CS GROUP

WORKING BY YOUR SIDE
TO TACKLE YOUR CRITICAL CHALLENGES

SCO SPACE CLIMATE OBSERVATORY

STOCK WATER

Final Report

Université de Strasbourg

Sertit ICUBE

Table of signatures

FINAL REPORT	NAME	ROLE	DATE	SIGNATURE
Written by :	Sabrine Amzil Aurélien Bricier Thomas Ledauphin Gael Nicolas Benjamin Tardy Hervé Yesou	Project Team	20/01/2023	p.o. Gael Nicolas
Checked by :	**Frédéric Liège** **Santiago Pena Luque**	Head of Radar Simulation & Studies Team	20/01/2023	

Successive versions

VERSION	DATE	MOTIF
1.0	10/01/2022	Creation
1.2	22/02/2022	End of Phase I report
1.3	21/10/2022	End of Phase II report
1.9	07/12/2022	Final report, interim version
2.0	21/12/2022	Final Report_fr
2.1	24/01/2023	Final Report_fr (updated)
2.1	24/01/2023	Final Report_en
2.2	09/03/2023	Final synthetic Report_en

Table of contents

Table of figures

Table of tables

1 Introduction

This document presents the final report of the SCO STOCK WATER project.

2 Study objectives

2.1 Issues

The monitoring and understanding of our environment becomes ineluctably a major geopolitical issue, particularly critical in areas such as ecology, economy, energy, food security, land use planning, risk management or security. The exact knowledge in real time of the geophysical state of our planet and of its evolution in time becomes therefore a major stake.

From a strategic point of view, having a tool allowing the evaluation of the water resources of a third party with a relevant precision scale is an essential advantage. It is then possible to evaluate not only the water resources but also directly the evolution of the capacities of this country in terms of irrigation (food security, agricultural production capacities, and therefore imports/exports), energy production (hydroelectricity, power plant cooling...), industrial production (and therefore dependence) for certain sectors, supply of drinking water to the populations (conflicts over water, dimensioning of humanitarian aid, etc...).

The effects of the drought in the summer of 2022, in a tense geopolitical context, have particularly highlighted the sensitive nature of the management of this resource, for which there is currently no method for monitoring volumes on a large scale.

From a technical point of view, due to the criticality of the constraints and the extent of the data to be used, precise, automatic and systematic approaches are essential. It is therefore essential to develop tools for processing satellite data, in order to allow the exploitation and interpretation of these data in constrained time, with the aim of generating high value-added operational products to address these issues, for example the particular case of monitoring water resources.

2.2 Technical context

Beyond the issue of water surfaces, the management of captured resources (dams, reservoirs, etc.) is particularly complex insofar as there are more than 45,000 major reservoirs in the world, which cover critical needs (domestic and industrial water supply, irrigation, energy production, flood control, etc.), and that almost all the major rivers in the world are affected by the presence of dams.

A method allowing the monitoring in constrained time of the state of the volumes of water stored in these various infrastructures thus constitutes a major (and henceforth critical) stake. Such a tool would allow the monitoring of water management strategies (allocation of extreme climatic events...) at local, regional (or even larger) scales, but also the understanding of longer-term dynamics such as the consideration of climate change effects and the elaboration of adaptation strategies.

Water detection is a complex task, and various methods have been developed to extract water bodies from remote sensing data acquired in different observation domains. Hydrologists and hydraulic engineers have a strong demand for the ability to know in near-real time where water is and how water surfaces evolve. The SURFWATER chain, which uses Sentinel 1 and 2 data to generate an observation map of water surfaces over a given area of interest, provides an answer to this problem.

2.3 Objectives of the project

The objective of the SCO STOCKWATER project is to provide users (dam managers, experts, operational decision-makers) with a global solution for dynamic monitoring of dam water stocks in constrained time over the entire globe, at scales relevant to resource management.

From a technical point of view, this demonstration is based on a scaling up and generalization to other regions of the world of methods evaluated in France, at least insofar as relevant data are available before a possible leap towards an application adaptable to the national or even global level.

The operational methods previously developed and implemented in the framework of the analysis of water stocks by satellite tracking in France have demonstrated their relevance. The SCO STOCKWATER project aims to characterize and demonstrate the generality and adaptability of these methods by applying them to a very large number of sites, for various regions and conditions of use. If this demonstration proves conclusive, the method may open major perspectives in the possibility of generalization or adaptability for water resources monitoring at the global level.

The dynamic results of resource monitoring are made available to users through an open web platform

2.4 Constraints

2.4.1 Availability of in-situ data

Due to the large volume of sites to be processed, it is important to ensure the availability of relevant data for each of them. One of the major constraints of the project is to have enough valid sites, i.e., for which *in situ* data are available.

The list of available supplies at T0 includes *in situ* data in Occitania, Andalusia and India, but the project in its completed form is dependent on the number of sites covered by these data, as well as the availability of data from local partners in the other areas to be treated (Laos, Burkina-Faso and Tunisia), within a timeframe compatible with the completion of the project on the one hand, but beyond that, for updates within a timeframe that is relevant for operational use. Beyond the technical aspect, there is the problem of data accessibility, at relevant frequencies and precision, which cannot be ascertain, even for France (some data remain confidential).

In the event that such *in situ* data are not available, it may be necessary to study alternative solutions in other regions, or to modify the distribution of sites based on the other targeted study areas.

2.4.2 Data preparation

Regarding earth observation data, the study will exploit the input products already available in Theia platform (notably for France, Spain, India, Tunisia, Laos and Burkina-Faso). In the case of sites not covered by Theia, the implementation of MAJA will then be required to have the Sentinel-2 L2A data necessary for the Surfwater chain.

2.4.3 Processing chains evolution

Evolutions of the Surfwater-Postprocess chain are necessary, and improvements are to be brought for the HSV modeling, in order to pass from the state of Proof of Concept to that of the "dem4water" prototype production chain. These evolutions have been integrated from the Proof-of-Concept Phase I implementation.

The selection of the improvements to be implemented in the HSV modeling chain was made in agreement with the CNES, while ensuring the control of the development time with the aim of achieving a usable tool in ergonomic conditions compatible with a pre-operational tool and a feasibility demonstration, without compromising the project schedule.

2.4.4 Scope of study

Given the large initial number of sites (250) and the observation window (3 years) targeted, the study process initially proposed in the description of the offer has been adapted in order to:

- ✓ To ensure greater flexibility in data availability times,
- ✓ To be able to decouple the tasks and carry them out in parallel,
- ✓ Validate an end-to-end workflow on a subset of controlled and representative sites before scaling up,
- ✓ Allow rapid iterative evolutions on this subset of sites to accelerate reactivity and validation as an extension of the implementation of improvements,
- ✓ Decouple the uncertainties:
    - Of data availability on the initial 250 sites,
    - Functional validation of the end-to-end chain,
    - The complexity of scaling up through massive parallelization.

3 Organisation of the study

3.1 Methodology

In order to ensure the relevance of the method on a global scale, a controlled validation framework was set up prior to the generalization of the method to all the sites on the initial list. The purpose of this prior validation was to address the various areas of variability that are likely to emerge during the generalization of the methods implemented.

In particular, the following assumptions were taken into account

- ✓ The variability of the nature of the input data:
    - The different DTMs available present a necessary have a very great heterogeneity availability (in terms of nature, resolution, quality, and limitations). They have therefore been analyzed in order to identify their constraints and to identify the intrinsic limitations of each type of DTM that will be considered.
    - The availability of a relevant volume of observations to define image time series covering the changes to be observed. Some areas of the globe do not have sufficient coverage of non-cloudy Sentinel-2 optical observations necessary to monitor water surfaces, just as the availability of Sentinel-1 radar acquisitions is more or less dense depending on the geographical area.
- ✓ The variety of topologies of water storage sites, both in terms of the natural context and the techniques for setting up artificial reservoirs.

The implementation of the method for a limited (but representative) number of targets has allowed us to evaluate the performance of the method in conditions representative of the change of scale induced by the Phase II configuration and the final configuration, which will be encountered during the transition to the global scale. In order to be representative, these conditions must cover a relevant and representative number of study cases. The analysis of these results, and their validation, will in turn allow the transition to the global scale.

Basically, the methodology is based on three major steps necessary to monitor a site:

1. Modeling of the Height/Surface/Volume relationship of each dam, based on DTM analysis (dem4water chain),

2. Determination of the water surface of the site at each observation by Sentinel-1 and Sentinel-2 satellites (Surfwater chain) to collect the S(t) series,

3. Generation of the S(t), V(t) and TR(t) series (Surfwater-Postprocess chain).

These three steps will be applied to a number of sites sufficiently representative to allow a global scale.

3.2 Articulation of the study

3.1 Structure

Given the scope of the implementation and validation framework required (initially 250 sites), it was proposed to organize the tasks described in two consecutive phases:

Phase I (Proof of Concept): implementation of the entire methodology in end-to-end mode on a reduced number of sites, as heterogeneous and representative as possible, on which the data (input and validation) are already available, especially in the acquisition areas in Occitania, Andalusia and India, for which *in situ* data are available This first phase is articulated according to the following steps:

- ✓ Selection of a reduced number of targets, as representative as possible, depending on the availability of input and validation data,

- ✓ Preparation of the data,
- ✓ Generation of the water masks on the periods of availability of the *in situ* data,
- ✓ Modeling of the Height, Surface, Volume relationships (H/S/V modeling) for each of the available DTMs,
- ✓ Generation of time series S(t), V(t) and Fill Rate TR(t) for the reduced selection of sites,
- ✓ Validation on these sites:

By automatic comparative analysis to *in situ* data,

By comparison of the performances according to the DTMs used for the H/S/V modeling.

In the framework of this Phase I, the tools necessary for these steps (orchestrator, automatic comparison tool with *in situ* data, improvements and evolutions necessary to the existing tools...) have been implemented. At the end of this initial phase, the analysis of the quality of the DTMs allowed to classify them according to the quality of the generated models, and to exclude some of them from Phase II.

Phase II (Scaling): exploiting the environment implemented in Phase I to apply the methodology to all 250 sites. The steps comprising this Phase II were as follows:

- ✓ Selection of sites complementary to Phase I, in conjunction with the expanded availability of *in situ* data,
- ✓ Preparation of the data for these new sites, as well as the complementary data from the Phase I sites, ,
- ✓ Generation of water masks for all sites,
- ✓ H/S/V modeling for the new sites for each of the DTM types considered in this step,
- ✓ Generation of S(t), V(t) and TR(t) time series for all sites and for each model (one per DTM),
- ✓ Validation of results:
    - By automatic comparative analysis to *in situ* data;
    - By comparison of the performances according to the DTMs used for the H/S/V modeling;
- ✓ Restitution of the results in the form of a dashboard with sufficient functionalities for a pre-operational tool for the CNES partners, for generic needs at first.

4 Phase I process

4.1 Demonstrator development

4.1.1 Surfwater

The Surfwater chain detects continental water surfaces using observations from the Sentinel-2 satellites (L2A products) as optical data and Sentinel-1 (GRD L1C products) as radar data.

Several products are generated by the Surfwater chain, for both optical and radar modes:

- ✓ A detection mask of the water surfaces called "instantaneous" because relative to an observation,
- ✓ A mask of synthesis of the instantaneous masks on a period of 10 days,
- ✓ A monthly accumulation mask based on all the instantaneous masks of the month.

The generation of Surfwater products requires the preparation of input data over the required period for a selection of tiles. The methods of recovering S2 L2A optical data via the Amalthée API on the CNES Datalake and S1 L1C radar data via eodag (S1tiling) were quickly confronted with server limitations. These problems penalized the project and led to blocking points on this task.

A significant effort to improve and adapt the recovery methods was provided in relation to the limitations of the servers.

4.1.2 DEM4water

The DEM4water chain calculates the estimation of the relationship between the variables: water height H, water surface S and stored volume V of dam reservoirs from the information of a surrounding DEM. The chain also offers the possibility of comparison with a reference model if available.

Due to its prototype state, the chain did not meet both the generic production needs and the large volume of data required by the Stockwater project. Evolutions have been made in order to automate and industrialize the chain.

The size of the area of interest was set "by hand" in the code and was not necessarily suitable for the different shapes of dams, in addition to the unnecessary processing time if the area was oversized in order not to risk missing a part of the dam (without having the guarantee of a good result, because of the variability of the shape of the structures).

A tool was developed to define more precisely the bounding box of the dams from the envelope calculation and to provide the associated radius.

The parameterization of the list of dams to be processed has been implemented, replacing their names and identifiers written in hard copy in the script. For this purpose, a tool was developed to generate the list of IDs/DAMS from a database file in geojson format.

The processing, initially carried out in series, required for certain dams (India) a significant processing time. An optimization was carried out by parallelizing the processing by dam.

Among the evolutions, we note the implementation of tree structure according to the possibility of different DEMs, with the creation of generic and explicit names, the integration of the new format of the databases and its attributes, the modification of the principle of the ID within the chain with the possibility of choosing a specific ID, the passage in argument of reference model.

A consistency step between units has been made in the comparison module of estimated and reference models dealing with m2 and hectares, with the same for volumes. The units handled now match the international system.

4.1.3 Surfwater_Postprocess

The goal of this step is to use the Surfwater masks in conjunction with the estimated models to generate time series of input values such as water surface and volume for each dam.

The Surfwater_Postprocess chain has several methods of counting water surfaces, offering the possibility of choosing among the various Surfwater products (instantaneous, synthetic and monthly occurrence masks), available for both optical and radar.

A post-processing is carried out on the water masks, using geolocalisation of the the water bodies by a point in latitude and longitude, the pixels of the masks are processed and counted to provide a status relative to the temporal evolution of the surfaces of the dams. Their surface is converted into volume, thanks to an estimated law computed by the DEM4water chain.

The metrics extraction and comparison modules are based on the following input data

- ✓ a list of water masks computed by the Surfwater chain,
- ✓ the database of the dams of the studied site (provided by the SERTIT),
- ✓ the table of conversion of surfaces in volumes corresponding to the estimated law calculated by the chain DEM4water,
- ✓ the reference time series from the *in situ* references (provided by SERTIT).

Except for the Surfwater masks, these entries have been modified in terms of format, then iteratively in terms of content and fields, requiring successive adaptations of the Surfwater_Postprocess string.

The dam database was changed from shapefile to geojson format with new attributes.

The create_water_mask tool was therefore also adapted to generate the output file in the new format.

In order to maintain consistency with the file formats used in the Stockwater project, the estimated laws used as a lookup table were changed from csv to json format, with new attributes and unit consistency to remove heterogeneous uses of units (m2/hectares, m3/hm3) and to be based on the international system of units. The *in situ* reference time series have also evolved in terms of format, so it was necessary to adapt the internal management of the data tables and to homogenize the units to the international system.

A new option has been integrated: the possibility to select a field name as identifier. Several id fields are now present in the delivered databases (GranD id, SWOT id).

The metrics comparison module has evolved to integrate, in addition to the name of the dams, their identifier in the evaluation json files.

A new additional indicator has been integrated into the chain. It is the calculation of the water filling rate TR(t) of the dam, accompanied by statistics and visualization graphs. This rate indicates the estimated volume of water in relation to the maximum capacity of the dam observed over the period treated, applying the 98th quantile to eliminate outliers. In addition, a control of the extreme values of this index appears as a warning in the log.

The Surfwater_Postprocess chain has a module of metrics comparison (surface and volume), allowing to evaluate the time series of one or several water bodies compared to the reference time series, generating graphs and associated statistics. A new functionality has been developed, offering the possibility to compute reference time series TR(t) from reference time series V(t), if present, in order to qualitatively evaluate the water filling rate of dam,s as well as other quantities.

The parameterization of the resources linked to the PBS scheduler had to be readjusted several times according to the non-standard characteristics of some dams.

4.1.4 Stockwater Workflow

In order to ease the processing of production campaigns, a Stockwater orchestrator has been developed to carry out the management and the sequence of the treatments intended for the production on a site, for a set of tiles, over a given period of one or several years, for one or several DEMs.

The orchestrator allows the automatic setting up of a context, for a super site, in terms of parameterization, tree structure and data formatting for each of the following processed steps

- ✓ the generation of water masks from satellite observations allowing the collection of continental water surface series,
- ✓ the modeling of the HSV relationships for each of the DTMs selected in the study, followed by the comparison with the *in situ* reference models available for the studied structures
- ✓ generation of time series of surface S(t), volume V(t) and filling rate TR(t) for each reservoir based on a final selected DTM, followed by comparison of the metrics with the reference *in situ* time series.

The workflow between the Surfwater, DEM4Water and Surfwater_Postprocess chains is shown below:

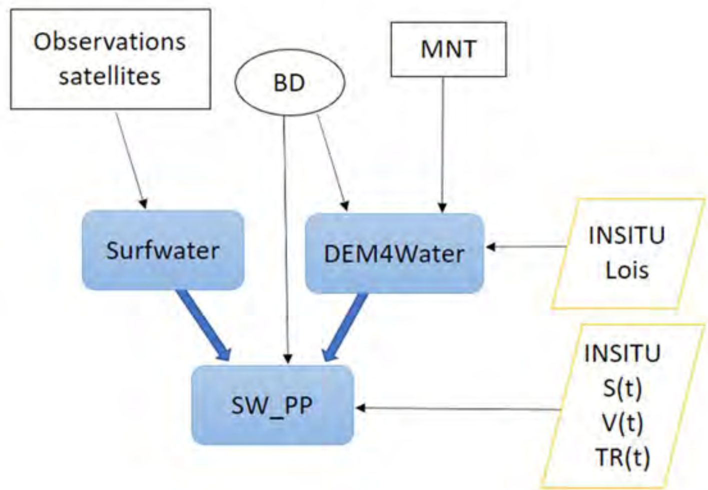


Figure 1 : Stockwater Workflow

An analysis of the results by an operator is strongly recommended between each step of chain launching. The resumption of processing between steps is simplified by the use of true/false keywords related to the activation of chains.

In a first step, the campaign tree for a site is automatically set up by the orchestrator.

1. The orchestrator also manages the successive launches of the Surfwater chain over several years. The operator will have to intervene on the one hand in terms of control of the results of production of the water masks for each tile (exhaustiveness of the masks, taking into account the differences of volume by site, possibly by year), and on the other hand in the final selection of the validated tiles.

2. The next step of creating the estimated models with the DEM4water chain requires, before its launching, some prerequisites which are realized by the orchestrator. The Surfwater products of all the years are gathered and a calculation of the occurrences is carried out on the whole period, then a vrt water map file is generated. The DEM4water chain is then launched to produce the estimated models and comparisons to *in situ* references -if available-. The orchestrator manages the possibility to process one or more DEMs. The estimated models produced are listed and concatenated into a single json file required for the next step. The operator will have to perform a qualitative control of the estimated models because of the variety of local topographies of the processed geographical areas and the shapes of the water reservoirs, which can met the methodological limitations of the chain and cause anomalies.

3. The time series S(t), V(t), TR(t) are generated by the Surfwater_Postprocess chain whose input parameter file is automatically produced by the orchestrator. The processing is performed for an estimated model chosen (via the parameter file) and compared to reference time series -if available-.

For example, for a site, the production of estimated models by the DEM4water chain, on a context of already produced water mask data, requires the following activation in the orchestrator parameter file:

- "run_surfwater" : false

- "run_dem4water" : true
- "run_sw_pp" : false

Launching the orchestrator

After updating the various fields required by the stockwater_params.json parameter file, the orchestrator is launched via the following command:qsub -v WD=$PWD,CONF=stockwater_params.json launch_stockwater.pbs

The environments are automatically installed successively by the pbs scripts according to the steps being processed.

Demonstration

The demonstration of the operation of the orchestrator was provided via a description of the launching procedure on a test set of the Occitanie site and its expected results.

4.2 Implementation of the database

4.2.1 Test sites detection

During the preliminary phase, a first sample of reservoirs was established, with sites selected to be as heterogeneous and representative as possible in terms of surface area, topographic configuration, etc.... and for which the input data were already available (availability of several DTMs, validation parameters, etc.).

A reference database describing these sites was set up during Phase I and adapted to the needs of the project as Phase II progressed.

As part of the planned complementary activities of this addendum, this initial list will be supplemented with sites selected and incorporated into the database during Phase II.

The selection of sites of interest is based on the analysis of the lake databases described below. The sites are considered compatible when the results are relevant (rights-of-way, metadata validity...), when the sites are easily exploitable (preferably on a single S1/S2 right-of-way), but also when exploitable *in situ* validation data are available.

The databases that were explored are the following:

- ✓ SWOT Lakes DB, made from Circa 2015 by Y. Sheng,
- ✓ GOOD2,
- ✓ GRanD,
- ✓ The GEODAR database, which integrates the GRanD and GOOD databases and complements them,
- ✓ The official Indian DB, from the Water Central Commission.

TABLE 1 : SYNTHESIS OF AVAILABLE DATABASES

DB	Geometry	attributes	NB Burkina	NB Inde	NB Laos	NB Andalousie
GOODD	Point	--	125	6785	25	98
GOODD unsnapped	Point	--	111	6210	8	93
GRAND	Point + Extent*	- Name - storage capacity - Height of dam - Alternative height of dam (update or secondary) - Depth - Altitude - Quality - ...	58	333	9	60
GeoDAR	Point + Extent**	- Storage capacity (only for GRanD features in input)	58	932	12	163

Specifically, the final selection criteria are as follows:

- ✓ The availability of *in situ* data, among others the H of the dam foot and the maximum filling volume on the reservoir, and their relevance to reality (e.g. mismatch between the dam height and the maximum filling level...),
- ✓ The availability of DTM or DSM data (Copernicus, CARTOSAT, Alos, SRTM, Stereo Pleiades...),
- ✓ The surface variability marked on the Global Surface Water (GSW, Pekel et al 2016) occurrence map,
- ✓ The good fit between the proposed contour of the Reservoir DBs and the GSW occurrence,

Where possible, a minimum size of 10 ha for Phase 1 (5 ha to be studied for Phase 2)

4.2.2 Phase I datasets built-up

The initial campaign, intended to constitute the datasets used for Phase I,1 focused on Andalusia, Occitania and India. Wherever possible, we selected the most relevant sites present in the GRanD database, which happens to be the most complete.

Reservoirs never fully covered by a tile were discarded as much as possible, reservoirs fully covered by several tiles were systematically selected in order to double the frequency of observation.

The following global methodology was adopted for the constitution of the dataset, with some adaptations according to the complexity and the unforeseen events encountered on the different areas:

After analysis of the different geo-referenced databases available, it turned out that the maximum extensions from GRanD were globally of poor quality with a tendency to underestimate these maximum extensions (see Figure 2).

We therefore chose to extract the maximum extensions of each reservoir from GSW 2020, which proved to be more accurate and more representative of the reality on the ground, as it takes into account the dynamics of the reservoirs over several years. A manual refinement of the maximum extensions from GSW was also carried out. (see Figure 3).

The attributes from GRanD were then associated with each of these extensions, when they were present in this database. The reference data (water level, maximum storage, ...) missing in GRanD were manually searched on Google Earth and other external sites.

For each of the reservoirs, the water location points and dam locations were all validated/corrected. Dam elevations were also corrected from Google Earth and the maximum elevations of available in-situ data.

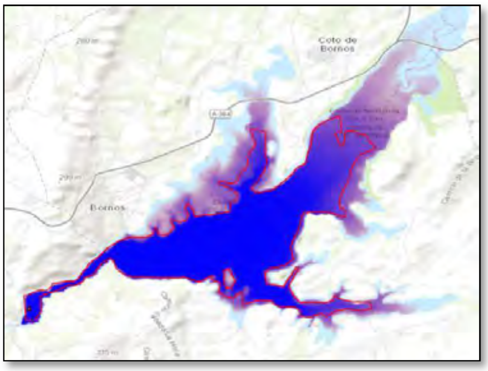


Figure 2 : Bornos / Guadalete River, Red outline for GRanD /GeoDAR extent, GSW occurrence in blue (1984-2020)

Figure 3: José Torán Reservoir in Sevilla, yellow outline for GRanD /GeoDAR extent, GSW occurrence in red (1984-2020)

For Phase I, the initial target was 50 sites. Finally, the stabilized and representative batch delivered includes 75 sites of which :

- ✓ 37 in Andalusia,
- ✓ 16 in Occitania,
- ✓ 22 in India.

4.2.3 Andalusia campaign

37 reservoirs, considered as the most promising based on the selection criteria, were selected for the Andalusia campaign.

20 reservoirs, of the 37 selected, have *in situ* data (regular measurements of water height and volume).

Figure 4 : Andalusia – data chart

4.2.4 Occitania campaign

For the Occitania campaign, 16 reservoirs present in GRanD were selected for which *in situ* data are available.

Figure 5 : Occitania – data chart

4.2.5 India campaign

For the India campaign, it was decided, in the first instance, to give priority to the areas of the scenes available on THEIA. (Based on this criterion and on the reservoirs present in GRanD, and/or having *in situ* data, 22 reservoirs were selected over India.

The selection of sites in India appeared to be the most complex, as a large part of the reservoirs present in the *in situ* data provided were not present in the georeferenced databases at our disposal (GRanD, GEODAR, ...). It was therefore necessary to manually search for the location of reservoirs present in the area covered by the THEIA data and listed in the *in situ* data files.

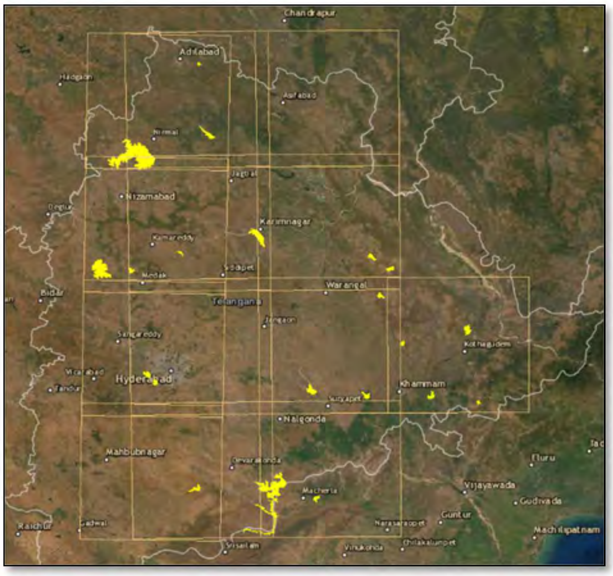


Figure 6 : India – data chart

Finally, only 8 of the 22 selected reservoirs have *in situ* data. Seven of the reservoirs with *in situ* data were present in the GRanD database and only one in GEODAR. Unfortunately, it was not possible to locate additional reservoirs that met the criteria (*in situ* data available and present in the area covered by THEIA).

TABLE 2 : INVENTORY OF AVAILABLE ***in situ*** DATA ON THE RESERVOIRS IDENTIFIED FOR PHASE I

Reservoir Name	DISTRICT	BASIN	DATABASE
Almatti Reservoir	BIJAPUR	KRISHNA	Hors emprises
Narayanapura Reservoir	BIJAPUR	KRISHNA	Hors emprises
Gobind SagarBhakra Reservoir	BILASPUR	INDUS (UP TO BORDER)	Hors emprises
BUGGAVAGU	GUNTUR	KRISHNA	GRanD
NAGARJUNA SAGAR	GUNTUR	KRISHNA	GRanD
PRAKASAM BARRAGE	GUNTUR	KRISHNA	Hors emprises
Srisailam Reservoir	GUNTUR	KRISHNA	Hors emprises
Srisailam Reservoir	GUNTUR	KRISHNA	Hors emprises
Lower ManairG.V.Sudhakar Rao Reservoir	KARIMNAGAR	GODAVARI	GRanD
MUNIYERU PROJECT	KRISHNA	KRISHNA	Hors emprises
TAMMILERU RESERVOIR	KRISHNA	EAST FLOWING RIVERS BETWEEN MAHANADI AND PENNAR	Hors emprises
VIJAYARAI ANICUT	KRISHNA	EAST FLOWING RIVERS BETWEEN MAHANADI AND PENNAR	Hors emprises
PULICHINTHALA PROJECT	NALGONDA	KRISHNA	GEODAR
Nizam sagar	NIZAMABAD	GODAVARI	GRanD
Sriram Sagar	NIZAMABAD	GODAVARI	GRanD
VENGALRAYA SAGARAM	VIZIANAGARAM	EAST FLOWING RIVERS BETWEEN MAHANADI AND PENNAR	Hors emprises
GONELAVAGU RESERVOIR	WEST GODAVARI	EAST FLOWING RIVERS BETWEEN MAHANADI AND PENNAR	Hors emprises
JALLERU RESERVOIR	WEST GODAVARI	EAST FLOWING RIVERS BETWEEN MAHANADI AND PENNAR	Hors emprises
KOVVADAKALVA RESERVOIR	WEST GODAVARI	GODAVARI	Hors emprises
YERRAKALVA RESERVOIR	WEST GODAVARI	EAST FLOWING RIVERS BETWEEN MAHANADI AND PENNAR	Hors emprises

4.2.6 Data preparation

4.2.6.1 Structure of the attributes file

This activity focused on the finalization of the database format, with the evaluation of 6 successive versions, to finally obtain the following structure during Phase II.

TABLE 3 : STRUCTURE OF DATASET ATTRIBUTES

Colum name	Type	Description
ID_SWOT	TEXT	SWOT ID
ID_DB	TEXT	Database ID
NAME_DB	TEXT	Database name
RES_NAME	TEXT	Reservoir name
DAM_NAME	TEXT	Dam name
NEAR_CITY	TEXT	Name of the nearest city
COUNTRY	TEXT	Name of the country
YEAR	INTEGER	Year when the dam was built
DAM_LVL_M	FLOAT	Elevation of reservoir surface in meters above sea level; value derived from EarthEnv- DEM90 data set (Robinson et al. 2014) at 15s resolution at point location of dam
CAP_MCM	FLOAT	Representative maximum storage capacity of reservoir in million cubic meters
DEPTH_M	FLOAT	Average depth of reservoir in meters; calculated as ration between storage capacity and surface area.
MAIN_USE	TEXT	Main use of the reservoir
LONG_DD	FLOAT	Longitude of the dam reservoir
LAT_DD	FLOAT	Latitude of the dam reservoir
LONG_WW	FLOAT	Longitude of a point in the reservoir water
LAT_WW	FLOAT	Latitude of a point in the reservoir water
POLY_SRC	TEXT	Reshaped polygon source
AREA_HA	TEXT	Area of the polygon in Ha

The NO DATA are filled in as -999 for numeric fields and NULL for text fields.

It should be noted that more than six versions were produced before agreeing on this structure, which is now considered final.

The different datasets on Andalusia, Occitania and India were then produced in shapefile, geojson and CSV formats.

4.2.6.2 *In situ* reference data

4.2.6.2.1 WATER LEVELS AND *in situ* VOLUMES

In order to validate the results obtained via STOCKWATER, a number of *in situ* data were provided by different organizations. These data have been systematically analyzed and significant irregularities have been identified.

- ✓ Occitania: Data provided by the DREAL Occitania.
- ✓ Andalusia: Data provided by the Confederacion Hidrografica del Guadalquivir.
- ✓ India: Data provided by ISRO.

An important activity of analysis, selection and correction (according to different methods) was carried out until late in Phase II, to generate operational datasets, compatible with the operation of the PoC STOCKWATER.

Beyond the actions carried out during Phase II, the work necessary for the Valuation Phase showed that anomalies (or at least causes of malfunctioning) remain and affect in particular the functioning of the dashboard. A new action of systematic verification and correction of the database will be carried out within the framework of the planned complementary activities.

However, after analysis, selection (sites with usable *in situ* data) and correction, it was possible to generate a dataset sufficient for the realization of Phase I of the project:

DAM_NAME	Z_IN_SITU	V_IN_SITU
Pareloup	X	X
Saint Geraud	X	X
Avene	X	X
Salagou	X	X
Saints Peyres	X	X
Cammazes	X	X
Astarac	X	X
La Gimone	X	X
Montbel	X	X
Agly	X	X
Laparan	X	X
Pla de Soulcem	X	X
Vinca	X	X
Puyvalador	X	X
Villeneuve la Raho	X	X
Matemale	X	X

TABLE 4 : ***in situ*** DATA - OCCITANIA

In situ data (water height, volume) are available for all the 16 selected sites. These data come from different managers. They were aggregated in the same format by the MTE and sent through the DREAL service.

DAM_NAME	Z_IN_SITU	V_IN_SITU
Cuevas de Almanzora	X	X
Beznar	X	X
Rules	X	X
Guadalhorce	X	X
Guadalteba	X	X
Conde de Guadalhorce	X	X
La Vinuela	X	X
Limonero	X	X
Guadarranque	X	X
Charco Redondo	X	X
La Concepcion	X	X
San Rafael de Navallana	X	X
Los Hurones	X	X
Almodovar	X	X
Zahara	X	X
Bornos	X	X
Barbate	X	X
Arcos de la Frontera	X	X
Guadalcacin 2	X	X
Celemin	X	X
Tranco de Beas	X	X
Arenoso	X	X
Yeguas	X	X
Los Bermejales	X	X
Giribaile	X	X
Jose Toran	X	X
Vadomojon	X	X
Iznajar	X	X
Puebla de Cazalla	X	X
La Brena II	X	X
Los Melonares	X	X
Puente Nuevo	X	X

TABLE 5 :DONNÉES ***in situ*** ANDALOUSIE

In situ data (water height, volume) are available for the 33 selected reservoirs. These data are from two different sources: SAIH Guadalquivir and SAIH Hidrosur.

TABLE 6 : *in situ* DATA - INDIA

DAM_NAME	DB_1			DB_2			DB_3			DB_4		
	Z_IN_SITU	V_IN_SITU	S_IN_SITU	Z_IN_SITU	V_IN_SITU	S_IN_SITU	Z_IN_SITU	V_IN_SITU	S_IN_SITU	Z_IN_SITU	V_IN_SITU	S_IN_SITU
Buggavagu	X	X								X	X	
Dindi					X	X						
Himayat Sagar					X	X				X	X	
Kaddam					X	X						
Kinnersani					X	X						
Lakhnavaram					X	X						
Lankasagar					X	X						
Large Tank Bayya					X	X						
Lower Manair	X	X			X	X	X	X				
Musi					X	X						
Nagarjuna Sagar	X	X			X	X	X	X				
Nizam Sagar	X	X			X	X	X	X				
Osman Sagar					X	X				X	X	
Pakhal					X	X						
Palair					X	X						
Pocharam					X	X						
Pulichinthala Project	X	X									X	
Ramappa					X	X						
Sathnala					X	X						
Sriram Sagar	X	X			X	X	X	X				
Upper Manair					X	X						
Wyra					X	X						

Concerning the India sites, 4 *in situ* data submissions from different sources were made by Sylvain FERRAND and his PhD student Abhilash Kumar Paswan.

- ✓ DB1, DB3 and DB4 are daily data produced by the Central Water Commission (CWC, federal agency), which are regularly updated with each arrival of data from state or private agencies. These data are derived from field measurements and tank gauging,
- ✓ DB2: product of remote sensing and more precisely of LISS IV and Cartosat-1.

It should be noted that several months passed between the sending of DB1 and DB4 which was provided on 20/10/2021.

4.2.6.2.2 DATA SELECTION

The reference data provided, present significant differences and are not necessarily homogeneous. A comparison analysis was done on the volumes of the 4 tanks common to the delivered batches. This analysis shows considerable differences between the data sets. (see Figure 7, 8, 9, 10).

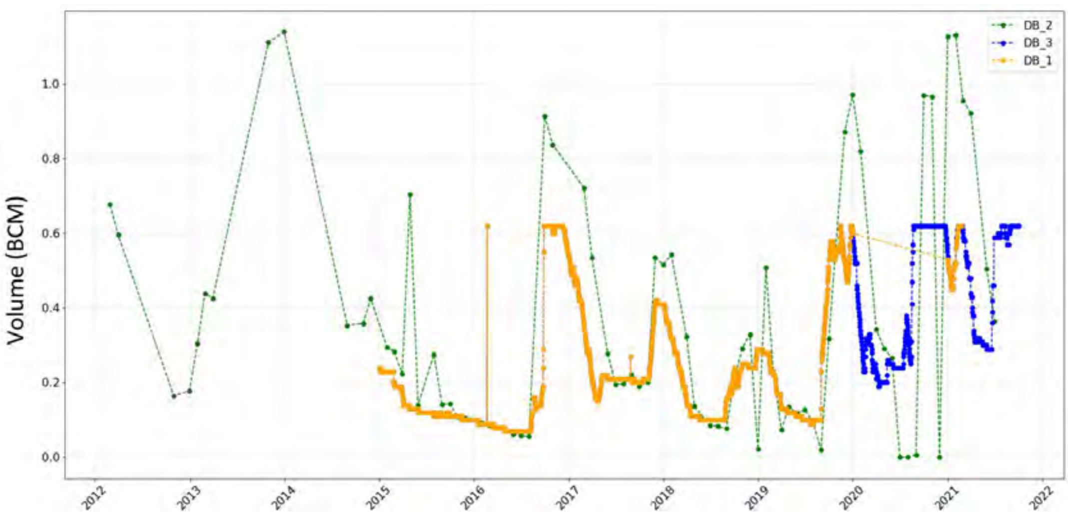


Figure 7 : Volume Comparison - Lower Manair

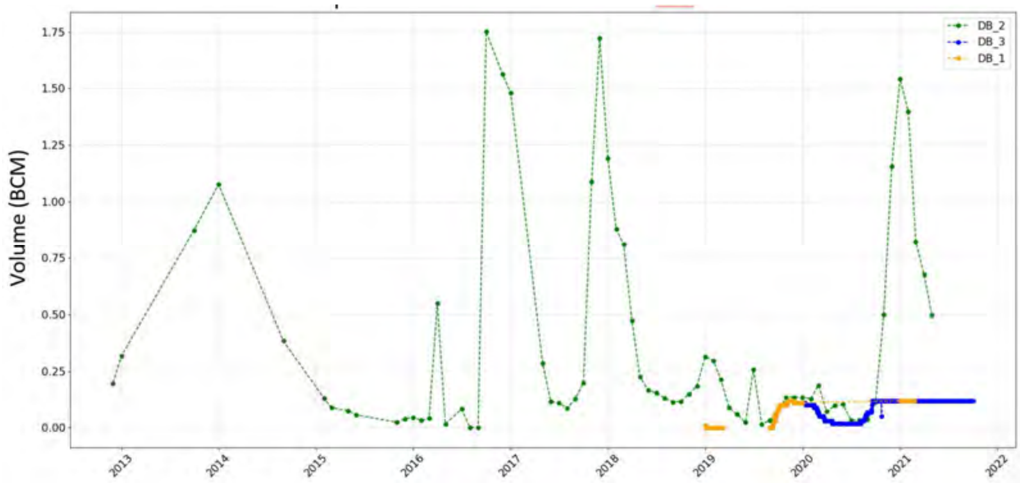


Figure 8 : volume comparison - Nizam Sagar

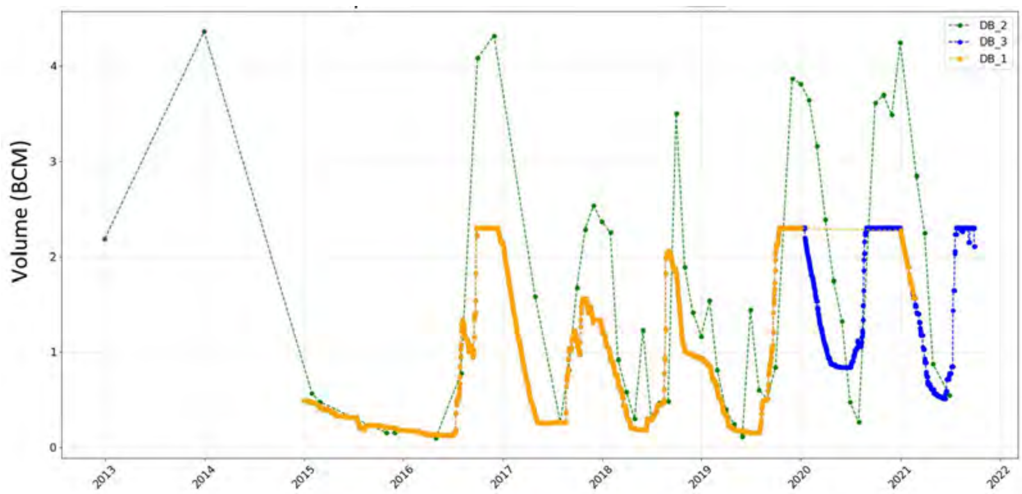


Figure 9: volume comparison- Sriram Sagar

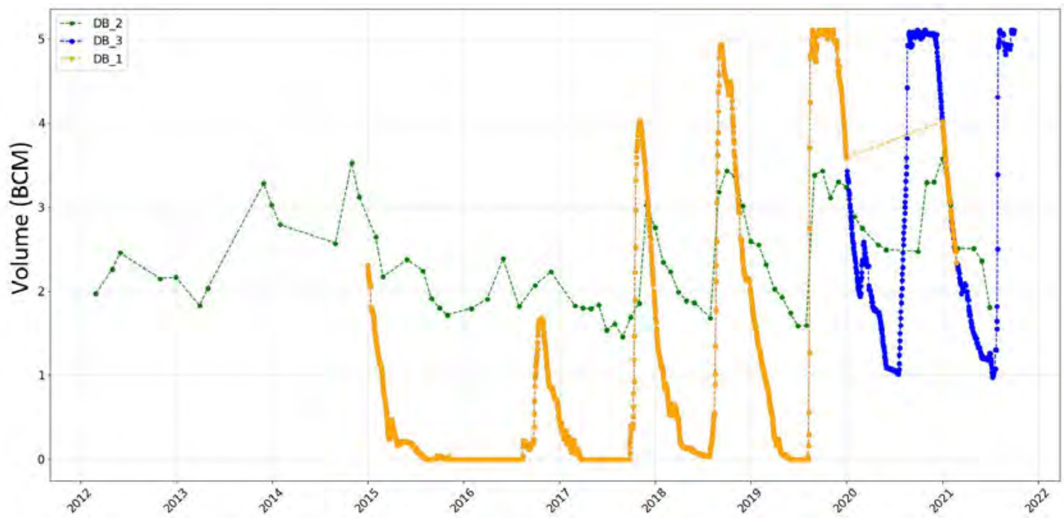


Figure 10 : volume comparison - Nagarjuna Sagar

The following remarks have been made from the previous comparison graphs:

- ✓ Databases 1 and 3 have identical volume values,
- ✓ There are significant differences between databases 1/3 and 2,
- ✓ Greater data density in databases 1 and 3 than in 2,
- ✓ Database 3 is more complete (more values than 1),
- ✓ In view of the values in database 2 and the fact that it is derived from satellite solutions makes it difficult to use as a quality reference.

In the end, taking into account the above remarks, the following *in situ* data were selected for 8 reservoirs. Despite the fact that there are still inconsistencies in some datasets, it was decided to keep these datasets in order to observe the behaviour of the law generation algorithms when faced with this type of poor-quality data, which can be discarded later on.

		Values	
SWOT_ID	DAM_NAME	Z_IN_SITU	V_IN_SITU
4530313823	Buggavagu	DB_4	DB_4
4530118323	Dindi		
4530212143	Himayat Sagar	DB_4	DB_4
4530430393	Kaddam		
4530382953	Kinnersani		
4530421653	Lakhnavaram		
4530387933	Lankasagar		
4530422183	Large Tank Bayya		
4530433353	Lower Manair	DB_3	DB_3
4530422623	Musi		
4530347283	Nagarjuna Sagar	DB_3	DB_3
4530187293	Nizam Sagar	DB_3	DB_3
4530212043	Osman Sagar	DB_4	DB_4
4530404463	Pakhal		
4530417683	Palair		
4530197583	Pocharam		
4530308913	Pulichinthala Project	DB_1	DB_1
4530406983	Ramappa		
4530185803	Sathnala		
4530240913	Sriram Sagar	DB_3	DB_3
4530197543	Upper Manair		
4530419843	Wyra		

TABLE 7 : SYNTHESIS OF SELECTED ***in situ*** DATA

Nota bene **:** Additional information/*in situ* data can be extracted from local databases (Figures 11 and 12), but some of the sites provided are anomalous because seasonal variability is not taken into account, and these data do not fully overlap with the list of sites provided. They have therefore been discarded as they are difficult to use in this study.

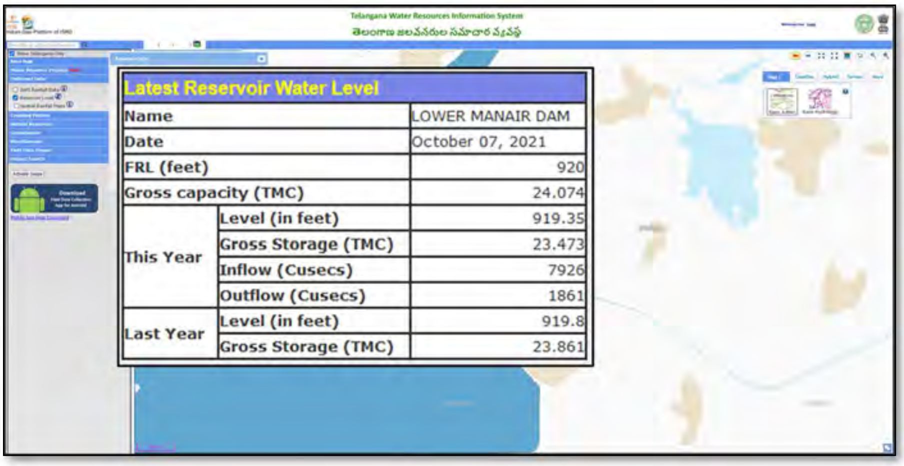


Figure 11: Telangana Water Ressources Information System

4.2.6.2.3 QUALITY ANALYSIS

After selection of the *in situ* data (Water Heights and Volumes) considered as the most relevant, a first generation of the *in situ* India parameterised laws, including also the calculation of the Surfaces derived from the *in situ* Water Heights and Volumes, was carried out with the help of the CNES tool "timeseries_tool.py". This first step allowed us to detect inconsistencies in these reference Water Levels and Volumes data.

Indeed, we noticed that the volumes and heights *in situ* India are globally of poor quality and can be difficult to use. This leads to inconsistencies in the generation of *in situ* surfaces generated by the CNES script, as shown in the following analyses.

The *in situ* volume (volumes) and level (water height) data for reservoirs 4530187293 - Nizam Sagar and 4530308913 - Pulichinthala Project are of very poor quality with a large number of inconsistencies (missing data and staircase data (several level values for a given volume) (see Figure 12).

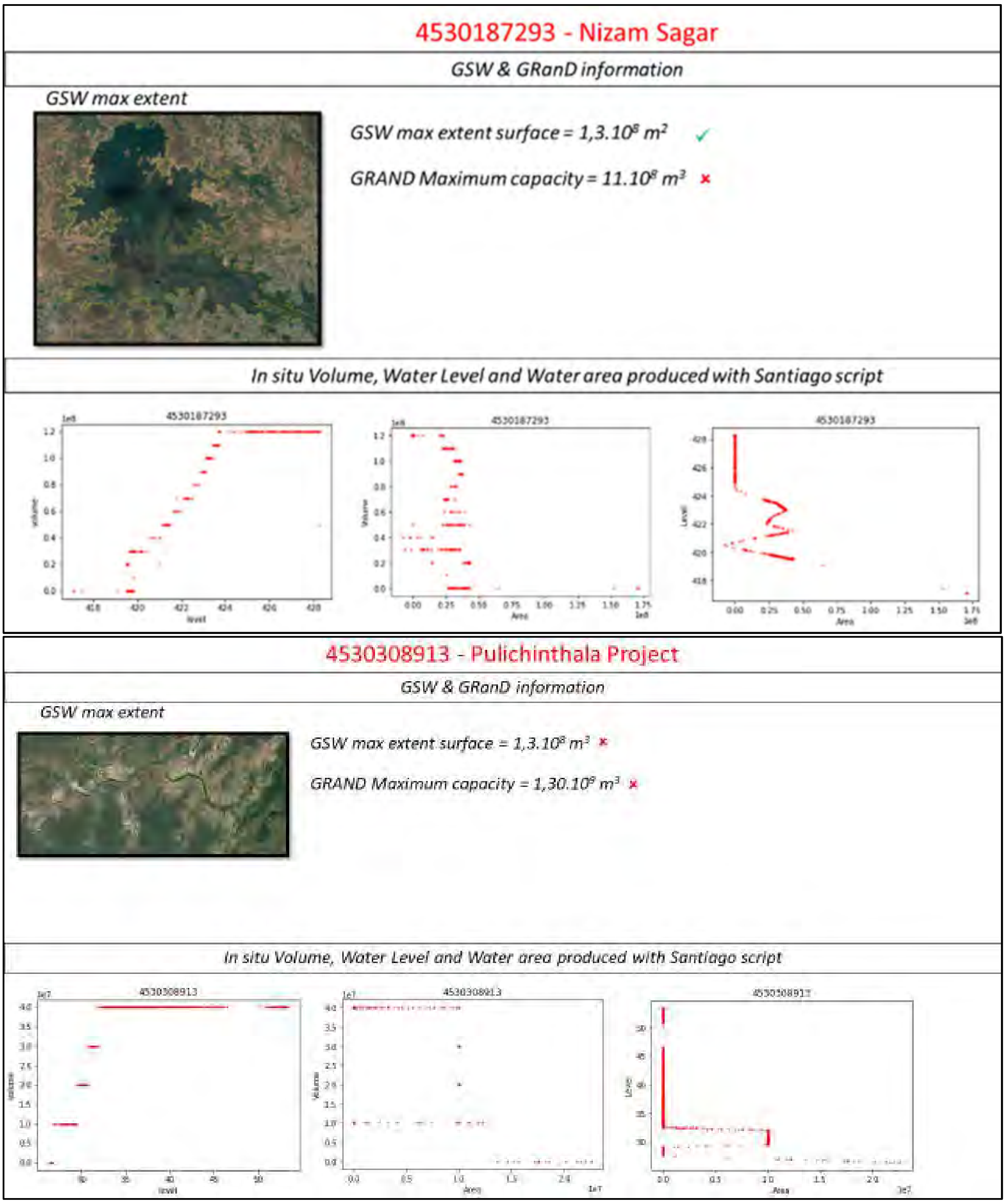


Figure 12 : ***in situ*** data for Nizam Sagar & Pulichintala Project

Les données volume, level et surface *in situ* pour les réservoirs 4530212043 - Osman Sagar et 4530212143 - Himayat Sagar semblent correctes, malgré certains comportements étranges des level et surface lorsque le nombre de mesures/données est moins dense. (*cf.* Figure 13).

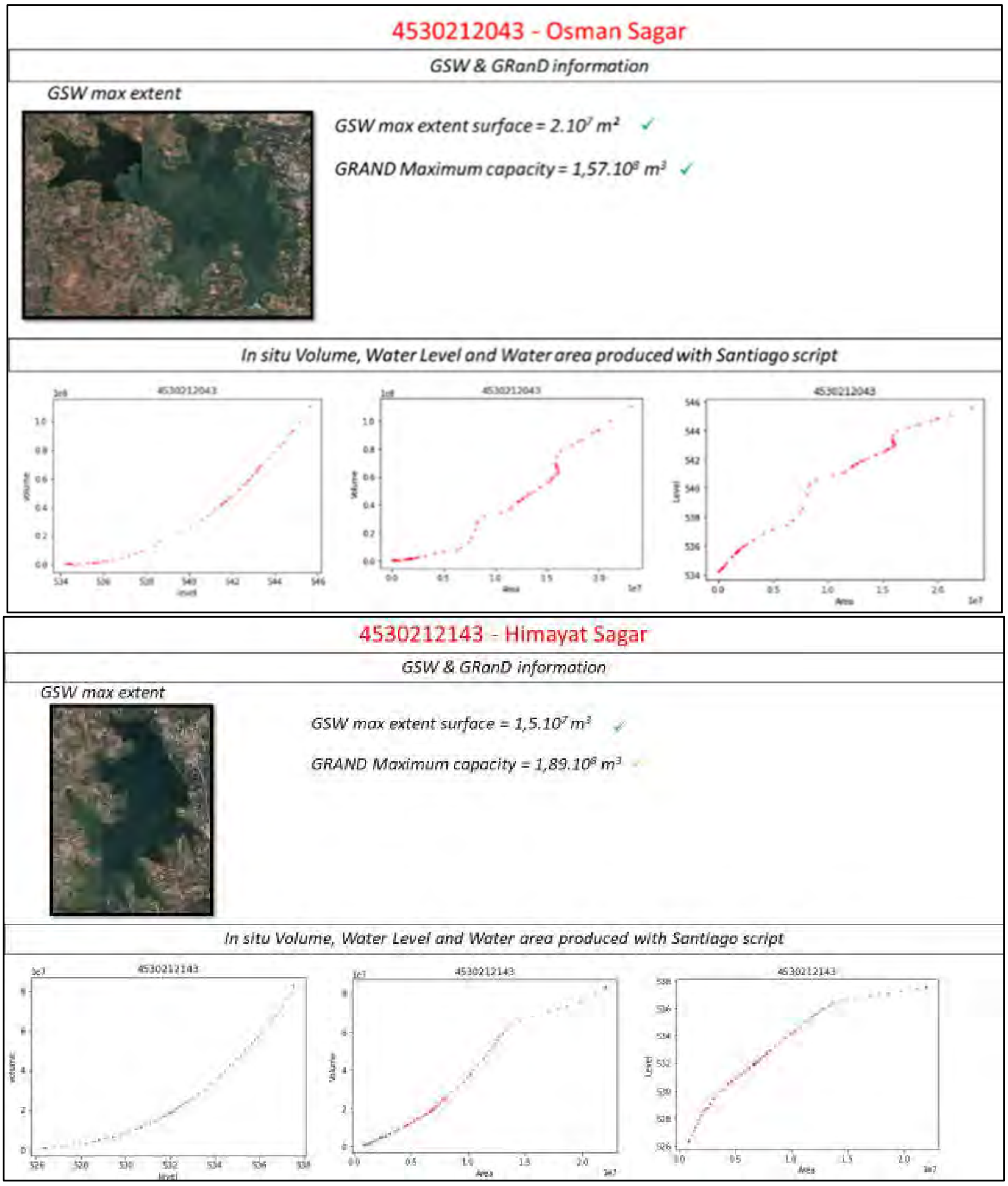


Figure 13 : Données ***in situ*** Osman Sagar & Himayat Sagar

The *in situ* volume, level and surface data for reservoirs 4530212043 - Osman Sagar and 4530212143 - Himayat Sagar appear to be correct, despite some strange behaviour of level and surface when the number of measurements/data is less dense (see Figure 14).

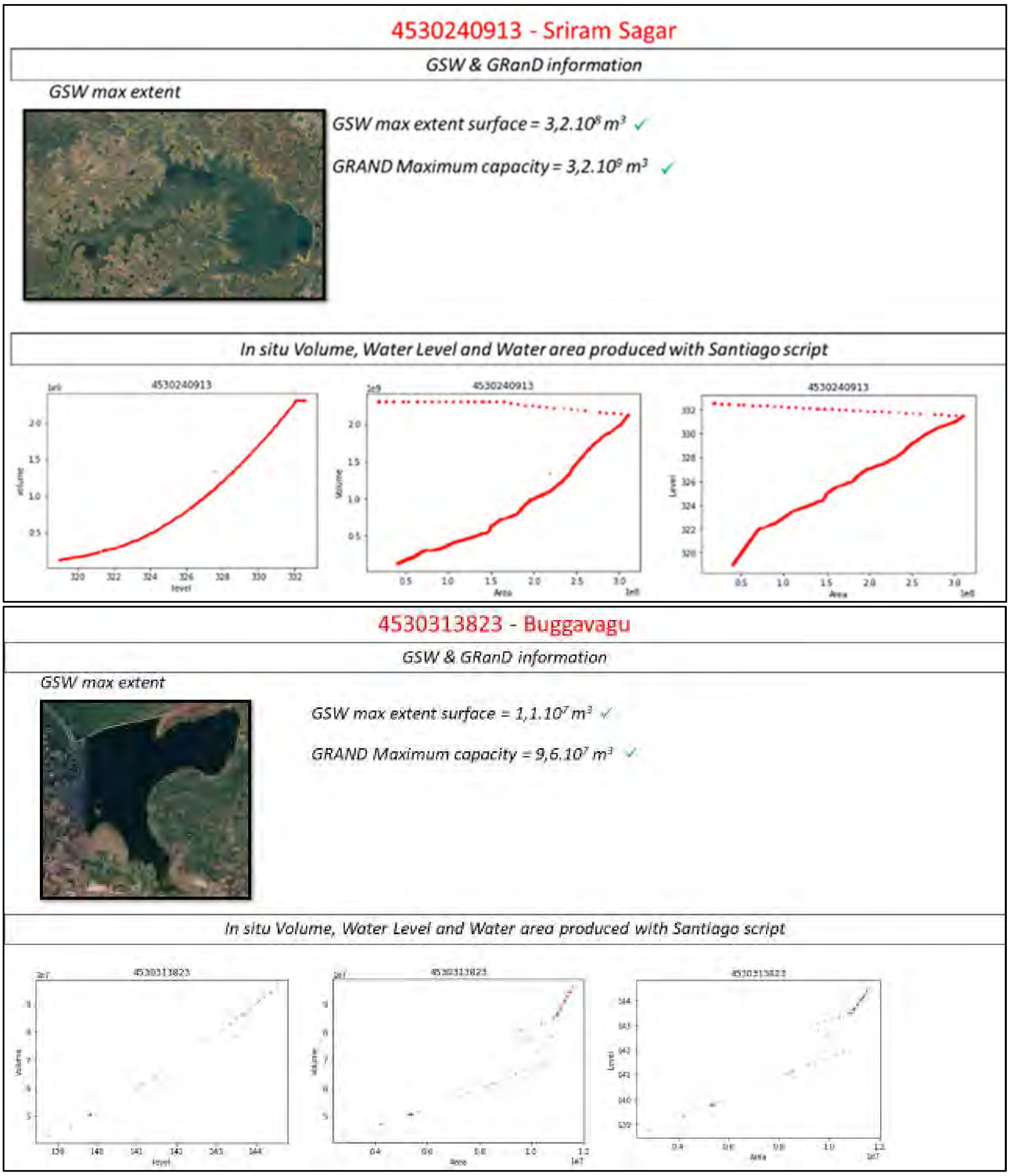


Figure 14 : ***in situ*** data for Sriram Sagar & Buggavu

There are a large number of inconsistencies in the *in situ* volume, level and surface data for reservoirs 4530433353 - Lower Manair and 4530347283 - Nagarjuna Sagar, with "staircase" data, a number of outliers, as well as "trays" (several level values for a given volume) (see Figure 15).

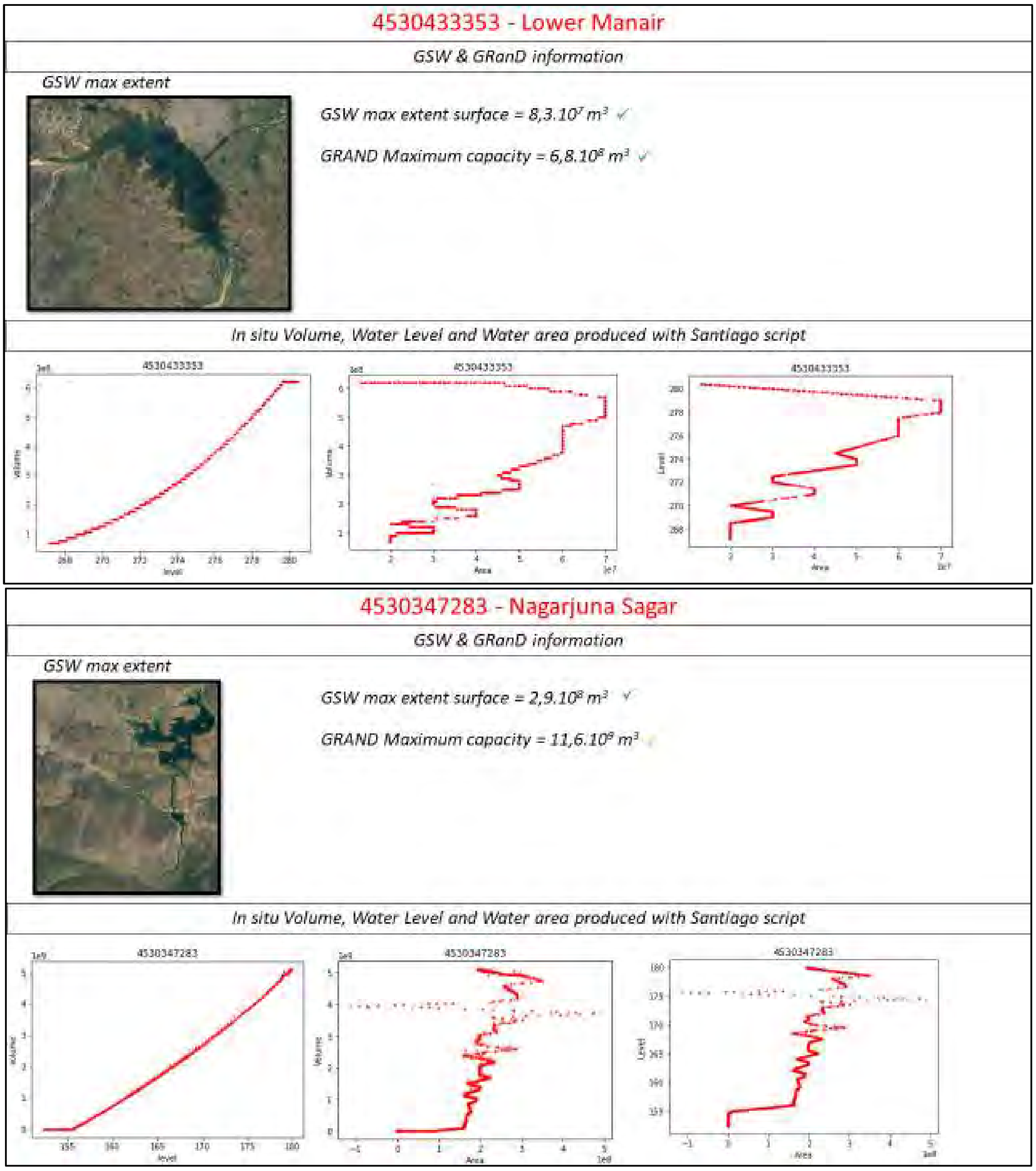


Figure 15 : ***in situ*** data for Lower Manair & Nagarjuna Sagar

4.2.6.2.4 *in situ* DATA IMPROVEMENT

Considering the problems observed on these data described above, a tool was set up to "improve" the quality of the data of some reservoirs and make them usable for the steps of generation of *in situ* laws.

This tool works in 3 steps:

- ✓ Step 1: Removal of "trays" / non-variable data. This means the removal of measurements for which more than X values of level are observed for a given volume or more than X values of volume for a given level (see Figure 16),

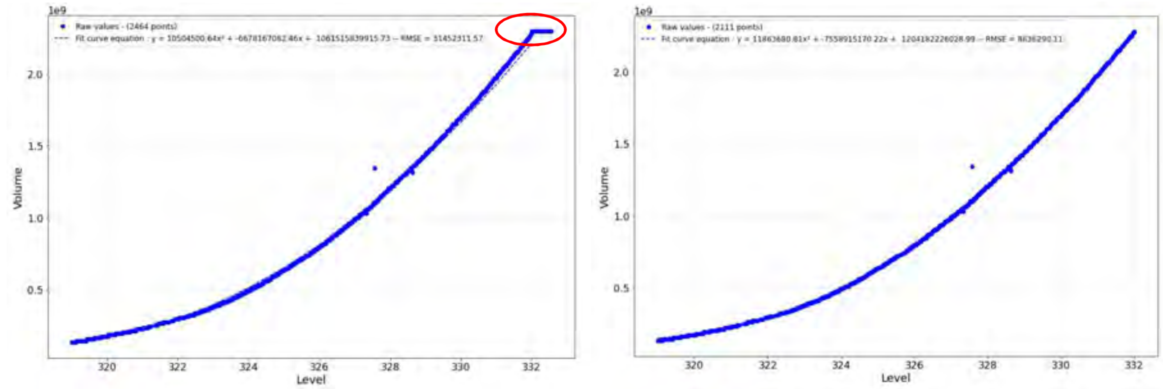


Figure 16 : "trays » removal - Sriram Sagar

✓ Step 2: Removal of outliers (see Figure 17).

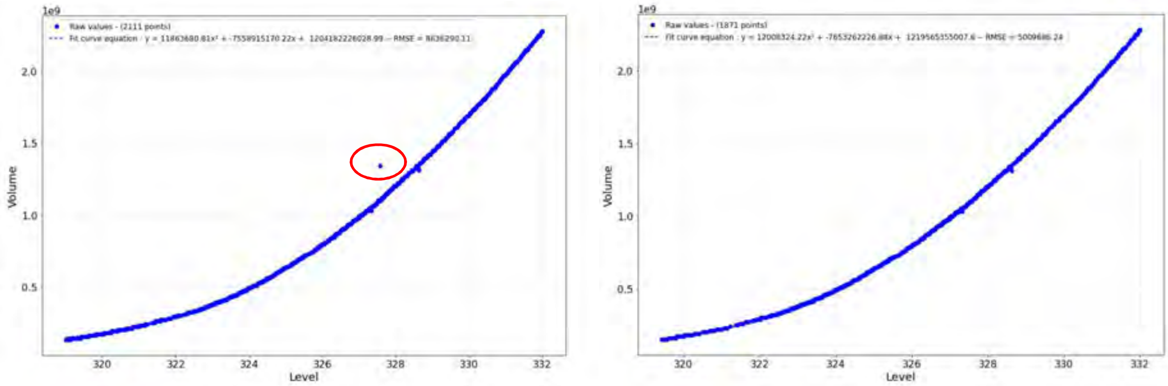


Figure 17 : outliers removal - Sriram Sagar

✓ Step 3: Adjust/Fit the staircase values according to the equation of the V(L) curve (with V = Volume and L = Level) calculated from the data in step 3. This consists in recalculating the measurements for which more than 3 values of level are observed for a given volume or more than 3 values of volume for a given level according to the equation of the V(L) curve (with V = Volume and L = Level). (see Figure 18).

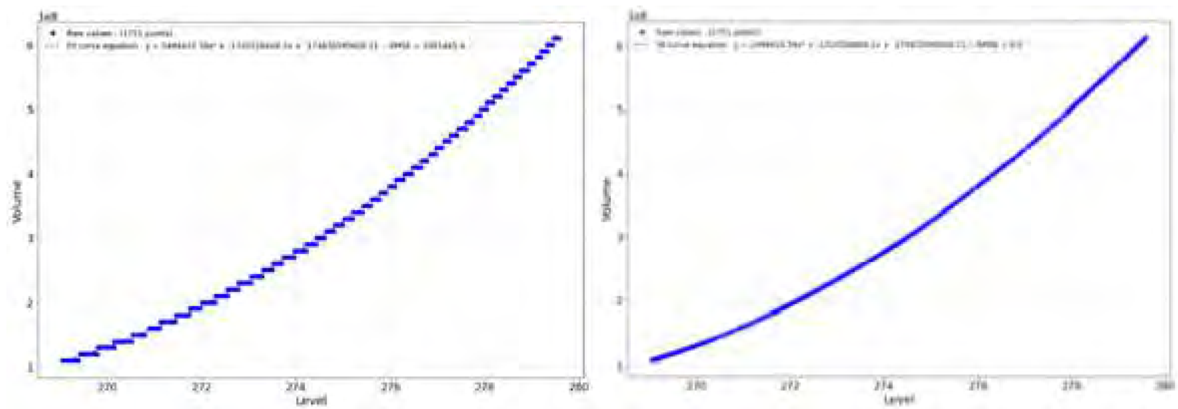


Figure 18 : Adjustment of « staircases » values - 4530433353 - Lower Manair

These 3 steps were applied to the "volume" and "level *in situ* " of the following reservoirs:

- ✓ 4530240913 - Sriram Sagar,

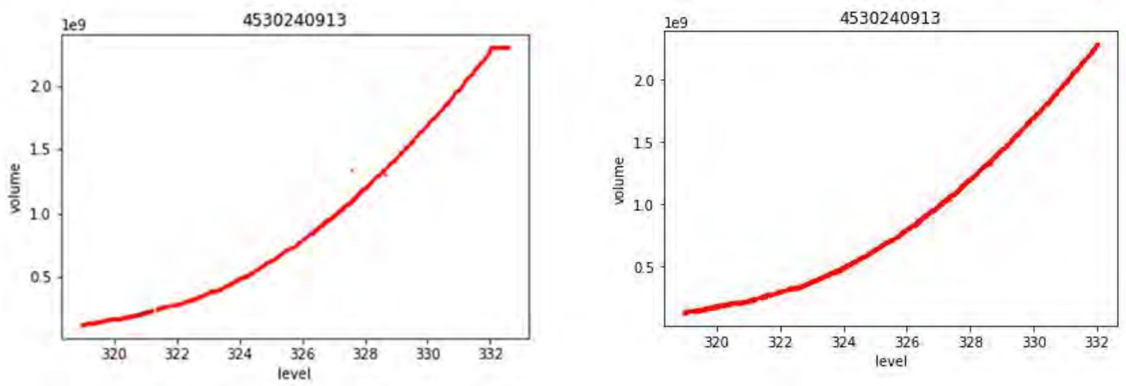


Figure 19 : Raw (left) and improved (right) ***in situ*** data - 4530240913 - Sriram Sagar

- ✓ 4530313823 – Buggavagu

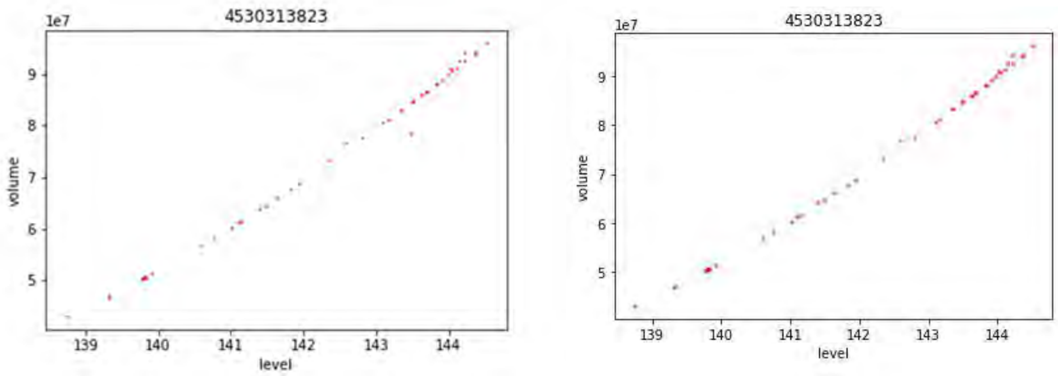


Figure 20 : Raw (left) and improved (right) ***in situ*** data - 4530313823 – Buggavagu

- ✓ 4530347283 - Nagarjuna Sagar

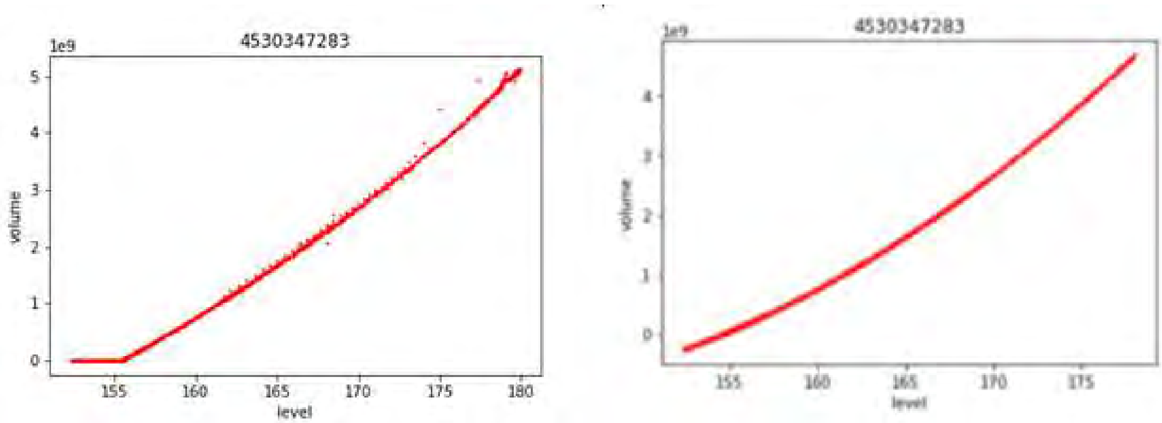


Figure 21 : Raw (left) and improved (right) ***in situ*** data - 4530347283 - Nagarjuna Sagar

- ✓ 4530433353 - Lower Manair

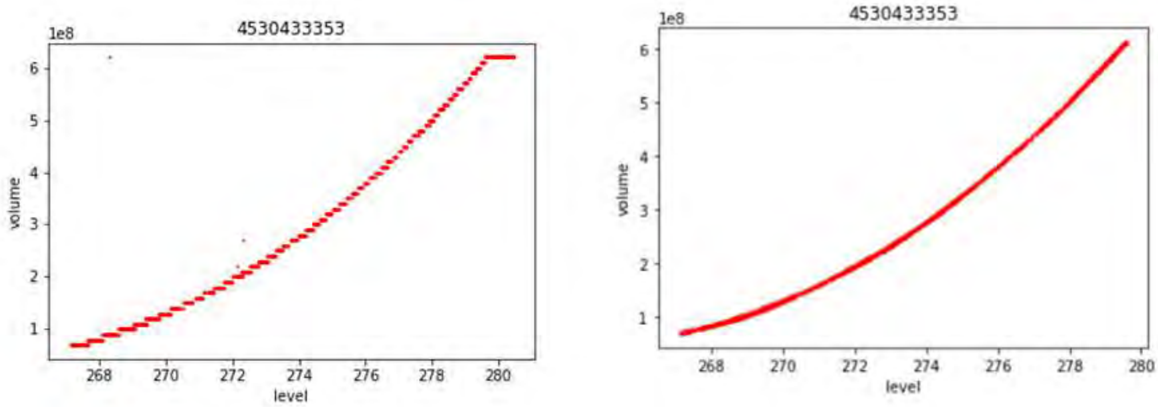


Figure 22 : Raw (left) and improved (right) ***in situ*** data - 4530433353 - Lower Manair

The *in situ* data from reservoirs 4530212043 - Osman Sagar and 4530212143 - Himayat Sagar were not changed because their quality was deemed sufficient.

It was decided to remove the *in situ* data from reservoirs 4530187293 - Nizam Sagar and 4530308913 - Pulichinthala Project from the reference dataset due to their poor quality.

For some datasets, the measurements are strongly modified/impacted, especially by step 3. It is therefore necessary to reflect on whether certain datasets that have been too heavily transformed should be discarded or whether this "transformation" is considered "acceptable". Example for tanks 4530433353 - Lower Manair and 4530347283 - Nagarjuna Sagar.

Nota Bene *:* In order to improve the quality of the India *in situ* data, other smoothing methods (Savitzky-Golay filter and Lowess smoother) were also tested in order to "clean" the in stu data. After comparing the results from these methods with the results obtained via the previously described method, we decided to retain the previously described method which offered better results.

The first smoothing method is based on LOESS (LOcally Estimated Scatterplot Smoothing) local regression.

This is a non-parametric regression method that has the advantage of calculating as many local functions as there are data segments using the k-nearest neighbor method in order to best fit the curves. However, it requires a large amount of data sets to generate good models. Therefore, the V(Z) series correction could not be achieved with this first approach.

The second method consists in applying the Savitzky-Golay filter on the data V(t) and Z(t) independently. This filter approaches the function V(Z), keeping only the important features and getting rid of the meaningless fluctuations. Successive subsets of points are fitted with a polynomial function that minimizes the fitting error.

As for the input parameters, the data to be filtered should be filled in, the window size corresponding to the number of points used to compute the fit should be chosen, as well as the order of the polynomial function.

The following figures show some results varying according to the selected window size indicated by the "Smoothing factor" parameter.

- ✓ 4530212043 - Osman Sagar,

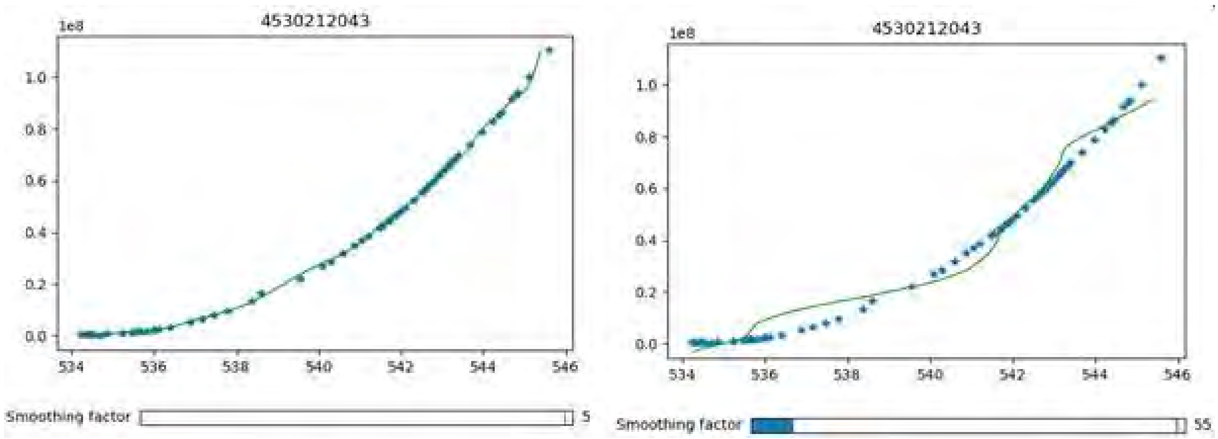


Figure 23 : Raw (blue) and improved (green) ***in situ*** data - 4530212043 - Osman Sagar

- ✓ 4530433353 - Lower Manair,

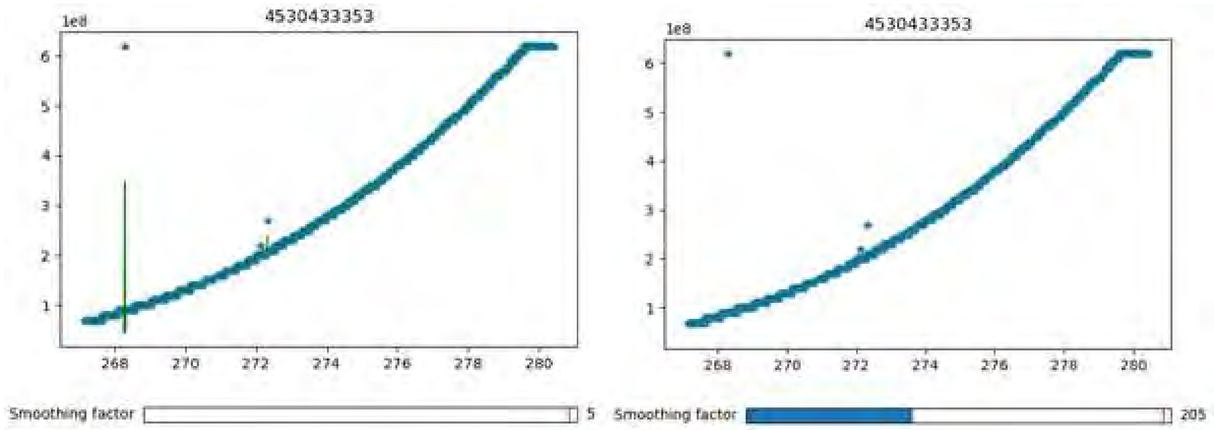


Figure 24 : Raw (blue) and improved (green) ***in situ*** data - 4530433353 - Lower Manair

- ✓ 4530347283 - Nagarjuna Sagar,

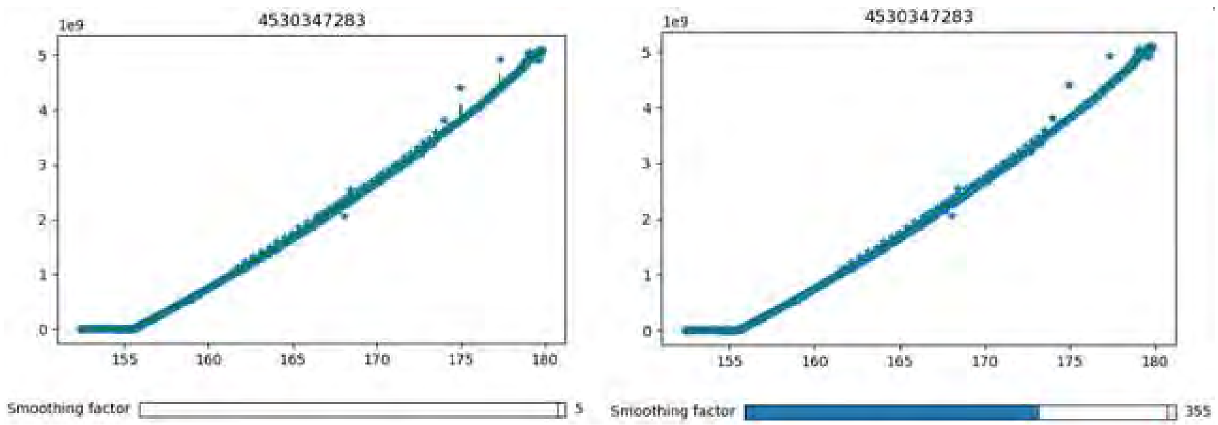


Figure 25 : Raw (blue) and improved (green) ***in situ*** data - 4530347283 - Nagarjuna Sagar

- ✓ 4530187293 - Nizam Sagar

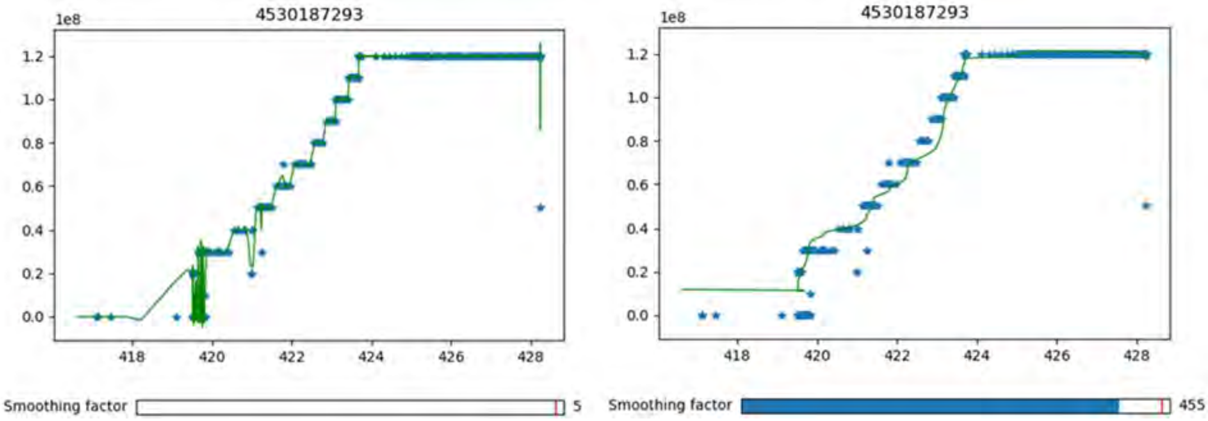


Figure 26 Raw (blue) and improved (green) ***in situ*** data - 4530187293 - NIZAM SAGAR

The Savitzky-Golay filter allows to follow the trend of the original function V(Z) and to eliminate the outliers but does not act completely on the "trays" / non-variable data at the beginning or end of the series. This will always require a pre-cleaning of this part of the data. In addition, the "staircase" effect along the curve (see Figure 27) is generally damped, which prevents a smooth curve. The results can be improved by choosing the most optimal window size.

This method works on all the Indian reservoirs tested, except for reservoirs 4530187293 - Nizam Sagar (see Figure 27) and 4530308913 - Pulichinthala Project for which curve smoothing was inconclusive due to poor quality of the raw data.

4.2.6.3 Generation and comparison of *in situ* laws (see Appendix B)

First, it was necessary to generate the set of *in situ* laws from the *in situ* water volumes and heights data (modified or not) on the 3 sites (Andalusia: 22 reservoirs with *in situ* data, Occitania: 16 reservoirs with *in situ* data and India: 6 reservoirs with *in situ* data). These laws were generated using the CNES "timeseries_tool.py" tool.

It should be noted that several evolutions of the tool were made during phase 1, which led to the regeneration of the *in situ* laws a number of times (more than 3 versions for each site).

In order to evaluate the dem4water laws it is necessary to:

- ✓ Evaluate S(Z), V(S)
- ✓ Determine the sources to be evaluated:
    - Reference: *In situ* relations (no laws),
    - *In situ* laws,
    - Dem4water laws,
    - Modification of *in situ* laws by adding Zmin, Zmax, Smin, Smax or Elevation(dam);

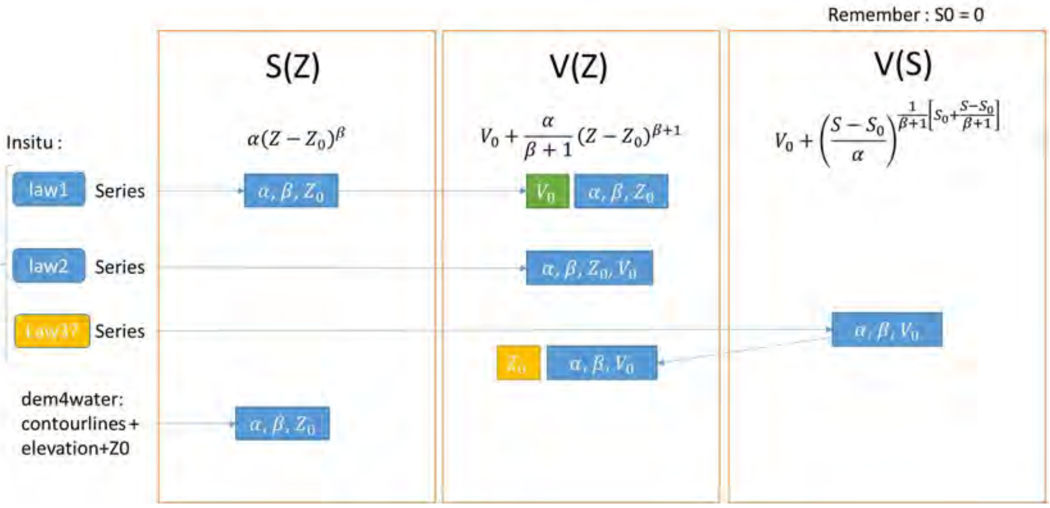


Figure 27 : Relationship between laws and associated parameters

4.2.7 DEM quality analysis

In order to determine the best choice of this DTM to launch the production, a qualification analysis of the DTMs is made with DemCompare on a S2 tile, taking as reference the IGN 5m DTM and that according to the following parameters:

- ✓ Geometric aspect (coregistration),
- ✓ Simple difference,
- ✓ Slope analysis.

The analysis is done on the Occitanie region on the following DTMs:

- ✓ Copernicus 10m,
- ✓ Copernicus 30m,
- ✓ SRTM 30m,
- ✓ Alos Dem 25m.

Nota Bene : SERTIT provided as an illustration this Copernicus 10m DEM on a restricted region of Occitania, knowing that it was only available on Europe, and that this DTM is licensed; CNES has a license but it was not open for Stockwater.

We can see differences in altitude between the DEMs and the reference (see Figure 29). The ALOS DEM has a larger error range, and larger geolocation errors of the decametric or even multi-decametric errors, while on the Copernicus and SRTM DEMs, these errors are multi-metric. In addition, on the ALOS DEM we observe stripping effects, which induce a band depression of the altitude of this DEM.

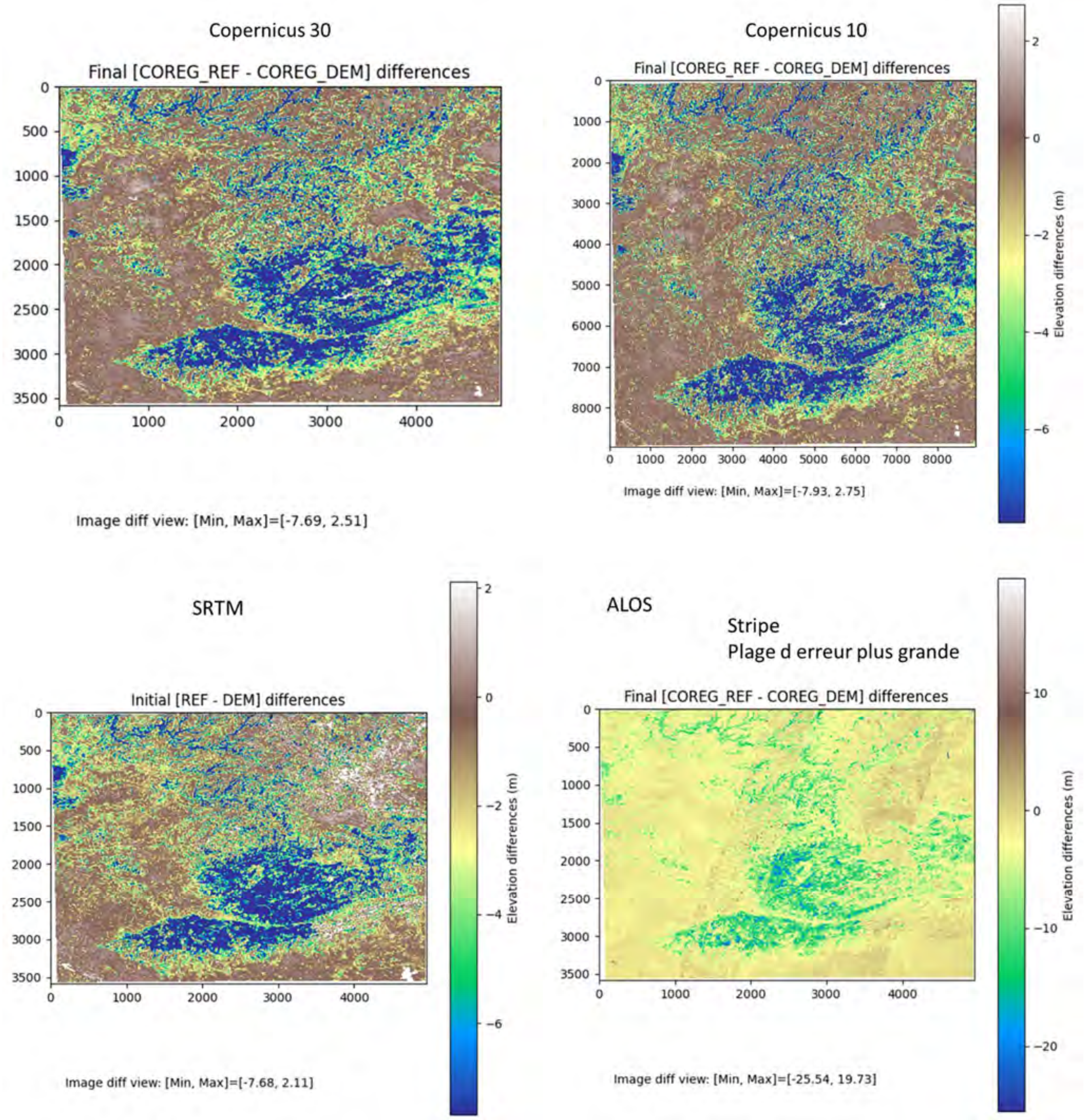
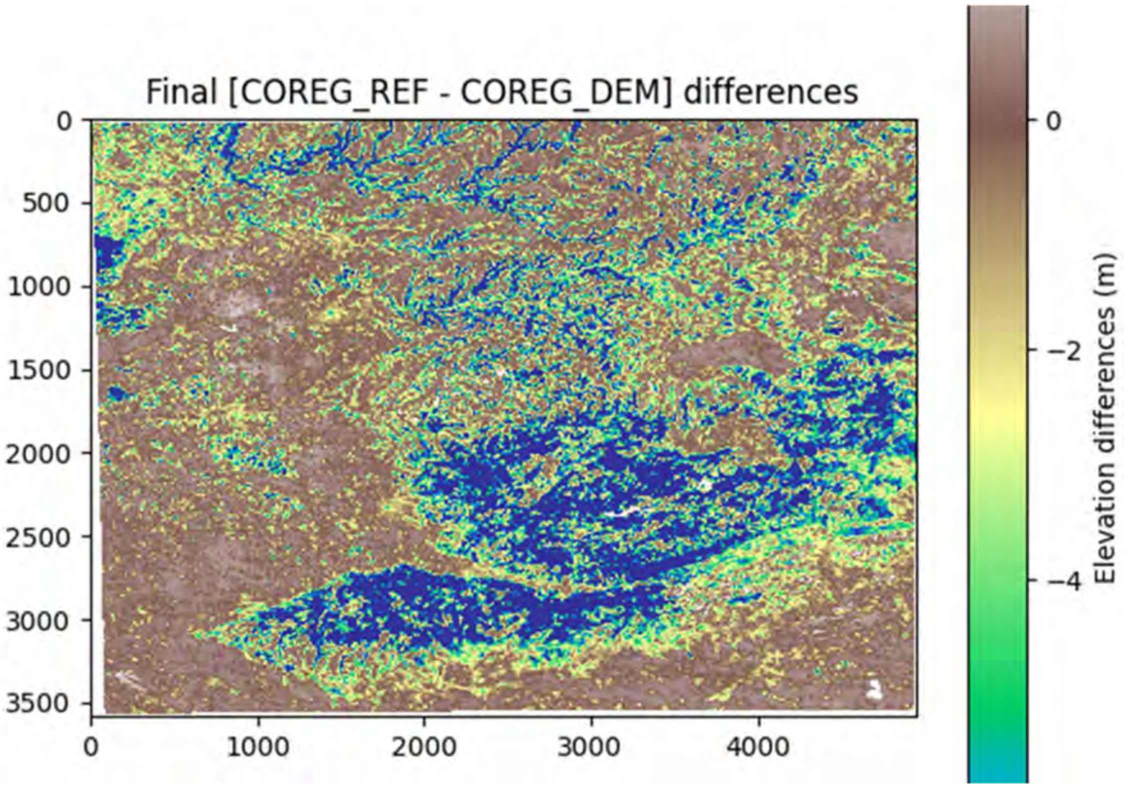
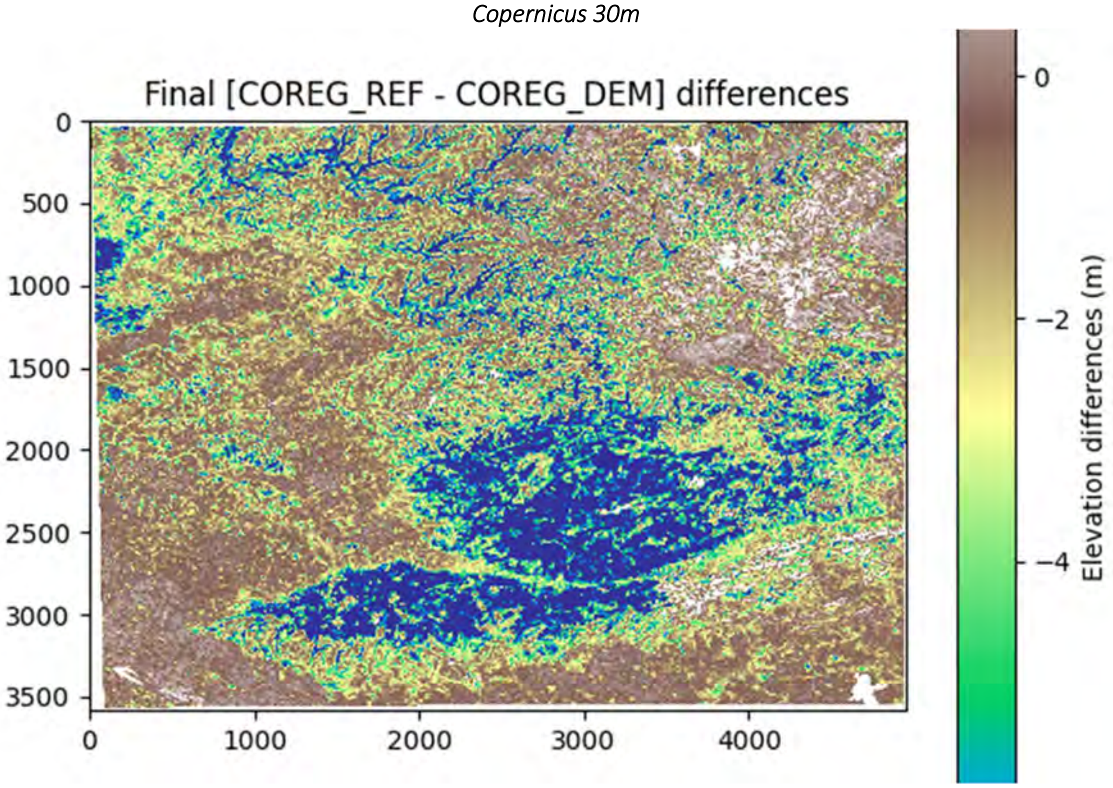


Figure 28: Elevation difference between Copernicus DEM 30 and 10, ALOS DEM, SRTM and IGN 5m DTM reference



Copernicus 30m



SRTM

Figure 29: Difference between DEM SAR 30m, SRTM, Copernicus and IGN reference DTM

The second parameter used to evaluate the quality of the DTMs is the slope distribution, always taking the IGN 5m DTM as a reference (see Table 7).

As could be expected from the results of the altitude errors, the ALOS DTM has the highest Max and Min errors for all slope classes. These errors are 4 to 5 times higher than those obtained for the Copernicus and SRTM DEMs. This is reflected in the mean error which is up to 10 times higher for the ALOS DEM for low slopes than for the Copernicus DEM 30 and 10m. The average SRTM error compared to those obtained for the Copernicus DEMs for the two lowest slope classes, 0-5 and 5-10, are also much larger, but with a much smaller ratio.

TABLE 8 : Slope distribution

COP10M	% Of Valid Points	Max Error	Min Error	Mean Error	Error std	RMSE	Median Error	NMAD	90 percentile
All classes considered	96,43828457	13,41682816	-18,59852791	-2,239506245	4,664140224	**5,173932552**	0,005352224	1,552446643	8,432321548
[0; 5]	19,1103051	13,41676903	-18,59840584	-0,106106669	2,392163277	2,394515038	0,566456258	0,534566256	2,235988283
[5; 10]	18,03569473	13,41523552	-18,59848595	-0,721338034	3,241569281	3,320858717	0,45078361	0,70845231	3,992949009
[10; 25]	33,00777298	13,41602993	-18,59852028	-2,003983498	4,423974037	4,856697559	0,004147011	1,505596485	7,772279835
[25; 45]	19,15678592	13,41484165	-18,59852791	-4,76706171	5,478144169	7,261885166	-3,376380444	5,399457703	9,776192093
[45; inf[	7,127725843	13,41682816	-18,59852409	-6,098513126	5,54611063	8,243249893	-5,5501647	5,974422323	10,2746871

COP30M	% Of Valid Points	Max Error	Min Error	Mean Error	Error std	RMSE	Median Error	NMAD	90 percentile
All classes considered	96,34543086	12,70356464	-17,88702202	-2,254799843	4,45730257	**4,995161533**	-0,126697123	1,536936498	8,039367867
[0; 5]	19,47098568	12,70356464	-17,88685608	-0,133531764	2,262515545	2,266452074	0,555375755	0,451975444	2,059969854
[5; 10]	19,0808532	12,70200729	-17,88663101	-0,774144113	3,127536058	3,221921921	0,416776448	0,658015264	3,85698967
[10; 25]	34,07180121	12,69761944	-17,88698959	-2,183660507	4,300228119	4,822896481	-0,219439864	1,620043743	7,589035988
[25; 45]	18,16111421	12,69782925	-17,88702202	-5,053526402	5,233130932	7,274872303	-3,887000084	5,440121972	9,378242779
[45; inf[	5,560676559	12,70005894	-17,88695717	-6,058505535	5,160805702	7,958605766	-5,483187675	5,471671774	9,558903694

SRTM	% Of Valid Points	Max Error	Min Error	Mean Error	Error std	RMSE	Median Error	NMAD	90 percentile
All classes considered	96,68088919	11,51245022	-17,1117897	-2,570799828	4,296855927	**5,007194519**	-1,426728725	3,053117746	7,488256454
[0; 5]	19,41542775	11,51068687	-17,11167336	-0,956427693	2,515786886	2,691456556	-0,537604868	1,557664882	3,382747984
[5; 10]	19,03275916	11,51111889	-17,11172104	-1,591539025	3,341993809	3,701609373	-0,862389803	2,131008445	5,051651955
[10; 25]	34,10669282	11,5099659	-17,1117897	-2,700713396	4,285223007	5,065274239	-1,697534323	3,34579797	7,292960358
[25; 45]	18,50349057	11,51155567	-17,1117897	-4,297511101	5,137206078	6,697724342	-3,761078358	5,221416146	9,172075653
[45; inf[	5,622518895	11,51245022	-17,11154556	-4,989738464	5,617648602	7,51368475	-4,995785236	5,860637662	10,65087337

ALOS	% Of Valid Points	Max Error	Min Error	Mean Error	Error std	RMSE	Median Error	NMAD	90 percentile
All classes considered	98,215589[illegible]3	64,91111755	-70,58598328	-3,620969772	5,[illegible]8136549	**6,321233273**	-1,971942663	2,56707281	8,338098812
[0; 5]	19,5833883[illegible]	[illegible]4,84627533	-69,96243286	-1,647333264	[illegible],830607653	3,275063992	-1,22998631	1,434907854	3,224345589
[5; 10]	19,28440745	64,9111[illegible]	[illegible]	[illegible],204917669	3,593720436	4,216216087	-1,461113691	1,720932062	4,399199438
[10; 25]	34,64179483	64,57949066	-70,29610443	-3,457961798	4,894978046	5,993188381	-2,029710293	2,575174381	7,812008762
[25; 45]	18,93485384	64,40345001	-70,58598328	-6,276783466	6,522634029	9,052224159	-4,605126381	5,623174806	10,52302246
[45; inf[	5,771145289	54,87805557	-69,65821075	-7,314857483	6,801689625	9,988499641	-6,310265541	6,428858248	10,9352932

It is possible to analyze the altitude errors from the approximate Gaussian curves of the histograms of the distribution of the altitude errors according to the slope classes. These graphs show, once again, differences or similarities in terms of quality between the DEMs analyzed. Thus, it appears that for the Copernicus DEMs, the differences observed between the two DEMs 30 and 10 m are minimal, and they concern a better respect of the steep slopes (i.e., 25-40%).

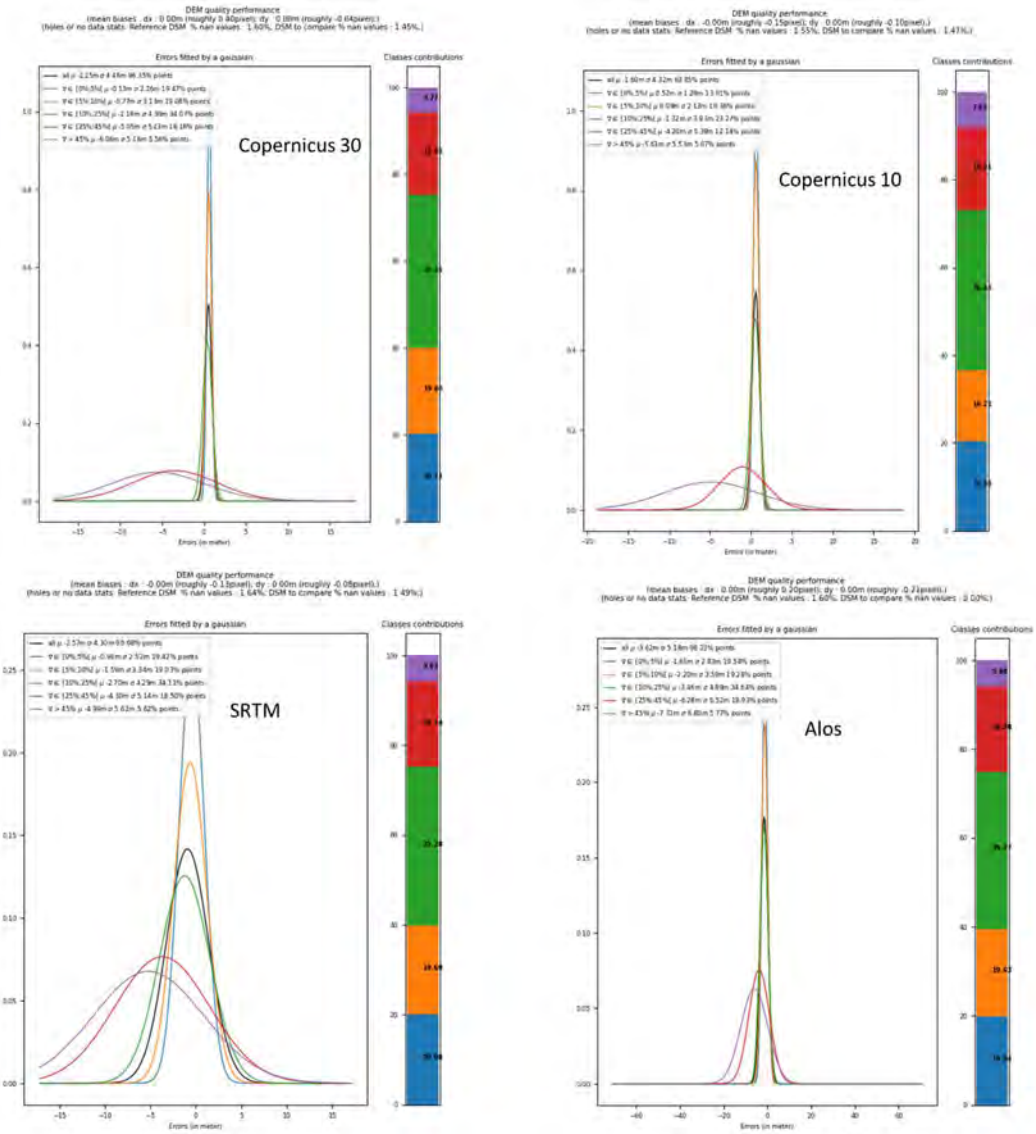


Figure 30: approximate gaussian curves of histograms of the distribution of elevation errors according to slope, classes

In conclusion of this analysis of the quality of DEMs and their ability to account for slopes, we can see that the Copernicus 10 and 30 data are quite similar in terms of performance, although Copernicus 30 drifts somewhat on the highest slope values. Overall Copernicus 30 and SRTM are quite similar, while ALOS shows the highest variability.

Copernicus 10 having been considered only in order to illustrate/understand a DEM of higher resolution, the analysis of the global DEMs in relation to a reference, the IGN 5 m, show that the Sar DEMs take better account of the slopes than the ALOS DEM, and present differences in altitude of a multi-metric order in relation to the reference, whereas the ALOS DEM presents decametric or even multi-decametric differences.

These results match with those obtained during the work carried out in the framework of the SERTIT support to SWOT activities (Yésou et al., 2018), and also with what is described in the literature (for example in Schumann and Bates, 2020)

For these reasons, we propose for further work, based on this analysis, to qualify only the SRTM and Copernicus 30 Global DEMs, having a slight qualitative advantage, and to reject the ALOS optical DEM.

4.3 Analysis and validation

For Phases I and II, the validation on the selected sites was performed:

- ✓ By comparative automatic analysis based on *in situ* data,
- ✓ By evaluation of the relevance of the different SURFWATER products in terms of representativeness of the reservoir dynamics. For this purpose, an independent extraction of the water surfaces was carried out on a reduced number of reservoirs, using the SERTIT EXtractEO chain, and the dynamics as described by the SURFWATER products were compared to those obtained from this chain,
- ✓ By comparison of the performances according to the DTMs used for the H/S/V modeling,
- ✓ By comparative or cross-analysis of some results obtained on different sites or groups of sites compared to the expected references or to the improvements made during phase I.

4.3.1 Qualitative assessment of the water extent

4.3.1.1 Methodology

In order to evaluate the quality of the water extent generated by Surfwater, several comparisons were performed on two tiles (30SUH and 30STF). During this step of validation of water extent, we have only analyzed the quality of results generated by Surfwater from Sentinel-2 optical data (Instant-mask and Water-mask).

For this we compared the surfaces extracted over a period of 3 years both with the ExtractEO software and with Surfwater.

ExtractEO is a software developed by ICube-SERTIT, implementing automated end-to-end chains on satellite data. ExtractEO can process different satellites, Sentinel-2, Landsat, Sentinel-1... and has several chains implemented, from water and fire extraction to cloud detection. The objective of ExtractEO is to work both in a time series context and in large study areas (large spatial and temporal windows). Water surfaces are detected using a multilayer perceptron algorithm (neural network) and integrating the Global Surface Water database for sampling.

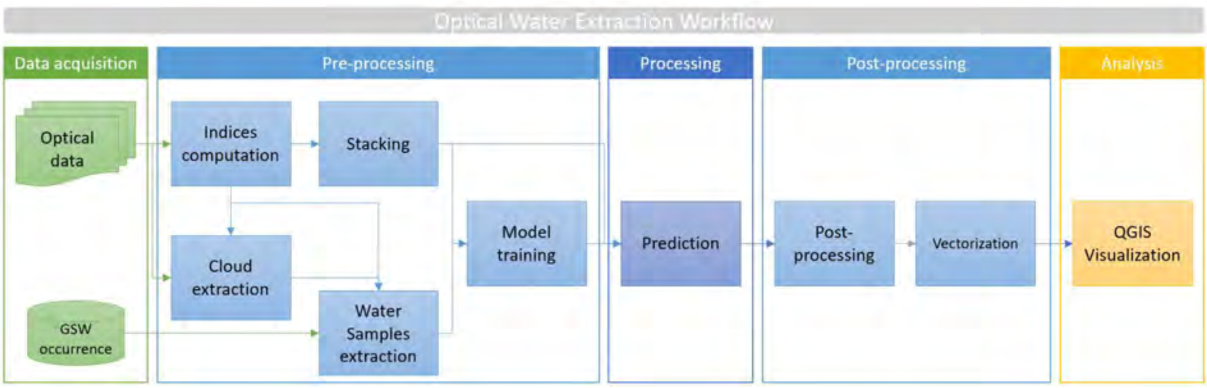


Figure 31 : architecture of the EXTRACTEO software suite

For the 30SUH tile, we compared the water extent generated from the SERTIT ExtractEO tool with the water extent generated by Surfwater (Instant-mask and Water-mask).

The results from Surfwater (Optical Random Forest based on Global Surface Water as reference) were compared with the ExtractEO results.

For tile 30SUH, we compared the water extent generated from the SERTIT ExtractEO tool with the water extent generated by Surfwater (Instant-mask and Water-mask), but also with the *in situ* water extent generated by the CNES "timeseries_tool.py" tool, based on *in situ* water volumes and heights. Only the results from SURFWATER with SW configuration in optical (random Forest mode), were analyzed for this step.

For these different comparisons, we compared the water extent generated at the dates common to ExtractEO and SURFWATER while removing the water extent for which the presence of clouds had been identified by one of the two tools (Nodata MAJA for SURFWATER and Cloud mask for ExtractEO)

4.3.1.2 30SUH tile analysis

This 30SUH tile includes 4 reservoirs ranging in size from 581 ha to 31811 ha.

Figure 32 : Location map Sentinel-2 (tile 30SUH)

We find a good consistency between the SURWATER Water Mask and ExtractEO data, despite a gap between the values, which is explained by a tendency to commission by ExtractEO at the level of the shadows of the reliefs and problems of classification on the part of SURFWATER which creates omission.

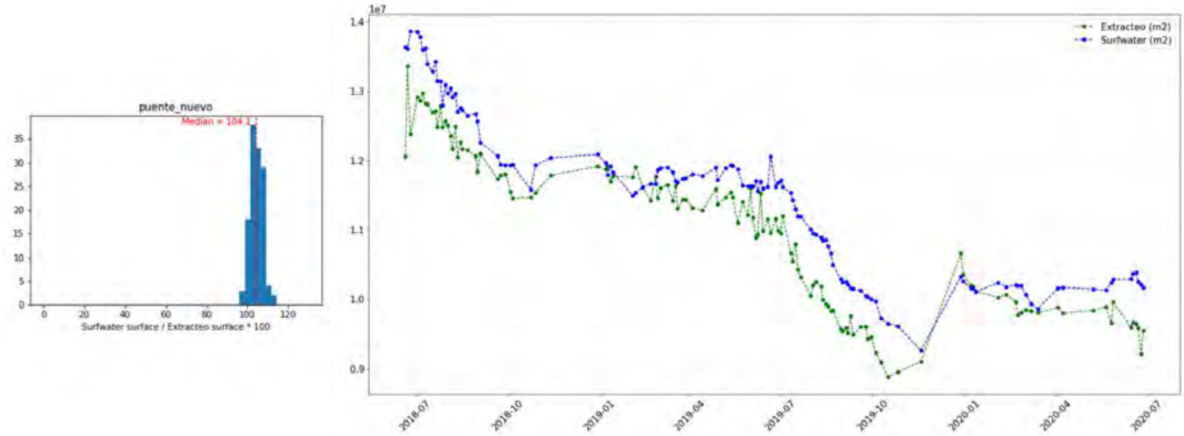


Figure 33 : Puento Nieve - Water Mask

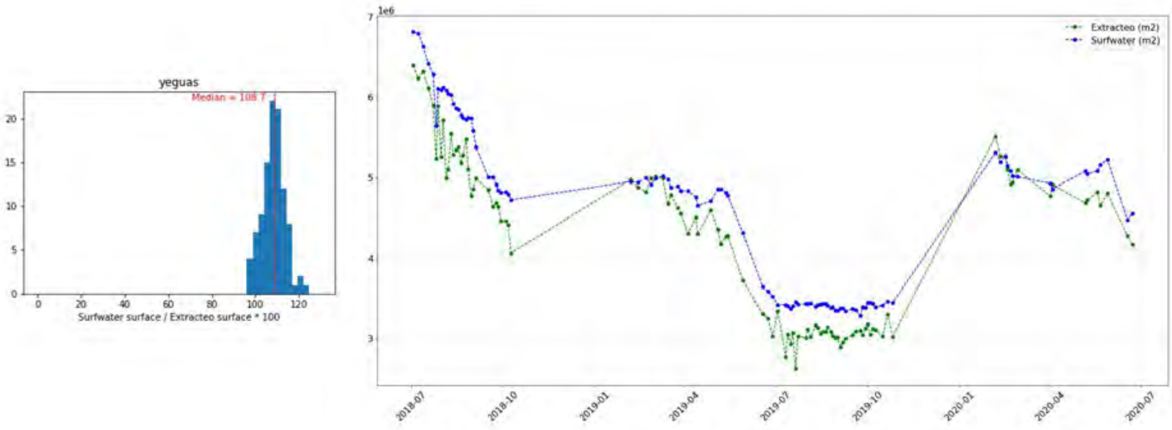


Figure 34 : Yeguas - Water Mask

4.3.1.3 30STF tile analysis

This tile includes 9 reservoirs for which it was possible to generate *in situ* water extent from the volume/height data provided by the providers. The size of the reservoirs varies from 63 ha to 3193 ha.

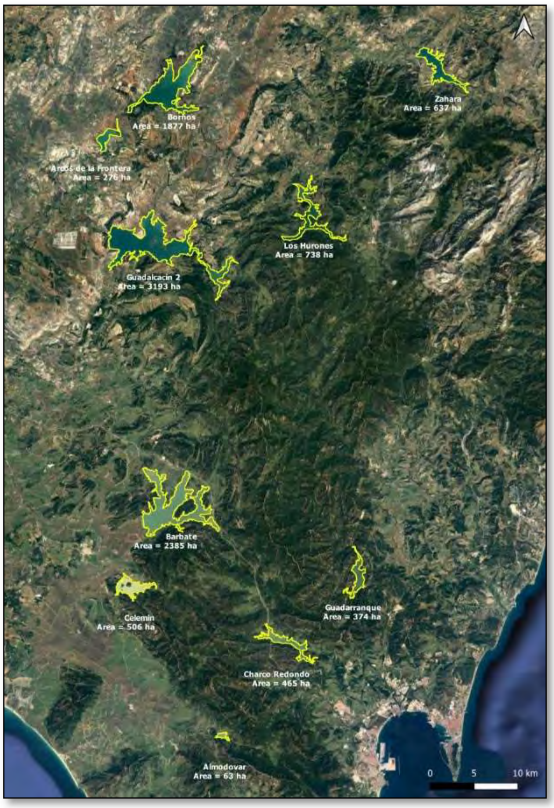


Figure 35 : Location map Sentinel-2 (tile 30 STF)

The analysis leads to the following observations:

- ✓ Overall, better results are obtained with the Multiple date SURFWATER Water Mask than with the Instant Mask. Unfortunately, it sometimes deteriorates the quality of the results (see Figures 40 and 41),

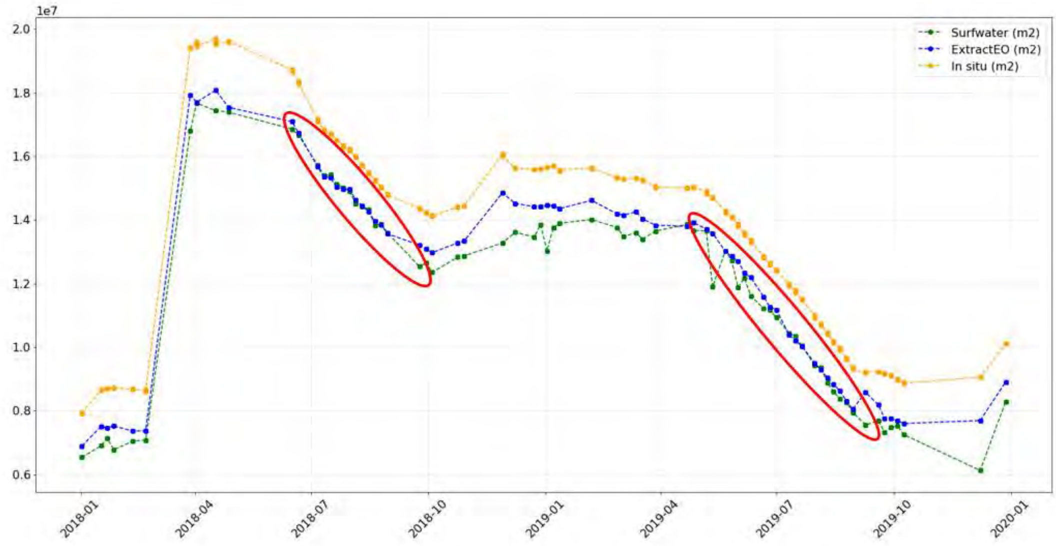


Figure 36 : Bornos - Instant Mask

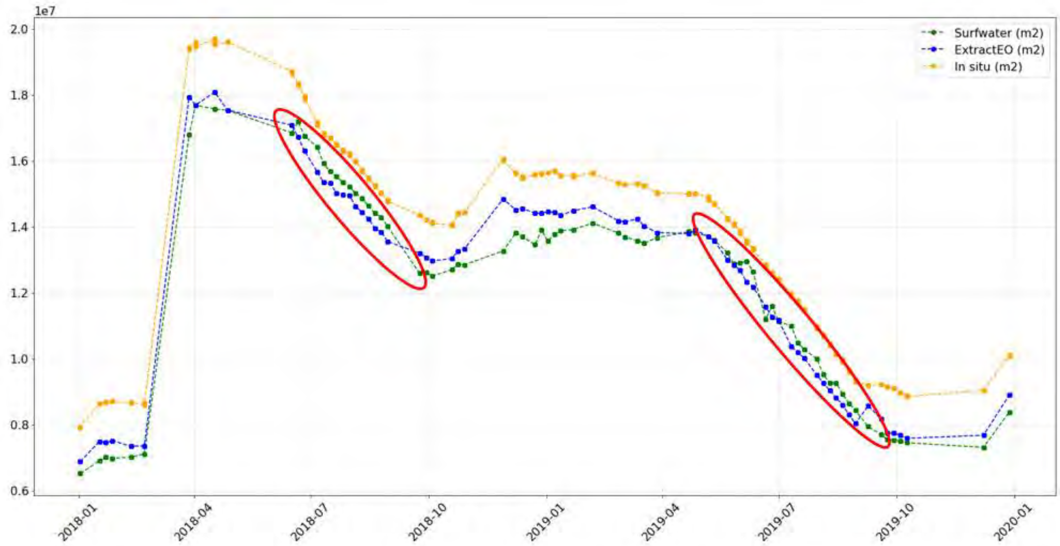


Figure 37 : Bornos - Water Mask

- ✓ ExtractEO is often closer to the *in situ* data (see Figures 42, 43, 44). Classification problems are identified in summer for SURWATER, around July, probably due to sunglint (see Figure 42).

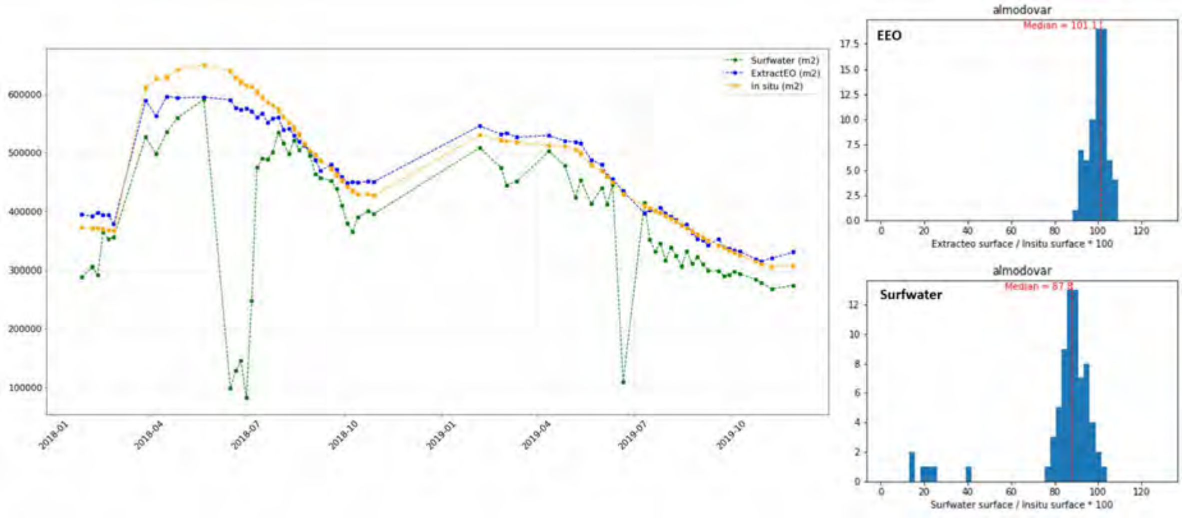


Figure 38 : Almodovar - Water Mask

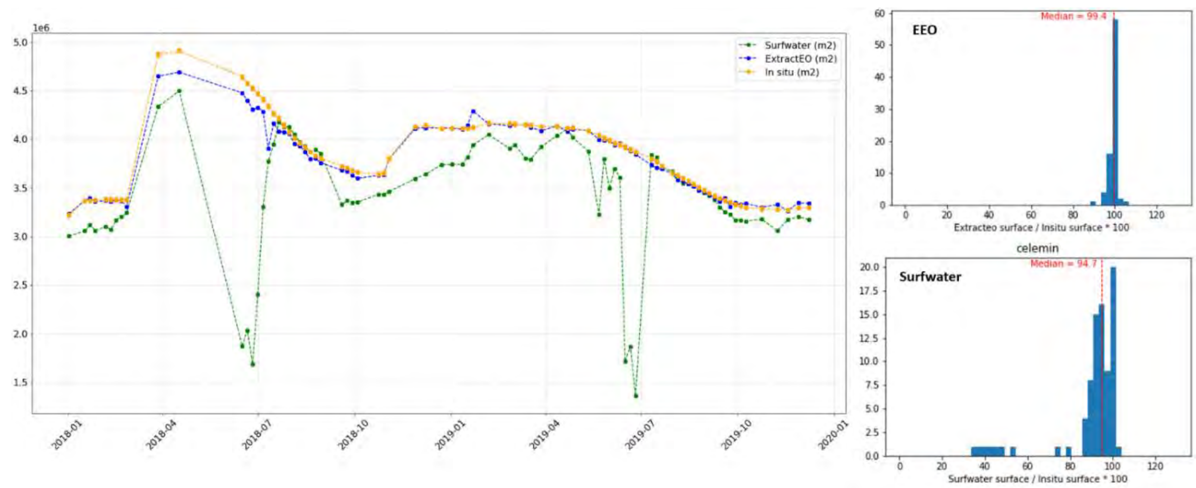


Figure 39 : Cemelin - Water Mask

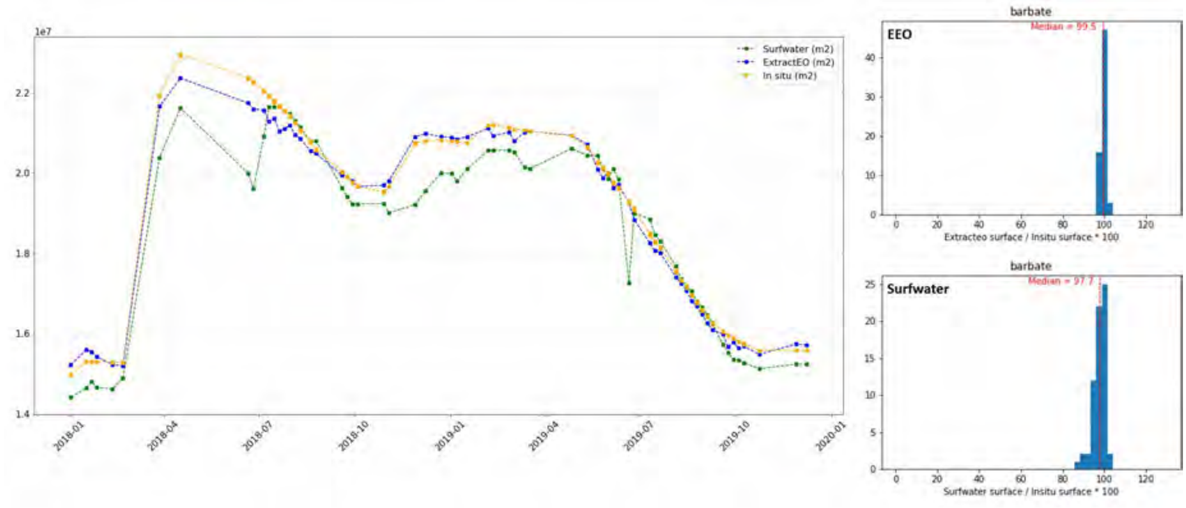


Figure 40 : Barbate - Water mask

In order to analyze the artifacts observed in the SURFWATER series, a comparative analysis of masks from SURFWATER and ExtracEO was performed on data from the Barbate reservoir, which showed an anomalous drop in June 2018 (see Figure 45). The analysis of the Sentinel-2 data shows that the water surface has a set of textures/tints, likely caused by the effect of wind and solar illumination (sunglint). Surfwater is more sensitive to these variations in surface conditions, and part of the water surface is not recognized as such, hence the negative peak observed. (see Figure 41).

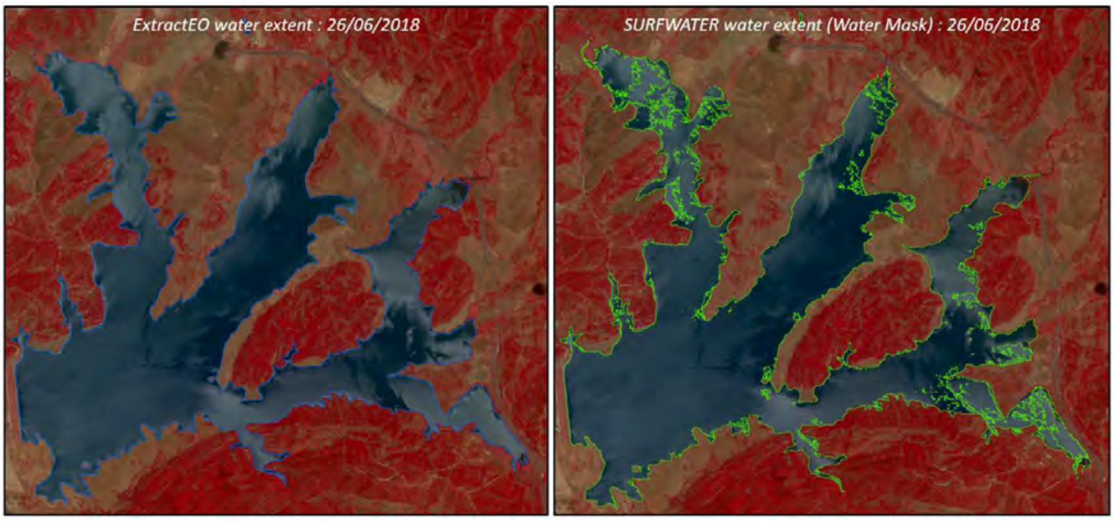


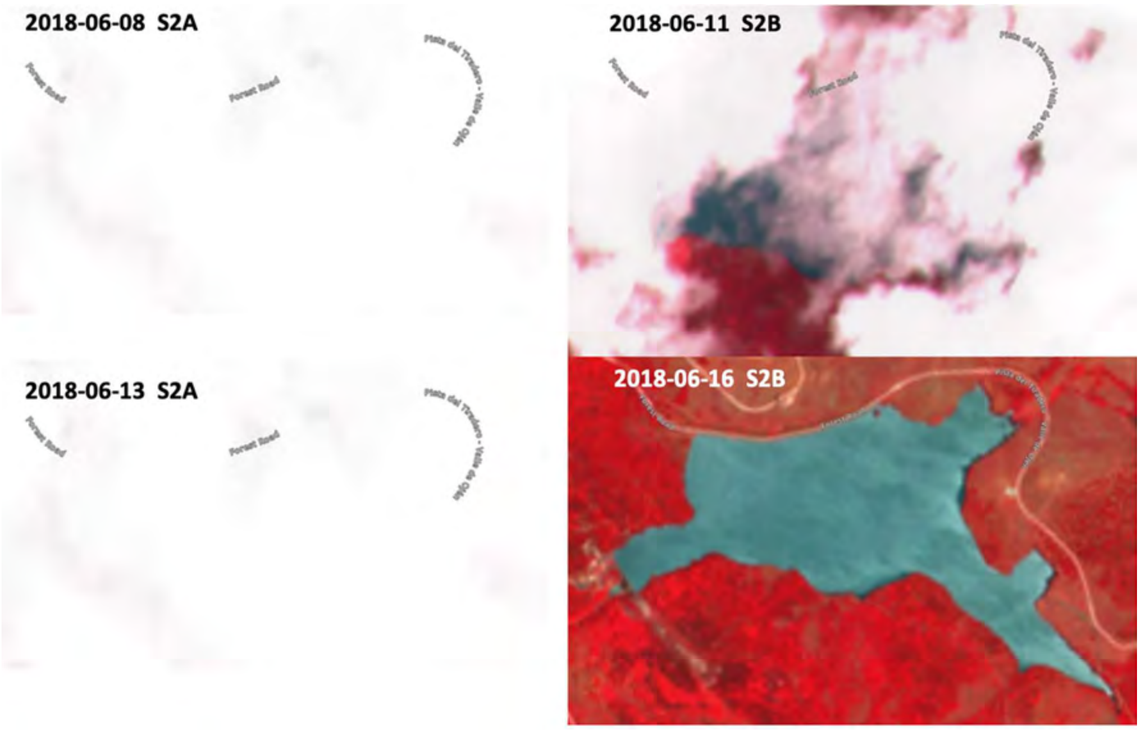
Figure 41 : Water Extent EEO & SURFWATER - Barbate - 26/09/2018

The two Sentinel-2 satellites (A and B) acquire images over the southern tip of Spain, more or less completely. The Sentinel-2B track covers the entire tip, while Sentinel-2A, more offset to the east, covers only partially this area. It is assumed that this is the reason why data acquired by Sentinel-2B have been selected for this study.

However, it is interesting to note that the effects seem to be stronger when the targets are located in the center of the image. As a result, the Sentinel-2B data are more subject to sunglint effects than the data acquired by Sentinel-2A. These effects are illustrated on the Almodovar reservoir, but a similar demonstration could have been made on other reservoirs for example that of Charco Retondo.

Figure 42 : Sentinel-2B footprint (left) and Sentinel-2A (right)

It is also interesting to compare the Sentinel-2B data with each other. Indeed, the dates for which the Surfwater data had the most difficulties to extract the water surfaces, are 2018-06-16, 2018-06-21, 2018-06-26, 2018-07-01. These images acquired during this period have the brightest water surfaces. The water surfaces returning to a "normal" look at the end of the presented series (late July 2018).



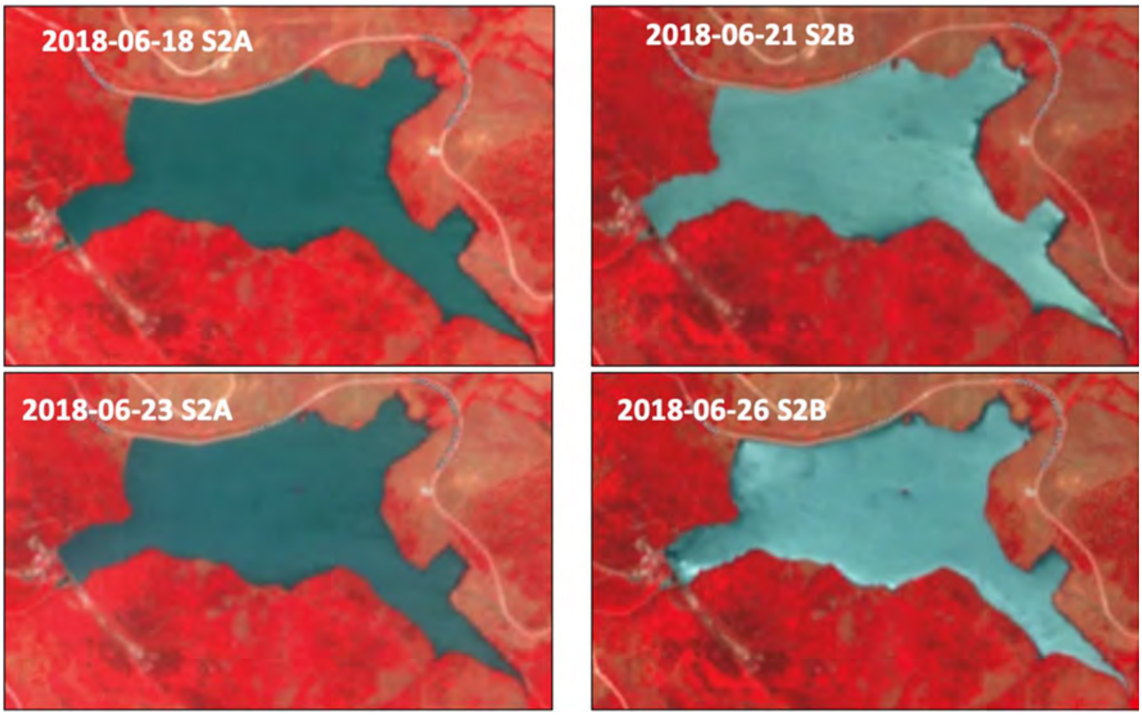


Figure 43 : Reflectance more or less impacted by sunglint effects depending on the trace location. On the left Sentinel-2A, and on the right Sentinel-2B. Example of the Almodovar reservoir. Colored composition S2, B8, B4, B3 in RGB.

On the Los Hurones, Bornos, and Char Retondo reservoirs, a gap (regular shift) is observed between the ExtractEO/Surfwater data and the so-called "*in situ* " surface data, with a similar overall behavior of the series (see Figures 51, 52, 53). This shift could have been explained by a systematic omission of part of the water bodies. A comparison between the images, the SURFWATER and ExtractEO extractions was carried out for the Bornos reservoir and on three dates, corresponding to three levels of filling (full, intermediate, and low water level) (see Figures 54, 55, 56). The water surfaces extracted on the three dates are very close/nearly similar between the SURFWATER and ExtractEO approaches and appear to be representative of the observed water surface. There are no under- or over-lit areas, which could have been missed, and there is no significant vegetation on the edges of the lake that would have caused shadows, resulting in strong omissions.

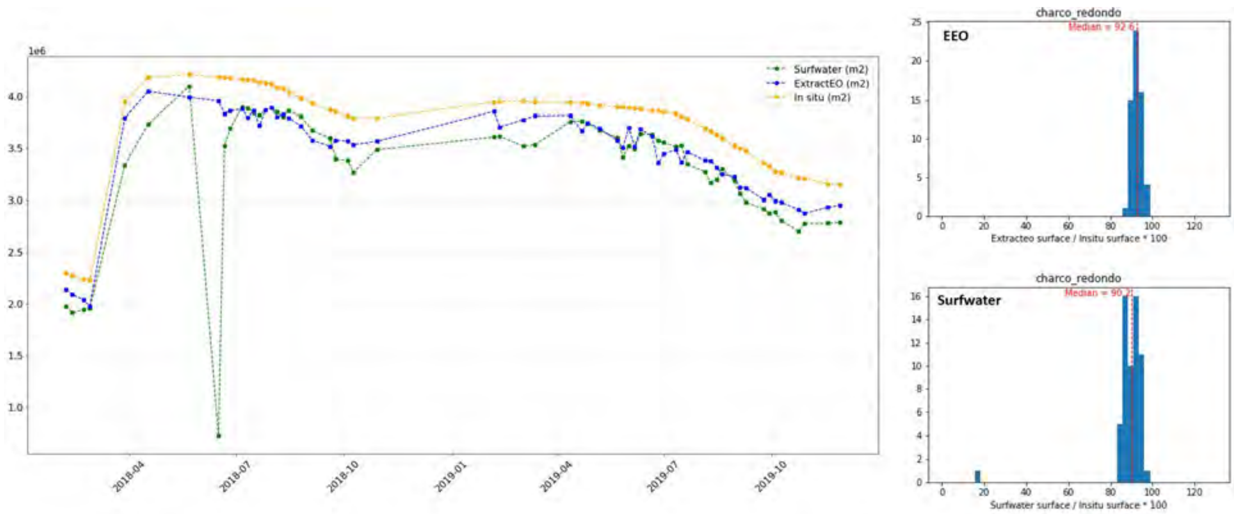


Figure 44 : Water mask - Charco Redondo

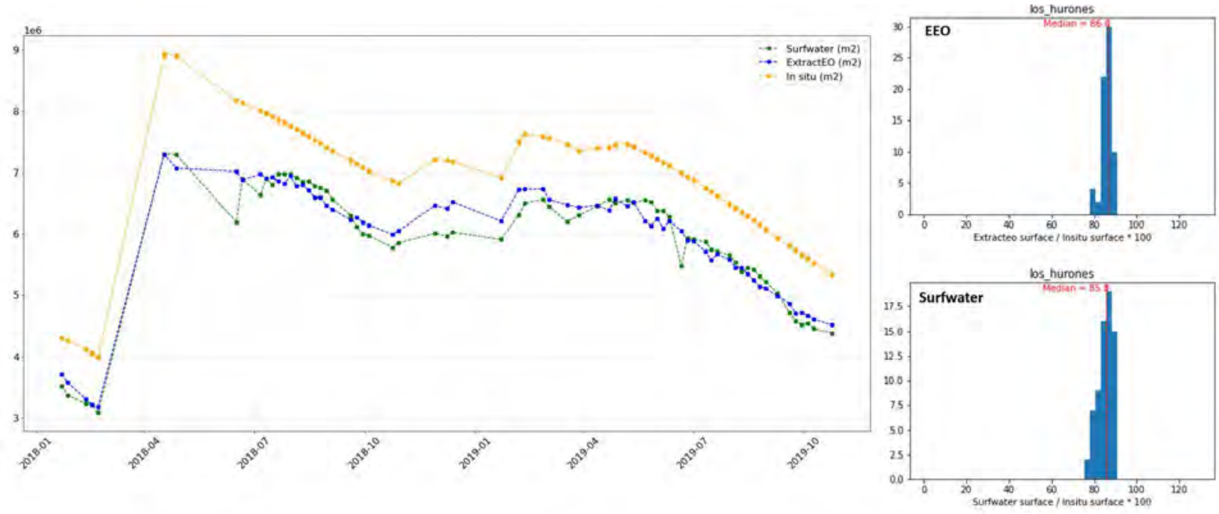


Figure 45 : Water Mask - Los Hurones

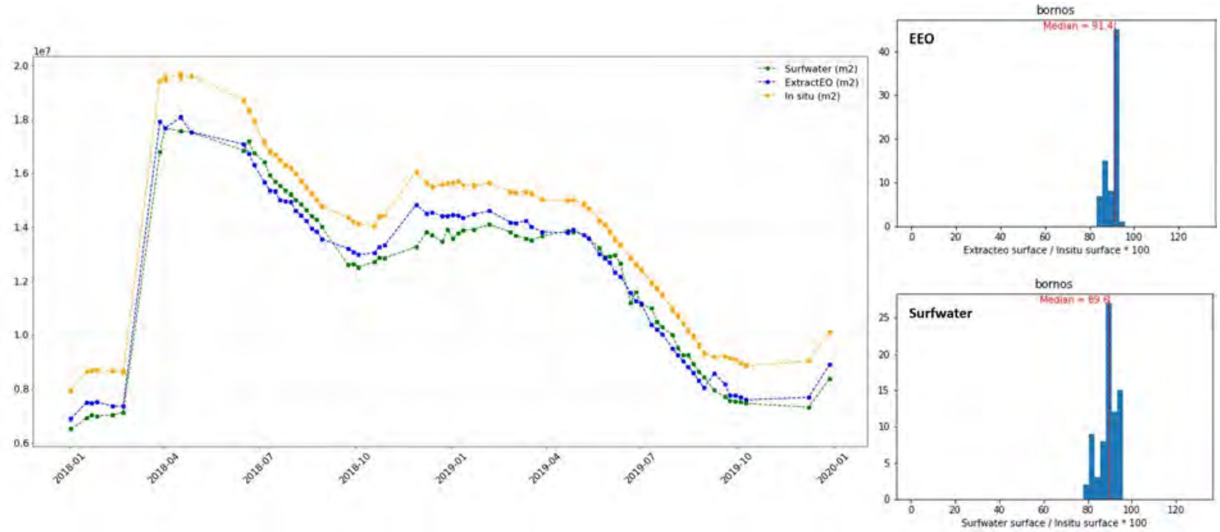


Figure 46 : WATER MASK – Bornos

Figure 47 : Water Extent EEO & STOCKWATER - Bornos - 27/04/2018

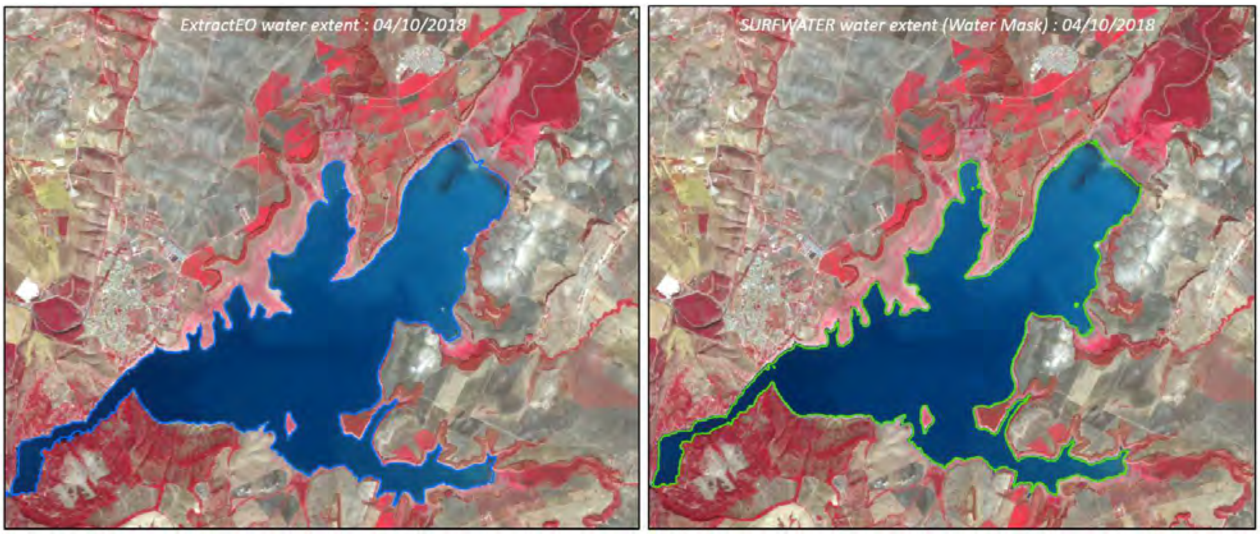


Figure 48 : WATER EXTENT EEO & STOCKWATER - Bornos - 04/10/2018

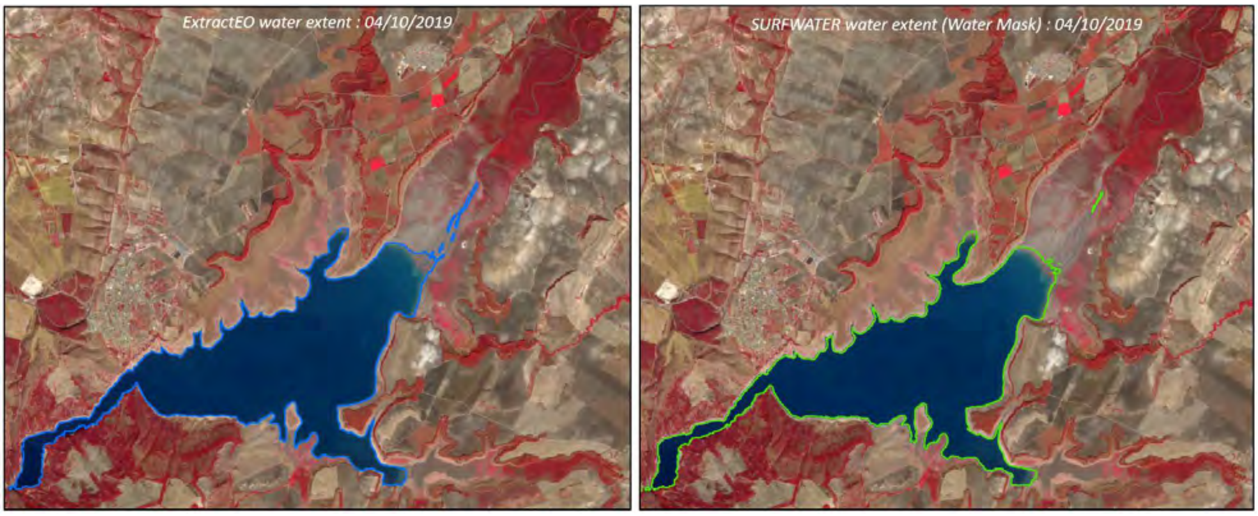


Figure 49 : WATER EXTENT EEO & STOCKWATER - Bornos - 04/10/2019

Also, if we consider the SURFWATER and ExtractEO extractions as valid and optimal in terms of representation of the water surface, this leads to question the quality of the so-called reference data.

The constant shift observed between the Surfwater and ExtractEO data was calculated for the Charco Redondo, Los Hurones and Bornos reservoirs with values of 7, 12 and 8% respectively. It is interesting to note that if we subtract this bias from the so-called reference values, we obtain an almost perfect agreement between the surfaces from the satellite data and the "corrected" *in situ* values (*cf.* Figures 57, 58, 59).

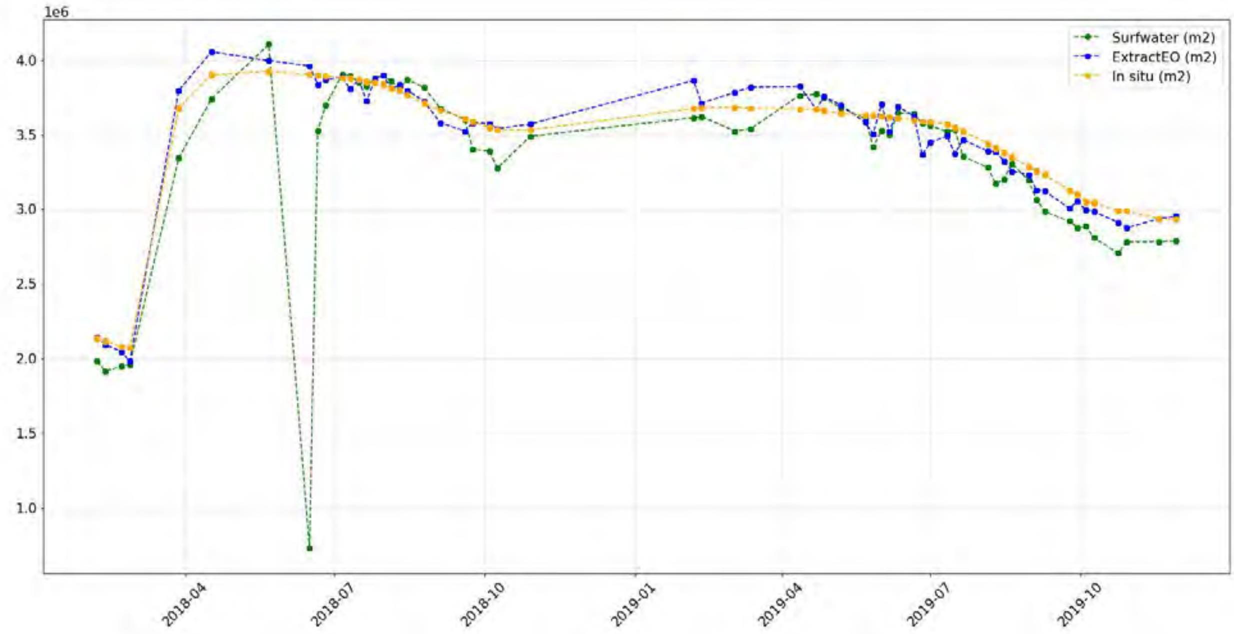


Figure 50 : Comparison after subtraction of 7% from the reference area – Charco Redondo

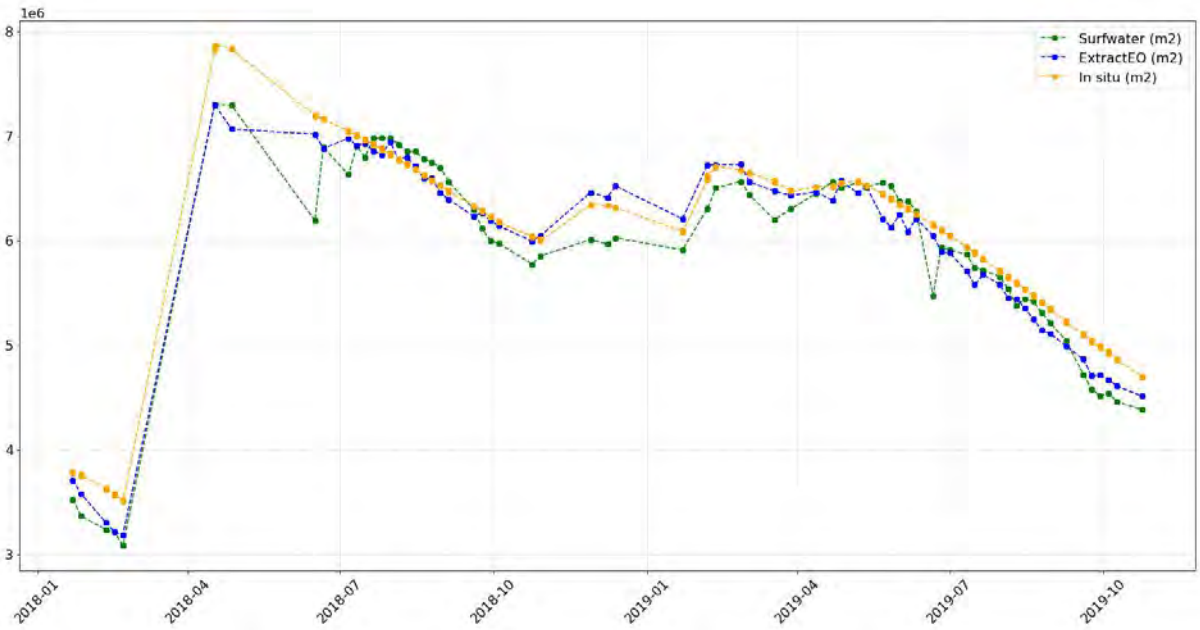


Figure 51 : Comparison after subtraction of 12% from the reference area – Los Hurones

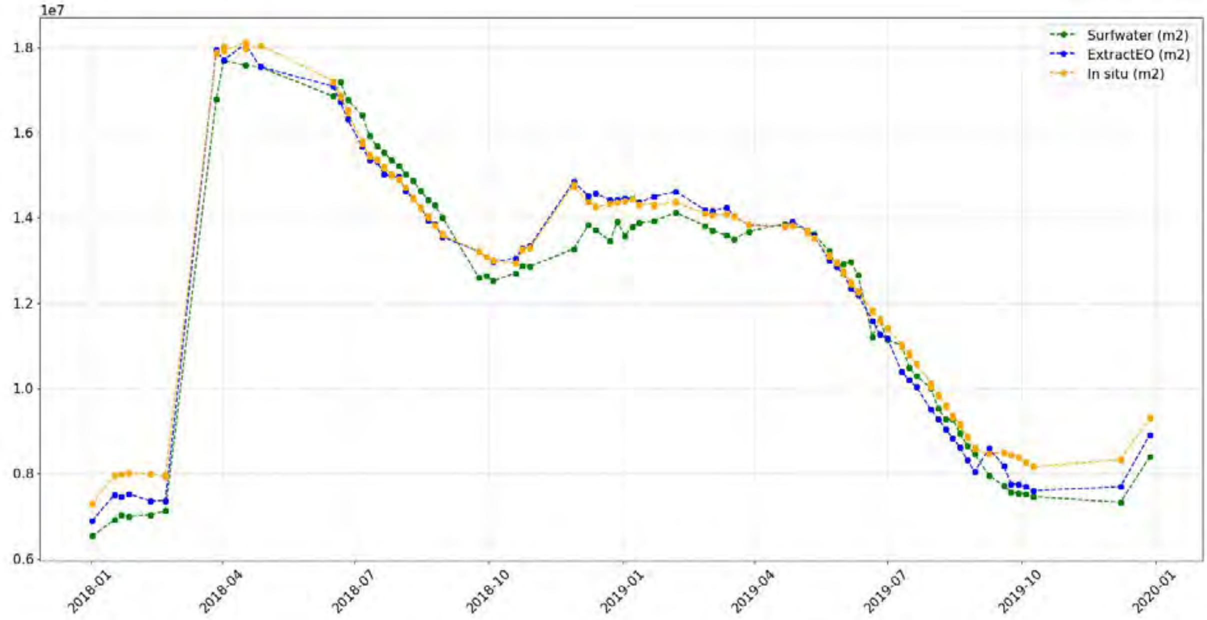


Figure 52: Comparison after subtraction of 8% from the reference - Bornos

This confirms the idea that we must question the quality of the *in situ* data. Indeed, this so-called reference or *in situ* surface data are in fact derived from an H/S/V model. At least two possibilities could therefore lead to an overestimate of the *in situ* surfaces:

- ✓ The first would be to question the quality of the model/calculation method used to generate the surfaces from the *in situ* Volume and Level,
- ✓ The second leads us to reflect on the quality of the data provided by the dam managers. Indeed, contrary to the heights, which are measured physical data, the volumes are quantities resulting from modeling. If the site manager's modeling is erroneous, this could explain these "inconsistent" reference surfaces,
- ✓ One way to overcome this problem could be to work not in terms of absolute volume but of volume variations.

4.3.1.4 Comparison of Reference Area, ExtractEO Area and SURFWATER Area

On two tiles, 30SUH and 30STF comprising respectively 4 and 9 reservoirs of sizes ranging from a few hundred hectares to several thousand, a comparison of water extent generated by SURFWATER and the Sertit tool was conducted. Moreover, on the 30SUH tile, two parameterizations of the Surfwater suite were tested. We observe a clear improvement by going through the second setting.

In general, the water surfaces generated by the ExtractEO tool are slightly higher than those obtained by Surfwater and seem to be more realistic. Moreover, we observe on the SURFWATER series a certain number of dysfunctions leading to variations/anomalies of the water surfaces (with very low values obtained). These malfunctions are related to more or less marked sunglint phenomena.

Surfwater shows a strong sensitivity to these phenomena which are reinforced on the L2C products compared to the L1C products. This could partly explain the fact that the ExtractEO suite of SERTIT using L1C data is less sensitive to these phenomena which significantly affect the series from Surfwater. The use of the Water mask rather than the Instant mask seems to partially overcome these problems.

4.3.1.5 Conclusions

The comparison with the so-called *in situ* surface series (modelling from the heights/surfaces/volume relationship) leads to several considerations:

- ✓ Surfaces extracted from Sentinel-2 data follow the same trends as those observed from *in situ* data, re producing the reservoir filling/emptying dynamics,
- ✓ The ExtractEO surfaces are in all cases closer to the so-called *in situ* values, than the SURFWATER values,
- ✓ In almost a third of the cases, there is a systematic bias between the surfaces derived from the Sentinel-2 data and the references. The analysis of the images and extractions does not suggest that there was a significant omission in the extraction of water surfaces by SURFWATER or ExtractEO. This bias may be due to a less efficient operation of the "*in situ* " surface modeling for some reservoirs, or to an erroneous modeling of the *in situ* volume by the managers of these reservoirs. To overcome this problem, it might be interesting to work not in terms of absolute volume, but rather in terms of volume variations.

4.3.2 Analysis of the REF laws

During the production of the estimated HSV laws on the Andalusia, India and Occitania sites, anomalies were encountered, reflecting in some cases the limitations of the dem4water automatic estimators and also sometimes those of the input data.

Indeed, the variety of the local topography of the various sites studied, the singular shapes of certain water reservoirs and the imprecision of the input data have put the dem4water chain to the test. The list of anomalies encountered consists of a bad estimation of dam feet, a partially looped cut-off line, water crossing, complex cases of multiple spillways with several structures on the same dam.

Nevertheless, most of such errors (Dam footing, cut-off line loop, water crossing and multi weirs cases) are simply corrected by manual adjustment.

Automatically estimated Dam footings

From the longitudinal topographic profile downstream of the dam and the determination of the point of slope break, the automatic estimation of the height of the foot of the dam for the Pla de Soulcem reservoir in Occitanie gives an erroneous estimation of 29 m of the foot of the dam with a height of 1528 m instead of 1499 m. This difference in Z0 has a significant impact on S(Z) as can be seen in the graph below between the estimated model S(Z) in green and the reference *in situ* model in red.

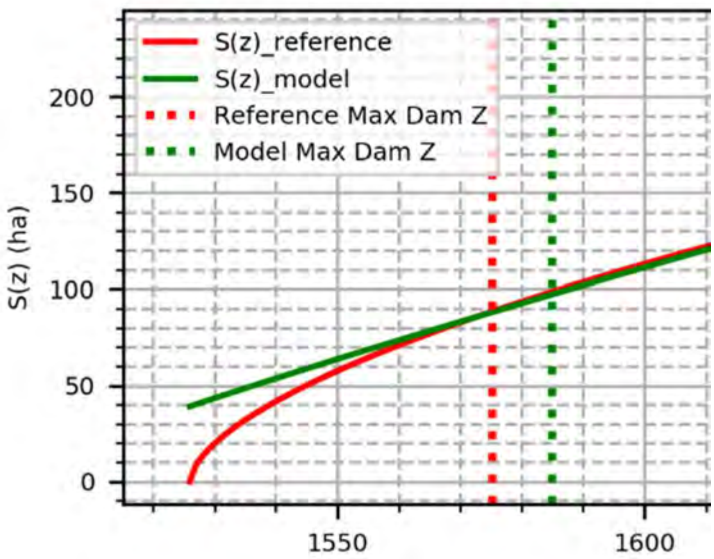


Figure 53 : Comparison of S(Z) with the reference

Consequently, this error has a strong impact on the estimation of the V(S) law as shown below with the estimated model in green and the reference *in situ* model in red.

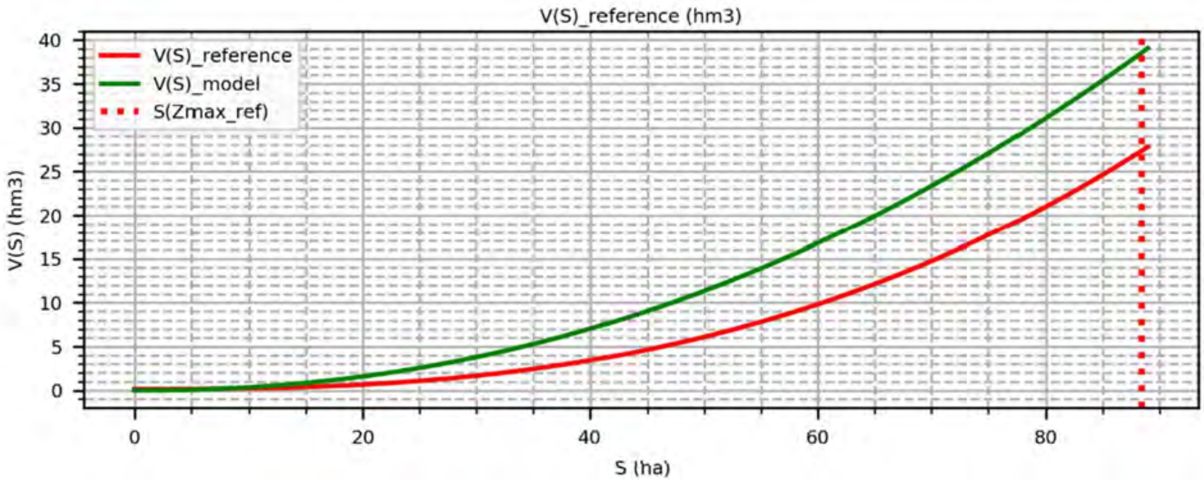


Figure 54 : V(S) comparison with the reference

Cut-off line loop

The case of the Yeguas dam in Andalusia shows how the automatic determination of the valley cutoff line, calculated from the DTM exploitation and the dam location via the reservoir database, gets lost in the loop. However, this is not a blocking anomaly, with a minor impact on the generation of the HSV law of this dam.

Water crossing

An example of a water crossing on the cut-off line for the Jose Toran dam. Considering the total size of the dam, this is not a blocking anomaly for the calculation of the HSV laws, especially since the water surface concerned is very small.

The Andevalo site cumulates the 2 previous anomalies on the valley cut-off line.

Multi-weirs

The Guadalteba reservoir has the peculiarity of having several structures, the valley cut-off line identifies only one structure, since it is geolocated by a single point in latitude and longitude provided by the reservoir database.

4.3.3 S(t), V(t), Tr(t) Time series analysis

During phases I and II, an analysis of the time series from the Optical and Radar sliding windows was performed. Indeed, these types of surface extraction proved to be the most relevant and robust in this study.

For this analysis of the results, a typical summary table was set up, to obtain a rendering allowing to clearly structure the results coming from the Dem4water and SURFWATER_Postprocess chains and thus to be able to evaluate the quality of the results more easily

A tool (Notebook) has been set up to automatically generate this table and to quickly obtain a summary of the quality of the generated time series and V(S).

This methodology and these tools will also be used for the analysis of the results obtained in the scope of the complementary activities.

Description of the columns:

- ID_SWOT: SWOT identifier.
- Dam_name: Name of the dam.
- V(S)-(%) : Average of the 'V(S)' deviations on three slices (low = 0-33%, medium= 33%-66% and high= 66%-100% of the Z(t) observed insitu) allows to evaluate the quality of the V(S) Dem4Water law
- TR_MO2-(%) : {Median, Quantile 75, Quantile 90} of the absolute relative deviations of the filling rate for the optical sliding window (MO2).
- TR_MR2-(%) : {Median, Quantile 75, Quantile 90} of the absolute relative deviations of the fill rate for the radar sliding window (MR2).
- Volume_MO2-(%) : {Median, Quantile 75, Quantile 90} of the absolute relative deviations of the volume for the optical sliding window (MO2).
- Volume_MR2-(%) : {Median, Quantile 75, Quantile 90} of the absolute relative volume deviations for the radar sliding window (MR2).
- Surface_MO2-(%) : {Median, Quantile 75, Quantile 90} of the absolute relative deviations of the surface for the optical sliding window (MO2).
- Surface_MR2-(%) : {Median, Quantile 75, Quantile 90} of the absolute relative deviations of the surface for the radar sliding window (MR2).

A tool (Notebook) has been set up to automatically generate this table and thus to quickly obtain a summary of the quality of the generated time series and V(S).

4.3.3.1 Andalusia

TABLE 9 : MEDIAN-ANDALUSIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2310018272	Almodovar	33.547278	8.640105	3.807054	40.293719	18.260598	5.042347	7.093556
2310018492	Arcos de la Frontera	nan	5.521457	4.643713	32.912032	45.559400	nan	nan
2310022062	Arenoso	nan	4.705129	11.294963	20.165054	50.695183	4.737193	30.101512
2310018473	Barbate	16.224382	5.009977	4.445845	15.518853	17.247211	1.485256	2.501627
2160004183	Beznar	6.439839	3.028011	5.291297	12.182245	8.471982	6.363693	4.166509
2310018403	Bornos	16.921904	7.341080	12.701782	2.421764	6.860859	10.018735	11.931699
2160004602	Casasola	29.030360	9.898731	18.340475	11.575996	12.454832	8.292030	12.438131
2310020223	Celemin	1.479886	10.264586	13.782729	4.546566	12.004184	1.779907	4.751051
2160004643	Charco Redondo	6.796969	4.017612	3.403600	4.325230	11.507426	2.014981	10.876315
2160004403	Conde de Guadalhorce	13.902395	6.083551	6.982693	8.589586	12.696954	3.855485	5.032415
2160003673	Cuevas de Almanzora	39.476015	38.753236	12.720898	52.093551	14.749333	7.200243	26.465142
2310022963	Giribaile	11.548433	2.445960	5.252608	14.620347	8.787124	2.979537	2.512043
2310020153	Guadalcacin 2	11.479984	3.043650	1.903974	10.508936	13.890824	1.848481	2.074174
2310022023	Guadalen	89.275800	34.556496	42.533256	89.253653	89.669937	4.318011	8.785087
2160004373	Guadalhorce	nan	6.286298	6.668644	24.949633	23.894376	3.687836	4.531548
2160004383	Guadalteba	73.066744	3.518998	4.683328	17.793655	14.621797	5.629889	3.597679
2160004623	Guadarranque	6.155025	3.488211	4.599676	11.881497	19.627326	3.326524	7.104867
2310024073	Iznajar	5.355205	1.950412	2.015796	4.592357	9.262872	5.383316	8.295655
2310023523	Jose Toran	28.495275	28.821432	24.805022	3.638160	4.842784	14.054625	17.213996
2310024253	La Brena II	3.068848	57.968239	60.310976	2.737888	10.002299	2.952571	9.679181
2160004663	La Concepcion	11.107910	10.583083	8.502378	5.955679	5.762209	6.606121	7.977125
2160004413	La Vinuela	2.902015	2.181131	9.361156	5.925951	6.836616	2.180723	3.474610
2160004532	Limonero	6.493850	4.060364	9.072407	20.501719	32.388372	10.836801	19.994032
2310022903	Los Bermejales	5.789899	2.750491	9.798661	4.757764	10.205666	1.443207	6.795085
2310018233	Los Hurones	23.623679	17.428371	17.182404	7.429911	4.461895	9.389569	13.754018
2310027322	Los Melonares	6.400711	3.434543	2.415320	13.669876	21.180346	6.232305	12.169398
2310024113	Puebla de Cazalla	0.928407	4.441831	5.084491	20.210513	27.205604	12.055494	16.645726
2310027683	Puente Nuevo	16.699458	9.704795	3.120744	16.711627	9.288584	1.598177	3.812201
2160004253	Rules	3.300350	1.809185	5.521387	14.965960	42.369308	7.072566	27.813227
2310000173	San Rafael de Navallana	3.159786	8.531998	10.216608	7.349391	18.920846	4.681572	13.248553
2310020933	Tranco de Beas	10.442158	5.431621	10.986788	8.311613	6.292428	2.869281	2.560959
2310023943	Vadomojon	12.727538	3.071674	2.899687	17.156901	24.485631	4.083173	9.842282
2310022743	Yeguas	5.100191	3.592094	13.396673	3.757266	17.468366	5.192712	12.503133
2310018393	Zahara	0.450902	2.907220	2.376253	14.379610	22.930431	8.763362	14.442386

TABLE 10 : SUMMARY OF MEDIAN RESULTS - NUMBER OF TANKS PER CLASS - ANDALUSIA

	V(S)-(%)	TR_MO2(%)	TR_MR2(%)	Vol_MO2(%)	Vol_MR2(%)	Surf_MO2(%)	Surf_MR2(%)
<= 10%	15	27	22	14	10	29	19
> 10 & <=25%	10	3	10	16	18	4	11
> 25%	6	4	2	4	6	0	3

4.3.3.2 Occitania

TABLE 11 : MEDIAN-OCCITANIA

	ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
6	2160030553	Agly	5.711032	6.851354	5.714414	4.820893	19.933550	3.033697	16.028295
15	2320039133	Astarac	13.847274	15.275600	20.175701	14.474329	29.130089	5.768444	12.346579
5	2160030183	Avene	nan	5.132743	9.099241	33.354524	61.188980	9.007280	39.957780
10	2320030823	Cammazes	3.318017	6.568112	11.956606	9.444146	46.505842	8.848530	33.560017
14	2320038733	La Gimone	9.497278	4.387762	3.148246	12.496536	5.115862	1.726755	7.331327
9	2320030233	Lac de Montbel	87.039512	7.453731	16.919095	81.291950	82.359563	19.627945	19.695025
11	2320031293	Laparan	69.055075	19.416749	49.986447	23.498691	49.576424	9.767403	42.483917
1	2160028013	Matemale	34.228502	3.074687	6.378637	29.581491	38.336876	4.761081	3.917078
7	2320028893	Pareloup	33.081892	4.771804	8.269476	22.882545	18.973725	3.536807	5.827029
12	2320031303	Pla de Soulcem	67.837554	6.676346	27.837119	33.909342	28.877195	4.761080	23.321128
3	2160029943	Puyvalador	19.861078	7.702238	8.341651	23.046999	21.236073	4.663609	5.480270
8	2320028933	Saint Geraud	nan	9.594513	13.209795	12.366989	42.145512	14.963699	19.935873
13	2320033043	Saints Peyres	99.987579	10.573200	17.287488	99.983535	99.990778	4.935649	18.799789
0	2160026973	Salagou	90.151847	1.261693	1.783709	91.573992	92.073186	1.194278	2.338390
2	2160029873	Villeneuve la Raho	99.957016	8.034100	6.902714	99.897874	99.911546	10.128008	6.528470
4	2160030123	Vinca	77.570885	10.814732	15.102630	77.634535	81.491445	13.146897	20.817741

TABLE 12 : SUMMARY OF MEDIAN RESULTS - NUMBER OF TANKS PER CLASS - OCCITANIA

	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Vol_MO2(%)	Vol_MR2(%)	Surf_MO2(%)	Surf_MR2(%)
<= 10%	3	12	8	2	1	12	6
> 10 & <=25%	2	4	6	6	3	4	7
> 25%	9	0	2	8	12	0	3

4.3.3.3 India

TABLE 13 : MEDIAN - INDE

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
4530212143	Himayat Sagar	44.017309	16.824274	35.168083	40.409705	44.481561	24.233585	16.480837
4530347283	Nagarjuna Sagar	89.965918	204.033804	155.473813	573.274159	389.458117	19.510134	17.121652
4530212043	Osman Sagar	115.957147	312.575013	101.380842	123.039879	42.071812	24.134926	16.404753

4.3.3.4 Results analysis

There are generally 2 types of errors:

1) Errors due to a bad estimation of the law V(S). These bad estimations of the V(S) law by Dem4water can have several causes :

- ✓ Multi-weirs problems (presence of several dams on a reservoir, etc)
- ✓ Wrong automatic estimation of the foot of the dam
- ✓ Wrong automatic calculation of the valley cut-off line

Note that it is sometimes difficult to identify V(S) estimation problems only from the average of the absolute 'V(S)_quality_low_mean', 'V(S)_quality_mid_mean' and 'V(S)_quality_high_mean' deviations, because a compensation of errors at the different levels is possible. It is therefore sometimes necessary to take into account the error values obtained at each of the 3 levels (Low, Mean, High).

2) Errors due to significant differences between the SURFWATER surfaces and the reference surfaces. These differences can be explained by:

Classification errors on the part of SURFWATER. This type of error is illustrated by a dispersion of deviations with variable deviations throughout the series

Poor quality of *in situ* data. These errors can be identified by a constant deviation along the series with a low dispersion of deviations and a good correlation of deviations between surface extraction methods (see Figure 69).

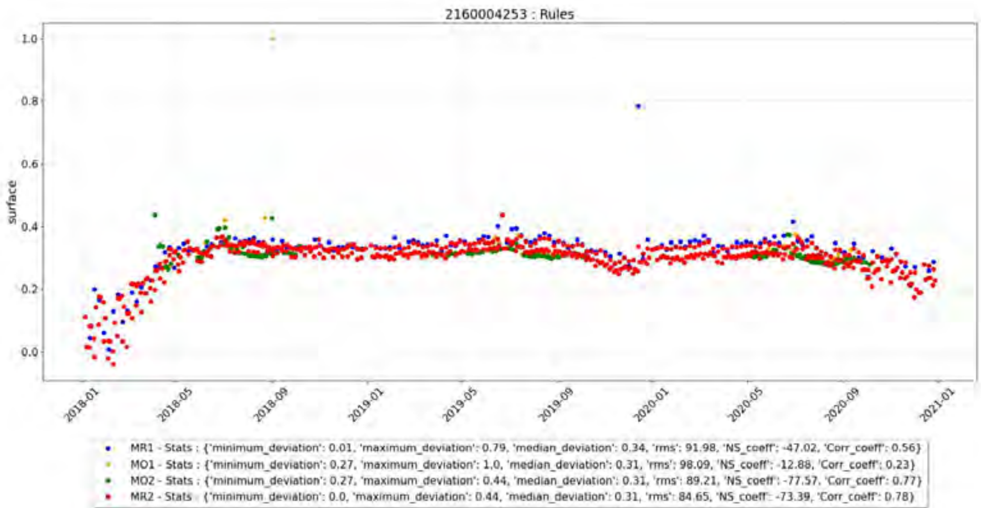


Figure 55 : Time series of absolute relative deviations – Rules (Andalusia)

These two types of errors almost systematically lead to poor estimates of the volumes that are directly derived from these laws and SURFWATER surfaces. (see tables 14, 15, 16).

TABLE 14 : EXAMPLE OF AREA ESTIMATION ERRORS IMPACTING VOLUME ESTIMATION - ANDALUSIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2160004643	Charco Redondo	8.924094	6.879428	3.807375	12.284053	12.131625	11.979577	11.281453
2160004183	Beznar	7.205803	4.577590	7.053454	23.446642	20.065446	14.647451	12.467572
2160004253	Rules	3.473714	3.727214	5.746094	47.162529	47.282534	31.131217	31.446644

TABLE 15 : EXAMPLE OF V(S) ERRORS AFFECTING VOLUME ESTIMATION - ANDALUSIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2310020153	Guadalcacin 2	11.383586	4.442626	1.739500	17.196794	14.206992	5.552399	2.413985
2310022963	Giribaile	13.827539	3.648985	7.985531	11.991661	11.547925	2.298460	2.529000

TABLE 16 : EXAMPLE OF V(S) ERRORS AFFECTING VOLUME ESTIMATES - OCCITANIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2160026973	Salagou	90.151847	1.984592	2.206573	92.350677	92.622686	3.262701	5.133346

Note that these errors sometimes compensate each other. Poor results for the estimation of the V(S) law and the extraction of surfaces by SURFWATER can therefore produce volumes that appear correct (see table 17).

TABLE 17 : EXAMPLE OF ERROR COMPENSATION - ANDALUSIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2310018233	Los Hurones	24.467642	19.112084	21.553131	5.349799	6.620677	17.061411	16.065529
2310018403	Bornos	17.176650	6.960333	13.919609	3.263753	7.815188	13.315401	12.187291

Fill rate :

Concerning the filling rates, we note that there are generally correct differences between the filling rates obtained from SURFWATER post_process and the filling rates obtained from reference data (cf. Figures 73, 75).

However, these results must be treated with caution, because the SURFWATER post-process and reference filling rates are based on their respective maximum volumes. They are therefore data relative to a maximum filling volume that must be observed within the observation window (normally 3 years are sufficient to observe these filling periods).

The above remarks are based on the analysis of the medians of the absolute relative deviations. If we use more restrictive metrics, such as the 75% quantiles, we observe a fill rate error of 13.02% in optical (MO2) and 11.56% in radar (MR2). At the 90% quantile, we observe a very strong increase in the deviations, with errors almost systematically higher than 25%.

At the 75% quantile, the RT errors are as follows:

Sensor = MO2

15.168395843765497 % (Average 75% quantile of RT for each dam)

13.024173096475 % (Average 75% quantile of RT for each dam)

Sensor = MR2

12.614921980481494 % (Average 75% quantile of RT for each dam)

11.56787438345% (75% quantile of TR of all measures combined)

One of the explanations for the increase in deviations as a function of the quantile is the persistence of errors in the extraction of surfaces from SURFWATER and particularly from optical imagery (used with the non-optimal configuration - parameterization 1). Thus, the majority of the surfaces extracted by SURFWATER are close to the so-called reference surfaces, but we observe a non-negligible population with strong differences. The influence of this population will thus have a more or less important impact on the analysis of the results according to the quantile (weak to moderate according to an analysis of the median and strong to preponderant for the quantiles 75 and 90) (cf. Appendix A).

To improve the robustness of the approach, a weekly indicator was evaluated using the median values of each natural week. This hardly changes the quality of the measurements (does not completely eliminate outliers), but allows a cleaner tracking of trends (especially in radar). The following figure shows such monitoring indicators as Volume and Fill Rate and allows to observe the compensation effects using the Fill Rate.

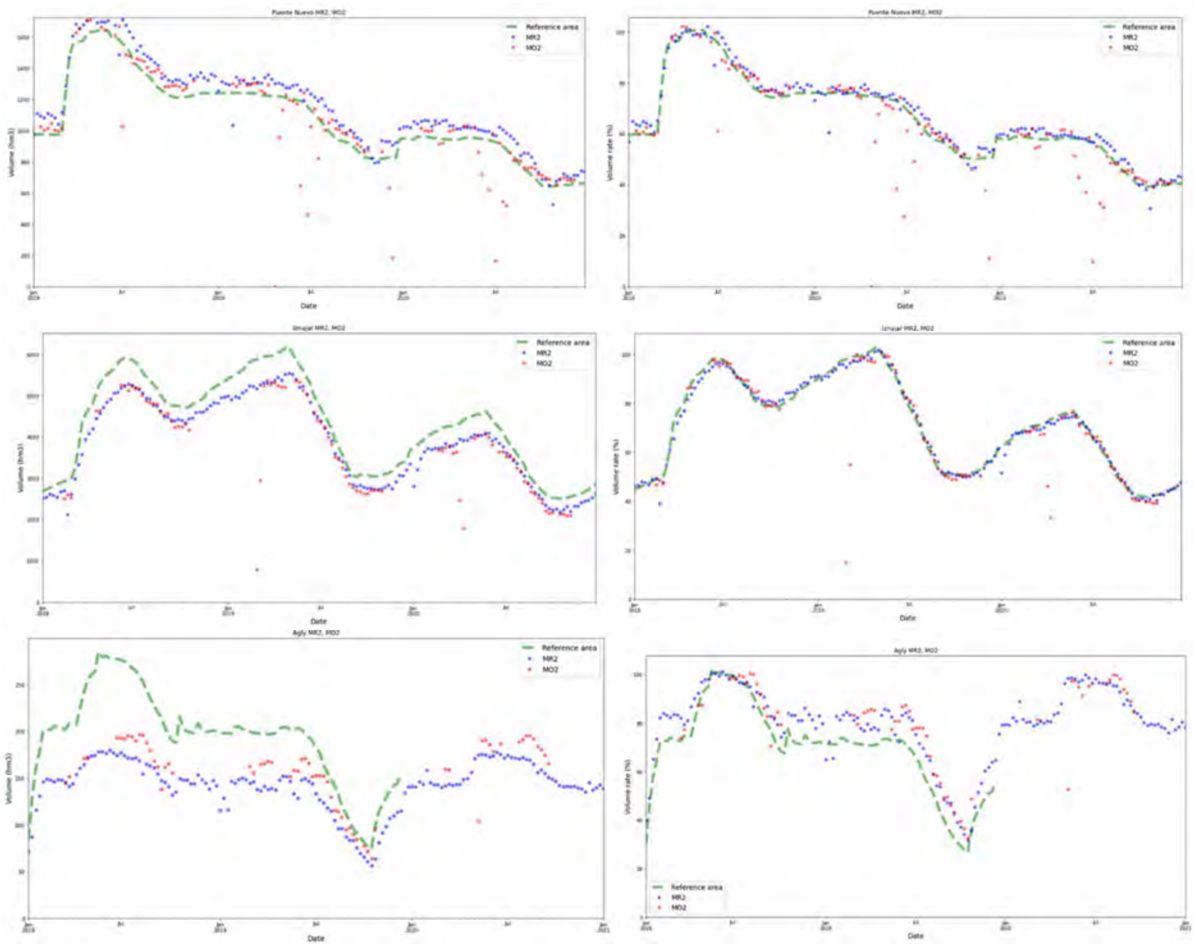


Figure 56 : Comparison of weekly volume estimates (left graphs) with filling rate estimates (right graphs) for the Pueblo Nuevo, Iznajar and Agly dams. In red the optical estimates (problematic configuration - MO2), in blue the Radar (MR2) and green line the reference (volume / filling rate)

5 Progress of Phase II

During the previous phase, the methodology was developed and implemented on the super sites of Andalusia, India and Occitania, selected on the one hand for the experience already acquired in these areas and on the other hand for the accessibility of reference data.

This first phase will have raised several problems/limitations at different levels of the process, access and quality of *in situ* data, access to image data, quality of optical extractions, robustness of the modeling, etc.

The methodology has therefore been refined as a whole and new solutions are being implemented in Phase II:

- ✓ improvement of the quality of the optical classification by applying a more adapted parameterization to the Surfwater chain,
- ✓ resolution of limitations on the generation of models by the DEM4water chain.

In addition to the replay of some steps of the super sites of Phase I, a scaling up is done with the application of this environment extended to the new super sites Tunisia, Burkina Faso, Laos, and Brazil.

5.1 Development of the demonstrator

5.1.1 Surfwater

New retrieval problems appeared on tiles of the Brazil super site with no Sentinel-2 data retrieved (same for the whole year 2021), although these data are present on the datalake. This anomaly was reproduced on the previous commit of the Amalthee channel by Robin Vermes. The problem identified is due to the version difference between our Amalthee client and the server, which do not define the same final paths for products that have the V3-0 extension. Moreover, Amalthee does not check the version of the client in the request made and no warning is returned if the versions are not compatible.

5.1.2 DEM4water

DEM4Water is basically a prototype whose initial objective was to demonstrate the feasibility of automating the modeling of surface/height and especially surface/volume relationships in artificial reservoirs. The upstream objective is to allow the monitoring by remote sensing of the quasi-real time capacities of these artificial reservoirs.

The improvements were mainly functional and algorithmic:

- ✓ Decrease the reprocessing time by implementing a system of caches of the level lines generated from the DEM,
- ✓ Improving the parameterization of the chain by allowing the user to substitute the automatically computed elements with inputs expertly corrected by the user. In this way, it is possible, in cases where the chain produces unsatisfactory results due to automatic detection of poor quality, to manually overcome these limitations:
    - In the case of a bad detection of the dam foot, which impacts the quality of the model estimation.
    - In the case of the cut-off line:
        - To handle cases where a reservoir has several dams,
        - The algorithm fails to reach sufficient altitudes to allow the collection of enough samples S(Z_i) on which the model estimation is based.
- ✓ Derive a satisfactory cut-off line in cases where:
    - The algorithm for determining the cutoff line generates geometries that are too complex (MultiLineString).
    - The cut-off line contains loops (red dotted line), they are then naively simplified
- ✓ Improve edge case management, with better reporting:
    - Management of comparison launches to a ground truth when it does not exist (induced by the way the chain is used by the Stockwater orchestrator, ideally, without ground data, the model quality calculation should not be launched),
    - Handling cases where the foot of the dam detection algorithm does not converge.

5.1.3 Surfwater_Postprocess

A new indicator has been developed within the Surfwater_Postprocess chain generating the time series. This weekly indicator of median type is applied on the time series of surface S(t), volume V(t) and filling rate TR(t) on one hand, and on the other hand on the differences with the reference as well as on their absolute values. The metrics comparison module has also evolved to generate the graphs related to this weekly indicator for each of the quantities as well as the result files in json format.

The actions on the extraction of surfaces realized by P.Tysebaert have been integrated in the Surfwater_Postprocess chain. Technical support was provided by CS GROUP-France for this work, as well as the creation of a complete test set with the procedure made available.

5.1.4 Stockwater orchestrator

The Stockwater orchestrator allows to realize the sequence and the resumption of the treatments for a studied super site. Some steps require an operator validation before the next one is launched, such as the production of Surfwater masks and the generation of estimated models. The creation/setting up of the context automatically managed by the orchestrator greatly simplifies the resumption of processing after the operator control steps.

Some additional evolutions were brought to the orchestrator: the addition in the parameter file of the possibility of having recourse to the occurrence map and an associated *percentile* following the evolution of the Surfwater_PostProcess chain, the calculation of the occurrence maps over the studied period for the radar mode, the generation of the vrt of the occurrence maps for the radar mode.

The auxiliary data composed of the databases, the estimated laws, the reference laws and the reference time series are updated according to the corrections brought to the data or the addition of new studied super site.

As part of the stage of valorization of the results of the Stockwater project via the web platform, it was necessary to develop tools to extract the dams individually from the files of estimated time series on the one hand, and reference time series on the other hand.

A formatting was carried out to obtain a csv file of time series by dam identifier, for a method directory (MO1/MR1/MO2/MR2) with a modification of format finally to put the header to the standards (fields "date, value") and a nomenclature of file of type: [dam_id]_[dam_name]_[quantity]_raw_[method].csv

The reference time series, in the form of a single file per super site, also required formatting for data extraction by dam identifier.

Subsequently, data filtering was applied to the estimated time series:

- ✓ remove quantity values at 0 as well as values for which the no-data are 0,
- ✓ if there are several measurements for the same date, keep the largest value.

A filtering was performed on the reference time series:

- ✓ restrict to the observation period to that of the Stockwater study,
- ✓ keep only one measurement per day: at 00:00:00,
- ✓ insert in the file nomenclature the name of the dam in addition to the identifier and delete the name of the super site.

5.2 Implementation of the database

5.2.1 Site selection

The selection of targets is a continuation of Phase I. The areas analyzed in this Phase II are in Brazil, Burkina Faso, Tunisia and, to a lesser extent, Laos.

Tunisia

Discussions with LISAH (Jean-Stéphane Bailly) and INRAE (Jérôme Molénat) were taken into account in the selection of the 10 sites illustrated in Figure 74. 6 sites are located on the Cap Bon peninsula (Lebna, Chiba, Kamech, Abid, Bezirk, Masri) and 2 in the vicinity (El Hamma, Er Rmal), *in situ* data are expected for these 8 sites. The site of Kamech is the smallest (~10 ha), it is a hill lake monitored in the framework of OMERE (www.obs-omere.org).

The 2 other sites selected (Sidi Salem and Sidi Saad) are large reservoirs (more than 10 km²) and are listed in the BD GRanD. The geometries of maximum extension of water surfaces are based on Global Surface Water, Geodar and GRanD. However, the geometries of 6 sites had to be adjusted by SERTIT.

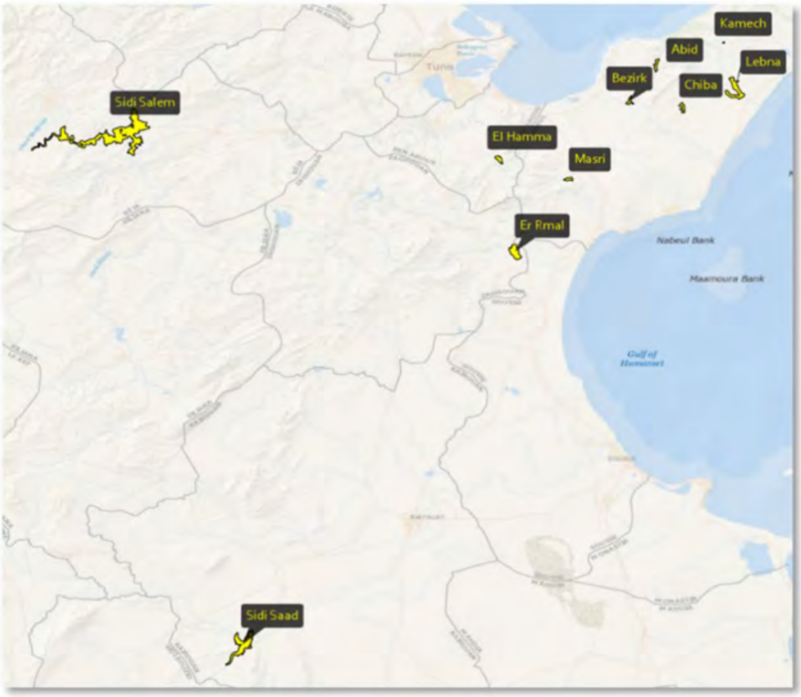


Figure 57: Sites in Tunisia

Brasil

Discussions with Marielle Gosset (IRD) and the Brazilian teams were considered in the selection of the 14 sites located in the Ceara province (Figure 75). The Bonito site is the smallest (~9 ha) and is kept as a test site despite its small size. 8 sites are under the SWOT 1 day orbit. Maximum water surface extension geometries are based on Global Surface Water, Geodar, and GRanD. However, the geometry of one site had to be adjusted by SERTIT.

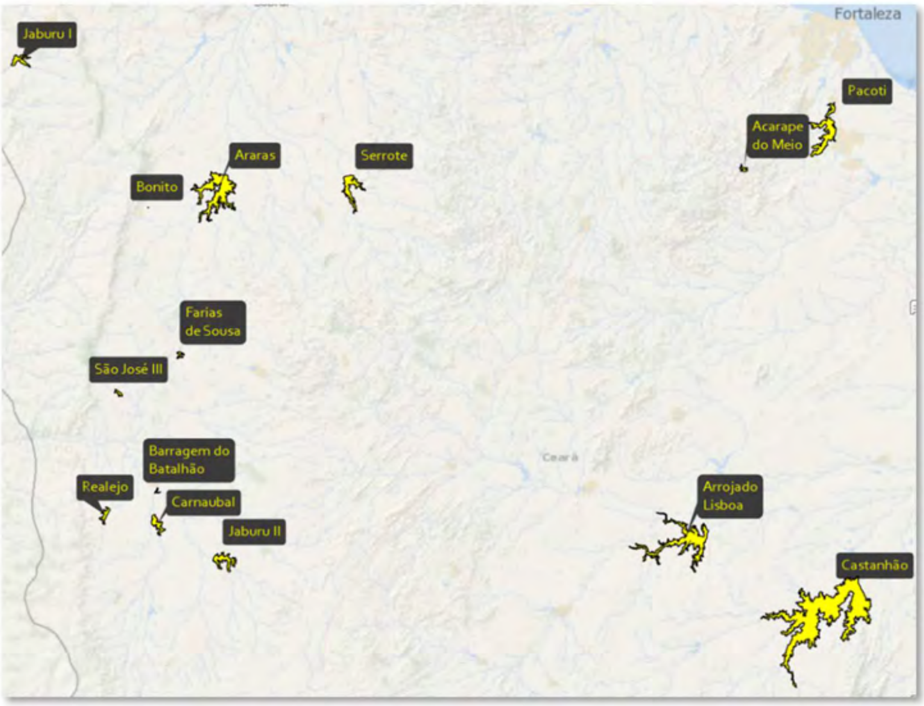


Figure 58 : Sites in Brasil

Burkina Faso

The priorities of the TEP (Manuela Grippa) were considered when selecting the 9 sites shown in Figure 76. The Tanvi2 site is the smallest (~56 ha). 3 selected sites are not listed in the GRanD. The geometries of maximum water surface extension are based on Global Surface Water and GRanD. However, the geometries of 2 sites had to be adjusted by SERTIT.

Figure 59 : Sites in Burkina Faso

Laos

The Lao site is the 540 km² Nam Theun 2 reservoir shown in Figure 77.

Figure 60 : Site in Laos

5.2.2 Data preparation

At the end of Phase I, a protocol was set up for the preparation of the data sets (Dam Database). This protocol was followed for the preparation of Laos, Brazil and Burkina Faso data.

5.2.2.1 Structure of the attribute file

At the end of Phase I, a protocol was set up for the preparation of the data sets (Dam Database). This protocol was followed for the preparation of Laos, Brazil and Burkina Faso data. (see Figure 78).

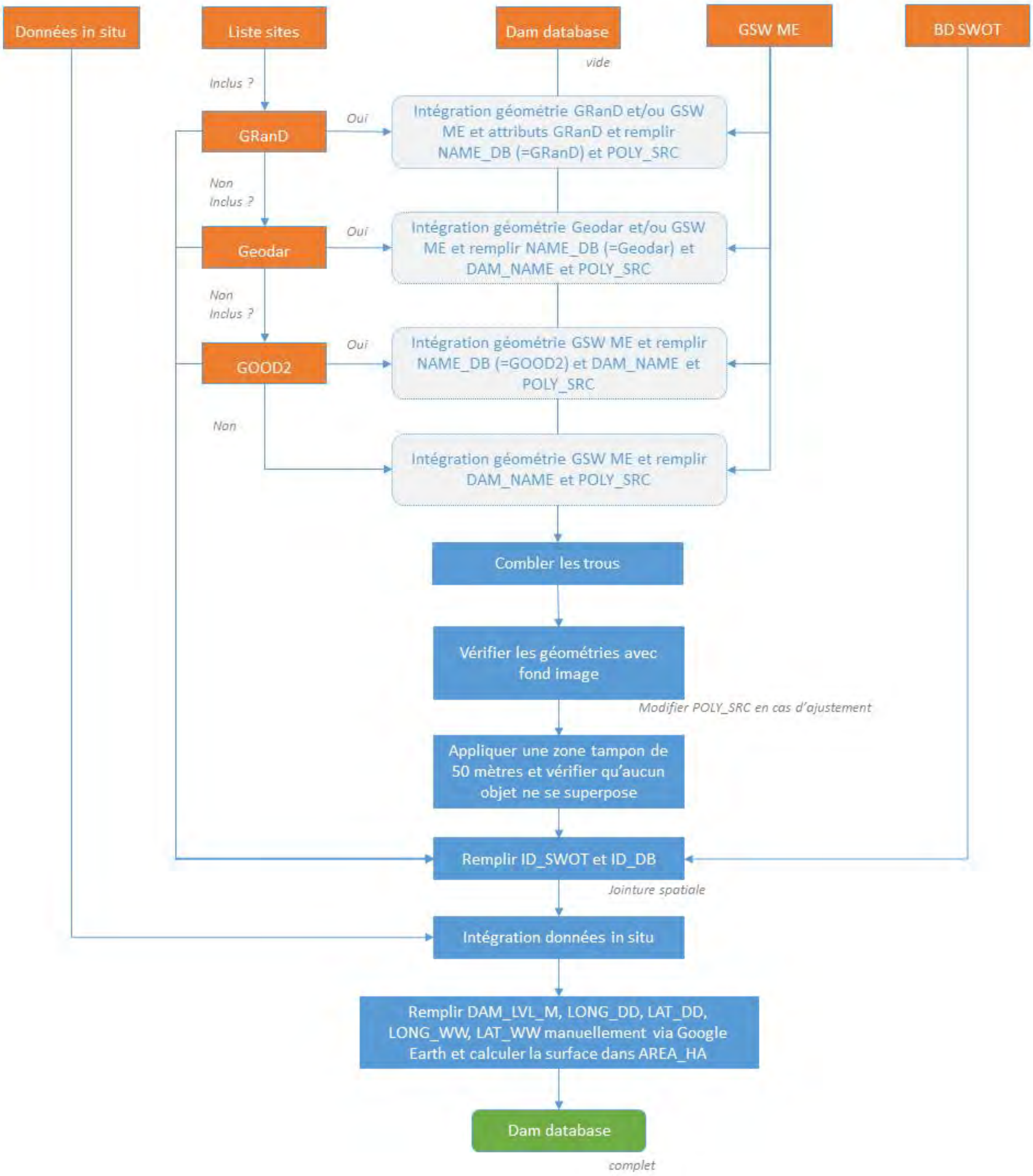


Figure 61 : Workflow for the preparation of a Dam Database dataset

Geometry

This section aims to obtain the maximum extents of the reservoirs from different databases (GRanD, Geodar and Global Surface Water Maximum Extent). These extension polygons are integrated into a Dam Database reservoir database in vector format, which has its own data model that must be respected (see Phase I). Certain attributes associated with the extents must be filled in when importing the vector geometries in order to keep the genealogy. The workflow presented in figure 78 takes up the elements presented in this section in a more visual and succinct way.

- ✓ Get the list of selected sites,
- ✓ Check their presence in GRAND with the name in the attributes. Integrate those present in the

Dam Database vector by keeping the attributes (automatically recognized). Fill in the fields NAME_DB = GRanD for these objects,

- ✓ For the others, check their presence in GEODAR (location via google or other mapping). Integrate those present in the Dam Database vector and fill in the NAME_DB = Geodar and DAM_NAME fields,
- ✓ For those which are neither in GRAND nor in GEODAR, check their presence in GOOD2 (location via google or other mapping). There is no associated surface geometry. The GSW ME object can be directly integrated into the Dam Database vector and fill in the fields NAME_DB = GOOD2 and DAM_NAME,
- ✓ For reservoirs that are not listed at all, integrate the GSW ME geometry associated with the Dam Database vector,
- ✓ For geometries from GRAND/GEODAR, merge them with the GSW ME geometries or even replace them if they are of poor quality (visual analysis),
- ✓ Fill in the gaps,
- ✓ Manually expand the geometries by checking with a background image to be sure to cover the maximum area of the reservoirs,
- ✓ Apply a buffer zone of 50 meters around the tanks,
- ✓ Check that no objects are overlapping in the Dam Database vector. If so, reduce the buffer zone of the concerned objects.

Attributes

Some of the fields in the attribute table are already filled in during the previous step, but there is still information to be filled in. The workflow presented in Figure 78 repeats the elements presented in this section in a more visual and succinct way.

ID_ SWOT: Manual spatial intersection with BD SWOT. This can be automated (spatial join) but beware, our reservoir can intersect several objects from the SWOT DB. I manually took the largest corresponding object when it was the case,

- ID_DB : Manual spatial intersection with GranD, Geodar or GOOD2. Cascading spatial join possible,
- NAME_DB : Name of the DB used to fill ID_DB. GRandD, Geodar or GOOD2,
- RES_NAME : Sometimes filled by GRanD. For Burkina I have indicated the names given by GET (example Burkina Faso),
- DAM_NAME : Filled by GRanD. For those that are not included, I have indicated the names given by GET (example Burkina Faso),
- NEAR_CITY : Inherited from GRanD only,
- COUNTRY : Inherited from GRanD only. Otherwise filled manually,
- YEAR : Inherited from GRanD only,
- DAM_LVL_M: Altitude read from Google Earth and manually reported (or *in situ* data).

Note 1: The base source for GE heights is SRTM but high-resolution data from various sources are integrated locally as they become available. Globally, MERIT, NASADEM and ALOS DSM are also available in Google Earth Engine. It is difficult to know the source of the altitude read in Google Earth. To know the source, it would be necessary to read on a DEM in a GIS or in GEE.

Note 2: The vertical component (altitude) is measured from the vertical datum, which is the WGS84 EGM96 geoid. In general, elevation can be considered as a measurement in meters above sea level (MSL)

- CAP_MCM: Inherited from GRanD only or *in situ* data,
- DEPTH_M: Inherited from GRanD only or *in situ* data,
- MAIN_USE : Inherited from GRanD only,
- LONG_DD : Coordinate read in Google Earth and transferred manually,
- LAT_DD : Coordinate read in Google Earth and manually reported,
- LONG_WW : Coordinate read in Google Earth and reported manually,
- LAT_WW : Coordinate read in Google Earth and reported manually,
- POLY_SRC : GSW(+GRAND+GEODAR+Refined by XXX),
- AREA_HA : Calculated at the end.

5.2.2.2 *In situ* baseline data

5.2.2.2.1 WATER LEVELS AND *IN SITU* VOLUMES

Contrary to Phase I, during which a large amount of *in situ* data was available at the different sites, during Phase II, *in situ* water level and volume data were only made available for the Brazil site. As done in Phase I, a step of analysis, selection, and correction of these *in situ* data was performed, to generate operational datasets, compatible with the operation of the STOCKWATER PoC. The fourteen selected Brazilian sites have *in situ* time series of water level and volume.

TABLE 18 : SUMMARY TABLE OF AVAILABLE *in situ* DATA FOR THE BRAZIL SITE

Dam_name	ID_SWOT	Z_INSITU	V_INSITU
Sao Jose III	6320007102	X	X
Barragem do Batalhao	6320007172	X	X
Jaburu I	6320007492	X	X
Carnaubal	6320008892	X	X
Jaburu II	6320009412	X	X
Realejo	6320015192	X	X
Bonito	6330000001	X	X
Serrote	6330068463	X	X
Araras	6330068933	X	X
Farias de Sousa	6330070272	X	X
Arrojado Lisboa	6330095223	X	X
Acarape do Meio	6330098332	X	X
Pacoti	6330098612	X	X

5.2.2.2.2 QUALITY ANALYSIS AND IMPROVEMENT OF *IN SITU* DATA

As observed at the different sites during Phase I, it turned out that the *in situ* data (Water Heights and Volumes) provided for the Brazilian sites had inconsistencies, preventing their use as is for the generation of *in situ* laws. These inconsistencies were corrected, for the most part, using the scripts developed in Phase I.

The *in situ* data from reservoirs 6330098593 - Castanhão, 6320015192 - Realejo, 6320007102 - São José III, 6320007492 - Jaburu I and 6330068463 - Edson Queiroz were kept as they were, as they did not present any major inconsistencies.

Regarding reservoirs 6330098612 - Pacoti, 6330068933 - Araras and 6330098332 - Acarape do Meio, corrections were necessary to eliminate outliers and "plateaus" / non-variable data. Inconsistencies visible on the V(Z).

Despite the modifications made to the water level and volume data to obtain consistent V(Z) relationships, there are still inconsistencies in the S(Z) relationship. These inconsistencies are probably due to problems in generating the so-called *in situ* surfaces.

Indeed, unlike the water level which is a rather reliable *in situ* measurement from stations and the volume which is derived from a model that probably takes into account the bathymetry of the reservoirs, the *in situ* surface is a theoretical data from laws that do not take into account the morphology of the reservoir.

These observations should be considered when analyzing the time series output from STOCKWATER.

5.2.3 Time series analysis S(t), V(t), Tr(t)

As described in the analysis performed in Phase I, an analysis of time series from the Optical and Radar sliding windows was performed. Indeed, these types of surface extraction proved to be the most relevant and robust in this study.

Tables with structures identical to those presented in phase I were generated from the notebooks.

5.2.3.1 Brasil

TABLE 19 : MEDIAN - BRASIL

	ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
11	6330098332	Acarape do Meio	nan	6.393449	17.368976	29.287535	42.425615	13.820472	20.849680
8	6330068933	Araras	85.107299	7.425078	14.379832	89.090553	89.889159	26.471635	25.756121
10	6330095223	Arrojado Lisboa	30.563766	73.259262	33.765447	99.952994	73.117328	99.675989	56.331136
1	6320007172	Barragem do Batalhao	nan	28.504105	33.115264	73.423341	66.046339	46.176698	43.688167
6	6330000001	Bonito	nan	88.582267	95.440060	99.958559	99.998962	99.067503	99.800938
3	6320008892	Carnaubal	33.525627	24.993743	13.290283	78.574539	81.438330	58.218293	61.064527
9	6330070272	Farias de Sousa	nan	34.161818	31.684221	57.259142	48.901004	35.276432	27.648184
2	6320007492	Jaburu I	16.779143	9.249037	4.677136	9.575751	4.674653	2.719546	8.599022
4	6320009412	Jaburu II	nan	90.416946	96.352461	43.063579	35.727004	16.312191	15.331991
12	6330098612	Pacoti	15.995939	13.863894	23.613350	13.295391	15.563633	5.176296	12.096005
5	6320015192	Realejo	27.101354	17.890491	21.002019	9.971300	34.859547	6.591743	17.665416
0	6320007102	Sao Jose III	nan	18.810285	22.489237	14.683855	38.154769	4.402261	18.292245
7	6330068463	Serrote	21.332547	11.802185	17.077881	28.457980	33.695171	11.067083	14.664458

TABLE 20 : SUMMARY OF MEDIAN RESULTS - NUMBER OF TANKS PER CLASS - BRAZIL

	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
<= 10%	0	3	1	2	1	4	1
> 10 & <=25%	3	6	8	2	1	3	6
> 25%	4	4	4	9	11	6	6

5.2.3.2 Results analysis

The Bonito reservoir was not included in the analysis, because the in-situ data identified for this reservoir did not correspond to the morphology of the reservoir (surface * 10), the same is true for the Castanhau reservoir for which the process did not work.

If we rely on the previous tables (see Table 23 and Table 24) which illustrate the potential anomalies in the results from DEM4WATER and SURWATER post-process, we can see :

- ✓ That for the surface, only half of the reservoirs have median deviations with the *in situ* data less than 25%,
- ✓ Only two tanks have median deviations of less than 25% for volumes,
- ✓ And that 9 reservoirs out of 14 have fill rates with a median deviation of less than 25% with the *in situ* data.

Unfortunately, we cannot blindly rely on this table, because, as we pointed out in the previous section (see 5.2.2.2.2.) we notice inconsistencies in the behavior of the S(Z) relationship of the *in situ* data. These inconsistencies question the quality of the so-called *in situ* Surfaces.

This is for example the case for reservoirs 6330068933 - Araras, 6320008892 - Carnaubal and 6320009412 - Jaburu II for which, we notice that the so-called *in situ* surfaces sometimes exceed the polygon of counts, considered as the max-extent of the reservoir.

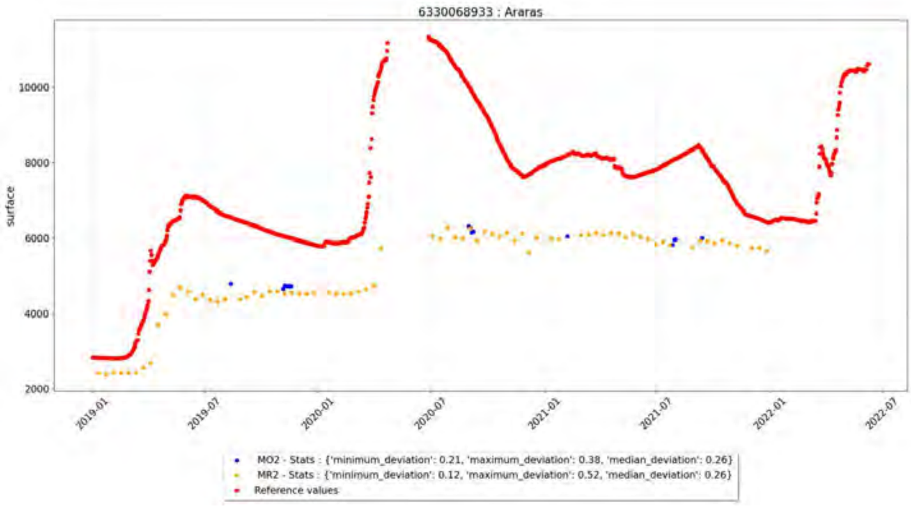


Figure 62 : Time series of MO2, MR2 and INSITU surfaces - Arraras

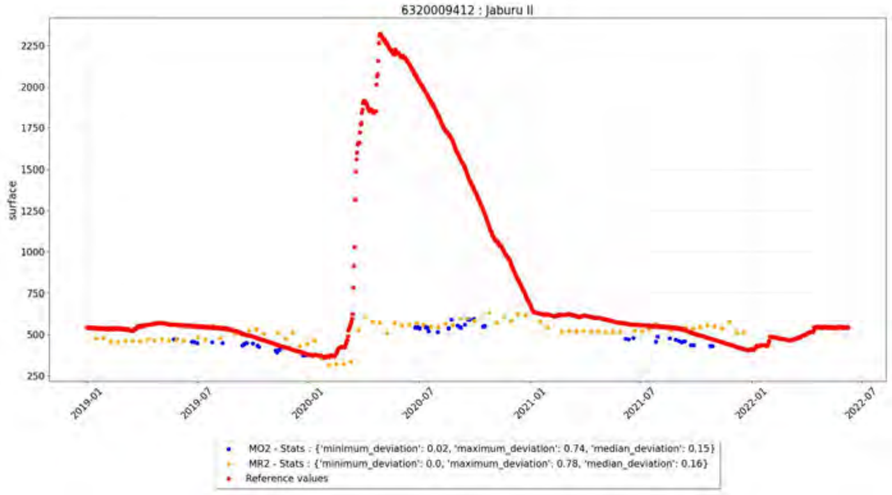


Figure 63 : Time series of M02, MR2 and INSITU surfaces - Jaburu II

In addition, other inconsistencies in the dynamics of the surface time series are noted. This is the case for reservoirs 6330068463 - Edson Queiroz, 6330068933 - Araras, 6330098332 - Acarape do Meio, 6320008892 - Carnaubal, 6320009412 - Jaburu II, 6330070272 - Farias de Sousa and 6330095223 - Banabuiu. Indeed, we notice some consistency in the V(Z) relations of these reservoirs, but inconsistencies in their S(Z) relations. This suggests a poor estimate of their so-called *in situ* surfaces. It is therefore difficult to assess the quality of Surfwater post-process surfaces for these reservoirs, as we cannot rely on their so-called *in situ* surfaces. A comparison with extractions from another surface extraction chain, such as ExtractEO, would have been more appropriate for the validation of these SURWATER surfaces.

Regarding the estimation of the quality of the volumes, we notice globally very important differences with the *in situ* data. These discrepancies may be partly due to poor quality *in situ* data, but, as presented in Phase I, the absolute volume tends to be underestimated by DEM4WATER, compared to the *in situ* volume data. It is more appropriate to rely on the fill rate, for which correct discrepancies are observed between the fill rates from SURFWATER post_process and the fill rates from reference data.

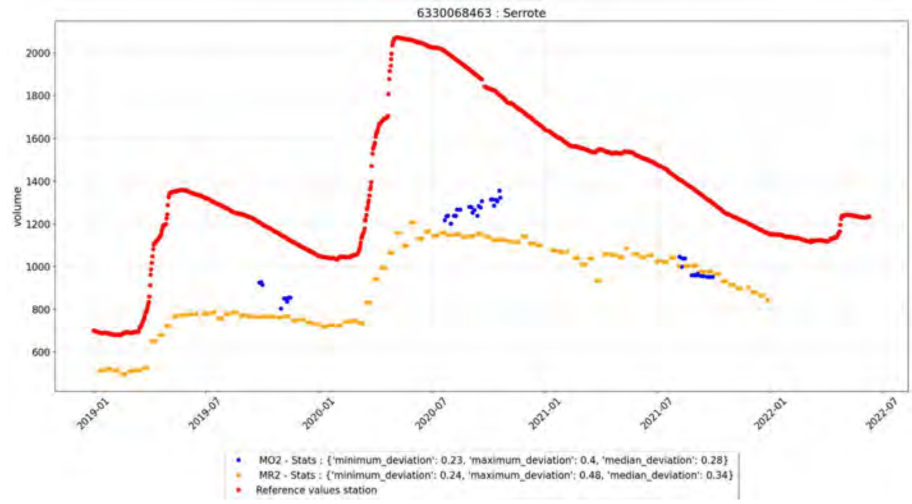


Figure 64 : Volumes temporary series M02, MR2 and INSITU - Serrote

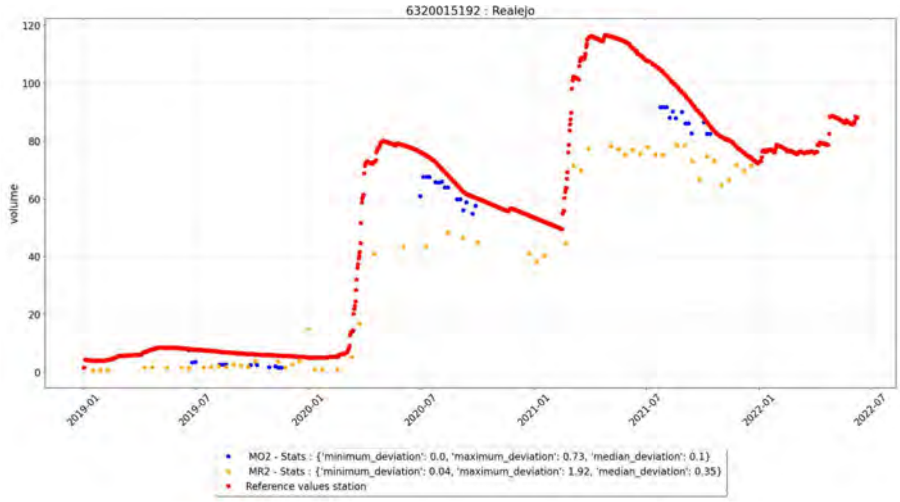


Figure 65 : time series of volumes M02, MR2 and INSITU - REALEJO

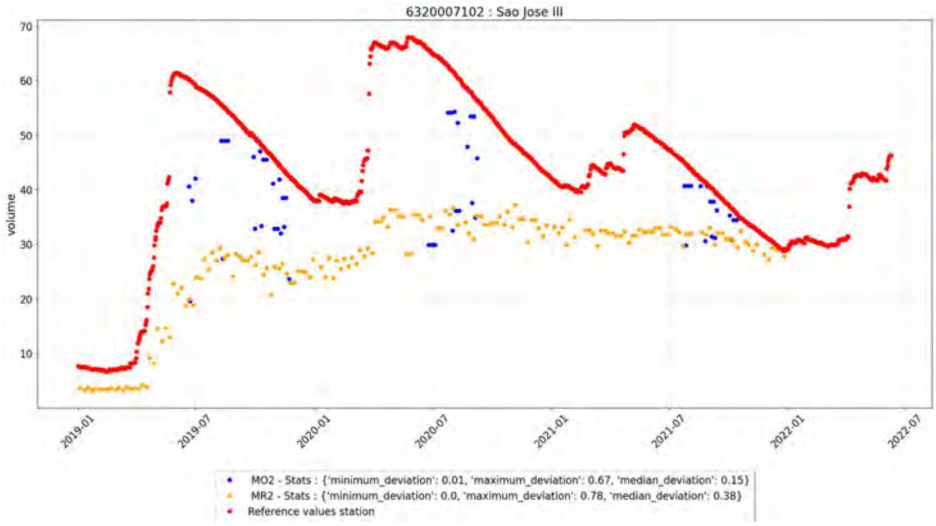


Figure 66 : time series of volumes M02, MR2 and INSITU - SAO JOSE III

As mentioned in Phase 1, these results should be considered with caution because the SURFWATER post-process and reference fill rates are based on their respective maximum volumes. They are therefore data relative to a maximum filling volume that must be observed within the observation window (normally 3 years are sufficient to observe these filling periods).

5.3 Production

5.3.1 Generation of water masks

The production campaigns of the Water Surface masks for the Andalusia, Occitania, India, Tunisia, Burkina Faso and Laos super sites have been carried out on the CNES HAL cluster for the period from 2018 to 2020, and for the Brazil super site from 2019 to 2021.

The following tiles have been processed:

- ✓ Andalusia: T29SPB, T30STF, T30STG, T30SUF, T30SUG, T30SUH, T30SVF, T30SVH, T30SWG, T30SWH,
- ✓ Occitania: T30TYP, T31TCH, T31TCJ, T31TDH, T31TDJ, T31TDK, T31TEJ,
- ✓ India: T43QHA, T43QHB, T43QHU, T43QHV, T44QKD, T44QKE, T44QKF, T44QKG, T44QLD, T44QLE, T44QLF, T44QME,
- ✓ Tunisia: T32SPF, T32SNF, T32SNE,
- ✓ Burkina Faso: T30PXU, T30PXV, T30PYT, T30PYU, T30PZT, T30PUT, T31PBN,
- ✓ Laos: T48QVE, T48QVF, T48QWE, T48QWF,
- ✓ Brazil: T24MTA, T24MUA, T24MTV, T24MUV, T24MWA, T24MVV, T24MWU.

The Surfwater mask production campaigns revealed limitations in the extraction of optical water masks that could be improved by using the Random Forest method instead of the Clustering Average method initially selected. The optical mode campaigns of the 3 super sites of Phase I, Andalusia, Occitania, India, were regenerated with this new parameterization in order to qualitatively improve the classification. Then the new optical data sets were merged with the radar data produced during Phase I.

The Surfwater campaigns of the new super sites Tunisia, Burkina Faso, Laos were carried out over the period 2018-2020 and over the period 2019-2021 for Brazil.

As part of the extension of Phase II, changes on the parameterization used have been made to remove the effects of sunglint observed on the water masks. As such, the Tunisia, Burkina Faso and Brazil sites were selected to be reprocessed. The optical datasets were regenerated over the 3-year period. The occurrences were also recalculated on each site. These new optical data sets were merged with the previous radar data sets.

One tile was problematic over Burkina Faso; the T30PXU tile had no Pekel data over this area, so no Surfwater product could be generated. The data produced by the Waterdetect chain over this area were provided by CNES. These water masks were integrated into the 3 existing datasets 2018, 2019 and 2020, with generation of monthly accumulations, calculation of occurrences over 3 years and regeneration of the global optical wmap.

5.3.2 Computing HSV relationships and comparison

This step consists in generating/improving the estimated HSV models with the Copernicus 30m DTM.

As described in Phase I, the methodological limitations of the DEM4water chain were reached on some reservoirs, with for example a bad estimation of the foot of the dam, the cut-off line partially looped, water crossing by the cut-off line, case of multiple spillways with several structures on the same water reservoir, etc.

The possibility of using manually filled files was implemented in order to correct some of these erroneous data, whether it be at the level of the location of the foot of the dams or the cut-off line.

Initially erroneous models were produced, while others were qualitatively improved on the Andalusia, Occitania and India sites. The Nam Theun 2 dam at the Laos super site required an increase in resources due to the unusual surface area of this reservoir, however the model result for this atypical reservoir is inconclusive. The Castanhao Reservoir in Brazil has posed similar problems due to its huge size.

TABLE 21 : number of models generated according to the studied reservoirs, by Super Site

Super site	Reservoirs	Automatically Generated Models	Correct Models	Models to be manually corrected
Andalusia	37	37	25	12
Occitania	16	16	6	10
India	23	16	9	8
Brasil	14	13	10	3
Tunisia	10	10	9	1
Burkina Faso	10	10	9	1
Laos	1	0	0	1

The following models are considered correctly generated:

- ✓ Andalusia: Arcos de la Frontera, Beznar, Bornos, Celemin, Charco Redondo, Cuevas de Almanzora, Guadalcacin 2, Guadalteba, Guadarranque, Iznajar, La Brena II, La Concepcion, La Vinuela, Limonero, Los Bermejales, Los Hurones, Piedras, Puebla de Cazalla, Puente Nuevo, Rules, San Rafael de Navallana, Tranco de Beas, Vadomojon, Yeguas, Zahara.
- ✓ Occitania: Agly, Astarac, Avene, Cammazes, La Gimone, Puyvalador.
- ✓ India: Dindi, Himayat Sagar, Kaddam, Musi, Nagarjuna, Pakhal, Palair, Pocharam, Sathnala.
- ✓ Brazil: Acarape do Meio, Araras, Arrojado Lisboa, Barragem do Balahao, Carnaubal, Farias de Sousa, Jaburu I, Jaburu II, Realejo, Serrote.
- ✓ Tunisia: Abid, Bezirk, Chiba, El hamma, Er rmal, Kamech, Masri, Sidi Saad, Sidi Saleem.
- ✓ Burkina Faso: Bagre, Kompiega, Lake Bam, Nagbangre.

Not benefiting from all the software improvements made during Phase II, some Phase I tanks (Andalusia, Occitania, India) require further manual corrections.

As far as the Burkina Faso sites are concerned, the Namsuguia, Samendeni, Tanvi 1 and Tanvi 2 reservoirs, whose models are listed below, have the particularity of being atypical due to their shallow depths, with the result that errors in the models are highly probable.

The Ziga reservoir has a weir problem with a jump at the end, which will require a correction of the cutline.

In Tunisia, the Lebna reservoir has a location error with the point located in the water.

Very large reservoirs like Castanhao in Brazil or Nam Theun 2 in Laos will require the development of a specific approach.

Sites with models to be manually improved (whether a more precise valley-cutline or Base of Dam geolocation):

- ✓ Andalusia: Almodovar, Arenoso, Barbate, Casasola, Conde de Guadalhorce, Giribaile, Guadalen, Guadalhorce, Jose Toran, Los Meloranes.
- ✓ Occitania: Laparan, Matemale, Montbel, Pareloup, Pla de soulcem, Saint Géraud, Saint Peyres, Salagou, Villeneuve de la Raho, Vinca.
- ✓ India : Buggavagu, Kinnersanni, Lakhnavaram, Lankasagar, Large Tank Bayya, Lower Manair, Nizam Sagar, Osram Sagar, Pulichinthala, Ramappa, Sriram Sagar, Upper Manair, Wyra.
- ✓ Brazil: Castanhao, Bonito, Pacoti.
- ✓ Tunisia: Lebna.
- ✓ Laos: Nam Theun 2.
- ✓ Burkina Faso: Ziga, Namsuguia, Samendeni, Tanvi 1, Tanvi 2.

5.3.3 Generation of S(t), V(t) and TR(t) series

The Surfwater_Postprocess chain was the subject of evolutions with the implementation of a reconstruction mode on the one hand, and on the other hand the improvement of the data processing speed. Moreover, new models, previously in error or under improvement during Phase I, were generated during Phase II.

The estimated time series of the relevant sites from Phase I were regenerated without reconstruction for the MO1/MR1/MO2/MR2 counting methods over the 2018-2020 period. In addition, these time series datasets were produced 2 other times in parallel with different reconstruction modes: at 1% and then at 2% for all dams of the Andalusia and Occitania super sites. The time series of the super sites Tunisia and Burkina Faso were generated twice each, without reconstruction and with reconstruction at 5%.

During Phase II, the optical time series of the Tunisia and Burkina Faso super sites were regenerated without reconstruction and with 2% reconstruction. The optical and radar time series of Brazil were regenerated without reconstruction and with 2% reconstruction.

Since the sites studied in Brazil are subject to strong clouding, the time series were again produced with the nodata parameter set to 0.9 (and 2% reconstruction).

As a last step, these time series were pre-filtered by a tool developed in the Stockwater project before being ingested into the recovery site. The data with nodata were removed, which gives a lower date frequency but much more relevant for the trends. As for the reference *in situ* time series, they are reprocessed to provide a cvs file per dam over a defined period.

5.4 Analysis and Validation of the results of the complementary phase

5.4.1 Absolute volume

The medians of all relative estimation errors are 15% for the MO2 optic and 22% for the MR2 radar. Looking at the average of the median of relative error of absolute volume per reservoir, we find 29% for MO2 and 33% for MR2. Of the 34 Andalusian reservoirs, 14 have absolute volumes estimated with a deviation of less than 10% from the reference, 20 with a deviation greater than 25%, 4 of which are more than 50%, for MO2.

In the case of Occitania, of the 16 reservoirs, 2 have a deviation of less than 10%, 6 have a deviation between 10 and 25%, the rest have a deviation greater than 25%. For the India Super Site, finally the process could only be completed for 3 reservoirs out of the 6 initial ones, and the discrepancies are always higher than 25%, with figures such as to indicate a problem at certain levels (*in situ* data, etc.). For Brazil, only 2 reservoirs have deviations below 10%, 2 between 10 and 25% and the remaining 9 have deviations of more than 25% from the reference. If we use more restrictive metrics, such as the 75 and 90 quantiles, we see a very large increase in deviations.

Several sources of error have been identified, related to poor estimation of the V(S) law in some cases where there are large discrepancies between Surfwater surfaces (optically on the parameterization 1, very erroneous, or biased detection problems by radar [Peña-Luque S. et al., 2021]) and reference surfaces (*in situ* data of questionable quality on some dams). Note that these errors sometimes offset each other. Poor results for the estimation of the V(S) law and the extraction of the surfaces by Surfwater can therefore generate volumes that seem correct.

The main source of error is an "ill" estimation of the V(S), which is detected by regarding the output elevation contourlines from DEM4Water processing. Such models may be manually corrected by replacement of the automatically detected cutline or base of dam geolocation by a manual one. This approach has been used in Phase 2 of the project, but not in Phase 1 (Andalousia, Occitania and India). When observing only models "without issues" on the automatic process, the Absolute Volume estimation is improved:

Median - absolute relative error %

	All Bathymetries		After Revision	
	Volume Optical	Volume Radar	Volume Optical	Volume Radar
Andalousie	14	18	12	16
Occitanie	29	47	18	38
Brazil	40	42	36	42
	Analysis on 66 sites		Analysis on 40 sites	

5.4.2 Fill rate

The results are positive. Regarding the totality of measurements, the Fill Rate - FR(t) estimates obtain a 75% quantile of measurements of 12.8% for optical measurements (MO2) and 17.3% for radar measurements (MR2). These results have a similar quality level to the fill rate monitoring of the previous CNES/MTE study over Occitania, which used a semi-manual V(S) law estimation from DTMs. This indicates that the growth/decrease trends of the total volume stored are well detected by the time series S(t) by Surfwater, and that the errors made on the V(S) estimation do not disturb too much the trend monitoring FR(t).

This indicator remains interesting for water managers, who generally know the total volumes that can be mobilized. On the other hand, this does not solve the need to know the real volume mobilized in basins that are not instrumented or not recorded in dam databases (GranD, GeoDAR, etc).

The following table summarize Filling Rate results for each instrumented region :

TABLE 22. FILLING RATE RELATIVE ERROR (%)

	Volume Rate Optical - Median	Volume Rate Radar - Median	Volume Rate Optical - Quantile 75%	Volume Rate Radar - - Quantile 75%
Andalousie	4	5	7	9
Occitanie	5	6	8	12
Brazil	11	12	27	23

Inde	27	25	45	39

5.5 Valorisation

In order to restitute these results, a web portal has been set up to allow consultation of the time series of the dams studied. Several proposals were made for the hosting of the valorization site. The following domain name was chosen: sco-stockwater.org

The code is centralized on a public Github account of CS: https://github.com/CS-SI/sco_stockwater

Description of the features

The valorization site is carto-centric, allowing to visualize the data of one or several reservoirs simultaneously.

It consists of 3 parts:

- ✓ the "header" part with the SCO logo, the name of the site and an icon that allows to change the theme,
- ✓ the "action" part, which is divided into two parts, the upper one that shows the selected tanks and the lower one that allows you to interact and update the requests,
- ✓ the "major" part, being the map.

Figure 67 : details of the home page.

By clicking on a reservoir, the major part is divided in two, the map is displayed on the upper part and the graph on the lower part. Three types of observation are available: optical, radar and reference.

The upper left part, framed in red in the figure below, shows the reservoirs selected on the map. This one is then displayed in another color (light blue) to distinguish it from the other reservoirs on the map.

The figure shows the time series of the filling rate of the Guadalteba reservoir in Andalusia, over the 3-year period from 2018 to 2020.

The estimated data from the radar and optical modes are displayed with the insitu reference.

In terms of product, the section "Observation depth", allows you to select the choice :

- " 1 day ": the observation of the day corresponding to the Surfwater single product,
- " 10 days " : the 10 days synthesis corresponding to the Surfwater multiple product.

Figure 68 : data displayed after selection on a reservoir

For a selected point on the curve, the following information is available concerning this date: date, name of the reservoir, type of observation, volume (in hm3).

Concerning the metrics to be visualized, a drop-down menu " Attributes " allows the selection of the surface, the volume or the filling rate.

It is also possible to have access to general information such as the identifier of the dam, the type of use, the nearby city, the coordinates of the dam.

The time series data can be downloaded as a csv file (icon " c " next to the dam name in the image below).

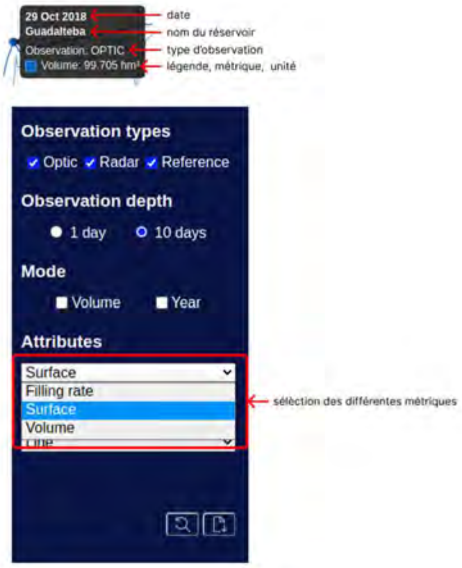


FIGURE 69 : item selection features. Right part: details of the item when adding a tank. Left part: anatomy of the graph tooltip and the selection of the different metrics.

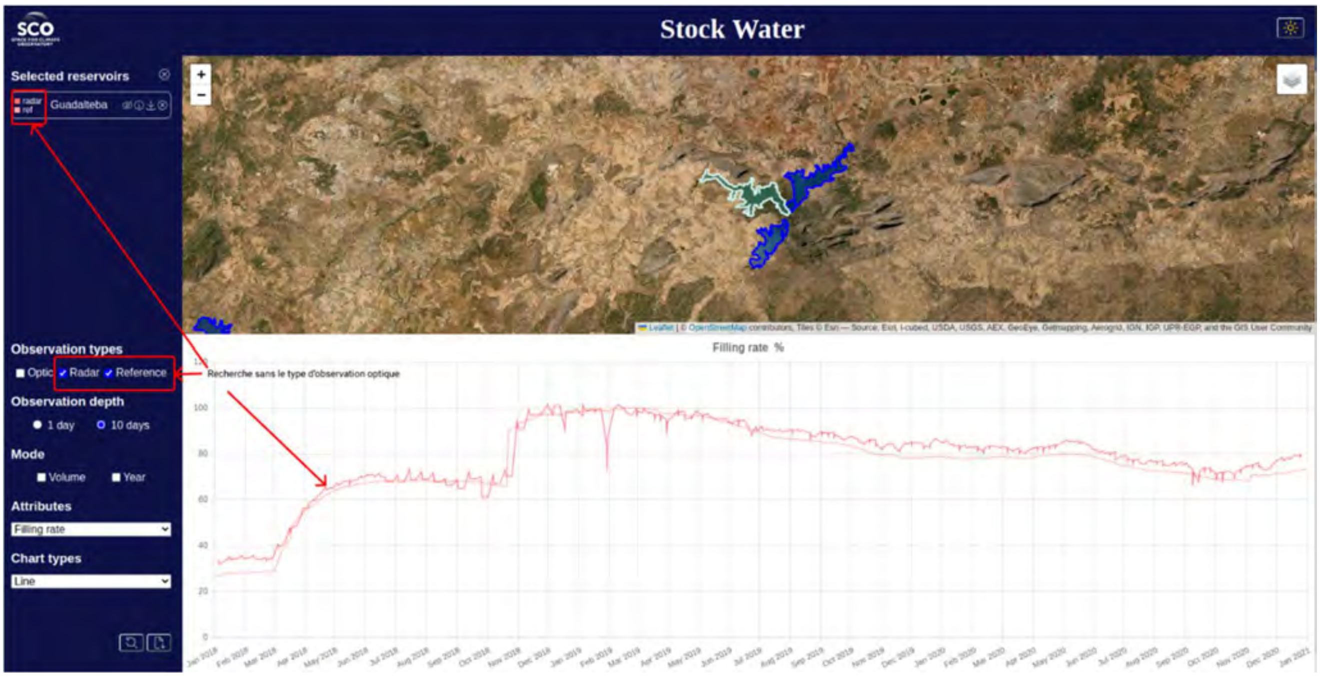


Figure 70 : example of a request by unchecking an observation type.

It is possible to visualize several dams at the same time, for example in the figure below 3 dams are presented: Guadalteba (in red), Guadalhorce (in green) and Zahara (in blue).

The time series data of the filling rate are in optical and radar mode for each dam, over a period of 3 years corresponding to the study, from 2018 to 2020.

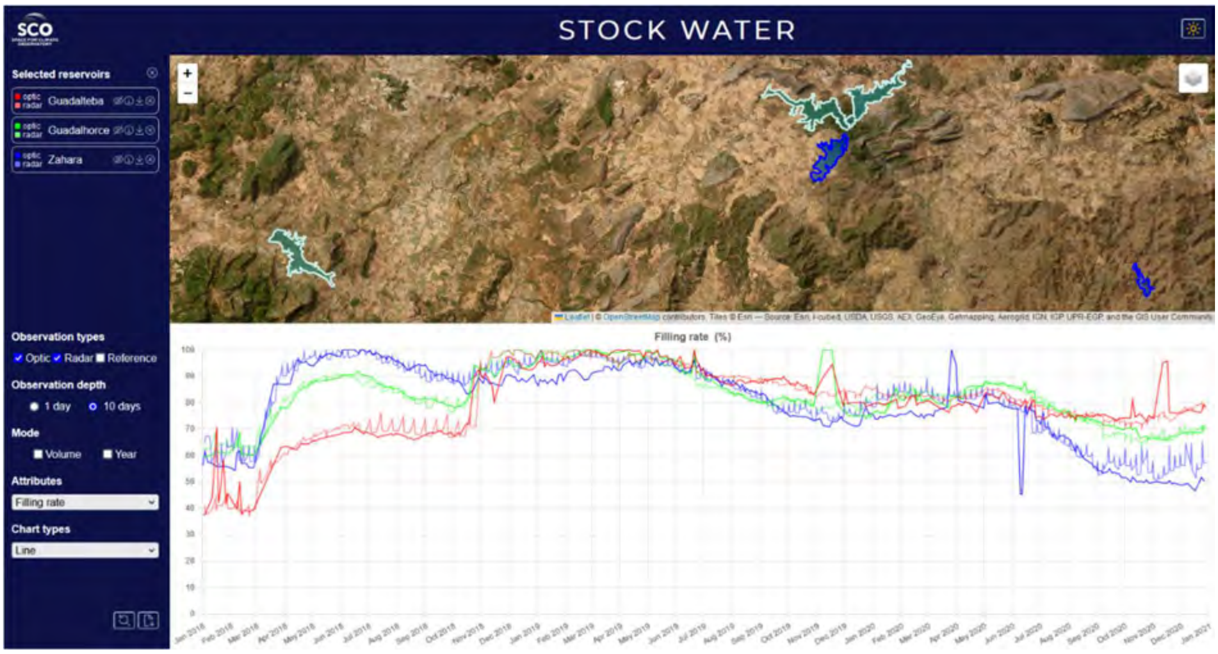


Figure 71 : example of a request on several reservoirs.

It is possible to focus on a particular period, the days are then displayed on the scale instead of the months initially.

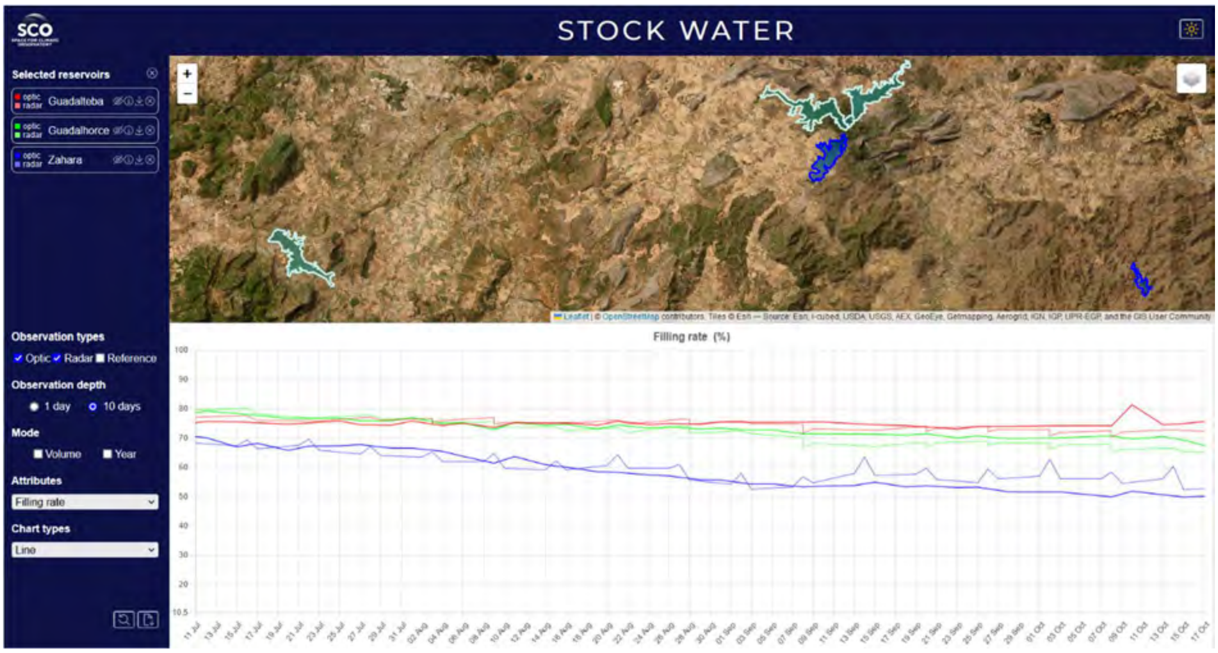


Figure 72 : example of a zoomed-in part of the graph.

Among the functionalities, the "Volume" and "Year" modes allow a specific visualization of the data.

The " Volume " mode calculates the sum of the volumes of the selected dams. The tooltip displays, for the day clicked on the curve, the total volumes in hm3 as well as the value of the individual volumes of each dam, as shown in the following figure.

The reference in-situ volume time series is displayed in light blue, the optical volume time series in dark blue, then the radar time series.

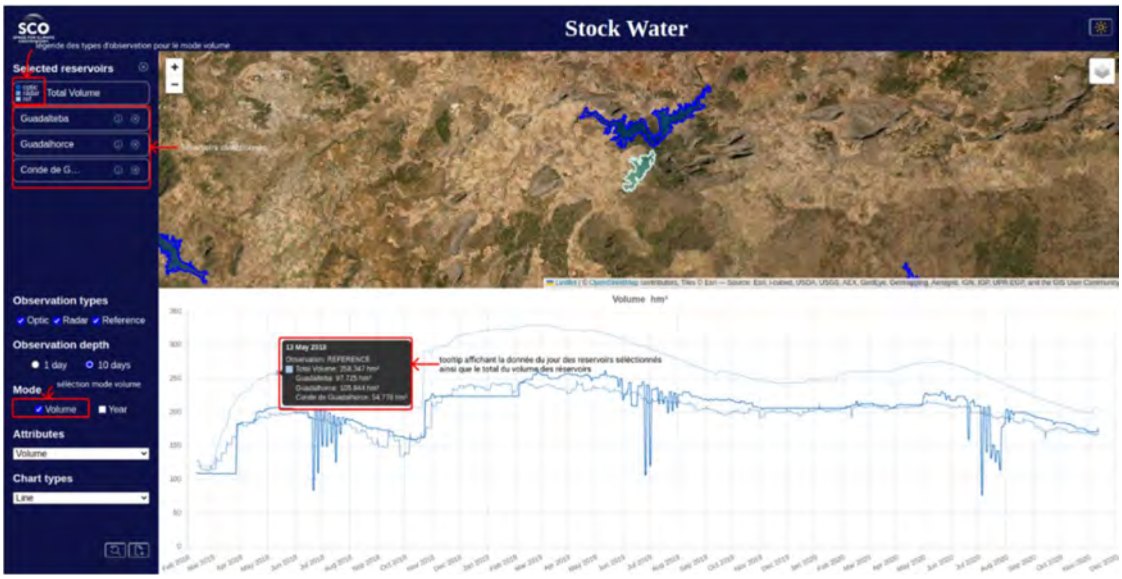


Figure 73 : details of the "Volume" mode.

The " Year " mode allows to compare on a 12-months scale (from January to December), the data of a same reservoir over several years. This mode of visualization makes it easier to compare the interannual evolution of a reservoir.

For example, here are the time series of the filling rate of the José Toran reservoir in Andalusia for the years 2018 (in red), 2019 (in orange) and 2020 (in blue).

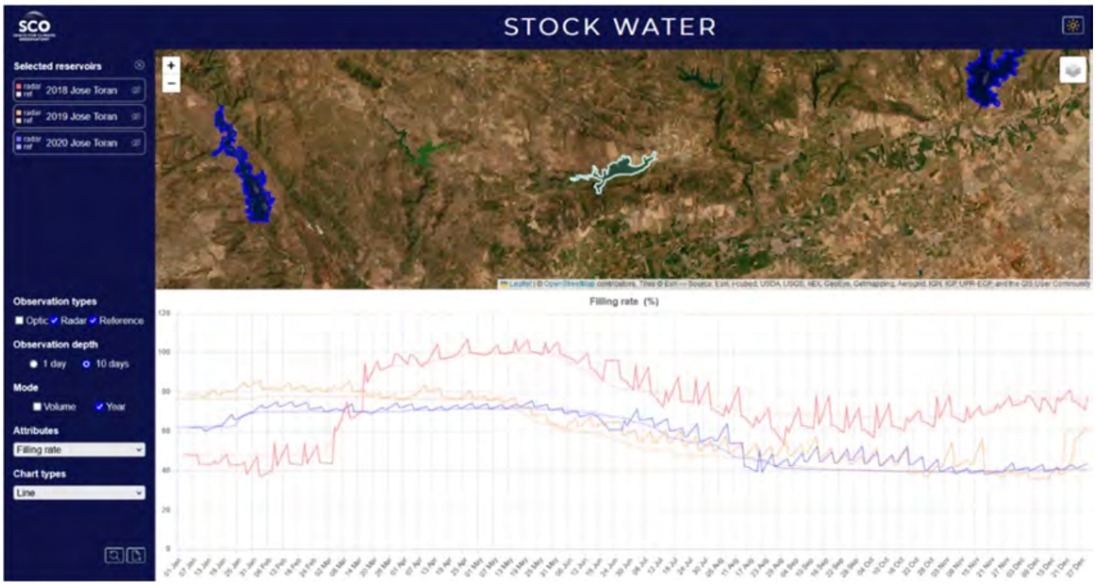


Figure 74 : details of the "Year" mode.

For information, there is the possibility to visualize the data via the drop-down menu "Chart types" either according to the "Line" style, as it has been presented on the previous examples, or with points according to the "Scatter" style.

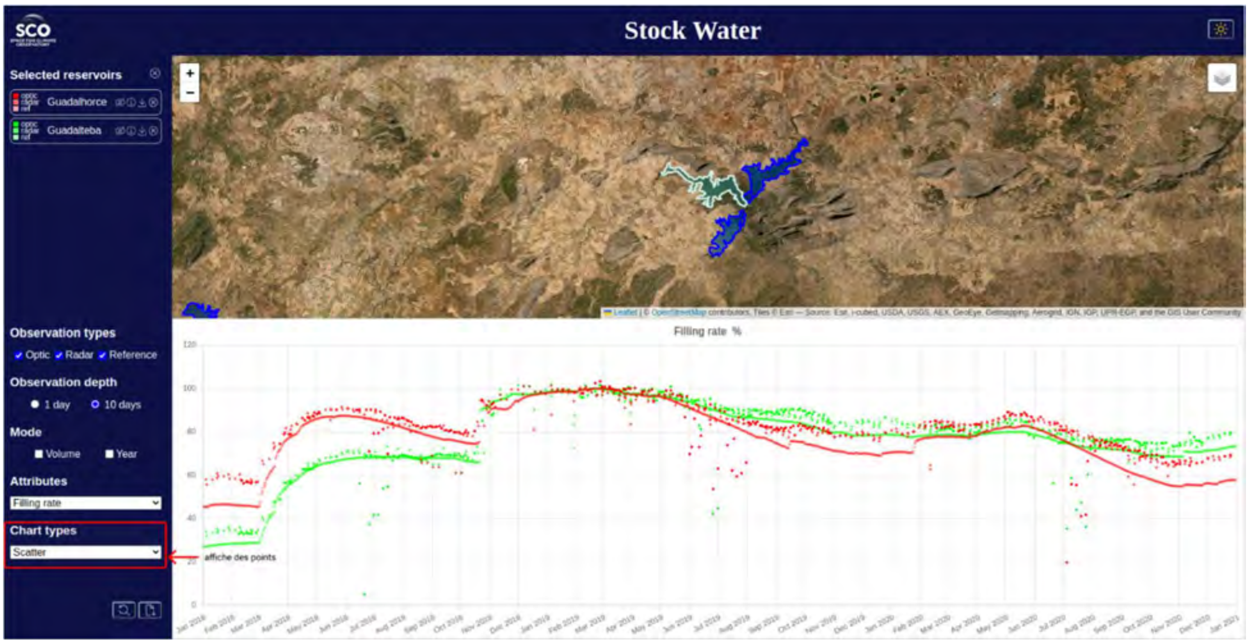


Figure 75 : example of a data display with the "Scatter" style.

6 Robustification of demonstrator and dashboard

As an extension of the activities of phases I and II, it was decided to improve the operationality of the tool in a complementary phase. The scope of the activities was established as follows:

- ✓ Finalization of data selection, validation, and preparation activities,
- ✓ Data selection and preparation activities according to the new perimeter defined by this amendment (CS GROUP - France),
- ✓ Integration of new corrections to the "Surfwater", "Surfwater-Postprocess" and "DEM4water" software (CS GROUP - France),
- ✓ Generation of new time series on all the sites of the perimeter defined within the framework of the present amendment (CS GROUP - France),
- ✓ Analysis and validation of results (SERTIT),
- ✓ Presentation to users (CS GROUP - France and SERTIT).

6.1 Improvements to the processing chain

In order to improve the quality of the results produced during Phases I and II on various aspects, evolutions were carried out on the chains Surfwater, DEM4water and Surfwater_Postprocess. A new production phase was carried out based on these software versions.

6.1.1 Surfwater

In the same perspective as in the previous phases, the productions were again generated enriched with the previously mentioned evolutions, with also the correction of the nodata borders and the update of the Pekel occurrence data.

6.1.2 DEM4water

The new version of the DEM4water chain integrates the following evolutions:

- ✓ preservation of contour lines data if they are already present in order to save processing time,
- ✓ modification of the management of the dam foot elevation in order to adapt to the manual correction mode,
- ✓ possibility to launch the calculation of S(Z_i) from the contour lines with manually defined information,
- ✓ implementation of a performance report.

6.1.3 Surfwater_Postprocess

The Surfwater_Postprocess chain has a new reconstruction mode. Its use and the associated parameterization have been chosen in agreement with the CNES. In parallel, other evolutions are considered, with a standardization of the use of the name of the reservoir and actions concerning the filtering of the time series.

6.1.4 Orchestrator

The Stockwater orchestrator was updated to reflect the new calls with respect to the DEM4water and Surfwater_Postprocess processes, in their modified versions.

6.2 Modification of the database

During the preliminary phase, about 50 reservoirs were selected to be as heterogeneous and representative as possible in terms of surface area, topographic configuration, etc.... and for which the input data are already available (availability of several DTMs, validation parameters, etc.). A reference database describing these sites was set up during Phase I. In a second selection phase, 50 additional reservoirs were identified and integrated into the reference database. At the end of the selection of the test sites, a database is available.

The format of this geo-referenced database is compatible with the format used in DEM4water to describe the dams, as well as the associated (estimated and reference) models. This ensures homogeneity throughout the method, as well as an efficient exploitation during the model comparison and validation phase.

6.3 Generation of new time series

6.3.1 Finalization of HSV calculations

From results obtained during Phase II, some models were improved by considering corrective files provided by the CNES, concerning the cut-off lines and/or other information. In particular, the possibility of improving the method by calculating the watershed to limit the surface of the contours was studied. The sites concerned cover the following countries:

- ✓ Burkina Faso (9 dams),
- ✓ Tunisia (10 dams),
- ✓ Brazil (14 dams).

6.3.2 Restarting SURFWATER calculations

The re-generation of Surfwater masks for the optical mode was performed on the Brazil super-site with the following parameterization:

- ✓ Selection of 14 dams out of the 67 available with *in situ* data,
- ✓ Deactivation of the snow detection filter for the 7 tiles,
- ✓ Period 2019-2021.

The new optical database will be merged with the previous Phase II radar database.The optical occurrences were regenerated:

- ✓ Annually on each of the 7 tiles, for the years 2019, 2020 and 2021,
- ✓ Over the period 2019-2021.

6.3.3 Restarting SURFWATER_Postprocess calculations

The production of the Surfwater-Postprocess chain was performed with the optical occurrence map reconstruction improvements. The regeneration of the time series of surface S(t), volume V(t) and filling rate Tr(t) concerns the super-sites:

- ✓ Andalusia,
- ✓ Burkina Faso,
- ✓ Tunisia,
- ✓ Brazil.

The time series of the 4 super-sites (Andalusia, Burkina Faso, Tunisia, Brazil) were generated for the optical :

- ✓ In "with reconstruction" mode for a value of 2%,
- ✓ In "no reconstruction" mode.

The data "without reconstruction" were not evaluated.

Time series were not regenerated for radar data.

The super sites Occitania, India, Laos having already the "without reconstruction" dataset generated, they have not been subject to any reprocessing.

6.3.4 Formatting the new time series

The task consists in filtering and formatting the optical datasets "with reconstruction" 2% of the new time series S(t), V(t) and Tr(t), with the objective of integrating them into the STOCKWATER dashboard, for the super sites:

- ✓ Andalusia,
- ✓ Burkina Faso,
- ✓ Tunisia,
- ✓ Brazil.

The new time series data sets have been archived and delivered.

6.4 Valorization

The initial version of the dashboard developed during Phase II was modified and robust according to the evolution of products and the needs expressed by CNES. This task also includes the final update of the data with their required format.

6.4.1 Dashboard Robustification

Initially, a code review and a "business need" review were carried out, with the aim of highlighting possible areas for improvement. According to the elements returned by this analysis, the effort focused on the main components in terms of code refactoring, stability, and performance. In this context, a better management of the state and data was implemented to fluidify the sequences and to correct the bugs of transitions. In terms of interface and navigability, the effort was mainly focused on

- ✓ The display of the environment data,
- ✓ Display ergonomics, in particular cognitive ergonomics, and business relevance with regard to product information,
- ✓ The improvement of the volume/year mode,
- ✓ Navigation from drop-down lists,
- ✓ Fixing bugs and conflicts in the display of products and captions,
- ✓ Fixed bugs in the selection/deselection of products in "year" mode,
- ✓ Improved map navigation and readability of information on a map background.

An analysis of the performance and stability of the hosting solutions was also carried out. The different budgets associated with the solutions will be evaluated in the perspective of an operational solution. The final version is identified as v1.0.0.

6.4.2 Data integration and final update

The task is to update the database feeding the STOCKWATER dashboard with the new optical time series datasets delivered.

7 Conclusions and Perspectives

7.1 Summary of the study

This methodology was based on three steps necessary to monitor a site:

- ✓ Determination of the water surface of the site at each observation performed by the Sentinel-1 and Sentinel-2 satellites,
- ✓ Modeling of the HSV relationship based on the analysis of Copernicus30, SRTM30 and Alos DEMs,
- ✓ Generation of time series S(t), V(t) and TR(t).

To carry out this first phase, four tasks were carried out in parallel or in succession, depending on the case. They are:

- ✓ The development of the demonstrator,
- ✓ The setting up of the database,
- ✓ The production,
- ✓ Analysis and validation of the results.

The development of the demonstrator also required the implementation of improvements on the Surfwater, DEM4water and Surfwater-Postprocess chains to adapt to the new input data and to propose functionalities necessary to the volume of production, to the variability of the data, but also to the analysis of results.

The implementation of the database includes the selection of the studied reservoirs, the definition and the filling of a fairly complete attribute table by study region and by reservoir, the recovery, analysis and formatting of the *in situ* H-V data, and even their corrections.

For this Phase I, it was decided to focus the study on three Super Sites, Occitania, because of the CNES experience in the area, Andalusia, and India, which seemed to be very mature in terms of data accessibility.

The selection of the reservoirs and the constitution of the database followed the recommendations of the CNES, and the elements found in the CS GROUP and SERTIT proposal. These steps are based on the exploitation of the SWOT, GOOD2, GRanD, GEODAR (GRanD + GOOD databases) and INDIAN Water Central Commission databases.

Once this information was collected, the attribute table was built, and many exchanges were necessary to arrive at the final version in terms of file type, format / order of fields, ... Thus, for each Super site, at least 6 versions were produced.

The work then focused on the recovery, formatting and analysis of the *in situ* data, which were to be made available via CNES by the end users. It turned out, in a certain number of cases, that these data which were to be used as reference, were delivered with different structures/formats, requiring pre-processing which it is not always possible to automate.

The last step of the data formatting concerned the generation of in situ laws. For this, several evolutions of the CNES tool "timeseries_tool.py" have been made, either by CNES or SERTIT. If the tool can be considered as usable, improvements are still necessary for further steps and operational version.

The implementation of the database was completed with the evaluation of the DTMs. Within the DTMs identified as potential inputs, one or two DTMs had to be qualified as unique reference data, in order to optimize the production and analysis phases and to guarantee a minimum of homogeneity and interoperability of the processing. The analysis of the results allowed to classify the DTMs in terms of performance, to retain Copernicus 30 and to reject the ALOS DTM, plagued with decametric errors.

The third task is the production, whose first step is the production of water masks. The production campaigns of the water surface masks for the sites Andalusia, India and Occitania were carried out on the HAL cluster of CNES for the period spanning 2018 to 2020. The difficulty of retrieving optical and radar input data was quickly identified as a major bottleneck of the process, and improvement efforts had to be made, more than previously expected.

The second step of production was the calculation of HSV relations and their comparisons, taking as input DTMs Copernicus 30 and SRTM on the three Super sites, and the ALOS DTM as additional data on Andalusia.

The last step of production corresponds to the generation of time series S(t), V(t) and TR(t), and over a period of three years. The *in situ* time series, available for some sites, were added as input for comparison and evaluation of the estimated results. Several production re-sets were necessary due to successive evolutions of some *in situ* time series, as well as to new implemented functionalities.

The fourth task corresponds to the analysis and validation of the results from the different steps of the process. In a first step, the outputs of reservoir extension, multiple and single Surfwater masks have been analyzed on the Super Site Andalusia, the first site treated. For this, the SURFWATER outputs were compared with those from the SERTIT EXTRACTEO chain, and the so-called reference surfaces generated via H/S/V laws from the *in situ* H and S data. A first set of Surfwater output from the optical chain was analyzed (Config 1 of Surfwater). Faced with unsatisfactory results, CNES proposed an adaptation/modification of the SURFWATER parameters, and a second set was analyzed (only over Andalusia, outside the generic analysis S(t), V(t), TR(t)). The quality of the results is greatly improved, but a number of anomalies are still observed in the optical data.

On the other hand, in almost a third of the cases, we observe a systematic bias between the surfaces derived from the Sentinel-2 data and the references. The analysis of the images and extractions does not suggest that there was a significant omission in the extraction of water surfaces by SURFWATER or ExtractEO. This bias, of 7 to 12%, could be due to a less efficient operation of the "*in situ* " surface modeling for certain reservoirs, or to erroneous *in situ* volume modeling by the managers of these reservoirs. To overcome this problem, it might be interesting to work not in terms of absolute volume, but rather in terms of volume variations, as these 30% of Andalusian reservoirs have unreliable *in situ* data, which could bias the final validations.

To analyze the quality of time series S(t), V(t), Tr(t) from the reports generated by DEM4Water and Surfwater_Postprocess, a tool (notebook) has been implemented to automatically generate a summary table to assess more simply the quality of time series of surface, volume, and filling rate.

7.2 A convincing initial approach

The objective of the SCO STOCKWATER project was to establish the potential of generalization of the method, by extending the processing initially performed in France to other regions of the world. In this context, we had the opportunity to study a panel of reservoirs with very different shapes, in various topographic contexts.

At the operational level, the demonstrator allows to automatically set up a context for each chain and inter-stage. The use is simple and flexible, with the possibility to restart very simply from a selected step if needed.

Processing times have been optimized by parallelizing processes whenever possible for mass processing purposes.

The organization in two stages will have allowed, in first phase, to dimension and to validate the methodology to pass to the generalization of the selected approach. It allowed to highlight a certain number of problems/limitations at different levels of the process, access, and quality of *in situ* data, access to image data, quality of extractions, etc.

From the tools point of view, the concept can be considered as valid. A "Proof of Concept" level demonstrator and related tools have been developed, but additional efforts are needed in terms of method and tool robustification to fully reach the stage of a true demonstrator.

The implementation of the database allows the selection of the studied reservoirs, the definition and the filling of a fairly complete attribute table by study region and by reservoir, the retrieval, the analysis and the formatting of the *in situ* H-V data, and even their corrections. Once this information was collected, the attribute table was built, but it took many exchanges to arrive at the final version in terms of file type, format, and field order.

The Surfwater, DEM4Water and Surfwater_Postprocess strings were both adapted to the new input data provided and functionally improved. As the project progressed, the integration and interoperability of these different components within the demonstrator were perfected.

Good performances were observed on the time series of filling rates, giving a relative variation. However, less good results were observed on absolute volume time series.

Finally, a SCO web portal has been set up, allowing the restitution of these results and access to the monitoring of reservoirs. The interface developed within the framework of the project meets the generic needs of the community gathered around the project. This solution ensures adaptability and portability at low cost to users, considering the upcoming prospect of service platforms for dissemination to users, carried out by the CNES.

The first feedback from the European partners who tested the platform is positive. However, this first tool obeys a logic of genericity, and must be adapted to specific uses and users.

7.3 Guidelines for improvement for an operational version

The complementary phase has enabled the robustification of all the components in a sufficiently stable manner so that end-to-end production is possible with human intervention limited to the recalculations linked to the refinement of the results.

However, it is not yet possible to consider that the pre-operational level has been reached. It is indeed necessary to continue the efforts of improvement, as well at the level of the acquisition and the preparation of the data as of the stability and the genericity of the results. These efforts will have to be translated into software improvements on the three main components Surfwater, DEM4water and Surfwater_Postprocess, as well as on their validation procedures.

7.3.1 Streamlining data handling

From the outset of the study, we were confronted with problems of data acquisition, heterogeneity and compatibility, with anomalies noted at various stages of the process. Significant efforts were made to circumvent the bulk of these problems with some success, but these anomalies made the qualitative assessments more complex, with a verification and preparation process that proved extremely time consuming.

In its current state, STOCKWATER meets the initial need and has the great merit of proposing a first approach to a large-scale water resources management tool that does not yet have an equivalent in Europe.

However, this situation is not compatible with a requirement of availability of data in constrained time and will require in the future to take technical, organizational and standardization measures. These measures are necessary to make the tool easy to use and with relevant results in order to preserve its attractiveness, which remains undeniable in view of the economic, environmental and safety issues.

However, it must be emphasized that this standardization will not be easy to achieve since the target community is global, covering diverse eco-climatic areas, with national or even local uses and problems of data collation and accessibility (even in France, the data from certain reservoirs and dams are not accessible neither to the public, nor the scientific community)

If basic organizational measures can be envisaged, the future system will undoubtedly have to be on one hand, as agnostic as possible, so as to be able to work with alternative input data (images, DTMs, support and reference data), and on the other hand, easily accessible to all users (open source?) so as to allow its use by scientists or users in countries, regions or fields of activity for which the data cannot be disseminated to the community

7.3.1.1 More robust input data

The capabilities of the tool, from a decision support perspective, are limited primarily by the difficulty of obtaining homogeneous and directly usable data. This problem of acquiring and preparing input data is a major bottleneck in the process.

The implementation of a unique file structure has largely solved the problem of the heterogeneity of the sources, but in the current situation a significant number of manual actions are still necessary during the data acquisition and preparation phase.

In any case, considering the particular context of the field in which STOCKWATER will evolve, a preliminary step of control and preparation of the data remains essential. However, this process remains rather long for the moment, and does not allow us to consider strictly an ARD (Analysis Ready Data) approach, and therefore the use of a Datacube (the eodag component is already a first approach of this organization but it does not work efficiently in the current conditions, which reduces its interest). It is therefore essential to consider an automation of the data acquisition and preparation process that is transparent to the user.

7.3.1.2 The issue of support and reference data

The same applies to the supporting data. A second bottleneck in the process is in the formatting and analysis of the *in situ* data. Inconsistencies were sometimes encountered in the *in situ* data, with problems matching volume values to reported elevations, missing data, etc.

This should not be considered abnormal in the context of the study. This was a first approach, using data from different sources or teams, with data sets often made under specific conditions, or according to other needs.

However, the importance of the efforts made to complete, harmonize and format the datasets, as well as to develop tools to "smooth" the quality of the data, is not compatible with an evolution towards an operational stage. It will therefore be necessary in the future to resort to organizational measures to ensure that the data sets provided meet common standards for all contributors.

7.3.1.3 *In situ* laws

The last step of the data formatting concerned the generation of *in situ* laws. For this, several evolutions of the CNES "timeseries_tool.py" tool were carried out, either by CNES or SERTIT. As a result, the *in situ* laws had to be generated a number of times. If the tool works today, improvements are still necessary to ensure a good functioning in the future.

7.3.1.4 DEMs

Although it was relatively easy to converge on generic DTMs within the scope of the study, it is clear that it is not possible to have the same DTM sources, in conditions of quality and homogeneous accuracy, depending on the region or country. This is the case in particular for countries for which the dissemination of elevation data is not free, or subject to significant constraints of dissemination. Also, il is not guaranteed to systematically get data with similar, if not identical, levels of accuracy and quality.

In this case too, it will be necessary to make the system as DTM-agnostic as possible, in order to allow users confronted with these situations to have access to a generic DTM that can be enriched or replaced partially or locally by local data (i.e. not shared with community or third-party). This functionality already exists in other CNES projects, and the evolution of the project will undoubtedly be inspired by it.

7.3.2 Towards an operational chain

In order to really reach a pre-operational step (demonstrator under development for the French government services), the efforts of robustification and adaptation of the various components must be continued, as well as a stabilization and a better accessibility of the associated documentation.

If the hypothesis of a facilitated distribution to the community is confirmed (extended or open source license), it is essential to carry out code audit actions, accompanied by possible corrective actions before any distribution. Steps in this direction are underway at CNES.

The development of the demonstrator required the implementation of improvements on the Surfwater, DEM4water and Surfwater-Postprocess chains to adapt to the new input data and to propose functionalities answering the required production volume, the variability of the data, but also the analysis of results. These efforts will have to be continued according to the evolution of the method and the tool until the realization of the operational version.

Regarding Surfwater, there are still inconsistencies to investigate between the optical and radar on some sites, especially in Brazil. On the other hand, good results are observed on Andalusia and Tunisia. Such a difference in results has already been encountered in other projects (IOTA2 in Canada and in Africa, for example) and can probably be explained by phenomena due to the different eco-climatic environment and the different illumination conditions of the targets. These difficulties can certainly be solved by an approach adapted to the context (as for IOTA2), but it will probably be necessary to accept a certain level of variability in the results compared to the basic method depending on the regions observed.

Concerning the DEM4water chain, in line with the improvements initiated during Phase II, work remains to be done to refine the cut-off lines, the accuracy of the location of the foot of the dam and the adjustment of the law. An automatic model validation tool based on a statistical report would also alleviate the validation task performed by the operator. A process for improving the tool is currently underway at CS GROUP, led by CNES.

In terms of tools, it would be interesting to consider a method of automatic ingestion of new time series on the SCO web platform, as the update is currently done via an operator procedure.

Regarding the operationality, work is in progress on a decision support service demonstrator for the French government, with 340 reservoirs currently studied and a perspective of 500 dams in total. We expect a lot of feedback from this action for the evolution of the future operational tool.

7.3.3 A validation phase to be stabilized

The major work of investigation, analysis and validation undertaken by SERTIT has been very beneficial within the scope of the study, and several sources of difficulties and errors have been identified and corrected thanks to this long-term work. However, such "advanced" approaches cannot be considered for operational use, or even at a pre-operational stage.

The importance of the effort provided, if it is relevant in the context of a Proof of Concept, is perfectly disproportionate in the perspective of an operational version, and one cannot demand a high level of competence from all future operators and users. The analysis and validation process must be easy to implement and accessible to non-expert operators and users.

An effort to stabilize and standardize the analysis and validation approaches, both upstream and downstream of production, must therefore be undertaken. However, it must be emphasized that these efforts are in parallel with other stabilization and organizational actions mentioned in the previous paragraphs.

7.3.4 An architecture and interface adapted to a wider user community

From the beginning of the project, it was decided to use the same technologies for the dashboard and the associated functionalities as those used for the development of the future CNES HYSOPE II/Hydroweb-Next platform, which will be dedicated to the dissemination of data and products for hydrological purposes. This choice ensures the compatibility of the STOCKWATER dashboard with the future platform, or at least its adaptation at lower cost. This should also facilitate the synergy of products with those from other processing chains or data sources (especially SWOT), while opening up, at the same time, prospects for evolution or interoperability of the tool.

Figure 96 below shows what could be the typical configuration of the STOCKWATER tool within the future architecture. This architectural vision is based on similar approaches already implemented or envisaged in the framework of other CNES projects, for example IOTA2.

However, to rebound on some of the considerations discussed in the previous chapters, it seems useful to us to propose a more flexible approach to making the tool available, with a view to reaching a larger community of users.

Indeed, as we have mentioned, the potential user community does not appear to be homogeneous in terms of familiarity with the tools, expertise in handling data, and use of and access to the data. It therefore seems to us necessary to propose a more diversified "commercial offer":

- ✓ An interface accessible via the HYSOPE II platform, for the benefit of operational or non-expert users of the products,
- ✓ As it is currently the case, to propose the manipulation of the data and the tool through a Jupyter notebook interface for the benefit of an expert public. This possibility already exists for hydrological product processing chains and associated functionalities (IOTA2, SWOT),
- ✓ Finally, a "stand alone" version, which can be used independently of the future CNES platform. This version could be deployed by users for the use of their own data or requiring reference or support data whose access is regulated or restricted, or who cannot access the data put online by CNES, ESA or other Agencies. Of course, such an option implies that CNES implements a licensing or software release policy that allows this (actions in this direction are underway). The essential advantage of this possibility is to allow the methodologies used to converge towards similar approaches and to obtain results that can be compared.

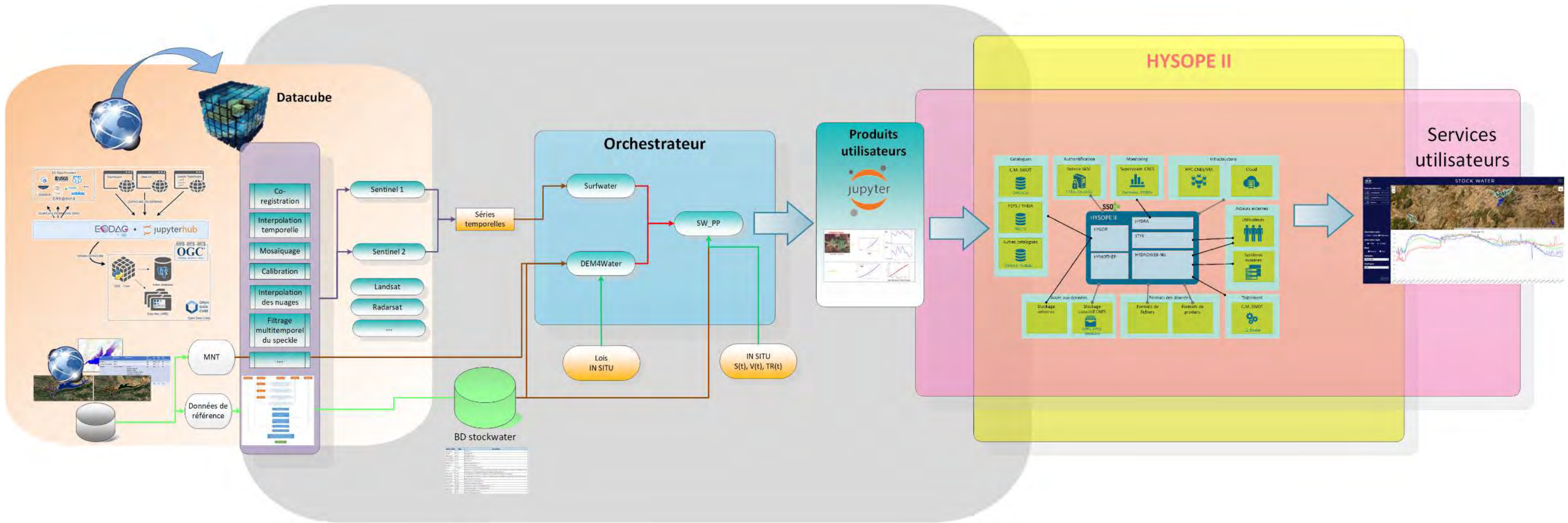


Figure 76 : STOCKWATER target operational architecture

8 References

- Circa 2015 UCLA Global Lake Database de Y. Sheng
- Pekel, J.F. & Cottam, A. & Gorelick, N. & Belward, A.S. (2016) “*High-resolution mapping of global surface water and its long-term changes*”, *Nature*, vol. 540, p. 418–422 (doi:10.1038/nature20584)
- Cleveland, W. S. (1979) “*Robust Locally Weighted Regression and Smoothing Scatterplots”*, *Journal of the American Statistical Association*, vol. 74, nº 368, p. 829–836 (DOI 10.2307/2286407, JSTOR 2286407)
- Savitzky, A. & Golay, M. J. E. “*Smoothing and Differentiation of Data by Simplified Least Squares Procedures”*, *Analytical Chemistry*, vol. 8, nº 36, 1964, p. 1627–1639 (DOI 10.1021/ac60214a047)
- Peña-Luque S. et al. (2021) “*Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level*” (https://doi.org/10.3390/rs13163279)
- Schumann and Bates, 2020: The need of a high accuracy, open free-acces global digital elevation model, Fontiers in Earth Sciences, doi: 10.3389/feart.2020.618194
- Yésou H., Maxant J., Briant J. & Studer M., 2018: SWOT, Support au Développement du Masque d’Eau de Référence et de la Base de Données Lacpour les Données SWOT HR, , Final report 7th December 2018, 122 pages

9 Glossary

ALCD	Active Learning for Cloud Detection
ALOS	Advanced Land Observing Datellite
API	Application Programming Interface
DB	Database
CIRCA-2015	Global Lake database, de Sheng
CNES	Centre National d'Étude Spatiale
DEM	Digital Elevation Model
DOM	Départements d'Outre-Mer
DREAL	Direction Régionale de l'Environnement, de l'Aménagement et du Logement
DSM	Digital Surface Model
EGM	Earth Gravitational Model
EODAG	Earth Observation Data Access Gateway
ERS	European Remote Sensing satellite
ESA	European Space Agency
ESRIN	European Space Research INstitute
ExtractEO	Chaine d'extraction multi thématique
GE	Google Earth
GEE	Google Earth Engine
GeoDAR	Georeferenced Global Dam and Reservoir
GEOSTORM	GEO Services plaTfORM
GOODD	GlObal geOreferenced Database of Dams
GRandD	Global Reservoir and Dam Database
GRD	Ground Range Detected
GRDC	Global Runoff *Data* Base Centre
GSWME	Global Surface Water Maximum Extent

GWS	Global Surface Water
HAND	Height Above Nearest Drainage (index)
HPC	High-Performance Computing
HR	High Resolution
H/S/V	Rapport Hauteur/Surface/Volume
HYSOPE	HYdrométrie Spatiale OPErationnelle
ICube	Laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie (INSA Strasbourg)
IGEDD	Inspection Générale de l'Environnement et du Développement Durable
IGN	Institut Géographique National
LEGOS	Laboratoire d'Etudes en Géophysique et Océanographie Spatiales
LOESS	LOcally Estimated Scatterplot Smoothing
LPS	Living Planet Symposium
MAE	Mean Absolute Error
MAJA	Maccs-Atcor Joint Algorithm
ME	Mean Error
MAPE	Mean Absolute Percentage Error
MERIT	Multi-Error-Removed Improved Terrain
MO	Mesure Optique
MR	Mesure Radar
MRF	Markov Random Fields
MSE	Mean Square Error
MSL	Metres above mean Sea Level
MTE	Ministère de la Transition Ecologique
MUSCATE	Atelier de production MUlti Satellite, multi-CApteurs, pour des données multi-TEmporelles
OTB	ORFEO Toolbox
PCA	Principal Component Analysis
PHR	Pleiades High Resolution

PoC	Proof of Concept
QGIS	Quantum Geographic Information System
R&D	Research & Development
RGE	Référentiel à Grande Echelle
RMSE	Root Mean Square Error
ROI	Region of Interest
SAIH	Sistema Automático de Información Hidrológica
SAR	Synthetic Aperture Radar
SCHAPI	Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations
SCO	Space Climate Laboratory
SERTIT	Service Régional de Traitement d'Image et de Télédétection
SRTM	Shuttle Radar Topography Mission
SURFWATER	High-resolution and near real time monitoring of SURFace WATER
SVM	Support Vector Machine
SWOT	Surface Water Ocean Topography Mission
THR	Très Haute Résolution (Very High Resolution)
TRL	Technology Readiness Level
WGS	World Geodetic System
WI	Water Index

Appendix A : Analysis of the time series S(t), V(t), Tr(t) according to the 75 and 90 quantiles

TABLE 23 : QUANTILE 75 - ANDALUSIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2310018272	Almodovar	33.547278	12.613832	7.433950	45.310303	22.744216	7.229447	9.967205
2310018492	Arcos de la Frontera	nan	7.931957	8.000759	34.623696	47.629828	nan	nan
2310022062	Arenoso	nan	8.430592	17.084015	23.869487	53.912905	7.391927	33.167396
2310018473	Barbate	16.224382	8.351728	7.580543	17.904286	19.545721	3.058606	5.032897
2160004183	Beznar	6.439839	5.084800	8.032180	14.289740	12.279271	8.102881	6.794264
2310018403	Bornos	16.921904	9.675219	22.093253	4.194782	11.557150	12.717003	14.778658
2160004602	Casasola	29.030360	20.916892	28.896923	22.762315	19.311958	10.896653	17.055283
2310020223	Celemin	1.479886	15.343878	22.600442	6.931381	16.844358	3.345681	6.993375
2160004643	Charco Redondo	6.796969	7.282636	6.018841	7.999891	15.397831	3.954294	12.504761
2160004403	Conde de Guadalhorce	13.902395	10.148412	17.489881	13.307706	24.221593	6.412347	7.402033
2160003673	Cuevas de Almanzora	39.476015	57.243247	22.727475	72.361268	22.225572	9.288500	30.302035
2310022963	Giribaile	11.548433	4.257335	9.226704	17.312267	14.230791	4.141705	5.745830
2310020153	Guadalcacin 2	11.479984	5.099615	3.572397	12.880846	15.689197	3.796062	4.044801
2310022023	Guadalen	89.275800	49.778196	60.259664	90.075324	90.684803	6.046917	11.174265
2160004373	Guadalhorce	nan	10.584282	13.542686	27.026843	26.850523	5.734296	7.091782
2160004383	Guadalteba	73.066744	6.285171	8.081185	19.836650	17.542344	7.342089	5.982575
2160004623	Guadarranque	6.155025	7.384085	7.446318	16.396164	22.540467	5.139108	9.063081
2310024073	Iznajar	5.355205	3.582882	2.989729	6.349167	10.750950	6.042059	9.601199
2310023523	Jose Toran	28.495275	32.121973	31.313776	6.168093	10.206507	15.755753	19.783636
2310024253	La Brena II	3.068848	62.774880	64.720208	6.991051	13.806722	4.989278	12.420633
2160004663	La Concepcion	11.107910	16.681813	14.588300	10.748466	9.553568	10.734830	10.207611
2160004413	La Vinuela	2.902015	3.671456	12.135066	8.297334	9.709391	3.458595	5.993356
2160004532	Limonero	6.493850	8.129216	14.034544	23.476724	39.124305	13.311018	25.056154
2310022903	Los Bermejales	5.789899	4.815453	15.146953	7.260066	21.667498	2.619843	12.793864
2310018233	Los Hurones	23.623679	19.514065	22.010100	8.874069	8.666416	11.344527	15.763449
2310027322	Los Melonares	6.400711	5.561896	3.958556	15.640671	22.573749	8.099304	14.004229
2310024113	Puebla de Cazalla	0.928407	7.178360	8.224102	23.805889	29.722002	14.415954	18.194139
2310027683	Puente Nuevo	16.699458	14.544177	5.648255	21.860100	11.961196	3.100448	5.111897
2160004253	Rules	3.300350	3.038741	9.278179	16.249668	45.617980	7.916660	30.455821
2310000173	San Rafael de Navallana	3.159786	14.060453	14.186791	12.917356	22.506130	9.023380	16.000146
2310020933	Tranco de Beas	10.442158	14.919749	22.230213	11.805526	17.964783	4.156693	4.113448
2310023943	Vadomojon	12.727538	5.566702	6.153823	19.731759	27.881930	6.371155	11.142219
2310022743	Yeguas	5.100191	7.252294	22.286444	6.736309	25.940180	8.109317	16.085205
2310018393	Zahara	0.450902	6.428811	4.097839	17.485094	24.708207	10.893803	15.679677

TABLE 24 : SUMMARY OF RESULTS QUANTILE 75 - NUMBER OF TANKS PER CLASS - ANDALUSIA

	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
<= 10%	15	19	15	10	3	24	14
> 10 & <=25%	10	10	14	19	22	8	14
> 25%	5	3	3	3	7	0	4

TABLE 25 : QUANTILE 90 - ANDALUSIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2310018272	Almodovar	33.547278	17.936419	10.849978	51.020137	27.522464	10.831666	12.272512
2310018492	Arcos de la Frontera	nan	9.730069	9.935733	35.900511	48.732625	nan	nan
2310022062	Arenoso	nan	14.354944	23.134354	26.824378	57.275858	10.988898	36.477985
2310018473	Barbate	16.224382	11.554563	14.833166	20.160667	21.140744	5.182697	10.216127
2160004183	Beznar	6.439839	8.381177	11.331077	17.266432	14.595955	9.734574	9.044481
2310018403	Bornos	16.921904	11.736326	35.122093	5.742730	16.284561	15.224204	17.370509
2160004602	Casasola	29.030360	26.853583	39.845308	28.789612	27.074851	12.720796	21.319328
2310020223	Celemin	1.479886	23.393794	39.346170	24.816035	21.604767	11.059583	9.573977
2160004643	Charco Redondo	6.796969	10.085451	9.581984	10.394732	19.533066	5.863543	13.669956
2160004403	Conde de Guadalhorce	13.902395	13.757481	31.626050	16.436183	38.249013	10.255472	12.611306
2160003673	Cuevas de Almanzora	39.476015	62.025243	34.667788	77.603026	29.103723	11.533498	34.160851
2310022963	Giribaile	11.548433	5.845243	16.520062	20.373863	28.082608	5.184218	9.897005
2310020153	Guadalcacin 2	11.479984	7.925997	5.510279	15.475482	17.781682	4.975993	5.828911
2310022023	Guadalen	89.275800	65.538716	71.353356	90.674724	91.754435	7.704505	13.716034
2160004373	Guadalhorce	nan	17.454372	23.880236	30.070452	28.708252	9.155359	8.771758
2160004383	Guadalteba	73.066744	9.159217	14.245389	25.234582	19.467524	11.740012	7.687055
2160004623	Guadarranque	6.155025	13.990393	10.413735	22.359640	24.328738	7.912635	10.353802
2310024073	Iznajar	5.355205	4.902507	4.611717	7.449581	12.228099	7.239298	10.624894
2310023523	Jose Toran	28.495275	36.199244	39.407413	9.426691	14.623372	16.667551	23.345777
2310024253	La Brena II	3.068848	68.529999	69.419129	10.730517	17.558382	7.075393	14.824945
2160004663	La Concepcion	11.107910	30.138611	25.180196	34.845523	15.076067	27.596691	12.085839
2160004413	La Vinuela	2.902015	5.466106	16.750477	10.132060	20.246382	4.723017	14.754422
2160004532	Limonero	6.493850	13.367223	18.696572	26.989118	42.457403	15.889522	28.223692
2310022903	Los Bermejales	5.789899	7.266745	20.490825	9.734432	38.867503	3.965115	23.061441
2310018233	Los Hurones	23.623679	21.176675	28.802440	10.386191	15.276049	13.368208	17.988985
2310027322	Los Melonares	6.400711	6.996448	5.448143	17.165367	23.781239	10.929410	15.967739
2310024113	Puebla de Cazalla	0.928407	10.401633	11.875900	25.834317	32.568232	15.728989	20.226637
2310027683	Puente Nuevo	16.699458	17.046150	8.244389	24.521873	14.530015	4.240763	6.568781
2160004253	Rules	3.300350	4.740213	24.809523	17.593331	47.277557	9.250021	31.595989
2310000173	San Rafael de Navallana	3.159786	23.746792	20.215550	22.271840	27.950419	15.118598	20.227510
2310020933	Tranco de Beas	10.442158	21.426526	28.832333	14.643273	24.929009	5.092145	6.100586
2310023943	Vadomojon	12.727538	7.482829	9.776509	22.337802	31.857465	8.005253	12.395047
2310022743	Yeguas	5.100191	11.800159	26.841003	10.543568	30.280605	10.840667	19.580223
2310018393	Zahara	0.450902	9.243104	6.219889	19.966852	26.547078	12.530394	16.933147

TABLE 26 : SUMMARY OF 90TH QUANTILE RESULTS - NUMBER OF TANKS PER CLASS - ANDALUSIA

	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
<= 10%	15	11	7	4	0	15	8
> 10 & <=25%	10	16	14	19	17	16	20
> 25%	5	5	11	9	15	1	4

TABLE 27 : QUANTILE 75 – OCCITANIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2160030553	Agly	5.711032	11.260062	9.071860	8.363164	23.482556	5.485555	18.388214
2320039133	Astarac	13.847274	22.756544	34.310635	26.016390	39.713758	11.897968	21.111238
2160030183	Avene	nan	8.675368	17.405380	35.032678	64.080847	13.065877	44.861148
2320030823	Cammazes	3.318017	12.394415	21.045248	12.232018	52.683757	10.817895	38.141865
2320038733	La Gimone	9.497278	7.139942	6.066951	20.359937	9.823199	5.289311	9.641062
2320030233	Lac de Montbel	87.039512	13.447082	29.756094	81.949098	84.163010	32.476790	38.071736
2320031293	Laparan	69.055075	42.660677	60.856555	47.375434	60.535647	14.307189	51.031186
2160028013	Matemale	34.228502	8.883670	16.018256	33.793625	49.982046	6.668608	8.985165
2320028893	Pareloup	33.081892	6.716751	11.446477	24.830080	22.379380	4.767475	7.723774
2320031303	Pla de Soulcem	67.837554	16.461385	42.976499	48.474868	41.597659	6.264390	32.446296
2160029943	Puyvalador	19.861078	18.675278	19.528004	31.423249	27.856438	11.452413	11.493102
2320028933	Saint Geraud	nan	15.179250	19.606544	18.154650	47.917189	28.161158	26.673209
2320033043	Saints Peyres	99.987579	19.970420	37.812677	99.985982	99.993195	8.921933	22.268208
2160026973	Salagou	90.151847	2.147468	4.477391	91.649249	92.294452	2.271981	3.835921
2160029873	Villeneuve la Raho	99.957016	10.605282	11.730010	99.903771	99.914602	12.943376	9.700859
2160030123	Vinca	77.570885	24.462890	40.236460	79.875811	83.419786	15.839728	25.635605

TABLE 28 : SUMMARY OF QUANTILE 75 RESULTS - NUMBER OF TANKS PER CLASS - OCCITANIA

	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
= 10%	2	3	1	1	0	4	2
> 10 & <=25%	2	6	6	3	2	5	3
> 25%	4	1	3	6	8	1	5

TABLE 29 : QUANTILE 90 – OCCITANIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2160030553	Agly	5.711032	16.229235	11.970850	10.152143	26.340447	7.929359	20.741633
2320039133	Astarac	13.847274	32.500867	44.434044	29.102591	42.755979	14.057070	23.341384
2160030183	Avene	nan	15.492933	29.171672	36.691640	66.989245	16.578530	48.190875
2320030823	Cammazes	3.318017	18.317091	38.502775	15.524088	63.728796	12.814384	47.361179
2320038733	La Gimone	9.497278	12.447470	11.697148	59.448668	62.847226	41.406501	44.463432
2320030233	Lac de Montbel	87.039512	21.142715	60.134999	82.513946	85.034749	49.839138	50.468657
2320031293	Laparan	69.055075	99.434807	88.113805	99.638467	88.016359	93.614001	65.573049
2160028013	Matemale	34.228502	49.738514	70.132315	41.902392	89.334807	33.154117	46.997457
2320028893	Pareloup	33.081892	8.050474	17.109934	25.939124	29.398664	6.404208	9.234413
2320031303	Pla de Soulcem	67.837554	36.626677	56.325642	74.183296	59.421866	7.366489	40.696840
2160029943	Puyvalador	19.861078	83.869171	54.813589	86.999620	44.129200	65.079505	34.495404
2320028933	Saint Geraud	nan	19.833040	27.846759	21.647172	51.606140	36.695607	37.447414
2320033043	Saints Peyres	99.987579	34.113683	71.840665	99.988683	99.996919	11.186282	25.796473
2160026973	Salagou	90.151847	2.955908	7.972652	91.758436	92.636936	3.134522	5.688913
2160029873	Villeneuve la Raho	99.957016	12.850839	20.190582	99.910191	99.917453	14.555315	11.639616
2160030123	Vinca	77.570885	53.308665	72.637976	80.717322	84.498469	35.071769	45.256936

TABLE 30 : SUMMARY OF 90TH QUANTILE RESULTS - NUMBER OF TANKS PER CLASS - OCCITANIA

	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
<= 10%	4	1	1	0	0	1	2
> 10 & <=25%	2	4	4	1	1	4	2
>= 25%	8	10	10	14	14	10	11

TABLE 31: QUANTILE 75 – INDIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
4530212143	Himayat Sagar	44.017309	27.256249	36.725319	48.341240	46.024782	26.664732	19.535531
4530347283	Nagarjuna Sagar	89.965918	870.579259	553.581415	2120.444824	1393.448521	29.692703	35.739524
4530212043	Osman Sagar	115.957147	inf	549.813378	inf	174.348741	43.668839	35.026014

TABLE 32 : QUANTILE 90 – INDIA

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
4530212143	Himayat Sagar	44.017309	27.968040	45.740818	50.326562	51.672850	27.975620	24.231737
4530347283	Nagarjuna Sagar	89.965918	1165.226218	1608.183781	2794.524049	3748.929325	35.437006	45.709367
4530212043	Osman Sagar	115.957147	inf	inf	inf	inf	55.285021	45.896479