



ENGAGÉS À VOS CÔTÉS
AU SERVICE DE VOS ENJEUX CRITIQUES

Page 1 sur 123

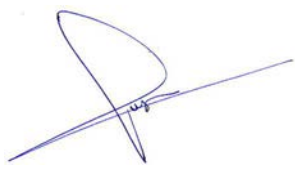


STOCK WATER



Rapport final

Tableau des signatures

RAPPORT FINAL	NOM	FONCTION	DATE	SIGNATURE
Ecrit par :	Sabrine Amzil Aurélien Bricier Thomas Ledauphin Gael Nicolas Benjamin Tardy Hervé Yesou	Equipe projet	24/01/2023	p.o. Gael Nicolas 
Vérifié par :	Frédéric Liège	Responsable d'activité "Etudes et Simulations Radar »	24/01/2022	

Versions successives

VERSION	DATE	MOTIF
1.0	10/01/2022	Création
1.2	22/02/2022	Rapport Phase I
1.3	21/10/2022	Rapport Phase II
1.9	07/12/2022	Rapport final, version intérimaire
2.0	21/12/2022	Rapport final_fr
2.1	24/01/2023	Rapport final_fr (prise en compte des remarques et derniers apports internes)

Table des matières

1	Objectifs de l'étude.....	10
1.1	Les Enjeux.....	10
1.2	Le Contexte technique.....	10
1.3	Objectifs du projet.....	10
1.4	Contraintes	11
1.4.1	Disponibilité des données <i>in situ</i> en adéquation avec le déroulement du projet.....	11
1.4.2	Préparation des données.....	11
1.4.3	Evolution des chaines	11
1.4.4	Périmètre de l'étude.....	11
2	Organisation de l'étude	12
2.1	Méthode	12
2.2	Articulation de l'étude.....	13
3	Déroulement de la Phase I	1
3.1	Développement du démonstrateur.....	1
3.1.1	Surfwater	1
3.1.2	DEM4water.....	1
3.1.3	Surfwater_Postprocess.....	2
3.1.4	Orchestracteur Stockwater	3
3.2	Mise en place de la base de données	4
3.2.1	Sélection de cibles	4
3.2.2	Constitution des datasets de la Phase I.....	5
3.2.3	Campagne Andalousie	7
3.2.4	Campagne Occitanie.....	7
3.2.5	Campagne Inde.....	8
3.2.6	Préparation des données.....	9
3.2.7	Analyse de la qualité des MNTs.....	25
3.3	Analyse et validation	30
3.3.1	Evaluation qualitative du Water extent.....	30
3.3.2	Analyse des lois REF.....	46
3.3.3	Analyse séries temporelles S(t), V(t), Tr(t).....	49
4	Déroulement de la Phase II	59

4.1	Développement du démonstrateur.....	59
4.1.1	Surfwater	59
4.1.2	DEM4water.....	59
4.1.3	Surfwater_Postprocess.....	60
4.1.4	Orchestracteur Stockwater	60
4.2	Mise en place de la base de données	61
4.2.1	Tache 2.1 – Sélection de cibles.....	61
4.2.2	Tache 2.2 – Préparation des données	64
4.2.3	Analyse séries temporelles S(t), V(t), Tr(t).....	74
4.3	Production	77
4.3.1	Génération des masques d'eau	77
4.3.2	Calcul de relations HSV et comparaison	77
4.3.3	Génération des séries S(t), V(t) et TR(t).....	79
4.4	Analyse/Validation des résultats de la phase complémentaire	79
4.4.1	Volume absolu	79
4.4.2	Taux de remplissage	80
4.5	Valorisation.....	81
5	Robustification du démonstrateur et du tableau de bord	87
5.1	Améliorations de la chaine	87
5.1.1	Surfwater	87
5.1.2	DEM4water.....	87
5.1.3	Surfwater_Postprocess.....	87
5.1.4	Orchestracteur	87
5.2	Modification de la base de données.....	88
5.3	Génération de nouvelles séries temporelles	88
5.3.1	Finalisation des calculs de lois ZSV	88
5.3.2	Relance des calculs SURFWATER.....	88
5.3.3	Relance des calculs SURFWATER_Postprocess.....	88
5.3.4	Formatage des nouvelles séries temporelles	89
5.4	Valorisation.....	89
5.4.1	Robustification du tableau de bord	89
5.4.2	Intégration des données et mise à jour finale.....	89
6	Conclusions et Perspectives	90
6.1	Synthèse de l'étude	90

6.2	Une première approche concluante.....	91
6.3	Des axes d'amélioration pour une version opérationnelle	92
6.3.1	Fluidifier la manipulation des données	92
6.3.2	Vers une chaine opérationnelle.....	94
6.3.3	Une phase de validation à stabiliser.....	94
6.3.4	Une architecture et une interface adaptées à une plus large communauté d'utilisateurs.....	95
7	Références	1
8	Glossaire	2
9	Annexe A : Analyse des séries temporelles $S(t)$, $V(t)$, $Tr(t)$ selon les quantiles 75 et 90.....	5
10	Annexe B : Analyse des lois REF	11
10.1	Analyse des lois ref vs <i>in situ</i>	11
10.2	Analyse des lois ref vs dem4water	12

Liste des figures

Figure 1 : Workflow Stockwater

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)

Figure 4 : Carte de localisation – Andalousie

Figure 5 : Carte de localisation Occitanie

Figure 6 : Carte de localisation – Inde

Figure 7 : Comparaison volume - Lower Manair

Figure 8 : comparaison volume - Nizam Sagar

Figure 9: comparaison volume - Sriram Sagar

Figure 10 : comparaison volume - Nagarjuna Sagar

Figure 11: Telangana Water Ressources Information System

Figure 12 : Données in situ Nizam Sagar & Pulichintala Project

Figure 13 : Données in situ Osman Sagar & Himayat Sagar

Figure 14 : Données in situ Sriram Sagar & Buggavu

Figure 15 : Données in situ Lower Manair & Nagarjuna Sagar

Figure 16 : Suppression des "plateaux" - Sriram Sagar

Figure 17 : Suppression des outliers - Sriram Sagar

Figure 18 : Ajustement des valeurs en escaliers - 4530433353 - Lower Manair

Figure 19 : Données in situ brutes (gauche) et modifiées (droite) - 4530240913 - Sriram Sagar

Figure 20 : Données in situ brutes (gauche) et modifiées (droite) - 4530313823 – Buggavagu

Figure 21 : Données in situ brutes (gauche) et modifiées (droite) - 4530347283 - Nagarjuna Sagar

Figure 22 : Données in situ brutes (gauche) et modifiées (droite) - 4530433353 - Lower Manair

Figure 23 : données *in situ* brutes (bleu) et modifiées (vert) - 4530212043 - Osman Sagar

Figure 24 : données *in situ* brutes (bleu) et modifiées (vert) - 4530433353 - Lower Manair

Figure 25 : données *in situ* brutes (bleu) et modifiées (vert) - 4530347283 - Nagarjuna Sagar

Figure 26 : données *in situ* brutes (bleu) et modifiées (vert) - 4530187293 - NIZAM SAGAR

Figure 27 : Relations entre les lois et les paramètres associés

Figure 28 : Représentation des lois $V(Z)$ et $V(S)$

Figure 29: Différence d'élévation entre les DEM Copernicus 30 et 10, ALOS DEM, SRTM et le MNT IGN 5m de référence

Figure 30: Différence entre DEM SAR 30m, SRTM , Copernicus et le MNT de référence IGN

Figure 31: Courbes des gaussiennes approximées des histogrammes de la répartition des erreurs d'altitudes en fonction des classes de pentes

Figure 32 : Architecture de la suite logicielle ExtractEO du SERTIT

Figure 33 : Location map tuile Sentinel-2 (30SUH)

Figure 34 : Arenoso - Instant Mask

Figure 35 : Arenoso - Water Mask

Figure 36 : Puente Nuevo - Water Mask

Figure 37 : Puento Nieve - Water Mask

Figure 38 : Yeguas - Water Mask

Figure 39 : Location map tuile Sentinel-2 (30 STF)

Figure 40 : Bornos - Instant Mask

Figure 41 : Bornos - Water Mask

Figure 42 : Almodovar - Water Mask

Figure 43 : Cemelin - Water Mask

Figure 44 : Barbate - Water mask

Figure 45 : Water Extent EEO & SURFWATER - Barbare - 26/09/2018

Figure 46 : Trace Sentinel-2B à gauche et Sentinel-2A à droite

Figure 47 : Réflectance plus ou moins impactée par les effets de sunglint selon la trace position dans la trace. A gauche Sentinel-2A, et à droite Sentinel-2B. Exemple du réservoir de Almodovar. Composition colorée S2, B8, B4, B3 en RVB, images acquises entre le 19/05/2018 et le 29/06/2018.

Figure 48 : Réflectance plus ou moins impactée par les effets de sunglint selon la trace position dans la trace. A gauche Sentinel-2A, et à droite Sentinel-2B. Exemple du réservoir de Almodovar. Composition colorée S2, B8, B4, B3 en RVB.

Figure 49 : Réflectance plus ou moins impactée par les effets de sunglint selon la trace position dans la trace. A gauche Sentinel-2A, et à droite Sentinel-2B. Exemple du réservoir de Almodovar. Composition colorée S2, B8, B4, B3 en RVB, images acquises entre le 28/08/2018 et le 16/07/2018.

Figure 50 : Réflectance plus ou moins impactée par les effets de sunglint selon la trace position dans la trace. A gauche Sentinel 2A, et à droite Sentinel 2B. Exemple du réservoir de Almodovar. Composition colorée S2, B8, B4, B3 en RVB, images acquises entre le 18/07/2018 et le 28/07/2018.

Figure 51 : Water mask - Charco Redond

Figure 52 : Water Mask - Los Hurones

Figure 53 : Water Mask – Bornos

Figure 54 : Water Extent EEO & STOCKWATER - Bornos - 27/04/2018

Figure 55 : WATER EXTENT EEO & STOCKWATER - Bornos - 04/10/2018

Figure 56 : WATER EXTENT EEO & STOCKWATER - Bornos - 04/10/2019

Figure 57 : Comparaison après soustraction de 7% à la surface de référence – Charco Redondo

Figure 58 : Comparaison après soustraction de 12% à la surface de référence – Los Hurones

Figure 59: Comparaison après soustraction de 8% à la surface de référence - Bornos

Figure 60 : Comparaison S(Z) avec la référence

Figure 61 : Comparaison V(S) avec la référence

Figure 62 : Boucle de la ligne de coupure

Figure 63 : Water crossing

Figure 64 : Boucle et water crossing de la ligne de coupure

Figure 65 : Multi-déversoirs de Guadalteba

Figure 66 : Série temporelle des écarts relatifs absolus – Yeguas (Andalousie)

Figure 67 : Série temporelle des écarts relatifs absolus – Arenoso (Andalousie)

Figure 68 : Série temporelle des écarts relatifs absolus – Arenoso (Andalousie)

Figure 69 : Série temporelle des écarts relatifs absolus – Rules (Andalousie)

Figure 70 : Série temporelle des écarts relatifs absolus – Los Melonares (Andalousie)

Figure 71 : Série temporelle des écarts relatifs absolus – Bornos (Andalousie)

Figure 72 : Comparaison d'estimation de volume hebdomadaire (graphiques gauche) avec l'estimation de Taux de remplissage (graphiques droite) pour les barrages Pueblo Nuevo, Iznajar et Agly. En rouge les estimations

optiques (configuration problématique - MO2) , en Bleu le Radar (MR2) et ligne verte la référence (volume / taux de remplissage)

Figure 73 : correction des anomalies de lignes de coupure

Figure 74: Sites Tunisie

Figure 75 : Sites Brésil

Figure 76 : Sites Burkina Faso

Figure 77 : Site Laos

Figure 78 : Workflow de la préparation d'un jeu de donnée Dam Database

Figure 79 : Représentation des lois $V(Z)$ et $S(Z)$ pour les données *in situ* non modifiées

Figure 80 : Représentation des Lois $v(Z)$ et $S(Z)$ pour les données *in situ* corrigées (cas de plateaux)

Figure 81 : Représentation des Lois $v(Z)$ et $S(Z)$ pour les données *in situ* corrigées (cas des données dupliquées)

Figure 82 : Séries temporelles des surfaces MO2, MR2 et INSITU - Arraras

Figure 83 : Séries temporelles des surfaces M02, MR2 et INSITU - Jaburu II

Figure 84 : SERIES TEMPORELLES DE VOLUMES M02, MR2 ET INSITU – Serrote

Figure 85 : SERIES TEMPORELLES DE VOLUMES M02, MR2 ET INSITU - REALEJO

Figure 86 : SERIES TEMPORELLES DE VOLUMES M02, MR2 ET INSITU – SAO JOSE III

Figure 87 : détails de la page d'accueil.

Figure 88 : données affichées après sélection sur un réservoir.

Figure 89 : fonctionnalités de sélection des items. Partie droite : détails de l'élément à l'ajout d'un réservoir. Partie gauche : anatomie de la tooltip du graphique et de la sélection des différentes métriques.

Figure 90 : exemple d'une recherche en décochant un type d'observation.

Figure 91 : exemple d'une recherche sur plusieurs réservoirs.

Figure 92 : exemple d'une partie zoomée du graphique.

Figure 93 : détails du mode « Volume ».

Figure 94 : détails du mode « Year ».

Figure 95 : exemple d'un affichage des données avec le style « Scatter ».

Figure 96 : architecture opérationnelle cible STOCKWATER

Figure 97 : Série temporelle des écarts relatifs absolus – Conde de Guadalhorce (Andalousie)

Figure 98 : Série temporelle des écarts relatifs absolus – Guadalhorce (Andalousie)

Figure 99 : Série temporelle des écarts relatifs absolus – Rules (Andalousie)

Liste des Tableaux

Tableau 1 : synthèse des bases de données disponibles
Tableau 2 : Inventaire des données in situ disponibles sur les réservoirs identifiés pour la Phase I
Tableau 3 : Structure attributs dataset
Tableau 4 : Données in situ Occitanie
Tableau 5 : Données in situ Andalousie
Tableau 6 : Données in situ Inde
Tableau 7 : Résumé des données in situ retenues
Tableau 8 : Répartition des pentes
Tableau 9 : Médiane – Andalousie
Tableau 10 : résumé résultats médiane - nombre réservoirs par classe - andalousie
Tableau 11 : Médiane - Occitanie
Tableau 12 : Résumé résultats médiane - Nombre reservoirs par classe - Occitanie
Tableau 13 : Médiane - Inde
Tableau 14 : Exemple d'erreurs d'estimation des surfaces impactant l'estimation des volumes - Andalousie
Tableau 15 : Exemple d'erreurs V(S) impactant l'estimation des volumes - Andalousie
Tableau 16 : Exemple d'erreurs V(S) impactant l'estimation des volumes - Occitanie
Tableau 17 : Exemple compensation des erreurs – Andalousie
Tableau 18 Médiane-Occitanie
Tableau 19 résumé résultats médiane - nombre réservoirs par classe - Occitanie
Tableau 20 Médiane-Andalousie
Tableau 21 Résumé résultats médiane - nombre réservoirs par classe - Andalousie
Tableau 22 : Tableau récapitulatif des données in situ disponibles pour le site Brésil
Tableau 23 : MEDIANE - Brésil
Tableau 24 : Résumé résultats médiane - Nombre de réservoirs par classe - Brésil
Tableau 25 : nombre de modèles générés en fonction des réservoirs étudiés, par Super site
Tableau 26 : Quantile 75 - Andalousie
Tableau 27 : Résumé résultats quantile 75 - Nombre réservoirs par classe - Andalousie
Tableau 28 : Quantile 90 - Andalousie
Tableau 29 : Résumé résultats quantile 90 - Nombre réservoirs par classe - Andalousie
Tableau 30 : Quantile 75 – Occitanie
Tableau 31 : Résumé résultats quantile 75 - Nombre réservoirs par classe – Occitanie
Tableau 32 : Quantile 90 – Occitanie
Tableau 33 : Résumé résultats quantile 90 - Nombre réservoirs par classe - Occitanie
Tableau 34: Quantile 75 – Inde
Tableau 35 : Quantile 90 – Inde

1 Objectifs de l'étude

1.1 Les Enjeux

La gestion et la compréhension de notre environnement deviennent un enjeu géopolitique majeur, particulièrement critique dans des domaines tels que l'écologie, l'économie, l'énergie, la sécurité alimentaire, l'aménagement du territoire, la gestion des risques ou la Sécurité. La connaissance exacte en temps réel de l'état géophysique de notre planète et de son évolution dans le temps devient donc un enjeu majeur.

D'un point de vue stratégique, pouvoir disposer d'un outil permettant d'évaluer les ressources en eau d'un pays tiers avec une échelle de précision pertinente, constitue un avantage essentiel.

Il est alors possible d'évaluer non seulement les ressources en eau mais également directement l'évolution des capacités en termes d'irrigation (sécurité alimentaire, capacités de production agricole et donc d'importations/exportations), de production d'énergie (hydroélectricité, refroidissement centrales...), de production (et donc de dépendance) industrielle pour certaines filières, d'alimentation des populations en eau potable (conflits sur l'eau, dimensionnement de l'aide humanitaire, etc...). Un tel outil permettrait le pilotage à des échelles pertinentes de stratégies de gestion de l'eau (allocation événements climatiques extrêmes...), mais également la compréhension de dynamiques à plus long terme telles que la prise en compte des effets du changement climatique et les adaptations nécessaires.

Les effets de la sécheresse de l'été 2022, dans un contexte géopolitique tendu, ont particulièrement mis en évidence le caractère sensible de la gestion de cette ressource, pour laquelle il n'existe actuellement pas de méthode permettant le suivi des volumes à grande échelle.

D'un point de vue technique, du fait de la criticité des contraintes, et de l'ampleur des données à exploiter, des approches précises, automatiques et systématiques sont indispensables. Il est donc indispensable de développer des outils de traitement des données satellitaires, afin de permettre l'exploitation et l'interprétation en temps contraint de ces données, dans le but de générer des produits opérationnels à haute valeur ajoutée permettant de répondre à ces problématiques, par exemple le cas particulier du suivi des ressources en eau.

1.2 Le Contexte technique

Au-delà de la problématique des surfaces en eau, celle de la gestion des ressources captées (barrages, retenues, réservoirs...) est particulièrement complexe dans la mesure où il existe plus de 45 000 retenues majeures dans le monde, qui couvrent des besoins critiques (l'approvisionnement en eau domestique et industrielle, irrigation, production d'énergie, contrôle des crues...), et que presque tous les grands cours d'eau du monde sont affectés par la présence de barrages.

La détection de l'eau est une tâche complexe, et diverses méthodes ont été mises au point dans le but d'extraire les plans d'eau des données de télédétection acquises dans différents domaines d'observation. Les hydrologues et hydrauliciens sont fortement demandeurs d'une capacité à connaître en temps quasi réel où se trouve l'eau et comment les surfaces en eau évoluent. La chaîne SURFWATER, qui exploite les données Sentinel 1 et 2 dans le but de générer une carte d'observation des surfaces en eau sur une zone d'intérêt donnée, permet de répondre à cette problématique.

1.3 Objectifs du projet

L'objectif du projet SCO STOCKWATER est de mettre à disposition des utilisateurs (gestionnaires de barrages, experts, décideurs) une solution globale de suivi dynamique en temps contraint des stocks d'eau de barrage sur l'ensemble du globe, à des échelles pertinentes pour la gestion de la ressource.

D'un point de vue technique, cette démonstration repose sur une mise à l'échelle et à une généralisation à d'autres régions du monde de méthodes évaluées en France, du moins dans la mesure où des données pertinentes sont disponibles, avant un éventuel saut vers une application adaptable au niveau

national voire mondial.

Les méthodes opérationnelles précédemment développées et mises en œuvre dans le cadre de l'analyse des stocks d'eau par suivi satellitaire en France ont démontré leur pertinence. Le projet SCO STOCKWATER s'attache à caractériser et démontrer la généricité et l'adaptabilité de ces méthodes en les appliquant à un très grand nombre de sites, pour des régions et des conditions d'utilisation variées. Si cette démonstration se révèle concluante, la méthode peut ouvrir des perspectives majeures dans la possibilité de généralisation ou d'adaptabilité pour le suivi des ressources en eau au niveau global.

Les résultats dynamiques de suivi des ressources sont mis à disposition des utilisateurs au travers d'une plateforme web ouverte.

1.4 Contraintes

1.4.1 Disponibilité des données *in situ* en adéquation avec le déroulement du projet

Du fait du grand volume de sites à traiter, il est important de s'assurer de la disponibilité des données pour chacun d'entre eux. Une des contraintes majeures du projet est de pouvoir disposer de suffisamment de sites valides, c'est-à-dire pour lesquels des données *in situ* sont disponibles.

La liste des fournitures disponibles inclut des données *in situ* en Occitanie, en Andalousie et en Inde, mais le projet dans sa forme aboutie est dépendant du nombre de sites couverts par ces données, ainsi que de la mise à disposition par les partenaires locaux de données sur les autres zones à traiter (Laos, Burkina-Faso et Tunisie) dans des délais compatibles avec la réalisation du projet d'une part, mais au-delà pour les mises à jour dans des délais pertinents pour une utilisation opérationnelle. Au-delà de l'aspect technique se pose donc la problématique de l'accessibilité des données, à des fréquences et des degrés de précision pertinents, et qui n'est pas acquise, même pour la France (certaines données restent confidentielles).

Dans le cas où ces données *in situ* ne seraient pas disponibles, il est donc nécessaire d'étudier des solutions alternatives dans d'autres régions, ou de modifier la répartition des sites en s'appuyant sur les autres zones d'étude ciblées.

1.4.2 Préparation des données

En ce qui concerne les données d'observation de la terre, l'étude exploite les produits d'entrée déjà disponibles dans Theia (notamment pour la France, l'Espagne, l'Inde, la Tunisie, le Laos et le Burkina-Faso). Dans le cas des sites non couverts par Theia, la mise en œuvre de MAJA sera alors requise pour disposer des données Sentinel-2 L2A nécessaires à la chaîne Surfwater.

1.4.3 Evolution des chaines

Des évolutions de la chaîne Surfwater-Postprocess sont nécessaires, et des améliorations sont à apporter pour la modélisation HSV, afin de passer de l'état de Preuve de Concept à celui de chaîne de production. Ces évolutions ont été intégrées dès la Phase I d'implémentation de la Preuve de Concept.

La sélection des améliorations à implémenter dans la chaîne de modélisation HSV s'est fait en accord avec le CNES, en s'assurant de la maîtrise des temps de développement dans le but d'aboutir à un outil utilisable dans des conditions d'ergonomie compatibles avec un outil pré-opérationnel et une démonstration de faisabilité, sans compromettre le planning de déroulement du projet.

1.4.4 Périmètre de l'étude

Etant donné le grand nombre initial de sites (250) et la fenêtre d'observation (3 ans) visés, le

déroulement de l'étude proposé initialement dans la description de l'offre a été adapté dans le but :

- ✓ De garantir une plus grande flexibilité dans les délais de disponibilité des données,
- ✓ De pouvoir découpler les tâches et les réaliser en parallèle,
- ✓ De valider un workflow end-to-end sur un sous-ensemble de sites maîtrisé et représentatif avant le passage à l'échelle,
- ✓ De permettre des évolutions itératives rapides sur ce sous-ensemble de sites pour accélérer la réactivité et la validation en prolongement de l'implémentation d'améliorations,
- ✓ De découpler les incertitudes :
 - De disponibilité de la donnée sur les 250 sites initiaux,
 - De validation fonctionnelle de la chaîne end-to-end,
 - De la complexité de passage à l'échelle par parallélisation massive.

2 Organisation de l'étude

2.1 Méthode

Afin de s'assurer de la pertinence de la méthode à l'échelle globale, un cadre de validation maîtrisé a été mis en place en préalable à la généralisation de la méthode à l'ensemble des sites visés dans la liste initiale. Cette validation préalable avait pour but d'adresser les différents axes de variabilités susceptibles d'émerger lors de la généralisation des méthodes mises en œuvre.

Il s'agit notamment de prendre en compte les hypothèses suivantes :

- ✓ La variabilité de la nature des données d'entrée :
 - Les différents MNTs disponibles présentent une très grande hétérogénéité (nature, résolution, qualité, limitations). Ils ont donc été analysés dans le but d'identifier leurs contraintes et limitations intrinsèques,
 - La disponibilité d'un volume d'observations pertinent pour définir des séries temporelles d'images couvrant les évolutions à observer. Certaines zones du globe ne bénéficient pas d'une couverture suffisante en observations optiques Sentinel-2 non nuageuses nécessaires au suivi des surfaces en eau, de même que la couverture en acquisitions radar Sentinel-1 est plus ou moins dense selon les zones géographiques,
- ✓ La variété des topologies des sites de stockage en eau, tant au niveau du contexte naturel que des techniques de mise en place de retenues artificielles.

La mise en œuvre de la méthode pour un nombre limité (mais représentatif) de cibles a permis, dans ce cadre de validation, d'évaluer les performances dans des conditions représentatives du changement d'échelle induit par la configuration de Phase II et la configuration finale.

Fondamentalement, la méthodologie repose sur trois étapes majeures nécessaires au suivi d'un site :

1. Modélisation de la relation Hauteur/Surface/Volume de chaque barrage, basée sur l'analyse de MNT (chaîne DEM4water),
2. Détermination de la surface en eau du site à chaque observation réalisée par les satellites Sentinel-1 et Sentinel-2 (chaîne Surfwater) afin de collecter la série $S(t)$,
3. Génération des séries temporelle $S(t)$, $V(t)$ et $TR(t)$ (chaîne Surfwater-Postprocess).

Ces trois étapes seront appliquées à un nombre de sites suffisamment représentatif pour permettre un passage à l'échelle globale.

2.2 Articulation de l'étude

Devant l'étendue du cadre de mise en œuvre et de validation demandé (initialement 250 sites), il a été proposé d'organiser les tâches décrites en deux phases consécutives :

- ✓ **Phase I (Preuve de Concept)** : mise en œuvre de la totalité de la méthodologie en mode end-to-end sur un nombre réduit de sites, le plus hétérogène et représentatif possible, sur lesquels les données (entrée et validation) sont déjà disponibles, notamment dans les zones d'acquisition en Occitanie, Andalousie et Inde, pour lesquelles les données *in situ* sont disponibles. Cette première phase s'articule selon les étapes suivantes :
 - Sélection d'un nombre réduit de cibles, aussi représentatives que possible, en fonction de la disponibilité des données d'entrée et de validation,
 - Préparation des données,
 - Génération des masques d'eau sur les périodes de disponibilité des données *in situ*,
 - Modélisation des relations Hauteur, Surface, Volume (modélisation H/S/V) pour chacun des MNT disponibles,
 - Génération des séries temporelles $S(t)$, $V(t)$ et Taux de remplissage $TR(t)$ pour la sélection réduite de sites,
 - Validation sur ces sites :
 - Par analyse automatique comparative aux données *in situ*,
 - Par comparaison des performances en fonction des MNT utilisés pour la modélisation H/S/V.

Dans le cadre de cette Phase I, les outils nécessaires à ces étapes (orchestrateur, outil de comparaison automatique aux données *in situ*, améliorations et évolutions nécessaires aux outils existants...) ont été implémentés. A l'issue de cette phase initiale, l'analyse de la qualité des MNTs a permis de les classer en fonction de la qualité des modèles générés, et d'en exclure certains de la Phase II.

- ✓ **Phase II (Mise à l'échelle)** : exploiter l'environnement mis en œuvre en Phase I pour appliquer la méthodologie à l'ensemble des 250 sites. Les étapes composant cette Phase II étaient les suivantes :
 - Sélection des sites complémentaires à la Phase I, conjointement à la disponibilité élargie de données *in situ*,
 - Préparation des données pour ces nouveaux sites, ainsi que les données complémentaires des sites de la Phase I,
 - Génération des masques d'eau pour l'ensemble des sites,
 - Modélisation H/S/V pour les nouveaux sites pour chacun des types de MNTs considérés à cette étape,
 - Génération des séries temporelles $S(t)$, $V(t)$ et $TR(t)$ pour tous les sites et pour chaque modèle (un par MNT),
 - Validation des résultats :
 - Par analyse automatique comparative aux données *in situ* ;
 - Par comparaison des performances en fonction des MNT utilisés pour la modélisation H/S/V ;
 - Restitution des résultats sous forme d'un tableau de bord présentant des fonctionnalités suffisantes pour un outil pré-opérationnel de consultation et de manipulation des résultats, à destination des partenaires du CNES, pour des besoins génériques dans un premier temps.

3 Déroulement de la Phase I

3.1 Développement du démonstrateur

3.1.1 Surfwater

La chaîne Surfwater détecte les surfaces en eaux continentales en exploitant les observations issues des satellites Sentinel-2 (produits L2A) pour les données optiques et Sentinel-1 (produits GRD L1C) pour les données radars.

Plusieurs produits sont générés par la chaîne Surfwater, pour le mode optique comme pour celui radar :

- ✓ Un masque de détection des surfaces en eau relatif à une observation appelé single,
- ✓ Un masque multiple, composé de la synthèse de masques single sur une période de 10 jours,
- ✓ Un masque de cumul mensuel basé sur l'ensemble des masques single du mois.

La génération des produits Surfwater nécessite la préparation des données d'entrée sur la période requise pour une sélection de tuiles. Les méthodes de récupération des données optiques S2 L2A via l'API Amalthée sur le Datalake du CNES et radar S1 L1C via eodag (S1tiling) ont été rapidement confrontées aux limitations des serveurs. Ces problématiques ont pénalisé le projet et conduit à des points bloquants sur cette tâche.

Un effort conséquent d'amélioration et d'adaptation des méthodes de récupération a été fourni par rapport aux limitations des serveurs.

3.1.2 DEM4water

La chaîne DEM4water réalise l'estimation de la relation entre les variables : hauteur d'eau H , superficie du plan d'eau S et volume stocké V de retenues de barrages à partir des informations d'un MNT (DEM) environnant. La chaîne offre également la possibilité de comparaison par rapport à un modèle de référence si celui-ci est disponible.

Du fait de son état de prototype, DEM4water ne répondait pas en l'état à la fois aux besoins de production générique et au large volume de données requis par le projet Stockwater. Des évolutions ont été apportées afin d'automatiser et d'industrialiser la chaîne.

La dimension de la zone d'intérêt était fixée « à la main » dans le code et ne convenait pas forcément aux différentes formes de barrages, en plus du temps de traitement inutile si la zone était surdimensionnée afin de ne pas risquer de manquer une partie du barrage (sans pour autant avoir la garantie d'un bon résultat, du fait de la variabilité de forme des ouvrages).

Un outil a été développé afin de définir plus précisément la zone de contour des barrages à partir du calcul de l'enveloppe et de fournir le radius associé.

Le paramétrage de la liste des barrages à traiter a été mis en place, remplaçant leur noms et identifiants inscrits en dur dans le script. A ce titre, un outil a été développé réalisant la génération de liste d'ID/DAMS à partir d'un fichier base de données au format geojson.

Les traitements, initialement réalisés en série, demandaient pour certains barrages (Inde entre autres) un temps d'attente conséquent. Une optimisation a été réalisée en parallélisant les traitements par barrage.

Parmi les évolutions, on notera la mise en place d'arborescence en fonction de la possibilité de différents MNTs, avec la création de noms génériques et explicites, l'intégration du nouveau format des bases de données et ses attributs, la modification du principe de l'ID au sein de la chaîne avec la possibilité de choisir un ID spécifique, le passage en argument de modèle de référence.

Une mise en cohérence entre les unités a été effectuée dans le module de comparaison des modèles estimés et de référence traitant en m2 et en hectares, idem pour les volumes. Les unités manipulées

sont maintenant celles du système international.

3.1.3 Surfwater_Postprocess

L'objectif de cette étape consiste à exploiter les masques d'eau Surfwater conjointement avec les modèles estimés, pour générer des séries temporelles de quantités telle que la surface et le volume d'eau pour chaque barrage.

La chaîne Surfwater_Postprocess dispose de plusieurs méthodes de comptage de surfaces en eau offrant la possibilité de choisir parmi les différents produits Surfwater (masques single, masques multiples et masques d'occurrence mensuels), disponibles pour l'optique comme pour le radar.

Un post-traitement est réalisé sur les masques d'eau, et à l'aide de données géolocalisant les plans d'eau par un point en latitude et en longitude, les pixels des masques sont traités et comptabilisés pour fournir un état relatif à l'évolution temporelle des surfaces des barrages. Leur superficie est traduite en volume grâce à une loi estimée calculée par la chaîne DEM4water.

Les modules d'extraction et de comparaison des métriques se fondent sur les données d'entrée suivantes :

- ✓ une liste des masques d'eau calculés par la chaîne Surfwater,
- ✓ la base de données des barrages du site étudié (fourniture par le SERTIT),
- ✓ la table de conversion des surfaces en volumes correspondant à la loi estimée calculée par la chaîne DEM4water,
- ✓ les séries temporelles de référence provenant des données *in situ* (fourniture par le SERTIT).

Mis à part les masques d'eau Surfwater, ces entrées ont été modifiées en termes de format, puis de manière itérative au niveau du contenu et des champs, nécessitant des adaptations successives de la chaîne Surfwater_Postprocess.

La base de données des barrages est passée du format shapefile au format geojson avec de nouveaux attributs. L'outil `create_water_mask` a donc aussi été adapté pour générer le fichier de sortie au nouveau format.

Dans l'objectif de garder une cohérence avec les formats des fichiers utilisés dans le cadre du projet Stockwater, les lois estimées utilisées comme table de conversion ont été passées du format csv au format json, avec de nouveaux attributs et une mise en cohérence des unités afin de supprimer les utilisations hétérogènes des unités (m²/hectares, m³/hm³) et de se baser sur le système international d'unités.

Les séries temporelles de référence *in situ* ont également évolué en ce qui concerne le format, et il a donc fallu adapter la gestion interne des tableaux de données, et homogénéiser les unités au système international.

Une nouvelle option a été intégrée : la possibilité de sélectionner un nom de champ en tant qu'identifiant. Plusieurs champs id sont maintenant présents dans les bases de données livrées (id Grand, id SWOT).

Le module de comparaison des métriques a évolué pour intégrer, en plus du nom des barrages, leur identifiant dans les fichiers json d'évaluation.

Un nouvel indicateur complémentaire a été intégré à la chaîne. Il s'agit du calcul du taux de remplissage TR(t) en eau du barrage, accompagné de statistiques et graphiques de visualisation. Ce taux indique l'estimation du volume d'eau par rapport à la capacité maximale du barrage observé sur la période traitée, en y appliquant le quantile 98 afin d'éliminer les valeurs aberrantes. En complément un contrôle des valeurs extrêmes de cet indice figure en tant que warning dans le log.

La chaîne Surfwater_Postprocess dispose d'un module de comparaison des métriques (surface et volume), permettant d'évaluer les séries temporelles d'un ou plusieurs corps en eau par rapport aux séries temporelles de référence, générant graphiques et statistiques associées. Une nouvelle fonctionnalité a été développée offrant la possibilité de calculer des séries temporelles TR(t) de référence à partir de séries temporelles V(t) de référence, si présentes, afin d'évaluer qualitativement le taux de remplissage en eau des barrages au même titre que les autres quantités.

Le paramétrage des ressources liées à l'ordonnanceur PBS a dû être réajusté en fonction des caractéristiques hors normes de certains barrages.

3.1.4 Orchestrateur Stockwater

Afin de faciliter la réalisation des campagnes de productions, un orchestrateur Stockwater a été développé pour réaliser la gestion et l'enchaînement des traitements destinés à la production sur un site, pour un ensemble de tuiles, sur une période donnée (d'une ou plusieurs années), pour un ou plusieurs MNTs.

L'orchestrateur permet la mise en place automatique d'un contexte, pour un super site, ceci en termes de paramétrage, d'arborescence et de mise en forme des données pour chacune des étapes traitées suivantes :

- ✓ la génération des masques d'eau à partir des observations satellites permettant de collecter les séries de surface en eau continentales,
- ✓ la modélisation des relations HSV pour chacun des MNT sélectionnés dans l'étude, suivi de la comparaison avec les modèles *in situ* de références disponibles pour les réservoirs étudiés,
- ✓ la génération des séries temporelles de surface $S(t)$, de volume $V(t)$ et de taux de remplissage $TR(t)$ pour chaque réservoir sur la base d'un MNT final sélectionné, suivi de la comparaison des métriques avec les séries temporelles *in situ* de référence.

Le workflow entre les chaines Surfwater, DEM4Water et Surfwater_Postprocess est représenté ci-dessous :

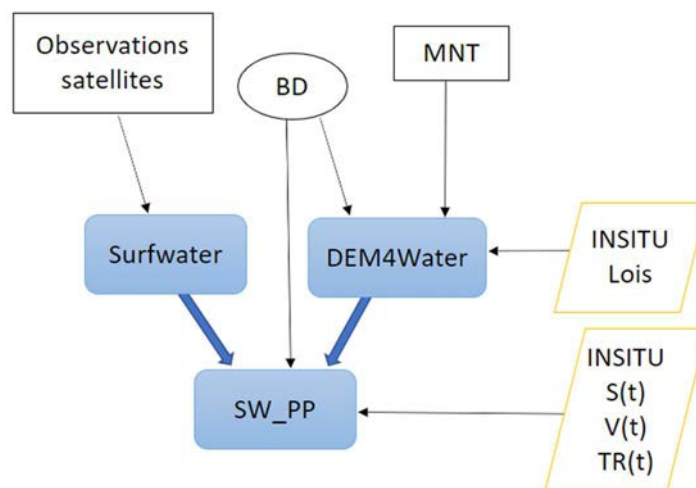


Figure 1 : Workflow Stockwater

Une analyse des résultats par un opérateur est fortement conseillée entre chaque étape de lancement de chaîne. La reprise des traitements entre étapes est simplifiée par l'utilisation de mots clefs true/false relatifs à l'activation des chaînes.

Dans un premier temps, l'arborescence de la campagne pour un site est automatiquement mise en place par l'orchestrateur.

1. Celui-ci gère également les lancements successifs de la chaîne **Surfwater** sur plusieurs années. L'opérateur devra intervenir d'une part en termes de contrôle des résultats de production des masques d'eau pour chaque tuile (exhaustivité des masques, prise en compte des différences de volume par site, éventuellement par année), et d'autre part dans la sélection finale des tuiles validées.
2. L'étape suivante de création des modèles estimés avec la chaîne **DEM4water** nécessite, avant son lancement, des prérequis qui sont réalisés par l'orchestrateur. Les produits Surfwater de toutes les années sont rassemblés et un calcul des occurrences est réalisé sur toute la période

de l'étude, puis un fichier vrt water map est généré. La chaîne DEM4water est ensuite lancée pour produire les modèles estimés et les comparaisons aux références *in situ* -si disponibles-. L'orchestrateur gère la possibilité de traiter un ou plusieurs MNTs. Les modèles estimés produits sont listés et concaténés en un unique fichier json nécessaire à l'étape suivante. L'opérateur devra effectuer un contrôle qualitatif des modèles estimés du fait de la variété des topographies locales des zones géographiques traitées et des formes des retenues d'eau pouvant atteindre les limitations méthodologiques de la chaîne et causer des anomalies dans les modèles.

3. Les séries temporelles $S(t)$, $V(t)$, $TR(t)$ sont générées par la chaîne **Surfwater_Postprocess** dont le fichier paramètre d'entrée est produit automatiquement par l'orchestrateur. Le traitement est effectué pour un modèle estimé choisi (via le fichier paramètre) et comparé à des séries temporelles de références -si disponibles-.

Par exemple pour un site, la production des modèles estimés réalisée par la chaîne DEM4water, sur un contexte de données de masques d'eau déjà produits, nécessite l'activation suivante dans le fichier paramètre de l'orchestrateur :

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

Lancement de l'orchestrateur

Après la mise à jour des différents champs requis par le fichier paramètre stockwater_params.json, le lancement de l'orchestrateur se fait via la commande suivante :

```
qsub -v WD=$PWD,CONF=stockwater_params.json launch_stockwater.pbs
```

Les environnements sont automatiquement installés successivement par les scripts pbs selon les étapes en cours de traitements.

Démonstration

La démonstration du fonctionnement de l'orchestrateur a été fournie (le 14/10/2021) via une description de la procédure de lancement sur un jeu de test du site Occitanie et de ses résultats attendus.

3.2 Mise en place de la base de données

3.2.1 Sélection de cibles

Lors de la phase préliminaire, un premier échantillon de réservoirs a été établi, avec des sites sélectionnés de façon à être le plus hétérogènes et représentatifs possible en termes de surfaces, de configuration topographique, ... et pour lesquels les données d'entrée, sont déjà disponibles (disponibilité de plusieurs MNTs, paramètres de validation, etc).

Une base de données de référence décrivant ces sites a été mise en place lors de la Phase I et adaptée aux besoins du projet au fur et à mesure de l'avancement de la Phase II.

Dans le cadre des activités complémentaires prévues du présent avenant, cette liste initiale est complétée à partir des sites retenus et intégrés dans la base de données lors de la Phase II.

La sélection des sites d'intérêt repose sur l'analyse des bases de données lacs décrites ci-dessous. Les sites sont considérés comme compatibles lorsque les résultats sont pertinents (emprises, validité des métadonnées...), que les sites sont facilement exploitables (de préférence sur une seule emprise S1/S2), mais également que des données de validation *in situ* exploitables soient disponibles.

Les bases de données ayant été explorées sont les suivantes :

- ✓ BD Lacs de SWOT, réalisée à partir de Circa 2015 de Y. Sheng,
- ✓ GOOD2,
- ✓ GRanD,
- ✓ La BD GEODAR qui intègre les BDs GRanD et GOOD et les complète,
- ✓ La BD indienne officielle, issue de la Water Central Commission 1.

Tableau 1 : synthèse des bases de données disponibles

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

Plus précisément, les critères de choix final sont les suivants :

- ✓ La disponibilité de données *in situ*, entre autres le H de pied de barrage et le volume de remplissage maximal sur le réservoir, et leur pertinence par rapport à la réalité (par exemple décalage entre la hauteur du barrage et le niveau maximum de remplissage...),
- ✓ La disponibilité de données MNT ou MNS (Copernicus, CARTOSAT, Alos, SRTM, Stéréo Pléiades...),
- ✓ La variabilité de surface marquée sur la carte d'occurrence de Global Surface Water (GSW, Pekel et al 2016),
- ✓ La bonne adéquation entre le contour proposé par les BD Réservoirs et l'occurrence de GSW,
- ✓ Dans la mesure du possible, une taille minimale de 10 ha pour la Phase I (5 ha à étudier pour la Phase II).

3.2.2 Constitution des datasets de la Phase I

La campagne initiale, destinée à constituer les *datasets* utilisés pour la phase I, a porté sur l'Andalousie, l'Occitanie et l'Inde. Dans la mesure du possible, nous avons sélectionné les sites les plus pertinents, présents dans la base de données GRanD, qui se trouve être la plus complète.

Les réservoirs qui ne sont jamais totalement couverts par une tuile ont été écartés au maximum, les réservoirs entièrement couverts par plusieurs tuiles ont été systématiquement sélectionnés afin de doubler la fréquence d'observation.

La méthodologie globale suivante a été adoptée pour la constitution du *dataset*, avec quelques

1 <http://cwc.gov.in/national-register-large-dams>

adaptations en fonction de la complexité et des imprévus rencontrés sur les différentes zones :

Après analyse des différentes bases de données géoréférencées disponibles, il s'est avéré que les extensions maximales issues de GRanD étaient globalement de mauvaise qualité avec une tendance à sous-estimer ces extensions maximales (cf. Figure 2).

Nous avons donc choisi d'extraire les extensions maximales de chaque réservoir à partir de GSW 2020, qui s'est avéré plus précis et plus représentatif de la réalité terrain, car prenant en compte la dynamique des réservoirs sur plusieurs années. Un raffinement manuel des extensions maximales issues de GSW a également été effectué (cf. Figure 3).

Les attributs issus de GRanD ont ensuite été associés à chacune de ces extensions, lorsque ceux-ci étaient présents dans cette base de données. Les données de référence (Hauteurs d'eau, Maximum Storage, ...) manquantes dans GRanD ont été recherchées manuellement sur Google Earth et autres sites externes.

Pour chacun des réservoirs, les points de localisation de l'eau et de localisations des barrages ont tous été validés/corrigés. Les altitudes des barrages ont également été corrigées à partir de Google Earth et des altitudes maximales des données *in situ* disponibles.

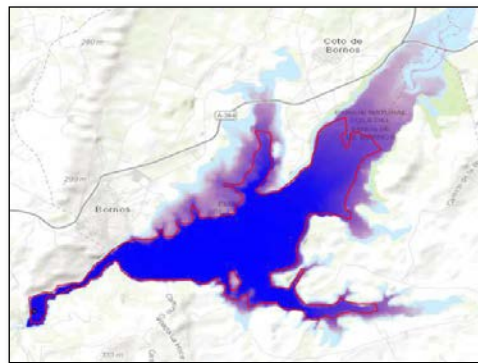


Figure 2 : Bornos / Guadalete River, Red outline for GRanD /GeoDAR extent, GSW occurrence in blue (1984-2020)



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

Pour la phase I, la cible initiale était de 50 sites. Pour finir, le lot stabilisé et représentatif livré regroupe 75 sites dont :

- ✓ 37 en Andalousie,
- ✓ 16 en Occitanie,
- ✓ 22 en Inde.

3.2.3 Campagne Andalousie

37 réservoirs, considérés comme les plus prometteurs en se basant sur les critères de sélections, ont été retenus pour la campagne Andalousie.

20 réservoirs, sur les 37 sélectionnés, disposent de données *in situ* (mesures régulières de Hauteur d'eau et volume).



Figure 4 : Carte de localisation – Andalousie

3.2.4 Campagne Occitanie

Pour la campagne Occitanie, 16 réservoirs présents dans GRanD ont été sélectionnés pour lesquels les données *in situ* sont disponibles.

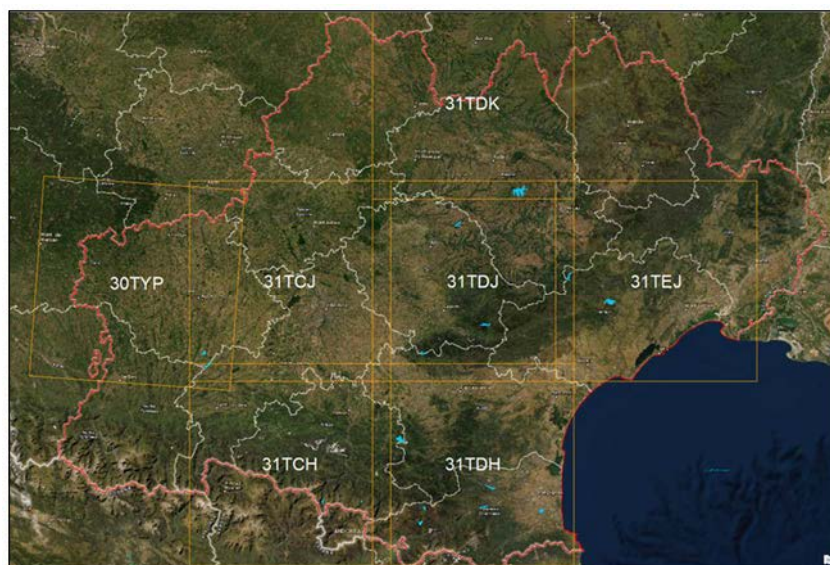


Figure 5 : Carte de localisation Occitanie

3.2.5 Campagne Inde

Pour la campagne Inde, il a été décidé, dans un premier temps, de privilégier les emprises des scènes disponibles sur THEIA. (cf. Figure 6) En se fondant sur ce critère et sur les réservoirs présents dans GRanD, et/ou disposant de données *In situ*, 22 réservoirs ont été sélectionnés sur l'Inde.

La sélection des sites Inde est apparue comme la plus complexe, car une grande partie des réservoirs présents dans les données *in situ* fournies n'était pas présents dans les bases de données géoréférencées à notre disposition (GRanD, GEODAR, ...). Il a donc fallu rechercher manuellement la localisation des réservoirs présents dans la zone couverte par les données THEIA et répertoriés dans les fichiers des données *in situ*.

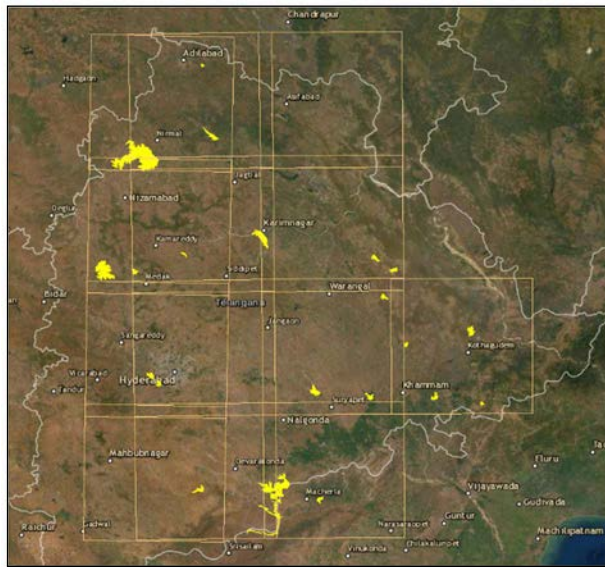


Figure 6 : Carte de localisation – Inde

Pour finir, seulement 8 réservoirs sur les 22 sélectionnés disposent de données *in situ*. Sept des réservoirs disposant de données *in situ* étaient présents dans la base GRanD et un seulement dans GEODAR. Il n'a malheureusement pas été possible de localiser des réservoirs supplémentaires répondant aux critères (données *in situ* disponibles et présent dans la zone couverte par THEIA).

Tableau 2 : Inventaire des données *in situ* disponibles sur les réservoirs identifiés pour la Phase I

Reservoir Name	DISTRICT	BASIN	DATABASE
Almatti Reservoir	BIJAPUR	KRISHNA	Hors emprises
Narayanapura Reservoir	BIJAPUR	KRISHNA	Hors emprises
Gobind SagarBakra Reservoir	BILASPUR	INDUS (UP TO BORDER)	Hors emprises
BUGGAVAGU	GUNTUR	KRISHNA	GRanD
NAGARJUNA SAGAR	GUNTUR	KRISHNA	GRanD
PRAKASAM BARRAGE	GUNTUR	KRISHNA	Hors emprises
Srisaillam Reservoir	GUNTUR	KRISHNA	Hors emprises
Srisaillam 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
VUAYARAI 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

3.2.6 Préparation des données

3.2.6.1 Structure du fichier attributs

Cette activité a porté sur la finalisation du format de la base de données, avec l'évaluation de 6 versions successives, pour finalement, obtenir la structure suivante au cours de Phase II.

Tableau 3 : Structure attributs dataset

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

Les NO DATA sont renseignés en -999 pour les champs numérique et NULL pour les champs texte.

On notera que plus de six versions ont été réalisées avant de convenir de cette structure, considérée aujourd'hui comme finale.

Les différents datasets sur l'Andalousie, l'Occitanie et l'Inde ont ensuite été produits aux formats shapefile, geojson et CSV.

3.2.6.2 Données de référence *in situ*

3.2.6.2.1 HAUTEURS D'EAU ET VOLUMES *IN SITU*

Afin de valider les résultats obtenus *via* STOCKWATER, un certain nombre de données *in situ* ont été fournies par différents organismes. Ces données ont fait l'objet d'une analyse systématique, et des irrégularités importantes ont été mises en évidence.

- ✓ Occitanie : Données fournies par la DREAL Occitanie.
- ✓ Andalousie : Données fournies par la Confederacion Hidrografica del Guadalquivir.
- ✓ Inde : Données fournies par l'ISRO.

Une activité importante d'analyse, de sélection et de correction (selon différentes méthodes) a été effectué jusque tardivement lors de la Phase II, pour générer des datasets opérationnels, compatibles avec le fonctionnement du PoC STOCKWATER.

Au-delà des actions menées lors de la Phase II, les travaux nécessaires à la Phase de valorisation ont montré que des anomalies (ou en tout cas des causes de dysfonctionnement) subsistent et nuisent

notamment au fonctionnement du tableau de bord. Une nouvelle action de vérification systématique et correction de la base de données sera menée dans le cadre des activités complémentaires prévues.

Cependant, après analyse, sélection (sites disposant de données *in situ* exploitable) et correction, il a été possible de générer un dataset suffisant pour la réalisation de la Phase I du projet :

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

Tableau 4 : Données in situ Occitanie

Nous disposons de données *in situ* (Hauteur d'eau, Volume) sur l'ensemble des 16 sites sélectionnés. Ces données sont issues de différents questionnaires. Elles ont été agrégées au même format par le MTE et envoyées à travers la DREAL.

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

Tableau 5 :Données in situ Andalousie

Nous disposons de données *in situ* (Hauteur d'eau, Volume) sur les 33 réservoirs sélectionnés. Ces données sont issues de deux sources différentes : SAIH Guadalquivir et SAIH Hidrosur.

Tableau 6 : Données in situ Inde

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						
Lakshnavaram					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						

Concernant les sites Inde, 4 envois de données *in situ* issues de sources différents ont été effectués par Sylvain FERRAND et son thésard Abhilash Kumar Paswan.

- ✓ DB1, DB3 et DB4 sont les données journalières produites par la Central Water Commission (CWC, agence fédérale), qui sont mises à jour régulièrement à chaque arrivage de données d'agences d'états ou privées. Ce sont des données issues de mesures de terrain et de tarage par réservoir,
- ✓ DB2 : produit issu de la télédétection et plus précisément de LISS IV et Cartosat-1.

On notera que plusieurs mois se sont passés entre l'envoi de la DB1 et la DB4 qui a été fournie le 20/10/2021.

3.2.6.2.2 SELECTION DES DONNEES

Les données de référence fournies, présentent des différences significatives et ne sont pas forcément homogènes. Une analyse de comparaison a été faite sur les volumes des 4 réservoirs communs aux lots livrés. Cette analyse montre des différences considérables entre les jeux de données. (cf. Figure 7, 8, 9, 10).

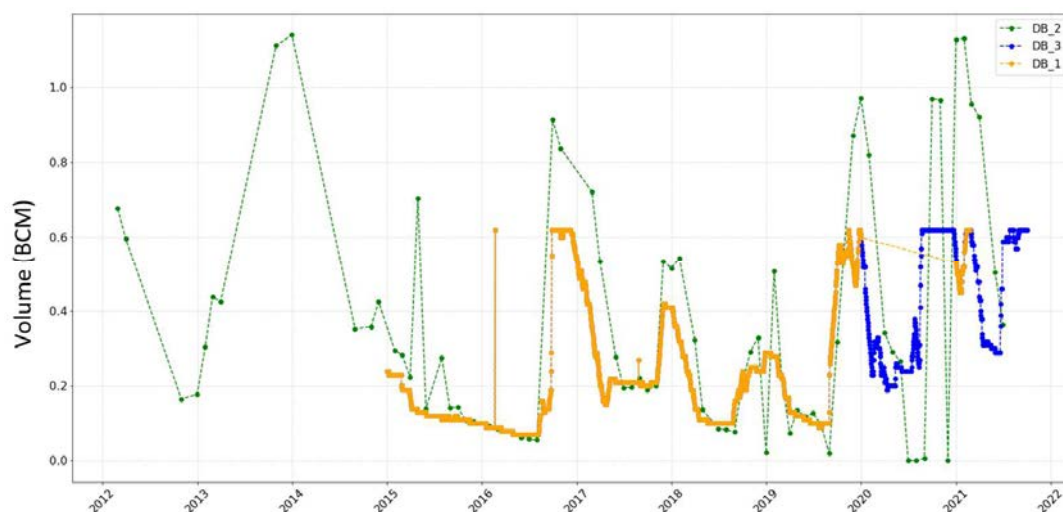


Figure 7 : Comparaison volume - Lower Manair

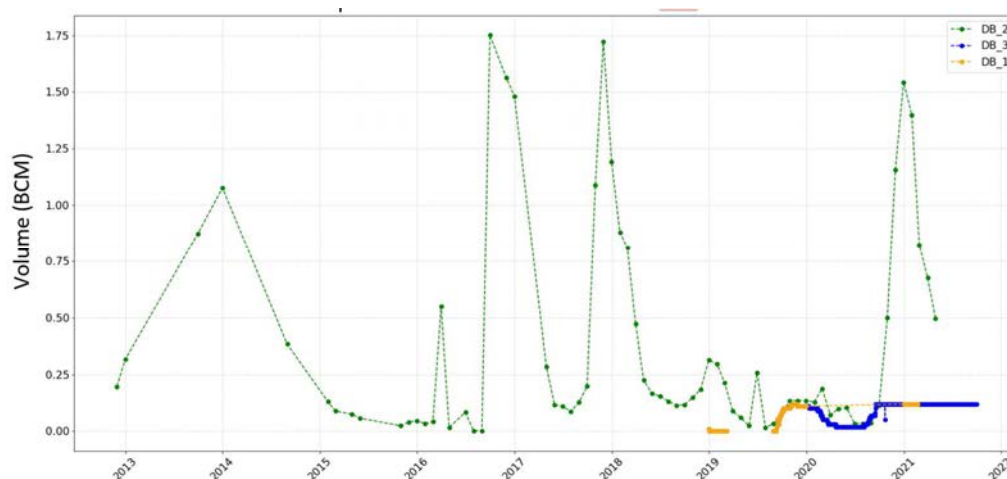


Figure 8 : comparaison volume - Nizam Sagar

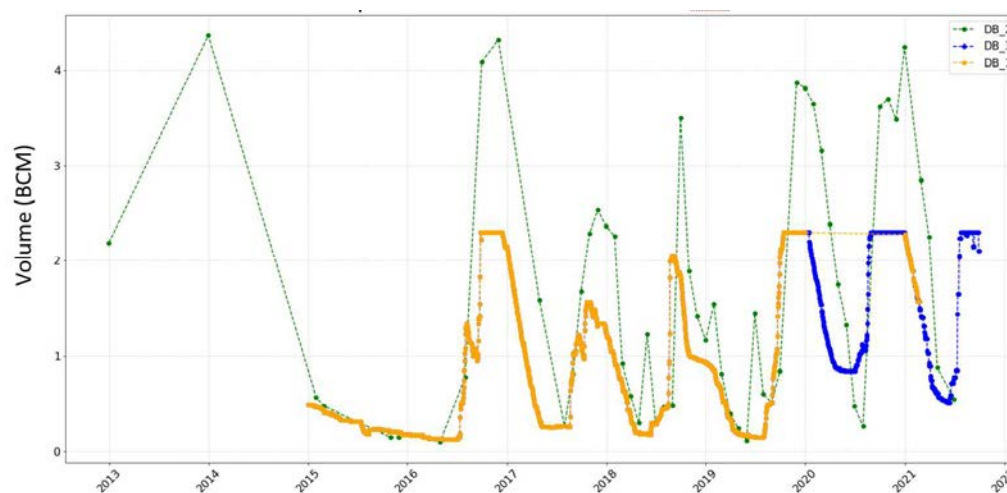


Figure 9: comparaison volume - Sriram Sagar

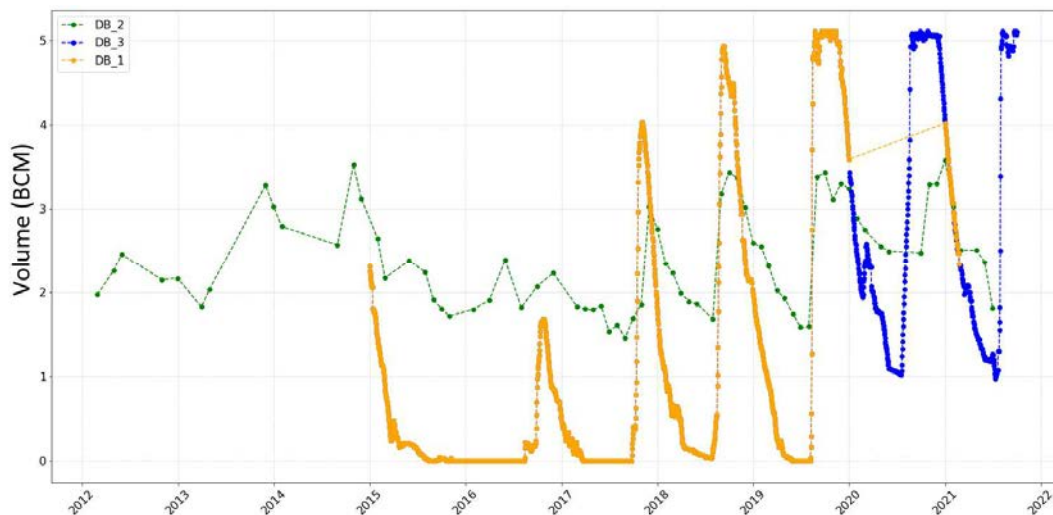


Figure 10 : comparaison volume - Nagarjuna Sagar

Les remarques suivantes ont été effectuées à partir des graphiques de comparaison précédents :

- ✓ Les bases de données 1 et 3 ont des valeurs de volumes identiques,
- ✓ On observe des différences importantes entre les bases de données 1/3 et 2,
- ✓ Plus grande densité de données dans les bases de données 1 et 3 que dans la 2,
- ✓ La base de données 3 est plus complète (plus de valeurs que la 1),
- ✓ En vue des valeurs issues de la base de données 2 et le fait qu'elle soit issue de solutions satellites la rend difficilement utilisable comme référence de qualité.

Au final, en prenant en compte les remarques précédentes, les données *in situ* suivantes ont été sélectionnées pour 8 réservoirs. Malgré le fait que l'on note encore des incohérences dans certains jeux de données, il a été choisi de conserver ces jeux de données afin d'observer le comportement des algorithmes de génération des lois face à ce type de données de mauvaise qualité, qui pourront être écartées par la suite.

SWOT_ID	DAM_NAME	Values	
		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	Lakshnavaram		
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		

Tableau 7 : Résumé des données *in situ* retenues

Nota bene : Des informations complémentaires/données *in situ* peuvent être extraites des bases de données locales (tableau 7 et figure 11), mais une partie des sites fournis présentent des anomalies car la variabilité saisonnière n'est pas prise en compte, et ces données ne recoupent pas intégralement la liste des sites fournis. Elles ont donc été écartées, car elles sont difficilement utilisables dans le cadre de cette étude.

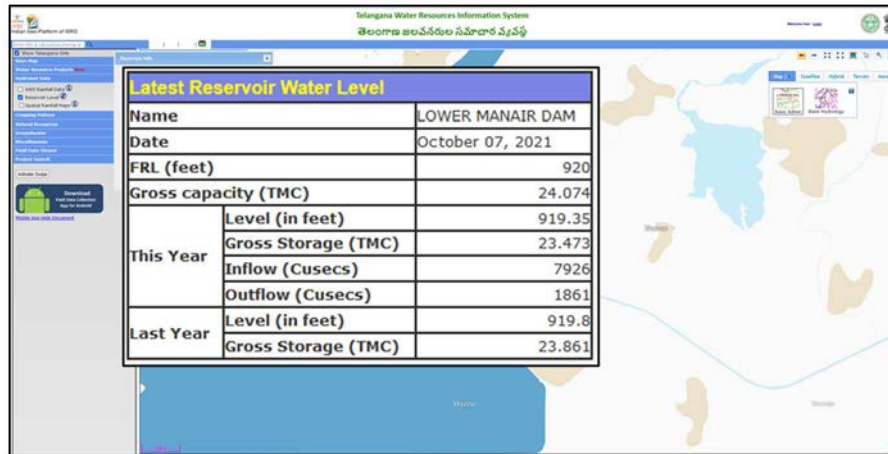


Figure 11: Telangana Water Ressources Information System

3.2.6.2.3 ANALYSE DE LA QUALITE

Après sélection des données *in situ* (Hauteurs d'eau et Volumes) considérées comme les plus pertinentes, une première génération des lois paramétrées *in situ* Inde, incluant également le calcul des Surfaces dérivées des Hauteurs d'eau et Volumes *in situ* a été réalisée à l'aide de l'outil « timeseries_tool.py » du CNES. Cette première étape a permis de déceler des incohérences dans ces données de Hauteurs d'eau et de Volumes de référence.

En effet, nous avons remarqué que les volumes et hauteurs *in situ* Inde sont globalement de mauvaise qualité et peuvent être difficilement utilisées. Cela mène à des incohérences au niveau de la génération des surfaces *in situ* générées par le script CNES, comme le montrent les analyses suivantes.

Les données *volume* (volumes) et *level* (hauteur d'eau) *in situ* pour les réservoirs 4530187293 - Nizam Sagar et 4530308913 - Pulichinthala Project sont de très mauvaise qualité avec un grand nombre d'incohérences (données manquantes et en escalier (plusieurs valeurs de level pour un volume donné) (cf. Figure 12).

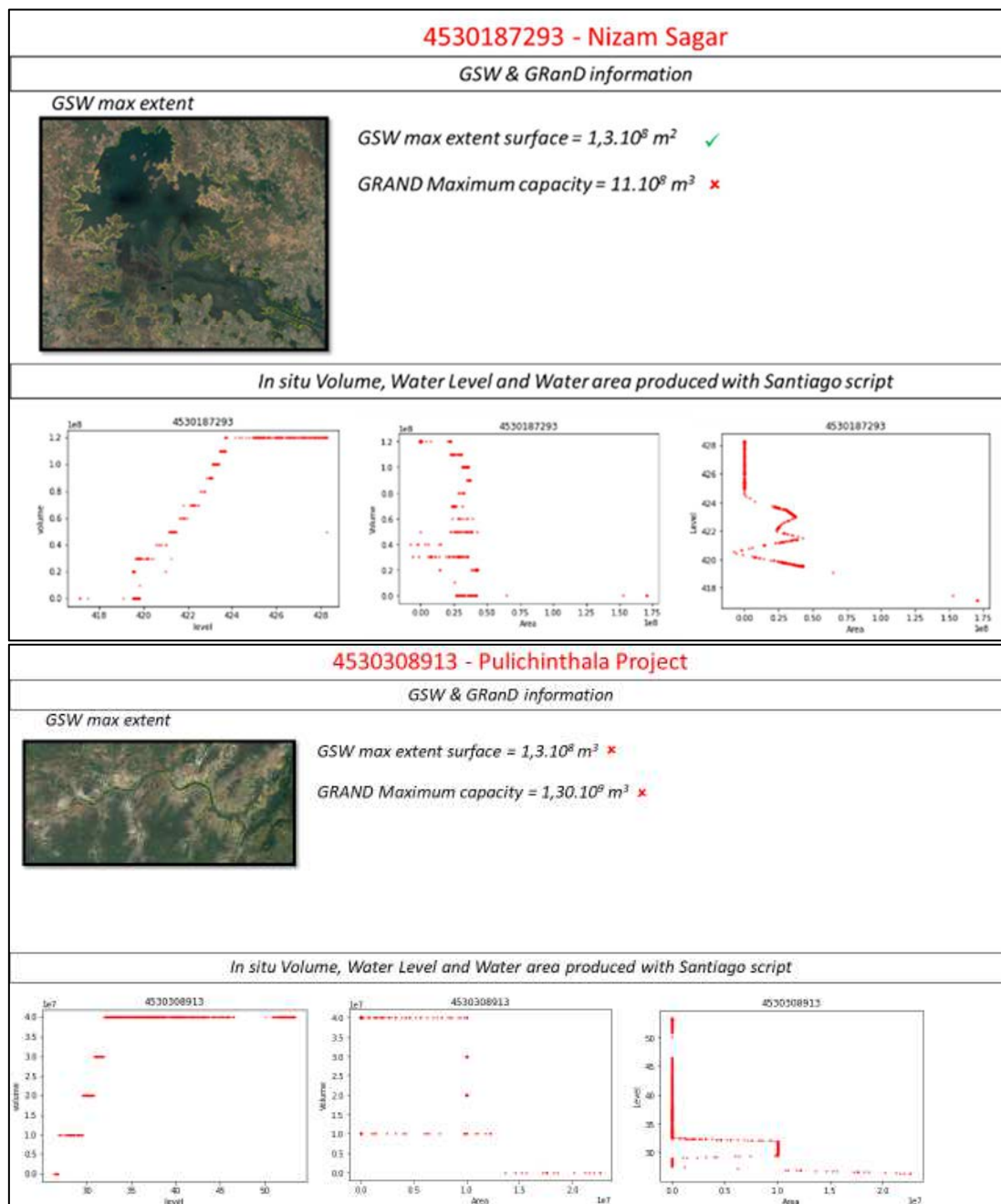


Figure 12 : Données in situ Nizam Sagar & Pulichintala Project

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).

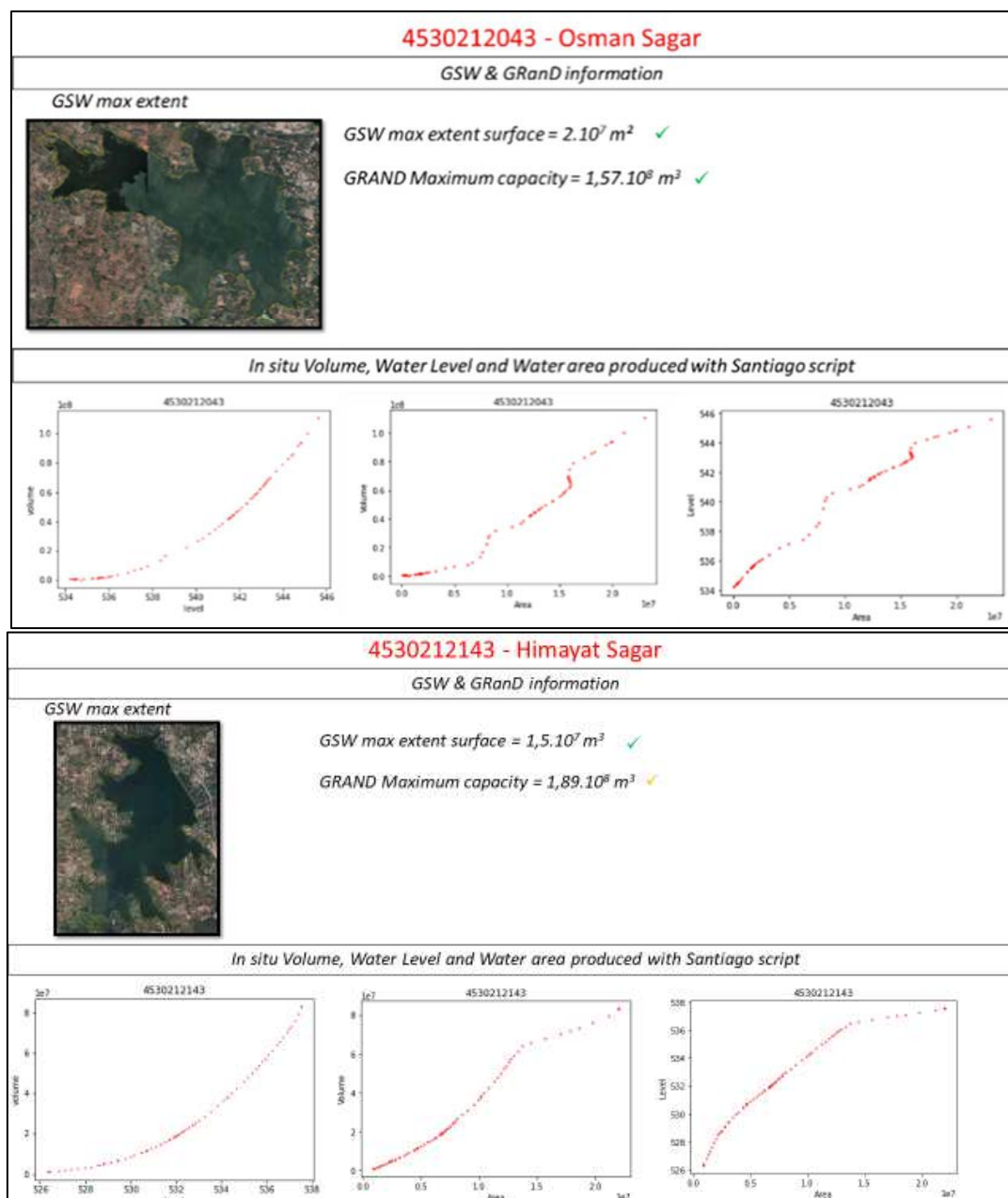


Figure 13 : Données in situ Osman Sagar & Himayat Sagar

Les données volume, level et surface *in situ* pour les réservoirs 4530240913 - Sriram Sagar et 4530313823 – Buggavagu semblent globalement correctes malgré la présence de quelques données incohérentes à ne pas prendre en compte (outliers, plusieurs valeurs de level pour un volume donné) (cf. Figure 14).

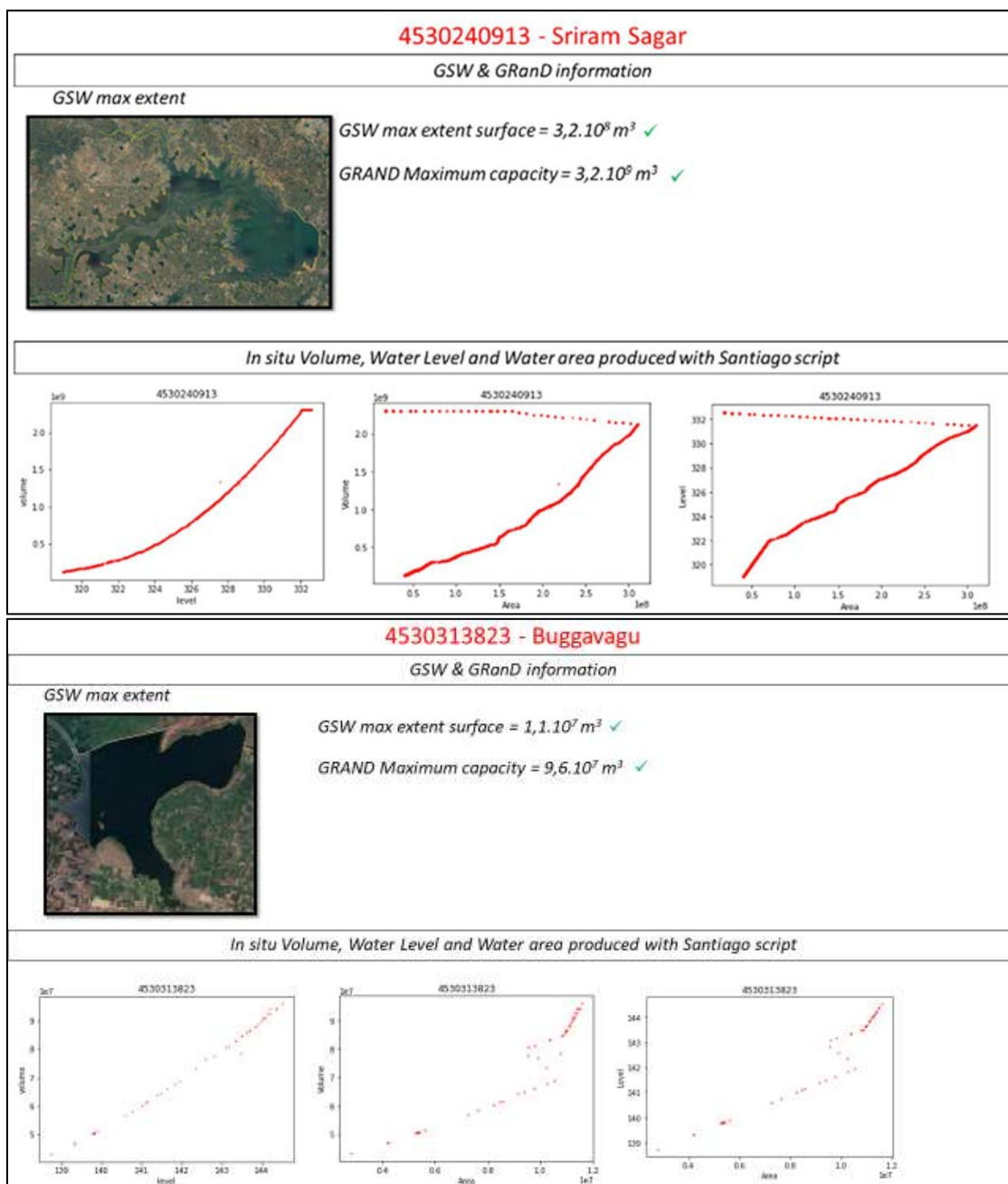


Figure 14 : Données in situ Sriram Sagar & Buggavu

On observe un grand nombre d'incohérences dans les données volume, level et surface *in situ* des réservoirs 4530433353 - Lower Manair et 4530347283 - Nagarjuna Sagar, avec des données en « escalier », un certain nombre d'outliers, ainsi que des « plateaux » (plusieurs valeurs de level pour un volume donné) (cf. Figure 15).

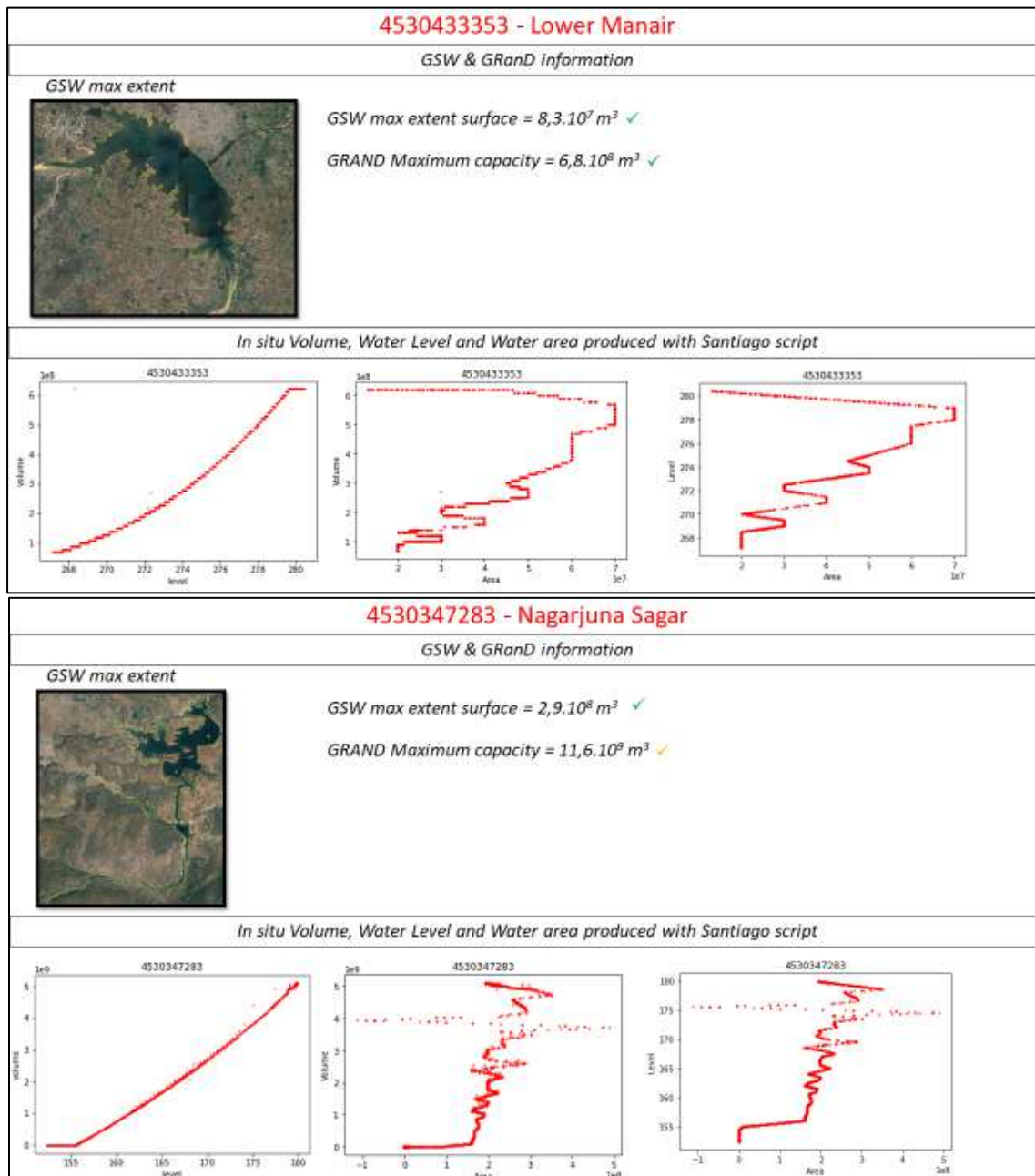


Figure 15 : Données in situ Lower Manair & Nagarjuna Sagar

3.2.6.2.4 AMELIORATION DES DONNEES *IN SITU*

Considérant les problèmes observés sur ces données décrits précédemment, un outil a été mis en place pour « améliorer » la qualité des données de certains réservoirs et les rendre utilisables pour les étapes de génération des lois *in situ*.

Cet outil fonctionne en 3 étapes :

- ✓ Etape 1 : Suppression des « plateaux » / données non-variables. On entend par cela la suppression des mesures pour lesquelles on observe plus de X valeurs de level pour un volume donné ou plus de X valeurs de volume pour un level donné (cf. Figure 16).

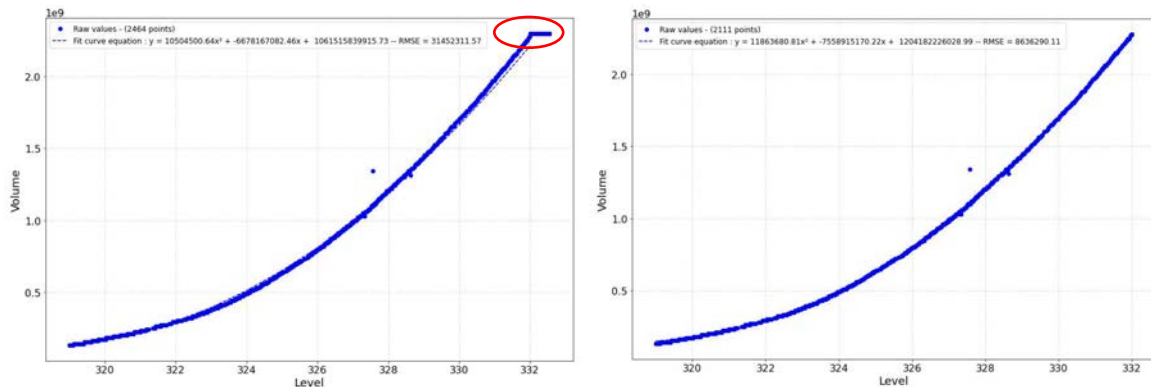


Figure 16 : Suppression des "plateaux" - Sriram Sagar

- ✓ Etape 2 : Suppression des *outliers* (cf. Figure 17).

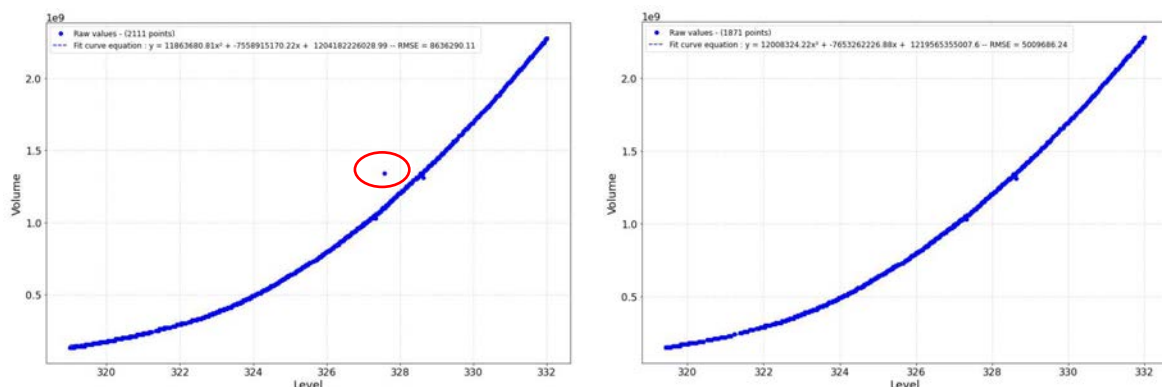


Figure 17 : Suppression des outliers - Sriram Sagar

- ✓ Etape 3 : Ajuster/*Fit* les valeurs en escaliers en fonction de l'équation de la courbe V(L) (avec V = Volume et L = Level) calculée à partir des données issues de l'étape 3. Cela consiste à recalculer les mesures pour lesquelles on observe plus de 3 valeurs de level pour un volume donné ou plus de 3 valeurs de volume pour un level donné en fonction de l'équation de la courbe V(L) (avec V = Volume et L = Level) (cf. Figure 18).

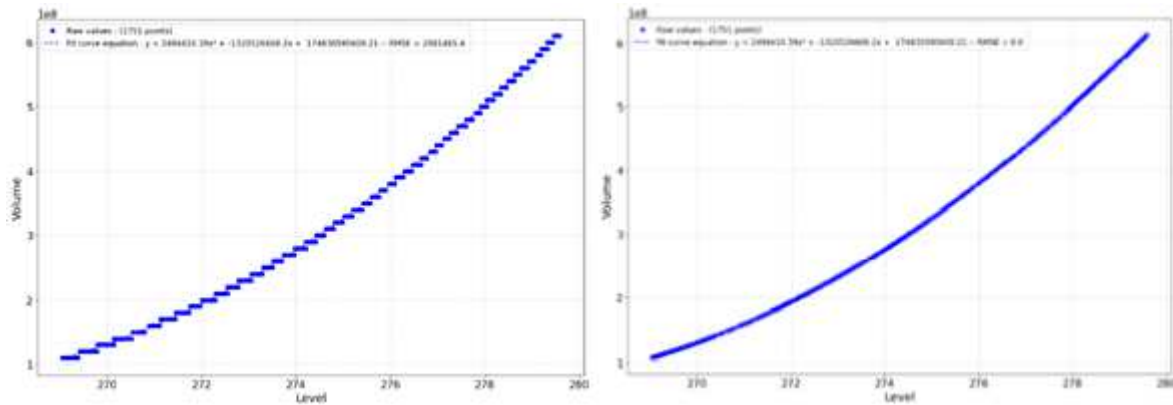


Figure 18 : Ajustement des valeurs en escaliers - 4530433353 - Lower Manair

Ces 3 étapes ont été appliquées sur les « volume » et « level in situ » des réservoirs suivants :

- ✓ 4530240913 - Sriram Sagar,

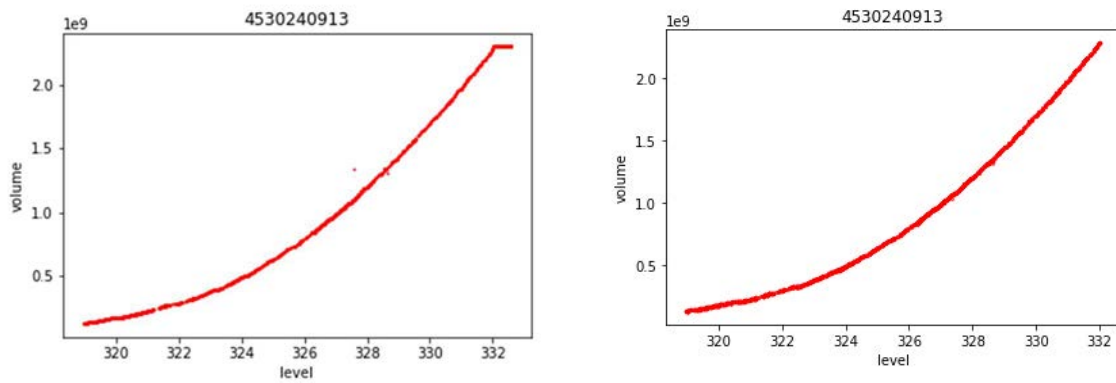


Figure 19 : Données in situ brutes (gauche) et modifiées (droite) - 4530240913 - Sriram Sagar

- ✓ 4530313823 – Buggavagu

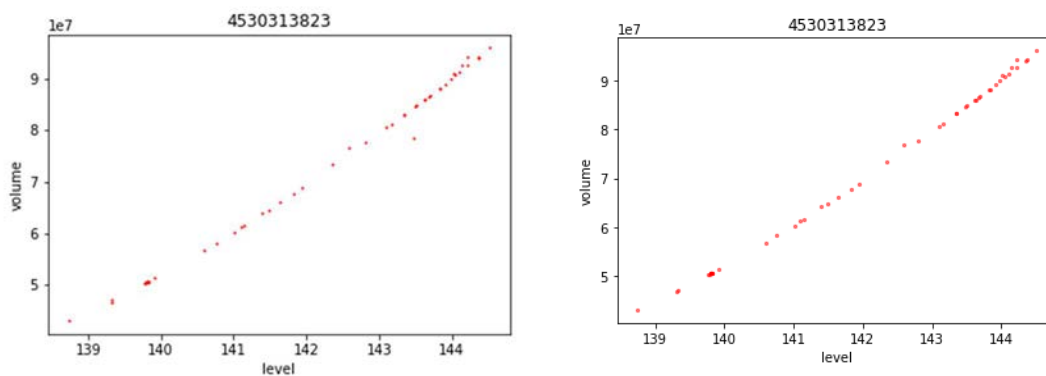


Figure 20 : Données in situ brutes (gauche) et modifiées (droite) - 4530313823 – Buggavagu

✓ 4530347283 - Nagarjuna Sagar

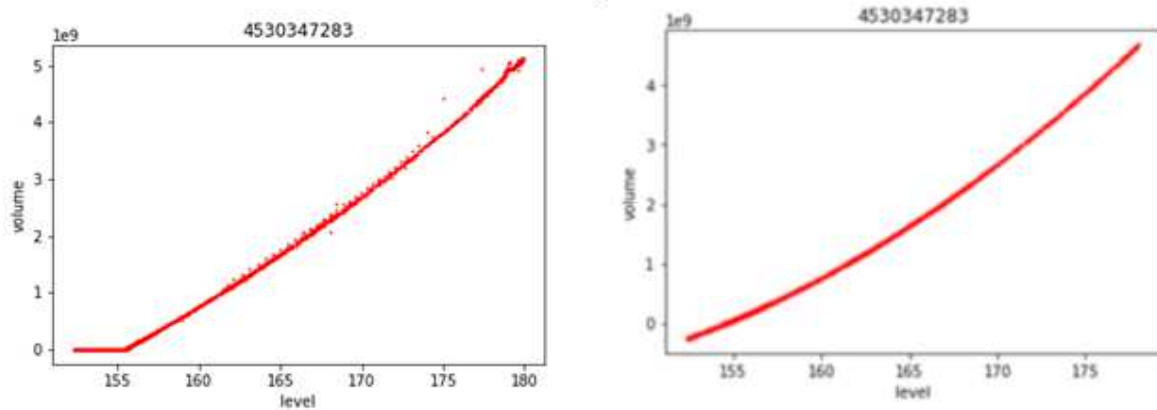


Figure 21 : Données in situ brutes (gauche) et modifiées (droite) - 4530347283 - Nagarjuna Sagar

✓ 4530433353 - Lower Manair

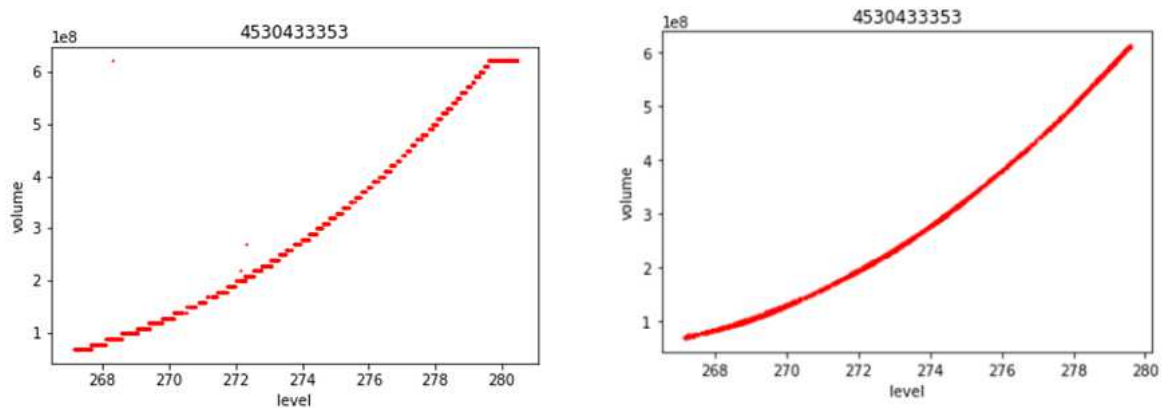


Figure 22 : Données in situ brutes (gauche) et modifiées (droite) - 4530433353 - Lower Manair

Les données *in situ* des réservoirs 4530212043 - Osman Sagar et 4530212143 - Himayat Sagar n'ont pas été modifiées car leur qualité a été jugée suffisante.

Il a été décidé d'écarter les données *in situ* des réservoirs 4530187293 - Nizam Sagar et 4530308913 - Pulichinthala Project du jeu de données de référence, du fait de leur trop mauvaise qualité.

Pour certains jeux de données, les mesures sont fortement modifiées/impactées, particulièrement par l'étape 3. Une réflexion est donc à mener sur le fait d'écarter certains jeux de données qui ont été trop fortement transformés ou si cette « transformation » est considérée comme « acceptable ». Exemple pour les réservoirs 4530433353 - Lower Manair et 4530347283 - Nagarjuna Sagar.

Nota Bene :

Afin d'améliorer la qualité des données *in situ* Inde, d'autres méthodes de lissage (*Savitzky-Golay filter* et *Lowess smoother*) ont également été testées afin de « nettoyer » ces données. Après comparaison des résultats issus de ces méthodes avec les résultats obtenus via la méthode décrite précédemment, nous avons décidé de retenir la méthode précédemment décrite qui offrait de meilleurs résultats.

La première méthode de lissage est basée sur la régression locale de LOESS « LOcally Estimated Scatterplot Smoothing ».

C'est une méthode de régression non paramétrique qui a l'avantage de calculer autant de fonctions locales qu'il y a de segments de données à l'aide de la méthode des k plus proches voisins afin d'ajuster au mieux les courbes. Cependant, elle exige une quantité de jeux de données importante pour générer

de bons modèles. Par conséquent, la correction des séries $V(Z)$ n'a pas pu être réalisée avec cette première approche.

La deuxième méthode consiste à appliquer le filtre Savitzky-Golay sur les données $V(t)$ et $Z(t)$ indépendamment. Ce filtre approche donc la fonction $V(Z)$, en ne conservant que les caractéristiques importantes et en se débarrassant des fluctuations sans signification. Des sous-ensembles successifs de points sont ajustés avec une fonction polynomiale qui minimise l'erreur d'ajustement.

En ce qui concerne les paramètres d'entrée, il faudrait renseigner les données à filtrer, choisir la taille de la fenêtre correspondant au nombre de points utilisés pour calculer l'ajustement ainsi que l'ordre de la fonction polynomiale.

Les figures suivantes montrent certains résultats variant en fonction de la taille de fenêtre sélectionnée indiqué par le paramètre « Smoothing factor ».

✓ 4530212043 - Osman Sagar,

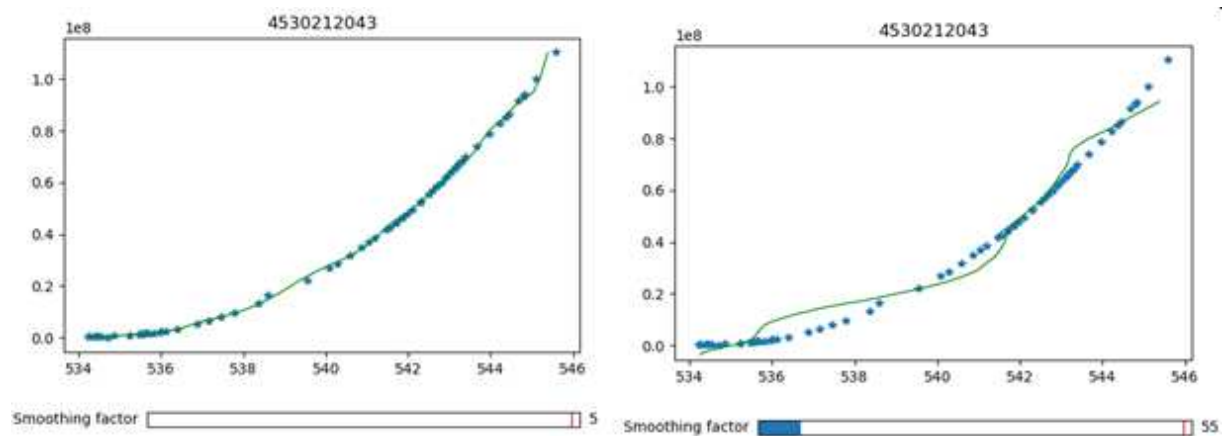


Figure 23 : données *in situ* brutes (bleu) et modifiées (vert) - 4530212043 - Osman Sagar

✓ 4530433353 - Lower Manair,

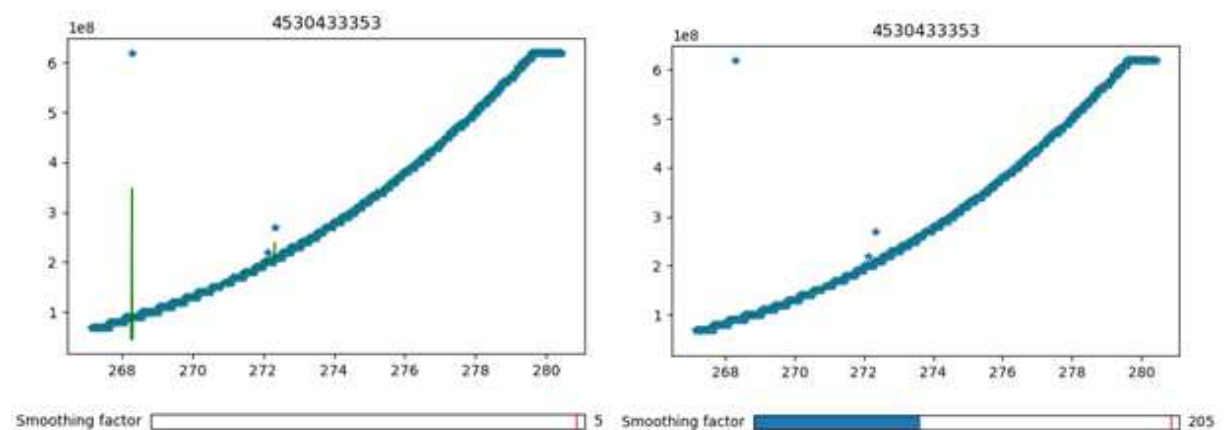


Figure 24 : données *in situ* brutes (bleu) et modifiées (vert) - 4530433353 - Lower Manair

✓ 4530347283 - Nagarjuna Sagar,

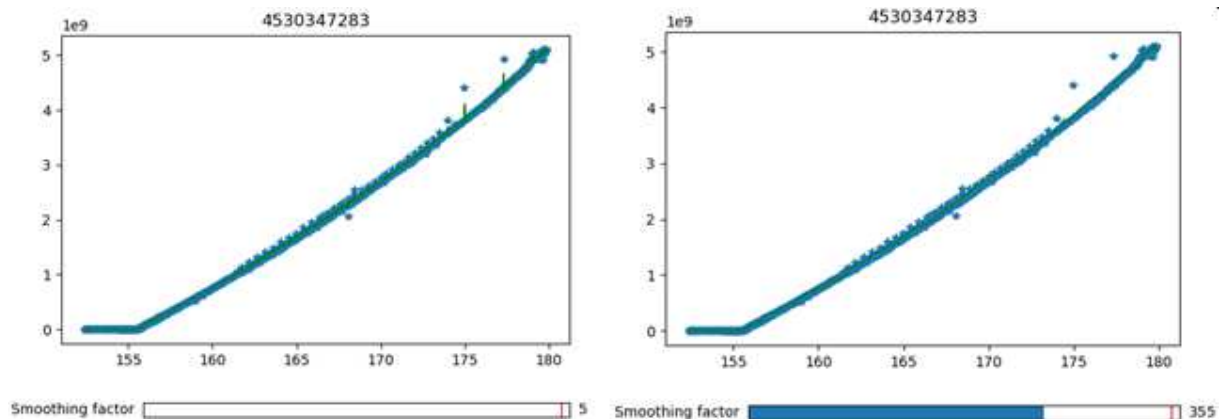


Figure 25 : données *in situ* brutes (bleu) et modifiées (vert) - 4530347283 - Nagarjuna Sagar

✓ 4530187293 - Nizam Sagar

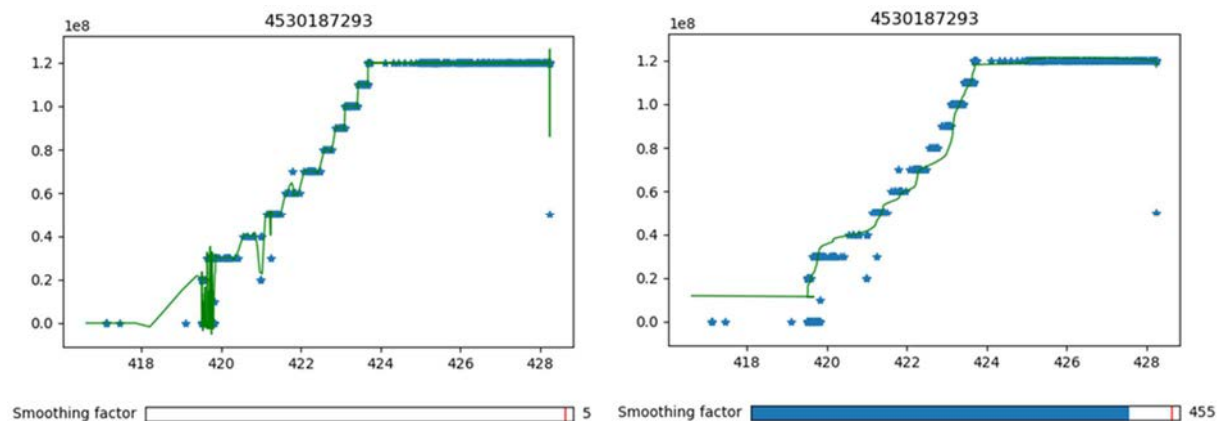


Figure 26 : données *in situ* brutes (bleu) et modifiées (vert) - 4530187293 - NIZAM SAGAR

Le filtre de Savitzky-Golay permet de suivre la tendance de la fonction originale $V(Z)$ et d'écarter les *outliers* mais n'agit pas complètement sur les « plateaux » / données non-variables en début ou fin de série. Ce qui nécessitera toujours un pré-nettoyage de cette partie de données. De plus, l'effet « escaliers » le long de la courbe (cf. Figure 24) est généralement amorti, ce qui empêche d'avoir une courbe bien lisse. Les résultats peuvent être améliorés en choisissant la taille de fenêtre la plus optimale.

Cette méthode fonctionne sur tous les réservoirs en Inde testés, sauf pour les réservoirs 4530187293 - Nizam Sagar (cf. Figure 27) et 4530308913 - Pulichintha Project pour lesquels le lissage des courbes n'a pas été concluant à cause de la mauvaise qualité des données brutes.

3.2.6.3 Génération et comparaison des lois *in situ* (voir Annexe B)

Dans un premier temps, il a été nécessaire de générer l'ensemble des lois *in situ* à partir des données de volumes et de hauteurs d'eau *in situ* (modifiées ou non) sur les 3 sites (Andalousie : 22 réservoirs avec données *in situ*, Occitanie : 16 réservoirs avec données *in situ* et Inde : 6 réservoirs avec données *in situ*). Ces lois ont été générées en utilisant l'outil « timeseries_tool.py » du CNES.

On notera que plusieurs évolutions de l'outil ont été effectuées durant la Phase I, ce qui a entraîné la régénération des lois *in situ* un certain nombre de fois (plus de 3 versions pour chaque site).

Dans le but d'évaluer les lois DEM4water il est nécessaire de :

- ✓ Evaluer $S(Z)$, $V(S)$
- ✓ Déterminer les sources à évaluer :
 - Référence : relations In situ (pas de lois),
 - Lois in situ,
 - Lois dem4water,
 - Modification des lois in situ par ajout des Z_{min} , Z_{max} , S_{min} , S_{max} ou Elévation(dam) et max area.

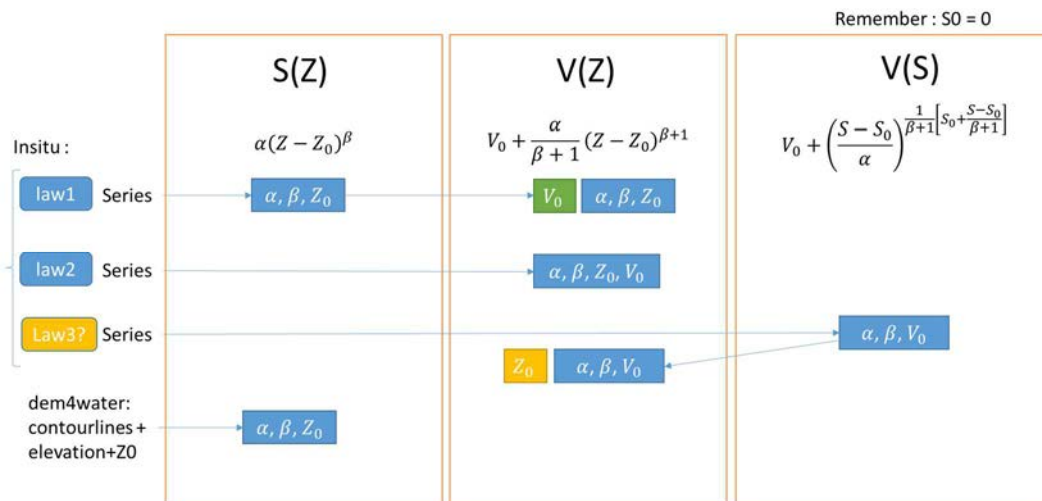


Figure 27 : Relations entre les lois et les paramètres associés

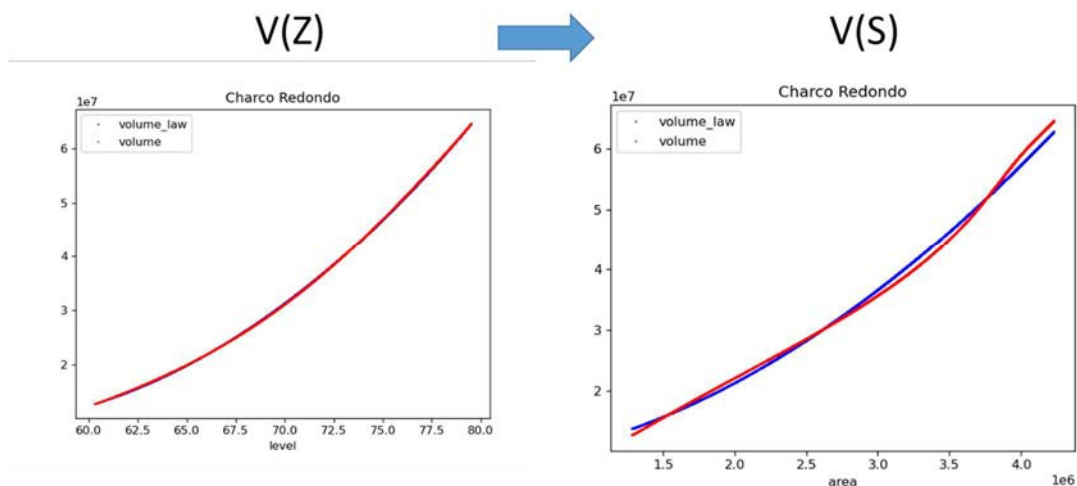


Figure 28 : Représentation des lois $V(Z)$ et $V(S)$

3.2.7 Analyse de la qualité des MNTs

Dans le but de déterminer le meilleur choix de MNT pour lancer la production, une analyse de qualification des MNT est faite avec DemCompare sur une tuile S2, en prenant comme référence le MNT IGN 5m et cela en fonction des paramètres suivants :

- ✓ Aspect géométrique (coregistration),
- ✓ Différence simple,
- ✓ Analyse des pentes.

L'analyse est faite sur l'Occitanie sur les MNTs suivants :

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

Nota Bene : Le SERTIT a fourni à titre d'illustration ce DEM Copernicus 10m sur une région restreinte d'Occitanie, sachant qu'il n'était d'une part disponible que sur Europe, et d'autre part que ce MNT est sous licence ; le CNES ayant bien une licence mais celle-ci n'était pas ouverte pour Stockwater.

On peut constater des différences d'altitude entre les DEM et la référence (cf. Figure 29). Le DEM ALOS présente une plage d'erreur plus importante, et de plus fortes erreurs de géolocalisation des erreurs décimétriques voire pluri décimétriques, alors que sur les DEM Copernicus et SRTM, ces erreurs sont pluri métriques. De plus, sur le DEM ALOS on observe des effets de *stripping*, qui induisent un affaissement en bande de l'altitude de ce DEM.

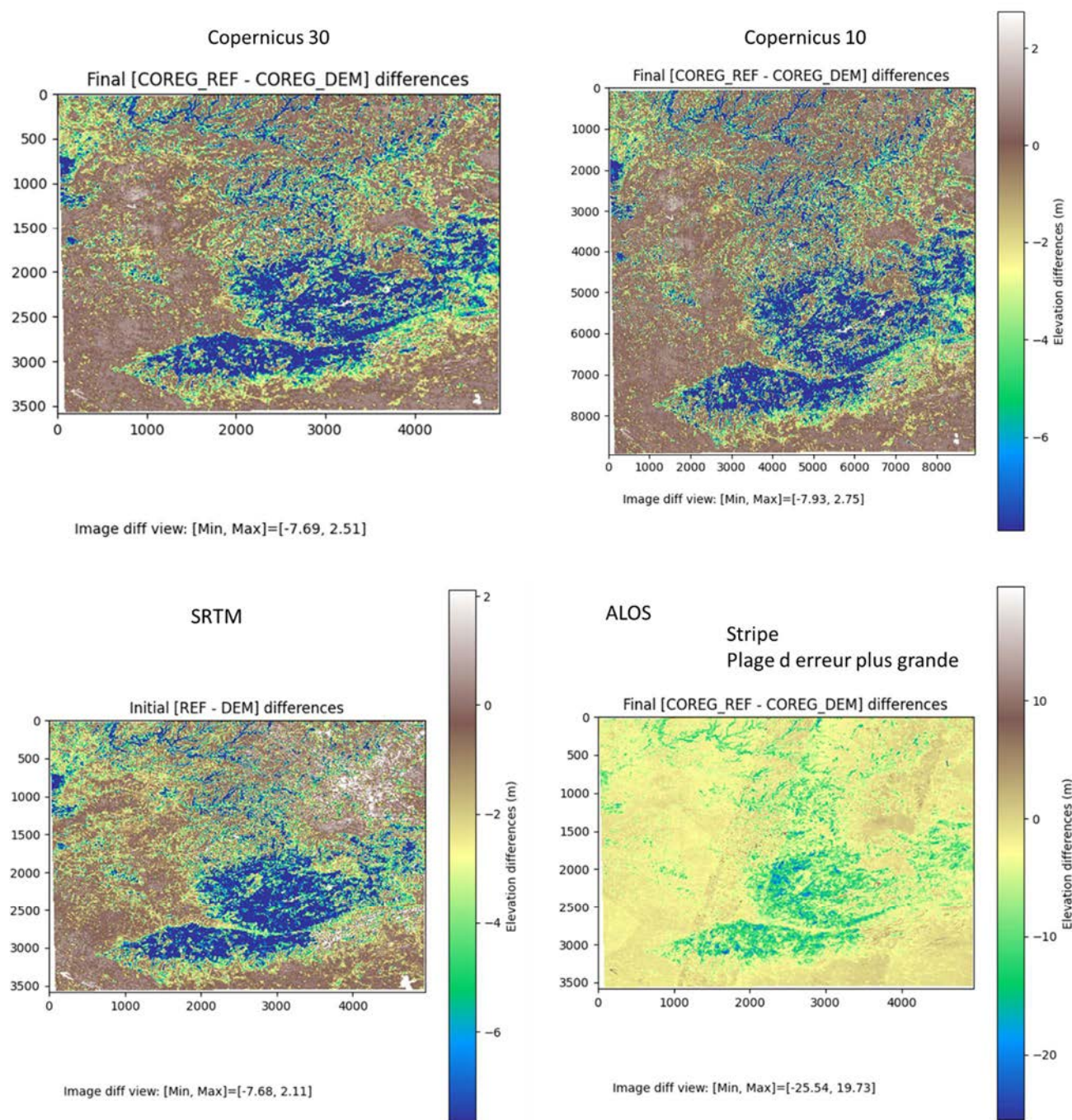


Figure 29: Différence d'élévation entre les DEM Copernicus 30 et 10, ALOS DEM, SRTM et le MNT IGN 5m de référence

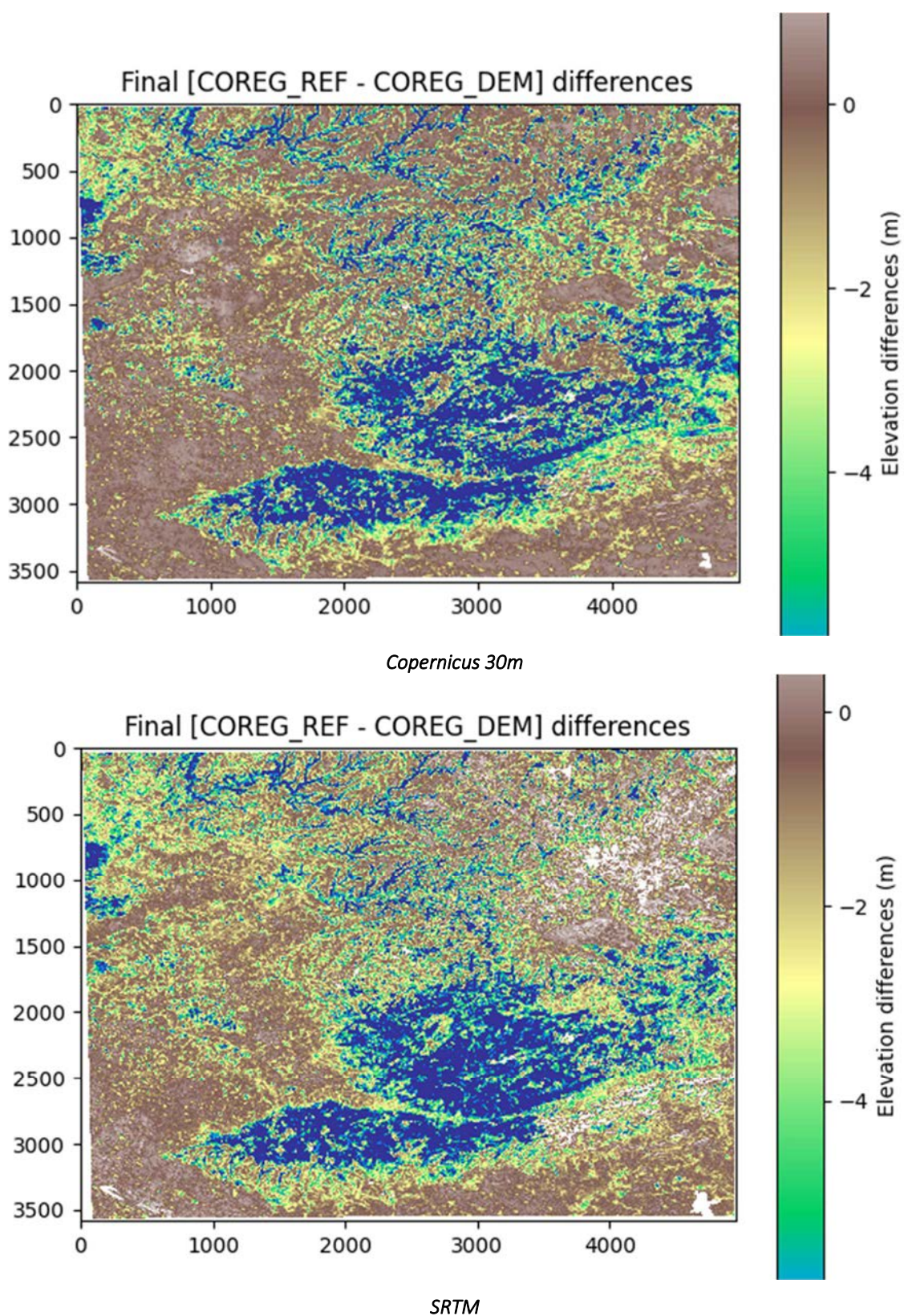


Figure 30: Différence entre DEM SAR 30m, SRTM , Copernicus et le MNT de référence IGN

Le second paramètre d'évaluation retenu de la qualité des MNT est celui de la répartition des pentes, en prenant toujours comme référence le MNT IGN 5m (cf. Tableau 8).

Comme on pouvait s'y attendre au vu des résultats des erreurs d'altitude, c'est le DEM ALOS qui présente les plus fortes erreurs Max et Min, et cela pour toutes les classes de pentes. Ces erreurs sont de 4 à 5 fois supérieures à celles obtenues pour les DEM Copernicus et SRTM. Cela se retrouve dans l'erreur moyenne qui est jusqu'à 10 fois supérieur dans le cas du DEM ALOS pour les faibles pentes que pour les DEM Copernicus 30 et 10m. L'erreur moyenne SRTM comparée à celles obtenues pour les DEM Copernicus pour les deux classes de pentes les plus faibles, 0-5 et 5-10, sont également bien plus importantes, mais avec un rapport bien moindre.

Tableau 8 : Répartition des pentes

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	5,99457703	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	0,55375755	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	64,91111755	-70,58598328	-3,620969772	5,18136549	6,321233273	-1,971942663	2,56707281	8,338098812
[0; 5]	19,58338832	64,84627533	-69,96243286	-1,647333264	3,830607653	3,275063992	-1,22998631	1,434907854	3,224345589
[5; 10]	19,28440745	64,91111755	-70,58598328	-2,204917669	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

Il est possible à partir de courbes des gaussiennes approximées des histogrammes de la répartition des erreurs d'altitudes en fonction des classes de pentes, d'analyser également les erreurs d'altitudes. Ces graphiques mettent là encore en évidence, des écarts ou des similarités en termes de qualité entre les MNTs analysés. Ainsi, il ressort que pour les DEM Copernicus les différences observées entre les deux DEM 30 et 10 m sont minimes/ voire infimes, elles portent sur un meilleur respect des fortes pentes (i.e. 25-40%).

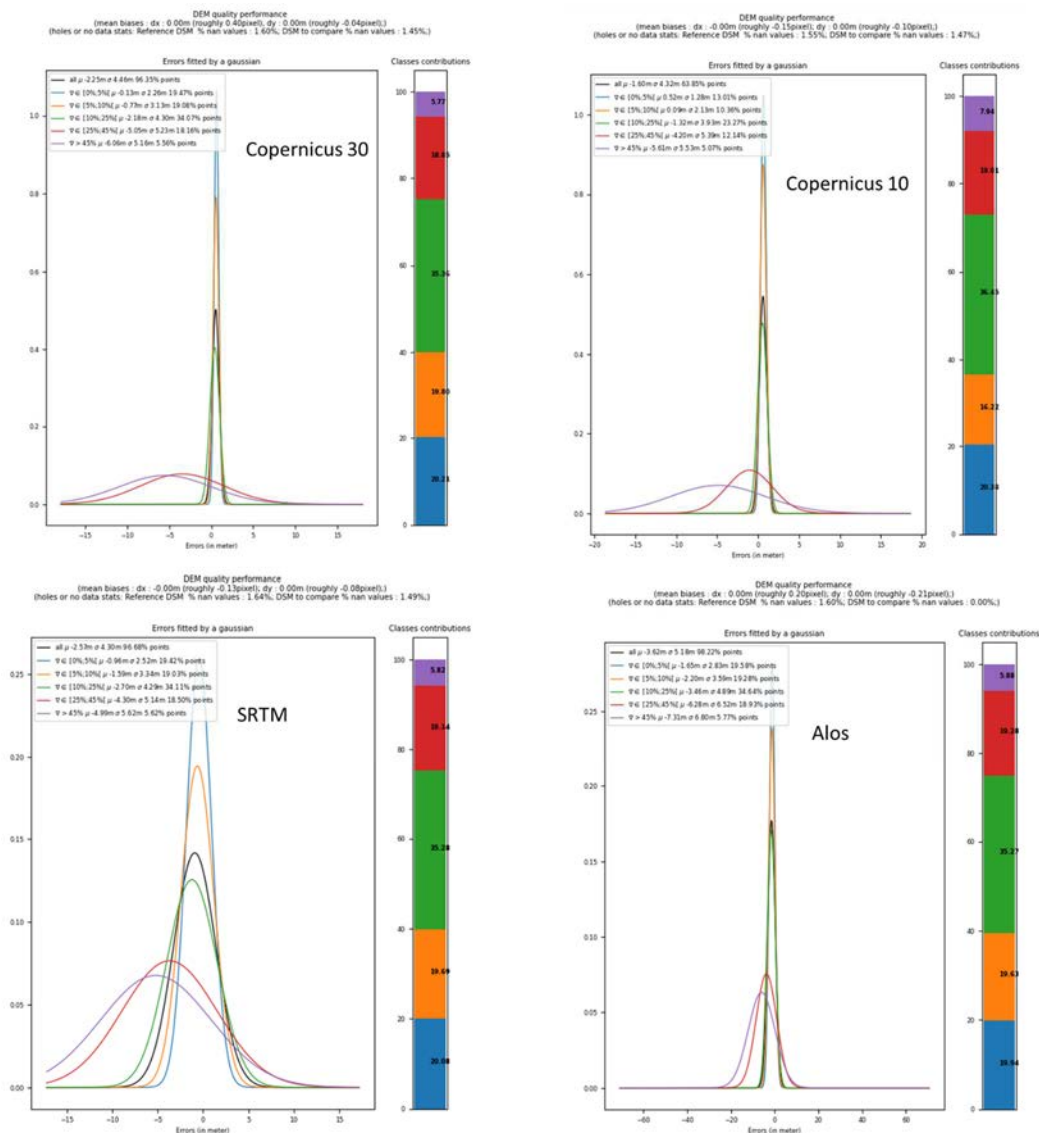


Figure 31: Courbes des gaussiennes approximées des histogrammes de la répartition des erreurs d'altitudes en fonction des classes de pentes

En conclusion de cette analyse de qualité des MNTs et de leurs aptitudes à rendre compte des pentes, on constate que les données Copernicus 10 et 30 sont assez semblables en termes de performances, bien que Copernicus 30 dérive quelque peu sur les plus fortes valeurs de pente. Globalement Copernicus 30 et SRTM sont assez similaires, alors qu'ALOS présente la plus forte variabilité.

Copernicus 10 n'ayant été pris en compte qu'afin d'illustration/compréhension d'un MNT de plus haute résolution, l'analyse des MNTs globaux par rapport à une référence, l'IGN 5m, montrent que les MNT Sar tiennent mieux compte des pentes que le DEM ALOS, et présentent des différences d'altitude d'ordre pluri-métrique par rapport à la référence, alors que le DEM ALOS présente lui des différences décimétriques voir pluri-décimétriques.

Ces résultats sont en accord avec ceux obtenus lors des travaux réalisés dans le cadre du support SERTIT aux activités SWOT (Yésou et al., 2018), et aussi avec ce qui est décrit dans la littérature (par exemple in Schumann and Bates, 2020)

Pour ces raisons, nous proposons pour la suite des travaux, sur la base de cette analyse, de qualifier uniquement les MNTs Globaux SRTM et Copernicus 30, ayant un léger avantage qualitatif, et de rejeter le MNT optique ALOS.

3.3 Analyse et validation

Pour les Phases I et II, la validation sur les sites retenus a été effectuée :

- ✓ Par analyse automatique comparative en fonction des données *in situ*,
- ✓ Par évaluation de la pertinence des différents produits SURFWATER en termes de représentativité de la dynamique des réservoirs. Pour cela, une extraction indépendante des surfaces en eau a été réalisée sur un nombre réduit de réservoirs, en exploitant la chaîne **EXtractEO** du SERTIT, et la dynamique telle que décrite par les produits SURFWATER a été comparée à celle issue de cette chaîne,
- ✓ Par comparaison des performances en fonction des MNTs utilisés pour la modélisation H/S/V,
- ✓ Par analyse d'expertise comparative ou croisée de certains résultats obtenus sur différents sites ou groupes de sites par rapport aux références attendues ou aux améliorations effectuées en cours de phase I.

3.3.1 Evaluation qualitative du Water extent

3.3.1.1 Méthodologie

Afin d'évaluer la qualité des *water extent* générés par Surfwater, plusieurs comparaisons ont été effectuées sur deux tuiles (30SUH et 30STF). Lors de cette étape de validation des *water extent*, nous avons uniquement analysé la qualité des résultats générés par Surfwater à partir de données optiques Sentinel-2 (Instant-mask et Water-mask).

Pour cela nous avons comparées les surfaces extraites sur une période de 3 ans, via notre suite logicielle ExtractEO, avec celles issues de Surfwater.

Cette activité sera renouvelée si nécessaire sur les jeux de données du périmètre des activités complémentaires.

ExtractEO est un logiciel développé par ICube-SERTIT, mettant en œuvre des chaînes automatisées de bout en bout sur des données satellitaires. ExtractEO est capable de traiter différents satellites, Sentinel-2, Landsat, Sentinel-1... et dispose de plusieurs chaînes implémentées, de l'extraction de l'eau et du feu à la détection des nuages. L'objectif d'ExtractEO est de travailler à la fois dans un contexte de séries temporelles et de larges zones d'études (larges fenêtres spatiales et temporelles). Les surfaces en eau sont détectées à partir d'un algorithme perceptron multicouche (réseau de neurones) et intégrant la base de données Global Surface Water pour l'échantillonnage.

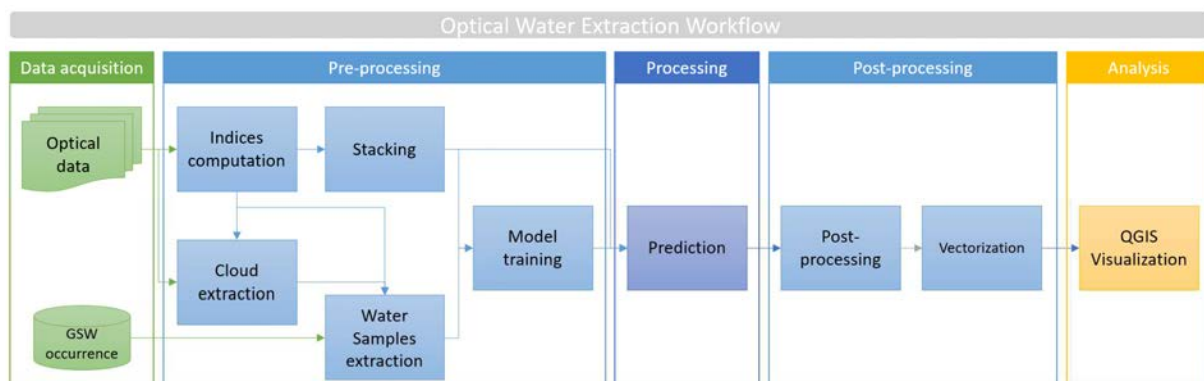


Figure 32 : Architecture de la suite logicielle ExtractEO du SERTIT

Pour la tuile 30SUH, nous avons comparé les *water extent* générés à partir de l'outil ExtractEO du SERTIT avec les *water extent* générés par Surfwater (Instant-mask et Water-mask).

Les résultats issus de Surfwater avec deux paramétrages différents ont été comparés avec les résultats ExtractEO:

- ✓ Paramétrage 1 : SW Optique basé sur algorithme Water Detect,
- ✓ Paramétrage 2 : SW Optique basé sur RandomForest + référence GSW.

Pour la tuile 30SUH, nous avons comparé les *water extent* générés à partir de l'outil ExtractEO du SERTIT avec les *water extent* générés par Surfwater (Instant-mask et Water-mask), mais également avec les *water extent in situ* générées par l'outil « timeseries_tool.py » du CNES, à partir des volumes et hauteurs d'eau in situ. Uniquement les résultats issus de SURFWATER avec configuration de SW en optique (mode random Forest), ont été analysés pour cette étape.

Pour ces différentes comparaisons, nous avons comparé les *water extent* générés aux dates communes à ExtractEO et SURFWATER tout en supprimant les *water extent* pour lesquels la présence de nuages avait été identifiée par un des deux outils (Nodata MAJA pour SURFWATER et Cloud mask pour ExtractEO).

3.3.1.2 Cas de la tuile 30SUH

Cette tuile 30SUH comprend 4 réservoirs dont la superficie varie de 581 ha à 31811 ha.

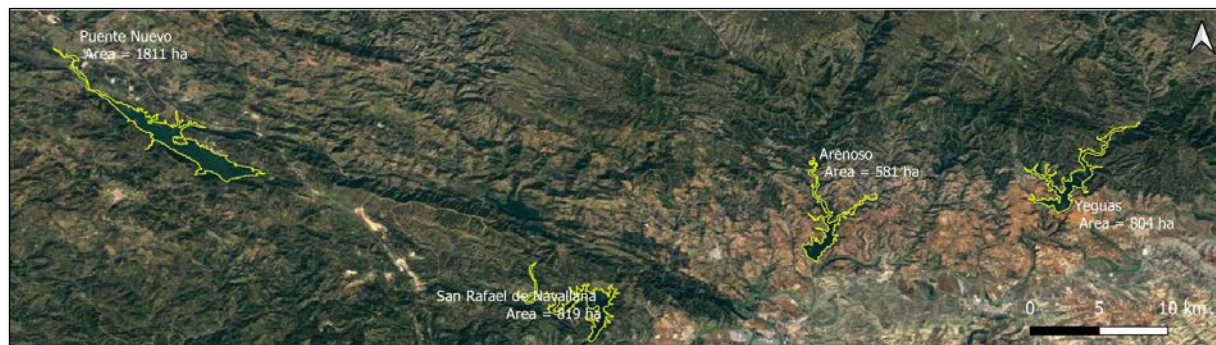


Figure 33 : Location map tuile Sentinel-2 (30SUH)

Paramétrage 1 :

- ✓ Globalement, on observe des tendances similaires entre les surfaces SURFWATER et ExtractEO, avec toutefois de nombreuses incohérences, artefacts, sur les séries SURFWATER, marquées par des chutes des surfaces incompréhensibles (problèmes de paramétrages?). Le passage de l'Instant mask au Water mask (10 jours) tend à gommer partiellement ces artefacts. De meilleurs résultats sont obtenus avec le Water Mask de SURFWATER, qui semblent plus robustes,

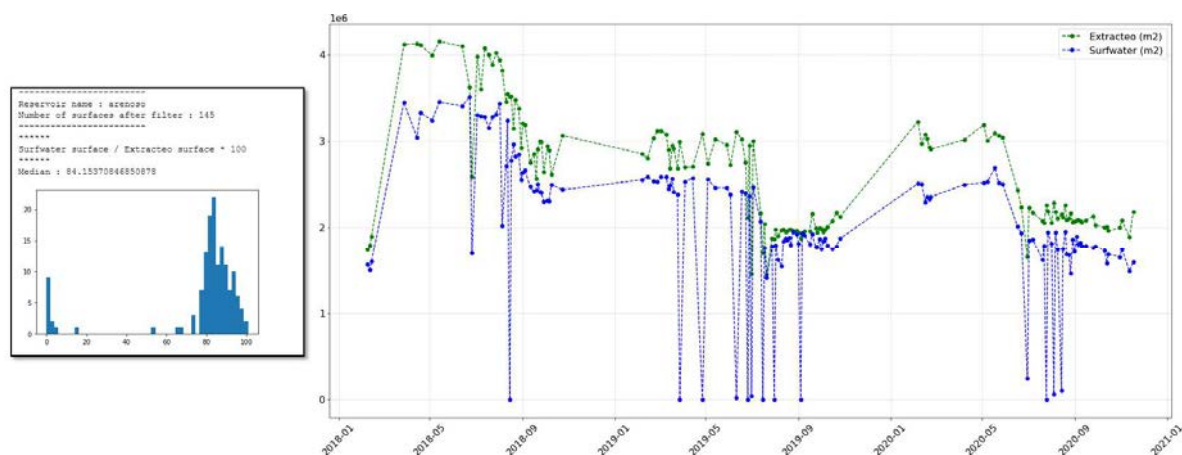


Figure 34 : Arenoso - Instant Mask

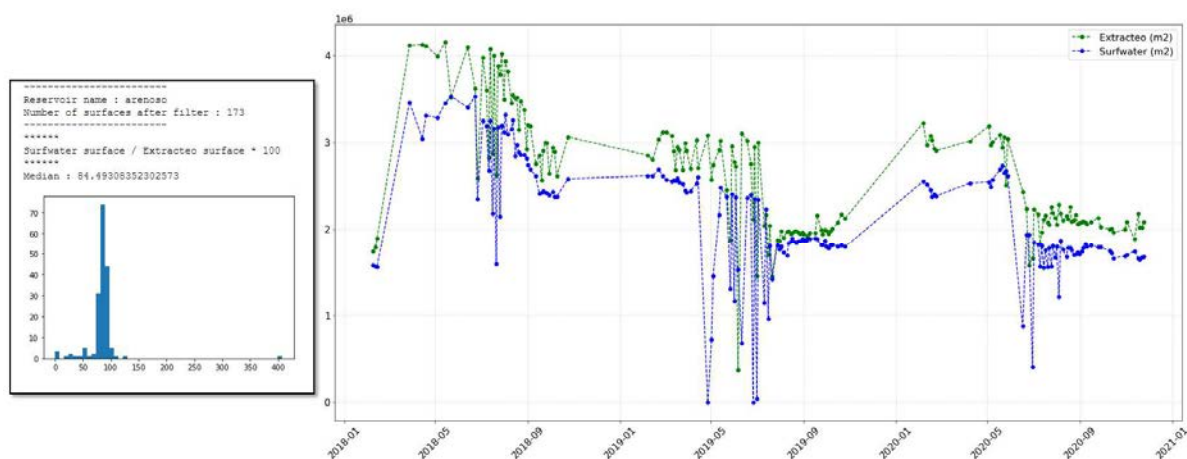


Figure 35 : Arenoso - Water Mask

- ✓ Des incohérences importantes, malgré la suppression des dates avec du nodata (masque nuage MAJA) (cf. Figure 36).

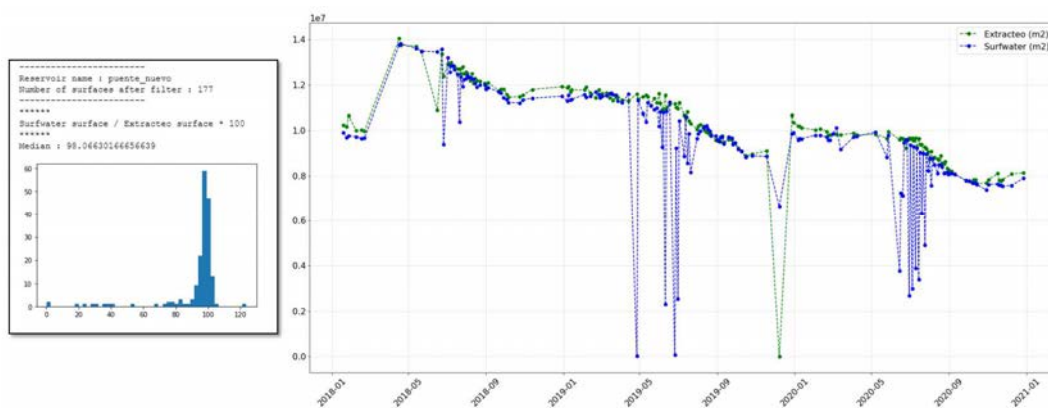


Figure 36 : Puente Nuevo - Water Mask

Paramétrage 2 :

On obtient de meilleurs résultats avec ce paramétrage. On retrouve une bonne cohérence entre les données SURWATER Water Mask et ExtractEO, malgré un gap entre les valeurs, qui s'explique par une tendance à la commission par ExtractEO au niveau des ombres des reliefs et des problèmes de classification de la part de SURFWATER qui crée de l'omission.

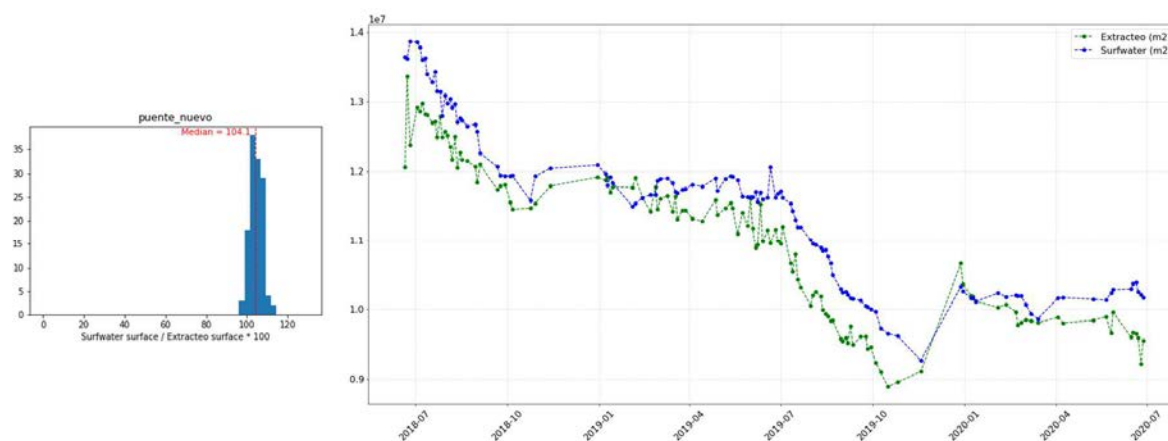


Figure 37 : Puento Nieve - Water Mask

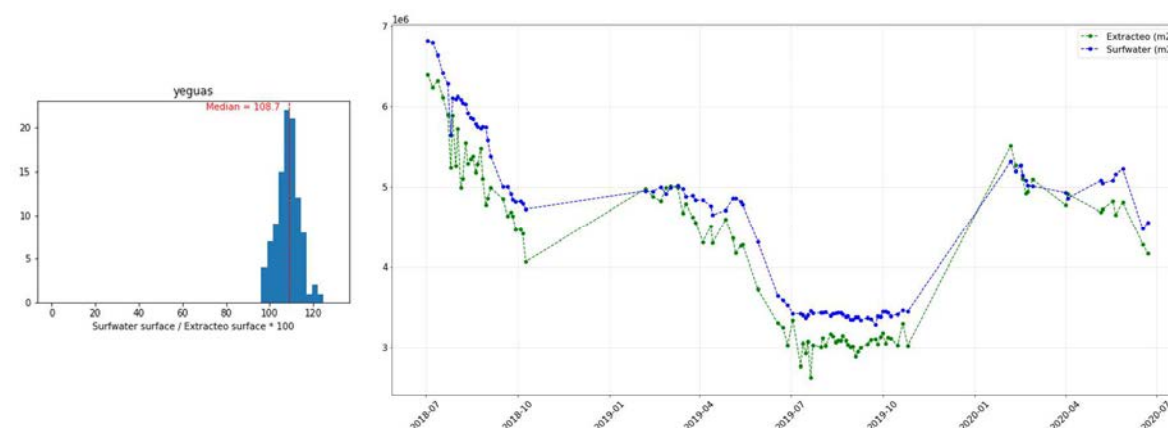


Figure 38 : Yeguas - Water Mask

3.3.1.3 Cas de la tuile 30STF

Cette tuile comprend 9 réservoirs pour lesquels il a été possible de générer des water extent *in situ* à partir des données volume/hauteur fournies par les prestataires. La taille des réservoirs varie de 63 ha à 3193 ha.

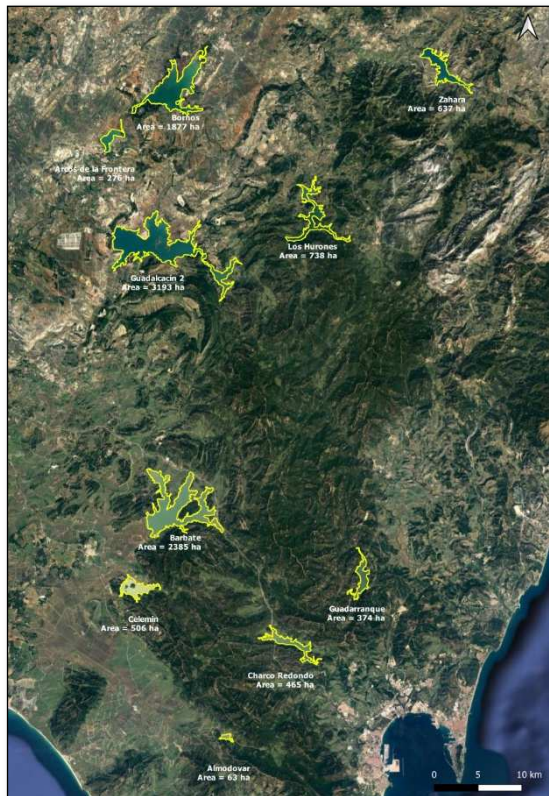


Figure 39 : Location map tuile Sentinel-2 (30 STF)

L'analyse a permis d'obtenir les observations suivantes :

- ✓ Globalement, de meilleurs résultats sont obtenus avec le *Water Mask* de SURFWATER, qu'avec l'*Instant Mask*. Malheureusement, il s'avère qu'il détériore parfois la qualité des résultats (cf. Figures 40 et 41),

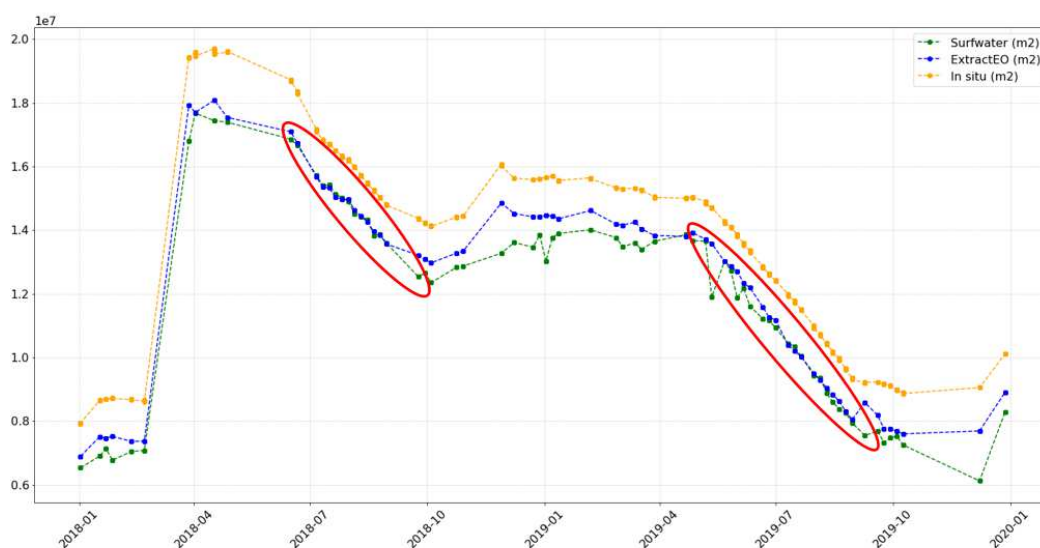


Figure 40 : Borlos - Instant Mask

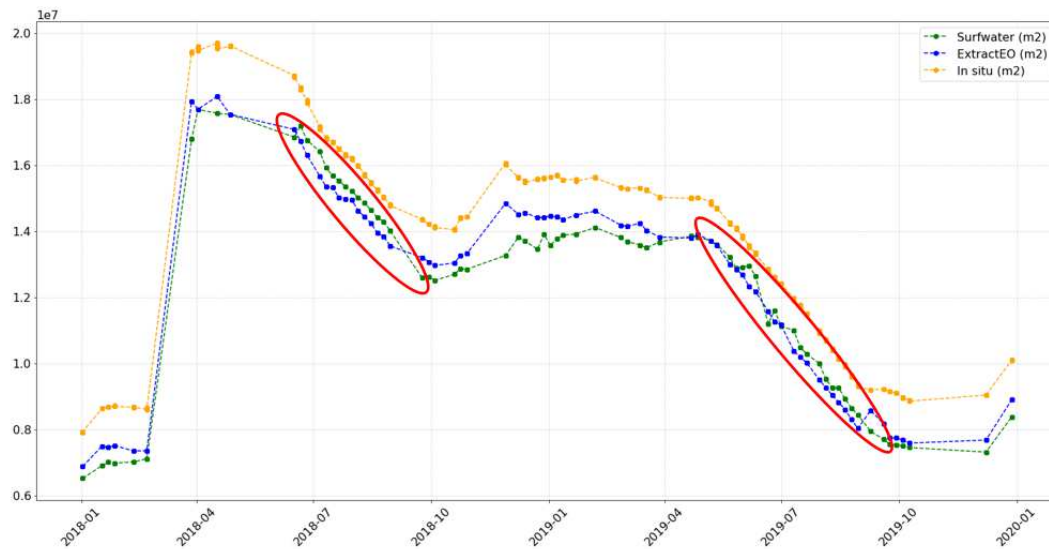


Figure 41 : Bornos - Water Mask

- ✓ ExtractEO est souvent plus proche des données *in situ* (cf. Figures 42, 43, 44). On identifie des problèmes de classification en été pour SURWATER, aux alentours du mois de juillet, probablement dû au sunglint.(cf. Figure 42).

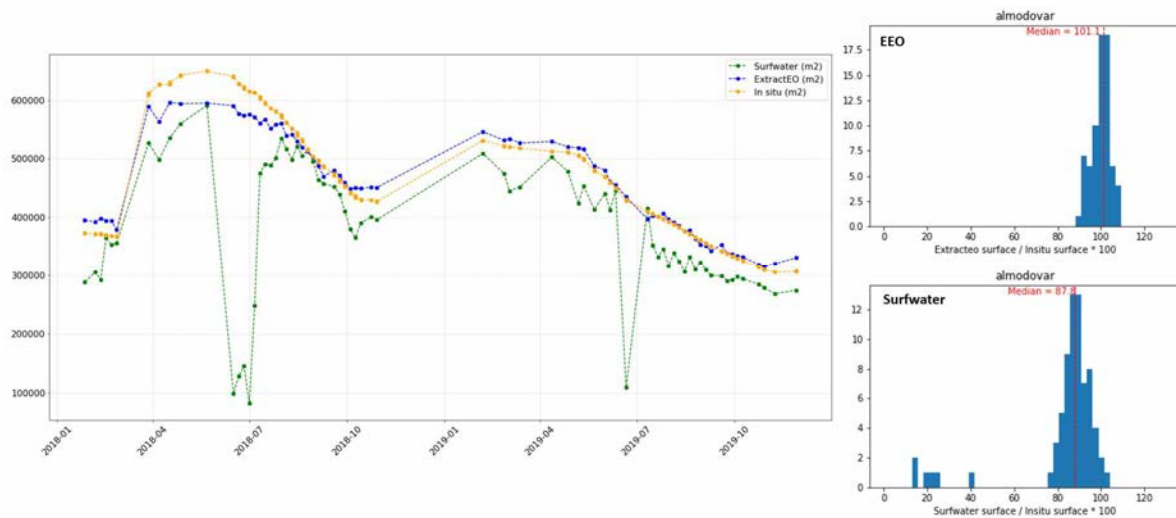


Figure 42 : Almodovar - Water Mask

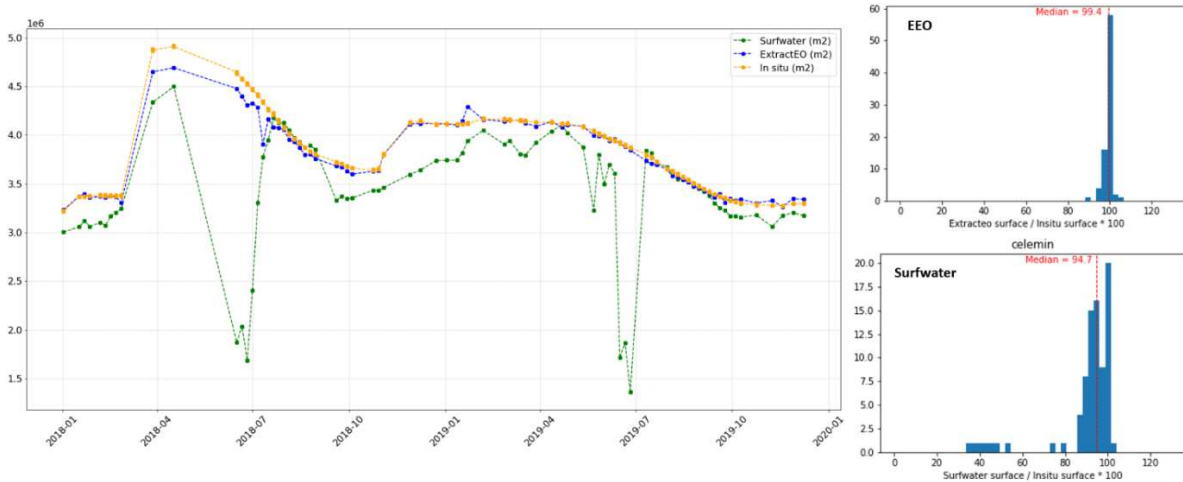


Figure 43 : Cemelin - Water Mask

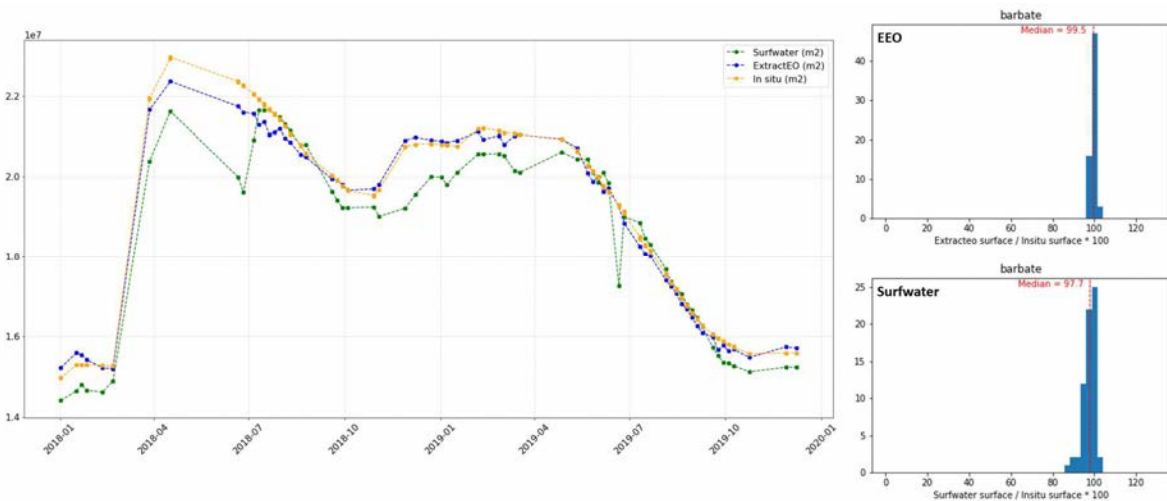


Figure 44 : Barbate - Water mask

Dans le but d'analyser les artefacts observés dans les séries SURFWATER, une analyse comparative des masques issus de SURFWATER et d'ExtractEO a été effectuée sur les données du réservoir de Barbate qui présentait une chute anormale en juin 2018 (cf. Figure 44). L'analyse des données Sentinel-2 montre que la surface d'eau présente un ensemble de textures/teintes, vraisemblablement causées par l'effet du vent et l'illumination solaire (sunglint). Surfwater est plus sensible à ces variations de conditions de surfaces, et une partie de la surface en eau n'est pas reconnue comme telle, d'où le pic négatif observé (cf. Figure 42).

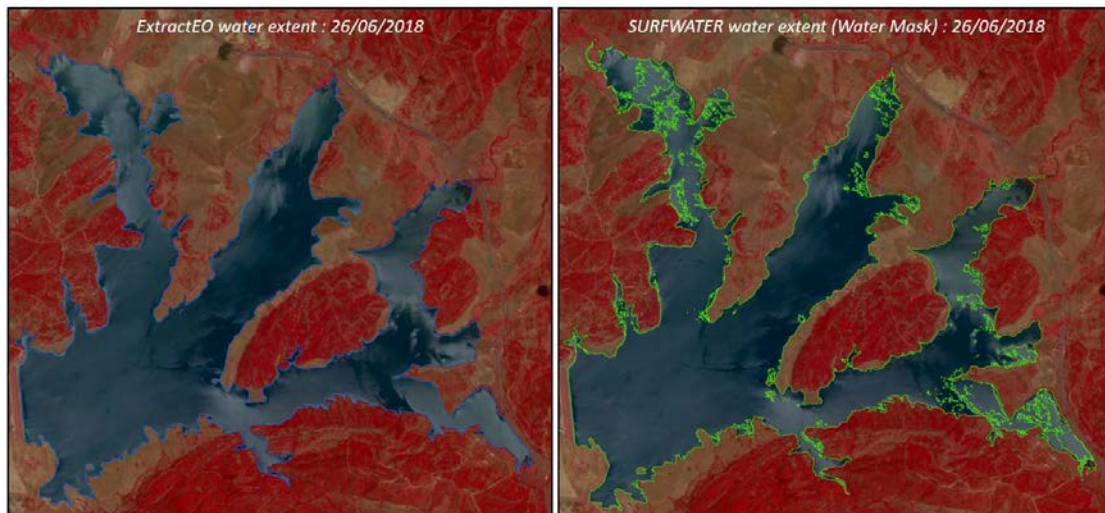


Figure 45 : Water Extent EEO & SURFWATER - Barbare - 26/09/2018

Les deux satellites Sentinel-2 (A et B) acquièrent des images au-dessus de la pointe Sud de l'Espagne, de façon plus ou moins complète. La trace Sentinel-2B couvre l'intégralité de la pointe, alors que Sentinel-2A, plus décalée à l'Est ne couvre que partiellement cette zone. On suppose que c'est la raison pour laquelle des données acquises par Sentinel-2B ont été sélectionnées pour cette étude.

Toutefois, il est intéressant de noter que les effets semblent plus importants lorsque les cibles sont localisées au centre de l'image. De ce fait, les données Sentinel-2B sont plus sujets aux effets de sunglint que les données acquises par Sentinel-2A. Ces effets sont illustrés sur le réservoir d'Almodovar, mais une démonstration similaire aurait pu être faite sur d'autres réservoirs par exemple celui de Charco Retondo.



Figure 46 : Trace Sentinel-2B à gauche et Sentinel-2A à droite

Il est également intéressant de comparer les données Sentinel-2B entre elles. En effet, les dates pour lesquelles les données Surfwater ont éprouvées le plus de difficultés à extraire les surfaces en eau, sont 2018-06-16, 2018-06-21, 2018-06-26, 2018-07-01. Ces images acquises durant cette période présentent des surfaces en eau les plus brillantes. Les surfaces en eau reprenant une allure « normale » à la fin de la série présentée (fin juillet 2018).

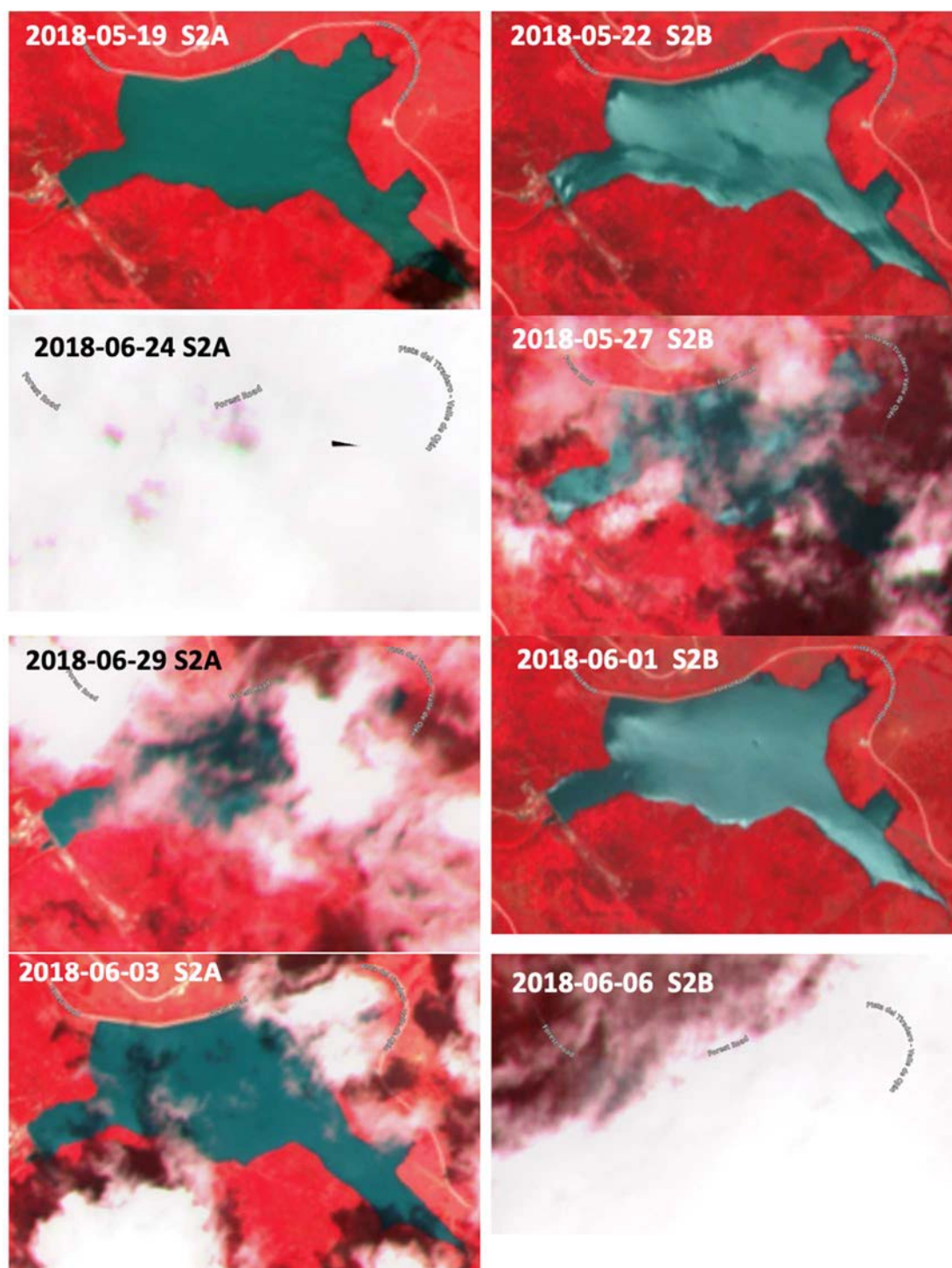


Figure 47 : Réflectance plus ou moins impactée par les effets de sunglint selon la trace position dans la trace. A gauche Sentinel-2A, et à droite Sentinel-2B. Exemple du réservoir de Almodovar. Composition colorée S2, B8, B4, B3 en RVB, images acquises entre le 19/05/2018 et le 29/06/2018.

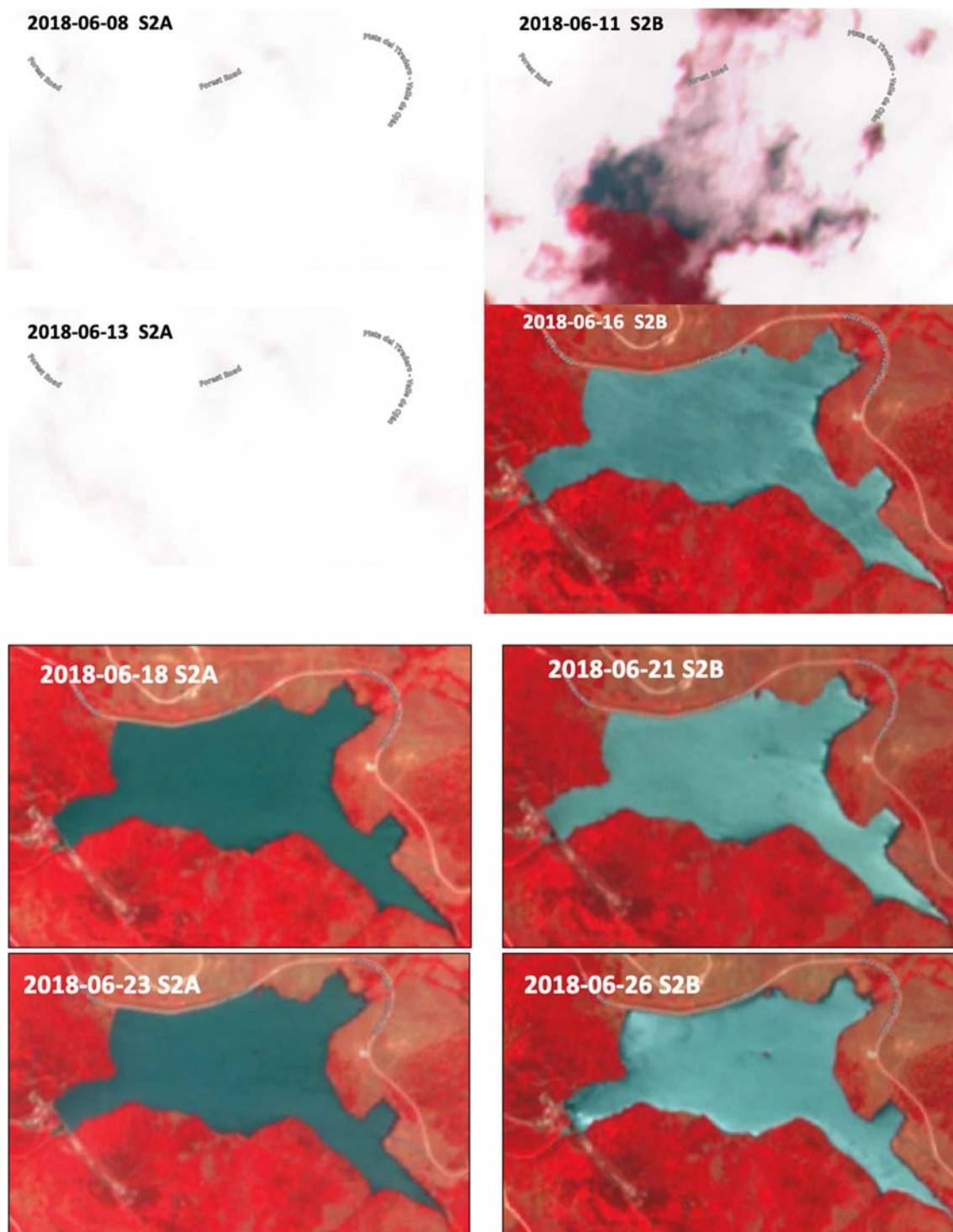


Figure 48 : Réflectance plus ou moins impactée par les effets de sunglint selon la trace position dans la trace. A gauche Sentinel-2A, et à droite Sentinel-2B. Exemple du réservoir de Almodovar. Composition colorée S2, B8, B4, B3 en RVB.

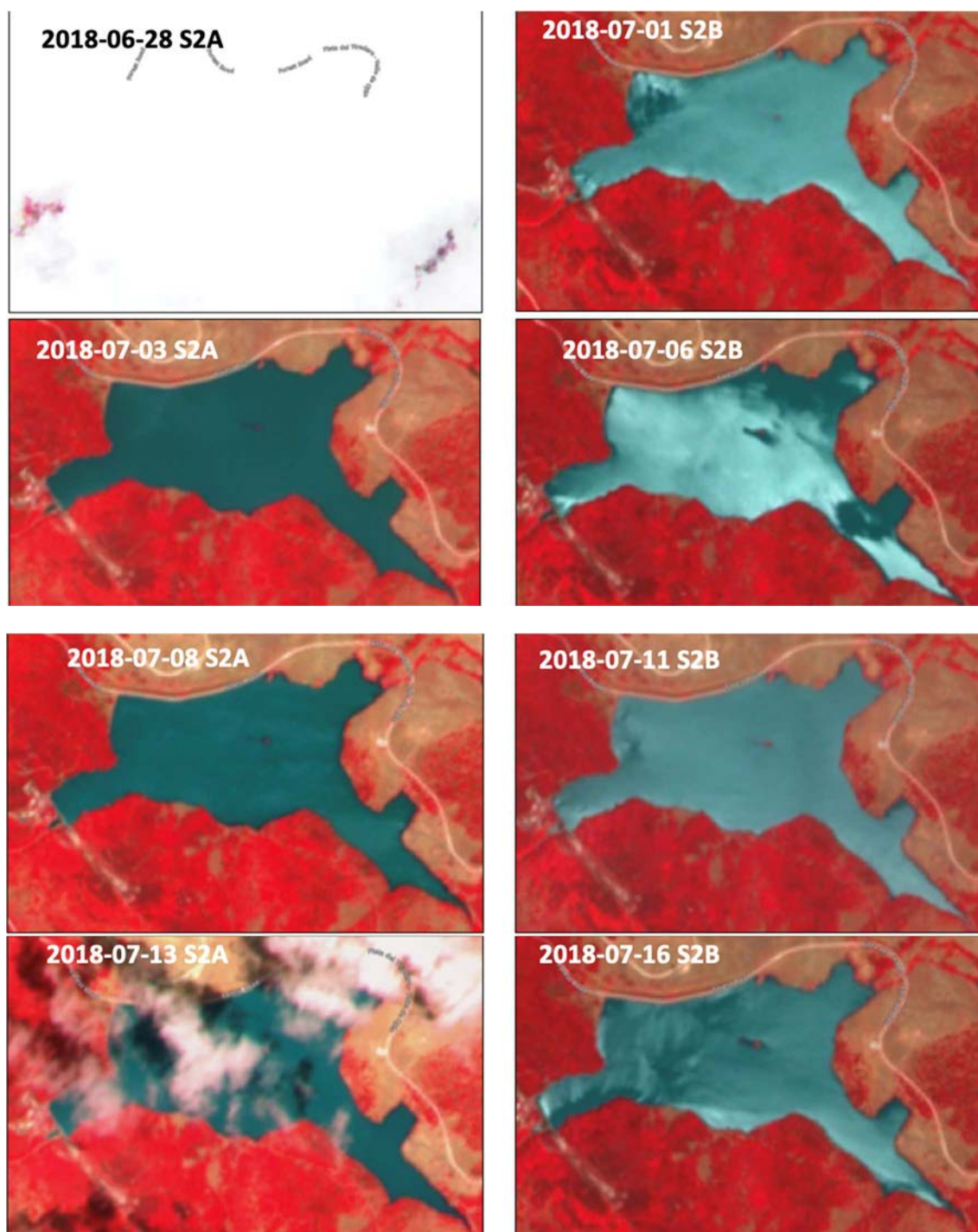


Figure 49 : Réflectance plus ou moins impactée par les effets de sunglint selon la trace position dans la trace. A gauche Sentinel-2A, et à droite Sentinel-2B. Exemple du réservoir de Almodovar. Composition colorée S2, B8, B4, B3 en RVB, images acquises entre le 28/06/2018 et le 16/07/2018.

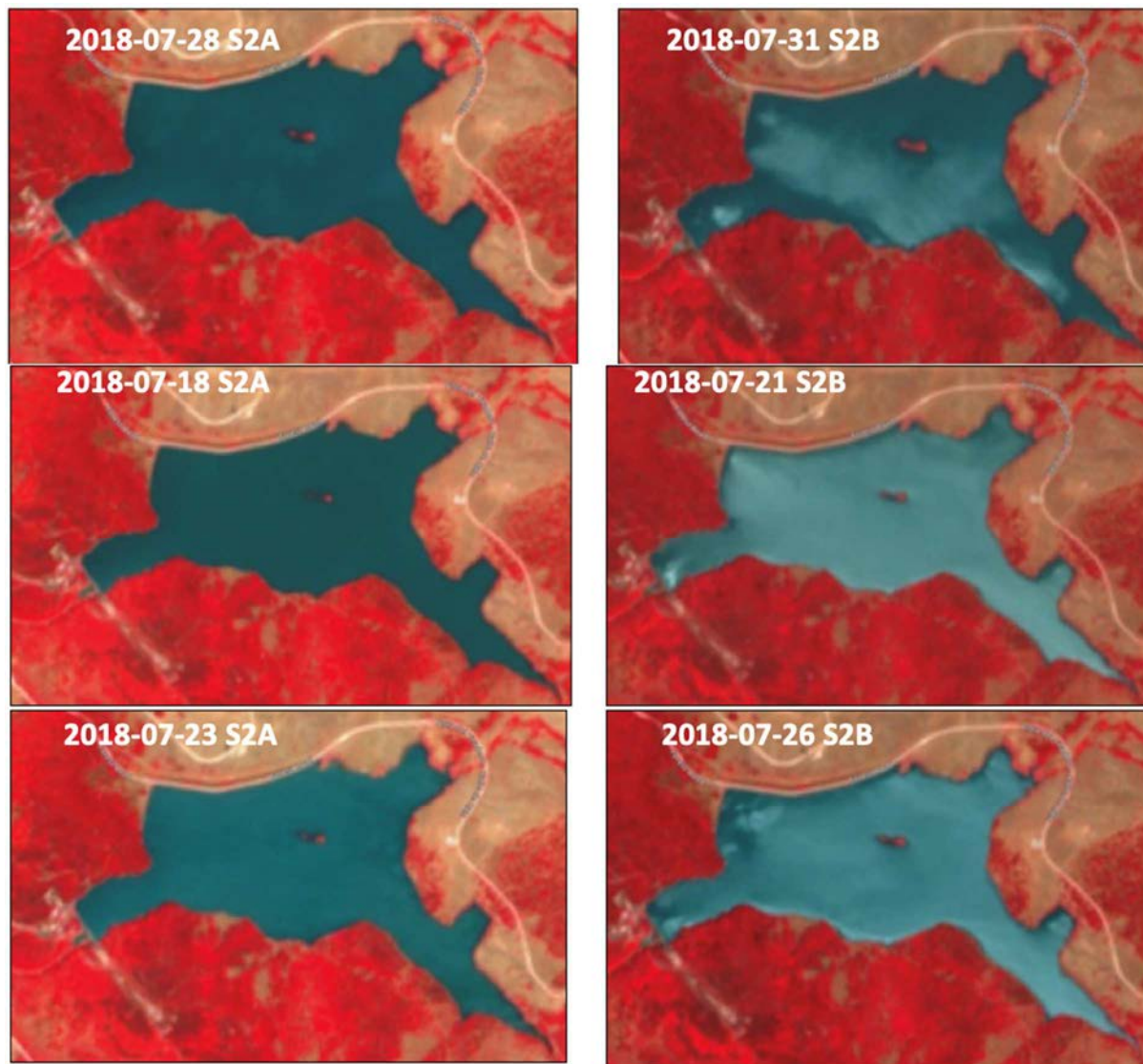


Figure 50 : Réflectance plus ou moins impactée par les effets de sunglint selon la trace position dans la trace. A gauche Sentinel 2A, et à droite Sentinel 2B. Exemple du réservoir de Almodovar. Composition colorée S2, B8, B4, B3 en RVB, images acquises entre le 18/07/2018 et le 28/07/2018.

Sur les réservoirs de Los Hurones, Bornos, et Char Retondo, on observe un gap (shift régulier) entre les données ExtractEO/Surfwater et les données surface dites « in situ », avec un comportement global similaire des séries (cf. Figures 51, 52, 53). Ce shift aurait pu être expliqué par une omission systématique d'une partie des plans d'eau. Une confrontation entre les images, les extractions SURFWATER et ExtractEO a été réalisée pour le réservoir de Bornos et à trois dates, correspondant à trois niveaux de remplissage (plein, intermédiaire, et bas niveau d'eau) (cf. Figures 54, 55, 56). Les surfaces en eaux extraites aux trois dates sont très proches/quasi similaires entre les approches SURFWATER et ExtractEO, et semblent être représentatives de la surface en eau observée. Il n'y a pas de zone sous ou sur éclairée, qui aurait pu être omise, on n'observe pas de végétation très importante sur les bords du lacs qui aurait entraîné des zones d'ombres, causant ainsi des omissions fortes.

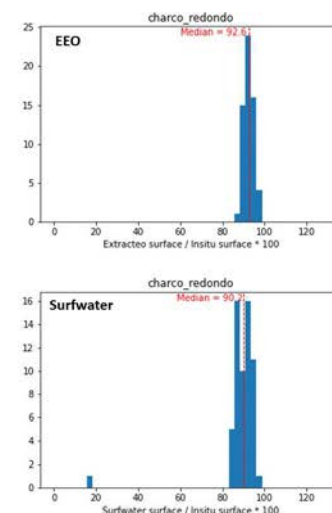


Figure 51 : Water mask - Charco Redond

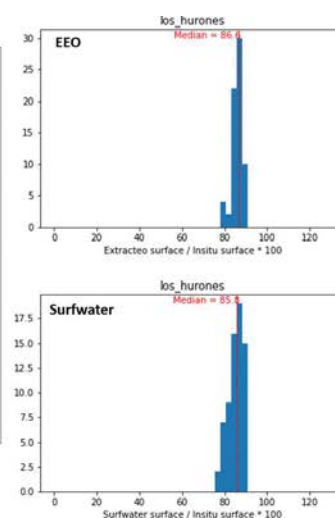
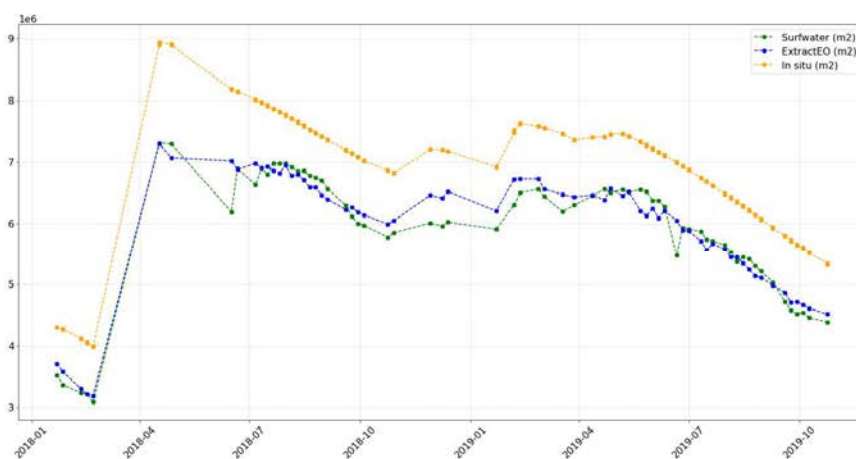


Figure 52 : Water Mask - Los Hurones

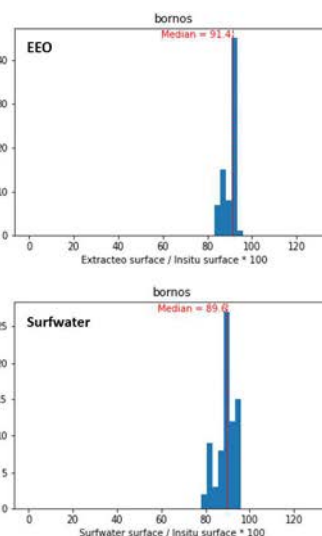
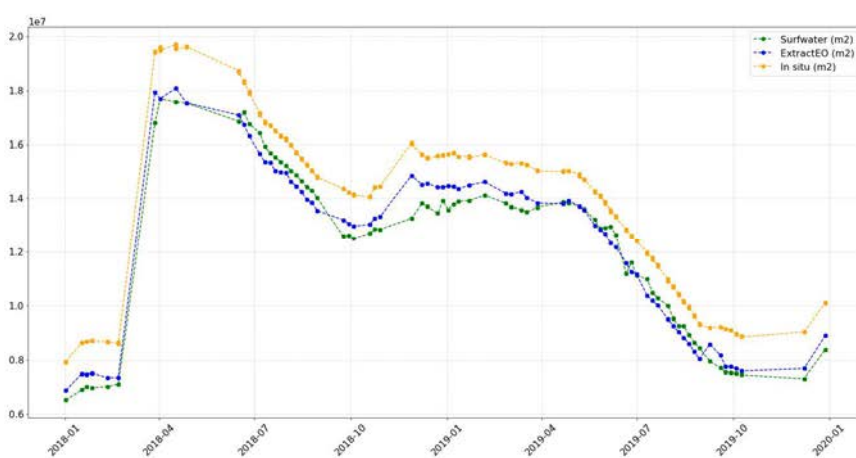


Figure 53 : Water Mask – Bornos

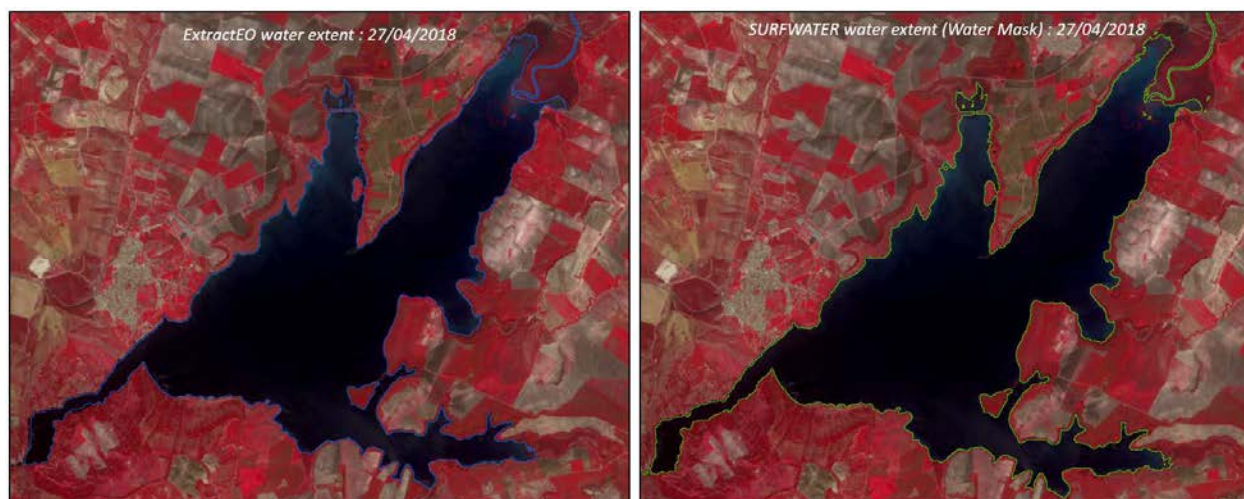


Figure 54 : Water Extent EEO & STOCKWATER - Bornos - 27/04/2018

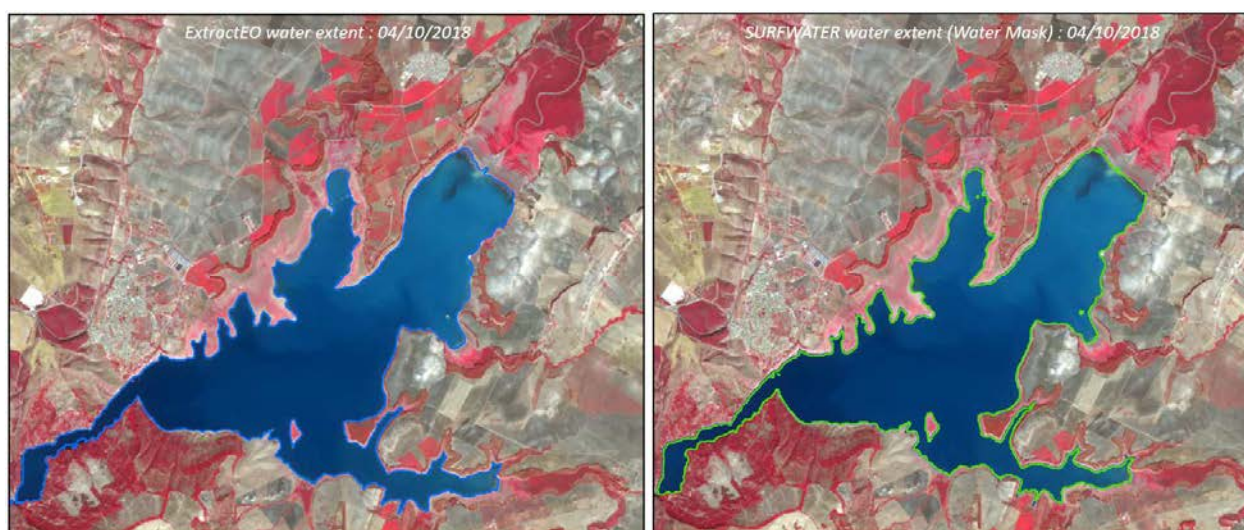


Figure 55 : WATER EXTENT EEO & STOCKWATER - Bornos - 04/10/2018

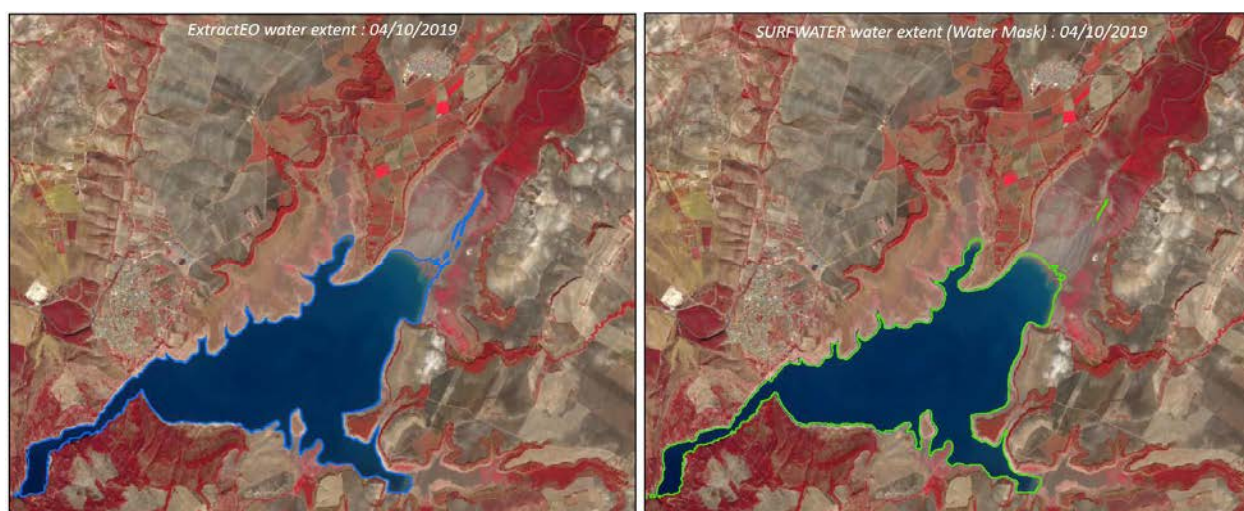


Figure 56 : WATER EXTENT EEO & STOCKWATER - Bornos - 04/10/2019

Aussi, si l'on considère les extractions SURFWATER et ExtractEO comme valides et optimales en termes de représentation de la surface en eau, cela amène à remettre en cause la qualité des données dites de référence.

Le shift constant observé entre les données Surfwater et ExtractEO a été calculé pour les réservoirs de Charco Redondo, de Los Hurones et de Bornos avec des valeurs respectivement de 7, 12 et 8%. Il est intéressant de noter que si l'on soustrait ce biais aux valeurs dites de référence, on obtient une quasi parfaite concordance entre les surfaces issues des données satellites et des valeurs *in situ* « corrigées » (cf. Figures 57, 58, 59).

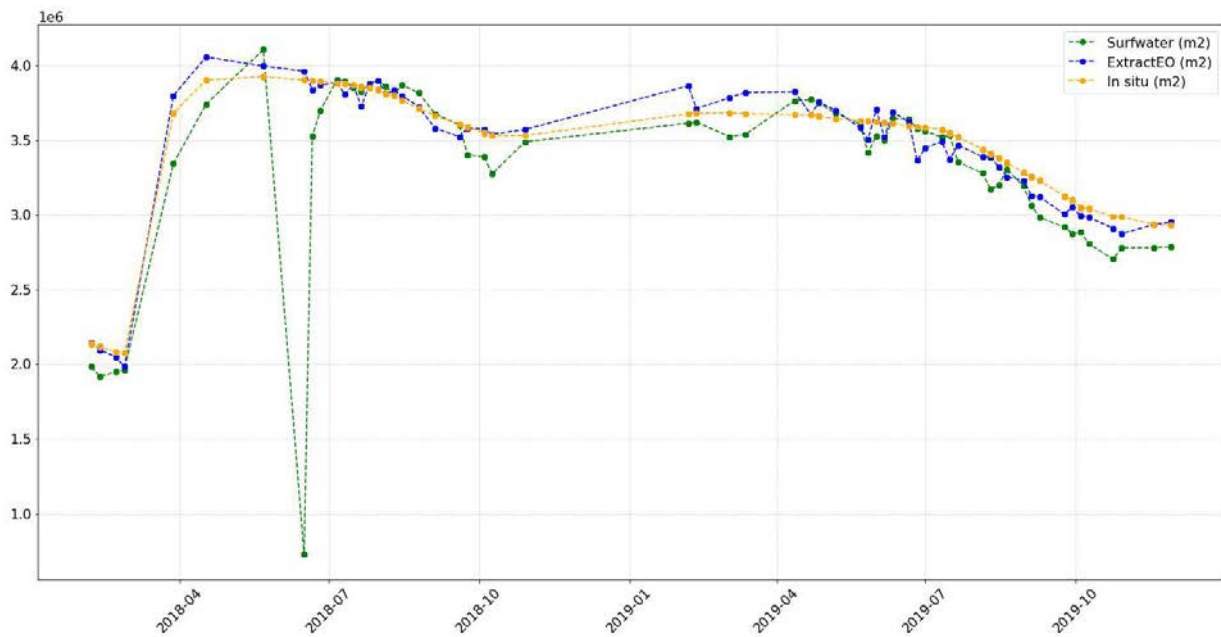


Figure 57 : Comparaison après soustraction de 7% à la surface de référence – Charco Redondo

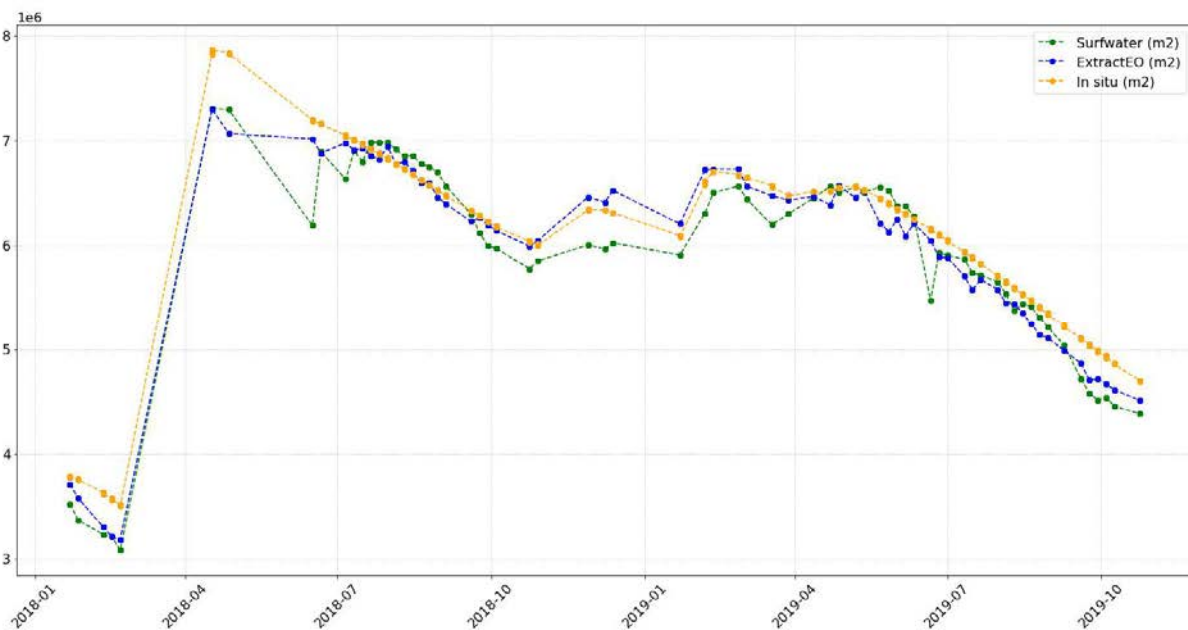


Figure 58 : Comparaison après soustraction de 12% à la surface de référence – Los Hurones

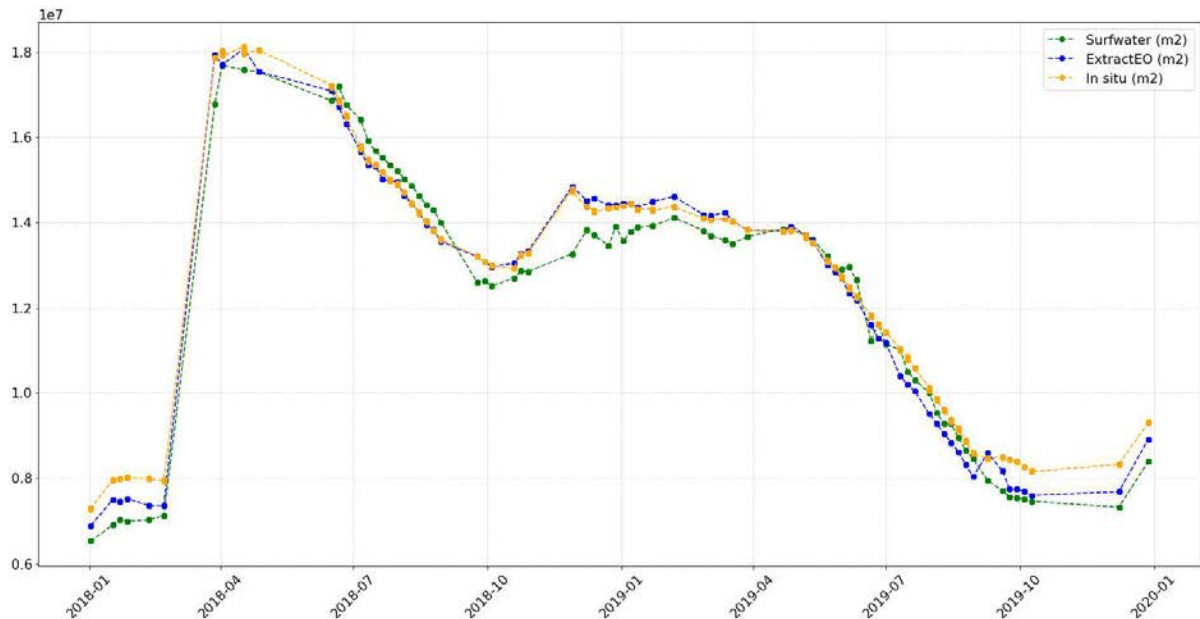


Figure 59: Comparaison après soustraction de 8% à la surface de référence - Bornos

Cela conforte l'idée qu'il faille émettre un doute quant à la qualité des données in situ. En effet, ces données de surfaces dites de référence ou *in situ* sont en fait issues d'un modèle H/S/V. Au moins deux possibilités pourraient donc engendrer une sur estimation des surfaces « *in situ* » :

- ✓ La première, serait de remettre en cause la qualité du modèle/de la méthode de calcul permettant de générer les surfaces à partir des Volume et Level *in situ*,
- ✓ La seconde amène à réfléchir sur la qualité des données fournies par les gestionnaires de barrage. En effet, contrairement aux hauteurs, qui sont des données physiques mesurées, les volumes sont des quantités issues de modélisation. Si la modélisation du gestionnaire de site est erronée, cela pourrait expliquer ces surfaces de référence « incohérentes »,
- ✓ Une façon de palier à ce problème pourrait être de travailler non pas en termes de volume absolu mais de variations de volume.

3.3.1.4 Comparaison Surface de référence, Surface ExtractEO et Surface SURFWATER

Sur deux tuiles, 30SUH et 30STF comprenant respectivement 4 et 9 et réservoirs de tailles allant de quelques centaines d'hectares à plusieurs milliers, une comparaison des water extent générés par SURFWATER et l'outil Sertit a été menée. De plus, sur la tuile 30SUH deux paramétrages de la suite Surfwater ont été testées. On observe une nette amélioration en passant par le deuxième paramétrage.

De façon générale, les surfaces en eau générées par l'outil ExtractEO sont légèrement supérieures à celles obtenues par Surfwater et semblent être plus réalistes. De plus, on observe sur les séries SURFWATER un certain nombre de dysfonctionnements entraînant des variations/anomalies des surfaces en eau (avec des valeurs obtenues très faibles). Ces dysfonctionnements sont à mettre en relations avec des phénomènes plus ou moins marqués de sunglint.

Surfwater montre une forte sensibilité à ces phénomènes qui sont renforcés sur les produits L2C par rapport aux produits L1C. Cela pourrait en partie expliquer le fait que la suite ExtractEO du SERTIT exploitant des données L1C soit moins sensible à ces phénomènes qui perturbent de façon non négligeable les séries issues de Surfwater. L'exploitation du Water mask plutôt que de l'Instant mask semble partiellement pallier ces problèmes.

3.3.1.5 Conclusions

La comparaison avec les séries de surfaces dites *in situ* (modélisation issue de la relation hauteurs/surfaces /volume) amène quant à elle à plusieurs considérations :

- ✓ Les surfaces extraites des données Sentinel-2 suivent les mêmes tendances que celles observées à partir des données *in situ*, re produisant la dynamique de remplissage/vidange des réservoirs,
- ✓ Les surfaces ExtractEO sont dans tous les cas plus proches des valeurs dites *in situ*, que les valeurs SURFWATER,
- ✓ Dans près d'un tiers des cas, on observe un biais systématique entre les surfaces dérivées des données Sentinel-2 et des références. L'analyse des images et des extractions ne permet pas de penser qu'il y ait eu une omission importante lors de l'extraction des surfaces en eau par SURFWATER ou ExtractEO. Ce biais peut être dû à un fonctionnement moins performant de la modélisation des surfaces « *in situ* » pour certains réservoirs, ou d'une modélisation du volume *in situ* erronée de la part des gestionnaires de ces réservoirs. Il pourrait être intéressant pour pallier ce problème de travailler non pas en termes de volume absolu, mais plutôt de variations de volume.

3.3.2 Analyse des lois REF

Lors de la production des lois estimées HSV sur les sites Andalousie, Inde et Occitanie, des anomalies ont été rencontrées traduisant sur certains cas les limites méthodologiques de la chaine dem4water et également parfois celles des données d'entrée.

En effet la variété des topographies locales des divers sites étudiés, les formes singulières de certaines retenues d'eau et l'imprécision des données d'entrée ont mis à l'épreuve la chaine dem4water. La liste des anomalies rencontrées consiste en une mauvaise estimation de pieds de barrage, une ligne de coupure partiellement en boucle, du water crossing, des cas complexes de multi-déversoirs avec plusieurs ouvrages sur une même retenue d'eau.

Cette activité sera renouvelée si nécessaire sur les jeux de données du périmètre des activités complémentaires.

Pieds de barrage

À partir du profil topographique longitudinal à l'aval du barrage et de la détermination du point de rupture de pente, l'estimation automatique de hauteur de pied de barrage pour le réservoir Pla de Soulcem en Occitanie donne une estimation erronée de 29 m du pied de barrage avec une hauteur de 1528 m au lieu de 1499 m. Cette différence de Z0 impacte notablement S(Z) comme on peut le constater sur le graphique ci-dessous entre le modèle estimé S(Z) en vert et le modèle *in situ* de référence en rouge.

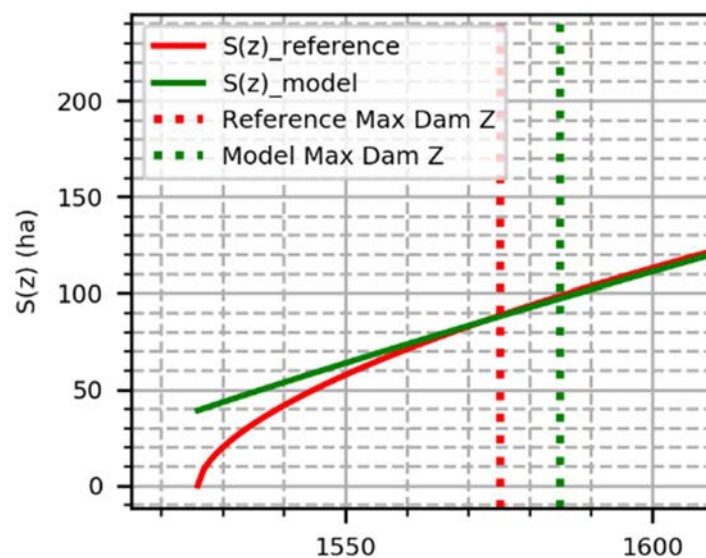


Figure 60 : Comparaison $S(Z)$ avec la référence

En conséquence cette erreur se répercute fortement sur l'estimation de la loi $V(S)$ comme présenté ci-dessous avec le modèle estimé en vert et le modèle *in situ* de référence en rouge.

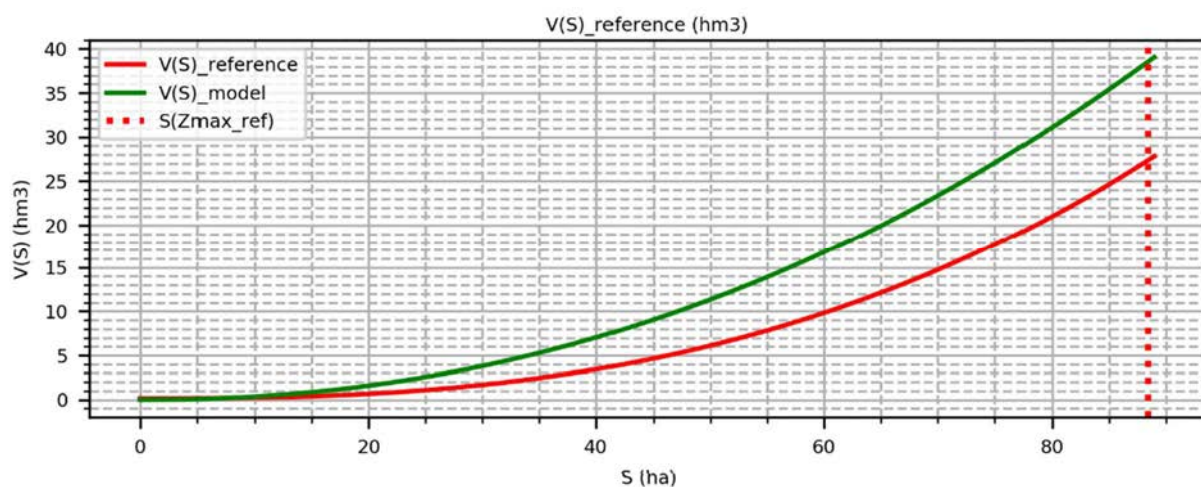


Figure 61 : Comparaison $V(S)$ avec la référence

Boucle ligne de coupure

Le cas du barrage Yeguas en Andalousie montre comment la détermination automatique de la ligne de coupure de la vallée, calculée à partir de l'exploitation du MNT et de la localisation du barrage via la base de données des réservoirs, se perd en boucle. Ce n'est cependant pas une anomalie bloquante, avec un impact mineur sur la génération de la loi HSV de ce barrage.



Figure 62 : Boucle de la ligne de coupure

Water crossing

Un exemple de cas de water crossing sur la ligne de coupure pour le barrage Jose Toran. Au regard de la dimension totale de la retenue d'eau, ce n'est pas une anomalie bloquante pour le calcul des lois HSV d'autant que la surface d'eau concernée est très faible.



Figure 63 : Water crossing

Le site Andevalo cumule les 2 précédentes anomalies sur la ligne de coupure de la vallée.



Figure 64 : Boucle et water crossing de la ligne de coupure

Multi-déversoirs

La retenue d'eau Guadálteba présente la particularité d'avoir plusieurs ouvrages, la ligne de coupure de la vallée n'identifie qu'un ouvrage, puisque celui-ci est géolocalisé par un unique point en latitude et en longitude fourni par la base de données des réservoirs.

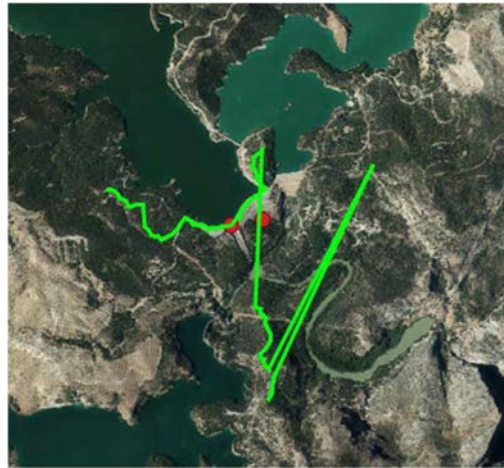


Figure 65 : Multi-déversoirs de Guadálteba

Les anomalies sont répertoriées sur la page Confluence.

3.3.3 Analyse séries temporelles $S(t)$, $V(t)$, $Tr(t)$

Lors des phases I et II, il a été effectué une analyse des séries temporelles issues des fenêtres glissantes Optique et Radar. En effet, ces types d'extraction de surfaces se sont avérés les plus pertinents et les plus robustes dans le cadre de cette étude.

Pour cette analyse des résultats, un tableau récapitulatif type a été mis en place, afin d'obtenir un rendu permettant de structurer clairement les résultats issus de la chaînes Dem4water et SURFWATER_Postprocess et ainsi de pouvoir évaluer la qualité des résultats plus facilement

Un outil (Notebook) a été mis en place afin de générer automatiquement ce tableau et ainsi pouvoir obtenir rapidement un résumé de la qualité des séries temporelles et $V(S)$ générées.

Cette méthodologie et ces outils seront utilisés également pour l'analyse des résultats obtenus dans le périmètre des activités complémentaires.

Descriptif des colonnes :

- ID_SWOT : identifiant SWOT.
- Dam_name : Nom du barrage.
- $V(S)$ -(%) : Moyenne des écarts ' $V(S)$ ' sur trois tranches (low = 0-33%, medium= 33%-66% et high= 66%-100% du $Z(t)$ observé insitu) permet d'évaluer la qualité de la lois $V(S)$ Dem4Water
- TR_{MO2} -(%) : {Médiane, Quantile 75, Quantile 90} des écarts relatifs absolus du taux de remplissage pour la fenêtre glissante optique (MO2).
- TR_{MR2} -(%) : {Médiane, Quantile 75, Quantile 90} des écarts relatifs absolus du taux de remplissage pour la fenêtre glissante radar (MR2).
- $Volume_{MO2}$ -(%) : {Médiane, Quantile 75, Quantile 90} des écarts relatifs absolus du volume pour la fenêtre glissante optique (MO2).
- $Volume_{MR2}$ -(%) : {Médiane, Quantile 75, Quantile 90} des écarts relatifs absolus du volume pour la fenêtre glissante radar (MR2).
- $Surface_{MO2}$ -(%) : {Médiane, Quantile 75, Quantile 90} des écarts relatifs absolus de la surface

pour la fenêtre glissante optique (MO2).

- Surface_MR2-(%) : {Médiane, Quantile 75, Quantile 90} des écarts relatifs absolus de la surface pour la fenêtre glissante radar (MR2).

Un outil (Notebook) a été mis en place afin de générer automatiquement ce tableau et ainsi pouvoir obtenir rapidement un résumé de la qualité des séries temporelles et V(S) générées.

3.3.3.1 Andalousie

Tableau 9 : Médiane – Andalousie

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2160003673	Cuevas de Almanzora	39.273590	52.000753	13.600589	25.626485	15.347581	8.618268	26.907596
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
2160004373	Guadalhorce	nan	9.399825	11.590819	28.829221	24.887802	8.656515	5.856785
2160004383	Guadaleba	7.770141	8.701750	14.059277	45.689051	39.708767	11.976490	5.518904
2160004403	Conde de Guadalhorce	12.626295	8.589965	11.507614	6.286464	7.820403	10.340930	6.243175
2160004413	La Vinuela	2.381152	3.931812	5.132870	11.746054	6.266542	5.998499	3.060634
2160004532	Limonero	4.718092	6.364851	9.277931	32.115855	32.920269	20.261746	21.441237
2160004623	Guadarranque	6.321110	5.881398	12.034120	27.475299	21.922930	11.707443	8.369460
2160004643	Charco Redondo	6.924094	6.879426	3.807375	12.284053	12.131625	11.979577	11.281453
2160004663	La Concepcion	9.723020	15.996454	19.292306	26.917818	17.914193	21.703991	15.974043
2310000173	San Rafael de Navallana	2.799966	5.660480	7.677122	20.582186	20.071743	14.871300	13.774601
2310018233	Los Hurones	24.467642	19.112084	21.553131	5.349799	6.620677	17.061411	16.065529
2310018272	Almodovar	99.484707	66.176565	53.459215	99.775559	99.736404	8.860020	7.155868
2310018393	Zahara	1.424229	2.901947	2.013763	23.950933	24.856572	14.576631	15.197746
2310018403	Bornos	17.176650	6.960333	13.919609	3.263753	7.015188	13.315401	12.187291
2310018473	Barbate	13.846101	6.252567	6.384302	18.964340	15.967077	2.838404	2.893043
2310018492	Arcos de la Frontera	nan	3.997866	4.010099	39.146240	47.179720	nan	nan
2310020153	Guadalcazin 2	11.383586	4.442626	1.738500	17.196794	14.206992	5.552099	2.413985
2310020223	Celemin	3.023574	15.508215	23.069501	6.908778	10.903963	4.458372	5.036346
2310020933	Tranco de Beas	8.357065	4.658674	9.123152	5.783947	5.216348	2.239665	2.672078
2310022062	Arenoso	nan	5.165107	10.133631	47.199526	51.039086	27.050108	30.487939
2310022743	Yéguas	4.896721	11.748693	6.590446	23.611800	19.062845	15.036658	13.316359
2310022903	Los Bermejales	7.020355	30.410327	7.619637	34.673715	7.290442	19.158726	5.569576
2310022963	Giribaile	13.827539	3.646985	7.985531	11.991661	11.547925	2.298460	2.529000
2310023523	Jose Toran	28.336494	22.374097	25.152428	11.117532	5.234120	22.683093	18.422614
2310023943	Vadomojon	12.321373	10.924405	5.427202	26.540506	26.313331	11.389361	11.640971
2310024073	Iznajar	5.649402	2.022242	1.833455	10.557061	9.409912	9.204915	8.307563
2310024113	Puebla de Cazalla	0.917271	8.384789	3.818534	32.865596	31.226055	20.386547	19.357538
2310024253	La Brena II	2.688674	60.031593	60.526026	13.413693	10.604680	11.238823	9.568880
2310027322	Los Melonares	78.990601	9.586464	8.863212	81.787265	81.746546	13.654065	13.308714
2310027683	Puente Nuevo	15.968427	3.134225	3.598467	5.259441	8.449434	6.742578	3.878412

Tableau 10 : résumé résultats médiane - nombre réservoirs par classe - andalousie

	V(S)- (%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
<= 10 %	17	22	19	6	9	11	15
> 10 & <=25 %	8	6	10	12	14	18	13
>= 25 %	4	4	3	14	9	2	3

3.3.3.2 Occitanie

Tableau 11 : Médiane - Occitanie

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
2160026973	Salagou	90.151847	1.994592	2.206573	92.350677	92.622686	3.262701	5.138346
2160028013	Matemale	34.228502	18.905880	9.413975	27.455387	29.428643	19.190365	6.643194
2160029873	Villeneuve la Raho	99.957016	5.180678	8.938131	99.912488	99.918831	5.886598	4.429572
2160029943	Puyvalador	19.861078	18.080310	8.243660	37.212586	32.124960	14.293708	7.439838
2160030123	Vinca	77.570885	13.870660	22.874459	82.489227	82.684684	18.136976	20.620443
2160030183	Avene	6.105639	6.572220	8.494587	46.905418	63.928362	21.491102	43.029072
2160030553	Agly	5.711032	15.352590	13.048004	18.743270	27.795971	15.390157	21.328310
2320028893	Pareloup	33.081892	9.048991	11.485365	12.965749	12.219161	9.349322	9.383501
2320028933	Saint Geraud	nan	13.680201	12.440003	25.141987	43.413075	23.696824	20.303942
2320030233	Montbel	87.039512	11.663044	24.563921	82.680066	83.491235	15.145812	17.984247
2320030823	Cammazes	3.318017	8.911244	9.341485	41.878625	60.128836	30.307358	44.532022
2320031293	Laparan	69.055075	79.299340	46.978939	41.603984	63.127562	37.708797	51.650149
2320031303	Pla de Soulcem	67.837554	25.321298	27.341403	27.991506	25.215194	14.790534	27.027582
2320038733	La Gimone	7.438983	6.294627	4.660892	17.290212	24.563354	8.902703	12.511576
2320039133	Astarac	13.847274	25.230449	26.171052	20.669356	31.967469	8.453292	14.001586

Tableau 12 : Résumé résultats médiane - Nombre reservoirs par classe - Occitanie

	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
<= 10%	4	6	7	0	0	5	5
> 10 & <=25%	2	6	5	4	2	8	6
>= 25%	8	3	3	11	13	2	4

3.3.3.3 Inde

Tableau 13 : Médiane - Inde

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
4530212043	Osman Sagar	115.957147	312.575013	101.380842	123.039879	42.071812	24.134926	16.404753
4530212143	Himayat Sagar	44.017308	16.824274	35.168083	40.409705	44.481561	24.233685	16.480837
4530347283	Nagarjuna Sagar	89.965929	204.033804	155.473813	573.274159	389.458117	19.510134	17.121652

3.3.3.4 Analyse des résultats

On retrouve globalement 2 types d'erreurs :

- 1) Erreurs dues à une mauvaise estimation de la loi $V(S)$. Ces mauvaises estimations de la loi $V(S)$ par Dem4water peuvent avoir plusieurs causes (voir <https://confluence.cnes.fr/pages/viewpage.action?pageId=98312301>) :
 - ✓ Des problèmes de multi-déversoirs (présences de plusieurs barrages sur un réservoir, etc)
 - ✓ Calcul automatique du pied de barrage
 - ✓ Calcul automatique de la ligne de coupure de la vallée

À noter qu'il est parfois difficile d'identifier les problèmes d'estimation de la loi $V(S)$ uniquement à partir de la moyenne des écarts 'V(S)_quality_low_mean', 'V(S)_quality_mid_mean' et 'V(S)_quality_high_mean' absolus, car une compensation des erreurs aux différents niveaux est possible. Il est donc parfois nécessaire de prendre en compte les valeurs des erreurs obtenues à chacun des 3 niveaux (Low, Mean, High).

- 2) Erreurs dues à des écarts importants entre les surfaces SURFWATER et les surfaces de références. Ces écarts peuvent s'expliquer par :
 - ✓ Des erreurs de classification de la part de SURFWATER. Ce type d'erreur est illustré par une dispersion des écarts avec des écarts variables tout au long de la série (cf. Figures 66, 67, 68),

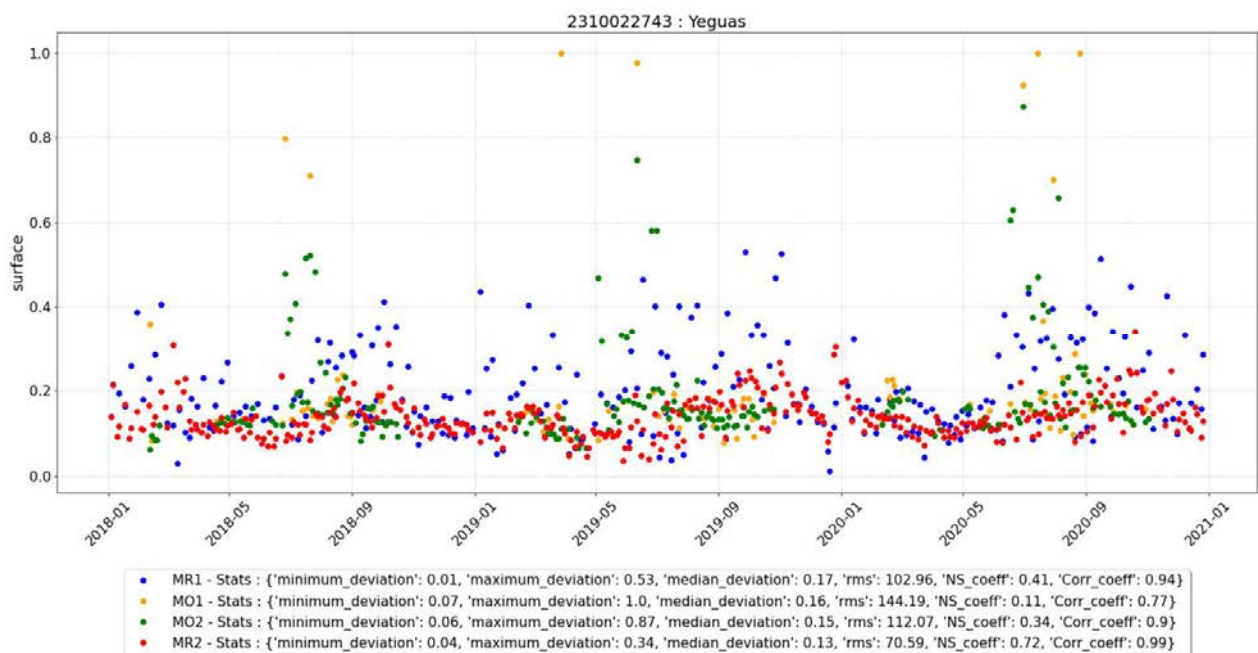


Figure 66 : Série temporelle des écarts relatifs absolus – Yeguas (Andalousie)

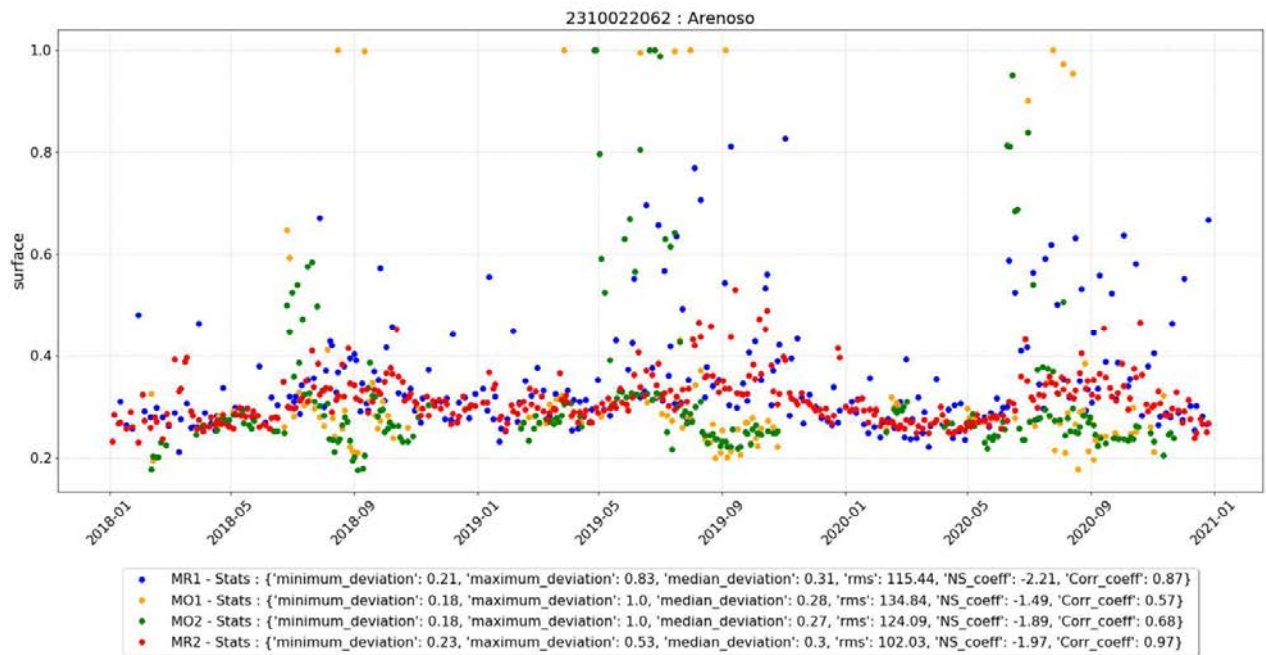


Figure 67 : Série temporelle des écarts relatifs absolus – Arenoso (Andalousie)

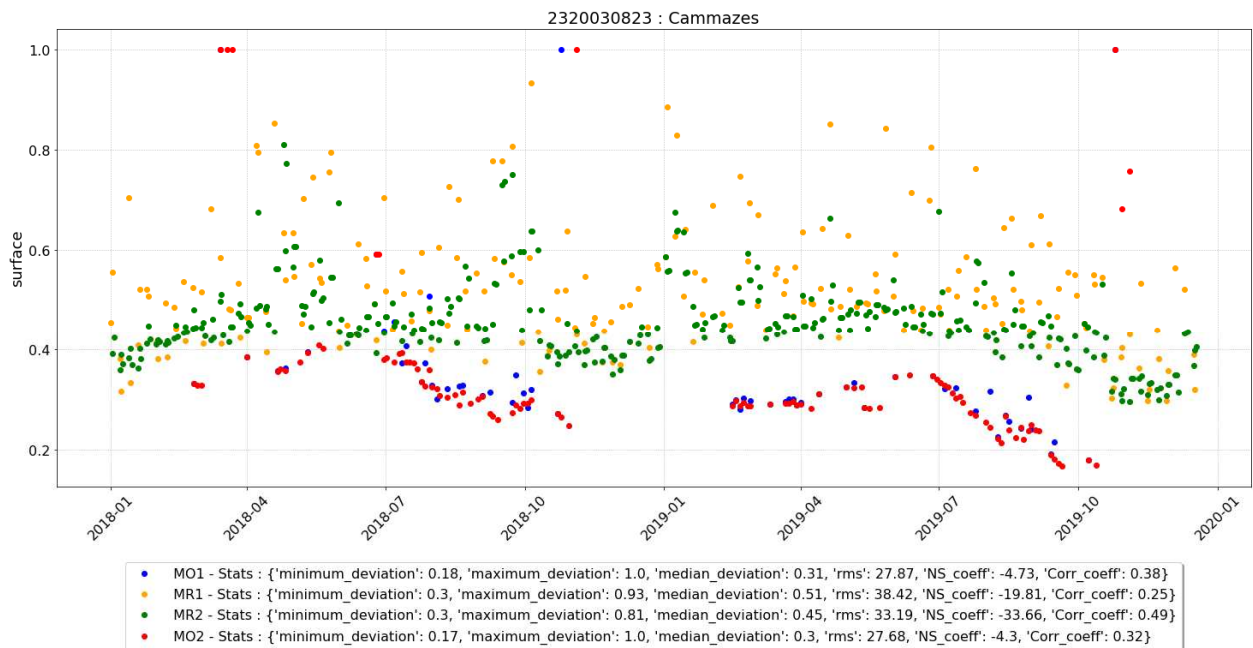


Figure 68 : Série temporelle des écarts relatifs absolus – Arenoso (Andalousie)

- ✓ Des données *in situ* de mauvaise qualité. Ces erreurs peuvent être identifiées par un écart constant le long de la série avec une dispersion faible des écarts et une bonne corrélation des écarts entre les méthodes d'extraction des surfaces (cf. Figures 69, 70, 71).

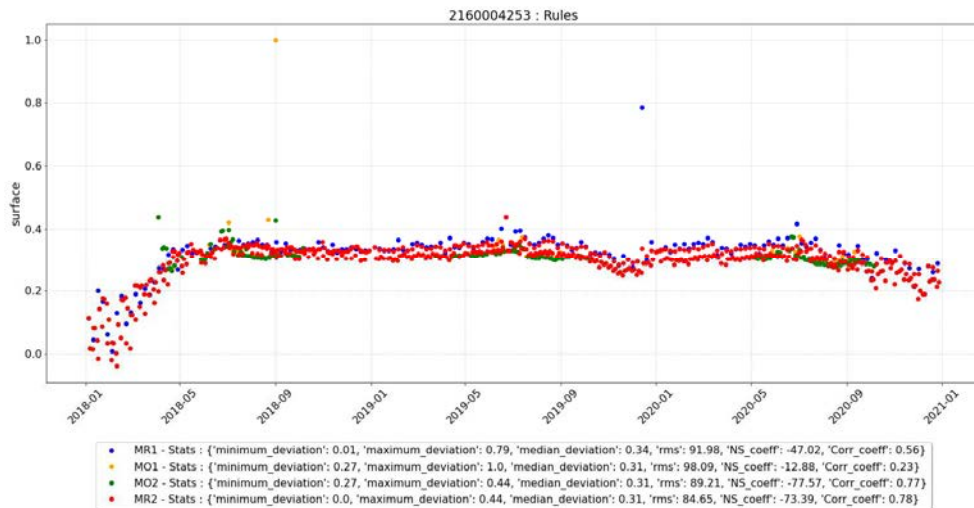


Figure 69 : Série temporelle des écarts relatifs absolus – Rules (Andalousie)

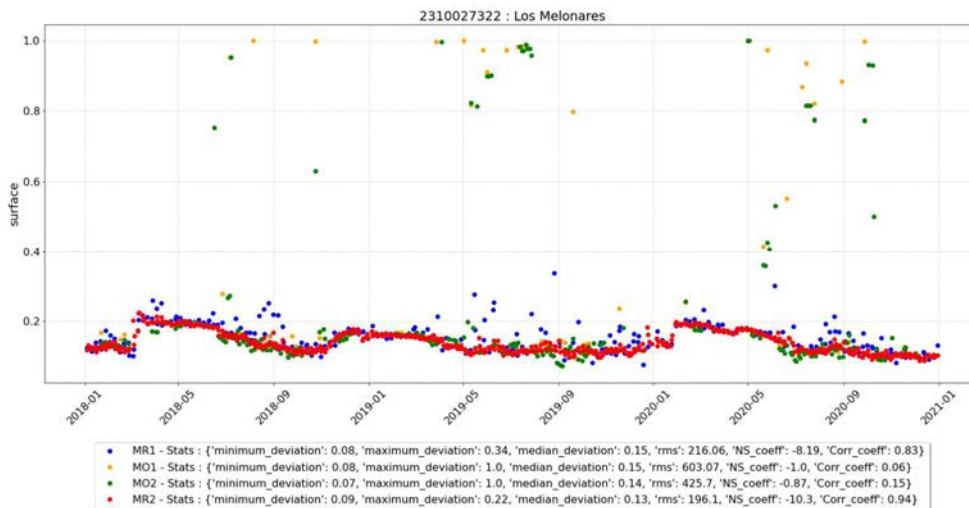


Figure 70 : Série temporelle des écarts relatifs absolus – Los Melonares (Andalousie)

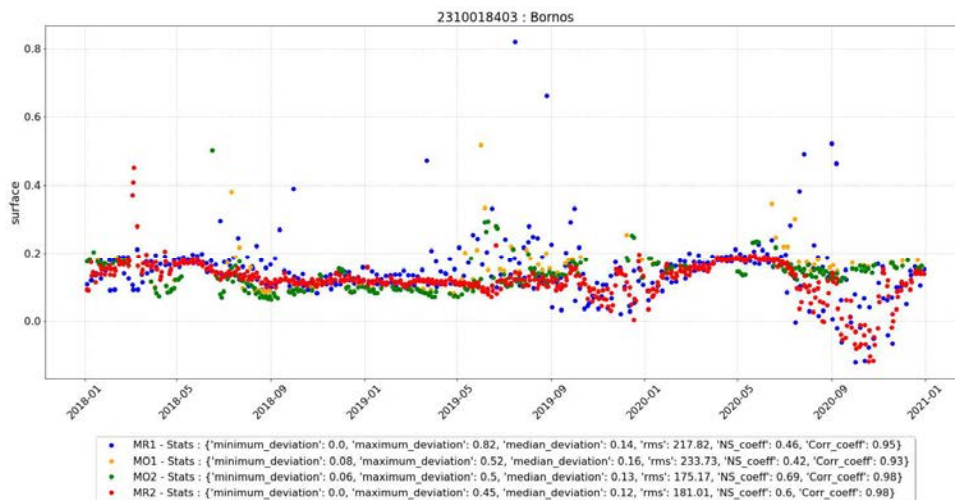


Figure 71 : Série temporelle des écarts relatifs absolus – Bornos (Andalousie)

Ces deux types d'erreurs entraînent quasi-systématiquement des mauvaises estimations des volumes qui sont directement dérivés de ces lois et surfaces SURFWATER. (cf. tableaux 14, 15, 16).

Tableau 14 : Exemple d'erreurs d'estimation des surfaces impactant l'estimation des volumes - Andalousie

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2160004643	Charco Redondo	6.924094	6.879426	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

Tableau 15 : Exemple d'erreurs V(S) impactant l'estimation des volumes - Andalousie

ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
2310020153	Guadalcazin 2	11.383586	4.442626	1.738500	17.196794	14.206992	5.552099	2.413985
2310022963	Giribaile	13.827539	3.646985	7.985531	11.991661	11.547925	2.298460	2.529000

Tableau 16 : Exemple d'erreurs V(S) impactant l'estimation des volumes - Occitanie

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	5.138346

À noter qu'il arrive parfois que ces erreurs se compensent. De mauvais résultats pour l'estimation de la loi V(S) et l'extraction des surfaces par SURFWATER peuvent donc engendrer des volumes qui semblent corrects. (cf. tableau 17)

Tableau 17 : Exemple compensation des erreurs – Andalousie

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	7.015188	13.315401	12.187291

Taux de remplissage :

Concernant les taux de remplissage, on remarque globalement des écarts corrects entre les taux de remplissage issus de SURFWATER post_process et les taux de remplissage issus de des données de références.

Il est tout même nécessaire de considérer ces résultats avec précautions car les taux de remplissage SURFWATER post_process et de référence se fondent sur leurs volumes maximums respectifs. Ce sont donc des données relatives à un volume maximum de remplissage qui doit être observé dans la fenêtre d'observation (normalement 3 ans suffisent pour observer ces périodes de remplissage).

Les remarques précédentes se fondent sur l'analyse des médianes des écarts relatifs absolus. Si on se fonde sur des métriques plus restrictives, tels que les quantiles 75%, on observe une erreur de taux de remplissage de 13.02 % en optique (MO2) et 11.56 % en radar (MR2). En quantile 90%, on constate une très forte augmentation des écarts, avec des erreurs quasi-systématiquement supérieures à 25%.

En quantile 75%, les erreurs de TR sont les suivantes:

Sensor = MO2

15.168395843765497 % (Moyenne de quantile 75% de TR pour chaque barrage)

13.024173096475 % (quantile 75% de TR de toutes les mesures confondues)

Sensor = MR2

12.614921980481494 %

(Moyenne de quantile 75% de TR pour chaque barrage)

11.56787438345% (quantile 75% de TR de toutes les mesures confondues)

Une des explications de l'augmentation des écarts en fonction du quantile est la persistance d'erreurs au niveau des extractions de surfaces issues de SURFWATER et particulièrement à partir de l'imagerie optique (utilisé avec la configuration - paramétrage 1 non optimale). Ainsi, la majorité des surfaces extraites par SURFWATER sont proches des surfaces dites de référence, mais on observe une population non négligeable présentant de fortes différences. L'influence de cette population aura donc un impact plus ou moins important sur l'analyse des résultats en fonction du quantile (Faible à modéré suivant une analyse de la médiane et fort à prépondérant pour les quantiles 75 et 90) (cf. [Annexe A](#)).

Pour améliorer la robustesse de l'approche, un indicateur hebdomadaire a été évalué en utilisant les valeurs médianes de chaque semaine naturelle. Ceci a peine changeait la qualité de mesures (n'élimine pas complètement les points aberrants), mais permet un suivi plus propre de tendances (notamment en radar). La figure suivante montre tels indicateurs de suivi de Volume et Taux de remplissage, et permet d'observer les effets de compensation en utilisant le taux de remplissage.

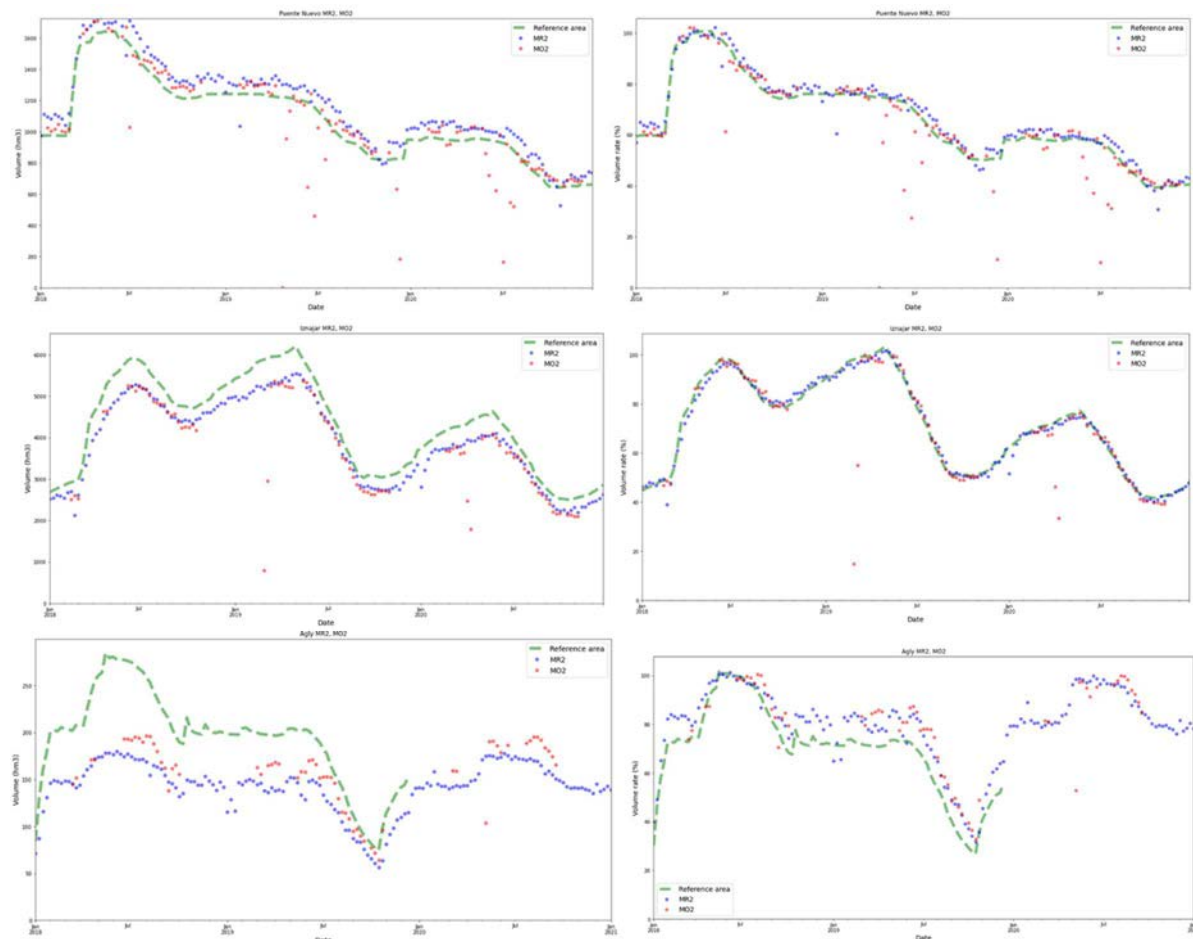


Figure 72 : Comparaison d'estimation de volume hebdomadaire (graphiques gauche) avec l'estimation de Taux de remplissage (graphiques droite) pour les barrages Pueblo Nuevo, Iznajar et Agly. En rouge les estimations optiques (configuration problématique - MO2), en Bleu le Radar (MR2) et ligne verte la référence (volume / taux de remplissage)

3.3.3.5 Amélioration des résultats

A la fin de la Phase I/début de Phase II, des corrections sur les polygones de comptages Occitanie et Andalousie ont été effectuées afin d'améliorer les résultats obtenus.

On remarque une nette amélioration de la qualité des résultats, et particulièrement pour les résultats issus de l'optique MO2.

En effet, pour l'Occitanie MO2, on constate que :

- ✓ 75% des barrages ont un écart médian de moins de 10% avec les surfaces in situ (contre 33% au préalable),
- ✓ 6% des barrages ont un écart médian de moins de 10% avec les volumes in situ (contre 0% au préalable),
- ✓ 75% des barrages ont un écart médian de moins de 10% avec les taux de remplissage in situ (contre 40% au préalable).

Tableau 18 Médiane-Occitanie

	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

Tableau 19 résumé résultats médiane - nombre réservoirs par classe - Occitanie

	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_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

Il en est de même pour l'Andalousie MO2, pour laquelle on constate que :

- ✓ 88% des barrages ont un écart médian de moins de 35% avec les surfaces in situ (contre 33% au préalable),

- ✓ 41% des barrages ont un écart médian de moins de 19% avec les volumes in situ (contre 0% au préalable),
- ✓ 79% des barrages ont un écart médian de moins de 69 avec les taux de remplissage in situ (contre 40% au préalable).

Tableau 20 Médiane-Andalousie

	ID_SWOT	Dam_name	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_MR2-(%)
14	2310018272	Almodovar	33.547278	48.866354	57.999633	99.718372	99.816166	5.042347	7.093556
18	2310018492	Arcos de la Frontera	nan	5.577141	4.709143	36.135734	48.276046	nan	nan
23	2310022062	Arenoso	nan	4.705129	11.294963	20.165054	50.695183	4.737193	30.101512
17	2310018473	Barbate	16.224382	5.009977	4.445845	15.518853	17.247211	1.485256	2.501627
1	2160004183	Beznar	6.439839	3.028011	5.291297	12.182245	8.471982	6.363693	4.166509
16	2310018403	Bornos	16.921904	7.341080	12.701782	2.421764	6.860859	10.018735	11.931699
8	2160004602	Casasola	29.030360	9.898731	18.340475	11.575996	12.454832	8.292090	12.438131
20	2310020223	Celemin	1.479886	10.264586	13.782729	4.546566	12.004184	1.779907	4.751051
10	2160004643	Charco Redondo	6.796969	4.017612	3.403600	4.325230	11.507426	2.014981	10.876315
5	2160004403	Conde de Guadalhorce	13.902395	6.083551	6.982693	8.589586	12.696954	3.855485	5.032415
0	2160003673	Cuevas de Almanzora	39.476015	38.753236	12.720896	52.093551	14.749333	7.200243	26.465142
26	2310022963	Giribaile	11.548433	2.889062	5.938083	16.203069	9.923382	2.979537	2.512043
19	2310020153	Guadalcanal 2	11.479984	3.043650	1.903974	10.508936	13.890824	1.848481	2.074174
22	2310022023	Guadalen	89.275800	34.556496	42.533256	89.253653	89.669937	4.318011	8.785087
3	2160004373	Guadalhorce	nan	4.737758	4.867605	37.616023	36.344581	3.687836	4.531548
4	2160004383	Guadaleba	73.066744	4.665842	6.392390	44.195736	41.905748	5.629889	3.597679
9	2160004623	Guadarranque	6.155025	3.488211	4.599676	11.881497	19.627326	3.326524	7.104867
29	2310024073	Iznajar	5.355205	1.950412	2.015796	4.592357	9.262872	5.383316	8.295655
27	2310023523	Jose Toran	28.495275	28.821432	24.805022	3.638160	4.842784	14.054625	17.213996
31	2310024253	La Brena II	3.068848	57.968239	60.310976	2.737888	10.002299	2.952571	9.679181
11	2160004663	La Concepcion	11.107910	10.583083	8.502378	5.955679	5.762209	6.606121	7.977125
6	2160004413	La Vinuela	2.902015	2.181131	9.361156	5.925951	6.836616	2.180723	3.474610
7	2160004532	Limonero	6.493850	4.060364	9.072407	20.501719	32.388372	10.836801	19.994032
25	2310022903	Los Bermejales	5.789899	2.750491	9.798661	4.757764	10.205666	1.443207	6.795088
13	2310018233	Los Hurones	23.623679	17.428371	17.182404	7.429911	4.461895	9.389569	13.754018
32	2310027322	Los Melonares	6.400711	7.315972	6.015562	80.499185	81.609172	6.232305	12.169398
30	2310024113	Puebla de Cazalla	0.928407	4.441831	5.084491	20.210513	27.205604	12.055494	16.645726
33	2310027683	Puente Nuevo	16.699458	9.704795	3.120744	16.711627	9.288584	1.598177	3.812201
2	2160004253	Rules	3.300350	1.809185	5.521387	14.965960	42.369308	7.072566	27.813227
12	2310000173	San Rafael de Navallana	3.159786	6.531996	10.216608	7.349391	18.920846	4.681572	13.248553
21	2310020933	Tranco de Beas	10.442158	5.431621	10.986788	8.311613	6.292428	2.869281	2.560959
28	2310023943	Vadomojon	12.727538	3.071674	2.899687	17.156901	24.485631	4.083173	9.842282
24	2310022743	Yeguas	5.100191	3.592094	13.396673	3.757266	17.468366	5.192712	12.503133
15	2310018393	Zahara	0.450902	2.907220	2.376253	14.379610	22.930431	8.763362	14.442386

Tableau 21 Résumé résultats médiane - nombre réservoirs par classe - Andalousie

	V(S)-(%)	TR_MO2-(%)	TR_MR2-(%)	Volume_MO2-(%)	Volume_MR2-(%)	Surface_MO2-(%)	Surface_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

On remarque également une amélioration moindre, mais tout de même notable pour les données MR2.

4 Déroulement de la Phase II

Lors de la précédente phase, la méthodologie dans son ensemble a été développée et mise en œuvre sur les super sites Andalousie, Inde et Occitanie, sélectionnés d'une part pour l'expérience déjà acquise sur ces certaines zones et d'autre part pour l'accessibilité aux données de référence.

Cette première phase aura permis de soulever un certain nombre de problèmes/limitations à différents niveaux du processus, accès et qualité des données *in situ*, accès aux données images, qualité des extractions optiques, robustesse de la modélisation, etc.

La méthodologie a donc été perfectionnée dans son ensemble et de nouvelles solutions sont mises en œuvre lors de la Phase II :

- ✓ amélioration de la qualité de la classification optique en appliquant un paramétrage plus adapté à la chaîne Surfwater,
- ✓ résolution de limitations sur la génération des modèles par la chaîne DEM4water.

Outre le rejeu des certaines étapes des super sites de la Phase I, une mise à l'échelle est effectuée avec l'application de cet environnement étendue aux nouveaux super sites Tunisie, Burkina Faso, Laos et Brésil.

4.1 Développement du démonstrateur

4.1.1 Surfwater

De nouveaux problèmes de récupération sont apparus sur des tuiles du super site Brésil ne présentant aucune donnée Sentinel-2 récupéré (idem pour toute l'année 2021), bien que ces données soient présentes sur le datalake. Cette anomalie a été reproduite sur le commit précédent de la chaîne Amalthee par Robin Vermes. Le problème identifié est dû à la différence de version entre notre client Amalthee et le serveur, ceux-ci ne définissant pas les mêmes chemins finaux pour les produits qui ont l'extension V3-0. De plus Amalthee ne contrôle pas la version du client dans la requête faite et aucun warning n'est renvoyé si les versions ne sont pas compatibles.

Une mise à jour de la libAmalthee dans l'environnement de la chaîne Surfwater a été faite, cependant des anomalies de droits sur les sites-packages ont dû être également identifiés et résolus, des bibliothèques ayant les droits inégalement répartis dont certaines en exécution mais pas en lecture.

4.1.2 DEM4water

Les activités ont porté sur la robustification de la chaîne DEM4water, afin d'améliorer son comportement et ses performances dans le cadre de son utilisation en production à large échelle.

DEM4Water est à la base un prototype dont l'objectif initial était de démontrer la faisabilité de l'automatisation de la modélisation des relations Surface/Hauteur et surtout surface/Volume dans les réservoirs artificiels. L'objectif amont est de permettre le suivi par télédétection des capacités quasi-temps réels de ces réservoirs artificiels.

Les améliorations ont été principalement fonctionnelles et algorithmiques :

- ✓ Diminuer les temps de retraitement en implémentant un système de caches des lignes de niveaux générées à partir du DEM,
- ✓ Améliorer la paramétrisation de la chaîne en permettant à l'utilisateur de substituer aux éléments calculés automatiquement des entrées corrigées de manière experte par l'utilisateur. Ce faisant, il est possible, dans les cas où la chaîne produit des résultats peu satisfaisants du fait de détection automatiques de mauvaise qualité, de palier manuellement à ces limitations :
 - Dans le cas d'une mauvaise détection du pied de barrage, qui impacte la qualité de l'estimation du modèle ;
 - Dans le cas de la ligne de coupure :

- Afin de gérer les cas où une retenue comporte plusieurs barrages ;
 - L'algorithme ne parvient pas à atteindre des altitudes suffisantes pour permettre la collection de suffisamment d'échantillons $S(Z_i)$ sur lesquels repose l'estimation du modèle.
- ✓ Dériver une ligne de coupure satisfaisante dans les cas où :
- L'algorithme de détermination de la ligne de coupure génère des géométries trop complexes (*MultiLineString*) ;
 - La ligne de coupure comporte des boucles (tracé rouge pointillé), elles sont alors *naïvement* simplifiées (Figure 73, ligne verte) :



Figure 73 : correction des anomalies de lignes de coupure

- ✓ Améliorer la gestion des cas limites, avec un meilleur reporting :
- Gestion des lancements de comparaison à une vérité terrain quand celle-ci n'existe pas (induit par la façon dont est utilisée la chaîne par l'orchestrateur de Stockwater, idéalement, sans données terrain, le calcul de la qualité du modèle ne devrait pas être lancé),
 - Gestion des cas où l'algorithme de détection du pied de barrage ne converge pas.

4.1.3 Surfwater_Postprocess

Un nouvel indicateur a été développé dans le cadre de la chaîne Surfwater_Postprocess générant les séries temporelles. Cet indicateur hebdomadaire de type médian est appliqué sur les séries temporelles de surface $S(t)$, de volume $V(t)$ et de taux de remplissage $TR(t)$ d'une part, et d'autre part sur les différences avec la référence ainsi que sur leurs valeurs absolues. Le module de comparaison des métriques a également évolué pour générer les graphiques relatifs à cet indicateur hebdomadaire pour chacune des quantités ainsi que les fichiers résultats au format json.

Les actions sur l'extraction des surfaces réalisés par P.Tysebaert ont été intégrés à la chaîne Surfwater_Postprocess. Un soutien technique a été apporté par CS dans le cadre de ces travaux, ainsi que la création d'un jeu de test complet avec mise à disposition de la procédure.

4.1.4 Orchestrateur Stockwater

L'orchestrateur Stockwater permet de réaliser l'enchaînement et la reprise des traitements pour un super site étudié. Certaines étapes nécessitent impérativement une validation opérateur avant que la

suivante ne soit lancée, comme la production des masques d'eau Surfwater et la génération des modèles estimés. La création/mise en place du contexte automatiquement géré par l'orchestrateur simplifie grandement la reprise des traitements après les étapes de contrôle opérateur.

Quelques évolutions complémentaires ont été apportées à l'orchestrateur : l'ajout dans le fichier paramètre de la possibilité d'avoir recourt à la carte d'occurrence et un percentil associé suite à l'évolution de la chaîne Surfwater_PostProcess, le calcul des cartes d'occurrence sur la période étudiée pour le mode radar, la génération du vrt des cartes d'occurrence pour le mode radar.

Les données auxiliaires composées des bases de données, des lois estimées, des lois de références et des séries temporelles de référence sont mises à jour en fonction des corrections apportées aux données ou de l'ajout de nouveau super site étudié.

Dans le cadre de l'étape de valorisation des résultats du projet Stockwater via la plateforme web, il a été nécessaire de développer des outils permettant d'extraire individuellement les barrages des fichiers de séries temporelles estimées d'une part, et d'autre part des séries temporelles de référence.

Un formatage a été réalisé pour obtenir un fichier csv de séries temporelles par identifiant de barrage, pour un répertoire méthode (MO1/MR1/MO2/MR2) avec une modification de format enfin de mettre le header aux normes (champs « date, value ») et une nomenclature de fichier de type : [dam_id]_[dam_name]_[quantity]_raw_[method].csv

Les séries temporelles de référence, sous forme de fichier unique par super site, ont également nécessité un formatage pour l'extraction des données par identifiant de barrage.

Par la suite, un filtrage des données a été appliqué sur les séries temporelles estimées :

- ✓ supprimer les valeurs des quantités à 0 ainsi que les valeurs pour lesquels les no-data sont à 0,
- ✓ s'il existe plusieurs mesures pour la même date, conserver la plus grande valeur.

Un filtrage a été réalisé sur les séries temporelles de référence :

- ✓ restreindre à la période d'observation à celle de l'étude Stockwater,
- ✓ garder une seule mesure par jour : à 00:00:00,
- ✓ insérer dans la nomenclature du fichier le nom du barrage en plus de l'identifiant et supprimer le nom du super site.

4.2 Mise en place de la base de données

4.2.1 Tache 2.1 – Sélection de cibles

La sélection des cibles est opérée dans la continuité de la Phase I. Les zones analysées dans cette Phase II sont au Brésil, au Burkina Faso, en Tunisie et, dans une moindre mesure au Laos.

Tunisie

Les échanges avec le LISAH (Jean-Stéphane Bailly) et INRAE (Jérôme Molénat) ont été pris en compte lors de la sélection des 10 sites illustrés par la figure 90. 6 sites sont localisés sur la péninsule du Cap Bon (Lebna, Chiba, Kamech, Abid, Bezirk, Masri) et 2 à proximité (El Hamma, Er Rmal), des données *in situ* sont attendues pour ces 8 sites. Le site de Kamech est le plus petit (~10 ha), c'est un lac collinaire suivi dans le cadre d'OMERE (www.obs-omere.org). Les 2 autres sites retenus (Sidi Salem et Sidi Saad) sont des réservoirs de grande taille (plus de 10 km²) et sont répertoriés BD GRanD. Les géométries d'extension maximale des surfaces en eau sont basées sur Global Surface Water, Geodan et GRanD. Les géométries de 6 sites ont toutefois dû être ajustées par le SERTIT.

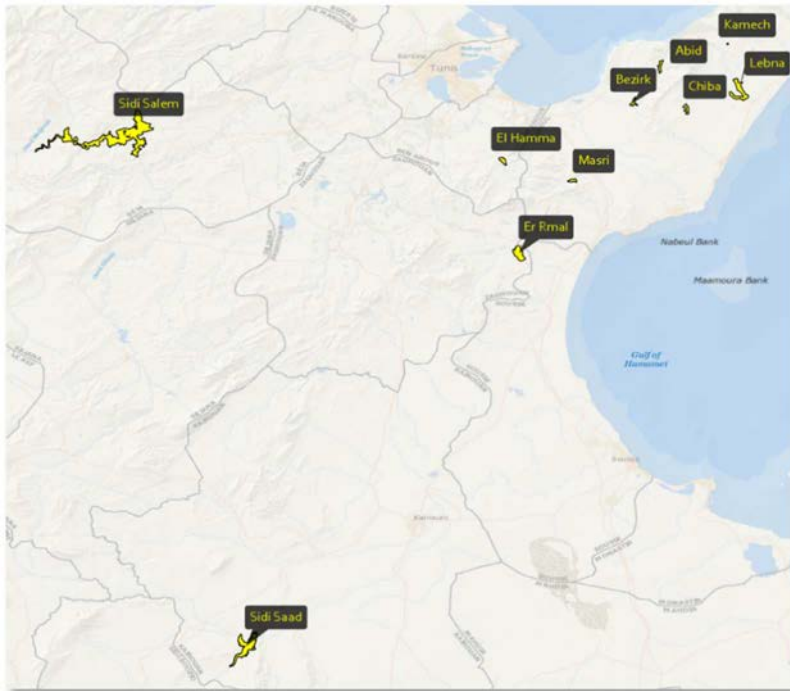


Figure 74: Sites Tunisie

Brésil

Les échanges avec Marielle Gosset (IRD) et les équipes brésiliennes ont été pris en compte lors de la sélection des 14 sites localisés dans la province Ceara (figure 75). Le site de Bonito est le plus petit (~9 ha), il est conservé en test malgré sa petite taille. 8 sites sont sous l'orbite SWOT 1 jour. Les géométries d'extension maximale des surfaces en eau sont basées sur Global Surface Water, Geodan et GRanD. La géométrie d'un site a toutefois dû être ajustée par le SERTIT.



Figure 75 : Sites Brésil

Burkina Faso

Les priorités du GET (Manuela Grippa) ont été prises en compte lors de la sélection des 9 sites illustrés par la figure 76. Le site Tanvi2 est le plus petit (~56 ha). 3 sites retenus ne sont pas répertoriés dans la BD GRand. Les géométries d'extension maximale des surfaces en eau sont basées sur Global Surface Water et GRand. Les géométries de 2 sites ont toutefois dû être ajustées par le SERTIT.

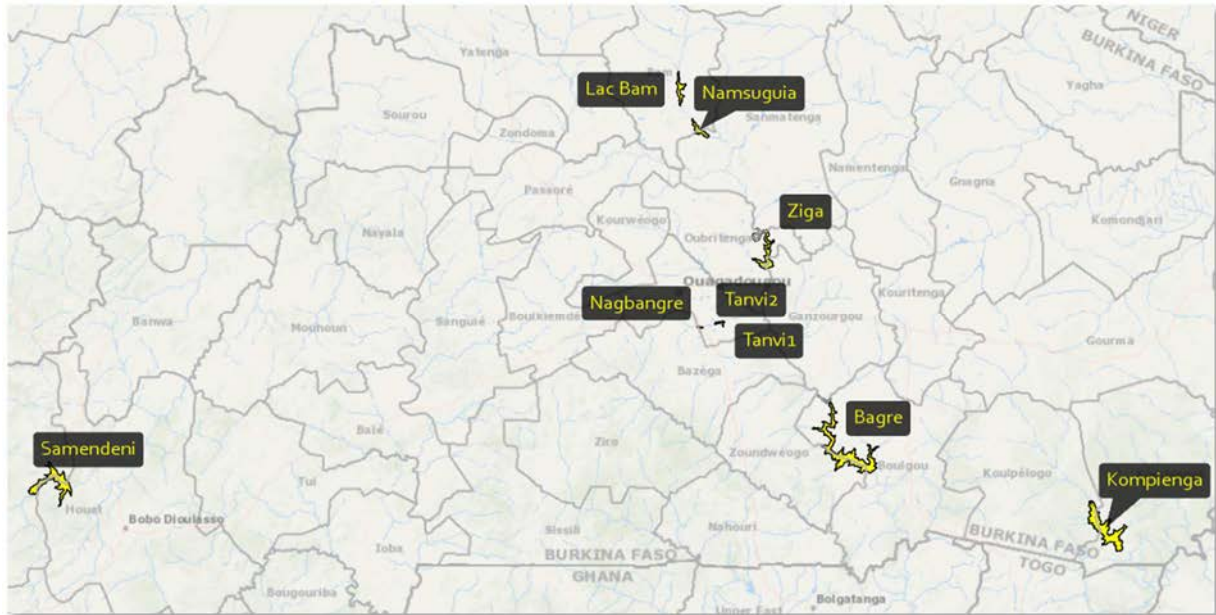


Figure 76 : Sites Burkina Faso

Laos

Le site Laos est le réservoir Nam Theun 2 de 540 km² illustré par la figure 77.

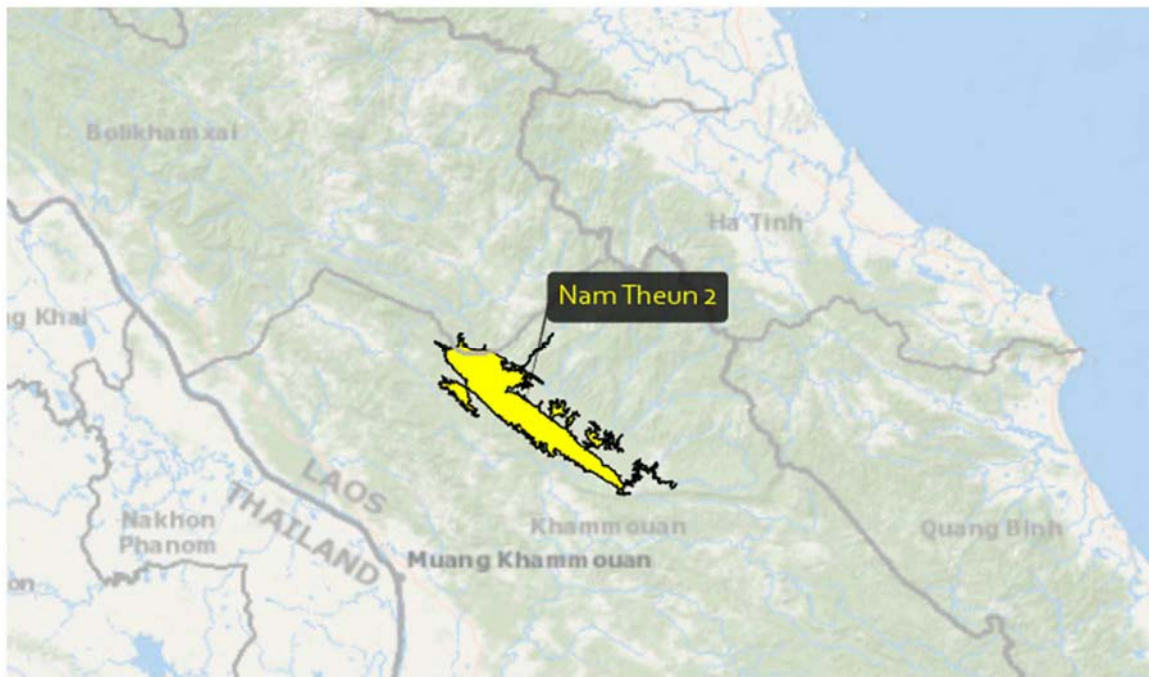


Figure 77 : Site Laos

4.2.2 Tache 2.2 – Préparation des données

A l'issue de la Phase I, un protocole a été mis en place pour la préparation des jeux de données (Dam Database). Ce protocole a été suivi pour la préparation des données Laos, Brésil et Burkina Faso.

4.2.2.1 Structure du fichier attributs

A l'issue de la Phase I, un protocole a été mis en place pour la préparation des jeux de données (Dam Database). Ce protocole a été suivi pour la préparation des données Laos, Brésil et Burkina Faso. (cf. Figure 78).

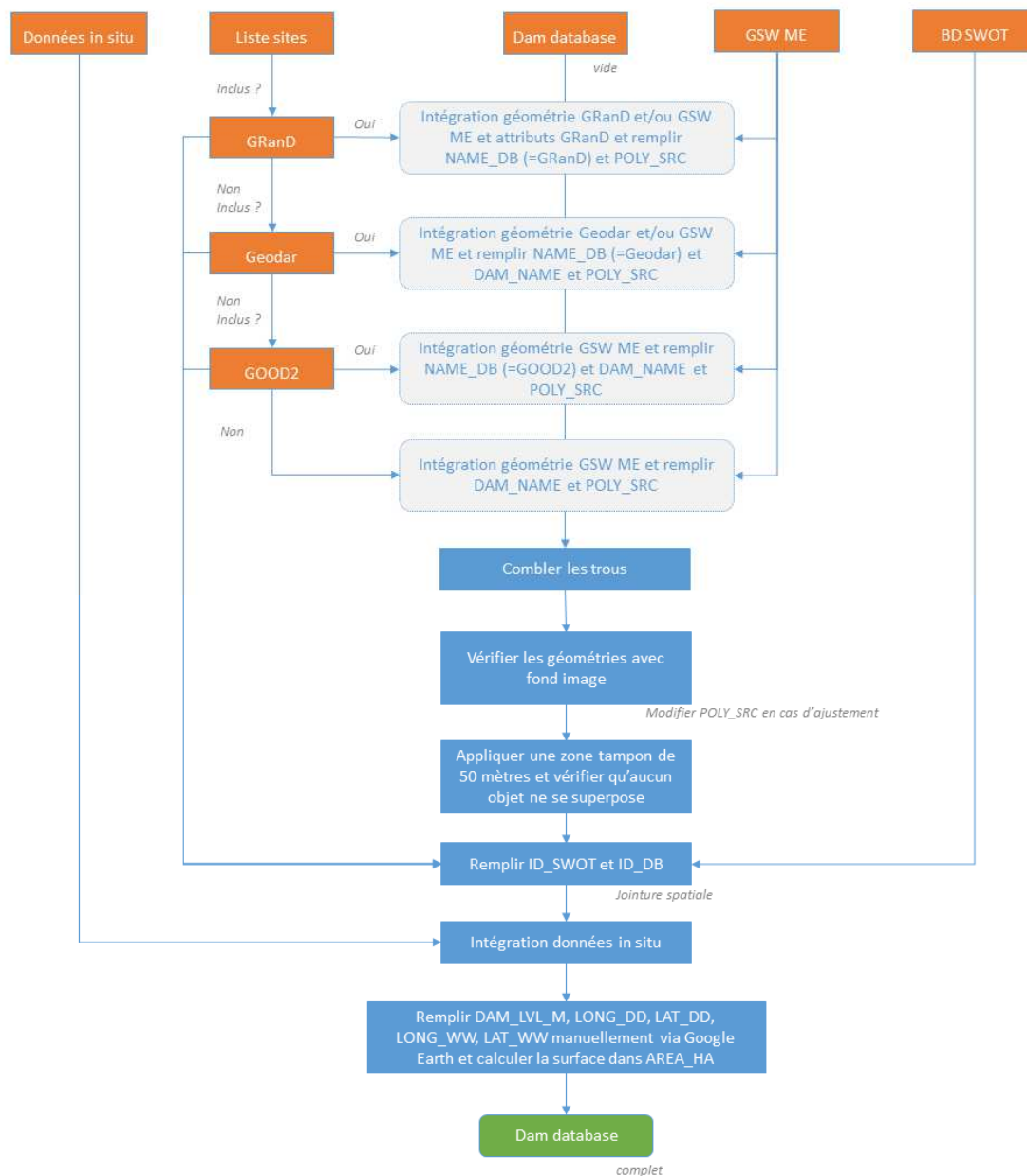


Figure 78 : Workflow de la préparation d'un jeu de donnée Dam Database

Géométrie

Cette section vise à obtenir les étendues maximales des réservoirs à partir de différentes bases de données (GRanD, Geodar et Global Surface Water Maximum Extent). Ces polygones d'extension sont intégrés à une base de données réservoirs *Dam Database* au format vector et qui dispose d'un modèle de données propre à respecter (cf. Phase I). Certains attributs associés aux étendues sont à remplir lors de l'import des géométries vectorielles pour garder la généalogie. Le workflow présenté dans la figure 78 reprend les éléments présentés dans cette section de manière plus visuelle et plus succincte.

- ✓ Disposer de la liste des sites sélectionnées ;
- ✓ Vérifier leur présence dans GRAND avec le nom dans les attributs. Intégrer ceux qui sont présent dans le vecteur *Dam Database* en conservant les attributs (automatiquement reconnus). Renseigner les champs NAME_DB = GRanD pour ces objets ;
- ✓ Pour les autres, vérifier leur présence dans GEODAR (repérage via google ou autre cartographie). Intégrer ceux qui sont présents dans le vecteur *Dam Database* et remplir les champs NAME_DB = Geodar et DAM_NAME ;
- ✓ Pour ceux qui ne sont ni dans GRAND ni dans GEODAR, vérifier leur présence dans GOOD2 (repérage via google ou autre cartographie). Il n'y a pas de géométrie surfacique associée. L'objet GSW ME peut être directement intégrée au vecteur *Dam Database* et remplir les champs NAME_DB = GOOD2 et DAM_NAME ;
- ✓ Pour les réservoirs qui ne sont pas répertoriés du tout, intégrer la géométrie GSW ME associée au vecteur *Dam Database* ;
- ✓ Pour les géométries issues de GRAND/GEODAR, il faut les fusionner avec les géométries GSW ME ou même les remplacer si celles-ci sont de mauvaise qualité (analyse visuelle),
- ✓ Combler les trous,
- ✓ Elargir manuellement les géométries en vérifiant avec un fond image pour être sûr de couvrir la surface maximale des réservoirs,
- ✓ Appliquer une zone tampon de 50 mètres autour des réservoirs,
- ✓ Vérifier qu'aucun objet ne se superpose dans le vecteur *Dam Database*. Si oui, réduire la zone tampon des objets concernés.

Attributs

Certains champs de la table attributaire sont déjà remplis lors de l'étape précédente mais il reste des informations à remplir. Le workflow présenté dans la figure 78 reprend les éléments présentés dans cette section de manière plus visuelle et plus succincte.

- ✓ **ID_SWOT** : Intersection spatiale manuelle avec BD SWOT. Cela peut être automatisé (jointure spatiale) mais attention, notre réservoir peut intersecter plusieurs objets de la BD SWOT. J'ai pris manuellement l'objet correspondant le plus grand quand c'était le cas,
- ✓ **ID_DB** : Intersection spatiale manuelle avec GranD, Geodar ou GOOD2. Jointure spatiale en cascade possible,
- ✓ **NAME_DB** : Nom de la BD utilisée pour remplir ID_DB. GRanD, Geodar ou GOOD2,
- ✓ **RES_NAME** : Parfois rempli par GRanD. Pour le Burkina j'ai indiqué les noms donnés par GET (exemple Burkina Faso),
- ✓ **DAM_NAME** : Rempli par GRanD. Pour ceux qui n'y figurent pas, j'ai indiqué les noms donnés par GET (exemple Burkina Faso),
- ✓ **NEAR_CITY** : Hérité de GRanD uniquement,
- ✓ **COUNTRY** : Hérité de GRanD uniquement. Sinon rempli manuellement,
- ✓ **YEAR** : Hérité de GRanD uniquement,

- ✓ **DAM_LVL_M** : Altitude lue dans Google Earth et reportée manuellement (ou données in situ).

Remarque 1 : La source de base pour les altitudes GE est SRTM mais sont intégrés localement des données à haute résolution provenant de diverses sources au fur et à mesure de leur disponibilité. En global, MERIT, NASADEM et ALOS DSM sont aussi disponibles dans Google Earth Engine. Il est difficile de connaître la source de l'altitude lue dans Google Earth. Pour connaître la source il faudrait lire sur un DEM dans un SIG ou dans GEE.

Remarque 2 : La composante verticale (altitude) est mesurée à partir du référentiel vertical, qui est le géoïde WGS84 EGM96. En général, l'altitude peut être considérée comme une mesure en mètres au-dessus du niveau de la mer (MSL)

- ✓ **CAP_MCM** : Hérité de GRanD uniquement ou donnée *in situ*,
- ✓ **DEPTH_M** : Hérité de GRanD uniquement ou donnée *in situ*,
- ✓ **MAIN_USE** : Hérité de GRanD uniquement,
- ✓ **LONG_DD** : Coordonnée lue dans Google Earth et reportée manuellement,
- ✓ **LAT_DD** : Coordonnée lue dans Google Earth et reportée manuellement,
- ✓ **LONG_WW** : Coordonnée lue dans Google Earth et reportée manuellement,
- ✓ **LAT_WW** : Coordonnée lue dans Google Earth et reportée manuellement,
- ✓ **POLY_SRC** : GSW(+GRAND+GEODAR+Refined by XXX),
- ✓ **AREA_HA** : Calculé à la fin.

4.2.2.2 Données de référence *in situ*

4.2.2.2.1 HAUTEURS D'EAU ET VOLUMES *IN SITU*

Contrairement à la Phase I, durant laquelle un grand nombre de données *in situ* était disponible sur les différents sites, lors de la Phase II, des données *in situ* de Hauteurs d'eau et Volumes ont uniquement été mises à disposition pour le site Brésil.

Comme réalisé lors de la Phase I, une étape d'analyse, de sélection et de correction de ces données *in situ* a été effectuée, pour générer des datasets opérationnels, compatibles avec le fonctionnement du PoC STOCKWATER.

Les quatorze sites brésiliens sélectionnés, disposent de séries temporelles *in situ* de hauteurs d'eau et de volumes.

Tableau 22 : Tableau récapitulatif des données *in situ* disponibles pour le site Brésil

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

4.2.2.2.2 ANALYSE DE LA QUALITE ET AMELIORATION DES DONNEES *IN SITU*

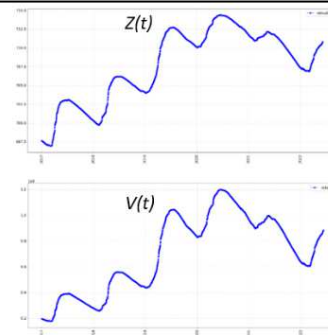
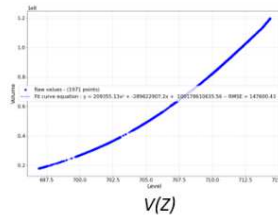
Comme constaté sur les différents sites lors de la Phase I, il s'est avéré que les données *in situ* (Hauteurs d'eau et Volumes) fournies pour les sites Brésil comportaient des incohérences, empêchant ainsi leur utilisation telles quelles pour la génération des lois *in situ*. Ces incohérences ont pu être corrigées, pour la plupart, grâce aux scripts mis en place lors de la Phase I.

Les données *in situ* des réservoirs 6330098593 – Castanhão, 6320015192 – Realejo, 6320007102 - São José III, 6320007492 – Jaburu I et 6330068463 - Edson Queiroz ont été conservées telles quelles, car elles ne présentaient pas d'incohérence majeure.

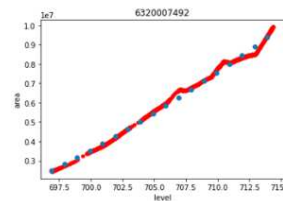
6320007492 - Jaburu I



Max extent = $1,2 \cdot 10^7 \text{ m}^2$



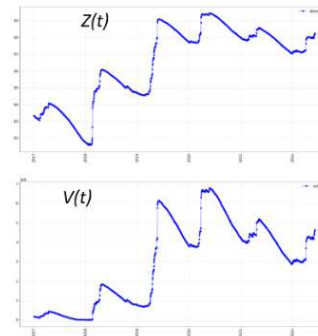
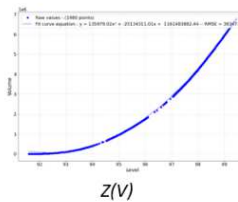
S(Z) LAW (blue) vs S(Z) TS (red)



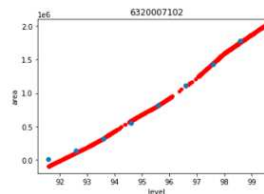
6320007102 - São José III



Max extent = $2,54 \cdot 10^6 \text{ m}^2$



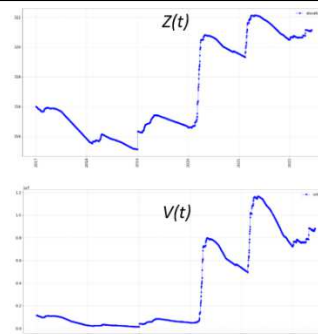
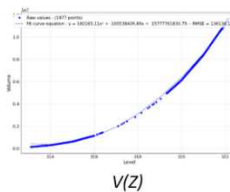
S(Z) LAW (blue) vs S(Z) TS (red)



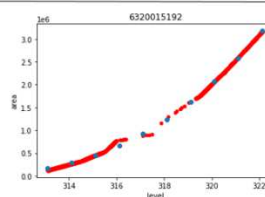
6320015192 - Realejo



Max extent = $6,92 \cdot 10^6 \text{ m}^2$



S(Z) LAW (blue) vs S(Z) TS (red)



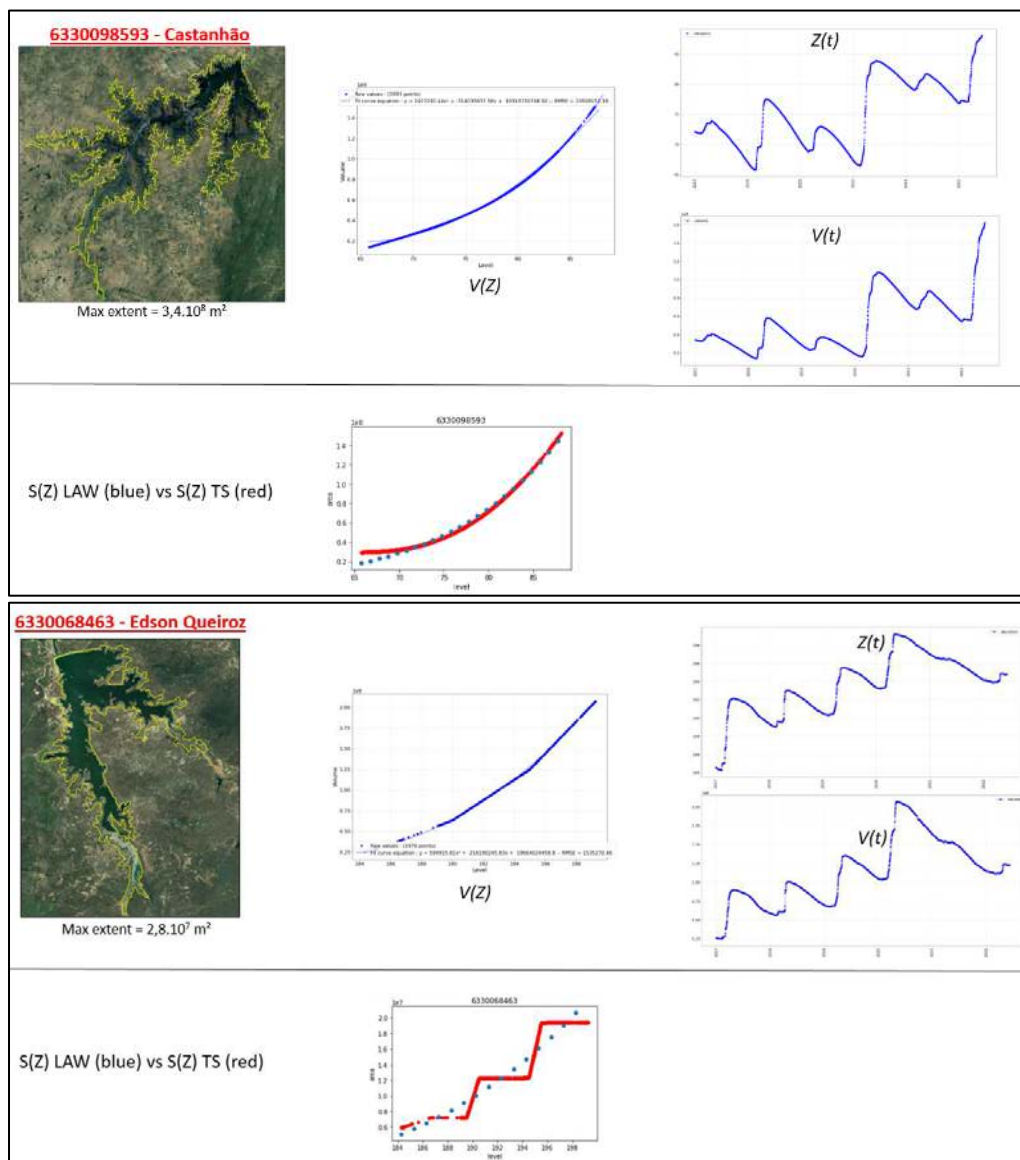
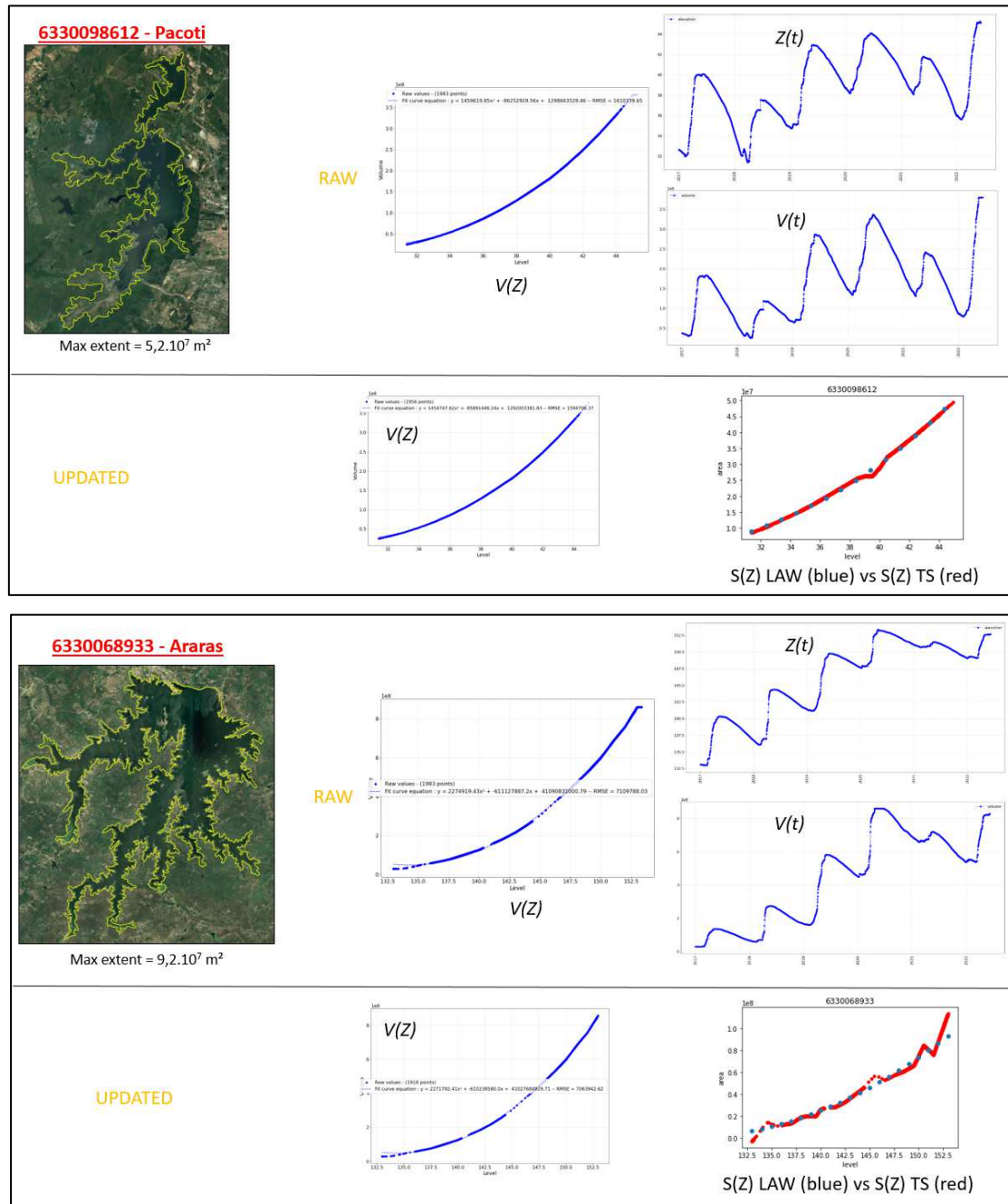


Figure 79 : Représentation des lois $V(Z)$ et $S(Z)$ pour les données *in situ* non modifiées

Pour les réservoirs 6330098612 – Pacoti, 6330068933 – Araras et 6330098332 - Acarape do Meio, des corrections ont été nécessaires pour éliminer les outliers et « plateaux » / données non-variables. Incohérences visibles sur les $V(Z)$.



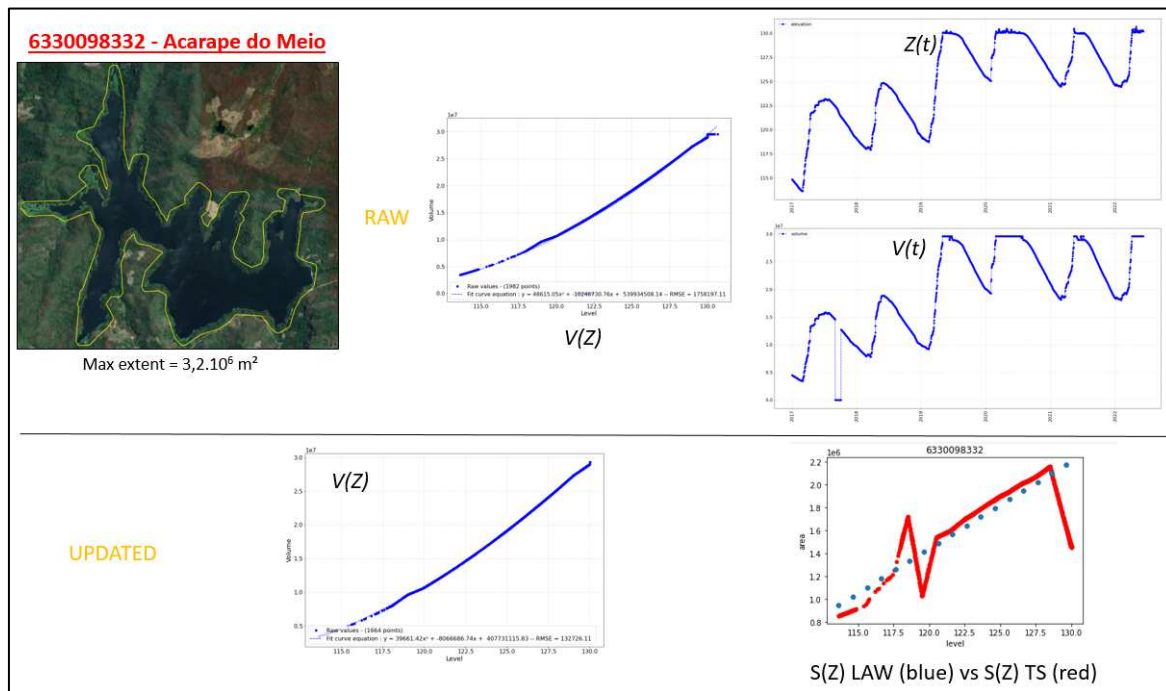
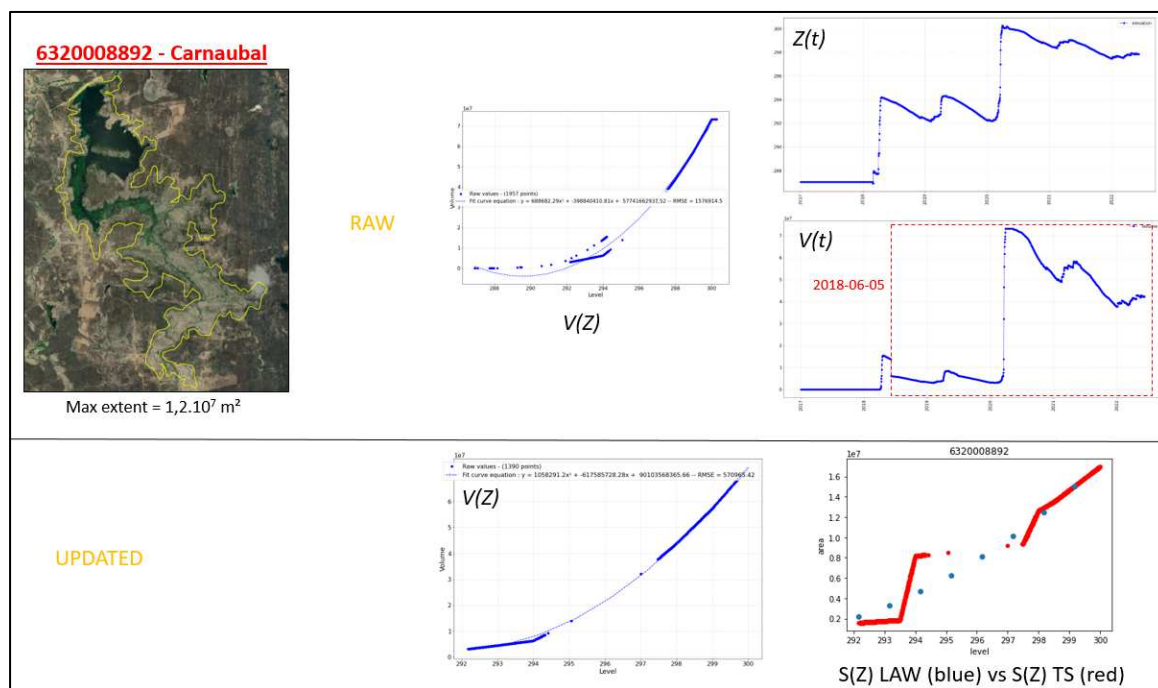


Figure 80 : Représentation des Lois $v(Z)$ et $S(Z)$ pour les données in situ corrigées (cas de plateaux)

Les données des réservoirs 6320008892 – Carnaubal, 6320009412 - Jaburu II, 6330070272 - Farias de Sousa, 6330095223 – Banabuiu et 6330000001 – Bonito présentent, quant à elles, des données dupliquées qui correspondent à des données ayant des volumes différents pour une même hauteur d'eau. Ces incohérences sont potentiellement dues à des mises à jour des modèles ou bathymétries des gestionnaires permettant de calculer les volumes à partir des hauteurs d'eau. Ces changements sont clairement identifiables lorsque l'on analyse simultanément le comportement des séries temporelles de Hauteurs d'eau et de Volumes.

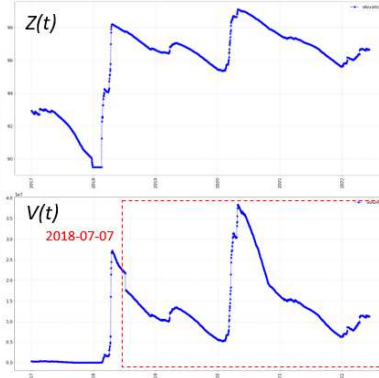
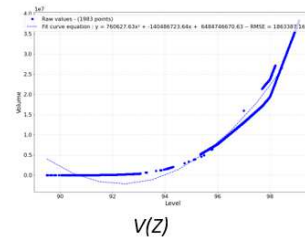


6320009412 - Jaburu II

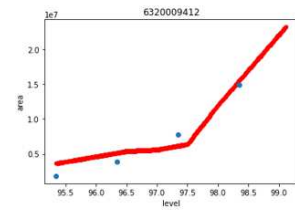
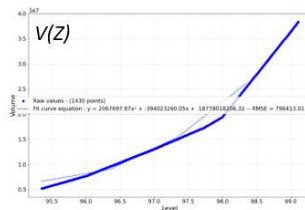


Max extent = $2,0.10^7 \text{ m}^3$

RAW



UPDATED



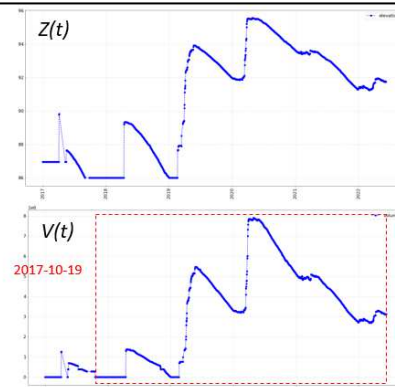
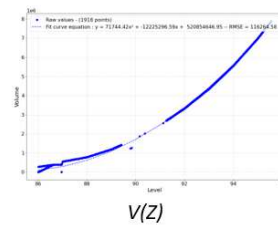
S(Z) LAW (blue) vs S(Z) TS (red)

6330070272 - Farias de Sousa

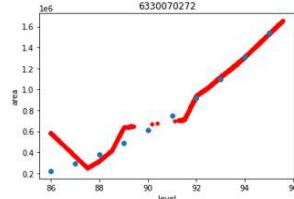
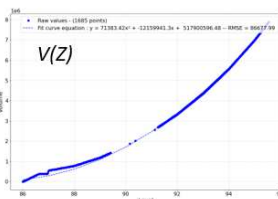


Max extent = $2,7.10^6 \text{ m}^3$

RAW



UPDATED



S(Z) LAW (blue) vs S(Z) TS (red)

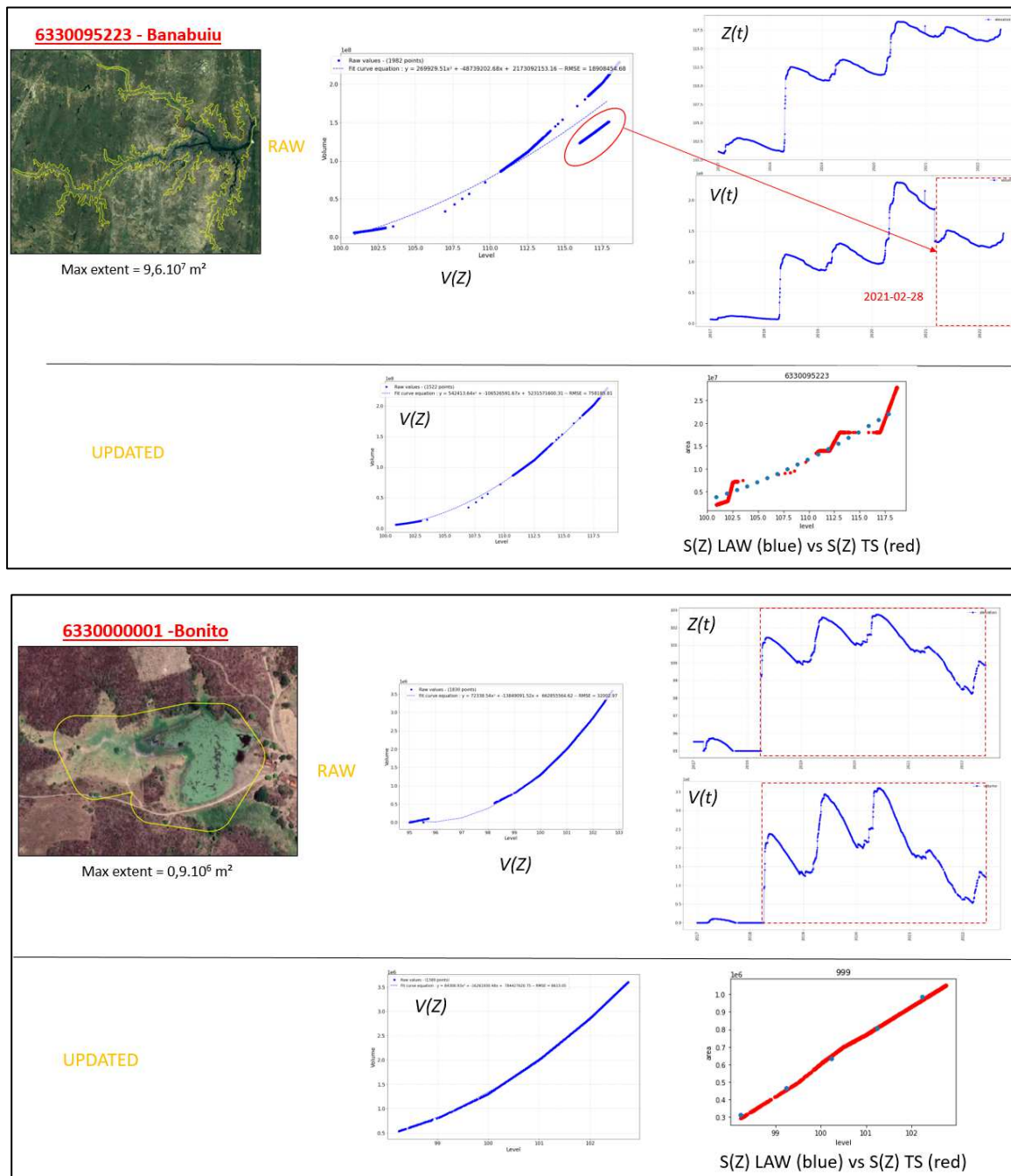


Figure 81 : Représentation des Lois $v(Z)$ et $S(Z)$ pour les données in situ corrigées (cas des données dupliquées)

Malgré les modifications apportées aux données de hauteurs d'eau et de volumes pour obtenir des relation $V(Z)$ cohérentes, on constate des incohérences qui subsistent dans la relation $S(Z)$. Ces incohérences sont probablement dues à des problèmes lors de la génération des surfaces dites *in situ*.

En effet, contrairement à la hauteur d'eau qui est une mesure *in situ* plutôt fiable issue de stations et le volume qui est issu d'un modèle prenant probablement en compte la bathymétrie des réservoirs, la surface *in situ* est, quant à elle une donnée théorique issues de lois ne prenant pas en compte la morphologie propre du réservoir.

Ces observations seront à prendre en compte lors de l'analyse des séries temporelles en sortie de STOCKWATER.

4.2.3 Analyse séries temporelles $S(t)$, $V(t)$, $Tr(t)$

Comme décrit lors de l'analyse effectuée lors de la phase I, une analyse des séries temporelles issues des fenêtres glissantes Optique et Radar a été effectuée. En effet, ces types d'extraction de surfaces se sont avérés les plus pertinentes et les plus robustes dans le cadre de cette étude.

Des tableaux avec des structures identiques à ceux présentés lors de la phase I ont été générés à partir des notebook mis en place.

4.2.3.1 Brésil

Tableau 23 : MEDIANE - Brésil

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
6320007102	Sao Jose III	nan	18.810285	22.489237	14.683855	38.154769	4.402261	18.292245
6320007172	Barragem do Batalhao	nan	28.504105	33.115264	73.423341	66.046339	46.176698	43.688167
6320007492	Jaburu I	16.779143	9.249037	4.677136	9.575751	4.674653	2.719546	8.599022
6320008892	Carnaubal	33.525627	24.993743	13.290283	78.574539	81.438330	58.218293	61.064527
6320009412	Jaburu II	nan	90.416946	96.352461	43.063579	35.727004	16.312191	15.331991
6320015192	Realejo	27.101354	17.890491	21.002019	9.971300	34.859547	6.591743	17.665416
6330000001	Bonito	nan	88.582267	95.440060	99.958559	99.998962	100.000000	100.000000
6330068463	Serrote	21.332547	11.802185	17.077881	28.457980	33.695171	11.067083	14.664458
6330068933	Araras	85.107299	7.425078	14.379832	89.090553	89.889159	26.471635	25.756121
6330070272	Farias de Sousa	nan	34.161818	31.684221	57.259142	48.901004	35.276432	27.648184
6330095223	Arrojado Lisboa	30.563766	73.259262	33.765447	99.952994	73.117328	99.675989	56.337778
6330098332	Acarape do Meio	nan	6.393449	17.368976	29.287535	42.425615	13.820472	20.849680
6330098612	Pacoti	15.995939	13.863894	23.613350	13.295391	15.563633	5.176296	12.096005
6330098593	Castanhao	nan	nan	nan	nan	nan	15.020484	21.025105

Tableau 24 : Résumé résultats médiane - Nombre de réservoirs par classe - Brésil

	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

4.2.3.2 Analyse des résultats

A noter : Le réservoir Bonito a été écarté de l'analyse, car les données in-situ identifiées pour ce réservoir ne correspondaient pas à la morphologie du réservoirs (surface * 10), il en est de même pour le réservoir Castanhao pour lequel le process n'a pas fonctionné.

Si on se fie aux tableaux précédents (cf. Tableau 23 et Tableau 24) qui illustrent les potentielles anomalies dans les résultats issues de DEM4WATER et SURWATER post-process, on constate :

- ✓ Que pour la surface, seulement la moitié des réservoirs ont des écarts médians avec les données *in situ* inférieures à 25%,
- ✓ Seulement deux réservoirs ont des écarts médians inférieurs à 25% pour les volumes,
- ✓ Et que 9 réservoirs sur 14 ont des taux de remplissage avec un écart médian de moins de 25% avec les données *in situ*.

Malheureusement, nous ne pouvons pas nous fier aveuglement à ce tableau, car, comme nous l'avons signalé dans la partie précédente (cf. 4.2.2.2.2.) on remarque des incohérences dans le comportement

de la relation $S(Z)$ des données *in situ*. Ces incohérences nous interrogent sur la qualité des Surfaces dites *in situ*.

C'est par exemple le cas pour les réservoirs 6330068933 - Araras, 6320008892 – Carnaubal et 6320009412 - Jaburu II pour lesquelles, on remarque que les surface dites *in situ* dépassent parfois le polygone de comptages de comptages, considéré comme le max-extent du réservoir.

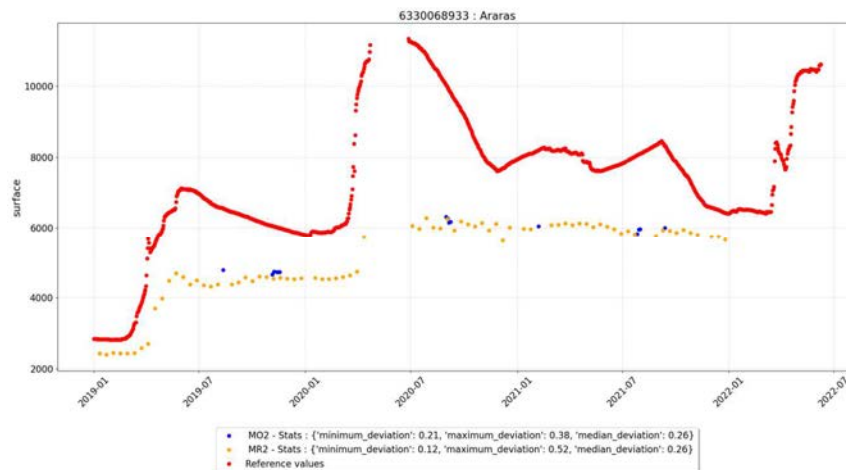


Figure 82 : Séries temporelles des surfaces MO2, MR2 et INSITU - Arraras

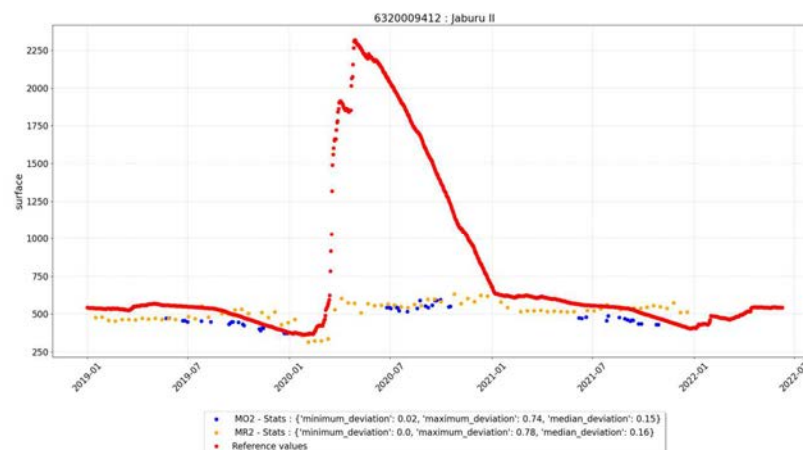


Figure 83 : Séries temporelles des surfaces MO2, MR2 et INSITU - Jaburu II

De plus, on remarque d'autres incohérences dans la dynamique des séries temporelles des surfaces. C'est le cas pour les réservoirs 6330068463 - Edson Queiroz, 6330068933 – Araras, 6330098332 - Acarape do Meio, 6320008892 – Carnaubal, 6320009412 - Jaburu II, 6330070272 - Farias de Sousa et 6330095223 – Banabuiui. En effet, on remarque une certaine cohérence dans les relations $V(Z)$ de ces réservoirs, mais des incohérences dans leurs relations $S(Z)$. Ce qui laisse à penser une mauvaise estimation de leurs surfaces dites *in situ*. Il est donc difficile d'évaluer la qualité des surfaces issues de Surfwater post-process pour ces réservoirs, ne pouvant nous fier à leurs surfaces dites *in situ*. Une comparaison avec des extractions issues d'une autres chaîne d'extraction de surfaces, type ExtractEO aurait été plus adaptée pour la validation de ces surfaces SURFWATER.

En ce qui concerne l'estimations de la qualité des volumes, on remarque globalement des écarts très importants avec les données *in situ*. Ces écarts peuvent, être en partie dus à des données *in situ* de mauvaise qualité, mais, comme présenté dans la Phase I le volume absolu a tendance à être sous-estimé par DEM4WATER, par rapport aux données de Volume *in situ*. Il est plus adapté de se fonder sur le taux de remplissage, pour lequel on observe des écarts corrects entre les taux de remplissage issus de SURFWATER post_process et les taux de remplissage issus de des données de références.

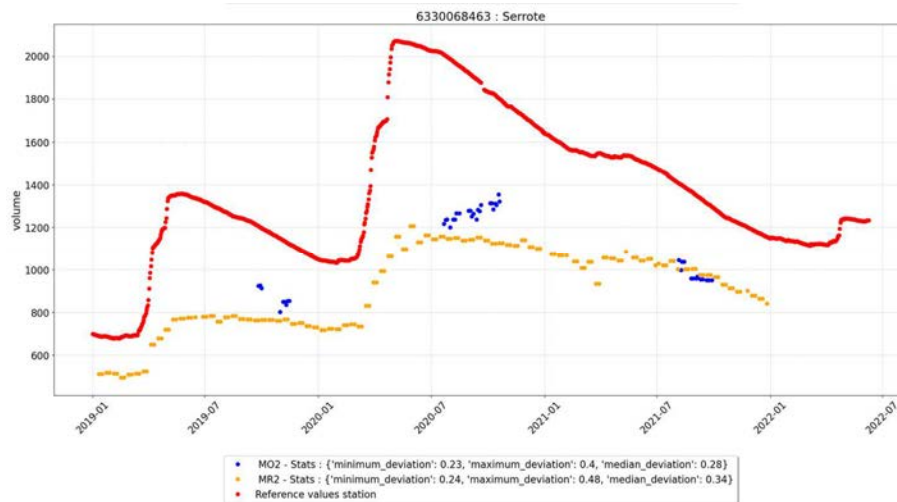


Figure 84 : SERIES TEMPORELLES DE VOLUMES M02, MR2 ET INSITU – Serrote

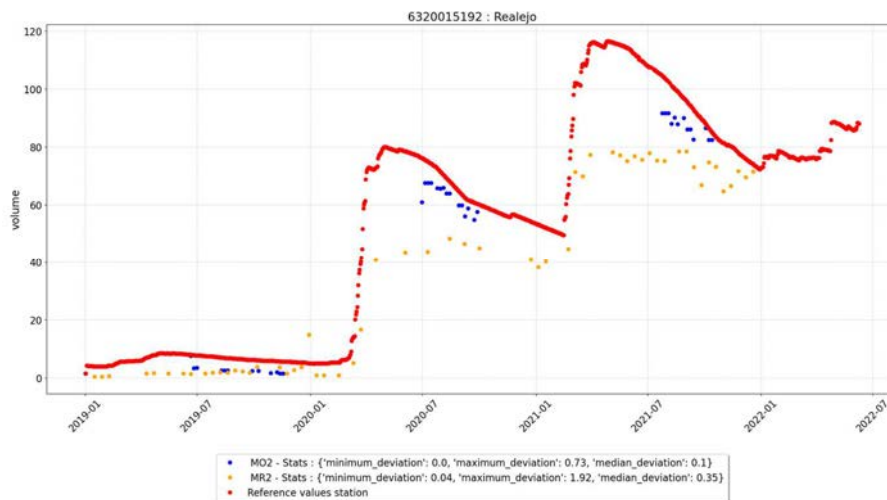


Figure 85 : SERIES TEMPORELLES DE VOLUMES M02, MR2 ET INSITU - REALEJO

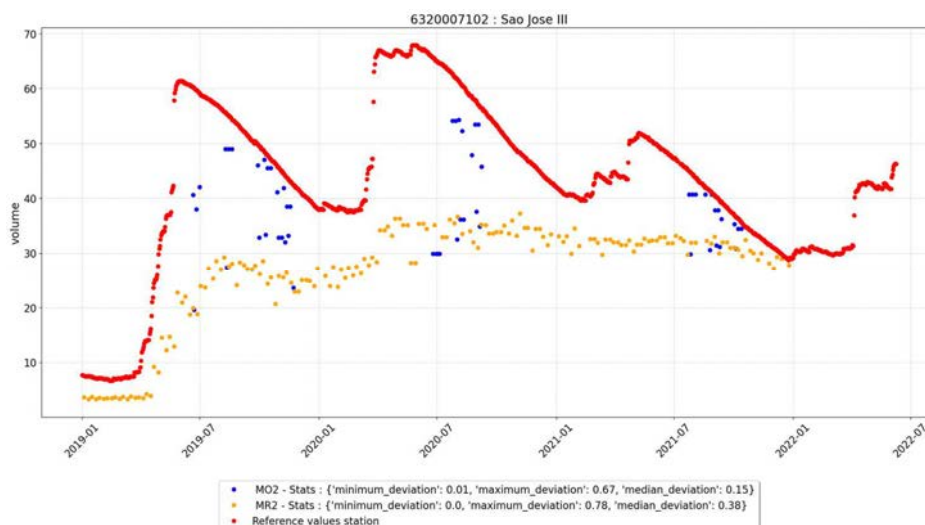


Figure 86 : SERIES TEMPORELLES DE VOLUMES M02, MR2 ET INSITU – SAO JOSE III

Comme évoque lors de la Phase I, il est tout même nécessaire de considérer ces résultats avec précautions car les taux de remplissage SURFWATER post_process et de référence se fondent sur leurs volumes maximums respectifs. Ce sont donc des données relatives à un volume maximum de remplissage qui doit être observé dans la fenêtre d'observation (normalement 3 ans suffisent pour observer ces périodes de remplissage).

4.3 Production

4.3.1 Génération des masques d'eau

Les campagnes de production des masques de Surface en eau pour les super sites Andalousie, Occitanie, Inde, Tunisie, Burkina Faso et Laos ont été réalisées sur le cluster HAL du CNES pour la période s'étalant de 2018 à 2020, et pour le super site Brésil de 2019 à 2021.

Les tuiles suivantes ont été traitées :

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

Les campagnes de production des masques Surfwater ont mis en évidence des limitations au niveau de l'extraction des masques d'eau optiques qu'il a été possible d'améliorer en utilisant la méthode Random Forest à la place de la méthode Clustering Average initialement sélectionnée. Les campagnes du mode optique des 3 super sites de la Phase I, Andalousie, Occitanie, Inde, ont été régénérés avec ce nouveau paramétrage afin d'améliorer qualitativement la classification. Puis les nouveaux jeux de données optiques ont été fusionnés avec ceux radars produits lors de la Phase I.

Les campagnes Surfwater des nouveaux super sites Tunisie, Burkina Faso, Laos ont été réalisées sur la période 2018-2020 et sur la période 2019-2021 pour le Brésil.

Dans le cadre du prolongement de la Phase II, des modifications sur le paramétrage utilisé ont été apportées afin de supprimer les effets de sunglint observé sur les masques d'eau. A ce titre les sites Tunisie, Burkina Faso et Brésil ont été sélectionnés pour être retraités. Les jeux de données optique ont été régénérés sur la période des 3 années. Les occurrences ont été également recalculées sur chacun des sites. Ces nouveaux jeux de données optiques ont été mergés avec les précédents jeux de données radar.

Une tuile a posé un problème sur Burkina Faso ; la tuile T30PXU n'ayant pas de donnée Pekel sur cette zone, aucun produit Surfwater n'a pu être généré. Les données produites par la chaîne Waterdetect sur cette zone ont été fournies par le CNES. Ces masques d'eau ont été intégrés dans les 3 jeux de données existants 2018, 2019 et 2020, avec génération des cumuls mensuels, calcul des occurrences sur 3 ans et régénération du wmap optique global.

4.3.2 Calcul de relations HSV et comparaison

Cette étape consiste à générer/améliorer les modèles HSV estimés avec le MNT Copernicus 30m.

Comme il a été décrit lors de la Phase I, les limitations méthodologiques de la chaîne DEM4water ont été atteintes sur certains réservoirs, avec par exemple une mauvaise estimation de pied de barrage, la ligne de coupure partiellement en boucle, du water crossing par la ligne de coupure, cas de multi-déversoirs avec plusieurs ouvrages sur une même retenue d'eau, etc.

Ces anomalies sont répertoriées par site sur les liens confluence ci-dessous :

- ✓ <https://confluence.cnes.fr/display/SCOSW/Validation+lois+zsv+-+dem4water+Final>
- ✓ <https://confluence.cnes.fr/display/SCOSW/20220302+CNES+-+DEM4Water+issues+table>

Il a été mis en place la possibilité d'utiliser des fichiers renseignés manuellement afin de corriger certaines de ces données erronées que ce soit au niveau de la localisation du pied de barrages ou de la ligne de coupure.

Des modèles initialement en erreur ont pu être produits, tandis que d'autres qualitativement améliorés sur les sites Andalousie, Occitanie et Inde.

Le barrage Nam Theun 2 du super site Laos a nécessité une augmentation des ressources du fait de la superficie hors norme de ce réservoir, cependant le résultat du modèle de ce réservoir atypique n'est pas concluant.

Le réservoir Castanhao du Brésil a posé des problèmes similaires en raison de sa taille gigantesque.

TABLEAU 25 : NOMBRE DE MODELES GENERES EN FONCTION DES RESERVOIRS ETUDIES, PAR SUPER SITE

Super site	Réservoirs	Modèles
Andalousie	37	37
Occitanie	16	16
Inde	23	16
Brésil	14	13
Tunisie	10	9
Burkina Faso	9	9
Laos	1	0

Les modèles suivants sont considérés comme correctement générés :

- ✓ **Andalousie** : Arcos de la Frontera, Beznar, Bornos, Celemin, Charco Redondo, Cuevas de Almanzora, Guadalquivir 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.
- ✓ **Occitanie** : Agly, Astarac, Avene, Cammazes, La Gimone, Montbel, Puyvalador.
- ✓ **Inde** : Dindi, Himayat Sagar, Kaddam, Musi, Nagarjuna, Pakhal, Palair, Pocharam, Sathnala.
- ✓ **Brésil** : Acarape do Meio, Araras, Arrojado Lisboa, Barragem do Balahao, Carnaubal, Farias de Sousa, Jaburu I, Jaburu II, Realejo, Serrote.
- ✓ **Tunisie** : Chiba, El hamma, Er rmal, Kamech, Masri, Sidi Saad, Sidi Saleem.
- ✓ **Burkina Faso** : Bagre, Kompiega, Lac Bam, Nagbangre.

Ne bénéficiant pas de toutes les améliorations logicielles apportées au cours de Phase II, certains réservoirs de la Phase I (Andalousie, Occitanie, Inde) nécessitent de plus amples corrections manuelles.

En ce qui concerne les sites du Burkina Faso, les réservoirs Namsuguia, Samendeni, Tanvi 1 et Tanvi 2, dont les modèles sont listés ci-dessous, ont la particularité d'être atypiques par leur faible profondeur avec de ce fait des erreurs fort probables dans les modélisations.

Le réservoir Ziga on a un problème de déversoir avec un saut à la fin, ce qui va nécessiter une correction

de la cutline.

En Tunisie, le réservoir Lebna a une erreur de localisation avec le point situé dans l'eau.

Les réservoirs de très grande taille comme Castanhao au Brésil ou Nam Theun 2 au Laos nécessiteront le développement d'une approche spécifique.

Sites dont les modèles sont à améliorer :

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

4.3.3 Génération des séries $S(t)$, $V(t)$ et $TR(t)$

La chaine Surfwater_Postprocess a fait l'objet d'évolutions avec la mise en place d'un mode de reconstruction d'une part, et d'autre part l'amélioration de la vitesse de traitement des données.

De plus, de nouveaux modèles, précédemment en erreur ou en cours d'amélioration lors de la Phase I, ont été générés au cours de la Phase II.

Les séries temporelles estimées des sites concernés de la Phase I ont été régénérés sans reconstruction pour les méthodes de comptage MO1/MR1/MO2/MR2 sur la période 2018-2020.

De plus, ces jeux de données des séries temporelles ont été produits 2 autres fois en parallèle avec différents modes de reconstructions : à 1% puis à 2% pour tous les barrages des super sites Andalousie et Occitanie.

Les séries temporelles des super sites Tunisie et Burkina Faso ont été générés 2 fois chacun, sans reconstruction et avec reconstruction à 5%.

Au cours du prolongement de la Phase II, les séries temporelles optiques des super sites Tunisie et Burkina Faso ont été régénérés sans reconstruction et avec reconstruction 2%.

Les séries temporelles optiques et radars du Brésil ont été régénérés sans reconstruction et avec reconstruction 2%.

Les sites étudiés sur le Brésil étant sujets à fort ennuagement, les séries temporelles ont été à nouveau produites avec le paramètre nodata à 0.9 (et reconstruction 2%).

En dernière étape, ces séries temporelles ont été préalablement filtrées par un outil développé dans le cadre du projet Stockwater avant d'être ingérées dans le site de valorisation. Les données avec du nodata y sont supprimées, ce qui donne une fréquence de date moins grande mais beaucoup plus pertinente au niveau des tendances.

Quant aux séries temporelles *in situ* de référence, elles sont retraitées afin de fournir un fichier cvs par barrage sur une période définie.

4.4 Analyse/Validation des résultats de la phase complémentaire

4.4.1 Volume absolu

Les résultats ne sont pas satisfaisants. Les médianes de toutes les erreurs relatives d'estimation sont

de 15% pour l'optique MO2 et 22% pour le radar MR2. Regardant la moyenne de la médiane d'erreur relative de volume absolu par réservoir, on retrouve 29% pour MO2 et 33% pour MR2. Sur les 34 réservoirs andalous, 14 présentent des volumes absolus estimés avec un écart de moins de 10% avec la référence, 20 présentant un écart supérieur à 25% dont 4 à plus de 50%, pour le MO2. Dans le cas de l'Occitanie, sur les 16 réservoirs, 2 présentent un écart de moins de 10%, 6 un écart entre 10 et 25%, le reste présente un écart supérieur à 25%. Pour le Super Site Inde, finalement le processus n'a pu aboutir que pour 3 réservoirs sur les 6 initiaux, et les écarts sont toujours supérieurs à 25%, avec des chiffres tels qu'ils indiquent un problème à certains niveaux (données *in situ*, etc.). En ce qui concerne le Brésil, uniquement 2 réservoirs ont des écarts inférieurs à 10%, 2 entre 10 et 25% et les 9 restants accusent des écarts de plus de 25% avec la référence. Si on se fonde sur des métriques plus restrictives, comme les quantiles 75 et 90, on constate une très forte augmentation des écarts.

Plusieurs sources d'erreurs ont été identifiées, liées à une mauvaise estimation de la loi $V(S)$ dans certains cas où il y a des écarts importants entre les surfaces Surfwater (en optique sur le paramétrage 1, très erroné, ou les problèmes de détection biaisés par radar [Peña-Luque S. et al., 2021]) et les surfaces de références (données *in situ* de qualité questionable sur certains barrages). À noter qu'il arrive parfois que ces erreurs se compensent. De mauvais résultats pour l'estimation de la loi $V(S)$ et l'extraction des surfaces par Surfwater, peuvent donc engendrer des volumes qui semblent corrects.

4.4.2 Taux de remplissage

Les résultats sont positifs. Sur la totalité de mesures, les estimations de $TR(t)$ obtiennent un quantile 75% de mesures de **12.8%** pour les mesures optiques (MO2) et **17.3%** de mesures radar (MR2). Ces résultats ont un niveau de qualité similaire au suivi de taux de remplissage de l'étude précédente avec le CNES/MTE sur l'Occitanie, qui utilisait une estimation de lois $V(S)$ semi-manuelle à partir de MNT. Ceci indique que les tendances de croissance/décroissance du volume total stocké sont bien détectées par les séries temporelles $S(t)$ par Surfwater, et que les erreurs commises sur l'estimation $V(S)$ ne perturbent pas trop le suivi de tendance $TR(t)$.

Cet indicateur reste intéressant pour les gestionnaires d'eau, qui connaissent en général les volumes totaux mobilisables. En revanche, ceci ne résout pas le besoin de connaître le volume réel mobilisé sur des bassins pas instrumentés ou pas enregistrés sur les bases de données de barrages (GranD, GeoDAR, etc).

4.5 Valorisation

Dans l'objectif de la restitution de ces résultats, un portail web a été mis en place permettant la consultation des séries temporelles des barrages étudiés.

Plusieurs propositions ont été faites en matière d'hébergement du site de valorisation. Le nom de domaine suivant a été choisi : sco-stockwater.org

Le code est centralisé sur un compte Github publique de CS:

https://github.com/CS-SI/sco_stockwater

Description des fonctionnalités

Le site de valorisation est carto-centré, permettant de visualiser les données d'un ou plusieurs réservoirs simultanément.

Il est constitué de 3 parties :

- ✓ la partie « en-tête » avec le logo du SCO, du nom du site et d'une icône qui permet de changer de thème ;
- ✓ la partie « action » qui elle-même est divisée en deux parties, celle supérieure qui montre les réservoirs sélectionnés et celle inférieure qui permet d'interagir et d'actualiser les recherches;
- ✓ la partie « majeure », étant la carte.

A titre d'illustration, quelques copies d'écran sont présentées ci-dessous afin de décrire les principales fonctionnalités du tableau de bord.



Figure 87 : détails de la page d'accueil.

En cliquant sur un réservoir, la partie majeure est divisée en deux, la carte est affichée sur la partie supérieure et le graphique sur la partie inférieure.

Trois types d'observation sont disponibles : optique, radar et référence.

La partie supérieure gauche, encadrée en rouge sur la figure ci-dessous, indique les réservoirs sélectionnés sur la carte. Celui-ci est alors affiché dans une autre couleur (bleu clair) afin de se distinguer des autres réservoirs présents sur la carte.

On peut voir sur la figure les séries temporelles du taux de remplissage du réservoir Guadalteba en

Andalousie, sur la période de 3 ans allant de 2018 à 2020.

Les données estimées des modes radar et optique sont affichées avec la référence *in situ*.

En termes de produit, la partie « Observation depth », permet de sélectionner au choix :

- ✓ « 1 day » : l'observation du jour correspondant au produit Surfwater single,
- ✓ « 10 days » : la synthèse sur 10 jours correspondant au produit Surfwater multiple.

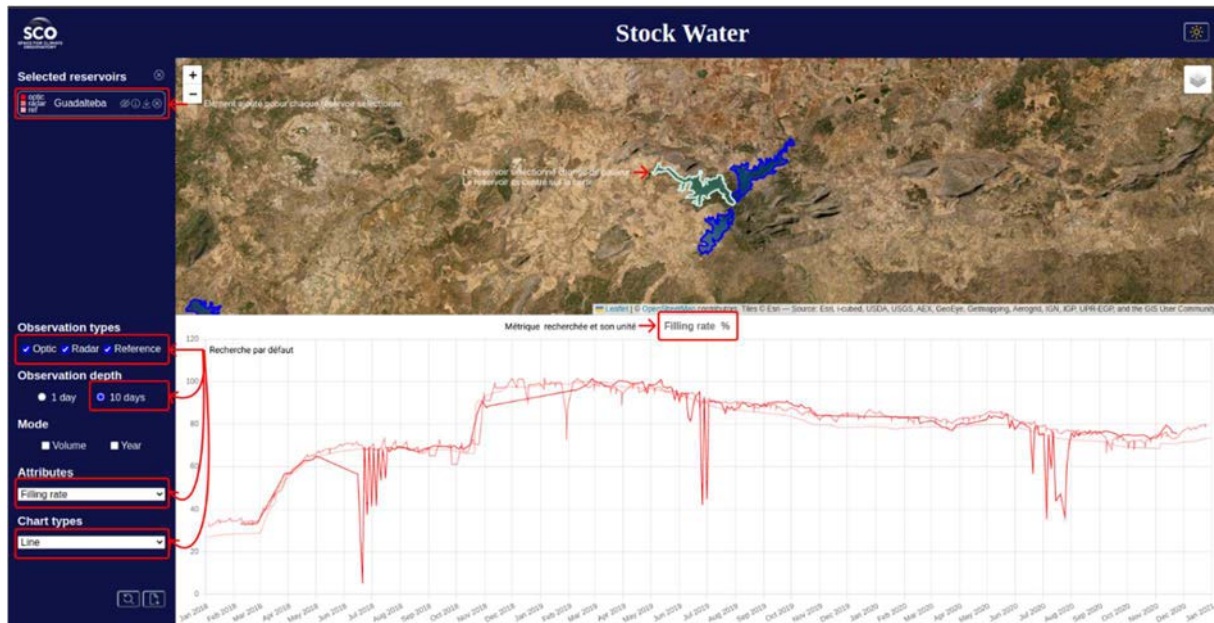


Figure 88 : données affichées après sélection sur un réservoir.

Pour un point sélectionné sur la courbe, on dispose des informations suivantes concernant cette date en particulier : date, nom du réservoir, type d'observation, volume (en hm3).

Concernant les métriques à visualiser un menu déroulant « Attributes » permet la sélection de la surface, du volume ou du taux de remplissage.

Il est possible également d'avoir accès à des informations générales telles que l'identifiant du barrage, le type d'utilisation, la ville à proximité, les coordonnées du barrage.

Les données de séries temporelles sont téléchargeables sous forme de fichier au format csv (icone « c » à côté du nom du barrage sur l'image ci-dessous).

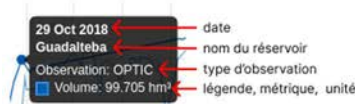


FIGURE 89 : fonctionnalités de sélection des items. Partie droite : détails de l'élément à l'ajout d'un réservoir. Partie gauche : anatomie de la tooltip du graphique et de la sélection des différentes métriques.

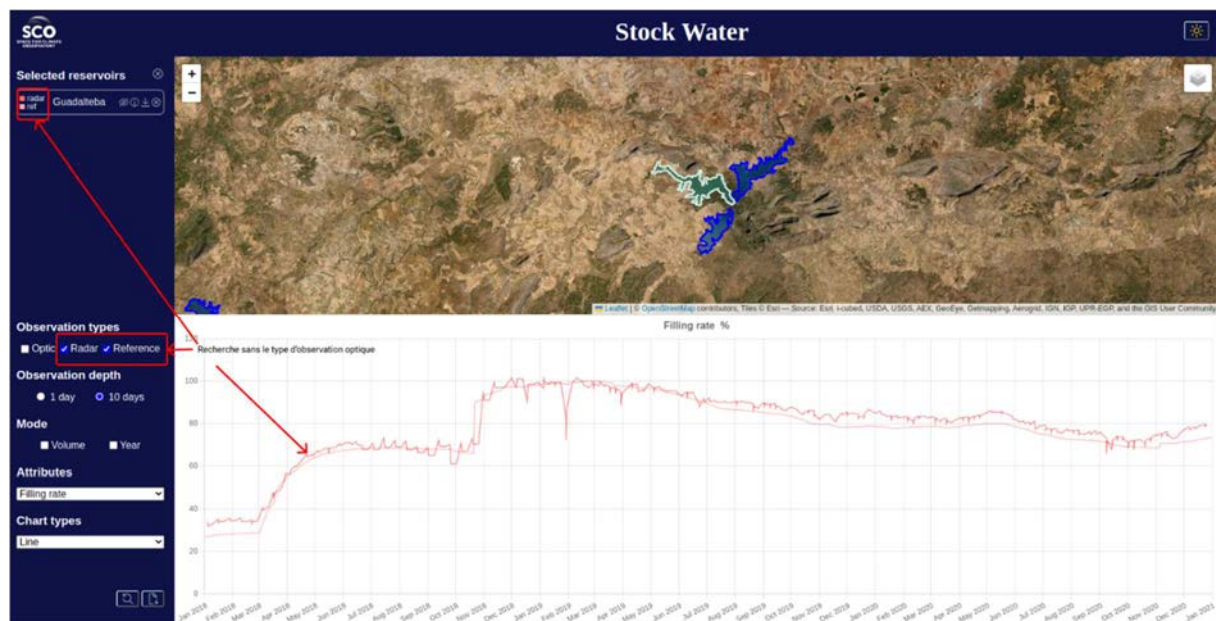
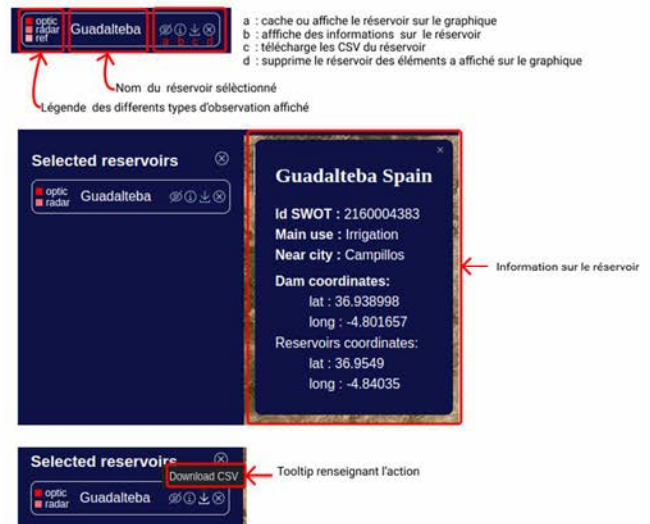
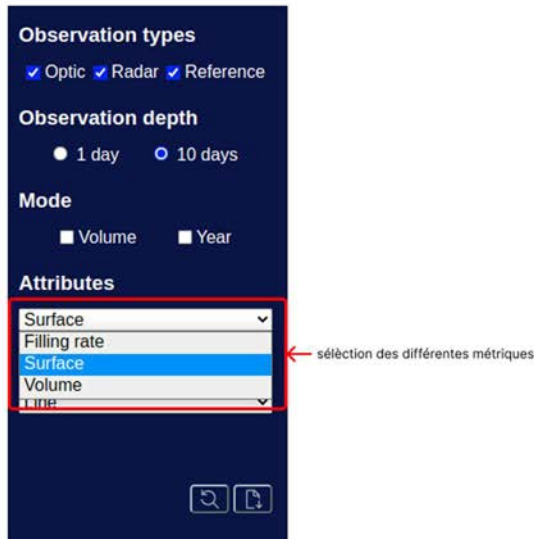


Figure 90 : exemple d'une recherche en décochant un type d'observation.

Il est possible de visualiser plusieurs barrages à la fois, par exemple sur la figure ci-dessous 3 barrages sont présentés : Guadalteba (en rouge), Guadalhorce (en vert) et Zahara (en bleu).

Les données des séries temporelles du taux de remplissage sont en mode optique et radar pour chaque barrage, sur une période de 3 ans correspondant à l'étude, de 2018 à 2020.

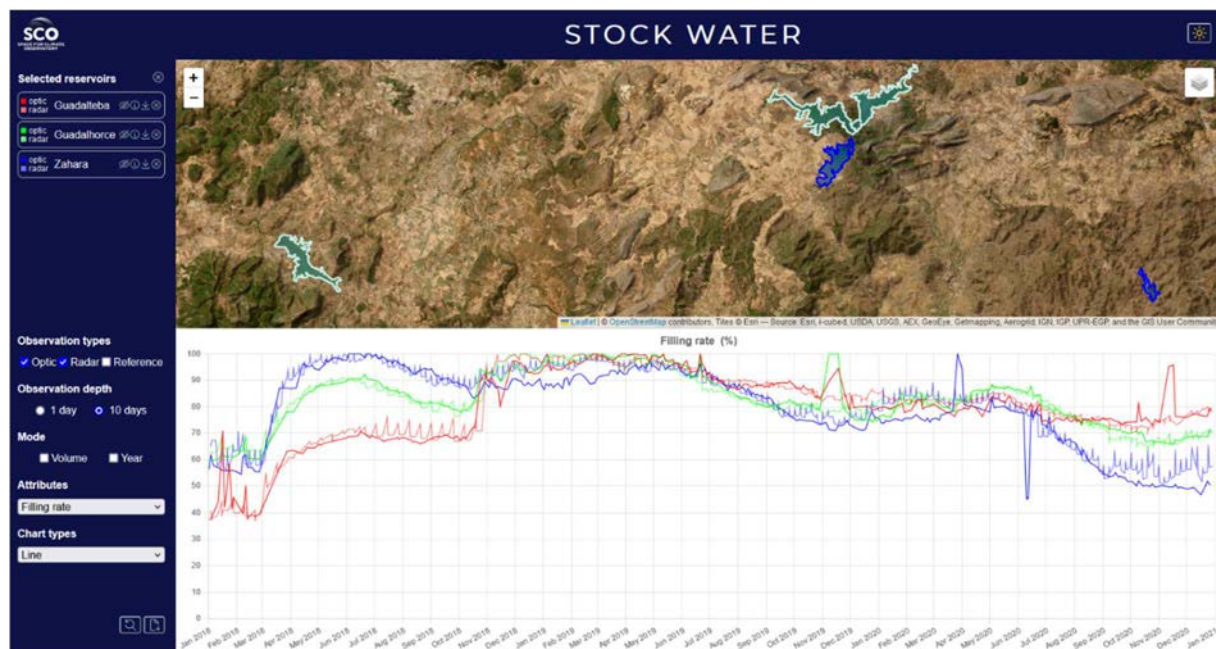


Figure 91 : exemple d'une recherche sur plusieurs réservoirs.

Il est possible de zoomer sur une période en particulier, les jours sont alors affichés sur l'échelle à la place initialement des mois.

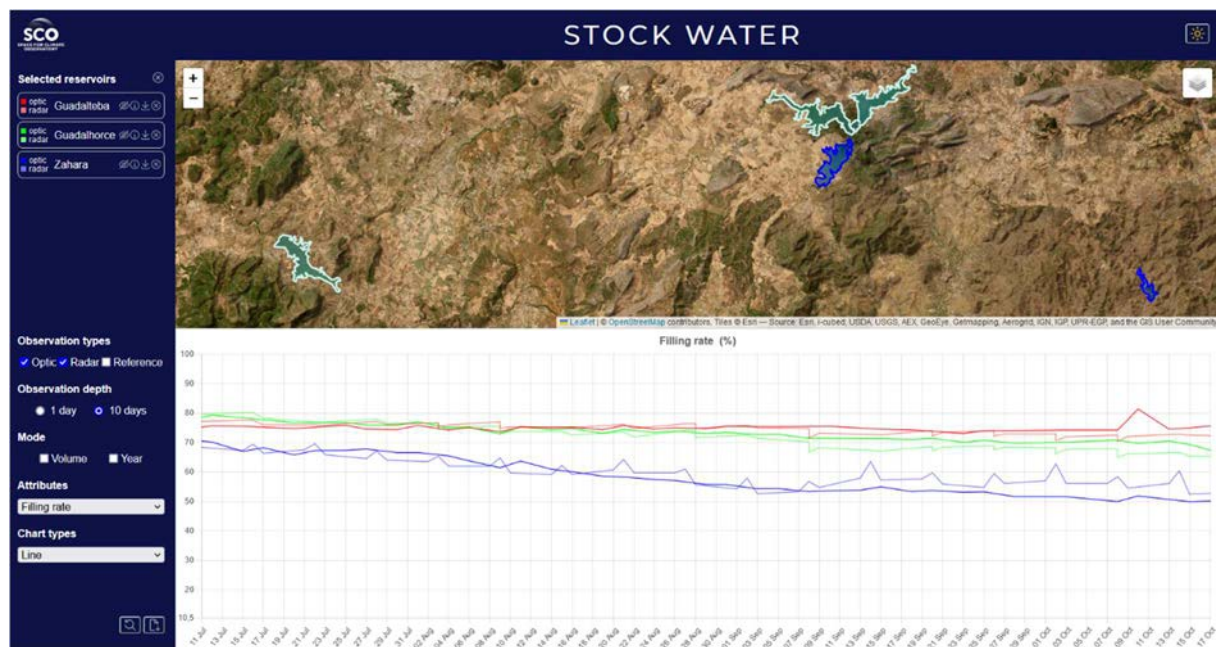


Figure 92 : exemple d'une partie zoomée du graphique.

Parmi les fonctionnalités, le mode « Volume » et le mode « Year » permettent une visualisation spécifique des données.

Le mode « Volume » calcule la somme des volumes des barrages sélectionnés. Le tooltip affiche, pour le jour cliqué sur la courbe, le total des volumes en hm3 ainsi que la valeur des volumes individuels de

chaque barrage, comme présenté sur la figure suivante.

La série temporelle de volume in-situ de référence est affichée en bleu clair, la série temporelles de volume optique en bleu foncé, puis celle du radar.

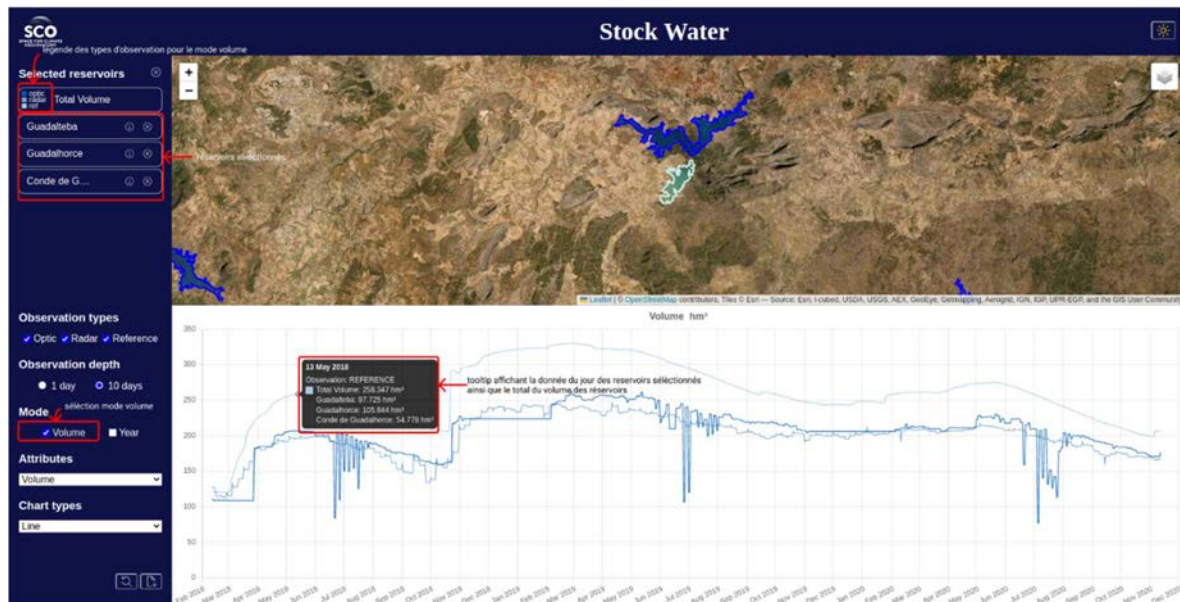


Figure 93 : détails du mode « Volume ».

Le mode « Year » permet de comparer sur une échelle de 12 mois (de janvier à décembre), les données d'un même réservoir sur plusieurs années. Ce mode de visualisation permet de faciliter la comparaison de l'évolution interannuelle d'un réservoir.

Voici par exemple les séries temporelles du taux de remplissage du réservoir José Toran en Andalousie pour les années 2018 (en rouge), 2019 (en orange) et 2020 (en bleu).

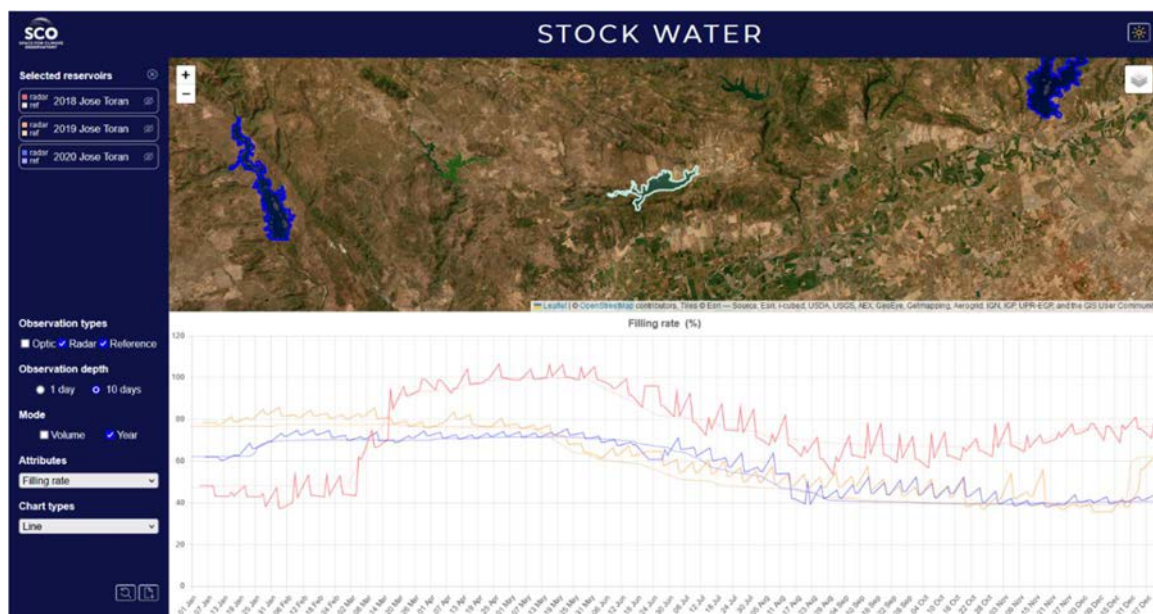


Figure 94 : détails du mode « Year ».

Pour information, il existe la possibilité de visualiser les données via le menu déroulant « Chart types » soit selon le style “Line”, comme il a été présenté sur les exemples précédents, soit avec des points selon le style “Scatter”.

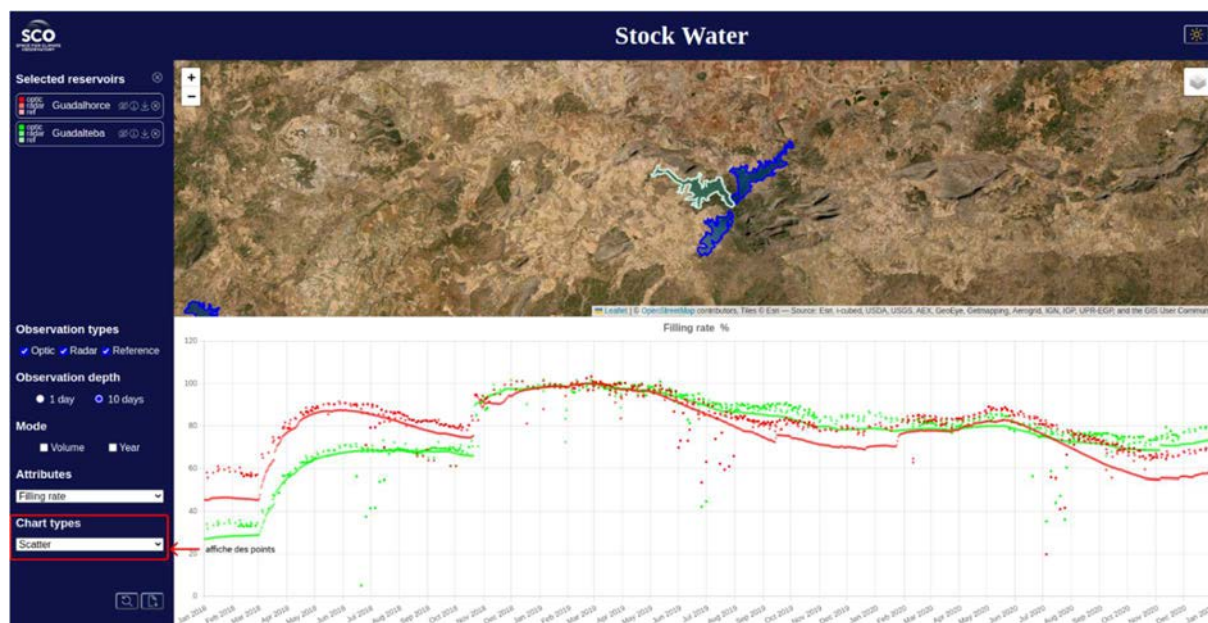


Figure 95 : exemple d’un affichage des données avec le style « Scatter ».

5 Robustification du démonstrateur et du tableau de bord

En prolongement des activités des phases I et II, il a été décidé de prolonger les activités dans le but d'améliorer l'opérationnalité de l'outil. Le périmètre des activités s'est établi comme suit :

- ✓ Finalisation des activités de sélection, validation et préparation des données,
- ✓ Activités de sélection et de préparation des données selon le nouveau périmètre défini par le présent avenant (CS GROUP – France),
- ✓ Intégration de nouvelles corrections de logiciels « Surfwater », « Surfwater-Postprocess » et « DEM4water » (CS GROUP -France),
- ✓ Génération de nouvelles séries temporelles sur l'ensembles des sites du périmètre défini dans le cadre du présent avenant (CS GROUP -France),
- ✓ Analyse et validation des résultats (SERTIT),
- ✓ Restitution aux utilisateurs (CS GROUP – France et SERTIT).

5.1 Améliorations de la chaine

Dans le but d'améliorer la qualité des résultats produits lors des Phases I et II sur différents aspects, des évolutions ont été réalisées sur les chaines Surfwater, DEM4water et Surfwater_Postprocess. Une nouvelle phase de production a été effectuée sur la base de ces versions logicielles.

5.1.1 Surfwater

Dans la même perspective que lors des précédentes phases, les productions ont été de nouveau générées enrichie des évolutions précédemment citées, avec également la correction des bordures nodata et la mise à jour des données d'occurrences Pekel.

5.1.2 DEM4water

La nouvelle version de la chaine DEM4water intègre les évolutions suivantes :

- ✓ conservation des données de lignes de contour si elles déjà sont présentes afin de gagner du temps de traitement,
- ✓ modification de la gestion de l'élévation de pied de barrage dans l'objectif de s'adapter au mode de correction manuel,
- ✓ possibilité de lancer le calcul des $S(Z_i)$ à partir des lignes de contour avec des informations définies manuellement,
- ✓ mise en place d'un rapport de performance.

5.1.3 Surfwater_Postprocess

La chaine Surfwater_Postprocess dispose d'un nouveau mode de reconstruction. Son utilisation et le paramétrage associé ont été choisis en accord avec le CNES. En parallèle, d'autres évolutions sont prises en compte avec une uniformisation de l'utilisation du nom du réservoir et des actions concernant le filtrage des séries temporelles.

5.1.4 Orchestrateur

L'orchestrateur de Stockwater a été mis à jour en fonction des nouveaux appels en ce qui concerne les chaines DEM4water et Surfwater_Postprocess, dans leurs versions modifiées.

5.2 Modification de la base de données

Lors de la phase préliminaire, une cinquantaine de réservoirs ont été sélectionnés de façon à être le plus hétérogènes et représentatifs possible en termes de surfaces, de configuration topographique, etc.... et pour lesquels les données d'entrée, sont déjà disponibles (disponibilité de plusieurs MNT, paramètres de validation...). Une base de données de référence décrivant ces sites a été mise en place lors de la Phase I.

Lors d'une seconde phase de sélection, 50 réservoirs supplémentaires ont été identifiés et intégrés à la base de données de référence. Au terme de la sélection des sites tests, une base de données finalisée est disponible. Le format de cette base de données géoréférencées et compatible avec le format utilisé dans DEM4water pour décrire les barrages, ainsi que les modèles (estimés et de référence) associés. Cela assure une homogénéité tout au long de la méthode, ainsi qu'une exploitation efficace en phase de comparaison et de validation des modèles.

5.3 Génération de nouvelles séries temporelles

5.3.1 Finalisation des calculs de lois ZSV

A partir des résultats obtenus lors de la Phase II, certains modèles ont été améliorés au travers de la prise en compte de fichiers correctifs fournis par le CNES, concernant les lignes de coupure ou/et autres informations. Il a été notamment étudié la possibilité d'améliorer la méthode par le calcul de bassin versant pour limiter la surface des courbes de niveau. Les sites concernés couvrent les pays suivants :

- ✓ Burkina Faso (9 barrages),
- ✓ Tunisie (10 barrages),
- ✓ Brésil (14 barrages).

5.3.2 Relance des calculs SURFWATER

La re-génération des masques d'eau Surfwater pour le mode optique a été réalisée sur le super-site Brésil avec le paramétrage suivant :

- ✓ Sélection de 14 barrages sur les 67 disponibles avec des données *in situ*,
- ✓ Désactivation du filtre de détection de la neige pour les 7 tuiles,
- ✓ Période 2019-2021.

La nouvelle base de données optique sera mergée avec la précédente base de données radar de la Phase II. Les occurrences optiques ont été régénérées :

- ✓ Annuellement sur chacune des 7 tuiles, pour les années 2019, 2020 et 2021,
- ✓ Sur la période 2019-2021.

5.3.3 Relance des calculs SURFWATER_Postprocess

La production de la chaîne Surfwater-Postprocess a été réalisée avec les améliorations de reconstruction par carte d'occurrence optique. La régénération des séries temporelles de surface $S(t)$, volume $V(t)$ et taux de remplissage $Tr(t)$ porte sur les super-sites :

- ✓ Andalousie,
- ✓ Burkina Faso,
- ✓ Tunisie,
- ✓ Brésil.

Les séries temporelles des 4 super sites (Andalousie, Burkina Faso, Tunisie, Brésil) ont été générés pour l'optique :

- ✓ En mode « avec reconstruction » pour une valeur de 2%,
- ✓ En mode « sans reconstruction ».

Les données « sans reconstruction » n'ont pas fait l'objet d'une évaluation. Les séries temporelles n'ont pas été régénérées pour les données radar. Les super sites Occitanie, Inde, Laos ayant déjà le jeu de données « sans reconstruction » généré, ils n'ont fait l'objet d'aucun retraitement.

5.3.4 Formatage des nouvelles séries temporelles

La tâche consiste à filtrer et formater les jeux de données optiques « avec reconstruction » 2% des nouvelles séries temporelles $S(t)$, $V(t)$ et $Tr(t)$, ceci dans un objectif d'intégration au tableau de bord STOCKWATER, pour les super sites :

- ✓ Andalousie,
- ✓ Burkina Faso,
- ✓ Tunisie,
- ✓ Brésil.

Les nouveaux jeux de données de séries temporelles ont été archivés et livrés.

5.4 Valorisation

La version initiale du tableau de bord développée lors de la Phase II a été modifiée et robustifiée en fonction de l'évolution de produits et des besoins exprimés par le CNES. Cette tâche inclue également la mise à jour finale des données avec leur mise au format requis.

5.4.1 Robustification du tableau de bord

Dans un premier temps, un code review et une revue « besoin métier » ont été réalisés, dans le but de mettre en lumière les axes d'amélioration possibles. En fonction des éléments retournés par cette analyse, l'effort a porté sur les principaux composants en termes de **refactorisation du code, de stabilité et de performance**. Dans ce cadre, une meilleure gestion du state et des données a été mise en place afin de fluidifier les enchaînements et de corriger les bugs de transitions. En termes **d'interface et de navigabilité**, l'effort a porté pour l'essentiel sur :

- ✓ L'affichage des données d'environnement,
- ✓ L'ergonomie de l'affichage, notamment l'ergonomie cognitive et de pertinence métier en ce qui concerne les informations sur les produits,
- ✓ L'amélioration du mode volume/year,
- ✓ La navigation à partir de listes déroulantes,
- ✓ La correction des bugs et des conflits d'affichage des produits et des légendes,
- ✓ La correction des bugs de sélection/désélection des produits en mode « year »,
- ✓ L'amélioration de la navigation cartographique et de la lisibilité des informations sur fond cartographique.

Une analyse des performances et de la stabilité en termes de solutions d'hébergement a été également effectuée. Les différents budgets associés aux solutions seront évalués dans la perspective d'une solution opérationnelle. La version finale est identifiée v1.0.0.

5.4.2 Intégration des données et mise à jour finale

La tâche consiste à mettre à jour la base de données alimentant le tableau de bord STOCKWATER avec les nouveaux jeux de données des séries temporelles optiques livrées.

6 Conclusions et Perspectives

6.1 Synthèse de l'étude

Lors de la Phase I, la **méthodologie en mode end-to-end** a été mise en œuvre sur un nombre réduit de sites comprenant l'Andalousie, l'Inde et l'Occitanie sur la période 2018 à 2020. Cette méthodologie a reposé sur trois étapes nécessaires au suivi d'un site :

- ✓ Détermination de la surface en eau du site à chaque observation réalisée par les satellites Sentinel-1 et Sentinel-2,
- ✓ Modélisation de la relation HSV basée sur l'analyse des DEMs Copernicus30, SRTM30 et Alos,
- ✓ Génération des séries temporelles $S(t)$, $V(t)$ et $TR(t)$.

Pour mener à bien cette première phase, quatre tâches ont été réalisées en parallèle ou à la suite, selon les cas. Il s'agit :

- ✓ Du développement du démonstrateur,
- ✓ De la mise en place de la base de données,
- ✓ De la production,
- ✓ De l'analyse et validation des résultats.

Le **développement du démonstrateur** a également nécessité l'implémentation d'améliorations sur les chaînes Surfwater, DEM4water et Surfwater-Postprocess afin de s'adapter aux nouvelles données d'entrée et de proposer des fonctionnalités nécessaires au volume de production, à la variabilité des données, mais aussi à l'analyse de résultats.

La **mise en place de la base de données** comprend, d'une part la sélection des réservoirs étudiés, la définition puis le remplissage par région d'étude et par réservoir d'une table attributaire assez complète, la récupération, analyse et mise en forme des données *in situ* H-V, voire leurs corrections. Pour cette Phase I, il a été décidé de centrer l'étude sur trois Super Sites, l'Occitanie, du fait de l'expérience acquise par le CNES dans le cadre des travaux avec le Ministère de la Transition Écologique, de l'Andalousie et de l'Inde qui semblaient présenter une grande maturité en termes d'accessibilité aux données.

La **sélection des réservoirs** et la constitution de la base de données ont suivi les recommandations du CCTP et les éléments de réponse apportés dans la réponse technique. Elles se fondent sur la prise en compte et l'exploitation, des bases de données Lacs de SWOT, GOOD2, GRanD, GEODAR (BDs GRanD + GOOD), INDIAN Water Central Commission.

Une fois ces informations rassemblées, la table attributaire a été construite, il a fallu de nombreux échanges pour arriver à la version définitive en termes de type de fichiers, de format / ordre des champs, ... Ainsi, pour chaque Super site, au minimum 6 versions ont été produites.

Le travail a ensuite porté sur la récupération, la mise en forme et l'analyse des données *in situ*, qui devaient être mises à disposition via le CNES par les utilisateurs finaux. Il s'est avéré, dans un certain nombre de cas, que ces données devant servir de référence, étaient livrées avec des structures/formats différents, nécessitant des pré-traitements qu'il n'est pas toujours possible d'automatiser.

La dernière étape de la mise en forme des données a porté sur la génération des lois *in situ*. Pour cela, plusieurs évolutions de l'outil « timeseries_tool.py » du CNES ont été effectuées, que ce soit par le CNES ou le SERTIT. Si l'outil fonctionne aujourd'hui, des améliorations sont encore nécessaires afin d'assurer un bon fonctionnement dans une version opérationnelle.

La mise en place de la base de données s'est terminée avec l'**évaluation des MNTs**. Il s'agissait, au sein des MNTs identifiés comme inputs potentiels de qualifier un ou deux MNTs en tant que donnée de référence unique, afin d'optimiser par la suite les phases de production et d'analyse et de garantir un minimum d'homogénéité et d'interopérabilité des traitements. L'analyse des résultats a permis de classer les MNTs en termes de performances, de retenir de Copernicus 30 et de proposer de rejeter le MNT ALOS présentant des erreurs décimétriques.

La troisième tâche, est celle de la **production**, dont la première étape est celle de la production des

masques d'eau. Les campagnes de production des masques de surface en eau pour les sites Andalousie, Inde et Occitanie ont été réalisées sur le cluster HAL du CNES pour la période s'étalant de 2018 à 2020. La difficulté de récupération des données d'entrée optiques et radar a été un point sur lequel des efforts d'amélioration ont été apportés. La seconde étape de production a porté sur le calcul de relations HSV et leurs comparaisons en prenant comme MNTs d'entrée Copernicus 30 et SRTM sur les trois Super sites, et le MNT ALOS en donnée additionnelle sur l'Andalousie.

La dernière étape de production correspond à la génération des séries temporelles $S(t)$, $V(t)$ et $TR(t)$ et sur une période de trois ans. Les séries temporelles *in situ*, disponibles sur certains ouvrages, ont été ajoutées en entrée à des fins de référence pour la comparaison et l'évaluation des résultats estimés. Plusieurs rejeux de production ont été nécessaires dues aux évolutions successives de certaines séries temporelles *in situ*, ainsi qu'à de nouvelles fonctionnalités implémentées dans la chaîne.

La quatrième tâche correspond à l'**analyse et à la validation des résultats** issus des différentes étapes du processus. Dans un premier temps, les sorties d'extension des réservoirs, les masques multiple et single Surfwater ont été analysées sur le Super Site Andalousie, premier site traité. Pour cela, les sorties SURFWATER ont été comparées avec celles issues de la chaîne du SERTIT EXTRACTEO, et les surfaces dites de références générées via des lois H/S/V à partir des données H et S *in situ*. Un premier jeu de sortie Surfwater issu de la chaîne optique a été analysé (Config 1 de Surfwater). Devant les résultats peu satisfaisants, le CNES a proposé une adaptation/modification des paramètres SURFWATER, et un second jeu a été analysé (uniquement sur l'Andalousie, en dehors de l'analyse générique $S(t)$, $V(t)$, $TR(t)$). La qualité des résultats est fortement améliorée, mais on observe toutefois, sur les données optiques, encore un certain nombre d'anomalies.

D'autre part, dans près d'un tiers des cas, on observe un biais systématique entre les surfaces dérivées des données Sentinel-2 et des références. L'analyse des images et des extractions ne permet pas de penser qu'il y ait eu une omission importante lors de l'extraction des surfaces en eau par SURFWATER ou ExtractEO. Ce biais, de 7 à 12 %, pourrait être dû à un fonctionnement moins performant de la modélisation des surfaces « *in situ* » pour certains réservoirs, ou d'une modélisation du volume *in situ* erronée de la part des gestionnaires de ces réservoirs. Il pourrait être intéressant pour pallier ce problème de travailler non pas en termes de volume absolu, mais plutôt de variations de volume car ces 30 % des réservoirs Andalous ont des données *in situ* peu fiables, risquant ainsi de biaiser les validations finales.

Pour analyser la qualité des séries temporelles $S(t)$, $V(t)$, $Tr(t)$ à partir des rapports générés par DEM4Water et Surfwater_Postprocess, un outil (notebook) a été mis en place afin de générer automatiquement un tableau récapitulatif permettant d'évaluer plus simplement la qualité des séries temporelles de surface, de volume et de taux de remplissage.

6.2 Une première approche concluante

L'objectif du projet SCO STOCKWATER était d'établir le potentiel de généricité de la méthode, en étendant les traitements initialement réalisés en France à d'autres régions du monde. Dans ce cadre, nous avons eu l'opportunité d'étudier un panel de réservoirs aux formes très variées, dans des contextes topographiques qui l'étaient tout autant.

Au niveau de l'opérationnalité, le démonstrateur permet de mettre en place automatiquement un contexte pour chaque chaîne et inter-étape. L'utilisation en est simple et flexible, avec la possibilité de relancer très simplement à partir d'une étape sélectionnée si besoin.

Les temps de traitement ont été optimisés avec la parallélisation des processus dès que cela était possible pour un objectif de traitement de masse.

L'organisation en deux étapes aura permis, en première phase, de dimensionner et de valider la méthodologie afin de passer à la généralisation de l'approche retenue. Elle a permis de mettre en évidence un certain nombre de problèmes/limites à différents niveaux du processus, accès et qualité des données *in situ*, accès aux données images, qualité des extractions, etc.

Du point de vue des outils, on peut considérer le concept comme valide. Un démonstrateur de niveau « Proof of Concept » et des outils annexes ont été développés, mais des efforts supplémentaires sont nécessaires en termes de robustification de la méthode et de l'outil pour pouvoir atteindre pleinement le stade d'un véritable démonstrateur.

La mise en place de la base de données permet la sélection des réservoirs étudiés, la définition puis le remplissage par région d'étude et par réservoir d'une table attributaire assez complète, la récupération, l'analyse et la mise en forme des données *in situ* H-V, voire leurs corrections. Une fois ces informations rassemblées, la table attributaire a été construite, mais il a fallu de nombreux échanges pour arriver à la version définitive en termes de type de fichiers, de format et d'ordre des champs.

Les chaines Surfwater, DEM4Water et Surfwater_Postprocess ont été, d'une part, adaptée aux nouvelles données d'entrée fournies et, d'autre part, améliorées fonctionnellement. Au fur et à mesure de l'avancement du projet, l'intégration et l'interopérabilité de ces différents composants au sein du démonstrateur ont été perfectionnées.

De bonnes performances ont été constatées sur les séries temporelles de taux de remplissage, donnant une variation relative. Cependant de moins bons résultats ont été observés sur les séries temporelles de volumes absolus.

Enfin, un portail web SCO a été mis en place permettant la restitution de ces résultats et l'accès au suivi des réservoirs. L'interface développée dans le cadre du projet répond aux besoins génériques de la communauté rassemblée autour du projet. Cette solution assure une adaptabilité et une portabilité à moindre coût aux utilisateurs, considérant la perspective prochaine de plateformes de services de diffusion aux utilisateurs, portées par le CNES.

Les premiers retours des partenaires européens ayant testé la plateforme sont positifs. En revanche, ce premier outil obéit à une logique de généricité, et doit être adapté à des usages et des utilisateurs spécifiques.

6.3 Des axes d'amélioration pour une version opérationnelle

La phase complémentaire a permis de robustifier l'ensemble des composants de façon suffisamment stable pour qu'une production end-to-end soit possible avec une intervention humaine limitée aux reprises de calculs liées à l'affinage des résultats. Les efforts consentis lors de cette phase complémentaire ont permis de passer à un stade TRL 4 confirmé.

Cependant, il n'est pas encore possible de considérer que le niveau pré-opérationnel est atteint. Il reste nécessaire de poursuivre les efforts d'amélioration, tant au niveau de l'acquisition et de la préparation des données que de la stabilité et de la généricité des résultats. Ces efforts devront se traduire par des améliorations logicielles sur les trois composants principaux Surfwater, DEM4water et Surfwater_Postprocess, ainsi que sur leurs procédures de validation.

6.3.1 Fluidifier la manipulation des données

Dès le début de l'étude, nous avons été confrontés à des problèmes d'acquisition, d'hétérogénéité et de compatibilité des données, avec des anomalies constatées à différents stades du processus. D'importants efforts ont été consentis pour circonvier l'essentiel de ces problèmes avec un succès certain, mais ces anomalies ont rendu les évaluations qualitatives plus complexes, avec une démarche de vérification et de préparation qui s'est révélée extrêmement chronophage.

Dans son état actuel, STOCKWATER répond au besoin initial et a le grand mérite de proposer une première approche d'outil de gestion des ressources en eau à grande échelle qui n'a pas encore d'équivalent en Europe.

Cependant, cette situation n'est **pas compatible avec une exigence de disponibilité des données en temps contraint** et nécessitera à l'avenir de prendre des mesures techniques, organisationnelles et de standardisation. Ces mesures sont nécessaires pour rendre l'outil facile d'utilisation et avec des résultats pertinents afin de préserver son attractivité, **qui reste indéniable au vu des enjeux économiques, environnementaux et de sécurité**.

Il faut cependant souligner que cette standardisation ne pourra être acquise facilement à partir du moment où la communauté visée est globale, couvrant des aires éco-climatiques diverses, avec des usages et des problématiques nationales, voire locales, de collationnement et d'accessibilité des données (même en France, les données de certaines retenues et barrages ne sont pas accessibles au public, ni même à la communauté scientifique).

Si des mesures organisationnelles de base peuvent être envisagées, le futur système se devra sans doute d'être, d'une part, aussi **agnostique** que possible, de façon à pouvoir fonctionner avec des données d'entrée alternatives (images, MNT, données de support et de référence), et d'autre part facilement accessible à tous les utilisateurs (open source ?) de façon à permettre son utilisation par des scientifiques ou des utilisateurs dans des pays, régions ou domaines d'activité pour lesquels les données ne peuvent être diffusées à la communauté.

6.3.1.1 Des données d'entrée plus robustes

Les capacités de l'outil, dans une perspective d'aide à la décision, restent en premier lieu limitées par la difficulté de disposer de données homogènes et directement utilisables. **Cette problématique d'acquisition et de préparation des données d'entrée constitue un premier goulot d'étranglement de la démarche.**

La mise en place d'une structure unique de fichier **a permis de résoudre en grande partie le problème de l'hétérogénéité des sources**, mais dans la situation actuelle un nombre important d'actions manuelles sont toujours nécessaires lors de la phase d'acquisition et de préparation des données.

Dans tous les cas, considérant le contexte particulier du domaine dans lequel va évoluer STOCKWATER, une démarche préalable de contrôle et de préparation des données reste incontournable. Cependant cette démarche reste pour l'instant assez longue, et ne permet pas en l'état d'envisager une **approche ARD** (Analysis Ready Data), et donc l'utilisation d'un **Datacube** (le composant eodag est déjà une première approche de cette organisation mais il ne fonctionne pas de façon efficace dans les conditions actuelles, ce qui en réduit l'intérêt). Il est donc indispensable d'envisager une réflexion quant à une automatisation du process d'acquisition et de préparation des données transparente pour l'utilisateur.

6.3.1.2 La problématique des données de support et de référence

La même remarque s'applique également sur les données de soutien. **Un deuxième goulot d'étranglement de la démarche se situe au niveau de la mise en forme et de l'analyse des données *in situ*.** Des incohérences ont été parfois rencontrées dans les données *in situ*, avec des problèmes d'adéquation entre les valeurs des volumes et les élévations reportées, des données manquantes, etc.

Cette situation ne doit pas être considérée comme anormale dans le contexte de l'étude. Il s'agissait d'une première approche, utilisant des données provenant de sources ou d'équipes différentes, avec des jeux de données souvent réalisés dans des conditions spécifiques, ou en fonction d'autres besoins.

Cependant, l'importance des efforts consentis pour compléter, harmoniser et formater les jeux de données, ainsi que développer des outils permettant de « lisser » la qualité des données n'est pas non plus compatible avec une évolution vers une étape opérationnelle. Il sera donc nécessaire à l'avenir de recourir à **des mesures organisationnelles** pour que les jeux de données fournis répondent en amont à des standards communs à l'ensemble des contributeurs.

6.3.1.3 Les lois *in situ*

La dernière étape de la mise en forme des données a porté sur la génération des lois *in situ*. Pour cela, plusieurs évolutions de l'outil « timeseries_tool.py » du CNES ont été effectuées, que ce soit par le CNES ou le SERTIT. De ce fait, les lois *in situ* ont dû être générées un certain nombre de fois. Si l'outil fonctionne aujourd'hui, **des améliorations sont encore nécessaires** afin d'assurer un bon fonctionnement à l'avenir.

6.3.1.4 Les MNTs

S'il a été relativement aisé de converger sur des MNTs génériques dans le périmètre de l'étude, il reste évident qu'il n'est pas possible de pouvoir disposer des mêmes sources de MNT, dans des conditions de qualité et de précision homogène, selon les régions ou les pays. C'est le cas notamment de pays pour lesquels la diffusion des données d'élévation n'est pas libre, ou soumise à des contraintes de

diffusion importantes (c'est d'ailleurs le cas en France pour certains sites). Il n'est pas non plus possible d'avoir la garantie de disposer systématiquement de données présentant des niveaux de précision et de qualité identiques.

Dans ce cas également, il sera nécessaire de rendre le système aussi agnostique que possible en termes de MNT, dans le but de permettre aux utilisateurs confrontés à ces situations de pouvoir disposer d'un MNT générique, mais pouvant être enrichi ou remplacé partiellement ou localement par des données locales. Cette fonctionnalité existe déjà dans d'autres projets du CNES et l'évolution du projet pourra sans doute s'en inspirer.

6.3.2 Vers une chaîne opérationnelle

Afin d'atteindre réellement un niveau TRL 5 puis un stade pré-opérationnel, les efforts de robustification et d'adaptation des différents composants doivent être poursuivis, ainsi qu'une stabilisation et une meilleure accessibilité de la documentation associée.

Si l'hypothèse d'une diffusion facilitée au profit de la communauté se confirme (licence élargie ou open source), il est indispensable de mener des actions d'audit de code, accompagnées d'éventuelles actions correctives avant toute diffusion.

Le développement du démonstrateur a nécessité l'implémentation d'améliorations sur les chaînes Surfwater, DEM4water et Surfwater-Postprocess afin de s'adapter aux nouvelles données d'entrée et de proposer des fonctionnalités répondant au volume de production requis, à la variabilité des données, mais aussi à l'analyse de résultats. Ces efforts devront être poursuivis en fonction de l'évolution de la méthode et de l'outil jusqu'à la réalisation de la version opérationnelle.

En ce qui concerne Surfwater, il reste des incohérences à investiguer entre l'optique et le radar sur certains sites, en particulier sur le Brésil. En revanche, on constate de bons résultats sur l'Andalousie et la Tunisie. Une telle différence de résultats a déjà été rencontrée sur d'autres projets (IOTA2 au Canada et en Afrique par exemple) et peut sans doute s'expliquer par des phénomènes dus aux différents environnement éco-climatiques et des conditions d'éclairement des cibles différentes. Ces difficultés peuvent certainement se résoudre par une démarche adaptée au contexte (comme pour IOTA2), mais il faudra sans doute accepter un certain niveau de variabilité des résultats par rapport à la méthode de base en fonction des régions observées.

Concernant la chaîne DEM4water, dans la lignée des améliorations entamées lors de la Phase II, il reste à poursuivre un travail d'affinage des lignes de coupure, de précision de localisation de pied de barrage et d'ajustement de la loi. Un outil de validation automatique des modèles basé sur un rapport de statistiques permettrait également d'alléger la tâche métier de validation réalisée par l'opérateur. Une démarche d'amélioration de l'outil est en cours actuellement à CS GROUP, pilotée par le CNES.

Au niveau des outils, il serait intéressant d'envisager une méthode d'ingestion automatique des nouvelles séries temporelles sur la plateforme web SCO, la mise à jour se faisant actuellement via une procédure opérateur.

En ce qui concerne l'opérationnalité, **des travaux sont en cours sur un démonstrateur de service d'aide à la décision pour le gouvernement français**, avec 340 réservoirs actuellement étudiés et une perspective de 500 barrages au total. Nous attendons beaucoup du retour d'expérience de cette action pour l'évolution du futur outil opérationnel.

6.3.3 Une phase de validation à stabiliser

L'important travail d'investigation, d'analyse et de validation entrepris par le SERTIT a été très bénéfique dans le périmètre de l'étude, et plusieurs sources de difficultés et d'erreur ont été identifiées et corrigées grâce à ce travail de longue haleine. Cependant, de telles démarches « poussées » ne peuvent être envisagées dans le cadre d'une utilisation opérationnelle, ni même à un stade pré-opérationnel.

Cependant l'importance de l'effort fourni, s'il est pertinent dans le contexte d'un Proof of Concept, **est complètement démesuré dans la perspective d'une version opérationnelle**, et l'on ne peut exiger un niveau de compétence élevé à l'ensemble des futurs opérateurs et utilisateurs. La démarche d'analyse et de validation devra être facile à mettre en œuvre, accessible à des opérateurs ou

utilisateurs non experts.

Un effort de stabilisation et de standardisation des démarches d'analyse et de validation, en amont et en aval de la production doit donc être entrepris. Il faut cependant souligner que ces efforts s'inscrivent en parallèle d'autres actions de stabilisation et organisationnelles évoquées dans les paragraphes précédents.

6.3.4 Une architecture et une interface adaptées à une plus large communauté d'utilisateurs

Dès la conception du projet, il a été décidé pour la réalisation du tableau de bord et des fonctionnalités associées, de s'appuyer sur les mêmes technologies que celles employées pour le développement de la future plateforme HYSOPE II/Hydroweb-Next du CNES, qui sera dédiée à la diffusion des données et de produits à vocation hydrologique. Ce choix assure d'emblée la compatibilité du tableau de bord STOCKWATER avec la future plateforme, ou du moins son adaptation à moindre coût. Cela devrait également faciliter la synergie des produits avec ceux issus d'autres chaînes de traitement ou sources de données (notamment SWOT), tout en ouvrant, dans un même temps, des perspectives d'évolution ou d'interopérabilité de l'outil. La figure 96 ci-dessous évoque ce qui pourrait être la configuration type de l'outil STOCKWATER au sein de la future architecture de l'ensemble. Cette vision d'architecture reprend des démarches similaires déjà mises en œuvre ou envisagées dans le cadre d'autres projets du CNES, par exemple IOTA2.

Cependant, pour rebondir sur certaines considérations abordées aux chapitres précédents, **il nous semble utile de proposer une approche plus souple de la mise à disposition de l'outil, dans la perspective de pouvoir atteindre une plus grande communauté d'utilisateurs.** En effet, comme nous l'avons évoqué, la communauté d'utilisateurs potentiels n'apparaît pas comme homogène tant en termes de familiarisation avec les outils, d'expertise dans la manipulation des données, d'usage et d'accès aux données. Il nous semble donc nécessaire de proposer une « offre commerciale » plus diversifiée :

- ✓ Une **interface accessible via la plateforme HYSOPE II**, au profit des utilisateurs opérationnels ou non experts des produits,
- ✓ Comme c'est le cas actuellement, proposer la manipulation des données et de l'outil au travers d'un **interface Jupyter notebook** au profit d'un public expert. Cette possibilité existe déjà pour des chaînes de traitements de produits hydrologiques et les fonctionnalités associées (IOTA2, SWOT),
- ✓ Enfin, **une version « stand alone »**, pouvant être utilisée de façon indépendante de la future plateforme du CNES. Cette version pourrait être déployée par des utilisateurs pour l'utilisation de leurs propres données, ou nécessitant des données de références ou de soutien **dont l'accès est réglementé ou restreint, ou qui ne peuvent accéder aux données mises en ligne par le CNES, l'ESA ou d'autres Agences.** Bien sûr, une telle option sous-entend que le CNES mette en œuvre une politique de licence ou de libération de logiciel qui permette de le faire (des actions en ce sens sont en cours). L'avantage essentiel de cette possibilité est de permettre aux méthodologies employées de **converger vers des approches similaires et d'obtenir des résultats pouvant être comparables.**



7 Références

- 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, n° 368, p. 829–836 ([DOI 10.2307/2286407](https://doi.org/10.2307/2286407), [JSTOR 2286407](https://www.jstor.org/stable/2286407))
- Savitzky, A. & Golay, M. J. E. “*Smoothing and Differentiation of Data by Simplified Least Squares Procedures*”, *Analytical Chemistry*, vol. 8, n° 36, 1964, p. 1627–1639 ([DOI 10.1021/ac60214a047](https://doi.org/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-access global digital elevation model, *Frontiers 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 Lac pour les Données SWOT HR, , Final report 7th December 2018, 122 pages.

8 Glossaire

ALCD	Active Learning for Cloud Detection
ALOS	Advanced Land Observing Datellite
API	Application Programming Interface
BD	Base de Données
CGTD HYSOPE	Centre de Gestion et de Traitement des Données HYSOPE
CIRCA-2015	Global Lake database, de Sheng
CNES	Centre National d'Étude Spatiale
CNRS	Centre National de la Recherche Scientifique
DOM	Départements d'Outre-Mer
DREAL	Direction Régionale de l'Environnement, de l'Aménagement et du Logement
EGM	Earth Gravitational Model
EODAG	Earth Observation Data Access Gateway
ERS	European Remotes Sensing satellite
ESA	European Space Agency
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 <i>Data</i> 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
IOTA²	Infrastructure pour l'Occupation des sols par Traitement Automatique incorporant les Toolbox Applications
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
MNT/DEM	Modèle Numérique de Terrain/Digital Elevation Model
MNS/DSM	Modèle Numérique de Surface/Digital Surface Model
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
PHR	Pleiades High Resolution

OTB	ORFEO Toolbox
PCA	Principal Component Analysis
PoC	Proof of Concept/Preuve de Concept
QGIS	Quantum Geographic Information System
R&D	Recherche et Développement
RGE	Référentiel à Grande Echelle
RMSE	Root Mean Square Error
ROI	Region of Interest (zone d'intérêt dans une scène, une image ou un produit)
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
SIG	Système d'Information Géographique
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
TRL	Technology Readiness Level
WGS	World Geodetic System
WI	Water Index

9 Annexe A : Analyse des séries temporelles $S(t)$, $V(t)$, $Tr(t)$ selon les quantiles 75 et 90

Tableau 26 : Quantile 75 - Andalousie

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
2160003673	Cuevas de Almanzora	39.273590	69.634367	25.262498	40.029554	22.702556	18.521919	30.566628
2160004183	Beznar	7.205803	7.232193	13.523263	25.633294	23.667781	16.265936	15.005196
2160004253	Rules	3.473714	7.824646	12.416312	48.645975	48.602602	32.197463	33.247820
2160004373	Guadalhorce	nan	14.948490	18.762001	32.586803	28.631116	12.783825	8.825927
2160004383	Guadaleba	7.770141	11.613687	17.396984	49.287851	43.025125	15.972954	10.028230
2160004403	Conde de Guadalhorce	12.626295	17.237102	23.169760	23.036856	18.564215	19.083749	9.320334
2160004413	La Vinuela	2.381152	9.982267	8.060661	17.376357	9.009762	9.800234	5.157789
2160004532	Limonero	4.718030	11.395863	13.534905	35.151046	39.146577	23.596322	25.661687
2160004623	Guadarranque	6.321110	15.250053	15.119430	36.236985	24.969292	16.205065	10.053247
2160004643	Charco Redondo	6.924094	19.268207	6.459137	23.953802	15.614171	16.442718	12.860564
2160004663	La Concepcion	9.723020	61.895135	33.280306	73.596325	23.529897	56.824201	19.261396
2310000173	San Rafael de Navallana	2.799966	10.376287	11.816731	25.482964	23.655597	17.831623	16.418404
2310018233	Los Hurones	24.467642	22.365676	28.000224	14.554464	11.180160	21.950395	17.876729
2310018272	Almodovar	99.484707	77.824589	68.981653	99.852396	99.824320	13.973426	10.324141
2310018393	Zahara	1.424229	6.666370	3.521626	27.034210	26.378593	16.647053	16.257869
2310018403	Bornos	17.176650	9.456584	23.453211	6.040597	11.555537	16.019051	15.244347
2310018473	Barbate	13.846101	10.875107	9.589038	22.324056	18.000645	4.581563	5.780386
2310018492	Arcos de la Frontera	nan	8.308328	7.286778	41.971351	49.105307	nan	nan
2310020153	Guadalcacín 2	11.383586	8.226538	3.193711	21.492027	15.823888	8.227413	4.286172
2310020223	Celemin	3.023574	20.084801	33.382165	14.111632	15.705840	6.908483	7.168804
2310020933	Tranco de Beas	6.357065	14.548354	19.546932	10.923267	12.389947	3.658432	4.202206
2310022062	Arenoso	nan	11.406555	15.590356	50.936348	54.012014	31.107953	33.424803
2310022743	Yeguas	4.898721	16.886822	15.777993	28.059240	27.226910	17.911251	16.142957
2310022903	Los Bermejales	7.020355	99.775133	12.554680	99.788910	16.121506	96.781198	11.060387
2310022963	Giribaile	13.827539	6.727916	10.676500	15.749389	17.096467	3.805143	5.898551
2310023523	Jose Toran	28.336494	28.126756	30.311665	22.180890	9.222132	30.391921	20.626510
2310023943	Vadomojon	12.321373	14.151773	7.517075	31.635119	29.859172	14.213474	12.766531
2310024073	Iznajar	5.648402	3.665308	2.784380	12.443092	10.785467	10.754651	9.742225
2310024113	Puebla de Cazalla	0.917271	12.332033	6.314248	35.984010	33.047369	22.714697	20.574692
2310024253	La Brena II	2.888674	63.444991	64.706405	18.614576	13.606263	15.398829	12.130331
2310027322	Los Melonares	78.990601	12.706291	12.401708	82.595984	82.572793	16.751845	16.414151
2310027683	Puente Nuevo	15.968427	5.286283	5.964485	7.771363	10.964043	8.024026	5.144806

Tableau 27 : Résumé résultats quantile 75 - Nombre réservoirs par classe - Andalousie

	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
<= 10 %	17	10	10	2	2	7	10
> 10 & <=25 %	8	16	15	12	18	19	17
>= 25 %	4	6	7	18	12	5	4

Tableau 28 : Quantile 90 - Andalousie

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
2160003673	Cuevas de Almanzora	39.273590	75.771921	39.223238	44.936581	29.589767	22.684926	34.415751
2160004183	Beznar	7.205803	10.815232	18.815587	29.945832	25.895014	19.839091	16.582658
2160004253	Rules	3.473714	11.144111	30.252100	50.497409	51.013377	33.993931	34.413085
2160004373	Guadalhorce	nan	24.130576	28.942023	49.282490	31.999836	29.142107	12.367770
2160004383	Guadaleba	7.770141	25.425723	22.288535	55.253089	46.803587	23.472038	12.765884
2160004403	Conde de Guadalhorce	12.626295	83.376343	36.943824	84.541274	33.792780	65.993266	12.173347
2160004413	La Vinuela	2.381152	66.998032	18.195274	69.708826	16.278979	52.312019	11.964875
2160004532	Limonero	4.718092	17.869724	19.225691	40.192569	42.323200	27.143030	29.112492
2160004623	Guadarranque	6.321110	80.218014	18.674301	85.116698	28.258785	58.931263	12.786906
2160004643	Charco Redondo	6.924094	76.853395	9.434289	78.196801	19.134945	65.107384	14.031351
2160004663	La Concepcion	9.723020	99.739653	48.122381	99.819739	28.119442	97.753385	21.708472
2310000173	San Rafael de Navallana	2.799966	30.044952	17.565087	41.968434	28.632219	32.651679	20.265848
2310018233	Los Hurones	24.467642	33.737287	36.884353	46.511112	18.032866	42.140014	20.403596
2310018272	Almodovar	99.484707	98.046271	74.574541	99.986909	99.855996	63.900304	13.552896
2310018393	Zahara	1.424229	40.314819	5.374990	53.329686	27.918423	36.537894	17.314441
2310018403	Bornos	17.176650	11.135477	36.393630	10.170110	16.982774	17.768505	17.632030
2310018473	Barbate	13.846101	32.801623	16.411680	47.326230	20.776882	22.257307	9.90167
2310018492	Arco de la Frontera	nan	20.225928	9.023298	49.513609	50.087414	nan	nan
2310020153	Guadalcacin 2	11.383586	46.227751	4.803087	54.127547	17.842818	34.588356	5.965327
2310020223	Celemin	3.023874	36.185103	50.342123	47.623230	23.962165	26.473676	10.959109
2310020933	Tranco de Beas	8.357065	20.997205	25.928364	16.629114	19.339062	5.666873	6.176513
2310022062	Arenoso	nan	40.758960	21.729639	67.237305	57.356813	53.306185	36.696280
2310022743	Yeguas	4.899721	37.870460	21.037516	46.222171	31.771467	33.098369	19.844605
2310022903	Los Bermejales	7.020255	99.995625	20.225275	99.996081	33.125222	99.725238	21.065150
2310022963	Giniballe	13.827539	10.028976	15.809078	17.844154	30.204752	5.621228	9.787963
2310023523	Jose Toran	28.336484	36.993440	36.715805	49.514527	13.851818	53.448524	24.035367
2310023943	Vadomojon	12.321373	18.618349	9.040777	38.104209	32.231528	19.002901	14.880000
2310024073	Iznajar	5.648402	5.402381	4.190654	14.618054	12.158534	13.057741	10.704050
2310024113	Puebla de Cazalla	0.917271	21.078934	9.589265	49.879616	35.271886	32.568818	22.343265
2310024253	La Brena II	2.688674	66.838629	68.745831	47.448927	16.409647	36.067528	13.984545
2310027322	Los Melonares	78.990601	43.834480	15.423247	90.450926	83.168623	70.408635	19.030635
2310027683	Puente Nuevo	15.968427	23.956419	8.581358	21.936482	13.476133	22.415963	6.812095

Tableau 29 : Résumé résultats quantile 90 - Nombre réservoirs par classe - Andalousie

	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
<= 10 %	17	1	8	0	0	2	5
> 10 & <=25 %	8	11	12	5	12	8	22
>= 25 %	4	20	12	27	20	21	4

Tableau 30 : Quantile 75 – Occitanie

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
2160026973	Salagou	90.151847	2.749054	3.790464	92.463121	92.763729	4.155655	5.660208
2160028013	Matemale	34.228502	89.981283	18.931893	88.400480	39.944874	85.463094	10.011978
2160029873	Villeneuve la Raho	99.957016	7.913326	13.802688	99.915531	99.921917	9.350958	7.274553
2160029943	Puyvalador	19.861078	67.548057	21.604671	67.990304	35.276138	64.321818	10.533315
2160030123	Vinca	77.570885	28.763371	54.220678	84.566401	84.898918	25.525163	28.971429
2160030183	Avene	6.105639	14.898682	17.383973	50.404032	66.564805	29.221380	47.543959
2160030553	Agly	5.711032	22.804493	16.728623	25.595276	31.864895	19.562341	24.271364
2320028893	Pareloup	33.081892	11.105148	16.511033	14.990158	17.284202	10.767114	12.040798
2320028933	Saint Geraud	nan	25.396988	19.549292	37.383647	49.397532	33.977126	27.625862
2320030233	Montbel	87.039512	19.907126	37.831722	83.986413	85.607431	27.299529	33.014135
2320030823	Cammazes	3.318017	15.815348	18.867381	46.620791	64.396390	35.468374	48.190286
2320031293	Laparan	69.055075	99.986913	57.093849	99.819699	69.738884	99.563273	56.689410
2320031303	Pla de Soulcem	67.837554	78.689393	40.287690	81.577290	43.447414	35.100945	35.586183
2320038733	La Gimone	7.438983	10.074721	11.810930	23.119811	28.196943	11.769807	14.308166
2320039133	Astarac	13.847274	34.643123	42.797443	37.352295	44.193718	19.311219	24.297113

Tableau 31 : Résumé résultats quantile 75 - Nombre réservoirs par classe – Occitanie

	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
<= 10%	4	2	1	0	0	2	2
> 10 & <= 25%	2	6	9	2	1	4	6
>= 25%	8	7	5	13	14	9	7

Tableau 32 : Quantile 90 – Occitanie

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
2160026973	Salagou	90.151847	3.352496	5.781294	92.560483	92.962589	4.586620	6.745676
2160028013	Matemale	34.228502	99.946688	71.320239	99.938276	79.917591	99.714950	45.458089
2160029873	Villeneuve la Raho	99.957016	12.566379	20.928797	99.919273	99.924129	12.277384	9.045657
2160029943	Puyvalador	19.861078	99.879951	51.229960	99.919695	45.103067	99.214568	20.613635
2160030123	Vinca	77.570885	47.942109	90.244952	85.441522	85.837309	27.723674	43.686622
2160030183	Avene	6.105639	20.177243	30.706810	54.856809	68.860863	34.430735	50.208933
2160030553	Agly	5.711032	32.348816	22.578932	30.009489	36.019948	22.786138	27.544790
2320028893	Pareloup	33.081892	12.540693	23.756103	16.721463	24.591535	13.287724	13.832924
2320028933	Saint Geraud	nan	37.421115	26.793436	48.564990	52.993178	41.319003	38.553009
2320030233	Montbel	87.039512	26.849787	72.488060	85.063356	86.178275	41.200159	45.603180
2320030823	Cammazes	3.318017	25.064148	34.625381	52.147801	72.696714	39.784307	55.627562
2320031293	Laparan	69.055075	102.616697	89.460504	99.997994	90.789949	100.000000	70.269088
2320031303	Pla de Soulcem	67.837554	99.992787	58.470918	99.992625	54.613805	94.978014	45.263941
2320038733	La Gimone	7.438983	15.357785	17.010273	69.742773	71.843788	45.613377	47.971406
2320039133	Astarac	13.847274	44.660640	53.401296	44.370157	46.768434	24.310680	26.362197

Tableau 33 : Résumé résultats quantile 90 - Nombre réservoirs par classe - Occitanie

	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

Tableau 34: Quantile 75 – Inde

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
4530212043	Osman Sagar	115.957147	inf	549.813378	inf	174.348741	43.668839	35.026014
4530212143	Himayat Sagar	44.017308	27.256249	36.725319	48.341240	46.024782	26.664732	19.535531
4530347283	Nagarjuna Sagar	89.965929	870.579259	553.581415	2120.444824	1393.448521	29.692703	35.739524

Tableau 35 : Quantile 90 – Inde

ID_SWOT	Dam_name	V(S)- (%)	TR_MO2- (%)	TR_MR2- (%)	Volume_MO2- (%)	Volume_MR2- (%)	Surface_MO2- (%)	Surface_MR2- (%)
4530212043	Osman Sagar	115.957147	inf	inf	inf	inf	55.285021	45.896479
4530212143	Himayat Sagar	44.017308	27.968040	45.740818	50.326562	51.672850	27.975620	24.231737
4530347283	Nagarjuna Sagar	89.965929	1165.226218	1608.183781	2794.524049	3748.929325	35.437006	45.709367

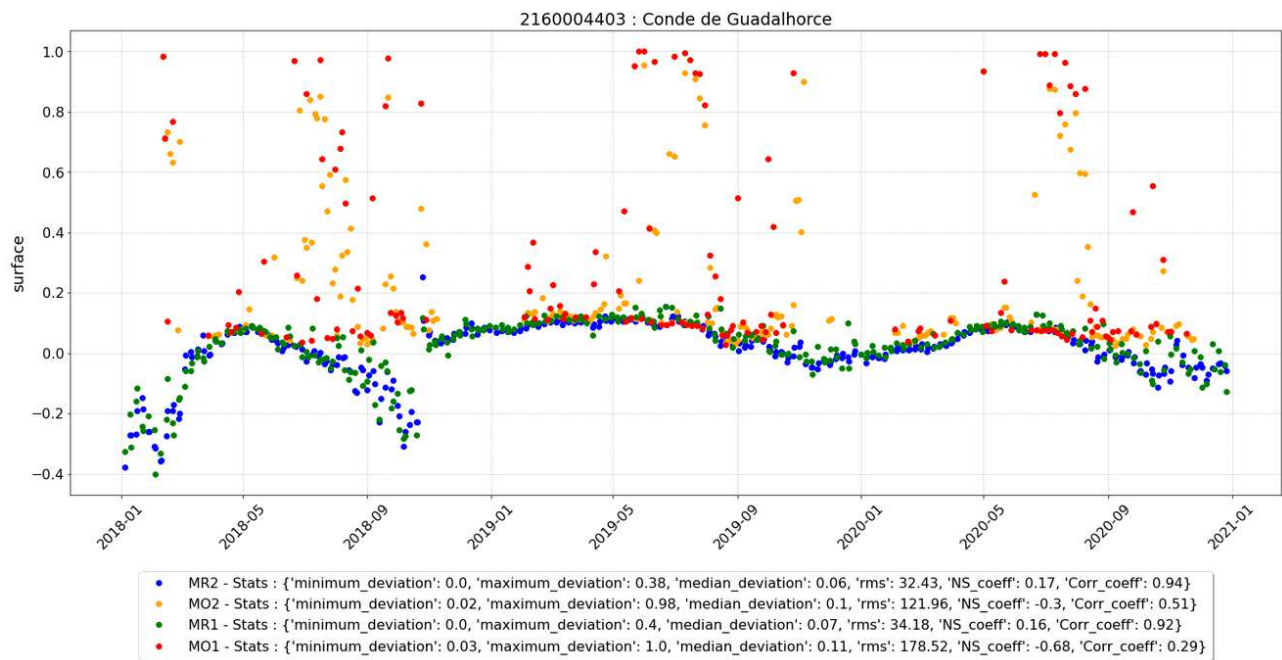


Figure 97 : Série temporelle des écarts relatifs absolus – Conde de Guadalhorce (Andalousie)

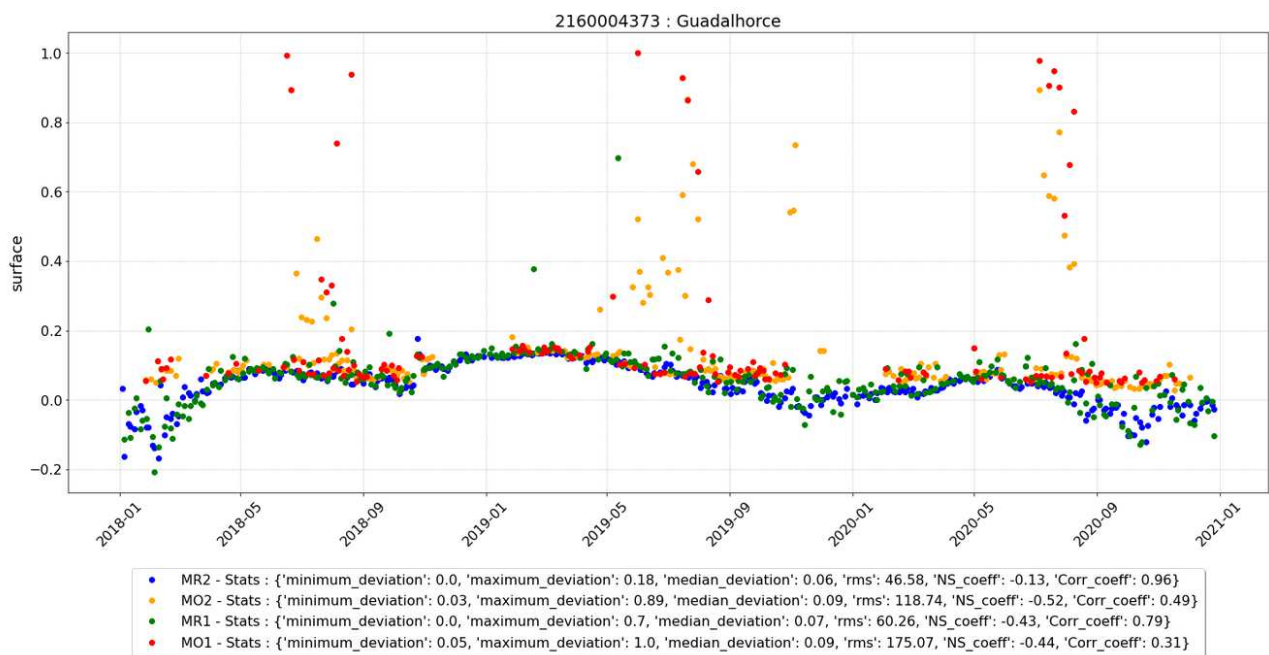


Figure 98 : Série temporelle des écarts relatifs absolus – Guadalhorce (Andalousie)

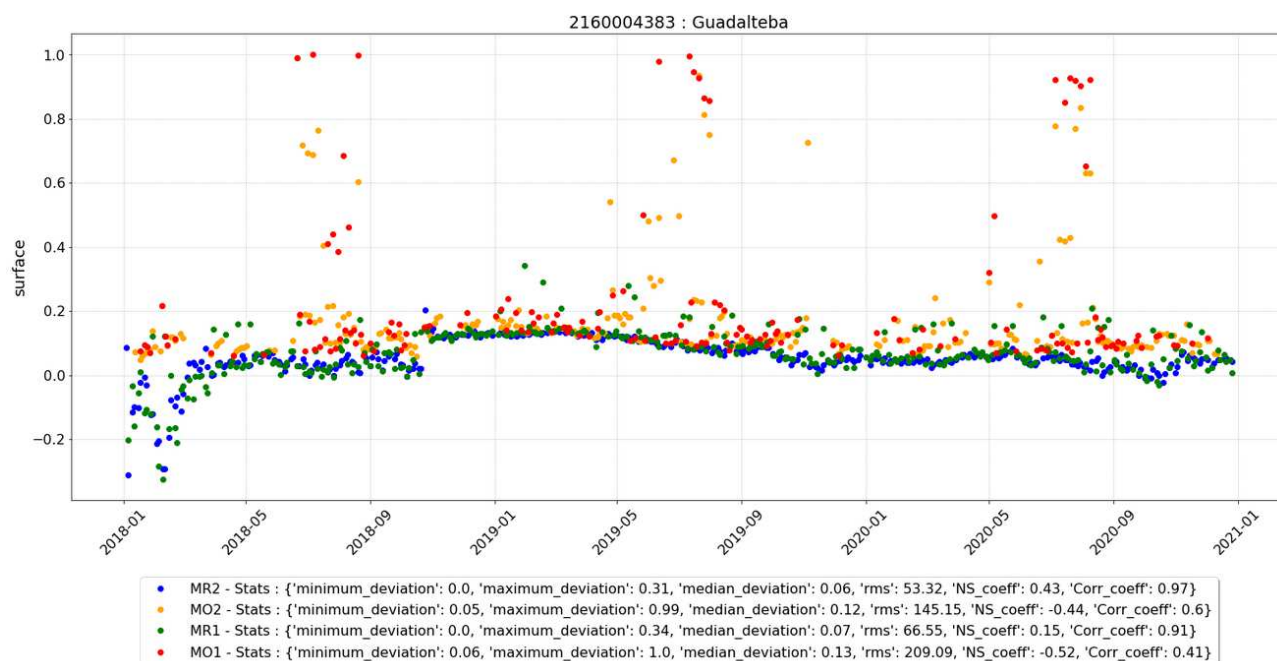
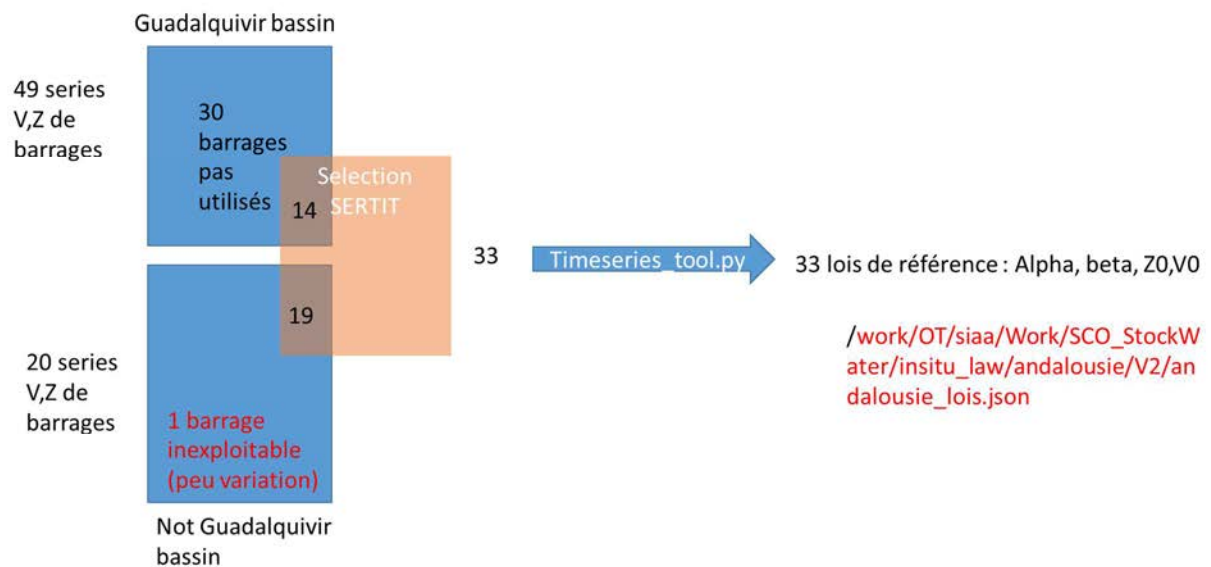


Figure 99 : Série temporelle des écarts relatifs absolus – Rules (Andalousie)

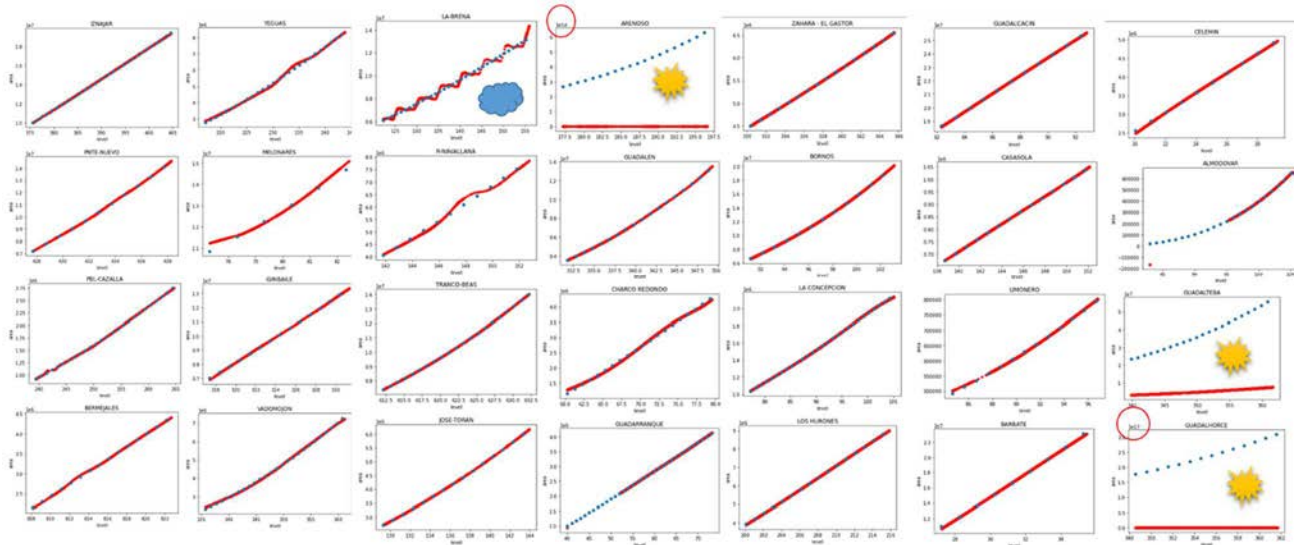
10 Annexe B : Analyse des lois REF

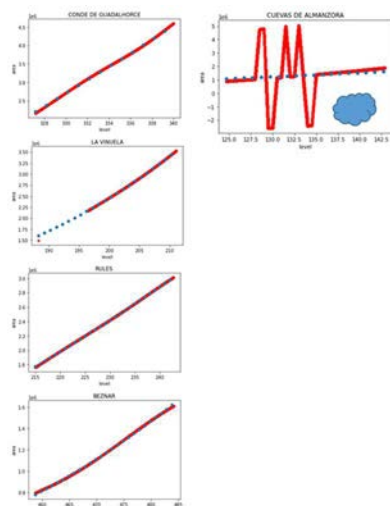
10.1 Analyse des lois ref vs *in situ*

Andalousie Reference Data



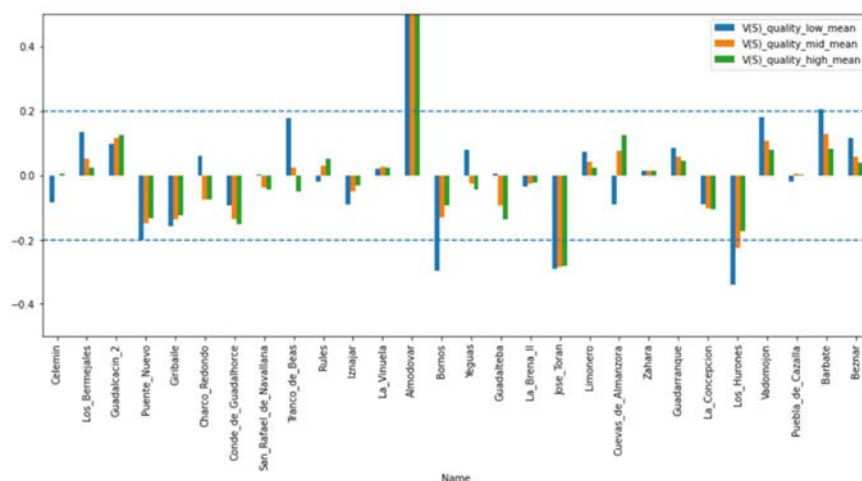
S(Z) Reference Lois vs Insitu Series





10.2 Analyse des lois ref vs dem4water

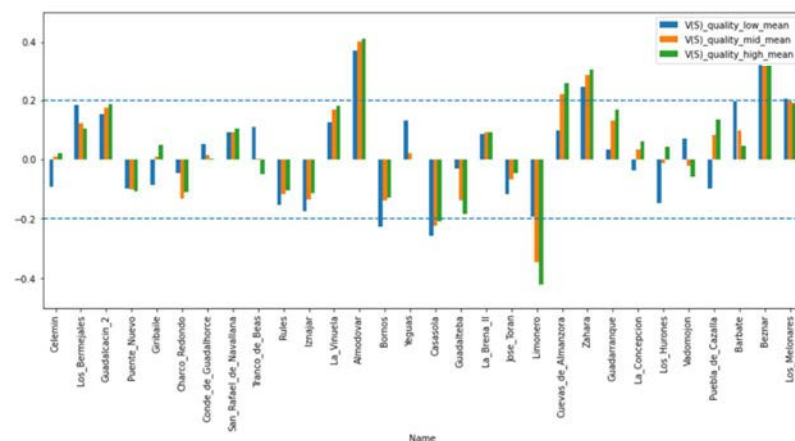
V(S) - Copernicus



28
Comparaisons

5 non
convergences
DEM4WATER à
vérifier

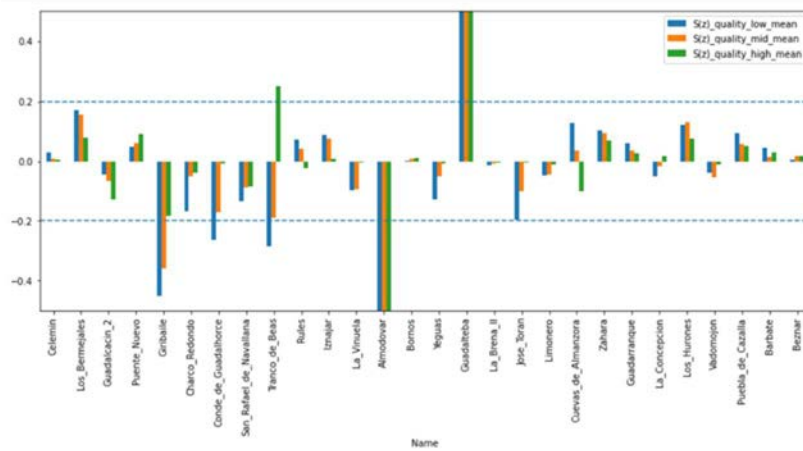
V(S) - SRTM



30
Comparaisons

3 non
convergences
DEM4WATER à
vérifier

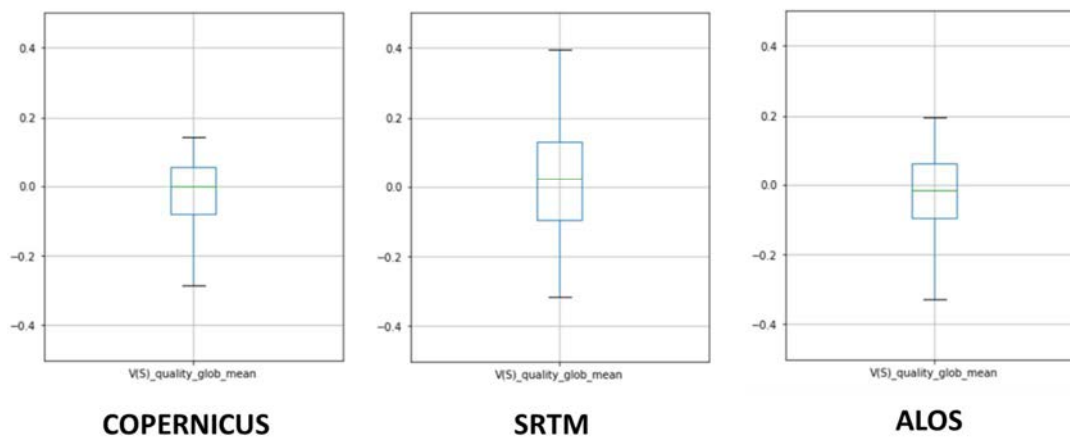
V(S) - ALOS



31
Comparaisons

2 non
convergences
DEM4WATER à
vérifier

Comparison of V(S) qualities



Mean of Absolute Relative Error V(S)

HIGH waters

Copernicus= 0.1101039174972788
SRTM= 0.14027042572961584
ALOS= 0.11366542392228193

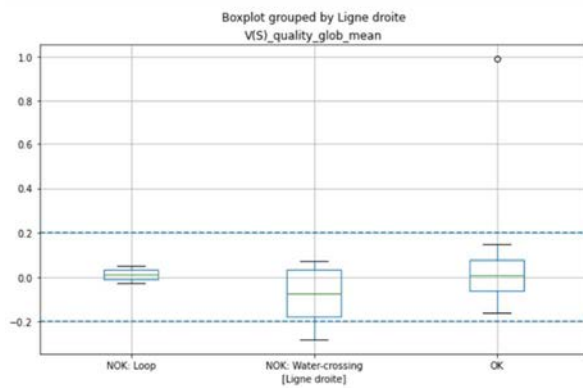
MID waters

Copernicus= 0.11414012125600383
SRTM= 0.13013243490559448
ALOS= 0.11259706488112078

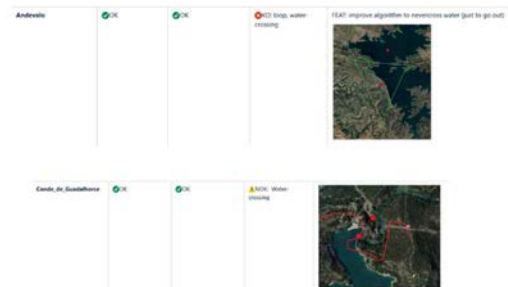
LOW waters

Copernicus= 0.11414012125600383
SRTM= 0.13013243490559448
ALOS= 0.11259706488112078

V(S) Copernicus – Mean relative Error vs dem4water errors



Error examples



V(S) Copernicus – Mean relative Error vs Z0 estimation

