

Journal of Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Spatial and spectral pattern identification for the automatic selection of high-quality MODIS images

Lluís Pesquer
Cristina Domingo-Marimon
Xavier Pons

SPIE.

Lluís Pesquer, Cristina Domingo-Marimon, Xavier Pons, "Spatial and spectral pattern identification for the automatic selection of high-quality MODIS images," *J. Appl. Remote Sens.* **13**(1), 014510 (2019), doi: 10.1117/1.JRS.13.014510.

Spatial and spectral pattern identification for the automatic selection of high-quality MODIS images

Lluís Pesquer,^{a,*} Cristina Domingo-Marimon,^a and Xavier Pons^b

^aGrumets Research Group, CREAM, Edifici C, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain

^bGrumets Research Group, Dep. Geografia, Edifici B, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain

Abstract. Remote sensing is providing an increasing number of crucial data about Earth. Systematic revisitation time allows the analysis of long time series as well as imagery utilization in the most interesting moments. Nevertheless, the current huge amount of data makes essential the usage of automatic methods to select the best captures, as many of them are not useful because of clouds, shadows, etc. Because of that, one of the characteristics of the more recent missions is the distribution, along with the spectral data, of a large amount of quality ancillary datasets. These datasets can act synergistically in the aim of selecting the best quality images, but the criteria they provide are not always enough. Indeed, these datasets are often used on a per pixel basis and the spatial pattern of the different spectral bands is forgotten, so ignoring the key information they can provide for our goals. With this aim, our work takes one of the most successful instruments in remote sensing, MODIS, and demonstrates, through geostatistical techniques, that the role of the spatial patterns of the spectral bands can effectively improve image selection in a complex (for climate, relief, and vegetation and crop phenology) region of 63,700 km². The results show that band 01 (red) is the preferred one, as it achieves a 13% higher success than when only using quality bands criteria: a 94% global accuracy (66 true classifications, and only four omissions and one commission error). A second, important finding, is that the geostatistical selection improves results when using any band, except for band 02 (NIR1), which makes our proposal potentially useful for most remote sensing missions. Finally, the method can be executed in a reasonable computing time due to previously developed high-performance computing techniques. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 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.13.014510](https://doi.org/10.1117/1.JRS.13.014510)]

Keywords: spatial and spectral pattern; quality bands; MODIS surface reflectance; automatic variogram; time series.

Paper 180581 received Jul. 16, 2018; accepted for publication Jan. 4, 2019; published online Feb. 4, 2019.

1 Introduction

The increasing interest shown by the scientific community in using sensors from the Earth Observation System (EOS) satellites, and specifically MODIS (MODerate Resolution Imaging Spectroradiometer), mainly arises from the need to better understand global changes by monitoring environmental and ecological dynamics at a global scale and at high temporal resolution. In this framework, the free and accessible images provided by MODIS data constitute a large set of ancillary data, essential for many remote sensing (RS) studies. Indeed, in a similar way to this companion onboard the Aqua satellite, the MODIS sensor onboard the Terra satellite, has been collecting data since 2000 and currently provides a long time series of daily images that constitutes a big data framework, currently making it possible to progress in global monitoring.¹ However, and due to this huge amount of data, MODIS data collection poses an interesting challenge for achieving both accurate and automatic processing.

MODIS is a passive sensor with 36 spectral bands and covers an important portion of the electromagnetic spectrum, from visible to thermal infrared (0.45 to 14.385 μm) at medium

*Address all correspondence to Lluís Pesquer, E-mail: l.pesquer@creaf.uab.cat

spatial resolution (from 250 m to 1 km at nadir).² Nevertheless, as this dataset can contain some images with specific issues that could reduce the quality of the acquisitions, a substantial effort must be made, as said, to automatically produce high quality subsets. For instance, several atmospheric effects interfere with the remotely sensed signal. In particular, day and nighttime acquisitions are affected by uncontrollable cloud coverage or aerosols (pollution, dust, sea salt, etc.), which makes difficult the measurement of land surface conditions. Among these atmospheric effects, the presence of clouds, and their projected shadows over the Earth, is a major constraint on applications that use optical RS.³ Their automatic detection is still a challenge to be overcome, especially when automatic methods for analyzing large volumes of data are used, as in Ref. 4 and in the present work. Moreover, MODIS has a singular field of view of 110 deg that, combined with its distance to the ground, achieves a wide swath covering 2330 km (across track) in one scan. This large swath causes a panoramic distortion whereas partially overlapping scans produce data repetition.^{5,6} This feature, known as the bow-tie effect, is additionally exacerbated by Earth's curvature (bands with a spatial resolution of 250 m at nadir have 1207 m at the image edges, while 1 km resolution bands have 4816 m at the image edges⁷).

In addition, these data might also have some inaccuracies derived from the different levels of postacquisition processing, such as detector noise or data faults.² Consequently, it is necessary to select a high-quality dataset before massive analysis. Fortunately, many of these issues can be filtered out using the MODIS ancillary data, but not all of them. Faced with such a large amount of data, advancing in more complete, yet automatic, methods for selecting high-quality images becomes necessary.

A range of methods for partially mitigating these effects have been explored in various studies. For example, authors of Ref. 8 developed a by-pixel filtering algorithm to produce relatively cloud-free single-day snow covered area (SCA) products from MODIS imagery using the previous day's imagery to estimate current SCA. Authors of Ref. 9 used a wavelet-based method to remove contaminated data from NDVI time-series, generating much better temporal data, while authors of Ref. 10 used kernel methods (linear and nonlinear) for cloud masking and cloud removal to obtain cloud-free time series. In addition, authors of Ref. 11 presented an interesting aerosol retrieval method. However, the present work is not focused on overcoming one particular atmospheric effect, but rather aims to select a subset of high-quality images while preserving their original values.

The present research is based on a geostatistical approach and a tailor-made set of spatial pattern analysis tools. The spatial and spectral pattern analysis is the core of the presented methodology. Exploring, describing, and providing indicators of spatial variation in images are among the main geostatistical applications in RS.¹² Indeed, geostatistics can provide parameters for describing spatial patterns,¹³ modeling uncertainty in spatial analyses,¹⁴ measuring the spatial autocorrelation¹⁵ and radiometric coregionalization,¹⁶ and can also support downscaling techniques¹⁷ and classification.¹⁸ Authors of Ref. 19 presented a method based on variogram analysis that analyzed the spatial pattern of MODIS images using variogram tools to identify the images with higher quality, adopting MODIS band 03 (459 to 479 nm, blue) as the band that defines the structural parameters of the variogram. Such approach was successfully used in the generation of pseudo (radiometrically) invariant areas.²⁰ However, different studies^{21,22} demonstrated that the results of the RS's derived analyses could be spectral dependent and, consequently, the selection of the band to be used in the method should be studied in depth, particularly due to the influence of the data spatial pattern on the analyses.²³ Likewise, the use of some combination of bands (e.g., orthogonal transformations, tasseled cap, panchromatic fusion, etc.) in different RS applications can achieve higher quality results than just using pure original bands.²⁴ Accordingly, such types of transformations are also considered in the present work.

This paper explores the influence of the spatial pattern of different MODIS spectral bands (bands 01, 02, 03, 04, 06, and 07) and the first component of the principal component analysis (PCA)²⁵ of two subsets of spectral bands on the variability of the variogram. The result to be reached is a criterion that aims to identify the most suitable spectral band, or transformation of bands (in this case, the selected variable is the first component of the PCA) to spatially characterize a given geographical region. Two complementary targets are to increase the quality of the set of selected images and to provide additional validations that increase the robustness

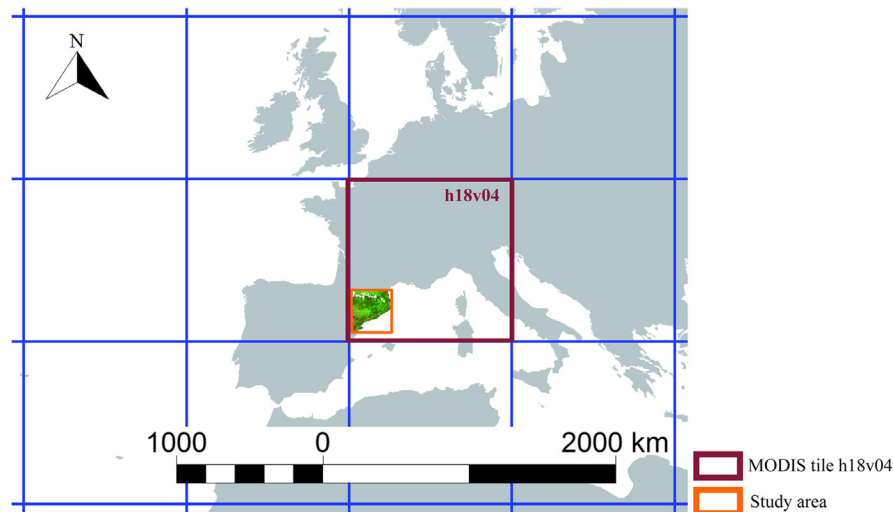


Fig. 1 Location of the study area.

and confidence on the assessed methodology. It is absolutely necessary to achieve this objective in an automatic and massive geoprocessing environment, implementing a chain of GIS and RS algorithms to manage huge series of RS data.

2 Data

2.1 Test Area

The presented method was tested in the NE of the Iberian Peninsula (SW Europe), in the area of Catalonia and neighboring regions, a 63,700 km² region covered by MODIS tile h18v04 (Fig. 1). The topography of the area stands out for its wide altitudinal range, from 0 m to more than 3000 m above sea level, configuring a heterogeneous landscape with several mountain ranges (including the pre-coastal mountain range, the coastal mountain range, and the Pyrenees). This area's interesting mosaicked land cover is composed of agriculture lands mixed with areas of natural and seminatural vegetation, including coniferous, sclerophyllous, and deciduous forests. These characteristics, together with its mainly mediterranean climate with a strong seasonality (snow cold winters with heavy rains and dry hot summers), make the selected area a perfect test region in terms of several types of potential problems that need to be solved when the spatial pattern variability of the variogram is analyzed.

2.2 MOD09GA Product Preprocessing

MOD09GA, the daily surface reflectance product, provides MODIS bands from 1 to 7 (459 to 479 nm (B03-blue), 545 to 565 nm (B04-green), 620 to 670 nm (B01-red), 841 to 876 nm [B02-NIR1 (near infrared)], 1230 to 1250 nm (B05-NIR2), 1628 to 1652 nm [B06-SWIR1 (short wave infrared)], and 2105 to 2155 nm (B07-SWIR2) at 500-m spatial resolution (nadir nominal spatial resolution of 463.31 m, 2413 m at the image edges) together with their quality assurance science data set (QA-SDS), which provides information about the quality of each pixel.² The MODIS land quality assessment (QA) products are essential for discarding potentially wrong pixel values and for applying high quality image selection filtering. However, the quality of these datasets could be affected by the presence of undetected poor quality pixels or artifacts corresponding to clouds and cloud shadows, imprecise snow coverage detection, atmospheric contamination, aerosols, technical problems of the acquisition, or algorithm problems.²⁶

A total of 364 images of the MOD09GA daily surface reflectance from 2009 (all bands except band 05 due to the stripe noise,²⁷ and an image from November 16, 2009 due to internal errors) were downloaded and used in this work. The original sinusoidal cartographic projection

was kept. Embedded QA-SDS were used to obtain meaningful image processing^{28,29} and data were masked accordingly. Pixels with clouds (state QA band bits 0-1 described as “cloudy,” “mixed,” or “not set”), cloud shadows (state QA band bit 2 described as “yes”), cirrus (state QA band bits 8 to 9 described as “small,” “average,” or “high”), and fire (state QA band bit 11 described as “no fire”) and snow (state QA band bit 12 and bit 15 with nonzero value) were discarded. Pixels with a sensor zenith angle of 35 deg or greater were removed in order to delete pixels subject to heavy geometric distortion and lower spatial resolution.³⁰ In addition, pixels classified with high aerosol categories (State QA band bits 6 to 7 described as “high,” “average,” or “climatology”) were also excluded because areas with high aerosol optical depth reduce the surface contrast and have a negative effect on surface reflectance.³¹ As another filter, only pixels classified as “corrected product at ideal quality” in the QC quality band (bits 0 to 1) were kept. Sea regions were also excluded for all images.

Moreover, to maintain informative power but to decrease computation efforts, it may also be useful to reduce the number of variables without losing too much information. Therefore, an orthogonal transformation (PCA) of the original MODIS bands (bands 01, 02, 03, 04, 06, and 07) was performed to generate a second type of variable, a combination set. The first principal component, which has the largest possible variance of the six spectral bands of the same acquisition, was included in the analysis and hereafter is referred to as PCA6. In addition, and to improve the results of this complete PCA combination, the first PCA component generated through only the two bands (bands 01 and 03) was also included and called PCA2. Band 01 and band 03 were selected because they were the ones that achieved the best results for the analysis on separated bands (no combination). In particular, as all bands were in reflectance units, it was considered more convenient not to apply a standardized PCA, but to run an unstandardized one. Altogether, the final dataset was formed by eight subsets (one per each individual spectral band and the two PCA combinations, PC6 and PC2) made up of 364 images.

3 Methodology

3.1 Geostatistical Approach: the Variogram Analysis

As previously mentioned, the QA-SDS is a very useful dataset that may have some limitations. A regionalized approach based on a geostatistical point of view has been developed in order to improve the quality of the information provided with the aim of complementing the pixel-by-pixel approach of the QA-SDS.

The geostatistical analysis was carried out following the image preprocessing flow described in Fig. 2. The method applied, the variogram analysis, plots the dependence of the spatial variance of a given variable (e.g., surface reflectance of a spectral band in a MODIS image) and therefore the spatial pattern can be analyzed.³² The empirical variogram (coming from samples, i.e., pixels, in the particular case of remotely sensed images) can be modeled and fitted by a continuous function to identify structural parameters (nugget, sill, and range) that characterize the spatial pattern for any quantitative variable distribution.³² The nugget is the variance near the origin (very short distances) and represents the component of the nonspatially correlated error. The sill represents the variance at the variogram saturation and the range is the spatial distance at which the variogram reaches this mentioned saturation, and represents the limit distance of autocorrelation. The concrete calculation of the nugget, sill, and range depends on the type of variogram function. This work used the exponential and the quadratic models (both with nugget effect) to obtain the fitted variogram and the corresponding structural parameters. The equations of these models are

$$\gamma(h) = c_0 + c_1 \left[1 - e^{-3\frac{h}{r}} \right] \quad \gamma(h) = \begin{cases} c_0 + c_1 \left[2\frac{h}{r} - \left(\frac{h}{r}\right)^2 \right] & h < r \\ c_0 + c_1 & h \geq r \end{cases}$$

Exponential variogram

Quadratic variogram

where c_0 is the nugget, c_1 is the sill, and r is the range.

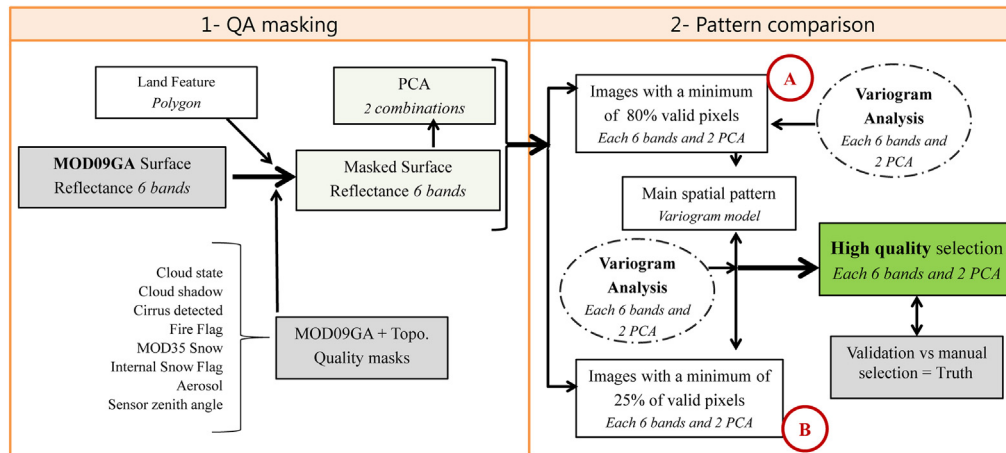


Fig. 2 Process workflow and its two steps: from the original MODIS 2009 daily series, to the high quality selected images. Groups A and B are defined according to the % of valid pixels (80% and 25%, respectively).

Therefore, these parameters characterize the spatial pattern of a concrete spectral band (or PCA combination) of a MODIS image. For instance, a homogeneous image has a higher range than a heterogeneous image. The variogram of a high quality image can characterize the landscape in a way that the variogram is spatially (in the extension of 63700 km² study region) and statistically invariant along time. The general spatial pattern differences among most of the images are caused by meteorological conditions (e.g., snow and clouds), geometrical distortions, or processing errors; seasonal phenology changes or urban growing are spectral band depending (e.g., vegetation) and less statistically relevant. For instance, an increase/decrease in the spatial pattern heterogeneity could be caused by the above-mentioned undetected poor quality pixels. Thus, the variogram analysis identifies the pixels in the latter case (such as an unusual increase in variability in some regions) and these images are filtered out (Fig. 2).

This geostatistical approach was applied using the eight subsets of MOD09GA spectral bands or PCA combinations, corresponding to each of the MODIS bands (bands 01, 02, 03, 04, 06, and 07), PCA6 and PCA2 as the variables to be modeled (Fig. 2). PCA6 uses all suitable bands and PCA2 is undefined at the current design step. PCA2 will be the combination of the two bands that will generate the lowest error in the individual assessment of the proposed methodology.

All images of these subsets were classified essentially in two groups (groups A and B) according to the number of valid pixels (pixels considered as of good quality according to MODIS QA-SDS and QC-SDS). The threshold used for group A was 80% valid pixels and at least 25% valid pixels for group B. Images with less than 25% valid pixels were discarded. It is important to note that the concrete values of the thresholds can be automatically adapted (in an iterative process) to any range of particular spatial patterns or landscapes of a given study region. It will basically depend on the presence and combination of possible sources of errors at the study region (detailed in the previous 2.2 MOD09GA product preprocessing section), outstanding as the main relevant factor the frequency of cloud cover events (Fig. 3). Additionally, these thresholds (the default and the range of possible values) could be customized for different purposes or requirements; for example, a lower value can be used when a higher time frequency data information is needed.

In the second step, group A images (more than 80% valid pixels) were geostatistically analyzed and, as a result, a representative variogram for each subset was obtained. This variogram is considered to be a model of the spatial pattern distribution of the study area and is the key for subsequent automatic rejections of images that might have different types of problems and distortions.

Next, group B images were also geostatistically analyzed and their variograms were compared with the representative variogram. The criterion for selecting an image of group B as having high quality was that the three variogram structural parameters (nugget, sill, and range)

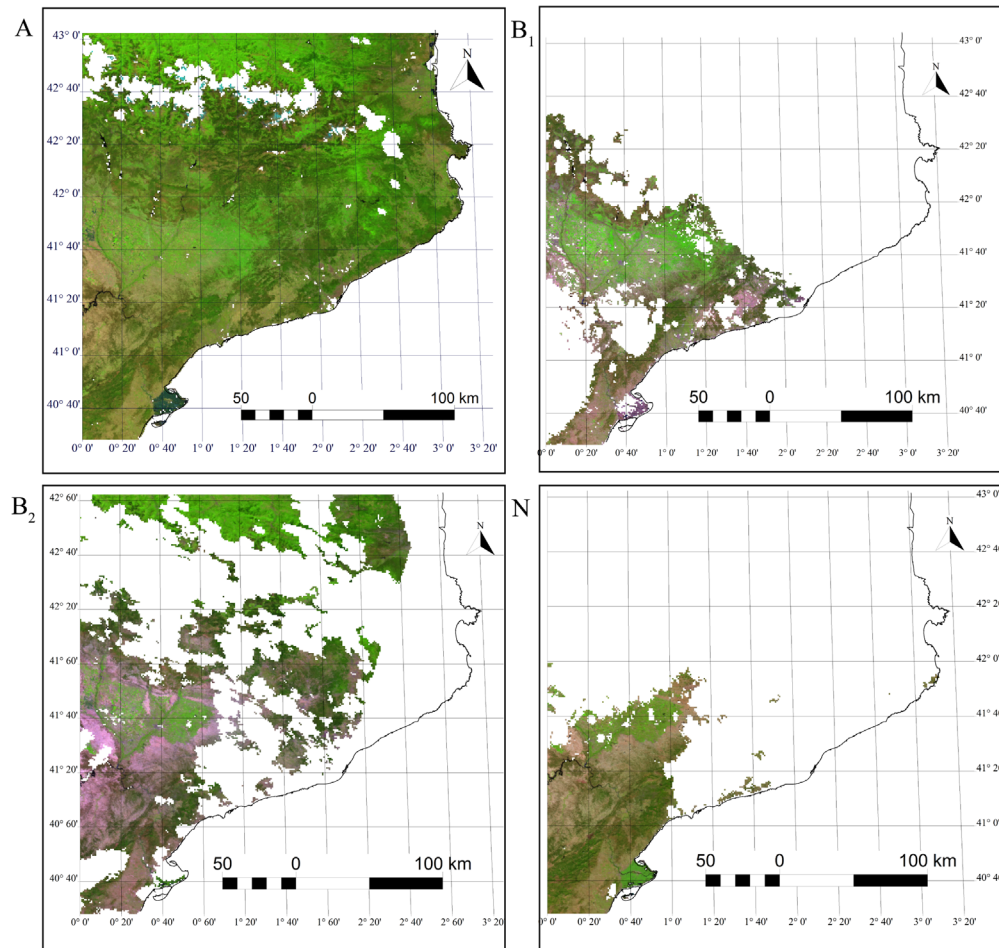


Fig. 3 (A) MOD09GA May 29, 2009 93.4% valid pixels (group A, higher than 80%); (B₁) MOD09GA April 20, 2009 26.8% valid pixels (selected in group B, higher than 25%); (B₂) MOD09GA June 28, 2009 44.5% valid pixels (nonselected, in group B); (N) MOD09GA September 4, 2009 19.6% valid pixels (nonselected: less than 25%). RGB compositions: 6,2,1. Note that B₂ presents more valid pixels than B₁ but will be automatically discarded according to the proposed procedure.

were within a range that is 2.75 times the standard deviation from the mean calculated from the representative variogram (Fig. 4).

The final selection for each band was validated by comparing it to an independent and non-automatic expert classification, quality controlled by an analyst, manually too. The validation was also performed by making a comparison with a set that had only been filtered by QA-SDS.

The free MiraMon GIS and RS software³³ was the main software used in this work. The authors developed auxiliary tools in MiraMon specifically addressed to fully automatic process long time series of RS images. These specialized geostatistical tools were built on the geostatistical libraries in Ref. 34 and the new developed options and adaptations for the present work are fully integrated in those libraries.

4 Results and Discussion

After masking the images (Fig. 2, step 1) only 7 out of 364 images had more than 80% valid pixels (group A) (MonthDay [MMDD]: 0506; 0529; 0718; 0725; 0817; 0819; 1013) (Table 1) and 71 images were classified in group B as images with at least 25% valid pixels. Figure 5 shows two examples of bands 01 and 02 of the resulting variograms for group A, which were geostatistically analyzed; for each subset, a representative variogram was obtained. As explained

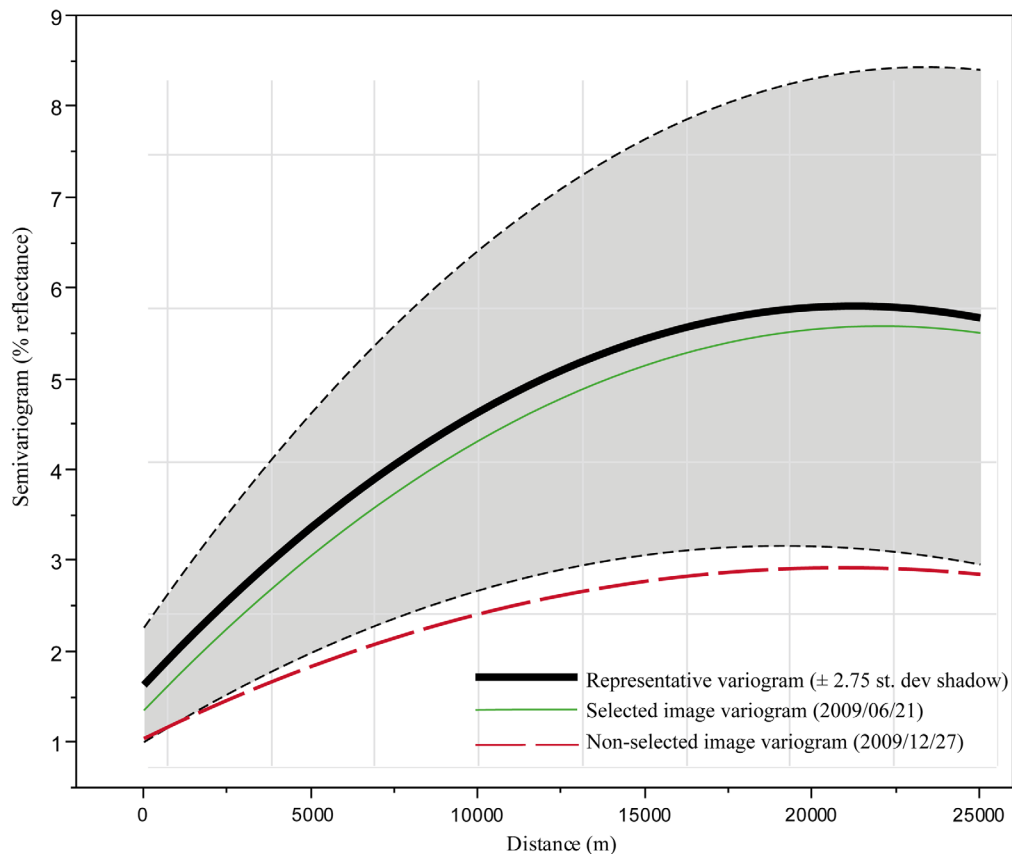


Fig. 4 Example of band 04 representative variogram (± 2.75 std. dev.) (black line and shadowed area), group B nonselected image variogram (red line) (December 27, 2009), and group B selected image variogram (green) (June 21, 2009). The variogram structure parameters of the selected image are within the valid parameter interval of the representative variogram.

above, this variogram is considered as a model of the spatial distribution pattern of the study area and is used for the automatic selection. Table 1 also shows the variogram structure parameters for each spectral band and for the two options of the PCA.

The expert and independent visual exploration classified a total of 59 of the 71 images of group B as high quality, and the remaining 12 images were considered to be poor quality images. That means that a simple filter based on the usage of MODIS QA-SDS quality masks disagreed in 12 images with regard to the expert manual classification. Therefore, using only MODIS QA-SDS for filtering data may result in the inclusions of 17% low quality data.

The variogram of each band of these 71 images was compared to the corresponding representative variogram (like in Fig. 4). Results are detailed in Fig. 6 and summarized in the global accuracy (%) and commission error (%) parameters, included in the two final rows. Global accuracy was computed as the ratio between the number of right classifications (considering two classes, high and poor quality images) and the total number of classifications (71); the commission error is the ratio between the number of wrongly selected images and the total number of images (12) not selected by the manual classification (truth). Although less important in our study, the omission error parameter was also included; this is the ratio of the number of wrongly nonselected images and the total number of images (59) selected by the manual classification. In most cases, the priority of the automatic selection is to achieve a low commission error on the quality filtering, thus this indicator could be as important as the global accuracy.

These results for the two main evaluation indicators show that band 01 (red) is the more appropriate to be used for the automatic selection because it achieves both the maximum global accuracy (92.96%) and the minimum commission error (8.33%). Band 03 (blue) and the PCA2 (composition of band 01 and band 03) achieve the same minimum commission error as band 01.

Table 1 Variogram parameters for images with at least 80% (group A) valid pixels, which will define the representative variogram. Values for each band correspond to the nugget, sill, and range. MM: month; DD: day; M: mean.

	MMDD	0506	0529	0718	0725	0817	0819	1013		
	<i>N</i> valid pixels	209547	238084	212591	216035	213553	237369	244990	$M - 2.75\sigma$	$M + 2.75\sigma$
B01	Nugget	2.28	2.14	3.03	2.88	2.13	2.37	2.25	1.51	3.37
	Sill	6.70	10.33	15.54	13.93	12.05	12.48	8.40	3.48	19.22
	Range	21996	22327	24815	24679	29119	25747	22981	18260	30787
B02	Nugget	8.94	9.13	7.55	7.42	5.52	6.94	11.82	3.06	13.31
	Sill	17.13	14.37	12.63	11.48	8.55	10.38	8.56	3.90	19.84
	Range	21486	21460	20871	20507	21563	21205	18379	17916	23647
B03	Nugget	0.75	0.95	0.88	0.80	0.60	0.68	0.68	0.45	1.08
	Sill	1.23	2.51	2.17	2.02	2.16	2.05	1.56	0.87	3.04
	Range	20028	20691	21089	21238	23116	21721	21225	18859	23744
B04	Nugget	1.53	1.78	1.92	1.80	1.27	1.48	1.70	1.07	2.21
	Sill	2.51	4.46	5.11	4.72	4.25	4.31	3.40	2.23	6.10
	Range	20517	20458	21094	21202	22857	21688	20818	19115	23351
B06	Nugget	7.86	7.46	11.17	11.72	10.65	11.23	14.18	4.72	16.50
	Sill	13.73	14.75	28.56	27.57	27.96	25.89	21.37	6.69	38.98
	Range	19861	19753	20511	20155	21207	20484	19474	18797	21704
B07	Nugget	6.93	6.19	9.41	9.22	8.45	9.30	9.03	5.09	11.64
	Sill	12.72	13.01	26.51	24.12	24.62	24.76	19.11	5.90	35.49
	Range	20382	20092	20511	20259	21325	20496	20098	19379	21525
PCA6	Nugget	20.49	19.36	27.80	28.16	23.74	26.08	32.90	13.50	37.51
	Sill	41.77	41.83	76.72	71.49	70.36	67.92	54.07	23.45	97.73
	Range	20164	19797	20591	20357	21538	20680	19741	18846	21974
PCA2	Nugget	3.69	4.09	5.14	4.80	3.43	3.99	3.73	2.54	5.71
	Sill	6.85	11.05	15.18	13.69	12.18	12.50	8.60	4.09	18.79
	Range	20759	20847	21453	21481	22964	21836	21103	19578	23406

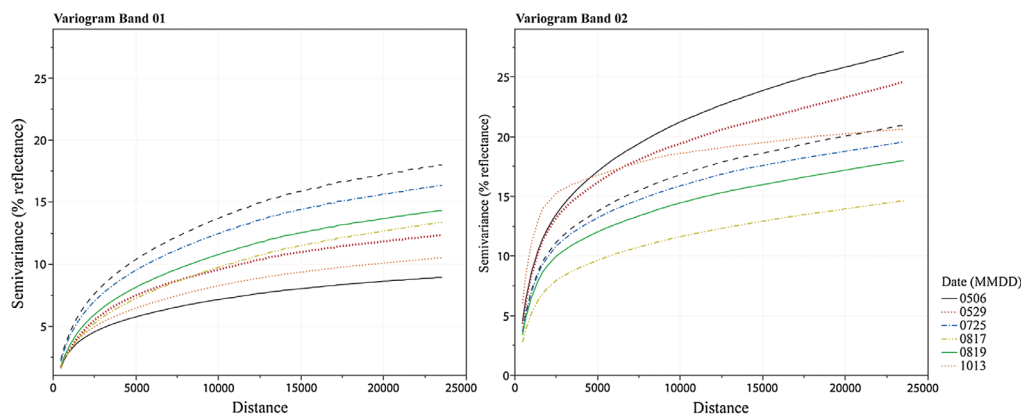


Fig. 5 Example of band 01 and band 02 variogram plots for selected images with at least 80% valid pixels (group A). MM: month; DD: day.

MMDD	Manual Selection (Truth)	Mask filter	B 01 Red	B 02 NIR1	B 03 Blue	B 04 Green	B 06 SWIR1	B 07 SWIR2	PCA 6bands	PCA 2bands
0121						N				
0125										
0204										
0213				R		R	R	R	R	
0217										
0224							R	R	r	r
0226										
0308				R						
0310				R		R		R	R	
0312				R						
0314						R			R	
0317										
0319				R						
0321										
0326				R						
0420				R,S						
0422				R,S						
0427										
0429										
0504				R,S			R	R	R	
0506										
0515				R						
0520										
0527				R						
0529										
0531										
0602				R						
0612										
0618										
0621										
0623										
0625										
0628										

Fig. 6 Detail of the automatic selection of high quality images for each band or band composite subset. Images selected are represented by filled cells, in gray for the truth column, in green when they were successful in the geostatistical pattern comparison, and in red for the wrong results. Images correctly nonselected are empty cells with a green cross-line, while wrongly nonselected cases are indicated with a red letter in relation to the variogram parameter (S, R, or N) at fault. Capitals indicate that the parameter is higher than the model while lowercase letters indicate that the parameter is smaller than the model (out of the interval). Dates in bold belong to group A.

This error is a result of considering image 0427 as a high quality image while the trained expert considered it as poor quality.

It is important to explain that all bands and band combinations, except band 02 (NIR1), achieve better results in terms of global accuracy than when only the MODIS mask filter is used (83.10% global accuracy of the QA-SDS method). Moreover, all bands, without exception, obtain lower commission errors than the mask filter.

Exploring the set of images, we found that, except for band 02 (NIR1), most errors were concentrated in a few dates: 0427 and 0520 are the scenes that increase the commission error for most bands, while 0716, 0730, and 0812 are the images that commit omission errors for most of bands. Figure 6 also shows that most of the not-selected images by the automatic method and in disagreement with the trained expert, were omitted due to a range (R) parameter of the variogram that was outside the model interval, either due to having a lower (few cases) or higher value (most of the cases) than the model. Therefore, this spatial autocorrelation parameter, the range, is mainly affected by the quality of the images.

Concerning the combination of bands, although the two PCA sets collected the main patterns of their corresponding images, this collection does not improve the best results obtained by some

0630				R						
0704										
0716			R	R	R		R	R	R	R
0718										
0720				R						
0723							R			
0725										
0727										
0730			R	R		R	R	R	R	R
0801										
0805										
0810										
0812			R		R,S	R	R	R	R	R
0815				R						
0817										
0819										
0821										
0826										
0828										
0831			R				R	R	r	R
0906										
0909										
0911								R		
0925										
0929										
1002								R		
1004										
1011										
1013										
1015										
1018								R		
1027				N,r	R			R		
1103								R		
1123				N,r						
1128				r						
1218										
1225						S				
1227				R		S				
N selected images	59	71	56	44	59	55	54	47	52	55
N success	N/A	59	66	46	65	61	60	57	60	65
N wrongly selected images	N/A	12	1	5	3	3	3	1	2	1
Omission error (%)	N/A	0.00	6.78	33.90	5.08	13.56	13.56	22.03	18.64	10.17
Commission error (%)	N/A	100.00	<u>8.33</u>	41.67	25.00	25.00	25.00	<u>8.33</u>	16.67	<u>8.33</u>
Global accuracy (%)	N/A	83.10	<u>92.96</u>	64.79	91.55	84.51	84.51	80.28	84.51	91.55

Fig. 6 (Continued)

of their single components: they collect most correct cases but also a few wrong cases. Thus, the single reflectance bands preserve better the spatial pattern of radiometric variability and the PCA combinations have less capacity to identify high quality images.

These results demonstrate that the spatial pattern characterized by each variogram is spectral dependent and landscape dependent because each variogram, and therefore, the pattern variability of the corresponding reflectance values, is different for each of the spectral bands (Fig. 5). The most accurate automatic selection of high quality images was obtained when the visible bands, i.e., red (band 01), blue (band 03), and the PCA2 (composition of bands 01 and 03), were used. Medium accuracy was achieved by the green band (band 04), both SWIR bands (06 and 07) and the PCA6 (composition of bands 01, 02, 03, 04, 06, and 07).

These results, clearly dependent on the spectral band, show that the interesting spatial pattern in this area was mostly influenced by vegetation, given that band 02 is the most sensitive to changes in phenology and vegetation water content.³⁵ In a mediterranean area such as Catalonia, where climate is characterized by strong seasonality and the landscape is fragmented, patches of vegetation that show significantly different phenological stages directly increment the variability in the spatial pattern, which is statistically reflected in the variogram parameters. As a result, the total sill is increased and the range decreased. Indeed, band 02 wrongly rejected images mainly in spring, the most active period for vegetation, when phenology events that are relevant for RS take place (mainly the emergence of leaves). PCA6 showed an intermediate accuracy; this decrease in accuracy was caused by the phenological variability noise introduced by bands 02 and 07 during its computation, and the better PCA2 results (without these “noisier” bands) confirm this point.

5 Conclusions

The presented geostatistical procedure was confirmed to be an excellent approach to automatically select high-quality images using a variogram analysis. In addition, this study confirmed that the spatial pattern characterized by the variogram is dependent on the spectral band. The highest accuracy of automatic selection is achieved using band 01 (93.0%) and band 03 (91.6%), corresponding to the red and blue visible spectrum bands, respectively. On the other hand, the least appropriated bands are band 02 (64.8%) and band 07 (80.3%), corresponding to the NIR1 and SWIR2 regions of the electromagnetic spectrum, respectively. These bands are the most sensitive to phenology and vegetation water content, both of which show strong seasonal variability that decreases the effectivity of the variogram analysis. Therefore, these results confirmed that the optimal band to be used in this geostatistical approach must follow the presented protocol for different study regions. This is especially recommended in those regions with strong seasonal vegetation changes where the methodology will automatically discard the bands presenting large anomalies regarding the global spatial pattern and it will appoint the band that will obtain a time series of high quality images.

Further analysis to reduce this seasonality component will be conducted in extreme dynamic landscapes, as it can be considered a potential limitation of this methodology; for instance, by applying the analysis at a 3-month time scale instead of at an annual time scale. In addition, we plan to extend the present spectral and spatial analysis to other satellite images; an especially interesting research will be a mixed time series obtained from different sensors: MODIS/MERIS/Sentinel-3 or other high frequency acquisition missions.

Finally, it was demonstrated that the approach is suitable for processing huge amounts of RS data, such as data provided by the MODIS Terra (or Aqua) full archive, with very low human dedication to the task and a modest computational effort.

Acknowledgments

This work has been partially funded by the Spanish MINECO Ministry through the ACAP research project (CGL2015-69888-P MINECO/FEDER), by the Catalan Government [SGR2017 1690] and by the ECOPotential research project. ECOPotential received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 641762. Xavier Pons is recipient of the ICREA Academia Excellence in Research Grant (2016–2020). We thank the anonymous reviewers of the previous version of the work, presented at the SPIE Europe Remote Sensing Conference 2017, for their careful reading of the manuscript and their many insightful comments and suggestions.

References

1. P. Xin and H. Sun, “A new approach for remote sensing image sample selection based on convex theory,” *Int. J. Online Eng.* **11**(4), 14–19 (2015).

2. E. F. Vermote, J. C. Roger, and J. P. Ray, "MOD09 (surface reflectance) user's guide, Version 1.4. MODIS land surface reflectance science computing facility," p. 34 (2015).
3. R. P. Armitage et al., "Probability of cloud-free observation conditions across Great Britain estimated using MODIS cloud mask," *Remote Sens. Lett.* **4**(5), 427–435 (2013).
4. Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. Environ.* **118**, 83–94 (2012).
5. R. Ren, S. Guo, and L. Gu, "Fast bowtie effect elimination for MODIS L1B data," *J. China Univ. Posts Telecommun.* **17**(1), 120–126 (2010).
6. A. H. Souiri and A. Azizi, "Removing Bowtie phenomenon by correction of panoramic effect in MODIS imagery," *Int. J. Comput. Appl.* **68**(3), 12–16 (2013).
7. X. Pons and A. Arcalís, *Diccionari Terminològic De Teledetecció*, 1st ed., Institut Cartogràfic de Catalunya: Enciclopèdia Catalana, Barcelona (2012).
8. B. F. Morriss et al., "Persistence-based temporal filtering for MODIS snow products," *Remote Sens. Environ.* **175**, 130–137 (2016).
9. X. Lu et al., "Removal of noise by wavelet method to generate high quality temporal data of terrestrial MODIS products," *Photogramm. Eng. Remote Sens.* **73**(10), 1129–1139 (2007).
10. L. Gomez-Chova et al., "Cloud masking and removal in remote sensing image time series," *J. Appl. Remote Sens.* **11**, 015005 (2017).
11. G. Wen et al., "Improvement of MODIS aerosol retrievals near clouds," *J. Geophys. Res.-Atmos.* **118**(16), 9168–9181 (2013).
12. P. J. Curran and P. M. Atkinson, "Geostatistics and remote sensing," *Prog. Phys. Geog.* **22**(1), 61–78 (1998).
13. S. Garrigues, D. Allard, and F. Baret, "Using first- and second-order variograms for characterizing landscape spatial structures from remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.* **45**(6), 1823–1834 (2007).
14. J. You and Z. Pei, "Error modeling based on geostatistics for uncertainty analysis in crop mapping using Gaofen-1 multispectral imagery," *J. Appl. Remote Sens.* **9**(1), 097096 (2015).
15. R. G. Craig, "Autocorrelation in Landsat data," in *Proc. Int. Symp. on Remote Sens. of Environ.*, Vol. 3, pp. 1517–1524 (1979).
16. M. Chica-Olmo and F. Abarca-hernandez, "Radiometric coregionalization of Landsat TM and SPOT HRV images," *Int. J. Remote Sens.* **19**(5), 997–1005 (1998).
17. P. M. Atkinson, E. Pardo-Iguzquiza, and M. Chica-Olmo, "Downscaling cokriging for super-resolution mapping of continua in remotely sensed images," *IEEE Trans. Geosci. Remote Sens.* **46**(2), 573–580 (2008).
18. J. L. Gil-Yepes et al., "Description and validation of a new set of object-based temporal geostatistical features for land-use/land-cover change detection," *ISPRS J. Photogramm. Remote Sens.* **121**, 77–91 (2016).
19. L. Pesquer, C. Domingo-Marimon, and X. Pons, "A geostatistical approach for selecting the highest quality MODIS daily images," in *Pattern Recognition and Image Analysis*, J. M. Sanches, L. Micó, and J. S. Cardoso, Eds., pp. 608–615, Springer, Berlin, Heidelberg (2013).
20. X. Pons et al., "Automatic and improved radiometric correction of Landsat imagery using reference values from MODIS surface reflectance images," *Int. J. Appl. Earth Obs. Geoinf.* **33**, 243–254 (2014).
21. P. Mausel, W. Kramber, and J. Lee, "Optimum band selection for supervised classification of multispectral data," *Photogramm. Eng. Remote Sens.* **56**(1), 55–60 (1990).
22. R. R. Joshi et al., "Monitoring changes in land use land cover of Yamuna riverbed in Delhi: a multi-temporal analysis," *Int. J. Remote Sens.* **32**(24), 9547–9558 (2011).
23. B. Qiu, C. Zeng, and C. Chen, "Comparative spatio-spectral heterogeneity analysis using multispectral and hyperspectral airborne images," *Geo-spatial Inf. Sci.* **16**(2), 83–90 (2013).
24. S. Madonsela et al., "Remote sensing of species diversity using Landsat 8 spectral variables," *ISPRS J. Photogramm. Remote Sens.* **133**, 116–127 (2017).
25. Y. Guo et al., "Remote sensing image classifier based on PCA," *J. Inf. Comput. Sci.* **5**(1), 429–433 (2008).
26. D. P. Roy et al., "The MODIS land product quality assessment approach," *Remote Sens. Environ.* **83**(1–2), 62–76 (2002).

27. R. Wang et al., "Terra MODIS band 5 stripe noise detection and correction using MAP-based algorithm," in *Int. Conf. on Remote Sens., Environ. and Transp. Eng. (RSETE)*, pp. 8612–8615 (2011).
28. R. R. Colditz et al., "Land cover classification with coarse spatial resolution data to derive continuous and discrete maps for complex regions," *Remote Sens. Environ.* **115**(12), 3264–3275 (2011).
29. A. M. Wilson, B. Parmentier, and W. Jetz, "Systematic land cover bias in collection 5 MODIS cloud mask and derived products: a global overview," *Remote Sens. Environ.* **141**, 149–154 (2014).
30. W. Yang and L. Di, "An accurate and automated approach to georectification of HDF-EOS swath data," *Photogramm. Eng. Remote Sens.* **70**(4), 397–404 (2004).
31. K. Grogan and R. Fensholt, "Exploring patterns and effects of aerosol quantity flag anomalies in MODIS surface reflectance products in the tropics," *Remote Sens.* **5**(7), 3495–3515 (2013).
32. M. A. Oliver and R. Webster, "Kriging: a method of interpolation for geographical information systems," *Int. J. Geog. Inf. Syst.* **4**(3), 313–332 (1990).
33. X. Pons, *MiraMon. Geographical Information System and Remote Sensing Software*, Centre de Recerca Ecològica i Aplicacions Forestals (CREAF), Bellaterra, Spain (2017).
34. L. Pesquer, A. Cortés, and X. Pons, "Parallel ordinary kriging interpolation incorporating automatic variogram fitting," *Comput. Geosci.* **37**(4), 464–473 (2011).
35. D. S. Steven, T. J. Malthus, and F. Baret, "Towards standardization of vegetation indices," in *Remotely Sensed Data Characterization, Classification, and Accuracies*, Vol. **1**, P. S. Thenkabail, Ed., CRC Press, Boca Raton (2015).

Lluís Pesquer has received his PhD in geography from UAB, his MSc degree computational science from UAB, and his BS degree in physics from UB. He is a researcher with GRUMETS at CREAM. His main research lines are spatial analysis, geostatistical analysis for remote sensing, parallel solutions in HPC, and interoperability in ICT water. He is an adjunct lecturer at the Geography Department of UAB in remote sensing and GIS master (UAB) for degrees in geography, smart cities, and environmental sciences. He coordinated the H2020-WaterInnEU project and he is participating in the H2020-ECOPotential, H2020-SMURBS, and ESA-HR_LandCover_CCI projects.

Cristina Domingo-Marimon received her PhD in geography, her MSc degree in remote sensing and GIS and BS biology from UAB. She is a postdoctoral researcher with GRUMETS at CREAM. She is currently working on the H2020 ECOPotential (improving future ecosystem benefits through Earth observations) project, H2020 BestMap (behavioral, ecological, and socioeconomic tools for modelling agricultural policy), and REGIME-SHIFT (drought regime shifts and global warming: impacts, mechanisms, forecasts, and mitigation strategies). She is interested in the use of remote sensing as a tool for the analysis of natural and climatic patterns.

Xavier Pons received his BS degree in biology, his master's degree in botany and geography, and his PhD full professor in geography at UAB in 2005. An ICREA Acadèmia researcher, from 2011 to 2015 and from 2016 to 2020, he is the principal investigator (PI) for the GRUMETS GIS&RS research group and creator in 1994 of the MiraMon software (+200,000 users, 47 countries). His methodological and applied work in RS (geometry, radiometry, classification, etc.) and GIS (climate models, standards, etc.) has been widely cited (+7300). He is the director of the masters in RS&GIS since 1996 and PI or participant in ~110 competitive projects and contracts.