Project Title:	ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS
Project number:	641762
Project Acronym:	ECOPOTENTIAL
Proposal full title:	IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS
Type:	Research and innovation actions
Work program topics addressed:	SC5-16-2014: "Making Earth Observation and Monitoring Data usable for ecosystem modelling and services"

Deliverable No: D4.1

EO data preprocessing and fusion modules

Version:	V1
Main Authors:	DELTARES, ISPRA, CERTH, CESBIO, DLR, CREAF, UAB

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

Contents

Co-funded by the European Union

1 Executive summary

The preprocessing of data is a crucial step in the remote sensing analytical workflow, and is often the most time consuming and costly. Examples of preprocessing tasks include geometrically correcting imagery to improve the positional accuracy, Lidar point cloud processing, spectral and spatial preprocessing are the main modules described in this deliverable.

In this document each modules present a short description and references to the methodology followed, the flowchart of the module specifying inputs/outputs need it. Also relevant results are presented to illustrate the module output.

Earth Observation data plays a major role in ECOPOTENTIAL, however filling the gaps, generated by the existence of clouds, rain, or simply due to incomplete track coverage of upper atmospheric layers, is one of the most common problems, faced during the processing of satellite data. Many methods have been tested out over the years for solving this problem, with different results, regarding the field of application and the expertise of the scientists involved. The approach tested and presented here is the Data Interpolating Empirical Orthogonal Functions (DINEOF) method.

One of the main goals of the ECOPOTENTIAL project is a better use of existing and incoming Earth Observation and field monitoring data, complemented by appropriate interpretation tools, data services and ecosystem models able to use these data. Today's data comes from multiple sources, which makes it difficult to link, match, cleanse and transform data across systems. However, it’s necessary to connect, find relationships and fusion data for a better understanding of the phenomena. Therefore this document describe a number of non-probabilistic data fusion methods, integration of in-situ data, integration of expert knowledge, integration of multisensory radiometry and a useful module to downscale waves.

Copernicus is a European system for monitoring the Earth and consists of a complex set of systems which collect data from multiple sources: earth observation satellites and in situ sensors such as ground stations, airborne and sea-borne sensors. ECOPOTENTIAL aims to benefit from this the services provide by Copernicus. Last section of this document describes different methods and techniques for assimilating data from INSPIRE spatial datasets and Copernicus core services/products such as integration of High Resolution Layers (HRL) of Copernicus Land Monitoring Services with Corine Land Cover (CLC) and national data sets and Copernicus Marine Environment Monitoring Service (CMEMS) products analysis for assessing marine food provision ecosystem services.

Co-funded by the European Union

1.1 Abbreviation and acronyms

ASAR	Advanced Synthetic Aperture Radar (ASAR)
CCI	Climate Change Iniziative
Chl-a	Chlorophyll - a
CLC	Corine Land Cover
CMEMS	Copernicus Marine Environment Monitoring Service
DO	Dissolved oxygen
DOW	Downscale Ocean Waves
DTM	Digital terrain Model
EMODnet	European Marine Observation and Data Network
ENVISAT	European Space Agency Environmental Satellite
EO	Earth Observation
EODESM	EO Data for Ecosystem Monitoring
EUNIS	European Nature Information System
EV	Essential Variable
FWHM	Full-Width-Half-Maximum
GEOWOW	GEOSS Interoperability for Weather, Ocean and Water
GOW	Global Ocean Waves
HRL	High resolution layers
INSPIRE	Infrastructure for Spatial Information in Europe
IPR	Intellectual Property Right
LiDAR	Light Detection And Ranging
LSMA	Linear Spectral Mixing Analysis
MDA	Maximum-dissimilarity algorithm
MIVIS	Multispectral infrared and Visible Imaging Spectrometer
MLLH	Maximum Likelihood Algorithm
PCA	Principal Component Analysis
RBF	Radial basis functions
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SST	Sea Surface Temperature
SWAN	Simulating Waves Nearshore
VLP	Virtual Laboratory Platform
XRD	X-Ray Diffraction

Co-funded by the European Union

1.2 Modules and tools attribution

Modules and Tools	**Authors**
Spatial preprocessing tools: clipping, mosaicking, georeferencing and projection change	CREAF-UAB
Radiometric correction tool	CREAF-UAB
Lidar point cloud processing	CESBIO
Speckle suppression fusing Sentinel 2-A RGB Optical data	CERTH
Image Time Series –Data Mining System	DLR
Gap Filling: Data Interpolating Empirical Orthogonal Functions	Deltares
Non probabilistic Fusion techniques	Deltares
Integration of in-situ data in Linear Spectral Mixing Analysis (LSMA) for the onshore coastal sediment characterization	ISPRA
Integration of expert knowledge, satellite multispectral and in-situ radiometry to map habitat types	ISPRA
LiDAR bathymetry integration with multisensory radiometry for seabed mapping in shallow waters	ISPRA
Maximum-dissimilarity algorithm (MDA) as fusion module to downscale waves on shallow waters from SAR wind fields, modeled wind and waves fields	ISPRA
Integration of High Resolution Layers (HRL) of Copernicus Land Monitoring Services with Corine Land Cover (CLC) and national data sets	ISPRA
Copernicus Marine Environment Monitoring Service (CMEMS) products analysis for assessing marine food provision ecosystem service ecopotential	ISPRA

Co-funded by the European Union

2 Introduction

Considering the project aims of addressing the data-to-information-to-decision-making process for ecosystem services and the subsequent objectives (i.e. making extensive use of Earth Observation data in combination with in situ monitoring; creating an Ecosystem Data Service related to the Copernicus space component – ECOPERNICUS-, providing a significant contribution to the research requirement for the Copernicus operational services and allowing simplified access to EO data and products and to ecosystem models outputs), this deliverable contributes – along with tasks 4.2, 4.3, 4.4 and 4.5 - to the achievement of the WP4 goal of designing and developing an EO Data for Ecosystem Monitoring (EODESM) that generates products, organizes and shares data for selected Protected Areas, models, and ecosystem and ecosystems services assessments.
To achieve this goal EO data preprocessing, integration and fusion tools/modules provided by involved partners were collected, in the view of providing comprehensive multi-source, multi-scale and multi-temporal EO products (services) for ecosystem monitoring to be distributed as an ECOPERNICUS service (WP10).

2.1 The overall framework that the preprocessing and fusion modules and tools will feed

- **EODESM (EO Data for Ecosystem Monitoring) (Figure 1):** a pre-operational multi-modular system that performs actions at Level 1, 2, 3 and 4 to quantify EBVs and other EV. EODESM tools/modules will:
  - use both pre-existing and new techniques;
  - be mainly focus on new EO services for ecosystem mapping/monitoring (retrieval bio-geophysical variables: terrestrial, marine and coastal; retrieval of land cover/use, ecosystem and habitats mapping), but will include also EO data pre-processing and integration modules;
  - be based on open source software and will represent a significant advancement of the state of the art.

- **Sandbox (Figure 2):** will enable access to EO data and products from existing archives, providing processing tools based on previous experience in GEOWOW (Developer Cloud Sandbox) for WP4 to achieve a comprehensive multisource, multi-scale and multi-temporal EO set for ecosystem monitoring. After testing and validation it will offer some EODESM modules developed in WP4 as processing services to WP10.

- **Virtual Laboratory Platform (VLP) (Figure 3):** VLP will build on the services published by ECOPERNICUS as well as others implemented via EODESM, providing in turn the virtual service platform for managing interdisciplinary research and applications.. Users will interact with the platform through a set of resource sharing services including: web-based data services, open archives, scientific models accessible as web services, semantic assets, and analytics services. The accessible resources are both provided by the project partners, in particular through the work carried out in WP3, WP4, WP5 and WP6 (such as models developed in WP6, or analytics services developed and made accessible by WP3), and external resources considered relevant for the ECOPOTENTIAL community. The platform provides four main service layers to empower the virtual laboratory capabilities:
  - discovery, access, and harmonized use of open EO and in-situ monitoring data, metadata, scientific models and results, semantic engines, and analytics' provision
  - discovery, access, and use of ecosystems modelling, including ecosystem services, upscaling and future protected areas
  - discovery, access, and use of the necessary knowledge to analyse ecosystems services as well as to obtain climate and land-use change scenarios, upscaling to larger areas and definition of the requirements of future protected areas
  - provision of specific applications and portal for different users.

- **ECOPERNICUS (Figure 4):** project activities will contribute to Copernicus and non-Copernicus contexts for ecosystems, and will create an Ecosystem Data Service for Copernicus (ECOPERNICUS), a new open-access, smart and user-friendly geospatial data and EO products retrieval portal and web coverage service using a dedicated online server.

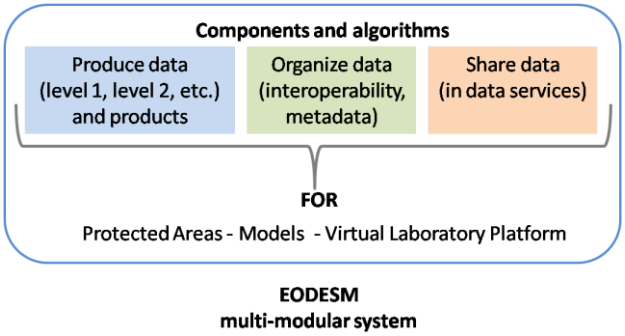


Figure 1 - EO Data for Ecosystem Monitoring

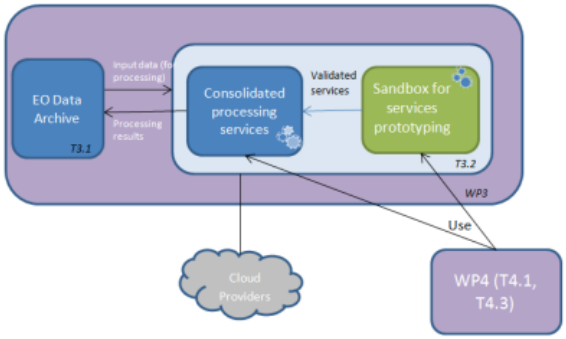


Figure 2.Scheme of the Sandbox approach for services prototyping and processing services for EO data

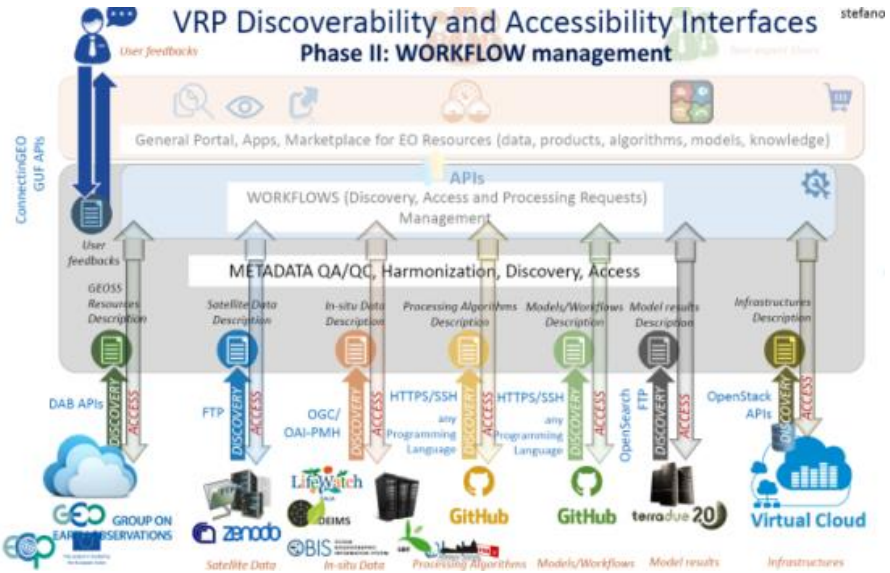


Figure 3. VLP/VRP as described in WP4-WP5 f2f meeting in Barcelona (April 2016)

Co-funded by the
European Union

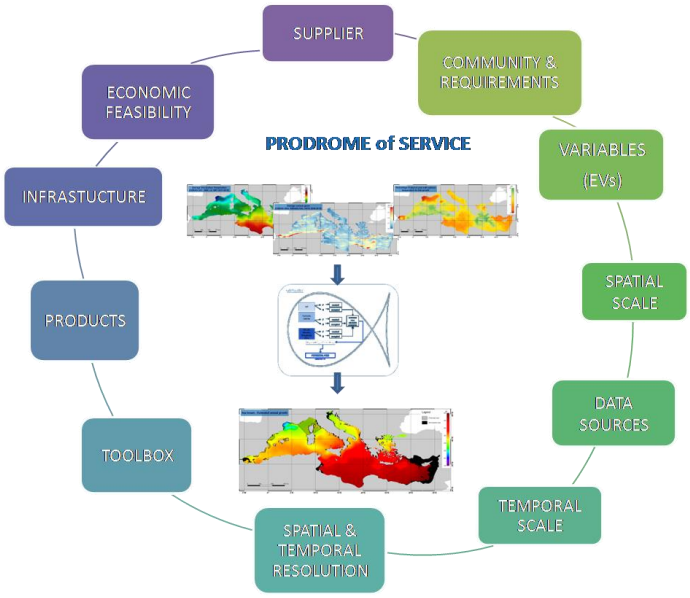


Figure 4. Example of the components of an Ecopernicus service for the assessment of ecosystem services

3 About the modules

3.1 Disclaimer

This deliverable is a collection of EO data integration tools and fusion modules in the form of workflows with the related description.

Each partner decided autonomously which tools and modules release to this deliverable and the degree of detail of each workflows description.

With regard to issues related to responsibility of the uses inconsistent with those stated in the project, each of contributing partner provided the module with the associated user license.

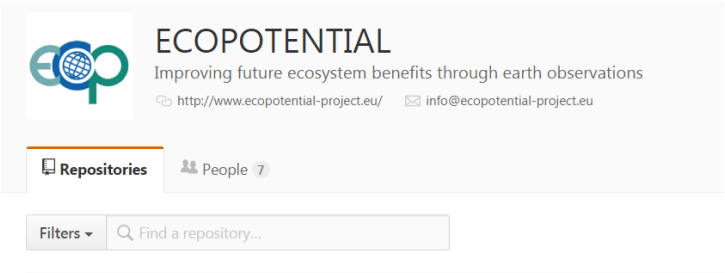
3.2 Restriction

By default, there will be no restriction on the use of the modules, unless explicitly stated otherwise by the partners.

3.3 Location

The latest working version of the modules will be uploaded to the following repository, unless the general ECOPOTENTIAL sharing policy change.

https://github.com/ec-ecopotential



3.4 Acknowledgements

We request that end users who make use of **EO data pre-processing and fusion modules** for deriving value added products, in written or oral presentations to add acknowledgment words of the use of EO data pre-processing and fusion modules therefore ECOPOTENTIAL project

Co-funded by the European Union

4 Preprocessing modules

Remote sensing images acquisition entails several disturbing factors due to different factors. From the sensor characteristics, atmosphere, weather, etc. to Earth surface geometry. Therefore a preparatory phase to improve image quality as a basis for further analysis is a must on the remote sensing processing chain. In this section, a set of preprocessing modules that are useful on the context of remote sensing data needed in the project will be presented. Modules are differenced among spatial preprocessing and spectral preprocessing. In the last part of the section, LiDAR point cloud processing and Speckle suppression fusing Sentinel 2-A RGB Optical data and gap filling will be discussed.

4.1 Spatial Preprocessing (CREAF-UAB)

In this section, we will explain spatial pre-processing techniques related to spatial data management which improves the positional accuracy and resolution, along with filtering techniques that serve to enhance or highlight the spatial characteristics of an image dataset. The topics are listed below.

4.1.1 Layer clipping

Remote sensing images from multiple sensors have different envelope and extension. It is thus sometimes necessary to extract a smaller defined area of the image. The module allows clipping by area defining X min X max Ymin and Ymax as an output enveloping coordinates, as well as using a vector mask, in which the external objects to the mask are eliminated. In this last case is also possible to indicate optionally a set of parameters defined by a selection.

The module is provided as part of the MiraMon Remote Sensing and GIS software available to all ECOPOTENTIAL partners any time on demand. Information regarding the inputs, parameters and outputs of the module are provided in the relevant Help of the software. MiraMon supports several parameters on the command line and allows writing Batch files that can run combining several chained orders. Moreover, combining Web Processing Service (WPS) standard and the ISO lineage model, MiraMon allows to easily edit and reproduce geospatial workflows described on the metadata, scaling them up to similar situations and executing them as remote processes.

4.1.1.1 References

[1] Pons, X. (2016). *MiraMon. Geographical information system and remote sensing software*. Centre de Recerca Ecològica i Aplicacions Forestals (CREAF). Retrieved from http://www.creaf.uab.es/miramon

4.1.2 Layer mosaicking

Given a set of images that cover a determined region sometimes it is necessary to mosaic them in order to form a larger image. This module generates a resulting layer by combining the original ones. The program can automatically determine the geographical extent of the set of layers or user can specify a given extent. It is also possible to specify a clip region through a vector mask. In this case the program examines each file to be mosaicked and used the useful part for the specified region. When a file is completely outside the region it is not considered. In case of superposition of rasters, if the metadata of preexisting files indicate that they contain NODATA, this value is not written in the destination file. This allows images that contain areas outside the region to be overlaid correctly. In the same way, when overlaying consists of values that are not background, the program DOES NOT makes sure that they are equal: the values of the last file that affects the region will be the

ones used. It is possible to specify the value of the NODATA of the output image. If it is not specified this will be the value of the NODATA of the first image. Original raster files must be coregistered.
The module is provided as part of the MiraMon Remote Sensing and GIS software available to all ECOPOTENTIAL partners any time on demand. Information regarding the inputs, parameters and outputs of the module are provided in the relevant Help of the software. MiraMon supports several parameters on the command line and allows writing Batch files that can run combining several chained orders. Moreover, combining Web Processing Service (WPS) standard and the ISO lineage model, MiraMon allows to easily edit and reproduce geospatial workflows described on the metadata, scaling them up to similar situations and executing them as remote processes.

References

[1] Pons, X. (2016). *MiraMon. Geographical information system and remote sensing software*. Centre de Recerca Ecològica i Aplicacions Forestals (CREAF). Retrieved from http://www.creaf.uab.es/miramon

4.1.3 Projecting Raster Data – Orthorectification (georeference)

Remotely sensed images present a series of geometric distortions produced by the rotation and curvature of the Earth, the irregularities of platform's orbit, etc. Moreover, sometimes datasets, although orthorectified, may also present undefined coordinate system. This module allows such images to be corrected to remove these distortions using conventional techniques based on first and second order polynomials. In case of remotely sensed images with a high level of detail or images of areas with a pronounced topography it is advisable to perform corrections taking into account the effect of the relief of the land surface. This module offers a specific method "polynomial" fit with heights" for satellite images with small pixels, such as Sentinel 2, Landsat, SPOT XS, etc. A set of ground control points which indicate a number of coordinate points in both the uncorrected original reference system and the final corrected reference system is needed. To increase the statistical reliability of the geometric correction process, the module allows two set of ground control points: one that is used to determine the transformation equations for the fit and the other, called *test points*, that is used to estimate the error of the fit based on an independent set of points. The criterion for deciding the output pixel value can be defined by user among the following three: nearest neighbor (will preserve the original radiometry), bilineal interpolation and bicubic interpolation. By orthorectifying an image, the distortions are geometrically removed, creating a planimetric image at every location with consistent scale across all parts of the image.

The images that have been corrected using information about the relief contained in a Digital Elevation Model (DEM) are known as orthoimages.

The module is provided as part of the MiraMon Remote Sensing and GIS software available to all ECOPOTENTIAL partners any time on demand. Information regarding the inputs, parameters and outputs of the module are provided in the relevant Help of the software. MiraMon supports several parameters on the command line and allows writing Batch files that can run combining several chained orders. Moreover, combining Web Processing Service (WPS) standard and the ISO lineage model, MiraMon allows to easily edit and reproduce geospatial workflows described on the metadata, scaling them up to similar situations and executing them as remote processes.

4.1.3.1 References

[1] Palà, V. & Pons, X. (1995) "Incorporation of relief into geometric corrections based on polynomials" Photogrammetric Engineering & Remote Sensing, 61(7):935-944.

[2] Pons, X. (2016). *MiraMon. Geographical information system and remote sensing software*. Centre de Recerca Ecològica i Aplicacions Forestals (CREAF). Retrieved from http://www.creaf.uab.es/miramon

4.1.4 Change Geographical reference system

Users that wish to change the projection system of an image once it has been georeferenced (do not confuse with the functions of the orthorectification that allows non-georeferenced layers to be georeferenced) will have to apply a transformation from one system to another setting the new projection parameters. The module can use different criterion to decide the output pixel value once the projection is changed: nearest neighbor (will preserve the original radiometry), bilineal interpolation (recommended for Digital Elevation Models) and bicubic interpolation (return smoothed results).

The module is provided as part of the MiraMon Remote Sensing and GIS software available to all ECOPOTENTIAL partners any time on demand. Information regarding the inputs, parameters and outputs of the module are provided in the relevant Help of the software. MiraMon supports several parameters on the command line and allows writing Batch files that can run combining several chained orders. Moreover, combining Web Processing Service (WPS) standard and the ISO lineage model, MiraMon allows to easily edit and reproduce geospatial workflows described on the metadata, scaling them up to similar situations and executing them as remote processes.

4.1.4.1 References

[1] Pons, X. (2016). *MiraMon. Geographical information system and remote sensing software*. Centre de Recerca Ecològica i Aplicacions Forestals (CREAF). Retrieved from http://www.creaf.uab.es/miramon

4.2 Spectral Pre-processing (CREAF-UAB)

4.2.1 Radiometric Correction

Radiometric correction is a prerequisite for generating high-quality scientific data, making it possible to discriminate between product artefacts and real changes in Earth processes. Image data recorder by satellite sensors can contain errors or interpretation problems due to, among other factors, the relative brightness over a pixel in a given band according to the solar position and characteristics of slope and exposure, and different atmospheric conditions between dates. Atmospheric correction is therefore recommended, and is particularly important in those cases where information that is going to be extracted will be primarily based on the spectral properties of the data, and/or adjacent or overlapping to scenes acquired under differing conditions. For instance, accurately production of change detection or land cover maps over a large area using multiple Sentinel images.

This module allows the radiometric correction (atmospheric and topographic corrections) of remote sensing images captured in the spectral regions visible and non-infrared. This correction allows expressing pixel values in reflectance units and reduces the number of undesired artifacts that are due to the effects of the atmosphere or to differential illumination which is, in turn, due to the time of the day, the location in the Earth and the relief (zones that are more illuminated than others, shadows, etc.). The radiometric correction model implemented in this module is based on two proposals by Pons *et al.* 1994 and Pons *et al.* 2014.
The model of radiometric correction takes several factors, among others, into account: relief (incidence angles and projected shadows), solar position, Earth-Sun distance in the moment of image capture, the decrease in radiation in its way down and up because of the atmosphere, and standard parameters for each channel such as atmospheric optical density, exoatmospheric solar irradiance and sensor calibration.

The topographic correction by default the module creates two temporary auxiliary images derived from the digital elevation model and the corresponding solar position at the moment of image acquisition with respect to surface slope, aspect and elevation: the first contains the cosine of the solar incidence angle for each pixel; the second contains for the solar azimuth with which the image was acquired, the solar heights causing shadows in each pixel.

The module can apply atmospheric correction based on two different approaches. The first one takes into account the subtraction of the lowest radiometric values for each channel –dark object subtraction (manual correction given that Kl – lowest value- has to be visually identified). The second approach performs an automatic correction by fitting tau and La parameters through the known reflectance pseudoinvariant areas (PIA) obtained from other remote sensors or field radiometers. The optical depth is obtained from MODTRAN simulations. This method has two essential advantages. On one hand the method guarantees radiometric homogeneity in large time-series and in the other hand it allows homogenous corrections among series of different sensors, such as for example Landsat and Sentinel-2.

Although both models work with incidence angles smaller than 90 degree, it is recommended and more realistic to assume that with angles that are close to this value the model cannot be assumed to be lambertian, being possible to indicate the angle from which the correction model will not be applied.
Sensors which are currently supported are: Sentinel 2, MSS (Landsat 1 - 5), TM (Landsat 4, 5), ETM+ (Landsat 7) [high and low gain; NLAPS and LPGS], OLI TIRS combined (Landsat 8), OLI (Landsat 8), HRV (SPOT 1, 2, 5), HRG (SPOT 5), ASTER VIS-NIR (Terra), ASTER SWIR (Terra), HRCC [High Resolution CCD Camera] (CBERS-1, CBERS-2, CBERS-2B).

The module is provided as part of the MiraMon Remote Sensing and GIS software available to all ECOPOTENTIAL partners any time on demand. Information regarding the inputs, parameters and outputs of the module are provided in the relevant Help of the software. MiraMon supports several parameters on the command line and allows writing Batch files that can run combining several chained orders. Moreover, combining Web Processing Service (WPS) standard and the ISO lineage model, MiraMon allows to easily edit and reproduce geospatial workflows described on the metadata, scaling them up to similar situations and executing them as remote processes.

4.2.1.1 References

[1] Pons, X. (2016). *MiraMon. Geographical information system and remote sensing software*. Centre de Recerca Ecològica i Aplicacions Forestals (CREAF). Retrieved from http://www.creaf.uab.es/miramon

[2] Pons X, Pesquer L, Cristóbal J, González-Guerrero Ò (2014) "Automatic and improved radiometric correction of Landsat imagery using reference values from MODIS surface reflectance images." International Journal of Applied Earth Observation and Geoinformation, 33:243-254.

[3] Pons X, Solé-Sugrañes L (1994) "A Simple Radiometric Correction Model to Improve Automatic Mapping of Vegetation from Multispectral Satellite Data." Remote Sensing of Environment, 48:191-204.

4.3 Lidar point cloud processing (CESBIO)

Lidar based retrieval of forest structural characteristics has become the method of choice over the past decade following the development of commercially viable airborne laser scanning (ALS) systems. Usually, ALS systems record echo pulses within small footprints (e.g., 15-30 cm) at average first-return pulse densities of 0.5 to 10 (sometimes more) p m^{-2} depending on sensor, flight characteristics and requirements. Proprietary software (i.e., sensor dependent) is usually used to extract discrete returns from raw ALS data. These returns are combined

with the navigation data to yield the geo-referenced point clouds provided by commercial operators. Accuracies of such procedures are of approximately 0.5 m horizontally and 0.15 m vertically, with higher accuracies within individual scans.

4.3.1 Description of the module

Subsequently, lidar-point-clouds are further processed by the beneficiary (user) through customized chains (Fig xy). For the purpose of Essential Variables generation over forested areas, the following steps were used within the ECOPOTENTIAL project. First, the data are examined for extent, point density, consistency, overlapping areas or gaps, and existing classification (if available) accuracy. Such issues are addressed in a second step by removing 'overlap' (e.g. points from adjacent flight lines) and 'noise' point classes to avoid deriving skewed metrics and speed up the classification process. In addition, the provided las tiles are checked for data gaps. Gap filling is achieved by merging data from adjacent flight lines (in overlapping areas) or from additional flight lines (if available). Open tools such as 'las2las' or 'lasmerge' from software suite Lastools are used to achieve such tasks. The next step is ground/non-ground classification, with all non-ground returns being considered vegetation in areas with no man-made features such as forests. Such classifications can be achieved with a range of algorithms (Montealegre et al. 2015). The open source Multiscale Curvature Classification (MCC) algorithm (Evans and Hudak 2007) was designed for forested environments on rough surfaces and was proved to produce the highest success rates at identifying ground and non-ground returns (Montealegre et al. 2015). Therefore, MCC algorithm is used to classify (or re-classify when the point cloud has been classified by the data provider) point cloud data. Depending on flight characteristics, ALS point densities and topography different *scale* and *curvature* parameters are used. Their values are determined for each PA by iterative tests starting from the default values.

With ground point correctly classified a digital elevation model is interpolated and used to compute the height above ground (i.e. normalized height) for the remaining points. The normalized point cloud is used to produce generic Lidar-based metrics (canopy closure, canopy density, percentiles, etc.). Such Lidar metrics represent proxies of forest structural characteristics and can be used with parametric (e.g., linear regression) or non-parametric (e.g., random forest, support vector machines) models to derive above ground forest biomass or other forest structural characteristics of interest. The Lidar metrics may be computed using Fusion and Fusion LTK (for large areas), open software developed and maintained by USDA Forest Service.

4.3.2 Flow chart

Once ALS–based layers (metrics) are generated post-processing tasks may include transformation to the desired file format (e.g. GeoTiff) and cartographic projection (e.g. UTM). For irregular (non-rectangular) ALS coverages one may need to subset the resulting raster files to eliminate areas interpolated over the missing tiles (Figure 6).

Co-funded by the European Union

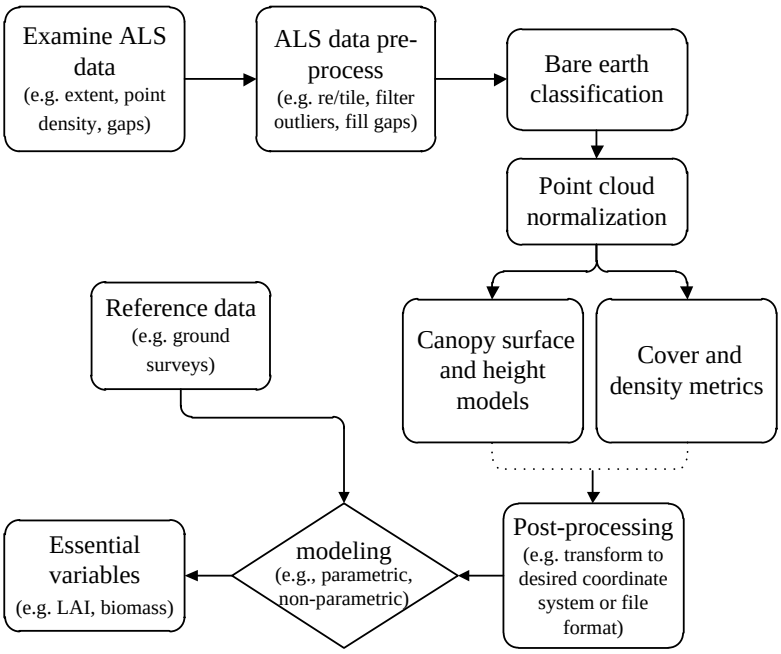


Figure 5 Typical flowchart for EV modelling from Lidar data (including ALS processing)

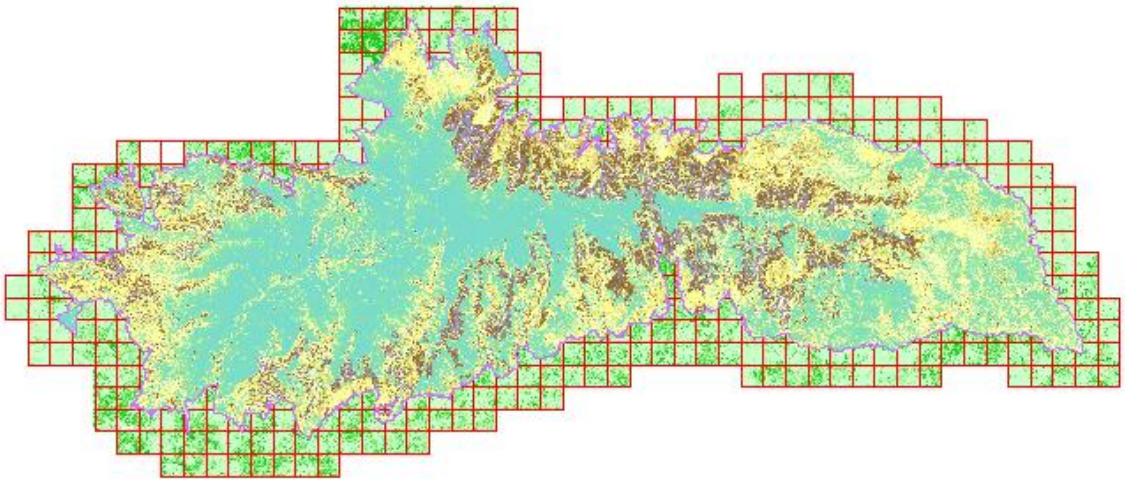


Figure 6 ALS coverage for Sierra Nevada (red tiles) together with the generated canopy height model (green background) and the clipped (using the national park limits, in purple) final raster (light blue to dark brown shades within the park limits).

4.3.3 References

[1] Evans, J.S., & Hudak, A.T. (2007). A multiscale curvature algorithm for classifying discrete return LiDAR in forested environments. IEEE Transactions on Geoscience and Remote Sensing, 45, 1029-1038

[2] Montealegre, A.L., Lamelas, M.T., & Riva, J. (2015). Interpolation Routines Assessment in ALS-Derived Digital Elevation Models for Forestry Applications. Remote Sensing, 7, 8631-8654

4.4 Pre-processing of SAR data (CERTH)

Many studies based on SAR data for biophysical parameters retrieval have been carried out at various frequencies and incidence angles, based on electromagnetic modeling and/or experimental data. Radar systems are sensitive to biophysical parameters over time and are not affected by clouds. Therefore, more robust,

accurate and frequently updated environmental monitoring is expected from the joint use of both types of measurements. Data from various sensors are used in the frame of the Ecopotential project, with Sentinel-1 the main sensor. All the Sentinel-1 imagery as well as the the ESA SNAP tool are made available for free by ESA and. The main SAR pre-processing steps that can be carried out in the SNAP toolbox are defined below and summarized in Figure 7.

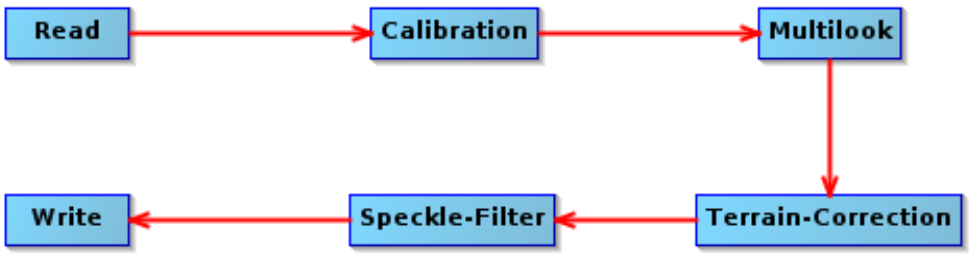


Figure 7 Summary of the SAR processing module to be uploaded to GitHub. Note that the flowchart may vary slightly depending on the user needs.

4.4.1 Calibration

When dealing with Single Look Complex (SLC) data, the first step of the processing is importing each TOPS (Terrain Observation with Progressive Scanning) burst/sub-swath using the ancillary information provided by ESA (annotation, calibration and noise files) as well as the data stored in GeoTiff format. The result of this process are three complex TOPS sub-swathes. When reading in the Sentinel-1 TOPS burst SLC data, radiometric calibration is automatically applied by using the calibration and noise information from the ancillary files. Level-1 Ground Range Detected (GRD) products consist of focused SAR that has been detected, multi-looked and projected to ground-range using an Earth ellipsoid model. The Sentinel-1 SLC or GRD data are calibrated using the following equation (S1 Report 2015):

$$\gamma^o = \frac{DN^2.tan(\theta_i)}{A_{dn}^2.K}$$

where DN stands for digital number, θ_i is the local incidence angle, A_{dn} the product final scaling from internal SLC to Level-1 GRD digital and K the calibration factor. Gamma nought is defined as the average radar cross section per unit area obtained by projecting the ground area into the plane perpendicular to the line of sight (Frey et al. 2013).

The SLC bursts in the TOPS mode data have to be mosaicked into a single SLC image. Data may be further multi-looked to reduce the speckle noise effect and the size of the images (that is of interest when dealing with a consequent amount of data from the Sentinels), at the cost of a degradation of the spatial resolution.

4.4.2 Co-registration

Images acquired are co-registered in order to be stackable. The resulting stack of data may be used for multi-image processing such as temporal speckle filtering or InSAR coherence computation. The co-registration step is straightforward when images are orthorectified. If not, offsets with respect to a reference image are estimated for the remaining images using a cross-correlation algorithm, such as in Werner et al. 2005. The offsets are modeled using least-squares regression and a third degree polynomial function that is subsequently used to correct a lookup table used to co-register each image to the reference image. A fast way of estimating the offsets is to compute the cross-correlation by simply multiplying images in the Fourier domain. After co-registration, images are stacked and further processing may be carried out.

Co-funded by the European Union

4.4.3 Topographic normalization

Terrain correction should be applied to geocode the images by correcting SAR geometric distortions using a producing ground projected images. Geometric distorsions include foreshortening (slope appearing compressed in the SAR image because the radar beam reached the base of a tall feature tilted towards the radar before it reaches the top), layover (the top of a tall feature lays over the base of the feature because the radar beam reaches the top of the feature before the base), and shadow (the radar beam is not able to illumnate the ground surface).

4.4.4 Speckle filtering

Speckle noise is an inherent problem in coherent imaging systems such as SAR. It creates strong intensity fluctuations due to positive and negatives interferences and hampers the analysis of images and the estimation of local radiometry. SAR processing chains thus include speckle reduction, often at the cost of a resolution loss. Spatial filtering is commonly used, but the preservation of point-like and fine structures and textures requires to adapt locally the estimation. However, the use of multi-image filtering is increasing with the availability of a high number of Sentinel temporal images and is therefore entirely adapted to the use of Sentinel-1 images.

4.4.4.1 Multi-image speckle filtering

A multi-image filter may be used to reduce speckle while preserving spatial resolution and the fine texture present in the image after multi-looking. The filter exploits the signatures of a scatterer in multiple images and it can be referred to as multi-looking in the temporal domain (Quegan et al. 2001). The method available in the SNAP toolbox, developed by Bruniquel and Lopes (1997) and Quegan and Yu (2001), provides new images with reduced speckle effects from multi-temporal and multi-polarised images. It is based on the following relation:

$$J_k(x,y) = \frac{\langle I_k(x,y)\rangle}{N}\sum_{i=1}^{N}\frac{I_i(x,y)}{\langle I_i(x,y)\rangle} with\ k = 1,\ldots,N$$

where $J_k(x, y)$ is the radar intensity of the output image k at pixel (x, y), $I_i(x, y)$ is the radar intensity of the input image i at pixel (x, y), $\langle I_i(x, y)\rangle$ is the local average intensity of the input image i at pixel (x, y) and N is the number of images. Because the operation consists of a pixel-to-pixel sum, the radiometric resolution is improved without damaging the spatial resolution. However, the images must be perfectly overlaid and all the statistical parameters well estimated. The resulting equivalent number of looks (ENL) is given by the following equation:

$$ENL = \frac{L.N.W}{N+W-1}$$

With L the initial number of looks and a fixed windiw size of W pixels.
An example of pre-processed Sentinel-1 image (calibration, multilooking, terrain correction and multi-image filtering) is shown in Figure 8.

Co-funded by the European Union

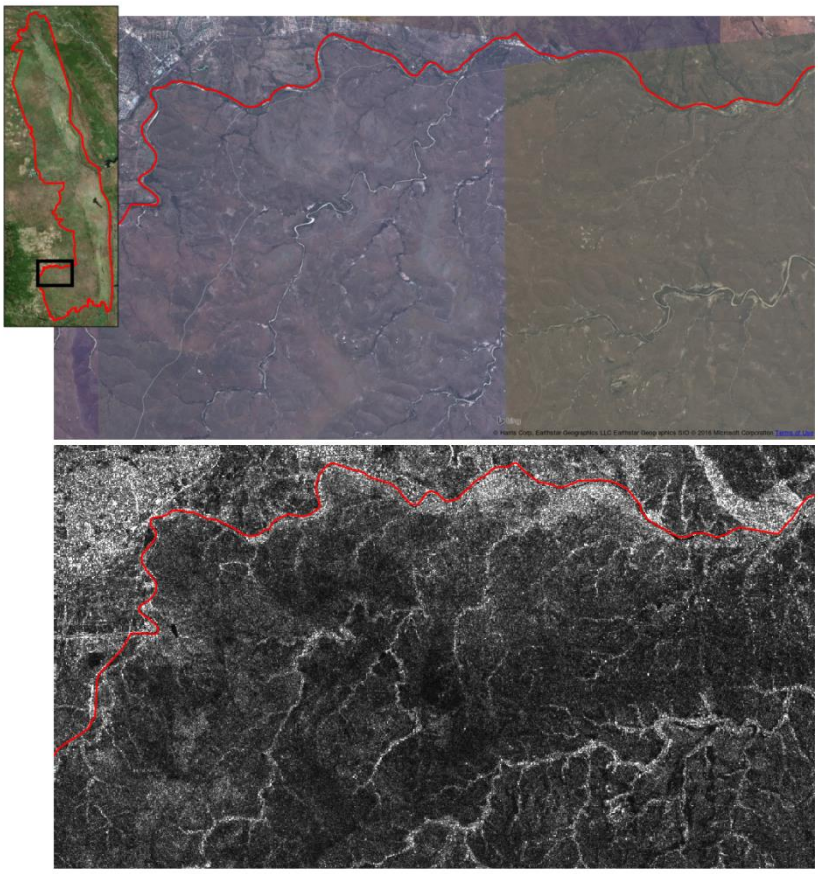


Figure 8 (Up) Very high resolution optical data from Google Earth and (Down) Pre-processed Sentinel-1 image over a 15 x 30 km area centred on 25.05oS and 31.35oE, in the Kruger National Park protected area (Kruger borders in red), after calibration, multilooking, terrain correction and multi-image filtering. The multi-image filter has been performed using 50 images: 25 dates (from 20 March 2015 to 7 April 2016) and 2 polarisations (VH and VV).

4.4.4.2 Spatial speckle filtering

Many approaches, relying on adaptive selection of pixels, has seen growing attention over the past years. Among the various methods that have been proposed to perform adaptive selection, the most used is the Lee filter (Lee et al. 1981) that locally select the best window among a few predefined windows (a rectangular window and eight edge-aligned oriented windows). Vasile et al. (2006) proposed to use region growing to build an adapted neighborhood restricted to similar pixels only. Adjacent pixels are incrementally aggregated based on their intensity. This approach is therefore more flexible than the use of predefined windows and leads to better resolution preservation than the Lee filter. Other approached are based on the work of Buades et al. (2005) who proposed the nonlocal means method: the relative importance of pixels can be weighted by comparing their surrounding neighborhoods rather than selecting pixels with similar intensities.

A new speckle reduction fusing Sentinel-1 and 2 filter has been developed in the frame of Ecopotential and is described in the following section.

The suppression of SAR image speckle noise is an important preprocessing step that favors the efficient exploitation of SAR data in a wide range of applications. This module aims to suppress speckle in theSAR Sentinel-1A GRD product by using guided image filtering. The guided filter computes the filtered output by considering the content of a guidance image. Guided filter has good edge-preserving property meanwhile it suppresses speckle noise.

4.4.4.2.1 Guided filter methodology

Co-funded by the European Union

Guided filter involves an input image I, a guidance image J and an output image $I'(p)$. The filtered output$I'(p)$at a pixel p is expressed as a weighted average:

$$\boldsymbol{I'(p) = \sum_{q} W(p,q) I(p)}$$

Equation 1

The filter weights $W(p,q)$ are given by:

$$W(p,q) = \frac{1}{|w_k|^2} \sum_{(p,q)\in w_k} (1 + (J(p) - \mu_k)^{\mathrm{T}} (\Sigma_k + \varepsilon U)^{-1} (J(p) - \mu_k)),$$

where$|w_k|$ is the total number of pixels in a support window w_k centered at pixel k and ε is a smoothness parameter. Σ_k and μ_k are the covariance and the mean of pixel colors within w_k A main advantage of the guided filter is that its computation cost is independent to the size of the selected support window, since *Equation 1* can be expressed a linear transform as follows:

$$I'(p) = \frac{1}{|w_k|} \sum_{p\in w_k} (a_k J(p) + b_k)$$

where:

$$a_k = (\Sigma_k + \varepsilon U)^{-1} \left(\frac{1}{|w_k|} \sum_{p\in w_k} J(p) I(p) - \mu_k \bar{I}(k) \right)$$

$$b_k = \bar{I}(k) - {a_k}^{\mathrm{T}} \mu_k$$

where$\bar{I}(k)$ is the mean of I within w_k. Moreover, the time complexity of the guided filter is independent to the support window radius.

In this module, the despeckling of SAR Sentinel-1A data is guided by the RGB bands of Sentinel-2. Therefore, the input image I is the SAR Sentinel-1A data and the guidance image$J$ is comprised of the RGB bands of Sentinel-2.

4.4.4.2.2 Inputs, parameters and outputs

The inputs to the module are a SAR Sentinel-1A image (with one or two polarization bands) and a Sentinel-2 RGB image (the inputs should have been acquired in a close date). Optionally, the user will be able to provide as input a cloud mask (corresponding to Sentinel-2) so that despeckling is not applied to the areas covered by clouds. In order to despeckle noise in SAR areas covered by clouds, the median filter is applied. In this way, despeckling is performed to all SAR regions independently to cloud coverage. Before running the module, the three inputs, provided in GeoTIFF format, are tested if they are projected to the same projection system and have the same resolution and extent. If the projection system and/or the resolution and/or the extent of the inputs are different, then the user is prompted to modify her/his input, prior applying the module.

The parameters required in the guided image filter are the radius “R” of the support window w_k (see the description of the guided filter methodology for the definition of w_k) and the smoothness parameter “eps” (which is denoted asεin the description of the guided image methodology), which is the criterion of separating a “flat patch” from a “high variance patch”. The patches with variance smaller than “eps” are smoothed whereas those with variance larger than eps are preserved. Default parameters are defined as {R,eps}={3,0.01}.

The output of the module is a raster of the despeckled SAR Sentinel-1A data at the same projection system, resolution and extent. The flowchart of the module is provided in Figure 9.

Co-funded by the
European Union

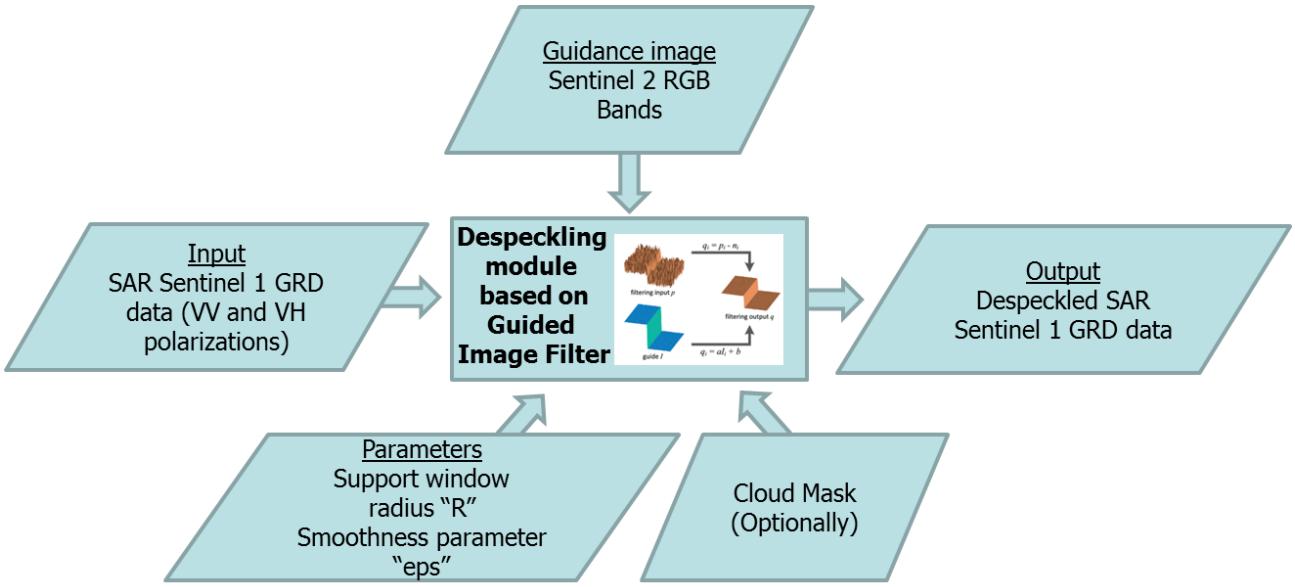
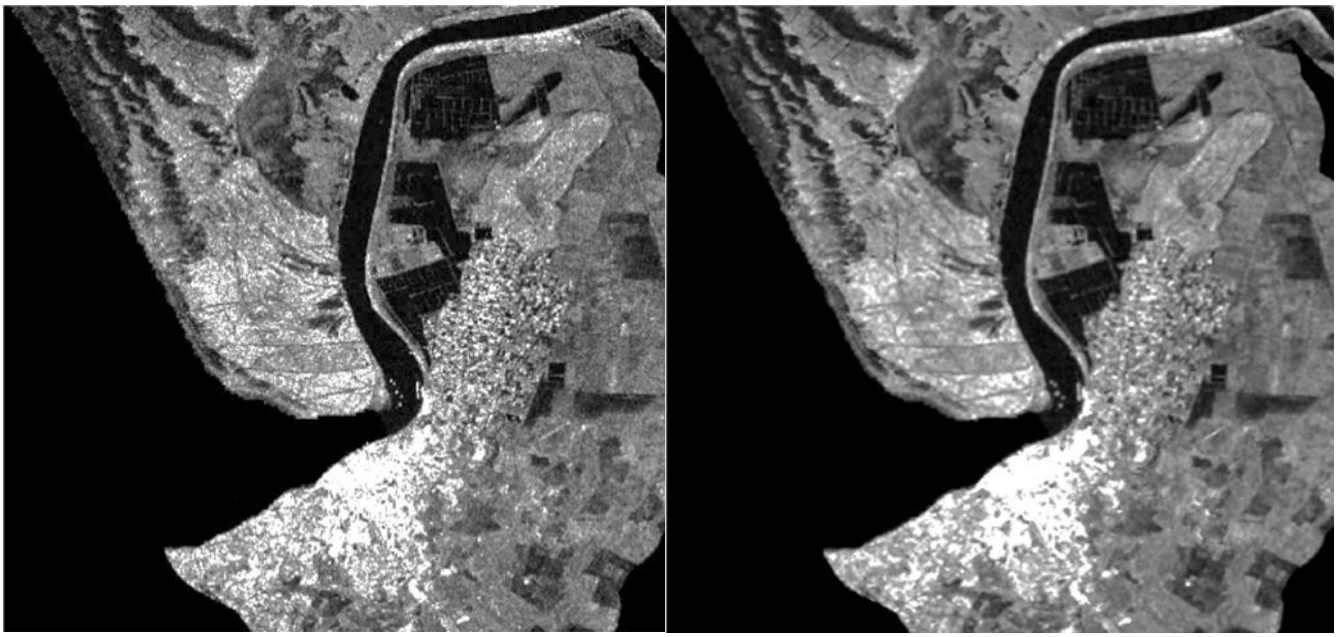


Figure 9. Flowchart of the module for Sentinel-1A SAR speckle suppression by fusing Sentinel 2-A RGB Optical data

(a) (b) (c)

Co-funded by the European Union



(d) (e)

Figure 10. Visual representation of: (a) SAR Sentinel 1 GRD VV Polarization before despeckling, (b) Sentinel 2 RGB bands used as guided image, (c) SAR Sentinel 1 GRD VV Polarization after despeckling, zoomed region of SAR data: (d) before and (e) after despeckling.

4.4.4.2.3 Experimental Results

Experimental results show that the proposed method can effectively suppress speckle and keep the texture structure. In Figure 9, a visual example from the Doñana protected area is provided. More specifically, Figure 9, (a) shows the SAR Sentinel-1A GRD VV polarization data before despeckling (date acquired 02/03/2016), Figure 9, (b) shows the Sentinel-2 RGB bands provided as guided image in the module (date acquired 08/03/2016), Figure 2 (c) shows the SAR data after despeckling. The parameters used for the guided image filtering were the default ones. Figure 9 (d), (e), show zoomed SAR regions, before and after despeckling, for more detailed visual comparison. The visual comparison shows that speckle noise has been significantly reduced, while the details have been preserved.

The results of this module are compared against the results of four despeckle filters, which are provided by the SNAP tool of ESA, namely: Lee [2], Median, Frost [3] and Boxcar (moving average) filters. The window size for the Lee, Median, Frost and Boxcar filters was set to 11. A quantitative and a visual comparison among the results are performed.

Co-funded by the European Union

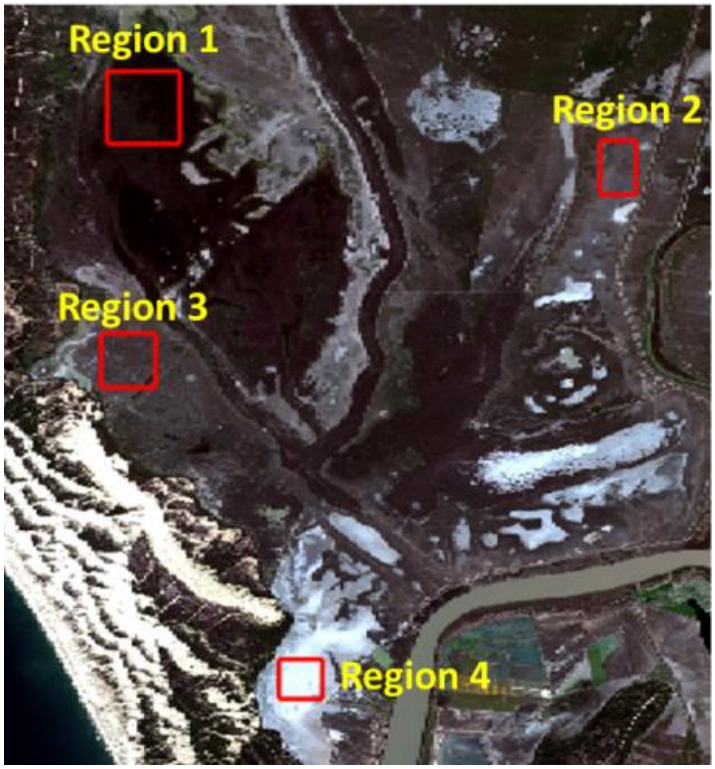


Figure 11.Four homogeneous areas inside Sentinel 2 RGB data.

4.4.4.2.4 Quantitative comparison

The equivalent number of looks (ENL), which is given by μ^2/σ^2 (μ and σ are the mean value and the standard deviation inside a SAR region, respectively) and is used to evaluate objectively the speckle reduction results [4]. The larger the ENL value, the better speckle noise has been suppressed. The ENL has been computed for four homogeneous regions inside Sentinel 2A RGB raster as visualized in Figure 11.

The ENL results for thefour homogeneous regions, depicted in Figure 11, are provided in Table 4.1 Obviously, guided filter suppresses more efficiently speckle noise than the compared approaches, since it gives larger ENL for all four homogeneous areas.

	ENL			
	Area 1	Area 2	Area 3	Area 4
Original SAR	1.1817	4.8157	3.8274	2.1446
Provided module	2.1429	60.9861	22.3346	6.9751
Lee	2.1429	56.5983	21.2264	6.4241
Median	2.0127	50.0371	19.7083	5.8892
Frost	1.9584	35.0789	15.2195	5.2231
Moving Average	1.8497	19.8006	11.0249	4.3307

Table 4.1 Methods performance comparison.

4.4.4.2.5 Texture information preservation

Following a visual comparison, Figure 12 shows zoomed despeckled SAR regions, which were generated by the compared despeckle approaches. , Figure 12 (a) maintains the edge details (e.g. efficiently despeckling while maintaining most possible texture information) in a higher or similar degree than , Figure 12 (b)-(e). Thus, by a first visual evaluation on an ECOPOTENTIAL PA, this module is able to preserve efficiently the SAR details.

Co-funded by the European Union

Furthermore, in an effort to quantitatively evaluate the performance for ECOPOTENTIAL sites to preserve real texture information, Donana PA ground data were retrieved by visual interpretation of a validation dataset taken by the equivalent Sentinel-2 image. Preliminary results indicate an approximately 5-10% improvement in comparison to SNAP offered statistical approaches. CERTH wishes to work with real ground truth data in the imminent future to fine tune the quantification of the performance of the suggested module/ approach.

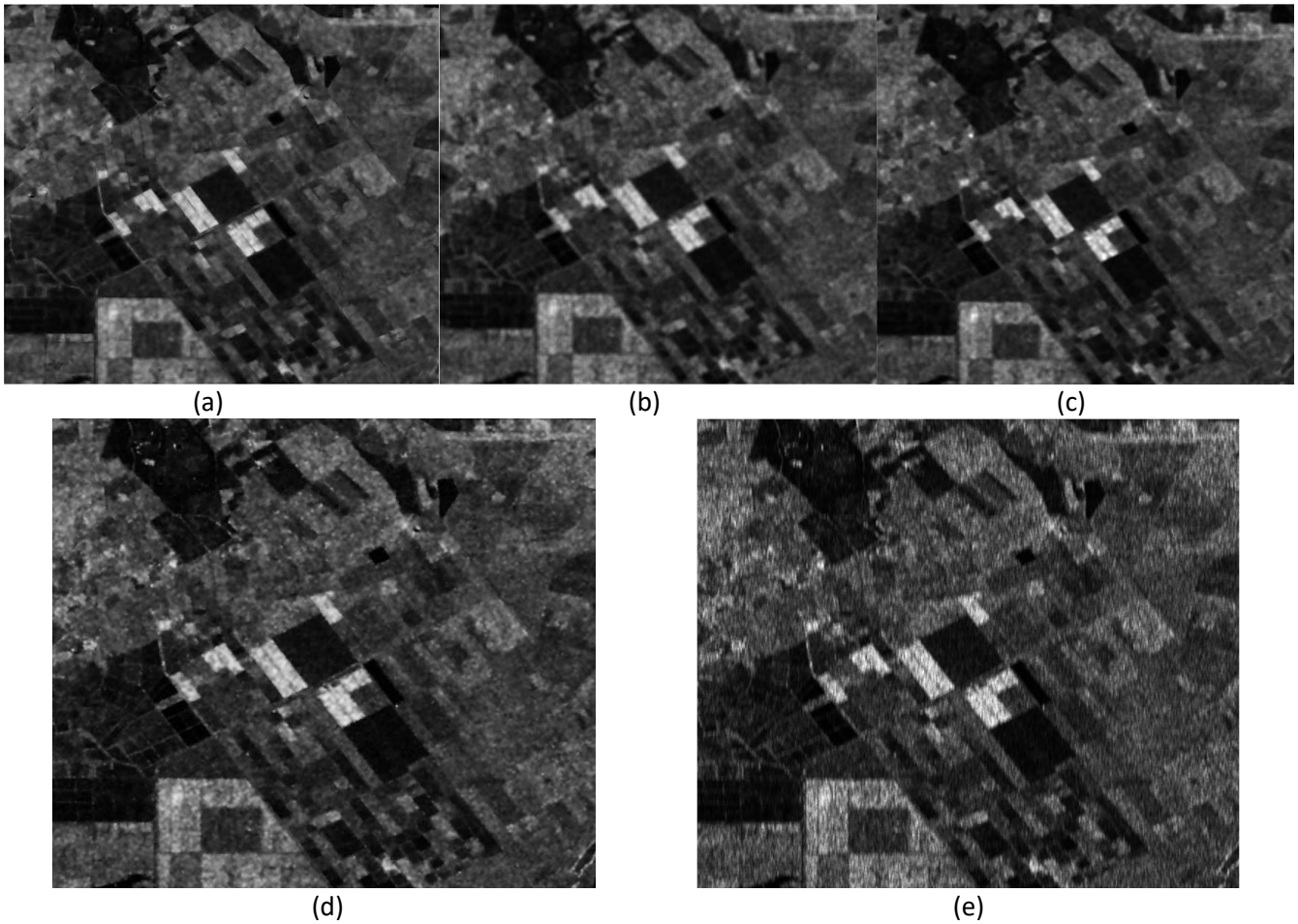


Figure 12. Zoomed region in SAR data after despeckling with (a) the proposed module, (b) Lee, (c) Median, (d) Frost, (e) Boxcar

4.4.4.2.6 Limitations of the module

The performance of this module depends on the quality of the Sentinel-2 data. Therefore, Sentinel-2 data should have been corrected for atmospheric effects and exhibit low cloud coverage. Additionally, Sentinel-1A data and Sentinel-2 data should be geometrically and timely (as possible) aligned.

4.4.4.2.7 Location to the module

The module is provided in a dedicated CERTH repository till its final version has been translated to open source code. It is available to all ECOPOTENTIAL partners any time on demand. Information regarding the inputs, parameters and outputs of the module are provided in the relevant ReadMe.docx file.

4.4.4.2.8 Disclaimer

The initial version of the module has been uploaded to CERTH repository and it is available to all ECOPOTENTIAL partners. The next step is to implement it in an open source code and upload it to GitHub to be broadly available.

4.4.5 Ortho-rectification

The final SAR pre-processing step is ortho-rectification to the desired coordinate system (e.g. Universal Transverse Mercator - UTM) using a lookup table based on digital elevation model (DEM) and the orbital information of the SAR data. To correct for possible inaccuracies in the input data, a refinement of the lookup table is applied in form of offsets between the reference SAR image (i.e. the first image of the data series) and a DEM-based simulated SAR image transformed to the radar geometry (Wegmüller et al. 2002).

4.4.6 References

[1] K. He, J. Sun, and X. Tang, "Guided image filtering," in Proc. ECCV, pp. 1–14, 2010.
[2] J. S. Lee, "Speckle analysis and smoothing of synthentic aperture radar images," Comput. Graph. Image Process., vol. 17, no. 1, pp. 24–32, 1981.
[3] V. S. Frost and J. A. Stiles, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," IEEE Trans. Pattern Anal. Mach. Intell., vol. 4, no. 2, pp. 157–166, 1982.
[4] F. Argenti, A. Lapini, L. Alparone, and T. Bianchi, "A tutorial on speckle reduction in synthetic aperture radar images," IEEE Geosci. Remote Sens.Mag., vol. 1, no. 3, pp. 6–35, 2013.
[5] Bruniquel, J., & Lopes, A. (1997). Multi-variate optimal speckle reduction in SAR imagery. International journal of remote sensing, 18(3), 603-627.
[6] Quegan, S., & Yu, J. J. (2001). Filtering of multichannel SAR images. IEEE Transactions on Geoscience and Remote Sensing, 39(11), 2373-2379.
[7] Vasile, G., Trouvé, E., Lee, J., & Buzuloiu, V. (2006). Intensity-driven adaptive-neighborhood technique for polarimetric and interferometric SAR parameters estimation. IEEE Transactions on Geoscience and Remote Sensing, 44(6), 1609-1621.
[8] Buades, A., Coll, B., & Morel, J. M. (2005). A review of image denoising algorithms, with a new one. Multiscale Modeling & Simulation, 4(2), 490-530.

4.5 Gap Filling: Data Interpolating Empirical Orthogonal Functions (DELTARES)

The rising need for complete datasets in several geophysical applications led to the development and improvement of techniques for the reconstruction of missing data in space and time. One of these is the Data Interpolating Empirical Orthogonal Functions (DINEOF) method. DINEOF achieves this using an EOF-based technique. In particular, it manages to fill gaps by iteratively decomposing the data field via Singular Value Decomposition (SVD) until a best solution is found. This is done by progressively considering more EOFs in the reconstruction of missing values until the minimization of error converges [Nikolaidis et al. 2013].

4.5.1 Description of the module

DINEOF determines the dominant spatial and temporal patterns of variation in the data set, called Empirical Orthogonal Functions, EOFs (principal components). By putting together these modes of variation one can obtain a so-called gap-filled reconstruction of the MERIS data at all grid locations and sample times
In an EOF analysis, the MERIS 3D hypercube "time series" [x, y, time] is decomposed into its statistically dominant patterns. This yields a total of p 2D maps [x,y]p with associated 1D time series [time]p. Multiplying a 2D map [x,y]p with a 1D time series [time]p will yield an approximation of the original 3D dataset [x, y, time]. Using more

modes p will yield a better approximation. However, the first few modes already contain most of the variance of the dataset. This is the variance associated with large scale patterns.
The figure below illustrates the decomposition of the MERIS data in the spatiotemporal EOF modes and the reconstruction (gap filling).

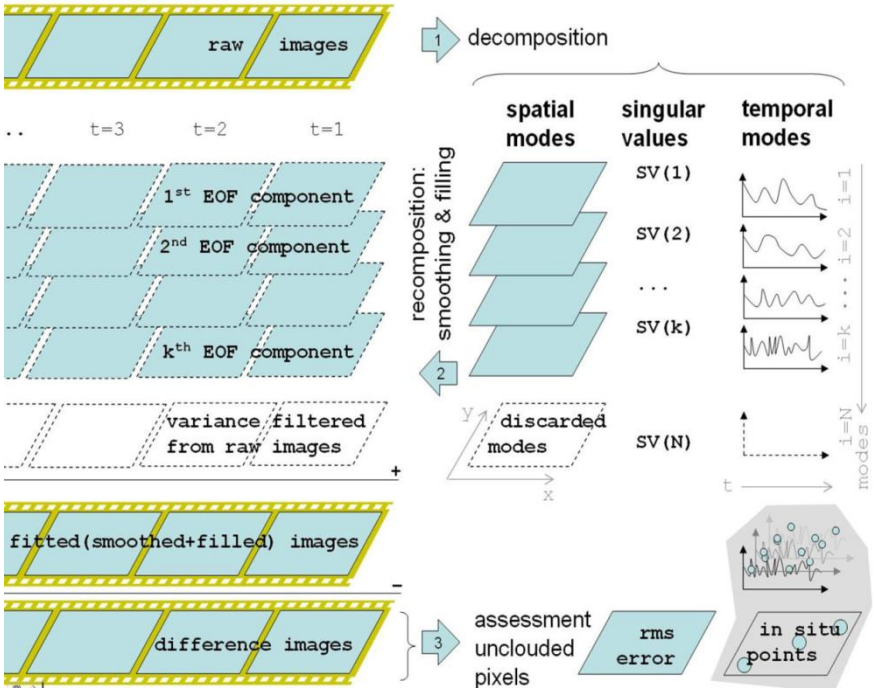


Figure 13 Schematic work flow of DINEOF analysis of a time series of RS images. [Src :CoBiosProject - FP7-SPACE-2010-1]

The input image archive is condensed in a two-dimensional matrix of columns of m elements (spatial dimension) and rows of n elements (temporal dimension). To construct this matrix, each scene acquired at given t_i time (where i=1,…,n) is stored as a one dimensional array, in the i-th column of the matrix. To build each one-dimensional array, the original two-dimensional scene is scanned in a column-major order (column by column and pixel by pixel). The module support only NetCDF files as input.

The calculation of the EOFs can be performed by a standard Singular Value Decomposition SVD technique, which provides a decomposition of the initial matrix in 3 matrices giving direct access to the spatial EOF modes and the temporal EOF modes. Since the calculation of EOFs requests a set of cloudless images and the interpolation of the missing data requests the knowledge of EOFs, this circular dependence has been solved by introducing an iterative method.

Therefore, the principle of the algorithm is to fill in the missing data of the initial matrix by using iterative cycles of SVD producing a set of EOFs and to find new missing data estimates in each iteration. The initialization of the iterative process is done by replacing all missing data with zero values in order to obtain a complete matrix. We first apply the reconstruction of X with a single EOF which allows to calculate the first prediction of the missing data which replaces the first guess made and allows thus for a second iterative cycle. This iteration continues to produce improved missing data and EOF until the changes observed in the missing data estimates between one iterative cycle and the previous one are insignificant. The convergence criterion is chosen by the user. This iterative cycle is then carried out considering two EOFs, before going to three EOFs, and so on. The optimal number of EOFs needed to fill in the missing data is determined by a cross-validation technique: before starting the whole procedure, a small percentage of data points are set apart and flagged as missing. Once convergence is reached for a fixed number of EOF modes, a RMS is calculated between the reconstructed values and the initial dataset aside. The number of modes that minimizes the error will be considered as the optimal one.

Co-funded by the European Union

4.5.2 Experimental Results

The results presented in this section make use of ESA's MERIS data acquired over the domain of interest: North Sea and Wadden Sea. The available images were recorded during the overpass of the Envisat satellite occurring nominally once per day between 10:00 and 12:00 AM UTC. Although MERIS recorded the optical imagery at a full spatial resolution of about 300x300 m2, the standard product for the North Sea is the Reduced Resolution (RR) 1x1 km2 resolution pixel values. From the optical reflectance spectra recorded by MERIS, concentrations of chlorophyll-a and SPM (suspended particulate matter), CDOM absorption (coloured dissolved organic matter) and spectral extinction of visible light (Kd) and estimates of the standard error in these values have been computed. These data are based on retrieval by means of the HYDROPT algorithm (Van der Woerd&Pasterkamp, 2008). The period consider for the results here presented is from January 2003 to December 2009.

Before building the data cubes from the DINEOF, MERIS data have been evaluated for outliers and flagged pixel values have been rejected according to the criteria explored in Gaytan et al. (2013) and Eleveld & De Reus (2010).It is important than within the cleaning procedure not only outliers but also images with extreme cloud coverage should be excluded from the dataset. Base in experience and literature (Alvera-Azcárate et al. 2005; Sirjacobs et al. 2010) it is recommended to use only images containing at least 2% of the coverage.

MERIS data used in this example has been projected onto the ZUNO-DD ("ZUidelijkeNOordzee, DomeinDecompositie") Delft3D-WAQ grid. This grid has a lower resolution than the satellite images. The reason for that is that Gap filled data will be used it in a later state in the water quality model.

To illustrate the gap filling module, two procedures are presented. The first one considers 7 individual data cubes, one per year (2003,…,2009) and de second consider one data cube for the all period (2003-2007). The purpose of those schemas y to illustrate how sensible is the reconstruction procedure to the amount of images. Figure 14 shows the images before the filling procedure and after the filling procedure for the 2 scheme.

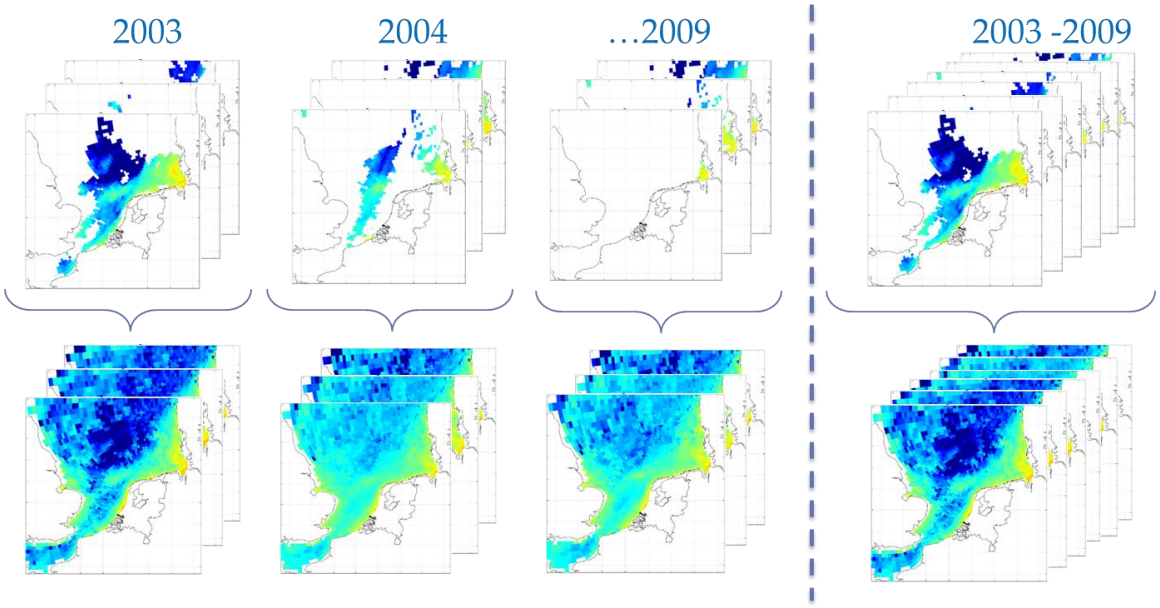


Figure 14 Illustration of the Gap filled data using yearly available MERIS data and longer period (2003-2009) of data

To validate the results presented in the Figure 14 available In-situ data was been used to evaluate the performance of the reconstructed images.

Chlorophyll-a measured values are acquired from the monitoring stations which belong to the Dutch water agency Rijkswaterstaat (RWS). The sampling is characterized by a limited frequency in time: the stations tend to have 10 to 20 measurements per year. Figure 15 shows the location of the monitoring stations. In total 24

Co-funded by the European Union

stations are used, with different distances for the coast, 12 of them are located less than 10 km far from the coast, while the most distant one is 235 km far from the coast; 6 out of 24 stations are located in the Wadden Sea.

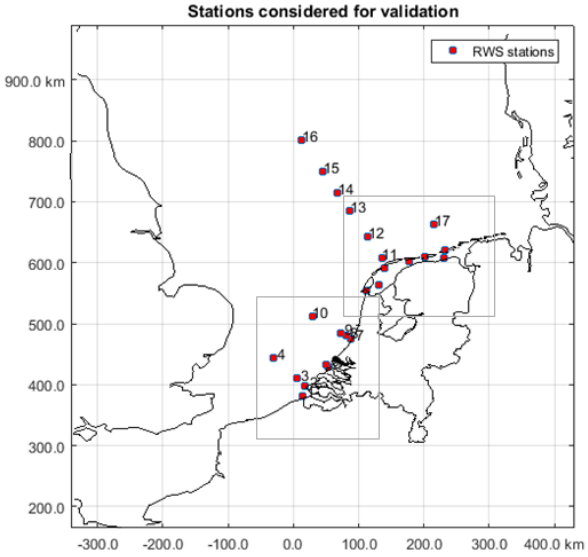


Figure 15 Location of the monitoring stations

In order to compute a goodness of fit, in this case the Bias and Root Mean Square Error (RMSE), it is necessary to match the samples in time. Unfortunately In-situ measurements and the MERIS information do not happen often at the same time. So a time window of 24 hours was consider for the data matching.

Table.2 shows the error statistics for the both schemas, also the number of images per year and the percentage of area cover along the year. The average coverage of the data set is about 85%. Spatially, regions with the highest missing data percentage are the coastal area. The accuracy of DINEOF algorithm under both schemas is very similar, also viable in Figure 16 . Nevertheless when considering larger imagery for the reconstruction the spread in the errors is reduced.

Year	Numb of images	% of cloud coverage	Yearly reconstruction		2003-2009	
			BIAS	RMSE	BIAS	RMSE
2003	234	84.59%	-2.476	6.976	-2.417	7.245
2004	234	85.23%	-0.578	9.460	-1.028	7.373
2005	249	87.39%	-0.998	12.011	-0.502	6.784
2006	238	86.69%	-2.347	6.890	-2.138	6.747
2007	205	85.04%	-2.338	6.820	-2.283	7.541
2008	162	86.77%	-2.280	9.414	-1.423	10.889
2009	181	88.25%	-1.405	8.082	-0.846	8.670

Table.2 Bias and Root Mean Square Error of the reconstructed images for the two schemas

Co-funded by the European Union

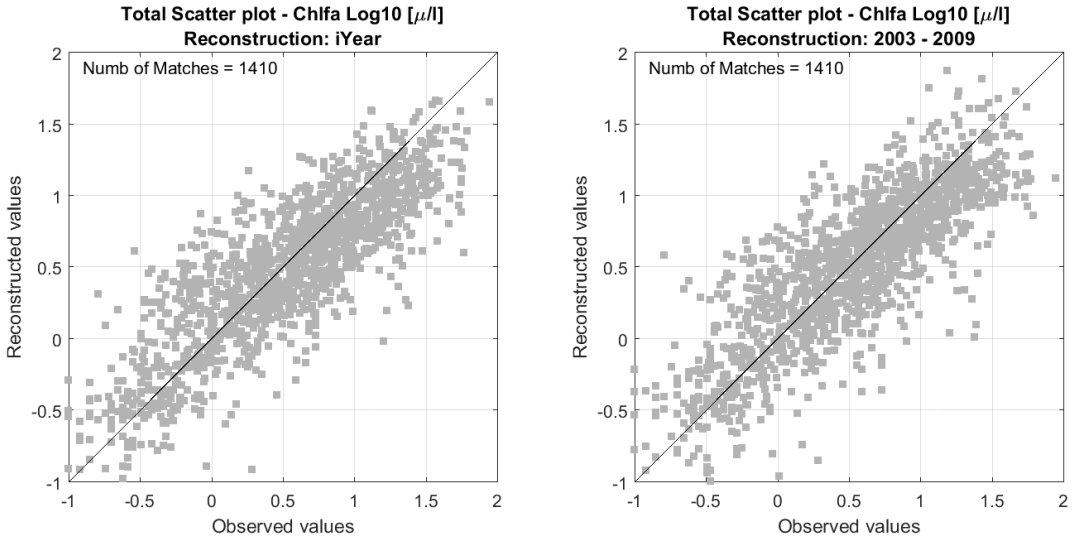


Figure 16 In-situ measurement VsRecostructed MERIS images. Plot in the left shows the yearly schema, plot in the right shows the comparison when considering long period imagery.

Finally, Figure 17 shows time series of reconstructions of chlorophyll-a data and the original images with missing values corresponding to cloudy conditions. The selected station has 22% of valid data along for the year 2004. It is clear that the reconstruction yields realistic chlorophyll-a values even for the highest percentages of cloud cover.

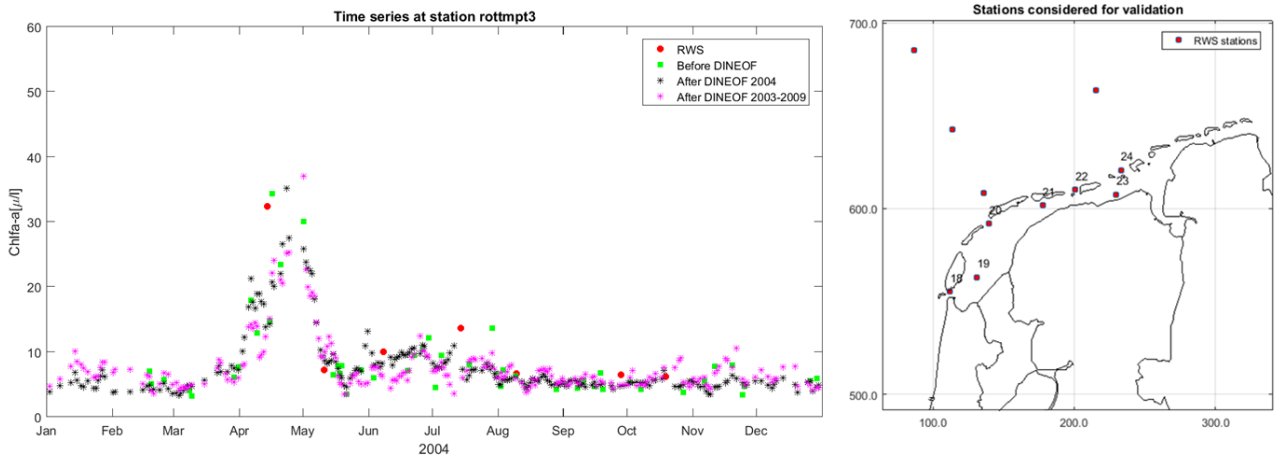


Figure 17 Time Series of original MERIS data and reconstructed MERIS. In-situ measurements are represented by red dots, original MERIS by green dots, MERIS reconstructed yearly by black dots and MERIS reconstructed by long period imagery by magenta dots.

4.5.3 Flow chart

For this module two main inputs are need it, a file containing the data cube and a file containing a mask of the area. The mask files contain binary information, 1 if the pixel should have in principle a value or 0 otherwise.

The inputs for the module are different layers or images of the same variable with some overlapping area. The resolutions/coarseness can vary but the geographical reference must be the same for all input layers. So far the module support only NetCDF files. The flow char of this module is presented in Figure 18.

The first step for the module is to read NetCDF files to produce the Data cube in a DINEOF input format. Moreover, it is possible to set the logarithmic transformation in order to convert the natural values of the studied variable into their corresponding logarithmic values.

The next step is to proceed with the gap filled process using DINEOf. If the user has specific values for the parameter to be used in the gap filling procedure, it is possible to add those specification, otherwise the module has some default setting for the DINEOF.

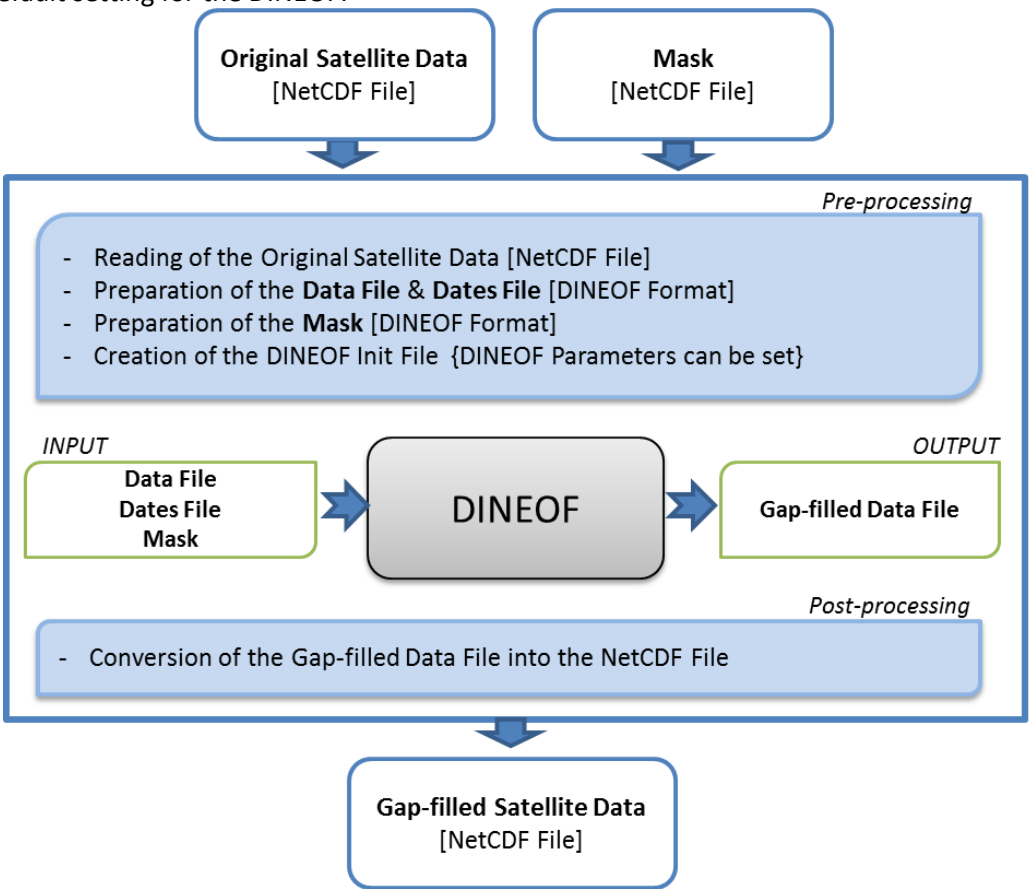


Figure 18Flowchart for the Gap filled module

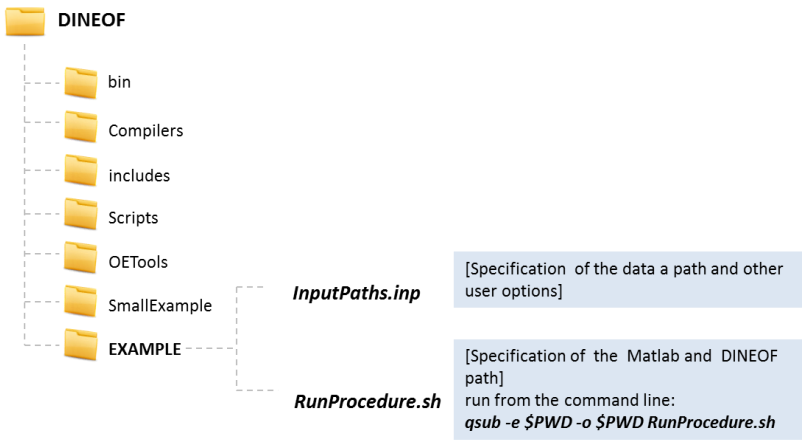


Figure 19 Structure of the DINEOF folder

The final step of the module is to convert the DINEOF data cube output format back to the NetCDF file format. If a logarithmic transformation was chosen, the output will be converted back to natural values .Figure 19 shows the structure of the folder and the files to be changed with the user paths and settings.

4.5.4 References

[1] Alvera-Azcarate, A., Barth, A., Rixen, M., and Beckers, J.-M.: Reconstruction of incomplete ´ oceanographic data sets using Empirical Orthogonal Functions. Application to the Adriatic Sea, Ocean Modelling, 9, 325–346, 2005.

[2] ALVERA-AZCÁRATE, A., BARTH, A., BECKERS, J.-M.& WEISBERG, R. H., 2007.Multivariate reconstruction of missing data in sea surface temperature, chlorophyll and wind satellite fields. Journal of Geophysical Research, 112, C03008, doi:10.1029/2006JC003660.

[3] BECKERS, J.-M., BARTH, A. & ALVERAAZCÁRATE, A., 2006.DINEOF reconstruction of clouded images including error maps.Application to the Sea Surface Temperature around Corsican Island. Ocean Science 2:(2), 183-199.

[4] BECKERS, J.-M. & RIXEN, M., 2003.EOF calculations and data filling from incomplete oceanographic data sets. Journal of Atmospheric and Oceanic Technology 20 (12): 1839-1856.

[5] MAURI, E., POULAIN, P. M. & JUZNICZONTAC, Z., 2007. MODIS chlorophyll variability in the northern Adriatic Sea and relationship with forcing parameters. Journal of Geophysical Research 112, C03S11, doi:10.1029/2006JC003545. MAURI, E., POULAIN, P. M. & NOTARSTEFANO, G., 2008. Spatial and temporal variability of the sea surface temperature in the Gulf of Trieste between January 2000 and December 2006.Journal of Geophysical Research. In Press, doi:10.1029/2007JC004537.

[6] http://modb.oce.ulg.ac.be/mediawiki/index.php/DINEOF

Co-funded by the European Union

5 Image processing

5.1 Indices and Band Ratios(CREAF-UAB)

Indices and band ratios are the most common form of spectral enhancement. The technique of rationing bands, involves, basically, dividing the spectral response value of a reflectance pixel in one band or image with the spectral value of reflectance of the corresponding pixel in another band or image. This is done with the objective of normalize varying effects such as soil difference or in order to maximize sensitivity to the feature of interest, for instance, the relative health of vegetation. The most common approaches are based on simple band division to include weighting or difference as well as introduction of new variables. Of these, the Normalized Difference Vegetation Index (NDVI) or the EVI (Enhanced Vegetation Index) are the most widely used.
With this module (CalcImg) mathematical and logical operations on images can be performed simultaneously on any type and number of layers. All rasters contained in the expression must have the same geographical extent, resolution and reference units in the same reference system. All calculations are made with double precision (15-16 significant digits, approx. range 1.7E-308 to 1.7E+308). There is no limit to the length and complexity of the expression (it will only depend on the user's hardware and operating system). The logical-mathematical formula to be calculated is entered in the first entry field and is formed by numbers, layers (raster and vector), operators and functions. Some functions and operators are already implemented. The resulting data type can be controlled and supports several types (integer, real, double,,,).

Moreover, to simplify the production of widely used vegetation indices, another module (Indexs) is provided to automatically compute several indices, using well-known formulas:

1) DVI (Differential Vegetation Index): Difference between near infrared (NIR) and red (R) bands of a sensor.

$$DVI_{i,j} = 2.4 \cdot \rho_{i,j,IRP} - \rho_{i,j,R}$$

2) RVI (Ratio Vegetation Index): Simple division between the near infrared and red bands. It is being used for biomass and leaf area index (LAI) assessments from 1969.

$$RVI_{i,j} = \frac{\rho_{i,j,IRP}}{\rho_{i,j,R}}$$

3) NDVI (Normalized Difference Vegetation Index): The NDVI vegetation index has been the most used in Remote Sensing of vegetation in the last decades. It was developed by Rouse et al. (Rouse J.W., Haas, R.H., Schell, J.A. i Deering, D.W. "Monitoring Vegetation Systems in the Great Plains with ERTS". Third ERTS Symposium, NASA SP-351, I:309-317 (1973)) from the analysis of Landsat-MSS data. They found that, although the simple quotient (RVI) could be used as a measure of relative greenness, the position and cyclical deviations could cause important errors. They defined the Normalized Difference Vegetation Index (NDVI) in the following way:

$$NDVI_{i,j} = \frac{\rho_{i,j,IRP} - \rho_{i,j,R}}{\rho_{i,j,IRP} + \rho_{i,j,R}}$$

4) TVI (Transformed Vegetation Index): Transform of the NDVI index that allows working without negative values.

$$TVI_{i,j} = \sqrt{NDVI_{i,j} + 0.5}$$

5) PVI (Perpendicular Vegetation Index)

$$PVI_{i,j} = \sqrt{(\rho_{i,j,IRP_s} - \rho_{i,j,IRP_v})^2 + (\rho_{i,j,R_s} - \rho_{i,j,R_v})^2} \equiv \frac{\rho_{i,j,IRP} - \gamma \cdot \rho_{i,j,R} - \delta}{\sqrt{(\gamma^2 + 1^2)}}$$

Co-funded by the European Union

where S indicate soil reflectance and V indicate vegetation reflectance.

6) WDVI (Weighted Difference Vegetation Index)

7) SAVI (Soil Adjusted Vegetation Index): Index that intends to minimize the soil effect on the vegetation index. This effect is important for partially covered surfaces.

$$SAVI_{i,j} = \frac{\rho_{i,j,IRP} - \rho_{i,j,R}}{\rho_{i,j,IRP} + \rho_{i,j,R} + L}(1+L)$$

L = constant that depends on the type of soil. A typical value is 0.5.

8) TSAVI (Transformed Soil Adjusted Vegetation Index): Readjustment of the SAVI index to consider the effect of the soil brightness.

$$TSAVI_{i,j} = \frac{\gamma \cdot (\rho_{i,j,IRP} - \gamma \cdot \rho_{i,j,R} - \delta)}{\gamma \cdot \rho_{i,j,IRP} + \rho_{i,j,R} - \gamma \cdot \delta + \chi \cdot (1+\gamma^2)}$$

9) MSAVI1 (Modified Soil Adjusted Vegetation Index - 1): The adjustment factor L of SAVI index is not constant, but an inverse function of the vegetation cover. There have been proposed two solutions:

$$MSAVI_{i,j,1} = \frac{\rho_{i,j,IRP} - \rho_{i,j,R}}{\rho_{i,j,IRP} + \rho_{i,j,R} + L}(1+L) \quad \text{on } L = 1 - 2 \cdot \gamma \cdot NDVI \cdot WDVI$$

10) MSAVI2 (Modified Soil Adjusted Vegetation Index - 2):

$$MSAVI_{i,j,2} = \frac{2 \cdot \rho_{i,j,IRP} + 1 - \sqrt{(2 \cdot \rho_{i,j,IRP} + 1)^2 - 8(\rho_{i,j,IRP} - \rho_{i,j,R})}}{2}$$

11) TWVI (Two-axis adjusted Vegetation Index): Studying the complexity of the soil line, it has been proposed an index that improves some aspects of PVI and SAVI indexes. This improvement is appreciable when the offset D between a pixel of soil (without vegetation) and the soils line is large:

$$TWVI_{i,j} = \frac{\rho_{i,j,IRP} - \rho_{i,j,R} - \Delta}{\rho_{i,j,IRP} + \rho_{i,j,R} + L}(1+L) \quad \text{on } \Delta = \sqrt{2} \cdot e^{-KLAI} \cdot D = \sqrt{2} \cdot (1 - (I_\lambda - I) / (I_\chi - I)) \cdot D$$

12) ARVI (Atmospheriaclly Resistant Vegetation Index):
Vegetation index similar to NDVI but improved in the way that it considers the atmospheric effects. To achieve this, the red band is substituted by a red-blue band that accounts for the different atmospheric dispersion in these two bands giving the following expression:

$$ARVI_{i,j} = \frac{\rho_{i,j,IRP} - \rho_{i,j,RB}}{\rho_{i,j,IRP} + \rho_{i,j,RB}}$$

where $\rho_{i,j,RB} = \rho_{i,j,R} - \eta(\rho_{i,j,A} - \rho_{i,j,R})$ is the combined reflectance of the red and blue bands. The constant is a calibration parameter that depends on the type of atmosphere. By default its value is 1.

13) SARVI (Soil Adjusted and Atmospheriaclly Resistant Vegetation Index)

14) TSARVI (Transformed Soil Atmospheriaclly Resistant Vegetation Index):

$$TSARVI_{i,j} = \frac{\gamma_{rb} \cdot (\rho_{i,j,IRP} - \gamma_{rb} \cdot \rho_{i,j,R} - \delta_{rb})}{\gamma_{rb} \cdot \rho_{i,j,IRP} + \rho_{i,j,R} - \gamma_{rb} \cdot \delta_{rb} + \chi \cdot (1+\gamma_{rb}^2)}$$

Where the soils line in the spectral range red, blue and near-infrared is defined by:

$$IRP = \gamma_{rb} RB + \delta_{rb}$$

15) AVI (Angular Vegetation Index): The AVI (Plummer et al. 1994) was chosen for the ATSR-2 sensor, and aims to use the availability of the green channel to normalize the effects of soil types and the atmosphere while maintaining sensitivity to the properties of the vegetation. The index is based on a geometric concept of the angle defined between the green, red and near-infrared wavebands, which characterizes chlorophyllthe absorption.

$$AVI_{i,j} = \tan^{-1}\left\{\frac{\lambda_3 - \lambda_2}{\lambda_2}(\rho_{i,j,IRP} - \rho_{i,j,R})^{-1}\right\} + \tan^{-1}\left\{\frac{\lambda_2 - \lambda_1}{\lambda_2}(\rho_{i,j,V} - \rho_{i,j,R})^{-1}\right\}$$

The λ correspond to the centers of each of the wavebands.

16) GEMI (Global Environment Monitoring Index): This index was proposed for NOAA-AVHRR sensor in order to have a global knowledgement of the vegetation on the Earth. It is a non linear index that minimizes the solar and atmospheric effects.

$$GEMI_{i,j} = \eta(1 - 0.25\eta) - (\rho_{i,j,R} - 0.125)(1 - \rho_{i,j,R})$$

$$\text{on } \eta = \left[2 \cdot (\rho^2_{i,j,IRP} - \rho^2_{i,j,R}) + 1.5 \cdot \rho_{i,j,IRP} + 0.5\rho_{i,j,R}\right] \cdot (\rho_{i,j,IRP} + \rho_{i,j,R} + 0.5)$$

17) NDWI-1 (Normalized Difference Water Index): The NDWI was created by McFeeters (1996) to delineate open-water features such as lakes and reservoirs, and improve their detection in satellite images. In this index water surfaces have positive values, while land and vegetation surfaces have 0 or negative values.

$$NDWI_{i,j} = \frac{\boldsymbol{\rho}_{i,j,GRENN} - \boldsymbol{\rho}_{i,j,IRP}}{\boldsymbol{\rho}_{i,j,GRENN} + \boldsymbol{\rho}_{i,j,IRP}}$$

McFeeters, SK (1996) "The use of the normalized difference water index (NDWI) in the delineation of open water features". INTERNATIONAL JOURNAL OF REMOTE SENSING. 17: 1425-1432.

18) NDWI-2 (Normalized Difference Water Index): The NDWI index has been developed in many different ways and some authors (Ouma & Tateishi, 2006) have proposed modifications. The NDWI2 uses the NIR and MIR bands to distinguish shoreline from water masses. This index makes it possible to produce rapid mapping of the shoreline.

$$NDWI_{i,j} = \frac{\boldsymbol{\rho}_{i,j,IRM} - \boldsymbol{\rho}_{i,j,IRP}}{\boldsymbol{\rho}_{i,j,IRM} + \boldsymbol{\rho}_{i,j,IRP}}$$

Ouma, YO; Tateishi, R. (2006). "A water index for rapid mapping of shoreline changes of five East African Rift Valley lakes: an empirical analysis using Landsat TM and ETM+ data". INTERNATIONAL JOURNAL OF REMOTE SENSING. 27: 3153-3181.

19) EVI (Enhanced Vegetation Index): The EVI obtains a response from the structural variations of the vegetation canopy including the leaf area index (LAI), the type and architecture of the canopy and plant physiognomy. The EVI was developed to optimize the vegetation signal with improved sensitivity for high biomass densities, which it achieves by separating the signal coming from the vegetation and from atmospheric influence.

Co-funded by the European Union

$$EVI_{i,j} = G \times \frac{\boldsymbol{\rho}_{i,j,IRP} - \boldsymbol{\rho}_{i,j,R}}{\boldsymbol{\rho}_{i,j,IRP} + C_1 \cdot \boldsymbol{\rho}_{i,j,R} - C_2 \cdot \boldsymbol{\rho}_{i,j,B} + L}$$

The EVI is an improved version of the NDVI with an L=1 soil correction factor, two C1=6 and C2=7.5 coefficients designed to correct atmospheric dispersion effects and a G=2.5 gain factor for freely-available data from the MODIS satellite.

20) AFRI (Aerosol Free Vegetation Index): The AFRI was developed from the NDVI and uses the bands of the MIR mid-infrared by establishing a correlation between these bands for MODIS (p2.1 and p1.6) and the red band.

$$AFRI_{i,j,2.1} = \frac{\boldsymbol{\rho}_{i,j,IRP} - \mathbf{0.5} \cdot \boldsymbol{\rho}_{i,j,2.1}}{\boldsymbol{\rho}_{i,j,IRP} + \mathbf{0.5} \cdot \boldsymbol{\rho}_{i,j,2.1}} \qquad AFRI_{i,j,1.6} = \frac{\boldsymbol{\rho}_{i,j,IRP} - \mathbf{0.66} \cdot \boldsymbol{\rho}_{i,j,1.6}}{\boldsymbol{\rho}_{i,j,IRP} + \mathbf{0.66} \cdot \boldsymbol{\rho}_{i,j,1.6}}$$

The superiority of the AFRI is due to the capacity of the MIR bands to penetrate the atmospheric column, even with the presence of smoke and sulphates. These indices are therefore useful for evaluating vegetation in the presence of smoke, anthropic pollution or volcanic eruptions. It is important to bear in mind that in clear sky conditions with low atmospheric influence these indices correlate highly with the NDVI and their values are very similar.

21) PRI (Photochemical Reflectance Index): The PRI is an index that makes it possible to estimate the RUE (Radiation-Use Efficiency) of vegetation and GPP (Gross Primary Productivity) in Mediterranean ecosystems. The evergreen Mediterranean forests display very low seasonal variation in the vegetation indices (NDVI and EVI) and very high seasonal variation in carbon absorption. In these cases this index makes it possible to detect what part of the absorption capacity of carbon actually occurred. The PRI is obtained from freely-available information from the MODIS satellite.

$$PRI = \frac{\boldsymbol{\rho}_{5.2} - \boldsymbol{\rho}_{5.4}}{\boldsymbol{\rho}_{5.2} + \boldsymbol{\rho}_{5.4}}$$

Where the p5.2 corresponds to band 11 (526-536 nm) and thep5.4 corresponds to band 12 (546-556 nm) of MODIS.
Garbulsky, MF; Penuelas, J; Papale, D & Filella, I (2008) Remote estimation of carbon dioxide uptake by a Mediterranean forest. GLOBAL CHANGE BIOLOGY. 14. 12: 2860-2867

The Indexs module asks for the names of the files corresponding to the red, infrared and blue bands (only atmospheric indexes). It also requires a set of parametres that depend on the index equation. If they are not given in the command line, the program will ask for them interactively, suggesting a possible value.

Both CalcImg and Indexs modules are provided as part of the MiraMon Remote Sensing and GIS software available to all ECOPOTENTIAL partners any time on demand. Information regarding the inputs, parameters and outputs of the module are provided in the relevant Help of the software. MiraMon supports several parameters on the command line and allows writing Batch files that can run combining several chained orders. Moreover, combining Web Processing Service (WPS) standard and the ISO lineage model, MiraMon allows to easily edit and reproduce geospatial workflows described on the metadata, scaling them up to similar situations and executing them as remote processes.

Pons, X. (2016). *MiraMon. Geographical information system and remote sensing software*. Centre de Recerca Ecològica i Aplicacions Forestals (CREAF). Retrieved from http://www.creaf.uab.es/miramon

Co-funded by the European Union

5.2 Image Time Series –Data Mining System (DLR)

We present a modular system offering Data Mining and Knowledge Discovery functionality for long time series data called Image Time Series- Data Mining (ITS-DM). It implements novel techniques for time series content exploration and exploitation.

The Image Time Series –Data Mining system is composed of several components: the Data Model Generation (DMG), Data Mining Database (DMDB), Data Mining and Knowledge Discovery (KDD), and the Query Engine and Analytics. These components offer functionalities such as ingestion and feature extraction from radiometrically and geometrically corrected Earth-Observation Data, metadata extraction, semantic definition of the image content based on machine learning and data mining methods, advanced querying of the image archives utilizing data content, metadata and semantic categories In the system the processing chain starts with the ingestion of Earth-Observation products, followed by knowledge discovery and data mining processes to obtain semantic annotations and, finally, to enable the end-user to perform multiple queries based on different parameters and compute analytics about the data content. In the following, we briefly describe the main components of ITS-DM.

In the Data Model Generation, the main steps performed are: 1) Metadata extraction by reading the annotation file and retrieving the relevant metadata entries as for example: locations, angles, mission information, etc. 2) Image tiling and quick-look generation in order to create a multiresolution tile pyramid (several grid levels with tiles of different size); 3) Feature extraction, which is based on tiles. When applying tile-based feature extraction, the methods consider the whole patch as one entity, using the relations between all pixels within the patch and not only single pixel information. The second component: Data Mining Database is a relational Database Management System (DBMS), which acts as the core of the system, interacting with all user-oriented components and supporting their functionality. The use of a DBMS provides some advantages such as a natural integration of different kinds of information, ensuring the integrity of relations, speed of operations, etc. The Knowledge Discovery in Databases (KDD) component is used for the semantic definition of the image content. This component is based on semi-supervised learning methods and comprises relevance-feedback in order to include human expertise in the annotation process. The last component, the Query Engine (QE) allows the end-user to find desired scenes and to retrieve images containing the required content. The image archive can be queried by using metadata and semantic definitions. In the case of queries based on given metadata, metadata entries such as geographical coordinates, acquisition angles and time, product types, etc. are used as query parameters. In the query engine is possible to compute analytics about the results of the query as well as about the content of the image archive.

Co-funded by the European Union

6 Fusion modules and data integration from multiple sources.

6.1 Non probabilistic fusion techniques (DELTARES)

This section, briefly describe a number of non-probabilistic data fusion methods. Data fusion systems are often complex combinations of sensor devices, processing and fusion algorithms. At the end of this section we highlight few example applications that bring out these features.

6.1.1 Description of the module

Multisensor data fusion is the process of combining observations from a number of different sensors to provide a robust and complete description of an environment or process of interest. The first step of this module is to associate pixels from different layers (data sets) to the target grid. Figure 20 shows the schematization of three basic approaches.

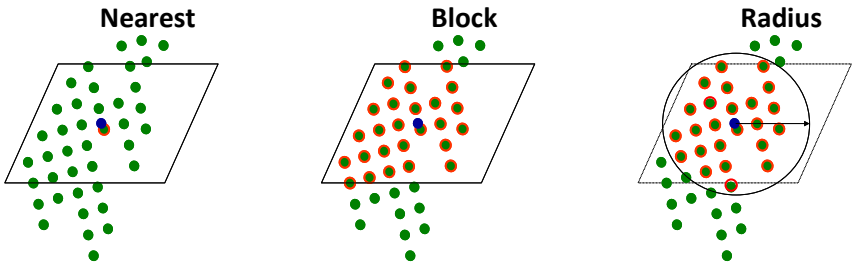


Figure 20 Schematization of the four basic approaches to association pixel data onto a grid. The parallelogram indicates a single grid cell. Green dots are pixel values form a single layer; red circles indicate which of these pixels are included in computing the representative value for the geographic target grid cell (blue circle). This association of pixels is done for the n layers under consideration.

6.1.1.1 Weighted averaging.

Distance Weighting is a type of deterministic method for multivariate interpolation with a known scattered set of points. The assigned values to unknown points are calculated with a weighted average of the values available at the known points. The name given to this type of methods was motivated by the weighted average applied, since it resorts to the inverse of the distance to each known point ("amount of proximity") when assigning weights.

The weighted average equation and some fundamental properties are described below.

$$\overline{X} = \sum_{i=1}^{N} w_i X_i$$

The sum of all the weights is equal to 1.

$$\sum_{i=1}^{n} w_i = 1$$

An average is greater than or equal to the smallest number and smaller than or equal to the largest number. Assuming wi ≥ 0 for all i,

$$X_{min} \leq \overline{X} \leq X_{max}$$

In addition, an overall average is equal to a weighted average of group sub-averages1.

The average could be considered as the center of gravity. Thus if a leverage is placed under the average value, then one can achieve a balance.

Co-funded by the European Union

As *wi* increases, $\overline{X} \rightarrow Xi$. The greater the value of the weight, the closer the average tends to the number in the series associated with the weight.

The weights given to each pixel in the modules presented here can be based on e_i and d_i. Where e_i is the estimation of the error for pixel i, and d_i is the distance between the target grid cell and pixel i.

6.1.1.2 Weighted Ensemble Mean

The Weighted Ensemble Mean is a variation of the weighted mean. This method assumes that the errors are normally distributed with standard deviation provided by the retrieval errors e_i. Then for each pixel sample n number of samples from that distribution and compute the weighted average and a new standard deviation for the ensemble errors. (See Blaas et al. 2012 for details.)

6.1.1.3 Nearest neighbor

This method assigns the value of the nearest point to each grid node. This method is useful when data are already evenly spaced, but need to be converted to a SURFER grid file. Alternatively, in cases where the data are close to being on a grid, with only a few missing values, this method is effective for filling in the holes in the data.

6.1.1.4 Multiple linear regression

Linear regression is an approach for modelling the relationship between a scalar dependent variable and one or more explanatory variables. If only one explanatory variable is used, the approach is called simple linear regression. If more than one explanatory variable is used, the approach is called multiple linear regression. The relationships are modelled using linear predictor functions whose unknown parameters are estimated from the data. These are called linear models. A linear model looks as follows:

$$Y = \beta_0 + \sum_{k=1}^{p} \beta_k X_k + \varepsilon$$

where Y is the dependent or response variable, Xk stands for the independent or predictor variable(s), β0 stands for the regression intercept, βj stands for the regression coefficient(s) and ε stands for the residual, also called the error term. The regression intercept is the expected mean value of Y when all X are zero. The regression coefficients represent the mean change in the dependent variable for one unit of change in the independent variable while holding other predictors constant.

6.1.1.5 Kriging

Kriging attempts to express trends suggested in the data, so that, for example, high points might be connected along a ridge rather than isolated by bull's-eye type contours. Kriging is a very flexible gridding method

The basic tool of geostatistics and kriging is the semivariogram. The semivariogram captures the spatial dependence between samples by plotting semivariance against separation distance. The premise of any spatial interpolation is that close samples tend to be more similar than distant samples (this is also called spatial autocorrelation). This property of spatial data is implicitly used in IDW. In kriging, one must model the spatial autocorrelation using a semivariogram instead of assuming a direct, linear relationship with separation distance. Semivariance equals one-half the squared difference between points separated by a distance d±Δd (assuming no direction preference). As the distance between samples increase, we expect the semivariance to also increase (again, because near samples are more similar than distant samples). This is true, however, only up to some given separation distance. For this distance and up, points are unrelated. Stated another way, if 50m is this critical separation distance, two points separated by 50m are likely to be just as similar (or dependent on one another) as samples separated by 100, 200, 300, or any distance greater than 50m.

The inclusion of secondary information, in the present work is summarise as followed (Serpa, 1998): Using the related secondary information with the main attribute (Z's (x)), known in all the study area the final estimated cartography Z*s (x) is equal to the kriged map Z*ok(x) * plus a deviation e(x):

$$Z_s'(x) = Z^*_{ok}(x)^* + e^*(x)$$

Co-funded by the European Union

Where e*(x) is the deviation values between the real values of the secondary variable (Z s (x)) and the kriged values Z*s(x) from the secondary information:

$$e(x) = \left[Z_s(x) - Z^*_s(x) \right]$$

Z*s (x) is the value kriged with the secondary data (soft) in the points where the data referring to the main information had been collected Xa:

$$Z^*_S(x) = \sum_a \beta_a Z_S(x_a)$$

The final map of estimated values Z*ok(x), shows always in the location of the experimental supports, the main information (index). Reversibly, when main information does not exist takes advantage the secondary information.

6.1.2 Experimental Results

Chlorophyll concentration provides an estimate of the live phytoplankton biomass in the surface layer. It is also referred to as the amount of the green pigment chlorophyll. For this example, Medium Resolution Imaging Spectrometer(MERIS) recordings of the southern North Sea are used. From the spectra of these recordings water quality data have been retrieved with the algorithm described in Doerffer (2007). MERIS for coastal applications are the spatial resolution of 300m.
Also the L2 ocean color products (chlorophyll-a concentration) generated by Moderate Resolution Imaging Spectroradiometer (MODIS)is used. Products derived from MODIS sensor have approximately 1 km-per-pixel resolution. The plotted values from MODIS and MERIS are presented in Figure 21.

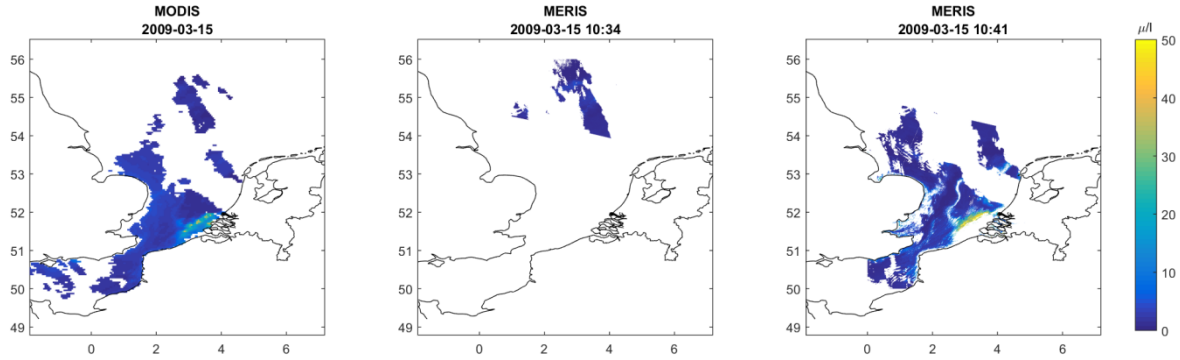


Figure 21 Chlorophyll-a values (µ/l) from MODIS[1] and MERIS for 15th March 2009.

For the fusion on the data sets above presented the Block schematization was chosen, meaning that the data used at each grid point are those pixels lying strictly with the cell grid. The results of testing the different fusion methods are presented in Figure 22.

[1]NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group; (2014): Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Ocean Color Data, NASA OB.DAAC. http://doi.org/10.5067/ORBVIEW-2/SEAWIFS_OC.2014.0. Accessed on 2016/02/29.

Co-funded by the European Union

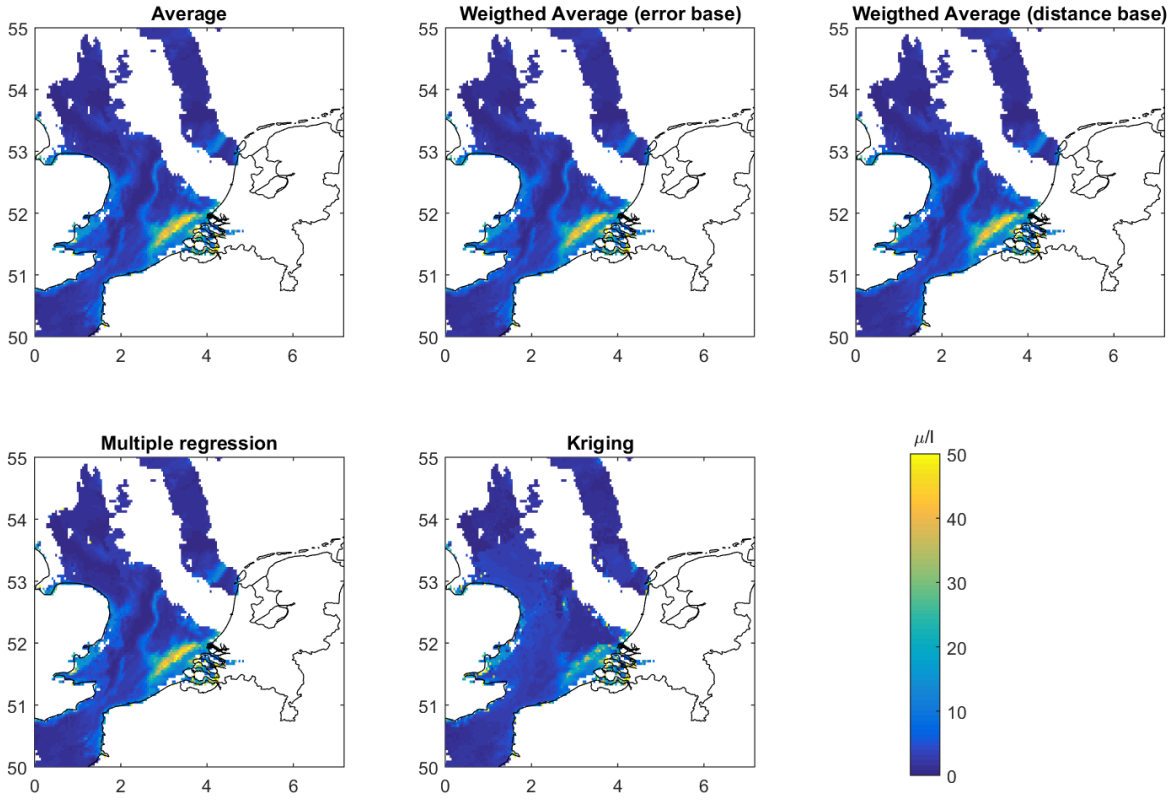


Figure 22 Different fusion technique for chlorophyll-a values in the North Sea area.

Another example is in the Xynthia area. Xynthia is a part of the Bay of Biscay at the Southwest coast in France. It is the area between 45.5 and 46.5 degrees North and 0.7 and 1.7 degrees West. As shows the figure below, different sets of bathymetry data are available for this area.

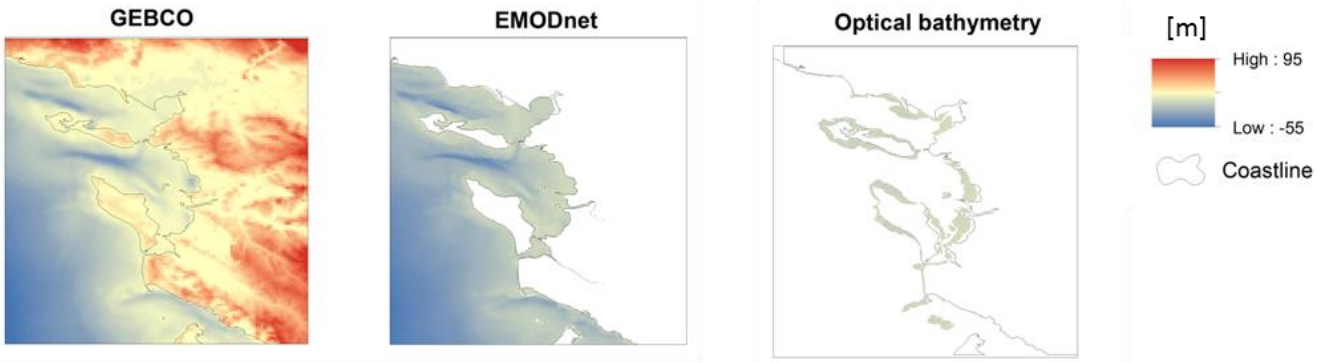


Figure 23 Different bathymetry datasets available in the Xynthia area: GEBCO_2014 Grid is presented at 30 arc second intervals of latitude and longitude, EMODnet with a resolution of 1/8*1/8 arc minutes, which is around 230 meters at the equator and bathymetry derived from Landsat 8 optical imagery by applying the MIP algorithm

The results of testing the different integration methods and validating the results are presented in Figure 24. It includes a validation of the individual base datasets that were available as well as a validation of the results of the different integration methods. The validation result is presented in the table below.

Co-funded by the European Union

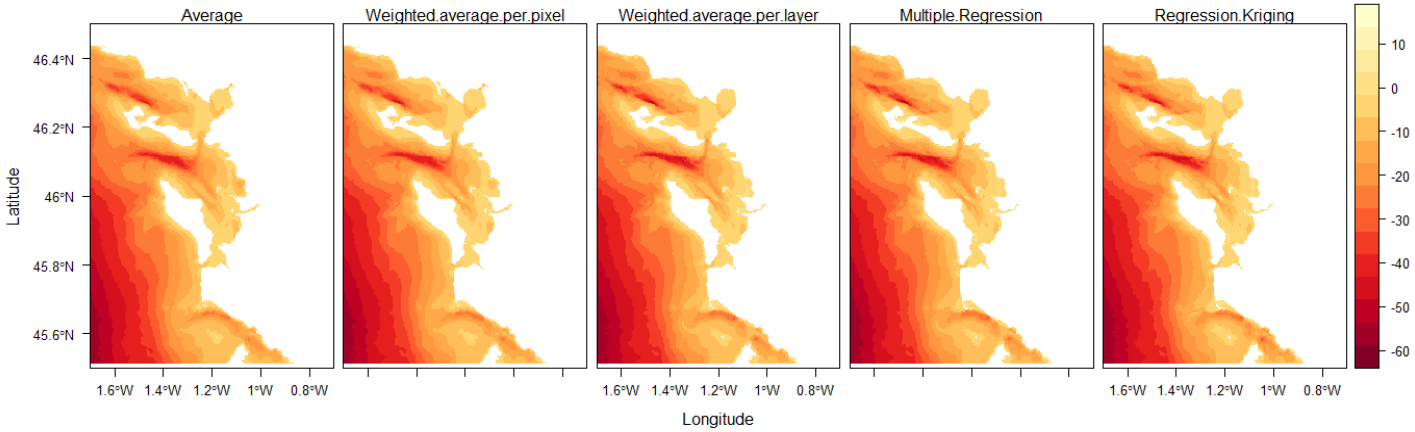


Figure 24 Different fusion technique for the Xynthia area: Average, weighed average per pixel , weighed average per layer, multiple regression and Kriging.

For this example, Table.1 shows that the Regression Kriging is the technique that produces the lowest error after the merging of the data. This is also the only technique that generates an output where the error is lower than in all of the input bathymetry datasets. Therefore Regression Kriging seems to be the technique that leads to the highest accuracy when combining those data sets

Data	RMSE	Error Mea
Average	1.3773286	-0.10440215
Weighted average per pixel	1.2608655	-0.09335774
Weighted average per layer	1.2137457	-0.07757859
Multiple regression	1.2164039	-0.10827557
Regression Kriging	0.7170053	-0.05279199

Table.1 Validation of different fusion technique.

6.1.3 Flow chart

The inputs for the module are different layers or images of the same variable with some overlapping area. The resolutions/coarseness can vary but the geographical reference must be the same for all input layers. So far the module support only NetCDF files but in principle can be extended to support GeoTIFF format for example. The flowchart of the module is presented in Figure 18.

Co-funded by the European Union

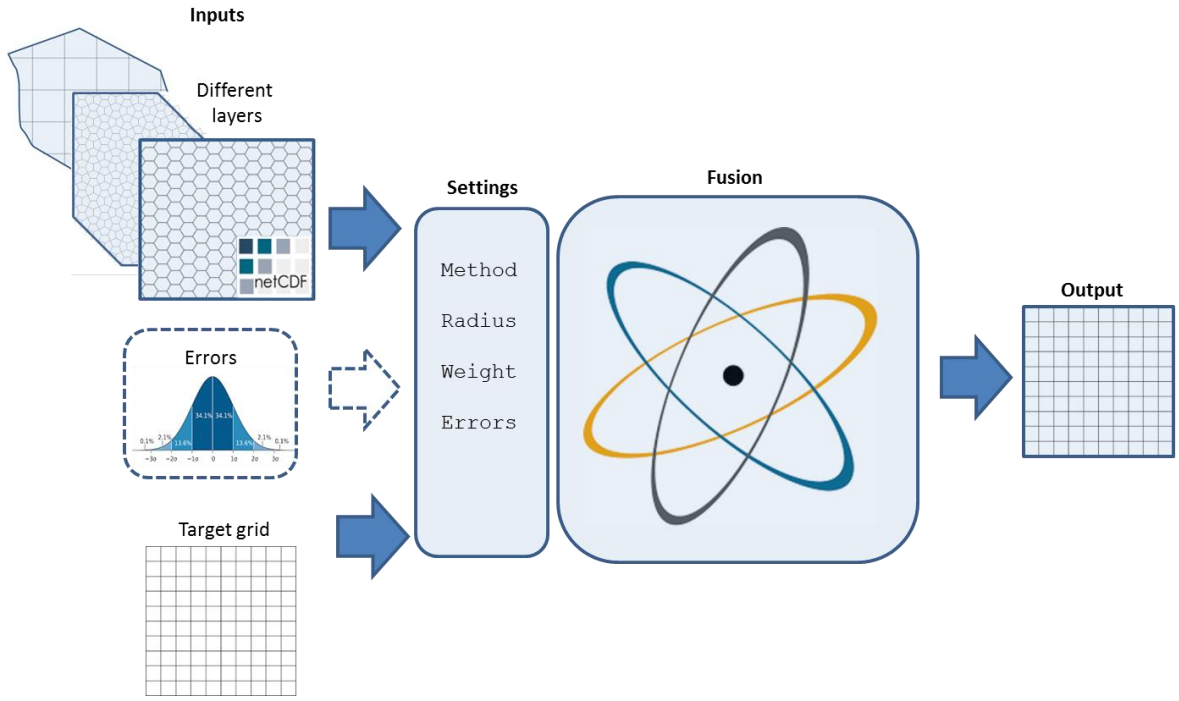


Figure 25 Flowchart for the fusion data module

For certain fusion methods it is required provide an estimation of the errors at the pixel level, or a "general" estimation of the error per layer over the all domain. Therefore if case of selecting weighted average based on errors, an extra layer of errors information should be provided also as an input.
The user must provide a target grid. This is the final grid where all layers will be projected and merge. This target grid can be one of the input layers or a complete new grid. The final product is a single layer with the merged data.

6.1.4 References

[1] Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., ... & Ladner, R. (2009). Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_PLUS. *Marine Geodesy*, *32*(4), 355-371.

[2] Danielson, J. J., &Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)* (No. 2011-1073).US Geological Survey.

[3] R. Doerffer& H. Schiller (2007) The MERIS Case 2 water algorithm, International Journal of Remote Sensing, 28:3-4, 517-535, DOI: 10.1080/01431160600821127

[4] EMODnet Lot 1 Bathymetry (2015) *Guidelines for metadata, data and DTM QA/QC*.Retrieved February 3, 2016, from http://www.emodnet-bathymetry.eu/documents/emodnet_hydrography/html_page/qa_qc_dtm_specifications_04022015.pdf.

[5] Federico Castanedo, "A Review of Data Fusion Techniques," The Scientific World Journal, vol. 2013, Article ID 704504, 19 pages, 2013. doi:10.1155/2013/704504

[6] Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barrand, N. E., Bell, R., ... & Catania, G. (2013). Bedmap2: improved ice bed, surface and thickness datasets for Antarctica. *The Cryosphere*, *7*(1).

[7] Jawak, S.D., Vadlamani, S.S. and Luis, A.J. (2015) A Synoptic Review on Deriving Bathymetry Information Using Remote Sensing Technologies: Models, Methods and Comparisons. *Advances in Remote Sensing, 4*, 147-162. http://dx.doi.org/10.4236/ars.2015.42013

[8] Pleskachevsky, A., Lehner, S., Heege, T., & Mott, C. (2011).Synergy and Fusion of Optical and Synthetic Aperture Radar Satellite Data for Underwater Topography Estimation in Coastal Areas.*Ocean Dynamics, 61(12)*, 2099-2120.

Co-funded by the European Union

[9] Smith, W. H. F., & Wessel, P. (1990).Gridding with continuous curvature splines in tension.*Geophysics*, *55*(3), 293-305.

6.2 Integration of in-situ data in Linear Spectral Mixing Analysis (LSMA) for the onshore coastal sediment characterization. (ISPRA)

6.2.1 Description of the module

Beach dune systems are important for coastal zone ecosystems: they represent natural sea defenses by dissipating wave energy and incrementing coastal protection. Synoptic maps of site-specific sediment composition, grain size and moisture content are a key input in ecological modeling. The extraction of such information requires adequate bio-geophysical and empirical models and analysis techniques, like Linear Spectral Mixing Analysis (LSMA).

Hyperspectral or multispectral data, integrated with in-situ measurements of hyperspectral radiometry, grain size, natural water content and mineralogical composition, can be used to estimate the spatial distribution of onshore coastal sediment characteristics (Innocenti et al., 2013; Manzo et al., 2015, Valentini, 2013).

The integrated approach showed in Figure 26 makes use of hyperspectral data (MIVIS), radiometrically and spatially corrected, with in-situ spectral signatures of collected sediment samples, supplied with laboratory analysis of mineralogical and moisture content. The workflow is divided in 3 main steps.

1) Hyperspectral data are atmospherically corrected and signal-to-noise estimation is used to remove bands whose radiance energy was low in relation to atmospheric absorption bands.
2) As water content in beach sediments is the main factor reducing the reflectance and modifying the spectral properties, water content characterization is performed in order to spatial divide the hyperspectral dataset in subsets. In-situ data of natural water content are compared with airborne hyperspectral data using multivariate linear regression analysis in order to generate a moisture map and remove the effect of light absorbance of water in beach sediments. Following the same approach, in-situ data of sediment grain size are compared to the airborne hyperspectral data to generate maps of grain size distribution.
3) LSMA technique (Small 2009; Taramelli et al., 2013; Valentini et al., 2015; Manzo et al., 2015) combines the linear spectral mixing space concept within hyperspectral or multispectral data, in order to identify the spectral endmembers representative of the spectral variability and identify their fraction of abundance for each pixel. The spectroscopy of sediments is analyzed in the mixing space obtained by Principal Component Analysis (PCA) of the image bands.
4) Spectral libraries of in-situ collected radiometry can be used either as spectral endmember in LSMA or as validation tool to characterize the spectral properties of the spectral endmember identified from the mixing space. Mineralogical composition, measured from sediment samples using XRD (Carbone et al., 2011), is finally used to characterize each spectral endmember and convert spectral fraction maps to minerals abundance maps.

Co-funded by the European Union

6.2.2 Flow chart

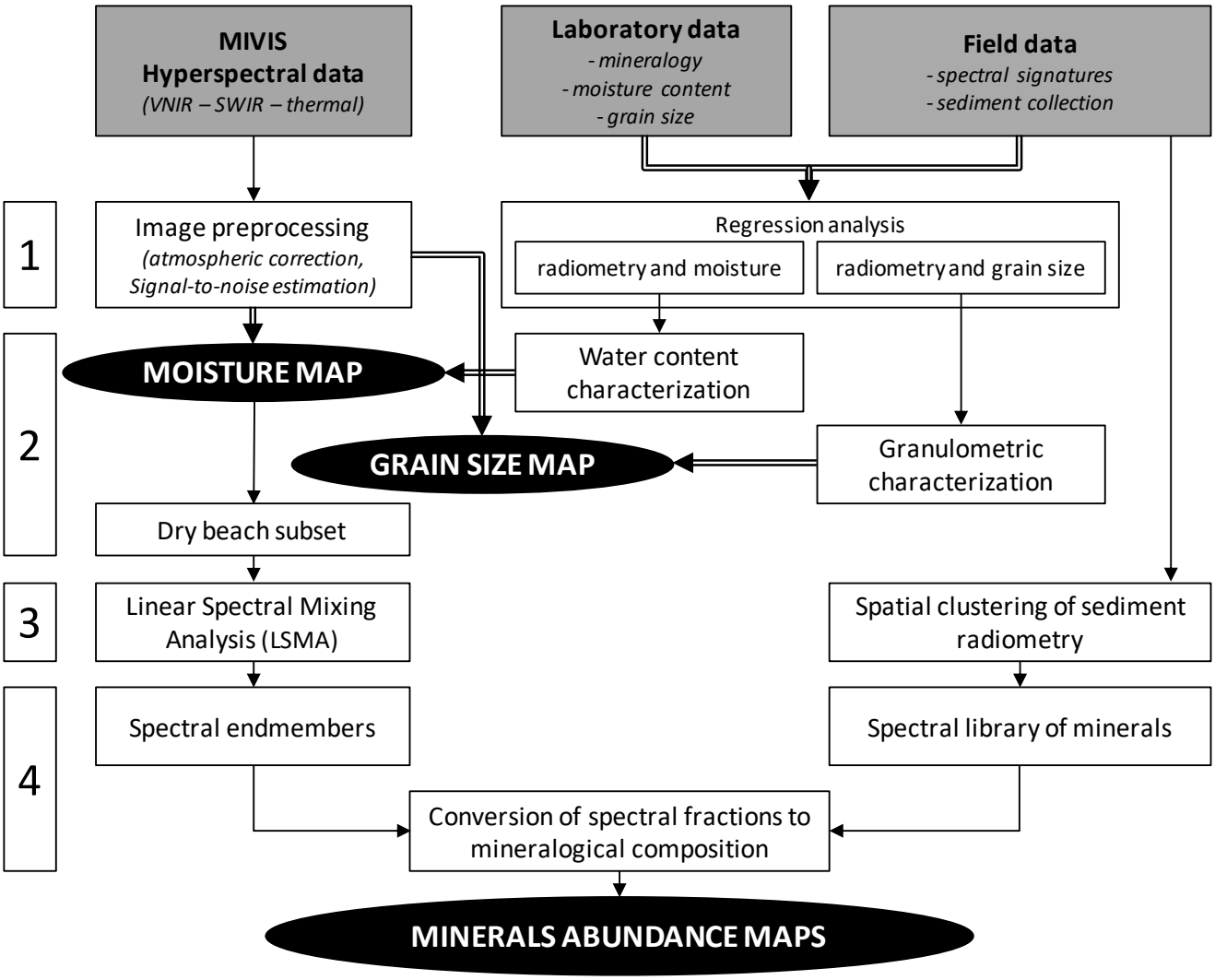


Figure 26. Workflow for mineralogical characterization of the beach sediment using a hyperspectral data integration tool. (Edited from: Manzo et al., 2015).

6.2.3 Experimental results

The integration of in-situ measurements with airborne hyperspectral data was successfully applied for the characterization of onshore coastal sediment in the area of Sabaudia (Italy) (Innocenti et al., 2013; Manzo et al., 2015).

Field measurement included the acquisition of spectral signatures of 350 coastal sediment and the collection of 57 samples to identify grain size, mineralogical composition and natural water content.

MIVIS hyperspectral data were atmospherically corrected using ATCOR4® based on the MODTRAN radiative transfer model (Matthew et al.,2000), and signal-to-noise estimation was used to remove bands whose radiance energy was low in relation to atmospheric absorption bands.

Water content (W%) is defined as the amount of water in the pore of the sediment, and it is the physical property characterized by soil moisture. The radiometric interval 1.45-1.70 µm and the MIVIS thermal band at 8.34 µm were used as the reference for sediment moisture analysis, focusing on a comparison of in-situ samples and MIVIS hyperspectral data (Manzo et al., 2015). Moisture information was used to generate a spatial subset from

MIVIS data, representing area of dry sediments, that was further analysed using empirical models and LSMA technique.

Grain size was spatially estimated by means of multiple linear regression model from in-situ samples (dependent variable) and the MIVIS data (independent variable), generated for all the MIVIS band combination and limited to two to avoid problem of overfitting (Innocenti et al., 2013). Laboratory and field analysis provided an empirical estimation to characterize sediment grain size and the moisture distribution related to natural water content, which masks the spectral features related to other sediment characteristics such as mineral composition or grain size. As result of the water content and granulometric analysis, moisture map and grain size map were generated from airborne MIVIS hyperspectral data (Figure 27).

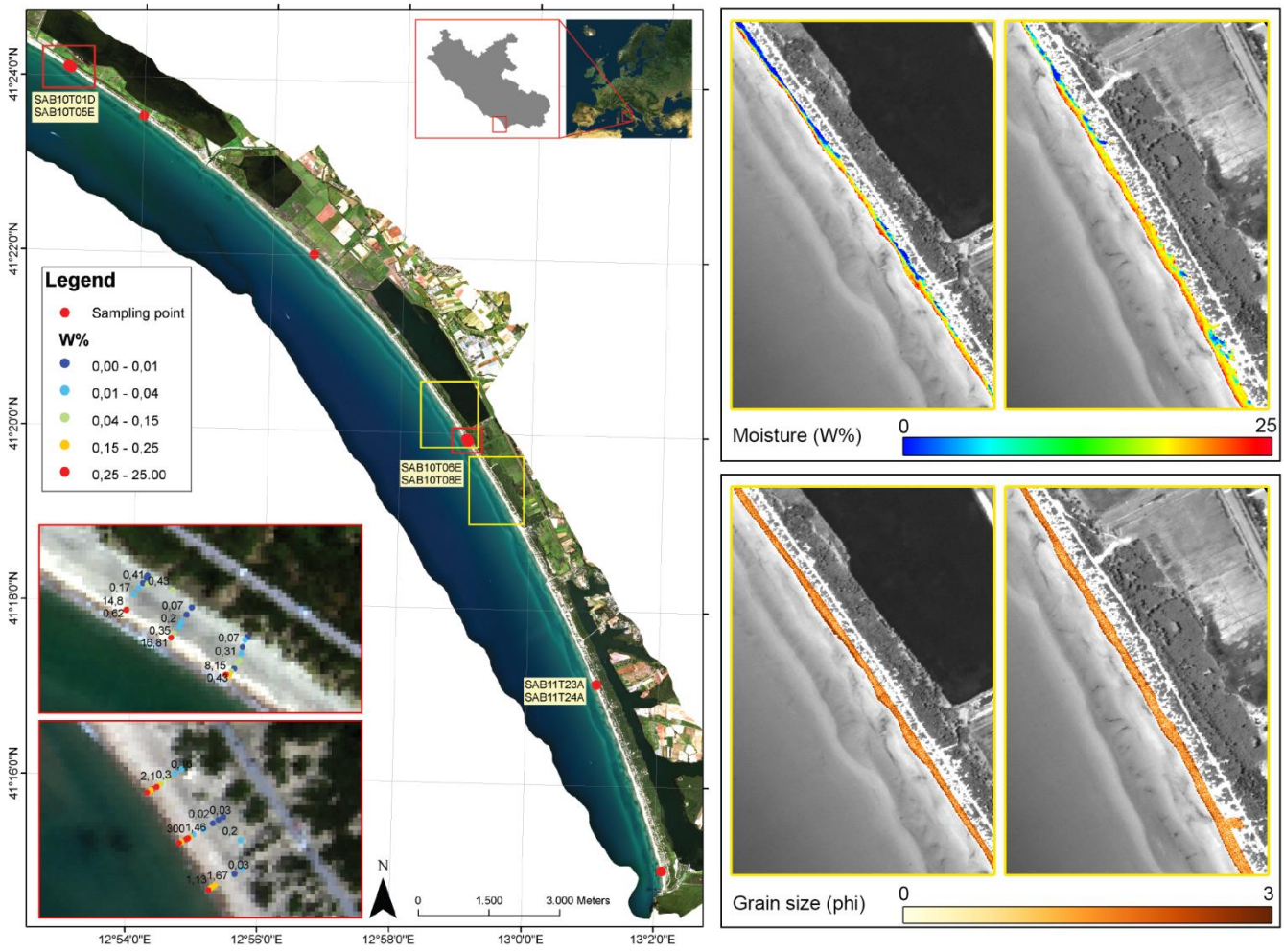


Figure 27. Sediment moisture and grain size estimated by integrating in situ measurement on beach sampling points with airborne MIVIS hyperspectral data.

LSMA technique was applied to the spatial subset from MIVIS data in order to identify the minerals abundance within the onshore coastal sediments.

Two significant spectral signatures were considered as representative for the characterization of coastal sediments from the MIVIS data in the study area, and were selected as endmembers in the LSMA analysis for the retrieval of abundances maps. The application of LSMA to the dry beach subset could be used to define the spatial distribution of spectral endmembers, characterized by different mineralogical compositions. The integration with the mineralogic composition measured using XRD was finally used to derive minerals abundance maps.

6.2.4 References

[1] Carbone, C., Marescotti, P., Lucchetti, G., Cauzid, J., Chalmin, E., 2011. Application of synchrotron radiation-based techniques (μ-XRD,-XRF, and μ-XANES) to study Fe-rich hardpans within waste-rock dump. Neues Jahrbuch fur Mineralogie-Abhandlungen 188 (1), 21–30.

[2] Innocenti, C., Filipponi, F., Valentini, E., Taramelli, A., 2013. Multisensory DataFusion Methods for the Estimation of Beach Sediment Features: Mineralogi-cal, Grain size and Moisture. IEEE, pp. 3064–3067, http://dx.doi.org/10.1109/IGARSS.2013.6723473.

[3] Matthew, M.W., Adler-Golden, S.M., Berk, A., Richtsmeier, S.C., Levine, R.Y.,Bernstein, L.S., et al., 2000. Status of atmospheric correction using a MODTRAN4-based algorithm. Proc. SPIE 4049, 199–207, http://dx.doi.org/10.1117/12.410341.

[4] Small, C., 2004. The Landsat ETM+ Spectral Mixing Space, Remote Sensing of Environment, 140, pp. 131-145.

[5] Taramelli, A., Valentini, E., Innocenti, C., Cappucci, S., 2013. FHYL: Field SpectralLibraries, Airborne Hyperspectral Images and Topographic and BathymetricLidar Data for Complex Coastal mapping. IEEE, pp. 2270–2273, http://dx.doi.org/10.1109/IGARSS.2013.6723270.

[6] Valentini E., 2013. PhD Thesis : 'A new paradigm in coastal ecosystem assessment: linking ecology and geomorphology implementing the FHyL approach.' http://hdl.handle.net/2067/2800

[7] Valentini E., Taramelli A., Filipponi F., Giulio S., 2015. An effective procedure for EUNIS and Natura 2000 habitat type mapping in estuarine ecosystems integrating ecological knowledge and remote sensing analysis. Ocean & Coastal Management, 108: 52-64, doi:10.1016/j.ocecoaman.2014.07.015.

6.3 Integration of expert knowledge, satellite multispectral and in-situ radiometry to map habitat types. (ISPRA)

6.3.1 Description of the module

Satellite optical multispectral images can be integrated with in-situ data and expert knowledge to generate maps of habitat types, following habitat classification standards (Valentini et al., 2015). Figure 28 shows the workflow for habitat mapping from the integration of satellite optical multispectral data, in-situ data and expert knowledge.

Satellite optical multispectral data are first accurately georeferenced and radiometrically corrected. In-situ spectral signatures are collected and spectrally resampled in order to match the spectral sensitivity of the satellite multispectral sensor. The spectral resampling is a procedure that convert the radiance response at given set of wavelengths to a target set of wavelengths, using a Gaussian model with Full-Width-Half-Maximum (FWHM) equal to target wavelengths spacing.

Habitat types mapping workflow is composed by 6 main steps (Figure 28):

1) First image analysis step is preprocessing (radiometric calibration, data reprojection) and a preliminary discrete classification using Maximum Likelihood Algorithm (MLLH), that is necessary to reduce the spectral heterogeneity of the image dataset.
2) A Principal Component Analysis (PCA) of the multispectral image is then performed to reduce the numerical variability of each pixel on the basis of its spectral composition. The resulting mixing space features the most pure pixels at the apexes of the components axis and the mixed pixels (mixture of cover typologies) along the trajectories between the apexes.
3) Supervised classification step, using Maximum Likelihood Algorithm (MLLH) method, is applied to the image in order to reduce the spectral variability.
4) Linear Spectral Mixture Analysis (LSMA), a spectral sub-pixel analysis, is then used to solve the linear combination model of mixed spectral contributions using spectral endmembers. The spectral

endmembers are selected from the multispectral mixing space (Small, 2004; Valentini, 2015; Valentini et al., 2015), and represent the most spectrally representative cover types.

5) Classification of cover types is performed from the abundance fractions generated using LSMA by means of a decision tree. Each pixel is classified on the base of its abundance values, and it is assigned to a natural target based on the strongest Pearson correlation score obtained between the correspondent pixel spectral profile and the in-situ spectral library. The decision tree was selected as the most flexible instrument to translate the expert knowledge from botanists and ecologists to a image processing analytical tool. It allows the quantification of specific thresholds across cover abundance maps to represent the relevant components of the environmental variability.
6) By consulting and interpreting the standard habitat classification databases, a table for the conversion of the standard product of cover classes can be used to generate the added value product of the habitat types. Validation of the results is done using a confusion matrix (Congalton, 1991; Congalton and Green, 1999) for the comparison of habitat type map and in-situ samplings.

6.3.2 Flow chart

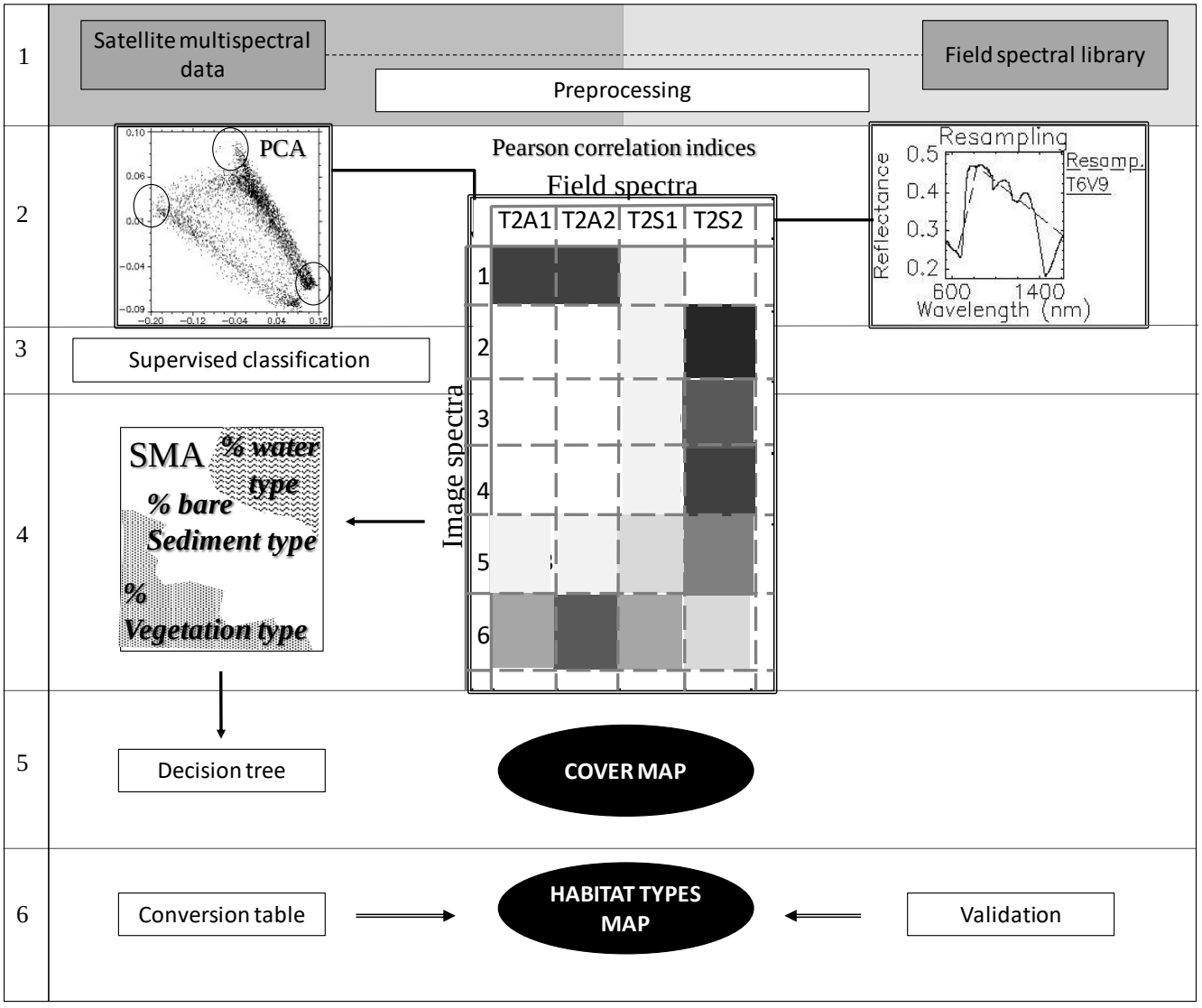


Figure 28. Standardized processing procedure for habitat types mapping integrating satellite optical multispectral data, in-situ data and expert knowledge. (Edited from: Valentini et al., 2015).

Co-funded by the European Union

6.3.3 Experimental results

Habitat types mapping was successfully done integrating satellite optical multispectral data, in-situ data and expert knowledge for the Po Delta (Italy) area (Valentini et al., 2015).

A multispectral satellite image was acquired on 11/09/2012 by the High Resolution Visible and InfraRed (HRVIR) sensor aboard the SPOT-4 (Satellite Pour l'Observation the la Terre), with spatial resolution 20 m. The image was geo-referenced and radiometrically calibrated to obtain reflectance values.

Field spectral measurements of sand and mixed vegetation targets of terrestrial and aquatic environment were acquired, specifically different sediment types, vegetation groups, intertidal and very shallow waters.
In addition, vegetation types and the relative cover percentage were identified to support the taxonomic identification. Sediment and water samples were analysed in laboratory to obtain information about Chl-a, degraded pigments, dissolved organic carbon and coloured dissolved organic matter content.
SPOT multispectral data was analyzed using Maximum Likelihood Algorithm (MLLH). PCA and LSMA analysis were used to generate fractional abundances from the multispectral data. The resulting cover map (Figure 30a) is composed of thirteen different cover classes obtained with the decision tree based on the fractional abundance from LSMA. Root Mean Square error was calculated for each image pixel, and reported in Figure 30b.

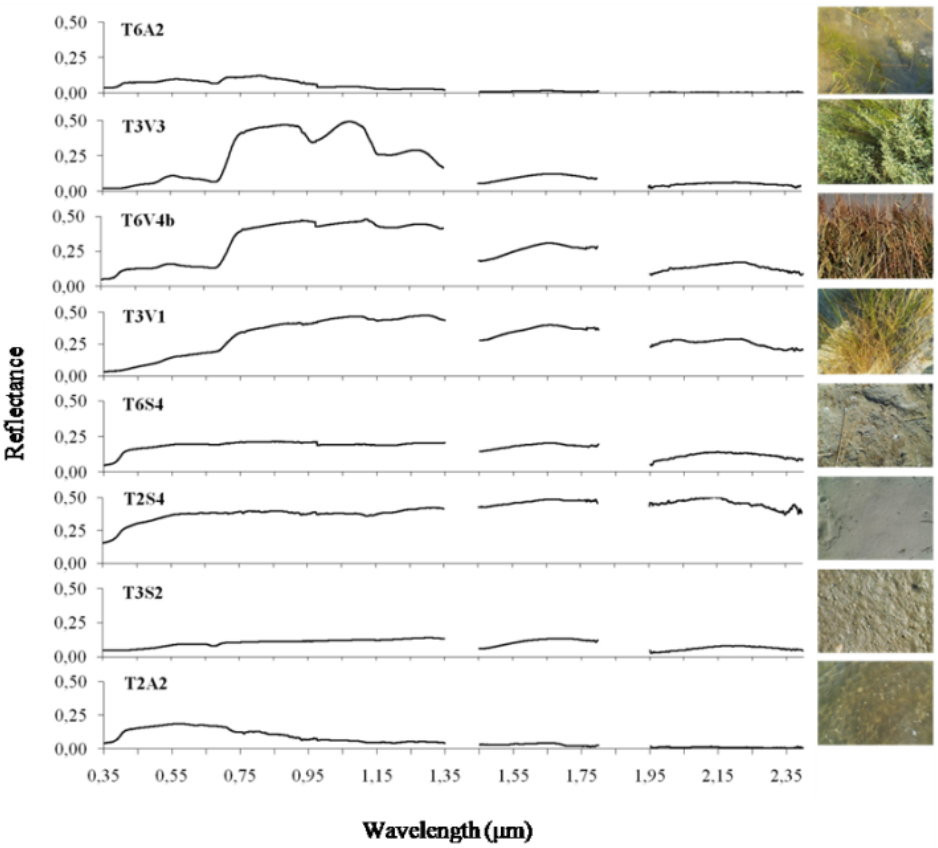


Figure 29. In situ spectral library

Co-funded by the European Union

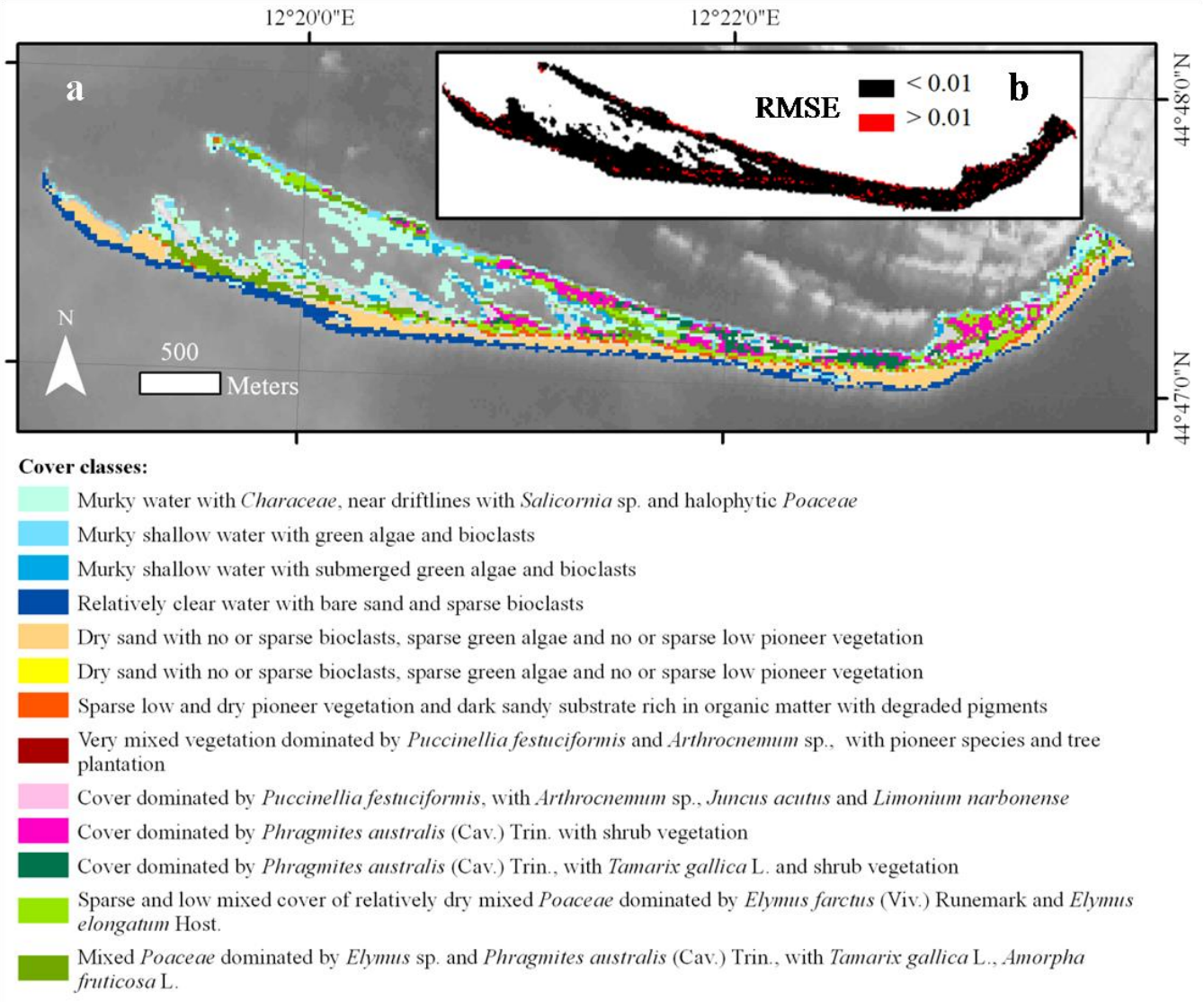


Figure 30. a) Cover map. b) Spatial distribution of Root Mean Square Error. (From: Valentini et al., 2015).

The identification of the composition of the mixed pixels was obtained by evaluating the Pearson correlation score between the field spectra and the representative spectra collected from the image (Figure 29).
The basic knowledge of a large scale description of the study area from the analyzed SPOT image can also be associated to classes of EUNIS and Natura 2000 habitat types. Each habitat type is associated with one or more cover classes based on the expert knowledge interpretation from botanists and ecologists to identify the contribution of more than one field observation.
An example of the applicability of this integration tool and the described classification analysis is showed in Figure 31, representing the potential nesting area of an endangered bird species, *Sterna albifrons* classified as 'Least Concern' in the IUCN Red List of Threatened Species. Nesting areas are derived from the classified cover map and correspond to sandy spits not inundated during high tides and characterized by the presence of sparse vegetation.

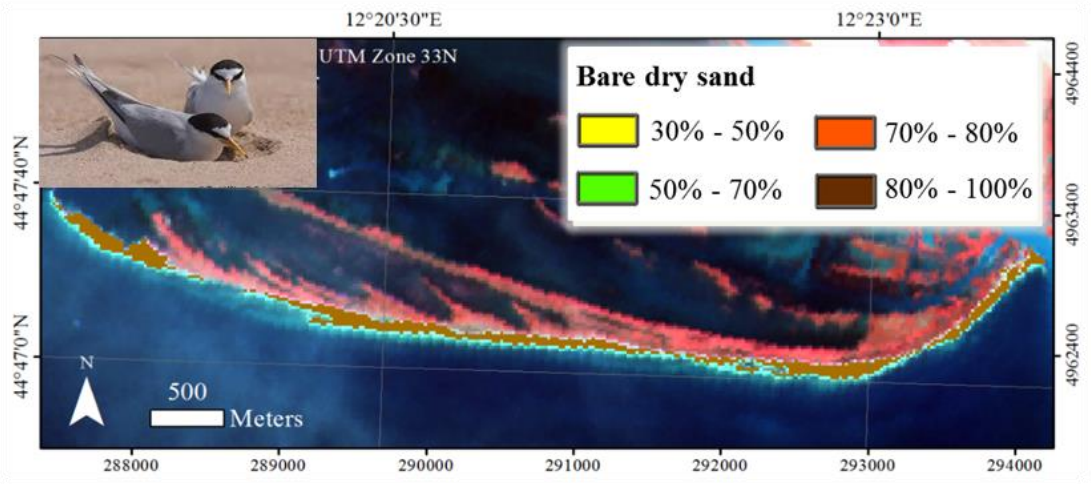


Figure 31. Potential nesting areas of *Sterna albifrons.*

6.3.4 References

[1] Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37 (1), 35-46.

[2] Congalton, R.G., Green, K., 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Lewis Publishers, Boca Raton, FL.

[3] Small, C., 2004. The Landsat ETM+ Spectral Mixing Space, Remote Sensing of Environment, 140, pp. 131-145.

[4] Valentini E., 2013. PhD Thesis : 'A new paradigm in coastal ecosystem assessment: linking ecology and geomorphology implementing the FHyL approach.' http://hdl.handle.net/2067/2800

[5] Valentini E., Taramelli A., Filipponi F., Giulio S., 2015. An effective procedure for EUNIS and Natura 2000 habitat type mapping in estuarine ecosystems integrating ecological knowledge and remote sensing analysis. Ocean & Coastal Management, 108: 52-64, doi:10.1016/j.ocecoaman.2014.07.015.

6.4 LiDAR bathymetry integration with multisensory radiometry for seabed mapping in shallow waters. (ISPRA)

6.4.1 Description of the module

Seagrass cover about 0.1-0.2% of the seafloor worldwide, but has a key role in coastal ecosystem (Duarte, 2002). Seagrass meadows are, in fact, among the most relevant components of marine coastal ecosystems (Koch et al., 2006) and represent a major focal point for both research and conservation strategy (Short et al., 2011). The presence of *Posidonia oceanica* meadow on sand stabilizes superficial sediment and contrasts resuspension of particles under the effect of waves and currents, increasing water transparency and mitigating coastal erosion processes. Deterioration of the meadow under the effect of human impacts affects the stability of the upper limit of *Posidonia oceanica* and the integrity of the meadow at shallow water.

The ability of mapping the seabed is depending on the radiometric characteristics of its reflectance and the water optical properties (i.e. transparency), while the quantification of cover types is dependent on the optical attenuation of the signal due to the water column properties (Silva et al., 2008, Lyzenga, 1985). In order to perform an accurate classification of the seabed cover from RS data, the influence of water column in reducing the electromagnetic energy should be taken into account, by decorrelating spectral response of seabed targets from the water column.

Co-funded by the European Union

This integration tool makes use of bathymetric information, derived either from LiDAR data processing or multibeam acquisitions, to correct optical multi and hyper-spectral data from water column signal attenuation and improve classification performances for seabed mapping purpose. Seabed mapping is done using a combination of optical multi and hyper-spectral data and in situ data with Linear Spectral Mixing Analysis (LSMA) technique. The workflow of this integration tool is divided in four main steps (Figure 32).

1) Optical multi and hyper-spectral data are preprocessed using atmospheric correction modules. Later the submerged portion of the image is masked and bright pixels, reporting higher reflectance values, are identified. The dataset is normalized using logarithmic transformation, in order to facilitate the finding of a linear relation between reflectances and the corresponding bathymetric values.
2) Data integration module consists in three parts: threshold identification, linear regression analysis and slope estimation for data rotation. Threshold identification is done by plotting reflectances values against the bathymetric values and finding the bathymetry minimum and maximum threshold values for which a good correspondence of the two parameters is obtained. Linear regression analysis is done in order to find the numerical relation between each band reflectance, the response variable, and the bathymetric value, the explanatory variable. Later the slope of the regression line generated for each spectral band is estimated.
3) The estimated slope is used to rotate the reflectance data in order to correct water column signal attenuation in both the optical multi and hyper-spectral data and in situ radiometry.
4) Linear Spectral Mixture Analysis (LSMA), a spectral sub-pixel analysis, is then used to solve the linear combination model of mixed spectral contributions using spectral endmembers. The spectral endmembers are selected from the multispectral mixing space, obtained by Principal Component Analysis (PCA) of the image bands (Small, 2004), or derive by spectrally resampling in situ radiometry (Valentini et al., 2015). LSMA technique identifies the most spectrally representative cover types and estimates their fraction of abundance for each pixel. Abundance fractions are finally classified using a decision tree in order to generate the seabed map.

Co-funded by the European Union

6.4.2 Flow chart

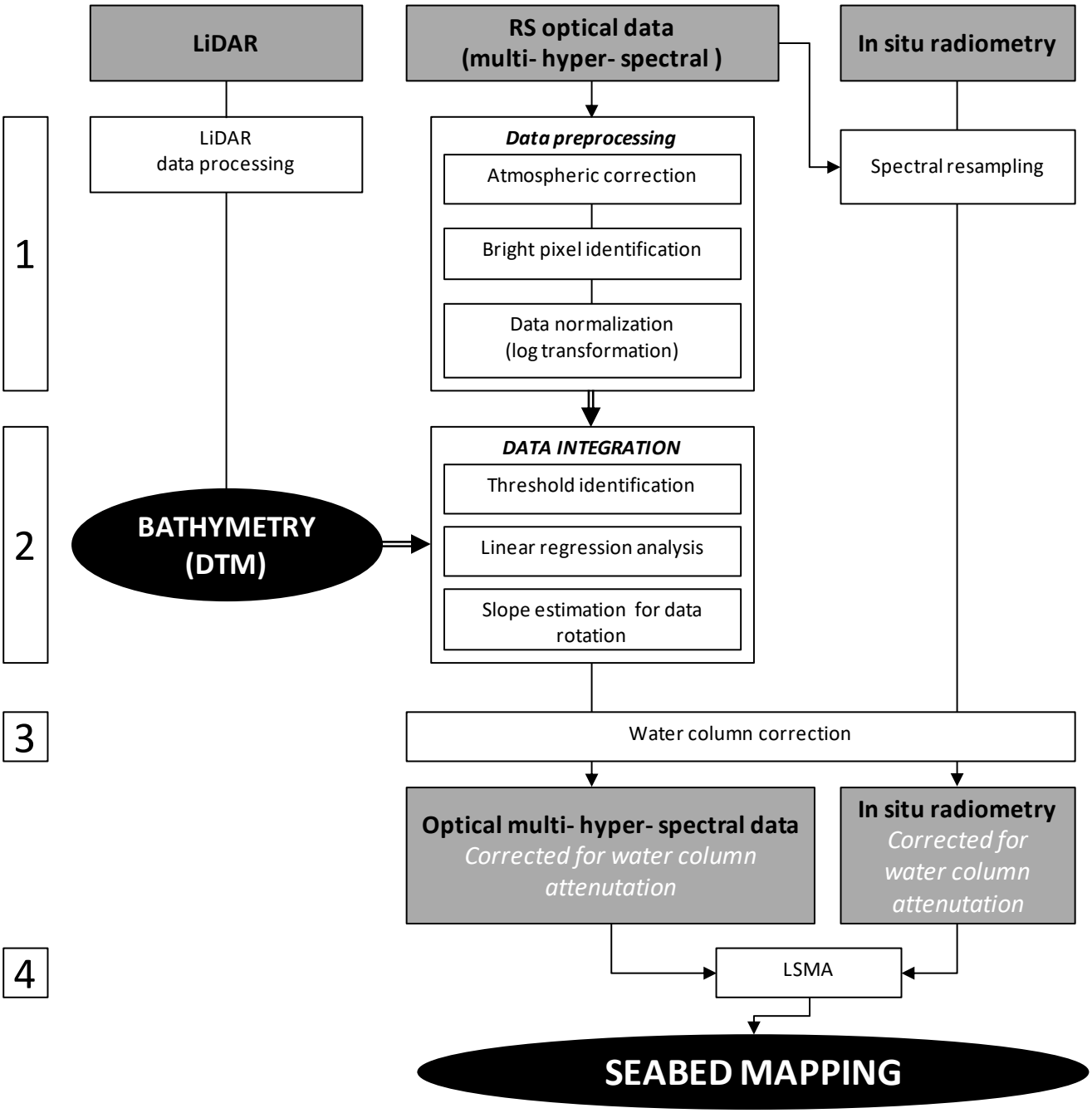


Figure 32. Seabed mapping classification workflow. DTM product generated from LiDAR active sensor is integrated with optical multi- hyper- spectral data and in situ radiometry to correct data for water column attenuation.

6.4.3 Experimental results

Seabed mapping was successfully performed integrating simultaneous acquisitions of airborne optical hyperspectral data, LiDAR and in-situ radiometry in two coastal sites in Italy: Favignana (Cappucci et al., accepted), Capo Rizzuto.

LiDAR data (acquired using Hawk Eye instrument) have been processed in order to obtain the bathymetry (Digital Terrain Model) of the area and perform water column correction. MIVIS hyperspectral data (20 spectral bands

Co-funded by the European Union

in the interval 0.44 – 0.81 µm, spatial resolution 3 m) were successfully corrected for signal optical attenuation (Figure 33). Lower reliability limit of seabed cover types classification was found to be -9 m (Favignana) and -15 m (Capo Rizzuto). At lower bathymetries the ability of classification was limited to the identification of bright seabed (e.g. sandy seafloor) and dark seabed (e.g. *Posidonia oceanica* meadows, rocky seafloor). Four seabed cover classes have been identified in the study sites (Figure 34b), and corresponds to: sandy bottom, seaweed, sparse seagrass, dense seagrass.

Co-funded by the
European Union

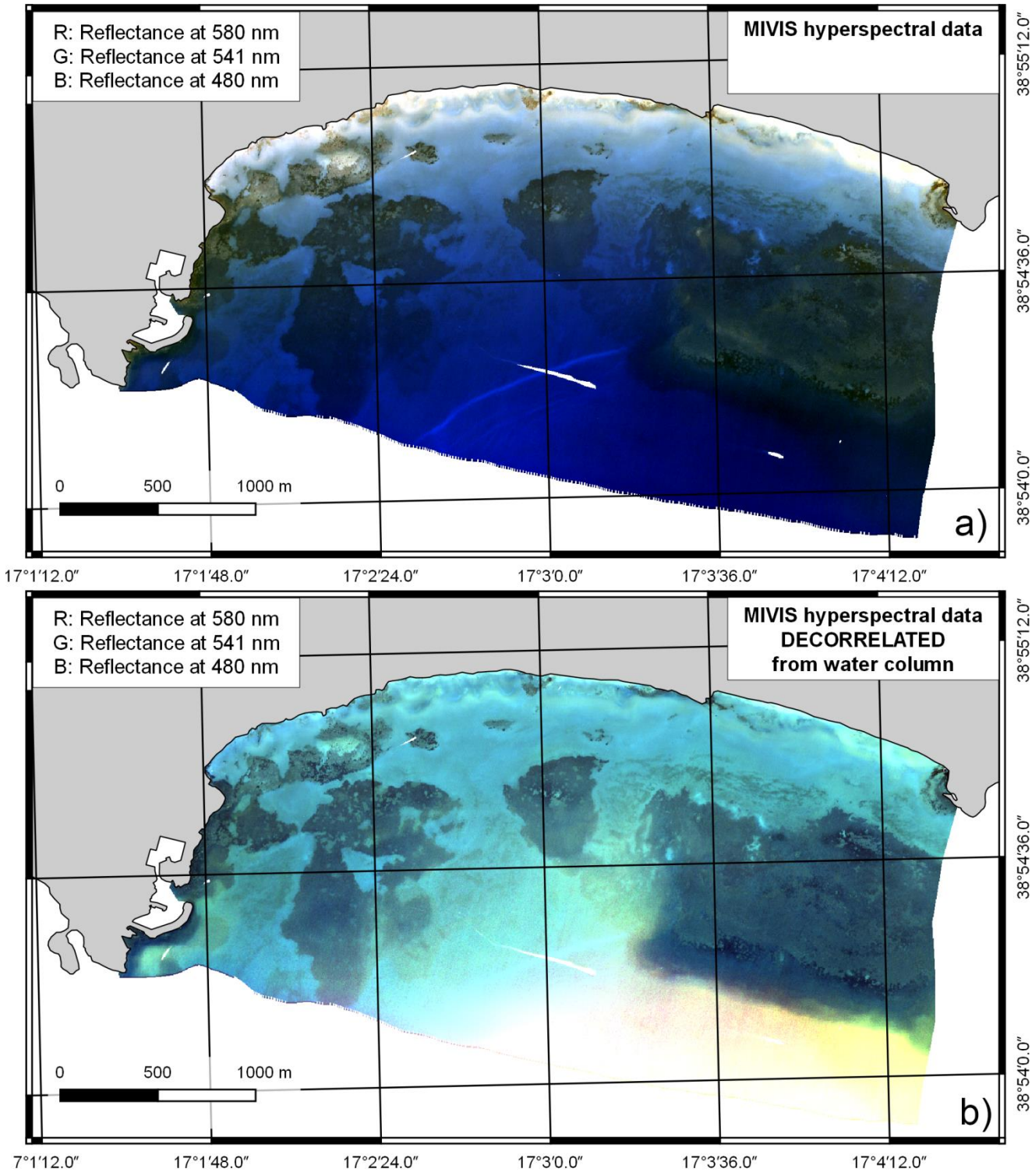


Figure 33. a) MIVIS hyperspectral data of Marine Protected Area Of Capo Rizzuto (Italy). b) MIVIS hyperspectral data decorrelated from water column to normalize the effect of signal optical attenuation.

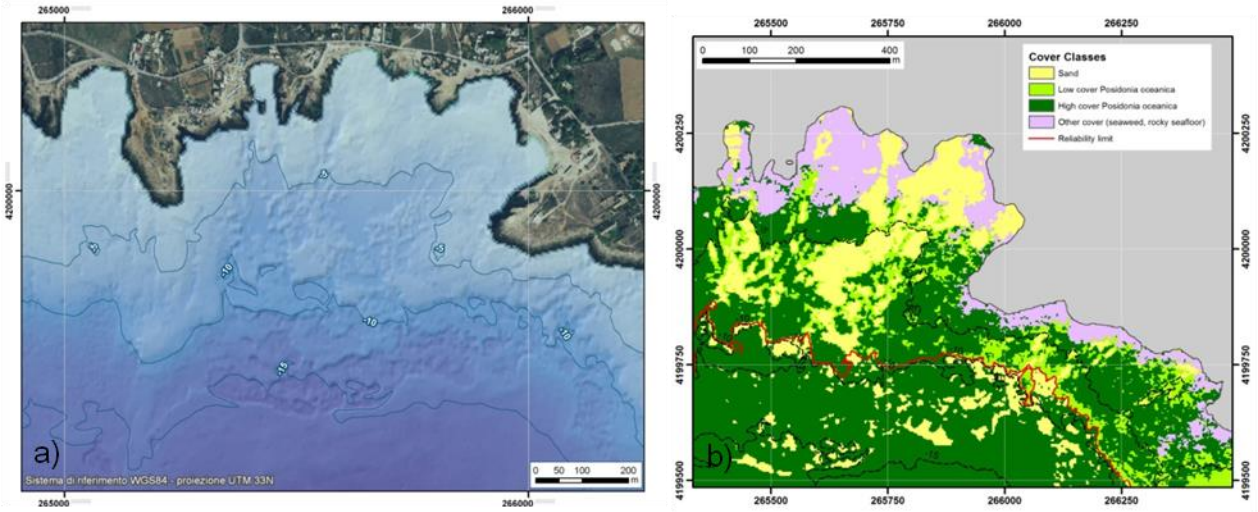


Figure 34. a) Bathymetric map of Favignana study site (Italy). b) Classified seabed map of Favignana study site (Italy). (From: Cappucci et al., accepted).

6.4.4 References

[1] Cappucci S., Valentini E., Del Monte M., Paci M., Filipponi F., Taramelli A., accepted. A novel framework for detection of natural and anthropic features.

[2] Duarte C.M. (2002). The future of seagrass meadows. Environmental Conservation, Vol.29, Issue 2: 192-206.

[3] Koch, E.W.; Sanford, L.P.; Chen, S.N.; Shafer, D.J., and Smith, J.M., 2006. Waves in seagrass systems: review and technical recommendations (No. ERDC-TR-06-15). Maryland University Cambridge Center for Environmental Science.

[4] Lyzenga, D.R., 1985. Shallow-water bathymetry using combined lidar and passive multispectral scanner data. International Journal of Remote Sensing, 6(1), 115-125.

[5] Short, F.T.; Polidoro, B.; Livingstone, S.R.; Carpenter, K.E.; Bandeira, S.; Sidik Bujang, J.; Calumpong, H.P.; Carruthers, T.J.B.; Coles, R.G.; Dennison, W.C.; Erftemeijer, P.L.A.; Fortes, M.D.; Freeman, A.S.; Jagtap, T.G.; Kamal, A.H.M.; Kendrick, G.A.; Judson Kenworthy, W.; La Nafie, Y.A.; Nasution, I.M.; Orth, R.J.; Prathep, A.; Sanciangco, J.C.; van Tussenbroek, B.; Vergara, S.G.; Waycott, M., and Zieman, J.C. (2011). Extinction risk assessment of the world's seagrass species. Biological Conservation, 144(7), 1961-1971.

[6] Silva, T.S.; Costa, M.P.; Melack, J.M., and Novo, E.M., 2008. Remote sensing of aquatic vegetation: theory and applications. Environmental monitoring and assessment, 140(1-3), 131-145.

[7] Small, C., 2004. The Landsat ETM+ Spectral Mixing Space, Remote Sensing of Environment, 140, pp. 131-145.

[8] Valentini E., Taramelli A., Filipponi F., Giulio S., 2015. An effective procedure for EUNIS and Natura 2000 habitat type mapping in estuarine ecosystems integrating ecological knowledge and remote sensing analysis. Ocean & Coastal Management, 108: 52-64, doi:10.1016/j.ocecoaman.2014.07.015.

[9] Valentini E., 2013. PhD Thesis : 'A new paradigm in coastal ecosystem assessment: linking ecology and geomorphology implementing the FHyL approach.' http://hdl.handle.net/2067/2800

Co-funded by the European Union

6.5 Maximum-dissimilarity algorithm (MDA) as fusion module to downscale waves on shallow waters from SAR wind fields, modeled wind and waves fields (ISPRA)

6.5.1 Description of the module

Wave downscaling in coastal area can benefit from the high spatial resolution (300-1000 m) of wind fields estimated from satellite Synthetic Aperture Radar (SAR), that can be used as forcing in the wave reconstruction. Wave downscaling is done using a hybrid methodology (Figure 35) that consists on the selection of a set of wave climate cases by means of the Maximum Dissimilitude Algorithm, the propagation of these cases, and finally the reconstruction of time series by means of Radial Basis Functions.

1) First, a calibration step corrects significant wave heights with instrumental data from satellite according to the mean wave direction (Mínguez et al., 2011).
2) Then a selection of representative scenarios of ocean conditions of the total database is done applying a Maximum Dissimilarity Algorithm (MDA), which has been proved to identify the most dissimilarity wave conditions on a reanalysis database including the extreme events (Camus et al., 2011a).
3) Propagation of the representative cases of wave climate, selected by MDA, are propagated to coastal areas using the numerical model Simulating WAves Nearshore (SWAN; Booij et al., 1999).
4) Finally the reconstruction of the time series of wave parameters on the DOW grid is done by means of a Radial Basis Functions (RBF) interpolation (Camus et al., 2011b).
5) Validation of the reconstructed wave time series is done using wave data acquired from coastal buoys.

Co-funded by the European Union

6.5.2 Flow chart

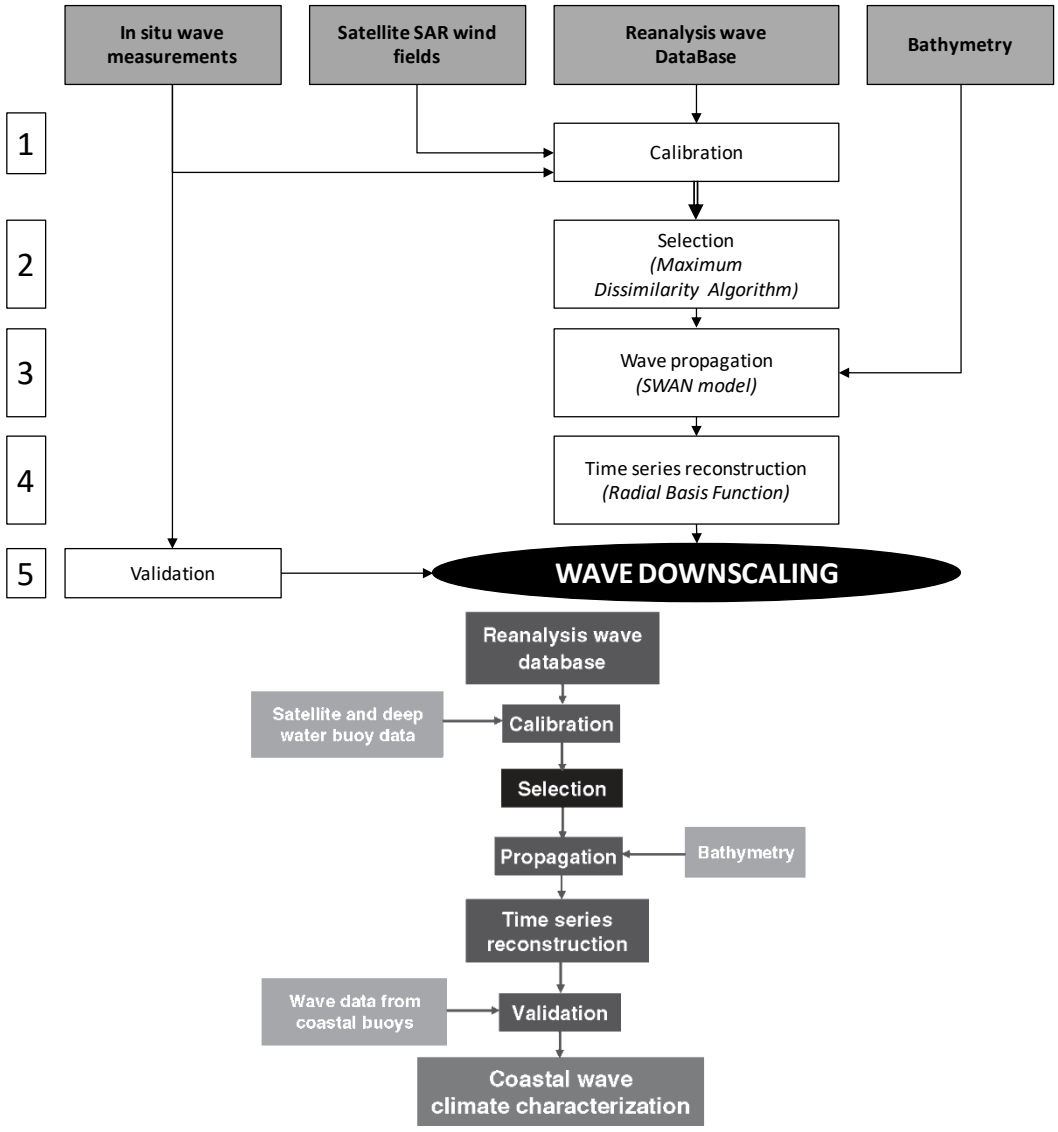


Figure 35. Coastal wave climate characterization workflow. The fusion module uses MDA to downscale waves on shallow waters from SAR wind fields, modeled wind and waves fields. (Edited from: Gutiérrez et al., 2016).

6.5.3 Experimental results

The fusion methodology of SAR wind fields and modeled wind fields for wave downscaling in coastal area was successfully applied to northern Adriatic Sea (Gutiérrez et al., 2016) using a selection method based on MDA (Figure 36).

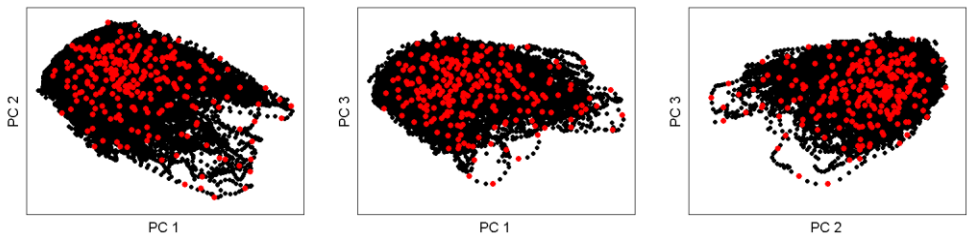


Figure 36. Selection of cases using MDA. (From: Gutiérrez et al., 2016).

Co-funded by the European Union

As inputs were used the GOW Mediterranean (Reguero et al., 2012), a wind field product generated using atmospheric models with a spatial resolution of about 20 km, the Seawind I database (Menéndez et al., 2013) with a spatial resolution of about 30 km, and estimated SAR wind fields, acquired by ENVISAT-ASAR with a spatial resolution of about 300 m (Figure 37). Wave climate definition for the open boundary of downscaling was obtained from GOW database.

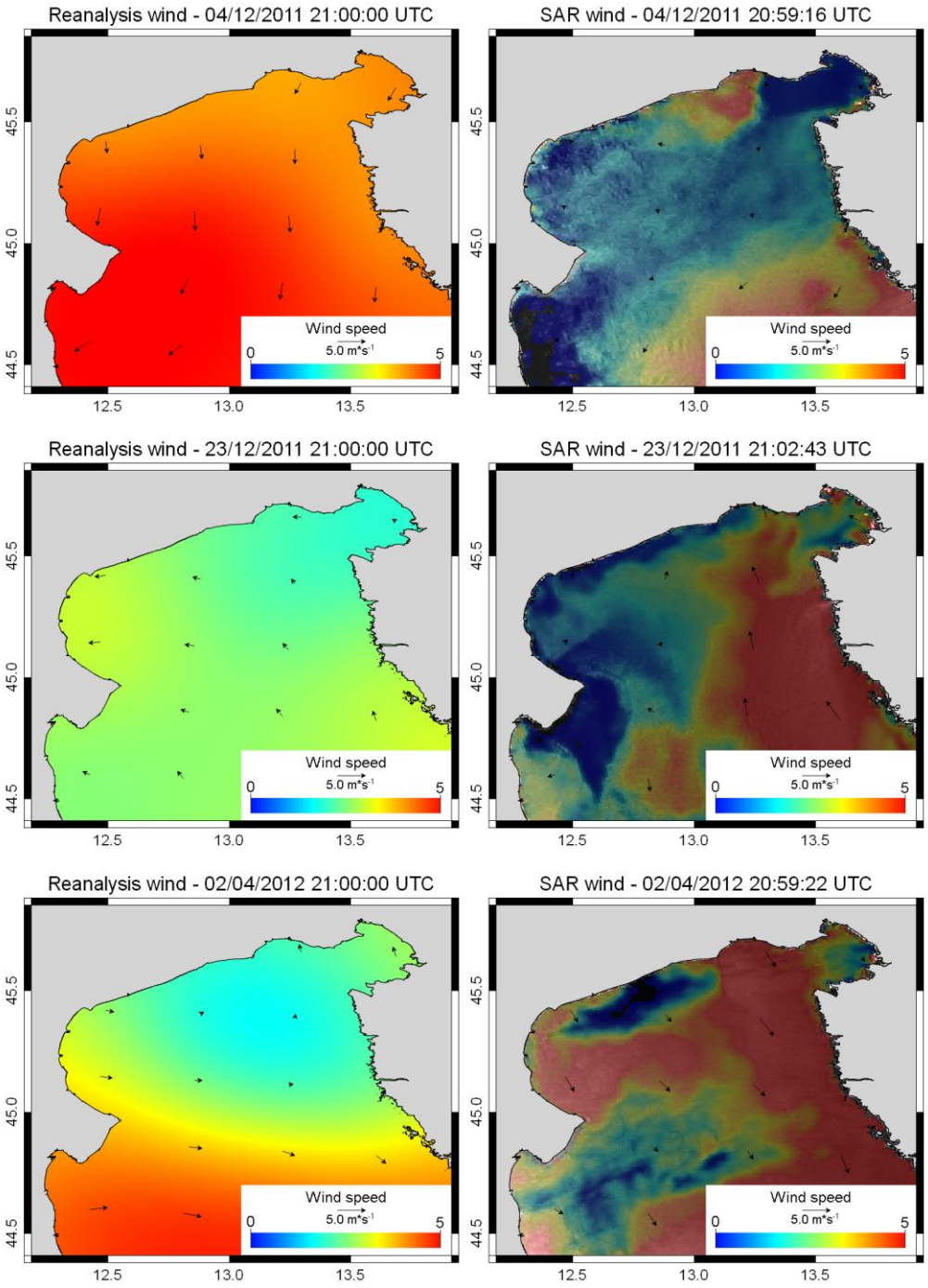


Figure 37. Comparison of reanalysis wind and SAR wind fields used as input to force wave downscaling. (From: Gutiérrez et al., 2016).

SAR wind fields were used in combination with modeled wind fields to force the numerical model and produce wave fields. These simulations were also forced with the corresponding wave climate through the open boundary.
Comparison with in situ instrumental data indicates the general good quality of the downscaled waves (Figure 38).

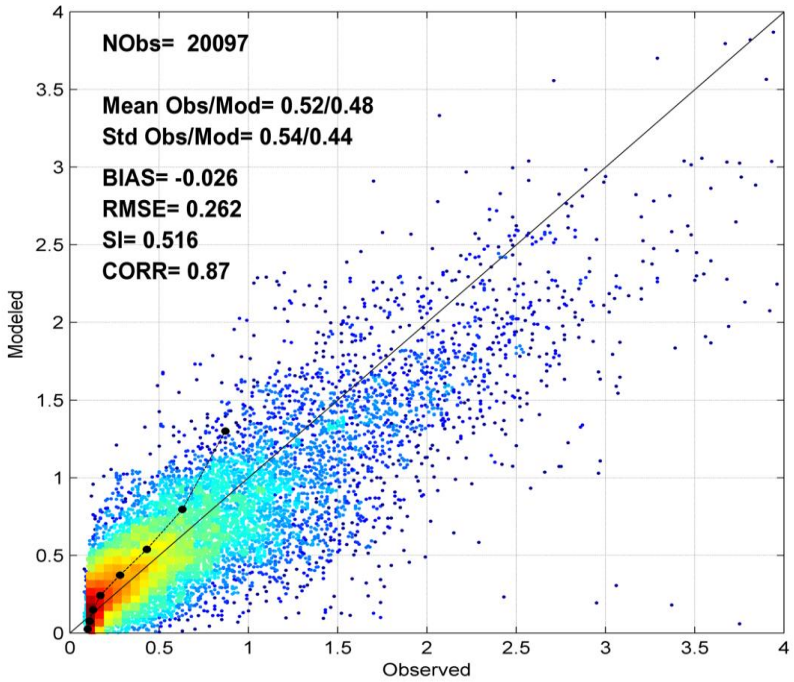


Figure 38. Comparison of observed and modeled Significant Wave Heights. (From: Gutiérrez et al., 2016).

Although there are differences between SAR and modeled wind fields, a good correlation was found on the downscaled waves forced with different wind fields compared with measured significant wave heights from buoy (Figure 38). In comparison with previous experimental research (Camus et al., 2011b), results carried out using the described data fusion using a hybrid methodology show the ability of SAR satellite data to force time series of wave fields by means of a radial basis functions (RBF) interpolation.

6.5.4 References

[1] Booij, N., Ris, R. C., and Holthuijsen, L. H.: A third-generation wave model for coastal regions. Part I: model description and validation, J. Geophys. Res., 104, 7649–7666, 1999.
[2] Camus, P., Méndez, F. J., Medina, R., and Cofiño, A. S.: Analysis of clustering and selection algorithms for the study of multivariate wave climate, Coast. Eng., 58, 453–462, 2011a.
[3] Camus, P., Méndez, F. J., and Medina, R.: A hybrid efficient method to downscale wave climate to coastal areas, Coast. Eng., 58, 851–862, 2011b.
[4] Gutiérrez O.Q., Filipponi F., Taramelli A., Valentini E., Camus P., Méndez F.J., 2016. On the feasibility of the use of wind SAR to downscale waves on shallow water. Ocean Science, 12, pp. 39-49, doi:10.5194/os-12-39-2016.
[5] Menéndez, M., García-Díez, M., Fita, L., Fernández, J., Méndez, F. J., and Gutiérrez, J. M.: High-resolution sea wind hindcasts over the Mediterranean area, Clim. Dynam., 42, 1857–1872, doi:10.1007/s00382-013-1912-8, 2013.
[6] Mínguez, R., Espejo, A., Tomás, A., Méndez, F. J., and Losada, I. J.: Directional calibration of wave reanalysis databases using instrumental data, J. Atmos. Ocean. Tech., 28, 1466–1485, doi:10.1175/JTECH-D-11-00008.1, 2011.

Co-funded by the European Union

7 Assimilating data from INSPIRE spatial datasets and Copernicus core services/products (ISPRA)

7.1 Integration of High Resolution Layers (HRL) of Copernicus Land Monitoring Services with Corine Land Cover (CLC) and national data sets

Copernicus is a European system for monitoring the Earth and consists of a complex set of systems which collect data from multiple sources: earth observation satellites and in situ sensors such as ground stations, airborne and sea-borne sensors. Copernicus provides users with reliable information through a set of services related, among others, to environmental issues.

One of the Copernicus services, the Land Monitoring Service, provides Land Cover data for the Pan-European component at 20 m resolution. Pan-European High Resolution Layers (HRL) provide information on specific Land Cover characteristics, using a classification system that is still considering only general themes as cover types, that in this section is named Level 1. Five themes have been identified for Copernicus HRL Level 1, corresponding to the level of sealed soil (imperviousness), tree cover density and forest type, (semi-) natural grasslands, wetlands and permanent water bodies.

The integration tool here presented, makes use of Corine Land Cover, data about population and OpenStreetMap database to generate a National Land Cover Map of Level 2 starting from Copernicus HRL of Level 1.

Data are integrated in a classification tree that is using rules to supply the HRL Level 1 class with more details and generate a Level 2 nomenclature for the Italian National area.
As result, a National Land Cover Map of Level 2 is generated with the following cover classes: main road network, residential areas, sport and leisure facilities, mineral extraction sites, dump sites, construction sites, other impervious areas, permanent crops, broadleaved forests, coniferous forests, agricultural areas, green urban areas, natural grasslands, wetlands, permanent water bodies, dunes and sands, unclassified.

Co-funded by the European Union

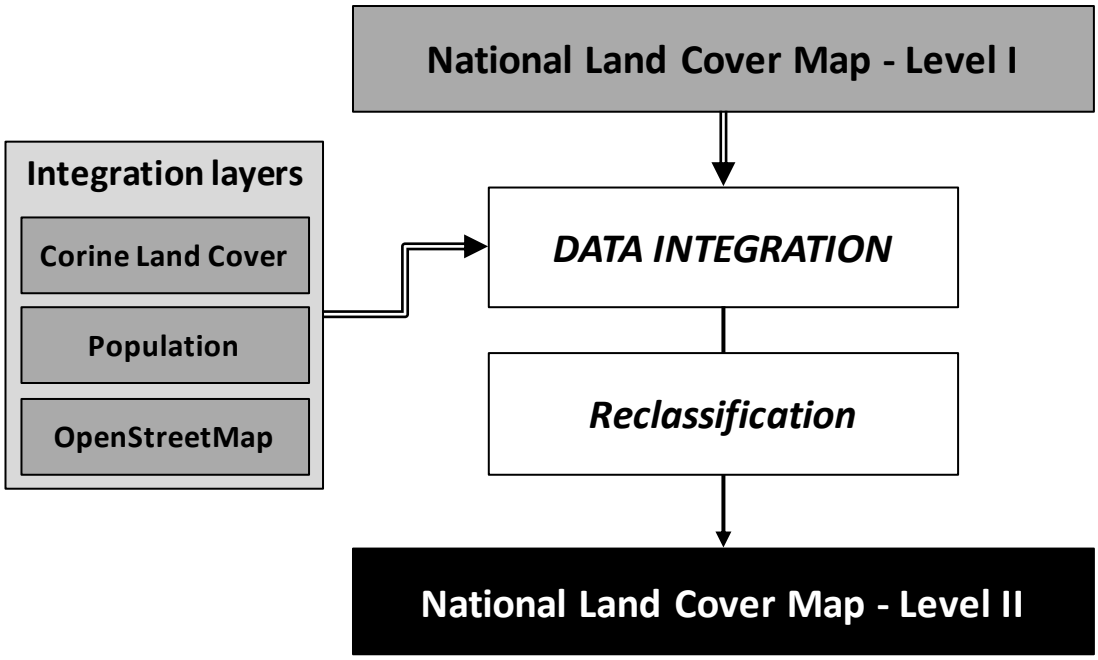


Figure 39. Flowchart of the integration tool to generate National Land Cover map of Level 2 from High Resolution Layers (HRL) of Copernicus Land Monitoring Services.

7.2 Copernicus Marine Environment Monitoring Service (CMEMS) products analysis for assessing marine food provision ecosystem service ecopotential [

7.2.1 Tool description

Physical, chemical and biological characteristics of seawaters are primary descriptors to assess the spatial and temporal dimensions of ecopotential productivity performances in terms of fish vitality, growth and stress. Seawater temperature can be considered the fundamental Essential Variable (EV) required for the characterization of the fish vitality and consequently the marine food provision potential. Fish growth is strongly dependent on the seawater temperature, and this EV is significantly influencing the rate of fish growth.

In addition, seawater temperature influences the variation of many other parameters (others marine EVs) and as consequence the entire life cycle of marine organisms, in terms of fish size, metabolism, sex determination, growth and population structure (Handeland et al., 2008; Person-Le Ruyet et al., 2004; Pavlidis et al., 2000; Sigholt & Finstad, 1990; Tandler et al., 1898; Jonsson et al., 1985). Parameters such as dissolved oxygen (DO), salinity level, chlorophyll-a (Chl-a) and nutrient compounds are necessary for an early stage evaluation of the fish food provision potential (Soto & Norambuena, 2004; Pitta et al., 2009).

The toolbox presented in this section and shown in Figure 40 identifies suitable fish growth conditions sites scenario and estimates the potential annual fish growth by integrating in-situ data, Earth Observation (EO) and modeling products (Valentini et al., 2016; Valentini et al., in press).
EO products allow the spatial and temporal representation of EVs required for the characterization of the fish vitality and for the assessment of the distribution of many water quality parameters (Gašparović, 2012; Silió-Calzada et al., 2008; Filipponi et al., 2015a; Filipponi et al., 2015b), including Sea Surface Temperature (SST) (Merchant et al., 2008) and Chlorophyll-a.

Co-funded by the European Union

The toolbox requires the following information in order to integrate the different datasets and generate the suitability scenario:
- sea basin of interest;
- time period of interest;
- EVs of interest;
- fish species for which vitality should be assessed;
- vitality ranges for the selected fish species related to each considered EV.

Vitality ranges of the selected fish species are associated with the EVs of interest in order to define the threshold values for each EV related to fish vitality.

With respect to the datasets, in-situ data can be collected from various sources, for example the EMODnet chemistry portal (http://www.emodnet-chemistry.eu) allows to search for in-situ measurement data in the European Seas. RS data are pre-processed in order to generate gridded EO products describing the selected EVs. A wide range of products describing biological, chemical and physical characteristics of marine seawater can be collected from Copernicus Marine Environment Monitoring Service (CMEMS; http://marine.copernicus.eu).

A first preprocessing analysis is done on the input dataset in order to clip the data on the extent of sea basin of interest and generate depth-averaged variables, in the case of CMEMS multiple depth layer dataset generated using numerical modeling.

The workflow of the assessment is divided in two main steps.
The first step of the toolbox analyses biologically, chemically and physically relevant seawater parameters to verify that values are within the vitality ranges of fish species. The analysis of seawater parameters integrates in-situ measured data, which accurately describe the temporal trend of the EVs, and spatio-temporal represented parameters, in order to identify the existence of exceeding values with respect to the defined species specific thresholds. When exceeding values are found from the in-situ measured data, either the existence of site specific condition or the presence of persistent seawater conditions not suitable for fish growth should be evaluated. The assessment based on spatially represented data evaluates the acceptable locations (pixels) not exceeding threshold values for the selected EVs, hence suitable for fish growth. The other locations are masked from further investigation.

The second step estimates the potential fish growth rate for the selected fish species for the suitable sites identified in the first step. The generated suitability scenario, pictured by a spatial representation of available suitable location for fish growth in the selected sea basin, is indeed further in order to estimate the annual fish growth for the considered time period. A fish growth equation (reported in Figure 40) is set up with species specific fish growth rate coefficients and gives ad result the annual growth in grams. Temporal SST profiles for each suitable location (resulting suitability scenario) is used to estimate the average annual fish growth during the considered time period. As result a spatial distribution of estimated annual fish growth is generated, for the assessment of marine food provision ecosystem service.

Co-funded by the European Union

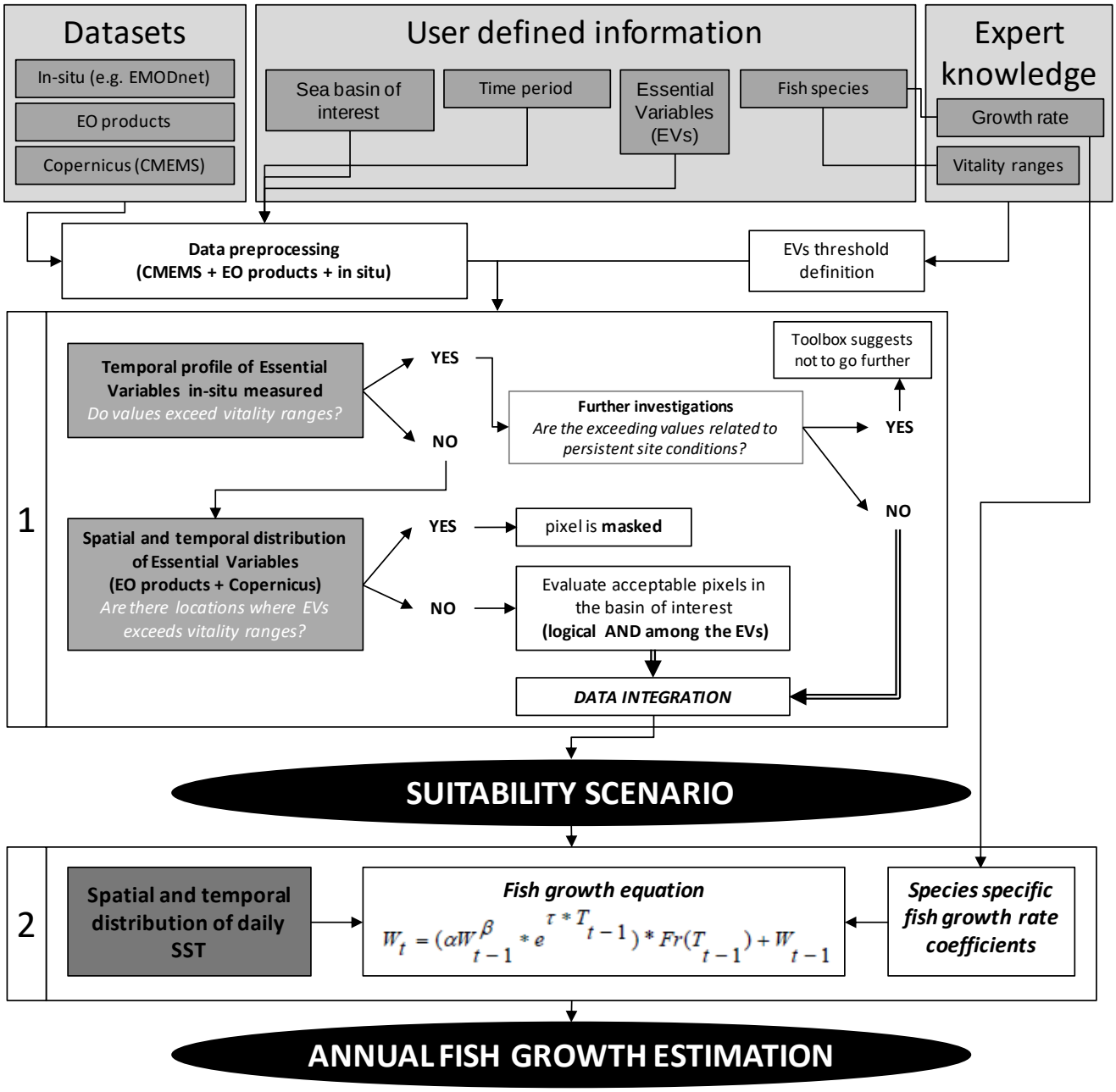


Figure 40. Data integration workflow of in-situ, EO products and CMEMS products for the assessment of marine food provision ecosystem service.

7.2.2 Application example

The developed toolbox to identify suitable fish growth conditions sites scenario and potential annual fish growth has been tested in the Mediterranean Sea (Valentini et al., in press) for the period 2008-2014 and the northern Adriatic Sea (Valentini et al., 2016) for the period 2002-2012. Among the identified suitable sites, the potential fish growth rate for two fish species, Sea bass and Sea bream, has been estimated.

The assessment of fish growth conditions was done using the Copernicus Marine Environment Monitoring Service (CMEMS) products, such as Sea Surface Temperature (SST), MERIS L2 FR products, like Total Suspended Matter (TSM) concentration, together with other modeled variables like currents, dissolved oxygen (DO), nitrates, phospates, etc.

Regarding the CMEMS products, for the assessment of the SST the satellite-derived gap-filled L4 SST maps at 1 km horizontal resolution, acquired by the infrared sensors of different satellite platforms (Nardelli et al., 2013) was selected. Analysis of Chlorophyll-a variable was done for gap-filled (Beckers & Rixen, 2003) monthly data estimated using MedOC4 algorithm (Volpe et al., 2007) from remote sensing reflectance (Rrs) spectra reprocessed within the ESA Climate Change Iniziative (CCI).

In-situ data, acquired from European Marine Observation and Data Network (EMODnet) database, were used to analyse the existence of suitable conditions for fish growth at high temporal resolution.

Seawater conditions enable the survival and the growth of the two selected fish species in almost all the Mediterranean Sea basin. Locations identified as not suitable for fish growth conditions, and masked from further analysis, are those reporting seawater temperature exceeding minimum threshold value of 5°C and maximum threshold value of 30°C and reporting a high percentage of annual days of stress temperatures. Other areas masked from further analysis (showed in black color in Figure 41 and Figure 42)are those with high salinity levels exceeding average value of 40 PSU, average current speeds lower than 0.02 m s-1 or higher than 0.51 m s-1, low dissolved oxygen average concentrations, and high nutrients and turbidity levels, mainly due to the river discharges and reduced circulation.

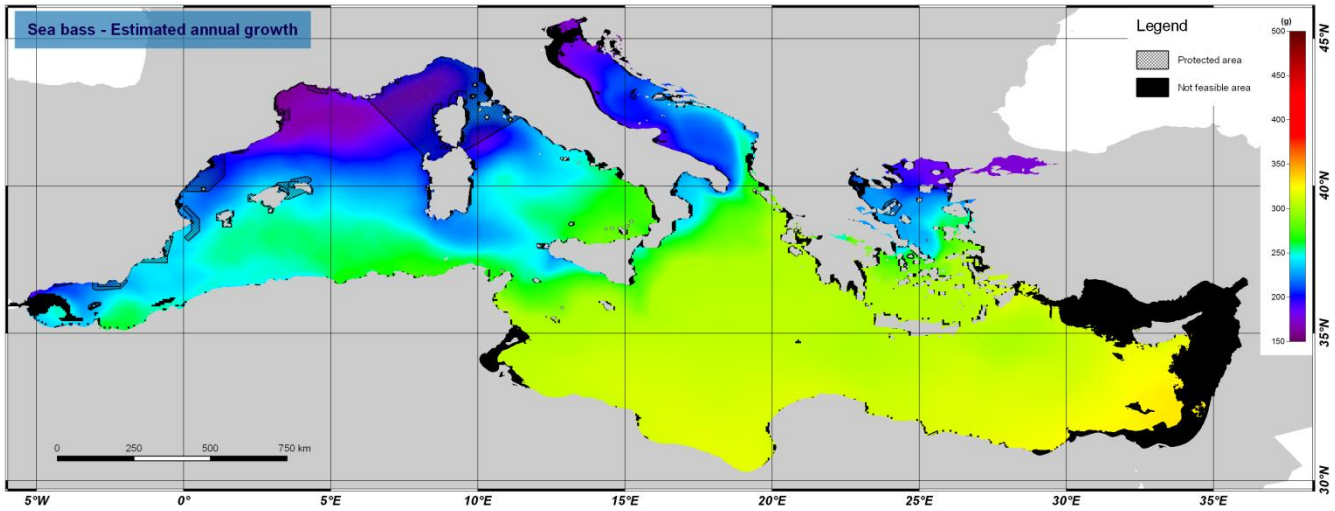


Figure 41. Estimated annual fish growth for sea bass within the suitability scenario.

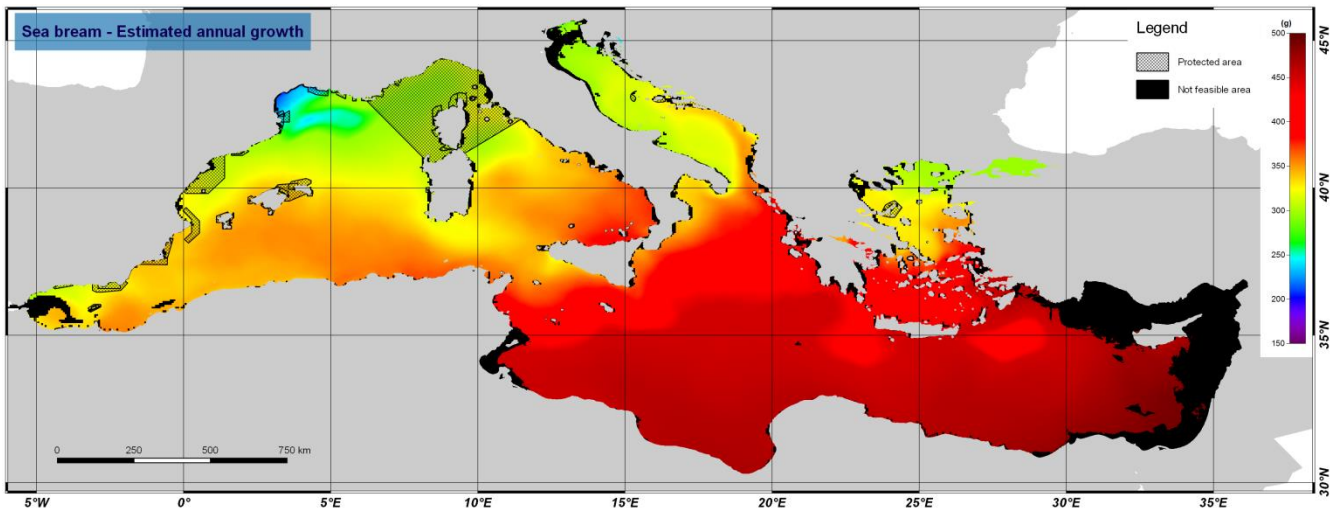


Figure 42. Estimated annual fish growth for sea bream within the suitability scenario.

As a result of the first step analysis, the toolbox output is a suitability scenario delimiting the area of the study basin in which, according to the considered parameters related to fish vitality, fish growth is suitable.

The sea bass and sea bream estimated annual growth shows the typical latitudinal gradient mostly due to the temperature factor in the growth function. Numerical results for the estimated annual fish growth show how there are essentially the same gradients in the spatial distribution (Figure 41 and Figure 42). Ecopotential productivity as annual growth in grams of sea bream, ranging among 250 and 500 g y^{-1}, is on average higher than the that of sea bass, that ranges among 160 and 360 g y^{-1}. The dual biophysical properties behavior detected in the analysis makes the two Mediterranean sub-basins differently suitable for food production, especially in terms of fish annual growth.

7.2.3 References

[1] Beckers, J. M., Rixen, M. (2003). EOF calculations and data filling from incomplete oceanographic datasets, J. Atmos. Oceanic Tech., 20, 12, 1839-1856.

[2] Filipponi F., Taramelli A., Zucca F., Valentini E., El Serafy G.Y., 2015a. Ten years sediment dynamics in northern Adriatic Sea investigated through optical Remote Sensing observations. International Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, pp. 2265-2268. DOI: 10.1109/IGARSS.2015.7326258. ISBN: 978-1-4799-7929-5/15/$31.00.

[3] Filipponi F., Zucca F., Taramelli A., Valentini E., 2015b. Total Suspended Matter (TSM) and maximum signal depth (Z90_max) for monitoring the evolution of sediment resuspension processes in shallow coastal environments. Proceedings of "Sentinel-3 for Science Workshop", Venice (Italy) 2-5 June 2015, ESA SP-734 (CD-ROM). ISBN: 978-92-9221-298-8.

[4] Gašparović, B. Decreased production of surface-active organic substances as a consequence of the oligotrophication in the northern Adriatic Sea. Estuar. Coast. Shelf S. 2012, 115, 33–39.

[5] Handeland, S.O.; Imsland, A.K.; Stefansson, S.O. The effect of temperature and fish size on growth, feed intake, food conversion efficiency and stomach evacuation rate of Atlantic salmon post-smolts. Aquaculture 2008, 283, 36–42.

[6] Jonsson, B.; Ruud-Hansen, J. Water temperature as the primary influence on timing of seaward migrations of Atlantic salmon (Salmo salar) smolts. Can. J. Fish. Aquat. Sci. 1985, 42, 593–595.

[7] Merchant, C.J.; Le Borgne, P.; Marsouin, A.; Roquet, H. Optimal estimation of sea surface temperature from split-window observations. Remote Sens. Environ. 2008, 112, 2469–2484.

[8] Nardelli, B. B., Tronconi, C., Pisano, A., & Santoleri, R. (2013). High and Ultra-High resolution processing of satellite Sea Surface Temperature data over Southern European Seas in the framework of MyOcean project. Remote Sensing of Environment, 129, 1-16.

[9] Pavlidis, M.; Koumoundouros, G.; Sterioti, A.; Somarakis, S.; Divanach, P.; Kentouri, M. Evidence of temperature-dependent sex determination in the European sea bass (Dicentrarchus labrax L.). J. Exp. Zool. 2000, 287, 225–232.

[10] Person-Le Ruyet, J.; Mahe, K.; Le Bayon, N.; Le Delliou, H. Effects of temperature on growth and metabolism in a Mediterranean population of European sea bass. Dicentrarchus labrax. Aquaculture 2004, 237, 269–280.

[11] Pitta, P.; Tsapakis, M.; Apostolaki, E.T.; Tsagaraki, T.; Holmer, M.; Karakassis, I. Ghost nutrients from fish farms are transferred up the food web by phytoplankton grazers. Mar. Ecol Prog Ser. 2009, 374, 1–6.

[12] Sigholt, T.; Finstad, B. Effect of low temperature on seawater tolerance in Atlantic salmon (Salmo salar) smolts. Aquaculture 1990, 84, 167–172.

[13]Silió-Calzada, A.; Bricaud, A.; Gentili, B. Estimates of sea surface nitrate concentrations from sea surface temperature and chlorophyll concentration in upwelling areas: A case study for the Benguela system. Remote Sens. Environ. 2008, 112, 3173–3180.

[14]Soto, D.; Norambuena, F. Evaluation of salmon farming effects on marine systems in the inner seas of southern Chile: A large-scale mensurative experiment. J. Appl. Ichtyol. 2004, 20, 493–501.

[15]Tandler, A.; Har'el, M.; Wilks, M.; Levinson, A.; Brickell, L.; Christie, S.; Avital, E.; Barr, Y. Effect of environmental temperature on survival, growth and population structure in the mass rearing of the gilthead seabream. Sparus aurata. Aquaculture 1989, 78, 277–284.

[16]Valentini E., Filipponi F., Nguyen Xuan A., Passarelli F.M., Taramelli A., 2016. Earth Observation for maritime spatial planning: measuring, observing and modeling marine environment to assess potential aquaculture sites. Sustainability 8(6), 519, doi:10.3390/su8060519.

[17]Valentini E., Filipponi F., Nguyen Xuan A., Taramelli A., 2016. Marine food provision ecosystem services assessment using EO products. Proceedings of "ESA Living Planet Symposium 2016", Prague (Czech Republic) 9-13 May 2016, ESA SP-740 (CD-ROM), in press.

[18] Volpe, G., Santoleri, R., Vellucci, V., Ribera d Acalà, M., Marullo, S., and D Ortenzio, F. (2007). The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates. Remote Sens. Environ., 107, 625-638.