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: 4.6

EO data quality elicitation and documentation

Due date of deliverable:	31st May 2018
Actual submission date:	30th May 2018
Version:	v1
Main Authors:	CREAF, UAB

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

Project ref. number	641762
Project title	ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS

Deliverable title	*EO data quality elicitation and documentation*
Deliverable number	D4.6
Deliverable version	V1
Contractual date of delivery	31 May 2018
Actual date of delivery	30 May 2018
Document status	Draft document
Document version	1.0
Online access	ECOPOTENTIAL website: http://www.ECOPOTENTIAL-project.eu
Diffusion	Public
Nature of deliverable	Report
Workpackage	WP4
Partner responsible	CREAF
Author(s)	Joan Masó (CREAF), Cristina Domingo-Marimon (CREAF), Alaitz Zabala (UAB), Núria Julià (CREAF)
Editors	Joan Masó (CREAF), Cristina Domingo-Marimon (CREAF)
Approved by	
EC Project Officer	Gaëlle Le Bouler

Abstract	Deliverable D4.6 provides an overview of Earth Observation (EO) data quality, elicitation and documentation done in WP4. It also evaluates the community needs on metadata quality as well as the existent metadata quality editors. Gaps have been identified and solutions proposed. For data quality we have selected the QualityML, an extension combining ISO geospatial metadata standards (e.g. ISO 19157) and UncertML. QualityML defines hierarchically structured concepts to precisely define quantitative and qualitative quality measures and relate it with quality classes and quality indicators. Besides, the present document includes a description on how to encompass this type of information in a metadata management tool based on XML. Additionally, the deliverable explains how to improve the final user interactions through two options: by accessing and query this information and by providing feedback using the OGC Geospatial User Feedback model. To integrate all these needs, we have created a map browser that incorporates data visualization, data analytics,

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

	metadata, user data quality, user feedback and data download in a single interface.
Keywords	Quality, metadata, documentation

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

Table of Contents

Table of figures

Co-funded by the European Union

1. Executive summary

ECOPOTENTIAL WP4 is responsible for the provisioning of Earth Observation (EO) data for the Protected Areas (PA) as well as to support other activities executed in WP6, WP7, WP8 WP9 and WP10.

In this document we define quality as the degree of congruency between recorded data and actual conditions that the historical date it represents. The quality of data generated is an essential aspect that influences the validity of the analysis performed using EO data. Quality data is also essential for several types of end users, ranging from simple interpretation of the data to modelers, which will determine the ability to provide realistic models of the future impacts of environmental change. Data quality elicitation is a compulsory requirement for all partners producing data for ECOPOTENTIAL. This information may be associated at dataset level or at pixel level, and ECOPOTENTIAL covers both alternatives as it is explained in Section **Errore. L'origine riferimento non è stata trovata.**. Nevertheless, quality must be a balance between the indicators provided by the producer and the feedback quality provided by the user.

The present study evaluates the community needs on metadata quality as well as the existent metadata quality editors. A review analysis was performed to metadata editor tools and outputs from the surveys performed in WP5 concerning quality metadata were taken into account. This resulted in a clear identification of gaps and concerns of ECOPOTENTIAL EO data end users. For instance, INSPIRE metadata tool online editor allows documenting metadata elements considered mandatory in the technical specifications of the INSPIRE European Directive; however its main limitation is related to the quality and validity section, in which user can not include quality parameters but only lineage descriptions. In order to cover these gaps, WP4 presents a two-fold solution based on the experience acquired during the FP7 GeoViQua project. On the one hand, ECOPOTENTIAL has designed a quality metadata management tool based on XML that is built on current metadata standards ISO19115, ISO19139 and ISO19157 and QualityML, a quality data model and vocabulary that encodes data quality. This solution includes an integrated model that allows user to edit the quality of data directly from the metadata components to which the quality applies, and not as separate elements. On the other hand, all the quality data generated in the project is accessible through the ECOPOTENTIAL map browser in conjunction with a Geospatial User Feedback system. With this implementation all layers available through the map browser can be evaluated by the users an then they can report on their experience through the User feedback option achieving the optimal quality balanced between producers and users.

Abbreviation and acronyms

AGB	Above Ground Biomass
CHM	Canopy Height Model
EO	Earth Observation
GAM	General Additive Model
GeMM	MiraMon metadata manager
GUF	Geospatial User Feedback
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
OGC	Open Geospatial Consortium
PA	Protected Area
VLAB	Virtual Laboratory

2. Introduction

Deliverable D4.6 provides an overview of Earth Observation (EO) data quality, elicitation and documentation done in WP4. Each EO dataset that has been generated during the project was produced following a rigorous quality control procedure based on the expertise of the producers. Nevertheless, rarely these procedures are communicating data quality in a useful way for the end users. In ECOPOTENTIAL, remote sensing data is an input (together with in-situ data) for a wider deployment of ecosystem services monitoring for Protected Areas (PA). Modellers, considered as an intermediate user, need this information in order to be able to propagate the quality in their models and therefore, estimate the uncertainty of such models. This is one of the reasons why it is important to have a standardized procedure to communicate about data quality and uncertainty.

This deliverable is not a guide on how to apply or control quality assessment to generate data quality indicators but on how such quality information can be formalized in an accessible way for end users in the metadata that describes the data. The deliverable provides a clear overview on the requirements concerning quality metadata especially for EO data. This deliverable focuses mainly on attribute quality (the degree of quality of the values provided in remote sensing products) and assumes that positional accuracy of derived products is inherited from the original source that is already known. Metadata is complex to generate and sometimes this results on poor metadata. This is particularly true with the metadata elements concerning quality. In ECOPOTENTIAL we agreed that these metadata must be conformant with international directives such as INSPIRE and this document explains in detail all the analysis done and solutions that have been implemented in order to improve data quality elicitation and documentation. It is worth notice that ISO quality model also considers linage as part of data quality but lineage is considered out of the scope of this document.

3. Quality elicitation

The ISO 9000:2005 defines Quality information as the degree to which a set of inherent characteristics fulfils requirements. Having complete metadata with such type of quality fields allows users to discern among products and select the best fit-for-purpose data. Nevertheless, quality measurements can be delivered as different levels of granularity, depending on statistical selected domain. One of the available options is to associate quality as a single metric for a product or for a whole dataset (dataset level). Alternatively a quality measure can be associated to each pixel of the image (pixel level). This section covers both alternatives.

3.1 Quality at product or dataset level

Traditionally, quality reports are associated to product specifications resulting in quality metrics common for the whole product. Besides, an individual scene of a given product may be evaluated against what is considered ground truth data, leading up to a new quality metric report. In the context of WP4, many products have been validated by the PA’s personnel, and where possible, uncertainty measures, in relation with the process of calculation of variables per se, have been computed. Hereafter, some of the metrics generated are listed. Samples of detailed reports of such metrics are included in Annex 1. Examples of Quality reports.

Quality for Doñana National Park Inundation maps

Inundation maps for Doñana National Park derived from Sentinel-2 were validated using as ground truth 7 Landsat based inundation maps coinciding with Sentinel-2. For each of the dates, accuracy metrics were estimated for water and non-water classes including producer’s accuracy, user’s accuracy, overall accuracy and kappa coefficient. An example on how this quality information is visualized and distributed is included in Section 5.2.1 (Figure 12). Here, we describe these examples.

Quality for Doñana National Park water bodies delineation and water turbidity for inland waters maps

Water bodies’ delineation product was cross validated obtaining and overall agreement and kappa index. Alternatively, the water turbidity for inland water product is distributed with a General Additive Model (GAM) explained variance for two areas, shallow marshlands and river estuary.

Quality for Samaria Land cover maps

Land cover maps obtained from classification of remote sensing data are frequently used as input layers to spatially explicit vegetation variables or environmental models, such as land cover fragmentation. It is therefore essential to provide information regarding the classification quality of these data. In fact, modelling quality of classification from satellite images is one of the most typical examples of modelling quality attributes of categorical data. In this example, overall accuracy for the land cover maps used in Samaria, for years 1985, 1995, 2000, 2005, 2010 and 2015 was generated together with the kappa coefficient.

Quality for Above Ground Biomass (Vegetation metric derived from Lidar)

Canopy height models (CHM) can be computed from Lidar data, and from these models derived products such as above ground biomass (AGB) can be generated. AGB was computed for several PA (Sierra Nevada, La Palma and Swiss National Park and Davos) and validated against in situ data when forest inventories where available (that is considered *the ground truth*). A quantitative attribute correctness metric, the root mean square error (in % and in t/ha) and the Pearson Coefficient correlation (*r*) were distributed as quality metrics in these datasets.

3.2 Pixel level quality

Remote sensing data is intrinsically informing about the spatial variability of a variable. However, the metrics generated at product or dataset level are validated against a small set of in-situ data that is considered representative for the dataset but is not dense enough to provide information regarding the spatial distribution and variability of the uncertainties. Although the most frequent granularity for distributing data quality is at product or dataset level, in some cases (e.g. algorithms that are capable of modelling the spatial distribution of an essential biodiversity variable) quality information at a more detailed level is required. In this case uncertainty can be represented as a map and can be propagated at high level models to derived products that are also reporting data quality in a more granular level. In fact, Earth Observation programs such as Landsat or MODIS are already including associated quality data (cloud masks, radiometry quality, etc.) at pixel level in some of the products they are distributing. However EO quality assessment at pixel level is a complex procedure and its use is not general yet. In ECOPOTENTIAL, users will find examples of quality data computed at pixel level.

Quality for Doñana National Park Hydroperiod maps

The annual hydroperiod maps are generated from a time series of inundation maps corresponding to the period of flooding, by accumulating the flooded information through the cycle. The quality metric of the hydroperiod are the mean of the inundation maps frequency, the standard deviation of the inundation maps frequency and the Gini index[1] that gives an assessment on the evenness of the frequency available inundation maps along the flooding cycle. The higher the Gini Index for a cycle the lower is the representativeness of the hydroperiod value per pixel for every flooding cycle.

Quality for Sierra Nevada National Park Vegetation Index maps

Many band indices derived from EO data are the result of algebraic combinations of bands. For instance, the NDVI (Normalized Difference Vegetation Index) is computed as $NDVI = \frac{IR-R}{IR+R} = \frac{M}{N}$ (add and subtract variables and then divided), and also the NDWI (Normalized Difference Water Index) is computed as $NDWI = \frac{IR-SWIR}{IR+SWIR} = \frac{M}{N}$. Indeed, each of the bands used to compute this type of indices present an associated uncertainty that depends on several factors, some of which have an special distribution; for example, the pixel orientation angle in relation to the sun and the nadir distance. All these factors can be modelled and the uncertainty at the pixel level, spatialized through error propagation formulas derived from the algebraic original equations (Figure 1). In fact, the error propagation

[1] *Díaz-Delgado, R.; Aragonés, D.; Afán, I.; Bustamante, J. Long-term monitoring of the flooding regime and 592 hydroperiod of Doñana marshes with Landsat time series (1974–2014). Remote Sensing, 2016, 8(9), 775.*

Co-funded by the European Union

methodology could be generalized to other EO quantitative products. In order to allow an adequate use of such relevant information by end users (i.e. modellers), this is documented in the metadata with the support of quality coverages (additional quality bands). Indeed, in Figure 1 we see the NDVI values represented as an image and a window that shows that for each pixel we can obtain the NDVI value (e.g 0.85) as well as the NDVI uncertainty (e.g ± 0.08) meaning that the actual value should have a probability of 68% of being between (0.85-0.08, 0.85+0.08) (assuming a normal distribution).

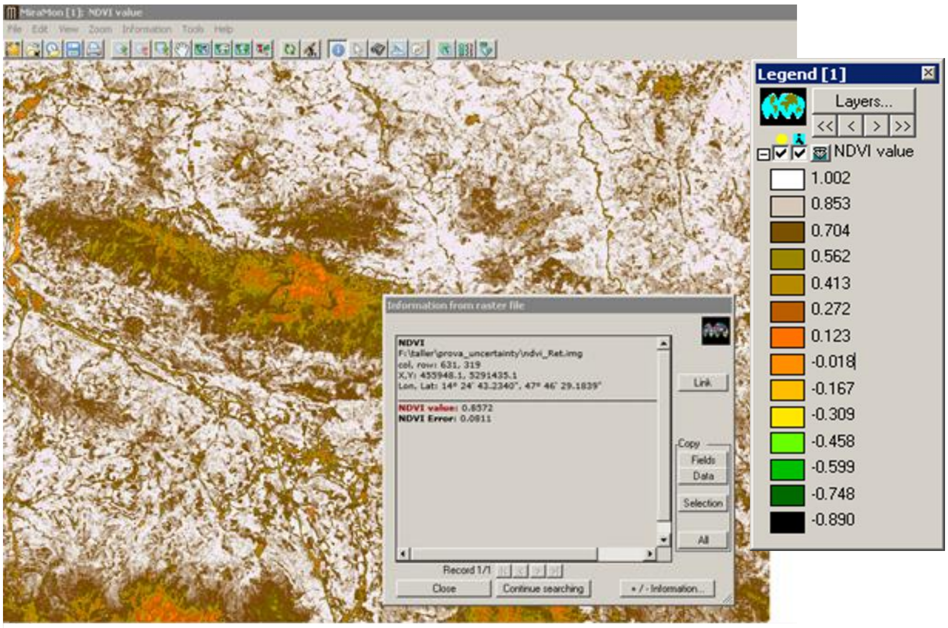


Figure 1. NDVI and uncertainty band associated. Information is documented in the Metadata.

4. Quality model

4.1 Data quality on Metadata and INSPIRE directive

There are some metadata standards that could be used to document metadata records generated by WP4 data producers. The INSPIRE profile of ISO 19115 has been the selected metadata framework to be followed, because it has a balance between having a normative background to follow in Europe and at the same time an reasonable amount of mandatory metadata elements to be filled in by the metadata producers. Another advantage is having a set of tools already available; some of them in open source. The Regulation as regards metadata (and subsequent amendments)[2] and Technical guidelines[3] set out the requirements for the creation and maintenance of this metadata. INSPIRE metadata implementing rules are the basis for WP4 metadata generation.

Concerning software, one of the existent tools that could be recommended is the INSPIRE online editor. It can be used to create a metadata template that can be modified using BATCH processing for generate metadata records for all the produced datasets. Unfortunately, this editor presents also some limitations on data quality aspects, as it will be explained in Section 5.1.

In fact, the description of the metadata model to use, the tools to work with it and also the extensions required to describe metadata regarding data quality, were the main topics of an internal session dedicated to the Metadata, that we named Metadatathon (the contraction of Metadata-Hackathon) that took place on November the 27th and

[2] https://inspire.ec.europa.eu/Legislation/Metadata/6541
[3] https://inspire.ec.europa.eu/Technical-Guidelines2/Metadata/6541

28th. The Metadatathon was jointly organized by UAB and CREAF and participated by many ECOPOTENTIAL partners and most of the WP4 participants. The main objective was to provide clear instruction on how to have a homogeneous set of metadata for each product and among partners. It started by a well balanced theoretical and practical lesson through a Webinar (27th Nov from 11:00 AM - 1:00 PM CET), devoted to the metadata understanding, production and use in the context of ECOPOTENTIAL project, following by INSPIRE metadata implementing rules as well as internal good practices. The first part targeted the whole ECOPOTENTIAL community and discussed about the usefulness and importance of different metadata sections for data producers and data users. How to read and exploit metadata records and where RS data and metadata are available was shown. The second part targeted mainly WP4 participants and was focused on tools for producing metadata and which are the main sections that ECOPOTENTIAL considers essential to populate, as well as guidelines on how to fill them in detail. It was based on a practical exercise using the INSPIRE metadata online editor for a real example. This webinar was complemented with an agenda of personalized support sessions where WP4 data producers attended to solve questions and work together to achieve the necessary quality level in their metadata records, as well as of course several emails and meetings to follow up the metadata creation by the data producers.

The materials distributed in this webinar are public available through:

http://twiki.grumets.uab.es/foswiki/pub/ECOPOTENTIAL_WP4/Metadatathon01/20171127_ECOP_Metadatathon_final_version.zip

and the video of the presentation can be found here:

https://www.gotostage.com/channel/a83bafe4fb324272a61f14f259a74a94/recording/7009120f0d324093bd3f703757add032/watch

The Metadatathon was very effective and the level o quantity and quality of the metadata improved considerably after the activity.

4.1.1 Recommended Metadata elements for remotely sensed observations

In general, metadata should include standard elements, but its format must be essentially adapted to end users, to increase its usefulness. After the review, ECOPOTENTIAL has identified and conformed some metadata elements to their own needs defining a new dimension fitted to existent standard elements which surely goes through existent gaps and helps in further metadata developments. In any case, metadata should also include a description of quality products to be distributed along with the data itself in a consistent and both human and machine readable ways. Guidelines on how content should be described were included in D5_2 Metadata for pre-existing datasets, but the elements related to quality will be detailed hereafter.

Dataset date

Definition: reference date for the cited dataset. Includes date and date type. This is especially relevant for temporal accuracy data quality.

Recommendation: at least dataset created date should be documented. Please note that this is not the date of the image/product (documented on the temporal extent) but the date while de dataset/product is created.

Origin: ISO 19115 md:MD_Metadata/gmd:identification Info/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date.

Example:	**date:**	2016-04-11
	dateType:	Creation

Temporal extent

Definition: The temporal extent defines the time period covered by the content of the dataset. This is especially relevant for temporal accuracy data quality.

Origin: INSPIRE 1205/2008 B.5 Temporal reference + ISO 19115 gmd:MD_Metadata/gmd:identificationInfo/ gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:temporalElement/gmd:EX_TemporalExtent/ gmd:extent/gml:TimeInstant/gml:timePosition.

Recommendation: date of the distributed image or that is the source of the distributed product is described as an individual date. If the product is derived as a result of several EO images, several individual dates may be described.

Example: *2003-07-16T10:02:00*

Spatial representation

Definition: digital mechanism used to represent spatial information. This information is especially relevant for the quality information of the spatial reference system.

Origin: ISO 19115 MD_SpatialRepresentation is an abstract and complex element including MD_VectorSpatialRepresentation and MD_GridSpatialRepresentation and, within the latter, MD_Georetified and MD_Georeferenceable: *i.e.* gmd:MD_Metadata/gmd:spatialRepresentationInfo/ gmd:MD_Georectified/gmd:axisDimensionProperties/gmd:MD_Dimension.

Recommendation: within WP4 products, MD_Georetified spatial representation applies. To fully describe this section, at least two axisDimensionProperties should be described, namely "column" and "row", and for each of them dimensionSize and resolution (including units) should be described.

Moreover, an id identifier is given to the full section, namely *spatialRepres*, in order to use it when describing the spatial representation of the CoverageResult for the DQ_QuantitativeAttributeAccuracy quality element (if any).

Example: ***gmd:MD_Georectified@id****:* spatialRepres
dimensionName *: row*
dimensionSize*: 2736*
resolution*:30 m*
dimensionName *: column*
dimensionSize*: 2951*
resolution*:30 m*

Corner points

Definition: earth location in the coordinate system defined by the Spatial Reference System and the grid coordinate of the cells at opposite ends of grid coverage along two diagonals in the grid spatial dimensions. This information is especially relevant for the quality information of the spatial reference system.

Origin: ISO 19115 gmd:MD_Metadata/gmd:spatialRepresentationInfo/gmd:MD_Georectified/ gmd:cornerPoints/gml:Point/gml:pos

Recommendation: within WP4 products, gmd:cornerPoints are used to describe the precise spatial extent in the spatial reference system of the dataset. Two corner points are described, with specific gml:id values: NW_corner and SE_corner.

Example: ***cornerPoint@gml:id****: NW_corner*
cornerPoint pos*: 411510 5329740*
cornerPoint@gml:id*: SE_corner*
cornerPoint pos*: 500010 5247690*

Quantitative Attribute Accuracy

Definition: closeness of the value of a quantitative attribute to a value accepted as or known to be true.

Origin: ISO 19157 gmd:MD_Metadata/gmd:dataQualityInfo/gmd:DQ_DataQuality/gmd:report/ gmd:DQ_QuantitativeAttributeAccuracy/gmd:result.

Recommendation: within WP4 products, whenever it's possible, a result describing the quantitative accuracy of the attribute should be provides as a quantitative result for the whole product or, even better, as a coverage result. If a coverage result is selected (*e.g.* having an "NDVI error" band), then spatial representation, content description and d format should be described as links to the previous metadata sections Semantically it should be related to standard descriptions of quality measures such as the ones on UncertML or QualityML (but this is still work in progress in task 4.6).

Example 1: NDVI RMS error: 0.2

Example 2: **QE_CoverageResult**
MD_SpatialRepresentationTypeCode: Grid
gmi:resultSpatialRepresentation: #spatialRepres"
gmi:resultContentDescription: #NDVI_error
gmi:resultFormat: #distribFormat"

4.1.2 QualityML

QualityML (http://www.qualityML.org) is a data model, a vocabulary and an encoding for data quality. QualityML vocabulary was developed as an extension of UncertML that includes all quality indicators described in ISO 19157 and others coming from other sectors in a unified model. It also proposes an encoding alternative for XML metadata documents (see http://www.QualityML.org/). QualityML was developed in FP7 GeoViQua project (2007-2013 agreement number 265178) and has been reviewed and extended as a result of OGC Testbed-12 Imagery Quality and Accuracy ER and for ECOPOTENTIAL task 4.6.

The main idea behind QualityML is that it can be used to describe quality elements in metadata in a standardized way that also include semantics (as it also points to the included definitions), allowing easy comparison of products. On one hand, the QualityML is a profile of the ISO geospatial metadata standards (e.g. ISO 19157) providing a set of rules for precisely documenting quality indicator parameters that is structured in 4 levels (indicators, measurements, domains and metrics). On the other hand, QualityML includes semantics and vocabularies for the quality concepts.

Whenever possible, QualityML uses statistic expressions from the UncertML dictionary encoding. However it also extends UncertML to provide list of alternative metrics that are commonly used to quantify quality beyond the uncertainty concept, for example the ones coming from ISO 19157. Quality metrics (most of them statistical operations) are used to compute the result of each quality measure value, when applied to a certain domain of uncertainty values. QualityML also provides a matrix of the combinations of indicators, measurements, domains and metrics commonly used.

The main idea behind this structure is to unlink measures, domains and metrics description, in order to **maximize generalization** of descriptions and increase coherence among several measures using the same metrics (even with different domains), or several quality indicators using the same measures. This was the original idea behind UncertML but UncertML lacks several common metrics necessary in ISO19157.

An effort has been done to unify different ISO Quality Basic Measures into a single QualityML Quality Metrics (with parameters). ISO 19157, in its Annex D, introduces the concept of data quality basic measure it has some repetitive definition of the same concept. Once these repetitions were identified in data quality measures that have certain commonalities, we were able to define the new metrics and separate them from the measurements. The <u>uncertainty-related data quality basic measures</u> are based on the concept of modeling the uncertainty of measurements with statistical methods. The measured quantity can be embedded in different dimensions. Depending on the dimension of the measured quantity, different types of data quality basic measures are used to construct data quality measures. The <u>counting-related data quality basic measures</u> are based on the concept of counting errors or correct items.

Uncertainty-related data quality basic measures:

Several uncertainty-related measures are described in ISO linked to a common Basic measure, for example there are three measures using LE50 or LE50(r) basic measure: linear error probable (Id. 33), time accuracy at 50% (Id. 55) and attribute value uncertainty at 50 % significance level (Id. 69)

QualityML goes one step further in this generalization effort and also group basic measures describing the same metrics with different parameters. For example, all the measures regarding "half length of the interval" are grouped in a single general metric called Half-lengthConfidenceInterval, that includes a parameter to describe the confidence level (or probability) of the true value being between the lower and the upper limit. Level has to be in the range [0,1]. This QualityML metric includes several ISO 19157 basic measures. This is done in QualityML not only for one dimensional random variables (Z, using "Half-length Confidence Interval" metrics), but also for two dimensional variables, including in a single definition several ISO metrics (confidence ellipse and uncertainty ellipse). In fact, the ISO description of confidence ellipse is general in the same way, as it has a parameter to describe the confidence (or significance) level.

The advantage of this generalization is not only the increase of coherence on the quality measures and metrics description, but also the possibility to describe any other confidence level interval in a standardized way.

Counting-related data quality basic measures:

A first grouping that QualityML defines beyond these ISO basic measures is related to the concept of counting items. All the measures related to the same quality measure are grouped and used a metrics call items which result can be expresses as a boolean number, a count or a rate.

In fact ISO 19157 slightly suggest several options, in this case for the rate elements, when recognizes that *"[Error rate / Correct items rate] can either be presented as* percentage *or as a* ratio. *The value unit in the quantitative result (see 7.5.4.2) can be used to specify that the result is presented in percentage or as a ratio."* To standardize these options for the rate as well as to combine the other two options (boolean and count), QualityML describe the Items metrics as a choice among "indicator" (for boolean), "count" or "rate". For the last one a parameter is described in order to include the maximum value of the rate. Thus, a value of 100 in this attribute will be used to express that the value is a percentage. Default value for this attribute is 1, representing a pure ratio.

Moreover, usually measures based on errors and on correct items are described in ISO 19157. Both definitions are exactly the same, only with the difference about "which elements" the measure is counting. This, in QualityML is described by the Domain of the Quality measure, allowing then a higher aggregation schema, setting the domain to Non-Conformance or Conformance.

4.2 Geospatial User Feedback (GUF)

Geospatial User Feedback is metadata that is mainly produced by the consumers of geospatial data products concerning the data they have bought or used, and for which they are experts on their advantages and gaps. The OGC standard allows for documenting feedback items such as ratings, comments, quality reports, citations, significant events, justifications of the ratings, etc. about the usage of the data. In fact, the user feedback is a boost for data producers and providers to improve their products.

The geospatial user feedback model, that intends to remain as simple as possible, describes the structure re-using some ISO quality and metadata elements of ISO 19115-1:2014. Then another OGC standard, describes how to implement the model in XML documents.

In order to implement this system a server side catalogue of feedback items is needed. The client application can integrate the feedback by communicating with the catalogue server. In the case of ECOPOTENTIAL, a map browser is used as *the client*. The user interface is designed to cover different levels of expertise on geospatial data usage, remaining as simple as possible but comprehensive enough. Details about the implementation can be found in Section 5.2.2 *Geospatial User feedback - NiMMbus* of the present document.

5. Quality documentation

5.1 Data quality on a Metadata management tool based on XML

In order to have a clear idea about the state of the art of metadata management tools, a review analysis of existent tools has been performed. One of the main interesting findings was related to data quality. Some of the tools partially or totally ignore the quality data, while some others do treat the quality data but, unfortunately in a non-friendly way, complicating the documentation and defaulting the understanding of quality metadata.

For instance, INSPIRE online metadata editor (available at http://inspire-geoportal.ec.europa.eu/editor/) allows a user to document the metadata elements defined as mandatory in the technical specifications of the INSPIRE European Directive. The online editor has a quality section and a validity section (Figure 2); however these sections do not allow users to document quality parameters but only lineage descriptions. The tool also documents the conformity with the INSPIRE metadata implementing rules.

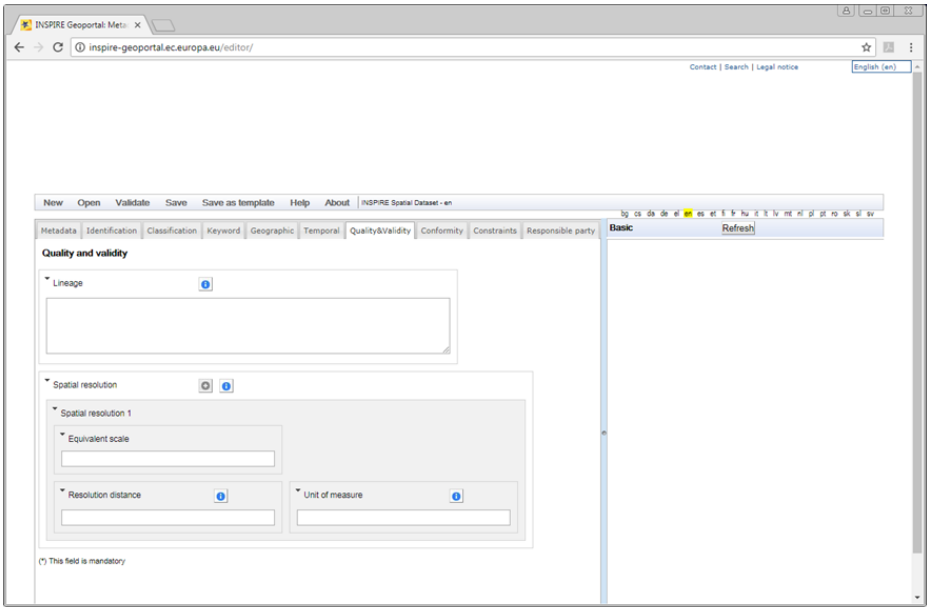


Figure 2. INSPIRE online metadata editor quality and validity section

Another metadata editor tool worth to mention is CatMDEdit available at http://catmdedit.sourceforge.net/. This tool allows a user to document all metadata elements defined in the ISO 19115 metadata standard. In this case, the non-intuitive interface makes difficult the use of the tool, transforming data documentation into a challenge both for the user and the producer (Figure 3).

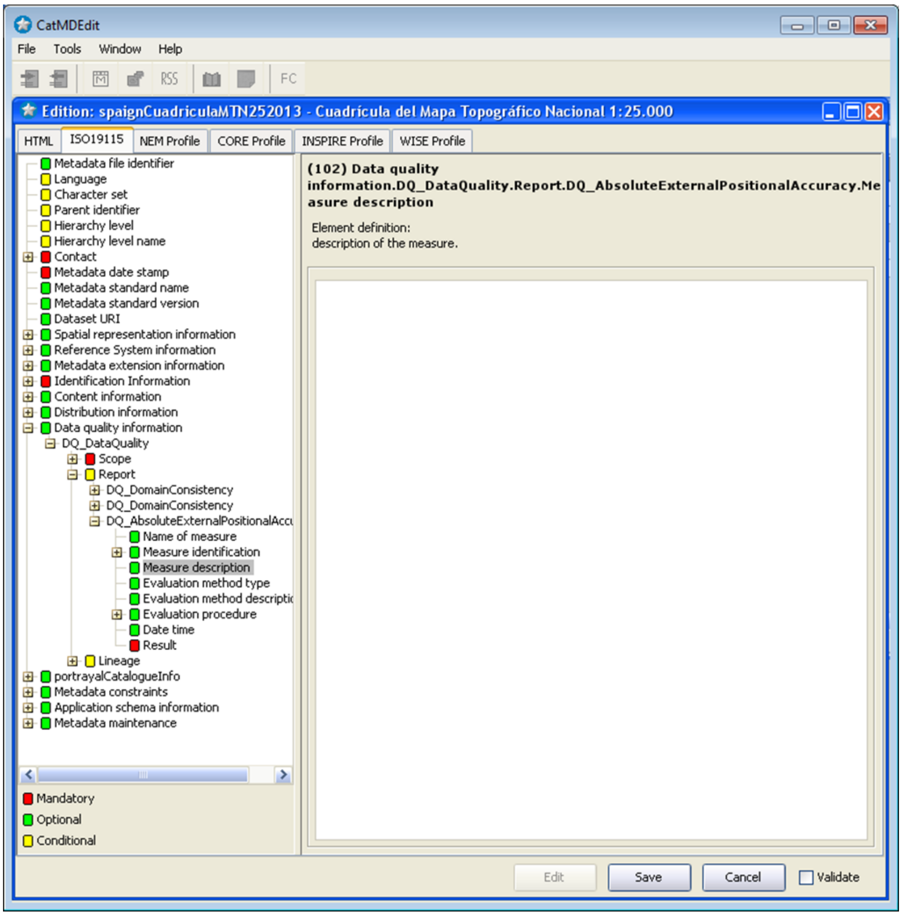


Figure 3. CatMDEdit tool allows quality metadata documentation. Unfortunately, the interface is not enough user-friendly.

All the tools reviewed so far suffer from the fact that either are incomplete or have non-intuitive interfaces. Also, most of metadata tools treat data quality in a separate window. In our experience, including quality in the context it applies improves the comprehension of this type of information.

Additionally, an online questionnaire concerning Metadata was performed (by WP5) to the ECOPOTENTIAL researchers. From the answers (54 responses) the most common concerns about quality metadata were extracted:

i. classify their position relating spatial data life cycle, mainly as data user or present equilibrium as double nature as data provider/supplier and data user/consumer;
ii. indicates their limited knowledge about (spatial) data quality/domain of spatial/geographic data concepts and related ISO standards as well as, INSPIRE, EML, DEIMS, Dublin and Darwin Core metadata standards/application schemas;
iii. At same time, the user show knowledge relating data quality elements according ISO19157:2013 and limited experience in implement quantitative/qualitative procedures of spatial data quality management along spatial data life cycle and consider or communicate in quality assurance (QA)/quality control (QC) in contrast of explicit high interests in know/use spatial data quality elements, in know/use spatial data quality assessment methods and tools as well as, in use/participate in spatial data quality management process;
iv. consider decisive data quality elements (ISO19157:2013) in order to discover and select input data for applying socio-ecological/environmental models and workflows, to explore the results of practical/ecological meaning of output models data and to communicate with end user/technical-political decision makers;

v. Indicates as important, critical and relevant spatial data elements/indicators to incorporate into a metadata profile which to support/facilitates data quality assessment and management in scientific collaborative network management.

In an attempt to fill the gaps mentioned above and taking into account the concerns of ECOPOTENTIAL users, a quality management tool has been designed and added to a pre-existing metadata tool called GeMM (MiraMon metadata manager). This new implementation is based on the current metadata standards ISO19115, ISO19139 and ISO19157 but also follows QualityML model described in section 4.1.2. This solution includes an integrated model that allows user to edit the quality of data directly from the metadata components to which the quality applies, and not as separate elements. The tool also groups the quality indicators depending on the metadata aspects. Organizing the quality information next to the metadata component improves user's quality indicators interpretation, because it simplifies the association between the quality indicator and its respective metadata component.

A generic quality indicator dialog box at dataset level has been implemented (Figure 4). This dialog box presents a set of quality elements related by the metadata component. The information of the quality elements (indicator, measure, domain and metric) definitions and their description, parameters, etc. is directly read from the QualityML website in JSON and HTML format. These information is displayed through the indicator side buttons ( ), similarly as it is displayed in the map browser (Figure 11).

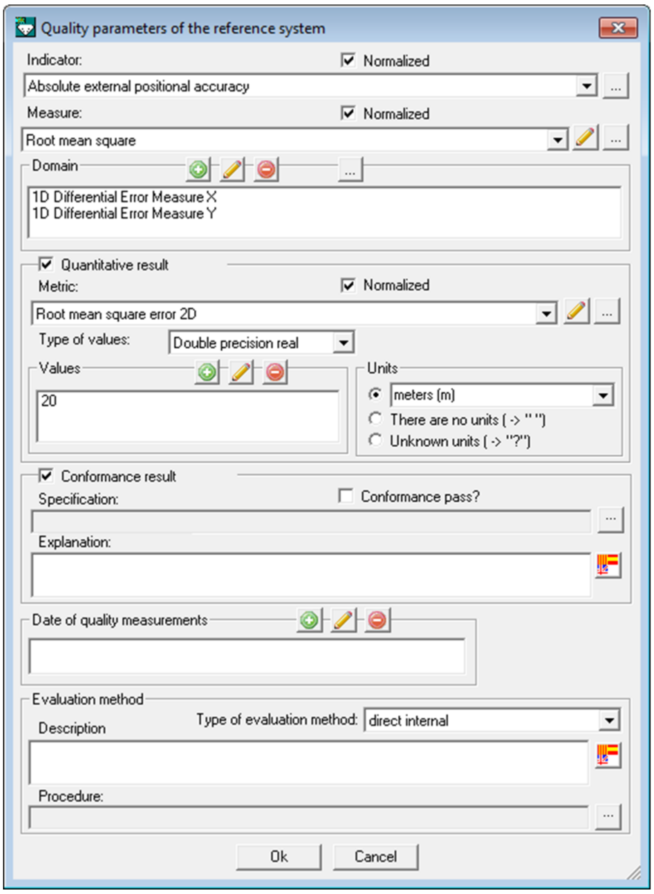


Figure 4. Example of the generic quality indicator dialog box at dataset level quality that has been developed.

Indeed, ISO19115 metadata standard defines 5 main types of quality classes: positional accuracy, thematic accuracy, temporal accuracy, completeness and logical consistency. Each of these classes is specialized in several quality indicators (Figure 5).

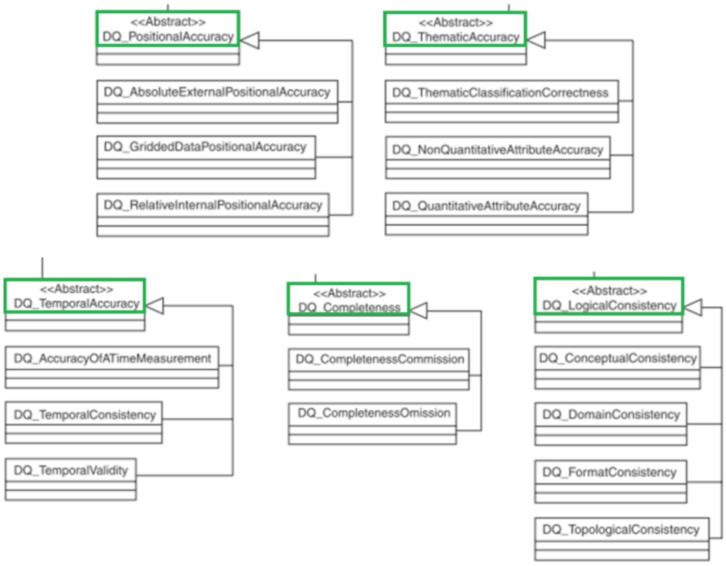


Figure 5. ISO19115 metadata standard main types of quality elements.

This implementation allows documenting all the quality metadata elements defined by the ISO19115 but the information of quality is not organized as a single group disconnected from the aspect it characterizes. Instead, the information has been grouped by the relation to metadata component it is relevant. Therefore, quality information of some indicators applies to all dataset, but in other cases quality information only applies to a field or to data subset. Examples will be illustrated below.

Quality indicators related to spatial reference system

The quality indicators related to spatial reference system mainly correspond to the Positional Accuracy type. This type includes the *absolute external positional accuracy*, the *gridded data positional accuracy* (for raster data) and the *relative internal positional accuracy* (Figure 6).

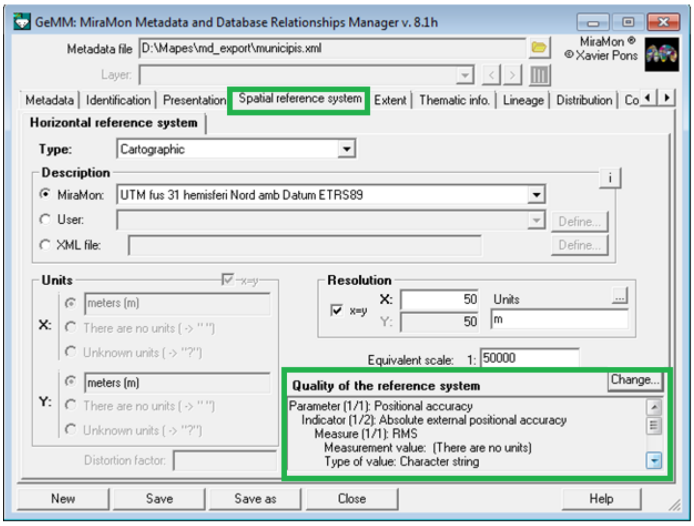


Figure 6. Positional Accuracy indicators included at GeMM

Quality indicators related to temporal extent

The Temporal Accuracy quality indicators are related to the quality of the time measurement. These include *temporal accuracy*, *temporal consistency* and *temporal validity*, and are presented in the temporal extent information section (Figure 7).

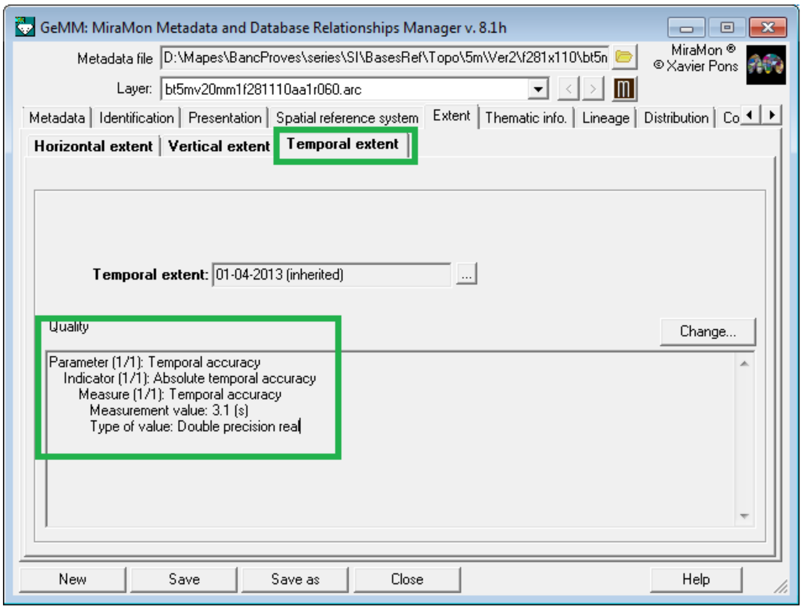


Figure 7. Temporal accuracy indicators included at GeMM

Quality indicators related to thematic data model

The quality indicators related to thematic attributes in the data model are defined by different types: the *thematic accuracy*, the *completeness* and *logical consistency* types. Several indicators are available, depending on the type of data. For categorical attributes, user could document indicators such *thematic classification correctness*, *non-quantitative attribute accuracy*, *completeness commission*, *completeness omission*, *domain consistency* and *format consistency*. On the other side, for quantitative attributes (Figure 8), the indicators implemented to be documented are *quantitative attribute accuracy*, *domain consistency* and *format consistency*.

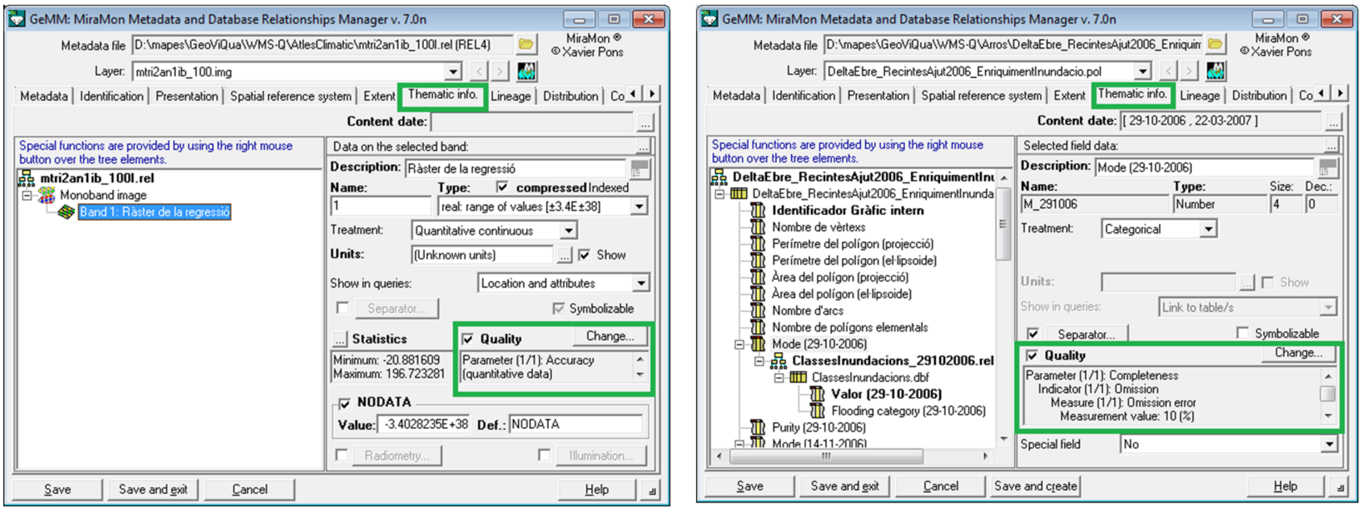


Figure 8. Thematic accuracy indicator for a quantitative dataset is included at the thematic information section in GeMM (Left). Completeness is also included in the thematic information section (Right).

Quality indicators related to content identification

The *conceptual consistency* indicator considers identification information content and therefore is included in the identification section (Figure 9).

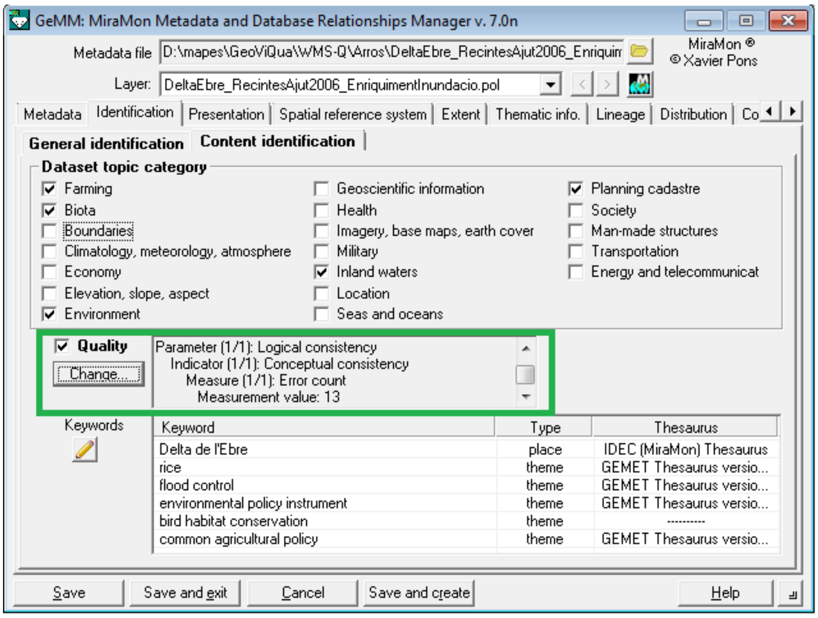


Figure 9. Conceptual consistency quality indicators included at GeMM are managed through the identification section.

Quality indicators related to data model

When data model is a vector file (geospatial features), the *topological consistency* indicators are included in the technical and model aspects section (Figure 10).

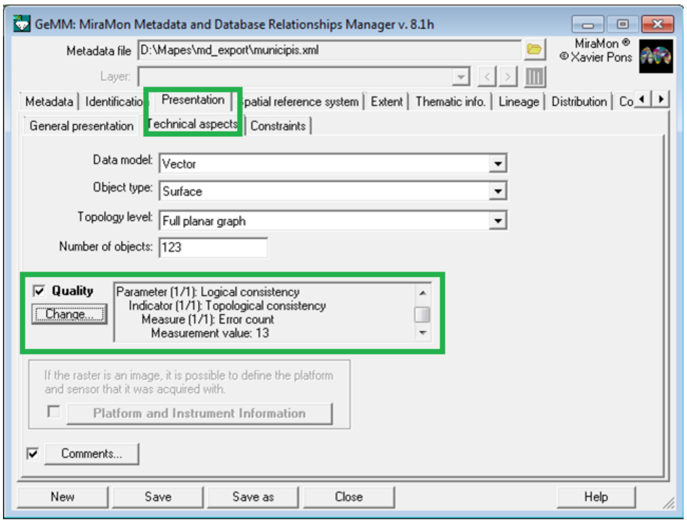


Figure 10. Topological consistency indicators are included in the technical and model aspects section.

5.2 Accessibility and visualization of metadata and quality data

ECOPOTENTIAL WP4 is the responsible for the provisioning of Earth Observation and in order to provide an easy way to visualize such data, a map browser app was created. This app features new functionalities specially tailored for ECOPOTENTIAL that provide more user interaction and exploitation of the data that goes beyond simple visualization by adds some basic analytical capabilities (details can be found in Deliverable D10.3). One of the functionalities that stand out is the possible to inquire about metadata. Metadata is essential for a correct interpretation of EO data and its derived products. In fact, metadata is a powerful tool that, once queried, allows users to locate specific data they need. Besides, quality metadata is a type of metadata that is more and more demanded because is one of the main elements giving the user criteria to decide on fit for purpose. On the one hand metadata describe the quality of the data per se. On the other hand, metadata are themselves quality components improving the data that is complementing. In any case, metadata and quality data must be accessible by any user.

5.2.1 Data Quality

Data quality indicators are available for those layers for which data quality has been quantified by the producer. This type of information has been carefully documented using QualityML (see section 4.1.2), The Map Bowser exposes the quality information in a dedicated window (http://maps.ECOPOTENTIAL-project.eu/).

Essentially, producer can report the dimension of the data quality measured, the indicator used, the type of measurement done, the uncertainties used to compute the numeric result (the domain) and the statistical expression (the metrics) used to summarize it. User can access this information through the map browser context menu: right click on the layer name in the legend and selecting the option "Quality" (Figure 11). In the dedicated window, each element is linked to the corresponding QualityML definition by clicking on the double blue arrow facilitating the user interpretation of each bit.

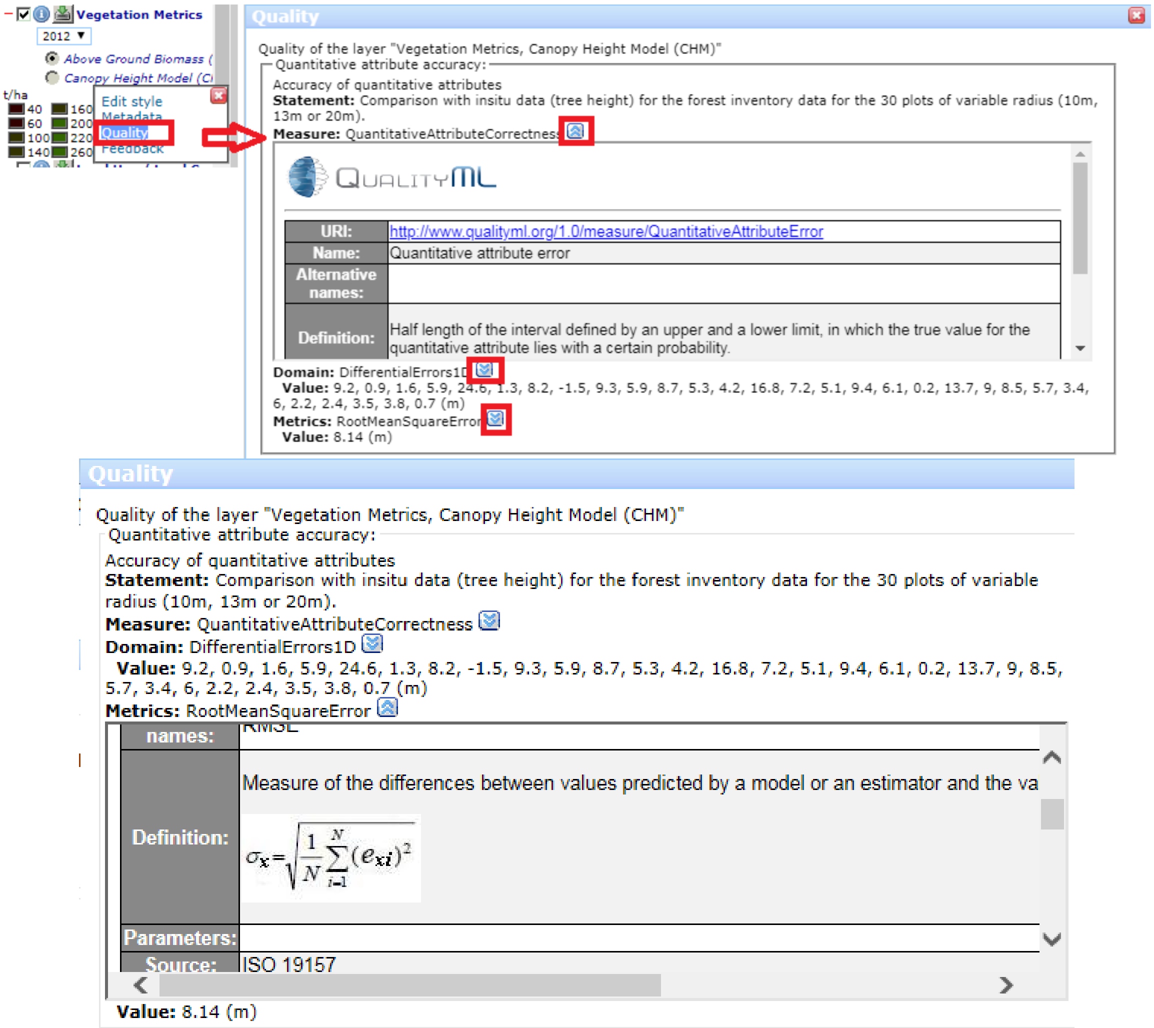


Figure 11. Data quality statements and indicators associated to QualityML

As an example, we show the Inundation maps for Doñana National Park generated by WP4 (CERTH) and previously introduced in section 3.1. The quality information has been incorporated into the map browser following QualityML and is presented to the user.

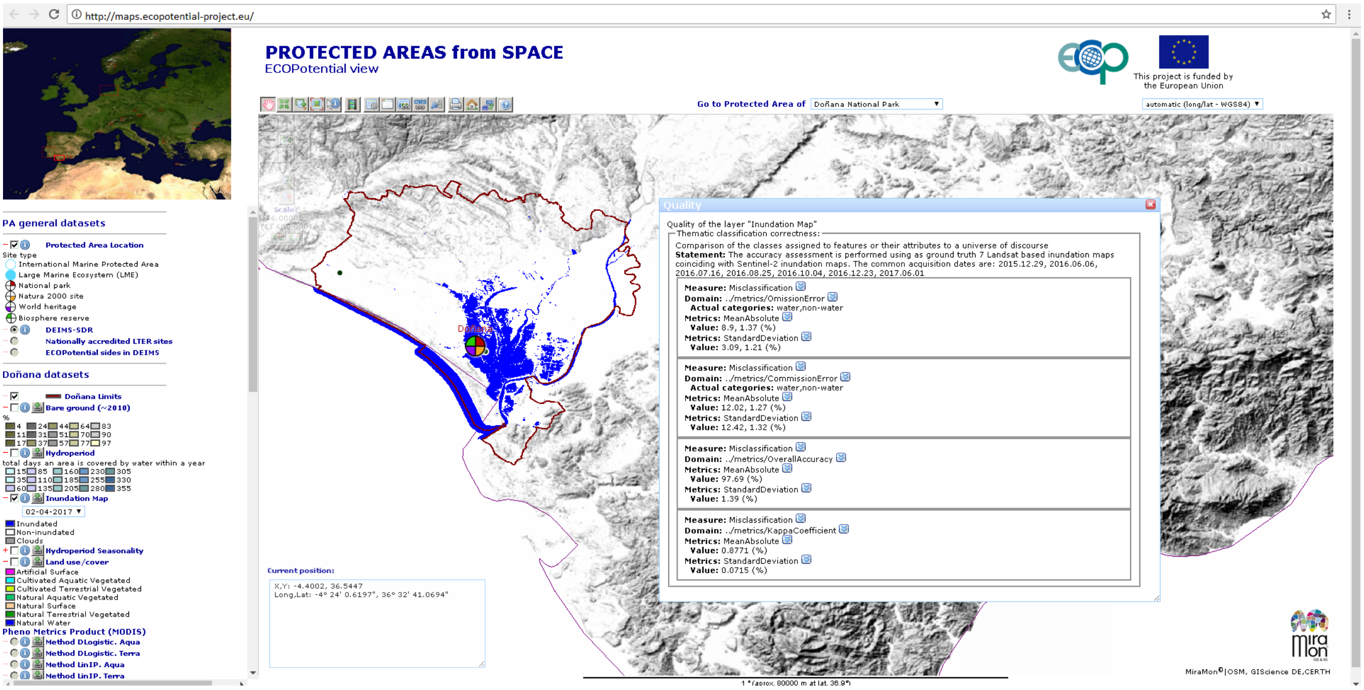


Figure 12. The layer corresponding to Inundation Maps for Doñana National Park was generated in the context of WP4. The associated quality can be consulted through the map browser. In this case the quality information provided correspond to omission errors, commission errors, overall accuracy and kappa coefficient

5.2.2 Geospatial User feedback - NiMMbus

As it has been explained in Section 4.2, it is more frequent to request feedback to users of geospatial data. Although producers do their best to communicate to the users who to interpret the data, sometimes mistakes or misinterpretations still happen. In other cases, users may find something interesting to comment in the data or may have a question. For these reasons it is essential to introduce a Geospatial User Feedback system (see section 4.2). With this implementation, all layers available through the map browser can be evaluated by the users through the *user feedback* option (context menu, right click on the layer name in the legend) (Figure 13 and Figure 14). The user feedback is implemented using the feedback catalogue NiMMbus that is developed in the context of H2020 NextGEOSS project. The system implements the Geospatial User Feedback standard developed in the OGC www.opengeospatial.org/standards/guf. The user feedback allows to provide comments, ratings and questions associated to a given geospatial dataset, in this case, the selected layer.

Users can provide feedback while exploring the data with the map browser and can be informed of other user's feedback items.

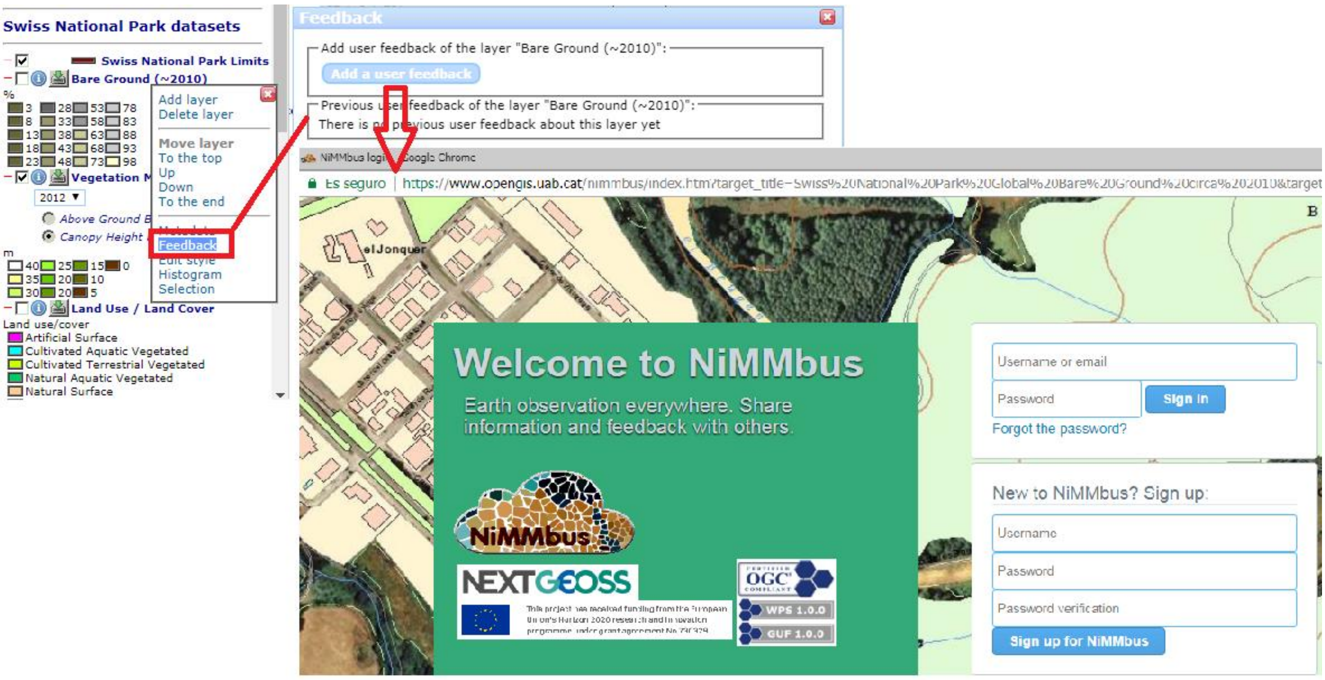


Figure 13. Adding new Geospatial User Feedback about a dataset

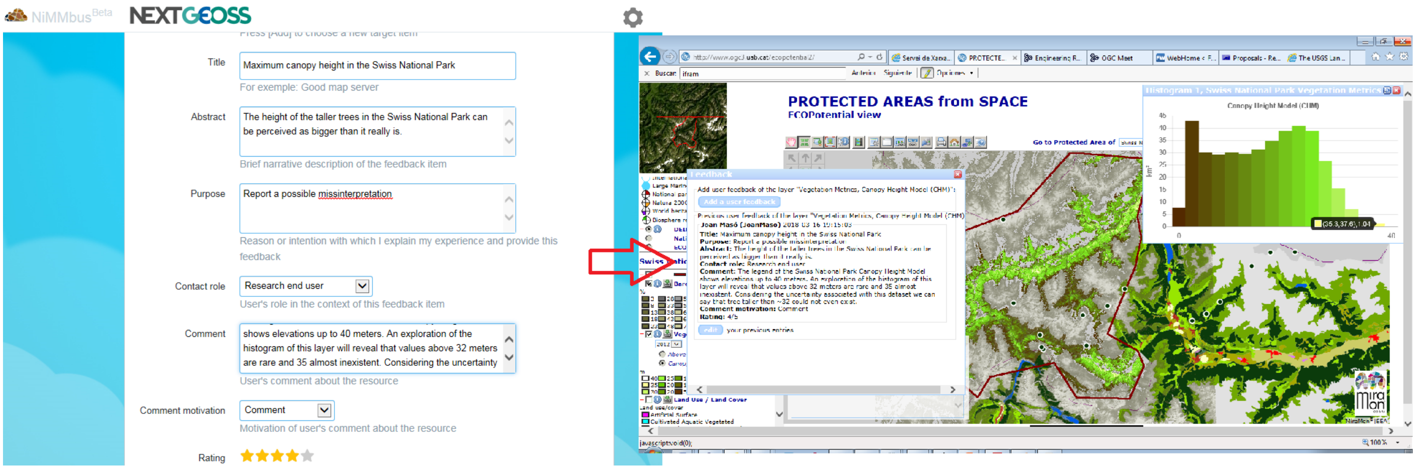


Figure 14. Showing Geospatial user Feedback about a dataset.

6. Conclusions

The use of geospatial data often deals with large amount of heterogeneous data derived from different sources. It is important to keep it organized in a well defined structure to simplify finding and reusing existing data. In addition metadata, data quality and user feedback inform the user and generate trust. Besides, quality data is a difficult notion to define precisely and therefore difficult to be generated without a clear set of concepts. QualityML provide a structured framework to classify and provide meaning to quantitative quality estimation. Data producers in ECOPotencial needed guidance and better tools on the production of metadata and data quality. An innovative Metadatathon session was organized and resulted in a more homogenous set of documentation and data quality reports. Those quality reports were transformed into QualityML summaries.

In general, metadata is exposed through a catalogue system but this keeps metadata separated from the data. In ECOPOTENTIAL we have opted by a more simple approach that makes the interaction with the data and metadata more natural: we have created a map browser that integrates data visualization, data analytics, metadata, user data quality, user feedback and data download in a single interface. This demonstrates that what was designed in the context of FP7 GeoViQua project is useful to be adopted at PA management level. At the end, QualityML and Geospatial User Feedback have become indispensable tools to deal with the complexity of metadata and quality generation.

7. Annex 1. Examples of Quality reports

A) Example: Quality report by CESBIO concerning Swiss National Park Vegetation metrics

EVs generation for Swiss national park and Davos wilderness areas

Prepared by Cesbio in the frame of the of ECOPOTENTIAL project

M. Tanase[1,2,3], Stéphane Mermoz[3], Alexandre Bouvet[3] and Thuy Le Toan[3]

[1] School of Ecosystem and Forest Sciences, Faculty of Science, The University of Melbourne, Victoria
[2] University of Alcala, Madrid, Spain
[3] Center for the Study of the Biosphere from Space (CESBIO)

Use of this document

1. This report is based on unpublished data and has not been peer-reviewed;
2. This report may be copied for distribution within the ECOPTENTIAL consortium
3. Distribution to a wider audience should be with permission of the authors.
4. Draft date: 22 November 2016

For further details regarding processing, modelling and cal/val activities please contact the CESBIO team (mihai@tma.ro, mermozs@cesbio.cnes.fr, Alexandre.Bouvet@cesbio.cnes.fr)

Data Use Policy

The data use is governed by the DATA USE POLICY. Data are furnished by individual scientists who decided to share products in the view of potential collaborations and joint activities. The data owners that decided to share these data rely on the ethics and integrity of the users to assure that the Data Use Policy is fully respected.

Data users are requested to inform the Data owners/authors about the planned activities and INVITE them to contribute to the work with additional intellectual inputs, analysis and discussion that would lead to a co-authorship. The invitation to contribute should be based on a first draft of results still open enough to be discussed and changed on the basis of the feedback from the Data owners/authors.

In all the cases (e.g. if the data owner is not interested to contribute) it is requested to properly acknowledge the data providers and the funding agencies. The data owner/authors have to be informed about publications that include these data, also to supply information about funding agencies. Downloaded data cannot be redistributed to others and must not be redistributed via other websites, databases or any other storage system to prevent circulation of different versions of the datasets.

Version control

Version 1	Uploaded on the ECOPTENTIAL repository (November, 23rd 2016)
Version 2	Uploaded on the ECOPTENTIAL repository (April, 4th 2017)

I. Input data (from ECOPOTENTIAL repository)

1. Remote sensing data

1.1 Ground-normalized digital surface model (DSM) from airborne laser scanning (ALS)
- Davos_nDSM.laz
- SNP_nDSM.laz

flight date: April 2003; provided as point cloud; clipped by forest extent; EPSG:21781.

1.2. Ground-normalized DSM (provided as point cloud) from high resolution stereo image matching (Leica ADS)
- Davos_ADS_nDSM.laz
- SNP_ADS_nDSM.laz

flight date: 2012; provided as point cloud; normalized with digital elevation model (DEM) derived from 2003 ALS flight; EPSG:21781.

1.3. Multiple returns cloud data from ALS for Dischma valley (part of Davos wilderness area, 9.87E, 46.78N)
- *wilderness_ultra_fine.laz (298 files)

flight date: 2015; provided as classified point cloud data (ground/non-ground); spatially coincident with in-situ measurements.

2. In-situ data
- forest inventory data (height and diameter at breast height) for the 31 plots of variable radius (10m, 13m or 20m)
- above ground biomass (AGB) for the 31

Plot 278 is located outside the ALS data acquired in 2015 (i.e. Dichma valley). Five additional plots located outside the forest perimeter (three on valley bottom and two on terrain slopes >15°) were added to account for zero biomass areas.

3. Ancillary data
- forest stand map for Davos wilderness area: stand_map_Davos2014.shp (cca. 2014)
- Landcover map for Swiss National Park (SNP): SNP_landcover.shp (source: Swisstopo VECTOR25).

II. Output data (delivered on the ECOPOTENTIAL repository)
- Davos_CHM_2003.tif
- Davos_AGB_2003.tif
- Davos_CHM_2012.tif
- Davos_AGB_2012.tif
- SNP_CHM_2003.tif
- SNP_AGB_2003.tif
- SNP_CHM_2012.tif
- SNP_AGB_2012.tif

*CHM-canopy height model; AGB- above ground biomass; SNP files masked using SNP_landcover.shp (*bush, *forest kept); Davos files masked using stand_map_Davos2014.shp; background value set to -9999; 10 m resolution; EPSGs:21781; CHMs at lower (2 and 5 m) and higher (20 m) spatial resolution are also available (contact CESBIO team for delivery).*

III. Methods

1. CHM generation for Davos and SNP areas:

CHMs were created using canopymodel.exe (part of USDA Fusion open software for ALS data processing) as exemplified below:

canopymodel /median:3 /ascii /class:0 /outlier:0,45 Davos_CHM_2003.dtm 10 m m 0 0 0 0 Davos_nDSM.laz

returns 'never classified' (class 0) were used; 3x3 median filter applied; heights above 45 m were considered outliers.

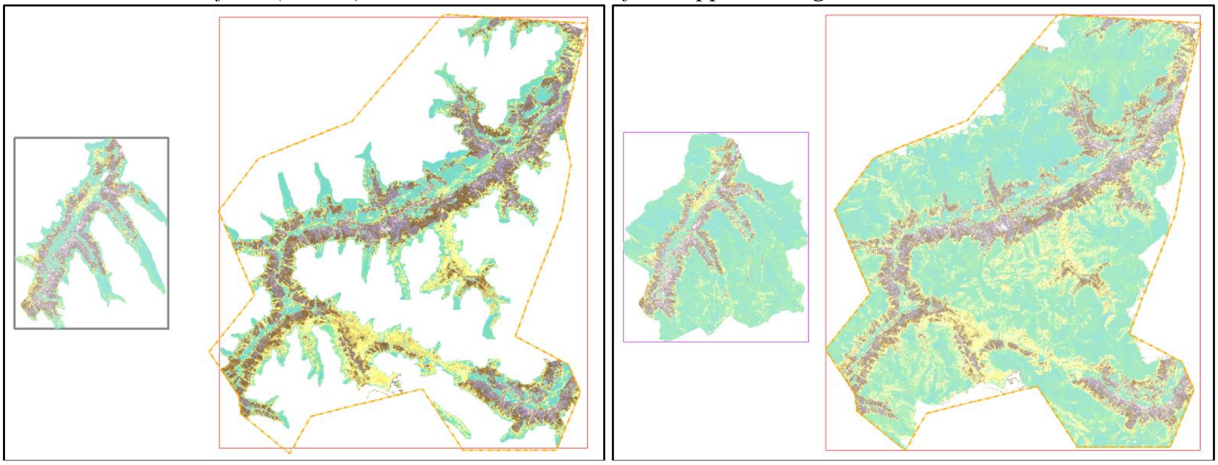


Fig. 1 CHMs derived for Davos and SNP using *_nDSM.laz data (left panel) and*ADS_nDSM.laz data (right panel). Data gaps in white. In each panel, Davos wilderness (9.83E, 46.78N) is at left and SNP at right (10.25E, 46.77N).

2. Above ground biomass retrieval

Two methods were trialed to retrieve AGB:

A) Through support machine regression (SVR) modeling of biomass as a function of CHM using as reference the in-situ AGB data (Fig. 2A).

B) Through intermediate AGB (irAGB) reference maps (Fig 2B). The irAGB maps were obtained using in situ data and full point cloud ALS data acquired in 2015 in Dischma valley (part of Davos wilderness area). Two intermediate reference AGB (irAGB) maps were generated using SVR modeling of biomass as a function of predictor variables describing the forest structure as derived from the 2015 ALS data:

1) Canopy height model and vegetation density for the 7 to 30 m strata (i.e., strata layers)

2) Average canopy height and percentage cover over 2m height (i.e., area processor (AP) derived layers)

The above layers were selected based on Akaike Information Criterion (AIC) values of linear regression models relating in situ biomass (log transformed) with up to three predictor variables (arcsine transformed) derived from the full point cloud dataset provided for Dischma valley. All possible combinations of the generated/selected predictor variables were analyzed.

The 2015 irAGB maps were subsequently used as reference to retrieve AGB for Davos and SNP areas through SVR modelling. As predictor variable, the canopy height model (2003 and 2012) was used. SVR models were trained using a set of 327 AGB/CHM value pairs extracted using a regularly spaced (100 m) grid.

Both methods were trialed on strata derived at 10 m and 20 m spatial resolution. The errors of irAGB maps derived from Strata or AP layers were cross-checked against the in-situ data to asses if significant difference exist between the two type of ALS strata generation. The error metrics (root mean squared error (RMSE), relative RMSE (RMSE%), and the correlation between predicted and observed values (r)) were used to help in deciding the optimal choice for the final AGB retrieval method.

Comparing the values in table 1 (Davos area) from methods A and B, it is evident that, at 20 m spatial resolution, the differences in AGB retrieval accuracy is small. Both methods show RMSE% values around 40%, RMSE varies between 96 and 102 t ha^{-1} while r value is the same (0.62). Apparently, the use of AP layers seemed to produce a more accurate irAGB map (check values for Dischma valley). However, this did not translate into a more accurate map during the second step in method B (Step 2: generalization). In fact, it seems the accuracy metrics for the final Davos AGB map are slightly worse when using the 'more accurate' AP-derived irAGB map. This might be related to the need to validate the AGB against the same in situ data used when for model calibration. The small in-situ data set available and the lack of independent data for validation hindered the error analysis step which in turn may have made the decision of modelling approach less objective. However, since the differences observed between the two methods are rather small the results might not different greatly when using either one of the two methods.

Comparing the accuracy metrics for AGB maps obtained at 10 and 20 m spatial resolution (Method A) one should notice a decrease in errors for the map obtained at the higher spatial resolution. Therefore, the final AGB maps were derived at 10 m resolution using method A.

Table 1 AGB validation metrics for the two different methods and intermediate steps.

Area (year)	AGB	ALS Proc.	Method (resolution)	Predictor	Validation against in situ data* RMSE (t/ha)	RMSE (%)	r
Davos (2012)	In situ	-	A (10 m)	CHM	87	35	0.68
Davos (2012)	In situ	-	A (20 m)	CHM	96	39	0.62
				Step 1: irAGB maps			
Dischma (2015)	In situ	Strata	B (20 m)	CHM Density 7_30m	96	39	0.65
		AP		Average canopy height 1st cover above 2m	79	31	0.75
				Step 2: generalization			
Davos (2012)	irAGB	Strata		CHM	98	40	0.62
		AP		CHM	102	41	0.62

* mapped AGB extracted and averaged over 314 to 1256 m^2 depending on field plot size

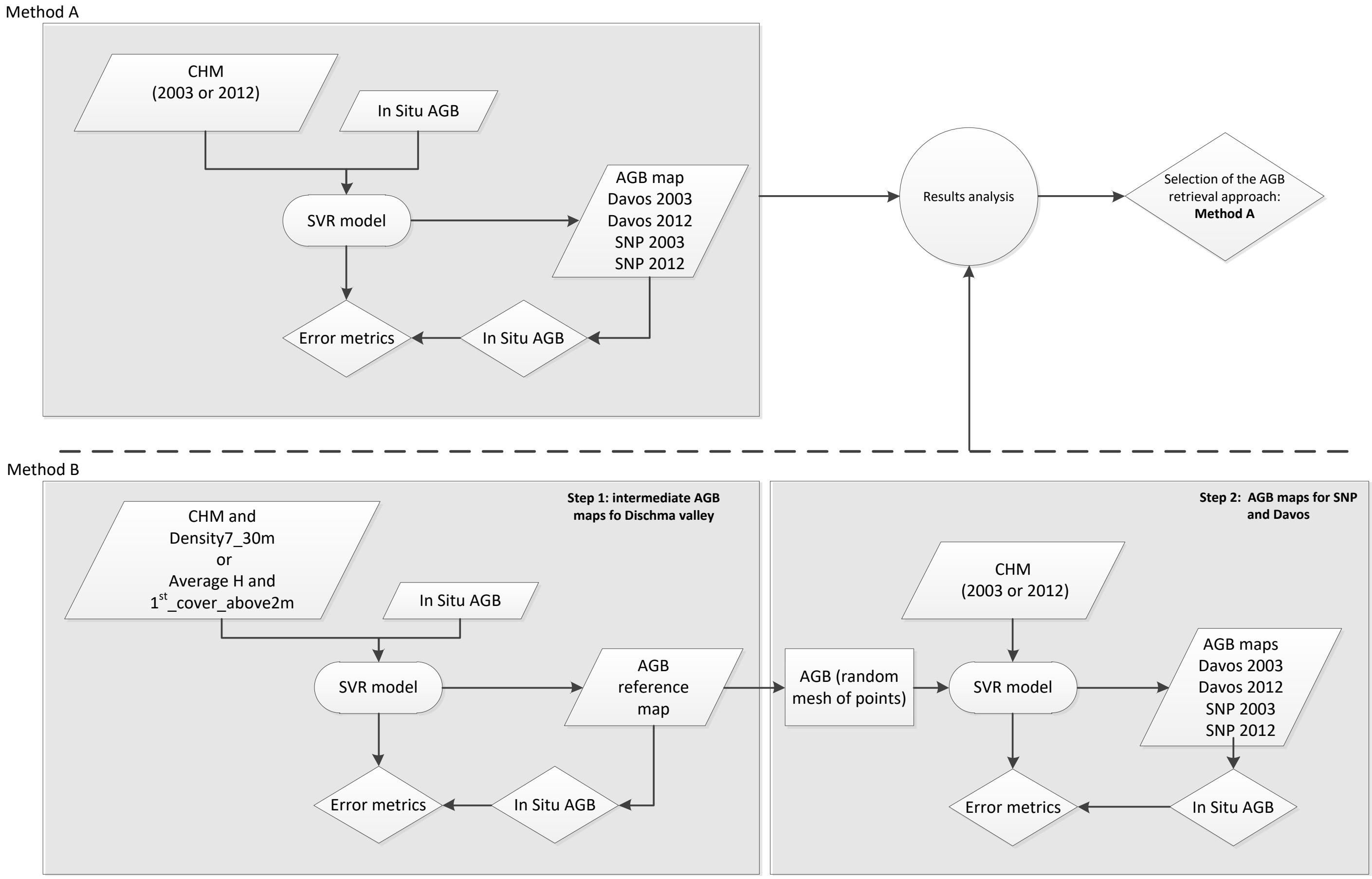


Fig.2 Flowchart for methods used to retrieve AGB in Davos and SNP areas

3. Post processing

The CHM and AGB maps were masked (using the ancillary data described at point I.5) to keep only areas classified as forests.

<u>IV) Preliminary analyses (results)</u>

1. Field data vs. CHM

Fig. 3 shows the consistency between CHM values (2015 vs. 2012 vs. 2003) extracted at the location of forest inventory (FI) plots. The agreement of the retrieved CHM heights is high (R^2 >0.9) which reassures temporal change analysis of forest cover/structure between 2003 and 2012. Thus, changes in forest structure between 2003 and 2012 might be derived by subtracting the two layers (e.g., CHM 2012 – CHM 2003). However, the agreement between field assessed biomass and the CHM is lower than expected (R^2_{linear} = 0.46). For a power type model the coefficient of determination increases to about 0.7 which should be explained by the functional relationship between biomass and height.

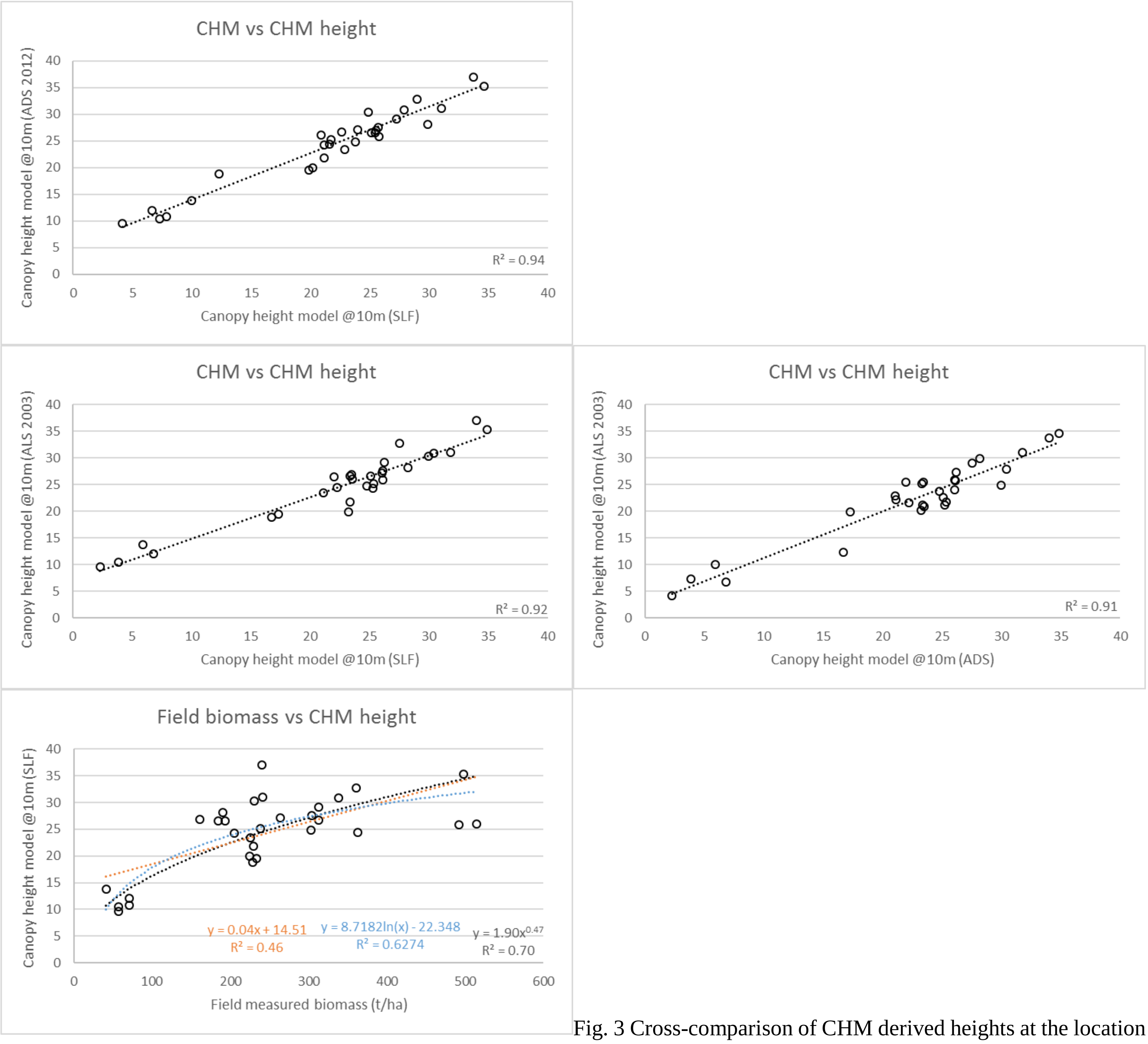
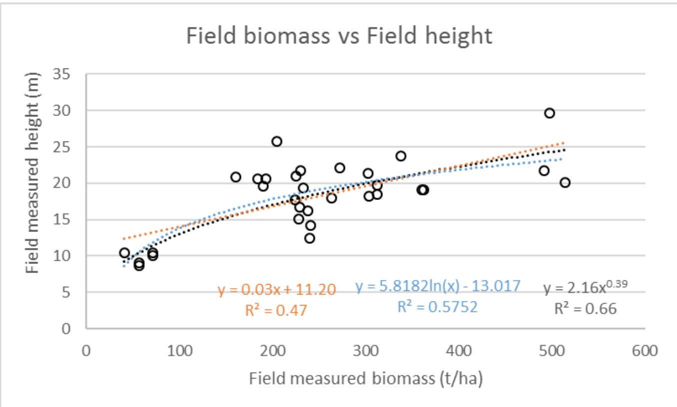


Fig. 3 Cross-comparison of CHM derived heights at the location of the field assessed plots in Dischma valley. Field derived height data is compared to ALS data acquired in 2015 (bottom right panel).

To briefly check the field data, the average field measured height was plotted against the biomass values derived from the forest inventory data in Fig. 4. The type of relationship observed in Fig. 3 (bottom right panel) is also observed for the field estimated height. However, one could notice several seemingly outlier values (plots 115,122,125, and 130, 147). Looking at the values for these plots one notices possible discrepancies between field average height/DBH and the corresponding biomass values (Table 2). In addition, two other plots (132,152) show significant discrepancies between field average height and CHM height. Therefore, the in-situ data was re-checked (by Ana Stritih, ETH Zurich) with the apparent discrepancies being attributed to heterogeneous forest structure (i.e., very dense young spruce forests on one hand, and rather open and heterogeneous stands at high elevation on the other hand). At plots 132 and 152 the forest is vertically structured, with many young trees in the lower canopy and some higher dominating trees, which might explain the discrepancy with the CHM. A second possible source of error may be related to small inaccuracies in the GPS location of the plots (up to ca. 5 m).

Table 2 Discrepancies between field measurements and the estimated biomass and between field and airborne measurements. In red between field H/DBH and computed biomass. In purple between CHM values and field measured height. In green expected H/DBH values for high biomass plots. Point 278 lays outside the 2015 ALS data, hence no value in the table.

Plot	SLF CHM	H (m)	DBH (cm)	AGB (t/ha)	Plot	SLF CHM	H (m)	DBH (cm)	AGB (t/ha)
89	27.2	18	24.7	263.77	144	29.1	19.7	34.1	312.63
91	9.6	8.7	12.6	56.76	147	26.9	20.8	35.9	160.51
95	12.0	10.4	13.2	71.31	148	19.5	19.3	22.2	233.53
103	26.5	20.6	35.8	184.05	152	32.8	19.1	27.5	360.33
115	37.0	12.4	14.0	240.42	155	25.2	16.2	24.0	238.52
119	10.4	9.1	14.3	56.97	156	28.1	19.6	27.8	189.95
121	26.6	18.4	25.8	312.77	158	35.3	29.6	35.8	497.8
122	24.3	25.8	51.6	205.03	159	13.8	10.4	11.3	41.09
124	27.5	18.2	21.2	304.07	259	26.6	20.6	39.2	193.12
125	26.0	20.1	25.3	514.01	278	N/A	22.1	32.1	271.99
126	30.4	21.7	36.7	230.4	293	19.9	17.7	37.1	224.26
129	24.4	19.1	20.8	362.33	294	23.4	21	37.4	225.82
130	25.9	21.7	18.3	491.75	314	24.8	21.3	41.4	302.69
132	31.0	14.2	19.4	241.27	317	18.9	15.1	42.9	228.8
135	30.9	23.7	39.7	338.21	318	10.8	10.1	26.4	71.37
138	21.8	16.7	25.8	229.14					



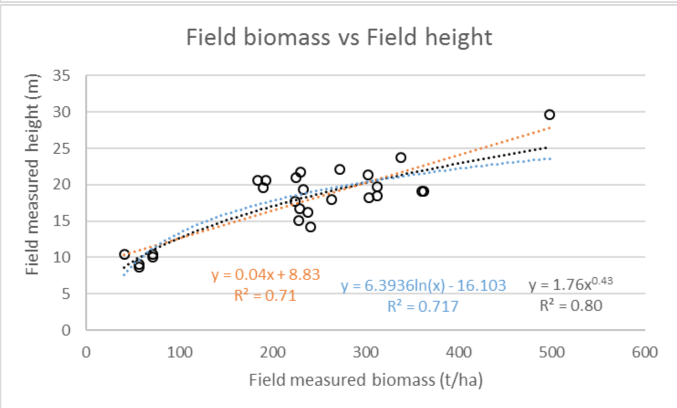


Fig. 4 Field measured biomass vs. field measured height. All plots (left panel) and without potentially erroneous plots (right panel).

2. CHM processing: filtered vs not-filtered

Initially, smoothed median filtered CHMs were produced. However, a closer examination revealed larger than expected changes between non-filtered and filtered data (Fig 5). Since higher differences were observed for the smoothed median filtered data it was decided to only use a median filter (3x3) when generating the CHMs. Completely unfiltered CHM have a noisy (salt and pepper) appearance particularly at resolutions below 20 m. The use of median filter incurs some loss of detail which is however compensated for by the significant noise reduction. The use of 'smooth' option alone produced sub-optimal results and its use was therefore discarded.

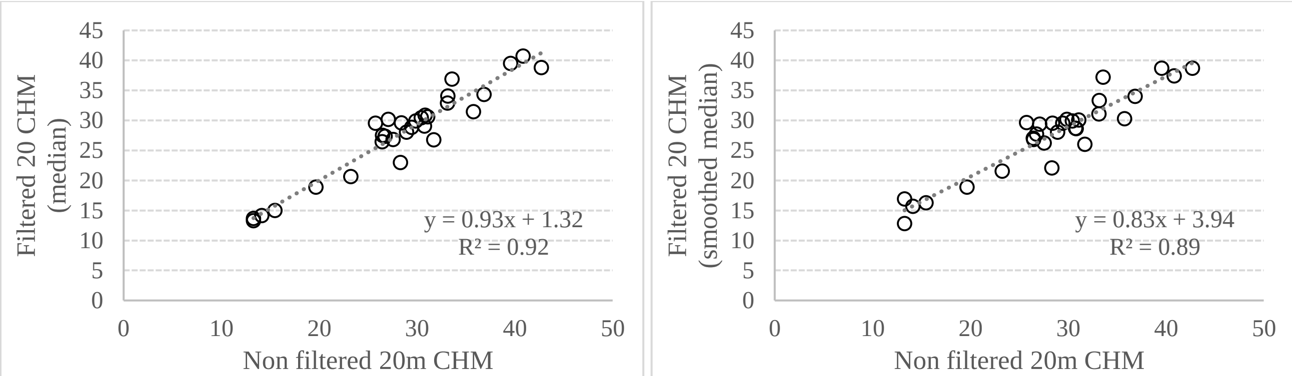


Fig. 5 The result of filtering options during CHM generation

<u>Intellectual property rights (IPR)</u>

These datasets are provided as <u>Permission by default</u>

Excerpt from ECOPOTENTIAL guidelines applying to IPR:

Permission by default: data are available to other ECOPOTENTIAL partners without the need for specific access permission. The custodian should be informed of each individual data download and intended use, and rules for co-authorship apply.

B) Example: Quality report by CERTH concerning Doñana Inundation maps

QUALITY ASSESSMENT OF CERTH's WP4 PRODUCTS in ECOPOTENTIAL
Status of 26.04.18
More info at imanakos@iti.gr

1. Overall statement

CERTH has used ground data, where available and provided by the PAs' actors. CERTH has repeatedly asked for validation by the PAs' personnel, but not all were responsive. CERTH has estimated, where possible, uncertainty measures in relation with the process of calculation of variables per se and has provided the results. CERTH has researched further and formulated (i) an approach for estimating uncertainty in the classification of inundated and aquatic vegetation areas, and (ii) a second one for assessing the uncertainty of the landscape biodiversity measures. The results showed a logical improvement in reporting and uncertainty quantifying potential; however, in lack of ground truth data, these have not been published or promoted as official project results.

2. Accuracy Assessment for Donana Inundation maps

The Donana inundation maps for 23 dates, based on CERTH's unsupervised approach, were uploaded to:

"/lustre/ECOPOTENTIAL/incoming/PAs/Donana/Inundation_maps_and_Hydroperiod_by_CERTH"

An inundation map has the form: YYYY_MM_DD_inundation_map_Donana_S2.tif, where 'YYYY' is the year, 'MM' is the month, and 'DD' is the day.

The accuracy assessment is performed using as ground truth 7 Landsat based inundation maps coinciding with Sentinel-2 inundation maps. The common acquisition dates are: 2015.12.29, 2016.06.06, 2016.07.16, 2016.08.25, 2016.10.04, 2016.12.23, 2017.06.01.

For each of the seven dates, Producer's Accuracy (PA), User's Accuracy (UA), Overall Accuracy (OA), and kappa coefficient (***k***) accuracy metrics are estimated for water and non-water classes. Then, for each statistic metric its mean value and standard (Std) deviation is estimated. In order to compensate for the uncertainty caused in the accuracy assessment due to the lower spatial resolution of the Landsat derived inundation maps, two accuracy assessment results are presented: (i) boundary regions between inundated and non-inundated pixels are excluded from the accuracy estimation (Case 1), and (ii) boundary regions are included (Case 2). The results for the two cases are the following:

Case 1	**PA (water class)**	**UA (water class)**	**PA (non-water class)**	**UA (non-water class)**	**OA**	***k***
Mean	91.1%	87.98%	98.63%	98.73%	97.69%	0.8771
Std	3.09%	12.42%	1.21%	1.32%	1.39%	0.0715

Case 2	**PA (water class)**	**UA (water class)**	**PA (non-water class)**	**UA (non-water class)**	**OA**	***k***
Mean	85.74%	83.79%	97.82%	98.04%	96.40%	0.8201
Std	4%	13.69%	1.76%	1.52%	1.8%	0.0767

The above accuracy assessment results may be extrapolated to the inundation maps corresponding to all Donana inundation maps, due to the lack of Landsat ground truth inundation maps corresponding to each date.

2. Statistics **for Hydroperiod maps generated for Donana and Camargue**

An annual hydroperiod is generated from series of inundation maps, which have been produced for dates falling in the time period between the starting and the ending date of hydroperiod, by applying the following interpolation approach. For two dates separated by n days, the occurrence of water is compared. If a pixel is inundated on both dates, then it is assumed inundated for n-days, but n/2 days otherwise. The total number of days of inundation is determined by accumulating the water masks throughout the year.

The quality of the Hydroperiod map depends on the frequency of the available inundation maps falling in each hydroperiod cycle. The quality metrics of hydroperiod are: (i) the mean of inundation maps frequency, (ii) the standard deviation of inundation maps frequency, (iii) Gini index (Díaz-Delgado et al. 2016) that gives an assessment on the evenness of the frequency of available inundation maps along the flooding cycle (values from 0 to 1). The higher the Gini Index for a cycle the lower is the representativeness of the hydroperiod value per pixel for every flooding cycle.

Díaz-Delgado, R.; Aragonés, D.; Afán, I.; Bustamante, J. Long-term monitoring of the flooding regime and 592 hydroperiod of Doñana marshes with Landsat time series (1974–2014). Remote Sensing, 2016, 8(9), 775.

2.1 Donana

Statistics on the two Donana Hydroperiod maps for 1-Sept-2015 to 31-Aug-2016 and 1-Sept-2016 to 31-Aug-2017 annual cycles relying on S-2 inundation maps (and Landsat inundation maps for the first annual cycle due to the lack of S-2 inundation maps) are given in the following:

The Donana Hydroperiod maps for 1-Sept-2015 to 31-Aug-2016 and 1-Sept-2016 to 31-Aug-2017 were uploaded to:

“/lustre/ECOPOTENTIAL/incoming/PAs/Donana/Inundation_maps_and_Hydroperiod_by_CERTH”

with filenames:

“Donana_Hydroperiod_from_1st_Sept_2015_to_31st_Aug_2016_using_Sentinel_2_and_Landsat_inundation_maps.tif”

“Donana_Hydroperiod_from_1st_Sept_2016_to_31st_Aug_2017_using_Sentinel_2_inundation_maps.tif”

Quality metrics of Donana hydroperiod map for 1-Sept-2015 to 31-Aug-2016: Mean Frequency: 29.8182 days, Standard Devation: 26.8843 days (mostly due to the winter conditions, which bring in seasonal restrictions for the availability of cloud free acquisitions), Gini Index: 0.4512 .

Quality metrics of Donana hydroperiod map for 1-Sept-2016 to 31-Aug-2017: Mean Frequency: 20.5882 days, Standard Devation: 21.0566 days (mostly due to the winter conditions, which bring in seasonal restrictions for the availability of cloud free acquisitions), Gini Index: 0.3966 .

2.2 Camargue

Statistics on the Camargue Hydroperiod map for 1-Sept-2016 to 31-Aug-2017 annual cycle relying on S-2 inundation maps.

The Camargue Hydroperiod map for 1-Sept-2016 to 31-Aug-2017 was uploaded to:

“/lustre/ECOPOTENTIAL/incoming/PAs/Camargue/Inundation_maps_and_Hydroperiod_by_CERTH”

with filename:

“Camargue_Hydroperiod_from_1st_Sept_2016_to_31st_Aug_2017_using_Sentinel_2_inundation_maps.tif”

Quality metrics of Camargue hydroperiod map for 1-Sept-2016 to 31-Aug-2017: Mean Frequency: 38.8889 days, Standard Deviation: 18.1758 (mostly due to the winter conditions, which bring in seasonal restrictions for the availability of cloud free acquisitions), Gini Index: 0.2254 .

3. Statistics **for Phenology metrics maps in Donana Marshes estimated for the period 1-Dec-2016 to 30-Nov-2017**

The files containing the estimated statistics and other relevant files are included in "Phenology Statistics" folder. The layer "Donana_Marshes_objects_in_10m_resolution.tif"containing habitat objects was generated based on the shapefile "Donana_habitat_map_marsh.shp" (the habitat map of the marshland) and it has resolution 10x10m. The Habitat Class IDs and their corresponding Habitat Class names can be found in columns "HABITAT_UE,C,103" and "ID_ASOCIAC,C,10", contained file "Donana_habitat_map_marsh.dbf".

Statistics on the Donana marshes Green up day layer generated for the period from 1-Dec-2016 to 30-Nov-2017

The Green up day layer with name "Greenup_day_Dec2015_Nov2016.tif" was uploaded to:

"/lustre/ECOPOTENTIAL/incoming/PAs/Donana/S2-PhenologyMetrics&PerPixelISODATAclassification_by_CERTH"

Based on the Green up day layer, which provides the Green up day per pixel, two outputs are generated. The first output provides the mean Green up day value and the standard deviation per habitat class ID as a list of values. The second output is a layer with two bands, where the first band gives the mean Green up day of pixels per object and the second band gives the standard deviation of the Green up day of pixels per object.

The two outputs are named as:

- "Statistics_of_Layer_Greenup_day_Dec2015_Nov2016.xlsx"
- "Mean_and_std_for_Layer_Greenup_day_Dec2015_Nov2016_per_object.tif"

Statistics on the Donana marshes Senescence day layer generated for the period from 1-Dec-2016 to 30-Nov-2017

The Senescence day layer with name "Senescence_day_Dec2015_Nov2016.tif" was uploaded to:

"/lustre/ECOPOTENTIAL/incoming/PAs/Donana/S2-PhenologyMetrics&PerPixelISODATAclassification_by_CERTH"

Based on the Senescence day layer, which provides the Senescence day per pixel, two outputs are generated. The first output provides the mean Senescence day value and the standard deviation per habitat class ID as a list of values. The second output is a layer with two bands, where the first band gives the mean Senescence day of pixels per object and the second band gives the standard deviation of the Senescence day of pixels per object.

The two outputs are named as:

- "Statistics_of_Layer_Senescence_day_Dec2015_Nov2016.xlsx"
- "Mean_and_std_for_Layer_Senescence_day_Dec2015_Nov2016_per_object.tif"

Statistics on the Donana marshes Highest NDVI day layer generated for the period from 1-Dec-2016 to 30-Nov-2017

The Highest NDVI day layer with name "Max_day_Dec2015_Nov2016.tif" was uploaded to:

"/lustre/ECOPOTENTIAL/incoming/PAs/Donana/S2-PhenologyMetrics&PerPixelISODATAclassification_by_CERTH"

Based on the Highest NDVI day layer, which provides the Highest NDVI day per pixel, two outputs are generated. The first output provides the mean Highest NDVI day value and the standard deviation per habitat class ID as a list of values. The second output is a layer with two bands, where the first band gives the mean Highest NDVI day of pixels per object and the second band gives the standard deviation of the Highest NDVI day of pixels per object.

The two outputs are named as:

- “Statistics_of_Layer_Highest_NDVI_day_Dec2015_Nov2016.xlsx”
- “Mean_and_std_for_Layer_Highest_NDVI_day_Dec2015_Nov2016_per_object.tif”

Statistics on the Donana marshes Number of NDVI peaks layer generated for the period from 1-Dec-2016 to 30-Nov-2017

The Number of NDVI peaks layer with name “Max_day_Dec2015_Nov2016.tif” was uploaded to:

“/lustre/ECOPOTENTIAL/incoming/PAs/Donana/S2-PhenologyMetrics&PerPixelISODATAclassification_by_CERTH”

Based on the Number of NDVI peaks layer, which provides the Number of NDVI peaks per pixel, two outputs are generated. The first output provides the mean Number of NDVI peaks value and the standard deviation per habitat class ID as a list of values. The second output is a layer with two bands, where the first band gives the mean Number of NDVI peaks of pixels per object and the second band gives the standard deviation of the Number of NDVI peaks of pixels per object.

The two outputs are named as:

- “Statistics_of_Layer_Number_of_peaks_Dec2015_Nov2016.xlsx”
- “Mean_and_std_for_Layer_Number_of_peaks_Dec2015_Nov2016_per_object.tif”

4. Statistics on the layer containing the total number of abrupt changes for the period 2007-2016 for Donana Marshes

The files containing the estimated statistics and other relevant files are included in “BFAST statistics” folder. The layer “Donana_Marshes_objects_in_30m_resolution.tif”containing habitat objects was generated based on the shapefile “Donana_habitat_map_marsh.shp” (the habitat map of the marshland) and it has resolution 30x30m. The Habitat Class IDs and their corresponding Habitat Class names can be found in columns “HABITAT_UE,C,103” and “ID_ASOCIAC,C,10”, contained file “Donana_habitat_map_marsh.dbf”.

The “Marshes_maximum_number_of_breaks_2007_to_2016.tif” layer containing the total number of abrupt changes for the period 2007-2016 was estimated relying BFAST algorithm and was uploaded to:

“/lustre/ECOPOTENTIAL/incoming/PAs/Donana/BFAST_Phenological_Change_Detection_in_Marshes_2007to_2016_using_Landsat_NDVI_series_by_CERTH/Output”

Two outputs are generated based on this layer. The first output provides the mean value of total number of abrupt changes and the standard deviation per habitat class ID as a list of values. The second output is a layer with two bands, where the first band gives the mean value of the total number of abrupt changes of pixels per object and the second band gives the standard deviation of the total number of abrupt changes of pixels per object.

The two outputs are named as:

- “Statistics_of_Layer_Marshes_maximum_number_of_breaks_Dec2015_Nov2016_per_object.xlsx”
- “Mean_and_std_for_Layer_Marshes_maximum_number_of_breaks_Dec2015_Nov2016_per_object.tif”

5. Accuracy of land cover maps used for estimating the landscape fragmentation metrics

Landscape fragmentation metrics could not be assessed with measurements on the ground. Thus, the accuracy provided is the combination of the a) accuracy accompanying the commonly approved statistics’ equations, and b) the accuracy of the input layers. For a), since measures are equivalent with ones provided by the publicly available software Fragstats (https://www.umass.edu/landeco/research/fragstats/fragstats.html), no other indication is given. For b) following subsections present existing information.

5.1 Samaria

For Samaria, the accuracy of the land cover maps per year used for estimating the landscape fragmentation measures is given in the following Table:

	1985	1995	2000	2005	2010	2015
Overall Accuracy (%)	88.8	89.2	89.2	87.8	85.7	92.3
Kappa Coefficient	0.85	0.85	0.85	0.83	0.8	0.89

The Fragmentation measures produced per year are given in the following folders:

1985:

- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_3x3m/1985/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/1985/Outputs

1995:

- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_3x3m/1995/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/1995/Outputs

2000:

- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_3x3m/2000/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/2000/Outputs

2005:

- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_3x3m/2005/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/2005/Outputs

2010:

- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_3x3m/2010/Outputs

- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/2010/Outputs

2015:

- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_3x3m/2015/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Samaria/PosterInputs/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/2015/Outputs

5.2 Lake Prespa

For Lake Prespa, a subset of the Corine Land Cover map 2012 was used for estimating the fragmentation measures. There is no accuracy assessment for this specific land cover subset, but for the Corine Land Cover map of a group of countries (e.g. Albania - Serbia - FYROM – Montenegro) containing this land cover map subset. The accuracy assessment results for this group of countries according to Corine Land Cover 2012 Final Validation Report are:

Blind Analysis: Overall Accuracy: 82.08%, Confidence interval of 95%: 3.87%
Plausibility Analysis: Overall Accuracy: 93.46%, Confidence interval of 95%: 2.18%

The Fragmentation measures produced for 2012 are given in the following folders:

2012:

- /lustre/ECOPOTENTIAL/incoming/PAs/Lakes Ohrid_Prespa/Landscape_biodiversity_indicators_by_CERTH/Resolution_3x3m/2012/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Lakes Ohrid_Prespa/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/2012/Outputs

5.3 La Palma

The La Palma land cover map of 2007 was manually drawn based on aerial images and expert knowledge. No accuracy assessment results have been provided for La Palma land cover map, following a specific request of ours to the respective Management Authority, which provided the map.

The Fragmentation measures produced for 2007 are given in the following folders:

2007:

- /lustre/ECOPOTENTIAL/incoming/PAs/Canary_Islands_La_Palma/Landscape_biodiversity_indicators_by_CERTH/Resolution_3x3m/2007/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Canary_Islands_La_Palma/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/2007/Outputs

5.4 Montado

For Montado, two land cover maps of 2007 and 2012 were used for estimating fragmentation measures. The land cover of 2007 was provided by Portuguese Geographic Institute, and the one for 2012 is a subset of Corine Land Cover 2012, where accuracy is given for the complete Corine Land Cover map 2012 of Portugal. The accuracy per land cover map is the following:

Accuracy of Land Cover for 2012 according to Corine Land Cover 2012 Final Validation Report:
Blind Analysis: Overall Accuracy: 75.91%, Confidence interval of 95%: 3.68%
Plausibility Analysis: Overall Accuracy: 84.80%, Confidence interval of 95%: 3.17%

Accuracy of Land Cover 2007:
Overall Accuracy (reported in the files' own metadata): 85.13%, Confidence interval of 95%: 2%.

The Fragmentation measures produced for 2007 and 2012 are given in the following folders:

2007:

- /lustre/ECOPOTENTIAL/incoming/PAs/Montado/Landscape_biodiversity_indicators_by_CERTH/Resolution_10x10m/2007/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Montado/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/2007/Outputs

2012:

- /lustre/ECOPOTENTIAL/incoming/PAs/Montado/Landscape_biodiversity_indicators_by_CERTH/Resolution_10x10m/2012/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Montado/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/2012/Outputs

6. Products **without Quality Assessment**

6.1 Sierra Nevada: Landscape fragmentation Measures

For Sierra Nevada, three land cover shapefiles were provided for estimating fragmentation measures were provided. Expect for the scale of the shapefiles which is 1:250001 for "SN_vege_1956_99_03_07_25000.shp" and "SN_vege_56_77_84_99_03_07_reduc_field.shp" shapefiles and 1:10000 for "SN_vege_1996_2006_10000.shp" shapefile, no additional accuracy information was provided so far. We forwarded you our email conversation with Ricardo Moreno from University of Granada. He sent us some links with content in Spanish to find further information on the generation of land cover shapefiles. You could have a look at the links in case you wish to search if there is available information on the accuracy, since we do not understand Spanish.

The mail sequence:

" -------------------

Hi Georgios:

You´re right regarding the scale.

There is no geographic accuracy assessment of the land cover map generated per year,The accuracy is defined by the scale used to photointerpret the images.

The information in table 14 is referred to the validation of the categorization of the land use or vegetation (thematic validation).

2018-04-24 12:47 GMT+02:00 Georgios Kordelas <kordelas@iti.gr>:

Dear Ricardo,

If we understood well, both shapefiles "SN_vege_1956_99_03_07_25000.shp" and "SN_vege_56_77_84_99_03_07_reduc_field.shp" have scale 1:25000, while "SN_vege_1996_2006_10000.shp" has scale 1:10000. Please confirm if this is correct.

Additionally, we looked for information regarding the accuracy assessment of the land cover map generated per year, such as the Overall Accuracy and/or additional metrics (i.e. the kappa coefficient). Please inform us if there is such a kind of information in the links you have sent us, since we do not know Spanish. We found some numbers in Tabla 14 of "http://www.juntadeandalucia.es/medioambiente/web/Bloques_Tematicos/Publicaciones_Divulgacion_Y_Noticias/Documentos_Tecnicos/mapa_usos_guia_tecnica/3_proceso_metodologico.pdf", but we are not sure which date and data layer do they account for. Could you please provide some additional support in retrieving the information?

23-Apr-18 10:47 AM, o Ricardo Moreno:

Dear Georgios:

The Map of Land Use and Coverage in Andalusia is a project carried out by the Regional Ministry of the Environment since 1987, whose main task is to make a cartographic and statistical follow-up of changes in the typology of occupation of the territory of the Autonomous Community of Andalusia.

This project has its precedents in the CORINE-Land Cover European Union Program, developed in Spain in 1987, whose main objective was the realization of a Land Occupancy Map at a European level, at a scale of 1 / 100,000, starting from the photointerpretation of satellite images. This European program is updated periodically, maintaining the same technical and methodological characteristics, which makes it possible to study and analyze changes in the occupation of European territory and its permanent monitoring.

The Map of Land Use and Coverage of Andalusian Land was born as an autonomous project based on the adaptation of the CORINE methodology to the physical-territorial reality of the Andalusian community. It is proposed with a four-year review period, with five updates to date: 1991-1995-1999 on a scale of 1 / 50,000, and 1999-2003-2007 on a scale of 1 / 25,000.

The land uses maps of the territory of the years 56_77_84_99_03_07 on a scale of 1: 25,000 using the level of maximum disaggregation of the legend of land use and vegetation cover of Andalusia. Cartography generated from the photointerpretation of photogrammetric flights and Landsat TM satellite images. Those polygons with uses that are forestry, in addition, are assigned the dominant and codominant species in the arboreal and shrub layer, if they existed, as well as the present vegetation formation.

More information in spanish you can find HERE.

Regarding Vegetation map at detail scale 1: 10,000, the work consisted in the recognition by photointerpretation of homogeneous vegetation units among themselves and the subsequent verification and assignment of information to each one of said units, with the support of numerous field works. In the methodology for its elaboration, we started with black and white and infrared aerial orthophotos and numerous field visits made by the work teams of the specialized Andalusian universities and companies that have been responsible for gathering information, experts knowledgeable about the vegetation of your work areas. More information in spanish you can find HERE.

--------------------"

In the following, information on the FTP folders containing the landscape measures is given.

The Fragmentation measures produced for 1956, 1999, 2003, 2007 based on shapefile “SN_vege_1956_99_03_07_25000” are given in the following folders:

1956:

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1956_99_03_07_25000/Resolution_3x3m/1956/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1956_99_03_07_25000/Resolution_30x30m/1956/Outputs

1999

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1956_99_03_07_25000/Resolution_3x3m/1999/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1956_99_03_07_25000/Resolution_30x30m/1999/Outputs

2003

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1956_99_03_07_25000/Resolution_3x3m/2003/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1956_99_03_07_25000/Resolution_30x30m/2003/Outputs

2007

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1956_99_03_07_25000/Resolution_3x3m/2007/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1956_99_03_07_25000/Resolution_30x30m/2007/Outputs

The Fragmentation measures produced for 1956, 1977, 1984, 1999, 2003, 2007 based on shapefile "SN_vege_56_77_84_99_03_07_reduc_field" are given in the following folders:

1956:

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_3x3m/1956
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_30x30m/1956/Outputs

1977:

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_3x3m/1977/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_30x30m/1977/Outputs

1984:

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_3x3m/1984/Outputs

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_30x30m/1984/Outputs

1999:

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_3x3m/1999/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_30x30m/1999/Outputs

2003:

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_3x3m/2003/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_30x30m/2003/Outputs

2007:

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_3x3m/2007/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_56_77_84_99_03_07_reduc_field/Resolution_30x30m/2007/Outputs

The Fragmentation measures produced for 1996_2006 based on shapefile "SN_vege_1996_2006_10000" are given in the following folders:

1996_2006:

- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1996_2006_10000/Resolution_3x3m/1996_2006/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Sierra_Nevada/Landscape_biodiversity_indicators_by_CERTH/SN_vege_1996_2006_10000/Resolution_30x30m/1996_2006/Outputs

6.2 Curonian Lagoon: Landscape **fragmentation Measures**

For Curonian Lagoon, land cover shapefile "Curonian_landuse_2013_2014" was provided for estimating fragmentation measures. No accuracy information accompanied the shapefile.

The produced Fragmentation measures are given in the following folders:

2013_2014:

- /lustre/ECOPOTENTIAL/incoming/PAs/Curonian_Lagoon/Landscape_biodiversity_indicators_by_CERTH/Resolution_3x3m/2013_2014/Outputs
- /lustre/ECOPOTENTIAL/incoming/PAs/Curonian_Lagoon/Landscape_biodiversity_indicators_by_CERTH/Resolution_30x30m/2013_2014/Outputs

6.3 Camargue Inundation maps

Inundation maps for Camargue and 10 dates, based on CERTH's unsupervised approach, were uploaded to:

"/lustre/ECOPOTENTIAL/incoming/PAs/Camargue/Inundation_maps_and_Hydroperiod_by_CERTH"

No validation data was provided by the Park Managers to perform accuracy assessment for the inundation maps.

6.4 Danube **Delta Inundation maps**

Inundation maps for Danube Delta and 10 dates, based on CERTH's unsupervised approach, were uploaded to:

"/lustre/ECOPOTENTIAL/incoming/PAs/Danube_Delta/Inundation_maps_and_Hydroperiod_by_CERTH"

No validation data was provided by the Park Managers to perform accuracy assessment for the inundation maps.

6.5 Danube Delta Hydroperiod

The Danube Delta Hydroperiod map for 1-Sept-2016 to 31-Aug-2017 was uploaded to:

"/lustre/ECOPOTENTIAL/incoming/PAs/Danube_Delta/Inundation_maps_and_Hydroperiod_by_CERTH"

with filename:

"Danube_Delta_Hydroperiod_from_1st_Sept_2016_to_31st_Aug_2017_using_Sentinel_2_inundation_maps.tif"

Not all pixels could use the data from all inundation maps falling within the time period from_1st_Sept_2016_to_31st_Aug_2017 to estimate the hydroperiod duration, as clouds were partially prohibiting time interpolation between dates. In such cases, cloud affected pixels were not considered and the hydroperiod was calculated based on the next previous and following non-cloud affected instances. Therefore, it was not feasible to estimate unique quality metrics of hydroperiod (e.g. mean inundation maps frequency, the standard deviation of frequency, Gini index) that will represent the complete hydroperiod map.

6.6 Donana Phenology Curves

Phenology curves for 66 classes containing vegetation, according to the detailed land cover map of Donana generated on 2011 by CSIC, have been generated for five selected indices and for year 2016. These curves can be found in FTP folder: /lustre/ECOPOTENTIAL/incoming/PAs/Donana/Donana_PhenologycurvesbyCERTH

No validation of the phenology curves has been performed.