavelet Analysis (RWA). This transform can be seen as a predictive scheme to reduce redundancy after wavelet analysis has been performed in the spectral dimension of hyperspectral images. RWA shares the principal properties of the DWT such as component scalability and low complexity, while yielding excellent performance for lossless coding. Specifically, RWA yields superior performance to that of the best comparable spectral decorrelation techniques like PCA. Significantly, RWA also overperforms the most recent and most competitive prediction-based hyperspectral coding technique, CCSDS-123 [51].

The paper is organized as follows. Section II describes and formalizes the proposed RWA scheme and selection of the regression model. Section III investigates several relevant features of RWA, including its redundancy reduction ability and computational complexity. In Section IV we assess the coding performance achieved by RWA for images corresponding to a wide range of remote sensing scenarios. Finally, Section V puts forward our conclusions.

## II. Regression Wavelet Analysis

To establish the necessary notation, this section begins with a review of the DWT. The discussion then proceeds to an analytical description of the RWA and the corresponding regression model. The section concludes by suggesting a fast implementation for RWA.

### A. Discrete Wavelet Transform

The discrete wavelet transform (DWT) provides a multiresolution decomposition of signals into approximation, $\mathbf{V}$, and details, $\mathbf{W}$. In what follows, we consider the 1D DWT in the component (spectral) direction of a multicomponent image. The DWT can be computed with a pyramidal algorithm based on convolution [52]. The algorithm is illustrated in Fig. 1. We begin by considering a general formulation for the DWT that maps real numbers to real numbers. We later consider transforms that map integers to integers.

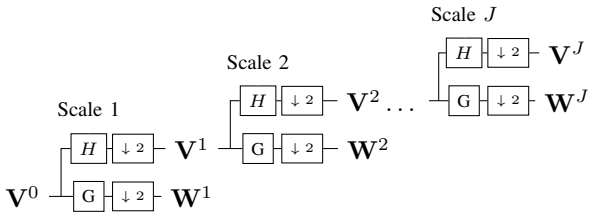



Figure 1. DWT decomposition with $J$ levels.

Suppose that an original multicomponent image $\mathbf{V}^0$ has $z{=}2^d$ components with each component having $m$ spatial samples. Then we write $\mathbf{V}^0{\in}\mathbb{R}^{m\times z}$ and

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

Then the wavelet representation of $\mathbf{V}^0$ with $J$ levels for $1\le J\le \log_2(z)$ is given by:

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

where the one level DWT decomposition of each $\mathbf{V}^{j-1}$ is given by:

$$\mathbf{DWT}(\mathbf{V}^{j-1},1)=\left(\mathbf{V}^j,\mathbf{W}^j\right) \tag{2}$$

with

$$\mathbf{V}^j(n)=\sum_k h(2n-k)\mathbf{V}^{j-1}(k) \tag{3}$$

and

$$\mathbf{W}^j(n)=\sum_k g(2n-k)\mathbf{V}^{j-1}(k). \tag{4}$$

In these expressions, $h(n)$ and $g(n)$ are the impulse responses of low pass and high pass analysis filters $H$ and $G$, respectively. At each scale or level $j$ the signal $\mathbf{V}^{j-1}{\in}\mathbb{R}^{m\times(z\cdot 2^{-j+1})}$ is decomposed into the approximation signal $\mathbf{V}^j{\in}\mathbb{R}^{m\times(z\cdot 2^{-j})}$ and the details signal $\mathbf{W}^j{\in}\mathbb{R}^{m\times(z\cdot 2^{-j})}$ at half resolution each. The approximation signal $\mathbf{V}^j$ usually contains most of the information of the previous signal $\mathbf{V}^{j-1}$, whereas the details signal $\mathbf{W}^j$ contains the information difference between $\mathbf{V}^{j-1}$ and $\mathbf{V}^j$. The decomposition is usually repeated in cascade form on the approximation signal $\mathbf{V}^j$ as shown in Fig. 1 and the maximum number of iterations that can be performed is $d=\log_2(z)$.

To reconstruct the original data, the DWT components are passed through low and high-pass synthesis filters to obtain:

$$\begin{aligned}\mathbf{V}^{j-1}(n)&=\mathbf{DWT}^{-1}\left((\mathbf{V}^j,\mathbf{W}^j),1\right)=\\ &\sum_k \tilde{h}(n{-}2k)\mathbf{V}^j(k){+}\sum_k \tilde{g}(n{-}2k)\mathbf{W}^j(k).\end{aligned} \tag{5}$$

If the original signal $\mathbf{V}^0$ has $z{=}2^d$ components, then the representations $\mathbf{V}^j$ and $\mathbf{W}^j$ have $z{\cdot}2^{-j}$ components each and the DWT representation $\left(\mathbf{V}^J,(\mathbf{W}^j)^{1\le j\le J}\right)$ has the same total number of components as the original signal $\mathbf{V}^0$. With proper boundary handling procedures, the number of components $z$ need not be a power of two. In this case $z{\cdot}2^{-j}$ is rounded up or down so that the total number of components at each level is always the same.

Rather than the convolution implementation described above, the DWT decomposition can be performed using a lifting scheme [53], [54]. This implementation facilitates the inclusion of rounding steps to obtain invertible transforms that map integers to integers. Such transforms are often called "reversible." In the wavelet literature, it has been demonstrated that the DWT approximately decorrelates some stochastic and non stochastic processes [55]. This means that any two distinct within-scale or between-scale coefficients are approximately decorrelated, with the correlation decaying as the separation between scales increases.

### B. Regression Wavelet Analysis

The proposed Regression Wavelet Analysis (RWA) scheme generalizes the discrete wavelet transform (DWT) by applying regression to tackle the redundancy that still remains in the DWT domain. At each scale $j$, each details component $\mathbf{W}^j(i)=\mathbf{W}_i^j\in\mathbb{R}^{m\times 1}$ is predicted from the information contained in the approximation components $\mathbf{V}^j\in\mathbb{R}^{m\times(z\cdot 2^{-j})}$ within the same scale $j$. This prediction $\widehat{\mathbf{W}}_i^j=f_i(\mathbf{V}^j)$ is then removed to obtain:

$$\mathbf{R}^j=\mathbf{W}^j-\widehat{\mathbf{W}}^j. \tag{6}$$

A regression model $f_i(\mathbf{V}^j)$ is used to estimate the conditional mean of each $\mathbf{W}_i^j{\in}\mathbb{R}^{m\times 1}$ (dependent variable) from some or all the approximation components $\mathbf{V}^j{\in}\mathbb{R}^{m\times(z\cdot 2^{-j})}$ (independent variables). Later in this paper, we will propose

three different models that address the issues concerning estimation accuracy, computational complexity and spectral scalability. Note that the prediction functions $f$ might be linear or nonlinear, and they might use a small or a large set of neighbor coefficients. These possibilities for prediction through regression lead to a range of particular implementations of RWA.

As illustrated in Fig. 2, the resulting RWA affects only the details components at each level of the transform. The approximation components are unchanged from those of the DWT.

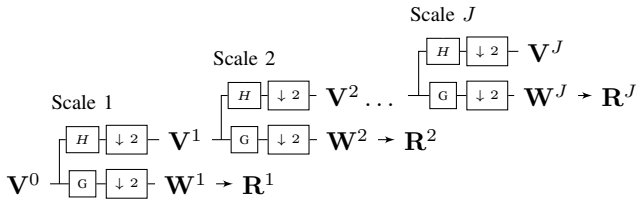



Figure 2. RWA decomposition with $J$ levels.

Following the notation of (Eq. 2) and (Eq. 1), the one level RWA and the $J$ level RWA become:

$$\mathbf{RWA}(\mathbf{V}^{j-1}, 1) = \left(\mathbf{V}^j, \mathbf{R}^j\right) \tag{7}$$

and

$$\mathbf{RWA}(\mathbf{V}^0, J) = \left(\mathbf{V}^J, (\mathbf{R}^j)^{1\leq j\leq J}\right), \tag{8}$$

respectively.

The RWA transform is easily inverted. At each scale $j$ the estimate $\widehat{\mathbf{W}^j}$ is computed from $\mathbf{V}^j$. The approximation at level $j-1$ is then obtained by computing:

$$\mathbf{W}^j = \mathbf{R}^j + \widehat{\mathbf{W}^j} \tag{9}$$

$$\mathbf{V}^{j-1} = \mathbf{DWT}^{-1}\left((\mathbf{V}^j, \mathbf{W}^j), 1\right). \tag{10}$$

By repeating in cascade for $j = J, J\text{-}1, \ldots, 1$, the reconstruction $\mathbf{V}^0$ is achieved from its RWA representation $\left(\mathbf{V}^J, (\mathbf{R}^j)^{1\leq j\leq J}\right)$.

### C. Regression Model

To find the estimate $\widehat{\mathbf{W}}^j$ of the details components $\mathbf{W}^j$ based on the approximation components $\mathbf{V}^j$, three different estimation models are discussed in this subsection. The first model, called the *Maximum model*, is the most general. It employs all of the approximation components from $\mathbf{V}^j$ in the computation of the prediction of each details component $\mathbf{W}_i^j$. The second model is called the *Restricted model*. For a given details component to be predicted, this model determines a subset of components from $\mathbf{V}^j$ to include in the prediction in order to preserve the component scalability of the original DWT. The third and final model is called the *Exogenous model*. This model has a low computational cost, and is similar to the Exogenous model proposed for the PCA [56] or for the orthogonal optimal spectral transform (OST) [21], [22].

#### 1) Maximum model

The *Maximum model* involves all $k{=}z{\cdot}2^{-j}$components of the approximation $\mathbf{V}^j{\in}R^{m\times(z\cdot 2^{-j})}$ in order to estimate each details component $\mathbf{W}_i^j{\in}\mathbb{R}^{m\times 1}$, $i{\in}I{=}\{1, ..., k{=}z{\cdot}2^{-j}\}$. The form of each estimator is given by:

$$\widehat{\mathbf{W}}_i^j = \boldsymbol{\beta}_{i,0}^j + \boldsymbol{\beta}_{i,1}^j\mathbf{V}_1^j + \cdots + \boldsymbol{\beta}_{i,k}^j\mathbf{V}_k^j, \quad \mathbf{V}_i^j{\in}\mathbb{R}^{m\times 1}. \tag{11}$$

Since all components of the approximation are included, this form provides the most general linear estimator possible in this formulation. In this model, the estimator parameters (regression coefficients) $\boldsymbol{\beta}^j$ are found for each individual image using the least-squares method [57] that minimizes the sum of squares of the distances between the original components and the estimated ones:

$$\textit{min: } \|\mathbf{W}_i^j - \widehat{\mathbf{W}}_i^j\|_2.$$

#### 2) Restricted model

The *Restricted model* employs only a small number of approximation components, with the goal of preserving as much as possible the component scalability of the original DWT. In other words, the number of transformed components that need to be accessed to reconstruct an image component should be as small as possible. In the DWT context, the number of needed transformed components depends on the width of the synthesis filters. As an example, consider the Haar DWT. For this simple transform, only two transformed components from level $j$ are required to reconstruct one approximation component at level $j-1$, one approximation component and one details component. More generally, for a given DWT, a number $|I_p|$ of approximation components $\mathbf{V}_{i\in I_p}^j$ and a number $|I_q|$ of details components $\mathbf{W}_{i\in I_q}^j$ from level $j$ are required to reconstruct an approximation component $\mathbf{V}_r^{j-1}$ at level $j-1$. We denote this by:

$$(\mathbf{V}_{i\in I_p}^j, \mathbf{W}_{i\in I_q}^j) \xrightarrow{Reconst.} \mathbf{V}_r^{j-1}.$$

From the opposite point of view, a given single details component $\mathbf{W}_i^j$ is involved in the reconstruction of $t$ components at level $j-1$. Let us then associate $\mathbf{W}_i^j$ with these $t$ components $\left(\mathbf{V}_{r1}^{j-1}, \ldots, \mathbf{V}_{rt}^{j-1}\right)$, and for each single component $\mathbf{V}_{ri}^{j-1}$, we associate it with the set of the approximation components $(\mathbf{V}_{i\in I_{ri}}^j)$ needed for its reconstruction:

$$\mathbf{W}_i^j \xrightarrow{Reconst.} \begin{cases} \mathbf{V}_{r1}^{j-1} \xleftarrow{Reconst.} & \mathbf{V}_{i\in I_{r1}}^j \\ \vdots & \vdots \\ \mathbf{V}_{rt}^{j-1} \xleftarrow{Reconst.} & \mathbf{V}_{i\in I_{rt}}^j. \end{cases}$$

For Haar filter this association is simple:

$$\mathbf{W}_i^j \xrightarrow{Reconst.} \left\{ \begin{array}{l} \mathbf{V}_{2i-1}^{j-1} \\ \mathbf{V}_{2i}^{j-1} \end{array} \right\} \begin{array}{l} \xleftarrow{Reconst.} \\ \xleftarrow{Reconst.} \end{array} \begin{array}{l} \mathbf{V}_i^j \\ \mathbf{V}_i^j. \end{array}$$

Now, if in the regression model to predict $\mathbf{W}_i^j$ we involve the approximation components $\mathbf{V}_{i\in\mathfrak{I}}^j$, where $\mathfrak{I} = \cup(I_{r1},\ldots,I_{rt})$ is the union of these $t$ sets of indices of approximation components in the scale $j$, then this model will largely preserve the scalability of the original DWT. Note that the intersection $\cap(I_{r1},\ldots,I_{rt})$ could be empty depending on the wavelet filter used. It is beyond the scope of this work to discuss the model that exactly preserves the component scalability for any arbitrary filter. Nevertheless, we propose a model that perfectly preserves the scalability for Haar filter, with $\mathfrak{I}=\{i\}$. It is given by:

$$\widehat{\mathbf{W}}_i^j = f_i[\mathbf{V}_i^j].$$

In this work we propose the following model for Haar, involving $\mathbf{V}_i^j$ and its second and third-order terms:

$$\widehat{\mathbf{W}}_i^j = \boldsymbol{\beta}_{i,0}^j + \boldsymbol{\beta}_{i,1}^j \mathbf{V}_i^j + \boldsymbol{\beta}_{i,2}^j (\mathbf{V}_i^j)^2 + \boldsymbol{\beta}_{i,3}^j (\mathbf{V}_i^j)^3. \qquad (12)$$

The *Restricted model* preserves the component scalability requirement of the original DWT by involving a restricted number of components in the prediction model. An added benefit due to the restricted number of components is a reduction of computational complexity. Higher-order terms of $\left(\mathbf{V}_i^j\right)^n$ can be added to improve the predictive power. Additional improvements may be had by adding more components $\mathbf{V}_{k\neq i}^j$, at the cost of a decrease in scalability and an increase in complexity and side information corresponding to the additional regression coefficients.

*3) Exogenous model*

The *Maximum model* employs a large number of components in each estimator, and accurate estimation is expected. However, this large model can result in large computational cost when computing the least squares solution for each image. This method may also result in excessive side information needed to be transmitted in order to inform the decoder of the model parameters. Since each hyperspectral image from the same sensor may have similar statistical relationships among its components, the *Exogenous model* computes the regression coefficients $\overline{\boldsymbol{\beta}}^j$ over a set of training images, and then uses these coefficients for all new images that come from the same sensor. As a consequence, the computational cost of RWA is reduced considerably to roughly the same as that of the DWT. Additionally, no side information needs to be stored for each individual image. The model is given by

$$\widehat{\mathbf{W}}_i^j = f_i\left[\mathbf{V}^j\right] = \overline{\boldsymbol{\beta}}_{i,0}^j + \overline{\boldsymbol{\beta}}_{i,1}^j \mathbf{V}_1^j + \ldots + \overline{\boldsymbol{\beta}}_{i,k}^j \mathbf{V}_k^j, \quad \mathbf{V}_i^j \in \mathbb{R}^{m\times 1}.$$

*4) Side information*

Table I reports the side information for the three proposed models of RWA in comparison with the side information required for PCA, consisting of the $z^2$ transform coefficients and the $z$ means used to center the data prior to transformation. At each level $j$, the prediction of each component $\mathbf{W}_i^j$ requires $1+z\cdot 2^{-j}$ regression coefficients ($\boldsymbol{\beta}_{i,.}^j$) when using the *Maximum model* (Eq. 11) and 4 coefficients when using the *Restricted model* (Eq. 12).

Table I
SIZE OF THE SIDE INFORMATION. $z$ IS THE NUMBER OF COMPONENTS AND $l$ IS THE NUMBER OF WAVELET DECOMPOSITION LEVELS.

| Model | Number of coefficients | $z$= 224 $l$=8 |
|---|---|---|
| RWA (*Maximum*) | $\frac{z^2}{3}(1-\frac{1}{2^{2l}})+z(1-\frac{1}{2^l})$ | 16947 |
| RWA (*Restricted*) | $2z(1-\frac{1}{2^l})$ | 446 |
| RWA (*Exogenous*) | – | – |
| PCA | $z^2+z$ | 50400 |

These three regression models are suggested to be used for different —possibly overlapping— situations. The *Maximum model* is suggested in most of the cases since it is expected to give the most accurate estimation, though the regression coefficients $\boldsymbol{\beta}$ are needed as side information. For situations where the computation cost or the size of the side information matter, the *Exogenous model* is suggested. Finally, the *Restricted model* is suggested for those cases where the computational cost matters or the spectral component scalability is required.

*D. Fast RWA*

The ordinary least squares (OLS) method estimates the regression coefficients with a complexity cost of $O(mk^2)$ [58], [59], where $k$ is the number of components used to form the prediction and $m$ is the number of spatial samples. In the *Maximum model* $k = z\cdot 2^{-j}$ at level $j$, where $z$ is the number of spectral components in the original image.The *Restricted model* uses less input components to the predictors leading to a reduced complexity $O(mk'^2)$ $(k'<k)$ depending on the wavelet filter. For example, for the Haar DWT, $k'=3$ (Eq. 12). However, assuming that $m\gg z$, the number of spatial samples $m$ will dominate the complexity. The fast RWA proposed here addresses this problem by randomly selecting a subset of $m'=\rho\cdot m$ samples ($m'\ll m$). The least-squares optimization is then carried out only on this small dataset to obtain the regression coefficients. For images with a large spatial dimension ($m \gg z$), the subsampling

employed in the fast RWA has minimal impact on the results. However, for images with small spatial dimension, the *Exogenous model* may be a better choice, especially when side information is considered.

## III. Relevant properties of RWA

On the one hand, the redundancy reduction ability of RWA for remote sensing imagery is a major advantage from the fundamental point of view. On the other hand, from the applied perspective, component scalability and the possibility of efficient integer representations are also major advantages of RWA.

### A. Reduction of Correlation and Mutual Information

Hyperspectral images usually exhibit significant redundancy along the spectral dimension. The spectral transform aims at exploiting this redundancy among components, so that they can be coded independently. In this section we analyze the effect of the transform on mutual information and correlation. Mutual information (MI) is not limited to linear relations between variables (as is the case with correlation), but it also takes into account eventual nonlinear relations [60]. As a result, mutual information is a more general description of statistical relations than correlation. Mutual information includes correlation and non-Gaussianity [61]. Therefore, changes in mutual information may come either from better decorrelation or also from removing higher-order statistical relations.

Here we measure correlation and mutual information between the coefficients for the three more representative transformations in this work: wavelets (Haar with eight decomposition levels), PCA and the proposed RWA (*Maximum model* using Haar filter). We have taken the Haar and PCA transformations because they provide extreme cases of previous work from the literature. In particular, the Haar transform has low complexity and provides fine grain scalability, but has modest compression performance. On the other hand, the PCA provides superior compression performance at the cost of high complexity and no scalability.

The following analysis employed the hyperspectral image "Hawaii" (314368 spatial samples, 224 spectral components). The matrices depicted as images in Fig. 3 represent the statistical relations between the coefficients of transformed versions of the image. The left column of matrices represent correlation while the right column of matrices represent mutual information. Each row of matrices corresponds to one (transformed) image representation. For a given matrix, the element $M_{kl}$ represents the quantity of interest (correlation or mutual information) between the *k-th* and *l-th* components.

The degree to which these matrices appear to be diagonal has been used to describe the suitability of a domain for scalar (coefficient-wise) coding [33], [62], [63]. Specifically, better representations for transform coding are those with small off-diagonal values. To quantify this property, the average value of the off-diagonal coefficients are depicted in the figure. The non-diagonal nature of the matrices in the top row shows that the original spectral domain is highly unsuitable for scalar coding. Subsequent rows indicate that significant improvements can be obtained via the application of transforms.

To compare the residual correlation that exists after application of the three transforms, we turn our attention to the left hand side of the figure. As can be seen there, the DWT provides significant decorrelation. However strong correlations still exist among and between the components of certain subbands in the transform domain. PCA achieves significantly better decorrelation than the DWT. By definition, it is designed to diagonalize the covariance matrix. However in practical applications total decorrelation is hard to achieve due to implementation constraints (e.g., reversible integer to integer implementation) and the difference between the sample and the actual covariance matrices. Thus, some residual correlation is still present. It can be seen that after applying RWA (*Maximum model*), almost full decorrelation is reached in the transform domain, as evidenced by the improved diagonal appearance and the smaller off-diagonal average. This illustrates the clear advantage of RWA over the usual Haar DWT and even a small advantage over PCA. Similar comments can be made for mutual information, as reported in bits, in the right side of the figure. These observations serve to explain why RWA obtains superior compression results compared to the usual DWT, as well as comparable performance to that of PCA. These statements are verified in the compression results presented below.

Note that mutual information can be expressed as a sum of a correlation-dependent index, a global negentropy, and a marginal negentropy [61]. Reducing the correlation implies reducing the mutual information in most cases. However, given the invariance of global negentropy to linear transforms, once the correlation has been removed (through PCA) and the marginal negentropy (or sparsity) has been maximized (through ICA), the global negentropy can only be reduced through nonlinear means. The results in Fig. 3 obtained using linear regression– suggest that the prediction in RWA increases the sparsity of the representation. More sophisticated (nonlinear) regression techniques could be included in the RWA framework (functions $f_i$ in Eq. 6) to reduce the global negentropy at the cost of training, complexity and side information.

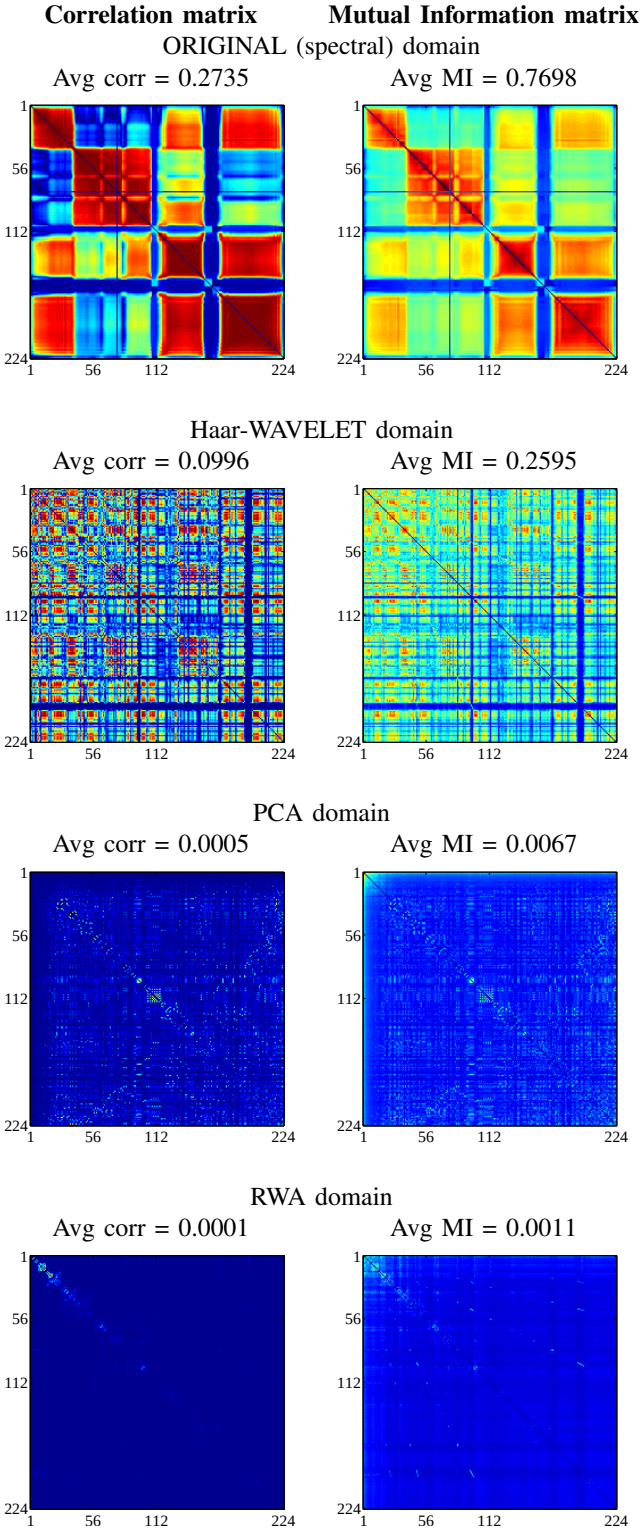



Figure 3. Correlation and mutual information matrices (left and right columns, respectively) for the coefficients of an illustrative hyperspectral image, "Hawaii" with 224 spectral bands. Each row of the figure corresponds to a different representation domain. The blue-to-red colormap indicates low-to-high dependence values. Average values are computed using the magnitude of the off-diagonal elements. MI is given in bits. Matrices in each column are normalized in the same way for fair comparison. For the purpose of visualizing the small off-diagonal values, their fourth roots are displayed.

### *B. Reversible Integer RWA and Dynamic Range*

For lossless coding, it is required that the redundancy reduction transform maps integer coefficients to integer coefficients. For wavelets, the integer versions can be achieved by applying the lifting scheme [64], [65]. For instance, integer versions of Haar filter are computed and reversed as follows:

$$\textit{Forward:} \left\{ \begin{array}{l} \mathbf{W}_i^j = \mathbf{V}_{2i}^{j-1} - \mathbf{V}_{2i-1}^{j-1} \\ \mathbf{V}_i^j = \mathbf{V}_{2i-1}^{j-1} + \lfloor \frac{1}{2}\mathbf{W}_i^j \rfloor, \end{array} \right. \tag{13}$$

$$\textit{Reverse:} \left\{ \begin{array}{l} \mathbf{V}_{2i-1}^{j-1} = \mathbf{V}_i^j - \lfloor \frac{1}{2}\mathbf{W}_i^j \rfloor \\ \mathbf{V}_{2i}^{j-1} = \mathbf{W}_i^j + \mathbf{V}_{2i-1}^{j-1}. \end{array} \right. \tag{14}$$

Given that, we propose a reversible integer version of RWA based on an integer DWT decomposition $\left(\overline{\mathbf{V}}^j, \overline{\mathbf{W}}^j\right)$, and a quantized estimation removal $\overline{\mathbf{R}}^j = \overline{\mathbf{W}}^j - Q(\widehat{\mathbf{W}}^j)$. The scalar quantization is here performed through a simple rounding operation. The reversible integer RWA representation is then given by:

$$\overline{\mathbf{RWA}}(\mathbf{V}^0, J) = \left(\overline{\mathbf{V}}^J, (\overline{\mathbf{R}}^j)^{1 \le j \le J}\right).$$

In general, RWA can be based on any wavelet transform. Nevertheless, as this paper focuses on lossless coding, and for faster implementation we focus exclusively on RWA based on the integer Haar transform (Eq. 13). This transform, besides its simplicity, provides another advantage, which consists of largely preserving the dynamic range of the original domain. Note that with the integer Haar filter, the dynamic range can be expanded by only 1 bit in the details signal. In some applications, large dynamic range expansion can lead to serious problems, especially for existing devices or systems that support only a limited bitdepth [66].

### *C. Component Scalability and Computational Cost*

Table II reports the scalability of several spectral transforms in terms of the number of transformed components required to reconstruct a single original component. For the *Restricted model*, the dependencies for RWA are the same as for the wavelet transform (Haar in this case).

The computational cost of RWA is dominated by the estimation of regression coefficients $\boldsymbol{\beta} = (\mathbf{V}^\intercal\mathbf{V})^{-1}\mathbf{V}^\intercal\mathbf{W}$ and the generation of the predictions $\widehat{\mathbf{W}}=\boldsymbol{\beta}\mathbf{V}$. In the case of the *Exogenous model*, estimation of $\boldsymbol{\beta}$ is performed off-line and does not contribute to the complexity of encoding. The computational cost of RWA is at its highest when using the *Maximum model*. However, the cost of other efficient transforms such as Principal Component Analysis (PCA) is still significantly higher. Furthermore, integer implementation of the lossless PCA is usually performed by a factorization of the transform matrix [67], [68]. The resulting

Table II
COMPONENT SCALABILITY FOR DIFFERENT TRANSFORMS. THE NUMBER OF REQUIRED TRANSFORM COMPONENTS TO RECONSTRUCT ONE SINGLE ORIGINAL COMPONENT. $z$ IS THE NUMBER OF COMPONENTS AND $l$ IS THE NUMBER OF TRANSFORM LEVELS

| Method | Requirement components | $z$= 224 $l$=3 |
|---|---|---|
| RWA (*Maximum model*) | $z(1-\frac{1}{2^l})$ | 196 |
| RWA (*Restricted model*) | $l+1$ | 4 |
| RWA (*Exogenous model*) | $z(1-\frac{1}{2^l})$ | 196 |
| PCA | $z$ | 224 |
| DWT 5/3 (even component) | $2l+1$ | 7 |
| DWT 5/3 (odd component) | $3l+2$ | 11 |
| Haar | $l+1$ | 4 |

implementation is typically not computationally efficient and does not offer a high degree of parallelization due to the recursiveness of the associated reconstruction process.

Table III details the computational cost of RWA in Floating-point Operations (FLOPs), Fig. 4 compares the computational cost of different transforms applied to a typical uncalibrated image from the AVIRIS sensor.

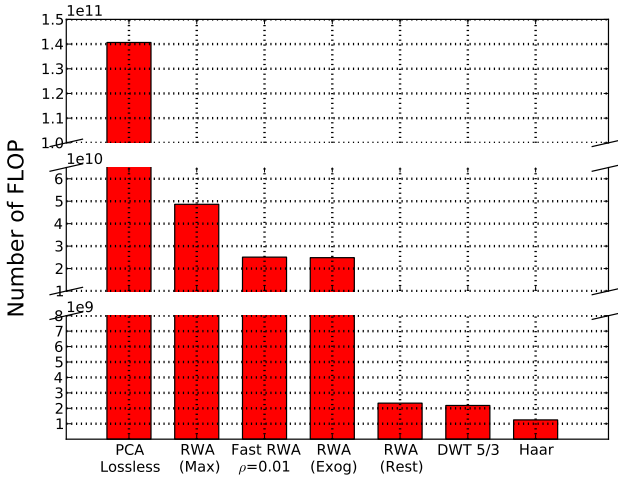



Figure 4. Cost comparison in FLOPs for different transforms applied to the uncalibrated Yellowstone image with 224 spectral components and a spatial resolution of $512 \times 680$. The values reported for PCA and DWT 5/3 are from [18].

## IV. EXPERIMENTAL RESULTS

This section presents experimental results for the proposed RWA applied to lossless coding of hyperspectral images. Comparisons are provided with the most prominent methods in the state of the art.

### A. Dataset and Coding Pipeline

Experimental evaluations were conducted using a set of images from three different sensors: AVIRIS [69], HYPERION [70] and IASI [71]. Table IV provides detailed information about these images. As described in the table, the AVIRIS images include 5 uncalibrated and 5 calibrated images corresponding to 5 scenes in Yellowstone. These images have a bit-depth of 16 bits per pixel per component (bpppc). Also included are two uncalibrated AVIRIS images (Maine and Hawaii) each having a bit depth of 12 bpppc. All of the AVIRIS images have 224 spectral components and 512 lines. The widths of these images vary between 614 and 680. The Hyperion sensor produces images with 242 spectral components each having a bit-depth of 12 bpppc and a width of 256. The number of lines varies from image to image. The Infrared Atmospheric Sounding Interferometer (IASI) is composed of a Fourier transform spectrometer and an Integrated Imaging Subsystem (IIS). IASI Level 0 images have 8359 spectral components, each having 1528 lines of width 60. Of Level 1 IASI images are of size $8461 \times 1530 \times 60$.

The proposed coding system pipeline is shown in Fig. 5. A one-dimensional transform is applied in the spectral dimension followed by 2-D JPEG2000 compression of each resulting transformed component. To this end, the Kakadu software implementation of JPEG2000 is employed with five levels of reversible 2-D DWT 5/3. When a wavelet-based transform (including the proposed RWA) is employed in the spectral dimension, the maximum possible number of levels ($\lceil \log_2(z) \rceil$) are employed. For comparison with RWA, experimental results are provided for four other reversible spectral transforms, including the Haar and 5/3 wavelet transforms, as well as PCA and POT (Pairwise Orthogonal Transform a low-complexity approximation of PCA [20]). Also for the purpose of comparison, experimental results are reported for the state-of-the-art predictive methods M-CALIC [72] and CCSDS-123 [51]. Rather than performing predictions in the transform domain as in the case of RWA, these methods perform prediction in the original (pixel) domain, and are of particular interest since both are deemed appropriate for on-board hyperspectral image coding [73].

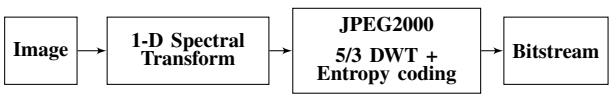



Figure 5. Proposed coding pipeline: 1-D spectral transform followed by JPEG2000 standard, which includes a 2-D spatial DWT followed by bit-plane coding.

As noted in the paragraph above, 5 levels of spatial DWT are applied as part of the 2-D JPEG2000 coding process applied to the transformed components. This was found to yield good

Table III
COMPUTATIONAL COMPLEXITY IN FLOPS FOR RWA. $z$ IS THE NUMBER OF SPECTRAL COMPONENTS, $m$ IS THE NUMBER OF SAMPLES PER COMPONENT, AND $l$ IS THE NUMBER OF WAVELET DECOMPOSITION LEVELS. $k_i$ IS THE NUMBER OF DETAILS COMPONENTS EMPLOYED IN THE PREDICTION AT LEVEL $i$.

| Operation | Floating-point operations (FLOP) for lossless RWA |
|---|---|
| Integer Haar | $8(1-\frac{1}{2^l})mz$ |
| $\boldsymbol{\beta}=(\mathbf{V}^\intercal\mathbf{V})^{-1}\mathbf{V}^\intercal\mathbf{W}$ | $\sum_{i=1}^{l}(2m-1)(k_i+1)^2+(k_i+1)^3+(\frac{z}{2^i})(k_i+1)\,[(2m-1)+(2k_i+1)]$ |
| $\widehat{\mathbf{W}}=\boldsymbol{\beta}\mathbf{V}$ | $2\sum_{i=1}^{l}(2k_i-1)m\frac{z}{2^i}$ |
| Apply and Remove | $2m(z-1)$ |

Table IV
DATASET INFORMATION FOR AVIRIS, HYPERION AND IASI SENSORS. $z$ IS THE NUMBER OF SPECTRAL COMPONENTS, $y$ IS THE HEIGHT AND $x$ IS THE WIDTH.

| Corpus | Image |
|---|---|
| AVIRIS (Uncal)<br>$z$=224, $y$=512<br>$x$=680 | Yellowstone, sc: 0, 3, 10, 11, 18<br>Hawaii ($x$=614)<br>Maine |
| AVIRIS (Cal)<br>$z$=224, $y$=512, $x$=677 | Yellowstone, sc: 0, 3, 10, 11, 18 |
| Hyperion (Uncal)<br>$z$=242<br>$x$=256 | ErtaAle ($y$=3187)<br>Lake Monona ($y$=3176)<br>Mt. St. Helens ($y$=3242) |
| Hyperion (Cal)<br>$z$=242<br>$x$=256 | Agriculture ($y$=3129)<br>Coral Reef ($y$=3127)<br>Urban ($y$=2905) |
| IASI<br>Level 0<br>$z$=8359<br>$y$=1528<br>$x$=60 | L0 1: 20091007093900Z<br>L0 2: 20091007143900Z<br>L0 3: 20100319050300S6<br>L0 4: 20120718075700Z<br>L0 5: 20130116133300Z<br>L0 6: 20130916080300Z |
| IASI<br>Level 1<br>$z$=8461<br>$y$=1530<br>$x$=60 | L1 1: 20130816230553Z<br>L1 2: 20130817004753Z<br>L1 3: 20130817041457Z<br>L1 4: 20130817055657Z<br>L1 5: 20130817073857Z |

performance for all images and all transformed techniques with a small number of notable exceptions. For the for AVIRIS and IASI images, the PCA and RWA (*Maximum Model*) spectral transforms were found to be so efficient that no spatial DWT was necessary. In fact, in these cases, application of the spatial DWT resulted in slightly inferior performance as compared to omitting the spatial DWT. For this reason, 0 levels of spatial DWT were performed as part of the 2-D JPEG2000 compression process in these cases.

Hyperion and IASI images are tall and narrow. That is, they have a large number of lines compared to columns. For this reason, experiments were performed on versions of these images that were spatially rotated 90° prior to compression. In the following section, results are reported for the rotation (0° or 90°) that provided the best performance, algorithm by algorithm. Specifically, for the uncalibrated Hyperion images, the 90° rotation was used for all algorithms except M-CALIC. For the calibrated Hyperion images, the 90° rotation was used for all algorithms except M-CALIC and CCSDS-123. For the IASI images, the 90° rotation was only used for the POT-based algorithm. In most cases, the differences between rotating and not rotating are minor. On the other hand, for CCSDS-123 and the POT-based algorithm, the differences can be significant. The POT is a line-based transform that produces side information for each line of the input image. Hence, the rotation results in a small number of lines giving rise to a reduced amount of side information.

For reproducibility of results, basic Matlab source code for our implementations of PCA, Haar, DWT 5/3 and RWA spectral transforms is available online at [74].

### B. Results

Table V provides lossless coding results for all systems under test in terms of the bitrate, in bits per pixel per component (bpppc). All necessary side information is included in the results reported in the table. The compression efficiency of each algorithm can be appreciated by observing the degree to which its resulting bitrate falls below the bit depth of the original image (e.g., 12 or 16 bpppc).

It can be observed that, overall, the proposed RWA outperforms the state-of-the-art methods included in the comparison. RWA, in one of its modalities, is the best in three out of the six data subsets (three sensors, calibrated and uncalibrated). When it is not the best, it yields results within 0.1 bpppc of the best technique. The worst such cases occur for Hyperion uncalibrated images that tend to have streaking artifacts due to the nature of the pushbroom sensors [70]. More detailed observations are now provided for each specific sensor.

For both uncalibrated and calibrated (radiance) AVIRIS images, on average, RWA (*Maximum Model*) outperforms PCA and POT by, respectively, more than 0.1 and 0.5 bpppc.

Table V
LOSSLESS BITRATE FOR HAAR, DWT 5/3, POT, PCA AND RWA SPECTRAL TRANSFORMS WITH JPEG2000; AND M-CALIC AND CCSDS-123 CODING TECHNIQUES IN TERMS OF BITS PER PIXEL PER COMPONENT (BPPPC). THE LOWER THE BITRATE, THE BETTER THE PERFORMANCE.

| Image / Transform or coding technique | Haar | DWT 5/3 | POT | PCA | RWA (Restricted) | RWA (Maximum) | Fast RWA ($\rho$) | RWA (Exogenous) | M-CALIC | CCSDS-123 |
|---|---|---|---|---|---|---|---|---|---|---|
| AVIRIS Uncalibrated ($\rho = 0.01$) | | | | | | | | | | |
| Yellowstone 00 | 7.52 | 7.16 | 6.64 | **6.05** | 6.67 | 6.07 | 6.11 | Train | 6.34 | 6.40 |
| Yellowstone 03 | 7.37 | 7.03 | 6.46 | **5.88** | 6.50 | 5.90 | 5.94 | 5.95 | 6.16 | 6.26 |
| Yellowstone 10 | 6.28 | 6.11 | 5.77 | **5.44** | 5.81 | 5.46 | 5.52 | 5.55 | 5.54 | 5.66 |
| Yellowstone 11 | 6.91 | 6.63 | 6.24 | **5.74** | 6.44 | 5.76 | 5.81 | 5.81 | 5.93 | 5.93 |
| Yellowstone 18 | 7.52 | 7.15 | 6.70 | **5.95** | 6.76 | 5.97 | 6.01 | 6.05 | 6.35 | 6.50 |
| Average 16 bpppc | 7.12 | 6.82 | 6.36 | **5.81** | 6.44 | 5.83 | 5.88 | 5.89 | 6.06 | 6.15 |
| Maine | 3.52 | 3.43 | 3.09 | 3.22 | 3.08 | **2.69** | 2.71 | Train | 2.89 | 2.77 |
| Hawaii | 3.37 | 3.27 | 2.98 | 3.20 | 2.95 | **2.55** | 2.56 | 2.82 | 2.84 | 2.70 |
| Average 12 bpppc | 3.44 | 3.35 | 3.03 | 3.21 | 3.01 | **2.62** | 2.63 | 2.68 | 2.86 | 2.73 |
| Global Average | 6.07 | 5.83 | 5.41 | 5.06 | 5.46 | **4.91** | 4.95 | 4.97 | 5.14 | 5.17 |
| AVIRIS Calibrated (Radiance, $\rho = 0.01$) | | | | | | | | | | |
| Yellowstone 00 | 5.13 | 4.75 | 4.45 | 3.85 | 4.45 | 3.74 | **3.73** | Train | 4.13 | 3.90 |
| Yellowstone 03 | 5.00 | 4.64 | 4.30 | 3.74 | 4.31 | **3.58** | 3.60 | 3.63 | 3.98 | 3.76 |
| Yellowstone 10 | 4.01 | 3.84 | 3.66 | 3.40 | 3.65 | **3.18** | 3.20 | 3.25 | 3.39 | 3.35 |
| Yellowstone 11 | 4.55 | 4.25 | 4.08 | 3.67 | 4.06 | **3.47** | 3.49 | 3.52 | 3.73 | 3.58 |
| Yellowstone 18 | 5.12 | 4.71 | 4.50 | 3.80 | 4.50 | **3.65** | 3.67 | 3.72 | 4.10 | 3.92 |
| Average 16 bpppc | 4.76 | 4.44 | 4.20 | 3.69 | 4.20 | **3.52** | 3.54 | 3.57 | 3.87 | 3.70 |
| Hyperion Uncalibrated (Rotated 90°, $\rho = 0.01$) | | | | | | | | | | |
| Erta Ale | 4.40 | 4.44 | **4.21** | 4.49 | 4.26 | 4.46 | 4.47 | Train | 4.76 | 4.24 |
| Lake Monona | 4.52 | 4.55 | 4.41 | 4.62 | 4.46 | 4.63 | 4.64 | 4.63 | 4.92 | **4.36** |
| Mt. St. Helens | 4.57 | 4.57 | 4.31 | 4.54 | 4.40 | 4.52 | 4.53 | 4.52 | 4.83 | **4.28** |
| Average | 4.54 | 4.56 | 4.31 | 4.55 | 4.38 | 4.54 | 4.55 | 4.54 | 4.83 | **4.29** |
| Hyperion Calibrated (Rotated 90°, $\rho = 0.01$) | | | | | | | | | | |
| Agricultural | 6.20 | 6.35 | 5.92 | 5.88 | 5.55 | **5.44** | **5.44** | Train | **5.44** | 5.73 |
| Coral Reef | 5.54 | 5.80 | 5.44 | 5.53 | 5.19 | 5.11 | 5.11 | 5.18 | **5.05** | 5.47 |
| Urban | 6.29 | 6.39 | 5.99 | 5.89 | 5.58 | 5.46 | 5.46 | 5.50 | **5.44** | 5.75 |
| Average | 6.01 | 6.18 | 5.78 | 5.76 | 5.44 | 5.34 | 5.34 | 5.37 | **5.31** | 5.65 |
| IASI Uncalibrated (Level 0, $\rho = 0.1$) | | | | | | | | | | |
| IASI L0 1 | 3.00 | 3.02 | 2.87 | – | 2.85 | 2.35+0.9 | 2.38+0.9 | Train | 2.89 | 2.90 |
| IASI L0 2 | 2.99 | 3.02 | 2.86 | – | 2.84 | 2.34+0.9 | 2.37+0.9 | **2.39** | 2.87 | 2.88 |
| IASI L0 3 | 2.98 | 3.01 | 2.86 | – | 2.84 | 2.35+0.9 | 2.38+0.9 | **2.40** | 2.87 | 2.90 |
| IASI L0 4 | 3.03 | 3.05 | 2.89 | – | 2.87 | 2.35+0.9 | 2.38+0.9 | **2.41** | 2.91 | 2.91 |
| IASI L0 5 | 2.96 | 2.99 | 2.86 | – | 2.89 | 2.35+0.9 | 2.38+0.9 | **2.40** | 2.86 | 2.90 |
| IASI L0 6 | 3.06 | 3.08 | 2.90 | – | 2.89 | 2.37+0.9 | 2.39+0.9 | **2.41** | 2.92 | 2.91 |
| Average | 3.00 | 3.02 | 2.87 | – | 2.85 | 2.35+0.9 | 2.38+0.9 | **2.40** | 2.88 | 2.90 |
| IASI Calibrated (Level 1, $\rho = 0.1$) | | | | | | | | | | |
| IASI L1 1 | 7.37 | 7.46 | 7.12 | 7.22+2.4 | 7.10 | 6.42+0.9 | 6.45+0.9 | Train | 6.86 | 6.60 |
| IASI L1 2 | 7.37 | 7.09 | 7.12 | 7.05+2.4 | 7.10 | 6.41+0.9 | 6.44+0.9 | **6.48** | 6.88 | 6.61 |
| IASI L1 3 | 7.38 | 7.11 | 7.14 | – | 7.13 | 6.42+0.9 | 6.45+0.9 | **6.49** | 6.88 | 6.59 |
| IASI L1 4 | 7.33 | 7.06 | 7.12 | – | 7.10 | 6.43+0.9 | 6.46+0.9 | **6,50** | 6.87 | 6.56 |
| IASI L1 5 | 7.40 | 7.12 | 7.15 | – | 7.13 | 6.43+0.9 | 6.46+0.9 | **6.50** | 6.87 | 6.60 |
| Average | 7.37 | 7.16 | 7.13 | 7.13+2.4 | 7.11 | 6.42+0.9 | 6.45+0.9 | **6.49** | 6.87 | 6.59 |

RWA also outperforms the predictive methods M-CALIC and CCSDS-123 by more than $0.2$ bpppc. When compared with the HAAR and DWT 5/3 wavelet spectral transforms, the coding gain of RWA is larger than $1.2$ and $0.9$ bpppc, respectively.

For uncalibrated Hyperion images, the coding performance of Haar, DWT 5/3, PCA and RWA (*Maximum Model*) is very similar. CCSDS-123 provides the best results, followed by POT and RWA (*Restricted Model*), which benefits from the 2-D spatial transform more than PCA or RWA (*Maximum Model*). In this case, RWA (*Restricted Model*) provides improvement over the (usual) wavelet spectral transforms by about $0.2$ bpppc with a low computational cost. For calibrated Hyperion images, RWA (*Maximum model*) achieves a coding performance similar to M-CALIC and outperforms POT and PCA by around $0.4$ bpppc. It outperforms CCSDS-123 by $0.3$ bpppc and improves over the wavelet transforms by more than $0.7$ bpppc.

For IASI images, which have more than 8000 components, lossless PCA encounters two serious problems. First is a huge increment in computational cost, which results in an implausible execution time. Second is a huge increase in the required side information. For instance, the PCA matrix for IASI Level 1 images has a size of $8461 \times 8461$, while the spatial dimension is only $1530 \times 60$, implying a side information of $2.4$ bpppc. For this reason, results for PCA applied to IASI images are only reported for a couple of examples. On the other hand, RWA (*Maximum Model*) does not suffer from extreme complexity, but side information plays a non-negligible role. This problem can be solved by using RWA (*Exogenous model*). As noted in the table, the results for this method were obtained using only one training image per sensor to learn the regression coefficients. For IASI Level 0 images, RWA *Exogenous model* outperforms POT, M-CALIC and CCSDS-123 by about $0.4$ bpppc. It is superior to the Haar and 5/3 wavelet transforms by about $0.5$ bpppc. For IASI Level 1 images, RWA *Exogenous model* achieves competitive coding results compared to CCSDS-123, while outperforming PCA, POT and wavelets by more than $0.5$ bpppc and M-CALIC by about $0.2$ bpppc.

With regards to the results for calibrated vs. uncalibrated images, we note that for the AVIRIS sensor, the calibrated images yield better performance, while for Hyperion and IASI, we find the opposite situation, where uncalibrated images achieve better performance. Similar results were observed in [20] for AVIRIS and Hyperion.

It is worth pointing out that Fast RWA is a straightforward approach to reduce the cost of computing the regression coefficients. For the AVIRIS and Hyperion sensors, only the $1\%$ of pixels were used for this computation, while achieving a performance almost identical to that obtained when using $100\%$ of the pixels. For the IASI sensor, $10\%$ of the pixels had to be used in order to approximately maintain the same performance due to the low degree of freedom ($m - z - 1$). It is also worth highlighting the fact that RWA *Exogenous model* eliminates the need for side information as well as the cost of computing the regression coefficients online.

## V. Conclusions

Remote sensing is becoming increasingly widespread and finds application in a growing number of fields. At the same time, the size of transmitted remote sensing data has increased and is foreseen to continue increasing, which begs for efficient ways to improve its storage and dissemination. Data compression plays a key role in this regard. In particular, lossless data compression has seen numerous recent advances, spanning prediction-based to transform-based techniques. In this regards, much work has focused on the spectral dimension of hyperspectral images. Principal Component Analysis usually provides the best performance among transform-based techniques, but entails some drawbacks. Several approaches have been proposed to partially tackle these drawbacks while, at the same time, yielding competitive coding performance.

In this paper we introduce a novel spectral redundancy reduction transform that builds upon the low-complexity Haar wavelet transform and exploits the remaining redundancy among wavelet-transformed components through regression analysis. The suggested Regression Wavelet Analysis (RWA) is capable of estimating, via regression, the details components from the approximation components resulting from the wavelet transform. RWA allows for an integer-to-integer implementation and perfect reconstruction, and thus, for lossless compression.

The regression model can be devised to account for finer estimation accuracy, for finer spectral component scalability, or for lower computational complexity, which give rise to the *Maximum model*, the *Restricted model*, and the *Exogenous model*, respectively. The first model involves all the approximation components in the regression; the second model involves only the approximation components as required to maintain the same component scalability as the baseline Haar wavelet transform; while the third model computes the regression coefficients based on training data from the appropriate hyperspectral sensor and employs the learned coefficients to all the other images in the same sensor. Yet another variant is a Fast RWA implementation, which performs a fast prediction using the least squares methods based only on spatially subsampled data, reducing the time to compute the regression coefficients.

Extensive experimental results for lossless compression of images from three different popular and widely used hyperspectral sensors have been carried out. Specifically, both

calibrated and uncalibrated images have been employed from the airborne AVIRIS sensor, as well as the satellite-based Hyperion and IASI sensors. The resulting coding performance suggests that RWA yields, overall, the best achievement as compared to other spectral transforms such as Haar, DWT 5/3, PCA or its low-complexity approximation POT. RWA also provides superior performance compared to other prominent prediction-based coding techniques like M-CALIC and the current CCSDS-123 standard.

To summarize, RWA provides a trade-off between computational complexity and coding performance that makes it an appealing approach for remote sensing lossless data compression. It offers additional desirable features such as limited dynamic range increase and superior component scalability.

## REFERENCES


[1] V. Laparra, S. Jiménez, D. Tuia, G. Camps-Valls, and J. Malo, "Principal polynomial analysis," *Intl. Journal of Neural Systems*, vol. 24, no. 07, p. 1440007, 2014.

[2] N. Amrani, V. Laparra, G. Camps-Valls, J. Serra-Sagrista, and J. Malo, "Lossless coding of hyperspectral images with principal polynomial analysis," in *ICIP 2014. Proceedings. International Conference on Image Processing (ICIP), 2014*, Oct 2014, pp. 4023–4026.

[3] V. Laparra, J. Malo, and G. Camps-Valls, "Dimensionality reduction via regression in hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1026–1036, Aug 2015.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[5] X. Tang and W. A. Pearlman, *Hyperspectral Data Compression*. Springer, 2006, ch. Three-Dimensional Wavelet-Based Compression of Hyperspectral Images, pp. 273–308.

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

[7] B. Penna, T. Tillo, E. Magli, and G. Olmo, "A new low complexity KLT for lossy hyperspectral data compression," *IGARSS 2006. IEEE International Conference on Geoscience and Remote Sensing Symposium, 2006.*, vol. 7, pp. 3525–3528, 2006.

[8] J. Zhang and G. Liu, "A novel lossless compression for hyperspectral images by context-based adaptive classified arithmetic coding in wavelet domain," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 461–465, Jul. 2007.

[9] 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, pp. 814–818, Oct. 2008.

[10] F. García-Vilchez and J. Serra-Sagristà, "Extending the CCSDS recommendation for image data compression for remote sensing scenarios," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, pp. 3431–3445, October 2009.

[11] 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, May 2007.

[12] ——, "Progressive 3-D coding of hyperspectral images based on JPEG2000," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 125–129, Jan. 2006.

[13] I. T. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer Verlag, 2002.

[14] I. P. Akam Bita, M. Barret, and D. Pham, "On optimal transforms in lossy compression of multicomponent images with JPEG2000," *Signal Processing*, vol. 90, no. 3, pp. 759–773, 2010.

[15] 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, 2004.

[16] P. Hao and Q. Shi, "Reversible integer KLT for progressive-to-lossless compression of multiple component images," in *ICIP 2003. Proceedings. 2003 International Conference on Image Processing, 2003.*, vol. 1, Sept 2003, pp. I–633–6 vol.1.

[17] Q. Du and J. Fowler, "Low-complexity principal component analysis for hyperspectral image compression," *International Journal of High Performance Computing Applications*, vol. 22, no. 4, pp. 438–448, 2008.

[18] I. Blanes and J. Serra-Sagristà, "Clustered reversible-KLT for progressive lossy-to-lossless 3d image coding," in *Data Compression Conf. 2009 (DCC 2009)*. IEEE Press, Mar. 2009, pp. 233–242.

[19] I. Blanes, J. Serra-Sagrista, M. Marcellin, and J. Bartrina-Rapesta, "Divide-and-conquer strategies for hyperspectral image processing: A review of their benefits and advantages," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 71–81, 2012.

[20] 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, Mar. 2011.

[21] 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, May 2011.

[22] I. P. Akam Bita, M. Barret, F. D. Vedova, and J.-L. Gutzwiller, "Lossy and lossless compression of MERIS hyperspectral images with exogenous quasi-optimal spectral transforms," *Journal of Applied Remote Sensing*, vol. 4, pp. 041 790–1–15, Jul 2010.

[23] I. P. Akam Bita, M. Barret, and D. Pham, "On optimal orthogonal transforms at high bit-rates using only second order statistics in multicomponent image coding with JPEG2000," *Signal Processing*, vol. 90, no. 3, pp. 753–758, Jan 2010.

[24] M. Barret, J.-L. Gutzwiller, I. P. Akam Bita, and F. D. Vedova, "Lossy hyperspectral images coding with exogenous quasi optimal transforms," in *Data Compression Conference, 2009. DCC'09.* IEEE, March 2009, pp. 411–419.

[25] P. Birjandi and M. Datcu, "Multiscale and dimensionality behavior of ICA components for satellite image indexing," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 1, pp. 103–107, 2010.

[26] G. Camps-Valls, D. Tuia, L. Gomez-Chova, S. Jimenez, and J. Malo, *Remote Sensing Image Processing*, ser. Synthesis Lectures on Image, Video and Multimedia Proc., A. Bovik, Ed. Austin, TX, USA: Morgan & Claypool Publ, 2011.

[27] V. Laparra and R. Santos-Rodriguez, "Spatial/spectral information trade-off in hyperspectral images," in *IEEE Intl. Geosci. Rem. Sens. Symp. (IGARSS)*, July 2015, pp. 1124–1127.

[28] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, ser. The Kluwer International Series in Engineering and Computer Science. Kluwer, 1992, no. 159.

[29] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, USA: John Wiley & Sons, 2001.

[30] X. Tang, W. A. Pearlman, and J. W. Modestino, "Hyperspectral image compression using three-dimensional wavelet coding," in *Electronic Imaging 2003*. International Society for Optics and Photonics, 2003, pp. 1037–1047.

[31] E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *Proc 31st Asilomar Conf on Signals, Systems and Computers*, vol. 1. Pacific Grove, CA: IEEE Computer Society, November 2-5 1997, pp. 673–678.

[32] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 8, no. 12, pp. 1688–1701, December 1999.

[33] J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli, "Non-linear image representation for efficient perceptual coding," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 68–80, 2006.

[34] J. Malo and J. Gutiérrez, "V1 non-linear properties emerge from local-to-global non-linear ICA," *Network: Comp. Neur. Syst.*, vol. 17, no. 1, pp. 85–102, 2006.

[35] J. Malo and V. Laparra, "Psychophysically tuned divisive normalization approximately factorizes the PDF of natural images," *Neur. Comp.*, vol. 22, no. 12, pp. 3179–3206, 2010.

[36] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[37] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, December 2000.

[38] J. Arenas-García, K. Petersen, G. Camps-Valls, and L. K. Hansen, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 16–29, 2013.

[39] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE journal*, vol. 37, no. 2, pp. 233–243, 1991.

[40] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[41] M. Scholz, M. Fraunholz, and J. Selbig, *Nonlinear principal component analysis: neural networks models and applications*. Springer, 2007, ch. 2, pp. 44–67.

[42] V. Laparra, G. Camps, and J. Malo, "Iterative gaussianization: from ICA to random rotations," *IEEE Transactions on Neural Networks*, vol. 22, no. 4, pp. 537–549, 2011.

[43] U. Ozertem and D. Erdogmus, "Locally defined principal curves and surfaces," *Journal of Machine Learning Research*, vol. 12, pp. 1249–1286, 2011.

[44] V. Laparra, S. Jimenez, G. Camps, and J. Malo, "Nonlinearities and adaptation of color vision from sequential principal curves analysis," *Neural Computation*, vol. 24, no. 10, pp. 2751–2788, 2012.

[45] V. Laparra and J. Malo, "Visual Aftereffects and Sensory Nonlinearities from a single Statistical Framework," *Front. Human Neurosci.*, vol. 9, no. 557, 2015. [Online]. Available: http://www.frontiersin.org/human_neuroscience/10.3389/fnhum.2015.00557/abstract

[46] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: existence and uniqueness results," *Neural Networks*, vol. 12, no. 3, pp. 429–439, 1999.

[47] S. Goyal and J. O Neal, "Entropy coded differential pulse-code modulation systems for television systems," *IEEE Trans. Communications*, no. 6, pp. 660–666, 1975.

[48] N. Farvadin and J. Modestino, "Rate-distortion performance of DPCM schemes for autoregressive sources," *IEEE Trans. Information Theory*, vol. 31, no. 3, pp. 402–418, 1985.

[49] M. Weinberger, G. Seroussi, and G. Sapiro, "LOCO-I: A low complexity, context-based, lossless image compression algorithm," *Proc. Data Compression Conference*, pp. 140–149, 1996.

[50] R. Rinaldo and G. Calvagno, "Image coding by block prediction of multiresolution subimages," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 909–920, 1995.

[51] Consultative Committee for Space Data Systems (CCSDS), *Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1*, ser. Blue Book. CCSDS, May 2012. [Online]. Available: http://public.ccsds.org/publications/archive/123x0b1ec1.pdf

[52] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[53] A. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Lossless image compression using integer to integer wavelet transforms," in *Image Processing, 1997. Proceedings., International Conference on*, vol. 1, Oct 1997, pp. 596–599 vol.1.

[54] J. Jing Zheng, J. Yun Fang, and C. de Han, "Reversible integer wavelet evaluation for DEM progressive compression," in *IEEE International Geoscience and Remote Sensing, Symposium, ,IGARSS 2009*, vol. 5, July 2009, pp. V–52–V–55.

[55] P. F. Craigmile and D. B. Percival, "Asymptotic decorrelation of between-scale wavelet coefficients," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 1039–1048, 2005.

[56] C. Thiebaut, E. Christophe, D. Lebedeff, and C. Latry, "Cnes studies of on-board compression for multispectral and hyperspectral images," in *Optical Engineering+ Applications*. International Society for Optics and Photonics, 2007, pp. 668 305–668 305.

[57] J. Nocedal and S. J. Wright, *Least-Squares Problems*. Springer, 2006.

[58] P. Bloomfield and W. L. Steiger, *Least absolute deviations: theory, applications, and algorithms*. Birkhäuser Boston, 1983.

[59] D. M. Bates and D. G. Watts, *Nonlinear regression: iterative estimation and linear approximations*. Wiley Online Library, 1988.

[60] T. M. Cover and J. A. Tomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[61] J. Cardoso, "Dependence, correlation and gaussianity in independent component analysis," *Journal of Machine Learning Research*, vol. 4, no. 7, pp. 1177–1203, 2003.

[62] R. Clarke, *Transform Coding of Images*. New York, USA: Academic Press, 1985.

[63] J. Malo, R. Navarro, I. Epifanio, F. Ferri, and J. Artigas, "Non-linear invertible representation for joint statistical and perceptual feature decorrelation," *Lect. Not. Comp. Sci.*, vol. 1876, pp. 658–667, 2000.

[64] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Applied and computational harmonic analysis*, vol. 3, no. 2, pp. 186–200, 1996.

[65] 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, 1998.

[66] I. Blanes, M. Hernandez-Cabronero, F. Auli-Llinas, J. Serra-Sagrista, and M. Marcellin, "Isorange pairwise orthogonal transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3361–3372, June 2015.

[67] P. Hao and Q. Shi, "Matrix factorizations for reversible integer mapping," *IEEE Transactions on Signal Processing,*, vol. 49, no. 10, pp. 2314–2324, Oct 2001.

[68] L. Galli and S. Salzo, "Lossless hyperspectral compression using KLT," *IEEE International Geoscience and Remote Sensing, Symposium, (IGARSS 2004)*, vol. 1-7, pp. 313–316, 2004.

[69] Jet Propulsion Laboratory, NASA, "Airborne Visible InfraRed Imaging Spectrometer website," http://aviris.jpl.nasa.gov/html/aviris.overview.html.

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

[71] "Infrared atmospheric sounding interferometer (IASI)," https://wdc.dlr.de/sensors/iasi/.

[72] 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, Jan. 2004.

[73] L. Santos, L. Berrojo, J. Moreno, J. Lopez, and R. Sarmiento, "Multispectral and hyperspectral lossless compressor for space applications (hyloc): A low-complexity fpga implementation of the ccsds 123 standard," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 9, no. 2, pp. 757–770, Feb 2016.

[74] N. Amrani and J. Serra-Sagristà, "RWA coding toolbox," http://gici.uab.es/GiciWebPage/downloads.php, 2015.

**Naoufal Amrani** received the B.S. degree in Mathematics from the Universitat de Barcelona in 2011. He received the M.S. degree in Computer Vision and the M.S. degree in High Performance Computer from the Universitat Autonoma de Barcelona in 2012 and 2013 respectively. Since 2013, he has been with the Group on Interactive Coding of Images, Universitat Autonoma de Barcelona, where he is currently a PhD student. He was the recipient of the Capocelli Prize for the 2016 Data Compression Conference.

**Joan Serra-Sagristà** (S'97–M'05–SM'11) received the Ph.D. degree in computer science from Universitat Autònoma Barcelona (UAB), Spain, in 1999. He is currently an Associate Professor at Department of Information and Communications Engineering, UAB. From September 1997 to December 1998, he was at University of Bonn, Germany, funded by DAAD. His current research interests focus on data compression, with special attention to image coding for remote sensing and telemedicine applications. He serves as Associate Editor of IEEE Trans. on Image Processing. He has co-authored over one hundred publications. He was the recipient of the Spanish Intensification Young Investigator Award in 2006.

**Valero Laparra** was born in València (Spain) in 1983, and received a B.Sc. degree in Telecommunications Engineering (2005), a B.Sc. degree in Electronics Engineering (2007), a B.Sc. degree in Mathematics degree (2010), and a PhD degree in Computer Science and Mathematics (2011). He is a postdoc in the Image Processing Laboratory (IPL) at Universitat de València. More details in `http://www.uv.es/lapeva/`.

**Michael W. Marcellin** (S'81–M'87–SM'93–F'02) received the B.S. in Electrical Engineering from San Diego State University in 1983, and the MS and PhD in Electrical Engineering from Texas A&M University in 1985 and 1987, respectively. Since 1988, Dr. Marcellin has been with the University of Arizona, where he currently holds the title of Regents' Professor, and is the International Foundation for Telemetering Chaired Professor. His research interests include digital communication and data storage systems, data compression, and signal processing. He has authored or coauthored more than two hundred publications in these areas. He has received numerous honors, including six teaching awards.

**Jesus Malo** received the M.Sc. degree in Physics in 1995 and the Ph.D. degree in Physics in 1999 both from the Universitat de Valncia. He was the recipient of the Vistakon European Research Award in 1994. In 2000 and 2001 he worked as Fulbright Postdoc at the Vision Group of the NASA Ames Research Center (A.B. Watson), and at the Lab of Computational Vision of the Center for Neural Science, New York University (E.P. Simoncelli). He came back to the NYU in 2013 for a semester. Currently, he serves as Associate Editor of IEEE Trans. Im. Proc. He is with the Image and Signal Processing Group and the Visual Statistics Group (VI(S)TA) at the Universitat de València. He is member of the Asociación de Mujeres Investigadoras y Tecnólogas (AMIT).

# Chapter 4

# Regression Wavelet Analysis for Progressive Lossy-to-Lossless Coding

# Regression Wavelet Analysis for Progressive-Lossy-to-Lossless Coding of Remote-Sensing Data

Naoufal Amrani*, Joan Serra-Sagristà*, Miguel Hernández-Cabronero‡ and Michael Marcellin†

*Dep. of Information and Communications Eng., Universitat Autònoma de Barcelona, Spain
‡Dep. of Computer Science , University of Warwick, UK
†Dep. of Electrical and Computer Engineering, University of Arizona, USA

**Abstract**

Regression Wavelet Analysis (RWA) is a novel wavelet-based scheme for coding hyperspectral images that employs multiple regression analysis to exploit the relationships among spectral wavelet-transformed components. The scheme is based on a pyramidal prediction, using different regression models, to increase the statistical independence in the wavelet domain. For lossless coding, RWA has proven to be superior to other spectral transform like PCA and to the best and most recent coding standard in remote sensing, CCSDS-123.0. In this paper we show that RWA also allows progressive lossy-to-lossless (PLL) coding and that it attains a rate-distortion performance superior to those obtained with state-of-the-art schemes. To take into account the predictive significance of the spectral components, we propose a Prediction Weighting scheme for JPEG2000 that captures the contribution of each transformed component to the prediction process.

## 1 Introduction

Remote-sensing data has become enormously important for a wide range of applications in many different fields of Earth's observation thanks to the technological evolution of optical sensors over the last years. As recent sensors can cover large areas of the Earth's surface, producing images with unprecedented spectral and spatial resolution, the need for efficient coding techniques to improve the capabilities of storage and transmissions emerges naturally. While there are some scenarios where only lossless coding is acceptable, there are other applications where lossy coding is indispensable because only a limited amount of data can be transmitted. Techniques allowing progressive lossy-to-lossless coding are thus gaining importance in remote-sensing as they provide similar rate-distortion performance to pure lossy methods at low bitrates, and also a competitive coding performance in the lossless regime.

Most efficient coding schemes for remote-sensing data include decorrelating transforms to exploit the spatial correlation between adjacent pixels within a component and the spectral correlation between adjacent components. Some schemes apply 3D transforms [1, 2], commonly based on wavelets, to handle the joint correlation in both spatial and spectral dimensions. Other somewhat different schemes apply separately a 1D spectral transform followed by a 2D spatial transform, using the most convenient one in each dimension. Generally, the correlation removal along the spectral dimension is considered critical in coding hyperspectral images due to its significant impact on the coding performance [3].

In particular, Principal Component Analysis (PCA) is the optimal decorrelating transform for Gaussian sources [4, 5]. It is widely applied for multicomponent images as a

spectral transform since it also yields an excellent coding performance. However, PCA is a data-dependent transform, entailing the need to compute it for each individual image before its application. As a consequence, its computational complexity is substantial due to the covariance matrix calculation, the extraction of eigenvectors, the matrix factorization and the integer mapping (integer implementation is needed for lossless coding [6, 7]). Hence, in scenarios where the data has a significant size in the spectral dimension, such as IASI images –with more than 8000 components–, PCA and its integer implementation are not feasible. Additionally, PCA is not a component-scalable transform, i.e., the recovery of any single image component depends on every transformed component.

Since the complexity reduction of PCA is demanded, a number of approaches have been proposed, including (i) approaches based on spatially and/or spectrally sub-sampling the dataset and estimating the covariance matrix using a reduced subset of coefficients [3, 8]; (ii) approaches based on divide-and-conquer strategies to approximate the PCA with a reduced computational complexity and some amount of component scalability [9, 10]; and (iii) approaches that consist in learning the transform on a set of images of one particular sensor and apply the learned transform on new images from the same sensor [11]. The the techniques of family (i) only alleviate the complexity of computing the covariance matrix while other complexity sources still remain, such as the integer mapping for lossless coding. On the other hand, the techniques of families (ii) and (iii) might produce significant deterioration in the coding performance compared to the full-complexity PCA.

Beyond computational issues, another limitation of classical PCA is its linear nature. Even though linear PCA yields competitive performance in a wide range of applications because of its invertibility and energy compaction, its efficiency is hampered by data exhibiting non-linear relationships. Thus, despite the simplicity and efficiency that linearity provides, non-linear methods could provide performance improvements by better exploiting data dependencies. Recently, several non-linear generalization of PCA have been proposed to deal with data of non-linear nature [4, 12, 13]. In the adaptation of these non-linear methods to image coding, there are two major considerations: first, the transform must be invertible; and second, the computational complexity and the memory consumption should be reasonable. These two requirements eliminate a large number of candidate approaches.

In [14] we explored lossless hyperspectral image coding using curvilinear techniques based on Principal Polynomial Analysis (PPA) [13]. In that work, it was shown that PPA achieves higher energy compaction and statistical independence than PCA. However, in practical hyperspectral image coding, appropriate handling of side information and the sequential error introduced by the integer mapping dramatically penalize the coding gain. Additionally, the computational complexity and memory requirements of these non-linear transforms were even larger than those of the original PCA, which is already computationally expensive [4].

Alternatively, rather than using sophisticated non-linear methods for remote-sensing coding, we focus here on theoretically suboptimal but simpler traditional transforms (such as the Discrete Wavelet Transform) while adopting predictive techniques [15, 16] to exploit any remaining *post transform* statistical dependence.

Following this line of work, in [17] we introduced the Regression Wavelet Analysis (RWA) transform for lossless coding. This transform can be seen as a predictive scheme to reduce redundancy after wavelet analysis has been performed in the spectral dimension of

hyperspectral images. RWA shares the principal properties of the DWT such as component scalability and low complexity, while yielding excellent performance for lossless coding. Specifically, RWA yields superior performance compared to that of the best spectral decorrelation techniques like PCA, and also compared to the most recent and most competitive prediction-based hyperspectral coding technique, CCSDS-123.0 standard [18].

In this paper, we extend the benefits of RWA to progressive lossy-to-lossless coding of remote-sensing data by considering the predictive significance of the transformed components. To address this issue, we propose a weighting scheme that assigns a weight to each spectral component depending on its contribution to the prediction process. Weighting strategies are useful to improve the quality of the recovered data by allocating more bits to the significant coefficients for an algorithm design [19]. Unlike pure lossless coding, where it is not necessary to consider these weights to achieve the aforementioned performance when using RWA, in progressive lossy-to-lossless the loss of information at intermediate or low bitrates may have an error propagation effect for coding schemes based on pyramidal prediction like RWA.

This paper is organized as follows. Section 2 introduces briefly the RWA algorithm and the regression model. Section 3 puts forward the proposed Prediction Weighting scheme. Section 4 reports the experimental results. Finally, section 5 concludes.

## 2 Regression Wavelet Analysis

This section begins with a review of the DWT, providing the necessary notation. It then goes on to describe the Regression Wavelet Analysis algorithm and the regression model.

### *2.1 Discrete Wavelet Transform*

Let us suppose that an original multicomponent image $\mathbf{V}^0$ has $z{=}2^d$ spectral components with each component having $m$ spatial samples. Then we write $\mathbf{V}^0 \in \mathbb{R}^{m\times z}$ and

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

The wavelet representation of $\mathbf{V}^0$ with $J$ levels, for $1 \leq J \leq \log_2(z)$, is given by

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

where the one level DWT decomposition of each $\mathbf{V}^{j-1}$ is given by

$$\mathbf{DWT}(\mathbf{V}^{j-1}, 1) = \left(\mathbf{V}^j, \mathbf{W}^j\right) \tag{2}$$

At each scale or level $j$ the signal $\mathbf{V}^{j-1}{\in}\mathbb{R}^{m\times(z\cdot 2^{-j+1})}$ is decomposed into the approximation signal $\mathbf{V}^j{\in}\mathbb{R}^{m\times(z\cdot 2^{-j})}$ and the details signal $\mathbf{W}^j{\in}\mathbb{R}^{m\times(z\cdot 2^{-j})}$ at half resolution each. The approximation signal $\mathbf{V}^j$ usually contains most of the information of the previous signal $\mathbf{V}^{j-1}$, whereas the details signal $\mathbf{W}^j$ contains the information difference between $\mathbf{V}^{j-1}$ and $\mathbf{V}^j$. The decomposition is usually repeated in cascade form on the approximation signal $\mathbf{V}^j$ as shown in Fig. 1.

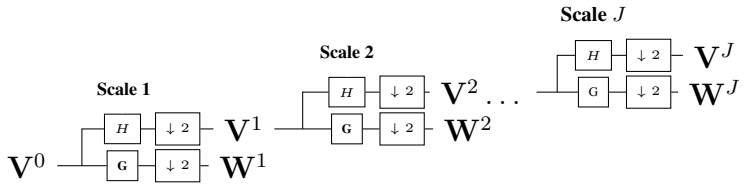



Figure 1: DWT decomposition with $J$ levels.

### *2.2 Regression Wavelet Analysis*

The Regression Wavelet Analysis (RWA) scheme generalizes the discrete wavelet transform (DWT) by applying regression to tackle the correlation that still remains in the DWT domain. At each scale $j$, each details component $\mathbf{W}^j(i) = \mathbf{W}_i^j \in \mathbb{R}^{m\times 1}$ is predicted from the information contained in the approximation components $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ within the same scale $j$. This prediction $\widehat{\mathbf{W}}_i^j = f_i(\mathbf{V}^j)$ is then removed to obtain a residual

$$\mathbf{R}^j = \mathbf{W}^j - \widehat{\mathbf{W}}^j. \tag{3}$$

As illustrated in Fig. 2, the resulting RWA affects only the details components at each level of the transform. The approximation components are unchanged from those of the DWT.

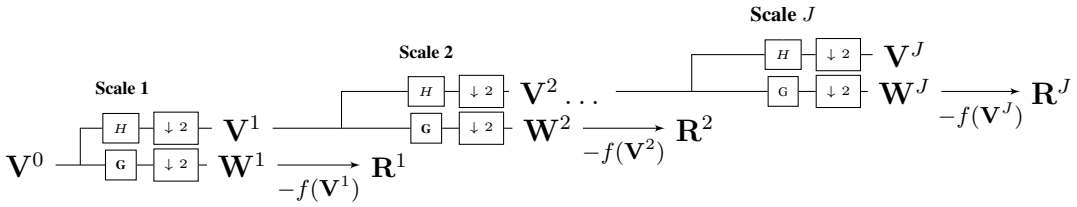



Figure 2: RWA representation with $J$ levels.

Following the same notation as in (Eq. 2) and (Eq. 1), the one level and the $J$ level RWA transform are denoted, respectively, as

$$\mathbf{RWA}(\mathbf{V}^{j-1}, 1) = \left(\mathbf{V}^j, \mathbf{R}^j\right) \tag{4}$$

$$\mathbf{RWA}(\mathbf{V}^0, J) = \left(\mathbf{V}^J, (\mathbf{R}^j)^{1\le j\le J}\right). \tag{5}$$

The RWA transform is easily inverted. At each scale $j$ the estimate $\widehat{\mathbf{W}^j}$ is computed from $\mathbf{V}^j$. The approximation at level $j-1$ is then obtained by computing

$$\mathbf{W}^j = \mathbf{R}^j + \widehat{\mathbf{W}^j} \tag{6}$$

$$\mathbf{V}^{j-1} = \mathbf{DWT}^{-1}\left((\mathbf{V}^j, \mathbf{W}^j), 1\right). \tag{7}$$

In general, RWA can be based on any wavelet filter, nevertheless, as we focus on progressive lossy-to-lossless (PLL) coding, and for faster implementation, hereafter, we will refer to RWA as the one based on integer Haar filter (S-transform) [20], which produces integer coefficients $(\mathbf{V}^j, \mathbf{W}^j)$ with a quantized estimation removal $\mathbf{R}^j = \mathbf{W}^j - Q(\widehat{\mathbf{W}}^j)$, where the quantization is applied through a simple rounding operation.

### *2.3 Regression Model*

Different prediction models have been proposed for RWA. The most general model is called the *Maximum model*. It employs all of the approximation components from $\mathbf{V}^j$ for the computation of the prediction of each details component $\mathbf{W}_i^j$. It is given by

$$\widehat{\mathbf{W}}_i^j = f_i\left[\mathbf{V}^j\right] = \boldsymbol{\beta}_{i,0}^j + \boldsymbol{\beta}_{i,1}^j \mathbf{V}_1^j + \cdots + \boldsymbol{\beta}_{i,k}^j \mathbf{V}_k^j, \quad \mathbf{V}_i^j \in \mathbb{R}^{m\times 1}. \tag{8}$$

For the *Maximum model*, the regression coefficients $\boldsymbol{\beta}^j$ are found for each individual image using the least-squares method [21] that minimizes the sum of squares of the distances between the original components and the estimated ones, i.e.,

$$\textit{min: } \|\mathbf{W}_i^j - \widehat{\mathbf{W}}_i^j\|_2.$$

Since *Maximum model* has a non-negligible computational complexity, to alleviate its cost we resort to the application of an approach in line with those of family (iii) (see Sect. 1). The proposed *Exogenous model* computes the regression coefficients $\overline{\boldsymbol{\beta}}^j$ over a set of training images, and then uses these coefficients for all new images that come from the same sensor. As a consequence, the computational cost of RWA for *Exogenous model* is reduced considerably to roughly the same as that of the DWT. Additionally, no side information needs to be stored for each individual image. The model is given by

$$\widehat{\mathbf{W}}_i^j = f_i\left[\mathbf{V}^j\right] = \overline{\boldsymbol{\beta}}_{i,0}^j + \overline{\boldsymbol{\beta}}_{i,1}^j \mathbf{V}_1^j + \cdots + \overline{\boldsymbol{\beta}}_{i,k}^j \mathbf{V}_k^j, \quad \mathbf{V}_i^j \in \mathbb{R}^{m\times 1}. \tag{9}$$

## 3 Prediction Weighting Scheme

According to the RWA design, the residuals $\mathbf{R}^j$ not only contribute to reconstruct the details components $\mathbf{W}^j$ (Eq. 6), but are also used to generate the estimates $\widehat{\mathbf{W}}^{j-1}$ in the next inverse steps. As a result, the distortion introduced to a given residual component $\mathbf{R}_i^j$ at scale $j$ will affect the reconstruction in all finer scales $j-1, \cdots, 1$.

The Prediction Weighting scheme that we propose aims at emphasizing the contribution of each spectral residual to the prediction process by assigning weights that capture the predictive significance. These weights will be used as input values to JPEG2000 to control the relative significance of the spectral components [19].

Consider the RWA representation $\left(\mathbf{V}^J, (\mathbf{R}^j)^{1\le j\le J}\right)$ at scale $J$. To invert the transform, we need to generate the estimates $(\widehat{\mathbf{W}}^j)^{1\le j\le J}$ used in (Eq. 6) and (Eq. 7). Let us analyze how the residuals $\mathbf{R}^j$ of each level $j$ contribute to the generation of these estimates. Given

that $\widehat{\mathbf{W}}_i^j = f_i\left[\mathbf{V}^j\right]$ and using (Eq. 7)

$$\begin{aligned}\widehat{\mathbf{W}}_i^{j-1} =& f_i\left[\mathbf{V}^{j-1}\right] = f_i\left[\mathbf{DWT}^{-1}\left(\mathbf{V}^j, \mathbf{W}^j\right)\right] \\ \widehat{\mathbf{W}}_i^{j-1} =& f_i\left[\mathbf{DWT}^{-1}\left(\mathbf{V}^j, \widehat{\mathbf{W}}^j + \mathbf{R}^j\right)\right].\end{aligned} \tag{10}$$

Applying (Eq. 10) for $j-2$

$$\widehat{\mathbf{W}}_i^{j-2} = f_i\left[\mathbf{DWT}^{-1}\left(\mathbf{V}^{j-1}, \widehat{\mathbf{W}}^{j-1} + \mathbf{R}^{j-1}\right)\right]. \tag{11}$$

If we substitute (Eq. 10) in (Eq. 11)

$$\widehat{\mathbf{W}}_i^{j-2} = f_i\left[\mathbf{DWT}^{-1}\left(\mathbf{V}^{j-1}, f_i\left[\mathbf{DWT}^{-1}\left(\mathbf{V}^j, \widehat{\mathbf{W}}^j + \mathbf{R}^j\right)\right] + \mathbf{R}^{j-1}\right)\right]. \tag{12}$$

By induction,

$$\begin{aligned}\widehat{\mathbf{W}}_i^1 = f_i\big[\mathbf{DWT}^{-1}\big(\mathbf{V}^2, f_i\big[\mathbf{DWT}^{-1}\big(\mathbf{V}^3, \cdots\cdots \\ \cdots\ f_i\left[\mathbf{DWT}^{-1}\left(\mathbf{V}^{J-1}, f_i\left[\mathbf{DWT}^{-1}\left(\mathbf{V}^J, \widehat{\mathbf{W}}^J + \mathbf{R}^J\right)\right] \cdots \mathbf{R}^3\right)\right] + \mathbf{R}^2\right)\right].\end{aligned} \tag{13}$$

From (Eq. 13) it can be seen that the generation of each estimate $\widehat{\mathbf{W}}_i^j$ at scale $j$ requires all the residuals of the coarser scales $\left(\mathbf{R}^k\right)^{j+1\leq k\leq J}$. In other words, each residual $\mathbf{R}_i^j$ at scale $j{>}1$ is involved in the prediction of all the estimates $(\widehat{\mathbf{W}}^k)^{1\leq k\leq j-1}$ in the finer scales $j-1, \cdots, 1$. Additionally, the estimates $\widehat{\mathbf{W}}^k{\in}\mathbb{R}^{m\times(z\cdot 2^{-k})}$ in each scale $k$ are formed by $z\cdot 2^{-k}$ components (Eq. 2).

Given that, let us consider a function $N(\cdot)$ that relates the residuals $\mathbf{R}^j$ with the number of predictions in which it is involved, i.e., the number of components of $\widehat{\mathbf{W}}^k{\in}\mathbb{R}^{m\times(z\cdot 2^{-k})}$ with $1\leq k\leq j-1$

$$N(\mathbf{R}^j) = \begin{cases} z\cdot\sum_{i=1}^{j-1}\left(2^{-i}\right) & ,\text{if } j>1 \\ 0 & ,\text{if } j=1 \end{cases}. \tag{14}$$

Then the relative proportion with respect to the overall number of components $z$ will be

$$PN(\mathbf{R}^j) = \frac{N(\mathbf{R}^j)}{z} = \begin{cases} \sum_{i=1}^{j-1} 2^{-i} & ,\text{if } j>1 \\ 0 & ,\text{if } j=1 \end{cases}. \tag{15}$$

We now define the weights for $\mathbf{R}^j$ as

$$\mathcal{W}(\mathbf{R}^j) = \frac{1}{PN(\mathbf{R}^j) - PN(\mathbf{R}^{j-1})} = 2^{j-1}. \tag{16}$$

Finally, the approximation $\mathbf{V}^J$ of RWA (in the case of applying the maximum possible number of levels $J{=}\log_2(z)$) consists of only one component $\mathbf{V}^J{\in}\mathbb{R}^{m\times 1}$, which is weighted by

$$\mathcal{W}(\mathbf{V}^J) = 2^J. \tag{17}$$

In this way, $\mathcal{W}(\cdot)$ weights the transformed components $\left(\mathbf{V}^J, (\mathbf{R}^j)^{1\leq j\leq J}\right)$ by different quantities, taking into account the predictive significance of each component and the components of the previous scale. These prediction weights play a role similar to that of the subband weighting factors in spatial or temporal wavelet transforms [19].

## 4 Experimental results

In this section we evaluate the RWA transform with the proposed Prediction Weighting scheme (*Weighted RWA*) on a set of hyperspectral images from two different sensors: AVIRIS [22] and IASI [23]. For comparison purposes, we provide results for other reversible spectral transforms including RWA without weighting, PCA and DWT 5/3.

The coding system pipeline is shown in Fig. 3 and illustrates the compression process. The compared spectral transforms are first applied in the spectral dimension followed by a 2D JPEG2000 compression, which may include a reversible 2D DWT 5/3 transform with 5 levels in the spatial dimension and a bitplane-based entropy coding. The spatial transform is not applied for PCA and RWA. In these cases, the spatial transform provides no improvement, as most of the transformed components already have low energy. The Kakadu software implementation of JPEG2000 has been used. For *Weighted RWA*, the weights were introduced using the option **Cweight**.

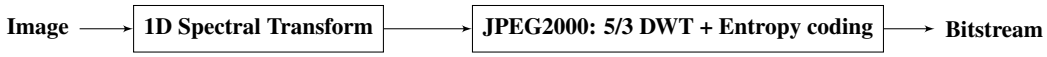
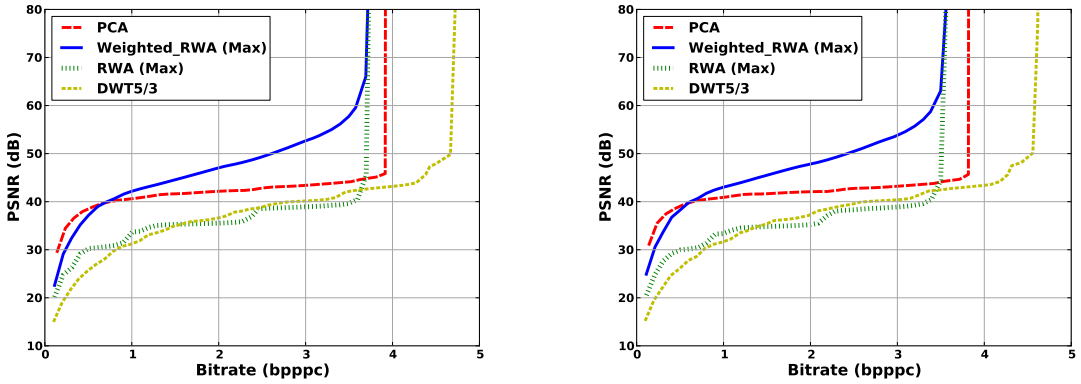



Figure 3: 1D spectral transform followed by JPEG2000 standard.



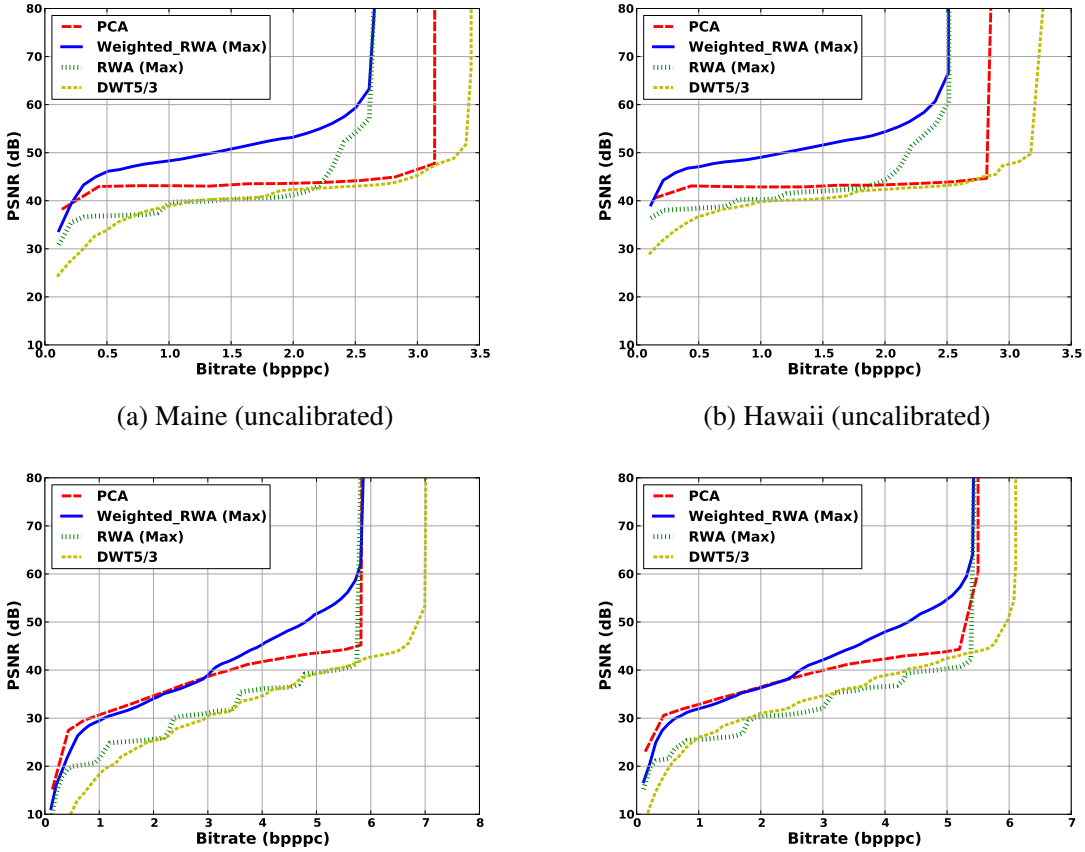

(a) Yellowstone 00 (radiance)

(b) Yellowstone 03 (radiance)

Figure 4: PLL Rate-Distortion for AVIRIS (radiance) with 224 spectral components, $512\times677$ spatial resolution and 16 bpppc bitdepth.

Figures 4, 5 and 6 provide the Rate-Distortion performance for the compared spectral transforms. These figures provide the relation between the target bitrate, measured in bits per pixel per component (bpppc) (including all required side information), and the reconstruction quality, reported as Peak Signal-to-Noise Ratio (PSNR), measured in $dB$.

As a general trend, *Weighted RWA* performs significantly better than RWA without weighting for progressive lossy-to-lossless coding. Additionally, RWA without weighting




(a) Maine (uncalibrated)

(b) Hawaii (uncalibrated)

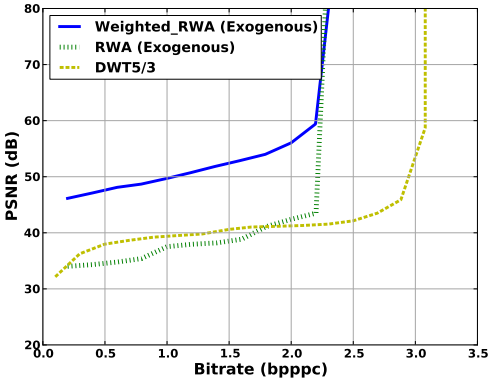
(c) Yellowstone 03 (uncalibrated)

(d) Yellowstone 10 (uncalibrated)

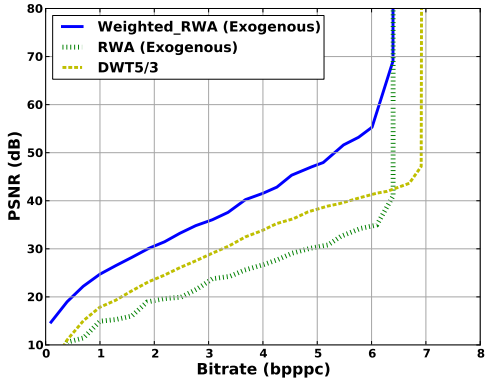
Figure 5: PLL Rate-Distortion for AVIRIS (uncalibrated) with 224 spectral components and $512\times680$ spatial resolution ($512\times614$ for Hawaii). Bitdepth: 16 bpppc for Yellowstone and 12 bpppc for Maine and Hawaii.

is heavily affected by the error propagation at all bitrates (low, intermediate and high bitrates), which produces multiple plateaus in the Rate-Distortion performance. *Weighted RWA* is also superior to the DWT 5/3 spectral transform, which has a performance similar to that of the RWA without weighting. In comparison to PCA, the performance of *Weighted RWA* for 16 bpppc AVIRIS calibrated (resp. uncalibrated) images is somewhat worse up to a bitrate of 0.7 (resp. 2.8) bpppc; from then onwards, PCA falls clearly behind, with a difference of over 8 $dB$ at 3 (resp. 5) bpppc. For 12 bpppc AVIRIS uncalibrated images, *Weighted RWA* is the most competitive spectral transform.

With regard to IASI images, as they have a large size in the spectral dimension (more than 8000 components), reversible PCA results in a prohibitive computational cost due to the recursiveness of the integer mapping [24]. For this reason, coding results for PCA on IASI images are not reported. The results of *Weighted RWA* and classical RWA (without weighting) are provided using the *Exogenous model* due to the non-negligible role of the side information (coefficients $\boldsymbol{\beta}$). It is worth noting that the *Exogenous coefficients* $\overline{\boldsymbol{\beta}}^{j}_{i,k}$




(a) 20091007143900Z (IASI L0).




(b) 20130817004753Z (IASI L1).

Figure 6: PLL Rate-Distortion for IASI images. IASI L0 and L1 have 8359 and 8561 spectral components, 1528 and 1530 lines respectively. The width for both is 60 columns.

were computed using only one training image from IASI Level 0 and another one from IASI Level 1. The analysis of the coding performance of the *Exogenous model* for PLL coding is similar to that reported before: *Weighted RWA* is clearly superior to RWA without weighting and to DWT 5/3.

## 5 Conclusions

We proposed a Prediction Weighting scheme for progressive lossy-to-lossless coding of hyperspectral images with Regression Wavelet Analysis algorithm. The suggested weights are able to capture the relative significance of the transformed spectral components in the prediction process, minimizing the effect of error propagation and providing a steady quality evolution. At a much lower computational complexity than reversible PCA, our proposal yields the most competitive rate-distortion coding performance published to date.

## Acknowledgements

This work has been partially supported by the Spanish Government (MINECO), by FEDER, by the Catalan Government and by Universitat Autònoma de Barcelona, under Grants TIN2015- 71126-R, TIN2012-38102-C03-03, 2014SGR-691, and UAB-PIF-472-03-1/2012.

## References

[1] X. Tang and W. A. Pearlman, *Hyperspectral Data Compression*. Springer, 2006, ch. Three-Dimensional Wavelet-Based Compression of Hyperspectral Images, pp. 273–308.

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

[3] 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, May 2007.

[4] I. T. Jolliffe, *Principal Component Analysis*. Berlin, Germany: Springer Verlag, 2002.

[5] 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, 2004.

[6] P. Hao and Q. Shi, "Reversible integer KLT for progressive-to-lossless compression of multiple component images," in *ICIP 2003. Proceedings. 2003 International Conference on Image Processing, 2003.*, vol. 1, Sept 2003, pp. I–633–6 vol.1.

[7] L. Galli and S. Salzo, "Lossless hyperspectral compression using KLT," *IEEE International Geoscience and Remote Sensing, Symposium, (IGARSS 2004)*, vol. 1-7, pp. 313–316, 2004.

[8] Q. Du and J. Fowler, "Low-complexity principal component analysis for hyperspectral image compression," *International Journal of High Performance Computing Applications*, vol. 22, no. 4, pp. 438–448, 2008.

[9] I. Blanes, J. Serra-Sagrista, M. Marcellin, and J. Bartrina-Rapesta, "Divide-and-conquer strategies for hyperspectral image processing: A review of their benefits and advantages," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 71–81, 2012.

[10] 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, Mar. 2011.

[11] 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, May 2011.

[12] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[13] V. Laparra, S. Jiménez, D. Tuia, G. Camps-Valls, and J. Malo, "Principal polynomial analysis," *Intl. Journal of Neural Systems*, vol. 24, no. 07, p. 1440007, 2014.

[14] N. Amrani, V. Laparra, G. Camps-Valls, J. Serra-Sagrista, and J. Malo, "Lossless coding of hyperspectral images with principal polynomial analysis," in *ICIP 2014. Proceedings. International Conference on Image Processing (ICIP), 2014*, Oct 2014, pp. 4023–4026.

[15] S. Goyal and J. O Neal, "Entropy coded differential pulse-code modulation systems for television systems," *IEEE Trans. Communications*, no. 6, pp. 660–666, 1975.

[16] N. Farvadin and J. Modestino, "Rate-distortion performance of DPCM schemes for autoregressive sources," *IEEE Trans. Information Theory*, vol. 31, no. 3, pp. 402–418, 1985.

[17] N. Amrani, J. Serra-Sagrista, V. Laparra, M. Marcellin, and J. Malo, "Regression wavelet analysis for lossless coding of remote-sensing data," 2015, manuscript submitted for publication to IEEE Transactions on Geoscience and Remote Sensing.

[18] Consultative Committee for Space Data Systems (CCSDS), *Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1*, ser. Blue Book. CCSDS, May 2012.

[19] M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards, and Practice*. Springer Science & Business Media, 2002, vol. 1.

[20] M. Adams, F. Kossentini, and R. Ward, "Generalized S transform," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2831–2842, Nov 2002.

[21] J. Nocedal and S. J. Wright, *Least-Squares Problems*. Springer, 2006.

[22] Jet Propulsion Laboratory, NASA, http://aviris.jpl.nasa.gov/html/aviris.overview.html.

[23] "Infrared atmospheric sounding interferometer (IASI)," https://wdc.dlr.de/sensors/iasi/.

[24] Y. She, P. Hao, and Y. Paker, "Matrix factorizations for parallel integer transformation," *IEEE Transactions on Image Processing*, vol. 54, no. 12, pp. 4675–4684, 2006.

# Chapter 5

# Unbiasedness of Regression Wavelet Analysis for Progressive Lossy-to-Lossless

# Unbiasedness of Regression Wavelet Analysis for Progressive Lossy-to-Lossless Coding

Naoufal Amrani, Joan Serra-Sagristà
Universitat Autònoma de Barcelona

Michael Marcellin
University of Arizona

***Abstract*—The recently proposed Regression Wavelet Analysis (RWA) scheme holds a great promise as a spectral transform for compressing hyperspectral images due to its low complexity, reversibility, and demonstrated superior coding performance. The scheme is based on a pyramidal prediction, using multiple regression analysis, to exploit statistical dependence in the wavelet domain. For lossless coding, RWA has proven to offer better performance than other spectral transforms like reversible PCA and also better than the best and most recent lossless coding standard in remote sensing, CCSDS-123.0. For progressive lossy-to-lossless coding, RWA also yields improved performance as compared to PCA. In this paper, we show that the RWA parameters are unbiased for lossy coding, where the regression models are used not with the original transformed components, but with the recovered ones, which lack some information due to the lossy reconstruction. As a byproduct, we also report that the *Exogenous* RWA model, a variant of RWA where the employed regression parameters have been trained using other images from the same sensor, still provides almost identical performance to that obtained by applying regression on the same image, thus showing that hyperspectral images with large sizes in the spectral dimension can be coded via RWA without side information and at a lower computational cost.**

## I. Introduction

Regression Wavelet Analysis (RWA) is a novel predictive scheme that we introduced in [1] to encode hyperspectral images. The method exploits the residual dependence among the data coefficients after wavelet analysis has been performed along the spectral dimension. Specifically, RWA generalizes the Discrete Wavelet Transform (DWT) by employing multiple regression analysis within a pyramidal scheme to tackle the dependency that still remains in the DWT domain. In the RWA scheme, any wavelet filter and any regression model might be used, with the reversible Haar (S-transform) and the *Maximum Model* being the proposed choices for the *general* RWA due to the reversibility and simplicity of the S-transform and due to the prediction accuracy provided by the *Maximum Model*. Other variants for the regression model have been proposed: *Fast* RWA reduces the computational complexity of the prediction by using a small percentage of the data to learn the regression parameters, and *Exogenous* RWA eliminates the need for side information by using regression parameters previously learned on trained data.

RWA and its variants share the principal properties of the DWT: low computational complexity, high component scalability, and competitive performance for lossless and progressive lossy-to-lossless coding. Moreover, for lossless coding, RWA outperforms the best comparable decorrelation techniques like PCA and the most competitive prediction-based hyperspectral coding technique, such as CCSDS-123.0. For progressive lossy-to-lossless (PLL), RWA also yields superior performance thanks to the Predictive Weighting Scheme (PWS) that we proposed in [2], which captures the predictive significance of the RWA coefficients. In contrast to other predictive methods [3], the Weighting Scheme allows RWA to provide improved performance not only at high bit-rates but also at intermediate and low bit-rates.

RWA uses the least squares method to compute the optimal regression parameters $\beta$ that minimize the sum of squared differences between the response component and its prediction provided by the regression model. In lossy coding, these parameters might be suboptimal, since the model components (regressors) are recovered with lack of some information (reconstruction error), which typically leads to a bias in the least squares parameters. By bias, we mean the difference between the regression parameters $\beta$ computed having the original transformed components and the regression parameters $\beta^*$ computed by regressing the response component on the available recovered components. In this case, the regression parameters $\beta^*$ are those that minimize the distance between the response component and its prediction, leading to a smaller error. The extent of the bias depends on the regression model and the reconstruction error distribution. This bias can be measured by comparing the reconstruction quality in the cases of using $\beta$ and $\beta^*$ to check whether $\beta$ are unbiased estimates of $\beta^*$ or not. This issue can be considered as a special case of measurement error problems in statistics studies [4], [5].

In this paper, we show that the regression parameters of RWA are unbiased when encoding at low bit-rate (high compression ratio) due to the low variance of the RWA components; hence, parameter adjustment has a negligible impact on coding performance. Following the same line, we illustrate the unbiasedness of *Exogenous* RWA for progressive lossy-to-lossless coding. We show that the trained parameters of only one image applied to many others from the same sensor yield almost identical compression performance to that of the model using the true parameters, while the computational complexity is significantly reduced and the need of side information is removed.

This paper is organized as follows. Section II introduces briefly the RWA algorithm, the regression model and the Predictive Weighting Scheme. Section III argues the unbiasedness of RWA. Section IV provides the experiment results and

Section V concludes the paper.

## II. Regression Wavelet Analysis

To provide the necessary notation, we begin with a review of the DWT, then we go on to describe RWA, the regression model and finally the Predictive Weighting Scheme.

### A. Discrete Wavelet Transform

Let us suppose that an original multicomponent image $\mathbf{V}^0$ has $z{=}2^d$ spectral components with each component having $m$ spatial samples. Then we write $\mathbf{V}^0 \in \mathbb{R}^{m\times z}$ and

$$\mathbf{V}^0{=}\left[\mathbf{V}_1^0,\ldots,\mathbf{V}_z^0\right], \quad \mathbf{V}_i^0 \in \mathbb{R}^{m\times 1}.$$

The wavelet representation of $\mathbf{V}^0$ with $J$ levels, for $1{\leq}J{\leq}\log_2(z)$, is given by

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

where the one level DWT decomposition of each $\mathbf{V}^{j-1}$ is given by

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

At each scale or level $j$, the signal $\mathbf{V}^{j-1} \in \mathbb{R}^{m\times(z\cdot 2^{-j+1})}$ is decomposed into the approximation signal $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ and the details signal $\mathbf{W}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ at half resolution each. The decomposition is usually repeated in cascade form on the approximation signal $\mathbf{V}^j$.

### B. Regression Wavelet Analysis

The RWA scheme generalizes the DWT by applying regression to tackle the correlation that still remains in the DWT domain. At each scale $j$, each details component $\mathbf{W}_i^j \in \mathbb{R}^{m\times 1}$ is predicted from the information contained in the approximation components $\mathbf{V}^j \in \mathbb{R}^{m\times(z\cdot 2^{-j})}$ within the same scale $j$. This prediction $\widehat{\mathbf{W}}_i^j = f_i(\mathbf{V}^j)$ is then removed to obtain a residual

$$\mathbf{R}^j = \mathbf{W}^j - \widehat{\mathbf{W}}^j. \tag{3}$$

As illustrated in Fig. 1, the resulting RWA affects only the details components at each level of the transform. The approximation components are unchanged from those of the DWT.

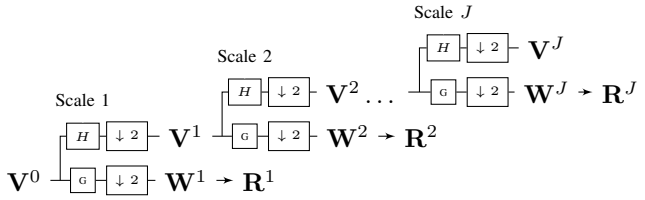



Fig. 1: RWA decomposition with $J$ levels.

Following the same notation as in (1) and (2), the one level and the $J$ level RWA transform are denoted, respectively, as

$$\mathbf{RWA}(\mathbf{V}^{j-1}, 1) = \left(\mathbf{V}^j, \mathbf{R}^j\right), \tag{4}$$

$$\mathbf{RWA}(\mathbf{V}^0, J) = \left(\mathbf{V}^J, (\mathbf{R}^j)^{1\leq j\leq J}\right). \tag{5}$$

The RWA transform is easily inverted. At each scale $j$ the estimate $\widehat{\mathbf{W}^j}$ is computed from $\mathbf{V}^j$ through $\widehat{\mathbf{W}}_i^j = f_i(\mathbf{V}^j)$. The approximation at level $j-1$ is then obtained by computing

$$\mathbf{W}^j = \mathbf{R}^j + \widehat{\mathbf{W}^j}, \tag{6}$$

$$\mathbf{V}^{j-1} = \mathbf{DWT}^{-1}\left((\mathbf{V}^j, \mathbf{W}^j), 1\right). \tag{7}$$

RWA can be based on any wavelet filter. One choice is the S-Transform [6], which operates with integer coefficients $\left(\mathbf{V}^j, \mathbf{W}^j\right)$, requiring that a quantized estimation $\mathbf{W}^j = \mathbf{R}^j + Q(\widehat{\mathbf{W}^j})$ be employed in (3) and (6). Here, the quantization is applied through a simple rounding operation.

### C. Regression Model

Different prediction models have been proposed for RWA. The most general model is called the *Maximum model*. It employs all the approximation components from $\mathbf{V}^j$ for the computation of the prediction of each details component $\mathbf{W}_i^j$. It is defined as

$$\widehat{\mathbf{W}}_i^j = \beta_{i,0}^j + \beta_{i,1}^j\mathbf{V}_1^j + \cdots + \beta_{i,k}^j\mathbf{V}_k^j, \quad \mathbf{V}_i^j \in \mathbb{R}^{m\times 1}. \tag{8}$$

For the *Maximum model*, the regression coefficients $\boldsymbol{\beta}^j$ are found for each individual image using the least-squares method [7] that minimizes the distances between the original components and the estimated ones, i.e., *min* $\|\mathbf{W}_i^j - \widehat{\mathbf{W}}_i^j\|_2$.

The *Exogenous model* computes the regression coefficients $\overline{\boldsymbol{\beta}}^j$ over a set of training images, and then uses these coefficients for all new images that come from the same sensor. As a consequence, the computational cost of RWA for the *Exogenous model* is considerably reduced. Additionally, no side information (the regression parameters) needs to be stored for each individual image. The model is defined as

$$\widehat{\mathbf{W}}_i^j = \overline{\beta}_{i,0}^j + \overline{\beta}_{i,1}^j\mathbf{V}_1^j + \cdots + \overline{\beta}_{i,k}^j\mathbf{V}_k^j, \quad \mathbf{V}_i^j \in \mathbb{R}^{m\times 1}. \tag{9}$$

### D. Predictive Weighting Scheme

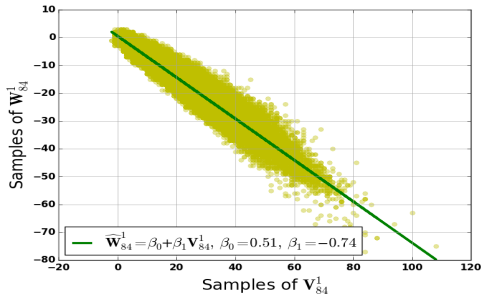
The Predictive Weighting Scheme assigns weights $\omega(.)$ to each component from the RWA in order to capture its significance for the rate-distortion optimization in lossy coding. The weights are defined as follows:

$$\omega(\mathbf{R}^j) = 2^{j-1} \quad \text{and} \quad \omega(\mathbf{V}^j) = 2^j. \tag{10}$$

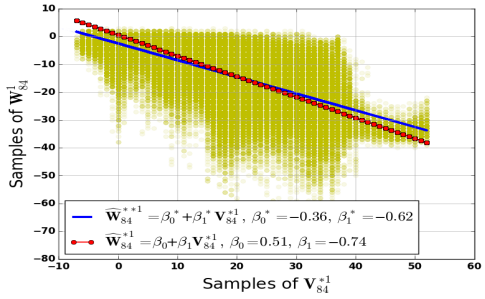
## III. Unbiasedness of RWA parameters in lossy coding

The regression goal in the RWA scheme is to maximize prediction accuracy. The ordinary least squares (OLS) provides unbiased and consistent parameters that minimize the Euclidean distance between a response component $\mathbf{W}_i^j$ and its prediction $\widehat{\mathbf{W}}_i^j$. In lossless coding, the RWA parameters have shown a strong unbiasedness for a given sensor via the *Exogenous model*, since the regression parameters $\overline{\boldsymbol{\beta}}^j$ of only one training image applied to many other images from the same sensor yield almost the same coding performance as the true parameters $\boldsymbol{\beta}^j$ [1], [2].

In this section we analyze the unbiasedness of the RWA parameters for lossy coding, where the regression model is




(a) Scatter of $\mathbf{V}_i^j$ and $\mathbf{W}_i^j$ ($r^2$=0.94) with the regression line.




(b) Scatter of $\mathbf{V}_i^{*j}$ and $\mathbf{W}_i^j$ ($r^2$=0.5) with two regression lines.

Fig. 2: Yellowstone sc 00 (radiance), $j$=1 and $i$=84.

used with the recovered components $\mathbf{V}^{*j}$ instead of the true ones $\mathbf{V}^j$ due to error introduced by the lossy reconstruction. Specifically, we analyze if the loss of information in the RWA components requires any adjustment of parameters in order to provide the best prediction accuracy using the available recovered components $\mathbf{V}^{*j}$. To illustrate this point, let us consider an example of simple linear regression with two variables. Fig. 2a depicts the relation between one details component $\mathbf{W}_i^j$ and one approximation component $\mathbf{V}_i^j$ with the regression line computed using the least squares method that finds $\boldsymbol{\beta}=(\beta_0,\beta_1)$. Fig. 2b depicts the relationship between the same details component $\mathbf{W}_i^j$ and the recovered approximation component $\mathbf{V}_i^{*j}$ after a lossy compression with a bit-rate of 0.1 bpp. Also included in the figure are two regression lines. The first (in blue) corresponds to the conditional mean $\widehat{\mathbf{W}}_i^{**j}$ as computed by regressing $\mathbf{W}_i^j$ on $\mathbf{V}_i^{*j}$ resulting in the parameters $\boldsymbol{\beta}^*=(\beta_0^*,\beta_1^*)$. The second regression line (in red) is plotted using the parameters $\boldsymbol{\beta}$, and is denoted by $\widehat{\mathbf{W}}_i^{*j}$. The second plot differs from the regression line in part (a) of the figure only in that it is "sampled" due to the quantization of $\mathbf{V}_i^j$. As can be seen, the $\boldsymbol{\beta}$ parameters yield biased estimates compared to those of the $\boldsymbol{\beta}^*$ parameters. Therefore $\widehat{\mathbf{W}}_i^{**j}$ leads to a smaller error than $\widehat{\mathbf{W}}_i^{*j}$. Note that correlation is significantly reduced by quantization. In particular, the squared correlation coefficient $r^2$ changes from 0.93 between $\mathbf{V}_i^j$ and $\mathbf{W}_i^j$ to 0.44 between $\mathbf{V}_i^{*j}$ and $\mathbf{W}_i^j$.

Consider the $J$ level RWA representation $\left(\mathbf{V}^J,(\mathbf{R}^j)^{1\le j\le J}\right)$, where $J=\log_2 z$. For this value of $J$, the approximation and the residual at level $J$ consist of exactly one component each, $\mathbf{V}^J,\mathbf{R}^J\in\mathbb{R}^{m\times 1}$. Let $\left(\mathbf{V}^J,(\mathbf{R}^{*j})^{1\le j\le J}\right)$ be the recovered RWA representation after encoding with a low bit-rate using the Predictive Weighting Scheme [2]. Here, $\mathbf{R}^j=\mathbf{R}^{*j}+\mathbf{U}^j$, where $\mathbf{U}^j$ is error introduced in $\mathbf{R}^j$ via lossy compression. We assume that the *single* component $\mathbf{V}^J$ is preserved losslessly. In what follows, we describe the RWA inverse.

### A. Inverse at scale $J$

In the first level of the inverse transform, we deal with two components, $\mathbf{V}^J,\mathbf{R}^{*J}\in\mathbb{R}^{m\times 1}$ and we perform two operations.

*1) **Estimates Generation**:* The first operation consists of generating $\widehat{\mathbf{W}}^J$ via:

$$\widehat{\mathbf{W}}^J=\beta_0^J+\beta_1^J\mathbf{V}^J. \tag{11}$$

Since we assume that the single approximation component $\mathbf{V}^J$ is decompressed without error, there is no need to adjust the regression parameters at this level. We then recover $\mathbf{W}^{*J}=\widehat{\mathbf{W}}^J+\mathbf{R}^{*J}$ instead of $\mathbf{W}^J=\widehat{\mathbf{W}}^J+\mathbf{R}^J$.

*2) **Wavelet Inverse**:* In the second operation, we invert the S-transform to obtain the two approximation components:

$$\mathbf{V}_1^{*J-1}=\mathbf{V}^J-\frac{\mathbf{W}^{*J}}{2}=\mathbf{V}_1^{J-1}+\frac{\mathbf{U}^J}{2}$$
$$\mathbf{V}_2^{*J-1}=\mathbf{V}^J+\frac{\mathbf{W}^{*J}}{2}=\mathbf{V}_2^{J-1}-\frac{\mathbf{U}^J}{2}$$

The expressions here and below are approximate, since we ignore the effect of rounding in the lifting steps for the sake of simplicity.

### B. Inverse at scale $J-1$

In the second level of the inverse transform, we deal with four components of scale $J-1$: the two approximation components from the previous level, $\mathbf{V}^{*J-1}\in\mathbb{R}^{m\times 2}$, and two residual components, $\mathbf{R}^{*J-1}\in\mathbb{R}^{m\times 2}$.

*1) **Estimates Generation**:* First, we generate the details estimates $\widehat{\mathbf{W}}^{**J-1}$ as:

$$\widehat{\mathbf{W}}_i^{**J-1}=\beta_{i,0}^{*J-1}+\beta_{i,1}^{*J-1}\mathbf{V}_1^{*J-1}+\beta_{i,2}^{*J-1}\mathbf{V}_2^{*J-1}. \tag{12}$$

We then recover $\mathbf{W}^{*J-1}$ instead of $\mathbf{W}^{J-1}$, where:

$$\mathbf{W}^{*J-1}=\widehat{\mathbf{W}}^{**J-1}+\mathbf{R}^{*J-1}.$$

*2) **Wavelet Inverse**:* We apply the inverse S-transform to the four components described above to obtain $\mathbf{V}^{*J-2}\in\mathbb{R}^{m\times 4}$. It is interesting to study the relationship between the parameters $\boldsymbol{\beta}^{*J-1}$ and $\boldsymbol{\beta}^{J-1}$, which are computed based on the available components $\mathbf{V}^{*J-1}$ and the original ones $\mathbf{V}^{J-1}$, respectively. In what follows, we denote by Var and Cov the variance and covariance, respectively. We can assume that $\mathrm{Cov}(\mathbf{V}^{*J-1},\mathbf{R}^{*J-1})=0$, since $\mathrm{Cov}(\mathbf{V}^{J-1},\mathbf{R}^{J-1})=0$ is satisfied by the least squares estimation in the forward transform. Under this assumption, the parameters $\boldsymbol{\beta}^{*J-1}$ can be derived as a function of the parameters $\boldsymbol{\beta}^{J-1}$ as described next.

First, let us assume that losses are introduced in only one component. Then the model that we want to fit is as follows:

$$\widehat{\mathbf{W}}_i^{***J-1} = \beta_{i,0}^{**J-1} + \beta_{i,1}^{**J-1}\mathbf{V}_1^{*J-1} + \beta_{i,2}^{**J-1}\mathbf{V}_2^{J-1}. \quad (13)$$

It is worth noting that the parameter $\beta_{i,2}^{**J-1}$ will be different from $\beta_{i,2}^{J-1}$, even though $\mathbf{V}_2^{J-1}$ is not changed, since correlation between $\mathbf{V}_1^{J-1}$ and $\mathbf{V}_2^{J-1}$ exists in most cases [4], [5]. Then, the least squares estimation of the parameters, computed as in [4], [5], will be:

$$\beta_{i,1}^{**J-1} = \lambda_1 \beta_{i,1}^{J-1}, \quad \lambda_1 = \frac{\mathrm{Var}(\mathbf{V_1}|\mathbf{V_2})}{\mathrm{Var}(\mathbf{V_1^*}|\mathbf{V_2})} \quad (14)$$

$$\beta_{i,2}^{**J-1} = \beta_{1,2}^{J-1} + \beta_{1,1}^{J-1}(1-\lambda_1)\gamma_1 \quad (15)$$

where $\mathrm{Var}(\mathbf{V_1}|\mathbf{V_2})$ is the residual variance of $\mathbf{V}_1^{J-1}$ estimated from $\mathbf{V}_2^{J-1}$ and $\gamma_1$ is the coefficient of $\mathbf{V}_2^{J-1}$ in the regression of $\mathbf{V}_1^{J-1}$ on $\mathbf{V}_2^{J-1}$.

Now, when introducing loss in both components of the approximation, (14) and (15) can be used in another step to yield:

$$\beta_{1,1}^{*J-1} = \beta_{1,1}^{J-1}\lambda_1 + \left(\beta_{1,2}^{J-1} + \beta_{1,1}^{J-1}(1-\lambda_1)\gamma_1\right)(1-\lambda_2)\gamma_2 \quad (16)$$

$$\beta_{1,2}^{*J-1} = \left(\beta_{1,2}^{J-1} + \beta_{1,1}^{J-1}(1-\lambda_1)\gamma_2\right)\lambda_2 \quad (17)$$

where $\lambda_2 = \frac{\mathrm{Var}(\mathbf{V_2}|\mathbf{V_1^*})}{\mathrm{Var}(\mathbf{V_2^*}|\mathbf{V_1^*})}$ and $\gamma_2$ is the coefficient of $\mathbf{V}_1^{*J-1}$ in the regression of $\mathbf{V}_2^{J-1}$ on $\mathbf{V}_1^{*J-1}$.

The attenuation factors $\lambda_i$ depend on the variance of the error introduced by lossy compression, which can be represented as the variance of $\mathbf{U}^J$, for instance:

$$\lambda_1 = \frac{\mathrm{Var}(\mathbf{V_1}|\mathbf{V_2})}{\mathrm{Var}(\mathbf{V_1^*}|\mathbf{V_2})} = \frac{\mathrm{Var}(\mathbf{V_1}|\mathbf{V_2})}{\mathrm{Var}(\mathbf{V_1}|\mathbf{V_2}) + \mathrm{Var}(\frac{\mathbf{U}^J}{2})}. \quad (18)$$

Hence, if $\mathrm{Var}(\frac{\mathbf{U}^J}{2})=0$, then $\lambda_1=\lambda_2=1$, and $\boldsymbol{\beta}^{*J-1}=\boldsymbol{\beta}^{J-1}$, as expected. This leads to an interesting conclusion for any residual subband that has a very low variance. Lossy compression systems typically result in such subbands being quantized to 0, i.e., $\mathbf{R}^{*J}=0$, or $\mathbf{U}^J=\mathbf{R}^J$. This in turn implies that $\mathrm{Var}(\frac{\mathbf{U}^J}{2})$ is very small so that $\lambda_1, \lambda_2 \simeq 1$ and $\boldsymbol{\beta}^{*J-1} \approx \boldsymbol{\beta}^{J-1}$.

### C. The General case: Adjusted RWA

In general, for lossy coding the bias in the regression parameters of RWA is highly related with the variance of its components $\mathbf{R}^j$. Small variances of $\mathbf{R}^j$ will lead to a small bias in the regression parameters. Fig. 3 shows the variances, sorted in descending order, of the RWA components compared to those of PCA, Haar and the original image for an AVIRIS image with 224 spectral components. It can be seen that almost all the RWA components have a low variance close to zero. From these variance values, we can expect a small bias in the regression parameters. The parameters bias is not easily described when dealing with multiple regressors (predictor components). Nevertheless, in lossy coding, the introduced loss is known at the encoder. Hence, at each compression with a given target bit-rate, the regression parameters of RWA $\boldsymbol{\beta}^j$ can be adjusted via equations (14) to (18) to give $\boldsymbol{\beta}^{*j}$ in order to use the regression model with the least squares parameters

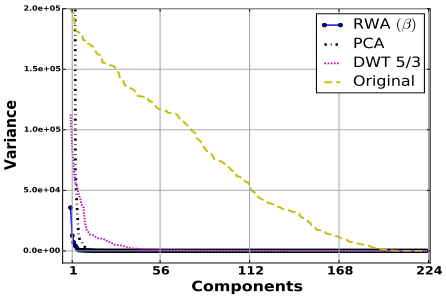



Fig. 3: Components variance for different spectral transform.

that minimize the sum of the squared distances between $\mathbf{W}^j$ and the prediction $\widehat{\mathbf{W}}^{**j}$. We refer to this adjustment as *Adjusted RWA*, which can be performed pyramidally by applying the two basic inverse operations (6) and (7) beginning at scale $J$.

## IV. Results

In this section we evaluate the unbiasedness of the RWA parameters for progressive lossy-to-lossless coding over a set of hyperspectral images from different sensors: calibrated AVIRIS [8] (224 spectral components) and uncalibrated IASI [9] (8359 spectral components). We provide the Rate-Distortion (R-D) performances of *Adjusted* RWA ($\boldsymbol{\beta}^*$), *general* RWA ($\boldsymbol{\beta}$) and *Exogenous* RWA ($\overline{\boldsymbol{\beta}}$). For comparison purposes, we also provide the results for other reversible transforms, including PCA and DWT 5/3. The coding system pipeline is

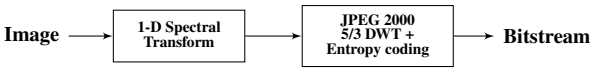
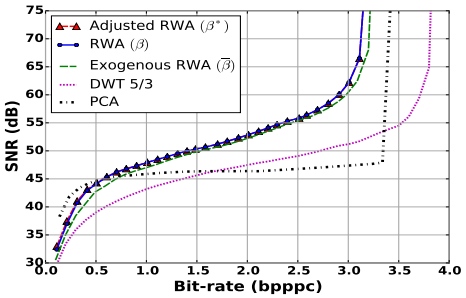



Fig. 4: Coding pipeline.

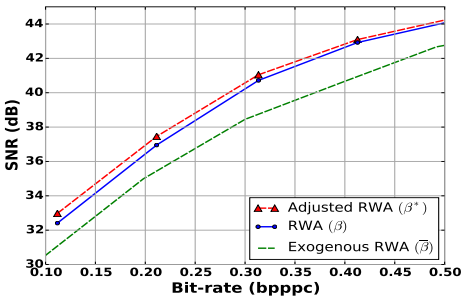
shown in Fig. 4. First, the spectral transform is applied, then a JPEG 2000 compression is carried out, which may include a reversible 2D DWT 5/3 transform with 5 levels in the spatial dimension. The spatial transform is not applied for RWA and PCA, since it does not provide any improvement due to the low energy of the components in these cases. The Kakadu software implementation has been used for JPEG 2000.

Fig. 5, 6 and 7 provide the Rate-Distortion performance for the compared spectral transforms. The Rate-Distortion provides the relation between the bit-rate in bits per pixel per component (bpppc) and the Signal-to-Noise Ratio (SNR), expressed in dB and computed as follows: $SNR=10\log_{10}(\frac{\sigma^2}{MSE})$, where $\sigma^2$ is the variance of the original image and $MSE$ is the mean square error. As training images for *Exogenous* RWA, Yellowstone sc 00 has been used for AVIRIS and IASI L0-*20091007093900Z* for the IASI sensor.

Theoretically, *Adjusted* RWA gives the best prediction, since for each target bit-rate, the parameters $\boldsymbol{\beta}$ are adjusted to give the optimal parameters $\boldsymbol{\beta}^*$. We observe that *general* RWA does yield diminished performance for progressive lossy-to-lossless




(a) Rate-Distortion performance.




(b) Zoom

Fig. 5: Yellowstone sc 10 (radiance)

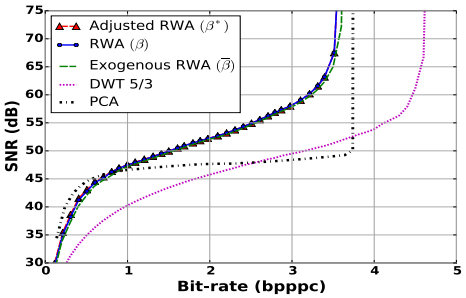



Fig. 6: Yellowstone sc 03 (radiance)

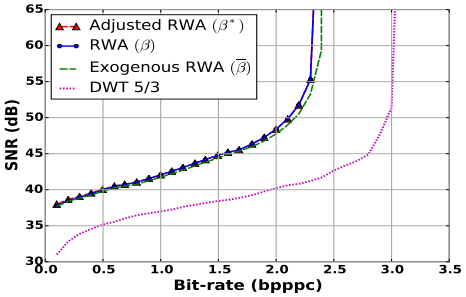



Fig. 7: IASI L0 20091007143900Z. The side information (about 0.9 bpppc) is not included for RWA and Adjusted RWA.

coding compared to *Adjusted* RWA. However, the performance sacrifice is small (less than about 0.2 dB) which can be seen by zooming in on the rate-distortion curves for low bit-rates (see Fig. 5b).

As a side result, *Exogenous* RWA yields diminished performance (less than about 2.0 dB) compared to *Adjusted* and *general RWA*. It is worth noting that *Exogenous RWA* does not require any side information and is computationally cheaper. This advantage is more relevant for images with large size in the spectral dimension, like IASI images with 8359 components, since the side information has a non-negligible role (about 0.9 bpppc). Side information has not been included in Fig. 7 for *Adjusted* or *general RWA* to facilitate a comparison with *Exogenous RWA*. Another major advantage is that *Exogenous* parameters $\overline{\beta}$, known by the decoder, were computed using only one image from each sensor. For brevity, we only report results for three images. Finally, the superiority of RWA over the spectral transforms PCA and DWT 5/3 is maintained by *Exogenous* RWA.

## V. Conclusions

In this paper we have shown that the regression parameters of RWA for progressive lossy-to-lossless coding are nearly unbiased. We show that correcting for the small existing bias can result in small performance increases. Additionally, we explain why *Exogenous* RWA provides a very competitive coding performance for large hyperspectral images at a lower computational cost and without side information.


## Acknowledgement

This work has been partially supported by the Spanish Government (MINECO), by FEDER, by the Catalan Government and by Universitat Autònoma de Barcelona, under Grants TIN2015-71126-R, TIN2012-38102-C03, 2014SGR-691, and UAB-PIF-472-03-1/2012.


## References


[1] N. Amrani, J. Serra-Sagrista, V. Laparra, M. Marcellin, and J. Malo, "Regression wavelet analysis for lossless coding of remote-sensing data," *IEEE Transactions on Geoscience and Remote Sensing, vol. 54, in press*, 2016.

[2] N. Amrani, J. Serra-Sagrista, M. Hernandez-Cabronero, and M. Marcellin, "Regression Wavelet Analysis for Progressive-Lossy-to-Lossless Coding of Remote-Sensing Data," in *Data Compression Conference*, March 2016.

[3] N. Amrani, V. Laparra, G. Camps-Valls, J. Serra-Sagrista, and J. Malo, "Lossless coding of hyperspectral images with principal polynomial analysis," in *ICIP 2014. Proceedings. International Conference on Image Processing (ICIP), 2014*, Oct 2014, pp. 4023–4026.

[4] R. J. Carroll, "Measurement error in epidemiologic studies," *Encyclopedia of biostatistics*, 1998.

[5] R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, *Measurement error in nonlinear models: a modern perspective*. CRC press, 2006.

[6] M. Adams, F. Kossentini, and R. Ward, "Generalized S transform," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2831–2842, Nov 2002.

[7] J. Nocedal and S. J. Wright, *Least-Squares Problems*. Springer, 2006.

[8] Jet Propulsion Laboratory, NASA. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.overview.html

[9] "Infrared atmospheric sounding interferometer (IASI)," https://wdc.dlr.de/sensors/iasi/.

# Chapter 6

# Low Complexity Prediction Model for Coding Remote-Sensing Data with Regression Wavelet Analysis

# Low Complexity Prediction Model for Coding Remote-Sensing Data with Regression Wavelet Analysis

Naoufal Amrani*, Joan Serra-Sagristà* and Michael Marcellin†

*Dep. of Information and Communications Eng., Universitat Autònoma de Barcelona, Spain
†Dep. of Electrical and Computer Engineering, University of Arizona, USA

## Abstract

Fast and efficient coding techniques are being increasingly required to meet the complexity restrictions of on-board satellite compression. The recently proposed Regression Wavelet Analysis (RWA) has proven to be highly effective as a spectral transform for coding remote sensing images. The algorithm is based on a pyramidal prediction, using multiple regression analysis, to tackle residual data dependencies in the wavelet domain. RWA combines low complexity and reversibility and has demonstrated competitive performance for lossless and progressive lossy-to-lossless compression superior to the state-of-the-art predictive-based CCSDS-123.0 and the widely used transform-based principal component analysis (PCA). In this paper we introduce a very low-complexity RWA approach, where prediction is based on only a few components, while the performance is maintained. When RWA computational complexity is taken to an extremely low level, careful model selection is necessary. Contrary to expensive selection procedures, we propose a simple and efficient strategy called *neighbor selection* for using small regression models. On a set of well-known and representative hyperspectral images, these small models maintain the excellent coding performance of RWA, while reducing the computational cost by about 90%.

## 1 Introduction

Recent optical remote sensing satellites provide a wealth of data, containing useful information that significantly contributes to a wide range of Earth Observation applications. Compression techniques aim at reducing large amounts of data to meet limited satellite storage and downlink capabilities, and also the speed/complexity requirements of some on-board applications. This gives rise to the need for low complexity, fast and efficient lossless and lossy compression techniques for remote sensing data.

In general, exploiting the data dependencies along the spectral dimension plays a key role for coding remote sensing data. This is, in most of the state-of-the-art coding techniques, conducted through either a transform or a prediction. For transform-based techniques, principal component analysis (PCA) is widely used due to its efficiency, as it is optimal in decorrelating Gaussian sources [1]. However, its computational complexity is substantial and it entails storing and transmitting a non-negligible amount of side information. Meanwhile, other transforms like the discrete wavelet transform (DWT) feature low complexity, but their decorrelation capabilities are suboptimal. On the other hand, for lossless or near lossless coding, predictive methods can yield a coding performance better than that provided by transform-based methods [2].

Recently, we proposed Regression Wavelet Analysis (RWA), a novel predictive scheme to encode hyperspectral images [3–5]. The method exploits, through regression analysis,

the residual dependence among the data coefficients after wavelet analysis has been performed along the spectral dimension. Any wavelet filter and any regression model might be used in a RWA scheme, with reversible Haar (S-transform) and the so-called *Maximum Model* being the proposed choices for the *general* RWA scheme. The *Maximum Model* includes all image components in the prediction and yields high accuracy.

At a much lower computational cost than PCA, the RWA algorithm yields the most competitive performance published to date (see, e.g., [3] and [4]) for for both lossless and progressive lossy-to-lossless compression. Nevertheless, the computational cost of RWA can be further reduced. The computational cost of a spectral decorrelation method is associated with the number of spectral components. In particular, the computational cost of RWA is dominated by the size of the regression model, i.e., the number of components used in the prediction. With that in mind, small models might be more convenient than, for instance, the *Maximum Model*.

On a related note, multicollinearity is usually present among the predictor components, i.e., the regressors are usually highly correlated. In general, determining the model structure is crucial in any practical data modelling problem, and a fundamental principle is that the model should be no more complex than what is required to capture the underlying data generating mechanisms. This concept, known as the parsimonious principle, is particularly relevant in nonlinear data modelling because the size of a nonlinear model can easily become explosively large [6]. However, as the goal of the regression in our algorithm is the prediction accuracy, this multicollinearity is not a harmful problem, since it does not affect the ability of the model to predict [7]. That said, multicollinearity strongly argues for using a small model that only includes a few important components and which still provides an accurate prediction as close as possible to that of the *Maximum Model*.

In this paper we address the problem of selecting small models with few components for RWA algorithm. Rather than using expensive selection strategies, we focus on simpler but efficient ones, based on divide and conquer strategies. We propose a *neighbor selection* approach, a regression model that uses a few adjacent spectral components to predict a given detail component.

This paper is organized as follows. Section 2 briefly reviews the RWA algorithm. Section 3 discusses the model selection and introduces our *neighbor selection* strategy. Section 4 provides experimental results and Section 5 concludes the paper.

## 2 Regression Wavelet Analysis

This section reviews Regression Wavelet Analysis.

### *2.1 Discrete Wavelet Transform*

Let an original multicomponent image $\mathbf{V}^0$ have $z{=}2^d$ spectral components, with each component having $m$ spatial samples. Then $\mathbf{V}^0 \in \mathbb{R}^{m\times z}$ and

$$\mathbf{V}^0{=}\left[\mathbf{V}_1^0,\ldots,\mathbf{V}_z^0\right], \quad \mathbf{V}_i^0 \in \mathbb{R}^{m\times 1}.$$

The wavelet representation of $\mathbf{V}^0$ with $J$ levels, for $1 \leq J \leq d = \log_2(z)$, is given by

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

where the one level DWT decomposition of each $\mathbf{V}^{j-1}$ is given by

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

At each scale or level $j$, signal $\mathbf{V}^{j-1} \in \mathbb{R}^{m \times (z \cdot 2^{-j+1})}$ is decomposed into approximation signal $\mathbf{V}^j \in \mathbb{R}^{m \times (z \cdot 2^{-j})}$ and details signal $\mathbf{W}^j \in \mathbb{R}^{m \times (z \cdot 2^{-j})}$ at half resolution each.

### *2.2 Regression Wavelet Analysis*

The Regression Wavelet Analysis (RWA) scheme generalizes the discrete wavelet transform (DWT) by applying regression to tackle any dependency that still remains in the wavelet domain. At each scale $j$, each details component $\mathbf{W}_i^j \in \mathbb{R}^{m \times 1}$ is predicted from the information contained in the approximation components $\mathbf{V}^j \in \mathbb{R}^{m \times (z \cdot 2^{-j})}$ within the same scale $j$. This prediction $\widehat{\mathbf{W}}_i^j = f_i(\mathbf{V}^j)$ is then removed to obtain a residual

$$\mathbf{R}^j = \mathbf{W}^j - \widehat{\mathbf{W}}^j. \tag{3}$$

As illustrated in Fig. 1, the resulting RWA affects only the details components at each level of the transform. The approximation components are unchanged from those of the DWT.

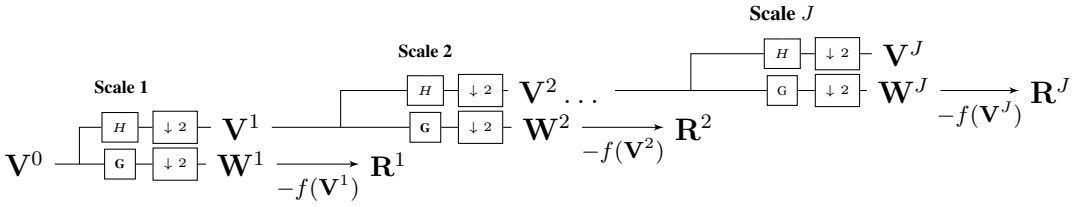



Figure 1: RWA representation with $J$ levels.

Following the same notation as in (Eq. 1) and (Eq. 2), the one level and the $J$ level RWA transform are denoted, respectively, as

$$\mathbf{RWA}(\mathbf{V}^{j-1}, 1) = \left(\mathbf{V}^j, \mathbf{R}^j\right) \tag{4}$$

$$\mathbf{RWA}(\mathbf{V}^0, J) = \left(\mathbf{V}^J, (\mathbf{R}^j)^{1 \leq j \leq J}\right). \tag{5}$$

The RWA transform is easily inverted. At each scale $j$ the estimate $\widehat{\mathbf{W}^j}$ is computed from $\mathbf{V}^j$. The approximation at level $j-1$ is then obtained by computing

$$\mathbf{W}^j = \mathbf{R}^j + \widehat{\mathbf{W}^j} \tag{6}$$

$$\mathbf{V}^{j-1} = \mathbf{DWT}^{-1}\left((\mathbf{V}^j, \mathbf{W}^j), 1\right). \tag{7}$$

RWA can be based on any wavelet filter. Our choice is the S-transform [8], which operates with integer coefficients $(\mathbf{V}^j, \mathbf{W}^j)$, requiring that a quantized estimation $\mathbf{R}^j = \mathbf{W}^j - Q(\widehat{\mathbf{W}}^j)$ be employed. Here, the quantization is applied through a rounding operation.

### 2.3 Regression Model

Different prediction models can be applied to RWA. The most general model is called the *Maximum Model*. It employs all of the approximation components from $\mathbf{V}^j$ for the computation of the prediction of each details component $\mathbf{W}_i^j$. It is given by

$$\widehat{\mathbf{W}}_i^j = f_i\left[\mathbf{V}^j\right] = \boldsymbol{\beta}_{i,0}^j + \boldsymbol{\beta}_{i,1}^j \mathbf{V}_1^j + \cdots + \boldsymbol{\beta}_{i,k}^j \mathbf{V}_k^j, \quad \mathbf{V}_i^j \in \mathbb{R}^{m\times 1}. \tag{8}$$

The regression coefficients $\boldsymbol{\beta}^j$ are found for each individual component and for each particular image using the least-squares method [9] that minimizes the sum of squares of the distances between the original components and the estimated ones, i.e.,

$$\underset{\beta^j}{\operatorname{argmin}} \|\mathbf{W}_i^j - \boldsymbol{\beta}^j \mathbf{V}^j\|_2. \tag{9}$$

The *Exogenous* RWA variant [3] uses fixed regression parameters $\overline{\boldsymbol{\beta}}^j$ obtained for a single image, then applies these parameters to all new images that come from the same sensor. As a consequence, the computational cost of RWA is considerably reduced. Additionally, no side information (the parameters $\boldsymbol{\beta}^j$) needs to be stored for each individual image. *Exogenous* RWA is given by

$$\widehat{\mathbf{W}}_i^j = f_i\left[\mathbf{V}^j\right] = \overline{\boldsymbol{\beta}}_{i,0}^j + \overline{\boldsymbol{\beta}}_{i,1}^j \mathbf{V}_1^j + \cdots + \overline{\boldsymbol{\beta}}_{i,k}^j \mathbf{V}_k^j, \quad \mathbf{V}_i^j \in \mathbb{R}^{m\times 1}. \tag{10}$$

### 2.4 Predictive Weighting Scheme

The Predictive Weighting Scheme [4] assigns weights $\omega(.)$ to each component from RWA in order to capture its significance for the rate-distortion optimization in lossy coding. The weights are defined as follows:

$$\omega(\mathbf{R}^j) = 2^{j-1} \quad \text{and} \quad \omega(\mathbf{V}^j) = 2^j. \tag{11}$$

### 2.5 Computational Cost

The computational cost of RWA is dominated by the estimation of regression coefficients $\boldsymbol{\beta} = (\mathbf{V}^\intercal\mathbf{V})^{-1}\mathbf{V}^\intercal\mathbf{W}$ and the generation of the predictions $\widehat{\mathbf{W}}{=}\boldsymbol{\beta}\mathbf{V}$ as detailed in Table 1.

| Operation | Floating-point operations (FLOPs) for RWA |
|---|---|
| S-transform | $8(1-\frac{1}{2^l})mz$ |
| $\boldsymbol{\beta} = (\mathbf{V}^\intercal\mathbf{V})^{-1}\mathbf{V}^\intercal\mathbf{W}$ | $\sum_{i=1}^{l}(2m-1)(k_i+1)^2+(k_i+1)^3+(\frac{z}{2^i})(k_i+1)\left[(2m-1)+(2k_i+1)\right]$ |
| $\widehat{\mathbf{W}} = \boldsymbol{\beta}\mathbf{V}$ | $2\sum_{i=1}^{l}(2k_i-1)m\frac{z}{2^i}$ |
| $\mathbf{W} \pm \widehat{\mathbf{W}}$ | $2m(z-1)$ |

Table 1: Computational complexity in FLOPs for RWA [3]. $z$ is the number of spectral components, $m$ is the number of samples per component, and $l$ is the number of wavelet decomposition levels. $k_i$ is the number of predictor components employed at level $i$.

## 3 Model Selection

The general problem to be analyzed in this section is as follows: from the $k$ candidate approximation components $\left(\mathbf{V}_1^j, \cdots, \mathbf{V}_k^j\right)$ that can be included in the regression model for estimating a given detail component $\mathbf{W}_i^j$, we seek to determine a small and efficient subset of $k' < k$ components, able to provide an accurate prediction as close as possible to that of the *Maximum Model*.

Traditionally, selection strategies in regression analysis focus on deciding whether a component should be added or not to the model, given a specific criterion. Such criterion is usually based on the residual variance when the regression goal is to predict. Common selection strategies are *stepwise regression* procedures, including forward selection, backward elimination and other generalizations [7]. Nonetheless, in an on-board scenario, these strategies are not feasible because of their high computational complexity when the number $k$ of candidate components is large.

Hence, there is a need for a simple selection method that can be easily carried out yet produces a reliable small model. Under this requirement, we propose a *neighbor selection* approach, which selects some adjacent components for the prediction of each detail component. Before describing our proposal in detail and to put in context the different selection strategies, we describe two other possible approaches.

### *3.1 Random Selection*

This approach randomly selects some approximation components to predict a given detail component. The performance, in terms of prediction accuracy, provided by such an approach helps us determine whether there is in fact any need for a selection strategy.

### *3.2 Forward Selection*

This approach starts with an empty set of approximation components. Then, the most correlated component with the dependent response is appended. Afterwards, and through an iterative process, the component that best improves the model following a specified criterion –usually based on the error variance– is appended. The iterative process stops after a determined number of iterations or when the prediction error is small enough. If the size of the *restricted model* does not exceed $k' < k$ components, the complexity of *forward selection* is less than or equal to $\mathcal{O}(k'mk)$, where $m$ is the number of training samples [10]. This method is not guaranteed to find the best model, although it can glean essentially the necessary information from the data to choose a reliable model [7].

### *3.3 Neighbor Selection*

Our proposed *neighbor selection* is, in principle, a divide-and-conquer strategy obtained by dividing the original image (spectrally) into multiple sub-images, each having the original spatial size but only a subset of the components. The one level RWA scheme with the *Maximum Model* is then applied independently to each sub-image with a complexity of $\mathcal{O}(mk')$, where $k' < \frac{z}{2}$.

In practice, the divide and conquer strategy discussed above can be seen to be essentially equivalent to applying the one level RWA to the whole image, but including only a small set of adjacent components in the regression model. The slight difference in the two strategies is due to the symmetric extension applied by the DWT at the boundaries between sub-images in the divide and conquer strategy. In what follows, we consider only the modified strategy that applies the RWA to the entire image, but with a *restricted model* that includes only a few components selected using the *neighbor selection* strategy. Specifically, for predicting a given detail component $\mathbf{W}_i^j$, the *restricted model* has as its inputs approximation components with index $i$ ranging over some number $k'$ of adjacent components:

$$\widehat{\mathbf{W}}_i^j = f_i'\left[\mathbf{V}^j\right] = \beta_{i,0}'^j + \beta_{i,1}'^j \mathbf{V}_{i-p}^j + \cdots + \beta_{i,2p+1}'^j \mathbf{V}_{i+p}^j, \quad k' = 2p+1 < k. \tag{12}$$

The size of the *restricted model* (the value of $p$) can be determined using a specified criterion. For instance, a reasonable criterion can be a threshold based on the number of floating-point operations.

## 4 Experimental Results

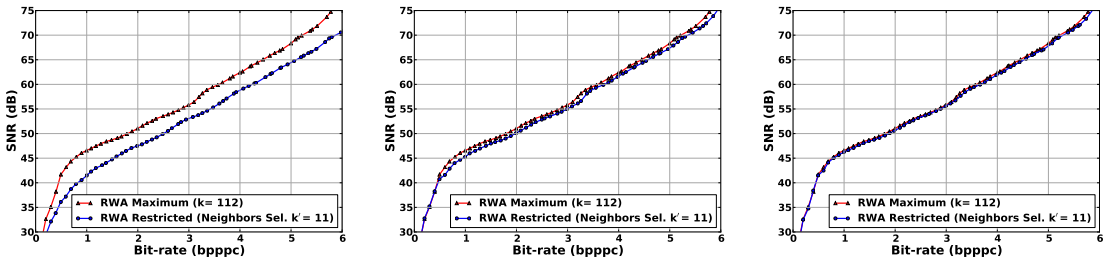
In this section we evaluate the performance of the RWA scheme with *restricted model* using the proposed *neighbor selection* strategy for progressive lossy-to-lossless coding (which includes the lossless regime). First, we compare the performance of RWA with *restricted model* using different selection strategies. Then we evaluate the coding performance of RWA and *Exogenous* RWA with *restricted model* determined by the *neighbor selection* strategy for different sizes. For comparison purposes, we also report the results of two other popular spectral decorrelation transforms, reversible PCA and DWT 5/3.

Coding performance is reported in terms of rate-distortion, showing the relation between the bit-rate in bits per pixel per component (bpppc) and the Signal-to-Noise Ratio (SNR), expressed in dB and computed as follows, $SNR = 10\log_{10}\left(\frac{\sigma^2}{MSE}\right)$, where $\sigma^2$ is the variance of the original image and $MSE$ is the mean squared error.

The coding system pipeline is as follows: first, a spectral transform is applied, then a JPEG 2000 compression is carried out, which may include a reversible 2D DWT 5/3 transform with 5 levels in the spatial dimension. The spatial transform is not applied for RWA and PCA, since it does not provide any improvement. The Kakadu software v7.7 implementation has been used for JPEG 2000. Performance evaluation is conducted on a set of 12 images from two different sensors: 5 calibrated and 5 uncalibrated Yellowstone scenes from AVIRIS [11] (224 components) and 2 uncalibrated images, *20091007093900Z* and *20091007143900Z*, from IASI Level 0 [12] (8359 components).

### *4.1 Influence of the Selection Strategy*

Fig. 2 reports the performance of RWA with *restricted model* with $k' = 11$ predictor components when using different selection strategies: *random selection*, *forward selection* and the proposed *neighbor selection*. Also included for comparison is the performance of RWA with the *maximum model*. As can be seen, *random selection* yields the worst performance, indicating that a careful selection of the predictor components is needed. On the other hand,




(a) Random selection (b) Forward selection (c) Neighbor selection

Figure 2: Comparison among different selection strategies for RWA with *restricted model* using $k' = 11$ components in the prediction. The average results are shown for the 5 uncalibrated scenes of the Yellowstone image with 224 components. The red curve is the same in each sub-figure.

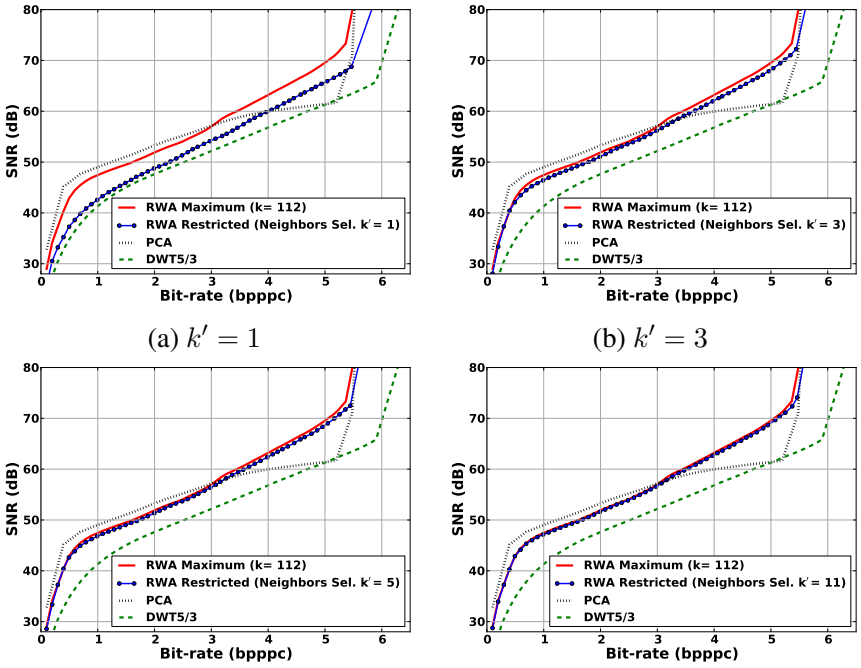
our proposed low-complexity *neighbor selection* can perform with equivalent or even better performance than the sophisticated and expensive *forward selection* method.

### *4.2 Rate Distortion*

Figures 3, 4 and 5 provide the rate-distortion performance for RWA using our proposed *restricted model* as selected with our *neighbor selection* strategy. The results for several choices of $k'$ are included. Also included for comparison are results for RWA with the *Maximum Model* and for the reversible 5/3 DWT. Results for the reversible KLT (PCA) are included for the AVIRIS images (224 components), but not for the IASI images (8359 components). The computational complexity of PCA is prohibitive for these latter images.

Results show that extreme reduction in the model size ($k'$ <3 for AVIRIS and $k'$ <40 for IASI) yields diminished coding performance compared to that obtained by the *Maximum Model*. However, the performance of the *Maximum Model* can be attained by the *restricted model* using modest values of $k'$. For instance, the *restricted model* with only $k'$=11 for AVIRIS (Fig. 3 (d), 4 (d)) or $k'$=81 for IASI (Fig. 5 (b)) yields a performance roughly equivalent to that of the *Maximum Model*. These choices still result in a significant reduction compared to the number of components employed by the maximum model, which includes $k$=112 components for AVIRIS and $k$=4180 for IASI in the first level. An additional consideration favoring the restricted model, is that the side information required for the parameters $\beta^j$ tends be negligible for small values of $k'$. For instance, use of the *Maximum Model* for IASI leads to a side information of about $0.9$ bpppc. In this case, the *Exogenous* version of the *Maximum Model* leads to comparable compression performance due to not needing any side information. In contrast, by using the proposed small model of size $k'$=81, side information is reduced to only $0.03$ bpppc.

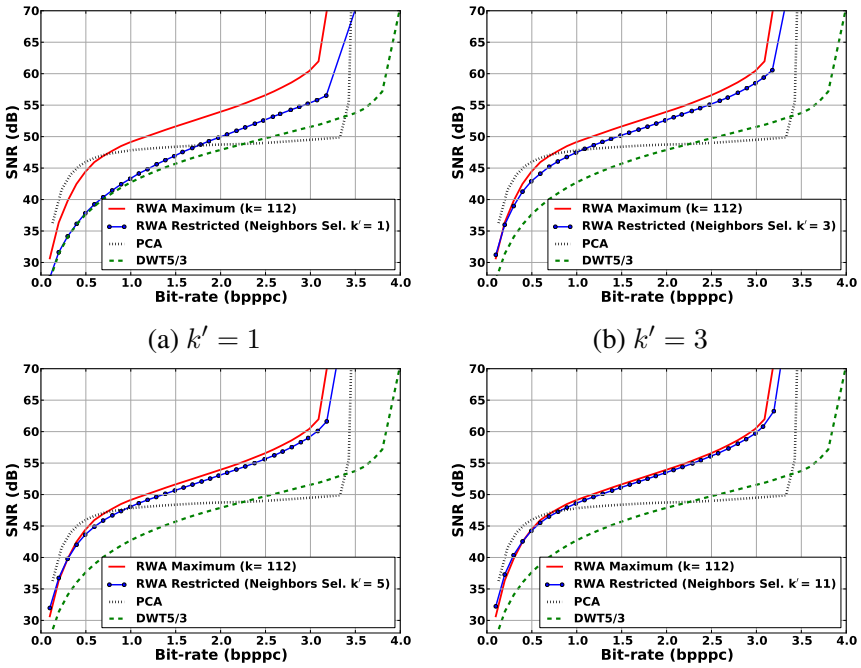
Finally, for the sake of completeness, Fig. 6 provides a comparison between RWA and *Exogenous* RWA with the same *restricted model*. Not surprisingly, the performance of the two systems is nearly identical. The same behavior was mentioned above (and reported previously [5]) in conjunction with the *Maximum Model*.




(a) $k' = 1$ (b) $k' = 3$

(c) $k' = 5$ (d) $k' = 11$

Figure 3: Average rate-distortion curves for 5 scenes of uncalibrated Yellowstone; 224 components.



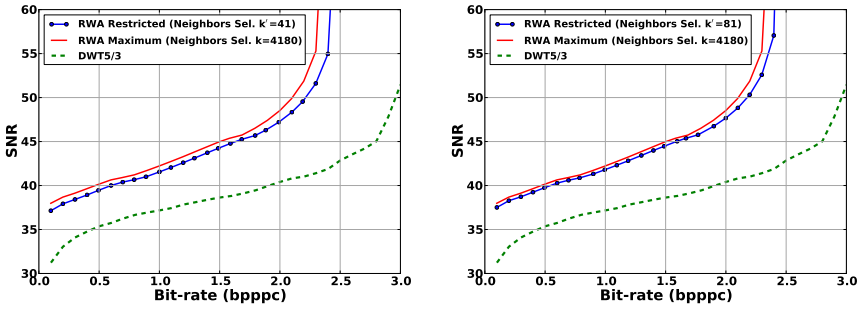

(a) $k' = 1$ (b) $k' = 3$

(c) $k' = 5$ (d) $k' = 11$

Figure 4: Average rate-distortion curves for 5 scenes of calibrated Yellowstone; 224 components.

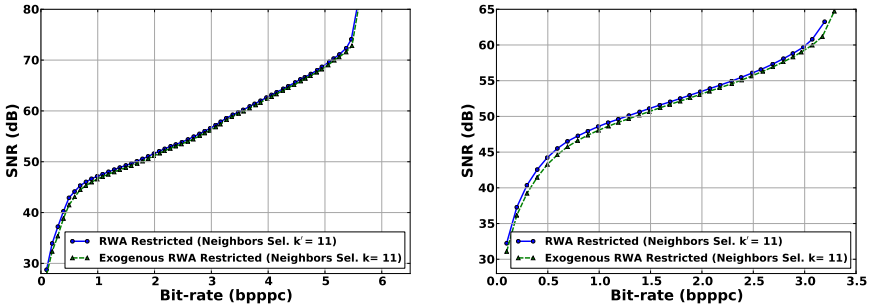
### *4.3 Computational Cost*

Fig. 7 compares the computational cost in floating-point operations for different spectral transforms, including RWA with different models. As can be seen, the difference be-




(a) $k' = 41$ (b) $k' = 81$

Figure 5: Average rate-distortion curves for two IASI L0 images; 8359 components.




(a) uncalibrated Yellowstone, $z = 224$ (b) calibrated Yellowstone, $z = 224$

Figure 6: Comparison between RWA and Exogenous RWA using *restricted model*.

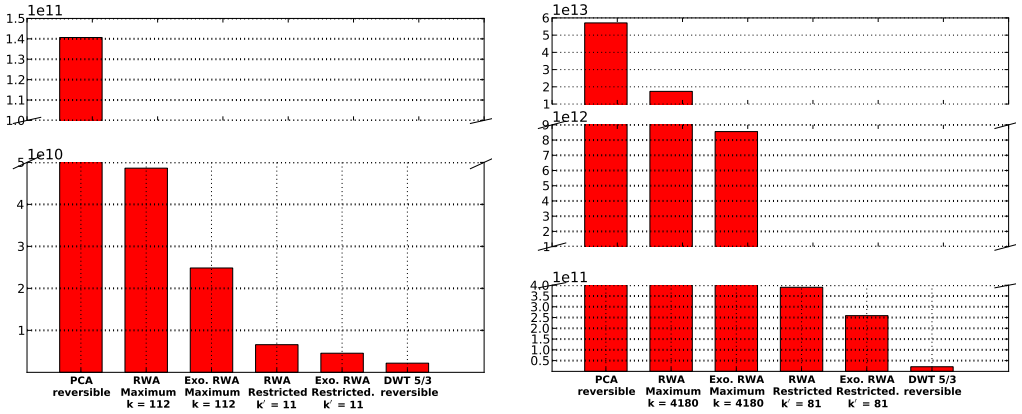
tween *Maximum Model* and *restricted model* is substantial. In particular, using $k' = 11$ for AVIRIS and $k' = 81$ for IASI, results in the computational cost being reduced by $86\%$ and $97\%$, respectively, compared to that of the *Maximum Model*. This is achieved with negligible loss in coding performance.

## 5 Conclusions

The encoding of imagery on-board remote sensing satellites asks for efficient, fast and low complexity approaches. We recently introduced RWA and reported its competitive performance. Here we proposed an extremely low complexity variant that yields the coding performance of the best performing PCA but with the computational cost of the affordable DWT 5/3. In fact, our proposal can be understood as a means to provide a reliable parsimonious model for the regression estimation at no cost. Additional benefits consist of a significant reduction of side information and scalability to any model size.

## Acknowledgements

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




(a) Yellowstone uncalibrated, $z = 224$ (b) IASI L0, $z = 8359$

Figure 7: Computational cost for different spectral transforms on AVIRIS and IASI sensors.

Government under Grant 2014SGR-691, and by Universitat Autònoma de Barcelona under Grant UAB-PIF-472-03-1/2012.

## References

[1] N. Amrani, V. Laparra, G. Camps-Valls, J. Serra-Sagrista, and J. Malo, "Lossless coding of hyperspectral images with principal polynomial analysis," in *International Conference on Image Processing (ICIP)*, Oct 2014, pp. 4023–4026.

[2] 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, Jan. 2004.

[3] N. Amrani, J. Serra-Sagrista, V. Laparra, M. 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, September 2016.

[4] N. Amrani, J. Serra-Sagrista, M. Hernandez-Cabronero, and M. Marcellin, "Regression wavelet analysis for progressive-lossy-to-lossless coding of remote-sensing data," in *Data Compression Conference*, March 2016.

[5] N. Amrani, J. Serra-Sagrista, and M. Marcellin, "Unbiasedness of regression wavelet analysis for progressive lossy-to-lossless coding," in *Picture Coding Symposium*, December 2016.

[6] S. Chen, "orthogonal-least-squares forward selection for parsimonious modelling from data," *Engineering*, vol. 2, pp. 55–74, 2009.

[7] D. Kleinbaum, L. Kupper, A. Nizam, and E. Rosenberg, *Applied regression analysis and other multivariable methods*. Cengage Learning, 2013.

[8] M. Adams, F. Kossentini, and R. Ward, "Generalized S transform," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2831–2842, Nov 2002.

[9] J. Nocedal and S. J. Wright, *Least-Squares Problems*. Springer, 2006.

[10] J. A. Khan, S. Van Aelst, and R. H. Zamar, "Robust linear model selection based on least angle regression," *Journal of the American Statistical Association*, vol. 102, no. 480, pp. 1289–1299, 2007.

[11] Jet Propulsion Laboratory, NASA. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.overview.html

[12] "Infrared atmospheric sounding interferometer (IASI)," https://wdc.dlr.de/sensors/iasi/.

# Chapter 7

# Results summary

In this chapter, we summarize the coding results of our Regression Wavelet Analysis (RWA) algorithm. First, we explain the system coding pipeline and the dataset of images. Then, we provide the results for pure lossless and progressive lossy-to-lossless respectively. We focus on three variants of RWA that represent its strength: **(1)** ***general* RWA** with ***Maximum Model***, **(2)** ***Exogenous* RWA** with ***Maximum Model*** and **(3) RWA** with ***restricted model*** based on our ***neighbor selection*** strategy. We compare RWA performance to the reversible transforms: DWT 5/3 and PCA, and to the predictive methods: M-CALIC [23] and CCSDS-123 [24].

## 7.1 Pipeline Coding and Dataset

The coding system pipeline is shown in Fig. 7.1. First, a one-dimensional transform is applied in the spectral dimension, then a JPEG 2000 compression is carried out, which may include a 2-D DWT 5/3 transform with 5 levels in the spatial dimension. The spatial transform is not applied for PCA and RWA. In these cases, the spatial transform provides no improvement, as most of the transformed components already have low energy. The Kakadu software implementation (v7.7) of JPEG 2000 has been used.

The experimental evaluations rely on a set of images from three different sensors: AVIRIS [1], HYPERION [32] and IASI [33]. Table 7.1 provides detailed information

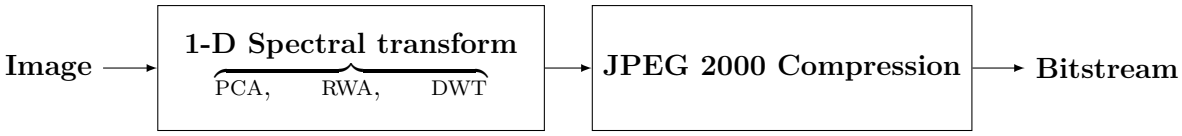



Figure 7.1: 1D spectral transform followed by JPEG 2000 compression.

Table 7.1: Dataset information for AVIRIS, Hyperion and IASI sensors. $z$ is the number of spectral components.

| Image | Scene | Year | Width | Height | Bitdepth | Datatype | Variance |
|---|---|---|---|---|---|---|---|
| **AVIRIS** $z = 224$ | | | | | | | |
| Yellowstone | 0, 3, 10, 11, 18 | 2006 | 680 | 512 | 16 | uncalibrated | average=3962160 |
| Yellowstone | 0, 3, 10, 11, 18 | 2006 | 677 | 512 | 16 | calibrated | average=2.8214e+05 |
| Hawaii | 1 | 2001 | 614 | 512 | 12 | uncalibrated | 1.2481e+04 |
| Maine | 10 | 2006 | 680 | 512 | 12 | uncalibrated | 7.5558e+04 |
| **Hyperion** $z = 242$ | | | | | | | |
| Lake Monona | – | 2009 | 256 | 3176 | 12 | uncalibrated | 7.7836e+04 |
| Mt. St. Helens | – | 2009 | 256 | 3242 | 12 | uncalibrated | 4.2316e+04 |
| Erta Ale | – | 2010 | 256 | 3187 | 12 | uncalibrated | 4.2982e+04 |
| Agricultural | – | 2002 | 256 | 3129 | 12 | calibrated | 1.4017e+06 |
| Coral Reef | – | 2003 | 256 | 3127 | 12 | calibrated | 7.5267e+05 |
| Urban | – | 2002 | 256 | 2905 | 12 | calibrated | 1.8807e+06 |
| **IASI Level 0** $z = 8359$ | | | | | | | |
| 20091007093900*Z* | – | 2009 | 60 | 1528 | 12 | uncalibrated | 1.2395e+04 |
| 20091007143900*Z* | – | 2009 | 60 | 1528 | 12 | uncalibrated | 1.1263e+04 |
| 20100319050300*S* | – | 2010 | 60 | 1528 | 12 | uncalibrated | 1.1839e+04 |
| 20120718075700*Z* | – | 2012 | 60 | 1528 | 12 | uncalibrated | 1.3169e+04 |
| 20130116133300*Z* | – | 2013 | 60 | 1528 | 12 | uncalibrated | 1.1531e+04 |
| 20130916080300 | – | 2013 | 60 | 1528 | 12 | uncalibrated | 1.2123e+04 |
| **IASI Level 1** $z = 8461$ | | | | | | | |
| 20130816230553*Z* | – | 2013 | 60 | 1530 | 12 | calibrated | 7.1551e+06 |
| 20130817004753*Z* | – | 2013 | 60 | 1530 | 12 | calibrated | 6.6356e+06 |
| 20130817041457*Z* | – | 2013 | 60 | 1530 | 12 | calibrated | 7.2839e+06 |
| 20130817055657*Z* | – | 2013 | 60 | 1530 | 12 | calibrated | 8.8519e+06 |
| 20130817073857*Z* | – | 2013 | 60 | 1530 | 12 | calibrated | 8.0731e+06 |

about these images. As described in the table, the AVIRIS images include 5 uncalibrated and 5 calibrated images corresponding to 5 scenes in Yellowstone. These images have a bit-depth of 16 bits per pixel per component (bpppc). Also included

are two uncalibrated AVIRIS images (Maine and Hawaii), each having a bit depth of 12 bpppc. All of the AVIRIS images have 224 spectral components and 512 lines. The Hyperion sensor produces images with 242 spectral components, each having a bit-depth of 12 bpppc and a width of 256. The number of lines varies from image to image. The Infrared Atmospheric Sounding Interferometer (IASI) is composed of a Fourier transform spectrometer and an Integrated Imaging Subsystem (IIS). IASI Level 0 and IASI Level 1 images have 8359 and 8461 spectral components respectively.

## 7.2 Lossless Coding

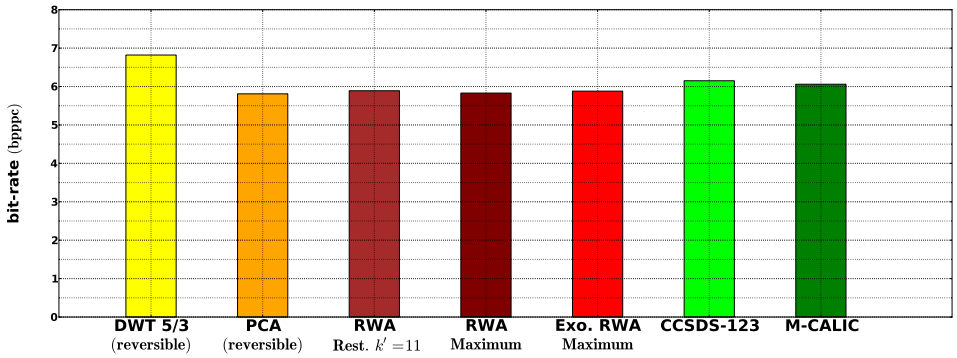
Table 7.2 provides the lossless coding results for all systems under test in terms of the bit rate, in bpppc. The figures 7.2, 7.3 and 7.4 show the average results for each sensor. All necessary side information is included in the results reported. The compression efficiency of each algorithm can be appreciated by observing the degree to which its resulting bit rate falls below the bit depth of the original image (e.g., 12 or 16 bpppc).

The summarized results show that, on average, RWA performs better than the state-of-the-art methods included in the comparison. RWA is either the best –in three out of the six datasets– or it yields results within 0.1 bpppc of the best technique. The worst such cases occur for Hyperion uncalibrated images that tend to have streaking artifacts due to the nature of the pushbroom sensors [32].

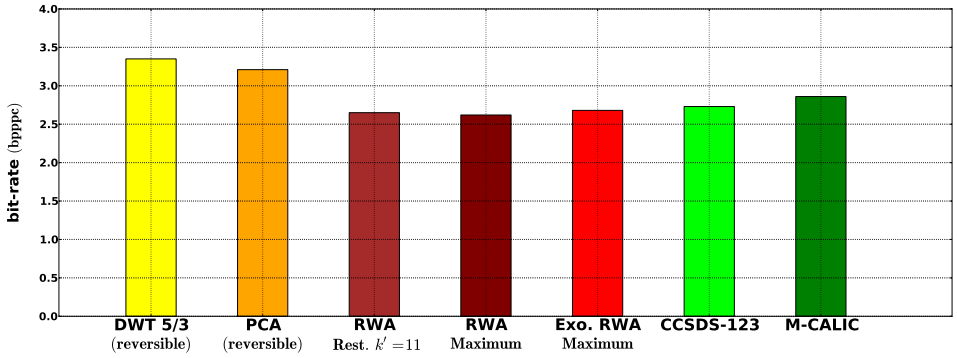
In comparison to transform-based techniques, the three RWA variants perform considerably better than DWT 5/3 by up to 1.2 bpppc and also in many cases can beat PCA by more than 0.2 bpppc. *Exogenous* RWA provides almost the same performance as RWA with a reduced complexity and without side information. This allows to apply *Maximum model* even to images with large size in the spectral dimension, such as IASI images. On the other hand, RWA with *restricted model* selected using our *neighbor selection* demonstrates that small regression models can give rise to a performance similar to that of RWA with *Maximum model* at a much lower complexity and memory requirement.

Table 7.2: Lossless bitrate for RWA in comparison to spectral transforms DWT 5/3 and PCA with JPEG 2000; and to predictive techniques M-CALIC and CCSDS-123. The results are in terms of bits per pixel per component (bpppc). Lower is better.

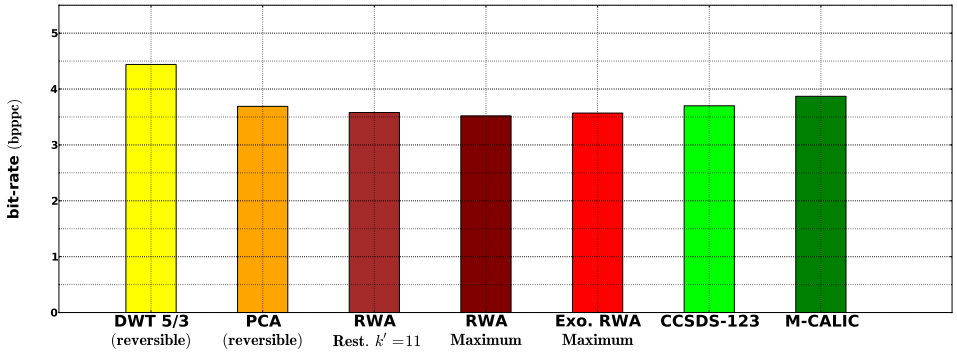
| Technique / Image | **DWT 5/3** | **PCA** | **RWA** *Restricted* ($k'$) | **RWA** *Maximum* | **Exo. RWA** *Maximum* | **M-CALIC** | **CCSDS -123** |
|---|---|---|---|---|---|---|---|
| AVIRIS Uncalibrated $k' = 11$ | | | | | | | |
| Yellowstone 00 | 7.16 | **6.05** | 6.13 | 6.07 | Train | 6.34 | 6.40 |
| Yellowstone 03 | 7.03 | **5.88** | 5.95 | 5.90 | 5.95 | 6.16 | 6.26 |
| Yellowstone 10 | 6.11 | **5.44** | 5.50 | 5.46 | 5.55 | 5.54 | 5.66 |
| Yellowstone 11 | 6.63 | **5.74** | 5.82 | 5.76 | 5.81 | 5.93 | 5.93 |
| Yellowstone 18 | 7.15 | **5.95** | 6.02 | 5.97 | 6.05 | 6.35 | 6.50 |
| Average 16 bpppc | 6.82 | **5.81** | 5.89 | 5.83 | 5.89 | 6.06 | 6.15 |
| Maine | 3.43 | 3.22 | 3.08 | **2.69** | Train | 2.89 | 2.77 |
| Hawaii | 3.27 | 3.20 | 2.58 | **2.55** | 2.82 | 2.84 | 2.70 |
| Average 12 bpppc | 3.35 | 3.21 | 2.65 | **2.62** | 2.68 | 2.86 | 2.73 |
| Global Average | 5.83 | 5.06 | 5.01 | **4.91** | 4.97 | 5.14 | 5.17 |
| AVIRIS Calibrated (Radiance) $k' = 11$ | | | | | | | |
| Yellowstone 00 | 4.75 | 3.85 | 3.81 | 3.74 | Train | 4.13 | 3.90 |
| Yellowstone 03 | 4.64 | 3.74 | 3.65 | **3.58** | 3.63 | 3.98 | 3.76 |
| Yellowstone 10 | 3.84 | 3.40 | 3.22 | **3.18** | 3.25 | 3.39 | 3.35 |
| Yellowstone 11 | 4.25 | 3.67 | 3.52 | **3.47** | 3.52 | 3.73 | 3.58 |
| Yellowstone 18 | 4.71 | 3.72 | 3.72 | **3.65** | 3.72 | 4.10 | 3.92 |
| Average 16 bpppc | 4.44 | 3.69 | 3.58 | **3.52** | 3.57 | 3.87 | 3.70 |
| Hyperion Uncalibrated $k' = 11$ | | | | | | | |
| Erta Ale | 4.44 | 4.49 | 4.65 | 4.46 | Train | 4.76 | 4.24 |
| Lake Monona | 4.55 | 4.62 | 4.85 | 4.63 | 4.63 | 4.92 | **4.36** |
| Mt. St. Helens | 4.57 | 4.54 | 4.70 | 4.52 | 4.52 | 4.83 | **4.28** |
| Average | 4.56 | 4.55 | 4.38 | 4.54 | 4.54 | 4.83 | **4.29** |
| Hyperion Calibrated $k' = 11$ | | | | | | | |
| Agricultural | 6.35 | 5.88 | 5.58 | **5.44** | Train | **5.44** | 5.73 |
| Coral Reef | 5.80 | 5.53 | 5.20 | 5.11 | 5.18 | **5.05** | 5.47 |
| Urban | 6.39 | 5.89 | 5.59 | 5.46 | 5.50 | **5.44** | 5.75 |
| Average | 6.18 | 5.76 | 5.46 | 5.34 | 5.37 | **5.31** | 5.65 |
| IASI Uncalibrated (Level 0) $k' = 81$ | | | | | | | |
| IASI L0 1 | 3.02 | – | 2.52 | $2.35_{+0.9}$ | Train | 2.89 | 2.90 |
| IASI L0 2 | 3.02 | – | 2.51 | $2.34_{+0.9}$ | **2.39** | 2.87 | 2.88 |
| IASI L0 3 | 3.01 | – | 2.52 | $2.35_{+0.9}$ | **2.40** | 2.87 | 2.90 |
| IASI L0 4 | 3.05 | – | 2.52 | $2.35_{+0.9}$ | **2.41** | 2.91 | 2.91 |
| IASI L0 5 | 2.99 | – | 2.51 | $2.35_{+0.9}$ | **2.40** | 2.86 | 2.90 |
| IASI L0 6 | 3.08 | – | 2.53 | $2.37_{+0.9}$ | **2.41** | 2.92 | 2.91 |
| Average | 3.02 | – | 2.52 | $2.35_{+0.9}$ | **2.40** | 2.88 | 2.90 |
| IASI Calibrated (Level 1) $k' = 81$ | | | | | | | |
| IASI L1 1 | 7.46 | $7.22_{+2.4}$ | 6.53 | $6.42_{+0.9}$ | Train | 6.86 | 6.60 |
| IASI L1 2 | 7.09 | $7.05_{+2.4}$ | 6.53 | $6.41_{+0.9}$ | **6.48** | 6.88 | 6.61 |
| IASI L1 3 | 7.11 | – | 6.54 | $6.42_{+0.9}$ | **6.49** | 6.88 | 6.59 |
| IASI L1 4 | 7.06 | – | 6.55 | $6.43_{+0.9}$ | **6,50** | 6.87 | 6.56 |
| IASI L1 5 | 7.12 | – | 6.55 | $6.43_{+0.9}$ | **6.50** | 6.87 | 6.60 |
| Average | 7.16 | $7.13_{+2.4}$ | 7.11 | $6.42_{+0.9}$ | **6.49** | 6.87 | 6.59 |




(a) AVIRIS uncalibrated

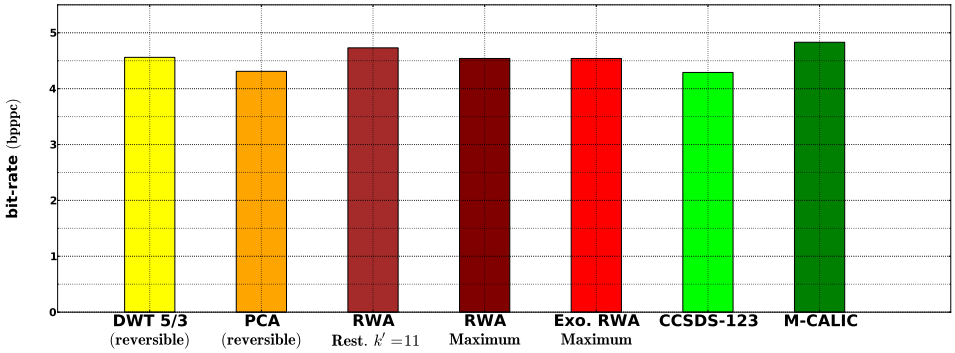



(b) AVIRIS uncalibrated 12 bits (Maine and Hawaii)

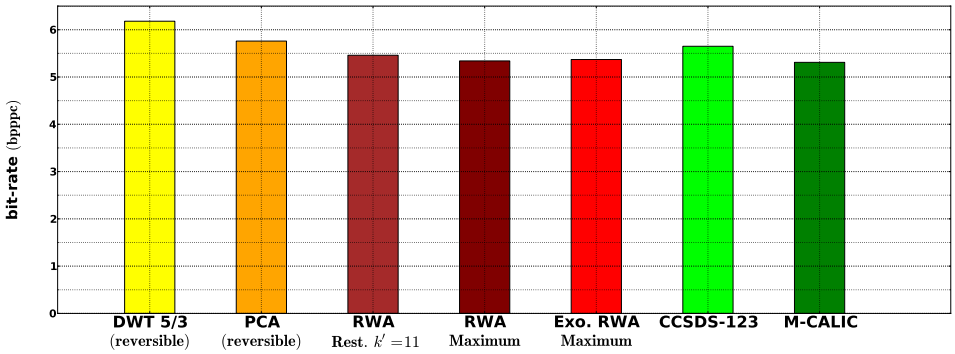



(c) AVIRIS radiance

Figure 7.2: Lossless coding performance for AVIRIS in bpppc (lower is better)

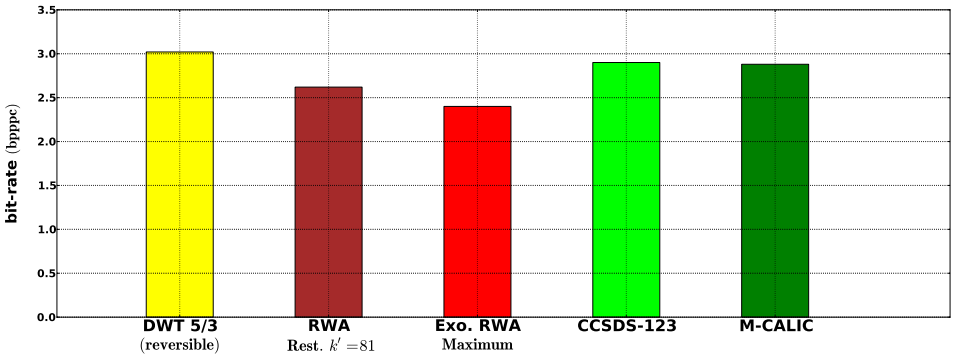



(a) Hyperion uncalibrated

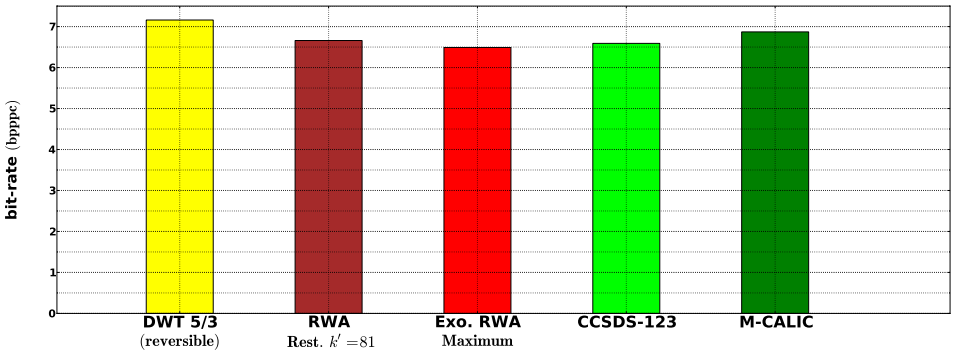



(b) Hyperion radiance

Figure 7.3: Lossless coding performance for Hyperion in bpppc (lower is better).




(a) IASI L 0




(b) IASI L 1

Figure 7.4: Lossless coding performance for IASI in bpppc (lower is better).

In relation to predictive methods, RWA performs with equivalent or better accuracy than M-CALIC and CCSDS-123. The difference in coding gain is higher for IASI images in favor of RWA, since RWA can benefit from a large number of available predictors.

## 7.3 Progressive Lossy-to-Lossless Coding

Figures 7.5, 7.6 and 7.7 summarize the Progressive lossy-to-lossless results in terms of the rate-distortion curves for RWA and the compared spectral transforms. The rate-distortion provides the relation between the bit-rate in bpppc and the Signal-to-Noise Ratio (SNR), expressed in dB and computed as follows: $SNR = \log_{10}\left(\frac{\sigma^2}{MSE}\right)$, where $\sigma^2$ is the variance of the original image and $MSE$ is the mean square error.

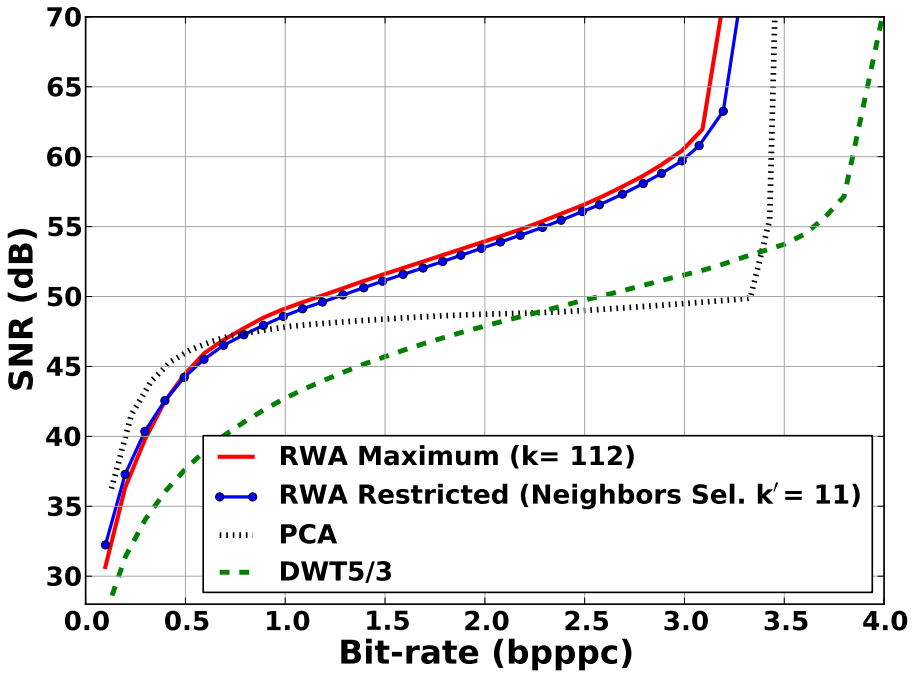



Figure 7.5: Average rate-distortion curves for AVIRIS Radiance (224 spectral components).

The results indicate that RWA offers a significant coding gain over DWT 5/3 and PCA. Even though RWA and PCA provide a comparable performance for lossless

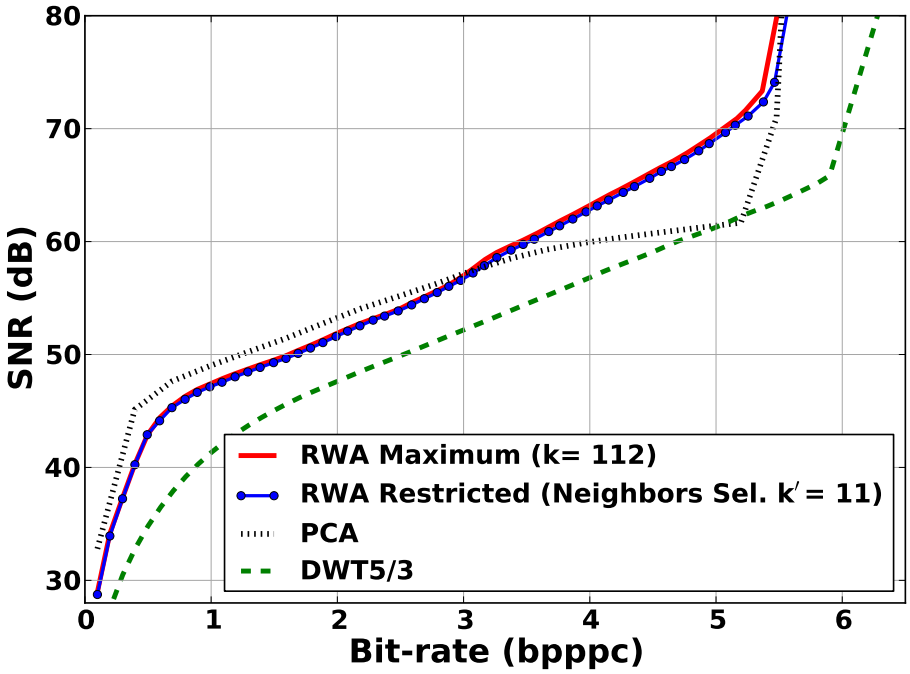



Figure 7.6: Average rate-distortion curves for AVIRIS Uncalibrated (224 spectral components).

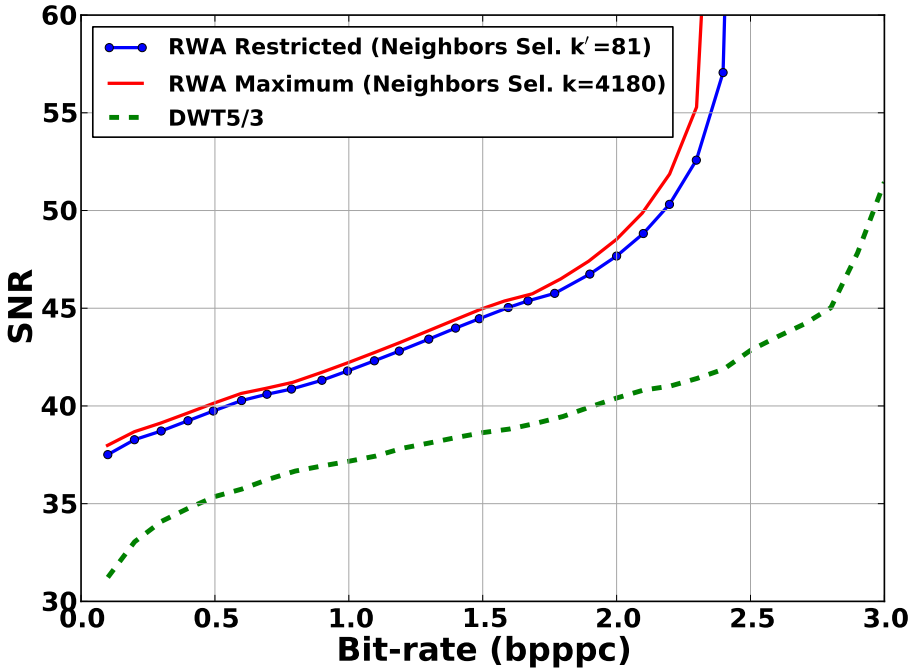



Figure 7.7: Rate-distortion curves for IASI L0 (8359 spectral components).

coding, the coding gain of RWA for low and intermediate bitrates is significantly higher over reversible PCA. On the other hand, RWA using *restricted model* with a small size is able to provide a similar performance to that obtained by *Maximum model*. Finally, *Exogenous* RWA, using only the regression parameters computed on one image, also provides almost identical performance to that obtained by *general* RWA.

## 7.4 Computational Complexity

Complexity and efficiency have motivated research into compression techniques that meet the restrictions of on-board satellite systems. To this end, a number of compression approaches aim at reducing the complexity while minimizing the performance sacrifice [9, 8, 13, 34, 7, 25]. In particular, RWA has the benefit of both advantages. In this section, we provide the computational complexity of RWA, when using different models and variants. In general, the complexity of RWA is dominated by the computation of the least squares parameters $\boldsymbol{\beta}$ and the estimates generation $\widehat{\mathbf{W}}$. In the case of *Exogenous* RWA, estimation of $\boldsymbol{\beta}$ is performed off-line and does not contribute to the complexity of encoding. Meanwhile, the complexity of the estimates generation $\widehat{\mathbf{W}}$ depends on the size of the regression model, and is at its highest when using the *Maximum* model. However, the cost of other efficient transforms such as Principal Component Analysis (PCA) is still significantly higher. Table 7.3 details the computational cost of RWA in Floating-point operations (FLOP), while figures 7.8, 7.9 and 7.10 compare the computational cost of different transforms applied to different images from the AVIRIS, Hyperion and IASI sensors.

| Operation | Floating-point operations (FLOP) for RWA |
|---|---|
| S-transform | $8(1-\frac{1}{2^l})mz$ |
| $\boldsymbol{\beta} = (\mathbf{V}^\intercal\mathbf{V})^{-1}\mathbf{V}^\intercal\mathbf{W}$ | $\sum_{i=1}^{l}(2m-1)(k_i+1)^2+(k_i+1)^3+(\frac{z}{2^i})(k_i+1)\left[(2m-1)+(2k_i+1)\right]$ |
| $\widehat{\mathbf{W}} = \boldsymbol{\beta}\mathbf{V}$ | $2\sum_{i=1}^{l}(2k_i-1)m\frac{z}{2^i}$ |
| $\mathbf{W} \pm \widehat{\mathbf{W}}$ | $2m(z-1)$ |

Table 7.3: Computational complexity in FLOPs for RWA [26]. $z$ is the number of spectral components, $m$ is the number of samples per component, and $l$ is the number of wavelet decomposition levels. $k_i$ is the number of predictor components employed at level $i$.

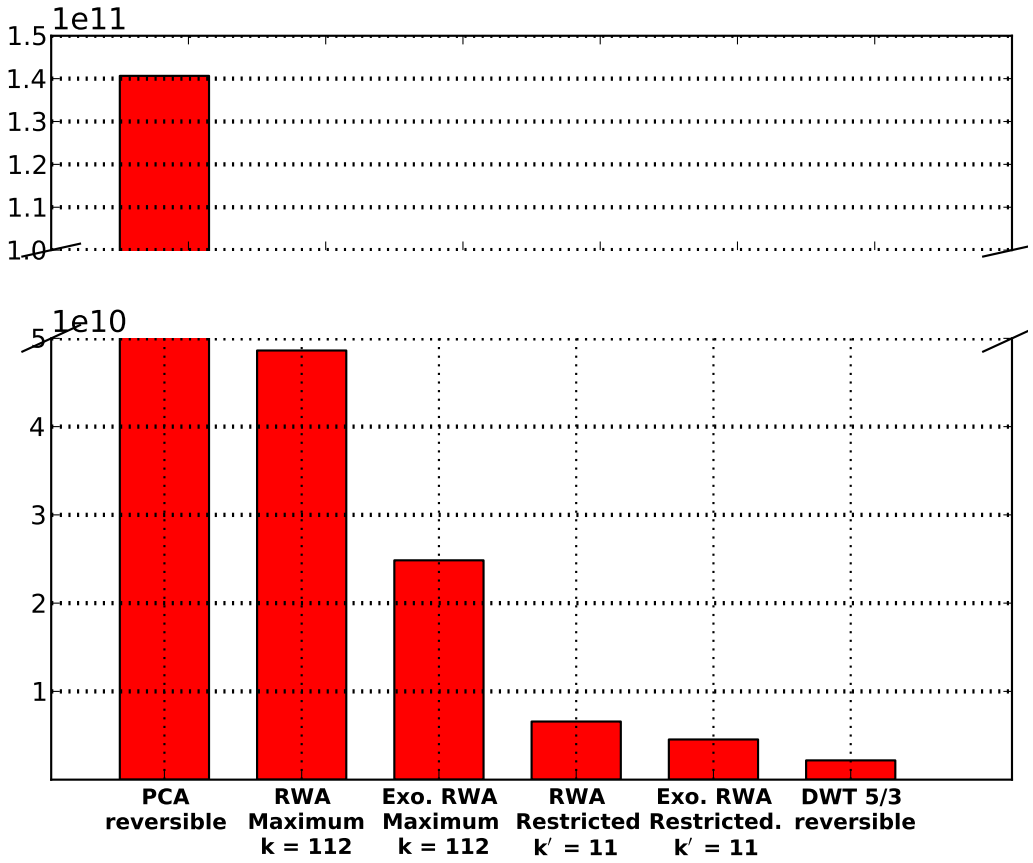



Figure 7.8: Cost comparison in FLOP for different transforms applied to the uncalibrated Yellowstone image from AVIRIS with 224 spectral components and a spatial resolution of $512 \times 680$. The values reported for PCA and DWT 5/3 are from [7] .

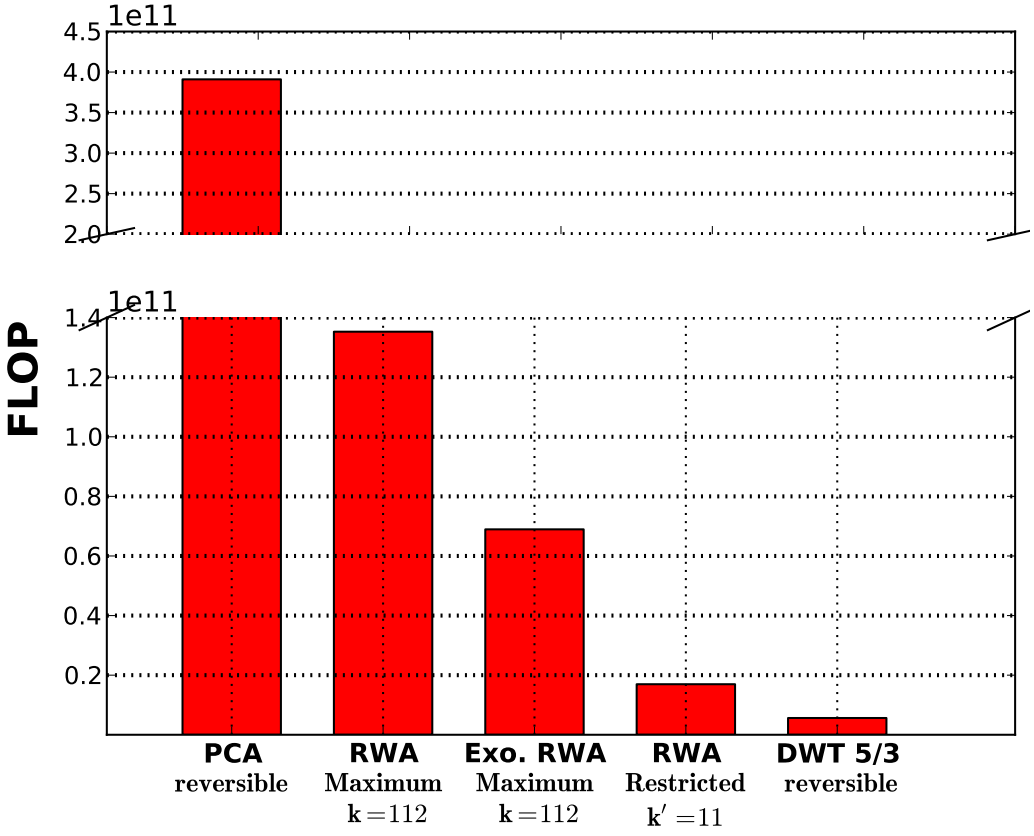



Figure 7.9: Cost comparison in FLOP for different transforms applied to the uncalibrated Mt. St. Helens image from Hyperion with 242 spectral components and a spatial resolution of $3242 \times 256$. The values reported for PCA and DWT 5/3 are from [7] .

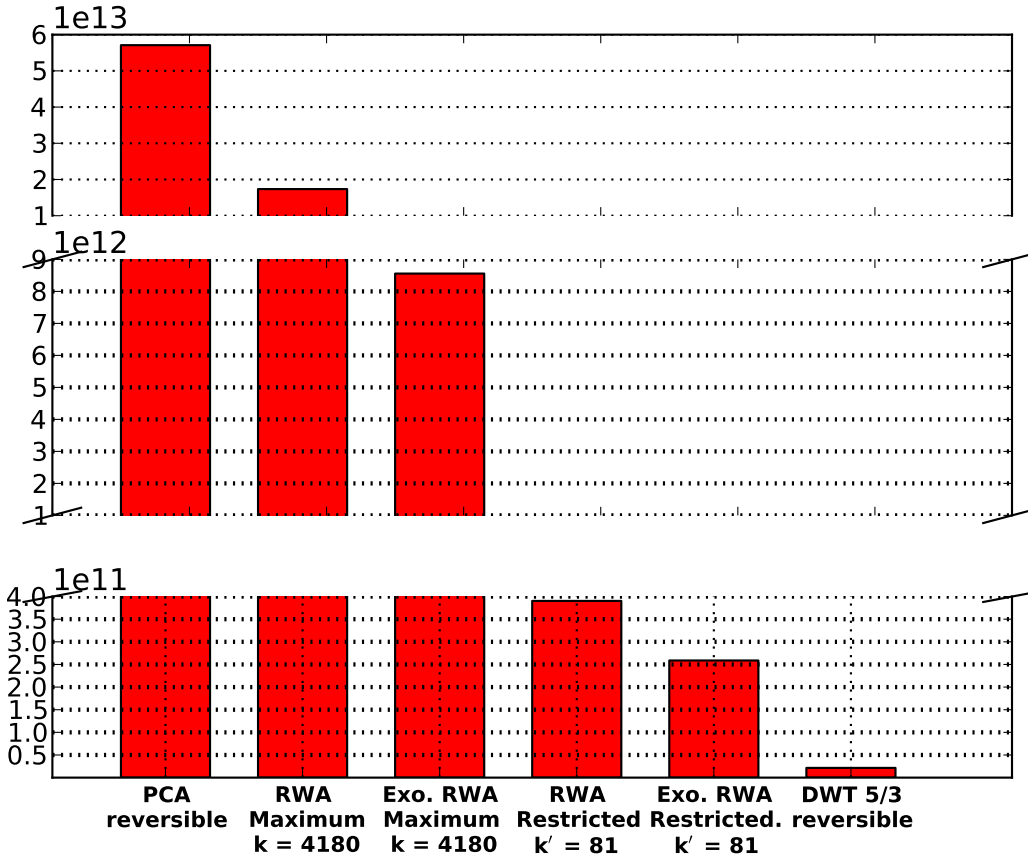



Figure 7.10: Cost comparison in FLOP for different transforms applied to a IASI L0 image with 8359 spectral components and a spatial resolution of $1528 \times 60$. The values reported for PCA and DWT 5/3 are from [7] .

# Chapter 8

# Conclusions

## 8.1 Summary

Remote sensing imagery has become extremely essential for today's applications. The development of the optical sensors is still increasing, leading to an unprecedented size of produced data, which challenges the capacity of storage and transmission. This calls for adopting more efficient and robust compression techniques than the existing ones. Traditionally, the research activities have focused on transforms and predictive techniques. Principal Component Analysis (PCA), which is optimal for decorrelating Gaussian sources, provides an excellent coding performance, but entails several drawbacks related with its expensive complexity. Meanwhile, the performance of other simple data-independent transforms such as DWT is not convincing. On the other hand, the predictive methods are mostly addressed for lossless or near-lossless coding, where they can outperform transform-based techniques.

At the beginning of this thesis, we have targeted to analyze transforms and predictive techniques, and the potential of their combination. First, we focused on PCA as a powerful transform that produces linearly uncorrelated components, while combining polynomial regression of higher orders to exploit the remaining non-linear dependencies among components. Even though this idea makes sense for some specific artificial data that truly feature polynomial relations, for practical hyperspectral data, it does not provide useful improvement due to the scarcity of polynomial dependencies. Our

analysis in Chapter 2 suggests that, over hyperspectral images, the principal components are not able to predict each other through polynomial regression, not even by using different models with multiple regressors. Furthermore, this results in a higher computational complexity and memory requirements, and also produces some coding penalizations due to the increase of side information and the sequential rounding error.

On the basis of the aforementioned analysis, it is concluded that the best combination of transform and prediction techniques for compression purposes does not necessarily rely on using the optimal transform for Gaussian sources, i.e., PCA. This conducted us to seek for simpler transforms that do not entail severe drawbacks, even if their coding performance is limited. Meanwhile, we adopt prediction techniques to improve the independence property, which usually results in an improvement in the coding performance. Based on that, in Chapter 3 we introduced Regression Wavelet Analysis (RWA) for lossless coding of remote sensing images. RWA is a pyramidal scheme that employs multiple regression analysis to exploit the dependencies among wavelet components. For lossless coding, RWA provides superior performance to PCA at much lower computational complexity, and also superior performance to the most prominent predictive techniques like M-CALIC and CCSDS-123. We proposed several variants of RWA to reduce its complexity and preserve the component scalability of the wavelet transform. The strongest and the most important variant is *Exogenous* RWA, which uses fixed parameters, previously computed, to each new image from the same sensor.

In our algorithm, the primary goal of regression is the prediction accuracy, thus, we only encode a small residual. Nevertheless, *Exogenous* RWA demonstrates that our regression scheme is also able to reliably describe the relationships among transformed components. This reliability is confirmed by the excellent performance obtained when applying fixed parameters, computed over only one image, to all new images coming from the same sensor. The resulting performance is almost identical to that obtained when using the optimal least squares parameters computed for each individual image. As a consequence, the complexity is significantly reduced at the cost of a negligible sacrifice in performance, and no side information is needed to be transmitted, since

the parameters are known at the encoder. This latter property is crucial for images with large size in the spectral dimension, such as IASI (8461 components), since the amount of the side information becomes unreasonable, gravely penalizing the coding gain.

The success of RWA for lossless coding motivated us to explore its benefits for lossy coding. In Chapter 4 we extend RWA to progressive lossy-to-lossless. Initially, we have seen that the loss of information in the RWA coefficients without weights leads to a poor performance with an unsteady rate-distortion curve. The reason of this problem is that RWA coefficients with similar magnitude may have extremely different impact on the reconstruction quality. Hence, the encoder needs to know the significance of these coefficients in order to appropriately allocate the bitrate. To this end, we proposed a Predictive Weighting Scheme that captures the prediction significance of each spectral components. With that, we showed that RWA can attain, also for progressive lossy-to-lossless, a rate-distortion performance superior to that provided by other state-of-the-art methods. As a conclusion, the predictive capability of RWA has demonstrated to be highly robust to the loss of information, providing a stable and competitive reconstruction quality.

On the basis of this satisfactory attainment for lossy coding, the impact of the data change on the prediction accuracy due to lossy coding has emerged naturally. Specifically, the bias in the regression parameters of RWA when coding with low bitrates has spurred our interest. In Chapter 5, we investigate the attenuation of the RWA parameters, when recovering components with loss of information. First, we explain the attenuation for the simple linear regression, which is the case of the first scale in the inverse RWA. Then, we proceed with the attenuation equations in the case of multiple regression with two regressors, where it is also illustrated the influence of the correlation among regressors. Finally, we illustrate the high relation between the bias in the regression parameters and the variance of the RWA components for the general case. We close this analysis by assessing the bias of RWA parameters for coding performance over a set of remote sensing images. For each lossy compression at a given ratio, the encoder adjusts the parameters according to the loss of information introduced. Then, the decoder inverts the RWA algorithm using old and new optimal

parameters to provide a performance comparison. The results show that at very low compression ratios, the introduced bias is very small, therefore, there is no need for parameters adjustment due to very small improvement that can be achieved in the reconstruction quality. This unbiasedness of RWA parameters for lossy coding allows us to conclude that our predictive scheme is well specified and accurately estimated, yielding unbiased predictions of the response when the loss of information is introduced in the data. The same conclusion is drawn for *Exogenous* RWA, where fixed parameters, computed over only one image, give unbiased predictions for all images coming from the same sensor. This attainment indicates that the predictive scheme of RWA accurately captures the relations among data components. Rarely a predictive method allows the luxury of using fixed parameters obtained through a straightforward training phase and giving a nearly unbiased coding performance to that obtained by computing the optimal parameters. Add to this the fact that no side information is needed to be stored or transmitted, resulting in a data-independent algorithm with a reduced complexity.

As a final stage of this thesis, we address in Chapter 6 the problem of selecting parsimonious models for the RWA scheme to meet the restricted limitation and requirements of on-board compression. Rather than using expensive selection strategies, we focus on simpler but efficient ones, based on divide and conquer strategies. We propose a neighbor selection approach, a regression model that uses a few adjacent spectral components to predict a given detail component. Our strategy selects at no cost reliable parsimonious models that provide similar coding performance to that obtained by *Maximum* model. Parsimonious models in the RWA scheme drive their complexity and memory requirements to the low extend, almost identical to the extend of the affordable DWT 5/3, while the coding performance is still superior to the reversible PCA and to the most sophisticated approaches in the state-of-the-art.

As a general conclusion, along this thesis, we have deeply analyzed the compression of remote sensing images and the state-of-the-art techniques. This allowed us to introduce Regression Wavelet Analysis, a novel algorithm that covers most aspects of the remote sensing compression. The first aspect is coding efficiency. For lossless coding, RWA outperforms the best comparable decorrelation transforms like PCA

and the most competitive prediction methods, such as CCSDS-123. For progressive lossy-to-lossless (PLL), RWA also yields superior performance as compared to PCA. The next aspect is data independence. RWA vastly reduces the correlation and the mutual information among components. It achieves almost full linear decorrelation and attains a small advantage over PCA in reducing mutual information. Another important aspect is the low complexity and memory requirements. The computational cost of RWA is significantly lower to other efficient transforms like PCA and close to affordable DWT 5/3. Unbiasedness for lossy coding is another important central point. RWA has proven to be robust to the data change due to lossy coding. Last but not least, straightforwardness and effectiveness of parameters learning: *Exogenous* RWA learns the least squares parameters on only one image, while their application to all new images coming from the same sensor is highly effective, as evidenced by the negligible sacrifice in the coding performance compared to the true optimal parameters.

## 8.2 Future work

Even though the provided solutions in this thesis for remote sensing compression were broad and diverse, some new important ideas arise for further contribution and research in relation to theoretical and practical topics.

First of all, data classification and dimensionality reduction are almost natural extensions of RWA application. So far, the quality assessment for lossy coding has been carried out by metrics based on Mean Squared Error (MSE). Some other applications may focus on classification accuracy as a quality metric. Therefore, it seems interesting to investigate models that optimize the classification accuracy rather than minimizing the sum of squared distances between the prediction and the response component. On the other hand, many other applications have as an objective to reduce the dimensionality of the data instead of reducing the size in memory. Thus, other approaches like supervised classification are learned and applied on a reduced number of features. As a future work, we aim at exploiting RWA for dimensionality reduction.

Regarding practical on-board compression, a number of satellite instruments are push broom sensors that obtain images by capturing separately spectral lines. Hence, a compression system on-board is constrained to encode each single or group of captured spectral lines. So far, our RWA algorithm only has considered the whole image with all the spatial samples. For the least squares estimation, the degree of freedom $(m - z - 1)$ plays an important role and may change the results. Although Fast RWA learns the least squares parameters on a sub-sampled set of spatial pixels, this sub-sampling has only been performed through a grid selection over the whole image. Therefore, it is worth analyzing the parameters learning when only a few number of spectral lines is available.

On the other hand, our RWA approach is successful for coding remote sensing images, but it can also be analyzed for other 3D images. For instance, video coding is one of the most interesting fields that rely on prediction techniques. However, extensions of RWA to motion data have not yet been investigated.

Beyond extensions to other fields, there are still some gaps and lacks in the research. Along this thesis, our work was based on spectral prediction and decorrelation, resulting in a one-dimensional approach. Nevertheless, remote sensing and other 3D images also feature reasonable redundancy along the spatial dimension. In our RWA algorithm we neglect the 2D decorrelation and we apply entropy coding without involving any other decorrelation step. We have seen that a spatial 2D wavelet transform does not provide any additional improvement to the RWA residuals due to their sparsity and low energy. However, the prediction step can use not only the spectral bands, but also the spatial pixels to predict the current ones. This idea is common in the state-of-the-art and may capture more correlation, leading to an improvement of the performance.

In a similar way, the temporal correlation could be a beneficial information to exploit. Some satellites on-board perform orbits in a repeated cycles, providing images of the same area for different period times.

Finally, and from a theoretical perspective, the prediction methods for lossy coding are an attractive research line that needs further investigation. We successfully extended our RWA scheme to lossy coding with an exhaustive analysis of the bias

in the least squares parameters when having loss of information. However, *imputation* is one of the attractive techniques that is still unexplored for prediction-based compression. *Imputation* is one of the most popular approaches to deal with loss of information that adopts the idea of replacing missing information with substituted values. This family of techniques is common and widely used in statistics to deal with missing data for prediction purposes.

# Bibliography

[1] Jet Propulsion Laboratory, NASA. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.overview.html

[2] N. Amrani, V. Laparra, G. Camps-Valls, J. Serra-Sagrista, and J. Malo, "Lossless coding of hyperspectral images with principal polynomial analysis," in *International Conference on Image Processing (ICIP)*, Oct 2014, pp. 4023–4026.

[3] P. Hao and Q. Shi, "Reversible integer KLT for progressive-to-lossless compression of multiple component images," in *ICIP 2003. Proceedings. 2003 International Conference on Image Processing, 2003.*, vol. 1, Sept 2003, pp. I–633–6 vol.1.

[4] B. Penna, T. Tillo, E. Magli, and G. Olmo, "A new low complexity KLT for lossy hyperspectral data compression," *IGARSS 2006. IEEE International Conference on Geoscience and Remote Sensing Symposium, 2006.*, vol. 7, pp. 3525–3528, 2006.

[5] ——, "Transform coding techniques for lossy hyperspectral data compression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1408–1421, May 2007.

[6] Q. Du and J. Fowler, "Low-complexity principal component analysis for hyperspectral image compression," *International Journal of High Performance Computing Applications*, vol. 22, no. 4, pp. 438–448, 2008.

[7] I. Blanes and J. Serra-Sagristà, "Clustered reversible-KLT for progressive lossy-to-lossless 3d image coding," in *Data Compression Conference. 2009.* IEEE Press, Mar. 2009, pp. 233–242.

[8] I. Blanes, J. Serra-Sagrista, M. Marcellin, and J. Bartrina-Rapesta, "Divide-and-conquer strategies for hyperspectral image processing: A review of their benefits and advantages," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 71–81, 2012.

[9] 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, Mar. 2011.

[10] 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, May 2011.

[11] I. P. A. Bita, M. Barret, F. D. Vedova, and J.-L. Gutzwiller, "Lossy and lossless compression of MERIS hyperspectral images with exogenous quasi-optimal spectral transforms," *Journal of Applied Remote Sensing*, vol. 4, pp. 041 790–1–15, Jul 2010.

[12] I. P. A. Bita, M. Barret, and D. Pham, "On optimal orthogonal transforms at high bit-rates using only second order statistics in multicomponent image coding with JPEG2000," *Signal Processing*, vol. 90, no. 3, pp. 753–758, Jan 2010.

[13] M. Barret, J.-L. Gutzwiller, I. P. A. Bita, and F. D. Vedova, "Lossy hyperspectral images coding with exogenous quasi optimal transforms," in *Data Compression Conference, 2009. DCC'09.* IEEE, March 2009, pp. 411–419.

[14] P. Birjandi and M. Datcu, "Multiscale and dimensionality behavior of ICA components for satellite image indexing," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 1, pp. 103–107, 2010.

[15] G. Camps-Valls, D. Tuia, L. Gomez-Chova, S. Jimenez, and J. Malo, *Remote Sensing Image Processing*, ser. Synthesis Lectures on Image, Video and Multimedia Proc., A. Bovik, Ed. Austin, TX, USA: Morgan & Claypool Publ, 2011.

[16] V. Laparra and R. Santos-Rodriguez, "Spatial/spectral information trade-off in hyperspectral images," in *IEEE Intl. Geosci. Rem. Sens. Symp. (IGARSS)*, July 2015, pp. 1124–1127.

[17] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, ser. The Kluwer International Series in Engineering and Computer Science. Kluwer, 1992, no. 159.

[18] 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, 2014.

[19] N. Farvadin and J. Modestino, "Rate-distortion performance of DPCM schemes for autoregressive sources," *IEEE Trans. Information Theory*, vol. 31, no. 3, pp. 402–418, 1985.

[20] J. Mielikainen and P. Toivanen, "Clustered dpcm for the lossless compression of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 12, pp. 2943–2946, 2003.

[21] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE transactions on Communications*, vol. 45, no. 4, pp. 437–444, 1997.

[22] ——, "Context-based lossless interband compression-extending calic," *IEEE Transactions on Image Processing*, vol. 9, no. 6, pp. 994–1001, 2000.

[23] 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, Jan. 2004.

[24] Consultative Committee for Space Data Systems (CCSDS), *Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1*, ser. Blue Book. CCSDS, May 2012.

[25] M. Klimesh, "Low-complexity lossless compression of hyperspectral imagery via adaptive filtering," *IPN Progress Report 42-163*, pp. 1–10, 2005. [Online]. Available: http://ipnpr.jpl.nasa.gov/progressreport/42-163/163H.pdf

[26] N. Amrani, J. Serra-Sagrista, V. Laparra, M. 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, September 2016.

[27] N. Amrani, J. Serra-Sagrista, M. Hernandez-Cabronero, and M. Marcellin, "Regression wavelet analysis for progressive-lossy-to-lossless coding of remote-sensing data," in *Data Compression Conference*, March 2016.

[28] N. Amrani, J. Serra-Sagrista, and M. Marcellin, "Unbiasedness of regression wavelet analysis for progressive lossy-to-lossless coding," in *Picture Coding Symposium*, December 2016.

[29] ——, "Low complexity prediction model for coding remote-sensing data with regression wavelet analysis," in *Data Compression Conference*, April 2017.

[30] V. Laparra, S. Jiménez, D. Tuia, G. Camps-Valls, and J. Malo, "Principal polynomial analysis," *Intl. Journal of Neural Systems*, vol. 24, no. 07, p. 1440007, 2014.

[31] M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards, and Practice*. Springer Science & Business Media, 2002, vol. 1.

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

[33] "Infrared atmospheric sounding interferometer (IASI)," https://wdc.dlr.de/sensors/iasi/.

[34] I. Blanes, M. Hernandez-Cabronero, F. Auli-Llinas, J. Serra-Sagrista, and M. Marcellin, "Isorange pairwise orthogonal transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3361–3372, June 2015.