

This is the **accepted version** of the article:

García-Sobrino, Joaquín; Laparra, Valero; Serra Sagristà, Joan; [et al.]. «Improved Statistically Based Retrievals via Spatial-Spectral Data Compression for IASI Data». IEEE transactions on geoscience and remote sensing, Vol. 57, Issue 8 (August 2019), p. 5651 - 5668. DOI 10.1109/TGRS.2019.2901396

This version is available at <https://ddd.uab.cat/record/215775>

under the terms of the  **IN**
COPYRIGHT license

Improved Statistically Based Retrievals via Spatial-Spectral Data Compression for IASI data

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

Abstract—In this paper we analyze the effect of spatial and spectral compression on the performance of statistically based retrieval. Although the quality of the information is not completely preserved during the coding process, experiments reveal that a certain amount of compression may yield a positive impact on the accuracy of retrievals. We unveil two strategies, both with interesting benefits: either to apply a very high compression, which still maintains the same retrieval performance as that obtained for uncompressed data; or to apply a moderate to high compression, which improves the performance. As a second contribution of this work, we focus on the origins of these benefits. On the one hand, we show that a certain amount of noise is removed during the compression stage, which benefits the retrievals performance. On the other hand, we analyze the effect of compression on spectral/spatial regularization (smoothing). We quantify the amount of information shared among the spatial neighbours for the different methods and compression ratios. We also propose a simple strategy to specifically exploit spectral and spatial relations and find that, when these relations are taken into account beforehand, the benefits of compression are reduced. These experiments suggest that compression can be understood as an indirect way to regularize the data and exploit spatial neighbours information, which improves the performance of pixel-wise statistics based retrieval algorithms.

Index Terms—Infrared Atmospheric Sounding Interferometer (IASI), Lossy Compression, Spectral Transforms, JPEG 2000, Statistically Based Retrieval, Kernel Methods, Regression.

I. INTRODUCTION

In recent decades, advances in remote sensing technology have made it possible to collect information from the electromagnetic spectrum with unprecedented accuracy and resolution [1]. Infrared sounder instruments usually produce large volumes of data, which is costly to manage in an operational context, i.e., for transmission, processing, and storage. An effective strategy to alleviate the problems derived from the data size is to compress the data according to the specific needs of the final users. However, in order to achieve high compression ratios, it is typically needed to go

through a lossy compression stage. Lossy compression entails to lose some information in the reconstructed data, which may compromise the quality of the products in later processing stages. Therefore, a careful evaluation of the impact of the compression process applied to the data is needed to determine the quality of the information and of the derived products.

Although lossy compression implies going through a distortion process, the quality of the recovered data can be still adequate for the intended specific use. If the amount of the inherent data acquisition noise is large, the signal removed during the coding process is mostly noise for certain compression ratios [2], [3]. Therefore, the amount of useful signal lost is usually low and the quality of the recovered products may be yet appropriate, or even better suited, to feed a subsequent information extraction stage. Of course, if the compression ratio is very high, a lot of useful information is removed and the quality of the reconstructed data is seriously compromised.

The impact of lossy compression on information extraction systems has been investigated in several application areas. Ryan and Arnold discussed the lossy compression of remotely sensed images using vector quantization in the context of image classification in [4]. The impact of compression was analyzed on the maximum likelihood classification of the recovered data concluding that the loss in the classification accuracy was not significant (less than 8%). Sánchez and Peronnin evaluated two lossy compression techniques, one based on product quantizers and another based on dimensionality reduction, to assess large-scale image classification in [5]. Results reported that compression ratios between 64:1 and 128:1 produced little loss in the classification performance. Similar results were obtained in [6], [7], [8], [9], [10], [11], [12], where lossy compression did not markedly reduce classification accuracy.

Besides, several studies have analyzed the effect of near-lossless and lossy compression techniques based on JPEG 2000 standard within the framework of feature extraction, classification, and anomaly detection tasks. Pal et al. analyzed the performance of supervised, unsupervised, and hybrid classification processes in [13]. Results reflected that classification accuracy was still reliable even at low bit-rates (high compression ratio). A compression scheme based on Principal Component Analysis (PCA) was proposed in [14] producing competitive results in terms of information preservation in anomaly detection tasks. Supervised and unsupervised classification of reconstructed data was evaluated in [15]. The experimental results showed competitive classification performance after the

Manuscript received December 17, 2019;

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: joaquin.garcia.sobrino@deic.uab.cat, joan.serra@uab.cat. VL and GCV are with the Image Processing Laboratory (IPL), Universitat de 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 and RTI2018-095287-B-I00, by the Catalan Government under Grant 2017SGR-463, and by the European Research Council under Consolidator Grant SEDAL ERC-2014-CoG 647423.

compression stage. Analogous results were reported in [16], where the proposed compression algorithm for hyperspectral images did not significantly reduce the performance of hard classification, linear spectral unmixing, and anomaly detection.

An interesting observation was reported in [17], where the impact of lossy compression of hyperspectral data was analyzed on multiclass classification, classification of mixed pixels via spectral unmixing, binary hard classification, and anomaly detection. Experimental results revealed that for multiclass classification and spectral unmixing, the Signal-to-Noise ratio (SNR) was a reasonable indicator of classification performance; for binary hard classification, the performance was little dependent on the SNR; and for anomaly detection, the compression algorithms that produced the best rate-distortion performance were not the best choice. Actually, the effect of lossy compression on supervised classification and spectral unmixing using support vector machines was quantified in [2]. The assessment reported that lossy compression can produce accurate results even at high compression ratios (above 16:1).

The estimation of atmospheric parameters from remote sensing data is an *inverse problem* where we retrieve the physical parameters given a set of observations. Inverting the physical (radiative transfer) model through look-up-tables or optimal estimation models (OEM) [18] are standard approaches, but they lead to much higher computational costs than those required for statistical approaches [19]. In the last decade, statistical model inversion based on machine learning has provided excellent performance in accuracy and efficiency terms. Statistically based atmospheric parameter retrieval from IASI and AIRS data was first conducted using vanilla neural networks trained with backpropagation [20], [21], [22]. In a set of previous comparisons [19], [23], [24], [25], [26] we used other forms of nonlinear and nonparametric machine learning regression like more advanced neural nets, kernel ridge regression and Gaussian processes. Other strategies to plain regression consider introducing a dimensionality reduction before the regression. Smart approaches consider spatially and noise-aware transforms like in [25], [27], or more advanced *nonlinear* dimensionality reduction approaches based on kernel machines [28].

The functional sliced inverse regression (FSIR) in [29] is a dimensionality reduction method that generalizes PCA by designing a spectral information driven by the parameters to be retrieved. FSIR is actually related to linear discriminant analysis (LDA) and partial least squares (PLS) approaches, being a kind of supervised dimensionality reduction approach, and provides noise-aware feature components (hence being related to minimum noise fraction -MNF- methods). Nevertheless, the standard PCA approach is still the most widely adopted and studied for dimensionality reduction in atmospheric parameter retrieval. In [30] the impact of spectral compression on statistically based retrievals was analyzed, investigating the use of Principal Components Analysis (PCA) based methods for compressing high-resolution infrared measurements before performing linear regression. It was shown that a compression ratio of 15:1 can be performed with low degradation of the temperature and water vapour retrievals. In

[31], [32] we observed that data that had gone through a near-lossless and lossy compression process enabled achieving, for certain compression ratios, improved statistically based retrieval performance compared to the retrieval on the original products. It was suggested that the spectral/spatial regularization and the noise filtering produced by the compression stage benefit the performance of the statistical retrieval algorithms. However, these counterintuitive observations were scarcely studied. In this paper, we provide appropriate responses to the observations reported in [31], [32], investigate the actual impact of compression in the retrievals, and supply results and discussion to understand why lossy compression may give rise to improved retrievals. This led us to lay out the main contributions of this paper:

- 1) While experiments in [31], [32] were carried out in an ideal scenario, where samples to define the training, validation and testing subsets were acquired under the same conditions, here, we assume a more realistic scenario, where samples used to define the training, the validation and the testing subsets come all from disjoint orbits. Therefore, conclusions are not biased by an eventual risk of model overfitting.
- 2) Three different multi-component transforms have been employed in the compression process, all bringing forth improved performance.
- 3) The causes of the benefits on the retrieval performance due to a compression stage are thoroughly investigated and discussed, and reasons of the statistically based retrieval improvement are provided.

In this paper, a wide spread lossy compression technique and two different statistical retrieval algorithms are evaluated on Infrared Atmospheric Sounding Interferometer (IASI) L1C data [33]. For lossy compression, JPEG 2000 standard [34] is paired along with three different spectral transforms: the Discrete Wavelet Transform (DWT) [35], the Pairwise Orthogonal Transform (POT) [36]), and the Multilevel Clustering Karhunen-Loève Transform (Multilevel Clustering KLT) [37]. For atmospheric variables prediction, a linear and a nonlinear statistically based retrieval algorithm, i.e., the Linear Regression (LR) and the Kernel Ridge Regression (KRR) [38], respectively, are employed in the retrieval of physical information (temperature and dew point temperature profiles) from the reconstructed data.

To investigate the origin of the improved retrieval performance produced by lossy compression, two approaches are followed. On the one hand, we analyze the noise level remaining in the recovered spectra. On the other hand, we compare the performance of the retrieval algorithms when a compressed (and reconstructed) data is employed, with a simple method which specifically takes into account information from the spatial and spectral neighbours. Experiments reveal that the improvement in retrieval accuracy is motivated by: 1) lossy compression performing some noise filtering, which typically improves regression and function approximation results; and 2) compression being an indirect way to exploit spatial feature relations, which, generally, helps in pixel-wise retrieval algorithms.

The remainder of the paper is organized as follows. Section II introduces the proposed sequential approach detailing the compression scheme and the statistically based retrieval methods used in the experiments. Section III reports the data collection used in the experiments. Section IV reports and analyzes the experimental results. Section V provides an extensive discussion. Finally, Section VI draws some conclusions.

II. METHODOLOGY

This section describes the methods and techniques used in the experiments. For the compression stage, the JPEG 2000 standard is paired along with three spectral transforms to exploit the high spectral redundancy present in IASI L1C data. For atmospheric prediction, two different statistically based retrieval algorithms, which have provided competitive performance in atmospheric parameter retrieval [39], are employed: a standard least squares LR and a kernel ridge regression (KRR) method.

Figure 1 illustrates the proposed sequential scheme. In a first stage, the data is lossy compressed. At the receiver side, data is decompressed to produce the reconstructed data. Finally, the statistical parameter retrieval is carried out on the reconstructed data.

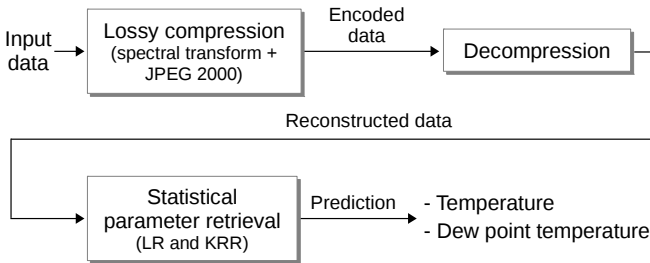


Fig. 1. Proposed sequential chain.

A. Data Compression

Compression of remote sensing products is an efficient strategy to reduce the large size of the collected data. In recent years, various compression strategies have been adopted in several standards and have been implemented both on board the satellites and on-the-ground processing stations [40], [41], [42], [43], [44].

Three compression approaches can be considered, namely, lossless, lossy, and near-lossless. Lossless compression is the mandatory approach when the original signal must be fully preserved, however, the achievable compression ratios are rather low (below 5:1). Lossy compression can be a desirable approach in scenarios where high compression ratios are required and losing some information in the signal might be admissible. Near-lossless compression is also a lossy compression approach, where some information is removed from the original data, but the compression ratios achieved are closer to lossless compression than to typical lossy compression. Near-lossless compression is an adequate approach when higher compression ratios compared to lossless compression are demanded and, at the same time, a specific fidelity criterion

(usually the peak absolute error) must be preserved in the recovered data.

Here we will focus on *lossy compression* because it is a widely accepted strategy when large volumes of remote sensing data need be compressed, as witnessed in the today-in-use pipeline for IASI data [45], [46] (recall that we will use IASI data for the experiments).

1) *Lossy Compression*: Lossy compression approach allows to achieve high compression ratios at the expense of recovering a reconstructed data not identical to the original ones. It is expected that the reconstructed data preserve enough information to be used for the intended specific purpose, while alleviating problems derived from the transmission, handling, and storage of large volumes of information.

The international standard JPEG 2000 [34] is used to carry out the lossy compression stage. JPEG 2000 was published by the *Joint Photographic Experts Group (JPEG)* in year 2000 as the successor of classical JPEG coding standard, and provides a wide diversity of features and functionalities in a single compressed code-stream: among others, lossless and lossy compression, progressive lossy-to-lossless coding, robustness to the presence of errors, region-of-interest coding, and progressive transmission. JPEG 2000 is employed in a wide range of applications, e.g., remote sensing, medical imagery, mobile applications, digital library, digital photography, etc.

To achieve competitive compression performance in data formed by thousands of spectral components, like IASI L1C products, it is of paramount importance to exploit the redundancy present in the spectral dimension. Three spectral transforms are paired along with JPEG 2000: the Discrete Wavelet Transform (DWT), the Pairwise Orthogonal Transform (POT), and the Multilevel Clustering Karhunen-Loève Transform (Multilevel Clustering KLT).

DWT decomposes the processed signal into different subbands, which allows to decorrelate the data to be encoded. When the transform is applied, the signal is decomposed into two set of coefficients: the low and high frequency subbands. The low frequency subbands (L) have a coarse frequency resolution, while the high frequency subbands (H) represent fine details of the data. The high frequency subbands contain coefficients that can be discarded or quantized to achieve efficient lossy compression. Often, competitive data quality can be achieved from a small amount of transform coefficients, which makes DWT a suitable transform for lossy compression [47]. DWT can be applied on the transform coefficients (low frequency subbands) in successive levels producing further decorrelation, although the optimal number of transform levels shall depend on the particularities of the data. DWT can be used too to transform two dimensional signals by applying a one-dimensional transform in the vertical direction and a one-dimensional transform in the horizontal direction, which usually improves the coding results.

Both POT and Multilevel Clustering KLT are affordable approximations of the computationally demanding Karhunen-Loève Transform (KLT) [48], which is optimal decorrelating Gaussian sources. A critical consideration when applying a spectral transform on data volumes with large spectral dimension is the computational complexity [49]. Some spectral trans-

forms like the KLT may prove unusable for large dimensions such as in IASI data. POT and Multilevel Clustering KLT rely on the implementation of a divide-and-conquer strategy to the KLT. While a classic KLT decorrelates all components with each other regardless of how much energy they share, a divide-and-conquer strategy implements a mechanism to decorrelate several spectral components with high shared energy and ignore the other components. The resulting transform is the composition of smaller KLT transforms, where each composition is processed as a classic KLT. If a multilevel mode is used, most of the data energy flows across the composition of transforms up to the last level, because most of the signal energy is grouped in the first few decorrelated components, which are further decorrelated in upper levels. POT uses a two-component KLT transform for every pair of consecutive components and works in a multilevel mode. In the case of Multilevel Clustering KLT, each composition of KLT can be formed by a number of spectral components between two and all the spectral components. Figure 2 illustrates the structure of a classical KLT, of a POT and of a Multilevel Clustering KLT.

B. Statistically Based Parameter Retrieval

In the last decades, statistically based retrieval has proved to be very useful to solve inverse problems using remote sensing data. Different algorithms have been considered, but mainly restricted to neural networks [20], [21], [22], [25] and kernel machines and Gaussian processes [19], [23], [24], [26]. In particular, kernel methods are well suited due to its ability to deal with high dimensional data [50]. Kernel methods generalize linear algorithms while still relying on linear algebra. This is why we focus on two approaches for (first guess) parameter estimation: least squares (regularized) linear regression and its generalization using kernels. We will use classical linear regression (LR) and kernel methods in this work to implement the step of retrieving physical parameters from the data. We will use the Kernel Ridge Regression (KRR) algorithm [38], [51], which has shown very good performance in prediction when used on IASI L1 data on different problems [39], [52], [53]. KRR was the method employed in [31], [32] to analyze the effect of compression on physical parameters retrieval at different compression ratios. The KRR has the advantage of generalizing least squares linear regression to the nonlinear regression case.

1) *Statistically Based Retrieval Methods:* Here we define the notation and review the formulation of KRR from the LR solution. We will denote $\mathbf{x}_i \in \mathbb{R}^{D_x}$ as the inputs and $\mathbf{y}_i \in \mathbb{R}^{D_y}$ as the desired outputs. In matrix notation, we will denote the training input samples as $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_N] \in \mathbb{R}^{N \times D_x}$, the training desired outputs as $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_N] \in \mathbb{R}^{N \times D_y}$, the test input samples as $\mathbf{X}^* = [\mathbf{x}_1^* \dots \mathbf{x}_N^*] \in \mathbb{R}^{M \times D_x}$, and the test desired outputs $\mathbf{Y}^* = [\mathbf{y}_1^* \dots \mathbf{y}_N^*] \in \mathbb{R}^{M \times D_y}$. Using the matrix notation, the application of the linear model is:

$$\hat{\mathbf{Y}}^* = \mathbf{X}^* \mathbf{W}_L, \quad (1)$$

where we have discarded the bias term and $\mathbf{W}_L \in \mathbb{R}^{D_x \times D_y}$ are the regression weights. We fit the weights using classical

least squares solution, which depends on the inversion of $(\mathbf{X}^\top \mathbf{X})$. In order to ensure inversion we use the classical Tikhonov regularized solution:

$$\mathbf{W}_L = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{Y}. \quad (2)$$

Kernel methods are based on defining a mapping function $\phi(\mathbf{x})$ for the input samples to a Hilbert space, \mathcal{H} , of very high (possibly infinite) dimensionality $D_{\mathcal{H}}$. KRR can be defined following the same procedure as in the LR case as a linear least squares regression in the Hilbert space. If we map the samples using the mapping function we have $\Phi = \phi(\mathbf{X}) \in \mathbb{R}^{N \times D_{\mathcal{H}}}$, and $\Phi^* = \phi(\mathbf{X}^*) \in \mathbb{R}^{M \times D_{\mathcal{H}}}$. Therefore the prediction model is given by:

$$\hat{\mathbf{Y}}^* = \Phi^* \mathbf{W}. \quad (3)$$

As in the LR case, we can apply the Tikhonov regularized solution to find the weights:

$$\mathbf{W} = (\Phi^\top \Phi + \lambda \mathbf{I})^{-1} \Phi^\top \mathbf{Y}. \quad (4)$$

Note that this problem is not solvable as the inverse runs on matrix $\Phi \Phi^\top$, which is of size $D_{\mathcal{H}} \times D_{\mathcal{H}}$, and Φ is in principle unknown. However, if we use a λ that ensures that the matrix can be inverted, the solution is equivalent to:

$$\mathbf{W} = \Phi^\top (\Phi \Phi^\top + \lambda \mathbf{I})^{-1} \mathbf{Y}. \quad (5)$$

If we summarize the right part of the equation by $\alpha = (\Phi \Phi^\top + \lambda \mathbf{I})^{-1} \mathbf{Y}$, the predictions for the text input samples is:

$$\mathbf{Y}_* = \Phi_* \Phi^\top \alpha. \quad (6)$$

Note that even though the mapping Φ is unknown, one can replace this inner product matrix with a similarity matrix between samples, which is known as the *kernel matrix* \mathbf{K} . In this case we can replace the Gram matrix as $\mathbf{K} = \Phi \Phi^\top$, and equivalently $\mathbf{K}_* = \Phi_* \Phi^\top$. Therefore the predictions can be computed by:

$$\hat{\mathbf{Y}}_* = \mathbf{K}_* \alpha. \quad (7)$$

In KRR we only need a kernel function, $k(\mathbf{x}_i, \mathbf{x}_j)$, according to the Mercer's theorem [54]. Kernel methods literature is full of examples of proper kernel matrices. Here we will use the most standard one, the Gaussian Function (Radial Basis Function, RBF) kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma^2))$, which has only one free parameter, σ . Therefore, in LR one parameter is tuned: the regularization parameter λ . In KRR, two free parameters are tuned: the regularization parameter λ and the kernel parameter σ .

In both cases we used a cross-validation strategy for their optimization using one half of the data for training and the other half for validating. A MATLAB implementation of KRR and other regression algorithms can be found at http://isp.uv.es/soft_regression.html.

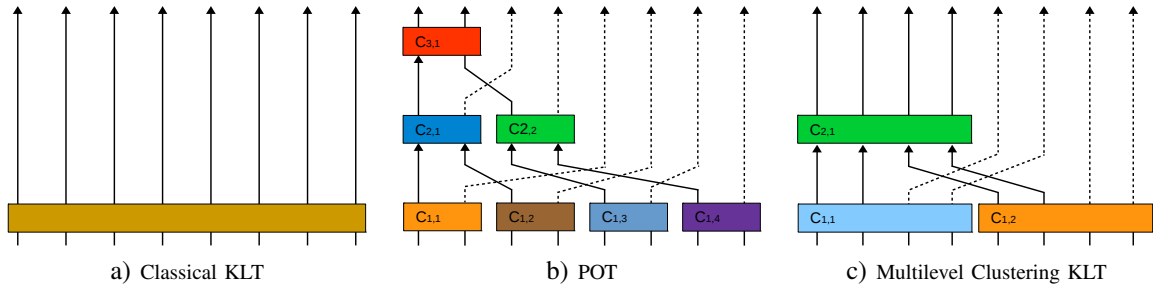


Fig. 2. Structure of a classical KLT, POT, and Multilevel Clustering KLT. This example decorrelates eight spectral components. Each arrow denotes a component and each coloured rectangle represents the computation of a KLT. In the case of POT, three levels of transform are applied, with a different number of clusters per level, and all clusters transforming two components. In the case of Multilevel Clustering KLT, two levels of multilevel clustering are applied and two clusters are defined in the first level.

III. DATA AND EXPERIMENTAL SETTING

This section introduces the data collection used in the experiments and the parameter configuration employed in the compression and in the statistically based regression stages. First, a description of the main characteristics of the IASI instrument is given, along with the specific products used in the experiments.

A. Data Collection

IASI is the main payload instrument carried on the MetOp satellite series [55]. Data provided by the instrument represent a significant improvement in the retrieval accuracy and vertical resolution of atmospheric parameters (temperature and water vapour concentration mostly) with respect to previous lower spectral resolution instruments, such as HIRS, SEVIRI, etc. This has in turn improved the output from numerical weather prediction models and atmospheric chemistry studies.

IASI produces rich spectral information to derive temperature and moisture profiles. The spectra is also used for the determination of trace gases such as CO, CO₂, CH₄, SO₂, N₂O, HNO₃, NH₃, OCS, and CF₄ [56], [57], [58], [59], [60], [61], land and sea surface temperatures, and cloud properties.

The instrument collects data with high spectral, spatial, and temporal resolution producing large volumes of information (about 16 Gigabytes per day generated by each of the IASI-A and IASI-B instruments). IASI covers the spectral range between 645 and 2760 cm^{-1} yielding 8461 spectral components. Such an amount of information is costly to manage, hence the need to search for efficient strategies to reduce the large size of the data for improved processing, transmission, and storage.

An effective strategy to alleviate the large volume of information produced by the IASI instrument is to compress the data according to the specific requirements of the end-user applications. As outputs of the models, we use the physical variables (temperature and dew point temperature) given by the analyses of the model of the European Centre for Medium-Range Weather Forecasts (ECMWF). The model provides estimations for 137 different pressure levels between $[10^{-2} \dots 10^3]$ hPa in the atmosphere and spatial resolution of 0.5 degrees. We co-registered the predictions supplied with the scenes acquired by IASI instrument with the analyses of the ECMWF.

The experiments have been conducted in a realistic context, where data from disjoint orbits are used for training and

testing. The training set is composed by seven orbits and the atmospheric parameter retrieval test is carried out on six different orbits. All data have been produced by the IASI-B instrument, implemented on the MetOp-B satellite. Table I reports the characteristics of the 13 IASI orbits used in the experiments. In order to isolate the results for cloud free and cloudy areas we use the fractional cloud cover data provided in the L2 level [62].

Usually, the Band 3 of the IASI spectrum and some channels from the Band 1 and from the Band 2 are not used for temperature and dew point temperature retrieval because they are influenced by solar radiation and trace gases such as CO, CH₄, etc. Following the feature selection in [39] we performed feature selection removing the most noisy bands and keeping 4699.

B. Setting and Parameter Configuration for Lossy Compression

Lossy compression is carried out through JPEG 2000 standard. It is meaningful to consider that, in hyperspectral data with such a large number of spectral components as IASI products, it is of utmost importance to exploit the high spectral redundancy inherent to the data. As mentioned, to achieve improved coding performance, JPEG 2000 is paired along with three spectral transforms: for DWT, 10 levels of 9/7 DWT are applied in the spectral dimension; POT is run using the default parameter settings; and for Multilevel Clustering KLT, 100 clusters are defined in the first level. To account for the spatial redundancy, a 9/7 DWT with five levels is applied in the spatial dimension. Nine target bit-rates are analyzed, distributed from 2 to 0.0025 bit per pixel per component (bpppc), corresponding to compression ratios from 8:1 to 6,400:1. Table II summarizes the main characteristics of the three compression settings studied in the experiments.

In the experiments, Kakadu software [63] has been used for JPEG 2000, Pairwise Orthogonal Transform software [64] has been run for POT, and Spectral Transform software [65] has been employed for Multilevel Clustering KLT. The JPEG 2000 standard provides a multicomponent extension in its Part 2 [66], which has been used to apply the DWT.

It is worth mentioning that, in our lossy compression approach, we apply a decoding stage prior to the statistically based retrieval, so that retrieval works with the reconstructed

TABLE I

IASI L1C PRODUCTS USED IN THE EXPERIMENTS. IDENTIFIERS AND SIZES ARE PROVIDED. M IS THE NUMBER OF SPECTRAL COMPONENTS, N_s IS THE NUMBER OF SCAN LINES, N-FORS IS THE NUMBER OF ELEMENTARY FIELDS OF REGARD (FOR) PER LINE, AND N-IFOVS IS THE NUMBER OF INSTANTANEOUS FIELDS OF VIEW (IFOV) PER FOR.

Orbit identifiers	Size (M × N _s × N-FORS × N-IFOVS)
Training set	
IASI_XXX_1C_M01_20131017012656Z_20131017030856Z_N_O_20131017021701Z	(8461 × 765 × 30 × 4)
IASI_XXX_1C_M01_20131017045352Z_20131017063552Z_N_O_20131017054413Z	(8461 × 765 × 30 × 4)
IASI_XXX_1C_M01_20131017081752Z_20131017095656Z_N_O_20131017090901Z	(8461 × 743 × 30 × 4)
IASI_XXX_1C_M01_20131017113856Z_20131017131752Z_N_O_20131017123024Z	(8461 × 742 × 30 × 4)
IASI_XXX_1C_M01_20131017145656Z_20131017163856Z_N_O_20131017155030Z	(8461 × 765 × 30 × 4)
IASI_XXX_1C_M01_20131017181752Z_20131017195952Z_N_O_20131017190946Z	(8461 × 765 × 30 × 4)
IASI_XXX_1C_M01_20131017214152Z_20131017232352Z_N_O_20131017223236Z	(8461 × 765 × 30 × 4)
Test set	
IASI_XXX_1C_M01_20131017030856Z_20131017045352Z_N_O_20131017035958Z	(8461 × 787 × 30 × 4)
IASI_XXX_1C_M01_20131017063552Z_20131017081752Z_N_O_20131017072637Z	(8461 × 765 × 30 × 4)
IASI_XXX_1C_M01_20131017095656Z_20131017113856Z_N_O_20131017104905Z	(8461 × 765 × 30 × 4)
IASI_XXX_1C_M01_20131017131752Z_20131017145656Z_N_O_20131017141023Z	(8461 × 743 × 30 × 4)
IASI_XXX_1C_M01_20131017163856Z_20131017181752Z_N_O_20131017173059Z	(8461 × 742 × 30 × 4)
IASI_XXX_1C_M01_20131017195952Z_20131017214152Z_N_O_20131017205049Z	(8461 × 765 × 30 × 4)

TABLE II

CONFIGURATION OF THE COMPRESSION SETTING USED IN THE EXPERIMENTS.

Spectral transform		Spatial transform	Compression technique	Target bit-rates (bpppc)	Compression ratios
DWT	10 levels	5 levels of DWT	JPEG 2000	2, 0.5, 0.25, 0.1, 0.05, 0.03, 0.01, 0.005, and 0.0025	8, 32, 64, 160, 320, 533, 1,600, 3,200, and 6,400
POT	Default parameter settings				
Multilevel Clustering KLT	100 clusters in the first level and multilevel mode				

data and not in the compressed domain. These reconstructed data have exactly the same spectral and spatial dimensions as the original data, i.e., no spectral or spatial dimensionality reduction occur. It is the quality of the observations that is being modified, not the size of the scenes.

C. Settings of Statistically Based Regression

The retrieval experiments have been carried out in a realistic scenario, i.e., data from disjoint IASI orbits have been used for training and testing. The statistical retrieval algorithms are applied on IASI data hyperpixels at different pressure levels considering all the spectral components at a particular spatial position in the regression.

In the training stage 8,000 samples are randomly selected from each of the training orbits (seven products), which produces a training set of 56,000 samples. To allow a fair comparison, the positions of the samples used for training are kept constant for all the experiments, i.e., all three multi-component transforms and all compression ratios. Six different orbits are used for testing. The prediction is carried out on the whole orbit. To assess the retrieval performance, the RMSE between the ground-truth values of temperature/dew point temperature and the predictions produced by the regression models is computed in each pressure level.

We trained one model for each coding setting and each compression ratio. Note that the input data to the retrieval process is the reconstructed spectra, i.e., compressed and decompressed data. The models were trained using a simple cross-validation scheme with half of the data used for training and the other half used for validating the parameters.

IV. EXPERIMENTAL RESULTS

This section presents the experimental results. We report results for atmospheric parameter retrieval over the reconstructed spectra and investigate the origin of the improvement in the retrievals produced using the reconstructed spectra. Results are evaluated in terms of noise removal and shared spatio/spectral information to assess the impact of compression on the retrievals. This study is conducted through several experiments:

- 1) We evaluate the performance of the retrievals produced using reconstructed spectra. Different compression settings and retrieval algorithms are used. The experiment is carried out in a realistic scenario where data used for testing comes from orbits different from those used for training.
- 2) We analyze the amount of noise left in the data after being processed with each compression setting and at different compression ratios.

- 3) We analyze the spectral and spatial information in the compression settings. First we quantify the amount of information shared among the spatial neighbours. We then propose an scheme which specifically enforces different levels of spatial and spectral relations in a simple way, and compare the performance of the retrieval algorithms when using data processed following this scheme and when using reconstructed data from the compression algorithms.

Results are reported for one IASI L1C orbit and one physical variable (dew point temperature) in most of the experiments due to restrictions in page length. Conclusions from the results for the remainder orbits and for temperature are similar. We report all these results in the supplementary material at http://isp.uv.es/spatio_spectral_compression.html.

A. Retrieval Assessment

This section reports the results of estimating atmospheric physical variables through LR and KRR using reconstructed IASI L1C spectra. Results are presented for the orbit IASI_XXX_1C_M01_20131017030856Z_2013-1017045352Z_N_O_20131017035958Z (see table I) and dew point temperature prediction. Results for all the other orbits and variables are very similar and reported as supplementary material.

A first assessment of the reconstructed radiances is reported in the Appendix at http://gici.uab.cat/pub/spatio_spectral_compression/appendix.html, where the normalized radiance residuals statistics and the spectral signature of the reconstructed radiances are analyzed. Next, Figure 3 shows the retrieval performance of dew point temperature prediction for the lossy compression settings proposed. Plots show the average of the RMSE prediction over the different pressure levels between 1100 and 100 hPa. We analyze four different scenarios, i.e., land and cloud free, land and cloudy, ocean and cloud free, and ocean and cloudy conditions. Cloud free conditions is equivalent to cloud fraction equals to 0%. Otherwise, it is considered cloudy conditions. Ocean is equivalent to land fraction equals to 0%. Otherwise, it is considered land. Table III reports the number of spectra analyzed in each scenario. The compression settings POT + JPEG 2000 and Multilevel Clustering KLT + JPEG 2000 achieve lower maximum compression ratios compared to DWT + JPEG 2000 because side-information needs to be transmitted in addition to the compressed data.

TABLE III
NUMBER OF SPECTRA IN EACH OF THE SCENARIOS ANALYZED. THE PERCENTAGE OF THE ORBIT IN EACH SCENARIO IS REPORTED IN BRACKETS.

Land - Cloud free	Land - Cloudy	Ocean - Cloud free	Ocean - Cloudy
10,218 (10.8%)	28,613 (30.3%)	6,658 (7.1%)	48,951 (51.8%)

One can see that while low compression ratios do not reduce the accuracy in the retrievals, the prediction results are improved as the compression ratio increases. Nonetheless, when

the compression ratio is very high, the retrieval performance begins to deteriorate because too much distortion is introduced in the reconstructed data. This behaviour occurs for both LR and KRR, although the retrieval improvements are larger for LR (in the sense that higher compression ratios can be applied before the retrieval performance starts to deteriorate). These results are consistent with the observations reported in [32]. However, note that here the situation is more realistic than in [32], since we use different orbits for training, validation, and testing. By doing so we ensure that no information about the true values due to the blurring introduced by the compression settings can be used in the predictions. Results are also consistent for the new compression settings evaluated.

An interesting analysis was introduced in [27] about the correlation between the dimensions of the predictions, where an index (i_D) of non-diagonality of the covariance matrix of the predictions for different pressure levels, $\Sigma_{\hat{Y}}$, was proposed:

$$i_D = \frac{B_{\hat{Y}}^d}{S_{\hat{Y}}} \quad (8)$$

where $S_{\hat{Y}} = \|\Sigma_{\hat{Y}}\|$, $\Sigma_{\hat{Y}}$ is the covariance matrix of the predictions, and $B_{\hat{Y}}^d$ is the biggest value in the diagonal of $\Sigma_{\hat{Y}}$ (see [27] for details). Figure 4 shows the value of the i_D index for the different compression configurations and the different regression algorithms taken into consideration in this paper. We provide the index without the normalization by the number of output dimensions. We can see that given a particular regression method, the differences in the correlation between the outputs for different compression ratios (and also for different compression settings) are negligible for the compression ranges with low prediction errors, i.e., until compression ratios about 300:1.

Several conclusions can be drawn:

- 1) *Statistical atmospheric parameter retrieval benefits from spectral/spatial lossy compression.* It is clear from Fig. 3 that using reconstructed data after a compression process allows to achieve improved statistically based retrieval performance. While moderate and medium compression ratios enable to achieve at least the same performance than the original spectra, when the distortion level introduced in the data is high, the accuracy in the retrievals decreases. In all the scenarios analyzed, a compression ratio of, approximately, 100:1 allows to achieve the same or improved retrieval results compared to retrieval results obtained from uncompressed spectra (original data).
- 2) *KRR produces more accurate predictions than LR.* For all plots, KRR consistently yields better retrieval performance than LR. These results are constant for all the data (original and reconstructed spectra) and scenarios analyzed. The difference is specially significant for cloudy conditions.
- 3) *Lossy compression enables more significant retrieval improvements for LR compared to KRR.* The retrieval improvement is more evident for LR than for KRR when a compression process has taken place, which suggests that the spectral/spatial regularization produced by the compression stage particularly benefits LR.

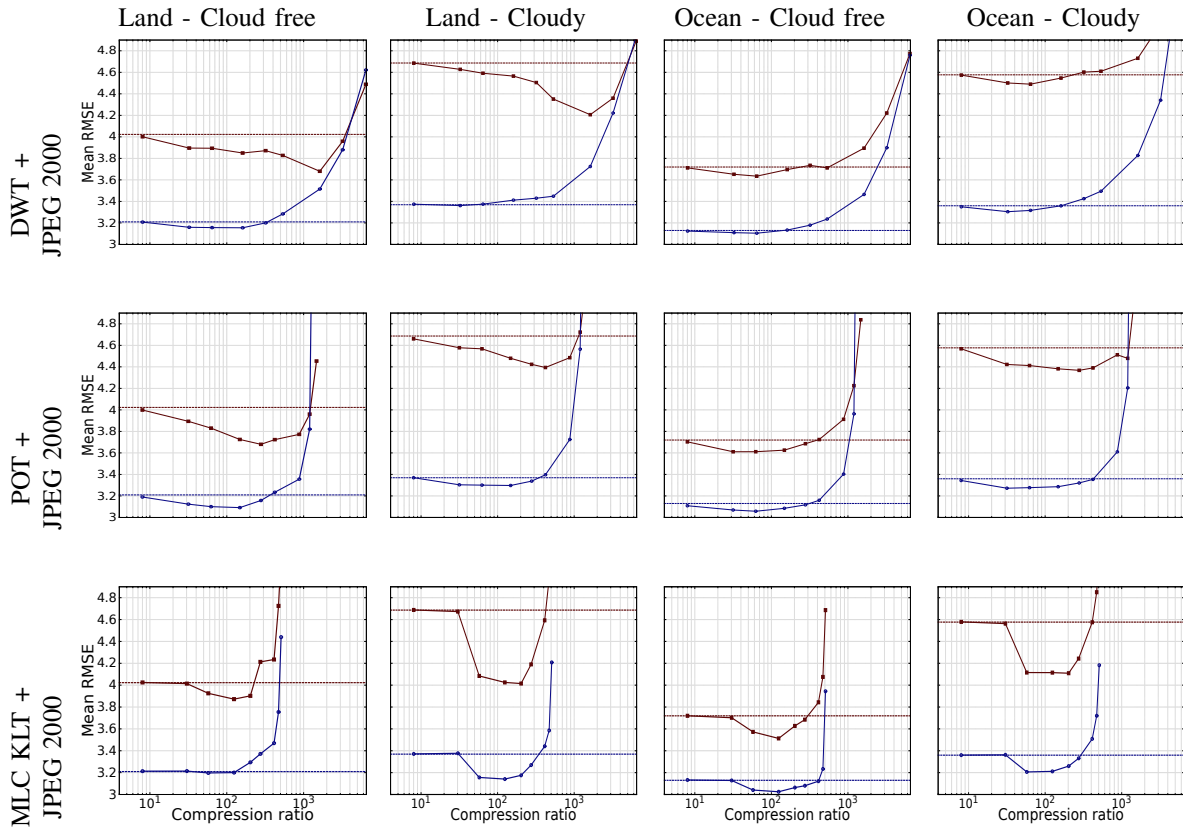


Fig. 3. Dew point temperature (in kelvin) retrieval performance for different lossy compression settings using LR (solid red lines) and KRR (solid blue lines). In all the plots, the vertical axis represents the averaged RMSE over the different pressure levels between 1100 and 100 hPa and the horizontal axis represents the compression ratio. Ranges are the same in all the plots to ease the comparison. Each row shows the results for a particular compression setting and each column shows the results for a particular scenario. Results using uncompressed spectra (original data) for LR (dashed red lines) and KRR (dashed blue lines) are plotted for comparison purposes.

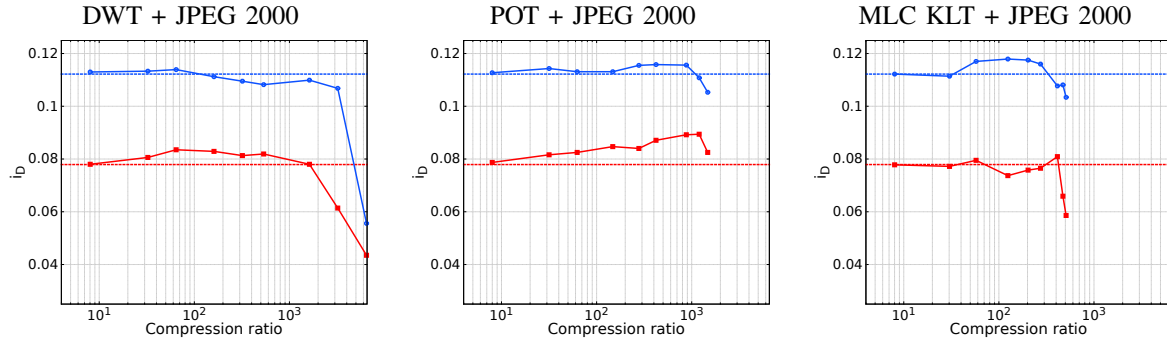


Fig. 4. Index i_D for different lossy compression settings using LR (solid red lines) and KRR (solid blue lines). In all the plots, the vertical axis represents the i_D index and the horizontal axis represents the compression ratio. Ranges are the same in all the plots to ease the comparison. Results using uncompressed spectra (original data) for LR (dashed red lines) and KRR (dashed blue lines) are plotted for comparison purposes.

- 4) *LR achieves competitive performance when the spectra have been compressed at higher compression ratios compared to KRR.* Fig. 3 illustrates that while KRR is able to achieve improved retrieval results when data compressed at a compression ratio of, approximately, 124:1 are used, LR still produces competitive performance when spectra compressed at a compression ratio of, approximately, 203:1 is employed.
- 5) *Multilevel Clustering KLT + JPEG 2000 leads to the best prediction results.* The compression setting Multi-

level Clustering KLT + JPEG 2000 leads to the best retrieval improvements for land and cloudy, ocean and cloud free, and ocean and cloudy conditions. For land and cloud free conditions, both DWT + JPEG 2000 and POT + JPEG 2000 slightly overcome Multilevel Clustering KLT + JPEG 2000. If the interest is to optimize the retrieval performance in all the scenarios, Multilevel Clustering KLT + JPEG 2000 is the best compression setting.

- 6) *POT + JPEG 2000 allows to achieve higher com-*

pression ratios leading at least to the same retrieval performance than the original spectra. If the interest is to achieve very high compression ratios, POT + JPEG 2000 would be the best compression setting. A compression ratio of 421:1 allows to achieve at least the same prediction results compared to the original data for all the scenarios analyzed. Higher compression ratios decrease the retrieval performance because useful information is missing in the recovered spectra.

- 7) *The effect of lossy compression is alike for all the compression settings and scenarios analyzed.* One can see in Fig. 3 that, depending on the compression setting applied and the scenario analyzed, the results can slightly vary, but the impact of lossy compression on retrieval performance is similar for all cases.
- 8) *The effect of the compression on the predictions correlation is negligible for useful compression ratios.* If we take in consideration the range of compression ratios for low prediction error (i.e. [0-300], see Fig. 3), the difference in correlation between the predicted outputs when using and when not using compression is very low (Fig. 4). The same effect can be observed if we compare this difference with the difference inferred by using different prediction methods.

Figure 5 shows the bias and the RMSE results for the range of pressure levels corresponding to the troposphere (between 1100 and 100 hPa). For each compression setting and each scenario, results for the compression ratio that produces the best prediction performance are reported.

Reconstructed spectra produce better (or at least the same) results compared to the original data for all pressure levels and all compression settings. Generally, improvements are more evident in the mid-low troposphere. As illustrated in Fig. 3, the improved prediction is more significant for LR. Results are usually better over ocean areas than over land, and better on cloud free situations rather than in cloudy scenarios.

The Mean Absolute Error (MAE) between the dew point temperature values provided by the ECMWF analyses (and taken here as the ground truth), and the predicted values produced by LR and KRR when uncompressed and reconstructed spectra are employed in the retrievals is illustrated in Fig. 6. Maps plot the MAE for all the analyzed orbits. Results are reported only for the original data and Multilevel Clustering KLT + JPEG 2000, which is the compression technique that produces the most competitive performance, due to restrictions in page length. As expected, KRR produces smaller errors than LR (more bluish areas), and maps show that the MAE is very similar when the original data and the reconstructed spectra are used. Errors occur in the same geographic areas and the error pattern is similar for both the original and the reconstructed data.

B. Noise Reduction (denoising) through Compression

This section is specifically dedicated to analyze the effect of the compression process on the noise present in the data. Lossy compression is known to be an indirect way of performing signal filtration and denoising. Compression aims for

a compact representation of the data by eliminating redundant information, preserving as much as possible the characteristics of the observation needed by the intended data user.

Figure 7 reports an estimate of the noise level present in the reconstructed spectra for the compression settings analyzed. In the experiments it is assumed that the noise model in the IASI data comes from an additive white Gaussian distribution, and also independent and identically distributed. This is a realistic assumption for an interferometer and is the model adopted in [46], [67], [68], [69]. To compute the noise standard deviation, the Anisotropic Nonparametric Image Restoration toolbox [70] was used. This method provides a robust estimate of the standard deviation based on the well-known median of the absolute deviation (MAD) [71]. This technique is used along with an orthogonal wavelet transform in such a way that the median estimator is applied on the high frequency subbands (fine details) of the transform domain to reduce the impact of the features present in the signal. In the experiments, a Daubechies wavelet transform was used. Specific and more detailed description of the adopted strategy can be found in [72], [73].

We can see how part of the improvement in the retrievals is due to the filtering of noise performed by the compression process. It is interesting to compare Fig. 7 with Fig. 3. It is clear from Fig. 7 that, as the compression ratio increases, the noise level in the reconstructed data decreases. While most of the noise is removed at moderate compression ratios, the noise level keeps unchanged for extremely high compression ratios because almost all the noise has already been removed from the data. When data is compressed in moderate to high compression ratios (i.e., [10:1-300:1]), most of the content removed by the compression process is noise, keeping the retrieval performance competitive or even getting better results than using uncompressed data. However, if the compression ratio is extremely high (i.e., [300:1-2,000:1]), useful information is also eliminated, degrading the accuracy of the retrievals. We can suppose that noise filtering is one of the reasons why reconstructed compressed spectra from high compression ratios can result in better performance compared to uncompressed data. It is also clear that increasing the compression ratio beyond a certain point has no effect on the elimination of noise and therefore the only content removed from now on is relevant information. Therefore, the retrievals performance decreases beyond this point. Note that the compression settings POT + JPEG 2000 and Multilevel Clustering KLT + JPEG 2000 achieve lower maximum compression ratios compared to DWT + JPEG 2000 because side-information needs to be transmitted in addition to the compressed data.

C. Exploiting Spectral/Spatial Relations through Compression

It is acknowledged that parameter retrieval and model inversion applications largely improve when exploiting spatial information [74], [75], [76]. Compression is an indirect way to stress relevant feature relations and to enforce smoothness in the reconstructed data, which can be exploited by pixelwise retrieval methods.

In this section we analyze the impact that the analyzed compression approaches have on sharing spectral and spatial

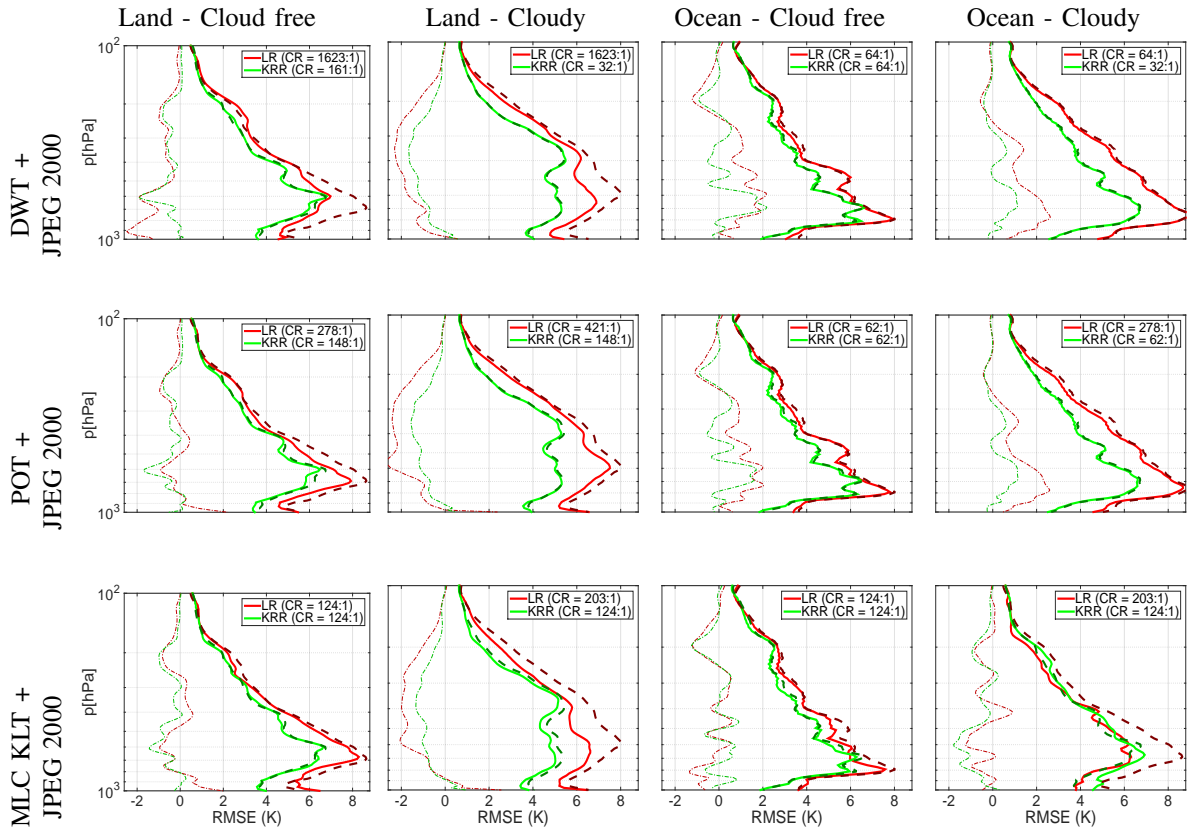


Fig. 5. Dew point temperature (in kelvin) RMSE profiles and bias. The compression ratio (CR) with the best average RMSE is reported for LR (thick, solid and red lines) and KRR (thick, solid and green lines). Results using the original data are shown as well for comparison purposes when LR (thick, dashed and red lines) and KRR (thick, dashed and green lines) are used for the predictions. The bias of the reconstructed data are plotted with thin and dash-dot lines for LR (red) and KRR (green).

information about the neighboring pixels and the effect it has in the retrieval performance. We first assess how much information is shared between spatial neighbours. Then we propose a simple method which specifically takes into account information from the spatial and spectral neighbours, and we compare these results with the results obtained when we do the retrieval over the recovered coefficients after compression. We also investigate how the benefits of compression (sec. IV-A) are affected when we have already exploited specifically the spatial/spectral information in the data.

1) *Assessing the amount of Spectral/Spatial Information shared through Compression:* Here we use information theory [77] measures in order to quantify the amount of information that spatial neighbours share when use different compression approaches and different compression ratios. In particular we compute the mutual information between one pixel and its neighbours. The same procedure has been used to measure the amount of information shared between spatial, spectral and orientation neighbours in the wavelet domain [78] and in the normalized domain [79].

In Fig. 8 we show the results when the spatial neighbours are taken into account. We computed the mutual information between each pixel and its neighbours in a neighbouring of 11×11 . For the experiment we used the data contained in the whole orbit, i.e., for the computation of the mutual information

we used 94,440 samples for each component and compute the average among all the components. Particular values are different for the three compression settings analyzed, however, the trend is the same for all of them. For low to high compression ratios (i.e. [30:1-500:1]) each coefficient contains more information about its neighbours than the original scenes. When the compression ratio is extremely high (i.e., higher than 500:1), the shared information decreases dramatically. This behavior is also easy to see in the detailed squares below the graphs. In these squares we show the spatial pattern of the mutual information. It is easy to see that closer coefficients are more related (as expected), and that only the global intensity of the relations changes (i.e., the pattern of the relations is similar for different compression ratios).

2) *A Simple Way to Exploit Spectral/Spatial Relations:* In the following we investigate the impact that specifically exploiting spectral and spatial information has in the retrievals. Elaborated models that take into account the dependence of the the spatial smoothness with the spectral channel [80], or avoid the neighbour pixels corrupted by clouds could be used. However, for simplicity we present results for the less complex way of computing the average between neighboring pixels. We choose here to simply convolve the data with two Gaussian filters, one over the spectral domain and one with spatial neighboring pixels. We define the σ_{se} and the σ_{sa}

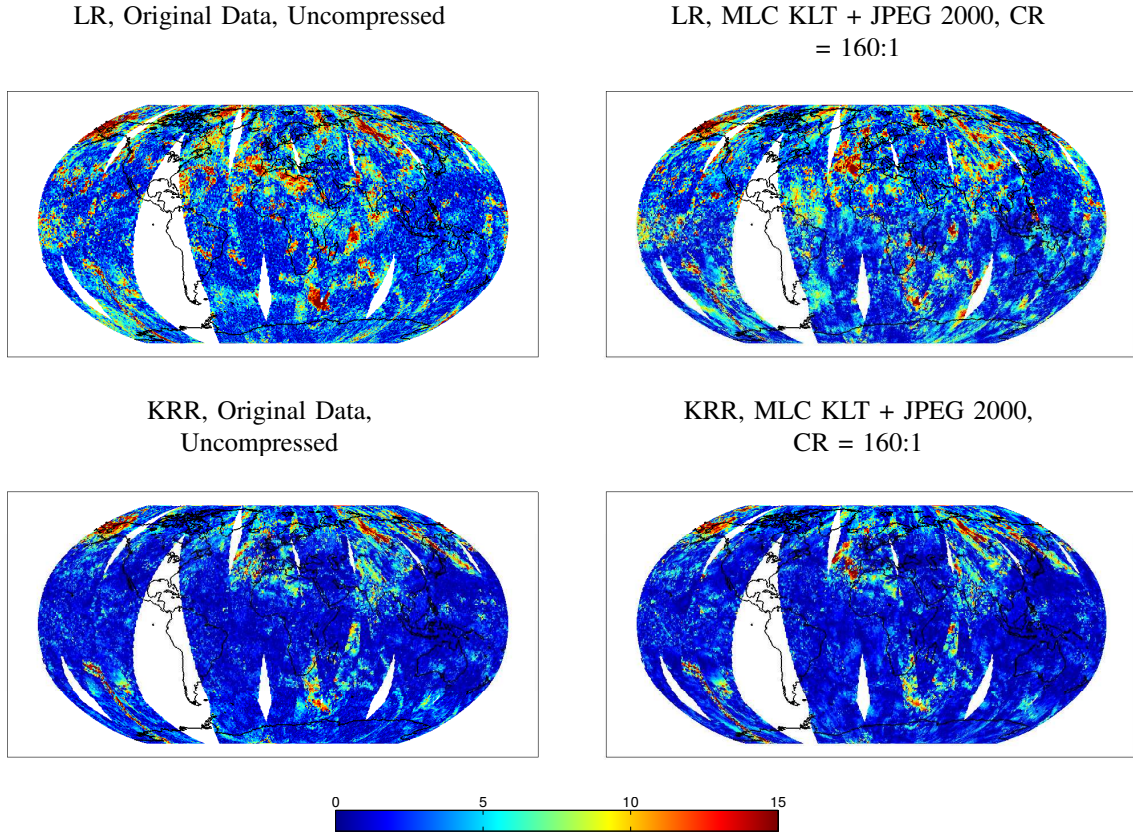


Fig. 6. MAE for dew point temperature over the whole profile. Plots report MAE for uncompressed and reconstructed spectra and LR and KRR algorithms. Larger errors are represented by red color and smaller errors by blue color. CR reports the compression ratio achieved in the compression stage.

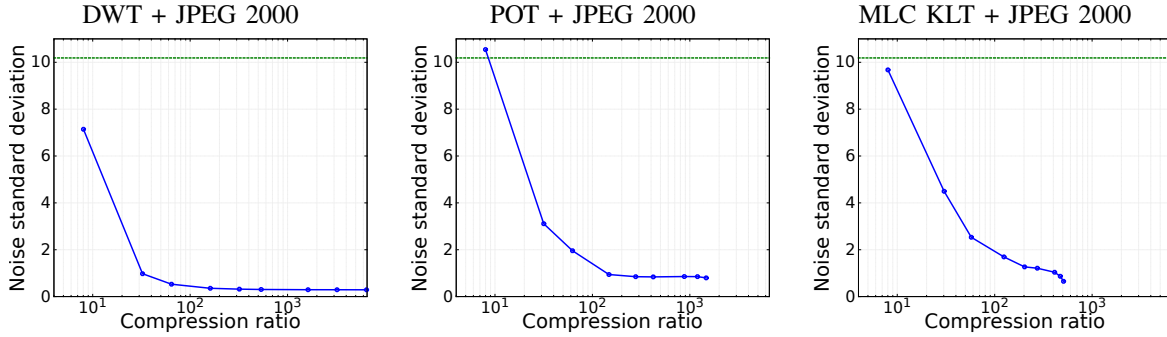


Fig. 7. For each compression setting, the Noise Standard Deviation is plotted (solid blue lines). In all the plots, the vertical axis represents the noise level (noise standard deviation of IASI raw data) in the reconstructed spectra and the horizontal axis represents the compression ratio. Noise estimation for uncompressed data (dashed green lines) is reported as well for comparison purposes.

parameters for the standard deviation of the Gaussian filter in the spectral and the spatial dimension, respectively. σ_{se} and σ_{sa} are defined in the image domain. In particular σ_{se} refers to distance in the spectral dimension, $\sigma_{se} = 1$ is equal to one channel distance, which corresponds to $\sim 0.5\text{cm}^{-1}$. σ_{sa} is defined regarding the distances between pixels, for instance $\sigma_{sa} = 1$ is equal to the distance between one pixel and the neighbouring one, which corresponds to $\sim 0.5^\circ$. Combinations of five different σ_{se} and seven different σ_{sa} are evaluated. The minimum values, $\sigma_{se} = 0.01$ and $\sigma_{sa} = 0.01$, correspond to a identity function, i.e., no spectral/spatial features are taken into account.

Note that the IASI data we are using is already apodised.

Apodization introduces spatial correlation, σ_{ap} . So the total spatial correlation would be $\sigma_{tot} = \sigma_{ap} + \sigma_{sa}$. Or equivalently we cannot start our experiment from zero correlation since the apodization process has already introduced some spatial correlation.

Table IV reports the retrieval performance for dew point temperature for different combinations of the parameters values. While exploiting both spectral and spatial features benefits the retrieval performance for LR, exploiting spectral features does not improve the results for KRR. In the case of LR, the performance is slightly improved when σ_{se} is large and σ_{sa} is small. However, for KRR the retrieval results are generally degraded when σ_{se} is increased.

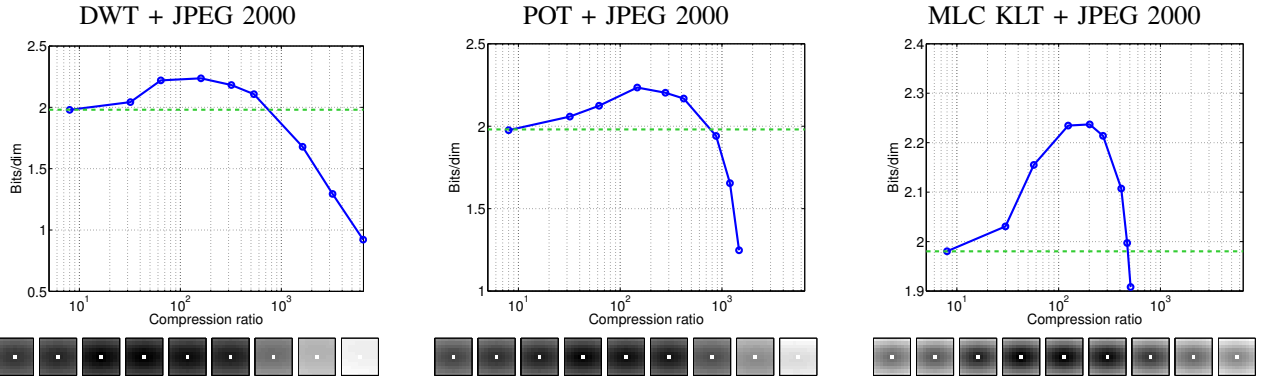


Fig. 8. Spatial information shared for each compression approach and different compression ratios. Blue solid curve shows the average amount of mutual information (in number of bits) shared between one coefficient and its spatial neighbours. Green dashed line represents the same quantity for the uncompressed image. Below each plot, the detailed shared information between the central pixel and each neighbour is reported. In each square, the amount of black corresponds to the mutual information between this spatial position and the central pixel. Each square corresponds to a different compression ratio in the plot above, and are sorted left to right as the dots in the blue solid curve in the plot.

TABLE IV

RMSE OF DEW POINT TEMPERATURE PREDICTIONS (IN KELVIN) FOR DIFFERENT COMBINATIONS OF σ_{sa} AND σ_{se} . THE IMPROVEMENT OVER THE ORIGINAL DATA IS REPORTED IN PERCENTAGE IN BRACKETS. BEST (GREEN) AND WORST (RED) PERFORMANCE ARE REPORTED.

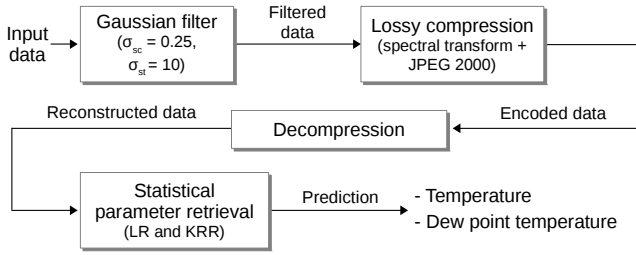
LR							KRR						
		$\sigma_{se}(spectral)$							$\sigma_{se}(spectral)$				
		0.01	0.25	1	5	10			0.01	0.25	1	5	10
$\sigma_{sa}(spatial)$	0.01	4.515	4.515 (0%)	4.540 (-0.6%)	4.370 (3.2%)	4.362 (3.4%)	$\sigma_{sa}(spatial)$	0.01	3.345	3.345 (0%)	3.376 (-0.9%)	3.389 (-1.3%)	3.379 (-1.0%)
	0.25	4.343 (3.8%)	4.343 (3.8%)	4.400 (2.6%)	4.254 (5.8%)	4.233 (6.2%)		0.25	3.248 (2.9%)	3.248 (2.9%)	3.276 (2.1%)	3.305 (1.2%)	3.266 (2.4%)
	1	4.061 (10.1%)	4.061 (10.1%)	4.097 (9.3%)	4.036 (10.6%)	3.996 (11.5%)		1	3.082 (7.9%)	3.082 (7.9%)	3.109 (7.1%)	3.113 (6.9%)	3.109 (7.1%)
	5	3.834 (15.1%)	3.835 (15.1%)	3.880 (14.1%)	3.809 (15.6%)	3.771 (16.5%)		5	2.994 (10.5%)	2.995 (10.5%)	3.017 (9.8%)	3.028 (9.5%)	3.021 (9.7%)
	10	3.742 (17.1%)	3.740 (17.2%)	3.795 (15.9%)	3.724 (17.5%)	3.711 (17.8%)		10	3.007 (10.1%)	3.007 (10.1%)	3.025 (9.6%)	3.035 (9.3%)	3.031 (9.4%)
	20	3.684 (18.4%)	3.683 (18.4%)	3.737 (17.2%)	3.697 (18.1%)	3.681 (18.4%)		20	3.069 (8.3%)	3.070 (8.2%)	3.108 (7.1%)	3.114 (6.9%)	3.095 (7.5%)
	30	3.669 (18.7%)	3.668 (18.8%)	3.736 (17.3%)	3.711 (17.8%)	3.694 (18.2%)		30	3.138 (6.2%)	3.138 (6.2%)	3.139 (6.2%)	3.160 (5.5%)	3.152 (5.8%)

It is clear that spatial features are more important than spectral features, as shown in [76] for hyperspectral data. It is interesting to note that, while the best prediction results for LR are yielded for large σ_{sa} , the best choice for KRR is significantly smaller σ_{sa} . Improvements for LR are more significant than for KRR. Note that even for this extremely simple averaging procedure, results confirm the suggestion reported in [32]: exploitation of spectral and spatial regularization (smoothing) improves the retrieval results.

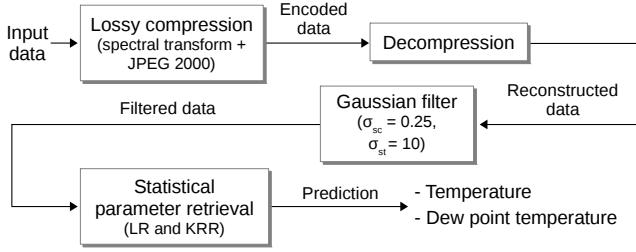
3) *Combining Compression Settings with the Specific Exploitation of Spectral/Spatial Relations:* In this section we show how the benefits due to compression are drastically reduced when we have already taken into account spatial/spectral information when pre-processing the data. We will show that compression is an effective, consistent, and cheap way to exploit important spectral and spatial data relations and removing noise in the reconstructed spectra. However, it is done in an indirect way and the benefits are limited.

In order to analyze the performance of compression paired along with the exploitation of spectral/spatial feature relations, we analyze several configurations. On the one hand we present results for the already explored strategies: retrieval over the original data, retrieval over the compressed and recovered data, and retrievals over the data processed using the method presented in the previous section. On the other hand we follow two extra strategies: 1) first, the Gaussian filter is applied to the original data and then, the filtered data are compressed; 2) first, the original data are compressed and then, the reconstructed spectra are filtered. Figure 9 illustrates the sequential chain of the last two strategies.

Figure 10 and Fig. 11 illustrate the retrieval performance of dew point temperature for LR and KRR, respectively. Plots show the average of the RMSE prediction over the different pressure levels. A configuration that produces competitive performance for both LR and KRR is used in the filtering stage, i.e., $\sigma_{se} = 0.25$ and $\sigma_{sa} = 10$. Results are reported for



a) Spectral/spatial regularization + compression. First, we filter the data to take into account the spatial and spectral relations. Then, the filtered data are compressed and decompressed, finally statistical parameter retrieval is carried out over the reconstructed spectra.



b) Compression + spectral/spatial regularization. First, a compression and decompression stage is performed. Then, feature relations are taken into account in the reconstructed spectra using a Gaussian filter, finally statistical parameter retrieval is carried out over the filtered data.

Fig. 9. Adopted sequential chain when a spectral/spatial regularization stage and a coding process are carried out.

four different scenarios: land and cloud free, land and cloudy, ocean and cloud free, and ocean and cloudy. The compression settings POT + JPEG 2000 and Multilevel Clustering KLT + JPEG 2000 achieve lower maximum compression ratios compared to DWT + JPEG 2000 because side-information needs to be transmitted in addition to the compressed data. Several conclusions can be extracted from this experiment:

- 1) *Pre-process the data to take into account the spatial and the spectral information about the neighbours compensates the effect of using recovered data from compression.* While using recovered data (maroon curve) improves the results obtained for the original data (red curve), this improvement vanishes when spectral/spatial information is exploited, i.e., using a compression method over data that have been already pre-processed to take into account spectral/spatial information (blue line) does not improve the performance (orange line). This observation indicates that when spectral/spatial regularization is carried out before compression, the spectral/spatial transform applied in the compression stage is not able to exploit extra features.
- 2) *Exploiting feature relations achieves improved retrieval performance.* As we have already seen in section IV-C2, exploiting extra features about the neighbouring pixels in the original spectra clearly benefits the prediction results. The consistency and homogeneity of the data is improved by pre-processing the data to take into ac-

count local relations between neighboring pixels, which benefits the regression algorithms. The improvement is more significant for cloudy conditions than for cloud free conditions.

- 3) *LR benefits more significantly from spectral/spatial regularization than KRR.* According to the results reported in the supplementary material at http://isp.uv.es/spatio_spectral_compression.html, it is clear from the individual plots that, when spectral/spatial regularization is added, larger improvements in the retrieval performance are achieved for LR compared to KRR. Gains produced for LR are, approximately, twice the gains produced for KRR.
- 4) *Filtering + compression yields better results than compression + filtering.* When the exploitation of features relations is paired along with compression, it is more efficient to perform first the spectral/spatial regularization and then compress the filtered data. When the compression stage is first carried out and the reconstructed spectra are filtered, the prediction performance is degraded as the compression ratio increases. The loss of performance is clear even at low compression ratios.
- 5) *Results are consistent for all the approaches and scenarios analyzed.* Conclusions are similar for all the plots, which suggests that the observed behaviour is consistent for all the approaches, compression settings, and scenarios analyzed.

Figure 12 illustrates the bias and the RMSE results for the range of pressure levels between 1100 and 100 hPa. The figure illustrates that the benefits of compression are compensated by pre-processing the data to take into account neighbouring spatial and spectral features. This supports the idea that one important advantage of using recovered compressed data for retrieval is that the process exploits information about the neighbours in the data.

Both approaches achieve the similar performance for all pressure levels and scenarios, which indicates that the benefits obtained from compression were mostly due to the indirect exploitation of spatial and spectral neighbouring pixels information.

V. DISCUSSION

Lossy compression techniques aim at reducing the size of the transmitted/stored information keeping the most informative part of the signal. As a consequence, the data information becomes more compact, which helps the subsequent information extraction steps. The multi-component transforms employed in the compression stage in this paper are based either in the Karhunen-Loeve transform (KLT) or the wavelet transform. Both transformations have the effect of translating the data to a domain where the statistical relations (or correlations) are reduced [35], [81], [82]. When compressing in these domains, the less relevant features are reduced, thus stressing the important ones. When transforming back to the original domain, the reconstructed data is a version of the original data where the less relevant features have also been reduced. From the signal processing point of view, it can be seen as a filtering process.

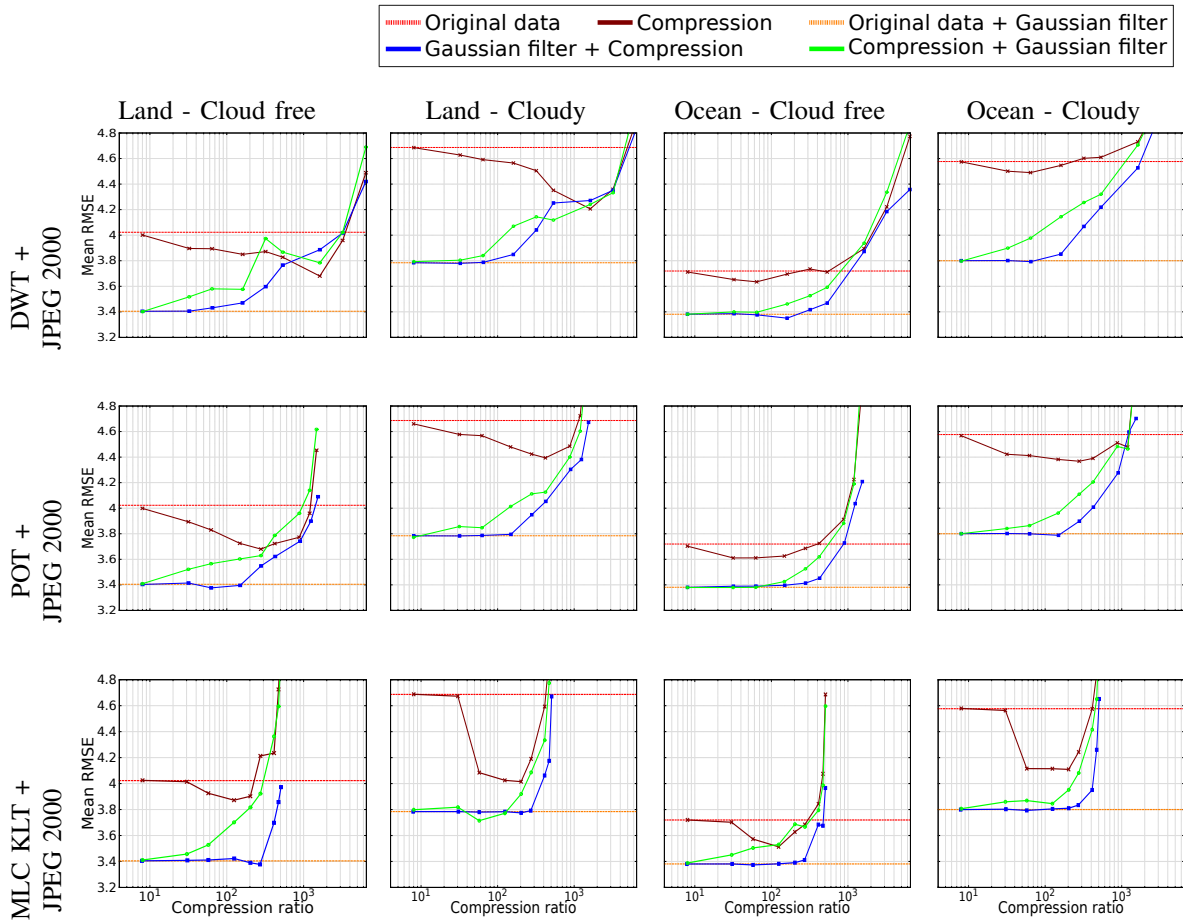


Fig. 10. Dew point temperature (in kelvin) retrieval performance for different lossy compression settings using LR. In all the plots, the vertical axis represents the averaged RMSE over the different pressure levels and the horizontal axis represents the compression ratio. Ranges are the same in all the plots to ease the comparison. Each row shows the results for a particular compression setting and each column shows the results for a particular scenario. Each plot compares five different approaches, i.e., original data, compression of the original data, original data + Gaussian filter, Gaussian filter + compression, and compression + Gaussian filter.

As an example, when using KLT on images, the transformation becomes very similar to the Fourier transformation [83], i.e., the signal is decomposed in different frequencies. The high frequencies of the images have low magnitude in this domain, which in general translates into a poor signal to noise ratio –this serves to explain why most coding techniques aim at reducing the energy of these frequencies–. As a consequence, the reconstructed image has less energy in the high frequencies, which stresses the contribution of the low frequencies.

We investigated how much each of these aspects affects the performance of the statistically-based retrieval methods for sounder data. In particular, we analyzed the retrieval of physical variables (temperature and dew point temperature) from IASI L1C data.

We first verified that applying a compression process to remote sensing scenes can improve the performance of the statistically based retrieval methods. We analyzed different compression settings combined with different retrieval methods in a realistic scenario. This was illustrated in Figure 3 and Figure 5. While low compression ratios kept the results almost unchanged, moderate and high compression ratios generally enabled improved retrievals. However, when the compression

ratios were very high, the retrieval performance was notably decreased, as expected.

Then we analyzed the first effect: how certain compression ratios help to remove useless information while keeping the relevant one. We analyzed the compression approaches as if they were denoising methods. When a signal is lossy compressed, the reconstructed (decompressed) signal has lost information. When this removed information is mostly noise and only a small amount of useful information is removed, compression has a positive impact on the regression results. We showed that for certain compression ratios and compression settings, most of the removed information was noise. Figure 7 reported the noise level present in the reconstructed spectra for different compression ratios and different compression settings. Plots illustrated that low compression ratios of approximately 8:1 removed small amounts of noise, specially in the case of POT + JPEG 2000 and Multilevel Clustering KLT + JPEG 2000. Moderate and high compression ratios were able to filter most of the noise present in the data. However, when extremely high compression ratios (i.e., above 500:1) were used in the compression stage, useful information was also removed and the quality of the reconstructed data was

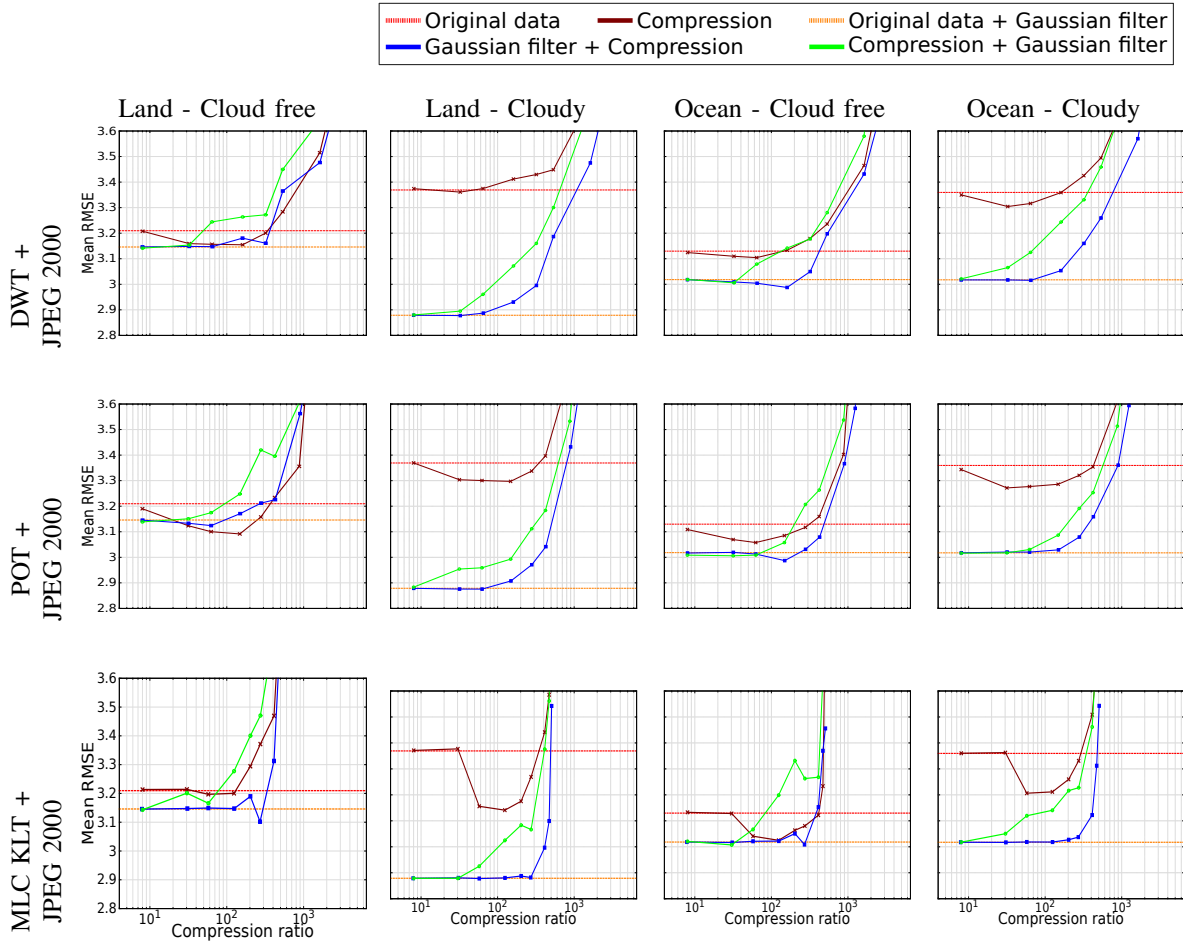


Fig. 11. Dew point temperature (in kelvin) retrieval performance for different lossy compression settings using KRR. In all the plots, the vertical axis represents the averaged RMSE over the different pressure levels and the horizontal axis represents the compression ratio. Ranges are the same in all the plots to ease the comparison. Each row shows the results for a particular compression setting and each column shows the results for a particular scenario. Each plot compares five different approaches, i.e., original data, compression of the original data, original data + Gaussian filter, Gaussian filter + compression, and compression + Gaussian filter.

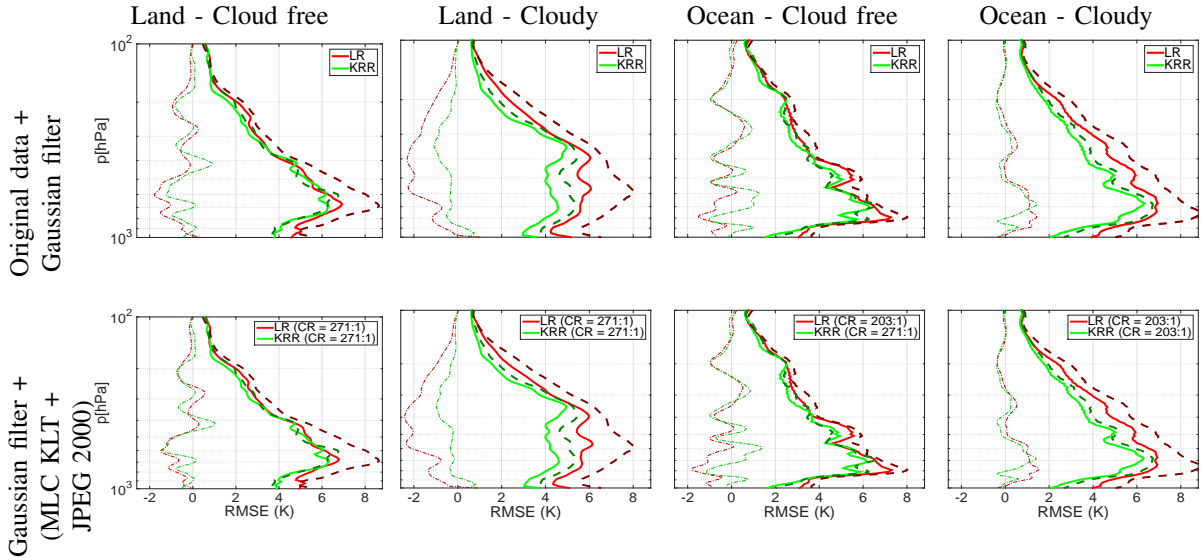


Fig. 12. Dew point temperature (in kelvin) RMSE profiles and bias. Two approaches are compared: original data + Gaussian filter, and Gaussian filter + Multilevel Clustering KLT + JPEG 2000. LR (thick, solid and red lines) and KRR (thick, solid and green lines). The bias of the each approach is plotted with thin and dash-dot lines for LR (red) and KRR (green). For the compression setting, only the compression ratio (CR) with the best average RMSE is reported. Results using the original data are also displayed for comparison purposes, LR (thick, dashed and red lines) and KRR (thick, dashed and green lines).

degraded. This explains why reconstructed spectra produce accurate retrieval performance when the original data have been compressed at moderate and high compression ratios (between 32:1 and 420:1, approximately, depending on the compression setting). When extremely high compression ratios were used, the removed signal was informative and therefore retrieval performance decreased.

The last part of the paper was focused on the second important effect of the lossy compression techniques. We posit that the compaction of the information introduced by the compression has the effect of exploiting spatial and spectral information about the neighbours in the reconstructed signal.

First, we specifically computed the amount of information about the spatial neighbours is introduced when compressing the data. After that, we showed that exploiting certain spectral/spatial regularization on hyperspectral data had a positive effect in the retrievals (Table IV). In particular, the spatial relations were more important than the spectral relations. Then we evaluated the impact of compression in the retrievals when the data was already pre-processed to take into account spectral/spatial feature relations. Results in Fig. 10 and Fig. 11 showed that when the data was pre-processed the effect of compression in retrieval performance are compensated. In other words, an important advantage of using recovered compressed data for retrieval is that the process exploits information about the data neighbours.

VI. CONCLUSIONS

This paper studied the impact of lossy compression on statistically based regression algorithms for retrieving atmospheric profiles of temperature and dew point temperature using infrared sounder data (IASI L1C). We analyzed the reasons behind the benefits produced by compression and provided recommendations for effective prediction performance and data compression.

Several compression settings were evaluated and a linear and a nonlinear regression algorithm were assessed in a realistic training/testing scheme. Results for the prediction of two physical variables (temperature and dew point temperature) and four different scenarios (land and cloud free, land and cloudy, ocean and cloud free, and ocean and cloudy) were reported.

Lossy compression was carried out through different compression settings, always within the scope of international standard JPEG 2000. To achieve competitive compression performance in hyperspectral data with a large number of spectral components (such as IASI data), three spectral transforms were paired along with JPEG 2000, i.e., DWT, POT, and Multilevel Clustering KLT. In the experiments, nine compression ratios were analyzed to search for an optimal trade-off between retrieval performance and data compression.

Experimental results revealed that reconstructed spectra have enough quality to achieve competitive retrieval performance because noise filtering is carried out in the compression stage, which allows to achieve high compression ratios while retaining as much features as possible. Another positive effect arises from the ability of compression to exploit spectral and

spatial feature relations in an indirect way, which benefits the retrieval methods (on average and using ECMWF analysis as ground truth). We observed that moderate-to-high compression ratios produced improved results in predicting temperature and dew point temperature in the different scenarios analyzed. As expected, when the compression ratio is extremely high, the benefits disappear because large amounts of useful information are removed from the data.

Results reported that exploiting spatial relations between neighboring pixels is more productive than exploiting spectral relations. While spectral regularization kept the results almost unchanged, spatial regularization improved the predictions by almost 20%. Spatial regularization is, hence, a key element to be exploited in pixelwise retrieval algorithms, where the spatial component is missing.

Experiments revealed that when the data is pre-processed to take into account the spectral and spatial feature relations before compression, the retrieval performance was improved compared to the results yielded by the original data even at high compression ratios. The compression setting Multilevel Clustering KLT + JPEG 2000 was able to significantly improve the atmospheric predictions at compression ratios higher than 200:1.

The proposed methodology can be applied to other retrieval methods and benefits the development of current and upcoming infrared sounding instruments. While current retrieval methods would benefit from efficiently compressed spectra, savings in data transmission and storage would involve operational improvements.

This study may have a deep impact on both currently flying infrared sounding instruments (e.g., IASI and CrIS) or upcoming (e.g., MTG-IRS).

REFERENCES

- [1] C. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. John Wiley & Sons, 2007.
- [2] F. García-Vilchez, J. Muñoz-Marí, M. Zortea, I. Blanes, V. González-Ruiz, G. Camps-Valls, A. Plaza, and J. Serra-Sagristà, "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.
- [3] S. E. Qian, M. Bergeron, I. Cunningham, L. Gagnon, and A. Hollinger, "Near lossless data compression onboard a hyperspectral satellite," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 3, pp. 851–866, 2006.
- [4] M. J. Ryan and J. F. Arnold, "Lossy compression of hyperspectral data using vector quantization," *Remote Sensing of Environment*, vol. 61, no. 3, pp. 419–436, 1997.
- [5] J. Sánchez and F. Perronnin, "High-dimensional signature compression for large-scale image classification," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2011, pp. 1665–1672.
- [6] J. Minguillon, J. Pujol, J. Serra-Sagristà, I. Ortuno, and P. Guitart, "Adaptive lossy compression and classification of hyperspectral images," in *Proc. of Image and Signal Processing for Remote Sensing VI*, January 2001, pp. 214–226.
- [7] 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.
- [8] A. Zabala, X. Pons, R. Díaz-Delgado, F. García, F. Aulí-Llinàs, and J. Serra-Sagristà, "Effects of JPEG and JPEG 2000 lossy compression on remote sensing image classification for mapping crops and forest areas," in *Proc. of IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS)*, July 2006, pp. 790–793.

- [9] I. Blanes and J. Serra-Sagrà, "Quality evaluation of progressive lossy-to-lossless remote-sensing image coding," in *16th IEEE International Conference in Image Processing (ICIP)*, 2009, pp. 3709–3712.
- [10] A. Zabala and X. Pons, "Effects of lossy compression on remote sensing image classification of forest areas," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 1, pp. 43–51, 2011.
- [11] —, "Impact of lossy compression on mapping crop areas from remote sensing," *International Journal of Remote Sensing*, vol. 34, no. 8, pp. 2796–2813, 2013.
- [12] S. Álvarez-Cortés, N. Amrani, M. Hernández-Cabronero, and J. Serra-Sagrà, "Progressive lossy-to-lossless coding of hyperspectral images through regression wavelet analysis," *Taylor and Francis International Journal of Remote Sensing*, vol. 39, no. 7, 2018.
- [13] 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.
- [14] 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.
- [15] I. Blanes, A. Zabala, G. Moré, X. Pons, and J. Serra-Sagrà, "Classification of hyperspectral images compressed through 3D-JPEG 2000," *Knowledge-Based Intelligent Information and Engineering Systems. Springer Berlin Heidelberg*, pp. 416–423, 2008.
- [16] 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.
- [17] 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.
- [18] C. D. Rodgers, *Inverse methods for atmospheric sounding: theory and practice*. World Scientific Publishing, 2000.
- [19] G. Camps-Valls, J. Muñoz-Marí, L. Gómez-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 PART 2, pp. 1759–1769, 2012.
- [20] F. Aires, "A regularized neural net approach for retrieval of atmospheric and surface temperatures with the IASI instrument," *Journal of Applied Meteorology*, vol. 41, pp. 144–159, 2002.
- [21] W. Blackwell, "A neural-network technique for the retrieval of atmospheric temperature and moisture profiles from high spectral resolution sounding data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 11, 2005.
- [22] W. Blackwell, M. Pieper, and L. Jairam, "Neural network estimation of atmospheric profiles using AIRS/IASI/AMSU data in the presence of clouds," in *Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications II, Proceedings of SPIE Vol. 7149*, Dec 2008.
- [23] V. Laparra, J. Malo, and G. Camps-Valls, "Dimensionality reduction via regression in hyperspectral imagery," *IEEE Journal on Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1026–1036, 2015, cited By 1. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84939542542&partnerID=40&md5=13304693b6bbca7255c26cec46d7d74e>
- [24] J. García Sobrino, Serra-Sagrà, V. Laparra, X. Calbet, and G. Camps-Valls, "Statistical atmospheric parameter retrieval largely benefits from spatial-spectral image compression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–12, 2017.
- [25] D. Malmgren-Hansen, V. Laparra, A. Nielsen, X. Calbet, and G. Camps-Valls, "Temperature retrieval with spatial noise-aware dimensionality reduction," in *2017 EUMETSAT Meteorological Satellite Conference*, Rome, Italy, 2–6 October 2017.
- [26] G. Camps, V. Laparra, J. Muñoz, L. Gómez, X. Calbet, T. Hultberg, and T. August, "Advances in non-linear retrievals for iasi and mtg-irs hyperspectral infrared sounders," in *EUMETSAT Meteorological Satellite Conference*, Geneva, Switzerland, September 2014, pp. 22–26. [Online]. Available: http://www.eumetsat.int/website/home/News/ConferencesandEvents/DAT_2076129.html
- [27] U. Amato, A. Antoniadis, I. De Feis, G. Masiello, M. Matricardi, and C. Serio, "Technical note: Functional sliced inverse regression to infer temperature, water vapour and ozone from iasi data," *Atmospheric Chemistry and Physics*, vol. 9, no. 14, pp. 5321–5330, 2009. [Online]. Available: <https://www.atmos-chem-phys.net/9/5321/2009/>
- [28] J. Arenas-García, K. Petersen, G. Camps-Valls, and L. 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, cited By 17.
- [29] U. Amato, A. Antoniadis, I. De Feis, G. Masiello, M. Matricardi, and C. Serio, "Technical note: Functional sliced inverse regression to infer temperature, water vapour and ozone from iasi data," *Atmospheric Chemistry and Physics*, vol. 9, no. 14, pp. 5321–5330, 2009.
- [30] H.-L. Huang and P. Antonelli, "Application of principal component analysis to high-resolution infrared measurement compression and retrieval," *J. Appl. Met.*, vol. 40, pp. 365–388, 2001.
- [31] J. García-Sobrino, I. Blanes, V. Laparra, G. Camps-Valls, and J. Serra-Sagrà, "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.
- [32] J. García-Sobrino, J. Serra-Sagrà, V. Laparra, X. Calbet, and G. Camps-Valls, "Statistical Atmospheric Parameter Retrieval Largely Benefits From Spatial-Spectral Image Compression," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 4, pp. 2213–2224, 2017.
- [33] EUMETSAT, "IASI Instrument," 2018, Available: <https://www.eumetsat.int/website/home/Satellites/CurrentSatellites/Metop/MetopDesign/IASI/index.html>.
- [34] D. Taubman and M. Marcellin, *JPEG 2000 image compression fundamentals, standards and practice: image compression fundamentals, standards and practice*. Springer Science & Business Media, 2002.
- [35] I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.
- [36] I. Blanes and J. Serra-Sagrà, "Pairwise orthogonal transform for spectral image coding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, pp. 961–972, March 2011.
- [37] —, "Cost and Scalability Improvements to the Karhunen-Loève Transform for Remote-Sensing Image Coding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2854–2863, 2010.
- [38] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
- [39] G. Camps-Valls, J. Muñoz-Marí, L. Gómez-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.
- [40] Surrey Satellite Technology Limited, "BILSAT-1: The Mission," Available: <https://www.sstl.co.uk/Missions/BILSAT-1-Launched-2003/BILSAT-1/BILSAT-1-The-Mission>.
- [41] Sharing Earth Observation Resources, "THEOS," Available: <https://directory.eoportal.org/web/eoportal/satellite-missions/t/theos>.
- [42] N. Ismailoglu, O. Benderli, S. Yesil, R. Sever, B. Okcan, and R. Oktem, "GEZGIN-2: an advanced image processing subsystem for earth-observing small satellites," in *Proceedings of 2nd IEEE International Conference on Recent Advances in Space Technologies, RAST 2005*, 2005, pp. 605–610.
- [43] Sharing Earth Observation Resources, "PICARD," Available: <https://directory.eoportal.org/web/eoportal/satellite-missions/p/picard>.
- [44] Geospatial Data Abstraction Library, "SENTINEL-2 Products," Available: http://www.gdal.org/frmt_sentinel2.html.
- [45] EUMETSAT, "IASI Level 1 PCC Product Generation Specification, EUM/OPS-EPS/SPE/08/0199," 2010.
- [46] —, "EPS Product Validation Report: IASI L1 PCC PPF, EUM/OPS-EPS/REP/10/0148," 2010.
- [47] D. Salomon, *Data compression: the complete reference*. Springer Science & Business Media, 2004.
- [48] L. Chang, C. M. Cheng, and T. C. Chen, "An efficient adaptive KLT for multispectral image compression," in *Proceedings of 4th IEEE Southwest Symposium on Image Analysis and Interpretation*, Austin, TX, 2000.
- [49] J. García-Sobrino, J. Serra-Sagrà, and J. Bartrina-Rapesta, "Hyperspectral IASI L1C data compression," *Sensors*, vol. 17, no. 6, p. 1404, 2017.
- [50] L. Gómez-Chova, J. Muñoz-Marí, V. Laparra, J. Malo-López, and G. Camps-Valls, *A Review of Kernel Methods in Remote Sensing Data Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 171–206. [Online]. Available: https://doi.org/10.1007/978-3-642-14212-3_10
- [51] B. Schölkopf and A. Smola, *Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press Series, 2002.
- [52] G. Camps-Valls, J. Verrelst, J. Muñoz-Marí, V. Laparra, F. Mateo-Jimenez, and J. Gomez-Dans, "A survey on gaussian processes for earth-observation data analysis: A comprehensive investigation," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 58–78, June 2016.

- [53] V. Laparra, J. Muñoz-Marí, L. Gómez-Chova, X. Calbet, and G. Camps-Valls, "Nonlinear statistical retrieval of surface emissivity from IASI data," in *2017 IEEE International Geoscience and Remote Sensing Symposium*, Fort Worth, Texas, USA, July 23–28 2017.
- [54] M. A. Aizerman, E. M. Braverman, and L. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and remote Control*, vol. 25, pp. 821–837, 1964.
- [55] F. Hilton, T. A. R. Armante, C. Barnet, A. Bouchard, C. Camy-Peyret, V. Capelle, L. Clarisse, C. Clerbaux, P. Coheur, A. Collard, C. Crevoisier, G. Dufour, D. Edwards, F. Faijan, N. Fourrié, A. Gambacorta, M. Goldberg, V. Guidard, D. Hurtmans, S. Illingworth, N. Jacquinet-Husson, T. Kerzenmacher, D. Klaes, L. Lavanant, G. Masiello, M. Matricardi, A. McNally, S. Newman, E. Pavelin, S. Payan, E. Péquignot, S. Peyridieu, T. Phulpin, J. Remedios, P. Schlüssel, C. Serio, L. Strow, C. Stubenrauch, J. Taylor, D. Tobin, W. Wolf, and D. Zhou, "Hyperspectral Earth Observation from IASI: Five Years of Accomplishments," *Bulletin of the American Meteorological Society*, vol. 93, no. 3, pp. 347–370, 2012.
- [56] M. George, C. Clerbaux, D. Hurtmans, S. Turquety, P. F. Coheur, M. Pommier, J. Hadji-Lazaro, D. P. Edwards, H. Worden, M. Luo, C. Rinsland, and W. McMillan, "Carbon monoxide distributions from the IASI/METOP mission: evaluation with other space-borne remote sensors," *Atmospheric Chemistry and Physics*, vol. 9, no. 21, pp. 8317–8330, 2009.
- [57] L. Clarisse, P. F. Coheur, A. J. Prata, D. Hurtmans, A. Razavi, T. Phulpin, J. Hadji-Lazaro, and C. Clerbaux, "Tracking and quantifying volcanic SO₂ with IASI, the september 2007 eruption at Jebel at Tair," *Atmospheric Chemistry and Physics*, vol. 8, no. 24, pp. 7723–7734, 2008.
- [58] C. Clerbaux, A. Boynard, L. Clarisse, M. George, J. Hadji-Lazaro, H. Herbin, D. Hurtmans, M. Pommier, A. Razavi, S. Turquety, C. Wespes, and P. F. Coheur, "Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder," *Atmospheric Chemistry and Physics*, vol. 9, no. 16, pp. 6041–6054, 2009.
- [59] C. Wespes, D. Hurtmans, C. Clerbaux, M. L. Santee, R. V. Martin, and P. F. Coheur, "Global distributions of nitric acid from IASI/MetOp measurements," *Atmospheric Chemistry and Physics*, vol. 9, no. 20, pp. 7949–7962, 2009.
- [60] M. Pommier, K. S. Law, C. Clerbaux, S. Turquety, D. Hurtmans, J. Hadji-Lazaro, P. F. Coheur, H. Schlager, G. Ancellet, J. D. Paris, P. Nédélec, G. S. Diskin, J. R. Podolske, J. S. Holloway, and P. Bernath, "IASI carbon monoxide validation over the arctic during polarcat spring and summer campaigns," *Atmospheric Chemistry and Physics*, vol. 10, no. 21, pp. 10655–10678, 2010.
- [61] G. Liuzzi, G. Masiello, C. Serio, S. Venafrà, and C. Camy-Peyret, "Physical inversion of the full ASI spectra: Assessment of atmospheric parameters retrievals, consistency of spectroscopy and forward modelling," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 182, pp. 128–157, 2016.
- [62] EUMETSAT, *IASI Level 2: Product Guide*, 2017, EUM/OPS-EPS/MAN/04/0033,.
- [63] D. S. Taubman, "Kakadu software," 2018, Available: <http://www.kakadusoftware.com/>.
- [64] GICI-UAB, "Pairwise Orthogonal Transform (POT) software," 2018, Available: <http://gici.uab.cat/GiciWebPage/downloads.php#pot>.
- [65] —, "Spectral Transform software," 2018, Available: <http://gici.uab.cat/GiciWebPage/downloads.php#spectral>.
- [66] JPEG, *JPEG 2000 Part 2 - Extensions*, 2004, document ISO/IEC 15444-2, 2004.
- [67] T. Hultberg, "IASI Principal Component Compression (IASI PCC) FAQ," 2009, Available: http://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET_FILE&dDocName=pdf_ipcc_faq&RevisionSelectionMethod=LatestReleased&Rendition=Web.
- [68] F. Aires, W. B. Rossow, N. A. Scott, and A. Chedin, "Remote sensing from the infrared atmospheric sounding interferometer instrument 1. Compression, denoising, and first-guess retrieval algorithms," *Journal of Geophysical Research: Atmospheres*, vol. 107, no. D22, 2002.
- [69] X. Calbet, J. Sanz, and M. A. Martínez-Rubio, "Scientific Report on noise characteristics dependency of the decompressed radiances when using different principal components compression methods and its application to MTG-IRS," 2017, Technical Report: NWC/CDOP3/MTG/AEMET/SCI/RP/PCnoise, AEMET.
- [70] LASIP, "Anisotropic Nonparametric Image Restoration Toolbox," 2018, Available: <http://www.cs.tut.fi/~lasip/2D/>.
- [71] F. R. Hampel, "The influence curve and its role in robust estimation," *Journal of the american statistical association*, vol. 69, no. 346, pp. 383–393, 1974.
- [72] F. Mosteller and J. Tukey, *Data analysis and regression: a second course in statistics*. Addison-Wesley Series in Behavioral Science: Quantitative Methods, 1977.
- [73] A. Foi, "Anisotropic nonparametric image processing: theory, algorithms and applications," Ph.D. dissertation, Dip. di Matematica, Politecnico di Milano, 2005.
- [74] 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.
- [75] 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.
- [76] V. Laparra and R. Santos-Rodríguez, "Spatial/spectral information trade-off in hyperspectral images," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 1124–1127.
- [77] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 623–656, 1948.
- [78] V. Laparra, J. Gutiérrez, G. Camps-Valls, and J. Malo, "Image denoising with kernels based on natural image relations," *Journal of Machine Learning Research*, vol. 11, pp. 873–903, 2010, cited By 9. [Online]. Available: <http://jmlr.org/papers/volume11/laparra10a/laparra10a.pdf>
- [79] J. Malo and V. Laparra, "Psychophysically tuned divisive normalization approximately factorizes the pdf of natural images," *Neural Computation*, vol. 22, no. 12, pp. 3179–3206, 2010, cited By 14. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-78649661352&partnerID=40&md5=427bcaa4858bfb2305575c83ee9a774d>
- [80] Y. Plokhenko, W. P. Menzel, H. E. Revercomb, E. Borbas, P. Antonelli, and E. Weisz, "Analysis of multispectral fields of satellite ir measurements: Using statistics of second spatial differential of spectral fields for measurement characterization," *International Journal of Remote Sensing*, vol. 29, no. 7, pp. 2105–2125, 2008.
- [81] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psych.*, vol. 24, 1933.
- [82] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, no. 4–5, pp. 411–430, May 2000. [Online]. Available: [http://dx.doi.org/10.1016/S0893-6080\(00\)00026-5](http://dx.doi.org/10.1016/S0893-6080(00)00026-5)
- [83] A. Hyvriinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.*, 1st ed. Springer Publishing Company, Incorporated, 2009.