

# Statistical Atmospheric Parameter Retrieval Largely Benefits from Spatial-Spectral Image Compression

Joaquín García-Sobrino, Joan Serra-Sagrístà, *Senior Member, IEEE*,  
Valero Laparra, Xavier Calbet, and Gustau Camps-Valls, *Senior Member, IEEE*

**Abstract**—The Infrared Atmospheric Sounding Interferometer (IASI) is flying on board of the Metop satellite series, which is part of the EUMETSAT Polar System (EPS). Products obtained from IASI data represent a significant improvement in the accuracy and quality of the measurements used for meteorological models. Notably, IASI collects rich spectral information to derive temperature and moisture profiles –among other relevant trace gases–, essential for atmospheric forecasts and for the understanding of weather. Here, we investigate the impact of near-lossless and lossy compression on IASI L1C data when statistical retrieval algorithms are later applied. We search for those compression ratios that yield a positive impact on the accuracy of the statistical retrievals. The compression techniques help reduce certain amount of noise on the original data and, at the same time, incorporate spatial-spectral feature relations in an indirect way without increasing the computational complexity. We observed that compressing images, at relatively low bit-rates, improves results in predicting temperature and dew point temperature, and we advocate that some amount of compression prior to model inversion is beneficial. This research can benefit the development of current and upcoming retrieval chains in infrared sounding and hyperspectral sensors.

**Index Terms**—Infrared Atmospheric Sounding Interferometer (IASI), Statistical retrieval, Kernel Methods, Lossy Compression, Near-Lossless Compression, JPEG 2000, M-CALIC, Spectral Transforms.

## I. INTRODUCTION

Temperature and water vapour atmospheric profiles are essential meteorological parameters for weather forecasting and atmospheric chemistry studies. Observations from high spectral resolution infrared sounding instruments on board of satellites provide unprecedented accuracy and vertical resolution of temperature and water vapour profiles. However, it is not trivial to retrieve the full information content from radiation measurements. Accordingly, improved retrieval algorithms are desirable to achieve optimal performance for existing and future infrared sounding instrumentation.

Manuscript received October 17, 2016;

JGS and JSS are with the Group on Interactive Coding of Images, Universitat Autònoma de Barcelona, Spain. Web: <http://www.gici.uab.cat>. E-mail: {FranciscoJoaquin.Garcia,Joan.Serra}@uab.cat.

VL and GCV are with the Image Processing Laboratory (IPL), Universitat de València, València, Spain. Web: <http://isp.uv.es>. E-mail: {valero.laparra,gustau.camps}@uv.es.

XC is with AEMET, C/ Leonardo Prieto Castro, 8, 28071 Madrid, Spain. E-mail: xcalbeta@aemet.es.

This work was supported in part by the Spanish Ministry of Economy and Competitiveness and by the European Regional Development Fund under Grants TIN2015-71126-R, TIN2012-38102-C03-00, by the Catalan Government under Grant 2014SGR-691, and by the European Research Council under Consolidator Grant SEDAL ERC-2014-CoG 647423.

## A. Atmospheric parameter retrieval with IASI

The use of Metop data in Numerical Weather prediction (NWP) accounts for 40% of the impact of all space based observations in NWP forecasts. The Infrared Atmospheric Sounding Interferometer (IASI) sensor is implemented on the Metop satellite series. Products obtained from IASI data are a significant improvement in the quality of the measurements used for meteorological models. In particular, IASI collects rich spectral information to derive temperature and moisture profiles, which are essential to the understanding of weather and to derive atmospheric forecasts. The sensor provides infrared spectra with high resolution between  $645\text{ cm}^{-1}$  and  $2760\text{ cm}^{-1}$ , from which temperature and humidity profiles with high vertical resolution and accuracy are derived. Additionally, these spectra are used for the determination of trace gases such as ozone, nitrous oxide, carbon dioxide and methane, as well as land and sea surface temperature and emissivity and cloud properties [1], [2].

EUMETSAT, NOAA, NASA and other operational agencies are continuously developing product processing facilities to obtain L2 atmospheric profile products from infrared hyperspectral radiance instruments, such as IASI. One of the retrieval techniques commonly used in L2 processing is based on linear regression, which is a valuable and very computationally efficient method. It consists of performing a canonical least squares linear regression on top of the data projected onto the first principal components or Empirical Orthogonal Functions (EOF) of the measured brightness temperature spectra (or radiances) and the atmospheric state parameters. To further improve the results of this scheme for retrieval, nonlinear statistical retrieval methods can be applied as an efficient alternative to more costly optimal estimation (OE) schemes. These methods have proven to be valid in retrieval of temperature, dew point temperature (humidity), and ozone atmospheric profiles when the original data are used [3].

## B. Impact of hyperspectral image compression

Given the orbit time of Metop satellites (101 minutes), the large spectral resolution (8461 components) and spatial resolution ( $60 \times 1530$  samples) of IASI L1C products, and that there are 2 active Metop satellites (A and B), about 41,3 GB of IASI L1C data are produced daily. This large volume of remote sensing data asks for efficient compression systems for both storage and transmission. Remote sensing data compression is a mature field attracting interest of space administrations, public bodies and private companies. Lossless, near-lossless

and lossy coding techniques are already in use in on-going satellite missions and have been adopted in several current standards. However, the *impact* of the (near-lossless and lossy) compression on the radiances can compromise the quality of posterior products, such as classification and detection maps, or bio-geo-physical parameter estimates as is the case of this work.

1) *Impact on image classification*: Effects of lossy data compression have been analyzed mostly in the scope of classification applications. Mercier et al. [4] proposed a vector quantization for joint classification and compression of hyperspectral data. This approach proved very competitive for compressing images, achieving compression ratios of approximately 70:1 and classification performance was not markedly reduced. Analogous results were achieved in [5]–[7], where high compression ratios yielded high classification accuracy.

Especially interesting is the observation that this behavior is more apparent when a spatial-spectral wavelet transform is applied. Penna et al. [8] reported a stimulating consideration: classification performance was not closely linked to the distortion levels introduced in the image, meaning that, in general, the best classification accuracy was not achieved after applying the coding algorithms/settings that yielded the best rate-distortion performance.

Du and Fowler [9] proposed a scheme based on principal component analysis (PCA) deployed in JPEG 2000 to provide spectral decorrelation. The proposed scheme produced competitive data-analysis performance, in terms of information preservation in an anomaly-detection task. Blanes et al. [10] conducted a study on compression of hyperspectral images through 3D-JPEG 2000, where supervised and unsupervised classification of the recovered images were evaluated. The results reflect that classification accuracy is still reliable after the compression stage. Similar results were achieved in [11], where the proposed lossy and near-lossless compression algorithm for hyperspectral images yielded good results for hard classification, spectral unmixing, and anomaly detection.

2) *Impact on image unmixing*: The impact of lossy compression in linear spectral unmixing and nonlinear hyperspectral image classification using support vector machines (SVMs) was investigated in [12]. The experimental results suggest that for some compression techniques, a higher compression ratio may produce more accuracy in classification results. Some authors have reported that lossy compression can preserve the needed information to estimate endmember fractional abundances in linear spectral unmixing even at low bit-rates [13], [14]. Particularly interesting is the observation reported in [15], [16], where it is noted that the joint use of spectral and spatial information in some stages of the linear spectral unmixing chain can produce smaller image reconstruction errors.

3) *Impact on bio-geo-physical parameter retrieval*: It is worth noting that, to the authors' knowledge, there are no works analyzing the impact of spatial-spectral data compression on bio-geo-physical parameter retrieval. We reported preliminary results in [17], which are further extended here with more data and in-depth analysis. Experimental results revealed an interesting issue: the performance of statistical

retrieval methods substantially improved after near-lossless compression. This observation leads us to define the main goal of this research.

### C. Goal and main findings

This paper is concerned with the performance of different (near-lossless and lossy) compression techniques and different (linear and non-linear) statistical retrieval algorithm on IASI L1C data. We will touch upon M-CALIC [18] coding technique and upon JPEG 2000 standard [19], [20]. In this latter case, two spectral transforms will be employed to better exploit the high spectral redundancy inherent to IASI data, namely, Pairwise Orthogonal Transform (POT) [21], [22] and Discrete Wavelet Transform (DWT), which have proved to achieve good data decorrelation in multi-, hyper-, and ultraspectral images. Then, two different statistical retrieval algorithms, Linear Regression (LR) and Kernel Ridge Regression (KRR) [23], will be evaluated in retrieval of temperature and dew point temperature (humidity) profiles from the recovered decoded images.

In short, the objective of this paper is to provide an outlook of the effects of near-lossless and lossy compression of IASI L1C data when statistical retrieval methods are employed to retrieve physical information from the reconstructed images. An important conclusion of the work is that retrieval methods may operate on data that do not have the same quality precision as that originally provided by IASI L1C products. Actually, though it may appear counter-intuitive at first glance, some amount of compression can improve the accuracy in atmospheric parameter profiles estimation. This analysis may have a deep impact in other infrared sounding instruments and hyperspectral sensors, both currently flying (e.g., AIRS) or upcoming (e.g., MTG-IRS). The observed effect has been widely exploited in signal and image processing problems (see e.g. [24]–[27]), and can be explained by noting that compression performs a sort of signal denoising and it generally constitutes an indirect way to include spatial-spectral feature relations, which ultimately help to pixel-wise retrieval algorithms. This turns to be an extremely simple, yet effective way to improve retrieval results, and comes at the price of including a compression step before retrievals, where the compression ratio needs to be properly adjusted.

### D. Outline of the paper

The remainder of the paper is organized as follows. Section II details the proposed approach, describing the different compression paradigms and statistical retrieval methods. Section III provides the experimental results. Section IV provides some discussion about the reasons behind the observed phenomena. Conclusions and further work are drawn in Section V.

## II. METHODS

This section introduces the experimental setting and methods. The proposed sequential approach includes first a compression stage and then a statistical parameter retrieval. Fig. 1 illustrates the adopted scheme. Two different compression

approaches will be considered. On the one hand, the maximum absolute error per pixel introduced in the reconstructed image shall be bounded using a near-lossless paradigm. On the other hand, the overall bit-rate shall be controlled using a lossy approach. Similarly, two different statistical retrieval algorithms will be studied: a standard least squares LR and a KRR method, which have provided very good performance in recent studies [3]. At a very early stage, a feature extraction/selection process is conducted to discard some less useful components, leading from the original 8461 spectral components on IASI LIC products to 4699 spectral components [3].

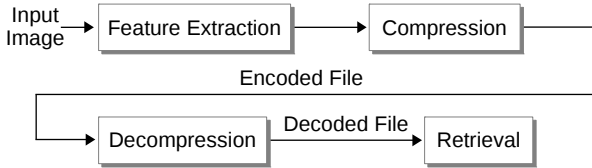


Fig. 1. Proposed sequential approach: after a feature extraction process, compression is applied first, followed by statistical parameter retrieval.

#### A. Near-lossless and lossy compression techniques

Lossless compression techniques allow to completely reconstruct all the original data in the reconstructed image after the decoding process is performed. However, compression ratios for lossless coding are limited, usually not going beyond 2:1. Near-lossless and lossy compression remove some information during the coding process, preventing the recovery of all the original data, but allowing to improve the compression ratios as the introduction of distortion becomes larger. These approaches rely on the assumption that the quality of the data after the coding process is still appropriate for the intended specific use.

Two recognized coding techniques, M-CALIC for near-lossless and JPEG 2000 for lossy compression, are discussed in this section.

1) *Near-Lossless Compression*: M-CALIC [18] has been selected for near-lossless compression since it provides a competitive performance when spectral redundancy is high, which is the case of IASI LIC products. M-CALIC is a lossless and near-lossless compression technique based on context-based adaptive lossless image coding (CALIC) [28]. CALIC was designed as a proposal for the ISO standard for lossless and near-lossless compression of 2D images, and although it was not finally selected because of its higher computational complexity as compared to LOCO-I [29] algorithm, CALIC provides a higher coding performance.

M-CALIC uses a multi-component spectral predictor, along with optimized model parameters and optimization thresholds. The algorithm exploits correlation among components by employing the two previous components of the current one in the prediction, which allows to capture most of the spectral correlation. The intensity of the pixel to be encoded is estimated based mostly on the intensity of co-located pixels in previous components, i.e., pixels at the same spatial position but in

components with a lower index. In fact, it is the prediction error that is encoded. For near-lossless compression, these prediction errors are quantized, ensuring that all distortion (peak absolute) errors fall below a given threshold. Fig. 2 illustrates the basic architecture of M-CALIC.



Fig. 2. M-CALIC basic architecture.

2) *Lossy Compression*: JPEG 2000 is an international standard developed by the *Joint Photographic Experts Group (JPEG)*. It was intended as the successor of JPEG due to JPEG limitations in many application areas as mobile communications, medical diagnostic imaging, enhanced Internet browsing, digital cinema, and multimedia. JPEG 2000 provides a wide range of features in a single compressed bit-stream for a large amount of applications. Figure 3 illustrates the basic architecture of JPEG 2000.

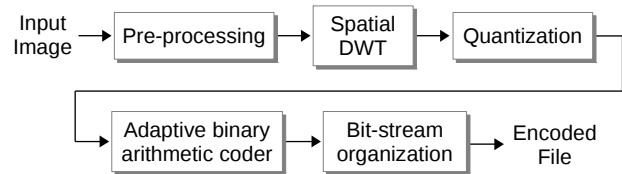


Fig. 3. JPEG 2000 basic architecture.

Two spectral transforms, namely POT and DWT, are applied along with JPEG 2000 to account for the large correlation among the spectral components.

Pairwise Orthogonal Transform (POT) [21], [22] relies on the implementation of a divide-and-conquer strategy to the Karhunen-Loève Transform (KLT), where the composition of smaller KLT transforms produces the resulting transform. Each composition of KLT is computed from only two image components. In a classic KLT, every component is decorrelated with each other, irrespective of how much energy they share. In its turn, POT implements a mechanism that decorrelates portions with high shared energy while ignoring the other portions. POT works in a multilevel mode, where a two-component KLT transform is computed for every pair of consecutive components at each level. This allows to accumulate most of the image energy in the first components since each composition is computed as a classic KLT. Most of the signal energy is grouped in one of the two resulting components allowing most of the image energy to flow across the composition of transforms up to the last level.

The Discrete wavelet transform (DWT) can be seen as successive levels of decomposition that allow to decorrelate the processed data. Every time the transform is applied, the signal is decomposed into two sets of coefficients, the low frequencies (L) and the high frequencies (H) sets. The low frequencies set is a representation of the input signal at a

coarser resolution, while the high frequencies set represents details.

### B. Statistical parameter retrieval

We aim at studying the impact of image compression on the retrieval of atmospheric profiles using standard and modern statistical retrieval algorithms. In particular, we will pay attention to the Kernel Ridge Regression (KRR) algorithm [23], [30], which generalizes least squares linear regression (LR). KRR has shown good performance in prediction of such variables using IASI L1 data in previous works [3]. We analyze the performance for both LR and KRR at different compression ratios.

Let us now fix the notation and review the formulation of KRR for regression. The KRR, also known as Least Squares SVM (LS-SVM), is the kernel version of a regularized linear regression (LR). Let  $\mathbf{x}_i \in \mathbb{R}^N$  (spectra) and  $y_i \in \mathbb{R}^M$  (state vectors), where  $i = 1, \dots, n$  indicates the index of the  $n$  training samples. We want to perform a linear least squares regression in a Hilbert space,  $\mathcal{H}$ , of very high (possibly infinite) dimensionality  $D_{\mathcal{H}}$ , where samples have been mapped through a mapping  $\phi(\mathbf{x}_i)$ . In matrix notation, the model is given by  $\mathbf{Y} = \Phi \mathbf{W} + \mathbf{b}$  and, under the assumption of an additive *i.i.d.* noise model,  $\hat{\mathbf{Y}} = \mathbf{Y} + \mathbf{E}$  with Gaussian noise  $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I})$  of zero mean and standard deviation  $\sigma_n$ .

Then, as in the regularized linear regression setting, we want to minimize the regularized squared loss function,  $\mathcal{L}_p = \|\mathbf{Y} - \Phi \mathbf{W}\|^2 + \lambda \|\mathbf{W}\|^2$  with respect to model weights  $\mathbf{W}$ . In what follows, we deliberately drop the bias term. Therefore, taking derivatives with respect to  $\mathbf{W}$  and equating them to zero, gives  $\mathbf{W} = (\Phi^T \Phi + \lambda \mathbf{I})^{-1} \Phi^T \mathbf{Y}$ , where  $\Phi$  is the matrix of mapped samples,  $[\phi(\mathbf{x}_1)^T, \phi(\mathbf{x}_2)^T, \dots, \phi(\mathbf{x}_n)^T]$ , whose size is  $n \times D_{\mathcal{H}}$ . Note that this problem is not solvable as the inverse runs on matrix  $\Phi \Phi^T$ , which is of size  $D_{\mathcal{H}} \times D_{\mathcal{H}}$ , and  $\Phi$  is in principle unknown. Here, we apply the Representer's theorem by which we can express the solution as a linear combination of mapped samples,  $\mathbf{W} = \Phi^T \alpha$ , and then the solution is expressed as a function of the dual weights  $\alpha$  (one per sample),  $\alpha = (\Phi \Phi^T + \lambda \mathbf{I})^{-1} \mathbf{Y}$ . Note that now the problem is solvable as we only need to compute the inverse of the (regularized) Gram matrix  $\mathbf{K} = \Phi \Phi^T$  of size  $n \times n$ . Even though the mapping is unknown, one can replace this inner product matrix with a similarity matrix between samples, which is known as the *kernel matrix*  $\mathbf{K}$ .

We finally need to show that we never actually require access to the mapped feature vectors, which could be of infinite dimension. What we need in practice is the predicted value for a new matrix of test examples,  $\mathbf{X}_*$ . This is computed by projecting it onto the solution  $\mathbf{W}$ :

$$\hat{\mathbf{Y}}_* = \Phi_* \mathbf{W} = \Phi_* \Phi^T \alpha = \mathbf{K}_* \alpha \quad (1)$$

where the matrix  $\mathbf{K}_*$  contains the similarities between all test and training samples, whose entries are  $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \phi(\mathbf{x}_j)^T$ . The important message here is of course that we only need access to the kernel function  $K$ . Examples of typical kernel functions are the linear  $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ , the polynomial  $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^d$ , or the one used

in our implementation, the Gaussian Function (Radial Basis Function, RBF) kernel  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma^2))$ . Therefore, in KRR, two free parameters are tuned: the regularization parameter  $\lambda$  and the kernel parameter  $\sigma$ . We used a cross-validation strategy for their optimization. For the interested reader, a MATLAB implementation of KRR and other regression algorithms can be found at [http://isp.uv.es/soft\\_regression.html](http://isp.uv.es/soft_regression.html).

## III. EXPERIMENTAL RESULTS

This section is devoted to report the experimental results. First, we will describe the IASI L1C data used in the experiments. Then, we will pay attention to the compression results obtained by a near-lossless (M-CALIC) and a lossy (JPEG 2000) coding technique, and finally we will evaluate the impact of compression on linear (LR) and nonlinear (KRR) retrieval of atmospheric parameters. Due to restrictions in the space, we show results for only one image and one biophysical variable in most of the experiments; please browse [http://isp.uv.es/coding\\_retrieval.html](http://isp.uv.es/coding_retrieval.html) for results for all the other images.

### A. Data collection

To conduct the experiments, several IASI L1C images were used. Table I provides characteristics for four images of this dataset. All four images were obtained from IASI-B instrument (implemented on MetOp-B satellite), are 16 bits per pixel per component (bpppc), and are stored as signed integers.

### B. Compression Results

All the considered IASI L1C images have gone through a feature (component) selection process before carrying out the compression stage. We selected a subset of spectral components based on the minimization of measurement errors [3], which discards components with negative radiance and high levels of estimated noise by applying thresholds on the estimated noise standard deviation per component.

For near-lossless compression, we used the M-CALIC software [31]. The tested images have been compressed using the default parameters configuration. Eleven different peak absolute errors (PAE),  $\delta \in \{1, 3, 7, 15, 31, 63, 127, 255, 511, 1023, 2047\}$ , are used in our experiments. Table II shows the performance for M-CALIC compressor on IASI L1C products. Of course, the higher the peak absolute errors, the lower the bit-rate (or, equivalently, the higher the compression ratio).

For lossy compression, we used the well-known JPEG 2000 Part 2 standard with two spectral transforms, DWT and POT. We analyze three different compression schemes: a) only a spatial transform (DWT); b) only a spectral transform (either DWT or POT); c) both a spectral transform (either DWT or POT) and a spatial transform (DWT). When applicable, schemes are evaluated considering 5 and 10 levels of DWT spectral transform, and 1, 3 and 5 levels of DWT spatial transform. Seventeen target bit-rates are analyzed, uniformly distributed from 0.001 to 2 bpppc. Kakadu software [32] was used for JPEG 2000 experiments. The standard provides a

TABLE I  
EXAMPLES OF IASI L1C IMAGES USED IN THE EXPERIMENTS. TECHNICAL NAMES, IDENTIFIERS AND SIZES ARE PROVIDED.

Technical name	Identifier	Size ( $x \times y \times z$ )
IASI_XXX_IC_M01_20130817004753Z_20130817022952Z_N_O_20130817013849Z	IASI_20130817004753Z	60×1530×8461
IASI_XXX_IC_M01_20130817041457Z_20130817055656Z_N_O_20130817050513Z	IASI_20130817041457Z	60×1530×8461
IASI_XXX_IC_M01_20130817055657Z_20130817073856Z_N_O_20130817064707Z	IASI_20130817055657Z	60×1530×8461
IASI_XXX_IC_M01_20130817073857Z_20130817092056Z_N_O_20130817082957Z	IASI_20130817073857Z	60×1530×8461

TABLE II  
M-CALIC COMPRESSION PERFORMANCE IN BIT-RATE (BR) (MEASURED IN BITS PER PIXEL PER COMPONENT, LOWER IS BETTER) AND SNR ENERGY (MEASURED IN DB, HIGHER IS BETTER).

IASI L1C	$\delta = 1$		$\delta = 3$		$\delta = 7$		$\delta = 15$		$\delta = 31$		$\delta = 63$		$\delta = 127$		$\delta = 255$		$\delta = 511$		$\delta = 1023$		$\delta = 2047$	
	BR	SNR	BR	SNR	BR	SNR	BR	SNR	BR	SNR	BR	SNR	BR	SNR	BR	SNR	BR	SNR	BR	SNR	BR	SNR
IASI_20130817004753Z	5.07	75.02	3.89	67.24	2.88	60.59	2.01	54.38	1.31	48.37	0.78	42.68	0.41	37.36	0.22	31.74	0.13	25.70	0.09	19.63	0.07	13.75
IASI_20130817041457Z	5.07	75.49	3.88	67.71	2.87	61.06	2.00	54.85	1.30	48.84	0.77	43.16	0.40	37.83	0.21	32.18	0.13	26.11	0.08	20.05	0.07	14.12
IASI_20130817055657Z	5.05	76.14	3.86	68.36	2.85	61.71	1.98	55.50	1.29	49.49	0.76	43.81	0.39	38.48	0.21	32.82	0.12	26.73	0.08	20.62	0.07	14.73
IASI_20130817073857Z	5.08	75.75	3.89	67.97	2.88	61.31	2.01	55.10	1.31	49.09	0.78	43.40	0.40	38.07	0.21	32.44	0.13	26.42	0.08	20.34	0.07	14.35
Average	5.06	75.55	3.88	67.77	2.87	61.12	2.00	54.91	1.30	48.90	0.77	43.22	0.40	37.90	0.21	32.26	0.13	26.20	0.08	20.12	0.07	14.18

multi-component extension in its Part 2 [33]. In order to compute the POT spectral transform, the Pairwise Orthogonal Transform software [34] was used.

Figure 4 shows the performance of the different lossy compression schemes for product IASI\_20130817055657Z. The results for the other images are practically identical, and hence not reported here for the sake of brevity. Actually, due to space constraints, only the best configuration for each scheme is plotted, i.e., scheme a): *Spatial DWT 5 Levels*; scheme b): *Spectral POT*; scheme c) both *Spectral POT + Spatial DWT 5 Levels* and *Spectral DWT 10 Levels + Spatial DWT 5 Levels*. To facilitate the performance comparison between near-lossless and lossy compression, a curve reporting the performance of M-CALIC is also plotted in Figure 4. For M-CALIC, only compression results for PAEs between 15 and 2047 are plotted, because lower PAEs imply bit-rates larger than those commonly employed for lossy compression.

One can see that, as expected, the best performance is achieved when a spectral transform followed by a spatial transform is applied. The performance comparison between POT and DWT as spectral transforms is very similar, and selecting one or the other may depend not only on the coding performance, but also on complexity issues. Also as expected, M-CALIC starts achieving competitive coding performance only at medium to high bit-rates.

### C. Retrieval Assessment

This section presents the results of estimating physical variables (temperature and dew point temperature) using compressed IASI L1C images. Results are reported for image IASI\_20130817055657Z for moisture prediction (related dew point temperature). Once more, results for the other images are practically identical, and hence not reported here for the sake of brevity. Similarly, reciprocal performance is obtained for temperature estimation; these results are not reported either for the sake of conciseness. As supplementary material, we encourage the reader to browse through <http://isp.uv.es/>

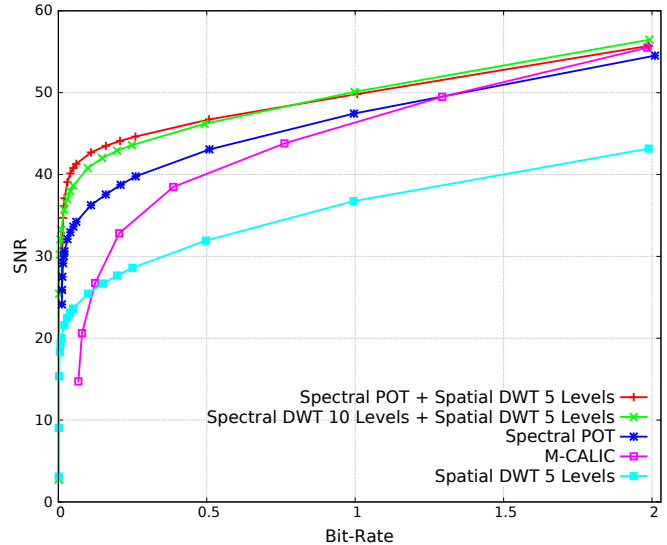


Fig. 4. **Rate-distortion performance for near-lossless and lossy compression.** Results show SNR energy (in dB, higher is better) versus bit-rate (in bppc) for image IASI\_20130817055657Z.

coding\_retrieval.html, reporting detailed results for four IASI L1C products, for both near-lossless and lossy compression, and for both temperature and dew point temperature.

The acquisition conditions vary depending on several factors and, globally, this affects the values of the radiances from overpass to overpass. We can identify three different strategies to mitigate the effect of training and testing on statistically different datasets: (1) adapt the regression method in each overpass, (2) build a sufficiently large and representative training dataset, or (3) adapt the feature representation (relative normalization) in each overpass. In this paper, we decided to start with the simpler first approach, hence the statistical retrieval algorithms (LR and KRR) are trained (and cross-validated) in each overpass, as previously done in [3]. The approach is

simple and efficient. The second approach implies building a representative big dataset, but this is challenging, suboptimal, and makes training methods more complex. The last approach would be certainly ideal, but still unsolved in signal processing and machine learning, where this is commonly referred to as *manifold alignment or domain adaptation*. Domain adaptation assumes that the test examples come from a (slightly) different distribution to those from the training set. Several works trying to accommodate such statistical differences between train-test distributions have been proposed in the recent literature [35]–[37]. Such complex methods are not considered in this work. In order to avoid the effect of this *dataset shift problem*, we assume an ideal situation: the acquisition conditions are exactly the same for both train and test (data points inside the same image are used for training and testing), but both datasets are disjoint, i.e., we report results in a test set never seen by the regression algorithms for training. This is the most common case in the literature [38].

In the experiments we follow a similar procedure as in [3]. We apply LR and KRR on IASI data hyperpixels to predict temperature and humidity at different pressure levels, i.e., all the spectral components at a particular spatial position are considered for the regression. The experiments are conducted on images compressed at different target bit-rates (different compression ratios). For each image, coding scheme and compression ratio we follow the same procedure. A set of 2,000 samples is used for training and a different set of 2,000 samples is used for testing. Note that the test samples are not used in any moment in the training process and that test samples are used only for the evaluation of the performance. Samples are first drawn following a random distribution; then, to allow a fair comparison, the position of the training samples and the position of the test samples are kept constant for all coding schemes. The procedure consists of two steps. First, the dimensionality of the data is reduced to 260 components by using the classical PCA/EOF transformation in the spectral domain. Then, LR and KRR models are applied to predict the temperature and the humidity profiles (i.e., the values at each pressure level) at the spatial position corresponding to each particular sample. LR and KRR are trained on input samples of 260 dimensions to predict output samples of 137 dimensions (one for each pressure level). We employ regularization for both methods. Also we train the parameters of both methods using cross-validation for minimizing the average RMSE of the predictions for all the pressure levels, one parameter for the LR method (the regularization parameter) and two parameters for the KRR method (the regularization parameter and the kernel RBF sigma parameter). Everything that needs to be trained (PCA and the regression parameters) is trained using just the training set. Once everything is learned, the same procedure is applied to the test data. To assess the performance we compute in the test set the root mean squared error (RMSE) between the predictions given by the regression models and the ground truth values of temperature and humidity in each pressure level.

1) *Retrieval Assessment for Near-lossless Compressed Images*: Figure 5 and 6 summarize the prediction results that can be obtained using the selected near-lossless coding technique,

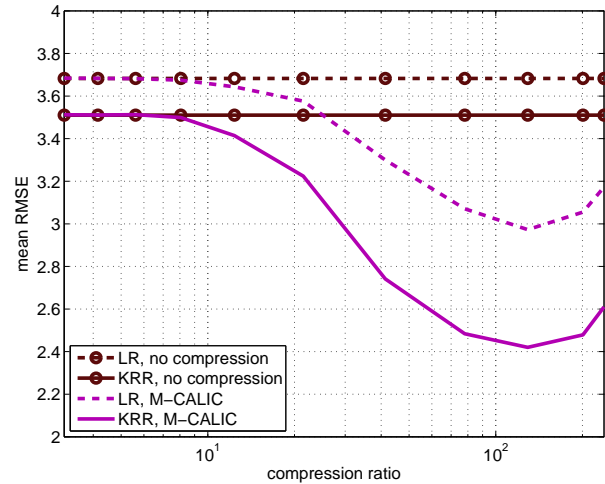


Fig. 5. **Estimation using LR or KRR for M-CALIC compression.** The horizontal axis represents the compression ratio, and the vertical axis represents the mean RMSE (dew point temperature in K) over the different pressure levels. The plot shows the results for M-CALIC compression when using LR (dashed line) and KRR (solid line) for predicting moisture as compared to retrieval results on original images.

M-CALIC, and the investigated regression methods, LR and KRR. Figure 5 shows the average root mean square error (RMSE) across the whole atmospheric column as a function of the compression ratio and the RMSE for each specific pressure level, respectively. One can see that while small compression ratios (left part of the plot) keep the results almost unchanged, larger compression ratios benefit the retrieval performance. Nevertheless, when the level of distortion inserted –due to the excessive compression– is increased (right-most part of the plot), the prediction results start to deteriorate. The effect is the same for both LR and KRR, but with improved accuracy for KRR. These results, although counter-intuitive, are consistent with facts and results reported before in the literature. We refer the reader to Section IV for a lengthy discussion providing explanations and examples for this behaviour.

Figure 6 reports the RMSE results for the different pressure levels (measured in hectoPascals) achieved after a near-lossless compression at compression ratio 129:1; again, we can observe that compression benefits the retrieval estimation and that KRR yields a higher performance than LR.

2) *Retrieval Assessment for Lossy Compressed Images*: As Figure 6 in the case of near-lossless compression, Figure 7 reports the RMSE results for the whole range of pressure levels. The conclusions above apply here too: introducing compression benefits the retrieval estimation and KRR is superior to LR.

Next, Figure 8 summarizes the prediction results that can be obtained using the selected lossy compression approach (JPEG 2000 with several compression configurations) and LR or KRR. The figure shows the average of the RMSE prediction over the different pressure levels. Several relevant conclusions can be extracted from this figure:

- *KRR always yields better results than LR.* It is clear from the individual plots that, for the same configuration of spatial and spectral compression methods, KRR always

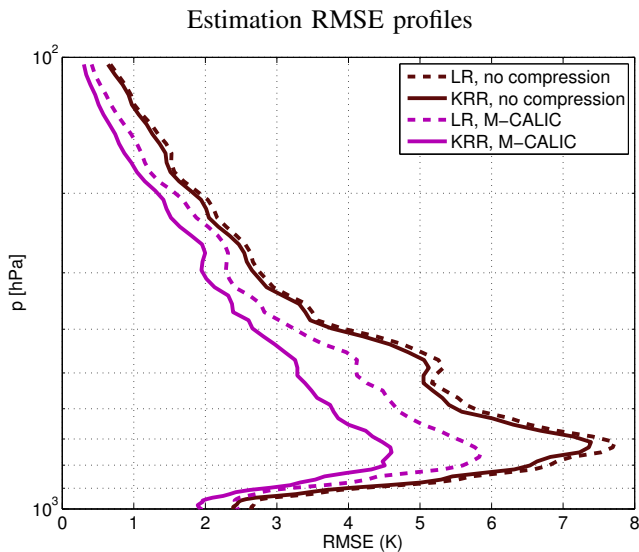


Fig. 6. **Moisture RMSE (dew point temperature in K) profiles for near-lossless compression.** We chose the PAE with the best average RMSE. For LR (dashed line) and KRR (solid line), the minimum averaged RMSE is obtained at a compression ratio of 129:1. Results when using original images are shown as well for comparison purposes.

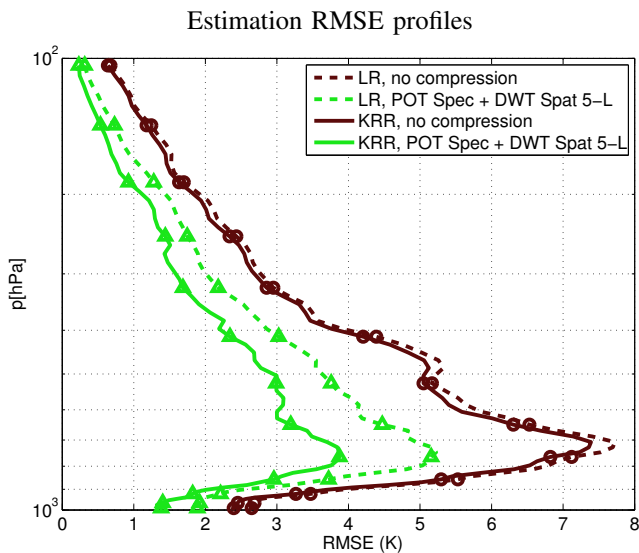


Fig. 7. **Moisture RMSE (dew point temperature in K) profiles for lossy compression.** The selected configuration for multi-component JPEG 2000 lossy compression is POT as spectral transform and 5 levels of spatial DWT. We chose the compression ratio with the best average RMSE. For LR, the minimum averaged RMSE is obtained at a compression ratio of 3200:1. For KRR, the minimum averaged RMSE is obtained at a compression ratio of 1600:1. Results when using original images are shown as well for comparison purposes.

produces less error than LR.

- *Using compressed images produces better results than using original images.* This statement is derived by comparing the top-left plot (original images) with the other plots in the Figure. Actually, this is confirmed within a wide bit-rate range. Although it could seem counter-intuitive at first glance (compressed images carry less information than original ones), compression algorithms capture and convey the important information, which

allows the prediction algorithms to focus on just *this spatial-spectral compact information*, and thus they can yield a better performance. Of course, when the bit-rate is very small (i.e., a high compression is applied) the remaining information is very small and the prediction performance decreases regardless of the capacity of the regression method (either LR or KRR). We argue that the positive influence is an effect of the spatial-spectral compression applied, which can be seen as an efficient (yet indirect) way to include spatial-spectral relations in the retrieval algorithm (cf. Section IV).

- *Spatial transform is more important than spectral transform.* If we look at the plots within the same row, going from left to right (from a lower to a larger number of levels of spatial wavelet transform) the performance of the retrieval algorithms increases. This effect is less noticeable across columns, i.e., using different spectral transforms or a different number of transform levels for the same wavelet-based spectral transform does not have significantly different effects on the retrieval performance. Note that prediction is performed using single hyperpixels, therefore the prediction models are not using spatial information directly, although they are indirectly using it through the compressed data. Actually, spatial transform fuse the spatial information in a way that the compressed hyperpixels carry some information about the close-by hyperpixels, while original hyperpixels only carry the spectral information. As such, spatial transform helps to introduce some spatial information in the prediction (cf. Section IV).
- *Using both spectral and spatial transform yields the best retrieval performance for low bit-rates.* Using an extremely low bit-rate (0.005 bpppc) produces a very good performance in retrieval accuracy, see, e.g., the right-most plot in the second row (green triangles) and in the last row (gray diamonds). These would be the best configurations if the interest is to optimize performance and compression ratio simultaneously.

Finally, Figure 9 compares the performance of the different compression schemes considering the compression ratio that yields the lower mean RMSE. Again, the conclusion is that using both spectral and spatial transform yields the best retrieval performance.

#### IV. DISCUSSION

As reported in Figures 5, 6, 7 and 8, compressing the images before carrying out the statistical retrieval for weather forecast brings a significant improvement. In order to understand this observed positive effect, we should first review some well-known facts in signal and image processing. First, it is acknowledged that signal coding is a way of performing signal filtering and denoising. This is a known fact reported elsewhere [24]–[27]. The connections between denoising and compression have been actually widely studied theoretically [39] and experimentally in general and for wavelet transforms in particular [26], [27], [40]. Therefore, it goes without saying that when one performs lossy signal compression (coding) the obtained signal is denoised.

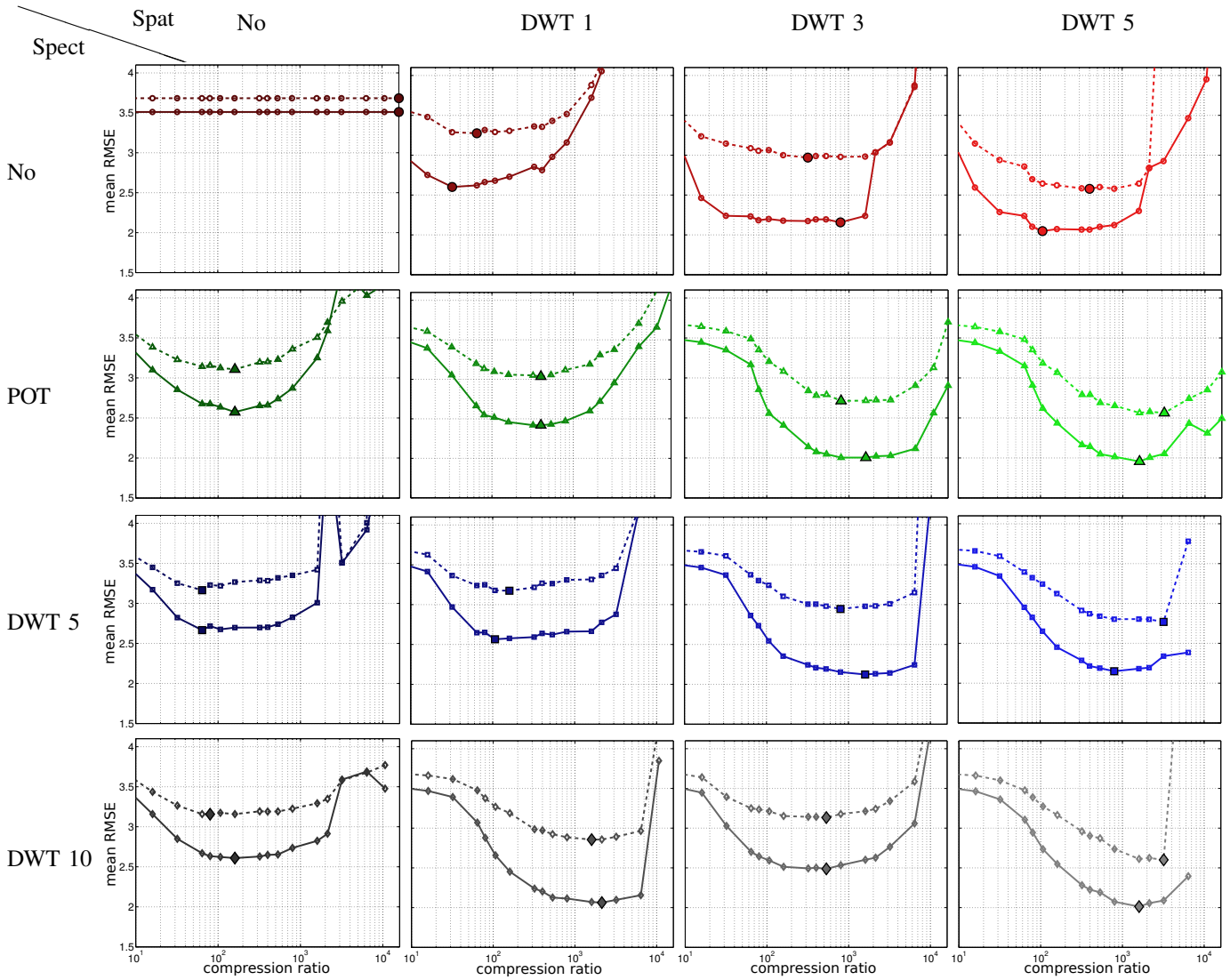


Fig. 8. Estimation using LR or KRR for different spatial and spectral configurations for lossy compression. In all the plots, the horizontal axis represents the compression ratio, and the vertical axis represents the averaged RMSE (dew point temperature in K) over the different pressure levels. Ranges are the same in all the plots to ease the comparison. Each plot shows the results for a particular compression configuration when using LR (dashed lines) and KRR (solid lines) for predicting moisture. Configurations using the same number of levels for the spatial wavelet transform are in the same column and configurations using the same spectral transform are in the same row (same color, multiple spatial nor spectral). The minimum averaged RMSE in each curve is indicated with a larger marker point.

An additional important observation, relevant to our discussion, is the fact that using a filtered (denoised) signal typically improves regression and function approximation results. The presence of noise in the observations obviously hampers estimation of the underlying signal. This is why very often when performing regression, one aims at estimating the noise (or its covariance) where the signal was buried. Such noise estimate is then used to discount the uninformative noise contributions in the observation, which definitely helps in recovering the signal. There is a vast literature in both noise (covariance) estimators and how one can embed these priors in, for example, Generalized Least Squares (GLS) [41], which has been extended for nonlinear regression under a Bayesian framework in [42] and the kernel framework in [43], [44]. However, noise estimation is typically difficult, especially

when signal-dependent noise (heteroscedasticity), structured domains (as in time series or images), or outlying samples are present. A much simpler and more practical approach is to just filter out the observed signal before applying the regression algorithm. Such filtering can be done with standard PCA projection on the top eigenvectors (thus assuming that noise is related to the higher frequency components), via wavelets (thresholding coefficients in particular scales or orientations whenever some knowledge about the signal is available), or via iterative thresholding methods (that iteratively refine the signal estimation). In conclusion, as mentioned before, compression schemes constitute an alternative, indirect, simple way to filter signals.

Actually, using wavelet-based and M-CALIC coding schemes yields an extremely useful by-product: they rein-



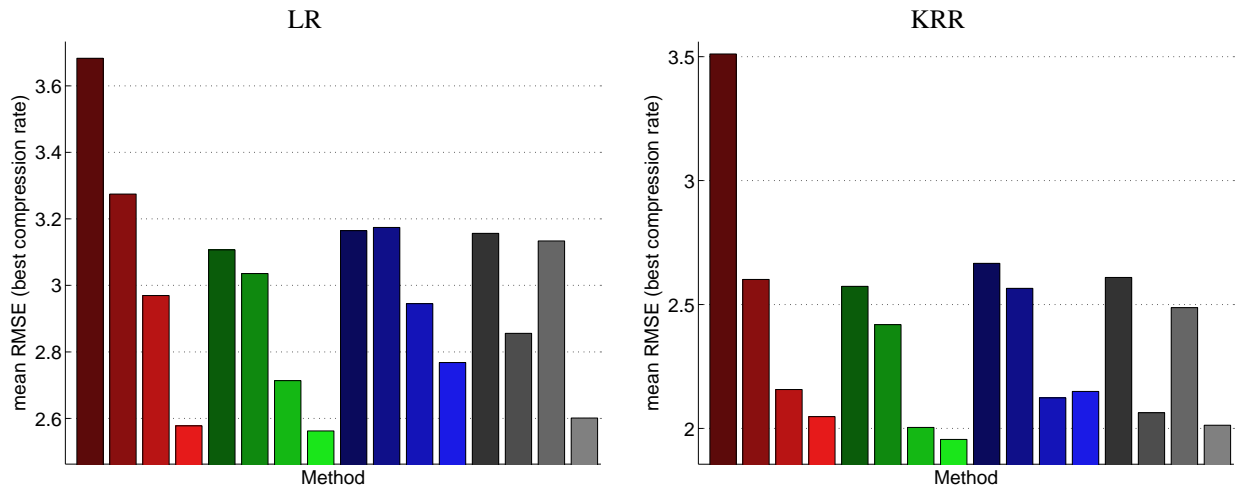


Fig. 9. Performance of the different lossy compression schemes for predicting moisture. The vertical axis reports the minimum averaged RMSE (dew point temperature in K) achieved for each compression scheme. The horizontal axis represents the different lossy compression configuration schemes; the ordering follows that of Figure 8, first by type of spectral transform (same color) and then by number of levels for the spatial wavelet transform (color shade).

force/encode spatial relations in the generate product. Coding schemes such as the ones used in this work (wavelet-based multi-component JPEG 2000 for lossy compression and M-CALIC for near-lossless compression) typically reduce the noise variance *but also* encode spatial smoothness in the compressed image. Compression algorithms aim to concentrate information in a reduced set of coefficients, which is done through ‘decorrelation’ of the observation space (either spatial, spectral or both), and very often assume stationarity (hence smoothness) implicitly. This is obviously an indirect, simple way to enforce local (spatial) relations between nearby pixels. Image compression therefore enforces pixel correlations and smoothness in the image plane, that can be exploited in pixel-wise regression algorithms to improve the spatial homogeneity/consistency of the prediction maps too. In fact, it is widely acknowledged that model inversion and parameter retrieval in remote sensing applications largely improve when including spatial information, see [45], [46] and references therein. Moreover, in recent years, it is being widely accepted that using the spectral information only (as done by pixel-wise algorithms), is not optimal, and that it is important to find a trade-off between the spatial and the spectral information [47].

Finally, we should mention that these observations are not new in signal and image processing. Several applications benefit from a coding-first scheme. There is a wide evidence that compressing the signal before any further processing is beneficial. We have shown this effect previously for several compression algorithms in hyperspectral image classification [48], and it has been a common strategy in other applications such as genomics –in particular for Next Generation Sequencing (NGS) [49], [50]–, seismic signal processing [51], [52], bioengineering [53] and communications [54]. Therefore, we posit that what we report is yet another case of an effect previously noticed in signal and image processing, here transported to the field of bio-geo-physical parameter estimation in atmospheric applications using infrared sounding data.

## V. CONCLUSIONS

This paper was concerned with how to include spatial-spectral information for retrieving atmospheric profiles of temperature and humidity using infrared sounders and statistical regression. While many sophisticated approaches can be deployed for this purpose (e.g., contextual and spatial-spectral feature extraction or filtering at a postprocessing stage), they typically involve hand-crafted features and engineering new and complicated retrieval algorithms, non-automatic processes, and tedious user intervention. Last but not least, such processes raise important issues about parameter tuning and the little control one has on the retrieval generalization capabilities. Alternatively, we approached the problem in an indirect way by compressing the hyperspectral cubes before performing retrieval. Compression techniques are well-understood and the impact on the quality of the radiances can be easily controlled by prescribing a compression ratio. A wide range of spatial, spectral and spatial-spectral image compression approaches were evaluated and two linear and nonlinear regression algorithms were compared. Results for prediction of both temperature and humidity can be found at [http://isp.uv.es/coding\\_retrieval.html](http://isp.uv.es/coding_retrieval.html).

Near-lossless compression was carried out through the M-CALIC coding technique, while lossy compression was based on multi-component JPEG2000 standard. Even though both compression approaches introduce certain level of distortion on the original data, they simultaneously incorporate the spatial-spectral feature relations, so there is no need to design specific retrieval algorithms that fuse such information. In our experiments, we searched for an acceptable compression ratio on IASI L1C products such that it has a positive impact on the accuracy of the statistical retrievals, i.e., an optimal trade-off between the positive effect of the compression, i.e., the introduction of spatial relations and some sort of denoising on the data, and its negative effect, i.e., the reduction of information as compared to the original image. Our first conclusion is interesting enough, and we think that the message can be of

high interest to the parameter retrieval community in general and to scientists and organizations dealing with atmospheric retrieval models in particular. We show that compression is a very easy, cheap, and consistent way to perform denoising and to introduce spatial-spectral image relations in the compressed product that largely benefits the current (point-wise) retrieval. We observed that high compression ratios yielded improved results in predicting dew point temperature (similar results for temperature were obtained—see supplementary material—) in all the experiments for moderate and high compression ratios. Of course when the compression ratio is extremely large, this effect vanishes, since the amount of information removed by the compression is too large.

Our second message we want to convey in this communication is that a spatial transform revealed itself to be more important than a spectral transform. This observation is related to the fact that the retrieval algorithms work in a pixel-wise manner, and hence the spatial component is missing. This is why the spectral-spatial image compression configuration results in such big gains in spectrum-based retrieval. We will explore in future works the trade-off between the gains obtained when using a scheme based on *spectral-spatial compression plus spectral retrieval* versus a more complete yet challenging *spectral-spatial compression plus spectral-spatial retrieval*. This second approach would require, however, re-designing retrieval algorithms, which in most of the cases are pixel-based.

As the main conclusion, we want to highlight that compression prior to model inversion is largely beneficial in retrieval problems in general. The benefit is not only of practical use but also computationally convenient and more accurate. It is *practical* because following the presented methodology does not change current operational approaches much: 1) current retrieval schemes would be still valid since they would be fed with efficiently compressed data before the retrieval instead of the original data, and 2) there are moderate-to-big savings in data storage and transmission because of the compression step. The benefit is also in terms of prediction accuracy, as we have observed consistent gains in all cases and images. We believe that the proposed methodology may benefit the development of current and upcoming infrared sounding and hyperspectral sensors to advance in bio-geo-physical parameter estimation schemes. It has not escaped our notice that the same procedure can be actually applied to other algorithms for retrieval, such as optimal estimation scheme or Bayesian approaches, and to other problems dealing with high resolution data, such as spectral unmixing. These issues are subject of further research studies.

## VI. ACKNOWLEDGEMENTS

We want to thank Dr. Thomas August (EUMETSAT, Germany) for the provided data and the fruitful discussions on retrievals and spatial-spectral coding.

## REFERENCES

[1] EUMETSAT, *IASI Level 1: Product Guide*, EUM/OPS-EPS/MAN/04/0032, 2014.

- [2] B. Tournier, D. Blumstein, F. Cayla, , and G. Chalon, "IASI level 0 and 1 processing algorithms description," in *Proc. of ISTCXII Conference*, 2002.
- [3] G. Camps-Valls, J. Muñoz-Mari, L. Gomez-Chova, L. Guanter, and X. Calbet, "Nonlinear statistical retrieval of atmospheric profiles from MetOp-IASI and MTG-IRS infrared sounding data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1759–1769, 2012.
- [4] G. Mercier, M. Mouchot, and G. Cazuguel, "Joint classification and compression of hyperspectral images," in *IEEE International and Remote Sensing Symposium (IGARSS)*, vol. 4, 1999, pp. 2035–2037.
- [5] M. Pal, C. Brislawn, and S. Brumby, "Feature extraction from hyperspectral images compressed using the JPEG 2000 standard," in *Fifth IEEE Southwest Symposium in Image Analysis and Interpretation*, 2002, pp. 168–172.
- [6] A. Kaarna, P. Toivanen, and P. Keränen, "Compression and classification methods for hyperspectral images," *Pattern Recognition and Image Analysis*, vol. 16, no. 3, pp. 413–424, 2006.
- [7] I. Blanes and J. Serra-Sagrìstà, "Quality evaluation of progressive lossy-to-lossless remote-sensing image coding," in *16th IEEE International Conference in Image Processing (ICIP)*, 2009, pp. 3709–3712.
- [8] 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, 2007.
- [9] Q. Du and J. E. Fowler, "Hyperspectral image compression using JPEG 2000 and principal component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 201–205, 2007.
- [10] I. Blanes, A. Zabala, G. Moré, X. Pons, and J. Serra-Sagrìstà, "Classification of hyperspectral images compressed through 3D-JPEG 2000," *Knowledge-Based Intelligent Information and Engineering Systems. Springer Berlin Heidelberg*, pp. 416–423, 2008.
- [11] G. Carvajal, B. Penna, and E. Magli, "Unified lossy and near-lossless hyperspectral image compression based on JPEG 2000," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 593–597, 2008.
- [12] F. García-Vílchez, J. Muñoz-Marí, M. Zortea, I. Blanes, V. González-Ruiz, G. Camps-Valls, A. Plaza, and J. Serra-Sagrìstà, "On the impact of lossy compression on hyperspectral image classification and unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 2, pp. 253–257, 2011.
- [13] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Kluwer, 2003.
- [14] Q. Du and C. I. Chang, "Linear mixture analysis-based compression for hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 875–891, 2004.
- [15] A. Plaza, P. Martínez, R. Pérez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2025–2041, 2002.
- [16] M. Zortea and A. Plaza, "Spatial preprocessing for endmember extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2679–2693, 2009.
- [17] J. García-Sobrino, I. Blanes, V. Laparra, G. Camps-Valls, and J. Serra-Sagrìstà, "Impact of near-lossless compression of IASI L1C data on statistical retrieval of atmospheric profiles," in *Proc. of On-Board Payload Data Compression Workshop (OBPDC)*, Venice, Italy, October 2014.
- [18] 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, January 2004.
- [19] D. S. Taubman and M. W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards, and Practice*. Springer Science & Business Media, 2012.
- [20] JPEG-Committee, *Standard JPEG 2000*, document ISO/IEC 15444. [Online]. Available: <http://www.jpeg.org/jpeg2000/documentation.html>
- [21] I. Blanes and J. Serra-Sagrìstà, "Pairwise orthogonal transform for spectral image coding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 961–972, March 2011.
- [22] I. Blanes, J. Serra-Sagrìstà, M. W. 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.
- [23] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
- [24] B. K. Natarajan, "Filtering random noise via data compression," in *Proceedings IEEE Data Compression Conference*, 1993, pp. 60–69.
- [25] —, "Filtering random noise from deterministic signals via data compression," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2595–2605, 1995.

- [26] P. L. Dragotti and M. Vetterli, "Footprints and edgeprints for image denoising and compression," in *Proceedings IEEE International Conference on Image Processing*, vol. 2, 2001, pp. 237–240.
- [27] —, "Wavelet and footprint sampling of signals with a finite rate of innovation," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004, pp. 941–944.
- [28] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Transactions on Communications*, vol. 45, no. 4, pp. 437–444, April 1997.
- [29] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Transactions on Image Processing*, vol. 9, no. 8, pp. 1309–1324, August 2000.
- [30] B. Schölkopf and A. Smola, *Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press Series, 2002.
- [31] E. Magli, "M-CALIC software," <http://www1.tlc.polito.it/oldsite/sas-ipl/download.php>.
- [32] D. Taubman, "Kakadu software," <http://www.kakadusoftware.com/>.
- [33] *JPEG 2000 Part 2 - Extensions*, document ISO/IEC 15444-2.
- [34] GICI-UAB, "Spectral transform," <http://gici.uab.cat/GiciWebPage/downloads.php>.
- [35] E. Izquierdo-Verdiguier, V. Laparra, L. Gómez-Chova, and G. Camps-Valls, "Encoding invariances in remote sensing image classification with svm," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 1–5, 2013.
- [36] M. U. Gutmann, V. Laparra, and A. H. J. Malo, "Spatio-chromatic adaptation via higher-order canonical correlation analysis of natural images," *PLoS ONE*, 2014.
- [37] D. Tuia and G. Camps-Valls, "Kernel manifold alignment for domain adaptation," *PLoS ONE*, vol. 11, no. 2, 2016.
- [38] S. Srivastava, P. K. Thapliyal, M. V. Shukla, J. S. H. Bisht, and D. Mitra, "Statistical retrieval of ozone and meteorological parameters using SHADOZ observations and radiative transfer model," *Journal of the Indian Society of Remote Sensing*, pp. 1–8, February 2016.
- [39] T. Weissman and E. Ordentlich, "The empirical distribution of rate-constrained source codes," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3718–3733, 2005.
- [40] S. G. Chang, B. Yu, and M. Vetterli, "Image denoising via lossy compression and wavelet thresholding," in *Image Processing, 1997. Proceedings., International Conference on*, vol. 1, Oct 1997, pp. 604–607 vol.1.
- [41] J. Johnston, *Econometric Methods*, 1972. [Online]. Available: <https://books.google.es/books?id=BZtVwZAGyV0C>
- [42] M. Lázaro-Gredilla, M. K. Titsias, J. Verrelst, and G. Camps-Valls, "Retrieval of biophysical parameters with heteroscedastic gaussian processes," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 4, pp. 838–842, April 2014.
- [43] G. Camps-Valls and L. Bruzzone, Eds., *Kernel methods for Remote Sensing Data Analysis*. UK: Wiley & Sons, Dec 2009.
- [44] L. Gomez-Chova and G. Camps-Valls, "Learning with the kernel signal to noise ratio," 2012, cited By 0. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84870658481&partnerID=40&md5=702f25c382d3d87f1d5c9b3509995474>
- [45] S. Jiménez and J. Malo, "The role of spatial information in disentangling the irradiance-reflectance-transmittance ambiguity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4881–4894, 2014.
- [46] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jiménez, and J. Malo, *Remote Sensing Image Processing. Synthesis Lectures on Image, Video, and Multimedia Processing*. Morgan & Claypool Publishers, 2011.
- [47] V. Laparra and R. Santos-Rodríguez, "Spatial/spectral information trade-off in hyperspectral images," in *Proceedings IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 1124–1127.
- [48] F. García-Vílchez, J. Muñoz-Marí, M. Zortea, I. Blanes, V. González-Ruiz, G. Camps-Valls, A. Plaza, and J. Serra-Sagrìstà, "On the impact of lossy compression on hyperspectral image classification and unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 2, pp. 253–257, 2011.
- [49] Y. W. Yu, D. Yorukoglu, and B. Berger, "Traversing the k-mer landscape of ngs read datasets for quality score sparsification," in *Research in Computational Molecular Biology, Springer Lecture Notes in Computer Science*, vol. LNBI 8394, 2014, pp. 385–399.
- [50] I. Ochoa, M. Hernaez, R. Goldfeder, T. Weissman, and E. Ashley, "Denoising of quality scores for boosted inference and reduced storage," in *Proceedings IEEE Data Compression Conference*, 2016.
- [51] M. F. Khene and S. H. Abdul-Jauwad, "Adaptive seismic compression by wavelet shrinkage," in *Statistical Signal and Array Processing, 2000. Proceedings of the Tenth IEEE Workshop on*, 2000, pp. 544–548.
- [52] J. R. Jr and F. G. Meyer, "Machine learning for seismic signal processing: Phase classification on a manifold," in *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, vol. 1, Dec 2011, pp. 382–388.
- [53] A. A. Petrosian and F. G. Meyer, Eds., *Wavelets in signal and image analysis: from theory to practice*, ser. Computational imaging and vision. Dordrecht, Boston, London: Kluwer Academic, 2001. [Online]. Available: <http://opac.inria.fr/record=b1098905>
- [54] L. Hanzo, F. C. Somerville, and J. Woodard, *Voice and Audio Compression for Wireless Communications*. John Wiley & Sons, Ltd, 2007. [Online]. Available: [http://dx.doi.org/10.1002/9780470516034\\_fmatter](http://dx.doi.org/10.1002/9780470516034_fmatter)

Post-print of: García-Sobrino, J. et al. "Statistical atmospheric parameter retrieval largely benefits from spatial-spectral image compression" in *IEEE transactions on geoscience and remote sensing*, vol. 55, issue 4 (april 2017), p. 2213-2224. DOI 10.1109/TGRS.2016.2639099 - Cop. 2016 IEEE. Personal use of this material is permitted. Permissions from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.