Proceedings of the 7th International Conference on the Application of Physical Modelling in Coastal and Port Engineering and Science (Coastlab18)

Santander, Spain, May 22-26, 2018

SHOREX: A NEW TOOL FOR AUTOMATIC AND MASSIVE EXTRACTION OF SHORELINES FROM LANDSAT AND SENTINEL 2 IMAGERY

JESÚS PALOMAR-VÁZQUEZ, JAIME ALMONACID-CABALLER, JOSEP ELISEU PARDO-PASCUAL, ELENA SÁNCHEZ-GARCÍA

1 GeoEnvironmental Cartography and Remote Sensing, Department of Cartographic Engineering, Geodesy and Photogrammetry, (CGAT-UPV) Universitat Politècnica de València, Spain - jmpalomav@upvnet.upv.es, jaime.almonacid.caballer@gmail.com, jepardo@cgf.upv.es, elena.sanchez.upv@gmail.com

ABSTRACT

A sub-pixel automated shoreline extraction workflow (SHOREX) from satellite images is presented. SHOREX was designed as a set of separated tools that have been integrated within a single python framework to be used even in production environments. The workflow can be separated into three phases: (1) downloading, (2) pre-processing and (3) processing, being the first step the most time consuming. For the downloading process, the usage of servers like Amazon S3 has been a great help. The internal structure of this server allows downloading both Sentinel-2 and Landsat-8 images separated by tiles or scenes and bands, rather than all bands in a single file. That involves higher flexibility and less time of process. The preprocessing phase consists of a range of tasks like conversion between image formats, clipping, TOA reflectance conversion, cloud filtering or band sub-pixel geo-referencing. Once all images are prepared, they are processed using SHOREX, being the result a point shapefile layer organized by date with some statistical attribute fields. This statistical information can be useful to detect potential anomalies or analyze special dynamics on the coast related with climate phenomena or direct human action.

KEWORDS: Beach monitoring, coastal management, shoreline subpixel detection, Landsat images, Sentinel images.

1 INTRODUCTION

Having an efficient method of coastal monitoring is a key issue for any government. The good or bad management of beaches and dune systems can influence directly on aspects such important as tourist development, environmental sustainability or the fight against climate change. In this way, shoreline line position can be used as a good indicator to evaluate the impact of storms (Pardo-Pascual, *et al.*, 2014) or coastal evolution (Almonacid-Caballer *et al.*, 2016). The best way to obtain quality shoreline positions is by means of high precision acquisition methods, like GPS geodesic receivers (Pardo-Pascual, *et al.*, 2005), or photogrammetric surveys (Sánchez-García *et al.*, 2017). These methods, despite of its obvious advantages, present a main drawback: they become very expensive to analyze long coastal sectors with high temporal resolution. Fortunately, the use of satellite imagery overcomes this fact, since satellite platforms cover wide areas of land with a relatively high temporal frequency. In this way, we can cite the case of Landsat platform, whose images, after being released in 2008 and with over 30 million images downloaded so far now, have contributed enormously to develop several fields of research, as natural resources management, forestry, ecology or climate change. Recently, in 2015, the European Spatial Agency (ESA) launched Sentinel program, which allows to obtain free of charge images that cover the whole Earth with 20 meters resolution (Sentinel-2). In this way, if both Landsat and Sentinel missions are joined, we would able to analyze long sectors of coast with high combined revisit time (16 days with Landsat and 10 or 5 days with Sentinel-2).

Once this data are at our disposal, having an inefficient method to download such a huge among of information might become a real problem in a processing workflow. Traditionally, although both USGS and ESA provide web platforms in order to download any satellite scene, to some needs they are not fast and flexible enough or not fully adapted to massive download. In this regards, some big technology companies offer services to store and manage satellite imagery apart from public institutions. This is the case of S3, the storage service from Amazon, that offers both Sentinel-2 (S2) and Landsat-8 (L8) images stored in a folder structure that allows to perform selective and massive download of image bands using any scripting language.

Although traditionally in the field of coastal studies, mid-resolution images have been considered as set of data not enough accurate because their spatial resolution (30 m for Landsat and 20 m for Sentinel-2), to take advantage of all the possibilities that these sort of images offer us, the Geo-Environmental Cartography and Remote Sensing research group (CGAT) from the Universitat Politècnica de València (UPV), Spain, has been developing a set of algorithms to extract the shoreline positons from Landsat and Sentinel images overcoming all previous cited disadvantages:

- Pixel resolution limitations. CGAT has been developing a set of algorithms at sub-pixel level that allow to obtain a more accurate shoreline position (Almonacid-Caballer, 2014, Ruiz et al. 2007).

- Massive downloading of images. CGAT has developed a management system thought to facilitate a flexible and fast downloading of images from S3 server.

The whole system has been named SHOREX (**SHOR**eline **EX**traction). It is not just a set of single pieces of software, but a real management system that allows us to face all phases within a production pipeline: downloading, format conversions, clipping, storing, georeferencing and shoreline extraction.

In sections 2 and 3 the methodology behind SHOREX and the program interface with its main input parameters are describe. In section 4, an application example is performed and analyzed. Finally, section 5 exposes the main conclusions.

2 METHODOLOGY

A sub-pixel automated shoreline extraction workflow (SHOREX) from mid-resolution satellite images is presented. SHOREX was born as a set of separated tools that have been integrated within a single python framework to become useful software even in production environments. The internal algorithm of SHOREX has been tested in previous studies on micro-tidal coasts (Pardo-Pascual et al, 2012), obtaining an accuracy (RMSE) around 5 m. This fact proves how powerful this method can be in beach evolution studies. The production pipeline (fig. 1) can be separated into three phases: (1) downloading, (2) pre-processing and (3) processing, being the first step the most time consuming. The process begins with the download of the images. In that sense, the usage of servers like Amazon S3 (AWS-S3) has been a great help. The internal structure of this server allows downloading both S2 and L8 images separated by tiles or scenes and bands, rather than all bands in a single file. That involves higher flexibility and less time of process. However, the new workflow is being adapted to process Landsat-5 and Landsat-7 images. The preprocessing phase consists of a range of tasks like conversion between image formats, clipping, TOA (top of atmosphere) reflectance conversion, cloud filtering or band sub-pixel georeferencing. To obtain better performance it is necessary to maintain a good directory structure, where input, intermediate and final data are stored. Once all images are prepared, they are processed using SHOREX, being the result a point layer in shapefile format organized by date, where some extra attribute fields containing statistical values have been added.

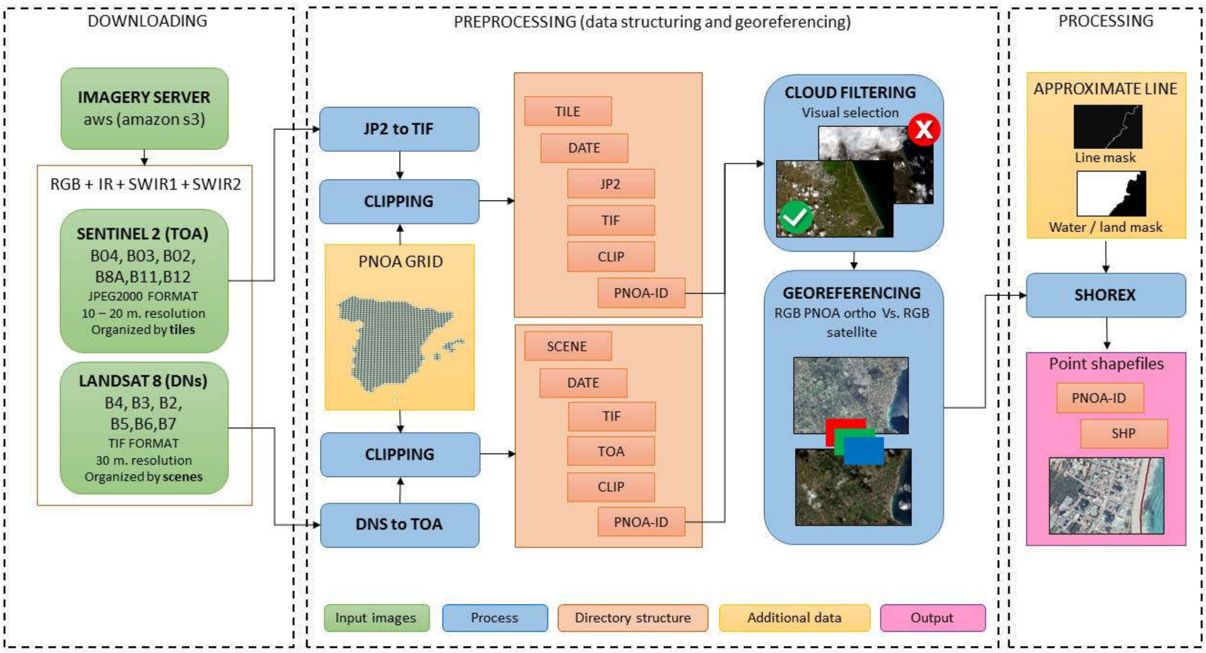

Figure 1. Production pipeline of SHOREX

2.1 Downloading

As we have seen previously, it is essential to have a fast and flexible way to download images. In our case, we take advantage of the storing structure that the Amazon S3 server provides, being the downloading process a bit easier for S2 images than for L8 images (figure 2).

- S2. In this case, the main advantage is the structure where these images are stored. This structure is organized using the military grid reference system (MGRS) where we can access directly to a specific band by means of using a regular URL link throughout any scripting language.

- L8. In this respect, we face two main drawbacks: (i) we have two datasets, collection and pre-collection, each one with a different way to be named, and (ii), we need more metadata to access directly to a URL link (in this case, we need the path, the row, the date of acquisition and the date of processing). Fortunately, AWS-S3 server offers two .csv files with all fields of information we need and its corresponding URL link to download each image. We have developed a set of tools to download these files and convert them into a database format in order to query them easily. The results are new text files with these URLs, used as inputs for other tool that download the images.

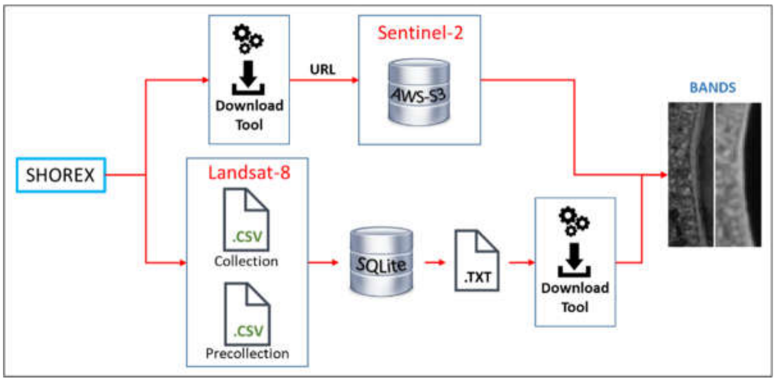

Figure 2. Python tool to download images from AWS-S3 server.

2.2 Pre-processing

The downloading process is not only in charge of getting each band we need, but also of storing these images into a consistent structure of folders, which will be used later to manage all this information more efficiently (figure 1). After downloading and storing, the pre-processing phase begins. It is necessary to adapt each band to every sub-process. In general and depending on the type of image (Landsat or Sentinel), these sub-processes consist in format and value transformations, clipping, cloud filtering and georeferencing.

- Format and value transformations. S2 images are downloaded in JPEG2000 format (.jp2) with pixels value in TOA, whereas L8 is in TIFF format (.tif), but its values are in DNs (digital numbers). To homogenize the final results and compare both type of images it is necessary to convert S2 from .jp2 to .tif format and also convert L8 values from DNs to TOA.
- Clipping. In order to better manage and organize such a big among of information, an homogeneous spatial division is proposed. In the case of Spain, we use each sheet the PNOA grid (Plan Nacional de Ortofotografía Aérea) as unit of processing. This helps to organize all products and sub-products obtained during the process and avoids possible problems (e.g. memory errors due to an excessive size of image). In this sense, SHOREX provides a master project in ArcMap where we can access to different layers of grids (PNOA, S2 and L8) and orthoimages (figure 3) helping to decide the type of information needed in every project.

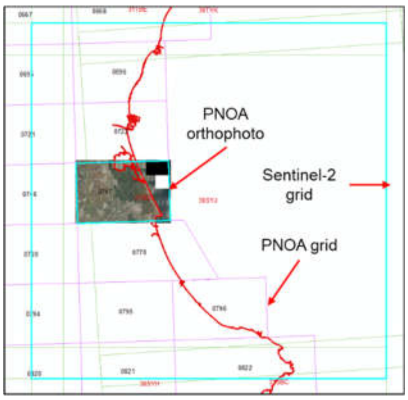

Figure 3. Spatial organization of SHOREX.

- Cloud filtering. Although we can filter cloudy S2 images using metadata files from AWS-S3 server, it is very possible that the distribution of clouds do not overlay with the analyzed area. Therefore, as general rule we download all available images, using later a visual tool to decide whether the image is useful or not. In figure 4 two examples of different situations are shown. In this example, the second image, although having lower percentage of cloud cover, is dismissed because a big part of the coast is cover by clouds. On the contrary, the first image that presents a similar level of cloud cover, they are located quite a distance of the analysis area, allowing the shoreline extraction without problems.

Figure 4. a) S2 image accepted (2016-4-1); b) S2 image dismissed (2016-4-11)

- Georeferencing. Once the cloud filtering has been performed, it is necessary to correct the images of possible georeferencing inaccuracies (Iron et al, 2012; Clerc, 2017). Despite both type of images reach a known accuracy in registration (0.4 pixels for L8 and 1.2 pixels for S2), the precision of this georreferencing may vary between sensors and scenes. Almonacid-Caballer 2017 et al 2017 demonstrates that sub-pixel phase correlation methods decrease the deviation of the registration to the tenth part of a pixel (3 and 2 m for Landsat and Sentinel respectively) improving the position of the obtained shorelines.Therefore, SHOREX uses this correlation method to georeferencing each band of each scene. That way, the RGB bands of each satellite scene and one PNOA orthophoto at the same spatial resolution are used for the registration. (figure 5).

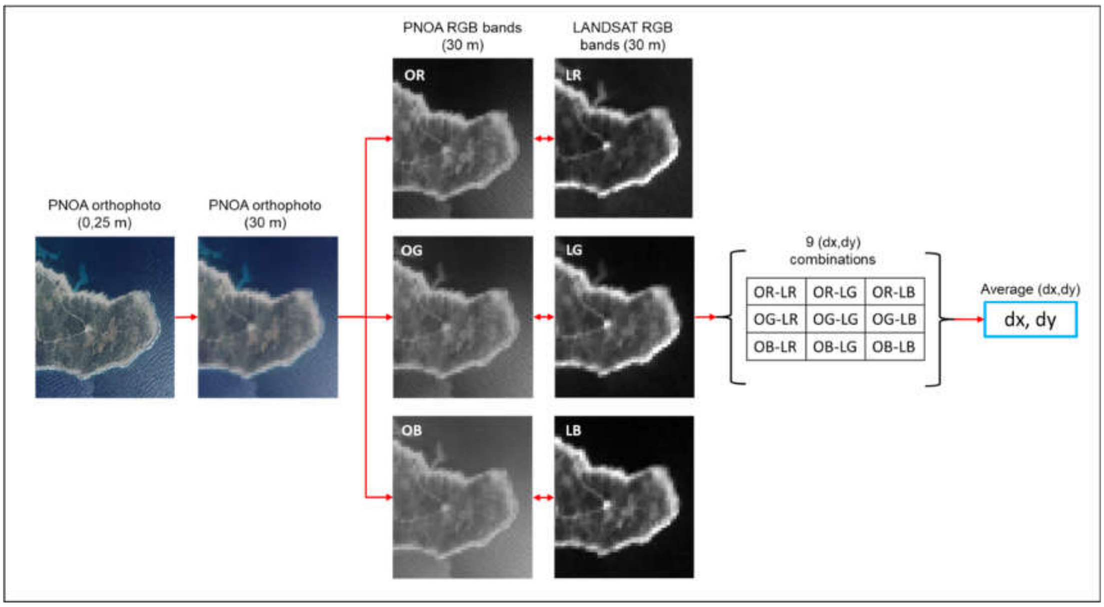

Figure 5. Georeferencing process. O: orthophoto. L: Landsat. R: Red channel. G: Green channel. B: Blue channel

2.3 Processing

All previous sub-processes are needed to prepare each band in a suitable way in order to be analyzed using the extraction algorithm of SHOREX. One of the first steps of this algorithm is to determine shoreline candidate pixels to be analyzed. In case there are not available approximate shorelines, SHOREX provides a semi-automated visual tool that uses thresholding segmentation and mathematical morphology techniques to obtain two masks, land/water mask and approximate shoreline mask. On the contrary, if we have an enough quality approximate shoreline, these previous mask can be obtained in an automate way, saving processing time. In the case of Spain, to obtain these mask, we use PNOA images (0.25 m pixel size) to digitalize approximate shorelines in ESRI shapefile format (.shp). These masks will be used to analyze only potential shoreline pixels, extract some useful statistical parameters and help to filter outliers points in the final result (figure 6).

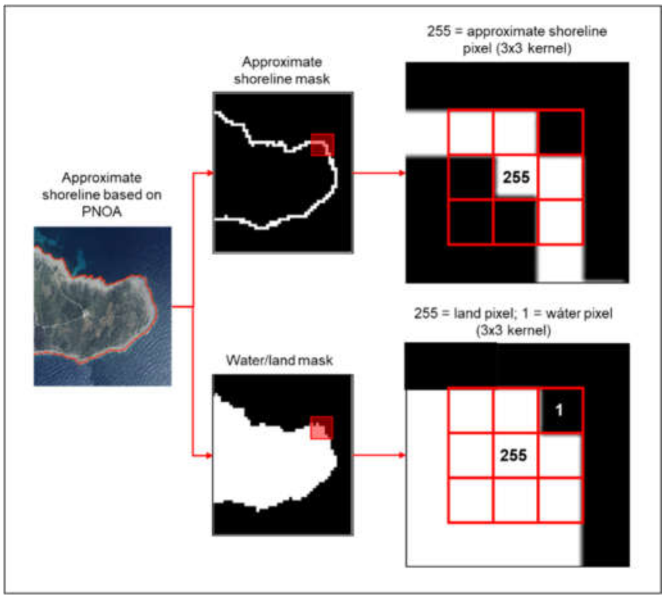

Figure 6. Approximate shoreline and water/land masks.

The central core of the SHOREX algorithm is based on taking a candidate pixel using a defined neighbourhood to analyze a surrounded area in order to determine a most accurate position between water and land areas. To obtain this position at sub-pixel level, the algorithm performs five steps: extraction of a sub-image form the original values within a kernel, sub-image resampling, surface polynomial function fitting, computation of the laplacian operator (Δ) from the previous polynomial function and finally, the extraction of the curve where laplacian value equals zero (Δ=0). This last condition indicates the inflection points in the mathematical fitted surface, and therefore, the potential sub-pixel shoreline position. The algorithm obtain a set of positions for every pixel analyzed that will be refined later in order to obtain a final average position (figure 7).

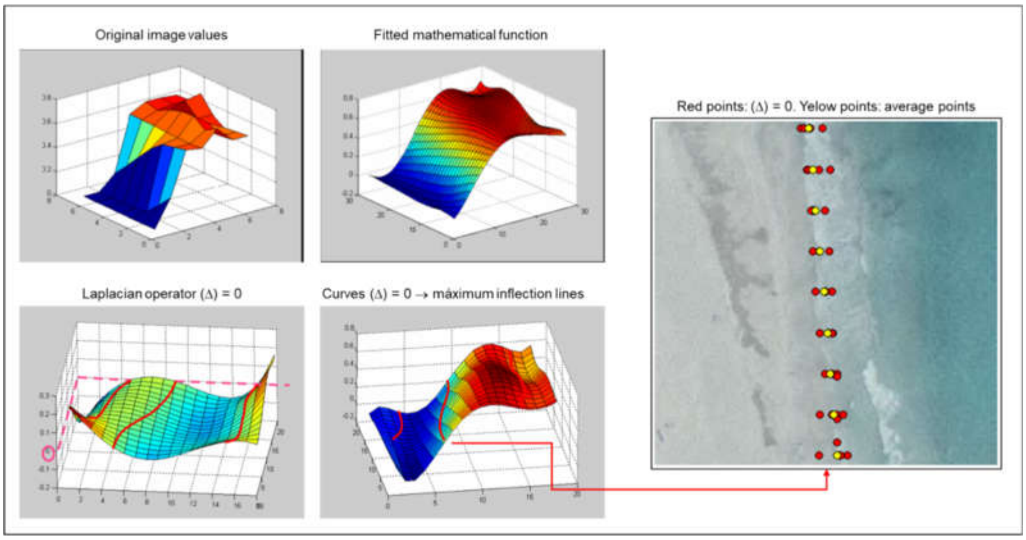

Figure 7. Maximum inflection lines detection and final average position.

The final part of the algorithm consists on enriching the information by adding some extra fields with statistical parameters that can help to analyze and interpret the results. In this sense, basic statistical parameters as average and standard deviation of pixel values for land and water areas, directional gradients, slope module or its direction are used to detect and assess problems in the algorithm or analyze coast areas with special features.

3 THE PROGRAM

This section describes the graphic user interface (GUI) for the automatic method to extract shoreline positions. Supposing we have a set of pre-processed bands and the masks with the approximate shoreline and the land/water segmentation, the program interface allows us to introduce several parameters to control the input data and the behavior of the algorithm. The GUI are divided in two sections: input files and parameters (figure 8).

- Input files. Water/land and approximate shoreline masks in TIFF format, the scene path that contains all bands to be analyzed, a file in .shp format to indicate de area of interest (if it is not used, the whole band will be analyzed).

- Parameters. The dates and bands to be processed, the type of image (Landsat or Sentinel), and finally, a set of parameters that directly control the behavior of the extraction algorithm: kernel size (depending on the nature of the coast), number of points per pixel (density of the final shoreline), degree used in the surface polynomial function (suitable to adapt different type of textures like vegetation-beach-water), cluster distance and minimum number of points in a cluster. The last two parameters are used to filter the points in order to obtain an average position.

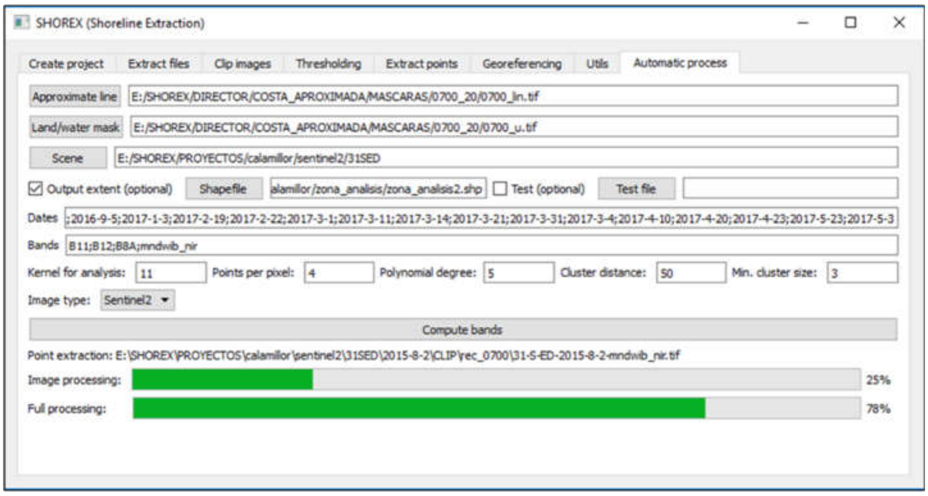

Figure 8. Graphic user interface of SHOREX.

4 APPLICATION EXAMPLE

A full-workflow processing has been carried out using SHOREX. The application example has been performed taken data from PNOA 0747 sheet, showing a coast sector of 22 kilometers long and sited in the southern part of the beach barrier of the Albufera de Valencia, Spain (figure 9). In order to measure the productivity of the SHOREX workflow, it all images available (L8 and S2) for a whole year (from 2016/09/01 to 2017/09/01) of PNOA 0747 sheet (18x28 km) have been processed. The number of available dates for both S2 and L8 images has been 81 and 23 respectively. From each date, RGB band combination and SWIR1 bands (B11 and B6 for S2 and L8 respectively) were downloaded. All these images were pre-processed and, after cloud filtering, the initial number of dates was reduced to 42 (28 for S2 and 14 for L8). The total processing time was 4 h and 34 min, being downloading phase the most consuming time. The information of the whole workflow has been summarized in tables 1 and 2. As we can see, the main idea we can extract from these tables is that image acquisition is the most time consuming process (83%), which enforce the fact that it is essential to dispose of an enough flexible and fast method to download these images.

Figure 9. Study area.

Table 1. Main descriptive information after processing by image.

Process		Sentinel-2 (30SYJ)	Landsat-8 (19933)
	Dates	81	23
	Bands	B02,B03,B04,B11	B2,B3,B4,B6,MTL
Donwloading	Band average size	100 Mb (RGB-10m) / 28 Mb (B11-20m)	60 Mb
	Files	324	115
	Time	2h 39' (include jp2-tif conversion)	1h 9' (include DNs-TOA)
Clipping	Time	4'	20''
Cloud filtering	Dates after filtering	28	14
Georeferencing	Time	5'	1'
Shoreline extraction	Time	28'	8'
Summary	Processing time	3h 16'	1h 18'

Table 2. Processing Summary.

Coast length		22,7 km
Number of dates		42
Number of files		439
	Total time	4h 34'
Processing time	Downloading	3h 48' (83%)
	Processing	46' 20'' (17%)

Regards to the analysis of the results, obviously, the study of the evolution of the shorelines over the time would be the first application of all this process. In that way, after comparing different shorelines in different dates, we can deduce the effects of a storm in the coast or compute the potential erosion or accumulation areas in long-term periods and try to figure out the cause (figure 10a). On the other hand, with regard to the shoreline files, in order to help to analyze the results and detect possible anomalies, a set of valuable extra information has been added to each file. In this sense, in figure 10b for instance, we can see how slope direction data can help to detect the main beach orientation to relate this information with other climate phenomena.

Figure 10. Extracted shorelines (10a) and main beach orientations (10b).

5 CONCLUSIONS

A methodology and full workflow to extract shorelines in massive way are presented. This workflow involves an important number of programs to perform the different processes. Among these programs, a deeply description of SHOREX, its algorithm and graphical user interface is showed.

In order to decrease the run-time extraction shoreline process and refine the results, different strategies have been taken in account: (i) having an initial approximate shoreline to teach the software where the potential shoreline pixels are, (ii) using the land/water mask to get some statistical values which can be used to analyze the results and detect errors, and (iii) designing a flexible system to download images. In this sense, the large amounts of images to be processed in a big project, reinforces the idea of the need of a flexible and fast system to download the information. In this sense, servers like AWS-S3, which has a public structure of storing, allows to download any single band of any single S2 or L8 scene using a scripting language. Despite of this, we can conclude that the images downloading is the main time-consuming process (83% in the application example).

The output data consist in a set of point layers in shapefile format that can be used directly to compare different dates to carry out both, long-term evolution studies detecting periods of erosion-accumulation phenomenon, or short-term effects, like those produced after storms. Besides, the extra statistical information added to these files can be useful to detect potential anomalies or deduce interesting features of the coast, like beach orientation, which can be used to be related with other climate phenomena.

SHOREX can be consider as a system designed to manage and process the extraction of shorelines from satellite imagery in a massive way. Within this system, the main time-consuming part is the data preparation, where process like downloading, transformation of formats, clipping, georeferencing, cloud filtering and approximate shoreline and mask creation are essential previous tasks to be consider before extracting the final shoreline.

ACKNOWLEDGEMENT

The authors appreciate the financial support provided by the Spanish Ministry of Economy and Conpetitiveness in the framework of project CGL2015-69906-R. This study is part of the PhD dissertation of the fourth author, which is supported by a grant from the Spanish Ministry of Education, Culture and Sports (I+D+i 2013-2016).

REFERENCES

J. Almonacid-Caballer, 2014. Extraction of shorelines with sub-pixel precision from Landsat images (TM, ETM+, OLI). Universitat Politècnica de València. doi:10.4995/Thesis/10251/48462.

J. Almonacid-Caballer, E. Sánchez-García, J.E. Pardo-Pascual, A. Balaguer-Beser, J. Palomar-Vázquez, 2016. Evaluation of annual mean shoreline position deduced from Landsat imagery as a mid-term coastal evolution indicator, *Marine Geology*, 372, 79-88.

J., Almonacid-Caballer,; J.E. Pardo-Pascual, L.A.; Ruiz, 2017. Evaluating Fourier Cross-Correlation Sub-Pixel Registration in Landsat Images, Remote Sens*ing*, 9, 1051.

S. Clerc, MPC Team. S2 MPC, 2017. Data quality report, reference S2-PDGS-MPC-DQR, issue 20. Available in https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2-702 Data-Quality-Report

J.R. Iron, J.L. Dwyer, J.A. Barsi, 2012. The next Landsat satellite: The Landsat Data Continuity Mission, Remote Sensing of Environment, 122, 11-21

J.E. Pardo-Pascual, L. García-Asenjo, J. Palomar-Vázquez, P. Garrigues-Talens. 2005. New methods and tools to analyze beach-dune system evolution using a Real-Time Kinematic Global Positioning System and Geographic Information Systems, *Journal of Coastal Research*, 49, 34-39.

J.E. Pardo-Pascual, J. Almonacid-Caballer, L.A. Ruiz, J. Palomar-Vázquez, 2012. Automatic extraction of shorelines from Landsat TM and ETM+ multi-temporal images with subpixel precision, *Remote Sensing of Environment*, 123, 1-11.

J.E. Pardo-Pascual, J. Almonacid-Caballer, L.A. Ruiz, J. Palomar-Vázquez, R. Rodrigo-Alemany. 2014. Evaluation of storm impact on sandy beaches of the Gulf of Valencia using Landsat imagery series, *Geomorphology*, 214, 388-401.

L.A. Ruiz, J.E. Pardo-Pascual, J. Almonacid-Caballer, B. Rodríquez. 2007. Coastline automated detection and multi-resolution evaluation using satellite images, *Proceedings of Coastal Zone,* 22 – 26 July, Portland, Oregon.

E. Sánchez-García, A. Balaguer-Beser, J.E. Pardo-Pascual. 2017. C-Pro: A coastal projector monitoring system using terrestrial photogrammetry with a geometric horizon constraint, *ISPRS Journal of Photogrammetry and Remote Sensing*, 128, 255-273.