

Journal of
Applied Remote Sensing

**Spatial pattern alterations from
JPEG2000 lossy compression of remote
sensing images: massive variogram
analysis in high performance computing**

Lluís Pesquer
Xavier Pons
Ana Cortés
Ivette Serral

Spatial pattern alterations from JPEG2000 lossy compression of remote sensing images: massive variogram analysis in high performance computing

Lluís Pesquer,^a Xavier Pons,^b Ana Cortés,^c and Ivette Serral^a

^aCREAF, Cerdanyola del Vallès 08193, Spain

l.pesquer@creaf.uab.cat

^bUniversitat Autònoma de Barcelona, Geography Department, Edifici B., E-08193 Bellaterra, Catalonia, Spain

^cUniversitat Autònoma de Barcelona, Computer Architecture and Operating Systems Department, Edifici Q, E-08193 Bellaterra, Catalonia, Spain

Abstract. We evaluate the implications of JPEG2000 lossy compression of remote sensing images for spatial analytical purposes. The main issue is to identify which cases and conditions in geostatistical studies are suitable for using lossy compressed images. For these purposes, an extensive test using Landsat, compact airborne spectrographic imager (CASI), and moderate resolution imaging spectroradiometer (MODIS) image series has been analyzed, through applying and comparing two-dimensional and three-dimensional (spectral and time domains) compression methods with a wide range of compression ratios for several dates, different landscape regions, and spectral bands. Due to the massive test bed and consequently to the high time consuming executions, a parallel solution was specifically developed. Variogram analyses showed that all the compression ratios maintain the variogram shapes, but high compression ratios (>20:1) degrade the spatial patterns of the remote sensing images. These alterations are lower for the three-dimensional compression method, which was a considerable improvement (25%) on the two-dimensional method for large three-dimensional series (CASI, MODIS). However, the two methods behave similarly in the Landsat case. Finally, the parallel solution in a distributed environment demonstrates that high performance computing offers a suitable scientific platform for highly demanding time execution applications, such as geostatistical analyses of remote sensing images. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.7.073595](https://doi.org/10.1117/1.JRS.7.073595)]

Keywords: geostatistics; remote sensing images; lossy compression; high performance computing.

Paper 12062P received Mar. 12, 2012; revised manuscript received Oct. 23, 2012; accepted for publication Dec. 14, 2012; published online Jan. 17, 2013.

1 Introduction and Objective

Remote sensing provides large amounts of data to the scientific community in terms of spatial, spectral, radiometric, and temporal resolutions¹ that can be used in a vast range of applications in agriculture,² climate change research,³ urban management,⁴ ecosystems monitoring,⁵ and thematic mapping,⁶ among others. It is a very interesting field for evaluating the quality of lossy compression techniques.

Lossy data compression facilitates accessing, sharing, and transmitting huge spatial datasets in environments with limited storage or with limited bandwidth. JPEG2000 is one of the possible lossy compression algorithms, and is currently an ISO standard.⁷ However, lossy data compression procedures modify the original information,⁸ and therefore rigorous studies are needed to understand the effects and consequences of this manipulation. Some previous works have

analyzed these alterations in remote sensing images in different fields and with varying objectives: spectral analysis,⁹ digital classification,¹⁰ texture analysis,¹¹ stereoscopy,¹² and multivariate regression,¹³ among others.

The present work aims at studying the effect of JPEG2000 lossy compression by determining the differences between the spatial pattern domains in the original image and the compressed image, and more particularly, the impact on the geostatistical usage of remotely sensed imagery. Comparing the geostatistical properties of compressed images before and after compression is a different and novel approach, explored here instead of the usual comparison of the global parameter peak signal to noise ratio (PSNR).¹⁴

Exploring and describing the spatial variation in images is one of the main applications of geostatistics in remote sensing¹⁵ as it can provide parameters for describing spatial patterns,¹⁶ measurements for spatial autocorrelation,¹⁷ procedures for downscaling images,¹⁸ tools for estimating continuous variables,¹⁹ data for radiometric coregionalization analysis,²⁰ and optimal sampling designs for ground surveys.²¹ The variogram is an appropriate tool often used in geostatistics to carry out exploratory analyses.²² Section 2.4 of the present work outlines this geostatistic tool and its specific characteristics when it is applied to remote sensing images. It is, however, a particularly slow procedure for processing large amounts of remotely sensed data. This computational constraint makes it necessary to use distributed environments, such as high performance computing (HPC). HPC provides methods and infrastructures for distributing computation in order to reduce the total execution time. Some examples of the benefits of HPC applied to remote sensing are hyperspectral image modeling,²³ fire monitoring,²⁴ meteorological applications,²⁵ image processing,²⁶ some classification methods²⁷ and web grid environments,²⁸ among others.

In this work, parallel computing plays an important role in allowing us to perform spatial pattern analyses within an acceptable period of time. Section 2.5 details the architecture, design paradigm and language, and programming libraries, as well as results in time execution performance in order to generate variogram analyses in an efficient scientific environment.

Many different cases were studied in order to obtain robust conclusions regarding the main questions raised: is the degradation similar at short and far distances applying different compression ratios? Do the analyzed compression methods modify patterns at different directions? Do possible pattern alterations depend on three or two dimension compression methods? Are the short and large remote imagery series behaviors similar in spectral and time dimension?

The rapid generation of variograms allowed applying these massive analyses to:

- Types of sensor images: multispectral series (Landsat), hyperspectral series (CASI), and large time series (MODIS) in a spatial resolution range from 3 to 250 m.
- Landscape regions with different spatial patterns.
- Scenes with different season phenology.
- A wide range of compression ratios.

This complete test bed is detailed in Sec. 2.1.

These different scenarios allow studying the possible spatial pattern alterations generated by the lossy compression methodologies and exploring the magnitudes and properties of these hypothetical changes. The conclusions of the work provide recommendations in each specific case about applying JPEG2000 lossy compression algorithms to remote sensing images for geostatistical analysis purposes.

2 Methodology

The proposed methodology is a chain of image processing, lossy compression procedures, and geostatistical analyses in an HPC environment. Figure 1 shows the different stages and elements of this methodology as well as the position of the parallel task within the processing chain. These stages are detailed in the following subsections.

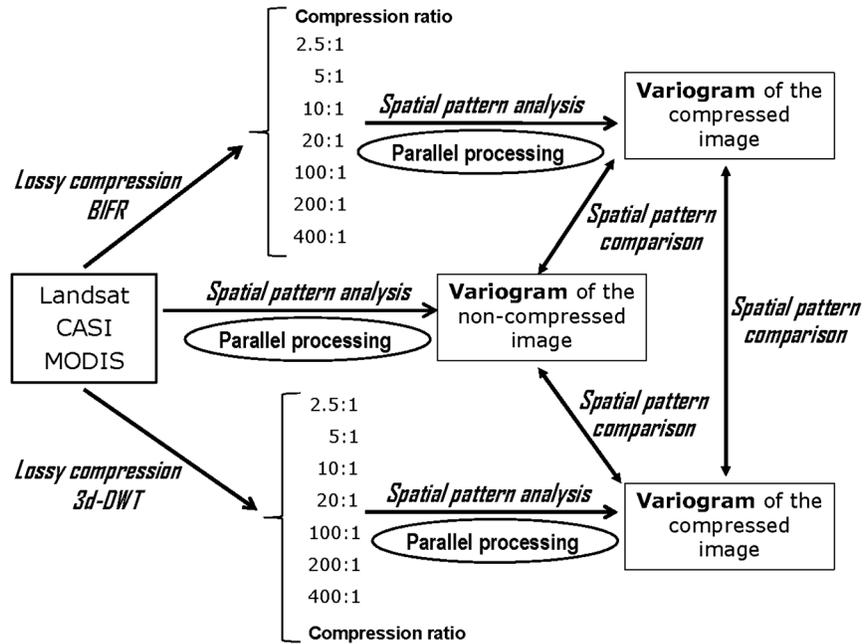


Fig. 1 Flowchart of the proposed methodology.

2.1 Materials

2.1.1 Landsat

In this study, four scenes of multispectral (seven bands) Landsat-5 Thematic Mapper (TM) images have been used.²⁹ The images selected for this work correspond to the dates 13-4-2006, 02-7-2006, 19-08-2006, and 11-9-2006, and cover two areas of about 15 km × 15 km (spatial resolution of 20 m) of the 197-031 and the 198-31 path-rows. They thus provide a set of images that is very representative of the main phenological cycle in this Mediterranean region. These two areas, in the Ripollès and Penedès counties, (Fig. 2, pink

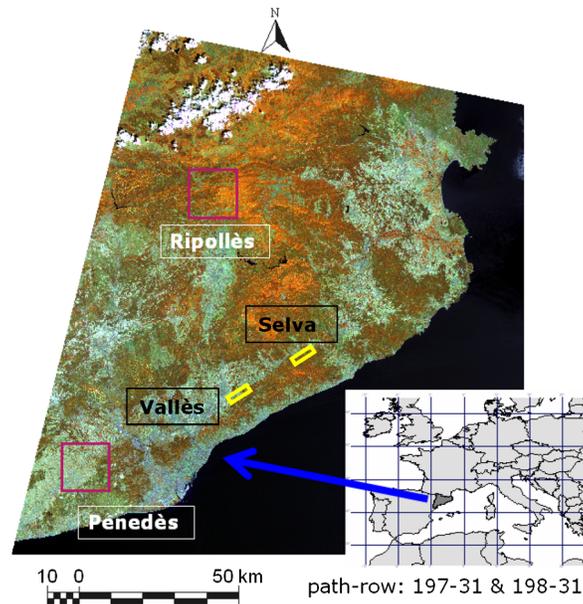


Fig. 2 Study regions: the Ripollès and Penedès areas for Landsat and MODIS are shown as pink rectangles and the Vallès and Selva areas for CASI images are shown as yellow rectangles.

squares) are located in Catalonia, a region of approximately 32,000 km² situated in the northeast Iberian Peninsula at the extreme southwest of Europe (map location Fig. 2). The main landscape of the Ripollès region is deciduous forest, while the Penedès region has a vineyard landscape. Landsat images were chosen because they are the remote sensing images most used for research purposes in different applications,³⁰ and they are also used in a large variety of applied projects.³¹

2.1.2 CASI

The compact airborne spectrographic imager (CASI) is an optic sensor for hyperspectral scanning based on a small CCD bar. In this study, two images (provided by the Institut Cartogràfic de Catalunya, ICC) with 72 spectral bands from red to near infrared (409.15 to 955.70 nm and an average bandwidth of 7.6 nm) were used.³² The two scenes selected correspond to 18-05-2007 and 19-06-2007. For these two dates, two different study areas of 4.3 km × 1.2 km at a spatial resolution of 3 m (Fig. 2, yellow rectangles) called Vallès and Selva located in the Montseny region, a mountain landscape in Catalonia. CASI images were chosen as an interesting test bed for three-dimensional compression methodologies (hyperspectral images are large radiometric third dimension examples).

2.1.3 MODIS

A moderate resolution imaging spectroradiometer (MODIS) Surface Reflectance Daily L2G Global 250 m product³³ was used in this study as an example of a large time series dataset used for analyzing the geostatistical patterns of compressed and noncompressed images. The time series include 73 selected cloud free images from 2007 (from 5-5-2007 to 30-09-2007) from the same regions as the two Landsat images analyzed (Penedès and Ripollès), but covering a more extensive area. Higher spatial resolution MODIS bands were used, i.e., the near infrared and red spectral bands. The region covered is a square of 25 × 25 km² centered in the 15 × 15 km² square of the corresponding Landsat dataset. The study area of MODIS images was expanded in order to obtain a number of pixels with a resolution of 250 m large enough for statistical purposes, in this case 100 columns × 100 rows.

In summary, this remote sensing material was selected because Landsat images are used most widely, CASI serves to extend the third spectral dimension, and MODIS serves to extend the third time dimension tests. Both CASI and MODIS provide large datasets, which improves the lossy compression analyses.

2.2 Image Processing

Different image processing methods were applied for each type of image depending essentially on the previous image processing level and on their particular characteristics.

Landsat images were acquired at the L1G processing level, and were then georeferenced and radiometrically corrected with MiraMon GIS software³⁴ according to the methodologies detailed in Refs. 35 and 36, and applied to the *SatCat* server.³⁷ The optical reflectance range of the corrected images extends from 0 to 10,000, representing the % of reflectance*100, and so numbers were written with two decimal places in a short integer. This factor and type of data (signed short integer) are also used to represent the possible range of ground temperatures in °C (derived from the radiometric correction) of the TM thermal band. All images use a -999 value to represent nodata regions, which in this case are normally caused by sensor problems or by post-processing artifacts. Defining the nodata value as -999 is not arbitrary because nodata values that are defined too far from the valid data range can negatively affect normalization compression procedures.³⁸ Consistent treatment of nodata values is one of the keys to performing correct image processing and subsequent data analysis.

However, CASI images were acquired geometrically georeferenced using the SISA system developed by ICC.³⁹ The flight orientation was SW-NE, which resulted in the geometrically corrected image being very large (18,366 columns × 11,918 rows), and therefore the CASI images were rotated -32.5 degrees with respect to the north projection in order to reduce the large amount of nodata pixels. The value range, data type and nodata definition in the

CASI images were unified to match those of the Landsat images, also using the MiraMon software.

Finally, the MODIS images were downloaded from the Warehouse Inventory Search Tool⁴⁰ as a georeferenced reflectance product. It was only necessary to clip them to the study regions already detailed in the previous section and unify the nodata values.

2.3 Lossy Compression

The reflectance product obtained was compressed at a wide range of different quality levels with the following compression ratios (CR): 2.5:1, 5:1, 10:1, 50:1, 100:1, 200:1, and 400:1 (from soft to hard compression). Two methodologies, based on the JPEG2000 lossy compression procedures, were evaluated in this work:

- The band-independent fixed-rate (BIFR) method is an independent compressor in which no inter-band redundancy is exploited and the bit-rate is split equally among all bands. BIFR is therefore a two-dimensional (space) compression method.
- Three-dimensional discrete wavelet transform (3d-DWT) is an inter-band decorrelation technique that exploits the spectral (or temporal) redundancy between multiple bands or scenes: the third dimension. This transform is applied to the input image to obtain a spectrally (or temporally) decorrelated image.

Both compression procedures were applied using the Kakadu software,⁴¹ which is a complete implementation of the JPEG2000 standard, Part 1, completed with a great deal of Part 2 and 3.⁴² The BOÍ software,⁴³ an implementation of the JPEG 2000 (Part 1) standard, was used for calculating the PSNR quality compression indicator in reference to noncompressed images.

2.4 Geostatistics

Geostatistics is a subset of statistics specialized in the analysis, interpretation, and inference of geographically referenced data.⁴⁴ It was initially developed by Matheron⁴⁵ and defined as a set of methods for studying the spatial distribution of a variable from a statistical approach in order to estimate the corresponding unknown value of a particular location or simulate its variability. When this variation has a spatial structure, the regionalized variable theory makes it possible to take spatial properties into account following a stochastic approach, and assumes a constant local mean and a stationary variance of the differences between regions separated by a given distance and direction.⁴⁶ The variogram (also referred to as semivariogram) is based on these previous assumptions and, as defined in Eq. (1), it plots the variance function depending on sample distance separation and, consequently, the spatial patterns analyzed. This expression implies, for remote sensing images, that pixel values variance is only a function of distance and direction between pixels; then, there is no dependence on their particularly locations and, in statistics sense, this variance pattern is stationary for all study regions. Therefore, the variogram describes the pattern distribution of image spatial resolution, and it can identify a study region, a subscene of remote sensing image in this work.

$$\gamma(h) = \frac{1}{2 \cdot n} \sum_{i=1}^n [z(\vec{x}_i) - z(\vec{x}_i + h)]^2. \quad (1)$$

Equation 1: variogram definition. Note that the dependence is on distance (h) and on the squared difference of z values (pixel values), and independent of particular positions (x_i and $x_i + h$).

The sampled variogram can be modeled and fitted by a continuous function to identify structural parameters that characterize the spatial pattern for any quantitative variable distribution, including, of course, those of remotely sensed images, as carried out here in this study and in other previous studies⁴⁷ (Fig. 3). These parameters are:

- Nugget: the variance near the origin, so at very low distances, it represents the fluctuations at scales smaller than the sampling interval (lag distance), and the component of the nonspatially correlated error.

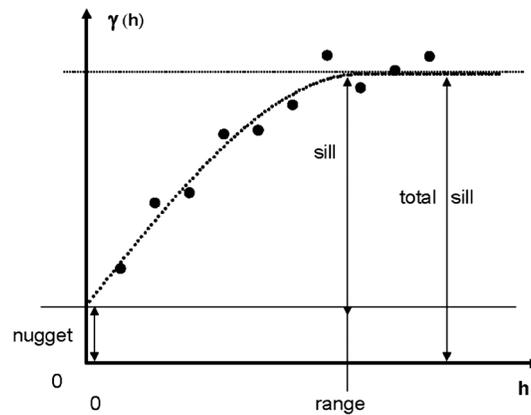


Fig. 3 Variogram structure parameters.

- Range: the distance at which the variogram reaches saturation. It means that variance becomes stabilized despite distances increase and it defines the limit distance of autocorrelation.
- Sill (partial): the variance at the variogram saturation excluding nugget variance, it means a measure for spatially correlated variability.
- Total sill: sum of the partial sill and nugget. It represents the global (random + spatially correlated) variability at the autocorrelation threshold.

For each specific geostatistical application, the key role parameter depends upon the focus of a particular spatial pattern analysis. For example, range is the most suitable for autocorrelation analysis, nugget for determining stochastic noise and sill for variability's measurements.

The total sill was chosen as the main quality measure shown in Sec. 3 for analyzing alterations in the spatial patterns at studied compression ratios and with different methodologies. Different selected plots were used for three sensors for several dates, bands, and regions, which made it possible to determine the complete variogram structure by comparing the theoretical plot in Fig. 3 with the experimental plots in Sec. 3.

2.5 Parallel Solution

Most variogram analyses are used to explore the spatial patterns of irregularly distributed samples.⁴⁸ Variogram analysis is usually a previous step to interpolation (kriging), which is a process with a high computational cost.⁴⁹

Nevertheless, the variogram analysis of remotely sensed images has higher computing demands to those of irregularly distributed samples. Indeed, a large number of pixels are involved even in geographically small scenes with a high or medium spatial resolution, and therefore a very large number of pairs are necessary to obtain the variogram. Table 1 compares four examples of the amount of data involved and the time necessary for executions running on a single, normal PC (Intel(R) Xeon (R) 3.0 GHz and 1 GB of RAM).

In this work, a large number of spatial pattern analyses have been executed. All spectral bands (the number of bands for Landsat, CASI, and MODIS are 7, 72, and 2, respectively) on several scene dates (4, 2, and 73) and in two different study regions were analyzed. In addition, the images were compressed at several compression ratios (eight in all cases) and compared with noncompressed images. This implied 504 analyses for Landsat, 2592 for CASI, and 2628 for MODIS. Performing all these analyses would be a very time-consuming process; a single CASI analysis takes 57,727 s (see Table 1). Therefore, a distributed environment was the most suitable solution for a complete and efficient study.

In order to reduce the execution time, the authors carried out a parallel implementation of the variogram analyses. The master/worker programming paradigm was used in which the master process schedules, distributes, and coordinates the tasks executed by the worker processes. The code was written in ANSI-C language, including MPI (message passing interface) as a message-passing library (version 1.2.7) that is used by the master process to communicate with the worker

Table 1 Comparison of processing time of variogram analyses between two irregularly distributed point measurements and remote sensing images. The sparse irregular distribution corresponds to a network of weather stations, and the dense, irregular, distribution to a GPS network of altimetry measurements, both detailed in Ref. 44; the remotely sensed image corresponds to two (Landsat and CASI) of the three examples used in this work.

Type	Extension (km ²)	<i>N</i> data	<i>N.</i> pairs examined	Time (s)
Sparse point irregular sampling	69,284.96	100	1070	26
Dense point irregular sampling	0.216	2855	1,627,350	946
Remotely sensed image: medium resolution	225	562,500	16,493,978,252	43,850
Remotely sensed image: high resolution	6.5	725,145	33,554,502,751	57,727

processes. Since MPI has become a standard de-facto message passing library,⁵⁰ this solution (ANSI-C + MPI) guarantees portability to different computer platforms.

The distributed load design⁵¹ is not specifically defined and optimized for the characteristics of the images used in this work. However, it is a flexible design, tested with a wide range of image dimension that maintains a good balance and scalability suitable for other studies with a wide range of image dimensions and cluster properties. The scalability solution (adapting point samples analysis from Ref. 49 to remote sensing image analysis in the present work) can be achieved by defining a relatively small load unit: of two image rows of pixels. Each worker analyzes all possible combinations of nonrepeated pairs between two rows of the image and returns the partial variance and distance of the pairs involved to the master. The master distributes tasks (rows of data) to workers, accumulates partial results and manages input (remotely sensed image) and output (resulting variogram) tasks.

3 Results

Two types of results are presented in this section: geostatistical and computational. The geostatistical results are a set of variograms that compare images compressed at different ratios and noncompressed images for three image types (Landsat, CASI, and MODIS) focusing on the spectral or time dimension and possible anisotropy. The total sill parameters of the theoretical variogram are shown in different tables, while the experimental variograms are shown in plots.

The computational results are focused on performance analyses of the parallel solution. The example shown in Sec. 3.2 corresponds to Landsat executions, but the performance results are very similar for all the image types analyzed.

3.1 Variogram Analysis

As explained above, lossy compression mechanisms modify the original data information. Figure 4 illustrates the effects of compression ratios on quality visualization of the images. In this figure, the central image is a noncompressed CASI example, the top image sequence corresponds to three BIFR compressed images; from left to right, it is shown increasing compression ratios and decreasing quality. The bottom image sequence corresponds to the same sequence improved using 3d-DWT techniques. This composition summarizes most of the results presented in this section, and the following tables and figures quantify these effects in a geostatistical sense using the variogram spatial pattern analysis.

3.1.1 Landsat

The first plot of Fig. 5 is a representative result of the alterations in the spatial pattern caused by the lossy compression (BIFR in this case) mechanisms. The main alteration to the spatial pattern due to lossy compression is the reduction in data variability at high compression ratios (more than 100:1), although at lower CRs (from 2.5:1 to 50:1) the variograms have very similar

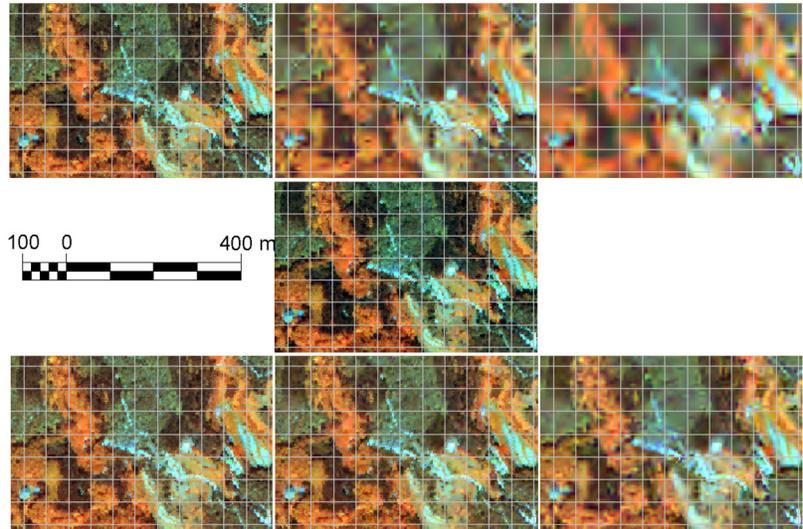


Fig. 4 The effects of compression on a CASI image fragment from the Selva region: noncompressed in the middle, upper sequence of BIFR compression ratios: 2.5:1, 10:1, and 400:1. The images in the bottom sequence were compressed at the same ratios but the 3d-DWT method was used. All of them correspond to a 4-5-3 band composite image.

shapes and structural parameters (nugget, range and sill). Table 2 is a summary of the behavior of the total sill, which is the most representative parameter in this study, for two dates and two regions for the BIFR method. Comparing the sill between compressed and noncompressed images represents a measurement of the loss of detail and variability caused by a compression method for all Landsat spectral bands. Table 3 shows the same information in relation to the 3d-DWT compression method. As seen in Fig. 6, there are only small differences between the two methods for Landsat images with the highest CR. Therefore, the following tables and figures corresponding to Landsat images only show the results for the BIFR method. The next cases (CASI and MODIS) demonstrate that 3d-DWT is a more useful method for image series that are larger than Landsat image series.

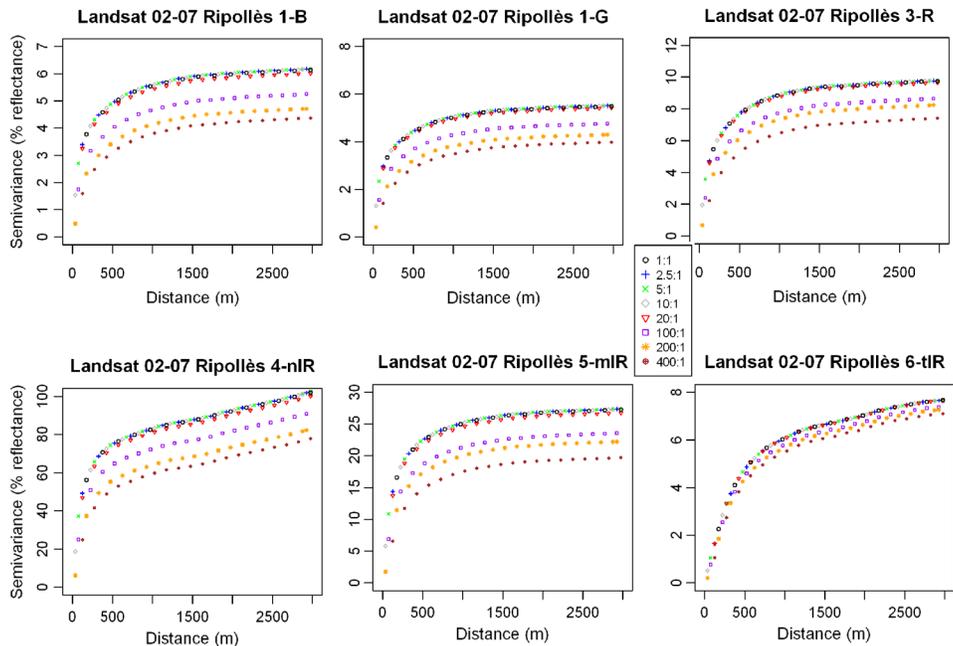


Fig. 5 Variogram results of the BIFR compression method on a Landsat image from the Ripollès region on 02-07-2006, a multispectral view.

Table 2 Two examples (Ripollès region on 02-07-2006, and Penedès region on 11-9-2006) of the variation in the variogram parameter total sill at different compression ratios for the BIFR method.

	Band/CR	1:1	2.5:1	5:1	10:1	20:1	100:1	200:1	400:1
Ripollès02-07-2006	1-B	5.92	5.95	5.96	5.95	5.80	5.04	4.51	4.17
	2-G	5.30	5.34	5.35	5.33	5.23	4.58	4.13	3.81
	3-R	9.40	9.44	9.44	9.43	9.31	8.3	7.90	7.07
	4-nIR	92.96	93.30	93.34	93.19	90.93	82.33	74.19	69.57
	5-mIR1	26.31	26.40	26.43	26.38	25.77	22.63	21.28	18.85
	6-tIR	7.10	7.10	7.11	7.11	7.09	6.84	6.72	6.53
	7-mIR2	15.74	15.78	15.81	15.69	15.40	13.8	13.41	12.50
Penedès11-9-2006	1-B	18.01	18.51	18.57	18.53	18.06	15.68	15.68	15.68
	2-G	17.30	17.73	17.79	17.77	17.33	14.96	14.96	14.95
	3-R	25.94	26.37	26.43	26.31	25.68	22.21	22.21	22.21
	4-nIR	36.49	37.8	38.00	37.49	36.23	30.33	30.33	30.32
	5-mIR1	44.61	45.53	45.67	45.11	43.76	37.43	37.43	37.43
	6-tIR	3.53	3.54	3.55	3.53	3.53	3.35	3.35	3.36
	7-mIR2	31.54	32.06	32.17	31.9	31.15	27.24	27.23	27.24

Table 3 Two examples (Ripollès region on 02-07-2006, and Penedès region on 11-9-2006) of the variation in the variogram parameter total sill at different compression ratios for the 3d-DWT method.

	Band/CR	1:1	2.5:1	5:1	10:1	20:1	100:1	200:1	400:1
Ripollès02-07-2006	1-B	5.92	5.95	5.97	5.93	5.88	5.54	5.35	5.02
	2-G	5.30	5.33	5.34	5.33	5.30	5.05	4.89	4.61
	3-R	9.40	9.42	9.45	9.45	9.37	9.06	8.87	8.65
	4-nIR	92.96	92.80	92.83	92.19	90.44	82.75	76.49	70.17
	5-mIR1	26.31	26.37	26.40	26.04	25.33	22.47	20.42	18.34
	6-tIR	7.10	7.08	7.10	7.20	7.29	7.31	7.28	7.16
	7-mIR2	15.74	15.76	15.80	15.80	15.62	14.37	13.63	12.84
Penedès11-9-2006	1-B	18.01	18.45	18.52	18.59	18.38	16.81	16.37	15.69
	2-G	17.30	17.64	17.72	17.71	17.53	16.33	15.76	14.93
	3-R	25.94	26.26	26.35	26.06	25.34	22.47	20.73	19.15
	4-nIR	36.49	37.62	37.77	37.40	35.99	31.00	27.46	23.84
	5-mIR1	44.61	45.33	45.43	44.98	43.53	39.05	36.35	32.06
	6-tIR	3.53	3.52	3.56	3.82	4.28	5.18	5.73	6.44
	7-mIR2	31.54	31.93	32.03	32.34	32.33	30.53	29.44	27.97

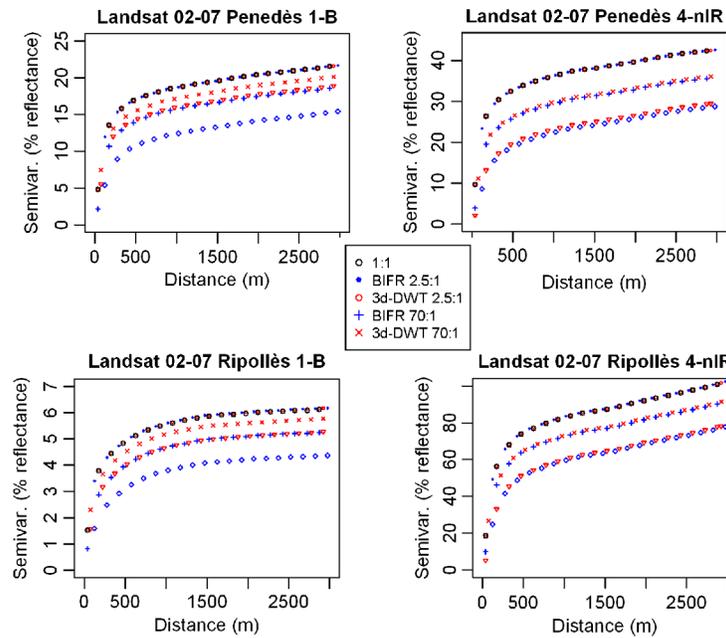


Fig. 6 Landsat variogram comparison between BIFR and 3d-DWT compression in relation to the original pattern.

The sequenced plot in Fig. 5 shows that there are no significant differences in the variogram shape between bands, although there are differences in band range variances, and consequently in structural parameters depending on the region of the landscape.

Table 4 and Fig. 7 reproduce the same variogram spatial pattern in a temporal sequence for an example of the near infrared (4-nIR) band.

Furthermore, lossy compression may produce different variance alterations in relation to the spatial direction. In order to analyze this possible anisotropy modification, variance pattern at different azimuth angles has been studied. Figure 8 shows that the lossy compression mechanisms used in this work maintain the expected directional patterns at all compression ratios; thus, this result proves that the previous variograms were always omnidirectional.

Table 4 Time comparison of variogram results (total sill) for BIFR compression of Landsat images from the Penedès region on the 1-B and nIR bands.

Date	Band/CR	1:1	2.5:1	5:1	10:1	20:1	100:1	200:1	400:1	
Penedès	13/04/2006	1-B	22.32	22.01	22.03	22.21	21.96	20.07	19.12	17.72
	4-nIR		38.46	38.48	38.48	38.23	36.88	31.36	28.10	24.63
	02/07/2006	1-B	20.51	20.12	20.12	20.29	20.21	18.68	18.37	17.42
	4-nIR		40.49	39.70	39.70	39.38	38.09	33.33	30.04	26.95
	19/08/2006	1-B	37.96	37.27	37.29	37.34	36.94	33.89	32.60	30.76
	4-nIR		54.20	53.35	53.38	53.18	51.74	45.71	42.01	37.39
	11/09/2006	1-B	18.01	18.45	18.52	18.59	18.38	16.81	16.37	15.69
	4-nIR		36.49	37.62	37.77	37.40	35.99	31.00	27.46	23.84

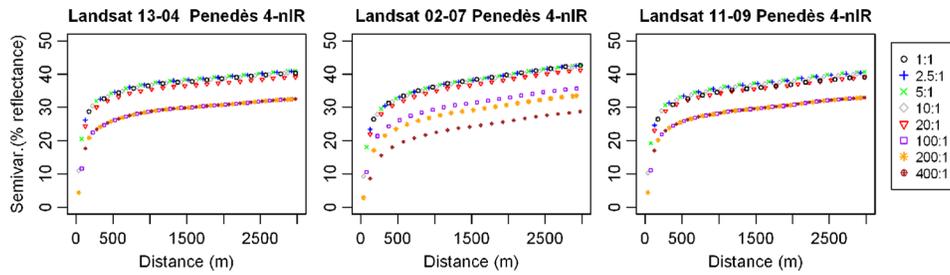


Fig. 7 Date comparison of variogram results for BIFR compression of Landsat images from the Penedès region on the near infrared (4-nIR) band.

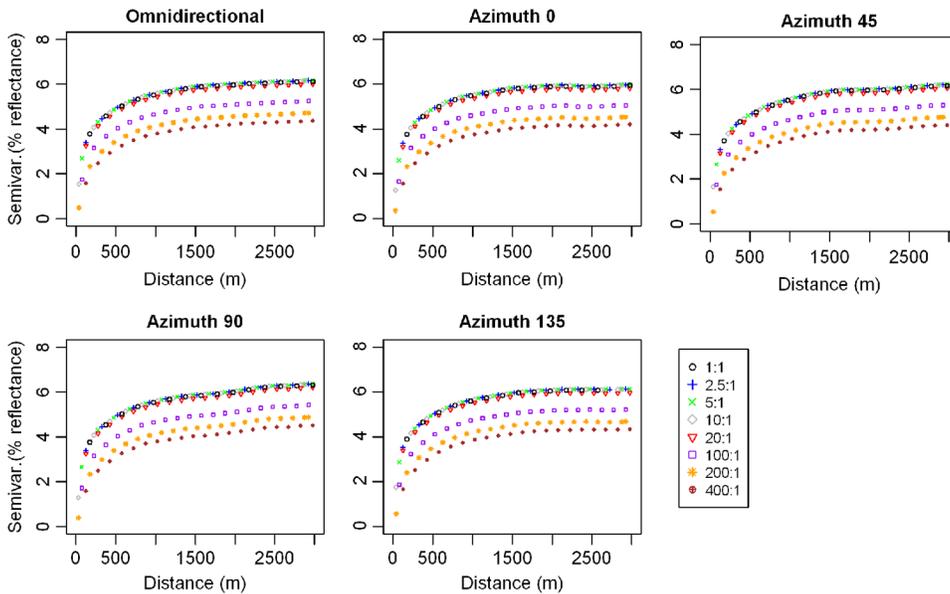


Fig. 8 Variogram behavior in relation to different azimuth directions. Landsat 02-07-2006 1-B band images.

3.1.2 CASI

The main characteristic of CASI images in relation to Landsat images is their higher spectral resolution (much more data in the spectral domain). The higher spatial and radiometric resolution are also notable, but they are not the focus of the present study. It is therefore an interesting test bed for exploiting three-dimensional compression methods. As it is impossible to show the results for 72 bands for two regions and two dates, the following figures display representative results for selected bands for different regions and dates. For example, Table 5 shows the total sill

Table 5 Total sill parameters for the k4p scene in the Selva region and for the complete sequence of compression ratios at four representative spectral bands.

Band/CR	1:1	2.5:1	5:1	10:1	20:1	100:1	200:1	400:1
SelvaK4p	10	14.07	14.07	14.07	13.82	13.91	12.25	10.37
	21	36.88	36.39	36.41	36.21	35.13	29.81	23.62
	34	40.92	41.18	41.20	40.97	40.26	37.71	31.88
	58	398.05	398.05	397.98	395.84	388.54	370.07	356.27

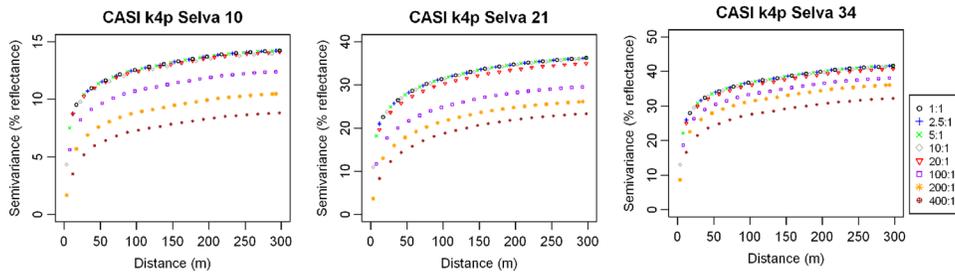


Fig. 9 Variogram results for the 3d-DWT compression method in the Selva region at selected CASI bands on 19-06-2007 (*k4p* scene).

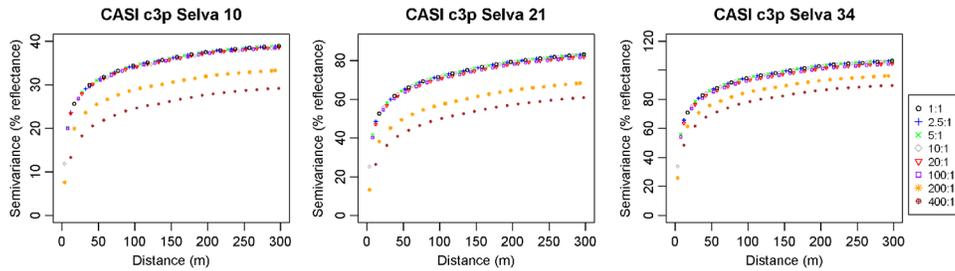


Fig. 10 Comparison of variogram results for the 3d-DWT compression method in the Selva region at selected CASI bands on 18-05-2007 (*c3p* scene).

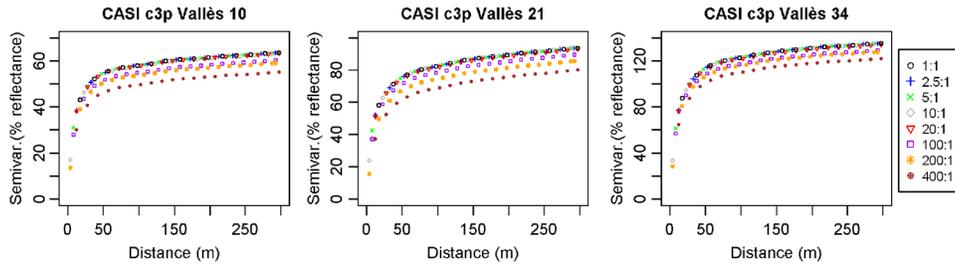


Fig. 11 Variogram results for the 3d-DWT compression method in the Vallès region at selected CASI bands.

parameter of the 3d-DWT method for four selected bands (10, 21, 34, and 58) in the *k4p* scene (19-06-2007), while Fig. 9 plots variograms at different compression ratios for three selected bands. Figures 10 and 11 show the variogram analyses of the *c3p* scene for the two regions applying the 3d-DWT compression method.

Table 6 shows the total sill parameter for the other scene (*c3p* on 18-05-2007) in the two study regions (Selva and Vallès) comparing the BIFR and 3d-DWT compression methods at selected CRs and bands (10, 21, 34). This table comparison demonstrates that spatial patterns have better fidelity with the 3d-DWT method than the BIFR method for CASI images. Figure 12 shows that 3d-DWT plots are closer to the noncompressed patterns, and that BIFR significantly alters the spatial patterns, especially for high compression ratios.

3.1.3 MODIS

The MODIS test bed allows a much larger amount of data to be explored in the time domain than the Landsat images because the MODIS daily revisit period makes it possible to obtain a large cloud-free time series. These series are especially appropriate for analyzing the pattern alterations of three-dimensional compression methods *versus* two-dimensional ones.

Table 6 Comparison between total sill parameters for the BIFR and 3d-DWT methods applied to the *c3p* scene in two regions at four representative compression ratios and three spectral bands.

<i>c3p</i>	Band/CR	1:1	2.5:1		5:1		20:1		100:1	
			BIFR	3d	BIFR	3d	BIFR	3d	BIFR	3d
Selva	10	38.65	38.65	38.66	38.69	38.67	37.24	38.31	30.59	38.31
	21	84.98	84.99	84.98	78.85	85.01	78.85	83.59	65.12	83.59
	34	105.63	105.97	105.93	101.46	106.01	101.46	103.88	85.42	103.88
Vallès	10	65.88	65.88	65.87	66.05	65.84	63.90	65.61	55.97	62.49
	21	99.38	99.38	99.38	99.40	99.36	98.46	98.46	82.55	95.26
	34	139.98	142.59	139.99	143.54	139.99	139.11	138.73	124.53	134.03

The results for the MODIS images are similar to those obtained for CASI images, but both are different from the Landsat images. In these cases, the 3d-DWT method improves the quality of BIFR compressions. Table 7 shows that, at low compression ratios, 3d-DWT and BIFR maintain the spatial variability patterns of the original images, and that, at high compression ratios, BIFR loses quality compared to the 3d-DWT method. The plots in Figs. 13 and 14 (corresponding to the same region but to different spectral bands) show a reduction in the total sill parameter at increasing compression ratios for the 3d-DWT method, while Fig. 15 shows that 3d-DWT produces the pattern variability more faithfully than BIFR.

In summary, these results quantify spatial pattern alterations for lossy compression images depending on compression ratios, compression methods, and series dimension through analyzing differences between structural variogram parameters (focusing on total sill) of compressed and noncompressed images. These results shows that, for compression ratios higher than 1 : 100, the variogram is clearly degraded, reducing variability and thus, some capabilities or accuracy in related applications. This degradation could be partially solved, in some cases, by 3d-DWT, but this method needs a large amount of redundant data for significantly improve the BIFR method. These improvements are similar for time series and spectral series.

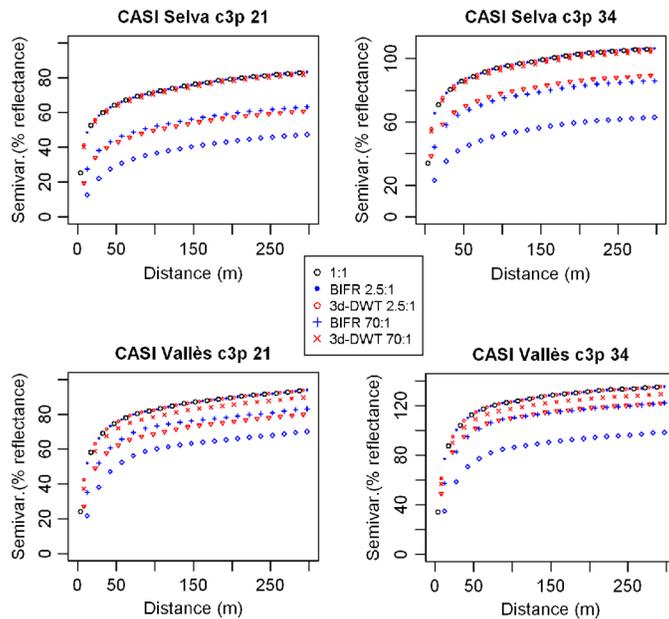


Fig. 12 Comparison of variogram results of the 3d-DWT and BIFR compression methods in both regions at two selected CASI bands.

Table 7 Comparison between the BIFR and 3d-DWT methods of the total sill for four selected dates in the Penedès region for two spectral bands at representative compression ratios.

Penedès	CR	1:1	2.5:1		20:1		70:1	
Date	Band		BIFR	3d	BIFR	3d	BIFR	3d
176	RED	20.04	20.06	20.02	19.10	19.10	12.50	19.45
	NIR	19.46	19.43	19.46	17.62	18.70	13.93	18.73
207	RED	20.23	20.27	20.21	20.20	20.20	11.67	19.70
	NIR	20.41	20.48	20.46	19.35	20.04	13.63	18.56
249	RED	19.77	19.86	19.92	19.90	19.90	11.05	18.52
	NIR	29.05	29.06	29.13	26.45	28.49	23.62	27.20
271	RED	16.43	16.42	16.43	15.83	15.83	10.13	13.97
	NIR	18.72	18.70	18.70	17.47	17.63	9.91	16.25

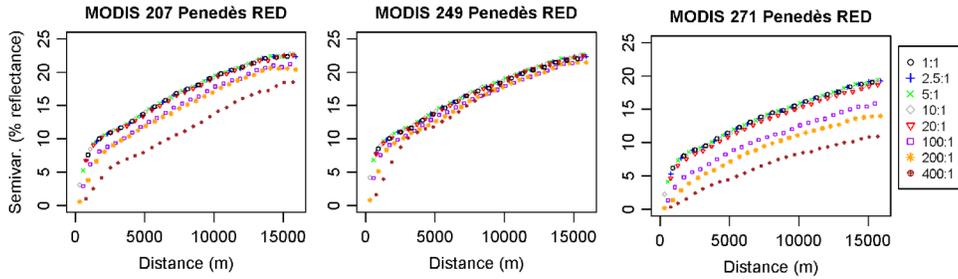


Fig. 13 Comparison of experimental variograms of the 3d-DWT compression method for the Penedès region on three selected dates applied to MODIS images in the red band.

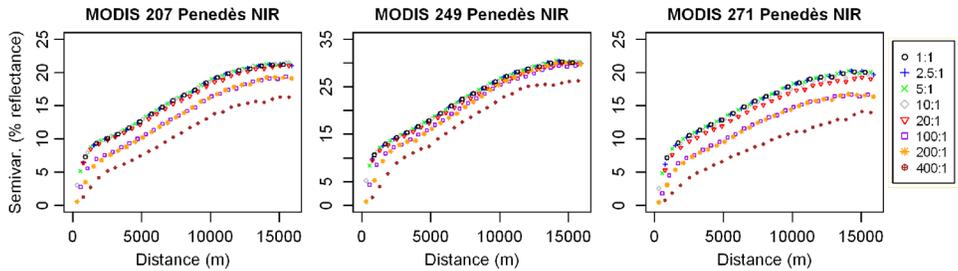


Fig. 14 Variograms of the 3d-DWT compression method for the Penedès region on three selected dates for MODIS images in the near infrared band.

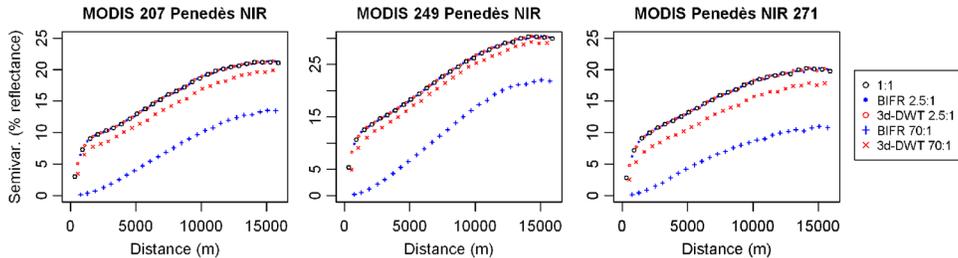


Fig. 15 Comparison of variograms for the BIFR and 3d-DWT methods for three selected dates in the Penedès region at representative compression ratios for MODIS images in the near infrared band.

This significant alteration (CR higher than 1:100) could be relevant in several geostatistical studies of remote sensing images, such as geostatistical applications referred in Sec. 1 of this work, and for example, in landscape scale classification based on variogram analysis.⁵² In this example, if someone uses compressed remote sensing images instead of original images, and this compression is not done with the appropriate parameters, it could significantly alter classification results and, consequently, its main goal of determining the optimal support size (i.e., optimal spatial resolution) for characterizing forest ecosystems.

3.2 Performance Results

All the variogram analyses were processed using the IBM cluster of the research laboratory of the Computer Architecture and Operating Systems Department at the Universitat Autònoma de Barcelona. The IBM cluster is formed by 32 nodes, each with two Dual Core Intel(R) Xeon (R) 3.0 GHz processors with 12 Gbyte of RAM, and communicated with an integrated dual gigabit ethernet. Table 8 shows the average time of 14 independent executions using a different number of workers, while Fig. 16 shows the graphical representation and corresponding speedup evaluation.⁵³

The proximity between the empirical speedup behavior and the theoretical linear speedup confirms that the parallel design and implementation provide a satisfactory solution for the computational problem at hand. These performance results demonstrate the validity of the proposed parallel solution, and the significant time reduction evidences the benefits of using distributed environments for processing large amounts of data, such as remotely sensed images. In fact, they provide an exhaustive test bed allowing running many executions with different parameters at

Table 8 Execution time and speedup using n (0 to 24) workers.

No. workers	Time (s)	Speedup
0	2975.03	
2	2112.33	1.41
4	1067.45	2.79
8	534.18	5.57
12	357.54	8.32
16	269.00	11.06
20	216.24	13.76
24	186.31	15.97

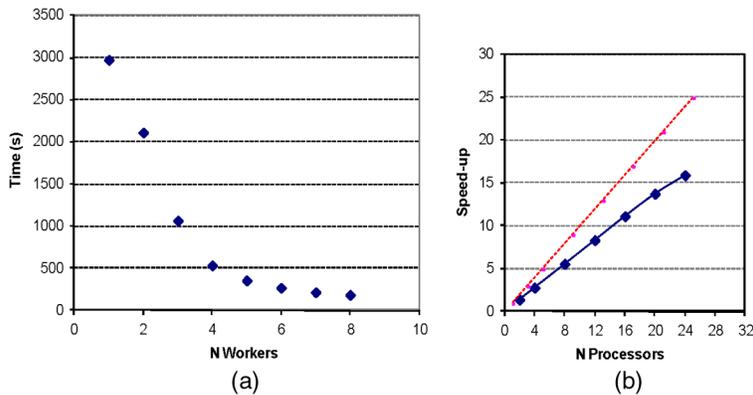


Fig. 16 Execution times (a) and speedup (b) with a different number of workers. The red dotted line corresponds to the theoretical linear speedup, while the blue line and points correspond to the empirical data.

several compression ratios in different regions on different dates, etc., and obtaining faster results and more agile comparisons.

4 Conclusions

The main conclusions of this study on the alterations in the spatial pattern produced by JPEG2000 compression methodologies applied to remote sensing images are drawn hereafter:

- Low compression ratios (less than 20:1 in the present study) maintain the radiometric image variability in all distance analyses, and therefore generate very similar variograms. This general behavior is slightly different for the 3d-DWT and BIFR methods: if the third dimension is large enough (CASI and MODIS) 3d-DWT is slightly more accurate to the noncompressed images than BIFR (less than a 0.5% reduction in variability).
- High compression ratios (over 20:1 in the present study) alter all spatial patterns of remote sensing images, but 3d-DWT is considerably better than BIFR for large three-dimensional images; 3d-DWT maintains about 25% more radiometric variability than BIFR.
- For Landsat images, seven spectral bands as a third spectral dimension are not enough to exploit three-dimensional compression, and therefore the BIFR method is preferable due to its simplicity.
- Despite the fact that pattern variability depends on the spectral band, all variogram alterations are very similar in all spectral bands and spatial directions.

Regarding to the computational methods developed in this work to reduce the execution time required for variogram analyses of remote sensing images:

- A distributed parallel solution based on a master/worker scheme and using MPI as the message-passing library obtains an efficient computing performance and provides a suitable environment for carrying out exhaustive analyses with very different compression ratios.

Acknowledgments

This work was partially supported by the Spanish government, by FEDER, and by the Catalan government, under Grants TIN2009-14426-C02-02, TIN2007-64974, PTA2010-3619-I, and SGR2009-1511. The authors are especially grateful to the Institut Cartogràfic de Catalunya, ICC, for providing the CASI data. Xavier Pons is a recipient of an ICREA Acadèmia Excellence in Research grant (2011 to 2015). Finally, the authors also wish to thank to the anonymous reviewers for their constructive comments that helped to improve the manuscript.

References

1. J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, 3rd ed., Prentice Hall Series in Geographic Information Science (2005).
2. P. Serra and X. Pons, "Monitoring farmers decisions on Mediterranean irrigated crops using satellite image time series," *Int. J. Rem. Sens.* **29**(8), 2293–2316 (2008), <http://dx.doi.org/10.1080/01431160701408444>.
3. P. Schneider and S. J. Hook, "Space observations of inland water bodies show rapid surface warming since 1985," *Geophys. Res. Lett.* **37**(22), L22405 (2010), <http://dx.doi.org/10.1029/2010GL045059>.
4. N. B. Chang et al., "Change detection of land use and land cover in an urban region with SPOT-5 images and partial Lanczos extreme learning machine," *J. Appl. Remote Sensing* **4**(1), 043551 (2010), <http://dx.doi.org/10.1117/1.3518096>.
5. K. Richter et al., "Plant growth monitoring and potential drought risk assessment by means of Earth observation data," *Int. J. Rem. Sens.* **29**(17), 4943–4960 (2008), <http://dx.doi.org/10.1080/01431160802036268>.

6. T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote Sensing and Image Interpretation*, 5th ed., John Wiley and Sons, New York (2004).
7. D. S. Taubman and M. W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards, and Practice*, pp. 778, Kluwer Academic Publishers, Dordrecht (2002).
8. T. Ranchin and L. Wald, "The wavelet transform for the analysis of remotely sensed images," *Int. J. Rem. Sens.* **14**(3), 615–619 (1993), <http://dx.doi.org/10.1080/01431169308904362>.
9. G. Martín et al., "Impact of JPEG2000 compression on endmember extraction and unmixing of remotely sensed hyperspectral data," *J. Appl. Rem. Sens.* **4**(1) 041796 (2010), <http://dx.doi.org/10.1117/1.3474975>.
10. A. Zabala and X. Pons, "Effects of lossy compression on remote sensing image classification of forest areas," *Int. J. Appl. Earth Observat. Geoinf.* **13**(1), 43–51 (2011), <http://dx.doi.org/10.1016/j.jag.2010.06.005>.
11. M. A. Chaudhry et al., "Optimization of wavelet bases for texture analysis," in *Proc. Fifth IEEE International Symposium on Signal Processing and Information Technology*, Athens, pp. 789–793 (2005).
12. Y. Li et al., "Adaptive compression of remote sensing stereo image pairs," *J. Appl. Rem. Sens.* **4**(1), 041777 (2010), <http://dx.doi.org/10.1117/1.3495716>.
13. A. Zabala et al., "Implications of JPEG2000 lossy compression on multiple regression modeling," *Proc. SPIE* **6749**, 674918 (2007), <http://dx.doi.org/10.1117/12.738028>.
14. M. Kiefner and M. Hahn, "Image compression versus matching accuracy," in *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIII, Part B2, pp. 316–323 (2000).
15. P. J. Curran and P. M. Atkinson, "Geostatistics and remote sensing," *Prog. Phys. Geogr.* **22**(1), 61–78 (1998), <http://dx.doi.org/10.1177/030913339802200103>.
16. C. S. A. Wallace, J. M. Watts, and S. R. Yool, "Characterizing the spatial structure of vegetation communities in the Mojave Desert using geostatistical techniques," *Comput. Geosci.* **26**(4), 397–410 (2000), [http://dx.doi.org/10.1016/S0098-3004\(99\)00120-X](http://dx.doi.org/10.1016/S0098-3004(99)00120-X).
17. R. G. Craig, "Autocorrelation in Landsat data," in *Proc. 13th International Symposium on Remote Sensing of Environment*, Ann Arbor, pp. 1517–1524 (1979).
18. P. M. Atkinson, E. Pardo-Igúzquiza, and M. Chica-Olmo, "Downscaling cokriging for super-resolution mapping of continua in remotely sensed images," *IEEE Trans. Geosci. Rem. Sens.* **46**(2), 573–580 (2008), <http://dx.doi.org/10.1109/TGRS.2007.909952>.
19. M. J. Pringle, M. Schmidt, and J. S. Muir, "Geostatistical interpolation of SLC-off Landsat ETM + images," *ISPRS J. Photogram. Rem. Sens.* **64**(6) 654–664 (2009), <http://dx.doi.org/10.1016/j.isprsjprs.2009.06.001>.
20. F. Chica-Olmo and M. Abarca-Hernandez, "Radiometric coregionalization of Landsat TM and SPOT HRV images," *Int. J. Rem. Sens.* **19**(5), 997–1005 (1998), <http://dx.doi.org/10.1080/014311698215838>.
21. M. A. Oliver, R. Webster, and K. Slocum, "Filtering SPOT imagery by kriging analysis," *Int. J. Rem. Sens.* **21**(4), 735–752 (2000), <http://dx.doi.org/10.1080/014311600210542>.
22. N. A. C. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, New York (1993).
23. A. J. Plaza and C. I. Chang, *High Performance Computing in Remote Sensing*, Chapman & Hall/CRC Computer, Boca Raton (2008).
24. A. Cortés, "Half-duplex dynamic data driven application system for forest fire spread prediction," in *Lecture Notes in Computer Science. High Performance Computing and Applications*, Vol. 5938, pp. 1–7, Springer (2010).
25. L. S. R. Froude, "Storm tracking with remote data and distributed computing," *Comput. Geosci.* **34**(11), 1621–1630 (2008), <http://dx.doi.org/10.1016/j.cageo.2007.11.004>.
26. J. Le Moigne, W. J. Campbell, and R. F. Crompt, "An automated parallel image registration technique based on the correlation of wavelet features," *IEEE Trans. Geosci. Rem. Sens.* **40**(8), 1849–1864 (2002), <http://dx.doi.org/10.1109/TGRS.2002.802501>.
27. M. K. Dhodhi et al., "D-ISODATA: a distributed algorithm for unsupervised classification of remotely sensed data on network of workstations," *J. Parallel Distributed Comput.* **59**(2), 280–301 (1999), <http://dx.doi.org/10.1006/jpdc.1999.1573>.

28. S. Wang et al., "Simple grid toolkit: enabling geosciences gateways to cyberinfrastructure," *Comput. Geosci.* **35**(12), 2283–2294 (2009), <http://dx.doi.org/10.1016/j.cageo.2009.05.002>.
29. National Aeronautics and Space Administration (NASA), Landsat Program <http://landsat.gsfc.nasa.gov>.
30. A. C. Newton et al., "Remote sensing and the future of landscape ecology," *Prog. Phys. Geogr.* **33**(4), 528–546 (2009), <http://dx.doi.org/10.1177/0309133309346882>.
31. X. Pons et al., "Ten years of local water resource management: integrating satellite remote sensing and geographical information systems," *Eur. J. Rem. Sens.* **45**(1), 317–332 (2012), <http://dx.doi.org/10.5721/EuJRS20124528>.
32. L. Martínez et al., "Atmospheric correction algorithm applied to CASI multi-height hyper-spectral imagery," in *Proc. RAQRS II Valencia*, pp. 170–173 (2006).
33. R. E. Wolfe, D. P. Roy, and E. Vermote, "MODIS land data storage, gridding, and compositing methodology: level 2 grid," *IEEE Trans. Geosci. Rem. Sens.* **36**(4), 1324–1338 (1998), <http://dx.doi.org/10.1109/36.701082>.
34. X. Pons, "MiraMon. Geographical information system and remote sensing software," Center for Ecological Research and Forestry Applications, CREAf, (2000), <http://www.creaf.uab.cat/MiraMon> (7 January 2013).
35. V. Palà and X. Pons, "Incorporation of relief into geometric corrections based on polynomials," *Photogram. Eng. Rem. Sens.* **61**(7), 935–944 (1995).
36. X. Pons and L. Solé-Sugrañes, "A simple radiometric correction model to improve automatic mapping of vegetation from multispectral satellite data," *Rem. Sens. Environ.* **48**(2), 191–204 (1994), [http://dx.doi.org/10.1016/0034-4257\(94\)90141-4](http://dx.doi.org/10.1016/0034-4257(94)90141-4).
37. <http://opengis.uab.cat/wms/satcat/index.htm> (7 January 2013).
38. A. Zabala et al., "JPEG2000 encoding of images with NODATA regions for remote sensing applications," *J. Appl. Rem. Sens.* **4**(1), 041793 (2010), <http://dx.doi.org/10.1117/1.3474978>.
39. R. Alamús, J. Talaya, and I. Colomina, "The SISA/0: ICC experiences in airborne sensor integration," in *Joint Workshop of ISPRS WG I/1, I/3 and IV/4*, Hannover (1999).
40. Warehouse Inventory Search Tool, National Aeronautics and Space Administration (NASA), <https://wist.echo.nasa.gov/wist-bin/api/ims.cgi?mode=MAINSRCH&JS=1> (26 May 2011).
41. <http://www.kakadusoftware.com> (7 January 2013).
42. JPEG2000 Part I: ISO/IEC/15444-1:2004, Part II: ISO/IEC/15444-2:2004, Part III: ISO/IEC/15444-3:2004, Geneva, Switzerland (2004).
43. F. Aulí-Llinàs et al., "J2K: introducing a novel JPEG 2000 coder," *Proc. SPIE* **5960**, 1763–1773 (2005), <http://dx.doi.org/10.1117/12.633226>.
44. P. Goovaerts, *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York (1997).
45. G. Matheron, *Traité de Géostatistique Appliquée*, Editions Technip, Paris (1962).
46. M. A. Oliver and R. Webster, "Kriging: a method of interpolation for geographical information systems," *Int. J. Geograph. Inf. Sci.* **4**(3), 313–332 (1990), <http://dx.doi.org/10.1080/02693799008941549>.
47. C. E. Woodcock, A. H. Strahler, and D. L. B. Jupp, "The use of variograms in remote sensing: I. Scene models and simulated images," *Rem. Sens. Environ.* **25**(3), 323–348 (1988), [http://dx.doi.org/10.1016/0034-4257\(88\)90108-3](http://dx.doi.org/10.1016/0034-4257(88)90108-3).
48. C. D. Lloyd, *Local Models for Spatial Analysis*, CRC Press, Boca Raton (2007).
49. L. Pesquer, A. Cortés, and X. Pons, "Parallel ordinary kriging interpolation incorporating automatic variogram fitting," *Comput. Geosci.* **37**(4), 464–473 (2011), <http://dx.doi.org/10.1016/j.cageo.2010.10.010>.
50. Anonymous, "Introduction to MPI," *Int. J. Supercomput. Appl. High Perform. Comput.* **8**(3–4), 169–173 (1994), <http://dx.doi.org/10.1177/109434209400800302>.
51. I. Foster, *Designing and Building Parallel Programs. Concepts and Tools for Parallel Software Engineering*, Addison-Wesley (1995).

52. P. Treitz and P. Howarth, "High spatial resolution remote sensing data for forest ecosystem classification: an examination of spatial scale," *Rem. Sens. Environ.* **72**(3), 268–289 (2000), [http://dx.doi.org/10.1016/S0034-4257\(99\)00098-X](http://dx.doi.org/10.1016/S0034-4257(99)00098-X).
53. M. Casas, R. Badia, and J. Labarta, "Automatic analysis of speedup of MPI applications," in *Proc. of the 22nd ACM International Conf. on Supercomputing*, Island of Kos, Greece, pp. 349–358 (2008).



Lluís Pesquer received his BS degree in physics in 1994 at the Universitat de Barcelona (UB) and a MS degree in computational science in 2009 at the Universitat Autònoma de Barcelona (UAB). He is researcher at CREAM since 2000 and he is currently the main developer of analysis tools, remote sensing modules and geodesy libraries of MiraMon GIS software. His main research interests are: spatial analysis, geostatistics, parallel computing in geoprocessing environments, remote sensing and applied geodesy to GIS.



Xavier Pons received his BS degree in biology in 1988, a MS degree in botany in 1990, a MS degree in geography in 1995, and a PhD degree in remote sensing and GIS in 1992, all from the Universitat Autònoma de Barcelona (UAB). His main work has been done in radiometric and geometric corrections of satellite imagery, in cartography of ecological and forest parameters from airborne sensors, in studies of the spectral response of Mediterranean vegetation and in GIS development, both in terms of data structure and organization and in terms of software writing. He has recently worked in descriptive climatology models, in modeling forest fire hazards and in analysis of landscape changes from long series of satellite images. He is Full professor at the Department of Geography of the Autonomous University of Barcelona and coordinates research activities in GIS and Remote Sensing.



Ana Cortés received both her first degree and her PhD in computer science from the Universitat Autònoma de Barcelona (UAB), Spain, in 1990 and 2000, respectively. She is currently associate professor of the Computer Architecture and Operating Systems Department at UAB. Her areas of interest are high performance computing, distributed and grid computing and also scheduling and load balancing in parallel systems. Her current research interest is focused on high performance computing applied to dynamic data driven applications systems for forest fire spread prediction.



Ivette Serral received her degrees in environmental sciences, MSc in remote sensing and GIS, both at the Universitat Autònoma de Barcelona (UAB), she is researcher at CREAM since 2005. The main objective of her work is to develop and manage GIS for the public administration: the Catalan marine and coastal information system, the Catalan healthcare information system, the Andorran environmental information system, etc. She developed new methodological tools in reference systems integration and multicriteria studies and Web geodata portals. She has been collaborating with secondary schools to disseminate GIS science and applications.